# ADDRESSING THE DATA-ASSOCIATION PROBLEM FOR MULTIPLE SOUND SOURCE LOCALIZATION USING DOA ESTIMATES

*Anastasios Alexandridis, Giorgos Borboudakis, and Athanasios Mouchtaris*[*]

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

## ABSTRACT

In this paper, we consider the data association problem that arises when localizing multiple sound sources using direction of arrival (DOA) estimates from multiple microphone arrays. In such a scenario, the association of the DOAs across the arrays that correspond to the same source is unknown and must be found for accurate localization. We present an association algorithm that finds the correct DOA association to the sources based on features extracted for each source that we propose. Our method results in high association and localization accuracy in scenarios with missed detections, reverberation, and noise and outperforms other recently proposed methods.

***Index Terms***— data-association, multiple microphone arrays, direction of arrival, wireless acoustic sensor networks, localization

## 1. INTRODUCTION

Wireless acoustic sensor networks (WASNs) consist of nodes that are microphones or microphone arrays and feature signal processing and wireless communication capabilities to perform computations and communicate with each other. WASNs offer richer sensing capabilities than a single microphone array and find use in applications such as hearing aids, ambient intelligence, hands-free telephony, and acoustic monitoring [1]. In such applications, information about the sources' locations is important for operations like noise reduction and speech enhancement. Location estimation is performed by fusing information from the arrays through the use of the full audio signals [2], the Steered Response Power function [3], or the construction of acoustic maps [4]. However, such approaches require the transmission of a significant amount of information through the network resulting in high bandwidth consumption.

The minimum transmission bandwidth can be attained when each node transmits a direction of arrival (DOA) estimate. In the single source case, location can be estimated as the intersection of DOA lines from the arrays [5–7]. However, when multiple sources are active, a key problem is that each array transmits the multiple DOA estimates and the central processing node receiving these DOA estimates cannot know to which source they belong. The correct association of DOAs 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*. Also, with multiple active sources, some arrays might underestimate their number, especially when the sources are close together [8, 9]. Thus, the DOAs of some sources from some arrays may be missing. This problem of missed detections is an important aspect which—to the best of our knowledge—has not been widely examined in localization.

Some approaches to data-association rely only on the DOA estimates from the arrays: the work in [10] is based on clustering the intersections of DOA lines from sensor pairs, but is suitable only for scenarios with no missed detections, while the work in [11, 12] considers only noiseless scenarios with perfect DOA estimates. The method in [13] incorporates the data-association to the localization stage by designing a non-linear location estimator that is applied to every possible DOA combination from the sensors and is expected to converge only to the locations of the real sources. However, as shown in [9], the performance of [13] severely degrades in the presence of missed detections. Recently, we proposed the Grid-based (GB) method [9, 14] for localizing multiple sources from DOA estimates, which estimates a location for every possible DOA combination and then decides which of the estimated locations correspond to real sources based on the location estimates and their corresponding DOAs. However, its performance still degrades when the number of arrays that exhibit missed detections increases. The method of [15] utilizes additional information, apart from the DOA estimates: each array computes binary masks in the frequency domain for source separation. The association of DOAs to the sources is found by comparing these binary masks. However, the method does not consider missed detections and works only for the limiting case of two arrays.

In this paper, we examine the data-association and localization problem focusing on more realistic scenarios with missed detections. In our approach, each array estimates features—in addition to the DOAs—for every source that it can detect. By comparing these features between the arrays, the correct association of DOAs can be found. We propose the use of a feature that describes how the frequencies of the captured signals in each array are distributed to the sources. We also propose an algorithm for associating those features. In contrast to [15] our association algorithm can handle multiple arrays. As our results in Section 4 indicate, our algorithm offers high association accuracy, our proposed features exhibit robustness to missed detections, while the amount of information that needs to be transmitted remains at low levels.

## 2. PROBLEM DEFINITION AND ASSUMPTIONS

We consider a WASN whose $M$ nodes are each equipped with a microphone array, as depicted in Fig. 1. We assume that $K$ sources are active ($K$ is assumed to be known). Missed detections can occur, e.g., when the angular separation of some sources is small. Thus, each array can detect *up to* $K$ sources and the number of sources each array detects can vary across the arrays. We also assume that each active source is detected by at least one array, which is a necessary condition to find a DOA association for all $K$ sources. Each array, locally estimates a DOA and a feature for each source it detects
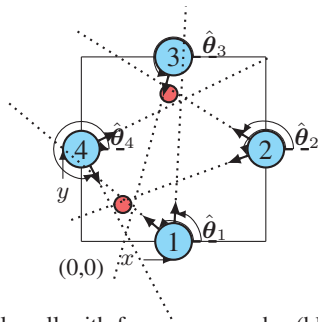
**Fig. 1**. Example cell with four sensor nodes (blue circles), and the estimated DOAs ($\hat{\theta}_1$–$\hat{\theta}_4$) to two sources (red circles).

---

**Algorithm 1** Association Algorithm

**Input**: Features $\mathbf{F}$ , Number of Sources $K$
**Output**: Assignment $\mathbf{G}$

$\quad \mathbf{S} \leftarrow \bigcup_i CreateInitialAssignment(\mathbf{F}_i, K)$ (Fig. 2(a))
$\quad$ **while** $|\mathbf{S}| > 1$ **do** (Fig. 2(b))
$\quad\quad (i, j) \leftarrow \arg\min_{i,j} Score(Merge(\mathbf{S}_i, \mathbf{S}_j))$
$\quad\quad \mathbf{S} \leftarrow \mathbf{S} \setminus (\mathbf{S}_i \cup \mathbf{S}_j) \cup Merge(\mathbf{S}_i, \mathbf{S}_j)$
$\quad$ **end while**
$\quad \mathbf{G} \leftarrow GetFinalAssignment(\mathbf{S})$ (Fig. 2(c))
$\quad$ **while** $\min_{p,q} Score(\mathbf{G}[F_p^i \leftrightarrow F_q^j]) < Score(\mathbf{G})$ **do**
$\quad\quad \mathbf{G} \leftarrow \mathbf{G}[F_p^i \leftrightarrow F_q^j]$
$\quad$ **end while**

---

and transmits them to the central processing node. The central node runs the association algorithm—based on the extracted features—to find the DOA combination from the arrays that correspond to the same source. When this association is found, a single source location estimator can be applied to the DOAs of each source.

## 3. PROPOSED ASSOCIATION METHOD

### 3.1. In-node processing and feature extraction

Apart from the DOAs, each array estimates an association feature for every source it can detect. Our proposed features describe how the frequencies of the captured signals are distributed to the detected sources. This is found by comparing the DOA estimate obtained in each frequency at a given time frame to the final DOAs of the sources that are estimated at that frame. The microphone signals in each (say the $m$th) array are transformed into the Short-Time Fourier Transform (STFT) domain, resulting in the signals $X_{m,i}(\tau, \omega)$ where $i$ is the microphone index, and $\tau$ and $\omega$ denote the time frame and frequency index, respectively. We also denote as $(\tau, \Omega)$ the set of frequencies $\omega$ for frame $\tau$ up to a maximum frequency $\omega_{\max}$. In the remainder, we omit $\tau$, as the procedure is repeated in each frame. Our method starts by estimating a DOA in each frequency $\omega \in \Omega$ resulting in the DOA estimates $\phi(\Omega)$. For DOA estimation in each frequency we use the method of [16], although any method for narrowband DOA estimation can be applied. We set $\omega_{\max}$ to the spatial-aliasing frequency above which ambiguous DOA estimates occur.

To estimate the DOAs of the sources for each frame, we use the method of [8]. This method is based on finding the number of sources the array is able to detect and their DOAs through a matching-pursuit algorithm applied on the histogram of DOA estimates obtained using all frequencies in $\Omega$. The output is the number of detected sources $K_m$ and their DOAs $\boldsymbol{\theta}_m = [\theta_1, \cdots, \theta_{K_m}]$ for time frame $\tau$. Then, the frequencies in $\Omega$ are assigned to the detected sources according to the following rule: The frequency point $\omega \in \Omega$ is assigned to source $p$ if the following conditions are met:

$$A(\phi(\omega), \theta_p) < A(\phi(\omega), \theta_q), \quad \forall q \neq p, \tag{1}$$

$$A(\phi(\omega), \theta_p) < \epsilon, \tag{2}$$

where $A(X, Y)$ denotes an angular distance function that returns the difference between $X$ and $Y$ in the range of $[0, \pi]$ [9]. Eqs. (1) and (2) suggest that a frequency is assigned to the source whose DOA is nearest to the DOA estimated in this frequency, as long as their distance does not exceed a predefined threshold $\epsilon$. If one of the conditions is not met, then the DOA estimate in this frequency is considered erroneous and is not assigned to any of the sources.

Using the frequencies in $\Omega$ for the current and $B$ previous frames, a histogram is formed for each detected source $p$ that counts

how many times each frequency was assigned to source at direction $\theta_p$. These histograms constitute the proposed association features which, along with the estimated DOAs $\boldsymbol{\theta}_m$, are transmitted to the central node for each time frame $\tau$. As all arrays record the same signal—albeit with relative phase differences—the extracted histogram features across the arrays that belong to the same source are expected to be "similar".

### 3.2. Data-association algorithm

Let $\mathbf{F}$ denote the set of all association features and $\mathbf{F}_m$ denote the set of features $F_{m,k}, k = 1, \ldots, K_m$ for all detected sources from the $m$th array. The problem is to find a set $\mathbf{G}$ that contains $K$ groups of features, denoted as $G_i, i = 1, \ldots, K, G_i \in \mathbf{G}$, such that: (a) features from the same array cannot be assigned to the same group, (b) each feature must be assigned to exactly one group, and (c) all groups contain features that are "similar" to each other. We call the set $\mathbf{G}$ an *assignment* of features to groups. As each feature corresponds to a source's DOA, the resulted $K$ groups provide the association of DOAs across the arrays for the $K$ sources. We proceed by proposing and defining a way to measure the quality of an assignment.

Let $D$ be a function measuring the dissimilarity of two features, taking values in $[0, 1]$. We define the score of each group $G_i$ as the maximum pairwise dissimilarity of its contained features:

$$Score(G_i) = \max_{p,q} D(F_p^i, F_q^i), \tag{3}$$

where $F_p^i$ denotes the $p$th feature of group $G_i$. Next, we define the overall score of an assignment $\mathbf{G}$ as the maximum score among the scores of its contained groups $G_i \in \mathbf{G}$ :

$$Score(\mathbf{G}) = \max_i Score(G_i). \tag{4}$$

Our goal is to find an assignment $\mathbf{G}$ that minimizes (4) and satisfies constraints (a) and (b) mentioned above. Formally, it is defined as:

$$\arg\min_{\mathbf{G}} Score(\mathbf{G}). \tag{5}$$

To compare two assignments with the same score, we compare the scores of their groups in descending order up to their maximum non-equal value. The intuition behind the above formulation is that we want to find an assignment where *all groups* contain features that are as similar as possible to each other. We propose a greedy algorithm to find such an assignment. It does not necessarily identify the optimal solution, but it is simple, fast, and finds good solutions in practice. The algorithm—shown in Algorithm 1—works as follows.

First, an assignment of features to $K$ groups is created for each set $\mathbf{F}_m, m = 1, \cdots, M$, where each group contains a single feature. If the number of features for an array is less than $K$, some
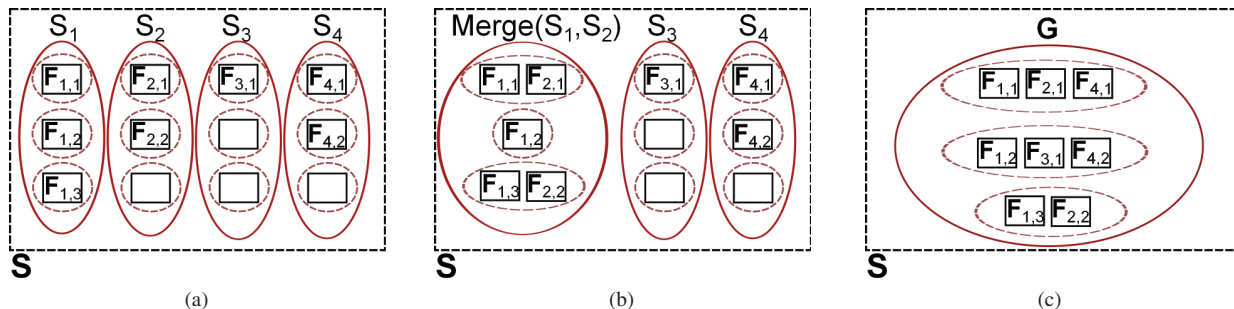
**Fig. 2**. Example of the association algorithm for $M = 4$ arrays and $K = 3$ sources. (a) First, an assignment of features to $K$ groups is created for the set of features $\mathbf{F}_m$ for each array. The empty boxes represent the empty groups, as the corresponding arrays have detected less than $K$ sources. (b) The algorithm finds the assignment (array 1 and 2 in this example) that, when merged, produce the best score according to (5) and merges them. (c) The merging operations stop when a single assignment $\mathbf{G}$ remains.

groups are left empty (Fig. 2(a)). Then, the algorithm tries to greedily merge those assignments, until only one remains (Fig. 2(b)). This is done by selecting two assignments that, when merged, produce the best possible assignment according to (5). The merging is done by considering all possible ways to merge them, which equals $K!$. This problem is known as the *Linear Bottleneck Assignment Problem* [17] which can be solved efficiently in polynomial time. In case $K$ is relatively small, a brute-force approach is often faster. When this step is finished, a single assignment $\mathbf{G}$ remains (Fig. 2(c)).

Finally, we perform a second greedy step to further refine this assignment. This step selects two features $F_p^i$ and $F_q^j$ from different groups and tries to swap them in order to further reduce the score of $\mathbf{G}$; for brevity, we use $\mathbf{G}[F_p^i \leftrightarrow F_q^j]$ to refer to the new assignment in which $F_p^i$ and $F_q^j$ are interchanged. We also allow one of them to be empty or, in other words, to move a feature from one group to another. The algorithm terminates if no such pair exists.

## 4. RESULTS AND DISCUSSION

We performed simulations on a cell of a WASN with dimensions of $V = 4$ m, with four nodes arranged according to the setup in Fig. 1. Each node 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 2 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.

We used the Image-Source method [18] 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 two and three simultaneously active sources. Each simulation was repeated 30 times and the sources were located within the cell with independent uniform probability. For processing, we used frames of 2048 samples with 50% overlap. The FFT size was set to 2048 and $\omega_{\max} = 4$ kHz which is the spatial-aliasing frequency for the given array geometry. The same frequency range was also used for the method of [15] that we used for comparison purposes. For our method we set $\epsilon = 10°$ and

$B = 5$ previous frames. Finally, to measure the dissimilarity of our proposed histogram features in (3) we use the Pearson Correlation Coefficient Distance defined as [19]:

$$D(F_p, F_q) = \frac{1 - r_{F_p, F_q}}{2}, \tag{6}$$

which takes values in the range of $[0, 1]$, where $r_{F_p, F_q}$ is the Pearson correlation coefficient between histogram features $F_p$ and $F_q$.

### 4.1. Evaluation metrics

To evaluate the association accuracy we use two metrics. The first (denoted as *Metric 1*) counts the percentage of frames where the correct association was found. We say that the association for a frame is correct if *all* sources in that frame are assigned the correct DOAs from *all* the arrays. When an error occurs, it means that some DOAs from some arrays are assigned to an erroneous source. As an example consider that in Fig. 2(c) the DOA that corresponds to feature $F_{4,1}$ was erroneously assigned to the source that corresponds to the first group. While this association is erroneous—according to Metric 1—there are pairs of DOAs from arrays that are associated correctly, such as the pairs $(F_{1,1}, F_{2,1})$, $(F_{1,2}, F_{3,1})$, $(F_{3,1}, F_{4,2})$, $(F_{1,2}, F_{4,2})$, and so on, while other pairs are associated erroneously, such as pairs $(F_{1,1}, F_{4,1})$, $(F_{2,1}, F_{4,1})$. The more the DOA pairs that are associated correctly, the lower the impact of the erroneous association is to the localization error. For this reason, our second metric (denoted as *Metric 2*) counts the percentage of correct pairwise associations between all pairs of arrays quantifying better how many parts of an association are correct.

### 4.2. Robustness to missed detections

First, we evaluate the efficiency of our method in handling missed detections. We compare the proposed features for data-association with the ones proposed in [15], which we modified to work with a circular array. As the association algorithm of [15] can only work for two arrays, we use our proposed association algorithm on the features (i.e., estimated source separation binary masks) extracted according to [15]. In this simulation, we assume the DOAs of the sources, i.e., vectors $\boldsymbol{\theta}_m$, at each array are known. To simulate missed detections, we define $C_s$ as the number of arrays that detected $s$ sources. We fix $C_s$ and remove some DOAs from some arrays until the desired value of $C_s$ is reached. The sources whose DOAs are removed as well as the arrays that exhibit the missed de-
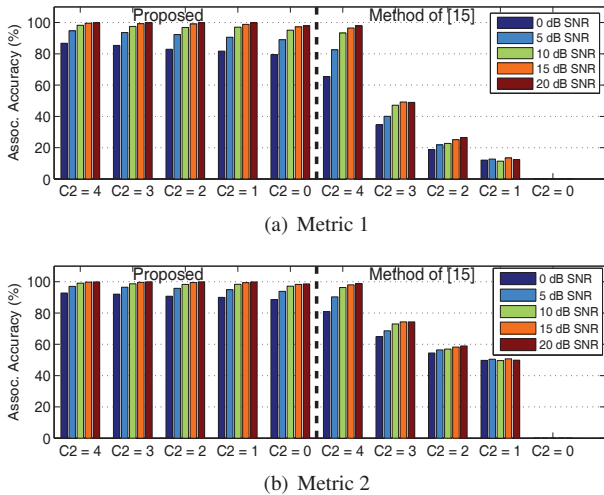
(a) Metric 1



(b) Metric 2

**Fig. 3**. Data association accuracy for two sources in an anechoic environment for different values of SNR and $C_2$.



(a) Two sources



(b) Three sources

**Fig. 4**. Data association accuracy of the proposed method for two and three sources in different reverberation and SNR scenarios.

tection are selected at random in every frame under the constraint that each source must be detected by at least one array (Section 2).

Fig. 3 shows the association accuracy in an anechoic environment with two sources and different SNR cases, for all values of $C_2$, i.e., the number of arrays that detected two sources. From Fig. 3 it is clear that our approach is robust to missed detections, achieving more than 80% accuracy with both metrics for all SNR and $C_2$ cases. The accuracy of association using the features from [15] when missed detections are present ($C_2 < 4$) is severely degraded, showing the method's inability to handle missed detections. When a source is not detected, the method of [15] erroneously assigns its frequencies to the other sources, degrading the association performance. In our method, we avoid such erroneous assignments through the use of (2). In the case of no missed detections ($C_2 = 4$) our proposed features are more robust to noise offering improved accuracy, especially at low SNR, compared to the features from [15].

Finally, the features of [15] cannot be used when $C_2 = 0$ (the corresponding area in Fig. 3 is empty). When all arrays detect one source—which is not necessarily the same—these features do not provide useful information for association as each array will estimate a binary mask with the value of one for every frequency. However, our proposed association features are able to discriminate between the sources providing high association accuracy even in this extreme case where all the arrays detected one source.

### 4.3. Data-association accuracy

We evaluate the association accuracy of our proposed method in a more practical setting where the sources' DOAs are estimated per frame as described in Section 3.1. The association accuracy for different SNR values and reverberation times, namely, anechoic, $T_{60} = 250$ ms, and $T_{60} = 400$ ms is shown in Fig. 4 for two and three simultaneously active sources. As expected, the performance degrades with increasing reverberation time, especially in the three sources case. However, note that while the overall association accuracy (Metric 1) is reduced, most of the DOAs between pairs of arrays are associated correctly (Metric 2) in all reverberation and SNR scenarios. This indicates that while association errors in frames occur more often, most DOA pairs are associated correctly.
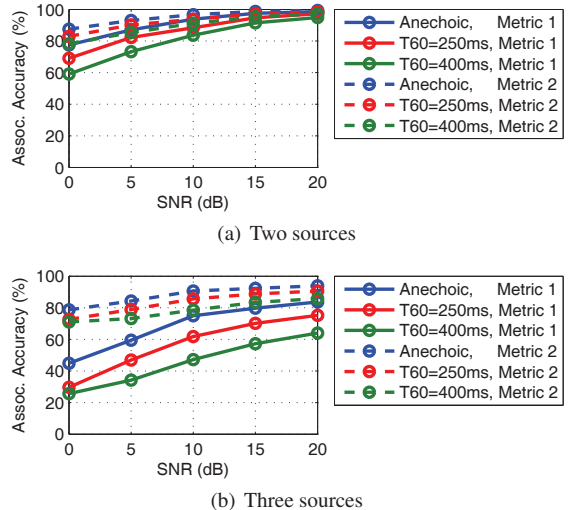
Note that in these results the values of $C_2$ and $C_3$ vary in each frame, in contrast to the results in Section 4.2 where the number of $C_s$ was fixed. These values depend on how many sources the DOA estimation in each array can detect. We observed that for the two sources case in approximately only 16% of the frames all four arrays detected two sources ($C_2 = 4$), in 24% of the frames $C_2 = 3$, in 35% of the frames $C_2 = 2$, in 15% of the frames $C_2 = 1$, and in 10% of the frames none of the arrays detected two sources. The values were approximately constant for all SNR cases. The problem of missed detections is more evident in the case of three sources, where in 95% of the frames $C_3$ was either zero or one and with the value of $C_2$ being either two or three for approximately 72% of the frames in all SNR cases. These numbers not only reveal the efficiency of our method, but also show how often missed detections can occur in practice, highlighting the need for a method that can handle missed detections to accurately localize the sources.

### 4.4. Localization experiments

Since the final goal of any data-association algorithm is to improve the localization accuracy, in the section we evaluate our proposed method in terms of localization error and compare it again with the use of association features extracted from [15]. Our previously proposed grid-based (GB) method for multiple sources [9], which uses no additional information apart from the DOA estimates, is also considered. For the method of [15] and our proposed method, the single source grid-based method from [9] is applied to localize the resulted DOA associations. The localization performance was measured in terms of the root-mean square error (RMSE) over all sources, all 30 different source configurations, and for all frames where each source was detected by at least one array. Fig. 5 depicts the localization performance for two and three simultaneously active sources in reverberant environments with reverberation time $T_{60} = 250$ ms and $T_{60} = 400$ ms. The localization error using the estimated DOAs but assuming that the correct association of DOAs to the sources is known (denoted as Perfect Association), is also included to represent the best-case scenario. It can be observed, that the proposed method outperforms the others providing location estimates close to
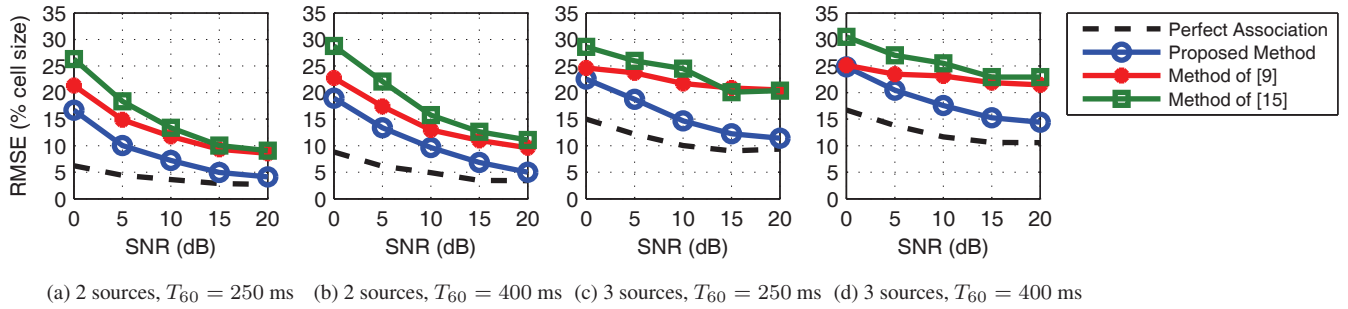
(a) 2 sources, $T_{60} = 250$ ms   (b) 2 sources, $T_{60} = 400$ ms   (c) 3 sources, $T_{60} = 250$ ms   (d) 3 sources, $T_{60} = 400$ ms

**Fig. 5**. Localization error as a percentage of cell size $V$ for different reverberation scenarios and number of active sources.

the best-case, especially at higher SNR values, for both reverberation scenarios and for both two and three sources. The other two methods perform worse than the proposed for all cases, with the performance degradation being more evident in the three sources case. The method of [15] performs worse or equally well as the GB method for multiple sources [9]. However, the method of [9] uses no additional information apart from the DOA estimates in contrast to [15].

### 4.5. Transmission requirements

Finally, we quantify the transmission requirements. Apart from the DOA estimates, each array $m$ has to transmit the association features. A simple and efficient scheme to transmit this information is to encode the DOA index that each frequency was assigned to. Given $K_m$ estimated sources, each frequency will belong to one of the sources or it will be considered erroneous. The central processing node can then construct the association features using the transmitted frequency assignment for the current and $B$ previous frames. Thus, for each frequency $\lceil \log_2(K_m+1) \rceil$ bits are required to encode the DOA indices, with $\lceil \cdot \rceil$ denoting the ceiling operator. A similar encoding scheme can be used for the method of [15], where each frequency requires $\lceil \log_2(K_m) \rceil$. Thus, our proposed method results in improved data-association and localization accuracy, while its transmission requirements remain at low levels.

## 5. CONCLUSIONS

We considered the data-association problem that occurs when localizing multiple sound sources in a WASN using DOA estimates from multiple microphone arrays. We presented a data-association approach that can associate the DOAs from the different arrays to the correct sound sources with high accuracy. Our approach was based on the construction of association features for each detected source in each array, which proved to be robust to scenarios with missed detections that occur very often in practice. Using simulations and comparisons with other methods, we confirmed the efficiency of our proposed method to correctly associate the sound sources with their corresponding DOAs and significantly improve localization performance, while keeping transmission requirements at low levels.

### 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] D. J. Mennill, M. Battiston, D. R. Wilson, J. R. Foote, and S. M. Doucet, "Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour," *Methods in Ecology and Evolution*, vol. 3, no. 4, pp. 704–712, 2012.

[3] H. Do and H.F. 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.

[4] P. Aarabi, "The fusion of distributed microphone arrays for sound localization," *EURASIP Journal of Applied Signal Processing*, vol. 2003, pp. 338–347, January 2003.

[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, pp. 817–828, 1992.

[6] Z. Wang, J. Luo, and X. Zhang, "A novel location-penalized maximum likelihood estimator for bearing-only target localization," *IEEE Trans. on Signal Processing*, vol. 60, no. 12, pp. 6166–6181, 2012.

[7] K. Doğançay, "Bearings-only target localization using total least squares," *Sig. Proc.*, vol. 85, no. 9, pp. 1695–1710, 2005.

[8] 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 Language Processing*, vol. 21, no. 10, pp. 2193–2206, 2013.

[9] A. Griffin, A. Alexandridis, D. Pavlidi, Y. Mastorakis, and A. Mouchtaris, "Localizing multiple audio sources in a wireless acoustic sensor network," *Sig. Proc.*, vol. 107, no. 0, pp. 54–67, 2015.

[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 MILCOM*, Nov 2008, pp. 1–7.

[11] A. Bishop and P. Pathirana, "A discussion on passive location discovery in emitter networks using angle-only measurements," in *IWCMC*. 2006, ACM.

[12] A. Bishop and P. Pathirana, "Localization of emitters via the intersection of bearing lines: A ghost elimination approach," *IEEE Trans. on Vehicular Technology*, vol. 56, no. 5, pp. 3106–3110, Sept 2007.

[13] 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, pp. 40–51.

[14] 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*, 2014.

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

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

[17] G. Carpaneto and P. Toth, "Algorithm for the solution of the bottleneck assignment problem," *Computing*, vol. 27, no. 2, pp. 179–187, 1981.

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

[19] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*, Academic Press, 4th edition, 2008.