A Source Counting Method using Acoustic Vector Sensor based on Sparse Modeling of DOA Histogram

Yang Chen, Wenwu Wang, *Senior Member, IEEE,* Zhe Wang, and Bingyin Xia

***Abstract*—The number of sources present in a mixture is crucial information often assumed to be known or detected by source counting. The exiting methods for source counting in underdetermined blind speech separation (UBSS) suffer from the overlapping between sources with low W-disjoint orthogonality (WDO). To address this issue, we propose to fit the direction of arrival (DOA) histogram with multiple von-Mises density (VM) functions directly and form a sparse recovery problem, where all the source clusters and the sidelobes in the DOA histogram are fitted with VM functions of different spatial parameters. We also developed a formula to perform the source counting taking advantage of the values of the sparse source vector to reduce the influence of sidelobes. Experiments are carried out to evaluate the proposed source counting method and the results show that the proposed method outperforms two well-known baseline methods.**

***Keywords*—source counting, AVS, DOA histogram, OMP.**

I. Introduction

Acoustic vector sensor (AVS) has drawn a lot of interest in recent years due to their ability in measuring the full sound field information benefiting from a co-located sensor structure [1], [2], [3], [4], [5], [6] . It has promising advantages over the conventional microphones and shows good performances on several applications, such as sound localization [7], speech enhancement [8] and separation [9], [10]. For example, the advantage of AVS over a linear microphone array for underdetermined blind speech separation (UBSS) has been shown recently in [11] where the DOA values at time frequency (TF) bins are assumed to follow the von Mises (VM) distribution and the contribution of a specific source is estimated at each TF point in the mixture.

In UBSS, the number of sources present in the mixture is crucial information often assumed to be known or detected by a source counting algorithm [12], [13], [14], [15]. In [12], an algorithm called DEMIX is proposed to count the number of sources, via the estimation of the steering vectors (SV) in the TF domain, and the clustering of the SVs by the basic sequential algorithmic scheme (BSAS), based on the assumption that each TF point is dominated by only one source. In [14], [15], the GMMEM algorithm is presented where the DOA values of the TF bins are represented by the Gaussian mixture model (GMM) with a Dirichelet prior. The parameters of the GMM are estimated by a maximum a posteriori (MAP) approach employing the expectation maximization (EM) algorithm [16]. The source clusters are counted by picking out the corresponding GMM components.

However, both methods suffer from the source overlapping issue caused by low W-disjoint orthogonality (WDO) which is common in practice due to the presence of multiple sources and room reverberations [17]. For DEMIX, the overlapping between sources leads to a low WDO and thus ambiguity in source clustering and false counts. For GMMEM, a low WDO leads to sidelobes between source clusters. This results in a large variance of the parameters of the Gaussian components and thus increased difficulty in distinguishing the components corresponding to the sources from those corresponding to the background.

To address this issue, a new source counting method is proposed where the DOA histogram generated by AVS is modelled using sparse representation with a dictionary matrix containing atoms formed using VM functions with different shape parameters. Sources are assumed to be sparsely distributed in space and represented as a spatial domain sparse vector. The value of each entry in this vector is either 1 or 0 which indicates the presence or absence of the source respectively. The DOA histogram is considered to be a combination of a small number of VM functions with different shape parameters. The sparse vector can be calculated using a sparse recovery algorithm such as Orthogonal Matching Pursuit (OMP) [18]. To our knowledge, formulating source counting as a sparse recovery problem is novel and has not been done in the literature. Furthermore a robust measure for source counting is derived from the sparse vector by taking into account the sidelobes of the source clusters. We also observed empirically that the proposed method is more robust against the violation of the WDO property as compared with baseline methods DEMIX and GMMEM.

The remainder of the letter is organized as follows. In section II, the signal model of the DOA histogram generated by AVS is introduced. Based on this model, our source counting method is proposed in section III. In section IV, the method is evaluated through experiments. The experiments are conducted by comparing the performance of the proposed method with

Y. Chen is with the Department of Information Science and Engineering, Changzhou University, Changzhou 213164, China, (e-mail: chenyang.heu@gmail.com).

W. Wang is with the Department of Electrical and Electronic Engineering, University of Surrey, Surrey GU27XH, United Kingdom, (e-mail: w.wang@surrey.ac.uk).

Z. Wang and B. Xia are with Media Lab, Huawei Technologies Co., Ltd. Beijing 100080, China, (e-mail: zhe.wang@huawei.com; xiabingyin@huawei.com).

the baseline methods in [12], [14].

II. Signal Model of DOA Histogram

Assuming N different speech signals are presented, the observations of AVS omitting the height information can be expressed as [3], [9], [11]

$$\mathbf{y}(t) = \begin{bmatrix} p(t) \\ g_x(t) \\ g_y(t) \end{bmatrix} = \sum_{n=1}^{N} \begin{bmatrix} h_p^n(t) \\ h_x^n(t) \\ h_y^n(t) \end{bmatrix} \otimes r_n(t) \tag{1}$$

where N is the number of sources, t is the discrete time index, and $p(t)$, $g_x(t)$, $g_y(t)$ are the pressure component, the pressure gradient corresponding to the x and y coordinate respectively, which can be obtained directly from the B-format recordings. $r_n(t)$ is the signal radiated from the nth source. $\otimes$ denotes convolution, and $h_p^n(t)$, $h_x^n(t)$, $h_y^n(t)$ represent the corresponding room impulse response (RIR) from the nth source to $p(t)$, $g_x(t)$, $g_y(t)$ respectively cascading the direct path as well as the multipath responses. In anechoic scenario, each source contributes to each channel only through the direct acoustic path without reflections, then we have $h_p^n(t) = 1$, $h_x^n(t) = \cos\mu_n$, $h_y^n(t) = \sin\mu_n$, where μ_n is the direction of the nth source.

Taking the short time Fourier transform (STFT) of each channel of the mixture, the mixing process can be modeled in the TF domain as

$$\mathbf{Y}(\omega, m) = \sum_{n=1}^{N} \mathbf{H}_n(\omega) R_n(\omega, m) \tag{2}$$

where ω and m are the frequency bin and time frame indices respectively. $\mathbf{Y}(\omega, m) = [P(\omega, m), G_x(\omega, m), G_y(\omega, m)]^T$ in which $P(\omega, m)$, $G_x(\omega, m)$, $G_y(\omega, m)$ are the STFT of $p(m)$, $g_x(m)$, $g_y(m)$ respectively. $\mathbf{H}_n(\omega) = [h_p^n(\omega), h_x^n(\omega), h_y^n(\omega)]^T$ is the frequency domain representation of the RIR from the nth source to the corresponding sensor element. T denotes transposition. $R_n(\omega, m)$ is the STFT of $r_n(t)$ which is approximately W-disjoint orthogonal for speech signals.

In the intensity based DOA algorithm [3], the direction of the intensity can be obtained by

$$\theta(\omega, m) = \arctan[\frac{Re\{P^*(\omega, m)G_y(\omega, m)\}}{Re\{P^*(\omega, m)G_x(\omega, m)\}}] \tag{3}$$

where * denotes the complex conjugating. Based on the estimation of $\theta(\omega, m)$ for each discrete bearing on a grid of possible bearings $\theta \in \{\theta_i\}_{i=1}^{\Theta}$, a histogram of all the directions is obtained. The VM function is often utilized to fit the DOA histogram of each source [9], [11]. It is circular within $[0, 360]^\circ$. Therefore, there is no need to deal with the wrapping problem. The DOA histogram of the observations is considered to be a combination of a small number of VM functions with different shape parameters.

$$g(\theta) = \sum_{n=1}^{N} f(\theta \mid \mu_n, \kappa_n) = \sum_{n=1}^{N} \frac{exp(\kappa_n \cos(\theta - \mu_n))}{2\pi I_0(\kappa_n)\rho} \tag{4}$$

where μ_n is the mean direction of the nth source. κ_n is the concentration. The pair (μ_n, κ_n) is here referred to as the shape parameters. $\rho = \int_0^{\pi} \frac{exp(\kappa_n \cos(\theta - \mu_n))}{2\pi I_0(\kappa_n)} \mathrm{d}\theta$ is used to normalize the VM function. $I_0(\kappa_n)$ is the modified Bessel function of order zero.

III. Proposed Source Counting Method

In this work, sources are assumed to be distributed in spatial domain with unknown shape parameters. As shown in Fig. 1, the DOA histogram from the observations $g(\theta)$ is a combination of the VM functions selected from all potential directions $\{\mu_j\}_{j=1}^{\Omega}$ and concentrations $\{\kappa_k\}_{k=1}^{K}$. A spatial domain sparse vector $\mathbf{s} = \{s_i\} \in \mathbb{R}^{\Omega K}$ is used to represent this selection. If there is a source at the direction μ_j with the concentration κ_k, then $s_{i=(j-1)\Omega+k} = 1$, otherwise $s_{i=(j-1)\Omega+k} = 0$. To form the relation between the sparse vector $\mathbf{s}$ and the DOA histogram, a dictionary $\mathbf{A} \in \mathbb{R}^{\Theta \times \Omega K}$ is constructed, whose $((j-1)\Omega + k)$th column is a vector comprising $f(\theta \mid \mu_j, \kappa_k)$ at $\{\theta_i\}_{i=1}^{\Theta}$

$$\mathbf{a}(\mu_j, \kappa_k) = \begin{bmatrix} f(\theta_1 \mid \mu_j, \kappa_k) \\ f(\theta_2 \mid \mu_j, \kappa_k) \\ \vdots \\ f(\theta_\Theta \mid \mu_j, \kappa_k) \end{bmatrix}. \tag{5}$$

So the model is expressed as

$$\mathbf{g} = \mathbf{A}\mathbf{s} \tag{6}$$

where $\mathbf{g}$ is a vector comprising $g(\theta_i)$.

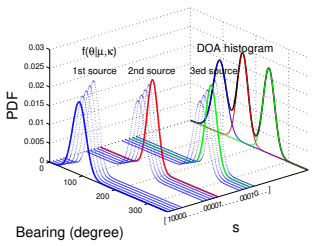


Fig. 1. The sparse representation of the DOA histograms ($\mathbf{g} = \mathbf{A}\mathbf{s}$).

Then with the presence of noise, the DOA histogram derived from the observations of AVS can be expressed as

$$\mathbf{x} = \mathbf{g} + \mathbf{n} = \mathbf{A}\mathbf{s} + \mathbf{n}, \tag{7}$$

where $\mathbf{n}$ is additive white Gaussian noise with zero mean and the variance of σ_n^2 representing errors caused by sidelobes, model mismatches and background noise. This is a typical underdetermined sparse recovery problem

$$\arg\min_{\mathbf{s}} \parallel \mathbf{x} - \mathbf{A}\mathbf{s} \parallel^2 \quad \text{subject to} \quad \parallel \mathbf{s} \parallel_0 \leq L \tag{8}$$

where L controls the sparsity of $\mathbf{s}$. L is unknown but here we assume it is larger than N. This problem can be solved by OMP [18] leading to an estimation of $\mathbf{s}$ denoted by $\hat{\mathbf{s}}$.

Because L is larger than N, the number of nonzero values in $\hat{\mathbf{s}}$ could be larger than the number of sources. We take advantage of amplitudes of the nonzero values in $\hat{\mathbf{s}}$ to carry out source counting. Considering $\hat{\mathbf{s}}$ to be an estimate of $\mathbf{s}$ with

error, the sum of nonzero values in $\hat{\mathbf{s}}$ should be close to N and the maximum nonzero value should be close to 1. Therefore the source number can be estimated directly by

$$C(\hat{\mathbf{s}}) = \lfloor \sum_{i=1}^{\Omega K} \hat{s}_i / \max_i \hat{s}_i \rfloor \tag{9}$$

where $\hat{s}_i$ is the ith element of $\hat{\mathbf{s}}$ and $\lfloor\rfloor$ rounds the argument to its nearest integer. Next, we show theoretically the accuracy of $C(\hat{\mathbf{s}})$. Define L nonzero values in $\hat{\mathbf{s}}$ as $\{1+\epsilon_1, 1+\epsilon_2, \ldots, 1+\epsilon_N, \epsilon_{N+1}, \ldots, \epsilon_L\}$, where $\{\epsilon_i\}_{i=1}^L$ is the estimating error. Define ϵ_m as the maximum of $\{\epsilon_i\}_{i=1}^L$.

$$\begin{aligned} C(\hat{\mathbf{s}}) &= \lfloor \frac{\sum_{i=1}^{N}(1+\epsilon_i) + \sum_{i=N+1}^{L}(\epsilon_i)}{(1+\epsilon_m)} \rfloor \\ &= \lfloor \frac{N - N\epsilon_m + (1-\epsilon_m)\sum_{i=1}^{L}\epsilon_i}{1-\epsilon_m^2} \rfloor \end{aligned} \tag{10}$$

If $|\epsilon_m| < 1$, then $\epsilon_m^2 << 1$. As a result, we have

$$C(\hat{\mathbf{s}}) = \lfloor N + (1-\epsilon_m)\sum_{i=1}^{L}\epsilon_i - N\epsilon_m \rfloor = N \tag{11}$$

when

$$|(1-\epsilon_m)\sum_{i=1}^{L}\epsilon_i - N\epsilon_m| < 1/2. \tag{12}$$

Define $\epsilon_e = \sum_{i=1}^{L}\epsilon_i/N$, then

$$\begin{aligned} |(1-\epsilon_m)\sum_{i=1}^{L}\epsilon_i - N\epsilon_m| &\le N|(1-\epsilon_m)||\epsilon_e| + N|\epsilon_m| \\ &< 4Nmax\{|\epsilon_m|, |\epsilon_e|\}. \end{aligned} \tag{13}$$

So if

$$max\{|\epsilon_m|, |\epsilon_e|\} \le 1/(8N), \tag{14}$$

equation (12) stands. This means the source counting result based on equation (9) will be correct provided that the estimation of $\mathbf{s}$ is accurate enough. It must be emphasized that equation (14) is a sufficient but unnecessary condition of equation (12) and is more restrictive. As a result, the counting formulation is bounded with limited estimating errors of the sparse vector due to the model mismatch.

The proposed algorithm is summarized in Algorithm 1.

Algorithm 1 Source Counting

Input: Sensor output, $p(m)$, $g_x(m)$ and $g_y(m)$
Output: Source number, $C(\hat{\mathbf{s}})$

1: Get $\mathbf{Y}(\omega, m)$ from segments of sensor output by STFT using equation (2);
2: Calculate the DOA of each TF bin $\theta(\omega, m)$ according to equation (3);
3: Derive the DOA histogram $\mathbf{x}$;
4: Estimate the sparse vector $\mathbf{s}$ according to equation (8);
5: Count the sources according to equation (9);
6: **return** $C(\hat{\mathbf{s}})$;

TABLE I. PERCENTAGE (%) OF CORRECT COUNT OF THE NUMBER OF SOURCES

Number of sources	2	3	4	5
The proposed method	91.5	90.0	86.5	82.0
GMMEM	86.5	73.5	65.0	56.5
DEMIX	79.0	52.5	35.0	31.5

IV. EVALUATIONS

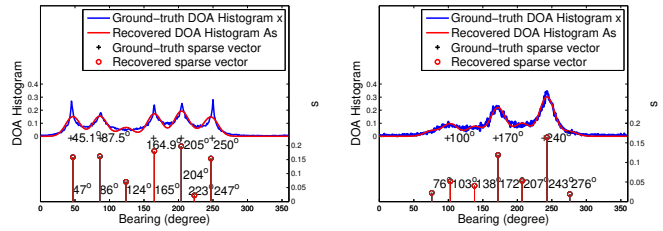
In this section, we conduct experiments comparing the proposed method with baseline methods DEMIX [12] and GMMEM [14] in both anechoic and reverberant environments. Both GMMEM and DEMIX have been briefly described in the first section. DEMIX introduces the local confidence measure to improve the clustering tendency of the SVs estimated in the TF domain. Equation (1) can be considered as an mixture model of N audio sources on 3 channels. Therefore DEMIX can be performed on the observations of an AVS.

For the anechoic environment, N sources are randomly chosen from 26 speech signals of 13 English speakers (males and females) from "TED Talks" to generate anechoic mixtures based on Equation (1). There is no sensor noise. The directions of sources are random but with a space no less than 40 degrees. The sampling frequency is 48kHz and the signal length is 10 seconds. The STFT frame size is 2048. A Hanning window with a half-window overlap is used. The hyper parameter in the Dirichlet distribution is $\phi = 0.9$. In the EM algorithm, we utilized the number of components $M = 7$ in the GMM model, whose initial means are uniformly distributed within $[0, 360]^\circ$, the initial standard deviations are the standard deviation of DOA values and the initial weights are $1/M$. The number of sources is determined by counting the number of Gaussians whose weights are larger than 0.02 and standard deviations are less than 20 as in [14]. In the proposed algorithm, the sparsity setting is $L = M = 7$ and κ is scanned with a grid of $[0:1:30]$. μ and θ are scanned with a grid of $[0:1:359]^\circ$. In DEMIX, the segment size and the thresholds are the same as those in [12]. 200 random tests are run for each trial.

Table I shows the percentage of correct source counts. It can be observed that the proposed method outperforms GMMEM and DEMIX, especially when the number of sources is increased. DEMIX relies on a clear clustering tendency of the estimated SVs. It suffers the most when the WDO level decreases due to the increase in the number of sources. Compared to other two methods, the proposed method is more robust when the number of sources is increased.

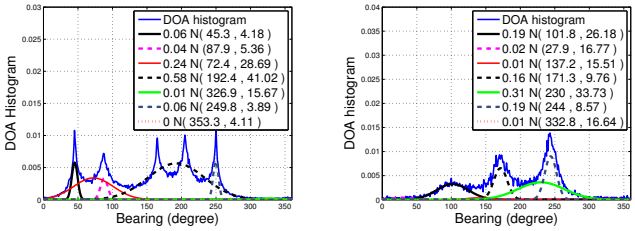
Fig. 2(a) shows an example of the proposed method where $N = 5$. Both the DOA histogram and sparse vector $\mathbf{s}$ can be recovered and the source number can be estimated correctly by equation (9). Despite the estimating error of $\mathbf{s}$ due to the sidelobes and the model mismatch, a correct counting can be achieved. Fig. 3(a) is the outcome of GMMEM with the same mixture. The variance of the fourth Gaussian component (the black dash line) fitting to the source cluster is similar to that of the third Gaussian component (the red solid line) fitting to the sidelobe. This leads to the difficulty in distinguishing different kinds of Gaussian components, and hence an inaccurate counting of the sources. Fig. 4(a) is the scatter

plot of estimated SVs weighted by the confidence measures in DEMIX. The overlapping of the sources makes the clustering tendency unclear and introduces outliers. The BSAS may consider the outliers as new clusters and thus leading to an incorrect clustering result.



(a) anechoic scenario (b) reverberant scenario

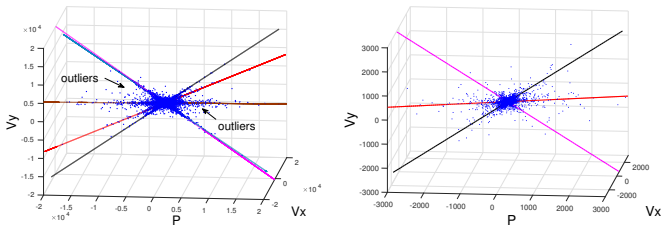
Fig. 2. The ground-truth and recovered DOA histograms and sparse vectors.



(a) anechoic scenario (b) reverberant scenario

Fig. 3. The estimated GMM components in GMMEM. The legend $\omega \mathrm{N}(\mu, \sigma)$ represents a Gaussian component weighted by ω, with the mean μ and the standard deviation σ.

For the reverberant scenario, we use RIRs recorded in an AudioBooth with RT_{60} of about 120ms in University of Surrey, where 4 loudspeakers are mounted at the same height with the directions of 100, 170, 240 and 310 degrees, clumped on a geodesic sphere structure, and a sound-field microphone was placed in the center [19]. In the mixture, N sources are randomly chosen and convolved with N RIRs respectively. No sensor noise is included. The parameters in the algorithms are the same as those for the anechoic mixtures. We run 200 random tests for each trial. The source counting results for this dataset are given in Table II.



(a) anechoic scenario (b) reverberant scenario

Fig. 4. The scatter plots of the estimated SVs weighted by the local confidence measures used in DEMIX. The colored lines show the mixture directions of the sources.

TABLE II. PERCENTAGE (%) OF CORRECT COUNT OF THE NUMBER OF SOURCES

Number of sources	2	3	4
The proposed method	63.0	59.5	50.5
GMMEM	60.5	57.5	48.0
DEMIX	43.5	22.0	13.0

From this table, we can see that the performance of the proposed method and the two baseline methods degrades with the increase in the number of sources. Fig. 2(b) is an example of the proposed method with 3 sources. Due to the presence of room reverberation, the source clusters become more spread as compared to the anechoic scenario. The value in the recovered sparse vector corresponding to the first left source cluster with low height is similar to those corresponding to the sidelobes. In this case, a miscounting is very likely to happen. For GMMEM as in Fig. 3(b), the reverberation causes problems in another way. Although the first Gaussion component (the black solid line) fits exactly to the first left source, the variance is close to that of the fifth Gaussion component (the green solid line) fitting to a sidelobe. This leads to a miscount. DEMIX is not suitable for reverberant environment as it is designed to work 'in an anechoic setting' [12]. Fig. 4(b) is the scatter plot of estimated SVs weighted by the confidence measures for the mixture in Fig. 2(b). Due to the presence of reverberation, the overlapping between the sources is increased, as a result, there may be no cluster tendency that is crucial for the success of DEMIX. This confirms the performance of DEMIX in Table II. In reverberant environment the performance of all the three methods degrades. However, the performance of the proposed method is significantly better than DEMIX and slightly better than GMMEM.

It should be noted that the von Mises distribution is a close approximation to the wrapped normal distribution, therefore, using the von Mises mixture model with a Dirichlet prior in the GMMEM algorithm is likely to give similar results to the use of GMM with the Dirichlet prior, however, a detailed experimental comparison between these two models is out of the scope of this work.

V. CONCLUSION

In this work, we modeled the DOA histogram as a sparse representation where the dictionary matrix contains atoms formed using VM functions with different shape parameters. A new formulation for source counting was then proposed based on the sparse source vector estimated using a sparse recovery algorithm. The proposed method has been evaluated with experiments and compared with two baseline methods. The results have shown that the proposed method gave a better performance in both anechoic and reverberant environments as compared with the baseline methods DEMIX and GMMEM.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under project 61501061, State Key Laboratory of Ocean Engineering (Shanghai Jiao Tong University) under project 1316 and Huawei Technology Ltd.

REFERENCES

[1] K. Ferrar, "Soundfield microphone: design and development of microphone and control unit," *Wireless World*, vol. 85, pp. 48–50, 1979.

[2] M. Shujau, C. H. Ritz, and I. S. Burnett, "Designing acoustic vector sensors for localisation of sound sources in air," in *European Signal Processing Conference*, no. Eusipco, 2009, pp. 849–853.

[3] M. Hawkes and A. Nehorai, "Acoustic vector-sensor processing in the presence of a reflecting boundary," *IEEE Transactions on Signal Processing*, vol. 48, no. 11, pp. 2981–2993, 2000.

[4] M. Hawkes, "Acoustic vector-sensor beamforming and capon direction estimation," *IEEE Transactions on Signal Processing*, vol. 46, no. 9, pp. 2291–2304, 1998.

[5] X. Zhong and A. B. Premkumar, "Particle filtering approaches for multiple acoustic source detection and 2-D direction of arrival estimation using a single acoustic vector sensor," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4719–4733, 2012.

[6] M. Shujau, C. H. Ritz, and I. S. Burnett, "Speech enhancement via separation of sources from co-located microphone recordings," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 137–140.

[7] M. Shujau and C. H. Ritz, "Using in-air acoustic vector sensors for tracking moving speakers," in *IEEE International Conference on Signal Processing and Communication Systems*, 2010, pp. 1–5.

[8] M. Shujau, "In air acoustic vector sensors for capturing and processing of speech signals," Ph.D. dissertation, University of Wollongong, 2011.

[9] B. Günel, H. Hachabibolu, and A. M. Kondoz, "Acoustic source separation of convolutive mixtures based on intensity vector statistics," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 4, pp. 748–756, 2008.

[10] M. Shujau, C. H. Ritz, and I. S. Burnett, "Separation of speech sources using an acoustic vector sensor," in *IEEE International Workshop on Multimedia Signal Processing*, 2011.

[11] X. Chen, W. Wang, Y. Wang, X. Zhong, and A. Alinaghi, "Reverberant speech separation with probabilistic time-frequency masking for B-format recordings," *Speech Communication*, vol. 68, pp. 41–54, 2015.

[12] S. Arberet, R. Gribonval, and F. Bimbot, "A robust method to count and locate audio sources in a multichannel underdetermined mixture," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 121–133, 2010.

[13] Z. Yang, B. Tan, G. Zhou, and J. Zhang, "Source number estimation and separation algorithms of underdetermined blind separation," *Science in China, Series F: Information Sciences*, vol. 51, no. 10, pp. 1623–1632, 2008.

[14] S. Araki, T. Nakatani, H. Sawada, and S. Makino, "Blind sparse source separation for unknown number of sources using Guassian mixture model fitting with Dirichet prior," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2009, pp. 33–36.

[15] S. Araki, T. Nakatani, and S. Makino, "Stereo source separation and source counting with MAP estimation with Dirichlet prior," in *Independent Component Analysis and Signal Separation*, 2009, pp. 742–750.

[16] M. Mandel, D. Ellis, and T. Jebara, "An EM algorithm for localizing multiple sound sources in reverberant environments," in *Advances in Neural Information Processing Systems*, 2007, pp. 953–960.

[17] S. Rickard, *The DUET Blind Source Separation*. Springer Press, 2007.

[18] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[19] L. Remaggi, P. J. Jackson, P. Coleman, and W. Wang, "Acoustic reflector localization: novel image source reversion and direct localization methods," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, no. 2, pp. 296–309, 2017.