ResearchGate

See discussions, stats, and author profiles for this publication at: http://www.researchgate.net/publication/276061802

Bioacoustic Approaches to Biodiversity Monitoring and Conservation in Kenya

CONFERENCE PAPER · MAY 2015

READS

73

1 AUTHOR:

Ciira wa Maina

Dedan Kimathi University of Technology

15 PUBLICATIONS **33** CITATIONS

SEE PROFILE

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.

Available from: Ciira wa Maina

Retrieved on: 12 November 2015

IST-Africa 2015 Conference Proceedings
Paul Cunningham and Miriam Cunningham (Eds)
IIMC International Information Management Corporation, 2015
ISBN: 978-1-905824-50-2

Bioacoustic Approaches to Biodiversity Monitoring and Conservation in Kenya

Ciira wa MAINA
Dedan Kimathi University of Technology, P.O. BOX 657-10100 Nyeri, Kenya,
Email: ciira.maina@dkut.ac.ke

Abstract: Kenya's rich biodiversity faces a number of threats including human encroachment, poaching and climate change. Since Kenya is a developing country, there is need to manage the sometimes competing interests of development, such as infrastructure development, and conservation. To achieve this, tools to effectively monitor the state of Kenya's various ecosystems are essential. In this paper we propose a biodiversity monitoring software tool that integrates acoustic indices of biodiversity, recognition of species of interest based on their vocalizations and acoustic census. This tool can be used by non-experts to determine the current state of their ecosystems by monitoring the state of bird species that serve as indicator taxa and whose abundance is related to the abundance of other terrestrial vertebrates including the "big five". The tool we propose exploits state-of-the art advances in signal processing and machine learning to perform biodiversity monitoring, bird species detection and census in a joint framework. Using publicly available data we demonstrate how current acoustic indices of biodiversity can be improved by incorporating machine learning based audio segmentation algorithms. We also show how open source toolkits can be used to build bird species recognition systems. Code to reproduce the experiments in this paper is available on Github at https://github.com/ciiram/BirdPy.

Keywords: Biodiversity, conservation, bird species recognition, open source software

1. Introduction

Biodiversity monitoring is important in determining the state of our environment and in guiding the use of scarce resources for conservation [1]. Traditional approaches to biodiversity monitoring involve the use of skilled experts with the ability to capture and determine the species present in an area (see for example [2]). This manual cataloguing is tiresome and can often only be applied in certain areas, often those already identified as suffering acute biodiversity loss. This has motivated efforts to develop effective methods to perform rapid biodiversity assessment (RBA) [3]. These include the selection of indicator taxa that are representative of species richness in a given area for sampling. The disadvantage of these methods is that they are invasive and still involve a huge amount of effort to sample the indicator taxa.

Since a large number of species produce vocalizations, recordings of these vocalizations can in principle be used to perform biodiversity assessment in a non intrusive and automatic way particularly when these vocalizations are produced by indicator taxa. A number of bird species serve as good indicator taxa which can be used to monitor the state of other terrestrial vertebrates in the same ecosystem [4]. Birds have the advantage of being easier to monitor than larger terrestrial vertebrates since even when they cannot be seen their vocalizations indicate their presence.

Sueur *et al.* [5] proposed the use of acoustic indices computed from recordings taken in the wild to perform biodiversity assessment. It has been observed that these acoustic measures are correlated to observed species richness [6]. In addition to biodiversity assessment, acoustic recordings can be used to identify the species responsible for the vocalizations present in the recording. A number of authors have used acoustic recordings to identify bird species [7, 8]. These approaches borrow from the well developed literature on speaker recognition using speech recordings and apply it to recordings of bird calls and songs. These systems derive features from the audio signal and use these features to identify the species in the recording. These features are often derived from the Short Time Fourier Transform (STFT) of the signal and include Mel frequency cepstral coefficients (MFCCs) which are popular in speaker recognition [9].

In order to train the classifiers, labelled training data is used and a number of such datasets are now freely available. For example, a recent bird species identification competition used data recorded in the H. J. Andrews Long-Term Experimental Research Forest in Oregon and contained recordings of 19 bird species [10]. A number of classification techniques are reviewed in [9] with most relying on hand crafted features derived from the audio signal. In the recent past, feature learning has been a key research area in machine learning. Feature learning involves the unsupervised extraction of relevant features directly from data and then using these features for classification. This approach has been successfully applied in audio classification and computer vision [11]. In recent work, an unsupervised feature learning method based on spherical k-means was applied to bird song recognition [8].

In addition to recognition of species present in audio recordings, a successful biodiversity monitoring tool needs to determine the number of animals. To do this, we can adapt technology developed for speaker diarization [12, 13]. This technology segments an audio recording according to 'who spoke when'. Given a recording of a meeting with several participants, speaker diarization systems segment the recording according to the speakers. In principle these systems can determine the number of speakers present in a recording. A number of approaches have been used to solve the clustering problem but the most successful are based on Gaussian Mixture Models (GMMs) with the number of clusters/speakers determined via the Bayesian information criterion (BIC). A significant recent advance was the use of i-vector features obtained from the audio recording as the input to clustering algorithm instead of traditional MFCC features [14]. An initial attempt at applying this technology to conduct census of bird species is reported in [15].

While significant progress has been made in automated biodiversity assessment and bird species identification, significant challenges remain. Firstly, current algorithms are susceptible to environmental noise and efforts to improve noise robustness are necessary. Secondly, the sampling of the species is sparse, for example comprehensive datasets for African bird species are lacking. This is particularly important since the tropics contain several endemic bird species. Finally, there is need to develop methods that scale with increase in data to allow us to deal with large datasets.

In this work we propose the use of acoustic indices of biodiversity in biodiversity monitoring by local citizens with the aim of empowering individuals to track the state of their ecosystems. Focussing on bird species which serve as indicator taxa [4], we show that traditional acoustic indices of biodiversity can be improved by segmenting the audio signal such that only areas with signals of interest are considered. We propose to create a system that incorporates acoustics indices, bird species identification and automatic census to achieve robust biodiversity monitoring. We also propose to integrate these systems with mobile technology to allow individuals without expert knowledge to perform biodiversity monitoring.

2. Objectives

The main objectives of this work are:

1. To design a mobile based software tool capable of biodiversity monitoring that incorporates traditional acoustic indices of biodiversity, automatic recognition of indicator taxa, in particular bird species and automatic methods of counting the number of vocalizing species given an audio recording.
2. To improve the performance of traditional acoustic indices of biodiversity by determining the sections of the audio signal containing vocalizations of interest.
3. To improve wildlife conservation efforts by developing technologies that allow non-experts to assess the state of their natural environment and determine when interventions are necessary.

3. Methodology

3.1 Acoustic Biodiversity Monitoring

In work by Sueur *et al* [5] it was shown that species abundance can be estimated in a non-invasive way by using acoustic indices. The main idea is that if we can determine the uncertainty associated with an audio recording, we can estimate the number of different species producing vocalizations in the recording. These indices are computed from sampled versions of the audio signal and the corresponding short time Fourier transforms (STFT). In information theory, the uncertainty associated with a random variable is given by its entropy. For a discrete random variable X drawn from an alphabet χ with probability mass function (p.m.f) $p(x) = \Pr\{X = x\}, x \in \chi$, its entropy is given by [16]

$$H(X) = -\sum_{x \in \chi} p(x) \log(p(x)).$$

When the logarithm is to the base 2, the entropy is measured in bits. The greater the entropy, the greater the uncertainty associated with the random variable. The entropy is bounded between zero and the logarithm of the cardinality of the set from which the random variables are drawn i.e. $0 \leq H(X) \leq \log|\chi|$. For example, for a binary random variable drawn from the set $\chi = \{0,1\}$, the entropy is bounded between zero and 1 bit.

To compute the uncertainty associated with the sampled audio signal s[n], we first compute the analytic signal of s[n] given by $\varphi[n] = s[n] + i s_h[n]$ where $i^2 = -1$ and $s_h[n]$ is the Hilbert transform of the s[n]. We then determine the probability mass function of the temporal envelope $P_t[n]$ as

$$P_t[n] = \frac{|\varphi[n]|}{\sum_{n=1}^{L} |\varphi[n]|}$$

where L is the length of the audio signal in samples. The temporal entropy is then given by

$$H_t = -\frac{1}{\log_2 L} \sum_{n=1}^{L} P_t[n] \log_2 P_t[n]$$

The factor $\frac{1}{\log_2 L}$ normalizes the temporal entropy to lie in the interval [0,1].

To compute the spectral entropy, we first compute the STFT of the audio signal using non-overlapping windows of size N, the mean spectral envelope is then determined and used to compute the spectral probability mass function. The spectral entropy H_s is then computed in a manner similar to the temporal entropy. Finally, the acoustic entropy index H is the product of the temporal entropy and the spectral entropy i.e. $H = H_s \times H_t$. It was shown in [6, 5] that these acoustic indices are correlated with species abundance and can be used for biodiversity assessment.

To improve the performance of the acoustic entropy index, we explored the segmentation of the audio signal into regions with and without vocalizations of interest. To test this we used the publicly available data released as part of the bird species recognition competition organised by the Machine Learning for Signal Processing (MLSP) workshop in 2013. These recordings were obtained in the H. J. Andrews Long-Term Experimental Research Forest in Oregon and contained recordings of 19 bird species (HJA dataset). Each of the recordings is 10 seconds long and contains vocalizations of up to six species. The recordings are sampled at 16 kHz with 16 bits used to represent each sample. To compute the spectral entropy a non-overlapping window size of 512 samples is used which corresponds to 32ms. In addition the audio signal are divided in regions with and without bird sounds using a random forest classifier to classify pixels of the spectrogram of each recoding into bird and not bird sounds [7]. We use this segmentation to determine regions of the audio recording that are dominated by bird sounds.

3.2 Bird Species Recognition

In this work we consider the automatic recognition of bird species based on audio recordings of their vocalizations. Most solutions to this problem consist of the following steps which are common to most classification problems in machine learning 1) feature extraction, 2) modelling and 3) classification. Appropriate features that are suitable for bird species recognition must be obtained from the raw audio recording. In this work we use Mel frequency cepstral coefficients (MFCCs) that are popular in speech processing applications. A number of authors have applied these features to bird species recognition with good results [17, 15].

The system we employ is based on the same Universal Background Model (UBM)-Gaussian Mixture Model (GMM) framework used in [17]. Here a UBM with 32 mixture components is trained using 19 dimensional MFCCs obtained from all files with multiple species in the HJA data set. Of the 645 recordings, 82 are recordings of single species. That leaves 563 files to train the UBM. Only frames that correspond to bird sounds are used to train the models. We use the open toolkit Bob [18] to compute the MFCCs and train the GMMs using K-means initialized expectation maximization. Bob is a versatile tool written in Python with underlying C++ implementation of key machine learning algorithms for speed. In this work we use Bob for all feature extraction, modelling and classification steps.

Once the UBM is trained, MFCCs from the individual species files are used to adapt this model via maximum a posteriori (MAP) estimation to obtain individual species models. Due to the limited amount of data we focus on 5 of the 19 species which have more than 7 single species recording each. The corresponding species IDs in the data set are 1, 6, 9, 10 and 18. For each species, 5 recordings are used to train the species specific GMM model.

For classification, MFCCs are obtained from the test recording and the corresponding species estimate is obtained by maximum likelihood. There are a total of 24 test recordings with between 2 and 7 per species.

3.3 Audio Diarization

Speaker diarization systems aim to assign segments of speech in an audio recording to the speakers that produced them. That is to solve the ‘who spoke when’ problem. These systems must segment the recording into regions with single speakers and then cluster these segments such that each cluster corresponds to one speaker. The most successful systems are based on agglomerative clustering where initially each segment is assigned an individual cluster. Clusters are then merged so that at the end only one cluster remains per speaker.

The system we propose here would incorporate an audio diarization system that would cluster segments according to the bird species that produced them in a straight forward extension of speaker diarization systems.

4. Technology Description

One of the aims of this work is to provide technology capable of biodiversity monitoring that can be used by non-experts. It is recognised that several biodiversity hotspots lie outside protected areas such as game reserves and it is therefore important to empower citizens with the tools to monitor their ecosystems, some of which are home to endangered species. For example the Hinde's Babbler is common in farm lands in Central Kenya [19].

The current generation of smart phones come with audio recording equipment and this together with their portability makes them ideal for acoustic biodiversity monitoring.

We envision that as prices of these devices continue to fall, they will become as common as feature phones and can be widely used even in rural communities.

The system illustrated in the schematic shown in Figure 1 consists of a smart phone loaded with appropriate software to process the audio signals in order to come up with estimates of species abundance. In this paper we examine the software that would be loaded on these smart phones. Future work will aim at creating a user friendly smart phone application that runs all this software. This will allow non-experts to use these systems with little or no training.

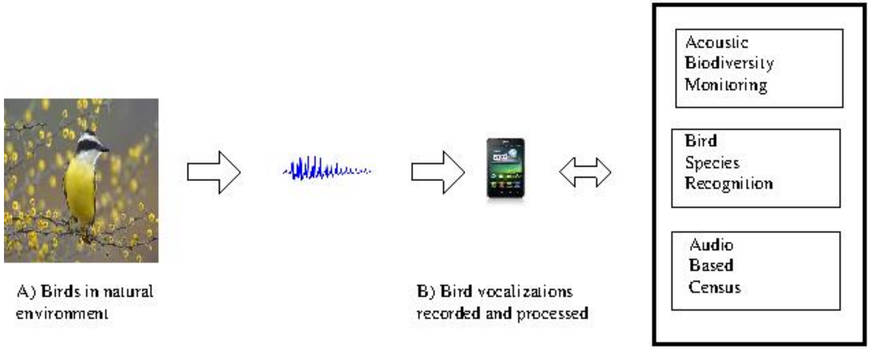
A) Birds in natural environment
B) Bird vocalizations recorded and processed
Acoustic Biodiversity Monitoring
Bird Species Recognition
Audio Based Census

Figure 1 Schematic diagram illustrating the proposed technology.
Birds serving as indicator taxa in their natural environment produce vocalizations (A) which are recorded and processed using appropriate software installed on a smart phone or computer (B)

5. Results

5.1 Acoustic Biodiversity Monitoring

For the proposed acoustic biodiversity monitoring technology to work well in real life environments, it is important to be able to segment audio signals into regions that contain and do not contain signals of interest. To do this, we determined the acoustic entropy index of the 179 recordings in the HJA dataset for which the number of species in the recording is available. We then determined the correlation of this acoustic index and the number of species present in the recording. Each recording has between one and six species.

The segmentation algorithm works by computing the spectrogram of the audio and treating it as an image. This spectrogram is computed using a window of 512 samples with 75% overlap between segments. Using 20 hand segmented recordings, a random forest classifier is trained to classify the spectrogram pixels as either bird or non-bird sound. Figure 2 shows the results of the segmentation algorithm on one of the files. From the frequency domain plot we see that regions containing bird sounds can be clearly seen. Moreover most of these regions are correctly classified as bird sound by the random forest classifier as seen on the top right panel of Figure 2.

This segmentation was applied to all the recordings in the HJA dataset and the acoustic entropy index computed for both the segmented and unsegmented audio files. Figure 3 shows a scatter plot of the acoustic entropy index H versus number of species present in the recording with and without segmentation. We also compute the mean acoustic entropy index for each number of species present. We see that with segmentation, the linear relationship between H and the number of species is more apparent. In fact when we compute the Pearson correlation coefficient between H and the number of species present in the recording for both cases we find that it is 0.89 when segmentation is applied and 0.17 without segmentation. This result clearly shows that for this technology to be useful, we must find the regions of the audio signal that contain vocalizations of the species of interest.

5.2 Bird Species Recognition

Using the species models trained as described in section 3.2, we examined the bird species classification performance for both individual audio frames and entire recordings. On the training data the frame classification accuracy was 90% while on the test data the performance decreased to just 37%. For the entire recordings, the classification accuracy was 68% on the training data and 50% on the test data. Code to reproduce these experiments is available on Github (https://github.com/ciiram/BirdPy).

The main reason for this low classification accuracy on the test data is likely to be lack of adequate training data to learn good species models. Since we use just five recordings per species each lasting 10s and with bird vocalizations for only a tiny fraction of the duration, the total number of frames per species is low (see Figure 2).

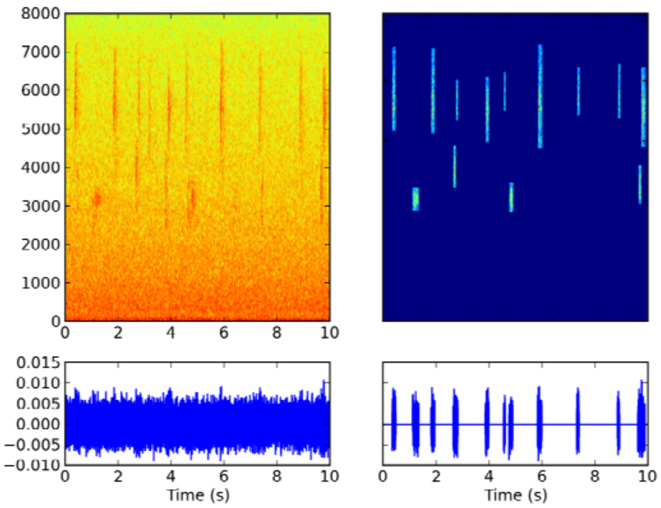

Figure 2: Spectrograms and time domain plots of the audio signal.
The panels on the right show the results of the segmentation algorithm.

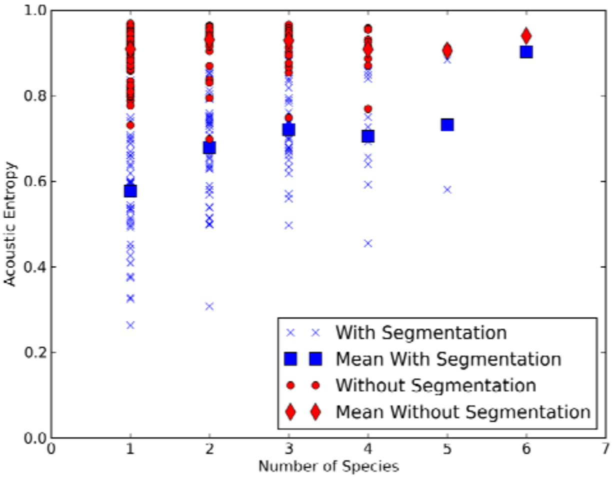

Figure 3: Scatter plot of the acoustic entropy versus the number of species with and without segmentation. The segmentation determines which sections of the audio signal correspond to bird calls.

6. Business Benefits

Technological solutions which allow us to monitor the state of our natural environment are necessary if developing countries are to industrialize in a sustainable manner. In addition, with greater awareness environmental resources can be exploited in a sustainable manner with the additional benefit of income generation.

Here we propose the use of mobile technology to monitor biodiversity in a non-invasive way using acoustic indices. We also propose a system which will be able to determine the abundance of bird species that serve as indicator taxa in biodiversity assessment. This work is likely to yield a number of financial benefits.

Firstly, with tools capable of identifying bird species, non-experts can determine if endangered species and other species of interest are present in their area. This information can be used to promote avi-tourism among both local and foreign tourists and this can serve as a lucrative source of income. Avi-tourism is one of the fastest growing tourism sub-sectors and attracts high end tourists. In addition, jobs for locals to act as guides will be created [20].

Secondly, one of the benefits of the introduction of a tool for biodiversity monitoring is that it would encourage farming practices that are environmentally sustainable. If farmers market the fact that they grow their crops using such sustainable practices they would attract higher prices especially among urban consumers.

To achieve these benefits, significant research work still remains to be done. We must collect sufficient training data of acoustic recordings of Kenyan bird species and develop reliable methods of bird species recognition. However, the preliminary results presented in this paper are encouraging.

7. Conclusions

In this work we have proposed an acoustic biodiversity monitoring tool that combines traditional acoustic indices of biodiversity, bird species recognition and audio diarization. We focus on bird species because they serve as indicator taxa and are present in most habitats. Furthermore even when birds cannot be seen they produce vocalizations which indicate their presence.

We demonstrate the performance of acoustic biodiversity indices and bird species recognition algorithms using a freely available data set recorded in the H. J. Andrews Long-Term Experimental Research Forest in Oregon which contains recordings of 19 bird species. The bird species recognition models in this work are based on the UBM-GMM framework and are built using open source tools.

Future work will involve collecting sufficient data from appropriate areas in Kenya and using these data to train models for local bird species. We will also adapt the system so that it can run on a smart phone in reasonable time.

References

[1] Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403: 853–858.

[2] Lawton JH, Bignell D, Bolton B, Bloemers G, Eggleton P, et al. (1998) Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. Nature 391: 72–76.

[3] Kerr JT, Sugar A, Packer L (2000) Indicator taxa, rapid biodiversity assessment, and nestedness in an endangered ecosystem. Conservation Biology 14: 1726–1734.

[4] Brooks T, Balmford A, Burgess N, Hansen LA, Moore J, et al. (2001) Conservation priorities for birds and biodiversity: do East African important bird areas represent species diversity in other terrestrial vertebrate groups. Ostrich 15: 3–12.

[5] Sueur J, Pavoine S, Hamerlynck O, Duvail S (2008) Rapid acoustic survey for biodiversity appraisal. PLoS One 3: e4065.

[6] Depraetere M, Pavoine S, Jiguet F, Gasc A, Duvail S, et al. (2012) Monitoring animal diversity using acoustic indices: implementation in a temperate woodland. Ecological Indicators 13: 46–54.

[7] Briggs F, Lakshminarayanan B, Neal L, Fern XZ, Raich R, et al. (2012) Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. The Journal of the Acoustical Society of America 131: 4640–4650.

[8] Stowell D, Plumbley MD (2014) Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. PeerJ 2: e488.

[9] Stowell D, Plumbley MD (2011) Birdsong and C4DM: A survey of UK birdsong and machine recognition for music researchers. Centre for Digital Music, Queen Mary, University of London, London, UK, Tech Rep C4DM-TR-09-12 181.

[10] Briggs F, Huang Y, Raich R, Eftaxias K, Lei Z, et al. (2013) The 9th annual MLSP competition: New methods for acoustic classification of multiple simultaneous bird species in a noisy environment. In: IEEE International Workshop on Machine Learning for Signal Processing (MLSP). pp. 1–8.

[11] Lee H, Pham P, Largman Y, Ng AY (2009) Unsupervised feature learning for audio classification using convolutional deep belief networks. In: Advances in neural information processing systems. pp. 1096–1104.

[12] Shum SH, Dehak N, Dehak R, Glass JR (2013) Unsupervised methods for speaker diarization: An integrated and iterative approach. IEEE Transactions on Audio, Speech, and Language Processing 21: 2015–2028.

[13] Vijayasenan D, Valente F, Bourlard H (2009) An information theoretic approach to speaker diarization of meeting data. IEEE Transactions on Audio, Speech, and Language Processing 17: 1382–1393.

[14] Dehak N, Kenny P, Dehak R, Dumouchel P, Ouellet P (2011) Front-end factor analysis for speaker verification. IEEE Transactions on Audio, Speech, and Language Processing 19: 788–798.

[15] Adi K, Johnson MT, Osiejuk TS (2010) Acoustic censusing using automatic vocalization classification and identity recognition. The Journal of the Acoustical Society of America 127: 874–883.

[16] Cover TM, Thomas JA (2012) Elements of information theory. John Wiley & Sons.

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

[18] Anjos A, Shafey LE, Wallace R, Günther M, McCool C, et al. (2012) Bob: a free signal processing and machine learning toolbox for researchers. In: 20th ACM Conference on Multimedia Systems (ACMMM), Nara, Japan. ACM Press.

[19] Njoroge P, Bennun LA (2000) Status and conservation of Hinde's Babbler turdoides hindei, a threatened species in an agricultural landscape. Ostrich 71: 69–72.

[20] Biggs D, Turpie J, Fabricius C, Spenceley A, et al. (2011) The value of avitourism for conservation and job creation-An analysis from South Africa. Conservation and Society 9: 80.