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 (referred 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.](Image)

2.1. Acoustic model
Assume a single sound source propagating from an unknown DOA, e.g., \((\theta_s, \phi_s)\) where \(0 < \theta_s < \pi\), \(0 < \phi_s < 2\pi\), with respect to the origin of a higher order microphone array.
(e.g., a spherical microphone array in Fig. 1). Let the microphone array has \(M\) microphones whose spherical coordinates are \(x_j = (r_j, \theta_j, \phi_j) (j = 1, \ldots, M)\), 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,

\[
p(x_j, n) = s(n) + a(x_j, n) + v(x_j, n), \quad j = 1, \ldots, M
\]  

(1)

where \(\ast\) is a convolution operation, \(s(n)\) is the source signal, \(a(x_j, n)\) is the room impulse response between the sound source and the \(j\)-th microphone, and \(v(x_j, n)\) 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,

\[
P(x_j, k) = Q(x_j, k) + V(x_j, k) = S(k)A(x_j, k) + V(x_j, k)
\]  

(2)

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

\subsection*{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,

\[
P(x_j, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k) Y_{nm}(\theta_j, \phi_j)
\]  

(3)

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

\[
Y_{nm}(\theta, \phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P^m_n(\cos \theta)e^{im\phi}
\]  

(4)

is the spherical harmonic function with order \(n\) and mode \(m\), \(P^m_n(\cdot)\) 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 \([26]\),

\[
\alpha_{nm}(k) = \frac{1}{j_n(kr)} \sum_{j=1}^{M} a_{j} P(x_j, k)^* Y_{nm}(\theta_j, \phi_j)
\]  

(5)

in which \(\ast\) 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,

\[
\alpha_{nm}(k) = y_{nm}(k) + \gamma_{nm}(k), \quad -n \leq m \leq n, 0 \leq n \leq N
\]  

(6)

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

\subsection*{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., \(y(k) = [y_{00}(k), \ldots, y_{nm}(k)]\), from the mixture measurements of \(\alpha_{nm}(k)\). Note that the noisy measurements in (6) can be rewritten as,

\[
\alpha_{nm}(k) = \beta_{nm}(k)y_{00}(k) + \gamma_{nm}(k)
\]  

(7)

where

\[
\beta_{nm}(k) = \frac{y_{nm}(k)}{y_{00}(k)}
\]  

(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., \(\beta(k) = [\beta_{00}(k), \ldots, \beta_{nm}(k)]^T\). (ii) recovering the \(y_{00}(k)\) using a beamformer. (iii) multiplying the estimated \(\beta(k)\) and \(y_{00}(k)\) to achieve accurate estimation of \(y(k)\).

\section*{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.

\subsection*{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,

\[
\frac{S_{n_{nm}n_{00}}(k)}{S_{n_{00}n_{00}}(k)} - \frac{S_{n_{nm}m_{00}}(k)}{S_{m_{00}m_{00}}(k)} = \frac{\beta_{nm}(k)}{\beta_{00}(k)}
\]  

(9)

where

\[
S_{n_{nm}n_{00}}(k) = \mathbb{E}\{\alpha_{nm}(k)\alpha_{00}^*(k)\} \quad S_{n_{00}n_{00}}(k) = \mathbb{E}\{\alpha_{00}(k)\alpha_{00}^*(k)\}
\]

\[
S_{n_{nm}m_{00}}(k) = \mathbb{E}\{\gamma_{nm}(k)\gamma_{m0}^*(k)\} \quad S_{m_{00}m_{00}}(k) = \mathbb{E}\{\gamma_{00}(k)\gamma_{00}^*(k)\}
\]  

(10)

and \(\mathbb{E}\{\cdot\}\) 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 \(S_{m_{00}m_{00}}(k)\) is unknown in practice. Some state-of-art power spectral density techniques, such as \([29, 30]\), are available to update the \(S_{m_{00}m_{00}}(k)\). For simplicity, we use the biased estimator by neglecting it in (9),

\[
\hat{\beta}_{nm}(k) \approx \frac{S_{n_{nm}n_{00}}(k)}{S_{n_{00}n_{00}}(k)}
\]  

(11)

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

\subsection*{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
sound source [20, 31, 32],
\[ S(k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \hat{\beta}^{m}_{nm}(\theta_{s}, \varphi_{s}) Y_{nm}(\theta_{s}, \varphi_{s}) \alpha_{nm}(k). \] (12)

The spherical harmonic coefficient of \( \tilde{y}_{00}(k) \) is equivalent to the received signal estimated in (12) [27],
\[ \tilde{y}_{00}(k) = S(k). \] (13)

However, the beamer approach in (12) requires the additional knowledge of source DOA, i.e., \((\hat{\theta}_{s}, \hat{\varphi}_{s})\), 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],
\[ y_{nm}(k) = S(k) 4\pi i^{n} Y_{nm}^{*}(\theta_{s}, \varphi_{s}) \] (14)
where \((\theta_{s}, \varphi_{s})\) is the source DOA. Following the definition in (8), we know its relative harmonic coefficients,
\[ \beta_{nm}(k) = 2\sqrt{\pi} i^{n} Y_{nm}^{*}(\theta_{s}, \varphi_{s}). \] (15)

Considering the cases up to the \(N\)-th order, we have a vector,
\[ \beta(\theta_{s}, \varphi_{s}) = [1, 2\sqrt{\pi} Y_{1,-1}(\theta_{s}, \varphi_{s}), \ldots, 2\sqrt{\pi} Y_{N N}(\theta_{s}, \varphi_{s})]^T. \] (16)
Assuming the continuous two dimension space \( \Phi = \{ (\theta_{s}, \varphi_{s}) : 0 < \theta_{s} < \pi, 0 < \varphi_{s} < 2\pi \} \) is sampled by \( S \) discrete candidates, we have a feature set using (16),
\[ \mathcal{H} = \{ \beta(\varphi_{1}, \varphi_{1}), \beta(\varphi_{2}, \varphi_{2}), \ldots, \beta(\varphi_{S}, \varphi_{S}) \} \] (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 \( \mathcal{H} \),
\[ \min_{(\theta_{s}, \varphi_{s})} \sum_{n=0}^{N} \sum_{m=-n}^{n} \left| \hat{\beta}^{m}_{nm} - 2\sqrt{\pi} i^{n} Y_{nm}^{*}(\theta_{s}, \varphi_{s}) \right|^2 \] (18)
where \( \hat{\beta}^{m}_{nm}(k) \) denotes the practical estimations at a single frequency, and \( \beta_{nm} \) 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,
\[ \bar{y}_{nm}(k) = \hat{\beta}^{m}_{nm}(k) \tilde{y}_{00}(k). \] (20)
Note that, at the \( k \)-th frequency bin, the \( \hat{\beta}^{m}_{nm}(k) \) is fixed, while the \( y_{nm}(k) \) is updated by the dynamic \( \tilde{y}_{00}(k) \) over the time-varying source signal. When the \( \{ \tilde{y}_{00}(k), \ldots, \tilde{y}_{NN}(k) \} \) over the entire STFT bins are estimated, we can then reconstruct spectrogram of any individual microphone on the array, using the spherical harmonic 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 \( 6 \times 4 \times 3 \), 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 white 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 2: Clean, noisy and enhanced soundfield over the microphone array when \( z = 0 \) (3 KHz, 5dB noise).
The metric for the received signal in the STFT domain is, the clean and estimated relative harmonic coefficients. Next, the estimation is evaluated by the mean absolute estimated error for harmonics coefficients, respectively. Accuracy of the DOA estimations for relative harmonic coefficients, DOA and received signal, respectively. Seeing the DOA estimation, we recognize the modified accuracy when the SNR level decreases. This signal, respectively. Table 1 depicts the accuracy of the estimations for relative harmonic coefficients under various SNR conditions ranging from 5 dB to 25 dB. However, the proposed method improves the accuracy of the spherical harmonic coefficients estimates are denoted 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 \((\varphi_s, \varphi_v) = (1.13, 3.94)\). We recognize that the enhanced soundfield generally gets rid of 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 \( y_00(k) \) 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.

<table>
<thead>
<tr>
<th>SNR level (dB)</th>
<th>Method</th>
<th>25</th>
<th>20</th>
<th>15</th>
<th>10</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR_{\beta} (dB)</td>
<td>-20.3</td>
<td>-15.0</td>
<td>-9.5</td>
<td>-5.5</td>
<td>-2.65</td>
<td></td>
</tr>
<tr>
<td>MAEE (degrees)</td>
<td>0.55</td>
<td>0.64</td>
<td>0.91</td>
<td>1.58</td>
<td>2.44</td>
<td></td>
</tr>
<tr>
<td>NMSE_{\beta0} (dB)</td>
<td>-15.1</td>
<td>-11.6</td>
<td>-7.9</td>
<td>-4.7</td>
<td>-2.2</td>
<td></td>
</tr>
</tbody>
</table>

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


