

Multiple source location estimation on a dataset of real recordings in a wireless acoustic sensor network

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

*Institute of Computer Science, Foundation for Research and Technology-Hellas, Heraklion 70013, Greece,
Department of Computer Science, University of Crete, Heraklion 70013, Greece
e-mail: analexan@ics.forth.gr, mouchtar@ics.forth.gr

[†]High Performance Computing Research Laboratory, Electrical and Electronic Engineering Department,
School of Engineering, Computer and Mathematical Sciences,
Auckland University of Technology, Auckland 1024, New Zealand
e-mail: anthony.griffin@aut.ac.nz

Abstract—Recently, wireless acoustic sensor networks (WASNs) have received significant attention from the research community and a variety of methods have been proposed for numerous applications, such as location estimation and speech enhancement. The lack of publicly available datasets with signals recorded in WASNs, presents difficulties in obtaining consistent performance indicators across the different approaches. In this paper, we present and release a dataset of real recorded signals in an outdoor WASN comprised of four microphone arrays. Our dataset consists of several speakers recorded at various locations within the WASN and can be used for benchmarking purposes. We also present location estimation results using our real recorded dataset. Our results can serve as a baseline indicator of localization performance of single and multiple sources in a real environment.

I. INTRODUCTION

In the last few decades, microphone arrays have received significant attention from the audio signal processing community. Their ability to perform signal enhancement based on spatial features from the sound source locations have made them the typical interface for robust signal acquisition in adverse environments. Recently, a new paradigm has emerged that further extends the abilities of traditional microphone arrays. The research community started to investigate configurations where multiple microphone arrays are present in the acoustic environment, forming a sensor network of acoustic nodes which is often referred to as a wireless acoustic sensor network (WASN) [1]. The nodes feature wireless communication capabilities in order to allow information exchange in the network and have processing power in order to locally perform operations on the captured signals.

Compared to traditional microphone arrays, WASNs can provide improved sampling of the acoustic environment, since the multiple acoustic nodes that are distributed over the monitored area increase the probability of finding a microphone that is close to the source of interest [1]. Their potential has been explored with promising results in numerous applications, such as DOA estimation [2], estimation of the exact location of the

acoustic sources [3]–[7], wildlife monitoring [8], and speech enhancement and beamforming [9]–[11].

However, so far, the evaluation of signal processing methods tailored for WASNs has been based on simulations or on real recorded signals which are rarely made publicly available. Moreover, the many parameters that can be configured differently, such as the number of microphone arrays, the geometry of each array and the array locations, makes it difficult to obtain a direct comparison between different methods proposed in the literature. The availability of datasets with signals recorded in WASNs would enable the comparison and benchmarking between a variety of methods for DOA estimation, location estimation, and signal enhancement in WASNs.

In this paper, we make an attempt to collect and release such a dataset¹. We consider an outdoor WASN cell comprised of four circular microphone arrays and we record several speakers at different spatial locations within the cell. Recently, we also presented another small-scale dataset of real recordings [12] in a two-node WASN where the nodes were circular microphone arrays. The dataset consisted of speech recordings made by two speakers at distinct locations when the microphone arrays were placed very close to the walls of a room and its purpose was to test localization performance in this challenging setup. The current dataset differs in the sense that it is a more general-purpose one, it consists of more speakers at more locations, and it is recorded in a larger WASN comprised of four nodes. We also present localization results on this dataset for scenarios of single and multiple sources. For localization we use our previously proposed method of [4], which infers the locations of the sources based on direction of arrival (DOA) estimates transmitted by the microphone arrays. Our results can serve as a baseline indicator of performance for the task of location estimation of acoustic sources.

II. THE DATASET

The dataset was recorded in an outdoor WASN square cell of dimensions of 400×400 cm with 4 sensor nodes

This research has been partly funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 644283, Project LISTEN.

¹The dataset and its complete documentation can be found at <https://github.com/spl-icsforth/OutdoorWASNRecordings>

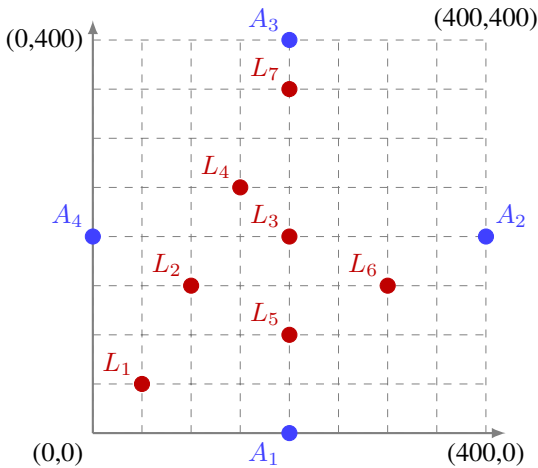


Fig. 1. Layout of the recording setup, where A_1 to A_4 are the microphone arrays, and L_1 to L_7 are the locations of the speakers. The coordinates are shown in cm.

that consist of uniform circular microphone arrays. Several speakers were recorded in seven distinct locations within the cell. Figure 1 depicts the recording configuration, i.e., the sensors’ and speakers’ locations. The duration of the utterances range from 8 to 12 seconds. The speakers read sentences from a Greek newspaper. The sentences were recorded with a microphone in a quiet office environment and then played back through a loudspeaker which was placed at the seven locations shown in Figure 1. The loudspeaker was pointed at the sky in order to give the sources a directivity pattern as close to omnidirectional as possible. Since the dataset was recorded outdoors, the signals do not have many reflections. However, they do suffer from various noise sources from the environment such as wind, cars passing by, and bird vocalizations.

The locations were measured as the x - and y -coordinates, while the microphone arrays and the loudspeaker were placed at approximately the same height. At each location, five different speakers (3 male, 2 female) were recorded, except for location L_1 and L_2 where four speakers were recorded. Table I reports the x - and y -coordinates of the speaker locations and the speaker IDs that were recorded at each location.

The microphone arrays used for recording are of two different types: two uniform circular arrays with 5 cm radius and two uniform circular arrays with 3 cm radius. All the arrays had 8 microphones. Table II summarizes the geometry and locations of the microphone arrays that comprise the WASN: A_2 and A_3 were comprised of analog Shure SM93 omnidirectional microphones. A_2 was connected to an M-Audio M-Track Eight sound card and A_3 was connected to a TASCAM US-2000 sound card. A_1 and A_4 contained digital MEMS omnidirectional microphones [13]. They were connected to the host PC using a USB cable thus not requiring an external multichannel sound card. The analog microphone arrays (A_2 and A_3) recorded signals at a sampling frequency

TABLE I
LOCATION COORDINATES AND IDS OF SPEAKERS THAT WERE RECORDED AT EACH LOCATION

Location ID	(x, y) coordinates in cm.	Speaker IDs (Bold=female)
L_1	(50, 50)	S01, S03, S04 , S05
L_2	(100, 150)	S02 , S03, S04 , S05
L_3	(200, 200)	S01, S02 , S03, S04 , S05
L_4	(150, 250)	S01, S02 , S03, S04 , S05
L_5	(200, 100)	S01, S02 , S03, S04 , S05
L_6	(300, 150)	S01, S02 , S03, S04 , S05
L_7	(200, 350)	S01, S02 , S03, S04 , S05

TABLE II
MICROPHONE ARRAY DETAILS

Array	coordinates (x, y) in cm	radius	mic. type	soundcard
A_1	(200, 0)	3 cm	MEMS	integrated in array [13]
A_2	(400, 200)	5 cm	analog	M-Audio M-Track Eight
A_3	(200, 400)	5 cm	analog	TASCAM US-2000
A_4	(0, 200)	3 cm	MEMS	integrated in array [13]

of 44.1 kHz, while the digital ones (A_1 and A_4) operated at 48 kHz. Thus, after recording, the signals from arrays A_1 and A_4 were downsampled to 44.1 kHz in order to keep the sampling frequency constant at all nodes in the WASN.

Each microphone array was connected to a different PC and provided sample-synchronized 8-channel signals. The channel to microphone correspondence is shown in Figure 2 which depicts the microphone placement of an array with respect to its center. In order to make all arrays start recording at the same time, we implemented a client-server approach: a central node notifies the four PCs (clients) when to start recording by sending a special packet through the network. When this packet was received by the clients they started to record the captured signals. At the start of each recording, we played back a sinusoidal signal through the loudspeaker that was used to further synchronize the signals among the different arrays. During processing, the start of the sinusoidal signal was detected by means of cross-correlation of the recorded signals with the known sinusoidal pulse. Since the duration of the pulse and the duration of the speakers’ signals were fixed, utterances were then easily segmented. Thus, in general, the signals between different arrays are expected to be synchronized to within a few milliseconds.

III. LOCALIZATION METHOD

The method used to evaluate the localization performance in this dataset was our previously proposed grid-based (GB) method [4], which belongs to the class of localization approaches using direction of arrival (DOA) estimates from the nodes. This framework assumes the availability of a central processor that performs the localization task. Each microphone array—whose location is known—estimates the DOAs of each source it can detect at each time instant and transmits the DOA estimates to the central processor that fuses them together

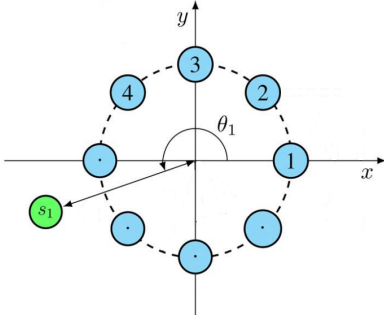


Fig. 2. Microphone array configuration and the coordinate system for DOA estimation. The numbers in the blue circles denote the channel number of the 8-channel recording.

in order to infer the locations of the sources. Note that, this approach results in reduced bandwidth usage as only the DOA estimates need to be transmitted within the network. Moreover, it does not require the arrays to be perfectly synchronized [3]. In this section, we briefly outline the grid-based method for a single and multiple sources.

A. Grid-based method for a single-source

Let $\hat{\theta}$ denote the $M \times 1$ vector of estimated DOAs from the M nodes (i.e., microphone arrays) at a given time instant. The method is based on constructing a grid with N grid points over the localization area. The resolution of the grid is determined by the grid point spacing G . For each grid point n , we calculate $\psi_{m,n}$ which denotes the DOA of the n -th grid point with respect to node m and is given by:

$$\psi_{m,n} = \arctan \frac{g_{y,n} - q_{y,m}}{g_{x,n} - q_{x,m}}, \quad (1)$$

where $\arctan(\cdot)$ denotes the four quadrant arctangent function that returns an angle in the range of $[0, 2\pi)$, $(g_{x,n}, g_{y,n})$ are the x - and y -coordinates of the n -th grid point and $(q_{x,m}, q_{y,m})$ are the x - and y -coordinates of the m -th microphone array which are assumed to be known. The location of the source is estimated as the grid point n^* whose DOAs most closely match the estimated DOAs by solving:

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

where $\hat{\theta}_m$ denotes the DOA from the m -th microphone array and $A(\cdot, \cdot)$ denotes an angular distance function that returns the difference between two angles in the range of $[0, \pi)$ [4].

In order to solve (2) efficiently without testing all possible grid points and still attain a high resolution in the grid (small grid point spacing) we employ an iterative approach which starts with a coarse grid (i.e, low N and large grid point spacing) over the localization area and once the best grid point is found—according to (2)—a new grid centered on this grid point is generated with a smaller spacing between the grid points but also a smaller scope. Then, the best grid point in this new grid is found and the procedure is repeated until

we reach the desired grid resolution. For more details on this iterative procedure the reader is referred to [4].

B. Grid-based method for multiple sources

For multiple sources, each microphone array estimates multiple DOAs (one for each detected source) and transmits them to the central processor. A key problem when localizing multiple sources using DOA estimates is that the central processor that receives the multiple DOAs from the arrays cannot know to which source they belong. Moreover, it is possible that some arrays may not be able to detect all of the sources, thus some DOA estimates from some arrays may be missing. Such missed detections can occur for several reasons, such as due to a challenging environment in terms of reverberation and noise, or because some sources are close together in terms of their angular separation with respect to some arrays. In fact, our previous work on localization of multiple sources using DOA estimates has shown that such missed detections occur quite often [5].

To localize multiple sources, the grid-based method starts by enumerating all possible unique combinations of DOA estimates from the nodes and estimating an initial location for each combination. Taking into account missed detections, the number of unique combinations is given by:

$$N_{\text{comb}} = \prod_{s=1}^S s^{C_s} \quad (3)$$

where S is the total number of sources and C_s denotes the number of microphone arrays that detected s sources.

For each DOA combination, the initial location is estimated by applying the single-source grid-based method to the DOAs of the combination. Let j enumerate the combinations and let $\hat{\theta}^{(j)}$ denote the $M \times 1$ vector of DOAs that belong to the j -th combination and were used to estimate the j -th initial location $\mathbf{p}^{(j)}$. Moreover, let $\hat{\theta}_m^{(j)}$ denote the DOA of node m in the j -th combination.

The grid-based method for multiple sources employs two different approaches (Brute-force and Sequential) [4] that test all initial location estimates and decide which of them correspond to the final location estimates. The two approaches exhibit quite similar performance with the Brute-force approach being much more computationally demanding. Thus in this paper we use the Sequential approach which is much more computationally efficient.

For each DOA combination, the Sequential approach computes the residual:

$$r_j = \sum_{m=1}^M \left[A \left(\hat{\theta}_m^{(j)}, \theta_m(\mathbf{p}^{(j)}) \right) \right]^2 \quad (4)$$

where $\theta_m(\mathbf{p})$ is a function that computes the DOA with respect to node m for location \mathbf{p} and can be implemented based on (1). The residual denotes the sum of squared distances between the DOAs of combination j and the DOAs that correspond to the estimated initial location that was found using the DOAs of the j -th combination. The final location estimates are chosen

TABLE III
DOA OFFSETS THAT CORRECT UNINTENDED ORIENTATION ERRORS IN
THE POSITIONING OF THE MICROPHONE ARRAYS

Array ID	offset
A_1	10.02°
A_2	6.46°
A_3	-1.11°
A_4	4.27°

as the locations that result from the DOA combinations with the minimum residual. The reader is referred to [4] for more implementation details of this approach.

IV. RESULTS AND DISCUSSION

We now present our localization results using our real recorded dataset and our grid-based method for one, two, and three simultaneously active sound sources.

A. DOA estimation

For DOA estimation we employ our previously proposed method of [14]. We measure the DOA with respect to the array center, as shown in Figure 2 for source s_1 . The method requires the setting of a parameter related to the maximum sources it can detect. Assuming that the number of sources is known, we set the maximum number of sources to the true number of sources for each scenario of one, two, and three active sources. However, while overestimation of the number of sources is avoided, some sources may not be detected. All other parameters were set according to [14]. For processing, we used a frame size of 2048 samples with 50% overlap, thus providing DOA estimates every 23 ms. The DOA estimation method requires one second of initialization. As a result, DOA estimates—and thus location estimates—are provided after the first second of each recording.

B. Calibration procedure

Since the orientation of the microphone arrays had some unintended errors due to the positioning of the arrays in the real setup, we perform a calibration procedure in order to correct such errors in the DOA estimates. For calibration we use location L_3 which is equidistant from all arrays and thus DOA estimation is expected to give similar performance for all microphone arrays. For each array, we perform single-source DOA estimation and we take the mean of the DOAs estimated at each time instant. We subtract the mean estimate from the true DOA in order to estimate an offset which corresponds to the unintended DOA error that occurs from the positioning of the array. The procedure was repeated using the signals from four different speakers and a mean offset was estimated from the offsets found from each individual speaker. Very little deviation was observed in the offsets of the individual speakers, with their difference being less than half a degree. The offsets are given in Table III. To correct the DOA estimates, we add the offset of each array to the final DOA estimates obtained from our DOA estimation procedure.

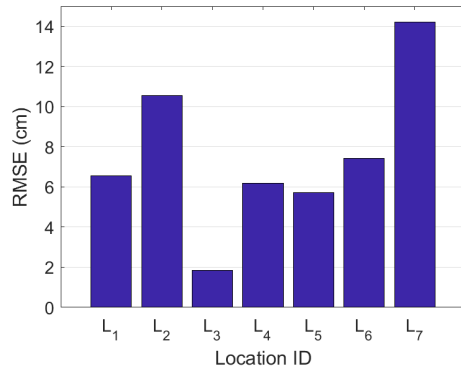


Fig. 3. Location estimation error for each location of our dataset for a single active sound source.

C. Location estimation

The localization area was defined as the square defined from point $(0, 0)$ to point $(400, 400)$ (see Figure 1). For the iterative grid-based method, the resolution of the grid is defined as the spacing between two adjacent grid points. In the first iteration the spacing was set to 50 cm, while in the final iteration the spacing was 1 cm. To generate scenarios with multiple sources, we added the microphone signals of different speakers at different locations.

Based on the array locations, the DOA estimates of each array can take values in specific ranges. For example, the DOAs of A_1 must belong to the first and second quadrants. However, because our DOA estimation method can estimate a DOA in the entire range of $[0, 2\pi)$, we remove DOA estimates that occur outside the valid range for each array.

As an error metric, we use the root mean square error (RMSE) between the estimated and true source locations. For multiple sources, we measure the error of each source from the location estimate that is closest to the source. However, due to missed detections there are cases where our method cannot produce estimates for all the sources (e.g., when all arrays detect one source in a two-sources scenario). In such cases, we use the same location estimate to estimate the error for both of the two sources, assuming that our method produced two estimates at the same location. The same procedure is carried out for the three-sources case: when our method produces two location estimates we duplicate the location estimate that is closest to the third undetected source, while when our method produces one location estimate we use this estimate three times to calculate the error of the three sources.

1) *Single-source case:* For the single source case, we tested all seven locations and at each location we tested all available speakers' signals. Figure 3 depicts the RMSE for each location over all time frames and all speakers at that location. It is clear that our method can accurately localize one source with the maximum error being approximately 14 cm for location L_7 , and that the localization performance varies across different locations. This result was also observed in our previous work with another dataset [12], and highlights the necessity to

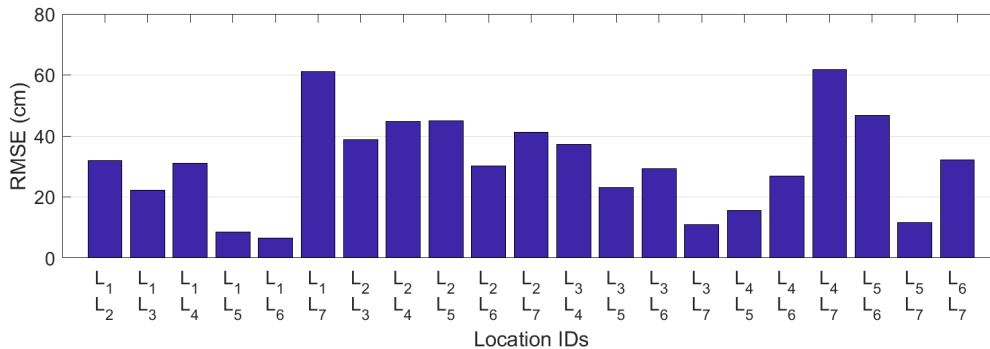


Fig. 4. Location estimation error for each location pair of our dataset for two active sound sources

evaluate localization performance in different locations within the cell. The overall RMSE for this single-source scenario, which was computed using all time frames, all locations, and all speakers, was measured to be 8.32 cm.

2) *Two-sources case*: For our experiments with two sources, we used speakers S03 and S04 and we tested all possible combinations of location pairs. This results in 21 possible location pairs, where S03 was placed at the first location and S04 was placed at the second location. For example, for location pair L_1 & L_2 , speaker S03 was at location L_1 and speaker S04 was at location L_2 . Figure 4 shows the RMSE of the two sources for each location pair. We again observe that the performance varies for each location pair. This is due to the specific setup of the sources, which affects the number of sources each microphone array is able to detect. To quantify missed detections, we measured the percentage of frames where each possible value of C_2 (i.e., the number of arrays that detected two sources) occurs. Overall, by taking into account all location pairs, we observed that in approximately 10% of the frames none of the arrays detected two sources (i.e., $C_2 = 0$), in 15% of the frames $C_2 = 1$, in the majority of frames (approximately 48%) $C_2 = 2$, in 26% of the frames $C_2 = 3$ and in only 1% of the frames all arrays detected two sources (i.e., $C_2 = 4$). Taking into account these numbers—which highlight how often missed detections can occur in practice—our method can perform quite accurate localization with the RMSE over all location pairs being 34.84 cm.

Finally, Figure 5 presents heat maps of the location estimates over all time frames of some representative location pairs. We can observe that Figures 5a & 5b achieve accurate localization, while performance is decreased in Figure 5c where one source is estimated further away from the true source. Finally, in Figures 5d & 5e the location estimates start to concentrate in the middle of the two sources. This occurs because the sources are too close together in terms of their angular separation with respect to almost all arrays and thus the value of C_2 is decreased. Note that in location pair L_3 & L_4 , in all of the frames none of the arrays detected two sources (i.e., $C_2 = 0$).

TABLE IV
OVERALL LOCATION ERRORS FOR THE DATASET

	RMSE
One source	8.32 cm
Two Sources	34.84 cm
Three Sources	55.36 cm

3) *Three-sources case*: Similar to the two-sources case, for our experiments with three sources, we used speakers S03, S04, and S01 and tested all possible combinations of three sources locations (35 different combinations). Due to space limitations, we report the location error in terms of RMSE over all location combinations which was measured to be 55.36 cm. As expected, in this three-sources scenario, missed detections are very prominent: in approximately 58% of the frames none of the arrays detected three sources ($C_3 = 0$), in 33% of the frames only one array detected three sources ($C_3 = 1$) while the value of C_3 was greater than one in only approximately 9% of the frames. Similarly, in approximately 45% of the frames only two arrays detected two sources, with the value of C_2 being greater than two in only 25% of the frames.

Figure 6 presents some heat maps of the location estimates, using all time frames of some representative location combinations for the three sources. We can observe quite accurate localization for the first three location combinations (Figures 6a–6c). However performance degrades in Figure 6d where one source has not been localized, and in Figure 6e, where two sources have been detected as one. Note that in these two last location combinations, in all time frames none of the arrays detected three sources (i.e., $C_3 = 0$). Finally, the overall RMSE for the case of one, two, and three active sound sources is summarized in Table IV.

V. CONCLUSION

In this paper, we presented a dataset of real recorded signals in a wireless acoustic sensor network (WASN). The dataset consists of various speakers recorded at several locations within the WASN and is made publicly available in order to be used for benchmarking purposes of various signal processing methods, such as DOA estimation, location estimation, and

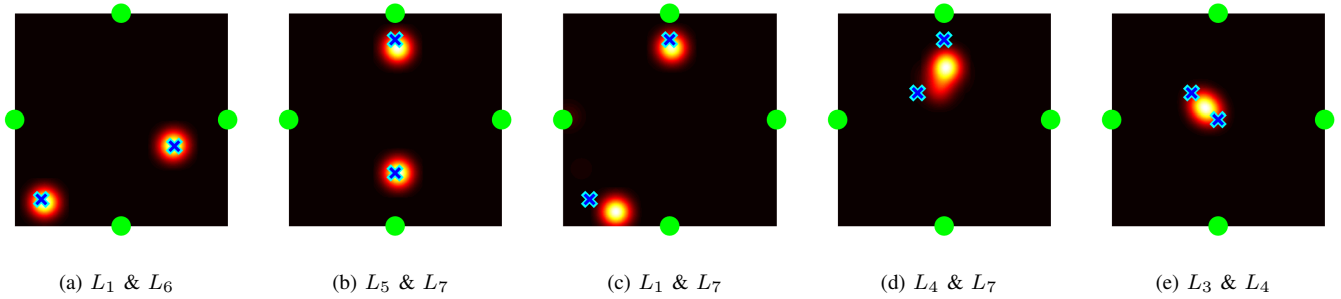


Fig. 5. Heat maps of the location estimates for representative location pairs of two active sources. The blue x's show the true sources' locations, and the green dots show the locations of the microphone arrays.

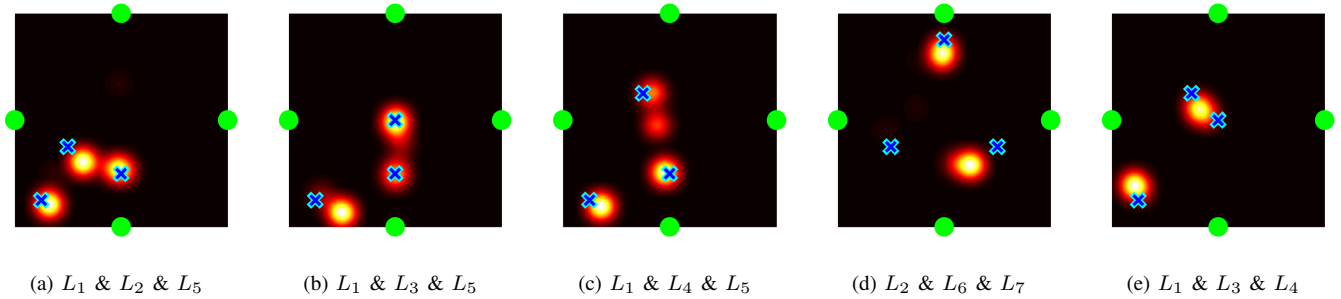


Fig. 6. Heat maps of location estimates for representative location combinations of three active sources. The blue x's show the true sources' locations, and the green dots show the locations of the microphone arrays.

speech enhancement in WASNs. We also presented results of location estimation of single and multiple sources using our previously proposed localization method which is based on direction of arrival estimates transmitted from the microphone arrays that comprise the WASN. Our results can serve as a baseline indicator of localization performance in our recorded dataset.

ACKNOWLEDGMENT

The authors would like to thank Emmanuil Lakiotakis for his valuable contribution during the recording of the dataset.

REFERENCES

- [1] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: a signal processing perspective," in *Proceedings of the 18th IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT)*, IBBT, Ghent, Belgium, Nov 2011.
- [2] A. Hassani, A. Bertrand, and M. Moonen, "Distributed node-specific direction-of-arrival estimation in wireless acoustic sensor networks," in *21st European Signal Processing Conference (EUSIPCO 2013)*, Sept 2013, pp. 1–5.
- [3] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," *Wireless Acoustic Sensor Networks and Applications*, 2017.
- [4] 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.
- [5] A. Alexandridis and A. Mouchtaris, "Multiple sound source location estimation in wireless acoustic sensor networks using DOA estimates: The data-association problem," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 2, pp. 342–356, Feb 2018.
- [6] P. Aarabi, "The fusion of distributed microphone arrays for sound localization," *EURASIP Journal of Applied Signal Processing*, vol. 2003, pp. 338–347, Jan 2003.
- [7] M. Compagnoni, P. Bestagini, F. 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, Sep 2012.
- [8] D. T. Blumstein, D. J. Mennill, P. Clemins, L. Girod, K. Yao, G. Patricelli, J. L. Deppe, A. H. Krakauer, C. Clark, K. A. Cortopassi, S. F. Hanser, B. McCowan, A. M. Ali, and A. N. G. Kirschel, "Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus," *Journal of Applied Ecology*, vol. 48, no. 3, pp. 758–767, 2011.
- [9] S. Markovich-Golan, A. Bertrand, M. Moonen, and S. Gannot, "Optimal distributed minimum-variance beamforming approaches for speech enhancement in wireless acoustic sensor networks," *Signal Processing*, vol. 107, pp. 4 – 20, 2015, special Issue on ad hoc microphone arrays and wireless acoustic sensor networks.
- [10] A. Bertrand and M. Moonen, "Distributed node-specific LCMV beamforming in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 233–246, Jan 2012.
- [11] A. Hassani, J. Plata-Chaves, M. H. Bahari, M. Moonen, and A. Bertrand, "Multi-task wireless sensor network for joint distributed node-specific signal enhancement, LCMV beamforming and DOA estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 3, pp. 518–533, April 2017.
- [12] A. Alexandridis, N. Stefanakis, and A. Mouchtaris, "Towards wireless acoustic sensor networks for location estimation and counting of multiple speakers in real-life conditions," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 6140–6144.
- [13] A. Alexandridis, S. Papadakis, D. Pavlidi, and A. Mouchtaris, "Development and evaluation of a digital MEMS microphone array for spatial audio," in *24th European Signal Processing Conference (EUSIPCO)*, Aug 2016, pp. 612–616.
- [14] 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.