Fast 3D AML-Based Bird Song Estimation

Jiawei Zhang*,‡,§, George Kossan†, Richard W. Hedley†, Ralph E. Hudson*, Charles E. Taylor†, Kung Yao*, Ming Bao‡

*Electrical Engineering Department, UCLA, Los Angeles, CA 90095, USA
†Ecology and Evolutionary Biology Department, UCLA, Los Angeles, CA 90095, USA
‡Lab of Communication Acoustics, IACAS, Beijing 100190, China

In this paper, we present simulations and experimentally collected bird song data collected using a modified Voxnet acoustic array node (with four microphones) to perform 3D direction-of-arrival (DOA) estimation of various bird sources. We used the Approximate Maximum-Likelihood (AML) algorithm to construct the steering matrix in the beamforming process for the estimation of the DOA of the bird signals. While the computational burden is high in the 3D scenario, various strategies have been developed to reduce the computational burden of the algorithm for potential real-time applications. Extensive simulations and experimentally collected data are used to validate the effectiveness of the AML algorithm for 3D estimations and the usefulness of the modified Voxnet node. Both the estimated azimuths and elevations have approximately plus and minus 10 degrees of errors.

Keywords: AML; 3D DOA; beamforming.

1. Introduction

Biologists frequently identify animals and study their behaviors through their vocalizations. In the field, the natural movement patterns of animals are unpredictable and animals are often located in environments with visual obstructions. In addition, more than one animal may emit sounds simultaneously making it even more difficult for biologists to identify animals from the mixed signals. In practice, it may be difficult for anyone to locate the animals rapidly and accurately in sufficiently complicated environments. An embedded acoustic array can often aid biologists to obtain the 3D direction-of-arrivals (DOAs) of the animals and separate the signals [1, 2].

An improved method for one or two source 3D DOA estimation and source separation for bird songs using a modified Voxnet acoustic array node is presented in this paper. The general Approximate Maximum Likelihood (AML) [3] method, whose performance approaches the Cramer–Rao Bound (CRB) under high signal-to-noise ratio (SNR), is used to estimate the DOAs of bird sources as well as generate the steering matrix. Then the beamforming method can be utilized to separate sources from the mixed acoustic signals, whereas, if azimuths and elevations of the two sources need to be estimated, the general AML algorithm has to search in a 4-dimensional grid which is time consuming. To reduce the complexity of the general 3D AML algorithm, the sparsity property of bird signals in time-frequency cells is utilized in this paper. It is very likely that the two birds may sing at different time and dominate different frequency bins as well. Thus most of the active time-frequency cells can be separated into two groups and the DOAs of the two sources can be accurately estimated in terms of these two groups by using a single source AML algorithm. In addition, a decoupled 3D AML algorithm [4], which estimates azimuth first and then finds the elevation, is used to further decrease the computational burden by utilizing the property of a 3D isotropic array [5] in a single source scenario. After obtaining the DOAs of two sources,
the steering matrix can be constructed to separate the two sources by implementing a beamforming algorithm.

This paper is organized as follows. Section 2 will give an introduction to DOA and localization of a bird source in 3D. The overview of general 3D AML is presented in Sec. 3. In order to reduce the complexity of the algorithm, we discuss a decoupled 3D AML algorithm in Sec. 4 and introduce a source separation method in Sec. 5. The beamforming algorithm is presented in Sec. 6. Section 7 describes the structure and parameters of the modified Voxnet node as well as the basic parameters of our algorithm. Extensive simulations and experimentally collected data are used to validate the effectiveness of the proposed 3D AML algorithm in Secs. 8 and 9, respectively. We draw some conclusions in Sec. 10.

2. Introduction to DOA and Localization of a Bird Source in 3D

Consider the estimation of the DOA and the location of an active source emitting some energy. In this manuscript, we assume the source is the bird vocalization. There are various known ways to find the DOA and the location of a source.

The Received-Signal-Strength (RSS) is a conceptually simple energy-based source localization method. An acoustic source in free space radiating omni-directionally will attenuate at a rate inversely proportional to the square of the distance. With multiple microphones at known locations, the trilateration method of using the intersection of the circles (in 2D) or spheres (in 3D) of known radii can find the location of the source. However, in practice with foliage shadowing and multiple reflections, the effective attenuation is not simply proportional inversely to the square of the distance, and the RSS method is not practical for animal source localization. In the Time-Difference-Of-Arrivals (TDOA) method, a group of synchronized microphones can estimate the difference of the time delays due to different propagation delays. If the speed of propagation is known, then these time differences are equal to the differences in distance from the source to the microphones, and the location of the source can be estimated again using the trilateration method. In practice, synchronization errors, variation in propagation speed, and multiple reflections, all make the TDOA method to be not practical for acoustic localization of bird sources [6].

For wide-band acoustic signals, we proposed the following maximum-likelihood (ML) array of N microphones for DOA estimation of M bird sources [3]. In this section, we will only introduce qualitatively this method (Sec. 3 provides a detailed quantitative explanation of this method). The received wideband signal is transformed into the frequency domain via the DFT (or FFT), where each frequency bin is then treated as a narrowband signal. Even if the original additive noise in the received waveform is not Gaussianly distributed, the noise in the frequency domain (due to the result of the Central Limit Theorem) will be modeled as Gaussianly distributed. Then the log-likelihood function can be expressed explicitly with a parameter which depends on the azimuth angle (in the 2D case) and depends on the azimuth and elevation angles (in the 3D case). We denote this ML algorithm as an Approximate ML (AML) algorithm, since with finite number of terms in the DFT (FFT) operation, edge effect makes a slight degradation from a true ML algorithm. A single source location (either only in azimuth angle or in both azimuth and elevation angles) can be estimated by performing a AML optimization on the log-likelihood function. For multiple source locations, a wideband signal extension of the narrowband signal Alternating Projection Algorithm (APA) [7], can be used. First, estimate the angle(s) of the strongest source using the AML algorithm. Second, search for the angle(s) of the second strongest source using the AML algorithm subject to constraining the estimate of the first source constant. Third, repeat for the next source, etc. At the M step, do an AML estimation for the location of the first source again. Typically, only few iterative cyclings of the above operations are needed. For M = 2, typically two cycling of the above APA operations yield converged estimated angle(s). Our prior experiences with the AML DOA estimation method show that the AML algorithm has been effective for both 2D as well as 3D estimations. Of course, in order to find the location of a source, we need to use several array nodes whose locations are known. Then the bearing crossings of the DOAs estimated from two or more such array nodes can yield the location of the sources. An earlier work of ours in 2D and preliminary 3D AML arrays appeared in [8].

In several well-known papers and books on the study of birds [9–11], an ability to estimate the location of the birds, particularly with respect to the height of the location, is important. As discussed in [11], both the Philadelphia Vireo (Vireo philadelphicus) and the Red-eyed Vireo (Vireo olivaceous) reside over the same area (i.e., similar azimuth angles). However, the Red-eyed vireos do their foraging in the mid-canopy levels while the Philadelphia vireos do their foraging in the upper-canopy levels. Thus, the use of a 2D DOA estimation array is not adequate to understand and study the different behaviors of these two kinds of vireos. The use of a 3D DOA estimation array can not only determine the heights of the two different vireos, but also the beamforming capability of the 3D AML array can provide SNR enhancement of the signal of one kind of vireo, while simultaneously it can provide nulling (or attenuation) of the signal from another kind of vireo at a different height (i.e., a different elevation angle) even when their azimuth angles...
may be similar. Thus, as expounded in [1], the availability of modern array processing with beamforming enhancement and nulling can provide significant capability in bird research not possible with a single microphone recording system.

3. Overview of General 3D AML

Consider a sensor array of $N$ omnidirectional microphones is impinged by $M$ sources $s^{(m)}$ from $M$ distinct directions in the far field. The data collected by the $n$th microphone is given by

$$x^{(n)}(t) = \sum_{m=0}^{M-1} s^{(m)}(t - t^{(n)}(\theta^{(m)}, \phi^{(m)})) + w^{(n)}(t),$$

in which $t^{(n)}(\theta^{(m)}, \phi^{(m)}) = (x^{(n)} \cos(\theta^{(m)}) \cos(\phi^{(m)}) + y^{(n)} \sin(\theta^{(m)}) \cos(\phi^{(m)}) + z^{(n)} \sin(\phi^{(m)}))/c$ is the time delay of the $m$th source in the $n$th microphone relative to the centroid of the array, $r^{(n)} = [x^{(n)}, y^{(n)}, z^{(n)}]^T$ is the location of $n$th microphone, $\theta^{(m)}$ and $\phi^{(m)}$ are azimuth and elevation of the $m$th source, $c$ is the speed of the sound which is assumed to be 345 m/s and $w^{(n)}$ is the noise.

If all the data are separated into $H$ successive frames, the array model signal of the $h$th frame in the $k$th frequency bin, which is defined as a time-frequency cell, is given by

$$x_k(h) = D_k(\theta) s_k(h) + w_k(h),$$

in which $x_k(h) = [s_k^{(0)}(h), \ldots, s_k^{(N-1)}(h)]^T$, $D_k(\theta) = [d_k(\theta^{(0)}, \phi^{(0)}), \ldots, d_k(\theta^{(M-1)}, \phi^{(M-1)})]^T$ is the steering matrix, $d_k(\theta^{(m)}, \phi^{(m)}) = \exp(-2\pi f_k \kappa r^{(n)}(\theta^{(m)}, \phi^{(m)})/K)$, $\exp(-2\pi f_k \kappa r^{(n)}(\theta^{(m)}, \phi^{(m)})/K)$, $\ldots$ is the $m$th steering vector, the DOAs of all the sources are $\Theta = [\theta^{(0)}, \phi^{(0)}], \ldots, \theta^{(M-1)}, \phi^{(M-1)}]^T$, the frequency components of sources are given by $s_k(h) = [s_k^{(0)}(h), \ldots, s_k^{(N-1)}(h)]^T$, the noise spectrum $w_k(h)$ is a zero mean i.i.d Gaussian vector and $F_k$ is the sampling rate of the array.

According to [3], the general 3D AML DOA estimation can be obtained by solving the following maximization problem

$$\max_{\Theta} J(\Theta) = \max_{\Theta} \sum_{k=0}^{K/2-1} \sum_{h=0}^{H-1} \| P_k(\Theta)x_k(h) \|^2,$$

in which $P_k(\Theta) = D_k(\Theta)D_k^T(\Theta)$ is the orthogonal projection and $D_k^T(\Theta) = (D_k(\Theta)D_k(\Theta))^(-1)$ is the pseudo inverse.

For the single source scenario, Eq. (3) can be simplified as

$$\max_{\theta, \phi} J(\theta, \phi) = \max_{\theta, \phi} \sum_{k=0}^{K/2-1} \sum_{h=0}^{H-1} \| D_k^T(\theta, \phi)x_k(h) \|^2.$$

4. Decoupled 3D AML

In the single source scenario, the general 3D AML algorithm needs to implement a 2D search which is time consuming. To reduce the calculation burden, [4] proposed a decoupled 3D AML algorithm if the array is 3D isotropic. According to [5], the necessary and sufficient condition for a 3D isotropic array is $B = \alpha I$, in which the array geometry matrix $B = \sum_{n=0}^{N-1} r^{(n)}r^{(n)^T}$, $\alpha$ is a positive constant and $I$ is the $3 \times 3$ identity matrix.

From [4] and [5], it is clear that the Fisher information matrix is diagonal for an isotropic array which means the azimuth and the elevation can be estimated separately. The decoupled 3D AML, which only needs two 1D searches, is proposed in [4]. The simulation and experimental results [4] show that the performance of decoupled 3D AML is close to that of general 3D AML while complexity is much lower.

However, the likelihood function (4) may lead to ambiguity result if the inter-spacing of the microphones is too large. (For narrow-band signals, the inter-spacing of the microphones should be less than half-wavelength of the signal to avoid ambiguity.) Even though the ambiguity problem may be mitigated by considering broader frequency components, it is still possible for decoupled 3D AML to find the local maximum point in the likelihood metric.

In order to obtain the global maximum point, $L$ initial elevations $e_0, \ldots, e_{L-1}$ are selected in the azimuth finding process which finds $L$ azimuths $\theta_0, \ldots, \theta_{L-1}$. Then $L$ elevations $\phi_0, \ldots, \phi_{L-1}$ can be obtained according to these $L$ azimuths. Since $J(\theta_0, \phi_0), \ldots, J(\theta_{L-1}, \phi_{L-1})$ are local maximum points in the likelihood metric, global maximum point can be found in these local maximum points if initial elevations are appropriately selected.

5. Source Separation Method

The general method to reduce the computational burden in the multiple sources scenario is to use APA [3, 7]. However decoupled 3D AML is not suitable for APA, since the azimuth estimation and the elevation estimation will influence each other in multiple sources scenario.

For the two birds DOA estimation problem, it is very likely that birds may sing at different times and dominate different frequency bins as well. Thus most of the active time-frequency cells, which have enough energy, can be separated into two groups and the DOAs of the two birds can be accurately estimated from these two groups by using a decoupled 3D AML algorithm.

By using source separation method, a two sources DOA estimation is transformed into two single source DOA.
estimation problem and the decoupled 3D DOA algorithm can be applied to reduce the computational burden. In addition, Eq. (4) instead of Eq. (3) is used to calculate the likelihood function which can further reduce the complexity of the algorithm, since we do not need to calculate the orthogonal projection of the steering matrix.

In order to separate the two sources, signal-to-signal-and-interference ratio (SSIR) [12] should be introduced

\[
\text{SSIR}(k, h, \hat{\theta}, \hat{\phi}) = \frac{\|d_k^H(\hat{\theta}, \hat{\phi})x_k(h)\|^2}{(d_k^H(\hat{\theta}, \hat{\phi})d_k(\hat{\theta}, \hat{\phi}))^T(x_k^H(h)x_k(h)),}
\]

in which \(\hat{\theta}\) and \(\hat{\phi}\) are the estimated azimuth and elevation angles from the group.

Algorithm 5.1 (Source separation method)

**Initialization**

(a) Initialize group label map \(L = 0\), source separation threshold \(\rho\) and iteration number \(I\).

(b) Select at most \(R\) time-frequency cells \(x_{k_{0}}(h), \ldots, x_{k_{R-1}}(h)\) from every frame as active cells and label all these cells as group 0 \(L(k, h) = 1\).

(c) Set \(i = 0\).

**Repeat** (Separate the cells that belong to source 0 and estimate DOA of source 0)

(d) Use the decoupled 3D AML algorithm to calculate DOA of source 0 \([\hat{\theta}_0, \hat{\phi}_0]\) according to the cells in group 0.

(e) Calculate SSIR \(SSIR(k, h, \hat{\theta}_0, \hat{\phi}_0)\) for every cell in group 0.

(f) If \(\sqrt{\text{SSIR}(k, h, \hat{\theta}_0, \hat{\phi}_0)} < \rho\), eliminate these cells from group 0 and label them as group 1 \(L(k, h) = 2\).

(g) \(i \rightarrow i + 1\).

while \(i < I\).

(h) Implement the same procedure from (c) to (g) for the cells in group 1 to separate the cells that belong to source 1 and estimate the DOA of source 1 \([\hat{\theta}_1, \hat{\phi}_1]\).

6. Beamforming

To separate the two sources from the mixed signals, beamforming is used in this paper. After all the DOAs \(\hat{\Theta} = [\hat{\theta}^{(0)}, \hat{\phi}^{(0)}, \ldots, \hat{\theta}^{(M-1)}, \hat{\phi}^{(M-1)}]^T\) are determined, the steering matrix \(\mathbf{D}_k(\hat{\Theta})\) can be constructed, and then the \(k\)th frequency component of all the sources in the \(h\)th frame can be estimated by

\[
\hat{s}_k(h) = \mathbf{D}_k(\hat{\Theta})x_k(h).
\]

7. Hardware and Algorithm Parameters

The modified Voxnet node, which is shown in Fig. 1 and Table 1, is an isotropic array. The node has a tetrahedral microphone array with 4-channel ADC and 48 kHz sampling rate per channel. The inter-spacing of the microphones is 62.9 mm. The original version of Voxnet nodes [13] needed to be modified since the sound reflected by top plate of the node will have an influence on the estimate of the elevation.

The parameters of our algorithm are described as follows. For every frame, \(K = 1024\) points FFT is calculated and \(R = 50\) most active frequency bins are selected. The initial elevations \([e_0, e_1, e_2]\) are \([0^\circ, 20^\circ, 40^\circ]\) in the decoupled 3D AML algorithm. Negative initial elevations are not considered since the bird signal cannot come below the ground. The resolution of the angle grid is \(2^\circ\). In the two sources scenario, the source separation threshold is \(\rho = 0.8\) and iteration number for every source is \(I = 3\). Five kinds of birds — Bewick’s Wren (Thryomanes bewickii) (BEWR), Cassin’s Vireo (Vireo cassinii) (CAVI), California Thrasher (Toxostoma redivivum) (CATH) and Black-headed Grosbeak (Pheucticus melanocephalus) (BHGB1 and BHGB2) are considered in the simulations and the experiments. The spectra of these bird songs are shown in Fig. 2.

8. Simulations

In this section, the simulated DOAs of a single source or two sources are estimated by the decoupled 3D AML algorithm...
and source separation method in noise free condition. Beamforming is used to separate signals in the two sources scenario. We also assume the simulated array has the same geometric placement of microphones as the modified Voxnet node in Fig. 1.

8.1. Single source scenario

Assume a BEWR signal arrives at the acoustic array from the direction \([\theta = 150^\circ, \phi = 50^\circ]\). Figure 3(a) shows the contour plot of the AML likelihood metric for the BEWR. It
Fig. 3. (Continued)

Fig. 4. Separation result. (a) Time-frequency cells of CATH, (b) time-frequency cells of BHGB2 and (c) classified CATH and BHGB2.

Fig. 5. Two sources 3D DOA estimation results from simulations. (a) Estimated DOA of CATH: $\theta_0 = 60^\circ$ and $\phi_0 = 30^\circ$ and (b) estimated DOA of BHGB2: $\theta_1 = 130^\circ$ and $\phi_1 = 10^\circ$. 
can be seen that there are several sidelobes in the plot, since the half-wavelength of dominant frequencies of the signal are smaller than the inter-spacing of the array. To avoid find local maximum point in the decoupled 3D AML algorithm, 3 initial elevations \(e_0, e_1, e_2 = [0^\circ, 20^\circ, 40^\circ]\) are selected in the azimuth finding procedure and we can find the azimuths \(\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2 = [22^\circ, 150^\circ, 150^\circ]\) in Fig. 3(b). Then elevations \(\hat{\phi}_0, \hat{\phi}_1 = [-14^\circ, 50^\circ]\) can be obtained according to these azimuths in Fig. 3(c). At last, the estimated DOA \(\hat{\theta} = 150^\circ\) and \(\hat{\phi} = 50^\circ\) can be obtained by comparing the value of likelihood function among these local maximum points. We do the same simulations for BHGB2 and the estimation results are also accurate in Figs. 3(d)–3(f).

### 8.2. Two sources scenario

Assume a CATH signal and a BHGB2 signal arrive at the acoustic array from the directions \([\theta_0 = 60^\circ, \phi_0 = 30^\circ]\) and \([\theta_1 = 130^\circ, \phi_1 = 10^\circ]\), respectively. Figure 4 demonstrates that the source separation method can identify the time-frequency cells of CATH or BHGB2, and the decoupled 3D AML algorithm can be used to separately estimate the DOAs of the two sources. It can be seen from Fig. 5 that the estimation results are accurate.

### 8.3. Beamforming

After the DOAs of two sources \([\hat{\theta}_0, \hat{\phi}_0, \hat{\theta}_1, \hat{\phi}_1]\) are estimated, then the steering matrix can be constructed. Then beamforming equation (6) can be used to separate simulated mixed signal Fig. 6(c) into Figs. 6(d) and 6(e) which are similar to the original BHGB2 signal in Fig. 6(a) and the original CATH signal in Fig. 6(b).

### 9. Experiments

Two speakers were used in the experiments. Single source and two sources scenarios were considered as in the previous section. In order to obtain the true DOAs of the speakers, a laser measurement device was used to obtain the relative positions between the modified Voxnet nodes and the speakers.
Table 2. Single source 3D DOA estimation results from experiments.

<table>
<thead>
<tr>
<th>BEWR</th>
<th>CATH</th>
<th>CAVI</th>
<th>BHGB1</th>
<th>BHGB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
</tr>
<tr>
<td>122°</td>
<td>16°</td>
<td>120°</td>
<td>12°</td>
<td>120°</td>
</tr>
</tbody>
</table>

9.1. Single source scenario

In single source scenario, one speaker played all five kinds of bird songs — BEWR, CATH, CAVI, BHGB1 and BHGB2 from the direction \([\theta = 120.0°, \phi = 15.6°]\) relative to the modified Voxnet. The DOAs are estimated by the decoupled 3D AML algorithm and Table 2 shows the estimation results. Since it is very difficult to obtain the true direction by the laser measurement, these errors are acceptable. In addition, the reflection signal from the ground will have an influence on the elevation estimate. Therefore, the estimated elevations are lower than the measured elevation.

9.2. Two sources scenario

In the two sources scenario, two speakers played bird signals from different azimuths and elevations at the same time and one modified Voxnet was used to record the data. speaker0 was from the direction \([\theta_0 = 120.0°, \phi_0 = 15.6°]\) and speaker1 was from the direction \([\theta_1 = 52.0°, \phi_1 = 6.4°]\). The estimation results are shown in Table 3 which are acceptable.

9.3. Beamforming

According to the DOAs of \([\hat{\theta}_0, \hat{\phi}_0, \hat{\theta}_1, \hat{\phi}_1]\) estimated by the 3D AML, two sources can be separated by beamforming. Figures 8(a) and 8(b) are the spectrums of original BHGB2 and original BEWR. Figure 8(c) is the spectrum of the signal recorded by the acoustic array. The separated spectrums in Figs. 8(d) and 8(e) show that the beamforming algorithm is quite effective in separating the two sources. The time

Table 3. Two sources 3D DOA estimation results from experiments.

<table>
<thead>
<tr>
<th>speaker0: BEWR speaker1: CATH</th>
<th>other signals</th>
<th>speaker0: BEWR speaker1: CAVI</th>
<th>other signals</th>
<th>speaker0: BEWR speaker1: BHGB1</th>
<th>other signals</th>
<th>speaker0: BEWR speaker1: BHGB2</th>
<th>other signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
</tr>
<tr>
<td>120°</td>
<td>14°</td>
<td>52°</td>
<td>0°</td>
<td>120°</td>
<td>14°</td>
<td>52°</td>
<td>0°</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>speaker0: CATH speaker1: BEWR</th>
<th>other signals</th>
<th>speaker0: CATH speaker1: CAVI</th>
<th>other signals</th>
<th>speaker0: CATH speaker1: BHGB1</th>
<th>other signals</th>
<th>speaker0: CATH speaker1: BHGB2</th>
<th>other signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
</tr>
<tr>
<td>118°</td>
<td>14°</td>
<td>52°</td>
<td>0°</td>
<td>118°</td>
<td>14°</td>
<td>52°</td>
<td>0°</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>speaker0: CAVI speaker1: BEWR</th>
<th>other signals</th>
<th>speaker0: CAVI speaker1: CATH</th>
<th>other signals</th>
<th>speaker0: CAVI speaker1: BHGB1</th>
<th>other signals</th>
<th>speaker0: CAVI speaker1: BHGB2</th>
<th>other signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
</tr>
<tr>
<td>120°</td>
<td>14°</td>
<td>52°</td>
<td>0°</td>
<td>120°</td>
<td>14°</td>
<td>52°</td>
<td>0°</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>speaker0: BHGB1 speaker1: BEWR</th>
<th>other signals</th>
<th>speaker0: BHGB1 speaker1: CATH</th>
<th>other signals</th>
<th>speaker0: BHGB1 speaker1: CAVI</th>
<th>other signals</th>
<th>speaker0: BHGB1 speaker1: BHGB2</th>
<th>other signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
</tr>
<tr>
<td>118°</td>
<td>12°</td>
<td>52°</td>
<td>0°</td>
<td>118°</td>
<td>12°</td>
<td>52°</td>
<td>0°</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>speaker0: BHGB2 speaker1: BEWR</th>
<th>other signals</th>
<th>speaker0: BHGB2 speaker1: CATH</th>
<th>other signals</th>
<th>speaker0: BHGB2 speaker1: CAVI</th>
<th>other signals</th>
<th>speaker0: BHGB2 speaker1: BHGB1</th>
<th>other signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
<td>(\hat{\theta})</td>
<td>(\hat{\phi})</td>
</tr>
<tr>
<td>120°</td>
<td>10°</td>
<td>52°</td>
<td>4°</td>
<td>120°</td>
<td>10°</td>
<td>52°</td>
<td>4°</td>
</tr>
</tbody>
</table>
Fig. 7. Beamforming result from experiment in time domain.

Fig. 8. Beamforming result from experiment in spectrum. (a) Spectrum of original BHGB2, (b) spectrum of original BEWR, (c) spectrum of mixed signal, (d) spectrum of separated BHGB2 and (e) spectrum of separated BEWR.
domain signal separation results also validate the effectiveness of the beamforming in Fig. 7.

10. Conclusion

In this paper, both simulated data and experimental data are used to validate the decoupled 3D AML algorithm and the source separation method. It takes a laptop less than 10 s to estimate the azimuths and the elevations whose errors are both less than 10°. In addition, since the modified Voxnet may receive the reflection signal from the ground, the estimated elevation is lower than the true elevation. Furthermore, beamforming is used to separate original bird signals from the mixed signal.

Acknowledgments

This work was supported by the NSF grant IIS-1125423, the National Natural Science Foundation of China (Grant No. 11174316) and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA06020201).

References


Jiawei Zhang received his B.E. degree in Electronic Information Engineering from the University of Science and Technology of China in 2011. He is currently a graduate student in Institute of Acoustics, Chinese Academy of Sciences, where he studies direction-of-arrival estimation.

George Kossan received his B.S. degree in Biology from the University of California, Los Angeles, in 2011. From 2012 to 2014, he was at the University of California, Los Angeles as a research associate in Dr. Charles Taylors lab. He is currently a research associate in the Department of Ecology and Evolutionary Biology, at the University of California, Los Angeles.

George Kossan is the author of over 7 technical publications. His research interests include bird communication, ecology and evolutionary biology, and linguistics.

Richard W. Hedley received his B.S. degree in Biology from the University of Victoria, Canada in 2010. Since 2012, he has been a Ph.D. student in the Ecology and Evolutionary Biology department at the University of California, Los Angeles, USA, where he studies the syntax and functions of bird song.

Ralph E. Hudson received his B.S. in Electrical Engineering from University of California, Berkeley, CA. He received his Ph.D. in optimal estimation and control from U.S. Naval Postgraduate School, Monterey, CA. Then he served as a Lieutenant Commander defining requirements, evaluated proposals and accelerated acquisition of electronic warfare systems at the U.S. Naval Air Systems Command, Arlington, VA. Afterwards, he was a Chief Scientist at Hughes Aircraft Company, El Segundo, California (1973–1993) designing air to ground radar systems for U-2, F-15E and B-2 aircrafts. Since 1993, he has been a Research Associate at UCLA performing research in signal processing in acoustics and seismic sensor arrays and spatial diversity for wireless data transmission.
Charles E. Taylor received his B.A. degree in Genetics from the University of California, Berkeley, and his Ph.D. degree in Ecology and Evolution from the State University of New York, Stony Brook, in 1968 and 1972, respectively. From 1975 to 1980, he was at the University of California, Riverside, as Assistant Professor. He is currently a Full Professor in the Department of Ecology and Evolutionary Biology, the University of California, Los Angeles. Charles Taylor is the author of over 150 technical publications. His research interests include artificial life, sensor arrays and bird communication.

Kung Yao received the B.S. (Summa Cum Laude) and Ph.D. degrees in E.E. from Princeton University. He was a NAS-NRC Post-Doctoral Fellow at the University of California, Berkeley. He has served as an Assistant Dean of the School of Engineering at UCLA. Presently, he is a Distinguished Professor in the E.E. Department at UCLA. His professional interests include wireless fading channel modeling, digital communication theory, acoustic beamforming, sensor array system, and simulation. He received the IEEE Signal Processing Society’s 1993 Award in VLSI Signal Processing, the 2008 IEEE Communications Society/Information Theory Society Joint Paper Award, and the 2012 JCN Journal Best Paper Award. He is a co-author of Detection and Estimation in Communication and Radar Systems”, Cambridge University Press, in 2013. He also has extensive practical system experiences in sensor networks, satellite-wireless communications, radar system, and systolic and microphone array designs. He is a Life Fellow of the IEEE.

Ming Bao received his B.E. degree from Wuhan University of Transportation Technology, Wuhan, China and his Ph.D. degree in signal and information processing from The Institute of Acoustics, Chinese Academy of Sciences (IACAS), Beijing, China in 1996 and 2008, respectively. From 1999 to 2013 he was a assistant professor and then an associate professor in IACAS. Since January 2014 he has been a full professor at IACAS, China. His main research areas are in pattern recognition, machine learning, intelligence acoustic signal processing, binaural interaction modeling and function sensing.