Acoustic Source Localization and Discrimination in Urban Environments

Manish Kushwaha, Xenofon Koutsoukos, Peter Volgyesi and Akos Ledeczi
Institute for Software Integrated Systems (ISIS)
Department of Electrical Engineering and Computer Science
Vanderbilt University
Nashville, TN 37235, USA
manish.kushwaha@vanderbilt.edu

Abstract − *Collaborative localization and discrimination of acoustic sources is an important problem for monitoring urban environments. Acoustic source localization typically is performed using either* signal-based *approaches that rely on transmission of raw acoustic data and are not suitable for resource-constrained wireless sensor networks or* feature-based *methods that result in degraded accuracy, especially for multiple targets. In this paper, we present a feature-based localization and discrimination approach for multiple acoustic sources using wireless sensor networks that fuses beamform and power spectral density data from each sensor. Our approach utilizes a graphical model for estimating the position of the sources as well as their fundamental and dominant harmonic frequencies. We present simulation and experimental results that show improvement in the localization accuracy and target discrimination. Our experimental results are obtained using motes equipped with microphone arrays and an onboard FPGA for computing the beamform and the power spectral density.*[1].

Keywords: Acoustic source localization, wireless sensor networks, Bayesian estimation, feature-level fusion.

1 Introduction

Acoustic source localization is an important problem in many diverse applications such as military surveillance and reconnaissance, underwater acoustics, seismic remote sensing, communications, and environmental and wildlife habitat monitoring. Recently, more innovative applications such as smart video-conferencing, multimodal sensor fusion and target tracking have been proposed to utilize multimodal source localization.

In wireless sensor networks (WSNs), collaborative source localization is needed, where the objective is to estimate the positions of multiple sources by fusion of observations from multiple sensors. There are two broad classes of methods for collaborative source localization. The first class of approaches, where the estimation is done by fusion of the sampled signals, is called *signal-based*, or *signal-level fusion* methods. The second class of approaches, where signal features are extracted at each sensor and estimation is done by fusion of the extracted features collected from all the sensors, is called *feature-based*, or *feature-level fusion* methods.

In this paper, we present a feature-level fusion method for collaborative source localization of multiple acoustic sources in WSNs. We use microphone arrays as sensors that compute beamforms and estimate power spectral density (PSD) as the signal features. The advantage of using the beamform over signal energy is that the beamform captures the angular variation of the signal energy, which results in better localization resolution. The use of PSD as another signal feature allows us to identify multiple sources under our harmonic signal assumption. The target tracking application in [1] demonstrated that the communication bandwidth available in sensor networks is sufficient to support wireless transmissions of such features.

Several signal-based methods using a microphone array for source localization have been proposed [2]. These approaches use time delay of arrival (TDOA) or direction of arrival (DOA) estimation, beamforming [2, 3] and maximum likelihood estimation [4]. The signal-based methods are not suited for WSNs because they require transmission of the raw signal. On the other hand, the feature-based methods are appropriate for WSNs due to its lower bandwidth and power requirements. An example of the feature-based method is energy-based localization (EBL), where signal energy is taken as the feature. EBL has been solved using various least squares [5, 6] and maximum likelihood [7] formulations. EBL suffers from poor localization resolution for multiple targets, where the resolution is defined as the ability of the localization algorithm to discriminate two closely spaced targets.

Localization algorithms based on least squares operate on a strict Gaussian noise assumption and are

[1]This work is partially supported by ARO MURI W911NF-06-1-0076.

not extensible to multiple sources; while those based on maximum likelihood are not extensible to tracking applications where data association across time becomes an issue. A Bayesian approach for source localization can handle non-Gaussianity and multiple sources, both stationary and moving. Several approaches based on graphical models [8] and Bayesian estimation [9, 10] have been proposed for multiple target localization and tracking. A graphical model based approach for audio-visual object tracking is presented in [8] that combines the audio and video data. A Bayesian approach for tracking the DOA of multiple targets using a passive sensor array is presented in [9]. Another Bayesian approach for multiple target detection and tracking for unknown number of targets, and a particle filter-based algorithms are proposed in [10].

An alternative to Bayesian statistics is Finite Set Statistics (FISST), which treats the multitarget state and multiple observations as finite sets. An approximate multitarget tracking approach based on FISST, called the Probability Hypothesis Density (PHD) filter, is proposed in [11].

We solve the localization problem using a graphical model that is generalization of Bayesian estimation. Graphical models provide a compact representation of joint probability density and facilitate the factorization of joint density into conditional densities [12]. Graphical models require generative models that describe the observed data in terms of the observation process and the hidden state variables. We present generative models for beamform and PSD data. The problem is divided into two steps as source separation and source localization. The idea is to separate the sources in frequency domain using the received PSD from the sensors, and then use the separated sources for localization. We propose a maximum likelihood (ML) estimation method for source separation and Bayesian estimation for localization.

We present simulation results for multiple source localization in a grid sensor network. Our algorithm is able to achieve an average of 25 cm localization error for three targets, 8 cm for two targets, and less than 5 cm for single target in a sensor network of four sensors. We are able to distinguish between two targets as close as 50 cm using our algorithm. Our results show that as the separation between targets increases, our algorithm is able to achieve higher localization accuracy, comparable to single target localization. Finally, we present results for an outdoor experimental setup with MICAz sensor nodes and real acoustic sources.

The rest of the paper is organized as follows. In Section 2, we present the acoustic source model and the acoustic sensor model. Section 3 describes the graphical model. Sections 4 and 5 describe the source separation and source localization, respectively. In Section 6, we present the Gibbs sampler and the initialization strategy. We present results for various simulation scenarios in Section 7, and the outdoor experiment setup and results in Section 8. We conclude in Section 9.

2 Source & Sensor Models

Consider a wireless sensor network of K acoustic sensors in a planar field. Each acoustic sensor is a microphone array with N_{mic} microphones each. Consider M far-field stationary acoustic sources coplanar with the sensor network. The acoustic wavefront incident on the sensors is assumed to be planar for far-field sources. Each sensor receives the signal and runs simple signal processing algorithms to compute the beamform and acoustic PSD. The goal of this work is to estimate the 2D position of all the sources given the beamform and the PSD from all sensors.

Acoustic Source Model The main assumptions made in this paper for acoustic sources are that they are (1) stationary point sources, (2) emitting stationary signals, (3) the source signals are harmonic, and (4) the cross-correlation between two source signals is negligible compared to the signal autocorrelations. Harmonic signals consist of a fundamental frequency, also called the first harmonic, and other higher-order harmonic frequencies that are multiples of the fundamental frequency. The energy of the signal is contained in these harmonic frequencies only. The harmonic source assumption is satisfied by a wide variety of acoustic sources [13]. In general, any acoustic signal originating due to the vibrations from rotating machinery will have an harmonic structure.

The state for the m^{th} acoustic source is given by, (1) the position, $\mathbf{x}^{(m)} = \left[x^{(m)}, y^{(m)}\right]^T$, (2) the fundamental frequency, $\omega_f^{(m)}$, and (3) the energies in the harmonic frequencies, $\psi^{(m)} = \left[\psi_1^{(m)}, \psi_2^{(m)}, \cdots, \psi_H^{(m)}\right]^T$, where H is the number of harmonic frequencies.

Acoustic Sensor Model The intensity of an acoustic signal emitted omni-directionally from a point sound source attenuates at a rate that is inversely proportional to the distance from the source [7]. The discrete signal received at the p^{th} microphone is given by

$$r_p[n] = \sum_{m=1}^{M} \frac{d_0}{\parallel \mathbf{x}_p - \mathbf{x}^{(m)} \parallel} s^{(m)}[n - \tau_p^{(m)}] + w_p[n] \quad (1)$$

for samples $n = 1, \cdots, L$, where L is the length of the acoustic signal, M is the number of sources, $w_p[n]$ is white Gaussian measurement noise such that $w_p[n] \sim \mathcal{N}(0, \sigma_w^2)$, $s^{(m)}[n]$ is the intensity of the m^{th} source measured at a reference distance d_0 from that source, and $\tau_p^{(m)}$ is the propagation delay of the acoustic signal from the m^{th} source to the p^{th} microphone. The microphone and source positions are denoted by $\mathbf{x}_p$ and $\mathbf{x}^{(m)}$, respectively. We define the multiplicative term in

Equation (1) as the attenuation factor, $\lambda_p^{(m)}$, given by, $\lambda_p^{(m)} = d_0 / \parallel \mathbf{x}_p - \mathbf{x}^{(m)} \parallel$.

Beamforming is a signal processing algorithm for DOA estimation of a signal source. In a typical delay-and-sum single source beamformer, the 2D sensing region is discretized into directions, or *beams* as $\alpha = i\frac{2\pi}{Q}$, where $i = 0, \cdots, Q-1$ and Q is the number of beams. The beamformer computes the energy of the reconstructed signal at each beam direction. This is acheived by delaying and summing the indiviual microphone signals. The beam energy is given by

$$B(\alpha) = \sum_{n=1}^{L} \left[\sum_{p=1}^{N_{mic}} r_p[n + t_{pq}(\alpha)] \right]^2 \qquad (2)$$

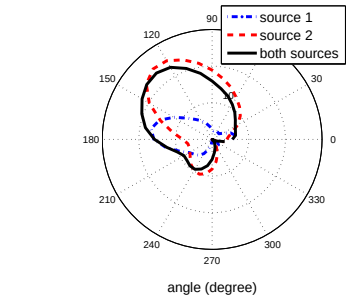
where α is the beam angle, $r_p[\cdot]$ is the received signal at the p^{th} microphone, q is the index of a reference microphone, and $t_{pq}(\alpha)$ is the relative time delay for the p^{th} microphone with respect to the reference microphone q, given by, $t_{pq}(\alpha) = d_{pq}cos(\alpha - \beta_{pq})f_s/C$, where d_{pq} and β_{pq} are the distance and angle between the p^{th} and q^{th} microphones, and f_s and C are signal sampling rate and speed of sound, respectively.

Beam energies are computed for each of the beams, and are collectively called the *beamform*. The beam with maximum energy indicates the DOA of the acoustic source. In case of multiple sources, there might be multiple peaks where the maximum peak would indicate the DOA of the highest energy source. Figure 1(a) shows a beamform for two acoustic sources. Advances in sensor network hardware and and FPGA integration has allowed us to implement real-time beamforming on MICAz sensor motes [1].

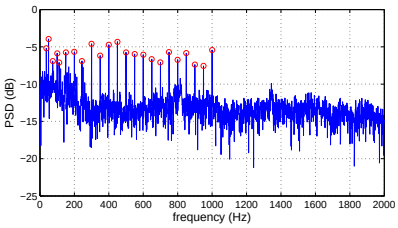
Acoustic PSD estimation is the estimation of the spectrum of the received acoustic signal, which describes how the power of the signal is distributed with frequency. We estimate the PSD as the magnitude of the discrete Fourier transform (DFT) of the signal. The PSD estimate can be written as

$$P(\omega) = Y(\omega) \cdot \overline{Y(\omega)} \qquad (3)$$

where $Y(\omega) = \mathbb{FFT}(r, N_{FFT})$ is the discrete Fourier transform of the signal $r[n]$, N_{FFT} is the length of the transform, and $\overline{Y(\omega)}$ is the complex conjugate of the transform. For real-valued signals, the PSD is real and symmetric, hence we need to store only half of the spectral density. In our implementation, we represent the spectral densities as the frequency–power pairs, (ω_j, ψ_j), for the N_{PSD} frequencies with the highest power values. Figure 1(b) shows an acoustic PSD estimate for a received signal when two harmonic sources are present.



(a)



(b)

Figure 1: (a) Acoustic Beamforms; The beamforms for single sources clearly show peaks at the source location but the beamform when both sources are present does not show two peaks. (b) Power spectral density (PSD); The highest PSD values are shown as empty circles. The PSD is compactly represented as pairs of the highest PSD values and corresponding frequencies.

3 Graphical Model Overview

Source separation and localization of multiple sources is performed using the graphical model shown in Figure 2. The nodes with clear backgrounds denote hidden

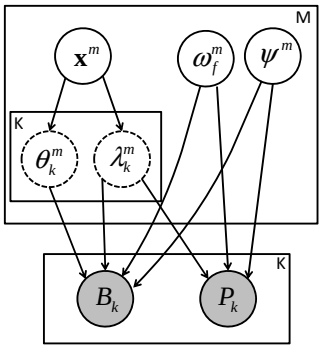


Figure 2: Graphical model (plate notation is used to represent the repetition of random variables).

state variables; $\mathbf{x}^{(m)}, \omega_f^{(m)}, \psi^{(m)}$ denote source position, fundamental frequency and harmonic energies for the m^{th} source, respectively. The nodes with shaded backgrounds denote observed variables; B_k and P_k denote the beamform and the PSD received at k^{th} sensor, respectively. Finally, the nodes with dotted outlines denote functions of random variables, or *auxiliary random*

variables, that capture the functional dependence of the observed variables on the hidden variables. These variables will be utilized in the generative models for the observed variables. The two auxiliary variables shown in the graphical model are the angle $\theta_k^{(m)}$ and the attenuation factor $\lambda_k^{(m)}$.

We perform multiple source localization in two steps. First, we use the PSD data only to separate the sources. Source separation, in our problem, refers to separating the PSDs of the sources. For harmonic sources, estimation of fundamental frequencies is sufficient for source separation, because all the dominant frequencies in the signal are multiples of the fundamental frequency. A ML estimation method is used for fundamental frequency estimation. The ML estimate is independent of the source location, which is intuitive because the dominant frequencies in the source signal are independent of the source location, as long as the source and sensor are stationary. We will support this intuition in next section via derivation of ML solution. In second step, we use the beamform data and the *separated source PSDs* to localize all sources.

We chose to perform multiple source localization in two steps instead of joint estimation because of the following two reasons, (1) estimation in two steps has lower computational complexity than joint estimation, and (2) during the simulations, we realized that the likelihood sensitivity for source separation and localization are different. In Monte Carlo context, joint estimation might cause slower convergence, and may require a large number of samples. Moreover, as mentioned earlier, the ML estimate for source fundamental frequencies is independent of the source locations.

4 Source Separation

An ML estimation method is presented for source separation. We begin by presenting the generative model and likelihood function for PSD data. We also present a result showing that the likelihood function at ML estimate of harmonic energies is independent of source positions.

Generative Model for PSD Data. For harmonic sources, the PSD can be given by

$$P_s^{(m)}(\omega) = \sum_{h=1}^{H} \psi_h^{(m)} \delta(\omega - h\omega_f^{(m)}) \tag{4}$$

where $m = 1, \cdots, M$ are source indices, ω is the frequency, $\omega_f^{(m)}$ is the fundamental frequency, $\psi_h^{(m)}$ is the energy in the h^{th} harmonic, H is the number of harmonics, and $\delta(\cdot)$ is the Dirac delta function. Using Equation (4), we derive a generative model for the PSD data received at a sensor node. Following proposition states the generative model for the PSD data.

Proposition 1. *For an arbitrary number of acoustic source signals, the power spectral density of the signal received at a sensor is given by*

$$\mathbb{P}(\omega) = \sum_{m=1}^{M}\sum_{n=1}^{M} \lambda^{(m)}\lambda^{(n)} \left(P_s^{(m)}(\omega)P_s^{(n)}(\omega)\right)^{\frac{1}{2}} \cos(\Phi^{(m)}(\omega) - \Phi^{(n)}(\omega)) \tag{5}$$

where M is the number of sources, $\lambda^{(m)}$ is the attenuation factor, and $\Phi^{(m)}(\omega)$ is the phase spectral density given by, $\Phi^{(m)}(\omega) = \phi^{(m)} - \parallel \mathbf{x}^{(m)} - \mathbf{x}_s \parallel \omega/C$, where $\phi^{(m)}$ is the phase of the source signal, $\mathbf{x}^{(m)}$ and $\mathbf{x}_s$ are the positions of the source and the sensor, respectively.

The proof for the proposition is given in [14]. Since we do not maintain the phase of the signal in the source model (see Section 2), we assume all the phases to be normally distributed with equal mean. The expected value of the cosine of the difference of two normally distributed angles is one, i.e. $E[cos(\Phi_i - \Phi_j)] = 1$. Using this, Equation (5) becomes

$$\mathbb{P}(\omega) = \left[\sum_{m=1}^{M} \lambda^{(m)} \left(P^{(m)}(\omega)\right)^{1/2}\right]^2 \tag{6}$$

Data Likelihood. Using Equation (6), the negative log-likelihood for PSD data at the kth sensor is defined as

$$\ell_k(\Omega_f, \Psi, \mathbf{X}) = \frac{1}{\sigma_P^2} \sum_{\omega_j} \parallel P_k(\omega_j) - \mathbb{P}_k(\omega_j) \parallel^2$$

where $\Omega_f = \left[\omega_f^{(1)} \cdots \omega_f^{(M)}\right]^T$, $\Psi = \left[\boldsymbol{\psi}^{(1)} \cdots \boldsymbol{\psi}^{(M)}\right]^T$, $\boldsymbol{\psi}^{(m)} = \left[\psi_1^{(m)} \cdots \psi_H^{(m)}\right]^T$, and $\mathbf{X} = \left[\mathbf{x}^{(1)} \cdots \mathbf{x}^{(M)}\right]^T$.

Proposition 2. *Likelihood for PSD data at ML estimate of harmonic energies is a function of source fundamental frequencies only and is independent of source positions. Mathematically,*

$$\ell_k(\Omega_f, \Psi^{ML}, \mathbf{X}) = \ell'_k(\Omega_f, \mathbf{X}) = \ell'_k(\Omega_f) \tag{7}$$

Proof. The maximum likelihood estimate of $[\Omega_f, \Psi, \mathbf{X}]^T$ can be obtained by minimizing the likelihood, $\partial\ell_k(\Omega_f, \Psi, \mathbf{X})/\partial\psi_h^{(m)} = 0$. This leads to

$$P_k(h\omega_f^{(m)}) = \mathbb{P}_k(h\omega_f^{(m)}) = \left(\sum_j^M \lambda_k^{(j)} \psi_{h_j}^{(j)^{1/2}}\right)^2 \tag{8}$$

where

$$\psi_{h_j}^{(j)} = \begin{cases} > 0 & \text{if } h_j = h\omega_f^{(m)}/\omega_f^{(j)} \in \mathbb{Z} \\ 0 & \text{otherwise.} \end{cases}$$

If the frequency $h\omega_f^{(m)}$ is *shared* by M' sources (or the

number of nonzero $\psi_{h_j}^{(j)}$ is M'), then Equation (8) becomes

$$P_k(h\omega_f^{(m)}) = \left(\sum_j^{M'} \lambda_k^{(j)} {\psi_{h_j}^{(j)}}^{1/2} \right)^2 \tag{9}$$

If we assume the energy contribution of all the sources to be same, i.e. $\lambda_k^{(j)} {\psi_{h_j}^{(j)}}^{1/2} = \bar{\psi}_h$, for $j = 1, \cdots, M'$, we have

$$P_k(h\omega_f^{(m)}) = \left(M'\bar{\psi}_h\right)^2 = {M'}^2 {\bar{\psi}_h}^2 = {M'}^2 {\lambda_k^{(m)}}^2 \psi_{h_m}^{(m)} \tag{10}$$

rearranging Equation (10), we have

$$\psi_h^{(m)^{ML}} = \frac{P_k(h\omega_f^{(m)})}{{M'}^2 {\lambda_k^{(m)}}^2} \tag{11}$$

Substituting the ML estimate for the energies (Equation (11)) in the negative log-likelihood (Equation (7)), we have a modified negative log-likelihood

$$\begin{aligned} \ell_k(\Omega_f, \hat{\Psi}^{ML}, \mathbf{X}) &= \ell'_k(\Omega_f, \mathbf{X}) \\ &= \sum_{\omega_j \notin \mathbb{H}} \parallel P(\omega_j) - \mathbb{P}(\omega_j) \parallel^2 \\ &\quad + \sum_{\omega_j \in \mathbb{H}} \parallel P(\omega_j) - \mathbb{P}(\omega_j) \parallel^2 \end{aligned}$$

where $\mathbb{H}$ is the harmonic set, which is the set of all harmonic frequencies for all sources, $\mathbb{H} = \bigcup_m \left[\omega_f^{(m)}, 2\omega_f^{(m)}, \cdots\right]^T$. The value of generative model $\mathbb{P}$ is zero at the frequencies *not in* the harmonic set, while it is exactly equal to the observed PSD at the frequencies *in* the harmonic set. Hence

$$\ell'_k(\Omega_f, \mathbf{X}) = \sum_{\omega_j \notin \mathbb{H}} (P_k(\omega_j))^2 \tag{12}$$

Equation (12) is the negative log-likelihood with the constraint of Equation (11) imposed. Equation (12) implies that the modified likelihood at the ML estimate of energies is independent of the source locations

$$\ell_k(\Omega_f, \Psi^{ML}, \mathbf{X}) = \ell'_k(\Omega_f, \mathbf{X}) = \ell'_k(\Omega_f)$$

□

Hence, according to proposition 2, source separation can be performed independent of source localization. The full negative log-likelihood for all sensors, $\ell'(\Omega_f)$ is defined as

$$\ell'(\Omega_f) = \frac{1}{K} \sum_{k=1}^{K} \ell'_k(\Omega_f)$$

Thus, the ML estimation of the fundamental frequencies can be obtained by minimizing $\ell'(\Omega_f)$

$$\hat{\Omega}_f^{ML} = \arg\min_{\Omega_f} \ell'(\Omega_f) \tag{13}$$

Since an exact ML estimation method for Equation (13) is not available we will use a Monte Carlo method described in Section 6 for estimation.

5 Source Localization

Source localization is performed by Bayesian estimation in the graphical model shown in Figure 2, and taking the *maximum a-posteriori* (MAP) estimate of the source positions. The posterior, $p(\mathbf{X}|\mathbf{B})$ of the source positions at the ML estimates for source fundamental frequencies and harmonic energies given the beamform data, is following

$$p(\mathbf{X}|\mathbf{B}) \propto \prod_k p(B_k|\mathbf{X}, \hat{\Omega}_f^{ML}, \hat{\Psi}^{ML}) p(\mathbf{X}) \tag{14}$$

where $p(B_k|\mathbf{X}, \hat{\Omega}_f^{ML}, \hat{\Psi}^{ML})$ is the likelihood function for beamform data. The likelihood function requires a generative model for the beamform data. In this section, we begin by presenting the generative model and three results pertaining to the generative model. Finally, we present the likelihood and MAP estimation.

Generative Model for Beamform. We start by developing a generative model for a beamform for a two-microphone array, single-source case. We will show that the beamform for an arbitrary microphone array and an arbitrary number of sources can be composed from the simple two-microphone array, single-source case.

Proposition 3. *Consider a microphone pair separated by distance d and the angle between the x-axis and the line joining the microphones is β. For an acoustic source at angle θ and range r with power spectral density $P(\omega)$, the beamform B at the microphone pair is given by*

$$B(\alpha) = 2\lambda^2(R_{ss}(0) + R_{ss}(\kappa_\alpha)) + 2R_\eta(0) \tag{15}$$

where $R_{ss}(\tau) = \mathbb{FFT}^{-1}(P(\omega))$ for $\tau \in [-\infty, +\infty]$ is the autocorrelation of the source signal, $R_{ss}(0)$ is the signal energy, $R_\eta(0)$ is the noise energy, λ is the attenuation factor, and $\kappa_\alpha = d(\cos(\alpha-\beta) - \cos(\theta-\beta))f_s/C$, where $\alpha \in [0, 2\pi]$ is the beam angle, f_s and C are sampling frequency and speed of sound, respectively.

The proof for the proposition is given in [14]. For an arbitrary microphone-array, the generative model can be extended using the model in Equation (15).

Proposition 4. *For an arbitrary microphone-array of N_{mic} microphones, the beamform is expressed in terms*

of pairwise beamforms as

$$B(\alpha) = \sum_{(i,j)\in \mathbf{pa}} B_{i,j}(\alpha) - N_{mic}(N_{mic}-2)(R_\eta(0)+\lambda^2 R_{ss}(0)) \tag{16}$$

where **pa** *is the set of all microphone pairs,* $R_{ss}(0)$ *is the signal energy,* $R_\eta(0)$ *is the noise energy,* λ *is the attenuation factor, and* $B_{i,j}$ *is beamform for the microphone pair* (i,j) *(Equation* (15)*).*

The proof for the proposition is given in [14]. For an arbitrary number of acoustic source, the generative model can be extended using the model in Equation (16).

Proposition 5. *For an arbitrary number of uncorrelated acoustic sources* M*, the beamform is expressed in terms of single source beamforms as*

$$B(\alpha) = \sum_{m=1}^{M} B_m(\alpha) - N_{mic}(M-1)R_\eta(0) \tag{17}$$

where $R_\eta(0)$ *is the noise energy and* B_m *is the beamform for* m^{th} *acoustic source (Equation* (16)*).*

The proof for the proposition is given in [14]. A general form of generative model for beamform for arbitrary microphone array and arbitrary number of sources can be obtained by substituting Equations (15) and (16) into Equation (17), which gives following

$$\begin{aligned}\mathbb{B}(\alpha) = 2\sum_{m=1}^{M} {\lambda^{(m)}}^2 \sum_{(i,j)\in\mathbb{P}} R_{ss}^{(m)}(\kappa_\alpha) \\ + N_{mic}\sum_{m=1}^{M} {\lambda^{(m)}}^2 R_{ss}^{(m)}(0) + N_{mic}R_\eta(0)\end{aligned} \tag{18}$$

Data Likelihood. Using Equation (18), the negative log-likelihood for beamform data is given as

$$-\ln p(B_k|\mathbf{X}) = \ell_k(\mathbf{X}) = \frac{1}{\sigma_B^2}\sum_{\alpha} \parallel B_k(\alpha) - \mathbb{B}_k(\alpha) \parallel^2$$

The MAP estimate of the source positions is given by

$$\hat{\mathbf{X}}^{MAP} = \arg\max_{\mathbf{X}} p(\mathbf{X}|\mathbf{B}) \tag{19}$$

Again, since an exact estimation method for Equation (19) is not available we will use the Monte Carlo method described in Section 6 for MAP estimation.

6 Bayesian Estimation

Due to the non-linearity of the observation model and non-Gaussianity of the probability densities, the use of exact methods for state estimation is not possible. We use Markov Chain Monte Carlo (MCMC) sampling algorithms, specifically Gibbs sampling and slice sampling [15] for approximate state estimation. The MCMC algorithms are more efficient in high-dimensions than Monte Carlo (MC) methods, also called particle filters, due to the fact that the samples in MC methods are drawn independently while in samples in MCMC are drawn from a Markov chain. The Gibbs sampler works on the idea that while the joint probability density is too complex to draw samples from directly, the *univariate* conditional densities – the density when all but one of the random variables are assigned fixed values – are easier to sample.

The choice of algorithm to sample from the univariate density determines the speed and convergence of the Gibbs sampler. We selected slice sampling [15] for its robustness in parameters such as step size and applicability toward non-log-concave densities, which is the case in our problem.

The likelihood in Equation (13) and the posterior density in Equation (19) are sampled using the Gibbs sampler to the estimate the ML estimate and MAP estimate, respectively.

Initialization Strategy. A good initialization of the state will ensure faster convergence of the Gibbs sampler. For source separation, the fundamental frequencies, Ω_f are initialized by doing a coarse resolution search to minimize the likelihood in Equation (13). During the source localization step, the source positions are initialized using one of the following methods, (1) the least-squares method for a single target, similar to one described in [6], or (2) the weighted-average of the sensor positions. Finally, the harmonic energies are initialized according to Equation (11).

7 Simulation Results

The scenarios considered here involve a wireless sensor network deployed in a grid topology. Typically, localization of an acoustic source is performed by the sensors that are close to the source because the signal-to-ratio (SNR) is lower for farther sensors. For this reason, we assume that even in a large sensor network, a source will be surrounded by a small number of sensors that will participate in the localization of that source.

Simulation Setup and Parameters. We consider a small sensor network of 4 acoustic sensors arranged in a grid of size $10m \times 5m$, wherein each sensor can detect all the sources. We simulate the sources according the acoustic source model (Section 2), simulate the data according to the observation process (Section 2), and finally check the output of source localization against the ground truth. The performance of the approach is measured in terms of localization error, which is defined as the root mean square (RMS) position error averaged over all the sources

$$E = \frac{1}{M}\sum_{m=1}^{M} ||\mathbf{x}^{(m)} - \tilde{\mathbf{x}}^{(m)}||$$

Table 1: Parameters used in simulations

Sampling frequency (f_s)	100kHz
Speed of sound (C)	350 m/sec
Downsampling factor	25
Audio data length (time)	1 sec
Maximum harmonic frequency (ω_{max})	1000Hz
SNR (dB)	25
Number of beams	36
Size of Fourier transform (N_{FFT})	4000
Number of Gibbs sample	40

where M is the number of source, and $\mathbf{x}^{(m)}$ and $\tilde{\mathbf{x}}^{(m)}$ are the estimated and ground truth positions for the m^{th} source, respectively. Table 1 shows the parameters used in the algorithm.

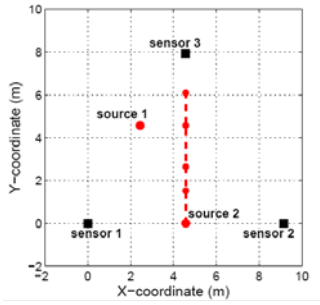
Simulation Scenarios. We study three simulation scenarios. In the first scenario, we increase the number of sources present in the sensing region gradually to see the effect on accuracy of detection. In the second scenario, we increase the average source SNR of two sources present in the sensing region. In the third scenario, we increase the separation between two sources present in the sensing region.

Figure 3(a) shows the localization error for the first scenario when the number of sources is increased from 1 to 4. The localization error increases approximately exponentially with the number of sources. Figure 3(b) shows the average localization error for the second scenario when source SNR for the two sources is increased from 7dB to 52dB. As expected, the localization error decreases with increasing SNR and remains approximately constant above 20dB. Figures 3(c) and 3(d) show the localization error for the third scenario when the source separation between two sources is increased from $0.1m$ to $8m$. For small source separations ($0.1m$ and $0.2m$), the localization error is of the same order as the separation. This indicates that the two sources cannot be disambiguated at such separations. For higher source separations (above $0.5m$), the localization error is a small fraction of the separation. This indicates that the two sources are successfully localized and disambiguated. In fact, for larger source separations (above $5m$), the average localization error for two sources is same as that for the single sources.

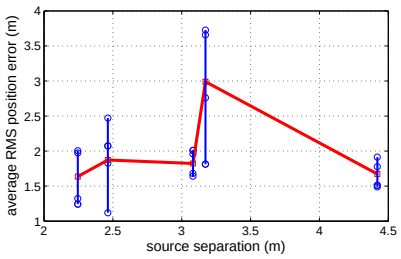
8 Outdoor Experiments

We implemented the beamforming and PSD estimation described in section 2 on an Xilinx XC3S1000 FPGA chip onboard the MICAz sensor motes. Both processes run at 4Hz. Beamforming utilizes 166 msec of audio data each cycle, while the PSD estimation module utilizes 1 sec of data with 75% overlap. The angular resolution of beamforming is 10 degrees while frequency resolution of PSD estimation is 1Hz. The PSD estimation module returns 30 PSD values.

We deployed a small sensor network of 3 MICAz-based acoustic sensor nodes in an equilateral triangle of side length 9.144m (15ft). Figure 4(a) shows the experimental setup and the location of the sources. We collected the sensor data and ran the algorithm offline. Figure 4(b) shows the localization error with source separation. The results follow the similar trend as that in Figure 3(c). For smaller source separations, the average error remains low but the algorithm is not able to disambiguate the two sources. For larger separations, the localization error decreases.



(a)



(b)

Figure 4: (a) Outdoor experimental setup. Source 1 is kept at the same location while source 2 is placed at different locations. (b) Localization error with source separation.

9 Conclusion

In this paper, we proposed a feature-based fusion method for localization and discrimination of multiple acoustic sources in WSNs. Our approach fused beamforms and PSD data from each sensor. The approach utilized a graphical model for estimating the source positions and the fundamental frequencies. We subdivided the problem into source separation and source localization. We showed in simulation and outdoor experiments that the approach can discriminate multiple sources using the simple features collected from the resource-constrained sensor nodes. As part of an ongoing work, we are working on target dynamics models to extend the approach for multiple source tracking. In the future, the use of graphical models will allow us to extend the approach to multimodal sensors.

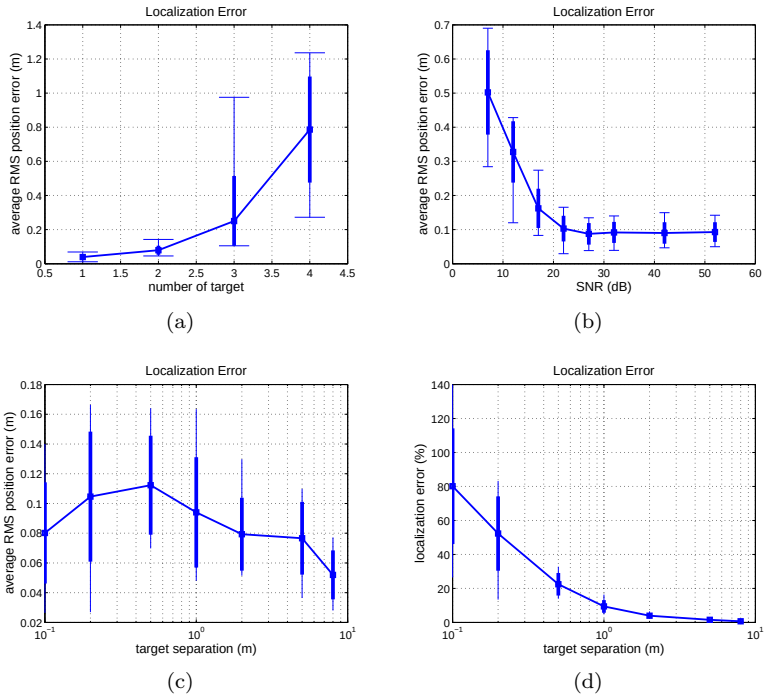


Figure 3: Localization error with (a) Source density, (b) Source SNR, and (c) Source separation. (d) Localization error as a percentage of source separation.

References

[1] M. Kushwaha, I. Amundson, P. Volgyesi, P. Ahammad, G. Simon, X. Koutsoukos, A. Ledeczi, and S. Sastry, "Multi-modal target tracking using heterogeneous sensor networks," in *ICCCN*, 2008.

[2] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications.* Springer, 2001.

[3] K. Yao, R. E. Hudson, C. W. Reed, D. Chen, and F. Lorenzelli, "Blind beamforming on a randomly distributed sensor array system," in *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, October 1998.

[4] J. C. Chen, K. Yao, and R. E. Hudson, "Acoustic source localization and beamforming: theory and practice," in *EURASIP J. Appl. Signal Process.*, April 2003.

[5] D. Li and Y. H. Hu, "Energy-based collaborative source localization using acoustic microsensor array," in *EURASIP J. Appl. Signal Process.*, 2003.

[6] C. Meesookho, U. Mitra, and S. Narayanan, "On energy-based acoustic source localization for sensor networks," in *IEEE Trans. Signal Process.*, vol. 56, 2008.

[7] X. Sheng and Y.-H. Hu, "Maximum likelihood multiple source localization using acoustic energy measurements with wireless sensor networks," in *IEEE Trans. Signal Process.*, vol. 53, 2005.

[8] M. J. Beal, N. Jojic, and H. Attias, "A graphical model for audiovisual object tracking," in *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, 2003.

[9] M. Orton and W. Fitzgerald, "A bayesian approach to tracking multiple targets using sensor arrays and particle filters," in *IEEE Trans. Signal Process.*, vol. 50, 2002.

[10] M. R. Morelande, C. M. Kreucher, and K. Kastella, "A bayesian approach to multiple target detection and tracking," in *IEEE Trans. Signal Process.*, vol. 55, 2007.

[11] R. Mahler, "Multi-target bayes filtering via first-order multi-target moments," in *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, 2003.

[12] F. V. Jensen, *Bayesian Networks and Decision Graphs.* Springer, 2001.

[13] C. Serviere and P. Fabry, "Blind source separation of noisy harmonic signals for rotating machine diagnosis," in *Journal of Sound and Vibration*, vol. 272, 2004.

[14] M. Kushwaha and X. Koutsoukos, "A graphical model approach to source localization in wireless sensor networks," ISIS, Vanderbilt University, Tech. Rep. ISIS-09-101, 2009.

[15] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, 2002.