Decoupled Direction-of-Arrival Estimations Using Relative Harmonic Coefficients

Yonggang Hu[1], Thushara D. Abhayapala[1], Prasanga N. Samarasinghe[1], Sharon Gannot[2]
[1]*Audio and Acoustic Signal Processing Group*, *Australian National University*, Canberra, Australia
[2]*Faculty of Engineering*, *Bar-Ilan University*, Ramat-Gan, Israel

***Abstract*—Traditional source direction-of-arrival (DOA) estimation algorithms generally localize the elevation and azimuth simultaneously, requiring an exhaustive search over the two-dimensional (2-D) space. By contrast, this paper presents two decoupled source DOA estimation algorithms using a recently introduced source feature called the *relative harmonic coefficients*. They are capable to recover the source's elevation and azimuth separately, since the elevation and azimuth components in the relative harmonic coefficients are decoupled. The proposed algorithms are highlighted by a large reduction of computational complexity, thus enable a direct application for sound source tracking. Simulation results, using both a static and moving sound source, confirm the proposed methods are computationally efficient while achieving competitive localization accuracy.**

***Index Terms*—Decoupled DOA estimation, relative harmonic coefficients, fast speed, DOA tracking.**

I. Introduction

In recent decades, source direction-of-arrival (DOA) estimators [1]–[3] have been essential for many spatial signal processing techniques and applications including source dereverberation, speech separation [4], automatic speech recognition [5] and automated camera steering [6].

Traditional DOA estimations can be broadly divided into three types: (i) Algorithms requiring a time difference of arrival (TDOA) between microphone pairs, such as the popular generalized cross-correlation phase transform (GCC-PHAT) [7]. (ii) Steered response power (SRP) [8] and SRP-phase transform (SRP-PHAT) [9] based techniques, searching the desired source's DOA with higher response power. (iii) Subspace methods, more suitable for multiple source DOA estimation, utilize the spatial covariance matrix of the multi-source recordings. The most well-known examples include multiple signal classification (MUSIC) [10] and its extension into the spherical harmonics domain [11], [12] (i.e., SHD-MUSIC). Overall, those methods achieve satisfying localization accuracy while associating with a significant computational complexity since they require a 2-D grid searching over all possible directions. Generally, a higher grid resolution to sample the directional space increases the accuracy of the algorithms at the cost of an additional higher computational expense.

Two-dimensional DOA estimators with a low-complexity and competitive localization accuracy, have attracted substantial interest in the research community. Huang et al. reduced the computational complexity for the traditional SHD-MUSIC by rewriting the steering vectors as the product of elevation and azimuth components [13]. Hafezi et al. used the intensity vectors for localization methods and increased the algorithm speed using an optimized gradient decent search [14]. Habets et al. used a pseudointensity vector for low computational complexity DOA estimators [15], [16]. Some DOA estimation procedures, such as [17], exploited the sparsity property of the speech recordings, achieving a faster convergence rate using $\ell_{1,2}-$norm minimization approaches [18]. Some other methods used parametric methods such as least-squares estimation [19]–[21]. Additional efforts saved the computational cost by rewriting the complex-valued formulations using real-valued ones [22].

The current contribution presents two novel solutions to achieve accurate 2-D DOA estimates which is attributed by a dramatically reduced computational complexity. The algorithms use a recently introduced source feature called *relative harmonic coefficients*, which are solely dependent on the source DOA. Moreover, the source's elevation and azimuth angles are decoupled in the relative harmonic coefficients, thus it enables us to estimate the source's elevation and azimuth in two separate stages. The proposed methods are finally validated using a static source localization as well as a direct application for tracking a moving sound source.

This project has received funding from the European Unions Horizon 2020 Research and Innovation Programme, Grant Agreement No. 871245. Yonggang Hu is sponsored by CSC agency for funding.

II. Problem Addressed

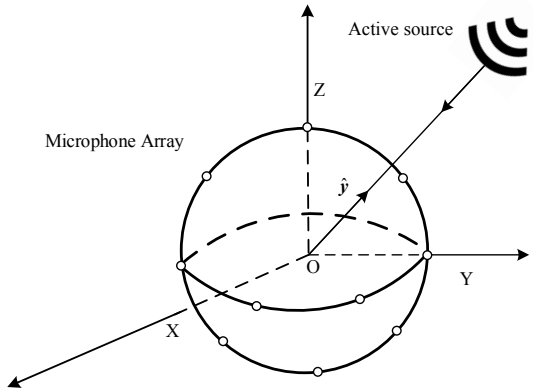

Fig. 1. DOA Estimation using a spherical microphone array.

Assume an active sound source propagating from an arbitrary DOA over the 2-D space, e.g., (ϑ_s, φ_s) where $0 < \vartheta_s < \pi$, $0 < \varphi_s < 2\pi$. The sound wave is impinging on a higher order microphone array (see Fig. 1). The array comprises of M microphones which polar coordinates are $\boldsymbol{x}_j = (r, \theta_j, \phi_j)$ $(j = 1, \ldots, M)$, with respect to its local origin O. The sound pressure, recorded by the array is represented in the frequency domain by

$$P(\boldsymbol{x}_j, k) = S(k)A(\boldsymbol{x}_j, k) + V(\boldsymbol{x}_j, k), \quad j = 1, \cdots, M \quad (1)$$

where $k = 2\pi f/c$ is the wavenumber, with f the frequency bin and c the speed of sound, $S(k)$ is the source signal, $A(\boldsymbol{x}_j, k)$ is the acoustic transfer function from the source to the j-th microphone, and $V(\boldsymbol{x}_j, k)$ is the additive noise signal. This paper aims to use the original noisy recordings to develop low-complexity DOA estimations, achieving competitive accuracy in comparison with the state-of-art approach.

III. Relative Harmonic Coefficients: a Review

This section reviews a spherical harmonics domain source feature of the relative harmonic coefficients, which are to be used by the proposed methods in the next section.

A. Definition of Relative Harmonic Coefficients

The sound pressure field in (1) can be represented in the spherical harmonics domain using the spherical harmonics expansion [23],

$$P(\boldsymbol{x}_j, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k)\, j_n(kr)\, Y_{nm}(\theta_j, \phi_j) \quad (2)$$

where $N = \lceil kr \rceil$ denotes the truncated order of soundfield which is set due to properties of the spherical Bessel functions [24], $j_n(\cdot)$ is the spherical Bessel function of the first kind,

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

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. In [25]–[27], the authors define the relative harmonic coefficients (RHC) as the ratios between the spherical harmonic coefficient $\alpha_{nm}(k)$ and $\alpha_{00}(k)$,

$$\boldsymbol{\beta}(k) = \left[1, \tfrac{\alpha_{1,-1}(k)}{\alpha_{00}(k)}, \cdots, \tfrac{\alpha_{nm}(k)}{\alpha_{00}(k)}, \cdots, \tfrac{\alpha_{NN}(k)}{\alpha_{00}(k)}\right]^T \quad (4)$$

which is a vector at the k-th frequency. Please refer to [25] for a source feature estimator under noisy conditions. Note that the defined relative harmonic coefficients in (4) are closely related to the relative-transfer function (RTF) decomposed into the spherical harmonic domain [25]. Early use of the relative harmonic coefficients was for a spherical harmonic domain noise reduction method [26].

B. Expression of Relative Harmonic Coefficients

This subsection reviews the expression of the relative harmonic coefficients assuming a free-field propagation [28]. We follow the common assumption to represent the sound pressure over the recording area using the plane waves modeling [29]–[31], whose spherical harmonic coefficients are,

$$\alpha_{nm}(k) = S(k) 4\pi i^n Y^*_{nm}(\vartheta_s, \varphi_s) \quad (5)$$

where $*$ represents the complex conjugate operator. Following (4), we have the explicit expression of the relative harmonic coefficients at order n and mode m,

$$\beta_{nm}(k) = 2\sqrt{\pi} i^n Y^*_{nm}(\vartheta_s, \varphi_s). \quad (6)$$

For the N-th order microphone array, the feature vector for the source propagating from (ϑ_s, φ_s) is defined as,

$$\boldsymbol{\beta}(\vartheta_s, \varphi_s) = \left[1, 2\sqrt{\pi} i Y^*_{1,-1}(\vartheta_s, \varphi_s), \cdots, 2\sqrt{\pi} i^N Y^*_{NN}(\vartheta_s, \varphi_s)\right]^T. \quad (7)$$

The properties of this feature vector are: (i) it is source signal independent while being solely dependent on the source DOA. (ii) assuming the continuous 2-D direction space $\boldsymbol{\Phi} = \{(\vartheta_s, \varphi_s) : 0 < \vartheta_s < \pi, 0 < \varphi_s < 2\pi\}$ is sampled by S discrete candidates, we can calculate a unique feature set as,

$$\mathcal{H} = \{\boldsymbol{\beta}(\vartheta_1, \varphi_1), \boldsymbol{\beta}(\vartheta_2, \varphi_2), \cdots, \boldsymbol{\beta}(\vartheta_S, \varphi_S)\} \quad (8)$$

which contains all possible feature vectors over the 2-D space when the S is large enough. (iii) it is frequency independent, however, as it may practically deviate from this expected independence, a frequency smoothing is suggested over a wide frequency band,

$$\bar{\boldsymbol{\lambda}}(\vartheta_s, \varphi_s) = \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{\lambda}(\vartheta_s, \varphi_s, k) \quad (9)$$

where $\boldsymbol{\lambda}$ denotes the practical measured feature vector (the analytical one is denoted using $\boldsymbol{\beta}$) at a single frequency bin and $\bar{\boldsymbol{\lambda}} = [\bar{\lambda}_{00}, \cdots, \bar{\lambda}_{NN}]^T$ denotes the smoothed vector over the K frequency bins. Please refer to [28] for more details about this subsection.

C. Source DOA Estimation

Assuming the relative harmonic coefficients due to a desired sound source are estimated using the feature estimator given in [25]. We then compute the practical smoothed vector, e.g., the $\bar{\boldsymbol{\lambda}}$ in (9), and compare this vector to the calculated/analytical set of $\mathcal{H}$ in (8) to recover the source's original DOA,

$$\text{argmin}_{(\vartheta_s, \varphi_s)} \sum_{n=0}^{N} \sum_{m=-n}^{n} \mid \bar{\lambda}_{nm} - 2\sqrt{\pi} i^n Y^*_{nm}(\vartheta_s, \varphi_s) \mid^2 \quad (10)$$

which uses a distance-based metric. However, this approach requires an exhaustive search over the 2-D directional set of $\mathcal{H}$, which is computational expensive. As explained in the next section, we show the relative harmonic coefficients are capable for faster DOA estimations, achieving a large reduction of the computational complexity.

IV. Decoupled DOA Estimations

This section exploits the relative harmonic coefficients to develop two decoupled source DOA estimations. Substituting (3) to (6), we have the detailed expression of the relative harmonic coefficients,

$$\beta_{nm}(\vartheta_s, \varphi_s) = \sqrt{\frac{(2n+1)(n-m)!}{(n+m)!}} P_{nm}(\cos\vartheta_s) i^n e^{-im\varphi_s} \quad (11)$$

where the associated Legendre function, $P_{nm}(\cdot)$, is a real-valued function. Table I lists exact expressions of (11) for the spherical harmonic modes up to $n = 2$ and $m = 1$. Both equation (11) and its specification in Table I imply that the components of source's elevation ϑ_s and azimuth φ_s are decoupled. Hence, assuming the relative harmonic coefficients are accurately estimated, we can recover the source's elevation and azimuth in two separate stages, as the methods explained in the sequel.

A. Method 1

This method estimates the source elevation in the first stage, which is then used to recover the azimuth in the second stage.

TABLE I
RELATIVE HARMONIC COEFFICIENTS UP TO THE 2ND ORDER.

(n, m)	$\beta_{nm}(k)$	(n, m)	$\beta_{nm}(k)$
(0, 0)	1	(2,-2)	$\sqrt{\frac{15}{8}}\sin^2(\vartheta_s)e^{2i\varphi_s}$
(1,-1)	$i\sqrt{\frac{3}{2}}\sin(\vartheta_s)e^{i\varphi_s}$	(2,-1)	$\sqrt{\frac{15}{8}}\sin(2\vartheta_s)e^{i\varphi_s}$
(1, 0)	$i\sqrt{3}\cos(\vartheta_s)$	(2, 0)	$\sqrt{\frac{5}{4}}\left(3\cos^2(\vartheta_s)-1\right)$
(1, 1)	$-i\sqrt{\frac{3}{2}}\sin(\vartheta_s)e^{-i\varphi_s}$	(2, 1)	$-\sqrt{\frac{15}{8}}\sin(2\vartheta_s)e^{-i\varphi_s}$

1) Stage One: Elevation Estimation: The magnitude of the relative harmonic coefficients in (11) is,

$$|\beta_{nm}(\vartheta_s)| = \sqrt{\frac{(2n+1)(n-m)!}{(n+m)!}}\,|P_{nm}(\cos\vartheta_s)| \tag{12}$$

which only depends on the elevation ϑ_s so that we omit its dependency on azimuth φ_s. Figure 2 demonstrates an example of the $|\beta_{1,-1}(\vartheta_s)|$ where the $0<\vartheta_s<\pi$. Combining (12) up to the N-th order, we have a vector of the magnitude,

$$|\boldsymbol{\beta}(\vartheta_s)| = [1, |\beta_{1,-1}(\vartheta_s)|, \cdots, |\beta_{NN}(\vartheta_s)|] \tag{13}$$

whose properties are summarized as follows: (i) Unique mapping: the vector of (13) has a unique mapping to the elevation when $0<\vartheta_s\leq\pi/2$. (ii) Calculated set: given the range of $0<\vartheta_s\leq\pi/2$, we have a unique set of (13) (see Fig. 2 for $|\beta_{1,-1}(\vartheta_s)|$):

$$\mathcal{H}_{\text{mag}} = \{|\boldsymbol{\beta}(\vartheta_1)|, |\boldsymbol{\beta}(\vartheta_2)|, \cdots, |\boldsymbol{\beta}(\vartheta_{S1})|\} \tag{14}$$

where $S1$ denotes the number of discrete elevation samples. Note that the set of $\mathcal{H}_{\text{mag}}$ is calculated analytically, not using any prior recordings. (iii) Symmetric: the vector of (13) is symmetric with respect to $\pi/2$ because $|P_{nm}(\cos\vartheta_s)| = |P_{nm}(\cos(\pi-\vartheta_s))|$ (see Fig. 2). Hence, we also have a unique mapping and set when $\pi/2<\vartheta_s<\pi$.

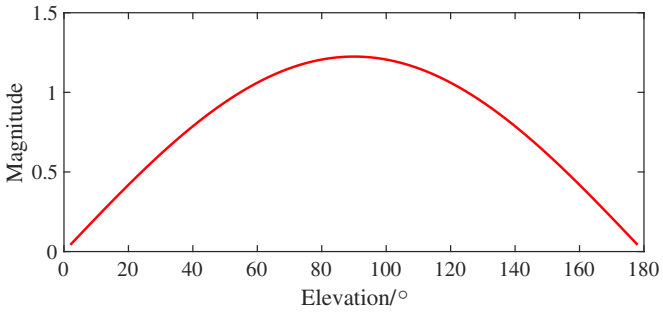

Fig. 2. $|\beta_{1,-1}(\vartheta_s)|$ for the elevation ranging from 0 to π.

Assuming the source's magnitude of relative harmonic coefficients are calculated, we show how to recover its elevation using the following two steps,

Step 1: Because of the symmetric w.r.t. $\pi/2$, the magnitude cannot distinguish whether the ϑ_s lies between $(0,\pi/2)$ or $(\pi/2,\pi)$. However, this can be known from the imaginary part of $\beta_{10}(k)$,

$$\text{Im}\{\beta_{10}(k)\} = \sqrt{3}\cos(\vartheta_s) \tag{15}$$

whose positive or negative characteristic only depends on the ϑ_s. Hence, we claim the estimated $\bar{\vartheta}_s$,

$$\begin{cases} 0<\bar{\vartheta}_s<\pi/2, & \text{if } \text{Im}\{\bar{\lambda}_{10}\}>0 \\ \pi/2\leq\bar{\vartheta}_s<\pi, & \text{if } \text{Im}\{\bar{\lambda}_{10}\}\leq 0. \end{cases} \tag{16}$$

Step 2: Exploiting the unique mapping between $(0,\pi/2)$ or $(\pi/2,\pi)$, we then use a distance-based metric to compare the measured magnitude vector with the set of $\mathcal{H}_{\text{mag}}$ in (14) to recover the source's elevation.

2) Stage Two: Azimuth Estimation: Since the source elevation has already been estimated, we can recover the source's φ_s by searching over all possible azimuths as follows,

$$\text{argmin}_{(\varphi_s)} \sum_{n=0}^{N}\sum_{m=-n}^{n} |\,\bar{\lambda}_{nm} - 2\sqrt{\pi}i^n Y^*_{nm}(\bar{\vartheta}_s,\varphi_s)\,|^2 \tag{17}$$

where γ_{nm} represents the smoothed relative harmonic coefficients measured in practice.

B. Method 2

Different from the above method, the second approach first estimates the azimuth and then recovers the source elevation.

1) Stage One: Azimuth Estimation: Let us define the ratio between the imaginary and real parts of the relative harmonic coefficients in (11),

$$\begin{aligned}\gamma_{nm}(\varphi_s) &= \frac{\text{Im}\{\beta_{nm}(\vartheta_s,\varphi_s)\}}{\text{Re}\{\beta_{nm}(\vartheta_s,\varphi_s)\}} \\ &= \begin{cases} -\tan(m\varphi_s), & \text{if } n=0,2,4,\cdots \\ -\cot(m\varphi_s), & \text{if } n=1,3,5,\cdots \end{cases}\end{aligned} \tag{18}$$

which only depends on the source's φ_s. Figure 3 exhibits $\gamma_{1,-1}(\varphi_s)$ where $0<\varphi_s<2\pi$. Combining the cases of (18) up to order N, we have a vector,

$$\boldsymbol{\gamma}(\varphi_s) = [\gamma_{00}(\varphi_s), \gamma_{1,-1}(\varphi_s), \cdots, \gamma_{NN}(\varphi_s)]. \tag{19}$$

which properties are: (i) Unique mapping: the vector has a unique mapping to the azimuth φ_s when $0<\varphi_s<\pi$ (Note that Fig. 3 is not fully unique when $0<\varphi_s<\pi$ but it only contains a single mode in the vector (19)). (ii) Calculated set: given the range when $0<\varphi_s<\pi$, we have a unique set as:

$$\mathcal{H}_{\text{tan}} = \{\boldsymbol{\gamma}(\varphi_1), \boldsymbol{\gamma}(\varphi_2), \cdots, \boldsymbol{\gamma}(\varphi_{S2})\} \tag{20}$$

where $S2$ denotes the number of discrete samples. (iii) Periodic: the vector in (19) is periodic by π because of the tan/cot functions in (18) (see Fig. 3). Therefore, we also have a unique mapping and set of (20) when $\pi<\varphi_s<2\pi$.

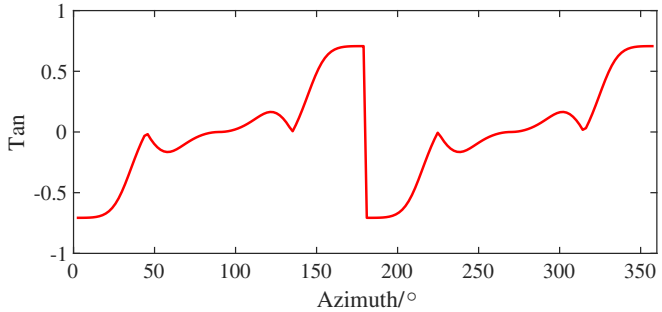

Fig. 3. $\gamma_{1,-1}(\varphi_s)$ for the azimuth ranging from 0 to 2π.

In the followings, we explain how to estimate the source's azimuth given the source's $\boldsymbol{\gamma}(\varphi_s)$ from an unknown direction.

Step 1: The source's $\boldsymbol{\gamma}(\varphi_s)$ cannot distinguish whether the sound source lies in $(0,\pi)$ or $(\pi,2\pi)$ because it is periodic by π. However, the real part of $\beta_{1,-1}(k)$ carries such information,

$$\text{Re}\{\beta_{1,-1}(k)\} = -\sqrt{\frac{3}{2}}\sin(\vartheta_s)\sin(\varphi_s). \tag{21}$$

Since $\sin(\vartheta_s) > 0$, we claim the estimated $\bar{\varphi}_s$,

$$\begin{cases} 0 < \bar{\varphi}_s \leq \pi, & \text{if } \mathrm{Re}\{\bar{\lambda}_{1,-1}\} \leq 0 \\ \pi < \bar{\varphi}_s < 2\pi, & \text{if } \mathrm{Re}\{\bar{\lambda}_{1,-1}\} > 0. \end{cases} \tag{22}$$

Step 2: Exploiting the mapping between $(0, \pi)$ or $(\pi, 2\pi)$, we also adopt a distance-based metric between the measured vector and the set of $\mathcal{H}_{\text{tan}}$ to recover its φ_s.

2) Stage Two: Elevation Estimation: Since the source azimuth is given by the first stage, we can recover the source ϑ_s by searching over all possible elevations,

$$\text{argmin}_{(\vartheta_s)} \sum_{n=0}^{N} \sum_{m=-n}^{n} \mid \bar{\lambda}_{nm} - 2\sqrt{\pi} i^n Y_{nm}^*(\vartheta_s, \bar{\varphi}_s) \mid^2 . \tag{23}$$

V. Simulations

This section uses extensive simulations to validate the proposed approaches. We conduct the evaluations in a reverberant room, whose size is $6 \times 4 \times 3$ m for the length, width and height, respectively. We simulate the incoming soundfield of an open-sphere spherical microphone array (with 32 channels and radius 4.2 cm). Note that, although a spherical microphone array is used, the proposed algorithms are equally applicable for other structured arrays as well such as planar microphone arrays [32]. We use an available toolbox,[1] that implements the image source method [33], to generate the room impulse response (RIR) from the sound source to the microphone array. We use the estimator ever discussed in [25] to estimate the relative harmonic coefficients. The original DOA estimation in (10) is taken as the baseline approach for comparisons. The source feature set of $\mathcal{H}$, $\mathcal{H}_{\text{mag}}$, and $\mathcal{H}_{\text{tan}}$ are computed by sampling both the elevation and azimuth with a resolution of 2 degrees. The algorithms use fifty frequency bins ranging from 1600 Hz to 2400 Hz, recording the soundfield up to the 2nd order ($N = \lceil kr \rceil$), so that the vector's dimension is 9. Lower frequency bins reduce the uniqueness of the feature set as the vector's dimension is reduced to 4, and higher frequency bins contain less speech components. Performance of our system is measured using the mean absolute estimated error (MAEE/°) between the estimated and original DOA,

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

A. DOA Estimations for a Static Sound Source

We first examine the performance of the methods using a static sound source, randomly positioned within the room. The evaluation results shown in the Table II and III denote the mean value over twenty trials. Table II presents the localization errors using the proposed algorithms in diverse reverberation levels. The increased reverberation level implies more negative impacts caused by the coherent reverberations, so that the localization accuracy degrades. Table III exhibits the localization errors under various noisy conditions (white Gaussian noise), where the SNR level ranges from 5 dB to 25 dB. Since the feature estimator has already taken the noise into account, we recognize only minor performance degradation when the SNR level decreases. Examinations of the proposed methods in both Table II and III confirm that, although our proposed methods only use a 1-D search, they still achieve competitive localization accuracy in comparison with the baseline which uses a 2-D search.

This paper emphasizes the speed of the proposed methods. For validations, we measure the computational complexity of the algorithms by directly recording the computational time, using a Matlab implementation on a standard desktop (CPU Intel Core i7-4790 Quad 3.6 GHz, RAM 16 GB). Table IV presents the computation time of all the algorithms. We observe that the proposed methods only take less than 0.4 ms (average time over twenty tests) to search the source DOA over the 2-D direction space, achieving dramatically improved speed compared to the baseline approach.

TABLE II
Source DOA estimation error under various reverberation levels where the SNR is 25 dB.

MAEE/°	T_{60} (s)				
Methods	0.2	0.3	0.4	0.5	0.6
Baseline	1.01	1.33	1.74	1.93	2.31
Method 1	1.00	1.38	1.76	1.91	2.34
Method 2	1.05	1.90	2.48	3.28	4.05

TABLE III
Source DOA estimation error under various SNR levels where $T_{60} = 0.4$ s.

MAEE/°	SNR level (dB)				
Methods	5	10	15	20	25
Baseline	1.87	1.64	1.82	1.64	1.74
Method 1	1.96	1.71	1.82	1.69	1.76
Method 2	2.84	2.66	2.32	2.40	2.48

TABLE IV
Execution time (averaged over 20 trials), searching the DOA.

Methods	Baseline	Proposed1	Proposed2
Time cost (ms)	309.5	7.9	8.1

B. DOA Tracking for a Moving Sound Source

Encouraged by the reduced complexity, we apply the proposed approaches to sound source tracking, as explained in the following three steps. Firstly, we generate the moving source's time-domain recordings in the above simulated room, using an available toolbox designed for sound source localization/tracking [34] (SNR level is 15 dB). Then, we split the measured recordings into the 0.5 s long frames, and use the decoupled algorithms to estimate the source DOA over the frames. Finally, we use two successive estimates (e.g., the t-th and $(t-1)$-th frames) to denote the eventual t-th estimate,

$$\bar{\Phi}(t) = w\Phi(t) + (1-w)\bar{\Phi}(t-1) \tag{25}$$

where t is the index of the frame, $\Phi(t)$ is the current estimated DOA of $(\vartheta_s^t, \varphi_s^t)$, and w denotes its weight (set at 0.7 in the simulation). Figure 4 and 5 exhibit the estimated source trajectory using the two proposed methods, respectively. Both the randomly generated trajectories by the moving source have been recovered accurately within a fast response time, which only takes about 0.1 s to process each instantaneous frame.

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

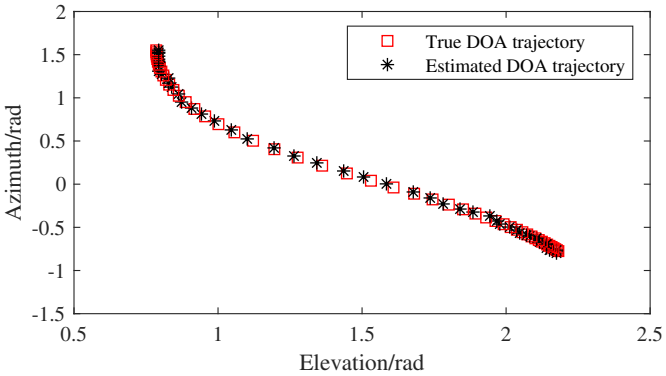

Fig. 4. Original and estimated source trajectory using the Method 1.

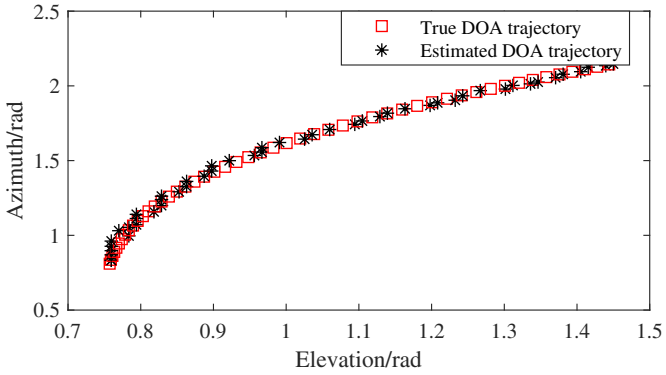

Fig. 5. Original and estimated source trajectory using the Method 2.

VI. Conclusion

This paper presented decoupled 2-D source DOA estimations using a recently introduced source feature of *relative harmonic coefficients*. Evaluations in both single source localization and tracking have confirmed the dramatic reduction in computational complexity, while achieving competitive accuracy. This paper mainly considered a far-field scenario, where the distance between the sound source and recording area was ignored. In the future, the authors intend to investigate the decoupled source localization under a near-field scenario and provide more analysis in a noisy and reverberate environment.

References

[1] 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.

[2] C. Evers, H. Loellmann, H. Mellmann, A. Schmidt, H. Barfuss, P. Naylor, and W. Kellermann, "The LOCATA challenge: Acoustic source localization and tracking," *arXiv preprint arXiv:1909.01008*, 2019.

[3] 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.

[4] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "PSD estimation and source separation in a noisy reverberant environment using a spherical microphone array," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018.

[5] F. Asano, M. Goto, K. Itou, and H. Asoh, "Real-time sound source localization and separation system and its application to automatic speech recognition," in *Seventh European Conference on Speech Communication and Technology*, 2001.

[6] Y. Hu, J. Benesty, and G. W. Elko, "Passive acoustic source localization for video camera steering," in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, pp. 909–912.

[7] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE transactions on acoustics, speech, and signal processing*, vol. 24, no. 4, 1976.

[8] K. Yao, J. C. Chen, and R. E. Hudson, "Maximum-likelihood acoustic source localization: experimental results," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, pp. 2949–2952.

[9] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*. Springer Science and Business Media, 2013.

[10] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[11] D. Khaykin and B. Rafaely, "Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 221–224.

[12] L. Birnie, T. D. Abhayapala, H. Chen, and P. N. Samarasinghe, "Sound source localization in a reverberant room using harmonic based MUSIC," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 651–655.

[13] 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.

[14] 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.

[15] D. P. Jarrett, E. A. Habets, and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *2010 IEEE 18th European Signal Processing Conference (EUSIPCO)*, pp. 442–446.

[16] A. Herzog and E. Habets, *On the Relation Between DOA-Vector Eigenbeam ESPRIT and Subspace Pseudointensity-Vector*, 2019.

[17] P. Stoica, P. Babu, and J. Li, "Spice: A sparse covariance-based estimation method for array processing," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629–638, 2010.

[18] B. Lin, J. Liu, M. Xie, and J. Zhu, "Direction-of-arrival tracking via low-rank plus sparse matrix decomposition," *IEEE Antennas and Wireless Propagation Letters*, vol. 14, pp. 1302–1305, 2015.

[19] O. Thiergart, W. Huang, and E. A. P. Habets, "A low complexity weighted least squares narrowband doa estimator for arbitrary array geometries," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 340–344.

[20] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 7, pp. 984–995, 1989.

[21] H. Teutsch and W. Kellermann, "Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5276–5279.

[22] K.-C. Huarng and C.-C. Yeh, "A unitary transformation method for angle-of-arrival estimation," *IEEE Transactions on Signal Processing*, vol. 39, no. 4, pp. 975–977, 1991.

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

[24] 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.

[25] 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.

[26] D. P. Jarrett and E. A. Habets, "On the noise reduction performance of a spherical harmonic domain tradeoff beamformer," *IEEE Signal Processing Letters*, vol. 19, no. 11, pp. 773–776, 2012 good.

[27] 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*, p. submitted, 2020.

[28] 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.

[29] Y. Hu, P. N. Samarasinghe, G. Dickins, and T. D. Abhayapala, "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.

[30] ——, "Modeling the interior response of real loudspeakers with finite measurements," in *2018 IEEE 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 16–20.

[31] B. Rafaely, *Fundamentals of spherical array processing*. Springer, 2015, vol. 8.

[32] 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.

[33] 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.

[34] C. Blandin, A. Ozerov, and E. Vincent, "Multi-source TDOA estimation in reverberant audio using angular spectra and clustering," *Signal Processing*, vol. 92, no. 8, pp. 1950–1960, 2012.