A multi-source localization method based on clustering and outlier removal

Shang Gao* Maoshen Jia† Changchun Bao~
*†~Faculty of Information Technology, Beijing University of Technology, Beijing China
*E-mail: gaoshang9795@163.com Tel: +86-10-67391642
†E-mail: jiamaoshen@bjut.edu.cn Tel: +86-10-67391642
~E-mail: chchbao@bjut.edu.cn Tel: +86-10-67391642

***Abstract*— Multiple sound source localization is a hot topic of concern in recent years. In this paper, a multi-source localization method based on weight clustering and outlier removal is proposed to deal with the multiple source localization in the environment with high reverberation time. In this kind of environments, there are always some T-F points consisting of components from multiple sources mixed in the detected spares components. These T-F points, which are called outliers, usually carry the wrong information of localization and could lead to the decline of localization accuracy. To solve this problem, the Point Offset Residual Weight (PORW) and Source Offset Residual Weight (SORW) are introduced to measure the contribution of each T-F point to the localization. The binary clustering is proposed to distinguish and remove the outliers. After that, a statistical histogram of DOA estimation is drawn using the composite weight to weaken the effect of components that interfere with the localization. Finally, the multi-source localization is conducted through peak searching. The objective evaluation of the proposed method is conducted in various simulated environments. The results show that the proposed method achieves a better performance compared with the reference methods in sources localization.**

Keywords--multiple sources localization, direction of arrival estimation, reverberation, sparsity, sound field microphone

I. INTRODUCTION

The task of multiple sound source localization aims to estimate the Direction of Arrival (DOA) of the sound sources without knowing the information about the recording environment and sound sources. The accurate DOA estimations play an important role in sound field analysis and the corresponding technology remains an active research subject with applications in a wide variety of fields, which include hearing aids [1], intelligent transportation [2], robotics [3]-[4] human-machine interaction [5]-[7], and so on.

It should be noted that in the actual recording environments, the presences of reverberation and noise make the DOA estimation of multiple and simultaneously sound sources a great challenge. Nowadays, the research on this problem has been well established, and the corresponding multi-source localization methods have formed their systems according to their different characteristics. Some notable examples like the Time Difference Of Arrival (TDOA) based methods [8][9], the subspace-based methods [10]-[11], the Direct Path Dominance (DPD) test based methods [12], and the Sparse Components Analysis (SCA) based methods [13]-[23].

Among the methods mentioned above, the SCA-based methods are favored for their outstanding localization performance in underdetermined conditions (i.e., the number of microphones is smaller than the number of sound sources). It should be noted that the traditional SCA-based methods rely on the W-Disjoint Orthogonally (W-DO) property [24], which means the representations of sound sources in the Time-Frequency (T-F) domain do not overlap. Each T-F component is consisting of direct-path signals from only one source. By transforming the recorded signals from the time domain to the T-F domain, the multi-source DOA estimation problem can be transformed to a single-source DOA estimation and the number of microphones required for localization is declined significantly. However, when there are more sources sound simultaneously, the W-DO assumption is hardly satisfied in the whole T-F plane. An extension of the W-DO assumption is proposed in Ref. [14]. Under this extended assumption, the whole T-F plane is divided into lots of tiny "time-frequency analysis zones (i.e., T-F zone, which is consists of a group of joint T-F points)." When the components of multiple sound sources are overlapped in T-F plane, there should be few T-F zones where only one source is dominant, these T-F zones are called "Single Source Zone" (SSZ). It has been proved that high localization accuracy can be obtained by making use of the T-F points within the detected SSZs. Moreover, Ref. [15] proposed a peak searching approach to jointly estimate the number of sources and their DOA, which has higher efficiency and accuracy of DOA estimation. Notably, the SCA-based methods include not only "zone-level" sparse components detection methods, but also "point-level" sparse components detection methods. These methods assume that there should always be some T-F points in the T-F plane whose components only consist of the direct-path signals of a single source. These T-F points are called the "Single Source Point" (SSP) and these methods are called the SSP-based methods [17][22]. Compared with SSZ-based methods, SSP-based methods focus more on the components of each T-F point instead of T-F zone, therefore, the T-F points with accurate clues of source localization are more likely to be selected.

Even though the performance of these SCA-based methods looks appealing, when the recording environment gets complex (includes but not limited to the situation with more sources sound simultaneously, higher reverberation time, and/or background noise level), the localization accuracy might heavily degrade.

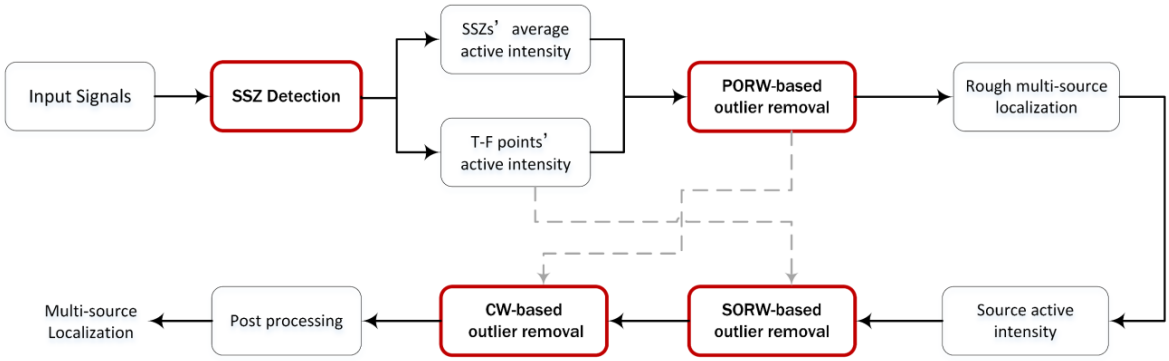

Fig. 1 The system block diagram of proposed method

One of the main reasons leads to this problem is that there are usually outliers which carry wrong localization clues mixed in the selected components of both two types of methods. In complex environments, the increasing number of outliers could directly affect the results of localization.

Aiming to settle this problem, this paper proposed a multi-source localization method based on weight clustering and outlier removal. By introducing Point Offset Residual Weight (PORW) and Source Offset Residual Weight (SORW) [23], the contribution of each T-F point in the direction of the actual source can be evaluated. Besides, instead of using the empirical threshold, an adaptive threshold, which can be derived through binary clustering, is applied through the whole method to achieve an accurate multiple source localization in complex environments.

The reminder of the paper is organized as follows: the proposed multi-source location method is introduced in Section 2. Next, the proposed method is evaluated through various experiments in Section 3. Finally, the conclusion is achieved in Section 4.

II. Proposed Localization Method

In this paper, a multi-source localization method based on weight clustering and outlier removal is proposed to deal with the localization accuracy decline problem caused by the outliers mixed in the detected sparse components. A soundfield microphone, which is consists of four closely placed microphone capsules, is used to record the speech signals. The system block diagram is shown in Fig. 1, the blocks marked in red indicate that the binary clustering is introduced in this step to obtain the adaptive thresholds. The whole method can be divided into four parts: Firstly, the signals recorded by soundfield microphone are transformed into T-F domain and split into T-F zones. The SSZs are selected from all the T-F zones. Secondly, the active intensity vectors of both SSZs and the T-F points within them are calculated. In order to measure the contribution of each T-F point to its corresponding SSZ and remove a part of outliers, PORWs are calculated for every T-F points. Then, a weighted histogram, which is used to conduct rough source localization, is drawn using PORW. Thirdly, the source active intensity is calculated based on the rough source localization. After that, the source active intensity is combined with the point active intensity to obtain the SORW of each T-F point. And the outliers are removed by clustering the SORW. Finally, the Composite Weight (CW) is obtained by combining the PORW and SORW, and the outliers are removed based on the clustering of CW. The accurate DOA estimation of sources can be achieved through the histogram drawn by CW with outliers removed. More details of these processes are described below:

A. *Modeling and SSZ detection*

In this paper, the signals from four channels of a soundfield microphone are chosen as the input of the system. Considering a situation that N sources sound simultaneously in an environment with reverberation and noise, the observed mixture can be modeled as:

$$x_i(n,k) = \sum_{p=1}^{N} h_{i,p}(k) \cdot s_p(n,k) + \dot{r}_i(n,k) + v_i(n,k) \quad (1)$$

where $i \in \{1,2,3,4\}$ represents the index of the four soundfield microphone channels, $h_{i,p}(k)$ is the transfer function from p-th source to the i-th microphone capsule. Ignoring the time delay during propagation, $s_p(n,k)$ is the T-F representation of the p-th source in n-th frame and k-th frequency point, and $x_i(n,k)$ is the signals received by i-th microphone channel. $\dot{r}_i(n,k)$ and $v_i(n,k)$ denote the reverberation components and noise components mixed in the signals of channel i, respectively.

Based on the co-located characteristic of four channels of soundfield microphone, the SSZ can be selected through the Normalized Cross-Correlation (NCC) between the signals from different channels [14], the formula of NCC is given as below:

$$r_{ij}(Z) = \frac{R_{ij}(Z)}{\sqrt{R_{ii}(Z) \cdot R_{jj}(Z)}} \quad (2)$$

where $i, j \in \{1,2,3,4\}\ (i \neq j)$ denote the indexes the microphone channels, Z denotes the a T-F zone, which is consist of a series of adjacent T-F points. $r_{ij}(Z)$ is the NCC calculated using the signals from channel i and j. $R_{ij}(Z)$ is cross-correlation coefficient between channel i and j, which is defined as follow:

$$R_{ij}(Z) = \sum_{(n,k)\in Z} \left| x_i(n,k) \cdot x_j(n,k) \right| \tag{3}$$

From Ref. [14], it can be found that if the signals from a T-F zone are only consist of the direct components from a single source, then the signals between channels should have similar waveforms, which means that the following formula should be satisfied:

$$r_{ij}(Z) > 1 - \varepsilon \tag{4}$$

where ε is threshold to select the SSZ. In the traditional methods, the threshold is set by users according to the recording environment. However, the information about recording environment is not always available and the empirical threshold is not suitable for the changing condition. Therefore, in this paper, the SSZ is selected by binary clustering the NCC and taking the cluster with higher values.

B. PORW-based outlier removal

It should be mentioned that due to the feature of SSZ-based methods, only the characteristic of the whole T-F zone, rather than each point within it, is considered in these methods. That leads to the inevitable mixing of outliers which consist of complex components and carry the inaccurate localization information in the selected SSZ. Therefore, the PORW is proposed to distinguish the outliers by measuring the contribution of each T-F point to their corresponding SSZ.

After selecting of SSZ, every T-F point within SSZ should be transformed into B-format [15] for further processing. The four channels of B-format signals are represented as $\{x_w, x_x, x_y, x_z\}$, where w represents the omnidirectional channel, x, y, z represent the are three Cartesian bi-directional channels. Then, the activity intensity for T-F point (n,k) is defined as:

$$\begin{cases} I_x(n,k) = \dfrac{\sqrt{2}}{\rho c}\left[Re\{x_w^*(n,k) \cdot x_x(n,k)\}\right] \\ I_y(n,k) = \dfrac{\sqrt{2}}{\rho c}\left[Re\{x_w^*(n,k) \cdot x_y(n,k)\}\right] \\ I_z(n,k) = \dfrac{\sqrt{2}}{\rho c}\left[Re\{x_w^*(n,k) \cdot x_z(n,k)\}\right] \end{cases} \tag{5}$$

where c is the velocity of sound, ρ is the density of the medium, $Re\{\cdot\}$ represents taking the real part of the signals, $[\cdot]^*$ denotes taking conjugation. The activity intensity vector can be formed as $\boldsymbol{I}(n,k) = [I_y(n,k), I_x(n,k)]$. The activity intensity vector for the SSZ can be calculated as:

$$\bar{\boldsymbol{I}}(Z) = \sum_{(n,k)\in Z'} \frac{\boldsymbol{I}(n,k)}{K \cdot \|\boldsymbol{I}(n,k)\|} \tag{6}$$

where Z' is a SSZ whose size is K.

Then, the localization information for the T-F points can be obtained using the activity intensity. For simplicity, the formula below only calculates the azimuth of T-F point (n,k):

$$\hat{\mu}(n,k) = \tan^{-1}\left(\frac{I_y(n,k)}{I_x(n,k)}\right) \tag{7}$$

In the following, the PORW is expounded based on the two characteristics below:

Characteristic 1: The outliers should be the minority in the SSZ and their directional information is randomly distributed.

Characteristic 2: The directional characteristic of the whole SSZ should have offset relative to the actual source due to the outliers, while it shouldn't be far away from it.

Based on these characteristics, the PORW is defined as:

$$W_p(n,k) = 1 - \cos^{-1}\left(\frac{\langle \bar{\boldsymbol{I}}(Z) \cdot \boldsymbol{I}(n,k) \rangle}{\|\bar{\boldsymbol{I}}(Z)\| \cdot \|\boldsymbol{I}(n,k)\|}\right) \cdot \frac{1}{\pi} \tag{8}$$

where $\|\cdot\|$ represents the Euclidean norm, and $W_p(n,k)$ is the PORW given to the T-F point $(n,k) \in Z$.

Since the directional information of the whole SSZ is much accurate than that of outlier, all the outliers within the SSZ should be given a lower value of PORWs. Therefore, the outliers can be filtered by binary clustering the PORWs and removing the cluster with lower PORWs.

The following process is performed based on the PORWs and the azimuth estimations. Firstly, the statistical source component equalization [16] is conducted to avoid the overwhelm effect between components from different sources. Then, the weighted histogram can be drawn for rough DOA estimation. Finally, the histogram is smoothed to remove the burrs and the peak searching is conducted to estimate the DOA of sources. For simplicity, the q-th estimated DOA is represented as $\hat{\mu}_q$.

C. SORW-based outlier removal

Although the PORW can distinguish a part of outliers, the difference between outliers and desired points is not obvious enough to filter out most of the outliers. These outliers can form pseudo-peaks in the DOA estimation histogram and lead to mistakes in source counting. Since the outliers are consist of multiple and complex components, they usually locate randomly in the histogram. That means the peaks formed by outliers usually locates differently in the histogram before and after the smoothing. Based on this characteristic, the SORW is then proposed to remove the outliers that failed to be removed by PORW. Before giving the definition of SORW, the source active intensity vector should be calculated:

$$\boldsymbol{I}_q = \boldsymbol{I}(n'_q, k'_q) \tag{9}$$

where $\boldsymbol{I}_q$ is the source active intensity vector of q-th estimated DOA, (n'_q, k'_q) is the T-F point whose azimuth estimation is closest to the estimated DOA and can be found by:

$$(n'_q, k'_q) = \arg\min_{(n,k)}(|\hat{\mu}_q - \hat{\mu}(n,k)|) \tag{10}$$

Then, the SORW is defined combining the active intensity vectors of sources and every T-F points:

$$W_s(n,k) = 1 - \cos^{-1}\left(\frac{\langle \boldsymbol{I}_q \cdot \boldsymbol{I}(n,k)\rangle}{\|\boldsymbol{I}_q\| \cdot \|\boldsymbol{I}(n,k)\|}\right) \cdot \frac{1}{\pi} \tag{11}$$

where $W_s(n,k)$ is the SORW given to the T-F point (n,k), and q is the index of DOA estimations corresponding to the current T-F point, which should satisfy

$$q = \arg\min_{a}(|\hat{\mu}_a - \hat{\mu}(n,k)|) \tag{12}$$

It can be found that the SORW gives the highest value to the T-F points whose estimated azimuths closest to the estimated DOA, and gives lower values to the T-F points around them. Since the peaks formed by outliers locate differently in the histogram before and after the smoothing, the outliers who form the local maximum in DOA statistical histogram can hardly be given the highest SORW. Therefore, the outliers can be distinguished through the binary clustering of the SORWs.

D. CW-based outlier removal and post-processing

In section II. B, and C, two kinds of weights are proposed to distinguish the outliers from the different ways. In this section, these weights are combined as a composite weight (CW) to jointly filter out the outliers, which can be represented as:

$$W_c(n,k) = W_p(n,k) \cdot W_s(n,k) \tag{13}$$

Although most of the outliers can be removed through PORW and SORW, the remaining points are not absolutely consisting of components of a single source. Therefore, each T-F point should be weighted according to its contribution to the localization. The histogram that used to perform localization is drawn using the composite weight:

$$Y(\mu) = \{\lceil \hat{\mu}(n,k) \rceil = \mu | \sum W_c(n,k)\} \tag{14}$$

where $\mu \in [1,360]$ represents the angle in the histogram, $\lceil \cdot \rceil$ represents round down to take an integer, and $Y(\mu)$ is the value of vertical coordinate at μ in the histogram. It should be mentioned that binary clustering of $Y(\mu)$ is still needed to be conducted. The peaks located at the angle that are not likely to be the azimuth of actual sources are removed. Finally, the post processing [16] including smoothing and peak searching is conducted to finish the multiple source localization.

III. Experimental Evaluation

In this section, the effectiveness of the proposed weight clustering and outlier removal based multi-source localization method has been verified via objective evaluation in simulated environments. The simulation room, which is a cube with a length of 6 meters, a width of 3 meters, and a height of 2 meters, is created using the ROOMSIM package. A soundfield microphone is set in the center of the room. All the sources are set around the microphone with 1-meter distance. The speech signals from the Chinese sub-database of NTT database have been chosen as the sound sources for analysis. Since a similar process is conducted for the estimation of azimuth and elevation except for the signals from different channels of the soundfield microphone are chosen as the input signals, the experimental evaluation only chooses azimuth to analyze for simplicity. The reference methods are selected from the proposed method without clustering, Statistical Source Component Equalization (SSCE) based method [16] and the SSP-based algorithm [22]. The experiments can be mainly divided into two parts according to the different settings of the room, the results and analysis are as follows:

A. The evaluation of the proposed method in different reverberation time

In this sub-section, three sources with a separation of 60° are set in the rooms with the reverberation time of {150ms, 300ms, 450ms}. One hundred simulation situations, including different sources, different positions of sources, are conducted for analysis. The average error of each sources' DOA estimation is calculated to evaluate the localization accuracy of each method. The results of the experiments are shown in Figure 2.

It can be found in the Figure that with the increase of reverberation time, the average error of all the methods increases. When the reverberation time is 150ms, only a few outliers are mixed in the detected sparse components. The average error of the proposed methods is only 2.8 and the average errors of reference methods are nearly 10. In this situation, the outliers can hardly form the peaks (i.e., pseudo-peaks) that can be confused with the real peaks. The difference in localization accuracy mainly lies in the effect of outliers on the true peaks. While when the reverberation time rises to 300ms, the number of outliers within the sparse components detected by reference methods increases, and the pseudo-peaks are formed in the histogram, which makes the localization accuracy decline significantly. Different from the reference methods, PORW and SORW can effectively distinguish the outliers and desired points, by weighting each T-F point according to their contribution to the localization, the influence of outliers can be weakened. The clustering of weight could provide guidance to the removal of outliers, which makes the effect of outliers on localization as low as possible. Therefore, the proposed method using weight clustering could always obtain the lowest average error in multi-source localization.

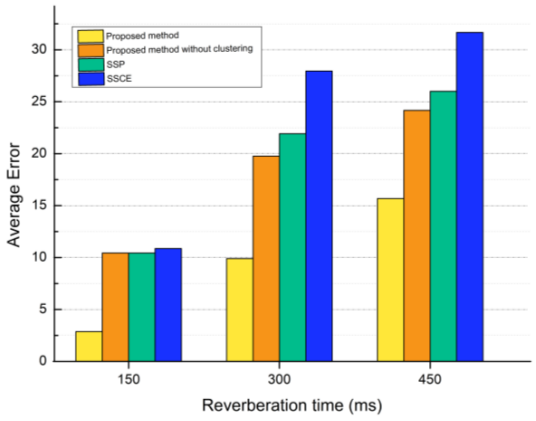

Fig. 2 Objective comparison on the errors among sources in different reverberation times

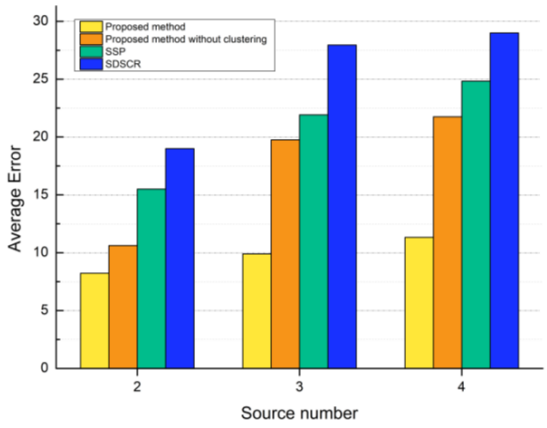

Fig. 3 Objective comparison on the errors among sources with different source number

B. The evaluation of the proposed method with different number of sources

In this subsection, the reverberation time of the room is set as 300ms, the numbers of sources are chosen from {2,3,4}, and the separation between sources is 60°. The results of one hundred simulation situations are shown in Figure 3.

Similar to the result in the last subsection, the obvious trends of the increasing number of sources lead to the increasing average error, and the proposed method using weight clustering can usually obtain the lowest average error. Besides, it can be found that as for the change of active sources' number, the outlier removal with clustering is more robust than the outlier removal with the empirical threshold. The reason is that for the method using an empirical threshold, as the number of simultaneous sources increases, the threshold with a constant value could increase the range that is considered as true peaks in the histogram, which introduces more outliers. To sum up, among all the situations, the proposed method can always achieve a better performance than reference methods, which implies the advantage of the proposed method.

IV. Conclusion

In this paper, the unavoidable problem of outliers' existence in the sparse components detected by SCA-based methods is addressed. In order to distinguish the outliers, PORW and SORW are proposed to measure the contribution of every T-F point to the localization. The binary clustering is conducted to remove the outliers, which can be applied in various blind source localization scenarios. Finally, the composite weight is calculated to weigh the T-F points in the histogram according to their contribution and weaken the effect of components that interfere with the localization. The proposed method has been proved to achieve better performance over experimental environments compared with the reference methods. Besides, the proposed method can be integrated into other localization frameworks making use of DOA histograms plotted by SCA-based methods.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grants (No. 61971015, 61831019) and the Cooperative Research Project of BJUT-NTUT (No. NTUT-BJUT-110-05).

References

[1] T. V. Bogaert, E. Carette, and J.Wouters, "Sound source localization using hearing aids with microphones placed behind-the-ear, in-the-canal, and in-the-pinna," Int. J. Audiol., vol. 50, no. 3, pp. 164–176, 2011.

[2] Zheng X, Ritz C, Xi J (2016) Encoding and communicating navigable speech soundfields. Multimedia Tools Appl. 75(9):5183–5204

[3] J. M. Valin, F. Michaud, B. Hadjou, and J. Rouat, "Localization of simultaneous moving sound sources for mobile robot using a frequency-domain steered beamformer approach," in Proc. IEEE Int. Conf. Robot. Automat., 2004, vol. 1., pp. 1033–1038.

[4] M. Togami, Y. Obuchi, and A. Amano, Automatic Speech Recognition of Human-Symbiotic Robot EMIEW. I-tech Education and Publishing, 2007, ch. 22, pp. 395–404.

[5] T. Latif, E. Whitmire, T. Novak, and A. Bozkurt, "Sound localization sensors for search and rescue biobots," IEEE Sensors J., vol. 16, no. 10, pp. 3444–3453, May 2016

[6] M. Togami, A. Amano, T. Sumiyoshi, and Y. Obuchi, "DOA estimation method based on sparseness of speech sources for human symbiotic robots," in ICASSP, 2009, pp. 3693–3696.

[7] Nikunj Raghuvanshi,John Snyder. Parametric wave field coding for precomputed sound propagation[J]. ACM Transactions on Graphics (TOG),2014,33(4).

[8] J. Chen, J. Benesty and Y. Huang, "Time delay estimation in room acoustic environments: An overview", EURASIP J. Adv. Signal Process., vol. 2006, Dec. 2006.

[9] Nesta F, Omologo M. Generalized state coherence transform for multidimensional TDOA estimation of multiple sources. IEEE Trans Audio Speech Lang Process 2012;20(1):246–60

[10] D. Ying, R. Zhou, J. Li and Y. Yan, "Window-Dominant Signal Subspace Methods for Multiple Short-Term Speech Source

Localization," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 4, pp. 731-744, April 2017, doi: 10.1109/TASLP.2016.2625458.
[11] S. Hafezi, A. H. Moore and P. A. Naylor, "Multiple DOA estimation based on estimation consistency and spherical harmonic multiple signal classification," 2017 25th European Signal Processing Conference (EUSIPCO), Kos, 2017, pp. 1240-1244, doi: 10.23919/EUSIPCO.2017.8081406.
[12] Rafaely B, Alhaiany K. Speaker localization using direct path dominance test based on sound field directivity[J]. Signal Processing, 2018, 143:42-47.
[13] Pavlidi D , Puigt M , Griffin A , et al. Real-time multiple sound source localization using a circular microphone array based on single-source confidence measures[J]. Acoustics, Speech, and Signal Processing, 1988. ICASSP-88. 1988 International Conference on, 2012.
[14] Pavlidi D, Griffin A, Puigt M, Mouchtaris A (2013) Real-time multiple sound source localization and counting using a circular microphone array. IEEE Trans Audio Speech Lang Process 21(10):2193–2206.
[15] Jia M , Sun J , Bao C . Real-time multiple sound source localization and counting using a soundfield microphone[J]. Journal of Ambient Intelligence & Humanized Computing, 2017, 8(6):829-844.
[16] Gao S, Jia M, Wu Y, et al. Multiple Sound Sources Localization by using Statistical Source Component Equalization[C]// ICCPR '19: 2019 8th International Conference on Computing and Pattern Recognition. 2019.
[17] A. Aissa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani and Y. Grenier, "Underdetermined Blind Separation of Nondisjoint Sources in the Time-Frequency Domain," in IEEE Transactions on Signal Processing, vol. 55, no. 3, pp. 897-907, March 2007, doi: 10.1109/TSP.2006.888877.
[18] M. Togami, T. Sumiyoshi, and A. Amano, “Stepwise phase difference restoration method for sound source localization using multiple microphone pairs,” in ICASSP 2007, vol. 1, April 2007, pp. I–117–I–120.
[19] S. Araki, H. Sawada, R. Mukai, and S. Makino, “DOA estimation for multiple sparse sources with normalized observation vector clustering,” in ICASSP 2006, vol. 5, May 2006.
[20] M. Togami and R. Scheibler, “Sparseness-aware DOA estimation with majorization minimization,” in Proc. Interspeech, Aug. 2020.
[21] Haijian Zhang, Guang Hua, Lei Yu, Yunlong Cai, and Guoan Bi, “Underdetermined blind separation of over-lapped speech mixtures in time-frequency domain with estimated number of sources,” Speech Communication, vol. 89, pp. 1 – 16, 2017.
[22] K. Wu, V. G. Reju and A. W. H. Khong, "Multisource DOA Estimation in a Reverberant Environment Using a Single Acoustic Vector Sensor," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 10, pp. 1848-1859, Oct. 2018, doi: 10.1109/TASLP.2018.2845121.
[23] Maoshen Jia, Shang Gao, Changchun Bao. Multi-source localization by using offset residual weight. EURASIP Journal on Audio, Speech and Music Processing. 2021, 23(2021):1-18.
[24] Yilmaz, O., & Rickard, S. (2004). Blind separation of speech mixtures via time-frequency masking. IEEE Transactions on signal processing, 52(7), 1830-1847.