A TWO-STEP APPROACH FOR NARROWBAND SOURCE LOCALIZATION IN REVERBERANT ROOMS

Wei-Ting Lai, Lachlan Birnie, Thushara Abhayapala, Amy Bastine, Shaoheng Xu, Prasanga Samarasinghe

Audio and Acoustic Signal Processing Group, The Australian National University, Canberra, Australia

ABSTRACT

This paper presents a two-step approach for narrowband source localization within reverberant rooms. In the first step, the sound field is decomposed into direct and reverberant components where the latter is estimated as an equivalent superposition of plane waves using Iteratively Reweighted Least Squares. The second step performs source localization by modeling the dereverberated component as a sparse representation of point-source distribution using Orthogonal Matching Pursuit. The proposed method enhances localization accuracy with fewer measurements, particularly in environments with strong reverberation. A numerical simulation in a reverberant room, using a uniform microphone array surrounding the region of interest, demonstrates real-world feasibility. Notably, the proposed method and microphone placement effectively localize sound sources within the horizontal plane without requiring prior knowledge of boundary conditions and room geometry, making it versatile for application in different room types.

***Index Terms*—** Source localization, reverberant environments, sparse representation, sound field decomposition

1. INTRODUCTION

Sound source localization plays a crucial role in various acoustic applications, such as speech enhancement [1, 2], source separation [3,4], and sound field translation [5]. In environments with strong reverberation, the challenge of source localization experiences a considerable escalation. Several source localization methods have been applied to reverberant environments, such as beamforming [6], MUSIC [7, 8], SRP-PHAT [9], and CLEAN [10]. These methods use statistical properties of signals to estimate source positions. However, their performance declines significantly when processing narrowband sources and correlated reflections.

Recently, some sparsity-based methods, such as LASSO [11], Orthogonal Matching Pursuit (OMP) [12], and Sparse Bayesian Learning [13], have been introduced to overcome the limitation of narrowband localization by assuming sparse distribution of sources in spatial domain. Nevertheless, these methods continue to struggle in strong reverberation due to the interference of room reflections. Overcoming this challenge often requires either prior knowledge of room geometry and boundary conditions [14] or the use of sound field decomposition [15]. Sound field decomposition entails separating the sound field into a direct component and reverberant component and modeling room reflections as a sum of plane waves or spherical-harmonics [16]. For source localization tasks, the latter approach models room reflections through a sum of plane waves and estimate the source positions based on the dereverberated sound field, such as wavefield separation projector processing (WSPP) [17] and sparsity-based spherical harmonics model (S-SH) [4]. These approaches can handle a wide range of scenarios without requiring prior information. However, these methods demand a large number of microphones for modeling the reverberant component and are restricted by the geometry of the microphone array.

In this paper, we propose a similar two-step approach of sparsity based source localization that uses fewer microphones. We first dereverberate the reverberant component of the captured sound field as an equivalent plane wave decomposition model [16]. The second step models the dereverberated sound field as superposition of the sparse point-sources and determine the source positions based on the sparse equivalent source method [18]. We verify the proposed method through simulations in a conference room scenario, using a microphone array around the middle horizontal plane of the wall. It is worth noting that the specified microphone placement is designed solely for the emulation of real-world environments. In practical applications, the proposed placement only needs to surround the region of interest and is not constrained by the room geometry. The results demonstrate that the two-step approach using separate algorithms improves source localization with fewer microphones, especially in rooms with strong reverberation.

2. PROBLEM FORMULATION

Consider N sound sources in a reverberant room, along with M microphones uniformly placed affixed to the walls, as illustrated in Fig. 1. The positions of the sound sources and microphones are defined as $\boldsymbol{y}_n \equiv (x_n, y_n, z_n)$ for $n = 1, 2, \ldots, N$ and $\boldsymbol{x}_m \equiv (x_m, y_m, z_m)$ for $m = 1, 2, \ldots, M$, respectively, with respect to the coordinate origin O at the front-left-bottom corner of the room. Note, this configura-

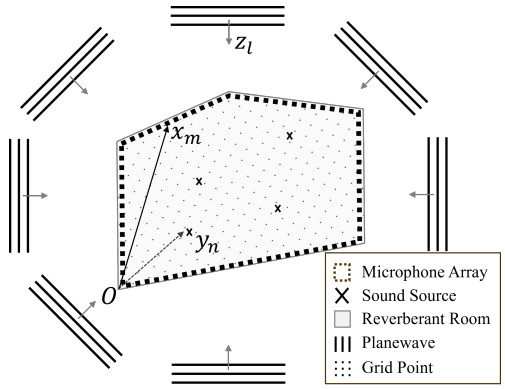

Fig. 1: Framework for the proposed method.

tion indicates that the N sources are positioned inside the microphone array.

The sound pressure received by the m^{th} microphone is:

$$s(k, \boldsymbol{x}_m) = \sum_{n=1}^{N} G(k, \boldsymbol{x}_m, \boldsymbol{y}_n)\alpha_n(k) + \mathcal{N}(k) \quad (1)$$

where $k = 2\pi f/c$ is the wave number, f is frequency, c is the speed of sound, $s(k, \boldsymbol{x}_m)$ represents the pressure, $G(k, \boldsymbol{x}_m, \boldsymbol{y}_n)$ denotes the room transfer function between the n^{th} source and the m^{th} microphone, $\alpha_n(k)$ denotes the signal produced by the n^{th} source, and $\mathcal{N}(k)$ denotes the noise term, which will be omitted for brevity in subsequent equations.

Based on sound field decomposition [15] and Vekua's theory [16, 19], any reverberant sound field can be partitioned into a sum of its particular and homogeneous solutions—equivalent to the direct and reverberant components, respectively. Within a bounded convex region, the reverberant sound field component can be well approximated using a finite number of plane wave functions distributed over a spherical region. Consequently, the sound field can be expressed as a linear combination of N direct-path point-source Green's function and L plane wave Green's function. Equation (1) can thus be decomposed as follows:

$$s(k,\boldsymbol{x}_m) \approx \underbrace{\sum_{n=1}^{N} G_0(k,\boldsymbol{x}_m,\boldsymbol{y}_n)\alpha_n(k)}_{\text{Direct}} + \underbrace{\sum_{\ell=1}^{L} W(k,\boldsymbol{x}_m,\hat{\boldsymbol{z}}_\ell)\beta_\ell(k)}_{\text{Reverberant}} \quad (2)$$

where $G_0(k, \boldsymbol{x}_m, \boldsymbol{y}_n) = e^{ik\|\boldsymbol{x}_m - \boldsymbol{y}_n\|}/(4\pi\|\boldsymbol{x}_m - \boldsymbol{y}_n\|)$ represents the direct-path Green's function between the n^{th} source and the m^{th} microphone in a free-field environment. $W(k, \boldsymbol{x}_m, \hat{\boldsymbol{z}}_\ell) = e^{-ik\hat{\boldsymbol{z}}_\ell \cdot \boldsymbol{x}_m}$ represents the ℓ^{th} plane wave Green's function at m^{th} microphone, with $\hat{\boldsymbol{z}}_\ell$ denoting the ℓ^{th} plane wave's incident direction for $\ell = 1, 2, \ldots, L$. The coefficient $\beta_\ell(k)$ represents the weight of the ℓ^{th} plane wave.

In order to find the source positions, the direct component is modeled using a dictionary of J point-sources within the room based on the sparse equivalent source method [18]. This method assumes that sound sources exhibit quantity sparsity in spatial domain, indicating $N \ll J$. Finally, equation (2) can be reformulated as follows:

$$s(k,\boldsymbol{x}_m) \approx \sum_{j=1}^{J} G_0(k,\boldsymbol{x}_m,\boldsymbol{y}_j)\alpha_j(k) + \sum_{l=1}^{L} W(k,\boldsymbol{x}_m,\hat{\boldsymbol{z}}_\ell)\beta_\ell(k) \quad (3)$$

where $G_0(k, \boldsymbol{x}_m, \boldsymbol{y}_j)$ denotes the free-field point-source Green's function between the j^{th} grid point and the m^{th} microphone. Hence, (3) is represented in matrix form as:

$$\mathbf{s} = \mathbf{G}_0\boldsymbol{\alpha} + \mathbf{W}\boldsymbol{\beta}, \quad (4)$$

where $\mathbf{s} \in \mathbb{C}^M$ denotes the measured pressure from M microphones, and $\mathbf{G_0} \in \mathbb{C}^{M\times J}$ and $\mathbf{W} \in \mathbb{C}^{M\times L}$ represent the dictionary matrices for point-sources and plane waves, respectively. The weight coefficient vectors for point-sources and plane waves are denoted as $\boldsymbol{\alpha} \in \mathbb{C}^J$ and $\boldsymbol{\beta} \in \mathbb{C}^L$.

Equation (4) is the sound field decomposition model in reverberant environments. As L becomes sufficiently large, most elements of $\boldsymbol{\beta}$ can be approximated as nearly zero. Additionally, relying on the spatial sparsity assumption for sound source distribution, the $\boldsymbol{\alpha}$ vector tends to also have few non-zero elements, given that N is significantly smaller than J. Therefore, the weight coefficients for point-sources $\boldsymbol{\alpha}$ and plane waves $\boldsymbol{\beta}$ can be determined through sparse optimization, as shown below [4]:

$$\underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\operatorname{argmin}} \|\boldsymbol{\alpha}\|_1 + \lambda\|\boldsymbol{\beta}\|_1 \quad \text{s.t. } \mathbf{s} = \mathbf{G}_0\boldsymbol{\alpha} + \mathbf{W}\boldsymbol{\beta}, \quad (5)$$

where λ is a regularization term.

The objective in this study is to estimate $\boldsymbol{\alpha}$ and determine the corresponding N source positions $\boldsymbol{y}_n$ by solving (4), while assuming that the number of sources N is known. We propose a two-step process to solve (4) next.

3. TWO-STEP SPARSE LOCALIZATION METHOD

3.1. Sound Field Decomposition

In the first step, we estimate the direct and reverberant components of the sound field. We start by rearranging (4) as:

$$\mathbf{s} = \boldsymbol{A}\boldsymbol{\gamma} \quad (6)$$

where $\boldsymbol{A} = [\mathbf{G_0}, \mathbf{W}] \in \mathbb{C}^{M\times(J+L)}$, and $\boldsymbol{\gamma} = [\boldsymbol{\alpha}, \boldsymbol{\beta}] \in \mathbb{C}^{(J+L)}$. In this context, we assume $M < (J+L)$, such that the number of measurements is fewer than the combined total modeled point-sources (grid points) and plane waves in the dictionary. Hence, the estimation of $\boldsymbol{\gamma}$ is an underdetermined problem.

We consider solving the linear regression problem (6) through Iteratively Reweighted Least Squares (IRLS) [20], exploiting an ℓ^p-norm approach by adding weights to ℓ^2-norm optimization that iteratively refine the solution's sparsity (where $0 < p \leq 2$):

$$\min_{\gamma} \sum_{i=1}^{\mathcal{L}} w_i\boldsymbol{\gamma}_i^2, \quad \text{subject to } \mathbf{s} = \boldsymbol{A}\boldsymbol{\gamma} \quad (7)$$

where $w_i = |\boldsymbol{\gamma}_i^{(v-1)}|^{p-2}$ are the weights computed from the previous iteration $\boldsymbol{\gamma}^{(v-1)}$. Hence, this iterative optimization is a ℓ^p objective function. The next iteration $\boldsymbol{\gamma}^{(v)}$ is as follows:

$$\boldsymbol{\gamma}^{(v)} = \boldsymbol{Q}_v \boldsymbol{A}^T (\boldsymbol{A}\boldsymbol{Q}_v \boldsymbol{A}^T)^{-1}\mathbf{s} \tag{8}$$

where Q_v is the diagonal matrix with $1/w_i = |\boldsymbol{\gamma}_i^{(v-1)}|^{2-p}$. We obtain the reverberant component of the sound field from the estimated weight coefficients of plane waves $\hat{\boldsymbol{\beta}}$, which is extracted from $\hat{\boldsymbol{\gamma}}$.

Therefore, the direct sound field component can be now estimated as:

$$\hat{\mathbf{s}}_0 = \mathbf{s} - \mathbf{W}\hat{\boldsymbol{\beta}} \tag{9}$$

which we use in the second step to perform source localization.

3.2. Source Localization

For the second step, we re-estimate $\boldsymbol{\alpha}$ from the dereverberated component using OMP. Given the spatial sparsity assumption for sound source distribution, it follows that $\boldsymbol{\alpha}$ is assumed to have N non-zero elements. Hence, we propose using OMP. OMP is a greedy algorithm, adapting an iterative process to select the most relevant sound source at each step [21]. By forcing inactive weights to zero, OMP improves the accuracy of source localization estimation. Utilizing the result from (9), we express the direct sound field component as:

$$\hat{\mathbf{s}}_0 = \mathbf{G}_0 \widetilde{\boldsymbol{\alpha}} \tag{10}$$

where $\widetilde{\boldsymbol{\alpha}}$ denotes the estimated point-source weight vector determined by OMP. In practice, source positions are found by selecting the weight coefficient with the highest correlation in each iteration. The OMP algorithm as follows:

Algorithm 1 Dereverberated OMP localization

Input: measurements $\hat{\mathbf{s}}_0$, point-source dictionary $\mathbf{G}_0$, number of sources N

Output: estimated weight coefficients $\widetilde{\boldsymbol{\alpha}}$, estimated source positions $\widetilde{\boldsymbol{y}}_{\boldsymbol{n}}$

$\boldsymbol{\Lambda}_0 = \emptyset, \boldsymbol{\Psi}_0 = \emptyset, \boldsymbol{g}_i = \mathbf{G}_0[:, i]$ for $i \in \{1, 2, \ldots, J\}$

for $n = 1$ to N **do**

$\widetilde{\boldsymbol{y}}_n \leftarrow \arg\max_{j \in \{1,2,\ldots,J\}} |\langle \hat{\mathbf{s}}_0, \boldsymbol{g}_j \rangle|$

$\boldsymbol{\Psi}_n \leftarrow \boldsymbol{\Psi}_{n-1} \cup \{\widetilde{\boldsymbol{y}}_n\}$

$\boldsymbol{\Lambda}_n \leftarrow \boldsymbol{\Lambda}_{n-1} \cup \{\boldsymbol{g}_j\}$

$\hat{\mathbf{s}}_0 \leftarrow \hat{\mathbf{s}}_0 - \boldsymbol{\Lambda}_n \boldsymbol{\Lambda}_n^{\dagger} \hat{\mathbf{s}}_0$

end for

$\widetilde{\boldsymbol{\alpha}}_{\boldsymbol{\Psi}_N} \leftarrow \boldsymbol{\Lambda}_N \boldsymbol{\Lambda}_N^{\dagger} \hat{\mathbf{s}}_0$

4. SIMULATION RESULTS

In this section, we evaluate the proposed method in a simulated reverberant room. To emulate practical scenarios, we focus on a conference room environment with multiple participants conversing. We assume that all sound sources are positioned at approximately 1.6 m height from the ground sitting around the conference table. Therefore, we evaluate the performance of the proposed method for horizontal plane localization as this minimizes the required microphones for a practical implementation.

We use the RIR generator toolbox [22] to simulate a $4.1 \times 6.2 \times 3.9$ m reverberant shoebox room. We position $N = 4$ sources within the height range of $1.55 \leq z \leq 1.65$ m. Then, we place a uniform rectangular microphone array with $M = 106$ microphones, spaced 0.2 m apart, affixed to the walls at a height of $z = 1.60$ m. The image source RIRs are generated for different T60s assuming that the reflection from ceilings and floor are inactive for the considered scenario. The RIRs are then convolved with clean speech signals to simulate the received source signals. We select two female and two male speech sources taken from the MS-SNSD dataset [23], each with a five-second length. The sampling frequency is 16 kHz. The measurements SNR are 30 dB. The STFT parameters are 16384 for the frame length with 50% overlap. The frequency bin we select is 1 kHz, as $k = 18.48$ while the speed of sound c is 340 m/s. For the reverberant sound field, we select $L = 3000$ plane waves and $J = 1600$ point-sources (grid points) on the $z = 1.60$ m plane, following the conference room scenario.

For comparison, we evaluate three other methods: non-dereverberation by OMP (ND-OMP), WSPP [17], and simultaneous dereverberation and localization by IRLS (D-IRLS). Specifically, ND-OMP estimates the source position directly using OMP without the dereverberation step. WSPP is a method based on plane wave decomposition using modified OMP, eliminating the ambient interference by a linear projection operator before localization. D-IRLS estimates both the direct and reverberant component simultaneously through IRLS, thereby determining $\boldsymbol{\alpha}$ by solving (6). Here, we select $L = 3000$ for D-IRLS and $L = 70$ for WSPP. Notably, WSPP is sensitive to the number of plane waves, requiring careful selection based on the number of microphones.

We first evaluate two specific cases, as shown in Fig. 2. We fix the four source positions as detailed in Fig. 2(a). We select two sets of wall reflection coefficients to compare the performance at different reverberation time (T_{60}): $[0.9, 0.93, 0.94, 0.94, 0, 0]$ and $[0.99, 0.98, 0.98, 0.99, 0, 0]$ with the reflection order of image source model as 30, equivalent to 0.75 s and 1.5 s T_{60}, respectively.

In Fig. 2, we present the estimated weight vector of point-sources $\widetilde{\boldsymbol{\alpha}}$ compared to the true source weights for both a high and extreme reverberation room. Fig. 2(a) shows the ground truth. Starting with ND-OMP in (b), we see that this method can successfully localize the sources without dereverberation when T_{60} is medium. However, the performance in (f) degrades when T_{60} is high owing to interference from room reflections. Although D-IRLS in (c) and (g) provides a

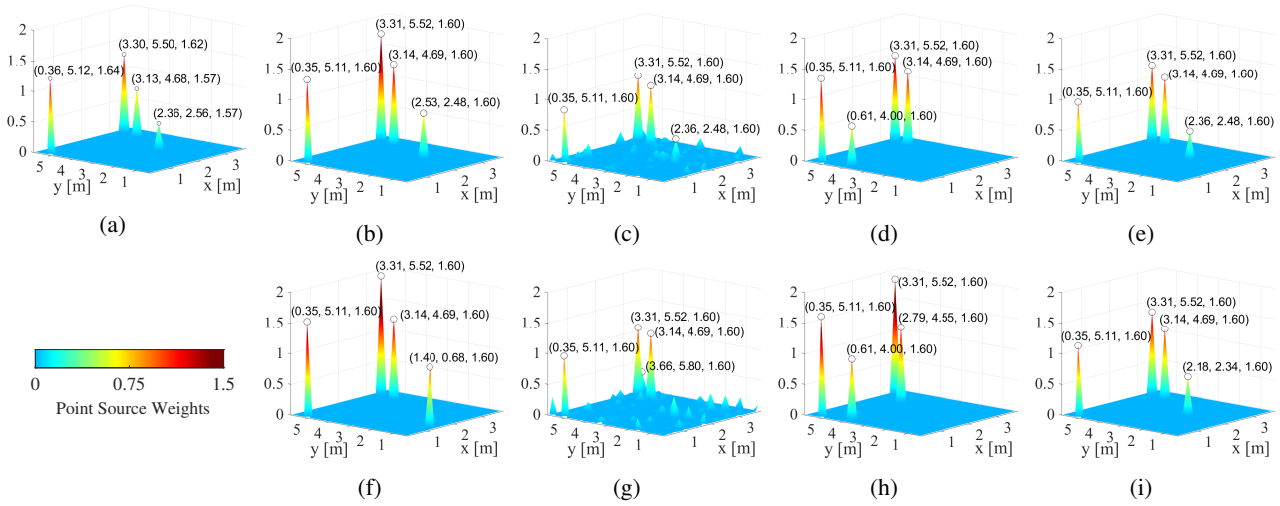

Fig. 2: Point-source weights at $f = 1000$ Hz in xy-plane of (a) ground truth, (b,f) ND-OMP, (c,g) D-IRLS, (d,h) WSPP, and (e,i) proposed, with (a) to (e) for $T_{60} = 0.75$ s and (f) to (i) for $T_{60} = 1.5$ s.

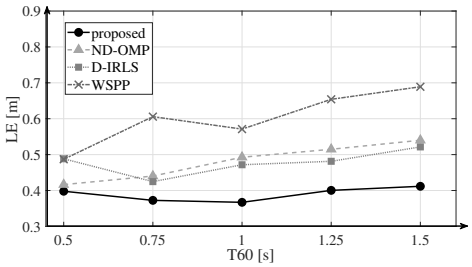

Fig. 3: Average localization errors for the four source locations at $f = 1000$ Hz averaged over 100 Monte Carlo tests.

good ability to cancel interference from room reflections, this method is unstable to obtain a sparse solution when estimating $\widetilde{\boldsymbol{\alpha}}$. In figure (d) and (h), the WSPP method is typically effective with a grid-point microphone array [12], but struggles in our setup due to its constrained microphone placements.

The proposed method as observed in (e) and (i) is seen to have the best performance. The two-step approach offers robust localization by combining the advantages of IRLS and OMP. IRLS excels at solving underdetermined systems, so it is not necessary to comply $L \simeq 2\lceil kr \rceil + 1$, where r is the measurement area radius, for optimal dereverberation [16]. This enhances dereverberation performance and prevents overfitting of the reverberant component, allowing flexibility in the number of plane waves to deal with different frequency bins. OMP then provides a strict sparse solution, enhancing source localization and enabling source loudness estimation.

In the second evaluation, we compare the average localization errors (LE) across 100 Monte Carlo test samples for different reverberation scenarios. We randomly placed the four sources inside the conference room at a height range $1.55 \leq z \leq 1.65$ m. The T_{60} is varied from 0.5 to 1.5 s. We define the LE between the estimated and true source positions to evaluate the performance of the localization methods as:

$$\mathbf{LE} = \tfrac{1}{N} \textstyle\sum_{n=1}^{N} \|\widetilde{\boldsymbol{y}}_{\boldsymbol{n}} - \boldsymbol{y}_n\|_2 \tag{11}$$

The results of Fig. 3 show that the proposed method provides robust source position estimation. As T_{60} increases, the method maintains stable LE values, while the performances of other methods degrade. However, we note that the average LE of the proposed method at $T_{60} = 0.5$ s remains relatively high, primarily due to the magnitude differences among different sources. Specifically, if the magnitude of one source is much lower than of the other sources, localizing this particular source becomes challenging because it might be regarded as noise when using sparse recovery methods. The presented results illustrated the robust performance achieved in horizontal-plane localization. However, it is worth to note that the proposed method can be extended to 3D localization with a 3D microphone array.

5. CONCLUSION

In this paper, we have introduced an enhanced method for narrowband source localization in reverberant environments. Based on sparse representation and plane wave decomposition, the two-step approach improves the localization accuracy by estimating the direct and reverberant component separately. The simulation results show that the proposed method can effectively localize multiple sources with a reduced number of measurements, particularly in scenarios with high T_{60}. Moreover, our method overcomes the overfitting problem in plane wave decomposition, allowing for a more flexible and straightforward determination of plane wave quantities. The scope of future work includes resolving the magnitude issue in sparse recovery algorithms and considering wideband scenarios to reduce microphones further.

6. REFERENCES

[1] Y. Peled and B. Rafaely, "Linearly-constrained minimum-variance method for spherical microphone arrays based on plane-wave decomposition of the sound field," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 12, pp. 2532–2540, 2013.

[2] A. Xenaki, J. B. Boldt, and M. G. Christensen, "Sound source localization and speech enhancement with sparse bayesian learning beamforming," *J. Acoust. Soc. Amer.*, vol. 143, no. 6, pp. 3912–3921, 2018.

[3] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "Sound field separation in a mixed acoustic environment using a sparse array of higher order spherical microphones," in *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)*. IEEE, 2017, pp. 151–155.

[4] M. Pezzoli, M. Cobos, F. Antonacci, and A. Sarti, "Sparsity-based sound field separation in the spherical harmonics domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 1051–1055.

[5] L. McCormack, A. Politis, T. McKenzie, C. Hold, and V. Pulkki, "Object-based six-degrees-of-freedom rendering of sound scenes captured with multiple ambisonic receivers," *J. Audio Eng. Soc*, vol. 70, no. 5, pp. 355–372, 2022.

[6] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[7] E. D. Di Claudio and R. Parisi, "Waves: Weighted average of signal subspaces for robust wideband direction finding," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2179–2191, 2001.

[8] L. Birnie, T. D. Abhayapala, H. Chen, and P. N. Samarasinghe, "Sound source localization in a reverberant room using harmonic based music," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2019, pp. 651–655.

[9] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone arrays: signal processing techniques and applications*, pp. 157–180. Springer, 2001.

[10] J. A. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astronomy and Astrophysics Supplement*, vol. 15, pp. 417, 1974.

[11] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.

[12] G. Chardon and L. Daudet, "Source localization in reverberant rooms using sparse modeling and narrowband measurements," *arXiv preprint arXiv:1307.4894*, 2013.

[13] G. Ping, E. Fernandez-Grande, P. Gerstoft, and Z. Chu, "Three-dimensional source localization using sparse bayesian learning on a spherical microphone array," *J. Acoust. Soc. Amer.*, vol. 147, no. 6, pp. 3895–3904, 2020.

[14] J. Le Roux, P. T. Boufounos, K. Kang, and J. R. Hershey, "Source localization in reverberant environments using sparse optimization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 4310–4314.

[15] S. Koyama and L. Daudet, "Sparse representation of a spatial sound field in a reverberant environment," *IEEE J. Selected Topics Signal Process.*, vol. 13, no. 1, pp. 172–184, 2019.

[16] A. Moiola, R. Hiptmair, and I. Perugia, "Plane wave approximation of homogeneous helmholtz solutions," *Zeitschrift für angewandte Mathematik und Physik*, vol. 62, pp. 809–837, 2011.

[17] G. Chardon, T. Nowakowski, J. de Rosny, and L. Daudet, "A blind dereverberation method for narrowband source localization," *IEEE J. Selected Topics Signal Process.*, vol. 9, no. 5, pp. 815–824, 2015.

[18] E. Fernandez-Grande, A. Xenaki, and P. Gerstoft, "A sparse equivalent source method for near-field acoustic holography," *J. Acoust. Soc. Amer.*, vol. 141, no. 1, pp. 532–542, 2017.

[19] I. N. Vekua, D. E. Brown, and A. B. Tayler, "New methods for solving elliptic equations," *North Holland*, 1967.

[20] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2008, pp. 3869–3872.

[21] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[22] E. A. P. Habets, "Room impulse response generator," *Technische Universiteit Eindhoven, Tech. Rep*, vol. 2, no. 2.4, pp. 1, 2006.

[23] C. K. A. Reddy, E. Beyrami, J. Pool, R. Cutler, S. Srinivasan, and J. Gehrke, "A scalable noisy speech dataset and online subjective test framework," *Proc. Interspeech*, pp. 1816–1820, 2019.