J Sign Process Syst (2009) 57:415–436
DOI 10.1007/s11265-008-0310-7

An Empirical Study of Collaborative Acoustic Source Localization

Andreas Mantik Ali · Shadnaz Asgari · Travis C. Collier · Michael Allen · Lewis Girod · Ralph E. Hudson · Kung Yao · Charles E. Taylor · Daniel T. Blumstein

Received: 15 November 2007 / Revised: 20 September 2008 / Accepted: 22 October 2008 / Published online: 26 November 2008

Abstract Field biologists use animal sounds to discover the presence of individuals and to study their behavior. Collecting bio-acoustic data has traditionally been a difficult and time-consuming process in which researchers use portable microphones to record sounds while taking notes of their own detailed observations. The recent development of new deployable acoustic sensor platforms presents opportunities to develop automated tools for bio-acoustic field research. In this work, we implement both two-dimensional (2D) and three-dimensional (3D) AML-based source localization algorithms. The 2D algorithm is used to localize marmot alarm-calls of marmots on the meadow ground. The 3D algorithm is used to localize the song of Acorn Woodpecker and Mexican Antthrush birds situated above the ground. We assess the performance of these techniques based on the results from four field experiments: two controlled test of direction-of-arrival (DOA) accuracy using a pre-recorded source signal for 2D and 3D analysis, an experiment to detect and localize actual animals in their habitat, with a comparison to ground truth gathered from human observations, and a controlled test of localization experiment using pre-recorded source to enable careful ground truth measurements. Although small arrays yield ambiguities from spatial aliasing of high frequency signals, we show that these ambiguities are readily eliminated by proper bearing crossings of the DOAs from several arrays. These results show that the AML source localization algorithm can be used to localize actual animals in their natural habitat using a platform that is practical to deploy.

Keywords Bioacoustics · Distributed signal processing · Source localization · Wireless sensor networks · Maximum likelihood

A. M. Ali (✉) · S. Asgari · T. C. Collier · R. E. Hudson · K. Yao · C. E. Taylor · D. T. Blumstein
University of California-Los Angeles, Los Angeles, USA
e-mail: amantik@ucla.edu

M. Allen
Coventry University, Coventry, UK

L. Girod
Massachusetts Institute of Technology, Cambridge, USA

1 Motivation

Field biologists use the vocalizations of animals to identify individuals, census species and to study the dynamics of acoustic communication [1, 2]. However, even experienced field biologists have difficulty accurately identifying and locating species acoustically, and most researchers are unable to identify more than a few distinctive individuals. Some acoustic phenomena such as alarm calling (where individuals produce specific vocalizations in response to predators [3]) are relatively rare thus difficult to study, while others, such as duetting (where two individuals interdigitate their vocalizations [4]) are extremely difficult to properly describe. Thus, field research of natural populations will benefit from the use of embedded sensor arrays that are constantly alert, and that are able to detect acoustic events, localize the sound's source, and identify the individual or species producing the sound.

Alarm calls form an ideal system for motivating and testing our technology because they are infrequent,

they are loud, and they are biologically important. The yellow-bellied marmots at the Rocky Mountain Biological Laboratory (RMBL), in Gothic, Colorado, have become a model system for studying alarm communication. Marmots communicate risk by emitting a simple single note alarm call and emit more calls and at a higher rate as risk increases [5]. However, the modal number of alarm calls produced is one, and it is remarkably difficult to identify the individual who produced the call (we are able to localize and identify only about 30% of callers—Blumstein, unpublished data). Calls are individually-specific and the adaptive utility of this individuality has been the focus of considerable study. We know that calls contain information about the age, sex and exact identity of the caller [6], and we know that marmots are able to discriminate individuals based solely on their calls [7].

Tools to remotely sense, record and automatically analyze acoustical bird sounds would be enormously helpful for studies in ecology, biodiversity and behavior. Our work with birds has focused on two paradigms that represent the problems involved: (1) Mexican Antbirds in tropical rainforests with very complex acoustical environments. To the human listener, the unfiltered sounds, many quite similar even between taxonomic groups, combine into a noisy and confusing cacophony, making it difficult to isolate specific acoustic signals. While general studies of rainforest acoustics have been reported, Embedded network sensing boxes (ENSBox) can be deployed to more thoroughly study the rainforest environment including species identities, spatial and temporal patterns, and joint utilization of frequency and patterning ranges by the acoustic community [8, 9]. (2) Acorn Woodpeckers (*Melanerpes formicivorus*) are a common, conspicuous and very chatty birds throughout the foothill and woodlands in much of California and Southwestern US down to Mexico and extending into Colombia. The species has received special attention from biologists because of two unique properties. One is its foraging habit. In the fall, woodpeckers store food, mainly acorns, in the holes drilled on a specialized storage trees called granaries. The other is its extreme sociality. In California, Acorn Woodpeckers live in family groups of up to 15 individuals of both sexes and all ages. Each family group shares and defends an all-purpose territory, inside of which there are one or more granaries for acorn storage. Groups roost communally and all breeding females of the group lay eggs in a single nest. The mating system of Acorn Woodpeckers varies between monogamy and polygynandry.

The sorts of questions that we would want to address with the Acorn Woodpeckers are much more detailed than those of the complex Mexican Antbirds, in large part because of the developed infrastructure and because of the detailed knowledge already accumulated about these populations. At Hastings Biological Field Station, Monterey, CA, we would like to address questions such as: (1) whether vocal traditions exist in this cooperatively breeding species; (2) whether individual variation in calls is more closely allied with kin or group associations, and whether individual change vocalizations when they move from one group to another; (3) whether there are sex differences in vocalizations that match the matrilineal societies often found in this species; and (4) the extent to which birds use vocalizations to discriminate among group and non-group members and among kin and non-kin.

Using the Acoustic ENSBox system, a multi-node distributed recording array, we evaluated the ability of an Approximate Maximum Likelihood (AML) source localization algorithm to correctly identify the location of naturally alarm calling marmots as well as recorded and re-broadcast alarm calls. Field tests allowed us to comprehensively evaluate all the features (node time synchronization, self-localization, event detection, and AML-based DOA bearing estimation) of the Acoustic ENSBox. Tests under field conditions are essential because animals move their heads while vocalizing, and because there is often substantial background noise through which the signals must be detected.

The main contributions of this paper are: (1) the implementation of a deployable on-line marmot call detection algorithm, (2) a centralized marmot call localization system based on AML bearing estimation, (3) an analysis of complexity and performance of the 3D AML algorithm for source localization, and (4) a thorough evaluation of the effectiveness of these algorithms based on a field study detecting real animals in their natural habitat.

2 Overview of Approach

Distributed source localization is a broad and active research area, and a diverse set of solutions have been proposed. These solutions fall into three categories, in which the localization solution is based on (1) differential signal amplitudes, (2) time-difference-of-arrivals (TDOA), and (3) comparison of direction-of-arrival (DOA) estimates. In general, characteristics of the application, the source signal, and the environment will determine which of these solutions performs best.

The characteristics of the environment and the nature of the source signals rule out some of these solutions. The first alternative, amplitude-based local-

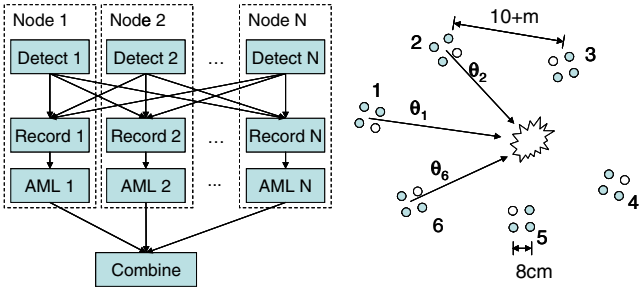


Figure 1 Block diagram of a DOA-based localization system. The AML algorithm was run centrally in the experimental results.

ization, is ruled out by foliage and terrain complexity, which yields non-isotropic signal attenuation patterns. Without discovering the complex model of the signal attenuation, received amplitude values are difficult to map to propagated distances.

The second alternative, TDOA-based localization, requires precise acquisition of the phases of the signals arriving at different nodes. However, this approach is not Maximum Likelihood. Our approach has therefore focused primarily on the third alternative, in which the location estimate is computed by combining DOA estimates assessed at a distributed set of locations. Our implementation employs a distributed set of small "sub-arrays", each capable of independently detecting the target signal and producing a DOA bearing estimate. The crossing of these bearing estimates are then combined to produce an estimate of the most likely source location. In the next subsections, we give a brief overview of this implementation, and highlight some key features of the platform. A detailed discussion of the processing algorithms follows in Section 3.

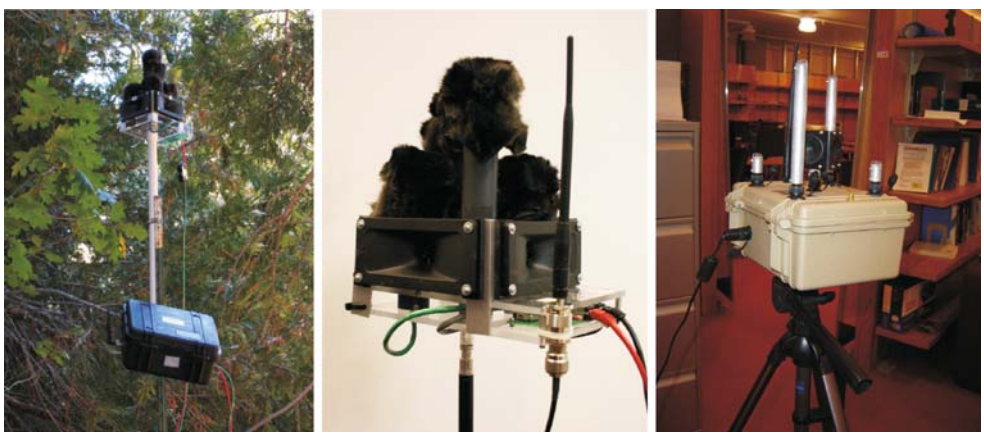
2.1 DOA-based Localization Using Sub-arrays

Figure 1 shows a high level diagram of a DOA-based localization system. To apply this method, we deploy a collection of sub-arrays surrounding a target of interest. The sub-arrays are typically deployed over a wide area relative to the size of each sub-array. In this paper, we use two different versions of the Acoustic ENSBox platform [10] shown in Fig. 2 and described in more detail in Section 2.3. The first version is used in RMBL, where each node hosts an 11.31 cm tetrahedral microphone sub-array rotated to form an 8 cm square when viewed from above. The second version is also a tetrahedral with 12 cm square when viewed from above. These sub-arrays are typically deployed at least 10 m apart, and often much farther: in the three sets of experiments presented in this paper, six sub-arrays are deployed surrounding a 70×140 m area (see Fig. 8). This large inter-node spacing means that any target source can be assumed to be in the far field of all but perhaps one of the nodes.

After deployment, the sub-arrays are automatically calibrated to determine the relative positions and orientations of the sub-arrays in the system. Next, software on the nodes begins implementing the detection and localization algorithms.

The detection software on each node performs a streaming analysis of the acoustic data in real time, identifying likely animal call events. Whenever any individual node's call detector is triggered, a radio message is sent to trigger all the nodes in the system to start recording that event and queue it for further processing. This approach enables optimization of the detection threshold such that only the nearest node to a source needs to be triggered.

Figure 2 The acoustic ENSBox platform.



(a) ENSBox Ver. 1. Left: Complete set at the field. Right: the microphone array.

(b) ENSBox Ver.2.

Once identified, segments of audio containing calls are analyzed using the (AML) algorithm described in Section 3.2. Based on the relative phases of signals recorded at the microphones in a given sub-array, this algorithm determines a likelihood metric describing the likely bearing to the source. These metrics are then collected centrally, placed on a map according to the location and orientation of each sub-array, and combined into a 2D or 3D pseudo-likelihood map of the source location. This map is formed by projecting each likelihood metric outwards from each node to form the joint approximate likelihood of a source at every point in the 2D or 3D space.

2.2 Performance Impact of Sub-array Size

The performance of the AML bearing estimation algorithm depends on characteristics of the source signal, and on the size and geometry of the array. The acoustic sources produced by different animals can vary significantly. In general, we consider these signals as wideband because the frequency ratio of the highest to the lowest is much larger than one. However, when a source may contain only few closely-spaced dominant frequencies, then it may behave more like a narrowband signal. This may presents a problem depending on the selection of the spacings among the sensors in a sub-array. When a narrowband source is present, there is a risk that the algorithm may return ambiguous likelihood metric results. Just as the Nyquist theorem states that to avoid aliasing, a signal must be sampled at least at two times the maximum frequency of the signal, an analogous property holds for *spatial* sampling. In order to measure the phase of an incoming signal by comparing two points in *space*, those two points must lie in the same half-wave. Energy in frequencies with wavelengths shorter than two times the sensor spacing will be aliased into lower frequencies. This implies that for two sensors with spacing D and signal propagation speed V_s, the maximum frequency detectable without aliasing is $F_c = V_s/(2D)$. For more than two sensors with maximum spacing D, the aliasing can theoretically disappear, but the sensitivity for some array configuration is quite high, thus a little noise can create ambiguous response.

The likelihood metric of the bearing estimate is typically represented as a polar plot for a 2D case and as a surface for a 3D case, where the likelihood value is plotted as a function of the bearing angle. In such a plot, the most likely bearing estimate is represented by the midpoint of the largest lobe of this metric. Array size has a two-fold impact on these results. As the array size increases, spatial aliasing can become a problem. Whenever the frequency content of the source signal is higher than the critical frequency F_c, spatial aliasing will produce *grating lobes*, (i.e., false lobes) that point in directions other than the true source bearing [11]. These grating lobes often have heights comparable to the true main lobe due to array geometry physical limitations. In the presence of noise, reverberation, or competing sources, grating lobes can severely complicate identification of the true DOA. However, as array size decreases, the width of the main lobe also increases. This tradeoff is depicted in Fig. 3 for a 2D case, which shows simulated beam patterns for two different size arrays detecting a 4 kHz source; the larger array has narrower main lobes, but more potential ambiguity.

In the analysis of a single array, there is usually a "sweet spot" for array size where the maximum value of the side lobes will be less than a certain fraction of the main lobe and thus can be excluded. However, for high frequency sources (e.g., 6 kHz and above for our implemented array), the array size sweet spot becomes quite small, producing a wide main lobe. Upon perturbation of an additive noise, the main lobe width directly impacts the variance of the DOA estimate, which often is investigated via a small perturbation lower bound such as the Cramér-Rao Bound (CRB). To avoid this problem, we ensure the main lobe width by using relatively large arrays and address the ambiguity problem by other means. When bearing data from multiple arrays is combined, some of the false lobe DOAs are rejected because they are inconsistent with the lobes from other nodes. While the "beam" projected from a false lobe has some chance of intersecting with other lobes, the DOAs of the true lobes from all the sub-arrays have a much higher likelihood of crossing near the true location of the source.

This approach has its limits; as the number of side lobes increases, this approach will eventually fail, hence

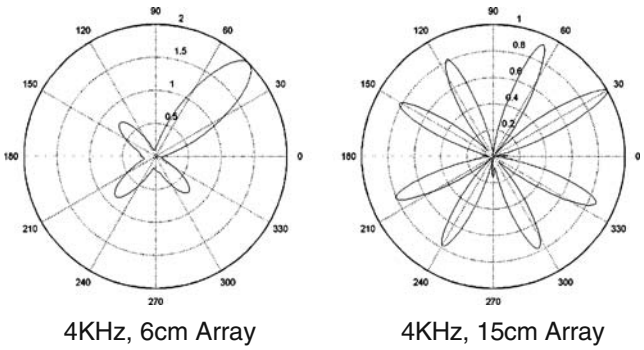


Figure 3 Simulated 2D beam patterns for two arrays detecting marmot alarm calls. Larger arrays yield narrower lobes, but more ambiguity.

the array size cannot be made arbitrarily large. For our purposes, we chose a convenient array size from a practical engineering implementation point of view, suitable for a class of sources of interest. These arrays are hosted by a deployable, general purpose sensing platform described in the next section.

2.3 Implementation of the Sub-array Nodes

The deployment described in this paper comprised 6 nodes, each an independent wireless processor hosting a sub-array. In undertaking this work, we were fortunate to be able to build upon an existing platform, the Acoustic ENSBox [10]. The ENSBox was specifically designed to support this type of application, and it has numerous features that make this type of deployment practical for the first time.

Packaging While packaging issues are far from novel, they are quite important in practice. In prior attempts to record using multiple arrays, our equipment was assembled with "off-the-shelf" components, dozens of microphone XLR cables and many batteries to support devices that were not tuned for low power consumption. These solutions are cumbersome. In contrast, the ENSBox is a wireless distributed sensor system. Each unit is a self-contained processor and array, has an internal battery with a lifetime of 5 h, is water-resistant, and has an array head that can be fitted to a tripod. The second generation of this ENSBox is smaller, lighter, wire-free, and easier to deploy. This box also has a GPS receiver and an accelerometer for tilt angles measurement.

Management As the size of a deployment grows, management rapidly becomes a critical concern. The larger the number of nodes deployed, the greater the likelihood that one or more nodes is faulty—and this is especially true for prototype systems. To facilitate deployment, the ENSBox supports a web-based management tool hosted on each node. By connecting to any one of the nodes and setting it into "master mode", the user can use that node as a gateway to centrally manage the rest of the network. Diagnostics available through this interface can identify problems with individual nodes and thus ensure that all nodes are functioning properly. Once the system is up, the web interface is also used to initiate and manage the application.

Self-configuration Self-configuration is another important factor in deployments. The ENSBox system features a self-configuring multi-hop wireless network, with network diagnostics available from the management gateway. It also features a sophisticated array self-calibration system that can establish precise positions and orientations for all of the arrays in the system. This system, described in detail in [10, 12], can compute relative array positions to within 10 cm over an 50×80 m field, and estimate array orientation to within 1.5°. Extensive testing has proved that the system is easy to operate, achieves detection ranges upwards of 100 m, is robust to noise and intervening foliage, and provides a consistency metric that immediately indicates whether the results are likely to be valid. By attending to the consistency metric and performing simple sanity-checks, we have yet to fail to get accurate self-localization results from a field deployment. This feature of the ENSBox is an enormous time-saver because it gives reliable results with low effort, and eliminates the need to carefully survey the deployment positions.

Software The final advantage of the ENSBox has been as an application development platform. The ENSBox provides a synchronized sampling framework that greatly simplifies the development of collaborative sensing application software [13]. The detector application described in this paper is 800 lines of C code and took about 1 week to develop within the EmStar software framework [14]; it detects marmots and triggers synchronized processing on all nodes in the system. Because the system is built on a 32-bit Linux processor, it has the additional resources to support rapid prototyping and minimize early optimization. In the next section, we describe some algorithms we have implemented and tested using this hardware.

3 Algorithms

The algorithms we describe herein are not wholly novel; in fact, the basic algorithms have all been introduced in prior work. Rather, the novelty in this work lies in the evaluation of these algorithms in the context of a real deployment and a real scientific application, and in the implementation details involved in tuning the algorithms for this application. In this section we will discuss the details of our marmot call detection algorithm and the AML bearing estimation and localization algorithm.

3.1 Marmot Event Detector

The marmot detection algorithm is an on-line statistical classifier for identifying the bursts of energy in a streaming signal in a frequency range of interest [15].

Assuming the noise distribution is Gaussian $N(\mu, \sigma^2)$, we can use a smoothing filter to compute on-line estimates $\bar{\mu}$ and $\bar{\sigma}$. From these estimates, we define a threshold value $\bar{\mu} + \beta\bar{\sigma}$, or β standard deviations above the mean energy value. Thus, if our noise model holds, any energy value exceeding this threshold is either part of the signal, or is noise with probability $1 - \text{erf}(\beta/\sqrt{2})$, which diminishes rapidly with β.

In practice, not all of the noise we would like to filter out is Gaussian. While the Gaussian distribution *is* a good model for ambient environmental noise, the noise caused by other animals and events in the environment will not fit that model. We counter this issue in two ways.

First, we apply a band pass filter that selects out only the frequency range used by our target signal (in this case, marmot alarm calls), and compute the energy metric over this band. Second, by adjusting the parameters of the smoothing filters, we select the adaptation rate for the noise estimator such that the model will adapt to signals that have a less abrupt onset than our target signal.

Figure 4 shows a block diagram of our marmot detection algorithm. The first three stages of the data flow implement a windowing operation and a decimation filter, resulting in windows of 32 points sampled at 24 kHz. To reduce the processing load, we only process only every fourth 64 point input window. Because marmot calls are approximately 0.04 s long, even skipping 3/4 of the windows we are still guaranteed to sample the marmot call. The next two stages sum the energy over the band of interest, by computing the Discrete Fourier Transform (DFT) and taking the magnitude of the sum of the frequency bins corresponding to the range 2.25–3.75 kHz. Next, this energy metric is fed into the marmot detector algorithm.

The energy metric feeds into the noise estimator, gated by whether the algorithm is currently "triggering". Whenever the detector triggers, new samples should not be added to the noise estimator until the signal of interest has passed—otherwise, the noise estimator would tend to adapt to the signal. In addition, there is typically a period of reverberation after an alarm call, during which the signal levels are higher than normal, but still below threshold. Thus, we define a hysteresis period such that the detector will remain in the triggered state for K_{min} samples after the last above-threshold sample. The sample ranges corresponding to periods of triggering are reported as the output of the detector.

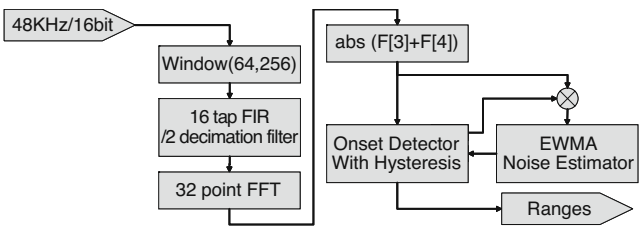


Figure 4 Block diagram of a marmot detection algorithm. This implementation is based on a similar algorithm for detecting bird calls.

The noise estimator itself is based on two *Exponentially Weighted Moving Average (EWMA)* smoothing filters, one estimating the mean μ, and the other estimating the variance σ^2. An EWMA is a simple feedback function that implements a smoothing function with very low computational complexity. The update function for a EWMA estimate $\bar{x}$ of x is

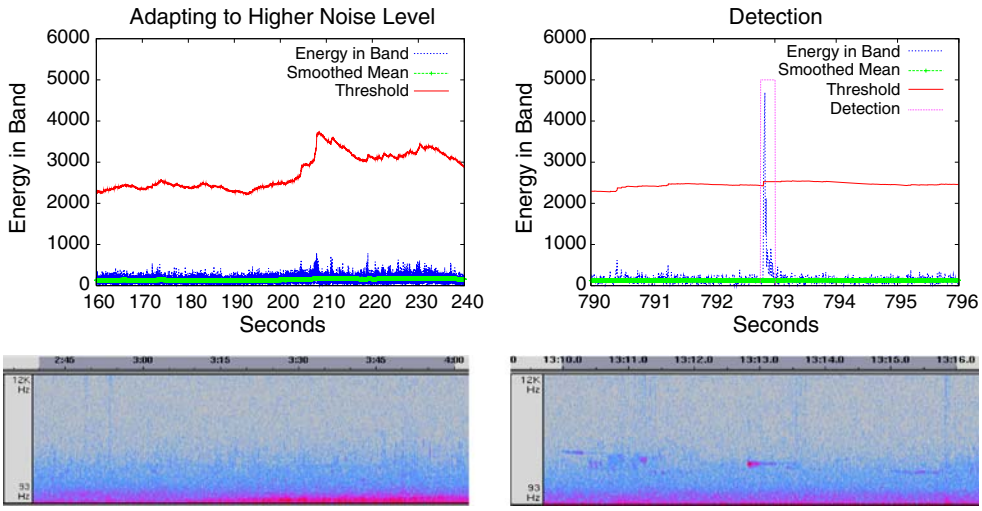
$$\bar{x}_{t+1} = \alpha x_t + (1 - \alpha)\bar{x}_t,$$

where α is the adaptation rate parameter. The detection threshold is then computed from $\beta\bar{\sigma}$; any sample above threshold is considered a detection.

While this covers the "streaming" case, there are two additional details of the algorithm: initialization and lockup detection. When the detector starts, there is no initial noise estimate, so the threshold cannot reliably be determined. Thus, we implement an initialization phase in which triggering is withheld for the first K_{init} samples while the noise estimator builds a model. The second detail is "lockup", a condition that can occur in the event of a sudden but permanent change in the noise level. If a permanent change in the noise level causes the detector to trigger, the detector will never un-trigger, because the noise estimates are not updated while triggering. To address this, we apply a heuristic that re-initializes the detector whenever triggering lasts for more than K_{max} samples.

Figure 5 shows the behavior of our on-line marmot call detector in two cases. Figure 5a represents a period of time in which the ambient noise level is gradually increasing, and the mean and threshold adapt to this change. Figure 5b shows the case where a marmot call exists and is detected. Note that other animal calls are present in the data, but are attenuated by the band pass filter and easily rejected by the detector. From our previous experience detecting birds, we expected to set a much lower β parameter; however, after analyzing initial recordings we found that the marmots were *surprisingly* loud—and therefore readily detected using a high value for β. We expect to continue to gain more experience with this algorithm in future deployments. We anticipate that there are a range of animal detection applications for which it is sufficient to simply select

Figure 5 Behavior of the marmot alarm call detector.



(a) Adaptation to an increase in noise level.

(b) Detection of a marmot call; rejection of other sounds.

different parameter values. The parameter values used in our implementation are given in Table 1.

3.2 AML Generalized to 3D

Approximate Maximum Likelihood (AML) is a likelihood-based algorithm that searches the event space for the most likely feature of the event [16]. When the source is in the far-field of the sub-array, the event of interest reduces to a DOA bearing estimation because the range becomes unobservable.

Suppose that we use the Cartesian coordinate system to express the position of a point $A_{\mathrm{RECT}} = [x, y, z]^T$ relative to the origin of the coordinate system O. According to Fig. 6, then position of A_{POL} in the spherical coordinate system can be expressed as $A = [R, \theta, E]$, where

$$R = \sqrt{x^2 + y^2 + z^2}, \quad \tan(\theta) = \frac{y}{x}, \quad \tan(E) = \frac{z}{\sqrt{x^2 + y^2}}. \tag{1}$$

Table 1 Parameter values.

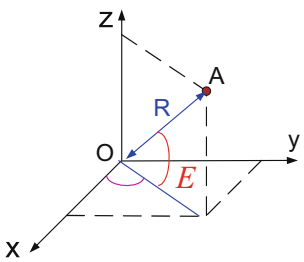
Parameter	Value
F_s	24,000 Hz
FFT points	32
Window feed	128
Frequency bins	3 or 4
α	0.999
β	32
K_{init}	300
K_{min}	40
K_{max}	120

Now, assume that the sensor array comprises P arbitrarily distributed, omni-directional sensors with identical behavior. Also assume that the array is centered at the origin. Each sensor is located at the position $r_p = [x_p, y_p, z_p]^T$, with $1 \le p \le P$. Similarly, assume that there are M wideband sources in the far-field of the array and at unknown locations $q_m = [x_m, y_m, z_m]^T$ with $1 \le m \le M$. Then the relative time delay of the mth source is given by

$$\begin{aligned} t_{cp}^{(m)} &= t_c^{(m)} - t_p^{(m)} \\ &= \frac{1}{v}\left[\left(x_p \cos\psi_m + y_p \sin\psi_m\right)\cos e_m + z_p \sin e_m\right] \\ &= \frac{R_p}{v}\left[\cos\left(\psi_m - \theta_p\right)\cos\left(e_m\right)\cos\left(E_p\right)\right. \\ &\qquad \left. + \sin\left(e_m\right)\sin\left(E_p\right)\right]. \end{aligned} \tag{2}$$

Note that $[R_p, \theta_p, E_p]^T$ is the position of the pth sensor and $[d_m, \psi_m, e_m]^T$ is the position of the mth source, both in the spherical coordinate system. $t_c^{(m)}$ and $t_p^{(m)}$ are the absolute time delays from the mth source to the centroid and to the pth sensor respectively, and v is the speed of sound (nominally 345 m/s at room temperature). The data received by the pth sensor at

Figure 6 Point A in the spherical coordinate system.



time $t = t_n$, henceforth, simply denoted by the index n, is given by $x_p(n) = \sum_{m=1}^{M} s^{(m)}(n - t_{cp}^{(m)}) + \omega_p(n)$, $n = 0, \ldots, N-1$ where N is the length of the data vector, $s^{(m)}(n)$ is the mth source signal at the array centroid, and ω_p is the zero mean white Gaussian noise with variance σ^2. Note that in the above equation, $t_{cp}^{(m)}$ can be any real-valued number.

The received wideband signal can be transformed into the frequency domain via the Discrete Fourier Transform (DFT), where a narrowband model can be attributed to each frequency bin. It is well-known that the circular shift property of the DFT has an edge effect problem for the actual linear time shift. These finite effects become negligible for a sufficient long data. Here, we assume the data length N is large enough to ignore the artifact caused by the finite data length. For the N point DFT transformation, the array data model in the frequency domain is given by $\mathbf{X}(\omega_k) = \mathbf{D}(\omega_k)\mathbf{S}(\omega_k) + \eta(\omega_k)$, $k = 0, \ldots, N-1$, where $\mathbf{X}(\omega_k) = [X_1(\omega_k), \ldots, X_P(\omega_k)]^T$ is the array data spectrum, and $\eta(\omega_k)$ is the complex Gaussian random vector of independent-identically-distributed (i.i.d) components, each with zero mean and variance $N\sigma^2$. Note that due to the transformation to the frequency domain, $\eta(\omega_k)$ asymptotically approaches a Gaussian distribution by the Central Limit Theorem, even if the actual noise has an arbitrary i.i.d distribution with bounded variance in the time domain. This asymptotic property in the frequency domain provides a more reliable noise model than the time domain model in some practical cases. The other terms on the right-hand side of the last equation are steering matrix $\mathbf{D}_{P\times M}(\omega_k) = [\mathbf{d}^{(1)}(\omega_k), \ldots, \mathbf{d}^{(M)}(\omega_k)]$ and the source spectrum $\mathbf{S}(\omega_k) = [S^{(1)}(\omega_k), \ldots, S^{(m)}(\omega_k)]^T$. Note that $\mathbf{d}^{(m)}(\omega_k) = [e^{-j2\pi k t_{c1}^{(m)}/N}, \ldots, e^{-j2\pi k t_{cp}^{(m)}/N}]^T$ is the steering vector. Throughout this manuscript, we denote the transposition and complex conjugate transposition operations with superscripts ^{T} and ^{H} respectively.

The AML estimator performs the data processing in the frequency domain. We define $Q(w_k) = D(w_k)S(w_k)$. Since zero frequency bin is not important and the negative frequency bins are merely mirror images, we can use only the positive $N/2$ frequency bins. Stacking them into a single column, we can rewrite the sensor data into an $(NP/2) \times 1$ space-temporal frequency vector as

$$X = G(\Theta) + \xi,\ G(\Theta) = \left[Q^T(w_k), \ldots, Q^T\left(w_{N/2}\right)\right]^T,$$
$$R_\xi = E\left[\xi\xi^H\right] = \left(N\sigma^2\right) I_{NP/2}. \tag{3}$$

We assume, initially, that the unknown parameter space is $\Theta = [\tilde{r}_s^T, S_0^{(1)^T}, \ldots, S_0^{(M)^T}]^T$, where the M source locations are denoted by $\tilde{r}_s = [r_{s_1}^T, \ldots, r_{s_M}^T]$ and the mth source signal spectrum is denoted by $S_0^{(m)} = [S^{(m)}(w_1), \ldots, S^{(m)}(w_{N/2})]^T$. So the log-likelihood function of the complex Gaussian noise vector ξ, after ignoring irrelevant constant terms is given by

$$\max_{\Theta} L(\Theta) = \max_{\Theta}\left(-||X - G(\Theta)||^2\right)$$
$$= \min_{\Theta} \sum_{k=1}^{N/2} ||X(w_k) - D(w_k)S(w_k)||^2. \tag{4}$$

This is equivalent to finding $\min_{(\tilde{r}_s, S(w_k))} f(w_k)$ for each frequency bin w_k, $k \in 1, \ldots, \frac{N}{2}$, where $f(w_k) = ||X(w_k) - D(w_k)S(w_k)||^2$. The minimal of $f(w_k)$ with respect to the source signal vector $S(w_k)$ must satisfy $\partial f(w_k)/\partial S^H(w_k) = 0$. Hence, the estimate of the source signal vector that yields the minimum residual at any source location is given by $\hat{S}(w_k) = D^\dagger(w_k)X(w_k)$, where $D^\dagger(w_k) = (D^H(w_k)D(w_k))^{-1}D^H(w_k)$ is the pseudo inverse of the steering matrix $D(w_k)$. Next, we define the orthogonal projection $P(w_k, \tilde{r}_s) = D(w_k)D^\dagger(w_k)$, so the complement orthogonal projection becomes $P^\perp(w_k, \tilde{r}_s) = I - P(w_k, \tilde{r}_s)$, where I is the identity matrix. By combining the last two equations, the minimization function becomes $f(w_k) = ||P^\perp(w_k, \tilde{r}_s)X(w_k)||^2$. After substituting the estimate of the $\hat{S}(w_k)$, the AML source location estimate can be obtained by finding

$$\max_{\tilde{r}_s} J(\tilde{r}_s) = \min_{\tilde{r}_s} \sum_{k=1}^{N/2} ||P^\perp(w_k, \tilde{r}_s)X(w_k)||^2$$
$$= \max_{\tilde{r}_s} \sum_{k=1}^{N/2} tr\left(P(w_k, \tilde{r}_s)R(w_k)\right). \tag{5}$$

Note that $R(w_k) = X(w_k)X^H(w_k)$ is the one snapshot covariance matrix. This multi-source AML algorithm performs signal separation by utilizing the physical separation of the sources, and for each source signal, the Signal-to-Noise-Ratio (SNR) is maximized in the ML sense. In general, no closed form solution can be obtained here.

3.2.1 Single Source Scenario

Suppose we have an array consisting of P sensors, where the array centroid is the reference point of the spherical coordinate system. It can be shown that if we have only one source represented with the vector $[d, \psi, e]^T$, then the maximum likelihood criterion in Eq. 5 can be simplified as $\max_{\tilde{r}_s} J(\tilde{r}_s) = \max_{\tilde{r}_s} \sum_{k=1}^{N/2} (X^H(w_k)D(w_k))^2$, where for the single source case, $D_{P\times 1}(w_k) = [e^{-j2\pi k t_{c1}^{(m)}/N}, \ldots, e^{-j2\pi k t_{cp}^{(m)}/N}]^T$ and the relative time delay between the centroid

and the pth sensor would be $t_{cp} = \frac{R_p}{v}[cos(\psi - \theta_p)\cos(e)\cos(E_p) + \sin(e_m)\sin(E_p)]$. This form of $J(\tilde{r}_s)$ makes the computational cost less than the original ML criterion form in Eq. 5. Even so, a fine grain grid search for the 3D AML algorithm is still computationally intensive. To obtain the estimated azimuth and elevation angle using the AML algorithm via a grid search, we have to do a search on all possible azimuth angles (from 0° to 360°) and all possible elevation angles (from −90° to 90°) to find the maximum $J(\tilde{r}_s)$.

3.2.2 Cramér-Rao Bound for 3D DOA

The CRB is most often used as a theoretical lower bound for any unbiased estimator. In this section, we derive the CRB directly from the signal model. We can construct the FIM (Fisher Information Matrix) [16, 17] from the signal model defined by $F = 2\Re[H^H R_\psi^{-1} H] = \frac{2}{N\sigma^2}\Re[H^H H]$. Note that $H = \frac{\partial G}{\partial \phi^T}$ and $\phi = [\psi, e]^T$. G and R_ψ are already defined in Eq. 3. For the single source, vector G can be simplified as

$$G_{(NP)/2\times 1} = \left[e^{-j2\pi t_{c1}/N} S_1(w_1), \ldots, e^{-j2\pi t_{cP}/N} S_1(w_1), \ldots, e^{(N/2)(-j2\pi t_{c1}/N)} S_1(w_{N/2}), \ldots, e^{(N/2)(-j2\pi t_{cP}/N)} S_1(w_{N/2})\right]^T \quad (6)$$

In this case,

$$F = \alpha T, \quad (7)$$

$$\alpha_{1\times 1} = \frac{2}{N\sigma^2}\sum_{k=1}^{N/2}\left(\frac{2\pi k|S_1(w_k)|}{N}\right)^2, \quad (8)$$

$$T = \sum_{p=1}^{P}\begin{bmatrix} \left(\frac{\partial t_{cp}}{\partial \psi}\right)^2 & \frac{\partial t_{cp}}{\partial \psi}\times\frac{\partial t_{cp}}{\partial e} \\ \frac{\partial t_{cp}}{\partial e}\times\frac{\partial t_{cp}}{\partial \psi} & \left(\frac{\partial t_{cp}}{\partial e}\right)^2 \end{bmatrix}. \quad (9)$$

As an example to see the performance of the proposed 3D AML algorithm relative to the CRB, suppose that we have a cube with a side length of 3 cm which is centered at the origin. This cube has six side planes and one sensor is placed in the middle of each side plane. This 3D array with six sensors is shown on Fig. 7a. Then the matrix ς whose columns show the position of the corresponding sensors in the array would be

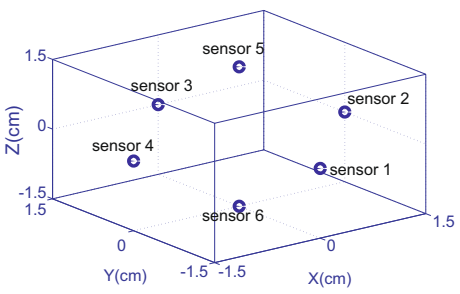
$$\varsigma = \begin{bmatrix} 0 & 1.5 & 0 & -1.5 & 0 & 0 \\ -1.5 & 0 & 1.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.5 & -1.5 \end{bmatrix}. \quad (10)$$

Assume our source signal is a male Dusky Antbird call which is located at an azimuth angle of 155° and an elevation angle of 60°. Figure 7b shows the mean square error (MSE) in degree square obtained by running the 3D AML algorithm for 1,000 different noise realizations and compares it to the corresponding CRB in both azimuth and elevation angles. As seen from this figure, the 3D AML performance is very close to the CRB and confirms the efficiency of the AML estimator algorithm.

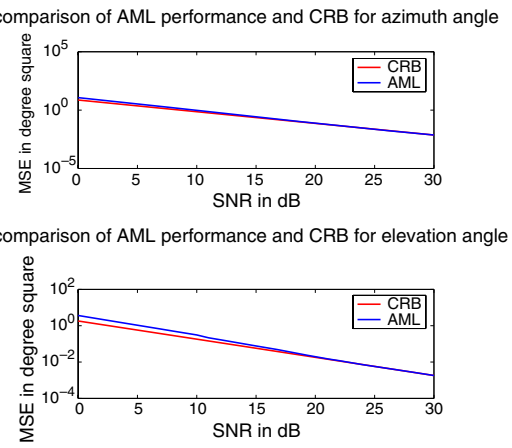
To reduce the 3D AML algorithm complexity, we use the concept of isotropic and non-isotropic arrays. An array is said to be isotropic if it has a constant mean square angular error (MSAE) for all azimuth and all elevation angles. Baysal and Moses [18] have proved the following theorem which gives the necessary and sufficient conditions for an array to be isotropic.

Theorem 1 *Suppose that an P element 3D array which is centered at the origin, is represented by the array geometry matrix B as*

$$B_{3\times 3} = \sum_{p=1}^{P} l_p l_p^T, l_p = r_p \times \frac{f_s}{v}, \quad (11)$$



(a) Cubic array with 6 sensors.



(b) The AML performance and its CRB.

Figure 7 The cubic array configuration and the 3D AML algorithm performance using the corresponding array with source azimuth at 155° and elevation at 60°.

where l_p is the normalized location of the pth sensor in the Cartesian coordinate system and f_s is the sampling frequency. Then the array is isotropic, if and only if $B = kI_{3\times 3}$ where I is the identity matrix.

It can be shown that the FIM for an isotropic array is given by

$$F = \alpha \begin{bmatrix} k\cos^2(e) & 0 \\ 0 & k \end{bmatrix}, \tag{12}$$

where α is the scalar defined in Eq. 8. So the CRB for an isotropic array becomes $\mathrm{CRB}(\psi) = \frac{1}{\alpha k \cos^2 e}$, $\mathrm{CRB}(e) = \frac{1}{\alpha k}$. Now, if we remove the scaling of the CRB of the azimuth angle, then the CRB of the elevation angle and the CRB of the azimuth angle become equal and as the result, the MSAE becomes constant for all azimuth and elevation angles ($\mathrm{MSAE}(B) = \cos^2(e)\mathrm{CRB}(\psi) + \mathrm{CRB}(e) = 2/(\alpha k)$).

Since the Fisher information matrix for an isotropic array is diagonal, it is possible to decouple the azimuth and the elevation angle estimation [18]; hereby, reducing processing time significantly from a 2D search to several 1D searches. The search can employ alternating maximization technique, whereas one time we fix the elevation angle and find the azimuth angle, and on the other time we fix the azimuth angle and find the elevation angle. A complete cycle of searching the azimuth and the elevation is counted as one iteration. In each iteration, the reduction in the complexity is approximately $120/\lambda$ times less than the original 3D AML algorithm, where λ is the angular accuracy of our DOA estimation in degree. If $\lambda = 1^\circ$, then the complexity of the decoupled version of AML algorithm is approximately 1/120 of the original 3D AML algorithm. Note that the decoupling property of isotropic arrays with an arbitrary initial point can be used only if the isotropic array size is selected appropriately. If the isotropic array size is not selected appropriately then the ML criterion can have more than one peak where one of these peaks is the global maximum and the others are just local maximums. Then, converging to the global maximum is not always guaranteed unless we choose a very good initial point; one that is close to the global maximum. Low complexity algorithms for non-isotropic arrays exists, but since it is outside the focus of this paper, we will omit them. Interested reader can find more information about 3D array issues in [19].

A natural generalization for this algorithm is the multi-source localization and tracking. The simultaneous vocalization of multiple species of interest is a possible reality as well as moving target when vocalizing. Even though some of the issues has been investigated in [20–22] and their respective references on theoretical level, the complexity prevents them to approach close to real time implementation. Further research will be needed to optimize these algorithm for low power devices.

4 Experiments

In this section we will describe a series of experiments we performed July 15–20, 2006 at the Rocky Mountain Biological Laboratory (RMBL), in Gothic, CO, and a series of test we performed at UCLA. In RMBL, we deployed six nodes in several locations where marmots are normally present. One of these, "Marmot Meadow", is shown in Fig. 8. We tested the performance of our marmot detector and the AML-based source localization under realistic field conditions, detecting real animals as sources. We also performed a controlled test of the sensitivity of the AML estimator to distance and source orientation, using pre-recorded audio for 2D analysis in the same field environment. For 3D DOA accuracy test, we conducted the experiment at UCLA with pre-recorded bird calls at different elevated locations. Then we performed a set of controlled experiments on a grassy field at UCLA using pre-recorded marmot calls and a white noise as audio source. In this test, we deployed six nodes and the source at a predetermined location, and we place the source at two typical positions: (1) inside and (2) outside the array convex hull.

Figure 8 "Marmot Meadow" location at Rocky Mountain Biological Laboratory, in Gothic, Colorado. The node locations correspond to the wide deployment described in Section 4.5.

4.1 Marmot Detector Performance

To test the performance of the marmot detector, we ran the detection software on a network of five nodes. This software implemented the algorithm described in Section 3.1, running in real time. Whenever a detection range (the beginning and ending of the signal) was determined, that range was broadcast to all nodes in a packet. The ENSBox's integrated synchronized sampling API was used by this application to synchronously record segments of audio corresponding to the detection range on every node in the system. These snippets were then stored to flash for further processing.

After the test, the results were compared with field notes taken during the experiment. Figure 9 shows the result of this comparison. Several nodes detected every call present, and a few nodes reported false alarms. Node 106 was not functioning properly and only reported false alarms. As we saw in Fig. 5b, marmot calls are very loud and the detector had very little trouble identifying them. In fact, most of the false alarms were introduced by researchers walking through the field manipulating the nodes.

4.2 2D DOA Accuracy Testing

Real animals move their heads while vocalizing. To assess the consistency of DOA estimates from a single sub-array in the field when the source's direction changes, we conducted a series of playback tests. The source was a marmot call broadcast from a powered speaker (Advent 570 Powered Partner) in the same meadow site where live marmot experiments were conducted. This speaker reproduces the calls faithfully, though at a significantly lower volume than marmots naturally produce.

A single node was placed as it was for the live marmot experiments, with the sub-array raised approximately 1.5 m above ground level. The source was aligned by eye at a bearing of approximately 180° relative to the coordinate system of the sub-array. The source marmot call was repeated 15 times during each playback experiment.

The playback was repeated with the source at three different distances from the sub-array: 12.5, 25, and 50 m. We also rotated the direction the source speaker was facing to test any possible effect that may have, since marmots often turn their heads between or even during a call. Three speaker facings were done at each distance: pointing directly at the sub-array, pointing perpendicular to the sub-array, and pointing away from the sub-array.

We applied the AML algorithm to a 0.07 s segment of each marmot call, 135 trials in total. The algorithm was limited to a 1 kHz band of frequencies centered on 3 kHz, which is a typical range over which marmot calls have their maximum power. The results are

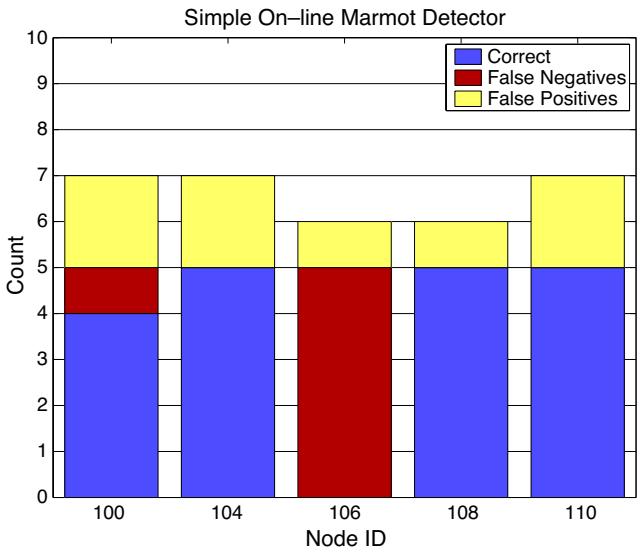


Figure 9 Performance of the marmot detection algorithm from 20 min of audio. The deployment laydown was similar to the compact deployment described in Section 4.4. For each node, the *bars* show the number of correct detections, missed detections, and false alarm. Note that if any node detects, all nodes will capture that call.

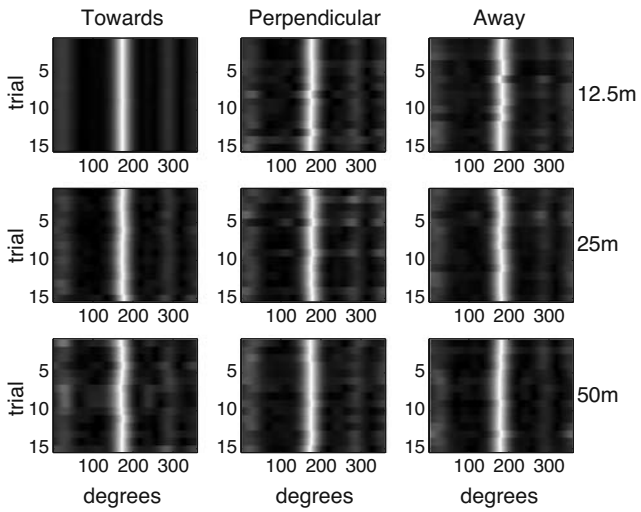


Figure 10 Results from a set of controlled DOA experiments in the field. A speaker was placed on the ground in a meadow at a set of specified distances (12.5, 25, and 50 m), and oriented to face either towards, perpendicular, or away from the array. Fifteen trials were performed in each configuration. The results above show the nine cases tested: each row of images is a different distance; each column a different facing of the source speaker. In each image, a row represents a separate trial, showing the likelihood metric as a function of angles.

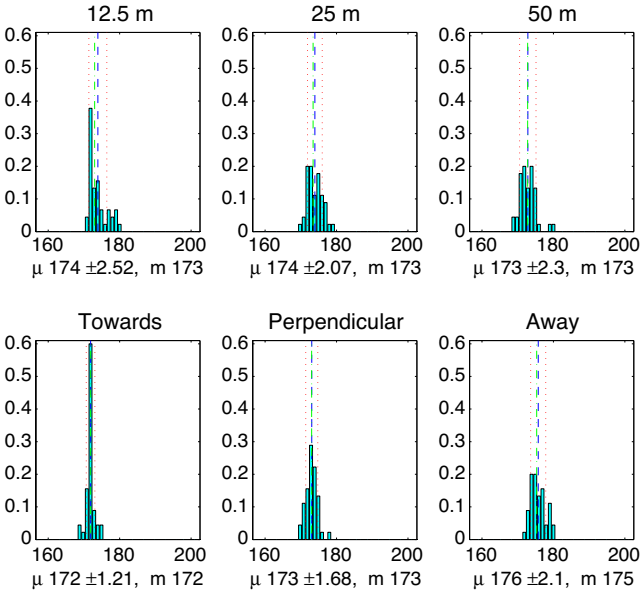


Figure 11 Histograms of the maximum DOA estimates for the same data shown in Fig. 10, grouped by distance and grouped by source facing. Each plot shows mean $\mu \pm$ the standard deviation and the median m.

shown in Fig. 10, where each subplot shows the AML estimates at each bearing for a different experimental condition. The side lobes common in all the figures are due to the array geometry effects explained in Section 2.2. The peculiar side lobes such as in trial 6 at 12.5 m are due to background noises, especially White-crowned Sparrows whose calls overlap marmot calls in frequency.

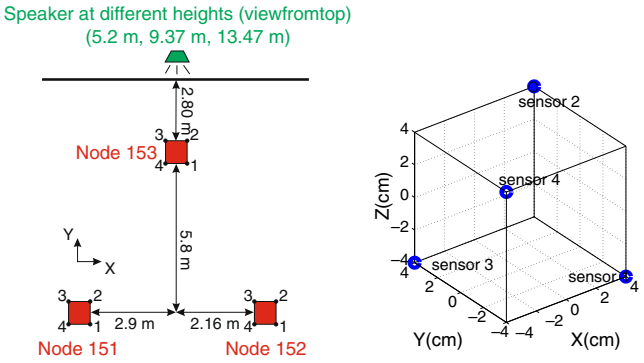
Figure 11 shows the statistical distribution of DOA estimates (the bearing with the maximum AML value for each trial). For all the data combined, the mean DOA estimate is 173.55 with a standard deviation of 2.35°. Distance from the source and the facing of the source have a significant effect on precision. The dominant effect appears to be the power of the signal arriving at the sub-array. As distance increases, the power of the signal drops. In addition, there is a large drop in power when the speaker faces perpendicular to or away from the sub-array. In our experimental setup, speaker direction caused a larger drop than distance (data not shown).

4.3 3D DOA Accuracy Testing

To test the effectiveness of the proposed 3D methods in a relatively non-open environment, we conducted various experiments in the central courtyard of UCLA Boelter Hall. The experimental location was surrounded by four sides of the tall quadrangle buildings, in addition to some trees and bushes. There was a persistent noise from ventilation systems of the surrounding buildings, and there was also a wind from the air flowing in between the buildings. The surrounding buildings causes reverberation that is further enhanced by the presence of trees. In order to manage these in the deployment, we place the three nodes within the source's line of sight and as far apart from each other as physically possible. The location of the experiment and the acoustic arrays are shown in Figs. 12 and 13.

We have three arrays denoted as node 151, 152 and 153. Each array consists of four microphones positioned at four corners of a cube with side length of 8 cm. The picture of one of the arrays and its isotropic configuration is shown in Fig. 13. The subarrays are placed at least 5 m apart. At the time, we rely only on a compass to orient our array with reliability within 2–4°. The loss in accuracy is mainly from the mitigation effort from the magnetic element effect in the speaker when reading the compass. A computer speaker is positioned at different floors to simulate different elevation angles

Figure 12 3D DOA accuracy testing experiment performed at the Boelter Hall in UCLA.



(a) The node and source positions viewed from the top.

(b) The isotropic array.

Figure 13 3D DOA accuracy testing experiment setup.

that will playback the pre-recorded audio source. The audio file is 23 s long and consists of several repeating Acorn Woodpecker and Mexican Antthrush calls. The data collected by each node is then processed using 3D AML algorithm to get the DOA estimate. Note that since the array size is 8 cm and the bird calls are mostly between 2 and 5 kHz, then the ML criterion would have several large lobes. So the decoupled 3D AML with an arbitrary initial point can't be used to process the collected data. Since the original 3D AML algorithm does 2D grid search on all possible azimuth and all possible elevation angles, it is not very efficient. A more efficient solution is to run the decoupled 3D AML algorithm with more than one initial point and then compare the obtained maximum from each initial point and select the global maximum. The required number of initial points to guarantee converging to the global maximum is completely dependent on the isotropic array configuration and the frequency characteristics of the signal of interest. For the specified isotropic array configuration and the audio calls used in these experiments, having five initial points (with the elevation angles of 5, 25, 45, 65 and 85 in degrees) guarantee to find the global maximum. This way, we would still have low complexity and good performance.

For the first test, we placed the speaker on the first floor of Boelter Hall with height of 5.2 m. We processed the received signal from each subarray using the proposed 3D AML algorithm discussed in Section 3.2. As a result, we were able to estimate an azimuth angle and an elevation angle for each of the calls. Table 2 shows the mean and the standard deviation of the angular error for each node in degrees. For the first experiment, we lost node 153 due to a technical problem, so we only have data collected from the two other nodes. In the second and third experiments, the speaker is located on the second and third floor of the Boelter Hall with the height of 9.37 and 13.47 m, respectively.

Note that some of the results are biased, especially the results from node 151. We believe this bias is caused by array orientation error which we know only accurate within a few degrees. The rest of the tables show that the estimated DOAs are fairly close to the true DOAs, which is especially true for elevation angle estimates. This confirms the practicality of the proposed 3D AML algorithms in real life scenario.

4.4 AML Localization, Compact Deployment

The goal of this experiment is to verify whether the system is capable of performing source localization based on actual animal calls in the field by performing properly all the features of node time synchronization, self-localization of the nodes, event detection, and AML-based DOA bearing estimation as considered above.

The setup of the system was as follows. Six subarray nodes were spread over a region which surrounds the "Spruce Burrow", a location where most marmots alarm-called. Figure 14 shows the location of Spruce

Table 2 Angular error's mean and standard deviation for various source elevation. The entry (a,b)/(c,d) represents a and b as mean and standard of azimuth angle, c and d as means and standard deviation of elevation angle, and all units are in degree.

Node	Mean/Std. dev. (Azim., Elev.) degree	
	Acorn woodpecker	Mexican antthrush
5.2 m above ground		
151	(−0.36, 1.39)/(0.4, 1.01)	(−0.49, 2.29)/(1.93, 1.35)
152	(2.60, 2.95)/(0.74, 1.78)	(−1.47, −1.3)/(2.92, 3.73)
9.37 m above ground		
151	(−1.53, −2.68)/(1.41, 0.74)	(2.36, −1.26)/(1.37, 1.52)
152	(1.47, 1.42)/(1.18, 0.82)	(2.69, −1.44)/(2.38, 1.82)
153	(−0.09, −0.01)/(1.69, 1)	(−3.21, 2.03)/(2.8, 0.22)
13.47 m above ground		
151	(−2.59, −1.44)/(1.75, 0.95)	(−3.4, −3.9)/(1.98, 1.4)
152	(−3.68, 0.62)/(3.05, 0.46)	(−3.2, −1.2)/(3.3, 2.21)
153	(−1.55, 0.93)/(2.03, 0.87)	(2.4, 2.09)/(8.07, 0.53)

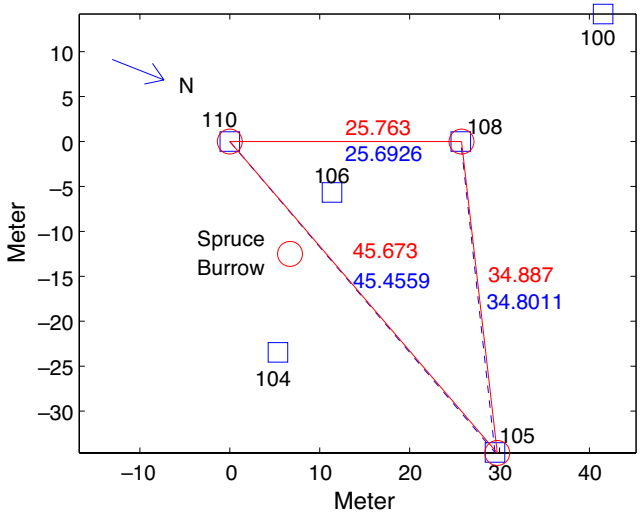


Figure 14 Compact deployment comparison between node self-localization and GPS data. GPS position is denoted by *round circles* connected in a *solid line*. *Top and bottom numbers pair* are self-localization and GPS distances respectively.

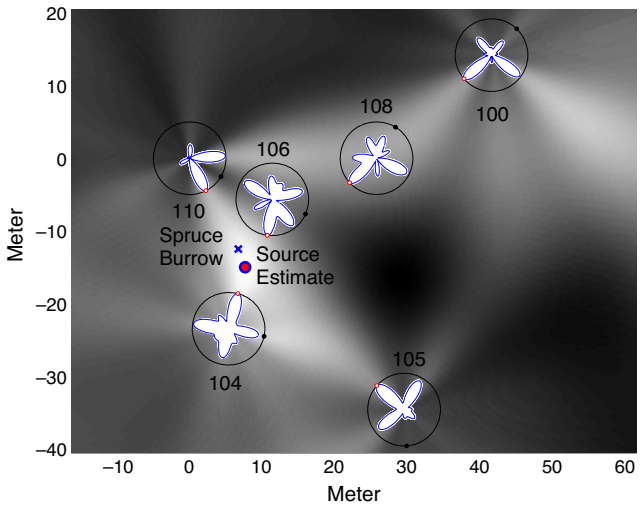


Figure 15 A pseudo-likelihood map from the compact deployment. Main lobes are denoted by the *circle on the log-likelihood ring*, and the *solid dots on these rings* are the array zero-degrees.

Burrow relative to the sensor nodes. To align the position and orientation of the nodes relative to the true Earth, we took three GPS readings on node 110, 108 and 105. Although two was sufficient, three enhanced our confidence. Figure 14 was made by translating node 110 GPS and self localization position on top each other and rotating node 110 to 108 GPS orientation to fall onto the self localization orientation. On each line the top numbers is the distance based on GPS coordinate, and the bottom is based on self localization. We know our self localization accuracy is within 5 cm in the open field [12]. The small difference between these numbers indicate that the GPS result is accurate well within its accuracy in the open field, which is about 1 m.

The position of the source is estimated using a pseudo-likelihood map. This map is generated in the following way. Each node runs the AML algorithm on the marmot calls to produce the bearing likelihood. Then, the 2D map is divided into uniform grid of positions, where the pseudo-likelihood of each position is computed by summing the bearing log-likelihood values of each nodes that points to that position. The estimate is chosen based on the most likely position. Figure 15 displays the pseudo-likelihood map with the bearing likelihood of each node in a polar coordinate.

In this experiment, error analysis is difficult to address. The marmots rarely call at the same location, and their position when making the calls is difficult to precisely record. Figure 15 displays the estimate in relation with the nodes and the burrow. This result is consistent with the notes made at the field, which can only indicate that a marmot was observed nearby the burrow. More analysis can only be done if caller localization can be improved by ground truth, or if the marmot doesn't move while making several calls; this is what happened next.

4.5 AML Localization, Wide Deployment

Our goal in this experiment was to investigate the limiting capability of the system by stretching the array

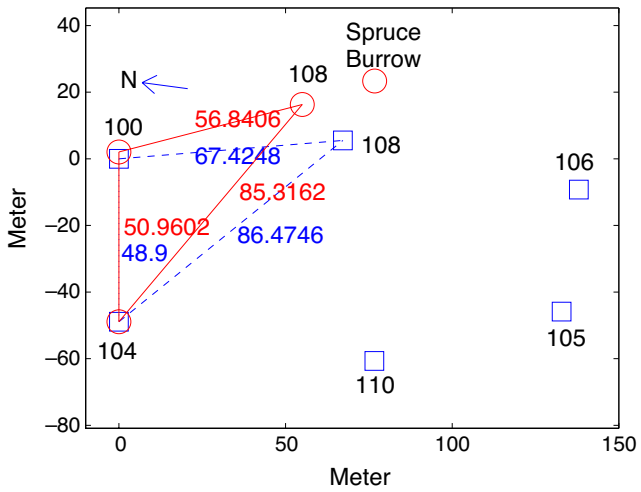


Figure 16 Wide deployment comparison between node self-localization and GPS data. GPS position is denoted by *round circles connected in a solid line*. *Top and bottom numbers pair* are self-localization and GPS distances respectively.

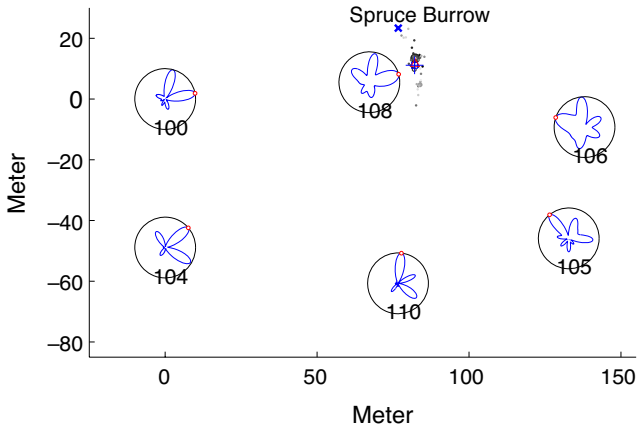


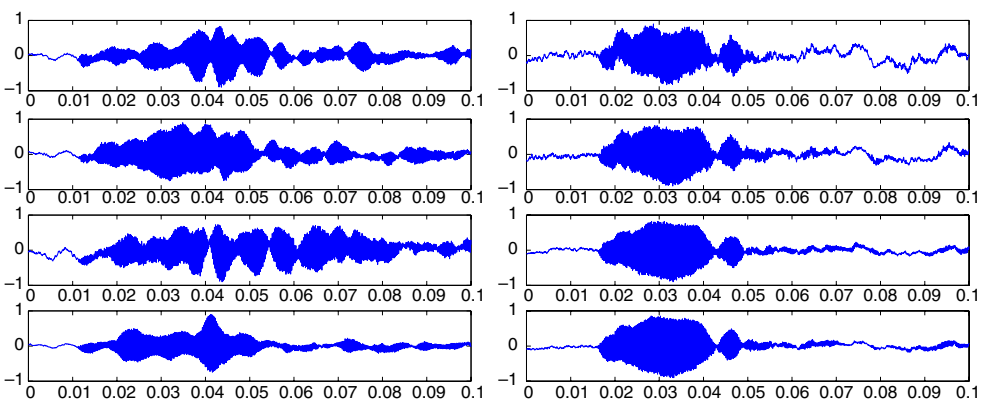
Figure 17 Scatterplot of location estimates from the wide deployment, relative to the position of the node 104 using all six nodes. μ = (82.1, 11.1), and σ = (2.9, 5.6) m denoted by *square* with cross hairs. Median is denoted by a *triangle*.

as large as the wireless link allows while performing the same task. Fortunately, the marmots chirps more frequently at a single spot, thus gives us opportunity to address more detailed error analysis.

The array setup was similar to the compact deployment except the distance between nodes, which was enlarged with the longest distance from node 104 to 106 spans 143.6 m. Similarly, to orient the node with respect to true Earth, we take three GPS readings on node 100, 104 and 108. Figure 16 displays the node positions according to self-localization and GPS coordinates, by aligning the corresponding points for nodes 100 and 104. In this analysis we encountered an interesting problem: as the graph clearly shows, the GPS and self-localization results show a discrepancy in which the position of node 108 differs by 10 m. Unfortunately, this problem was not discovered until we had packed up the equipment and left the field location, so we could not collect additional GPS data. At first, this might seem like a fatal error because we don't know which GPS value to rely to align self localization results to the true Earth. However, based on the estimated call positions, the burrow location and because the self-localization results are derived from an over-constrained system, we were able to produce a convincing argument that the GPS value for node 108 was flawed.

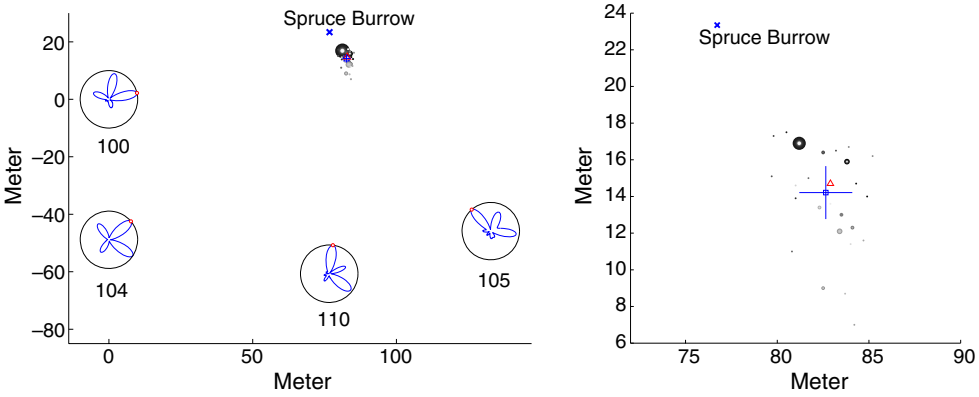
The marmot rapid, cyclic mode, chirping is called a *bout*, with one chirp every few seconds. In this type of behavior the marmot stands still while vocalizing, although it may move its head around to scan the surrounding area. This means we can use these data set to get the estimate distribution produced by our algorithms. With the estimation process follows a similar procedure as before, we generate the pseudo-likelihood map for each chirp and use 0.1 s duration in processing. Each estimated position is then collected and plotted in a scatter plot as seen in Fig. 17. The estimated positions are arranged in grayscale gradient from dark to light representing the progression in time from earlier to later event. To simplify computation, we divide the map into a square grid with a specified precision. The precision of the map we choose is 0.1 m, hence anything that falls within the 0.1×0.1 square is treated as the same location. In the graph, we indicate multiple estimates by making the dot size larger. The mean and standard deviation is shown as the square and the cross hairs, and the median is shown as triangle. The plot axis are set relative to node 100 and the mean (μ) and standard deviation (σ) are (82.1, 11.1) and (2.9, 5.6) m respectively.

The scatter plot has a "banana" shape that suggests the presence of correlated errors. Examining the results from nodes 108 and 106, we observed that the bearing



(a) Node 108. (b) Node 110.

Figure 18 A time domain segment containing a single marmot chirp from two nodes at different locations from the wide deployment.

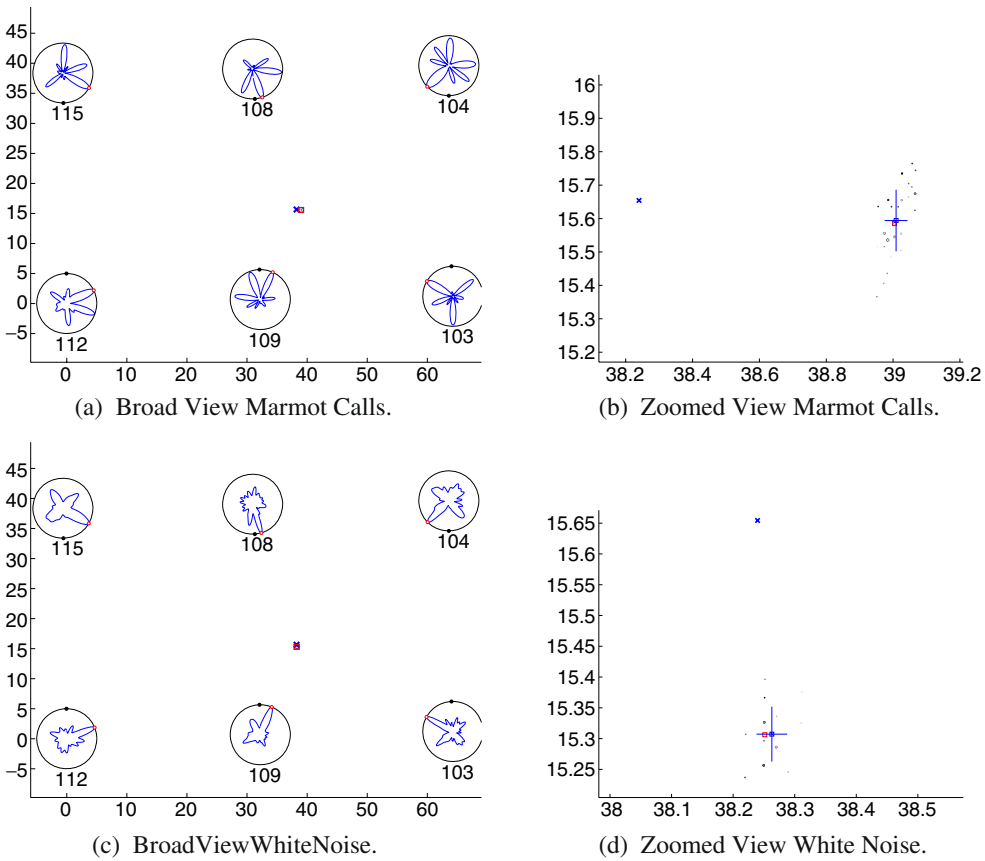


(a) Relative to node 100. (b) Zoomed in.

Figure 19 Scatterplot of location estimates from wide deployment after removing node 108 and 106. μ = (82.6, 14.2), and σ = (1.4,2.8) m denoted by square with cross hairs. Median is denoted by a *triangle*.

likelihoods reported by these nodes have broad, misshapen lobes that introduced error into the position estimate. Figure 18 shows time domain plots of one particular chirp recorded at nodes 108 and 110. From these plots it is clear that the data recorded at 110 is quite clean, whereas at 108 there was a great deal of distortion. We believe that this distortion is mainly caused by reverberation from trees. Generally, nodes that are closer to trees would be more susceptible to this problem. In addition, nodes close to the source can contribute more dispersion because their side-lobes are more likely to intersect near the true source location.

In Fig. 19a, we show the results of our localization after removing data from 108 and 106. This results in a tighter distribution, with mean (μ) (82.6, 14.2) m and standard deviation (σ) (1.4, 2.8) m. Clearly the sensors that are affected by reverberation yield poor estimate, and by selecting sensors that aren't affected, we can



(a) Broad View Marmot Calls. (b) Zoomed View Marmot Calls.

(c) BroadViewWhiteNoise. (d) Zoomed View White Noise.

Figure 20 Scatterplot of position estimates for source inside the convex hull. For marmot, μ = (39.01, 15.59), and σ = (0.03,0.09) m. For noise, μ = (15.87, -38.03), and σ = (0.04, 0.03) m. μ and σ are denoted by *square* with cross hairs respectively.

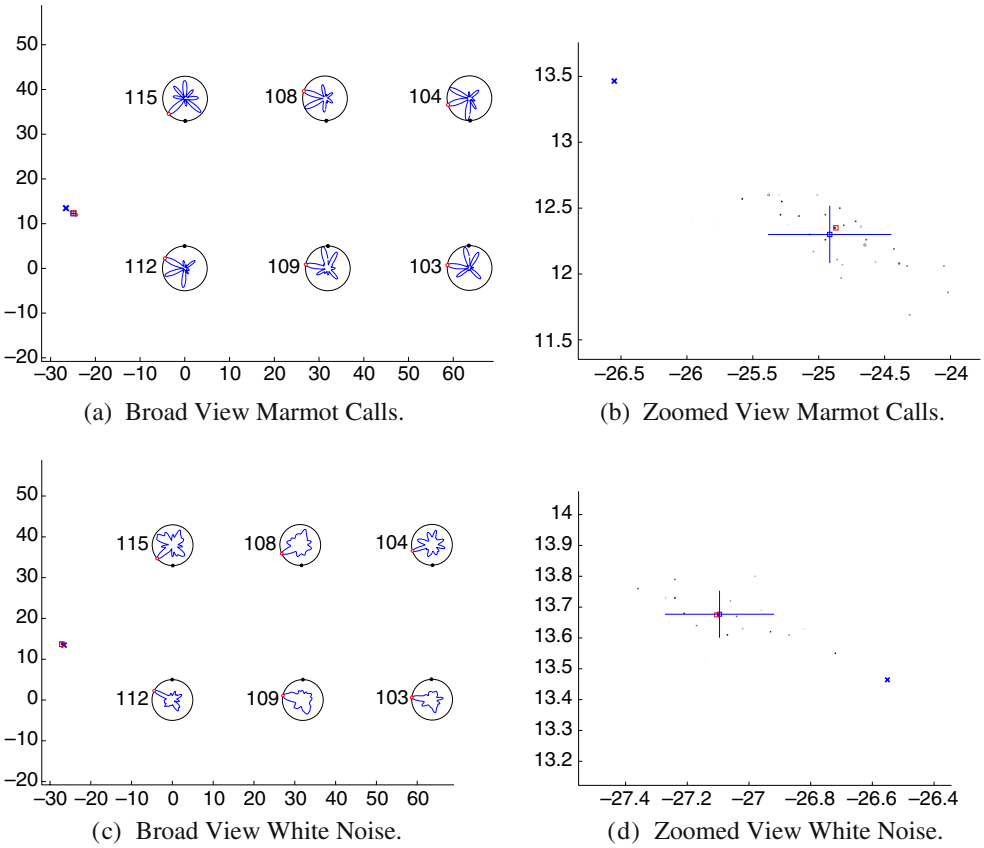


Figure 21 Scatterplot of position estimates for source inside the convex hull. For marmot, μ = (-24.92, 12.30), and σ = (0.47, 0.22) m. For white noise, μ = (−27.10, 13.68), and σ = (0.18, 0.08) m. *mu* and σ are denoted by *square* with cross hairs respectively.

minimize the variance of our estimate. A seemingly more reasonable approach is to apply multi-source estimation algorithm to estimate both the source and its reverberation. For this approach, the AML is equipped to do so [16]; however, the computation complexity can be impractical to implement in a sensor node with limited computational power. To asses this problem, our future work will investigate methods for automatically selecting nodes that achieves minimum variance, and investigate low complexity algorithms to handle reverberant case.

Figure 19b shows a zoomed-in version of Fig. 19a. Here we see there are two cluster of dots, one on the upper left and another on the bottom right. Confirmations from notes revealed that there are truly two call locations of the same marmot observed in this test (nicknamed "Smiley Face").

4.6 AML Localization, UCLA

In this experiment our goal is to characterize the performance of both localization and direction of arrival accuracy that are missing in RMBL from the lack of ground truth. However, we do not want to remove the self-localization step to test the accuracy of AML alone. Since we know the self-localization position estimate is accurate within 5 cm in the open field, we can use an extra node that can act both as a source and participate in the self-localization to get its positions. We further confirm this readings by measuring distances to surrounding nodes using a laser ranger. We also test two source positions that typically represents the usual deployment scenario: (1) when the nodes deployed are able to surround the source or inside the convex hull (ICH) of the nodes deployed, and (2) when the deployment cannot surround the target of interest or outside the convex hull (OCH), usually due to physical limitations such as wireless link, or GPS that performs better in the open area.

Table 3 Summary of position error statistics for each cases.

Case	Mean	Std. dev.	RMS dist.
ICH Marmot	(39.01, 15.59)	(0.03,0.09)	0.78
ICH Noise	(38.27, 15.31)	(0.02,0.04)	0.35
OCH Marmot	(−24.92, 12.30)	(0.47, 0.22)	2.07
OCH Noise	(−27.10, 13.68)	(0.18, 0.08)	0.61

Table 4 Summary of angular error statistics for each node.

Case	103	104	108	109	112	115
Mean						
ICH Marmot	−0.83	0.47	1.44	−2.69	3.50	1.43
ICH Noise	−0.03	1.51	−0.51	−0.54	−0.59	−0.19
OCH Marmot	1.19	2.40	3.71	3.13	−0.21	0.93
OCH Noise	2.24	2.46	2.11	1.49	−0.05	−1.34
Standard deviation						
ICH Marmot	0.23	0.49	0.13	0.07	0.27	1.12
ICH Noise	0.08	0.36	0.11	0.08	0.10	0.38
OCH Marmot	0.30	0.33	0.32	0.17	0.35	0.36
OCH Noise	0.34	0.19	0.44	0.47	0.13	0.25

Six nodes are deployed in a rectangular manner with each vertical pair faces each other. The nodes are raised up approximately 1.2 m above the ground and leveled with a water bubble leveler. Seven vertical and horizontal distances are measured with a laser distance ranger, and the orientation with a compass. At each source we played both a white noise and a series of pre-recorded marmot chirps. Each analysis is using 0.1 s data that is passed to a high pass filter to remove the low frequency wind and humming noise. For the marmot analysis the data is band pass filtered at marmot's dominant frequency to remove high frequency interferers such as crickets.

Figure 20 display the scatter plot of the inside convex hull case with pre-recorded marmot call playback. The dots shows 40 position estimates, and the beam patterns are taken from the last estimate. The cross represents the source position at (38.24, 15.65) m. The mean and standard deviation of the estimates are (39.01, 15.59) and (0.03,0.09) m respectively and are represented by the square and the crosshair. The root-mean-square (RMS) distance error is 0.78 m. Given the tightness of the standard deviation, the estimate has a slight bias. It seems at first that this is caused by orientation error in the self-localization, but several fact that against this is first the white noise analysis shown in Fig. 20c shows better and consistent estimates. Furthermore, the individual node angular error of white noise source does decrease as shown in Table 4; therefore cannot be dismissed as coincidence. We also investigate the non-uniform SNR across sensors, and when we remove a sensor that has low SNR, the angular error change is still within one standard deviation. Upon correlating the data with the original signal we found two high peaks that consistent if a reverberant comes from the nearest buildings, but processing the data before the reverberant arrives also resulted in angular change within one standard deviation. Another possibility is reflection from the ground which translates to 10–15 samples delay which is difficult to detect. Since the sensor arrangement are 3D, the non-perfect tiltness can also contribute to this bias. Furthermore, all the imperfection from self-localization result may also add up.

For marmot source outside the convex hull, the position estimate suffers from not having total beam crossings greater than 90°. The estimate resulted in a mean of (−24.92, 12.30), standard deviation (0.02, 0.04) m, and RMS distance error of 2.07 m, where the actual position from self-localization is (−26.55, 13.46) m (Fig. 21). Viewing the angular error from Table 4 reveals that neither the marmot case and the white noise case are very accurate. However, the white noise angular error happens to cancel each other giving RMS distance error of 0.61 m.

Summaries of these results can be seen in Tables 3 and 4. Very tight variances indicates the AML algorithm produces repeatable and consistent estimates. These results suggest that for inside the convex hull, we can get accuracy less than 1 m, and the accuracy for outside convex hull is around 2 m. Given the size of this array deployment is approximately 35 × 65 m, the error ratio to the smaller side is 2.9% and 5.7% respectively.

5 Conclusion

The ability to deploy an automated system to detect, localize, and record animal vocalizations in the field enables a host of new observations and approaches to biological questions. The successful deployment of the Acoustic ENSBox based system at the Rocky Mountain Biological Laboratory to study marmot alarm calls provides a powerful proof of concept. Our results indicate that it is tractable to localize marmot alarm calls to within at least a few meters, despite a noisy and geographically rough environment.

The self-localization feature of the Acoustic ENSBox nodes is a necessity for practical field deployment. This capability provides a robust and high-accuracy alternative to reliance on GPS and simple distance measurements. When combined with GPS and distance measurements, the different measurement modes provide invaluable error and consistency checking, as well as providing calibration to the speed of sound and registration of the node map with respect to global coordinates. Given the unexpected vagaries encountered in the field and the difficulty of always checking values in real time, independent redundant measurements are extremely valuable.

The tests of the on-line marmot detector demonstrated that a streaming detector could be developed quickly and deployed on the ENSBox nodes without excessive optimization. Demonstrating in-network data reduction, we showed that this detector could prefilter the data to meet the requirements of on-line localization algorithms which cannot run streaming in real time. Our experience also motivated the need for interactive development in the field. We anticipate that in future deployments: (1) initially, samples of raw data will need to be collected and analyzed, and (2) parameters—and in some cases, algorithms—will need to be tuned in the field in response to the particular conditions observed in the deployment. We are currently pursuing an interactive, query-oriented approach to these needs in the context of the WaveScope project [23].

Our controlled experiments with 2D AML based bearing estimation showed that pre-recorded playback tests under actual field conditions produces results consistent to within a few degrees out to 50 m, even though the volume of playback is significantly lower than a live marmot call. This result is especially encouraging since in the most extreme case with the source speaker facing away from the sub-array 50 m away, the call itself is nearly inaudible over the background noise to the human ear. The 3D AML based bearing estimation also shown to have satisfactory results despite the low array azimuth orientation accuracy.

Source localization experiments of actual marmots in their natural environment proved quite successful. Combining AML based DOA likelihoods from multiple nodes effectively overcomes the problem of ambiguity produced from a relatively large sub-array size. As demonstrated by the wide deployment experiment, redundancy provided by multiple nodes can be used to identify and exclude sub-arrays which have especially poor data due to reverberations, multi-path, or other practically unavoidable problems.

Finally, a controlled source localization experiment proved the system to be very consistent and accurate within a few meters. Redundancy in the measurements has provide convincing case for node position and orientation accuracy.

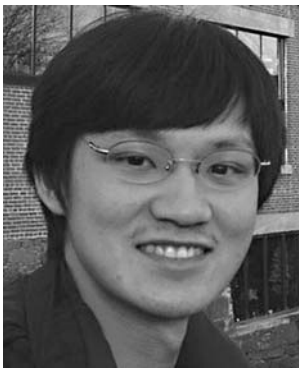
Continued development of this system will further reduce its size and weight, make it more weather resistant, and increase its sensitivity and accuracy. Upcoming field applications include further work with marmots at RMBL to test hypotheses regarding selfishness and trust when making and responding to alarm calls. The 3D array analysis, design, simulation, and experimental measurement at UCLA discussed in the paper are all aimed for our future deployment to study tropical birds in the rainforests of Chajul, Mexico.

Acknowledgements The authors would like to thank Chiao-En Chen, Wei-Ho Chung, and Yuan Yao for assistance in data collections at UCLA, Vlad Trifa and Martin Lukac for providing feedbacks on a draft of this paper, and RMBL for hosting us during the experiment. The authors would also like to thank the reviewers for their time and constructive suggestions. This work is partially supported by National Science Foundation (NSF) Center of Embedded and Network Sensing program under Cooperative Agreement CCR-012, NSF grant EF-0410438, University of California Discovery grant sponsored by ST Microelectronics, a UCLA faculty research grant to DTB, and the MIT WaveScope project (NSF).

References

1. Bradbury, J., & Vehrencamp, S. (1998). *Principles of animal communication.* Sunderland: Sinauer.
2. McGregor, P., Peake, T., & Gilbert, G. (2000). Communication behavior and conservation. In L. Gosling, & W. Sutherland (Eds.), *Behaviour and conservation* (pp. 261–280). Cambridge: Cambridge University Press.
3. Blumstein, D. (2007). The evolution of alarm communication in rodents: Structure, function, and the puzzle of apparently altruistic calling in rodents. In J. Wolff, & P. Sherman (Eds.), *Rodent societies*. Chicago: U. Chicago Press.
4. Hall, M. L. (2004). A review of hypotheses for the functions of avian duetting. *Behavioral Ecology and Sociobiology, 55*, 415–430.
5. Blumstein, D. T., & Armitage, K. B. (1997). Alarm calling in yellow-bellied marmots: The meaning of situationally-specific alarm calls. *Animal Behavior, 53*, 143–171.
6. Blumstein, D. T., & Munos, O. (2005). Individual, age and sex-specific information is contained in yellow-bellied marmot alarm calls. *Animal Behavior, 69*, 353–361.
7. Blumstein, D. T., & Daniel, J. C. (2005). Yellow-bellied marmot discriminate between the alarm calls of individuals and are more responsive to the calls from juveniles. *Animal Behavior, 8*, 1257–1265.
8. Vilches, E., Escobar, I., Vallejo, E., & Taylor, C. (2006). Data mining applied to acoustic bird species recognition. *International Conference on Pattern Recognition, 3*, 400–403.

9. Trifa, V., Girod, L., Collier, T., Blumstein, D. T., & Taylor, C. E. (2007). Automated wildlife monitoring using self-configuring sensor networks deployed in natural habitats. In *The 12th international symposium on artificial life and robotics (AROB)*.
10. Girod, L., Lukac, M., Trifa, V., & Estrin, D. (2006). The design and implementation of a self-calibrating distributed acoustic sensing platform. In *ACM SenSys.* Boulder, CO.
11. Wang, H., et al. (2005). Acoustic sensor networks for woodpecker localization. In *SPIE conference on advanced signal processing algorithms, architectures and implementations,* 5910, 591009.1–591009.12.
12. Girod, L. (2005). *A self-calibrating system of distributed acoustic arrays.* Ph.D. thesis, Univerity of Caliornia at Los Angeles.
13. Elson, J., Girod, L., & Estrin, D. (2002). A wireless time-synchronized COTS sensor platform, part i: System architecture. In *IEEE CAS workshop on wireless communications and networking.*
14. Girod, L., et al. (2004). Emstar: A software environment for developing and deploying wireless sensor networks. In *Proceedings of the 2004 USENIX technical conference.* USENIX Association, Boston, MA.
15. Trifa, V. (2006). *A framework for bird songs detection, recognition and localization using acoustic sensor networks.* Master's thesis, École Polytechnique Fédérale de Lausanne.
16. Chen, J., Yao, K., & Hudson, R. (2002). Maximum-likelihood source localization and unknown source localization estimation for wideband signals in the near-field. *IEEE Transactions on Signal Processing, 8*, 1843–1854.
17. Kay, S. (1993). *Fundamentals of statistical signal processing: Estimation theory.* New Jersey: Prentice-Hall.
18. Baysal, U., & R.M. (2003). On the geometry of isotropic arrays. *IEEE Transactions on Signal Processing, 51*(6), 1469–1478.
19. Asgari, S., Ali, A., Collier, T., Yao, Y., Hudson, R., Yao, K., et al. (2007). Theoretical and experimental study of doa estimation using aml algorithm for an isotropic and non-isotropic 3d array. In *SPIE conference on advanced signal processing algorithms, architectures, and implementations.* SPIE, (vol. 6697, pp. 66970I-1–66970I-12).
20. Chen, C. E., Wang, H., Ali, A. M., Lorenzelli, F., Hudson, R. E., & Yao, K. (2006). Particle filtering approach to localization and tracking of a moving acoustic source in a reverberant room. In *IEEE ICASSP06*.
21. Chen, C. E., Ali, A. M., Wang, H., Asgari, S., Park, H., Hudson, R. E., et al. (2006). Design and testing of robust acoustic arrays for localization and enhancement of several bird sources. In *Symposium on information processing in sensor networks* (IPSN06) (pp. 268–275). ACM Press.
22. Chen, C. E., Lorenzelli, F., Hudson, R. E., & Yao, K. (2008). Maximum likelihood DOA estimation of multiple wideband sources in the presence of nonuniform sensor noise. *EURASIP Journal on Advances in Signal Processing, 2008*, 12. doi:10.1155/2008/835079.
23. Girod, L., Jamieson, K., Mei, Y., Newton, R., Rost, S., Thiagarajan, A., et al. (2007). The case for WaveScope: A signal-oriented data stream management system (position paper). In *Proceedings of third biennial conference on innovative data systems research* (CIDR07) (pp. 397–406).



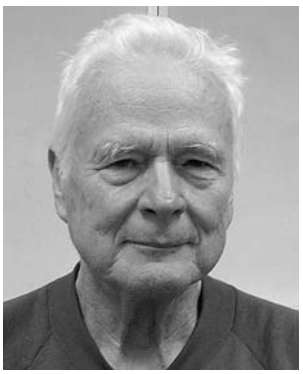
Andreas Mantik Ali received his B.S. (Highest Honors) and M.S. degrees in Electrical Engineering from University of California, Los Angeles. He is currently a Ph.D. candidate under advisory of Professor Kung Yao and in close collaboration with the Center for Embedded Network Sensors. His research interests include source localization, array processing, multiple target tracking and wireless sensor networks.

Shadnaz Asgari received her B.S. degree in Electrical Engineering from Sharif University of Technology in 1999. Since 2002 she has been a graduate student in the University of California, Los Angeles (UCLA) where she is currently doing research under the Prof. Yao's supervision. Her areas of interests include array signal processing, source localization and beam-forming in the sensor networks.

Travis C. Collier holds a B.S. in Engineering and Applied Sciences from Caltech, and is currently a graduate student researcher in Ecology and Evolutionary Biology at UCLA in close collaboration with the Center for Embedded Network Sensors. His research interests have included Artificial Life, population modeling, self-organizing systems, language evolution, and sensor networks.

Lewis Girod received the B.S. and M.E. degrees in Computer Science from MIT in 1995. He received his Ph.D. in Computer Science from University of California at Los Angeles in December 2005. After working at LCS for 3 years in the area of Internet naming infrastructure, he joined Deborah Estrins group in 1998. He is currently working as a Postdoctoral Researcher at the MIT Computer Science and AI Laboratory, on the WaveScope project. His research focus is the development of robust networked sensor systems, specifically physical localization systems that use multiple sensor modalities to operate independently of environment and deployment.

Michael Allen received the B.S. degree in Computer Science from Coventry University, UK in 2005. He is currently a Ph.D. student with the Cogent Applied Research Centre at Coventry University, under the supervision of Dr. Elena Gaura. His research interests include source and node localization in wireless sensor networks.



Ralph E. Hudson BSEE, UC, Berkeley, 1960, Ph.D. US Naval Postgraduate School, 1966. US Navy 1960 to 1973 naval aircraft electronic warfare systems. Hughes Aircraft Co. 1973 to 1993 airborne radar digital signal processing. Consultant and research associate at UCLA, 1993 to present. Research interest signal and array processing and synthetic array radar system design and signal processing.

Kung Yao received his B.S. (Highest Honors) and Ph.D. degrees in Electrical Engineering from Princeton University. Presently, he is a Distinguished Professor in the Electrical Engineering Dept. of UCLA. He was an Assistant Dean at the UCLA School of Engineering and Applied Science from 1985 to 1988. From May to June 2007, he was a Royal Society Kan Tong Po Visiting Professor at HK Polytechnic Univ. His research interests include acoustic/seismic array processing, sensor system, wireless communication theory and system, and systolic array algorithm and systems. He has published over 250 journal and conference papers. Dr. Yao received the IEEE Signal Processing Society's 1993 Senior Award in VLSI Signal Processing. He has served as Associate Editors of IEEE Trans. on Information Theory, IEEE Trans. on Signal Processing, IEEE Trans. on Circuits and Systems, IEEE Communication Letters, etc. He is a Life Fellow of IEEE.

Charles E. Taylor received his A.B. from the University of California, Berkeley in Genetics and his Ph.D. in Ecology and Evolution from the State University of New York, Stony Brook. Presently, he is a professor of Ecology and Evolutionary Biology at UCLA. He was co-director of the UCLA Cognitive Science Research Program from 1991 to 1999. His principle research areas are: (1) in adaptive sensor arrays, used to identify and localize birds—an outgrowth of his background in artificial life; and population genetics of *Anopheles gambiae*, the principal vector of malaria throughout much of Africa. Prof. Taylor has been a co-editor of 3 books and an author of 125+ articles and conference papers. He has served as co-editor of the journal Artificial Life (MIT Press) from 1997 to 2001, and has been a member of the editorial boards of Artificial Life, High Integrity Systems, IEEE Transactions on Evolutionary Computation, Artificial Life and Robotics and of the International Journal of Distributed Sensor Networks.

Daniel T. Blumstein is an Associate Professor of Ecology and Evolutionary Biology at UCLA. He received his B.S. (Magna Cum Laude—Environmental, Population and Organismic Biology; Cum Laude—Environmental Conservation) from the University of Colorado Boulder, and his M.S. and Ph.D. in Animal Behavior at the University of California Davis. He has studied alarm communication and bioacoustics in marmots for nearly two decades, and has written over 150 articles and book chapters, and two books, about these and other subjects. He is an Editor of the journal Animal Behavior, and on the editorial boards of Behavioral Ecology, and Biology Letters.