ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10445940

SPARSE SOUND FIELD REPRESENTATION USING COMPLEX ORTHOGONAL MATCHING PURSUIT

Shaoheng Xu[1*], *Jihui (Aimee) Zhang*[1,2], *Thushara D. Abhayapala*[1], *Amy Bastine*[1], *Wei-Ting Lai*[1], *Prasanga N. Samarasinghe*[1]

[1]The Australian National University, Australia
[2] University of Southampton, UK
[*]Shaoheng.Xu@anu.edu.au

ABSTRACT

Spatial audio reproduction and translation for virtual, augmented, and extended reality applications require an efficient representation of the recorded sound fields. In this paper, we investigate the possible sparse representations of the sound field recorded by multiple microphones in reverberant environments. We combine the Complex Orthogonal Matching Pursuit (COMP) algorithm with the concept of distributed virtual sound sources to propose a sparse sound field representation. The technique uses recordings from a grid of microphones and transforms them into a sparse representation featuring a selected set of active virtual sources. Using simulation, we evaluate the proposed COMP approach with LASSO and IRLS methods in a reverberant room of regular size with a ceiling-mounted microphone array.

Index Terms— Sparse Representation, Sound Field Reproduction, Sparse Sound Field Decomposition, Complex Orthogonal Matching Pursuit

1. INTRODUCTION

With the increasing popularity of the metaverse and virtual reality (VR) applications, there is a growing demand for capturing the real-world acoustic environment and reproducing it on portable VR devices, enhancing users' experience in navigating audio-visual scenes [1]. Hence, in this paper, our focus lies on finding a compact representation of the recorded spatial audio environment for the reproduction of the sound field. While there are existing methods in this area [2, 3, 4], the research on combining a limited number of microphone recordings and reproducing the sound field at multiple frequencies over a large region in a regular-size reverberant room needs further investigation.

Recording and reproducing a three-dimensional sound scene in a reverberant room with several sound sources is a challenging task. This task requires dense spatial sampling of sound waves across the room using numerous distributed microphones arranged in some specific configurations. The problem becomes more impractical at high frequencies, as it is infeasible to employ a great number of microphones to satisfy the sampling theorem [5]. That is where sparse algorithms come into play, addressing the under-sampled reproduction problem with an efficient representation of the sound field.

Existing works on sound field translation and reproduction have employed several sparse representation methods. In [6], the authors propose using the least-absolute shrinkage and selection operator (LASSO) for sound field synthesis. In [2], the reverberant sound field is represented by the summation of weighted plane waves on a two-dimensional plane. In [7], the authors measure the room transfer function over a pre-defined spatial region using a few loudspeakers and microphones, which is then parameterized by a finite set of coefficients. In [3], the authors propose a compressive sensing (CS) reconstruction method for inferring the sound field outside the area of measurement. In [4], a mix of near-field and far-field virtual sources with sparse constrained expansion is used for binaural sound field reproduction, validated by perceptual tests in [8]. In [9], point source decomposition is conducted with multiple sparse algorithms, including LASSO, iteratively reweighted least squares (IRLS) [10] and matching pursuit (MP). However, none of the above works consider finding a compact and efficient representation of the sound field over multiple frequencies.

Orthogonal matching pursuit (OMP) [11] is another useful algorithm used by the computer science community for sparse signal recovery. It finds widespread applications in medical imaging processing [12], Multiple-Input Multiple-Output (MIMO) communication system [13], and polynomial chaos expansions [14]. However, OMP only takes real-valued inputs, making it unsuitable for complex signal processing. Hence, in [15], the authors propose a complex OMP algorithm (COMP) to sparsely solve decomposition problems for complex cases. Since COMP outputs a compact representation of the input signals, it is adopted in wireless body sensor network [16, 17] and the compact representation of speech signals [18]. However, it has never been used for sound field processing.

Therefore, in this paper, we introduce COMP [15] into the field of sound field reproduction. We aim to find a sparse and compact representation of the recordings from spatially distributed microphones using COMP, and then estimate the audio scenes over a large spatial area in a reverberant room.

2. PROBLEM FORMULATION

Consider a sound field generated by R real sound sources sparsely distributed in the room. We use Q omnidirectional microphone devices to measure the sound field, and the q^{th} device is located at position $\mathbf{y}_q$, $q = 1, \ldots, Q$. The received signals at the microphones are defined as $\mathbf{s}(k) = [s^{(1)}(k),\ s^{(2)}(k),\ \cdots,\ s^{(Q)}(k)]^T$, where $(\cdot)^T$ represents the transpose operator, and k is the wave number.

To estimate the acoustic scene over a larger region of interest (ROI), we use a grid of distributed virtual point sources to represent the sound field. The estimated sound field generated by the virtual sources at the q^{th} microphone can be represented by

$$\hat{s}^{(q)}(k) = \mathbf{G}_q(k) \cdot \mathbf{w}(k), \tag{1}$$

where $\mathbf{w}(k) = [w^{(1)}(k),\ w^{(2)}(k),\ \cdots,\ w^{(P)}(k)]^T$ are the complex weights for the virtual sources in the frequency domain, $\mathbf{G}_q(k) = [G_{q,1}(k),\ \cdots,\ G_{q,p}(k),\ \cdots,\ G_{q,P}(k)]$ is the transfer functions between virtual sources and the q^{th} microphone, and P is the number of virtual sources.

As the real sound sources are sparsely distributed within the room, the solution of $\mathbf{w}(k)$ should be sparse as well. Therefore, we minimize the squared errors between the recorded sound field and the estimated sound field, with a constraint on the number of non-zero elements in complex weights $\mathbf{w}(k)$,

$$\min_{\mathbf{w}} \|\mathbf{s}(k) - \mathbf{G}(k) \cdot \mathbf{w}(k)\|_2^2, \quad \text{s.t.} \quad \|\mathbf{w}(k)\|_0 \leq T_{\mathbf{w}}, \tag{2}$$

where $T_{\mathbf{w}} \in \mathbb{Z}^+$ is the pre-defined sparsity.

Hence, the sparse sound field representation problem addressed in this paper can be represented as: Given microphone recordings $\mathbf{s}(k)$ and transfer function $\mathbf{G}(k)$, find a suitable sparse method to solve the minimization problem in (2).

In this work, we distribute the virtual sources inside and encompassing the room to model both direct-path and reflections. This arrangement simplifies the estimation process by only considering the free-field propagation from the virtual sources to the microphone devices, where [19]

$$G_{q,p}(k) = \frac{e^{(ik \cdot \|\mathbf{y}_q - \mathbf{x}_p\|)}}{4\pi \|\mathbf{y}_q - \mathbf{x}_p\|}. \tag{3}$$

3. SOUND FIELD REPRODUCTION USING COMPLEX ORTHOGONAL MATCHING PURSUIT METHOD

There are many existing methods for solving the optimization problem in (2) [10] [6]. However, when the frequency of interest is relatively high, the wavelength becomes smaller, and thus the results of sound field estimation are more sensitive to the locations of the active virtual point sources. Therefore, the separation between virtual sources should be smaller, and more virtual sources are required in the reproduction process. The number of required virtual point sources would be significantly larger than the number of microphone devices, i.e., $P \gg Q$. In other words, the problem addressed in (2) becomes an under-determined regression problem with a sparse solution.

Therefore, OMP [11] becomes a suitable method for solving the regression problem in (2). It can recover a high-dimensional sparse signal from a few numbers of linear measurements with noise [11]. However, OMP only takes real inputs and generates real outputs. For sound field decomposition in the frequency domain, we need to divide the complex inputs into real and imaginary parts, and then rewrite them into larger vectors or matrices. As a result, OMP evaluates the effects of real and imaginary parts of inputs separately. Most of the complex outputs from OMP are assigned with complex values that only have non-zero real or imaginary part, which reduces effectiveness in sound field decomposition.

For that reason, we replace the original OMP algorithm by COMP in sound field reproduction [15]. The latter accepts complex values for inputs and outputs, and provides a sparse approximation of $\mathbf{w}(k)$ with complex weights in the frequency (complex) domain.

COMP is an iterative greedy algorithm. It returns a sparse solution by activating the chosen virtual sources one-by-one in an iterative manner. Each selection of virtual sources is made based on the optimization of the information from the previous iteration. In each iteration, the algorithm consists of two main steps: the "Matching" step and the "Orthogonal" step.

Firstly, during the "Matching" step of the i^{th} iteration, the algorithm computes the correlations between $\mathbf{r}_{i-1}$ (the previous residual of sound pressure) and each column of $\mathbf{G}$. In this way, it estimates which column $\mathbf{g}_{\lambda_i}$ and its corresponding virtual sound source contributes the most to the previous residual of sound pressure $\mathbf{r}_{i-1}$. Then, it picks up the index number λ_i of the selected virtual source for the i^{th} iteration,

$$\lambda_i = \underset{\lambda_i}{\operatorname{argmax}} |{\mathbf{g}_{\lambda_i}}^H \cdot \mathbf{r}_{i-1}|, \tag{4}$$

where $\mathbf{g}_{\lambda_i}$ is the λ_i^{th} column of $\mathbf{G}$, and $(\cdot)^H$ represents the conjugate transpose operator.

Secondly, in the "Orthogonal" step of the i^{th} iteration, the algorithm updates the index set of selected virtual sources Λ_i with the selected λ_i from the previous "Matching" step,

$$\Lambda_i = \Lambda_{i-1} \cup \{\lambda_i\}. \tag{5}$$

Thirdly, COMP extracts the selected column $\mathbf{g}_{\lambda_i}$ of the current iteration from $\mathbf{G}$ and updates the current residual $\mathbf{r}_i$ by removing effects of $\mathbf{g}_{\lambda_i}$ on $\mathbf{r}_{i-1}$, i.e., removing the effects of the i^{th} selected virtual source on the remaining sound pressure,

$$\mathbf{M}_i = \mathbf{G}(\Lambda_i) \cdot \left(\mathbf{G}(\Lambda_i)^H \mathbf{G}(\Lambda_i)\right)^{-1} \cdot \mathbf{G}(\Lambda_i)^H, \tag{6}$$

$$\mathbf{r}_i = (\mathbf{I} - \mathbf{M}_i) \cdot \mathbf{s}, \tag{7}$$

where $\mathbf{M}_i$ is a Q-by-Q matrix that represents the projection onto the linear vector space spanned by the columns of $\mathbf{G}(\Lambda_i)$, $\mathbf{I}$ denotes the identity matrix, and $(\cdot)^{-1}$ denotes the inverse of a square matrix.

Lastly, when the number of iterations reaches the desired sparsity of $\mathbf{w}(s)$, $T_{\mathbf{w}}$, or when the ℓ_2 norm of $\mathbf{r}_i$ is less than the desired threshold value $\tau \in \mathbb{R}^+$, the algorithm exits the loop and estimates the complex weights of the active virtual sources using the existing selections,

$$\mathbf{w}(\Lambda_{T_{\mathbf{w}}}) = \mathbf{G}(\Lambda_{T_{\mathbf{w}}})^\dagger \cdot \mathbf{s}, \tag{8}$$

where $(\cdot)^\dagger$ represents the pseudo-inverse.

To use the COMP algorithm, we need to initialize the residual $\mathbf{r}_0$, the index set Λ_0, and the iteration counter i as follows:

$$\{\mathbf{r}_0, \Lambda_0, i\} = \{\mathbf{s}, \emptyset, 1\}. \tag{9}$$

Selecting a suitable value for $T_{\mathbf{w}}$ is essential to obtain a useful sparse solution from COMP. Specifically, $T_{\mathbf{w}}$ should be less than Q, and approximately equal to the number of real sources and their image sources.

Algorithm 1: Orthogonal Matching Pursuit in Complex Frequency Domain

Data: $\mathbf{s}(k)$, $\mathbf{G}(k)$, $T_{\mathbf{w}}$, and τ.
Result: $\mathbf{w}(k)$.
Steps:
1. Initialize $\mathbf{r}_0$, Λ_0 and i according to Eq. (9).
2. **for** $i = 1$ **to** $T_{\mathbf{w}}$ **do**
   a. Select the column from $\mathbf{G}(k)$ using Eq. (4) and update Λ_i using Eq. (5).
   b. Update $\mathbf{r}_i$ using Eq. (6) and Eq. (7).
   if $i = T_{\mathbf{w}}$ *OR* $||\mathbf{r}_i||_2 \leq \tau$ **then**
      c. Exit the **for** loop.
3. Obtain the sparse solution of $\mathbf{w}(k)$ using Eq. (8).

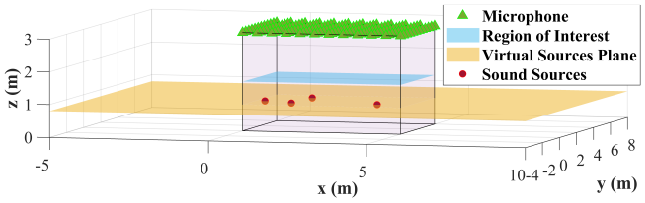

Fig. 1: The setup of simulation, where the green triangles represent microphone devices, the blue plane represents the area of interest, the yellow plane represents the virtual sources plane, and the red circles represent the sound sources.

4. SIMULATION STUDY

We firstly evaluate the performance of the COMP algorithm in sound field reproduction in a regular-size room at a specific frequency, and then compare its results with other sparse algorithms. After that, we evaluate and compare the performance of different algorithms over a large range of frequencies.

4.1. Simulation Setups

Consider a rectangular room with 5 meters in width (x-axis), 4 meters in length (y-axis), and 3 meters in height (z-axis) (as shown in Fig. 1). The 4 vertical walls are assigned with equal reflection coefficients of $[0.6, 0.6, 0.6, 0.6]$, while that of the ceiling and floor are assigned to be $[0, 0]$. We adopt the image-source method [20] to simulate the room acoustic and the depth of reflection is set to be 1.

In the room, there is a ceiling-mounted microphone array featuring a 12-by-12 rectangular grid of microphone devices, spaced at regular intervals. Additionally, 4 point sources are sparsely distributed, representing a real-world scenario with 3 people sitting beside a noise source (e.g., a noisy computer) and engaging in conversation. These point sources are located at: $[0.3, 1.5, 0.8]$, $[1.3, 0.9, 0.8]$, $[1.4, 3.0, 0.8]$ and $[4.0, 1.0, 0.8]$, respectively, with each assigned an equal weight of 1.

Virtual point sources are positioned on the same plane as the real sources at a height of 0.8 meters. They form a 120-by-120 rectangular grid evenly distributed over a 15m-by-12m area, which is centered at $[2.5, 2, 0.8]$ (the yellow plane in Fig. 1). This setup ensures the virtual sources plane can include all real sources and their potential image sources.

While the proposed method is capable of sound field reproduction throughout the entire room, for this study, we set the ROI to be the middle horizontal plane of the room, i.e., 5m-by-4m area centred at $[2.5, 2, 1.5]$ (the light-blue plane in Fig. 1), to provide a clearer view of the results.

As introduced above, to simplify computational complexity, all real and virtual sources are assumed to be located on the same plane, 0.8 meters above the floor. This setup simulates a conference room scenario where a few people are sitting and conversing together. For the convenience of this paper, we remove the reflections on the floor and ceiling, and only simulate the first-order reflections on the 4 vertical walls, reducing the number of required virtual sources. If the reflection coefficients of the floor and ceiling are non-zero, or if the real sources are not located on the same plane, we would need a 3D cubic grid of virtual sources, which has a significantly greater number of points than the current rectangular grid used for this simulation.

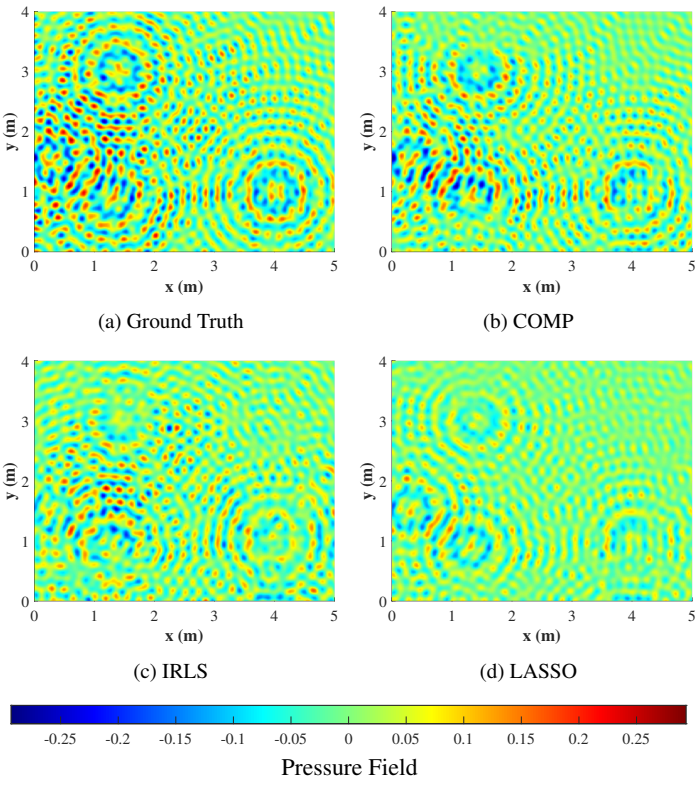

Fig. 2: The results of sound field reproduction using different methods at 2000 Hz: (a) The ground truth of sound field on the ROI, the estimated sound field on the ROI using (b) COMP, (c) IRLS, and (d) LASSO.

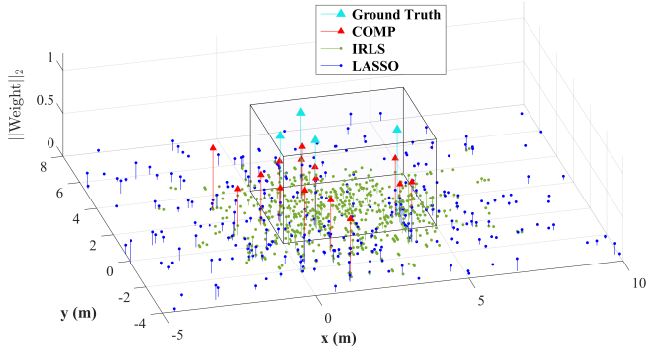

Fig. 3: The results of active virtual point sources using different methods at 2000 Hz, where the ground truth are marked with cyan triangles, the results using COMP are marked with red triangles, the results using IRLS are marked with green dots, and the results using LASSO are marked with blue dots. For clarity and precision, virtual sources with energy values less than 10^{-3} have been excluded from this plot.

4.2. Results Analysis

Initially, simulations are conducted at a single frequency of 2000 Hz, and we compare the performance of COMP algorithm with two other sparse algorithms, IRLS [10] and LASSO [6]. The sparsity $T_{\mathbf{w}}$

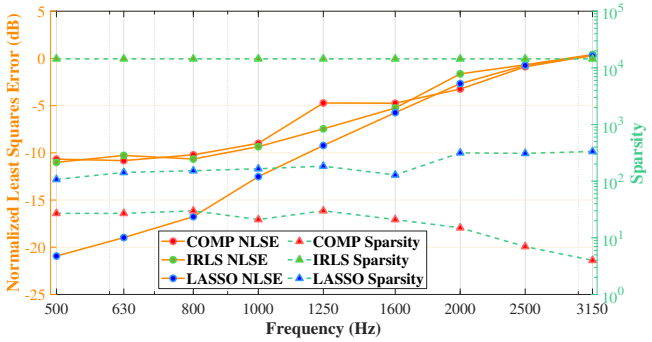

Fig. 4: The normalized least square error of sound field reproduction (Left) and the sparsity (Right) for different algorithms at different frequencies.

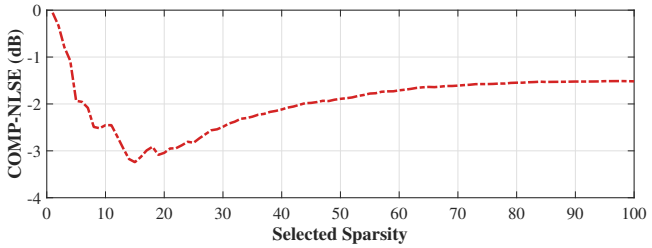

Fig. 5: The NLSE of COMP with various selected sparsity $T_{\mathbf{w}}$ at 2000 Hz.

of COMP is heuristically set to be 15, the norm of IRLS is set to be 1, and the tuning parameter that controls the bias-variance trade-off of LASSO, α_{LASSO}, is chosen by estimating its best value via 10-fold cross-validation.

As shown in Fig. 2b, the reproduction results using COMP preserve most of the directional patterns of the original sound field (Fig. 2a) over the entire ROI. We can clearly observe the direct path, as well as reflections of the 4 real sources, and the strength of the sound field is similar to the ground truth. In contrast, the IRLS (Fig. 2c) and LASSO (Fig. 2d) algorithms can partially replicate the patterns around the top-right corner and the area directly above the real sources. However, the strengths of the reproduced sound fields are less than the original one.

The comparisons between the real sources and the active virtual sources at 2000 Hz using different algorithms are plotted in Fig. 3. This result corroborates the reproduction performance seen in Fig. 2. As shown in Fig. 3, COMP (in red triangles) sparsely activates a few virtual sources around the real sources with similar energy. It also activates a few points outside the room, representing the image sources of the real ones. In contrast, IRLS (in green dots) and LASSO (in blue dots) activate a large number of virtual sources with smaller energies over the virtual sources plane. For low frequencies, this may be acceptable, as the wavelength is relatively large, and the adjacent active points can be considered integratedly as a large point source. However, for high frequencies, the wavelength is shorter than the separation between virtual sources. Consequently, the adjacent active virtual sources from IRLS and LASSO algorithms may cancel out each other in the sound field reproduction.

We also calculate the normalized least square errors (NLSE) of different algorithms over the entire ROI at different frequencies using the formula:

$$\epsilon(f) = 20\log_{10}\left(\frac{\sum_{\ell=1}^{L}\|P_{\text{ROI}}(\mathbf{z}_\ell, f) - \hat{P}_{\text{ROI}}(\mathbf{z}_\ell, f)\|_2}{\sum_{\ell=1}^{L}\|P_{\text{ROI}}(\mathbf{z}_\ell, f)\|_2}\right), \quad (10)$$

where f is the simulated frequency, L is the number of simulated points on the ROI, $\mathbf{z}_\ell$ is the ℓ^{th} point on the ROI with $\ell = 1, \ldots, L$, P_{ROI} is the ground truth of sound field, and $\hat{P}_{\text{ROI}}$ is the estimated sound field over the ROI.

We evaluate the three algorithms using (10) at various frequencies ranging from 500 Hz to 3150 Hz, which together form an octave band with a $1/3$ octave middle frequency. The parameters of different algorithms are optimized heuristically. At each frequency, we repeat the simulations 5 times and record the averaged NLSE for each algorithm. The results are summarised in Fig. 4. Additionally, we record the number of active virtual sources for different methods over the frequency range in Fig. 4.

The NLSE of all three algorithms increases as the simulated frequency increases. At low frequencies (before 1600 Hz), LASSO provides the smallest NLSE, while that of IRLS and COMP are roughly the same. After 1700 Hz, COMP slightly outperforms LASSO and IRLS in terms of NLSE. As the simulated frequency increases, the differences between each method in NLSE are minimized. Additionally, as shown in Fig. 4, the solution using COMP is more compact. COMP only activates a few virtual sources across the entire frequency range, compared to LASSO, which uses hundreds of virtual sources (varies from 110 to 336), and IRLS, which uses thousands of virtual sources (constant at 14400).

We also evaluate the effects of the selected sparsity $T_{\mathbf{w}}$ on NLSE in Fig. 5. Initially, as $T_{\mathbf{w}}$ increases from 1 to 15, the NLSE drops quickly from -0.05 dB to -3.24 dB. After reaching its minimum at $T_{\mathbf{w}} = 15$, NLSE gradually increases as $T_{\mathbf{w}}$ continues to rise. Hereafter, if $T_{\mathbf{w}}$ is greater than a specific value (e.g., 65), the reproduction results will not change significantly if we further increase $T_{\mathbf{w}}$.

The simulation results above prove that COMP is a useful tool for sound field reproduction as it preserves most of the main directional patterns of the sound field (Fig. 2) with acceptable NLSE across the frequency range of interest (Fig. 4). Additionally, COMP provides a more compact representation of the recorded sound field by activating a few virtual sources with approximately equal energy around the sound sources (Fig. 3). Moreover, while COMP has an optimum sparsity, beyond a certain value, its NLSE is not sensitive to the chosen sparsity $T_{\mathbf{w}}$ (Fig. 5), making it more robust across the frequency range. In contrast, for IRLS and LASSO, users need to carefully set parameters to avoid notable NLSE at specific frequencies. Finally, the computation time of COMP is notably faster than that of IRLS and LASSO, which are approximately 2.33s, 29.26s, and 4.41s, respectively, on our simulation device (5900X at 4.5 GHz and DDR4 RAM at 3600 MHz). This illustrates the potential of implementing the COMP algorithm on low-computing-power devices.

5. CONCLUSION

In summary, this paper presents a novel method to obtain a sparse representation of complex sound fields from microphone recordings using COMP. Unlike traditional methods, COMP has a special orthogonalization mechanism tailored for complex dictionaries. Our experiment simulates a reverberant room with a ceiling-mounted microphone array and distributed sound sources. We evaluate performance based on the estimated pressure field across the ROI, NLSE, sparsity of solutions, and computation time. The results demonstrate the efficacy of our approach, making it a promising solution for spatial sound field reproduction with limited numbers of microphones, reduced run-time, and a more compact representation.

6. REFERENCES

[1] B. Rafaely, V. Tourbabin, E. Habets, Z. Ben-Hur, H. Lee, H. Gamper, L. Arbel, L. Birnie, T.D. Abhayapala, and P.N. Samarasinghe, "Spatial audio signal processing for binaural reproduction of recorded acoustic scenes - review and challenges," *Acta Acust. United Acust.*, vol. 6, 10 2022.

[2] W. Jin and W.B. Kleijn, "Theory and design of multizone soundfield reproduction using sparse methods," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 12, pp. 2343–2355, 2015.

[3] S.A. Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3770–3779, 06 2018.

[4] L. Birnie, T.D. Abhayapala, P.N. Samarasinghe, and V. Tourbabin, "Sound field translation methods for binaural reproduction," in *IEEE Workshop Appl. Signal Process. Audio Acoust.* IEEE, 2019, pp. 140–144.

[5] P.N. Samarasinghe, T.D. Abhayapala, and M. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 3, pp. 647–658, 2014.

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. Ser. B.*, vol. 58, no. 1, pp. 267–288, 1996.

[7] P.N. Samarasinghe, T.D. Abhayapala, M. Poletti, and T. Betlehem, "An efficient parameterization of the room transfer function," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 12, pp. 2217–2227, Dec 2015.

[8] L. Birnie, T.D. Abhayapala, V. Tourbabin, and P.N. Samarasinghe, "Mixed source sound field translation for virtual binaural application with perceptual validation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 1188–1203, 2021.

[9] Y. Hu, P.N. Samarasinghe, T.D. Abhayapala, and G. Dickins, "Modeling characteristics of real loudspeakers using various acoustic models: Modal-domain approaches," in *Int. Conf. Acoust., Speech, Signal Process.*, 2019, pp. 561–565.

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

[11] T.T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.

[12] J. Liu, Y. Hu, J. Yang, Y. Chen, H. Shu, L. Luo, Q. Feng, Z. Gui, and G. Coatrieux, "3d feature constrained reconstruction for low-dose ct imaging," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 5, pp. 1232–1247, 2018.

[13] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R.W. Heath, "Hybrid mimo architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, vol. 4, pp. 247–267, 2016.

[14] J.D. Jakeman, M.S. Eldred, and K. Sargsyan, "Enhancing $\ell 1$-minimization estimates of polynomial chaos expansions using basis selection," *J. Comput. Phys.*, vol. 289, pp. 18–34, 2015.

[15] R. Fan, Q. Wan, Y. Liu, H. Chen, and X. Zhang, "Complex orthogonal matching pursuit and its exact recovery conditions," *arXiv:1206.2197*, 2012.

[16] A. Ravelomanantsoa, H. Rabah, and A. Rouane, "Systemc-ams based virtual prototyping of wireless body sensor network using compressed sensing," in *Int. Conf. Microelectron.*, 2013, pp. 1–4.

[17] A. Ravelomanantsoa, H. Rabah, and A. Rouane, "Simple and efficient compressed sensing encoder for wireless body area network," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 12, pp. 2973–2982, 2014.

[18] L.-C. Kwek, A. W.-C. Tan, H.-S. Lim, C.-H. Tan, and K.A. Alaghbari, "Sparse representation and reproduction of speech signals in complex fourier basis," *Int. J. Speech Technol.*, vol. 25, no. 1, pp. 211–217, 2022.

[19] E.G. Williams and J.A. Mann III, "Fourier acoustics: sound radiation and nearfield acoustical holography," 2000.

[20] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950, 04 1979.