CLOSED-FORM SINGLE SOURCE DIRECTION-OF-ARRIVAL ESTIMATOR USING FIRST-ORDER RELATIVE HARMONIC COEFFICIENTS

Yonggang Hu[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

The *relative harmonic coefficients* (RHC), recently introduced as a multi-microphone spatial feature, demonstrates promising performance when applied to direction-of-arrival (DOA) estimation. All existing RHC-based DOA estimators suffer from a resolution limitation due to the inherent grid-based search. In contrast, this paper utilizes the first-order RHC to propose a closed-form DOA estimator by deriving a direction vector, which points towards to the desired source direction. Two objective metrics, namely localization accuracy and algorithm complexity, are adopted for the evaluation and comparison with existing RHC-based and intensity based localization approaches, in both simulated and real-life environments.

Index Terms— Direction-of-arrival estimation, first-order relative harmonic coefficients, closed-form, intensity vector.

1. INTRODUCTION

In recent decades the problem of acoustic source direction-of-arrival (DOA) estimation using multi-microphones has attracted a significant amount of attention in the acoustic signal processing community. The DOAs of the sound sources are essential components in many spatial signal processing techniques and applications, such as dereverberation, beamforming, speech separation, enhancement, automatic speech recognition and sound event detection [1].

Many early single source localization approaches exploit the time difference of arrival (TDOA) between microphone pairs [2]. Other popular approaches are based on the beamforming, and a typical example is the steered response power (SRP) method [3], which scans all possible directions over the two-dimensional (2-D) directional space to search for areas with higher response power. In recent decade, we have witnessed a growing interest in the use of circular/spherical microphone arrays for sound source localization [4–7]. The two- or three-dimensional (3-D) soundfield can be analyzed by its decomposition into the spherical harmonics domain (SHD, also called modal domain) using a set of orthogonal spatial functions called spherical harmonics functions [8]. SRP-based method, implemented in the spherical harmonics domain, was initially studied in [9]. In [10], the SHD-based minimum variance distortionless response (SHD-MVDR) was proposed to localize a near-field sound source, where the distance between the source and the microphone array was also determined. Subspace-based localization methods, originally defined in the measurement space, in [11], were extended to the SHD, resulting in the SHD Multiple Signal Classification (SHD-MUSIC) method. An improved version of the SHD-MUSIC [12] achieved higher localization accuracy using a frequency smoothing technique to decorrelate the coherent source signal. Other improvements of the SHD-MUSIC based methods, more suitable to reverberant and noisy environments, were recently proposed in [5] and [13]. Another improvement with reduced algorithm complexity is the eigenbeam estimation of signal parameters via rotational invariance (EB-ESPRIT) [14].

Spatial features, such as the above-mentioned TDOA, SRP and spatial covariance matrix, used by the subspace methods, are vital factors influencing the accuracy of the localization algorithms, as they contain relevant cues of the source(s) to be localized. Recently, Hu *et al.* proposed a new spatial feature denoted *relative harmonic coefficients* (RHC) [15], which is formally defined as the normalization between spherical harmonic coefficients and the corresponding reference harmonic coefficients at the zero order. The normalization cancels out the contribution of the source signal, such that the RHC solely depends on the source position in a static acoustic environment. The RHCs are the SHD counterparts of the relative transfer functions (RTFs) [16–18]. The RHCs are spatially more discriminative than the RTFs as the spherical harmonic decomposition enhances the spatial resolution over space at the cost of additional computations. A recent study [19] confirmed that semi-supervised RHC-based source localization method outperforms the RTF-based method under equally noisy and reverberant conditions. Related work [15, 19, 20] highlight the promising properties of the RHC as source features for sound source localization, including (i) the independence of the time-varying source signal; (ii) the dependence on the source's position; (iii) simple estimation procedure from the noisy recordings; and (iv) the ability to detect the single-source components within the mixed recordings thus handling the challenging multi-source localization problems.

Several RHC-based single source localization techniques have been proposed, including (i) a 2-D estimator using an exhaustive grid search over the 2-D directional space [20]; (ii) a decoupled estimator which localizes the source elevation and azimuth using separate one-dimensional (1-D) search over the elevation and azimuth space, respectively [21]; and (iii) an optimized estimator using a gradient-descent search over the 2-D directional space [22]. All three RHC-based DOA estimators suffer from resolution limitations, as they all require a grid-search over the directional space, while using different search strategies. Alternatively, this paper proposes a closed-from RHC-based DOA estimator by deriving a direction vector, which points towards to the desired source direction, thus circumvents the exhaustive search over the elevation or azimuth directional space, while achieving equivalent localization accuracy. In addition to RHC-based methods, the proposed algorithm is also analyzed and compared with another source localization approach that uses the acoustic intensity vector as the input feature [23, 24], which also solely utilizes the first-order spherical harmonics coefficients in closed-form solution. In the sequel, we first introduce the first-order relative harmonic coefficients. Then, in Section 3 we present the new localization algorithm. Extensive experimental results are reported in Section 4. Finally, Section 5 concludes the work.

2. FIRST-ORDER RELATIVE HARMONIC COEFFICIENTS

This section introduces some prerequisite knowledge on the first-order relative harmonic coefficients, used by the proposed localization algorithm.

Assume a single sound source propagating from an unknown DOA, e.g., $\Phi = (\vartheta_s, \varphi_s)$ where $0 < \vartheta_s < \pi$, $0 < \varphi_s < 2\pi$, with respect to the origin of a microphone array (e.g., a spherical microphone array). The microphone array comprises M microphones whose spherical coordinates are $\boldsymbol{x}_j = (r, \theta_j, \phi_j)$ $(j = 1, \ldots, M)$. The sound pressure in the frequency domain, as measured by the j-th channel, is given by,

$$P(\boldsymbol{x}_j, k) = S(k)A(\boldsymbol{x}_j, k) + V(\boldsymbol{x}_j, k) \tag{1}$$

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, $A(\boldsymbol{x}_j, k)$ denotes the room transfer function between the sound source and the j-th microphone, and $V(\boldsymbol{x}_j, k)$ denotes the additive noise signal. The sound pressure, measured at the j-th microphone within the recording area, can be further decomposed into the spherical harmonics domain [8],

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

where $n(\geqslant 0)$ and m are integers, $N = \lceil kr \rceil$ is the truncated order of the soundfield [25], and

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

is the spherical harmonic function which is orthogonal over the 2-D directional space, i.e.,

$$\int_{\mathbb{S}^2} Y_{nm}(\hat{\boldsymbol{\phi}}) Y^*_{m'n'}(\hat{\boldsymbol{\phi}}) d(\hat{\boldsymbol{\phi}}) = \delta_{mm'}\delta_{nn'} \tag{4}$$

where $[\cdot]^*$ denotes the complex conjugate operator, $\hat{\boldsymbol{\phi}}$ denotes the angular direction and $\delta_{nn'}$ denotes the Kronecker delta function

$$\delta_{nn'} = \begin{cases} 1, & n = n' \\ 0, & n \neq n'. \end{cases} \tag{5}$$

$P_{nm}(\cdot)$ represents the associated Legendre function, and $b_n(\cdot)$ denotes the function based on the configuration of the microphone array,

$$b_n(kr) = \begin{cases} j_n(kr), & \text{for an open array} \\ j_n(kr) - \frac{j_n^{'}(kR)}{h_n^{'}(kR)} h_n(kr), & \text{for a rigid array} \end{cases} \tag{6}$$

where R denotes the radius of the spherical microphone array, $j_n^{'}(\cdot)$ and $h_n^{'}(\cdot)$ denote the partial derivatives of spherical Bessel and Hankel functions, respectively.

The term $\alpha_{nm}(k)$ in (2) denotes the spherical harmonic coefficients, containing compact information characterizing the captured soundfield in the spherical harmonics domain. In general, we assume the received signals in (1) to follow a plane wave model (i.e., far-field scenarios) [26], since the aperture of the recording area is much smaller compared to the source-array distance. In this scenario $\alpha_{nm}(k)$ due to the direct sound source follows,

$$\alpha_{nm}(k) = S(k) 4\pi i^n Y^*_{nm}(\vartheta_s, \varphi_s) \tag{7}$$

where (ϑ_s, φ_s) denotes the source DOA. The RHCs are formally defined in [15, 19, 20] as the ratio between the spherical harmonic coefficient $\alpha_{nm}(k)$ and $\alpha_{00}(k)$, so that its expression with order n and mode m takes the form,

$$\beta_{nm}(k) = 2\sqrt{\pi} i^n Y^*_{nm}(\vartheta_s, \varphi_s). \tag{8}$$

Based on the above definition, the first-order components of the RHC[1] are,

$$\boldsymbol{\beta}(\vartheta_s, \varphi_s) = \left[2\sqrt{\pi} i Y^*_{1,-1}(\vartheta_s, \varphi_s), 2\sqrt{\pi} i Y^*_{1,0}(\vartheta_s, \varphi_s), 2\sqrt{\pi} i Y^*_{1,1}(\vartheta_s, \varphi_s)\right]^T, \tag{9}$$

which has some unique properties summarized as follows [20]: (i) source signal independent; (ii) involve spatial information: its only degree-of-freedom coincides with the source DOA; and (iii) frequency independent: it is expected to be consistent over a wide frequency band. The frequency index k in (9) is omitted since the terms are frequency-independent. Based on first-order RHC, this paper aims at a new DOA estimator that localizes the unknown source elevation ϑ_s and azimuth φ_s using a closed-form solution, as elaborated in the next section.

3. PROPOSED CLOSED-FORM DOA ESTIMATOR USING FIRST-ORDER RHC

This section uses the first-order RHC to propose a DOA estimator with a closed-form solution, then summarizes its implementation procedure in practice, and finally highlights some promising properties of the algorithm.

3.1. Direction vector

Let us now investigate the expanded formula of the first-order RHC by substituting the exact expression of $Y^*_{nm}(\vartheta_s, \varphi_s)$ into the feature definition in (8) and decomposing the complex exponential function into their imaginary and real components, respectively,

$$\begin{aligned} \beta_{1,-1} &= i\sqrt{3/2}\sin(\vartheta_s) e^{i\varphi_s} \\ &= -\sqrt{3/2}\sin\vartheta_s\sin\varphi_s + i\sqrt{3/2}\sin\vartheta_s\cos\varphi_s \\ \beta_{1,0} &= i\sqrt{3}\cos(\vartheta_s) \\ \beta_{1,1} &= -i\sqrt{3/2}\sin(\vartheta_s) e^{-i\varphi_s} \\ &= -\sqrt{3/2}\sin\vartheta_s\sin\varphi_s - i\sqrt{3/2}\sin\vartheta_s\cos\varphi_s. \end{aligned} \tag{10}$$

These terms are weighted nonlinear combinations of the source elevation and azimuth angles, as arguments of the $\sin(\cdot)$ and $\cos(\cdot)$ functions, respectively.

Theorem: Let the estimated first-order RHC, denoted as $\bar{\beta}_{1,-1}$,

[1] Theoretically, the RHC comprise higher orders, depending on the structure and capacity of the microphones. By contrast, the proposed algorithm focuses on the first-order RHC, imposing less restrictions on the microphone array, e.g. the number of microphones can be relatively small (e.g., the microphone numbers). For example, the B-format based microphone array is able to capture the first-order harmonics using four point microphones.

$\bar{\beta}_{1,0}$, and $\bar{\beta}_{1,1}$.[2] Derive the *direction vector*:

$$\bar{\boldsymbol{I}} = \begin{bmatrix} \mathrm{Im}\{\bar{\beta}_{1,-1} - \bar{\beta}_{1,1}\} \\ \mathrm{Re}\{\bar{\beta}_{1,-1} + \bar{\beta}_{1,1}\} \\ \mathrm{Im}\{\bar{\beta}_{1,0}\} \end{bmatrix} \otimes \begin{bmatrix} \sqrt{1/6} \\ -\sqrt{1/6} \\ \sqrt{1/3} \end{bmatrix} \tag{11}$$

where $\otimes$ denotes the element-wise multiplication operator, and $\mathrm{Im}\{\cdot\}$ and $\mathrm{Re}\{\cdot\}$ denote the imaginary and real parts of the complex-valued input, respectively. If the estimated first-order RHC coefficients are equal to their theoretical values, the direction vector $\bar{\boldsymbol{I}}$ points towards the source direction.

Proof: Substituting the $\bar{\beta}_{1,-1}$, $\bar{\beta}_{1,0}$, and $\bar{\beta}_{1,1}$ with their theoretical values in (10), the direction vector $\bar{\boldsymbol{I}}$ in (11) simplifies to:

$$\begin{aligned} &\sqrt{1/6} \times \mathrm{Im}\{\bar{\beta}_{1,-1} - \bar{\beta}_{1,1}\} = \sin\vartheta_s \cos\varphi_s \\ &-\sqrt{1/6} \times \mathrm{Re}\{\bar{\beta}_{1,-1} + \bar{\beta}_{1,1}\} = \sin\vartheta_s \sin\varphi_s \\ &\sqrt{1/3} \times \mathrm{Im}\{\bar{\beta}_{1,0}\} = \cos\vartheta_s. \end{aligned} \tag{12}$$

Hence, the vector $\bar{\boldsymbol{I}} = \left[\sin\vartheta_s \cos\varphi_s, \sin\vartheta_s \sin\varphi_s, \cos\vartheta_s\right]^T$ represents the unit vector exactly pointing towards the source DOA.

The next subsection shows how to estimate the first-order RHC in the vector $\bar{\boldsymbol{I}}$ and briefly summarizes the procedure to implement the algorithm in practice.

3.2. Proposed algorithm

The implementation of the algorithm comprises six steps as summarized below:

(i) Assume a spherical microphone array is measuring the soundfield of an unknown single sound source propagating from an arbitrary direction.

(ii) Transform the time-domain multichannel recordings to the short-time Fourier transform (STFT) domain, with $t \in \{1, \ldots, T\}$ and $k \in \{1, \ldots, K\}$, where t and k denote the time and frequency index, respectively.

(iii) Decompose the sound pressure of the multichannel signals into the spherical harmonics domain and take the first-order spherical harmonic coefficients using the spherical microphone array:

$$\bar{\alpha}_{nm}(t,k) = \frac{1}{b_n(kr)} \sum_{j=1}^{M} a_j P_{\boldsymbol{x}_j}(t,k)\, Y^*_{nm}(\theta_j, \phi_j) \tag{13}$$

where a_j denotes the weight of each microphone that minimizes the error between the measured and theoretical quantities.

(iv) Use $[\bar{\alpha}_{00}(t,k), \bar{\alpha}_{1,-1}(t,k), \bar{\alpha}_{1,0}(t,k), \bar{\alpha}_{1,1}(t,k)]^T$ from (13) to calculate the RHC in noisy environments using the following estimator:

$$\bar{\beta}_{nm}(k) \approx \frac{S_{\bar{\alpha}_{nm}\bar{\alpha}_{00}}(k)}{S_{\bar{\alpha}_{00}\bar{\alpha}_{00}}(k)}, \quad n = 1, m = -1, 0, 1 \tag{14}$$

where

$$\begin{aligned} S_{\bar{\alpha}_{nm}\bar{\alpha}_{00}}(k) &= \frac{1}{T} \sum\nolimits_{t=1}^{T} \left\{\bar{\alpha}_{nm}(t,k)\bar{\alpha}^*_{00}(t,k)\right\} \\ S_{\bar{\alpha}_{00}\bar{\alpha}_{00}}(k) &= \frac{1}{T} \sum\nolimits_{t=1}^{T} \left\{\bar{\alpha}_{00}(t,k)\bar{\alpha}^*_{00}(t,k)\right\} \end{aligned} \tag{15}$$

denote the power spectral density (PSD) and CPSD (cross PSD) of the measured signals over the T time-varying source frames [19].

[2] To distinguish from the analytical RHCs denoted $\beta_{1,-1}$, $\beta_{1,0}$, and $\beta_{1,1}$ in (10), the practical measured first-order RHCs are denoted $\bar{\beta}_{1,-1}$, $\bar{\beta}_{1,0}$, and $\bar{\beta}_{1,1}$, respectively.

(v) Apply frequency-smoothing to the $\bar{\boldsymbol{\beta}}(k) = [\bar{\beta}_{1,-1}(k), \bar{\beta}_{1,0}(k), \bar{\beta}_{1,1}(k)]^T$ from (14), as the features may slightly differ from the frequency-independence property, i.e.,

$$\bar{\boldsymbol{\beta}} = \sum_{k=1}^{K} \gamma(k) \bar{\boldsymbol{\beta}}(k) \tag{16}$$

where $\gamma(k)$ denotes a weighting function over the K frequency bins with $\gamma(k) = 1/K$ as a default value, and $\bar{\boldsymbol{\beta}}$ the smoothed RHC of the sound source.

(vi) Substitute $\bar{\boldsymbol{\beta}} = [\bar{\beta}_{1,-1}, \bar{\beta}_{1,0}, \bar{\beta}_{1,1}]^T$ from (16) into the closed-form solution (11), and normalize, as the estimated direction vector may deviate from the unit-norm property, and obtain a practical direction vector

$$\hat{\boldsymbol{I}} = \frac{\bar{\boldsymbol{I}}}{||\bar{\boldsymbol{I}}||} \tag{17}$$

where $\|\cdot\|$ stands for an ℓ_2-norm.

3.3. Algorithm properties

A few comments on the theory and implementation procedures of the proposed localization algorithm:

- *Usability with other microphone constellations*: although a spherical microphone array is used by this paper, the proposed algorithm is independent of the specific microphone constellation provided the array facilitates first-order spherical harmonics decomposition of the soundfield. Hence, the algorithm is equally applicable to other structured arrays such as B-format array [27], planar array [28] and multiple circular arrays [29], by adjusting the estimates of spherical harmonics coefficients in (13).
- *Computational-efficiency*: the efficiency of the proposed algorithm stems from two reasons: (i) In (14) and (16) we applied time averaging and frequency smoothing, respectively, thus circumventing the need to localize the source for each time-frequency bin; and (ii) Eq. (11) is a closed-form expression for localizing the source, circumventing the tedious grid search applied in baseline algorithms.
- *Robustness to noise*: the algorithm is less sensitive to noise than baseline methods due to the time-averaging in (15) that reduce the negative impact of the noise signals.

4. EXPERIMENTS

This section presents experimental results for the proposed localization algorithm using both simulated and real-life source recordings and a comparison with baseline methods.

4.1. Metric and baseline approaches

For a comprehensive evaluation of the proposed method, we adopt two baseline localization approaches. One is an RHC-based decoupled DOA estimator, which localizes the source elevation and azimuth using two separate stages [21]. This method requires a search over the directional space, so that both the elevation and azimuth space are sampled with a dense one-degree resolution. The second baseline is the intensity based approach using the first-order spherical harmonic coefficients [23, 27]. Note that, similar to (16), the intensity-based approach also averages its vector over a wide frequency band. For convenience, the three DOA estimators above are referred to by their abbreviations, namely 'Intensity', 'Decoupled', and 'Proposed', respectively.

The following simulations examine and compare the three algorithms under diverse noisy and reverberant conditions. For a given reverberation and SNR level, we repeat the experiments $\mathcal{M}_{\text{tot}} > 1$ times, each using a source driven by the same signal while located

Table 1: Localization accuracy for various SNR levels. $\mathcal{M}_{\text{tot}} = 100$ tests in an anechoic room.

SR/$\overline{\text{MAEE}}$	Localization methods		
SNR level	Intensity	Decoupled	Proposed
5 dB	82%/3.54°	100%/0.35°	100%/0.33°
15 dB	98%/1.21°	100%/0.17°	100%/0.12°
25 dB	100%/0.33°	100%/0.15°	100%/0.07°

Table 2: Localization accuracy for various reverberation levels. $\mathcal{M}_{\text{tot}} = 100$ tests and SNR = 10 dB.

SR/$\overline{\text{MAEE}}$	Reverberation time (T_{60})		
Methods	150 ms	350 ms	550 ms
Decoupled	100%/1.05°	99%/3.56°	88%/4.69°
Proposed	100%/0.97°	99%/3.26°	93%/4.62°

at a randomly selected DOA over the directional space. For each of the m-th ($m = 1, \ldots, \mathcal{M}_{\text{tot}}$) experiments, we examine the localization accuracy using the mean absolute estimated error (MAEE/°) between the estimated and true source DOA defined as:

$$\text{MAEE}^m = \frac{1}{2}\left(|\theta^m_{\text{true}} - \theta^m_{\text{est}}| + |\phi^m_{\text{true}} - \phi^m_{\text{est}})\,|\right). \tag{18}$$

To fairly evaluate the algorithms concerning all cases, we use two qualitative metrics to measure the performance: (i) the success-ratio (SR/%) over the $\mathcal{M}_{\text{tot}}$ cases:

$$\text{SR} = \frac{\mathcal{M}_{\text{suc}}}{\mathcal{M}_{\text{tot}}} \times 100\% \tag{19}$$

where $\mathcal{M}_{\text{suc}}$ denotes the number of cases that succeed in localizing the sound source as its $\text{MAEE}^m \leq 10°$. (ii) the average MAEE over the $\mathcal{M}_{\text{suc}}$ successful cases,

$$\overline{\text{MAEE}} = \frac{1}{M_{\text{suc}}}\left(\sum_{q=1}^{M_{\text{suc}}} \text{MAEE}^q\right) \tag{20}$$

where $q = 1, \ldots, \mathcal{M}_{\text{suc}}$ denotes the index of the successful cases.

4.2. Simulated environments

We utilize an available toolbox [30] to generate the room impulse response (RIR) from the sound source to an open-sphere spherical microphone array (32 channels and radius 4.2 cm) in a $6 \times 4 \times 3$ m rectangular room. The multichannel time-domain recordings are produced by convolving the RIR with speech signals drawn from the TIMIT database (with signals re-sampled to 8kHz). Then, the recordings are transformed using the STFT with a 32 ms window, 50% overlap, and 1024-point discrete Fourier transforms (DFT). One hundred frequency bins ranging from 300 Hz to 1200 Hz are considered, to match the first-order spherical harmonics calculation.

We first evaluate all DOA estimators in various noisy environments. The speech recordings are contaminated by additive Gaussian white noise, with SNR levels in the set $\{5, 15, 25\}$ dB. Table 1 depicts the localization accuracy. It can be verified that although the 'Intensity' method achieves satisfactory results, it is clearly outperformed by the RHC-based methods, especially in the MAEE measure. This advantage of the RHC methods might be attributed to the time-averaging applied in (14) that reduces the noise contribution. Both RHC-based approaches achieve SR $= 100\%$, and the 'Proposed' estimator outperforms the 'Decoupled' estimator in the MAEE measure. The underlying reason is that the proposed estimator does not require a search over the grid space, so that it is able to localize the source irrespective of the grid resolution.

We will now examine the robustness of the proposed method to noisy and reverberant conditions. Table 2 depicts the localization accuracy as a function of reverberation level for $T_{60} = \{150, 350, 550\}$ ms with SNR $= 10$ dB. The results confirm that both RHC-based DOA estimators accurately localize the sound source in typical reverberant environments. Although sound reflections are not considered in their model, both RHC methods are only gradually degrading with increased reverberation level. The 'Proposed' closed-form estimator exhibits slightly higher accuracy than the 'Decoupled' estimator that is matching between the estimated and the theoretical RHCs. Note that the results for intensity-based estimator are not reported in Table 2, since it is severely impacted by the reverberation and noise considered here.

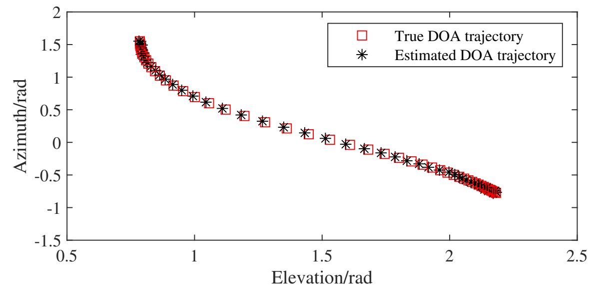

Fig. 1: DOA tracking using the closed-form localization algorithm.

Finally, we tested the run-time of the algorithms to validate our claims for computational efficiency of the closed-form method. The two RHC-based methods were implemented using MATLAB with CPU Intel Core i7-7500U 2.7 GHz, RAM 16 GB. The total runtime to complete the localization results in Table 2 by the 'Proposed' estimator is just 2.9 ms, in contrast to the 578 ms it takes the 'Decoupled' estimator. Motivated by the computational-efficiency of the 'Proposed' estimator, we further address single source tracking problem. The moving source recordings are simulated by adopting the method in the localization and tracking algorithm from [31]. Figure 1 depicts the true and estimated time-varying DOAs, respectively, demonstrating an accurately recovered moving trajectory.

4.3. Real-life environments

Next, we examine the performance of the proposed algorithm in real-life scenarios. The real recordings are measured in the acoustic lab in the Australian National University with dimensions $[6.7, 3.6, 2.8]$ m, and reverberation level approximately $T_{60} = 250$ ms. The practical implementation uses the same parameters as for the simulated environments study. We examined both the 'Proposed' and 'Decoupled' estimators to separately localize ten static speakers, played from loudspeakers at various directions with respect to a 32-channel EigenMike, both achieving a success-ratio of SR $= 100\%$ and average accuracy of $\overline{\text{MAEE}} = 4.91°$. Note that the localization results are inferior to the results obtained in simulations, due to non-negligible errors encountered in practical scenarios, such as deviations in the loudspeaker positions. Again, the proposed method is much more efficient. While it takes 0.015 ms on average for the 'Proposed' method to localize a source, the 'Decoupled' estimator requires 1.137 ms to accomplish the same accuracy.

5. CONCLUSION

This paper proposed a new algorithm with closed-form solution to estimate the source DOA in noisy and reverberant environments using first-order relative harmonic coefficients. The closed-form solution is compared with a grid-search method also utilizing first-order RHC and with the intensity based localization method, achieving better localization accuracy and reduced computational complexity. Extensive experimental study confirmed the effectiveness of the proposed algorithm. In the future, we plan to extend this method to the more challenging multiple speaker localization problem.

6. 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] T. 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.

[3] 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.

[4] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, "Real-time multiple sound source localization and counting using a circular microphone array," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2193–2206, 2013.

[5] L. I. Birnie, T. D. Abhayapala, and P. N. Samarasinghe, "Reflection assisted sound source localization through a harmonic domain music framework," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 279–293, 2020.

[6] C. Evers, A. H. Moore, and P. A. Naylor, "Multiple source localisation in the spherical harmonic domain," in *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 258–262.

[7] S. Chakrabarty and E. A. P. Habets, "Multi-speaker DOA estimation using deep convolutional networks trained with noise signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 8–21, 2019.

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

[9] B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, and E. Fisher, *Spherical microphone array beamforming*, pp. 281–305, Springer, 2010.

[10] L. Kumar and R. M. Hegde, "Near-field acoustic source localization and beamforming in spherical harmonics domain," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3351–3361, 2016.

[11] T. D. Abhayapala and H. Bhatta, "Coherent broadband source localization by modal space processing," *ICT 2003: 10th International Conference on Telecommunications*, pp. 1617–1623, 2003.

[12] 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.

[13] Y. Hu, T. D. Abhayapala, and P. N. Samarasinghe, "Multiple source direction of arrival estimations using relative sound pressure based MUSIC," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 253–264, 2021.

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

[15] 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.

[16] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, 2001.

[17] B. Laufer-Goldshtein, R. Talmon, and S. Gannot, "Semi-supervised source localization on multiple manifolds with distributed microphones," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 7, pp. 1477–1491, 2017.

[18] D. Cherkassky and S. Gannot, "Successive relative transfer function identification using blind oblique projection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 474–486, 2020.

[19] 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*, vol. 28, pp. 3108–3123, 2020.

[20] 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.

[21] Y. Hu, T. D. Abhayapala, P. N. Samarasinghe, and S. Gannot, "Decoupled direction-of-arrival estimations using relative harmonic coefficients," in *2020 28th European Signal Processing Conference (EUSIPCO)*, pp. 246–250.

[22] Y. Hu, P. N. Samarasinghe, S. Gannot, and T. D. Abhayapala, "Evaluation and comparison of three source direction-of-arrival estimators using relative harmonic coefficients," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 815–819.

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

[24] 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.

[25] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 9, no. 6, pp. 697–707, 2001.

[26] 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.

[27] D. Levin, E. A. P. Habets, and S. Gannot, "On the angular error of intensity vector based direction of arrival estimation in reverberant sound fields," *The Journal of the Acoustical Society of America*, vol. 128, no. 4, pp. 1800–1811, 2010.

[28] 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.

[29] 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.

[30] E. A. Habets, "Room impulse response (RIR) generator," 2006. [Online]. Available: https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator.

[31] 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.