Regularized Multi-Measurement Sparse Decomposition for Time-Domain Ambisonics Upscaling

Gyanajyoti Routray*, Priyadarshini Dwivedi, and Rajesh M. Hegde
SRM University AP Amaravati*, Indian Institute of Technology Kanpur India.
Email: gyanajyoti.r@srmap.edu.in*, {priyadw, rhegde}@iitk.ac.in

***Abstract*—Ambisonics is a powerful technique for capturing and reproducing spatial sound within the spherical harmonic domain. While low-order Ambisonics (B-format) works well at lower frequencies, its spatial resolution diminishes as frequencies increase. In contrast, higher-order Ambisonics (HOA) offers greater spatial accuracy but requires more complex hardware and increased storage. To overcome these limitations, this study investigates upscaling B-format signals to HOA using the Regularized Multiple Mearurement FOCal Underdetermined System Solver (RM-FOCUSS) approach. Regularization plays a pivotal role in addressing the sparse inverse problem, preventing overfitting, and enhancing the robustness and efficiency of the method compared to traditional sequential forward techniques. By incorporating regularization, the proposed technique significantly improves the upscaling process by minimizing diversity measures. The method's performance is assessed through mean square error metrics and validated via subjective listening tests, showing superior spatial resolution and overall performance compared to existing state-of-the-art methods.**

***Index Terms*—Higher order Ambisonics, Spatial audio imaging, FOCUSS, Diversity regularization.**

I. Introduction

Spatial audio is recorded and played back in 3D space using Ambisonics, which utilizes spherical harmonics (SH) coefficients to represent the spatial sound field [1], [2]. Higher-order Ambisonics (HOA) provides more detailed spatial information but faces challenges related to hardware complexity and storage requirements. The spatial resolution of the reconstructed sound field is limited by the HOA signal sequence and the source frequency. The "sweet spot," where low-frequency sounds are accurately reproduced, diminishes in radius and spatial resolution as frequency increases. This relationship is described by $N = \lceil kr_0 \rceil$, where N is the HOA order, r_0 is the sweet spot radius, $k = \frac{2\pi f}{c}$ is the wave number, and $c = 343$ m/s is the speed of sound [3]. Consequently, the HOA order at specific frequencies affects spatial sharpness.

Upscaling the order of Ambisonics is crucial for several reasons: (i) Higher-order Ambisonics recordings captures the spatial cues effectively to enhancing spatial accuracy. (ii) Upscaling increases the immersive quality of the sound field. (iii) Higher order improves sound source localization, which embedded the directional cues in rendering and vital for immersive audio experiences. (iv) The higher-order system captures the contents of the lower order preserving fidelity or spatial accuracy. Emphasis operator is utilized to improve the spatial resolution [4]–[6] or by increasing the order of encoded Ambisonic signals [7]–[17]. Emphasis operators localize the sources and conserve the energy [4], [5]. Sparse methods are discussed in [7]–[16], where solutions are obtained using either uniform overcomplete spherical harmonics basis matrices or non-uniform spherical harmonics dictionaries [16]. These approaches rely on dictionary matrix conditioning. Additionally, learning-based techniques, such as those presented in [17] and [18], use neural networks with dual path transform functions to upscale sampling rates and layers, though they are computationally intensive.

This work is the extension of [19], [20] by introducing regularization on multiple measurement vector based diversity minimization technique FOCUSS for upscaling. Regularization effectively handles the sparse inverse problem and improve robustness and efficiency. Compared to frequency domain upscaling [7], this technique is implemented in time-domain, which is computationally efficient and well-suited for broadband audio sources. The proposed approach offers advantages in efficiency, flexibility, and adaptability, with improved estimation quality through iterative optimization. These features are particularly useful in applications requiring source isolation. The key contributions of this approach is summarized as

i The regularization within M-FOCUSS allows for better handling of the sparse inverse problem, resulting in enhanced spatial accuracy and sharpness, especially at higher frequencies.
ii Regularization prevents overfitting, enabling more robust upscaling with improved generalization across diverse acoustic environments.
iii The technique effectively reduces diversity in signal estimation, ensuring a more consistent and precise reconstruction of the spatial sound field.
iv RM-FOCUSS offers an efficient and scalable solution for upscaling B-format signals to HOA, addressing hardware and storage constraints without compromising on sound quality.

This method outperforms traditional approaches in both objective and subjective evaluations, demonstrating its potential for advancing spatial audio reproduction.

The rest of the paper is organised as follows. Section II provides the signal model for lower-order ambisonics and formulates the system model for upscaling using the RM-FOCUSS method for the plane wave decomposition. Section-III evaluated the performance and compares it with the state-

of-art methods. Section-IV concludes the paper.

II. Data Model and Problem Formulation

The ambisonics encoding represents the sound scene using the ambisonics coefficients and the source signals as [21]

$$\mathbf{B} = \mathbf{Y}\mathbf{s} \tag{1}$$

where, $\mathbf{B} = [\mathbf{b}_{00}, \mathbf{b}_{1-1}, \ldots, \mathbf{b}_{NN}]^T$ is the encoded ambisonics, $\mathbf{s} = [s_1, \ldots s_Q]^T$ is the source signals in Q distinct direction $(\theta_{s_1}, \phi_{s_1}), \ldots, (\theta_{s_Q}, \phi_{s_Q})$. (θ, ϕ) denotes the elevation and azimuth respectively. $\mathbf{Y} \in \mathbb{C}^{(N+1)^2 \times Q}$ is the SH basis coefficient matrix, which contains the SH basis coefficients corresponding to the source direction and given as

$$\begin{aligned} \mathbf{Y} &= [\mathbf{y}(\theta_1, \phi_1), \mathbf{y}(\theta_2, \phi_2), \ldots, \mathbf{y}(\theta_Q, \phi_Q)] \\ \mathbf{y}(\theta, \phi) &= [Y_{00}(\theta, \phi), Y_{1-1}(\theta, \phi), \ldots, Y_{NN}(\theta, \phi)]^T \end{aligned} \tag{2}$$

where $Y_{nm}(\theta, \phi)$ is the spherical harmonics basis function of order n, $n = 0, \ldots, N$ and degree m, $m = -n, \ldots, n$, defined as [22]

$$Y_{nm}(\theta, \phi) = \sqrt{\frac{(2n+1)(n-|m|)!}{4\pi(n+|m|)!}} P_{nm}(\cos\theta) e^{im\phi} \tag{3}$$

The function $P_{nm}(\cdot)$ denotes the nth order mth degree associated Legendre function. Due to design and space constraints, the order is limited to N_l (lower order). The lower-order ambisonics signals are then processed through a time-domain overlapping sub-band filter of length w. Plane-wave decomposition is performed on each sub-banded encoded ambisonics signal to estimate the source signal and its direction for HOA order upscaling. The decomposition of the signal encoded in the τth sub-band is represented as

$$\mathbf{B}_\tau = \mathbf{Y}^D \hat{\mathbf{s}} \tag{4}$$

The matrix $\mathbf{Y}^D \in \mathcal{C}^{(N_l+1)^2 \times D}$ represents an overcomplete dictionary, constructed similarly to (2), where D is chosen such that $D \gg (N_l + 1)^2$. The rows of $\hat{\mathbf{s}}$ correspond to the decomposed plane waves, and the directions are inferred from the columns of $\mathbf{Y}^D$, as depicted in Figure 1. After this, the upscaling of the ambisonics order is performed. Since $D \gg (N_l + 1)^2$, and only a few plane waves are required to represent the encoded ambisonics coefficients, $\hat{\mathbf{s}}$ is sparse. The following assumptions apply to $\hat{\mathbf{s}}$

- Every column of $\hat{\mathbf{s}}$ must share the same sparsity profile, meaning the location of nonzero entries is consistent across columns.
- Fewer plane waves are needed to describe the dominant source, ensuring that $\hat{\mathbf{s}}$ remains sparse [23].

When a large number of sample points is used, $\mathbf{Y}^D$ becomes a full row-rank matrix, i.e., $\text{rank}(\mathbf{Y}^D) = (N+1)^2$. Moreover, (4) represents an underdetermined yet consistent system, meaning there are always multiple solutions. The challenge is to find the best possible solution. This becomes more complex when the nonzero row indices vary across subbands. In such cases, a sub-optimal solution is chosen to balance complexity and optimality.

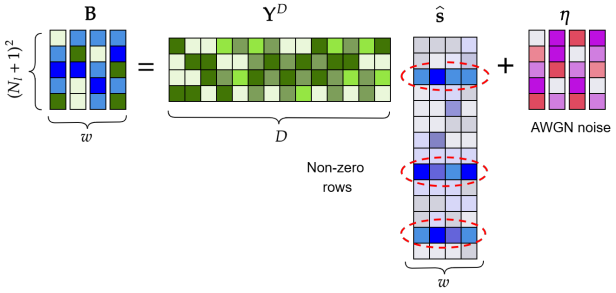

Fig. 1: Illustration of multiple measurement vector-based sparse representation for plane-wave decomposition.

A. Regularized M-FOCUSS algorithm for sparse plane wave decomposition

In contrast to forward sequential methods such as those proposed in [12], [13], the regularized M-FOCUSS algorithm is a diversity minimization technique used for sparse plane wave decomposition developed on our previous work [19]. The algorithm begins by selecting all columns of the dictionary matrix and iteratively removes columns until only a few remain. For ease of understanding, the algorithm assumes a noiseless scenario and an exact diversity solution, given as

$$\bar{\mathbf{s}} = \mathbf{Y}^{D^\dagger}\mathbf{B}, \quad \mathbf{Y}^{D^\dagger} = \mathbf{Y}^{D^H}(\mathbf{Y}^D\mathbf{Y}^{D^H})^{-1} \tag{5}$$

where $\mathbf{Y}^{D^\dagger}$ is the Moore–Penrose pseudo-inverse, representing the L_2-norm solution. For simplicity, the term τ is omitted. The diversity measure is defined as:

$$E^p(\hat{\mathbf{s}}) = \sum_{i=1}^{D} |\hat{\mathbf{s}}(i)|^p, \quad 0 \le p \le 1 \tag{6}$$

where $\hat{\mathbf{s}}(i) = [\hat{s}^1(i), \ldots, \hat{s}^w(i)]$ is the i-th row of $\hat{\mathbf{s}}$, and this resembles an L_p-norm, referred to as the p-norm diversity measure. For multiple measurements, the diversity measure is generalized as

$$\mathcal{J}^{p,q}(\hat{\mathbf{s}}) = \sum_{i=1}^{D} (\|\hat{\mathbf{s}}(i)\|_q)^p, \quad 0 \le p \le 1, q \ge 1 \tag{7}$$

where $\|\hat{\mathbf{s}}(i)\|_q = \sum_{l=1}^{w} |\hat{s}^l(i)|^q$. For simplicity, we set $q = 2$, resulting in

$$\mathcal{J}^p(\hat{\mathbf{s}}) = \sum_{i=1}^{D} \left(\sum_{l=1}^{w} |\hat{s}^l(i)|^2 \right)^{p/2}, \quad 0 \le p \le 1 \tag{8}$$

The cost function is chosen with two objectives: (i)As $p \to 0$, it identifies the number of non-zero rows of $\hat{\mathbf{s}}$, providing a sparse solution. (ii)It reduces computational complexity. The objective function is thus

$$\arg\min_{\hat{\mathbf{s}}} \mathcal{J}^p(\hat{\mathbf{s}}), \quad \text{s.t. } \mathbf{Y}^D\hat{\mathbf{s}} = \mathbf{B} \tag{9}$$

To account for noise and overfitting, regularized M-FOCUSS incorporates an additional regularization term, ensuring a stable and generalizable solution. The objective function for the regularized version is defined as:

$$\arg\min_{\hat{\mathbf{s}}} \left[\mathcal{J}^p(\hat{\mathbf{s}}) + \lambda \|\hat{\mathbf{s}}\|_2^2 \right], \quad \text{s.t. } \mathbf{Y}^D\hat{\mathbf{s}} = \mathbf{B} \tag{10}$$

where λ is the regularization parameter controlling the balance between sparsity and fitting. Representing this in Lagrange

form:

$$\mathcal{L}(\hat{\mathbf{s}}, \Lambda) = \mathcal{J}^p(\hat{\mathbf{s}}) + \lambda \|\hat{\mathbf{s}}\|_2^2 + \Lambda(\mathbf{Y}^D \hat{\mathbf{s}} - \mathbf{B}) \tag{11}$$

where Λ is the vector of Lagrange multipliers. At the stationary point

$$\nabla_{\hat{\mathbf{s}}} \mathcal{L}(\hat{\mathbf{s}}^*, \Lambda^*) = \nabla_{\hat{\mathbf{s}}} \mathcal{J}^p(\hat{\mathbf{s}}) + 2\lambda \hat{\mathbf{s}}^* + \mathbf{Y}^{D^H} \Lambda^* = 0 \tag{12}$$

$$\nabla_{\Lambda^*} \mathcal{L}(\hat{\mathbf{s}}^*, \Lambda^*) = \mathbf{Y}^D \hat{\mathbf{s}}^* - \mathbf{B} = 0 \tag{13}$$

The partial derivative of $\mathcal{L}(\hat{\mathbf{s}}^*, \Lambda^*)$ with respect to $\hat{\mathbf{s}}$ is given as:

$$\nabla_{\hat{\mathbf{s}}} \mathcal{J}^p(\hat{\mathbf{s}}) = |p| \Pi(\hat{\mathbf{s}}) \hat{\mathbf{s}} \tag{14}$$

where $\Pi(\hat{\mathbf{s}}) = \text{diag}(\|\hat{\mathbf{s}}\|^{p-2})$. Substituting into the stationary condition, the optimal solution is:

$$\hat{\mathbf{s}}^* = \Pi^{-1}(\hat{\mathbf{s}}^*) \mathbf{Y}^{D^H} (\mathbf{Y}^D \Pi^{-1}(\hat{\mathbf{s}}^*) \mathbf{Y}^{D^H} + 2\lambda \mathbf{I})^{-1} \tag{15}$$

with $\Pi^{-1}(\hat{\mathbf{s}}) = \text{diag}(\|\hat{\mathbf{s}}\|^{2-p})$. The solution is obtained iteratively, expressed as

$$\hat{\mathbf{s}}^{\nu+1} = \Pi^{-1}(\hat{\mathbf{s}}) \mathbf{Y}^{D^H} (\mathbf{Y}^D \Pi^{-1}(\hat{\mathbf{s}}) \mathbf{Y}^{D^H} + 2\lambda \mathbf{I})^{-1} B \tag{16}$$

For practical computation, the regularized M-FOCUSS algorithm is reformulated as

$$W^{\nu+1} = \text{diag}\left(\|\hat{\mathbf{s}}(i)\|^{1-p/2} \right), \quad p \in [0, 2] \tag{17a}$$

$$\mathbf{Y}^{D^H}{}^{\nu+1} = \mathbf{Y}^D W^{\nu+1} \tag{17b}$$

$$X^{\nu+1} = \mathbf{Y}^{D^{H\nu+1}} (\mathbf{Y}^{D^{\nu+1}} \mathbf{Y}^{D^{H\nu+1}} + \lambda I)^{-1} \mathbf{B} \tag{17c}$$

$$\hat{\mathbf{s}}^+ = W^+ X^+ \tag{17d}$$

The algorithm converges when:

$$\frac{\|\hat{\mathbf{s}}^+ - \hat{\mathbf{s}}\|_F}{\|\hat{\mathbf{s}}\|_F} < \delta \tag{18}$$

where $\|\cdot\|_F$ is the Frobenius norm. The regularization parameter λ plays a crucial role in controlling the trade-off between sparsity and the effect of noise, ensuring a robust solution even in practical, noisy environments.

B. Upscaling HOA

The upscaled spherical harmonics basis dictionary matrix $\mathbf{Y}^{D^u}$ is computed, and the upscaled higher-order ambisonics (HOA) is obtained as

$$\mathbf{B}^u = \mathbf{Y}^{D^u} \hat{\mathbf{s}} \tag{19}$$

where $\mathbf{B}^u$ is the encoded ambisonics of order $N_u > N_l$.

III. Performance Evaluation

The performance of the proposed upscaling method is assessed using metrics including upscaled error analysis, reproduced sound field evaluation, and subjective testing. Additionally, the method's performance is compared to state-of-the-art techniques such as compressed sensing (CS) [7], matching pursuit (MP) [12], order recursive matching pursuit (ORMP) [13], and M-FOCUSS [19].

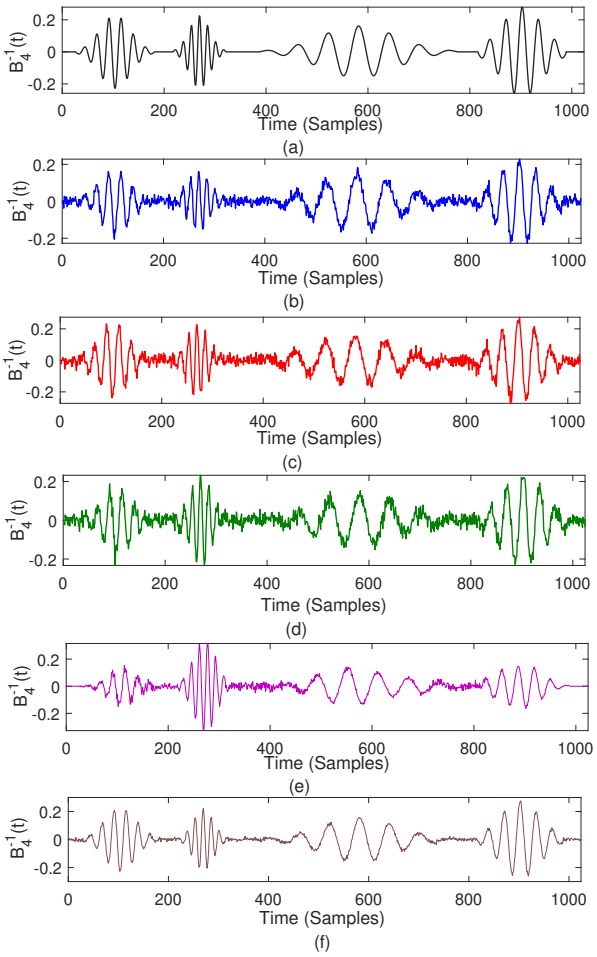

Fig. 2: Illustration of the B_{-1}^4 channel of upscaled ambisonics signal, (a) reference signal, and upscaled using (b) CS, (c) MP, (d) ORMP, (e) FOCUSS, (f) RM-FOCUSS methods in a diffused noise scenario (10 dB).

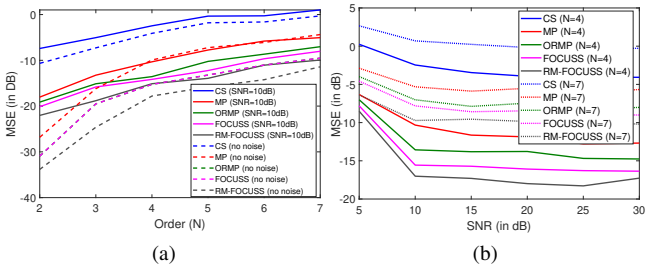

Fig. 3: Illustration of the MSE between the reference and upscaled ambisonics signals with (a) order variation from 2 to 7 and (b) SNR variation from 5 dB to 30 dB.

TABLE I: Run time complexity of the upscaling methods

Methods	Run time complexity (in Sec)			
	SNR=5dB	SNR=10dB	SNR=15dB	SNR=20dB
CS	65.5188	73.6874	75.0317	71.8285
MP	0.0069	0.0099	0.0097	0.0087
OMP	0.0056	0.0062	0.0064	0.0046
M-FOCUSS	0.0032	0.0058	0.0031	0.0026
RM-FOCUSS	0.0010	0.0014	0.0009	0.0009

A. Upscaling Error Analysis

For the upscaling error analysis, a sound scene was generated consisting of four Gaussian-modulated sinusoidal sources with core frequencies of 0.8 kHz, 1.5 kHz, 2 kHz, and 3 kHz, respectively, as described in previous works. The signals were encoded in B-format using first-order ambisonics, then processed using a 1024-sample rectangular window with 50% overlap. A spherical harmonic basis matrix for sparse plane wave decomposition was created using 512 directions from uniform sphere sampling. The reference upscaled signals, along with those upscaled using CS, MP, ORMP, FOCUSS, and RM-FOCUSS, were compared. The B^4_{-1} channel signal for one window is illustrated in Fig. 2, where the top image shows the reference signal, and subsequent figures display the upscaled signals under anechoic conditions for the different methods.

To further assess the robustness of the FOCUSS-based upscaling in diffuse conditions, uncorrelated white noise was added to the B-format signal at a noise level of 10 dB. The mean squared error (MSE) of the upscaled signals was computed by comparing them to the reference, varying both the order from 2 to 7 and the signal-to-noise ratio (SNR) from 5 dB to 30 dB. The results are obtained for all the methods and illustrated in Fig. 3. The results showed that the FOCUSS approach achieved a lower MSE compared to CS, MP, and ORMP, with the MSE decreasing as the SNR increased, demonstrating the effectiveness of the FOCUSS-based upscaling in noisy environments.

B. Complexity Analysis

The complexity analysis of HOA upscaling methods reveals significant differences in run-time performance. The experiments were conducted on a PC equipped with an Intel(R) Core(TM) i9-10900 CPU 2.80GHz and 32 GB RAM. As shown in Table I, the runtime complexity varies across different methods and signal-to-noise ratios (SNRs). CS exhibits the highest runtime complexity, ranging from 65.5 seconds at 5 dB SNR to 73.7 seconds at 10 dB SNR. In contrast, MP and ORMP methods demonstrate significantly lower runtimes, with MP ranging from 0.007 seconds to 0.010 seconds and OMP from 0.006 seconds to 0.007 seconds. The proposed M-FOCUSS method also shows lower complexity, with runtimes between 0.003 seconds and 0.006 seconds. Notably, the RM-FOCUSS method achieves the lowest runtime complexity, ranging from 0.001 seconds to 0.0014 seconds across all SNR levels. This substantial reduction in runtime demonstrates the efficiency of RM-FOCUSS, making it a highly effective choice for HOA upscaling in computationally demanding scenarios.

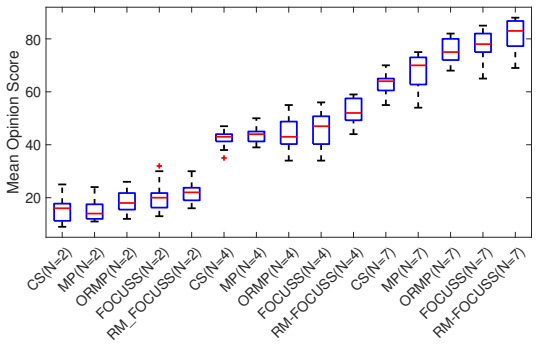

Fig. 4: Box plot showing the perception score for upscaled sound scene using CS and MP

C. Subjective Evaluation

For the subjective evaluation, a listening test was conducted using a B-format (order 1) ambisonics encoded signal that was upscaled to higher orders using CS, MP, ORMP, FOCUSS, and the proposed RM-FOCUSS techniques. The upscaled signals were decoded and played through an adequately spaced loudspeaker array for orders 2, 4, and 7, utilizing 10, 25, and 64 equally-spaced loudspeakers respectively. Fifteen subjects assessed the spatial audio quality based on their listening experience, rating the sound quality on a scale from 0 to 100, where 0 represents the worst quality and 100 represents the best.

The evaluation results, depicted in Fig. 4, reveal that the proposed RM-FOCUSS method generally provides a superior listening experience compared to CS, MP, ORMP and FOCUSS. Specifically, RM-FOCUSS achieved opinion scores of approximately 20 for order 2, 50 for order 4, and 83 for order 7. The box plots in Fig. 4 illustrate that the spatial audio quality improves with higher ambisonics orders. Overall, the proposed RM-FOCUSS method not only outperforms the other techniques but also enhances the listening experience as the order of ambisonics increases.

IV. Conclusion

In conclusion, the RM-FOCUSS is an effective approach to upscaling lower-order ambisonics, enhancing spatial resolution and delivering an improved immersive audio experience for listeners. By employing the FOCUSS algorithm and sub-band filtering in the time domain, RM-FOCUSS achieves efficient computational performance while maintaining high audio quality. Comparative analysis shows that RM-FOCUSS outperforms traditional methods like CS, MP, and ORMP, particularly in both clear and diffused noisy environments. It demonstrates superior spatial resolution with lower error and better subjective listening scores. Future research should focus on developing even more efficient, low-complexity upscaling methods with real-time implementation capabilities to further enhance the practicality of spatial audio systems.

ACKNOWLEDGEMENT

This work was supported by the ANRF under Project No. ANRF/ARG/2025/007629/ENS.

REFERENCES

[1] Michael A Gerzon, "Periphony: With-height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973.

[2] Jérôme Daniel, Sebastien Moreau, and Rozenn Nicol, "Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging," in *Audio Engineering Society Convention 114*. Audio Engineering Society, 2003.

[3] Darren B Ward and Thushara D Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on speech and audio processing*, vol. 9, no. 6, pp. 697–707, 2001.

[4] Franz Zotter and Matthias Frank, "All-round ambisonic panning and decoding," *Journal of the audio engineering society*, vol. 60, no. 10, pp. 807–820, 2012.

[5] Franz Zotter, Hannes Pomberger, and Markus Noisternig, "Energy-preserving ambisonic decoding," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 37–47, 2012.

[6] W. B. Kleijn, "Directional emphasis in ambisonics," *IEEE Signal Processing Letters*, vol. 25, no. 7, pp. 1079–1083, July 2018.

[7] Andrew Wabnitz, Nicolas Epain, Alistair McEwan, and Craig Jin, "Upscaling ambisonic sound scenes using compressed sensing techniques," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*. IEEE, 2011, pp. 1–4.

[8] Maximilian Kentgens, Shahd Al Hares, and Peter Jax, "On the upscaling of higher-order ambisonics signals for sound field translation," in *2021 29th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 81–85.

[9] A. Wabnitz, N. Epain, and C. T. Jin, "A frequency-domain algorithm to upscale ambisonic sound scenes," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 385–388.

[10] Nicolas Epain and Craig T Jin, "Super-resolution sound field imaging with sub-space pre-processing," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),*. IEEE, 2013, pp. 350–354.

[11] Shoichi Koyama, Naoki Murata, and Hiroshi Saruwatari, "Sparse sound field decomposition for super-resolution in recording and reproduction," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3780–3795, 2018.

[12] G. Routray and R. M. Hegde, "Sparse plane-wave decomposition for upscaling ambisonic signals," in *2020 International Conference on Signal Processing and Communications (SPCOM)*, 2020, pp. 1–5.

[13] Gyanajyoti Routray, Sumit Kumar Sahu, and Rajesh M Hegde, "Upscaling hoa signals using order recursive matching pursuit in spherical harmonics domain," in *2022 IEEE International Conference on Signal Processing and Communications (SPCOM)*. IEEE, 2022, pp. 1–5.

[14] N. Murata, S. Koyama, N. Takamune, and H. Saruwatari, "Sparse sound field decomposition with parametric dictionary learning for super-resolution recording and reproduction," in *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2015, pp. 69–72.

[15] N. Ueno, S. Koyama, and H. Saruwatari, "Three-dimensional sound field reproduction based on weighted mode-matching method," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 1852–1867, Dec 2019.

[16] M. Samarawickrama, N. Epain, and C. Jin, "Super-resolution acoustic imaging using non-uniform spatial dictionaries," in *2014 International Conference on Audio, Language and Image Processing*, July 2014, pp. 973–977.

[17] Gyanajyoti Routray, Priyadarshini Dwivedi, and Rajesh M. Hegde, "Binaural reproduction of hoa signal using sparse multiple measurement vector projections," in *2021 National Conference on Communications (NCC)*, 2021, pp. 1–6.

[18] Yiwen Wang, Xihong Wu, and Tianshu Qu, "Up-wgan: Upscaling ambisonic sound scenes using wasserstein generative adversarial networks," in *Audio Engineering Society Convention 152*, May 2022.

[19] Gyanajyoti Routray, Priyadarshini Dwivedi, and Rajesh M Hegde, "Diversity minimization technique for multiple measurement vector-based super-resolution spatial audio imaging," in *Forum Acusticum, 10th Convention of the European Acoustics Association*, 2023, pp. 615–621.

[20] Yue Qiao, Vinay Kothapally, Meng Yu, and Dong Yu, "Neural ambisonic encoding for multi-speaker scenarios using a circular microphone array," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

[21] Andrew Wabnitz, Nicolas Epain, André van Schaik, and Craig Jin, "Time domain reconstruction of spatial sound fields using compressed sensing," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 465–468.

[22] Boaz Rafaely, *Fundamentals of spherical array processing*, vol. 8, Springer, 2015.

[23] Svein Berge and Natasha Barrett, "High angular resolution planewave expansion," in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics May*, 2010, pp. 6–7.