Semi-automatic long-term acoustic surveying: A case study with bats

Andreassen, Tórur*,† Surlykke, Annemarie† Hallam, John‡

Abstract

Increasing concern about decline in biodiversity has created a demand for population surveys. Acoustic monitoring is an efficient non-invasive method, which may be deployed for surveys of animals as diverse as insects, birds, and bats. Long-term unmanned automatic monitoring may provide unique unbiased data from a whole season, but the large amount of data presents serious challenges for the automatic processing of the measurements. To demonstrate feasibility of automatic multi-channel surveying using new prototype hardware, we carried out a two-month study of echolocating bats requiring high data sampling rates (500 kHz). Using a sound energy threshold criterion for triggering recording, we collected 236 GiB ($\mathrm{Gi} = 1024^3$) of data at full bandwidth. We implemented a simple automatic method using a Support Vector Machine (SVM) classifier based on a combination of temporal and spectral analyses to classify events into bat calls and non-bat events. After experimentation we selected duration, energy, bandwidth, and entropy as classification features to identify short high energy structured sounds in the right frequency range. The spectral entropy makes use of the orderly arrangement of frequencies in bat calls to reject short noise pulses, e.g. from rain. The SVM classifier reduced our dataset to 162 MiB of candidate bat calls with an estimated accuracy of 96% for dry nights and 70% when it was raining. The automatic survey revealed calls from two species of bat not previously recorded in the area, as well as an unexpected abundance of social calls. The recordings provide data which can be used to correlate bat activity with rain, temperature, and sunset/sunrise. We discuss future applications, achieving higher accuracy in classifying bat calls and the possibility of using trajectory-tracking data to determine bat behavior and correct for the bias toward loud bats inherent in acoustic surveying.

*Corresponding author: thor@biology.sdu.dk

†Institute of Biology, University of Southern Denmark, Campusvej 55 5230 Odense M Denmark

‡The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Campusvej 55 5230 Odense M Denmark

1 Introduction

In the past decade, worries about decline of biodiversity have fostered growing interest in technology for environmental monitoring, using surveys of animals as diverse as insects, birds and bats, where one can automate species recognition based on acoustic communication [1]. For bats and birds, special concerns have been linked to the growing number of large scale wind turbine parks [2], making large spatial and temporal scale surveillance particularly interesting.

Bats are protected and some species are endangered; loss of foraging grounds and roosting places poses a serious threat to their welfare [3]. Major new building projects not only disturb local bats during construction, but may also add to the decline in suitable habitat. The University of Southern Denmark is undergoing major re-construction with plans for tripling the built-up area and introducing substantial changes to the surroundings. We are interested to know what effect such changes may have on the local bat populations.

Many methods of surveying are currently in use: carcass counting [4, 5], radar tracking [6, 7], infrared thermography tracking [8] and passive acoustical measurements, usually with one or more bat-detectors [9, 10]. Each of these methods has some drawbacks: carcass counting is done too late and does not reveal the bats' flight behavior; radar tracking and stereo infrared thermography can give suitable data, at least for large species, but availability and price make them infeasible for wide deployment; multiple bat-detectors could, with some effort, be used to count passing bats automatically and estimate their species — but unless they are synchronized one cannot easily combine information from the independent detectors, for example to localize an individual bat precisely relative to the recording sites. Thus a synchronized multi-channel solution is to be preferred. Furthermore, to obtain a fair picture of local bat activity one needs to survey over a long period, ideally continuously.

Acoustical surveys, especially ultrasonic surveys, are limited today by the available equipment. This problem concerns both the area that can be surveyed and the duration over which the survey can be performed. To ameliorate these shortcomings, we have developed an inexpensive scalable array recording system using commodity-computing components [11]. The system can be deployed for months while automatically recording and saving data on ultrasonic signals in its vicinity.

The goals of this study are thus three-fold:

1. to demonstrate that the new automatic array recording system can be deployed long-term in a real bat-surveying application;

2. to verify that the voluminous data thereby collected can be analyzed effectively, without unreasonable demands for human judgment — this is the key technical challenge of the proposed approach to surveying; and

3. to investigate and document local bat population behavior at a site in the vicinity of the University of Southern Denmark.

Since our primary goal (goal 1) is demonstration of the feasibility of conducting automatic acoustic surveys, we have not spent effort on optimizing all aspects of the survey: our primary goal is to demonstrate the adequacy and potential of the new hardware. Nevertheless, the survey carried out yielded novel and interesting biological observations. We demonstrate that the long-term multi-channel recording technology is effective, that the data processing challenges it raises can be answered effectively, and we note that it has the potential, beyond what we show in this paper, for some very interesting capabilities: bat flight-paths can be recovered, multiple bats can be separated in the recordings, we would be able to determine number of specie per volume per time.

In this paper we only consider acoustical surveillance with principal focus on automated long-term multi-channel recordings. This approach allows us to record bat activity over months, and to correlate bat behavior over time to several parameters including change of season, day/night, sunset/sunrise, lunar luminance, and weather data. One can also study the distribution of social versus echolocating calls, and in principle identify sequences which predict particular behaviors (such as foraging) and see how these are distributed over time. There is some evidence that monitoring the local bat population can work as a bioindicator [12].

2 Materials and Methods

To assess the function of the recording array, we chose to carry our a biologically relevant survey and use this as a vehicle for obtaining a good quality estimate of the recorded acoustic data. Our technical objective was to discover and classify as many bat calls as possible in the dataset, with a high accuracy and a low false positive rate, without trying to optimize the data processing more than was necessary to make feasible data inspection by bat biologists.

2.1 Study site

Our goals were both to investigate local bat activity but also to exercise and test the novel prototype recording array, so convenient access was important. Therefore we placed the system at a site on the Odense campus of the University of Southern Denmark (55°21'59.63"N, 10°25'54.10"E) with low anthropogenic activity, high bat activity, power grid access, and convenient physical access. The campus is located outside the city of Odense in Denmark in an open area with patches of old forest, large open fields and small streams. The array was placed on the roof of a garage building, about 5 meters off the ground, close to a small old forest where we had often observed bats of various species flying along the tree-line and inside the forest.

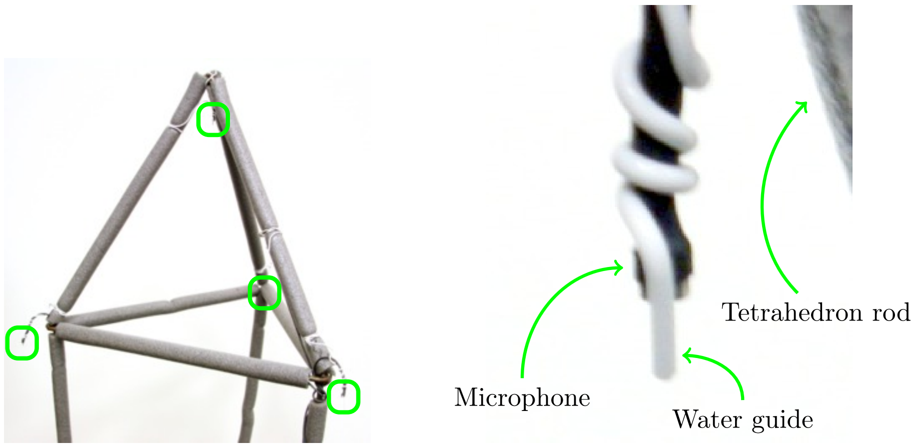


Figure 1: **Left:** The microphone assembly: a tetrahedron mounted with four Knowles microphones with water-guides. The microphones are marked by green rounded rectangles. **Right:** A close-up of the topmost microphone, showing its water-guide (the wire that projects below the bottom of the microphone).

2.2 The acoustic recording system

Four miniature microphones, type FG23329–PO7 (Knowles Acoustics, Itasca, Illinois, USA), were mounted on a tetrahedral structure with 30 cm sides. The tetrahedron can be seen in the left image of figure 1. The omnidirectional microphones were directed downwards and protected by a water-guide (right image in figure 1). The recording system comprises a standard PC running GNU/Linux equipped with a PCI-bus analog-to-digital board DAQ-2010 (Adlink Technologies inc., Taipei City, Taiwan). This board is capable of simultaneously recording 4 analog channels with a sample rate of 2 MHz per channel. The microphones interface to the analog-to-digital converter board via custom-made signal-conditioning and programmatically controllable amplification electronics. The recording system components fit comfortably in a sturdy IP65 compliant plastic box (Fibox type EK-W). Power to the system was supplied directly from the grid to avoid changing batteries. The system was configured to record four channels with a 500 kHz sampling rate per channel at 14 bits precision, the channels being synchronized to within one sample period. See [11] for further technical details of the array recording system.

2.3 Recording paradigm

We configured the system to wake up and start recording at 8:00 PM, set through the Real Time Clock on the computer motherboard. The system automatically switched itself off at 6:00 AM using the operating system scheduler. Upon return from power-failure the previous wake-up/shutdown cycle was recovered. The system was configured with a repeating trigger and a two-second pre-trigger buffer. The trigger was activated when the average energy of the signal exceeded a threshold for more than half a millisecond, where average energy is computed within a 1 ms moving window of samples. The trigger threshold was set to be three times the noise floor, which in turn was set to the average absolute amplitude per sample of the first two seconds of data seen after system startup. The trigger is deactivated again when the average signal energy has been below threshold for more than two seconds. The recording system saves the pre-trigger buffer and up to and including the final two seconds of sub-threshold average signal energy data to permanent storage each time the trigger is activated. Each such recording generates one data file on the recording system computer.

Upon the completion of this survey, the system had recorded 236 GiB of data. The volume of data is the key technical problem with our approach: to analyze such an amount of data requires a systematic (and automatic, if possible) approach.

2.4 Event detection and classification

The goal of automatic processing is to identify interesting segments of the recorded signals, which we shall call *events*, within which bat calls are likely to be found. Note that in general not every bat call will result in an event, since not every call will pass the event-energy threshold we set.

2.4.1 Event detection and marking

The first step in detecting events in the recorded files was to correct for DC shift. Then we squared and convolved the data with a normalized Gaussian moving average filter of length 100 μs, to construct a measure of the local (short-timescale) average intensity of the signal at each sample time. The filter does not remove interesting events as those are all much longer than the filter length.

An empirically estimated energy threshold was determined by selecting those portions of a recording where the signal power was in the top Θ%. This was found by sorting the samples in the recording into ascending order and setting the threshold to the value of the sample at the $1 - \Theta$ percentile. This is closely related to finding the cumulative distribution function. Figure 3 shows a typical sorted cumulative density of a recording. We want the inverse because we are mapping a specific density Θ to an energy sample value. In our analyses we used $\Theta = 0.1$, i.e. 10%.

Events were detected based on the energy threshold. Those samples of the energy signal that exceed a chosen threshold mark the presence of candidate events. Events are defined as starting at a mark preceded by a 500 μs interval without marks. The end of an event is defined by a mark followed by a 500 μs interval without marks. In this way, a cluster of high energy values, each generating a mark, are grouped into a single event. Furthermore, each event has a front and back "porch" (region) where the signal energy is below the event threshold. So events are clusters of marks with preceding and following "silence"

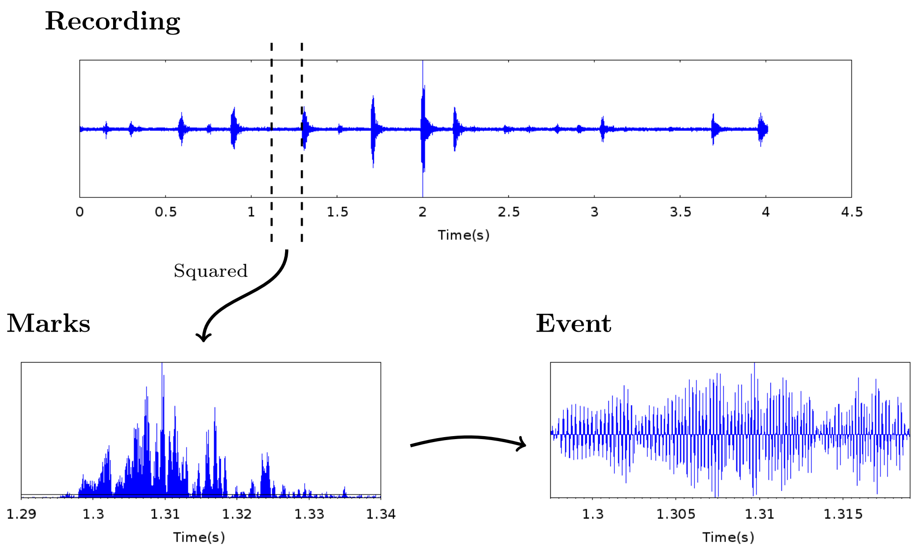


Figure 2: The figure shows how a region of a recording is squared. The marked samples are those exceeding the threshold (horizontal line). The cluster of marks that are closer than 500 μs define an event region.

regions. Figure 2 shows how a region of a recording has been squared, marked samples are those exceeding the energy threshold. The event is defined by the start and ending marked samples $\pm 500 \mu s$ of "silence".

Thus the final form of an event is an interval of the recorded data that begins 500 μs before the start mark and finishes 500 μs after the end mark. Further analysis of these events will classify them as comprising noise alone, or one or more bat calls. Events shorter than 1100 μs (including the front and back porches) were ignored.

2.4.2 Event quality measures

Given the results of the event extraction the problem becomes one of classification. We define a set of quantitative features that capture relevant aspects of the properties of bat call spectrograms and use vectors of their values as inputs to a suitable classification algorithm. The chosen features are in agreement with what other studies have found to be good predictors of bat echolocation and social calls [13, 14, 15, 16].

There are many alternatives to event classification. However, a comparison of methods for classifying bird and amphibian calls [17] showed that Support vector Machines (SVMs) performed best (though [16] showed that ensembles are slightly better relative to SVMs). We have chosen to use the SVM [18] to classify the events into bat and non-bat classes, since it is not relevant to our present goals to develop new, or deploy optimal, classification methods.

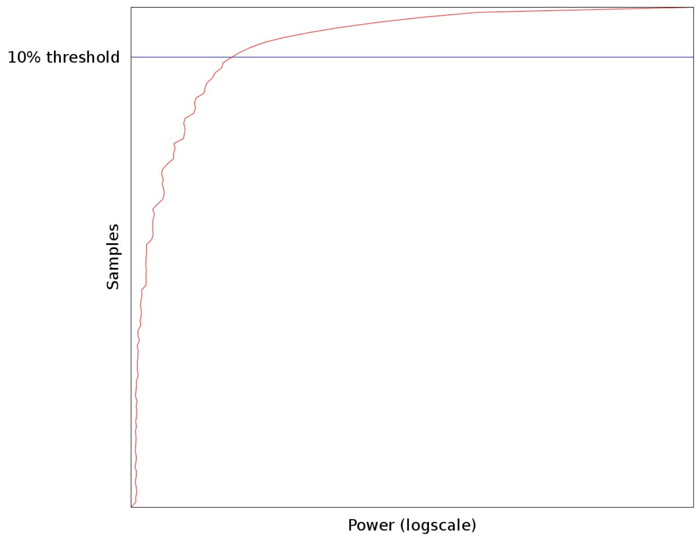


Figure 3: A typical sorted cumulative density plot. The x-axis runs from zero to one and shows for given position the proportion of samples in the set exceeded by that value. The x-axis is the sample value.

We used the LIBSVM[19] implementation of the method, which is a library that supports training and applying support vector machine models to a dataset. For the two-class case that we are considering, the SVM model can be envisioned as a hypersurface that separates a parameter space into two hypervolumes, one volume for each class. The parameter space in our case consists of the features we choose to describe events: *the event quality measures*. To classify the events we used a combination of temporal and spectral analyses. Bat calls are characterized by their relatively short duration and stereotypical spectrographic structure [20]. Consequently, features such as duration, energy, bandwidth and spectro-temporal structure might be expected to provide good discriminants. After some experimentation, we found the measures described below to be good enough classifier features for our purposes. Three of them are plotted in figure 4 and give a feel for how well they separate manually identified events. The quality measures were combined in a composite bat-call classifier. First, we formally define the quality measures used.

Duration The duration of event j was calculated in milliseconds as

$$d_j = \frac{\mathrm{end}(ev_j) - \mathrm{start}(ev_j)}{sr/1000} \tag{1}$$

where sr is the recording sample rate; $\mathrm{start}(\cdot)$ and $\mathrm{end}(\cdot)$ return the indices of the first and last marks belonging to the event supplied as argument, i.e. not

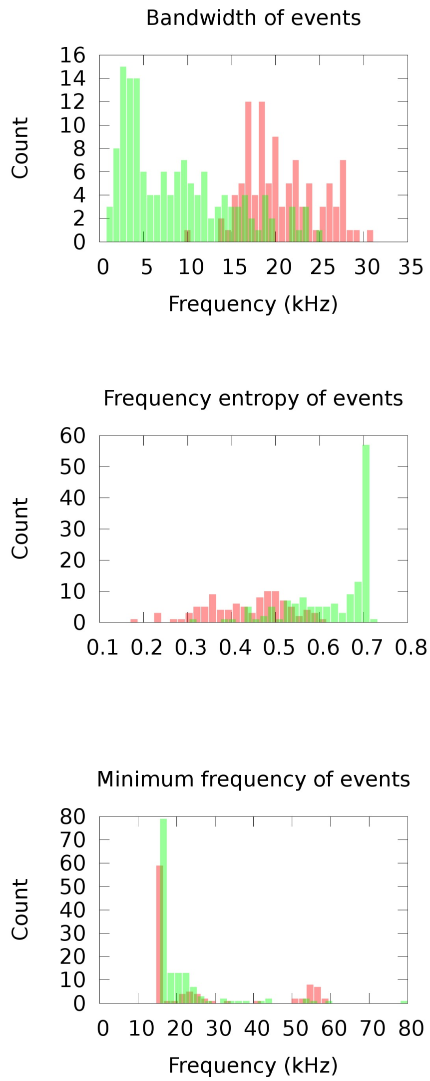


Figure 4: How three of the parameters or quality measures — bandwidth, frequency entropy, and minimum frequency — are divided amongst manually identified bat (red) and non-bat (green) events. The less overlap the better separation the quality measure carries. Note that the red bars are offset to the left to make them more visible.

including the samples from the porch sections.

Average power The average power of event j was calculated as the sum of squares of the samples comprising the event divided by the duration of the event:

$$W_j = \frac{1}{d_j} \sum_{i=\text{start}(ev_j)}^{i=\text{end}(ev_j)} ev_j(i)^2 \tag{2}$$

where $ev_j(i)$ is the i^{th} sample of event j.

Minimum and maximum frequency The most energetic frequency content was determined in a similar manner to the marks in the time domain. The FFT of the event is constructed so that each frequency bin is always 50 Hz wide, which makes events more spectrally comparable. The resulting frequency bins are thresholded to mark the top 10%. Adjacent marked bins are merged into the same cluster if they are less than 300 Hz apart. The minimum f_j^{min} and maximum f_j^{max} are determined by the start and end marks of the largest spectral cluster.

Bandwidth Bandwidth follows directly from the minimum and maximum detected frequencies:

$$bw_j = f_j^{\text{max}} - f_j^{\text{min}} \tag{3}$$

Spectral entropy In order to detect the bat calls even under noisy conditions, we devised a feature that highlights the much more orderly spectro-temporal structure of bat calls relative to noise, similar to the technique used to detect marine mammals in [15]. The feature enabled us to detect bat calls even when masked by moderate noise in the same bandwidth. Entropy is a generic statistical measure [21] that is high for unordered data and low for ordered data. As noise tends to have a broad unordered spectrum, entropy is a potentially good measure for separating bat calls from noise. Even broadband bats are distinguishable from broadband noise as they use a subset the spectral bins, assuming the sample rate is high enough.

The spectral entropy was calculated by first determining the probability distribution of spectral power. This was done by creating 5 kHz frequency bins, summing the power in each bin and dividing by total power of the event:

$$P_j(m) = \frac{1}{d_j W_j} |EV_j(m)|^2 \tag{4}$$

where m indicates the frequency bin. So $|EV_j(i)|^2$ is the power of event j in the i^{th} frequency bin. The entropy of the event is then calculated as:

$$E_j^{\text{freq}} = -\sum_{m=1}^{M} P_j(m) \log P_j(m) \tag{5}$$

2.4.3 Training the model

Classification works by building a model based on vectors of quality measures of manually classified events to determine an appropriate separatrix function [18]. The events are treated as labeled points in a high-dimensional space, and the classifier training algorithm determines a surface that separates the two clouds of labeled points, assuming this to be possible. The SVM algorithm in particular also maximizes the distance from the separatrix to the members of the two classes [18], producing thereby the "best" separating surface in the sense of allowing maximal clearance between the surface and the class members. The distance to the separatrix is a measure of how "bat-like" an event is.

Building the classifier requires labeled data, i.e. samples of input grouped into each of the classes to be learned. This data was prepared manually, thusly: the events were randomized, the quality measures were calculated for all the events, these and accompanying spectrogram, spectrum and amplitude plots were visually inspected, and events which were clearly identifiable as bat calls that exceeded 20 dB signal-to-noise ratio were marked as bat events. During this process many non-bat events were encountered and marked as such, to provide the training set entries for the non-bat class. In total the training set consisted of 99 bat call events and 141 non-bat events, 240 in total. The feature value vectors and their labels were provided to the LIBSVM training algorithm to construct the SVM classifier.

The actual model used by the SVM is based on a Radial Basis Function (RBF) kernel. The RBF is a real-valued function whose value depends only on the distance between a feature vector and a specified center for the function. It is the most common kernel function used with SVMs [19]. The RBF layer in LIBSVM has two parameters, *gamma* and *cost*. The optimal value for these is found by running the *svm-grid* program and adjusting *gamma* and *cost* based on the topographic map produced. This process optimizes the separation into correct classes.

When assessing the accuracy of the SVM classifier, N-fold cross-validation is used. The manually classified training events are divided into N blocks. N SVM classifiers are constructed by training using $N-1$ blocks of the data, while the final block is used for testing the performance of the model constructed on the other $N-1$. Each block is therefore used as testing set, while the rest are used as the basis for the model, giving N independent estimates of the classifier's accuracy and an indication of the variability of accuracy as a function of training data.

The goal of the classifier is to identify, with as high an accuracy as possible, events that contain bat calls. The best classifier obtained from training and cross-validation is used to classify all available events into candidate bat-calls, meriting attention from a biologist, and non-bat events (which may in fact contain bat calls, of course). From that point onward, in the current system design, data analysis requires human judgment. Since all positives potentially demand human attention, we prefer that the classifier generate false negatives rather than false positives for this survey.

2.4.4 Parameter scaling

As the individual feature value ranges are quite disparate when compared to each other, feature rescaling is often recommended to improve the classification model's accuracy. Three classifiers were built using three scaling approaches: no scaling, relative linear scaling, and customized absolute scaling. In the first case, the raw feature values were used unaltered. Relative linear scaling finds the global maximum and minimum values of each feature in the entire dataset, and uses that to scale the feature component in the SVM input parameter vector to the range -1 to 1. This scaling depends on the data set available for training (hence "relative") and it will differ for different data sets. A model trained on one set will thus not necessarily be applicable to another dataset if the minimum and/or maximum differ from the original. The third scaling type rescales each feature based on our *a priori* knowledge of the expected range of values for that feature. It is independent of the actual feature values in the training data set, hence "absolute".

In this third approach, there are three distinct cases. First, if the feature has a fixed minimum and maximum (for instance, spectral entropy lies between 1 and $\log M$) then the feature value is scaled linearly to ± 1. Second, if the feature has one fixed (e.g. lower) bound and one soft (e.g. upper) bound — that is, the *a priori* minimum value is never crossed while the maximum might be exceeded (though that is unlikely) — the feature value is rescaled using the right half of a sigmoid logistic squashing function to map the minimum to -1 and the soft maximum to 0.9. Outliers are thereby compressed into the $(0.9, 1]$ range. For the case of fixed upper and soft lower bound, the same scaling is used with the opposite half of the sigmoid. Finally, third, a feature with two soft bounds is scaled using a full sigmoid so that the bounds map to ± 0.9.

The performance of the three scaling approaches can be compared using the receiver operating curve plots for the classifiers built from those feature vectors. The greater the area under the curve, the better is the performance of the classifier. The ROC curves for unscaled, relative linear scaling, and customized absolute scaling are shown in figure 5. These ROC curves were calculated from the labeled training set, and suggest that the raw feature values are such that feature rescaling provides no advantage for this problem.

3 Results

3.1 General notes and observations

The system was first set up for automatic data capture from 20:00 to 6:00 on June 23rd 2011. The top plot in figure 6 shows the distribution of recorded data volume over time. The main panel in the top plot shows data volume plotted with respect to day and time-of-night counted using 2 minute time bins; the top panel shows the total data volume for that day in bytes. The bottom plot shows a histogram of the sizes of data capture files; notice that most files are four seconds in length or equivalently 4 MiB (pre-trigger buffer plus post-trigger

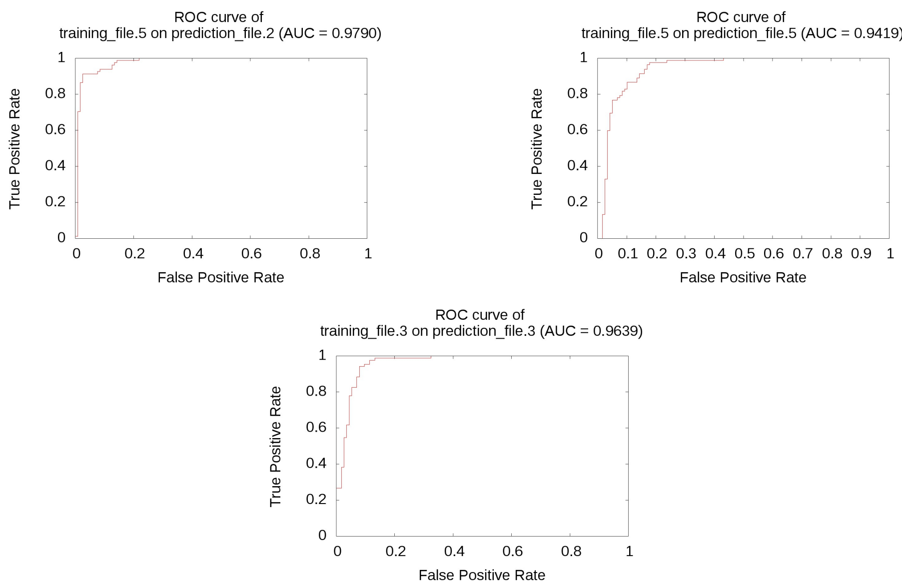


Figure 5: ROC curves comparing the accuracy of SVM models based on unscaled, relative linear scaled, and custom absolute scaled SVM parameters. **Left:** Unscaled, **Right:** Relative linear scaling, **Bottom:** Custom absolute scaling (§ 2.4.4).

silence), but there is a long tail of much bigger files shown in more detail in the inset plot. Overall, the system captured 236 GiB of raw data in total, in 11059 files of varying sizes.

Unfortunately the system was accidentally disconnected from the grid on June 27th 2011. This happened again on July 6th 2011 and this time the system remained off-line for almost a month. This is reflected in figure 8 by the many empty days during July. These disconnections always happened during working hours, i.e. between 8:00 h and 17:00 h, so the system was never interrupted during a recording run. The result is that we have sporadic recording days in June and July. When the system was restarted on August 1st 2011 it remained operational until the survey period was over on September 27th 2011. As we are interested in patterns over time, we have chosen to focus on the last period only, i.e. August 1st through September 27th, in the results and analysis below.

Although we recorded four channels in this survey, this paper only considers data from channel one as that is sufficient for our analyses here. There are improvement possibilities and other analyses available with multi-channel recordings, which are discussed in more detail in Section 4.

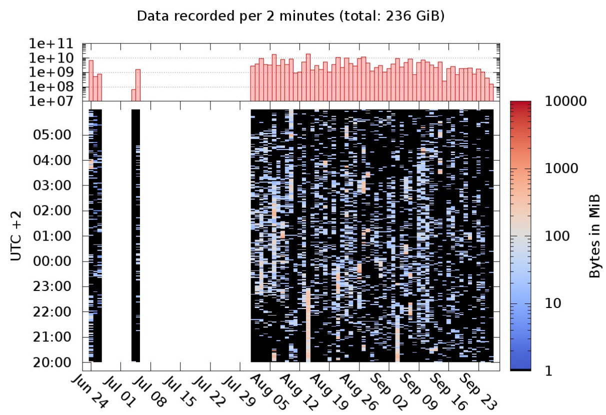
3.2 Event detection and classification

After applying the threshold and event marking methods described the data can be viewed as an event distribution with a total of 8,413,249 events. These automatically-identified regions of the recordings are where we hope to find bat calls. The construction of the events and their classification took a total of 10 hours on a dual core 1.6 MHz Intel Atom machine.

Figure 7 shows how the data files separate into number of events per second (top panel) and the percentage of the whole recording that events cover (bottom panel). The top panel shows that all of the recordings have fewer than 500 events per second, while the bottom panel shows that in most files the events cover less than 15 % of the samples.

A plot of the event data analogous to the temporal distribution of recordings is shown in figure 8 (top panel). The color of each block now represents the number of events occurring during each 2 minute bin. Note that this overview map plots all events, bat and non-bat alike: at this stage, we have not classified the events. The lines plotted on the figure are the times for nautical sunset/sunrise (gold colored) and end of/start of nautical twilight (gray colored) for the given night. There are clear high incidence areas both in the recordings plot (figure 6, top panel) and the event plot (figure 8, top panel) — for instance early evening on August 14th — and there are also areas which have few events, such as the later night and early morning of that same session. The distribution of events seems to peak within the inner sunset-sunrise lines, apart from busy periods such as during August 10th morning, 14th evening and September 4th evening.

Calculating the feature vectors for each event takes about 10 hours for the two month dataset. When the feature vectors are ready we apply the SVM classifier to the detected events which results in the bottom plot in figure 8, a total of 115,456 events. These remaining events are all candidate bat events, their num-



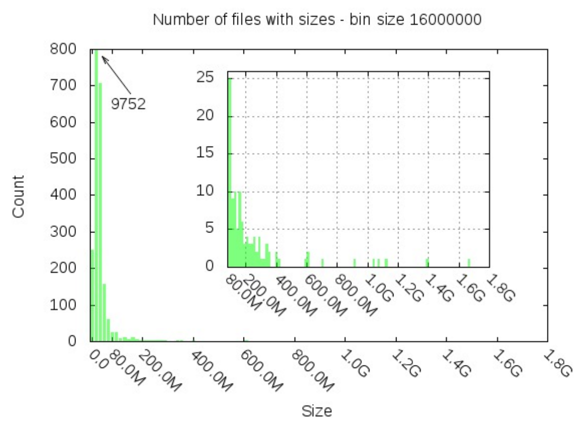


Figure 6: **Top:** Data distribution over time and size. The histogram along the top shows daily totals in number of bytes. Each slice in a day represents a two minute period and its color the number of bytes recorded during those minutes. Slices containing files longer than 2 minutes are inflated because the count depends on the start of a recording. **Bottom:** Histogram of recording file sizes. It shows that most of the recordings were the size of two post-trigger buffers, 4 seconds or roughly 4 MiB. There is a long “tail” on the histogram of a few long recordings. This part is zoomed in by the inset plot, with the size range limited from 80 MiB to 1.8 GiB.

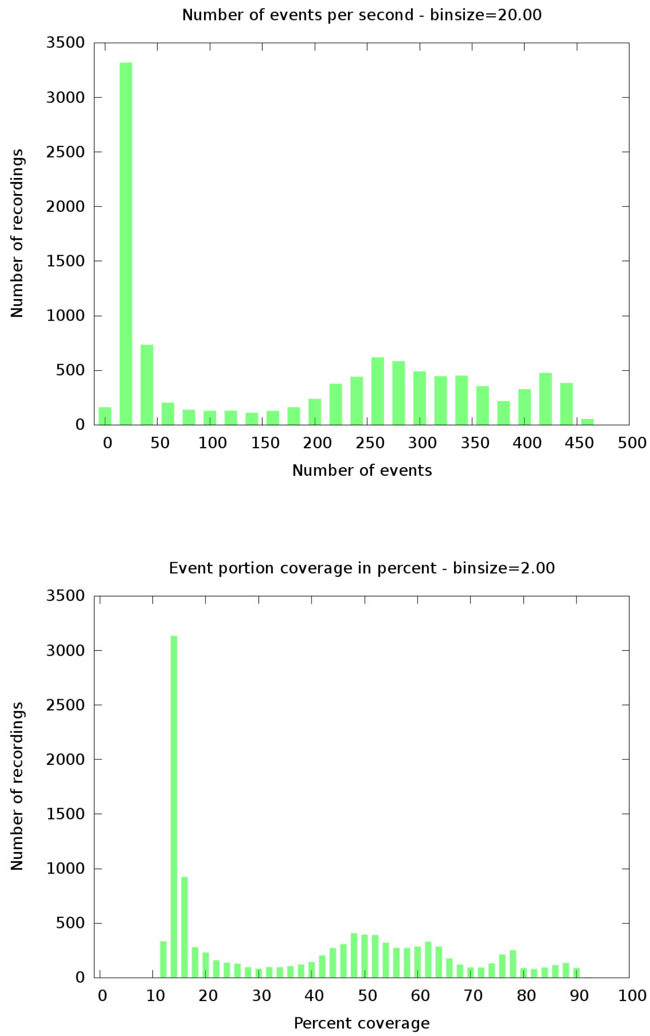


Figure 7: **Top:** Histogram of number of events per recording per second. **Bottom:** Histogram of event coverage in each recording file, i.e. what percentage of samples belong to an event.

	Aug 3	Aug 11	Aug 24	Aug 26	Sep 4	Sep 10	Sep 14
Eptesicus serotinus	145	0	34	58	0	20	0
Nyctalus noctula	12	0	31	6	0	3	1
Pipistrellus pygmaeus	16	10	38	13	0	47	31
Pipistrellus nathusii	0	0	2	0	0	5	1
Myotis daubentonii	0	0	3	0	0	0	0
Myotis dasycneme	0	0	2	0	0	1	0
Non-bat event	16	0	36	104	31	50	62

Table 1: Manually identified bat events in a sample of events labeled as candidate bat events by the custom-scaled SVM classifier.

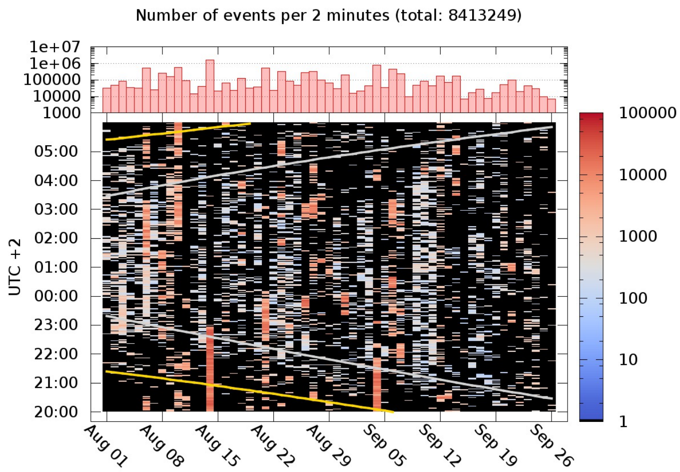
ber is considerably smaller, and many of the bright areas that were present in the top plot have disappeared. Inspection of the events in the previously-bright areas reveals that the majority of them are noise recordings, and correlation with weather data for the region suggests that the noise is caused by local rain. The SVM classifier rejects a large proportion of the rain-noise events, though given their density and the accuracy of the classifier it is unreasonable to expect it to eliminate all of them.

There are several interesting features to notice in the distribution plots in figure 8. Much of the noise has been removed, but regions of relatively high activity remain. Sunset and sunrise appear to have a substantial impact on activity since the candidate bat events tend to occur between the sunset and sunrise boundaries. Throughout the recording period there appear to be both high activity days and low activity days.

As yet we do not know what proportion of the candidate bat calls that are actual bat calls. To address this question, we inspected a number of events by looking at their spectrograms, signal-to-noise ratio relative to the rest of the recording, and duration. We looked at 7 different days and only every n^{th} event in each day, n was set to keep number of events to examine below 200. In total we reviewed just fewer than 800 events. Table 1 shows the results from this tally where we have identified a number of different bat species with a high degree of confidence. Figure 10 shows some of the bat events counted in the table.

There is considerable variability between the different days, which seems to be due to precipitation; see figure 9. To quantify this, two confusion matrices were counted: one for dry nights and one for rainy nights. 200 random events from four dry nights were visually inspected, 100 classified as bat and 100 classified as non-bat. The same was done for four rainy nights. The results can be seen in table 2. This confirms that rain has a major effect on the accuracy of the classification which should be handled for further analysis of this data and for future surveys.

Clearly, there are still many non-bat events present in the reduced collection,



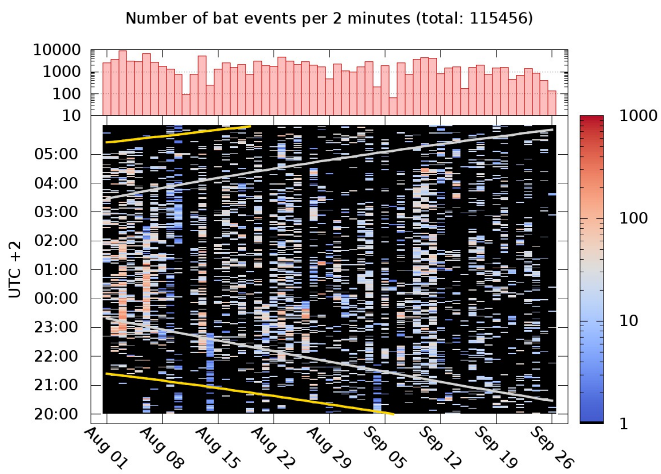


Figure 8: **Top:** Temporal distribution of events. **Bottom:** temporal distribution of classified bat events. Notice that the color bar scales are different by two orders of magnitude. The slanted horizontal lines show sunset/sunrise and nautical twilight end/start, gold and gray colored respectively. The histogram along the top shows daily total of events.

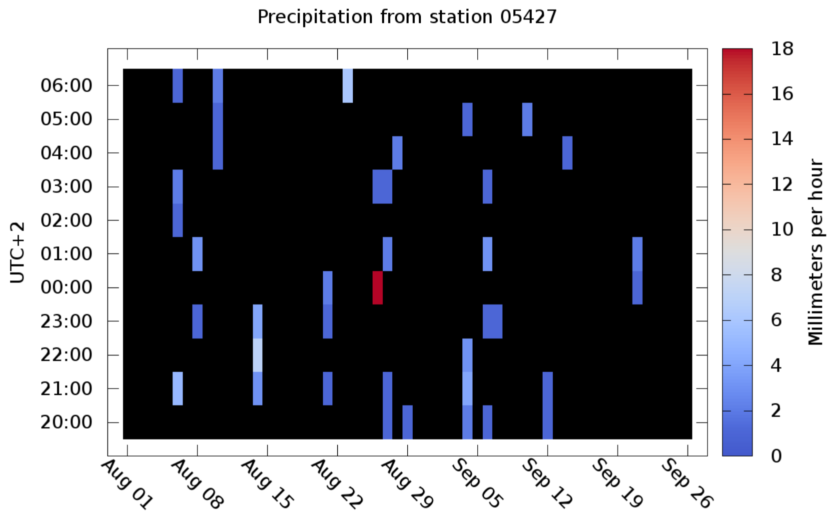


Figure 9: Hourly average precipitation measured 2.5 km away from the survey site.

		Predicted class	
		Bat	Non-bat
Actual class	Bat	94	3
	Non-bat	6	97

		Predicted class	
		Bat	Non-bat
Actual class	Bat	40	0
	Non-bat	60	100

Table 2: Confusion matrix for bat and non-bat events. **Top:** Dry nights. **Bottom:** Rainy nights.

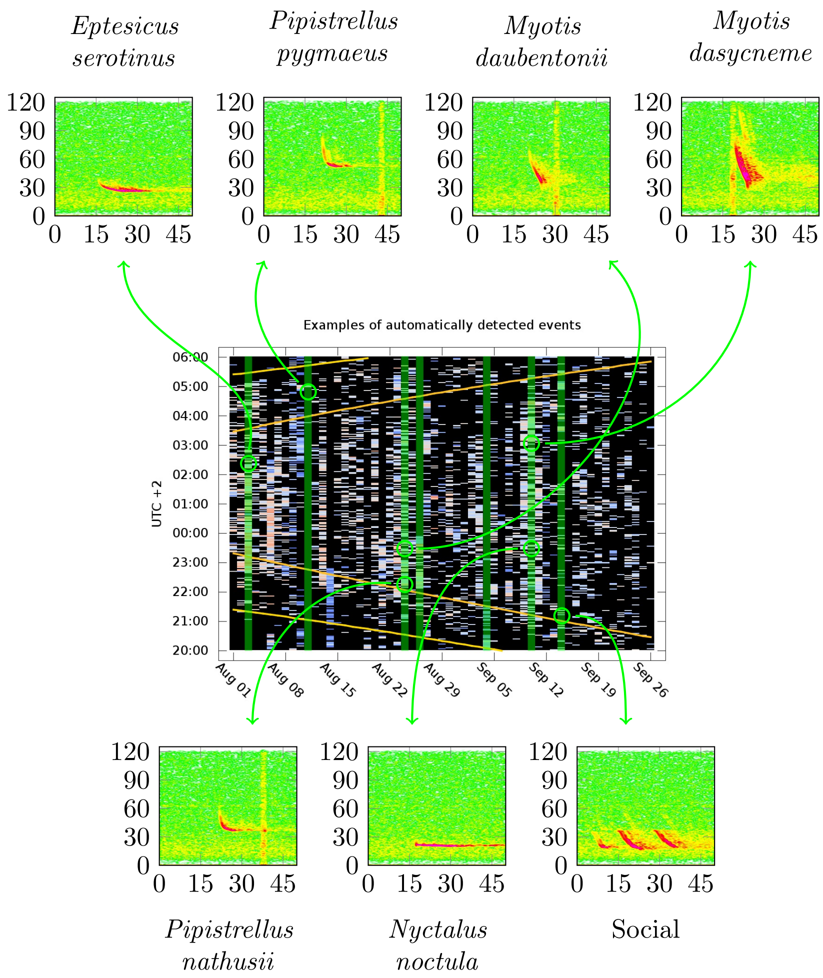


Figure 10: Example of typical events classified during the tally.

but when looking chronologically through the events, it is obvious that there are clusters of non-bat events, presumably once again related to rain. We are convinced that these can be spotted and removed fairly easily by improved automatic analysis.

Another interesting feature of this tally is that we found evidence of two species that have not been registered before in the survey area, namely *P. nathusii* and *M. dasycneme* [22].

4 Discussion

The survey described, and the methodology used to effect it, suggest a number of interesting discussion points, both biological and technological. We discuss some of them below.

4.1 Technical viewpoint

Firstly, from a technical viewpoint, the data processing has been designed to favor simplicity whenever that provides adequate functionality: optimal processing is unnecessary for our goal of demonstrating the potential of the new array recording technology. There are, therefore, a number of sensible improvements which — though unnecessary to achieve our present objectives — would yield a substantial advantage in future deployments of the system. Most or all of them can also be applied retrospectively to the data already collected.

For example, the single-threshold approach to event marking tends to label isolated regions of a recording file, and the mark clustering then aggregates nearby marks to make the events. A dual-threshold approach — a high threshold for the initial mark, and a lower threshold to determine the start and end of the candidate event — might be preferred because this could include low intensity calls from the beginning of a sequence of calls from a passing bat.

The support vector machine was chosen as a robust and versatile classifier, but there may be techniques better suited to the specific class discrimination problem posed by our data. In particular, adjusting the classifier parameters depending on the presence or absence of rain would be sensible, given the large number of rain events highlighted by the initial marking process: the prior distribution of the two classes is markedly different on rainy and dry days.

The question of scaling the feature vector for the classifier is also open. For the training data set, unscaled data clearly results in a very good classifier (figure 5, left); absolute rescaling causes a small loss of performance but incorporates more biological knowledge about the reasonable range of values for the chosen features – for example duration has a hard minimum and a soft but fairly predictable maximum (figure 5, bottom). It remains to be seen whether it is better to use the absolute scaling trained classifier when running on the bulk of events, as we did here, or to keep to the unscaled data trained classifier that performs better on the training set.

One might also envisage refinements of the chosen features, or further features that would increase the discriminating power of the classification process. The model can also be improved by adding more labeled training data using all the candidate bat events examined by the biologists, since their true class is known after inspection. After identifying likely candidates it is possible to search for quiet calls when good examples have been identified. The good candidates could be used as templates for extracting weaker signals, for example by matching the signal to the template with cross-correlation.

The analysis presented here used data from a single channel in the recorded data and also treated events as isolated occurrences. Given the four channels recorded, one can in principle recover the spatial position of the sound source relative to the recording array using time-of-arrival-difference techniques [23, 24]. This would allow the system automatically to recover spatial data for the bats detected. Correlating event marking between the four channels might improve event detection. Temporal patterns in the single channel event stream could also be extracted: bats do not generally emit isolated calls, but structure their call patterns temporally to reflect the activities in which they are engaged. One might therefore recover data not only about which bat species were present but also what they were doing, e.g. insects caught using buzz sequence counting. These analyses entail technical difficulties relating to signal quality and signal matching, but there are many approaches in the literature worth trying [25, 26, 27]. Note that recovery of buzz sequences with the automatic method still needs work, because of the high repetition rate and the short duration of buzz phase emissions, especially the termination buzz which typically comprises calls of down to 0.3 ms duration and 20 dB fainter than other (e.g. search) calls [28, 29].

Finally, it would be advantageous to augment the survey recording equipment with a small weather station able to monitor at least local temperature, humidity, illumination and rainfall. The former two data are needed to compute an accurate value for the speed of sound — for calculating spatial data — while all data are also relevant to bat behavior. This would certainly improve measurements of localized phenomena. Adding a sensor suite that measures some of these values together with the acoustic recordings has already been tested [11].

4.2 Biological viewpoint

From a biological viewpoint, there are a number of interesting ecological results from the survey. First, the variety of species sampled at the recording site includes frequent appearances by individuals of at least two species believed not to be present in the area, *Pipistrellus nathusii* and *Myotis dasycneme* [22]. Thus this technique proves very promising for surveying species that are rare, and thus rarely sampled in short random surveys or transects. One advantage of a longitudinal survey such as this is that one potentially observes everything that occurs, rather than relying on careful judgment and some luck to be surveying at the right time. This will prove useful for estimating the true occurrence of less common species and also for determining areas of importance for transient events such as, for example, migration.

Another, surprising, observation from the subset of survey data inspected so far is that social calls were frequent compared to the observed rates in a large number of echolocation studies. There are relatively few systematic studies of social calls in the field in temperate areas [30, 31, 32, 33, 34, 35]. Recordings of social calls emitted by bats may provide data for assessing cryptic diversity and investigating reproductive strategies and foraging ecology. Further, social calls reveal areas of importance for reproduction and roosting. Occupancy may help evaluate reaction to anthropogenic impact, not only through exploitation of land resources, but also through climate change and other changes of habitat.

One may consider the extent to which the survey is a reliable indicator of bat population and activity. A full answer to this question awaits a more complete analysis of the collected data, but initially we can say that the observed bat activity certainly underestimates what is really present. The automatic processing focuses on minimizing false positives, but may generate many false negatives to be caught by more sophisticated automatic processing not yet implemented. Thus, while inspections suggest that around 30% of candidate bat-call events proposed by the current signal-processing chain will turn out to be misclassified noise, there are certainly bat calls present in the large bulk of data not highlighted for human examination.

It is apparent in the bottom plot of figure 8 that the bat activity apparently varies across the sunrise and sunset periods in a systematic way, and one could construct plots of the density of bat calls (in total or per species) against the time relative to sunrise and sunset to investigate those phases of the bat's night. This also shows that the survey length should be adjusted to the length of the day instead of a fixed number of hours.

Of course, no survey methodology is unbiased [36]. In our case, the principal bias favors loud bats since their calls more easily pass our energy thresholds. This is partially an artifact of the simple signal processing chain we have presented. Having extracted all the loud bats, one could continue and search for quieter ones in the remaining data — especially if given models of the types of call made by species of interest. Such a search could also be automated, but would be much more time-consuming than the process adopted in this survey. On the other hand, technology aside, louder bats will always be detectable in a larger volume than quieter bats and this will inevitably bias population estimates in favor of the louder species. However, given spatial data for each call, extracted as outlined above, one could straightforwardly normalize for volume and derive bat population density measurements in units of individuals per cubic meter. Such analysis would be hard or impossible to do without the array recording technique underpinning the present survey.

A rather more subtle (absence of) bias is suggested by the observation that bat calls are detected in our survey very close to the end of periods of rain. Given that weather can be a very local phenomenon and that the nearest weather recording station is 2 km from our survey site, we can only speculate; but it appears that bats may either fly immediately after rain or indeed during rain. Careful analysis of the data from rainy periods may reveal bat calls partially masked by the rain noise. It could perhaps be that the generally held belief that bats do not fly in rainy weather reflects a preference of biologists (for not recording in the wet) rather than of bats. The serious point is that the methodology we demonstrate avoids this kind of potential observer bias, simply by observing all the time. Long-term automatic surveying not only permits sampling of data that are un-biased by weather conditions, but also are less biased by difficulty of access to a potential bat habitat, since one would require access only once or a few times.

Correlation with weather and ephemeris information suggests that further interesting observations may be possible. For instance, the data plotted in the lower panel in figure 8 shows nights with little bat activity followed by nights with much activity (e.g. September 8–11). Correlation with weather data (temperature and wind, principally), lunar phase and cloud cover (measured some 15 km from the survey site) suggests that the quiet nights in question were cloudless and therefore cool with bright moonlight, while the following nights were cloudy and warmer. This type of analysis would be facilitated by adding a weather station.

It is clear that in the second half of the survey (from around September 1st) the recording time should have been expanded to capture the longer nights in late summer and fall. This will be done in future surveys.

As a final comment, the principal difficulty of the survey methodology we advocate here is the sheer volume of data generated. A full season of recording bats, from say April to October, would generate nearly 1 TiB of recorded data. Storing such a data volume is trivial nowadays, but analyzing it by hand is infeasible. The automated processing developed here, with the improvements discussed, can reduce this data volume to something more manageable for human inspectors. For specific questions such as species density it may be possible to automate the answers completely. Our survey demonstrates the growing possibility of carrying out ecologically interesting comprehensive long-baseline studies of acoustically active animals without dedicating a lot of human resources on the enterprise.

5 Conclusion

The goals of this survey were three: to test and demonstrate that an automatic array recording system could be deployed long-term in a real bat-surveying application, to verify that the resulting voluminous data could be analyzed effectively, and to investigate and document local bat population behavior in the vicinity of the University of Southern Denmark.

Our inexpensive scalable array recording system [11] was used for the survey, and was at the survey site for just over three months. Despite some unexpected power interruptions, the system functioned largely as intended and we were able to capture two continuous months' worth of recorded data totaling nearly 240 GiB. Automatic processing of the data successfully identified around 115,000 candidate bat "events" so-labeled by a support vector machine classification algorithm, of which we expect some 94% to be recordings of actual bat calls, assuming that the effects of rain can be detected and eliminated. This represents a small percentage of the total volume of data which requires human judgment for further analysis.

From a biological perspective, the bat survey also reveals some interesting and novel results. Examination of a sample of events from specific days shows a variety of bat species observed at the survey site, including species not observed in the survey area before. A surprisingly large proportion of the total calls inspected turned out to be social calls rather than echolocation calls.

There are suggestions of correlations, some unexpected, between bat activity and local weather: busy activity on warm nights especially following cloudless moonlit nights, and bat activity surprisingly close to the end of rain showers. These tentative observations require further detailed analysis to confirm them, and the data collected in the survey will support such future analyses.

We therefore conclude that our goals have been met: the prototype equipment has been demonstrated in a real survey application that yielded biologically interesting results; and the simple data processing approach take proved surprisingly effective in dealing with the volume of data recorded by the deployed system. We conclude that automatic long-term recording is both feasible and practical, given our experience with this system. A number of plausible improvements to the automatic processing, mentioned in the technical discussion section, may eliminate more of the data from human consideration; conversely, improvements may also retrieve less conspicuous bat calls from the data rejected by the current analytic methods. Finally, implementation of trajectory tracking in the data processing chain will allow new analyses of survey data, such as discrimination of current bat behavior and normalization for species call amplitude.

Acknowledgments

Part of this project was funded by the EU 7th Framework Programme (FP7), through the ChiRoPing project (ICT-2007-1 STREP project 215370) and The Danish Council for Independent Research, Natural Sciences. A special thanks goes to the OSS community which has brought about many good and useful things, many of which have been used in this project.

References

[1] D. T. Blumstein, D. J. Mennill, P. Clemins, L. Girod, K. Yao, G. Patricelli, J. L. Deppe, A. H. Krakauer, C. Clark, K. A. Cortopassi, et al., Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus, Journal of Applied Ecology 48 (3) (2011) 758–767.

[2] T. H. Kunz, E. B. Arnett, W. P. Erickson, A. R. Hoar, G. D. Johnson, R. P. Larkin, M. D. Strickland, R. W. Thresher, M. D. Tuttle, Ecological impacts of wind energy development on bats: questions, research needs, and hypotheses, Frontiers in Ecology and the Environment 5 (6) (2007) 315–324.

[3] P. M. Cryan, J. W. Jameson, E. F. Baerwald, C. K. Willis, R. M. Barclay, E. A. Snider, E. G. Crichton, Evidence of late-summer mating readiness and early sexual maturation in migratory tree-roosting bats found dead at wind turbines, PloS one 7 (10) (2012) e47586.

[4] J. Fiedler, Assessment of bat mortality and activity at buffalo mountain windfarm, eastern tennessee, Ph.D. thesis, University of Tennessee, Knoxville (2004).

[5] R. Barclay, E. Baerwald, J. Gruver, Variation in bat and bird fatalities at wind energy facilities: assessing the effects of rotor size and tower height, Canadian Journal of Zoology 85 (3) (2007) 381–387.

[6] B. Bruderer, A. G. Popa-Lisseanu, Radar data on wing-beat frequencies and flight speeds of two bat species, Acta Chiropterologica 7 (1) (2005) 73–82.

[7] T. Kunz, E. Arnett, B. Cooper, W. Erickson, R. Larkin, T. Mabee, M. Morrison, M. Strickland, J. Szewczak, Assessing impacts of wind-energy development on nocturnally active birds and bats: A guidance document, The Journal of Wildlife Management 71 (8) (2007) 2449–2486.

[8] N. Hristov, M. Betke, T. Kunz, Applications of thermal infrared imaging for research in aeroecology, Integrative and Comparative Biology 48 (1) (2008) 50–59.

[9] I. Ahlén, L. Bach, H. J. Baagøe, J. Pettersson, Bats and offshore wind turbines studied in southern scandinavia., Tech. rep., Swedish Environmental Protection Agency (2007).

[10] E. F. Baerwald, R. M. Barclay, Geographic variation in activity and fatality of migratory bats at wind energy facilities, Journal of Mammalogy 90 (6) (2009) 1341–1349.

[11] T. Andreassen, A. Surlykke, J. Hallam, D. Brandt, Ultrasonic recording system without intrinsic limits, The Journal of the Acoustical Society of America 133 (2013) 4008–4018.

[12] P. Stahlschmidt, C. A. Brühl, Bats as bioindicators–the need of a standardized method for acoustic bat activity surveys, Methods in Ecology and Evolution 3 (3) (2012) 503–508.

[13] N. VAUGHAN, G. JONES, S. HARRIS, Identification of british bat species by multivariate analysis of echolocation call parameters, Bioacoustics 7 (3) (1997) 189–207.

[14] K. V. Melendez, D. L. Jones, A. S. Feng, Classification of communication signals of the little brown bat, The Journal of the Acoustical Society of America 120 (2006) 1095.

[15] C. Erbe, A. R. King, Automatic detection of marine mammals using information entropy, The Journal of the Acoustical Society of America 124 (2008) 2833.

[16] R. D. Redgwell, J. M. Szewczak, G. Jones, S. Parsons, Classification of echolocation calls from 14 species of bat by support vector machines and ensembles of neural networks, Algorithms 2 (3) (2009) 907–924.

[17] M. A. Acevedo, C. J. Corrada-Bravo, H. Corrada-Bravo, L. J. Villanueva-Rivera, T. M. Aide, Automated classification of bird and amphibian calls using machine learning: A comparison of methods, Ecological Informatics 4 (4) (2009) 206–214.

[18] B. E. Boser, I. M. Guyon, V. N. Vapnik, A training algorithm for optimal margin classifiers, in: Proceedings of the fifth annual workshop on Computational learning theory, ACM, 1992, pp. 144–152.

[19] C.-C. Chang, C.-J. Lin, Libsvm: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology (TIST) 2 (3) (2011) 27.

[20] H.-U. Schnitzler, E. K. Kalko, Echolocation by insect-eating bats: We define four distinct functional groups of bats and find differences in signal structure that correlate with the typical echolocation tasks faced by each group, Bioscience 51 (7) (2001) 557–569.

[21] C. E. Shannon, W. Weaver, A mathematical theory of communication (1948).

[22] H. J. Baagøe, Danish bats (mammalia: Chiroptera): Atlas and analysis of distribution, occurence, and abundance, Steenstrupia (26, 1) (2001) 1–117.

[23] Y. Chan, K. Ho, A simple and efficient estimator for hyperbolic location, Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on] 42 (8) (1994) 1905–1915.

[24] P. Madsen, M. Wahlberg, Recording and quantification of ultrasonic echolocation clicks from free-ranging toothed whales, Deep Sea Research Part I: Oceanographic Research Papers 54 (8) (2007) 1421–1444.

[25] M. J. O'Farrell, B. W. Miller, W. L. Gannon, Qualitative identification of free-flying bats using the anabat detector, Journal of Mammalogy (1999) 11–23.

[26] P. M. Gorresen, A. C. Miles, C. M. Todd, F. J. Bonaccorso, T. J. Weller, Assessing bat detectability and occupancy with multiple automated echolocation detectors, Journal of Mammalogy 89 (1) (2008) 11–17.

[27] S. L. Skalak, R. E. Sherwin, R. M. Brigham, Sampling period, size and duration influence measures of bat species richness from acoustic surveys, Methods in Ecology and Evolution 3 (3) (2012) 490–502.

[28] E. K. Kalko, H.-U. Schnitzler, The echolocation and hunting behavior of daubenton's bat, myotis daubentoni, Behavioral Ecology and Sociobiology 24 (4) (1989) 225–238.

[29] A.-M. Seibert, J. C. Koblitz, A. Denzinger, H.-U. Schnitzler, Scanning behavior in echolocating common pipistrelle bats (pipistrellus pipistrellus), PloS one 8 (4) (2013) e60752.

[30] K. Lundberg, R. Gerell, Territorial advertisement and mate attraction in the bat pipistrellus pipistrellus, Ethology 71 (2) (1986) 115–124.

[31] K. E. Barlow, G. Jones, Differences in songflight calls and social calls between two phonic types of the vespertilionid bat pipistrellus pipistrellus, Journal of Zoology 241 (2) (1997) 315–324.

[32] K. E. Barlow, G. Jones, Function of pipistrelle social calls: field data and a playback experiment, Animal behaviour 53 (5) (1997) 991–999.

[33] D. Russo, G. Jones, The social calls of kuhl's pipistrelles pipistrellus kuhlii (kuhl, 1819): structure and variation (chiroptera: Vespertilionidae), Journal of Zoology 249 (4) (1999) 476–481.

[34] J. Russ, P. Racey, Species-specificity and individual variation in the song of male nathusiusâĂŹ pipistrelles (pipistrellus nathusii), Behavioral Ecology and Sociobiology 61 (5) (2007) 669–677.

[35] P. Georgiakakis, D. Russo, The distinctive structure of social calls by hanák's dwarf bat pipistrellus hanaki, Acta Chiropterologica 14 (1) (2012) 167–174.

[36] C. Flaquer, I. Torre, A. Arrizabalaga, Comparison of sampling methods for inventory of bat communities, Journal of Mammalogy 88 (2) (2007) 526–533.