An Empirical Study of Collaborative Acoustic Source Localization

Andreas M. Ali,
Kung Yao
Electrical Engineering
University of California, Los Angeles

Travis C. Collier,
Charles E. Taylor, Daniel T. Blumstein
Ecology and Evolutionary Biology
University of California, Los Angeles

Lewis Girod
Computer Science and AI Laboratory
Massachusetts Institute of Technology

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 individual 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 an AML-based source localization algorithm, and use it to localize marmot alarm-calls. We assess the performance of these techniques based on results from two field experiments: (1) a controlled test of direction-of-arrival (DOA) accuracy using a pre-recorded source signal, and (2) an experiment to detect and localize actual animals in their habitat, with a comparison to ground truth gathered from human observations. 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.**

I. 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, and are 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 the degree of 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 1, 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].

Using the Acoustic ENSBox, 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 system based on a Constant False Alarm Rate (CFAR) algorithm, (2) a centralized marmot call localization system based on AML bearing estimation, and (3) a thorough evaluation of the effectiveness of these algorithms based on a field study detecting real animals in their natural habitat.

II. 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 localization, 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.

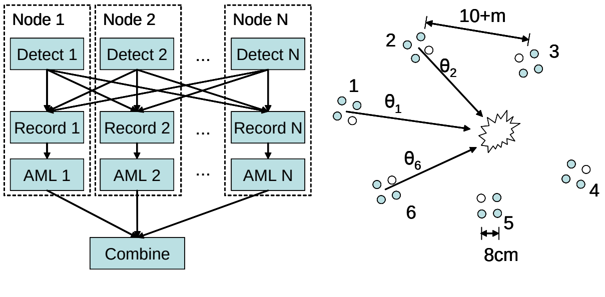


Fig. 1. Block diagram of a DOA-based localization system. Note, for the experimental results in this paper, the AML algorithm was run centrally.

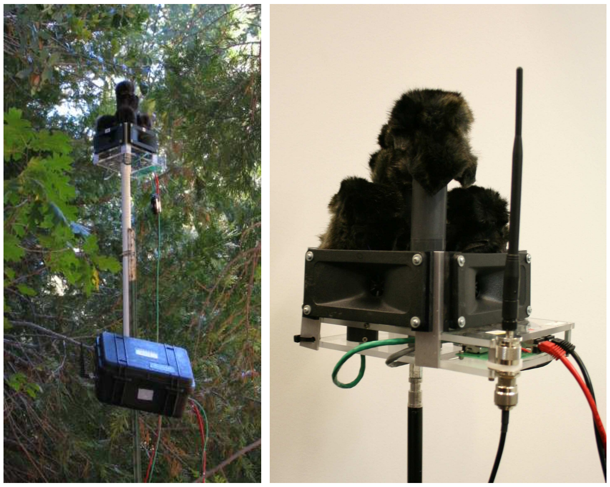


Fig. 2. The Acoustic ENSBox Platform.

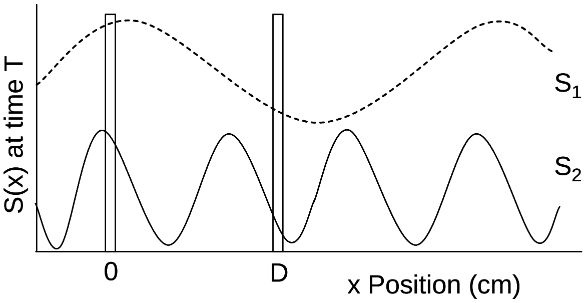


Fig. 3. The spatial aliasing problem. Sensors at position 0 and D observe the same phase offset for signals S_1 and S_2. When measured phase offsets are used to deduce direction of arrival, ambiguities result for signals with $\lambda < 2D$.

However, the features of animal calls that are most readily extracted tend to be narrow-band, making precise phase acquisition difficult in a noisy environment. Furthermore, techniques based on coherent processing of data from different nodes (e.g., correlation) are limited by the coherence properties of the acoustic environment and may not work well when the nodes are separated over tens of meters.

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 sections, 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 III.

A. DOA-based localization using distributed sub-arrays

Fig. 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 the Acoustic ENSBox platform [8] shown in Fig. 2 and described in more detail in Section II-C. Each node in the system hosts a 11.31 cm tetrahedral microphone sub-array, rotated to form an 8 cm square when viewed from above. These sub-arrays are typically deployed at least 10 m apart, and often much farther: in the three-set of experiments presented in this paper, 6 sub-arrays are deployed surrounding a 70x140 m area (see Fig. 14). 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.

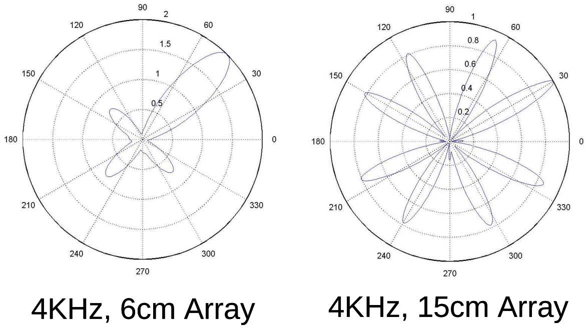
The detection software on each node performs a streaming analysis of the acoustic data in real time, identifying likely animal call events. Whenever any individual node's call detector is triggered, a radio message is sent to trigger all the nodes in the system to start recording that event and queue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

Once identified, segments of audio containing calls are analyzed using the Approximate-Maximum-Likelihood (AML) algorithm described in Section III-B. 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 2-D 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 space.

B. 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 the 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. The reason for this is shown in Fig. 3. Just as the Nyquist theorem states that to avoid



4KHz, 6cm Array

4KHz, 15cm Array

Fig. 4. Simulated beam patterns for two arrays detecting marmot alarm calls. Larger arrays yield narrower lobes, but more ambiguity.

aliasing, a signal must be sampled at least at 2x the maximum frequency of the signal, an analogous property holds for *spatial* sampling. In order to measure the phase of an incoming signal by comparison from two points in *space*, those two points must lie in the same half-wave. Energy in frequencies with wavelengths shorter than 2x the sensor spacing will be aliased into lower frequencies. This implies that for a sensor spacing of D and signal propagation speed V_s, the maximum frequency detectable without aliasing is

$$F_c = V_s/(2D).$$

The likelihood metric of the bearing estimate is typically represented as a polar plot, 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 [9]. These grating lobes often have heights comparable to the true main lobe due to various random imperfections. 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 resolution achievable with the array also decreases. This resolution is a function of many factors, including the resolution of a sample and the accuracy of the array calibration. Lower resolution increases the width of the main lobe, increasing the uncertainty in the estimate. This tradeoff is depicted in Fig. 4, which shows simulated beam patterns for two different size arrays detecting a 4 KHz source; the larger array has narrower 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, low resolution, main lobe.

To avoid this problem, we ensure high resolution 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 probability of intersecting with few 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 side lobes become increasingly broad, this approach will eventually fail, hence 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.

C. 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 [8]. 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.

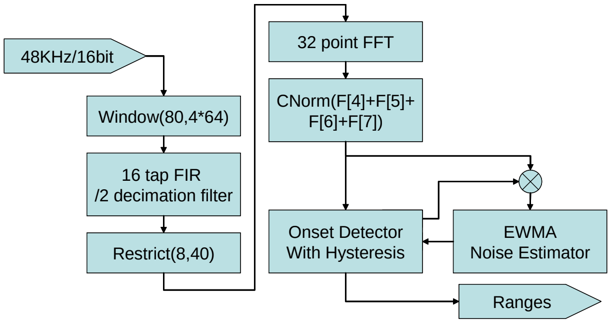


Fig. 5. Block diagram of a CFAR marmot detection algorithm. This implementation is based on a similar algorithm for detecting bird calls [10].

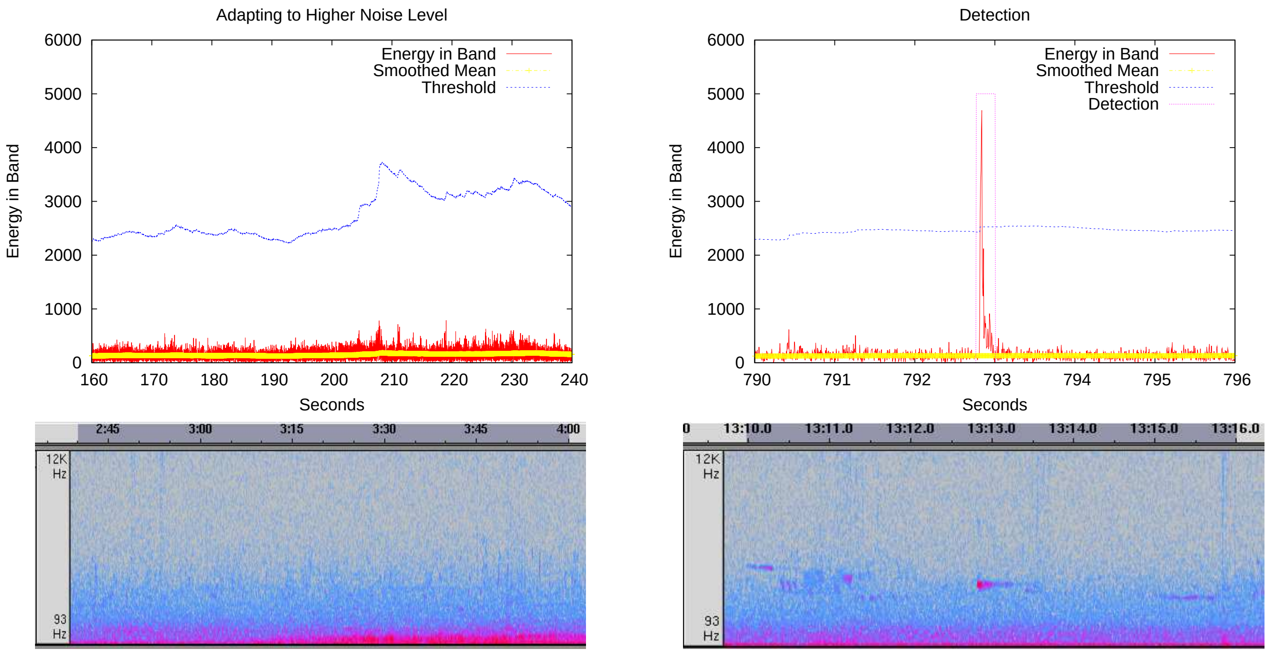
a) Packaging: While packaging issues are far from novel, they are quite important in practice. In prior attempts to record using multiple arrays, our equipment was cobbled together from "off-the-shelf" components and a surfeit of wires, plus heavy batteries to support devices that were not tuned for low power consumption. These solutions proved too cumbersome to be practical. In contrast, the ENSBox is a wireless distributed sensor system. Each unit is a self-contained processor and array, with an internal battery and a lifetime of 10 hours. The unit is water-resistant and the array head can be fitted to a tripod.[1]

b) 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.

c) Self-configuration: Another important factor in deployments is self-configuration. 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 [8], [11], can compute relative array positions to within 10 cm over an 50x80m field, and estimate array orientation to within 1.5 degrees. Extensive testing has proved that the system is easy to operate, achieves detection ranges upwards of 100m, and 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.

d) Software API: The final advantage of the ENSBox has been as an application development platform. The ENSBox provides a synchronized sampling API that greatly simplifies the development of collaborative sensing application software [12]. The detector

[1] We are currently developing version 2 of the ENSBox, which will be smaller, lighter, completely free of wires, and easier to deploy.



(a) Adaptation to an increase in noise level.

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

Fig. 6. Behavior of the CFAR-based marmot alarm call detector.

application described in this paper is 800 lines of C code and took about 1 week to develop within the EmStar software framework [13]; 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.

III. 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.

A. CFAR Event Detection

The *Constant False Alarm Rate (CFAR)* detection algorithm is an on-line statistical method for identifying the bursts of energy in a streaming signal [10]. The CFAR algorithm is based on the assumption that the ambient noise can be modeled as a Gaussian distribution $N(\mu, \sigma^2)$. Given this assumption, we can use a smoothing filter to compute on-line estimates $\bar{\mu}$ and $\bar{\sigma}$. Given 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 $1 - \mathrm{erf}(\beta/\sqrt{2})$, 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.

Fig. 5 shows a block diagram of our marmot detection algorithm. The first three stages of the data flow implement a decimation filter and windowing operation, resulting in windows of 32 points sampled at 24 KHz. To reduce the processing load, we only process every fourth 32 point window. Because marmot calls are approximately 0.04 seconds 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 CFAR 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.

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.

Fig. 6 shows the behavior of our on-line marmot call detector in two cases. Fig. 6(a) represents a period of time in which the ambient noise level is gradually increasing, and the mean and threshold adapt to this change. Fig. 6(b) 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 different parameter values. The parameter values used in our implementation are given in the following tables.

Parameter	Value
F_s	24000 Hz
FFT points	32
Window Feed	128
Frequency Bins	3,4

Parameter	Value
α	0.999
β	32
K_{init}	300
K_{min}	40
K_{max}	120

B. AML

Approximate-Maximum-Likelihood (AML) is a likelihood-based algorithm that searches the event space for the most likely feature of the event [14]. 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 estimate becomes unreliable. For simplicity, we assume both the source and the sensor array lie in the same plane (i.e., a 2-D space) as shown in Fig. 7, although the physical configuration of the four microphones on each node allows for a 3-D operation (which will not be considered in the paper).

Let there be M wideband sources, each at an angle θ_m from the array with the reference direction pointing to the east. The sensor array consists of P randomly distributed sensors, each at position $\mathbf{r}_p = [x_p, y_p]^T$. The sensors are assumed to be omni-directional and have identical response. The array centroid position is given by $\mathbf{r}_c = \frac{1}{P}\sum_{p=1}^{P}\mathbf{r}_p = [x_c, y_c]^T$. We use the array centroid as the reference point and define a signal model based on the relative time-delays from this position. The relative time-delay of the mth source is given by $t_{cp}^{(m)} = t_c^{(m)} - t_p^{(m)} = [(x_c - x_p)\cos\theta_m + (y_c - y_p)\sin\theta_m]/v$, in where $t_c^{(m)}$ and $t_p^{(m)}$ are the absolute time-delays from the mth source to the centroid and the pth sensor, respectively, and v is the speed of propagation. In a polar coordinate system, the above relative time delay can also be expressed as $t_{cp}^{(m)} = r_p\cos(\theta_m - \phi_p)/v$, where r_p and ϕ_p are the range and angle of the p sensor with respect to the array centroid. The data received by the pth sensor at time n is then

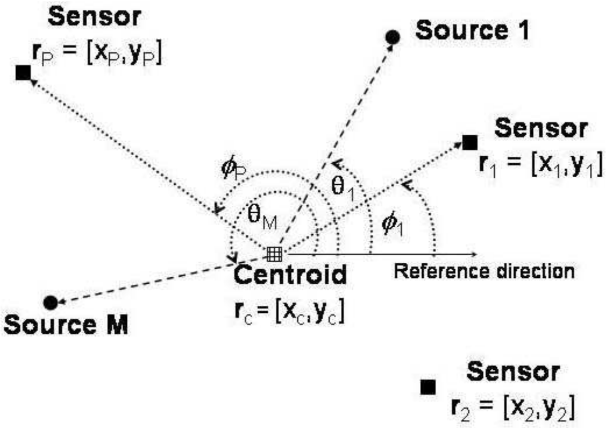


Fig. 7. Far-field notations for sources, sensors, and sensor array centroid

$$x_p(n) = \Sigma_{m=1}^{M} S^{(m)}(n - t_{cp}^{(m)}) + w_p(n), \tag{1}$$

for $n = 0, ..., N-1$, $p = 1, ..., P$, and $m = 1, ..., M$, where N is the length of the data vector, $S^{(m)}$ is the mth source signal arriving at the array centroid position, $t_{cp}^{(m)}$ is allowed to be any real-valued number, and w_p is the zero mean white Gaussian noise with variance σ^2.

For the ease of derivation and analysis, the received wideband signal can be transformed into the frequency domain via the DFT, where a narrowband model can be given for each frequency bin. However, 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 N-point DFT transformation, the array data model in the frequency domain is given by

$$\mathbf{X}(\omega_k) = \mathbf{D}(\omega_k)\mathbf{S}(\omega_k) + \eta(\omega_k), \tag{2}$$

for $k = 0, ..., N-1$, where the array data spectrum is $\mathbf{X}(\omega_k) = [X_1(\omega_k), ..., X_P(\omega_k)]^T$, the steering matrix $\mathbf{D}(\omega_k) = [\mathbf{d}^{(1)}(\omega_k), ..., \mathbf{d}^{(M)}(\omega_k)]$, the steering vector is given by $\mathbf{d}^{(m)}(\omega_k) = [d_1^{(m)}(\omega_k), ..., d_P^{(m)}(\omega_k)]^T$, $d_p^{(m)} = e^{-j2\pi k t_{cp}^{(m)}/N}$, and the source spectrum is given by $\mathbf{S}(\omega_k) = [S^{(1)}(\omega_k), ..., S^{(m)}(\omega_k)]^T$. The noise spectrum vector $\eta(k)$ is zero mean complex white Gaussian distributed with variance $N\sigma^2$. Note, 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 time-domain noise has an arbitrary i.i.d. distribution (with bounded variance). This asymptotic property in the frequency-domain provides a more reliable noise model than the time-domain model in some practical cases. Throughout this paper, we denote superscript ^{T} as the transpose, and ^{H} as the complex conjugate transpose.

The AML estimator performs the data processing in the frequency domain. The maximum-likelihood estimation of the source DOA and

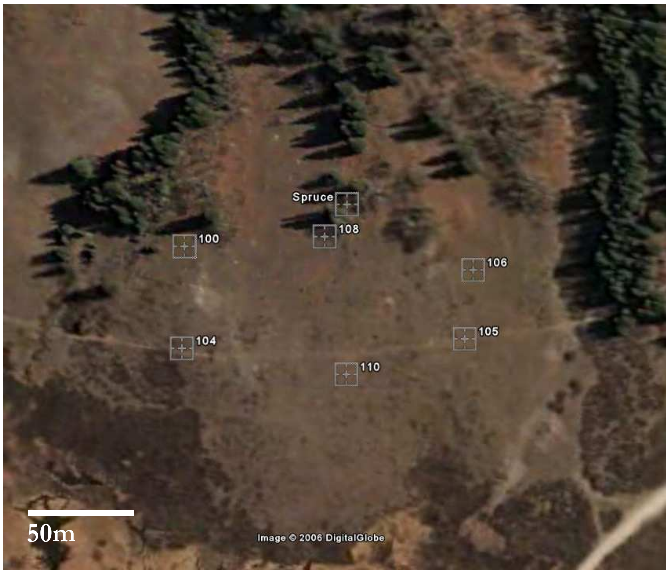


Fig. 8. "Marmot Meadow" location at Rocky Mountain Biological Laboratory, in Gothic, Colorado. The node locations correspond to the wide deployment described in Section IV-D.

source signals is given by the following optimization criterion [14]

$$\max_{\mathbf{\Theta},\mathbf{S}} L(\mathbf{\Theta},\mathbf{S}) = \min_{\mathbf{\Theta},\mathbf{S}} \sum_{k=1}^{N/2} ||\mathbf{X}(\omega_k) - \mathbf{D}(\omega_k)\mathbf{S}(\omega_k)||^2, \quad (3)$$

which is equivalent to a nonlinear least square problem. Using the technique of separating variables [15], the AML DOA estimate can be obtained by solving the following likelihood function

$$\max_{\mathbf{\Theta}} J(\mathbf{\Theta}) = \max_{\mathbf{\Theta}} \sum_{k=1}^{N/2} \text{tr}(\mathbf{P}(\omega_k,\mathbf{\Theta})\mathbf{R}(\omega_k)), \quad (4)$$

where $\mathbf{P}(\omega_k,\mathbf{\Theta}) = \mathbf{D}(\omega_k)\mathbf{D}^\dagger(\omega_k)$, $\mathbf{D}^\dagger = (\mathbf{D}(\omega_k)^H\mathbf{D}(\omega_k))^{-1}\mathbf{D}(\omega_k)^H$ is the pseudo-inverse of the steering matrix $\mathbf{D}(\omega_k)$ and $\mathbf{R}(\omega_k) = \mathbf{X}(\omega_k)\mathbf{X}(\omega_k)^H$ is the one snapshot covariance matrix. Once the AML estimate of $\mathbf{\Theta}$ is found, the estimated source spectrum can be given by

$$\hat{\mathbf{S}}^{ML}(\omega_k) = \mathbf{D}^\dagger(\omega_k, \hat{\mathbf{\Theta}}^{ML})\mathbf{X}(\omega_k). \quad (5)$$

The AML algorithm performs signal separation by utilizing the physical separation of the sources, and for each source signal, the Signal to Interference-plus-Noise Ratio (SINR) is maximized in the ML sense. Note that no closed-form solution can be obtained in eq. (4). In the multiple sources case, the computational complexity of the AML algorithm requires multi-dimensional search, which is much higher than the MUSIC type algorithm that requires only 1-D search. Various numerical solutions were proposed to obtain the AML estimate. These include the Alternating Projection (AP), Gauss-Newton (GN) and Conjugate-Gradient (CG). For detail derivation of these methods see [16].

IV. 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. In these tests, we deployed 6 nodes in several locations at RMBL where marmots are normally present. One of these, "Marmot Meadow", is shown in Fig. 8. We tested the performance of our CFAR-based 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 in the same field environment.

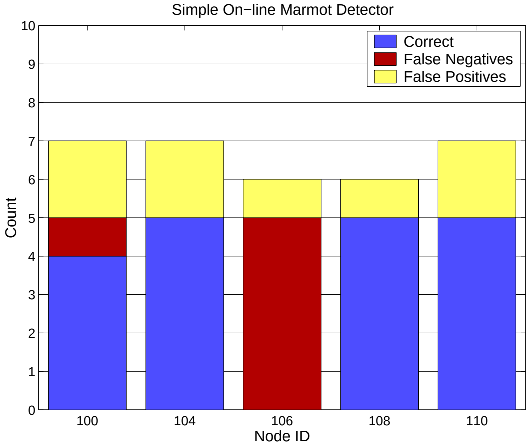


Fig. 9. Performance of the marmot detection algorithm from 20 minutes of audio. The deployment laydown was similar to the compact deployment described in Section IV-C. For each node, the bars show the number of correct detections, missed detections, and false positives. Note that if any node detects, all nodes will capture that call.

A. Marmot Detector Performance

To test the performance of the marmot detector, we ran the detection software on a network of 5 nodes. This software implemented the algorithm described in Section III-A, running in real time. Whenever a detection range 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 snippets 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. Fig. 9 shows the result of this comparison. Several nodes detected every call present, and a few nodes reported false positives. Node 106 was not functioning properly and only reported false positives. As we saw in Fig. 6(b), marmot calls are very loud and the detector had very little trouble identifying them. In fact, most of the false positives were introduced by researchers walking through the field manipulating the nodes.

B. 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 degrees relative to the coordinate system of the sub-array. The

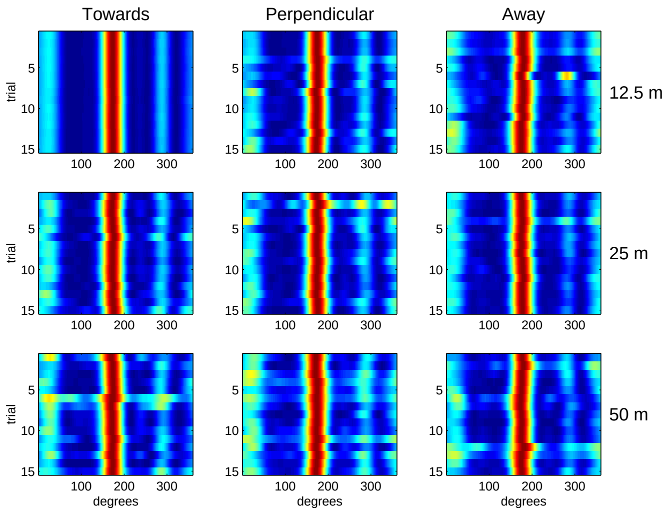


Fig. 10. Results from a set of controlled DOA experiments in the field. A speaker was placed on the ground in a meadow at a set of specified distances (12.5 m, 25 m, and 50 m), and oriented to face either towards, perpendicular, or away from the array. 15 trials were performed in each configuration. The results above show the 9 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.

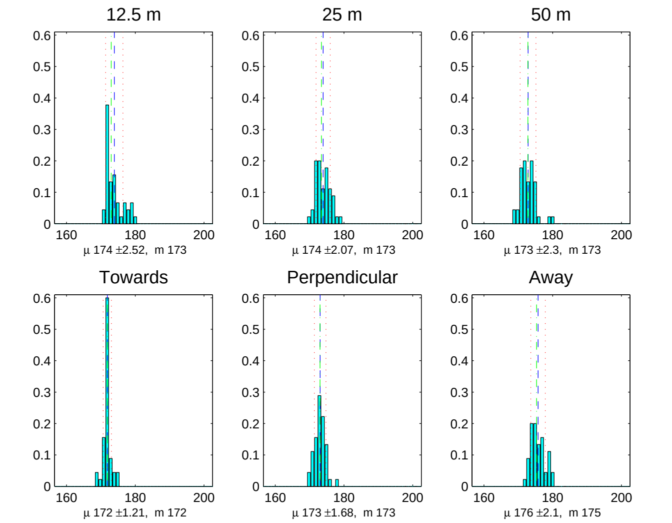


Fig. 11. Histograms of the maximum DOA estimates for the same data shown in figure 10, grouped by distance and grouped by source facing. Each plot shows mean $\mu \pm$ the standard deviation and the median m.

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 m, 25 m, 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 second section 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 shown in figure 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 II-B. 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 degrees. 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).

C. AML Localization in 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 combinations of these hardware and software operations are thus applied to real-life applications.

The setup of the array was as follows. Six sub-array nodes were spread over a region which surrounds the "Spruce Burrow", a location where marmots alarm-called. Fig. 12 shows the location of Spruce Burrow relative the sensor nodes. To align the position and orientation of the nodes relative to the true earth, we took GPS measurements on three nodes. Although two was sufficient, three enhanced our confidence. Fig. 12 was made by assuming node 110 position, and the orientation from node 110 to 108 are accurate. The GPS positions were then translated and oriented to fit the self localization positions based on these assumptions. On each line the top numbers is the distance based on GPS coordinate, and the bottom is based on self localization. The difference is well within the GPS accuracy in the open field, which are about 3 meters.

The position of the source is estimated using a pseudo-likelihood map. Each node runs the AML algorithm on the marmot calls to produce the bearing likelihood. Then, each position in a pseudo-likelihood map is generated by summing the bearing log-likelihood values that point to that position. The estimate was chosen based on the most likely position. Fig. 13 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. In this case we only know that a marmot was observed nearby the burrow. The least precise localization of the points in this figure is that of Spruce Burrow, because of difficulties getting good GPS locations in the woods. Nonetheless, the estimate of location in Fig. 13 is near the GPS measure we obtained at Spruce Burrow, and all the main lobes are pointing at the same directions. This suggests that the GPS measure we obtained actually is accurate. 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.

D. AML Localization, Wide Deployment

In this experiment, our goal was to investigate the limiting capability of the system by stretching the array as large as the wireless link allows while performing the same task.

The array setup was similar to the compact deployment experiment except the distance between nodes was much greater. The longest distance was from node 104 to node 106 (143.6 meters). In order to validate the self-localization results, we also used a tape measure to record the distance from nodes 100-104, and four GPS readings: node 100, 104, 108, and Spruce Burrow.

Fig. 14 displays the node positions according to both 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 meters. 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. However, 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 graph in Fig. 14 shows the distances between the GPS points, which can be compared to the distances between the points in the self-calibration solution. From this we see that while the distances 100-104 and 104-108 are within 1 meter, the distance 100-108 reports 10 meters shorter in the GPS data compared with the range measured from the acoustics. From this observation, we can form two alternate hypotheses. The first hypothesis is that the acoustic range measurement is long by 10 meters. This can happen in cases where the line of sight path is blocked and a reflected path is acquired. The alternate hypothesis is that the GPS position for either 108 or 100 is inaccurate, such that two legs of the triangle are relatively accurate while the third is wrong. In this case, the most likely culprit was 108, because of the presence of nearby trees which prevented the GPS unit from detecting the maximum number of satellites and possibly caused reflections.

To test these hypotheses, we analyzed the self-localization data more carefully. First, we re-ran the original data set and examined the residual values from the constraint system solution. This system is composed of 54 constraint equations, with a maximum range residual of 6.75 cm and an average residual value of 3.2 cm. Next, supposing that the GPS data is more accurate, we modified the source data by editing the range for 100-108 to match the distance derived from GPS. Re-running the system, we found that the modified system was highly inconsistent. In the new solution, the maximum range residual was 235 cm, and the average residual was 111 cm, with 6/10 node pairs registering large residuals. In addition, one of those high-residual pairs was 100-104, for which our tape-measure data point already corresponded well with the acoustic results. This result suggests that one incorrect acoustic range would not be sufficient to explain the discrepancy—more range errors and more nodes would need to be involved. Given that the original dataset achieved such a high degree of consistency, we determined that the GPS error was the more likely hypothesis.

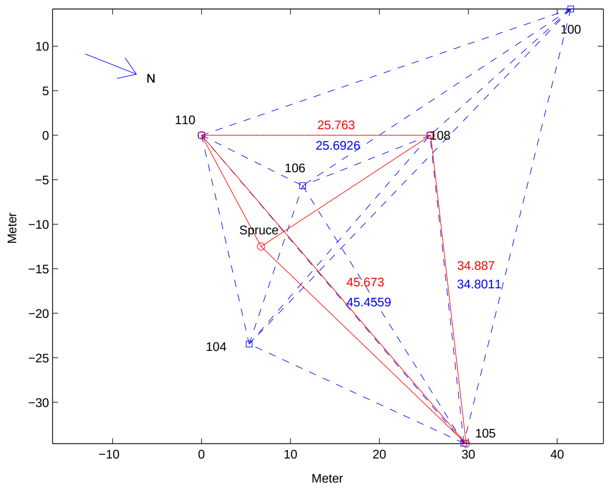


Fig. 12. Compact deployment comparison between node self-localization and GPS data. GPS data is denoted by round circles connected in a solid line.

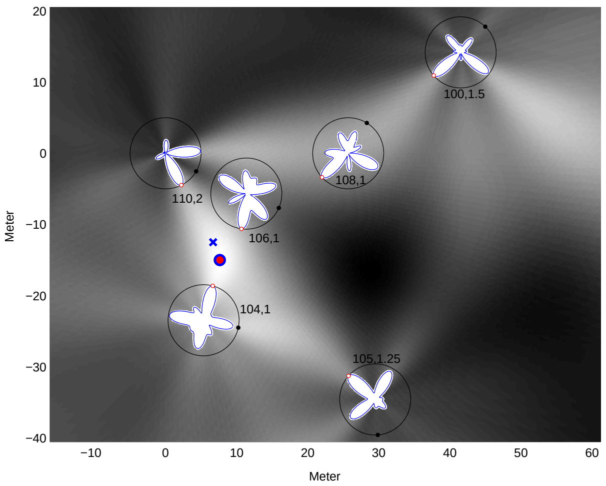


Fig. 13. A pseudo-likelihood map generated based on the log-likelihood of all nodes, taken from the compact deployment. Main lobes are denoted by small dark gray circle on the log-likelihood ring. Black dots points to the array zero degrees.

Based on this determination, we analyzed the rest of the data with the assumption that 100 and 104 were accurate and dropped the GPS data for 108. This analysis generated source estimates below and to the right of Spruce Burrow, as shown in Fig. 14. These results corresponded well with the reports from human observers at the site.

In this data set, the marmot was chirping in a rapid, cyclic mode, 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 consider this set of data points to indicate the distribution of errors we see from our algorithms.

To perform the estimation, we follow similar a procedure as before. We generate the pseudo-likelihood map for each chirp and use 0.1 sec duration in processing. Each estimated position is then collected and plotted in a scatter plot as seen in Fig. 15. The plots are arranged in time from dark to light as shown in the colorbar on the right hand side. The precision of the map is 0.1 meter, hence anything that falls within the 0.1 x 0.1 square is treated as the same location. Larger dot size indicates multiple estimates at that location. The mean and standard deviation is shown as the square and cross hairs, and the median is shown as triangle. The plot axis are set relative to node 104 and the mean (μ) and standard deviation (σ) are (82.1, 11.1)

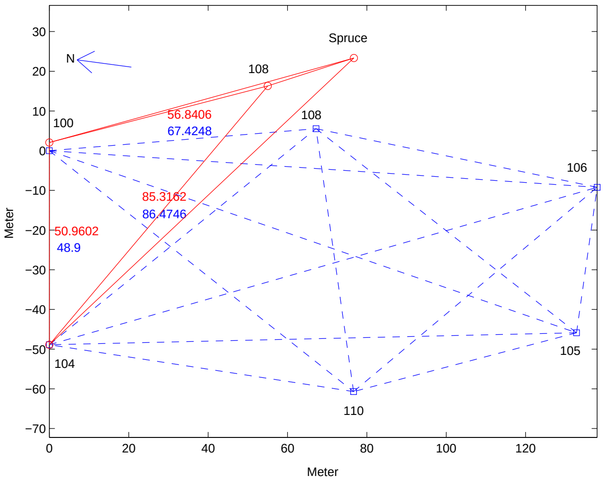


Fig. 14. Wide deployment comparison between node self-localization and GPS data. GPS data is denoted by round circles connected in a solid line.

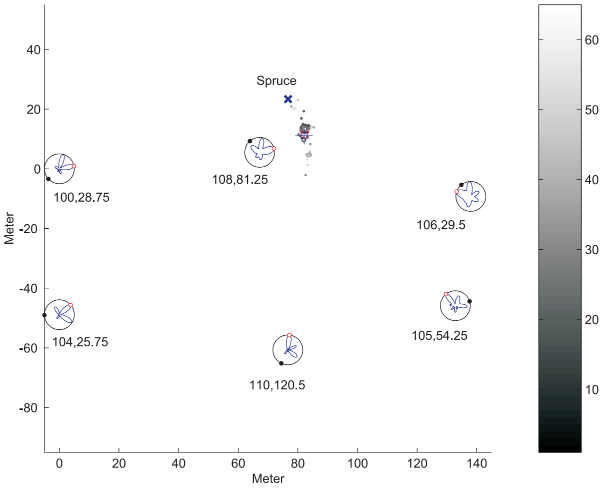
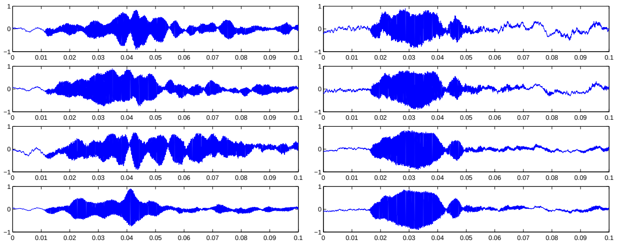


Fig. 15. Scatterplot of location estimates from the wide deployment, relative to the position of the node 104 using all six nodes. μ = (82.1, 11.1), and σ = (2.9, 5.6) meters.

meters and (2.9, 5.6) meters 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 likelihoods reported by these nodes have broad, mis-shapen lobes that introduced error into the position estimate. Fig. 16 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 reverberation and cancellation. We believe that this distortion is mainly caused by reflections 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. 17a, we show the results of our localization after removing data from 108 and 106. This results in a much tighter distribution, with mean (μ) (82.6, 14.2) meters and standard deviation (σ) (1.4, 2.8) meters. Clearly there is much benefit in detecting the reverberant conditions that yield poor results; we intend in future work to investigate methods for automatically selecting the nodes that minimize dispersion.



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

Fig. 16. A time domain segment containing a single marmot chirp from two nodes at different locations from the wide deployment.

Fig. 17b shows a zoomed-in version of Fig. 17a. 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").

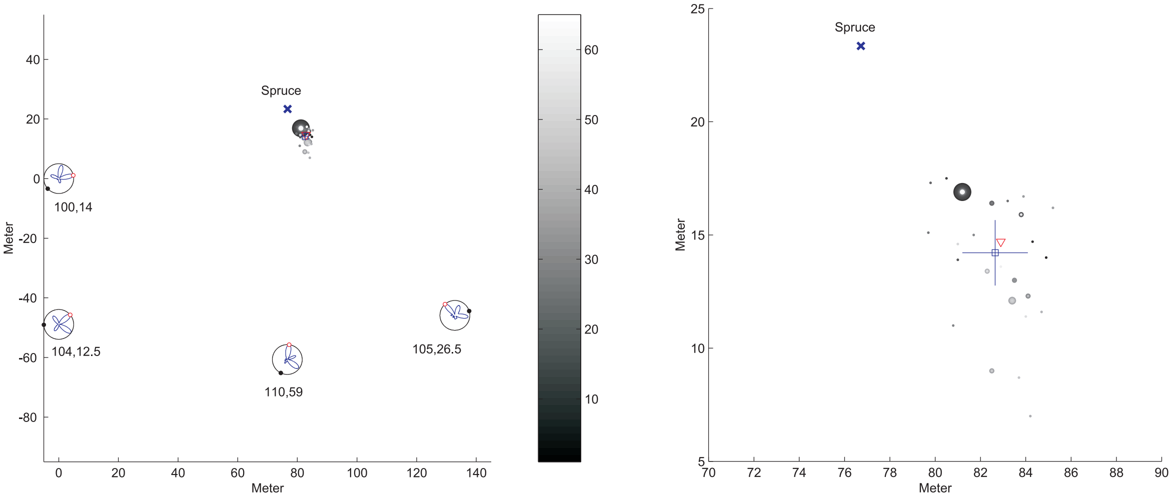
V. Conclusion

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 relatively quickly and deployed on the ENSBox nodes without excessive optimization. Demonstrating in-network data reduction, we showed that this detector could pre-filter 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 [17].

Our controlled experiments with 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.



(a) Relative to node 104.

(b) Zoomed in.

Fig. 17. Scatterplot of location estimates from wide deployment after removing node 108 and 106. μ = (82.6, 14.2), and σ = (1.4, 2.8) meters.

Finally, 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.

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, as well as studying tropical birds in the rainforests of Chajul, Mexico.

Acknowledgments

This work is partially supported by NSF CENS program under Cooperative Agreement CCR-012, NSF grant EF-o410438, UC Discovery grant sponsored by ST Microelectronics, a UCLA faculty research grant to DTB, and the MIT WaveScope project (NSF). The authors would like to thank Vlad Trifa and Martin Lukac for providing feedback on a draft of this paper. The authors would also like to thank RMBL for hosting us during the experiment, and to thank the marmots for their participation in this study.

References

[1] J. Bradbury and S. Vehrencamp, *Principles of animal communication*. Sinnauer, 1998.

[2] P. McGregor, T. Peake, and G. Gilbert, “Communication behavior and conservation,” in *Behaviour and conservation*, L. Gosling and W. Sutherland, Eds. Cambridge University Press, 2000, pp. 261–280.

[3] D. Blumstein, “The evolution of alarm communication in rodents: structure, function, and the puzzle of apparently altruistic calling in rodents,” in *Rodent societies*, J. Wolff and P. Sherman, Eds. U. Chicago Press, 2007.

[4] M. L. Hall, “A review of hypotheses for the functions of avian duetting,” *Behav. Ecol. Sociobiol.*, pp. 415–430, 2004.

[5] D. T. Blumstein and K. B. Armitage, “Alarm calling in yellow-bellied marmots:1. the meaning of situationally-specific alarm calls,” *Anim. Behav.*, vol. 53, pp. 143–171, 1997.

[6] D. T. Blumstein and O. Munos, “Individual, age and sex-specific information is contained in yello-bellied marmot alarm calls,” *Anim. Behav.*, vol. 69, pp. 353–361, 2005.

[7] D. T. Blumstein and J. C. Daniel, “Yellow-bellied marmot discriminate between the alarm calls of individuals and are more responsive to the calls from juveniles,” *Anim. Behav.*, vol. 68, pp. 1257–1265, 2005.

[8] L. Girod, M. Lukac, V. Trifa, and D. Estrin, “The design and implementation of a self-calibrating distributed acoustic sensing platform,” in *ACM SenSys*, Boulder, CO, Nov 2006.

[9] H. Wang, C.-E. Chen, A. Ali, S. Asgari, R. E. Hudson, K. Yao, D. Estrin, and C. Taylor, “Acoustic sensor networks for woodpecker localization,” in *SPIE Conference on Advanced Signal Processing Algorithms, Architectures and Implementations*, Aug. 2005.

[10] V. Trifa, “A framework for bird songs detection, recognition and localization using acoustic sensor networks,” Master’s thesis, École Polytechnique Fédérale de Lausanne, 2006.

[11] L. Girod, “A self-calibrating system of distributed acoustic arrays,” Ph.D. dissertation, Univerity of Caliornia at Los Angeles, 2005.

[12] J. Elson, L. Girod, and D. Estrin, “A wireless time-synchronized COTS sensor platform, part i: System architecture,” in *IEEE CAS Workshop on Wireless Communications and Networking*, 2002.

[13] L. Girod, J. Elson, A. Cerpa, T. Stathopoulos, N. Ramanathan, and D. Estrin, “Emstar: a software environment for developing and deploying wireless sensor networks,” in *Proceedings of the 2004 USENIX Technical Conference*. Boston, MA: USENIX Association, 2004.

[14] J. Chen, K. Yao, and R. Hudson, “Maximum-likelihood source localization and unknown source localization estimation for wideband signals in the near-field,” *IEEE Transactions on Signal Processing*, no. 8, pp. 1843–1854, 2002.

[15] I. Ziskine and M. Wax, “Maximum likelihood localization of multiple sources by alternating projection,” *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.

[16] L. Yip, J. Chen, R. Hudson, and K. Yao, “Numerical implementation of the aml algorithm for wideband doa estimation,” in *Proc. SPIE*, Aug. 2003, pp. 164–172.

[17] L. Girod, K. Jamieson, Y. Mei, R. Newton, S. Rost, A. Thiagarajan, H. Balakrishnan, and S. Madden, “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)*, 2007.