INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Cost Effective Acoustic Monitoring of Bird Species

Ciira wa Maina

Dedan Kimathi University of Technology, Nyeri, Kenya.
ciira.maina@dkut.ac.ke

Abstract

Climate change and human encroachment are some of the major threats facing several natural ecosystems around the world. To ensure the protection of ecosystems under threat, it is important to monitor the biodiversity within these ecosystems to determine when conservation efforts are necessary. For this to be achieved, technologies that allow large areas to be monitored in a cost effective manner are essential. In this work we investigate the use of acoustic recordings obtained using a low cost Raspberry Pi based recorder to monitor the Hartlaub's Turaco in central Kenya. This species is endemic to East Africa and faces habitat loss due to climate change. Using simple features derived from the spectrograms of the recordings, a Gaussian mixture model classifier is able to accurately screen large data sets for presence of the Hartlaub's Turaco call. In addition, we present a method based on musical note onset detection to determine the number of calls within a recording.

Index Terms: Biodiversity monitoring, bird species recognition, Raspberry Pi.

1. Introduction

Ecosystems around the world face a number of threats including climate change and human encroachment. In order to protect these habitats, there is need to develop technologies to monitor biodiversity within ecosystems to detect any damage before it is too late. Traditional approaches to biodiversity monitoring involve the survey of species richness within the ecosystem of interest [1]. This requires trained experts to spend a significant amount of time in the ecosystem in order to carry out the survey. This approach is costly and time consuming and can not scale to the large number of ecosystems that require monitoring. This has led to efforts to develop more efficient approaches to biodiversity monitoring which include rapid biodiversity assessment (RBA) where the species survey is limited to a few indicator species [2]. While RBA is faster than a complete survey, it still requires an expert to go to the field and this is not always possible.

One alternative to traditional biodiversity monitoring approaches is to use acoustic signals recorded in the ecosystems of interest to try and infer the biodiversity of the ecosystem. Given an acoustic recording, one can in principle determine the species vocalising in the recording and thereby determine species presence and richness. The recordings can either be annotated by an expert or processed using species identification software to determine the species present in the recording. Acoustic approaches to biodiversity monitoring have been investigated by a number of authors and found to be a promising and scalable approach [3, 4]. Some of the advantages of this approach include: 1) Audio recordings can be archived to serve as a permanent record of the state of the ecosystems at the time of the survey. 2) Experts are not needed to collect the data as the recordings can be collected by people trained to handle the recording equipment only. 3) Acoustic recordings can be used to perform analysis at various levels starting from a large scale biodiversity assessment [3, 4] to a survey of a particular species [5].

Despite these advantages, a number of shortcomings remain. Firstly, recording equipment can be costly to obtain and setup. Second, audio recording generates a large amount of data which can be time consuming to analyse especially if we must listen to these recordings manually. As a result, efforts to make bioacoustic approaches more cost effective and less time consuming are important. In [6], the authors describe the use of a low cost audio recorder to monitor diverse ecosystems. In [7] the authors describe a system to screen large datasets for the vocalisation of a particular species, the Screaming Piha (*Lipaugus vociferans*) in a tropical forest in French Guiana.

In this work we develop and test a low cost acoustic monitoring system based on the Raspberry Pi microprocessor for use in acoustic monitoring of bird species. The Raspberry Pi is a low cost programmable microprocessor with much of the functionality of a modern computer and this makes the devices suitable for both recording and processing of the recordings. We use the recordings to detect the presence of the Hartlaub's Turaco (*Tauraco hartlaubi*), a ubiquitous forest species with a distinct call. This species is endemic to East Africa and faces habitat loss due to climate change and human encroachment.

2. Data Collection

2.1. Study Area

The study was conducted at the 120 acre Dedan Kimathi University Wildlife Conservancy (DeKUWC) located at 0°23'17.0"S 36°57'43.2"E. The conservancy has three ecological zones namely open grassland, undisturbed indigenous forest and aquatic zones. See the map in Figure 2.

2.2. Equipment

The audio recordings were collected using a cheap microphone connected to a Raspberry Pi (RPi). To make the recordings, we use the open source sound processing software SoX. Figure 1 shows the acoustic sensor system in the lab.

2.3. Acoustic Survey Protocol

Four acoustic recorders were left at locations within the DeKUWC on 5th January, 2016 with recording starting between 10am and 12 noon. The acoustic recorder locations are labelled 1-4. The locations are shown on the map in Figure 2. The points were separated by approximately 100-200 meters and the recorders were left at these points for approximately 28 hours and were programmed to record for one minute at five minute intervals. This produced approximately 340 minute long

http://dx.doi.org/10.21437/Interspeech.2016-746

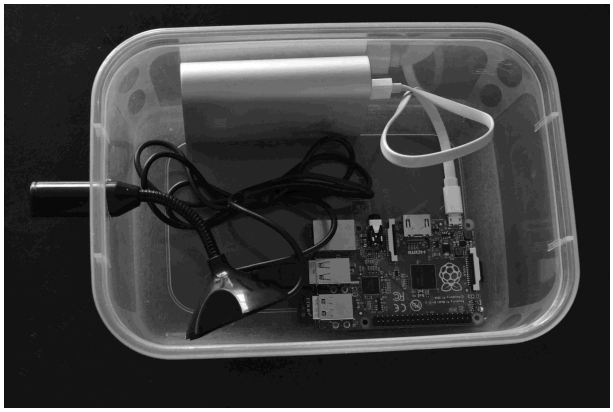

Figure 1: *The Raspberry Pi based acoustic sensor system.*

recordings per site. We set the sampling rate of the recorders to 16kHz at 16 bit resolution.

Figure 2: *A map of the DeKUWC with locations of the acoustic recorders indicated.*

2.4. Hartlaub's Turaco Dataset

From the recordings obtained from the fourth RPi recorder (location 4), 12 recordings were chosen which contained vocalisations of the Hartlaub's Turaco and 21 recordings without the Hartlaub's Turaco call but containing vocalisations from other bird species. Each of the recordings was split into six, ten second long recordings for use in training and testing our Hartlaub's Turaco call classifier.

3. Experiments

3.1. Acoustic Features

Mel frequency cepstral coefficients (MFCCs) are commonly used in speech processing applications and a number of authors have used them to recognise bird species from recordings [8]. In this work we use simpler features which have been used in music genre [9, 10] and acoustic scene classification [11, 12]. These features are: 1) *Band Energy Ratio*: This is the ratio of the total energy in a particular spectral band to the total energy of a frame. We use six logarithmically spaced spectral bands. 2) *Spectral Flux*: This is the squared difference between the normalized magnitude spectra of successive frames. We use a variation of the spectral flux by computing the flux in each of the six spectral bands 3) *Spectral centroid and bandwidth*: The spectral centroid is a measure of the frequency around which spectral energy is centered while bandwidth measures the dispersion of spectral energy around this centroid frequency. 4) *Spectral Rolloff*: This is the the frequency below which a given percentage of the spectral energy is contained. In this work we use 85%.

These features are computed from the spectrogram of the acoustic recording resulting in a 15-dimensional feature vector per frame. To obtain the spectrograms, we divide the signal into frames of 512 samples each (32 ms at 16kHz) with 50% overlap and compute the magnitude of the FFT of each frame. Figure 3 shows a typical spectrogram obtained from a recording containing the Hartlaub's Turaco vocalisation.

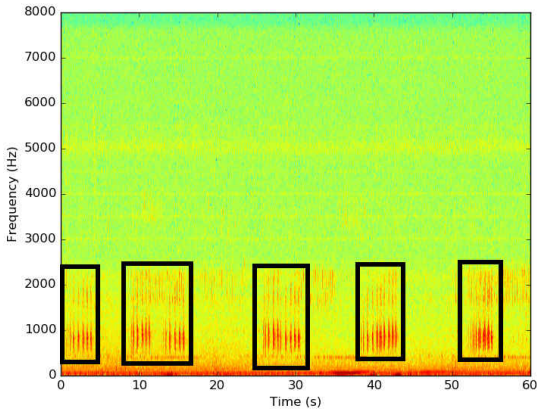

Figure 3: *A typical spectrogram obtained from a recording containing the Hartlaub's Turaco vocalisation. Regions with the vocalisation are indicated using the black rectangles.*

3.2. Classification

In order to classify the species, we train models using the features obtained from the acoustic recordings. We use Gaussian mixture models (GMMs) with 32 mixture coefficients and classification is by maximum likelihood. The GMMs are initialized by applying K-means clustering to the training data. The models and classifiers were implemented using Bob, an open source machine learning toolkit written in python [13].

3.3. Counting Hartlaub's Turaco Call Events

In addition to detection of recordings containing the call of the Hartlaub's Turaco, we are interested in counting the number of calls in a particular recording. This will indicate how long particular individuals are present in the vicinity of the recording device and also serve as a surrogate for the number of individuals present. In this work we use an approach similar to that used in note onset detection in music [14]. Here we first derive a detection function from the acoustic recording. Peaks in the detection function correspond to locations of the Hartlaub's Turaco call. The detection function we use is derived from the recordings using the following steps. 1) Divide the signal into frames of 512 samples each (32 ms at 16kHz) with 50% overlap

and extract the features used to classify the Hartlaub's Turaco call. 2) Compute the log likelihood ratio of each frame using the models trained to identify turaco calls. 3) Smooth this log likelihood ratio by computing the average log likelihood of the 300 previous blocks. This is so that we only consider blocks in a 5 second window which is the approximate duration of a call. 4) Fit a polynomial of degree 16 to this smoothed function and extract peaks that fall within the signal duration. These peaks correspond to the detected calls.

Figure 4 shows the spectrogram of an audio recording containing the Hartlaub's Turaco call and the detection function used to extract call locations. From this we see that the detection function (green curve) peaks at the end of the call and the polynomial fit (red curve) is able to detect the peaks of this function. In this case six call events are detected including one spurious detection as there are actually five call events.

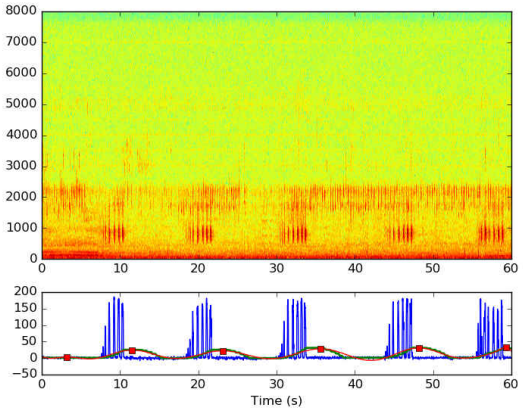

Figure 4: *The spectrogram of an audio recording containing the Hartlaub's Turaco call (top panel) and the detection function used to extract call locations (bottom panel).*

4. Results

4.1. Recognition of The Hartlaub's Turaco Call

We trained GMM models with 32 mixture coefficients to detect the presence and absence of the Hartlaub's Turaco call. The data were divided in half and one half used for training and the other half for testing. Figure 5 show the receiver operating characteristic of the GMM Hartlaub's Turaco call classifier. The area under the curve is 0.97 and the selected operating point corresponds to a true positive rate of 93% and a false positive rate of 7%. Code and data to reproduce the experiments are available on Github `https://github.com/ciiram/BirdPy`

4.2. Screening The Dataset

Using the GMM model trained using the training data and the threshold chosen for the classifier to operate at a true positive rate of 93% and a false positive rate of 7%, we run the GMM model on all the recordings obtained at location 4 to determine which files had the Hartlaub's Turaco call. For each frame, we compute the log likelihood ratio. As our detection function, we use the average log likelihood of the previous 300 frames (approximately 5 seconds) to classify the recording (see Figure 4). If the mean of the detection function is above the threshold,

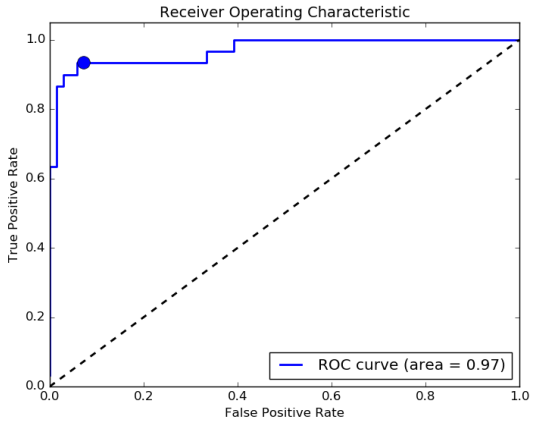

Figure 5: *Receiver operating characteristic of the GMM Hartlaub's Turaco call classifier.*

the recording is classified as having the Hartlaub's Turaco call. Of the 332 files screened, 122 were classified as having putative Hartlaub's Turaco calls including the 12 files used to train the classifier. Each one minute file was processed in approximately 5 seconds. Figure 6 shows the spectrogram of one of the files classified as containing the call. It is the spectrogram of the recording obtained at 06:15 on 6th January, 2016. Visual inspection of the spectrograms classified as containing the call showed that the classification was accurate.

We also compute the number of Turaco calls detected in each one minute file. Figure 7 shows a plot of the number of calls per minute detected at location 4 between 12:35 on January 5th, 2016 and 16:10 on January 6th, 2016. We see that calls are detected throughout the day at this location with the latest call detected at 18:45 and the earliest call detected at 06:05.

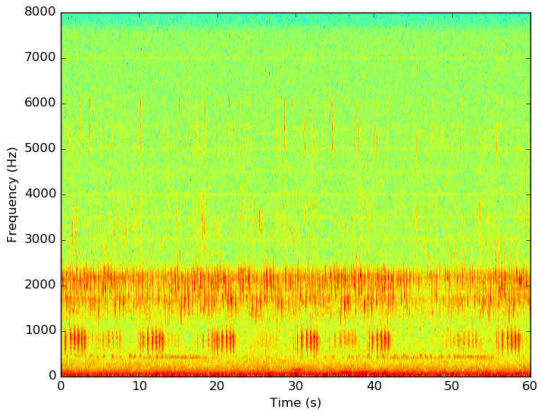

Figure 6: *The spectrogram of one of the files classified as containing the Hartlaub's Turaco call.*

5. Discussion and Conclusions

To form an accurate picture of the state of an ecosystem using acoustic recordings, it may be important to build recognizers for particular species known to be common in the ecosystem.

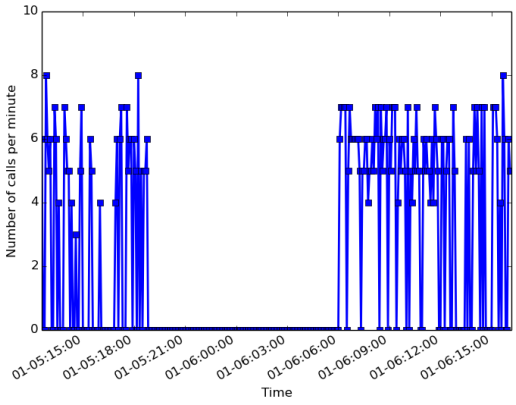

Figure 7: *Number of Hartlaub's Turaco calls per minute detected at location 4 between 12:35 on January 5th, 2016 and 16:10 on January 6th, 2016.*

This is an approach that has been successfully used in a number of studies such as detecting calls of the Screaming Piha in French Guiana [7] and detecting Nightjar calls in Northumberland, UK [5]. Here we successfully demonstrate the screening of a large audio dataset for calls of the Hartlaub's Turaco which is a ubiquitous species in montane forests in Kenya with a distinct call. The screening is based on a GMM classifier trained to achieve a true positive rate of 93% and a false positive rate of 7%. In addition to detecting presence of calls in a recording, we are able to detect the number of such calls. There are 332 one minute recordings obtained at location 4, approximately 5.5 hours of audio. We are able to screen each recording in less that 5 seconds with the entire dataset screened in approximately 20 minutes. Building automatic systems such as the one described here makes screening of these large datasets feasible.

6. Acknowledgements

The Kenya Education Network (KENET) provided financial support to purchase recording equipment. I would like to thank Mr. David Muchiri, the conservancy coordinator and all the wardens at the Dedan Kimathi University Wildlife conservancy especially Rasheed and Kimathi for support during fieldwork.

7. References

[1] J. H. Lawton, D. Bignell, B. Bolton, G. Bloemers, P. Eggleton, P. Hammond, M. Hodda, R. Holt, T. Larsen, N. Mawdsley *et al.*, "Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest," *Nature*, vol. 391, no. 6662, pp. 72–76, 1998.

[2] J. T. Kerr, A. Sugar, and L. Packer, "Indicator taxa, rapid biodiversity assessment, and nestedness in an endangered ecosystem," *Conservation Biology*, vol. 14, no. 6, pp. 1726–1734, 2000.

[3] J. Sueur, S. Pavoine, O. Hamerlynck, and S. Duvail, "Rapid acoustic survey for biodiversity appraisal," *PLoS One*, vol. 3, no. 12, p. e4065, 2008.

[4] N. Pieretti, A. Farina, and D. Morri, "A new methodology to infer the singing activity of an avian community: the Acoustic Complexity Index (ACI)," *Ecological Indicators*, vol. 11, no. 3, pp. 868–873, 2011.

[5] M. C. Zwart, A. Baker, P. J. McGowan, and M. J. Whittingham, "The use of automated bioacoustic recorders to replace human wildlife surveys: an example using nightjars," *PloS one*, vol. 9, no. 7, p. e102770, 2014.

[6] A. Farina, P. James, C. Bobryk, N. Pieretti, E. Lattanzi, and J. McWilliam, "Low cost (audio) recording (LCR) for advancing soundscape ecology towards the conservation of sonic complexity and biodiversity in natural and urban landscapes," *Urban ecosystems*, vol. 17, no. 4, pp. 923–944, 2014.

[7] J. S. Ulloa, A. Gasc, P. Gaucher, T. Aubin, M. Réjou-Méchain, and J. Sueur, "Screening large audio datasets to determine the time and space distribution of screaming piha birds in a tropical forest," *Ecological Informatics*, vol. 31, pp. 91–99, 2016.

[8] M. Graciarena, M. Delplanche, E. Shriberg, A. Stolcke, and L. Ferrer, "Acoustic front-end optimization for bird species recognition," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2010, pp. 293–296.

[9] G. Agostini, M. Longari, and E. Pollastri, "Musical instrument timbres classification with spectral features," *EURASIP Journal on Applied Signal Processing*, vol. 2003, pp. 5–14, 2003.

[10] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[11] S. Chu, S. Narayanan, and C. J. Kuo, "Environmental sound recognition with time–frequency audio features," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1142–1158, 2009.

[12] A. J. Eronen, V. T. Peltonen, J. T. Tuomi, A. P. Klapuri, S. Fagerlund, T. Sorsa, G. Lorho, and J. Huopaniemi, "Audio-based context recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 321–329, 2006.

[13] A. Anjos, L. E. Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel, "Bob: a free signal processing and machine learning toolbox for researchers," in *20th ACM Conference on Multimedia Systems (ACMMM), Nara, Japan*. ACM Press, Oct. 2012.

[14] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1035–1047, 2005.