MULTIPLE SOUND SOURCE LOCATION ESTIMATION AND COUNTING IN A WIRELESS ACOUSTIC SENSOR NETWORK

Anastasios Alexandridis *Athanasios Mouchtaris**

FORTH-ICS, Signal Processing Laboratory, Heraklion, Crete, Greece, GR-70013
University of Crete, Department of Computer Science, Heraklion, Crete, Greece, GR-70013

ABSTRACT

In this work, we consider the multiple sound source location estimation and counting problem in a wireless acoustic sensor network, where each sensor consists of a microphone array. Our method is based on inferring a location estimate for each frequency of the captured signals. A clustering approach—where the number of clusters (i.e., sound sources) is also an unknown parameter—is then employed to decide on the number of sources and their locations. The efficiency of our proposed method is evaluated through simulations and real recordings in scenarios with up to three simultaneous sound sources for different signal-to-noise ratios and reverberation times.

Index Terms— localization, DOA estimates, distributed microphone arrays, source counting, wireless acoustic sensor networks

1. INTRODUCTION

Wireless acoustic sensor networks (WASNs) represent a new paradigm for acoustic signal acquisition. They typically consist of nodes that are microphones or microphone arrays and feature signal processing and communication capabilities. WASNs find use in several applications, such as hearing aids, ambient intelligence, hands-free telephony, and acoustic monitoring [1]. A fundamental requirement with significant research interest for WASNs is to estimate the positions of the multiple active sound sources using the data acquired from the spatially distributed sensors. Location information is crucial in many signal processing tasks, such as noise reduction, source separation, and echo cancellation.

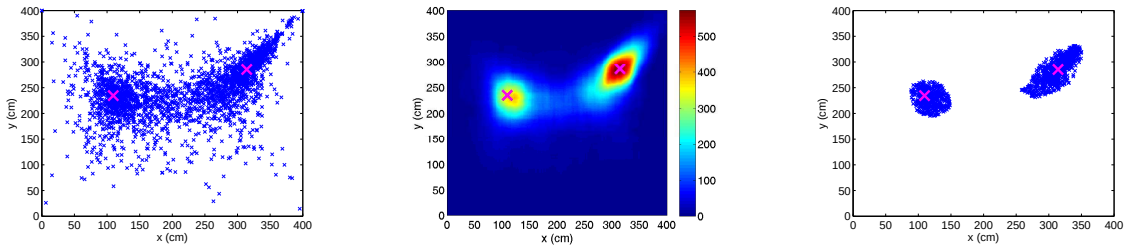
Typically, localization methods are divided into two classes: the direct and the indirect approaches [2]. Indirect approaches consist of a two-step procedure: each sensor is usually a microphone array that estimates time-differences of arrival (TDOAs) or direction of arrival (DOA) estimates. The estimates are then combined at a special node (the "fusion center") to infer the source position. The source locations are estimated through triangulation by intersecting DOA lines from the sensors [3–5] or by estimating the intersection of hyperbolas which are defined by the estimated TDOAs [6, 7]. An advantage of such methods is that they maintain transmission requirements at low levels, as only the TDOAs or DOAs of the active sources need to be transmitted at each time instant. However, when multiple sources are active, a key problem is that each sensor transmits the multiple TDOA/DOA estimates and the fusion center receiving these estimates cannot know to which source they belong. Moreover, in realistic scenarios missed detections can occur and the TDOA/DOA estimates of some sources from some sensors may be missing. The correct association of estimates across the arrays that correspond to the same source has to be found, otherwise location estimation will result in ghost-sources, i.e., locations not corresponding to real sources. This is known as the *data-association problem*. Solutions are proposed in [8–10] for scenarios with no missed detections and in our recent work of [11] which considers localization from DOA estimates—obtained using a DOA estimation method, e.g., [12]—in scenarios with missed detections

On the other hand, direct approaches are based on scanning a set of possible source locations and constructing a spatial likelihood map that describes the plausibility that a source is active at each candidate location. Approaches to construct likelihood maps are based on the Global Coherence Field (GCF) [13,14], the Steered Response Power (SRP) [15, 16], or the different level of access that each sensor has to different spatial positions [17]. The source location is then estimated from the peak of the spatial likelihood map. The work in [18] localizes multiple sources—whose number is known—by sequentially estimating the strongest peak in the likelihood map and then removing its contribution from the map, until all sources are found. However, with an increasing number of sources, the likelihood maps exhibit many local maxima some of which correspond to ghost sources, degrading the method's performance.

All the approaches above require the number of sources to be known *a priori*, which is not the case in a realistic scenario. In this work, we present a method —that shares common characteristics from both direct and indirect approaches—to estimate both the number of sources and their locations. It is based on estimating a location for each frequency of the captured signals, using the DOA estimates from the arrays at that frequency. An outlier rejection scheme is proposed to reject erroneous estimates due to noise and/or reverberation. Assuming that the location estimates are generated by a Gaussian mixture, we apply the Bayesian K-means clustering algorithm [19] to estimate the means of the Gaussians, as well as their number. The source number is given by the estimated number of Gaussians, while the sources' locations are given by their means.

A conceptually similar approach to counting is presented in [20] that employs a variational bayesian framework to model one-dimensional DOA estimates with a Gaussian Mixture. Their algorithm starts with a large number of Gaussian components (overdetermined case) and through an Expectation-Maximization (EM) algorithm and appropriate thresholing on the mixture weights some components are eliminated. After convergence, the number of sources is given by the number of components with non-zero weight. On the same spirit, a distributed approach to location estimation is presented in [21]. Other variational EM approaches focus on DOA estimation and counting and utilize mixtures of Watson distributions to model the Fourier coefficients of the captured signals [22] or incorporate tracking [23]. Compared to these works, our approach for online per-frame localization and counting oper-

*This research has been partly funded by the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 644283.



(a) (b) (c)

Figure 1: The effect of the outlier rejection process. (a) The per-frequency location estimates obtained using a history length of 1 sec. (b) The corresponding smoothed histogram obtained by filtering with a rectangular window of length $h_X = h_Y = 50$ cm in the $x-$ and $y-$ dimension. (c) The location estimates that remained after the outlier rejection process with $q = 0.35$. The X's represent the sources' true locations.

ates on a different way: it starts with one cluster (Gaussian component) and successively splits or merges clusters, an approach which eliminates the need for thresholding the estimated parameters.

2. PROPOSED METHOD

We consider a WASN with M nodes at known locations, each equipped with a circular microphone array. We assume that an unknown number of K sources are active at unknown locations.

2.1. In-node processing

The signals received at the ith microphone of each (say the mth) array are first transformed into the Short-Time Fourier Transform domain, resulting in the signals $X_{m,i}(\tau, \omega)$ where τ and ω denote the time frame and frequency index, respectively. We denote as (τ, Ω) the set of frequencies ω for frame τ up to a maximum frequency ω_{max}. In the remainder, we omit τ, as the procedure is repeated in each frame. Each array m estimates a DOA in each frequency $\omega \in \Omega$ resulting in the DOA estimates $\Theta_m(\Omega)$. For DOA estimation in each frequency we use the method of [24]. However, note that our proposed method for location estimation and counting is not restricted to a specific DOA estimation method or array geometry. The per frequency DOA estimates in $\Theta_m(\Omega)$ are then transmitted to the fusion center, which performs the localization and counting.

2.2. Processing at the fusion center

The fusion center estimates the number of active sources and their corresponding locations based on the DOA estimates in $\Theta_m(\Omega)$.

2.2.1. Per-frequency location estimation

First, a location is estimated for each frequency, based on the transmitted DOA estimates from the arrays at that frequency. For location estimation, we use the single-source version of our recently proposed grid-based (GB) method, which is a computationally efficient non-linear least squares estimator with high accuracy [11, 25]. The GB method constructs a grid of L grid points over the area of interest and finds the grid point whose DOAs most closely match the estimated DOAs. The location for each frequency ω is estimated as the co-ordinates of the grid point $\ell^*(\omega)$ that satisfies:

$$\ell^*(\omega) = \arg\min_{\ell} \sum_{m=1}^{M} [A(\Theta_m(\omega), \psi_{m,\ell})]^2 \qquad (1)$$

where $\psi_{m,\ell}$ is the DOA of the ℓth grid point to the mth array and $A(X, Y)$ denotes an angular distance function that returns the difference between X and Y in the range of $[0, \pi]$ [11]. We create a block of location estimates that contains the estimates of the current frame and B previous frames—also referred to as history length. Assuming that the signals are sufficiently sparse so that at most one source is dominant at each time-frequency point [26], we expect that the histogram of location estimates will have K clusters.

2.2.2. Outlier rejection

To remove erroneous estimates occurred due to noise and/or reverberation, we construct a two-dimensional histogram from the set of location estimates obtained from the previous step. We smooth the histogram by applying an averaging filter with a rectangular window $w(\cdot, \cdot)$ of length h_X and h_Y in the $x-$ and $y-$dimension, respectively. Erroneous estimates are expected to be of low cardinality in the smoothed histogram. Thus, we remove any location estimates whose cardinality is less than q times the maximum cardinality of the histogram, where $q \in [0, 1]$ is a pre-defined constant. The effect of the outlier rejection is shown in Figure 1 for a case of two active sources at 20 dB signal-to-noise ratio.

2.2.3. Clustering

The location estimates that remained are used for localization and counting. To do this, we employ the Bayesian K-means clustering algorithm proposed in [19]. The algorithm estimates the number of clusters and their centroids, which in our case correspond to the number of active sound sources and their locations, respectively. In the following, we briefly describe the Bayesian K-means approach.

Let $\boldsymbol{p}_n$, $n = 1, \ldots, N$ be the location estimates in the $D = 2$ dimensions, after the outlier rejection step. The algorithm assumes that the data have been generated by a mixture of C Gaussians:

$$p(\boldsymbol{p}_n|\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \sum_{c=1}^{C} \alpha_c \mathcal{N}(\boldsymbol{p}_n|\boldsymbol{\mu}_c, \boldsymbol{\Lambda}_c) \qquad (2)$$

where $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_C\}$ are the mixing coefficients, $\mathcal{N}(\cdot)$ is the normal distribution, and $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_C\}$ and $\boldsymbol{\Lambda} = \{\boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_C\}$ are the set of means and precision (inverse covariance) matrices of the Gaussians. We also define a C-dimensional binary cluster assignment variable $\boldsymbol{z}_n$ so that $z_{nc} = 1$ if the nth location estimate is assigned to cluster c (i.e., generated from the c-th Gaussian component) and $z_{nj} = 0$ for $j \neq c$.

In a Bayesian treatment of the mixture model, we place conjugate priors on the unknown parameters:

$$p(\boldsymbol{\alpha}) = \mathcal{D}(\boldsymbol{\alpha}, \phi_0), \quad p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{m_0}, \xi_0\boldsymbol{\Lambda})\, \mathcal{W}(\boldsymbol{\Lambda}|\eta_0, \boldsymbol{B}_0) \tag{3}$$

where $\mathcal{D}(\cdot)$ is the Dirichlet distribution and $\mathcal{W}(\cdot)$ is the Wishart distribution. The priors depend on the non-random hyper-parameters $\{\boldsymbol{m}_0, \boldsymbol{B}_0, \phi_0, \xi_0, \eta_0\}$.

The Bayesian K-means objective is to minimize the following function, jointly over assignment variables $\boldsymbol{z} = \{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_N\}$ and the number of clusters C:

$$\mathcal{F}(\boldsymbol{z}, C) = \sum_{c=1}^{C} \left[\frac{DN_c}{2}\log\pi + \frac{1}{2}\log\frac{\xi_c}{\xi_0} + \frac{\eta_c}{2}\log|\boldsymbol{B}_c| - \log\frac{\Gamma(\phi_c)}{\Gamma(\phi_0)} - \frac{\eta_0}{2}\log|\boldsymbol{B}_0| - \log\frac{\Gamma_D\left(\frac{\eta_c}{2}\right)}{\Gamma_D\left(\frac{\eta_0}{2}\right)} + \frac{1}{C}\log\frac{\Gamma(N + C\phi_0)}{\Gamma(C\phi_0)} \right] \tag{4}$$

where the dependence on $\boldsymbol{z}$ is through the cluster-dependent quantities $\boldsymbol{B}_c, \xi_c, \eta_c, \phi_c$ that are described in the following, N_c denotes the number of locations that belong to cluster c, $\Gamma_D(x) = \pi^{\frac{D(D-1)}{4}} \prod_{i=1}^{D} \Gamma\left(x + \frac{1-i}{2}\right)$, $\Gamma(\cdot)$ is the Gamma function, and $|\cdot|$ denotes the determinant of a matrix.

Update rules: Given C clusters, the algorithm performs the cluster assignment by iteratively minimizing the cost function:

$$\mathcal{C} = \sum_{n=1}^{N}\sum_{c=1}^{C} z_{nc}\gamma_c(\boldsymbol{p}_n) \tag{5}$$

with

$$\gamma_c(\boldsymbol{p}_n) = \frac{\eta_c}{2}(\boldsymbol{p}_n - \boldsymbol{m_c})^T\boldsymbol{B}_c^{-1}(\boldsymbol{p}_n - \boldsymbol{m}_c) + \frac{1}{2}\log|\boldsymbol{B}_c| + \frac{D}{2\xi_c} - \frac{1}{2}\sum_{d=1}^{D}\Psi\left(\frac{\eta_c + 1 - d}{2}\right) - \Psi(\phi_c) \tag{6}$$

where $\Psi(\cdot)$ is the "digamma" function and the cluster quantities $\{\boldsymbol{B}_c, \boldsymbol{m}_c, \xi_c, \eta_c, \phi_c\}$ are calculated as:

$$\begin{aligned} \boldsymbol{m}_c &= \frac{N_c\bar{\boldsymbol{p}}_c + \xi_0\boldsymbol{m}_0}{\xi_c} & \eta_c &= \eta_0 + N_c & \xi_c &= \xi_0 + N_c \\ \boldsymbol{B}_c &= \boldsymbol{B}_0 + N_c\boldsymbol{S}_c + \tfrac{N_c\xi_0}{\xi_c}(\bar{\boldsymbol{p}}_c - \boldsymbol{m}_0)(\bar{\boldsymbol{p}}_c - \boldsymbol{m}_0)^T & \phi_c &= \phi_0 + N_c \end{aligned} \tag{7}$$

where $\bar{\boldsymbol{p}}_c$ and $\boldsymbol{S}_c$ are the sample mean and sample covariance of cluster c, respectively. The algorithm—being similar to K-means—alternates between calculating the parameters $\{\boldsymbol{B}_c, \boldsymbol{m}_c, \xi_c, \eta_c, \phi_c\}$ and updating the cluster assignment according to:

$$z_{nc} = \begin{cases} 1 & \text{if } c = \arg\min_j \gamma_j(\boldsymbol{p}_n) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

The clustering assignment updates converge when the cost in Eq. (5) is kept constant between iterations.

Split and merge procedures: To search over different number of clusters, Bayesian K-means introduces split and merge operations which are based on the work of [27]. For each cluster a split score is calculated, based on the Kullback-Leibler divergence between the local empirical probability density around that cluster and the Gaussian mixture model. Moreover, for each pair of clusters i and j a merge score is calculated based on the cosine distance of the $N-$dimensional vectors that contain the posterior probabilities of the data points for the ith and jth Gaussian. The reader is referred to [19, 27] for more details on the merge and split scores, which are omitted here due to space limitations.

Algorithm 1 Bayesian K-Means Clustering [19]

Input: Location estimates $\boldsymbol{p}_n, n = 1, \ldots, N$

Output: Number of clusters C, cluster centroids $\boldsymbol{m}_c, c = 1, \ldots, C$

1. *Initialization:* Set $C = 1$, perform clustering using Eq. (5)-(8) until convergence, and evaluate Eq. (4) for this clustering assignment
2. *Split operations:* Calculate the split score for each cluster and sort them in descending order of scores.
   (i) Split the cluster with the highest score into two with centroids $\boldsymbol{m} \pm \boldsymbol{d}$ where $\boldsymbol{m}$ is the centroid of the cluster to be split and $\boldsymbol{d} = \boldsymbol{s}\sqrt{\lambda}$ with $\boldsymbol{s}$ being the principal eigenvector of the sample covariance matrix $\boldsymbol{S}$ of the cluster to be split and λ its corresponding eigenvalue.
   (ii) Perform clustering using Eq. (5)-(8) until convergence.
   (iii) Evaluate Eq. (4) for the new clustering assignment.
      a. If Eq. (9) holds, accept split and repeat STEP 2.
      b. Otherwise, reject split and go to STEP 2(i) to split the cluster with the next highest score.
3. *Merge operations:* Calculate the merge score for each pair of clusters and sort them in descending order of scores.
   (i) Merge the data points of the pair of clusters with the highest score into one cluster.
   (ii) Perform clustering using Eq. (5)-(8) until convergence.
   (iii) Evaluate Eq. (4) for the new clustering assignment.
      a. If Eq. (9) holds, accept merge and go to STEP 2.
      b. Otherwise, reject merge and go to STEP 3(i) to merge the pair of clusters with the next highest score.
4. *Terminate* when no more split/merge operations satisfy Eq. (9).

The complete algorithm is described in Algorithm 1. After each split and merge procedure the objective function in (4) is evaluated so as to accept or reject the split/merge operation. To avoid overestimation on the number of sources caused when the algorithm tries to overfit the data with a complex model with many clusters, we accept a split/merge operation only when the difference in the objective function is greater than a predefined threshold. More formally if we denote as $\mathcal{F}^b(\boldsymbol{z}, C)$ and $\mathcal{F}^a(\boldsymbol{z}, C)$ the value of (4) before and after a split/merge operation, we accept the operation iff

$$\frac{\mathcal{F}^b(\boldsymbol{z}, C) - \mathcal{F}^a(\boldsymbol{z}, C)}{\mathcal{F}^b(\boldsymbol{z}, C)} > t_{\text{accept}} \tag{9}$$

The algorithm terminates when no more split/merge operations satisfy (9), and outputs the number of clusters and their centroids, which denote the number of active sources and their locations.

3. RESULTS

3.1. Simulation Results

We performed simulations on a square cell of a WASN with dimensions of $V = 4$ m with four microphone arrays placed on the corners of the cell. Each array is a uniform circular array with $N = 8$ omnidirectional microphones and a radius $r = 0.05$ m. In each simulation, the sound sources were speech recordings of 3 seconds sampled at 44.1 kHz and had equal power when located at the center of the cell. The signal-to-noise ratio (SNR) was measured as the ratio of the power of each source signal when located at the center of the cell to the power of the noise signal. To simulate different SNR values we added white Gaussian noise at each microphone, uncorrelated with the source signals and the noise at the other microphones. Note that this framework results in different SNR at each array depending on how close the source is to the arrays.

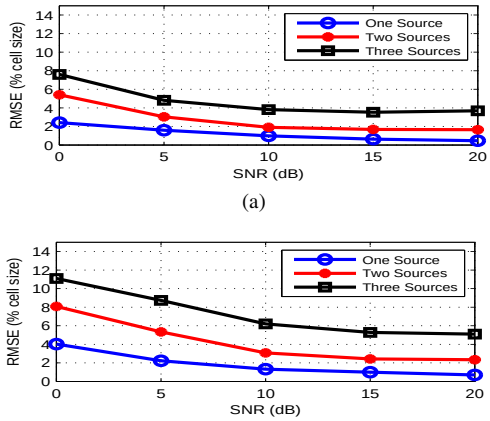


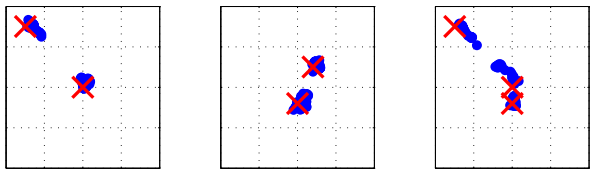
Figure 2: Location error as a percentage of the cell size $V = 4$ m for various number of active sound sources and reverberation time (a) $T_{60} = 250$ ms and (b) $T_{60} = 400$ ms

Table 1: Source counting success rates for various active sources in different SNR and reverberation conditions.

	$T_{60} = 250$ ms			$T_{60} = 400$ ms		
SNR	one source	two sources	three sources	one source	two sources	three sources
0 dB	98%	83%	51%	91%	68%	28%
5 dB	99%	89%	67%	97%	77%	46%
10 dB	98%	95%	88%	97%	93%	67%
15 dB	99%	98%	92%	96%	96%	80%
20 dB	98%	99%	94%	97%	97%	85%

We used the Image-Source method [28] to simulate a room of dimensions $10 \times 10 \times 3$ meters and produce signals of omnidirectional sources at various reverberation times. The WASN cell was placed in the middle of the room with the arrays and the sources being at 1.5 m height. We considered scenarios of up to three simultaneous sources. Each simulation was repeated 30 times and the sources were placed within the cell with independent uniform probability at a distance of at least 1 m away from each other and at least 0.5 m away from the arrays. For processing, we used frames of 2048 samples with 50% overlap, windowed with a Hamming window. The FFT size was 2048 and $\omega_{\text{max}} = 4$ kHz which is the spatial-aliasing frequency for the given array geometry. For the block processing we used a history length of 1 second which corresponds to $B = 43$ frames and the rectangular window for the smoothing of the histogram was of length $h_X = h_Y = 50$ cm. For outlier rejection we set $q = 0.35$, thus removing location estimates whose cardinality in the smoothed histogram is less than 35% the maximum cardinality. The histogram bin size was 1 cm^2. For the Bayesian K-means, the hyper-parameters were set to $\phi_0 = 2$, $\xi_0 = 0.1$, $\eta_0 = D = 2$, $\boldsymbol{m}_0 = \bar{\boldsymbol{p}}$, and $\boldsymbol{B}_0 = 2d_0^2\boldsymbol{S}/\text{trace}(\boldsymbol{S})$, where trace$(\cdot)$ denotes the trace of a matrix, $\bar{\boldsymbol{p}}$ and $\boldsymbol{S}$ are the sample mean and sample covariance matrix of the data, and d_0 is determined by computing the closest distance for 10% of the location estimates and averaging between the 3 closest pairs. Finally, we set t_{accept} to 0.01.

Table 1 depicts the source counting success rates as the percentage of frames for all 30 different source configurations—excluding the first $B-1$ frames for the history block initialization—where the correct number of sources was found for one, two, and three simultaneously active sources for different SNR levels and reverberation time of $T_{60} = 250$ ms and $T_{60} = 400$ ms. The method almost always identifies the correct number of sources in the single source scenario. It also yields accurate source counting performance for two and three sources, especially at the higher SNR cases for both reverberation conditions. Figure 2 shows the corresponding root-mean square error (RMSE) as a percentage of the cell size V, over all sources, all 30 different source configuration for all frames where the correct number of sources was detected. It is evident that the proposed method achieves quite accurate localization for all cases especially at higher SNR levels.



(a) RMSE = 4.62 % Counting Rate = 97%
(b) RMSE = 3.53 % Counting Rate = 74%
(c) RMSE = 9.14 % Counting Rate = 14%

Figure 3: Location errors (the blue cloud of estimates) throughout a 4-node square cell for real recordings of two and three active sources (the red X's). Each figure reports the location error (RMSE as % of the cell size V) and the source counting success rate.

3.2. Results of Real Measurements

We also performed some real recordings of acoustic sources in a 4-node square cell with sides $V = 4$ meters long. The nodes were 4-element circular microphone arrays of 2 cm radius. The sources were recorded speech signals of approximately 5 seconds duration, played back through loudspeakers at different locations and their SNR at the center of the cell was measured to be about 10 dB. The parameter setting for our method was the same as reported in Section 3.1. Figure 3 shows the results for different locations of two (Figure 3(a) & (b)) and three (Figure 3(c)) active sources. It can be seen that for the two source cases the method results in accurate counting and localization. For the three source case, the counting success rate is low, which is however also due to the fact that the two sources in the middle of the cell are located too close together. It should also be highlighted that these recordings took place outdoors, and while they may not have many reflections, there was a significant level of distant noise sources, such as dogs barking and cars passing by. Moreover, the locations and orientations of the arrays were not finely calibrated and had unintended offsets of a few centimetres and degrees. Thus, the conditions were far from ideal, making the results of our proposed localization and counting method even more encouraging.

4. CONCLUSIONS

In this work, we considered the joint problem of source counting and location estimation in a wireless acoustic sensor network, where each sensor is a microphone array that transmits per-frequency DOA estimates to the fusion center. We presented a method that performs outlier rejection and incorporates a bayesian clustering approach that have been proposed when the number of clusters is also unknown. Through simulations and experiments with real recordings we showed the effectiveness of our method to count the number of active sound sources and accurately estimate their locations.

5. REFERENCES

[1] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," in *IEEE Symp. on Communications and Vehicular Technology in the Benelux*, 2011, pp. 1–6.

[2] N. Madhu and R. Martin, "Acoustic source localization with microphone arrays," in *Advances in Digital Speech Transmission*. Wiley, 2008, pp. 135–166.

[3] M. Gavish and A. J. Weiss, "Performance analysis of bearing-only target location algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 817–828, 1992.

[4] K. Doğançay, "Bearings-only target localization using total least squares," *Signal Processing*, vol. 85, no. 9, 2005.

[5] L. M. Kaplan, Q. Le, and N. Molnar, "Maximum likelihood methods for bearings-only target localization," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, 2001, pp. 3001–3004.

[6] A. Canclini, E. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 2, pp. 439–443, Feb 2013.

[7] M. Compagnoni, P. Bestagini, E. Antonacci, A. Sarti, and S. Tubaro, "Localization of acoustic sources through the fitting of propagation cones using multiple independent arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 1964–1975, 2012.

[8] J. Reed, C. da Silva, and R. Buehrer, "Multiple-source localization using line-of-bearing measurements: Approaches to the data association problem," in *IEEE MILCOM*, Nov 2008, pp. 1–7.

[9] L. M. Kaplan, P. Molnar, and Q. Le, "Bearings-only target localization for an acoustical unattended ground sensor network," in *Proc. SPIE*, vol. 4393, 2001, pp. 40–51.

[10] M. Swartling, N. Grbić, and I. Claesson, "Source localization for multiple speech sources using low complexity non-parametric source separation and clustering," *Signal Processing*, vol. 91, no. 8, pp. 1781–1788, 2011.

[11] A. Griffin, A. Alexandridis, D. Pavlidi, Y. Mastorakis, and A. Mouchtaris, "Localizing multiple audio sources in a wireless acoustic sensor network," *Signal Processing*, vol. 107, pp. 54 – 67, 2015, Special Issue on ad hoc microphone arrays and wireless acoustic sensor networks. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168414003764

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

[13] M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 1994*, Apr 1994, pp. II/273–II/276 vol.2.

[14] A. Brutti, M. Omologo, and P. Svaizer, "Speaker localization based on oriented global coherence field." in *INTERSPEECH*. ISCA, 2006.

[15] H. Do and H. Silverman, "A Fast Microphone Array SRP-PHAT Source Location Implementation using Coarse-To-Fine Region Contraction (CFRC)," in *IEEE WASPAA*, Oct 2007, pp. 295–298.

[16] M. Cobos, A. Marti, and J. J. Lopez, "A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling," *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 71–74, Jan 2011.

[17] P. Aarabi, "The fusion of distributed microphone arrays for sound localization," *EURASIP Journal on Applied Signal Processing*, vol. 2003, pp. 338–347, Jan. 2003. [Online]. Available: http://dx.doi.org/10.1155/S1110865703212014

[18] A. Brutti, M. Omologo, and P. Svaizer, "Multiple source localization based on acoustic map de-emphasis," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2010, pp. 1–17, 2010.

[19] K. Kurihara and M. Welling, "Bayesian k-means as a "maximization-expectation" algorithm," *Neural Computation*, vol. 21, no. 4, pp. 1145–1172, Apr 2009.

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

[21] Y. Dorfan, G. Hazan, and S. Gannot, "Multiple acoustic sources localization using distributed expectation-maximization algorithm," in *Hands-free Speech Communication and Microphone Arrays (HSCMA), 2014 4th Joint Workshop on*, May 2014, pp. 72–76.

[22] L. Drude, A. Chinaev, D. H. T. Vu, and R. Haeb-Umbach, "Source counting in speech mixtures using a variational em approach for complex watson mixture models," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014, pp. 6834–6838.

[23] N. Madhu and R. Martin, "A scalable framework for multiple speaker localization and tracking," in *in Proceedings of the International Workshop for Acoustic Echo Cancellation and Noise Control (IWAENC 2008*, 2008.

[24] A. Karbasi and A. Sugiyama, "A new DOA estimation method using a circular microphone array," in *Proc. of EUSIPCO*, 2007, pp. 778–782.

[25] A. Griffin, A. Alexandridis, D. Pavlidi, and A. Mouchtaris, "Real-time localization of multiple audio sources in a wireless acoustic sensor network," in *Proc. of EUSIPCO*, Sept 2014, pp. 306–310.

[26] S. Rickard and O. Yilmaz, "On the approximate w-disjoint orthogonality of speech," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 1, May 2002, pp. I–529–I–532.

[27] N. Ueda, R. Nakano, Z. Ghahramani, and G. E. Hinton, "SMEM Algorithm for Mixture Models," *Neural Computation*, vol. 12, no. 9, pp. 2109–2128, Sep 2000.

[28] E. Lehmann and A. Johansson, "Diffuse reverberation model for efficient image-source simulation of room impulse responses," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, Aug. 2010.