UNIVERSITY OF SOUTHAMPTON

Smartphone-Powered Citizen Science for Bioacoustic Monitoring

by

Davide Zilli

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the

Faculty of Physical Sciences and Engineering
Electronics and Computer Science

June 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

by Davide Zilli

Citizen science is the involvement of amateur scientists in research for the purpose of data collection and analysis. This practice, well known to different research domains, has recently received renewed attention through the introduction of new and easy means of communication, namely the internet and the advent of powerful "smart" mobile phones, which facilitate the interaction between scientists and citizens. This is appealing to the field of biodiversity monitoring, where traditional manual surveying methods are slow and time consuming and rely on the expertise of the surveyor.

This thesis investigates a participatory bioacoustic approach that engages citizens and their smartphones to map the presence of animal species. In particular, the focus is placed on the detection of the New Forest cicada, a critically endangered insect that emits a high pitched call, difficult to hear for humans but easily detected by their mobile phones. To this end, a novel real-time acoustic cicada detector algorithm is proposed, which efficiently extracts three frequency bands through a Goertzel filter, and uses them as features for a hidden Markov model-based classifier. This algorithm has permitted the development of a cross-platform mobile app that enables citizen scientists to submit reports of the presence of the cicada. The effectiveness of this approach was confirmed for both the detection algorithm, which achieves an F_1 score of 0.82 for the recognition of three acoustically similar insects in the New Forest; and for the mobile system, which was used to submit over 11,000 reports in the first two seasons of deployment, making it one of the largest citizen science projects of its kind.

However the algorithm, though very efficient and easily tuned to different microphones, does not scale effectively to many-species classification. Therefore, an alternative method is also proposed for broader insect recognition, which exploits the strong frequency features and the repeating phrases that often occur in insects songs. To express these, it extracts a set of modulation coefficients from the power spectrum of the call, and represents them compactly by sampling them in the log-frequency space, avoiding any bias towards the scale of the phrase. The algorithm reaches an F_1 score of 0.72 for 28 species of UK Orthoptera over a small training set, and an F_1 score of 0.92 for the three insects recorded in the New Forest, though with higher computational cost compared to the algorithm tailored to cicada detection. The mobile app, downloaded by over 3,000 users, together with the two algorithms, demonstrate the feasibility of real-time insect recognition on mobile devices and the potential of engaging a large crowd for the monitoring of the natural environment.

Contents

List of Figures

List of Tables

Acknowledgements

First of all I would like to thank Dr Merrett and Prof Rogers for their excellent supervision, continued support and thorough feedback throughout my PhD programme, as well as Dr Costanza and Dr Chesmore for their valuable corrections. A big thanks to Prof Bullock and the ICSS team for running such an outstanding doctoral training programme.

I would like to thank Dr Trilar and Prof Gogala who have kindly supported our work, providing both materials and guidance, and Paul Brock, Sarah Henshall, Bjorn Beckmann, Jim Mitchell and everyone who taught us the ways of entomology.

Many thanks also to the two groups that hosted me as visiting researcher, Carlo Ratti and Yaniv Turgeman at SENSEable City Lab and Lucas Joppa, Drew Purves and Stephen Emmott at Microsoft Research Cambridge. To all these colleagues and mentors I owe a great deal of personal and professional growth, as well as some of the happiest months in my life.

Many thanks to my finaceé Chantelle and my brother Daniele for putting up with me during these years. I guarantee this saves you at least 3 years of purgatory, if that's of any consolation.

Many thanks to all the colleagues who touched my life in these years, and made sure that I didn't have enough time to make a better job of this thesis. Valentina, Woody, Nima, Domenico, Kier, Tom, Rob, the building 32 crew, thank you all. And all the other good friends in Southampton, Arinze, Aurore, Verity, Joe, John, Tsiloon, Issy—gosh I'm gonna be in trouble if I miss somebody—the Verbum Dei sisters and everyone at CathSoc, as well as those far away, Riccardo, Ankita, Vanessa, Piero. You surely made this journey one of a kind.

Last but not least I would like to thank my parents for their continued support for so many years. Your example as scientists has been a great motivation for me to undertake this journey.

This work was supported by an EPSRC Doctoral Training Centre grant (EP/G03690X/1) and by the ORCHID project (http://orchid.ac.uk).

I am strong when I am on your shoulders,
You raise me up to more than I can be.

Chapter 1

Introduction

Historically, science has been a matter for a few qualified professionals who had the education, tools, time and money to dedicate to pursuing further knowledge. However, as education has spread to a wider range of citizens, more people have gained access to the tools that allow them to participate in scientific research. A notable example is the collection of star observations by amateur astronomers, started in 1911 with the foundation of the American Association of Variable Star Observers (Saladyga, 2012b). The participation of a larger crowd in the scientific process has thereupon favoured faster progress, as more data could be collected, analysed and evaluated in a shorter time (Raddick et al., 2009). This process, often referred to as *citizen science* due to the involvement of the general public in scientific research, has been widely present for at least a century, but has received extraordinary interest in recent years.

This renewed interest may be mostly attributed to the introduction of novel methods of communication and tools for sharing knowledge worldwide, in particular the internet and the World Wide Web. These tools have been key to the development of citizen science as they have enabled global real-time communication between people who could discover communities of users with similar interests in order to share their findings. This has also effectively eliminated many economical barriers, as anyone with even sporadic access to the internet could participate in this effort.

Most recently, the uptake of powerful mobile phones has further facilitated this renewed interest for citizen science. Modern *smartphone* devices are equipped with a range of different hardware sensors (such as a camera, microphone, accelerometer, proximity sensor, compass, gyroscope and GPS) which, when combined with high computational power and internet connectivity, produce an incredibly powerful, environment-aware device (Lane et al., 2010). In addition to this, short-range wireless networking interfaces (Wi-Fi, Bluetooth, infrared) permit further enhancement through the connection to other gadgets. This hardware, originally designed to add simple user-experience improvements to standard mobile phones, such as the rotation of the interface upon tilting the device, has allowed, in recent years, the development of a range of highly varied applications.

These applications are often found not just in those fields for which these devices were designed, such as social media and communications, but also in many other areas, such as pervasive health care systems (De Jager et al., 2011), traffic monitoring (Mohan et al., 2008; Raddick et al., 2009), environmental sensing (Mun et al., 2009) and e-learning (Kukulska-Hulme and Traxler, 2005), to mention but a few. This use was promoted by opening development through public APIs, and further facilitated by "app markets", such as the Apple's "App Store"[1] and "Google Play"[2] (formerly known as "Android Market"), where developers can share applications with users, often making a profit from selling the software or additional services for a few pence (Lane et al., 2010). The number of individuals in possession of such a device is steadily increasing, providing a vast user-base (Sedghi, 2011) and most of these individuals carry their devices with them wherever they go (Landrey, 2012).

This explosion of use has not been limited to first-world countries. In developing nations, mobile phones are also widely used, and they often represent the main tool with which people connect to the internet for communication and educational purposes (Traxler and Kukulska-Julme, 2005; Kulkarni and Agrawal, 2008). There, the dissemination of smartphones is not as high due to their cost and for this reason more traditional *feature phones* are standard. However, basic sensors such as microphones and low-resolution cameras may still be present on the device and, as cheap smartphones[3] are becoming available on the market, more access to advanced functionality will be available even in developing countries.

[1] Apple's App Store at http://itunes.apple.com/gb/genre/ios/id36?mt=8

[2] Google Play at https://play.google.com/store

[3] As an example, Nokia, Alcatel, ZTE and Samsung offer smartphones with different operating systems for under £50 in the UK, as of the first quarter of 2015. Lowest-end Android tablets and laptops (often called "netbooks") are sold for a similar price.

This potential has therefore had a great impact on the field of citizen science. Citizens now find themselves equipped with very powerful sensors that allow them to perform all sorts of data collection and processing tasks. This provides a highly favourable environment for both the scientist and the citizen, where the former has access to an unprecedented amount of data, collected across a varied range of potentially thousands of individuals, while the latter is exposed to scientific research without necessarily being part of an academic institution. This process of outsourcing a data collection task to the general public falls under the wider umbrella of *crowdsourcing* (Howe, 2006); a technique that makes use of the intelligence and work force of a large number of generally untrained individuals (namely a *crowd*), as opposed to a restricted group of highly skilled specialists, to accomplish a task (Brabham, 2008). In this context, citizen science can be considered as the branch of crowdsourcing concerned with large data collection and analysis tasks. According to Wiggins and Crowston (2011), citizen science is different from any other form of participation in scientific research in that it involves active engagement of the participants. On the contrary, projects such as *SETI@Home* (Anderson and Cobb, 2002)—an effort to search for extraterrestrial intelligence with computation distributed across millions of volunteers' computers—only request computational resources from its participants, but no active involvement.

Due to this requirement of participants' active engagement, a clear challenge in crowdsourcing is the difficulty in motivating people's involvement. Several studies analyse the presence of intrinsic and extrinsic motivations for people to undertake such tasks (Rogstadius et al., 2011; Kaufmann et al., 2011). Monetary reward is often an option as demonstrated in successful crowdsourcing experiments such as the *DARPA Red Balloon Challenge* (Tang et al., 2011) and the *Tag Challenge 2012* (Rahwan et al., 2012). The former was a $40,000 challenge posed by the US Defence Advanced Research Projects Agency (DARPA) in 2009 to celebrate the 40$^{\text{th}}$ anniversary of the internet. The aim was to prove how modern social media could assist in finding 10 floating balloons scattered in undisclosed locations around the US. The MIT-based winning team achieved the target in less than nine hours, devising an incentive mechanism with which every participant who found a balloon would be rewarded $1000, the recruiter of that participant half the amount, and so forth up the recruitment tree. The Tag challenge 2012 was a competition in which participants were encouraged to locate five suspects walking around five known cities with easily-identifiable clothing, given only a picture of each suspect. This simulated law-enforcement exercise posed the challenge of coordinating different people around the world (as no one could be in all of the cities at the same time) in a time-constrained task, motivated by a $5000 reward. The winning team had a similar strategy to that devised for the Red Balloon Challenge, in which each person who verifiably identified a suspect was rewarded $500. In addition, each identifier's referer was awarded $100 and the first 2000 recruits conferred their recruiter a further $1 prize. The lack of a scientific target to these synthetic challenges make their nature a matter of pure *crowdsourcing*, rather than citizen science, as their only character of scientific interest is the strategy used by different teams to solve the task.

However, while social media have played an important role in the evolution of crowdsourcing and citizen science and have offered the possibility for the introduction of novel applications, these techniques have been around in more traditional forms for decades. One eminent example is found in the field of biodiversity monitoring. A long lasting tradition of bird watching and systematic reporting dates back to 1900 when an event called the *Christmas Bird Count* was initiated by Frank Chapman (Silvertown, 2009). In a 24 hour period, teams of volunteers would each gather to report the sighting of birds in a 15 mile radius. This data, collected over more than one hundred years and now digitally stored and freely available online, has since been used to compile statistics about the population dynamics of different species.

Such conservation biology tasks represent an ideal application of crowdsourcing. The task can be challenging for small groups of employed specialists, who can only cover a limited territory, but becomes easier if a large part of the population is involved in the effort. However, detecting the presence of some species may not be trivial, especially if these are small, nocturnal, elusive or in general difficult to spot at sight. Moreover, expertise may be required to identify the exact species of the specimen observed, to then report the observation accurately.

To overcome these challenges, one approach to detecting the location of animal species is to rely on the sound they make (Baptista and Gaunt, 1997). Under certain circumstances, in fact, many species emit a unique and characteristic sound, for example when they are in danger or when attracting a desirable mate, allowing for very precise identification of the species. Growing interest in this discipline, often referred to as *ecoacoustics* as a branch of *bioacoustics*, has been particularly shown by communities concerned with animals that emit very distinctive sounds, such as bats, birds and insects (Chesmore, 2004; Riede, 1998; Parsons and Jones, 2000). For example, bats use an ultrasonic system of echolocation of great intensity ($\approx$ 130 dB, louder than the maximum noise level permitted in a night club), despite humans being unable to hear it. Although it is necessary to use a specialist device to aid detection of their ultrasounds, bats can be more easily identified by this signal than by eyesight. An existing project on participatory bat monitoring presents the possibility of using a smartphone-enhanced bat detector to geo-tag and centrally collect observations of these mammals (Jones et al., 2009). Nevertheless, this approach still requires the purchase of supplementary hardware, i.e. the ultrasonic detector itself. This, however, is not the case when studying other animals such as birds and insects, as their call can often be heard by humans—and/or their smartphones—without the aid of a supplementary device (research by Gogala and Trilar (2004) show that certain insects are in fact more easily recognised by their call than by their morphology). From the application developer's perspective, a large number of users are available without being required to purchase any other hardware or enforced to carry an additional gadget. The microphone records sound that can be processed and elaborated to provide immediate feedback to the user, while geo-tagged data can be sent over the internet to be collected on a wider scale and to allow for further processing of the samples. Moreover, the automation of this process can decrease the level of expertise required from the user, as the task is delegated to an algorithm or jointly executed by the interaction of the user with the algorithm.

To this extent, the present work aims to show that crowdsourced biodiversity monitoring can be far more effective than its traditional counterpart when combined with automation techniques that can only be enabled by the use of modern equipment with sufficient computational power and rich in sensors. To prove this, it is here presented the case study of the New Forest cicada (Section 1.1), an endangered insect native to the New Forest in Hampshire, UK, which must be urgently rediscovered to prevent its extinction. The aims of this investigation are presented in Section 1.2, followed by the challenges and contributions of this work, presented in Section 1.3 and 1.4.

1.1 The New Forest Cicada

The New Forest cicada (*Cicadetta montana* Scopoli 1772, referred to hereon as simply 'the cicada') is an insect of the order Hemiptera, suborder Auchenorrhyncha and the only specimen of the *Cicadidae* family native to the United Kingdom. Here, it is only found in the New Forest, from which its name originates. Around Europe, it is known to be present in several countries, including France, Northern Italy, Slovenia, Germany and Poland (Trilar and Hertach, 2008; Sueur and Puissant, 2007b).

Male cicadas are equipped with sound producing organs, called tymbals, with which they emit high-pitch signals for the purposes of mating and locating. Female cicadas can respond to these sounds with short wing clicks. The sound produced is difficult to detect by humans, being at the boundaries of our hearing range (at a frequency centred around 14 kHz). While children are known to be capable of hearing this pitch, the natural decline in human hearing leads to most adults above 40 years in age being unable to detect this call (Hertach, 2007). Nonetheless, the intensity of the sound is very high, to the point that those who can capture the frequency are thought to be able to hear the sound even 60 metres away (Pinchen and Ward, 2010).

In England, cicadas have only been heard singing between late May and early July, and are commonly found in sunny, south facing, open deciduous woodland, with few bushes and wide clearings (Pinchen and Ward, 2010). However, no sighting has been confirmed since 1991, and the cicada is now considered highly endangered and potentially already extinct.

Nevertheless, an active effort has been made to search for the New Forest cicada. An official UK BAP (Biodiversity Action Plan) report, compiled by JNCC (Joint Nature Conservation Committee) for DEFRA (Department for Environment, Food and Rural Affairs) states the urgency of intervention necessary to save the New Forest cicada (Joint Nature Conservation Committee, 2010). The observed decline in the species' population is acknowledged to be an indicator for a "threatened habitat or conservation issue". The report highlights a need to explore different potential sites, as well as to collect information on surveying techniques and habitat management to preserve, or improve, the autecological requirements.

However, the lack of funding allocated to autecological monitoring has made it difficult in recent years to cover the entire area of the New Forest (not to mention the possibility of exploring new areas) for the few entomologists that were assigned to the task. To address this issue, the use of a crowdsourcing approach is proposed here, so as to involve a large number of citizens in the search. In fact, the New Forest is visited by millions of people every year (13.5 millions of day trip visitors alone according to the New Forest District Council) who, if provided with the right tools, could help in monitoring the presence of the cicada by covering a number of sites potentially never searched before.

1.2 Research Aims

Therefore, this research focuses on four key contributions:

1. Construct the appropriate bioacoustical tools for detecting the New Forest cicada, consisting primarily in an efficient machine-learning algorithm that can recognise its call.
2. Develop a system for citizen scientists to search for the insects, reporting to one central authority. Form a community around this endeavour to inform users and promote collaboration.
3. Prove the effectiveness of this method as a generic technique for the monitoring of species. The attention of this work is directed towards all British singing Orthoptera and related insects.
4. Develop a set of tools for the collaboration of wildlife sound classification experts to compare and benchmark different identification methods.

Finally this research aims to rediscover the presence of the New Forest cicada in order to protect its habitat and conserve its population. This is, of course, beyond the control and means of the methodology used, but will certainly be welcomed as a positive outcome.

These goals present numerous challenges, summarised in the following section.

1.3 Research Challenges

To allow the involvement of the *crowd*, the first step to be taken is to provide the end users with the appropriate tools. To this end, a visitor to the New Forest will have to be equipped with a method to detect and record the location of the cicada. The use of mobile phones to accomplish this task is enticing, as they provide all the necessary components and do not require visitors to buy or carry additional hardware. Hence, a smartphone application (or *app*) is required that aims to provide the following features:

1. Record sound from the built-in microphone. Issues are represented by microphone's sensitivity, user's interaction with the smartphone (e.g. where the device is kept while recording, if the instructions are followed by the user, etc.) and battery drain.

2. Detect the presence of a cicada call in real time and provide immediate feedback to the user. This is affected by recording quality, computational power of the device and effectiveness of detection algorithm.

3. Transmit observations to a central server to aggregate results and for further processing of sound samples. Recordings should be geo-tagged and time-stamped. The ability to relay the recording to the server should not depend on the presence of an active data connection at the time of the sample collection.

This outlines the additional requirement for a server-side infrastructure to collect and analyse recordings. From these prerequisites, numerous challenges arise. Firstly, the technical hurdles involved in developing a cross-platform app that can serve the largest possible number of users, that is engaging and easy to use for a non-technical public; and to power this with a web back-end capable of collecting observations, processing recordings and collating a community of citizen scientists.

Secondly, the implementation of a robust algorithm to detect the cicada call. This may not be easily distinguishable from background noise or other insects' calls, especially if the specimen is far, the quality of the recording is insufficient, or other sound sources are much stronger, filling the available bandwidth. To this extent, the signal should be analysed to reveal markers that characterise the call, that is its frequency, amplitude and duration in time. In species identification studies, these features are analysed to define the species *volume*, i.e. the amount of inter-species variability and the intra-species limits (Sueur, 2006). In the present work these features will be used to differentiate the call from overlapping sounds, in particular those generated by similar insects. One should consider that directional microphones are normally used for field recordings in bioacoustical studies, but the limit here is imposed by the recording capabilities of the hardware considered, i.e. smartphones of varying price and quality, potentially held with the microphone facing away from the sound source or even covered.

Thirdly, the analysis of the data collected to understand if any conclusive answer can be extracted. In particular, it may be possible to map locations where the cicada is not present, and use this information to potentially identify where it might be present. However, a key issue remains as to how the population can be motivated to use the tools and become involved in the project. Major challenges faced by current crowdsourcing models are related to motivating users to return to the project after the initial participation and to recruit other participants (Bell et al., 2008). The difficulty in understanding the emerging behaviour of people interacting with the system and with other users may require the investigation through a simulation model, which would facilitate the task of exploring avenues for incentive models. It should be noted that special care is required by this project on the incentive front, as an excessive invasion of the known sites where the cicada could be found is also not advisable, since that could disrupt its habitat with the potential for a disastrous effect on this endangered insect.

Furthermore, the trustworthiness of the data collected must be taken into account to avoid false reports (extensive research has been conducted in this area, for a recent example see Yu et al., 2012). For instance, malicious users could record the sound of the cicada in a different place, reproduce it in the New Forest and claim to have found the local insect. This becomes a more concrete risk when monetary incentives are offered, and therefore much more interest for malicious reporting arises.

When extending this process to classifying a larger set of species, a method must be found that balances maximum information gain with minimal number of features, in order to be able to represent compactly sounds that are very different. It is therefore paramount to understand what set of features can maximise the information content across all the species considered. Towards this goal, several classification algorithms can, given a set of features of a known set of samples, learn a profile for each class to be used to distinguish unknown samples.

1.4 Research Contributions

Of the outlined objectives and challenges, this research has contributed to the field in the following ways:

1. A novel real-time detection algorithm for insect classification capable of running on a smartphone has been developed, addressing the gap in the literature of a mobile-based insect detection system with immediate feedback to the user. The algorithm is designed to be efficient to avoid draining the mobile device's battery and to be responsive in real-time even on more constrained handsets. It is capable of detecting the presence of the New Forest cicada even from low-quality smartphone recordings, and to distinguishing it from two other species of insects, which are found at the same time of the year in similar habitats. The algorithm is designed in such a way that permits extension and adaptation to different scenarios, by employing an established machine learning technique for the classification of different sound sources in the signal. This contribution has been published in the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013) (Zilli et al., 2013), where the paper has received the *Outstanding student paper award*.

2. Around this algorithm, a citizen science platform has been built and deployed to engage the general public in the search for the New Forest cicada. Users are able to submit positive and negative results of a survey for the insect in a specific location and can interact with the system by analysing the surveys they have completed and submitted. This system hopes to serve as a reference for future citizen science projects around the fields of computational sustainability and biodiversity monitoring. This contribution, together with the aforementioned algorithm, has been reported extensively in the award-winning track of

the Journal for Artificial Intelligence Research (JAIR) (Zilli et al., 2014). The interaction of citizen scientists with the system and its effects on the development of similar technologies has also been described in a paper published in the ACM CHI Conference on Human Factors in Computing Systems 2014 (Moran et al., 2014) and in the 3rd IEEE International Workshop on the Social Implication of Pervasive Computing for Sustainable Living (SIPC '14) (Pantidi et al., 2014).

3. The sensitivity of different smartphones' microphones and their ability to detect the cicada has also been compared so as to compile a data set of mobile devices to be used as a reference in future similar projects. This lead to the ability to tune the algorithm to match a specific device, improving the overall classification.

4. To extend this approach to multiple species, a generic insect classification algorithm has also been developed, which demonstrates the application of the same principles to all British singing Orthoptera. This also constitutes a platform to benchmark different classification approaches in the wider wildlife sounds classification literature, which will be released to the community upon completion of this research.

1.5 Thesis Outline

This thesis presents related work in Chapter 2, focusing on the two key research areas, that of smartphone-assisted citizen science applications in Section 2.1 and that of bioacoustics in Section 2.2, while reviewing signal processing techniques for sound analysis and machine learning methods for insect classification (Section 2.2.3). The ecology of the New Forest cicada is also documented in Section 2.3, together with a brief note on similar species Section 2.4. It then reports in Chapter 3 the achievements of this research towards an efficient algorithm for automated cicada classification, which resulted in a hidden Markov model-based classifier, assessed for accuracy and performance against a state-of-the-art approach. Chapter 4 shows how this algorithm has been ported to a cross-platform mobile app, now deployed and in use by hundreds of citizen scientists. The chapter also reports an analysis of the users, their devices and the locations surveyed for the presence of the cicada, as well as a comparison of modern smartphones based on the sensitivity of their microphone. Chapter 5 presents a novel adaptation of the literature on bird classification to the domain of insect calls, testing the approach on different data sets. This extends the classification of cicada calls to encompass all Orthoptera in the UK. Chapter 6 concludes and suggests some avenues for future work. The appendices provide additional information about the Application Programming Interfaces (APIs) of the *app* (Appendix A); a sample report of the data collected (Appendix B); a description of the data sets collected and used in the benchmarking of the algorithms presented (Appendix C); the original software requirements (Appendix D); and finally a list of awards and media engagements of the project spawned by this research (Appendix E).

Chapter 2

Background

This chapter describes relevant related work in the three broad areas in which this project is involved: citizen science, bioacoustic techniques and the ecology of the New Forest cicada.

2.1 Crowdsourcing and Citizen Science

This research investigates whether the involvement of citizens in biodiversity monitoring is feasible and could bring substantial benefits to the field. This mass involvement falls under the umbrella terms of *crowdsourcing* and *citizen science*, terms that are often used interchangeably. However, these are not synonyms and some key differences should be drawn to better define the problem.

Crowdsourcing is a newer term, coined by Jeff Howe in 2006. Born as a portmanteau of *crowd* and *outsourcing*, it represents the process by which the intelligence and workforce of a crowd can be exploited to accomplish a task. Companies and institutions that normally assign employees to a job choose to offer that same job, in the form of an open call, to the wider public (Howe, 2006). This is then often performed through the distributed collaboration of several peers, but sometimes even just by single individuals.

The value of the crowd has been demonstrated at many different levels. Firstly, a large collection of users often performs better then a selected few, especially when the task is simple and easily divided. The best solution is then not the average of individual solutions, but the aggregation of all of them (Brabham, 2008). Secondly, the value of the solution could lie in the abundance of different options. This is the case of remarkable crowdsourcing successes such as *Threadless*, a company that calls for t-shirt designs which is to then print and sell a selection of the best ones, rewarding the designers with a prize considerable in value (US$1,500) but only a fraction of the profit made; or *iStockPhoto*, a stock photography website that collects photographs and videos from users. These are sold for much cheaper than a professional service could offer (low and medium resolution are normally between US$1 and US$5 respectively), of which the photographer receives 20%. Thirdly, users are often geographically distributed and therefore have free and easy access to a vast area, while moving employees from one place to another would be expensive. This is particularly the case when the problem has a highly distributed nature, for instance monitoring of a condition or an event across a wide area. For example, after the Fukushima nuclear disaster, citizens in Japan received, bought or built Geiger counters to monitor radiation levels (Plantin, 2011). This proved more effective than the sporadic information they received from the authorities, and, as readings were published on Pachube[1], an open stream data gatherer, the collation of distributed data enabled anyone around the world to monitor the radiation levels in Japan.

Citizen science is research conducted by amateur or non-professional scientists. It generally involves the collection and/or analysis of data by volunteers who dedicate their time and resources to a scientific investigation. It may happen that the core project is run by professional scientists and that amateurs collaborate to provide a contribution, however this is not always the case. In this respect, citizen science can be viewed as a branch of crowdsourcing, in that the mass involvement is targeted to data collection and analysis for scientific purposes. Active participation in the project undertaken distinguishes other types of public engagement in scientific research from citizen science; an example being the aforementioned SETI@home, where users only made their machine's resources available for computation but had no active role in the research (Wiggins and Crowston, 2011).

The practise of citizen science is, however, older than that of crowdsourcing, although the use of this term is relatively new (Silvertown, 2009). Early signs of this custom date

[1] *Pachube* has since been renamed *Cosm* first, and later *xively* http://xively.com/

back to the beginning of the twentieth century, especially in fields such as astronomy and ecology. The American Association of Variable Star Observers was founded in 1911 (Saladyga, 2012a) and since then 21 million variable star observations where made by amateur and professional astronomers, resulting in a plethora of publications and journals (Saladyga, 2012b). Similarly, in 1900 Frank Chapman started an annual *Christmas Bird Count* where teams of volunteers conducted coordinated monitoring of bird species and individuals in North America (Silvertown, 2009). Again this data, collected over a number of years and freely available online, has been used to observe trends of bird populations and their numbers.

However, it was only recently that citizen science became a highly widespread movement. A profusion of different projects have spawned in fields such as conservation biology, water quality monitoring, protein unfolding, population ecology and several other monitoring tasks. To provide an indication of the extent of this explosion, a few examples of major citizen science efforts are presented below.

Zooniverse (zooniverse.org) is a web portal that hosts different citizen science projects; started in 2007 with *Galaxy Zoo*, a successful astronomical endeavour to classify galaxies from telescopic survey data, it has now grown to incorporate some of the largest citizen science projects in different areas. In the field of biodiversity monitoring, several different competing websites attempt to collect and map the presence of different species around the World: iSpot (www.ispot.org.uk), developed by the Open University, Project Noah (projectnoah.org), Bug Guide (bugguide.net), Wild Lab (thewildlab.org), Evolution MegaLab (evolutionmegalab.org) and more. In environment monitoring, NoiseTube (noisetube.net) is an example of mobile-phone aided noise pollution monitoring. Foldit (fold.it) is a computer game that harnesses humans' problem-solving abilities to tackle one of today's hardest problems in biology, i.e. that of protein folding.

Different factors have facilitated the flourishing of citizen science activity. Silvertown (2009) identifies three main causes. Firstly, the progress in technology that made the collection and sharing of data accessible to everyone. The internet is here, of course, the principal player, but Silvertown predicted that mobile computing will also play an important role. Some years later it can be observed that this phenomenon is already happening on a large scale. A second factor is the realisation among professionals that citizen scientists constitute a "free source of labours, skills, computational power and even finance" (Silvertown, 2009). This becomes particularly prominent in the case of geographically distributed projects. Thirdly, research councils now require a substantial component of 'science outreach' in any funding they assign so as to justify the large use of taxpayers' money. Citizen science becomes an opportunity to demonstrate public involvement and wide applicability of a given research. Although some may argue that this devalues scientific enquiry, it is clear that councils strive to encourage applicable research, as opposed to only speculative theories.

However, despite the abundance of available citizens[2] and the ease of data collection provided by modern technologies, key concerns for citizen science project initiators are the recruitment, retention and motivation of users. The first step to achieve this is to obtain visibility for the project, independently from the recruiting strategy used (Bell et al., 2008). While print and broadcasting publicity can be expensive, local and online media can be cost effective and incisive. Once users are recruited, they then need to be motivated to actively participate, delivering good quality and reliable work. A number of strategies have been proposed and used to this extent; monetary rewards are common and effective. Examples that successfully use this strategy are the DARPA Red Balloon Challenge and the Tag Challenge described in Chapter 1. The nature of these tasks made a monetary reward ideal as the task was very simple and users could see a real potential of obtaining the reward. However, Nov et al. (2011) warn against monetary incentives as they believe they do not motivate users to provide the best quality work, but rather any result that would guarantee them the reward with the least amount of effort. A different strategy to motivate users is to foster competition amongst each other. CollabMap (Stranders et al., 2011) is a collaborative tool for human computation developed to generate geo-spacial data for evacuation routes; users are required to augment a map by a) drawing the outline of a building, b) drawing evacuation routes and c) verifying routes drawn by others. Here the developers have devised a strategy, built on top of that suggested by Bernstein et al. (2010), by which users receive a reward in terms of reputation for any of their activities that are positively verified by other users. Users with a high reputation are then awarded a monetary return, and this not only motivates users to return to the

[2]The reasons behind why people decide to dedicate their time to citizen science projects or other crowdsourcing projects, a notable example being *Wikipedia*, is a matter of fascinating discussions and much research, which however lies beyond the scope of this thesis.

website, but also improves the level of trustworthiness of the inputted data. Extensive experiments run by Bell et al. (2008) show that volunteers are best motivated by a combination of social, cognitive and emotional drivers.

To effectively motivate users it may be desirable, however, to consider the context in which the problem is set or in which the users are acting. The human participation component of citizen science projects generates uncertainty about the outcomes as infinitely different possible reactions of players generate a complex non-linear scenario. Simulating the experiments in software may therefore be essential to explore in advance some of the emergent behaviours that could be observed in the real-world experiment. This requirement is further discussed in Chapter 6, where the need for a simulation model is debated together with future goals of this research.

2.2 Biodiversity Monitoring Using Sound

As introduced in Chapter 1, the use of sound to monitor biodiversity can be particularly useful for those animals that are elusive and difficult to spot at sight. This section presents related work in bioacoustics, with particular focus on the techniques used and the methods for automation of the task.

Animals often produce sound for communication purposes, known as *non-incidental* sounds, or as a result of their activities and movements in the surrounding environment, such as eating or flying, referred to as *incidental* sounds (Chesmore and Ohya, 2004). Bioacoustics is the discipline that, combining biology and acoustics, studies the production, dispersion and reception of these sounds. An application of this discipline is the acoustic identification of species, a practise that has received formal attention from at least the late 1970s (Sueur, 2006). Several studies confirm that for the purposes of speciation the analysis of animal sounds can, in certain cases, be even more accurate than that of morphological traits (see, for example, Sueur and Puissant, 2007b; Gogala and Trilar, 2004). Moreover, detecting these sounds also proves to be particularly useful to spot animals that are difficult to see, but make a distinctive noise which sets them apart from other animals. This is the case, for instance, of birds, bats and insects, where this identification has aided surveying, monitoring and mapping of different species, which in turn is used for habitat conservation (e.g. Planitz et al., 2009; Riede, 1998; Laiolo, 2010). However, it is also the case for animals that are much larger and easy to see, such as elephants, as their low-frequency vocalisations can travel for several miles (see for example Payne et al., 2003; Clemins et al., 2005).

Bats emit a sound as a means to identify the location of and orient towards obstacles and prey. This system, called *echolocation*, is used by several other animals and partly by humans. The sound that bats produce is normally outside the range of human hearing (ultrasound), and therefore it is not possible to detect without dedicated instrumentation. In 1997, Vaughan and Jones presented a method for identifying bat species with multivariate analysis of their echolocation call, where they considered time-expanded recordings of known species, offering an initial classification according to the duty cycle of the call. The multivariate analysis performed on the recordings, called discriminant function analysis (DFA), was used to categorise species (the dependent variables), and is now common practise in several other studies (e.g. Obrist and Flückiger, 2004; Papadatou et al., 2008).

The sound produced by insects is often a byproduct of their movement, for example generated by the flapping of wings, such as in bees and flies. However, certain insects intentionally emit a sound, often called *song* or *call*, for the purposes of courtship or locating each other. An example of this is the cicada, an insect of the order Hemiptera widely spread around the world. Bioacoustic identification is largely used for cicadas, and has permitted in the past to distinguish morphologically similar species that were thought to be the same (Gogala and Trilar, 2004; Sueur and Puissant, 2007a; Hertach, 2007; Trilar and Hertach, 2008).

Ultimately, the automated identification of species is the target that bioacoustics is aiming to achieve in the context of systematics. For bats, but also for insects such as Hemiptera and Orthoptera, this has been an ongoing effort for more than a decade (see for example Parsons et al., 2000; Chesmore and Ohya, 2004). A number of common problems recurring in several bioacoustic applications—and the techniques currently used to solve them—are summarised below, starting from those reviewed by Parsons et al. (2000). The focus is placed on those tools that will be valuable for this work, the application of which is later presented in Chapters 3 and 5.

2.2.1 Processing Animal Calls

Different tools and techniques are used in bioacoustical research to analyse the sounds produced by animals. This section presents frequency transformation, amplitude demodulation and spectral analysis, and in the context of the latter, it introduces two techniques for feature extraction: the Goertzel algorithm and mel frequency cepstral coefficients (MFCCs).

2.2.1.1 Transforming Frequency

Frequency transformation is required when the sound under analysis is inaudible by ear. This is certainly the case for bats, where the typical frequency range varies between 12 and 160 kHz (note that humans can only hear frequencies between 20 Hz and 20 kHz), but it may also be necessary for some insects, such as the *Cicadetta montana*, which produces a sound of dominant frequency roughly between 12 and 17 kHz, difficult to hear for people over 40 years of age. The most common techniques for transforming frequency are heterodyne, frequency division and time expansion (Parsons et al., 2000).

Heterodyning is a signal processing technique by which two or more frequencies are combined to form new ones. In particular, the aim in bioacoustics is to produce a lower frequency signal that maintains most of the original properties while being audible to humans. This is achieved by mixing (i.e. multiplying) the original signal F_{orig} by one of similar frequency F_{osc} produced by an oscillator. Assuming F_{orig} and F_{osc} are two simple sine wave signals, $sin(2\pi f_1 t)$ and $sin(2\pi f_2 t)$, their product is:

$$\frac{1}{2}cos[2\pi(f_1 - f_2)t] - \frac{1}{2}cos[2\pi(f_1 + f_2)t]$$

due to the trigonometric identity:

$$sin(\alpha)sin(\beta) = \frac{1}{2}cos(\alpha - \beta) - \frac{1}{2}cos(\alpha + \beta)$$

This result is the sum of two frequencies, one at the sum of the two original ones and one at their difference. A low-pass filter is then applied to the output so that only the low-frequency signal is preserved. This technique works with a narrow-band input signal, a fact that constitutes its main advantages and disadvantages. The narrow-band in fact provides good signal-to-noise ratio and high sensitivity to the input, which is useful in survey work. However it also makes species identification difficult, for example in the case of bats, because different species emit echos at different frequencies and may be difficult to recognise after the signal has been processed. Moreover, this may lead to undersampling of species as the heterodyne receiver can only be tuned to detect one particular frequency, and therefore only species in that band. (Parsons et al., 2000).

Another technique for transforming frequency is *frequency division*. This approach simply divides the frequency of the input signal f_{orig} by a predefined value n in order to lower its frequency:

$$f_{out} = \frac{f_{orig}}{n}$$

The output is produced by counting the zero crossings of the input and allowing only every n^{th} cycle to pass through. The amplitude is kept constant and therefore it does not reflect that of the input signal. Noise can be reduced by filtering out those cycles with an amplitude greater than a given threshold. To reintroduce the amplitude of the input, in the final stage the output is then multiplied by the envelope of the input amplitude. Among the advantages of this technique is the fact that, compared to heterodyning, it is not limited to capturing specific frequencies and can therefore detect all sounds in the spectrum (Parsons et al., 2000). Several disadvantages are also experienced, most prominently the fact that no harmonic information is present in the output, that division ratio must be carefully chosen (as it determines what frequencies will be heard) and that information is lost in those cycles not represented in the output.

An additional method is provided by *time expansion*, which replays the input signal slower than the recording speed. This method is becoming increasingly popular as no information is lost in the output signal, making it ideal for spectral analysis. High cost of equipment, slow recording and processing rate and size and weight of equipment are so far some of the most important limitations of this technique. However, the increase of computational power in small devices (such as mobile phones) is making this method more appealing.

2.2.1.2 Amplitude Demodulation

Modulation of a signal is the process by which a fast changing periodic waveform (the carrier wave, $c(t)$) is combined with a slow changing waveform carrying information (the modulator, $m(t)$), resulting in a signal $y(t)$ that is:

$$y(t) = m(t)c(t)$$

The resulting signal can then be demodulated, extracting information by separating it from the carrier signal. The modulating signal can be varied in all the three principal parameters of a periodic waveform, i.e. frequency, phase and amplitude. Here amplitude modulation is considered, since the cicada call, as shown in Chapter 3, reveals that important information is carried in this component.

One technique to demodulate an amplitude-modulated (AM) signal is the use of an envelope detector, which can be implemented in hardware or in software. In software, a common method is the use of a Hilbert transform, a mathematical linear operator widely used in signal processing. This is a procedure applied to a real signal $x_r(t)$, yielding to a new real signal $x_{ht}(t)$ which is a 90-degree phase-shifted representation of the original $x_r(t)$ (Lyons, 2004). The real continuous time-domain signal $x_r(t)$ can be associated with a complex signal $x_c(t)$ such that:

$$x_c(t) = x_r(t) + jx_i(t)$$

called the *analytic signal*, where the imaginary part $x_i(t)$ is the Hilbert transform of the original $x_r(t)$ and j the 90-degree phase shift. From this, the instantaneous envelope $E(t)$ can be measured as:

$$E(t) = |x_c(t)| = \sqrt{x_r(t)^2 + x_i(t)^2}$$

meaning that the envelope is equal to the the magnitude of the original $x_c(t)$ (Lyons, 2004). This technique has been used in preliminary work to classify a feature of the New Forest cicada call, i.e. the presence of a pattern repeating roughly every 8 ms, as shown in Section 3.2.

Other techniques for demodulating a signal include the square and low-pass method (SLP) and probabilistic amplitude demodulation (PAD) (Turner and Sahani, 2011). The former isolates the modulator signal to low frequencies, so that they can then be extracted by a low-pass filter. The latter is a method that uses probabilistic inference to estimate the modulator signal. This technique appears to be robust to noise and capable of considering prior user-specific knowledge to adapt to different signals. These benefits come, however, with a greater computational cost.

2.2.1.3 Spectral Analysis

Useful information about an animal call can also be found in its spectral content. Three techniques are commonly used to analyse the frequency in bats' echolocation calls: zero-crossing analysis, Fourier analysis and instantaneous frequency (Parsons et al., 2000).

The first is a simple method which involves counting the number of times the input wave crosses the x-axis. This permits one to quickly convert a signal from the time domain to the frequency domain, as the signal will cross the x-axis twice for every cycle. This method is, however, very sensitive to noise and will only take into account the main component of the signal. Nevertheless, it is still a widely used approach in bioacoustics (Chesmore, 2001).

A much more common technique is *Fourier analysis*, which is based on Joseph Fourier's intuition that representing a function in the form of a series decomposes it into simpler components, easier to analyse. A notable example is the exponential function, which can be represented as the series:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} + \ldots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Here the first few terms give a good approximation and are easier to deal with (for example to integrate or differentiate) (James, 2011). In particular, a periodic function $f(t)$ with period $T = 2\pi/\omega$ can be represented as a Fourier series of the form:

$$f(t) = A_0 + \sum_{n=1}^{\infty} A_n \sin(n\omega t + \phi_n)$$

By expressing a function $f(t)$ as its Fourier series expansion, this is decomposed into its harmonic or frequency components. Given a period T, the function has frequency components at discrete frequencies:

$$\omega_n = \frac{2\pi n}{T} = n\omega_0 \quad (n = 0, 1, 2, 3, \ldots)$$

where ω_0 is the frequency of the parent function $f(t)$, called the *fundamental frequency*. This forms a discrete spectrum, providing a frequency-domain representation of the signal, useful to highlight prominent frequency components.

This technique for transforming a discrete function in the time domain into another function in the frequency domain is known as the discrete Fourier transform (DFT). The definition of the DFT describes a naïve implementation which is inefficient to compute, namely of computational order $O(n^2)$. However, an efficient algorithm to compute the DFT is the fast Fourier transform (FFT), which reduces the complexity to order $O(nlog(n))$.

Among the several advantages of the Fourier transform is that no information is lost from the original to the output signal (except for floating-point errors), and that it is a reversible process. Moreover, this method is not particularly sensitive to noise, providing a useful alternative when analysing animals' calls in an environment such as a forest, where overlapping sounds are present (Parsons et al., 2000).

Finally, a lesser known approach used in bioacoustic is that of estimating *instantaneous frequency*, a technique falling under the branch of time–frequency analysis in signal processing. This practice is better suited to short and rapidly changing signals (such as those emitted by bats), where classical Fourier analysis assumes an infinite periodic signal. Among the advantages is the fact that no information is lost in the process and that frequency and time can be analysed simultaneously and at high definition. This method is, however, very sensitive to noise and demands high computational power (Parsons et al., 2000).

2.2.1.4 The Power Spectrum

The spectrum of frequencies in a sound or other signal is often a good indicator of what source generates that signal. Taking an FFT over the entire signal reveals this range of frequencies, and for a discrete signal its resolution is proportional to the number of samples used for the FFT. Sometimes, however, it is effective to understand how those frequencies change over time, and therefore a short-term power spectrum is considered. This method consists of segmenting the signal into shorter time series and taking an FFT over each segment. This is referred to as short-time Fourier transform (STFT). The visual representation of the short-time power spectrum takes the name of *spectrogram*. This thesis, as the literature, makes wide use of this visualisation tool to explain a sound.

The power spectrum of a discrete signal is itself a discrete signal sampled at linear intervals across the frequency domain. Each sample corresponds to a value on the Hertz scale. However, this scale is at times transformed into a non-linear scale that emphasises particular frequencies. One such scale is the mel scale, which endeavours to match human hearing by expanding (allocating more samples to) the lower frequencies, and compressing higher ones. Introduced in 1937 by Stevens et al., this scale is often used in speech recognition, which is particularly concerned with lower frequencies, typical of human speaking and hearing (Plannerer, 2005). It is also common in birds and insect classification, as the wildlife sound community has often borrowed methods for the speech recognition literature. In particular, the mel scale is used in a representation of the power spectrum called the mel-frequencies cepstrum, described in more detail below.

2.2.1.5 The Goertzel Algorithm

Granted that these techniques can aid in analysing animal sounds, the need arises to extract the maximum information from the signals considered at the lowest cost and to represent it in the most compact way, whether that be for real-time feedback or for batch processing of large data sets. Many strategies have been devised for this purpose, and the choice of the appropriate one largely depends on domain knowledge.

The DFT analyses the entire spectrum of a signal. However, many insects, including the *Cicadetta montana*, sing at a specific frequency with minimal variance between individuals. Therefore a more efficient method for analysing a sound for the presence of this tone would be a technique for *single tone detection*. An efficient algorithm for this purpose is the Goertzel filter, which effectively computes a sparse FFT (Lyons, 2004). Using this filter one can avoid computing the entire transform, the majority of which would be discarded to only keep the output relative to the frequency of interest.

An efficient implementation of the Goertzel algorithm requires two steps. The first step produces a coefficient that can be pre-computed and cached to reduce CPU cycles:

$$c = 2\cos\left(\frac{2\pi f}{f_s}\right) \tag{2.1}$$

where f is the central frequency in question and f_s the sampling rate of the recording.

The second step consists of iteratively updating the values of a temporary sequence y with any incoming sample s_n such that:

$$y_n = \text{hamming}(s_n) + (c \cdot y_{n-1}) - y_{n-2} \tag{2.2}$$

where the samples are passed through a Hamming filter, given by:

$$\text{hamming}(s_n) = 0.54 - 0.46\cos\left(\frac{2\pi s_n}{N-1}\right) \tag{2.3}$$

and the length of the sequence of samples N determines the bandwidth B of the Goertzel filter, such that:

$$B = 4\frac{f_s}{N} \tag{2.4}$$

A shorter sequence length N yields a larger bandwidth, at the cost of a noisier output. In practice, we use multiples of 128 samples to match a typical smartphone's audio recording buffer size. For example, a block size of $N = 128$ samples gives a bandwidth of just under 1.4 kHz. The magnitude m of the frequency band centred at f and with bandwidth B in time slice t is then given by:

$$m_{t,f} = \sqrt{y_N^2 + y_{N-1}^2 - c \cdot y_N \cdot y_{N-1}} \tag{2.5}$$

In terms of computational complexity, this approach shows a considerable benefit compared to the single-bin DFT. As mentioned above, the FFT has a complexity of $O(NlogN)$, while the Goertzel algorithm is only of order $O(N)$, where N is the number of samples per window. Moreover, the sample update described in Equation (2.5)

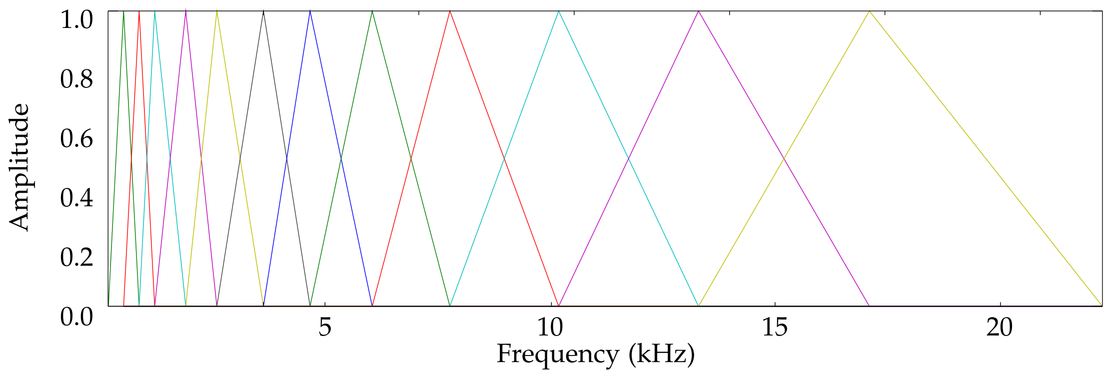


FIGURE 2.1: Mel filterbank with triangular filters

can be processed in real-time, eliminating the need for an independent background thread on the smartphone app and the need to store sample values.

2.2.1.6 Mel-Frequency Cepstral Coefficients

Another approach used to summarise the power spectrum of a sound signal is the mel frequency cepstrum, a representation based on the linear cosine transform of the log power spectrum on the mel scale of the frequency. The discrete set of coefficients that make up the cepstrum are called mel frequency cepstral coefficients (MFCCs). First introduced by Davis and Mermelstein in 1980, they have proved to be very robust against noise, and have been used in variety of different domains.

The extraction of MFCCs requires the following procedure (Lyons, 2014):

1. Frame the signal into short windows;
2. Take the FFT of each window;
3. Convert the powers of the frequencies obtained to the mel scale with the use of triangular overlapping windows, summing the energy in each filter;
4. Take the logarithm of the filterbank energies, that is the energy of each mel frequency bin;
5. Take the discrete cosine transform (DCT) of the log filterbank energies;
6. Select a number of coefficients of the DCT; the first one is normally discarded, and a number of the remaining ones is selected, for example 2-13.

MFCCs have a number of benefits as they have been modelled on the human cochlea. For starters, humans cannot detect variation in very similar frequencies, so a range of similar frequencies can be compressed into one value. On the mel scale, this becomes more pronounced in the higher part of the spectrum, where these differences are even more difficult to detect. Secondly, small changes in amplitude are also difficult to detect, and hence a change in the logarithm of the energy value resembles more closely the difference in amplitude we can detect. Thirdly, since the filterbanks are overlapping, they are quite strongly correlated with each other; the DCT decorrelates them, which ensures that they can be used with less bias in a classifier.

Having discussed some techniques to analyse sound and extract useful features from it, the use of these methods in the state-of-the-art literature will now be reviewed.

2.2.2 Automated Identification

Manual identification of species from their songs through expert surveys is common in bioacoustics and has been employed for several years (Chesmore, 2000). However, these surveys have strong limitations, namely the fact that they are time consuming and rely on the expertise of selected surveyors. Due to this, they tend to be performed sporadically, often leading to a lack of information on population trends (Chesmore, 2004). For years, automated identification of individuals and species has therefore been at the centre of research in systematics to solve problems such as group discrimination and intergroup characterisation (MacLeod, 2007). Possible techniques towards this goal are DNA barcoding and morphological image recognition, facilitated recently by the use of powerful computers and even powerful mobile devices. The potential of achieving this through sound on mobile devices has been identified by Chesmore (2004), who proposed a signal recognition system called IBIS (Intelligent Bioacoustic signal Identification System). This provides a time-domain analysis combined with an artificial neural network to recognise British Orthoptera.

Towards the goal of detecting the presence of a species from its sound, much can be learnt from the literature on Automated Taxon Identification (ATI) systems. In fact, the structure of such systems is defined as a common pattern recognition system (Chesmore, 2007), exemplified in Figure 2.2. A similar structure has been used in this work to identify the presence of the New Forest cicada. Chesmore (2007) describes

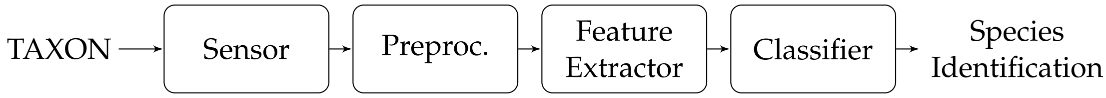


FIGURE 2.2: Block diagram of an Automatic Taxon Indentification system, reproduced from (Chesmore, 2007)

important parameters to consider for acoustic sensors: the frequency response and directionality of the microphone; the sampling frequency (which should be at least twice the maximum frequency of the input signal, called the Nyquist frequency); the accuracy to which the amplitude of the signal has to be represented (quantization); storage space concerns, as standard-quality uncompressed audio recording can weigh in the order of 5–10 megabytes per minute; and interference issues, in particular due to other animal sounds and the surrounding environment.

With the automation of the collection of large data sets of sound recordings, the need for automated techniques to analyse these signals efficiently becomes even more prominent. The algorithms to do so typically range from those that operate solely in the time domain, such as time domain signal coding (Chesmore, 2004; Chesmore and Ohya, 2004), to those inspired by the literature of human speech recognition (for example Potamitis et al., 2006; Pinhas et al., 2008). The latter typically use a hidden Markov model (HMM) for classification (Leqing and Zhen, 2010), and perform a number of pre-processing stages to extract features from the raw recording. For example, Chaves et al. (2012) present a state-of-the-art approach that pre-processes the recorded sound to remove un-sounded periods where no insect call is detected, that maps the raw frequencies to the mel scale; then it converts it into the *cepstrum*, the pseudo-time domain described above, by calculating a number of MFCCs, that are used as features for the HMM classification with just one state per species. Such approaches have been shown to classify insects to very high levels of accuracy from clean recordings collected using purpose-built hardware.

However, the present research argues that the use of the mel scale is not always beneficial for animal sound recognition. This is mainly due to the fact that this scale is designed to mimic human hearing and is therefore well suited to human speech, emphasising low frequencies where voice is more present. Similarly, it may be beneficial for the classification of a set of animals that emit low frequency signals, such as elephants and other large mammals, as evidence suggests that the majority of mammalian reception and vocal production systems are very similar (Clemins et al., 2005). On the contrary, insects sound cover a much wider range of frequencies, and compressing the higher end of the spectrum is not only unnecessary but potentially detrimental. This intuition is corroborated by a pilot system for automatic identification of insect songs by Ganchev et al. (2007), which uses linearly spaced filters between 2 and 22 kHz, spaced at 100 Hz from one another, arguing that insects calls can cover the entire spectrum of audible frequencies (and at times ultrasounds). Hence, the authors use linear frequency cepstral coefficients (LFCCs) as features to their classification system. Conversely, even with simpler features, HMMs are an efficient and scalable tool for the purposes of the system here proposed, and are therefore reviewed in greater detail in the following section.

In the wildlife sound classification domain, particularly relevant to the present research is a thorough investigation by Stowell and Plumbley (2014) on bird sounds classification. The authors compare three feature extraction techniques, MFCCs, mel spectra and learnt features through spherical k-means. Based on the established k-means clustering algorithm (Lloyd, 1982), the latter is an unsupervised method to extract information from a data set from its characteristics, without enforcing domain knowledge (hence *learning* the relevant features). This algorithm (Dieleman and Schrauwen, 2013; Coates and Ng, 2012) searches for unit vectors that minimise the angular distance, rather than the Euclidean distance, between data points (Stowell and Plumbley, 2014). The authors evaluate these techniques across four large data sets ranging between 0.8 and 77.8 hours, with single and multiple labels (i.e. more than one class present in each recording), with 77 to 501 different classes, and they find not only that feature learning through spherical k-means is beneficial, but also that the use the raw mel spectrum performs considerably better than MFCCs. This motivates our work, described in Chapter 5, to extend the approach proposed for cicada classification to a larger number of species in a scalable fashion. Stowell and Plumbley (2014) moreover summarise the features extracted in three different ways: by their mean and standard deviation across time, by their maximum and by modulation coefficients. The latter are calculated with a STFT along the time axis, which captures the temporal evolution of the features. They identify that in the case of birds the use of these modulation coefficients is not beneficial. In contrast we argue that, with a small improvement in the summarisation of this feature that consists in sampling the FFT spectrum on a logarithmic space, modulation coefficients can be useful for insects, whose sounds express a strong feature in the repetition of phrases at regular intervals. The coefficients can capture this repetition, improving the classification accuracy. The result of this investigation is reported in Chapter 5.

2.2.3 Classification Techniques

The process of classifying sounds starts with the extraction of the appropriate features, which has been covered in the previous section. The features obtained can fed to a classifier or estimator that will make a judgement as to what class a set of features related to an individual sample belongs to. This section reviews some of the relevant classification techniques, which have been selected at the intersection of state-of-the-art techniques and most appropriate tools for the problem this research aims to solve, that is the real-time classification of crowdsourced insect sound recordings.

2.2.3.1 Hidden Markov Models

A hidden Markov model consists of a Markov chain of discrete latent variables and a sequence of continuous observed variables, each of which is dependent upon one discrete variable's state (Blasiak and Rangwala, 2011). Figure 2.3 shows the graphical structure of a HMM, where the discrete, hidden variables are represented by the sequence $z_1, \ldots, z_T$, and the continuous, observed variables are represented by the sequence $x_1, \ldots, x_T$. The value of each discrete variable z_t corresponds to one of K states, while each continuous variable can take on the value of any real number.

The behaviour of a hidden Markov model is completely defined by the following three parameters. First, the probability of each state of the hidden variable at $t = 1$ is

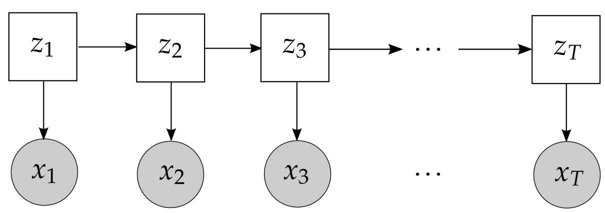


FIGURE 2.3: A hidden Markov model. Unshaded square nodes represent hidden discrete variables, while shaded circular nodes represent observed continuous variables.

represented by the vector $\boldsymbol{\pi}$ such that:

$$\pi_k = p(z_1 = k) \tag{2.6}$$

Second, the transition probabilities from state i at $t-1$ to state j at t are represented by the matrix $\mathbf{A}$ such that:

$$A_{i,j} = p(z_t = j | z_{t-1} = i) \tag{2.7}$$

Third, the emission probabilities that describe the observed feature, $\mathbf{x}$, given parameters $\boldsymbol{\phi}$, follow a log-normal distribution such that:

$$x_t | z_t, \boldsymbol{\phi} \sim \ln \mathcal{N}(\mu_{z_t}, \sigma^2_{z_t}) \tag{2.8}$$

where $\boldsymbol{\phi} = \{\boldsymbol{\mu}, \boldsymbol{\sigma}^2\}$, and μ_{z_t} and $\sigma^2_{z_t}$ are the mean and variance of the Gaussian distribution for state z_t.

Equations 2.6, 2.7 and 2.8 can then be used to calculate the joint likelihood of a hidden Markov model:

$$p(\mathbf{x}, \mathbf{z} | \boldsymbol{\theta}) = p(z_1 | \boldsymbol{\pi}) \prod_{t=2}^{T} p(z_t | z_{t-1}, \mathbf{A}) \prod_{t=1}^{T} p(x_t | z_t, \boldsymbol{\phi}) \tag{2.9}$$

where the model parameters are defined by $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\phi}\}$.

The most likely sequence of hidden states for a given observation sequence can be found with the max-sum algorithm, known in HMMs as the Viterbi algorithm (Bishop and Nasrabadi, 2006; Viterbi, 1967). This is different from the sequence of most probable states, which may have zero probability, should two adjacent states, individually most likely, not have any possibility of being connected. From the joint distribution above, we can obtain the probability of the most likely sequence $\mathbf{z}_1, \ldots, \mathbf{z}_T$ producing the observations $\mathbf{x}_1, \ldots, \mathbf{x}_T$, called here $\omega(\mathbf{z}_T)$, by taking the natural logarithm and exchanging maximisation and summation, such that:

$$\omega(\mathbf{z}_T) = \max_{\mathbf{z}} \ln p(\mathbf{x}, \mathbf{z}). \tag{2.10}$$

Where the model parameters $\boldsymbol{\theta}$, fixed when finding the most probable sequence, were omitted for clarity. This can be computed recursively (Storm, 2012) as:

$$\omega(\mathbf{z}_t) = p(\mathbf{x}_t|\mathbf{z}_t) \max_{\mathbf{z}_{t-1}} \omega(\mathbf{z}_{t-1}) p(\mathbf{z}_t|\mathbf{z}_{t-1}) \tag{2.11}$$

where the basis is represented by:

$$\omega(\mathbf{z}_1) = p(\mathbf{x}_1, \mathbf{z}_1) \tag{2.12}$$

This method, efficient because growing only linearly with the length of the feature, has been applied to construct an efficient algorithm for cicada detection, described later in Chapter 3.

2.2.3.2 Decision Tree Learning

A simple way of selecting a class for a specific input sample is by traversing a binary tree with a sequential decision-making process based on a threshold for each feature value (Bishop and Nasrabadi, 2006). This non-parametric supervised learning method is called a *decision tree*.

Decision trees have many advantages. Firstly, they can be visualised easily, are therefore easy to interpret, and require little effort in preparing the input data, even though some effort must be made to balance classes that are too dominant. They can also handle multi-output problem, such as the classification of sound recordings where multiple species are singing in each sound sample. They can handle both numerical and categorical data, and their accuracy can be assessed with statistical tests.

They also have, however, some disadvantages. They are susceptible to over-fitting, meaning that they can learn over-complex structures that do not generalise to subsequent data they are given. In terms of learning, finding the optimal decision tree is NP-complete, and implementations of the algorithm are therefore not guaranteed to find a global optimum. Furthermore, they can be unstable, as small differences in the input data can generate drastically different trees (Witten and Frank, 2005).

However, some of these drawbacks can be solved or mitigated by using an *ensemble* of estimators, that is a number of different decision trees or other classifiers, whose output values are aggregated to improve robustness and accuracy. The aggregation can be performed either by taking an average of independent results, in which case the ensemble is said to be using a *bagging* method, or by building a series of estimators that improve on the previous result, and taking a weighted average of the those, which constitutes a *boosting* method.

A random forest classifier is one such example of an ensemble that averages on a number of decision trees. However, a random forest also selects a subset of the features in the input space during the learning process, so as to reduce the correlation between trees. In fact, if one particular feature is a strong predictor of the outcome variable, many trees will select this feature, introducing a correlation among themselves. By selecting a random subset of features for each tree, the algorithm decreases the correlation between trees.

Both HMMs and decision trees have been used in the present work for the classification of sound recordings, and their performance in our setting is described and evaluated in Chapter 3 and 5.

2.2.3.3 Other classification techniques

Several other machine learning techniques have been used in the literature to classify animal sounds. Although not directly used in the present research for the reasons highlighted above and further discusses in Chapter 3, a brief outline of the most commonly occurring alternatives is provided below.

Linear discriminant analysis (LDA) is a statistical method with a linear decision boundary. A discriminant function is one such function that assigns an input vector $\mathbf{x}$ to a class $\mathcal{C}_k$ of a set of K possible classes (Bishop and Nasrabadi, 2006). Among its advantages are the inherent support for multi-class decisions, the easy computability and the absence of parameters to be tuned. Due to its simplicity, the method is widely used across the literature (for example in Simmonds et al. (1996) for fish and in Parsons and Jones (2000) for bats).

Support vector machines (SVMs), on the contrary, are not inherently multi-class. However, due to the common need for more-than-binary decisions, different methods have been proposed to combine two-class SVMs in order to provide a multi-class classifier (Bishop and Nasrabadi, 2006). The advantages of SVMs are found mostly in memory efficiency, versatility and effectiveness in higher dimensions, even when the number of dimensions is greater than that of samples (though in such cases the accuracy decreases significantly). The output estimates of SVMs are calculated through cross-validation, which can be computationally expensive. Examples of successful use of SVMs are found in classification of amphibians and birds (Acevedo et al., 2009; Fagerlund, 2007) and bats (Redgwell et al., 2009).

Artificial Neural Networks (ANNs) are a class of machine learning algorithms also commonly found in species sounds classification. Among these, the feed-forward neural network, or *multilayer perceptron*, is considered the most successful model (Bishop and Nasrabadi, 2006) and consists of multiple, fully-connected layers of nodes in a directed graph. For classification, the model is trained with a technique called *error back-propagation*. The models generated by training such a neural network are often more compact than an SVM, at the cost of more expensive training (Bishop and Nasrabadi, 2006). However, it is often acceptable to have a costly training procedure in order to produce a compact model that performs more efficiently on classifying new data. Once again, examples of using neural networks in bioacoustics are found in the classification of bats (Parsons and Jones, 2000; Redgwell et al., 2009; Walters et al., 2012), insects (Chesmore, 2001) and birds (McIlraith and Card, 1997).

Having described the classification techniques used in this research and briefly mentioned alternative approaches, the remainder of this chapter will introduce the ecology of the insect that motivates this research, the New Forest cicada.

2.3 The New Forest Cicada and Other Insects

New Forest cicada is the common name given to the only species of *Cicadidae* found in England, the *Cicadetta montana sensu stricto* (Scopoli 1772). First seen in 1812 in the south of the New Forest, it has since only been observed there (except for a few sightings in Surrey), and from there it received its name. Despite their rarity in England, *Cicadidae* are widely distributed outside the UK, predominantly in Southern Europe and Asia, where due to their abundance they sometimes reach the status of

FIGURE 2.4: *Cicadetta montana*, photograph by Jaroslav Maly, reproduced with permission.

pest (Pinchen and Ward, 2002). On the contrary, in England the presence of the cicada has always been sporadic, and the largest group ever reported was of 100 singing males in 1962, while the last confirmed sighting dates back to 1993. For this reason, it is now considered highly endangered (Joint Nature Conservation Committee, 2010), and some arguably consider it already extinct. However, lack of reports between 1941 and 1962 also resulted in its believed extinction until a colony was discovered on the northern edge of the New Forest.

Literature regarding the life cycle of these cicadas in the New Forest is sparse. The adult phase typically lasts between two and four weeks, in a period that extends to at most from late May to mid July. Cicadas occupy the scrub layer, but males often fly into the canopy (even as far as 12 m high) to sing. This makes them difficult to see and might hinder their detection with standard microphones. In contrast females normally bask on stems, where they feed on twigs, leaving traces of their presence. Cicadas are known to feed on oaks, beech, birches, hawthorn, small leaved lime or bracken, sucking the phloem of these plants (Pinchen and Ward, 2002). Another sign of their presence is represented by the oviposition marks; cicadas lay their eggs in small-diameter stems of herbaceous plants, creating W-shaped marks where they insert and hide their eggs.

Depending on the external temperature, eggs hatch after 50 to 125 days, after which nymphs find a suitable root, excavate an underground chamber around it, and live in

(A) Closed turret (B) Turret exit hole

FIGURE 2.5: Two *potential* turrets spotted in the summer 2014 in the New Forest. No other sign of cicada was detected in those locations. Photographs by Paul Brock.

shelter sucking its sap. Nymphs can stay underground for years before deciding to emerge. In the spring of the year they emerge, they construct a *turret*, a conical structure at the entrance of their chamber (Figure 2.5). This constitutes another detectable indicator of their presence. The use of this turret is still unknown, but it has been speculated that it could provide insulation to the chamber. When they emerge, early in the morning, they remove the turret and position themselves on the surrounding vegetation, where they *ecdyse* to adults.

During their adulthood, male cicadas sing to attract a partner. Two distinct calls are produced, one for locating each other and one for courting purposes, which differ mainly in duration. The call is produced in the thoracic chamber by a pair of tymbals rapidly clicking a drum membrane. The locating call is composed of two short warming-up chirrups of 2–3 seconds, the length of which may vary. The courtship song can last several minutes, although it is normally in the range of 30-40 seconds. It starts slowly at low amplitude and then increases progressively in volume, eventually stopping quite abruptly, as can be observed from the oscillogram in Figure 2.6.

The call has been described as a faint, high-pitched ringing buzz (Pinchen and Ward, 2002), with a frequency starting quietly at 4 kHz to then increase in intensity and frequency, stabilising around 16 kHz, where the majority of the call is produced. It is similar to the call emitted by the Roesel's bush cricket (*Metrioptera roeselii*) and it is inaudible to most people above the age of 40 (although some experts suggest even above the age of 25 (Trilar, 2012)). Males sing from an elevated position, only if the temperature is above 20° Celsius and only in the sunshine. In the New Forest, they

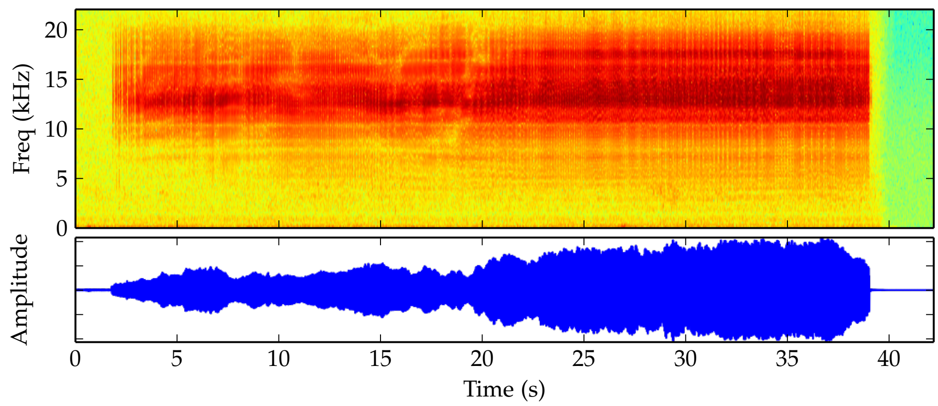


FIGURE 2.6: New Forest cicada call spectrogram (above) and waveform (below), of a recording made in June 1971 by Jim Grant, sourced from *The Wildlife Sounds Collection of the British Library* (Grant, 1971).

normally sing between 10:30am and 6pm, with a peak around 2–2:30pm. If cicadas detect any danger, if the sky is cloudy or in the presence of a cooling breeze, the singing stops immediately. Sometimes, however, they may continue singing while moving from one spot to another. The cicada call can generally be heard, and has traditionally been revealed, with a bat detector tuned to around 16 kHz.

The growing rarity of cicada sightings in the New Forest has generally been attributed to three factors. Firstly, an intensification of grazing policies, which downsized the natural habitat of the cicada and made turrets and underground chambers more likely to be trodden by grazing animals (Pinchen and Ward, 2010). Secondly, a change of weather in the last few decades may have destroyed nymphs, as frequent and abundant rain may have flooded the chambers, causing the nymphs to drown (Daponte, 2004). Thirdly, changes in felling and other forestry practises have reduced the space for scrubby woodland edges where this cicada is most likely to be found.

Some monitoring and research work has been performed in the past few years by the New Forest authorities and their partners and temperature monitoring loggers, which provide crucial information on eggs hatching, have been deployed since 1995. An event called *New Forest BioBlitz* is also held yearly to involve the population in mapping the presence of animals and plant species. However, most of the current knowledge about this insect can be attributed to a handful of entomologists that have collected observations throughout the last century (of particular value is the work of Lyle (1910, 1911, 1913) and Jim Grant between 1963 and 1990). In 2013, the Forestry

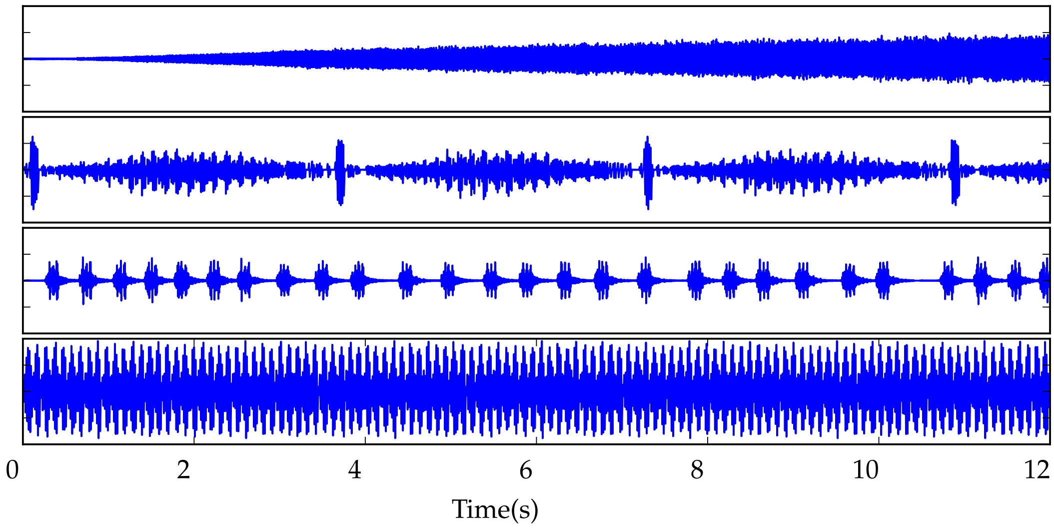


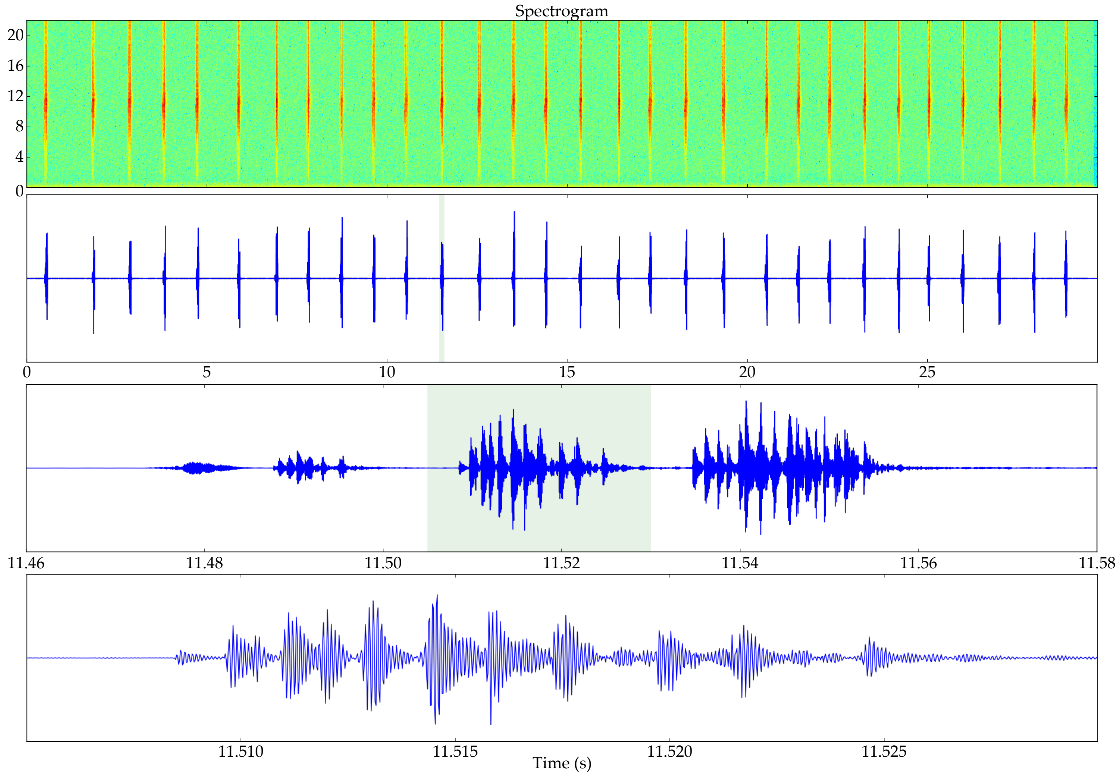
FIGURE 2.7: Waveforms of similar insects. From top to bottom, beginning of *Cicadetta montana* song, common field grasshopper, field cricket and Roesel's bush-cricket. Although at first sight they may look very different, they can be confused when heard. The call has been repeated where necessary to match the length of other samples. Sounds sourced from Jim Grant (Grant, 1971) and www.junglewalk.com

Commission has assigned funding to an insect charity named *BugLife* to perform habitat surveys. This has led to three entomologists extensively looking for the cicada across the summer 2013 in several of the known sites (Henshall, 2013). Nevertheless, the lack of experts monitoring this insect, despite the great interest showed by conservation bodies, the park authorities and the wider community, constitutes a founding motivation for this research.

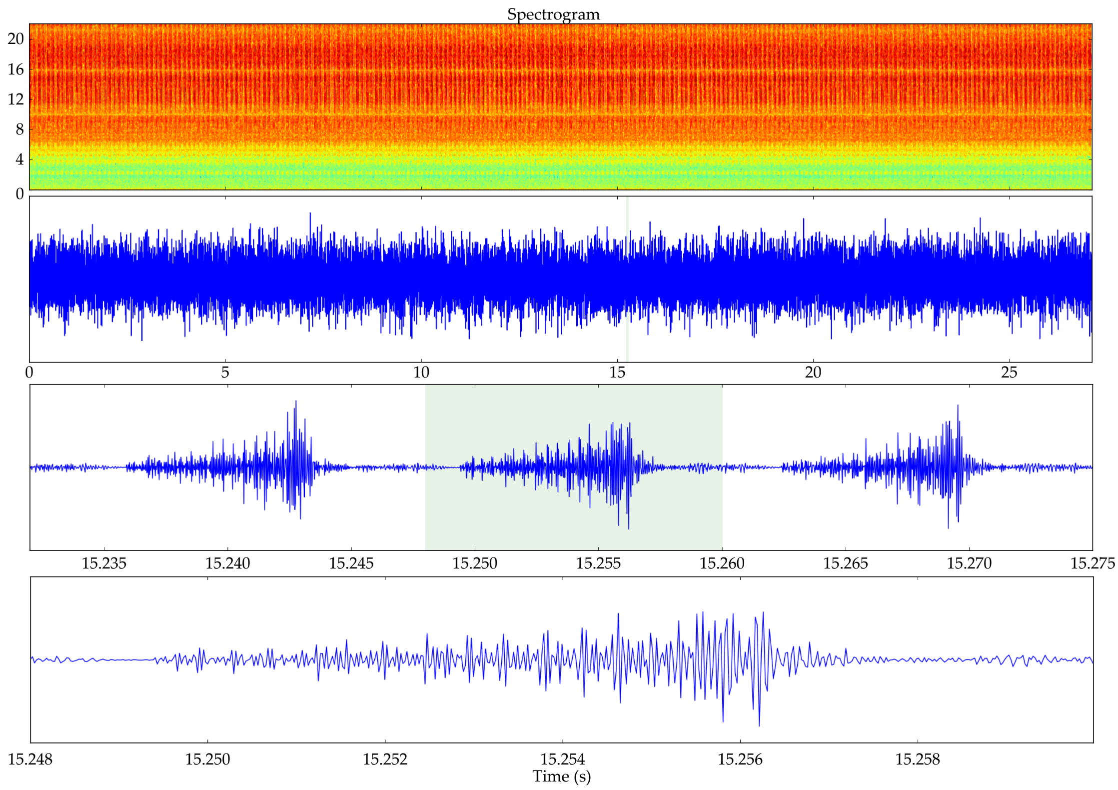
A few other singing insects have been heard in the New Forest, and the knowledge of their call is paramount to this research, as it can be confused for that of the cicada. For this reason, and for the purposes of extending this work to other insect species, described later in Chapter 5, the following section gives an overview of related insect species.

2.4 Crickets, Grasshoppers and Related Insects

The most archetypal, and perhaps best studied order of singing insects is that of the Orthoptera, which includes crickets and grasshoppers. The United Kingdom has 28 native species of these, divided in seven families (Benton, 2012), most of which emit

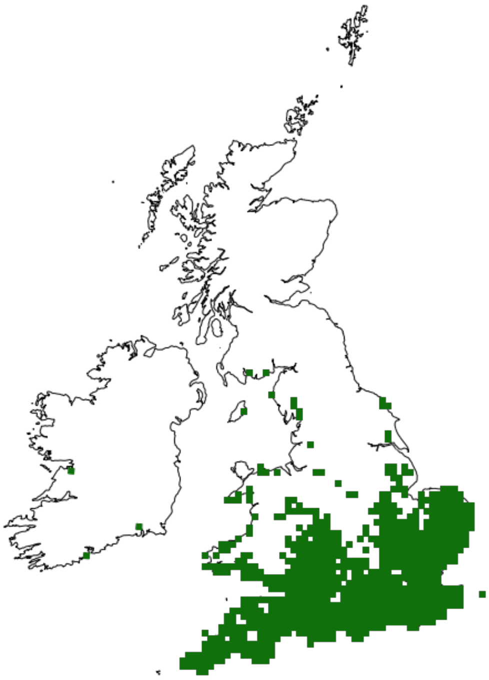


(A) Dark bush-cricket

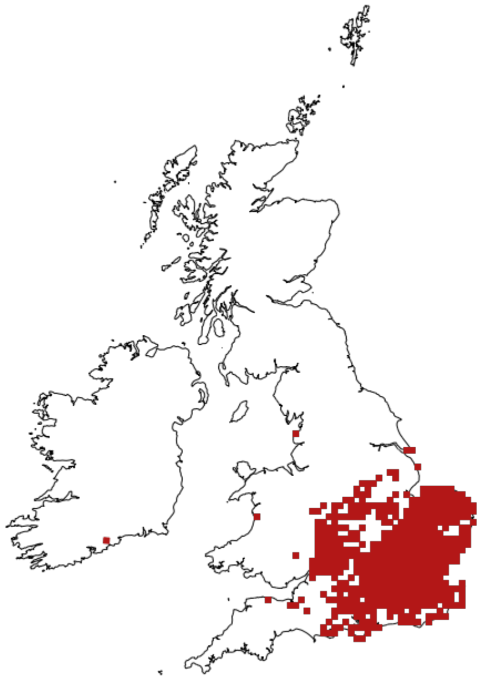


(B) Roesel's bush-cricket

FIGURE 2.8: Spectrogram and waveform of the call of the two bush-crickets, the waveform shown at three different zoom levels. Reproduced from Rogers (2014).



(A) dark bush-cricket



(B) Roesel's bush-cricket

FIGURE 2.9: Distribution of Dark and Roesel's bush-crickets across the entire country. Sourced from orthoptera.org.uk and data.nbn.org.uk

a sound that is useful for identification. An early but complete account is given by Raggae (1965), which includes all species in the country with the exception of the ones of recent discovery. Raggae (1965), and later Baldock (1999), also present a diagram of the calls of all these known species. A more up-to-date and detailed account is given on the 'Orthoptera and allied insects' website (Orthoptera Recording Scheme, 2015), which reports phylogenetic information about the various families and species as well as their conservation status and presence across the country.

Only a small number of these are present in the New Forest, and have a call that can be mistaken for that of the cicada. Among these are the wood cricket, the field grasshoppers, the Roesel's bush-cricket and the dark bush-cricket (Pinchen, 2012). The call of the latter two is particularly similar to the New Forest cicada's call in the frequency domain, although they differ significantly in the time domain. Figure 2.7 displays waveforms for these species, and although from these graphs certain differences appear very clearly, in noisy recordings the characters are less pronounced, and cause difficulty in the classification. However, these insects are active (i.e. in their adulthood and singing) at different times of the day and the year, though with some overlap, and in a real scenario this can help to distinguish from one another. Our data collection in the New Forest has revealed an abundance of dark bush-cricket and Roesel's bush-cricket (their presence in the country is shown in Figure 2.9), and therefore they have become particularly significant to this research.

The dark bush-cricket (*Pholidoptera griseoaptera*) has a dominant frequency around 10 kHz, with a very wide spectrum (Figure 2.8a). Individuals are found singing in proximity to each other, so often recordings contain more than one specimen. They are mainly nocturnal, and mostly heard singing in the evening (Benton, 2012). In the time domain, their call displays a short, $\approx$ 100 ms chirp, which itself is composed of shorter repeating patters. The alternating of adjacent males (Benton, 2012) makes it difficult to exploit the length of the pause in between chirps for the purposes of automated classification.

The call of the Roesel's bush-cricket (*Metrioptera roeselii*) also covers a very wide spectrum, at it can be heard as high-pitched buzz (Benton, 2012). Its prolonged, continuous $\approx$ 10 ms bursts (Figure 2.8b) are so fast that they may appear as a continuous call, and for this reason it is the insect that most resembles the cicada song, a key difference being the constant amplitude (as opposed to the cicada that starts quietly and becomes louder). The dominant frequency is even higher than the cicada's, around 23 kHz, with components still clearly visible around 60 kHz. The Roesel's bush-cricket's adulthood starts only a few weeks after the cicada's, around late June.

Though the most common, it should be noted that not only Orthoptera sing with a loud, distinctive call. The lesser water boatman (*Micronecta scholtzi*), is a peculiar example of a small bug, only 2 mm long, more closely related to the *Cicadetta montana* than crickets and grasshoppers—being part of the same order (Hemiptera)—that emits a mating stridulation of up to 99 dB by rubbing its penis against its abdomen (Sueur et al., 2011). This call, despite being underwater, is so loud that it can be heard from the surrounding environment. Similarly, many other insects produce incidental and non-incidental sounds (see Section 2.2).

2.5 Summary

This chapter has reviewed related work in the two key areas for this research, smartphone-based citizen science and bioacoustics, with a brief overview of some machine learning techniques for wildlife sound classification. In particular, the difference between crowdsourcing and citizen science has been analysed in the context of a number of examples from both domains, and the major issues related to crowd involvement and incentive mechanisms have been highlighted. From the bat monitoring and the automated taxa identification communities, knowledge has been drawn on state-of-the-art methods of call detection and classification.

However, no tool exists today that could aid the citizen scientist to find the New Forest cicada, nor any similar insect. A need for this is expressed by the combination of a growing community of citizen scientists, who have helped solve many similar problems in the past (Solon and Lanxon, 2012), the spreading of smartphone devices across the population, ecological reports requiring intervention for species protection, such as the one on the New Forest cicada (Joint Nature Conservation Committee, 2010) and finally the lack of funding for expert ecologists to carry out manual surveys.

In light of this, the present research identifies the need to address these shortcomings by devising a system to automatically detect and identify animal songs, reporting the findings to a centralised database. Chapter 3 presents the bioacoustic components that have been developed to support this system, and in particular the HMM-based cicada detector algorithm proposed by this research, together with acoustic analysis that has lead to the introduction of this algorithm. A study on the sensitivity of different smartphone's microphones is introduced in conclusion to this work. Chapter 4 then describes the implementation and deployment of the system on a real smartphone app, fully functional and currently in use by hundreds of citizen scientists. Chapter 5 extends on these methods to broaden the classification to all known species of British singing Orthoptera.

Chapter 3

HMM-Based Acoustic Cicada Detector

> Nuit et jour à tout venant
> Je chantais, ne vous déplaise.
>
> La Cigale et la Fourmi,
> traditional French fable

In order to address the shortcomings outlined in Chapter 2, the implementation of a system that would both provide the tools for searching for the New Forest cicada and act as a test bench for other citizen science bioacoustic projects was deemed necessary. The proposed system allows users to navigate around the New Forest and record the presence of the insect, giving immediate feedback about the surrounding environment. To this extent, an automated real-time low-power smartphone-based algorithm is required to classify the cicada call and report back to the user. This chapter presents *a*) the analysis performed to understand the features of the call; *b*) an initial, efficient algorithm based on a simple frequency feature; *c*) a more advanced algorithm based on a hidden Markov model (HMM), exploiting frequency-domain and time-domain features of the call; *d*) a critical evaluation of the proposed algorithm in comparison to a state-of-the-art technique for batch insect classification and an additional evaluation against alternative methods, also informed by the literature. Chapter 4 then presents how this algorithm has been ported to a fully-deployed mobile system.

3.1 Initial Recordings and Tools Used

The development of such a system requires that two conditions are met. In the first place, it must determined whether the sound of the cicada can be detected by a smartphone; if this holds true, the characteristics of the insect's call must be analysed to be able to automate its classification. To this extent, recordings of male *Cicadettae montanae* have been taken by the author in Kranjska Gora, Slovenia (see Figures 3.1 and 3.2), where a substantial population is known to be singing every year.

Sound recordings have been made with a range of different devices. From the top end to the bottom end of the price spectrum: a Fostex FR-2LE field recorder with a Telinga Stereo Dat-Mic microphone on a Pro 8 handle, in a 1mm flexible parabolic dish; a Sony PCM-M10 portable recorder, an iPhone 4S, an HTC One X, a Samsung Galaxy Tab GT-P1000, and two Samsung Galaxy Mini. Table 3.1 outlines the relevant characteristics of these devices, and the apps used for recording on the smartphones. No difference has been noted in the recording capabilities of the different apps, all able to capture uncompressed audio from the microphone, sampled at up to 44,100 Hz (Android) or 48,000 Hz (iPhone). In addition to a sound recording app, the smartphones used an instantaneous spectrum analyser to locate the high-frequency peaks generated by the cicada.

The recordings at two different sites near Kranjska Gora, collected over two days (more than 100 tracks in total), have been analysed for the presence of the cicada. Figure 3.3 shows an example of a good recording, made with the Fostex FR-2LE, where several songs have been captured. The song is visible as a band, centered around 13.5 kHz, lasting for roughly 30 seconds, starting quietly and becoming progressively louder (the warmer colors on the spectrogram) to then interrupt abruptly. The high-intensity sound at the bottom of the spectrum (low frequencies) is background noise, mostly

Device	*OS*	*Sampling*	*App*	*Approx. Cost*
Fostex FR-2LE + Telinga Pro 8	N/A	96,000 Hz, 24bit	N/A	£2000
Sony PCM-M10	N/A	48,000 Hz, 16bit	N/A	£150
iPhone 4S	iOS	48,000 Hz, 16bit	Recorder Pro	£450
HTC One X	Android 4.0	44,100 Hz, 16bit	Hertz, Tape-a-Talk	£500
Samsung Galaxy Tab GT-P1000	Android 2.3	44,100 Hz, 16bit	Hertz, Tape-a-Talk	£400
Samsung Galaxy Mini	Android 2.2	44,100 Hz, 16bit	Hertz, Tape-a-Talk	£100

TABLE 3.1: List of devices used for recording and relative *apps* and settings, where applicable. The approximate cost is accurate to the date of purchase (2012).

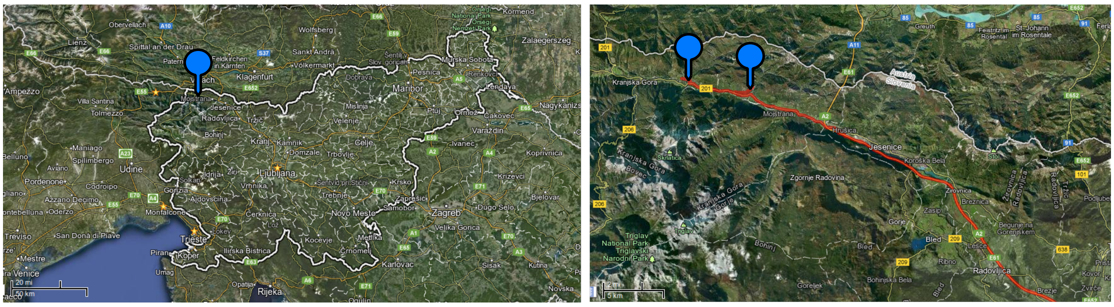


FIGURE 3.1: Visited sites in Kranjska Gora. On the left, the map of Slovenia. On the right, the two sites visited.

FIGURE 3.2: Typical habitat of *Cicadetta montana* in Kranjska Gora. In the New Forest, this habitat is slightly different, tending more towards open deciduous woodland. On the left, the Fostex recorder with the Telinga microphone in action; on the right, Faberacoustical's SignalScope (http://www.faberacoustical.com/ios_apps/signalscope/) showing an FFT of the microphone's input, with no sign of a cicada singing.

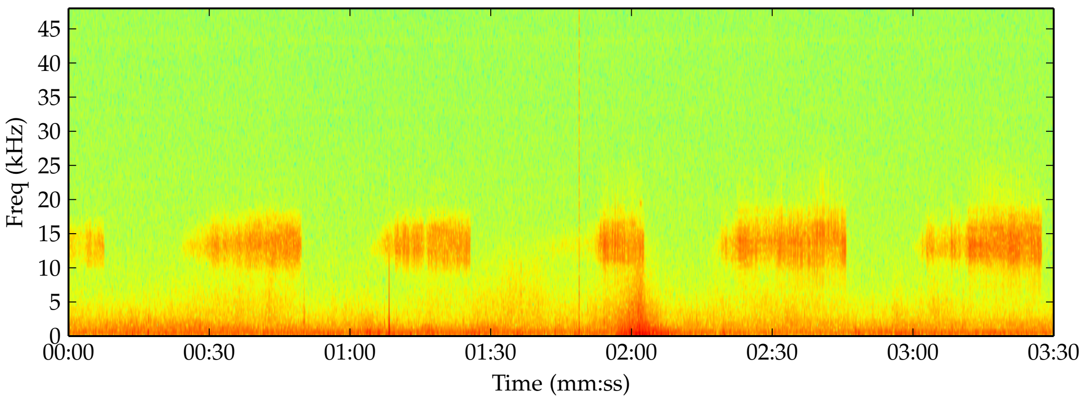


FIGURE 3.3: Spectrogram of a recording in Kranjska Gora, Slovenia, taken with the Fostex FR-2LE at 96 kHz. At least five calls are clearly visible, although one is interrupted by a vehicle passing by (the low-frequency band around 02:00).

represented by wind noise or nearby road traffic, and occasionally people speaking. It should be noticed how in moments where a high background noise was present (e.g., in Figure 3.3, the car passing by around 02:00m into the recording) the call becomes more difficult to hear, but it remains clearly visible on the spectrogram.

Beside providing a plethora of song recordings, the expedition highlighted different aspects of the problem. Firstly, the different recording capabilities of the devices tested. These are described in more detail in Section 4.4. Secondly, the difficulty at performing automated detection when more than one male is singing, as many of the features of the call are lost if two similar sounds are overlapping. However, experts say that in the case of the *Cicadetta montana* this is an unlikely possibility, as populations are small and rarely two males sing together (Trilar, 2012). Thirdly, the fact that a directional microphone, such as the Telinga used in the experiment, is excellent for a good quality recording and if one knows where the cicada could be, but it is not equally good for reconnoitring as the directionality impedes a wide-range search. These considerations further motivated this research and have been taken into account during the development of the tools.

3.2 Sound Analysis of Existing Recordings

The first step towards an automated detection of the cicada call is the analysis of its features. To this extent a high-quality recording of *Cicadetta montana*, provided by Dr Tomi Trilar and Prof Matija Gogala from the Slovenian Museum of Natural History, as well as a sound file from the wildlife recordings archive at the British Library, have been studied to discover key features. While the former constitutes a recording of *Cicadetta montana* of excellent quality in a different country, the latter is the only available recording of the cicada in the New Forest (to the best of the author's knowledge), dating back to 1971. The waveform and spectrogram of the call are shown in Figure 3.4. Unless stated otherwise, the characteristics here reported are found to apply to most song recordings.

In the frequency domain the most notable trait is the prevalence of a component between 12 and 17 kHz, particularly strong around 13–14 kHz. An FFT of the signal confirms this observation across the entire sample (Figure 3.5).

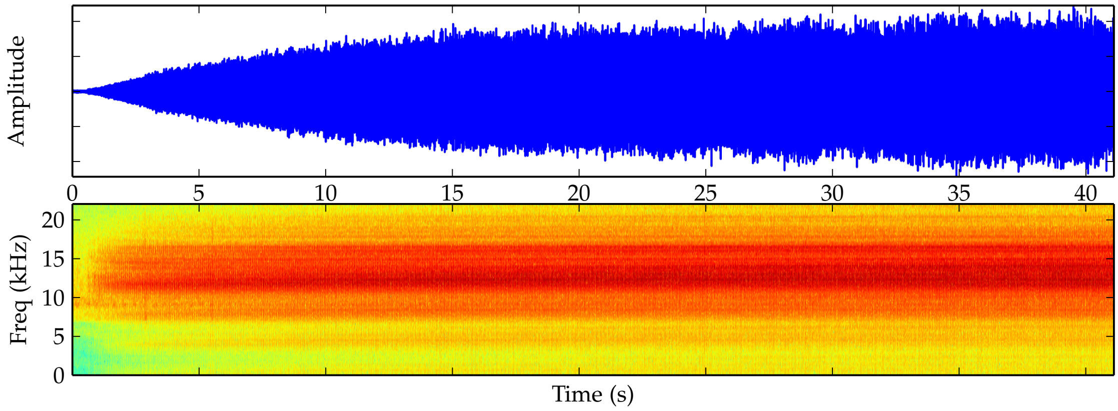


FIGURE 3.4: Waveform and Spectrogram of a high quality recording. In the top-right corner, a detail of the waveform shows an 8 ms repeating pattern (the size of the detail does not match the size of the box in the expanded waveform).

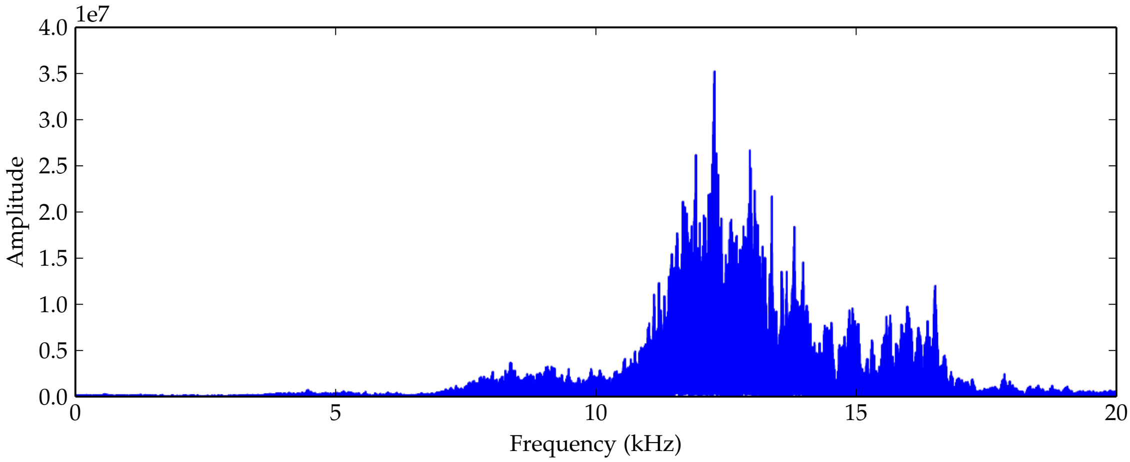


FIGURE 3.5: FFT across the entire 40 seconds sample.

In the time domain the most prominent behaviour is the increasing intensity of the call, which starts low (and is difficult to notice), to then become increasingly loud before stopping abruptly. This behaviour could be classified using a probabilistic model, whereby the sudden interruption of the sound can be a strong indicator of the presence, in the previous moment, of the call. This can be reversed using a Hidden Markov Model, as the information at a given time-step is correlated to the previous instant. The use of a Probabilistic Graphical Model (PGM) to classify the sound may be considered as future work (see Chapter 6).

Another feature is the presence of a repeating 7–8 ms amplitude modulation pattern. To extract the modulating waveform, the signal has been rectified and passed through a low-pass filter which retains this slower component of the signal (roughly 130 Hz). A better result is, however, achieved with a standard amplitude demodulation technique,

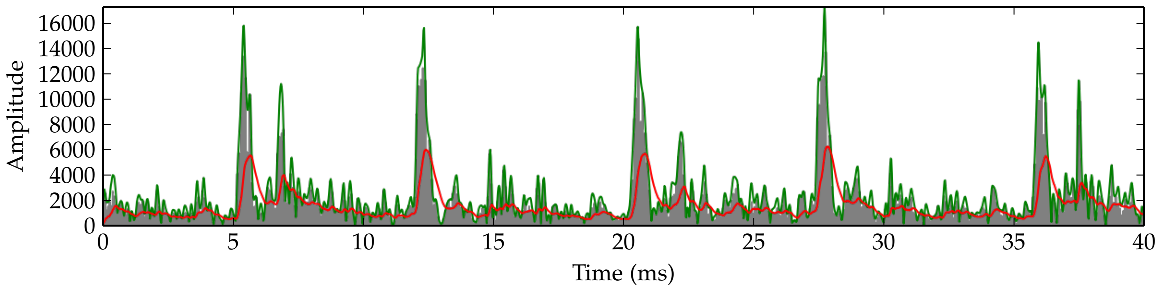


Figure 3.6: Comparison between a low-pass Butterworth filter (red) and a Hilbert transform (green) plotted on the positive side of the original sound, in a 40 ms window ($\approx$ 1800 samples)

envelope detection, obtained with a Hilbert transform. The comparison of the two is presented in Figure 3.6, where the green line represents the envelope computed by the Hilbert transform, while the red one the low-pass filter (in this case, a Butterworth filter), applied on a 40 ms window.

Running an FFT on the envelope shows the presence of a strong 130 Hz component, with several harmonics, as expected (Figure 3.7). Moreover, a 65 Hz component appears quite prominently, which may indicate the fact that there is also a 16 ms repetition, in which the two 8 ms windows are slightly different. This is further confirmed by a subsequent test, in which a sliding window of samples traversed the envelope of the signal looking for similar sets of samples. The result, of which a particular is shown in Figure 3.8, exhibits the two repeating patterns mentioned above (here lower score means closer matching to the sampling window).

However, an analysis of different sound samples demonstrates that this behaviour only manifest itself occasionally, probably in the highest quality recordings. This feature is therefore not to be relied upon and this needs to be considered especially

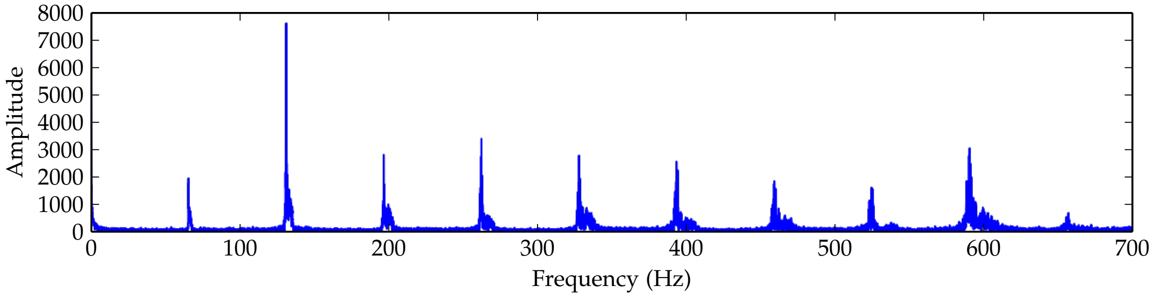


Figure 3.7: FFT of the envelope of the sound. The highest peak is on 0 Hz, the second on $\approx$ 130 Hz, with relative harmonics, and the third one at $\approx$ 65 Hz, with relative harmonics, showing the two 8 and 16 ms patterns.

when recording with mobile phones, where the response of the microphone may not be optimal.

The carrier wave can be obtained by dividing the signal by its envelope, as shown in blue in Figure 3.9, where the original signal in represented in black and its enevelope in red. Provided that this carrier is sufficiently clean, a simple method to calculate its frequency is counting the zero crossings. This has been performed in windows of 250 ms, and represented in Figure 3.10, which shows how the carrier wave increases to reach an equilibrium around 13.5 kHz, though still gradually incrementing until it stops.

3.3 Frequency-based Classifier

With the results of the analyses considered, a first classifier has been built, based purely on the frequency domain. This in fact exhibits clear separation between the cicada call and background noise, with a clean 13.5 kHz-centred peak, consistent across all available recordings. On the contrary, the time domain exhibits a clear character only in certain recordings—consider for example a recording where the microphone has been moved closer and further away from the singing cicada; the amplitude will vary and the feature will be lost (for an example, refer back to Grant's recording in Figure 2.6).

The algorithm therefore takes the signal, divides it into one second overlapping rectangular windows and calculates an FFT of the window. Within that spectrum, it computes the ratio between the sum of the frequency components from 11 to 17 kHz and from 8 to 9 kHz. This is based on the visual intuition that the frequency spectrum rarely shows any strong component above 8 kHz, while much of the background noise lies at lower frequencies (e.g. wind-generated noise, human voice, road traffic, etc.). However, between 11 and 17 kHz a high intensity noise is present during the cicada call, and therefore the difference between this and the 8–9 kHz range differs significantly in the presence of a cicada. The model can be expressed as:

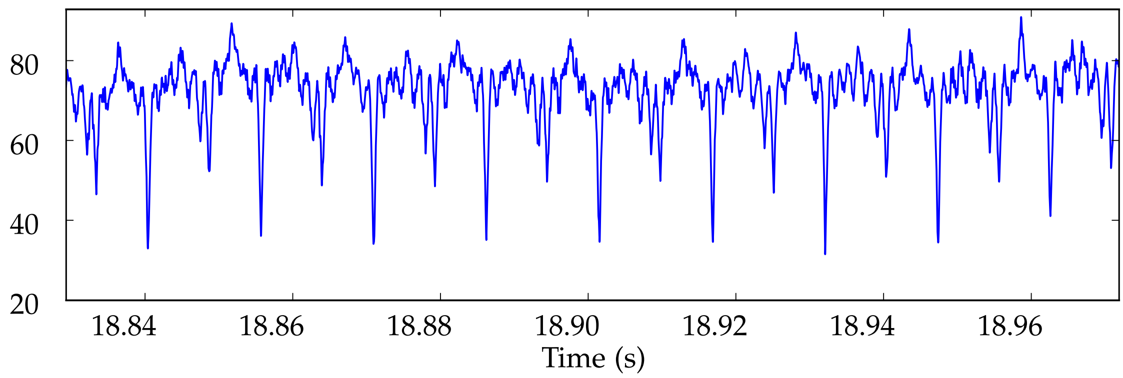


FIGURE 3.8: Pattern matching using recursive traversing of the sound sample. A 140 ms-long detail is shown here, where two repeating patterns are visible. Lower values mean closer matching.

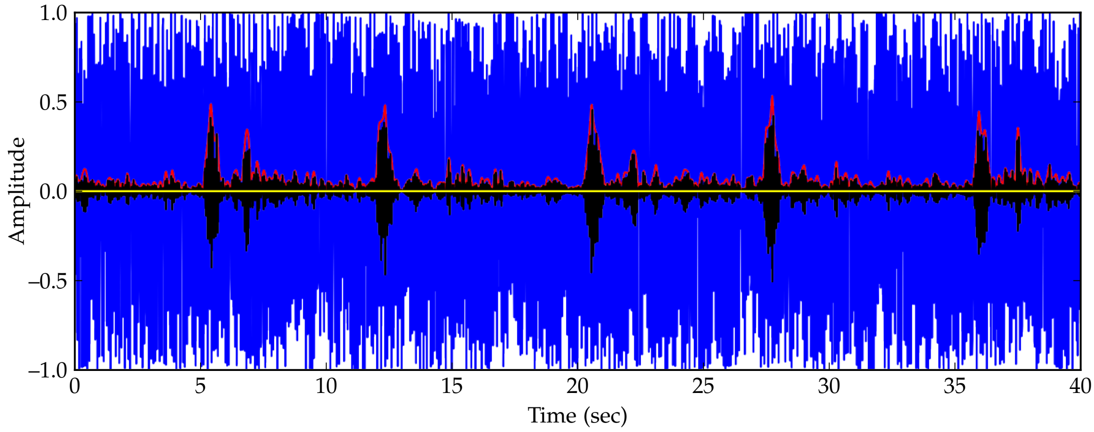


FIGURE 3.9: Carrier wave of the signal (in blue), obtained dividing the original signal (in black) by its Hilbert envelope (in red).

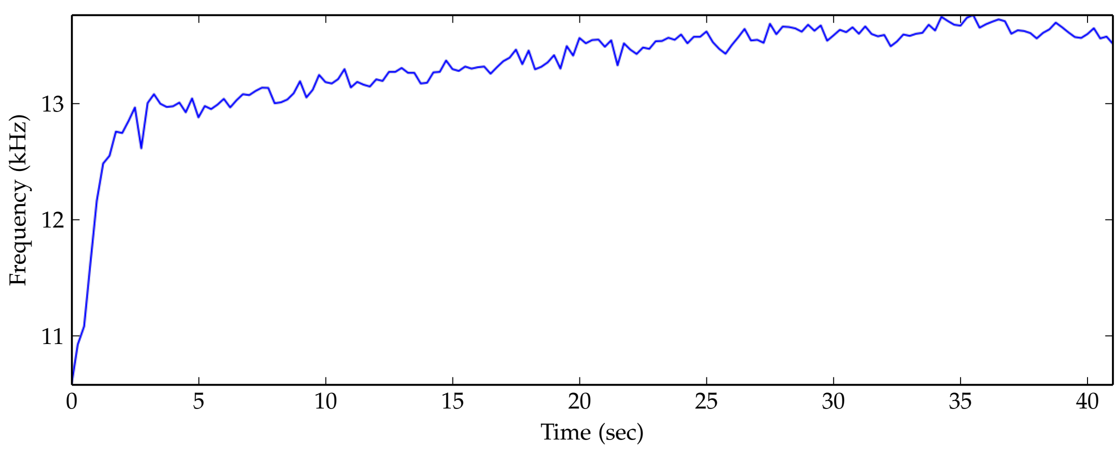
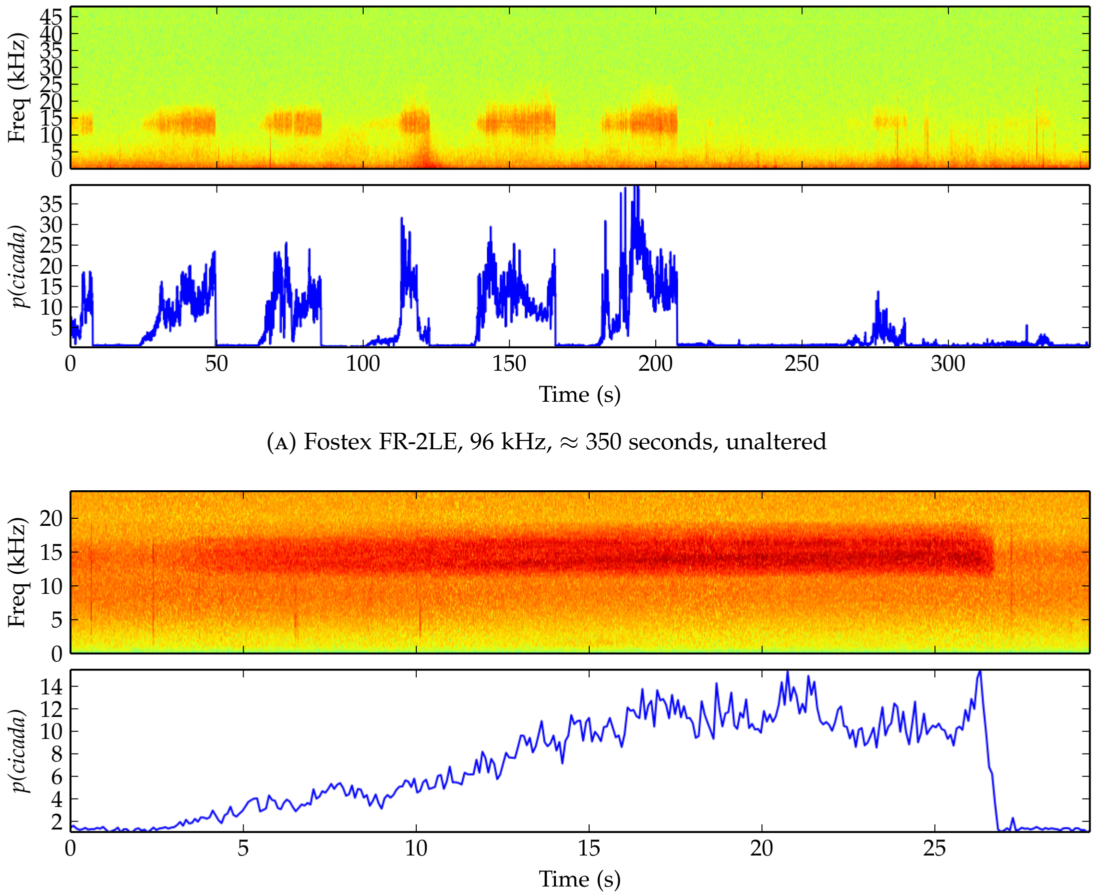


FIGURE 3.10: Frequency of the carrier wave obtained by counting zero crossings in 250 ms windows.

$$p(c) = \frac{\sum_{i=11e^3}^{17e^3} \Omega(i)}{\sum_{i=8e^3}^{9e^3} \Omega(i)}$$

where $p(c)$ is the likelihood, at each window, of a cicada being singing and $\Omega(i)$ the amplitude of the i^{th} frequency component of the spectrum. The value of $p(c)$ is therefore a measure of the acoustic energy in the range 11 to 17 kHz compared to 8 to 9 kHz. This can be normalised across samples to provide a consistent measurement.

The same calculation can be performed by extracting frequency bins with a Goertzel filter rather than an FFT. This is more efficient (for a sufficiently small number of bins, in this case 2), and does not require the signal to be divided into windows. The output of the filter can in fact be updated with each new sample (see Section 2.2.1.5). This



(A) Fostex FR-2LE, 96 kHz, ≈ 350 seconds, unaltered

(B) iPhone 4S, 48 kHz, ≈ 29 seconds, filtered

FIGURE 3.11: Output of the classifier for two different recordings. On top, one with several cicada songs; at the bottom one with one song only. The latter is high-pass filtered at 12 kHz and amplified.

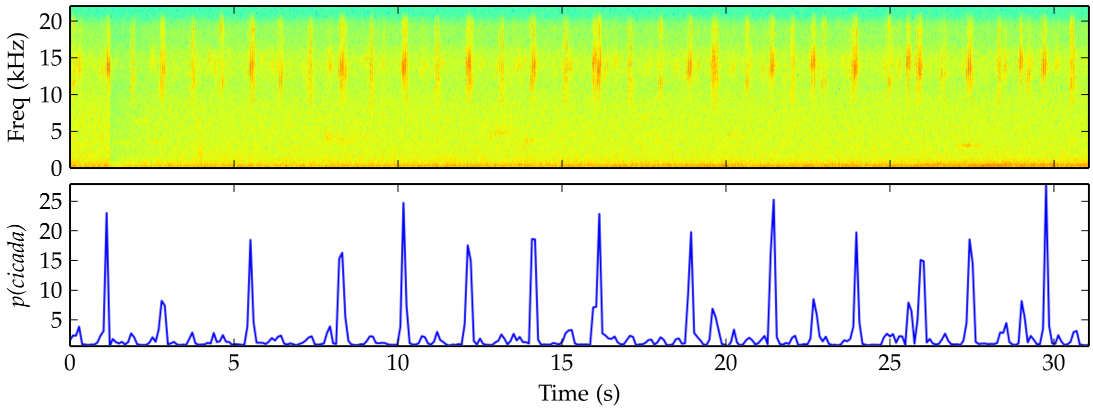


FIGURE 3.12: A dark bush-cricket triggers the frequency-based classifier, appearing like a short burst of cicada call.

method constitutes the foundations of the feature extraction process for the HMM-based classifier presented in the following section.

The output of this simple classifier, represented in Figure 3.11, shows good performance both on a high-quality sample (Figure 3.11a) and on a medium-quality smartphone sample (Figure 3.11b), and is robust to different types of noise, such as human voice, road traffic and similar low-frequency sounds. However, it is not robust to other insects' calls, as exemplified by the output of the algorithm run on a dark bush-cricket's recording (see Figure 3.12).

To address this issue, the algorithm can be extended to consider the notion of time, so that other insects would be rejected as the combination of frequency and time is likely to produce a distinctive signature of the insect. In order to implement this, a variable may count the amount of continuous samples of this features that have been observed thus far, and relate those to the length of each insect's call. While being very tractable and computationally efficient, this method is not very robust and does not scale well to multiple insects, which may have different combination of call's duration and frequency. Therefore, a robust classifier that combines all these features in a structured model is proposed and described below.

3.4 Real-Time Insect Detection Using Hidden Markov Models

To address the robustness and scalability limitations of the previous method, while still maintaining computational efficiency and maximising the detection accuracy, a novel method is presented in two stages. First, an efficient extraction of individual terms of a DFT from the raw audio recordings using the Goertzel algorithm, the implementation of which has already been presented in Section 2.2.1.5, and the combination of two or more of these terms to produce continuous feature vectors that are robust to noise. Then, the classification of insects at each sample of these vectors using a multi-state HMM.

3.4.1 Feature Extraction Using Goertzel Algorithm and Filter Ratio

As previously noted, the magnitude of the frequency component at 14 kHz is a good indicator of the presence of a New Forest cicada, robust against generic background noise, which is normally contained in the lower 5 kHz of the frequency spectrum. However, it may be sensitive to white noise that covers the entire frequency spectrum, such as handling noise. Therefore, in order to reduce this sensitivity, the magnitude of this feature is divided by the magnitude observed around 8 kHz. This band is outside the range of both the cicada call and environmental noise. Hence, this ratio will be high in the presence of a cicada and tend to zero when either no sound is detected in the cicada range or if sound is present across both bands. However, it will not be able to discriminate between the calls of the New Forest cicada and the Roesel's bush-cricket, both of which exhibit a prolonged call at a similar frequency. Therefore, an additional 19 kHz band is extracted, holding a block size $N = 128$ samples, which leads to a bandwidth of just under 1.4 kHz. Hence, the three frequency bands are as follows: $m_{t,8}$ which represents the 8 kHz frequency which is outside the range of both the cicada call and environmental noise, $m_{t,14}$ which represents the 14 kHz frequency of both the New Forest cicada and the dark bush-cricket, and $m_{t,19}$ which represents the 19 kHz frequency of only the dark bush-cricket and the Roesel's bush-cricket. We then take ratios of these frequencies to produce two features:

$$x_{t,1} = \frac{m_{t,14}}{m_{t,8}},\ x_{t,2} = \frac{m_{t,19}}{m_{t,14}} \tag{3.1}$$

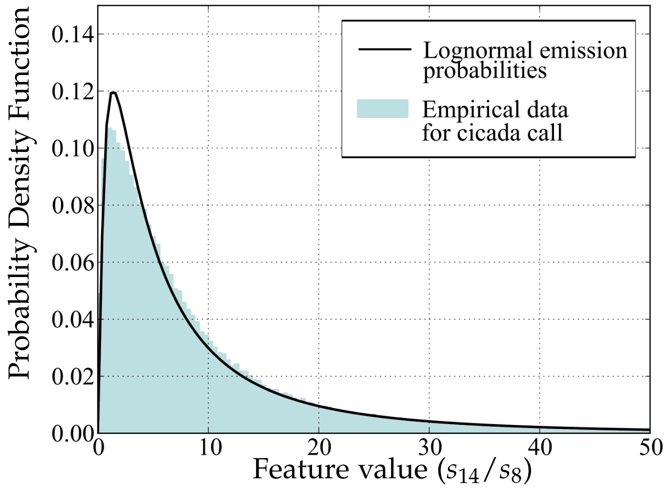


FIGURE 3.13: Log-normal distribution of the extracted feature for the cicada call

As such, at any point t, $x_{t,1}$ will be high in the presence of any of the insects considered and tend to one when either no sound is detected in the cicada range or if sound is present across both bands. In addition, $x_{t,2}$ will be high in the presence of the dark bush-cricket, and tend to zero in the presence of the New Forest cicada. These two features form a T-by-2 feature vector which is used for classification.

It is worth noting that the difference between these two features has also been considered as opposed to the ratio, as well as the individual frequency bins fed independently to the classifier, as reported later in the evaluation of this method. The ratio, however, has the benefit of acting as a *normaliser* for the amplitude of the two bands, providing a feature that is less dependent on the recording device.

With this, in order to obtain real-time computationally efficient insect identification, an HMM-based approach to classification is adopted.

3.4.2 Distribution of the Features

Figure 3.13 shows a histogram of data generated by a cicada's song, along with a log-normal distribution fitted to the data. A log-likelihood ratio test on a normal, log-normal and exponential distributions fitted to a data set of cicada songs shows that the log-normal distribution matches the data better than the normal ($F = 3512.13, p < 0.001$) and exponential ($F = 1516.06, p < 0.001$) distributions. However, despite its long tail, the log-normal distribution still has poor support for data of unusually high magnitude, as are often generated by handling noise. In order to prevent the model from strongly favouring a certain state when a data point is in the extreme of the log-normal distribution, the emission probabilities are capped to capture cases where the features are likely to be poorly represented by this model. The outcome of this is that the likelihood that such data points result from the correct state may be so low that the model triggers a state change even though the transition probability strongly discourages it (by being itself very low). Therefore, the emission probability of such data points are capped such that there is a maximum ratio, initially 100, with which any state can be preferred to another.

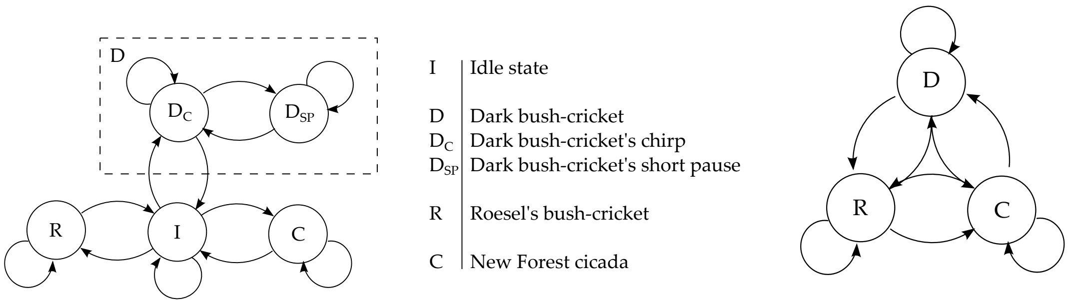
3.4.3 Multi-State Finite State Model of Insect Call

Therefore, a five-state HMM for cicada detection—hereon referred to as cicada detection algorithm (CDA)—is proposed, in which the states consist of: an *idle* state in which no insect is singing (I), a *cicada* singing state (C), a state where the *dark bush-cricket* is *chirping* (D_C), a *short pause* in between the dark bush-cricket's chirps (D_{SP}) and a state in which the *Roesel's bush-cricket* is calling (R). The emission parameters, i.e. the location a and scale b of the log-normal distribution, are learned empirically using:

$$a = \ln\left(\frac{\mu^2}{\sqrt{\sigma^2 + \mu^2}}\right), \; b = \sqrt{\ln\left(1 + \frac{\sigma^2}{\mu^2}\right)} \tag{3.2}$$

where μ represents the mean and σ^2 represents the variance of the data. This manual estimation was originally based on the few recordings the authors had gathered from historical archives, and has therefore been improved with recordings obtained by the deployment of this work, described in the following chapter.

The transition matrices describing the dynamics of a Markovian process can be represented graphically using finite state machines. Figure 3.14a shows the five states described above and all possible transitions, where those with non-zero probability are represented by arrows connecting two states. The model explicitly represents the silence between the dark bush-cricket's chirps, which is essential information for distinguishing between the calls of the New Forest cicada and dark bush-cricket. This is in contrast to existing batch classification methods which remove such silent periods of a recording in order to improve the computational cost of the operation and classify



(A) 5-state model used by the proposed approach.

(B) 3-state model.

Figure 3.14: Comparison of finite state machines

only sounded periods of the sample file (Chaves et al., 2012). These methods also employ a feature extraction process whereby they compute a number of mel-frequency cepstral coefficients for each species in the model, making the process scalable to several insects, at the cost of higher computational complexity. In contrast, this method is more closely tailored to the requirements of the present scenario, producing the improvement in efficiency necessary for a mobile application. Figure 3.14b shows a variant of the approach where the silent states have been removed, against which the model here proposed is evaluated in the following section. Furthermore, the HMM could be arranged so as to be fully-connected, allowing transitions between states that are otherwise disconnected (for example between a Roesel's Bush-cricket and a Dark bush-cricket). However, this confuses the model between states that have very similar emission probabilities, without providing any improvement in accuracy. Hence this variation has been excluded from the comparison in the following section. The entire classifier is summarised in Figure 3.15.

The Viterbi algorithm (Section 2.2.3.1) is used to infer the most likely sequence of hidden states given the features described. Despite the fact that the number of possible paths grows exponentially with the length of the chain, this algorithm efficiently finds the most probable sequence with a cost that grows only linearly with the length of the chain.

3.5 Evaluation of the algorithm Using Smartphone Recordings

The proposed approach is evaluated in three different ways. First, it is compared against a state-of-the-art approach, replicated from Chaves et al. (2012). Secondly, individual components of this method and other practises informed by the literature are used in turn to test if they improve the performance of the proposed algorithm. These two comparisons are presented in this section. Thirdly, the accuracy of this model is compared with a more generic insect recognition system, modelled on the bird classification algorithm presented in Stowell and Plumbley (2014). The latter is presented and assessed in Chapter 5.

3.5.1 Evaluation against the State of the Art

The benchmark system proposed by Chaves et al. (2012) works in the following way. The signal is firstly stripped of un-sounded areas and segmented to extract individual calls. It is then pre-processed by removing the DC offset, dividing it into frames, emphasising high frequencies, and passing it through a windowing function. The windows are then run through a FFT and converted into the mel frequency scale, from which the MFCCs are generated. The implementation of this process, replicated from the paper here considered, is summarised in Figure 3.16, where the feature extraction process is run on a recording with several dark bush-cricket's calls and three cicada calls. The input signal has already been stripped of unsounded periods. The output of the process are the MFCCs shown in the last plot of the figure. These are used as

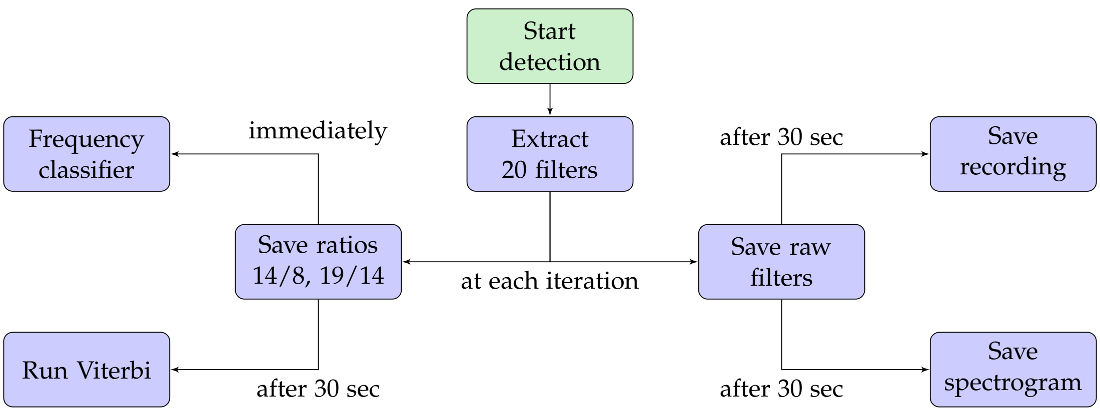


FIGURE 3.15: Detector flow, as implemented in the app. Saving the spectrogram and a sound recording is desirable as later discussed in Chapter 4, at the cost of saving all twenty filters instead of just three.

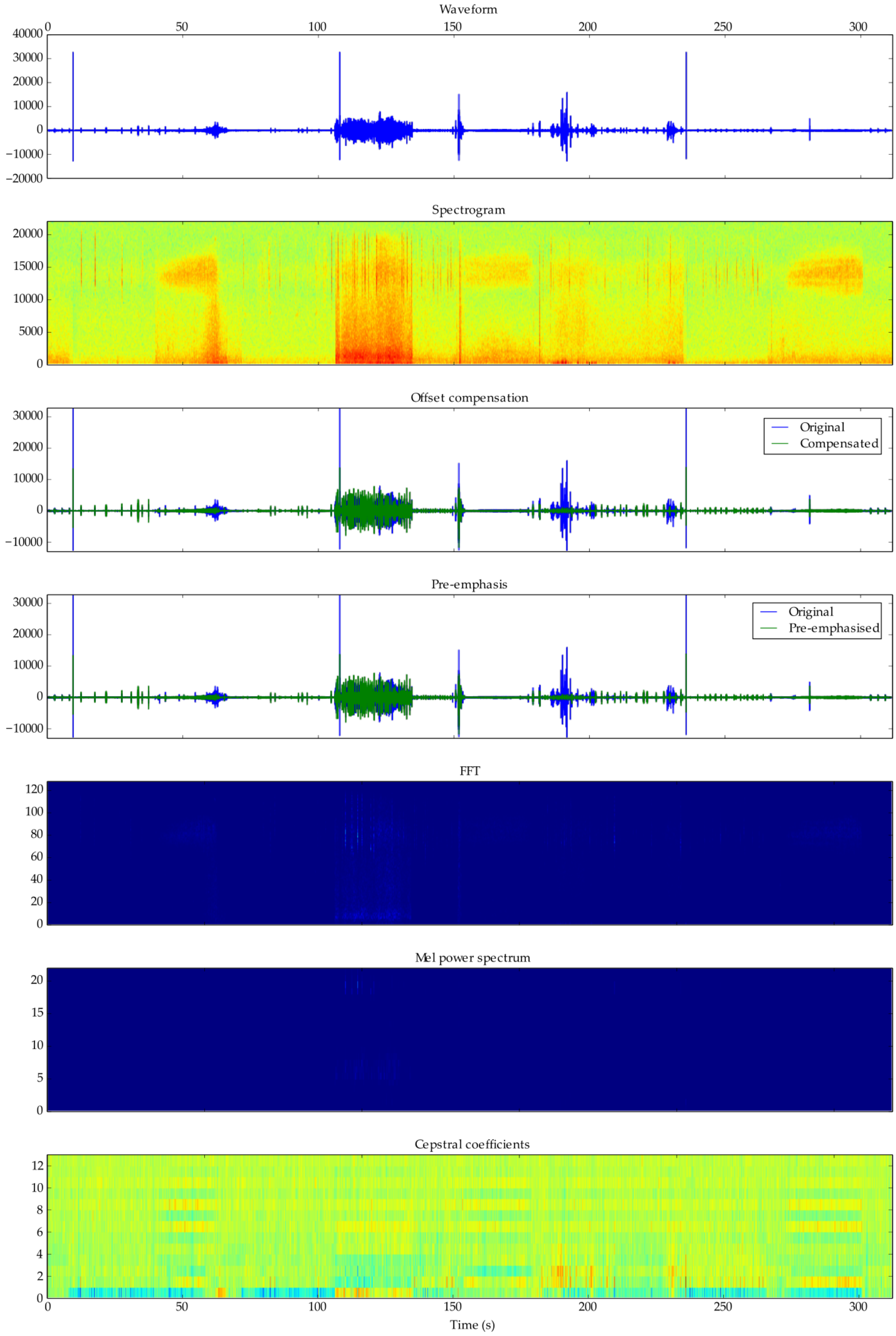


FIGURE 3.16: Implementation of the feature extraction process described by Chaves et al. (2012) on a recording that includes several dark bush-cricket's calls and three cicada calls. The figure does not include the pre-processing stage, so the calls have already been stripped of unsounded periods. The steps are labelled on top of each plot.

individual features for a simple HMM. For the recording in analysis, this consists of two states, one for the cicada and one for the bush-cricket, with a feature vector of 24 cepstral coefficients, each assumed to be normally distributed. No state for silence is considered, as this has been removed during the pre-processing stage.

To evaluate the accuracy of the two approaches, recordings of the New Forest cicada from the known habitat in Slovenia and the dark bush-cricket from the New Forest were collected using an Apple iPhone 4S. In contrast to existing recording libraries, this data set represents the quality of crowdsourced data that the system is likely to encounter, exhibiting significant noise and insect calls of varying amplitude depending on the proximity of the recording device to the specimen. Since this evaluation compares recordings at each time step (as opposed to classifying an entire recording as one insect), for the sake of clarity no Roesel's bush-cricket is considered in this instance, limiting the model to two insects and four states.

Figure 3.17 shows a comparison of the two approaches using a concatenation of three cicada calls and several instances of the dark bush-cricket call intertwined. Figure 3.17a shows a spectrogram with the time domain on the x-axis, and the frequency domain on the y-axis, with the magnitude of the frequency bins varying with the colour of the plot. The three cicada calls can be identified as the prolonged strong component in the high frequency band. The chirping calls are visible as thin vertical bars on the top half of the spectrum. Note that the different recordings, merged together into this data set, have varying background noise, identifiable particularly as high magnitude components at the bottom of the spectrum. Figure 3.17b shows the ground truth, labelled manually, i.e. the correct classification of the different insects. The states are labelled as in Figure 3.14a, where I represents the un-sounded idle state, C represents the cicada's song and D represents both the dark bush-cricket's chirping and short pause states. Figure 3.17c shows the output of the model from Chaves et al. (2012). For this approach, areas identified as idle have been removed from the feature by the pre-processing stage, but have been reintroduced in the output for the sake of comparison. On the plot they are marked as idle, although the model itself does not account for an idle state. Since the comparison is focused on the discernment of the two insects rather than the detection of sounded and un-sounded areas, the sounded and un-sounded areas were manually labelled. Finally, Figure 3.17d shows the output of the model proposed in this thesis. The two states used to identify the dark bush-cricket's call are merged into one, again as represented in Figure 3.14a. It is immediately apparent how closely the proposed approach matches the ground truth in comparison to Chaves et al. (2012).

From this it can be concluded that removing silence between calls also removes the time-domain features crucial at discerning these two insects. The output of the HMM in Figure 3.17c displays confusion between the chirping call and the prolonged call and is unable to identify them correctly. The visual intuition is confirmed by the accuracy measures described below and reported in Table 3.2. On the contrary, the proposed model is able to take advantage of the clear time-domain feature and, despite the emission probabilities of the two sounded and the two un-sounded states being identical, the transition probabilities ensure that prolonged periods of silence are classified as the idle state. To this extent, the backward pass of the Viterbi algorithm ensures that any mistakes due to a state having the highest local probability are corrected to provide the most likely overall path. Furthermore, this approach can be readily extended to calls of more complexity by further increasing the number of sub-states attributed to each insect.

The accuracy by which each approach can correctly classify the cicada is assessed using the standard precision, recall and F_1 score metrics. The precision represents the fraction of time slices in which the approach detected the cicada as singing when it was in fact singing, while the recall represents the fraction of time slices in which the

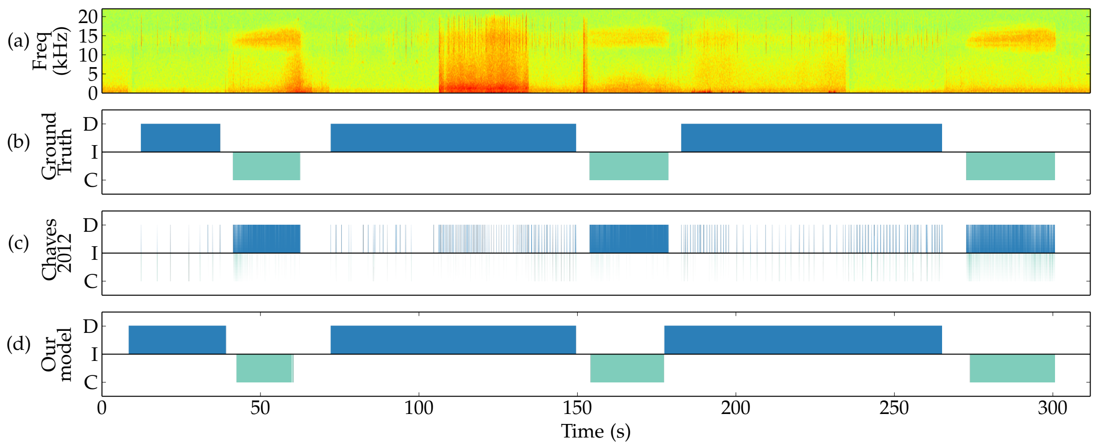


FIGURE 3.17: The proposed model, run on a recording with several dark bush-cricket's calls and three cicada songs. *I*, *C* and *D* represent the idle, cicada and dark bush-cricket states respectively, as in Figure 3.14a. *D* encompasses both the dark bush-cricket's chirping (D_C) and short pause (D_{SP}) states.

Approach	Precision	Recall	F_1-score
Proposed approach	1.000	0.914	0.955
Chaves et al. (2012)	0.563	0.071	0.126

TABLE 3.2: Accuracy metrics of cicada detection

cicada was singing that were correctly detected. Precision and recall are defined as:

$$precision = \frac{tp}{tp + fp} \tag{3.3}$$

$$recall = \frac{tp}{tp + fn} \tag{3.4}$$

where tp represents the number of correct cicada song detections (true positives), fp represents the number of cicada song detections when it was actually not singing (false positives), and fn represents the number of cicada songs which were not detected (false negatives). At this stage, this work is not concerned by the accuracy of the cricket's detection. We also use the F_1 score, which represents a weighted combination of precision and recall, defined as:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{3.5}$$

Table 3.2 shows the precision, recall and F_1 score metrics both for the approach described here and that used by Chaves *et al.* 2012 over a much larger data set of over 30 different cicada songs. It is clear that the approach proposed by Chaves et al. (2012) fails to distinguish between the cicada's song and the bush-cricket's chirp, resulting in poor precision and recall statistics. Conversely, both the precision and recall metrics for the proposed approach are close to 1, as a result of the model's ability to use the periods between the bush-cricket's chirps to differentiate between the two songs. Furthermore, the vastly greater precision and recall metrics for this approach have resulted in a greater F_1 score. This can be interpreted as a suitable trade off between false detections and missed detections.

It is also worth comparing the computational efficiency of the approach used by Chaves et al. (2012) to the approach described here. In the Chaves et al. (2012) model, the two most costly operations, namely the sound detection algorithm and the computation of the cepstral coefficients, both require an order $O(NlogN)$ to compute, with N being the number of samples in the recording. In comparison, the entire feature extraction process in the proposed model only requires $O(N)$ operations. This complexity corresponds to a computation time of 537s for the Chaves et al. (2012) approach, while the present approach takes 45s to process the recording of length 311s, shown in Figure 3.17. Since the Chaves et al. (2012) method takes longer to run than the length of the recording, clearly it is not efficient enough to run in real time. In comparison, the present approach processed the whole recording in one seventh of the recording time, and therefore is suitable to run in real time. These values, although dependent on implementation details, corroborate the hypothesis that the former model has a considerably higher computational complexity, as shown in Section 3.4. This, together with the increased robustness to noise shown by the accuracy metrics, allows us to conclude that the proposed model is better suited to real-time detection than the state of the art for insect classification. The execution times of both approaches were evaluated on a mid-range modern computer (Intel Core 2 Duo CPU, 2.4 GHz, 8 GB RAM), with the software entirely written in Python. This evaluation has been published in the proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI) 2013 (Zilli et al., 2013).

3.5.2 Evaluation against Variants

Further to the comparison above, this section introduces three variants of the approach described thus far that, selecting components and practices informed by the literature, may improve the cicada detection algorithm. For this test, the recordings of Roesel's bush-cricket calls were also used so as to match the requirement to recognise this insect, observed after the first season of deployment of the Cicada Hunt app (see Chapter 4)

The three variants are as follows. The first one uses the three raw frequencies described above (8, 14 and 19 kHz), as opposed to their ratio, directly as features (CDA raw frequencies). The second variant removes un-sounded periods from the recording and, as such, segments it into individual calls. It then applies the 3-state model shown in Figure 3.14b to classify the insects (CDA silence removed). The third approach does not apply a HMM at all, and instead uses the ratio of frequencies to directly identify

Approach	Precision	Recall	F_1-score
CDA	0.66	0.78	0.82
CDA raw frequencies	0.46	0.94	0.62
CDA silence removed	0.62	0.99	0.75
Mixture model	0.61	0.65	0.67

Table 3.3: Accuracy metrics of cicada detection

the most likely state, given only the instantaneous emission probabilities of the features. As such, this method can be considered as a mixture model, since each time slice is classified independently. This method is considerably more computationally efficient, at the cost of losing the information of the time domain.

The accuracy of each approach is evaluated using a collection of 235 recordings taken by citizen scientists using smartphones from the New Forest and by the authors of this paper in Slovenia over the summer of 2013. Each recording is 30 seconds long, and in most cases contains a call of either the New Forest cicada (from Slovenia), a dark bush-cricket or a Roesel's bush-cricket (from the New Forest). Some recordings contain different types of noise, including people speaking, walking, calls of birds, handling noise and even people mimicking the sound of the cicada. As discussed before, this data set represents the quality to be expected in real, crowdsourced recordings. Each recording was later labelled by domain experts as containing either one or none of the insects of interest. Although multiple insects in the recordings will not make the classification fail, only one singing insect per recording is here considered. If more than one is present, the ground truth is set across the 30-second recording as the longest or loudest singing insect, therefore taking the state active for the longest period as the outcome of the model. Since the emission probabilities in the model are purposely tuned, no training data is required, and hence the entire data set is used for testing.

Table 3.3 shows the precision, recall and F_1 score metrics of the proposed approach compared to the three variants over the data set of recordings from the New Forest and Slovenia. Similarly, Figure 3.18 reports the true and false positives, with real values along the y axis and predicted class along the x axis. It can be seen that the proposed approach (CDA) achieves an F_1 score of 0.82, and as such outperforms each benchmark variant, visually apparent from the darkness along the main diagonal in Figure 3.18a. In contrast, the variant which uses the raw frequency measurements

	I	C	D	R
I	.62	.38	0	0
C	0	.78	0	.22
D	.25	0	.50	.25
R	.07	.13	.01	.79

(A) CDA

	I	C	D	R
I	.75	.25	0	0
C	.38	.60	.01	.02
D	.75	.25	0	0
R	.43	.43	.09	.05

(B) CDA raw frequencies

	I	C	D	R
I	0	1	0	0
C	0	.99	.01	0
D	0	.88	.12	0
R	0	.65	.35	0

(C) CDA silence removed

	I	C	D	R
I	.62	.38	0	0
C	.07	.58	.34	.02
D	.25	.25	.50	0
R	.07	.14	.77	.02

(D) Mixture model

FIGURE 3.18: Confusion matrices for the four variants of the detection algorithm. On the y-axis, the actual class; on the x-axis, the predicted class.

as the HMM feature vector (CDA raw frequencies) receives an F_1 score of 0.62. This is a result of the approach's lack of robustness to noise, such as handling noise, as shown by the high number of false positives in Figure 3.18b. Furthermore, the variant of the CDA which removes the silent periods (CDA silence removed) receives an F_1 score of 0.75. Although this appears as positive result, Figure 3.18c highlights its lack of ability to discriminate between the dark bush-cricket and the New Forest cicada. This method, as well as the raw frequencies approach, favours the New Forest cicada, scoring a good true positive rate but consequently also a high false postive rate. Finally, the mixture model method receives an F_1 score of 0.67 because the lack of transition probabilities leaves the decision to the emission probabilities only, not utilising the information contained in the time domain, making the number of true and false positives more equally distributed (Figure 3.18d). Insects with similar emission probabilities, such as the Roesel's bush-cricket and the dark bush-cricket, will therefore be difficult to classify with this method. It should be noted however that this approach is considerably more computationally efficient, as it decides on the most likely state instantaneously and without traversing the entire recording.

Figures 3.19, 3.20, 3.21 and 3.22 show a comparison of the four approaches over a sample recording for each of the four species in the recordings analysed. The top plot of each figure shows a spectrogram with the time domain on the x-axis, and the frequency domain on the y-axis, with the magnitude of the frequency bins varying with the colour of the plot. Subsequently, the figure shows the most likely state identified by each approach. In each plot, the states are labelled as in Figure 3.14a, where I represents the un-sounded idle state (if present), C represents the cicada's

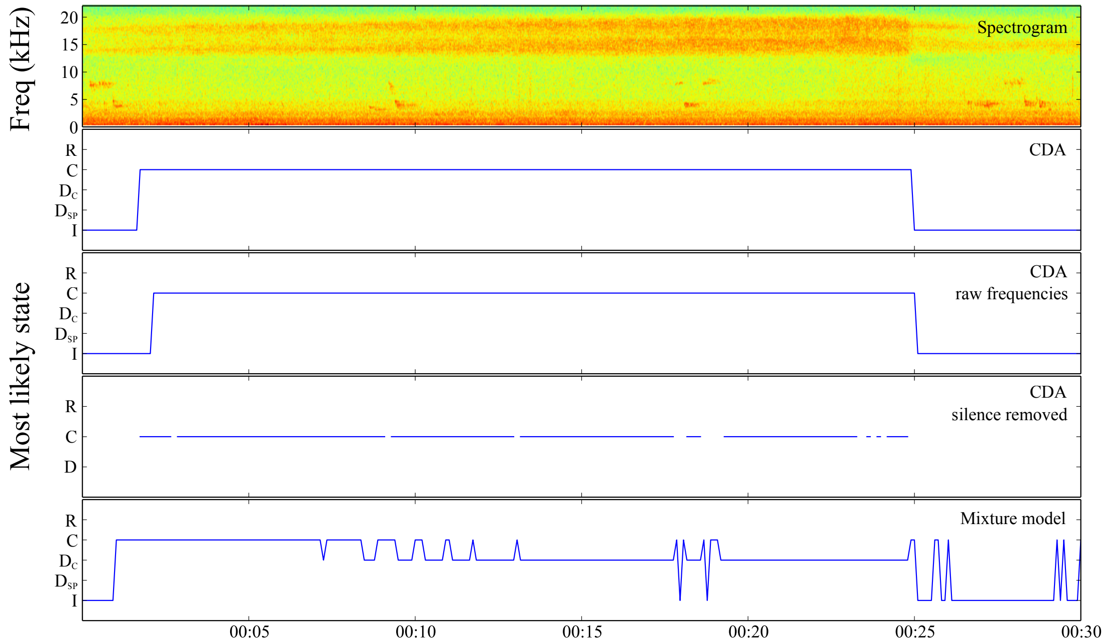


FIGURE 3.19: Model comparison on a New Forest cicada recording

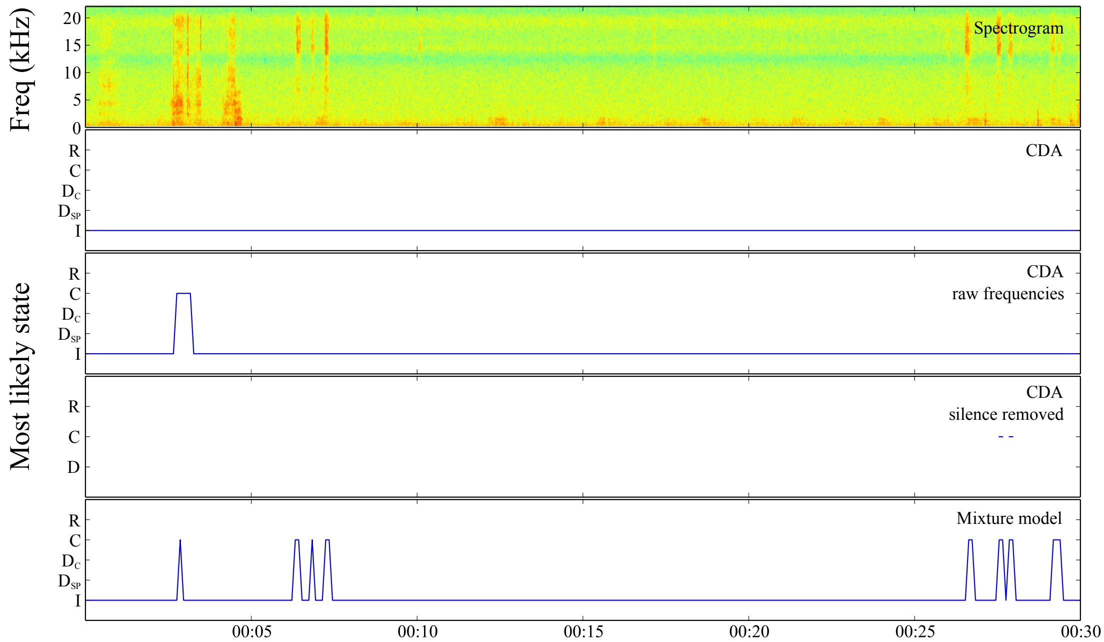


FIGURE 3.20: Model comparison on a recording with no singing insect

song, R represents the Roesel's bush-cricket and D_C and D_{SP} the dark bush-cricket's chirping and short pause states, respectively. The gaps in the silence-removed variant correspond to unsounded periods.

Figure 3.19 shows that classifying the cicada is easier for the HMM-based methods, as the call lasts for a long period without interruption and is clearly distinct from

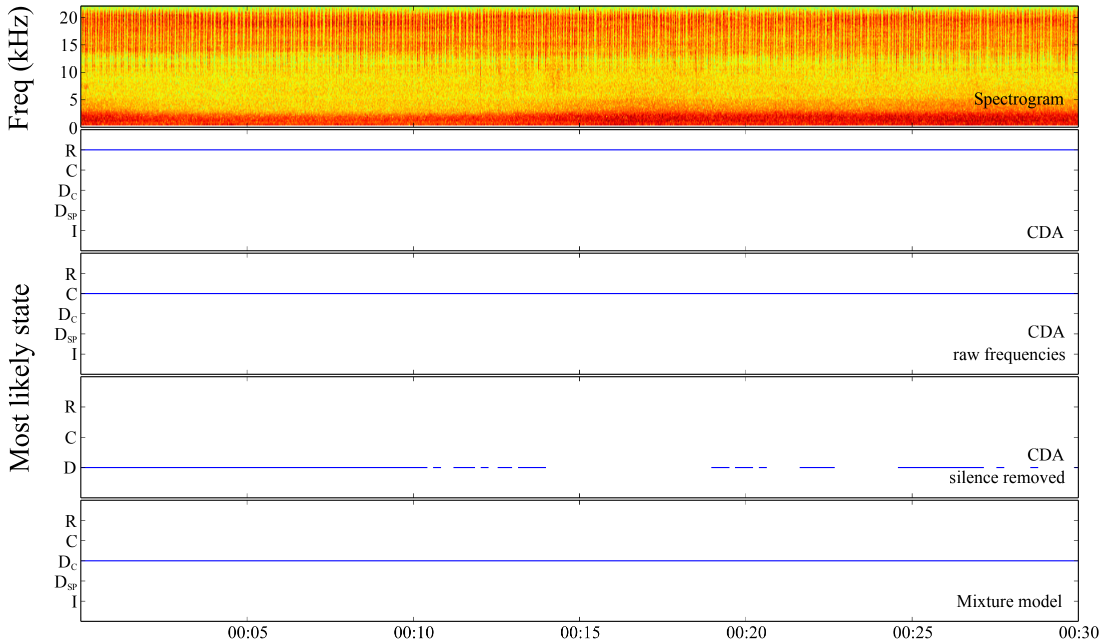


FIGURE 3.21: Model comparison on a Roesel's bush-cricket recording

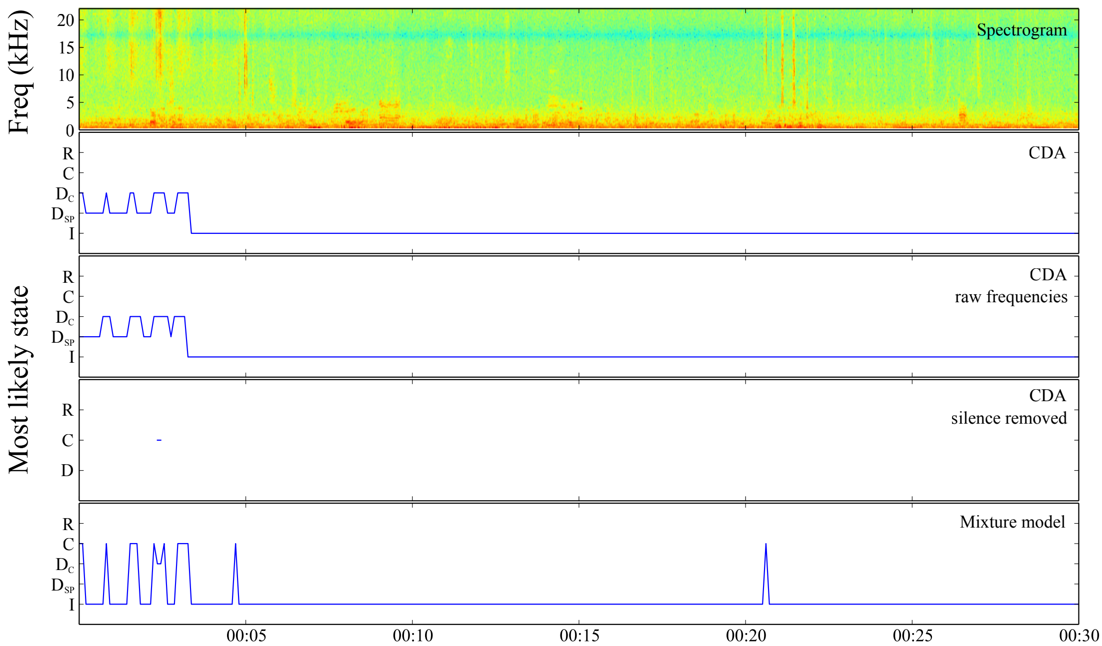


FIGURE 3.22: Model comparison on a dark bush-cricket recording

background noise. A more noisy recording would cause the raw-frequency approach to fail. The mixture model approach struggles to distinguish between the cicada and the dark bush-cricket call, since they are similar in features but different in the time domain, which this model does not capture. Figure 3.20 shows how the variants are more sensitive to noise than the CDA for different reasons. The raw frequencies approach doesn't filter out background noise, while the mixture model triggers a cicada state even for a very short noise in the right frequency band. The silence-removed method is only active in the short period of higher background noise, and not having an idle state, it is forced to classify the sound as any of the sounded states. Figure 3.21 shows how, when silence is removed, a Roesel's bush-cricket becomes very similar to a dark bush-cricket, having very similar emission probabilities. The same condition is observed by the mixture model, that doesn't have a perception of time. Similarly, Figure 3.22 shows that the dark bush-cricket is difficult to classify for the mixture model and the approach with silence removed, as explained thus far. Moreover, it shows how a trade-off between a very quiet insect (visible throughout the recording) and no insect must be made, as the insect could be at any distance from the microphone, and thus there is no limit to how quiet it may be.

The analysis and output of the 235 recordings is reported on the project's web site, with a page for *each* recording, together with the parameters of the HMM, the audio file, and information about the recording device[1]. This enables other researchers to replicate this method and compare results for each individual input in the data set. This comparison was also published in the Journal of Artificial Intelligence Research (JAIR) (Zilli et al., 2014).

3.6 Summary

This chapter has first presented a simple threshold-based classifier, efficient but not robust to noise. It has then extended the method by selecting a better set of frequency features, strong against noise and capable of indicating the presence of insects competing for a similar sound space to that of the New Forest cicada. It has proposed a

[1]Result at `http://www.newforestcicada.info/devdash`. The data can be used free of charge, provided that the New Forest Cicada Project is attributed according to the Creative Commons Attribution (BY) licence.

classification algorithm, based on a hidden Markov model that uses these frequency features to distinguish between New Forest cicada, dark bush-cricket and Roesel's bush-cricket, taking advantage of their signature in the time domain and fulfilling the requirement of this investigation to be able to detect the presence of the New Forest cicada in the wild. The chapter has then evaluated the approach against a state-of-the-art method, showing that the proposed system considerably outperforms the state-of-the-art. It has also introduced three variants that, informed by current practises in the literature, could have improved the classification. It resulted that the proposed method is still more accurate by a small margin. Finally, as this approach is tailored to power a smartphone-based acoustic classifier, the frequency response of some common smartphone models has been analysed, leading to the ability to tune the parameters of the HMM to match the specific device.

With that in mind, this thesis will now proceed to reporting the development, deployment and outcomes of the smartphone-based crowdsourced acoustic cicada detector that motivated the creation of the algorithm here described.

Chapter 4

Searching with Citizens

> The mantis stalks the cicada,
> unaware of the oriole behind.
>
> Writings of Zhuang-Zhou,
> Chinese philosopher, 4th century BC

Chapter 3 described the design and implementation of an algorithm for real-time detection of the New Forest cicada. Ultimately, the objective of this research is to produce a system that can help citizen scientists to detect the presence of the New Forest cicada. Therefore, this chapter describes the architecture of the system envisioned, its development, deployment and the initial results after two seasons of data collection. The citizen science project that was created around this system has been named "The New Forest Cicada Project" and its activities are collected and organised on the website `www.newforestcicada.info`.

4.1 System Requirements and Architecture

The system required is formed of different components, summarised in Figure 4.1. At the user's end, a smartphone client—an *app*—is required to collect observations. These consist of a record of whether a cicada was found or not, and require at least a time stamp and an accurate location in order to determine their validity. These two parameters alone are useful in a number of different ways. Firstly, knowing where

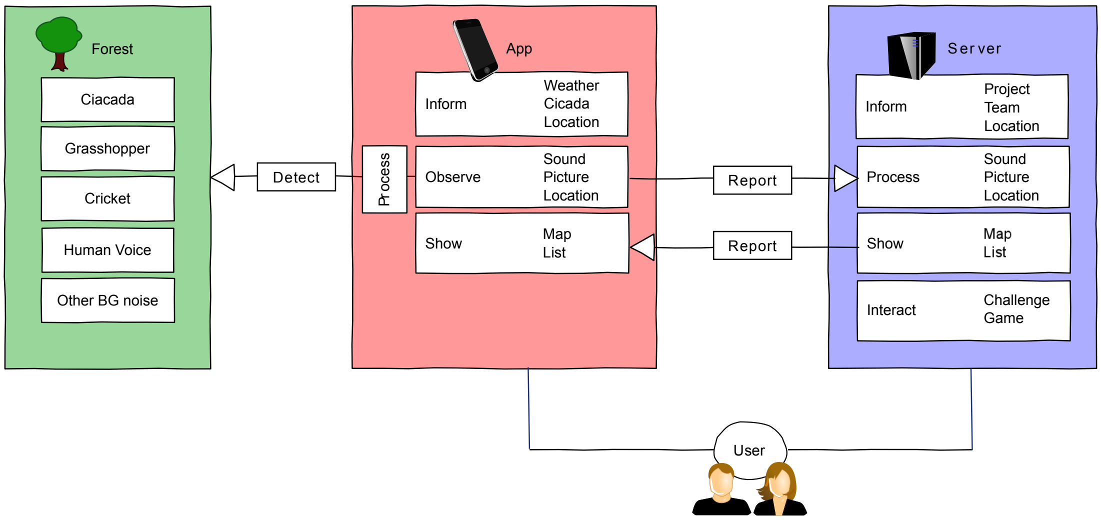


FIGURE 4.1: Architecture diagram of the system.

the cicada is not present at different times will generate a knowledge of the places where it is less likely to be found, and will give a degree of confidence that the insect can generally be considered missing from that area. Secondly, this knowledge can be represented on a map (for example as a probabilistic heat-map layer), with which users are recommended locations to visit. Thirdly, the data can be analysed to understand what parts of the forest are normally covered by users and potentially provide incentives to move to different areas. A model may be built with this information to understand respectively *a*) how people move, where do they go and how they can be motivated to go elsewhere, in order to cover more surface of the forest and *b*) with what confidence level it can be established that the New Forest cicada is really not present in the area considered, should it not be found.

To determine the presence of a cicada, the app processes the signal collected from the in-built microphone. Ideally, the sound should be stored for further processing, but as uncompressed audio uses large amounts of space, a selection must be made on what files to keep and what to discard. Finally, the collection of this information, i.e. time, location, likelihood of the presence of the cicada and sound recording, is hereon called a *survey* and the storage of this a *survey report*.

A second major component of the system is a server *back end* to store the information collected from the citizens' mobile devices. This requires a database infrastructure and an API to allow it to communicate with the mobile *app*. Additionally, further sound processing can be performed on the server, where more computational power is available. A front end to this database will also permit users to visualise and manage their contribution, promoting the establishment of an on-line community.

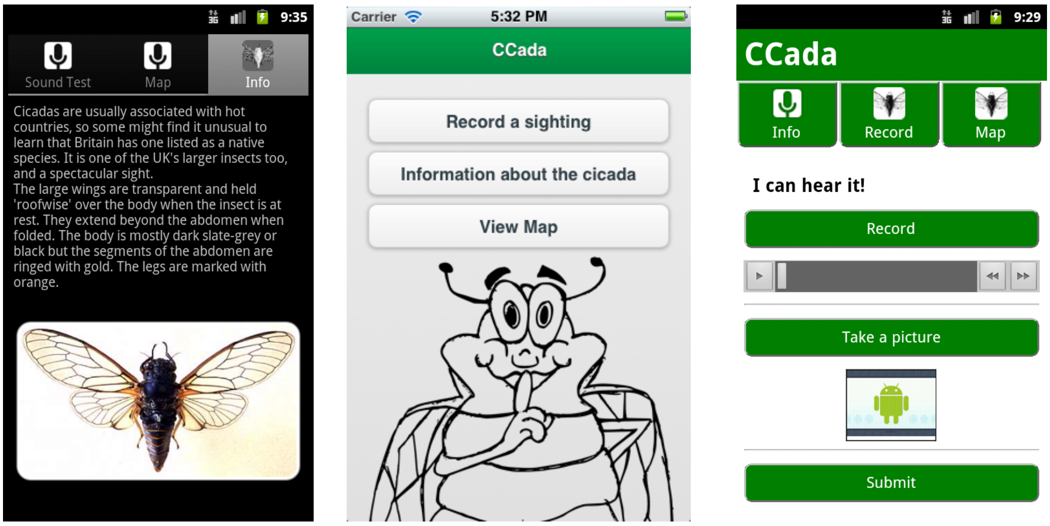
The combination of these components form the technical foundation required to address some key issues of this research, namely:

- to locate the vanishing animal so that it can be preserved and protected;
- to address the shortcomings of manual surveying, which is time-consuming and requires high expertise from the surveyor.
- to equip a large number of enthusiasts with inexpensive tools to perform this search.
- to test incentive mechanism in the field of biodiversity monitoring, well-established in terms of citizen science activism;
- to create a real time smartphone-based bioacoustic platform to act as a model for other applications, for example for the monitoring of different animal species.

The remainder of this chapter will describe how these components have been designed and implemented to meet the requirements outlined thus far.

4.2 Mobile Client

The mobile client is the part of the system with which users interact the most. Therefore, it is essential that careful design choices are followed in accordance to sound human-computer interaction principles to recruit and retain the largest number of users. Moreover, since the mobile development landscape is very varied with tens of different platforms and thousands of versions available, it is necessary for the success of the citizen science endeavour to target the appropriate devices in order to maximise coverage while minimising cost of development. This section outlines the design choices that were taken in relation to the principles followed.



(A) Native Android (B) PhoneGap on iOS (C) PhoneGap on Android

FIGURE 4.2: Prototype apps, on three different pages. On the left, the information page about the cicada in native Android; in the middle, the home page on an iOS device and on the right the observation page, which allows to take a recording, a picture or a combination of both and submit them to the servers.

4.2.1 Initial App

An initial *app* has been developed for Android and iPhone devices. Two different approaches have been implemented for prototyping, a native client that uses the platform's APIs directly and a PhoneGap[1] client, a cross-platform framework that allows HTML5 development, exposing platform calls through a Javascript API (see Figure 4.2).

Furthermore, the PhoneGap client has been designed with two different interfaces, aiming to look native on both iOS and Android. The former uses jQuery Mobile, another cross-platform HTML5 framework that provides a javascript library and iOS-like styling, optimised for touch-screen devices. The latter uses xUI, a lightweight javascript library similar to jQuery Mobile but with no UI styling. It strikes clearly that a compromise between portability and native look must be found, as using different UI development frameworks impedes portability, but native look is important to provide a professional appearance and a more responsive interface. However, the majority of the core functionality of the application needs to be implemented in platform-specific

[1]PhoneGap, `http://phonegap.com/` is also known as *Project Cordova*, acquired by the Apache Software foundation and currently in the Apache Incubator (`http://incubator.apache.org/`).

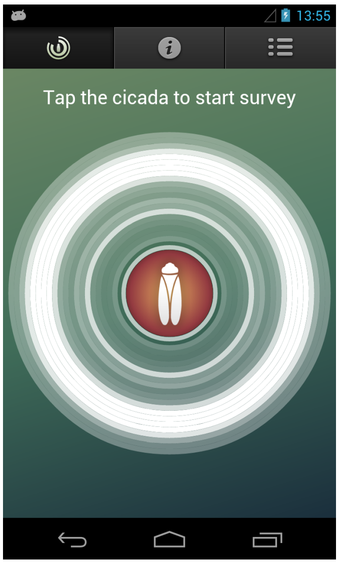
language to access lower level functionality and obtain maximum performance, especially when processing audio signals in real-time.

4.2.2 The App Deployed

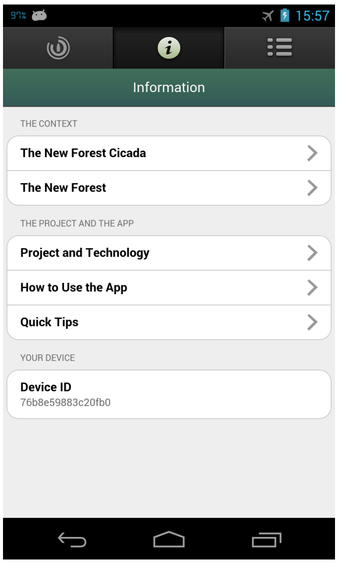
After several iterations, the mobile client was designed according to the following principles. The user interaction is controlled by a cross-platform HTML5, CSS and javascript-powered interface, which communicates with the underlying platform through the PhoneGap framework, separating the components according to a Model-View-Controller (MVQ) pattern. The app is developed for Android and iOS and all computationally expensive tasks, including the sound analysis, run in a PhoneGap *plug-in* (PhoneGap Development Team, 2013), implemented in the underlying platform's development kit language (Objective C for iOS, Java for Android)[2]. This ensures high performance while maintaining the cost for the front-end development low. The app was released to the markets in early June 2013 (see Section 4.5) under the name of *Cicada Hunt* and the API used to communicate between the front-end and the back-end is reported in Appendix A. The user interaction can be grouped in three areas, which correspond to the three tabs in the main screen, exemplified by screen captures in Figure 4.3.

Detector page Firstly, the app presents the detector page. A crucial difficulty for a human to detect the New Forest cicada's call is the fact that the pitch is too high for most people to hear, at the edges of the hearing range of the average 40 years old. To address this issue, this tab presents a visualisation of the sound drawn as a circular spectrogram. In the centre, the cicada logo lights up when a call is being detected, triggered by the instantaneous output of the Goertzel filter described in Section 3.3, updated every 128 samples from the microphone. Twenty concentric circles around it represent twenty frequency bands of the spectrum, centred from 1 to 20 kHz with bandpass of 1.4 kHz, which ensures rapid updating of the filter used for detection. Each of these becomes brighter with a higher signal strength (i.e. a louder sound at

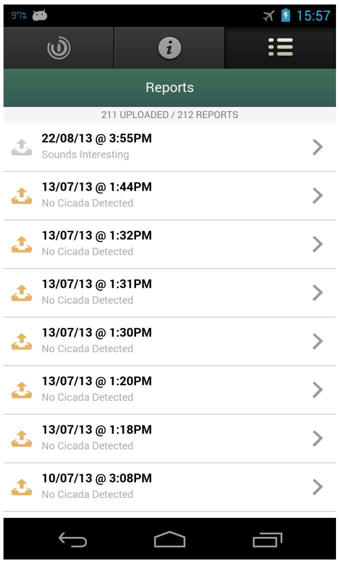
[2]The front-end was developed in collaboration with an external company; the iOS plug-in was developed by Prof Alex Rogers, while the Android plug-in by the author of this thesis.



(A) Detector page, with cicada being detected



(B) Information page.

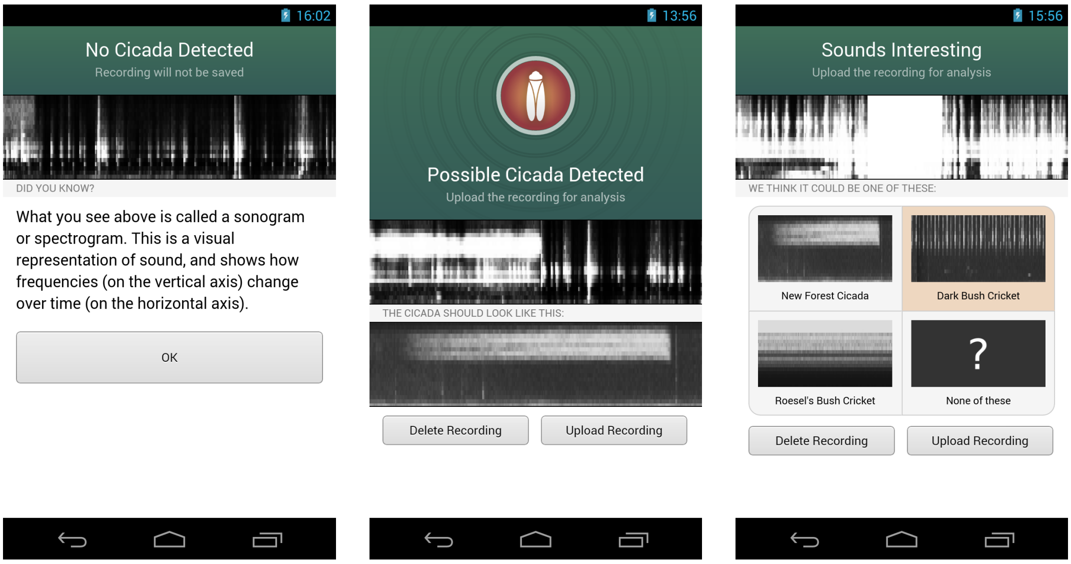


(C) Reports page.

FIGURE 4.3: The three main screens of *Cicada Hunt* on Android.

that pitch), paler when the band is quieter. The outer ones, roughly from 12 to 18 kHz are those triggered by the cicada call, producing the distinctive effect shown in Figure 4.3a. Tapping the centre of the app starts a 30-second *survey*, where the sound is recorded and then analysed by the HMM-based algorithm described in Section 3.4. This interaction is core to the interface, as it encourages users to stop and wait in silence, thus maximising the chances of detecting the required sound.

As the survey finishes, the user is presented with one of three cases, as shown in Figure 4.4. If nothing was detected, a fact about the cicada, its habitat, the New Forest or the technology behind the app provide an informative notion, encouraging the user to try again (Figure 4.4a). This intends to both support the morale of the user who is receiving negative results, and to provide educational content, so that citizen scientists receive some information in exchange for the work they have performed. If a cicada was found, a positive message informs the user of the *potential* discovery, allowing for the algorithm to be in error or to be triggered by, say, the recording of a cicada call instead of a call itself (Figure 4.12b). The third case is provoked by the detection of another insect, similar in call to the New Forest cicada. At present, the app encompasses three other insects present in the New Forest: a dark bush-cricket, a Roesel's bush-cricket and a field grasshopper, though it does not convey to the user an authoritative distinction between these three. Instead, the user is shown a spectrogram



(A) End of survey where no cicada or other insect was found

(B) End of survey screen where cicada was found

(C) End of survey screen where the user can select options

FIGURE 4.4: Three states of the detector page, before and after a survey.

of a typical call of these insects, as well as a spectrogram of what they have just recorded (Figure 4.4c), and they are asked to select to which insect their recording looks most similar. This promotes the involvement of the user in the process, who would otherwise be passively observing the mobile agent execute the detection.

Reports page The end of a survey produces a *report*, which is initially saved locally. The report is geo-tagged and time-stamped, and saves a unique identifier of the phone as well as basic information about the device. As soon as an internet connection becomes available, the report is uploaded to the project's servers, where it is available to the research team to analyse further. This apparently minor precaution is actually crucial, as a data connection is often missing or unreliable in the forest. Users are also allowed to manage their reports by logging onto the project's website and registering their device with the online system (described later in Section 4.3). The report also saves an uncompressed 44.1 kHz 16 bit PCM WAV sound recording in the case the cicada or another insect are found, and provided that the user has granted permission to do so. The file takes 2.7 MB on disk, so it is deleted as soon as the report is sent to the server. Lastly, a low-resolution spectrogram is saved in all cases, constructed from the combination of the output of the 20 Goertzel filters over the 30 seconds survey, saved every 128 samples. This constitutes the easiest way for a human to check for the presence of the cicada and provides no privacy concerns (speech could not be reconstructed from such spectrograms). Moreover, the payload of the image file, saved in Base64 (Josefsson, 2006), is around 15 KB and therefore much lighter than a full uncompressed sound recording.

Information page An information tab presents more extensive material on the New Forest, the cicada, the algorithm used and the device itself, as well as some tips on how to best use the app. This educational aspect is also very common in citizen science projects, where the time invested by the user is rewarded with information to learn more (Cooper et al., 2007). The "Tips and tricks" page is presented as a cartoon so as to be easy to read and accessible to the largest audience. Finally, the information page reports a unique identifier for the device, with which users can link their handset to an online account.

Finally, Figure 4.5 sums up the interaction of the user with the app, showing the flow one would follow once the client on their device has launched.

4.2.3 Other platforms

The principal target platforms are iOS and Android, as the two platforms in 2011, when the development was started, held the 75.6% of the smartphone market share (Go-Gulf, 2012) (note that by the last quarter of 2014 this value shot up to 96.3% of the market (International Data Corporation, 2015)). However, a simple, feature-reduced application may be developed in the future for other platforms to ensure the best coverage of the population. In fact, it is difficult to assess beforehand who the users of such a citizen science effort would be. 80% of the world population has a mobile phone, 20% of those have a smartphone, 89% of whom use their smartphone throughout the day (Go-Gulf, 2012). These statistics are not, however, linear with age, and the penetration of mobile phone and smartphone users decreases in the older population. An analysis of the users performed through the data collected is presented later in Section 4.5.

A minimal platform for the app could work with as a little as the GPS only, allowing users to indicate where they think they have detected a cicada by ear. Alternatively, in

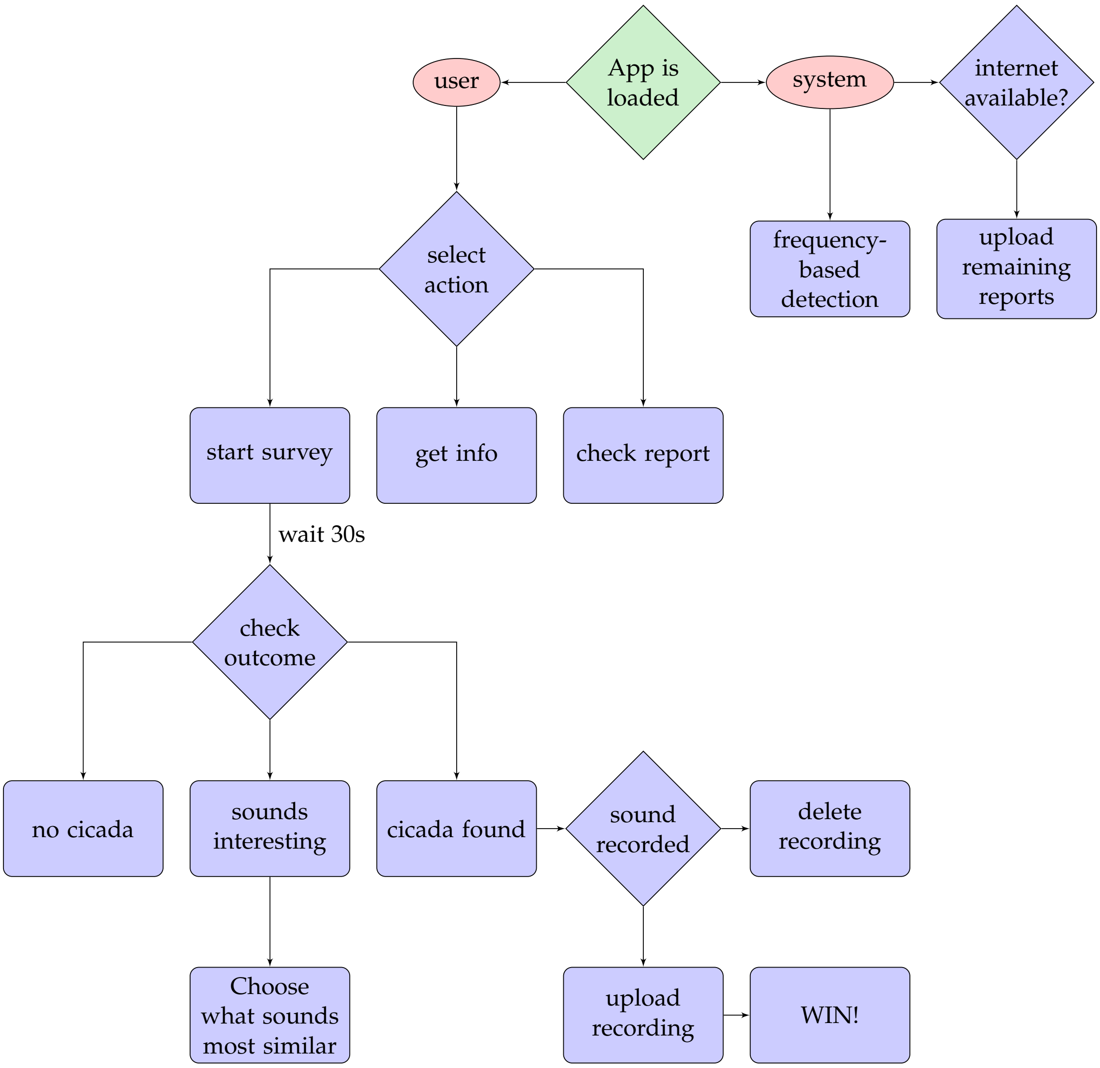
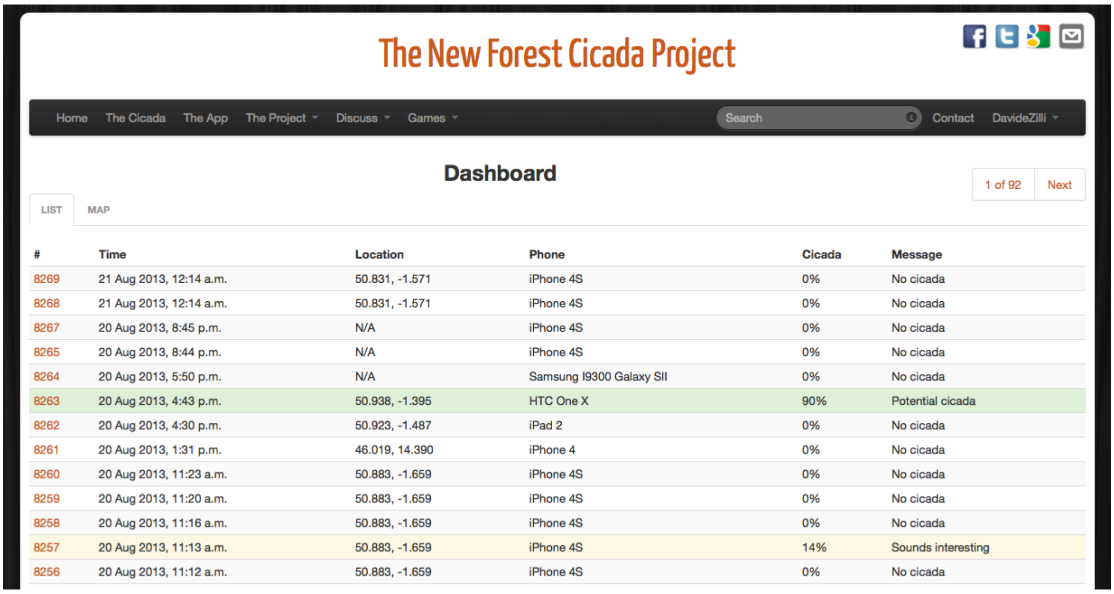


FIGURE 4.5: App flow, as experienced by the user.

order to include feature phones, an even simpler approach would consist in allowing users to text their position to the best of their knowledge to report a potential sighting. Finally, it should be mentioned that the platform need not be constrained to a mobile phone. A custom-built device with a low-power microcontroller, a microphone and a battery can be built for a few pounds, and deployed stand-alone in the forest, as part of a wireless sensor network or even embedded in a wrist band for users. These options have been initially considered, but their implementation lies beyond the scope of this thesis and is therefore discussed in Section 6.1 as possible future work.



#	Time	Location	Phone	Cicada	Message
8269	21 Aug 2013, 12:14 a.m.	50.831, -1.571	iPhone 4S	0%	No cicada
8268	21 Aug 2013, 12:14 a.m.	50.831, -1.571	iPhone 4S	0%	No cicada
8267	20 Aug 2013, 8:45 p.m.	N/A	iPhone 4S	0%	No cicada
8265	20 Aug 2013, 8:44 p.m.	N/A	iPhone 4S	0%	No cicada
8264	20 Aug 2013, 5:50 p.m.	N/A	Samsung I9300 Galaxy SII	0%	No cicada
8263	20 Aug 2013, 4:43 p.m.	50.938, -1.395	HTC One X	90%	Potential cicada
8262	20 Aug 2013, 4:30 p.m.	50.923, -1.487	iPad 2	0%	No cicada
8261	20 Aug 2013, 1:31 p.m.	46.019, 14.390	iPhone 4	0%	No cicada
8260	20 Aug 2013, 11:23 a.m.	50.883, -1.659	iPhone 4S	0%	No cicada
8259	20 Aug 2013, 11:20 a.m.	50.883, -1.659	iPhone 4S	0%	No cicada
8258	20 Aug 2013, 11:16 a.m.	50.883, -1.659	iPhone 4S	0%	No cicada
8257	20 Aug 2013, 11:13 a.m.	50.883, -1.659	iPhone 4S	14%	Sounds interesting
8256	20 Aug 2013, 11:12 a.m.	50.883, -1.659	iPhone 4S	0%	No cicada

FIGURE 4.6: The *dashboard* of survey reports, as visible by an entomologist or a staff member on the website. In green a potential cicada discovery, in yellow the report of another insect.

4.3 Server Back-end

The mobile client requires a central server to collect survey reports. This has been implemented in Django, a Python web framework, and powered by a PostgreSQL database. The implementation details of the server are not relevant to this thesis, but the API that governs communications between the client app and the server are reported in Appendix A.1.

The data collected from the app is reported in a *dashboard*, publicly available at `http://newforestcicada.info/dashboard` (see Figure 4.6). Here users can register their devices and see where they have been surveying the forest, accessing a list of their personal reports. At the same time, a small set of project authors and collaborating entomologist have access to the list of all reports submitted and can classify recordings by replaying the sound and analysing the spectrogram.

The report page, an example of which is depicted in Figure 4.7, gives a detailed description of all the data linked to a species survey, in particular recording and uploading date and time, device model, app and framework version, surveyor (if known), a map of the location and the low-resolution spectrogram generated by the app with the output of the 20 Goertzel filters. If a sound file is attached to the survey, the page will

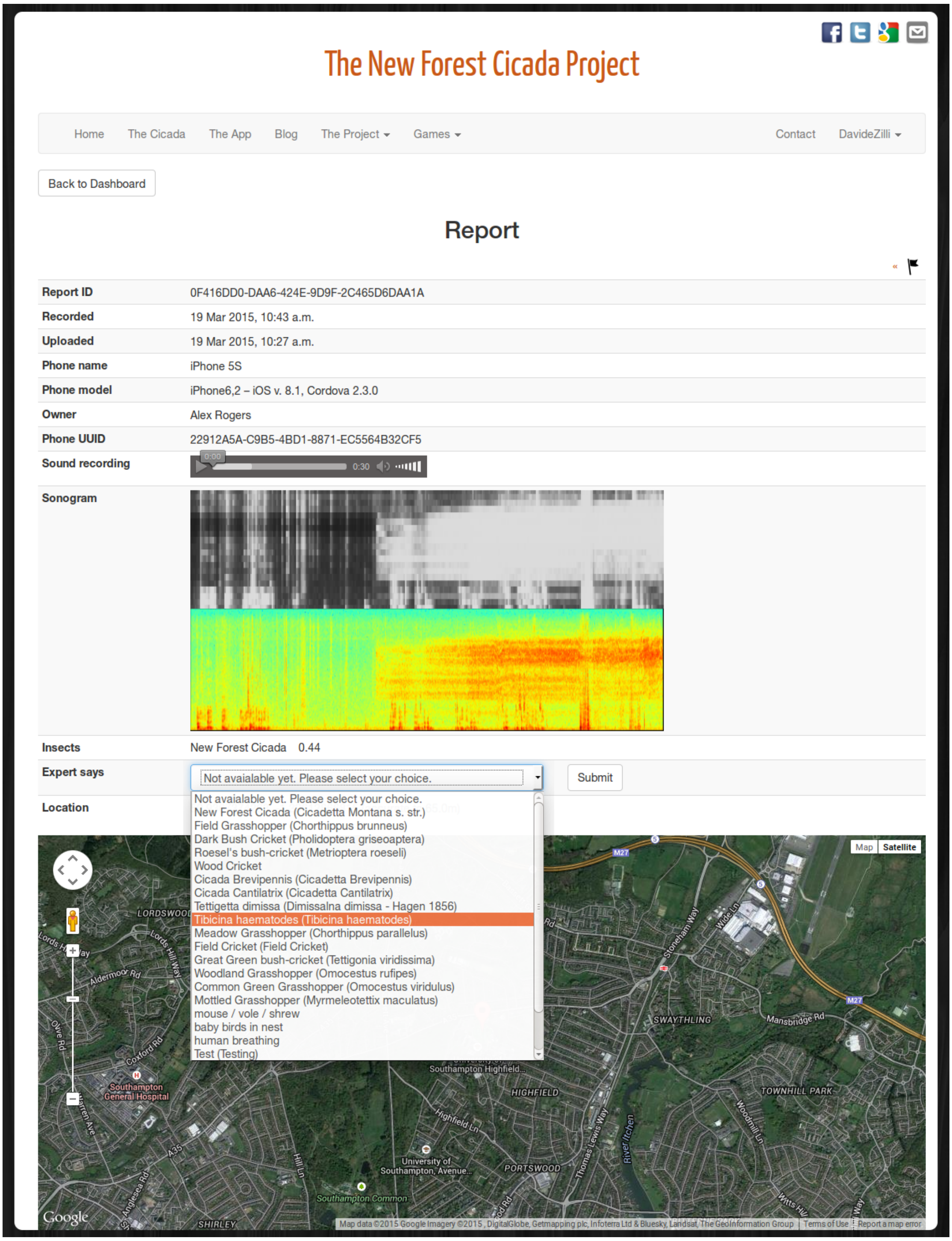


FIGURE 4.7: A report page in the dashboard, outlining some of the features. In this instance, a cicada recording was played in the microphone of the iPhone 5 at the University of Southampton.

also produce a high quality spectrogram as in Figure 4.7, as well as giving controls to play the sound. These are instrumental for an entomologist to manually confirm the classification.

Towards this purpose, a drop-drown menu allows entomologist to override the classification given by the app, providing an expert review of the recording. This has proved crucial for the correction of errors in the algorithm and the training of a more accurate model between the first and the second season of deployment.

4.4 Evaluation of Microphone Frequency Response

A major challenge faced in developing the app so as to be suitable for the largest possible number of devices is the difference in hardware components. Different screen sizes, for example, require either a tailored design, or an adaptive interface that can scale to small smartphones as well as large tablets. Even more challenging, however, is the difference in sensitivity of the built-in microphones.

Empirical tests reveal that some smartphones are equipped with a very sensitive microphone, while others have strong limitations. This generally relates well to the price of the device, but it is not always the case. To this respect, smartphones can be broadly divided into three categories; a) those with a high quality microphone, very sensitive to cicada call (generally quite expensive devices) b) those with a low quality microphone and not so sensitive, but still capable of detecting the cicada call when very loud (generally cheap devices); c) those with a sensitive microphone, but filtered in hardware, and therefore not capable of identifying the call at all (also often quite expensive handsets).

In order to quantify these claims, a range of different devices has been tested. The test consisted in reproducing four types of sound for at least 2 seconds each. Silence, white noise, a frequency sweep from 50 to 20,000 Hz, and the cicada call. These were reproduced in a custom-built sound-proof chamber, itself placed in a quiet location, with a Visaton KE 25 SC 8 Ohm tweeter. The phones were all arranged with the microphone facing the speaker and all equally distant from it. The sound volume was calibrated so that the volume of the cicada call was equivalent to that likely to be

Device	*Filtered*	*Silence (SEM)*	*Cicada (SEM)*	*Ratio*
iPhone 4	No	1.623 (0.075)	13.047 (0.327)	8.041
iPhone 5	No	1.897 (0.076)	14.793 (0.388)	7.800
iPhone 4S	No	1.466 (0.050)	10.549 (0.337)	7.196
iPhone 3	No	1.469 (0.047)	10.539 (0.430)	7.173
Telinga	No	1.500 (0.044)	7.658 (0.233)	5.104
HTC Desire	No	0.844 (0.041)	4.255 (0.265)	5.041
Xperia Mini	No	2.480 (0.155)	10.190 (0.262)	4.109
Moto A953	No	2.015 (0.104)	5.845 (0.148)	2.901
Galaxy S3	No	1.374 (0.038)	3.279 (0.088)	2.387
Xperia Z	No	0.951 (0.032)	1.971 (0.059)	2.072
HTC One S	No	1.466 (0.040)	2.915 (0.085)	1.988
Nexus 4	No	0.675 (0.025)	1.314 (0.026)	1.946
HTC Desire X	No	1.243 (0.054)	1.817 (0.075)	1.462
Galaxy Ace 2	No	1.953 (0.063)	2.162 (0.059)	1.107
Galaxy S2	No	1.916 (0.085)	2.101 (0.031)	1.097
Nexus One	Yes	1.514 (0.051)	1.568 (0.045)	1.036
HTC One X	Yes	1.933 (0.062)	1.732 (0.052)	0.896
HTC Wildfire S	No	2.032 (0.088)	1.683 (0.063)	0.828

TABLE 4.1: Comparison of popular smartphone devices. Values are means of ratios of 14 and 8 kHz Goertzel filters, sampled every ≈ 3 ms (128 samples at 44,100 kHz). In brackets, standard error of the mean.

detected by the phone in the wild, using the measurements obtained by recording cicada calls in Slovenia. The synthetic white noise and and frequency sweep were tuned accordingly. Finally, to collect test recordings and to automate their transmission to the server[3], an auxiliary app called *SoundCheck* has been implemented. This is capable of either downloading the up-to-date benchmarking sound and self-testing the device through its own speaker, or recording from an external speaker. The latter option is more accurate as it is independent of the quality of the device's speaker, and therefore it was the solution used for the tests here performed.

A comparison of the sensitivity of the microphones based on how well they are capable of detecting the cicada call in the test environment described above is hence reported here. Table 4.1 summaries the outcome of the test, reporting the standard error of the mean (SEM) of the ratio between the 14 kHz and 8 kHz bands extracted with the Goertzel filter when no sound was played (marked as *Silence*), when the cicada call was played (*Cicada*), and the ratio between these two. A higher value of the latter means a clearer indication of the cicada call, and therefore a greater confidence in the detection. Figure 4.8 shows the reference sound played to the phone, together

[3]The database of recording can be found at http://newforestcicada.info/phonetest/list

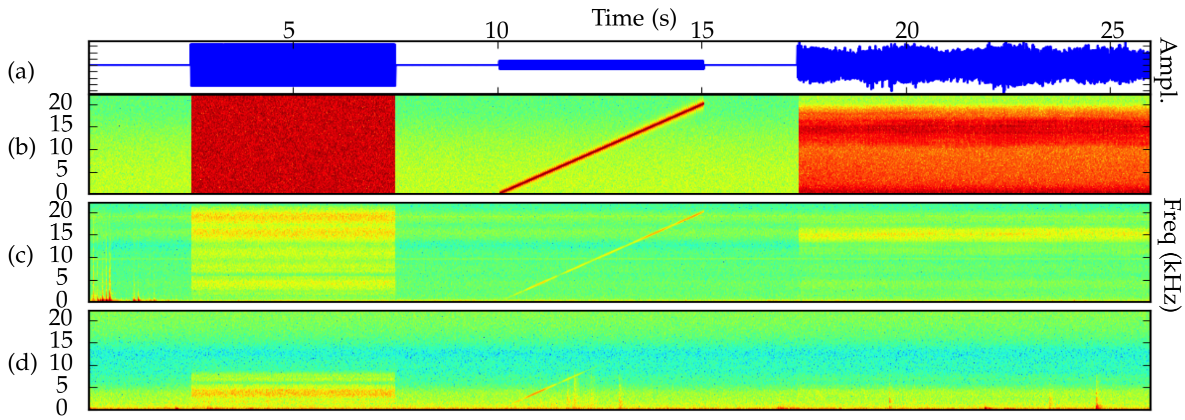


FIGURE 4.8: Comparison of two phones. At the top, the waveform (a) and spectrogram (b) of the sample calibration file. At the bottom, the very sensitive iPhone 5 (c) and the hardware-filtered HTC One X (d), both top-end devices for iOS and Android.

with two high-end devices; the Apple iPhone 5, detecting the cicada call very clearly, and the HTC One X, hardware filtered and therefore incapable of detecting the insect's call. Table 4.1 confirms this, as the two devices score values at opposite ends of the scale.

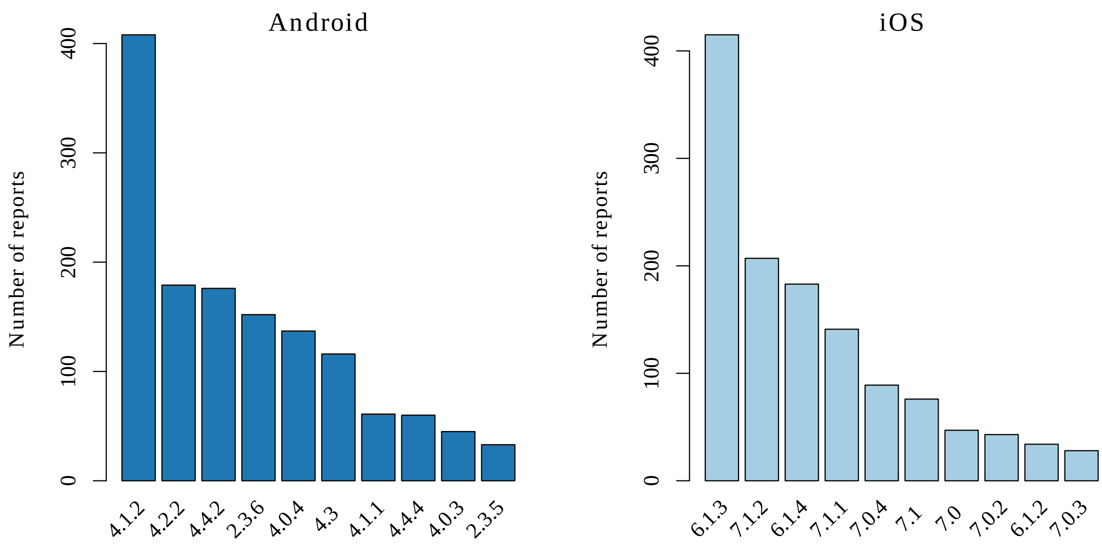
This comparison results in the ability to calibrate the emission probabilities of the HMM to the specific phone model. For devices not yet calibrated, the app selects a set of generic parameters that are suitable for most microphones, slightly skewed to discourage an abundance of false positives that would mislead the user. The app downloads a list of calibration parameters at run time, so that new devices can be calibrated without having to push an update of the app to the respective markets.

4.5 System Deployment and Users

The system presented thus far was officially launched on June 8th, 2013 at the New Forest BioBlitz, an annual event held by the New Forest National Park Authorities to engage the public into mapping biodiversity.

The user base this project is aiming to involve can be broadly divided in three categories.

Bug enthusiasts. The most technical group, these users are willing to set out to search for the endangered insect. The group also includes entomologists who are passionate or even paid to work on the species. Most likely, the smallest group.



(A) Android versions

(B) iOS versions

FIGURE 4.9: Operating system versions of the devices that submitted reports between June 2013 and March 2015.

Locals. Potentially the most assiduous group. This includes people living or working locally in the New Forest, who have a general interest in the park and its ecology. They may interact with the system often and for short periods of time.

Tourists and visitors. The most numerous group. The New Forest reports 13.5 million day visitors every year (New Forest District Council). They may be the least active, using the app only once or a few times, but scattered across various areas of the park and available in the right season for the insect's adulthood—hot sunny days, between May and July. They are also likely to be present around camp sites, some of which are historical know sites of cicada emergence.

Since its launch, the app has been downloaded around 3000 times. 2968 unique devices have submitted reports, in most days of the cicada adulthood period. Over 11000 reports were submitted, and of these 1303 have an audio track attached. 2577 reports were taken within the boundaries of the New Forest, but an additional 3000 does not have a GPS location, either because the user decided not to share it or because a GPS fix could not be obtained in time. Overall, 1482 devices (50%) were running iOS, while 1486 (50%) were running Android—a very different proportion from the first season alone, when over 65% of devices were running iOS. Figure 4.9 shows a breakdown of the most popular operating system versions for the two platforms. With a small

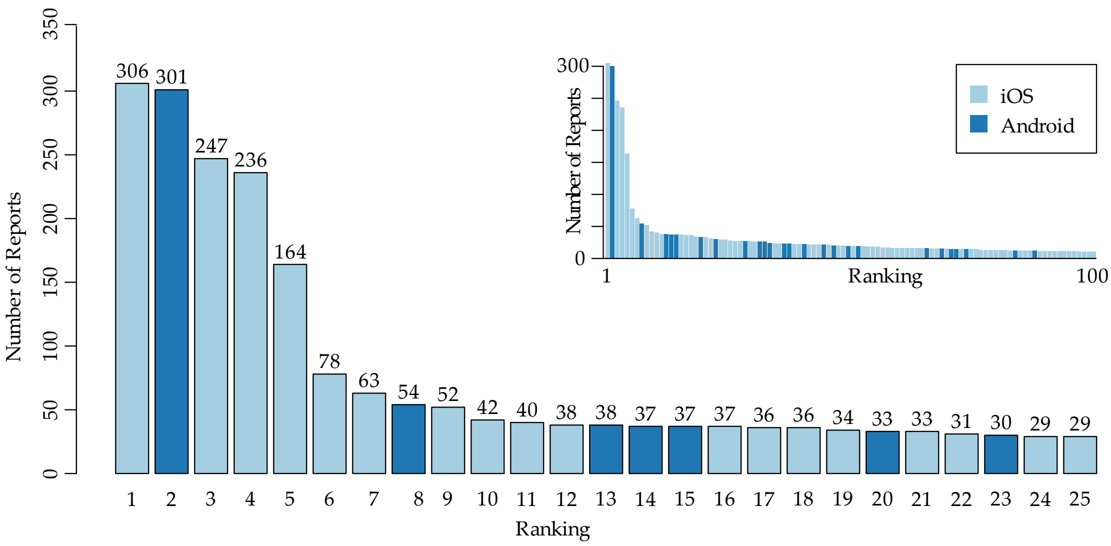


FIGURE 4.10: Reports per user by operating system for the top 25 users (right, trend of the top 100).

amount of resources to be allocated to development, these analytics are paramount to understanding what devices should be targeted first and to provide support for those that will enable the widest participation.

Figure 4.10 shows a bar graph of the number of reports uploaded by the top 25 contributors, with the trend for the top 100 users displayed in the top-right corner. It should be noted that among these, 5 are entomologists and member of the research team. However, these users only covered specific areas of the forest, in particular those where the cicada had historically been observed. In contrast, the citizen scientists submitted far fewer reports per user, but the reports were more evenly distributed across the New Forest, as shown in Figure 4.11. This shows the crucial difference that this distributed approach can make, as entomologists cannot be ubiquitously present in different areas of the forest when the conditions are favourable, and can only cover a limited territory, while visitors, though contributing individually less, can help rediscover the cicada if it has moved to different sites, as it is currently suspected. At the same time, while entomologists have the tools and the knowledge to recognise insects' calls, the general public must be equipped with an accessible method. In this space, the implementation and deployment of this automated acoustic insect detection algorithm has succeeded to bring to the public the possibility to contribute to the distributed monitoring of insect species, as shown by the large number of downloads of the app and submitted reports.

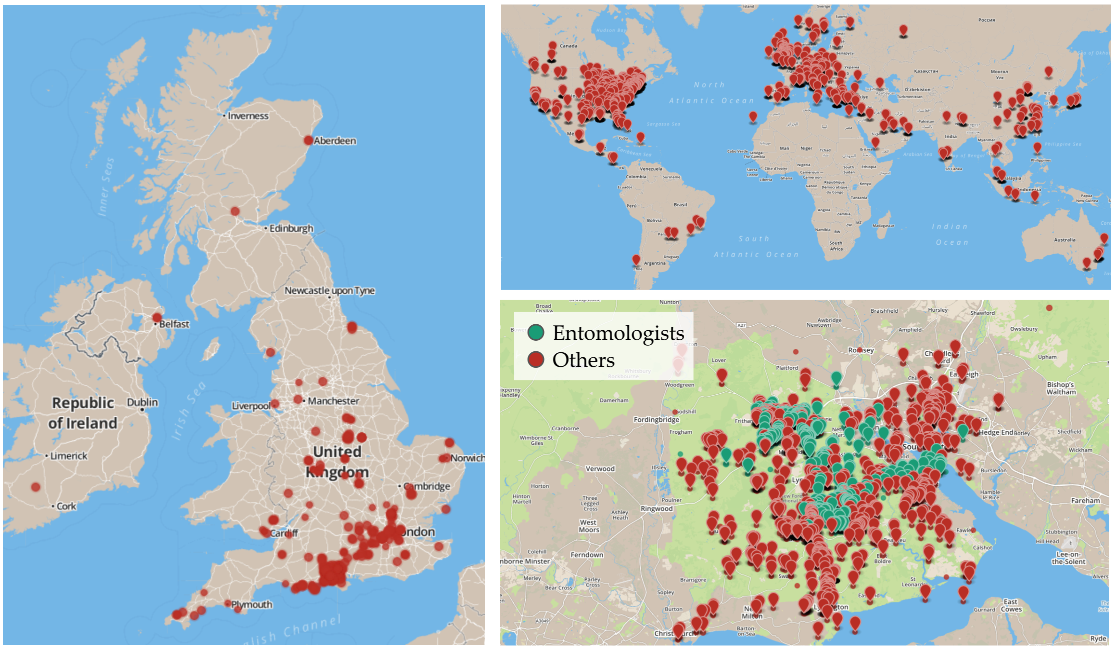


FIGURE 4.11: Location of submitted reports. In 2014 the 22nd brood of the North-American periodical cicadas (*Magicicada*) emerged, and attracted great interest on the topic of cicadas in general, hence the number of reports in North America.

A full report, up to date at the time of writing, is presented in Appendix Appendix B. No *Cicadetta montana* has yet been found in the New Forest.

4.5.1 System Testing

With no individual of cicada found, the need arose to test the system in a way that could prove its correct operation. This verification can be separated into two stages. Firstly, the assessment of the HMM-based algorithm powering the app; this has already been showed in comparison to the state-of-the-art approach for batch classification, and its better performance in the current scenario reported in Section 3.5. Secondly, the system may fail in the implementation of the algorithm onto the mobile client and in the integration of the infrastructure as a whole.

To test the latter, the research team once again travelled to Slovenia, where the first recordings of real *Cicadettae montanae* were made. There, it could be showed that the system was perfectly capable of detecting the cicada, and this resulted in the collection of hundreds of recordings. Due to the lack of ground truth, it is not possible to assess precision and recall of the system, but photographic and videographic evidence of

FIGURE 4.12: Female *Cicadetta montana* being held by entomologist Tomi Trilar, and Cicada Hunt detecting the presence of the insect, both in Slovenia.

the successful detection is reported on the project's website, and the full database of recordings is available upon request.

Furthermore, the interaction of users with the deployed system has been investigated to understand usage patterns and barriers in the uptake of the system by the wider community. The outcome of a number of interviews was reported in a paper appeared in the Proceedings of the 32nd annual ACM conference on Human factors in computing systems (CHI 2014) (Moran et al., 2014) and reported a certain level of resistance against the use of smartphone in the field, as well as providing guidance for the implementation of future systems in a similar space based on the expertise gathered in this experiment.

4.6 Publicity and Public Engagement

Significant user participation is necessary for a citizen science project to succeed. To expand the number of users, a certain level of publicity and public engagement is therefore indispensable. The New Forest Cicada Project was involved in the Science and Engineering Day and Solent Big Bang Fair at University of Southampton; the New Forest BioBlitz 2013 and Wood Fair 2013 and 2014; the British Science Festival 2013 in Newcastle; the BBC Summer of Wildlife 2013 in Birmingham, as well as talks in several conferences, meetings and schools. This resulted in articles or mentions on, among others, BBC News, The Guardian, The Daily Echo, BBC Wildlife Magazine, EnvioNews, Wildlife Extra, BugLife, Gigaom, DEFRA Magazine, the New Scientist and an interview on BBC Radio (a list of all media engagement is presented in Appendix E. This, together with presence on social media, contributed to the recruitment of hundreds of users.

4.7 Summary

This chapter has presented a novel, mobile citizen science platform for searching for an endangered insect—the first of its kind to the best of the author's knowledge. The system is composed of a mobile app for iOS and Android devices, catering to the vast majority of smartphone users, and a website that collects all the users' observations and permits the involvement of field experts. It showed how careful design is imperative for the participation of a broad audience, and a simple, intuitive interface can engage hundreds of citizen scientists. Over 3000 users have downloaded the app to date, and more than 11000 reports were submitted, making it one of the largest citizen science projects of its kind. While the effectiveness of the system was confirmed through a field trip in the Slovenian Alps, where the species is present, no *Cicadetta montana* was reported in the New Forest during the two mating seasons in which the system was deployed, and it therefore remains missing in the UK. However, the app is also used in survey work by professional entomologists in different locations across Europe. Finally, a question remains as to whether a similar system can be used for a broader set of insects and for wildlife in general. The following chapter attempts to address this matter.

Chapter 5

Broader Insect Sound Recognition

What would be left of our tragedies
if an insect were to present us his?

Emile Cioran,
Romanian philosopher, 20th century

Thus far, this research has demonstrated the usefulness and applicability of the usage of smartphones and citizen science to searching and monitoring the presence of one animal species in the wild. However, the app developed and the classification algorithm used are tailored to the New Forest cicada, and as such are not able to assist in the monitoring of other wildlife, if not for very few similar insects. Therefore, in order to extend this method for wider applicability, the present chapter describes a different classification algorithm, modelled on state-of-the-art techniques used in the bird classification literature, inspired particularly by the work presented by Stowell and Plumbley (2014). Since the focus remains on insects, the method here presented attempts to exploit two features commonly found in most species, that is a dominant frequency or set of frequencies, and a repetition of phrases at regular intervals. The following chapter will then evaluate this method against different data sets.

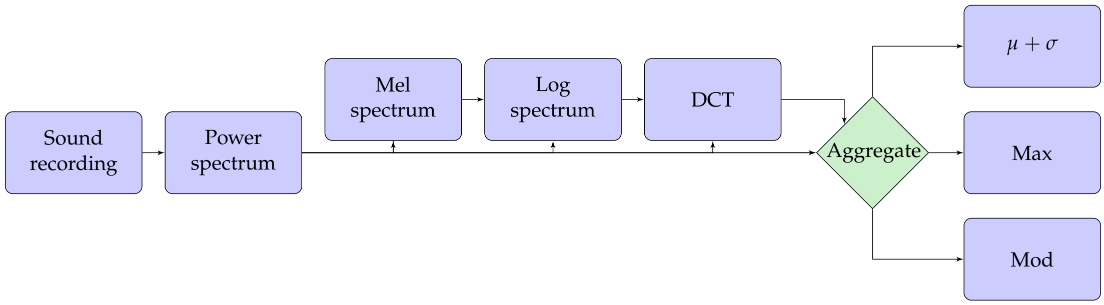


FIGURE 5.1: Flow of the possible combinations of feature extraction parameters and aggregation methods.

5.1 Feature Extraction

Many classification methods follow a similar procedure. The signal in analysis is initially (and optionally) pre-processed, where it is cleaned from noise and segmented into more tractable chunks. Subsequently, a set of features is extracted from it, generally based on some domain knowledge of the signal. Last comes an optional aggregation stage, where the features are down-sized to a more compact format, making them more tractable for the classifier. This procedure is repeated for all the signals in the data set to be classified. At a later stage, in order to assess the model, the data set is partitioned in two sets, so that the classifier can be trained on one part, and assessed on the other. The features of the training data are therefore fed to the classifier together with the ground truth (also known as labels) of the data. Once trained, the classifier is given the features extracted from the test set, from which it predicts the most likely class for each input sample. This is then matched to the ground truth of the latter data set to assess the accuracy of the classifier.

Figure 5.1 describes the feature extraction procedure in this investigation. First, the entire power spectrum is extracted, and optionally converted to the mel scale. This permits a comparison between the Hertz and the mel spectra, as it is argued here that the latter, although commonly used, is not beneficial for insect classification. The spectrum is then transformed into cepstral coefficients, but each step is disabled in turn. In fact, as seen in Section 2.2.1.6, the cepstral coefficients are the results of a discrete cosine transform (DCT) of the log of the power spectrum, and hence for each of the two spectra considered, the logarithm of the spectrum is optionally considered, and consequently the DCT, resulting in $2^3 = 8$ permutations of parameters.

The power spectrum is a two dimensional matrix, the size of which depends on the length of the window of the STFT and the length of the recording. For a 30 seconds recording sampled at 44,100 Hz, with a window size of 256 samples and no overlap between windows, the power spectrum is a matrix of 128 by $44100 \times 30/256 = 5167$ samples. Since classifiers generally require a single vector of features per sample input, the matrix must be stacked into an array of $5167 \times 128 = 661376$ features. Although modern computers can handle such a large feature space, it is often convenient and certainly much more efficient to reduce these dimensions to a more tractable number, which also mitigates the problem over overfitting the model when there a too few input samples.

Several techniques can be used to downscale these features. Firstly, one can assume that the evolution of the frequencies over time does not contain much information, so that the average strength of a frequency bin is a sufficient indicator of the type of sound in the recording. This appears to be a reasonable assumption for insects, the sounds of which are often generated by scratching, rubbing, tapping and hence do not tend to modulate in frequency. If this assumption holds, the mean of the magnitude of each frequency bin provides sufficient information on the dominant frequencies, while the standard deviation gives an indication of whether the amplitude varies in time or remains constant. The two combined are, moreover, very compact—in the example above, if one were to aggregate the features by their mean and standard deviation, they would result in 2×128 samples. Stowell and Plumbley (2014) argue that for birds, where often recordings contain a large amount of silence and only brief sounded periods, the mean of a frequency bin is diluted and approaches the value of the unsounded periods. In that case, summarising by taking the maximum value may be a better indicator. Therefore, the present approach (see Figure 5.1) summarises by both mean and standard deviation ($\mu + \sigma$) and maximum (*max*).

If, however, the pattern with which frequencies are observed in the recording exposes a character of the singing species, both these approaches fall short. Consider, for example, three tones of a certain frequency; one is present every second for half a second at amplitude x, the second is present every second for a quarter of a second, at amplitude $2x$, and the third constantly present at amplitude $2x$. The *max* would not be able to distinguish between the second and the third, while the mean would not distinguish the first from the second, although the standard deviation may reveal the difference. In a real-world setting where the quality of the recording can very significantly, this shortcoming might be even stronger. Therefore, an alternative solution is to take an additional FFT over each bin of the spectrum to detect the presence of repeating patterns. The discrete values generated by the FFT will here be called *modulation coefficients (mod)* to match the nomenclature in Stowell and Plumbley (2014).

5.1.1 Summarising by Modulation Coefficients

The FFT of the power spectrum captures the magnitude of both the frequency at which an insect is stridulating, and of the frequency of slower, repeating patterns. For example, the dark bush-cricket, common in the New Forest, chirps at a frequency of approximately 1 Hz, but each chirp is composed of three phrases, repeating at about 40 Hz, and in a clean recording each of these reveals a further 600 Hz amplitude modulation (for a clear example refer to Figure 2.8a in Chapter 2). Averaging over time or taking the maximum value of each frequency bin would discard this information. Figure 5.2 shows an artificial signal that exemplifies the power of this feature extraction method. The sample signal is composed of two sine waves, one at 8 kHz and one at 16 kHz, silenced 6 and 10 times per second respectively. The FFT of the mixed signal shows, like the clean carrier wave, the two peak frequencies, and the spectrogram also highlights the alternating chirps. However, these four frequencies are perfectly summarised only in the FFT of the power spectrum, which in Figure 5.2 is capped to the first 50 Hz. On the y-axis, the frequency domain is represented as in the spectrogram, while the x-axis shows the frequency of the repeated patterns (6 and 10 Hz). Finally Figure 5.2 shows the same feature, but represented with one line per frequency band, capped at 300 and 30 Hz respectively.

This artificial signal exemplifies the effectiveness of the FFT of the power spectrum. However, a question remains as to how to represent this feature compactly while remaining maximally general about the features extracted—the method should capture, for example, all the 1, 40 and 600 Hz modulations of the dark bush-cricket's call, and equally whatever other insect is given to the classifier. To do so, this work proposes to uniformly resample the FFT bins in the log-frequency space. This is motivated by the fact that determining the probability with which a bin contains the feature is

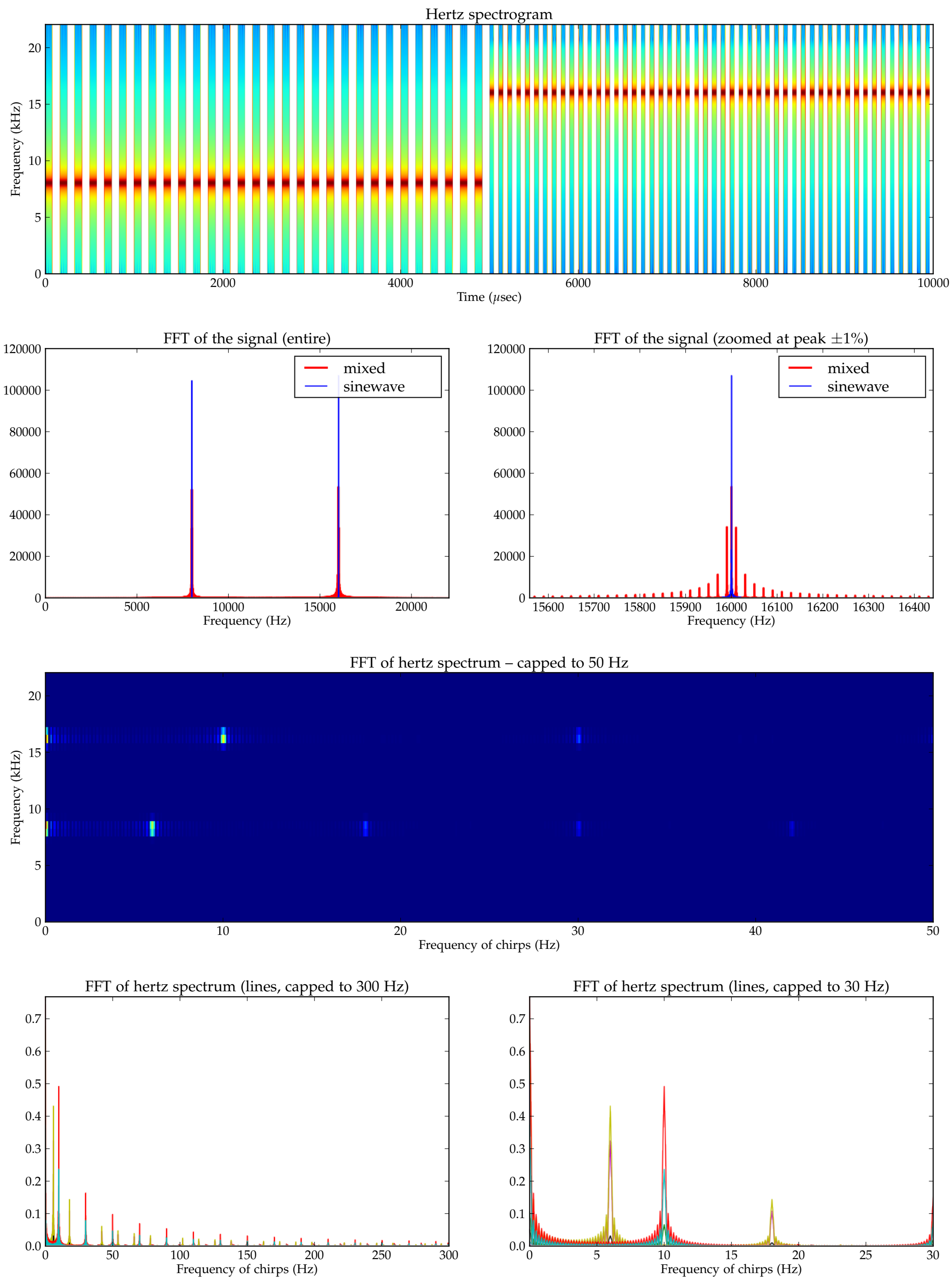


FIGURE 5.2: Modulation coefficients of a sample sine wave. The signal is composed of 5 seconds of a 8 kHz and 5 seconds of a 16 kHz sine waves, repeating at 6 and 10 Hz respectively. The FFT shows a clear peak at the two carrier frequencies, but only the FFT of spectrum shows both the carrier and the modulating frequencies.

equivalent to trying to determine an uninformative prior in Bayesian statistics. Maximum entropy theory says that the correct uninformative prior for a scale parameter, such as a time or a length, is given by a probability distribution proportional to $1/x$ (Jaynes, 2003). Therefore, since it cannot be determined *a priori* whether the repetition of phrases happens at say, 1, 40 or 600 Hz, resampling the FFT bins uniformly in the log-frequency space in the range provided by the power spectrum and the FFT discards minimal information, independently from the scale.

A few parameters determine the final output feature. Firstly, the maximum frequency f_{max} that the modulation coefficients can detect is determined by the window size w of the STFT and the sampling rate f_s of the recording, such that:

$$f_{max} = \frac{f_s}{2 \cdot w}$$

In the example above, with sampling rate of 44,100 Hz and a window size of 256, f_{max} is only $\approx$ 86 Hz, which is insufficient for many of the insects considered. Reducing the window size degrades the accuracy of the spectrum, but tests have shown that a window size of 64 samples, which gives $\approx$ 344 Hz output, performs well for the data sets analysed by this research (see Section 5.4 below). A second parameter is the lowest frequency of interest f_{min}. This is limited by the length of the recording considered, as the lowest component a recording will contain is inversely proportional to its length s, such that:

$$f_{min} = \frac{1}{2 \cdot s}$$

Therefore, a 1-second recording will not contain any component below 0.5 Hz and a 30-second recording will not contain anything below $1/60$ Hz. However, since this value must remain consistent across the feature set given to the classifier, the lowest useful value will be selected, albeit being wasteful in sampling space for shorter recordings. Finally, the number of modulation coefficients for the log-frequency space can be selected arbitrarily, with the lowest number being maximally efficient and the entire range of bins in the input being maximally accurate. Empirical tests show that 48 is a good compromise on a modern desktop computer. Figure 5.3 shows the log-frequency modulation coefficients for a sample call of New Forest cicada, dark bush-cricket and Roesel's bush-cricket side-by-side. The three insects are quite similar in this feature space, though clear differences can still be seen.

In conclusion, the combination of these feature extraction and aggregation methods is used in the system here proposed. The power spectrum is used either raw or logged, cosine-transformed and translated onto the mel scale, as well as each permutation of these transformations. It is then summarised by mean and standard deviation over time, maximum value over time, or with the modulation coefficients sampled on the log-frequency scale describe thus far, leading to a total 24 different feature sets. For each of these, the entire set of recordings is classified both with a decision tree and with a random forest classifier.

5.2 Classification with Decision Trees

Decision trees have been selected as classification algorithm for a number of reasons. First of all, they are commonly used in the literature, and therefore allow for easy comparison of the method here proposed against the state of the art. For example, they are used by Stowell and Plumbley (2014) for bird classification, and since the data sets used by this paper are available to download, the method has been benchmarked against the same data, and the comparison is described below. Secondly, they allow for multi-output classification, that is a scenario where multiple classes (in this case species) are present in the same signal (recording). Thirdly, the learning process to train decision trees is intuitive and can be easily visualised, making them compelling for this exploratory part of the present research. Lastly, many of the limitation of decision trees are mitigated by the usage of an ensemble algorithm, such as a random forest classifier, which is therefore also used for comparison.

5.3 Engineering of the Model

The computational methods used to develop this insect classification model have been carefully designed, and it is therefore worth spending a few words on the engineering of the software. This program aims to allow collaboration and contribution of different developers in the field of wildlife sound classification, so as to provide a common benchmarking platform, and as such it will be released upon completion of this thesis under an open-source licence.

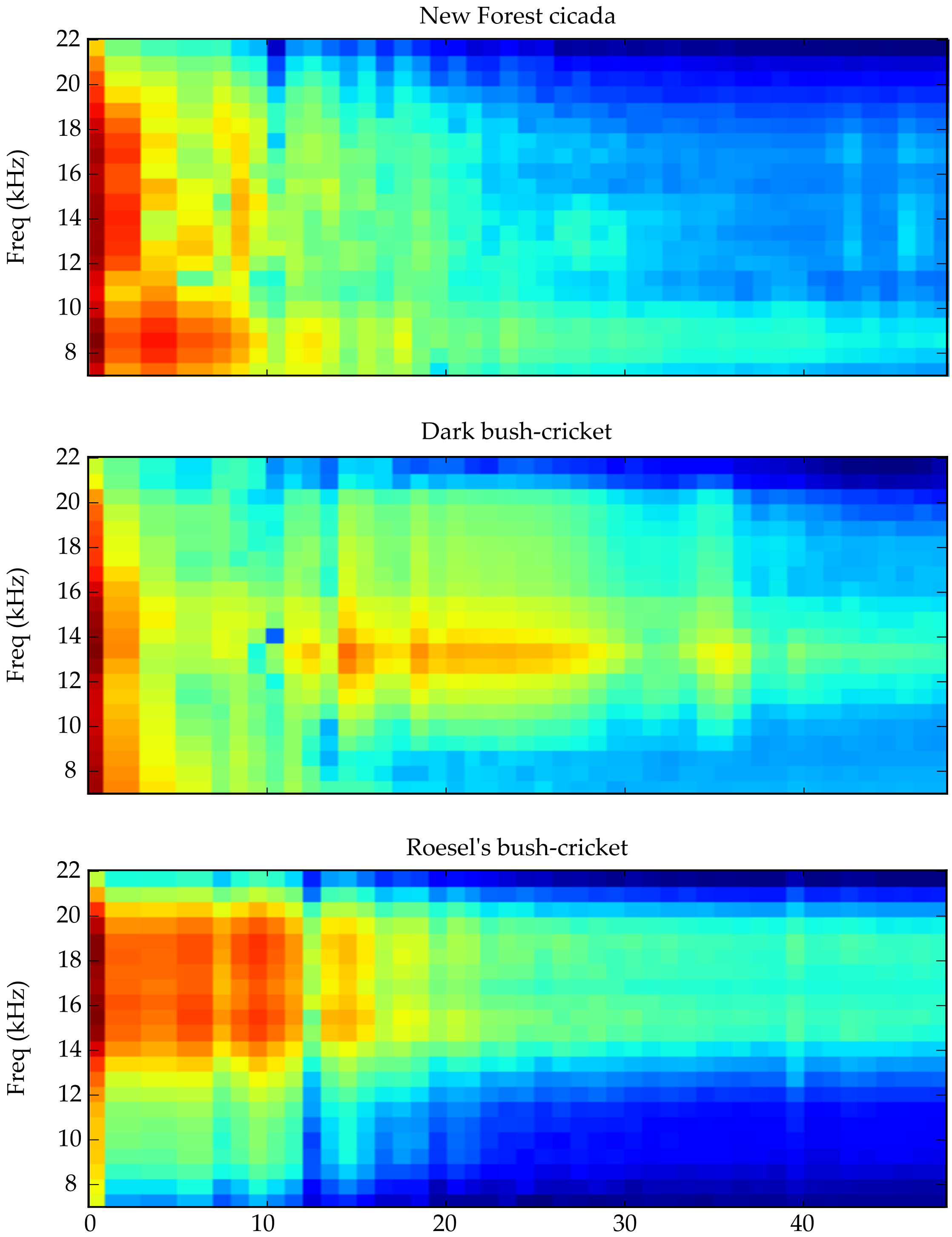


FIGURE 5.3: Comparison of log-frequency modulation coefficients for New Forest cicada, dark bush-cricket and Roesel's bush-cricket, with 7-22 kHz in the spectrum and 48 coefficients.

The software is written in Python and offers the following features:

- Easily extensible object-oriented design
- Separation of model and controller
- Multi-core operation that can process recordings, train and validate results in parallel (through the `multiprocessing` library)
- Serialisation of intermediate steps, to avoid recomputing identical steps
- Storage of results in a `mongo` database
- Serving of results in HTML via an HTTP Server
- Dynamic comparison of different parameters (also saved to database and served in HTML)
- Logging of model operation with different levels of verbosity
- Utilities for Comma-separated values (CSV) formatting of input data
- Separation of settings configuration to customise all parameters in one place.

In order to include a model for a new data set, a developer need only extend the abstract model class and provide an input file in CSV format with the files to classify and their respective label. Optionally, two different files can be provided as test and train sets. A plethora of different features extraction methods are already provided, but more can be implemented by extending the feature extraction routine. Similarly, two classifiers are currently included, but more can be added.

The classification is built on the powerful and widely used `scikits-learn` machine learning library (Pedregosa et al., 2011). The results are outputted in colour on the console and in the HTML reports, and served through the lightweight `flask` web framework. Five data set models are provided as examples.

Additionally, to further develop the collaborative approach to insect classification that this research wants to promote, an initial version of the model has also been implemented on Microsoft Azure Machine Learning (*Azure ML*[1]), a novel cloud-based data

[1] http://azure.microsoft.com/en-us/services/machine-learning

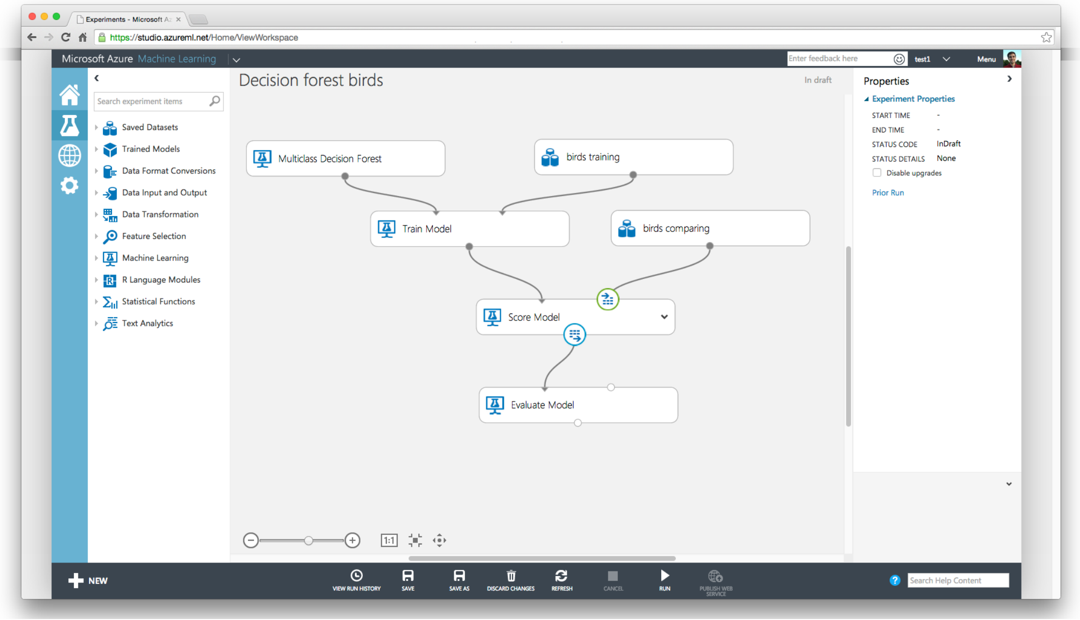


Figure 5.4: Wildlife classification system on Azure ML, tested on an birds data set.

mining and machine learning platform developed by Microsoft. This allows developers to collaborate remotely on the same model, and execute it on powerful pay-per-use servers, while also sharing large data sets. The platform is currently in its infancy, and it charges for usage and computational time, but it is actively developed and offers free plans for academic purposes. The interface of the system with an early version of the model proposed is exemplified in Figure 5.4, where it is tested on a bird calls data set.

5.4 Evaluation on Large Scale Data Sets

Once again, the method proposed is evaluated against three very different data sets, each with their own benefits and drawbacks. The limitations are outlined below, but can be broadly grouped into *a*) the quality of the recordings *b*) the number of samples per species *c*) the number of different devices used for recording. A more detailed description of these data sets with a list of species is reported in Appendix C.

The execution of the model proposed observes the following procedure. Each recording is re-sampled at 44,100 Hz, where necessary. The power spectrum is then extracted with an STFT window size of 64 samples and with no overlap, and the lower frequency bands are discarded to filter out background noise, with cut-off frequency dependent on the data set. The volume of each recording is normalised by dividing the power spectrum by the root-mean-square (RMS) amplitude. It is then optionally transposed on the mel spectrum with 40 mel filters. Subsequently it is also optionally logged and cosine-transformed to output 13 coefficients. The features extracted when the spectrum is logged and cosine-transformed correspond to the cepstral coefficients.

As described above, the spectrum is then summarised in one of three ways: either with the mean and standard deviation, providing 64 features (32 each for mean and standard deviation, corresponding to half the STFT window size) or less, in case the lower components are discarded; with the maximum of each frequency bin, providing again 32 features or less; or finally through the extraction of modulation coefficients. For the log-frequency sampling of the FFT, 48 bins are used, minimum and maximum frequencies of $1/60$ Hz and 344 Hz respectively, as discusses in the previous section. The first modulation coefficient corresponds to the DC component of the signal, and therefore resembles the values of the mean value across time, as exemplified later in Figures 5.5 and 5.6.

If only one set of data has been provided, this is split into two parts for training and testing (a method called *hold-out*). The results reported here use a 50% split, but 66% for training is also commonly used. The accuracy of the trained model is assessed with the F_1 score, described in Chapter 3, and by measuring the area under the receiver operating characteristic (ROC) curve. The ROC curve is a well-established technique that shows the rate of true positives against the rate of false positives. This measure is used in both the published literature and in wildlife sound classification challenges (for example the NIPS 2013 multi-label bird species classification challenge[2]).

Alternatively, a robust way of assessing the model trained is by using *cross-validation*. This method divides the training set in a number of chunks (or *folds*), for example 10. Nine of these are used for training, and one for testing, and in turn each fold is kept aside for testing, while the rest is used for training, which reduces the variance of the estimate.

[2] NIPS4B is assessed on kaggle at https://www.kaggle.com/c/multilabel-bird-species-classification-nips2013

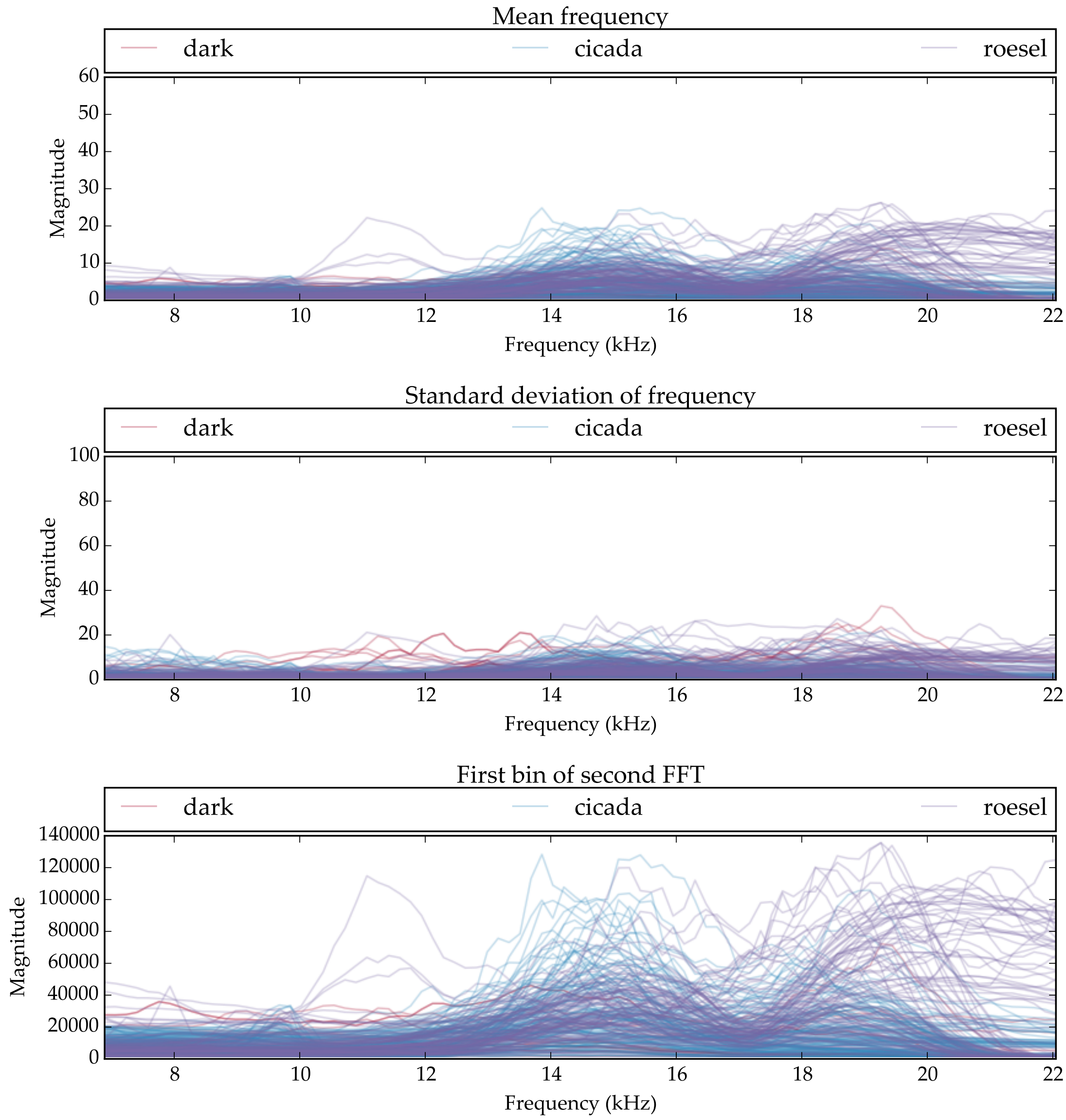


FIGURE 5.5: Mean, standard deviation and first modulation coefficient of the signal between 7 and 22.05 kHz for the three insects in the data set.

The results of the validation for each data set are reported below with F_1 score, ROC AUC score and score of a 10-fold cross validation. Unless specified otherwise, the parameters used are those described thus far.

5.4.1 Cicada Hunt Recordings

The first data set (`NFcrowd`) is a subset of 54 out of the 235 recordings collected by users in the New Forest and in Slovenia in the first season of deployment of the mobile system, representing three species: New Forest cicada, dark bush-cricket and

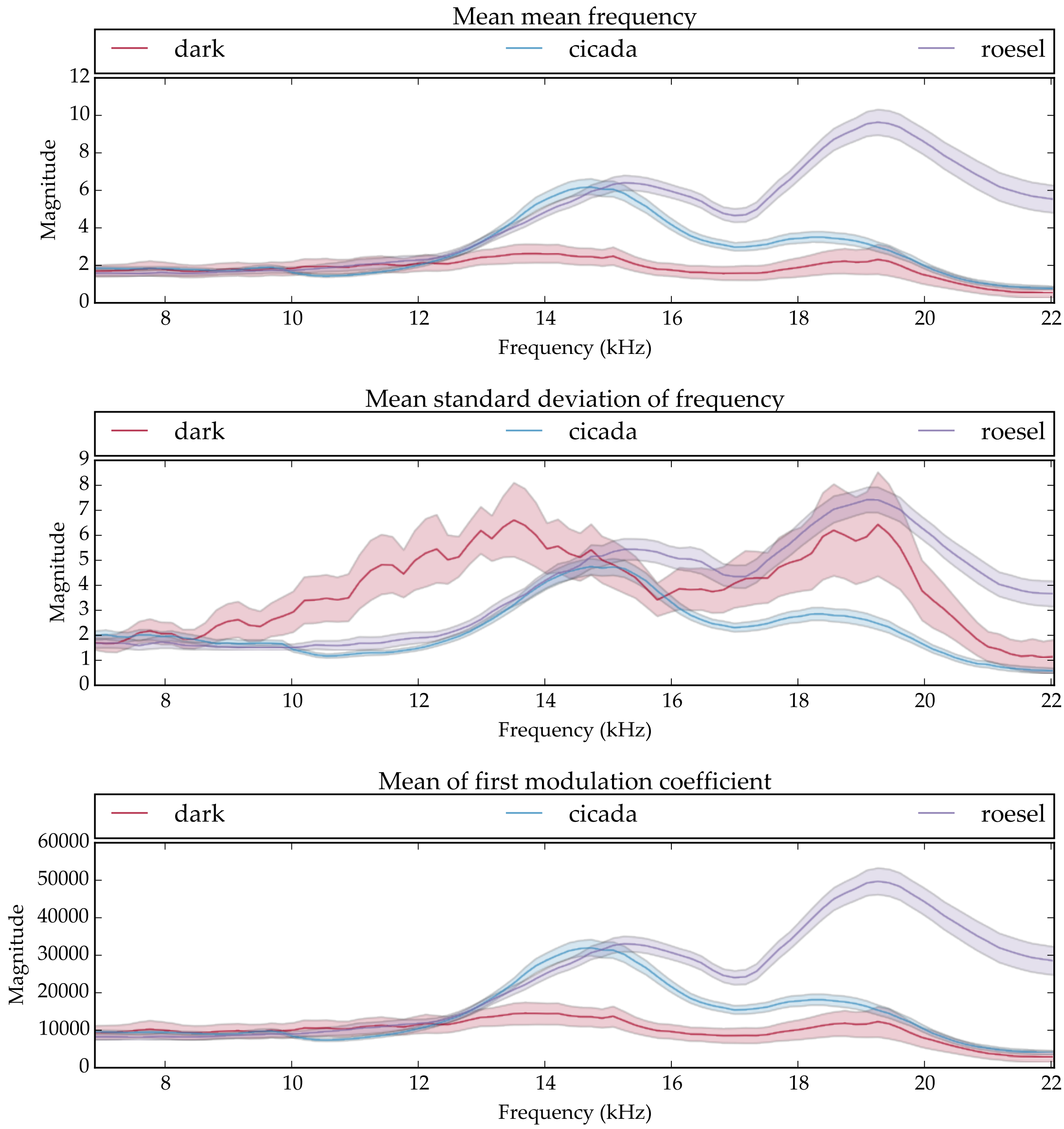


FIGURE 5.6: Mean of mean, standard deviation and first modulation coefficient of the signal between 7 and 22.05 kHz for the three insects in the data set.

Roesel's bush-cricket, with 18 samples each. The limit of 18 is imposed by the number of recordings of the least occurring species, the dark bush-cricket, to which the other species were matched. This removes one aspect of bias for the classifier, which may otherwise learn the frequency of occurrence of one species and score a good result by simply always selecting the most frequently occurring one.

The importance of this data sets lies chiefly in the quality of the recordings. In fact it represents real, crowdsourced data, and is therefore the collection that most closely matches the purpose of this investigation. This not only means that the recording device is different, but also that many forms of background noise are present, that potentially parts of the call are missing and that the singing insect is not cleanly centred in the recording, as it happens in most library recordings that are manually processed. It also constitutes the same set for which the cicada detection algorithm (CDA), presented in Chapter 3, was developed and on which is was calibrated, allowing for a direct comparison. On the contrary, the small number of species and of samples constitute its main limitations.

In this evaluation, the frequencies below 7 kHz have been discarded. This is due to the fact that all cicada recordings, taken with similar devices and in a similar location, presented an artifact around 6 kHz that was producing a bias in the classifier. The learning algorithm would therefore learn the artifact instead of the song of the insect, scoring very good results. Removing these frequencies solved the problem, yet highlighted an unpredicted limitation of the data set. Figures 5.5 and 5.6 show the mean, standard deviation and first modulation coefficient of all the recordings in the set, the former in the form of one line per recording, the latter as the average across all the recordings, surrounded by the standard deviation. The model was also trained with the first modulation coefficient only, which is very similar to the mean as discussed above, as well as with linear modulation coefficients. The combination of all three transforms and five aggregation methods is depicted in Figure 5.7, where all the powers have been log-scaled for clarity.

Table 5.1 presents the 20 permutation of parameters that performed best at classifying this data set, sorted by their F_1 score. The log-frequency modulation coefficients (`logmod`) perform generally better than mean and standard deviation ($\mu + \sigma$), and taking the logarithm of the values also has a generally positive effect. It is less clear whether the mel or the Hertz spectrum performs better, with the mel scoring slightly higher results. The reason for this is that, with the lower 7 kHz removed and the three species being fairly similar in frequency, the alleged drawbacks of using the mel spectrum are not noticeable. As expected, no benefit has been observed in the first coefficient as opposed to mean and standard deviation, and equally for linear-frequency coefficients as opposed to log-frequency ones; therefore for the sake of clarity these results have been omitted from the table.

A further comparison can be drawn also with the previous cicada detection algorithm

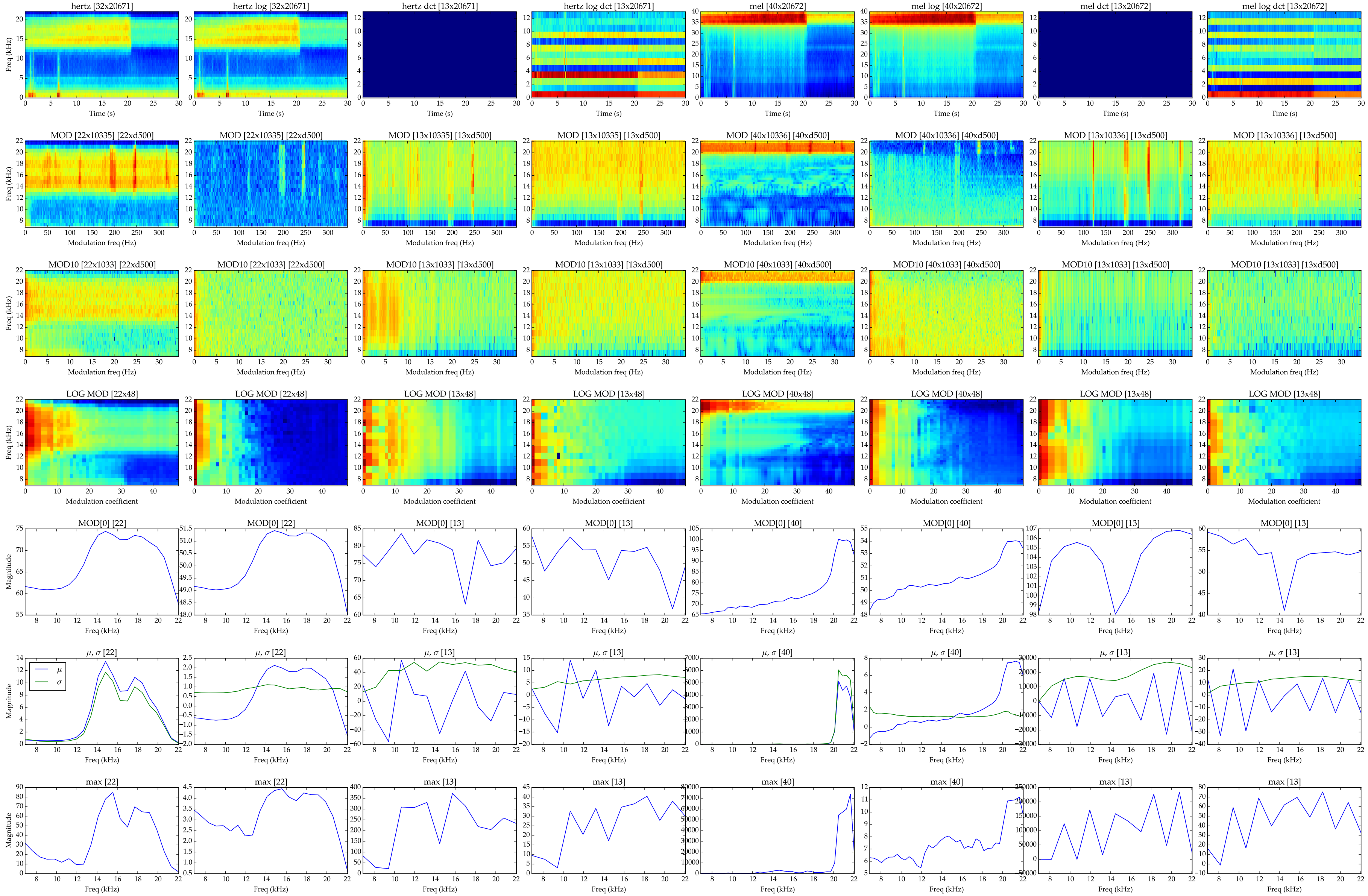


FIGURE 5.7: Feature extraction of a *Cicadetta montana* with all permutations of parameters. The columns correspond to Hertz/mel spectrum, logged and not logged, and DCT-transformed or not. The rows are the summary features, linear modulation coefficients (MOD), only the first 10 coefficients (MOD10), on the log-frequency scale (LOG MOD), only the first coefficient (MOD0), mean and standard deviation (μ, σ), and maxium (max).

Classifier	Mel	Log	DCT	MOD	μ	σ	Max	Agg	Features	CV score	F1 score	ROC AUC	Correct
decisiontree	•	•	•	•				logmod	624	0.800	0.927	0.944	92.593%
randomforest	•	•		•				logmod	1920	0.900	0.926	0.980	92.593%
randomforest	•	•	•	•				logmod	624	0.833	0.888	0.975	88.889%
decisiontree		•		•				logmod	1056	0.867	0.810	0.861	81.481%
randomforest		•	•	•				logmod	624	0.733	0.809	0.914	81.481%
randomforest	•	•	•				•	max	13	0.700	0.769	0.864	77.778%
decisiontree	•			•				logmod	1920	0.583	0.743	0.806	74.074%
randomforest		•		•				logmod	1056	0.900	0.733	0.908	74.074%
randomforest	•	•			•	•		μ+σ	80	0.700	0.710	0.737	70.370%
randomforest		•			•	•		μ+σ	44	0.700	0.705	0.861	70.370%
decisiontree	•		•	•				logmod	624	0.767	0.704	0.778	70.370%
decisiontree			•	•				logmod	624	0.700	0.701	0.778	70.370%
randomforest	•				•	•		μ+σ	80	0.633	0.693	0.851	70.370%
randomforest		•	•		•	•		μ+σ	26	0.617	0.691	0.885	70.370%
decisiontree	•	•		•				logmod	1920	0.933	0.690	0.778	70.370%
decisiontree	•		•		•	•		μ+σ	26	0.617	0.668	0.750	66.667%
decisiontree	•						•	max	40	0.650	0.668	0.750	66.667%
decisiontree	•	•					•	max	40	0.650	0.668	0.750	66.667%
randomforest					•	•		μ+σ	44	0.800	0.666	0.853	66.667%
decisiontree					•	•		μ+σ	44	0.700	0.659	0.750	66.667%

TABLE 5.1: Summary of results for the `NFcrowd` data set.

(CDA) model from Chapter 3, which classified the same data set (although on all 235 samples). The F_1 scored by the CDA reached a value of 0.82, while the model here described with the optimal set of parameters scores $F_1 = 0.92$. The difference is notable and is to be attributed to the hand-tuning of parameters performed in the CDA. In fact, with the little data for which that method was designed, the hand-tuning was not only beneficial but necessary (no phone model had been tested, and very few cicada recordings were available to the authors), while with a good set of samples as in the present evaluation the model is able to learn the necessary parameters. In the future, a further comparison may use this data to learn the parameters of the HMM for the CDA, for example through an expectation-maximisation (EM) algorithm such as the Baum–Welch algorithm, therefore adapting the parameters to the data.

5.4.2 British Orthoptera Recordings

While the `NFcrowd` data set provides a good comparison to the existing system, it does not assess the model against the purpose it was built for, that is many-species classification. To address this shortcoming, another data set containing all 28 known species of Orthoptera in the UK is used here, and hereon referred to as `UKorthoptera`. The set contains 70 recordings, with two to four recordings per species, which for certain species also represent different types of calls. The recordings are very clean and of good quality, which simplifies classification but can also be considered a limitation for the present scenario. The small number of recording per species constitutes the other main limitation of this data set.

Given the small amount of data, the recordings are segmented into 2 second chunks that are treated as individual samples. Despite dramatically improving the accuracy, this trick has two side-effects, namely that a) the model could be tested and trained on parts of the same recording, giving it a slightly positive bias, and b) that unsounded periods of a recording may happen to be treated as samples for an insect, giving the model a slightly adverse bias.

Results in Table 5.2 show good overall performance, given the large number of classes and the small number of samples, with F_1 score just under 0.72. While the best performing method uses log-frequency modulation coefficients, it is not clear that these are significantly better than aggregating by mean and standard deviation, as the latter

	1	3	4	5	6	7	8	10	12	13	15	16	18	19	20	21	22	26	29	30	31	32	35	37	40	42	44	45
1	.92	0	0	0	0	0	0	0	.08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	.89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.06	0	.06	0	0
4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	.89	0	0	0	0	0	0	0	0	.05	0	0	0	0	0	0	0	0	0	0	0	.05	0	0
7	0	0	0	0	0	.90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.10	0	0	0	0
8	0	0	0	0	0	0	.95	0	0	0	0	.05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	.05	0	0	0	0	0	0	0	0	.82	0	0	0	0	0	0	0	0	0	0	0	.09	0	.05	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	.08	.92	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	.50	0	0	0	0	0	0	0	0	0	.50	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.75	0	0	0	0	0	0	0	0	.25
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.50	0	.50	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.08	0	.92	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	.14	0	0	0	0	0	0	0	0	.14	0	.14	.43	0	.14	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.29	0	0	0	.71	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.89	0	.11	0	0
40	0	.33	0	0	0	0	0	0	.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.33	0	0	0
42	0	0	0	0	.15	0	0	0	0	0	0	0	0	0	0	0	0	.08	0	0	0	0	0	.38	0	.38	0	0
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
45	0	0	0	0	0	0	0	0	0	0	.75	0	0	0	0	0	0	0	.25	0	0	0	0	0	0	0	0	0

FIGURE 5.8: Confusion matrix of `UKorthoptera` data set. The labels correspond to the Dutch orthoptera atlas identification numbers, as reported in the table below.

ID	Latin name	English name	ID	Latin name	English name
1	Phaneroptera falcata	Sickle-bearing bush-cricket	20	Nemobius sylvestris	Wood cricket
3	Leptophyes punctatissima	Speckled bush-cricket	21	Oecanthus pellucens	Tree cricket
4	Meconema thalassinum	Oak bush-cricket	22	Gryllotalpa gryllotalpa	Mole cricket
5	Meconema meridionale	Southern Oak bush-cricket	26	Stethophyma grossum	Large marsh grasshopper
6	Conocephalus dorsalis	Short-winged conehead	29	Stenobothrus lineatus	Stripe-winged grasshopper
7	Conocephalus discolor	Long-winged conehead	30	Stenobothrus stigmaticus	Lesser Mottled Grasshopper
8	Tettigonia viridissima	Great green bush-cricket	31	Omocestus viridulus	Common green grasshopper
10	Decticus verrucivorus	Wartbiter	32	Omocestus rufipes	Woodland grasshopper
12	Platycleis albopunctata	Grey bush-cricket	35	Chorthippus vagans	Heath grasshopper
13	Metrioptera brachyptera	Bog bush-cricket	37	Chorthippus brunneus	Field grasshopper
15	Metrioptera roeselii	Roesel's bush-cricket	40	Chorthippus albomarginatus	Lesser marsh grasshopper
16	Pholidoptera griseoaptera	Dark bush-cricket	42	Chorthippus parallelus	Meadow grasshopper
18	Gryllus campestris	Field cricket	44	Myrmeleotettix maculatus	Mottled grasshopper
19	Acheta domesticus	House cricket	45	Gomphocerippus rufus	Rufous grasshopper

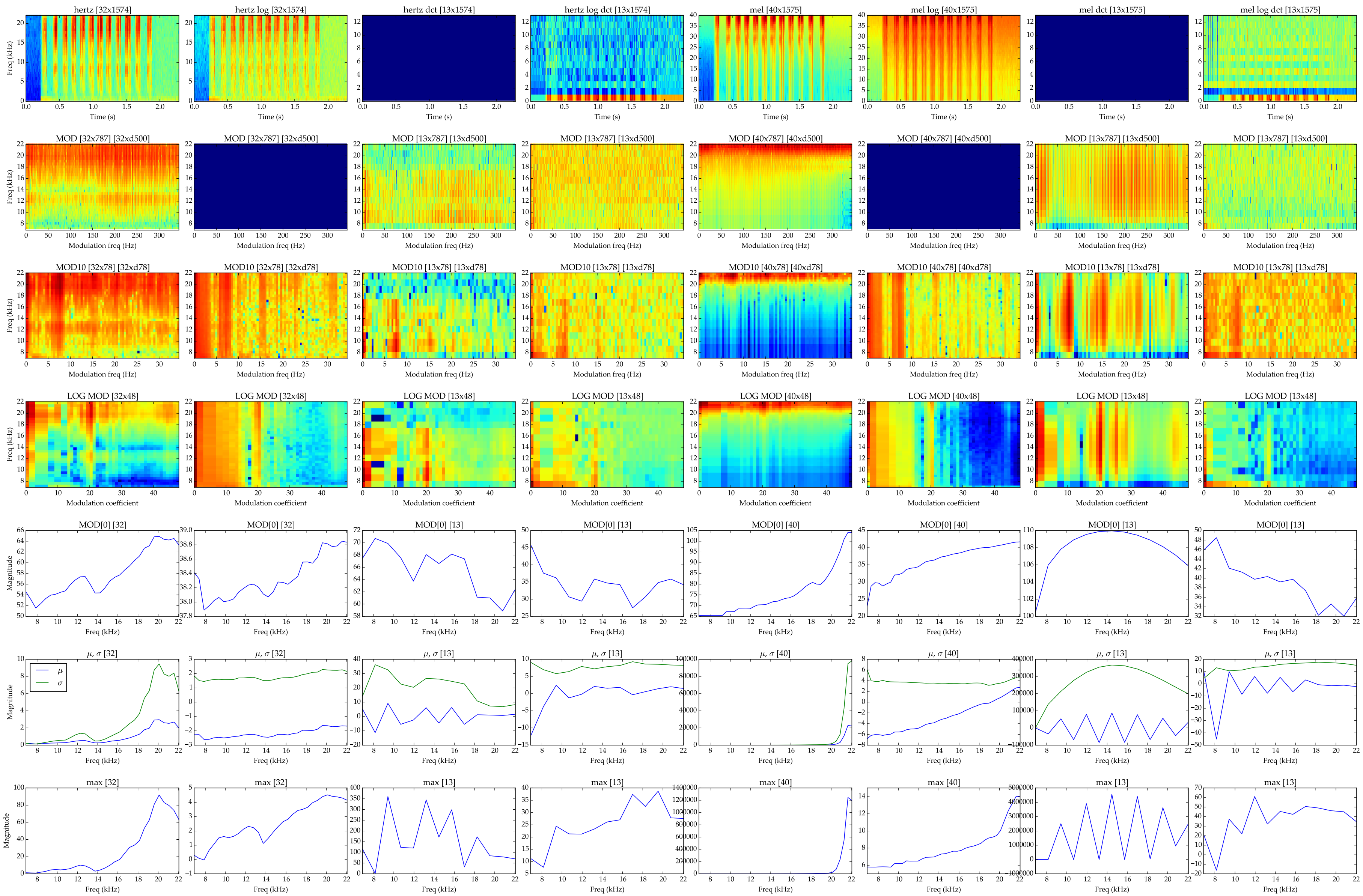


FIGURE 5.9: Feature extraction of a *Sickle-bearing bush-cricket* with all permutations of parameters. The columns correspond to Hertz/mel spectrum, logged and not logged, and DCT-transformed or not. The rows are the summary features, linear modulation coefficients (`MOD`), only the first 10 coefficients (`MOD10`), on the log-frequency scale (`LOG MOD`), only the first coefficient (`MOD0`), mean and standard deviation (μ, σ), and maxium (`max`).

Classifier	Mel	Log	DCT	MOD	μ	σ	Max	Agg	Features	CV score	F1 score	ROC AUC	Correct
randomforest		•	•		•	•		μ+σ	26		0.864	0.497	87.702%
randomforest	•	•	•		•	•		μ+σ	26		0.864	0.497	88.026%
randomforest			•		•	•		μ+σ	26		0.857	0.521	87.379%
randomforest					•	•		μ+σ	64		0.856	0.492	87.055%
decisiontree		•	•		•	•		μ+σ	26		0.847	0.499	84.790%
randomforest				•				logmod	1536		0.838	0.509	85.761%
randomforest							•	max	32		0.825	0.500	83.819%
randomforest		•					•	max	32		0.825	0.499	83.819%
randomforest	•		•		•	•		μ+σ	26		0.814	0.512	82.201%
randomforest		•			•	•		μ+σ	64		0.809	0.500	82.524%
decisiontree					•	•		μ+σ	64		0.804	0.499	80.259%
randomforest			•				•	max	13		0.802	0.513	81.877%
decisiontree			•		•	•		μ+σ	26		0.793	0.500	80.906%
decisiontree	•				•	•		μ+σ	80		0.793	0.505	79.612%
randomforest	•				•	•		μ+σ	80		0.790	0.489	80.259%
randomforest	•		•	•				logmod	624		0.783	0.505	79.935%
randomforest	•			•				logmod	1920		0.781	0.490	79.288%
decisiontree	•		•		•	•		μ+σ	26		0.780	0.502	78.964%
decisiontree	•	•	•		•	•		μ+σ	26		0.769	0.506	77.023%
decisiontree		•			•	•		μ+σ	64		0.764	0.507	77.023%

TABLE 5.2: Summary of results for the `UKorthoptera` data set.

score very good results, with better computational efficiency. However, the random forest ensemble mostly outperforms the decision tree, which highlights the instability of the latter to training with little data.

5.4.3 NIPS4B Bird Classification Challenge

Much of the literature on insect sound classification assesses algorithms against a specific set of recordings. If these data sets are not made available, it becomes hard to benchmark the methods used against each other. Recently, in order to mitigate this problem, the community has organised challenges in which competitors are given a labelled set of recordings to train their algorithms against, and an unknown set of test data. Results against the test data are submitted to a common platform, which assesses the performance on the results obtained.

Therefore, one such data set was required to evaluate the approach proposed in this chapter against the state of the art. Given that these competitions are not yet common for insect classification and that this approach is modelled on the bird sounds recognition literature, a data set of bird calls was selected, provided by the Neural Information Processing Scaled for Bioacoustics (NIPS4B) bird song competition. Competitors are given a set of 678 recordings with labels, containing one *or more* of 87 species of birds. They are also given an additional 1000 recordings with no labels, which they have to classify and the results of which are submitted by each entrant and compared to others. Hence, no ground truth is given for these, but upon submission an overall score, in terms of ROC AUC score, and a ranking are returned to the competitor.

Therefore, the most prominent advantages of this data set are a) an easy comparison against other implementations, b) a large number of samples and c) the presence of separate, unlabelled data for testing, which ensures that no information can be learnt from scoring the model (and therefore eliminates any such bias). The main shortcoming is of course the fact that the species classified are birds and not insects, so the device introduced to model insect calls may not perform as well as existing implementations.

Consequently, two sets of results are presented. The first one, reported in Table 5.3, shows once again very good overall performance, this time sorted by ROC AUC score to match the competition's assessment. As expected, the modulation coefficients do not perform better in this scenario, as they are tailored to strong, regular chirps that are not often present in birds. It is interesting to note how here the random forest classifier consistently outperforms the simple decision tree. The second assessment of this data set is the comparison with other competitors in the challenge. This is extremely beneficial to test that the implementation matches what is expected, and to avoid any voluntary or involuntary tampering of the results. The method here presented scores a ROC AUC on the 1000 unlabelled samples of 0.87, similar to Stowell and Plumbley (2014) on which this system is modelled. This confirms the integrity of the results, and gives confidence of the validity of the model in the global research community.

5.4.4 Additional Data Sets

In order to address the limitation of sample size of the insect data sets here presented, a further set of all insect and related sounds is currently being collected, and already amounts to over 14 GB of recordings. This has being provided by the British Library's Wildlife Recording Scheme and contains samples for each British Orthoptera species, together with a small number of birds and frogs. The processing of this data set, in which many of the insects are "introduced" by a voice that describes the environment and credits the authors, is ongoing effort and its scope lies beyond that of this thesis. Initial results on a subset of these files show promising results through the system described in this chapter.

5.5 Summary

This chapter has presented a novel insect classification algorithm based on a random forest classifier, which extracts a discrete set of log-frequency modulation coefficients from the power spectrum of a recording and uses them as features for classification. The model shows excellent performance on three insect species commonly found in the New Forest, and it classifies a large data sets of 87 birds species with accuracy comparable to the state-of-the-art in the field. It also achieves a high score on a small data set of 28 insect species, though this limitation calls for a larger data base of insect

Classifier	Mel	Log	DCT	MOD	μ	σ	Max	Agg	Features	CV score	F1 score	ROC AUC	Correct
randomforest			•		•	•		μ+σ	26	0.224	0.159	0.788	11.034%
randomforest	•	•	•		•	•		μ+σ	26	0.183	0.108	0.787	7.931%
randomforest		•	•		•	•		μ+σ	26	0.202	0.146	0.784	9.655%
randomforest					•	•		μ+σ	64	0.215	0.151	0.783	10.517%
randomforest	•				•	•		μ+σ	80	0.191	0.142	0.782	10.862%
randomforest	•		•		•	•		μ+σ	26	0.195	0.156	0.775	10.862%
randomforest		•			•	•		μ+σ	64	0.198	0.090	0.770	5.862%
randomforest	•	•			•	•		μ+σ	80	0.173	0.100	0.766	7.241%
randomforest							•	max	32	0.179	0.081	0.750	5.345%
randomforest		•					•	max	32	0.179	0.081	0.748	5.345%
randomforest	•			•				logmod	1920	0.160	0.078	0.730	5.517%
randomforest	•	•					•	max	40	0.156	0.037	0.726	2.414%
randomforest				•				logmod	1536	0.160	0.044	0.726	2.586%
randomforest	•						•	max	40	0.156	0.035	0.724	2.241%
randomforest	•		•	•				logmod	624	0.159	0.035	0.722	2.586%
randomforest	•	•		•				logmod	1920	0.154	0.026	0.680	1.724%
randomforest			•				•	max	13	0.162	0.074	0.677	5.000%
randomforest		•	•				•	max	13	0.157	0.055	0.663	3.793%
randomforest	•		•				•	max	13	0.150	0.026	0.660	1.897%
randomforest			•	•				logmod	624	0.157	0.015	0.659	0.862%

TABLE 5.3: Summary of results for the `nips4b` data set.

recordings, which is being collected at the time of writing and will continue to grow over the years. The algorithm surpasses the previously proposed cicada detection algorithm in terms of accuracy and scalability, though it remains more computationally expensive. The entire sound recording classification system is also currently being deployed in an Orthoptera reporting app, developed by the Orthoptera Recording Scheme at the Centre for Ecology and Hydrology, which will provide an unprecedented set of Orthoptera recordings, similar to that of the New Forest Cicada Project, but on a national and international scale.

Chapter 6

Conclusions and Future Work

This thesis documented novel research on the application of citizen science to biodiversity monitoring. Surveying animal species through the sound they produce has long been an established technique, especially for those animals that emit a very distinctive call, such as bats, birds and insects. The involvement of the general public in this effort has been effective in past years, but the introduction of new technologies recently brought citizen involvement to a new level. Smartphones have been integral to this shift; the large number of sensors with which they are equipped, their widespread presence among the population and their ease of use make them the perfect vehicle for people's involvement in scientific research. This work exploits these emerging technologies to rediscover the New Forest cicada, a highly endangered insect that produces a high-pitched song, difficult to hear for humans but easily detected by mobile phones. The conception of a machine learning algorithm that would detect the presence of this insect in real time on a mobile device and the implementation of the infrastructure apt to search for it are the core objectives of this research, which eventually aims to generalise this method to any application of citizen biodiversity monitoring.

More specifically, this thesis has reviewed in Chapter 2 the literature in the two key areas for this research: citizen science and bioacoustics. The former has been described in the wider context of crowdsourcing applications, collecting a range of examples from which experience has been drawn. The latter has been discussed in relation to automated identification of taxa, of which a specific area is aimed to be advanced with

FIGURE 6.1: The app's main screen on Android and iOS.

this work. Chapter 2 has also reviewed tools, techniques and algorithms commonly used in bioacoustics, as well as introducing the ecology of the New Forest cicada, the knowledge of which is essential to an effective search.

Chapter 3 has then presented a novel approach to efficiently detecting the presence of this insect, built on an HMM-based approach. In order to devise this method, the call of the cicada and of few similar insects have been analysed, and two strong features where observed in the call of the former. A constant 13.5 kHz-centred signal in the frequency domain, and an increasing intensity of the sound in the amplitude domain, with a sudden stop after 30–40 seconds. A weaker feature also consists of an 8 ms and a 16 ms pattern that repeat in the call, but these only become noticeable in high quality recordings. Results of the evaluation of this method against a data set of over 200 smartphone recordings of the New Forest cicada, the Roesel's bush-cricket and the dark bush-cricket, collected by the research team in Slovenia and by citizen scientists in the New Forest show high accuracy, with an F_1 score of 0.82 for the cicada call, despite the variable quality of the recordings.

Chapter 4 then illustrated how this novel approach has been ported to a mobile system, with a client for Android and iOS devices and a server to collect user observations. These form the basis of a large citizen science endeavour, called the *New Forest Cicada Project*, which has been fully deployed and can boast 3000 users collecting thousands of observations. With this system, citizens have already surveyed the New Forest during two seasons. At present, no specimen has been found, but the correct operation of the system has been verified by trialling the app in Slovenia, where the species was previously found. There, the client has been used to collect hundreds of call samples, which have later been used to increase the robustness of the system against other insects. In addition, the chapter compared 17 of the most widespread mobile devices in terms of the sensitivity of their microphone and their ability to detect the cicada.

Finally, Chapter 5 investigated an alternative approach to automated insect sounds classification, based on the extraction of log-frequency cepstral coefficients and the classification with decision forests and random forest ensembles. The former is an innovative, compact representation of the spectrum that exploits the strong frequency components in insect calls and the repetition of phrases, while assuming no prior knowledge of the calls in analysis. It therefore scales better than then HMM-based approach to a many-species scenario, at the cost of higher computational complexity. This algorithm is in the process of being included in an Orthoptera recording app developed by the Centre for Ecology and Hydrology.

In conclusion, the research presented in this thesis has constructed the tools, both methodological and practical, to detect the presence of an endangered insect with the involvement of the general public, and has proposed methods to apply the same methodology to the broader monitoring of wildlife with sound. The publication of this research in international conferences and journals, the awards received and most importantly the strong participation of citizen scientists in this endeavour vouch for the need for similar efforts to be embraced for the protection of the natural environment.

6.1 Future Work

The effectiveness of the methods proposed call for further research in the areas explored here, as well as for the practical implementation and deployment of similar tools in different domains. A few avenues for future work are discussed below.

For the purposes of the present and other smartphone-based acoustic projects, further extensive tests may be performed on mobile devices and the frequency response of their microphone. A database that collects this type of information is not yet present, to the best of the author's knowledge, and may be valuable to similar projects within and outside the biodiversity context. This may even be achieved through a collaborative effort, where users can send recordings of a test sound that are then centrally collected and analysed.

With the infrastructure of the mobile client in place and a community of citizen scientists already interested in the subject, further research may also be performed on incentive mechanisms that can motivate people's participation. This can be obtained through the involvement of local schools, businesses, tourist information centres and park authorities who may be interested in attracting customers while raising public awareness on the issues of biodiversity monitoring. A local café may offer, for example, a free beverage to the user who submitted the most number of reports in a day, so that the business and the citizen science project both benefit from this mutual interaction. Gamification, that is the engagement of users through a game in a non-game context, is also an appealing route as *a*) hundreds of millions of people worldwide play electronic games (McGonigal, 2011), making them a very large community to appeal to and *b*) the computational sustainability scenario attracts much public attention these days when more and more emphasis is put on safeguarding the environment.

Moreover, this work has already collected thousands of reports worldwide, the analysis of which lies beyond the scope of this thesis, but providing an excellent starting point for future work. Many observation of different insects have in fact been submitted to the system, some mistakenly classified as one of the known insects (for example the periodical cicadas (*Magicicadae*) recorded in America, where many reports were submitted); other were submitted in places where the *Cicadetta montana* is present, such as Portugal and Spain. The analysis of the existing reports and the engagement with local communities then proves a matter of entomological interest.

The system, however, should not be restricted to smartphones alone. The entomologists who survey the New Forest for the presence of the cicada often return to the same sites where the insect was historically found. The inspection of these sites is less suitable to citizen scientists, who may gather in large numbers and disrupt the habitat. At the same time, the few entomologists struggle to cover all sites in the sunny, warm days in which the insect sings. A hardware-based, standalone cicada detector would therefore be desirable to monitor the sites at regular intervals. The device should have a long-lasting battery, potentially harvesting energy from the environment, for example with a solar panel. It should be low-cost to be widely deployed and to allow for the risk of being damaged by the elements. Ideally, it may be networked to report any relevant observation to a base station, but in its simplest form it may save data to a memory card, which can be occasionally collected and analysed.

Furthermore, the same techniques and algorithm used in this research may be applied outside the biodiversity monitoring domain by using citizen science and smartphones to monitor other environmental factors using sound. Example applications would be the monitoring of soundscapes and tranquillity around urban parks or the detection of faults in electrical equipment, such as alternators, that emit a distinctive noise when close to failure.

The distribution of the source code produced by this research—in particular the mobile app and the classification algorithms—under an open-source licence proposed upon completion of this programme should further facilitate the employment of similar methods in the wider contexts and scenarios outline above.

Appendix A

Application Programming Interfaces

A.1 App to Server Back End API

A.1.1 Fields

`$SERVER=http://newforestcicada.info/api`

A.1.1.1 Methods

`GET $SERVER/get_auth_token`

- return

  `auth_token`

  - type: `char(64)`
  - description: unique authentication token from the server

`POST $SERVER/upload`

- params:

  `auth_token`:

- required
- type: `char(64)`
- description: unique authentication token

`wav_file`

- optional
- type: `file` (only WAV or FLAC)
- description: sound recording

`recording_titestamp`

- required
- type: `datetime`
- description: timestamp of the observation time (NOT the uploading time)

`latitude`

- optional
- type: `float`
- description: latitude of the observation

`longitude`

- optional
- type: `float`
- description: longitude of the observation

`photo_file`

- optional
- type: `file` (only PNG or JPG)
- description: photograph of the observation.

`description`

- optional
- text

- description: anything that the user wants to add to the observation

`observation_id`

- optional
- type: `int`
- description: database ID, only needed to modify an existing entry

- return

  `JSON array {exit_status, observation_id}`

  `exit_status`

  - type: ‘int‘:

    `0:  OK`

    >0: error

  `observation_id`

  - type: `int`
  - description: database ID of the entry

A.2 Internal App’s Front-end to Back-end API

A.2.0.2 Method Summary

```
boolean execute(java.lang.String~action,
org.json.JSONArray~args,
org.apache.cordova.api.CallbackContext~callbackContext)}
```

The only method ever to be called from the javascript interface.

`double getAmplitude()`

Get a value of the amplitude from the microphone.

`double getCicada()`

Get the estimate of the presence of the cicada, in a float value between 0 and 1.

`double[] getFrequencies()`

Get array of frequency magnitudes, one per frequency bin.

`org.json.JSONArray getReport(int id)`

Retrieve a survey report.

`void initialiseDetector(org.apache.cordova.api.CallbackContext~callbackContext)}`

Initialise the audio system.

`void startDetector()`

Start buffering audio sample for the benefit of the cicada detector.

`void startWhiteNoise()`

Emit white noise from the default output device.

`void stopDetector()`

Gracefully stop and destroy the audio analysis system.

`void stopWhiteNoise()`

Stop emitting white noise.

`java.lang.String writeRecording()`

Write the current buffer to file.

A.2.0.3 Method Detail

execute

```
boolean execute(java.lang.String action,
org.json.JSONArray args,
```

`org.apache.cordova.api.CallbackContext callbackContext)`

`throws org.json.JSONException`

The only method ever to be called from the javascript interface. The call will be in the following format:

```
exec(<successFunction>, <failFunction>, <service>, <action>, [<args>]);
```

where `<service>` will be the name of the class implementing this interface and `<action>` one of the private methods below.

Throws: `org.json.JSONException`

initialiseDetector

`void initialiseDetector(org.apache.cordova.api.CallbackContext callbackContext)`

Initialise the audio system. This should be called before any other call to the audio system is made, including detecting the cicada or requesting an amplitude value.

startDetector

`void startDetector()`

Start buffering audio sample for the benefit of the cicada detector. Once this function is called, it is safe to retrieve values of the cicada estimate through `getCicada()`.

stopDetector

`void stopDetector()`

Gracefully stop and destroy the audio analysis system. A call to `startDetector()` is sufficient to restart the process.

getAmplitude

```
double getAmplitude()
```

Get a value of the amplitude from the microphone.

Returns: a single floating point value between 0 and 1.

getFrequencies

```
double[] getFrequencies()
```

Get array of frequency magnitudes, one per frequency bin. The number of frequency bins will be proportional to the sampling frequency, but would normally be 20, representing frequencies between 1 and 20 kHz. This number **will** however vary and one should not rely on it being 20.

Returns: double array of frequency magnitudes

getCicada

```
double getCicada()
```

Get the estimate of the presence of the cicada, in a float value between 0 and 1.

Returns: the estimated value

startWhiteNoise

```
void startWhiteNoise()
```

Emit white noise from the default output device.

stopWhiteNoise

```
void stopWhiteNoise()
```

Stop emitting white noise. A call to `startWhiteNoise()` is sufficient to restart the noise generation.

writeRecording

```
java.lang.String writeRecording()
```

Write the current buffer to file. The filename is currently determined internally

Returns: the path to the file written.

getReport

```
org.json.JSONArray getReport(int id)
```

Retrieve a survey report. If id is null, then retrieve the latest report.
The JSON Array will be in the form: `{id: {<insect_id> : {name : <value>}, ...}, recording: <true|false>}}`

Appendix B

Sample Survey Report

The document below is a sample report that can be automatically generated for single users, groups, or for the entire body of citizen scientists surveying for the New Forest cicada. This report is generated by an *R* script that connects to the project's back-end PostgreSQL database and generates a reStructuredText document, then compiled to HTML or PDF. The document below is the complete version of all surveyors.

New Forest Cicada Project

Summary of Reports

Updated on Fri Mar 27 15:34:27 2015

Table of Contents

Number of Reports

There are **11386** reports. Of these, 3025 don't have location and **2578** have a location within the New Forest area (approximately). These are reported on the map below.

- Total: 11386
- In the New Forest: 2578
  - BugLife: 577 in the New Forest
  - Others: 2001 in the New Forest

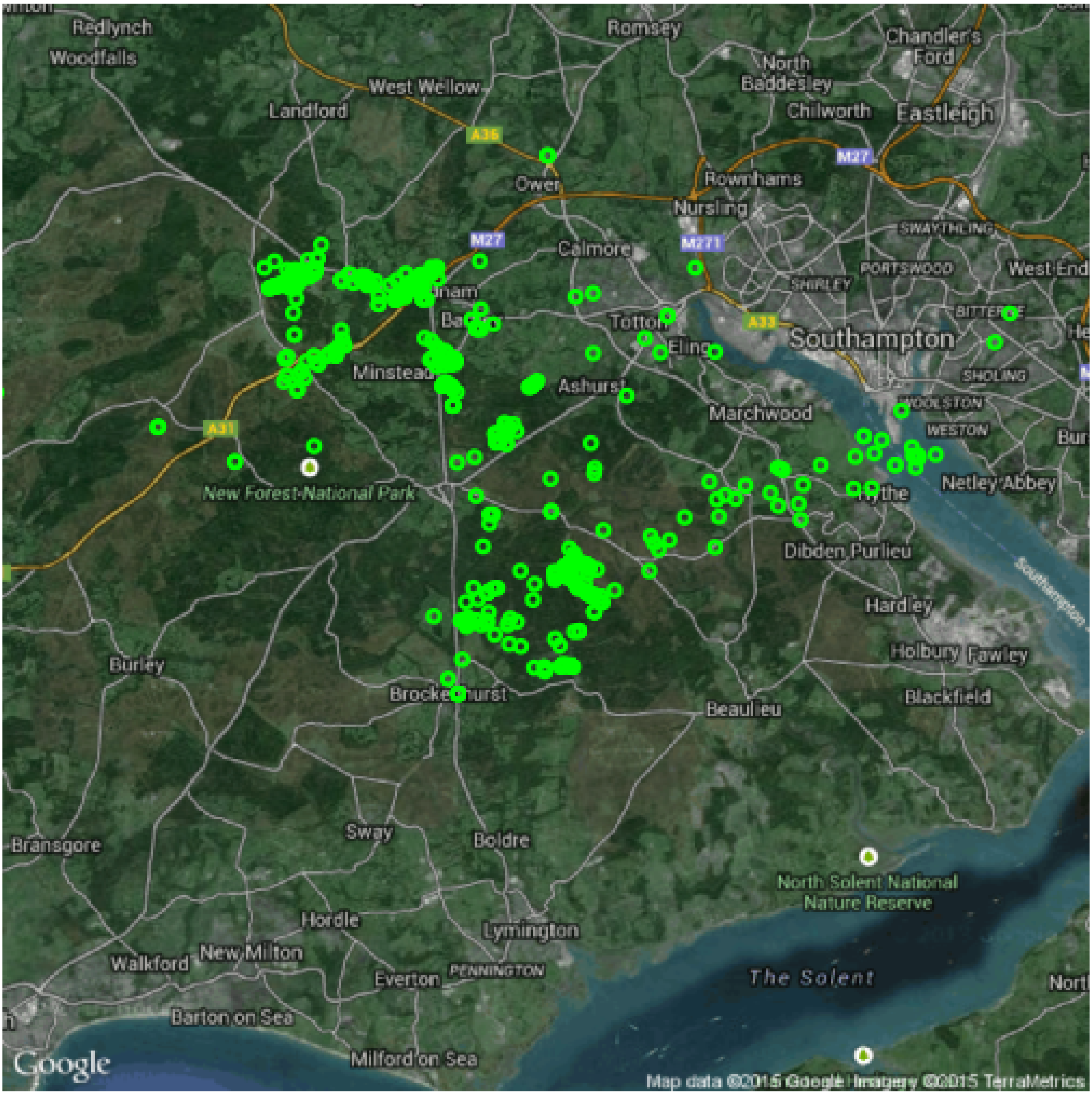


Figure 1: Buglife

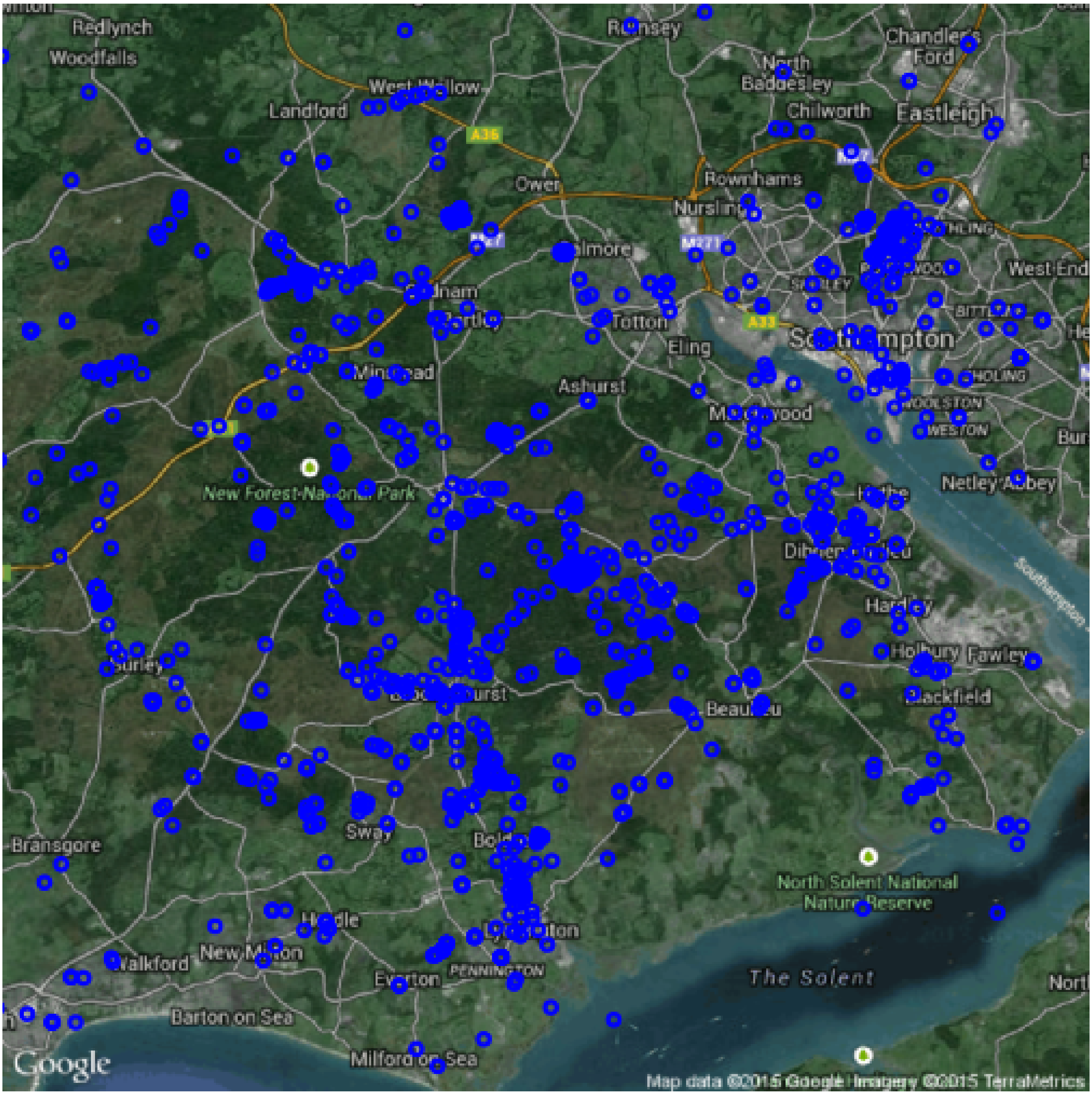


Figure 2: Others

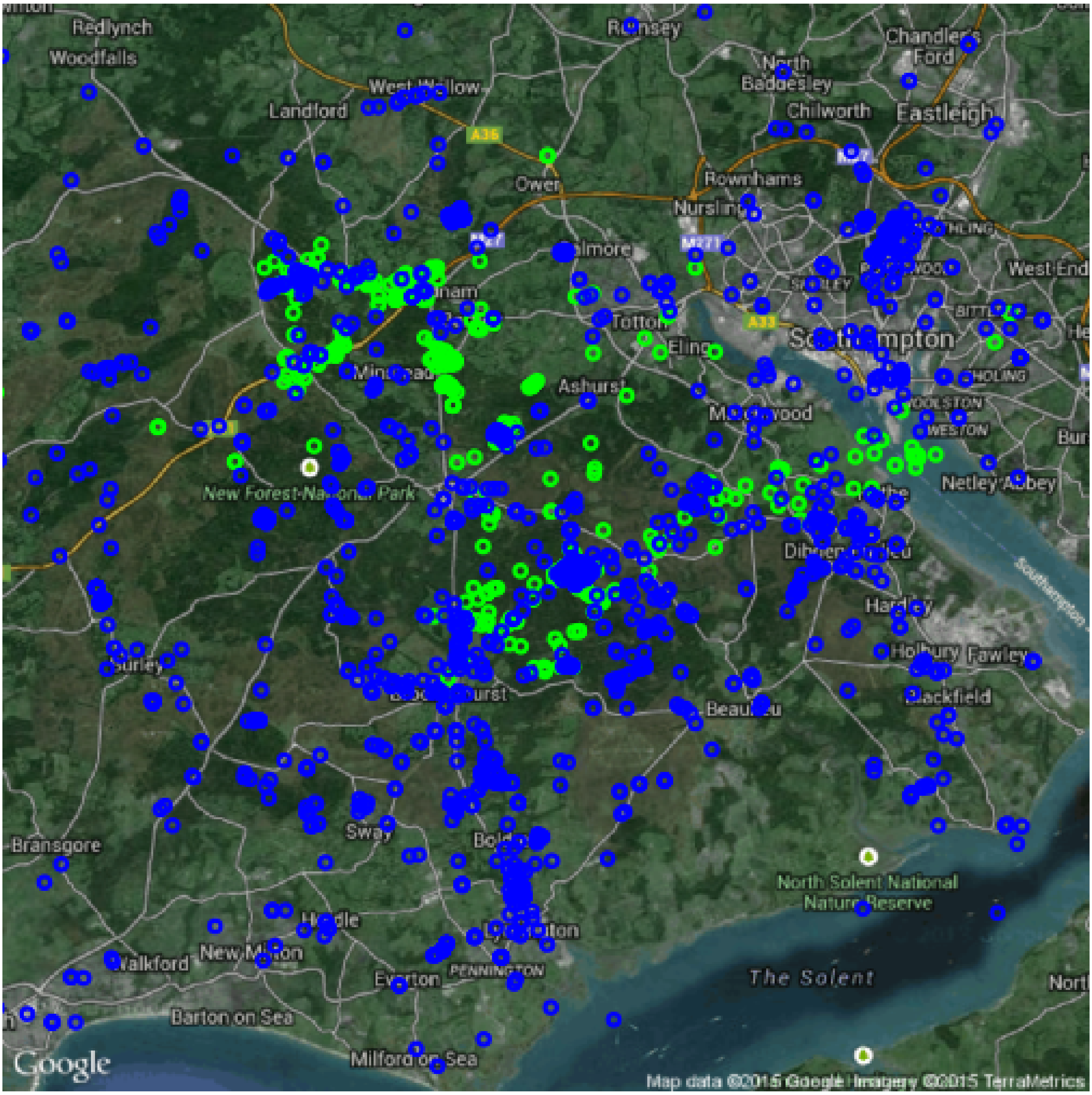


Figure 3: Everyone

The mean location accuracy is 1077. This means that the GPS was mostly prrr disabled.

Survey Dates

Reports have been submitted on 654 days:

31 May 2013	11 Jul 2013	24 Aug 2013	25 Feb 2014
01 Jun 2013	12 Jul 2013	25 Aug 2013	11 Mar 2014
03 Jun 2013	13 Jul 2013	26 Aug 2013	13 Mar 2014
04 Jun 2013	14 Jul 2013	27 Aug 2013	14 Mar 2014
05 Jun 2013	15 Jul 2013	28 Aug 2013	15 Mar 2014
06 Jun 2013	16 Jul 2013	29 Aug 2013	16 Mar 2014
07 Jun 2013	17 Jul 2013	30 Aug 2013	17 Mar 2014
08 Jun 2013	18 Jul 2013	31 Aug 2013	20 Mar 2014
09 Jun 2013	20 Jul 2013	01 Sep 2013	24 Mar 2014
10 Jun 2013	21 Jul 2013	02 Sep 2013	25 Mar 2014
11 Jun 2013	22 Jul 2013	03 Sep 2013	27 Mar 2014
12 Jun 2013	23 Jul 2013	05 Sep 2013	04 Apr 2014
13 Jun 2013	24 Jul 2013	06 Sep 2013	10 Apr 2014
14 Jun 2013	25 Jul 2013	07 Sep 2013	13 Apr 2014
15 Jun 2013	26 Jul 2013	09 Sep 2013	14 Apr 2014
16 Jun 2013	27 Jul 2013	11 Sep 2013	19 Apr 2014
17 Jun 2013	28 Jul 2013	12 Sep 2013	25 Apr 2014
18 Jun 2013	29 Jul 2013	14 Sep 2013	26 Apr 2014
19 Jun 2013	30 Jul 2013	17 Sep 2013	06 May 2014
20 Jun 2013	31 Jul 2013	19 Sep 2013	07 May 2014
21 Jun 2013	01 Aug 2013	22 Sep 2013	09 May 2014
22 Jun 2013	03 Aug 2013	25 Sep 2013	10 May 2014
23 Jun 2013	04 Aug 2013	27 Sep 2013	11 May 2014
24 Jun 2013	05 Aug 2013	29 Sep 2013	14 May 2014
25 Jun 2013	06 Aug 2013	02 Oct 2013	15 May 2014
26 Jun 2013	07 Aug 2013	03 Oct 2013	16 May 2014
27 Jun 2013	08 Aug 2013	05 Oct 2013	17 May 2014
28 Jun 2013	09 Aug 2013	07 Oct 2013	18 May 2014
29 Jun 2013	10 Aug 2013	10 Oct 2013	22 May 2014
30 Jun 2013	11 Aug 2013	25 Oct 2013	24 May 2014
01 Jul 2013	13 Aug 2013	26 Oct 2013	28 May 2014
02 Jul 2013	14 Aug 2013	07 Nov 2013	29 May 2014
03 Jul 2013	15 Aug 2013	10 Dec 2013	01 Jun 2014
04 Jul 2013	17 Aug 2013	13 Dec 2013	02 Jun 2014
05 Jul 2013	18 Aug 2013	03 Jan 2014	05 Jun 2014
06 Jul 2013	19 Aug 2013	08 Jan 2014	06 Jun 2014
07 Jul 2013	20 Aug 2013	23 Jan 2014	07 Jun 2014
08 Jul 2013	21 Aug 2013	07 Feb 2014	08 Jun 2014
09 Jul 2013	22 Aug 2013	11 Feb 2014	09 Jun 2014
10 Jul 2013	23 Aug 2013	23 Feb 2014	10 Jun 2014

11 Jun 2014	13 Jul 2014	10 Aug 2014	05 Oct 2014
12 Jun 2014	14 Jul 2014	14 Aug 2014	06 Oct 2014
13 Jun 2014	15 Jul 2014	15 Aug 2014	11 Oct 2014
14 Jun 2014	16 Jul 2014	18 Aug 2014	12 Nov 2014
15 Jun 2014	17 Jul 2014	19 Aug 2014	16 Nov 2014
17 Jun 2014	18 Jul 2014	20 Aug 2014	17 Nov 2014
18 Jun 2014	19 Jul 2014	21 Aug 2014	22 Nov 2014
19 Jun 2014	20 Jul 2014	23 Aug 2014	24 Nov 2014
21 Jun 2014	21 Jul 2014	24 Aug 2014	25 Nov 2014
22 Jun 2014	22 Jul 2014	29 Aug 2014	02 Dec 2014
23 Jun 2014	23 Jul 2014	02 Sep 2014	04 Dec 2014
24 Jun 2014	24 Jul 2014	05 Sep 2014	16 Dec 2014
25 Jun 2014	25 Jul 2014	06 Sep 2014	19 Dec 2014
27 Jun 2014	26 Jul 2014	08 Sep 2014	05 Jan 2015
28 Jun 2014	28 Jul 2014	09 Sep 2014	11 Jan 2015
29 Jun 2014	29 Jul 2014	10 Sep 2014	18 Jan 2015
30 Jun 2014	30 Jul 2014	13 Sep 2014	22 Jan 2015
01 Jul 2014	31 Jul 2014	14 Sep 2014	30 Jan 2015
02 Jul 2014	01 Aug 2014	16 Sep 2014	03 Feb 2015
03 Jul 2014	02 Aug 2014	18 Sep 2014	05 Feb 2015
05 Jul 2014	03 Aug 2014	20 Sep 2014	24 Feb 2015
09 Jul 2014	04 Aug 2014	28 Sep 2014	03 Mar 2015
10 Jul 2014	06 Aug 2014	02 Oct 2014	04 Mar 2015
11 Jul 2014	08 Aug 2014	03 Oct 2014	05 Mar 2015
12 Jul 2014	09 Aug 2014	04 Oct 2014	

Devices

There are 3014 unique devices. Of these, 1492 (50) are iOS, 1520 are Android.

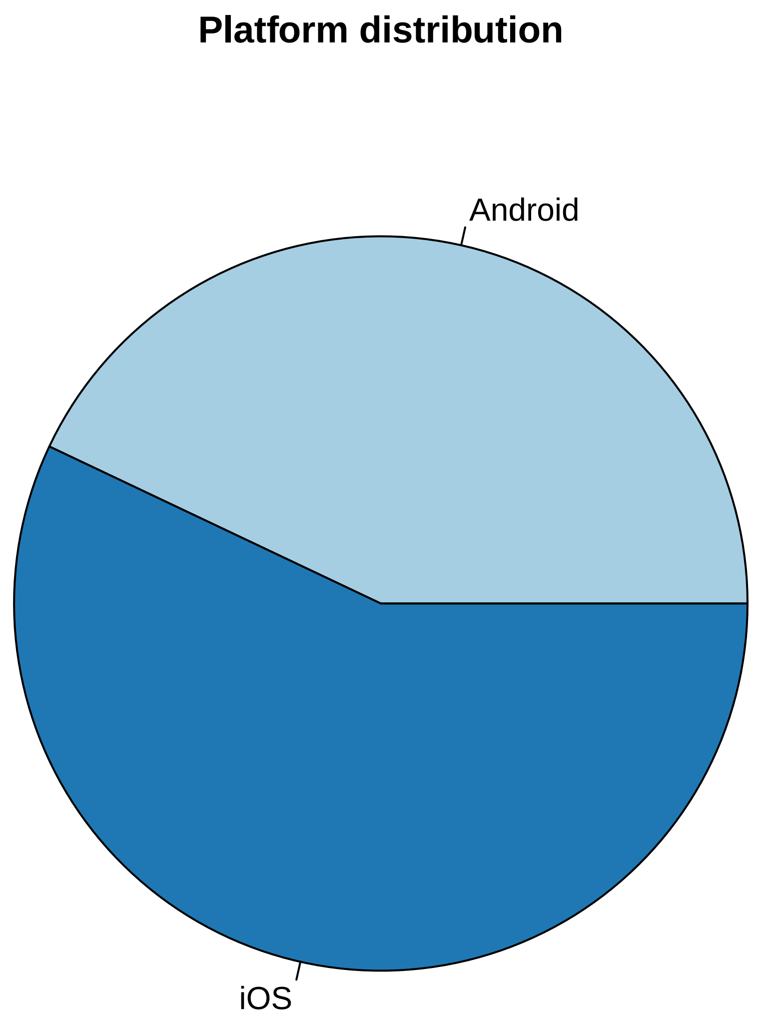


Figure 4: Device distribution

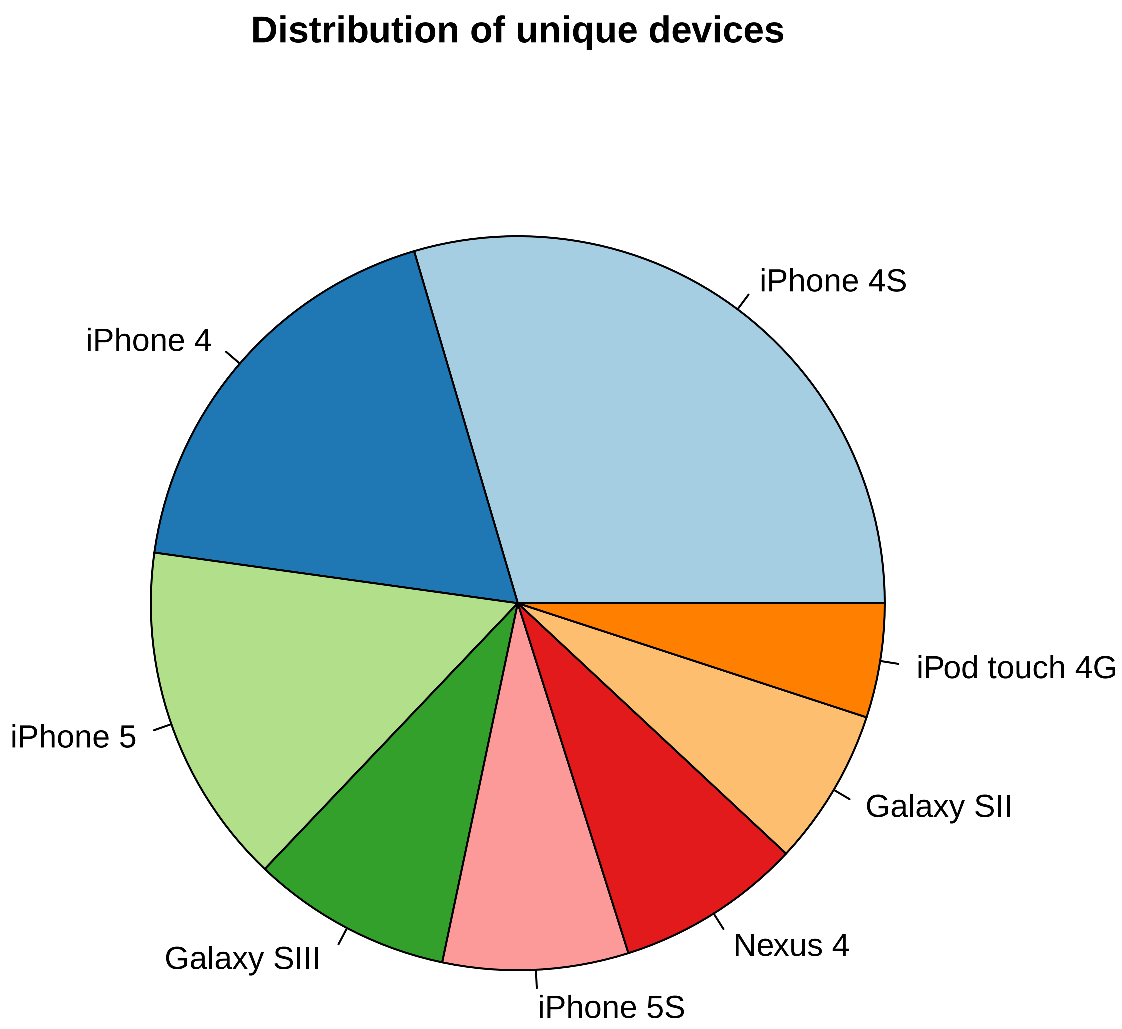


Figure 5: Device distribution

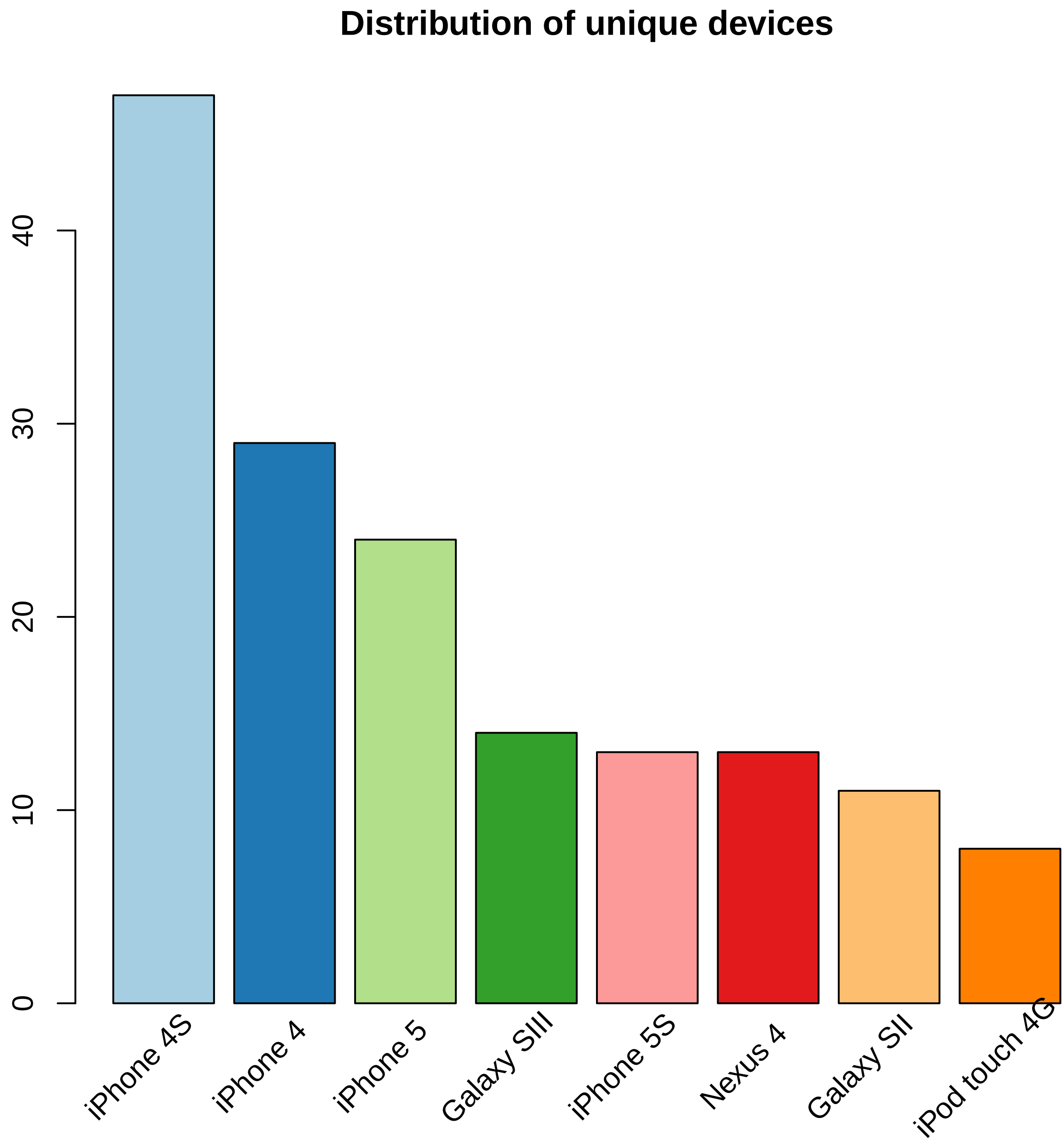


Figure 6: Device distribution

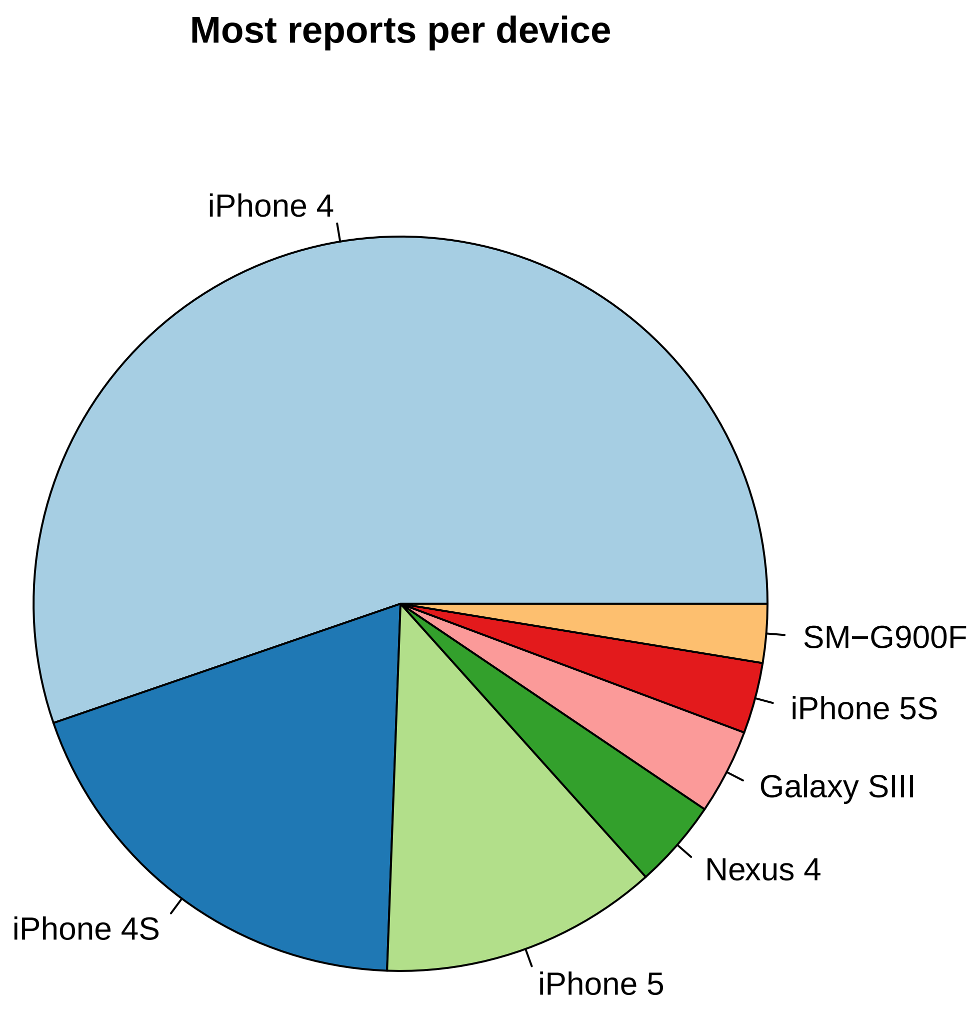


Figure 7: Device distribution

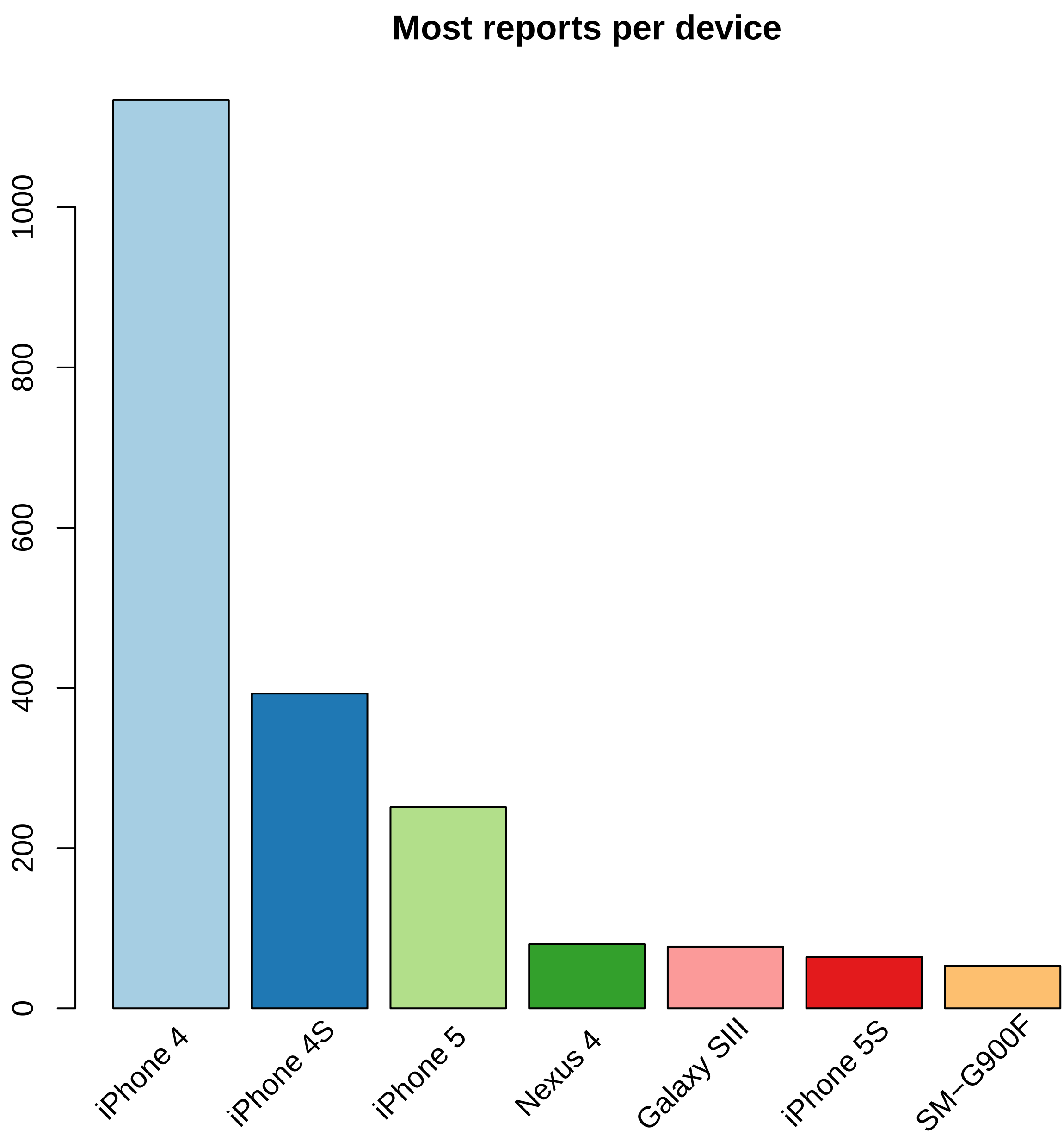


Figure 8: Device distribution

Users

We cannot determine the number of users, as we don't keep any information about them, unless they register on the website. Most users have one device, so we can approximate the number of users to that of devices (see above). Below we report statistics about users assuming a single device per user.

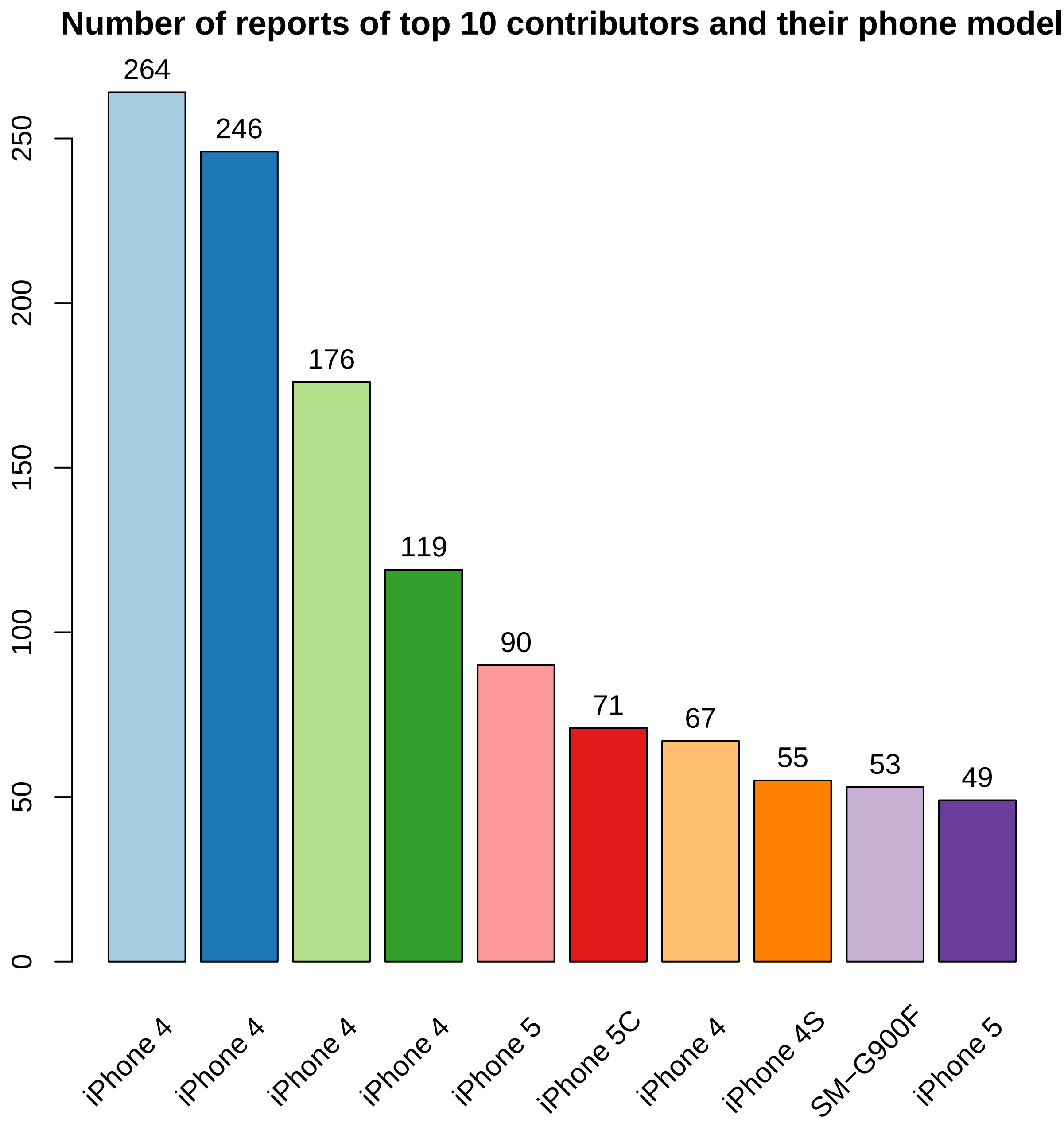


Figure 9: Top 10 contributors

Appendix C

Wildlife Sounds Data Sets

Four core data sets have been used throughout this investigation. These summarised below.

C.1 `NFcrowd`

The `NFcrowd` data set is composed of 235 of New Forest cicada, Roesel's bush-cricket and dark bush-cricket collected with smartphones around the New Forest and the Slovenian Alps. All recordings are 30-seconds long, sampled at 44,100 Hz, taken with the Cicada Hunt application.

C.2 `UKorthoptera`

The `UKorthoptera` data set is composed of 70 recording by Baudewijn Ode from the Dutch Orthoptera Atlas and selected by entomologist Bjorn Beckmann to represent the 28 species of Orthoptera in the UK. Follows a list of descriptive file names and spectrogram of the calls. All rights remain with Baydewijn Ode.

Sickle-bearing bush-cricket - call type 1 - 1 echeme
Sickle-bearing bush-cricket - call type 1 - 1 echeme - example 2
Sickle-bearing bush-cricket - call type 2 - series of syllables
Speckled bush-cricket - call of a few males

Speckled bush-cricket - call of a male on the right, female replying on the left

Speckled bush-cricket - call of one male

Oak bush-cricket - a male drumming on side of container

Oak bush-cricket - a male drumming on side of container - example 2

Southern oak bush-cricket - a male drumming on side of container

Southern oak bush-cricket - a male drumming on side of container - example 2

Short-winged conehead

Short-winged conehead - example 2

Short-winged conehead - example 3

Long-winged conehead

Long-winged conehead - example 2

Great green bush-cricket - at low temperature, others calling in the background

Great green bush-cricket

Great green bush-cricket - example 2

Wartbiter

Wartbiter - example 2

Grey bush-cricket - many echemes - at low temperature

Grey bush-cricket - many echemes

Bog bush-cricket - many echemes - at low temperature

Bog bush-cricket - many echemes

Bog bush-cricket - many echemes - example 2

Roesel's bush-cricket - at low temperature

Roesel's bush-cricket - long echeme

Roesel's bush-cricket - short echemes

Dark bush-cricket - 2 males alternating

Dark bush-cricket - at low temperature, Great green bush-cricket in the background

Dark bush-cricket - many echemes

Dark bush-cricket - many echemes - example 2

Field cricket - at low temperature

Field cricket - chorus of many males

Field cricket - many echemes

Field cricket - many echemes - example 2

House cricket - many echemes

House cricket - several echemes - less active

Wood cricket - chorus of many males

Wood cricket - long echemes

Wood cricket - short echemes

Tree cricket - various echemes

Mole cricket

Mole cricket - example 2

Large marsh grasshopper - 1 series

Large marsh grasshopper - 1 series - example 2

Stripe-winged grasshopper - 1 echeme

Stripe-winged grasshopper - 1 echeme - example 2

Lesser Mottled Grasshopper - 1 echeme

Lesser Mottled Grasshopper - 1 echeme - example 2

Lesser Mottled Grasshopper - 1 echeme - example 3

Common green grasshopper - 1 echeme

Common green grasshopper - 1 echeme - example 2

Woodland grasshopper - 1 echeme

Woodland grasshopper - 1 echeme - example 2

Woodland grasshopper - 1 echeme - example 3

Heath grasshopper - 1 echeme

Heath grasshopper - 1 echeme - example 2

Heath grasshopper - 1 echeme - example 3

Field grasshopper - 1 series of echemes

Field grasshopper - 1 series of echemes - example 2

Field grasshopper - 4 echemes - at low temperature

Lesser marsh grasshopper - 1 series of echemes

Lesser marsh grasshopper - 1 series of echemes - example 2

Meadow grasshopper - at lower temperature - 2 echemes - other males in the background

Meadow grasshopper - several echemes

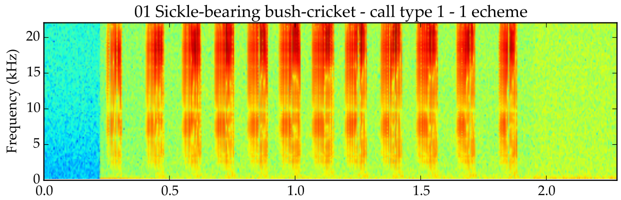
Mottled grasshopper - 1 series of echemes

Mottled grasshopper - 1 series of echemes - example 2

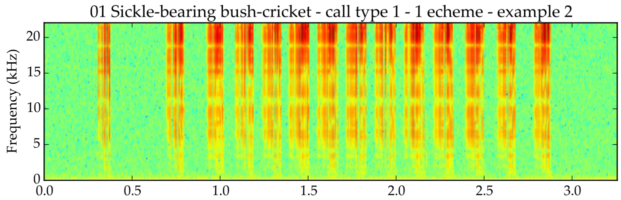
Rufous grasshopper - 1 verse

Rufous grasshopper - 1 verse - example 2

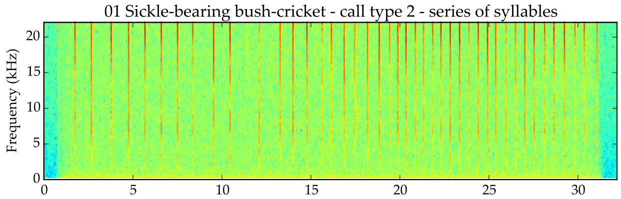
01 Sickle-bearing bush-cricket - call type 1 - 1 echeme



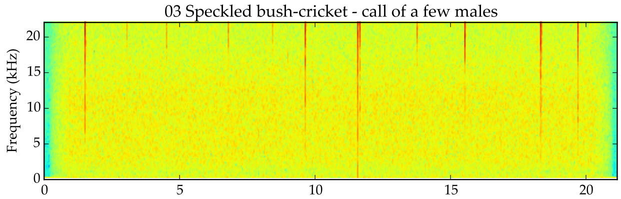
01 Sickle-bearing bush-cricket - call type 1 - 1 echeme - example 2



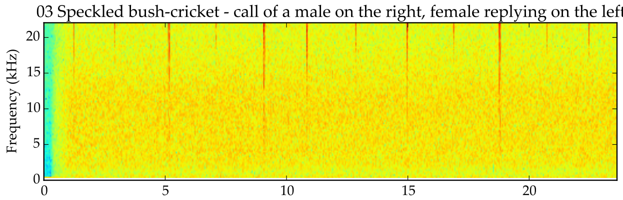
01 Sickle-bearing bush-cricket - call type 2 - series of syllables



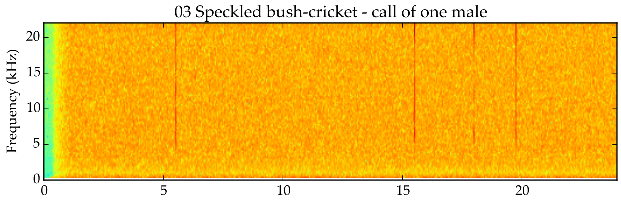
03 Speckled bush-cricket - call of a few males



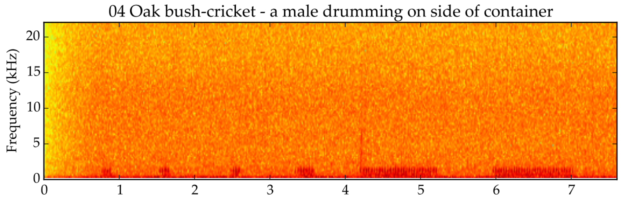
03 Speckled bush-cricket - call of a male on the right, female replying on the left



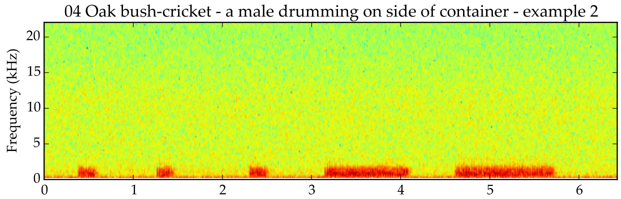
03 Speckled bush-cricket - call of one male



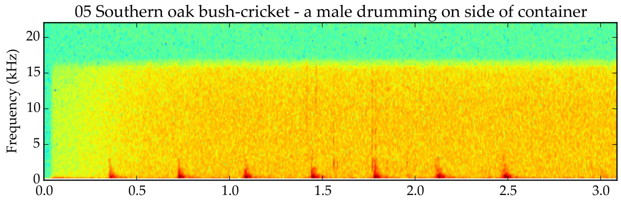
04 Oak bush-cricket - a male drumming on side of container



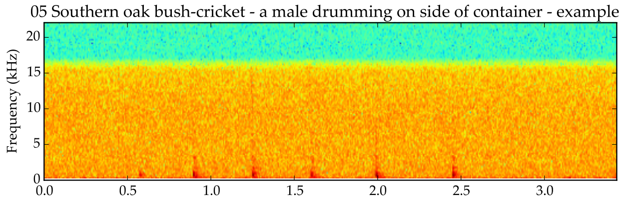
04 Oak bush-cricket - a male drumming on side of container - example 2



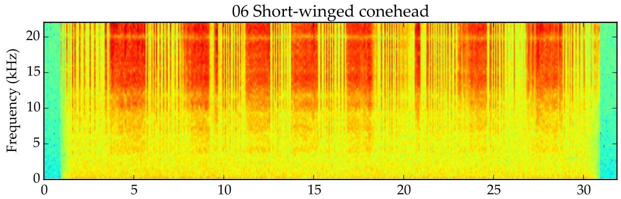
05 Southern oak bush-cricket - a male drumming on side of container



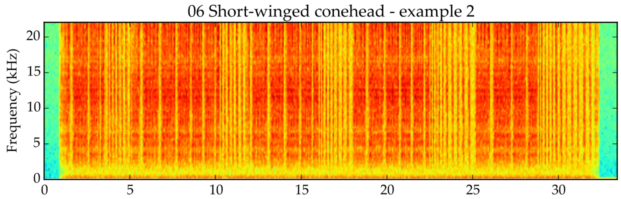
05 Southern oak bush-cricket - a male drumming on side of container - example 2



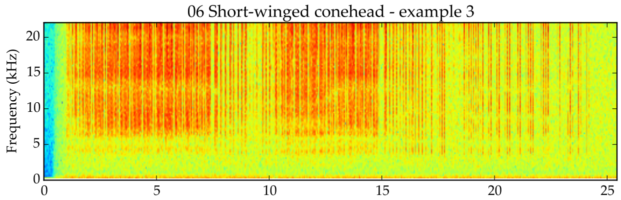
06 Short-winged conehead



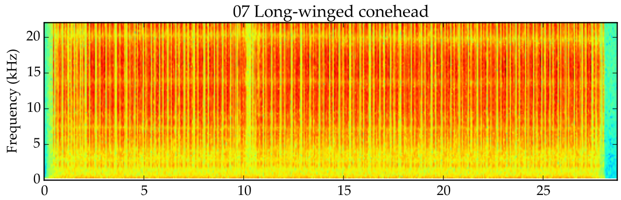
06 Short-winged conehead - example 2



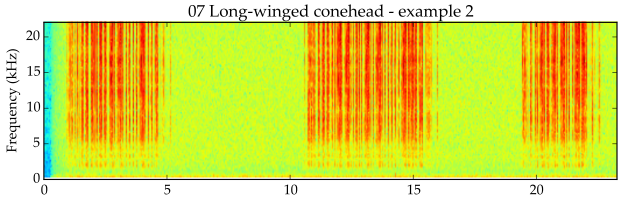
06 Short-winged conehead - example 3



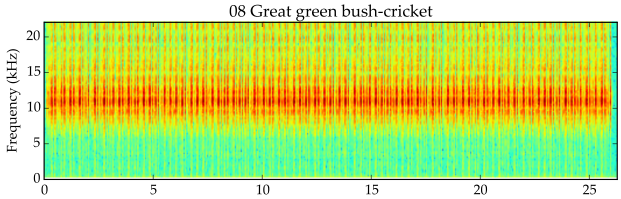
07 Long-winged conehead



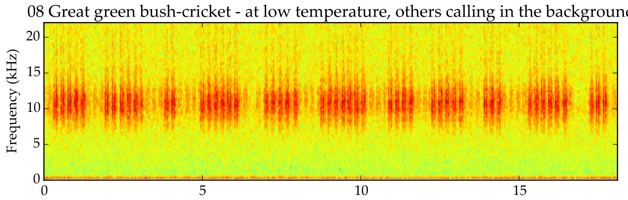
07 Long-winged conehead - example 2



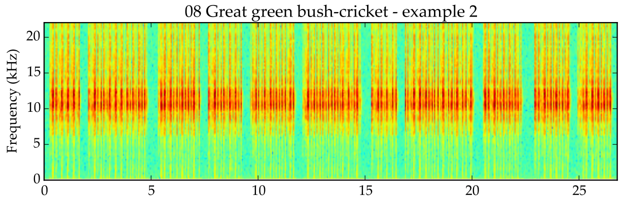
08 Great green bush-cricket

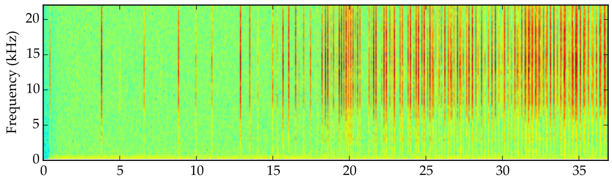


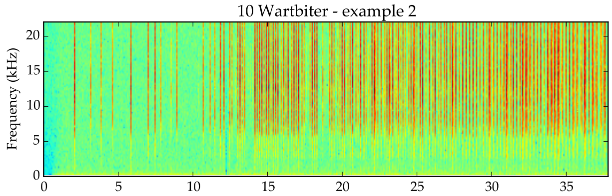
08 Great green bush-cricket - at low temperature, others calling in the background

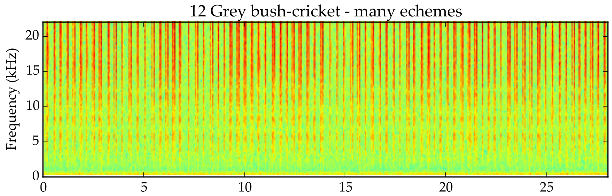


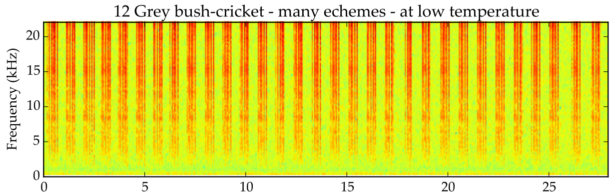
08 Great green bush-cricket - example 2

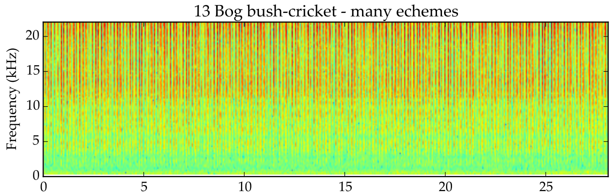


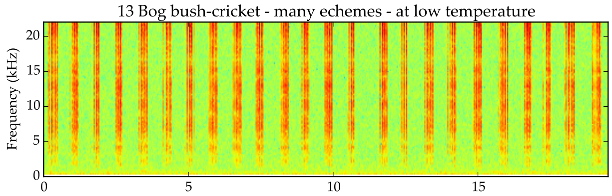
10 Wartbiter


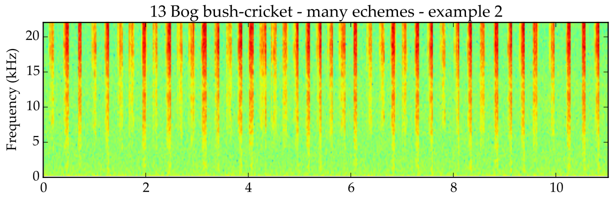
10 Wartbiter - example 2


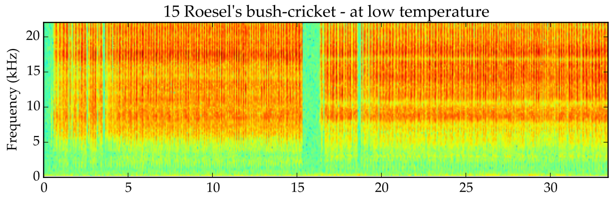
12 Grey bush-cricket - many echemes


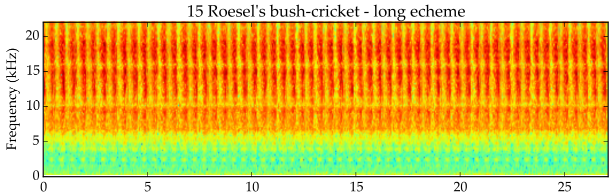
12 Grey bush-cricket - many echemes - at low temperature


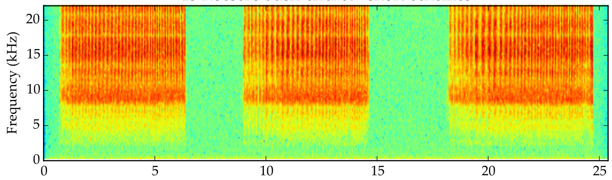
13 Bog bush-cricket - many echemes


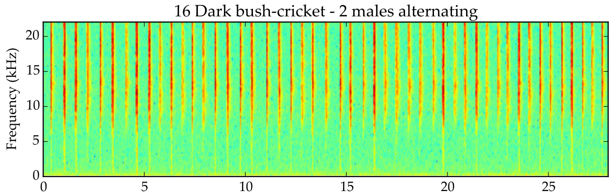
13 Bog bush-cricket - many echemes - at low temperature


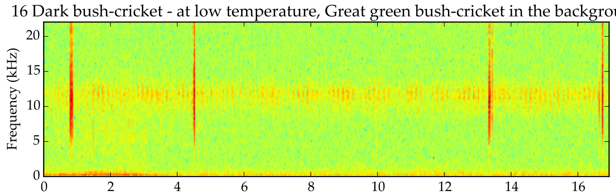
13 Bog bush-cricket - many echemes - example 2


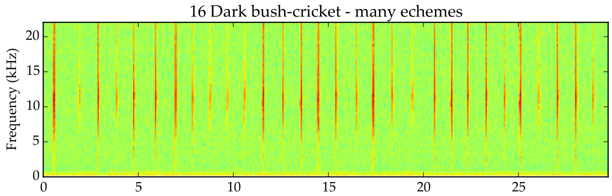
15 Roesel's bush-cricket - at low temperature


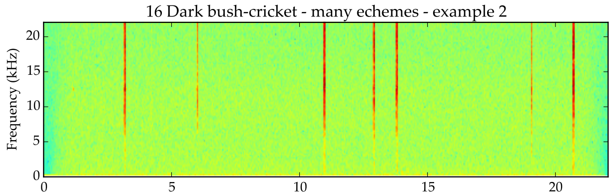
15 Roesel's bush-cricket - long echeme


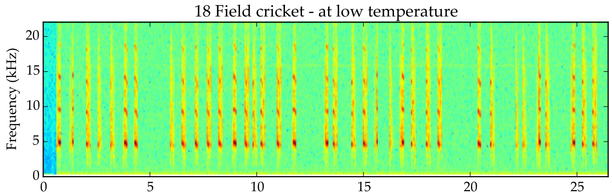
15 Roesel's bush-cricket - short echemes


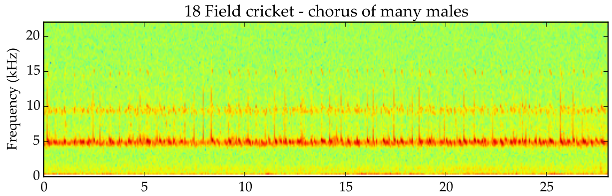
16 Dark bush-cricket - 2 males alternating


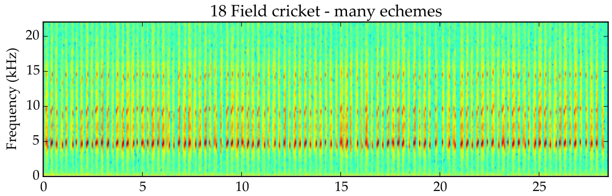
16 Dark bush-cricket - at low temperature, Great green bush-cricket in the background


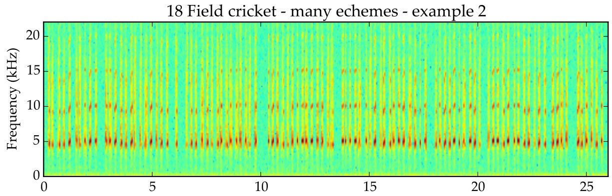
16 Dark bush-cricket - many echemes


16 Dark bush-cricket - many echemes - example 2


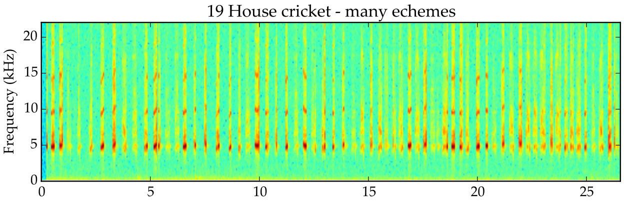
18 Field cricket - at low temperature


18 Field cricket - chorus of many males


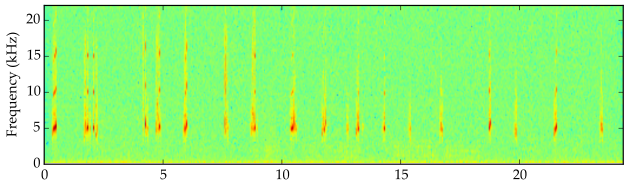
18 Field cricket - many echemes


18 Field cricket - many echemes - example 2


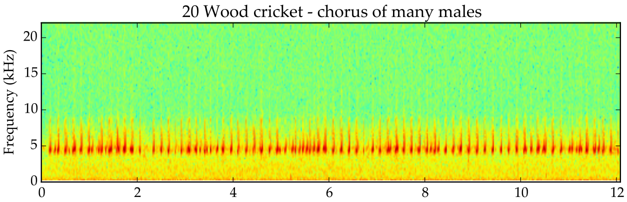
19 House cricket - many echemes



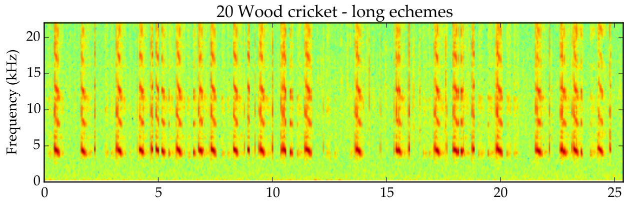
19 House cricket - several echemes - less active



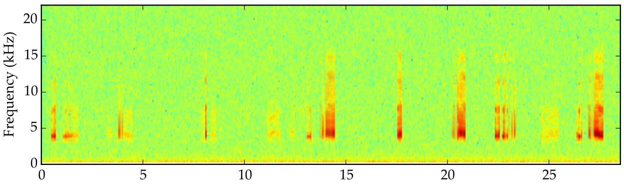
20 Wood cricket - chorus of many males



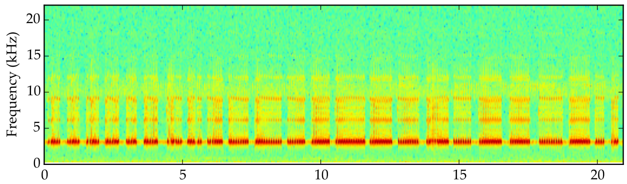
20 Wood cricket - long echemes



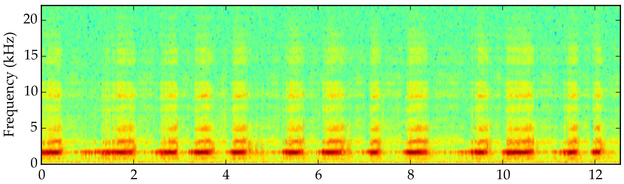
20 Wood cricket - short echemes



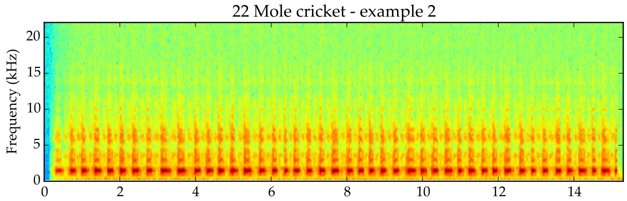
21 Tree cricket - various echemes



22 Mole cricket



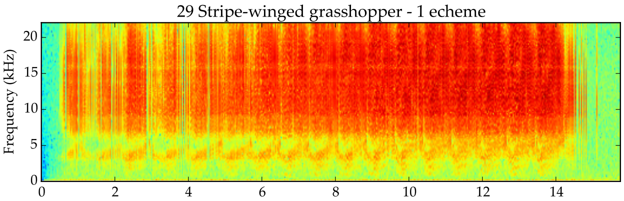
22 Mole cricket - example 2



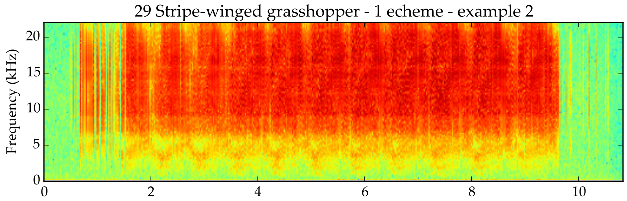
26 Large marsh grasshopper - 1 series

26 Large marsh grasshopper - 1 series - example 2

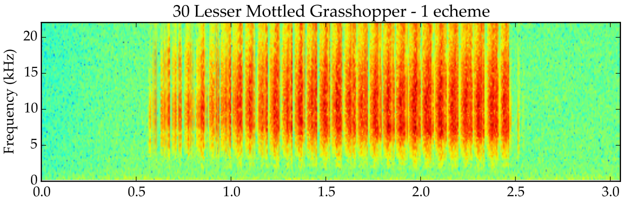
29 Stripe-winged grasshopper - 1 echeme



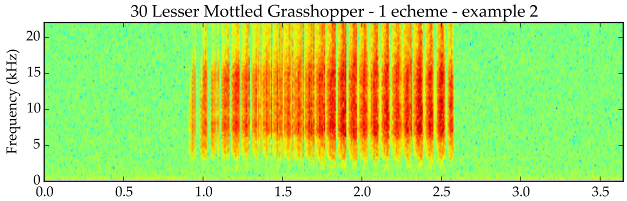
29 Stripe-winged grasshopper - 1 echeme - example 2



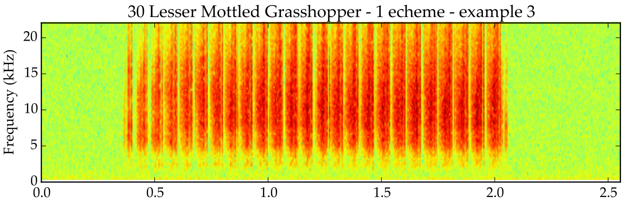
30 Lesser Mottled Grasshopper - 1 echeme



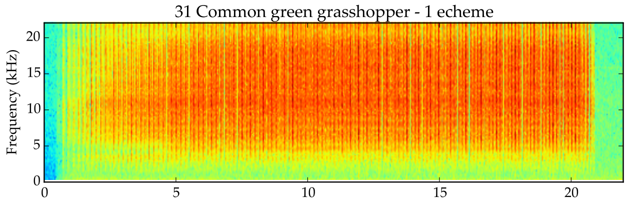
30 Lesser Mottled Grasshopper - 1 echeme - example 2



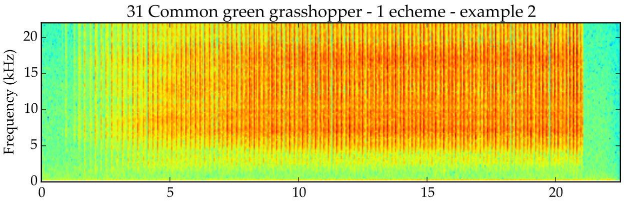
30 Lesser Mottled Grasshopper - 1 echeme - example 3



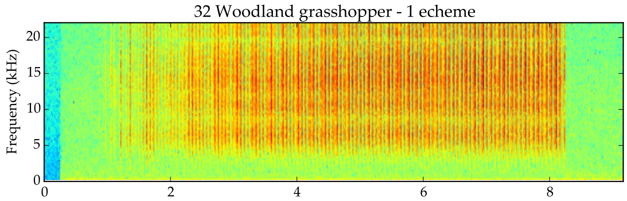
31 Common green grasshopper - 1 echeme



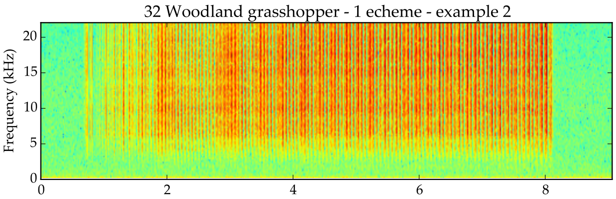
31 Common green grasshopper - 1 echeme - example 2



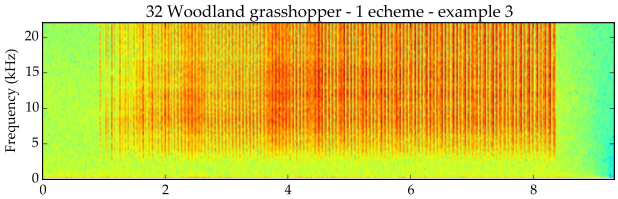
32 Woodland grasshopper - 1 echeme



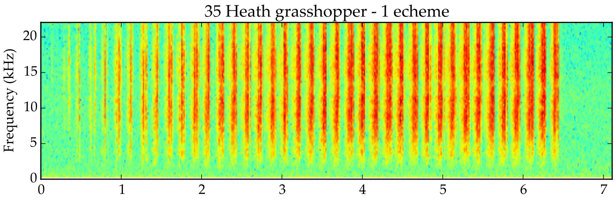
32 Woodland grasshopper - 1 echeme - example 2



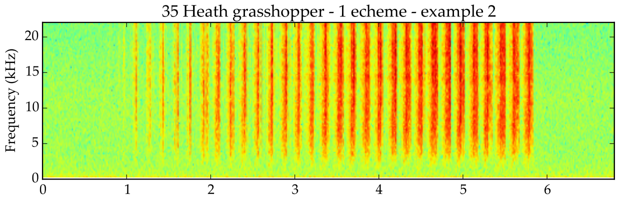
32 Woodland grasshopper - 1 echeme - example 3



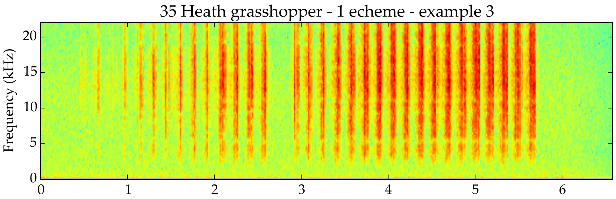
35 Heath grasshopper - 1 echeme



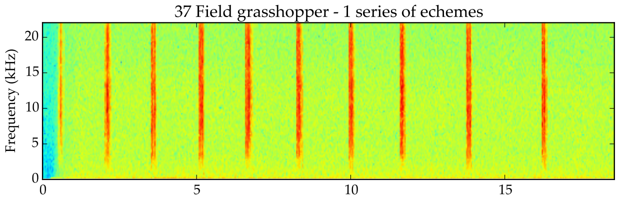
35 Heath grasshopper - 1 echeme - example 2



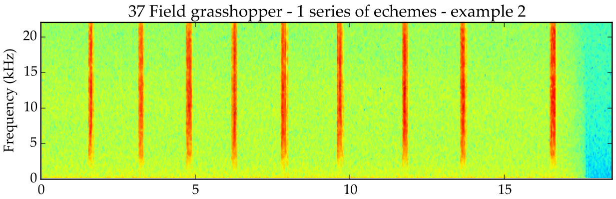
35 Heath grasshopper - 1 echeme - example 3



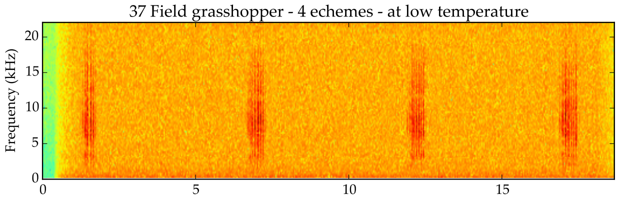
37 Field grasshopper - 1 series of echemes



37 Field grasshopper - 1 series of echemes - example 2

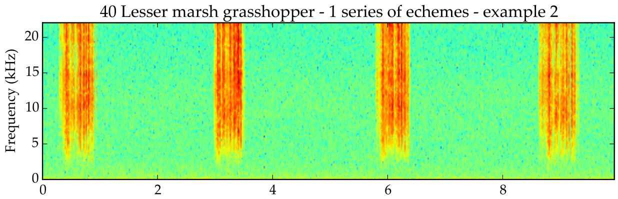


37 Field grasshopper - 4 echemes - at low temperature

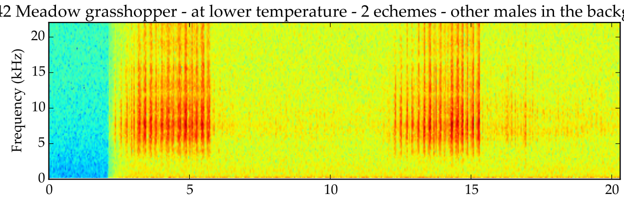


40 Lesser marsh grasshopper - 1 series of echemes

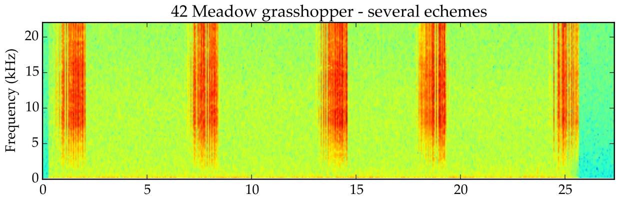
40 Lesser marsh grasshopper - 1 series of echemes - example 2



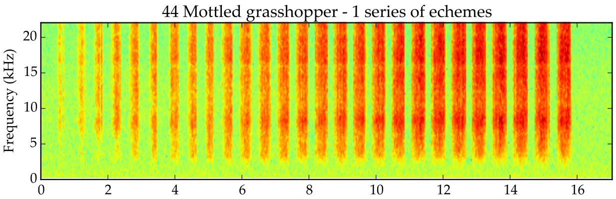
42 Meadow grasshopper - at lower temperature - 2 echemes - other males in the background



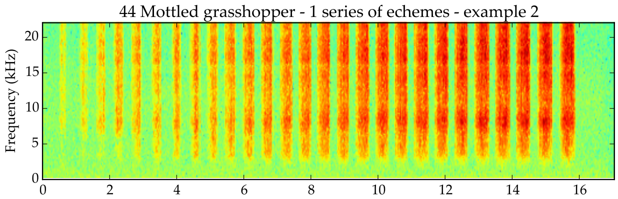
42 Meadow grasshopper - several echemes



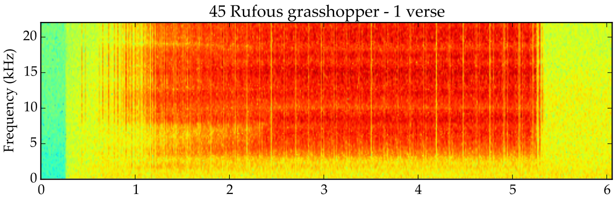
44 Mottled grasshopper - 1 series of echemes



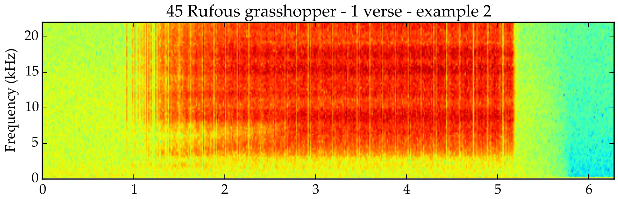
44 Mottled grasshopper - 1 series of echemes - example 2



45 Rufous grasshopper - 1 verse



45 Rufous grasshopper - 1 verse - example 2



C.3 `nips4b`

The `nips4b` data set has been assembled by the BIOTOPE society, which collects bird recordings around Europe. It was made available for the NIPS 2013 multi-label bird species classification challenge and remains available after the competition has ended. More details about the data set can be found on the competition's submission page: https://www.kaggle.com/c/multilabel-bird-species-classification-nips2013.

C.4 `BLorthoptera`

Finally, a large data set of Orthoptera recordings and related sounds has been collected by the author of this thesis, and provided by the curator of the British Library's wildlife and environmental sounds division. A brief description of the files has is reported schematically below.

Family	**Samples**
Gryllidae	184
Tettigoniidae	96
Acrididae	24
Discoglossidae	16
Meliphagidae	4
Alaudidae	3
Turdidae	2
Ranidae	1
Strigidae	1
Not confirmed	2

TABLE C.1: Recordings by family for the `BLorthoptera` data set

Species	**Samples**
Field Cricket	49
Mole Cricket	30
Sword-bearing Conehead	27
Fire-bellied Toad	16
Black-horned Tree Cricket	16
Carolina Ground Cricket	15
Allard's Ground Cricket	14
Spring Field Cricket	13
Broad-winged Bush Katydid	12
Snowy Tree Cricket	11
Common True Katydid	10
Four-spotted Tree Cricket	9
Narrow-winged Tree Cricket	8
Fall Field Cricket	7
Gladiator Meadow Katydid	6
Meadow Grasshopper	6
Common Meadow Katydid	6
Froggatt's Buzzer Grasshopper	5
Striped Ground Cricket	4
Oblong-winged Katydid	4
Tui	4
Bush Cricket	4
Mottled Grasshopper	4
Bush Cricket sp. nov.	4
Wart-biter	3
Woodlark	3
Cricket	3
Great Green Bush Cricket	3
Grasshopper	3
Dark Bush Cricket	3
Roesel's Bush Cricket	2
Dusky-faced Meadow Katydid	2
Wood Cricket	2
White-fronted Wart-biter	2
Northern Meadow Locust	2
Texas Bush Katydid	2
Say's Bush Cricket	2
Common Field Grasshopper	2
Common Green Grasshopper	2
Southern Field Cricket	2
Long-winged Cone-head	1
Bog Bush Cricket	1
Bluethroat	1
Tree Cricket	1
Curve-tailed Katydid	1
Tawny Owl	1
Iberian Marsh Frog	1
Slender Conehead	1
Nimble Meadow Katydid	1
Nightingale	1
Least Shieldback	1

TABLE C.2: Recordings by species for the `BLorthoptera` data set

Appendix D

Software Requirements

This appendix provides an early version of the formal software requirements of the New Forest Cicada Project Formal, which coordinated the development of the system required for this work. Part of the tasks have been developed in cooperation with a student intern, working on the project over a three month period.

New Forest Cicada Project
Formal Software Requirements
v. 0.1

Davide Zilli

Started: June 25, 2012
Last updated: July 18, 2012

The New Forest Cicada project is based on Crowdsourcing and Citizen Science techniques. As such, it requires *a*) a strong web presence and *b*) a mobile app to provide information and collect data from users. In this context the development of the software necessary to satisfy these requirements is not an implementation excercise or mere publicity effort, but a founding block of the project itself. To an extent, novel research will only be possible once these tools are ready to use.

The development of the tools can, however, eventually drain a large part of the project resources. To limit this, the cooperation of different developers will be required. This document outlines the software requirements in order to facilitate the collaboration of developers.

The document is by nature work in progress and the latest version should always be considered.

Contents

1 Web presence and Backend

The website is made of two separate components:

1. a collection of information pages about the project, the researchers and the the New Forest Cicada, with dynamic content generation.
2. a dynamic backend to the app, providing user management, data collection/storage, data visualisation, additional user participation features (such as a "Cicada game").

The two should however be implemented on the same system because of *a*) a slight overlap (e.g. an info page about the project would be required on both) and *b*) because it would facilitate users moving from one to the other.

1.1 Features

Follows a list of features for the web site. The order in which they should be implemented is expressed in subsection 1.2.

Theme

Design of a CSS/JS theme for the frontend. Can potentially reuse an existing freely available one.

Basic static website Progress: 10%

Static web pages including:

- Information page about the project and authors.
- Information page about the cicada. The species found in the New Forest, a sonogram and oscillogram of the call, a sample audio of the call.
- A selection of media: photos, videos
- Information page about the app.

More dynamic pages include:

- Weather feed on the home page, to attract people to go look for the cicada. It should be targeted at "promoting" sunny warm days.
- Downloads of the app.
- Blog entries (see below) on project updates

Blog

A blog of the highlights of the project. Features should include:

- Subscription (RSS or similar, Facebook, Twitter)

- Commenting on blog entries
- Easy input of new contents (requires user management)

The use of existing platforms, such as wordpress, should be considered.

Social media integration in the website

- Embed Twitter feed
- Facebook/Google+ *like* and sim.
- Share this page
- Email this page

Twitter feed

- Creation of a Twitter account

Facebook page

- Creation of a facebook page
- Detail about the project, link to the website on the info section

User Management

Django integrates easy user management. This will be required for the Blog (posting, commenting) and for the app.

OpenID and social media login should be allowed. It should include at least OpenID, Google and Facebook authentication, as well as custom registration/login.

Sound Game (iHear)

Implement a simple game platform. Rules for the game are:

- users are given a sound file on a page that also displays the location of the recording, the time, an oscillogram and a spectrogram.
- a player has to listen to the sound and tell what animal they can hear
- a points system rewards them for:
    1. guessing the correct animal, if known
    2. guessing the correct animal, if unknown. Points will be awarded after the identification has been confirmed
    3. extra points will be awarded for more specific classification (species, subspecies, . . .)
- initially, known recordings will be presented, so that the user can be trained and assessed.

The dynamics of the games are still loosely defined. A precise design will be required before implementation. Potentially it could include:

- Levels. At initial levels the user plays with known recordings. Afterwards, known and unknown will be mixed.
- Different sorts of animals. The target is cicadas, but nothing prevents us to use any sound emitting animals: ducks, elephants, monkeys, ... The implementation should not change.

Basic Backend

Receive recordings from the mobile app

Manipulate and show recordings

- display recordings' oscillogram, spectrogram, source, location
- display a map of observations

1.2 Roadmap

SY = systems, FE = front end, BE = back end, EX = extra

```
~~0001. [SY] django project setup~~                                [DONE]
~~0002. [SY] redmine setup~~                                       [DONE]
0003. [SY] git repo setup
0004. [FE] website mockup/system diagram
0005. [FE] theme
0006. [FE] static info pages
0016. [BE] OpenID auth
0017. [SY] Website analytics
0007. [FE] homepage weather feed
0008. [FE] blog
0009. [BE] receive recordings from mobile app
0010. [BE] display list of recordings
0011. [BE] display recordings' oscillogram, spectrogram, source, loc
0012. [BE] display recordings on a map
0013. [FE] create facebook page (only once website is running)
0014. [EX] game design and mockup
0015. [EX] game implementation
```

2 Mobile app

2.1 Features

Client Setup

- User interface mockup
- Formal app requirements
- framework choice (native API, HTML5 framework?)
- storage model

WAV recording

- Neat, polished, un-copyrighted WAV recording.
- 44.1 kHz sampling rate or higher
- Storage
    1. sqlite db? Are there better options?
    2. Keep important files until able to send them
    3. What if you run out of space?
    4. Keep files that user wants to keep anyway.
    5. Delete files that have been transmitted and are not important.
- 30 seconds windowing for continuous monitoring
- Extensive testing.

File sharing system

- Choice of method/protocol. HTTP? Are there better options?
- Cross site request forgery protection
- Authentication
- Coupling with backend

Cross Platform port

- What are the target platforms? Mandatory: Android, iPhone. Desirable: Windows mobile, Blackberry, Symbian.
- Can an easier implementation be made for other platforms?

Icon Set

Design an icon set for both the website and the apps. The icons will recall the cicada, but present it as a pleasant character rather than a scary insect.

Social aspect

Share through social media:

- Findings, observations (both sound and pictures)
- amount of the forest covered
- "Now in the forest"

Game App

A game to publicize the New Forest Cicada, building on examples such as "Angry Birds". The implementation of this will require an immense development effort and is left as future work.

Integration with PlanetOrchid

Provide a Human-Agent agile-teaming game to look for electronic cicadas, based on the PlanetOrchid platform. The interface should be easy for vast public engagement, including engagement of children in schools.

2.2 Roadmap

```
0001. Android client setup
0002. GUI mockup
0003. home page - buttons
0004. info page - the new forest cicada - text
0005. info page - the new forest cicada - sound
0006. record in WAV
0011. storage model
0007. send recording to server (file sharing)
0008. iOS client port
0009. other platforms port
0010. icon set
```

3 Systems

Version Control

Software is under `git` version control, and the address is:

`git://git.ccada.co.uk`

Tools and Frameworks

Component	Language or tool
Web server	apache running `mod_wsgi`.
Web framework	Python Django
App	A combination of native iOS dev kit (objective C) and native Android APIs (Java) plus any other required by additional platforms

Table 1: Languages and frameworks

Appendix E

Awards and Media Engagement

E.1 Awards

The project has received the following awards:

- *Silver Medal for Engineering* at the *SET for Britain 2015* research competition (http://www.setforbritain.org.uk/2015winners.asp)
- *Best Student Paper Award* at the *International Joint Conference for Artificial Intelligence (IJCAI) 2013, Beijing, China*
- *Winner* of the Faculty heat of the *Three Minute Thesis (3MT®) Competition*, University of Southampton.

E.2 Media Presence

The project has featured several times in local and national mass media publication. An up-to-date list of the most important publications can be found at http://newforestcicada.info/press/. The list below summarises the most relevant ones at the time of writing:

- BBC Radio Solent

  Dr Alex Rogers presents the project on air, 05 June 2013

- Get ready to find the British cicada

  DEFRA Biodiversity News, Issue #64, page 19, April 24, 2014

- Hunt for cicada only the young can hear

  The Telegraph, April 24, 2014

- Elusive Forest insect is one of UK's most endangered species

  New Forest Post, March 13, 2014

- New Forest cicada is named one of country's most endangered creatures by Species Recovery Trust

  James Franklin, Daily Echo, March 11, 2014

- The Unsuspecting Naturalist

  Paul Marks, New Scientist, Issue 2932, Aug. 29, 2013

- Bits and Bugs—Making the most of Technology in Entomology

  Alexander Hay, Software Sustainability Institute Bulletin of the Royal Entomological Society, Aug. 22, 2013

- BBC Wildlife—3 Things we love this month

  BBC Wildlife, Volume 31, number 8, July 15, 2013

- Shazam for cicadas: an app that helps scientists pick up the tune of a rare bug

  Lauren Hockenson, June 25, 2013

- Searching for Cicada Song: A crowdsourcing project

  Samuel, OpenSignal, June 25, 2013

- Search goes on for elusive New Forest species not seen for ten years

  Daily Echo, June 12, 2013

- The hunt for the New Forest Cicada

  BugLife, June 11, 2013

- Has the UK's only cicada disappeared?

  Wildlife Extra News, June 6, 2013

- Smartphone app launched to track down rare insect

  Daily Echo, June 5, 2013

- App to aid rare New Forest cicada hunt

  BBC News, 31 May 2013 www.bbc.co.uk/news

- How a smartphone could become an endangered cicada detector

  The Guardian, 20 September 2012, by Duncan Graham-Rowe guardian.co.uk

- New App Developed to Detect Endangered Cicada

  Enviro News & Business, 28 September 2012, by Rebecca Watson

Bibliography

M. A. Acevedo, C. J. Corrada-Bravo, H. Corrada-Bravo, Villanueva-Rivera L. J., and Aide T. M. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, 4(4):206 – 214, 2009.

D. P. Anderson and J. Cobb. SETI@home: an experiment in public-resource computing. *Communications of the ACM*, 45:56–61, 2002.

D. W. Baldock. *Grasshoppers and crickets of Surrey.* Surrey Wildlife Trust, 1999.

L. F. Baptista and S. L. L. Gaunt. Bioacoustics as a tool in conservation studies. In Janine Rhea Clemmons and Richard Buchholz, editors, *Behavioral approaches to conservation in the wild*, chapter 9, pages 212–242. Cambridge Univ Press, 1997.

S. Bell, M. Marzano, J. Cent, H. Kobierska, D. Podjed, D. Vandzinskaite, H. Reinert, A. Armaitiene, M. GrodziÅĎska-Jurczak, and R. Muršič. What counts? Volunteers and their organisations in the recording and monitoring of biodiversity. *Biodiversity and Conservation*, 17(14):3443–3454, 2008.

T. Benton. *Grasshoppers & crickets*. HarperCollins Publishers, 2012.

M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, D. Crowell, and Katrina P. Soylent. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10*, page 313, New York, New York, USA, 2010. ACM Press.

C. M. Bishop and N. M. Nasrabadi. *Pattern recognition and machine learning*, volume 1. springer New York, 2006.

S. Blasiak and H. Rangwala. A hidden markov model variant for sequence classification. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, pages 1192–1197. Barcelona, Catalonia, Spain, 2011.

D. C. Brabham. Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Convergence: The International Journal of Research into New Media Technologies*, 14(1):75–90, 2008.

V. A. E. Chaves, C. M. Travieso, A. Camacho, and J. B. Alonso. Katydids acoustic classification on verification approach based on MFCC and HMM. *Proceedings of the 16th IEEE International Conference on Intelligent Engineering Systems (INES)*, pages 561–566, 2012.

D. Chesmore. Methodologies for automating the identification of species. In *1st BioNet-International Working Group on Automated Taxonomy*, 2000.

D. Chesmore. Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals. *Applied Acoustics*, 62(12):1359–1374, 2001.

D. Chesmore. Automated bioacoustic identification of species. *Anais da Academia Brasileira de Ciências*, 2004.

D. Chesmore. The Automated Identification of Taxa: Concepts and Applications. In Norman MacLeod, editor, *Automated Taxon Identification in Systematics. Theory, Approaches and Applications*, chapter 6, pages 83–100. CRC Press, 2007.

D. Chesmore and E. Ohya. Automated identification of field-recorded songs of four British grasshoppers using bioacoustic signal recognition. *Bulletin of Entomological Research*, 94(04):319–330, 2004.

Patrick J Clemins, Michael T Johnson, Kirsten M Leong, and Anne Savage. Automatic classification and speaker identification of African elephant (Loxodonta africana) vocalizations. *The Journal of the Acoustical Society of America*, 117(2):956–963, 2005.

A. Coates and A. Y. Ng. Learning feature representations with k-means. In *Neural Networks: Tricks of the Trade*, pages 561–580. Springer, 2012.

C. B. Cooper, J. Dickinson, T. Phillips, and R. Bonney. Citizen science as a tool for conservation in residential ecosystems. *Ecology and Society*, 12(2):11, 2007.

R. Daponte. New Forest Fact File. Technical report, Forestry Commission, 2004.

S. Davis and P. Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), 1980.

D. De Jager, A. L. Wood, G. V. Merrett, B. M. Al-hashimi, K. O'Hara, N. Shadbolt, and W. Hall. A low-power, distributed, pervasive healthcare system for supporting memory. *System*, 2011.

S. Dieleman and B. Schrauwen. Multiscale Approaches To Music Audio Feature Learning. *Proceedings of the 14th international society for music information retrieval conference*, pages 116–121, 2013.

S. Fagerlund. Bird species recognition using support vector machines. *EURASIP Journal on Applied Signal Processing*, 2007(1):64–64, 2007.

T. Ganchev, I. Potamitis, and N. Fakotakis. Acoustic monitoring of singing insects. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 4, pages IV–721 –IV–724, 2007.

Go-Gulf. Smartphone Users Around the World - Statistics and Facts, 2012.

M. Gogala and T. Trilar. Bioacoustic investigations and taxonomic considerations on the Cicadetta montana species complex (Homoptera: Cicadoidea: Tibicinidae). *Anais da Academia Brasileira de Ciências*, 76(2):316–24, 2004.

J. Grant. Cicadetta montana. The wildlife sounds collection of the British Library, 1971.

S. Henshall. Personal communication, 2013. (Lead Entomologist for BugLife, insect charity.).

T. Hertach. Three species instead of only one: Distribution and ecology of the Cicadetta montana species complex (Hemiptera: Cicadoidea) in Switzerland. *MITTEILUNGEN-SCHWEIZERISCHE*, (2004):37–61, 2007.

J. Howe. The Rise of Crowdsourcing. *Wired Magazine*, 14(6):1–4, 2006.

International Data Corporation. Smartphone OS Market Share 2014, 2013, 2012, and 2011, 2015.

G. James. *Advanced Modern Engineering Mathematics*. Pearson Education Limited, 4 edition, 2011.

E. T. Jaynes. *Probability theory: the logic of science*. Cambridge university press, 2003.

Joint Nature Conservation Committee. UK priority species pages âĂŞ Cicadetta montana (New Forest Cicada). Technical report, 2010.

Kate E. Jones, Jon Russ, Colin Catto, Charlotte Walters, Abigel Szodoray-Paradi, Farkas Szodoray-Paradi, Elena Pandourski, Ivan Pandourski, and Tilova Pandourski. Monitoring bat biodiversity : indicators of sustainable development in Eastern Europe Darwin Initiative âĂŞ Final Report. Technical report, 2009.

S. Josefsson. The base16, base32, and base64 data encodings. 2006.

N. Kaufmann, T. Schulze, and D. Veit. More than fun and money. Worker Motivation in Crowdsourcing âĂŞ A Study on Mechanical Turk. *Proceedings of the Seventeenth Americas Conference on Information Systems*, (2009):1–11, 2011.

Agnes Kukulska-Hulme and John Traxler. *Mobile Learning. A Handbook for Educators and Trainers*, volume 57 of *The Open and Flexible Learning Series*. Routledge, 2005.

S. Kulkarni and P. Agrawal. Smartphone driven healthcare system for rural communities in developing countries. In *Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*, page 8. ACM, 2008.

P. Laiolo. The emerging significance of bioacoustics in animal species conservation. *Biological Conservation*, 2010.

J. Landrey. "IT in the Toilet" survey Press Release, 2012.

N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell. A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150, 2010.

Z. Leqing and Z. Zhen. Insect Sound Recognition Based on SBC and HMM. *2010 International Conference on Intelligent Computation Technology and Automation*, pages 544–548, 2010.

S. Lloyd. Least squares quantization in pcm. *Information Theory, IEEE Transactions on*, 28(2):129–137, 1982.

G. T. Lyle. Cicadetta montana. *The Entomologist*, 43(560):1–6, 1910.

G. T. Lyle. Further Notes on The British Cicada, Cicadetta Montana (Hemiptera). *The Entomologist*, 44:332–334, 1911.

G. T. Lyle. Cicadetta Montana in the New Forest. *The Entomologist*, 46:301–305, 1913.

J. Lyons. Mel Frequency Cepstral Coefficient, 2014.

R. G. Lyons. *Understanding Digital Signal Processing*. Prentice Hall PIR, 2004.

Norman MacLeod. *Automated Taxon Identification in Systematics: Theory, Approaches and Applications*. CRC Press, 2007.

J. McGonigal. *Reality is Broken: Why Games Make Us Better and How They Can Change the World*. Random House, 2011.

A. L. McIlraith and H. C. Card. Bird song identification using artificial neural networks and statistical analysis. In *Electrical and Computer Engineering, 1997. Engineering Innovation: Voyage of Discovery. IEEE 1997 Canadian Conference on*, volume 1, pages 63–66 vol.1, 1997.

P. Mohan, V. N. Padmanabhan, and R. Ramjee. Nericell : Rich Monitoring of Road and Traffic Conditions using Mobile Smartphones. *Proceedings of the 6th ACM conference on Embedded network sensor systems*, (c):323–336, 2008.

S. Moran, N. Pantidi, T. Rodden, A. Chamberlain, C. Griffiths, D. Zilli, G. V. Merrett, and A. Rogers. Listening to the forest and its curators: lessons learnt from a bioacoustic smartphone application deployment. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 2387–2396, 2014.

M. Mun, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, R. West, and P. Boda. Peir, the personal environmental impact report, as a platform for participatory sensing systems research. In *Proceedings of the 7th international conference on Mobile systems, applications, and services*, pages 55–68. ACM, 2009.

New Forest District Council. New Forest Tourism and Travel.

O. Nov, O. Arazy, and D. P. Anderson. Dusting for science: motivation and participation of digital citizen science volunteers. In *Proceedings of the 2011*, 2011.

M. K. Obrist and P. F. Flückiger. Variability in echolocation call design of 26 Swiss bat species : consequences , limits and options for automated field identification with a synergetic pattern recognition approach. 68(4), 2004.

Orthoptera Recording Scheme. Orthoptera and Allied Insects, 2015.

N. Pantidi, S. Moran, K. Bachour, T. Rodden, D. Zilli, G. Merrett, and A. Rogers. Field Testing a Rare Species Bioacoustic Smartphone Application : Challenges and Future Considerations. In *Third IEEE International Workshop on the Social Implication of Pervasive Computing for Sustainable Living*, 2014.

E. Papadatou, R. K. Butlin, and J. D. Altringham. Identification of bat species in Greece from their echolocation calls. *Acta Chiropterologica*, 10(1):127–143, 2008.

S. Parsons, A. M. Boonman, and M. K Obrist. Advantages and disadvantages of techniques for transforming and analyzing chiropteran echolocation calls. *Journal of Mammalogy*, 2000.

S. Parsons and G. Jones. Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *Journal of Experimental Biology*, 2000.

K. B. Payne, M. Thompson, and L. Kramer. Elephant calling patterns as indicators of group size and composition: The basis for an acoustic monitoring system. *African Journal of Ecology*, 41:99–107, 2003.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

PhoneGap Development Team. Plugin development guide. `http://docs.phonegap.com/en/edge/guide_hybrid_plugins_index.md.html#Plugin%20Development%20Guide`, 2013. [Online; accessed 21-August-2013].

B. Pinchen. Personal communication, 2012. (Coordinator of the New Forest Cicada Steering Group).

B. J. Pinchen and L. K. Ward. The history, ecology and conservation of the New Forest Cicada. *British Wildlife*, 13(4):258–266, 2002.

B. J. Pinchen and L. K. Ward. The New Forest cicada and other invertebrates. In Adrian C Newton, editor, *Biodiversity in the New Forest*, pages 58–64. Pisces Publications, 2010.

J. Pinhas, V. Soroker, A. Hetzoni, A. Mizrach, M. Teicher, and J. Goldberger. Automatic acoustic detection of the red palm weevil. *Computer and Electronics in Agriculture*, 63: 131–139, 2008.

B. Planitz, P. Roe, J. Sumitomo, M. Towsey, I. Williamson, and J. Wimmer. Listening to nature: techniques for large-scale monitoring of ecosystems using acoustics. In *3rd eResearch Australasia Conference*, 2009.

B. Plannerer. An Introduction to Speech Recognition. Technical report, IEEE, 2005.

J. C. Plantin. 'The Map is the Debate': Radiation Webmapping and Public Involvement During the Fukushima Issue. pages 1–16, 2011.

I. Potamitis, T. Ganchev, and N. Fakotakis. Automatic acoustic identification of insects inspired by the speaker recognition paradigm. In *Interspeech 2006*, pages 2126–2129, 2006.

M. J. Raddick, G. Bracey, P. L. Gay, C. J. Lintott, P. Murray, K. Schawinski, A. S Szalay, and J. Vandenberg. Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. *Astronomy Education Review*, 9(1):15, 2009.

D. R. Raggae. *Grasshoppers, Crickets and Cockroaches of the British Isles*. Frederick Warne and Co. London, 1965.

I. Rahwan, D. Sohan, A. Rutherford, V. Naroditskiy, J. Mcinerney, M. Venanzi, N. R. Jennings, and M. Cebrian. Global Manhunt Pushes the Limits of Social Mobilization. pages 1–12, 2012.

R. D. Redgwell, J. M. Szewczak, G. Jones, and S. Parsons. Classification of echolocation calls from 14 species of bat by support vector machines and ensembles of neural networks. *Algorithms*, 2(3):907–924, 2009.

K. Riede. Acoustic monitoring of Orthoptera and its potential for conservation. *Journal of Insect Conservation*, 1998.

A. Rogers. British Orthoptera Species Report. Orthoptera Recording Scheme, 2014.

J. Rogstadius, V. Kostakos, A. Kittur, and B. Smus. An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, number Gibbons 1997, pages 321–328, 2011.

M. Saladyga. AAVSO - Vision and Mission, 2012a.

M. Saladyga. AAVSO - What We Do, 2012b.

A. Sedghi. Addicted to smartphones: the latest Ofcom communications results | News | guardian.co.uk, 2011.

J. Silvertown. A new dawn for citizen science. *Trends in ecology & evolution*, 24(9): 467–71, 2009.

J. E. Simmonds, F. Armstrong, and P. J. Copland. Species identification using wide-band backscatter with neural network and discriminant analysis. *ICES Journal of Marine Science: Journal du Conseil*, 53(2):189–195, 1996.

O. Solon and N. Lanxon. We need more smartphone-based citizen science (Wired UK), 2012.

S. S. Stevens, J. Volkmann, and E. B. Newman. A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3): 185–190, 1937.

C. Storm. Hidden markov models. implementing the forward-, backward- and viterbi-algorithms using log-space and scaling. `http://users-cs.au.dk/cstorm/courses/PRiB_f12/slides/hidden-markov-models-2.pdf`, 2012. Retrieved on Aug 28, 2013.

D. Stowell and M. D. Plumbley. Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. 2014.

R. Stranders, S. D. Ramchurn, B. Shi, and N. R. Jennings. Collabmap: Augmenting maps using the wisdom of crowds. In *The Third Human Computation Workshop*, 2011.

J. Sueur. Insect Species and Their Songs. In S. Drosopoulos and M. Claridge, editors, *Insect Sounds and Communication: Physiology, Behaviour, Ecology, and Evolution*, chapter 15, pages 207–217. Taylor & f edition, 2006.

J. Sueur, D. Mackie, and J. F. C. Windmill. So small, so loud: Extremely high sound pressure level from a pygmy aquatic insect (corixidae, micronectinae). *PloS one*, 6 (6):e21089, 2011.

J. Sueur and S. Puissant. Breve note–Short note Biodiversity eavesdropping: bioacoustics confirms the presence of Cicadetta montana (Insecta: Hemiptera: Cicadidae) in France. *zoologie.umh.ac.be*, 43(1):126–128, 2007a.

J. Sueur and S. Puissant. Similar look but different song : a new Cicadetta species in the montana complex. *Zootaxa*, 68:55–68, 2007b.

J. C. Tang, M. Cebrian, N. A. Giacobe, H. Kim, T. Kim, and D. Wickert. Reflecting on the DARPA Red Balloon Challenge. *Communications of the ACM*, 54(4):78, 2011.

J. Traxler and A. Kukulska-Julme. Mobile learning in developing countries. *Commonwealth of Learning*, 2005.

T. Trilar. Personal communication, 2012. (Head of Invertebrate Department, Museum of Natural History, Ljubljana).

T. Trilar and T. Hertach. Three species of Mountain cicadas Cicadetta montana (sensu lato) found in northern Italy. *Bulletin of Insectology*, 61(1):185–186, 2008.

R. E. Turner and M. Sahani. Demodulation as Probabilistic Inference. XX(X):1–14, 2011.

N. Vaughan and G. Jones. Identification of British bat species by multivariate analysis of echolocation call parameters. *Journal Of Animal Sound And Its*, 1997.

A. J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *Information Theory, IEEE Transactions on*, 13(2):260–269, 1967.

C. L. Walters, R. Freeman, A. Collen, C. Dietz, M. Brock Fenton, G. Jones, M. K. Obrist, S. J. Puechmaille, T. Sattler, B. M. Siemers, S. Parsons, and K. E. Jones. A continental-scale tool for acoustic identification of European bats. *Journal of Applied Ecology*, 49: 1064–1074, 2012.

A. Wiggins and K. Crowston. From Conservation to Crowdsourcing: A Typology of Citizen Science. *2011 44th Hawaii International Conference on System Sciences*, pages 1–10, 2011.

I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

H. Yu, Z. Shen, C. Miao, and B. An. Challenges and opportunities for trust management in crowdsourcing. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 02*, WI-IAT '12, pages 486–493, Washington, DC, USA, 2012. IEEE Computer Society.

D. Zilli, O. Parson, G. V. Merrett, and A. Rogers. A Hidden Markov Model-Based Acoustic Cicada Detector for Crowdsourced Smartphone Biodiversity Monitoring. In *International Joint Conference for Artificial Intelligence*, Beijing, China, 2013. AAAI Press.

D. Zilli, O. Parson, G. V. Merrett, and A. Rogers. A Hidden Markov Model-Based Acoustic Cicada Detector for Crowdsourced Smartphone Biodiversity Monitoring. *Journal of Artificial Intelligence Research*, 51:805–827, 2014.