Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

WAVELET-BASED BIRDSONG RECOGNITION FOR CONSERVATION

A thesis presented in partial fulfilment of the requirements for the degree of
Doctor of Philosophy
in
Computer Science
at Massey University, Palmerston North,
New Zealand.

Nirosha Priyadarshani

2017

To Sumudu, for everything, forever...

Contents

Abstract **xiii**

Acknowledgements **xv**

1 Introduction **1**
1.1 Motivation 1
1.2 The Problem 2
1.3 Aims and Objectives 2
1.4 Thesis Outline and Research Contributions 3

2 Automated Birdsong Recognition in Complex Acoustic Environments: A Review **7**
2.1 Introduction 8
2.2 Birdsong Processing and Recognition 11
2.2.1 Sound Recording and Storage 11
2.2.2 Visual Representation of Birdsong 13
2.2.3 Noise Reduction 14
2.2.4 Performance Measures 18
2.2.5 Call Detection and Segmentation 19
2.2.6 Feature Choice and Extraction 21
2.2.7 Recognition and Classification 25
2.3 Current Birdsong Recognition Related Software 37
2.4 Conclusion and Recommendations 44

3 The Impact of Environmental Factors in Birdsong Acquisition using Automated Recorders **49**
3.1 Introduction 50
3.2 Materials and Methods 51
3.2.1 Study Species 51
3.2.2 Study Sites 55

3.2.3 Experimental Setup . . . 57
3.2.4 Wind Direction . . . 59
3.2.5 Data Extraction from Continuous Recordings . . . 60
3.2.6 Dependent Variable and Covariates . . . 60
3.2.7 Statistical Method . . . 61
3.3 Results . . . 63
3.3.1 Analysis I . . . 63
3.3.2 Analysis II . . . 73
3.3.3 Analysis III . . . 76
3.4 Discussion . . . 79
3.4.1 Sound Attenuation, Frequency, and Habitat . . . 79
3.4.2 Sound Attenuation and the Transmission Height . . . 82
3.4.3 Sound Attenuation and the Directionality of the Bird and the Recorder . . . 84
3.5 Conclusions . . . 84
3.6 Supporting Information . . . 85
3.7 Acknowledgments . . . 85

4 Birdsong Denoising Using Wavelets **87**
4.1 Introduction . . . 87
4.2 Bird Vocalisation, Categorisation and Spectrogram Patterns . . . 88
4.2.1 Spectrogram Analysis . . . 89
4.3 Bird Recording and Noise . . . 91
4.3.1 Types of Noise . . . 91
4.3.2 Noise Filtering . . . 92
4.4 Wavelets . . . 94
4.4.1 Wavelet Packet Decomposition . . . 95
4.4.2 Previous Uses of Wavelets for Bioacoustic Denoising . . . 97
4.4.3 Our Algorithm . . . 98
4.5 Experimental Evaluation . . . 100
4.5.1 Datasets . . . 100
4.5.2 Evaluation Metrics . . . 102
4.6 Results . . . 104
4.6.1 Extensions . . . 109
4.7 Discussion . . . 109
4.8 Supporting Information . . . 118
4.9 Acknowledgments . . . 118

5 Birdsong Recognition in Continuous Field Recordings **121**
5.1 Introduction . . . 122
5.2 Materials and Methods . . . 124
5.2.1 Datasets . . . 124
5.2.2 Our Algorithm: Wavelet-Based Call Detection (Segmentation by Wavelet Filtering) . . . 132
5.2.3 Comparison Between Segmentation by Wavelet Filtering and Other Methods . . . 138
5.3 Results . . . 138
5.3.1 The Effect of Distance and Direction of Bird on Song Capture and Automatic Detection . . . 138
5.3.2 Results on Non-Experimental Continuous Field Recordings . . . 141
5.3.3 Comparison With Other Methods . . . 143
5.3.4 Results on the RMBL Robin Database . . . 145
5.4 Discussion . . . 146
5.5 Supporting Information . . . 149

6 Concluding Remarks **151**
6.1 Pre-processing of Field Recordings . . . 151
6.2 Segmentation and Recognition . . . 152
6.3 Protocols for Data Acquisition . . . 152
6.4 Making Birdsong Recognition Practical: Recommendations . . . 153
6.4.1 Shared Data . . . 153
6.4.2 Inventory Management of Surveys . . . 154
6.4.3 Community Participation in Conservation . . . 154

A Generalised Linear Models **157**
A.1 Analysis I - Model effects . . . 158
A.2 Analysis II . . . 173
A.3 Analysis III . . . 174

Bibliography **177**

List of Tables

1.1 Manuscripts published or in preparation for submission from material in this thesis. 5

2.1 Noise reduction methods in descending order of effectiveness in regards to automatic analysis of recordings. 17
2.2 Call detection and segmentation methods in descending order of their effectiveness to automatic detection of putative calls. 22
2.3 Feature extraction and recognition in the descending order of their effectiveness to automatic recognition of bird sounds. 28
2.4 A summary of currently available software. 40

3.1 The bird sounds used in the experiment. 53
3.2 Summary of the playbacks and re-captures one bird sound produced. . . 58
3.3 Summary of the trials carried out. 59
3.4 GLM model development – goodness of fit in each model was measured using the Deviance. 62
3.5 Analysis I: The main effects found in each model (for each bird sound) at α=0.01. 63

4.1 List of species, their call types and frequency range 102
4.2 List of species introduced to the secondary dataset and their song characteristics. 103
4.3 Experimental Results – primary dataset 108
4.4 Experimental Results for the species introduced to the secondary dataset. 110
4.5 Comparing the denoising results – series of calls against their segmented calls. 111

5.1 The three training examples used to train the bittern detector. 135
5.2 The effect of the distance and the direction of the bird to the recorder in automated call detection. 140

5.3 Detection results on the complete test dataset of the four species of birds used in this study. The recall depends on the quality of the bird vocalisations. 143
5.4 Detection results on the complete test dataset of four species of birds. . 144
5.5 Recall, precision, specificity, and accuracy from our algorithm, energy thresholding, and median clipping on the complete test dataset of four species of birds. 145

A.1 Model effects. Dependent variable is SnNR. 158
A.2 EMM – Open vs Forest – overall test results 163
A.3 EMM – Day vs Night – overall test results 164
A.4 EMM – Low vs High transmission height – overall test results 165
A.5 EMM – Distance (20m, 25m, 50m, 100m, 120m) – individual test results 166
A.6 EMM – Distance (20m, 25m, 50m, 100m, 120m) – overall test results . . 168
A.7 EMM – Open/Forest*Day/Night interaction – overall test results 169
A.8 EMM – Open/Forest*Low/High transmission interaction – overall test results 170
A.9 EMM – Day/Night*Low/High transmission interaction – overall test results 171
A.10 EMM – Open/Forest*Distance interaction – overall test results 172
A.11 EMM (Analysis II) – Recorder direction – overall test results 173
A.12 EMM (Analysis III) – Recorder direction – overall test results 174
A.13 EMM (Analysis III) – Wind level – individual test results 175
A.14 EMM (Analysis III) – Wind level – overall test results 176

List of Figures

2.1 Representation of the full work process required for the use of acoustic recorders in wildlife management, including the development of protocols for the deployment of recorders. 12
2.2 Spectrogram representations of various bird species showing some of the typical appearances of sounds 15

3.1 (a) Recorders and playback setup, (b) the Kestrel wind meter and a half line of recorders (5 recorders) in open habitat, (c) speaker setup in the forest habitat, and (d) recorders mounted at eye level. 52
3.2 The experimental sites: (a) the open site and (b) the forest site. 56
3.3 The actual wind direction was measured for each bird sound in each trial by post processing Kestral data. 59
3.4 Estimated marginal means of SnNR for day vs night. 65
3.5 Estimated marginal means of SnNR for two different sites. 66
3.6 Estimated marginal means of SnNR for interaction effect of site and time of the call 67
3.7 Estimated marginal means of SnNR for two transmission heights. 68
3.8 Estimated marginal means of SnNR and interaction effect of the transmission height and the habitat. 69
3.9 Re-captured sounds from morepork broadcasts of more-pork (mp) and trill (trilH). In all cases the speaker was facing the recorders and the wind was calm. 70
3.10 Estimated marginal means of SnNR and interaction effect of the transmission height and the time of the call. 71
3.11 Estimated marginal means of SnNR against distance. 72
3.12 Re-captured birdsong that were transmitted from 3m to the ground in the forest during the day. Speaker was facing the recorders (20m, 50m, and 120m) positioned in one direction. (a) North Island saddleback (sad1), (b) North Island saddleback (sad2), (c) North Island robin, (d) hihi, (e) tui, and (f) North Island kākā. 74

3.13 The change of the SnNR when the speaker was facing Dir1 (Analysis II). 75
3.14 Estimated marginal means of SnNR against the four recorder directions when the bird calls to all four directions equally (Analysis III). The dataset includes 'calm', 'moderate', and 'windy' data in the open site. . 76
3.15 Re-captured little spotted kiwi call (lskm1) transmitted close to the ground 50m from the recorder in the open site under different wind levels when the speaker was facing the wind direction and away from the wind direction. 77
3.16 Estimated marginal means of SnNR against the different wind levels (Analysis III). 78
3.17 Re-captured male kiwi call (bm2) that was transmitted close to the ground in the forest and the open field at night. 81

4.1 Non-stationarity 90
4.2 Examples of bird calls with various degrees of noise, the effect of band-pass filtering and power spectrum of white and pink noise 93
4.3 Wavelets and their relation to time-frequency resolution and wavelet packet decomposition 96
4.4 Different mother wavelets produce different results 99
4.5 An example of kākāpō *chinging* used in the experiment 101
4.6 Denoising different types of noise. (a) White noise, (b) pink noise, and (c) brown noise 105
4.7 Bird call examples of before, after filtering, after denoising using wavelets as described in the text, and after denoising and classical filtering 107
4.8 Box plot view of the results in (a) Table 4.3 and (b) Table 4.4 112
4.9 Denoising entire songs and long series of calls 113
4.10 Box plot view of Table 4.5: (a) call series and (b) segmented calls 114
4.11 Denoising overlapped songs 116
4.12 A deliberate denoising example 117

5.1 Close-range call excerpts from the unattended field recordings used to evaluate the proposed method. Spectrogram settings for Australasian bittern are given in (O'Donnell and Williams, 2015). Note that the scale of the vertical axis differs between the calls. 125
5.2 Re-recorded '*trill*' sound of morepork at different distances (columns) when the bird (speaker) was (a) facing the recorder, (b) at 90° to the recorder, and (c) facing away from the recorder. 127
5.3 An excerpt from the RMBL dataset with its annotation. All American robin instances are labelled. 130

5.4 Call excerpts from the brown kiwi test dataset illustrating the different levels of quality of male whistles in field recordings. Note that all these examples were successfully detected by our kiwi detector. 133
5.5 Optimised wavelet packet decomposition tree for bittern detection. The white nodes are the ones that are sensitive to bittern booms. The filtered nodes occupy 62–250 Hz. 135
5.6 Detection of bittern calls using a wavelet node from (a) the first 5 min of the test dataset (Table 5.4). (b) The energy curve (shown in yellow) generated over the wavelet coefficients (node 10) from (a) and (c) the binary output of call availability. 137
5.7 Receiver Operating Characteristic (ROC) curve of each species based on the non-experimental continuous field recordings. 142
5.8 Original and denoised bittern boom examples used in this study ranging from very close (first row) to extremely faded (last row). On the left, the spectrogram settings are as given in (O'Donnell and Williams, 2015). On the right, another window preset with window size (sharpness) 400 was defined in Raven software to visualise the denoised bittern booms focusing on the bird's frequency range. 148

Abstract

According to the International Union for the Conservation of Nature Red Data List nearly a quarter of the world's bird species are either threatened or at risk of extinction. To be able to protect endangered species, we need accurate survey methods that reliably estimate numbers and hence population trends. Acoustic monitoring is the most commonly-used method to survey birds, particularly cryptic and nocturnal species, not least because it is non-invasive, unbiased, and relatively time-effective. Unfortunately, the resulting data still have to be analysed manually. The current practice, manual spectrogram reading, is tedious, prone to bias due to observer variations, and not reproducible.

While there is a large literature on automatic recognition of targeted recordings of small numbers of species, automatic analysis of long field recordings has not been well studied to date. This thesis considers this problem in detail, presenting experiments demonstrating the true efficacy of recorders in natural environments under different conditions, and then working to reduce the noise present in the recording, as well as to segment and recognise a range of New Zealand native bird species.

The primary issues with field recordings are that the birds are at variable distances from the recorder, that the recordings are corrupted by many different forms of noise, that the environment affects the quality of the recorded sound, and that birdsong is often relatively rare within a recording. Thus, methods of dealing with faint calls, denoising, and effective segmentation are all needed before individual species can be recognised reliably. Experiments presented in this thesis demonstrate clearly the effects of distance and environment on recorded calls. Some of these results are unsurprising, for example an inverse square relationship with distance is largely true. Perhaps more surprising is that the height from which a call is transmitted has a significant effect on the recorded sound. Statistical analyses of the experiments, which demonstrate many significant environmental and sound factors, are presented.

Regardless of these factors, the recordings have noise present, and removing this noise is helpful for reliable recognition. A method for denoising based on the wavelet packet decomposition is presented and demonstrated to significantly improve the quality of recordings. Following this, wavelets were also used to implement a call detection

algorithm that identifies regions of the recording with calls from a target bird species. This algorithm is validated using four New Zealand native species namely Australasian bittern (*Botaurus poiciloptilus*), brown kiwi (*Apteryx mantelli*), morepork (*Ninox novaeseelandiae*), and kakapo (*Strigops habroptilus*), but could be used for any species. The results demonstrate high recall rates and tolerate false positives when compared to human experts.

Acknowledgements

Undertaking this PhD has been a life-changing experience for me, inspired by the support and encouragement from researchers, conservation groups, Department of Conservation (DOC), and interesting individuals, combined with the best supervisory panel one could expect to have. Thinking of all the help I received reminds me of how fortunate I am, it would not have been possible to do this project without the support and guidance I received from those people.

I would like to express my sincere gratitude to Prof. Stephen Marsland and A/Prof. Isabel Castro for your effective supervision, understanding, patience, encouragement, criticisms, and for always being there to offer advice. Working with you was enjoyable as you lived in the project and enjoyed it. Thank you very much for promptly reading the manuscripts and providing valuable comments to improve my writing and my work in general. Stephen, thank you for your eternal and active support, you always had a solution when I struggled with a problem, it was sometimes challenging to achieve, but you always guided me to think outside the box. Most importantly, you were able to pull me back to the right path (even without me knowing) when I was lost. Thank you for independently implementing most of the methods we tried helping me to boost my confidence and also to fix bugs! I have learned a lot from you and appreciate your simple explanations given to complex concepts. I thank you for writing that machine learning book and allowing me to follow most of its content through your lectures during the first year of this study although our initial machine learning experiments are not in the thesis. The Latex template of Massey thesis that I used to create this document made my life easier during the last few months of the study. I also appreciate your financial support that helped me to focus on studies. Overall, I was blessed to have a supervisor like you. Isabel, thank you for introducing me to the avian world and to the field of conservation, your input, time, and energy into the project has been invaluable. I do believe that you were the main force to shape the project towards conservation opening a fairly large network of ornithologists, bird conservationists, and community groups. Thank you for guiding me in the field, particularly to handle the manual recorders in the field and to carefully get close to the birds for close-range recordings. It was certainly a fun time that I spent in the field with you. Thanks you so much for being

so generous with your time and your hours spent on spectrogram readings. Thank you for encouraging me to face the challenges, being there with me, and giving me moral support as a supervisor as well as a friend.

I owe a big thank to Dr. Amal Punchihewa, co-supervisor, for accepting me as a PhD student in first place, initiating this project, helping me all the way to Massey. It was you who gave me the initial idea of this research project. Subsequently, you helped me to develop the idea and design this project. Although when I arrive in New Zealand you transferred your role as the main supervisor to Stephen and moved overseas, you continued to support me remotely by helping me in various ways. You have been reachable throughout the project despite your tight schedule and time difference between the countries. You also helped me to improve my presentation skills by listening to my trial presentations via Skype and making constructive comments. I am really grateful to you for your all encouragement, kind advice and guidance throughout the project. I must also thank the referees who reviewed the journal article, as well as to the conference attendees who posed important questions and suggestions related to my research.

I am grateful to the volunteers who turned up with a short notice and helped me in the field with playback re-capture experiments, Natasha, Tim, Sumudu, Catherine, Kim, Julia, Ross Bell, Myung Jong, Anindya, Asif, Tobby, Emma, Giulian, Janna, Kelly, and Shari. Again thank you Natasha, Tim, Sumudu, and Catherine for your ability to be positive even when lots of supplejacks are ahead and need to be passed to make a straight line of recorders! Thank you Sara Treadgold (DOC, Whanganui branch), for lending me acoustic recorders, Dave Bell (Coordinator, NZ Falcon Survey) for connecting me to the right people (and also for donating us a Song Meter recorder with all the accessories), and Emma Williams for lending me few more acoustic recorders. Thank you, Paul Barrett and Cleland Wallace for your great technical support, for making the essential setups in-house. A big thank to Gary Mack (Massey Grounds Manager) not only for facilitating my use of the rugby field for the playback experiments but also to the kind efforts taken to minimise the machine noise by pausing the maintenance work during the trials. I am thankful to James Lambie (Science Coordinator, Horizons Regional Council) who provided access to the Pohangina Reserve and all the details about the tracks and safety matters in the forest. I would like to thank Ellen Schooner (Ecology Group) for connecting me to James and also for facilitating me to join with the Mokoia Island field visit. Without all your help Chapter 3 would simply not exist in this thesis.

Thank you Emma Williams, our bittern consultant, for welcoming me to the Hatuma lake and facilitating me (and my family) to spend several nights there and feel/-hear the amazing bittern booms; thanks for the inspirational moments you shared with

me. You helped me throughout the thesis providing time-stamped bittern recordings and opening new connections to the project. I saw your true dedication to save bitterns and you motivated me, which was so useful throughout this journey. To Rebecca, I was fascinated by your desperate work to save Samoan birds, particularly the Manumea (tooth-billed pigeon) and Ma'oma'o. They led me to test and confirm our methods on Manumea recordings, although this is not included in this thesis. I also thank Alex Brighten for giving me the opportunity to tune the selected commercial software to process a year-around continuous morepork recordings made in Ponui Island. Even though it took lots of my time, it was a great experience that convinced me of the difficulties associated with my research. Also, I highly appreciate the support given in the field during my visit to Ponui and providing me the whole morepork dataset to run my experiments. During the software tuning for morepork, Jeff Knewstubb gave me a great support as an experienced user and I appreciate that. I would like to thank Andrew Digby (Science Advisor for kākāpō and takahē, DOC) for your input to this project in several ways. Your nicely organised little spotted kiwi recordings (both manual and autonomous) and kākāpō recordings (collaborating with Bruce C. Robertson, University of Otago) were very useful for this thesis. I appreciate your suggestions and also networking me with other potential groups within DOC. I am also thankful to Les McPherson who kindly shared some birdsong from his Natural History Unit Sound Archive (http://www.archivebirdsnz.com/) with me at the beginning of the study.

I thank my home university, University of Kelaniya, for offering me an ideal environment in which I felt free to concentrate on this research. I acknowledge the funding received from Higher Education for the Twenty-first Century project (HETC), Sri Lanka. I am also grateful to the funding received from SEAT, Massey University. Financial support and the Conservation Innovation Award 2014 received from WWF facilitated and encouraged me to complete this thesis. I also appreciate the recent research fund granted from J S Watson Trust 2016, Forest & Bird to continue software implementations extending this thesis, particularly for kiwi conservation involving with community groups. I thank Richard Witehira (Blandy) and Jason Taiaroa for helping to setup the community groups for the latter. Thanks to Dilantha Punchihewa, Michele Wagner, Linda Lowe, and Karen Pickering for all the administrative support throughout the research.

I thank my office-mates and friends, Pramila, Ishani, Nadee, and Nisansala. Last, but not least, I am very grateful to my family – parents for their support, upbringing, and believing in me – loving husband Sumudu and our little princess, Dinara, for all your dedication during this research. Sumudu not only took the responsibilities at home providing a perfect environment for me, but also proofread the entire thesis. Dinara, you made me happy during the hard times – I love your surprise picnics!