# SOUND SOURCE LOCALIZATION USING RELATIVE HARMONIC COEFFICIENTS IN MODAL DOMAIN

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

Audio and Acoustic Signal Processing Group, Australian National University, Canberra, Australia

## ABSTRACT

This paper proposes a data-driven source localization approach under a noisy and reverberant environment, using a newly defined feature named relative harmonic coefficients (RHC) in the modal domain. Being independent of the source signal, the RHC is capable of localizing a sound source(s) located at unknown position(s). Two distinctive multi-view Gaussian process (MVGP), (i) multi-frequency views and (ii) multi-mode views, are developed for Gaussian process regression (GPR) to reveal the mapping function from the RHC to the corresponding source location. We evaluate the effectiveness of the algorithm for single source localization while the underlying concepts proposed can be extended to acoustic scenarios where multiple sources are active. Experimental results, using a spherical microphone array, confirm that the proposed algorithm has a faster speed and achieves competitive performance in comparison to the state-of-art algorithm.

***Index Terms***— Source localization, Relative harmonic coefficients, multi-view Gaussian process, Gaussian process regression.

## 1. INTRODUCTION

In the past few decades, sound source localization [1] has been an active research area in the field of acoustic signal processing, playing a fundamental role in many applications involving teleconferencing systems [2], source separation [3] and automatic speech recognition [4].

Generally, traditional localization methods can be categorized into three main types: (i) algorithms using maximization of the steered response power (SRP) [5], (ii) high-resolution spectral estimations, such as multiple signal classification (MUSIC) [6] and estimation of signal parameters via rotational invariance (ESPRIT) [7], and (iii) dual-stage methods requiring a time difference of arrival (TDOA) estimation. Nevertheless, due to the fact that the traditional algorithms mainly rely on assumptions of acoustic propagation and statistical characteristics, accurate source localization in real-world scenarios is still far to be realized as those theoretical assumptions hardly hold in practice.

By contrast, source localization using data-driven approaches have attracted greater attention recently because of their impressive localization accuracy with presence of background noise and reverberation. Provided abundant training data, those localization algorithms directly learn the mappings from the feature vectors taken as the inputs to the desired source positions, by making use of the prior knowledge from the labeled training samples. For these approaches, the features and their accurate estimations under noisy conditions are vital to the localization performance as they contain unique characteristics of the acoustic system.

With measurements from a single microphone, Talmon et al. [8, 9] proposed a supervised source localization algorithm that employs manifold learning to recover the fundamental controlling parameter of the acoustic impulse response (AIR), which coincides with the source position in a static environment. Using a pair of microphones, Laufer-Goldshtein et al. [10, 11] initially exploits relative transfer function (RTF) for single source localization. Assuming a static acoustic environment, the positions of the sound source are the only varying degree-of-freedom of the system. Therefore, the RTF-based localization algorithms are capable of recovering the unknown source locations. Thereafter, for improved accuracy, the setup of a single microphone pair is extended to that using multiple microphone pairs scattered within the recording area at the cost of the speed of algorithms [12].

Recently, source localization using higher order microphone arrays (e.g. spherical or circular arrays [13, 14, 15, 16, 17]) has been popular due to the effectiveness of such arrays. Therefore, given a higher order microphone array, this paper aims to address the issue to accurately localize a sound source under a noisy and reverberant environment.

Motivated by the proposals of RTF between two individual microphones [11, 12], this paper proposes a novel feature vector, named as Relative Harmonic Coefficients (RHC), which is appropriate for higher order microphone arrays. The RHC are defined and computed using the spherical harmonic coefficients over the microphone array in modal domain [18, 19]. Assuming a static acoustic environment, the RHC can be used as an effective feature vector to localize the sound sources because the only degree of freedom of RHC coincides with the position of the desired source. To reveal the mapping function, this paper develops two distinctive multi-view Gaussian process (MVGP) used for Gaussian process regression (GPR) [12, 20], i.e., multi-frequency views and multi-mode views, where each single view represents one exact practice to localize the sound source while the multi-views fuse the individuals efficiently for higher accuracy.

## 2. PROBLEM FORMULATION

In this section, we formulate the problem to be addressed. Typically, any arbitrary soundfield at a point $\boldsymbol{x} = (r, \theta, \phi)$ within a spatial region can be represented using spherical harmonic coefficients in modal domain [18],

$$P(\boldsymbol{x}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k)\, j_n(kr)\, Y_{nm}(\theta, \phi) \qquad (1)$$

in which $k = 2\pi f/c$ is the wavenumber, $N = \lceil kr \rceil$ indicates the order of soundfield [21], $j_n(\cdot)$ stands for spherical Bessel function

*Thanks to Australian Research Council Linkage Grant Projects funding scheme (project no. LP160100379). Yonggang Hu is sponsored by CSC agency for funding.

of the first kind, and $Y_{nm}(\theta, \phi)$ denotes the well-known spherical harmonic functions with order $n$ and degree $m$. The $\alpha_{nm}(k)$ represents the spherical harmonic coefficients, which can be captured using a higher order microphone array (e.g. a spherical microphone array (SMA) [19], multiple circular microphone arrays [22], and a planar microphone array [23]).

Based on the spatial analysis of the soundfield, this paper aims to localize the exact position of a single sound source under a noisy and reverberant environment, by proposing a data-driven localization approach in modal domain. An example to localize an active speaker using a SMA is presented in Fig. 1. To realize it, two vital problems need to be addressed. Firstly, an effective feature vector shall be exploited, which ought to be fit for the higher order microphone array and contain relevant relations or clues to localize the source. Secondly, the mapping function that reveals the underlying relations from the feature vectors to source positions shall be formulated while it would manage to fuse the muti-channel recordings effectively.

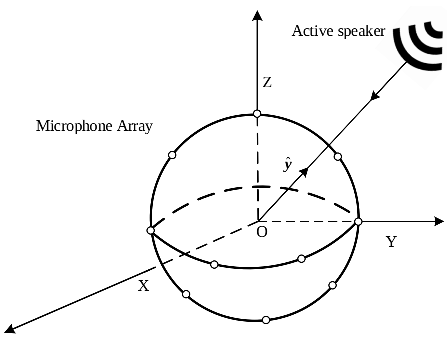


Figure 1: Source localization using a spherical microphone array.

## 3. RELATIVE HARMONIC COEFFICIENTS (RHC)

This section attempts to define a new kind of feature vector named as RHC in modal domain, which fits a higher order microphone array. Thereafter, an effective RHC estimator that is applicable under noisy conditions is proposed.

### 3.1. Definition of RHC

Given spherical harmonic coefficients over all individual modes $[\alpha_{00}(k), \cdots, \alpha_{NN}(k)]$ in (1), we formally define the RHC as the ratio between spherical harmonic coefficient $\alpha_{nm}(k)$ and $\alpha_{00}(k)$,

$$\boldsymbol{\beta}(k) = \left[1, \frac{\alpha_{1,-1}(k)}{\alpha_{00}(k)}, \cdots, \frac{\alpha_{nm}(k)}{\alpha_{00}(k)}, \cdots, \frac{\alpha_{NN}(k)}{\alpha_{00}(k)}\right]. \quad (2)$$

The ratio between the harmonic coefficients contributes by removing the effects of the input or driving signal.

Assume the omni-directional point source to be localized is located at $(r_s, \theta_s, \phi_s)$ in polar coordinates. Then, the closed form expression of the spherical harmonic coefficients $\alpha_{nm}(k)$ in free field takes form as,

$$\alpha_{nm}(k) = s(k) i k h_n(kr_s) Y_{nm}(\theta_s, \phi_s) \quad (3)$$

in which the $s(k)$ is the driving signal of the point source, $i$ denotes the imaginary number and $h_n(kr_s)$ represents the spherical Hankel function. Its relative harmonic coefficients $\beta_{nm}(k)$ are:

$$\beta_{nm}(k) = h_n(kr_s)Y_{nm}(\theta_s, \phi_s)/(h_0(kr_s)Y_{00}(\theta_s, \phi_s)). \quad (4)$$

Note that the only degree of freedom in $\beta_{nm}(k)$ is the desired source location $(r_s, \theta_s, \phi_s)$.

### 3.2. Estimator of RHC

In this section, we propose a method to estimate the RHC under noisy conditions. Given the RHC denoted by $\beta_{nm}(k)$, let us define a spatial function at $\boldsymbol{x}_j = (r, \theta_j, \phi_j)$ as,

$$\begin{aligned} Q(\boldsymbol{x}_j, k) &= \sum_{n=0}^{N} \sum_{m=-n}^{n} \beta_{nm}(k)\, j_n(kr)\, Y_{nm}(\theta_j, \phi_j) \\ &= \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{\alpha_{nm}(k)}{\alpha_{00}(k)}\, j_n(kr)\, Y_{nm}(\theta_j, \phi_j) \\ &= \rho \frac{\sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k)\, j_n(kr)\, Y_{nm}(\theta_j, \phi_j)}{\sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{nm}(k)\, j_n(0)\, Y_{nm}(0,0)} \\ &= \rho \frac{P(x_j, k)}{P(x_o, k)} = \rho \frac{s(k)A(x_j, k)}{s(k)A(x_o, k)} = \rho \frac{A(x_j, k)}{A(x_o, k)} \end{aligned} \quad (5)$$

where $\rho = Y_{00}(0, 0)$ is a fixed constant and neglected for convenience, $s(k)$ denotes the driving signal, $A(x_j, f)$ and $A(x_o, f)$ represents the acoustic transfer function (ATF) from the sound source to the $j$-th microphone and the one located at origin of the array, respectively. From above simplification, it is apparent that $Q(\boldsymbol{x}_j, k)$ coincides with the Relative Transfer Function (RTF) between the point microphone at $\boldsymbol{x}_j = (r, \theta_j, \phi_j)$ and $\boldsymbol{x}_o = (0, 0, 0)$. Therefore, if one has enough measurements of $Q(\boldsymbol{x}_j, k)$, the RHC $\beta_{nm}(k)$ can be calculated using spherical harmonic theory. In practice, using a higher order microphone array, such as a SMA with $M$ microphones, we can estimate pressure at origin, i.e. $P(x_o, k)$, by the addition of all the recordings. Afterwards, the ratio between each microphone recording $P(x_j, k)$ where $j = 1, \cdots, M$ and that at the origin $P(x_o, k)$ can give a total of $M$ number of RTFs.

However, under real scenarios, all recordings are contaminated by noise. In this case, the recordings can be represented by,

$$\begin{aligned} P(x_j, f) &= S(f)A(x_j, f) + V_j(f) \\ P(x_o, f) &= S(f)A(x_o, f) + V_o(f) \end{aligned} \quad (6)$$

in which $S(f)$ represents the driving signal of the sound source, $V(x_j, f)$ denotes the interfering noise recorded by microphone, and $f$ is the frequency index. In order to relieve the negative effects due to the noise when calculating the RTF, it is common approach in [11, 12] to exploit a biased estimator of RTF using the power spectral density (PSD) and CPSD (cross PSD) of the measured signals,

$$Q(x_j, f) = \frac{S_{p_j p_o}(f)}{S_{p_o p_o}(f) - S_{v_o v_o}(f)} = \frac{S_{ss}(f)A(x_j, f)A^*(x_o, f)}{S_{ss}(f)|A(x_o, f)|^2} \quad (7)$$

where $S_{p_j p_o}(f)$ is the CPSD between $P(x_j, f)$ and $P(x_o, f)$, $S_{p_o p_o}(f)$ is the PSD of $P(x_o, f)$, $S_{v_o v_o}(f)$ is the PSD of the noise in the reference microphone, and $S_{ss}(f)$ is the PSD of the source signal $S(f)$. Using the biased estimator, $Q(x_j, f)$ is estimated by neglecting the noise PSD $S_{v_o v_o}(f)$ in the denominator,

$$Q(x_j, f) \approx \frac{S_{p_j p_o}(f)}{S_{p_o p_o}(f)}. \quad (8)$$

Finally, once the $M$ samples of $Q(x_j, f)$ are available using the above procedures, the $\beta_{nm}(k)$ can be estimated by,

$$\beta_{nm}(k) = \frac{1}{j_n(kr)} \sum_{j=1}^{M} a_j\, Q(x_j, k)\, Y_{nm}^*(\theta_j, \phi_j) \quad (9)$$

where $Y_{nm}^*(\cdot)$ is the complex conjugate operator of $Y_{nm}(\cdot)$ and the formula holds provided that $j_n(\cdot) \neq 0$. The $a_i$ works as the weight of each microphone for the equality of the right and left side in (9).

## 4. LOCALIZATION USING MVGP BASED GPR

In this section, we develop a mapping algorithm to find the desired or testing source location given the estimated feature vector RHC. Firstly, we model variable of source position using multi-view Gaussian process (MVGP). Secondly, MVGP based Gaussian process regression (GPR) is introduced to predicate the testing source position. Finally, two distinctive ways to form MVGP are proposed.

### 4.1. Multi-view Gaussian Process (MVGP)

Since two distinctive MVGP are to be discussed, the current subsection introduces the MVGP in a general manner. As indicated in section 1, each view represents one exact practice to localize the sound source. Hence, we adopt and fuse $V$ views in total for the sake of improved accuracy. Considering an individual $v$-th view, we assume the variable of source position $p^v(\mathbf{h})$ from any source follows a Gaussian process,

$$p^v \sim \mathcal{GP}(0, \hat{k}) \tag{10}$$

where $\mathbf{h}$ is the feature vector due to the source, and notation of $\hat{k}$ represents the kernel or covariance function that specifies the defined Gaussian process. Through a series of testing, we select the squared exponential (SE) covariance function due to its effective performance, which is defined using an exponential function that takes feature vector of two sound sources as its inputs,

$$\hat{k}(\mathbf{h}_i, \mathbf{h}_j) = \sigma_x^2 \exp\left(-\frac{\|\mathbf{h}_i - \mathbf{h}_j\|^2}{2\sigma_y^2}\right) \tag{11}$$

where $\sigma_x$ works as a scale factor for the exponential function and $\sigma_y$ denotes a characteristic length-scale hyper-parameter.

The MVGP fuses all the $V$ views by modelling source position $p$ as the mean of the Gaussian processes between all individual views, i.e., the position $p_l$ for the $l$-th source is the average of the estimations at each view,

$$p_l = \frac{1}{V}\left(p_l^{(1)} + p_l^{(2)} + \cdots + p_l^{(V)}\right). \tag{12}$$

Due to the assumption that the processes are jointly Gaussian, the $p_l$ follows a zero-mean Gaussian process as well, whose covariance between two variables of source positions is defined as,

$$\begin{aligned}\operatorname{cov}(p_r, p_s) &= \frac{1}{V^2}\operatorname{cov}\left(\sum_{z=1}^{V} p_r^{(z)}, \sum_{w=1}^{V} p_s^{(w)}\right)\\ &= \frac{1}{V^2}\sum_{z,w=1}^{V} \hat{k}_z(\mathbf{h}_r^{(z)}, \mathbf{h}_s^{(z)})\hat{k}_w(\mathbf{h}_r^{(w)}, \mathbf{h}_s^{(w)})\end{aligned} \tag{13}$$

in which the $\hat{k}(\cdot)$ is the kernel function defined in (11). The subscript $r$ and $s$ represents the index of the sound sources, while the $z$ and $w$ stands for the index of views.

### 4.2. Gaussian Process Regression Using MVGP

On the basis of the MVGP derived above, localization of the unknown positions of the testing sources can be reviewed as a Bayesian analysis of a regression model, i.e., a Gaussian process regression,

$$p(\mathbf{H}_l) = f(\mathbf{H}_l) + \varepsilon_l, \;\; l = 1, \cdots, L \tag{14}$$

where $l$ is the index of sources, matrix of $\mathbf{H}_l$ refers to the feature vectors of all the views, function $f(\cdot)$ represents the latent function that relates the feature matrix $\mathbf{H}_l$ to the source position $f(\mathbf{H}_l)$, and symbol $\varepsilon_l$ stands for the zero mean Gaussian noise, $\varepsilon_l \sim \mathcal{N}(0, \sigma^2)$. Note that $\varepsilon_l$ denotes calibration inaccuracies that mainly originate from imprecise measurements of source positions and microphone noise.

For simplicity, we ignore the intermediate procedures (refer to details in [24]) and directly provide the closed-form probability distribution for the testing source conditioned on the training set, i.e., $Pr(f^\star|\mathbf{H}^\star, \mathbf{H})$, which follows Gaussian distribution as well.

$$\mathcal{N}\left(\mathbf{K}^*(\mathbf{K}+\sigma^2\mathbf{I})^{-1}\mathbf{p}, \mathbf{K}^{**} - \mathbf{K}^*(\mathbf{K}+\sigma^2\mathbf{I})^{-1}\mathbf{K}^{*T}\right) \tag{15}$$

where $\mathbf{K}^*$ is the matrix that contains covariance of any two vectors between the training feature set $\mathbf{H} = [\mathbf{H}_1, \cdots, \mathbf{H}_L]$ and testing feature set $\mathbf{H}^*$, and symbols of $\mathbf{K}$ and $\mathbf{K}^{**}$ denote the covariance matrix with respect to the $\mathbf{H}\times\mathbf{H}$ and $\mathbf{H}^*\times\mathbf{H}^*$, respectively. The $\sigma^2$ is the noise variance defined in (14), and the $\mathbf{I}$ is the Identity matrix. Following the maximum a posteriori probability (MAP) estimator, the unknown positions of the testing sources shall coincide with the mean value of the Gaussian distribution as the $Pr(f^\star|\mathbf{H}^\star, \mathbf{H})$ reaches its global maximum,

$$p^* = \mathbf{K}^*(\mathbf{K}+\sigma^2\mathbf{I})^{-1}\mathbf{p}. \tag{16}$$

Marginal likelihood conditioned on the training recordings and hyperparameters is to be used to update the unknown hyperparameters, i.e., parameters of the kernel function and noise covariance [24].

### 4.3. Multi-frequency Views and Multi-mode Views

Given a wide frequency band, the RHC, due to $l$-th sound source, belong to a two dimension matrix, $\mathbf{H}_l \in C^{(N+1)^2\times F}$, where $F$ denotes the number of frequency bins and $N$ is the harmonic order of soundfield over the band of frequency. It implies that there has two manners to generate the feature vector, i.e., the vector at a specific frequency $[\beta_{00}(k_f), \cdots, \beta_{NN}(k_f)]$ defined in (2) and the vector at a specific mode $[\beta_{nm}(k_1), \cdots, \beta_{nm}(k_F)]$. For comprehensive analysis, this paper adopts both kinds of vectors, and name the two distinctive MVGP using the two diverse feature vectors as multi-frequency views and multi-mode views, respectively.

Theoretically, we can localize the source position using just one specific frequency bin. However, for the wide-band case, the multi-frequency views contribute to improving the accuracy as it fuses more features over various frequency bins efficiently, avoiding certain drawbacks due to single frequency. For example, source signal at that frequency bin may be contaminated or distorted seriously. Similarly, due to the directivity patterns of spherical harmonic function, one specific mode view over multiple frequency bins may not characterize the sound sources effectively, especially when the desired source originates from a direction that is on the null of the directivity patterns of that mode. Hence, the multi-mode views contribute to overcome such drawbacks by using all the modes so that more directions over space can be covered.

## 5. EXPERIMENTS

This section presents experimental results for single source localization under noisy and reverberant conditions, using simulated recordings under various SNR (Signal-to-noise ratio) and reverberation levels. The experiment setup consists of 3 stages. Firstly, we use a SMA to record the soundfield due to both the training and testing sources. Secondly, the raw recordings are used to derive the feature vectors of RHC using the proposed estimator. Finally, the proposed localization algorithms are used to localize the testing sources propagating from unknown positions.

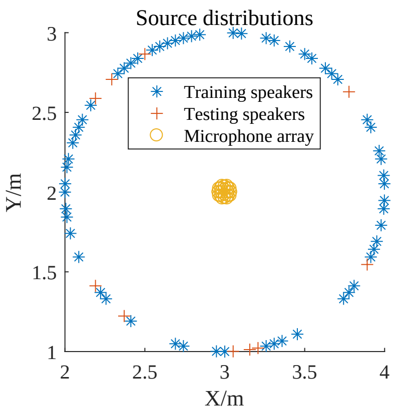


Figure 2: Top view of the setup for simulations.

As shown in Fig. 2, the simulation setup in a $6 \times 4 \times 3$ m room comprises of a 32-channel EigenMike (radius is 0.042 m) and a circular source area (radius is 1 m). The circular source area is irregular sampled, which has 60 training sources and 10 testing sources whose positions are unique from each other. An efficient implementation [25] of the image source method [26] is employed to generate time-domain room impulse respond from each source to the EigenMike. Each source uses a unique speech signal lasting 3 seconds, drawn from the public TIMIT database and is re-sampled to 8 KHz. Ten frequency bins approximately ranging from 3 KHz to 4 KHz are used, which manages to record the sound field up to 3-th order (due to $N = \lceil kr \rceil$). Source components at lower frequencies are abandoned because the estimations of RHC would suffer with more errors due to small amplitudes of Bessel functions [27, 28].

Another recently proposed MVGP based source localization approach [12], i.e., multi-node view, is adopted as a baseline for comparisons. We adjust and estimate the RTFs between microphones on the surface of EigenMike and the virtual microphone at the origin. We mainly localize the $X$ coordinate of the sources as localization of other coordinates just follows a similar repetition. To evaluate the accuracy, average distance between the true and estimated coordinate is calculated as $\sum_{i=1}^{T} |x_{\text{true}}(i) - x_{\text{est}}(i)|/T$, where $T$ is the number of testings sources. Finally, note that the initial values of hyperparameters has a certain impact on the accuracy due to the non-convexity of marginal likelihood. To tackle this issue, we use an approach of cross-validation (CV) [24] that requires to randomly assign multiple groups of initial hyperparameters and finally select a certain groups of the ones that provide satisfied accuracy.

Localization errors using the proposed algorithms under various reverberation levels are presented in Fig. 3. Note that the recordings are contaminated by white Gaussian noise (WGN) at the SNR level of 15 dB. It is observed that a higher reverberation level exerts direct negative influence on the accuracy. The increased reverberation level implies improved complexity of the acoustic environment within the room, so that the performance degrades. Performance of the proposed algorithms are evaluated under various noisy conditions as well, where the SNR level ranges from 5 dB to 25 dB and the reverberation level is around 300 ms. As reported in Fig. 4, we recognize the degraded performance when the SNR level increases.

Results of source localization under diverse acoustic environments in both Fig. 3 and Fig. 4 confirm the superiority of the algorithm using multi-frequency views over the others, especially at low SNR levels. On the other hand, algorithm using multi-mode views has more errors because of varied feature vector of RHC at

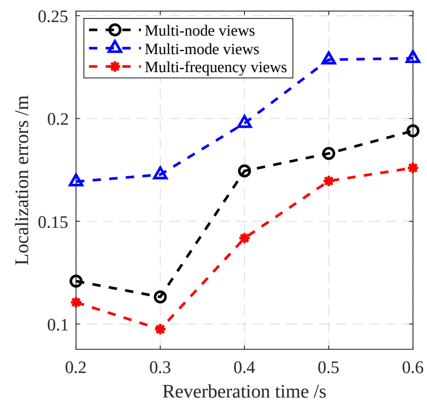


Figure 3: Localization errors under various reverberation levels.

each individual mode, which makes it harder to update and select proper hyperparameters under the same conditions.

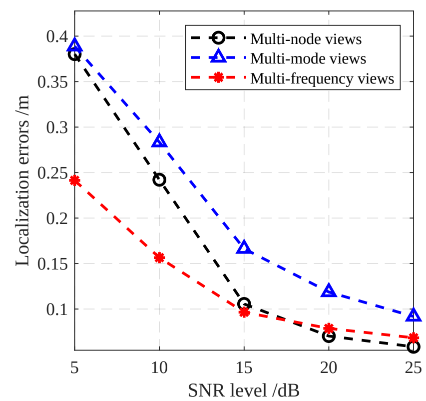


Figure 4: Localization errors under various SNR levels.

Finally, we analyze the computational complexity of the proposed data-driven algorithms, by measuring the average time of the algorithms using a Matlab implementation on a standard PC (CPU Intel Core i7-7500U 2.7 GHz, RAM 16 GB). As shown in Table 1, the speed of the proposed algorithms are faster than that of the baseline because both the proposed algorithms have a relatively smaller number of views, which implies that there remains less groups of hyperparameters to be updated.

Table 1: Speed of all the algorithms.

| Types | Number of views | Average time |
|---|---|---|
| Multi-node | 32 | 133.2 s |
| Multi-mode | 16 | 55.5 s |
| Multi-frequency | 10 | 42.5 s |

## 6. CONCLUSION

This paper has proposed a feature named as relative harmonic coefficients, defined in modal domain and successfully applied it in single source localization under diverse acoustic environments. Through simulated experiments, we have presented a comprehensive analysis over the strengths and weaknesses of the proposed algorithms for both multi-frequency views and multi-mode views. In the near future, the authors will explore applications of the feature in more challenging tasks, e.g. multiple source localization, and provide more in-depth analysis under real scenarios.

## 7. REFERENCES


[1] J. C. Chen, Y. Kung, and R. E. Hudson, "Source localization and beamforming," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 30–39, 2002.

[2] Y. Hu, J. Benesty, and G. W. Elko, "Passive acoustic source localization for video camera steering," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, vol. 2. IEEE, Conference Proceedings, pp. II909–II912 vol. 2.

[3] K. Itakura, Y. Bando, E. Nakamura, K. Itoyama, K. Yoshii, and T. Kawahara, "Bayesian multichannel audio source separation based on integrated source and spatial models," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 4, pp. 831–846, 2018.

[4] 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*, Conference Proceedings.

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

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

[7] 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.

[8] R. Talmon, I. Cohen, and S. Gannot, "Supervised source localization using diffusion kernels," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*. IEEE, 2011, pp. 245–248.

[9] R. Talmon, D. Kushnir, R. R. Coifman, I. Cohen, and S. Gannot, "Parametrization of linear systems using diffusion kernels," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1159–1173, 2012.

[10] B. Laufer-Goldshtein, R. Talmon, and S. Gannot, "Relative transfer function modeling for supervised source localization," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2013 IEEE Workshop on*. IEEE, 2013, pp. 1–4.

[11] ——, "Semi-supervised sound source localization based on manifold regularization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 8, pp. 1393–1407, 2016.

[12] ——, "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.

[13] T. D. Abhayapala and H. Bhatta, "Coherent broadband source localization by modal space processing," in *10th International Conference on Telecommunications, 2003. ICT 2003.*, vol. 2. IEEE, Conference Proceedings, pp. 1617–1623.

[14] B. Jo and J. Choi, "Direction of arrival estimation using nonsingular spherical esprit," *The Journal of the Acoustical Society of America*, vol. 143, no. 3, pp. EL181–EL187, 2018.

[15] O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, no. 10, pp. 1494–1505, 2014.

[16] 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.

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

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

[19] 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, Proceedings*, pp. 1949–1952, 2002.

[20] B. Laufer-Goldshtein, R. Talmon, I. Cohen, and S. Gannot, "Multi-view source localization based on power ratios," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 71–75.

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

[22] 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.

[23] 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.

[24] C. E. Rasmussen and C. K. Williams, *Gaussian process for machine learning*. MIT press, 2006.

[25] E. A. Habets, "Room impulse response generator," *Technische Universiteit Eindhoven, Tech. Rep*, vol. 2, no. 2.4, p. 1, 2006.

[26] 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.

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

[28] P. 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.