

Clustering based Multi-object Positioning System

Viet-Hung Dang
Department of Computer
Engineering, KyungHee
University, Gyeonggi-do,
446-701, Korea. Email:
dangviethung@oslab.khu.ac.kr

Thanh-Phuong Phan
Department of Mathematics,
Nguyen Trai School,
Hoi An, Quang Nam,
Vietnam. Email:
phuonghuevn2002@yahoo.com

Ba-Vui Le
Department of Computer
Engineering, KyungHee
University, Gyeonggi-do,
446-701, Korea. Email:
lebavui@oslab.khu.ac.kr

Young-Koo Lee *and*
Sungyoung Lee
Department of Computer
Engineering, KyungHee
University, Gyeonggi-do,
446-701, Korea. Emails:
sylee@oslab.khu.ac.kr,
yklee@oslab.khu.ac.kr

Abstract—Acoustic source positioning plays an important role in military tracking unwelcome objects. A system for this application must be capable of dealing with the input recorded convolved mixture signals while minimizing the high communication and computation cost. This paper describes a distributed system for positioning multiple independent moving sources relying on acoustic signals. The sensors pre-process the sensed data to obtain the frequency features before compressing and sending it to the base. At the base, the source positioning are carried out via two clustering stages and an optimization method. Analysis and simulation results show that our system provides high accuracy and needs neither much communication nor complex computation in a distributed manner. It is robust even when there exists high noise with Rayleigh multi-path fading under Doppler effect and when the number of independent sources is greater than the microphone number.

I. INTRODUCTION

Previous works are mainly based on finding the relative angles between the sound sources and the receiving sensor arrays, called angles of arrival (AOAs). For techniques that use AOA scheme, most of acoustic approaches so far can only give solutions to one tracked object [1][2][3][4][5], and just few are for multi-object tracking [6][7]. To solve multi-object tracking with non-array sensors, the idea of using independent component analysis (ICA) comes naturally since ICA is a powerful algorithm to separate and restore the original source data provided that these sources are statistically independent. In reality, an acoustic signal takes different time delays for propagating to the sensors, generating the convolved mixture data. Some methods have been developed to deal with this problem on time domain [8] and on frequency domain [9]. However their computation load are either too high [10] or too complicated especially when the finite impulse response (FIR) Linear algebra is used for ICA on the complex field [11]. In addition, all the related techniques so far generally require a centralized algorithm, making the communication load too big to apply to wireless sensor networks (WSNs).

Our preliminary method [12], based on ICA applied on frequency domain, can deal with convolved mixture data [8][9] and overcome many disadvantages of convolved mixture ICAs on both time and frequency domains. Nevertheless, it can only position still sources and does not adequately tolerate the noise. Fortunately, we observe that independent sources allow

for replacing the high-cost and low-reliability ICA techniques with clustering methods. The positioning problem can be solved relying on the information of magnitude ratios each of which is calculated from the energies of an f-component at different sensors. Localization method based on this information will be described in the paper.

II. PROPOSED METHOD FOR MULTI-OBJECT TRACKING

A. Problem statement

Consider M objects emitting continuous zero-mean acoustic signals and N location-known sensors. The signals are denoted by $s_j(t), j = 1, \dots, M$ while at each sensor i , the received data are denoted by $x_i(t)$ and modeled as in [2]

$$x_i(t) = \sum_{j=1}^M a_{ij} s_j(t - \tau_{ij}(t)), \quad i = 1, \dots, N \quad (1)$$

where $a_{ij} > 0$, is the amplitude gain of the signal from source j measured at sensor i and $\tau_{ij}(t)$ is the propagation time of this signal. When the sources move, these parameters change over time and cause different shifts to different f-components at the receivers. That phenomenon is called Doppler effect [13]:

$$f_{ij} = \left(\frac{v_c}{v_c + v_j \cos(\theta_{ij}(t))} \right) f_j, \quad (2)$$

where f_j is some f-component of source j , f_{ij} is the shifted version of f_j at sensor i , and $\theta_{ij}(t)$ is the immediate angle between \vec{i}_j and \vec{v}_j .

The issue is: With the received data and the only knowledge that the delayed versions of the sources are statistically independent of one another, the source positions must be indicated.

B. Distance information extraction

Applying Short Time Fourier Transformation (STFT) to the sampled data at the sensor i , the time-delay τ_{ij} only affects the phase spectral image, not the magnitude spectral image (so-called frequency image). Since the continuous form of STFT is not suitable for computing and storing, the Discrete Fourier Transformation (DFT) is replaced for calculation at sensors. Also note that when the speeds of the source are not zero, a source's magnitude frequency images calculated at different

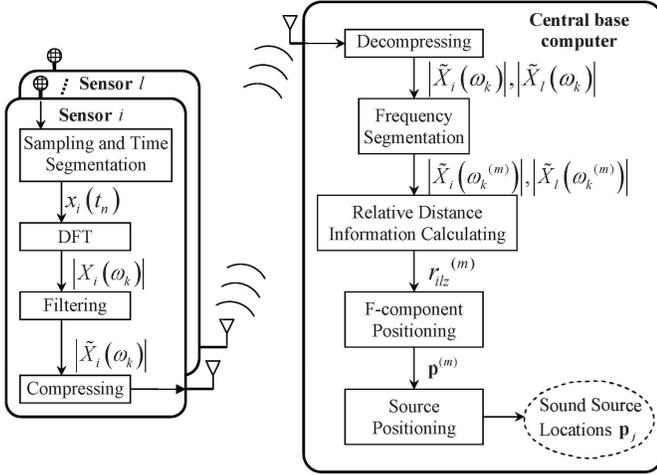


Fig. 1. Sensor architecture and Base architecture of the proposed system.

sensors do not have the same form, thus the results of DFT for recorded data are

$$|X_i(\omega_k)| = \sum_{j=1}^M |a_{ij}| |S_{ij}(\omega_k)|, \quad i = 1, \dots, N. \quad (3)$$

$X_i(\omega_k)$ in the above equation is the DFT results of $x_i(t)$ and k represents the discrete index. Meanwhile, $|S_{ij}(\omega_k)|$ is the discrete frequency image of the signal emitted by source j and recorded by sensor i . Now consider a particular interval on the frequency domain (ω_a, ω_b) containing all shifted versions of some f-component of source z without any interference from other sources' shifted f-components, the frequency images in this interval are

$$\begin{aligned} |X_i(\omega_k^{(m)})| &= \sum_{j=1}^M |a_{ij}| |S_{ij}(\omega_k^{(m)})| \\ &= |a_{iz}| |S_{iz}(\omega_k^{(m)})|, \quad i = 1, \dots, N, \end{aligned} \quad (4)$$

where $\omega_k^{(m)} \in (\omega_a, \omega_b)$ and m is the index of the f-component. Although this f-component has different shifted versions, its energy is unchanged since the magnitude of the signal on the time domain is the same, or

$$\left| S_{iz}(\omega_k^{(m)}) \right|^T \left| S_{iz}(\omega_k^{(m)}) \right| = \left| S_{lz}(\omega_k^{(m)}) \right|^T \left| S_{lz}(\omega_k^{(m)}) \right|, \quad (5) \\ i \neq l.$$

Based on the fact from (4) and (5), if an f-component belongs to source z , then all relative distance relationships are

$$r_{ilz}^{(m)} = \frac{|a_{iz}|}{|a_{lz}|} = \frac{d_{lz}}{d_{iz}} = \sqrt{\frac{\left| \tilde{X}_i(\omega_k^{(m)}) \right|^T \left| \tilde{X}_i(\omega_k^{(m)}) \right|}{\left| \tilde{X}_l(\omega_k^{(m)}) \right|^T \left| \tilde{X}_l(\omega_k^{(m)}) \right|}}, \quad i \neq l \quad (6)$$

where d_{lz} and d_{iz} are the distances from source z to sensor l and to sensor i respectively; $\tilde{X}_i(\omega_k)$ is the result after the step of noise filtering $X_i(\omega_k)$, and $\tilde{X}_i(\omega_k^{(m)})$ is the frequency

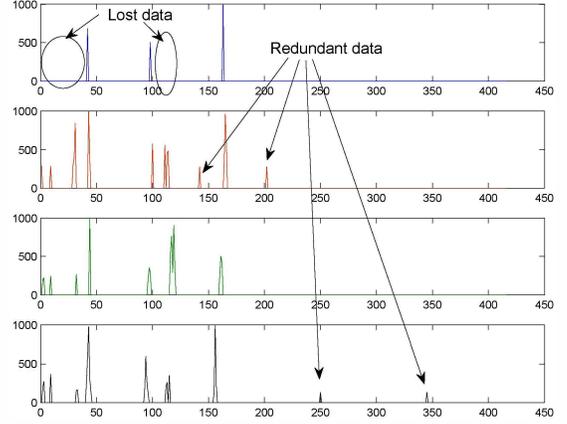


Fig. 2. An example of data after filtering on Fourier domain

image of $\tilde{X}_i(\omega_k)$ on the segment (ω_a, ω_b) (see Fig.1). That means, for each f-component m within the frequency interval, a set of constraints is calculated and the location of the source having these components can be estimated. Thus, two clustering steps are needed, one for grouping the shifted frequency components to determine the segment (ω_a, ω_b) , and the other for grouping f-component positions to calculate source locations after f-component positions are computed. The advantages of this system are: (a) it is more robust than our previous system even when the sources are fixed, (b) it works well with moving sources and tolerates the co-existence of Doppler effect and Rayleigh multi-path fading, (c) it is considered to be a distributed method since the computation load is shared among the sensors and the communication cost is low, and (d) it is not constrained by the condition that the sensor number is greater than the source number.

III. PROPOSED SYSTEM ARCHITECTURE

The architecture design of the acoustic tracking system is displayed in Fig.1 due to the extraction of distance information method as mentioned in Subsection II-B.

A. Sensor architecture

On frequency domain, the Gaussian noise level can be detected and all low f-components can be forced to zero (see Fig.2). Filtering step keeps only several dominant f-components, so the data to be transmitted from a sensor to the base computer is reduced significantly. This is one of the key ideas for compressing the communication load so that the method can be applied into WSNs.

The computation load at the sensors is high with DFT transformations of lengthy frames. However, as one can see in Fig.2, a sensor can skip calculating the frequency bins where the probability of major f-components' existence is low according to the feedback from the base. Thus, the computation load at sensors is reduced considerably, too.

B. Central base architecture

The work flow at the base is straightforward as described in Fig.1 and in Section II-B. Received data are decompressed and fed into the ‘‘Frequency-Segmentation’’ stage. This process marks dominated f-components as well as the corresponding segments that contain the components with the index m . Then the block ‘‘Relative Distance Information Calculate’’ calculates a set of $r_{il}^{(m)}$ for each component. These sets are then input into the ‘‘F-component Positioning’’ process to estimate the output position of each dominant f-component $\mathbf{p}^{(m)}$. Frequency leakages, setting noise, Doppler effect and Rayleigh fading influence the detection result and make f-components belonging to the same source j not have the same position. Therefore, the final stage ‘‘Source Positioning’’ is necessary to cluster those $\mathbf{p}^{(m)}$ and estimate \mathbf{p}_j under the averaging mechanism.

1) *Frequency Segmentation* : This stage is used to indicate every of frequency intervals which includes all shifted versions of a dominant f-component. It performs a clustering task that groups an f-component’s shifted versions and determines that frequency segment. Doppler effect influences the f-components differently, the higher is the frequencies, the larger is the shift. From (2), an f-component of source j at f_0 has shifted versions within $(\frac{v_c}{v_c+v_j}f_0, \frac{v_c}{v_c-v_j}f_0)$. This frequency interval varies depending on f_0 on the frequency scale, however, it is fixed on the logarithmic scale as can be seen below

$$\begin{aligned} \Delta f(\text{dB}) &= \log_{10}\left(\frac{v_c}{v_c-v_j}f_0\right) - \log_{10}\left(\frac{v_c}{v_c+v_j}f_0\right) \\ &= \log_{10}\left(\frac{v_c+v_j}{v_c-v_j}\right). \end{aligned} \quad (7)$$

As the result, clustering task should be performed on the $\log_{10}(\cdot)$ scale of the frequency image under following criteria: (a) the width of each segment is not larger than $\Delta f(\text{dB})$ (see 7); (b) the number of nonzero f-components within the grouped interval is greater than 2 so that the number of constraints is at least 3; and (c) the average energy of an f-component received at the sensors must be larger than the detected noise level. A sliding window with the width $\Delta f(\text{dB})$ is then used to detect the frequency segments that hold (b) and (c). As the result, the number of sources can be larger than that of sensors. Moreover, the total loss of some f-components due to filtering is acceptable and the redundant f-component will hardly be taken into account.

2) *F-component Positioning* : All constraint ratios r_{ilz} are computed (see (6)) in ‘‘Relative Distance Information Calculating’’ process before being fed into the ‘‘F-component Positioning’’ process. The error in the constraints is unavoidable due to frequency leakage and the setting noise, so the solution for the position of f-component m should be a vector $\mathbf{p}^{(m)}$, $\mathbf{p}^{(m)} \in \mathbf{R}^2$ that compromises the constraints. We propose an objective function for this compromise

$$\mathbf{F}_j = \sum_i^N \sum_{l,t \neq i}^{N-1} (d_{ij} - r_{ilj}d_{lj})^2, \quad 0 < r_{ilj} < \infty \quad (8)$$

and the solution for source j will be

$$\mathbf{p}^{(m)} = \arg \min_{\mathbf{p}^{(m)}} \mathbf{F}_j. \quad (9)$$

The simple negative gradient method is chosen for this optimization problem.

3) *Source Positioning* : This stage groups f-component positions and computes the source coordinates as the mean values of f-component groups. The nearest-neighbor clustering, or d-min clustering technique [14] is used in this ‘‘Source Positioning’’ process. Equations (10) estimates the position of a source that includes the f-components whose positions are in the group j' , denoted by $D_{j'}$:

$$\hat{\mathbf{p}}_{j'} = \frac{\sum_{\mathbf{p}^{(m)} \in D_{j'}} \mathbf{p}^{(m)}}{N^{(m)}}. \quad (10)$$

where $N^{(m)}$ is the number of f-components in the group.

IV. EXPERIMENTS AND DISCUSSIONS

Two main experiment sets are conducted via simulations in this section for system working demonstration and system evaluation.

A. Experiment setup and modeling

The deployed area is [0m,12m]x[0m,12m]. Five simulated sources ($M = 5$), which imitate the sounds of different vehicles, motors and a siren (see Fig.3), are generated. They are parametrically determined so that Doppler effect can be generated properly (see (1)). If a group sources are close together, they can be seen as one sound source; thus it is necessary for the sources to be distinct from one another for evaluation (5 meters in this simulation). Four sensors ($N = 4 < M$) are deployed around the corners of the deployed area (see Figure 4). The energies of the line-of-sight signals propagating to the sensors decrease according to the inverse square law at the sound speed of $c = 343\text{m/s}$. The sampling frequency is $F_s = 16.384\text{KHz}$ and the time segment length T_f is 0.2s. The background noise for simulations of this paper is Gaussian and its level is the same at all sensors.

Received acoustic data in practice always includes the effects of shadowing and fading due to multiple path reflections along with the received line-of-sight signals and Doppler effect. Therefore, we examine the situation under the existence of a Rayleigh fading channel. Since generating multiple paths for each source takes much computing time especially when Doppler effect is present, the Young model [15], which generates a Rayleigh channel with two arrays of Gaussian random variables and the inverse-DFT (IDFT) technique, is applied. If the ranges overlap, then the number of generated complex values is the number of overlapped ranges, and the Rayleigh noise at the bin is the sum of these values. In order for the result of IDFT to be real, the array representing the Rayleigh fading effect on the frequency domain, denoted by $R(k)$, must be conjugately symmetric, or

$$\begin{cases} R(0) = 0 \\ R(k) = R^*(N-k), \quad k = 1, \dots, N-1 \end{cases} \quad (11)$$

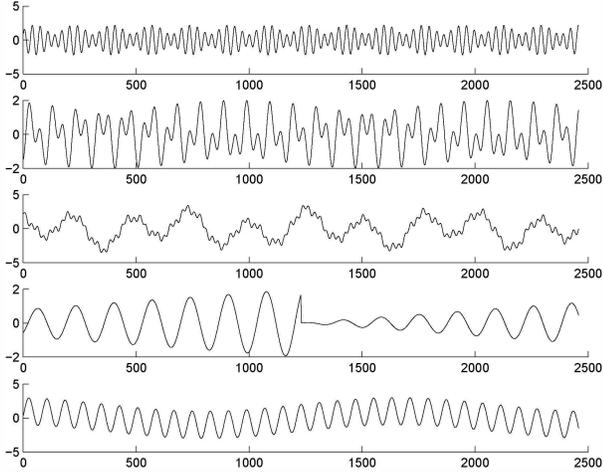


Fig. 3. Signals of sources which are parametrically generated to imitate sounds of vehicles, motors and a siren.

The parameter Signal to Noise Ratio SNR is chosen for evaluating results:

$$SNR = \frac{E_{mean}}{\alpha_{RL}E_{Rayleigh} + (1 - \alpha_{RL})E_{Background}}. \quad (12)$$

where E_{mean} is the mean value of the average signal energies received at the four sensors, E_{noise} is the sum noise energy of the background noise $E_{Background}$ and the Rayleigh noise $E_{Rayleigh}$, while $\alpha_{RL} \in (0, 1)$ represents the percentage of Rayleigh noise energy in the total noise energy.

B. System working demonstration

At the sensors, after ‘‘F-component Positioning’’ process, each f-component’s position is determined and plotted with a circle (see Fig.4). Those f-components whose estimated positions are close to one another are grouped together as described in Section III-B3, in which $d_{min} = 3.5m$.

Fig.4 displays the results of positioning task when the system attempts to localize five independent sources in the time segment T_f of $0.2s$ and the Rayleigh fading contribution of 20%. The estimated positions of the sources are calculated based on the groups’ f-component positions due to equations (10). Fig.4 is for the position estimation result when the speeds of the sources are all $40km/h$. It can be seen that the system is capable of locating the sources even when the source number is greater than that of the sensor. The Root Mean Square Errors (RMSEs) for the clustering in Fig.4 is less than 1.6 meters, an acceptable level especially when the speeds of the sources are high (in $0.2s$, the trail lengths are around $2.2m$). Obviously, positioning based on f-component localization is a good approach to deal with multiple acoustic source positioning in situations affected by high Gaussian noise, multi-path fading and Doppler effect. The SNR here is 2.51, the highest noise level of simulations in this paper.

C. System performance evaluation

The simulation in this section tries to examine the impacts of time segment T_f , the ratio SNR and the speed v_j on the

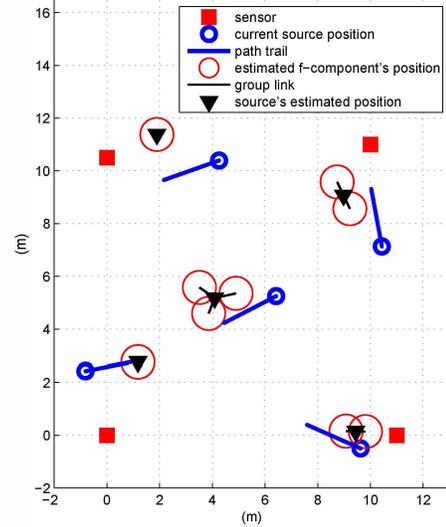


Fig. 4. Estimation results of an example using the highest level of noise in the simulation set and $\alpha_{RL} = 0.2$, source speeds are $40km/h$.

result accuracy. The chosen T_f is $0.2s$, while the speed varies from 0 to $40km/h$ in increments of $8km/h$. The SNR values are generated based on the linear increment of the standard deviation of Gaussian noise. Fig.5 illustrates the distance errors under the impacts of noise level SNR , the percentage of Rayleigh multi-path fading noise α_{RL} and the speed of sources v_j when the time segment is $0.2s$. Each plotted error value is the average result of RMSEs of 1000 trials.

It can be seen that higher noise and higher velocity lead to higher source positioning error. It is because the higher noise level results in more error in the f-component positions due to the increased error in the constraint ratios, especially if the f-components have low magnitude. Meanwhile, higher speeds lengthen the pathtrails, increasing the uncertainty of positions.

From Fig.5, one can see that the RMSE increases almost linearly with v_j . It is obvious since higher speed sources leave longer path trails. In addition, Rayleigh multi-path fading caused by high v_j affects the accuracy less than that caused by low v_j (compare the sub-figures). An f-component at a low v_j produces noise in a narrow and condensed Doppler shift range on the frequency domain. As a result, at the same level of Rayleigh noise, the received energy of an f-component through the line-of-sight path is corrupted more by a narrow shift range than by a wide shift range, causing poor accuracy at low v_j when there exists Rayleigh fading. Meanwhile, when v_j is high, the energy of Rayleigh fading noise is spread wider and thinner on the frequency domain and less affects the line-of-sight f-component.

It is noticeable that the system can not obtain the ideal results when there exists no noise ($SNR = \infty$). The accuracy of estimated f-component positions suffers because of the limited time length of a frame which produces unavoidable spectral leakage. Moreover, the errors are also caused by the influence of the second original source, whose f-components appear all

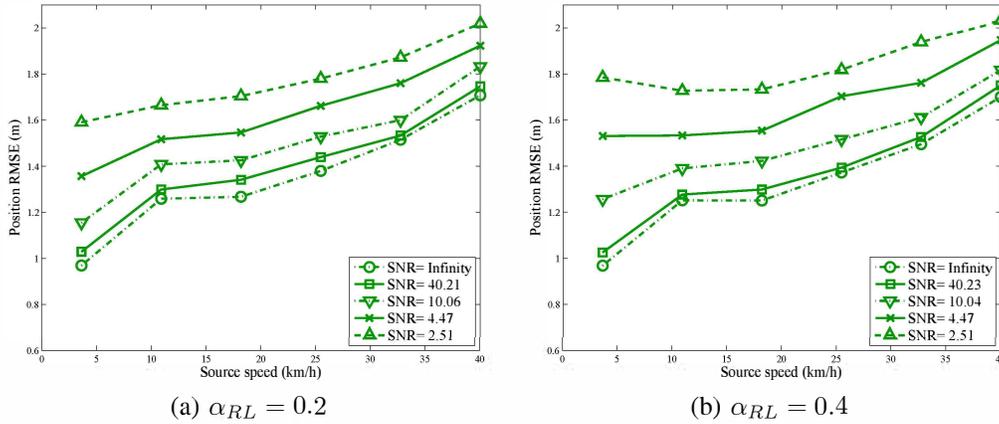


Fig. 5. RMSE error results of source locations under influences of speeds of sources, Gaussian noise and Rayleigh multi-path fading when T_f 0.2s.

over the frequency domain, to other sources' f-components. However, when the noise level increases quickly, the system can tolerate the noise well with little error increment. The estimation error increases an average amount of 0.6m when the SNR decreases 16 times from around 40 to 2.5. Comparing two sub figures of Fig.5, it is evident that the higher is the contribution of Rayleigh fading noise to the same level of total noise, the worse is the f-component clustering result.

V. CONCLUSIONS

We have described a distributed system for independent acoustic source positioning in which the separation is performed based on the ratios of f-component energy values received at the sensors with clustering and optimization methods. The simulation conditions are made realistic with Doppler effect and Rayleigh multi-path fading to illustrate how well the system solves the problem of multiple moving source positioning. The results show that the system gives high accuracy and requires low communication cost for a large data set. The proposed system can be regarded as a design for the future generations of WSNs because it requires powerful sensors for performing DFT on long segments of data. Nevertheless, strong computing ability is not essential because with the feedback from the base, the sensors only perform the full DFT once and then focus on calculating DFT at bins in several frequency segments which contain the dominant f-components. The system is actually more useful than just positioning multiple sources. It can also output the characteristics of sources for further position estimation refinement and recognition since most acoustic features are on the frequency domain.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Kyung Hee University in 2010, (KHU-10201372).

REFERENCES

[1] Y. Sasaki, S. Kagami, and H. Mizoguchi, "Multiple sound source mapping for a mobile robot by self-motion triangulation," in *International Conference on Intelligent Robots and Systems*. IEEE, Oct 2006, pp. 380–385.

[2] M. Stanacevic, "Micropower gradient flow acoustic localizer," *IEEE Transactions on Circuits and systems*, vol. 52, pp. 2148–2157, 2005.

[3] S. T. Birchfield and D. K. Gilmore, "A unifying framework for acoustic localization," in *IEEE International Conference*. IEEE Computer Society, 2001.

[4] A. Mahajana and M. Walworth, "3-d position sensing using the differences in the time-of-flights from a wave source to various receivers," *IEEE Transaction on Robotics and Automation*, vol. 17, pp. 91–94, 2001.

[5] J.-M. Valin, F. Michaud, J. Rouat, and D. Letourneau, "Robust sound source localization using a microphone array on a mobile robot," in *In Proc. ICASSP98*, 2003, pp. 1228–1233.

[6] F. Antonacci, D. Riva, D. Saiu, A. Sarti, M. Tagliasacchi, and S. Tubaro, "Tracking multiple acoustic sources using particle filtering," in *European Signal Processing Conference, EUSIPCO-2006*, 2006.

[7] K. Kaouri, "Left-right ambiguity resolution of a towed array sonar," in *Ph.D dissertation*. MSc in Mathematical Modelling and Scientific Computing, 2000.

[8] A. Ciaramella and R. Tagliaferri, "Separation of convolved mixtures in frequency domain ica," *International Mathematical Forum*, vol. 16, pp. 769–795, 2006.

[9] P. Smaragdakis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, pp. 22:21–34, 1998.

[10] D. B. Ward and R. C. Williamson, "Particle filter beamforming for acoustic source localization in a reverberant environment," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2002.

[11] E. BINGHAM and A. HYVARINEN, "A fast-point algorithm for independent component analysis of complex valued signals," *International Journal of Neural Systems*, vol. 10, pp. 1–8, 2000.

[12] V.-H. Dang, T. Le-Tien, Y.-K. Lee, and S. Lee, "Acoustic multiple object positioning system," in *International Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks, PE-WASUN-2010*. ACM, 2010.

[13] R. Serway and R. Beichner, *Physics for Scientists and Engineers*, 5th ed. Brooks/Cole Publishing Company, 1999.

[14] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley & Son, 2001.

[15] D. J. Young and N. C. Beaulieu, "The generation of correlated rayleigh random variates by inverses discrete fourier transform," *IEEE Transactions on Communications*, vol. 48, pp. 1114–1127, 2000.