REAL-TIME LOCALIZATION OF MULTIPLE AUDIO SOURCES IN A WIRELESS ACOUSTIC SENSOR NETWORK

Anthony Griffin[⋆], *Anastasios Alexandridis*[⋆†], *Despoina Pavlidi*[⋆†], *and Athanasios Mouchtaris*[⋆†]

[⋆]FORTH-ICS, Heraklion, Crete, Greece, GR-70013
[†]University of Crete, Department of Computer Science, Heraklion, Crete, Greece, GR-70013
Email: doctorgriffin@yahoo.co.uk, {analexan, pavlidi, mouchtar}@ics.forth.gr

ABSTRACT

In this work we propose a grid-based method to estimate the location of multiple sources in a wireless acoustic sensor network, where each sensor node contains a microphone array and only transmits direction-of-arrival (DOA) estimates in each time interval, minimizing the transmissions to the central processing node. We present new work on modeling the DOA estimation error in such a scenario. Through extensive, realistic simulations, we show our method outperforms other state-of-the-art methods, in both accuracy and complexity. We present localization results of real recordings in an outdoor cell of a sensor network.

Index Terms— Acoustic sensors, acoustic source localization, location estimation, microphone arrays, wireless acoustic sensor networks

1. INTRODUCTION

The popularity of microphone arrays has led to increased interest in wireless acoustic sensor networks (WASNs), where a number of microphones are distributed over an area to provide better spatial coverage. WASNs are attractive due to their wide variety of applications in hearing aids, hands-free telephony, acoustic monitoring, and ambient intelligence [1].

For such applications, information about the location of the source is an essential, but challenging task due to several constraints inherent in a sensor network (time-synchronization, power and bandwidth limitations, etc). By allowing increased computational resources in the nodes, the absolute minimum transmission bandwidth can be attained when each sensor node transmits only a direction-of-arrival (DOA) estimate to the central processing node [2, 3]. Moreover, localization using DOA estimates from multiple microphone arrays "relaxes" the time-synchronization constraints as the individual nodes need not be perfectly synchronized.

Solutions to the single source localization problem include linear estimators [4], and maximum likelihood (ML) approaches [5–7]. However, in many realistic scenarios multiple sources may co-exist in an area and the location of all sources may need to be known. The localization of multiple acoustic sources poses many challenges. First of all, there is the so-called *data association problem*, where the central node receiving DOA estimates for multiple sources from the different sensors cannot know to which source they belong. Erroneous DOA combinations will result in "ghost sources" that do not correspond to real sources. A solution to this problem was given in [8] but has been found to be *NP-hard* when the number of sensors is greater than two, and the solution of [9] is only suitable for noiseless scenarios. The work in [10] proposes a solution based on statistical clustering of the intersection of bearing lines. However, they consider idealized scenarios of no missed detections and no spurious measurements. A method using non-linear least squares that tries to overcome the data association problem is discussed in [11]. However, ghost sources are not eliminated, leading to severe performance degradation.

Our previous experience with DOA estimation [12, 13] revealed that when the sources are close together some arrays may only detect one source, an observation made from experiments using real recorded signals. As a result, the DOAs of some sources from some sensors might be missing. This problem of missing DOA estimates as a function of the sources' locations is an important aspect which—to the best of our knowledge—has not been widely examined.

Our work in [14] considered a method for localizing two sources using far-field DOA measurements in an outdoor WASN. This paper extends [14] to more than two sources, and proposes a novel grid-based approach which is an alternative solution to the non-linear least squares (NLS) estimator. We solve the data association problem using a sub-optimal—yet computationally efficient—method which relies on the estimated locations and the corresponding DOA combinations. Our approach is real-time and as our simulations and real experiments show, it loses very little in accuracy.

Our simulations use new results that we present here to model the DOA estimation error of the algorithm of [13], and consider the problem of missing DOAs as a function of source location, which makes them more realistic than simulations considered previously. The problem of missing DOAs when the sources are close together occurs often in practice as our real experiments in this paper suggest.

2. THE FRAMEWORK

Our framework is a wireless sensor network whose M nodes are each equipped with a microphone array—which we will also refer to as a sensor. This enables each node to generate a DOA estimate for any sources that it can "hear" (any sources whose signal-to-noise ratio (SNR) at the node is high enough to be detected). Fig. 1 illustrates an example square cell with four nodes—separated by V—and the DOA estimates to the sources. Note that each node estimates a direction only with no range information, thus one node's DOA estimates are not sufficient to obtain absolute positions for the sources. We assume that the source's signal radiates as a spherical wave, and the attenuation it experiences in travelling from r_1 meters from the

This research has been co-financed by the European Union and Greek national funds through the National Strategic Reference Framework (NSRF), Research Funding Program: "Cooperation-2011", Project "SeNSE".

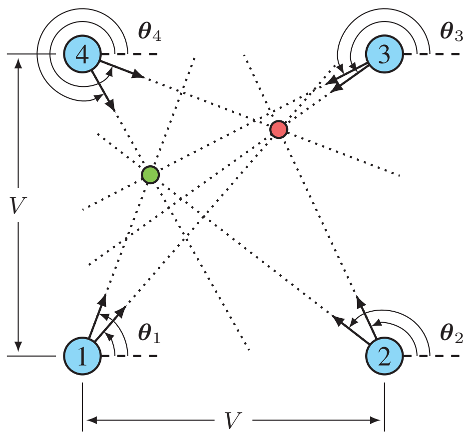


Fig. 1. Example cell with four sensor nodes (blue circles, 1 to 4), and the DOAs ($\boldsymbol{\theta}_1$–$\boldsymbol{\theta}_4$) to two sources (the red and green circles).

source to r_2 meters from the source is [15]

$$a = 20 \log_{10} \frac{r_2}{r_1} \quad \text{dB}. \tag{1}$$

Thus by specifying a signal-to-noise ratio (SNR) at (or very near) the source, we can determine the SNR at each sensor.

3. LOCALIZING SOURCES FROM MULTIPLE DOA ESTIMATES

By considering Fig. 1 again, it is clear that in the ideal case—i.e., perfect DOA estimates—the sources could be localized by finding the points where four DOA vectors intersect. In practice—or any realistic simulation—the DOA estimates will not be perfect, and will not all intersect at the same point. In previous work [14], we proposed an algorithm to solve this based on the centroid of intersection points of DOA vectors.

Here, we consider a grid-based (GB) method, which is an alternative formulation of the NLS estimator of [7], and tries to alleviate the major weaknesses of that approach, namely the need for a good initial point to ensure the estimator does not converge to a local minimum, and the computational burden of the minimization procedure.

Our approach is based on discretizing the area of interest into a grid of N points, and then finding the grid point whose DOAs most closely match the estimated DOAs. Moreover, since our measurements are angles, we propose the use of an Angular Distance—defined in the following—as a more proper measure of "similarity" than the absolute distance utilized in [7]. As we will show, this approach is much more computationally efficient and sacrifices no accuracy. Our grid is thus an $(M \times N)$ matrix, $\boldsymbol{\Psi}$, whose elements, $\psi_{m,n}$ are the DOAs from the m-th sensor to the n-th grid point.

3.1. Single-source localization

Localizing a single source with this method is conceptually very straightforward: we find the grid point whose DOAs most closely match the estimated DOAs. However—due to their inherent modulo-2π nature—in order to properly compare angles, we must first define an "angular distance" function, $A(X, Y)$, whose output is constrained to the range $[0, \pi]$. An elegant—if somewhat inefficient—implementation of $A(X, Y)$ is given by

$$A(X, Y) = 2 \sin^{-1} \frac{|\exp(jX) - \exp(jY)|}{2}. \tag{2}$$

With this defined, our problem may then be expressed as

$$n^* = \arg\min_{n} \sum_{m=1}^{M} \left[A(\hat{\theta}_m, \psi_{m,n}) \right]^2, \tag{3}$$

where $\hat{\theta}_m$ is the DOA estimate from the m-th sensor. The source position estimate $\hat{\boldsymbol{p}}_{\text{GB}}$ is simply given as the co-ordinates of the n^*-th grid point.

As a source may be located anywhere within the area of interest, it should be clear that the resolution of the grid —determined by the number of grid points, N—will determine the position estimation error. However, increasing N to decrease the position estimation error will also increase the complexity of the algorithm. A good compromise is to use an iterative algorithm that starts with a coarse grid, and once the best grid point is found, a new grid centered on this point is generated, with a smaller spacing between grid points, but also a smaller scope. Then the best grid point in the new grid is found. This may be repeated until the desired accuracy is obtained, while keeping the complexity under control.

3.2. Multiple-source localization

As previously discussed, the multiple-source case introduces further challenges. The processing node receiving the DOA estimates cannot know to which source they belong, and the localization algorithm must take this into account. An additional complication is that some sensor nodes may underestimate the true number of sources, as the sources' DOAs may be too close together for that node to discriminate between them. We call this the *minimum angular source separation* (MASS) of a sensor node, i.e., if the angular distance between two sources is less than the MASS, then the sensor node will only detect one source. Thus our localization algorithm must deal with the ambiguity that each DOA estimate may originate from either source, and that some (or even all) of the sensor nodes may underestimate the number of sources. Let C_s be the number of sensors that detect s sources, and S be the highest integer for which $C_s \neq 0$, i.e., the highest number of sources detected by at least one sensor. Our multiple source localization consists of a two-step procedure: in the first step, an initial candidate location is estimated for each possible combination of DOA measurements, while in the second step, the final S source locations must be chosen from the candidate locations.

Let $\mathcal{J}$ denote the set of all possible unique combinations of DOA estimates and j enumerate the combinations. The cardinality of $\mathcal{J}$ depends on the number of sources each sensor is able to detect and can be computed as:

$$|\mathcal{J}| = \sum_{s=1}^{S} s^{C_s} \tag{4}$$

As the correct association of the DOAs of each sensor to the sources cannot be known, the single-source GB method of Section 3.1 is applied to each element of $\mathcal{J}$ and the set $\mathcal{L}$ of candidate source locations is formed with $|\mathcal{L}| = |\mathcal{J}|$. Note that this multiple source localization algorithm increases complexity by at least $|\mathcal{J}| - 1$ times that of the single source algorithm, highlighting the need for a computationally efficient method to perform the single-source localization of each DOA combination. As we show in Section 4.3, using a standard NLS method here would significantly increase the computational burden without gaining any increase in accuracy over our GB method of Section 3.1. In the next step, the final S source locations must be identified from the set of candidate locations $\mathcal{L}$ by solving the data association problem.

3.2.1. *Brute-force approach*

A brute-force solution [10] to the data association problem is to perform an exhaustive search over all possible S-tuples of DOA combinations and select the most likely one. An S-tuple of DOA combinations is defined as the list of S DOA combinations (elements of $\mathcal{J}$) each of them being an $M \times 1$ vector of DOA measurements from the M sensors. Moreover, in forming an S-tuple, each sensor must contribute to each of the S DOA combinations with a different estimate, as the same DOA cannot belong to more than one source. In the case where a sensor has not detected all sources the same DOA can be repeated. This approach suffers from very high complexity as the number of tuples that need to be tested can grow as high as $\mathcal{O}\left((S!)^M\right)$, making this method highly impractical even for a moderate number of sources and sensors.

3.2.2. *Sequential approach*

Here we propose a computationally efficient approach to solve the data association problem. This sub-optimal approach relies on a sequential procedure to find the S DOA combinations without testing all the possible S-tuples. Let $\hat{\boldsymbol{\theta}}^{(j)}$ be the $M \times 1$ vector of DOAs for the j-th combination, and let $\hat{\theta}_m^{(j)}$ denote the DOA of the m-th sensor for the j-th combination. Each $\hat{\boldsymbol{\theta}}^{(j)}$ is associated with a candidate source location $\boldsymbol{p}^{(j)}$, and $\theta_m\left(\boldsymbol{p}^{(j)}\right)$ is the DOA of the m-th sensor from $\boldsymbol{p}^{(j)}$. Our approach can then be stated as:

1. Create a set $\mathcal{J}' = \mathcal{J}$.
2. For each DOA combination j in the set $\mathcal{J}'$ compute the residual:
$$r_j = \sum_{m=1}^{M} \left[A\left(\hat{\theta}_m^{(j)}, \theta_m\left(\boldsymbol{p}^{(j)}\right)\right)\right]^2. \tag{5}$$
3. Choose the DOA combination j^* with the minimum residual and output the corresponding location $\boldsymbol{p}^{(j^*)}$ as the location of one of the sources.
4. Update $\mathcal{J}'$ by subtracting all DOA combinations that contain DOAs that are part of the previously chosen combination j^*. Only DOAs of the sensors that have not detected all sources are allowed to take part in other combinations.
5. Repeat steps 2–4 until $\mathcal{J}' = \emptyset$, i.e., all S sources have been found.

Note that this approach does not test all possible S-tuples of DOA combinations, significantly reducing the computational burden to that of testing $\mathcal{O}(|\mathcal{J}|)$ DOA combinations. From (4), this will be $\mathcal{O}(S^M)$ when the MASS is $0°$, and decrease as the MASS increases.

4. RESULTS AND DISCUSSION

To investigate the performance of our proposed method, we performed simulations and real measurements of a square 4-node cell of a WASN, similar to that of Fig. 1. Although this is just a study of one cell in a larger sensor network, it is a reasonable assumption that a cell's performance will dominate the whole network's performance.

4.1. DOA Estimation Error Modeling

The DOA estimation error at each sensor was assumed to be normally distributed with a zero mean and a variance dependent only on the SNR at each sensor, which was in turn determined by the length of the path from the source to the sensor. Following the DOA estimation method of [13], we performed simulations to characterize the DOA estimation error, using sensors consisting of 4-element circular microphone arrays with a radius of 2 cm. We assumed an anechoic environment and simulated various SNR cases ranging from -5 dB to 20 dB. For each SNR, the simulation was repeated with the source rotated in $1°$ increments around the array to avoid any orientation biasing effects. Fig. 2 shows the standard deviations obtained when the DOA estimation error at each SNR was fitted with a Gaussian distribution. The fitted curve in Fig. 2 is given by $f(x) = 1.979\exp(-0.2815x) + 1.884$.

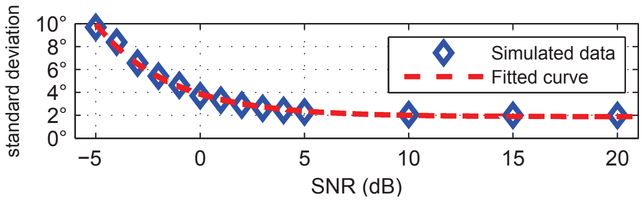


Fig. 2. Modeling the effect of SNR on DOA estimation error standard deviation for a circular microphone array.

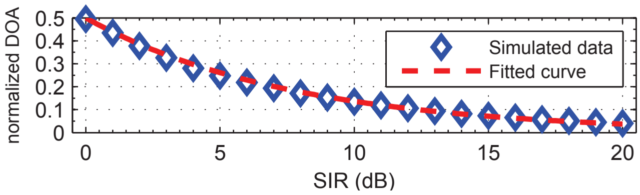


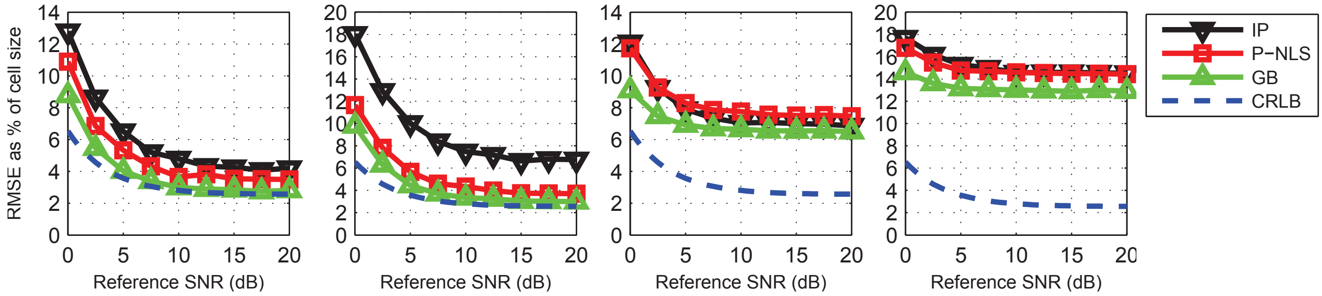
Fig. 3. Modeling the effect of MASS and SIR on DOA estimation error for a circular microphone array.

By specifying a reference SNR at the center of the cell, the SNR at each sensor can then be calculated through geometry and the use of (1), the DOA estimation standard deviation is then taken from the fitted curve of Fig. 2. It must be emphasized here that our framework results in a different SNR and, therefore, a different DOA estimation error standard deviation at each sensor.

It was also important to model the effect on DOA estimation when two sources were within the MASS of a sensor. We performed a simulation study where two sources were set at various separations of up to $20°$—below the MASS of the method of [13]—and the energy of the second source was incrementally decreased so the signal-to-interferer ratio (SIR) seen by the first source varied from 0 dB to 20 dB. These simulations were then repeated with the sources being rotated around the array in $1°$ increments—whilst preserving their angular separation—to avoid any orientation biasing effects. In all simulations only one source was detected and Fig. 3 shows the results of these simulations, where the DOA error has been normalized by the separation between the sources. The fitted curve of Fig. 3 is given by $f(x) = 0.5\exp(-0.12987x)$. It is clear that the detected source's DOA is estimated in the middle of the true DOAs when the sources have equal energy, and moves towards the dominant source as the weaker source decreases in energy. We used the fitted curve of Fig. 3 in all simulations involving more than one source.

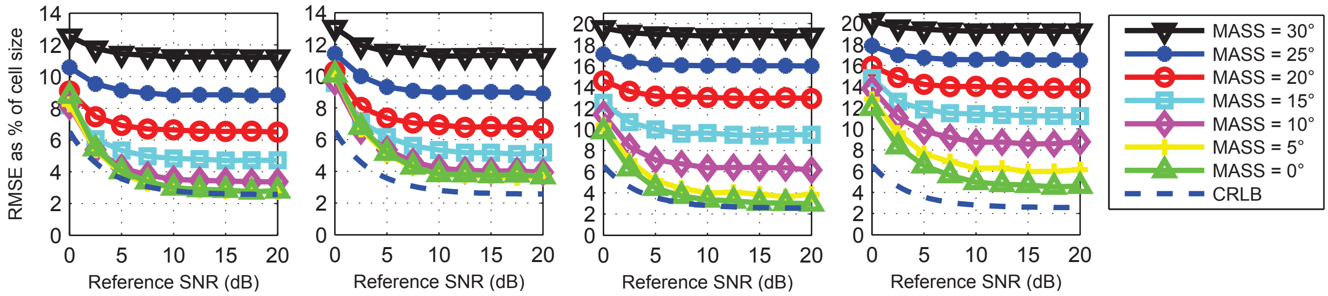
4.2. Simulation Results

In all simulations, the sources were located anywhere within the cell with independent uniform probability and the error measurement used was the root mean square error (RMSE) between the estimated positions and the true source positions. The methods compared were: (GB) the proposed grid-based method; (IP) our intersection point method of [14]; and (P-NLS) the position non-linear least squares method of [11], which we extended by using the final step approaches of Sections 3.2.1 & 3.2.2. We also calculated the Cramér-Rao lower bound (CRLB) of [3], which we extended to multiple sources, but could not include here due to space restrictions.



(a) 2 sources, MASS = 0° (b) 3 sources, MASS = 0° (c) 2 sources, MASS = 20° (d) 3 sources, MASS = 20°

Fig. 4. Position estimation error as a percentage of cell size V in a square 4-node cell.



(a) 2 sources, brute force (b) 2 sources, sequential (c) 3 sources, brute force (d) 3 sources, sequential

Fig. 5. Position estimation error as a percentage of cell size V in a square 4-node cell using the proposed method and the final step approaches of Sections 3.2.1 & 3.2.2.

Fig. 4(a) & (b) present the results of our simulations of two and three sources for the idealized case of MASS$=0^\circ$. Both the P-NLS and GB methods used the brute force approach of Section 3.2.1 for the final source location selection, and the GB method used initial and final grids with grid point spacing of 12.5% and 0.25% of the sensor spacing, respectively. It is very encouraging to see how close the performance of the GB method is to the lower bound.

Any realistic sensors and DOA estimation algorithm will have a non-zero MASS, and the performance of all localization algorithms is expected to degrade significantly as the MASS increases. This is due to the fact that the accuracy of the algorithms degrades as C_S decreases, and an increasing MASS directly decreases C_S, especially as the number of sources increases. Another way to think of this is that as the MASS increases, the accuracy of the DOA estimates from each sensor is much more likely to degrade significantly, due to the "merging" effect illustrated in Fig. 3. In the extreme case, C_S will be zero—i.e., no sensors will detect the true number of sources—and the localization algorithm will underestimate the number of source locations. A more realistic case of 20° MASS is presented in Fig. 4(c) & (d), and the degrading effect of the increased MASS is clear, particularly for the three source case. Note again, that the GB method consistently performs the best.

Fig. 5 presents the difference in performance for the two approaches of Sections 3.2.1 & 3.2.2 with the GB method for two and three sources. It is clear that very little performance is lost using the sequential approach particularly at the higher—and more realistic—values of MASS. The loss in performance is higher at low values of MASS, and for the three source case. Although not shown here due to space considerations, because the P-NLS method must use either the brute force or the sequential approach, it too suffers a similar performance loss to that of the GB method. Fig. 5 also illustrates the effect of MASS on the RMSE, highlighting the importance that the DOA estimation method used has a low MASS, if high accuracy is required by the system.

Table 1. Mean execution times in milliseconds for one set of DOA estimations

	MASS $=0^\circ$		MASS $=20^\circ$	
Method	two sources	three sources	two sources	three sources
IP	6.9	44.5	5.3	16.2
GB (& BF)	36.0	2961.6	19.2	214.3
GB (& Seq.)	29.4	162.8	16.8	26.7
P-NLS (& BF)	382.0	5033.4	205.1	509.6
P-NLS (& Seq.)	375.3	2238.8	202.7	322.1

4.3. Complexity

All the localization algorithms simulated were implemented in Matlab on a Windows laptop with a Core i5 CPU running at 2.53 GHz with 4 GB RAM, and their mean execution times are presented in Table 1. Note that while the absolute execution times may be highly dependent on the machine, we are only interested here in the relative times between the methods. The IP method is the fastest, and the P-NLS method is clearly the slowest, due to the non-linear optimization it requires. Table 1 also highlights the dramatic reduction in complexity when using the sequential rather than the brute force approach, particularly in the three source case. Also evident is the effect of the MASS on the complexity, however these complexity savings are offset by a reduction in accuracy as the MASS increases, as discussed in Section 4.2.

These results, together with those of Section 4.2, strongly suggest that the GB method with the sequential approach of Section 3.2.2 is the best choice given its accuracy and moderate com-

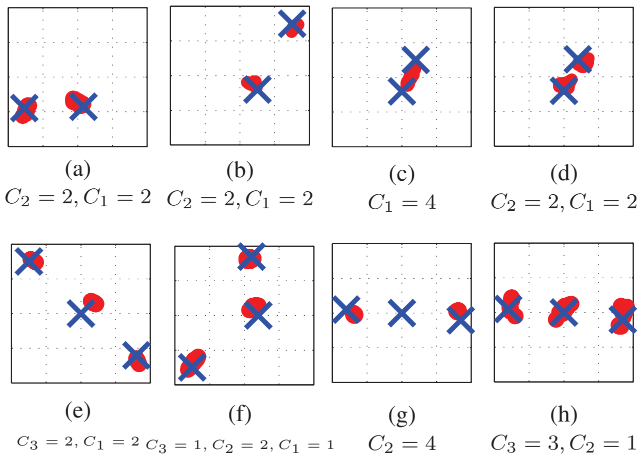


Fig. 6. Position estimates (the red clouds) in a square 4-node cell, of real recordings of two or three sources (the blue X's).

plexity. To further verify this suitability, we implemented the GB method with the sequential approach in C++ and measured that it only consumed 25% of the available processing time, making it an excellent candidate for a real-time system.

4.4. Results of Real Measurements

We also performed some real recordings of acoustic sources in a 4-node square cell with sides 4 meters long. The sensors on the nodes were circular 4-element microphone arrays with a radius of 2 cm, and the DOA estimation was performed by our real-time system of [12, 13]. The sources were recorded speech, played back simultaneously through loudspeakers at different locations, and their SNR at the center of the cell was measured to be about 10 dB. Although a 4×4 metre square is not a particularly large area, since we measure our reference SNR at the center of the cell, these results should be scalable to larger cells. Fig. 6 shows the position estimates from the real recordings using the proposed grid-based method for different layouts of two and three sources. The red dots are the cloud of estimates over about 5 seconds, and show quite accurate localization. All of the plots *except* (d) and (h) used the standard parameter set of [12, 13] which has a MASS of around 20°, and it is clear that in (c) and (g) the source positions are underestimated. By modifying some of the parameters of the DOA estimation, we were able to decrease the system's MASS so that all the sources in (d) and (h) could be localized, albeit with a greater variance in the estimates.

It should be noted that these recordings took place outdoors, and as such did not have many reflections, but there was a significant level of distant noise sources, such as cars and dogs barking. Furthermore, the orientations of the sensors were not finely calibrated, and the DOA estimates likely have unintended offsets of a few degrees. Thus the conditions were far from ideal, making the results of our proposed localization method even more encouraging.

5. CONCLUSIONS

In this work we have considered the challenge of localization in a WASN where each sensor node only transmits direction-of-arrival estimates, minimizing the transmissions to the processing node. We considered some real problems in this scenario, such as accurately-modeled DOA estimation error, and the merging of two DOA estimates that are very close together. We presented a real-time, grid-based method to perform the position estimation of multiple sources along with a sequential approach to the final source location selection. Through extensive simulations and measurements we showed that our proposed method outperforms the other state-of-the-art methods considered in both accuracy and computational complexity.

REFERENCES

[1] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," in *IEEE Symp. on Comm. and Veh. Tech. in the Benelux*, 2011.

[2] A. Ledeczi, G. Kiss, B. Feher, P. Volgyesi, and G. Balogh, "Acoustic source localization fusing sparse direction of arrival estimates," in *Int. Workshop on Intelligent Solutions in Embedded Systems*, 2006.

[3] A. Farina, "Target tracking with bearings-only measurements," *Signal Processing*, vol. 78, no. 1, 1999.

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

[5] M. Gavish and A. J. Weiss, "Performance analysis of bearing-only target location algorithms," *IEEE Trans. on Aerospace and Electr. Syst.*, vol. 28, no. 3, 1992.

[6] A. N. Bishop, B. D. O. Anderson, B. Fidan, P. N. Pathirana, and G. Mao, "Bearing-only localization using geometrically constrained optimization," *IEEE Trans. on Aerospace and Electr. Syst.*, vol. 45, no. 1, 2009.

[7] L. M. Kaplan, Q. Le, and N. Molnar, "Maximum likelihood methods for bearings-only target localization," in *IEEE Int. Conf. on Acoust., Speech, and Sig. Proc. (ICASSP)*, 2001.

[8] S. Deb, M. Yeddanapudi, K. Pattipati, and Y. Bar-Shalom, "A generalized S-D assignment algorithm for multisensor-multitarget state estimation," *IEEE Trans. on Aerospace and Electr. Syst.*, vol. 33, no. 2, April 1997.

[9] A.N. Bishop and P.N. Pathirana, "Localization of emitters via the intersection of bearing lines: A ghost elimination approach," *IEEE Trans. on Veh. Tech.*, vol. 56, no. 5, Sept 2007.

[10] J.D. Reed, C.R.C.M. da Silva, and R.M. Buehrer, "Multiple-source localization using line-of-bearing measurements: Approaches to the data association problem," in *IEEE Mil. Comm. Conf. (MILCOM)*, 2008.

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

[12] D. Pavlidi, M. Puigt, A. Griffin, and A. Mouchtaris, "Real-time multiple sound source localization using a circular microphone array based on single-source confidence measures," in *Proc. of IEEE Int. Conf. on Acoust., Speech, and Sig. Proc. (ICASSP)*, 2012.

[13] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, "Real-time multiple sound source localization and counting using a circular microphone array," *IEEE Trans. on Audio, Speech, and Lang. Proc.*, vol. 21, no. 10, 2013.

[14] A. Griffin and A. Mouchtaris, "Localizing multiple audio sources from DOA estimates in a wireless acoustic sensor network," in *Proc. of IEEE Work. on App. of Sig. Proc. to Audio and Acoust. (WASPAA)*, October 2013.

[15] Malcolm J. Crocker, *Handbook of Acoustics*, Wiley, 1998.