Acoustic localization of antbirds in a Mexican rainforest using a wireless sensor network

Travis C. Collier, a Alexander N. G. Kirschel, and Charles E. Taylor
Department of Ecology and Evolutionary Biology, University of California Los Angeles, 621 Charles Young Drive South, Los Angeles, California 90025

(Received 7 July 2009; revised 15 March 2010; accepted 9 April 2010)

Acoustic localization is a promising method to passively observe vocal animal species, but remains difficult and time consuming to employ. To reduce the labor intensity and impact of deployment, an acoustic localization system has been developed consisting of battery powered wireless sensor nodes. The system also has the ability to perform an acoustic self-survey, which compares favorably in accuracy to global positioning system survey methods, especially in environments such as forest.

The self-survey and localization accuracy of the system was tested in the neotropical rainforest of Chiapas, Mexico. A straight-forward and robust correlation sum localization computation method was utilized and is described in detail. Both free-ranging wild antbird songs and songs played from a speaker were localized with mean errors of 0.199 m and 0.445 m, respectively. Finally, additional tests utilizing only a short segment of each song or a subset of sensor nodes were performed and found to minimally affect localization accuracy. The use of a wireless sensor network for acoustic localization of animal vocalizations offers greater ease and flexibility of deployment than wired microphone arrays without sacrificing accuracy.

© 2010 Acoustical Society of America. [DOI: 10.1121/1.3425729]

PACS number(s): 43.60.Fg, 43.60.Jn, 43.60.Qv [WMC]

I. INTRODUCTION

Acoustic monitoring and localization provides a powerful tool for noninvasively studying vocal animals. As the capability and affordability of hardware increases, arrays of acoustic sensors are being developed and applied in terrestrial environments for applications including surveying birds and endangered frogs, studying duetting behavior of wrens, and group dynamics of elephants, monitoring invasive cane toads, and localizing woodpeckers.

In an environment such as the rainforest where visual observation of animals is very difficult, acoustic sensor arrays offer several advantages. Observations may be unattended, which avoids disturbing the animals being studied and allows the researcher to conduct other types of observations or activities simultaneously. Also, the entire “soundscape” is captured, allowing interpretation of behaviors within context. However, the rainforest is also a noisy environment with numerous practical challenges where it is difficult to run cables, perform careful surveys, rely on the global positioning system (GPS), or even rely on staying dry.

An acoustic localization system (ALS) has been developed using a wireless sensor network paradigm. This system differs from other terrestrial ALS because it consists of an array of weather-resistant, self-contained recording nodes which are wirelessly networked together to form the overall microphone array. This approach allows for much greater ease of deployment as well as adding additional capabilities such as remote monitoring and control. The ALS system also has an acoustic self-surveying facility which can determine node locations automatically without relying on GPS.

To evaluate the performance of this system, we deployed an array of eight nodes in the neotropical rainforest of Chiapas, Mexico and asayed the accuracy of self-localization, two-dimensional localizations of antbird songs played from a portable speaker (referred to as playbacks), and two-dimensional localizations of wild bird songs. Testing the ability of the system to localize actual birds singing from positions relevant to their typical behavior from the forest undergrowth was a particular focus.

II. METHODS

A. Study site and species

We focused our study on the Mexican Anthrush (Formicarius moniliger, MAT) at the Chajul Tropical Biology Station of the Montes Azules Biosphere Reserve in Chiapas, Mexico. The Mexican antthrush is a suboscine passerine in the family Formicaridiae whose range extends from the Southern Mexican states of Veracruz and Oaxaca eastwards to Central Honduras. It inhabits tropical rainforests, where it is terrestrial, foraging on the forest floor for invertebrates.

The Mexican antthrush is an attractive study system for this analysis because it sings mostly from the ground, allowing two-dimensional localization. Its song is fairly loud and simple, occupying a narrow frequency range which can be isolated using a bandpass filter to reduce the effects of ambient noise. It also often sings from relatively open parts of the forest where it can be observed singing, aiding in surveying its exact location while singing.
B. The acoustic localization system

The acoustic localization system developed is an array of self-contained recording nodes. Each node has a sub-array of four microphones spaced 16.97 cm apart in a tetrahedral configuration as shown in Fig. 1. Nodes also contain an embedded computer, a wireless 802.11 network interface, a GPS, an acoustic emitter, and a smart lithium ion battery pack and is packaged in a weather resistant Pelican 1150 case 25 cm by 20 cm by 12 cm in size.

The wireless network allows the system to be monitored and controlled remotely from a laptop or hand-held device such as a smart-phone. It also allows the nodes to maintain time synchronization on the order of 10^-9 s without relying on a central timing source. The four microphone sub-array allows each node to estimate the direction of arrival (DOA) as well as time of arrival (TOA) of sounds. The node locations can be determined automatically using an acoustic self-survey system. This self-survey is more accurate and reliable than is sometimes available from GPS, especially in forested or mountainous environments and also solves for the orientation of each node.

The ALS system allows for cooperative event detection and near real-time processing. However, for this deployment each node was set to continuously record four channel audio data which was saved internally to a compact-flash card and analyzed off-line.

C. Field methods

1. Deployment

From December 9 to 14, 2008, an array of eight nodes was deployed in a fairly flat area of primary tropical rainforest near the field station around a section of the Circuitos Norte Este trail as shown in Fig. 2.

The nodes were placed on tripods approximately 1.5 m above ground level and positioned opportunistically about 15–20 m from the nearest other nodes. The mean distance between nodes was 39 m. A more comparable metric is array density, computed as the number of nodes divided by the area enclosed within convex hull of the array. Localization accuracy rapidly degrades as sources move outside this convex hull for time difference of arrival (TDOA) methods. This deployment of 8 nodes enclosed an area of 1855 m^2 for a density of 43.13 nodes/ha.

Recording sessions were started just after dawn and continued for about 2 h. Then an acoustic self-survey was performed, after which the nodes were taken back to the field station to be recharged and to copy the recordings from them. The tripods were left in place so subsequent self-surveys at the same location would produce replicated results. On several occasions a second set of recordings was

\[FIG. 1. A recording node deployed in the primary forest showing the arrangement of microphones.\]

\[FIG. 2. Overview of the deployment. Recording nodes are denoted by squares and acoustic source locations by circles. Trails are shown by the gray lines in (a), the playback locations along the trail are numbered and denoted by \(\oplus\). The nodes used in the 4 node subset analysis are shown by \(\blacksquare\). In (b), positions of surveyed wild bird songs are shown by \(\circ\). The Circuitos Norte Este trail extends from north to south with a small side trail extending from the center of the figures to the south-east.\]
done in the late afternoon before removing the nodes. Each node recorded 4 channels at 48000 16 bit samples per second.

During recording sessions, playbacks were used to elicit responses from targeted species. The locations of birds seen singing were flagged and the time noted. Birds were mist netted and color banded to enable identification of individuals.

2. Playbacks for localization error analysis

To assess localization error, bird songs from several species which typically occur at the field site were repeatedly played back through a portable speaker from six different locations across the deployment as shown in Fig. 2(a). Playbacks were standard single channel uncompressed audio files (16 bit at 44.1 kHz) from a portable player. The source sounds for the playbacks were songs from individuals recorded on previous trips to this site using a directional microphone and normalized. The MAT songs used were from individuals previously caught, color banded, and genetically sexed. The MATs whose songs were included in this study were those color banded and labeled as KRG, YYG, KYC, GGR, and GWW.

At each playback location, the speaker was facing down the trail roughly SSE. Eight individual songs from three different species of antbirds native to the site were used. In addition to MAT songs, songs of the Dusky Antbird (Cercomacra tyroamina, DAB) and Great Antshrike (Taraiba major, GAS) were included. Each song was repeated 26 times, except the GAS song which was repeated 20 times, for a total of 202 songs from each playback location. Due to heavy rain, playbacks from location 6 [Fig. 2(a)] were cut short and include only 17 repeats of MAT GGR and omit MAT GWW.

The composition of these playbacks is shown in Table I and spectrograms of each song are shown in Fig. 3.

3. Ground truth measurement

Since GPS availability and accuracy at the field site was severely limited, ground truth positions of the nodes, playback locations, and observed bird singing locations were determined using range and bearing measurements. Range measurements were taken using a laser range meter and bearing measures with a magnetic sighting compass. Range and bearing measures were taken from two tiers of survey points, a primary survey point assigned position (0,0) and secondary survey points at each node location.

A map was produced from the ground truth range and bearing measures using an iterative multilateration procedure similar to the multilateration system used in the automated self-survey. Initial position estimates were calculated using the minimum number of measurements from the most connected point. Each position was iteratively refined using a non-linear maximum likelihood fit minimizing the weighted sum of range and bearing residuals from all measures for that position. The iterated refinements were stopped when variation in the solutions stabilized below 10 cm.

The ground truth map is restricted to 2D due to inaccuracies attempting to measure declination or offset along the Z axis in the field. Most point to point ground truth measures were only possible through a small window in the vegetation. The overall layout of nodes was very close to planar (<2 m difference in Z over 80 m in X-Y). Therefore, accurate localization results are restricted to the X-Y plane and are not sensitive to small variations in Z as expected theoretically and shown experimentally in a similar context.

4. Acoustic self-survey

The ALS system has the ability to perform an acoustic self-survey. Each node emits a ‘ranging chirp’ sound in turn which is detected by the other nodes. Range and bearing measurements between nodes from repeated cycles of ranging chirps are then used to solve for node positions and orientations.

The ranging chirp is a 0.11 s pseudo-random sequence which is readily detectable and timed using a filter and correlation method. The node emitting the chirp also detects its own chirp, and transmits via radio the precise time the chirp was emitted. Range is determined by the time of flight of the ranging chirp, and bearing from the delays between the four microphones at each node.

<table>
<thead>
<tr>
<th>Species</th>
<th>Sex</th>
<th>Individual</th>
<th>Duration (s)</th>
<th>Peak frequency (Hz)</th>
<th>Repeats</th>
<th>dB (1 m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAB</td>
<td>♀</td>
<td>Unknown</td>
<td>1.53</td>
<td>2838</td>
<td>26</td>
<td>87.6</td>
</tr>
<tr>
<td>DAB</td>
<td>♂</td>
<td>Unknown</td>
<td>1.72</td>
<td>2858</td>
<td>26</td>
<td>84.2</td>
</tr>
<tr>
<td>MAT</td>
<td>♀</td>
<td>KRG</td>
<td>1.79</td>
<td>2105</td>
<td>26</td>
<td>85.8</td>
</tr>
<tr>
<td>MAT</td>
<td>?</td>
<td>YYG</td>
<td>2.16</td>
<td>2099</td>
<td>26</td>
<td>84.6</td>
</tr>
<tr>
<td>MAT</td>
<td>♂</td>
<td>KYC</td>
<td>1.99</td>
<td>1990</td>
<td>26</td>
<td>86.0</td>
</tr>
<tr>
<td>MAT</td>
<td>♂</td>
<td>GGR</td>
<td>2.52</td>
<td>1985</td>
<td>26</td>
<td>87.8</td>
</tr>
<tr>
<td>MAT</td>
<td>♂</td>
<td>GWW</td>
<td>2.02</td>
<td>2067</td>
<td>26</td>
<td>85.0</td>
</tr>
<tr>
<td>GAS</td>
<td>♂</td>
<td>Unknown</td>
<td>3.87</td>
<td>1100</td>
<td>20</td>
<td>86.6</td>
</tr>
</tbody>
</table>
The range and bearing measures along with a quality metric for each measure are fed to a centralized multilateration algorithm which solves for X, Y, and Z positions as well as azimuth, zenith, and yaw angles of each node. During setup, each node is leveled using a spirit level which minimizes the zenith and yaw angles. This results in the coordinate system of the self-survey results also being leveled.

The self-survey facility was improved for this deployment by adding a phase transform (PHAT) pre-whitening filter\textsuperscript{21} to the ranging chirp filter-detector. The quality threshold required for a value to be included in the multilateration solve was also lowered compared with previous results reported for more open environments.\textsuperscript{13}

Experiments with the previous version of the hardware in a temperate forested environment showed an average self-survey positioning error in X-Y of 9 cm with a maximum error of less than 20 cm.\textsuperscript{13} Error along the Z axis for a close to planar arrangement was slightly less than 0.5 m.

Self-surveys using ten rounds of ranging chirps were performed on five days. All localizations reported here were computed using the results of these self-surveys for microphone positions, not the ground truth measurements.

E. Registering self-survey with ground truth

The self-survey position results are registered to geographic coordinates by scaling and rotating the self-survey results to fit ground truth measures. Only a few such measures are required, though here we used all available.

By default, the self-survey facility assumes a value of 340 m/s for the speed of sound and produces position results in a coordinate frame anchored on one of the nodes. Small variations in the speed of sound from the assumption can be accounted for by scaling the self-survey position results. For a ground truth range between two nodes \( r_{ij} \) and a distance between nodes measured from the self-survey position results \( \hat{r}_{ij} \), the scaling factor equals \( \frac{\hat{r}_{ij}}{r_{ij}} \). The median of all scaling factors (one for each ground truth distance measure available) is used to scale the self-survey positions.

Similarly, the rotation around the Z axis of the self-survey map is determined from ground truth bearing measures \( b_{ij} \) between pairs of nodes and the values computed from the self-survey results \( \hat{b}_{ij} \). The self-survey results are rotated Z by \( \text{median}(b_{ij} - \hat{b}_{ij}) \).

Finally, the self-survey results are translated in X, Y, and Z to best fit the ground truth map minimizing the squared residuals. This translation is the only step using the solved ground truth map as opposed to individual ground truth range and bearing measurements. Alternatively, this translation could be performed to align the self-survey results with one or more fixed points such as GPS coordinates taken at a node.

The speed of sound can be determined from the scaling factor. For example, on 2008–12–12 the scaling factor was 1.01 implying a speed of sound of 343.4 m/s. Computing from recorded temperature and humidity values (18.5 °C and 88.5%RH using a Lascar EL-USB-2-LCD datalogger) gives a value of 343.49 m/s.\textsuperscript{22} These values are equal within the variation due to the 0.5 °C precision of temperature measurements.

F. Localization computation

The array of 8 nodes produced a total of 32 tracks of synchronized audio data. From this data, we solved for source location in the laboratory using a variation of the correlation sum\textsuperscript{3,23}–TDOA method (also known as accumulated correlation).

Signals were filtered using a Chebyshev type 1 IIR bandpass filter.\textsuperscript{24} For MAT and DAB songs, the band between 1.5 and 3.5 kHz was passed. For the GAS song, the bandpass filter range was 0.75 to 3 kHz. The speed of sound \( v \) is determined by the scaling factor returned by the self-survey registering process. From the self-survey positions and orientations of the nodes, the distance \( r_{ij} \) between each pair of microphones \( i \) and \( j \) is computed.

Correlation sum involves two basic steps: computing cross-correlations between tracks and determining the most probable source location.

1. Cross-correlations

Call the set of all tracks \( T=[1, \ldots, 32] \). The sound of interest is identified in a single track \( k \) and start time \( t_k \) and duration \( s_k \) are noted. This segment of signal \( X_k \) is used as the “key signal.”

1. \( k \) is removed from \( T \).
2. For every track \( i \) in \( T \)
   (a) The corresponding segment \( X_i \) starting from \( t_i - d_{ij} / v \) of length \( s_k + 2(d_{ij} / v) \) is extracted.
   (b) \( X_k \) is zero-padded to the length of \( X_i \) and the cross-correlation function \( C(l)_{ik} \) between \( X_k \) and \( X_i \) is computed.
   (c) Additionally, the Hilbert envelope \( \hat{C}(l)_{ik} \) of \( C(l)_{ik} \) is computed.
3. The track with the largest maximum \( C(l)_{ik} \) value is selected as \( k \) and the time delay (lag) for that value as \( l_{\text{max}} \).
4. \( t_k \) is set to \( t_k + l_{\text{max}} \) and the process is repeated again from 1 until \( T \) is empty.

This process results in computing the cross-correlation and the envelope of the cross-correlation between every pair of tracks. The track with the highest cross-correlation value to the previous key signal is selected as the next key signal and masked in time assuming the lag at the correlation peak is correct. This masking and the bandpass filter enhance robustness to the presence of other sounds.\textsuperscript{3}

2. Source location determination

For a point in space \( P \), the 3D Euclidean distance between microphone \( i \) and that point is \( r_{ij} \).

The correlation sum at point \( P \) is

\[
S(P) = \sum_{ij} C_{ij} \left( \frac{r_{ij} - r_{ij}}{v} \right)
\]

Collier et al.: Localizing antbirds with wireless acoustic sensors 185
where the sum is over each cross-correlation computed for a pair of microphones $i,j$. The most probable source location is determined by searching $S(P)$ for the maximum value.

This is equivalent to determining the point which maximizes the energy output of delay-and-sum beamforming, but has two drawbacks in practical application due to spatial aliasing of signals which have a wavelength longer than twice the distance between sensors. The function $S(P)$ has many local maxima spaced tightly together and the function is very sensitive to errors in the relative locations of the microphones. To address these problems, the Hilbert amplitude envelope is used in place of $C_{ij}(l)$ when $i$ and $j$ are on different nodes. Another common approach is to compute the cross-correlations between spectrograms instead of actual signals. For the frequency ranges of interest here, the Hilbert envelope is chosen since it potentially provides higher accuracy than spectrogram cross-correlations and does not require choosing additional parameters.

$S(P)$ is evaluated over a regular 2D lattice of points covering a predefined area being monitored at a fixed Z value. The resolution of the lattice was 0.25 m. For the playbacks, the Z value chosen was 0 corresponding to the approximate plane of the nodes. For localizations of actual birds, the Z plane was set to −1 m which is just above ground level. The monitored region covered the entire area shown in Fig. 2. After determining the point with the maximum $S(P)$ value over the 0.25 m lattice, a second refinement search was performed over the $2 \times 2$ m$^2$ area around that point at a resolution of 0.01 m. The $P$ value maximizing $S(P)$ in this refinement search was the final localization result.

III. RESULTS

A. Self-survey accuracy

The acoustic self-survey system produced positional results which differed from ground truth by $0.155 \pm 0.091$ m in the XY plane ($n=40$, mean ± standard deviation for all nodes on all five days self-survey was performed). The self-survey range measures only differed by $0.026 \pm 0.022$ m ($n=48$) from the ground truth range measures, and the error in bearing estimation was $0.737^\circ \pm 0.721^\circ$ ($n=48$).

These errors are higher than previously reported X-Y errors of 0.05 m in an open courtyard and 0.09 m in a temperate forest. However, this is unsurprising as the rainforest environment is much less open.

B. Localization accuracy

Four sets of analyses were performed to assess the accuracy of localizations. The first three of these are playback experiments from the six numbered locations shown in Fig. 2(a). These playback experiments are: (1) localizations using the full duration of each playback song; (2) using only the first 0.2 s of each playback song; and (3) using the full song durations but only the four nodes denoted by in Fig. 2(a). Finally, (4) we localized 10 songs of wild MAIs which were observed while singing. The positions of these songs were surveyed and are shown in Fig. 2(b).

A one-way repeated measures analysis of variance (ANOVA) shows no significant difference between the localization errors of the four sets of analyses ($F_{3,147}=1.092$, $p=0.327$). Each analysis is presented in more detail below and the overall results are summarized in Fig. 4. All errors are the Euclidean distance in 2D (X-Y) between the ground truth surveyed point and the solved localization. Statistics were computed using R.

1. Full playback songs

The overall mean positional localization error for the playback points using full duration of the signals was $0.199 \pm 0.064$ m ($n=1179$). The median error value was 0.205 m. The localization error is only slightly larger than the self-survey error, which is the floor of what is achievable.

The distribution of localization errors for each different source signal and each different playback location is shown in Figs. 5(a) and 5(b). The variation in mean error between different source sounds is less than 0.1 m.

A two-way repeated measures ANOVA computed from a linear mixed effects model fit shows a significant difference
in localization error for different playback locations \( F_{5,34} = 3.4335, p = 0.0128 \), but no significant difference between different source signals \( F_{5,34} = 0.1515, p > 0.99 \). Post-hoc Tukey contrasts using the “honest significant difference” method\(^\text{30} \) reveal that the only significant pairwise comparison between playback locations is location 6 vs. 3 \( (z = -3.927, p = 0.0012) \). The smaller localization errors for playbacks at location 6 are likely due to that location being in a relatively open area with line of sight to five recording nodes. This reinforces the conclusion that the results presented here are not a best-case scenario and that better accuracies are possible under more favorable conditions.

### 2. Localizing short segments of playbacks

It is sometimes desirable to perform localizations on only a part of a song when there is interfering noise or another song overlapping it. Duetting is one such case, where in species such as the DAB and the rufous-and-white wren \((Thryothorus rufalbus)\) one participant’s song may overlap the end of the other’s song.\(^\text{3,1} \)

To test the accuracy of the system on short segments of songs, localizations of the playbacks were also done using only the first 0.2 s of the signals. The positional error was 0.214 ± 0.151 m \((n=1179)\) with a median value of 0.188 m. The overall mean error is only 7.5% larger than for localizations using the entire signal duration. Increased sensitivity to other noises likely contributes to the larger variance and long tail of the error distribution of short segments relative to localizations performed on the full songs seen in Fig. 4.

### 3. Localizing playbacks using a subset of nodes

It may be desirable to monitor a larger area with the same number of nodes or use fewer nodes to monitor the same area. Localizations using only four of the nodes were done to assess the effects of a more sparse deployment. This was achieved by using only the four nodes denoted by ■ in Fig. 2(a) for the localization computation and keeping all else the same as the full song playback analysis.

The mean distance between nodes for the four node subset was 42.108 ± 12.749 m with a median value of 43.027 m. These four nodes enclosed 1228.3 m\(^2\) within the 2D convex hull for a deployment density of 32.56 nodes/ha. A point of comparison is a wired 24 microphone array used to study lekking of Greater Sage-Grouse \(^\text{32} \) \((Centrocercus urophasianus)\) which has a density of 36.25 nodes/ha.

The localization error was 0.244 ± 0.220 m \((n=1179)\) with a median value of 0.196 m. The mean error is 10.6% larger than the results using all 8 nodes. Fig. 4 shows an increased tail of larger errors compared with the full song localizations.

### 4. Localizations of wild birds

The singing locations of actual birds in the study area were noted and surveyed when possible. Ten MAT singing locations were observed. The ground truth survey positions of these songs are shown in Fig. 2(b).

The localizations all 10 wild MAT songs differed from the surveyed positions by a mean of 0.445 ± 0.500 m, with a median error of 0.258 m. The mean localization error of wild MAT songs compared to the playbacks is only slightly higher and not statistically significant from the three playback analyzes \( F_{3,147} = 1.092, p = 0.327 \).

Additional tests comparing the localization error of just the full song playbacks to the wild bird songs give the same results. The full playback error data is first collapsed to eliminate pseudo-replication by taking the mean error for each unique combination of playback location and source signal. Both a Welch two sample t-test \( t_{9,667} = 1.5677, p = 0.1511 \) and a Kruskal-Wallis rank sum test \( \chi^2 = 3.1805, p = 0.07452 \) comparing the collapsed full playback and the wild MAT localization errors show no significant difference.

### C. Illustrative application: Tracking birds over time

For animals which vocalize frequently, acoustic localization may be used to track the movements of individuals even when they cannot be seen. Fig. 6 shows an example of this application. The localization results of 20 songs over about 160 s from two different individual MATs are shown. Calls were assigned to each bird by connecting points which are closest in time and space. Each bird had distinctive features of their songs which allowed them to be identified to verify the song assignments. These two birds were singing in response to playback stimulus of a male MAT outside the figure to the south-west.

### IV. DISCUSSION

The acoustic localization system provided accurate localization results for both playbacks (0.199 m error) and wild bird songs (0.445 m error) in the rainforest without the use of GPS. The difference in localization accuracy of playbacks vs. wild birds was not statistically significant, which is reflected more clearly in the median errors of 0.205 m and 0.258 m respectively. MATs are terrestrial birds which sang...
from the ground, often near or in undergrowth. These are difficult situations for acoustic localization due to reflections and reverberations. These effects can be seen in the longer tail of the localization error distribution of wild bird songs compared with the full playbacks. Still, the results are quite reliable with all errors being less than 0.9 m except a single error of 1.70 m, which was a song outside the array, in undergrowth, and on a slope [the southern most point in Fig. 2(b)]. The system also produced reliable and accurate results localizing short 0.2 s segments of playback songs and when only a four node subset of the array was used. These demonstrate the applicability of the system for studying duetting (e.g., in DABs) and for use in less dense deployments.

The error estimates presented here are conservative since they assume the surveyed locations of playbacks and singing birds were the true locations. The accuracies achieved compare favorably with those reported for several other microphone array ALS. These include localization errors from sources within the arrays applied to localizing playbacks of wrens in tropical forest (2.82 m), great tit song playbacks in deciduous woodland (0.281 m), and songbirds in a re-grown agricultural field (0.82 m).

Direct comparison between different ALS may be misleading. Each node in the system presented here contains four closely spaced microphones, which adds a level of redundancy to localization computations not present in the other systems cited. The distance between nodes also varies between different systems and different deployments of the same system. On the other hand, the songs localized here were narrow-band and relatively unimodal, which have been found to be more difficult to localize than frequency modulated songs. The results presented here were in the presence of normal rainforest background noise and not individually adjusted or optimized. The only pre-filtering performed was a fairly permissive bandpass filter.

A large component of the increased localization accuracy of the system is likely due to the high accuracy of the acoustic self-survey facility. Localization accuracy is fundamentally limited by the accuracy of microphone position measurements. Determining microphone positions using GPS can be unreliable in many environments such as dense forest and mountains, and surveying using theodolites or total stations can be impractical. Both accuracy and the ability for rapid deployment were greatly facilitated by the acoustic self-survey capacity of the system. The self-survey error was 0.155 m, which contrasts to the 1.31 m accuracy for GPS in a similar environment reported by Mennill et al. despite their “extensive surveying across multiple years.” Advances in GPS receiver technology may mitigate this somewhat in the future, but even when GPS is available, the acoustic self-survey provides an independent measure and an estimate of the speed of sound.

The localization computation used here was a straightforward correlation sum method. This method follows the principle of least commitment and does not require the “tweaking” of many parameters to achieve good results. Using the amplitude envelopes of the cross correlation functions eliminates the requirement that the sounds being localized are low frequency, and trades some theoretical precision to eliminate ambiguity and increase reliability. It naturally extends to 3D, but it is computationally intensive and scales as the square of the number of microphones used. Therefore, correlation sum is not generally suitable for real-time application. However, this is not a primary concern for many behavioral biology and ecology studies.

Other localization computation methods have successfully been demonstrated with the ALS hardware used for this study, and the method presented here may be applied to other localization systems including wired microphone arrays. A more computationally efficient approximate-maximum-likelihood direction of arrival (AML-DOA) method has been used to localize playbacks of antbird and woodpecker songs, as well as marmot alarm calls. This method has been used with the ALS system and run in-situ to provide near real-time localizations in the field. Additionally, there exist promising methods which can efficiently estimate locations and additional parameters such as variations in speed of sound without linearization.

The self-contained wireless nature of the ALS nodes made the deployment of the system faster and easier than would be required for system of microphones wired to a central digitizer which typically require several people hours to setup and survey. Generally, wired systems remain a better choice for long term fixed deployments, but can be impractical for shorter term studies or locations without the appropriate infrastructure. A new version of the ALS hardware is currently in development focusing on reducing cost, improving battery life, reducing size, and supporting larger inter-node spacing. The authors anticipate that this platform can be made widely available to researchers to facilitate the addition of acoustic localization to the toolbox of standard observation methods.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under Grant No. 0410438. The authors would like to thank L. Girod and the Center for Embedded Network Sensing at UCLA who were instrumental in developing the ALS system. We also thank Z. Harlow, E. Vallejo, the Chajul Tropical Biology Station for assistance in the field, and G. Patricelli and A. Krakauer for detailed information on their microphone array system.

---
