A Graphical Model Approach to Source Localization in Wireless Sensor Networks

Manish Kushwaha and Xenofon Koutsoukos

Institute for Software Integrated Systems (ISIS)
Department of Electrical Engineering and Computer Science
Vanderbilt University
Nashville, TN 37235, USA
Email: manish.kushwaha@vanderbilt.edu

Abstract. Collaborative localization and discrimination of acoustic sources is an important problem in wireless sensor networks. The two broad approaches for source localization are *signal-based* and *feature-based* methods. The signal-based methods are not suitable for collaborative localization because they require transmission of raw acoustic data, which is costly due to limited bandwidth and power available on wireless sensors. In feature-based methods, signal features are extracted at each sensor and estimation is done by multisensor fusion of the extracted features collected from all the sensors. The feature-based methods are suitable to sensor networks due to its lower bandwidth requirements as compared to the signal-based methods.

In this paper, we present a feature-based localization and discrimination approach for multiple acoustic sources in wireless sensor networks. The approach uses beamform and power spectral density (PSD) from each sensor as the features for multisensor fusion, localization and discrimination. 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 PSD.

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, environmental and wildlife habitat monitoring [1]. Recently more innovative applications such as smart video-conferencing [2], multimodal sensor fusion and target tracking [3] 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. There are two broad classes of methods for collaborative source localization. The first class of approaches where the estimation is done using the sampled signals are called *signal-based methods*, while the second class of approaches where signal features are extracted at each sensor and estimation is done using the extracted signal features collected from all the sensors are called *feature-based methods*.

Several signal-based methods using a microphone array for source localization have been proposed [4]. These approaches use time delay of arrival (TDOA) or direction of arrival (DOA) estimation, beamforming [4, 5] and maximum likelihood estimation [6]. The signal-based methods are not suited for WSNs because they require transmission of the raw signal, which is costly due to limited bandwidth and power. On the other hand, the feature-based methods are appropriate for WSNs due to its lower bandwidth and power requirements. An example of feature-based method is energy-based localization (EBL), where signal energy is taken as the feature. EBL has been solved using various least squares [7, 8] and maximum likelihood [9] 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.

In this paper, we present a feature-based approach to collaborative source localization for multiple acoustic sources in a wireless sensor network. 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 as the feature is that the beamform captures the angular variation of 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. Target tracking application in [10] demostrated that the communication bandwidth available in sensor networks is sufficient to support wireless transmissions of such features.

Localization algorithms based on least squares work on strict Gaussian noise assumption and are generally not extensible to multiple sources; those based on beamforming have poor performance for multiple source; while those based on maximum likelihood are not extensible to tracking applications where data association across time become an issue. A Bayesian approach for source localization can handle non-Gaussianity and multiple sources, both stationary and moving.

We solve the localization problem using graphical models that are generalization of Bayesian estimation. Graphical models are graphs in which nodes represent random variables, and the (lack of) arcs captures conditional independence of random variables. They provide a compact representation of joint probability density and facilitate the factorization of joint density into conditional densities [11]. 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 acoustic PSD. We present a localization algorithm based on Gibbs sampling for approximate state estimation. Finally, 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.

The main assumption made in this paper for acoustic sources are that they are (1) stationary point sources, (2) emitting a stationary signal, (3) with a harmonic power spectral density, and (4) the cross-correlation between two sources is negligible compared to the autocorrelation. The assumptions made on sensors are that they are (1) coplanar with the sources, and (2) have 2D array of microphones. In addition, we also assume that the acoustic signal attenuates at a rate that is inversely proportional to the distance from the source.

The rest of the paper is organized as follows. In Section 3, we present the acoustic source model and the acoustic sensor model. Section 4 describes the graphical model. Sections 5 and 6 describe the source separation and source localization, respectively. In Section 7, we present the Gibbs sampler and the initialization strategy. We present results for various simulation scenarios in Section 8, and the outdoor experiment setup and results in Section 9. We conclude in Section 10.

2 Related Work

Signal-based methods for acoustic source localization methods typically make use of time delay of arrival (TDOA) and direction of arrival (DOA). An overview of theoretical aspects of TDOA acoustic source localization and beamforming is presented in [6], along with a localization algorithm based on maximum likelihood (ML) estimation. For multiple acoustic source localization, an approximate maximum likelihood (AML) algorithm based on alternative projection method is presented. An empirical study of collaborative acoustic source localization based on an implementation of the AML algorithm is shown in [1]. Among feature-based methods, energy-based acoustic source localization methods are presented in [7–9]. A least-squares approach is taken to solve the localization problem in [7]. Several other least-squares formulations are proposed that are efficient and achieve better accuracy in [8]. A maximum likelihood formulation with capability for multiple source localization is presented in [9]. A multiresolution search algorithm and an expectation-maximization (EM) like iterative algorithm are presented for ML estimation.

Several approaches based on graphical models [3, 12] and Bayesian estimation [13–15] have been proposed for multiple target localization and tracking. A graphical model based approach for audiovisual object tracking is presented that combines the audio and video data in [3]. The paper presents a graphical model for the audio-video data, and a Bayesian inference algorithm for tracking. A graphical model formulation for self-localization of sensor networks is presented in [12]. A technique called nonparametric belief propagation (NBP) that is a generalization of particle filtering, is proposed for Bayesian inference. The NBP approach has the advantage of being amenable to distributed implementation, can include a wide variety of statistical models, and can represent multimodal uncertainty. A Bayesian approach for tracking the direction-of-arrival (DOA) of multiple targets using a passive sensor array is presented in [13]. The paper includes a constant velocity target dynamics model, a data model for uniform linear sensor array, and a tracking algorithm based on particle filtering. The work in [13] is extended in [14] by considering maneuvering wideband acoustic targets. A Bayesian approach for multiple target detection and tracking, and particle filter-based algorithms are proposed in [15]. The paper constructs and computes the joint multitarget probability density (JMPD) for multitarget state estimation. For unknown number of target, the multitarget state is extended to include a random variable corresponding to the number of target.

3 Multiple Acoustic Source Localization

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 wave front incident on sensor is assumed to be planar for far-field sources. Each sensor receives the signal and runs simple signal processing algorithms to compute 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.

3.1 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 [16]. In general, any acoustic signal originating due to the vibrations from rotating machinery will have an harmonic structure.

In practice, some of the assumptions made in this model may not be always true and, hence, may affect its accuracy. For example, the engine sound of a vehicle may not be omni-directional and will be biased toward the side closer to the engine. The physical size of the acoustic source may be too large to be adequately modeled as a point source for sensors very close to the source. In an outdoor environment, strong background noise, including wind gusts, may be encountered during operation. In addition, the gain of individual microphones will need to be calibrated to yield consistent acoustic energy readings. Perhaps the most restrictive assumption is that the source signals are harmonic. In addition to the harmonic components, the engine sound signal may contain other frequency components, which when not accounted for, may cause localization to deteriorate.

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.

3.2 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 [9]. The discrete signal received at the p^{th} microphone is given by

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

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

$$\lambda_p^{(m)} = \frac{d_0}{\parallel \mathbf{x}_p - \mathbf{x}^{(m)} \parallel} \tag{2}$$

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. For each beam, assuming the source is located in that direction, the microphone signals are delayed according to the relative time delay and summed together into a composite signal as

$$r[n] = \sum_{p=1}^{N_c} r_p[n + t_{pq}(\alpha)] \tag{3}$$

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

$$t_{pq}(\alpha) = d_{pq} cos(\alpha - \beta_{pq}) f_s / C \tag{4}$$

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. The L_2 norm of the composite signal is the beam energy

$$B(\alpha) = \left(\sum_{n=1}^{L} r[n]^2 \right)^{1/2} \tag{5}$$

Beam energies are computed for each of the beams, and are collectively called the *beamform*, B. 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 [10].

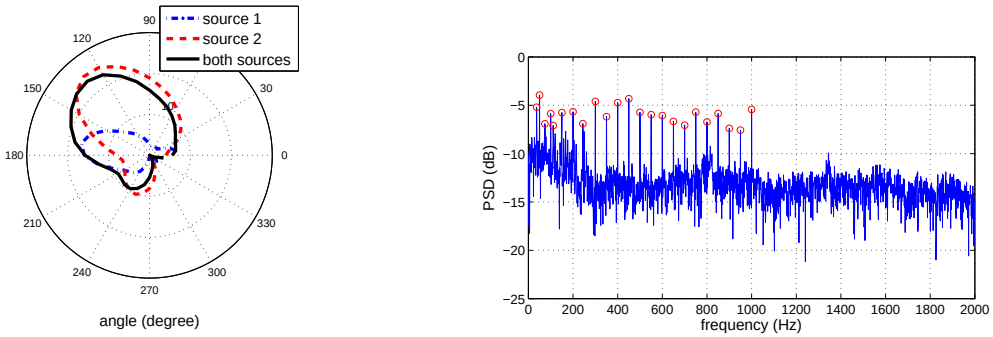


Fig. 1. (a) Acoustic Beamforms; The beamforms for single sources clearly show peaks at the source location but the beamform when both sources are present does not show two peaks. (b) Power spectral density (PSD); The highest PSD values are shown as empty circles. The PSD is compactly represented as pairs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

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

$$P(\omega) = Y(\omega) \cdot \overline{Y(\omega)} \tag{6}$$

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.

4 Overview

Source separation and localization of multiple sources is performed using the graphical model[1] shown in Figure 2. The nodes with clear backgrounds denote hidden state variables; $\mathbf{x}^{(m)}, \omega_f^{(m)}, \psi^{(m)}$ denote

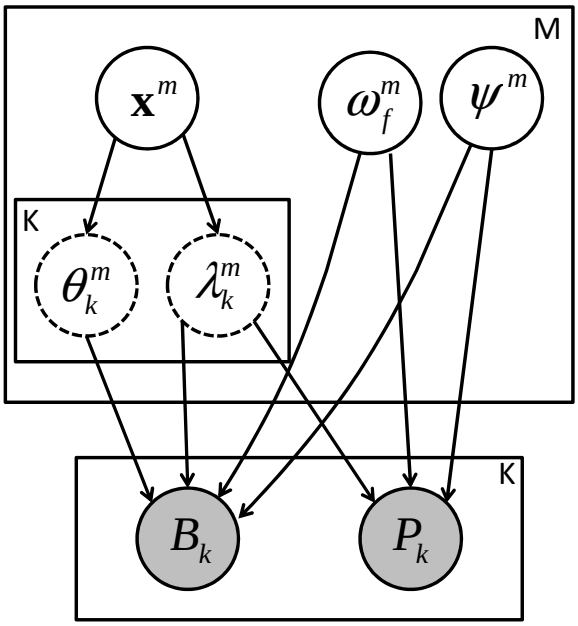


Fig. 2. Graphical model

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 $\lambda_k^{(m)}$.

Source separation, in the present context, means separating the PSDs of the sources. Source separation is done using fundamental frequency estimation, also called $F0$ estimation. 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. Source localization, in this context, means source position estimation. We use the *separated source PSDs* in Bayesian estimation for source localization.

[1] We use the plate notation to represent the repetition of random variables.

5 Source Separation

Source separation problems in signal processing are those in which several signals have been mixed together and the objective is to find the original signals. Several approaches based on principal component analysis (PCA) and independent component analysis (ICA) have been proposed for source separation in simplified cases. In the present context, source separation means separating the source PSDs, which is performed by estimating the fundamental frequencies. In this section, we introduce the generative model for PSD and the ML estimation method for fundamental frequencies.

5.1 PSD generative model

The PSD for the m^{th} source is given by

$$P_s^{(m)}(\omega) = \sum_{h=1}^{H} \psi_h^{(m)} \delta(\omega - h\omega_f^{(m)}) \tag{7}$$

where 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. We have the following proposition for the PSD estimate at a sensor.

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

$$\begin{aligned}\mathbb{P}(\omega) = \sum_{m=1}^{M}\sum_{n=1}^{M} \lambda^{(m)}\lambda^{(n)} \left(P_s^{(m)}(\omega)P_s^{(n)}(\omega)\right)^{\frac{1}{2}} \\ \cos(\Phi^{(m)}(\omega) - \Phi^{(n)}(\omega))\end{aligned} \tag{8}$$

where M is the number of sources, $\lambda^{(m)}$ is the attenuation factor, and $\Phi^{(m)}(\omega)$ is the phase spectral density, which is 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 appendix.

Since we do not maintain the phase of the signal in the source model (see Section 3), we assume all the phases to be uniformly distributed. The expected value of the cosine of the difference of two uniformly distributed angles is zero, i.e. $E[cos(\Phi_i - \Phi_j)] = 1$. Using this, Equation (8) can be approximated as

$$\mathbb{P}(\omega) \approx \left[\sum_{m=1}^{M} \lambda^{(m)} \left(P^{(m)}(\omega)\right)^{1/2}\right]^2 \tag{9}$$

5.2 Data likelihood

The negative log-likelihood at the k^{th} sensor is defined as

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

where the vectors of the unknown parameters are given by

$$\begin{aligned}
\Omega_f &= \left[\omega_f^{(1)}, \cdots, \omega_f^{(M)}\right]^T \\
\Psi &= \left[\psi^{(1)}, \cdots, \psi^{(M)}\right]^T \\
\psi^{(m)} &= \left[\psi_1^{(m)}, \cdots, \psi_H^{(m)}\right]^T \\
X &= \left[\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(M)}\right]^T
\end{aligned}$$

Thus, the maximum likelihood parameter estimation of $[\Omega_f, \Psi, X]^T$ can be obtained by minimizing $\ell_k(\Omega_f, \Psi, X)$. To minimize $\ell_k(\Omega_f, \Psi, X)$, the solution must lie where the partial derivatives of the likelihood w.r.t. to the parameters are zero. The partial derivative w.r.t. the energies is

$$\frac{\partial}{\partial \psi_h^{(m)}} \ell_k(\Omega_f, \Psi, X) = 0$$

which leads to the following

$$P_k(h\omega_f^{(m)}) = \mathbb{P}_k(h\omega_f^{(m)}) = \left(\sum_{m'}^{M} \lambda_k^{(m')} {\psi_{h_{m'}}^{(m')}}^{1/2}\right)^2$$

where

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

If the frequency $h\omega_f^{(m)}$ is *shared* by M' sources (or the number of nonzero $\psi_{h_{m'}}^{(m')}$ is M'), then we have

$$P_k(h\omega_f^{(m)}) = \left(\sum_{m'}^{M'} \lambda_k^{(m')} {\psi_{h_{m'}}^{(m')}}^{1/2}\right)^2$$

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

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

or

$$\psi_h^{(m)^{ML}} = \frac{\bar{\psi}_h^2}{{\lambda_k^{(m)}}^2} = \frac{P_k(h\omega_f^{(m)})}{{M'}^2 {\lambda_k^{(m)}}^2} \tag{10}$$

Substituting the solution for the energies in the negative log-likelihood equation, we have a modified negative log-likelihood

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

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

$$\ell'_k(\Omega_f, X) = \sum_{\omega_j \notin \mathbb{H}} \left(P_k(\omega_j)\right)^2 \tag{11}$$

Equation (11) is the log-likelihood with the constraint of Equation (10) imposed. Equation (11) implies that the modified likelihood at the ML estimate of energies is independent of the source locations.

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

Hence, source separation (f0 estimation) 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 parameter estimation of the fundamental frequencies can be obtained by minimizing $\ell'(\Omega_f)$

$$\hat{\Omega}_f^{ML} = \arg\min_{\Omega_f} \ell'(\Omega_f) \tag{12}$$

Since an exact ML estimation method for Equation (12) is not available we will use an approximate method described in Section 7 for ML parameter estimation.

6 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, $p(X|B)$ of the source positions conditioned on the observed beamforms and ML estimates for fundamental frequencies and energies, is given by

$$p(X|B) \propto \prod_k p(B_k|X, \hat{\Omega}_f^{ML}, \hat{\Psi}^{ML}) p(X) \tag{13}$$

6.1 Beamform generative model

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 2. *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*

$$B^2(\alpha) = 2\lambda^2(R_{ss}(0) + R_{ss}(\kappa_\alpha)) + 2R_\eta(0) \tag{14}$$

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 appendix.

The beamform expression in equation (14) is composable for arbitrary microphone-array and arbitrary number of acoustic sources.

Proposition 3. *For an arbitrary microphone-array of N_{mic} microphones, the beamform is expressed in terms of pairwise beamforms as*

$$B^2(\alpha) = \sum_{(i,j)\in\mathbf{pa}} B^2_{i,j}(\alpha) - N_{mic}(N_{mic} - 2)(R_\eta(0) + \lambda^2 R_{ss}(0)) \tag{15}$$

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 (14)).*

The proof for the proposition is given in appendix.

Proposition 4. *For an arbitrary number of uncorrelated acoustic sources M, the beamform is expressed in terms of single source beamforms as*

$$B^2(\alpha) = \sum_{m=1}^{M} B^2_m(\alpha) - N_{mic}(M - 1)R_\eta(0) \tag{16}$$

where $R_\eta(0)$ is the noise energy and B_m is the beamform for m^{th} acoustic source (Equation (15)).

The proof for the proposition is given in appendix. Substituting Equations (14) and (15) into Equation (16), and rearranging and simplifying gives the generative model for beamform as following

$$\begin{aligned}\mathbb{B}^2(\alpha) = 2\sum_{m=1}^{M} {\lambda^{(m)}}^2 \sum_{(i,j)\in\mathbb{P}} R^{(m)}_{ss}(\kappa_\alpha) \\ + N_{mic}\sum_{m=1}^{M} {\lambda^{(m)}}^2 R^{(m)}_{ss}(0) + N_{mic}R_\eta(0)\end{aligned} \tag{17}$$

6.2 Data likelihood

The negative log-likelihood is given as

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

The MAP estimate of the source positions is given by

$$\begin{aligned}\hat{X}^{MAP} &= \arg\max_X p(X|B) \\ &= \arg\max_X \prod_k p(B_k|X, \hat{\Omega}_f^{ML}, \hat{\Psi}^{ML})p(X) \end{aligned} \tag{18}$$

Again, since an exact estimation method for Equation (18) is not available we will use the approximate method described in Section 7 for MAP estimation.

7 Bayesian Estimation using Gibbs Sampling

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 [17] 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.

7.1 Univariate conditional density

Let's rewrite the state vector such that $X = \left[x^{(1)}, x^{(2)}, \cdots, x^{(D)}\right]^{\mathrm{T}}$, where D is the number of state variables. The joint density $p(X_t|X_{t-1}, Y_t)$ is sampled using Gibbs sampler by sequentially sampling univariate conditional densities given by

$$x_t^{(k,j)} \sim p(x^{(j)}|X_t^{(k,-j)}, X_{t-1}, Y_t) \tag{19}$$

where k is the index of the sample, $j = 1, \cdots, n$ is the index of state variable currently being sampled, and $X_t^{(k,-j)} = \left[x_t^{(k,1)}, \cdots, x_t^{(k,j-1)}, x_t^{(k-1,j+1)}, \cdots, x_t^{(k-1,D)}\right]^{\mathrm{T}}$ is the set of all state variables except $x^{(j)}$. The choice of algorithm to sample from univariate density in Equation (19) determines the speed and convergence of the Gibbs sampler. We selected slice sampling [17] for its robustness in parameters such as step size and applicability toward non-log-concave densities, which is the case in our problem due to non-Gaussian and multimodal probability densities.

The likelihood in Equation (12) and the posterior density in Equation (18) are sampled using the Gibbs sampler to the estimate the ML estimate and MAP estimate, respectively.

7.2 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 (12). 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 [8], or (2) the weighted-average of the sensor positions. Finally, the harmonic energies are initialized according to Equation (10).

8 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.

8.1 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 3), simulate the data according to the observation process (Section 3), 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

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

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.

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

8.2 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

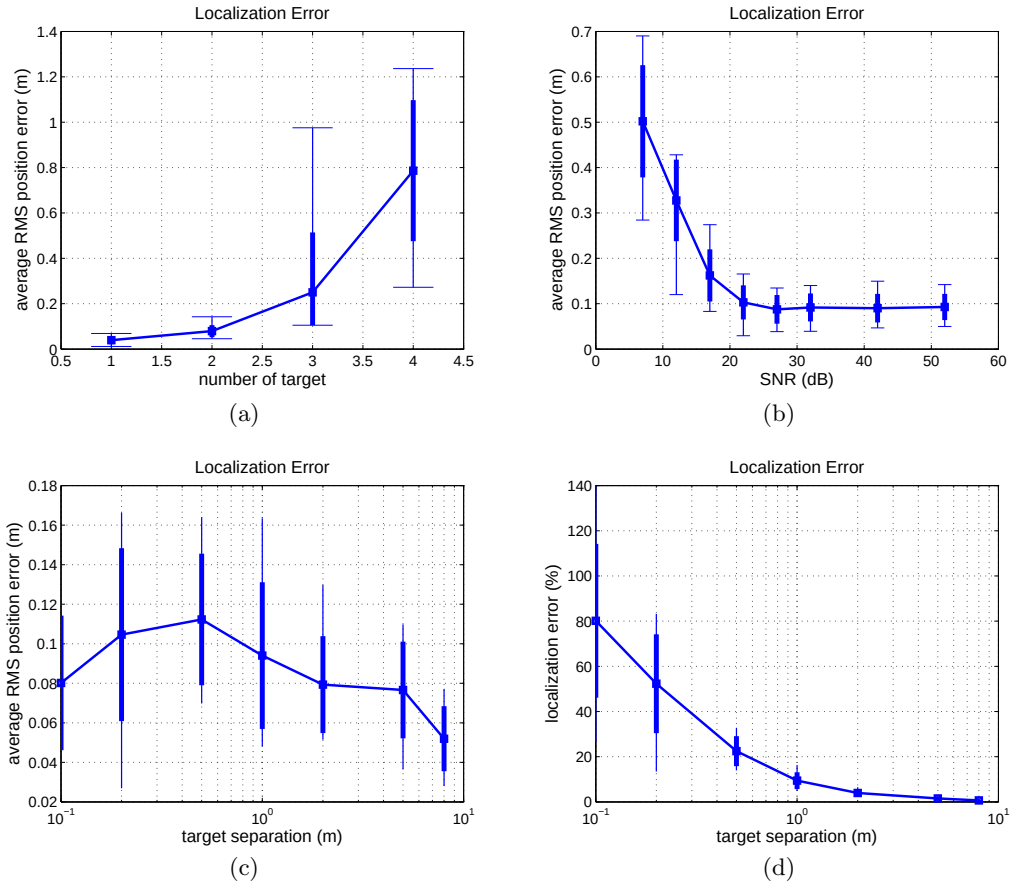


Fig. 3. Localization error with (a) Source density, (b) Source SNR, and (c) Source separation. (d) Localization error as a percentage of source separation.

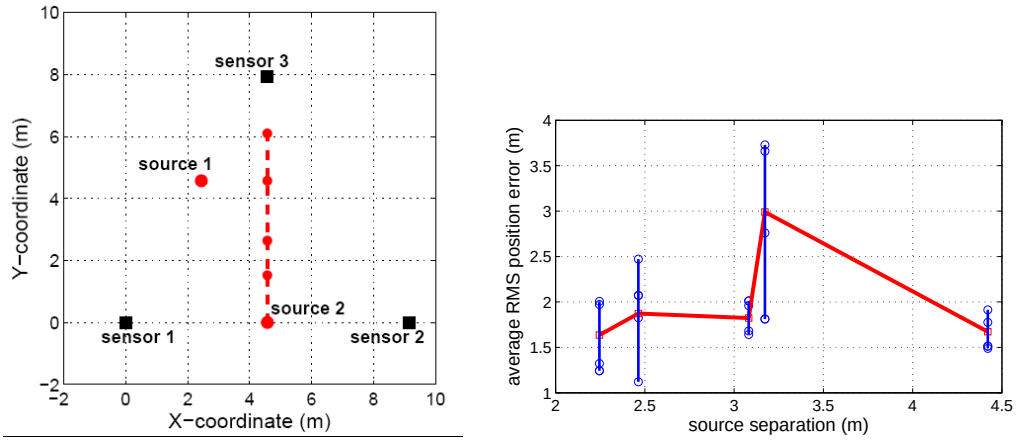
for the two sources is increased from 7dB to 52dB. As expected, the localization error decreases with increasing SNR and remains approximately constant above 20dB. Figures 3(c) and 3(d) show the localization error for the third scenario when the source separation between two sources is increased from $0.1m$ 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.

9 Outdoor Experiments

We implemented the beamforming and PSD estimation described in section 3 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)

Fig. 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.

10 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.

References

1. Ali, A.M., Yao, K., Collier, T.C., Taylor, C.E., Blumstein, D.T., Girod, L.: An empirical study of collaborative acoustic source localization. In: IPSN '07: Proceedings of the 6th international conference on Information processing in sensor networks. (2007) 41–50
2. Yoshimi, B.H., Pingali, G.S.: A multimodal speaker detection and tracking system for teleconferencing. In: ACM Multimedia '02. (2002)
3. Beal, M.J., Jojic, N., Attias, H.: A graphical model for audiovisual object tracking. Volume 25. (2003) 828–836
4. Brandstein, M., Ward, D.: Microphone Arrays: Signal Processing Techniques and Applications. Springer (2001)
5. Yao, K., Hudson, R.E., Reed, C.W., Chen, D., Lorenzelli, F.: Blind beamforming on a randomly distributed sensor array system. In: IEEE Journal on Selected Areas in Communications. Volume 16. (October 1998) 1555–1567
6. Chen, J.C., Yao, K., Hudson, R.E.: Acoustic source localization and beamforming: theory and practice. In: EURASIP Journal on Applied Signal Processing. (April 2003) 359–370
7. Li, D., Hu, Y.H.: Energy-based collaborative source localization using acoustic microsensor array. In: EURASIP Journal on Applied Signal Processing. Volume 2003. (2003) 321–337
8. Meesookho, C., Mitra, U., Narayanan, S.: On energy-based acoustic source localization for sensor networks. In: IEEE Transactions On Signal Processing. Volume 56. (2008) 365–377
9. Sheng, X., Hu, Y.H.: Maximum likelihood multiple source localization using acoustic energy measurements with wireless sensor networks. In: IEEE Transactions On Signal Processing. Volume 53. (2005) 44–53
10. Kushwaha, M., Amundson, I., Volgyesi, P., Ahammad, P., Simon, G., Koutsoukos, X., Ledeczi, A., Sastry, S.: Multi-modal target tracking using heterogeneous sensor networks. In: In International Conference on Computer Communications and Networks (ICCCN 2008). (2008)
11. Jensen, F.V.: Bayesian Networks and Decision Graphs. Springer (2001)
12. Ihler, A.T., Fisher, J.W., Moses, R.L., Willsky, A.S.: Nonparametric belief propagation for self-localization of sensor networks. In: IEEE Journal on Selected Areas in Communications. Volume 23. (2005) 809–819
13. Orton, M., Fitzgerald, W.: A bayesian approach to tracking multiple targets using sensor arrays and particle filters. In: IEEE Transactions on Signal Processing. Volume 50. (2002) 216–223
14. Cevher, V., McClellan, J.H.: Tracking of multiple wideband targets using passive sensor arrays and particle filters. In: Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop. (2002) 72–77
15. Morelande, M.R., Kreucher, C.M., Kastella, K.: A bayesian approach to multiple target detection and tracking. In: IEEE Transactions on Signal Processing. Volume 55. (2007) 1589–1604
16. Serviere, C., Fabry, P.: Blind source separation of noisy harmonic signals for rotating machine diagnosis. Volume 272. (2004) 317–339
17. MacKay, D.J.C.: Information Theory, Inference and Learning Algorithms. Cambridge University Press (2002)

A Appendix

A.1 Proof of Proposition 1

Proof. We prove Proposition 1 by derivation. Let there be M sources emitting source signals $s_m[n]$, for $m = 1, 2, \cdots, M$. Using Equation (1), the received signals at a microphone is given by

$$y[n] = \sum_{m=1}^{M} \lambda_m s_m[n - \tau_m] + w[n]$$

where τ_m is the propagation delay, and λ_m is the attenuation factor. Taking FFT of the received signal, we have

$$\begin{aligned} Y(\omega) &= \mathbb{FFT}\big(y[n]\big) \\ &= \mathbb{FFT}\left(\sum_{m=1}^{M} \lambda_m s_m[n - \tau_m] + w[n]\right) \\ &= \sum_{m=1}^{M} \lambda_m \mathbb{FFT}\big(s_m[n - \tau_m]\big) + \mathbb{FFT}\big(w[n]\big) \\ &= \sum_{m=1}^{M} \lambda_m S_m(\omega) + W(\omega) \end{aligned} \tag{20}$$

where $S_m(\omega)$ is the Fourier transform of m^{th} source signal, and $W(\omega)$ is the Fourier transform of noise. The power spectral density (PSD) of a signal is given by

$$\begin{aligned} P(\omega) &= Y(\omega) \cdot \overline{Y(\omega)} \\ &= \left(\sum_m \lambda_m S_m(\omega) + W(\omega)\right) \cdot \overline{\left(\sum_m \lambda_m S_m(\omega) + W(\omega)\right)} \\ &= \left(\sum_m \lambda_m S_m(\omega) + W(\omega)\right) \cdot \left(\sum_m \lambda_m \overline{S_m(\omega)} + \overline{W(\omega)}\right) \\ &= \sum_m \sum_n \lambda_m \lambda_m S_m(\omega) \cdot \overline{S_n(\omega)} + \sum_m \lambda_m S_m(\omega) \cdot \overline{W(\omega)} \\ &\quad + \sum_m \lambda_m \overline{S_m(\omega)} \cdot W(\omega) + W(\omega) \cdot \overline{W(\omega)} \end{aligned} \tag{21}$$

The Fourier transform $S(\omega)$ can also be written in terms of the PSD ($P(\omega)$) and phase spectral density ($\Phi(\omega)$), as

$$S(\omega) = P(\omega)^{1/2} e^{+i\Phi(\omega)} \tag{22}$$

which gives

$$\begin{aligned} S_m(\omega) \cdot \overline{S_n(\omega)} &= \big(P_m(\omega) P_n(\omega)\big)^{1/2} e^{+i(\Phi_m(\omega) - \Phi_n(\omega))} \\ S_m(\omega) \cdot \overline{W(\omega)} &= \big(P_m(\omega) P_\eta(\omega)\big)^{1/2} e^{+i(\Phi_m(\omega) - \Phi_\eta(\omega))} \\ \overline{S_m(\omega)} \cdot W(\omega) &= \big(P_m(\omega) P_\eta(\omega)\big)^{1/2} e^{-i(\Phi_m(\omega) - \Phi_\eta(\omega))} \\ W(\omega) \cdot \overline{W(\omega)} &= \big(P_\eta(\omega) P_\eta(\omega)\big)^{1/2} = P_\eta(\omega) \end{aligned}$$

where $P_\eta(\omega)$ and $\Phi_\eta(\omega)$ are PSD and phase spectral density of the noise signal. Rewriting Equation (21), we have

$$P(\omega) = \sum_m \sum_n \lambda_m \lambda_m \big(P_m(\omega) P_n(\omega)\big)^{1/2} e^{+i(\Phi_m(\omega) - \Phi_n(\omega))} + \sum_m \lambda_m \big(P_m(\omega) P_\eta(\omega)\big)^{1/2} e^{+i(\Phi_m(\omega) - \Phi_\eta(\omega))} + \sum_m \lambda_m \big(P_m(\omega) P_\eta(\omega)\big)^{1/2} e^{-i(\Phi_m(\omega) - \Phi_\eta(\omega))} + P_\eta(\omega) \tag{23}$$

Assuming that PSD for noise is negligible compared to actual source signals, we have

$$P(\omega) = \sum_m \sum_n \lambda_m \lambda_m \big(P_m(\omega) P_n(\omega)\big)^{1/2} e^{+i(\Phi_m(\omega) - \Phi_n(\omega))} \tag{24}$$

We know that PSD of real-valued signals is real-symmetric, hence the imaginary component in Equation (24) is zero. Hence, we have

$$P(\omega) = \sum_m \sum_n \lambda_m \lambda_m \big(P_m(\omega) P_n(\omega)\big)^{1/2} \cos\left(\Phi_m(\omega) - \Phi_n(\omega)\right) \tag{25}$$

□

A.2 Proof of Proposition 2

Proof. We prove Proposition 2 by derivation. Let the source be present at an angle θ emitting a source signal $s[n]$. Using Equation (1), the received signals at the microphones are given by

$$r_p[n] = \lambda_p s[n - \tau_p] + w_p[n]$$

for $p = 1, 2$, where τ_p is the propagation delay, and λ_p is the attenuation factor. For far-field case, the distances between the source and the closely-spaced microphones will be approximately same for all microphones, hence $\lambda_1 \approx \lambda_2 = \lambda$.

Using Equation (3), the composite microphone signal for the beam angle α is given by

$$\begin{aligned} r[n] &= r_1[n] + r_2[n + t_{12}(\alpha)] \\ &= \lambda s[n - \tau_1] + \lambda s[n + t_{12}(\alpha) - \tau_2] + w_1[n] + w_2[n + t_{12}(\alpha)] \end{aligned}$$

where $t_{12}(\alpha) = t_2(\alpha) - t_1(\alpha) = d\cos(\alpha - \beta) f_s / C$ is relative sample delay. The beam energy is given by

$$\begin{aligned} B^2(\alpha) &= \sum_n r[n]^2 \\ &= \sum_n \left(\lambda s[n - \tau_1] + \lambda s[n + t_{12} - \tau_2] + w_1[n] + w_2[n + t_{12}]\right)^2 \\ &= \lambda^2 \sum_n s[n - \tau_1]^2 + \lambda^2 \sum_n s[n + t_{12} - \tau_2]^2 + \sum_n w_1[n]^2 + \sum_n w_2[n + t_{12}]^2 \\ &\quad + 2\lambda^2 \sum_n s[n - \tau_1] s[n + t_{12} - \tau_2] + 2 \sum_n w_1[n] w_2[n + t_{12}] \\ &\quad + 2\lambda \sum_n (w_1[n] + w_2[n + t_{12}])(s[n - \tau_1] + s[n + t_{12} - \tau_2]) \end{aligned}$$

Rewriting the above expression in terms of signal and noise autocorrelation and cross-correlation, we have

$$\begin{aligned}B^2(\alpha) &= \lambda^2 R_{ss}(0) + \lambda^2 R_{ss}(0) + R_{w_1w_1}(0) + R_{w_2w_2}(0) \\ &+ 2\lambda^2 R_{ss}(t_{12} - \tau_2 + \tau_1) + 2R_{w_1w_2}(t_{12}) + 2\lambda R_{w_1s}(-\tau_1) + 2\lambda R_{w_2s}(t_{12} - \tau_1) \\ &+ 2\lambda R_{w_1s}(t_{12} - \tau_2) + 2\lambda R_{w_2s}(\tau_2)\end{aligned}$$

Now, assuming that the noises at the microphones are statistically same (i.e. $R_{w_1w_1}(0) = R_{w_2w_2}(0) = R_\eta(0)$) and the noises are uncorrelated (i.e. $R_{w_1w_2}[m] = 0$), and the noise and signal are also uncorrelated (i.e. $R_{w_ks}[m] = 0$), we have

$$B^2(\alpha) = 2\lambda^2 R_{ss}(0) + 2R_\eta(0) + 2\lambda^2 R_{ss}(t_{12} - \tau_{12})$$

Denoting $\kappa_\alpha = t_{12} - \tau_{12} = d(\cos(\alpha - \beta) - \cos(\theta - \beta))f_s/C$ and rearranging, we have

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

□

A.3 Proof of Proposition 3

Proof. We prove Proposition 3 by derivation. Let the source be present at an angle θ emitting a source signal $s[n]$. Using Equation (1), the received signals at the microphones are given by

$$r_p[n] = \lambda_p s[n - \tau_p] + w_p[n]$$

for $p = 1, 2, \cdots, N_{mic}$, where τ_p is the propagation delay, and λ_p is the attenuation factor. For far-field case, the distances between the source and the closely-spaced microphones will be approximately same for all microphones, hence $\lambda_1 \approx \lambda_2 \approx \cdots = \lambda$.

Using Equation (3), the composite microphone signal for the beam angle α is given by

$$\begin{aligned}r[n] &= \sum_{p=1}^{N_{mic}} r_p[n + t_{1p}(\alpha)] \\ &= \sum_{p=1}^{N_{mic}} \lambda s[n + t_{1p}(\alpha) - \tau_p] + w_p[n + t_{1p}(\alpha)]\end{aligned}$$

where $t_{1p}(\alpha) = t_p(\alpha) - t_1(\alpha) = d_{1p}\cos(\alpha - \beta_{1p})f_s/C$ is relative sample delay between the p^{th} and 1^{st} microphone. Let's denote $\phi_p = n + t_{1p}(\alpha) - \tau_p$ and $\psi_p = n + t_{1p}(\alpha)$ for clarity and brevity. The

beam energy is given by

$$
\begin{aligned}
B^2(\alpha) &= \sum_n r[n]^2 \\
&= \sum_n \left[\sum_p \lambda s[\phi_p] + w_p[\psi_p] \right]^2 \\
&= \sum_n \left[\left(\sum_p \lambda s[\phi_p] \right)^2 + \left(\sum_p w_p[\psi_p] \right)^2 + 2 \sum_p \sum_q \lambda s[\phi_p] w_q[\psi_q] \right] \\
&= \sum_n \left(\sum_p \lambda s[\phi_p] \right)^2 + \sum_n \left(\sum_p w_p[\psi_p] \right)^2 + 2 \sum_p \sum_{q, q \neq p} \underbrace{\sum_n \lambda s[\phi_p] w_q[\psi_q]}_{R_{w_q s}(\tau) = 0}
\end{aligned} \tag{26}
$$

The last term is signal-noise cross-correlation which is zero for uncorrelated signal and noise. The first two term in Equation (26) are expanded to

$$
\begin{aligned}
\sum_n \left(\sum_p \lambda s[\phi_p] \right)^2 &= \lambda^2 \sum_n \sum_p s^2[\phi_p] + 2\lambda^2 \sum_n \sum_p \sum_{q, q \neq p} s[\phi_p] s[\phi_q] \\
&= \lambda^2 \sum_p \underbrace{\left(\sum_n s^2[\phi_p] \right)}_{R_{ss}(0)} + 2\lambda^2 \sum_p \sum_{q, q \neq p} \underbrace{\left(\sum_n s[\phi_p] s[\phi_q] \right)}_{R_{ss}(\phi_p - \phi_q)} \\
&= \lambda^2 N_{mic} R_{ss}(0) + 2\lambda^2 \sum_{p,q,p \neq q} R_{ss}(\phi_p - \phi_q)
\end{aligned} \tag{27}
$$

and

$$
\begin{aligned}
\sum_n \left(\sum_p w_p[\psi_p] \right)^2 &= \sum_n \sum_p w_p^2[\psi_p] + 2 \sum \sum_p \sum_{q, q \neq p} w_p[\psi_p] w_q[\psi_q] \\
&= \sum_p \underbrace{\left(\sum_n w_p^2[\psi_p] \right)}_{R_{w_p w_p}(0)} + 2 \sum_p \sum_{q, q \neq p} \underbrace{\left(\sum_n w_p[\psi_p] w_q[\psi_q] \right)}_{R_{w_p w_q}[\phi_p - \phi_q] = 0} \\
&= N_{mic} R_\eta(0)
\end{aligned} \tag{28}
$$

The second term in Equation (28) is zero due to uncorrelated noises on different microphones. Substituting Equation (27) and Equation (28) back into Equation (26), we have

$$
B^2(\alpha) = \lambda^2 N_{mic} R_{ss}(0) + 2\lambda^2 \sum_{p,q,p \neq q} R_{ss}(\phi_q - \phi_p) + N_{mic} R_\eta(0) \tag{29}
$$

Rearranging the terms and denoting $\kappa_{pq} = \phi_q - \phi_p = t_{pq}(\alpha) - \tau_{pq}$

$$
B^2(\alpha) = N_{mic}(\lambda^2 R_{ss}(0) + R_\eta(0)) + 2\lambda^2 \sum_{p,q,p \neq q} R_{ss}(\kappa_{pq}) \tag{30}
$$

Adding and subtracting the term $2\frac{N_{mic}(N_{mic}-1)}{2}(\lambda^2 R_{ss}(0)+R_\eta(0))$, we have

$$\begin{aligned}
B^2(\alpha) &= N_{mic}(\lambda^2 R_{ss}(0)+R_\eta(0)) + \underbrace{2\lambda^2\sum_{p,q,p\neq q} R_{ss}(\kappa_{pq}] + 2\frac{N_{mic}(N_{mic}-1)}{2}(\lambda^2 R_{ss}(0)+R_\eta(0))} \\
&\quad - 2\frac{N_{mic}(N_{mic}-1)}{2}(\lambda^2 R_{ss}(0)+R_\eta(0)) \\
&= \sum_{p,q,p\neq q} \underbrace{2\left(\lambda^2 R_{ss}(\kappa_{pq})+\lambda^2 R_{ss}(0)+R_\eta(0)\right)}_{B^2_{pq}(\alpha) \text{ from Proposition 2}} - N_{mic}(N_{mic}-2)(\lambda^2 R_{ss}(0)+R_\eta(0)) \\
&= \sum_{(p,q)\in\mathbb{P}} B^2_{pq}(\alpha) - N_{mic}(N_{mic}-2)(\lambda^2 R_{ss}(0)+R_\eta(0))
\end{aligned}$$

□

A.4 Proof of Proposition 4

Proof. We prove Proposition 4 by derivation. Let there be M sources present at angles θ_m emitting source signals $s_m[n]$, for $m = 1, 2, \cdots, M$. Using Equation (1), the received signal at the p^{th} microphone is given by

$$r_p[n] = \sum_{m=1}^{M} \lambda_{mp} s_m[n-\tau_{mp}] + w_p[n]$$

where $p = 1, 2, \cdots, N_{mic}$, τ_{mp} is the propagation delay, and λ_{mp} is the attenuation factor. For far-field case, the distances between a source and the closely-spaced microphones will be approximately same for all microphones, hence $\lambda_{m1} \approx \lambda_{m2} \approx \cdots = \lambda_m$.

Using Equation (3), the composite microphone signal for the beam angle α is given by

$$\begin{aligned}
r[n] &= \sum_{p=1}^{N_{mic}} r_p[n+t_{1p}(\alpha)] \\
&= \sum_{p=1}^{N_{mic}} \sum_{m=1}^{M} \lambda_m s_m[n+t_{1p}(\alpha)-\tau_{mp}] + w_p[n+t_{1p}(\alpha)]
\end{aligned}$$

where $t_{1p}(\alpha) = t_p(\alpha) - t_1(\alpha) = d_{1p}\cos(\alpha-\beta_{1p})f_s/C$ is relative sample delay between the p^{th} and 1^{st} microphone. Let's denote $\phi_{mp} = n + t_{1p}(\alpha) - \tau_{mp}$ and $\psi_p = n + t_{1p}(\alpha)$ for clarity and brevity.

The beam energy is given by

$$
\begin{aligned}
B^2(\alpha) &= \sum_n r[n]^2 \\
&= \sum_n \left[\sum_p \sum_m \lambda_m s_m[\phi_{mp}] + w_p[\psi_p] \right]^2 \\
&= \sum_n \left[\left(\sum_p \sum_m \lambda_m s_m[\phi_{mp}] \right)^2 + \left(\sum_p w_p[\psi_p] \right)^2 + 2 \sum_p \sum_q \sum_m \lambda_m s_m[\phi_{mp}] w_q[\psi_q] \right] \\
&= \sum_n \left(\sum_p \sum_m \lambda_m s_m[\phi_{mp}] \right)^2 + \underbrace{\sum_n \left(\sum_p w_p[\psi_p] \right)^2}_{N_{mic} R_\eta(0) \text{ using Equation (28).}} \\
&\quad + 2 \sum_p \sum_q \sum_m \underbrace{\sum_n \lambda_m s_m[\phi_{mp}] w_q[\psi_q]}_{R_{w_q s_m}(\tau)=0}
\end{aligned}
\tag{31}
$$

The first term in Equation (31) is expanded to

$$
\begin{aligned}
\sum_n \left(\sum_p \sum_m \lambda_m s_m[\phi_{mp}] \right)^2 &= \sum_n \left(\sum_m \sum_p \lambda_m s_m[\phi_{mp}] \right)^2 \\
&= \sum_n \left(\sum_m \left(\sum_p \lambda_m s_m[\phi_{mp}] \right)^2 + 2 \sum_{m_1} \sum_{m_2} \lambda_{m_1} s_{m_1}[\phi_{m_1 p}] \lambda_{m_2} s_{m_2}[\phi_{m_2 p}] \right) \\
&= \sum_m \left(\sum_n \left(\sum_p \lambda_m s_m[\phi_{mp}] \right)^2 \right) + 2 \sum_{m_1} \sum_{m_2} \underbrace{\sum_n \lambda_{m_1} s_{m_1}[\phi_{m_1 p}] \lambda_{m_2} s_{m_2}[\phi_{m_2 p}]}_{R_{s_{m_1} s_{m_2}}(\tau)=0} \\
&= \sum_m \underbrace{\left(\sum_n \left(\sum_p \lambda_m s_m[\phi_{mp}] \right)^2 \right)}_{\text{substitute from Equation (27)}} \\
&= \sum_m \lambda_m^2 \left(N_{mic} R_{s_m s_m}(0) + \sum_{p,q,p \neq q} R_{s_m s_m}(\phi_{mp} - \phi_{mq}) \right)
\end{aligned}
\tag{32}
$$

Denoting $\kappa_{mpq} = \phi_{mq} - \phi_{mp} = t_{pq}(\alpha) - \tau_{mpq}$, and substituting Equation (32) in Equation (31), we have

$$
B^2(\alpha) = \sum_m \lambda_m^2 \left(N_{mic} R_m(0) + \sum_{p,q,p \neq q} R_m(\kappa_{mpq}) \right) + N_{mic} R_\eta(0)
$$

Adding and subtracting the term $\sum_m N_{mic} R_\eta(0)$, we have

$$
\begin{aligned}
B^2(\alpha) &= \sum_m \lambda_m^2 \left(N_{mic} R_m(0) + \sum_{p,q,p\neq q} R_m(\kappa_{mpq}) \right) + \sum_m N_{mic} R_\eta(0) - \sum_m N_{mic} R_\eta(0) + N_{mic} R_\eta(0) \\
&= \sum_m \underbrace{\lambda_m^2 \left(N_{mic} R_m(0) + \sum_{p,q,p\neq q} R_m[\kappa_{pq}] \right)}_{B_m^2(\alpha), \text{ using Equation (30)}} + N_{mic} R_\eta(0) - M N_{mic} R_\eta(0) + N_{mic} R_\eta(0) \\
&= \sum_m B_m^2(\alpha) - N_{mic}(M-1) R_\eta(0)
\end{aligned}
$$

□