A Two-Step Approach to Blindly Infer Room Geometries

Jason Filos, Emanuël A. P. Habets and Patrick A. Naylor†
Department of Electrical and Electronic Engineering, Imperial College London, UK
Email: {jason.filos, e.habets, p.naylor}@imperial.ac.uk

***Abstract*—An approach to blindly infer the geometry of an acoustic enclosure by estimating the location of reflective surfaces based on at least three channel estimates and relative microphone positions is presented. While blind system identification processes can be used to estimate the impulse responses of a single-input-multiple-output system, the propagation time of the direct sound cannot be directly observed. Because this propagation time is essential to infer the location of reflective surfaces, we propose a two-step approach. As a first step the time differences of arrival of the direct-paths are used to estimate the propagation time of the direct sound from the source to a reference microphone. Subsequently, the propagation times of all other arrivals can be inferred. In the second step the propagation times associated with the path of each first-order reflection are constrained to estimate the location of the reflectors and hence infer the geometry of the room. Preliminary results obtained using simulated acoustic impulse responses show that the proposed approach can be used to infer the room geometry when the propagation times of the first-order reflections can be identified.**

I. Introduction

Recently, the problem of reconstructing the geometry of an acoustic enclosure using microphone arrays has become an active area of research [1], [2], [3]. Knowledge gained about the acoustic environment, such as the location of sound reflectors, can be advantageous for applications such as sound source localization [1], [4] and adaptive echo cancellation by assisting in tracking environment changes and helping the initialization of such algorithms.

Recently Antonacci et al. [3] proposed a method to reconstruct the geometry of an environment through successive acquisitions of a controlled emission. Specifically, a loudspeaker is moved along a pre-defined trajectory. For each position a single-input-single-ouput system is identified and used to extract the time of arrivals (TOAs) of reflected signals. This information is subsequently used to determine the location of reflectors.

In contrast, for the applications we consider in our work, it is desirable to locate the reflectors passively, i.e., without controlled emissions. Due to the duality of the acoustic system, we can use a single source and multiple sensors (i.e. a single-input-multiple-output (SIMO) system) to deduce the location of reflectors. Such a SIMO system can be identified blindly using techniques described in [5], [6]. The missing but important piece of information in this scenario is the source range, i.e. the distance from the sound source to the closest microphone, which is initially unknown. However, the source range can be extracted from the identified acoustic SIMO system using the time differences of arrival (TDOAs) associated with the direct paths.

In this paper we consider the problem of localizing the reflective boundaries (i.e. walls) of a two-dimensional (2-D) enclosure, based on estimates of at least three acoustic impulse responses (AIRs) between stationary microphones located at arbitrary but known relative positions and a sound source located at an unknown position. Using a two-step approach we first estimate the range and location of the source relative to the position of a reference microphone using a closed-form estimator [7]. Secondly, we use the relative position of the source along with estimates of the TOAs related to first-order reflections, to constrain the possible reflector locations. Assuming that the reflective surfaces are at a unique distance from each microphone, the propagation delay of the sound associated with each reflector will appear as a distinct peak in each AIR. By grouping together peaks that are associated with a specific reflector, it is possible to localize this reflector. Finally we show that it is possible to infer the geometry of the acoustic environment from the AIRs and relative microphone positions alone.

†The authors acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 226007 SCENIC.

II. Proposed Approach

Working in a 2-D space, we assume one spatially stationary source, at an unknown location, and multiple stationary microphones, arranged at arbitrary but known relative positions inside the acoustic enclosure. Our task then is to infer the geometry of the enclosed space. The emitted wavefront undergoes various reflections off walls, and other reflective surfaces, and so each microphone will receive the sum of the direct-path signal and scaled replicas of the source signal with various time delays. However, the estimation accuracy of blind system identification processes for a SIMO system is limited, and it is important to note that the identified AIRs do not directly disclose the propagation time of the direct path between the source and the closest microphone. This propagation time is however directly linked to the source-microphone distance (i.e. range) which can be estimated using at least three TDOAs of the direct-paths. Thus, by first estimating the range with respect to one microphone we are able to infer the time of arrival of the direct propagation to each other microphone.

Every peak in the estimated AIR, except for the first peak that is related to the direct-path, is associated with a reflector present in the room. The TOA of a first-order reflection is composed of the propagation time associated with the path between source and reflector along with the path between reflector and microphone. Note that the angle of reflection and incidence are assumed equal from Snell's law. This means that we can geometrically constrain the possible trajectories of the sound and observe that the locus of candidate reflection points is an ellipse. By using the estimated source range along with the TOA of a reflective path we can specify the scaling of the major and minor axes of each ellipse. Furthermore, using the information about the relative position between source and microphones, it is possible to translate and rotate these estimated ellipses within our reference framework. We note that if the reflective points lie on a line (i.e. a wall), then we can estimate this reflector line as the common tangent to all associated ellipses [3]. If we assume that only the walls are the reflective boundaries of the enclosure, then our channel estimates contain the TOAs associated with these walls. Assuming a rectangular room we obtain four distinct TOAs due to first-order reflections in each AIR, from which the location of the walls and thus the geometry of the room can be inferred. Our current method considers only first-order reflections though extension to higher-order is potentially feasible.

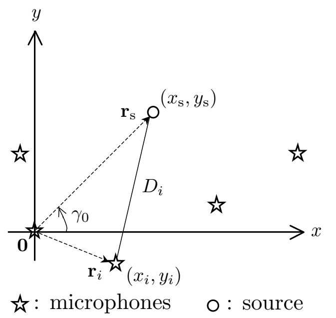

Fig. 1: Localization in a 2-D plane.

In Section III we will show how the range estimate between the source and a reference microphone can be obtained. The associated TOAs of direct-path and first-order reflections can be estimated using this range. Section IV outlines how the reflectors can be localized based on the TOA estimates. In Section V we present experimental results and finally draw conclusions in Section VI.

III. Range Estimation

In this section we show how the sound source can be localized using three or more microphones. The passive source localization algorithm outlined in [8] is based on a least squares estimator employing a spherical least squares (LS) error criterion defined in three-dimensional (3-D) space. For our purposes this algorithm is modified for a 2-D space. Consequently, the spherical LS error function is modified to a circular LS error criterion.

Assume that there are M microphones distributed arbitrarily in a 2-D plane located at positions

$$\mathbf{r}_i \triangleq [x_i \; y_i]^T, \quad i = 0, \ldots, M-1 \tag{1}$$

with the reference microphone ($i = 0$) placed at the origin of the coordinate system, i.e., $\mathbf{r}_0 = [0 \; 0]^T$, and a source located at $\mathbf{r}_s \triangleq [x_s \; y_s]^T$. The distances from the origin to the ith microphone and the source are denoted by R_i and R_s, respectively, where

$$R_i \triangleq \|\mathbf{r}_i\| = \sqrt{x_i^2 + y_i^2}, \quad i = 1, \ldots, M-1$$
$$R_s \triangleq \|\mathbf{r}_s\| = \sqrt{x_s^2 + y_s^2}.$$

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, denoted $\tau_{i,j}$. If the speed of sound is η, then

$$d_{i,j} = \eta \cdot \tau_{i,j}. \tag{2}$$

The acoustic source location, as well as its range, can be estimated using least-squares. We observe that the correct source location should be at the intersection of a group of circles. The center of each circle is equal to the location of the microphone and the radius of each circle is related to the source-microphone distance. Therefore, the best estimate of the source location will be the point that yields the shortest distance to those circles defined by the range differences and the hypothesized source range. From [8] we establish the distance D_i from the ith microphone to the source

$$\hat{D}_i = R_s + \hat{d}_{i0}, \tag{3}$$

where $(\hat{\cdot})$ denotes an observation based on the measured range difference. The error function is then defined as the difference between the measured and true values, which when putting the $M+1$ errors together and writing them in a vector form gives

$$\mathbf{e}(\mathbf{r}_s) = \mathbf{A}\theta - \mathbf{b} \tag{4}$$

where

$$\mathbf{A} \triangleq \left[\mathbf{S} \mid \hat{\mathbf{d}}\right], \quad \mathbf{S} \triangleq \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_N & y_N \end{bmatrix}$$

$$\theta \triangleq \begin{bmatrix} x_s \\ y_s \\ R_s \end{bmatrix}, \quad \mathbf{b} \triangleq \frac{1}{2} \begin{bmatrix} R_1^2 - \hat{d}_{10}^2 \\ R_2^2 - \hat{d}_{20}^2 \\ \vdots \\ R_N^2 - \hat{d}_{N0}^2 \end{bmatrix}$$

and $\mathbf{S} \mid \hat{\mathbf{d}}$ indicates that $\mathbf{S}$ and $\hat{\mathbf{d}}$ are stacked side-by-side with $\hat{\mathbf{d}} = [\hat{d}_{10}, \hat{d}_{20}, \ldots, \hat{d}_{N0}]^T$. The corresponding LS criterion is then given by

$$J = \mathbf{e}^T\mathbf{e} = [\mathbf{A}\theta - \mathbf{b}]^T [\mathbf{A}\theta - \mathbf{b}]. \tag{5}$$

The solution for θ is given by [8]

$$\hat{\theta} = \mathbf{A}^{\ddagger}\mathbf{b}, \tag{6}$$

where $(\cdot)^{\ddagger}$ defines the pseudo-inverse. Note that the final output in (6) can be additionally improved using an iterative method as shown in [8]. Using the estimated range between the source and the first microphone (given by R_s) and the estimated AIRs, the TOA of each first-order reflection can be computed.

IV. Reflector Localization

In the following section we use the TOAs of the first-order reflections to determine the location of the reflectors. First we determine the parameters of the ellipses that describe possible reflector locations for a given source-microphone pair and reflection. On the basis of these ellipses we can search for a common tangent that represents a reflector present in the acoustic enclosure. Finally we estimate which of these tangents yield a true solution to the localization problem of the four walls of a rectangular room.

A. Ellipse Parametrization

In homogenous coordinates we can define a conic in two dimensions using the parameters $\{a, b, c, d, e, f\}$ as [9]

$$C = \left\{(x, y) \in \mathbb{R}^2 | ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0\right\}. \tag{7}$$

By setting $\mathbf{v} = [x \; y \; 1]^T$ and $\mathbf{C} = \begin{bmatrix} a & b & d \\ b & c & e \\ d & e & f \end{bmatrix}$ this can be written as

$$\mathbf{v}^T\mathbf{C}\mathbf{v} = 0. \tag{8}$$

This is an ellipse after constraining

$$\det(\mathbf{C}) \neq 0, \begin{vmatrix} a & b \\ b & c \end{vmatrix} > 0, \det(\mathbf{C})/(a+c) < 0. \tag{9}$$

Using the results of the range estimation from Section III we wish to directly assign values to the parameters a,b,c,d,e,f. In order to do this we first note that the points on a unit circle satisfy

$$\mathbf{v}^T\mathbf{C}_I\mathbf{v} = 0, \tag{10}$$

where $\mathbf{C}_I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$. We observe that for matrix $\mathbf{C}_I$ the positive index of inertia is equal to two and that the negative index of inertia is equal to one [10]. In other words $\mathbf{C}_I$ has one negative and two positive eigenvalues. This justifies the following decomposition for the ellipse associated with the ith microphone ($i \in \{1, \cdots, M\}$), and the kth reflector ($k \in \{1, \cdots, N\}$):

$$\mathbf{C}_{i,k} = \mathbf{T}_i^{-T} \mathbf{R}_i^{-T} \mathbf{S}_{i,k}^{-T} \mathbf{C}_\mathrm{I} \mathbf{S}_{i,k}^{-1} \mathbf{R}_i^{-1} \mathbf{T}_i^{-1}, \tag{11}$$

where we can define a translation, rotation and scaling matrix such that

$$\mathbf{T}_i = \begin{bmatrix} 1 & 0 & \Delta x_i \\ 0 & 1 & \Delta y_i \\ 0 & 0 & 1 \end{bmatrix}, \tag{12}$$

$$\mathbf{R}_i = \begin{bmatrix} \cos\gamma_i & -\sin\gamma_i & 0 \\ \sin\gamma_i & \cos\gamma_i & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{13}$$

$$\mathbf{S}_{i,k} = \begin{bmatrix} Q_{i,k}^{\text{maj}} & 0 & 0 \\ 0 & Q_{i,k}^{\text{min}} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{14}$$

The quantities Δx_i, Δy_i, γ_i, $Q_{i,k}^{\text{maj}}$ and $Q_{i,k}^{\text{min}}$ are defined as follows. The point at $(\Delta x_i, \Delta y_i)$ can be seen as the geographic midpoint between $\mathbf{r}_\mathrm{s}$ and $\mathbf{r}_i$ and is defined by

$$\Delta x_i \triangleq x_\mathrm{s} + \frac{D_i \cos(\gamma_i)}{2}; \; \Delta y_i \triangleq y_\mathrm{s} + \frac{D_i \sin(\gamma_i)}{2},$$

with $\gamma_i \triangleq \tan^{-1}\left(\frac{y_\mathrm{s} - y_i}{x_\mathrm{s} - x_i}\right)$. If we assume N reflectors in our environment, then every channel estimate contains information about the N TOAs due to the reflective sound path. Consequently we can define $p_{i,k}$ as the TOA associated with the ith microphone and the kth reflector. The scaling of the semi-major and semi-minor axes of each ellipse is then given by

$$Q_{i,k}^{\text{maj}} \triangleq \frac{\sqrt{(\eta \cdot p_{i,k})^2 - D_i^2}}{2}; \; Q_{i,k}^{\text{min}} \triangleq \frac{\eta \cdot p_{i,k}}{2},$$

respectively. For N reflectors we can therefore construct $M \times N$ ellipses for all source-microphone pairings.

B. Common Tangent Estimation

We can define a line as [9]

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

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

$$\mathbf{l}^T \mathbf{v} = 0. \tag{16}$$

As stated in [9] this line is tangential to an ellipse if

$$\mathbf{l}^T \cdot \text{adj}(\mathbf{C}) \cdot \mathbf{l} = 0, \tag{17}$$

where

$$\text{adj}(\mathbf{C}) = \det(\mathbf{C}) \cdot \mathbf{C}^{-1}. \tag{18}$$

The line that is tangential to a set of ellipses coincides with the reflector. The three unknown parameters of $\mathbf{l}$ can be estimated by minimizing the following cost function for a particular $k \in \{1, \cdots, N\}$:

$$J_\mathrm{e}\left(\mathbf{l}, \{\mathbf{C}_{i,k}\}_{i=1}^M\right) = \sum_{i=1}^{M} \left\|\mathbf{l}^T \cdot \text{adj}(\mathbf{C}_{i,k}) \cdot \mathbf{l}\right\|^2, \tag{19}$$

where $M \geq 3$, with

$$\hat{\mathbf{l}}_k = \arg\min_{\mathbf{l}} J_\mathrm{e}\left(\mathbf{l}, \{\mathbf{C}_{i,k}\}_{i=1}^M\right). \tag{20}$$

Note however that the objective function in (19) is non-convex. Consequently, when employing an optimization algorithm to find $\mathbf{l}$, it is possible to get trapped in a local minimum rather than finding the global minimum. As proposed in [3], this problem can be alleviated by imposing that l_1 and l_2 lie on a circle of radius 1:

$$l_1 = \cos(\alpha), \; l_2 = \sin(\alpha). \tag{21}$$

We can find $\hat{\mathbf{l}}_\alpha = [\cos(\alpha), \sin(\alpha), l_3]^T$ by minimizing J_e in (19) using $\mathbf{l}_\alpha$ rather than $\mathbf{l}$ and thus estimate the location of the reflector as the common tangent to all ellipses associated with that particular reflector.

C. Localization

Prior knowledge is needed to correctly group together related ellipses, as our search space spans up to $\frac{(N \cdot M)!}{M!(N \cdot M - M)!}$combinations of possible ellipses used in (19). Since we assume that our channel estimates only yield the TOAs due to first-order reflections, then for a rectangular room (considering only four walls) we obtain four distinct TOAs (i.e. $N = 4$) for each source-microphone pair. We can therefore construct $M \times 4$ ellipses using (11). However in order to estimate the location of a particular reflector, only the ellipses associated with that particular reflector should be employed in (19). We therefore iteratively localize each reflector, starting with the geometrically closest reflector to the reference, gradually minimizing the total search space by discarding ellipses associated with already localized reflectors.

The TOA associated with the closest wall to the reference microphone ($\mathbf{r}_0$) is described by $p_{0,1}$. The ellipse constructed on basis of this TOA needs to be grouped together with all combinations of ellipses due to other source-microphone pairings and their associated TOAs, i.e.

$$p_{i,k}, \; i \in \{1, \cdots, M-1\}; \; k \in \{1, \cdots, 4\}.$$

This results in $\frac{(4 \cdot M)!}{M!(4 \cdot M - M)!}$ possible combinations. We can directly use (19) for each combination. The combination with the smallest value for J_e is obtained when all ellipses belong to the same reflector (i.e. $p_{i,1}$). All ellipses associated with that particular reflector can henceforth be discarded from the search space for subsequent iterations. This means that the reflector associated with $p_{0,2}$ can be estimated by employing (19) for the reduced set of $\frac{(3 \cdot M)!}{M!(3 \cdot M - M)!}$ different combinations. Consequently the search space for the third reflector, associated with $p_{0,3}$, is reduced to $\frac{(2 \cdot M)!}{M!(2 \cdot M - M)!}$ combinations while the final reflector, associated with $p_{0,4}$, is localized using the last M ellipses.

In this way we can iteratively localize all common reflective lines, i.e. the walls of the room.

V. Performance Evaluation

In order to simulate real acoustic environments, the source-image method [11] is employed to simulate the room response for a chosen geometry with an arbitrary source and microphone arrangement. The floor and ceiling are assumed completely absorbing such that only the walls are reflecting. In this experimental study only the first-order reflections are simulated.

A. Experimental Setup

To evaluate the performance of the proposed algorithm, we carried out Monte Carlo simulations for blind source localization and geometry reconstruction of an arbitrary rectangular room. Two sets of experiments are conducted using an omnidirectional source and four to five omnidirectional microphones.

- Room A: (Room dimension; 4×3 m).
- Room B: (Room dimension; 3×6 m).

The test environment is depicted in Figure 2. The sound source ($\mathbf{r}_\mathrm{s}$) and the microphones ($\mathbf{r}_i$) are uniformly distributed within the shaded

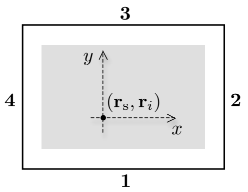

Fig. 2: Geometry of test environment.

region inside the room, constraining the positions to be at a distance of at least 0.5 m away from each wall and with each microphone being kept at a minimum distance of 0.5 m away from the source.

B. Results

We assess the accuracy of our method by averaging the results of 100 Monte Carlo runs. For the first step of the algorithm (source localization), the Euclidean distance between the estimated source location and the actual source location is used as an error measure. The mean and variance of this error, denoted $e_{\rm d}$, are calculated. The second part of the algorithm, namely the wall localization, is assessed as outlined in [3]. Namely if $\hat{\mathbf{l}}$ and $\mathbf{l}$ are, respectively, the actual and estimated reflector lines, we measure $e_{\rm rl}{=}(\hat{\mathbf{l}}^T\mathbf{l})/(\|\hat{\mathbf{l}}\|\|\mathbf{l}\|)$: a value close to 1 indicates that the angle between the lines is small. Similarly, mean and variance for this error are calculated for each wall separately.

The results of the source localization are presented in Table I. The source is well localized for both test environments, with an average error of only a few cm's. As the number of microphones is increased, we observe a concurrent increase of localization accuracy. As expected from [8], the mean and variance are larger for room B, which is bigger than room A, particularly for $M = 4$. However the results are almost identical for $M = 5$.

TABLE I: Source localization results.

Room	M	$\mu(e_{\rm d})$ [cm]	$\sigma(e_{\rm d})$ [cm]
A	4	3.51	7.67
A	5	1.25	2.10
B	4	4.37	10.34
B	5	1.34	2.01

TABLE II: Reflector localization results.

Room	M	$\mu(e_{\rm rl})$	$\sigma(e_{\rm rl})$
A	4	0.983	0.059
A	5	0.986	0.052
B	4	0.978	0.054
B	5	0.985	0.051

The accuracy of the wall localization is first presented by averaging $e_{\rm rl}$ for all four walls. These results are shown in Table II and are comparable to [3]. We additionally considered a particular arrangement of source and microphones and assessed the error for each wall (as indexed in Figure 2) separately. As an example, the exact localization results for room A are shown in Table III. It is observable that the deviations from the ground-truth values are very small. Figure 3 shows that the walls are accurately localized for this particular arrangement of source and microphones in room A.

VI. Conclusion

We presented a two-step approach for blindly estimating the sound source position and reconstructing the geometry of an acoustic enclosure using channel estimates from multiple microphones. Preliminary simulation results demonstrate that we can accurately identify the reflectors when the TDOAs associated with the first-order reflections can be detected. On basis of these promising results our approach can be used as a starting point for future research in which blindly estimated AIRs and higher order reflections are used and non-rectangular room geometries are considered.

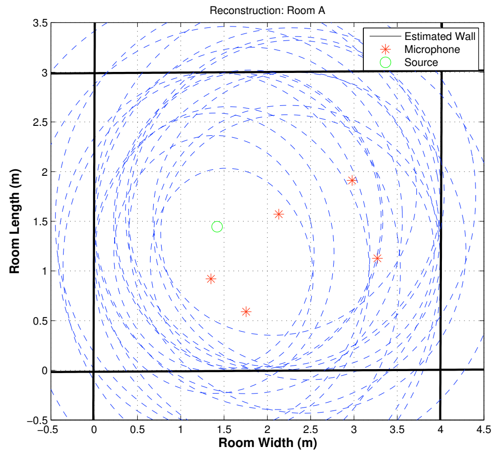

Fig. 3: A particular reconstruction result for room A.

TABLE III: Comparison of actual and estimated localization results for a particular experiment in room A.

Wall	$(\alpha\,[°]\,, l_3[m])$	$(\hat{\alpha}\,[°]\,, \hat{l_3}[m])$	$\mathbf{e}_{\rm rl}$
1	(90,0)	(90.282, 0.015)	0.999
2	(180,4)	(179.751, 3.998)	1.000
3	(-90,3)	(-89.663, 2.991)	1.000
4	(0,0)	(0.249,0.008)	1.000

References

[1] P. Aarabi and B. Mungamuru, "Scene reconstruction using distributed microphone arrays," in *Proc. Intl. Conf. on Multimedia and Expo (ICME)*, vol. 3, 2003, pp. 53–56.

[2] A. O'Donovan, R. Duraiswami, and D. Zotkin, "Imaging concert hall acoustics using visual and audio cameras," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2008.

[3] F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic reconstruction of the geometry of an environment through successive acquisitions of a controlled emission," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2009.

[4] F. Antonacci, D. Lonoce, M. Motta, A. Sarti, and S. Tubaro, "Efficient source localization and tracking in reverberant environments using microphone arrays," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, Mar. 2005, pp. 1061–1064.

[5] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[6] Y. Huang, J. Benesty, and J. Chen, "Adaptive blind multichannel identification," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Springer-Verlag, 2007, ch. 13, pp. 259–279, part B.

[7] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 12, pp. 1661–1669, Dec. 1987.

[8] 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.

[9] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2001.

[10] A. J. Laub, *Matrix Analysis for Scientists and Engineers*. Society for Industrial and Applied Mathematics (SIAM), 2005, p. 103.

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