DOI: 10.61782/fa.2023.1181

forum acusticum 2023

DIVERSITY MINIMIZATION TECHNIQUE FOR MULTIPLE MEASUREMENT VECTOR-BASED SUPER-RESOLUTION SPATIAL AUDIO IMAGING

Gyanajyoti Routray [1] **Priyadarshini Dwivedi** [1*] **Rajesh M. Hegde** [2]

[1] Department of Electrical Engineering, Indian Institute of Technology Kanpur, India.
[2] Department of Electrical Engineering, Indian Institute of Technology Dharwad, India.

ABSTRACT

Ambisonics is an efficient spatial sound acquisition and reproduction technique in the spherical harmonic domain. At low frequencies, lower-order ambisonics reproduction is accurate, but at high frequencies, the spatial resolution suffers. An increase in frequency shrinks the radius of the error-free region and degrades the spatial resolution. Higher-order ambisonics (HOA) provided better spatial resolution in this context. However, sound spatial acquisition in HOA is constrained by hardware complexity and storage space, in contrast to low-order ambisonics (B-format). So, it is worthwhile to acquire the sound scene at low order to reduce hardware complexity and storage requirement and upscale to a higher order while reproducing to improve the spatial resolution. This work investigated algorithms based on minimizing the diversity measures for obtaining higher-order ambisonics from the B-format signals. In particular, we are interested in the FOCUSS (FOCal Underdetermined System Solver) class of algorithms, which is an alternative and complementary approach to the sequential forward method to solve the sparse inverse problem. The performance of the proposed upscaling method is evaluated using the mean square error metrics. The subjective evaluation is performed using a listening test and compared with state-of-art methods.

*__Corresponding author__: *priyadw@iitk.ac.in.*

Keywords: *Higher-order ambisonics, Super resolution, Upscaling, Diversity minimization*

1. INTRODUCTION

Sound can be recorded and played back in a 3D space using ambisonics. HOA provides more detailed spatial information and uses the spherical harmonics (SH) basis to represent the spatial sound field [1, 2]. The hardware complexity and storage space requirement hinder the performance of this method, though. The sequence of the HOA signal and the source frequency limits the spatial resolution of the reconstructed sound field. The sweet spot is the zone where low-frequency sounds are reproduced faithfully. However, the sweet spot radius and spatial resolution decrease as frequency increases. The relationship between the size of the "sweet spot" and the order of the HOA is shown by the equation $N = \lceil kr_0 \rceil$ [3]. Where $\lceil \cdot \rceil$ is rounded to the nearest integer, r_0 is the radius of the sweet spot, $k = \frac{2\pi f}{c}$ is the wave number that goes with the frequency f, and $c = 343$ m/sec is the speed of sound. Because of this, the order at specific frequencies directly affects the spatial sharpness. The following reasons make it imperative to upscale the order of ambisonics.

- Higher order Ambisonics can record more sound field details, improving spatial accuracy. Increasing Ambisonics order improves spatial resolution, making sound sources and their places in reproduced audio more accurate.
- Upscaling Ambisonics can improve listening immersion. Recording and recreating spatial detail can generate a more immersive sound field.

- Higher order Ambisonics improves sound source localization. Increased order makes sound source direction and places easier to identify. Virtual and augmented reality requires excellent spatial awareness for immersive and participatory experiences.
- Upscaling Ambisonics allows lower-order content to be adapted to higher-order reproduction systems without losing fidelity or spatial accuracy as technology advances.

Using an emphasis operator [4–6] or increasing the order of encoded ambisonic signals [7–17] can improve spatial resolution. In [4, 5], an objective estimator is built to localize sources while conserving energy. Methods based on sparsity are presented in [7–16]. In [7–12], the solution matrix is obtained by using a uniform overcomplete spherical harmonics basis matrix, while in [16], a non-uniform spherical harmonics basis dictionary is employed. These sparse approaches rely on the conditioning of the dictionary matrix for their effectiveness. In [17], the authors presented a learning-based technique for upscaling ambisonics. A neural network with a dual path transform function is proposed to upscale the sampling rate and layers in [18]. However, such learning-based techniques are computationally intensive.

This study constructs a framework for plane-wave decomposition by investigating the sparseness of the source signals. Further, we extend our earlier work [12,15] by exploring a diversity minimization class of algorithms called FOCUSS. Compared to frequency domain upscaling [7], this method's implementation in the time domain is also computationally efficient and applicable to broadband audio sources. In addition, the FOCUSS approach has benefits in computing efficiency, flexibility, adaptability, and the estimation quality gained by iterative optimization of the estimation process. These features are helpful in various signal-processing applications that require identifying and isolating individual sources. Since the solution matrix must be evaluated for each frequency bin, the computational cost of frequency domain methods is considerable. In addition, a uniform sub-band filter is used as a window on the lower-ordered encoded signals to speed up calculations. Error in upscaled HOA signals and examination of replicated sound field performance matrices are used to evaluate the suggested FOCUSS approach. A hearing test is used for the subjective evaluation, and the results are compared to those obtained using state-of-the-art techniques.

The rest of the paper is structured as follows. A sparse framework decomposes the lower order ambisonics signal in Section-2. Later, the FOCUSS method is adopted to solve the sparse plane wave decomposition problem. Section-3 compares the suggested approach to the CS and MP methods. Section-4 finishes the paper.

2. PROBLEM FORMULATION

The ambisonics encoded signals are represented using the ambisonics coefficients and given as [19]

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

where, $\mathbf{s} = [s_1, \dots s_Q]^T$ is the source vector comprises of Q distinct source in the direction $(\theta_{s_1}, \phi_{s_1}), \dots, (\theta_{s_Q}, \phi_{s_Q})$, (θ, ϕ) denotes the elevation and azimuth. $\mathbf{Y} \in \mathbb{C}^{(N+1)^2 \times Q}$ is the SH basis coefficient matrix corresponding to the source and given as

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

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

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

$P_{nm}(\cdot)$ is the nth order mth degree associated Legendre function. Based on the design and space constraints, the order is limited to N_l (lower order). Subsequently, the lower-order ambisonics signals are passed through the time-domain overlapping sub-band filter of length w. Plane-wave decomposition is applied to each sub-banded encoded ambisonics signal. The decomposition of the signal encoded in the τth sub-band is represented as

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

where $\mathbf{Y}^D \in \mathcal{C}^{(N_l+1)^2 \times D}$, represents the overcomplete dictionary defined similar to (2) and D is chosen such that $D \gg (N_l + 1)^2$. The rows of $\hat{\mathbf{s}}$ represent the decomposed plane waves, and the corresponding directions are determined from the columns of $\mathbf{Y}^D$. The upscaling of the order of ambisonics is carried out after that. Since

$D \gg (N_l+1)^2$ and few numbers of plane waves are sufficient to represent the encoded ambisonics coefficient, $\hat{\mathbf{s}}$ is sparse. The following assumptions are made about $\hat{\mathbf{s}}$

- Each $\hat{\mathbf{s}}$ column must have the same sparsity profile, meaning that the nonzero entry index is unrelated to the column in which it appears.
- Fewer plane waves are needed to characterize the primary source; hence $\hat{\mathbf{s}}$ is sparse. [20].

Considering a high number of sample points transforms $\mathbf{Y}^D$ into a matrix with full row rank, i.e., $(rank(\mathbf{Y}^D) = (N+1)^2)$. In addition, (4) stands for an under-determined and consistent system so that there are always many solutions. The goal now is to identify the best option available. The situation gets more complicated when the nonzero row indices shift between subbands. A sub-optimal solution is selected in such a scenario as a compromise between complexity and optimality.

2.1 FOCUSS algorithm for sparse plane wave decomposition

In contrast to the forward sequential approaches [12, 15], the FOCUSS algorithm is a diversity minimization algorithm used for sparse plane wave decomposition. In this technique, all the columns of the SH dictionary matrix are selected, and an iterative strategy is followed to remove the columns until a few remain. This method begins with the noiseless issue and assumes an exact diversity solution that fulfils (4). The solution is expressed as $\bar{\mathbf{s}} = \mathbf{Y}^{D^\dagger}\mathbf{B}$, $\mathbf{Y}^{D^\dagger} = \mathbf{Y}^{D^H}(\mathbf{Y}^D\mathbf{Y}^{D^H})^{-1}$ (Moore–Penrose pseudo-inverse), and represents the norm-2 solution to (4). The term τ is dropped for the ease of representation. To begin with, the diversity measure is defined as

$$E^p(\hat{\mathbf{s}}) = \sum_{i=1}^{D} |\hat{\mathbf{s}}(i)|^p, \quad 0 \le p \le 1 \tag{5}$$

where $\hat{\mathbf{s}}(i) = [\hat{s}^1(i), \ldots, \hat{s}^w(i)]$ is the ith row of $\hat{\mathbf{s}}$. It is similar to l_p norm and called $p-$norm diversity measure. The diversity measure for multiple measurements is

$$\mathcal{J}^{p,q}(\hat{\mathbf{s}}) = \sum_{i=1}^{D} (\|\hat{\mathbf{s}}(i)\|_q)^p, \quad 0 \le p \le 1, q \ge 1 \tag{6}$$

where $\|\hat{\mathbf{s}}(i)\|_q = \sum_{l=1}^{w}\left(\|\hat{\mathbf{s}}^l(i)\|_q\right)^{1/q}$. For simplicity $q = 2$, and the diversity measure is written as $\mathcal{J}^p$ and expressed as

$$\mathcal{J}^p(\hat{\mathbf{s}}) = \sum_{i=1}^{D}\left(\sum_{l=1}^{w} |\hat{\mathbf{s}}^l(i)|^2\right)^{p/2}, \quad 0 \le p \le 1 \tag{7}$$

The choice of cost function is inspired in two direction

- If $p \to 0$, it finds the number of nonzero rows of $\hat{\mathbf{s}}$ and obtains a sparse solution to (4).
- Minimizing the computational complexity.

The objective function is defined as

$$\arg\min_{\hat{\mathbf{s}}} \ \mathcal{J}^p(\hat{\mathbf{s}}), \quad \text{s.t. } \mathbf{Y}^D\hat{\mathbf{s}} = \mathbf{B} \tag{8}$$

Representing the objective function in the Lagrange form

$$\mathcal{L}(\hat{\mathbf{s}}, \Lambda) = \mathcal{J}^p(\hat{\mathbf{s}}) + \Lambda(\mathbf{Y}^D\hat{\mathbf{s}} - \mathbf{B}) \tag{9}$$

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

$$\begin{aligned}\nabla_{\hat{\mathbf{s}}}\mathcal{L}(\hat{\mathbf{s}}^*, \Lambda^*) &= \nabla_{\hat{\mathbf{s}}}\mathcal{J}^p(\hat{\mathbf{s}}) + \mathbf{Y}^{D^H}\Lambda^* = 0\\ \nabla_{\Lambda^*}\mathcal{L}(\hat{\mathbf{s}}^*, \Lambda^*) &= \mathbf{Y}^D\hat{\mathbf{s}}^* - \mathbf{B} = 0\end{aligned} \tag{10}$$

The partial derivative of $\mathcal{L}(\hat{\mathbf{s}}^*, \Lambda^*)$, w.r.t $\hat{\mathbf{s}}$ in the tractable form is expressed as [21]

$$\nabla_{\hat{\mathbf{s}}}\mathcal{J}^p(\hat{\mathbf{s}}) = |p|\Pi(\hat{\mathbf{s}})\hat{\mathbf{s}} \tag{11}$$

where, $\Pi(\hat{\mathbf{s}}) = \text{diag}(\|(\hat{\mathbf{s}})\|^{p-2})$ and is independent of the columns. Now substituting in the stationary condition the optimal solution is obtained as

$$\hat{\mathbf{s}}^* = \Pi^{-1}(\hat{\mathbf{s}}^*)\mathbf{Y}^{D^H}(\mathbf{Y}^D\Pi^{-1}(\hat{\mathbf{s}}^*)\mathbf{Y}^{D^H}) \tag{12}$$

where, $\Pi^{-1}(\hat{\mathbf{s}}) = \text{diag}(\|(\hat{\mathbf{s}})\|^{2-p})$ and the solution $\hat{\mathbf{s}}^*$ is obtained iterative and expressed as

$$\hat{\mathbf{s}}^+ = \Pi^{-1}(\hat{\mathbf{s}})\mathbf{Y}^{D^H}(\mathbf{Y}^D\Pi^{-1}(\hat{\mathbf{s}})\mathbf{Y}^{D^H}) \tag{13}$$

For computation purposes, the FOCUSS algorithm is expressed as

$$W^+ = \text{diag}\left(\|\hat{\mathbf{s}}(i)\|^{1-p/2}\right) \tag{14a}$$

$$\mathbf{Y}^{D^{H+}} = \mathbf{Y}^D W^+ \tag{14b}$$

$$X^+ = \mathbf{Y}^{D^{H+}}\mathbf{B} \tag{14c}$$

$$\hat{\mathbf{s}}^+ = W^+X^+ \tag{14d}$$

The algorithm when the convergence attains the stopping criteria and given as

$$\frac{\|\hat{\mathbf{s}}^{+} - \hat{\mathbf{s}}\|_F}{\|\hat{\mathbf{s}}\|_F} < \delta \tag{15}$$

where, $\|\cdot\|_F$ being the Frobenius norm.

2.2 Upscaling HOA

Upscaled spherical harmonics basis dictionary matrix $\mathbf{Y}^{D^u}$ is calculated and the upscaled HOA is derived as

$$\mathbf{B}^u = \mathbf{Y}^{D^u}\hat{\mathbf{s}} \tag{16}$$

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

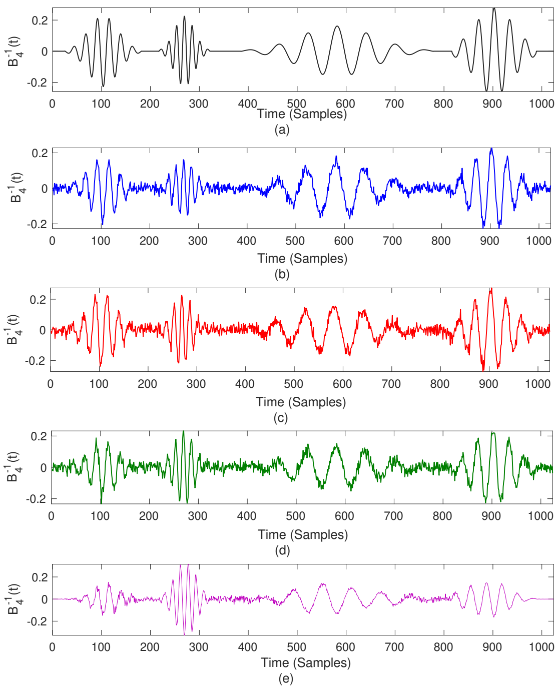

Figure 1: Illustration of the upscaled ambisonics signal for B_4^{-1} channel, (a) reference signal, Upscaling using (b) CS, (c) MP, (d) ORMP, and (e) FOCUSS method for diffused noise (10 dB) scenario.

3. PERFORMANCE EVALUATION

The proposed upscaling method's performance is evaluated utilizing metrics such as upscaled error analysis, reproduced sound field analysis, and subjective evaluations.

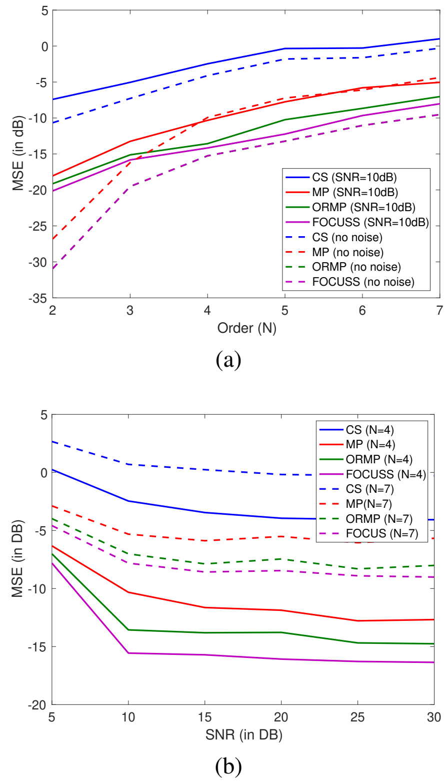

Figure 2: Illustration of MSE between the reference and upscaled ambisonics signal with (a) order variation form 2 to 7 (left), and (b) SNR variation from 5 dB to 30 dB (right).

Following that, the proposed approach's performance is compared to the state-of-the-art compressed sensing (CS) [7], matching pursuit (MP) method [15], and order recursive matching pursuit (ORMP) [12] methods.

3.1 Upscaling Error Analysis

A sound scene has been generated for upscaling error analysis, as shown in [15]. It is made up of four different Gaussian-modulated sinusoidal sources. The core frequencies of the transmitted source signal are 0.8 kHz, 1.5 kHz, 2 kHz, and 3kHz, respectively. The signals are encoded in B-format using order-1 ambisonics. The encoded signal is windowed using a 1024 sample-size rectangular window with 50% overlap. The dictionary spherical harmonic basis matrix is developed for sparse plane wave decomposition using 512 directions derived from uniform sphere sampling.

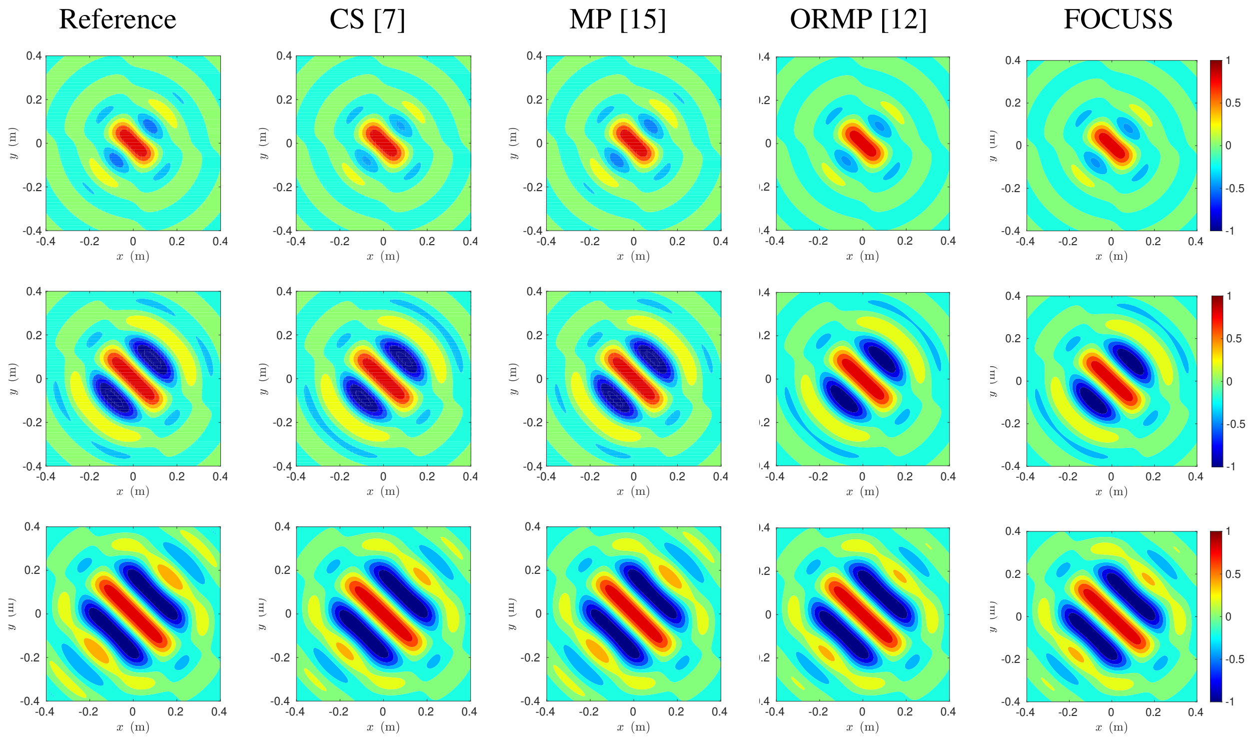

Figure 3: Sound pressure plots for order $n = 2, 4,$ and 7 (top-bottom)) for a frequency 2000 Hz. Column-1 represents the reference sound field, Column 2 -5 represent the sound field for the CS, MP ,ORMP, and proposed FOCUSS based upscaled sound fields respectively.

CS [7], MP technique [15], ORMP [12], and proposed FOCUSS-based algorithms are used to upscale B-format signals. For comparison, reference upscaled signals are obtained. Figure 1 depicts the B_4^{-1} channel signal across a single window. At the top is the upscaled reference signal for the B_4^{-1} channel. Figure 1(b)-(e) depicts the upscaled signal of the B_4^{-1} channel for anechoic conditions utilizing the CS, MP, ORMP, and FOCUSS techniques. Furthermore, the robustness of the suggested FOCUSS-based upscaling approach for diffuse conditions is investigated by introducing uncorrelated white noise to the B-format signal, as described in [7, 12, 15]. The diffused noise level is adjusted to 10dB in this scenario. The mean square error of the upscaled signals is calculated by comparing them to the reference and altering the order from 2 to 7 and the signal-to-noise level from 5 dB to 30 dB, as shown in Figure 2. The MSE for the suggested FOCUSS approach is the lowest compared to the CS, MP, and ORMP, as shown in Figure 2. In the diffused noise condition, the MSE falls as the signal-to-noise ratio grows. In this scenario, the MSE is the smallest possible for the proposed FOCUSS based upscaling.

3.2 Analysis of Reconstructed Sound Fields

To evaluate the performance, the upscaled sound field is recreated using the CS, MP, ORMP, and the proposed FOCUSS methods and analyzed. This analysis is carried out by considering a plane wave with a frequency of 2 kHz and an amplitude of 1 in the direction $(\pi/4, \pi/4)$. The sound pressure plots are obtained as

$$p(\mathbf{r}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} 4\pi i^n j_n(kr) Y_{nm}^*(\theta_s, \phi_s) s(k) Y_{nm}(\theta, \phi) \quad (17)$$

where $\mathbf{r} = (r, \theta, \phi)$ represents the observation point and $j(kr)$ represents the spherical Bessel function. The spherical harmonics coefficient for the source direction is repre-

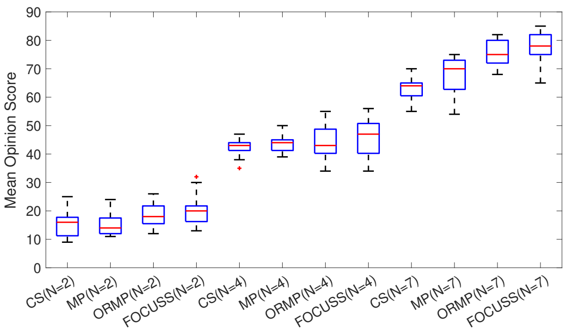

Figure 4: Box plot showing the perception score for upscaled sound scene using CS and MP

sented as $Y^*_{nm}(\theta_s, \phi_s)$. The CS, MP, and proposed ORMP methods calculate the upscaled spherical harmonics coefficient. These coefficients are used to calculate sound pressures, and the sound fields are presented in Fig. 3. The plane wave's sound field is reconstructed on a horizontal plane of 0.64 m^2 area. The reconstructed sound field for $N = 2$ (top row), $N = 4$ (middle row), and $N = 7$ (bottom row) is shown in Fig.3. The first column represents the reference sound field. The reconstructed sound field employing the CS, MP, ORMP, and FOCUSS methods is shown in the second to fifth columns. The spatial resolution improves with the order, as shown in Fig. 3.

3.3 Subjective Evaluation

For the subjective evaluation, an audition is performed. First, a B-format (order 1) ambisonics encoded signal is obtained and upscaled to higher orders using CS, MP, ORMP, and the proposed FOCUSS techniques, taking it to orders 2, 4, and 7. An adequately spaced loudspeaker array is used to decode these higher-order encoded signals. Orders 2, 4, and 7 of the decoded HOA signals are played through 10, 25, and 64 equally-spaced loudspeakers. Fifteen subjects evaluated the spatial audio quality from their listening experience. They are to assess the sound quality of the provided audio recordings on a scale from 0 to 100, with 0 representing the worst possible quality and 100 representing the best possible quality. Box plots of the results are shown in Fig. 4. Figure 4 shows that the listeners' aural acuity improves as the list is sorted. The proposed technique provides a better listening experience than the CS, MP, and ORMP methods. Further, it is observed that the listening experience improves with the increase in the order of ambisonics.

4. CONCLUSION

Sparse plane wave decomposition of lower-order ambisonics is developed using the diversity minimization algorithm FOCUSS. The lower-order ambisonics is upscaled to improve the spatial resolution, such as to give the listeners an improved immersive spatial audio experience. Time-domain upscaling algorithms use sub-band filters for efficient computing. CS, MP, ORMP, and FOCUSS-based methods are used to upscale lower-order encoded ambisonics signals. The replicated sound field, signal error, and listening test evaluate upscaling procedures. The proposed technique outperforms CS, MP, and ORMP in clear, diffused noisy situations and improves spatial resolution. Research is needed to produce low-complex, efficient spatial resolution, and real-time implementation approaches.

5. ACKNOWLEDGMENTS

This work is supported by SERB-MATRICS project under project no. SERB/EE/2021477.

6. REFERENCES

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

[2] J. Daniel, S. Moreau, and R. 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] D. B. Ward and T. 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] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *Journal of the audio engineering society*, vol. 60, no. 10, pp. 807–820, 2012.

[5] F. Zotter, H. Pomberger, and M. 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, pp. 1079–1083, July 2018.

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

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

[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)*, pp. 385–388, March 2012.

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

[11] S. Koyama, N. Murata, and H. 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, S. K. Sahu, and R. 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)*, pp. 1–5, IEEE, 2022.

[13] 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)*, pp. 69–72, Dec 2015.

[14] 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, pp. 1852–1867, Dec 2019.

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

[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*, pp. 973–977, July 2014.

[17] G. Routray, S. Basu, P. Baldev, and R. M. Hegde, "Deep-sound field analysis for upscaling ambisonic signals," in *EAA Spatial Audio Signal Processing Symposium*, (Paris, France), pp. 1–6, Sept. 2019.

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

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

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

[21] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2477–2488, July 2005.