IMPERIAL COLLEGE LONDON

THESIS

Using acoustic monitoring and machine learning to automate the detection of illegal activity in Costa Rica

---

Author:
Jacob GRIFFITHS

Supervisors:
Dr. Cristina BANKS-LEITE
Dr. James ROSINDELL

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science at Imperial College London

Formatted in the journal style of

Submitted for the MSc in Computational Methods in Ecology and Evolution

Department of Life Sciences

August 23, 2019

Declaration of Authorship

I, Jacob GRIFFITHS, declare that this thesis titled, "Using acoustic monitoring and machine learning to automate the detection of illegal activity in Costa Rica" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

IMPERIAL COLLEGE LONDON

Abstract

Faculty of Natural Sciences
Department of Life Sciences

Master of Science

Using acoustic monitoring and machine learning to automate the detection of illegal activity in Costa Rica

by Jacob GRIFFITHS

Abstract written here

Acknowledgements

Write acknowledgements here

Contents

List of Figures

List of Tables

List of Abbreviations

PAM	**P**assive **A**udio **M**onitoring
MFC	**M**el **F**requency **C**epstrum
MFCCs	**M**el **F**requency **C**epstrum **C**oefficients
CNN	**C**onvolutional **N**eural **N**etwork

Introduction

1.1 Biodiversity loss

It is well documented that global biodiversity loss is accelerating, with anthropogenic factors often touted as the leading cause, either directly through deforestation and hunting, or indirectly through climate change and the introduction of invasive species, often rodents (Chiarucci, Bacaro, and Scheiner, 2011; Doherty et al., 2016; Newbold et al., 2015). A meta-analysis by De Vos et al., 2015 gave a conservative estimate of current biodiversity loss as 1,000 times greater than the background 'natural' rate, with a magnitude of 10,000 times greater also plausible (Ceballos et al., 2015). This alarming rate of loss does not only affect the lost taxa themselves, it can also lead to extensive reduction in ecosystem multifunctionality, typically affecting poorer human communities (Chiarucci, Bacaro, and Scheiner, 2011; Allan et al., 2015; Fanin et al., 2018).

Rainforests are no exception to this trend and in addition to the adverse effects of deforestation on biodiversity, its though that other forms of anthropogenic disturbance, such as vehicles, hunting, and light pollution, can double the biodiversity loss caused by deforestation alone (Barlow et al., 2016).

1.2 Difficulties of measuring biodiversity and its loss

Whilst it is universally agreed that biodiversity is being lost and almost universally agreed that anthropogenic factors are accelerating this process, measuring biodiversity and its loss can be inconsistent and even erroneous. Firstly, only about 15% of all species have been described so we have no data concerning biodiversity loss for the vast majority of species on Earth. Secondly, of those that have been described, very little is known of their distribution, population size, ecology and life histories, with many species only known due to a single observation. Finally, of those species that have been described and documented to a greater extent, there are inconsistencies in survey methods and often a lack of baseline measures to compare to (Society, 2003). The 'best' survey method is often specific to a certain level of organisation and spatial scale of interest (e.g. satellite imagery and ground surveys for rainforest plant surveys). The rapid acceleration of biodiversity loss makes urgent the development

of programmes to assess and monitor biodiversity suitable for large-scale ecological questions (Chiarucci, Bacaro, and Scheiner, 2011).

1.3 Passive audio monitoring and its advantages

Passive audio monitoring (PAM) is becoming an increasingly popular method for large-scale biodiversity monitoring, primarily due to its relatively low cost (Browning et al., 2017). This involves deploying sound recorders in an environment and having them record for days or weeks at a time to either track a vocal species directly or to use a vocal species as a proxy for another species or the ecosystem as a whole. Previously, methods such as PAM have been greatly limited by lack of digitisation, high implementation costs and small data storage capacity (Merchant et al., 2015). However, advances over the last 10-15 years have improved these constraints dramatically and one audio sensor in particular that has been developed recently by a collaborative project between the University of Oxford and University of Southampton, AudioMoth, is making PAM not only viable option for monitoring biodiversity loss, but one of the best methods (Hill et al., 2018). More recently, multiple-microphone arrays are being used to spatially locate vocal species, improving population censoring (Blumstein et al., 2011; Stevenson et al., 2015).

In addition to the direct monitoring of biodiversity with PAM, PAM can also be used to track other acoustics which may be relevant to conservationists, such as gunshots which are generally associated with illegal hunting, particularly in protected areas. Astaras et al., 2017 used PAM in a national park in Cameroon to successfully monitor the rates of hunting in the area. They found most (68.6%) hunting occurred at night when ranger patrols were minimal and that there was more illegal activity during the week, probably due to the typical Saturday market days, implying this hunting was for the illegal mate trade rather than for sustinance or sport. However, Astaras et al., 2017 noted that the cost of equipment for PAM was quite high, and that this may limit its implementation in other national parks. The recent development of much cheaper audio sensors by Hill et al., 2018 may aid the spread of these techniques in conservation.

1.4 Analysis of audio data

Sound is the propagation of waves of pressure through a medium. When a gun is fired, the vibrations produced alternately compress and rarefy the medium, leading to waves of high and low pressure that propagate in all directions (Bradbury and Vehrencamp, 2011). Over time and distance, these waves attenuate, that is their amplitude reduces as energy dissipates into the environment (Russ, 2013).

To capture this data electronically, sound waves are transduced into an electrical signal with amplitude proportional to the amplitude of the sound wave through the vibration of the diaphragm of a microphone. Previously, the transduced electrical signal was then recorded directly onto an analogue tape but digitisation has become far more frequently used due to it allowing much longer recording times, with the data then being in a more appropriate format for computer analysis. To record digitally, the analog signal is sampled at a certain rate (typically measured in thousands of samples per second, kHz) and bit-depth (number of possible amplitude levels, typically 16-bit), with both parameters being important for determining frequency and amplitude resolution respectively. The signal information is then electronically recorded in the time-amplitude domain and can be processed mathematically using a fast Fourier transform (FFT) to convert the amplitude data into frequency data. For a given time window in the recording, the FFT calculates the frequency components of the signal and their relative amplitudes, producing a frequency spectrum. For a visual representation of the whole recording, an FFT is calculated with an overlapping short sliding window across the length of the recording, producing a spectrogram (Society, 2003) This entire process is outlined in Figure 1.1.

It is common in audio classification to plot a variant of the traditional spectrogram, the mel-frequency spectrogram. This involves calculating the mel-frequency cepstrum (MFC) which is a representation of the sound's power spectrum after the frequency has passed through a mathematical function. Mel-frequency cepstral coefficients (MFCCs) are the constiuent coefficients of an MFC (Xu et al., 2004). They are calculated from a cepstral representation of the sound, with frequency bands equally spaced on the mel scale (Stevens, Volkmann, and Newman, 1937).

Spectrograms are fundamental to the analysis of acoustic data as they allow very specific sounds (e.g. spider monkey call, gunshot) to be visually identified and labelled, either manually or automatically.

1.5 Machine learning

Once audio data are collected, ecological information can be extracted manually or automatically. Manual extraction involves an expert either auditorily or visually inspecting the data and classifying the desired sounds, naturally incurring some bias based on their skill (Heinicke et al., 2015). This may be a viable option with a skilled ecologist and a small dataset but the latter is becoming increasingly rare with technological advances, therefore the need for automated techniques is growing. Fortunately, automated techniques are improving rapidly in accuracy and efficiency, largely due to the use of machine learning (Digby et al., 2013). Most automated tools utilise supervised machine learning and related methods, including artificial

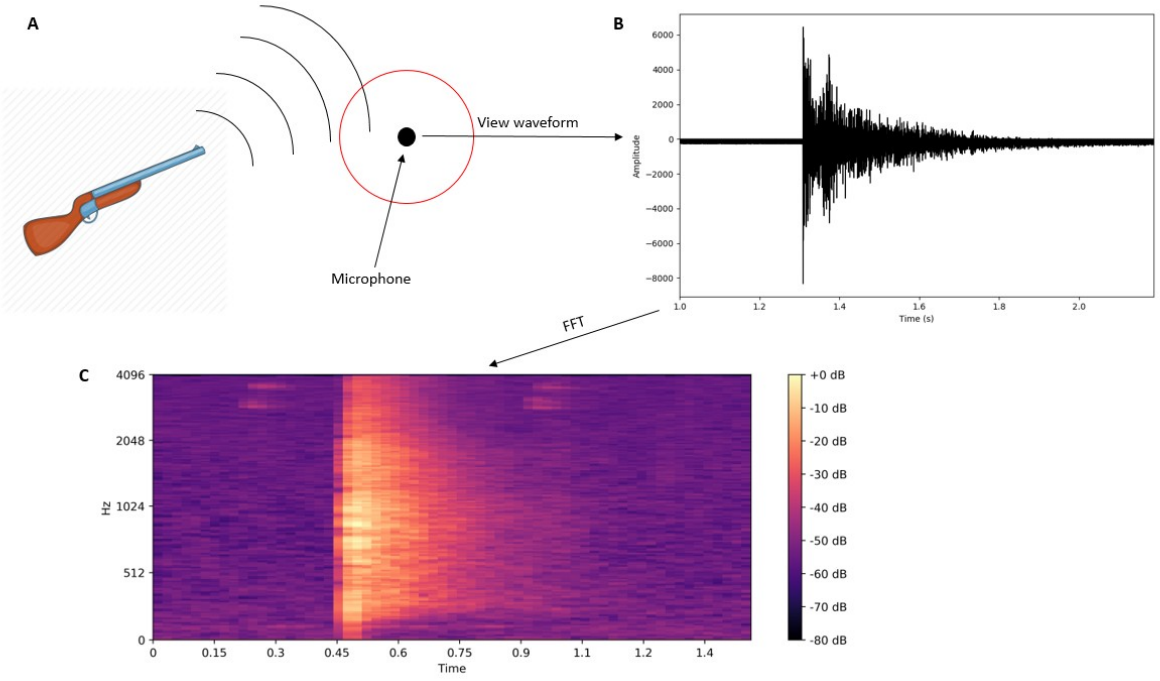

FIGURE 1.1: Recording and analysis of an acoustic signal. The emitted sound is transduced into an electrical signal by a microphone (**A**). In digital recording, the sound can be reconstructed in the time-amplitude domain (**B**) using the specified sampling rate (kHz). A frequency spectrum can then be produced using a fast Fourier transform (FFT), which calculates the signal's frequency components and their relative amplitudes. Calculating FFT within a sliding window across the recording produces a spectrogram, with time on the x-axis, frequency on the y-axis, and with amplitude (energy) shown as colour intensity (**C**)

neural networks (Walters et al., 2012), random forest (Zamora-Gutierrez et al., 2016), Hidden Markov Models (Zilli et al., 2014), and support vector machines (Heinicke et al., 2015). These methods commonly use libraries of species calls or other sounds to facilitate detection when presented with new recordings. Currently, the accuracy of these systems is rarely good enough to enable full automation and manual validation is often required (Kalan et al., 2016). However, new methods such as unsupervised feature extraction (Stowell and Plumbley, 2014) and deep convolutional neural networks (Goëau et al., 2016) can learn to classify directly from spectrogram data, often making them more robust and resistant to noise. At present, the main limitation for deep convolutional neural networks is large, clean datasets to train on.

1.6 Costa Rica and conservation

Costa Rica is no exception to the aforementioned trend of biodiversity loss, with the Osa Peninsula being an area of focus for recent studies as it contains a mixture of protected, partially-protected and unprotected land (Lawson, 2019). Whilst protected areas have benefited some species, others such as the Geoffroy's spider monkey, *Ateles Geoffroyi*, are struggling due to their diet and need for mature trees which is increasingly limiting their range and may be isolating populations, limiting their survival and genetic variability (Chapman, Chapman, and McLaughlin, 1989). *A. geoffroyi* is classified as endangered by the IUCN due to a 50% reduction in numbers of the last 45 years (Cuarón et al., 2008). This reduction may be having a negative impact on other species as *A. geoffroyi* is known to disperse the seeds of up to 150 different tree species (Roosmalen, 1985). As well as this habitat fragmentation, *A. geoffroyi* is being subjected to hunting in both protected and non-protected areas. Aquino et al., 2013 found hunted populations of *A. geoffroyi* in Peru were 70-80% less dense than non-hunted populations. However, the monitoring and prevention of hunting in protected areas is often difficult in large reserves with limited resources available to rangers and conservationists, but a recent study by Hill et al., 2018 demonstrated gunshots can be detected with acoustic sensors up to 1 km from the source, opening a potentially cheaper and more effective mitigation strategy.

1.7 Aims

1. To use data provided by Hill et al., 2018 to train a deep convolutional neural network that can detect gunshots in acoustic data
2. To investigate the effectiveness of using machine learning in cases such as this
3. To identify the presence of any spatio-temporal patterns of hunting in the Osa Peninsula

Methods

2.1 Data collection and preprocessing

2.1.1 Study site

The audio data used in this study was collected primarily by Jenna Lawson at a study site in the Osa Peninsula, Costa Rica (Figure 2.1). This 2,500 km^2 site sits in a particularly diverse region, containing approximately 2.5% of the world's species on less than 0.001% of its land area. Three national parks, Corcovado, Piedras Blancas and the Terreba-Sierpe wetlands, and one forest reserve, are represented within this study site. However, protection is not continuous throughout this area as the expanding human population has led to the inevitable landscape alteration brought about by agriculture and urbanisation.

2.1.2 Sampling

The Costa Rican government have produced a grid system to aid scientific research and conservation efforts in this area, consisting of 240 4x4 km^2 zones. This system was extended by Jenna to include the wetland region and 52 zones were removed as we are only interested in the zones that fall within or around the national parks and forest reserve. From the remaining 188 zones, 45 were randomly selected in a stratified manner to ensure good coverage of the region and appropriate representation of the national parks, forest reserve and unprotected land. In all 45 locations, 10 audio recording devices were installed at a minimum distance of 500m from each other to ensure independence of recordings. These 10 locations were also selected using random stratified sampling, ensuring fair representation from each habitat type. Where possible, trails and other areas of high human density were avoided. Each location was assigned to one of ten overall regions within the study site.

2.1.3 Recording

Audio recordings were gathered using AudioMoth devices and were set to record for six consecutive days. Each device was set to record for three periods a day, 0500-0930, 1400-1830 and 2100-0300, chosen to coincide with the peak activity levels of *A. geoffroyi* and their associated poaching. Sound was recorded continuously during

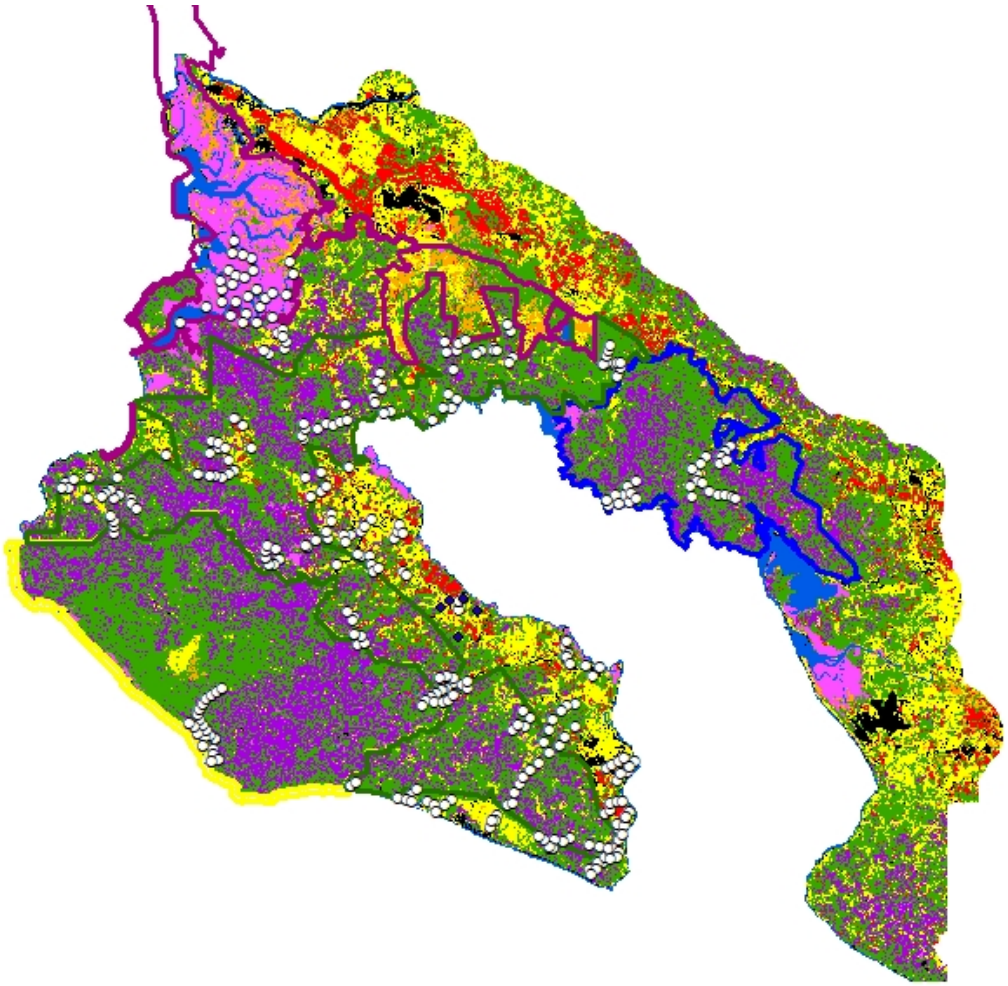

FIGURE 2.1: Map of the study site in the Osa Peninsula Lawson, 2019
(NEED KEY/AREAS MARKED)

these periods at a sample rate of 48 kHz and devices were contained in waterproof casing to avoid damage. Every minute of audio data was saved in a separate file with the filename being a Unix hexadecimal timestamp.

2.1.4 Subsetting

Of the ten regions captured in this dataset, six were selected for this analysis to ensure adequate processing time was available. The selected regions are known to be particular hotspots for illegal activity. Similarly, to minimise computational and manual expense, two sensors were selected at random from these six regions to be used for this analysis. This subset still includes 960 hours of audio data and represents 60% of the available regions and 40% of the available audio sensors within those regions.

2.1.5 Training data

Hill et al., 2018 provided the labelled audio data used in their study which was used in this study to train the CNN. The data was collected using 36 AudioMoth devices in the tropical rainforests in Pook's Hill Reserve, Belize. The devices covered 13 sites and were placed 200 m apart from each other. Controlled gunshots were set off by the research team and were labelled as positives in the dataset through visual inspection of spectograms. The detection algorithms used were able to detect 66% of gunshots up to 500 m away and 50% up to 1 km away. Devices facing towards the gunshot were 80% more likely to detect it than those facing away.

2.2 Machine learning

All coding was carried out using either Python (3.6.8) or bash (4.4.20), on an Ubuntu (18.04.3) operating system. For the training data, each audio file was split into four second clips using the pydub (0.23.1) Python module. Each clip was then plotted as a 60x60 mel-spectrogram using the librosa (0.7.0) Python module.

The CNN was then trained using the Keras (2.2.4) Python module. Mel-spectrogram images were imported and the pixel values were normalised to take values between 0 and 1 (rather than 0 and 255) as this has been shown to improve convergence and stability (Liao and Carneiro, 2016). The data was split randomly into training and validation data in the ratio 3:1 as this has been shown to be appropriate for a dataset of this size (Guyon, 1997). The seed for random number generation was set to ensure repeatability.

2.2.1 Convolutional Neural Network

The CNN was constructed using Chollet, 2016's tutorial as a template. The first layer was a convolutional 2D kernel which is convolved with the input layer to produce a tensor of outputs, the input layer being n spectrograms of 64x64x3 size. It contained 32 nodes which determines the dimensionality of the output space and a 3x3 convolutional window. A 'relu' activation function was then applied as this was shown by Glorot, Bordes, and Bengio, 2011 to enable better training of deep neural networks, compared to other common activation functions such as the logistic sigmoid. The identified features are then passed through a max pooling layer which determines the most activated presence of each feature within a 2x2 cluster of features. Put simply, this filters out the less important features and keeps the most important ones to be passed to the next layer. In this model, these three layers were repeated in the same order, two further times, with the model then totalling nine layers. The features were then passed to a 'flattening' layer. This is a layer that takes a two-dimensional matrix of features and transforms them into a vector of features than can fed to a neural network classifier, in this case a 'dense' layer composed of 64 fully-connected neurons. These neurons linearly take all the inputs from the previous layer, apply a weight to them and output to the next layer which in this CNN was another 'relu' activation layer. The activation layer output is then passed through a 'dropout' layer, in this case that involved randomly discarding 50% of nodes in an effort to minimise overfitting. The remaining nodes are put through another 'dense' layer, this time with 2 nodes as this CNN was only being trained in a binary 'gunshot' or 'no gunshot' manner. The output of this 'dense' layer was passed through a final 'softmax' activation layer which is similar to logistic regression but usually used in multi-classification problems. However, it has been shown to be more effective than logistic regression even in binomial classification (FIGURE NEEDED).

2.2.2 Training the model

Initially, the model was trained on the data provided by Hill et al., 2018 by randomly splitting the data into training and test data in a ratio of 70:30. The training and test data was comprised of equal numbers of positive (gunshots) and negative (no gunshot) spectra as imbalanced ratios have been previously shown to be ineffective (Kim and Kim, 2018) and preliminary experimentation on this dataset confirmed this. This model was then fed the subset of data from the Osa Peninsula and the returned 'gunshots' were manually checked for authenticity. The validated gunshots were used to retrain the model in the same manner, before being fed data from the Osa Peninsula that had not already been used to train the model. The model was retrained a further two times, once with a combined training dataset of both Hill

et al., 2018 and Osa Peninsula, and another with the Osa Peninsula data, this time the negatives used were the false positives identified originally.

Results

3.1 Model comparison

The number of 'gunshots' found by each variant of the model is highlighted in Figure 3.1. As the only way to validate the authenticity of these 'gunshots' is through manual checking, there was only time to validate the model that was trained on Hill et al., 2018's data from Belize. After validation, 252 'gunshtos' were confirmed to be authentic, meaning 6,912 'gunshots' were deemed to be false positives and giving a model accuracy of 3.65%. Although the other models haven't been manually validated, it seems a logical assumption, particularly for the Osa Peninsula and Combined models, that the vast majority of 'gunshots' are false positives. For example, if the 'gunshots' returned by the Combined model were all authentic, this would imply that a gunshot was being fired in that area approximately every 16 seconds (24% of clips) which seems highly unlikely, especially as 1.53% of audio clips have been manually checked and of these, only 0.08% contain a confirmed gunshot. Therefore, working on the assumption that the majority of 'gunshots' are false positives, The False Positives and Belize models are the best as they returned the lowest number of 'gunshots'.

3.2 Spatio-temporal analysis

Taking the authenticated gunshots from the Belize model, further analysis was undertaken to explore the temporal patterns of hunting (Figure 3.2). A chi-square test of goodness-of-fit was performed to determine whether gun frequency was independent of the day of the week. Gunshot frequency was equally distributed across the weekdays, X^2 (25, N = 252) = 42, p = 0.227.

Further analysis was carried out on the time of day gunshots occured in (Figure 3.3). Gunshots were assigned to either 'morning', 'afternoon' or 'night', mirroring the three periods of the day the audio sensors were set to record. A chi-square test of goodness-of-fit was performed to determine whether gun frequency was independent of time of day. Gunshot frequency was equally distributed across the three periods of the day, X^2 (4, N = 252) = 6, p = 0.199.

Area	Belize	Osa Peninsula	Combined	False Positives
SQ258	1179	2063	18080	1118
SQ283	1594	1343	12188	1576
SQ282	936	2404	22975	863
La_Balsa	758	5230	11108	729
Rancho_bajo	1514	3471	18003	1486
Indigenous_reserve	931	22565	27924	887

FIGURE 3.1: Number of returned 'gunshots' from each version of the convoluted neural network (CNN). The 'Belize' model was trained only on data provided by Hill et al., 2018, 'Osa Peninsula' was trained on data collected by Jenna Lawson (Lawson, 2019), 'Combined' was trained on a combination of the previous two datasets, and 'False Positives' was trained on Jenna's data but all negatives provided were previously identified false positives.

Gunshot frequency between the six study areas was also compared (Figure 3.4). A chi-square test of goodness-of-fit was performed to determine whether gun frequency was independent of Area. Gunshot frequency was equally distributed across the six area, X^2 (25, N = 252) = 30, p = 0.224.

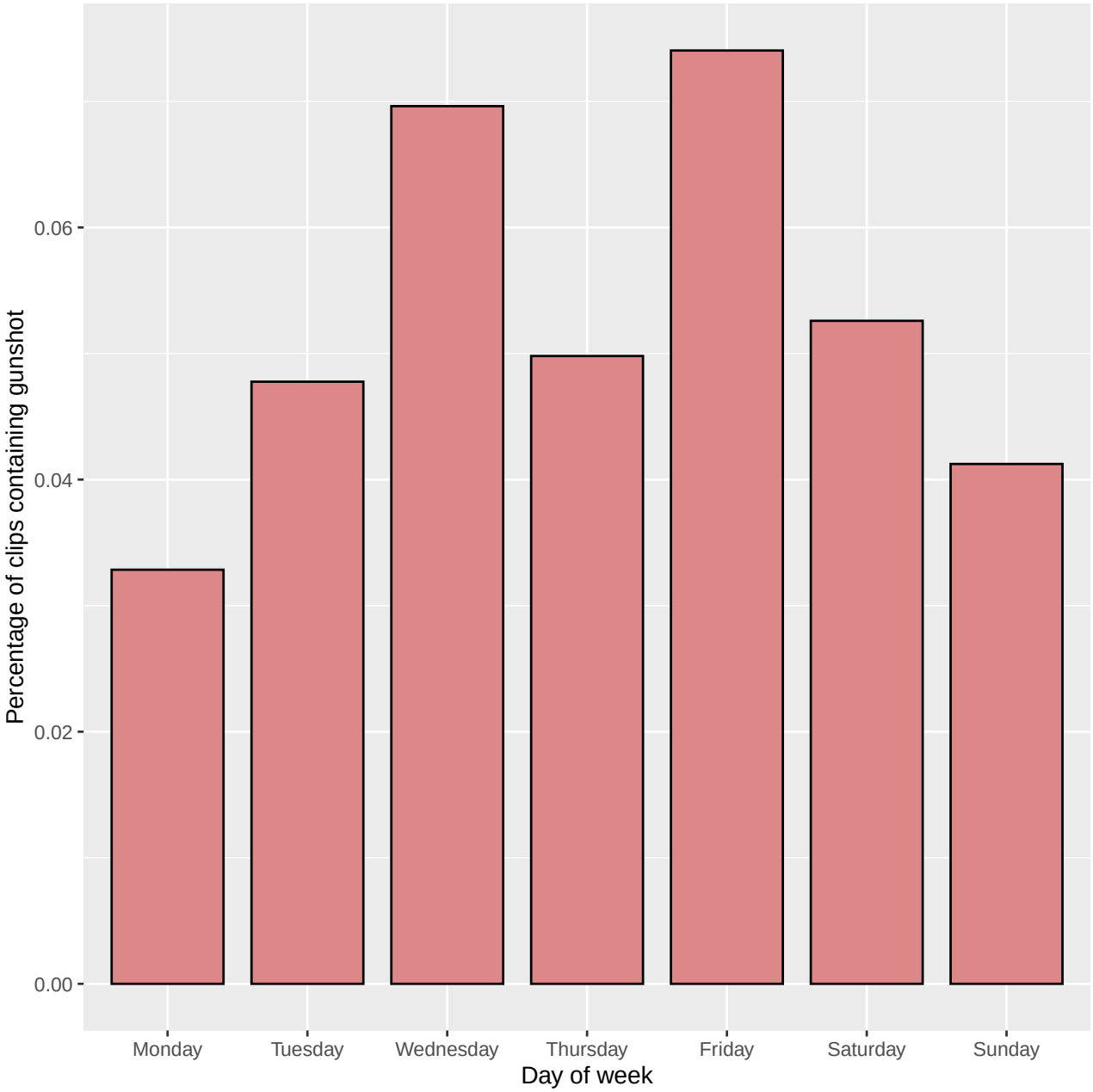

FIGURE 3.2: Percentage of clips containing a gunshot on each day of the week.

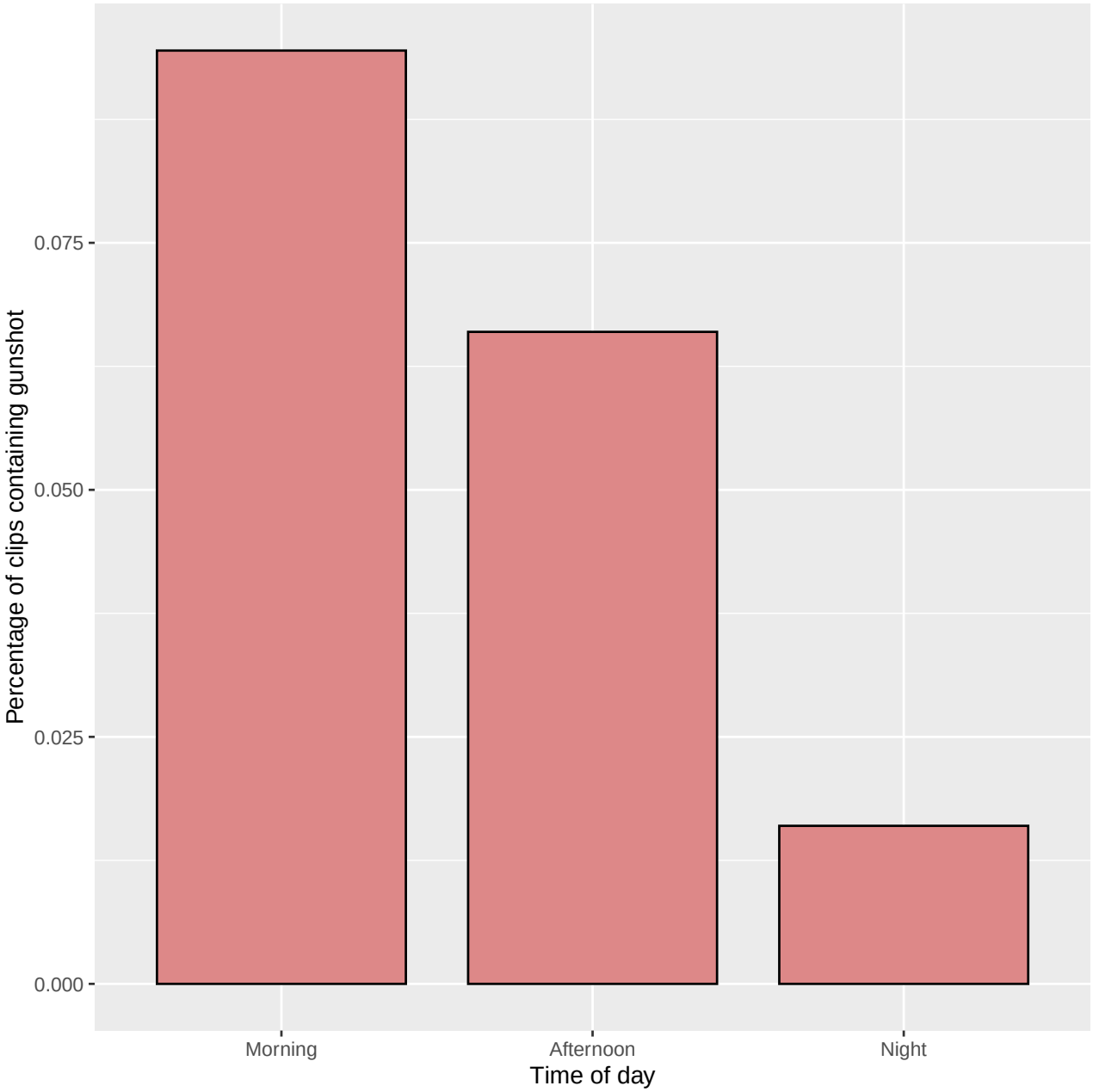

FIGURE 3.3: Percentage of clips containing a gunshot in each period of the day. 'Morning' was 0500-0930, 'Afternoon' was 1400-1830, and 'Night' was 2100-0300.

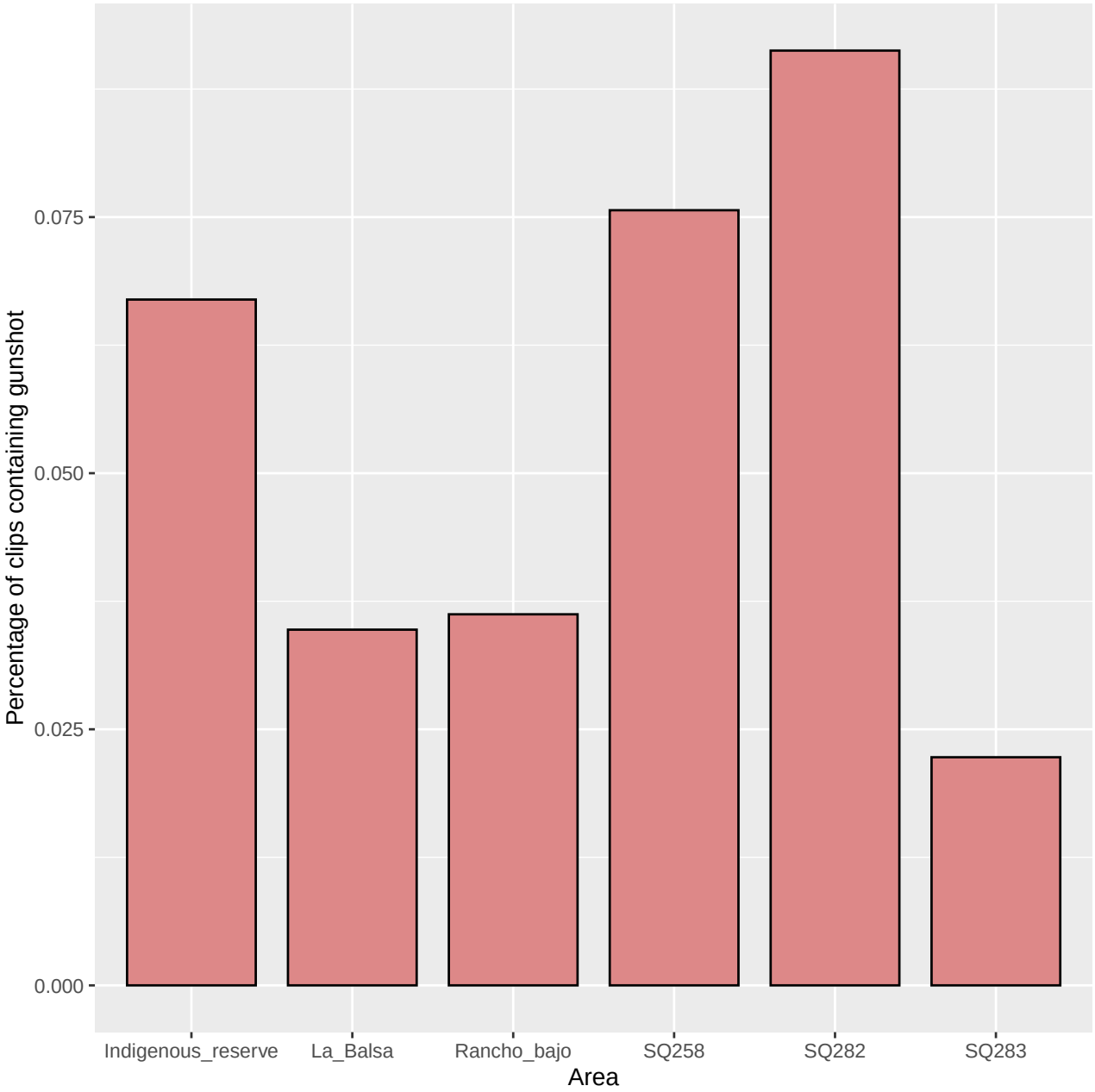

FIGURE 3.4: Percentage of clips containing a gunshot within each of the six regions within the study site.

Discussion

4.0.1 Machine learning and its effectiveness

Perhaps surprisingly, training the CNN on the data provided by Hill et al., 2018 was quite effective despite this data coming from a different country at a different time. This success was marred by low precision despite the potentially high recall. However it cannot be said for certain if the recall was high as that would require manual validation of the whole dataset to indentify all the gunshots and thus whether they were successfully identified by the CNN or not. That being said, due to the surprisingly high volume of gunshots identified, it seems logical to assume that the recall was high. Re-training the CNN on both the output of the Belize model and a combination of the Belize and Osa Peninsula data proved ineffective in reducing the volume of false positives returned. However, re-training the CNN on the Osa Peninsual data with the negatives provided being previously identified false positives, seemed to bring the precision in line with, or slightly better than, the Belize model. This strongly suggests that the primary factor in determining the effectiveness of a CNN model in studies such as this is the quality of the dataset used to train the model on. If time had permitted, the logical next step of this study would have been to increase the size of the dataset and continue to retrain the model on an ever larger, more accurate dataset. This may be the only way to improve the precision going forward.

4.0.2 Spatio-temporal hunting patterns on the Osa Peninsula

Correlation between gunshot frequency and the day of the week proved to be insignificant but visual inspection of the data does seem to suggest a non-unfiform distribution across the days and a relatively low p-value suggests this may prove significant if more data was gathered. Similarly, there was no significant correlation between time of day and gunshot frequency but again, chi-squared test results returned a low p-value so the size of the dataset is probably playing a part again. The gunshot frequency was around four times lower at night which seems to suggest non-uniformity and contrasts completely with the findings of Astaras et al., 2017. This may be due to infrequent patrolling in these areas. Jenna Lawson says that patrols in the national parks are at best once per day in the mornings and usually

only cover a small percentage of the total park area so perhaps this is a negligable hinderance to illegal hunters.

4.0.3 Conclusion

Bibliography

Allan, Eric et al. (2015). "Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition". In: *Ecology Letters* 18.8, pp. 834–843. ISSN: 14610248. DOI: 10.1111/ele.12469.

Aquino, Rolando et al. (2013). "Distribution and abundance of white-fronted spider monkeys, Ateles belzebuth (atelidae), and threats to their survival in Peruvian Amazonia". In: *Folia Primatologica* 84.1, pp. 1–10. ISSN: 00155713. DOI: 10.1159/000345549.

Astaras, Christos et al. (2017). "Passive acoustic monitoring as a law enforcement tool for Afrotropical rainforests". In: *Frontiers in Ecology and the Environment* 15.5, pp. 233–234. ISSN: 15409309. DOI: 10.1002/fee.1495.

Barlow, Jos et al. (2016). "Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation". In: *Nature* 535.7610, pp. 144–147. ISSN: 14764687. DOI: 10.1038/nature18326. URL: http://dx.doi.org/10.1038/nature18326.

Blumstein, Daniel T et al. (2011). "Acoustic monitoring in terrestrial environments using microphone arrays : applications , technological considerations and prospectus". In: *Journal of Applied Ecology* 48, pp. 758–767. DOI: 10.1111/j.1365-2664.2011.01993.x.

Bradbury, Jack and Sandra Vehrencamp (2011). *Principles of Animal Communication*. 2nd ed. OUP, USA.

Browning, Ella et al. (2017). *Passive acoustic monitoring in ecology and conservation*. Tech. rep., pp. 1–75.

Ceballos, Gerardo et al. (2015). "Accelerated modern human-induced species losses: Entering the sixth mass extinction". In: *Science Advances* 1.5, pp. 1–5. ISSN: <null>.

Chapman, C. A., L. J. Chapman, and R. L. McLaughlin (1989). "Multiple central place foraging by spider monkeys: travel consequences of using many sleeping sites". In: *Oecologia* 79.4, pp. 506–511. ISSN: 1432-1939. DOI: 10.1007/BF00378668.

Chiarucci, Alessandro, Giovanni Bacaro, and Samuel M. Scheiner (2011). "Old and new challenges in using species diversity for assessing biodiversity". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 366.1576, pp. 2426–2437. ISSN: 09628436. DOI: 10.1098/rstb.2011.0065.

Chollet, Francois (2016). *Building powerful image classification models using very little data*. URL: https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html.

Cuarón, A. D. et al. (2008). *Ateles geoffroyi , Geoffroy's Spider Monkey*. URL: http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T2279A9387270.en{\%}0ACopyright: (visited on 07/22/2019).

De Vos, Jurriaan M. et al. (2015). "Estimating the normal background rate of species extinction". In: *Conservation Biology* 29.2, pp. 452–462. ISSN: 15231739. DOI: 10.1111/cobi.12380.

Digby, Andrew et al. (2013). "A practical comparison of manual and autonomous methods for acoustic monitoring". In: *Methods* 4, pp. 675–683. DOI: 10.1111/2041-210X.12060.

Doherty, Tim S. et al. (2016). "Invasive predators and global biodiversity loss". In: *Proceedings of the National Academy of Sciences* 113.40, pp. 11261–11265. ISSN: 0027-8424. DOI: 10.1073/pnas.1602480113.

Fanin, Nicolas et al. (2018). "Consistent effects of biodiversity loss on multifunctionality across contrasting ecosystems". In: *Nature Ecology and Evolution* 2.2, pp. 269–278. ISSN: 2397334X. DOI: 10.1038/s41559-017-0415-0. URL: http://dx.doi.org/10.1038/s41559-017-0415-0.

Glorot, Xavier, Antoine Bordes, and Yoshua Bengio (2011). "Deep Sparse Rectifier Neural Networks Xavier". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Vol. 15, pp. 315–323. ISBN: 1532-4435. DOI: 10.1.1.208.6449. arXiv: 1502.03167. URL: http://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf.

Goëau, Hervé et al. (2016). "LifeCLEF Bird Identification Task 2016: The arrival of Deep learning." In: *Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum*, pp. 440–449. URL: https://hal.archives-ouvertes.fr/hal-01373779/document.

Guyon, Isabelle (1997). *A scaling law for the validation-set training-set size ratio*. Tech. rep., pp. 1–11. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.33.1337{\&}rep=rep1{\&}type=pdf.

Heinicke, Stefanie et al. (2015). "Assessing the performance of a semi-automated acoustic monitoring system for primates". In: *Methods in Ecology and Evolution* 6, pp. 753–763. DOI: 10.1111/2041-210X.12384.

Hill, Andrew P. et al. (2018). "AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment". In: *Methods in Ecology and Evolution* 9.5, pp. 1–13. ISSN: 2041210X. DOI: 10.1111/2041-210X.12955.

Kalan, Ammie K et al. (2016). "Passive acoustic monitoring reveals group ranging and territory use : a case study of wild chimpanzees (Pan troglodytes)". In: *Frontiers in Zoology* 13.34, pp. 1–11. ISSN: 1742-9994. DOI: 10.1186/s12983-016-0167-8. URL: http://dx.doi.org/10.1186/s12983-016-0167-8.

Kim, Jinseok and Jenna Kim (2018). "The impact of imbalanced training data on machine learning for author name disambiguation". In: *Scientometrics* 117.1, pp. 511–526. ISSN: 15882861. DOI: 10.1007/s11192-018-2865-9.

Lawson, Jenna (2019). "Sounds of the Spider Monkey : Using acoustics to investigate biodiversity, and land use and threats to the Geoffroy's spider monkey (Ateles geoffroyi) in Costa Rica".

Liao, Z. and G. Carneiro (2016). "On the Importance of Normalisation Layers in Deep Learning with Piecewise Linear Activation Units". In: *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–8.

Merchant, Nathan D. et al. (2015). "Measuring acoustic habitats". In: *Methods in Ecology and Evolution* 6.3, pp. 257–265. ISSN: 2041210X. DOI: 10.1111/2041-210X.12330.

Newbold, Tim et al. (2015). "Global effects of land use on local terrestrial biodiversity." In: *Nature* 520.7545, pp. 45–50. ISSN: 1476-4687. DOI: 10.1038/nature14324. URL: http://www.ncbi.nlm.nih.gov/pubmed/25832402.

Roosmalen, M. G. M. van (1985). "Habitat preferences, diet, feeding strategy and social organization of the black spider monkey (Ateles paniscus paniscus Linnaeus 1758) in Surinam". In: *Acta Amazonica* 15.

Russ, Jon (2013). *British Bat Calls: A Guide to Species Identification*. 1st ed. Pelagic Publishing.

Society, The Royal (2003). *Measuring biodiversity for conservation*. Tech. rep., pp. 1–65.

Stevens, S. S., J. Volkmann, and E. B. Newman (1937). "A Scale for the Measurement of the Psychological Magnitude Pitch". In: *Journal of the Acoustical Society of America* 8.3, pp. 185–190.

Stevenson, Ben C et al. (2015). "A general framework for animal density estimation from acoustic detections across a fixed microphone array". In: *Methods in Ecology and Evolution* 6, pp. 38–48. DOI: 10.1111/2041-210X.12291.

Stowell, Dan and Mark D. Plumbley (2014). "Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning". In: *PeerJ* 2, pp. 1–31. DOI: 10.7717/peerj.488.

Walters, Charlotte L. et al. (2012). "A continental-scale tool for acoustic identification of European bats". In: *Journal of Applied Ecology* 49.5, pp. 1064–1074. ISSN: 00218901. DOI: 10.1111/j.1365-2664.2012.02182.x.

Xu, Min et al. (2004). "HMM-based audio keyword generation". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 3333, pp. 566–574. ISSN: 03029743.

Zamora-Gutierrez, Veronica et al. (2016). "Acoustic identification of Mexican bats based on taxonomic and ecological constraints on call design". In: *Methods in Ecology and Evolution* 7.9, pp. 1082–1091. ISSN: 2041210X. DOI: 10.1111/2041-210X.12556.

Zilli, Davide et al. (2014). "A hidden Markov model-based acoustic cicada detector for crowdsourced smartphone biodiversity monitoring". In: *Journal of Artificial Intelligence Research* 51, pp. 805–827. ISSN: 10769757.