J Supercomput (2012) 61:438–462
DOI 10.1007/s11227-011-0600-x

A distributed design for multiple moving source positioning

Viet-Hung Dang · Sungyoung Lee · Young-Koo Lee

Published online: 8 April 2011

Abstract Acoustic source localization has many important applications particularly for military tracking foreign objects. Even though Wireless Sensor Networks (WSNs) have been developed, this localization problem remains a big challenge. A system for solving source localization must have the ability to deal with the problems of recorded convolved mixture signals while minimizing the high communication and computation cost. This paper introduces a distributed design for positioning multiple independent moving sources based on acoustic signals in which we focus on utilizing the relative information of magnitudes recorded at different sensors. The sensors perform preprocessing on the sensed data to capture the most important information before compressing and sending extracted data to the base. At the base, the data is uncompressed and the source locations are inferred via two clustering stages and an optimization method. Analysis and simulation results lead to the conclusion that our system provides good accuracy and needs neither much communication nor complex computation in a distributed manner. It works well when there exists high noise with Rayleigh multipath fading under Doppler effect and even when the number of independent sources is greater than the number of microphone sensors.

Keywords Multiple moving source · Positioning · Acoustic · Doppler effect · Rayleigh

V.-H. Dang · S. Lee (✉) · Y.-K. Lee
Department of Computer Engineering, KyungHee University, Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, 446-701, Korea
e-mail: sylee@oslab.khu.ac.kr

V.-H. Dang
e-mail: dangviethung@khu.ac.kr

Y.-K. Lee
e-mail: yklee@oslab.khu.ac.kr

1 Introduction

One of the most popular purposes of Wireless Sensor Networks (WSNs) is the determination of object locations. This ability is necessary especially in the military when WSNs have been deployed to detect objects and their locations within the deployed region. Unlike the active devices that emit radio or ultrasonic signals to detect objects based on the reflected waves, sensors in WSNs are usually passive and only record the signals from objects. Thus, most developed tracking techniques with passive sensors are based on known communications channels. In other words, they are only suitable for localizing objects that are designed to be monitored, not for the very important security and defense application of localizing foreign objects. These objects have no prior channels to communicate with the system and the only detected signals that the system can capture are the images and the sounds. For image signals, however, cameras cannot be deployed randomly or casually in a large number due to high costs, view blockage and the limited angle of view. Moreover, the communication cost would be very high for image transmission while the computation for positions needs a centralized method. Therefore, the easiest and most convenient way to monitor an object is using acoustics. This article describes a full design for positioning multiple moving objects that emit sound signals and the analysis for its performance. The design overcomes the drawbacks of both previous works and the previous version of the system. The input data is not simply the convolved mixtures, which are just the combinations of signals experiencing different time delays, but is more complicated mixtures in which Doppler effect and Rayleigh multipath fading coexist. Extracted information from the mixtures in this work is not of time-delay differences or directions of arrivals but of the ratios between Received Signal Strengths (RSS) from each object to different sensors. We remark that we are the first to address the multiobject positioning problem by focusing on the ratios of source magnitudes extracted from the complicated convolved mixture data which includes Doppler effect and Rayleigh multipath fading. In addition, the sensors deployed in the area are not the arrays of microphones as in many previous works but the isotropic acoustic recorders. More importantly, aiming for a low-cost method, we make the method applicable into WSNs in a distributed way, where the whole computation load is shared on the sensors with a low communication cost for data collection. Distributed approaches are categorized into *data decomposition*, *process decomposition* or *data-and-process decomposition*, depending on how the algorithm is shared on different computers [1]. In this regard, our system is designed for working in *process and data decomposition* manner whose details are discussed in Sect. 4. Sensed data is preprocessed, compressed, and sent to the base. The base computer decompresses the data and extracts the information of ratios between the energies of each dominant frequency component (f-component) using a clustering method on the frequency domain. These ratios are then used to estimate the positions of all dominant f-components, and the source location estimations are computed by clustering these f-component positions.

2 Related works

Technically, most current acoustic approaches which utilize passive sensors in WSNs can only deal with one tracked object [2–6], and few works have studied multiple object tracking problem [7, 8]. Moreover, the mentioned methods are strongly based on the techniques for one target tracking. The common focus of those techniques is on the Direction of Arrival (DoA), or the relative angles between sound sources and sensor arrays. After that, particle filter or hidden Markov model may be used to enhance the tracked routes of the sources [7, 9]. Except for the beamforming technique [8] used in passive towed array sonar system, the numbers of tracked sources used in previous works are small, two is common, and at most three [7, 10]. Towed array sonar system is an array of sensors deployed along a tow which is usually considered to be straight. Based on delay times and linear phase differences recorded at sensors, it gives proper directions of arrival signals if the sources are far from the tow but it gives poor range estimation [8, 11]. Such sonar system, after all, is a well developed DoA sensor array. In this work, we do not use the nodes each of which has a sensor array because this kind of node also has the problem when being randomly deployed. Instead we design an isotropic acoustic recorder for each node, so randomly deploying sensors is not a problem and the load of sensed data is also reduced. The direction of arrival information is then not the feature for solving the source localization because of the change of the input data. For the scenario where a single source is monitored by several sensors, calculation of localization is usually based on Time Differences Of Arrival (TDOA) directly [5] or indirectly by maximizing the cross correlation [6]. However, for a large number of sources to be located with the isotropic sensors, we would intuitively consider some techniques that can be applied to the problem, including Principal Component Analysis (PCA) and Independent Component Analysis (ICA) techniques [12] which are capable of solving a class of Blind Source Separation (BSS) problems [13]. When the sources are recovered, the different time delays from sources to sensors can be obtained for location estimating. Nevertheless, computing the delay from sources to sensors proves to be difficult since the signal from a source takes different time delays to reach the sensors, so the observed data is a set of convolved mixtures. Although convolved mixture ICAs have been developed to deal with this kind of complex data, their disadvantages include the excessive computation cost because the Finite Impulse Response (FIR) Linear Algebra model must be used [14, 15]. The necessity of a centralized manner makes the communication load become so big that it is difficult, if not impossible, to apply convolved mixture ICAs to WSNs. Moreover, related works on BBS so far cannot deal with the data that has Doppler effect, not to mention the interference of high noise and Rayleigh multipath fading. Dealing with convolved mixtures under the influence of Doppler effect and Rayleigh multipath fading for localization, which has not mentioned before, will be discussed in this paper. We have introduced the early version of the system design for source positioning based on ICA technique on the frequency domain in order to eliminate the influence of time delays and to extract the ratios between original source energies for position estimating [16]. Our preliminary method can deal with convolved mixture data [14, 17] and avoid many drawbacks of convolved mixture ICAs on both time and frequency domains [2, 9, 18]. Nevertheless, it can only be used for positioning still sources and does not adequately tolerate the noise. It fails to recover the frequency images, or magnitude spectral images, of moving sources when Doppler effect causes different spectral shifts on the frequency domain especially when Rayleigh multipath fading is experienced. Using ICA on the data of frequency images, we observe that independent sources allow for replacing the high-cost and low-reliability ICA techniques with clustering methods. The localization can be solved relying on the information of magnitude ratios each of which is calculated from energies of an f-component at different sensors. Localization method based on this information has not appeared in any previous works before either and will be presented in the following section.

3 Proposed method for multiobject tracking

3.1 Problem statement

Assume that there are M objects emitting continuous zero-mean acoustic signals and N location-known sensors; these signals can be considered as $s_j(t)$, $j = 1, \dots, M$. At each sensor i, the received data is denoted as $x_i(t)$. The data received at each sensor is the actual signal with continuous values of delay, similar to the model in [3], and is calculated according to

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

where a_{ij} is a real positive number representing 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 are fixed, the delays $\tau_{ij}(t)$ are constants. However, if the sources move, these parameters change at different time points,

$$\tau_{ij}(t) = \frac{d_{ij}(t)}{v_c}, \tag{2}$$

where $d_{ij}(t)$ is the immediate distance from sensor i to source j and v_c is the velocity of acoustic propagation. Source j's movement with velocity v_j causes $\tau_{ij}(t)$ to increase or decrease over time, resulting in a stretched or compressed image of the source signal on the time domain at receiver i. As a result, different shifts are caused to different f-components at the receivers. This phenomenon is known as Doppler effect [19, Chap. 17] and is expressed as

$$f_{ij} = \left(\frac{v_c}{v_c + v_j \cos(\theta_{ij}(t))}\right) f_j, \tag{3}$$

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

The problem can be stated as: Without prior knowledge about the sources except for the received data at the sensors and the information that delayed versions of the sources are statistically independent of one another, the locations of sources must be determined.

As mentioned before, the feature $\tau_{ij}(t)$ is actually the best measurement, giving good results of object location estimation. However, it is difficult to extract this feature and a big load of communication is required to transmit all of the data to the base in order to perform the algorithm. In this work, we mainly focus on extracting the relative information among the magnitudes of a source signal at different sensors.

3.2 Distance information extraction

Note that if there is only one active fixed source and the others are inactive, or emitting no sound, applying Short Time Fourier Transformation (STFT) to the sampled data $s_i(t_k)$ at this source and sampled data at each sensor, we obtain the same magnitude spectral images. The different parts of the STFT results are the scalar coefficients and the phase spectral images. Obviously, the time-delay τ_{ij} only affects the phase spectral image. Therefore, for multiple fixed and independent sources, if STFT is applied at each sensor for each mixture, the result is:

$$X_i(\omega) = \sum_{j=1}^{M} a_{ij} S_j(\omega) e^{-2\pi\tau_{ij}}, \quad i = 1, \dots, N, \tag{4}$$

$$\left|X_i(\omega)\right| = \sum_{j=1}^{M} |a_{ij}|\left|S_j(\omega)\right|, \quad i = 1, \dots, N. \tag{5}$$

As usual, the continuous form of STFT is difficult to compute and store, so the Discrete Fourier Transformation (DFT) form is the best choice of replacement. Equation (5) is rewritten in the discrete form as

$$\left|X_i(\omega_k)\right| = \sum_{j=1}^{M} |a_{ij}|\left|S_j(\omega_k)\right|, \quad i = 1, \dots, N. \tag{6}$$

Evidently, the magnitude spectrum data takes the form of instantaneous mixtures. The sound signals are zero-mean and mainly composed of sinusoid waves. Also, the delayed versions of source signals are statistically independent of one another. That leads to the fact that if an f-component is included in one source, it would not be present in the others. As a result, the magnitude spectra of different $|S_j(\omega_k)|$ are orthogonal to each other or

$$\left|S_u(\omega_k)\right|^T \left|S_v(\omega_k)\right| = 0, \quad u \neq v. \tag{7}$$

As a result, when the number of sources is less than or equal to the number of sensors, $|S_j(\omega_k)|$ in (6), or the magnitude spectral image of $s_i(t_n)$, can be restored using a standard ICA [16]. Note that ICA cannot restore the magnitudes of the original Independent Components (ICs). Instead of providing the exact $|a_{iz}||S_z(\omega_k)|$, ICA results in $b_z|S_z(\omega_k)|$, $b_z \in R$. $|X_i(\omega_k)|$ is the linear combination of orthogonal vectors $|S_j(\omega_k)|$, thus the inner product of each IC vector $|b_z||S_z(\omega_k)|$ and each magnitude spectral image $|X_i(\omega_k)|$ contains the information of energy of this IC observed by sensor i.

$$\left(|b_z|\left|S_z(\omega_k)\right|\right)^T \left|X_i(\omega_k)\right| = |b_z|\left|S_z(\omega_k)\right|^T \sum_{j=1}^{M} |a_{ij}|\left|S_j(\omega_k)\right| \quad \text{or} \tag{8}$$

$$\left(|b_z|\left|S_z(\omega_k)\right|\right)^T\left|X_i(\omega_k)\right| = |b_z a_{iz}|\left|S_z(\omega_k)\right|^T\left|S_z(\omega_k)\right|. \tag{9}$$

Therefore, for each IC z, for each pair of magnitude spectral images of observed data at sensors i and l, the ratio $|\frac{a_{iz}}{a_{lz}}|$ can be achieved as

$$\left|\frac{a_{iz}}{a_{lz}}\right| = \frac{(|b_z||S_z(\omega_k)|)^T|X_i(\omega_k)|}{(|b_z||S_z(\omega_k)|)^T|X_l(\omega_k)|}. \tag{10}$$

Meanwhile,

$$\left|\frac{a_{iz}}{a_{lz}}\right|^2 = \frac{a_{iz}^2|s_j(t_n)|^T|s_j(t_n)|}{a_{lz}^2|s_j(t_n)|^T|s_j(t_n)|} = \frac{E_{iz}}{E_{lz}}, \tag{11}$$

where E_{iz} is the energy sent by source z and received by sensor i in an interval of time. Since the absorption of gas molecules is insignificant, due to the inverse square law, the energy of sound decreases proportionally to the inverse square of the distance [19, Chap. 17]. In other words,

$$\frac{E_{iz}}{E_{lz}} = \frac{(\frac{1}{d_{iz}})^2}{(\frac{1}{d_{lz}})^2}. \tag{12}$$

From (10), (11), and (12), we have the relationships of all pairs of distances from any tracked object j to all sensors,

$$r_{ilz} = \left|\frac{a_{iz}}{a_{lz}}\right| = \frac{d_{lz}}{d_{iz}} = \frac{(|b_z||S_z(\omega_k)|)^T|X_i(\omega_k)|}{(|b_z||S_z(\omega_k)|)^T|X_l(\omega_k)|}, \quad i \neq l. \tag{13}$$

Based on these relationships, the locations of all of the sources will be inferred.

However, it should be remarked that ICA achieves poor separation in the presence of noise. Moreover, when the sources move, frequency shifts occur and ICA can no longer produce ICs $b_z|S_z(\omega)|$. Therefore, we extend the meaning of "independent" for sources to "being in the state in which a shifted major f-component of a source does not overlap the shifted major f-components of other sources." The interferences of minor f-components among the sources are considered as noise. For this situation, instead of using ICA, we apply a more robust method, the main idea of this study, which involves less computation based on clustering techniques. Equation (6) then is rewritten as

$$\left|X_i(\omega_k)\right| = \sum_{j=1}^{M}|a_{ij}|\left|S_{ij}(\omega_k)\right|, \quad i = 1, \ldots, N, \tag{14}$$

where $|S_{ij}(\omega_k)|$ is the discrete frequency image of the signal emitted by source j under the "view" of sensor i. Remark that for different sensors, S_j is not the same as in (6) due to different frequency shifts caused by Doppler effect.

Now considering a specific segment 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' shiftedf-components, we have

$$\begin{aligned}\left|X_i\left(\omega_k^{(m)}\right)\right| &= \sum_{j=1}^{M} |a_{ij}|\left|S_{ij}\left(\omega_k^{(m)}\right)\right| \\ &= |a_{iz}|\left|S_{iz}\left(\omega_k^{(m)}\right)\right|, \quad i = 1, \dots, N,\end{aligned} \tag{15}$$

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. The reason is while a signal is stretched and compressed by Doppler effect on the time domain, its amplitude at the source keeps unchanged, or

$$\left|S_{iz}\left(\omega_k^{(m)}\right)\right|^T \left|S_{iz}\left(\omega_k^{(m)}\right)\right| = \left|S_{lz}\left(\omega_k^{(m)}\right)\right|^T \left|S_{lz}\left(\omega_k^{(m)}\right)\right|, \quad i \neq l. \tag{16}$$

Based on this fact, if an f-component belongs to source z, then all relative distance relationships in (13) are computed according to

$$r_{ilz}^{(m)} = \left|\frac{a_{iz}}{a_{lz}}\right| = \frac{d_{lz}}{d_{iz}} = \sqrt{\frac{|\tilde{X}_i(\omega_k^{(m)})|^T |\tilde{X}_i(\omega_k^{(m)})|}{|\tilde{X}_l(\omega_k^{(m)})|^T |\tilde{X}_l(\omega_k^{(m)})|}}, \quad i \neq l, \tag{17}$$

where $\tilde{X}_i(\omega_k)$ is the result of noise filtering $X_i(\omega_k)$ and $\tilde{X}_i(\omega_k^{(m)})$ is the frequency image of $\tilde{X}_i(\omega_k)$ on the segment (ω_a, ω_b) (see Fig. 1). This implies that, for each f-component m of a source within the frequency segment, a set of relative distance relations can be computed and the position of the source having these components can be estimated. Therefore, there are necessarily two clustering stages, 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. This is the main idea of the design which is described in more details in the next section. 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 coexistence of Doppler effect and Rayleigh multipath 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 larger than the source number.

4 Multiobject tracking system architecture

The design of the system for acoustic tracking is depicted in Fig. 1 based on the key idea mentioned in Sect. 3.2 about distance information. The figure describes both acoustic sensor work flow and central base computer work flow. Because of the nature of WSNs, the input data is already decomposed at the sensors and hence can be utilized to build up a *data decomposition* algorithm. It also can be seen in the figure that the source localization computing is decomposed into two different processes, one at the sensors and the other at the central base. A powerful enough computer

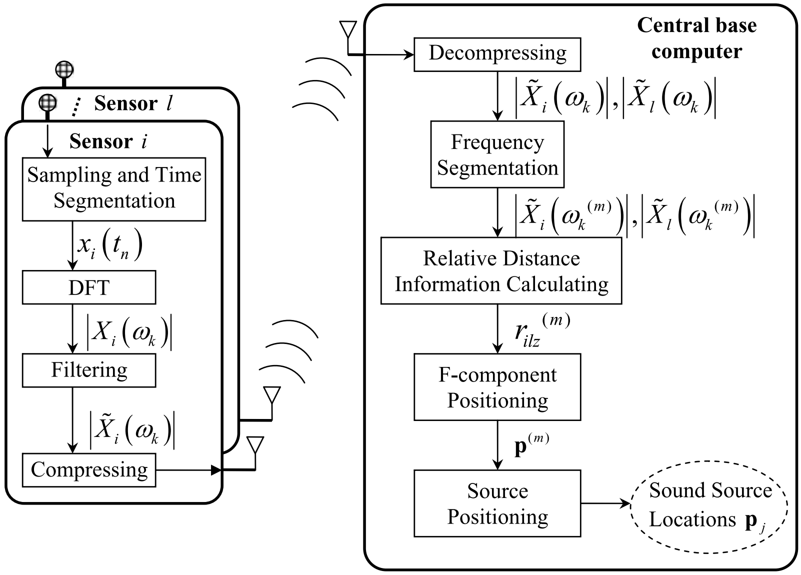


Fig. 1 Sensor architecture and Base architecture of the proposed system

is used at the base to solve its flow in a short time, not longer than the sum of the computing time at the sensors and data transmission time. Then although the process at the sensors must be executed before that at the base, with continuous input data, these two processes function in a pipeline manner, or the system works with *process decomposition* method. In other words, the design can be considered to be *data-and-process decomposition*. Since the data is decomposed into portions at sensors and the sensors have the same work flow, the process at the sensors obviously uses SIMD (Single Instruction on Multiple Data stream) approach [1]. Details of the work flows are following.

4.1 Acoustic sensor architecture

At the sensors, the acoustic signal is sampled and synchronically segmented into half-overlapped frames in the "Sampling and Time Segmentation" stage. A frame at sensor i is denoted by $x_i(t_n)$ and the time length of the frame is also called time segment, denoted by T_f.

Since our method is based on the energies within frequency intervals, we need to minimize the spectral leakage effect [20] caused by limited length DFT transformation. Thus, each frame of sensed data with length L is multiplied by the Hamming window whose weights are defined as [21]

$$w_{Hamming}(k) = 0.54 - 0.46\cos\left(\frac{2\pi(k-1)}{L-1}\right), \quad k = 1, \ldots, L. \tag{18}$$

The input data from the recorders always includes additional Gaussian noise whose spectrum is spread over the frequency domain. We can remove the noise by detecting its level and forcing all low f-components to zero. Figure 2 demonstrates the

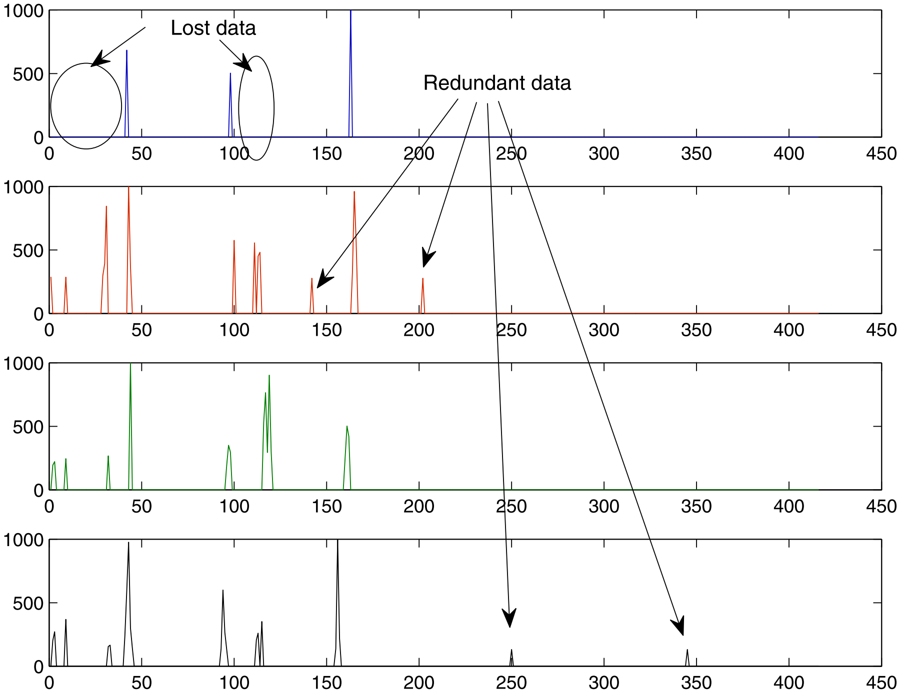


Fig. 2 An example of data after filtering on Fourier domain

results after the filtering process. The filtered data may have some lost f-components and may include redundant f-components; this is tolerated by the system as we will see in later analysis. Filtering step allows for the retention of only several dominant f-components, and as a result, the amount of data to be transmitted from a sensor to the base computer is reduced considerably. Moreover, only half of the frequency image length is needed owing to the symmetric property of the image. For example, instead of sending 1638 values in every time segment of 0.2 s with sampling frequency $F_s = 16.384$ KHz, each sensor sends only the compressed data containing less than twenty f-components (pairs of values and indexes). This is one of the key ideas for reducing the communication cost so that the method can be implemented into WSNs.

As the trade-off to this low communication cost, the computation cost at the sensors is high with DFT transformations of lengthy frames. However, as can be seen in Fig. 2, a sensor can skip calculating the frequency bins where the probability of major f-components' existence is low due to the feedback from the base. Only the frequency bins in the high energy ranges are computed. Therefore, the computation cost at sensors are also reduced considerably.

4.2 Central base computer architecture

At the central base, the flow is straightforward and consistent with what we analyze in Sect. 3. Data received from the sensors needs to be decompressed and fed into the "Frequency-Segmentation" process. This stage marks dominated f-components as well as the segments that contain the components with the index m. Then the process "Relative Distance Information Calculate" computes 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 $\vec{\mathbf{p}}^{(m)}$. Ideally, the f-components that belong to the same source j should have the same position $\vec{\mathbf{p}}_j$. However, factors that affect the frequency images will influence the detection result and make f-components belonging to the same source j not have the same position but have the positions that are close to the real position of source j. Some factors do not much influence the frequency image, like the measurement error due to sampling resolution, the heat noise of the electric system and the DFT frequency leak. DFT frequency leak, which depends on the length of frame (window length), obviously changes the frequency image but is reduced considerably by Hamming window multiplication. Others factors are more serious like setting noise, Doppler effect and Rayleigh multipath fading effect. Setting noise, or background noise is usually very high and affect much the measurements. Meanwhile, Doppler effect causes the shifted f-component, which increases the probability of wrong grouping on frequency domain. Rayleigh multipath fading is even more serious than background noise since it affects directly on the regions that cover dominant f-components. With the same level of noise, Rayleigh multipath fading gives more changes on the frequency images than other kinds of noise. How background noise, Doppler effect and Rayleigh multipath fading effect influence the localization can be seen in simulation results of Sect. 5. Due to those factors, f-components of a source have positions that are close to the real source position. Therefore, the final stage "Source Positioning" is necessary to cluster those $\vec{\mathbf{p}}^{(m)}$ and estimate $\vec{\mathbf{p}}_j$ using the averaging mechanism. Details of the three main processes are described in the following subsections.

4.2.1 *Frequency segmentation*

"Frequency Segmentation" determines the frequency segments on which all shifted components of a dominant f-component are included. As can be seen in Fig. 2, data from the "Decompressing" block may have some missing and redundant f-components. Frequency segmenting is actually a clustering task which groups shifted components of an dominant f-component and determines the frequency segment based on the cluster. Doppler effect influences the f-components differently, the higher is the frequencies, the larger is the shift. From (3), 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. Indeed,

$$\Delta f \text{ (dB)} = \log_{10}\left(\frac{v_c}{v_c - v_j} f_0\right) - \log_{10}\left(\frac{v_c}{v_c + v_j} f_0\right) = \log_{10}\left(\frac{v_c + v_j}{v_c - v_j}\right), \quad (19)$$

Therefore, clustering is performed on the $\log_{10}(.)$ scale of the frequency image according to the criteria: (a) the width of each segment is not larger than Δf (dB) as defined in (19); (b) the number of nonzero f-components within the clustered 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 Δf (dB) is used here to detect the frequency intervals that satisfy (b) and (c). With such a mechanism, the number of sources can be larger than that of sensors. The total loss of some f-components due to filtering is acceptable because a source can be positioned with only one of its f-components. In addition, the redundant f-component will hardly be taken into account, and the groups with missing data can still be considered for position estimating. The maximum number of sources now depends on the sampling frequency, the speeds of the sources and the number of dominant f-components in each source. If the minimum frequency of all f-components is $f_{\min}$ and the mean number of f-components in a source is $\overline{n_f}$, then the possible number of sources can be up to

$$M_{\max} = \frac{\left(\frac{\log_{10}\frac{F_s}{2} - \log_{10} f_{\min}}{\Delta f\ (\mathrm{dB})}\right)}{\overline{n_f}}, \tag{20}$$

where the numerator is the possible total number of f-components caused by all sources. If the maximum source velocity decreases, then the maximum number of sources increases. Obviously, the possible number of sources is not related to the number of sensors since the locations of sources are determined by their dominant f-components. The accuracy is improved with a large number of sensors but it is not if the number of sources is changed, providing that the shifted versions of dominant f-components of sources do not interfere with each other on the frequency domain.

4.2.2 F-component Positioning

All of the ratios r_{ilz} of different pairs of distances from a source to sensors are calculated (see (17)) in "Relative Distance Information Calculating" process before being fed into the "F-component Positioning" process. Each ratio defines a constraint, or a position curve to which the source belongs. As illustrated in Fig. 3, if $r_{ilj}^{(m)} = 1$, the curve is the perpendicular bisector of the line segment connecting sensor i and sensor l; otherwise, the curve is a circle.

The additional noise and spectral leakage always exist in the recorded data and cannot be completely removed. Moreover, when a source is very close to a sensor, its signal may dominate at this sensor but be at a level lower than noise level at other sensors producing large errors in the constraints. Therefore, the error in the constraints is unavoidable, and the solution for the position of f-component m should be a vector $\vec{\mathbf{p}}^{(m)}$, $\vec{\mathbf{p}}^{(m)} \in \vec{\mathbf{R}}^2$ that compromises the constraints. We propose an objective function in a quadratic sum form for this compromise

$$\mathbf{F}_j = \sum_{i}^{N} \sum_{l,l\neq i}^{N-1} (d_{ij} - r_{ilj} d_{lj})^2, \quad 0 < r_{ilj} < \infty \tag{21}$$

and the solution for source j will be

$$\vec{\mathbf{p}}^{(m)} = \underset{\vec{\mathbf{p}}^{(m)}}{\arg\min}\, \mathbf{F}_j. \tag{22}$$

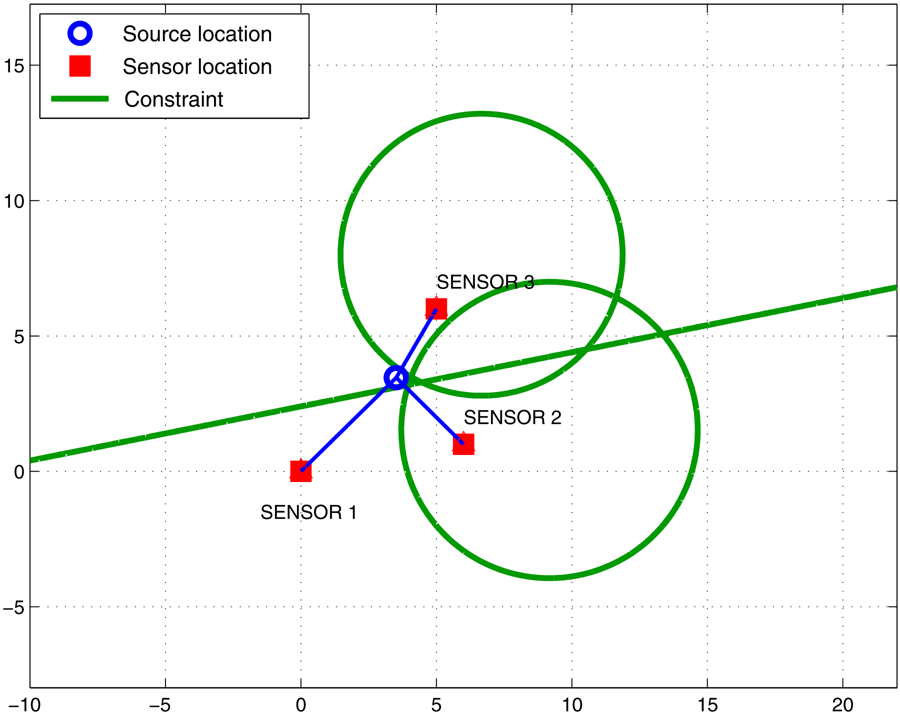


Fig. 3 An example of constraints obtained through a set of relative distance relations in which the number of involved sensors is 3

In fact, the objective function in (21) is not the quadratic form with respect to $\mathbf{p}^{(m)}$, but each constraint gives a quadratic form with respect to the minimum distance from $\mathbf{p}^{(m)}$ to the curve that satisfies $d_{ij} = r_{ilj}d_{lj}$. As depicted in Fig. 4, a combination of constraints may cause a valley whose bottom is slightly sloped. This situation takes the simple negative gradient method much time to converge because upon reaching the bottom of the valley, the search process oscillates back and forth on the sides with little progress to the optimum position. Based on the analysis in [22] (p. 149) and the results in [23], we choose the Fletcher–Reeves conjugate-gradient method for this optimization problem. The line search is performed using the Fibonacci segment search [22] and the first-order derivative used in the optimization is

$$\frac{d\mathbf{F}_j}{d\mathbf{p}^{(m)}} = \left[\sum_{i}^{N}\sum_{l,l\neq i}^{N-1}\left(2\left(\frac{(p_{jx}-m_{ix})}{d_{ij}} - \frac{r_{ilj}(p_{jx}-m_{lx})}{d_{lj}}\right)(d_{ij}-r_{ilj}d_{lj})\right)\right.$$
$$\left.\times\sum_{i}^{N}\sum_{l,l\neq i}^{N-1}\left(2\left(\frac{(p_{jy}-m_{iy})}{d_{ij}} - \frac{r_{ilj}(p_{jy}-m_{ly})}{d_{lj}}\right)(d_{ij}-r_{ilj}d_{lj})\right)\right]. \tag{23}$$

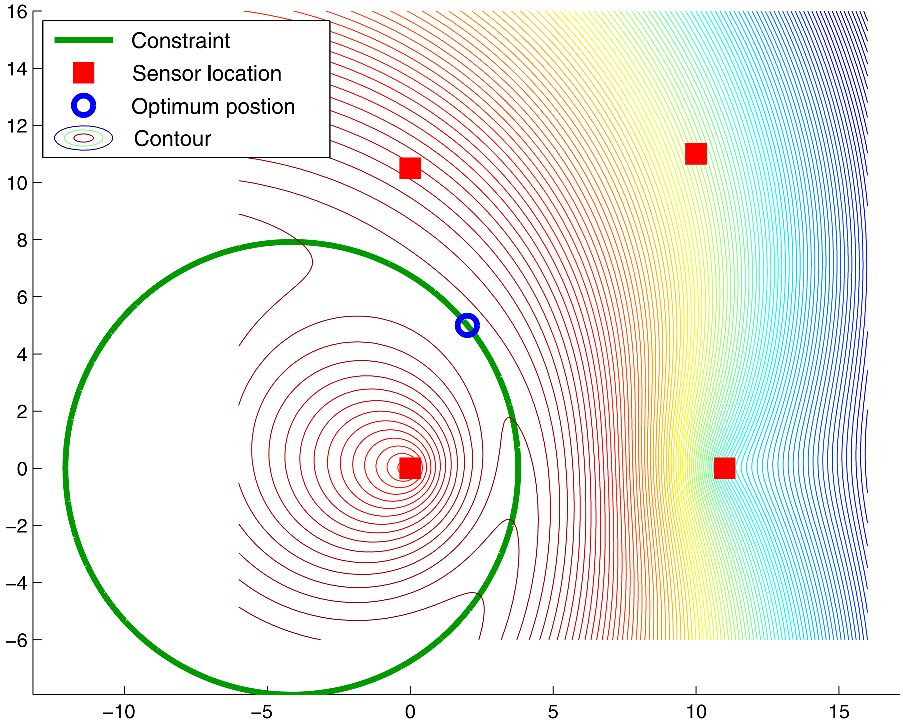


Fig. 4 Illustration of the rigid problem, one of the reasons that make gradient-based convergence slow

One remarkable note is that since our method is derived from the mutual information from every pair of active sensors, when the sensor number is greater than 3, the constraint number becomes much higher than the sensor number. For instance, if the sensor number is 5, then the constraint number is 10. The more constraints are there, the more accurate will be the results due to the averaging mechanism.

4.2.3 Source positioning

Data from the "F-component Positioning" stage includes the positions of the f-components. The source to which a specific f-component belongs must be next determined, and the answer can be found based on the estimated positions of these f-components. This stage clusters 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 [24] is used in this "Source Positioning" process with some modification compared to the typical version. The reason for this modification is that the resulted positions of different f-components have different levels of error depending on their energy to noise ratios. The higher is the energy, the more reliable is the resulted position. Therefore, each f-component position should be assigned with a weight. Then, instead of being randomly chosen as an input of d-min clustering, the positions are considered in the order of their weights. Moreover, the weights are also used to calculate the final source locations so that the computed f-component positions with higher energy play more important roles. The weight, $w^{(m)}$, is the mean values of the f-component's energies at the sensors. Equations (24) and (25) estimate the position of a source that includes the f-components whose positions are in the group j', denoted by $D_{j'}$:

$$\widehat{\mathbf{p}}_{j'} = \frac{\sum_{\mathbf{p}^{(m)} \in D_{j'}} w^{(m)} \mathbf{p}^{(m)}}{\sum_{\mathbf{p}^{(m)} \in D_{j'}} w^{(m)}}, \tag{24}$$

where

$$w^{(m)} = \frac{1}{N} \sum_{i}^{N} E_i^{(m)}. \tag{25}$$

$E_i^{(m)}$ is the energy of the f-component m received at sensor i. The distance from an f-component's position to a current group j' is the distance from it to the current estimated $\widehat{\mathbf{p}}_{j'}$.

5 Experiments and discussions

For the system working demonstration and the system evaluation, two main experiment sets are conducted via simulations in this section. They will be presented in details after we introduce the experiment setup and modeling in the next subsection.

5.1 Experiment setup and modeling

The monitored area is within [0 m, 12 m] × [0 m, 12 m]. We generate five simulated sources ($M = 5$), four of which imitate the sounds of different vehicles and motors while the rest mimics the sound of a siren (see Fig. 6). They are parametrically determined so that the received signals affected by Doppler effect can be generated properly using (1). These continuous sources have equivalent power levels, none have overlapped dominant f-components, and each is at least 5 meters from the others. It is reasonable to consider a group of close sources as one sound source; thus in order to illustrate how the system works, it is necessary for the sources to be distinct from one another. If the system functions properly when the sources are separated, then the source characteristics can be determined, memorized, and then used for future localization even when they are close to one another. The independence condition that requires no overlapped f-components among the sources is an elastic requirement. If the interference of other f-components is small, then it is considered as noise. In order to illustrate this point, we let one source, the second one in Fig. 6, have a wide spectrum which interferes into other frequency segments. The wave files of received signals are available on the website [25] as examples where the source speed is 40 km/h in 3 seconds at the highest noise level used in these simulation sets. Four sensors ($N = 4 < M$) are deployed around the corners of the deployed area in which the sources are set randomly (see Fig. 7). A source can move into and out of the monitored area during the monitoring time. The energies of the line-of-sight signals propagating to the sensors decrease according to the inverse square law at the sound speed of $c = 343$ m/s. The sampling frequency is $F_s = 16.384$ KHz and the time segment length is not less than 0.2 seconds. Because the sound signals of vehicles and motors are mainly composed by low frequency components, F_s and time segment T_f are set to be high so that different f-components can be separated as much as possible. The background noise is considered to be produced by the surrounding environment and the microphones, thus the noise is chosen to be Gaussian and its level is the same at all sensors.

In reality, received acoustic data always includes the effects of shadowing and fading due to multiple path reflections besides the received line-of-sight signals and Doppler effect. Therefore, we examine the situation under the existence of a Rayleigh fading channel and consider Rayleigh multipath signals as another kind of noise. Generating multiple paths for each source takes much computing time especially when Doppler effect is present, due to: (a) for each source, a large number of random paths are needed and for each path a new signal with respect to the direction of the path has to be generated, and (b) for each sensor, we have to generate different sets of simulated data as in (a). To reduce the simulation time for generating input data, the Young model [26], which generates a Rayleigh channel with two arrays of Gaussian random variables and the inverse-DFT (IDFT) technique, is applied. The model in [26] is for single f-component signals whereas the signals in this simulation set are multi f-component signals. Moreover, different f-components have different Doppler shift ranges and the Rayleigh energy in a range must be statistically proportional to the energy of the corresponding f-component. Therefore, in these simulations, after the ranges of Doppler shifts are determined, for each frequency bin within the range, a complex value is generated. The value's real and imaginary parts are Gaussian variables with the same standard deviation. 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 $R(k)$ must satisfy

$$\begin{cases} R(0) = 0, \\ R(k) = R^*(N-k), \quad k = 1, \ldots, N-1, \end{cases} \tag{26}$$

where the asterisk denotes the conjugate of the complex number. This makes the magnitude of the frequency frame $|R(k)|$ symmetrical over the dash line at $k = N/2$ (see Fig. 5). Figure 5 illustrates the generated Rayleigh fading signal on the frequency domain. The higher the frequency of the f-component, the wider is the Doppler shift range and the smaller are the frequency magnitudes within the spread range.

We increase the level of the total noise and choose the parameter Signal to Noise Ratio SNR for evaluating noise influence.

$$\begin{aligned} SNR &= \frac{E_{\text{mean}}}{E_{\text{noise}}} \\ &= \frac{E_{\text{mean}}}{\alpha_{\text{RL}} E_{\text{Rayleigh}} + (1-\alpha_{\text{RL}}) E_{\text{Background}}}, \end{aligned} \tag{27}$$

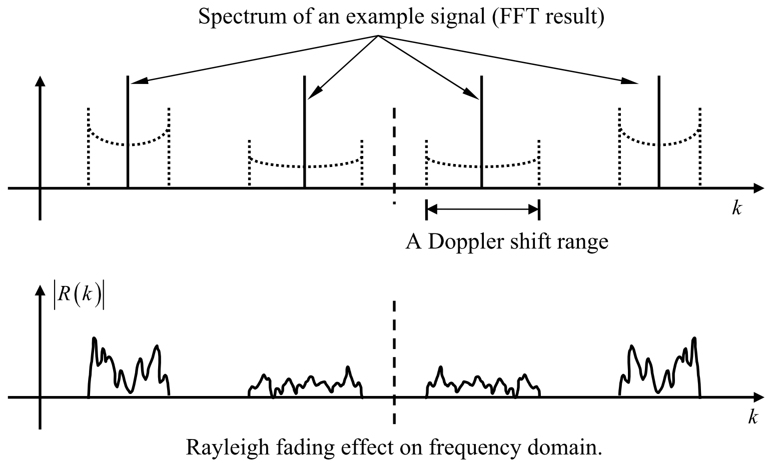


Fig. 5 Each bin within a Doppler shift range is generated with two Gaussian random variables to produce Rayleigh fading effect

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 $E_{\text{Background}}$ and the Rayleigh noise E_{Rayleigh}, while $\alpha_{\text{RL}} \in (0, 1)$ represents the percentage of Rayleigh noise energy in the total noise energy.

5.2 System working demonstration

Four sensors in the deployed area record the signals transmitted from fives sources, as described in Sect. 5.1. The processing steps of the system are basically consistent with those previously mentioned. The recorded data is sampled and segmented into overlapped frames. In this paper, we do not concentrate on continuously monitoring and refining the position estimations but only on positioning. Therefore, overlapped frames are not meaningful in these simulations. Each frame is multiplied by a Hamming window whose length is equal to the frame length (see (18)). DFT transformation is applied to convert data to the frequency domain in which the Gaussian noise filtering occurs. The filtering task can be conducted by detecting the maximum value of the frequency magnitude in the high frequency region having no f-components from the sources. Another simple method to detect Gaussian noise level is to determine the maximum magnitude of f-component on the frequency domain in advance when no sources are monitored. After the "F-component Positioning" stage, each f-component's position is determined and plotted with a circle. The radius of the circle is proportional to the f-component's weight or to the mean energy that the sensors receive from that f-component. Those f-components whose estimated positions are close to one another are grouped together as described in Sect. 4.2.3, in which $d_{\min} = 3.5$ m (new cluster is generated if an f-component's location is more than $d_{\min}$ from all others).

Figure 7 demonstrates the results of source positioning when the system tries to localize five independent sources in the time segment T_f of 0.2 seconds and the

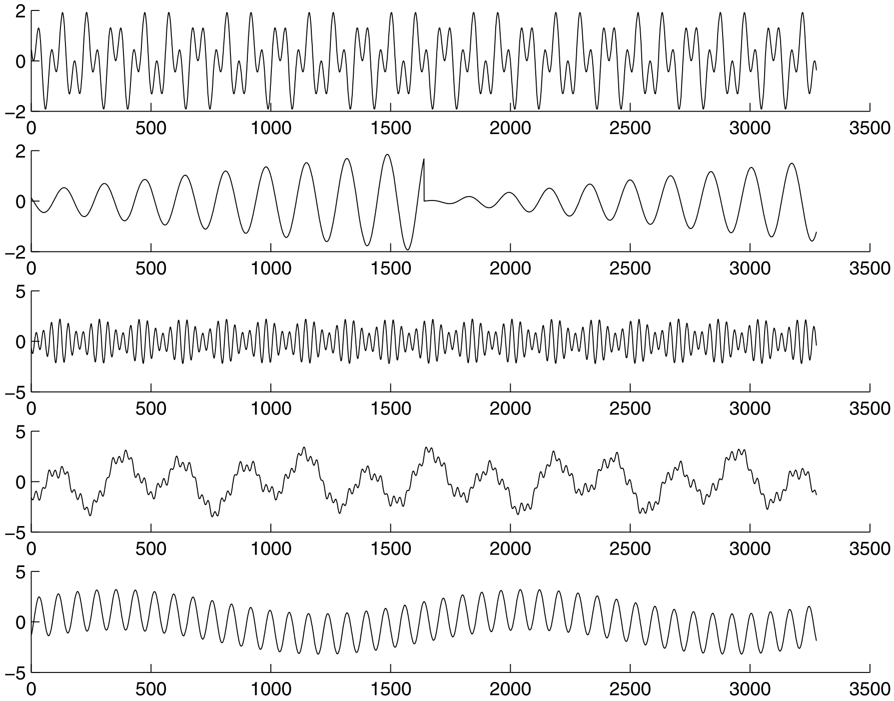
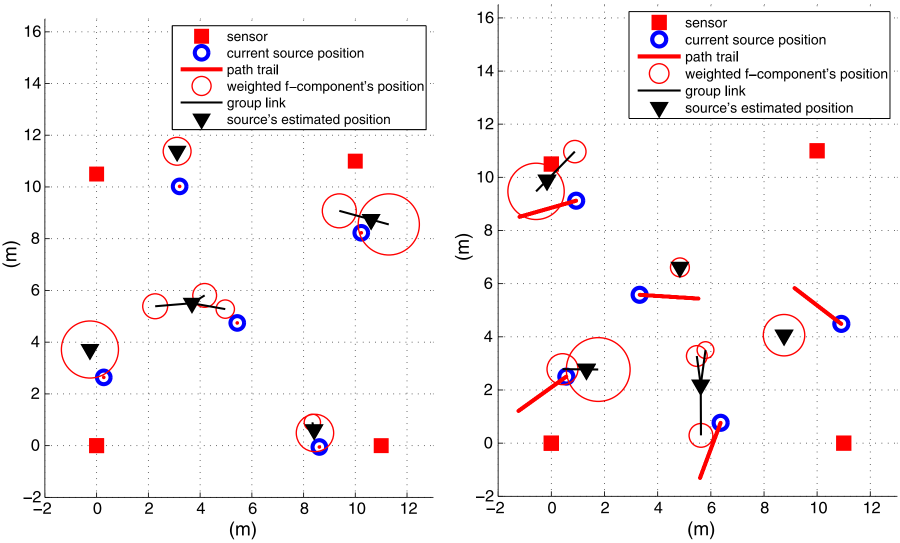


Fig. 6 Signals of sources which are parametrically determined to imitate sounds of vehicles, motors and a siren

Rayleigh fading noise energy contributes 20% of the total noise energy. The path trails that the sources leaves within the time segment of 0.2 seconds are denoted with the bold lines. The estimated positions of the sources are calculated based on the groups' f-component positions and their weights according to (24) and (25). Figure 7a shows the position estimation results when the speeds of the sources are zero, so the path trails are the small dots. Meanwhile, Fig. 7b presents the position estimation result when the speeds of the sources are all 40 km/h. It can be seen that the system is able to locate the sources even when the source number is greater than the sensor number. The Root Mean Square Errors (RMSEs) for the successful clustering in both Figs. 7a and b are less than 1.6 meters, an acceptable level especially when the speeds of the sources are high (in 0.2 s, the trail lengths are around 2.2 m). Obviously, positioning based on f-component localization is a good method to deal with multiple acoustic source positioning in situations affected by high Gaussian noise, multipath fading, and Doppler effect. The *SNR* here is 2.51, the highest level of all simulations.

To demonstrate the results of clustering, Table 1 shows the percentages of the group number resulted after f-component positions are clustered in the same condition as in the above example ($T_f = 0.2$ s, $\alpha_{\mathrm{RL}} = 0.2$). This task is repeated for 1,000 trials for each pair of total noise level and source speed. Since we focus on statistically evaluating the performance of the system during positioning, we do not try to

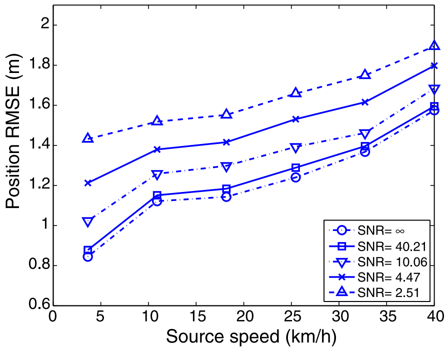


(a) (b)

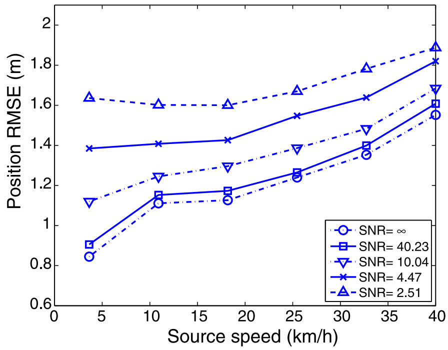
Fig. 7 (**a**) and (**b**) are the estimation results of two examples using the highest level of noise in the simulation set and $\alpha_{\mathrm{RL}} = 0.2$. One (**a**) locates five sources with source speeds of zero and the other (**b**) positions five sources with source speeds of 40 km/h

Table 1 Percentages (%) of resulting clusters when $T_f = 0.2$ s and $\alpha_{\mathrm{RL}} = 0.2$

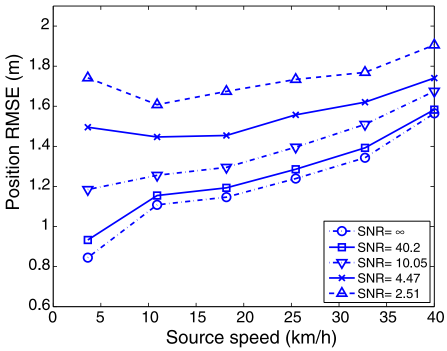
SNR	Speed (km/h)					
	0	8	16	24	32	40
Percentages of resulting into 5 clusters						
∞ (no noise)	91.7	85.9	86.8	84.3	84.1	86.8
40.21	89.5	83.2	80.6	83.5	85.1	81.6
10.05	86.5	79.3	77.7	74.4	79.0	76.8
4.46	72.6	71.5	67.1	67.5	67.8	69.4
2.51	66.3	64.8	63.7	62.6	63.2	64.8
Percentages of resulting into 4 clusters						
∞ (no noise)	6.4	9.2	10.2	11.9	11.7	8.9
40.21	8.4	12.2	15.3	12.4	10.5	13.1
10.05	10.0	14.4	16.4	16.7	14.1	14.0
4.46	17.8	18.5	18.3	19.7	19.1	17.6
2.51	19.4	20.9	21.0	19.4	19.34	19.6
Percentages of resulting into 6 clusters						
∞ (no noise)	1.9	4.9	2.7	3.6	4.1	4.2
40.21	2.1	4.6	3.9	3.8	4.1	4.8
10.05	2.9	5.0	4.8	7.2	6.4	8.2
4.46	9.1	8.8	13.5	11.4	11.0	11.5
2.51	12.6	13.0	13.7	16.3	15.1	14.2



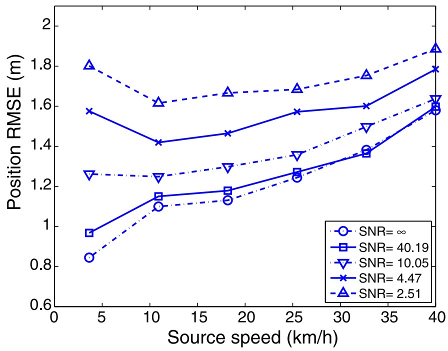
(a) $\alpha_{RL} = 0.2$



(b) $\alpha_{RL} = 0.4$



(c) $\alpha_{RL} = 0.6$



(d) $\alpha_{RL} = 0.8$

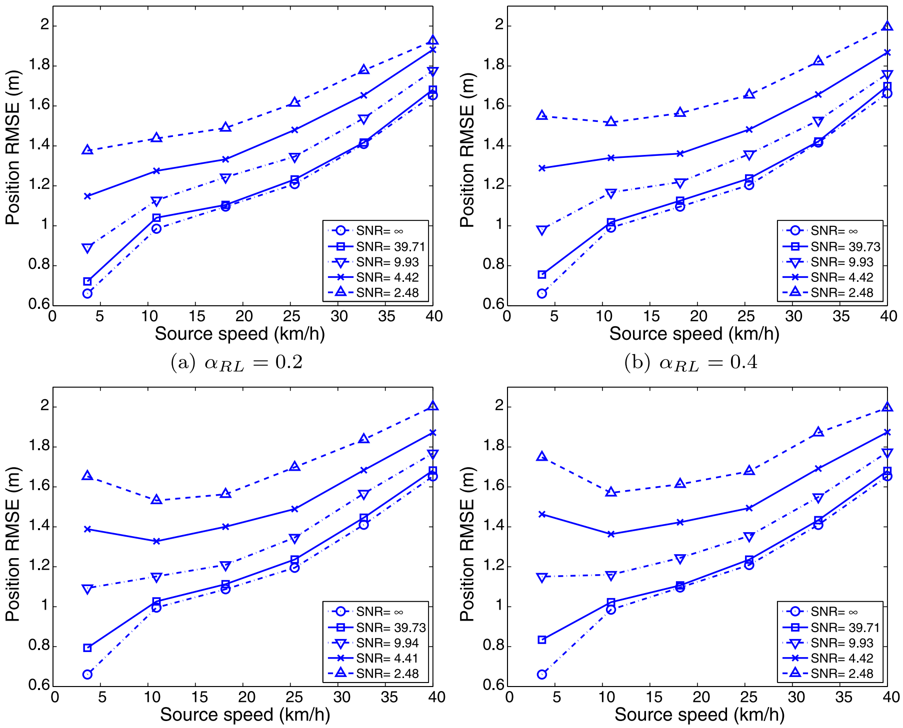
Fig. 8 RMSE error results of source locations under influences of speeds of sources, Gaussian noise, and Rayleigh multipath fading when the time segment is 0.2 s

monitor the sources in sequences of frames or to refine the location estimations over time. Instead, for each trial, we randomly set the positions and the moving directions of the sources and let the system perform positioning with only one frame. As one can expect, the percentages of exactly grouping into 5 sources decreases with the rise of noise and source speed, while the percentages of resulting in 4 and 6 groups after clustering increase. However, in the worst case, the percentage of clustering into 5 groups is still high, not less than 62%.

5.3 System performance evaluation

This simulation aims at examining the influences of time segment T_f, signal-noise ratio *SNR* and speed v_j on the results of positioning. The chosen T_f is either 0.2 s or 0.25 s, while the speed value varies from 0 to 40 km/h in increments of 8 km/h. The signal-noise ratios are generated based on the linear increment of the standard deviation of Gaussian noise. The set of *SNR* values are maintained statistically the same for the purpose of comparison, even when T_f and α_{RL} vary.

Figure 8 shows the results of distance errors under influences of noise level *SNR*, the percentage of Rayleigh multipath fading noise α_{RL} and the speed of sources v_j when the time segment is 0.2 seconds. Meanwhile, Fig. 9 illustrates the results when



(a) $\alpha_{RL} = 0.2$ (b) $\alpha_{RL} = 0.4$ (c) $\alpha_{RL} = 0.6$ (d) $\alpha_{RL} = 0.8$

Fig. 9 RMSE error results of source locations under influences of speeds of sources, Gaussian noise, and Rayleigh multipath fading when the time segment is 0.25 s

the time segment is 0.25 seconds. Each plotted error is the average value of RMSEs of 1,000 trials. At each trial, the RMSE is computed from the shortest distances between the estimated source positions and the actual source positions. When the number of estimated positions is different from that of actual positions, the number of distance pairs used to compute RMSE is the smaller number.

Table 2 shows the percentage results for clustering into 5 groups, 4 groups, and 6 groups at different conditions of noise levels and different source speeds. It is constructed when the $T_f = 0.2$ s and $\alpha_{\mathrm{RL}} = 0.8$, corresponding to Fig. 8d. Meanwhile, Table 3 shows the percentage results of clustering with the same format as in Tables 1 and 2 when the $T_f = 0.25$ s and $\alpha_{\mathrm{RL}} = 0.8$, corresponding to Fig. 9d.

From the results, it can be seen that at any speed value, higher noise and higher velocity lead to higher source positioning error. 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 path trails, increasing the uncertainty of positions. Consequently, the error in the f-component positions affects the f-component clustering results, causing error in the final source position estimations.

The relationships between v_j and RMSE can be considered to be positive linear when v_j is high (see Figs. 8 and 9). It is obvious since higher speed sources

Table 2 Percentages (%) of resulting clusters when $T_f = 0.2$s and $\alpha_{\mathrm{RL}} = 0.8$.

SNR	Speed (km/h)					
	0	8	16	24	32	40
Percentages of resulting into 5 clusters						
∞ (no noise)	91.7	86.3	85.7	85.0	84.9	86.1
40.19	86.8	84.1	81.3	82.1	81.7	83.2
10.05	73.4	75.4	74.0	73.8	74.0	76.6
4.47	58.1	65.1	61.9	66.3	67.0	65.2
2.51	52.5	55.3	58.5	60.3	61.1	58.4
Percentages of resulting into 4 clusters						
∞ (no noise)	6.4	9.2	10.8	10.6	10.9	10.0
40.19	11.0	11.9	14.3	13.3	13.8	12.4
10.05	19.5	18.4	18.3	19.3	19.0	17.9
4.47	25.0	23.4	27.4	22.0	21.4	22.8
2.51	24.9	27.0	25.1	23.3	24.9	25.7
Percentages of resulting into 6 clusters						
∞ (no noise)	1.9	4.3	3.3	4.2	3.7	3.6
40.19	2.0	3.6	3.7	4.5	4.3	4.2
10.05	5.9	5.1	6.7	5.7	6.4	4.5
4.47	13.1	9.5	8.2	9.7	8.9	9.7
2.51	16.7	13.6	13.0	12.9	10.6	11.7

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 subfigures). The accuracy of estimated f-component positions decreases when the source velocities increase, leading to the increased error in the source location estimations. On the other hand, when the *SNR* is high and v_j is low, the accuracy decreases rapidly because of Rayleigh multi-path fading. In the case where $SNR = 2.5$ and $\alpha_{\mathrm{RL}} = 0.8$, the RMSE error is up to around 1.8m (see Figs. 8d and 9d). 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. The fading noise can also be partly filtered by the Gaussian filtering stage. Therefore, Rayleigh fading noise in high v_j cases does not affect the accuracy as much as it does when v_j is low. Generally, the results in Figs. 8 and 9 show that the more Rayleigh fading noise contributes to the total noise level, the worse result the system achieves especially if the source speeds are low.

One can easily notice that the system could not give the ideal result when there is no noise ($SNR = \infty$). The f-component location errors are caused by a limited time segment of a frame which produces unavoidable spectral leakage. In addition, there exists the influence of the second original source, whose f-components ap-

Table 3 Percentages (%) of resulting clusters when $T_f = 0.25$s and $\alpha_{\mathrm{RL}} = 0.8$.

SNR	Speed (km/h)					
	0	8	16	24	32	40
Percentages of resulting into 5 clusters						
∞ (no noise)	94.3	85.3	86.7	87.1	88.4	89.7
39.71	90.2	82.3	86.0	84.8	86.1	87.7
9.93	70.3	75.8	77.5	76.5	78.6	77.7
4.42	60.4	67.3	67.2	68.6	69.2	68.2
2.48	50.1	57.5	59.4	61.0	63.4	65.0
Percentages of resulting into 4 clusters						
∞ (no noise)	5.6	13.3	11.8	11	9.1	8.8
39.71	8.8	15.3	12.1	13.2	11.0	9.9
9.93	19.5	18.2	16.1	17.1	14.9	16.3
4.42	21.3	19.8	20.3	21.3	18.4	21.5
2.48	23.2	21.3	22.8	20.5	21.3	20.6
Percentages of resulting into 6 clusters						
∞ (no noise)	0.1	1.2	1.3	1.7	2.1	1.4
39.71	1	1.9	1.7	1.4	2.8	2.3
9.93	8.2	5.6	5.6	5.5	5.2	5.7
4.42	14.5	10.6	10.8	8.7	10.4	9.3
2.48	22.6	17.7	14.9	15.7	12.1	10.9

pear and contribute interference all over the frequency domain, to other sources' f-components. As a consequence, the errors in final estimations are unavoidable. Nevertheless, when the noise level increases quickly, the system can tolerate the noise well with little error increment. The estimation error increases an amount of 0.6 m in average when the *SNR* decreases 16 times from around 40 to 2.5. Comparing subfigures by pair from Figs. 8 and 9, we can see that a bigger time segment T_f improves the accuracy when v_j is low (including the ideal case). However, it decreases the accuracy if v_j is high. Obviously, although bigger T_f reduces the spectral leakage, this improvement cannot compensate the increased error when the path trails are longer, producing high error results when v_j is high (>30 km/h).

Tables 2 and 3 show that f-component clustering relied on f-component locations is closely associated with the estimation accuracy. An incorrect number of clusters is caused by missing f-component locations due to noise filtering or by using too poor the f-component position estimations. Higher percentage of grouping into exactly 5 clusters increases the estimation accuracy. Contrariwise, this percentage decreases at high levels of noise and high source speeds.

Comparing Tables 1 and 2 both of which result from the same T_f of 0.2 s, one can see 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. This is consistent with the above analysis of Rayleigh fading influence on the RMSE. Meanwhile, comparing Tables 2 and 3 shows that in most of the cases, longer monitoring time reveals higher

Table 4 Total percentages (%) of clustering 4 groups, 5 groups and 6 groups

SNR	Speed (km/h)					
	0	8	16	24	32	40
			$T_f = 0.2s;\ \alpha_{\mathrm{RL}} = 0.2$			
∞ (no noise)	100.0	100.0	99.7	99.8	99.9	99.9
39.71	100.0	100.0	99.8	99.7	99.7	99.5
9.93	99.4	98.7	98.9	99.4	99.5	99.0
4.42	99.5	98.8	98.9	98.6	97.9	98.5
2.48	98.3	98.7	98.4	98.3	97.6	98.6
			$T_f = 0.2s;\ \alpha_{\mathrm{RL}} = 0.8$			
∞ (no noise)	100.0	99.8	99.8	99.8	99.5	99.7
39.71	99.8	99.6	99.3	99.9	99.8	99.8
9.93	98.8	98.9	99.0	98.8	99.4	99.0
4.42	96.2	98.0	97.5	98.0	97.3	97.7
2.48	94.1	95.9	96.6	96.5	96.6	95.8
			$T_f = 0.25s;\ \alpha_{\mathrm{RL}} = 0.8$			
∞ (no noise)	100.0	99.8	99.8	99.8	99.6	99.9
39.71	100.0	99.5	99.8	99.4	99.9	99.9
9.93	98.0	99.6	99.2	99.1	98.7	99.7
4.42	96.2	97.7	98.3	98.6	98.0	99.0
2.48	95.9	96.5	97.1	97.2	96.8	96.5

clustering accuracy and higher percentage of clustering into 5 groups. This is because most f-components of vehicle and motor sources appear in the low frequency range, so a longer time segment helps separate the f-components more clearly and decrease the spectral leakage. However, increased T_f makes the final estimation results at high source speeds not as good as those at low source speeds (see Figs. 8 and 9) since the path trails at high v_j are longer. This reflects the uncertainty law between time and frequency: low frequency needs a long observation time to be indicated. Obviously, there is a trade-off between f-component separation and the path trail length when T_f varies.

Now we consider the harsh conditions, when data is collected from high speed sources under serious noise and when low speed sources move in the setting where Rayleigh fading noise accounts for most of the total noise. Although the percentage of clustering into 5 groups is quite low (in the worse case, only 50.1%), the percentage of clustering results for all 4, 5, and 6 groups are very high. Table 4 displays the high values of total percentage of clustering for 4, 5, and 6 groups in three critical conditions in this simulation set (the least is 94.1%, see Table 4), indicating that a good inference mechanism module can be used in conjunction with this positioning system to combine the information of cluster number and the relationships between f-components and clusters. The continuous combination over time can better clustering results, measure the velocities of the sources and refine the position estimations by predicting their next positions. This module, however, requires much effort in both analysis and design, and is beyond the scope of this article.

6 Conclusions

This paper introduces a distributed system for independent acoustic source localization in parallelism of *data-and-process decomposition*. The information for separation is the ratios of f-component energy values received at the sensors. In order to obtain these ratios, the received signals are first segmented into frames and transformed to frequency frames with DFT transformation at the sensors. The Gaussian noise filtering and compressing also take place at the sensors before the preprocessed data is sent to the base computer. At the base, after decompressed, the data is used for f-component clustering, or determining the segments that contain the different f-components. All of the relative distance ratios are computed in order to establish an objective function for each f-component. Minimizing these objective functions reveals the location estimations of the f-components. A d-min clustering approach is used to group close f-component positions and to compute the final source locations. The simulations in the article have been made as realistic as possible with the coexistence of Doppler effect and Rayleigh multipath fading to illustrate how the system solves the problem of multiple moving source positioning. The results show that the method gives high accuracy and requires a very low communication cost for a large data set. The proposed system can be considered as the acoustic source localization design for the future generation of WSNs since it needs sensors with high computing ability in order to perform DFT on a long segment of data. However, powerful computing ability is not crucial because using 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 containing major f-components. The system is actually more useful than just positioning multiple independent acoustic sources. It can also provide characteristics of the acoustic sources which can be utilized for further position estimation refinement and for recognition or classification since most acoustic features are related to the frequency domain. As the task of continuously positioning and refining the estimations of source locations requires much effort in both analysis and simulation, we let it be our next work in future.

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

References

1. Arabnia Hamid R (1990) A parallel algorithm for the arbitrary rotation of digitized images using process-and-data-decomposition approach. J Parallel Distrib Comput 10:188–192
2. Sasaki Y, Kagami S, Mizoguchi H (2006) Multiple sound source mapping for a mobile robot by self-motion triangulation. In: International conference on intelligent robots and systems, IEEE, pp 380–385
3. Stanacevic Milutin (2005) Micropower gradient flow acoustic localizer. IEEE Trans Circuits Syst 52:2148–2157
4. Birchfield ST, Gilmor DK (2001) A unifying framework for acoustic localization. In: IEEE international conference, IEEE Comput Soc, Los Alamitos
5. Mahajana A, Walworth M (2001) 3-d position sensing using the differences in the time-of-flights from a wave source to various receivers. IEEE Trans Robot Autom 17:91–94
6. Valin J-M, Michaud F, Rouat J, L´etourneau D (2003) Robust sound source localization using a microphone array on a mobile robot. In: Proc ICASSP98, pp 1228–1233

7. Antonacci F, Riva D, Saiu D, Sarti A, Tagliasacchi M, Tubaro S (2006) Tracking multiple acoustic sources using particle filtering. In: European signal processing conference, EUSIPCO-2006
8. Kaouri K Left-right ambiguity resolution of a towed array sonar (2000) Ph.D dissertation, M.Sc. in mathematical modelling and scientific computing
9. Ward DB, Williamson RC (2002) Particle filter beamforming for acoustic source localization in a reverberant environment. In: IEEE international conference on acoustics, speech, and signal processing, IEEE Press, New York
10. Roman N, Wang DL (2008) Binaural tracking of multiple moving sources. IEEE Trans Speech Audio Process 16:728–739
11. Gisolf A, Simons DG (2006) Adaptive motion compensation in sonar array processing, 1st edn. Netherlands Organization for Applied Scientific Research (TNO)
12. Hyvarinen A, Oja E (2000) Independent component analysis: a tutorial. Neural Netw 13:4–5
13. Mutihac R, Van Hulle MM (2004) Comparison of principal component analysis and independent component analysis for blind source separation. IEEE Trans Speech Audio Process 56:20–32
14. Smaragdis P (1998) Blind separation of convolved mixtures in the frequency domain. Neurocomputing 22:21–34
15. Lambert RH (1996) Multichannel blind deconvolution: Fir matrix algebra and separation of multipath mixtures. Ph.D. dissertation, University of Southern California, EE Dept
16. Dang V-H, Le-Tien T, Lee Y-K, Lee SY (2010) Acoustic multiple object positioning system. In: International symposium on performance evaluation of wireless ad hoc, sensor, and ubiquitous networks, PE-WASUN-2010. ACM, New York
17. Ciaramella A, Tagliaferri R (2006) Separation of convolved mixtures in frequency domain ica. Int Math Forum 16:769–795
18. Birchfield ST, Gilmor DK (2002) Fast Bayesian acoustic localization. In: IEEE international conference on acoustics, speech, and signal processing, IEEE Press, New York
19. Serway R, Beichner R (1999) Physics for scientists and engineers, 5 edn. Brooks/Cole, Pacific Grove
20. Haddad RA, Parsons TW (1991) Digital signal processing theory, applications and hardware. Computer Science Press, New York
21. Van Den Enden AWM, Verhoeckx NAM (1989) Discrete-time signal processing: an introduction, 1 edn. Prentice Hall, New York
22. Wismer DA, Chattergy R (1979) Introduction to nonlinear optimization. North-Holland, Amsterdam
23. Van Trier J Some nonlinear optimization methods (http://sepwww.stanford.edu/public/docs/sep51/)
24. Duda RO, Hart PE, Stork DG (2001) Pattern Classification., 2nd edn. Wiley, New York
25. Dang V-H Generated examples of multi-object sound data (http://uclab.khu.ac.kr/ext/dvhung/)
26. Young DJ, Beaulieu NC (2000) The generation of correlated Rayleigh random variates by inverses discrete Fourier transform. IEEE Trans Commun 48:1114–1127