NEW WIRELESS ACOUSTIC ARRAY NODE FOR LOCALIZATION, BEAMFORMING AND SOURCE SEPARATION FOR BIO-COMPLEXITY BIRD DATA COLLECTION AND STUDY


#: IACAS, Beijing 100190, China
+: MIT, Cambridge, MA 02142
^: UCLA, Los Angeles, CA 90095
*: Zhejiang University, Hangzhou 310027, China

ABSTRACT

In this paper, we present simulation and experimental studies of localization and source separation for a bio-complexity bird study based on a new wireless acoustic array node named Voxnet. The Approximate Maximum Likelihood (AML) method is used to estimate blindly the direction-of-arrivals (DOAs) of sources which can be used to both locate the sources using a DOA-based localization method and generate the steering vectors in order to separate the sources via beamforming. Simulation and measured data confirmed the proper operations of the AML-based 2-D/3-D localization and 2-D source separation, as well as the Voxnet hardware nodes.

Index Terms—Voxnet, AML, Localization, beamforming, Source separation

1. INTRODUCTION

Animal vocalizations provide an important way for biologists to discover the presence of animals and to study their behavior. In practice, the natural movement patterns of animals are unpredictable and animals need to be located even in environments characterized by visual obstructions. Often, more than one animal/species will emit sounds simultaneously. Biologists want to separate the mixed acoustic signals, find out the accurate locations of the animals/species and identify the source. However, even experienced biologists may have difficulty identifying the animals/species from the mixed signals. It’s even impossible, even for humans, to accurately locate the animals/species in a complicated environment. Thus, the use of an embedded acoustic array can aid biologists in the difficult and time-consuming tasks of detection, localization, and separation, in order to focus on their detailed bio-complexity research.

DOA-based localization method using a distributed set of nodes is one of the most common ways to locate an acoustic source. Each node is able to detect the sources and produce DOA bearing estimation independently. Then the crossing of these bearing estimates are combined to compute an estimate of the most likely source location. In such a DOA-based localization system, the accurate relative locations among the nodes is a very important issue. The system performance may suffer dramatically due to sub-arrays’ positions uncertainty (due to unreliable measurement in random deployment). In our implementation, we used the Voxnet acoustic array node which is a newly developed wireless, self-localization system that can be easily used in the field for collecting, detecting, locating, and separating acoustic signals. The highly accurate self-localization feature (less than 4 cm error for a node) of the system is utilized.

This paper describes an application on multiple source DOA estimation, localization, and source separation for bird songs using Voxnet acoustic array nodes. The AML and Alternating Projection (AP) algorithms [1] are used to estimate the DOAs of multiple bird sources and generate the steering vectors of the DOAs to perform beamforming to separate the mixed acoustic signals of the multiple sources. By combining the crossing of the DOA bearings estimated by a set of nodes, the most likely source location can be obtained. We evaluated the performance of the AML beamforming and AML localization-based method first by simulation and then from measured data collected by Voxnet acoustic array nodes.

2. OVERVIEW OF APPROACH

2.1. Approximate maximum likelihood

For an arbitrarily distributed array of J microphones, the data collected by the j-th microphone can be given by

$$x_j(t) = \sum_{m=1}^{M} s^{(m)}(t - t_j^{(m)}) + n_j(t),$$

where $M$ is the number of sources, $s^{(m)}$ is the m-th source signal, $t_j^{(m)}$ is the fractional time-delay and $n_j$ is the noise.

In the frequency domain, the model is given by

$$X(k) = D(k)S(k) + N(k),$$

for $k = 0, ..., L - 1$, where the array data spectrum is given by $X(k) = [x_1(k), ..., x_J(k)]^T$, the steering matrix is given by $D(k) = [d^{(1)}(k), ..., d^{(M)}(k)]$, the steering vector is given...
by \( d^{m}(k) = [d^{m}_1(k), ..., d^{m}_j(k)]^{T}, \) the source spectrum is given by \( S(k) = [S^{1}(k), ..., S^{M}(k)]^{T} \) and the noise spectrum \( N(k) \) is a zero mean i.i.d. Gaussian vector. The AML source locations estimate can be obtained by solving the following maximization problem:

\[
\max J(\tilde{r}) = \max \sum_{k=1}^{L} \| P(k, \tilde{r}) X(k) \|^{2}, 
\]

where the orthogonal projection is given by

\[
P(k, \tilde{r}) = D(k) D^{\dagger}(k), \quad D^{\dagger}(k) = (D^{*}(k) D(k))^{-1} D^{*}(k).
\]

\( J(\tilde{r}) \) is the AML matrix and \( \tilde{r} \) is the source locations [2]-[3].

2.2. Beamformer

Beamforming method is used to separate the mixed signals after the AML estimate of \( \tilde{r} \) is obtained. Data from \( M \) sources collected by the \( J \) microphones of (1) are used to form the beamformer in the frequency domain given by

\[
Y(k) = D^{\dagger}(k, \tilde{r}) X(k),
\]

where \( Y(k) \) is the beamformed spectrum vector for the \( M \) sources. Information on the DOAs of sources obtained by AML estimation is contained in \( D^{\dagger}(k, \tilde{r}) \). The AML beamformer performs source separation by exploiting the wavefront differences of the multiple sources.

2.3. DOA-based Localization

We assume each microphone receives only one direct path ray from a source. Using the AML method, a node can estimate the DOA of the source. If another node is not colinear with the first one and the source, then the crossing of these two DOAs yields the location of the source. In practice, there will be an uncertainty region about the location of the source on account of uncertainties in the DOA estimation. However, by using three of more noncolinear sub-arrays and appropriate least-squares estimation technique [4], the uncertainty region can be reduced. This technique can be extended to multiple sources.

3. VOXNET ACOUSTIC ARRAY NODE

3.1. Hardware

The new version of the Voxnet node is a compact, self-contained package with integrated GPS, radio, temperature and humidity sensors, and speakers, as shown in Figure 1. The node is based on an embedded computer (running Linux) and also contains a low-power micro-controller (a MSP430) which acts as a supervisory coprocessor. Each node is a wireless recording and processing device which has a tetrahedral microphone array (63.5 mm between microphones) with four-channel 16 bits ADC and 48 kHz sampling rate per channel. The node has a 433 MHz packet radio module and a GPS module. Both of them can offer a precise time-synchronization for the whole system. Two 50mm*50mm, 5W speakers which can be used for self-localization of the node are mounted at two sides of a laser-cut plastic box. The node is powered by a 14.4 V and 2.9 Ah lithium-ion battery working for up to 20 hours. Additional details are given in [5]-[6].

Figure 1. Voxnet node and array geometry

3.2. Self-localization service

The self-localization service of Voxnet node includes two steps: range measurement and localization. We use Beepbeep, a two-way ranging scheme in the range measurement process [7]. For simplicity, we introduce the ranging scheme between two nodes here, but can extend it to multiple nodes easily based on the same concept. Firstly, node A emits a beep signal through its speaker. The signal will be recorded by both its own microphone and the other node B. Then node B emits another beep signal through its speaker. This signal is also recorded by both nodes. Afterwards, the two nodes count the number of samples between these two beeps and exchange the time duration information by radio. At last, both nodes can derive the two-way time of flight of the beeps at the granularity of sound sampling rate.

After obtaining the range information between nodes, simple geometric methods can be used to compute the relative positions among the nodes.

4. SIMULATION

In the simulation, we assumed a four-channel node with the same geometric position of microphone sensors as the Voxnet node. We used recordings of songs from Bewick’s Wren (Thryomanes bewickii) (BW) and Black-headed Grosbeak (Pheucticus melanoleucus) (BHG) as the sources.

4.1. Localization

Assume two nodes arranged at locations of (0m,0m) and (20m,0m), a BW sings at the location (10m,20m). We obtain the true DOAs of the source relative to the two nodes.
at 63.4° and 116.6°. Using the AML method to estimate the source by the two nodes independently, the estimated results are shown in the Figure 2. The estimated DOAs are quite close to the true DOAs. Then the position of the source is obtained by crossing the two estimated DOAs, and the localization result is shown in the Figure 3. The error of the localization in the simulation is about 30 cm.

![Figure 2. DOA estimation results of two nodes by AML](image)

![Figure 3. Localization result by crossing the AML DOAs](image)

4.2. Source separation
Assume a BW signal and a BHG signal arrive at the four-channel node from directions 180° and 320°. The estimated AML DOA is shown in Figure 4. This simulation estimated the DOAs to be the same as the true DOAs. From the estimated DOAs, we generated steering vectors and were able to separate the sources by beamforming. The source separation results are shown in Figure 5. The 4th and 5th sub-pictures are the separated signals from the mix signal (3rd sub-picture) by beamforming, and are quite similar to the original signals (1st and 2nd sub-pictures).

![Figure 4. DOA estimation results by AML in simulation](image)

![Figure 5. Sources separation results from simulation](image)

5. FIELD MEASUREMENTS
In the field measurement, 3 nodes were used for 2D estimation (shown in Figure 6); 4 nodes were used for 3D estimation. The largest range between two nodes is around 13 m. The field measurement experiment falls into 3 steps:
1) Locate the nodes by using self-localization of the nodes;
2) One node plays a bird song, other two locate the source;
3) Two nodes play bird songs, third one separate the sources.

![Figure 6. Field measurement scene](image)

5.1. Range measurement and self-localization
Two processes were needed in the first step:
Firstly, measure the ranges between the nodes. We also use a laser measurement device to obtain more accurate ranges among the nodes and check the performance of range measurement operation of the nodes. The range measurement results were shown in the Table 1(using the laser) and Table 2 (using the nodes). We note both the laser data and the nodes data are quite stable, and the average error between the laser data and the nodes data is less than 4 cm. We can conclude the range measurement operation of our nodes performs accurately with low variations.
Secondly, by using some simple geometry, ranges among the nodes are obtained, and the locations of the nodes can be obtained as shown in the Figure 7.
## 5.2. 2-D and 3-D Localizations

For 2-D localization, at the second step, node3 played the BW song as a source, node1 and node2 recorded the acoustic signal, then estimated the DOA of the source independently. Then combined the DOA information from the two recorders and located the source by crossing the DOA bearings. The localization result is shown in Figure 8, the error between the true and the estimated source location is about 33.8 cm.

Now, consider the 3-D localization [3] of a source (located at 270° azimuth and 0° elevation), by playing the BHG bird song at node 3. Nodes 1, 2, and 4 recorded the song of about 99 seconds long. The maximum value in Figure 9 shows the result of the 3-D AML-based estimated source location at 273° azimuth and 8° elevation.

## 5.3. 2D source separation

At the third step, node1 and node3 played the songs of BHG and BW simultaneously, node2 recorded the mixed acoustic signals. Then node2 estimated the DOAs of the two sources using AML/AP algorithms and generated the steering vectors from the DOA information. Finally, node2 separated the two sources using the beamforming method. Figure 10 shows the source separation results, the separated signals (4th and 5th sub-pictures) are quite close to the original signals (1st and 2nd sub-pictures). Thus the two bird sources were separated from the mixed signal recorded by the nodes.

## 6. CONCLUSIONS

In this paper, we presented both simulation/field collected data for bird song 2-D localization and source separation and 3-D localization by using the AML-based methods utilizing a set of Voxnet nodes.

## 7. ACKNOWLEDGMENTS

This work was supported by the NSF grant IIS-1125423, NSFC grant No.61273079 and No.61104208, and the “Strategic Priority Research Program” of the Chinese Academy of Sciences, Grant No. XDA06020201.
8. REFERENCES


