INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Acoustic Signal Enhancement Using Relative Harmonic Coefficients: Spherical Harmonics Domain Approach

Yonggang Hu[*], *Prasanga N. Samarasinghe*, *Thushara D. Abhayapala*

Audio and Acoustic Signal Processing Group, Australian National University, Canberra, Australia
yonggang.hu@anu.edu.au, prasanga.samarasinghe@anu.edu.au, thushara.abhayapala@anu.edu.au

Abstract

Over recent years, spatial acoustic signal processing using higher order microphone arrays in the spherical harmonics domain has been a popular research topic. This paper uses a recently introduced source feature called the *relative harmonic coefficients* to develop an acoustic signal enhancement approach in noisy environments. This proposed method enables to extract the clean spherical harmonic coefficients from noisy higher order microphone recordings. Hence, this technique can be used as a pre-processing tool for noise-free measurements required by many spatial audio applications. We finally present a simulation study analyzing the performance of this approach in far field noisy environments.

Index Terms: spatial acoustic signal enhancement, relative harmonic coefficients, spherical harmonics domain, far-field

1. Introduction

In the past decade, higher-order microphone arrays (e.g., spherical and circular microphone arrays) that are capable of recording and analyzing the soundfield over a spatial area, have been widely used in spatial acoustic processing techniques [1–5]. Up until now, soundfield recording using different structured higher-order microphone arrays has been developed, including a spherical microphone array [6], multiple circular microphone arrays [7], and a planar microphone array [8]. Samarasinghe et al. used an array of higher-order microphones (circular arrays in [9], spherical arrays in [10]) to develop the measuring techniques that are more suitable for large spatial regions. These measuring techniques generally use a spherical Fourier transform [11] to decompose the multi-channel recordings into the spherical harmonics domain (i.e., wave domain in [12]), where the soundfield is then characterized/represented by the spherical harmonic coefficients.

A common problem when acquiring spatial soundfield recordings is environmental, thermal and other forms of noise that hinders an accurate acquisition of desired soundfield in noisy environments. The noise signal causes more erroneous estimations in the spherical harmonics domain because it can be easily amplified (refereed as the “Bessel zero problem” in [10]). Some special structured microphone arrays have been designed to alleviate the “Bessel zero problem”, such as the dual and rigid spherical microphone array [13–15], at the cost of more complicated microphone array requirements. On the other hand, several noise reduction techniques have been developed to alleviate the noise. Early approaches in [16–18] used an optimal beamformer to extract the clean received signal at the origin of the microphone array from the noisy microphone recordings. However, those techniques cannot estimate the spherical harmonic coefficients up to the full soundfield order. By contrast, a recent research in [19] presented a methodology that fully recovers the spherical harmonic coefficients due to the original sound source, while preserving the spatial acoustic cues of the original soundfield. However, this work requires an additional localization technique [20] to estimate the unknown source direction of arrival (DOA) [21–23].

This paper develops a novel spherical harmonics domain enhancement approach in noisy environments, using a recently introduced source feature called the *relative harmonic coefficients* [24, 25]. According to the definition of relative harmonic coefficients, the spherical harmonic coefficients can be represented as a multiplication of relative harmonic coefficients and the received signal at the origin of the array (call as received signal below). Hence, this paper achieves the spherical harmonics coefficients estimations using the following three steps. Firstly, we estimate the relative harmonic coefficients. Secondly, we use a beamformer to estimate the received signal. Finally, we recover the original spherical harmonics coefficients by multiplying the estimated relative harmonic coefficients and received signal. We emphasize that, different from above approaches in [18, 19], this developed approach is self-contained because the relative harmonic coefficients contain the information of source DOA, thus the beamformer in the second step does not require any additional localization techniques. Extensive simulations, using the spherical microphone array measurements from a far-field speaker, confirm the effective denoising performance of this method in noisy environments.

2. System Model

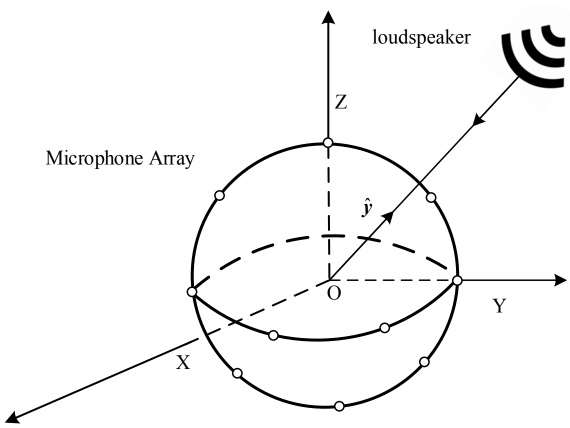

Figure 1: *Recording using a spherical microphone array.*

2.1. Acoustic model

Assume a single sound source propagating from an unknown DOA, e.g., (ϑ_s, φ_s) where $0 < \vartheta_s < \pi$, $0 < \varphi_s < 2\pi$, with respect to the origin of a higher order microphone array

[*]Yonggang Hu is sponsored by CSC agency for funding.

http://dx.doi.org/10.21437/Interspeech.2020-2316

(e.g., a spherical microphone array in Fig. 1). Let the microphone array has M microphones whose spherical coordinates are $\boldsymbol{x}_j = (r, \theta_j, \phi_j)$ $(j = 1, \cdots, M)$, respectively. The recording in the time domain, measured by the j-th channel, is represented as the sum of desired signal and the noise signal,

$$p(\boldsymbol{x}_j, n) = s(n) * a(\boldsymbol{x}_j, n) + v(\boldsymbol{x}_j, n), \;\; j = 1, \cdots, M \quad (1)$$

where $*$ is a convolution operation, $s(n)$ is the source signal, $a(\boldsymbol{x}_j, n)$ is the room impulse response between the sound source and the j-th microphone, and $v(\boldsymbol{x}_j, n)$ denotes the additive noise signal at the j-th microphone. The time domain expression is represented in the frequency domain using a Fourier transform as,

$$\begin{aligned} P(\boldsymbol{x}_j, k) &= Q(\boldsymbol{x}_j, k) + V(\boldsymbol{x}_j, k) \\ &= S(k)A(\boldsymbol{x}_j, k) + V(\boldsymbol{x}_j, k) \end{aligned} \quad (2)$$

where $k = 2\pi f/c$ is the wavenumber, f is the frequency bin, c is the speed of sound, $P(\boldsymbol{x}_j, k)$, $S(k)$, $A(\boldsymbol{x}_j, k)$, and $V(\boldsymbol{x}_j, k)$ denote the Fourier transforms of $p(\boldsymbol{x}_j, n)$, $s(n)$, $a(\boldsymbol{x}_j, n)$ and $v(\boldsymbol{x}_j, n)$, respectively.

2.2. Spherical harmonic representation of the soundfield

The sound pressure of (2) within the recording area can be further decomposed into the spherical harmonics domain [11] as,

$$P(\boldsymbol{x}_j, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k)\, j_n(kr)\, Y_{nm}(\theta_j, \phi_j) \quad (3)$$

where $N = \lceil kr \rceil$ is the truncated order of soundfield [26], $j_n(\cdot)$ is the spherical Bessel function of the first kind,

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

is the spherical harmonic function with order n and mode m, $P_{nm}(\cdot)$ is the associated Legendre function, and $\alpha_{nm}(k)$ is the spherical harmonic coefficient. It is of common practice to measure the spherical harmonic coefficients for a spherical microphone array by decomposing the sound pressure as [26],

$$\alpha_{nm}(k) = \frac{1}{j_n(kr)} \sum_{j=1}^{M} a_j P(\boldsymbol{x}_j, k)\, Y^*_{nm}(\theta_j, \phi_j) \quad (5)$$

in which $*$ is a conjugate transpose and a_j is the weight of the microphones to ensure the right side in (5) equals to the left side. For the noisy recordings by (2), the spherical harmonic coefficients also contain noise,

$$\alpha_{nm}(k) = y_{nm}(k) + \gamma_{nm}(k), \; -n \le m \le n, 0 \le n \le N \quad (6)$$

where $y_{nm}(k)$, $\gamma_{nm}(k)$ denote the spherical harmonic coefficients of $Y(\boldsymbol{x}_j, k)$, and $V(\boldsymbol{x}_j, k)$, respectively.

2.3. Problem addressed

This paper aims at acoustic signal enhancement in spherical harmonics domain by accurately estimating the vector of spherical harmonic coefficients, i.e., $\boldsymbol{y}(k) = [y_{00}(k), \cdots, y_{nm}(k)]$, from the mixture measurements of $\alpha_{nm}(k)$. Note that the noisy measurements in (6) can be rewritten as,

$$\alpha_{nm}(k) = \beta_{nm}(k) y_{00}(k) + \gamma_{nm}(k) \quad (7)$$

where

$$\beta_{nm}(k) = \frac{y_{nm}(k)}{y_{00}(k)} \quad (8)$$

denotes the relative harmonic coefficients defined in [24]. The acoustic model in (7) implies that the desired signal of $y_{nm}(k)$ can be represented as the multiplication of $\beta_{nm}(k)$ and $y_{00}(k)$. Moreover, the $y_{00}(k)$ is equivalent to the received signal at the array origin [27], i.e., $S(k)$. Hence, we intend to achieve our goal through the following three steps: (i) estimating the vector of relative harmonic coefficients, i.e., $\boldsymbol{\beta}(k) = [\beta_{00}(k), \cdots, \beta_{nm}(k)]^T$. (ii) recovering the $y_{00}(k)$ using a beamformer. (iii) multiplying the estimated $\boldsymbol{\beta}(k)$ and $y_{00}(k)$ to achieve accurate estimation of $\boldsymbol{y}(k)$.

3. Proposed Approach

This section presents the novel approach to estimate the clean spherical harmonic coefficients in the noisy environments, as explained in the following steps. Note that we mainly focus on the estimation at a single spherical harmonics mode, i.e., $y_{nm}(k)$, as estimations over the full soundfield order follow a similar process.

3.1. Relative harmonic coefficients estimation

This subsection reviews an estimator of the relative harmonics coefficients, i.e. $\beta_{nm}(k)$ in (7). Preliminary study in [24, 25, 28] confirms that $\beta_{nm}(k)$ only depends on the source DOA, thus is invariant over the time-varying source signal. In order to calculate the $\beta_{nm}(k)$ while alleviating the negative effects due to the noise, we exploit the power spectral density (PSD) and CPSD (cross PSD) of the measured signals,

$$\frac{S_{\alpha_{nm}\alpha_{00}}(k)}{S_{\alpha_{00}\alpha_{00}}(k) - S_{\gamma_{00}\gamma_{00}}(k)} = \frac{\beta_{nm}(k) S_{y_{00}y_{00}}(k)}{S_{y_{00}y_{00}}(k)} = \beta_{nm}(k) \quad (9)$$

where

$$\begin{aligned} S_{\alpha_{nm}\alpha_{00}}(k) &= \mathbb{E}\left\{\alpha_{nm}(k)\alpha^*_{00}(k)\right\} \\ S_{\alpha_{00}\alpha_{00}}(k) &= \mathbb{E}\left\{\alpha_{00}(k)\alpha^*_{00}(k)\right\} \\ S_{\gamma_{00}\gamma_{00}}(k) &= \mathbb{E}\left\{\gamma_{00}(k)\gamma^*_{00}(k)\right\} \\ S_{y_{00}y_{00}}(k) &= \mathbb{E}\left\{y_{00}(k)y^*_{00}(k)\right\} \end{aligned} \quad (10)$$

where $\mathbb{E}\{\cdot\}$ denotes the statistical expectation operator. Note that (9) exploits the un-correlation between the spherical harmonic coefficients of sound source signal and additive noise signal due to the un-correlation between their sound pressure. However, the noise PSD of $S_{\gamma_{00}\gamma_{00}}(k)$ is unknown in practice. Some state-of-art power spectral density techniques, such as [29, 30], are available to update the $S_{\gamma_{00}\gamma_{00}}(k)$. For simplicity, we use the biased estimator by neglecting it in (9),

$$\bar{\beta}_{nm}(k) \approx \frac{S_{\alpha_{nm}\alpha_{00}}(k)}{S_{\alpha_{00}\alpha_{00}}(k)}. \quad (11)$$

The reader is referred to [28] for more details about the estimator in the presence of noise.

3.2. Estimation of the received signal at the origin

This subsection estimates the received signal of $y_{00}(k)$ in (7) by steering a beamformer. Since the aperture of the recording area is much smaller when compared to its distance to the sound source, we use a simple and straightforward method called as the maximum directivity beamformer toward to the far-field

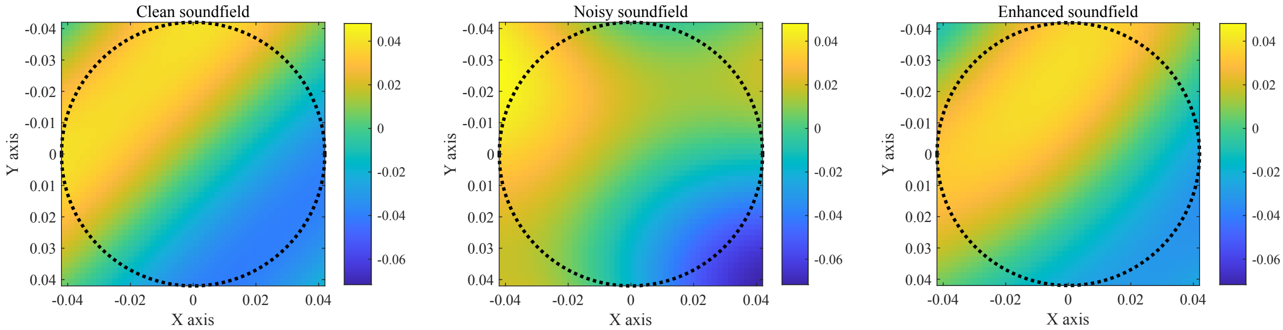

Figure 2: *Clean, noisy and enhanced soundfield over the microphone array when $z = 0$ (3 KHz, 5dB noise).*

sound source [20, 31, 32],

$$S(k) = \sum_{n=0}^{N}\sum_{m=-n}^{n}\frac{i^{-n}}{(N+1)^2}Y_{nm}(\vartheta_s,\varphi_s)\alpha_{nm}(k). \quad (12)$$

The spherical harmonic coefficient of $y_{00}(k)$ is equivalent to the received signal estimated in (12) [27],

$$\bar{y}_{00}(k) = S(k). \quad (13)$$

However, the beamformer approach in (12) requires the additional knowledge of source DOA, i.e., (ϑ_s,φ_s), which is hardly known in practice. We emphasize that estimation of the relative harmonic coefficients in (11) has an additional advantage as they enable us to recover the unknown source DOA. Hence, this proposed algorithm is self-contained as it does not require any additional localization techniques. Let us introduce the DOA estimation using the relative harmonic coefficients in the following subsection.

3.3. Direction-of-arrival estimation

This subsection introduces the DOA estimation approach, initially proposed in [25], using the estimated relative harmonic coefficients. The spherical harmonic coefficients due to the far-field sound source follow as [33, 34],

$$y_{nm}(k) = S(k)4\pi i^n Y^*_{nm}(\vartheta_s,\varphi_s) \quad (14)$$

where (ϑ_s,φ_s) is the source DOA. Following the definition in (8), we know its relative harmonic coefficients,

$$\beta_{nm}(k) = 2\sqrt{\pi}i^n Y^*_{nm}(\vartheta_s,\varphi_s). \quad (15)$$

Considering the cases up to the N-th order, we have a vector,

$$\boldsymbol{\beta}(\vartheta_s,\varphi_s) = \left[1, 2\sqrt{\pi}iY^*_{1,-1}(\vartheta_s,\varphi_s),\cdots,2\sqrt{\pi}i^N Y^*_{NN}(\vartheta_s,\varphi_s)\right]^T. \quad (16)$$

Assuming the continuous two dimension space $\boldsymbol{\Phi} = \{(\vartheta_s,\varphi_s) : 0 < \vartheta_s < \pi, 0 < \varphi_s < 2\pi\}$ is sampled by S discrete candidates, we have a feature set using (16),

$$\mathcal{H} = \{\boldsymbol{\beta}(\vartheta_1,\varphi_1),\boldsymbol{\beta}(\vartheta_2,\varphi_2),\cdots,\boldsymbol{\beta}(\vartheta_S,\varphi_S)\} \quad (17)$$

which is calculated without any prior recordings. Since the relative harmonic coefficients due to the unknown source are given by (11), we can recover its DOA by calculating its distance to the elements of $\mathcal{H}$,

$$argmin_{(\bar{\vartheta}_s,\bar{\varphi}_s)}\sum_{n=0}^{N}\sum_{m=-n}^{n} \mid \bar{\beta}_{nm} - 2\sqrt{\pi}i^n Y^*_{nm}(\vartheta_s,\varphi_s)\mid^2 \quad (18)$$

where

$$\bar{\beta}_{nm} = \frac{1}{K}\sum_{k=1}^{K}\bar{\beta}_{nm}(k) \quad (19)$$

where $\bar{\beta}_{nm}(k)$ denotes the practical estimations at a single frequency, and $\bar{\beta}_{nm}$ denotes the smoothed vector over K frequency bins as the (15) implies this feature is source-frequency independent. Please refer to [25] for more details about this DOA estimation algorithm.

3.4. Spherical harmonic coefficients estimation

Multiplying the estimations in (11) and (13), we finally recover the clean spherical harmonic coefficients as,

$$\bar{y}_{nm}(k) = \bar{\beta}_{nm}(k)\bar{y}_{00}(k). \quad (20)$$

Note that, at the k-th frequency bin, the $\bar{\beta}_{nm}(k)$ is fixed, while the $y_{nm}(k)$ is updated by the dynamic $\bar{y}_{00}(k)$ over the time-varying source signal. When the $[\bar{y}_{00}(k),\cdots,\bar{y}_{NN}(k)]$ over the entire STFT bins are estimated, we can then reconstruct spectrogram of any individual microphone on the array, using the spherical harmonics representation in (3).

4. Simulations

This section uses speech recordings along with a theoretical room model to evaluate the proposed acoustic enhancement approach in diverse noisy environments. When obtaining the following results, we first estimate the relative harmonic coefficients using (11). Then, we estimate the source DOA using (18). After that, we use (12) to extract the received signal. Finally, we use (20) to recover the spherical harmonic coefficients.

We simulate a room whose dimensions are $6\times4\times3$, and use an efficient implementation[1] of the image source method [35] to generate the room impulse response (RIR) between the sound source and a spherical microphone array (with 32 channels and radius 4.2 cm). The sound source uses a unique speech sentence lasting 3 seconds, drawn from the publicly available TIMIT database and is re-sampled at 8 KHz. We convolve the RIR and speech signal to obtain the received recordings, which are then contaminated by while Gaussian noise. Fifty frequency bins ranging from 1600 Hz to 2400 Hz, translating to a 2nd order spatial soundfield ($N = \lceil kr \rceil$), are exploited for the DOA estimation. The unique feature set of $\mathcal{H}$ in (16) is computed by sampling the elevation and azimuth angles with a resolution of 2 degrees.

Accuracy of the signal estimations is measured by the nor-

[1] https://www.audiolabs-erlangen.de/fau/professor/habets/software/signal-generator

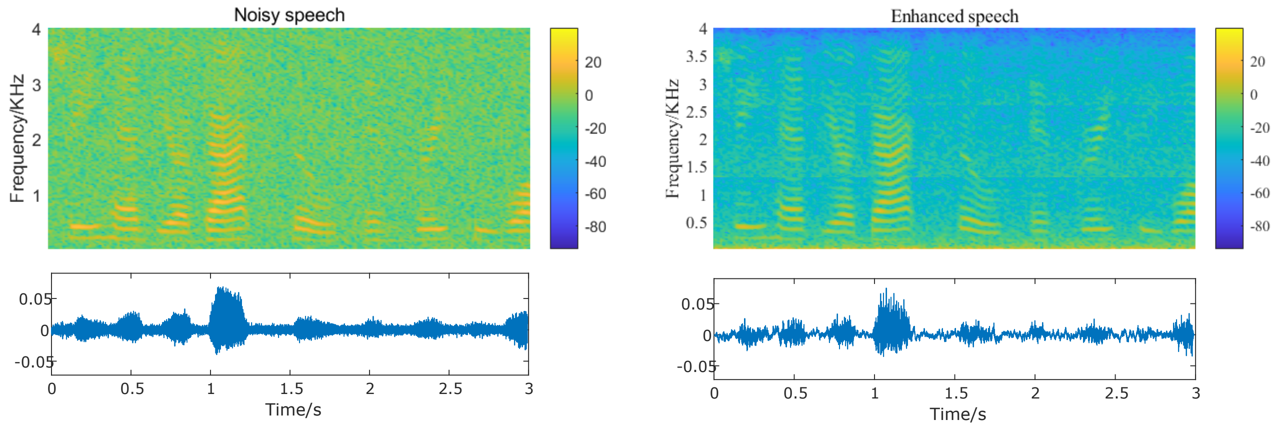

Figure 3: *Noisy and enhanced speech signal of the 1st microphone on the spherical microphone array (5dB noise).*

malized mean squared error (NMSE/dB) of the original and estimated signal in noisy environments. We first define the metric for the relative harmonic coefficients,

$$\mathrm{NMSE}_{\boldsymbol{\beta}} = 10\log_{10}\left(\frac{1}{K}\sum_{k=1}^{K}\frac{||\boldsymbol{\beta}(k)-\bar{\boldsymbol{\beta}}(k)||_2^2}{||\boldsymbol{\beta}(k)||_2^2}\right) \tag{21}$$

where $||\cdot||_2$ denotes the ℓ_2 norm, $\boldsymbol{\beta}(k)$ and $\bar{\boldsymbol{\beta}}(k)$ denote the clean and estimated relative harmonic coefficients. Next, the metric for the received signal in the STFT domain is,

$$\mathrm{NMSE}_{\boldsymbol{y}_{00}} = 10\log_{10}\left(\frac{1}{K}\sum_{k=1}^{K}\frac{||\boldsymbol{y}_{00}(k)-\bar{\boldsymbol{y}}_{00}(k)||_2^2}{||\boldsymbol{y}_{00}(k)||_2^2}\right) \tag{22}$$

where $\boldsymbol{y}_{00}(k)$ and $\bar{\boldsymbol{y}}_{00}(k)$ denote the vector of clean and estimated received signal, respectively. Finally, the metric for spherical harmonics coefficients is,

$$\mathrm{NMSE}_{\boldsymbol{y}} = 10\log_{10}\left(\frac{1}{K}\sum_{k=1}^{K}\frac{||\boldsymbol{y}(k)-\bar{\boldsymbol{y}}(k)||_2^2}{||\boldsymbol{y}(k)||_2^2}\right) \tag{23}$$

where $\boldsymbol{y}(k)$ and $\bar{\boldsymbol{y}}(k)$ denote the clean and estimated spherical harmonics coefficients, respectively. Accuracy of the DOA estimation is evaluated by the mean absolute estimated error (MAEE/degrees) between the estimated and original DOA,

$$\mathrm{MAEE} = \frac{1}{2}\left(|\vartheta_{\mathrm{ori}} - \vartheta_{\mathrm{est}}| + |\varphi_{\mathrm{ori}} - \varphi_{\mathrm{est}}|\right) \tag{24}$$

where $|\cdot|$ denotes the absolute operator.

Table 1: *Accuracy of estimations at various SNR levels, including relative harmonic coefficients, DOA and received signal.*

SNR level (/dB)	25	20	15	10	5
$\mathrm{NMSE}_{\boldsymbol{\beta}}$ (dB)	-20.3	-15.0	-9.5	-5.5	-2.65
MAEE (/degrees)	0.55	0.64	0.91	1.58	2.44
$\mathrm{NMSE}_{\boldsymbol{y}_{00}}$ (/dB)	-15.1	-11.6	-7.9	-4.7	-2.2

To achieve consistent results, we use ten groups of sound sources, each propagating from a randomly unknown DOA. Thus, the values of both NMSE and MAEE, presented in the followings, denote the mean value over all the cases. We evaluate the proposed method under various SNR conditions ranging from 5 dB to 25 dB. Table 1 depicts the accuracy of the estimations for relative harmonic coefficients, DOA and received signal, respectively. Seeing the DOA estimation, we recognize little degraded accuracy when the SNR level decreases. This is due to the accurate estimations of the relative harmonic coefficients in noisy environments. Accurate DOA estimations contribute to accurate estimations of the received signal as well. We then multiply the estimated relative harmonic coefficients and received signal to obtain the spherical harmonic coefficients. Table 2 depicts the NMSE where the original values are taken as the baseline. As expected, we observe that a stronger noisy environment exerts a direct negative influence on the received signal. However, the proposed method improves the NMSE by around 3 dB. Figure 2 exhibits the clean, noisy and enhanced soundfield over the microphone array whose coordinates are $z = 0$, due to the sound source propagating from $(\vartheta_s, \varphi_s) = (1.13, 3.94)$. We recognize that the enhanced soundfield generally gets rid off the distortions caused by the noise signal. Note that the Figure 2 shows the soundfield at a single STFT bin. By contrast, Figure 3 presents the speech spectrogram in the entire STFT domain and time domain recordings using an ISTFT. Most of the noise is alleviated and the enhanced speech has satisfying intelligibility. While the above results are promising, a limitation of this approach lies in the beamformer in (12), which is designed for far-field propagation. Therefore, in near-field soundfields, the estimation performance of $\bar{y}_{00}(k)$ in (13) will degrade unless an appropriate radial focused near-field beamformer is used.

Table 2: *Accuracy of the spherical harmonic coefficients estimations under various SNR levels.*

$\mathrm{NMSE}_{\boldsymbol{y}}$(/dB)	SNR levels (/dB)				
Method	25	20	15	10	5
Original	-11.9	-7.8	-4.8	-2.6	-1.2
Enhanced	-16.4	-11.8	-7.7	-4.4	-2.2

5. Conclusion

This paper uses the relative harmonic coefficients to present an acoustic enhancement approach for cleaning noisy higher-order microphone recordings. This method enables to estimate the spherical harmonics coefficients up to the whole soundfield order, while not requiring any additional localization techniques. Extensive results using noisy speech measurements of a spherical microphone array confirmed the effective denoising performance. However, current approach is only limited to a single static sound source under a far-field scenario. Hence, one potential future work is to extend the underlying theory into a denoising scheme for multi-source cases in a more dynamic environment including near-field propagation.

6. References

[1] J. Y. Hong, J. He, B. Lam, R. Gupta, and W.-S. Gan, "Spatial audio for soundscape design: Recording and reproduction," *Applied Sciences*, vol. 7, no. 6, p. 627, 2017.

[2] W. Zhang, P. N. Samarasinghe, H. Chen, and T. D. Abhayapala, "Surround by sound: A review of spatial audio recording and reproduction," *Applied Sciences*, vol. 7, no. 5, p. 532, 2017.

[3] Y. Takida, S. Koyama, N. Ueno, and H. Saruwatari, "Robust gridless sound field decomposition based on structured reciprocity gap functional in spherical harmonic domain," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 581–585.

[4] Q. Huang, L. Zhang, and Y. Fang, "Two-stage decoupled doa estimation based on real spherical harmonics for spherical arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 11, pp. 2045–2058, 2017.

[5] H. Sun, H. Teutsch, E. Mabande, and W. Kellermann, "Robust localization of multiple sources in reverberant environments using EB-ESPRIT with spherical microphone arrays," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 117–120.

[6] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vols I-Iv*, pp. 1949–1952, 2002.

[7] T. D. Abhayapala and A. Gupta, "Spherical harmonic analysis of wavefields using multiple circular sensor arrays," *IEEE Transactions on Audio Speech and Language Processing*, vol. 18, no. 6, pp. 1655–1666, 2010.

[8] H. Chen, T. D. Abhayapala, and W. Zhang, "Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis," *Journal of the Acoustical Society of America*, vol. 138, no. 5, pp. 3081–3092, 2015.

[9] P. N. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, "Spatial soundfield recording over a large area using distributed higher order microphones," *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 221–224, 2011.

[10] P. N. Samarasinghe, T. Abhayapala, and M. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," *IEEE-ACM Transactions on Audio Speech and Language Processing*, vol. 22, no. 3, pp. 647–658, 2014.

[11] E. G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*. Academic press, 1999.

[12] H. Teutsch, *Modal array signal processing: principles and applications of acoustic wavefield decomposition*. Springer, 2007, vol. 348.

[13] I. Balmages and B. Rafaely, "Open-sphere designs for spherical microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 727–732, 2007.

[14] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on speech and audio processing*, vol. 13, no. 1, pp. 135–143, 2005.

[15] ——, *Fundamentals of spherical array processing*. Springer, 2015, vol. 8.

[16] S. Yan, H. Sun, U. Svensson, X. Ma, and J. M. Hovem, "Optimal modal beamforming for spherical microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 361–371, 2011.

[17] D. P. Jarrett, E. A. Habets, J. Benesty, and P. A. Naylor, "A tradeoff beamformer for noise reduction in the spherical harmonic domain," in *IWAENC 2012; International Workshop on Acoustic Signal Enhancement*, pp. 1–4.

[18] D. P. Jarrett, M. Taseska, E. A. Habets, and P. A. Naylor, "Noise reduction in the spherical harmonic domain using a tradeoff beamformer and narrowband DOA estimates," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, no. 5, pp. 967–978, 2014.

[19] C. Borrelli, A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "A denoising methodology for higher order ambisonics recordings," in *2018 16th IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2018, pp. 451–455.

[20] D. P. Jarrett, E. A. Habets, and P. A. Naylor, *Theory and applications of spherical microphone array processing*. Springer, 2017, vol. 9.

[21] T. G. Dvorkind and S. Gannot, "Time difference of arrival estimation of speech source in a noisy and reverberant environment," *Signal Processing*, vol. 85, no. 1, pp. 177–204, 2005.

[22] S. Gannot and T. G. Dvorkind, "Microphone array speaker localizers using spatial-temporal information," *EURASIP journal on applied signal processing*, vol. 2006, pp. 174–174, 2006.

[23] S. Gannot, M. Haardt, W. Kellermann, and P. Willett, "Introduction to the issue on acoustic source localization and tracking in dynamic real-life scenes," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 3–7, 2019.

[24] Y. Hu, P. N. Samarasinghe, and T. D. Abhayapala, "Sound source localization using relative harmonic coefficients in modal domain," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 348–352.

[25] Y. Hu, P. N. Samarasinghe, T. D. Abhayapala, and S. Gannot, "Unsupervised multiple source localization using relative harmonic coefficients," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 571–575.

[26] 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.

[27] S. Hafezi, A. H. Moore, and P. A. Naylor, "Augmented intensity vectors for direction of arrival estimation in the spherical harmonic domain," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2017.

[28] Y. Hu, P. N. Samarasinghe, S. Gannot, and T. D. Abhayapala, "Semi-supervised multiple source localization using relative harmonic coefficients under noisy and reverberant environments," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, submitted.

[29] J. Taghia, J. Taghia, N. Mohammadiha, J. Sang, V. Bouse, and R. Martin, "An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4640–4643.

[30] J. K. Nielsen, M. S. Kavalekalam, M. G. Christensen, and J. Boldt, "Model-based noise PSD estimation from speech in non-stationary noise," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5424–5428.

[31] B. Rafaely and M. Kleider, "Spherical microphone array beam steering using wigner-d weighting," *IEEE Signal Processing Letters*, vol. 15, pp. 417–420, 2008.

[32] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "PSD estimation of multiple sound sources in a reverberant room using a spherical microphone array," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 76–80.

[33] Y. Hu, P. N. Samarasinghe, G. Dickins, and T. D. Abhayapala, "Modeling the interior response of real loudspeakers with finite measurements," in *2018 IEEE 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 16–20.

[34] ——, "Modeling characteristics of real loudspeakers using various acoustic models: Modal-domain approaches," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 561–565.

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