



HAL
open science

Assessment of animal acoustic diversity in neotropical forest

Juan Ulloa Chacón

► **To cite this version:**

Juan Ulloa Chacón. Assessment of animal acoustic diversity in neotropical forest. Ecology, environment. Université Paris Saclay (COMUE), 2018. English. NNT : 2018SACLS120 . tel-03221791

HAL Id: tel-03221791

<https://theses.hal.science/tel-03221791>

Submitted on 10 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation de la diversité acoustique animale en forêt néotropicale

Thèse de doctorat de l'Université Paris-Saclay préparée à
l'Université Paris-Sud et le Muséum national d'Histoire naturelle

École doctorale n°568 Signalisations et réseaux
intégratifs en biologie (BIOSIGNE)
Spécialité de doctorat: Sciences de la vie et de la santé

Thèse présentée et soutenue à Paris, le 11 Juin, par

Juan Sebastian Ulloa Chacón

Composition du Jury :

| | |
|---|-----------------------|
| Jane Lecomte Professeure, Université Paris-Saclay, France | Présidente |
| Jérôme Chave Directeur de Recherche, Université Paul Sabatier, France | Rapporteur |
| Gianni Pavan Professeur, Università di Pavia, Italie | Rapporteur |
| Pierre-Michel Forget Professeur, Muséum national d'Histoire naturelle, France | Examineur |
| Sandrine Pavoine Maître de Conférences, Muséum national d'Histoire naturelle, France | Examineur |
| Thierry Aubin Directeur de Recherche, Université Paris-Saclay, France | Directeur de thèse |
| Jérôme Sueur Maître de Conférences, Muséum national d'Histoire naturelle, France | Co-Directeur de thèse |

Titre : Estimation de la diversité acoustique animale en forêt néotropicale

Mots clés : suivi de biodiversité, écoacoustique, apprentissage automatique, écologie tropicale

Résumé : La communication par émission sonore est un trait comportemental répandu chez les animaux terrestres. Les riches textures sonores de la forêt néotropicale nous suggèrent que la faune est non seulement abondante, mais aussi diverse et dynamique. Cette facette de la biodiversité peut révéler des informations précieuses sur les communautés animales qui habitent les milieux tropicaux, mais reste largement méconnue. Comment mesurer la diversité acoustique tropicale pour aborder des questions écologiques ? Dans le cadre de l'écoacoustique, nous avons cherché à révéler des structures dissimulées dans le paysage sonore de la forêt néotropicale, et tenter d'expliquer leurs présences à travers les processus écologiques sous-jacents. Tout d'abord, nous avons suivi la dynamique spatio-temporelle d'une empreinte sonore amazonienne, le chant de l'oiseau tropical

Lipaugus vociferans, montrant une activité liée à des caractéristiques spécifiques d'habitat. Puis, nous nous sommes intéressés aux communautés d'amphibiens. L'analyse de variables acoustiques et météorologiques nous a permis de mieux comprendre les causes, patrons et conséquences du comportement reproductif explosif. Enfin, nous avons adapté de nouveaux outils de calcul, issus des disciplines de l'apprentissage automatique et de la reconnaissance de formes, pour proposer une analyse efficace, objective et facilement reproductible de grands jeux de données acoustiques. L'écoacoustique, renforcée par des algorithmes informatiques, émerge comme une approche clé pour les programmes de suivis de biodiversité à large échelle, permettant de mieux comprendre et valoriser la diversité de formes de vies unique abritée par la forêt tropicale.

Title : Estimating animal acoustic diversity in neotropical forest

Keywords : biodiversity monitoring, ecoacoustics, machine learning, tropical ecology

Abstract : Acoustic signalling is a common behavioural trait among terrestrial animals. The rich sound textures of neotropical forest echo that wildlife is not only abundant, but also diverse and dynamic. This facet of biodiversity can reveal valuable insights of animal communities inhabiting tropical environments, yet remains poorly understood. How to best measure tropical acoustic diversity to address ecological questions? Based on the ecoacoustic framework, we explored the soundscape of neotropical forest, revealing patterns and investigating the ecological underlying processes. First, we tracked the spatiotemporal dynamics of an amazonian soundmark, the song of the bird *Lipaugus vociferans*, showing

activity patterns related to specific habitat features. Then, we investigated amphibian communities with very brief reproduction periods. Coupling acoustic and environmental variables, we shed light on the causes, patterns and consequences of explosive breeding events. Finally, we adapted novel computational tools from the machine learning and pattern recognition disciplines to provide an efficient, objective and replicable analysis of large acoustic datasets. Ecoacoustics, powered with computer algorithms, emerge as a suitable approach to scale-up biodiversity monitoring programs, allowing to better understand and cherish the unique diversity of life sustained by tropical forest.



**Estimating animal acoustic diversity
in lowland neotropical forest**

*To my grand parents
and their unconditional love*

CONTENTS

| | |
|--|-----|
| Acknowledgments | 7 |
| General introduction | 9 |
| Chapter 1. Screening large audio datasets to determine the time and space distribution of Screaming Piha birds in a tropical forest | 32 |
| Chapter 2. Explosive breeding in tropical anurans: environmental triggers, community composition and acoustic structure | 62 |
| Chapter 3. Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis | 93 |
| Chapter 4. Towards an end-to-end framework for sound classification in ecoacoustic datasets | 123 |
| General discussion | 149 |
| Appendix | 170 |

ACKNOWLEDGMENTS

Many people and institutions made this study possible. I would like to begin by acknowledging Jérôme Sueur and Thierry Aubin for the wealth of advice and feedback they gave me during my research. I had the chance to work side by side with two brilliant and thoughtful supervisors and I'm deeply grateful for the opportunity they gave me. I hope that the research pursued with them in France, will be just the beginning of many other projects and collaborations, at the rhythm of *currulao*, *cumbia* and *guaguancó* in Colombia.

I would like to thank the members of the jury for kindly accepting to assess my thesis. In particular, I would like to thank Jérôme Chave and Gianni Pavan for accepting the hard work of reviewing the manuscript, and Sandrine Pavoine, Pierre-Michel Forget and Jane Lecomte for accepting to participate in the evaluation process. I would also like to thank the members of my thesis committee, Olivier Adam, Charles Bouveyron and Cedric Gervaise, their valuable advices certainly focused my research on the best path.

Various institutions made my research possible. This work was supported by COLCIENCIAS (doctorados en el exterior, convocatoria 646 de 2014), SABIOD MASTODONS Big Data (CNRS MI project 2012-2017), Labex CEBA (Centre d'Étude de la Biodiversité Amazonienne, 2015), the *Institut de Systématique Evolution et Biodiversité* (ISYEB) of the *Muséum national d'Histoire naturelle*, and the *Équipe Communication Acoustique* of the Université Paris-Sud.

During my journey I had the chance to assist to multiple conferences, seminars and scientific meetings where I encountered many who were generous with their time and expertise. In particular, I would like to thank Dan Stowell, Hervé Glotin, Stéphane Mallat, Sandrine Pavoine, Charles Bouveyron, Maxime Réjou-Méchain, Antoine Fouquet, Diego Llusia, Laure Desutter-Grandcolas, Yves Bas, Amandine Gasc, and Alexandra Rodriguez.

Fieldwork in the lowland tropical rainforest was a living dream. This was largely because I was fortunate in having people that assisted me during my journey. For their hospitality, I would like to thank Philippe Gaucher, Elodie Courtois, Dorothee Deslignes, from the CNRS Cayenne, Patrick Châtelet, from the Nouragues Research Station, and Jeannot and Odette, from the *Gîte Patawa*.

Finding a place to live in Paris can be hazardous, yet can lead you to find some amazing people. I would like to thank Claudia Valencia, Bérénice Sevestre, Benoit, Harbyn Patiño, Daniel Donneys, Gisela Gottschalk, Lorena Villegas and Hugo Sacuto, their companionship, tolerance and encouragement certainly led me to give the best on my research. In addition, I would like to thank the members of *La Tromba Futbolera* and *El Pájaro Rojo*, rainy, windy, icy or sunny, under all circumstances a good plan was always assured with them.

I am grateful to the *Institut de Systématique Evolution et Biodiversité* (ISYEB) of the *Muséum national d'Histoire naturelle* (MNHN) for hosting me during my PhD. Here, I had the chance to learn every day from the numerous seminars, conferences and expositions. I also had the luck of meeting great friends, Mónica Arias, Diego Llusia, Camille Desjonquères, Maram Caesar, Pablo Bolaños, Bruno Dastillung, Sylvain Hauptert, Manon Ducrettet, Vitor Dias Tarli, Adrienne Gastineau, Ana Cristina Torres, Gabrielle Martin and Léo Martin. I was also warmly welcomed, and invited to stay as long as I felt at the Orsay laboratory *Equipe de Communication Acoustique*. I would like to thank Isabelle Charrier, Hélène Courvoisier, Chloé Huetz, Fanny Rybak and Olivier Adam.

Finally, the continual support of my family and my parents, long before I started this research, was certainly the key ingredient of this work. They showed me how to work with dedication and love.

GENERAL INTRODUCTION

"Close your eyes, prick your ears, and from the softest sound to the wildest noise, from the simplest tone to the highest harmony, from the most violent, passionate scream to the gentlest words of sweet reason, it is by Nature who speaks, revealing her being, her power, her life, and her relatedness so that a blind person, to whom the infinitely word is denied, can grasp an infinite vitality in what can be heard"

– Johann Wolfgang von Goethe (translated by Wulf, 2015)

1. Sound, a window to the animal community

The sound of tropical environments has long been fascinating the human mind. Some of the detailed written impressions come from early European explorers who noted a rich and vibrant soundscape. Alexander von Humboldt, after his expedition through the Orinoco river with Aimé Bonpland, wrote in his *Personal Narrative of a Journey to the Equinoctial Regions of the New Continent*: "When we lend an attentive ear to the most feeble sounds transmitted by the air we hear a dull vibration, a continual murmur, a hum of insects, that fill, if we may use the expression, all the lower strata of the air. Nothing is better fitted to make man feel the extent and power of organic life. Myriads of insects creep upon the soil, and flutter round the plants parched by the ardour of the Sun. A confused noise issues from every bush, from the decayed trunk of trees, from the clefts of the rocks, and from the ground undermined by the lizards, millipedes, and cecilians. There are so many voices proclaiming to us, that all nature breathes; and that under a thousand different forms, life is diffused throughout the cracked and dusty soil, as well as in the bosom of the waters, and in the air that circulates around us" (Humboldt and Bonpland, 1853).

Florence Hercule, a French illustrator embarked on the Langsdorff expedition into the Amazon river mouth in the early 19th century, noted that the multiplicity of beating patterns pouring from the forest were a fundamental part of the animal communities (Toledo and de Araujo, 2017). Not satisfied with fine silhouettes and

colourful illustrations of the natural world, he developed a method to apprehend the sounds of multiple birds, amphibians and insects. A method that he called Zoophonia (Florence, 1876). His field notes might be the most accurate sound chronicles of the Amazonian and Atlantic forests 200 years ago.

Not surprisingly, indigenous people living in tropical forests have a tacit but deep understanding of the soundscape, an intimate link with sound which ensures their survival. In an ethnobiological work, Lescure (1980) noted that the Wayapi, native Americans of the Guyanese region, classified amphibians according to their calls, a classification that closely matched the taxonomical studies. Lescure (1980) also noted that amphibians have not only a symbolic role, but they had also an ecological role. Since amphibians are sensitive to minuscule variations of humidity, they could announce the weather and the seasons, thereby, the call of the toad *Leptodactylus pentadactylus* would mark the proximity of the rainy season. The sounds of the forest were not only a reference, a soundmark, but also a highlight in their imagination, in their stories, in their ways to understand their environment.

*

The acoustic environment is a window to study the animal communities that inhabit the tropical environments. The numerous insect, amphibian, bird and mammal sounds are mixed down into a single time series. What factors have shaped the tropical acoustic environments and what does that tells us about the ecological communities that inhabit these habitats? How to decipher the tapestry of beating patterns in order to quantify and characterise the acoustic environment? In other words, how to link acoustic with ecology to get valuable indicators for science and conservation?

2. Tropical ecology and conservation

The diversity of life is distributed heterogeneously on Earth (Gaston, 2000). One of the most striking pattern, known as the latitudinal diversity gradient (LDG), is the increase in number of species when moving from polar to equatorial regions. Between the Cancer and Capricorn latitudes around the globe, tropical environments teem with life.

To date, studies estimate that tropical rainforests may hold more than half of the Earth biodiversity with only 7% of its surface (Kricher, 2011; Myer, 1988). Although ecologists still dispute what are the main drivers of the spatial pattern of species richness, it is undeniable that tropical environments present complex biological interactions imposing great challenges and opportunities for science and conservation biology.

The unparalleled diversity of tropical environments puts forward the forces that shape biological diversity. Indeed, tropical environments presents unique opportunities to experience and understand how organisms exist, adapt and interact with their biotic and abiotic environment. It is therefore not surprising that tropical environments have lured several generations of biologists, including Alexander von Humboldt, Charles Darwin, Alfred Russel Wallace and Henry Walter Bates. Among others, it was here that Humboldt and Aimé Bonpland observed that plants and animal communities changed as climate, by elevation, what is now termed *life zone concept* (Holdridge, 1967); here that Bates realized that unrelated species of butterflies, palatable and unpalatable, can look alike to gain protection from predators, a form of mimicry now termed *Batesian mimicry* (Waldbauer, 1988). In tropical environments biotic constraints are better evidenced than in temperate zones. Our understanding of biological forces would not be the same without tropical ecosystems. As Jansen (1986) noted more than 30 years ago, the future of ecology depends on how we as humans manage to preserve tropical ecosystems.

From the fundamental oxygen we breath and the freshwater we drink, to the invaluable aesthetic pleasure of tropical landscapes, humans well-being rely on tropical biodiversity. To highlight the amenities that we get from nature for granted, conservationists have framed the term ecosystem services (Daily, 1997; Balmford et al., 2002; Hassan et al., 2005). The list is long for tropical ecosystems, which harbour most of the resources that humans rely on and provide fundamental regulation of biogeochemical cycles. Plant and animals, including humans, depend upon the well functioning of ecosystems for their survival and well being, and probably nowhere else this is more evident than in the tropics.

Unfortunately, the concentrated resources that bring forth tropical ecosystems have become its torment. Tropical forest now suffer from human pressures, with destruction and degradation at exceptional and accelerating rates (Morris, 2010).

Alteration of land use (Sala, 2000), climate and biogeochemical cycles are only some examples of the major drivers of biodiversity loss. The need to understand biological communities and how to manage trade-offs between immediate human needs and the integrity of ecosystems seems more urgent than ever (Kricher, 2011).

To counter global biodiversity loss, there have been several international political commitments. The Aichi Targets for 2020 by parties of the United Nations and The Convention on Biological Diversity (CBD) might be the most notable examples (CBD, 2010). Averting biodiversity change and reducing the rate of biodiversity loss are the main targets envisaged. But, as mentioned by Purvis and Hector (2000) "we cannot even begin to look at how biodiversity is disturbed, or how fast is disappearing unless we can put units on it". In other words, assessing progress towards these commitments requires the development of effective methods to systematically and repeatedly measure and map changes in biodiversity. Such monitoring would allow not only to set protected areas based on priorities for species and ecosystems, but also to better understand what are the underlying causes of biodiversity loss and how to formulate effective mitigation actions.

3. Biodiversity assessment

It is impossible to conceive ecology or conservation biology without a systematic and standardized collection of data. Monitoring refers to the process of collecting periodic assessment of a system state at multiple points in space and time, which might lead to draw inferences about changes in system state over time (Yoccoz et al 2011). Related to biological diversity, the system of interest can be framed at multiple scales (species, population, communities or ecosystem), and the variable of interest might include one of the several measures of biological diversity (Magurran 2004; Pavoine et al., 2005; Gotelli and Chao, 2013). To date, many programs for biodiversity monitoring have been implemented to understand and manage ecosystems, integrating science and conservation to improve future decisions. Monitoring is a fundamental part of the scientific process, continuously adapting to the available technology, allowing to tackle new and more challenging questions (Collen et al., 2013; Chave, 2008).

3. 1. *Human-based surveys*

Traditionally, ecological data has been gathered by manual field surveys. In this type of monitoring, the data collected is sensed by the multisensory human perception integrated with our capacity for abstraction. When coupled with appropriate training and standardized methods, manual field survey constitute the most comprehensible way to monitor biodiversity. Indeed, this has been the main data collection process that led to the fundamental ecological hypotheses under discussion to date. Unfortunately, manual field surveys by experts are not scalable to large temporal and spatial dimensions. With the emergent electronic technology, research have focused on the use of autonomous sensors to capture different facets of plant and animal diversity from both perspectives, *ex situ* and *in situ*.

3. 2. *Remote ex situ sensing*

Powered by satellites or aircrafts equipped with devices to sense electromagnetic waves, remote sensing seems an ideal tool to gather data at global scale (Lillesand et al., 2015). Remote sensing provides a systematic and synoptic view of the Earth at regular intervals, and has been widely used to estimate biotic and abiotic factors to model biodiversity (Turner et al., 2003; Geller et al., 2017). Satellite images can provide estimates on multiple habitat variables, such as topography, soil, climatic conditions and land cover, allowing to draw inferences about present and future distribution of species (Nagendra, 2001; Nagendra et al., 2013). With the increasing resolution available, it is even possible to directly identify plant species and communities based upon unique spectral signatures. Remote sensing thus plays a central role in biodiversity conservation applications (Pereira et al., 2013; Geller et al., 2017). Yet, it is clear that models derived from remote sensing need to have accurate species occurrence information that has to be collected in the field. Remote *ex situ* sensing needs to be networked with *in situ* observations.

3. 3. In situ *biodiversity sensing*

The use of automated sensors for *in situ* data collection is a recent area that has flourished with the miniaturisation of electronic devices, making embedded systems increasingly affordable and autonomous. Transmission tags (Rutz and Hays, 2009), collection on environmental DNA (Bohmann et al., 2014) and camera traps (Steenweg et al., 2017) can provide solutions to logistical constraints, allowing to increase the temporal and spatial scales of biodiversity monitoring (Le Galliard et al., 2012; Turner, 2014). These *in situ* observations provide information at the level of genes, species, communities and landscape at the fine-scale that remains hidden to remote sensing. Another recent, innovative and non-invasive approach is to track animals using passive acoustic sensors. This is the heart of this work, and the topic of the upcoming sections.

4. Acoustic monitoring

Acoustic signalling is a common behavioural trait among terrestrial animals, such as insects, amphibians, birds and mammals. These signals are species specific and propagates through the environment, which can be employed as a direct way to retrieve ecological data about species presence, abundance, status and distribution of animal communities (Obrist et al., 2010; Browning et al., 2017). In addition, some animals use their calls and the resulting echoes to detect, localize, and classify objects, so that they can forage and navigate through their territories. The acoustic environment is full of information about the animal communities, yet how to best exploit this information-rich data in an effective way requires theoretical and practical knowledge.

Acoustic cues have been used extensively for monitoring a broad range of vocal species. In terrestrial habitats, protocols to study bats, birds and amphibians have usually included species-specific calls as a fundamental evidence to determine presence or absence of taxa. The American Amphibian Monitoring Program (Weir and Mossman, 2005), the Tropical Ecology Assessment and Monitoring network to the bird monitoring protocol (Lacher, 2008), the Acoustic Monitoring of Night-Migrant Birds (Farnsworth and Russell, 2007), and the Indicators Bats Program (iBats; Jones et al., 2013), are just

some examples of the established protocols and increasing interest to survey animals using sound.

With the development of autonomous acoustic sensors, monitoring through acoustics can be applied to cover larger scales with lower human effort (Acevedo and Villanueva-Rivera, 2006; Obrist et al., 2010; Servick, 2014). These new digital acoustic sensors are weather proof and have low energy consumption, allowing to deploy them in the field for long periods of time. Usually, microprocessors are equipped with omnidirectional microphones, batteries, memory cards, and a digital clock to settle recording schedules. While sometimes continuous recording can be useful, the use of programmed recording periods allows to limit the energy and memory consumption, hence extending the sampling period. Coupled with solar panels and wireless transmission of data, such devices are increasingly autonomous. With lowering prizes of technology, arrays of passive acoustic sensors are now being deployed worldwide to capture temporal and spatial dynamics of animal communities (Browning et al., 2017).

Such advances technology and tools also require a conceptual framework to move forward. Ecoacoustics has recently emerged as a new discipline which searches for links between ecology and acoustics at different organisation levels in terrestrial, freshwater and marine ecosystems (Sueur and Farina 2015). Enlarging the possibilities of acoustic monitoring to assess presence-absence of species, ecoacoustics offers a conceptual framework, such as the acoustic adaptation hypothesis (Marten et al., 1977; Morton, 1975) or the acoustic niche hypothesis (Krause, 1993), which take into account also the shifts of calling patterns and interactions between species calls within a community. Moreover, the analysis of the effects of anthropogenic noise pollution are also an important contribution of ecoacoustics for conservation biology (Pavan, 2017). In summary, ecoacoustics is a discipline that expands the possibilities of biodiversity assessments to decipher the detailed information of the acoustic environment (Farina and Gage, 2017).

In tropical environments, ecoacoustic studies have revealed increasing possibilities for tracking multiple taxa and understanding the widespread acoustic signalling behaviour of animals. For example, acoustic sensors helped to assess links between anuran calling activity and changes in precipitation and temperature (Ospina et al., 2013), estimate occupancy of wild primates populations (Kalan et al., 2015), estimate elephant population size (Thompson et al., 2010), and detecting anthropogenic

noises associated with poaching (Wrege et al., 2017). In studies of multispecies acoustic assemblages, organisation with spectral segregation has been documented in tropical crickets (Schmidt et al., 2013), cicadas (Sueur, 2002) and frogs (Villanueva-Rivera, 2014), nevertheless, clustering patterns in tropical bird songs and bird signalling behaviour have also been documented, probably revealing inter-species complex communication network in neotropical forests (Tobias et al., 2014). At a more holistic and global approach, tropical acoustic communities were found to be heterogeneous in space and subtly structured in time (Pekin et al. 2012; Rodriguez et al. 2014). In particular, a diel pattern was found with a pronounced difference in intensity between the day and the night, the night having higher levels of activity.

Little by little the tropical acoustic environment is being unfolded, revealing new questions for science and new tools for conservation practices, but we are just decrypting the tip of the surface. With data pouring from passive acoustic sensors, the collection of sounds labelled as unclassified accumulates, evidencing that this facet of animal communities is still poorly understood. The raw data is continuously being collected, yet the analysis of such data remains a prominent hindrance for the wider application of acoustic sensors in biodiversity monitoring.

5. Analysis of passive audio recordings

Animal communities living in tropical environments produce complex sounds. The multiple sound sources are scattered in space and the signals show great deal of variety that often overlap. In addition, the intricate habitat structure modifies the signal by absorption, reverberation, and scattering (Marten and Marler, 1977; Richards and Wiley, 1980). Locating and identifying sounds of interest in outdoor recordings collected by acoustic sensors is a challenging task that can be done either manually or automatically.

5. 1. Manual analysis

In manual analysis, the recordings are usually scanned aurally and visually, using computer software that provides ways to play back the sound and a to visualise it

as a spectrogram, a time x frequency x amplitude graphical projection. Manual analysis by a skilled biologist, with extensive field-experience, can lead to precise transcriptions of field recordings into formatted data that can be exploited by standard numerical ecology methods, for instance a community presence-absence matrix. Such approach, even if tedious and time-consuming, has been the primary way to decrypt audio recordings.

As in any human observational process, the observer is a central figure and the quality of the data depends on his or her skills and experience, generally known as inter-observer variability. Inter-observer variability can take different forms, from mechanical errors during annotation to perceptual errors that can lead to spurious inferences (Fitzpatrick et al., 2009). While observer bias can be reduced with appropriate training, the number of experts seems extremely limited compared to the deluge of acoustic data collected. It would require large numbers of highly motivated assistants, and finding and training them on a regular basis seems difficult. These constraints motivated the emergence of automated alternatives.

5. 2. Automated analysis

For large datasets, such as outdoor recordings coming from passive acoustic sensors, the analysis can be facilitated in multiple ways using automated computational tools. A varied set of general tools developed for speech and music analysis is readily available. Yet, most of these tools were particularly designed to analyse man-made environments and can not be translated directly to analyse audio recordings made in nature environments. When analysing outdoor audio recordings, specific challenges arise: animal sound sources can (1) have a higher frequency range than human hearing, (2) show a great variation in their signal-to-noise ratio due to background noise and signal overlapping, (3) be mobile and (4) be very diverse within and between species including differences in overall amplitude, frequency modulations, amplitude modulations, temporal pattern and frequency content.

The cross-correlation of spectrograms has been proved to be of important use in ecoacoustics (Brunelli, 2009). Cross-correlation is simple measure of similarity between a target sound and an audio recording taken at different time lags. Since a variety of animal sounds are stereotyped, this approach can be effective in varied situations.

However, when the sound sources vary, or when the ambient noise generates multiple false positive detections, more sophisticated techniques should be envisaged.

Based on artificial intelligence, machine learning deals with the development of statistical classifiers that can learn from and make prediction on data. Machine learning algorithms offer a world of possibilities that have proved to rise to another level a broad range of scientific areas that also have to deal with massive amounts of data, such as genomics (Libbrecht and Noble, 2015), astronomy (Way, 2012) and medicine (Wernick et al., 2010). Machine learning techniques are categorized into supervised and unsupervised learning (Bishop, 2006). In supervised learning the statistical model is build using labelled data, which is obtained by manually annotations. Unsupervised learning, or clustering, is more flexible since it does not require labelled data, instead the classification is made based only on the data attributes.

In the field of bio- and ecoacoustics the use of machine learning tools is a more recent approach that is developing fast (Stowell, 2018). Current methods to detect and classify animal sounds use supervised learning. First approaches focused on the classification of sound segments. These approaches adapted techniques coming from speech recognition research (Skowronski and Harris, 2006). While these algorithms proved to deliver high performances on signals with high signal to noise ratio, they performed poorly with field recordings, which have varying signal to noise ratio and a multitude of interferences at multiple frequencies. With the growing data coming from unattended recordings, the community has more recently focused on analysing such scenarios.

The combination of acoustic sensors and supervised learning classification has allowed to track dynamics of a wide range of taxa, including amphibians (Aide et al., 2013), primates (Heinicke., 2015), birds (Goyette et al., 2011), elephants (Wrege et al., 2010) and bats (Walters et al., 2012). Yet, the methods envisaged were developed separately and is still unclear how to design a standardized framework to adapt these techniques to other scenarios. Moreover, most of these approaches were developed by large projects that already accounted with manual labelled datasets to train the classifiers. While it seems to be possible to find automatically sounds of interest from outdoor audio recordings, the multiplicity of methods clearly indicates that the trail is devious.

5.3. *Acoustic indices*

The acoustic environment can be characterised and analysed at a global scale, without subdividing the sound scene. Researchers realized that the structure of the acoustic environment that emerge from mixed signals incorporates information and could be characterised by acoustic indices (Sueur et al., 2014; Pieretti et al., 2011). Acoustic indices were designed to assist ecological investigation, allowing to track global dynamics of the acoustic environment without identifying species-specific calls. Following the biodiversity indices from ecological literature, two classes of indices were derived: within-group diversity (alpha indices) and between group diversity (beta indices). These indices aim to compute objective acoustic parameters related a characteristic of the acoustic community, such as amplitude, evenness, richness or heterogeneity.

Acoustic indices have been applied to multiple environments, including terrestrial (Depraetere et al., 2012; Lellouch et al., 2014; Towsey et al., 2014) and underwater scenarios (Harris et al., 2016). They showed to correlate to changes in bird species richness in woodland habitats (Depraetere et al., 2012) and spatio-temporal dynamics of the soundscape in tropical forest (Rodriguez et al., 2014). The indices have the advantage of being easy and fast to compute, they hence provide a rapid overview of the data captured by acoustic sensors. However, they also showed to be sensitive to transitory or permanent background noise, variation of calling rate and distance of the signalling animals, which makes indices hard to interpret and link with ecological processes. Studies related to acoustic indices open new perspectives to analyse passive acoustic recordings and emphasizes that the acoustic environment is more than the sum of its individual parts.

*

The diversity of methods to analyse field recordings that have been proposed show the complexity, but also the rich amount of information that can be derived from passive acoustic recorders. Such diversity of methods might seem confusing, detracting a large number of users from using acoustic sensors. Indeed, there is still limited standard protocol to analyse passive acoustic recordings (Knight et al., 2017), and hence this task

is limited to a personnel with high skills managing massive volumes of data (Browning et al., 2017). In addition, sounds heard on tropical environments are still poorly documented, which pose particular challenges to implement automated methods to analyse field recordings.

6. Working hypotheses

The aim of this work is to propose new ways to assess patterns of biodiversity at multiple organization levels in the tropical environments using passive acoustic sensors. Based on standardised and systematic passive acoustic monitoring, we tackled key ecological questions related to animal acoustics at the population and community levels. Furthermore, we focused on important methodological aspects regarding the efficient, objective and accurate analysis of acoustic recordings.

Passive acoustic monitoring seems to be a valuable tool to monitor rich animal communities of the tropical forest at large temporal and spatial scales. Yet, multiple challenges remain to be addressed. How to better quantify the acoustic environment? How to get valuable ecological information from acoustic sensors? Facing the data deluge from acoustic sensors, how to adapt rapid techniques to analyse the data? Moreover, how to address poorly documented environments such as the tropical forest? We to shed light on these questions, giving possible perspectives to best monitor and understand biodiversity in tropical environments through ecoacoustic monitoring. This manuscript is based on four research studies pursued in the lowland tropical forest of French Guyana.

First, we studied one of the most remarkable and distinguished animal sounds in the Guyanese and Amazonian rainforests, the song of the bird *Lipaugus vociferans*. This emblematic sound provides a prominent acoustic relief for the travellers of the dense forest. While the sound can be easily distinguished by humans, this song is mixed down with the tropical acoustic environment which is replete with intricate animal sounds. In Chapter 1, we asked if it was possible to track a population of *L. vociferans* coupling acoustic sensors and pattern recognition algorithms. The results of the acoustic monitoring were confronted with hydrology and vegetation maps, providing novel ecological information on this tropical bird.

Second, we examined the dynamics of the acoustic diversity at the community level. Neotropical rainforest are known to hold the highest diversity of amphibians and the calls of such communities are an essential mark of seasons in tropical acoustic environments. Since amphibians are the most endangered group of vertebrate, designing new techniques to understand and forecast impacts on this taxon is an urgent task. In Chapter 2 we investigated how these communities can be tracked using acoustic sensors. In particular we focused on the elusive explosive breeding communities and asked, what are the causes, patterns and consequences of such unique communities? We collected for the first time acoustic and environmental data to monitor simultaneously and regularly multiple explosive breeding events in tropical anuran communities, revealing coordinated changes in the anuran community at multiple spatiotemporal scales which was tightly linked to particular environmental patterns. The co-occurrence of several signalling species in such a tight spatio-temporal window further allowed us to reveal acoustic signatures at the community level with potential heterospecific signalling functions.

Third, we searched for computational methods to estimate automatically the acoustic diversity in passive acoustic recordings. Tropical acoustic environments are rich in sound shapes and textures, but the large majority these sounds have not being referenced. How to get valuable ecological information from environmental sounds that are unknown? In Chapter 3, we propose an new method, named Multiresolution Analysis of Acoustic Diversity (MAAD), to automate the detection of relevant structure in audio data. MAAD design adapts tools from the unsupervised learning field and aims to decompose the acoustic community into few elementary components (soundtypes) based on their time-frequency attributes. The method proved to be robust, deriving very similar partitions compared to human annotations in two distinct tropical environments. Such framework, based on unsupervised learning techniques, opens new perspectives for ecoacoustic monitoring in poorly documented habitats.

And fourth, we investigated multiple strategies to search for sounds of interest in audio recordings. Presence-absence data is a common format to analyse populations and communities in ecology. While the number of tools available for pattern recognition are rapidly increasing, there is still little practical guidance for the application of such algorithms for a broad audience. There remains the need of an end-to-end generic framework to classify sounds, from raw recordings to file presence-absence that

delivers accurate estimates of detection performance. In Chapter 4, we addressed the question: how to best estimate the presence or absence of target species from passive acoustic recordings? We raised particular attention to the sampling method to build datasets employed for tuning and testing statistical classifiers. In addition we compared manual, semi-supervised and supervised learning methods to analyse the data. We revealed strengths and weakness that shed light on how to combine human reasoning and computer algorithms in a standardized framework to deliver robust and accurate estimates of presence-absence.

The manuscript concludes with an overview of the current possibilities that acoustic monitoring offers to address ecological questions as well as conservation issues.

References

- Acevedo, M.A., Villanueva-Rivera, L.J., 2006. Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. *Wildlife Society Bulletin* 34, 211–214.
- Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1, e103. <https://doi.org/10.7717/peerj.103>
- Balmford, A., Bruner, A., Cooper, P., Costanza, R., Farber, S., Green, R.E., Jenkins, M., Jefferiss, P., Jessamy, V., Madden, J., Munro, K., Myers, N., Naeem, S., Paavola, J., Rayment, M., Rosendo, S., Roughgarden, J., Trumper, K., Turner, R.K., 2002. Economic reasons for conserving wild nature. *Science* 297, 950–953. <https://doi.org/10.1126/science.1073947>
- Bishop, C., 2006. *Pattern Recognition and Machine Learning, Information Science and Statistics*. Springer-Verlag, New York.
- Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Yu, D.W., de Bruyn, M., 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution* 29, 358–367. <https://doi.org/10.1016/j.tree.2014.04.003>
- Browning, E., Gibb, R., Glover-Kapfer, P., Jones, K.E., 2017. *Passive acoustic monitoring in ecology and conservation (No. 1(2))*, WWF Conservation Technology Series. WWF-UK, Woking, United Kingdom.
- Brunelli, R., 2009. *Template matching techniques in computer vision: theory and practice*. John Wiley and Sons, Ltd.
- Chave, J., 2008. Spatial variation in tree species composition across tropical forests: pattern and process, in: Carson, W.P., Schnitzer, S.A., Smithsonian Tropical Research Institute (Eds.), *Tropical Forest Community Ecology*. Wiley-Blackwell Pub, Chichester ; Malden, MA.
- Convention on Biological Diversity, 2010. Decision X/2, The Strategic Plan for Biodiversity 2011–2020 and the Aichi Biodiversity Targets. Nagoya, Japan.
- Daily, G.C. (Ed.), 1997. *Nature's services: societal dependence on natural ecosystems*. Island Press, Washington, DC.

- Depraetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S., Sueur, J., 2012. Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland. *Ecological Indicators* 13, 46–54.
- Farina, A., Gage, S.H. (Eds.), 2017. *Ecoacoustics: The Ecological Role of Sounds*. John Wiley & Sons, Ltd, Chichester, UK. <https://doi.org/10.1002/9781119230724>
- Farnsworth, A., Russell, R.W., 2007. Monitoring flight calls of migrating birds from an oil platform in the northern Gulf of Mexico. *Journal of Field Ornithology* 78, 279–289. <https://doi.org/10.1111/j.1557-9263.2007.00115.x>
- Fitzpatrick Matthew C., Preisser Evan L., Ellison Aaron M., Elkinton Joseph S., 2009. Observer bias and the detection of low-density populations. *Ecological Applications* 19, 1673–1679. <https://doi.org/10.1890/09-0265.1>
- Florence, H., 1876. Zoophonia. *Revista Trimestral do Instituto Histórico e Ethnographico do Brasil* XXXIX.
- Gaston, K.J., 2000. Global patterns in biodiversity. *Nature* 405, 220–227. <https://doi.org/10.1038/35012228>
- Geller, G.N., Halpin, P.N., Helmuth, B., Hestir, E.L., Skidmore, A., Abrams, M.J., Aguirre, N., Blair, M., Botha, E., Colloff, M., Dawson, T., Franklin, J., Horning, N., James, C., Magnusson, W., Santos, M.J., Schill, S.R., Williams, K., 2017. Remote Sensing for Biodiversity, in: *The GEO Handbook on Biodiversity Observation Networks*. Springer, Cham, pp. 187–210. https://doi.org/10.1007/978-3-319-27288-7_8
- Gotelli, N.J., Chao, A., 2013. Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data, in: *Encyclopedia of Biodiversity*. Elsevier, pp. 195–211.
- Goyette Jennifer L., Howe Robert W., Wolf Amy T., Robinson W. Douglas, 2011. Detecting tropical nocturnal birds using automated audio recordings. *Journal of Field Ornithology* 82, 279–287. <https://doi.org/10.1111/j.1557-9263.2011.00331.x>
- Harris, S.A., Shears, N.T., Radford, C.A., 2016. Ecoacoustic indices as proxies for biodiversity on temperate reefs. *Methods Ecol Evol* 7, 713–724. <https://doi.org/10.1111/2041-210X.12527>
- Hassan, R.M., Scholes, R.J., Ash, N., Millennium Ecosystem Assessment (Program) (Eds.), 2005. *Ecosystems and human well-being: current state and trends: findings of the Condition and Trends Working Group of the Millennium Ecosystem*

- Assessment, The millennium ecosystem assessment series. Island Press, Washington, DC.
- Heinicke, S., Kalan, A.K., Wagner, O.J.J., Mundry, R., Lukashevich, H., Kühl, H.S., 2015. Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods in Ecology and Evolution* 6, 753–763. <https://doi.org/10.1111/2041-210X.12384>
- Holdridge, L.R., 1967. Life zone ecology. Life zone ecology.
- Humboldt, A. von, Bonpland, A., 1853. Personal Narrative of Travels to the Equinoctial Regions of America: During the Years 1799-1804. Henry G. Bohn.
- Jones, K.E., Russ, J.A., Bashta, A.-T., Bilhari, Z., Catto, C., Csoz, I., et al., 2013. Indicator Bats Program: A System for the Global Acoustic Monitoring of Bats, in: Biodiversity Monitoring and Conservation: Bridging the Gap between Global Commitment and Local Action. Wiley-Blackwell, Oxford, UK, pp. 213–247.
- Kalan, A.K., Mundry, R., Wagner, O.J.J., Heinicke, S., Boesch, C., Kühl, H.S., 2015. Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. *Ecological Indicators* 54, 217–226. <https://doi.org/10.1016/j.ecolind.2015.02.023>
- Knight, E.C., Hannah, K.C., Foley, G.J., Scott, C.D., Brigham, R.M., Bayne, E., 2017. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conservation and Ecology* 12. <https://doi.org/10.5751/ACE-01114-120214>
- Krause, B.L., 1993. The niche hypothesis: a virtual symphony of animal sounds, the origins of musical expression and the health of habitats. *The Soundscape Newsletter* 6, 4–6.
- Kricher, J., 2011. Tropical Ecology. Princeton University Press.
- Lacher, T., 2008. Tropical Ecology, Assessment and Monitoring (TEAM) Network. Avian Monitoring Protocol Version 3.1. Conservation International, Washington, DC.
- Le Galliard, J.-F., Guarini, J.-M., Gaill, F., Institut Écologie et environnement (France), 2012. Sensors for ecology: towards integrated knowledge of ecosystems. CNRS, Paris.

- Lellouch, L., Pavoine, S., Jiguet, F., Glotin, H., Sueur, J., 2014. Monitoring temporal change of bird communities with dissimilarity acoustic indices. *Methods in Ecology and Evolution* 5, 495–505. <https://doi.org/10.1111/2041-210X.12178>
- Lescure, J., Grenand, F., Grenand, P., 1980. Les amphibiens dans l'univers Wayàpi. *Journal d'agriculture traditionnelle et de botanique appliquée* 27, 247–261.
- Libbrecht, M.W., Noble, W.S., 2015. Machine learning applications in genetics and genomics. *Nat Rev Genet* 16, 321–332. <https://doi.org/10.1038/nrg3920>
- Magurran, A.E., 2004. *Measuring biological diversity*. Blackwell Pub, Malden, Ma.
- Marten, K., Marler, P., 1977. Sound transmission and its significance for animal vocalization. *Behav Ecol Sociobiol* 2, 271–290. <https://doi.org/10.1007/BF00299740>
- Marten, K., Quine, D., Marler, P., 1977. Sound transmission and its significance for animal vocalization. *Behav Ecol Sociobiol* 2, 291–302. <https://doi.org/10.1007/BF00299741>
- Morris, R.J., 2010. Anthropogenic impacts on tropical forest biodiversity: a network structure and ecosystem functioning perspective. *Phil. Trans. R. Soc. B.* 365, 3709–3718. <https://doi.org/doi:10.1098/rstb.2010.0273>
- Morton, E.S., 1975. Ecological sources of selection on avian sounds. *The American Naturalist* 109, 17–34.
- Myer, N., 1988. Tropical forests and their species, in: *Biodiversity*. National Academy Press.
- Nagendra, H., 2001. Using remote sensing to assess biodiversity. *International Journal of Remote Sensing* 22, 2377–2400. <https://doi.org/10.1080/014311601300229872>
- Nagendra, H., Lucas, R., Honrado, J.P., Jongman, R.H.G., Tarantino, C., Adamo, M., Mairota, P., 2013. Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators* 33, 45–59. <https://doi.org/10.1016/j.ecolind.2012.09.014>
- Obrist, M.K., Pavan, G., Sueur, J., Riede, K., Llusia, D., Márquez, R., 2010. Bioacoustics approaches in biodiversity inventories, in: Eymann, J., Degreef, J., Häuser, C., Monje, J.C., Samyn, Y., Vandenspiegel, Y.D. (Eds.), *Manual on Field Recording Techniques and Protocols for All Taxa Biodiversity Inventories and Monitoring*. ABC Taxa, pp. 18–48.

- Ospina, O.E., Villanueva-Rivera, L.J., Corrada-Bravo, C.J., Aide, T.M., 2013. Variable response of anuran calling activity to daily precipitation and temperature: implications for climate change. *Ecosphere* 4, art47. <https://doi.org/10.1890/ES12-00258.1>
- Pavan, G., 2017. Fundamentals of Soundscape Conservation, in: Farina, A., Gage, S.H. (Eds.), *Ecoacoustics*. John Wiley & Sons, Ltd, Chichester, UK, pp. 235–258. <https://doi.org/10.1002/9781119230724.ch14>
- Pavoine, S., Ollier, S., Pontier, D., 2005. Measuring diversity from dissimilarities with Rao's quadratic entropy: Are any dissimilarities suitable? *Theoretical Population Biology* 67, 231–239. <https://doi.org/10.1016/j.tpb.2005.01.004>
- Pekin, B.K., Jung, J., Villanueva-Rivera, L.J., Pijanowski, B.C., Ahumada, J.A., 2012. Modeling acoustic diversity using soundscape recordings and LIDAR-derived metrics of vertical forest structure in a neotropical rainforest. *Landscape Ecology* 27, 1513–1522. <https://doi.org/10.1007/s10980-012-9806-4>
- Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential Biodiversity Variables. *Science* 339, 277–278. <https://doi.org/10.1126/science.1229931>
- Pieretti, N., Farina, A., Morri, D., 2011. A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). *Ecological Indicators* 11, 868–873. <https://doi.org/10.1016/j.ecolind.2010.11.005>
- Purvis, A., Hector, A., 2000. Getting the measure of biodiversity. *Nature* 405, 212–219. <https://doi.org/10.1038/35012221>
- Richards, D.G., Wiley, R.H., 1980. Reverberations and amplitude fluctuations in the propagation of sound in a forest—implications for animal communication. *The American Naturalist* 115, 381–399.
- Rodriguez, A., Gasc, A., Pavoine, S., Grandcolas, P., Gaucher, P., Sueur, J., 2014. Temporal and spatial variability of animal sound within a neotropical forest. *Ecological Informatics, Ecological Acoustics* 21, 133–143. <https://doi.org/10.1016/j.ecoinf.2013.12.006>

- Sala, O.E., 2000. Global Biodiversity Scenarios for the Year 2100. *Science* 287, 1770–1774. <https://doi.org/10.1126/science.287.5459.1770>
- Sala, O.E., Chapin, F.S., Iii, Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, M., Poff, N.L., Sykes, M.T., Walker, B.H., Walker, M., Wall, D.H., 2000. Global Biodiversity Scenarios for the Year 2100. *Science* 287, 1770–1774. <https://doi.org/10.1126/science.287.5459.1770>
- Schmidt, A.K.D., Römer, H., Riede, K., 2013. Spectral niche segregation and community organization in a tropical cricket assemblage. *Behavioral Ecology* 24, 470–480. <https://doi.org/10.1093/beheco/ars187>
- Servick, K., 2014. Eavesdropping on Ecosystems. *Science* 343, 834–837. <https://doi.org/10.1126/science.343.6173.834>
- Skowronski, M.D., Harris, J.G., 2006. Acoustic detection and classification of microchiroptera using machine learning: lessons learned from automatic speech recognition. *Journal of the Acoustical Society of America* 119, 1817–1833.
- Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J.T., Burton, C., Townsend, S.E., Carbone, C., Rowcliffe, J.M., Whittington, J., Brodie, J., Royle, J.A., Switalski, A., Clevenger, A.P., Heim, N., Rich, L.N., 2017. Scaling-up camera traps: monitoring the planet’s biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment* 15, 26–34. <https://doi.org/10.1002/fee.1448>
- Stowell, D., 2018. Computational Bioacoustic Scene Analysis, in: *Computational Analysis of Sound Scenes and Events*. Springer, Cham, pp. 303–333. https://doi.org/10.1007/978-3-319-63450-0_11
- Sueur, J., 2002. Cicada acoustic communication: potential sound partitioning in a multispecies community from Mexico (Hemiptera: Cicadomorpha: Cicadidae). *Biological Journal of the Linnean Society* 75, 379–394. <https://doi.org/10.1046/j.1095-8312.2002.00030.x>
- Sueur, J., Farina, A., 2015. Ecoacoustics: the Ecological Investigation and Interpretation of Environmental Sound. *Biosemiotics* 8, 493–502. <https://doi.org/10.1007/s12304-015-9248-x>
- Thompson, M.E., Schwager, S.J., Payne, K.B., 2010. Heard but not seen: an acoustic survey of the African forest elephant population at Kakum Conservation Area,

- Ghana. *African Journal of Ecology* 48, 224–231. <https://doi.org/10.1111/j.1365-2028.2009.01106.x>
- Tobias, J.A., Planqué, R., Cram, D.L., Seddon, N., 2014. Species interactions and the structure of complex communication networks. *PNAS* 111, 1020–1025. <https://doi.org/10.1073/pnas.1314337111>
- Toledo, L.F., de Araujo, C.B., 2017. *Zoophonie: les origines de la Bioacoustique*, in: *Hercules Florence: Le Nouveau Robinson*, Humboldt Books. pp. 264–285.
- Towsey, M., Parsons, S., Sueur, J., 2014. Ecology and acoustics at a large scale. *Ecological Informatics, Ecological Acoustics* 21, 1–3. <https://doi.org/10.1016/j.ecoinf.2014.02.002>
- Turner, W., 2014. Sensing biodiversity. *Science* 346, 301–302. <https://doi.org/10.1126/science.1256014>
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution* 18, 306–314. [https://doi.org/10.1016/S0169-5347\(03\)00070-3](https://doi.org/10.1016/S0169-5347(03)00070-3)
- Villanueva-Rivera, L.J., 2014. Eleutherodactylus frogs show frequency but no temporal partitioning: implications for the acoustic niche hypothesis. *PeerJ* 2, e496. <https://doi.org/10.7717/peerj.496>
- Waldbauer, G.P., 1988. Aposematism and Batesian Mimicry, in: *Evolutionary Biology, Evolutionary Biology*. Springer, Boston, MA, pp. 227–259. https://doi.org/10.1007/978-1-4613-0931-4_5
- Walters, C.L., Freeman, R., Collen, A., Dietz, C., Brock Fenton, M., Jones, G., Obrist, M.K., Puechmaille, S.J., Sattler, T., Siemers, B.M., Parsons, S., Jones, K.E., 2012. A continental-scale tool for acoustic identification of European bats. *Journal of Applied Ecology* 49, 1064–1074. <https://doi.org/10.1111/j.1365-2664.2012.02182.x>
- Way, M.J. (Ed.), 2012. *Advances in machine learning and data mining for astronomy*, Chapman & Hall/CRC data mining and knowledge discovery series. Chapman and Hall/CRC, Boca Raton, Fla.
- Weir, L.A., Mossman, M.J., 2005. North American Amphibian Monitoring Program (NAAMP) 307–313.

- Wernick, M.N., Yang, Y., Brankov, J.G., Yourganov, G., Strother, S.C., 2010. Machine Learning in Medical Imaging. *IEEE Signal Processing Magazine* 27, 25–38. <https://doi.org/10.1109/MSP.2010.936730>
- Wrege Peter H., Rowland Elizabeth D., Keen Sara, Shiu Yu, Matthiopoulos Jason, 2017. Acoustic monitoring for conservation in tropical forests: examples from forest elephants. *Methods in Ecology and Evolution* 8, 1292–1301. <https://doi.org/10.1111/2041-210X.12730>
- Wrege Peter H., Rowland Elizabeth D., Thompson Bruce G., Batruch Nikolas, 2010. Use of Acoustic Tools to Reveal Otherwise Cryptic Responses of Forest Elephants to Oil Exploration. *Conservation Biology* 24, 1578–1585. <https://doi.org/10.1111/j.1523-1739.2010.01559.x>
- Wulf, A., 2015. *The invention of nature: the adventures of Alexander von Humboldt, the lost hero of science*. John Murray, London.
- Yoccoz, N.G., Nichols, J.D., Boulinier, T., 2001. Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution* 16, 446–453. [https://doi.org/10.1016/S0169-5347\(01\)02205-4](https://doi.org/10.1016/S0169-5347(01)02205-4)

CHAPTER 1

Screening large audio datasets to determine the time and space distribution of Screaming Piha birds in a tropical forest

Juan Sebastian Ulloa

Muséum national d'Histoire naturelle, France

Université Paris-Sud, France

Amandine Gasc

Muséum national d'Histoire naturelle, France

Philippe Gaucher

CNRS USR 3456 Guyane, France

Thierry Aubin

Université Paris-Sud, France

Maxime Réjou-Méchain

Université Paul Sabatier France

French Institute of Pondicherry, India

Jérôme Sueur

Muséum national d'Histoire naturelle, France

OVERVIEW

Acoustic monitoring has proved to be an efficient approach to monitor wildlife, notably in environments with limited visibility, such as tropical rainforests. Today, recording equipment allows acoustic data to be gathered in remote areas at wide spatial and temporal scales. The resulting datasets are large and the use of automated processing systems to extract relevant information can greatly facilitate their analysis. Here, we have developed signal processing techniques to reveal the spatio-temporal dynamics of an emblematic bird voice of the neotropical forest: the song of the Screaming Piha (*Lipaugus vociferans*). Using recordings made in a French Guianan lowland forest, with an array of 24 microphones in a three dimensional space, we implemented a detection system based on spectrogram cross-correlation to trace the vocalisations of *L. vociferans*. We tuned the detection system based on the percentage area under the receiver operating characteristic curve, finding a maximum of 95.88%. To strictly minimise false positives, we set the operating point to have 34.9% true positives and 0% false positives. We detected a total of 12,735 songs attributed to the study bird during 25 study days. We found that spatial patterns of lower activity corresponded to a zone having smaller trees and more tree gaps – a known liana forest patch – suggesting that Screaming Piha birds tend to avoid non-mature primary forests. The sampling sites near the creeks had more detections than the sites further away, suggesting that the lek mating arenas might be distributed strategically to be near to a source of water. We also found a marked temporal pattern. The lek was active during the whole day, from sunrise to sunset, with two peaks of activity shifted by more than two hours from the dawn and dusk chorus. The approach described here can be tested using other conspicuous and stereotyped sounds that occur within a heterogeneous and noisy background. To decipher the complex interacting sounds of the tropical forest, these focal studies on specific acoustic elements should be complemented with community or soundscape analysis, to demonstrate the human impact on the ecosystem and to provide guidelines for natural resource management.

*"Descubrimos también el extraño silencio, lleno de todos los sonidos de la selva.
El rumor blando del agua, los gritos de los pájaros,
coda uno con su ritmo y su timbre,
nos parecieron música"*
– William Ospina (2012)

1. Introduction

Monitoring the dynamics of biodiversity is a key challenge for ecology and biological conservation (Magurran, 2004) or for the new ecoacoustics discipline (Sueur and Farina, 2015). In particular, there is an essential need to map the distribution of species in space and time over local or regional scales and during circadian or longer temporal cycles. To date, this challenge has been mainly tackled with the help of field-based data collected by human observers (Hill et al., 2005). The emergence of new sampling methods based on remote sensors, which automatically acquire environmental information at a regular rate, can supplement human observations, potentially increasing the accuracy of biodiversity monitoring data (Le Gaillard et al., 2012).

One of the most recent and original automatic approaches to track biodiversity changes is based on the detection of sounds produced by animals during communication (Towsey et al., 2014). New audio technology allows to deploy robust acoustic sensors, which collect data over long periods of time in remote areas. Acoustic monitoring based on these devices has already proved to be an efficient technique that could complement other data acquisition methods, notably in environments with limited visibility (Aide et al., 2013; Farnsworth and Russell, 2007; Frommolt and Tauchert, 2014; McDonald and Fox, 1999; Mellinger et al., 2007; Obrist et al., 2010; Yack et al., 2013). These

monitoring programmes can produce large volumes of data. The analysis of such unsupervised datasets by human observers is tedious and time-consuming, whereas the implementation of automated or semi-automated signal processing systems can greatly facilitate data management, data visualisation, and data analyses.

In recent years, numerous pattern recognition methods such as Gaussian mixture models (Cheng et al., 2010; Skowronski and Harris, 2006), hidden Markov models (Kogan and Margoliash, 1998), artificial neural networks (Chesmore, 2004) and random forest (Armitage and Ober, 2010; Potamitis, 2014; Briggs et al. 2012), have been applied to acoustic databases to detect and classify animal sounds. All these approaches are valuable advances to improve ecoacoustic monitoring. However, reliable recognition in complex audio scenes remains difficult (Bardeli, 2010; Potamitis, 2014; Towsey et al., 2012) because unsupervised recordings can contain significant background noise such as wind and rain, and species may vocalise simultaneously. To improve recognition rates and cover larger areas, Bardeli et al. (2010) proposed the development of a specific algorithm for each target species. The problem of detecting a known signal among noise has been previously studied for radar systems (Skolnik, 2001). One of the resulting methods is cross-correlation (Smith, 2003). This is a widespread approach, as it works as an optimal linear operation for detecting a deterministic signal corrupted by white Gaussian noise (Brunelli, 2009). Cross-correlation gives a measure of similarity between the template and the objective signal at shifted positions. Cross-correlation has been adapted to audio data screening by searching for areas of spectrograms that match with a template. Spectrogram cross-correlation has already been applied to identify the sound produced by a focused species in a multi-source audio recording as illustrated in several marine (Erbe et al., 1999; Mellinger and Clark, 2000) and terrestrial (Borker et al., 2014; Fitzsimmons et al., 2008; Frommolt and Tauchert, 2013; Llusia et al., 2013; Clark and Fristrup, 2009) studies. Moreover, a recent computational study included this technique in a multi-label classification system for the recognition of bird vocalisation (Potamitis, 2015).

Tropical forests pose a great challenge to global biodiversity conservation, as they suffer from destruction and degradation by human activity at exceptional rates (Morris, 2010). Surprisingly, large ecoacoustic monitoring programmes that focus on such tropical environments are rather rare. The coexistence of many tropical species generates a fascinating and complex acoustic environment, described as the soundscape

(Pijanowski et al., 2011), which remains poorly understood in its structure and dynamics. Recent studies have revealed rough temporal and spectral properties of the soundscape (Ellinger and Hödl, 2003; Hammer and Barret, 2001; Pekin et al., 2012; Riede, 1993; Rodriguez et al., 2014; Slabbekoorn, 2004), but the dynamics of the elementary items are still unknown. In order to understand better the emergent patterns of tropical soundscapes, it is necessary to decipher the dynamics of specific acoustic elements, in particular of soundmarks. A soundmark, as defined by Shafer (1977), refers to a unique sound that is specifically distinguished for its qualities by the people of a locality. Species associated with these iconic sounds might provide a focus for rising awareness and action – a flagship species in terms of conservation biology. In the South American tropical forest, the loud and characteristic song of the tropical bird *Lipaugus vociferans* is well known by locals and visitors and is undoubtedly a soundmark.

In a previous study, an array of automated recording systems was established in the primary lowland forest in French Guiana (Rodriguez et al, 2014). A large audio dataset was collected, replete with intricate animal sounds that compete and overlap, creating an overwhelming dataset. This audio data set was initially parameterised using a global approach, namely, by considering the files as unit samples without trying to identify the species they contained. This first approach provided important insights into the spatio-temporal organisation of the forest soundscape. However, it quickly appeared that analyses at a finer scale were required, to understand better the underlying factors that shape the dynamics of the tropical soundscape. To decrypt the complex acoustic environment of a tropical forest, we reconsidered our dataset by focusing on a salient acoustic element, the song of the *Lipaugus vociferans*.

The objectives of this study were (1) to adapt template matching techniques to find the song of the *L. vociferans* in the tropical acoustic environment and (2) to generate spatio-temporal maps of the distribution of this species. Here, we present a methodology to implement and tune template-matching techniques on ecoacoustic databases. The results of the acoustic monitoring are confronted with hydrology and vegetation maps through LiDAR data, providing novel ecological information on *L. vociferans*.

2. Materials

2.1. Study site

The study site was located in French Guiana at the CNRS Nouragues Research Station (4°05'N; 54°40'W), in a primary lowland rainforest (Figure 1). The Nouragues station is located in an inhabited region and is only accessible by river, 60 km from the nearest village, or by helicopter. Close to the equator line, the mean temperature oscillates between 26°C and 27°C and the humidity remains high during the whole year, between 80% and 90%. The rainfall is 2,861 mm year⁻¹ (average amount of rainfall from 1992 to 2012), with a 2-month dry season (<100 mm month⁻¹) occurring from September to October, and a shorter dry season in March. The study was conducted over 25 days, from 12 November to 6 December 2010, at the beginning of the wet season.

2.2. Acoustic sampling protocol

Sounds were recorded with Song Meter 2 (SM2) systems from Wildlife Acoustics Inc. The SM2 can be programmed to record automatically on a schedule. The device comprised two omnidirectional microphones, an analogue-to-digital converter and a controller inside a waterproof enclosure. Twelve recorders were deployed in a 4 by 3 grid 200 meters apart. The grid was formed by linear trails previously named with a letter and a Roman number. Each of our recording sites was named after its corresponding column (K, M, O) and row (XI, XIII, XV, XVII) coordinates (Figures 1b and 1c). Each recorder was set to sample the audio at 44.1 kHz at a 16-bit resolution. The left-channel microphone was placed at the canopy level (20 m above the ground) and the right-channel microphone was placed at the understory level 1.5 m above the ground. The devices were programmed to record for 1 minute every 15 minutes during the 24 h day/night cycle. The database used here accounts for 25 days of these recordings and represents a total of 960 hours of audio data.

1. SCREENING LARGE AUDIO DATASETS

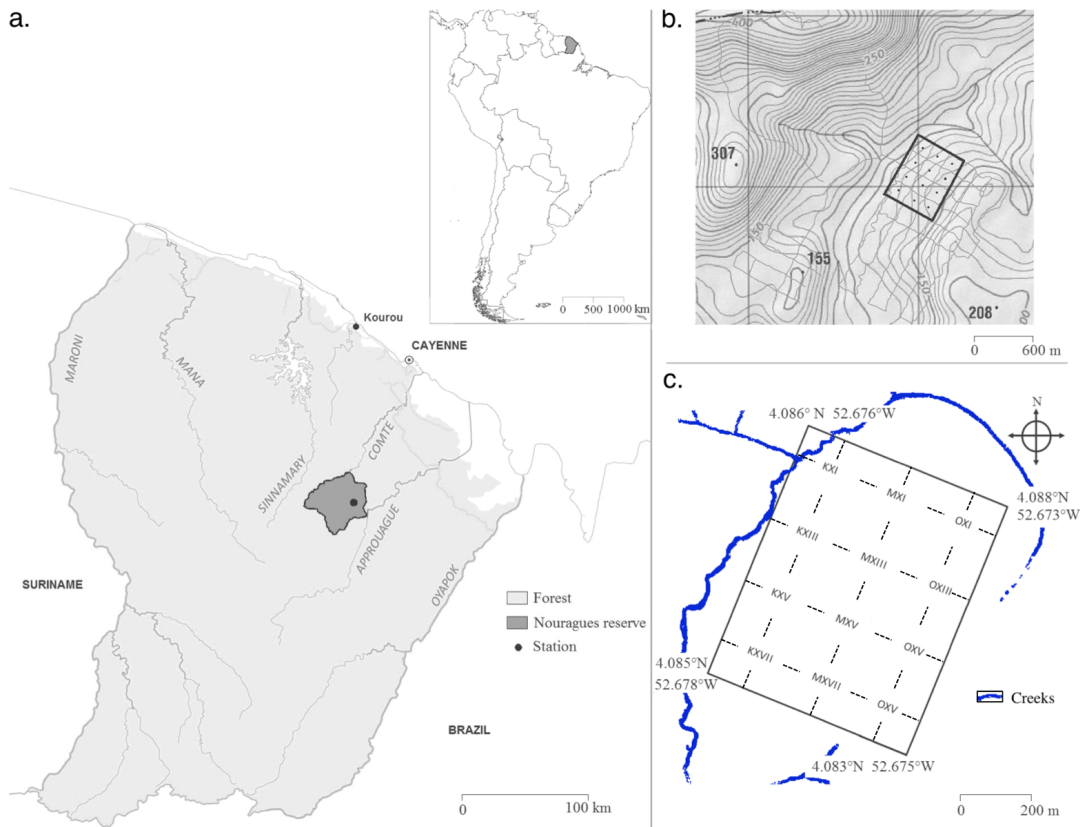


Figure 1. (a) Location of French Guiana and the Nouragues reserve. (b) Topographical map of the study area. (c) Schematic representation of the study area with the layout of the location of the 12 autonomous recording systems. The sensors were installed at the cross of perpendicular trails, in the understory (1.5 m) and at the canopy level (20 m). Letters and Roman numbers refer to trail names.

2.3. *Study species*

Despite its cryptic plumage, the Screaming Piha (*Lipaugus vociferans*; hereafter, “Piha”) is one of the most representative species of the neotropical forest. The commonly heard song of this bird is a remarkable soundmark known by anyone who has visited this region (electronic supplementary material, Audio S1). This species is regarded as one of the birds that possesses the loudest song in the world, with a mean sound-pressure level of 111.5 dB *re* $2 \cdot 10^{-5}$ Pa estimated at 1 m distance (Nemeth, 2004), and can spend 70% of the day time calling (Snow, 1961). Males gather in leks of around 25 individuals, where they highly compete vocally to mate with selecting females (Tostain et al., 1992). The Piha is common in the primary forest, but is highly sensitive to habitat degradation (Stotz et al., 1996).

The Piha is a suboscine passerine, therefore, it is thought that the song is inherited genetically and hence, stereotyped. Recent studies have shown that small variations in its song can encode individual distinctiveness and bear a lek signature (Fitzsimmons et al., 2008).

Perching on horizontal branches on his territory, the Piha calls with a typical song composed of two parts: (1) varied and quiet introductory notes, *groo groo*, and (2) a louder and highly modulated whistled *pee-haw*. For our study, we focused only on the louder part of the song. The first syllable, *pee*, is a rapidly ascending frequency modulation from 1.3 to 5.0 kHz. The second syllable, *haw*, can be considered as being composed of a whistle with three connected parts: an upswing, a downswing and a final constant frequency tune at 1.5 kHz (Figure 2 and electronic supplementary material, Audio S1). There is also another whistling call, *wee-oo*, made by the birds holding territories, but it is a very occasional vocalisation (Snow, 1961).

2.4. *Canopy height detection*

An acquisition of small-footprint discrete-return LiDAR (Light Detection and Ranging) was conducted in the Nouragues research area in March 2012 and covered an area of 2,400 hectares, including our study zone. The acquisition was based on a portable Riegl laser rangefinder (LMS-Q560) mounted on a Falcon aircraft at a speed of approximately 45 m s^{-1} about 400 m above the ground. The system had a multiple returns registering

capacity and the mean laser-point density was approximately 20 impulsions per m^2 . We constructed a canopy elevation model, i.e., the maximum local canopy height, using a procedure implemented in FUSION (McGaughey 2012). The canopy elevation model was constructed at 1-m resolution using a 1-m resolution digital elevation model (DEM) and the “CanopyModel” procedure implemented in FUSION. A 3×3 neighbour window median filter was finally used to smooth the surface and thus avoid local unrealistic maxima or minima. A full description of the LiDAR dataset and of the construction of the DEM and of the canopy model is given in Réjou-Méchain et al. (2015).

The median canopy height (in meters) of the surrounding 1 ha was calculated for each microphone to avoid overlap between neighbouring sample sites (Figure 3). The median of the canopy height has been shown to be a good proxy for the aboveground tree biomass, and hence of the whole forest structure (Réjou-Méchain et al., 2015).

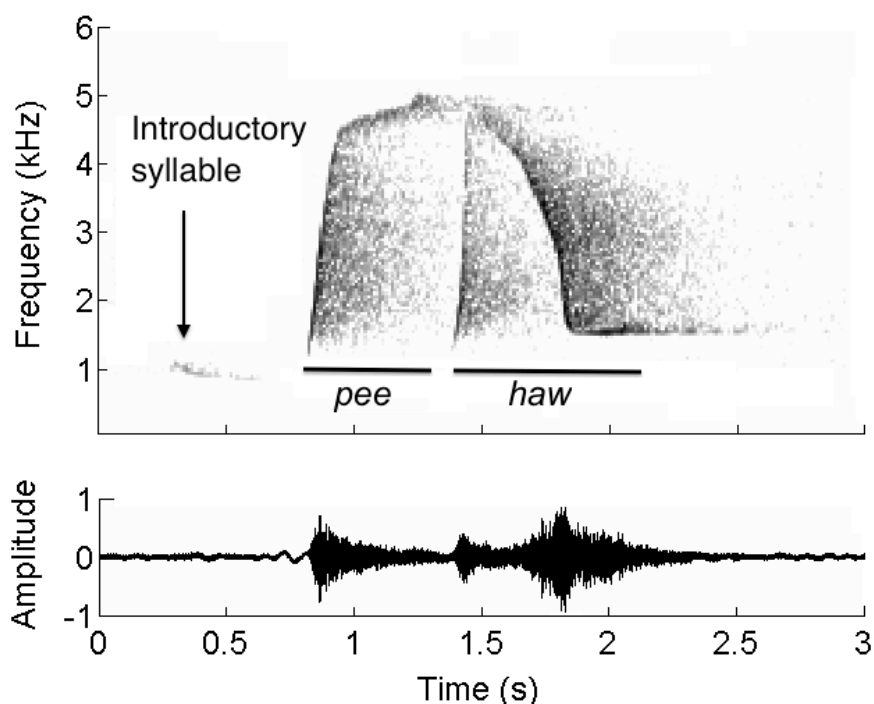


Figure 2. Spectrogram and waveform of a typical *Lipaugus vociferans* song composed of a two-syllable pee-haw. The high-energy regions are at the beginning and at the end of the song, at about 1.5 kHz. The short-time Fourier transform was calculated with a Hamming window of 1,024 samples, no zero padding and 50% overlap between windows. A dynamic range of 40 dB was set to display the spectrogram.

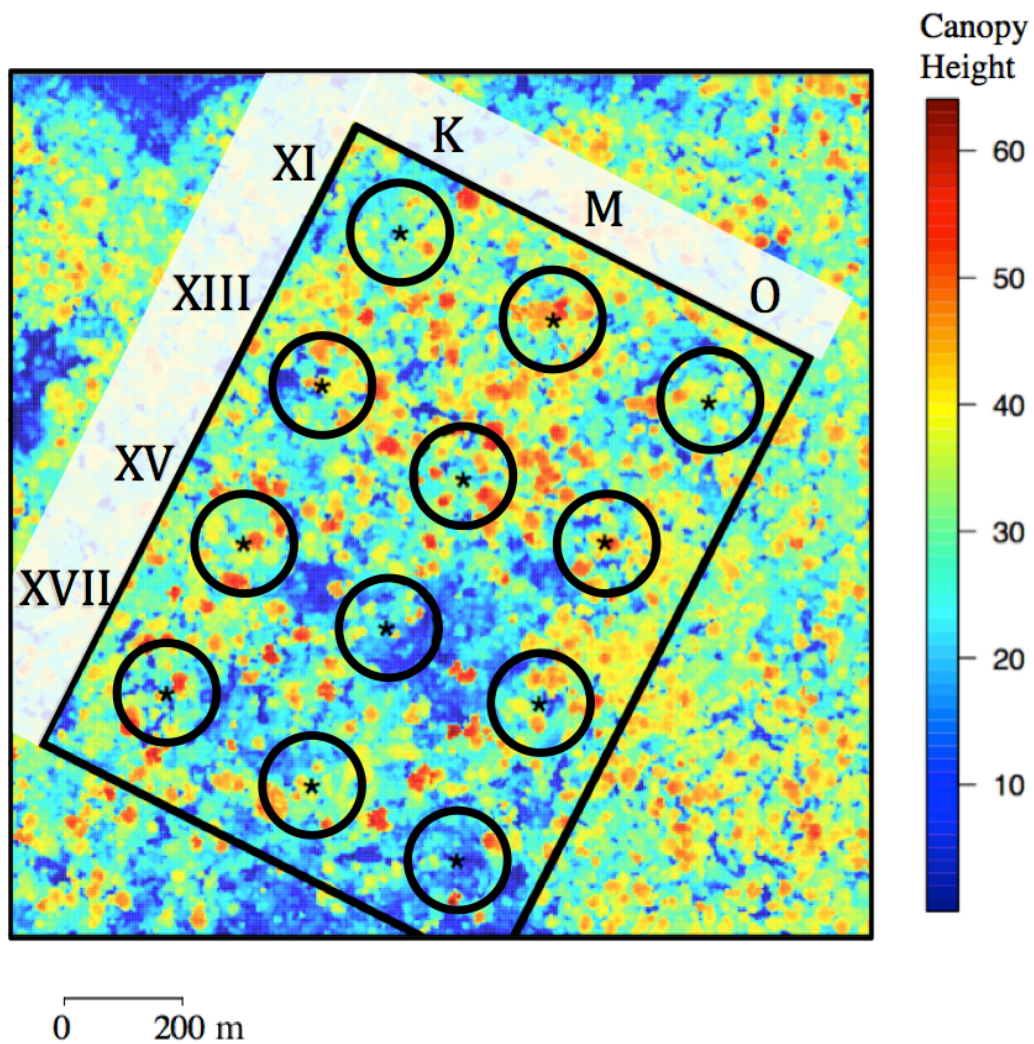


Figure 3. Height of the canopy of the study site using LiDAR technology. Stars indicate the position of the recorders and circles delimit an area of 1 ha surrounding the recorders where the median tree height was estimated. Letters and Roman numbers refer to trail names.

3. Methods

3.1. Test database

As a first step, we created a test database (training database) to develop a detection system. The acoustic environment in the tropical forest is heterogeneous in space and time. Previous statistical analyses on the recordings showed an acoustic pattern of four main sound environments during one day: (1) a morning period from 1:00 to 9:00, (2) a

day period from 9:00 to 18:00, (3) an evening period from 18:00 to 19:00, and (4) a night period from 19:00 to 1:00 (Rodriguez et al., 2014). To tune correctly the detection system, the test database should contain samples with soundscapes from the different time-periods. To account for the temporal variation, we selected one day every four days over the 25 days of study, and for each day, we chose one sample from each time period. To account for the spatial variation, we selected samples from all 12 recording sites. The resulting test database contained 336 files of one minute in length ($4 \text{ files/day/site} \times 7 \text{ days} \times 12 \text{ sites}$). Two of us (J.S and J.S.U.) listened with headphones and visually inspected the spectrogram of each file to annotate the occurrence of the Piha's songs in the test database. The open-source software Audacity (<http://audacity.sourceforge.net/>) was used for the manual analysis and annotations. Manually annotations contained Piha songs mixed with sounds produced by other species. The songs due to different Piha individuals vocalising at the same time could also overlap. Within the resulting test database, we found a total of 262 songs of the Piha. We used this material as the ground truth to tune and test the detection system (see Section 3.4.).

3.2. Signal Processing

Noise is unwanted sound that interferes with the desired signal and as stated in the introduction, it can be divided into background noise and interference. The recording equipment introduces unwanted energy and distortions; electromagnetic and thermal noise from the electronic circuit (-115 dBV equivalent input noise for the SM2 recorders), and quantisation distortion by the conversion of the signal from analogue to digital. Outdoor recordings always contain a significant amount of background noise that fluctuates in intensity. Geophonic sounds, such as wind and rain, add background noise to the signal as well as incidental anthroponic noise such as helicopter noise. Interferences are due to sounds produced by other vocalising species, but which are not of interest for the study. In the tropical forest, there is a large number of vocalising species that generate many sounds that can potentially interfere with the targeted signal. In signal detection, the background noise limits the detection range of the system, whereas the interference decreases the accuracy of the system by increasing the false-positive detection rate (Skolnik, 2001). In our study, the target signal is the song of the

Piha and all other sounds are regarded as noise. For example, we present a field recording that includes various sources of sounds, including the target signal (supplementary material, Audio S2).

Template matching is a set of techniques used to identify a pattern in a large database. A particular method of implementing template matching is to compute the cross-correlation function. Cross-correlation is performed by sliding a template signal over an objective signal, and calculating a correlation value at each time offset. The output signal is a measure of similarity between the template and the objective signal at a different time-lag. This method has been used in image processing to determine the position of a pattern in an image (Ding et al., 2001).

Audio data can be represented in a matrix by applying the short-time Fourier transform (STFT). The signal is divided into m overlapping windows of n samples and then the n -point discrete Fourier transform (DFT) is calculated for each windowed signal. The results are organised into a $m \times n$ matrix; the frequency frames m , the time frames n and the amplitude as cell values. The matrix is plotted as a colourmap, and the resulting image is called a spectrogram. In audio signal processing, the two-dimensional cross-correlation allows the search for patterns using features in the time and frequency domain.

Normalised cross-correlation (NCC) is an effective method for template matching. This approach is robust to differences in relative signal amplitude between the template and the objective signal, such as those caused by signal absorption (Brunelli, 2009). The two-dimensional normalised cross-correlation (NCC) is calculated by the following equation:

$$NCC(u, v) = \frac{\sum_{x,y} [f(x, y) - \bar{f}_{u,v}] [t(x - u, y - v) - \bar{t}]}{\sqrt{\sum_{x,y} [f(x, y) - \bar{f}_{u,v}]^2 \sum_{x,y} [t(x - u, y - v) - \bar{t}]^2}}$$

where f is the objective matrix and the sum is over the region x, y under the template t shifted at position u, v . $\bar{f}_{u,v}$ denotes the mean value of $f(x, y)$ within the region under the template and \bar{t} is the mean value of the template matrix.

3.3. Template selection

Firstly, we computed the signal-to-noise ratio (SNR) of the 262 Piha vocalisations that occurred in the test database. To calculate this ratio, we used the following equation:

$$SNR_{dB} = 20 \log_{10} \left(\frac{A_{signal}}{A_{noise}} \right)$$

where A is the root-mean-square of the sample. We selected for the background noise 0.7 s before and 0.7 s after the target song. The duration of the background noise (1.4 s) is equal to the mean duration of a Piha vocalisation. The SNR values range from -22 dB to +22 dB.

Then, to ensure the detection of nearby vocalisations, we selected the ten samples with the highest SNR (from +14 dB to +22 dB) and we standardised them. We aligned the samples in the time domain, starting at 0.2 s before the beginning of the vocalisation and with a total length of 1.6 s, and we normalised each sample between 0 and 1.

Finally, we calculated the mean over the ten standardised samples, to derive a single mean template. This template maintained the most salient parts of the signal and accounted for the small but inevitable variations in the song.

3.4. Tuning the spectrogram parameters

To tune the system, we compared the output of the NCC with manual annotations. The detection task treated showed an uneven distribution between classes, positive outcomes being rare compared to negative outcomes. As discussed in Provost & Fawcett (1997) and Daskalaki et al. (2006), the classification accuracy is an inappropriate performance measure in the particular case of unbalanced outcomes. We therefore used a methodology from signal detection theory, the Receiver Operating Characteristic (ROC) curve together with the Area Under the Curve (AUC) to summarise the rate of false and true detections. ROC curves describe the predictive behaviour of a class independent of class distribution, so they decouple classification performance from these factors (Provost et al., 1997). The AUC has two main advantages when evaluating the

performance of a detector: it is threshold-independent and it is invariant to *a priori* class probabilities (Bradey, 1997).

The main parameters of a spectrogram are the fast Fourier transform (FFT) window length, the overlap between successive FFT windows, the zero padding of the FFT and the shape of the window used. All these parameters change the time-frequency resolution of the spectrogram and therefore, change the result of the detection system. Preliminary observations have shown that the parameters that change the detection performance most drastically are the window length and the overlap between windows. We systematically measured the AUC metric associated with different combinations of these two parameters to find the optimal value. We changed the value of the window length from 128 to 8,192 samples by 2^k (where k is an integer from 7 to 13) steps, and the overlap from 0% to 75% by 25% steps. The performance of the system varied by 2.61% and we attained the optimal settings with a 256 window length and 0% overlap (Table 1).

The song of the Piha had a bandwidth of 4 kHz on average, ranging from 1.3 to 5 kHz. Thus, we narrowed our search around the frequencies of the signal of interest. We computed the correlation for different frequency bandwidths with bounds from 1 to 6 kHz by steps of 1 kHz. This ensured that the most suitable frequency band was found, for an optimised performance of the detection system. The performance of the system for the different bands varied by 9.72 % and the optimal setting was found for the frequency band from 1 to 6 kHz (Table 2).

Table 1. Variation in the system performance with the window FFT size and overlap of the short-time Fourier transform. The evaluation metric is the AUC (area under the curve) as a percentage.

| Overlap | FFT window size (number of samples) | | | | | | |
|---------|-------------------------------------|--------------|-------|-------|-------|-------|-------|
| | 128 | 256 | 512 | 1024 | 2048 | 4096 | 8192 |
| 0 % | 94.68 | 95.88 | 95.58 | 94.67 | 95.03 | 94.15 | 93.48 |
| 25 % | 94.27 | 95.17 | 95.11 | 94.83 | 94.43 | 94.21 | 93.27 |
| 50 % | 94.49 | 95.18 | 94.78 | 94.97 | 94.51 | 94.16 | 93.86 |
| 75 % | 94.34 | 94.99 | 94.65 | 94.92 | 94.64 | 94.01 | 93.84 |

Table 2. Variation in the system performance with the frequency bands of the search. The evaluation metric is the AUC (area under the curve) expressed as a percentage.

| Upper frequency bound (kHz) | Lower frequency bound (kHz) | | | | | |
|--------------------------------|-----------------------------|-------|-------|-------|-------|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 6 | 95.88 | 94.44 | 91.75 | 91.14 | 86.40 | |
| 5 | 92.74 | 91.23 | 89.40 | 89.84 | | |
| 4 | 90.22 | 89.06 | 86.16 | | | |
| 3 | 89.39 | 91.43 | | | | |
| 2 | 88.45 | | | | | |
| 1 | | | | | | |

3.5 Threshold setting

The corresponding ROC curve of the detection system selected is presented in Figure 4. The ROC curve displays the true positive rate (TP) of the detection system versus the false positive rate (FP) in relation with a variation of the discrimination threshold. A threshold must be defined to turn the results of cross-correlation into a series of discrete detection events. An event was registered when a peak of amplitude of the cross-correlation exceeded the threshold. The choice of such a threshold depends on the application. Increasing the threshold increases the selectivity of the detection system while lowering the threshold increases the sensitivity.

As the piha produces a loud song, the array of sensors could capture this sound in more than one recording site. In our particular case, we focused on the detection of the nearby vocalisations to avoid any overlap between neighbouring sites, which could overestimate the number of local vocalisations. We therefore chose a high and selective threshold value of 0.3. With this setting, the metrics of the detection system for the test database were: 0% FP and 34.9% TP. All analyses were programmed with Matlab (The MathWorks, Inc., Natick, MA).

4. Results

Prior to automated analysis, the manual detections were removed from the annotation table. Scanning the complete database took 17.4 h with a laptop running with a 2.8 GHz processor and a 8 GB memory. The detection system found a total of 12,735 Piha songs. From these, 62% of the vocalisations were detected in recordings made by the canopy sensors and the remaining 38% of the recordings were made by the understory sensors. Fewer detections were found in the understory in all recording sites (Figure 5). The site M-XI had by far the highest number of detections, followed by O-XI and K-XI. In contrast, almost no songs were detected in the recording sites M-XV, O-XV and O-XVII, where the forest had the lowest canopy (Figure 3) and which corresponds to a zone known to be dominated by lianas (Tymen et al., in press). All the other sites had a relatively low and similar number of detections.

In the time domain, there was a mean \pm SD of 509.4 ± 284.3 songs detected per day over the whole study area (Figure 6). The maximum activity was reached on the fourth day, with 1,405 songs, and the minimum activity on day 23, with 74 songs. Indeed, a decreasing trend was noted from the beginning to the end of the study. A simple linear regression model, with log-transformed response, showed that the number of vocalizations decreased approximately 5% by day (estimate \pm SE: -0.053 ± 0.013 , adjusted $R^2=0.4$, $F_{1,23}=17.09$, $P=0.0004$). Diagnostics of model validity and stability (Cook's distance, DFBetas, DFFits, and leverage; distribution of residuals plotted against fitted values) did not indicate obvious influential cases, nor deviations from the assumptions of normality and homogeneity of residuals. The understory and the canopy showed almost the same profile of temporal activity, and as observed before, the canopy always had a higher number of detections for each day.

A diurnal activity pattern appeared clearly (Figure 7): the calling activity began at sunrise (6:15) and increased slowly, reaching a first peak at 8:30. A reduction in the activity, with fluctuations, was observed between 10:00 and 14:45. The peak in activity was observed at 15:30 and then activity decreased progressively until sunset (18:10).

Finally, the complete set of results was summarised in a density plot that combined time (24-h cycle and days), space (horizontal and vertical) and Piha vocal activity (Figure 8). Zones of high (K-XVII, K-XV, K-XIII, K-XI, M-XI and O-XI) and low (M-XV, O-XV and OXVII) activity can be observed. In addition, the temporal

circadian activity is clearly emerging, showing the activity constrained by the sunrise and sunset dashed lines.

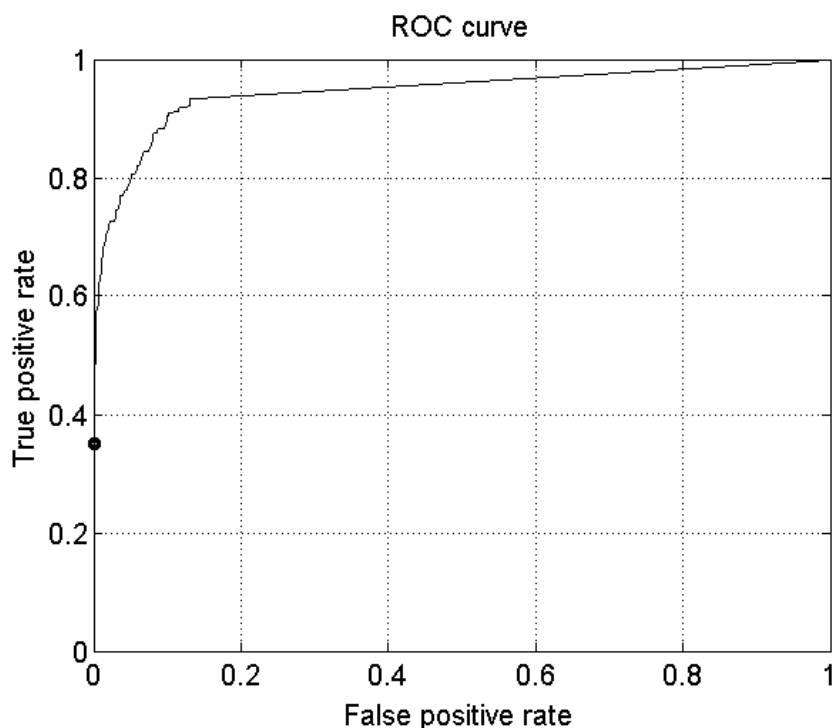


Figure 4. Receiver Operating Characteristic (ROC) curve of the detection system. The point on the curve indicates the operating point selected: false-positive rate = 0% and true-positive rate = 34.9%.

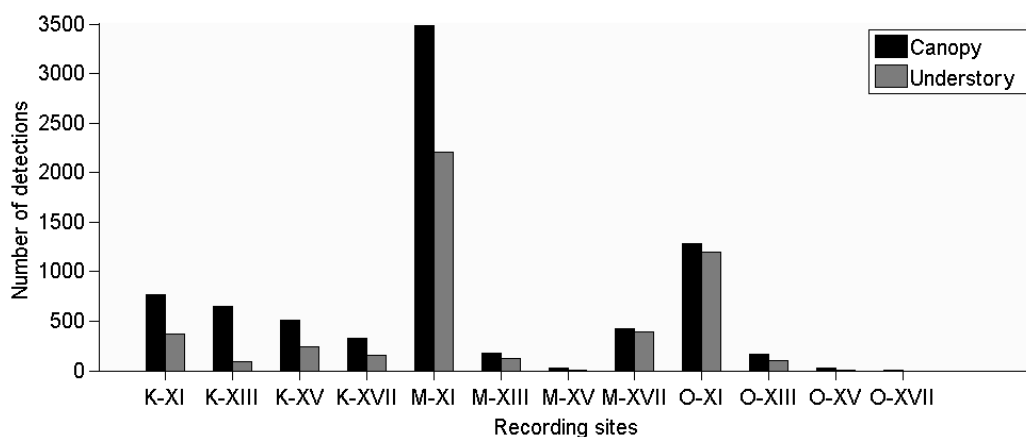


Figure 5. Barplot of the number of songs detected for each recording site at the understory and canopy levels.

1. SCREENING LARGE AUDIO DATASETS

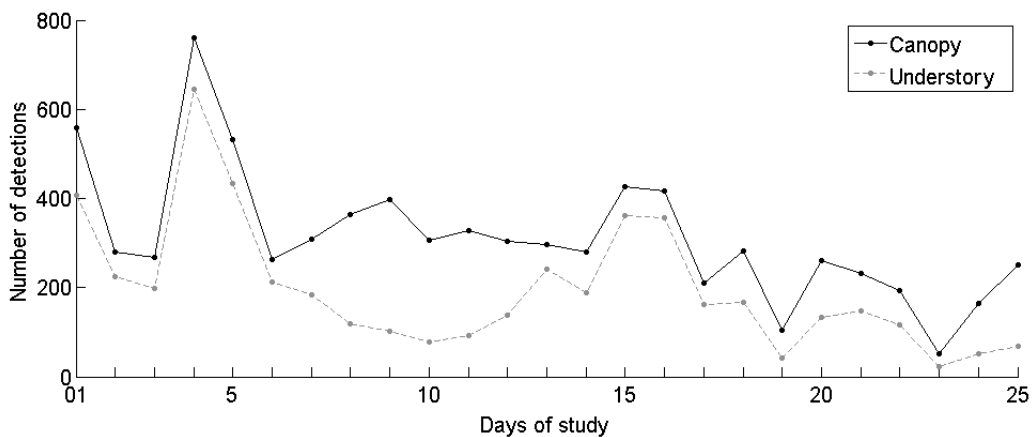


Figure 6. Sum of the songs detected for the understory and canopy levels for the 25 days of the study.

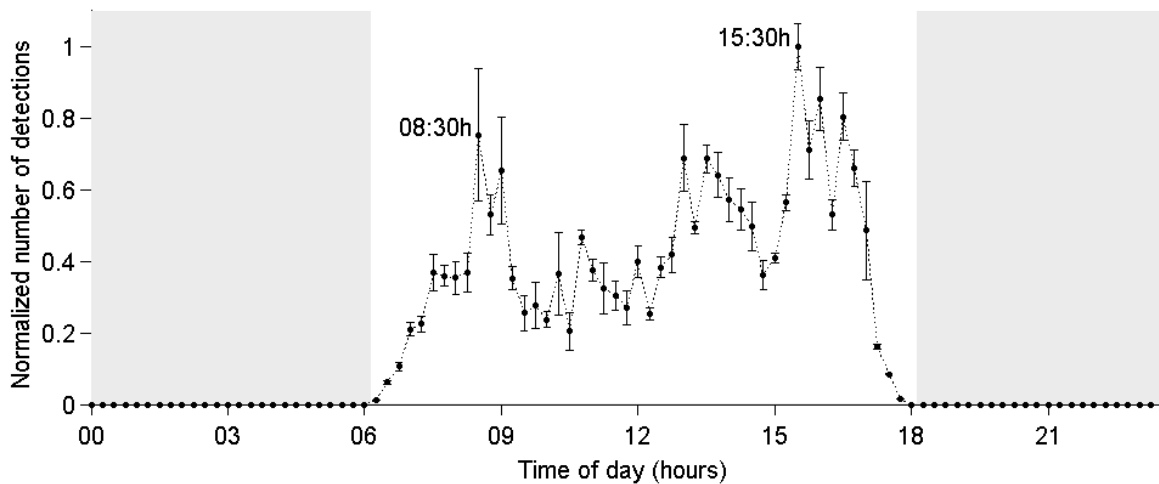


Figure 7. Circadian evolution of the Piha calling activity. The six recording sites with the highest activity were used to build the trend. The results of each recording site were normalised and then averaged. A final normalisation was performed to show the results from 0 to 1. Filled round marks represent the average and the bars represent the variance between the sites.

1. SCREENING LARGE AUDIO DATASETS

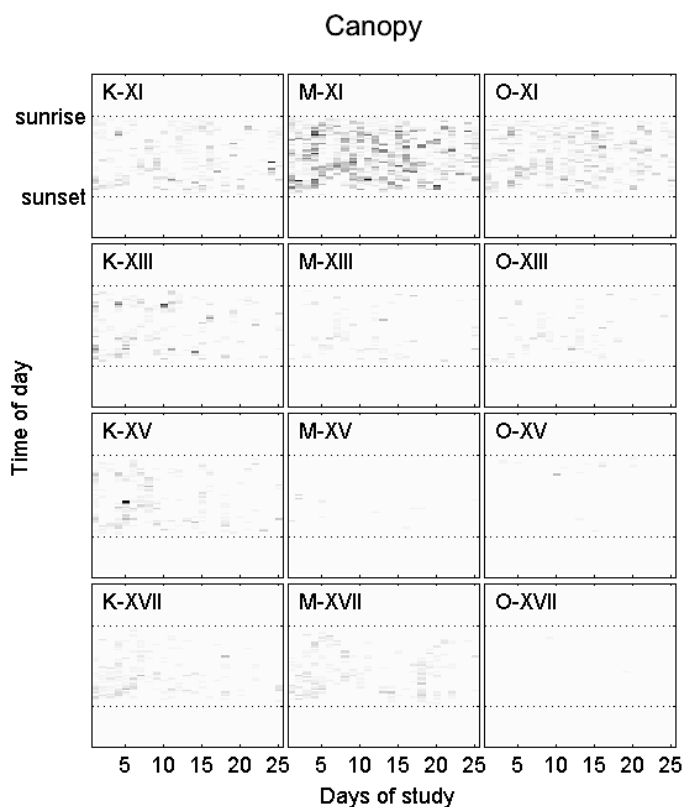
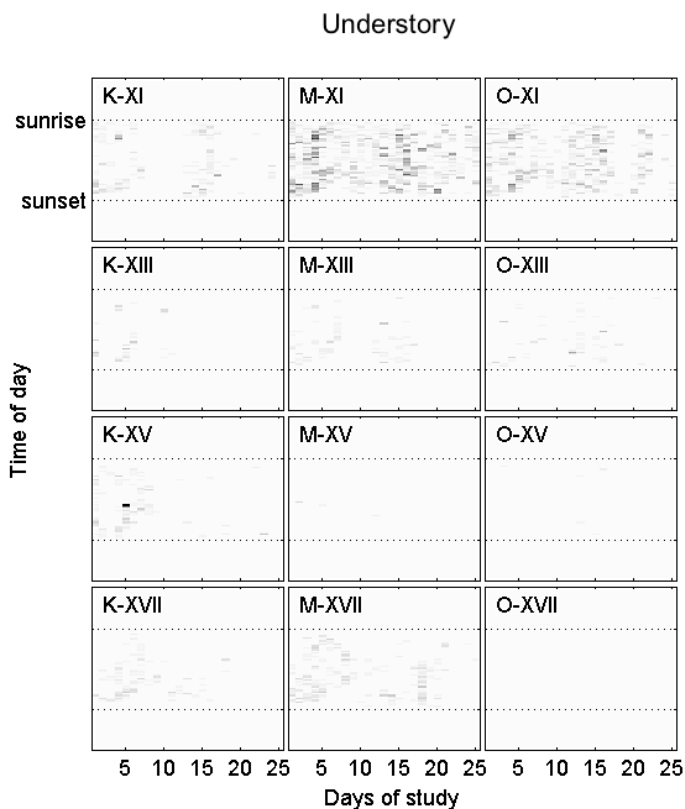


Figure 8. Representation of the spatio-temporal calling activity in an intensity image for the understory (a) and canopy (b) sensors. Each box represents a recording site. The y-axis represents the 24-h cycle from 0:00 am to 23:45, with sunrise (18:15) and sunset (18:10) indicated by a dashed line. Sunrise and sunset were considered constant during the sampling period (Earth equatorial line). The x-axis follows the activity during each of the 25 days of study. The grey scale of the pixels indicates the number of detections; the darker points represent more detections. The scale ranges from white (0 detections) to black (41 detections).



5. Discussion

Assessing and monitoring animal diversity in tropical forests is a challenging task, due to the immense number of co-existing species and the structural complexity of the environment. Acoustic monitoring can offer new tools to remotely detect and locate species of interest in space and time (Blumstein et al., 2011). In addition, acoustics might be involved in the global assessment of communities or ecosystems through the development of diversity or landscape indices (Gage & Axel, 2014; Pieretti et al., 2011; Sueur et al., 2014, 2008; Towsey et al., 2014). The sampling protocol used here was designed to define the acoustic dynamics of a patch of a neotropical forest (Rodriguez et al., 2014). This dataset embeds a unique amount of information concerning the behaviour, distribution and ecology of vocalising species. This information should be retrieved using methods adapted to large audio datasets. We therefore took the opportunity of this audio sampling to track one of the most conspicuous sounds of the forest in space and time – the song of the bird *Lipaugus vociferans* – the Piha.

5.1. Template matching in complex soundscapes

The Piha sings in an acoustic scene with several other species that can produce sound concomitantly. Detecting the song of the Piha in almost 1,000 hours of audio was therefore a challenge for signal analysis within large audio data. To achieve this task, we used a relatively standard template-matching approach through spectrogram cross-correlation. Using this technique, which has previously been used to identify vocal individuality in the Piha (Fitzsimmons et al., 2008), we could build an objective classification system with a relatively small test database. In addition, we could screen the complete database within an acceptable processing time (<18h) with a standard laptop. Therefore, we showed that the cross-correlation of spectrogram is a robust and rapid technique to detect soundmarks such as the Piha song. This success was primarily due to the loudness and stereotypy of the Piha song, therefore, we expect that our approach might also perform well with other conspicuous and stereotyped sounds such as some amphibian vocalisations and insect stridulations, thereby opening perspectives in multi-species monitoring approaches.

The performance of spectrogram cross-correlation varies, depending on the focused signal, the background noise, the interferences and the input parameters (spectrogram specifications and the template(s) used). The parameters of spectrogram cross-correlation must therefore be tuned to consider all the sources of variation as we did, by testing the features of the Fourier window, the frequency bandwidth of interest and the detection threshold value. The spectrogram cross-correlation technique is now incorporated into different software, facilitating its handling by non-experts in signal analysis. However, the features that tune the detection system are rarely found in software documentation, which reduces the possibility of reproducibility.

The spectrogram cross-correlation is definitely not universal and other methods should be invoked for less stereotyped songs. As suggested by Towsey et al. (2012), each species might require a specific tool to be identified automatically, making the unsupervised monitoring of several species extremely difficult. The use of more generic features describing intrinsic call structure, such as ridge features (Dong et al. 2015), could provide a better performance and the flexibility to detect species with more variable songs.

5.2. Remote-sensing, behaviour and ecology of the Piha

We initially found that the number of Piha vocalisations detected in the canopy was higher than in the understory. The canopy microphone was placed at a height of 20 m and the understory microphone at 1.5 m. Two facts might explain this vertical difference: (1) Piha singing perches are closer to the canopy than to the ground; (2) the acoustic properties of the habitat favour signal transmission towards the canopy. Regarding the first alternative, the literature is consistent in defining the perch height of the Piha in about the middle understory and the lower canopy (Stotz et al., 1996; Thiollay, 1994; Thiollay et al., 2001) with estimates between 6 and 16 m or higher (Snow, 1961) or 7 and 25 m (Pearson, 1971). Lower-boundary estimates are consistent, but upper-boundary ones vary significantly between observations. Personal observations (J.S.U) indicated that the perch height of the Piha was closer to the canopy microphone (i.e., 20 m) in our study site. Regarding the second explanation, the few studies on signal transmission in tropical environments have evaluated signal degradation based on source height (Marten et al., 1977; Morton, 1975) or in specific layers (Ellinger and

Hödl, 2003; Jain and Balakrishnan, 2012, Nemeth et al., 2001) but not on receiver height. Therefore, it would be too speculative to draw conclusion at this point. In summary, given our field observations (see above), it is more likely that the Piha sings closer to the canopy microphone at this specific study site. Transmission experiments and calibration methodologies at ecoacoustic monitoring sites would be needed, however, to confirm this supposition. Some methodological techniques that might be applicable for future studies are presented in the work of Llusia et al. (2011).

Considering a wider spatial scale, the results showed that the recording site M-XI produced by far the highest number of vocalisations (5,696), followed by the site O-XI (2,476) on all days. The high activity at this site is probably due to the proximity of a lek, where a higher number of individuals were concentrated.

The horizontal spatial activity pattern corresponded to the structure of the canopy estimated by LiDAR techniques. In particular, the sites with the lowest activity had the lowest canopy height. These sites are located in a forest dominated by lianas where canopy openness and dynamics is much greater than in the surrounding tall forest (Tymen et al. in press). This reflects that the habitat of the Piha is the high mature primary forest and confirms observations made by Stotz et al. (1996). In addition, Piha songs were more frequently detected in the recording spots K-XVII, K-XV, K-XIII, K-XI, M-XI and O-XI, which border the main creek. This suggests that the lek mating arenas are distributed strategically near to a source of water for the singing males. However, further analysis and data would be needed to confirm this habitat preference.

Regarding the temporal organisation of the acoustic activity of the Piha, the screening of the database showed a well-defined circadian rhythm over a 12-h period, starting at 6:00 and ending at 18:00. It appears that the Piha avoids calling at dawn and at dusk, at the transition between day and night, when soundscapes with more acoustic activity could be observed (Rodriguez et al., 2014). The morning peak of Piha singing activity appeared more than two hours after dawn and the afternoon peak of activity was reached two and a half hours before dusk. During the rest of the day, the vocal activity was less intense, but still present, making the Piha a relatively active bird that dominates the day soundscape and confirms its status as a soundmark of the forest.

On a larger temporal scale, the number of detections decreased from the beginning to the end of the study. However, the sampling period of only 25 days, even though important, prevents a seasonal trend from being interpreted. Only longer

sampling protocols, in terms of development, over months and possibly years, could provide the required data to establish a seasonal trend for this bird and for the other sounds of the forest.

6. Conclusion

This species-centred study represents a contribution towards biodiversity assessment at a large scale of a taxon-rich ecosystem – the lowland tropical forest. Using data collected in a neotropical forest with a tree-dimension array of 24 microphones and implementing signal processing techniques, we could describe the spatial and temporal distribution of the tropical bird, *L. vociferans*. We provide detailed, objective and quantitative results, which are three crucial qualities for the advancement of biodiversity-monitoring strategies.

This study reused a protocol that aimed to sense the global acoustic features of the forest. It is therefore possible to use the sample dataset to obtain information on tropical forest acoustics at different scales, from populations to communities or landscapes. We advocate the development and use of both bottom-up and top-down approaches for research that deals with individual species of interest through automatic identification, and for research that zooms out on a group of species through the use of diversity indices (Sueur et al., 2014).

The tropical acoustic environment is composed of numerous sounds that interact and remain to be analysed at narrow and wide ranges. Even when the focal point of a study is a particular sound, it is important to remember that this sound occurs in a specific acoustic context that might have shaped the focused sound. Understanding the diversity of sounds in the tropical forest is a more difficult task than to excise individual elements and examine them independently; the scene and its characters should be examined together.

Acknowledgments

This research was supported by Fondation pour la Recherche sur la Biodiversité grant (Biosound), CNRS grant (Amazonie) and SABIOD MASTODONS Big Data CNRS MI project 2012-2017. We would like to warmly thank Olivier Bohner and Patrick Châtelet for their help during our stay at the Nouragues research station, Olivier Adam and Alexandra Rodriguez for advice on the dataset. J.S.U. was supported by COLFUTURO (Posgrados en el exterior, 2012) and COLCIENCIAS (Doctorados en el exterior, convocatoria 646 de 2014) . M.R-M was supported by a post-doctoral grants from CNES (0101544) and by two 'Investissement d'Avenir' grants managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-2501; TULIP, ref. ANR-10-LABX-0041). The authors would like to greatly thank two anonymous referees for their helpful remarks and comments on the manuscript.

Supplementary Information

Supplementary data to this chapter can be found online at:

<https://doi.org/10.1016/j.ecoinf.2015.11.012>

Audio S1. Focused recording of the typical vocal display of a *Lipaugus vociferans*. Recorded on November 16, 2010, at 1:30 pm, in site M-XI with the canopy microphone.

Audio S2. Soundscape of the tropical forest. Among the numerous sounds, a typical song of the *Lipaugus vociferans* can be heard between second 9 and 11. Despite the interferences, this song was detected by our detection system without false positives. Recorded on November 20, at 6:00 am in site M-XI with the ground microphone.

References

- Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1, e103. <https://doi.org/10.7717/peerj.103>
- Armitage, D.W., Ober, H.K., 2010. A comparison of supervised learning techniques in the classification of bat echolocation calls. *Ecological Informatics* 5, 465–473. <http://dx.doi.org/10.1016/j.ecoinf.2010.08.001>
- Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K.-H., Frommolt, K.-H., 2010. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters, Pattern Recognition of Non-Speech Audio* 31, 1524–1534. <https://doi.org/10.1016/j.patrec.2009.09.014>
- Blumstein, D.T., Mennill, D.J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J.L., Krakauer, A.H., Clark, C., Cortopassi, K.A., Hanser, S.F., McCowan, B., Ali, A.M., Kirschel, A.N.G., 2011. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology* 48, 758–767. <https://doi.org/10.1111/j.1365-2664.2011.01993.x>
- Borker, A.L., Mckown, M.W., Ackerman, J.T., Eagles-Smith, C.A., Tershy, B.R., Croll, D.A., 2014. Vocal Activity as a Low Cost and Scalable Index of Seabird Colony Size: Automated Acoustic Seabird Monitoring. *Conservation Biology* 28, 1100–1108. <https://doi.org/10.1111/cobi.12264>
- Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30, 1145–1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
- Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X.Z., Raich, R., Hadley, S.J., Hadley, A.S., Betts, M.G., 2012. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *J. Acoust. Soc. Am.* 131, 4640–4650.
- Brunelli, R., 2009. *Template matching techniques in computer vision: theory and practice*. John Wiley and Sons, Ltd.

- Cheng, J., Sun, Y., Ji, L., 2010. A call-independent and automatic acoustic system for the individual recognition of animals: a novel model using four passerines. *Pattern Recognition* 43, 3846–3852.
- Chesmore, D., 2004. Automated bioacoustic identification of species. *Anais da Academia Brasileira de Ciências* 76, 435–440.
- Clark, C.W., Fristrup, K.M., 2009. Advanced technologies for acoustic monitoring of bird populations (SERDP Project RC-1461). Cornell University.
- Ding, L., Goshtasby, A., Satter, M., 2001. Volume image registration by template matching. *Image and Vision Computing* 19, 821–832.
- Dong, X., Towsey, M., Truskinger, A., Cottman-Fields, M., Zhang, J., Roe, P., 2015. Similarity-based birdcall retrieval from environmental audio. *Ecological Informatics* 29, Part 1, 66–76. <https://doi.org/10.1016/j.ecoinf.2015.07.007>
- Ellinger, N., Hödl, W., 2003. Habitat acoustics of a neotropical lowland rainforest. *Bioacoustics* 13, 297–321.
- Erbe, C., King, A.R., Yedlin, M., Farmer, D.M., 1999. Computer models for masked hearing experiments with beluga whales (*Delphinapterus leucas*). *Journal of the Acoustical Society of America* 105, 2967–2978. <http://dx.doi.org/10.1121/1.426945>
- Farnsworth, A., Russell, R.W., 2007. Monitoring flight calls of migrating birds from an oil platform in the northern Gulf of Mexico. *Journal of Field Ornithology* 78, 279–289. <https://doi.org/10.1111/j.1557-9263.2007.00115.x>
- Fitzsimmons, L.P., Barker, N.K., Mennill, D.J., 2008. Individual variation and lek-based vocal distinctiveness in songs of the Screaming Piha (*Lipaugus vociferans*), a suboscine songbird. *Auk* 125, 908–914.
- Frommolt, K.-H., Tauchert, K.-H., 2014. Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird. *Ecological Informatics* 21, 4–12. <http://dx.doi.org/10.1016/j.ecoinf.2013.12.009>
- Gage, S.H., Axel, A.C., 2014. Visualization of temporal change in soundscape power of a Michigan lake habitat over a 4-year period. *Ecological Informatics, Ecological Acoustics* 21, 100–109. <https://doi.org/10.1016/j.ecoinf.2013.11.004>
- Hammer, Ø., Barrett, N., 2001. Techniques for studying the spatio-temporal distribution of animal vocalizations in tropical wet forest. *Bioacoustics* 12, 21–35.

- Hill, D.A., Fasham, M., Tucker, G., Shewry, M., Shaw, P. (Eds.), 2005. Handbook of biodiversity methods: survey, evaluation and monitoring. Cambridge University Press, Cambridge ; New York.
- Jain, M., Balakrishnan, R., 2012. Does acoustic adaptation drive vertical stratification? A test in a tropical cricket assemblage. *Behavioral Ecology* 23, 343–354. <https://doi.org/10.1093/beheco/arr191>
- Kogan, J.A., Margoliash, D., 1998. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. *Journal of the Acoustical Society of America* 103, 2185–2196.
- Le Galliard, J.-F., Guarini, J.-M., Gaill, F., Institut Écologie et environnement (France), 2012. Sensors for ecology: towards integrated knowledge of ecosystems. CNRS, Paris.
- Llusia, D., Márquez, R., Beltrán, J.F., Benítez, M., do Amaral, J.P., 2013. Calling behaviour under climate change: geographical and seasonal variation of calling temperatures in ectotherms. *Glob Chang Biol* 19, 2655–2674. <https://doi.org/10.1111/gcb.12267>
- Llusia, D., Márquez, R., Bowker, R., 2011. Terrestrial sound monitoring systems, a methodology for quantitative calibration. *Bioacoustics* 20, 277–286. <https://doi.org/10.1080/09524622.2011.9753651>
- Magurran, A.E., 2004. Measuring biological diversity. Blackwell Pub, Malden, Ma.
- Marten, K., Quine, D., Marler, P., 1977. Sound transmission and its significance for animal vocalization. II. Tropical forest habitats. *Behavioral Ecology and Sociobiology* 2, 291–302.
- McDonald, M.A., Fox, C.G., 1999. Passive acoustic methods applied to fin whale population density estimation. *Journal of the Acoustical Society of America* 105, 2643–2651. <https://doi.org/10.1121/1.426880>
- Mellinger, D.K., Clark, C.W., 2000. Recognizing transient low-frequency whale sounds by spectrogram correlation. *Journal of the Acoustical Society of America* 107, 3518–3528.
- Mellinger, D.K., Stafford, K.M., Moore, S.E., Dziak, R.P., Matsumoto, H., 2007. An overview of fixed passive acoustic observation methods for cetaceans. *Oceanography* 20, 36–45. <http://dx.doi.org/10.5670/oceanog.2007.03>

- Morris, R.J., 2010. Anthropogenic impacts on tropical forest biodiversity: a network structure and ecosystem functioning perspective. *Phil. Trans. R. Soc. B.* 365, 3709–3718. <https://doi.org/doi:10.1098/rstb.2010.0273>
- Morton, E.S., 1975. Ecological sources of selection on avian sounds. *The American Naturalist* 109, 17–34.
- Nemeth, E., 2004. Measuring the sound pressure level of the screaming piha *Lipaugus vociferans*: one of the loudest birds in the world? *Bioacoustics: The International Journal of Animal Sound and its Recording* 14, 225–228.
- Obrist, M.K., Pavan, G., Sueur, J., Riede, K., Llusia, D., Márquez, R., 2010. Bioacoustics approaches in biodiversity inventories, in: Eymann, J., Degreef, J., Häuser, C., Monje, J.C., Samyn, Y., Vandenspiegel, Y.D. (Eds.), *Manual on Field Recording Techniques and Protocols for All Taxa Biodiversity Inventories and Monitoring*. ABC Taxa, pp. 18–48.
- Ospina, W., 2012. *El país de la canela*. Penguin Random House Grupo Editorial Colombia.
- Pearson, D.L., 1971. Vertical stratification of birds in a tropical dry forest. *The Condor* 73, 46–55.
- Pekin, B.K., Jung, J., Villanueva-Rivera, L.J., Pijanowski, B.C., Ahumada, J.A., 2012. Modeling acoustic diversity using soundscape recordings and LIDAR-derived metrics of vertical forest structure in a neotropical rainforest. *Landscape Ecology* 27, 1513–1522. <https://doi.org/10.1007/s10980-012-9806-4>
- Pijanowski, B.C., Farina, A., Gage, S.H., Dumyahn, S.L., Krause, B.L., 2011. What is soundscape ecology? An introduction and overview of an emerging new science. *Landscape Ecol* 26, 1213–1232. <https://doi.org/10.1007/s10980-011-9600-8>
- Potamitis, I., 2015. Unsupervised dictionary extraction of bird vocalisations and new tools on assessing and visualising bird activity. *Ecological Informatics* 26, Part 3, 6–17. <http://dx.doi.org/10.1016/j.ecoinf.2015.01.002>
- Potamitis, I., 2014. Automatic classification of a taxon-rich community recorded in the wild. *PLoS ONE* 9, e96936. <https://doi.org/10.1371/journal.pone.0096936>
- Provost, F., Fawcett, T., 1997. Analysis and Visualization of Classifier Performance: Comparison under Imprecise Class and Cost Distributions, in: *In Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*. AAAI Press, pp. 43–48.

- Provost, F., Fawcett, T., Kohavi, R., 1997. The Case Against Accuracy Estimation for Comparing Induction Algorithms, in: In Proceedings of the Fifteenth International Conference on Machine Learning. Morgan Kaufmann, pp. 445–453.
- Réjou-Méchain, M., Tymen, B., Blanc, L., Fauset, S., Feldpausch, T.R., Monteagudo, A., Phillips, O.L., Richard, H., Chave, J., 2015. Using repeated small-footprint LiDAR acquisitions to infer spatial and temporal variations of a high-biomass Neotropical forest. *Remote Sensing of Environment* 169, 93–101. <https://doi.org/10.1016/j.rse.2015.08.001>
- Riede, K., 1993. Monitoring Biodiversity: Analysis of Amazonian Rainforest Sounds. *Ambio* 22, 546–548.
- Rodriguez, A., Gasc, A., Pavoine, S., Grandcolas, P., Gaucher, P., Sueur, J., 2014. Temporal and spatial variability of animal sound within a neotropical forest. *Ecological Informatics, Ecological Acoustics* 21, 133–143. <https://doi.org/10.1016/j.ecoinf.2013.12.006>
- Schafer, R., 1993. *The soundscape : our sonic environment and the tuning of the world*. Destiny Books Distributed to the book trade in the United States by American International Distribution Corp, Rochester, Vt. United States.
- Skolnik, M.I., 2001. *Introduction to radar systems*. McGraw Hill, Boston.
- Skowronski, M.D., Harris, J.G., 2006. Acoustic detection and classification of microchiroptera using machine learning: lessons learned from automatic speech recognition. *Journal of the Acoustical Society of America* 119, 1817–1833.
- Slabbekoorn, H., 2004. Habitat-dependent ambient noise: consistent spectral profiles in two African forest types. *Journal of the Acoustical Society of America* 116, 3727. <https://doi.org/10.1121/1.1811121>
- Smith, S.W., 2003. *Digital Signal Processing, a practical guide for engineers and scientists*. Newnes.
- Snow, B., 1961. Notes on the behavior of three Cotingidae. *Auk* 78, 153–157. <https://doi.org/10.2307/4082129>
- Stotz, D., Fitzpatrick, J.W., Parker, T.A., Moskovits, D.K., 1996. *Neotropical birds : ecology and conservation*. University of Chicago Press, Chicago.
- Sueur, J., Farina, A., 2015. Ecoacoustics: the Ecological Investigation and Interpretation of Environmental Sound. *Biosemiotics* 8, 493–502. <https://doi.org/10.1007/s12304-015-9248-x>

- Sueur, J., Farina, A., Gasc, A., Pieretti, N., Pavoine, S., 2014. Acoustic Indices for Biodiversity Assessment and Landscape Investigation. *Acta Acustica united with Acustica* 100, 772–781. <https://doi.org/10.3813/AAA.918757>
- Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S., 2008. Rapid Acoustic Survey for Biodiversity Appraisal. *PLoS ONE* 3. <https://doi.org/10.1371/journal.pone.0004065>
- Thiollay, J.-M., 1994. Structure, density and rarity in an Amazonian rainforest bird community. *Journal of Tropical Ecology* 10, 449–481. <https://doi.org/10.449-481>
- Thiollay, J.M., Jullien, M., Théry, M., Erard, C., 2001. Bird species (from Nouragues inselberg to arataye river), in: *Nouragues: Dynamics and Plant-Animal Interactions in a Neotropical Rainforest*. Kluwer Academic Publishers, Dordrecht Boston.
- Tostain, O., Dujardin, J.L., Érard, C., Thiollay, J.M., 1992. *Oiseaux de Guyane*. Société d'études ornithologiques.
- Towsey, M., Parsons, S., Sueur, J., 2014. Ecology and acoustics at a large scale. *Ecological Informatics, Ecological Acoustics* 21, 1–3. <https://doi.org/10.1016/j.ecoinf.2014.02.002>
- Towsey, M., Planitz, B., Nantes, A., Wimmer, J., Roe, 2012. A Toolbox for Animal Call Recognition. *Bioacoustics* 21, 107–125.
- Tymen, B., Réjou-Méchain, M., Dalling, J.W., Fauset, S., Feldpausch, T.R., Norden, N., Phillips, O.L., Turner, B.L., Viers, J., Chave, J., 2015. Evidence for arrested succession in a liana-infested Amazonian forest. *J Ecol* n/a-n/a. <https://doi.org/10.1111/1365-2745.12504>
- Yack, T.M., Barlow, J., Calambokidis, J., Southall, B., Coates, S., 2013. Passive acoustic monitoring using a towed hydrophone array results in identification of a previously unknown beaked whale habitat. *Journal of the Acoustical Society of America* 134, 2589–2595. <https://doi.org/10.1121/1.4816585>

CHAPTER 2

Explosive breeding in tropical anurans: environmental triggers, community composition and acoustic structure

Juan Sebastian Ulloa

Muséum national d'Histoire naturelle, France

Université Paris-Sud, France

Thierry Aubin

Université Paris-Sud, France

Diego Llusia

Universidad Autónoma de Madrid, Spain

Élodie A. Courtois

Université de Guyane, France

Antoine Fouquet

Université Paul Sabatier, France

Université de Guyane, France

Philippe Gaucher

Université de Guyane, France

Sandrine Pavoine

Muséum national d'Histoire naturelle, France

Jérôme Sueur

Muséum national d'Histoire naturelle, France

OVERVIEW

Amphibians are endangered worldwide by a diversity of threats, including habitat loss, diseases and climate change. These threats could be pronounced for explosive breeding communities, which breed massively in limited-resource areas during brief periods. Tropical explosive breeding events generally involve highly-diverse anuran communities. Many aspects of these events are still largely unknown, probably because their study remains challenging due to their ephemeral nature. Using automated sensors, we collected acoustic and environmental data to monitor five ponds in French Guiana during a four-month period. We assessed acoustic dynamics in the anuran communities before and during explosive breeding events and confronted these changes with environmental variables. We detected in each pond two explosive breeding events, lasting between 24 and 70 hours. The rainfall during the previous 48 hours was the most important factor predicting the emergence of these events. During explosive breeding events, we identified a temporal factor that clearly distinguished pre- and mid-explosive communities. A common pool of explosive breeders co-occurred in most of the sites, namely *Chiasmocleis shudikarensis*, *Trachycephalus coriaceus* and *Ceratophrys cornuta*. Nevertheless, the species composition was remarkably variable between and within ponds. The acoustic structure of explosive breeding communities had outlying levels of amplitude and unexpected low acoustic diversity, significantly lower than the communities preceding explosive breeding events. Explosive breeding communities were tightly linked with specific rainfall patterns. With climate change increasing rainfall variability in tropical regions, such communities may experience significant shifts in their timing, distribution and composition. In structurally similar habitats, located in the same region without obvious barriers, our results highlight the diversity of explosive breeding communities. The characteristic acoustic structure of explosive breeding events stands out from the circadian acoustic environment being easily detected at long range, probably reflecting behavioural singularities and conveying heterospecific information announcing the availability of short-lived breeding sites. Our results shed light on the causes, patterns and consequences of anuran explosive breeding events. Combining the use of acoustic with environmental sensors would allow to establish a comprehensible and cost-effective framework to understand and manage rich amphibian communities of tropical forests.

*"Ao cair a noite, éramos incommodados pelo
coachar dos sapos, tão forte que imitava os
sons de um tambor de batuque de negros."*

– Hercule Florence (1876)

1. Introduction

Amphibians are currently the most endangered group of vertebrates, with more than 32% of its species classified as threatened (Stuart et al., 2004; Hoffman et al., 2010). Habitat loss, emerging infectious diseases, and climate changes are the primary identified sources of amphibian decline and extinction worldwide (Colin and Storfer, 2003; Beebee and Griffith, 2005; Hof et al., 2011). Understanding and forecasting global change impacts on this taxon is thus an urgent task in biology, particularly in mega-diverse tropical regions where high rate of environmental degradation and biodiversity loss are concomitantly taking place (Pounds et al., 2006; Alroy, 2015).

Amphibian anurans largely rely on acoustic communication for sexual selection and reproduction (Gerhardt and Huber, 2002; Narins et al., 2007). In these species, the temporal patterns of calling and breeding activity are influenced by multiple environmental factors, such as temperature, humidity or light intensity (Brooke et al., 2000; Oseen and Wassersug, 2002; Llusia et al., 2013b). While some anurans show long periods of calling activity and mating, known as prolonged breeders, others concentrate their reproduction during short time windows, even a few hours per year, and are known as explosive breeders (*sensu* Wells, 1977). In tropical regions, massive aggregations of explosive breeders may involve simultaneously multiple species, leading to highly-diverse anuran communities. Such phenomena typically occurs in ephemeral ponds,

2. EXPLOSIVE BREEDING

which are relatively rare in tropical forests and are likely triggered under particular weather conditions.

The structure and dynamics of these unique acoustic communities are still largely unknown probably because of their ephemeral nature, density and complexity. To our best knowledge, only a few studies have documented broad and generic patterns of calling activity of explosive neotropical anurans, linking them to weather conditions (Aichinger, 1987; Duellman, 1995) or habitat use (Prado et al. 2005), and a single study has analysed fine scale dynamics of explosive-breeding species in tropical regions (Gottsberger and Gruber, 2004). In this previous study, human observations through a four-month fieldwork in French Guiana identified two explosive breeding events with detailed information on the co-occurrence of multiple species. Gottsberger and Gruber (2004) focused on a single pond, limiting the interpretation of the results to this case study. However, replications at spatial and temporal dimensions are crucial to examine the constitution and diversity of these communities, to decipher their dynamics and to identify their link with environmental factors. Calling individuals gathering around breeding points form dense choruses characterized by a complex acoustic structure, broad masking interference and high sound pressure level. Choruses formed by tropical anurans in explosive breeding events are extreme on these features due to the extraordinary species diversity and density of calling males. Such assemblages constitute unique examples of multi-species choruses presumably eliciting complex inter-species interactions.

The technical difficulty in monitoring simultaneously these ephemeral communities has been one of the reasons for the lack of a wider geographic coverage. Traditional field-based observations are unfortunately not scalable. Alternative for cost-effective methods that measure changes in biological communities are thus needed. The emergence of new sampling techniques based on passive acoustic recorders provides a suitable approach to track such complex animal communities (Acevedo and Villanueva-Rivera, 2006; Obrist et al., 2010; Servick, 2014). These weatherproof acoustic sensors can be programmed to record for days or even months in a non-invasive and cost-efficient way, so that replication in time and space is now possible. Most anuran amphibians produce loud, stereotyped, and species-specific advertisement calls for mate attraction. These acoustic signals can be therefore remotely recorded to monitor populations as testified by several studies on temperate zone (e.g., Bridges and Dorcas,

2000; Oseen and Wassersug, 2002; Llusia et al., 2013b) and tropical species (e.g., Brooke et al., 2000; Hilje et al., 2012; Janzen et al., 2016; Pereyra et al., 2016).

Using automated sensors, we collected for the first time acoustic and environmental data to monitor simultaneously and regularly multiple explosive breeding events in tropical anuran communities, at five temporary ponds located along the Kaw Mountains in French Guiana. This systematic passive acoustic monitoring allowed us to tackle key questions related to the causes, patterns, and consequences of such a striking phenomenon. We specifically addressed four questions: (1) What are the main meteorological factors that trigger the emergence of explosive breeders? (2) Which species co-occur before and during explosive breeding events? (3) What is the variation in the acoustic community composition within and between sites? (4) What are the main acoustic patterns, spectral characteristics and diversity before and during the explosive breeding, that may shed light on the potential selective pressures shaping these complex acoustic communities?

2. Materials and methods

2.1. Study site

Explosive breeding assemblages were monitored in the lowland tropical rainforest of French Guiana, along the Kaw mountain (4°36'N; 52°16'W). As in most regions located close to the equator line, seasonal climatic variations in the study site were primarily due to changes in rainfall and humidity. The climate regime is characterized by two periods of rainfall: the main rainy season takes place from mid-November to the end of February and a less marked rainy season occurs from April to July. For this study, acoustic and environmental data were collected from the end of the dry season (10 November 2015) to the end of the main rainy season (7 March 2016).

The sampling focused on five seasonal ponds along a 30.4 km transect corresponding to the departmental road D6 (Figure 1). These temporary shallow water bodies are flooded during the rainy seasons and then dry out during periods of low rainfall. The ponds were surrounded by dense tropical forest, located between 236 and 313 meters above the sea level, and had distinct sizes, from 224.8 m² to 2240.2 m² (Table 1).

2. EXPLOSIVE BREEDING

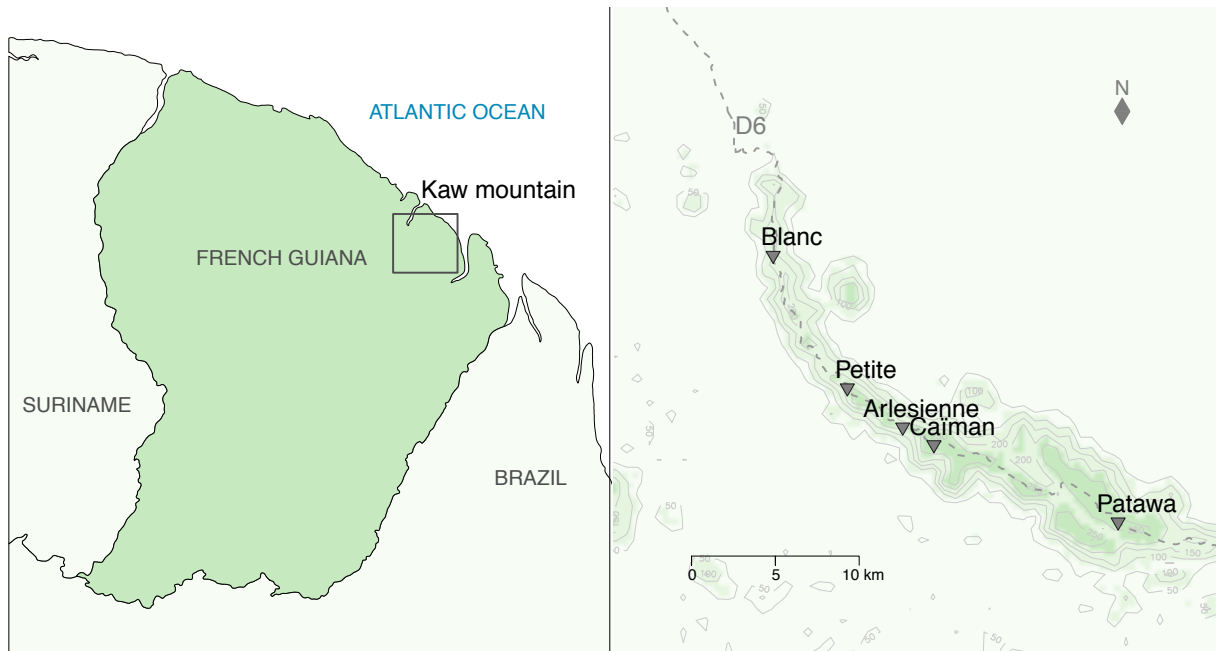


Figure 1. Location of the study area. On the left, location of the Kaw mountain in French Guiana. On the right, location of the five study sites along a 30.4 km transect on the departmental road D6.

Table 1. Altitude, location, and area of the five study ponds. Altitude is given in meters above sea level (m a.s.l.) and area in m².

| Local name | Code name | Altitude | GPS coordinates | Area |
|------------|-----------|----------|-----------------------|--------|
| Caïman | Ca | 313 | 4°34'10"N; 52°13'11"W | 1192.3 |
| Blanc | Bl | 236 | 4°40'14"N; 52°18'22"W | 399.5 |
| Patawa | Pa | 295 | 4°31'41"N; 52°07'14"W | 2240.2 |
| Arlesienne | Ar | 269 | 4°32'44"N; 52°14'11"W | 672.0 |
| Petite | Pe | 289 | 4°35'59"N; 52°15'59"W | 224.8 |

2.2. *Sampling protocol*

We monitored anuran calling activity and weather conditions simultaneously in each pond using automated sensors with a regular sampling schedule. To record the calling activity of focal species, we placed on the edge of each pond at breast height an automated sound recorder equipped with an omnidirectional microphone (SM2, Wildlife Acoustics, Inc., Concord, MA, USA). The device was set up to record data 1 minute every 29 minutes, at 44.1 kHz and 16 bit resolution, so that we obtained 5,616 recordings for each pond.

To register local abiotic environmental data, we placed next to the sound recorder a data logger (H21-002, Onset) equipped with sensors to measure three weather variables: rainfall (Onset, S-RGB-M002), temperature, and relative humidity (Onset, S-THB-M008). In addition, we retrieved two global environmental variables, atmospheric pressure (PTB220, Vaisala) and solar radiation (CMP6, Kipp and Zonen), from the nearest weather station at the Félix Eboué airport (4°50'N; 52°22'W), 19 km from the study site.

2.3. *Time-series analysis*

Because of the emergence of a great number of males from multiple amphibian species, explosive breeding events are known to produce a tremendously loud chorus. Therefore, we identified the occurrence of explosive breeding events in the audio recordings by searching for amplitude peaks lasting more than 24 hours. The overall amplitude of each recording was measured by computing the root-mean-square of the signal. Then, we applied a median filter with a 24-hour window and we searched for outliers in the smoothed time series. The outliers were defined as values distributed one-and-a-half times the inter-quartile range (IQR) above the third quartile ($Q3 + 1.5 \times IQR$). Every outlier event was inspected by listening to the recordings to confirm the presence of an explosive breeding event.

The explosive events detected on the time series showed a clear and steep increase on the calling activity from anuran assemblages. While the beginning of the explosive breeding events exhibited constant and exceptionally high call rate during around 24

2. EXPLOSIVE BREEDING

hours, calling activity later presented multiple oscillations before ceasing or returning to common levels. In order to have homologue sections for each event and compare pre- and mid-explosive breeding communities, we focused our subsequent analysis on a 48 h window, starting 24 h before the onset and ending 24 h after the onset of explosive breeding events.

We used a machine-learning framework to test whether the occurrence of the explosive breeding events could be predicted by abiotic factors. Weather conditions were considered as predictor variables and the triggering dates of the explosive breeding events as a binary response variable. The abiotic variables comprised low-level and high-level features. Low-level features were the raw quantitative meteorological measurements from the on-site sensors and the weather station, namely temperature, temperature variation, relative humidity, rainfall, atmospheric pressure, atmospheric pressure variation and solar radiation. Since the emergence of the breeding events can also be due to previous environmental conditions, we also included high-level features in the statistical analyses calculated based on the raw climatic data. These high-level features were the lagged-variables, previous 24, 48, and 72 hours, and past-cumulative variables from the previous 48 and 72 hours. The final predictor matrix included 42 variables with 466 observations. We measured prediction accuracy and variable importance on classification using the Random Forest statistical classifier (Breiman, 2001). The importance of the predictor variables was assessed by comparing the difference in misclassification error, mean decrease accuracy, between the original data and a modified set of data. The modified data for each predictor variable consisted in randomly permuted observations that are passed down the Random Forest. The higher the decrease in accuracy between the original and the modified data, the higher the importance of the predictor variable (Cutler et al., 2007).

2.4. Community diversity analysis

We investigated temporal and spatial variation on the diversity and composition of the acoustic communities of each explosive breeding event. We defined a community as the set of species observed at a given time, on a given pond. For each event, we systematically discretized the temporal gradient of 48 hours into four temporal periods of 12 hours. A first period (t1) ranged from 24 to 12 hours before the explosive breeding

2. EXPLOSIVE BREEDING

event, a second period (t2) ranged from 12 hours before to the onset of the event, a third period (t3) enclosed the first 12 hours of the event, and a fourth period (t4) ranged from 12 to 24 hours after the onset of the event.

We then sub-sampled our database by choosing one recording every two hours, for a total of 240 recordings of 60 seconds. Three of us (EC, AF and PG), who are highly trained in aural identification of anuran species of French Guiana, scrutinized each recording and annotated the occurrence of calling species. A final presence-absence vector was derived for each recording by majority voting. Thereby, potential observer bias was prevented while the accuracy of the annotations enhanced. This phase led to the identification of a total of 25 species.

Finally, we used the crossed-DPCoA (Pavoine et al., 2013), an ordination method that provides an approach for analysing the effects of crossed factors on the diversity of communities. Here we analysed the effect on the species composition of amphibian communities of the time period before or after the event (t1, t2, t3, t4), and the event (an event is one of the two breeding explosions observed at a given pond). The time period and the event are two crossed factors. The aim of crossed-DPCoA is to visualize the pattern of diversity due to a factor A knowing the existence of a crossed factor B. It helps to visualize the main effect of factor A, here on species composition, and the effect of the interaction between A and B, removing the main effect of factor B. The method first defines a space where species, communities and the levels of the two factors are visualized as points. In our first application of the crossed-DPCoA, species are equidistant in this space, then, the communities are positioned at the centroid of their constitutive species, and the levels of the factors at the centroid of communities associated with them. The method then searches for principal axes of the levels of factor A, retaining potential effects of the interaction between A and B, but removing the main effect of factor B. We analyzed first the effect of the events on the species composition of amphibian communities given the time period and then the effect of the time periods given the event.

2.5. *Acoustic diversity analysis*

To compare the anuran acoustic assemblages of the pre- and mid-explosive breeding events, we repeated the crossed DPCoA (Pavoine et al., 2013). However, here we did

2. EXPLOSIVE BREEDING

not consider species as equidistant in the space of the crossed DPCoA. Instead we used the acoustic properties of the calls of the species to define acoustic dissimilarities between pairs of species.

We estimated the acoustic dissimilarity between two species using focal recordings of each species-specific call available from personal field recordings (PG, EC, AF, JSU; n=17) and from commercial recordings (Marty and Gaucher, 2003; n=8). We selected recordings that met two criteria: (1) the call had to be emitted by an isolated individual, and (2) the signal-to-noise-ratio (SNR) of the signal had to be higher than 30, where $SNR = 20 \log_{10}(RMS_{signal}/RMS_{noise})$ and RMS is the root-mean-square amplitude of the signal. Then, the spectral composition of each call was quantified by computing a short-time Fourier transform (FFT length of 512, no overlap, Hanning window), averaging the columns of the subsequent matrix (the temporal dimension), and applying a log-transformation. The acoustic dissimilarity between the species call was assessed by computing the cumulative dissimilarity of the spectral distributions (Lellouch et al., 2014).

We also analyzed the spectral profiles of the recordings collected in the field to investigate the changes in the acoustic environment before and during the explosive breeding events. We first calculated the mean spectrum of each file. Then, we compared the spectral profiles at different moments of the explosive breeding event using a Random Forests procedure. We quantified and evaluated the classification accuracy and the importance of each feature, here each spectral profile, for the classification using the Random Forests importance measure.

Finally, we estimated the α diversity of each acoustic community by computing the species richness, the Gini-Simpson coefficient, and the quadratic entropy. The richness is the number of species in the community. The Gini-Simpson index takes into account the number of species and their proportions (Gini, 1912; Simpson, 1949). The quadratic entropy, or Rao's diversity coefficient (Rao, 1982), is based on the number of species, their proportions and incorporates a between-species dissimilarity matrix (here the pair-wise acoustic dissimilarities). For each diversity index, we tested the differences among periods of the explosive breeding event (i.e. t1, t2, t3 and t4) and between events (i.e. the first and second event per site), as well as the interaction between both factors, with a repeated-measures ANOVA. Shapiro-Wilk and Mauchly tests revealed no violation of the assumptions of normality and sphericity, respectively,

2. EXPLOSIVE BREEDING

when using ANOVA tests (in all cases: $W > 0.76$, $df = 5$, $p > 0.05$; $X^2 < 11,3$, $df = 5$, $p > 0.05$).

Acoustic and statistical analyses were computed using the R software (R Development Core Team, 2017). In particular, spectral audio features and dissimilarity matrices were computed using the Seewave R-package (Sueur et al., 2008), community and diversity ordination analyses were calculated with the adiv R-package (Pavoine, 2017), and statistical classification was computed with the randomForest R-package (Liaw and Wiener, 2002).

3. Results

3.1. Time series analysis

Applying an amplitude filter, we detected in each pond two major explosive events, i.e. 10 in total, lasting between 24 and 70 hours, later confirmed by aural evaluation. Using the combined meteorological variables (instant, lagged and past-cumulative) and the Random Forest classifier, we were able to accurately predict the emergence of all (100%) explosive breeding events with a low false positive rate of 9.6 % (out-of-the-bag results). Variable importance ranking showed that rainfall was the most influential weather determinant, in particular, the amount of rain during the previous 24 hours and most importantly the past-cumulative rainfall during the previous 48 to 72 hours (Fig 2). The rest of the variables (temperature, relative humidity, atmospheric pressure and solar radiation) had minor predictive power.

2. EXPLOSIVE BREEDING

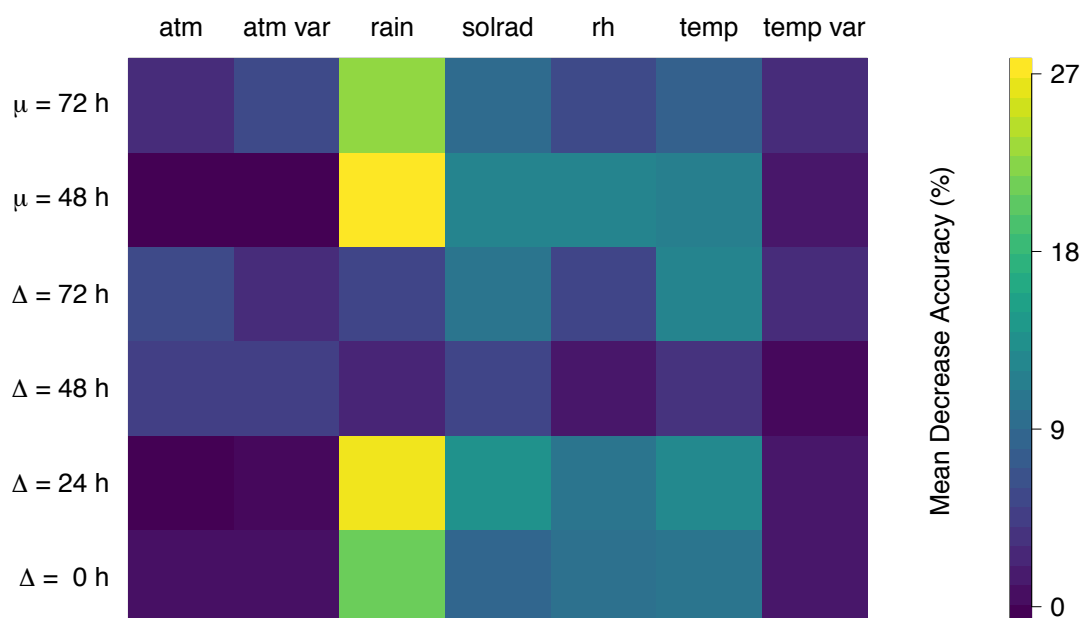


Figure 2. Variable importance measure (mean decrease accuracy) from Random Forest classification used for predicting the start of the explosive breeding events. Variables with higher values are more important for the classification. A total of 42 environmental variables were evaluated based on the combination of seven direct measurements and six derived variables. The variables directly measured were: atmospheric pressure (atm), atmospheric pressure variance (atm var), rainfall (rain), solar radiation (solrad), relative humidity (rh), temperature (temp) and temperature variance (temp var). The derived variables were based on their delay (Δ) and persistence (μ) along the time (0, 24, 48 and 72 hours).

3.2. Community diversity analysis

We first analysed the species composition of explosive breeding events using crossed-DPCoA, which allowed us to focus on the explosive breeding events removing the effect of the crossed factor linked to the time period before or after the event. The first two principal axes expressed respectively 34.8% and 30% of the main effect variability of the factor site (Figure 3a). Neither the first nor the second axis presented a particular pattern, the explosive breeding events having largely overlapping communities. Nevertheless, some sites (Patawa, Arlesienne and Petite) presented a high within-event diversity, each explosive breeding event having a particular and unique combination of species (Figure 3b). Inter-site and intra-site variability of the explosive breeding events for these sites had the same order of magnitude.

Then, to reveal the temporal variability in the acoustic signal of the events, we eliminated the crossed effect of factor 'event' in the space of the DPCoA. The calling activity of the anuran communities was clearly structured along the temporal dimension (Figure 4a). The first axis of the DPCoA, with 84.3 % of variance explained, clearly discriminated two assemblages: the pre-explosive community (t1 and t2 on the negative side) and a characteristic explosive breeding community (t4 on the positive side). A transitional community with species from both sides appeared near the origin (t3). While the pre-explosive communities (t1 and t2) were partly similar in their species composition, t3 and t4 had clear and unique species composition. The species that characterized the pre-explosive community (t1 and t2) were *Phyllomedusa tomopterna*, *Leptodactylus mystaceus*, and *Dendropsophus counani* (Figure 4b). Because they had positive coordinates on the first axis, the species that characterized the explosive breeding community (t4) were *Chiasmocleis shudikarensis*, *Trachycephalus coriaceus* and *Ceratophrys cornuta* (Figure 4b). The transitional community (t3) showed an intermediate place on the ordination; these communities had a balanced mixed of pre-explosive and explosive breeding species.

2. EXPLOSIVE BREEDING

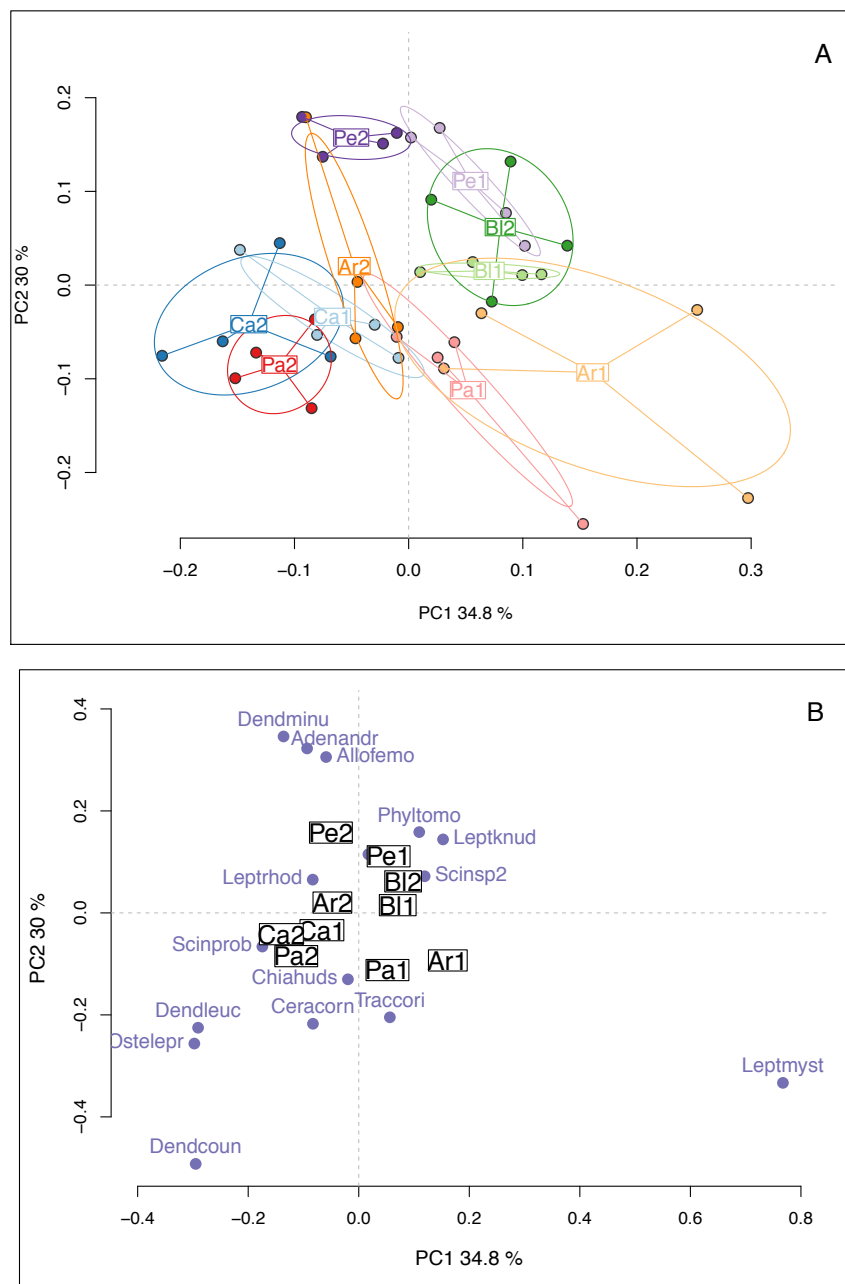


Figure 3. Diversity of the species composition in explosive breeding events. The two principal axes (64.8 % of variance explained) of the crossed DPCoA analysis are plotted. (A) Diversity between and within communities. Each point is a community. The communities were color-coded with the levels of the factor event. Code names for events are: Ar = Arlesienne, Bl = Blanc, Ca = Caïman, Pa = Patawa, Pe = Petite. The number that follows the code name distinguishes the explosive breeding event in each site. (B) Coordinates of the constitutive species in the axes. Each point is a species. Only the species that had the highest values on the axes were named. Code names for the species are: *Adenomera andreae* = Adenandr, *Allobates femoralis* = Allofemo, *Ceratophrys cornuta* = Ceracorn, *Chiasmocleis hudsoni* = Chiahuds, *Dendropsophus counani* = Dendcoun, *Dendropsophus leucophyllatus* = Dendleuc, *Dendropsophus minutus* = Dendminu, *Leptodactylus knudseni* = Leptknud, *Leptodactylus mystaceus* = Leptmyst, *Leptodactylus rhodomystax* = Leptrhod, *Osteocephalus leprieurii* = Ostelepr, *Phyllomedusa tomopterna* = Phyltomo, *Scinax sp2* = Scinsp2, *Trachycephalus coriaceus* = Traccori.

2. EXPLOSIVE BREEDING

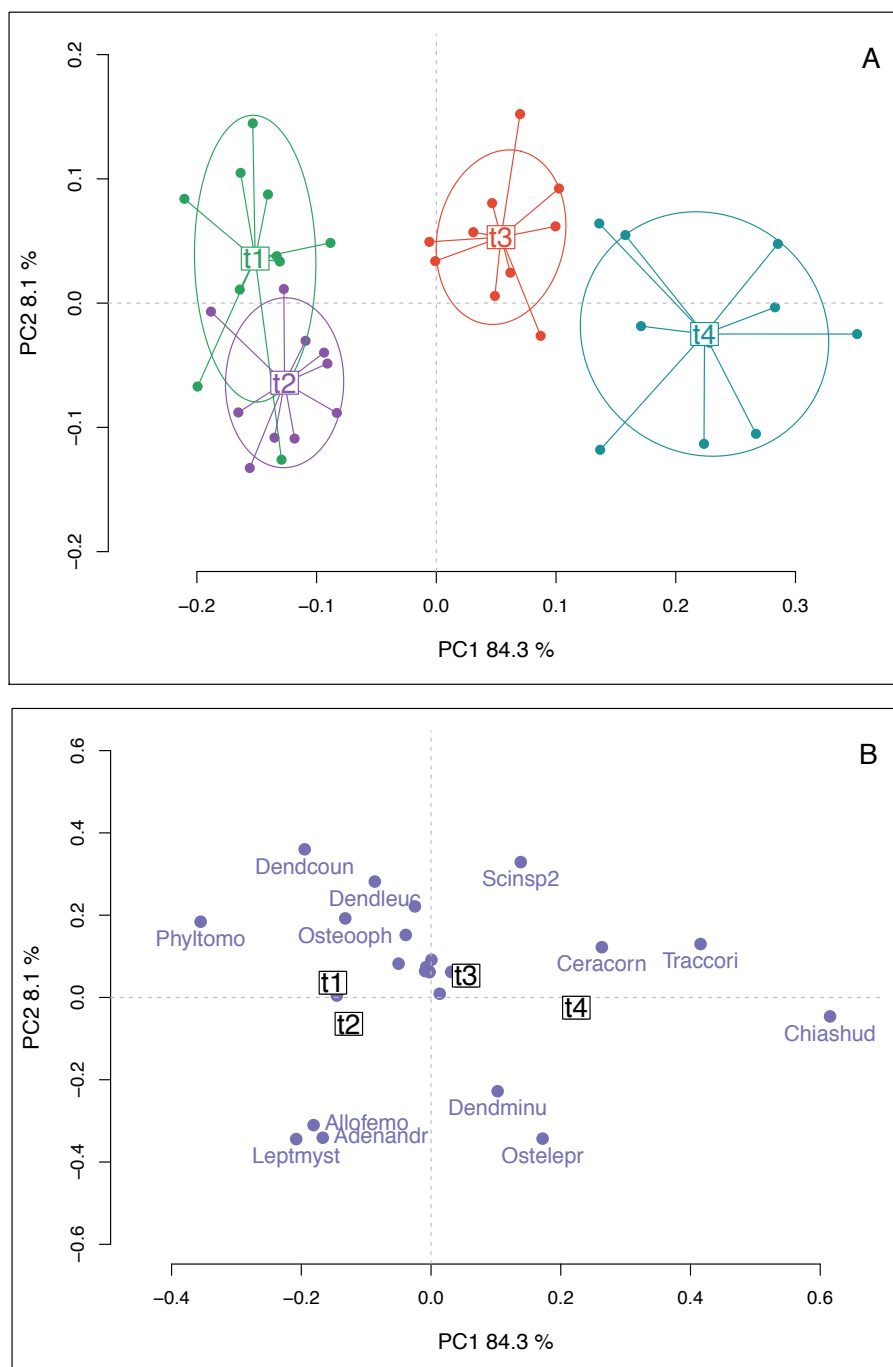


Figure 4. Temporal effect on the species composition in explosive breeding events. The two principal axes (92.4 % of variance explained) of the crossed DPCoA analysis are plotted showing (A) Diversity between and within communities along the time. Each point is a community. The communities were color-coded with the levels of factor time: t1, t2, t3, and t4. (B) Coordinates of the constitutive species in the axes. Each point is a species. Only the species that had the highest values on the axes were named: *Adenomera andreae* = Adenandr, *Allobates femoralis* = Allofemo, *Ceratophrys cornuta* = Ceracorn, *Chiasmocleis shudikarensis* = Chiashud, *Dendropsophus counani* = Dendcoun, *Dendropsophus leucophyllatus* = Dendleuc, *Dendropsophus minutus* = Dendminu, *Leptodactylus mystaceus* = Leptmyst, *Osteocephalus leprieurii* = Ostelepr, *Osteocephalus oophagus* = Osteooph, *Phyllomedusa tomopterna* = Phyltomo, *Scinax sp2* = Scinsp2, *Trachycephalus coriaceus* = Traccori.

3.3. Acoustic diversity analysis

The species were organised according to the characteristics of their calls when treated with a crossed-DPCoA embedding the acoustic dissimilarity matrix. The first and second axes were strongly correlated with the peak frequency of the calls ($r = 0.91$ and $r = 0.96$). For both axes, low frequency calls lied on the negative and high frequency calls on the positive side (Figure 5a). The levels of the temporal factor were positioned at the centroid of the acoustic community.

As in the previous anuran community analysis, the acoustic community was structured along the temporal dimension, as revealed by a crossed DPCoA (Figure 5a). The first axis of the ordination analysis, with 60.1 % of explained variance, showed a progression from t1 (negative side) to t4 (positive side). The sounds that characterized, by their higher proportions, the explosive breeding event acoustics were the calls of *C. shudikarensis* and *T. coriaceus* (Figure 5b). The calls of these anurans were in the middle range of the acoustic community, 3.4 kHz and 1.8 kHz for *C. shudikarensis* and *T. coriaceus* respectively.

The levels t1 and t2 presented elongated ellipses, showing a dispersed range of frequency calls, with low and high pitched sounds (Figure 5a). This elongated shape was much less pronounced for levels t3 and t4, which were mainly characterized by calls in the mid frequency range. Indeed, the dominance of mid-frequencies showed to be a distinctive spectral trait of explosive breeding events (Fig 6a).

We were able to classify the frequency spectra of explosive breeding events with a high accuracy (Random Forests, 89% out-of-the-bag accuracy). The feature importance analysis showed that mid frequencies, between 3 and 4.4 kHz, were clearly the most important predictor variables (Fig 6b).

2. EXPLOSIVE BREEDING

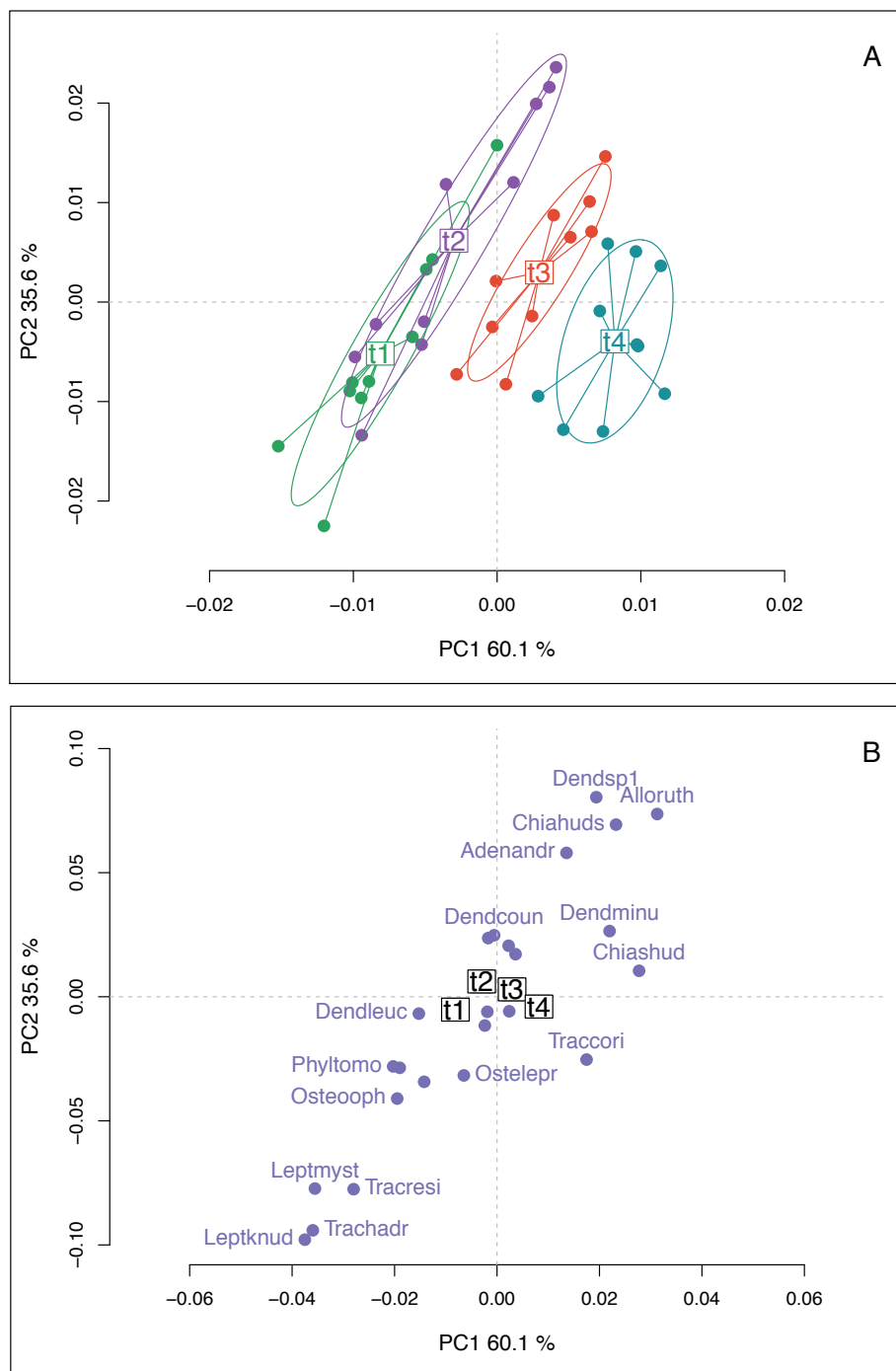


Figure 5. Acoustic diversity of anuran communities. The two principal axes (95.7 % of variance explained) of the crossed DPCoA analysis were plotted. (A) Diversity between and within communities. Each point is a community. Communities were color-coded with levels of factor time: t1, t2, t3, t4. (B) Coordinates of the constitutive species-specific calls in the principal axes. Each point is a species. Only the calls with higher values on the axes were named: *Adenomera andreae* = Adenandr, *Allophryne ruthveni* = Alloruth, *Chiasmocleis hudsoni* = Chiahuds, *Chiasmocleis shudikarensis* = Chiashud, *Dendropsophus counani* = Dendcoun, *Dendropsophus leucophyllatus* = Dendleuc, *Dendropsophus minutus* = Dendminu, *Dendropsophus sp1* = Dendsp1, *Leptodactylus knudseni* = Leptknud, *Leptodactylus mystaceus* = Leptmyst, *Osteocephalus leprieurii* = Ostelepr, *Osteocephalus oophagus* = Osteoph, *Phyllomedusa tomopterna* = Phyltomo, *Trachycephalus coriaceus* = Traccori, *Trachycephalus hadroceps* = Trachadr, *Trachycephalus resinifictrix* = Tracresi.

2. EXPLOSIVE BREEDING

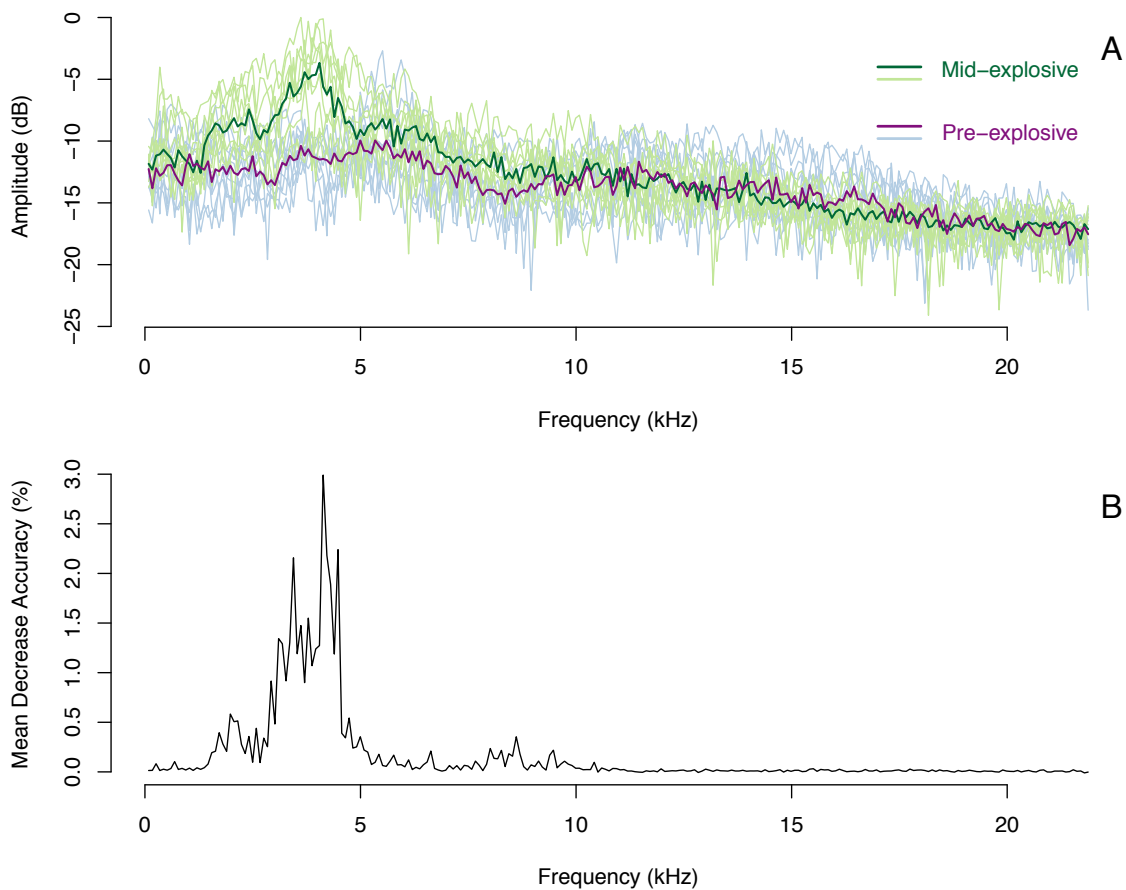


Figure 6. Illustrative comparison between pre-explisive (pre) and explosive (mid) breeding frequency spectra. (A) In light colours, 10 random samples of each acoustic community. In dark blue (pre-explisive) and black (mid-explisive), the median spectrum of the samples. (B) Variable importance measure (mean decrease accuracy) from Random Forest classification used for discriminating the acoustics of the explosive breeding events. Frequencies with higher values are more important for the classification.

2. EXPLOSIVE BREEDING

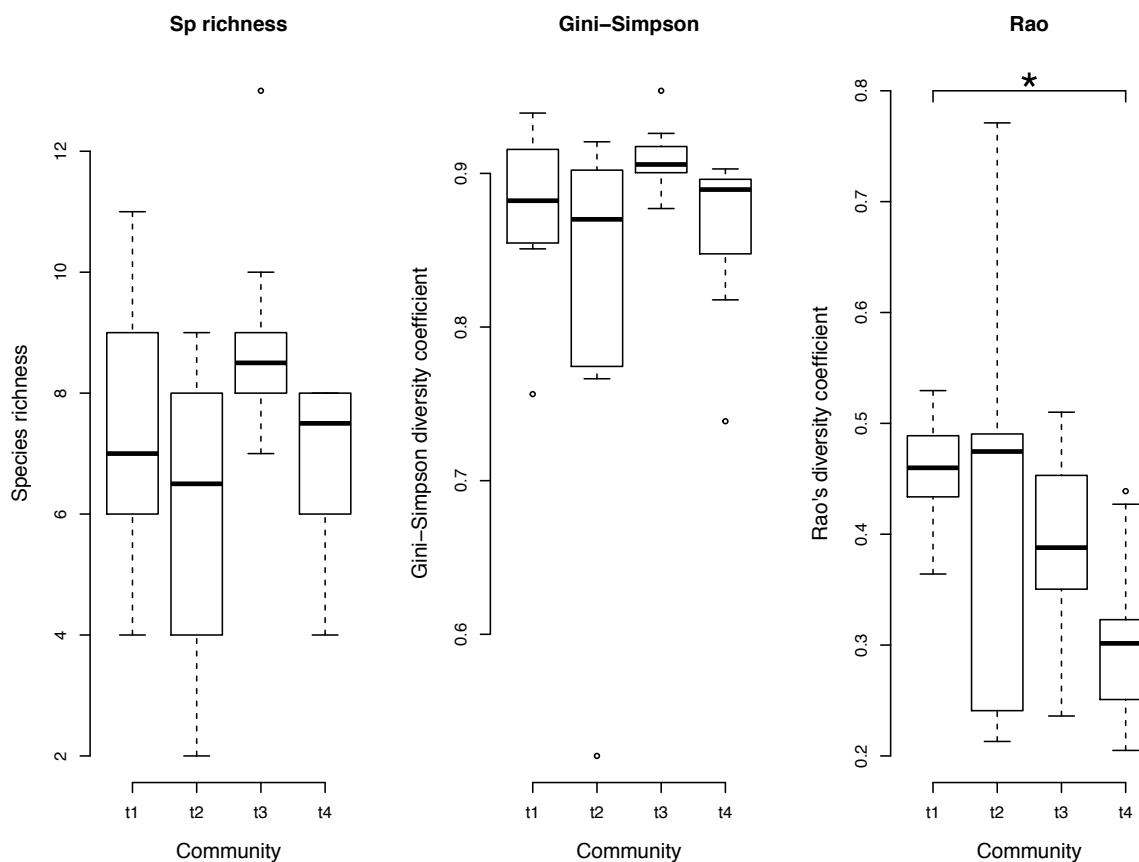


Figure 7. Diversity measures within each of the temporal community (t1, t2, t3 and t4). Three diversity indices are compared: species richness, Gini-Simpson diversity and Rao's diversity coefficient.

The temporal pattern observed using the Gini-Simpson index and species richness was similar, with maximal values during the first hours of the explosive breeding event (period t3; Figure 7). Differences in acoustic diversity among periods were marginally significant when measured by Gini-Simpson index (ANOVA, $F_{3,12} = 3.21$, $p = 0.062$) and significant, at a nominal type-I error of 5%, when measured as species richness (ANOVA, $F_{3,12} = 5.86$, $p = 0.010$). Rao's diversity coefficient, which includes the acoustic dissimilarity matrix, varied according to periods (ANOVA, $F_{3,12} = 5.72$, $p = 0.011$), indicating a significant decrease in the acoustic diversity as the explosive breeding community predominates (Figure 7). No effect of the season nor its interaction with the periods of the event were identified in all cases (ANOVA, $F_{1,4} < 4.48$, $p >$

0.101), and hence the two explosive breeding events recorded per site, during each of the two rainy seasons, were equivalent in terms of acoustic diversity.

4. Discussion

4.1. Time series analysis

We found that environmental variables could predict the emergence of explosive breeding events in the studied communities, with rain as the most important predictor variable. While rain is abundant during the whole season, it is relevant to note that explosive breeding species respond to two specific patterns of rain: consistency during the previous 48 to 72 hours and amount during previous 24 hours. Our results are in agreement with those of Gottsberger and Gruber (2004), who found that rainfall for the previous 24 hours contributed the best, among other environmental variables, to explain the calling activity of the explosive breeding species. As we included more derived variables of the rain in our analyses, we complement previous results asserting that the consistency of the rain is also crucial. Having replicated this observation at several sites, we confirm that species participating in explosive breeding events are highly tuned to specific rainfall patterns.

This apparently high dependency of explosive breeders' reproduction not only to the amount of precipitation but also to the timing of rain events raises the question of the vulnerability of explosive breeders to climate changes. Recent investigations on the causes of amphibian declines have studied the role of climate change as a global impact (Carey and Alexander, 2003; Lips et al., 2008; Bellard et al., 2012). In addition to the climate-linked epidemic hypothesis, research has focused on the effect of climate change on behaviour, reproduction and distribution of amphibians (Araujo et al., 2006; Llusia et al., 2013a). As ectotherms, alterations on temperature and rainfall regimes might strongly affect key aspects of amphibian life cycles, even jeopardizing their survival (Duarte et al., 2012). Both theoretical and experimental studies suggest that low latitude ectothermic species are more vulnerable to climate changes than their higher latitude counterparts (Deutsch et al., 2008). Tropical species indeed tend to have narrower thermal tolerance (Bonetti and Wiens, 2014) and their actual habitat

2. EXPLOSIVE BREEDING

temperatures are closer to their upper thermal limit (Sunday et al, 2011; Duarte et al., 2012). Even slight changes in environmental conditions might therefore have a strong effect on tropical species (Foden et al., 2013).

While other factors such as programmed annual migration might be involved in triggering explosive breeding events, our study suggests that the two specific patterns of rain (i.e. consistency during the previous 48 to 72 hours and amount during previous 24 hours) are key parameters for the initiation of reproduction. With climate change increasing rainfall variability in tropical regions (Feng et al., 2013), reproductive events might be shifted or disrupted. Moreover, these species rely on very specific habitats (reproductive ponds) for their reproduction, another factor that is known to increase vulnerability to climate changes (Foden et al., 2013). Finally, the high number of individuals from several species at the time of reproduction might increase probability of intra and inter-species infection at the breeding ponds and therefore increase the sensitivity of these species to emerging infectious disease. These combined factors, may lead to significant shifts in the timing, distribution and composition of explosive breeding communities, which may desynchronize phenology and other biological responses throughout several trophic levels in the ecosystem (Schwartz, 2013).

4.2. Acoustic community diversity analyses

In structurally similar habitats, located in the same region without obvious barriers, we expected to have homogeneous amphibian communities. Yet, our results highlight the variability of species composition in explosive breeding events. The ordination diagram showed differences in species composition both between ponds and for a given pond, between the two observed events. In other words, each explosive breeding event, while often sharing a common pool of species, had a unique combination of species. When controlling for the differences between explosive breeding communities, a clear temporal factor structured the acoustic community during explosive breeding events, showing pronounced differences between pre-explosive and explosive breeding communities. The main species characterising the explosive breeding event, *C. shudikarensis*, *T. coriaceus* and *C. cornuta*, were also found as predominant species in explosive breeding events in the Arataï river, more than 100 km away from our study site (Gottsberger and Gruber, 2004). While other species are also present during these

2. EXPLOSIVE BREEDING

aggregations, these species seem particularly representative of the acoustic community. It remains to understand how to explain the species turnover between events in space (ponds) and time (for each pond between the first and the second event). This turnover could be related to multiple combined factors, such as ecological and behavioural traits.

4.3. Acoustic diversity analyses

Regarding the acoustic environment of explosive breeding events, we found outlying levels of activity with a characteristic spectral signature. This signature stand out from the circadian acoustic environment and can be easily detected at long range. Acoustic signatures convey information that could be exploited by conspecifics (or heterospecific) for general orientation within a landscape (Slabbekoorn and Bouton, 2008). Fish and crustacean larvae (Montgomery et al. 2006), birds (Mönkkönen et al., 1990), and frogs (Gerhardt and Klump 1988; Bee 2007) are known to use sounds in the environment for spatial orientation. Indeed, acoustic cues might gain importance for anuran explosive breeding species since sounds may signal availability, in space and time, of short-lived breeding sites (Bee 2007, Swanson et al, 2007).

Alpha diversity indices, measured with species richness and Gini-Simpson, showed temporal communities with similar values between pre-explosive (t1-t2) and the explosive breeding community (t4). The transitional community (t3) had higher values, probably because it had species from both communities, pre- and explosive breeding. More surprisingly, Rao's diversity coefficient showed a significant diminution of the spectral diversity during the explosive breeding event (t4). Even when the number of singing species was similar, we observed more frequency overlap in signals for the explosive breeding community than for the pre-explosive community. Species belonging to a community may compete to access acoustic resources, that is to a free acoustic channel. It has been therefore hypothesized that species calling in a chorus should exhibit few frequency overlap. Formulated under the acoustic niche hypothesis, organisms would have evolved to occupy specific spectro-temporal 'niches', decreasing the risk of heterospecific mating and information masking (Krause 1993). Acoustic partitioning has been observed in multiple taxa, such as insects (Sueur et al., 2008, Schmidt et al., 2013), birds (Planqué and Slabbekoorn, 2008) and amphibians (Amézquita et al., 2011). However, recent studies also presented limitations of such

2. EXPLOSIVE BREEDING

hypothesis, showing no significant spectral divergence in cricket assemblages (Schmidt et al., 2016) and more similarity in signal design than expected by chance for tropical forest birds (Tobias et al., 2014). Our results are in line with these last studies; contrary to our prediction, the species did not show frequency dispersion but frequency overlap.

In our study, the frequency overlap is higher for the explosive breeding community than for the pre-explosive community, which might be explained by behavioural differences between these communities. As discussed by Wells (1977), the social behaviour of prolonged and explosive breeding species is distinct and might concern the form of male-male competition. While for prolonged breeders, female choice might be crucial for determining male reproductive success, for explosive breeders, females would have few opportunities to select among potential mates. An extremely short breeding period places selective premium on rapid acquisition of males, hence the selection for males by acoustic signals would be of low intensity. Because of the short time for exchange of signals between individuals, males would compete physically and not acoustically. Similar patterns may occur in other organisms which engage in group displays: in all such organisms, weak sexual selection would lead to less structured acoustic communities. Additional data including, for instance phylogenetic or functional traits, could increase our knowledge of this striking ecological event shedding light on the selective pressure driving widespread chorusing behaviour.

Acoustic signatures could be used by humans as a suitable way to monitor wildlife, not only at the individual or population level, but also at the community level. Our findings indicate that multiple spatial and temporal scales should be considered for precisely monitoring these communities. Moreover, the changes in the community were clearly reflected on a change of the spectrum of the acoustic environment. Under the conceptual framework of ecoacoustics (Sueur and Farina, 2015), recent studies have identified important changes in communities using acoustic indices (Lellouch et al., 2014). Most of the acoustic indices are fast and easy to compute, and hence could provide a straightforward method to track dynamics of explosive breeding species. In this study, we coupled biotic and abiotic variables, revealing changes in the anuran community at multiple spatiotemporal scales and their tight link with the environment. Such data provides a baseline against which future changes can be measured, contributing to a better understanding and hopefully to a better management of such

2. EXPLOSIVE BREEDING

unique communities. A more widespread use of standardized methods combining passive acoustic recorders with a monitoring of key environmental parameters would become a comprehensible and cost-efficient framework to improve our knowledge and manage rich animal communities of tropical forests.

Acknowledgments

This research was supported by the Labex CEBA (Centre d'Étude de la Biodiversité Amazonienne, call for proposals 2015) and the SABIOD MASTODONS Big Data (CNRS MI project 2012–2017). We would like to warmly thank Jeannot and Odette for their help during our stay at the *Gîte Patawa*. JSU was supported by COLCIENCIAS (Doctoral Scholarship of the Colombian government, call for proposals 2014, number 646).

References

- Acevedo, M.A., Villanueva-Rivera, L.J., 2006. Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. *Wildlife Society Bulletin* 34, 211–214.
- Aichinger, M., 1987. Annual activity patterns of anurans in a seasonal neotropical environment. *Oecologia* 71, 583–592.
- Alan Pounds, J., Bustamante, M.R., Coloma, L.A., Consuegra, J.A., Fogden, M.P.L., Foster, P.N., La Marca, E., Masters, K.L., Merino-Viteri, A., Puschendorf, R., Ron, S.R., Sánchez-Azofeifa, G.A., Still, C.J., Young, B.E., 2006. Widespread amphibian extinctions from epidemic disease driven by global warming. *Nature* 439, 161–167. <https://doi.org/10.1038/nature04246>
- Alroy, J., 2015. Current extinction rates of reptiles and amphibians. *PNAS* 112, 13003–13008. <https://doi.org/10.1073/pnas.1508681112>
- Amézquita, A., Flechas, S.V., Lima, A.P., Gasser, H., Hödl, W., 2011. Acoustic interference and recognition space within a complex assemblage of dendrobatid frogs. *PNAS* 108, 17058–17063. <https://doi.org/10.1073/pnas.1104773108>
- Araújo, M.B., Thuiller, W., Pearson, R.G., 2006. Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography* 33, 1712–1728. <https://doi.org/10.1111/j.1365-2699.2006.01482.x>
- Bee, M.A., 2007. Selective phonotaxis by male wood frogs (*Rana sylvatica*) to the sound of a chorus. *Behav Ecol Sociobiol* 61, 955–966. <https://doi.org/10.1007/s00265-006-0324-8>
- Beebee, T.J.C., Griffiths, R.A., 2005. The amphibian decline crisis: A watershed for conservation biology? *Biological Conservation* 125, 271–285. <https://doi.org/10.1016/j.biocon.2005.04.009>
- Bellard, C., Thuiller, W., Leroy, B., Genovesi, P., Bakkenes, M., Courchamp, F., 2013. Will climate change promote future invasions? *Global Change Biology* 19, 3740–3748. <https://doi.org/10.1111/gcb.12344>
- Bonetti, M.F., Wiens, J.J., 2014. Evolution of climatic niche specialization: a phylogenetic analysis in amphibians. *Proceedings of the Royal Society B: Biological Sciences* 281, 20133229–20133229. <https://doi.org/10.1098/rspb.2013.3229>

2. EXPLOSIVE BREEDING

- Bridges, A.S., Dorcas, M.E., 2000. Temporal variation in anuran calling behavior: implications for surveys and monitoring programs. *Copeia* 2000, 587–592. [https://doi.org/10.1643/0045-8511\(2000\)000\[0587:TVIACB\]2.0.CO;2](https://doi.org/10.1643/0045-8511(2000)000[0587:TVIACB]2.0.CO;2)
- Brooke, P.N., Alford, R.A., Schwarzkopf, L., 2000. Environmental and social factors influence chorusing behaviour in a tropical frog: examining various temporal and spatial scales. *Behav Ecol Sociobiol* 49, 79–87. <https://doi.org/10.1007/s002650000256>
- Carey, C., Alexander, M.A., 2003. Climate change and amphibian declines: is there a link? *Diversity and Distributions* 9, 111–121. <https://doi.org/10.1046/j.1472-4642.2003.00011.x>
- Collins, J.P., Storfer, A., 2003. Global amphibian declines: sorting the hypotheses. *Diversity and Distributions* 9, 89–98. <https://doi.org/10.1046/j.1472-4642.2003.00012.x>
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. *Ecology* 88, 2783–2792. <https://doi.org/10.1890/07-0539.1>
- Deutsch, C.A., Tewksbury, J.J., Huey, R.B., Sheldon, K.S., Ghalambor, C.K., Haak, D.C., Martin, P.R., 2008. Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences* 105, 6668–6672.
- Duarte, H., Tejedó, M., Katzenberger, M., Marangoni, F., Baldo, D., Beltrán Juan Francisco, Martí Dardo Andrea, Richter-Boix Alex, Gonzalez-Voyer Alejandro, 2011. Can amphibians take the heat? Vulnerability to climate warming in subtropical and temperate larval amphibian communities. *Global Change Biology* 18, 412–421. <https://doi.org/10.1111/j.1365-2486.2011.02518.x>
- Duellman, W.E., 1995. Temporal Fluctuations in Abundances of Anuran Amphibians in a Seasonal Amazonian Rainforest. *Journal of Herpetology* 29, 13. <https://doi.org/10.2307/1565080>
- Feng, X., Porporato, A., Rodriguez-Iturbe, I., 2013. Changes in rainfall seasonality in the tropics. *Nature Climate Change* 3, 811–815. <https://doi.org/10.1038/nclimate1907>
- Florence, H., 1876. Zoophonia. *Revista Trimestral do Instituto Histórico e Ethnographico do Brasil* XXXIX.

- Foden, W.B., Butchart, S.H.M., Stuart, S.N., Vié, J.-C., Akçakaya, H.R., Angulo, A., DeVantier, L.M., Gutsche, A., Turak, E., Cao, L., Donner, S.D., Katariya, V., Bernard, R., Holland, R.A., Hughes, A.F., O'Hanlon, S.E., Garnett, S.T., Şekercioğlu, Ç.H., Mace, G.M., 2013. Identifying the World's Most Climate Change Vulnerable Species: A Systematic Trait-Based Assessment of all Birds, Amphibians and Corals. *PLoS ONE* 8, e65427. <https://doi.org/10.1371/journal.pone.0065427>
- Gerhardt, H.C., Huber, F., 2002. Acoustic communication in insects and anurans: common problems and diverse solutions. University of Chicago Press, Chicago.
- Gerhardt, H.C., Klump, G.M., 1988. Phonotactic responses and selectivity of barking treefrogs (*Hyla gratiosa*) to chorus sounds. *J. Comp. Physiol.* 163, 795–802. <https://doi.org/10.1007/BF00604056>
- Gini, C., 1912. Variabilità e mutabilità, in: *Contributo Allo Studio Delle Distribuzioni e Delle Relazioni Statistiche*. C. Cuppini, Bologna.
- Gottsberger, B., Gruber, E., 2004. Temporal partitioning of reproductive activity in a neotropical anuran community. *Journal of Tropical Ecology* 20, 271–280. <https://doi.org/10.1017/S0266467403001172>
- Hilje, B., Aide, T.M., 2012. Recovery of amphibian species richness and composition in a chronosequence of secondary forests, northeastern Costa Rica. *Biological Conservation* 146, 170–176. <https://doi.org/10.1016/j.biocon.2011.12.007>
- Hof, C., Araújo, M.B., Jetz, W., Rahbek, C., 2011. Additive threats from pathogens, climate and land-use change for global amphibian diversity. *Nature* 480, 516–519. <https://doi.org/10.1038/nature10650>
- Hoffmann, M., Hilton-Taylor, C., Angulo, A., Böhm, M., Brooks, T.M., Butchart, S.H.M., Carpenter, K.E., Chanson, J., Collen, B., Cox, N.A., Darwall, W.R.T., Dulvy, N.K., Harrison, L.R., Katariya, V., Pollock, C.M., Quader, S., Richman, N.I., Rodrigues, A.S.L., Tognelli, M.F., Vié, J.-C., Aguiar, J.M., Allen, D.J., Allen, G.R., Amori, G., Ananjeva, N.B., Andreone, F., Andrew, P., Ortiz, A.L.A., Baillie, J.E.M., Baldi, R., Bell, B.D., Biju, S.D., Bird, J.P., Black-Decima, P., Blanc, J.J., Bolaños, F., Bolivar-G, W., Burfield, I.J., Burton, J.A., Capper, D.R., Castro, F., Catullo, G., Cavanagh, R.D., Channing, A., Chao, N.L., Chenery, A.M., Chiozza, F., Clausnitzer, V., Collar, N.J., Collett, L.C., Collette, B.B., Fernandez, C.F.C., Craig, M.T., Crosby, M.J., Cumberlidge, N., Cuttelod, A.,

2. EXPLOSIVE BREEDING

- Derocher, A.E., Diesmos, A.C., Donaldson, J.S., Duckworth, J.W., Dutson, G., Dutta, S.K., Emslie, R.H., Farjon, A., Fowler, S., Freyhof, J., Garshelis, D.L., Gerlach, J., Gower, D.J., Grant, T.D., Hammerson, G.A., Harris, R.B., Heaney, L.R., Hedges, S.B., Hero, J.-M., Hughes, B., Hussain, S.A., M, J.I., Inger, R.F., Ishii, N., Iskandar, D.T., Jenkins, R.K.B., Kaneko, Y., Kottelat, M., Kovacs, K.M., Kuzmin, S.L., Marca, E.L., Lamoreux, J.F., Lau, M.W.N., Lavilla, E.O., Leus, K., Lewison, R.L., Lichtenstein, G., Livingstone, S.R., Lukoschek, V., Mallon, D.P., McGowan, P.J.K., McIvor, A., Moehlman, P.D., Molur, S., Alonso, A.M., Musick, J.A., Nowell, K., Nussbaum, R.A., Olech, W., Orlov, N.L., Papenfuss, T.J., Parra-Olea, G., Perrin, W.F., Polidoro, B.A., Pourkazemi, M., Racey, P.A., Ragle, J.S., Ram, M., Rathbun, G., Reynolds, R.P., Rhodin, A.G.J., Richards, S.J., Rodríguez, L.O., Ron, S.R., Rondinini, C., Rylands, A.B., Mitcheson, Y.S. de, Sanciangco, J.C., Sanders, K.L., Santos-Barrera, G., Schipper, J., Self-Sullivan, C., Shi, Y., Shoemaker, A., Short, F.T., Sillero-Zubiri, C., Silvano, D.L., Smith, K.G., Smith, A.T., Snoeks, J., Stattersfield, A.J., Symes, A.J., Taber, A.B., Talukdar, B.K., Temple, H.J., Timmins, R., Tobias, J.A., Tsytsulina, K., Tweddle, D., Ubeda, C., Valenti, S.V., Dijk, P.P. van, Veiga, L.M., Veloso, A., Wege, D.C., Wilkinson, M., Williamson, E.A., Xie, F., Young, B.E., Akçakaya, H.R., Bennun, L., Blackburn, T.M., Boitani, L., Dublin, H.T., Fonseca, G.A.B. da, Gascon, C., Lacher, T.E., Mace, G.M., Mainka, S.A., McNeely, J.A., Mittermeier, R.A., Reid, G.M., Rodriguez, J.P., Rosenberg, A.A., Samways, M.J., Smart, J., Stein, B.A., Stuart, S.N., 2010. The Impact of Conservation on the Status of the World's Vertebrates. *Science* 1194442. <https://doi.org/10.1126/science.1194442>
- Jansen, M., Plath, M., Brusquetti, F., Ryan, M.J., 2016. Asymmetric frequency shift in advertisement calls of sympatric frogs. *Amphibia-Reptilia* 37, 137–152. <https://doi.org/10.1163/15685381-00003038>
- Krause, B.L., 1993. The niche hypothesis: a virtual symphony of animal sounds, the origins of musical expression and the health of habitats. *The Soundscape Newsletter* 6, 4–6.
- Lellouch, L., Pavoine, S., Jiguet, F., Glotin, H., Sueur, J., 2014. Monitoring temporal change of bird communities with dissimilarity acoustic indices. *Methods in Ecology and Evolution* 5, 495–505. <https://doi.org/10.1111/2041-210X.12178>

- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. *R News* 2, 18–22.
- Lips, K.R., Diffendorfer, J., Iii, J.R.M., Sears, M.W., 2008. Riding the Wave: Reconciling the Roles of Disease and Climate Change in Amphibian Declines. *PLOS Biology* 6, e72. <https://doi.org/10.1371/journal.pbio.0060072>
- Llusia, D., Márquez, R., Beltrán, J.F., Benítez, M., do Amaral, J.P., 2013a. Calling behaviour under climate change: geographical and seasonal variation of calling temperatures in ectotherms. *Glob Chang Biol* 19, 2655–2674. <https://doi.org/10.1111/gcb.12267>
- Llusia, D., Márquez, R., Beltrán, J.F., Moreira, C., Amaral, J.P. do, 2013b. Environmental and social determinants of anuran lekking behavior: intraspecific variation in populations at thermal extremes. *Behav Ecol Sociobiol* 67, 493–511. <https://doi.org/10.1007/s00265-012-1469-2>
- Marty, C., Gaucher, P., 1999. Sound guide to the tailless amphibians of French Guiana. Ceba (Centre Bioacoustique Alpin), France.
- Mönkkönen, M., Helle, P., Soppela, K., 1990. Numerical and behavioural responses of migrant passerines to experimental manipulation of resident tits (*Parus* spp.): heterospecific attraction in northern breeding bird communities? *Oecologia* 85, 218–225. <https://doi.org/10.1007/BF00319404>
- Montgomery, J.C., Jeffs, A., Simpson, S.D., Meekan, M., Tindle, C., 2006. Sound as an orientation cue for the pelagic larvae of reef fishes and decapod crustaceans. *Adv. Mar. Biol.* 51, 143–196. [https://doi.org/10.1016/S0065-2881\(06\)51003-X](https://doi.org/10.1016/S0065-2881(06)51003-X)
- Narins, P.M., Feng, A.S., Fay, R.R., Popper, A.N. (Eds.), 2006. Hearing and Sound Communication in Amphibians, *Springer Handbook of Auditory Research*. Springer New York. <https://doi.org/10.1007/978-0-387-47796-1>
- Obrist, M.K., Pavan, G., Sueur, J., Riede, K., Llusia, D., Márquez, R., 2010. Bioacoustics approaches in biodiversity inventories, in: Eymann, J., Degreef, J., Häuser, C., Monje, J.C., Samyn, Y., Vandenspiegel, Y.D. (Eds.), *Manual on Field Recording Techniques and Protocols for All Taxa Biodiversity Inventories and Monitoring*. ABC Taxa, pp. 18–48.

- Oseen, K.L., Wassersug, R.J., 2002. Environmental factors influencing calling in sympatric anurans. *Oecologia* 133, 616–625. <https://doi.org/10.1007/s00442-002-1067-5>
- Pavoine, S., 2017. *adiv: Analysis of Diversity*, version 1.1.
- Pavoine, S., Blondel, J., Dufour, A.B., Gasc, A., Bonsall, M.B., 2013. A New Technique for Analysing Interacting Factors Affecting Biodiversity Patterns: Crossed-DPCoA. *PLoS ONE* 8, e54530. <https://doi.org/10.1371/journal.pone.0054530>
- Pereyra, L.C., Akmentins, M.S., Sanabria, E.A., Vaira, M., 2016. Diurnal? Calling activity patterns reveal nocturnal habits in the aposematic toad *Melanophryniscus rubriventris*. *Canadian Journal of Zoology* 94, 497–503. <https://doi.org/10.1139/cjz-2015-0197>
- Planqué, R., Slabbekoorn, H., 2008. Spectral Overlap in Songs and Temporal Avoidance in a Peruvian Bird Assemblage. *Ethology* 114, 262–271. <https://doi.org/10.1111/j.1439-0310.2007.01461.x>
- Prado, C., Uetanabaro, M., Haddad, C., 2005. Breeding activity patterns, reproductive modes, and habitat use by anurans (Amphibia) in a seasonal environment in the Pantanal, Brazil. *Amphibia-Reptilia* 26, 211–221.
- R Core Team, 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rao, C.R., 1982. Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology* 21, 24–43. [https://doi.org/10.1016/0040-5809\(82\)90004-1](https://doi.org/10.1016/0040-5809(82)90004-1)
- Schmidt, A.K.D., Riede, K., Römer, H., 2016. No phenotypic signature of acoustic competition in songs of a tropical cricket assemblage. *Behav Ecol* 27, 211–218. <https://doi.org/10.1093/beheco/arv141>
- Schmidt, A.K.D., Römer, H., Riede, K., 2013. Spectral niche segregation and community organization in a tropical cricket assemblage. *Behavioral Ecology* 24, 470–480. <https://doi.org/10.1093/beheco/ars187>
- Schwartz, M.D., 2013. *Phenology: an integrative environmental science*.
- Servick, K., 2014. Eavesdropping on Ecosystems. *Science* 343, 834–837. <https://doi.org/10.1126/science.343.6173.834>
- Simpson, E.H., 1949. Measurement of Diversity. *Nature* 163, 688.

- Slabbekoorn, H., Bouton, N., 2008. Soundscape orientation: a new field in need of sound investigation. *Animal Behaviour* 76, e5–e8. <https://doi.org/10.1016/j.anbehav.2008.06.010>
- Stuart, S.N., Chanson, J.S., Cox, N.A., Young, B.E., Rodrigues, A.S.L., Fischman, D.L., Waller, R.W., 2004. Status and Trends of Amphibian Declines and Extinctions Worldwide. *Science* 306, 1783–1786. <https://doi.org/10.1126/science.1103538>
- Sueur, J., 2002. Cicada acoustic communication: potential sound partitioning in a multispecies community from Mexico (Hemiptera: Cicadomorpha: Cicadidae). *Biological Journal of the Linnean Society* 75, 379–394. <https://doi.org/10.1046/j.1095-8312.2002.00030.x>
- Sueur, J., Aubin, T., Simonis, C., 2008. Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics* 18, 213–226.
- Sueur, J., Farina, A., 2015. Ecoacoustics: the Ecological Investigation and Interpretation of Environmental Sound. *Biosemiotics* 8, 493–502. <https://doi.org/10.1007/s12304-015-9248-x>
- Sunday, J.M., Bates, A.E., Kearney, M.R., Colwell, R.K., Dulvy, N.K., Longino, J.T., Huey, R.B., 2014. Thermal-safety margins and the necessity of thermoregulatory behavior across latitude and elevation. *Proceedings of the National Academy of Sciences* 111, 5610–5615. <https://doi.org/10.1073/pnas.1316145111>
- Swanson, E.M., Tekmen, S.M., Bee, M.A., 2007. Do female frogs exploit inadvertent social information to locate breeding aggregations? *Canadian Journal of Zoology* 85, 921–932. <https://doi.org/10.1139/Z07-074>
- Tobias, J.A., Aben, J., Brumfield, R.T., Derryberry, E.P., Halfwerk, W., Slabbekoorn, H., Seddon, N., 2010. Song divergence by sensory drive in amazonian birds: sensory drive in birdsong evolution. *Evolution* 2820–2839. <https://doi.org/10.1111/j.1558-5646.2010.01067.x>
- Tobias, J.A., Planqué, R., Cram, D.L., Seddon, N., 2014. Species interactions and the structure of complex communication networks. *PNAS* 111, 1020–1025. <https://doi.org/10.1073/pnas.1314337111>
- Wells, K.D., 1977. The social behaviour of anuran amphibians. *Animal Behaviour* 25, 666–693. [https://doi.org/10.1016/0003-3472\(77\)90118-X](https://doi.org/10.1016/0003-3472(77)90118-X)

CHAPTER 3

Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis

Juan Sebastian Ulloa

Muséum national d'Histoire naturelle, France

Université Paris-Sud, France

Thierry Aubin

Université Paris-Sud, France

Diego Llusia

Muséum national d'Histoire naturelle, France

Université Paris-Sud, France

Charles Bouveyron

Université Paris Descartes & Sorbonne Paris Cité, France

Jérôme Sueur

Muséum national d'Histoire naturelle, France

OVERVIEW

Ecoacoustic monitoring has proved to be a viable approach to capture ecological data related to animal communities. While experts can manually annotate audio samples, the analysis of large datasets can be significantly facilitated by automatic pattern recognition methods. Unsupervised learning methods, which do not require labelled data, are particularly well suited to analyse poorly documented habitats, such as tropical environments. Here we propose a new method, named Multiresolution Analysis of Acoustic Diversity (MAAD), to automate the detection of relevant structure in audio data. MAAD was designed to decompose the acoustic community into few elementary components (soundtypes) based on their time-frequency attributes. First, we used the short-time Fourier transform to detect regions of interest (ROIs) in the time-frequency domain. Then, we characterised these ROIs by (1) estimating the median frequency and (2) by running a 2D wavelet analysis at multiple scales and angles. Finally, we grouped the ROIs using a model-based subspace clustering technique so that ROIs were automatically annotated and clustered into soundtypes. To test the performance of the automatic method, we applied MAAD to two distinct tropical environments in French Guiana, a lowland high rainforest and a rock savanna, and we compared manual and automatic annotations using the adjusted Rand index. The similarity between the manual and automated partitions was high and consistent, indicating that the clusters found are intelligible and can be used for further analysis. Moreover, the weight of the features estimated by the clustering process revealed important information about the structure of the acoustic communities. In particular, the median frequency had the strongest effect on modelling the clusters and on classification performance, suggesting a role in community organisation. The number of clusters found in MAAD can be regarded as an estimation of the soundtypes richness in a given environment. MAAD is a comprehensive and promising method to automatically analyse passive acoustic recordings. Combining MAAD and manual analysis would maximally exploit the strengths of both human reasoning and computer algorithms. Thereby, the composition of the acoustic community could be estimated accurately, quickly and at large scale.

"It was indeed a strange situation, to find no silence in the solitude of the woods. In the inns of Spain we dread the sharp sounds of guitars from the next apartment; in those of the Oroonoko, which are an open beach, or the shelter of a solitary tree, we are afraid of being disturbed in our sleep by voices issuing from the forest."

– Alexander von Humboldt (1825)

1. Introduction

The diversity of life forms is an invaluable biological resource threatened by anthropogenic environmental change (Pimm et al., 1995; Thomas et al., 2004). Given the pace of this change, there is an imperative need to develop quantitative indicators that provide specific information on the state of biodiversity (Pereira et al., 2013). With the advent of new sensor technology it is possible to remotely collect environmental data, assisting to determine, and eventually buffer, the pressures on biological diversity and ecosystem services (Petrou et al., 2015). In particular, the use of passive acoustic sensors in ecological research, or ecoacoustics (Sueur and Farina, 2015), has proved to be a viable method for biodiversity assessment that can be scaled up at multiple spatial and temporal scales (Towsey et al., 2014). The environmental sounds collected by these automated sensors usually include a large combination of both biotic and abiotic sounds, which are mixed down into a single time series. Such interlaced audio data needs to be unravelled in order to extract and to decipher ecological meaningful information, which represents to date a prominent bottleneck for the application of acoustic sensors in biodiversity monitoring.

A significant proportion of animal species produce sounds for social interaction, navigation or predator-prey encounters (Fletcher, 2014). Most of these acoustic signals have a species-specific signature that can be exploited for the remote identification of

species. The use of these signatures is a direct way to retrieve ecological data about species presence, abundance, status and distribution. Manual species identification by experts can be carried on audio datasets, but for large collections, the analysis can be facilitated by automatic pattern recognition methods such as supervised learning (Kershenbaum et al., 2016). Supervised learning is a method to build a statistical classifier based on labelled training data (Webb and Copsey, 2011). An increasing number of supervised learning tools have been adapted to identify automatically single species (Dugan et al., 2013; Ganchev et al., 2015; Ulloa et al., 2016) or several species (Briggs et al., 2012; Potamitis, 2014; Heinicke et al., 2015; Dong et al., 2015; Xie et al., 2016; Ruiz-Muñoz et al., 2016). The application of supervised learning is limited by the large reference datasets required to ‘train’ the classifiers and the high acoustic similarity sometimes observed between closely related taxa. The available sound libraries, even if providing thousands of samples, still cover only a small fraction of the animal sound diversity, at both population and species scales.

An alternative to species identification consists in characterising the acoustic community or the soundscape with the use of acoustic indices (Sueur et al., 2014). Rather than focusing on target species, acoustic indices aim to describe the global structure of the soundscape. A variety of indices have been proposed and applied to terrestrial (Lellouch et al., 2014; Farina et al., 2015; Fuller et al., 2015) and underwater habitats (Parks et al., 2014; Desjonquères et al., 2015; Harris et al., 2016; Buscaino et al., 2016). These indices revealed, for example, changes in bird species richness among woodland habitats (Depraetere et al., 2012) or dynamics of the soundscape across different temporal scales (Rodriguez et al., 2014). However, they also showed to be sensitive to transitory or permanent background noise, variation in the distance of the animals to the sensor, and the relative sound amplitude or the calling rate of the signalling animal (Gasc et al., 2015; Kendrick et al., 2016).

More recently, methods based on unsupervised learning have been adapted to audio recordings achieved in natural environments. Unsupervised learning searches for structures or patterns in a dataset without using labels. This approach has been extensively used to draw inferences in areas where labelled data is inaccessible or too expensive, such as astronomy (Way, 2012), genetics and genomics (Libbrecht and Noble, 2015). In an innovative work, Eldridge et al., (2016) adapted sparse-coding and source separation algorithms to extract shift-invariant spectro-temporal “atoms” from

environmental recordings. However, the authors did not establish a clear link between the spectro-temporal “atoms” and ecological or biological processes. Unsupervised learning has also been used as a pre-processing step for the classification task, significantly improving the classification performance on species recognition (Stowell and Plumbley 2014). In their approach, Stowell and Plumbley (2014) first decomposed the sounds into “atoms” with spherical k-means, and then used the “atoms” as features for the supervised learning framework. Thus, unsupervised learning offers new means to characterise sounds and may provide insights on the acoustic communities of diverse and threatened ecosystems, such as those of tropical regions (Pekin et al., 2012; Rodriguez et al., 2014).

The present work emerges from the question: how to best measure, quantify and characterise environmental sounds (from biotic and abiotic sources) in passive acoustic recordings to get valuable ecological indicators? We propose a new data-driven method, named Multiresolution Analysis of Acoustic Diversity (MAAD), to automate the discovery of plausible and interpretable patterns in passive acoustic recordings. To build a generalized method for multiple conditions and environments, we adapted methods from the unsupervised learning field. We estimated acoustic diversity by detecting regions of interest in sound recordings and grouping them into soundtypes based on the value of their time-frequency attributes at different scales. To test the flexibility and robustness of the method, we applied MAAD to two distinct night tropical environments in French Guiana, a lowland high rainforest (HF) and a rock savanna (RS). The RS is inhabited by a distinct and likely less diverse animal community in comparison with the HF (Bongers et al., 2001) so that it was expected to find contrasting acoustic communities between these two tropical environments. We compared manual and automated annotations to (1) evaluate the model selection procedure; (2) assess the relevance of different features in the clustering process; and (3) quantify the overall similarity between manual and MAAD soundtypes. To conclude, we give practical advices and discuss how MAAD can potentially be transferred to other environments in order to track the state and dynamics of animal communities for biodiversity studies.

2. Material and methods

The workflow of the proposed method (MAAD) followed four main steps: (1) passive acoustic recordings were transformed into the time-frequency domain using the windowed short-time Fourier transform and the Fourier coefficients were filtered to remove noise and to highlight sounds that can be delimited in time and frequency, here defined as regions of interest (ROIs); (2) each ROI was then characterised by features in the time-frequency domain using 2D wavelets; (3) the ROIs with their attributes were used to automatically estimate clustering hyper-parameters; and (4) the hyper-parameters and the attributes of the ROIs were passed to a clustering algorithm that formed homogenous groups of ROIs, namely soundtypes (Figure 1). This led to an automatic partitioning and characterization of soundtypes, which can be used to determine their presence, relative abundance and diversity within acoustic communities. To validate the proposed approach, the automatic partitioning provided by MAAD was compared to expert manual annotations using the adjusted Rand index (ARI).

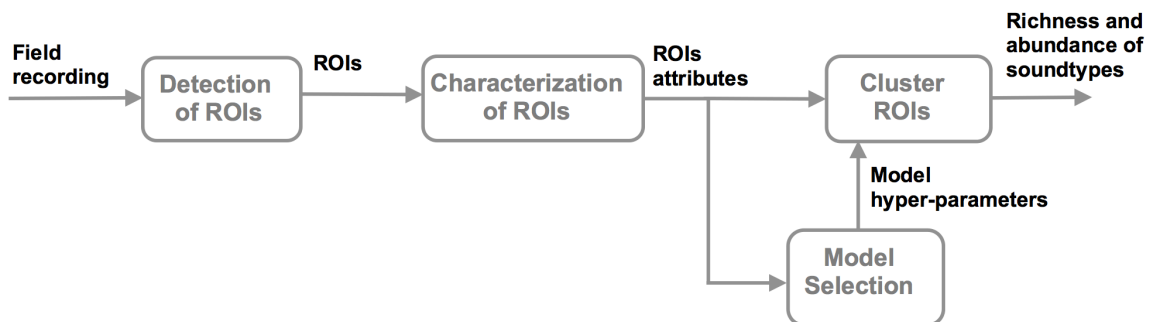


Figure 1: Block diagram of MAAD. Each step of the workflow is depicted as a grey box. Input and output after each step are indicated in black. Model selection is an optional step. Model hyper-parameters can also be set based on prior information about the acoustic community.

2.1. *Audio dataset*

Audio data were collected in French Guiana at the CNRS Nouragues Research Station (4°05'N; 54°40'W). The station is mainly occupied by a lowland high rainforest (HF) and a rock savanna (RS), among other ecosystems. The HF dominates on lower parts of the station (40-100 metres above sea level), has a fairly open understory and is closed on top by a dense canopy elevating at 25-35 m. The tree density in HF is high, with a total basal area ranging from 30 to 45 m²/ha, and the floristic composition is heterogeneous, since no species dominates the site. The RS is found on a granite hill about 400 m above sea level that rises abruptly and overtops the forest. Due to high microclimatic fluctuations and poorly developed soils, the RS is only partially colonised by vegetation, being its floristic composition drastically different from the surrounding forest. Small trees, xerophytic herbs and shrubs in scattered patches separated by rock areas covers nearly half of the area of the granite hill (Sarhou, 2001).

Environmental sounds were gathered using automated acoustic sensors (Songmeter SM2, Wildlife Acoustics Inc., Concord, MA, USA) equipped with omnidirectional microphones (PUI Audio POM-3535L-3-R, frequency response 50 Hz – 20 kHz ± 4 dB). A single acoustic sensor was placed at each environment, HF (04°05'15"N; 52°40'42"W) and RS (04°05'33"N; 52°40'40"W), and recorded one minute every 30 minutes from sunset to sunrise for 10 consecutive nights (5-15 December 2014). Each sensor was set to sample the audio at 44.1 kHz with a 16-bit resolution (mono, WAV format). This audio database was subsampled by selecting two one-minute samples per night, one four hours after sunset (22 h 17 min UTC/GMT -3 hours) and one four hours before sunrise (02 h 24 min UTC/GMT -3 hours). These environmental audio recordings were deposited at the sound library of the Muséum national d'Histoire naturelle (www.sonotheque.mnhn.fr, Table S2). At both sites, two files recorded during heavy rain were removed. The final audio database included 36 one-minute files in which 9403 ROIs were detected.

2.2. *Detection of regions of interest (ROIs)*

A region of interest is an isolated region in the time-frequency domain with a high density of energy. The automated detection of ROIs followed a four-step process

computed with MATLAB (The MathWorks, Inc., Natick, MA) using the signal processing toolbox. First, we computed a spectrogram of the audio signal using the windowed short-time Fourier transform, (1024 FFT length, 50% overlap, Hamming window). Second, we applied a denoising method, namely *spectral subtraction* (Boll, 1979; Yu et al., 2008), which allows to highlight transitory sounds by removing stationary noise found in the background. Third, we used a 2D rotationally symmetric Gaussian filter to remove small impulsive noise and to join close-by regions of high-density energy (5 by 5 element size, 0.5 standard deviation). As a final step, we applied a linear amplitude threshold to select the regions that were in the foreground. Since the spectrogram gives a sparse representation of the acoustic environment, regions of high density of energy can be identified as observations distant from the low-density background. Hence, the linear threshold (lth) was set for each recording by evaluating the dispersion of the spectrogram values and selecting values of the spectrogram distributed one-and-a-half times the inter-quartile range (IQR) above the third quartile ($lth = Q_3 + 1.5 \times IQR$). The use of quartiles gives a robust measure of central tendency and spread effective to non-normal data (Tukey, 1977).

Thereby, each detected ROI was a frame of variable size in the time-frequency domain, delimited by a start and end time, and a minimum and maximum frequency. The number of ROIs found in the RS and the HF audio files were respectively 4028 and 5375, for a total of 9403.

2.3. Characterization of ROIs

Automated measurements on the frequency and the time-frequency shape of each ROI were performed. To measure the frequency, a single feature was calculated: the median frequency, which is the value that divides the ROI into two frequency intervals of equal energy. This is a robust measurement that does not vary much based on the exact time-frequency bounds.

To measure the shape of the ROI in the time-frequency domain a wavelet analysis was used. The purpose of this procedure was to decompose the signal into coefficients that can be saved and manipulated to better represent the information in the signal. The wavelet transform is the result of filtering the signal with a bank of specific filters (or ‘wavelets’). Each analysing wavelet can be visualised as a kernel of fixed

scale that moves along the data. When the wavelet encounters a feature in the data with similar shape and scale, the analysis returns a high value for the wavelet coefficient. Then, the operation is repeated at a different scale with a new dilated or contracted wavelet. In this way the wavelet transform allows a multiresolution analysis and can represent hierarchical structures of the data. This scale-by-scale analysis is particularly suited for the detection of local features in aperiodic data. Wavelets can be extended to the two dimensional case (2D), in particular to process images (Mallat, 2009). In 2D, wavelets are dilated as in the one-dimensional analysis, and in addition rotated. A 2D wavelet transform of a spectrogram allows finding co-occurrence of time-frequency elements at different scales.

First, the high frequencies were recovered by convolution with the wavelet filters. By rotating and dilating the wavelet, we obtained rotation and scale covariant coefficients, which allowed discriminating the differences in shape of the different ROIs. Then, each filtered signal was averaged with a rotation-invariant low-pass filter. The rotation-invariant low-pass filter removed small differences between similar ROIs, forming homogeneous groups. The operation on a 2D signal x is formalised as:

$$Sx = (|x * \psi_{j,\theta}| * \phi)$$

where the symbol $*$ denotes spatial convolution, ϕ is a gaussian low-pass filter and $\psi_{j,\theta}$ is a wavelet dilated by 2^j and rotated by an angle θ . The filter bank used consisted of wavelets of the Morlet family, at 16 scales and 8 angles: horizontal (0°), vertical (90°) and diagonals (22.5° , 45° , 67.5° , 112.5° , 135° , 157.5°). In this way, a total of 128 shape features were calculated. An illustrative subset of the 2D filters is presented in Figure 2. The filter bank and the coefficients were computed with MATLAB (The MathWorks, Inc., Natick, MA) using the ScatNet toolbox (Sifre and Mallat, 2013).

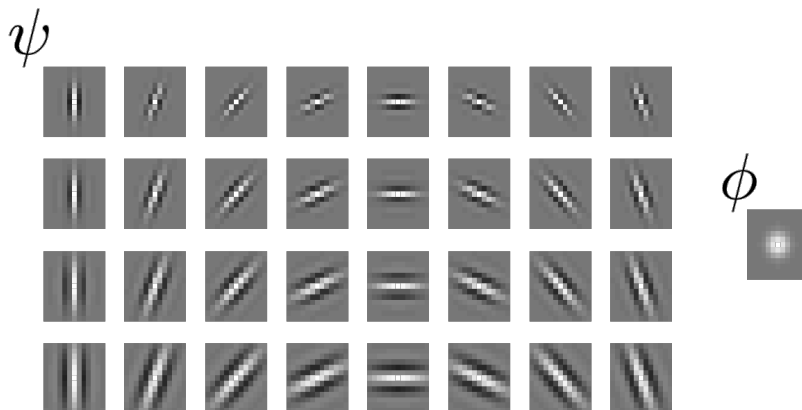


Figure 2. Subset of the 2D wavelet filter bank used to capture spectro-temporal features of the signal. On the left, Morlet wavelets ψ at four scales (along rows) and eight angles (along columns) are illustrated. On the right, the gaussian low-pass filter ϕ is represented.

2.4. High dimensional clustering

Clustering is an unsupervised learning analysis that aims at grouping objects into homogenous groups or clusters. As opposed to supervised learning, clustering is more flexible since no groups need to be defined a priori, *i.e.* the groups are formed based on the value of the attributes of the data. If available, labelled data can be used to estimate whether the groups found are suitable classes. To group the ROIs in homogeneous groups, a method suited to the multidimensional attributes of the ROIs was used. This method, named *High Dimensional Data Clustering* (HDDC), is a clustering technique based on a family of twelve parsimonious Gaussian mixture models adapted to multivariate high-dimensional data (Bouveyron et al., 2007). The mixture model-based clustering (on which HDDC is based) is defined in a probabilistic framework and has two particular advantages: (1) it is known to be a robust approach to deal with unbalanced datasets and (2) it is interpretable from a statistical point of view (Fraley and Raftery, 2002). The mixture model is naturally robust to unbalanced data sets because of the parameter π_k , which correspond to the weight of the group component in the mixture (see Equation 1, Text S6). The additional advantage of the mixture model is that it is a comprehensible statistical model and therefore allows to use model selection techniques, such as the Slope heuristics which we use later in the proposed framework.

The models proposed in HDDC have different regularizations that control the complexity of the clustering. The most complex model is $a_{kj}b_kQ_kd_k$, all the parameters

are class-specific and the dimension is specific to each cluster. The simplest model is abQkd, all the parameters are common between classes and the dimension of the class subspace is common. The properties of the parsimonious models in HDDC are detailed in Text S6.

A model selection procedure was implemented to estimate the hyper-parameters that control the complexity of the model. These hyper-parameters are the model M , the number of groups K , and the threshold value th to find the intrinsic dimensionality of each class. Classical model selection methods, namely AIC (Akaike, 1974) and BIC (Schwarz, 1978) criteria, are asymptotic (they assume that n tends to infinity) and therefore might not be appropriate. More recently, Birgé and Massart (2006) proposed a data-driven technique that alleviates this assumption and was used in different situations (Baudry et al., 2011), including model-based clustering (Bouveyron et al., 2015). The method, called *slope heuristics* (SHC), of the model M is defined as follows:

$$SHC(M) = l(\hat{\theta}) - 2\hat{s}\xi(M)$$

where θ is the set of parameter values that maximize the log-likelihood function $l(\hat{\theta})$, $\xi(M)$ is the number of free parameters of the model, and \hat{s} is the slope of the linear part of $l(\hat{\theta})$ with regard to the number of parameters. SHC follows the same rationale than other model selection criteria such as BIC and AIC, the likelihood of the fitted model is penalised by a function. Yet, SHC criterion has been found to be more consistent than BIC for model selection in HDDC (Bouveyron et al., 2015). A detailed overview and practical implementation advice of the SHC can be found in Baudry et al., (2011).

Slope heuristics were calculated for the twelve models implemented in HDDC (see Text S6), at ten different thresholds (0.0001, 0.0005, 0.001, 0.01, 0.03, 0.05, 0.07, 0.1, 0.15, 0.2), for 39 values of K (from 2 to 40, by steps of one). Since HDDC has a random initialization, the returned log-likelihood can vary between executions. Hence, the slope heuristics value was calculated ten times for each combination of hyper-parameters. The mean value was stored and the maximum was selected to find the hyper-parameters of the HDDC models.

With the hyper-parameters fixed, the model was fitted ten times with random initialisation. Random initialisation is a standard method to initiate the Expectation-maximization algorithm. This method correctly explores the parameter space to reach

the global maximum of the likelihood (Biernacki et al., 2003). Among the ten trials, only the model with the highest likelihood was selected for feature importance analysis and validation. Feature importance was calculated by multiplying the vector of estimated variances by the corresponding orthogonal matrix of orientations Q_k on each cluster. The total weight of the features is the average of the feature importance on all clusters. HDDC and slope heuristics were both computed using the R package HDclassif (Bergé et al., 2012).

2.5. Validation of system performance

To evaluate and determine the performance of MAAD, proper ground truth was established and a quantitative method to compare ground truth and the system output was used. ROIs ($n = 9403$) of sound recordings, which were automatically detected, were examined manually using the software Raven (Bioacoustic Research Program 2014). Aural and visual inspection of spectrograms, plus manual measurements on the temporal (duration and pulse rate) and spectral (median frequency and bandwidth) domain were made. Based on this combined examination, ROIs were categorised into distinct homogeneous groups, here referred as soundtypes. If the amplitude of the sound was too low and the features could not be inspected correctly, the ROI was marked as undetermined.

The automatic annotations output by MAAD were compared with the manual annotations using the adjusted Rand index (ARI). The ARI is a similarity measure between two partitions (Hubert and Arabie, 1985). Given two partitions, U and V, derived from a set of n objects, the ARI is computed according to:

$$ARI = \frac{\binom{n}{2}(a + d) - [(a + b)(a + c) + (c + d)(b + d)]}{\binom{n}{2} - [(a + b)(a + c) + (c + d)(b + d)]}$$

where a denotes objects in a pair placed in the same group in U and in the same group in V; b denotes objects in a pair placed in the same group in U and in different groups in V; c denotes objects in a pair placed in the same group in V and in different groups in U and; and d denotes objects in a pair placed in different groups in U and in different

groups in V . This index, bounded between ± 1 , was derived from the popular Rand index but has the advantage of being adjusted for chance with respect to the null hypothesis and can be interpreted as the difference between probabilities of concordance and discordance. Independently of the number of clusters and samples, the ARI has a value of -1 when the partitions are opposed, close to 0 for random labelling, and exactly 1 when the two partitions are identical.

Clustering analyses, cluster validation and graphs were achieved with R version 3.2.0 (R Core Team, 2015).

3. Results

We manually identified 35 soundtypes in the HF and 18 in the RS dataset. The relative soundtype abundance was unbalanced in both datasets (Figure S1). On average manual annotation required 25 to 35 minutes per file. Manual annotations were used only for performance validation purposes, that is, to interpret the return of MAAD at different settings. Two main tests were performed. The first one consisted in changing the hyperparameter K of the model, from 2 to 40 by unitary steps. The second one consisted in using different subsets of features: diagonal wavelets (16 scales \times 6 = 96 features), horizontal and vertical wavelets (16 scales \times 2 = 32 features), shape (32 + 96 = 128 features), median frequency (1 feature), and the full set (128 + 1 = 129 features).

3.1. Model selection

To begin with the cluster analysis, the most adequate model hyper-parameters were identified by observing the trend of the slope heuristics criterion. On both datasets, RS and HF, slope heuristics attained a maximum value with the model $a_{kj}b_kQ_kd_k$, the most complex one (a full covariance matrix for each group), and a threshold value of 0.0005. As expected, the suitable number of clusters K was different for each habitat. The curve showing the evolution of the slope heuristics value for different K peaked between 10 and 15 with a maximum at 11 on the RS dataset and peaked between 15 and 20 with a maximum at 17 on the HF dataset (Figure 3).

3. UNSUPERVISED MULTIREOLUTION ANALYSIS

Using the manual annotations, different settings of the hyper-parameter K (*i.e.* the number of clusters) were tested to analyse the response of MAAD. The response at different values of K was similar in both datasets (Figure 4), the performance of the clustering increments at the beginning and after reaching a peak, the performance begins to drop progressively. The peak value differs for the two habitats, 9 for the RS and 15 for the HF.

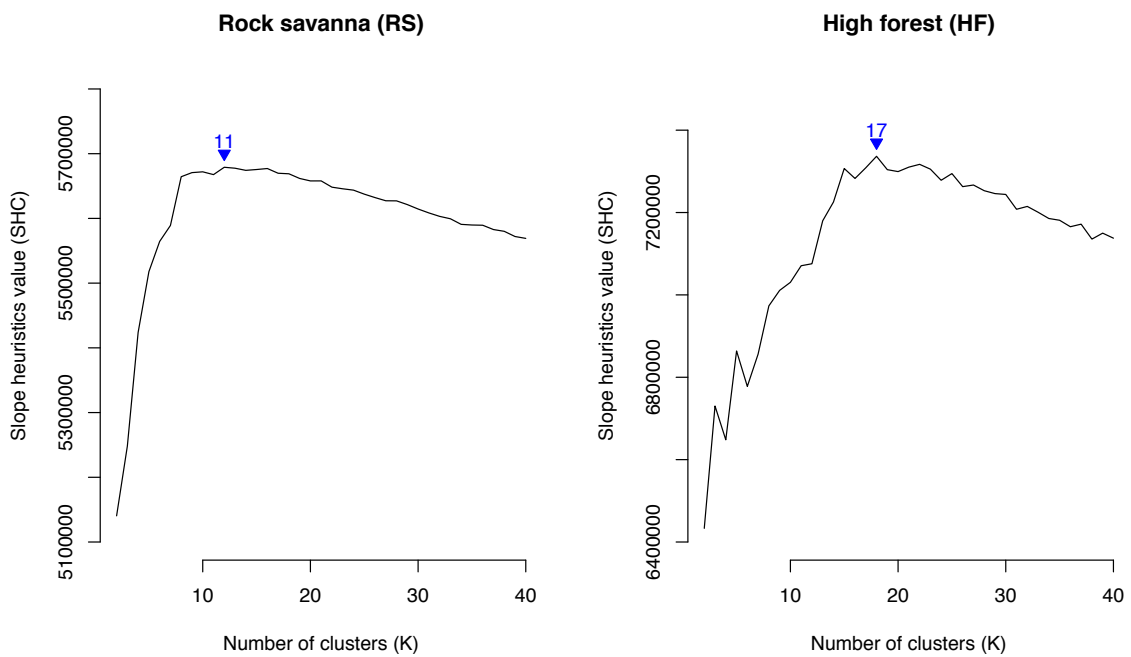


Figure 3. Model selection using slope heuristics. Variation of the slope heuristics criterion with respect to the number of clusters (K) for the RS and HF datasets. Slope heuristics find its maximum for RS at 11 clusters, and for HF at 17 clusters. This maximum is found for RS and HF with the same mixture model $(a_{kj}b_kQ_kd_k)$ and threshold value (0.0005).

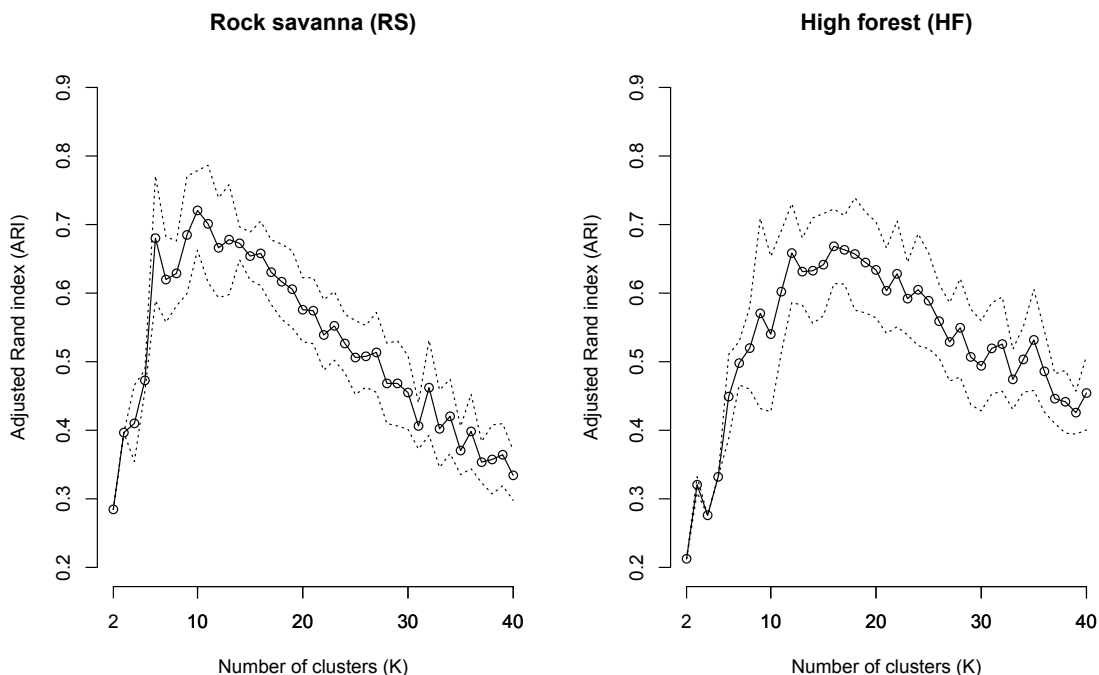


Figure 4. Variation of the performance of MAAD with respect to different values of the hyper-parameter K , number of clusters. The performance of MAAD is measured with the adjusted Rand index (ARI). The returned ARI value was calculated for 10 random initialisations of the clustering. The solid line represents the mean value and the dashed lines indicate the standard deviation of the results.

3.2. Feature relevance

Using the hyper-parameters found with the slope heuristics, the ROIs were automatically clustered based on their computed time-frequency attributes (129 features). Before evaluating the clustering results, the weight of the features estimated by the clustering process were analysed. Interestingly, a single feature, the median frequency, accounted for most of the variation in the data, 39.4 % and 51.0 % for the RS and HF respectively (Figure 5). The rest of the variation was associated to the combined wavelet features ($n = 128$) related to the time-frequency shape of the sound. The relative importance of each of the 128 wavelet features was plotted on an intensity map, a graphical representation of a matrix where each cell is highlighted according to its value. The intensity map showed that in both habitats the same two orientations explained best the data variance, the vertical (90°) and horizontal (0°) orientations (Figure 5). However, different scales are emphasized in each habitat, medium and large scales in the RS, and small scales in the HF. Wavelet features at diagonal angles (22.5° ,

3. UNSUPERVISED MULTIREOLUTION ANALYSIS

45°, 67.5°, 112.5°, 135° and 157.5°), which are related in the time-frequency domain to upsweeps and downsweeps, explained the residual data variance.

In light of the weights of the features estimated during the clustering process, different subsets of features were tested (diagonal wavelet orientations, perpendicular wavelet orientations, all wavelets, median frequency and the full set) and contrasted with the manual annotations to further examine the response of MAAD. The global return on both datasets, RS and HF, was the same (Figure 6). The model including only the diagonal wavelets (22.5°, 45°, 67.5°, 112.5°, 135° and 157.5°) gave the lowest ARI value. A higher ARI was obtained when using the horizontal and vertical wavelets (0° and 90°). By including all the wavelet features the result was improved again. By using only the median frequency of the ROIs, the results were even better than using all the 128 wavelet features. Finally, the best result (ARI value of 0.77 for RS and 0.85 for HF) was obtained by combining all the features, shape and median frequency.

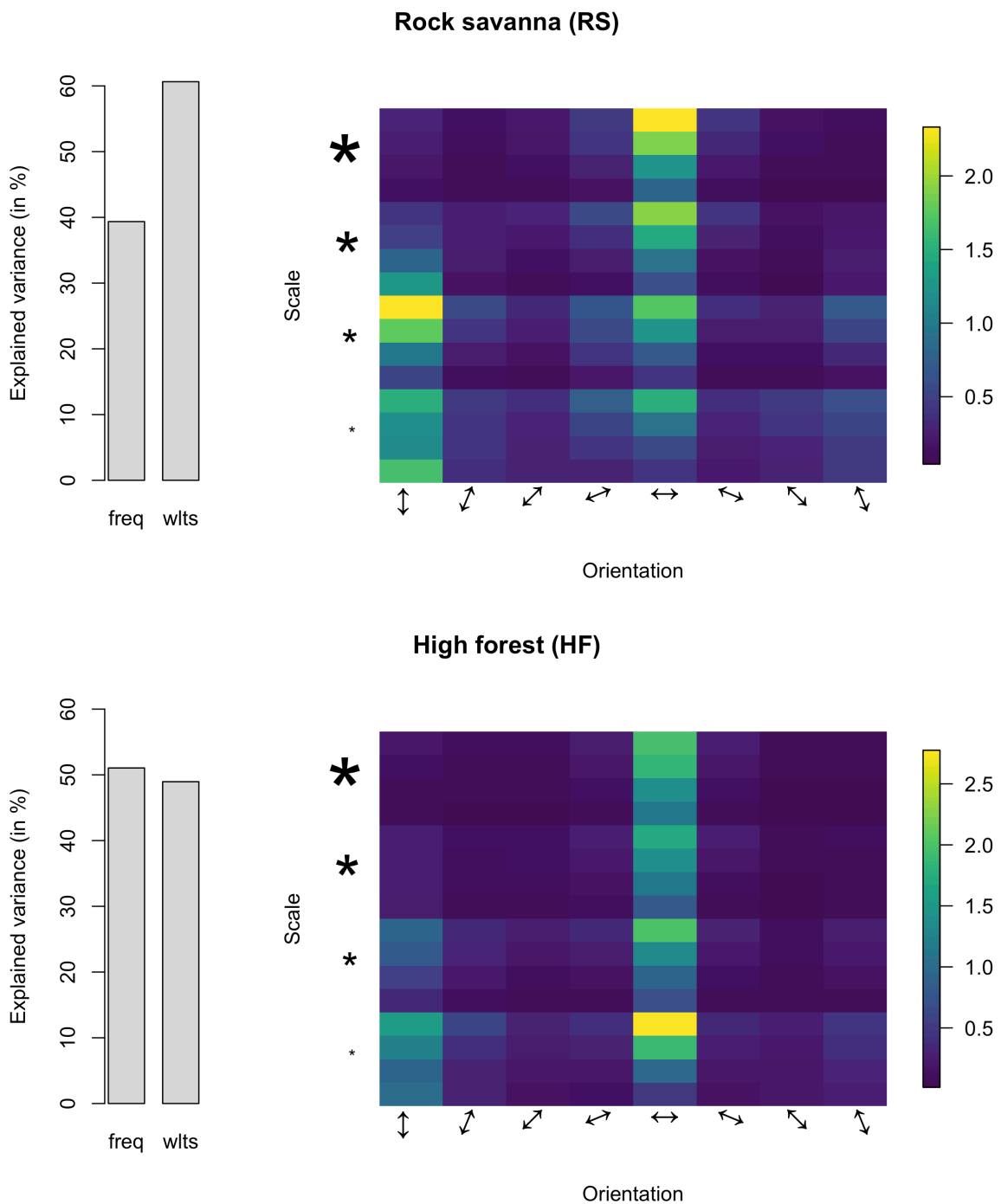


Figure 5. Representation of the amount of variance accounted for each of the 129 features used on the clustering process. The bar diagram (left) compares the median frequency (freq) and the sum of the 128 wavelets features (wlts). The intensity map (right) compares the relative importance of wavelets features at different angles and sizes, with dark blue indicating lowest value, and bright yellow the highest value.

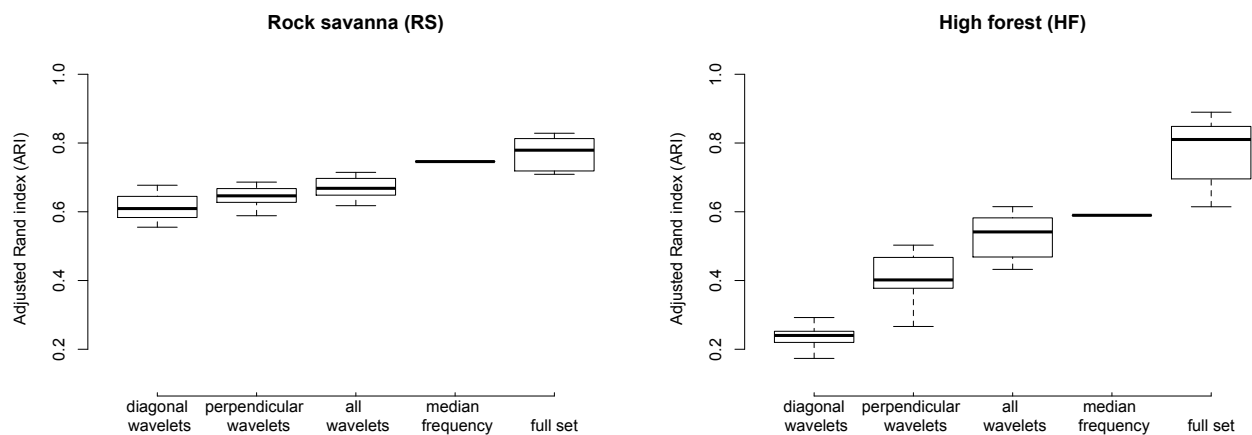


Figure 6. Response of MAAD using different feature sets: time frequency shape described by diagonal wavelets (96 non-perpendicular features), perpendicular wavelets (32 features), all wavelets (96 + 32 = 128 features), the median frequency (1 feature), and the full set of features (129 features). The performance was measured with the ARI metric computed over 10 trials. All but one feature set, the median frequency, had random initialisation. There is no box for the median frequency because univariate clustering had deterministic initialisation.

3.3. Clustering results

ROIs were grouped into soundtypes through an unsupervised framework using the hyper-parameters returned by the slope heuristics criterion and providing the full set of spectro-temporal features. Comparative analysis showed a high concordance between manual and automatic partitions with an ARI of 0.77 and 0.85 for the RS and the HF environments respectively (Figure 7). In general, the random initialisation of the clustering algorithm induced a relatively small variation on the result (s.d. < 0.13), compared to the possible variation of the ARI index (from -1 to +1). Detailed analysis by soundtype showed that most of the errors were due to clusters splitting (Table S4). A visual example of the final output is depicted on Figure S3.

The average computing time to process a one-minute file through the complete pipeline was 45.67 seconds on a desktop computer (3.4 GHz Intel Core i5 processor, 8 GB memory). Automatic annotation was on average forty times faster than human annotation.

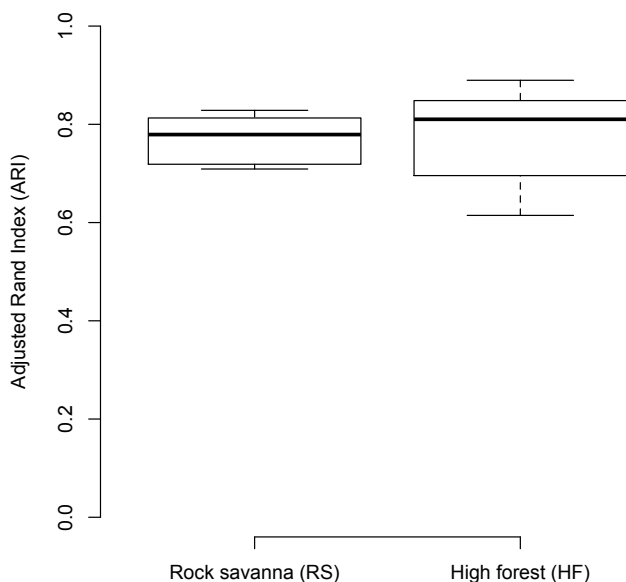


Figure 7. Global classification performance of MAAD for RS and HF datasets measured with the ARI metric computed over 10 trials with random initialisation. The ARI is bounded between ± 1 , has a value close to 0 for random labelling and exactly 1 when the two partitions are identical.

4. Discussion

The animal acoustic diversity is known to potentially carry relevant ecological information related to the species diversity (Riede, 1993; Krause and Farina, 2016). However, it is still challenging to use automated statistical tools to analyse and extract ecological meaningful information from passive acoustic recordings. MAAD was designed to overcome this barrier enabling to analyse environmental audio recordings by automatically decomposing the acoustic community into few elementary components based on their time-frequency attributes. Our experiments showed that the partitions derived by MAAD in distinct tropical acoustic communities were highly similar to the ones obtained by meticulous manual (aural and visual) inspection. In addition, MAAD showed that some specific features were more informative for the clustering model, revealing potential structures that partition the acoustic community.

4.1. Model selection

The number of soundtypes in an assemblage (*i.e.* the acoustic richness) is a common measure of the acoustic diversity. Slope heuristics indicated that the most appropriate model for decomposing the HF dataset had to include more clusters ($K = 17$) than the RS dataset ($K = 11$). A higher hyper-parameter K represents higher acoustic diversity in the HF, which is a result that matches our expectations and manual annotations. However, more soundtypes were found manually than automatically. In a closer look, we observed that common soundtypes were clustered correctly (e.g. A_1 , A_5 on the RS, and A_3 , A_5 on the HF), while rare soundtypes with less than 20 samples were not identified (e.g. A_{14} , A_{16} on the RS, and A_{24} , A_{33} on the HF). Slope heuristics makes a balance between the likelihood and the complexity of the model. As rare soundtypes are represented by a small number of samples (less than 20 samples), they do not increment the likelihood considerably to represent new clusters; instead, they are absorbed by larger clusters. Therefore, the number of clusters found in MAAD has to be regarded as the richness of common soundtypes in a given environment. In other words, soundtypes with infrequent presence in the recorded time series are expected to have low likelihood to be detected. As in many other sampling techniques in ecology, rare and elusive species are difficult to detect.

It is also important to note the resemblance between the slope heuristics trend and the response of the system with respect to increment of the hyper-parameter K , the number of clusters. In particular, the value of K selected by slope heuristics (11 and 17 for RS and HF respectively) is close to the value of K with the highest ARI value (9 and 15 for RS and HF respectively). Slope heuristics allows finding automatically a plausible number of clusters in relation to clustering performance, meaning that this criterion seems to be a suitable alternative to the human supervision.

4.2. Feature relevance

Generative modelling, such as HDDC, builds a full model of the distribution of features in each group. These models can be analysed to understand what group properties are the most important for clustering the objects. In our framework, the weight of the features estimated by the clustering process revealed important information about the

structure of the acoustic community. The median frequency had the strongest effect on modelling the clusters. In other words, frequency predicted soundtype identity better than all the shape features. Partitioning the transmission channel in different frequency bands appears to be a common strategy to avoid masking, although other mechanisms may also generate the same pattern. Our results are congruent with frequency partitioning, which has been previously observed on assemblages of crickets (Schmidt et al., 2013), cicadas (Sueur, 2002) and amphibians (Villanueva-Rivera, 2014). Frequency over dispersion allows a great number of co-occurring signals to be accommodated in a limited acoustic space. Formulated under the acoustic niche hypothesis, organisms would have evolved to occupy specific spectro-temporal 'niches', decreasing the risk of heterospecific mating and information masking (Krause, 1993). Alternatively, many other selective pressures might be responsible for signal divergence and acoustic partitioning, such as those related to body size or female preferences (Gerhardt, 1994).

The acoustic space has multiple dimensions and the frequency is just one of them. Other dimensions, such as the shape features, had a lower but significant impact. The shape features derived with the 2D wavelets were also important features to derive the clusters. In particular, vertical and horizontal wavelets (0° and 90°) had a significant effect on the clustering process. These features are based on the spectrogram representation of the signal; therefore, most of the sounds were clustered based on variations in the duration of the sound and variations in the frequency bandwidth. Diagonal wavelets had less importance in the model learned. This outcome was expected since insects and amphibians, which dominated the studied acoustic communities, are known to produce sounds with few frequency modulations (Gerhardt and Huber, 2002).

4.3. Clustering results

Signalling animals produce redundant and species-specific sounds, which result in intuitive clusters. Based on this observation, MAAD was designed to give a representation based on a combination of elementary components (soundtypes) to form a whole (the acoustic community). To our knowledge, only Eldridge et al., (2016) attempted a similar approach to characterise the acoustic communities or soundscape.

Both approaches, Eldridge et al (2016) and ours, are based on unsupervised learning techniques, however, the aim and the evaluation of the result differ significantly. Eldridge et al (2016) measured the ability to reconstruct a soundscape based on few spectro-temporal 'atoms' as a way to measure the 'decomposability' of a scene, and the evaluation was completely visual. In contrast, we aimed at finding ecologically plausible and interpretable 'atoms' or soundtypes. We evaluated our approach by comparing manual versus automated partitioning by using an objective measure of similarity, the ARI. Unfortunately, the differences in the objectives and on the evaluation procedure make our work hard to be confronted to this previous work.

MAAD was tested under two contrasting scenarios and gave robust clustering results, with high and consistent similarity between manual and automated partitions. This suggests that the elementary time-frequency components found by MAAD are interpretable and that the output can be used in further analysis for studies in ecology and evolution. For instance, the number of items in each cluster corresponds to the relative abundance of each soundtype. This information can be used to compute diversity indices such as Shannon, Simpson or Whittaker indices (Magurran, 2004), returning an estimation of local acoustic diversity. Alternatively, after processing with MAAD, a manual inspection of detected soundtypes may enable to establish a direct link between MAAD clusters and species. For example, HF cluster number B_5 could be identified as stridulations of the cricket *Lernecella minuta*, and RS cluster number B_6 could be identified as vocalisations of the amphibian *Hypsiboas boans* (Figure S5). This semi-supervised framework would allow to annotate efficiently large sound databases for deeper analyses.

The clustering errors were mainly due to the division of major ground-truth clusters into homogeneous subgroups. The marked unbalanced nature of the dataset played an important role on this clustering subdivision. Clusters with many observations have a stronger weight maximising the overall likelihood of the model than clusters with rare observations. Since the parameters of the model were estimated so as to maximise the global likelihood of the model, the likelihood was incremented by splitting large clusters instead of creating new small clusters. Cluster splitting was also observed in the study of the response of the model to the variation of K . After reaching a peak, at a lower K than the true number of groups, the performance measured by the ARI dropped in both datasets. The ARI measures the number of ROIs correctly

partitioned and hence the performance measure was mainly impacted by the splitting of large clusters and less by the wrong clustering of small ones. This also explains why the clustering results were highly accurate even if the soundtype richness found by the unsupervised procedure was lower than that by the manual one. Interestingly, the division of clusters with large observations still resulted in homogeneous groups, which could be assessed and combined by manual inspection. Further research is necessary to evaluate the performance on other scenarios in order to validate this method across the diverse acoustic communities found in practice. These tests would also be valuable to assess the error propagation of the system, identifying the potential sources of error and exploring how they influence the results.

MAAD is an adaptable framework that can be coupled with expert knowledge. An advantage of model-based clustering, which is used by MAAD, is that the uncertainty for an observation to belong to a cluster is measured by a posterior probability. Observations with probabilities drifting from 1 could be subsequently flagged and assessed by an expert. Combining MAAD and manual analysis would maximally exploit the strengths of both human reasoning and computer algorithms. Thereby, the composition of the acoustic community could be estimated accurately, quickly and at large scale.

Acknowledgements

This research was supported by the Labex CEBA (Centre d'Étude de la Biodiversité Amazonienne) and the SABIOD MASTODONS Big Data (CNRS MI project 2012–2017). We would like to warmly thank Philippe Gaucher, Elodie Courtois and Patrick Châtelet for their help during our stay at the Nouragues research station. We specially thank Julio Pedraza who provided valuable advices on the use of high performance computing clusters to run batch process. We also would like to thank Laure Desutter-Grandcolas and Elodie Courtois for their help in identifying clusters to a species-specific acoustic signal. DLL was supported by the Fyssen Fondation (Post-doctoral Grant). JSU was supported by COLCIENCIAS (Doctoral Scholarship of the Colombian government). The authors declare no conflict of interest. We finally wish to acknowledge the valuable help of two anonymous referees for their remarks and comments on the manuscript.

Data accessibility

The environmental audio recordings were deposited at the sound library of the Muséum national d'Histoire naturelle (www.sonothèque.mnhn.fr). The collection number of each file is presented on Table S2. Source codes (Matlab and R) are available at: https://github.com/juansulloa/maad_matlab. A step by step instruction to run the analysis is provided in Text S7.

References

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Baudry, J.-P., Maugis, C., Michel, B., 2011. Slope heuristics: overview and implementation. *Stat Comput* 22, 455–470. <https://doi.org/10.1007/s11222-011-9236-1>
- Bergé, L., Bouveyron, C., Girard, S., 2012. HDclassif: An R Package for Model-Based Clustering and Discriminant Analysis of High-Dimensional Data. *Journal of Statistical Software* 46. <https://doi.org/10.18637/jss.v046.i06>
- Biernacki, C., Celeux, G., Govaert, G., 2003. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis, Recent Developments in Mixture Model* 41, 561–575. [https://doi.org/10.1016/S0167-9473\(02\)00163-9](https://doi.org/10.1016/S0167-9473(02)00163-9)
- Bioacoustic Research Program, 2014. Raven Pro: Interactive Sound Analysis Software. The Cornell Lab of Ornithology, Ithaca, NY.
- Birgé, L., Massart, P., 2006. Minimal Penalties for Gaussian Model Selection. *Probab. Theory Relat. Fields* 138, 33–73. <https://doi.org/10.1007/s00440-006-0011-8>
- Boll, S.F., 1979. Suppression of acoustic noise in speech using spectral subtraction. *Acoustics, Speech and Signal Processing, IEEE Transactions on* 27, 113–120.
- Bongers, F., Charles-Dominique, P., Forget, P.M., Théry, M., 2001. *Nouragues: dynamics and plant-animal interactions in a neotropical rainforest*. Kluwer Academic Publishers, Dordrecht Boston.
- Bouveyron, C., Côme, E., Jacques, J., 2015. The discriminative functional mixture model for a comparative analysis of bike sharing systems. *The Annals of Applied Statistics* 9, 1726–1760.
- Bouveyron, C., Girard, S., Schmid, C., 2007. High-dimensional data clustering. *Computational Statistics & Data Analysis* 52, 502–519. <https://doi.org/10.1016/j.csda.2007.02.009>
- Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X.Z., Raich, R., Hadley, S.J., Hadley, A.S., Betts, M.G., 2012. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *J. Acoust. Soc. Am.* 131, 4640–4650.

- Buscaino, G., Ceraulo, M., Pieretti, N., Corrias, V., Farina, A., Filiciotto, F., Maccarrone, V., Grammauta, R., Caruso, F., Giuseppe, A., Mazzola, S., 2016. Temporal patterns in the soundscape of the shallow waters of a Mediterranean marine protected area. *Scientific Reports* 6, 34230. <https://doi.org/10.1038/srep34230>
- Depraetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S., Sueur, J., 2012. Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland. *Ecological Indicators* 13, 46–54.
- Desjonquères, C., Rybak, F., Depraetere, M., Gasc, A., Le Viol, I., Pavoine, S., Sueur, J., 2015. First description of underwater acoustic diversity in three temperate ponds. *PeerJ* 3, e1393. <https://doi.org/10.7717/peerj.1393>
- Dong, X., Towsey, M., Truskinger, A., Cottman-Fields, M., Zhang, J., Roe, P., 2015. Similarity-based birdcall retrieval from environmental audio. *Ecological Informatics* 29, Part 1, 66–76. <https://doi.org/10.1016/j.ecoinf.2015.07.007>
- Dugan, P., Pourhomayoun, M., Shiu, Y., Paradis, R., Rice, A., Clark, C., 2013. Using High Performance Computing to Explore Large Complex Bioacoustic Soundscapes: Case Study for Right Whale Acoustics. *Procedia Computer Science* 20, 156–162. <http://dx.doi.org/10.1016/j.procs.2013.09.254>
- Eldridge, A., Casey, M., Moscoso, P., Peck, M., 2016. A new method for ecoacoustics? Toward the extraction and evaluation of ecologically-meaningful soundscape components using sparse coding methods. *PeerJ* 4, e2108. <https://doi.org/10.7717/peerj.2108>
- Farina, A., Ceraulo, M., Bobryk, C., Pieretti, N., Quinci, E., Lattanzi, E., 2015. Spatial and temporal variation of bird dawn chorus and successive acoustic morning activity in a Mediterranean landscape. *Bioacoustics* 24, 269–288. <https://doi.org/10.1080/09524622.2015.1070282>
- Fletcher, N.H., 2014. Animal Bioacoustics, in: Rossing, T.D. (Ed.), *Springer Handbook of Acoustics*. Springer New York, pp. 821–841. https://doi.org/10.1007/978-1-4939-0755-7_19
- Fraley, C., Raftery, A.E., 2002. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association* 97, 611–631.

- Fuller, S., Axel, A.C., Tucker, D., Gage, S.H., 2015. Connecting soundscape to landscape: Which acoustic index best describes landscape configuration? *Ecological Indicators* 58, 207–215. <https://doi.org/10.1016/j.ecolind.2015.05.057>
- Ganchev, T.D., Jahn, O., Marques, M.I., de Figueiredo, J.M., Schuchmann, K.-L., 2015. Automated acoustic detection of *Vanellus chilensis lampronotus*. *Expert Systems with Applications* 42, 6098–6111. <https://doi.org/10.1016/j.eswa.2015.03.036>
- Gasc, A., Pavoine, S., Lellouch, L., Grandcolas, P., Sueur, J., 2015. Acoustic indices for biodiversity assessments: Analyses of bias based on simulated bird assemblages and recommendations for field surveys. *Biological Conservation* 191, 306–312. <https://doi.org/10.1016/j.biocon.2015.06.018>
- Gerhardt, H.C., 1994. The Evolution of Vocalization in Frogs and Toads. *Annual Review of Ecology and Systematics* 25, 293–324. <https://doi.org/10.1146/annurev.es.25.110194.001453>
- Gerhardt, H.C., Huber, F., 2002. *Acoustic communication in insects and anurans: common problems and diverse solutions*. University of Chicago Press, Chicago.
- Heinicke, S., Kalan, A.K., Wagner, O.J.J., Mundry, R., Lukashevich, H., Kühl, H.S., 2015. Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods in Ecology and Evolution* 6, 753–763. <https://doi.org/10.1111/2041-210X.12384>
- Hubert, L., Arabie, P., 1985. Comparing partitions. *Journal of classification* 2, 193–218.
- Humboldt, A. von, 1825. *Personal narrative of travels to the equinoctial regions of the new Continent*. Longmans Hurst.
- Kershenbaum, A., Blumstein, D.T., Roch, M.A., Akçay, Ç., Backus, G., Bee, M.A., Bohn, K., Cao, Y., Carter, G., Căsar, C., Coen, M., DeRuiter, S.L., Doyle, L., Edelman, S., Ferrer-i-Cancho, R., Freeberg, T.M., Garland, E.C., Gustison, M., Harley, H.E., Huetz, C., Hughes, M., Hyland Bruno, J., Ilany, A., Jin, D.Z., Johnson, M., Ju, C., Karnowski, J., Lohr, B., Manser, M.B., McCowan, B., Mercado, E., Narins, P.M., Piel, A., Rice, M., Salmi, R., Sasahara, K., Sayigh, L., Shiu, Y., Taylor, C., Vallejo, E.E., Waller, S., Zamora-Gutierrez, V., 2016. Acoustic sequences in non-human animals: a tutorial review and prospectus: Acoustic sequences in animals. *Biological Reviews* 91, 13–52. <https://doi.org/10.1111/brv.12160>

- Krause, B.L., 1993. The niche hypothesis: a virtual symphony of animal sounds, the origins of musical expression and the health of habitats. *The Soundscape Newsletter* 6, 4–6.
- Lellouch, L., Pavoine, S., Jiguet, F., Glotin, H., Sueur, J., 2014. Monitoring temporal change of bird communities with dissimilarity acoustic indices. *Methods in Ecology and Evolution* 5, 495–505. <https://doi.org/10.1111/2041-210X.12178>
- Libbrecht, M.W., Noble, W.S., 2015. Machine learning applications in genetics and genomics. *Nat Rev Genet* 16, 321–332. <https://doi.org/10.1038/nrg3920>
- Magurran, A.E., 2004. *Measuring biological diversity*. Blackwell Pub, Malden, Ma.
- Parks, S.E., Miksis-Olds, J.L., Denes, S.L., 2014. Assessing marine ecosystem acoustic diversity across ocean basins. *Ecological Informatics, Ecological Acoustics* 21, 81–88. <https://doi.org/10.1016/j.ecoinf.2013.11.003>
- Pekin, B.K., Jung, J., Villanueva-Rivera, L.J., Pijanowski, B.C., Ahumada, J.A., 2012. Modeling acoustic diversity using soundscape recordings and LIDAR-derived metrics of vertical forest structure in a neotropical rainforest. *Landscape Ecology* 27, 1513–1522. <https://doi.org/10.1007/s10980-012-9806-4>
- Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Petteorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential Biodiversity Variables. *Science* 339, 277–278. <https://doi.org/10.1126/science.1229931>
- Petrou, Z.I., Manakos, I., Stathaki, T., 2015. Remote sensing for biodiversity monitoring: a review of methods for biodiversity indicator extraction and assessment of progress towards international targets. *Biodivers Conserv* 24, 2333–2363. <https://doi.org/10.1007/s10531-015-0947-z>
- Pimm, S.L., Russell, G.J., Gittleman, J.L., Brooks, T.M., 1995. The Future of Biodiversity. *Science* 269, 347–350. <https://doi.org/10.1126/science.269.5222.347>
- Potamitis, I., 2014. Automatic classification of a taxon-rich community recorded in the wild. *PLoS ONE* 9, e96936. <https://doi.org/10.1371/journal.pone.0096936>
- R Core Team, 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Riede, K., 1993. Monitoring Biodiversity: Analysis of Amazonian Rainforest Sounds. *Ambio* 22, 546–548.
- Rodriguez, A., Gasc, A., Pavoine, S., Grandcolas, P., Gaucher, P., Sueur, J., 2014. Temporal and spatial variability of animal sound within a neotropical forest. *Ecological Informatics, Ecological Acoustics* 21, 133–143. <https://doi.org/10.1016/j.ecoinf.2013.12.006>
- Ruiz-Muñoz, J.F., Castellanos-Dominguez, G., Orozco-Alzate, M., 2016. Enhancing the dissimilarity-based classification of birdsong recordings. *Ecological Informatics* 33, 75–84. <https://doi.org/10.1016/j.ecoinf.2016.04.001>
- Sarthou, C., 2001. Plant Communities on a granitic outcrop, in: *Nouragues: Dynamics and Plant-Animal Interactions in a Neotropical Rainforest*. Kluwer Academic Publishers, Dordrecht Boston, pp. 64–77.
- Schmidt, A.K.D., Römer, H., Riede, K., 2013. Spectral niche segregation and community organization in a tropical cricket assemblage. *Behavioral Ecology* 24, 470–480. <https://doi.org/10.1093/beheco/ars187>
- Schwarz, G., 1978. Estimating the Dimension of a Model. *The Annals of Statistics* 6, 461–464. <https://doi.org/10.1214/aos/1176344136>
- Stowell, D., Plumbley, M.D., 2014. Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. *PeerJ* 2, e488. <https://doi.org/10.7717/peerj.488>
- Sueur, J., 2002. Cicada acoustic communication: potential sound partitioning in a multispecies community from Mexico (Hemiptera: Cicadomorpha: Cicadidae). *Biological Journal of the Linnean Society* 75, 379–394. <https://doi.org/10.1046/j.1095-8312.2002.00030.x>
- Sueur, J., Farina, A., 2015. Ecoacoustics: the Ecological Investigation and Interpretation of Environmental Sound. *Biosemiotics* 8, 493–502. <https://doi.org/10.1007/s12304-015-9248-x>
- Sueur, J., Farina, A., Gasc, A., Pieretti, N., Pavoine, S., 2014. Acoustic Indices for Biodiversity Assessment and Landscape Investigation. *Acta Acustica united with Acustica* 100, 772–781. <https://doi.org/10.3813/AAA.918757>
- Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A.,

3. UNSUPERVISED MULTIREOLUTION ANALYSIS

- Townsend Peterson, A., Phillips, O.L., Williams, S.E., 2004. Extinction risk from climate change. *Nature* 427, 145–148. <https://doi.org/10.1038/nature02121>
- Towsey, M., Parsons, S., Sueur, J., 2014. Ecology and acoustics at a large scale. *Ecological Informatics, Ecological Acoustics* 21, 1–3. <https://doi.org/10.1016/j.ecoinf.2014.02.002>
- Tukey, J.W., 1977. *Exploratory data analysis*, Addison-Wesley series in behavioral science. Addison-Wesley Pub. Co, Reading, Mass.
- Villanueva-Rivera, L.J., 2014. Eleutherodactylus frogs show frequency but no temporal partitioning: implications for the acoustic niche hypothesis. *PeerJ* 2, e496. <https://doi.org/10.7717/peerj.496>

CHAPTER 4

Towards an end-to-end framework for sound classification in passive acoustic recordings

Juan Sebastian Ulloa

Muséum national d'Histoire naturelle, France

Université Paris-Sud, France

Thierry Aubin

Université Paris-Sud, France

Diego Llusia

Universidad Autónoma de Madrid, Spain

Jérôme Sueur

Muséum national d'Histoire naturelle, France

OVERVIEW

Presence-absence data is a common format to analyse populations and communities in ecology. Currently, the translation from audio recordings to species presence-absence is largely done manually, but can be assisted by automated classification. The number of available tools for pattern recognition are rapidly increasing, yet there is still little practical guidance for the application of such algorithms for a broad audience. Moreover, most statistical classifiers require representative datasets that are not readily available in poorly documented habitats. We addressed the question: how to best estimate the presence-absence of target species from passive acoustic recordings? We designed experiments to (1) test two sampling strategies to effectively build a representative dataset to train and test automated classifiers, namely random and stratified sampling; and (2) compare three approaches to estimate presence-absence of target sound of interest, manual expert annotation, a semi-supervised and a supervised learning approach. The experiments were based on a dataset that included 528 one-minute files with heterogeneous sounds, collected in French Guiana. Our target sound of interest was the call of the tropical frog *Leptodactylus knudseni*. We demonstrate that stratified sampling, compared to naive random sampling, can be more effective to assemble a representative dataset in heterogeneous environments with unbalanced classes. Regarding presence-absence labelling, expert identifications were fairly similar, but presented non-negligible variation between observers. Semi-supervised learning required few manual effort but was the less accurate method. Supervised learning, while needing more manual annotations to estimate model parameters, showed results with a good consensus between the expert labelling. To conclude, we underline that solutions aimed at combining manual and automated analysis in a well framed annotation protocol should be envisaged to deliver robust and accurate estimates of presence-absence. Such framework is fundamental to assess poorly documented habitats, where calls are yet to be described, annotated and saved in sound libraries.

*“Far better an approximate answer to the right question,
which is often vague,
than an exact answer to the wrong question,
which can always be made precise”*
– John Wilder Tukey (1962)

1. Introduction

Passive acoustic monitoring is an emergent technique in ecology, evolution and behaviour research (Blumstein et al., 2011; Fristrup and Mennitt, 2012; Aide et al., 2013; Browning et al., 2017). Based on automated recording of animal sounds, passive acoustic monitoring is growingly being applied in both terrestrial and marine environments, providing cost-efficient methods for surveying biodiversity and opening new scientific pathways (Towsey et al., 2014; Sueur and Farina, 2015). Yet, scaling up acoustic sampling to big data, analysis of automated recording has proved challenging, and it represents nowadays the primary bottleneck that constraints the expansion of this technique and slow down its application (Eldridge et al., 2016; Stowell, 2018).

Facing the data flood coming from passive acoustic sensors, numerous tools for automated pattern recognition have been developed with applications in the ecoacoustic and bioacoustics fields. The combination of signal processing techniques and machine learning algorithms has opened the possibility to screen large audio datasets to detect and classify sounds as already illustrated for insects (Brandes et al., 2006), amphibians (Aide et al., 2013; Xie et al., 2017), birds (Briggs et al., 2009; Ulloa et al., 2016; Knight et al., 2017) or mammals (Heinicke et al., 2015; Keen et al., 2017). Hence, the ecoacoustic analysis toolbox now includes several options to deal with the monitoring and analysis of marine, freshwater and terrestrial environments (Stowell, 2018; Towsey

et al., 2012). Despite of these recent advances, there is still little practical guidance for the end-users, including researchers and practitioners, limiting significantly the use of these innovative methods in and out the academic system (Browning et al., 2017).

While a single software with a standard and universal classifier for all sounds might be desirable, the development of such computer facility have proved to be much more challenging than expected (Goëau et al., 2014; Stowell, 2018), and probably no single machine learning method can perform with a satisfactory efficiency on all possible problems (Wolpert and Macready, 1997). A more realistic scenario, achievable in the short term, would be to derive a generic and well-framed workflow that combines strengths of both human and computer algorithms. Supervised learning has a well-established framework with best practices and straightforward implementation offering a multiplicity of statistical classifiers (Theodoridis and Koutroumbas, 2006; Bishop, 2006; Webb and Copsey, 2011; Zaki and Meira, 2014). Many supervised learning algorithms are available in freeware software, such as R (Lantz, 2013) or Python (Pedregosa et al., 2011), already in the hands of ecologists.

The first step of supervised classification consists in selecting a reference dataset to train, tune and validate the subsequent automated classifier. The reference dataset plays a key role in the signal classification since it strongly influences the learning curve of the classifier, its performance and, ultimately its universality and degree of applicability. For the performance estimation phase, the classification estimates are computed in relation to the validation dataset and future prediction interpretation will likely be biased if the validation dataset is biased. Thereby, the extent and strength of the inferences drawn will vary depending on the sampling used for the training phase. Without methodical procedures to select the reference dataset, it would be unlikely to attain a robust and generic framework to analyse passive acoustic recordings that yields repeatable results.

The importance of having a rich reference dataset that incorporates the variations of the signal of interest in frequency, duration and strength, has been highlighted in several studies (Knight et al, 2017; Ranjard et al 2016; Priyadarshani et al., 2018). Unfortunately, only large scale monitoring programs dispose of important corpus of annotated datasets ready to be used (Dugan et al., 2013; Keen et al., 2017). In most studies, training and testing of automated classifiers is made on a selection among recordings freshly collected. To our knowledge, there is no clear procedure to select the

samples for the reference dataset. This selection should be achieved through a controlled sampling process aimed at reducing bias or error due to the under-representation or over-representation of observations, or sounds (Thompson, 1992). Sampling theory suggests that error can be reduced by (1) designing an appropriate sampling protocol, and (2) increasing the number of samples (Rubin and Babbie, 2008). The diverse sampling strategies developed to efficiently collect data avoiding biased estimates might provide methodological insights to frame an end-to-end generic and efficient framework for the analysis of ecoacoustic recordings.

Annotating field recordings is a demanding task and might require the involvement of an expert, as a result a very small proportion of data is labelled. On the other hand, passive acoustic recorders can easily collect large amounts of unlabelled data. To deal with these kind of situations, semi-supervised and unsupervised statistical classifiers have been developed in the machine learning field (Bishop., 2006; Theodoridis and Koutroumbas, 2006). Such set of tools were designed to exploit the information that resides in the unlabelled data and might circumvent the issue of having small manual annotations datasets (Chapelle et al., 2006, Ulloa et al., 2018).

Here, we propose a set of strategies combining human analysis and machine automation to estimate species presence-absence data from passive acoustic recordings. Such strategies were specially conceived to enhance acoustic analysis from biologically diverse and poorly documented habitats (*e.g.* tropical forests). We addressed on two main challenges: (1) how to build a dataset with representative samples to train and test automated classifiers? (2) How to use manual annotation, semi-supervised and supervised learning to succeed in the classification task? And, (3) how to combine them to derive efficient and robust estimates of the presence and absence of sounds of interest? To answer these questions, we performed experimental tests using field audio recordings collected by autonomous recorders during the rainy season in French Guiana. We focused on the challenge of detecting the presence-absence of the calling males of Amazonian toad-frog, *Leptodactylus knudseni*, in a heterogeneous acoustic environment composed by more than 30 other anuran species. To conclude, we propose best practices for sampling and discuss on the strengths and challenges of the different options, shedding light on how to best combine human reasoning and computer algorithms on analysing passive acoustic recordings.

2. Methods

A series of tests were designed to study the effect of two methodological procedures in the performance of signal classifiers: (1) sampling strategy for selecting training and validating datasets and (2) presence-absence labelling of field recordings. For the sampling strategy, we tested two widely used sampling methods, random sampling and stratified sampling, and examined the learning curves of the statistical classifiers (Figure 1). The classification on this first experiment was evaluated at the level of acoustic signals delimited in time-frequency, or regions of interest (ROIs). Next, we compared three approaches to estimate the presence-absence of the advertisement calls of a tropical frog (*Leptodactylus knudseni*): manual expert annotation, a semi-supervised approach and a supervised learning approach (Figure 2). For this second experiment, the computational analysis was done at the level of ROIs, but the classification (presence-absence) was evaluated at the level of a one-minute audio recording. For both set of tests, the raw field recordings were processed on a generic framework to find, characterise, and cluster ROIs.

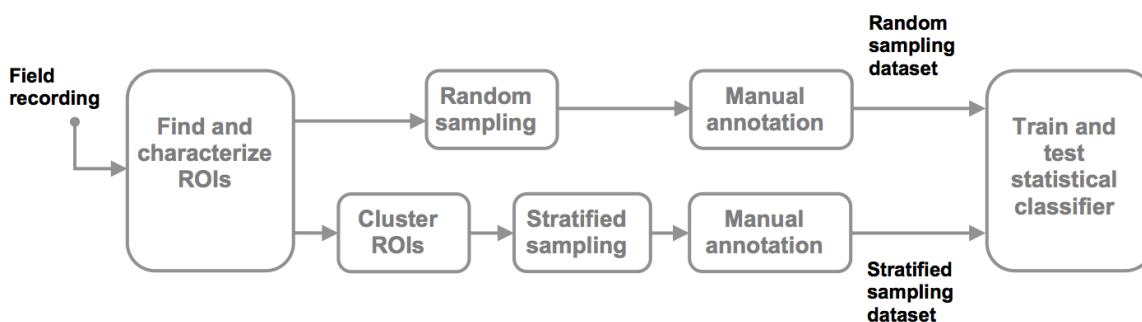


Figure 1. Flow diagram of sampling test. From the same audio recordings, two distinct datasets are derived: random and stratified sampling. Both datasets consist of regions of interest (ROIs) on the time-frequency domain, but the selection of observations follow different stages. Simple random sampling is composed by ROIs selected randomly from the full set of ROIs. Stratified sampling first structures the data into homogeneous groups and then samples the observations according to each stratum. Both collections of ROIs are manually annotated with presence-absence of the sound of interest. The ROIs, associated features and labels are used to train and test a statistical classifier.

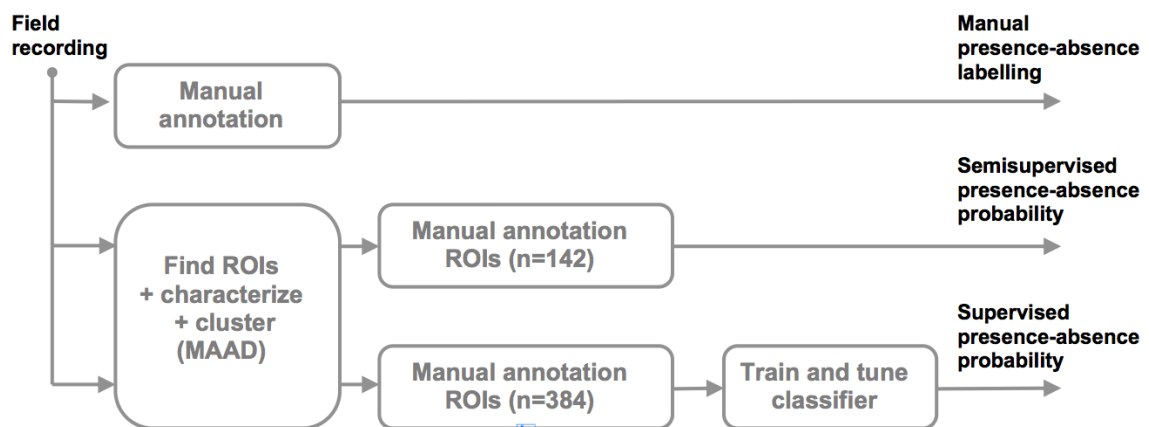


Figure 2. Flow diagram of presence-absence test. Field recordings are annotated based on the presence or absence of the call of *Leptodactylus knudseni*. Three different strategies are compared: manual labelling by three experts, semi-supervised learning and supervised learning. For expert presence-absence, the field recording is annotated with binary presence-absence. The computational analysis was done at the level of regions of interest (ROIs) delimited in the time-frequency, but the classification (presence-absence) was evaluated at the level of a one-minute audio recording. The semi-supervised flow follows a two stage process, first the ROIs are clustered into homogeneous groups and then, a sample of 32 ROIs is selected from every cluster ($n=192$) to assign a class to each cluster. Supervised learning probability is derived using a more extensive dataset of ROIs than semi-supervised learning ($n=384$) and has an additional step; the training and tuning of the classifier. The output of each process was compared using standard detection metrics, namely the area under the receiver operating curve.

2.1. Acoustic environment and signal of interest

A key element of the tropical acoustic environment arises from the vocalisations of amphibians. In particular, several frog and toad species exhibit a striking acoustic collective behaviour by emerging and calling in massive number in a limited area and during a few hours per year (Wells, 2007, 1977), a phenomenon called explosive breeding. During explosive breeding events it is possible to find hundreds of individuals of diverse species signalling acoustically, which results in signals with variable overlapping levels, from clear isolated calls to intense multi-specific choruses (Figure 3). A common species that aggregates around water bodies and vocalizes before and during these events is the tropical anuran *Leptodactylus knudseni*. The advertisement call of *L. knudseni* consists of a low frequency upswing, from 0.2 to 0.6 kHz that lasts approximately 0.4 s. At a finer temporal scale, the call present multiple frequency and amplitude modulations that give a particular vibrato timbre to the call (Figure 3, top

left). Such fine texture disappears when multiple signalling individuals aggregate forming choruses with overlapping signals (Figure 3, top right). Due to the background with multiple species and the presence of this focal signal at various chorusing levels, automated detection result in a complex problem.

Our experimental protocol and data analysis is based on passive acoustic recordings collected at five temporary ponds located along a 35 km transect in the Kaw mountains of French Guiana (4°57'N; 52°22'W). In each temporary pond, automated acoustic sensors (Song Meter 2, Wildlife Acoustics Inc.) were set to record one minute every 29 minutes from 10 November 2015 to 07 March 2016, with a digitization depth of 16 bits and a 44.1 kHz sampling rate. A subset of recordings from eleven days was selected during anuran pre-explosive and explosive breeding periods. The complete audio dataset included 528 one-minute files. This dataset was divided in two equal parts, training and validation, using random sampling of recordings. The training dataset was used for the experiments related to the sampling strategy of selecting the reference dataset. The validation dataset was subsequently used to compare different methods to estimate the presence-absence of sounds of interest.

Spectrograms and characteristics of the call of *L. knudseni* were computed using R version 3.4.3 (R Core Team, 2017) and the package *seewave* version 2.0.4 (Sueur et al., 2008).

2.2. Analysis of acoustic diversity

To find, characterise and cluster sounds, Multiresolution Analysis of Acoustic Diversity (MAAD, Ulloa et al., 2018) was computed throughout the recordings. The analysis operates by first segmenting the audio recordings into time-frequency regions of interest (ROIs) that are later clustered into homogeneous sections. MAAD workflow follows four main steps: detection of ROIs, characterisation of ROIs, model selection and clustering of ROIs. Here, we adopted the exact same process for the last three steps (characterisation, model selection and clustering), yet we modified the first part (detection of ROIs). Instead of analysing the full range of signals of the acoustic environment, we focused the analysis on sound candidates that approximate the time-frequency parameters of our sound of interest, the call of *L. knudseni*. In particular we searched for sound candidates by (1) applying a kaiser-windowed frequency band-pass

filter between 0.2 and 0.7 kHz with a stop-band attenuation of 120 dB, and (2) detecting temporal patterns with pronounced energy lasting 0.4 s using a Mexican Hat wavelet with a size of 0.4 s.

The process detected a total of 11 954 ROIs in the training dataset and 12 789 in the validation dataset, that is a mean of 47 ROIs per recording. Acoustic diversity analyses were computed in Python 3.6.3 (Python Software Foundation, 2017). To illustrate the acoustic diversity, the ROIs were visualized with a non-metric dimensionality reduction technique called t-distributed Stochastic Neighbour Embedding (t-SNE; Maaten and Hinton, 2008). This dimensionality reduction method was designed to capture local structures and global structures of high dimensional data, thus revealing clusters at several scales.

4. TOWARDS AN END-TO-END FRAMEWORK

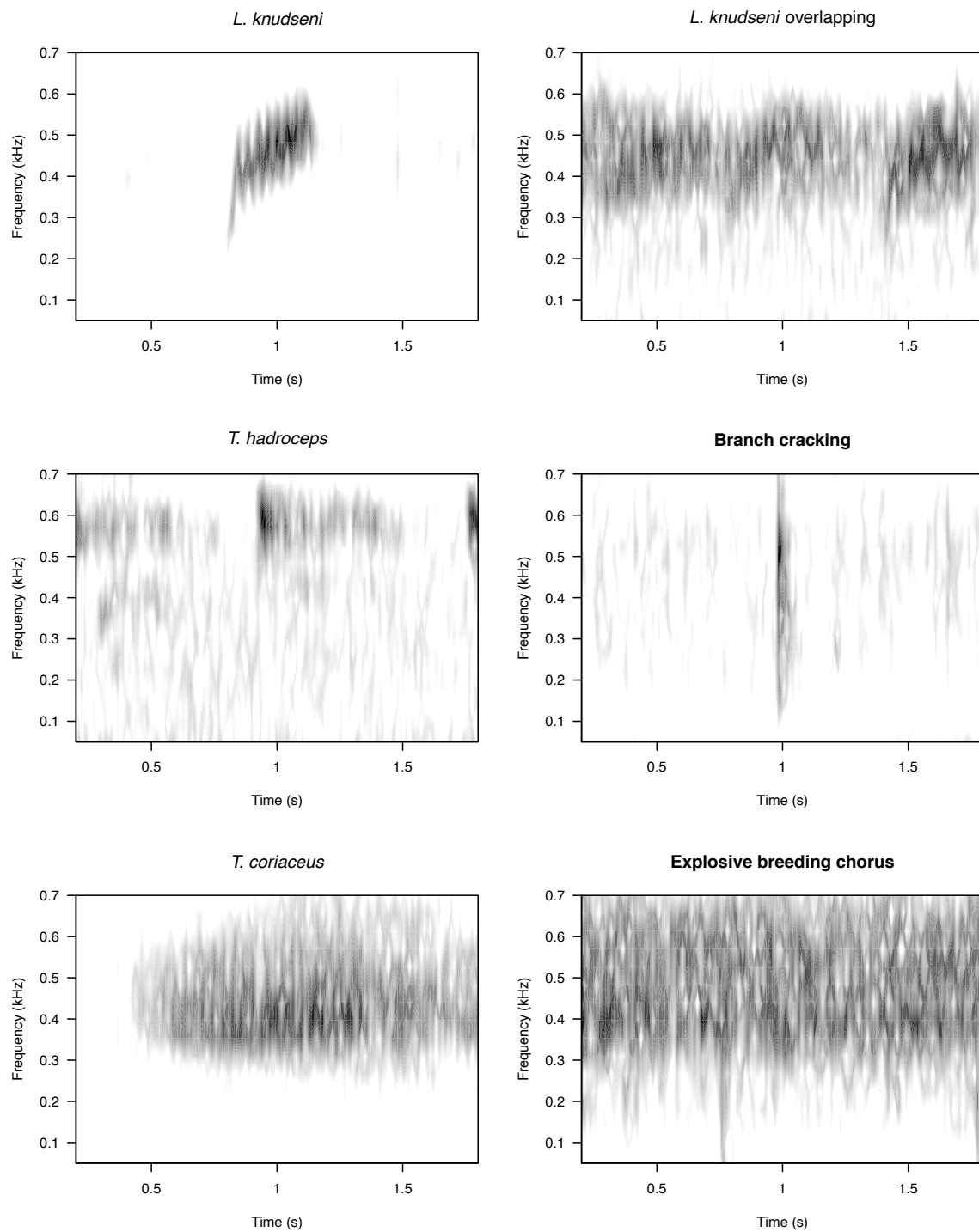


Figure 3. Spectrogram of illustrative ROIs detected automatically. Individual calls of *Leptodactylus knudseni*, calls of *Leptodactylus knudseni* with high overlap, calls of *Trachicephalus hadroceps*, sounds of branch cracking, calls of *Trachicephalus coriaceus*, and explosive breeding chorus.

2.3. Sampling strategies

The first test consisted in comparing different sampling strategies to select the ROIs that would compose the training dataset and be annotated by experts. Two sampling strategies were implemented: simple random sampling and stratified sampling. Simple random sampling consists in selecting observations with equal probability from the population without replacement. This random sampling is expected to provide unbiased estimates of the complete dataset. Inherent sampling error can be reduced by incrementing the sample size, or by implementing an alternative sampling design.

In scenarios with previously known and heterogeneous data, a widely used sampling strategy is stratified sampling (Thompson, 2012). Stratified sampling consists first in partitioning the data into homogeneous strata, including observations as similar as possible, and then performing a random sampling within each stratum. Here stratified sampling was applied on sound segments by considering the clusters returned by MAAD as strata. MAAD identified six clusters. We selected from each cluster a random sample of 64 ROIs (3.21 % of the total number of ROIs).

We selected a overall sample of 384 ROIs with each method. Once we build the reference dataset, we tested how a statistical classifier performed, by computing the scores from training and cross validation, when increasing from 5% to 100% the size of the training dataset. Training score refers to how well the model fits the observed data and cross-validation score specifies how well the model predicts the class of a new observation. A convergence between these two scores indicates that the classifier has the capacity to generalize to new data. The F-score metric was used to quantify the classification performance, which combines precision and recall into a single value (Van Rijsbergen, 1979; Hripcsak, 2005). We used a Gaussian naïve Bayes classifier to compare the sampling strategies. Naïve Bayes is a simple statistical classifier based on Bayesian networks (Frank et al., 2000). Simple naïve Bayes under Gaussian distribution requires no hyper-parameter tuning, which makes it an ideal candidate for such a test.

Naïve Bayes statistical classifiers and learning curves were computed in Python 3.6.3 using the toolbox scikit-learn version 0.19.1 (Pedregosa et al., 2011).

2.4. Presence-absence labelling

Three different classification methods (manual classification, semi-supervised, supervised learning classification) were tested to detect the presence-absence of the calls of *L. knudseni* in each recording of the validation dataset (n=264; Figure. 2). Manual annotation consisted in aural identification of the species by three experts who identified yet in a previous study (Chapter 2) every species heard in the present recordings. Semi-supervised learning consisted in a two-stage process that combined automated and manual processes. The automated process first detected regions of interest and arrange them into homogeneous regions (see previous subsection). Then, a random sample of 32 ROIs from each cluster was manually inspected to assign a '1' if the sound corresponded to the call of *L. knudseni* and a '0' if the sound could not be attributed to *L. knudseni*. Supervised learning approach consisted in sampling twice the number of ROIs and in the training and tuning of a supervised learning algorithm. The Random Forest statistical classifier was selected, as it slightly outperformed the basic Naive Bayes classifier in the preliminary tests. Models for both learning algorithms, semi-supervised learning and supervised learning, were built using a separate training dataset.

Performance metrics related to classification score were compared with the aim of identifying the strengths and limits of classification methods. To estimate the classification accuracy we computed the receiver operating characteristic curve (ROC) and the associated summary statistics area under the curve (AUC). ROC and AUC allow a classifier performance comparison independent of the threshold probability and have been acknowledge as suitable performance metrics for binary classification, even when the classes are unbalanced (Bradley, 1997). The ground truth is usually provided by human manual annotations when comparing automated classifiers. Since we had annotations from multiple experts, we calculated performance metrics against each of the expert annotations. A paired comparison was performed between manual annotations.

Random forest statistical classifier was computed with the RandomForest version 4.6 (Liaw and Wiener, 2002) package for R. All computational analyses were performed with a desktop computer (3.4 GHz Intel Core i5 processor, 8GB memory).

3. Results

3.1. Sampling strategies

MAAD partitioned the 11 954 ROIs into six different clusters. The size of the clusters varied from 72 ROIs (0.6 %) for the smallest cluster to 7 005 ROIs (58.6 %) for the biggest one. The 2D t-SNE projection of the ROIs showed heterogeneous regions, with one large homogeneous region and multiple small regions patchily distributed appearing as islands (Figure 4). This heterogeneity was clearly reflected by the two sampling strategies tested. Random sampling showed selected observations regularly distributed and equally covering the whole dataset (Figure 4, top). In contrast, stratified sampling showed that regions with more heterogeneity were sampled more intensely than the large uniform sector (Figure 4, bottom).

After manual inspection of the ROIs for both random and stratified sampling, we observed that the advertisement calls of *L. knudseni* were less numerous for the random sampling than for the stratified sampling, with proportions of respectively 14% and 39.1%. In particular, most of the samples on the random sampling dataset came from short and intense sounds produced by branch cracking and raindrops (40.6 %). These types of sounds were present in the stratified sample at lower proportions (18.52 %), similar to the proportions of clear *L. knudseni* vocalizations.

Statistical classifiers responded differently to the datasets collected with random and stratified sampling (Figure 5). The learning curves for the random sampling showed a wide gap between training and validation score. Once the size of the training dataset increases this gap narrowed slowly and only with more than 250 examples the classifier appears to converge between training and cross validation performance at a score of 0.68. The learning curves for the stratified sampling showed a smaller gap between training and cross-validation curves. The two curves remained stable with only 150 samples, converging at a score of 0.91.

4. TOWARDS AN END-TO-END FRAMEWORK

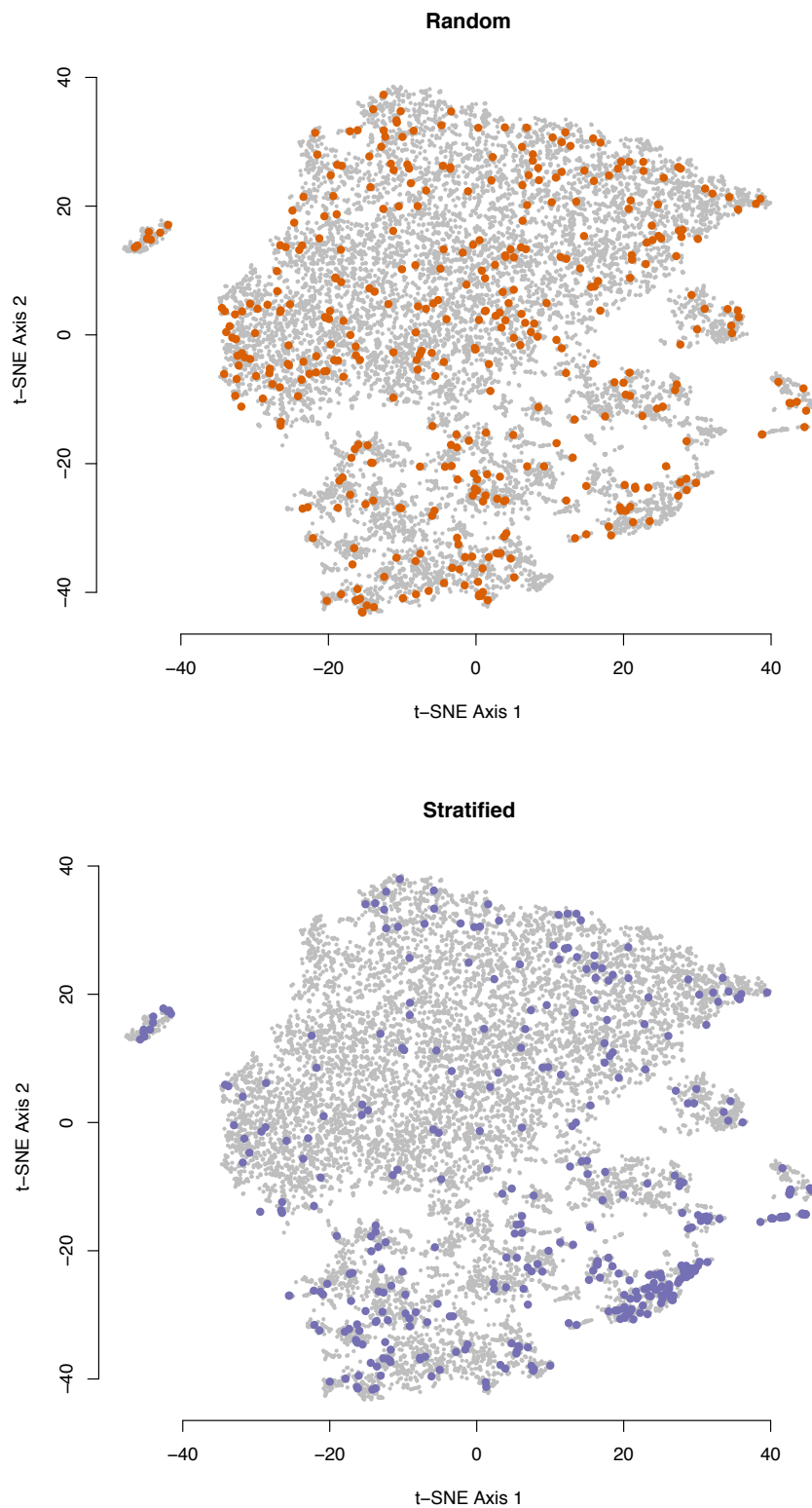


Figure 4. Non-metric multidimensional scaling (t-SNE) of the regions of interest showing random versus stratified sampling. Top, random sampling assumes homogeneity on all observations and hence the sampling is uniformly distributed. Bottom, stratified sampling first divide the observation into homogeneous groups and then samples randomly each subgroup; the sampling is less intense in homogeneous sectors than in heterogeneous sectors.

4. TOWARDS AN END-TO-END FRAMEWORK



Figure 5. Learning curves of a naïve Bayes classifier. The graph compares the performance of a model on training and cross-validation data over an increasing number of samples. Random sampling dataset is contrasted with the stratified sampling dataset. The F-score metric was used to quantify the classification performance.

3.2. Determining presence-absence data

The identification done by the experts showed a characteristic triangular shape due to the binary nature of the data. The three experts gave fairly similar classifications, but presented non-negligible variation between observers as shown by different ROC curves (Figure 6A). The AUC for the experts had a median value of 0.916 (Q1-Q3=0.062).

The semi-supervised approach showed no false positives at true positive rate of 0.4 (Figure 6B). While lowering the threshold to detect the rest of the occurrences, the false positive rate increased rapidly; at a true positive rate of 0.9, the false positive rate was higher than 0.4. Compared with the other approaches, semi-supervised learning showed the lowest AUC with a value of 0.85 (Q1-Q3=0.039).

Supervised learning showed a good consensus between the three expert annotations. This approach returned a true positive rate of 0.6 with a negligible false positive rate (Figure 4C). By lowering the threshold to capture more occurrences, we observed an increase on false positives that was much lower than for the supervised

4. TOWARDS AN END-TO-END FRAMEWORK

approach. At a true positive rate of 0.9, the false positive rate was lower than 0.2. The median AUC summary statistics for the supervised learning was 0.94 (Q1 – Q3 = 0.017).

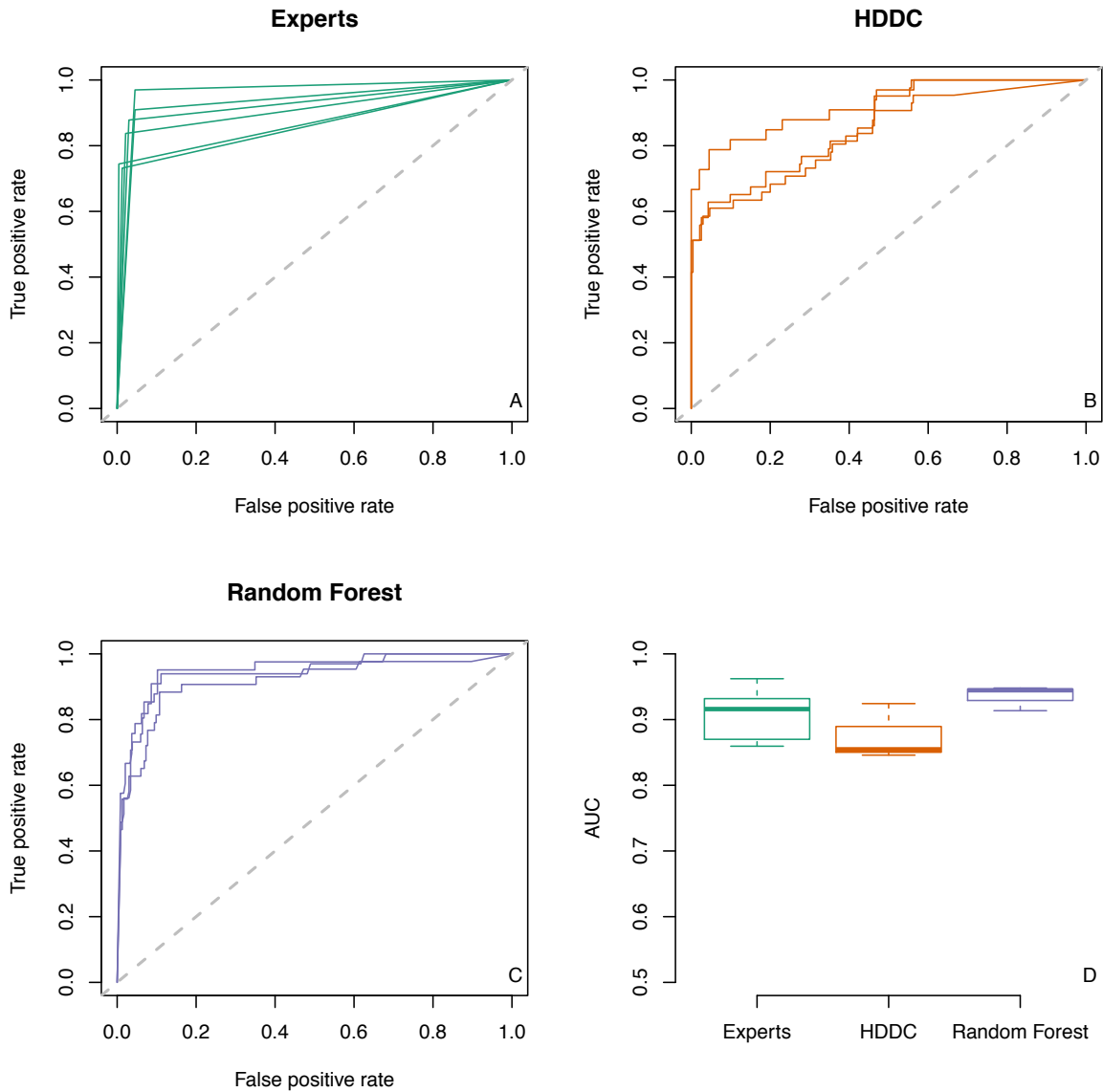


Figure 6. Receiver operating characteristic curve (ROC) on the detection of *Leptodactylus pentadactylus*. Three methods are contrasted: (A) expert manual annotation, (B) HDDC semi supervised learning, and (C) Random Forest supervised learning. A boxplot of the area under the ROC curve is presented as a univariate performance metric (D).

4. Discussion

Acoustic heterogeneity of tropical environments, coupled with a prominent lack of annotation datasets, calls for the development of innovative analytical frameworks. Manual analysis seems to be unscalable to large datasets and a generic classifier for every animal sound appears to be infeasible, in particular in highly diverse tropical acoustic environments. Therefore, it seems to be more suitable to frame a standard end-to-end protocol that incorporates both human expertise and computer algorithms. Heading for this purpose, we examined two key aspects of this analytical procedure: (1) the sampling methodology to select observations for tuning and training statistical classifiers, and (2) the protocol to annotate presence-absence of target signals in passive audio recordings.

4.1. Sampling strategies

Getting reliable statistical classifiers for automated detection can require large amounts of manual annotations, so an improvement in the classifier performance with a reduction in manual effort is desirable. The two strategies tested, random and stratified sampling, showed different responses in the implementation of a statistical classifier. Compared to random sampling, a statistical classifier trained with stratified sampling data showed learning curves converging earlier in their training and cross validation scores. Thereby, the statistical classifier based on stratified sampling was able to generalize to new observations with less than half of the samples.

A primary goal of predictive models is to train models that are able to forecast to unseen data (Bishop, 2006). For this purpose, the training data should be representative of the whole dataset, and particularly the sounds collected and annotated for training should have the similar characteristics than the rest of the unlabelled data. An homogeneous population produces samples with smaller sampling error than an heterogeneous one (Rubin and Babbie, 2008). Increasing the sampling is costly hence undesired, but it seems plausible to structure the data in order to have homogeneous subpopulation from which to sample later. Rather than selecting a sample from the total population at large, we ensure that appropriate number of elements is drawn from homogeneous subsets of that population. Stratified sampling guarantee the proper

representation of the stratification variables and this, in turn, enhances the representation of other variables related to them. Taken as a whole, a stratified sample will be more representative than a simple random sample in heterogeneous data. Stratified sampling is actually one of the most commonly used methods in social sciences (Rubin and Babbie, 2008) and ecology (Thompson, 2012). By clustering the sounds into homogeneous regions, unsupervised learning allowed to adopt a stratified sampling, which was more efficient than random sampling to train a statistical classifier.

As a simple and unbiased method, random sampling is well suited to validate classifier's performance. This type of selection has the advantage of removing human sources of bias, such as intentional or unintentional tendencies to select signals with particular spectro-temporal characteristics, for instance clear calls or ear-catching sounds. Moreover, the proportions of positive and negative observations of the whole dataset will be reflected in the random sample, giving performance estimates related to the natural distribution where the classifier will operate upon.

When screening for a target sound in long time-series of acoustic recordings, the problem can be illustrated by the popular saying "looking for a needle in a haystack". Positive observations are far less abundant than the negative observations, hence the task of obtaining representative samples for the call of interest require increasing proportionally the negative observations. Yet, with a structured dataset, a multiplicity of sampling techniques can be applied to obtain representative samples. In our case, unsupervised learning assisted in the constitution of a representative training dataset for supervised learning by forming homogeneous groups. An appropriate combination of unsupervised learning and sampling theory might lead to efficient, precise and robust statistical classifiers. Alternatively, the arrangement of sounds delivered by unsupervised learning combined with more complex sampling strategies, such as sequential or adaptive sampling, could be used to find and compose datasets of sounds of rare and elusive species (Thompson, 2004).

4.2. Determining presence-absence data

Determining the presence-absence of signals of interest is one of the most common aims in passive acoustic monitoring. Investigating multiple solutions to this challenge, we

tested three different strategies: expert manual analysis, semi-supervised learning and supervised learning. We used a set of 264 recordings to search for the presence of *Leptodactylus knudseni* in a heterogeneous acoustic environment.

Expert identifications were fairly similar, but presented non-negligible variation between observers. Observer bias is inevitable when collecting data manually (Fitzpatrick et al., 2009). By using multiple observers this bias can be prevented and/or estimated to some extent, but at the expense of increasing effort in manual annotation. By checking the annotations, we noted that the differences between observers were due to missed detections instead of false positives cases. Since the most common case is to have an absence of the species, it seems reasonable to assume that it is more likely that an expert incurs more on false negative cases (mark absence when there is a presence) than on false positive ones (interpret a sound from another species as the target signal, in this case, *L. knudseni*). Nevertheless it should be noticed that explosive breeding events in tropical anuran communities is particularly challenging in terms of signal identification due to the extremely high abundance and diversity of calling individuals simultaneously signalling.

The semi-supervised framework allowed to combine computer algorithms with a minor manual effort. At the initial step, unsupervised learning does not need any predefined label to structure the data. The manual effort only comes into play to assign labels for the cluster. Here, we selected randomly only 32 samples to identify the cluster as calls of *L. knudseni*, and hence, the manual effort was targeted. In general, the classification performance is expected to strongly depend on the correct definition of the boundaries of each cluster. In our example, two clusters were identified as calls of *L. knudseni*. While a first cluster (C1) had almost exclusively clear calls of *L. knudseni*, the other cluster (C3) had calls of dense multi-specific choruses, including calls of *L. knudseni*, but also of other species, such as *T. coriaceus*. As a result, the unsupervised boundary derived from an acoustic perspective differed slightly from the objective boundary: presence of *L. knudseni*.

While the cluster and the objective boundaries might differ, the results show that is possible to obtain high similarity between species and acoustic clusters. This is expected since each species has a species-specific call. Yet, the background and interferences (calls from the same species or from other sound biotic or abiotic sound sources) introduces noise and hence the cluster boundaries based on field recordings

might fluctuate and diverge from the desired output. More sophisticated semi-supervised learning techniques should be investigated, such as transductive support vector machines and label propagation (Chapelle et al., 2006). Such algorithms allow to draw decision boundaries that take into account both, the unsupervised arrangement and the few available labels.

The supervised learning approach requires, as the name implies, a higher degree of manual control and effort. This in turn helps to define boundaries that match the objectives pursued. To correctly train and tune a supervised learning algorithm, a considerable dataset that includes manual labelling is usually required. For the present study we used a dataset of 384 ROIs. Compared to the semi-supervised approach, the manual and computational effort are increased. However, this increased effort is reflected on the match between expert annotation and automated methods. Indeed, the supervised learning framework allowed to have expert level classification on the presence-absence of *L. knudseni*.

In order to train, tune and test a classifier, the selected observations that compose the annotated dataset have an essential role. While it is clear that a biased validation dataset can lead to biased detection estimates, we further showed that an adequate sampling strategy can lead to statistical classifiers able to generalize with less samples. Considering the possibilities to arrange the data with unsupervised learning and the multiple sampling strategies, the scope and potential to build statistical classifiers efficiently widens considerably. This in turn might lead to a definition of a simple and standardized protocol to find sounds of interest in large audio datasets. More standardized methods based on best practices from machine learning and sampling will ensure to have repeatable procedures with scientifically defendable results.

The present work sheds light into how to best combine methods to obtain accurate and efficient ecological data from passive acoustic recorders. We suggest to apply together human reasoning and computer algorithms to establish an end-to end framework that is generic, standardized and robust. We used an unsupervised learning approach to partition the data into homogeneous groups and apply a stratified sampling, thereby collecting a representative dataset of the acoustic diversity. This sample included not only the diversity of the sound of interest, but also the multiple interferences that are present in the recordings. Manual expert annotation can be included into the analytical procedure to train a semi-supervised or a supervised

classifier in order to carefully delineate a decision boundary. This dataset is expected to train a statistical classifier cost-effectively. Finally, a held-out dataset collected with random sampling should be used to validate the classification performance and give useful estimates of detection accuracy.

The above framework should be seen as a workflow with a feedback loop according to a conditional step: does the performance comply with the requirements imposed by the intended use of the system? To improve the performance, three basic not-exclusive alternatives can be envisaged: (1) collect more data, (2) add new features, or/and (3) try a different statistical classifier. In order to collect data efficiently, or to take advantage of structuring the data, unsupervised and semi-supervised learning methods are the ideal candidates. Yet, they are still an unexplored area on ecoacoustics. Novel methods to combine strengths of human reasoning and computational algorithms will likely become increasingly important to advancing passive acoustic monitoring.

Acknowledgments

We would like to thank the three experts that accepted the hard work of annotating the dataset: Elodie Courtois, Antoine Fouquet and Philippe Gaucher. We also thank Sylvain Hauptert and Chloé Dubset for their work on the transcription of functions from MATLAB to the open source Python programming language.

References

- Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1, e103. <https://doi.org/10.7717/peerj.103>
- Bishop, C., 2006. *Pattern Recognition and Machine Learning*, Information Science and Statistics. Springer-Verlag, New York.
- Blumstein, D.T., Mennill, D.J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J.L., Krakauer, A.H., Clark, C., Cortopassi, K.A., Hanser, S.F., McCowan, B., Ali, A.M., Kirschel, A.N.G., 2011. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology* 48, 758–767. <https://doi.org/10.1111/j.1365-2664.2011.01993.x>
- Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30, 1145–1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
- Brandes, T.S., Naskrecki, P., Figueroa, H.K., 2006. Using image processing to detect and classify narrow-band cricket and frog calls. *J. Acoust. Soc. Am.* 120, 2950–2957. <https://doi.org/10.1121/1.2355479>
- Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X.Z., Raich, R., Hadley, S.J., Hadley, A.S., Betts, M.G., 2012. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *J. Acoust. Soc. Am.* 131, 4640–4650.
- Browning, E., Gibb, R., Glover-Kapfer, P., Jones, K.E., 2017. *Passive acoustic monitoring in ecology and conservation (No. 1(2))*, WWF Conservation Technology Series. WWF-UK, Woking, United Kingdom.
- Chapelle, O., Schölkopf, B., Zien, A. (Eds.), 2006. *Semi-supervised learning, Adaptive computation and machine learning*. MIT Press, Cambridge, Mass.
- Dugan, P., Pourhomayoun, M., Shiu, Y., Paradis, R., Rice, A., Clark, C., 2013. Using High Performance Computing to Explore Large Complex Bioacoustic Soundscapes: Case Study for Right Whale Acoustics. *Procedia Computer Science* 20, 156–162. <http://dx.doi.org/10.1016/j.procs.2013.09.254>

- Eldridge, A., Casey, M., Moscoso, P., Peck, M., 2016. A new method for ecoacoustics? Toward the extraction and evaluation of ecologically-meaningful soundscape components using sparse coding methods. *PeerJ* 4, e2108. <https://doi.org/10.7717/peerj.2108>
- Fitzpatrick, M.C., Preisser, E.L., Ellison, A.M., Elkinton, J.S., 2009. Observer bias and the detection of low-density populations. *Ecological Applications* 19, 1673–1679. <https://doi.org/10.1890/09-0265.1>
- Frank, E., Trigg, L., Holmes, G., Witten, I.H., 2000. Technical Note: Naive Bayes for Regression. *Machine Learning* 41, 5–25. <https://doi.org/10.1023/A:1007670802811>
- Fristrup, K.M., Mennitt, D., 2012. Bioacoustical Monitoring in Terrestrial Environments. *Acoustics Today* 8, 16–24.
- Goëau, H., Glotin, H., Vellinga, W.-P., Planqué, R., Rauber, A., Joly, A., 2014. LifeCLEF Bird Identification Task 2014 (No. CLEF: Conference and Labs of the Evaluation Forum). Sheffield, United Kingdom.
- Heinicke, S., Kalan, A.K., Wagner, O.J.J., Mundry, R., Lukashevich, H., Köhl, H.S., 2015. Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods in Ecology and Evolution* 6, 753–763. <https://doi.org/10.1111/2041-210X.12384>
- Hripcsak, G., 2005. Agreement, the F-Measure, and Reliability in Information Retrieval. *Journal of the American Medical Informatics Association* 12, 296–298. <https://doi.org/10.1197/jamia.M1733>
- Keen, S.C., Shiu, Y., Wrege, P.H., Rowland, E.D., 2017. Automated detection of low-frequency rumbles of forest elephants: A critical tool for their conservation. *The Journal of the Acoustical Society of America* 141, 2715–2726. <https://doi.org/10.1121/1.4979476>
- Knight, E.C., Hannah, K.C., Foley, G.J., Scott, C.D., Brigham, R.M., Bayne, E., 2017. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conservation and Ecology* 12. <https://doi.org/10.5751/ACE-01114-120214>
- Lantz, B., 2013. *Machine Learning with R*. Packt Publishing Ltd.
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. *R News* 2, 18–22.

- Maaten, L. van der, Hinton, G., 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 2579–2605.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Priyadarshani, N., Marsland, S., Castro, I., 2018. Automated birdsong recognition in complex acoustic environments: a review. *Journal of Avian Biology*. <https://doi.org/10.1111/jav.01447>
- R Core Team, 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ranjard, L., Reed, B.S., Landers, T.J., Rayner, M.J., Friesen, M.R., Sagar, R.L., Dunphy, B.J., 2017. MatlabHTK: a simple interface for bioacoustic analyses using hidden Markov models. *Methods in Ecology and Evolution* 8, 615–621. <https://doi.org/10.1111/2041-210X.12688>
- Rubin, A., Babbie, E.R., 2008. *Research methods for social work*, 6th ed. ed. Thomson/Brooks/Cole, Belmont, CA.
- Sokolova, M., Japkowicz, N., Szpakowicz, S., 2006. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation, in: *AI 2006: Advances in Artificial Intelligence*, Lecture Notes in Computer Science. Presented at the Australasian Joint Conference on Artificial Intelligence, Springer, Berlin, Heidelberg, pp. 1015–1021. https://doi.org/10.1007/11941439_114
- Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 45, 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Stowell, D., 2018. Computational Bioacoustic Scene Analysis, in: *Computational Analysis of Sound Scenes and Events*. Springer, Cham, pp. 303–333. https://doi.org/10.1007/978-3-319-63450-0_11
- Sueur, J., Aubin, T., Simonis, C., 2008. Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics* 18, 213–226.
- Theodoridis, S., Koutroumbas, K., 2006. *Pattern Recognition, Third Edition*. ed. Academic Press.

- Thompson, S.K., 2012. *Sampling*, 3rd Edition. ed. Wiley-Blackwell, Hoboken, N.J.
- Thompson, W.L. (Ed.), 2004. *Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters*. Island Press, Washington.
- Towsey, M., Parsons, S., Sueur, J., 2014. Ecology and acoustics at a large scale. *Ecological Informatics, Ecological Acoustics* 21, 1–3. <https://doi.org/10.1016/j.ecoinf.2014.02.002>
- Towsey, M., Planitz, B., Nantes, A., Wimmer, J., Roe, 2012. A Toolbox for Animal Call Recognition. *Bioacoustics* 21, 107–125.
- Tukey, J.W., 1962. The Future of Data Analysis. *The Annals of Mathematical Statistics* 33, 1–67.
- Ulloa, J.S., Aubin, T., Llusia, D., Bouveyron, C., Sueur, J., 2018. Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis. *Ecological Indicators* 90, 346–355. <https://doi.org/10.1016/j.ecolind.2018.03.026>
- Ulloa, J.S., Gasc, A., Gaucher, P., Aubin, T., Réjou-Méchain, M., Sueur, J., 2016. Screening large audio datasets to determine the time and space distribution of Screaming Piha birds in a tropical forest. *Ecological Informatics* 31, 91–99. <https://doi.org/10.1016/j.ecoinf.2015.11.012>
- Van Rijsbergen, C.J., 1979. *Information retrieval*, 2d ed. ed. Butterworth-Heinemann, Newton, MA, USA.
- Webb, A.R., Copsey, K.D., 2011. *Statistical pattern recognition*, 3rd ed. ed. Wiley, Hoboken.
- Wells, K.D., 2007. *The ecology & behavior of amphibians*. The University of Chicago Press, Chicago.
- Wells, K.D., 1977. The social behaviour of anuran amphibians. *Animal Behaviour* 25, 666–693. [https://doi.org/10.1016/0003-3472\(77\)90118-X](https://doi.org/10.1016/0003-3472(77)90118-X)
- Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1, 67–82. <https://doi.org/10.1109/4235.585893>
- Xie, J., Towsey, M., Zhang, J., Roe, P., 2016. Adaptive frequency scaled wavelet packet decomposition for frog call classification. *Ecological Informatics* 32, 134–144. <https://doi.org/10.1016/j.ecoinf.2016.01.007>

4. TOWARDS AN END-TO-END FRAMEWORK

Zaki, M.J., Meira, W., 2014. Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press, New York, NY.

GENERAL DISCUSSION

*"Yo digo que no hay más canto
que el que sale de la selva
y que será el que lo entienda
fruto del árbol más alto"*
– Silvio Rodríguez (1994)

By sensing the acoustic environment we have access to a unique facet of biodiversity. The rich information found in every single second of audio echoes that wildlife at tropical latitudes is not only abundant, but also diverse and dynamic. Arrays of passive acoustic recorders are now deployed worldwide to monitor animal communities, but will this quantitative increase in data lead to a qualitative change in the way we do ecological research? Can acoustic monitoring become an effective tool for exploring patterns of biodiversity? The data collection problem has been leverage by passive acoustic sensors, but a new challenge have emerged: the examination and interpretation of data through quantitative methods. The standardized and efficient analysis of such data might be the first step to derive valuable ecological information for science and conservation.

1. Analysis of field audio recordings

Mixed methods have been envisaged and implemented to numerically analyse and derive valuable ecological information collected by passive acoustic sensors. Probably the most straightforward analysis consists in identifying species-specific calls in the recordings, by manual and/or automated annotations. Alternatively, the acoustic community can be analysed by examining its global structure. All these are fruitful

methodologies and should help to exploit the deluge of data being recorded in the natural environments.

1.1. Manual annotations

Since humans have a long experience interpreting environmental sounds it is relatively simple to gain the skills to scan and recognise few target sounds in audio recordings. When dealing with multiple target sounds in heterogeneous acoustic environments the problem gets more complex, training might require much more investment and variability between observers will likely increase (Chapter 4). When possible, the classification of complex soundscapes can be further refined by having multiple observers. Automated methods can greatly reduce the time required to scan through audio files, but their correct design, implementation, tuning and testing, might require significant amounts of time and probably give less reliable results than manual methods. Hence suitability of using automated versus manual scanning should be considered carefully according to the study (Swiston & Mennill, 2009). For example, Knight et al. (2017) compared implementation time to build and run automated acoustic classifiers against human listening and found that for datasets smaller than 36 hours, human listening was faster. Despite being more costly, traditional manual analysis might be the best option if dealing with small and manageable datasets.

Targeted at precise locations of interest and combined with sampling strategies, manual work can be optimized. As shown in our explosive breeding study (Chapter 2), manual annotation can be combined with acoustic indices (Sueur et al., 2014). The indices allow to detect strong changes in the acoustic community (Lellouch et al., 2014), and then a targeted subsample of the total recordings can be selected. In this way it is possible to obtain a manually workable dataset for expert annotations. Expert annotations allow to extract diverse and accurate information related to the species occurrences, getting all the benefits of human knowledge into the analysis. A clear objective in the analysis of the data is key to optimize expensive resources, such as expert manual annotation.

Manual annotations are valuable but time-consuming, and hence should be considered for re-use. We manually annotated a total of 902 recordings, equivalent to

more than 15 hours of continuous audio for different purposes (Chapters 1, 2 and 3). Nowadays, many researchers are being confronted to manually annotate their field audio recordings. Such information can be used, for example, to train, test and assess the performance of automated classifiers. For instance, the expert annotations of presence-absence of anuran species in Chapter 2 was reused to test and assess multiple automated methods in Chapter 4. Unfortunately, most of this data stays in laboratory shelves being neither shared nor accompanied with metadata. Currently, personal datasets are available from recordings which are saved in curated sound libraries, such as *The Macaulay Library of Natural Sounds* (<https://www.macaulaylibrary.org/>, Cornell University), *La Sonothèque* (<https://sonotheque.mnhn.fr/>, Muséum national d'Histoire naturelle) or the *Colección de Sonidos Ambientales* (<http://humboldt.org.co>, Instituto Humboldt Colombia), or are part of crowdsourcing projects, such as the *Xeno-canto* foundation (<http://www.xeno-canto.org>). However, manually annotated recordings from passive acoustic sensors are still rarely found in such libraries. Ecologists are collectively producing increasing amounts of audio classification data that can turn to be the building blocks of training libraries for automated methods. Indeed, a culture of transparent data exchange would favour the development of new pattern recognition tools particularly suited to ecoacoustic monitoring, and in turn, it will be possible to address questions at larger scales.

For large scale ecoacoustic monitoring projects innovative ideas should be envisaged to get manually labelled dataset, for example by distributing the task through crowdsourcing initiatives. In the closely related task of classifying images from hundreds of camera traps, researchers have build a citizen-science website (<http://www.snapshotserengeti.org>) where more than 28,000 registered users contributed to classify as much as 10.8 million images (Swanson et al., 2015). The data is currently used to address large-scale questions in community ecology and community dynamics. Collaborative annotations on some of the many charismatic sounds collected by acoustic sensors seem a sensible alternative to be considered to scale-up manual classification.

1.2. Computer assisted classification

Since passive acoustic recorders are recent devices in ecology, well-defined tools and protocols to implement automated analysis of data is still scarce – and urgently needed. The acoustic environment of the tropical environment presents two major challenges for automated analysis: (1) highly diverse types of sounds co-occur unceasingly, and (2) a large proportion of sounds are not identified at the species level. These particular constraints led us to develop innovative methods for the analysis of field audio recordings. We adapted, designed and tested methodologies to overcome these challenges: the cross-correlation of spectrogram (Chapter 1), the Multiresolution Analysis of Acoustic Diversity (Chapter 3) and a protocol to combine manual and automatic approaches within the machine learning framework (Chapter 4).

Simple, robust and straightforward techniques are especially interesting to test and adapt to tropical environments. They are simpler to explain and understand and reduce the chance of computational mistakes. We showed that the spectrogram cross-correlation could screen more than fifty thousand recordings and detect vocalizations of a target sound in the lowland rainforest, an intricate acoustic environment. The detections were then used to describe the spatial and temporal dynamics of the vocal activity of a population of *Lipaugus vociferans*. The cross-correlation was robust to the multiple interferences from other animal sounds and needed few manually annotated templates to run. This effective technique was found while searching for a fast, relatively simple and robust approach. The spectrogram cross-correlation is now incorporated into different software, facilitating its handling by non-experts. Yet, to derive meaningful results, it is fundamental to tune and test the system following best practices (Knight et al., 2017). This is why we comprehensively described our methodology, guiding others to follow a mapped road and hence contributing to building a body of literature of automated signal recognition for wildlife assessment in tropical environments. A significant proportion of birds (*e.g.* suboscines), anuran amphibians and insects have the stereotyped vocalizations and the procedure detailed in Chapter 1 is a guide to test thoroughly the spectrogram cross-correlation.

To date, one of the strongest constraints to analyse acoustic communities, is the lack of comprehensive datasets able to relate sounds with species. When labels are costly or simply not available, the parameters of supervised learning cannot be

estimated. Alternatives to overcome manual labelling include unsupervised learning, which can exploit information from unlabelled data by searching for structures on the data attributes (Bishop, 2006; Theodoridis and Koutroumbas, 2006; Hastie et al., 2009). Using unsupervised learning, we designed a method to decompose the acoustic community into few elementary spectro-temporal components, namely the Multiresolution Analysis of Acoustic Diversity (MAAD, Chapter 3). When confronted to manual annotation, MAAD gave strikingly similar partitioning. Tested to be robust under two contrasting scenarios, our method seems to be flexible and generalizable to other datasets. Indeed, MAAD was designed to be data-driven, which gives the possibility to adapt to different situations, offering new possibilities to analyse other poorly documented environments, such as freshwater, corral reefs and deep seas.

Previous studies involving automated analysis have mainly focused on acoustic indices or supervised learning. MAAD uses statistical classification in a new direction, opening further possibilities to ecoacoustics. For instance, MAAD can be used to give a global characterisation of the soundscape using common ecological measures of diversity. The number of clusters can be used as an estimate of soundtype richness, and since the groups have associated abundances and distances between groups, other diversity indices can be calculated, such as Shannon, Gini-Simpson or Rao's quadratic entropy (Magurran, 2004; Pavoine et al., 2005; Gotelli and Chao, 2013). Alternatively, MAAD can be seen as a general method to organize the constellation of environmental sounds, optimizing the search through large collections. Suppose you are searching for a specific set of books. Now, where would prefer to go, to a library where the books are randomly distributed, or where books are organized by subject and author name? MAAD opens new perspectives to analyse passive acoustic recordings, facilitating posterior analyses that combine machine speed and human reasoning.

To date, most of the techniques to analyse passive acoustic sensors are either manual or fully automated. Much effort has been placed at building an all-purpose classifier. Multiple competitions have been organised to solve the problem of automated classification of field audio recordings, MLSP 2013 Bird Classification Challenge (<https://www.kaggle.com/c/mlsp-2013-birds/data>), ICML 2013 Bird Challenge (<https://www.kaggle.com/c/the-icml-2013-bird-challenge>), LifeCLEF Bird Identification Task 2014 (Goëau et al., 2014), to name a few. These competitions have certainly contributed to highlight the most robust and feasible machine learning

approaches, yet it is unlikely that a single algorithm will classify all the sounds in real field situations. Machine learning studies agree that an increase in sample size will very likely incur in better classifiers, yet with the current available annotated datasets it seems to be unlikely to derive a classifier for all species on all scenarios.

Instead of looking for fully automated methods, we underline that solutions aimed at combining manual and automated analysis should be envisaged. Such procedure seems even more important for poorly documented habitats, where calls are yet to be described, annotated and saved in sound libraries. We framed a protocol based on machine learning that sheds light on how to best integrate the strengths from both sides. By screening the audio dataset and arrange it with unsupervised learning, we manage to select a representative sample of heterogeneous sounds efficiently (Chapter 4). We showed that an alternative at hand to most practitioners could be to frame a well-defined protocol that includes building and testing classifiers. While most ecologists are not well trained to manage the datasets from passive acoustic recorders, the task is little by little facilitated by open-source packages and documentation. For now, the tools are in the hands of computer scientists, but the questions are posed by ecologists. A better integration between researchers will likely benefit both sides and in general, the field of ecoacoustics.

1.3. Global characterisation of the soundscape

With the aim at analysing acoustic patterns at a larger scale, multiple acoustic indices have been designed (e.g. Sueur et al., 2008; Pieretti et al., 2011; Villanueva-Rivera et al., 2011; Kasten et al., 2012; for a review see Sueur et al., 2014). For populations or communities of organisms that signal acoustically in large aggregations, the use of acoustic indices seems particularly suitable. In dense choruses, such as the ones investigated in Chapter 2, the signals end-up intermingled with each other, losing the temporal envelope. However, the spectral characteristics are summed between individual sounds, which results in a clear spectral mark in the frequency domain of the acoustic environment, perceived as a hum or a hiss by our auditory system. During our studies, we inspected the overall amplitude of audio recordings by computing the root-mean-square of the signal (Chapter 2). Such quantification of the signal, which can be viewed as an acoustic index, was effective to strong changes in the anuran community

related to explosive breeding events. Moreover, using a statistical classifier to discriminate between global spectrums, we highlighted a characteristic spectral signature of these peculiar breeding events, rising multiple questions related to the community acoustic structures found in tropical forests (see next subsection).

While acoustic indices were not explicitly mentioned in our final methods or results, a global analysis of the acoustic environment was a fundamental exploratory step in our data. Some studies have searched for the link between acoustic indices and ecological processes (Gasc et al., 2012; Towsey et al., 2014; Harris et al., 2016), and others have sought to highlight their value for exploratory analyses (Phillips et al., 2018). Either way, the use of acoustic indices has opened new possibilities for ecoacoustics numerical analyses. Indeed, all of the aforementioned machine learning possibilities can be envisaged at a global scale, focusing the analysis at the soundscape level. After all, it is fundamental to examine the differences and similarities of the diversity of sounds, without ever losing the sight of the whole.

*

We developed techniques that could be flexible and easily adapted beyond our own datasets, contributing to the efforts on systematic, repeatable, objective and scalable ecoacoustic studies. We opened paths for the analysis of field audio recordings towards multiple directions, including supervised and unsupervised learning, combined with manual analysis and acoustic indices. The possibilities are large and the methods are clearly in their infancy. The full potential of ecoacoustic monitoring lies on the advances of methodological procedures to accurately and efficiently scrutinize the information condensed in digital format. Better integration of the multiple tools from machine learning must continue to develop and enable us to understand the complexities of tropical acoustic environments. Computational tools not only have the potential in assisting to the automated classification, but also might help us gain insights from the available data through data-mining (Stowell, 2018). These algorithms represent a major set of computational tools for knowledge discovery in large databases, which will be increasingly essential in the era of data-intensive ecoacoustics. Better integration and collaboration might be the answer to the growing need for classification

tools and sound libraries that facilitate robust ecological research (Browning et al., 2017).

A major endeavour of biodiversity monitoring resides in the standardization of methods that allow scaling up to large scales. Clearly, the use of passive acoustic sensors coupled with algorithmic analyses provide an objective, repeatable and cost-efficient alternative. Generalized methods for ecologists in a set of standard packages should become increasingly available to everyone, everywhere (Borcard et al., 2011). New hardware and software are currently being designed, this work is part of this movement that poses new foundations for biodiversity monitoring through ecoacoustics.

2. Patterns and dynamics of the tropical acoustic environment

The acoustic environment is more than the sum of its elementary sounds. Structural patterns have been documented in acoustic communities of birds, amphibians and insects, probably reflecting complex heterospecific interactions. Our studies revealed temporal patterns of the *Lipaugus vociferans* with peaks of activity shifted from times where the soundscape had higher levels of acoustic activity (Chapter 1), and a strong effect of the median call frequency explaining clusters found in the nocturnal acoustic community (Chapter 3). Such observations might be explained in the light of ecoacoustic models. The acoustic niche hypothesis (ANH), based on the ecological niche concept (Hutchinson, 1957), is linked to inter-specific acoustic competition (Krause, 1993). As the acoustic environment is a shared resource, organisms would have evolved to occupy specific spectro-temporal niches, decreasing the risk of heterospecific mating and information masking.

When assessing the ephemeral communities of explosive breeding amphibians, we expected to find similar frequency partitioning patterns. Amphibians strongly rely on acoustic communication for sexual selection, and during explosive breeding events multiple species are calling intensively at the same time (Gerhardt & Huber, 2002; Gottsberger & Gruber, 2004). Since the temporal window was the same for the co-occurring species, we expected to find a spectral structure in the acoustic community that would allow each species to exchange signals, and thus support the ANH. Yet, our results showed the opposite pattern: comparing pre-explosive and explosive breeding

acoustic communities, we found significant diminution in the acoustic diversity. The diminution was not evident when comparing species richness, but only when integrating a spectral distance between species. This unexpected result could be explained by sexual selection pressures and behavioural differences. While for prolonged breeders female choice is crucial in determining male reproductive success, during explosive breeding there would be no time to exchange acoustic signals, leaving females from explosive breeding species few opportunities to choose (Wells, 1977). Such weak female sexual selection would lead to acoustic communities with more overlapping signals. Clustering in acoustic signalling has been observed in tropical birds, suggesting that this signalling behaviour may facilitate the exchange of signals between species (Tobias et al., 2014). Regarding explosive breeding amphibians, such potential convergence might be used as a heterospecific cue that signals the availability of short-lived breeding sites (Bee, 2007; Swanson et al., 2007).

The contrasted results in acoustic community structures found in our studies highlight the importance of behavioural and experimental studies to better understand the underlying causes of the patterns observed in the soundscape. In addition, history might be an important predictor factor and hence phylogenetic data should be taken into account. In turn, the documented patterns in the soundscape can raise new behavioural and ecological questions. At the species level, studies have revealed that acoustic signatures incorporate cues to individuality, sex, age, body size and even physiological state (Aubin et al., 2007; Catchpole and Slater, 2008). At the community level, acoustic structures may encode information related to ecological processes. For example, disruption on the acoustic signature produced by cricket communities in New Caledonia showed to be indicative of the presence of *Wasmania auropunctata*, an invasive and devastating ant (Gasc et al., 2017). With a better comprehension of the underlying processes of community acoustic signatures, effective and non-invasive procedures will be emerging to improve biodiversity monitor. The better we link the soundscape with ecological processes, the better we will be able to monitor biodiversity and infer sound decision policies.

3. Acoustic surveys to monitor biodiversity at large scales

A fundamental part of conservation biology is focused in assessing the state and trends of biodiversity. To this end, monitoring studies are necessary to accumulate baseline data, which are still lacking for many tropical environments. Integration of research, management and monitoring is key to gain reliable knowledge about ecological systems. Tropical environments, which are highly diverse, complex, poorly understood and showing increasing human pressures, clearly pose challenges for conservation (Groom et al., 2012). Monitoring frameworks with efficient sampling designs are needed to address the multiple questions posed by local, regional and international commitments, such as the Convention on Biological Diversity (Collen et al., 2013; Turner, 2014; Steenweg et al., 2017).

Using acoustic sensors we succeeded to follow animal dynamics at multiple points simultaneously, in a standardized and at cost-efficient way. We tracked a population of the tropical bird *Lipaugus vociferans* at 24 plots during 25 days, and assemblages of anurans at five sites during four months. Over the last decade passive acoustic monitoring has emerged as an influential tool for studying wildlife, proving to be effective at sampling terrestrial (Aide et al., 2013; Blumstein et al., 2011; Fristrup and Mennitt, 2012; Kalan et al., 2015; Thompson et al., 2010; Wrege Peter H. et al., 2017) and underwater environments (Miksis-Olds et al., 2010; Sousa-Lima et al., 2013; Tregenza et al., 2016; Caruso et al., 2017). Our study adds up to this increasing list, demonstrating the feasibility and suitability of using acoustic sensor arrays to monitor key organisms of tropical forests, and evidencing that the full potential of acoustic monitoring is yet to be harnessed.

By coupling acoustic with environmental sensors and information related to habitat structure, we managed to tackle key ecological questions related to the studied species. The recent emergence of acoustic sensors may mirror a wider trend, the extensive use of electronic sensors in ecology. Remote triggered traps, environmental DNA, satellite and airborne imaging provide complementary information about biodiversity (Collen et al., 2013; Turner, 2014). Integration of such data would allow building up a harmonised observation system.

Quantitative sampling and monitoring of multiple groups is essential to best determine management programs for the conservation of biodiversity (Collen et al.,

2013). To cover a larger range of animals, probably the most evident integration for *in situ* automated data collection consists in coupling camera traps and acoustic monitoring. Such technology seems complementary, while camera traps is sensitive to the large, and often silent, ground dwelling animals, acoustic sensors are able to detect loud animals, which have often cryptic coloration or are small. Finally, data collected *in situ* should be used in concert with *ex situ* satellite or aircraft imaging to reach global scales. The digital era for ecology present great possibilities, but also needs innovation in the analysis of data to standardize and combine the procedures, a great challenge on its own (Turner, 2014).

Passive acoustic sensing is clearly scaling-up and going global. Emergent alternatives on recording equipment, software analysis, and the growing ecoacoustic monitoring literature highlight how popular this technique has become. A great amount of data is being collectively being produced, but is still fragmented and little standardization procedures currently exist. While monitoring design should be followed in order to answer research objectives, reporting metadata allows transparent exchange of information allowing larger-scale analyses (Hampton et al., 2013). As for other data collection protocols in ecology, a better coordination and collaboration between research groups would allow to combine efforts for broader biodiversity assessment and novel scientific insights (Reichman et al., 2011).

4. Rainforest soundscapes to promote biodiversity conservation

Cost-effective monitoring and a better understanding of animal ecology would hopefully lead to conservation action and policies, but public engagement is also crucial (Novacek, 2008). Lengthy audio field recordings collected during ecoacoustic monitoring, appropriately curated, may also be used in alternative ways to convey importance, wonder and relevance of biodiversity to the general public. As an experimental work, the artist Bérénice Sevestre and myself edited a short film based on videos and audio recordings collected during the fieldwork in French Guiana (Figure S1 on Appendix). The five-minute documentary participated in the film festivals: *Les Chercheurs Font Leur Cinéma*, where it was prized with the public and jury awards. It was further presented in the international film festival *Parisciences* and was part of the

exposition la *Nature Monte Le Son*. The film is now freely available on the Internet, where it reached more than 2,000 visualizations in a few weeks. Although this is clearly an amateur audio-visual work, it is an example that evidences how people are particularly lured to hearing the natural world.

Sonic experiences have a unique surrounding character, which envelopes us, providing our sense of place. Information-rich soundscapes, composed of multiple textures, insistent rhythms and repetitions and unbreakable sequences, seem to lure, not only tropical travellers of the last two hundred years, but also the general public. Sonic experiences of the natural world are demanded and valued, the evidence is clear from the multiple expositions, e.g. *Le Grand Orchestre des Animaux* (Fondation Cartier pour l'art contemporain, 2016) and *La Nature Monte le Son* (Muséum Vauban, 2018), and audio albums, e.g. *Voices of the Rainforest* (Feld, 2011) and *Why do whales and children sing* (Dunn, 1999), to give but a few examples.

Public engagement is key to the future of conservation science and to face the biodiversity crisis in general (Novacek, 2008; Kudavidanage et al., 2012). Besides logical explanations on the importance of preserving proper ecosystem functioning, there is also the need to touch people sensibility. Emotions play a central role in the decisions we make, it is therefore not a surprise that the scientific community has increased its interest into artistic projects to question and reach a wider public (Jacobson et al., 2007). Indeed, artists have the sensibility to talk in a universal language. Interdisciplinary work between artists, conservationists and ecologists seem most pressing in the context of anthropogenic climate change (Moser & Dilling, 2011), and biodiversity issues (Novacek, 2008). The sound-art project *Fragments of extinction*, which explores the acoustic environment of the remaining primary tropical forests, is an example of such interdisciplinary possibilities in the general framework of ecoacoustics (Monacchi, 2013). The experience of listening is often one of perceiving the inseparability of phenomena. Our sensibility to sonic experiences, combined with field recordings might have the ability to promote personal ways of thinking about the natural world, building new relationships between humans, the natural environment and the rest of species that inhabit it. After all, it was probably a poet, Wolfgang von Goethe, that moved and inspired Humboldt to explore and study equinoctial regions (Wulf, 2015).

*

Two hundreds years ago, Humboldt talked about hums and buzzes filling up the air of Neotropical forests. Some of these sounds have been documented, and many behavioural studies have revealed the underlying mechanisms for the diversity of these sounds. Yet, most of the constellations of sounds we now hear are yet to be meticulously searched and identified. When turning to calls that are nocturnal or far from our reach, such as the canopy of tropical forests, we can realize how little we know about tropical soundscapes. Sometimes, event prevalent sounds remain enigmatic, hard to even classify in broad categories. Regarding a sound that remained mysterious for 10 years of field work, (Dias et al., 2017) asked: "is it a bird, is it a frog, or a bush cricket?". It turned out to be the stridulations of the katydid *Paracycloptera grandifolia*. As more sounds are identified and structural patterns of the acoustic environment are linked with ecological processes, the horizons for ecoacoustics are expanded. Nature has abundant secrets to whisper; thus, our understanding of life and how to best preserve it will thrive if we lend an attentive ear.

References

- Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. *PeerJ* 1, e103. <https://doi.org/10.7717/peerj.103>
- Aubin, T., Mathevon, N., Staszewski, V., Boulinier, T., 2007. Acoustic communication in the Kittiwake *Rissa tridactyla*: potential cues for sexual and individual signatures in long calls. *Polar Biology* 30, 1027–1033. <https://doi.org/10.1007/s00300-007-0262-6>
- Bee, M.A., 2007. Selective phonotaxis by male wood frogs (*Rana sylvatica*) to the sound of a chorus. *Behav Ecol Sociobiol* 61, 955–966. <https://doi.org/10.1007/s00265-006-0324-8>
- Bickford, D., Posa, M.R.C., Qie, L., Campos-Arceiz, A., Kudavidanage, E.P., 2012. Science communication for biodiversity conservation. *Biological Conservation* 151, 74–76. <https://doi.org/10.1016/j.biocon.2011.12.016>
- Bishop, C., 2006. *Pattern Recognition and Machine Learning, Information Science and Statistics*. Springer-Verlag, New York.
- Blumstein, D.T., Mennill, D.J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J.L., Krakauer, A.H., Clark, C., Cortopassi, K.A., Hanser, S.F., McCowan, B., Ali, A.M., Kirschel, A.N.G., 2011. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology* 48, 758–767. <https://doi.org/10.1111/j.1365-2664.2011.01993.x>
- Borcard, D., Gillet, F., Legendre, P., 2011. *Numerical ecology with R, Use R!* Springer, New York.
- Browning, E., Gibb, R., Glover-Kapfer, P., Jones, K.E., 2017. *Passive acoustic monitoring in ecology and conservation (No. 1(2))*, WWF Conservation Technology Series. WWF-UK, Woking, United Kingdom.
- Caruso, F., Alonge, G., Bellia, G., Domenico, E.D., Grammauta, R., Larosa, G., Mazzola, S., Riccobene, G., Pavan, G., Papale, E., Pellegrino, C., Pulvirenti, S., Sciacca, V., Simeone, F., Speziale, F., Viola, S., Buscaino, G., 2017. Long-Term Monitoring of Dolphin Biosonar Activity in Deep Pelagic Waters of the Mediterranean Sea. *Scientific Reports* 7, 4321. <https://doi.org/10.1038/s41598-017-04608-6>

- Catchpole, C., Slater, P.J.B., 2008. Bird song: biological themes and variations. Cambridge University Press, Cambridge.
- Collen, B., Pettorelli, N., Baillie, J.E.M., Durant, S.M. (Eds.), 2013. Biodiversity Monitoring and Conservation: Bridging the Gap between Global Commitment and Local Action. Wiley-Blackwell, Oxford, UK. <https://doi.org/10.1002/9781118490747>
- Dias, I.R., Chamorro-Rengifo, J., Solé, M., 2017. Is it a bird, is it a frog or a bush cricket? On an enigmatic nocturnal calling song recorded at different locations in southern Bahia, Brazil. *Spixiana* 40, 189–192.
- Dunn, D., 1999. Why do whales and children sing: a guide to listen in nature. Earth Ear, Santa Fe.
- Feld, S., 2011. Voices of the rainforest. Smithsonian Folkways, Washington, D.C.
- Fondation Cartier pour l'art contemporain, 2016. Le Grand Orchestre des Animaux. Paris.
- Fristrup, K.M., Mennitt, D., 2012. Bioacoustical Monitoring in Terrestrial Environments. *Acoustics Today* 8, 16–24.
- Gasc, A., Anso, J., Sueur, J., Jourdan, H., Desutter-Grandcolas, L., 2017. Cricket calling communities as an indicator of the invasive ant *Wasmannia auropunctata* in an insular biodiversity hotspot. *Biol Invasions* 1–13. <https://doi.org/10.1007/s10530-017-1612-0>
- Gasc, A., Sueur, J., Jiguet, F., Devictor, V., Grandcolas, P., Burrow, C., Depraetere, M., Pavoine, S., 2012. Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? *Ecological Indicators* 279–287.
- Gerhardt, H.C., Huber, F., 2002. Acoustic communication in insects and anurans: common problems and diverse solutions. University of Chicago Press, Chicago.
- Goëau, H., Glotin, H., Vellinga, W.-P., Planqué, R., Rauber, A., Joly, A., 2014. LifeCLEF Bird Identification Task 2014 (No. CLEF: Conference and Labs of the Evaluation Forum). Sheffield, United Kingdom.
- Gotelli, N.J., Chao, A., 2013. Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data, in: *Encyclopedia of Biodiversity*. Elsevier, pp. 195–211.

- Gottsberger, B., Gruber, E., 2004. Temporal partitioning of reproductive activity in a neotropical anuran community. *Journal of Tropical Ecology* 20, 271–280. <https://doi.org/10.1017/S0266467403001172>
- Groom, M.J., Meffe, G.K., Carroll, C.R., 2012. *Principles of Conservation Biology*. Sinauer.
- Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller Archer L, Duke Clifford S, Porter John H, 2013. Big data and the future of ecology. *Frontiers in Ecology and the Environment* 11, 156–162. <https://doi.org/10.1890/120103>
- Harris, S.A., Shears, N.T., Radford, C.A., 2016. Ecoacoustic indices as proxies for biodiversity on temperate reefs. *Methods Ecol Evol* 7, 713–724. <https://doi.org/10.1111/2041-210X.12527>
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, 2nd ed*, Springer Series in Statistics. Springer-Verlag, New York.
- Hutchinson, G.E., 1957. Concluding Remarks. *Cold Spring Harb Symp Quant Biol* 22, 415–427. <https://doi.org/10.1101/SQB.1957.022.01.039>
- Jacobson, S.K., Mcduff, M.D., Monroe, M.C., 2007. Promoting conservation through the arts: outreach for hearts and minds. *Conservation Biology* 21, 7–10. <https://doi.org/10.1111/j.1523-1739.2006.00596.x>
- Kalan, A.K., Mundry, R., Wagner, O.J.J., Heinicke, S., Boesch, C., Köhl, H.S., 2015. Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. *Ecological Indicators* 54, 217–226. <https://doi.org/10.1016/j.ecolind.2015.02.023>
- Kasten, E.P., Gage, S.H., Fox, J., Joo, W., 2012. The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology. *Ecological Informatics* 12, 50–67. <https://doi.org/10.1016/j.ecoinf.2012.08.001>
- Knight, E.C., Hannah, K.C., Foley, G.J., Scott, C.D., Brigham, R.M., Bayne, E., 2017. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conservation and Ecology* 12. <https://doi.org/10.5751/ACE-01114-120214>

- Krause, B.L., 1993. The niche hypothesis: a virtual symphony of animal sounds, the origins of musical expression and the health of habitats. *The Soundscape Newsletter* 6, 4–6.
- Lellouch, L., Pavoine, S., Jiguet, F., Glotin, H., Sueur, J., 2014. Monitoring temporal change of bird communities with dissimilarity acoustic indices. *Methods in Ecology and Evolution* 5, 495–505. <https://doi.org/10.1111/2041-210X.12178>
- Magurran, A.E., 2004. *Measuring biological diversity*. Blackwell Pub, Malden, Ma.
- Miksis-Olds, J.L., Nystuen, J.A., Parks, S.E., 2010. Detecting marine mammals with an adaptive sub-sampling recorder in the Bering Sea. *Applied Acoustics, Proceedings of the 4th International Workshop on Detection, Classification and Localization of Marine Mammals Using Passive Acoustics and 1st International Workshop on Density Estimation of Marine Mammals Using Passive Acoustics* 71, 1087–1092. <https://doi.org/10.1016/j.apacoust.2010.05.010>
- Monacchi, D., 2013. Fragments of Extinction: acoustic biodiversity of primary rainforest ecosystems. *Leonardo Music Journal* 23–25. https://doi.org/10.1162/LMJ_a_00148
- Moser, S.C., Dilling, L., 2011. Communicating climate change: closing the science-action gap, in: *The Oxford Handbook of Climate Change and Society*. OUP Oxford, pp. 161–174.
- Muséum Vauban, 2018. *La Nature Monte le Son*.
- Novacek, M.J., 2008. Engaging the public in biodiversity issues. *PNAS* 105, 11571–11578. <https://doi.org/10.1073/pnas.0802599105>
- Pavoine, S., Ollier, S., Pontier, D., 2005. Measuring diversity from dissimilarities with Rao's quadratic entropy: Are any dissimilarities suitable? *Theoretical Population Biology* 67, 231–239. <https://doi.org/10.1016/j.tpb.2005.01.004>
- Phillips, Y.F., Towsey, M., Roe, P., 2018. Revealing the ecological content of long-duration audio-recordings of the environment through clustering and visualisation. *PLOS ONE* 13, e0193345. <https://doi.org/10.1371/journal.pone.0193345>
- Pieretti, N., Farina, A., Morri, D., 2011. A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). *Ecological Indicators* 11, 868–873. <https://doi.org/10.1016/j.ecolind.2010.11.005>

- Reichman, O.J., Jones, M.B., Schildhauer, M.P., 2011. Challenges and Opportunities of Open Data in Ecology. *Science* 331, 703–705. <https://doi.org/10.1126/science.1197962>
- Rodriguez, A., Gasc, A., Pavoine, S., Grandcolas, P., Gaucher, P., Sueur, J., 2014. Temporal and spatial variability of animal sound within a neotropical forest. *Ecological Informatics, Ecological Acoustics* 21, 133–143. <https://doi.org/10.1016/j.ecoinf.2013.12.006>
- Rodriguez, S., 1994. *Yo Digo que las Estrellas*. Fonomusic, Spain.
- Sousa-Lima, R.S., Norris, T.F., Oswald, J.N., Fernandes, D.P., 2013. A Review and Inventory of Fixed Autonomous Recorders for Passive Acoustic Monitoring of Marine Mammals. *Aquatic Mammals* 39, 23–53.
- Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J.T., Burton, C., Townsend, S.E., Carbone, C., Rowcliffe, J.M., Whittington, J., Brodie, J., Royle, J.A., Switalski, A., Clevenger, A.P., Heim, N., Rich, L.N., 2017. Scaling-up camera traps: monitoring the planet’s biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment* 15, 26–34. <https://doi.org/10.1002/fee.1448>
- Stowell, D., 2018. Computational Bioacoustic Scene Analysis, in: *Computational Analysis of Sound Scenes and Events*. Springer, Cham, pp. 303–333. https://doi.org/10.1007/978-3-319-63450-0_11
- Sueur, J., Farina, A., Gasc, A., Pieretti, N., Pavoine, S., 2014. Acoustic Indices for Biodiversity Assessment and Landscape Investigation. *Acta Acustica united with Acustica* 100, 772–781. <https://doi.org/10.3813/AAA.918757>
- Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S., 2008. Rapid Acoustic Survey for Biodiversity Appraisal. *PLoS ONE* 3. <https://doi.org/10.1371/journal.pone.0004065>
- Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., Packer, C., 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data* 2, 150026. <https://doi.org/10.1038/sdata.2015.26>
- Swanson, E.M., Tekmen, S.M., Bee, M.A., 2007. Do female frogs exploit inadvertent social information to locate breeding aggregations? *Canadian Journal of Zoology* 85, 921–932. <https://doi.org/10.1139/Z07-074>

- Swiston, K.A., Mennill, D.J., 2009. Comparison of manual and automated methods for identifying target sounds in audio recordings of Pileated, Pale-billed, and putative Ivory-billed woodpeckers. *Journal of Field Ornithology* 80, 42-50. <https://doi.org/10.1111/j.1557-9263.2009.00204.x>
- Thompson, M.E., Schwager, S.J., Payne, K.B., 2010. Heard but not seen: an acoustic survey of the African forest elephant population at Kakum Conservation Area, Ghana. *African Journal of Ecology* 48, 224–231. <https://doi.org/10.1111/j.1365-2028.2009.01106.x>
- Tobias, J.A., Planqué, R., Cram, D.L., Seddon, N., 2014. Species interactions and the structure of complex communication networks. *PNAS* 111, 1020–1025. <https://doi.org/10.1073/pnas.1314337111>
- Towsey, M., Parsons, S., Sueur, J., 2014a. Ecology and acoustics at a large scale. *Ecological Informatics, Ecological Acoustics* 21, 1–3. <https://doi.org/10.1016/j.ecoinf.2014.02.002>
- Towsey, M., Wimmer, J., Williamson, I., Roe, P., 2014b. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. *Ecological Informatics* 21, 110–119. <https://doi.org/10.1016/j.ecoinf.2013.11.007>
- Tregenza, N., Dawson, S., Rayment, W., Verfuss, U., 2016. Listening to Echolocation Clicks with PODs, in: *Listening in the Ocean, Modern Acoustics and Signal Processing*. Springer, New York, NY, pp. 163–206. https://doi.org/10.1007/978-1-4939-3176-7_7
- Turner, W., 2014. Sensing biodiversity. *Science* 346, 301–302. <https://doi.org/10.1126/science.1256014>
- Villanueva-Rivera, L.J., Pijanowski, B.C., Doucette, J., Pekin, B., 2011. A primer of acoustic analysis for landscape ecologists. *Landscape Ecol* 26, 1233. <https://doi.org/10.1007/s10980-011-9636-9>
- Wells, K.D., 1977. The social behaviour of anuran amphibians. *Animal Behaviour* 25, 666–693. [https://doi.org/10.1016/0003-3472\(77\)90118-X](https://doi.org/10.1016/0003-3472(77)90118-X)
- Wrege Peter H., Rowland Elizabeth D., Keen Sara, Shiu Yu, Matthiopoulos Jason, 2017. Acoustic monitoring for conservation in tropical forests: examples from forest elephants. *Methods in Ecology and Evolution* 8, 1292–1301. <https://doi.org/10.1111/2041-210X.12730>

GENERAL DISCUSSION

Wulf, A., 2015. *The invention of nature: the adventures of Alexander von Humboldt, the lost hero of science*. John Murray, London.

APPENDIX

Appendix Contents

Related publications and oral communications
 Supplementary information to Chapter 1
 Supplementary information to Chapter 2
 Supplementary information to Chapter 3
 Supplementary information to Chapter 4
Supplementary information to General discussion
 Synthèse

Related publications and oral communications

Publications

Ulloa, J.S., Aubin, T., Llusia, D., Sueur, J. Towards an end-to-end classification framework for passive acoustic recordings. *In preparation*.

Desjonquères, C., Rybak, F., **Ulloa, J.S.**, Kempf, A., Bar-Hen, A., Sueur, J. Monitoring the acoustic activity of an aquatic insect population in relation to temperature, vegetation and noise. *Freshwater Biology*, *Under review*.

Ulloa, J.S., Aubin, T., Llusia, D., Courtois, É., Gaucher, P., Fouquet, A., Pavoine, S., Sueur, J. Explosive breeding in tropical anurans: environmental triggers, community composition and acoustic structure. *BMC Ecology*. *To be submitted*.

Ulloa, J.S., Aubin, T., Llusia, D., Bouveyron, C., Sueur, J., 2018. Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis. *Ecological Indicators* 90, 346–355. <https://doi.org/10.1016/j.ecolind.2018.03.026>

Ulloa, J.S., Gasc, A., Gaucher, P., Aubin, T., Réjou-Méchain, M., Sueur, J., 2016. Screening large audio datasets to determine the time and space distribution of Screaming Piha birds in a tropical forest. *Ecological Informatics* 31, 91–99. <https://doi.org/10.1016/j.ecoinf.2015.11.012>

Oral communications

Ulloa, J.S., Aubin, T., Llusia, D., Courtois, É., Gaucher, P., Fouquet, A., Pavoine, S., Sueur, J., 2018. Causes and consequences of explosive breeding amphibian communities revealed by remote sensors. Presented at the European Conference of Tropical Ecology, Paris, France.

- Ulloa, J.S.**, Aubin, T., Sueur, J., 2017a. Monitoring animal communities and populations using acoustic sensors. Presented at the 13th Meeting on Ecology and Behaviour, Chizé, France.
- Ulloa, J.S.**, Aubin, T., Sueur, J., 2017b. Monitoring animal populations and communities in the tropical forest using acoustic sensors: case studies in French Guiana. Presented at the 28th International Congress for Conservation Biology, Cartagena, Colombia.
- Ulloa, J.S.**, Bouveyron, C., Llusia, D., Aubin, T., Sueur, J., 2017d. Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis. Presented at the 26th International Bioacoustics Congress, Haridwar, India.
- Ulloa, J.S.**, Gasc, A., Gaucher, P., Réjou-Méchain, M., Aubin, T., Sueur, J., 2015. Temporal and spatial dynamics of a tropical soundmark: the song of the Screaming piha *Lipaugus vociferans*. Presented at the 25th International Bioacoustics Congress, Murnau, Germany.

Supplementary information to Chapter 1

Supplementary audio data can be found online at:

<https://doi.org/10.1016/j.ecoinf.2015.11.012>

Audio S1. Focused recording of the typical vocal display of a *Lipaugus vociferans*. Recorded on November 16, 2010, at 1:30 pm, in site M-XI with the canopy microphone.

Audio S2. Soundscape of the tropical forest. Among the numerous sounds, a typical song of the *Lipaugus vociferans* can be heard between second 9 and 11. Despite the interferences, this song was detected by our detection system without false positives. Recorded on November 20, at 6:00 am in site M-XI with the ground microphone.

Supplementary information to Chapter 2

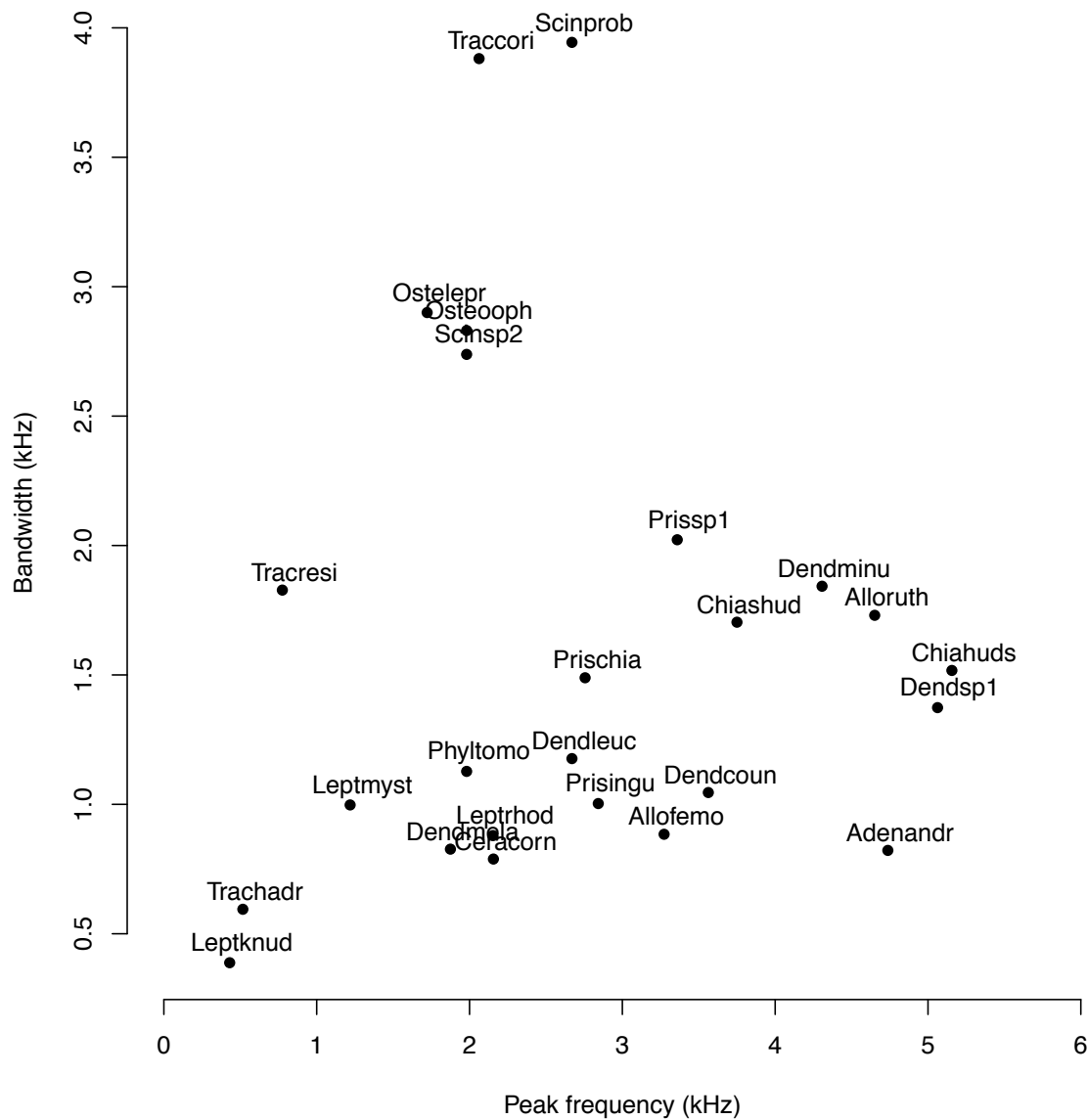
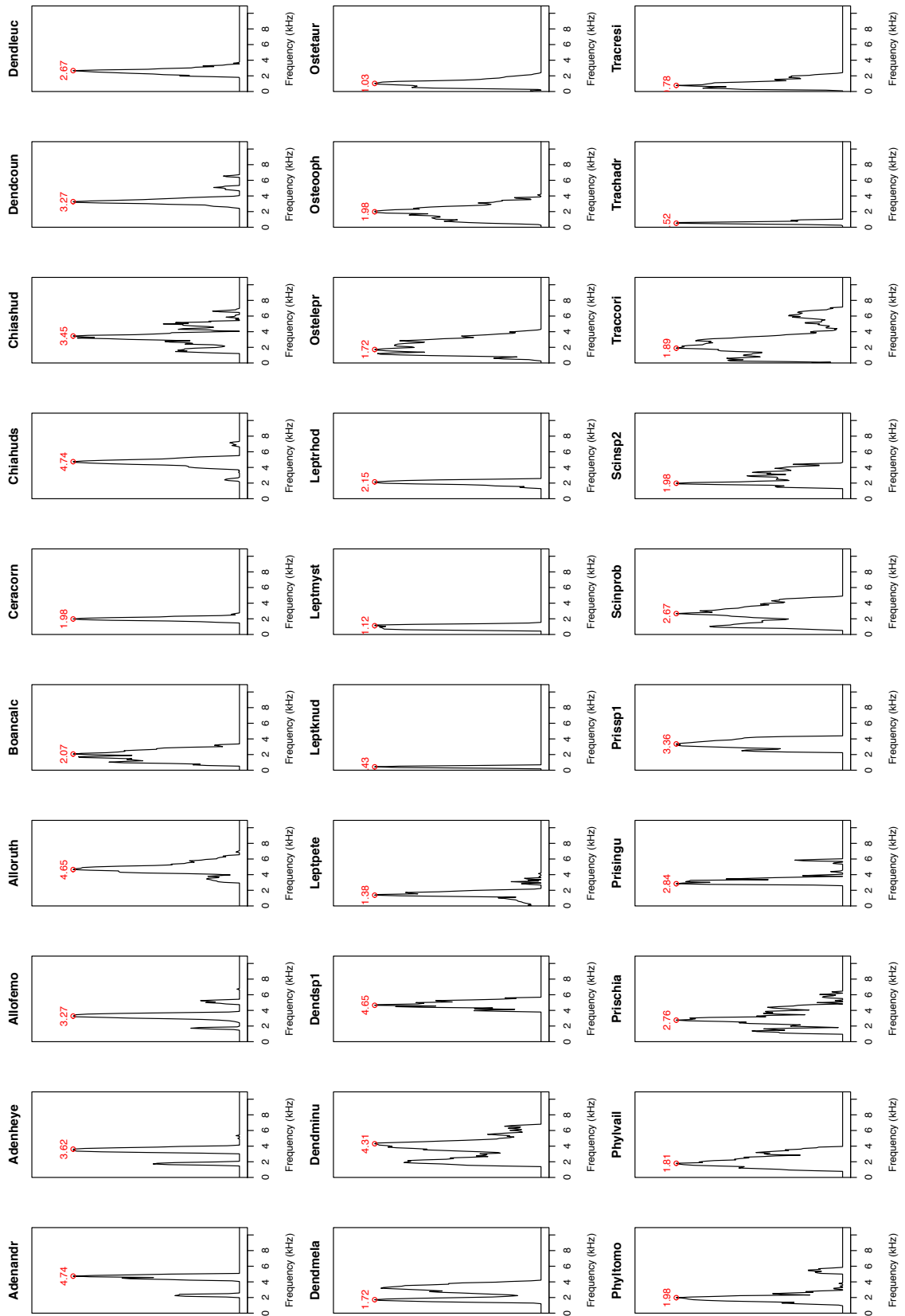


Figure S1. Acoustic features of calls from the 25 species of anuran found during explosive breeding events. The code names for the species are: *Adenomera andreae* = Adenandr, *Allobates femoralis* = Allofemo, *Allophryne ruthveni* = Alloruth, *Ceratophrys cornuta* = Ceracorn, *Chiasmocleis hudsoni* = Chiahuds, *Chiasmocleis shudikarensis* = Chiashud, *Dendropsophus counani* = Dendcoun, *Dendropsophus leucophyllatus* = Dendleuc, *Dendropsophus melanargyreus* = Dendmela, *Dendropsophus minutus* = Dendminu, *Dendropsophus sp1* = Dendsp1, *Leptodactylus knudseni* = Leptknud, *Leptodactylus mystaceus* = Leptmyst, *Leptodactylus rhodomystax* = Leprhod, *Osteocephalus leprieurii* = Ostelepr, *Osteocephalus oophagus* = Osteooph, *Phyllomedusa tomopterna* = Phyltomo, *Pristimantis chiastonotus* = Prischia, *Pristimantis inguinalis* = Prisingu, *Pristimantis sp1* = Prissp1, *Scinax proboscideus* = Scinprob, *Scinax sp2* = Scinsp2, *Trachycephalus coriaceus* = Traccori, *Trachycephalus hadroceps* = Trachadr, *Trachycephalus resinifictrix* = Tracresi

Figure S2. Spectrum of the calls of the of calls from the 25 species of anuran found during explosive breeding events. The code names for the species are given in Figure S1.



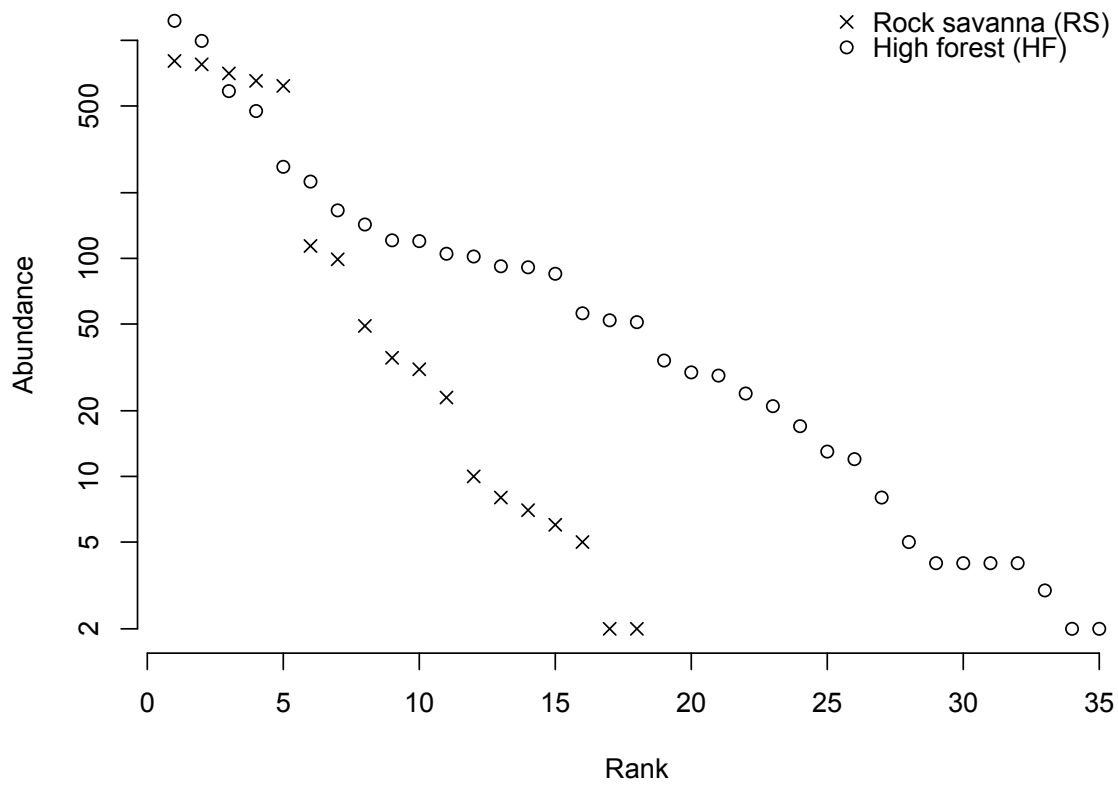
Supplementary information to Chapter 3

Figure. S1 Rank/abundance plot illustrating the relative abundance of soundtypes for the rock savanna (RS) and the high forest (HF). The y axis indicates the abundance using a logarithmic scale while the x axis ranks each soundtype in order from most to least abundant.

Table S2 List of environmental audio recordings deposited at the sound library of the Muséum national d'Histoire naturelle (www.sonotheque.mnhn.fr).

| Collection number | Site | Day | Time |
|--------------------------|-------------|-------------|-------------|
| MNHN-SO-2016-14366 | HF | 05-Dec-2014 | 22:17 |
| MNHN-SO-2016-14367 | HF | 06-Dec-2014 | 02:17 |
| MNHN-SO-2016-14368 | HF | 06-Dec-2014 | 22:18 |
| MNHN-SO-2016-14369 | HF | 07-Dec-2014 | 02:18 |
| MNHN-SO-2016-14370 | HF | 07-Dec-2014 | 22:18 |
| MNHN-SO-2016-14371 | HF | 08-Dec-2014 | 02:18 |
| MNHN-SO-2016-14372 | HF | 08-Dec-2014 | 22:18 |
| MNHN-SO-2016-14373 | HF | 09-Dec-2014 | 02:18 |
| MNHN-SO-2016-14374 | HF | 09-Dec-2014 | 22:19 |
| MNHN-SO-2016-14375 | HF | 10-Dec-2014 | 02:19 |
| MNHN-SO-2016-14376 | HF | 10-Dec-2014 | 22:19 |
| MNHN-SO-2016-14377 | HF | 11-Dec-2014 | 02:19 |
| MNHN-SO-2016-14378 | HF | 11-Dec-2014 | 22:20 |
| MNHN-SO-2016-14379 | HF | 12-Dec-2014 | 02:20 |
| MNHN-SO-2016-14380 | HF | 12-Dec-2014 | 22:20 |
| MNHN-SO-2016-14381 | HF | 13-Dec-2014 | 02:20 |
| MNHN-SO-2016-14382 | HF | 13-Dec-2014 | 22:21 |
| MNHN-SO-2016-14383 | HF | 14-Dec-2014 | 02:21 |
| MNHN-SO-2016-14384 | HF | 14-Dec-2014 | 22:21 |
| MNHN-SO-2016-14385 | HF | 15-Dec-2014 | 02:21 |
| MNHN-SO-2016-14386 | RS | 05-Dec-2014 | 22:17 |
| MNHN-SO-2016-14387 | RS | 06-Dec-2014 | 02:17 |
| MNHN-SO-2016-14388 | RS | 06-Dec-2014 | 22:18 |
| MNHN-SO-2016-14389 | RS | 07-Dec-2014 | 02:18 |
| MNHN-SO-2016-14390 | RS | 07-Dec-2014 | 22:18 |
| MNHN-SO-2016-14391 | RS | 08-Dec-2014 | 02:18 |
| MNHN-SO-2016-14392 | RS | 08-Dec-2014 | 22:18 |
| MNHN-SO-2016-14393 | RS | 09-Dec-2014 | 02:18 |
| MNHN-SO-2016-14394 | RS | 09-Dec-2014 | 22:19 |
| MNHN-SO-2016-14395 | RS | 10-Dec-2014 | 02:19 |
| MNHN-SO-2016-14396 | RS | 10-Dec-2014 | 22:19 |
| MNHN-SO-2016-14397 | RS | 11-Dec-2014 | 02:19 |
| MNHN-SO-2016-14398 | RS | 11-Dec-2014 | 22:20 |
| MNHN-SO-2016-14399 | RS | 12-Dec-2014 | 02:20 |
| MNHN-SO-2016-14400 | RS | 12-Dec-2014 | 22:20 |
| MNHN-SO-2016-14401 | RS | 13-Dec-2014 | 02:20 |
| MNHN-SO-2016-14402 | RS | 13-Dec-2014 | 22:21 |
| MNHN-SO-2016-14403 | RS | 14-Dec-2014 | 02:21 |
| MNHN-SO-2016-14404 | RS | 14-Dec-2014 | 22:21 |
| MNHN-SO-2016-14405 | RS | 15-Dec-2014 | 02:21 |

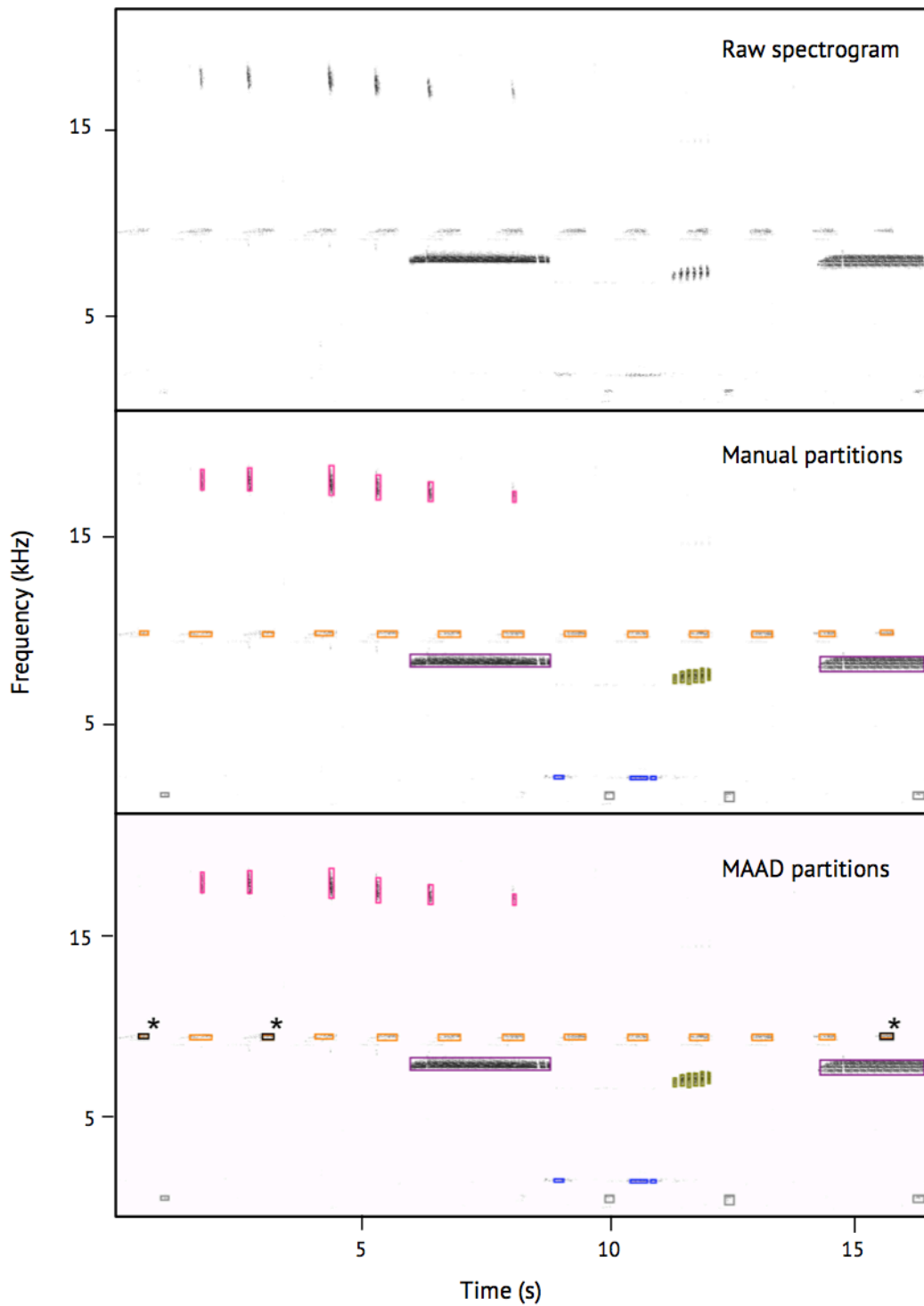


Figure. S3 Visual example of ROIs partitioned by manual analysis and by MAAD. ROIs are represented as boxes on the spectrogram and the ROIs belonging to the same group have the same colour. The asterisk symbol in the bottom spectrogram indicate a dissimilarity between manual and MAAD partitions.

Table S4a Comparison between partitions derived by manual and automated analysis for the rock savanna environment (RS).

| MAAD partition | Manual partition | | | | | | | | | | | | | | | | | |
|-------------------|------------------|----------------|----------------|----------------|----------------|----------------|-----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | A ₁ | A ₂ | A ₃ | A ₇ | A ₈ | A ₄ | A ₁₀ | A ₅ | A ₆ | A ₉ | A ₁₃ | A ₁₂ | A ₁₁ | A ₁₄ | A ₁₅ | A ₁₆ | A ₁₇ | A ₁₈ |
| B ₆ | 610 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| B ₁₀ | 0 | 776 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 |
| B ₈ | 0 | 0 | 323 | 0 | 0 | 29 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| B ₁ | 0 | 0 | 329 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 87 | 0 |
| B ₅ | 0 | 0 | 32 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 |
| B ₇ | 0 | 0 | 0 | 0 | 0 | 461 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 |
| B ₂ | 0 | 0 | 1 | 0 | 0 | 132 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| B ₄ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 773 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 |
| B ₃ | 0 | 0 | 20 | 0 | 0 | 29 | 0 | 2 | 23 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| B ₉ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 29 | 0 | 46 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 |
| B ₁₁ | 8 | 0 | 0 | 31 | 1 | 0 | 0 | 0 | 0 | 0 | 97 | 0 | 0 | 0 | 7 | 0 | 0 | 0 |

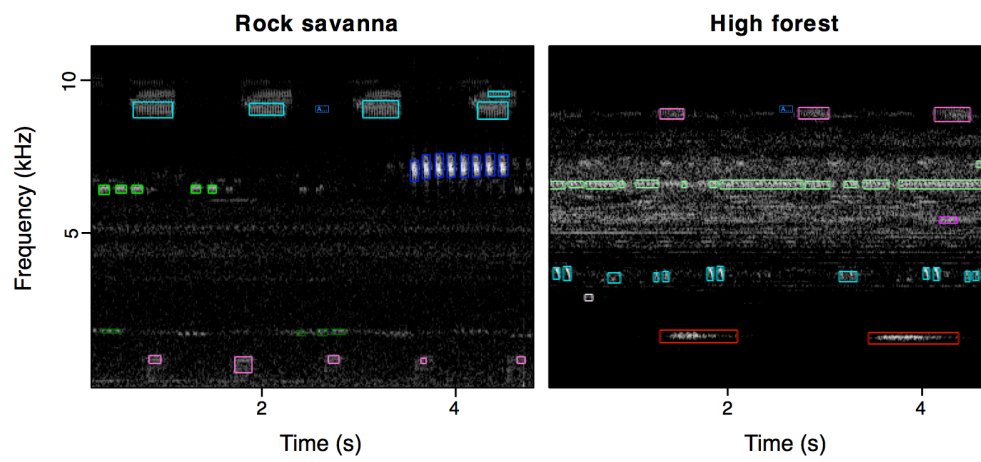


Fig. S5

Illustrative example of automatically annotated spectrograms. Some of the clusters could be linked to species sounds. On the rock savanna, cluster B₆ (pink boxes) could be identified as vocalisations of the amphibian *Hypsiboas boans*. On the high forest, cluster B₅ (cyan boxes) could be identified as stridulations of the cricket *Lernecella minuta*.

Text S6. Statistical model properties of HDDC

The mixture model-based clustering (on which HDDC is based) is defined in a probabilistic framework (Fraley & Raftery 2002): it is assumed that each group is modeled by a specific distribution $f_k(x)$, $k = 1, \dots, K$, which is often supposed to be Gaussian $f_k(x) = \phi(x; \mu_k, \Sigma_k)$ where ϕ is the Gaussian probability distribution function, μ_k its mean and Σ_k its covariance matrix. Assuming that the prior probability of the groups is such that $P(G_k) = \pi_k$, the marginal distribution of the random variable X is a Gaussian mixture model (GMM):

$$p(x) = \sum_{k=1}^K \pi_k \phi(x; \mu_k, \Sigma_k). \tag{1}$$

This modeling has two particular advantages: (1) it is known to be a robust approach to deal with unbalanced datasets, due to the presence of the parameters π_k , and (2) it is interpretable from a statistical point of view. Bouveyron *et al* (2007) proposed to constraint the GMM model through the eigen-decomposition of the covariance matrix Σ_k of the k th group:

$$\Sigma_k = Q_k \Lambda_k Q_k^t, \tag{2}$$

where Q_k is a $p \times p$ orthogonal matrix which contains the eigenvectors of Σ_k and Λ_k is a $p \times p$ diagonal matrix containing the associated eigenvalues (sorted in decreasing order). The key idea of the work of Bouveyron *et al.* is to reparametrize the matrix Λ_k , such as Σ_k has only $d_k + 1$ different eigenvalues:

$$\Delta_k = \left(\begin{array}{ccc|ccc} \boxed{\begin{matrix} a_{k1} & & 0 \\ & \ddots & \\ 0 & & a_{kd_k} \end{matrix}} & & \mathbf{0} & & & \\ & & & & & \\ \hline & & \mathbf{0} & \boxed{\begin{matrix} b_k & & 0 \\ & \ddots & \\ 0 & & b_k \end{matrix}} & & \\ & & & & & \end{array} \right) \left. \begin{array}{l} \} \\ \} \end{array} \right\} \begin{array}{l} d_k \\ (p - d_k) \end{array}$$

where the d_k first values a_{k1}, \dots, a_{kd_k} parametrize the variance in the group-specific subspace and the $p - d_k$ last terms, the b_k 's model the variance of the noise and $d_k < p$. With this parametrization, these parsimonious models assume that, conditionally to the groups, the noise variance of each cluster k is isotropic and is contained in a subspace which is orthogonal to the subspace of the k th group. Following the classical parsimony strategy, the authors proposed a family of parsimonious models from a very general model, the model $[a_{kj} b_k Q_k d_k]$, to very simple models (Table 1).

Model inference is performed using the Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977) which maximizes the likelihood by iteratively computing the conditional expectation of the complete-data likelihood and then optimizing it over the model parameters. The estimation of the intrinsic dimensions d_k , $k = 1, \dots, K$, relies on the scree test of Cattell (1966) which looks for a break in the eigenvalue scree of the empirical covariance matrix of each group. The break in the eigenvalue scree is detected as the dimension for which all differences between consecutive eigenvalues are smaller than a threshold th after this dimension. This strategy allows to find K intrinsic dimensions which can be different

Table 1: Properties of the parsimonious Gaussian mixture models implemented in HDDC

| Variances | Free dimensions | | Common dimensions | |
|-----------|----------------------|-----------------|----------------------|---------------|
| | Class specific noise | Common noise | Class specific noise | Common noise |
| Free | $a_{kj}b_kQ_kd_k$ | $a_{kj}bQ_kd_k$ | $a_{kj}b_kQ_kd$ | $a_{kj}bQ_kd$ |
| Isotropic | $a_kb_kQ_kd_k$ | $a_kbQ_kd_k$ | $a_kb_kQ_kd$ | a_kbQ_kd |
| Homosced. | $ab_kQ_kd_k$ | abQ_kd_k | ab_kQ_kd | abQ_kd |

using a common threshold value. A model selection procedure was finally implemented to estimate the hyper-parameters that control the complexity of the model. These hyper-parameters are the model M , the number of groups K , and the threshold value th to find the intrinsic dimensionality of each class.

Multiresolution Analysis of Acoustic Diversity *

Juan Sebastian Ulloa

Museum national d'Histoire naturelle

Université Paris-Saclay

February 13, 2018

The present document reports a set of basic instructions to run the Multiresolution Analysis of Acoustic Diversity (MAAD) on audio recordings. This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License <http://www.gnu.org/licenses/>. While due care has been taken and it is believed accurate, its use is solely the responsibilities of the user.

1 System requirements

The basic system requirements are:

- Matlab R2014b or later with ScatNet (v 0.2) toolbox. Scatnet toolbox can be downloaded from <http://www.di.ens.fr/data/software/scatnet/download/>
- R version 3.3.2 (2016-10-31) or later with package HDclassif (v 2.0.2)

2 Code description

To run the analysis you need to switch between two software environments. For preprocessing, detection and characterization of ROIs, and visualization you need to use a Matlab console. For the clustering step, you need to use the R console. The software is also indicated at each step by a commented line.

Open a **Matlab console**, load audio and default options for the analysis:

```
% MATLAB %  
run ./default_options.m  
s=audioread('demo.wav');
```

Transform passive acoustic recordings into the time-frequency domain using the windowed short-time Fourier transform. The Fourier coefficients are filtered to remove noise and to highlight sounds that can be delimited in time and frequency, here defined as regions of interest (ROIs):

```
% MATLAB %  
[s_filt,im,im2]=preprocess_audio(s,fs,preproc_opt,spectro_opt,ss_opt);  
[~,rois_ij]=find_rois(im2,imfilt_opt);
```

Visualize results:

```
% MATLAB %  
imshow_rois(im,rois_ij,[]);
```

Characterize ROIs with features in the time-frequency domain using 2D wavelets and the median frequency.

*Text S7. Manual for Matlab and R scripts. Supporting information for article Ulloa *et al.* Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis. *Ecological Indicators*, under review.

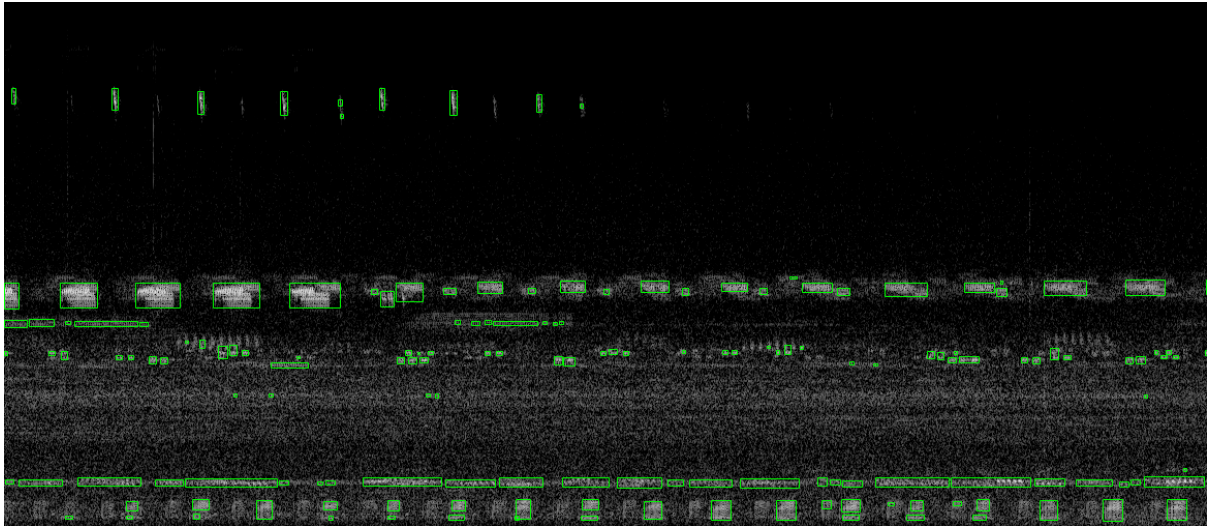


Figure 1: Example of the ROIs found in the audio recording. Spectrogram of the audio signal in gray scale colormap overlaid with ROIs. ROIs are represented as green rectangles

```
% MATLAB %
shape_features=calc_features('scatnet_op2',s,im,rois_ij,fs,filt_opt, []);
frequency_feature=calc_features('spectral_centroid',s,im,rois_ij,fs, [], ...
                                spectro_opt);
```

Organize the features in a table and save the output to a csv file. The csv file is used to transfer the data to the R software environment.

```
% MATLAB %
rois_features=table(shape_features,frequency_feature);
writetable(rois_features,'rois_features.csv','Delimiter','');
```

Cluster the ROIs into homogeneous groups based on their attributes. This step requires to open a R console and run the following commands.

```
# R #
library(HDclassif)
rois_features=read.table('rois_features.csv',sep=',',header=T)
set.seed(1234) # for repeatable example
data_hddc = hddc(rois_features,K=6,threshold=0.2,nb.rep = 10)
write.table(data_hddc$class,
            file = 'rois_group.csv',
            row.names = F,
            col.names = T)
```

Load and plot results

```
% MATLAB %
rois_group=readtable('rois_group.csv');
imshow_rois(im,rois_ij,table2array(rois_group));
```

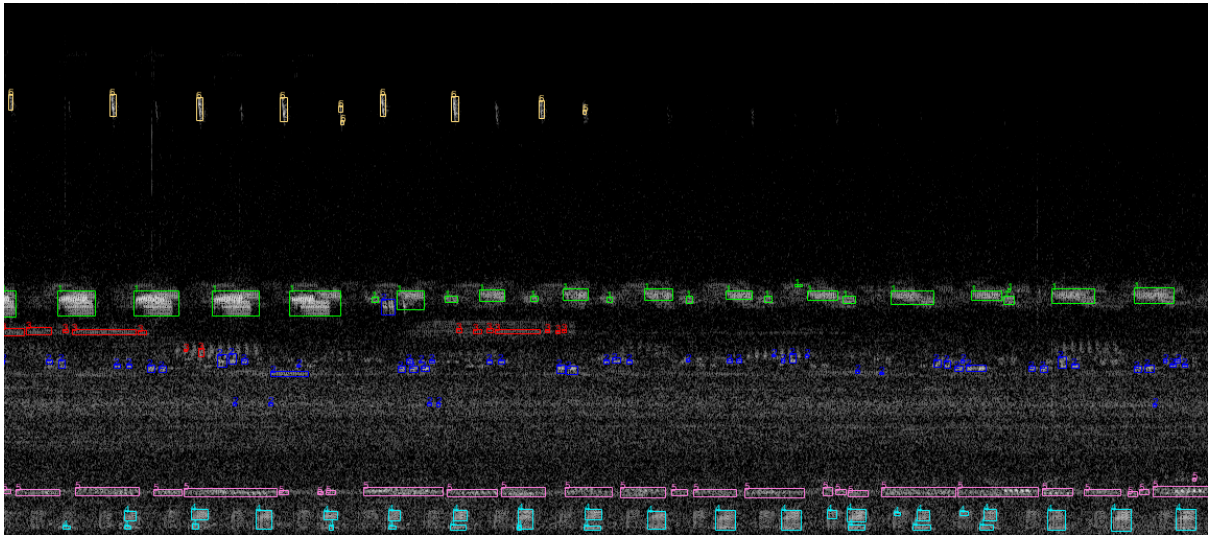


Figure 2: Example of the grouped ROIs found in the audio recording. Spectrogram of the audio signal in gray scale colormap overlaid with the ROIs. ROIs are represented as rectangles coloured according to the MAAD partitioning.

Supplementary information to Chapter 4

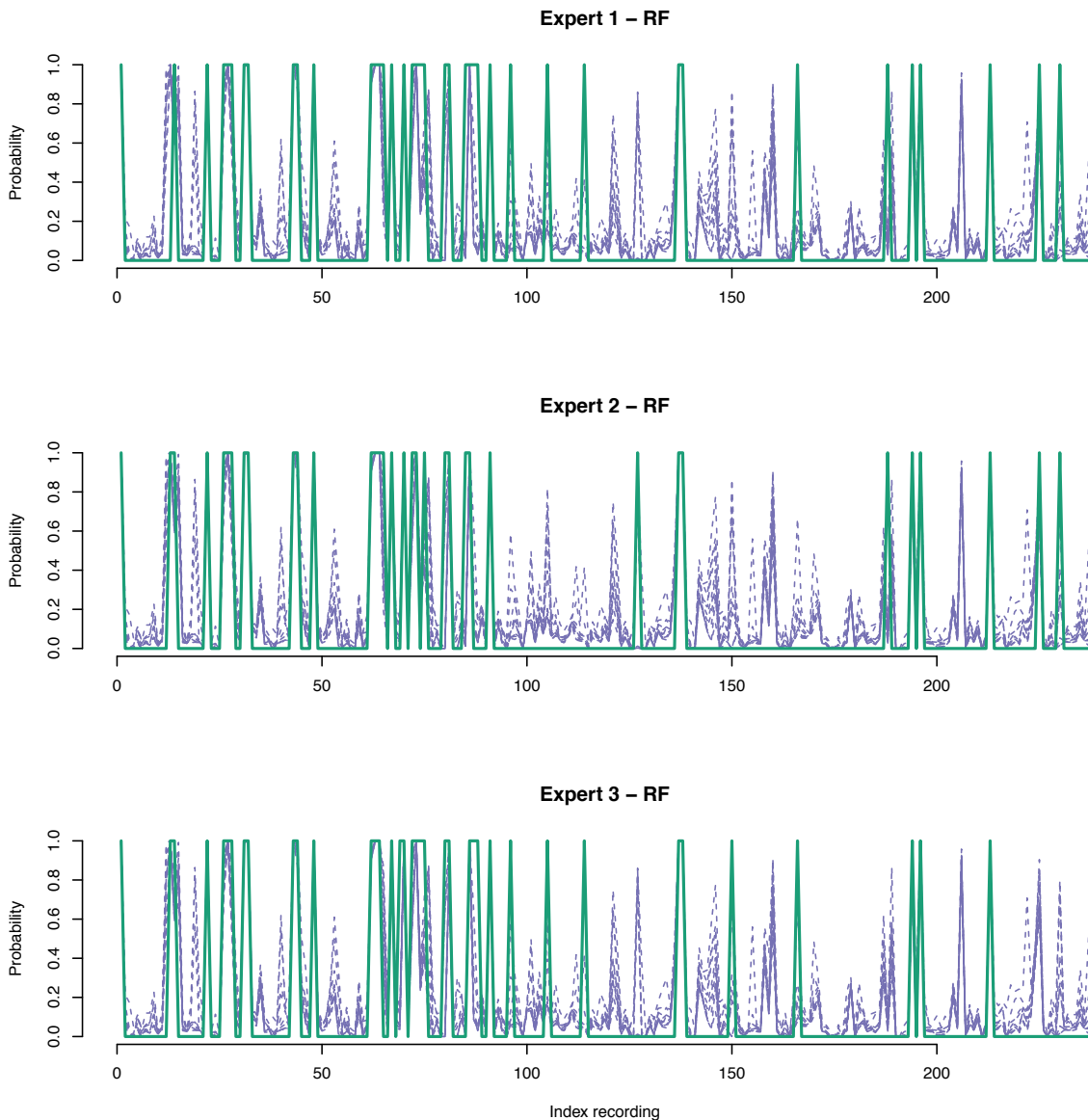


Figure S1. Comparison between manual and automatic annotation for presence absence of *Leptodactylus knudseni* in each recording of the test dataset . Presence-absence manual annotations were obtained from three different experts (green solid lines) and automatic annotations were derived from a probability response of a Random Forest (RF) statistical classifier (violet dashed lines). Since multiple regions of interest can be found in a recording, multiple RF probabilities are computed for each recording. Note that some recordings with high probabilities were not identified by all the experts, for example recording number 104, 125 and 150.

Supplementary information to General discussion



Figure S1. Right, poster of the short documentary *Paysage Sonore de la Forêt Tropicale*, a film by Juan Sebastian Ulloa and Bérénice Sevestre. The five-minute documentary participated in the film festivals: *Les Chercheurs Font Leur Cinéma*, where it was prized with the public and jury awards. It was further presented in the international film festival *Parisciences* and was part of the exposition *la Nature Monte Le Son*. The film is now freely available on the Internet. Left, poster of the 10th edition of the film festival *Les Chercheurs Font Leur Cinéma*.

Synthèse

La communication par émission sonore est un trait comportemental répandu chez les animaux terrestres, tels que les insectes, amphibiens, oiseaux et mammifères. Les riches textures sonores de la forêt néotropicale suggèrent que la faune est non seulement abondante, mais aussi diverse et dynamique. Cette facette de la biodiversité, qui peut révéler des informations précieuses sur les communautés animales habitant les milieux tropicaux, reste largement méconnue. Comment mesurer la diversité acoustique tropicale pour aborder des questions écologiques ? Dans le cadre de l'écoacoustique, nous avons cherché à révéler des structures dissimulées dans le paysage sonore de la forêt néotropicale, et tenter d'expliquer leurs présences à travers les processus écologiques sous-jacents. Cette recherche est basée sur quatre études menées dans la forêt ombrophile sempervirente de la Guyane Française.

Tout d'abord, à l'échelle de la population, nous avons suivi la dynamique spatio-temporelle d'une marque sonore amazonienne, le chant de l'oiseau tropical *Lipaugus vociferans*. À l'aide d'un réseau préalablement mis en place de 24 microphones distribués dans un espace tridimensionnel, nous avons détecté automatiquement avec des outils de traitement du signal le chant de cette espèce. Durant les 25 jours d'étude, nous avons identifié 12 735 vocalisations. Dans le domaine spatial, nous avons trouvé que l'activité vocale était moins importante dans les zones de forêt de lianes, ce qui semble indiquer que le *L. vociferans* a tendance à préférer la forêt à haute canopée. Les sites d'échantillonnage près des criques ont également présenté plus d'activité que les sites plus éloignés, ce qui suggère que les arènes de reproduction, ou leks, pourraient être distribués près d'une source d'eau. Dans le domaine temporel, nous avons identifié un cycle circadien très marqué, présentant une activité tout au long de la journée, entre le lever et le coucher du soleil, et deux pics d'activité décalés de deux heures correspondant à des chœurs matinaux et crépusculaires. Nous avons ainsi fourni des résultats détaillés, objectifs et quantitatifs sur la dynamique de cette population, trois qualités essentielles pour l'avancement des stratégies de suivi de la biodiversité.

Afin de déchiffrer le paysage sonore de la forêt tropicale, il est essentiel de considérer plusieurs échelles écologiques. Dans un deuxième temps, nous nous sommes donc intéressés aux dynamiques acoustiques au niveau de la communauté. La forêt

néotropicale est connue pour contenir la plus haute diversité d'amphibiens, dont les vocalisations constituent une marque essentielle du paysage sonore, notamment pendant la saison des pluies. En particulier, nous nous sommes intéressés aux communautés qui concentrent leur reproduction durant des périodes très brèves, c'est-à-dire les communautés composées d'espèces dites à reproduction explosive. Nous avons collecté des données acoustiques et environnementales pour suivre simultanément et régulièrement cinq mares temporaires de la montagne de Kaw pendant quatre mois. Les données acoustiques nous ont permis d'identifier dix événements à reproduction explosive, deux par site, principalement caractérisés par la cooccurrence des espèces *Chiasmocleis shudikarensis*, *Trachycephalus coriaceus* et *Ceratophrys cornuta*. Les précipitations au cours des 48 heures précédentes ont été le facteur le plus important pour prédire l'émergence de ces événements. Par ailleurs, les communautés acoustiques d'amphibiens à reproduction explosive ont montré une structure avec un fort chevauchement fréquentiel des vocalisations allant à l'encontre de l'hypothèse de partitionnement de l'espace sonore ou de niches sonores. Le chœur dense qui résulte des événements à reproduction explosive a une forte amplitude et une signature acoustique caractéristique qui pourrait annoncer la disponibilité de sites de reproduction éphémères clairsemés dans la forêt. Notre suivi acoustique a permis ainsi de mieux comprendre les causes, structures et conséquences de ce comportement collectif particulier. Puisque les amphibiens constituent le groupe de vertébrés le plus menacé, les nouvelles techniques de suivis basées sur des capteurs acoustiques semblent pertinentes, permettant de mieux les comprendre et de prévoir leur réponse aux changements globaux.

Les capteurs acoustiques automatiques permettent de suivre les dynamiques des communautés animales et leur utilisation peut être envisagée à de grandes échelles spatiales et temporelles. À ce jour, des réseaux d'enregistreurs acoustiques sont en effet déployés dans le monde entier. Cependant pour pouvoir gérer les grandes quantités de données collectées, l'analyse standardisée et efficace des enregistrements devient une nécessité. Dans un troisième temps, nous avons adapté des outils de calcul puissants, issus des disciplines de l'apprentissage automatique et de la reconnaissance de formes, pour proposer des nouvelles façons de quantifier et estimer la diversité acoustique qui émerge de la forêt tropicale. Nous avons développé une nouvelle méthode pour caractériser et automatiser la détection de structures dans le paysage sonore, nommée *Multiresolution Analysis of Acoustic Diversity* (MAAD). Cette analyse a été conçue

pour décomposer la communauté acoustique en plusieurs constituants élémentaires. MAAD repose sur des techniques d'apprentissage non supervisé qui permettent de classifier les sons en groupes homogènes sans phase préalable d'annotation manuelle et fastidieuse, ouvrant ainsi de nouvelles perspectives d'analyse pour l'écoacoustique.

Enfin, dans un quatrième temps, nous avons évalué plusieurs techniques pour annoter la présence ou absence de sons d'intérêt dans des enregistrements sonores. Nous avons comparé l'annotation manuelle, l'apprentissage semi-supervisé et l'apprentissage supervisé pour révéler les forces et les faiblesses de chaque approche, et ainsi mieux saisir comment combiner à la fois le raisonnement humain et les algorithmes informatiques. Nous avons utilisé 528 enregistrements qui contiennent un environnement acoustique tropical hétérogène et nous avons cherché à identifier la présence d'une espèce d'amphibien *Leptodactylus pentadactylus*. Nous montrons que les annotations manuelles faites par des experts sont convergentes, mais présentent néanmoins des variations individuelles non-négligeables. L'approche semi-supervisée nécessite très peu d'effort manuel, mais produit des résultats moins précis. L'apprentissage supervisé, tout en nécessitant plus d'annotations manuelles pour estimer les paramètres du modèle, a montré des résultats ayant un bon consensus avec l'étiquetage des experts. Nous soulignons que des solutions visant à combiner l'analyse manuelle et automatisée dans un protocole d'annotation bien encadré devraient être envisagées pour fournir des estimations robustes et précises de présence-absence de sons d'intérêt. Un tel cadre est fondamental pour évaluer les habitats mal documentés, où les sons animaux doivent encore être décrits, annotés et enregistrés dans des sonothèques.

Les environnements tropicaux, très diversifiés dans leur structure et complexes dans leur fonctionnement, subissent de plus en plus la pression humaine. Ces environnements uniques imposent des défis majeurs en écologie et biologie de la conservation. L'écoacoustique, associée à l'informatique, émerge comme une nouvelle approche clé permettant de renforcer les programmes de suivis de biodiversité à large échelle et ainsi de mieux comprendre et valoriser la diversité de formes de vies unique abritée par la forêt tropicale.

DETAILED TABLE OF CONTENTS

| | |
|--|-----------|
| ACKNOWLEDGMENTS | 6 |
| GENERAL INTRODUCTION | 9 |
| 1. Sound, a window to the animal community | 10 |
| 2. Tropical ecology and conservation | 11 |
| 3. Biodiversity assessment | 13 |
| 3.1. Human-based surveys | 14 |
| 3.2. Remote ex situ sensing | 14 |
| 3.3. In situ biodiversity sensing | 15 |
| 4. Acoustic monitoring | 15 |
| 5. Analysis of passive audio recordings | 17 |
| 5.1. Manual analysis | 17 |
| 5.2. Automated analysis | 18 |
| 5.3. Acoustic indices | 20 |
| 6. Working hypotheses | 21 |
| References | 24 |
| CHAPTER 1. Screening large audio datasets to determine the time and space distribution of Screaming Piha birds in a tropical forest. | 32 |
| Overview | 33 |
| 1. Introduction | 34 |
| 2. Materials | 37 |
| 2.1. Study site | 37 |
| 2.2. Acoustic sampling protocol | 37 |
| 2.3. Study species | 39 |
| 2.4. Canopy height detection | 39 |
| 3. Methods | 41 |
| 3.1. Test database | 41 |
| 3.2. Signal Processing | 42 |
| 3.3. Template selection | 44 |
| 3.4. Tuning the spectrogram parameters | 44 |
| 3.5. Threshold setting | 46 |
| 4. Results | 47 |

| | |
|---|-----------|
| 5. Discussion | 51 |
| 5.1. Template matching in complex soundscapes..... | 51 |
| 5.2. Remote-sensing, behaviour and ecology of the Piha..... | 52 |
| 6. Conclusion..... | 54 |
| Acknowledgments | 55 |
| Supplementary Information..... | 55 |
| References | 56 |
| CHAPTER 2. Explosive breeding in tropical anurans: environmental triggers, community composition and acoustic structure..... | 62 |
| Overview | 63 |
| 1. Introduction | 64 |
| 2. Materials and methods | 66 |
| 2.1. Study site..... | 66 |
| 2.2. Sampling protocol..... | 68 |
| 2.3. Time-series analysis..... | 68 |
| 2.4. Community diversity analysis..... | 69 |
| 2.5. Acoustic diversity analysis | 70 |
| 3. Results | 72 |
| 3.1. Time series analysis | 72 |
| 3.2. Community diversity analysis..... | 74 |
| 3.3. Acoustic diversity analysis | 77 |
| 4. Discussion | 81 |
| 4.1. Time series analysis | 81 |
| 4.2. Acoustic community diversity analyses..... | 82 |
| 4.3. Acoustic diversity analyses..... | 83 |
| References | 86 |
| CHAPTER 3. Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis | 93 |
| Overview | 94 |
| 1. Introduction | 95 |
| 2. Materials and methods | 98 |
| 2.1. Audio dataset | 99 |
| 2.2. Detection of regions of interest (ROIs)..... | 99 |
| 2.3. Characterization of ROIs | 100 |

| | |
|---|------------|
| 2.4. High dimensional clustering | 102 |
| 2.5. Validation of system performance | 104 |
| 3. Results | 105 |
| 3.1. Model selection | 105 |
| 3.2. Feature relevance | 107 |
| 3.3. Clustering results | 110 |
| 4. Discussion | 111 |
| 4.1. Model selection | 112 |
| 4.2. Feature relevance | 112 |
| 4.3. Clustering results | 113 |
| Acknowledgements | 116 |
| Data accessibility | 116 |
| References | 117 |
| CHAPTER 4. Towards an end-to-end framework for sound classification in passive acoustic recordings | 123 |
| Overview | 124 |
| 1. Introduction | 125 |
| 2. Methods | 128 |
| 2.1. Acoustic environment and signal of interest | 129 |
| 2.2. Analysis of acoustic diversity | 130 |
| 2.3. Sampling strategies | 133 |
| 2.4. Presence-absence labelling | 134 |
| 3. Results | 135 |
| 3.1. Sampling strategies | 135 |
| 3.2. Determining presence-absence data | 137 |
| 4. Discussion | 139 |
| 4.1. Sampling strategies | 139 |
| 4.2. Determining presence-absence data | 140 |
| Acknowledgments | 143 |
| References | 144 |
| GENERAL DISCUSSION | 149 |
| 1. Analysis of field audio recordings | 150 |
| 1.1. Manual annotations | 151 |
| 1.2. Computer assisted classification | 153 |

| | |
|--|------------|
| 1.3. Global characterisation of the soundscape..... | 155 |
| 2. Patterns and dynamics of the tropical acoustic environment | 157 |
| 3. Acoustic surveys to monitor biodiversity at large scales | 159 |
| 4. Rainforest soundscapes to promote biodiversity conservation | 160 |
| References | 163 |
| APPENDIX | 170 |
| Related publications and oral communications | 171 |
| Publications..... | 171 |
| Oral communications | 171 |
| Supplementary information to Chapter 1 | 173 |
| Supplementary information to Chapter 2 | 174 |
| Supplementary information to Chapter 3 | 176 |
| Supplementary information to Chapter 4 | 187 |
| Supplementary information to General discussion..... | 188 |
| Synthèse..... | 189 |

Titre : Estimation de la diversité acoustique animale en forêt néotropicale

Mots clés : suivi de biodiversité, écoacoustique, apprentissage automatique, écologie tropicale

Résumé : La communication par émission sonore est un trait comportemental répandu chez les animaux terrestres. Les riches textures sonores de la forêt néotropicale nous suggèrent que la faune est non seulement abondante, mais aussi diverse et dynamique. Cette facette de la biodiversité peut révéler des informations précieuses sur les communautés animales qui habitent les milieux tropicaux, mais reste largement méconnue. Comment mesurer la diversité acoustique tropicale pour aborder des questions écologiques ? Dans le cadre de l'écoacoustique, nous avons cherché à révéler des structures dissimulées dans le paysage sonore de la forêt néotropicale, et tenter d'expliquer leurs présences à travers les processus écologiques sous-jacents. Tout d'abord, nous avons suivi la dynamique spatio-temporelle d'une empreinte sonore amazonienne, le chant de l'oiseau tropical

Lipaugus vociferans, montrant une activité liée à des caractéristiques spécifiques d'habitat. Puis, nous nous sommes intéressés aux communautés d'amphibiens. L'analyse de variables acoustiques et météorologiques nous a permis de mieux comprendre les causes, patrons et conséquences du comportement reproductif explosif. Enfin, nous avons adapté de nouveaux outils de calcul, issus des disciplines de l'apprentissage automatique et de la reconnaissance de formes, pour proposer une analyse efficace, objective et facilement reproductible de grands jeux de données acoustiques. L'écoacoustique, renforcée par des algorithmes informatiques, émerge comme une approche clé pour les programmes de suivis de biodiversité à large échelle, permettant de mieux comprendre et valoriser la diversité de formes de vies unique abritée par la forêt tropicale.

Title : Estimating animal acoustic diversity in neotropical forest

Keywords : biodiversity monitoring, ecoacoustics, machine learning, tropical ecology

Abstract : Acoustic signalling is a common behavioural trait among terrestrial animals. The rich sound textures of neotropical forest echo that wildlife is not only abundant, but also diverse and dynamic. This facet of biodiversity can reveal valuable insights of animal communities inhabiting tropical environments, yet remains poorly understood. How to best measure tropical acoustic diversity to address ecological questions? Based on the ecoacoustic framework, we explored the soundscape of neotropical forest, revealing patterns and investigating the ecological underlying processes. First, we tracked the spatiotemporal dynamics of an amazonian soundmark, the song of the bird *Lipaugus vociferans*, showing

activity patterns related to specific habitat features. Then, we investigated amphibian communities with very brief reproduction periods. Coupling acoustic and environmental variables, we shed light on the causes, patterns and consequences of explosive breeding events. Finally, we adapted novel computational tools from the machine learning and pattern recognition disciplines to provide an efficient, objective and replicable analysis of large acoustic datasets. Ecoacoustics, powered with computer algorithms, emerge as a suitable approach to scale-up biodiversity monitoring programs, allowing to better understand and cherish the unique diversity of life sustained by tropical forest.

