

CARACAL: A versatile passive acoustic monitoring tool for wildlife research and conservation

Authors: Matthew Wijers², Andrew Loveridge², David W. Macdonald² and Andrew Markham¹

¹ Department of Computer Science, University of Oxford, United Kingdom

² Wildlife Conservation Research Unit, Department of Zoology, The Recanati-Kaplan Centre, University of Oxford, United Kingdom.

Corresponding Author: Matthew Wijers (matthew.wijers@zoo.ox.ac.uk)

Abstract

Acoustic localization technology has been widely tested and applied for passive acoustic monitoring and ecological research, however, hardware costs of commercially available devices limit scalability. Furthermore, few studies have explored its use with low-density arrays. We present a low-cost, custom-designed hardware and software system termed 'CARACAL' that is able to extract and localize weak acoustic signals. The key to this is the use of four microphones on each logger, allowing for phase-based measurements and the ability to enhance signal-to-noise ratio through beamforming. As a proof of concept, we test the functionality of the CARACAL system by conducting a gunshot localization experiment and demonstrate animal call detection and localization from a lion predation event. Results show the system could locate gunshots with an average accuracy of $33.2 \pm 15.3\text{m}$ within an array of 7 stations 500m apart. When applied to animal call positioning, we show long range (> 1 km) localization of three different species' calls, Cape buffalo, chacma baboon and spotted hyaena. With a cost of approximately £150 per unit, the CARACAL system provides a cost-effective solution for acoustic localization over large areas. The system is open source and can be customized to suit a variety of wildlife research applications.

Keywords

passive acoustic monitoring; acoustic localization; bioacoustics; acoustic array; animal vocalization.

Introduction

Ecosystem monitoring is a key aspect of effective conservation strategies that provides baseline evidence for determining wildlife population trends and habitat use (Gibbs et al. 1999). Monitoring is equally important for facilitating the detection of environmental disturbances threatening the viability of these populations or the ecosystem as a whole (Astaras et al. 2017). Numerous survey methods exist for monitoring wildlife, however, in many cases there are fundamental challenges with scaling across space and time (Noon et al. 2012). National parks and other wildlife areas often span several thousand square kilometres and lack the infrastructure to facilitate frequent access for repeatable surveys. Similarly, obtaining long-term datasets with high temporal resolution for these areas can be constrained by cost and human physical limitations. One technology which has the potential to address some of the challenges faced by traditional surveys is passive acoustic monitoring (Wrege et al. 2017).

Passive acoustic monitoring (PAM) is an emerging tool in ecology and resource management which collects and processes environmental audio data for research and monitoring purposes. Though the challenges of marine mammal monitoring prompted most of the early development of PAM technology, terrestrial PAM usage has been stimulated by substantial increases in digital recorder capabilities, reductions in recorder costs, and the emergence of adaptable analytical software (Gibb et al. 2018; Marques et al. 2013). The majority of terrestrial PAM research has been conducted on birds (McGregor et al. 1997; Mennill et al. 2006; Frommolt & Tauchert 2014; Sebastián-González et al. 2015), anurans (Crouch and Paton 2002; Ospina et al. 2013), insects (Mankin et al. 2002; Pinhas et al. 2008; Riede 1998) and bats (O'Farrell et al. 1999; Newson et al. 2017). PAM systems have a number of ecological applications such as determining species occupancy from presence and absence data (Heinicke et al. 2015), population assessment (Adi et al. 2010; Marques et al. 2013), monitoring of environmental disturbances and their effects on species behaviours (Hatch et al. 2008; Pirota et al. 2015; Wrege et al. 2017), animal movement and territory use (Clark et al. 1996; Kalan et al. 2016) and investigating animal vocal behaviour (Payne et al. 2003). While the functionality of PAM systems varies

according to research and monitoring objectives, several studies have outlined the importance of further development on automated sound detection and localization (Ali et al. 2009; Kalan et al. 2016; Marques et al. 2013).

Automated sound signal detection is critical for processing the vast quantities of data produced from multi-sensor acoustic arrays. Common detection techniques include hidden Markov models (Zilli et al. 2014), cross-correlation (Mellinger and Clark 2000) and supervised machine learning such as decision trees and support vector machines (Heinicke et al. 2015; Sebastián-González et al. 2015; Newson et al. 2017). Variable performance results have been reported for detection algorithms with recall rates (proportion of correctly classified sound signals) often below 60% (Digby et al. 2013; Heinicke et al. 2015; Swiston and Mennill 2009). Performance is likely to be species related as some animal vocal signals may be more amenable to detection than others due to the acoustic characteristics of the calls (Mellinger et al. 2007; Newson et al. 2017). Acoustic signal quality also affects the probability of successful detection and may be influenced by the presence of overlapping, non-target sound signals from both natural and anthropogenic sources, the distance of the animal from a sensor and local environmental factors (Gibb et al. 2018).

Acoustic localization, while not as common as acoustic detection, is generally achieved using an array of three or more spatially separated audio sensors where differences in time of arrival (ToA) of an animal call signal can be used to calculate signal source location (Ali et al. 2009; Wilson et al. 2014). Other methods for determining location are based on angle of arrival of the signal (on a multichannel recorder) and differential signal strength (Clark 1980; Ali et al. 2009). Acoustic localization accuracy can be affected by a number of environmental and array-specific factors. Dense vegetation, variable topography and the presence of large temperature and humidity gradients affect sound speed and attenuation and can be difficult to account for, thereby reducing the accuracy of location estimates (Darras et al. 2016). The degree of accuracy is also an intricate function of the locations of microphones in an array. More spatially extensive measurements of the arriving wavefronts leads to greater accuracy, yet larger spacing between microphones can decrease received signal strengths and

coherence, degrading the accuracy of those wavefront measurements. In practice, location accuracy often decreases when the sound source originates from outside the convex hull of the sensor array (McGregor et al. 1997; Ali et al. 2009; Frommolt and Tauchert 2014). Locating sound sources over large distances, even in flat, open areas, has therefore been challenging (Frommolt and Tauchert 2014).

The availability of audio recording devices capable of capturing data suitable for acoustic localization has been limited and thus the majority of studies have relied on commercial products such as the Wildlife Acoustics Song Meters as used by Mennill et al. (2012). These modern, GPS equipped devices cost upwards of US\$930 (Wildlife Acoustics 2019) which restricts their use to smaller, well-funded surveys. With the recent advances and availability of microcontroller technology, many researchers are now turning to custom-made devices to suit their own needs at a fraction of the cost (Beason et al. 2018; Hill et al. 2018; Whytock and Christie 2017). Most of the recent open-source innovations are, however, incapable of offering acoustic localization functionality.

In this study we present a low-cost, open-source hardware platform called CARACAL (Conservation at Range through Audio Classification And Localization) and associated signal processing software that is able to detect, match and localize sound emitters. Improved signal detection performance is facilitated by a novel multi-microphone hardware design which enables coherence processing. We demonstrate the functionality of this system by using a sparse sensor array to locate gunshots and animal calls from distances of more than 1 km. Long range acoustic positioning of vocal signals has rarely been tested on land as the majority of studies relying on acoustic localization have focused on species with relatively short-range vocalizations (McGregor et al. 1997; Bower and Clark 2005; Ali et al. 2009; Collier et al. 2010). The CARACAL system is intended to provide researchers and wildlife managers with a cost-effective tool to study and monitor large ecosystems remotely.

Materials and Methods

Sensor design

CARACAL circuit boards consisted of an ARM (Advanced RISC Machines) M4 Cortex microcontroller, interfaced to four PDM (digital) MEMS (microelectro-mechanical systems) microphones (64 dBA SNR and -32 dB FS sensitivity based on 94 dB, 1kHz signal), placed at the cardinal compass points on an 8.4 cm diameter circle. Operating power consumption is approximately 375 mW. Data are saved to micro-SD cards and accurately timestamped using GPS PPS signals. The prototype unit costs under £150 to fabricate with hardware designs and firmware openly available at: <https://github.com/OpenWild>.

Signal processing pipeline

A complete software pipeline was developed to automatically process audio files recorded by the CARACAL hardware (Fig. 1). Acoustic events are extracted from the raw multichannel audio based on a threshold of magnitude squared coherence (MSC) between pairs of channels. The MSC between two signals (x and y) is defined by:

$$C_{xy}(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)}$$

where $S_{xy}(f)$ is the cross spectrum density and $S_{xx}(f)$ is the power spectral density of signal x (Ramírez et al. 2008). The quadrophonic (four separate channels) audio per station are fused into a single (mono) channel using generalized cross correlation (GCC) beamforming (Knapp and Carter 1976). Beamforming involves coherent addition of channels by estimating the relative delay (equivalent to a phase shift) per microphone. Signals correlated across all four microphones are enhanced by superposition, whereas uncorrelated signals (e.g. white noise, wind) are non-coherently summed. In an ideal case, the overall signal to noise ratio is boosted, for a four-microphone array, by a factor of 4 (6dB). However, the microphones on the CARACAL circuit board are spaced so closely that many sound pressure fluctuations in the audio range will be correlated to some degree which will reduce the actual boost in signal to noise ratio.

Extracted events are then matched across all other station audio files to determine the relative time-delay with a metric measuring the confidence of the match. Matches are determined based on the

spectral envelope of a region of sound in which a corresponding signal is most likely to occur (considering the known positions of each logger). Thereafter, matched events are passed to a locator function which uses time of arrival differences to output an estimated position of the event. This is done by selecting a random point in a 2-D plane and calculating the ToAs that would be expected from each station and subsequently comparing them to the actual ToAs with an estimation of likelihood. This process is repeated many times (e.g. 1000) and the best location and likelihood retrieved for the position estimate. A low log-likelihood (LLH) value (e.g. < 0) indicates a high degree of uncertainty in the estimated position whereas a high LLH (e.g. 2-4) suggests strong confidence in the estimate. All functions are contained in a Python library and demonstrations made with Python notebooks which are openly available in the GitHub repository at <https://github.com/OpenWild>

Field deployment

Each CARACAL circuit board was mounted on a 3D printed plastic frame, fixed within a simple clip-seal plastic container (Fig. 2). We drilled three rows of 3 mm holes around the side of the container to allow free sound signal propagation to the microphones and secured polyester wadding material against the inner sides to reduce noise from wind. A 30 cm diameter, 3 mm thick acrylic disk, mounted on a 40 mm wooden cube block was fixed to the top of the container to provide protection from rain and direct sunlight. The containers, housing the circuit boards, were then mounted on top of 5 m high telephone poles (Fig. S1 - Supporting Information).

We deployed a total of 8 CARACAL recording stations in a grid layout across a $\sim 6 \text{ Km}^2$ study area in the south western section of the Bulyebe Valley Conservancy (BVC), a privately managed wildlife area in southern Zimbabwe (Fig. S2 - Supporting Information). The BVC hosts a large number of vocally active mammals such as African lion (*Panthera leo*), spotted hyaena (*Crocuta crocuta*), African bush elephant (*Loxodonta africana*), Cape buffalo (*Syncerus caffer*) and chacma baboon (*Papio ursinus*) making it an ideal study site to test the functionality of the CARACAL system with long-distance animal vocalizations. Each station was separated by a distance of $\sim 500 \text{ m}$ from the nearest neighbouring

station. Care was taken to ensure all stations were horizontally orientated with one microphone consistently facing North. While CARACAL circuit boards can be powered with a variety of battery options (3-12 V), we used a battery bank consisting of six AA 2500 mAh rechargeable batteries in series (~8 V) which was secured on top of the plastic container, below the acrylic disk. Audio data were written to a 64 Gb microSD card which could store 48 hours of continuous recording (44.1 kHz, 16 bit, 4 channel).

Case Study: Gunshot localization

This study aimed to test the accuracy of sound signal localization using the CARACAL system. Gunshots were chosen as the test signal as the characteristic gunshot blast is easily identifiable and propagates over long distances. Additionally, gunshots often reveal the presence of illegal poaching activities in wildlife areas where early detection and accurate localization could reduce reaction times to such incidents.

Using a LM-5 assault rifle, we fired 28 controlled gunshots along roads with a spacing of ~250 m between shot locations. For each shot, GPS coordinates were recorded using a handheld Garmin etrex GPS. All recordings were downloaded and imported into Audacity software, version 2.1.1 (Audacity Team 2019), where the arrival time of each gunshot was marked and recorded for each station by visual inspection of spectrograms. The ToA data were passed through the CARACAL localization algorithm with the Euclidean distance in two dimensions between the estimated and true location calculated for each shot. When assessing localization accuracy, it is necessary to distinguish between shots originating from inside and outside the convex hull (CH) of the array as signals outside the array are generally localized poorly in comparison to those within the array. Since the localization results were not normally distributed based on a Shapiro-Wilcoxon normality test, we tested whether accuracy differed between the two regions by carrying out a Mann-Whitney U-test.

We also sought to investigate the localization performance using ToA data from between 3 and 7 stations as sparser deployments are likely to be more cost effective but may impact localization

performance. For every combination of each number of stations, the localization error was calculated for each gunshot occurring within the CH of the tested array. A cumulative distribution function of the gunshot localization errors was plotted for each number of stations.

Case Study: Animal call localization

This study was based on a lion predation event, where a pride of 6 lionesses attacked, and later killed, a Cape buffalo within the study area. We aimed to demonstrate the ability of the CARACAL system to localize a variety of long-distance animal calls and to compare manual annotation of calls with the automated processing pipeline.

Only four CARACAL stations were deployed on the night of the predation event (stations 7, 8, 10 and 14) but still provided sufficient data for localization. We selected 300 seconds of continuous audio starting from the beginning of the event (first bellow of the buffalo in distress). We also included a 100 second period of audio from ~ 40 minutes after the start which contained additional long-distance animal calls. Following the method stated above for gunshot localization, spectrograms of the audio associated with the predation event were labelled manually with the arrival times of each animal call. Each call was also manually classified according to species. Using the CARACAL localization algorithm and ToA data, position estimates for each call were mapped relative to the array and surrounding landscape features (river and conservancy boundary).

Using the manually labelled data as a benchmark, we then assessed the performance of the CARACAL audio processing pipeline. All audio files were passed through the pipeline for event detection, matching and localization. We first tested the performance of the CARACAL signal detector (coherence detector) using a receiver operating characteristic (ROC) analysis and included a naïve energy detector to provide a comparative measure of performance. Energy detectors are often used with more typical, single-channel recordings. We then tested the efficacy of the locator function using a high LLH threshold ($LLH > 0$) in order to minimize the number of incorrectly localized events (false positives). The resulting output of localized events were compared with the manually labelled audio.

Results

Gunshot localization

Time synchronized audio data were collected from seven out of the eight stations as station 6 malfunctioned during the test. Gunshots originating within the CH (convex hull) of the functioning 7-station array were localized with significantly lower errors than those from outside the CH ($U = 32.0$, $n_{\text{inside}} = 12$, $n_{\text{outside}} = 16$, $p < 0.01$). The mean localization error for gunshots inside the CH was 33.2 ± 15.3 m (mean \pm 1 sd), while shots outside were localized with a mean error of 83.3 ± 58.1 m (mean \pm 1 sd). The median localization errors for each region were 32.0 m and 61.8 m respectively. Locations of the true and estimated locations of gunshots were plotted relative to the array stations along with the distributions of the errors in each region (Fig. 3).

When all 7 functional stations were used to localize gunshots within the convex hull of the array, 100% of shots were localized within 100 m of the true location. When stations were experimentally reduced to 6, 5, 4 and 3, the percentage of shots localized within 100 m reduced to 99%, 89%, 80% and 53% respectively (Fig. S3 - Supporting Information).

Animal call localization for predation event

Manual processing

Detailed manual inspection of the audio associated with the predation event revealed 64 distinct animal calls in the vicinity of the array. Twenty-four of these calls were identified as Cape buffalo bellows, 29 as chacma baboon alarm calls and 11 as spotted hyaena calls (Examples can be found in the Supporting Information, Audio 1-3). A temporal progression of the localization of all 64 calls (Fig. 4) showed an initial cluster of buffalo bellows closely distributed around co-ordinates $x = 250$ m, $y = 900$ m. Less than two minutes later, two clusters of baboon alarm calls were estimated to have occurred near station 7 (600 m, 950 m) and ~ 100 m south of station 14 (0 m, -50 m). Two separated bouts of spotted hyaena calls were then recorded, one at a time of 3 minutes and another at 40 minutes. The first bout was localized south east of the array (900 m, -250 m) and the second outside

the Conservancy boundary (-900 m, -100 m). Although no ground truth was available against which estimated positions could be compared, a fine scale plot of the estimated positions of each buffalo bellow showed these locations to be distributed within ~ 70 m of the carcass (Fig. S4 - Supporting Information).

Automated processing

Using the manually labelled dataset as a benchmark, we assessed the performance of the automated processing pipeline. The ROC analysis (Fig. 5) revealed high discriminatory power for the coherence detector (area under the curve (AUC): 0.95) against comparatively poorer discriminatory power for the energy detector (AUC: 0.52). The events detected by the coherence detector were then passed to the signal matching and locator function. With a LLH threshold > 0 , the locator function produced position estimates for 53.1% of the manually labelled events. Of these positions, 88.2% were found to have been localized correctly. Overall, the automated processing pipeline was able to detect and accurately localize 46.9% of the true acoustic events (Fig. 6) Manual classification of these events showed species specific recall of 83.3% for buffalo, 13.8% for baboon and 54.5% for spotted hyaena. The processing rate for the automated pipeline is approximately 3 minutes of 4-channel audio per minute. For an 8-station array, the system would take approximately 2 hours 40 minutes to process a 1-hour period of audio which is equivalent to 32 channel hours of audio (8 stations x 4 channels x 1 hour).

Discussion

A major factor limiting the scalability of PAM systems has been the hardware costs associated with commercially available audio recording products (Gibb et al. 2018; Mennill et al. 2012). Only recently have cheaper, opensource devices, such as the Audiomoth, become accessible to researchers and facilitated larger deployments (Beason et al. 2018; Hill et al. 2018). However, the main focus of Audiomoth to date has been on quantifying the acoustic soundscape and automated signal detection rather than localization, due to its lack of precise time-synchronization across loggers. At a cost price

of ~ £150 per unit, the CARACAL system fits into this category and provides weak signal detection and localization through the use of a time-synchronised, multi-microphone hardware design. Using a sparse array of CARACAL stations, we found that relatively accurate acoustic localization can be achieved over large ranges (> 1 km), offering valuable information on disturbance events, such as gunshots, and animal species' presence and distributions.

Although gunshot localizing and reporting technology already exists in modern cities (e.g. "ShotSpotter" (ShotSpotter 2019)), their size, cost and power requirements prohibit their use in remote, natural environments where the necessary supporting infrastructure does not exist (Hill et al. 2018). Several studies have tested the application of small, autonomous audio recorders to gunshot detection as the use of high powered guns for poaching is relatively common occurrence in many wildlife areas (Astaras et al. 2017; Hill et al. 2018; Wrege et al. 2017). Using data from the seven functioning stations, and with no information on environmental variables (eg. temperature and humidity), the CARACAL system was able to localize gunshots occurring inside the array within 33.2 m of the true location on average. The majority of studies that have conducted acoustic localization have done so using station separation distances of between 15 and 75 m and have reported localization errors of less than 3 m (McGregor et al. 1997; Bower and Clark 2005; Mennill et al. 2006; Ali et al. 2009). While closely spaced detectors may be favourable for some species in small areas, such configurations would not be practical or cost-effective for monitoring large areas (e.g. >1000 km²). Our results are, however, similar to those reported by Kershenbaum et al. (2019) who used commercially available recording devices spaced 1-3 km apart to localize wolf howls and achieved an average distance error of 83 m. Although the localization errors for sparse deployments are likely to be considerably higher than those for dense arrays, highly accurate (< 3.0 m error) localizations may not be necessary for monitoring species with large home ranges.

Minimising the number of stations available in an area to detect and locate animal calls is likely to be a priority when designing sparse acoustic arrays. Although a minimum of three detectors is required for localization, results from the simulated removal of stations showed considerable improvements in

localization accuracy when ToA data from more than three stations were available. The relatively poor localization performance observed for three stations is likely a result of the dual localization solutions produced for some combinations of arrival time differences in this configuration (Spencer 2007). Acoustic localization systems that are dependent on ToA would therefore benefit from deployment patterns that enable signal detection from at least four stations. In order to reduce array density, the use of techniques that facilitate detection of signals from greater distances would be advantageous. The multi-microphone design used in CARACAL enabled the use of beamforming; a method which enhances the signal to noise ratio of sound from certain directions, which effectively increases the acoustic detection area (Mellinger et al. 2007). In addition to this feature, The CARACAL system utilizes a coherence detector which provides a significant improvement over standard signal strength (energy) detection where it can be difficult to set detection thresholds. Signal coherence is also more consistent across varying degrees of signal strength and is especially effective for weak signals (Fig. 7). It is important to note that signal coherence relies on the availability of signals from two or more spatially close microphones which is a core design feature of the CARACAL hardware. A further benefit of the quadrophonic microphone design is the ability to estimate the Angle of Arrival through the delay matrix estimated during the beamforming process. This can be used in combination with ToA to reduce density further and enhance the capacity to distinguish between calls from multiple sources, offering a significant improvement over single sensors for point count applications. We therefore suggest this as a promising future direction of research.

Many species actively emit vocal signals for communication, and by doing so, offer information on their presence at a particular location (McGregor 1993). We demonstrate that long-range localization of animal calls can provide valuable information on biological events, such as predation, which would otherwise be difficult to detect. Although the CARACAL audio processing pipeline demonstrated good detection performance, it failed to accurately localize many of the baboon alarm calls and a whole bout of spotted hyaena calls. The majority of the buffalo bellows were, however, localized accurately. The poor performance for baboon calls is likely related to the short, repetitive vocal behaviour of the

species which creates challenges when matching signals from each station as one call may be incorrectly matched with another. Furthermore, different species' calls frequently overlapped with each other which also impacted the matching of calls between stations. The pipeline is likely to function effectively when animal calls do not overlap with those of conspecifics or those from other species. The presence of overlapping signals is, however, a common issue for most extant detection algorithms which leads to low recall (Gibb et al. 2018). Despite these limitations, the CARACAL processing pipeline can act as an effective initial filter by scanning through many hours of audio and identifying locations with high vocal activity which can then be investigated further. The alternative approach of manually processing long periods of raw audio would be unfeasible given the time and effort required to locate and match signals across several stations. The speed and functionality of the processing pipeline can also be enhanced through the use of parallel computing and cloud technologies (Ahmad et al. 2018; Varghese and Buyya 2018), as well as with the addition of sound classification techniques that assign species labels to extracted events (Mielke and Zuberbühler 2013). Acoustic localization allowed us to recreate a biological event by offering considerably more contextual information than would be available from other forms of monitoring technology used in wildlife research such as single channel acoustic recorders (Whytock and Christie 2017; Hill et al. 2018) and camera traps (Kays et al. 2011). Data obtained from such events could be used to investigate a variety of animal behaviour patterns including inter- and intraspecific interactions and competition, habitat use and species vocal communication. Our study also highlights the potential for acoustic detection and localization to complement conventional GPS tracking which still remains a popular method for studying animal movement behaviour and habitat use (Kays et al. 2015). The procedures required to fit GPS tracking devices to study animals are often invasive and can require substantial resources and veterinary expertise thereby limiting the number of individuals that can be tagged. The CARACAL system can overcome some of these challenges by facilitating remote localization, over large areas, for any vocalizing individual. Kershenbaum et al. (2019) showed this to be possible for wolves

in Yellowstone National Park but used costly commercial recording devices which would likely limit scalability.

While acoustic localization may be advantageous in several ways, there are several limitations with this approach. Firstly, localization is reliant on the emission of a sound signal from the animal of interest and therefore may not be a suitable method where regular and consistent location information is required or where study species do not vocalize at all. Secondly, being able to identify individual animals is a fundamental requirement for studies investigating animal movements or estimating density from spatially explicit mark recapture. Several studies have reported finding unique vocal signals for individuals of certain species, however, vocal individuality remains largely untested for most taxa (Gibb et al. 2018; Mathevon et al. 2010; Soltis et al. 2005). Thirdly, and perhaps a more common problem, is that sound signals are often indistinct as a result of high noise levels and overlapping signals from other sources (Frommolt and Tauchert 2014). Further work is required to build accurate signal extraction, separation and classification algorithms that will provide useful data to end users.

Weaknesses associated with the CARACAL system in its current form are mainly related to the lack of on-board processing. With ~ 30 GB of raw data produced by each station every day, SD cards had to be replaced every second day which was costly and time consuming. This could be extended to a week or more by using larger SD cards (e.g. 512 Gbyte) but these typically cost more than the logger itself. Retrospective examination of the audio also resulted in large time lags between event occurrence and detection, limiting its use for reacting to time-critical events such as poaching. The weaknesses identified from this study will be used to inform further development of the CARACAL system. Future iterations will combine hardware and software developments that will facilitate on-device and in-network data processing. Acoustic detection and localization can be improved by incorporating angle of arrival and signal strength variables, and by accounting for sound wave diffraction around the array structure (Gillett et al. 2008). Coupled with power supply improvements, through integration with

solar panels for example, these enhancements will increase the autonomy and independence of the system and deliver real-time alerting.

Marques et al. (2013), suggested that future research on PAM systems should focus on, among others, the development of reliable and low-cost autonomous sensor arrays that are capable of ranging. Similarly, Kalan et al. (2016) noted that PAM systems could be improved by automated individual caller identification, acoustic localization and real-time data transmission. We have demonstrated the ability of the CARACAL system to address some of these gaps by providing cost-effective acoustic detection and source localization over large ranges using novel hardware configurations and data processing techniques. CARACAL can be customised and developed further or used in its current form to investigate a variety of biological questions where knowledge of the locations of sound sources are required.

Acknowledgements

We are grateful to the Ivan Carter Wildlife Conservation Alliance (ICWCA) and The Beit Trust for funding this research. We thank the shareholders and management staff of the Buby Valley Conservancy for giving us access to their property. We also thank W. Gush for his assistance with the set-up and maintenance of the CARACAL array.

Declaration of Interest

No conflict of interest is reported by the authors.

Data Accessibility

Due to its large size, the data that support the findings of this study are available from the corresponding author, MW, upon reasonable request.

References

Adi K, Johnson MT, Osiejuk TS. 2010. Acoustic censusing using automatic vocalization classification and identity recognition. *J. Acoust. Soc. Am.* 127(2): 874–883.

352 Ahmad A, Paul A, Din S, Rathore MM, Choi GS, Jeon G. 2018. Multilevel Data Processing Using
 353 Parallel Algorithms for Analyzing Big Data in High-Performance Computing. *Int. J. Parallel*
 354 *Program.* 46(3): 508–527.

355 Ali AM, Asgari S, Collier TC, Allen M, Girod L, Hudson RE, Yao K, Taylor CE, Blumstein DT. 2009. An
 356 empirical study of collaborative acoustic source localization. *J. Signal Process. Syst.* 57(3): 415–
 357 436.

358 Astaras C, Linder JM, Wrege P, Orume RD, Macdonald DW. 2017. Passive acoustic monitoring as a
 359 law enforcement tool for Afrotropical rainforests. *Front. Ecol. Environ.* 15: 233–234.

360 Audacity Team. 2019. Audacity(R): Free Audio Editor and Recorder [Computer application] version
 361 2.1.1. Available from: <https://audacityteam.org/>.

362 Beason RD, Riesch R, Koricheva J. 2018. AURITA: an affordable, autonomous recording device for
 363 acoustic monitoring of audible and ultrasonic frequencies. *Bioacoustics* 4622: 1–16.

364 Bower JL, Clark CW. 2005. A field test of the accuracy of a passive acoustic location system.
 365 *Bioacoustics* 15(1): 1–14.

366 Clark CW. 1980. A real-time direction finding device for determining the bearing to the underwater
 367 sounds of Southern Right Whales, *Eubalaena australis*. *J. Acoust. Soc. Am.* 68(2): 508-511.

368 Clark CW, Charif R, Mitchell S, Colby J. 1996. Distribution and behavior of the bowhead whale,
 369 *Balaena mysticetus*, based on analysis of acoustic data collected during the 1993 spring
 370 migration off Point Barrow, Alaska. *Reports Int. Whal. Comm.* 46: 541–552.

371 Collier TC, Kirschel ANG, Taylor CE. 2010. Acoustic localization of antbirds in a Mexican rainforest
 372 using a wireless sensor network. *J. Acoust. Soc. Am.* 128(1): 182–189.

373 Crouch WB, Paton PWC. 2002. Assessing the Use of Call Surveys to Monitor Breeding Anurans in
 374 Rhode Island. *J. Herpetol.* 36(2): 185–192.

375 Darras k, Pütz P, Fahrurrozi, Rembold K, Tschardt T. 2016. Measuring sound detection spaces for

376 acoustic animal sampling and monitoring. 201: 29-37.

377 Digby A, Towsey M, Bell BD, Teal PD. 2013. A practical comparison of manual and autonomous
378 methods for acoustic monitoring. *Methods Ecol. Evol.* 4(7): 675–683.

379 Frommolt KH, Tauchert KH. 2014. Applying bioacoustic methods for long-term monitoring of a
380 nocturnal wetland bird. *Ecol. Inform.* 21: 4–12.

381 Gibb R, Browning E, Glover-Kapfer P, Jones KE. 2018. Emerging opportunities and challenges for
382 passive acoustics in ecological assessment and monitoring. *Methods Ecol. Evol.* 10: 169–185.

383 Gibbs JP, Snell HL, Causton C. 1999. Effective Monitoring for Adaptive Wildlife Management: Lessons
384 from the Galapagos Islands. *J. Wildl. Manage.* 63(4): 1055–1065.

385 Gillett P, Johnson M, Carneal J. 2008. Performance benefits of sperical diffracting arrays versus free
386 field arrays, in: *IEEE International Conference on Acoustics, Speech and Signal Processing -*
387 *Proceedings.* pp. 5264-5267

388 Hatch L, Clark C, Merrick R, Van Parijs S, Ponirakis D, Schwehr K, Thompson M, Wiley D. 2008.
389 Characterizing the Relative Contributions of Large Vessels to Total Ocean Noise Fields: A Case
390 Study Using the Gerry E. Studds Stellwagen Bank National Marine Sanctuary. *Environ. Manage.*
391 42(5): 735–752.

392 Heinicke S, Kalan AK, Wagner OJJ, Mundry R, Lukashevich H, Kühl HS. 2015. Assessing the
393 performance of a semi-automated acoustic monitoring system for primates. *Methods Ecol.*
394 *Evol.* 6(7): 753–763.

395 Hill AP, Prince P, Piña Covarrubias E, Doncaster CP, Snaddon JL, Rogers A. 2018. AudioMoth:
396 Evaluation of a smart open acoustic device for monitoring biodiversity and the environment.
397 *Methods Ecol. Evol.* 9(5): 1199–1211.

398 Kalan AK, Piel AK, Mundry R, Wittig RM, Boesch C, Kühl HS. 2016. Passive acoustic monitoring reveals
399 group ranging and territory use: A case study of wild chimpanzees (*Pan troglodytes*). *Front.*

400 Zool. 13(1): 1–11.

401 Kays R, Crofoot MC, Jetz W, Wikelski M. 2015. Terrestrial animal tracking as an eye on life and
 402 planet. *Science*. 348: 1-9

403 Kays R, Tilak S, Kranstauber B, Jansen PA, Carbone C, Rowcliffe M, Fountain T, Eggert J, He Z, 2011.
 404 Camera Traps as Sensor Networks for Monitoring Animal Communities. *Int. J. Res. Rev. Wirel.*
 405 *Sens. Networks* 1(2): 811–818.

406 Kershenbaum A, Owens JL, Waller S. 2019. Tracking cryptic animals using acoustic multilateration: A
 407 system for long-range wolf detection. *J. Acoust. Soc. Am.* 145(3): 1619–1628.

408 Knapp CH, Carter GC. 1976. The Generalized Correlation Method for Estimation of Time Delay. *IEEE*
 409 *Trans. Acoust. Speech Signal Process.* 24(4): 320–327.

410 Mankin RW, Osbrink WL, Oi FM, Anderson JB. 2002. Acoustic detection of termite infestations in
 411 urban trees. *J. Econ. Entomol.* 95(5): 981–988.

412 Marques TA, Thomas L, Martin SW, Mellinger DK, Ward JA, Moretti DJ, Harris D, Tyack PL. 2013.
 413 Estimating animal population density using passive acoustics. *Biol. Rev. Camb. Philos. Soc.*
 414 88(2): 287–309.

415 Mathevon N, Koralek A, Weldele M, Glickman SE, Theunissen FE. 2010. What the hyena’s laugh tells:
 416 sex, age, dominance and individual signature in the giggling call of *Crocuta crocuta*. *BMC Ecol.*
 417 10(1): 1-9.

418 McGregor PK. 1993. Signalling in territorial systems: a context for individual identification, ranging
 419 and eavesdropping. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 340: 237–244.

420 McGregor PK, Dabelsteen T, Clark CW, Bower JL, Tavares JP, Holland J. 1997. Accuracy of a passive
 421 acoustic location system: empirical studies in terrestrial habitats. *Ethol. Ecol. Evol.* 9: 269–286.

422 Mellinger D, Stafford K, Moore S, Dziak R, Matsumoto H. 2007. An Overview of Fixed Passive
 423 Acoustic Observation Methods for Cetaceans. *Oceanography* 20(4): 36–45.

424 Mellinger DK, Clark CW. 2000. Recognizing transient low-frequency whale sounds by spectrogram
 425 correlation. *J. Acoust. Soc. Am.* 107(6): 3518–3529.

426 Mennill DJ, Battiston M, Wilson DR, Foote JR, Doucet SM. 2012. Field test of an affordable, portable,
 427 wireless microphone array for spatial monitoring of animal ecology and behaviour. *Methods*
 428 *Ecol. Evol.* 3(4): 704–712.

429 Mennill DJ, Burt JM, Fistrup KM, Vehrencamp SL. 2006. Accuracy of an acoustic location system for
 430 monitoring the position of duetting songbirds in tropical forest. *J. Acoust. Soc. Am.* 119(4):
 431 2832–2839.

432 Mielke A, Zuberbühler K. 2013. A method for automated individual, species and call type recognition
 433 in free-ranging animals. *Anim. Behav.* 86(2): 475–482.

434 Newson SE, Bas Y, Murray A, Gillings S. 2017. Potential for coupling the monitoring of bush-crickets
 435 with established large-scale acoustic monitoring of bats. *Methods Ecol. Evol.* 8(9): 1051–1062.

436 Noon BR, Bailey LL, Sisk TD, Mckelvey KS. 2012. Efficient Species-Level Monitoring at the Landscape
 437 Scale. *Conserv. Biol.* 26(3): 432–441.

438 O’Farrell MJ, Miller BW, Gannon WL. 1999. Qualitative Identification of Free-Flying Bats Using the
 439 Anabat Detector. *J. Mammal.* 80(1): 11–23.

440 Ospina OE, Villanueva-Rivera LJ, Corrada-Bravo CJ, Mitchell Aide T. 2013. Variable response of
 441 anuran calling activity to daily precipitation and temperature: Implications for climate change.
 442 *Ecosphere* 4(4): 1–12.

443 Payne KB, Thompson M, Kramer L. 2003. Elephant calling patterns as indicators of group size and
 444 composition: The basis for an acoustic monitoring system. *Afr. J. Ecol.* 41(1): 99–107.

445 Pinhas J, Soroker V, Hetzroni A, Mizrach A, Teicher M, Goldberger J. 2008. Automatic acoustic
 446 detection of the red palm weevil. *Comput. Electron. Agric.* 63(2): 131–139.

447 Pirotta E, Merchant ND, Thompson PM, Barton TR, Lusseau D. 2015. Quantifying the effect of boat

448 disturbance on bottlenose dolphin foraging activity. *Biol. Conserv.* 181: 82–89.

449 Ramírez D, Via J, Santamaría I. 2008. A generalization of the magnitude squared coherence spectrum
 450 for more than two signals: Definition, properties and estimation, in: IEEE International
 451 Conference on Acoustics, Speech and Signal Processing - Proceedings. pp. 3769–3772.

452 Riede K. 1998. Acoustic Monitoring of Orthoptera and its Potential for Conservation. *J. Insect*
 453 *Conserv.* 2: 217–223.

454 Sebastián-González E, Pang-Ching J, Barbosa JM, Hart P. 2015. Bioacoustics for species management:
 455 Two case studies with a Hawaiian forest bird. *Ecol. Evol.* 5(20): 4696–4705.

456 ShotSpotter. 2019. Home - ShotSpotter. ShotSpotter, USA. Available from:
 457 <https://www.shotspotter.com/> [Accessed 6 Apr. 2019].

458 Soltis J, Leong K, Savage A. 2005. African elephant vocal communication II: Rumble variation reflects
 459 the individual identity and emotional state of callers. *Anim. Behav.* 70(3): 589–599.

460 Spencer SJ. 2007. The two-dimensional source location problem for time differences of arrival at
 461 minimal element monitoring arrays. *J. Acoust. Soc. Am.* 121(6): 3579–3594.

462 Swiston KA, Mennill DJ. 2009. Comparison of manual and automated methods for identifying target
 463 sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers. *J. F.*
 464 *Ornithol.* 80(1): 42–50.

465 Varghese B, Buyya R. 2018. Next generation cloud computing: New trends and research directions.
 466 *Futur. Gener. Comput. Syst.* 79: 849–861.

467 Whytock RC, Christie J. 2017. Solo: an open source, customizable and inexpensive audio recorder for
 468 bioacoustic research. *Methods Ecol. Evol.* 8(3): 308–312.

469 Wildlife Acoustics. 2019. Wildlife Acoustics Online Store page. Wildlife Acoustics, Maynard. Available
 470 from <https://www.wildlifeacoustics.com/store> [Accessed 6 Apr. 2019].

471 Wilson DR, Battiston M, Brzustowski J, Mennill DJ. 2014. Sound Finder: A new software approach for

472 localizing animals recorded with a microphone array. *Bioacoustics* 23(2): 99–112.

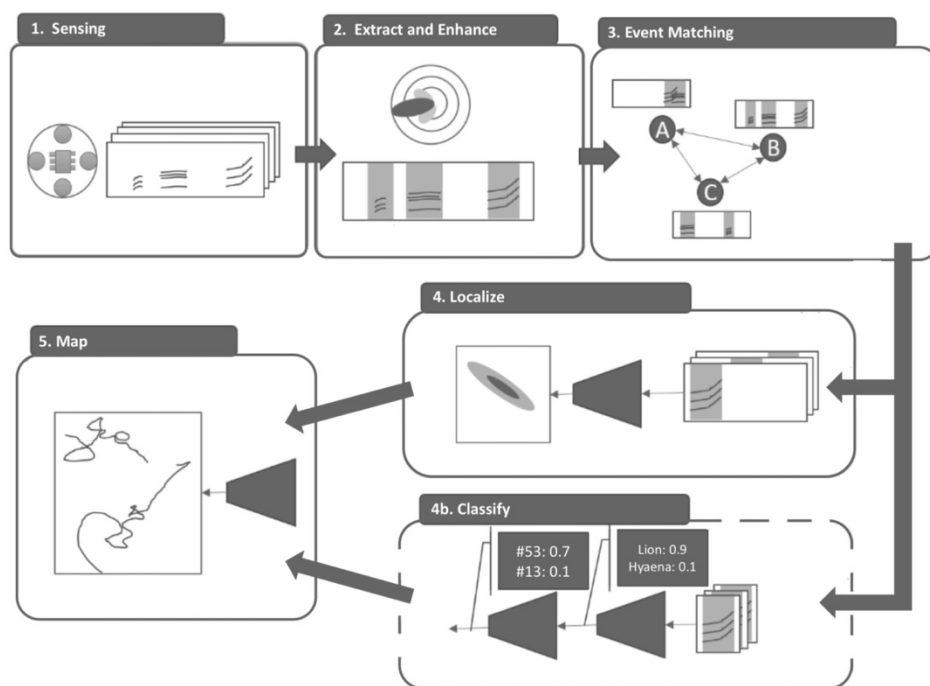
473 Wrege PH, Rowland ED, Keen S, Shiu Y. 2017. Acoustic monitoring for conservation in tropical

474 forests: examples from forest elephants. *Methods Ecol. Evol.* 8(10): 1292–1301.

475 Zilli D, Parson O, Merret, Geoff V, Rogers A. 2014. A Hidden Markov Model-Based Acoustic Cicada

476 Detector for Crowdsourced Smartphone Biodiversity Monitoring. *J. Artif. Intell. Res.* 51: 805–

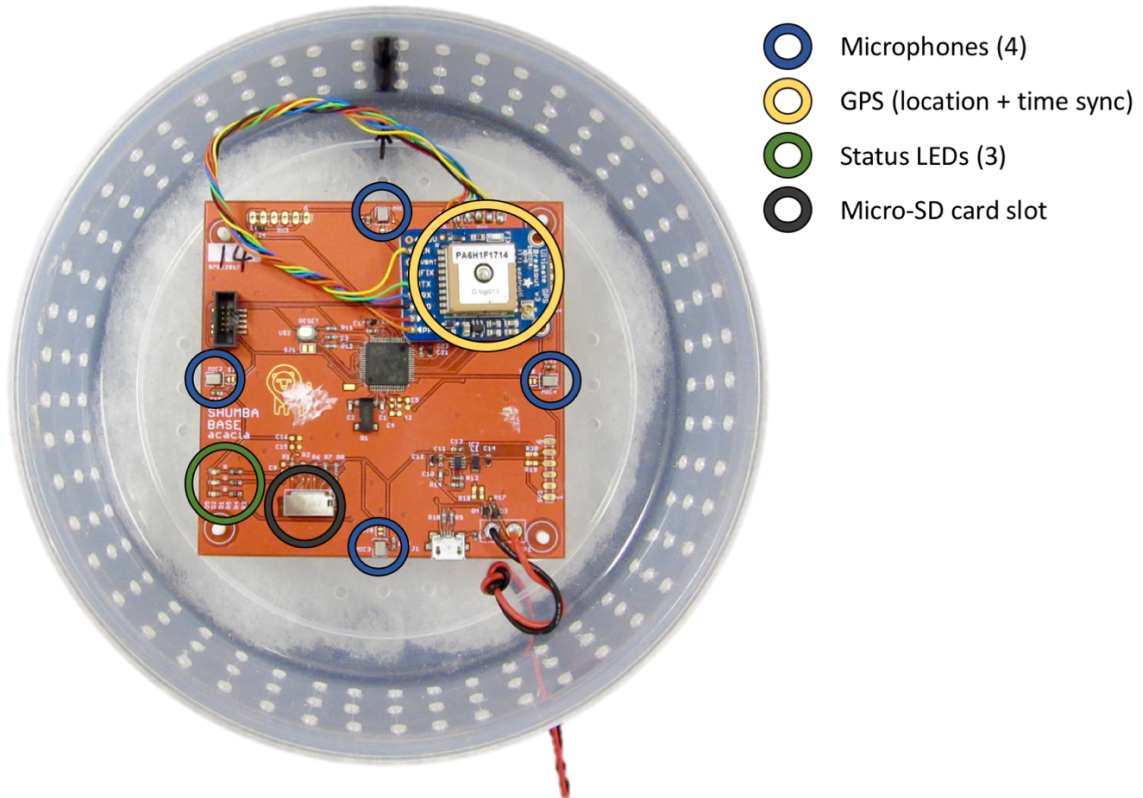
477 827.



478

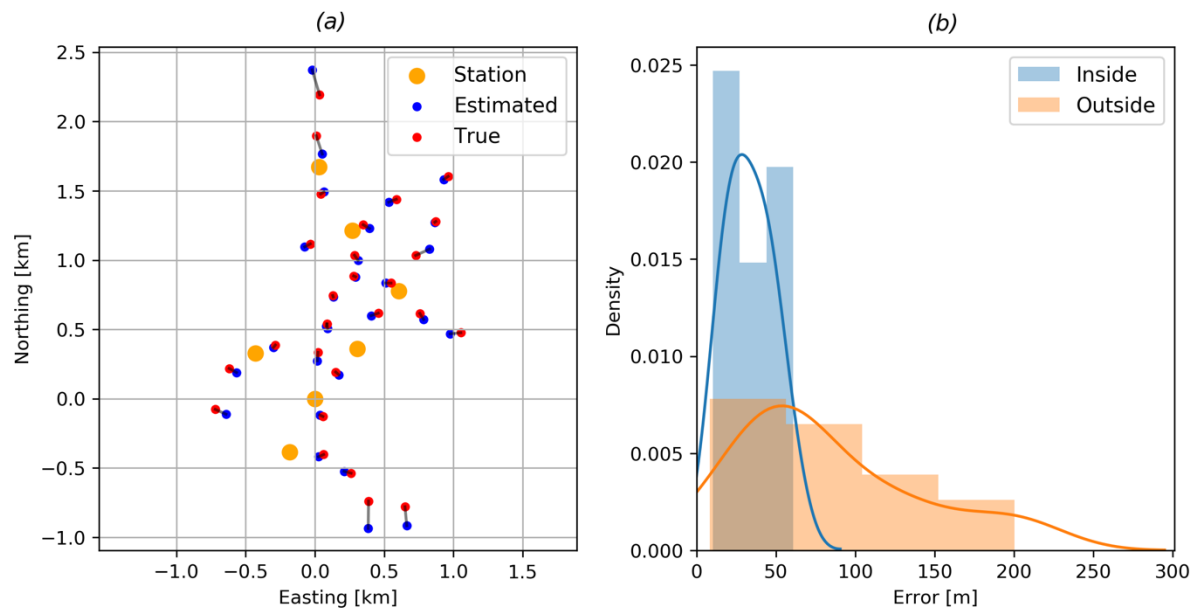
479 Figure 1. Flow diagram illustrating the sequential steps of the CARACAL audio processing pipeline. Step

480 4b can be included as an additional step in the processing sequence, but is not used in this study.



481

482 Figure 2. Image showing CARACAL circuit board with major components, secured within a plastic
483 container.



484

485 Figure 3. Relative positions of true and estimated locations of gunshots (a), and corresponding

486 distribution of localization errors for shots originating inside and outside the CH (b).

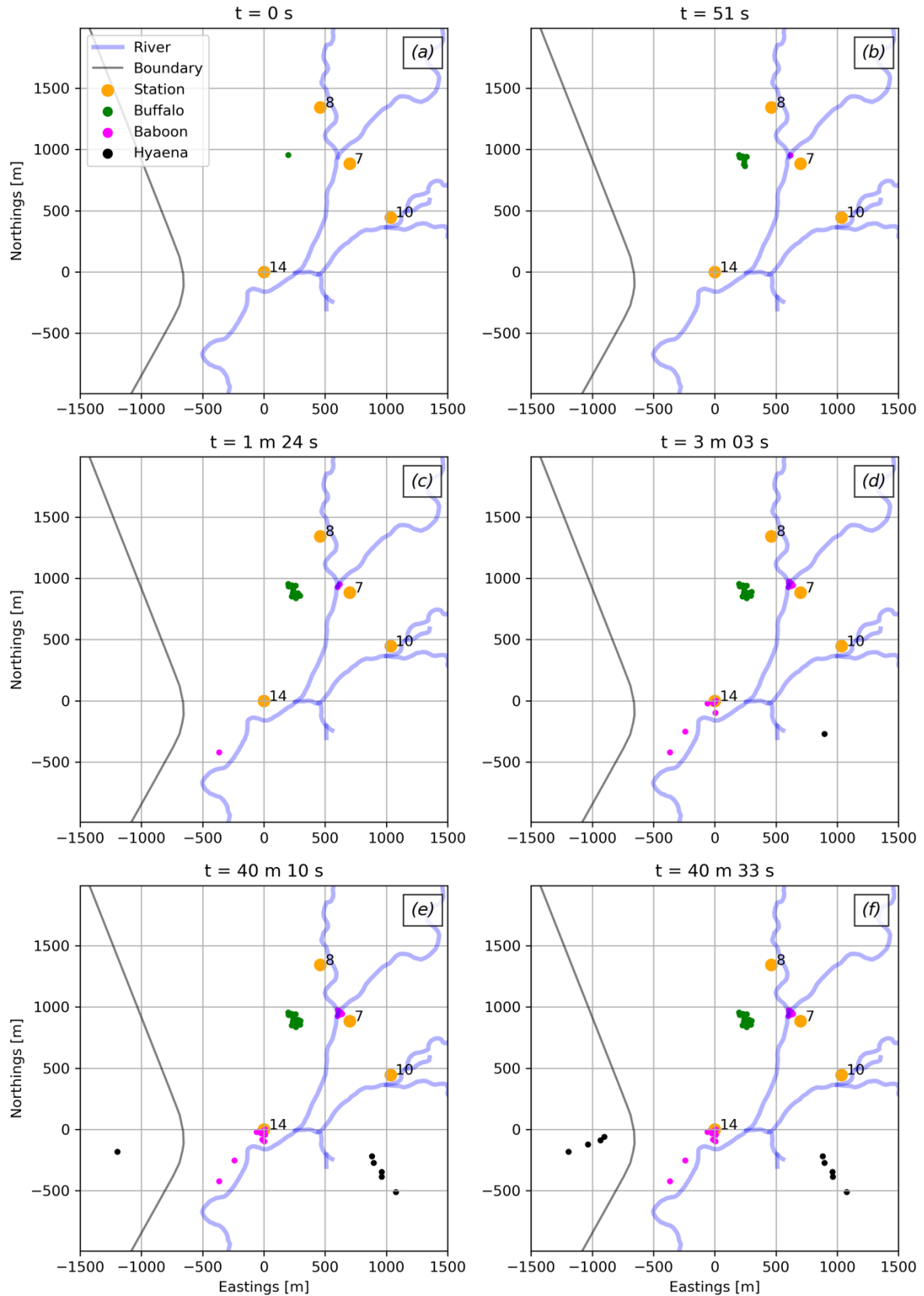
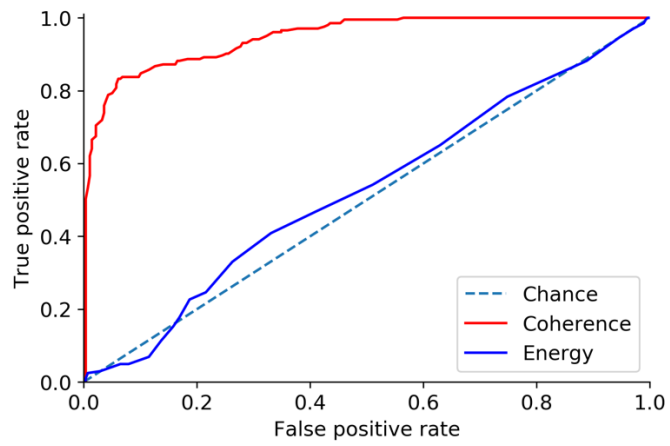


Figure 4. A series of maps showing overlaid localizations of animal calls: First bellows from buffalo being attacked by lions (a); First alarm call from baboon troop near station 7 (b); First alarm call from

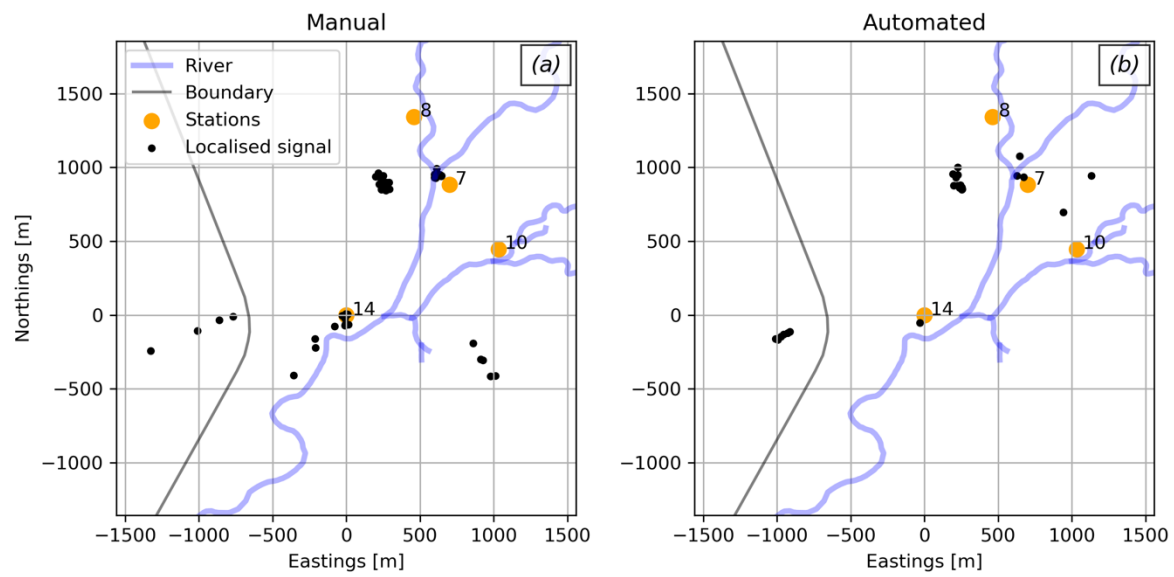
490 baboon troop near station 14 (c); First spotted hyaena call south east of the array (d); First spotted
491 hyaena call from a second animal outside the conservancy boundary, south west of the array (e); Last
492 call from spotted hyaena outside the boundary (f).



493

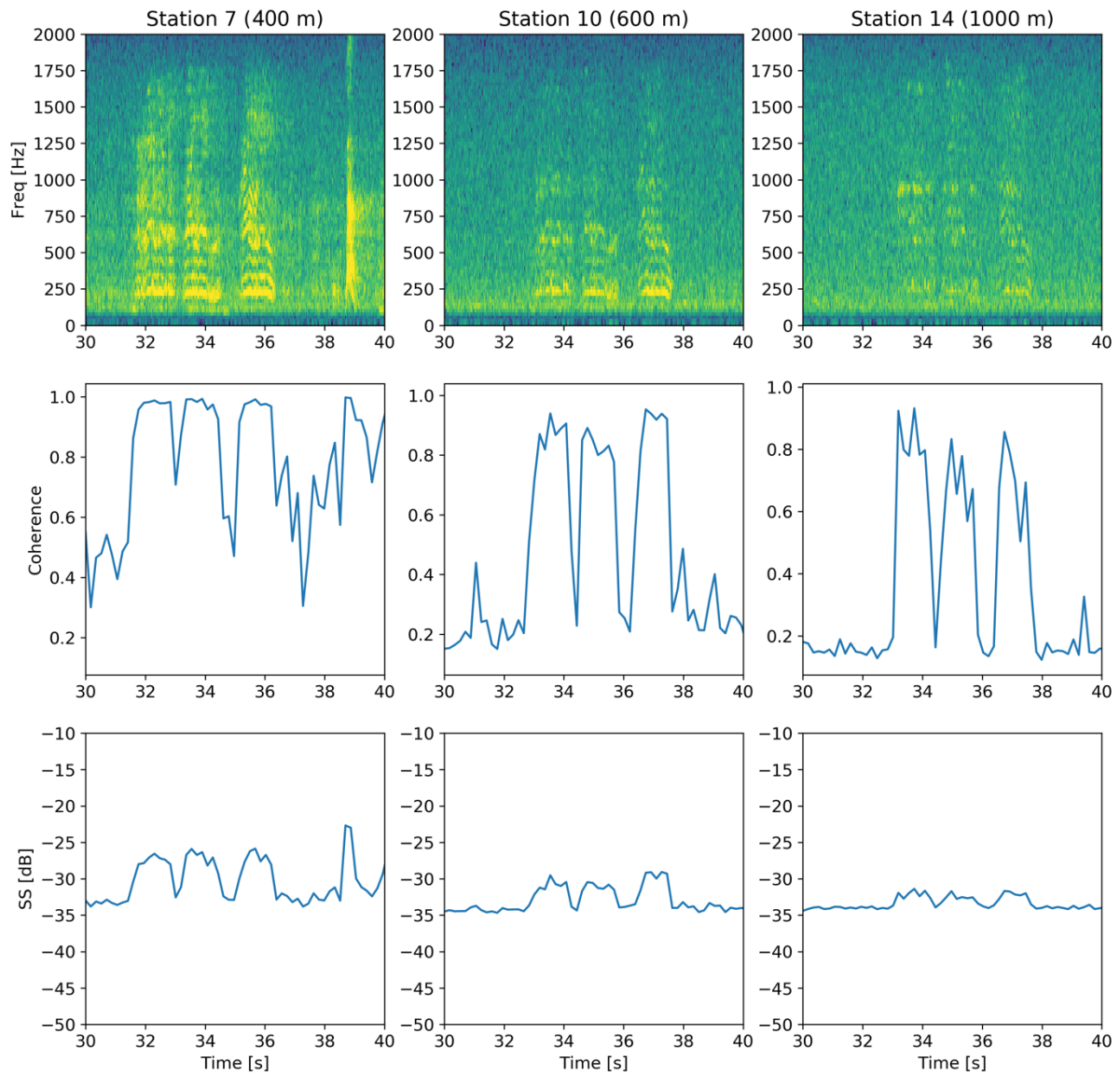
494 Figure 5. ROC curve comparing detection performance of the coherence detector used by the

495 CARACAL system against that of an energy detector.



496

497 Figure 6. Maps showing position estimates of animal calls identified manually (a) and using the
 498 automated signal processing pipeline (b).



499

500

501

Figure 7. Comparison between signal coherence and signal strength for progressively weaker buffalo call signals (left to right) recorded at three different stations with increasing distance from the animal.

Figure Captions

Figure 1. Flow diagram illustrating the sequential steps of the CARACAL audio processing pipeline. Step 4b can be included as an additional step in the processing sequence, but is not used in this study.

Figure 2. Image showing CARACAL circuit board with major components, secured within a plastic container.

Figure 3. Relative positions of true and estimated locations of gunshots (a), and corresponding distribution of localization errors for shots originating inside and outside the CH (b).

Figure 4. A series of maps showing overlaid localizations of animal calls: First bellows from buffalo being attacked by lions (a); First alarm call from baboon troop near station 7 (b); First alarm call from baboon troop near station 14 (c); First spotted hyaena call south east of the array (d); First spotted hyaena call from a second animal outside the conservancy boundary, south west of the array (e); Last call from spotted hyaena outside the boundary (f).

Figure 5. ROC curve comparing detection performance of the coherence detector used by the CARACAL system against that of an energy detector.

Figure 6. Maps showing position estimates of animal calls identified manually (a) and using the automated signal processing pipeline (b).

Figure 7. Comparison between signal coherence and signal strength for progressively weaker buffalo call signals (left to right) recorded at three different stations with increasing distance from the animal.

Supporting Information Captions

Figure S1. Image showing field deployment of CARACAL logger.

Figure S2. Map of the study area showing locations of CARACAL loggers. Logger 06 malfunctioned and is marked with a 'X'. Inset map shows location of BVC in Zimbabwe.

Figure S3. Cumulative distribution functions for the gunshot localization errors for each number of stations.

- 524 Figure S4. High resolution map showing estimated locations of buffalo bellows relative to the kill site.
- 525 Audio 1. Audio sample of buffalo bellows.
- 526 Audio 2. Audio sample of baboon alarm calls.
- 527 Audio 3. Audio sample of spotted hyaena calls.