Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **1**, January **2018**

Multiple Source Localization by Using Improved Single Source Bins Detection

Jundai Sun, Maoshen Jia*, Changchun Bao

Beijing Key Laboratory of Computational Intelligence and Intelligent System
Faculty of Information Technology
Beijing University of Technology, Beijing, China
sunjundai@emails.bjut.edu.cn; jiamaoshen@bjut.edu.cn

Received August, 2017; revised October, 2017

Abstract. *In our previous work, the outline of a novel multiple source localization scheme based on single source bins(SSBs) detection is presented. By only retaining the Direction of Arrival (DOA) estimates among single source bins (SSBs). The detection of SSBs is completed based on the DOA convergence assumption by using a clustering algorithm. However, in the selection of clustering algorithm, the distribution characteristics of DOA estimates are ignored. This makes detection results may contain some pseudo single source points caused by the algorithm itself. To solve this issue, an improved SSBs detection method is proposed by applying k-means++ algorithm. For the post processing of DOA estimation, an intra-cluster distance weighting algorithm is proposed to improve the accuracy of the localization. Experimental results demonstrate the localization accuracy of the proposed method outperforms existing single source zone or SSBs detection based localization approaches.*

Keywords: Multiple source localization; Single source bin; B-format microphone

1. Introduction.

Everyday listening scenarios are with multiple concurrent sound sources. Multiple sound source localization aims to estimate the direction of arrival (DOA) of each source, which is important to many applications such as hearing aid design, robust automatic speech recognition (ASR) and reconstruction of the sound scene [1–4].

Many researches has been conducted in the field of DOA estimation and classical methods include: time difference of arrival (TDOA) [5], multiple signal classification (MUSIC) [6], and independent components analysis (ICA) [7]. However, most of these methods either require excessive microphones or previous knowledge about the number of actual sources to improve the reliability. Considering the sparse property of the speech signals in the time-frequency (T-F) domain, many W-disjoint Orthogonality (W-DO) based localization approaches draw more attention due to its accuracy and computational efficiency. But the W-DO of speech signal on which these methods are relying is less accuracy with the number of actual sources raises, i.e., the multiple sources usually overlap in T-F domain and DOA estimates are no longer correspond to a true source.

To solve the localization problem of multiple sources, a "single source" zone (SSZ) detecting method has been proposed based on the assumption that there always exist some SSZs where there are only one active source [8]. By applying the SSZ detecting method, the multiple source localization problem is converted to a single one. Later, [9] has verified that the SSZ assumption is reasonable, and the phenomenon is much more evident among the soundfield microphone recordings. However, the T-F bins in the detected SSZ is not

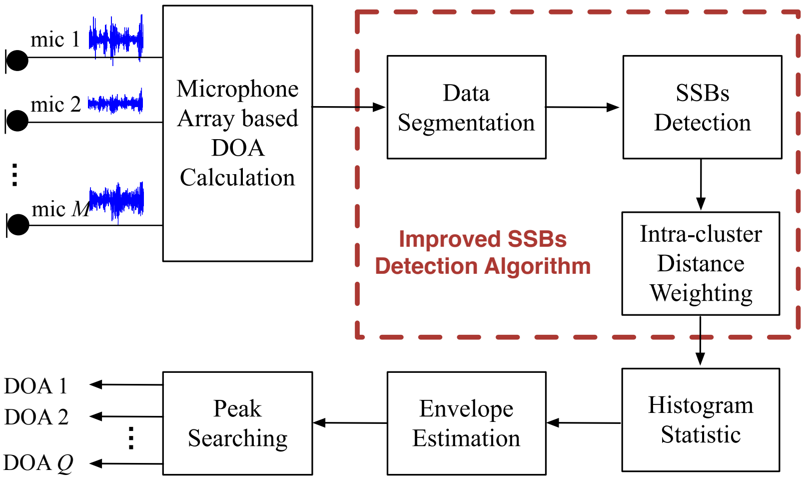

FIGURE 1. The block diagram for the proposed method.

all strictly derived from one source, there are still a few overlapped T-F bins which will affect the accuracy of localization badly especially when there are more sources than microphones. To solve this issue, a single source bins (SSBs) detection based multiple source localization scheme is proposed [10]. This scheme is based on detecting the SSBs in mixture signals that are only derived from one source. Specifically, after proposing a DOA convergence assumption, K-means clustering algorithm is used for SSBs detecting. However, in the selection of clustering algorithm, the distribution characteristics of DOA estimates are ignored. This makes detection results may contain a few pseudo single source points caused by the algorithm itself which will affect the localization results.

In this paper, aiming to solve this problem, we propose an improved SSBs detection based localization method. Specifically, by applying a k-means++ algorithm, the iterations is reduced which makes the method more computationally efficient. The "pseudo" SSBs is eliminated by selecting a closest datapoint to replace the calculated centroid. Besides, for the post-processing of the SSBs detection based methods, we propose an intra-cluster distance weighting algorithm, which makes the peaks of DOA histogram more evident for peak searching.

The remainder of the paper is organized as follows: Section II introduces the proposed localization method. Experimental results are presented in Section III, while conclusions are drawn in Section IV.

2. **Proposed Method.** Aiming to improve the robustness of the SSBs detection based multiple source localization methods, two aspects are taken into account. One is the clustering algorithm itself and the other is the corresponding post processing of these methods. The system block diagram of the proposed localization scheme is shown in Fig. 1. It can be seen that the proposed method is not relying on any microphone array and DOA estimates can be calculated according to specific arrays or algorithms. Some frames of the DOA estimates will be firstly proceeded by a data segmentation and some small zones of the DOA estimates will be obtained. Considering the phenomenon that the effect of K-means clustering algorithm may be affected by some noisy data, a k-means++ algorithm will be applied for SSBs detection in this paper. In addition, all the DOA estimates within SSBs will be processed by a intra-cluster distance weighting algorithm and then used for the histogram statistic. Final DOA estimation will be obtained by some other post-processings including envelop estimation and peak searching. More details will be described below.

2.1. **Improved SSBs detection algorithm.** Based on research of [10], the SSBs detection can be achieved by applying a clustering algorithm. The clustering method chosen in [10] is k-means algorithm, it always begins with an arbitrary set of cluster centers, which may significantly affect the computational efficiency of the SSBs detection method. To improve the efficiency of SSBs detection, a k-means++ algorithm is proposed in this section. On the one hand, a specific way of choosing these centers is applied, i.e., the distance between the initial cluster centroid should be as far as possible. On the other hand, the centroid is no more the mean value of the points in current cluster, but the value of data which is the most closest to the mean value.

Specifically, the core idea of the proposed centroid initialized algorithm is first to randomly select an data point as the first initial cluster center, and the other centers should be as far away from the existing cluster centers as possible. In detail, let $D(x)$ denote the shortest distance from a data point x to the closest center we have already chosen. Suppose the first center is c_1 which is randomly selected from $\boldsymbol{X}$, the next center c_i is selected in $\boldsymbol{X}$ based on a distance weighted probability, i.e.,

$$p(c_i = x_i) = \frac{D(x_i)^2}{\sum_{x\in X} D(x)^2} \tag{1}$$

Then, during each iteration, next centroid is calculated by the mean value of current cluster in k-means algorithm, while it is replaced by the nearest data point, i.e.,

$$c_i^{'} = \arg \underset{\boldsymbol{X}}{min} \, |x_i - c_i| \tag{2}$$

where $c_i^{'}$ denotes the updated centroid of c_i. The upper algorithm is mainly to solve the iterations and improve the computational efficiency, while the latter improvement aims to eliminate the influence of noise or singular values. The improved SSBs detection procedure is conducted as **Algorithm1**.

Algorithm 1 Improved SSBs detection

Input: Sample set, i.e., the DOA estimates in $\mathcal{Z}_p$: $\boldsymbol{X} = \{\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_{L(Z)}\}$; number of clusters: $k = 2$.
Initialize Choose an initial center c_1 uniformly at random from $\boldsymbol{X}$.
Repeat
Select the next center $c_i \in \boldsymbol{X}$ with the probability calculated by (1)
Until we have chosen a total of 2 centers.
Repeat
  Set $C_i = \varnothing (i = 1, 2)$;
  for m=1, 2, ..., $L_{(\mathcal{Z})}$ **do**
    Calculate the distance between $\hat{\mu}_m$ and each centroid $\bar{\mu}_i (i = 1, 2)$: $d_{mi} = ||\hat{\mu}_j - \bar{\mu}_i||_2$;
    Confirm the cluster flag of $\hat{\mu}_j$ based the closest centroid: $\lambda_m = arg\ min_{i\in 1,2} d_{mi}$;
    Put $\hat{\mu}_m$ into corresponding cluster: $C_{\lambda_m} = C_{\lambda_m} \cup \hat{\mu}_m$
  end for
  for $i = 1, 2$ **do**
    Calculate the new centroids:$\bar{\mu}_i^{'} = \frac{1}{|C_i|} \sum_{x\in C_i} x$
    if $\bar{\mu}_i^{'} \neq \bar{\mu}_i$ **then**
      Update the centoid $\bar{\mu}_i$ by the closest data point respecting to the value of $\bar{\mu}_i^{'}$
  end for
Until the value of the centoids do not change
Output:$\boldsymbol{C} = \boldsymbol{C}_1, \boldsymbol{C}_2$

2.2. **Improved SSBs detection method.** After conducting cluster algorithm among each T-F zone, the clusters around two centroid through a number of iterations can be obtained which are denoted by $[\bar{\mu}_1, \boldsymbol{C}_1]$ and $[\bar{\mu}_2, \boldsymbol{C}_2]$, respectively. If all the DOA estimates in Ω remains approximate stable which means that all the DOAs in this zone would belong to one cluster, if not, the DOA estimates in Ω would be variable and the obtained two clusters might be with a similar cardinality. Thereafter, a necessary and sufficient condition for a zone Ω where all TF bins are SSBs is

$$min\{||\boldsymbol{C}_1||_0, ||\boldsymbol{C}_2||_0\} = 0 \tag{3}$$

where $||\cdot||_0$ counts the number of non-zero components in its argument. Here, (4) is the ideal case, in practical, there will inevitably exist some singular values. So we search for T-F zone that satisfy the following inequality:

$$\frac{min\{||\boldsymbol{C}_1||_0, ||\boldsymbol{C}_2||_0\}}{max\{||\boldsymbol{C}_1||_0, ||\boldsymbol{C}_2||_0\}} \leqslant \epsilon \tag{4}$$

where ϵ is a sufficient small threshold to deal with the actual situation in which exists a few abnormal DOA estimates caused by residual non-SSBs. That is, if the cardinality of one cluster is much larger than the other one, it means that the DOA estimates in this T-F zone is stable which implies that most of the T-F components correspond to only one source. It is the zone which contains many SSBs that we aim to detect. Thereafter, the set of SSBs can be naturally obtained by retain the much larger cluster over the T-F zone, i.e., $max\{||\boldsymbol{C}_1||_0, ||\boldsymbol{C}_2||_0\}$. Naturally, all non-SSBs which leads to inaccurate DOA estimation are eliminated and the multiple source localization problem is then rewritten as a single source one among the SSBs. By applying the proposed SSBs detection based localization scheme, most of the existing algorithms may achieve an excellent performance for multiple source localization.

2.3. **Intra-cluster distance weighting.** In order to get a more accurate localization of multiple sources, a intra-cluster distance weighting algorithm is proposed in this paper. Specifically, we propose that in the same TF zone, the closer of the point to the cluster centroid, the greater contribution it will make to the statistical histogram. Suppose the lagger cluster which contains SSBs is $\boldsymbol{C}_1$, by using the logistic function, we mapped the distance between the DOA estimate in each cluster and the corresponding centroid to the range of 0-1, i.e.,

$$Weight(\mu_m) = \frac{1}{1 + e^{\alpha(ICD(\mu_m)-\beta)}}, \mu_m \in \boldsymbol{C}_1 \tag{5}$$

where β is a threshold for the contribution equals 0.5, α is a parameter which can be optimized according to experiment results and $ICD(\mu_m)$ is the intra-cluster distance between data point x and the centroid of its corresponding cluster i.e.,

$$ICD(\mu_m) = |\mu_m - \bar{\mu}_i| \tag{6}$$

By processing N frames of the data, and suppose we obtained $\mathcal{K}$ SSBs. Since, for azimuth μ_m, the corresponding intra-cluster distance weighted DOA histogram bin $hist_e^{\mu_m}$ is given by:

$$hist_w(\mu_m) = Weight(\mu_m) \cdot hist(\mu_m) \tag{7}$$

where $hist_w(\mu_m)$ is the weighted histogram processed by the proposed method, $hist(\mu_m)$ is a histogram statistic value at μ_m and can be obtained by summing the $D(\mu_m)$ which indicates if the current time-frequency component has the same DOA estimation as μ_m:

$$D(\mu_m) = \begin{cases} 1, \mu(n,k) = \mu_m \\ 0, otherwise \end{cases} \tag{8}$$

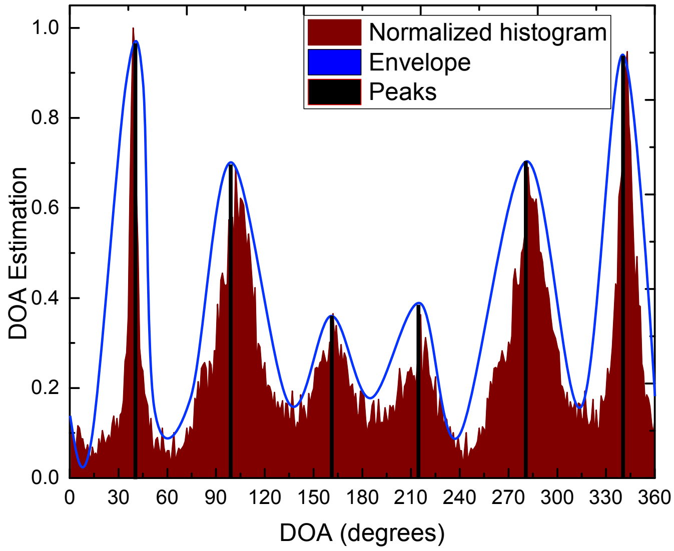

FIGURE 2. Example of DOA estimation with six sources in anechoic room with 95% confidence intervals.

It has to be mentioned that higher accuracy of DOA estimation will be obtained over more frames, we should balance the accuracy and delay according to the actual applications. Moreover, the peak searching process [9] is accurate and computational-efficient so that it can be done in real-time.

An example of the DOA estimation result of six sources at $45°, 105°, 165°, 225°, 285°$, and $345°$ is shown in Fig. 2. Where the soundfield microphone signal was recorded in Anechoic room, the width of "single-source" zone was 64, and 9 (4 look-ahead, 1 current and 4 look-back) frames are used for statistic. It can be seen in Fig. 2, there are six obvious peaks whose corresponding angles are particularly close to the ones of real sound sources (only the peak at $225°$ shift about $8°$), i.e., all peaks representing six sources can be identified. The peak value of sound source depends on its sparsity respecting to the other sources. Specifically, if the time-frequency components of a certain source are isolated with a higher probability, its peak will be more significant; on the contrary, if the time-frequency components are overlapped with the other sources, it is difficult to find a corresponding "single-source" region that will lead a lower peak value. Above all, it can be concluded that the distortion between multiple sound sources will be greatly weakened. As a result, the accuracy of DOA estimation is improved, besides, it means that the source number estimated by this method increases as well.

3. **Experimental results.** In the evaluation, we used Roomsim [11] to simulate a room of $6.25 \times 3.75 \times 2.5$ meters, where $\{6.25, 3.75, 2.5\}$ represented the length, width, height and we took them as the x, y, z axis, respectively. In this paper, we choose the soundfield microphone [9] as the recording array for example, but it should be mentioned that the proposed scheme is not specific to the chosen one. Moreover, the proposed scheme can be applied to any DOA estimation method. The soundfield microphone was placed in the center of the room paralleling with z-axis and the power of sound sources from different directions was equal in each simulation. A total of 36 sentences from the NTT database were selected for testing. Mean absolute estimated error ($MAEE$) [9] was utilized to measure the performance of the proposed method, which calculated the difference between the true DOA and the estimated DOA.

TABLE 1. Experimental parameters

Parameter	Notation	Value
sampling frequency of speech source	fs	16 kHz
source distance	r	1.5m
time-frequency zones width		64
overlapping in frequency		50%
"single-source" threshold	ϵ	0.8
number of bines in the histogram	L_h	720

TABLE 2. Parameters of testing room

Simulated Room	Absorption of the wall	T_{60}(ms)	Reflection order
Anechoic Room	1	0	0
Quiet Room	0.9	0	0
Room1	0.75	250	18
Room2	0.75	450	18
Room3	0.75	600	18

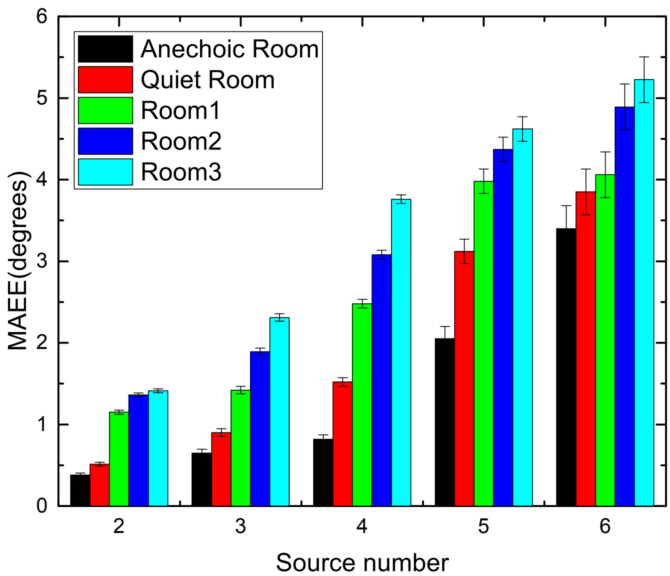

FIGURE 3. $MAEE$ versus separation between adjacent sources in anechoic room with 95% confidence intervals.

To investigate the performance of the proposed localization method in anechoic or reverberant environments, we measured the $MAEE$ [8] of the estimated DOA among different source numbers and separations in five simulated rooms {Anechoic, Quiet room, Room1, Room2, Room3}. In each scenario, we conducted 72 different orientations of each source number and separation. To investigate the localized accuracy, the separation was set as 30°, the source number was $\{2, 3, 4, 5, 6\}$ and the major parameter set is shown in Table I. It must be noted that all the orientations covered the whole 360 and the angle between sources was 30° for all the groups. For example, if the source number was 6, in the first group, the sources located at $0°, 30°, 60°, 90°, 120°, 150°$, and then shifted by

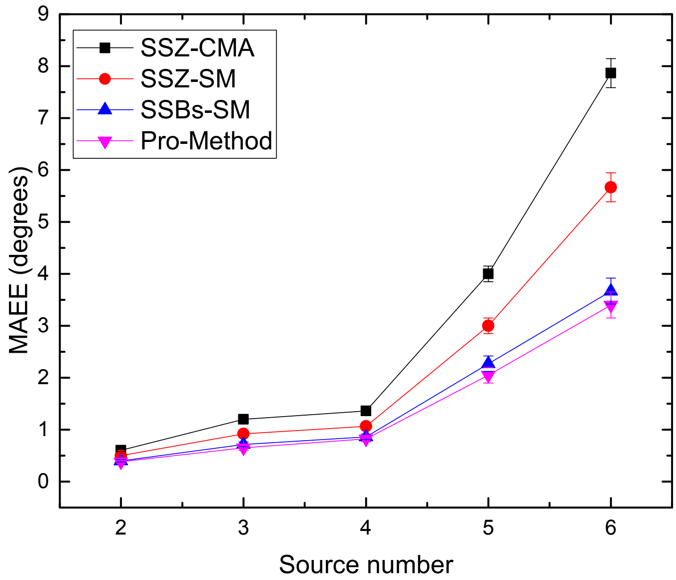

FIGURE 4. The accuracy of proposed source counting method with 95% confidence intervals

10° steps in next group, but remained the same separation. Then the $MAEE$ results are shown in Fig. 3.

It can be concluded that the $MAEE$ rises with the increase of source number but maintains blow 5.5° under all test conditions. When in anechoic room, the $MAEE$ is below 3.5° for all source number.

To compare the localization accuracy of the proposed algorithm with other methods. We chose three of the most effective localization method as the reference method, one is the SSZ detection based localization method by using a circular microphone array (8 microphones) (SSZ-CMA) [8], one is the SSZ detection based method by using a soundfield microphone (SSZ-SM) [9] and another one is the SSBs detection based method by using a soundfield microphone (SSBs-SM) [10]. The comparison result in anechoic is shown in Fig.4, condition Pro-Method refers the proposed localization scheme by using a soundfield microphone. It can be concluded that the proposed method achieves a higher localization accuracy compared with the existing localization approaches based on SSZ detection and SSBs detection. More comparisons in other scenarios will be presented in further work.

4. **Conclusion.** This work presented a novel localization approach for multiple sources based on improved SSBs detection. A k-means++ algorithm was applied to improve the computational efficiency by reducing the iterations of choosing centroid of the cluster. Besides, the influence of noisy point was eliminated by selecting a closest data point to represent the calculated centroid. Then an intra-cluster distance weighting algorithm was proposed as a post processing for the SSBs detection based methods, which made the peak in histogram more evident for peak searching. The evaluations reveal that the proposed method achieves a higher localization accuracy compared with the reference methods. In future work, we will investigate the performance of the proposed method in various scenarios involving more sources with closer DOAs.

Acknowledgment. This work has been supported by China Postdoctoral Science Foundation funded project(2017M610731), the Project supported by Beijing Postdoctoral Research Foundation and "Ri xin" Training Programme Foundation for the Talents by Beijing University of Technology.

REFERENCES

[1] S. Argentieri and P. Danes, Broadband variations of the music high-resolution method for sound source localization in robotics, in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pp. 2009–2014, Oct 2007.

[2] V. D. B. Tim, C. Evelyne, and W. Jan, Sound source localization using hearing aids with microphones placed behind-the-ear, in-the-canal, and in-the-pinna., *International Journal of Audiology*, vol. 50, no. 3, pp. 164–76, 2011.

[3] K. Nakadai, D. Matsuura, H. G. Okuno, and H. Kitano, Applying scattering theory to robot audition system: robust sound source localization and extraction, in *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, vol. 2, pp. 1147–1152 vol.2, Oct 2003.

[4] D. Bechler, M. S. Schlosser, and K. Kroschel, System for robust 3d speaker tracking using microphone array measurements, in *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 3, pp. 2117–2122 vol.3, Sept 2004.

[5] C. Knapp and G. Carter, The generalized correlation method for estimation of time delay, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, pp. 320–327, Aug 1976.

[6] R. Schmidt, Multiple emitter location and signal parameter estimation, *IEEE Transactions on Antennas and Propagation*, vol. 34, pp. 276–280, Mar 1986.

[7] P. Comon, Independent component analysis, a new concept?, *Signal Processing*, vol. 36, no. 3, pp. 287 – 314, 1994.

[8] 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, pp. 2193–2206, Oct 2013.

[9] M. Jia, J. Sun, and C. Bao, Real-time multiple sound source localization and counting using a soundfield microphone, *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, pp. 829–844, Nov 2017.

[10] M. Jia, J. Sun, F. Deng, and J. Sun, Single source bins detection-based localisation scheme for multiple speech sources, *Electronics Letters*, vol. 53, no. 6, pp. 430–432, 2017.

[11] D. Campbell, K. Palomaki, and G. Brown, A matlab simulation of" shoebox" room acoustics for use in research and teaching, *Computing and Information Systems*, vol. 9, no. 3, p. 48, 2005.