SINGLE-CHANNEL INDOOR MICROPHONE LOCALIZATION

Reza Parhizkar, Ivan Dokmanić and Martin Vetterli

School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
{reza.parhizkar, ivan.dokmanic, martin.vetterli}@epfl.ch

ABSTRACT

We propose a novel method for single-channel microphone localization inside a known room. Unlike other approaches, we take advantage of the room reverberation, which enables us to use only a single fixed loudspeaker to localize the microphone. Our method uses an *echo labeling* approach that associates the echoes to the correct walls. Echo labeling leverages the properties of the Euclidean distance matrices formed from the distances between the virtual sources and the microphone. Experiments performed in a real lecture room verify the effectiveness of the proposed localization algorithm.

Index Terms— Single-channel localization, microphone calibration, indoor localization, s-stress

1. INTRODUCTION

Most audio sensor array applications rely on the precise knowledge of the microphone positions. This motivated the development of several approaches for localization of microphones in an array. For example, in [1] the authors describe a closed-form method for calculating the relative geometry of multiple microphone arrays with known shapes. In [2], a maximum-likelihood approach is used to find the positions of microphones in an array. Multidimensional scaling is used to solve a similar problem in [3].

The topic is also highlighted by several papers at ICASSP-2013. An optimization approach to self-localization of ad-hoc arrays is presented in [4]. The solution does not require synchronization between the sources and the array. A characterization of cases when the solution exists is described in [5], as well as a minimal solver.

All of the above approaches involve multiple sources and receivers. Furthermore, the methods are independent of the fact that the localization is performed indoors (the reverberation is even considered detrimental). We note that localizing a single microphone inside a room is useful in its own right. Example applications are localization of household robots, people or smartphones, to name a few. Standard methods for microphone localization include triangulation and trilateration, both of which rely on multiple sources of sound (with the known relative geometry).

In contrast to previous approaches, in this paper we propose a single-channel method for microphone localization. Our method takes advantage of the room, by utilizing echoes from the walls to find the microphone position. The underlying assumption is that the room geometry is known, as well as the loudspeaker location. We do not require the detailed knowledge of the room structure—locations and orientations of principal reflectors (typically walls) suffice.

In [6], a method is described for determining the room geometry from acoustic echoes, using a single sound source, and a minimum of 4 microphones. The microphone localization method presented in this paper can be regarded as a dual of the room geometry reconstruction algorithm. Similarly to room reconstruction, we use *echo labeling* to associate echoes recorded by the microphone to the walls that generated them. Echo labeling is performed with the help of Euclidean distance matrices (EDM). The EDMs are used as a filter that reveals the correct combinations of echoes. This is formalized in Section 3. In Section 5 we present a real microphone localization experiment performed in a lecture room on EPFL campus. The results of the experiment verify the accuracy and robustness of the proposed localization algorithm.

We note that the microphone localization problem is similar to source localization in acoustics [7, 8]. Moreover, the method can be applied not only using sound, but also with any time-of-flight measuring pair of devices (ultrasound, light, UWB) [9, 10].

2. MODELING

We consider the known room to be a K-faced polygon. We work in three dimensions and note that the results are immediately valid in 2D as a special case. As the room shape is known, the location of wall vertices $\boldsymbol{p}_i \in \mathbb{R}^3$ are available. We also assume that there is one loudspeaker with known location $\boldsymbol{s} \in \mathbb{R}^3$ inside the room. We can model the sound propagation inside the room by the room impulse response (RIR). An RIR describes the acoustic channel between the source and the receiver inside the room. It depends on the shape of the room and locations of the loudspeaker and the

This work was partly supported by the ERC Advanced Grant – Support for Frontier Research – SPARSAM Nr: 247006.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

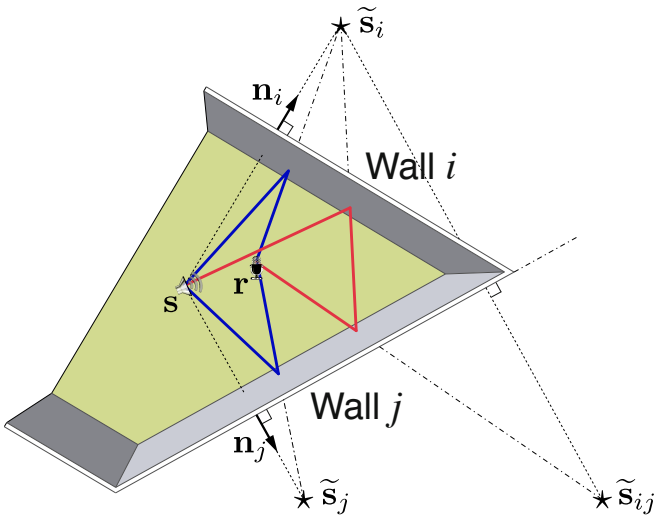


Fig. 1. Illustration of the image source model for first and second-order echoes. Image sources are shown by stars, and $\widetilde{\boldsymbol{s}}_{ij}$ is the image source corresponding to the second-order echo.

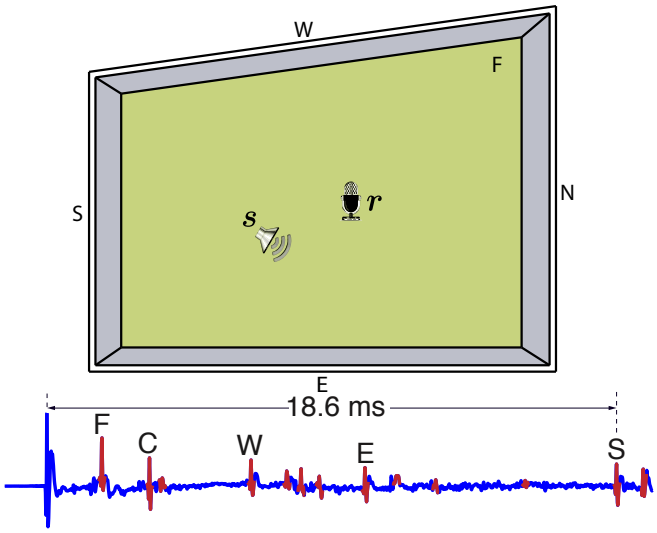


Fig. 2. Based on the shape of the room, position of the loudspeaker and also position of the microphone, echoes from walls may arrive in different orders.

microphone. Ideally it is a train of Diracs, each corresponding to an echo:

$$h(t) = \sum_i c_i \, \delta(t - t_i) \,,$$

where c_i and t_i are the amplitude and time of arrival of the ith echo. A microphone hears the convolution of the loudspeaker signal with the RIR. We are interested in estimating the unknown position of a microphone inside the room using the room impulse response between the microphone and the loudspeaker.

By measuring the room impulse response we access the echo times t_i. These echo times can be linked to the room geometry and microphone position with the image source model [11, 12]. According to this model, we can replace an echo from a wall by a virtual source behind the wall in a mirrored location of the original source.

As illustrated in Figure 1, virtual sources are mirror images of the true sources across the corresponding reflecting walls. The image $\widetilde{\boldsymbol{s}}_i$ of the source $\boldsymbol{s}$ with respect to the ith wall is computed as

$$\widetilde{\boldsymbol{s}}_i = \boldsymbol{s} + 2\langle \boldsymbol{p}_i - \boldsymbol{s}, \boldsymbol{n}_i \rangle \boldsymbol{n}_i, \tag{1}$$

where $\boldsymbol{n}_i$ is the unit normal to the ith wall. The time of arrival of the echo from the ith wall is $t_i = \|\widetilde{\boldsymbol{s}}_i - \boldsymbol{r}\|/c$, where c is the speed of sound and $\boldsymbol{r}$ is the location of the microphone. Assuming that the sound speed inside the room is fixed and known, we can relate the time of arrival of the echoes to the mutual distances of the microphone and the image sources.

As the geometry of the room and the position of the loudspeaker are known, we have access to the position of the image sources, $\widetilde{\boldsymbol{s}}_i$ (refer to (1)). If we were also to know the correspondences of the recorded echoes by the microphone with the image sources (which echo comes from which wall) we would be able to multilaterate the position of the microphone. However, we face two problems:

- Not all the extracted echoes from the impulse response correspond to first order image sources,
- The echoes arrive to the microphone in different orders based on the position of the microphone.

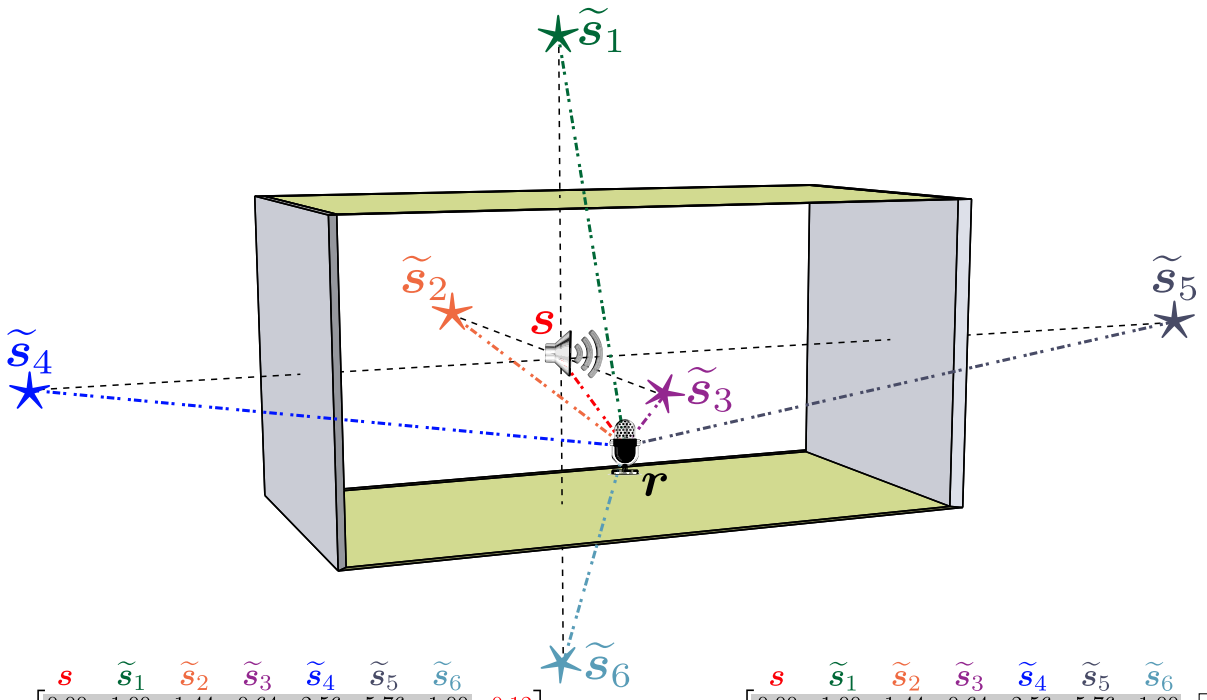
An example with real measurements is shown in Figure 2. As can be seen from the figure, many of the extracted peaks in the impulse response do not correspond to a valid image source. Thus, we are facing a labeling problem.

3. ECHO LABELING

With the echo labeling procedure we aim at first extracting the correct echoes from the impulse response and second finding the right assignment of these echoes to the walls. Consider the setup shown in Figure 3. Let $\boldsymbol{D} \in \mathbb{R}^{(K+1)\times(K+1)}$ be a matrix whose entries are as follows:

$$\boldsymbol{D}[i,j] = \begin{cases} \|\boldsymbol{s} - \widetilde{\boldsymbol{s}}_j\|^2 & i = 1 \\ \|\widetilde{\boldsymbol{s}}_i - \boldsymbol{s}\|^2 & j = 1 \\ \|\widetilde{\boldsymbol{s}}_{i-1} - \widetilde{\boldsymbol{s}}_{j-1}\|^2 & 2 \le i,j \le K+1 \end{cases}, \tag{2}$$

where $\widetilde{\boldsymbol{s}}_i$ are the locations of the first order image sources. As the geometry of the room and the location of the loudspeaker are known, $\boldsymbol{D}$ is a Euclidean distance matrix (EDM) with known entries. As the loudspeaker emits a sound, the microphone receives the direct sound (the first peak in its RIR)



$$\boldsymbol{D}_{\text{aug},1} = \begin{array}{c} \begin{matrix} \boldsymbol{s} & \widetilde{\boldsymbol{s}}_1 & \widetilde{\boldsymbol{s}}_2 & \widetilde{\boldsymbol{s}}_3 & \widetilde{\boldsymbol{s}}_4 & \widetilde{\boldsymbol{s}}_5 & \widetilde{\boldsymbol{s}}_6 & \end{matrix} \\ \begin{bmatrix} 0.00 & 1.00 & 1.44 & 0.64 & 2.56 & 5.76 & 1.00 & 0.12 \\ 1.00 & 0.00 & 2.44 & 1.64 & 3.56 & 6.76 & 4.00 & 1.52 \\ 1.44 & 2.44 & 0.00 & 4.00 & 4.00 & 7.20 & 2.44 & 2.04 \\ 0.64 & 1.64 & 4.00 & 0.00 & 3.20 & 6.40 & 1.64 & 0.44 \\ 2.56 & 3.56 & 4.00 & 3.20 & 0.00 & 16.0 & 3.56 & 3.32 \\ 5.76 & 6.76 & 7.20 & 6.40 & 16.0 & 0.00 & 6.76 & 4.92 \\ 1.00 & 4.00 & 2.44 & 1.64 & 3.56 & 6.76 & 0.00 & 0.72 \\ 0.12 & 1.52 & 2.04 & 0.44 & 3.32 & 4.92 & 0.72 & 0.00 \end{bmatrix} \end{array}$$

$$\boldsymbol{D}_{\text{aug},2} = \begin{array}{c} \begin{matrix} \boldsymbol{s} & \widetilde{\boldsymbol{s}}_1 & \widetilde{\boldsymbol{s}}_2 & \widetilde{\boldsymbol{s}}_3 & \widetilde{\boldsymbol{s}}_4 & \widetilde{\boldsymbol{s}}_5 & \widetilde{\boldsymbol{s}}_6 & \end{matrix} \\ \begin{bmatrix} 0.00 & 1.00 & 1.44 & 0.64 & 2.56 & 5.76 & 1.00 & \boxed{4.92} \\ 1.00 & 0.00 & 2.44 & 1.64 & 3.56 & 6.76 & 4.00 & 1.52 \\ 1.44 & 2.44 & 0.00 & 4.00 & 4.00 & 7.20 & 2.44 & 2.04 \\ 0.64 & 1.64 & 4.00 & 0.00 & 3.20 & 6.40 & 1.64 & 0.44 \\ 2.56 & 3.56 & 4.00 & 3.20 & 0.00 & 16.0 & 3.56 & 3.32 \\ 5.76 & 6.76 & 7.20 & 6.40 & 16.0 & 0.00 & 6.76 & \boxed{0.12} \\ 1.00 & 4.00 & 2.44 & 1.64 & 3.56 & 6.76 & 0.00 & 0.72 \\ \boxed{4.92} & 1.52 & 2.04 & 0.44 & 3.32 & \boxed{0.12} & 0.72 & 0.00 \end{bmatrix} \end{array}$$

Fig. 3. An example for echo labeling for microphone localization. The gray part of the matrices show the distances between the sources. We augment this matrix with a combination of echoes extracted from the microphone RIR. If the echoes are selected correctly and have the right order the augmented matrix is an EDM. The matrix $\boldsymbol{D}_{\text{aug},1}$ is an EDM. But since the echoes are not correctly ordered in $\boldsymbol{D}_{\text{aug},2}$, it is not an EDM.

and K first order echoes from the walls (consecutive peaks in its RIR). We propose a novel algorithm to extract these echoes from the RIR and label them according to their corresponding wall. To this end we use a fundamental property of EDMs: An EDM corresponding to a point set in $\mathbb{R}^n$ has rank at most $n+2$ [13]. Thus, in 3D its rank is at most 5. We start by the known EDM, $\boldsymbol{D}$ and augment it as follows: We choose $(K+1)$ echoes from the RIR of the microphone and augment $\boldsymbol{D}$ with them (we add an extra column and row to it). If these echoes are correctly assigned to the image sources, then they represent the distances of the microphone from these image sources and the augmented matrix $\boldsymbol{D}_{\text{aug}}$ is an EDM and thus will be low rank. However, if we did not choose the correct peaks or they do not have the right permutation, then the augmented matrix will not be an EDM. For example in Figure 3, since $\boldsymbol{D}_{\text{aug},1}$ contains the correct permutation of the echoes, it is an EDM, while $\boldsymbol{D}_{\text{aug},2}$ is not an EDM.

More formally, let $\boldsymbol{e}$ list the candidate distances computed from the RIR recorded by the microphone. We proceed by augmenting the matrix $\boldsymbol{D}$ with a combination of K unlabeled squared distances $\boldsymbol{d}_{(i_1,\ldots,i_K)}$ to get $\boldsymbol{D}_{\text{aug}}$,

$$\boldsymbol{D}_{\text{aug}}(\boldsymbol{d}_{(i_1,\ldots,i_K)}) = \begin{bmatrix} \boldsymbol{D} & \boldsymbol{d}_{(i_1,\ldots,i_K)} \\ \boldsymbol{d}^T_{(i_1,\ldots,i_K)} & 0 \end{bmatrix}.$$

The column vector $\boldsymbol{d}_{(i_1,\ldots,i_K)}$ is constructed as

$$\boldsymbol{d}_{(i_1,\ldots,i_K)}[k] = \boldsymbol{e}^2[i_k],$$

with $i_k \in \{1,\ldots,\text{length}(\boldsymbol{e})\}$. In words, we construct a candidate combination of echoes $\boldsymbol{d}$ by selecting K echoes out of all extracted echoes from the microphone RIR. Note that $\text{length}(\boldsymbol{e}) \neq K$ in general, meaning that we might pick more than K echoes from the RIR of the microphone.

If $\text{rank}(\boldsymbol{D}_{\text{aug}}) \leq 5$ or more specifically $\boldsymbol{D}_{\text{aug}}$ verifies the EDM properties, then the selected combination of echoes is the correct permutation.

4. PRACTICAL ALGORITHM

Both the measurements for $\boldsymbol{D}$ and $\boldsymbol{e}$ are often noisy. Thus, the rank test might not be enough for practical applications. Instead, we can check how close the augmented matrix $\boldsymbol{D}_{\mathsf{aug}}$ is to an EDM. We use multi-dimensional scaling (MDS) to define such measure of closeness. Introduced originally in psychometrics for data visualization [14], MDS tries to find the point set in a given dimension (here three) that produces an EDM closest to $\boldsymbol{D}_{\mathsf{aug}}$. Specifically, we use the s–stress criterion [15]. For each selection of echoes that results in $\widetilde{\boldsymbol{D}}_{\mathsf{aug}}$, $\mathsf{s\text{–}stress}(\widetilde{\boldsymbol{D}}_{\mathsf{aug}})$ is the value of the following optimization program,

$$\underset{\boldsymbol{D}_{\mathsf{aug}} \in \mathbb{EDM}^{(3)}}{\text{minimize}} \sum_{i,j} \left(\boldsymbol{D}_{\mathsf{aug}}[i,\, j] - \widetilde{\boldsymbol{D}}_{\mathsf{aug}}[i,\, j] \right)^2 . \tag{3}$$

By $\mathbb{EDM}^{(3)}$ we denote the set of EDMs generated by point sets in $\mathbb{R}^3$. We call $\mathsf{s\text{–}stress}(\widetilde{\boldsymbol{D}}_{\mathsf{aug}})$ the score of the matrix $\widetilde{\boldsymbol{D}}_{\mathsf{aug}}$, and use it to assess the likelihood that a permutation of echoes is correct. For optimizing (3) we use the method proposed in [16] which in almost every case finds the global minimum of the s–stress function.

The combination of echoes which results in the minimum value for the s–stress score is selected as the correct permutation. The algorithm is summarized as:

i. For every $\boldsymbol{d}_{(i_1,\ldots,i_K)}$,

$$\mathsf{score}[\boldsymbol{d}_{(i_1,\ldots,i_K)}] \leftarrow \mathsf{s\text{–}stress}(\widetilde{\boldsymbol{D}}_{\mathsf{aug}})$$

ii. Find the minimum score collected in score,

iii. Use the found echo combination and the image source locations to compute the microphone location.

Note that although the algorithm needs to check echo combinations and permutations, it is not necessary to test all echo combinations. The dimensions of the room together with the location of the loudspeaker define the size of a window in which all the first order echoes lie.

5. EXPERIMENTS

We ran an experiment in a lecture room on EFPL campus. Two walls of the room are glass windows, and two are gypsum-board partitions. The room is equipped with a perforated metal plate ceiling suspended below a concrete ceiling. We replaced one wall by a wall made of tables. We used a directional loudspeaker Genelec 8030A, and a non-matched omni-directional microphone Behringer ECM 8000. The RIR from the loudspeaker to the microphone was estimated by the sine sweep technique [17]. The room dimensions are known a-priori and the loudspeaker location was measured during the experiment. The experimental setup with the image sources of the loudspeaker are shown in Figure 4. As the loudspeaker

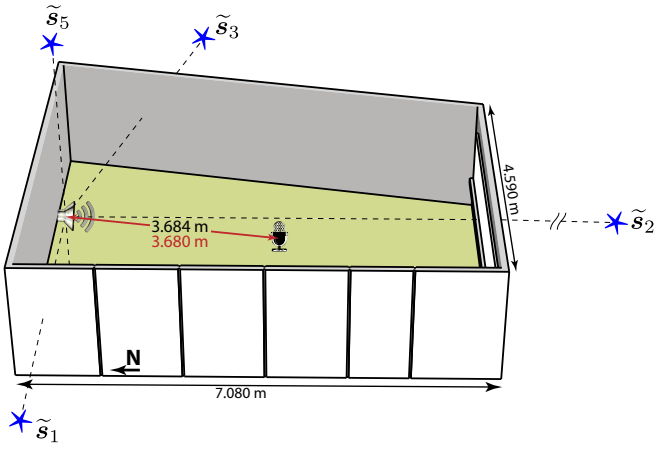


Fig. 4. Sketch of a room on EPFL campus where the in-room localization experiment is performed. The image sources of the loudspeaker are shown with stars. The image source of the floor ($\widetilde{s}_4$) is not shown for better visualization. The actual distance of the loudspeaker and the microphone is shown in red while the estimated distance is in black.

is placed against the north wall, we do not consider the image source for this wall. The matrix $\boldsymbol{D}$—defined in (2)—is

$$\boldsymbol{D} \approx \begin{bmatrix} 0.00 & 25.40 & 178.48 & 5.91 & 4.66 & 10.38 \\ 25.40 & 0.00 & 203.90 & 55.40 & 30.07 & 35.77 \\ 178.48 & 203.90 & 0.00 & 172.38 & 183.15 & 188.86 \\ 5.91 & 55.40 & 172.38 & 0.00 & 10.58 & 16.28 \\ 4.66 & 30.07 & 183.15 & 10.58 & 0.00 & 28.94 \\ 10.38 & 35.77 & 188.86 & 16.28 & 28.94 & 0.00 \end{bmatrix} .$$

We augment this matrix with 6-tuples of echoes selected from the microphone's RIR. For each combination we find the value of $\mathsf{s\text{–}stress}(\boldsymbol{D}_{\mathsf{aug}})$. The combination that results in the minimum score is selected as the correct combination and the microphone position is found using the estimated permutation of the echoes. As it is shown in Figure 4 the distance of the microphone from the loudspeaker is estimated with an error of less than 1 cm.

6. CONCLUSION

We proposed a new method for microphone localization inside a known room. Our method uses Euclidean distance matrices to detect the correct echo combinations. Experiments show that our algorithm can localize the microphone in a realistic scenario with the positioning error on the order of a centimeter in a room whose sides are several meters long.

Ongoing work includes the extension of the method to rooms with more general geometries (e.g. non-convex), performing joint source-microphone localization, and the integration of the method within a comprehensive indoor localization system.

References

[1] P. Pertila, M. Mieskolainen, and M. Hamalainen, "Closed-form self-localization of asynchronous microphone arrays," in *Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, 2011, pp. 139–144.

[2] V. C. Raykar and R. Duraiswami, "Automatic position calibration of multiple microphones," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2004, vol. 4, pp. 69–72.

[3] S. Birchfield and A. Subramanya, "Microphone array position calibration by basis-point classical multidimensional scaling," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1025–1034, 2005.

[4] N. D. Gaubitch, W. B. Kleijn, and R. Heusdens, "Auto-localization in ad-hoc microphone arrays," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2013, pp. 106–110.

[5] Y. Kuang, S. Burgess, A. Torstensson, and K. Astrom, "A complete characterization and solution to the microphone position self-calibration problem," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2013, pp. 3875–3879.

[6] I. Dokmanic, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," *Proceedings of the National Academy of Sciences*, vol. 110, no. 30, pp. 12186–12191, 2013.

[7] A. Beck, P. Stoica, and J. Li, "Exact and approximate solutions of source localization problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1770–1778, 2008.

[8] X. Sheng and Y.-H. Hu, "Energy based acoustic source localization," in *Information Processing in Sensor Networks*, F. Zhao and L. Guibas, Eds., vol. 2634 of *Lecture Notes in Computer Science*, pp. 285–300. Springer Berlin Heidelberg, 2003.

[9] R. M. Vaghefi and R. M. Buehrer, "Asynchronous time-of-arrival-based source localization," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 4086–4090.

[10] J. B. Harley and J. M. Moura, "Broadband localization in a dispersive medium through sparse wavenumber analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 4071–4075.

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

[12] J. Borish, "Extension of the image model to arbitrary polyhedra," *Journal of the Acoustical Society of America*, vol. 75, no. 6, pp. 1827–1836, 1984.

[13] J. Dattorro, *Convex Optimization & Euclidean Distance Geometry*, 2011.

[14] W. S. Torgerson, "Multidimensional scaling: I. Theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, Dec. 1952.

[15] Y. Takane, F. W. Young, and J. Leeuw, "Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features," *Psychometrika*, vol. 42, no. 1, pp. 7–67, Mar. 1977.

[16] R. Parhizkar, *Euclidean Distance Matrices: Properties, Algorithms and Applications*, Ph.D. thesis, Ecole Polytechnique Federale de Lausanne (EPFL), 2013.

[17] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, 2000, pp. 1–24.