

Politechnic School of Engineering

Department of Mechanic Engineering

AUTOMATIC ACOUSTIC IDENTIFIER FOR BRAZILIAN BIRD SPECIES

Monograph

Athos P. Damiani

Prof. PhD. Paulo do C. Hubert Jr

Prof. PhD. Linilson Padovese

University of São Paulo
São Paulo, Brazil

May 2020

Abstract

Recently, studies on bioacoustics have been used in several areas of Biology, from the Systematic, through Ethology, to the Conservation of Biodiversity. However, the potential of applying these tools in the study of endangered species and night fauna remains severely unexplored. This project aims to develop tools and a methodology for detection and analysis for night bird presence/absence data, in particular Strigidae family, through the development of acoustic detectors using machines learning, to apply for both large databases of acoustic landscapes scanning and embark in monitoring equipment installed in the field, developed by the Acoustics and Environment of Poli-USP. Several supervised learning approaches, such as convolutional neural networks and gradient boosting machines.

keywords: bioacoustics, ecology acoustics, landscape monitoring, signal processing, biodiversity conservation, machine learning, automatic species recognition, neural networks, gradient boosting machine, strigidae, owls

Contents

List of Tables	ii
List of Figures	iii
Literature Review	1
Chapter 1 Introduction	2
1.1 Notes	3
2 Objective	4
Chapter 3 The Data	5
3.1 Training/Testing data	5
3.2 Audio preparation	5
3.3 Wave file slicing	6
3.4 Data Labelling	6
3.5 Data Annotation	8
Chapter 4 Modelling	9
5 Results	10
6 Conclusion	11
References	12

List of Tables

3.1 MP3 files downloaded.	5
-----------------------------------	---

List of Figures

3.1 Screen shot of the wavesurfer::annotator_app()	8
--	---

Literature Review

To do.

Chapter 1 Introduction

Biodiversity analysis campaigns, mainly related to insects, birds, mammals and anurans, have traditionally been carried out in a “manual” manner, through the use of specialists in loco, using hearing and vision. These campaigns are time consuming, expensive, largely dependent on the experience and knowledge of each specialist, and limited in time and space. The presence or absence of a species in a certain site depends on the coincidence of both the species and the specialist in the same time and space.

In order to provide tools to increase the capacity for studying and analyzing biodiversity, both in terrestrial and underwater biomes, the concept of monitoring soundscapes has been used. For this, ambient acoustic recording equipment is used which, through microphones and hydrophones, capture environmental audio whether in the infrasound, audible or ultrasound bands. These devices can be autonomous in the sense that they have internal energy and storage capacity that allow them to work for a few days or several months. But they can also be permanent in the sense that they have no energy restrictions (either through the use of solar panels, or from another external energy source), with eventual transmission of information in real time. The generated database contains information that enables a great diversity of studies and analysis, such as: animal communication and behavior, landscape ecology and conservation biology, research in population dynamics (Campos-Cerqueira, 2016), communities monitoring (Dawson and Efford, 2009), detection of endemic, threatened or even unknown species (Vernier, 2012) and in studies regarding migration (Salamon et al, 2016).

Among these, the development of bird sound classification systems has been widely debated and reviewed in the literature, and some references consulted are listed at the end of the proposal. Several problems remain open, with the main one, perhaps, being about find ways of dealing with the immense variability of vocalizations and ambient acoustic conditions. Annually, transdisciplinary teams composed of researchers from different international organizations are mobilized to propose solutions to this type of challenge in an event called BirdCLEF¹, which is an important source of information about this subject.

Lacmam (Laboratory of Acoustics and Environment of EP-USP) has several years of

¹<https://www.imageclef.org/node/230>

experience in this activity, either in the development of its own hardware and software technology, or in conducting long-term campaigns for monitoring and analyzing soundscapes, both terrestrial and underwater (Figure 1 and Figure 2), as detailed in Methodology Session.

In this context, this project aims to take an additional step in innovation by developing technology for automatic detection of certain acoustic events. Of particular interest is the detection of some species of Strigiformes (owls), incorporating machine learning methods in the soundscape monitoring technology package already developed by the laboratory, in favor of the conservation of species and ecosystems. In the context of the species detection problem, there is a particularly challenging issue: detection of night birds due to the specialist's limited sight and locomotion caused by the biome.

On the other hand, 24 hours persistent acoustic monitoring equipments has relatively low cost and can be cover a wide region. Therefore an installation of automated detection software on these equipments, or using these softwares for analysis of large acoustic databases, enables an efficient and effective method/permanent monitoring in large areas at the same time that it provides quality information on ecological behavior.

Although the proposal's emphasis is on the automated recognition of some given bird species, the methodology to be developed may also be applied to other species of birds, anurans and, even in underwater fauna, such as marine mammals and chorus fish, among others.

This project was inspired by needs pointed out by managers from Federal and State Conservation Unities as well as by Environment Conservation NGOs. These are concrete demands to which we hope to contribute.

1.1 Notes

“There is the need for shared datasets with annotations of a wide variety of calls for a large number of species if methods that are suitable for conservation work are to be developed.” — Automated birdsong recognition in complex acoustic environments: a review

2 Objective

To do.

Chapter 3 The Data

3.1 Training/Testing data

Two main data sources with almost ready to use labels were gathered:

- Xeno-Canto.org
- Wikiaves.com.br

Those websites are the biggest repository of Brazilian bird species sounds. The table [3.1](#) show the counts of MP3 downloaded by species.

Table 3.1: MP3 files downloaded.

Species	#mp3
Megascops-choliba	803
Pulsatrix-coeniaswaldiana	334
Strix-hylophila	318
Megascops-atricapilla	258
Glaucidium-minutissimum	208
Total	1921

No providence was made if the same audio happened to appear at both sources.

The {warbleR} (Araya-Salas and Smith-Vidaurre [2017](#)) and {wikiaves} (Damiani [2020](#)) packages were used to retrieve all the mp3 available automatically. A tutorial to reproduce the extraction can be found in this link <https://athospd.github.io/mestrado/articles/data-gathering.html>.

3.2 Audio preparation

All the audios downloaded from Xeno-Canto and Wikiaves comes as mp3 files. Also, they can be stereo or mono, in potentially any sample rate and unnormalized. The first step of the data preparation standardize these properties.

Steps to apply to every mp3 file:

1. convert MP3 to WAV;
2. stereo to mono;
3. downsample to 16kHz;
4. normalize amplitude by rescaling to integers in $[-32767, 32767]$ (i.e. 16-bit).

3.3 Wave file slicing

The observations of the training set will be slices with a given length of the raw wave files. For instance, if an original wave file has duration of 55 seconds, then the slicing with interval of 1 second and no overlap will result in 55 disjoint 1 second long slices.

Table 3.2 shows the layout of the sliced dataset: the original audio file name as id, the start point of the slice and the end point of the slice. The final slice dataset is built by stacking the slices of all the files.

Table 3.2: Example of a dataset of slices from Glaucidium-minutissimum-1066225.wav audio file.

id	start	end
Glaucidium-minutissimum-1066225	0	1
Glaucidium-minutissimum-1066225	1	2
Glaucidium-minutissimum-1066225	2	3

The experiment tested slicing with 1 second long and with 2 seconds long, with no overlap. Table 3.3 shows the total counts of slices generated after the slicing process. There is a trade-off behind the window length choice. The smaller the interval, the more samples will be created. On the other hand, less information will fit in the frame causing issues for the learner. So there are a sweet spot to be discovered and this is the reason to study both 1 and 2 seconds frame datasets.

Table 3.3: Slices.

Species	#mp3	#1sec slices	#2sec slices
Megascops-choliba	803	30115	15259
Pulsatrix-coenobrya	334	14904	7529
Strix-hylophila	318	10812	5485
Megascops-atricapilla	258	10938	5536
Glaucidium-minutissimum	208	7339	3725
Total	1921	74108	37534

3.4 Data Labelling

Each slice can contain parts of bird songs in it or not. In this section, the annotation part is supposed to be already done. The annotation task is discussed in section [Data Annotation](#).

The table [3.4](#) show an example of the tables that were generated to store where the bird songs happened in each wave file and the respective bird species that made it. At the end of the labelling process the dataset had six distinct labels: five from bird species and one for “other events than bird songs”. This later label was called as “unknown” and these other, unknown events could be silence, noise, any other animal, or even other birds different from the five ones previous picked.

Table 3.4: The first four rows of the table storing the moments where the bird songs happen inside the *Glaucidium-minutissimum*-1066225.wav audio file along with the respective bird species label.

audio_id	region_id	start	end	label
Glauc...225.wav k6109vbm8ro		0.21	1.19	Glauc...mum
Glauc...225.wav upsuth6aang		6.10	7.40	Glauc...mum
Glauc...225.wav cgqlbd1m4r		17.22	18.36	Glauc...mum
Glauc...225.wav p97kav4lem8		23.90	24.88	Glauc...mum
Glauc...225.wav ft0i9775ejg		30.36	31.29	Glauc...mum

After build one of such tables for every wave file, a final dataset with all the slices labeled was made, one for one second long slices and other for two second long. Those tables have 74108 and 37534 rows respectively, as shown in Table [3.3](#). This dataset was made by joining [3.2](#) and [3.4](#) by interval. The slices id were conviniently named as

{bird-species}@{start}@{end}@.wav

so one can retrieve where this slices lays in the original wave file by separating the parts by @. For instance, Table [3.5](#) show that a *Glaucidium minutissimum* sang between the seconds 0 and 1, and also between the seconds 0 and 2 of the *Glaucidium-minutissimum*-1066225.wav audio file. The seconds 10 to 12 there was no bird song events of interest, so it was labeled as “unknown”.

Table
3.5:
The
first
four
rows
of
the
table

of all 1 second slices along with the bird species label.

audio_id	slice_id	label
Glaucidium-minutissimum-1066225.wav	<u>Glaucidium-minutissimum-1066225@0@1@.wav</u>	Glaucidium-minutissimum
Glaucidium-minutissimum-1066225.wav	<u>Glaucidium-minutissimum-1066225@1@2@.wav</u>	Glaucidium-minutissimum
Glaucidium-minutissimum-1066225.wav	<u>Glaucidium-minutissimum-1066225@10@11@.wav</u>	unknown
Glaucidium-minutissimum-1066225.wav	<u>Glaucidium-minutissimum-1066225@11@12@.wav</u>	unknown
Glaucidium-minutissimum-1066225.wav	<u>Glaucidium-minutissimum-1066225@12@13@.wav</u>	unknown

3.5 Data Annotation

The software used to annotate the audios was the {wavesurfer} R package (Damiani and Katspaugh 2020)². The parts containing bird songs were manually annotated by the humans.

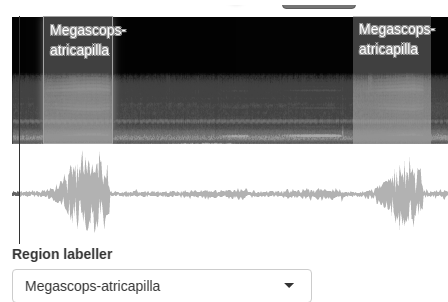


Figure 3.1: Screen shot of the wavesurfer::annotator_app().

Each one of the wave files were annotated manually by visual and by hearing. The software stored a dataset with the annotation information as shown in Table 3.4. Wave files with pure silence or damaged were discarded from the analysis.

²The {wavesurfer} annotator app is open-source and was developed by the authors of this monograph.

Chapter 4 Modelling

5 Results

To do.

6 Conclusion

To do.

References

- Araya-Salas, M., and G. Smith-Vidaurre. 2017. "WarbleR: An R Package to Streamline Analysis of Animal Acoustic Signals." *Methods in Ecology and Evolution*. <http://dx.doi.org/10.1111/2041-210X.12624>.
- Damiani, Athos. 2020. *Wikiaves: API to Retrieve Sounds from Wikiaves.com.br*.
- Damiani, Athos, and Katspaugh. 2020. *Wavesurfer: Create Interactive Sound Player and Visualizer in R*. <https://github.com/athospd/wavesurfer/issues>.

