Poster Abstract: A Beamforming Method for Multiple Source DOA Estimation, Spectrum Separation and Localization from Field Data

Juo-Yu Lee
Electrical Engineering Department
University of California, Los Angeles, CA, USA
juoyul@ucla.edu

Zac Harlow, Travis C. Collier,
Charles E. Taylor
University of California, Los Angeles, CA, USA
{zac.harlow, travcollier}@gmail.com,
taylor@biology.ucla.edu

Kung Yao
Electrical Engineering Department
University of California, Los Angeles, CA, USA
yao@ee.ucla.edu

ABSTRACT
In this abstract, we present a beamforming method for estimating the directions and locations of multiple sources and separating each source’s spectrum from field data collected by a wireless acoustic sensor network. Each acoustic sensor is equipped with four microphones that receive acoustic signals in a time-synchronized manner. The difference in time-of-arrival of proximal signals depends on the source direction with respect to the geometry of the microphone array. We show that by using beamforming in the frequency domain, the locations and Direction-Of-Arrivals (DOAs) of multiple 3D sources may be estimated, and the source spectrum may be separated from the audio data spectra.

Categories and Subject Descriptors
C.3.5 [Special-Purpose and Application-based Systems]: Signal processing systems

General Terms
Algorithms, Design, Measurement

Keywords
Acoustic sensor, DOA, Localization, Source Separation

1. INTRODUCTION
We consider the problem of localization, DOA estimation and source separation for multiple near-field wideband sources using a wireless acoustic sensor network and the maximum likelihood (ML) estimator. In our approach, we first apply Discrete Fourier transform (DFT) to the received wideband acoustic signals, and obtain the narrowband signals in each frequency bin [1]. We then explore all frequency bins of our interest and derive the optimal parametric ML solution to estimate DOAs of wideband sources. ML-based optimization for the source DOAs may be made possible without explicitly calibrating the relative time delay. The number of sources has been determined by techniques such as principal component analysis (PCA) independent of the underlying ML estimator. To reduce computation complexity, we only consider frequency bins of significant energy levels, which yield a reasonable approximate ML solution.

2. SYSTEM MODEL
Suppose the reference waveform \( s(\cdot) \) in our system model is a signal arriving at the origin of the coordinate system \( O \) at time zero. Without loss of generality, we consider an array (centered at the origin) of \( P \) microphones equipped on a sensor. In Cartesian coordinate system, each microphone is located at the position \( \mathbf{r}_p = [u_p, y_p, z_p]^T \), \( 1 \leq p \leq P \), where superscript \( T \) denotes matrix transposition. We assume \( M \) wideband sources in the near-field of the array are located at unknown points \([u_m, y_m, z_m] \), \( 1 \leq m \leq M \). The relative time delay of the \( m \)th source is given by

\[
t_{cp}^{(m)} = t_c^{(m)} - t_p^{(m)} = \frac{1}{c} \sqrt{(u_m - u_p)^2 + (y_m - y_p)^2 + (z_m - z_p)^2}
\]

where \( t_c^{(m)} \) and \( t_p^{(m)} \) are the absolute time delays from the \( m \)th source to the coordinate origin and to the \( p \)th microphone, respectively, and \( v \) is the speed of sound (nominally, 345 m/s at room temperature). Hence, the data received by the \( p \)th microphone at sample time \( t = t_n \) is given by

\[
x_p(t_n) = \sum_{m=1}^{M} s^{(m)}(t_n - t_{cp}^{(m)}) + w_p(t_n), n = 0, \ldots, N - 1,
\]

where \( N \) is the number of signal samples, \( s^{(m)}(n) \) is the \( m \)th source signal at the array center, and \( w_p(n) \) is the zero mean white Gaussian noise with variance \( \sigma^2 \). Note that in the above equation, \( t_{cp}^{(m)} \) can be an arbitrary positive or negative scalar.

The majority of the data processing performed by the ML estimator takes place in the frequency domain. Given the \( N \) point DFT \( X_p(\omega_0) \) and \( S^{(m)}(\omega_0) \) of \( x_p(n) \) and \( s^{(m)}(n) \), we aim to search for optimal \([u_m, y_m, z_m]^T \) by minimizing the residual error defined as follows.

\[
J([u_m, y_m, z_m]_{m=1}^{M}) = \sum_{\omega_0} \sum_{m=1}^{M} \left| X_p(\omega_0) - \sum_{n=1}^{M} e^{-j2\pi(n,u_m,y_m,z_m)} S^{(m)}(\omega_0) \right|^2
\]

We first cancel the residual error dependency on \( S^{(m)}(\omega_0) \) and solve for optimal locations, which can be used to solve for \( S^{(m)}(\omega_0) \), hence completing source separation. For each source, we perform search over all possible points in a hypothesized 3D grid to find the optimal \([u_m, y_m, z_m]^T \). We apply an efficient alternating
projection procedure [2] by sequentially estimating the location of one source while nulling out the estimate of other source locations from the previous iteration.

3. FIELD DATA COLLECTION

We deployed an acoustic sensor network of eight nodes for field data collection (Fig. 1). A self-survey procedure was activated to localize each sensor’s microphones [3]. Fig. 1 (left) shows the network topology projected onto a 2D plane. Fig. 1 (right) shows the interior view of an acoustic sensor. Each sensor receives audio signals through four microphones and stores the audio streams into WAV files with four channels. Each single channel audio stream was sampled at 48 kHz. The spectrum carrying significant acoustic energy ranges from 1295 Hz to 2259 Hz.

![Image](image1.png)

Figure 1. An acoustic sensor network of eight nodes. Left: network deployment. Right: an acoustic sensor node.

![Image](image2.png)

Figure 2. Location estimation (projected on one horizontal plane) for source A (left) and source B (right) based on the field data received by node #103. Actual localization results: source A is located at [8, 1, 11] (meters) and source B is located at [12, -11, 1] (meters).

4. RESULTS AND CONCLUSION

The ML estimator yields estimated locations based on segmented audio streams. Source DOAs may be computed, and source spectra may be separated. Three audio segments were chosen that contain overlapped multiple source audio streams. Figure 2 shows the localization results based on the field data collected by node #103 (in Fig. 1, left). The example described herein yields location estimates for one source at [8, 1, 11] (meters) and for a second source at [12, -11, 1] (meters). Given the received audio spectrum and localization results, one can separate source spectrum (as if measured at the estimated locations). Fig. 3 shows recovered and separated waveforms and spectra based on one sample audio segment (of 0.5 seconds) from the data set “Example 1” collected by Node 103. The DOA of source A (Table 1) with respect to node #103 is 85.6962 (degrees, azimuth)/33.1974 (degrees, elevation). The DOA of source B (Table 1) with respect to node #103 is -22.8344 (degrees, azimuth)/-33.7003 (degrees, elevation).

![Image](image3.png)

Figure 3. Source separation for two sources, based on data (“Example 1”) collected by node #103.

Table 1. DOAs of two sources (A and B in Figure 2) with respect to node #103.

<table>
<thead>
<tr>
<th>Source</th>
<th>Azimuth (degrees)</th>
<th>Elevation (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source A</td>
<td>85.6962</td>
<td>33.1974</td>
</tr>
<tr>
<td>Source B</td>
<td>-22.8344</td>
<td>-33.7003</td>
</tr>
</tbody>
</table>

5. ACKNOWLEDGMENT

This work was supported in part by the U.S. National Science Foundation (NSF) grant IIS1125423. The authors acknowledge the valuable comments from anonymous reviewers and Dr. Lewis Girod. Dr. Girod is one of the main developers of the acoustic nodes used in the experiments.

6. REFERENCES

