UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Seecology: Data Visualization Framework for Soundscape Ecology Applications

Clausius Duque Gonçalves Reis

Tese de Doutorado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Clausius Duque Gonçalves Reis

Seecology: Data Visualization Framework for Soundscape Ecology Applications

Thesis submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP – in accordance with the requirements of the Computer and Mathematical Sciences Graduate Program, for the degree of Doctor in Science. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Profa. Dra. Maria Cristina Ferreira de Oliveira

USP – São Carlos
July 2020

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

D375s Duque Gonçalves Reis, Clausius
Seecology: Data Visualization Framework for
Soundscape Ecology Applications / Clausius Duque
Gonçalves Reis; orientadora Maria Cristina Ferreira
de Oliveira. -- São Carlos, 2020.
110 p.

Tese (Doutorado - Programa de Pós-Graduação em
Ciências de Computação e Matemática Computacional) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2020.

1. Acoustic Ecology. 2. Soundscape. 3.
Visualization. 4. Feature Extraction. 5. Boat
Detection. I. Cristina Ferreira de Oliveira, Maria,
orient. II. Título.

Clausius Duque Gonçalves Reis

**Seecology: Um Framework de Visualização de Dados para
Aplicações em Ecologia Acústica**

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientadora: Profa. Dra. Maria Cristina Ferreira de Oliveira

**USP – São Carlos
Julho de 2020**

Dedico este trabalho aqueles que acreditando em mim tornaram possível vencer mais este desafio. Cujo apoio fortalece cada passo meu. Aos meus alicerces nesta jornada, meu pai, minha mãe e minha esposa.

I dedicate this work to those who, believing in me, made it possible to overcome this challenge. Whose support strengthens my every step. To my foundations on this journey, my father, my mother and my wife.

ACKNOWLEDGEMENTS

It is difficult to put the path taken during a doctoral research on one page, to remember everyone who was with us during the sleepless nights, difficult hours and moments of pleasure. Friends and family who understood that our absence had a purpose and cheered for our success.

I thank my parents, Efraim and Lúcia, for all the encouragement and reprimands, never giving up on the desire to see us grow and fly with our own wings, that with so much affection were nourished by years of dedication, with love and wisdom.

My wife Rachel, for supporting me in this task, dragging me out of my comfort zone, always resulting in funny stories and good memories. This beast always wants to be next to its beauty, I love you!

My advisor, Professor Maria Cristina, for the opportunity to carry out this work, for the lessons and example of professionalism. I thank you for making me see further, supported by the shoulders of giants.

To professors Bryan Pijanowsky and Linilson Padovese for providing the data and know-how necessary to conduct this research.

To friends Kamila and Diego who accompanied me in these hard years, sharing many laughs and moments of relaxation.

To CEFER friends, Gisele, Adilson and so many others who made me discover my sporty and partying side.

To VICG colleagues for the company and discussions and to all those who, even though they were not mentioned, were essential for my personal and professional growth.

To the Federal University of Viçosa, FAPESP and CNPQ for financial support.

É difícil colocar o caminho percorrido durante uma pesquisa de doutorado em uma página, lembrar de todos que estiveram conosco nas noites mal dormidas, horas de sufoco e momentos de prazer. Dos amigos e familiares que entenderam que nossa ausência tinha um propósito e torceram pelo nosso sucesso.

Agradeço aos meus pais, Efraim e Lúcia, por todo o incentivo e puxões de orelha, nunca desistindo da vontade de nos ver crescer e voar com asas próprias que com tanto carinho foram nutridas por anos de dedicação com amor e sabedoria.

A minha esposa Rachel, por me dar todo o apoio em mais essa etapa, me arrastando para fora da minha zona de conforto, sempre resultando em engraçadas histórias e boas lembranças. Esta fera quer sempre ficar ao lado de sua bela, te amo!

A minha orientadora, Professora Maria Cristina, pela oportunidade de realizar este trabalho, pelas lições e exemplo de profissionalismo. Agradeço por me fazer ver mais longe, apoiado pelos ombros de gigantes.

Aos professores Bryan Pijanowsky e Linilson Padovese por fornecerem os dados e know-how necessários para a condução dessa pesquisa.

Aos amigos Kamila e Diego que me acompanharam nesses duros anos, compartilhando de muitas risadas e momentos de descontração.

Aos amigos do CEFER, Gisele (Parça de São Pedro), Adilson (O homem da coxinha) e tantos outros que me fizeram descobrir meu lado esportivo e festeiro.

Aos colegas do VICG pela companhia e discussões e a todos aqueles que mesmo não sendo citados foram essenciais para o meu crescimento pessoal e profissional.

À Universidade Federal de Viçosa, FAPESP e CNPQ pelo apoio financeiro.

“Se eu vi mais longe, foi por estar sobre ombros de gigantes.”
(Isaac Newton, 1675)

RESUMO

REIS, C. D. G. **Seecology: Um Framework de Visualização de Dados para Aplicações em Ecologia Acústica**. 2020. 110 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

A área de Ecologia de Paisagens Sonoras (*Soundscape Ecology*) refere-se ao estudo de sons produzidos em ambientes naturais e como eles podem fornecer informações importantes sobre o meio ambiente, bem como possíveis impactos causados por alterações devido a influências externas. A análise e visualização de grandes quantidades de gravações ecológicas, juntamente com a produção de recursos para análises constituem um grande desafio. Meios para a extração de características dos áudios, bem como a caracterização de eventos acústicos de interesse, produzindo conjuntos de dados que representem as variações de frequências e eventos acústicos capturados nas gravações, ainda constituem um problema devido às soluções disponíveis não se mostrarem adequadas para análises de dados em pesquisas de ecologia acústica, envolvendo questões específicas do domínio e quantidades volumosas de registros de áudio coletados por longos períodos de tempo. Faz-se necessário o desenvolvimento de métodos e ferramental para extrair e representar a grande quantidade dos dados produzidos a partir de estudos de ecologia. Este trabalho tem por objetivo abordar problemas relacionados à extração de características dos áudios, auxiliando na seleção das mais significativas que representem as sutis variações nas gravações ecológicas, bem como auxiliar especialistas na geração de conjuntos de dados anotados pela caracterização de eventos acústicos por meio de visualizações exploratórias, e métodos para um problema específico, que é a detecção de embarcações em gravações subaquáticas. Um arcabouço nomeado Seecology é apresentado, englobando métodos e ferramentas adequados para dar suporte aos especialistas e estudiosos de análise ambiental. Estudos de caso foram realizados com o arcabouço em gravações terrestres e subaquáticas fornecidos por pesquisadores da área, produzindo conjuntos de dados a partir do extrator de características personalizado incluso no arcabouço. No caso do método desenvolvido para detecção de barcos em gravações subaquáticas, um estudo comparativo a outro método foi conduzido para determinar sua acurácia, além do estudo de caso para determinar sua eficácia. Os métodos propostos para extração de características, caracterização de eventos acústicos por meio de visualização exploratória e detecção de barcos, demonstraram sua eficácia para aplicações em ecologia acústica, sendo o arcabouço capaz de produzir conjuntos de dados multidimensionais sem custos computacionais excessivos. Dessa forma o usuário é capaz de gerar anotações nestes dados facilmente por meio das visualizações inclusas. O método de detecção de barcos obteve desempenho superior ao que foi comparado, tanto em velocidade quanto em acurácia, sendo capaz de detectar sinais fracos de barcos mesmo sob ruído extremo.

Palavras-chave: Ecologia Acústica, Soundscape, Framework, Visualização, Extração de características, Análise de Áudio, Detecção de barcos, Radviz.

ABSTRACT

REIS, C. D. G. **Seecology: Data Visualization Framework for Soundscape Ecology Applications**. 2020. 110 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

The field of Soundscape Ecology refers to the study of sounds produced in natural environments and how they can provide important information about the state of the environment, as well as on the potential impacts caused by changes due to external influences. The analysis and visualization of large amounts of ecological recordings, as well as the development of appropriate tools for audio analysis constitute a major challenge. Mechanisms for extracting audio features, as well as the characterization of acoustic events of interest, resulting in datasets that capture the frequency variations and the occurrence of acoustic events in the recordings, still constitute a problem due to available solutions do not prove adequate for data analysis in acoustic ecology research, involving domain-specific issues and voluminous amounts of audio records collected over long periods of time. This work aims to address problems related to the extraction of audio features, providing assistance through visualization to the selection of the most significant, that could represent the subtle variations in ecological recordings, as well as assisting specialists in the generation of annotated datasets by the characterization of acoustic events through exploratory visualizations, and methods for detecting vessels in underwater recordings. A framework named Seecology is presented, encompassing suitable methods and tools to supporting specialists and scholars of environmental analysis. Case studies were carried out with the framework in terrestrial and underwater recordings provided by acoustic ecology researchers, by producing datasets from the custom feature extractor included in the framework, and in the case of the method developed for detecting boats in underwater recordings, a comparative study to another method was conducted to determine its accuracy, in addition to the case study to determine its effectiveness. The presented methods for extracting characteristics, characterizing acoustic events through exploratory visualization and boat detection, demonstrated their effectiveness for applications in acoustic ecology, with the framework containing the methods capable of producing multidimensional datasets without excessive computational costs, allowing the user to easily generate annotations on this data through the included visualizations. The boat detection method performed better than the one it was compared, both in speed and accuracy, being able to detect weak signals from boats even under extreme noise.

Keywords: Acoustic Ecology, Soundscape, Framework, Visualization, Feature Extraction, Audio Analysis, Boat Detection, Radviz.

LIST OF FIGURES

Figure 1 – Audio spectrogram depicting frequencies up to 22kHz (Y axis). It depicts two acoustic events in the recording: insect noise (from second 0 to second 35) and cicada noise (from second 37 to the end). The colors map signal intensity, usually in decibels (dB) at each frequency, from lower (blue) to higher (red) values.	34
Figure 2 – Conceptual framework for Soundscape Ecology studies. (Extracted from Pijanowski <i>et al.</i> (2011b)).	35
Figure 3 – Workbench for Soundscape Information Systems with the major features of a system that could collect, manage, analyze and present Soundscape Ecology data. (Source: Pijanowski <i>et al.</i> (2011b)).	36
Figure 4 – Radviz visualization with 4 features placed as dimensional anchors along the dimensional axis. (A) Examples of samples with corresponding features array values are positioned inside the visualization. (B) A sample position calculated by the sum of vectors (Red lines), resulting in the vector that gives the sample position in the visualization (Blue line). The final sample position is represented by a green dot.	41
Figure 5 – DualRadviz overview with a Parallel Coordinates visualization (B), the DualRadviz visualization (C) with dimensional anchors placed in the central circle and sample classes distributed along the external circle. Moving the slider (A) will result in the transition between the feature and class projections, allowing to see the influence of each feature to a specific class. Source: Adapted from Correa and Carvalho (2016)	42
Figure 6 – Radviz visualization before and after histogram equalization. Source: Adapted from Wang <i>et al.</i> (2018)	42
Figure 7 – Concentric Radviz visualization. Source: Adapted from Ono <i>et al.</i> (2015)	43
Figure 8 – Feature Extraction Tool screen with some of the features available for analysis.	54
Figure 9 – Boat Signature Extraction Tool screen with available settings and progress bar.	54
Figure 10 – Timestamp mouting screen	56
Figure 11 – Directory structure containing the files with the feature vectors extracted, related to each audio sample, and visualization resources produced.	57

Figure 12 – Logarithmic scale spectrograms of two underwater recordings. The red squares mark frequency peaks due to boats and the blue squares mark peaks due to fish choruses. Both recordings include background noise due to acoustic emission by crustaceans, that can be perceived as vertical lines in the spectrograms. (A) 15-minute audio spectrogram with two types of fish chorus and a boat signature characterized by three peaks; (B) 15-minute spectrogram with a single boat signature characterized by 10 peaks and heavy background noise due to crustacean emissions.	59
Figure 13 – In both charts the blue lines depict, respectively, the frequency spectra (up to 2.7kHz) of an input signal in dB (A) and in dB^2 (B); the green and red dotted lines depict the spectral signal smoothed by convolution with a size 11 blackman window, scaled respectively by $1.5\times$ and by $3\times$ its standard deviation. The vertical watermarks in red identify the frequency peaks due to boat emissions.	63
Figure 14 – Spectrum signature (blue line), FAV signature (green line) and binary boat detection signature (red marks) of (A) a recording that includes fish chorus and a boat; and (B) a recording that includes a small boat at minute 11. The horizontal dotted purple lines indicate the threshold defined to associate frequency peaks with acoustic emissions by boats.	64
Figure 15 – Effect of the sigmoide function on sample spatial placement in the Concentric Radviz visualization. Original feature values are presented in the horizontal axis and resulting values after the use of the function are presented in the vertical axis. The values 0.3 and 0.7 were marked with a red dot and gray dotted lines to demonstrate the original and resulting values depending on the sigmoidal function values.	67
Figure 16 – Effect of the rotation of a dimentional axis over the dataset samples, represented with a black arrow originating from the original position of the first dimentional anchor. Concentric Radviz A presents the original sample placement, B with a clockwise rotation of 30° , C with a rotation of 60° and the final placement of the samples after a 90° clockwise rotation.	68
Figure 17 – Audio player associated with a spectrogram for the full recording and a line chart to visualize feature values variation on the samples. The sample window on both player was set to 2 seconds during the extraction process. The sample selected on each player is represented as a vertical red watermark. On A, a recording with cicadas is presented, and in B a recording with the presence of rain in different intensities.	69
Figure 18 – Sample selection on the visualization and list of each recording and sample of the respective selection.	69
Figure 19 – Working directory selection screen.	72

Figure 20 – Framework areas accessible from the home screen.	73
Figure 21 – Screen used to split audios into one-minute files.	74
Figure 22 – Line chart visualization of recordings from 2015-03-06 depicting the comparison of three features extracted (MFCC_0, MFCC_9 and RMS).	75
Figure 23 – Modal window with an audio player containing a spectrogram representation of the original recording for the sample clicked on the line chart visualization..	76
Figure 24 – Main screen of the Radviz Concentric visualization with feature selection on the left tab.	76
Figure 25 – Tabs for Feature selection (A), labels and filters (B), and sigmoidal function (C) to remap the sample values in the visualization.	77
Figure 26 – Boat detection visualization.	78
Figure 27 – 24-hour spectrogram (frequency band from 0 to 2KHz) of the audio recordings taken at LSMSP in February 17, 2015. The recordings captured three types of fish chorus (A, B and C), as well as long and short duration sound emissions by boats and ships (D). The line chart at the bottom shows the detections obtained with the proposed method minute-by-minute and the number of peaks that characterize the corresponding boat signature.	80
Figure 28 – 24-hour spectrogram (frequency band from 0 to 2KHz) of audio recordings taken at XJSP in February 10, 2015. The recordings captured three types of fish choruses (A, B and C) and sound emissions by multiple ships (D) appearing as long and short duration acoustic events. The detection line chart shown under the spectrogram reveals a concentration of short duration boat events in the period from around 9:00am to 11:00am; their corresponding signatures are characterized by up to 13 frequency peaks.	81
Figure 29 – 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at XJSP in February 4, 2015 at 05h:39m:52s. A boat has been detected in the interval from minutes 1 to 9; the corresponding boat detection line chart is shown at the bottom.	82
Figure 30 – 15-minute spectrogram of a recording taken at XJSP in February 4, 2015 at 21h:12m:44s (frequency band from 0 to 1KHz). A boat occurrence has been detected in the interval from minutes 7 to 12; the corresponding boat detection line chart is shown at the bottom.	83
Figure 31 – 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at XJSP in February 4, 2015 at 10h:40m:46s, in which a boat has been detected during minutes 5 and 6; the corresponding boat detection line chart is shown at the bottom.	84

Figure 32 – 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at XJSP in February 4, 2015 at 10h:10m:42s. Listening to the audio it is possible to identify the sound of a boat at minute 12, which the algorithm failed to detect, as observed in the corresponding detection line chart at the bottom.	85
Figure 33 – 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at LSMP in February 11, 2015 at 06h:27m:57s. The low amplitude signal from a boat has been detected in the intervals from minutes 0 to 3 and from 8 to 11, even though it is masked by crustacean noise, as observed in the boat detection line chart shown at the bottom.	86
Figure 34 – 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at LSMSP in February 11, 2015 at 15h:29m:54s. The signature of a boat has been detected along the duration of the recording, characterized by 35 up to 59 frequency peaks, as observed in the corresponding detection line chart.	87
Figure 35 – 15-minute spectrogram of a recording taken at LSMSP in February 4, 2015 at 04h:48m:55s, where overlapping sound emissions from two boats can be observed. The charts show the signatures computed of the audio samples corresponding to minutes 2, 3, 4 and 5, displaying the input spectrum signature (blue lines) with the corresponding boat detections highlighted by the red vertical lines.	88
Figure 36 – FAV and DEMON signatures obtained for a one-minute audio sample that includes crustacean acoustic activity overlaped with a persistent acoustic emission by a boat (spectrogram on the right). The dotted gray lines indicate the detection threshold ($1.5 \times Standard\ Deviation$). Both algorithms correctly detected the boat signature, with two frequency peaks. The red watermarks indicate the peaks detected in the FAV signature (purple line); the black dots indicate the detections in the DEMON signature (green line).	89
Figure 37 – FAV and DEMON signatures obtained for a one-minute audio sample that includes crustacean acoustic activity overlaped with a persistent acoustic emission by a boat (spectrogram shown on the right). The dotted gray lines indicate the detection threshold ($1.5 \times Standard\ Deviation$). In this case the boat signature is characterized by eight frequency peaks. The red watermarks indicate the boat peaks correctly detected in the FAV signature (purple line); the black dots indicate the detections in the DEMON signature (green line). The DEMON algorithm identified a single peak due to the boat, the other peak is actually a false-positive, not related to the boat.	90

Figure 38 – Acoustic events annotated during the exploratory analysis of the dataset, using 7 of the 12 components of the MFCC feature, distributed as dimensional anchors along two dimensional axis, depicted in "Layers". The samples were annotated for 8 acoustic events of interest, depicted in "Sample labels". Some acoustic events of the same group were divided into two groups, as in the case of rain and heavy rain.	93
Figure 39 – Experiments in segregating variations of the same acoustic event (A and B), and the resulting segregation response of adding more layers to the visualization (C).	93

LIST OF ALGORITHMS

Algorithm 1 – FAV signature computation	61
---	----

LIST OF TABLES

Table 1 – Available features for the terrestrial and underwater extraction process and its origin package.	55
Table 2 – Detection accuracy of FAV and DEMON algorithms on a database of 1440 one minute audio samples.	89
Table 3 – Boat detection accuracy and signature accuracy results obtained with the FAV and DEMON implementations, as well as false positive and false negative detection rates.	90
Table 4 – Times (in seconds) to compute the boat signatures of a 15 minute audio recording and of a one minute sample extracted from the same audio. Time measures were obtained from five executions of each (DEMON/FAV) implementation on a <i>Intel© CoreTM i7 – 7700 @ 3.60GHz</i> $\times 4$ with 16GB memory.	91

CONTENTS

1	INTRODUCTION	27
1.1	Problem Description	27
1.2	Motivation and Challenges	28
1.3	Objectives and Contributions	29
1.4	Research Questions	31
1.5	Thesis Organization	31
2	RELATED WORK	33
2.1	Basic Concepts	33
2.2	Acoustic Ecology and Visualization Frameworks	34
2.3	Acoustic Event Annotation	37
2.4	Radviz visualization and variations	40
2.5	Boat Detection	44
2.6	Final Considerations	46
3	MATERIALS AND METHODS	49
3.1	Equipment	49
3.2	Ecological Recordings and Datasets	50
3.3	External Libraries	51
3.4	Visualization Resources	52
3.5	Seecology: Data Visualization Framework for Soundscape Ecology Applications	52
3.5.1	<i>Feature Extraction Tool</i>	53
3.5.2	<i>Accessing Local Files In Web Environment</i>	58
3.6	Automatic Detection of Vessel Signatures in Audio Recordings with Spectral Amplitude Variation Signature	59
3.7	An Exploratory Approach For Annotating Acoustic Events	65
3.8	Final Considerations	70
4	RESULTS	71
4.1	Seecology: Data Visualization Framework for Soundscape Ecology Applications	71
4.1.1	<i>User Interface and Visualizations</i>	72
4.1.1.1	<i>Line Chart</i>	74

4.1.1.2	<i>Concentric Radviz</i>	75
4.1.1.3	<i>Boat Detection</i>	78
4.2	Automatic Detection of Vessel Signatures in Audio Recordings with Spectral Amplitude Variation Signature	79
4.2.1	Comparison: FAV vs DEMON	86
4.3	An Exploratory Approach For Annotating Acoustic Events	92
4.4	Final Considerations	94
5	CONCLUDING REMARKS	95
5.1	Conclusions and Contributions	95
5.2	Publications	98
5.3	Limitations and Future Work	99
	BIBLIOGRAPHY	101
	GLOSSARY	109

INTRODUCTION

The research field known as Soundscape Ecology refers to the study of sounds present in the environment in order to understand the dynamics of man-nature relationship in different spatial and temporal scales ([PIJANOWSKI *et al.*, 2011b](#)). According to [Pijanowski *et al.* \(2011b\)](#), sounds emanating from environments may vary both spatially and temporally. In an urban landscape, sounds of cars, planes, sirens and machinery, among others produced by humans (Anthrophony), prevail over sounds generated by the fauna and flora ([BOTTELDOOREN; COENSEL; MEUR, 2004](#); [RAIMBAULT; DUBOIS, 2005](#)). In contrast, sounds captured in a typical forestry landscape are dominated by animal sounds (Biophony), running water and leaves in the wind (Geophony) ([SWANSON *et al.*, 1988](#)). The integration of all these sound sources in an environment originates the term Soundscape.

1.1 Problem Description

The management and analysis of large amounts of audio recordings and associated data, so as to facilitate correlation with complementary environmental data obtained from other types of sensors or metrics, constitutes a major challenge in Soundscape Ecology. The hability to extract multiple audio features from ecological recordings, as well as to characterize acoustic events of interest in these recordings would provide experts with proper tools to identify relevant events, assess their impact and, at a wider scale, identify changes in environmental behavior, whether originating from Biophony, Geophony or Antrophony.

The study of soundscapes involves the analysis of massive volumes of data collected in raw format, which may correspond from days to years of audio recordings, as reported by [Pijanowski *et al.* \(2011a\)](#) and [Towsey *et al.* \(2014\)](#), where recordings were collected for over a year. The manual analysis of this information would require a huge effort in terms of time, manpower, financial and computer optimal resources, among other factors.

The analysis of audio records can be done manually or automatically, depending on the analyst's goals and tasks. If the purpose of the analysis is to identify the extent of recognized sounds (insects, birds, environment or subclasses within these groups), researchers can manually mark acoustic events, setting their time and duration. There are some online tools available to accomplish this task, such as Pumilio (VILLANUEVA-RIVERA; PIJANOWSKI, 2012) or Ecosounds¹, that provide a user-friendly interface to assist the specialist to account for the number of acoustic events annotated in the archives. If the purpose of the analysis is to identify the particular frequency characteristics or amplitude of the acoustic events, several libraries are available developed specifically to extract a wide range of characteristics from audio recordings, such as Yafee (MATHIEU *et al.*, 2010), Libxtract (BULLOCK, 2007), Essentia (BOGDANOV *et al.*, 2013) and Librosa (MCFEE *et al.*, 2015), that directly export the results of batch processing into files, facilitating the researcher's task. Moffat, Ronan and Reiss (2015) present a comparative analysis of 10 libraries for extracting audio features, reporting for each library a number of properties, such as the programming language it supports, the total number of features the library can extract, the effort required to use it and its performance.

Researchers in Acoustic Ecology and Soundscape Ecology often rely on ecological and acoustic indices to assess a landscape's acoustic environment. Such indices are calculated from the physical properties of the audio, aimed at capturing the characteristics and health of an environment from sounds recorded on site. The choice of index varies with each study, as some indices are more sensitive to sounds produced by humans, others are more suitable for studies of birds, indicating the places where birdsong is more pronounced (PIJANOWSKI *et al.*, 2011a; PIERETTI; FARINA; MORRI, 2011; FULLER *et al.*, 2015).

1.2 Motivation and Challenges

The motivation for collecting ecological and acoustic indices differs across the studies. While some authors focus on studying specific species (FARINA; PIERETTI; PICCIOLI, 2011; FROMMOLT; TAUCHERT, 2014; FIGUEIRA *et al.*, 2015), others attempt more embracing approaches and are concerned with representing temporal changes in the soundscape of entire environments (LAILOLO, 2010; DEPRAETERE *et al.*, 2012; STRÖM, 2013; FARINA; PIERETTI, 2014; SÁNCHEZ-GENDRIZ; PADOVESE, 2015), or even with comparing different locations (PIJANOWSKI *et al.*, 2011a; FULLER *et al.*, 2015). Patterns emerged from soundscape in different spatio-temporal scales can help specialists to determine the effects of the intrusion of anthropogenic sounds (MOORE *et al.*, 2012), as well as the health of an environment by changes in the sound frequencies.

Creating tools that are "soundscape-ready" is necessary to assist experts on soundscape studies (PIJANOWSKI *et al.*, 2011a), providing more effective mechanisms to extract features

¹ <<https://www.ecosounds.org/>>

from large collections of field recordings, process, analyze and annotate those recordings (PIJANOWSKI *et al.*, 2011b; PIJANOWSKI *et al.*, 2011a). Coupled with those previous steps, meaningful interactive visualizations may provide researchers novel approaches to look at the generated data in order to better assess the soundscape scenarios (PIJANOWSKI *et al.*, 2011a; TOWSEY; ZHANG, 2014; Truskinger *et al.*, 2014).

Another important task in the field of acoustic ecology is the identification of acoustic events in the audio recordings, where each distinct acoustic event may require different approaches or audio features in order to be identified. Several analysis methods can result in a variety of data types, annotations, and visualizations, which if performed manually can produce even greater data variety due to the large number of events that can occur simultaneously (Truskinger *et al.*, 2014). Another problem with manual inspection and annotation of ecological recordings is the amount of effort required, limiting the ability to provide large-scale spatiotemporal observations, since those tasks rely on the attention and accuracy of each person involved (WIMMER *et al.*, 2012).

1.3 Objectives and Contributions

Studies for automatic identification make use of annotated data for training machine learning algorithms (TRUSKINGER, 2015; TRUSKINGER; TOWSEY; ROE, 2015; BARCHIESI *et al.*, 2015; KHOLGHI *et al.*, 2018; BROWN; GARG; MONTGOMERY, 2018), which requires time and resources to be produced as ecological recordings tend to have long duration. In this work we combine the use of the Concentric Radviz visualization (ONO *et al.*, 2015) in conjunction with a custom soundscape feature extractor (REIS; SANTOS; OLIVEIRA, 2018), which creates multidimensional feature vectors describing the audio files. Such multidimensional data sets, described by a large number of attributes (features) pose many challenges for analysis and visualization. This custom feature extractor is capable of quickly producing the necessary audio features in a standardized format, employing a custom approach to extract audio features from small samples of the recordings, resulting in a detailed perception of the acoustic changes and acoustic events occurring on each recording.

We introduce a framework that relies on the Concentric Radviz visualization (ONO *et al.*, 2015) to support interactive inspection of the audio samples associated with acoustic events. This approach allows the identification of clusters of audio samples with common feature properties, making it easier to annotate similar acoustic events. Data manipulation tools have also been added to the Radviz visualization to assist in the data exploration task, such as using a sigmoidal function to reajust the sample spatial positions and prevent common problems encountered in the Radviz visualization. An audio player was also included in the visualization, to assist in the exploration of the recordings with their spectrogram representations and corresponding features extracted for each audio sample. This audio player can be useful to associate the audio samples

with possible acoustic events represented on them.

Besides the approaches for feature extraction and annotation of acoustic events, this research introduces a novel method for detecting boats in audio recordings. Ecological audio recorders and hydrophones offer a cost effective method for ocean monitoring in order to identify undesirable or threatening scenarios. This has motivated research efforts towards devising methods to automatically detect the presence of vessels in underwater acoustic recordings. This is a relevant problem because audio signals emitted by boats, ships and water crafts in general can strongly impact marine life. They may interfere with species communication (WITTEKIND; SCHUSTER, 2016), echolocation mechanisms (VEIRS; VEIRS; WOOD, 2016) and threaten fish population growth (JAIN-SCHLAEPFER *et al.*, 2018). Also, the presence of boats in conservation areas are, when not prohibited, at least controlled, aiming the prevention of illegal, unreported and unregulated fishing activities (PETROSSIAN, 2015).

While some solutions have employed classification approaches derived from the analysis of acoustic features (LEAL; LEAL; SANCHEZ, 2015), others have focused on the specific spectral signatures of the sounds emitted by the different types of vessels, e.g. training neural networks to detect a set of known signatures (POLLARA *et al.*, 2017; SLAMNOIU *et al.*, 2016; CHUNG *et al.*, 2011; HANSON *et al.*, 2008). Nonetheless, improving detection accuracy in noisy conditions and reducing the rate of false positives remain as challenges. A related relevant issue is automatic classification, e.g., categorizing the vessels detected, for example, according to their size (small, medium, large).

The *Detection of Envelope Modulation On Noise* (DEMON) is one of the most reliable methods for ship detection and classification (CHUNG *et al.*, 2011). However, it is reported to yield poor results on noisy signals. Moreover, it is highly dependable on a manual selection of the relevant frequency band pass filters for analysis, which must be tuned according to the type of target vessel (HANSON *et al.*, 2008).

We introduce a novel method to automatically detect the presence of boats or ships in acoustic underwater recordings. Similarly to the DEMON method, ours relies on computing acoustic signatures from the input signal. However, our method has lower computational cost and requires no calibration or human intervention. Moreover, we show that it can successfully detect vessels even in challenging conditions posed by very noisy signals, as in recordings that include crustacean acoustic activity.

This research addresses three problems in ecological soundscape data analysis, namely: (i) a framework for the extraction of features in ecological recordings, with approaches applicable to both terrestrial and underwater audios, (ii) a method to detect the presence of boats in underwater recordings and (iii) visualization methods for annotation of acoustic events and analysis of the recordings. While for the land studies, several features and assets for use in visualizations are produced, for the underwater recordings approach, adding to the features used in land studies, a boat presence detection method was also introduced and validated.

1.4 Research Questions

The purpose of this research project is to develop methods to assist in the identification and annotation of acoustic events of interest in ecological recordings, characterizing those events by means of audio features. A framework for Soundscape Ecology data visualization was developed, providing means for the visual analysis of data resulting from soundscape studies. As stated by Pijanowski *et al.* (2011a), although several libraries and tools have been developed for the analysis of soundscape data, most of those currently available to specialists are not entirely "soundscape-ready" and there is a lack of visualization frameworks specific for soundscape studies. In this work, we intend to fill this gap and introduce meaningful interactive visualizations integrated into an environment capable of assisting researchers in their data handling and analysis tasks. This includes an investigation on methods and techniques to assist feature extraction from audio recordings, considering different approaches. In order to fulfill these goals, this work will have to answer the following research questions:

- Q1: Can we formalize a visualization framework for soundscape ecology data that is robust enough to represent the wide range of information characterizing soundscapes, without incurring in the high computational processing costs of current approaches for audio data analysis?
- Q2: How can acoustic visualizations fueled by audio features support the identification of similar acoustic events of interest, so that specialists can explore the universe of acoustic events contained in ecological recordings, analyzing the relationship between acoustic events and audio features?
- Q3: It is possible to automatically detect the presence of boats and other types of vessels in noisy underwater recordings?

To help answering those questions, studies were developed applying several visualization techniques over ecological recordings data processed with a custom feature extraction tool. The boat detection method has also been compared with a standard method from the literature.

1.5 Thesis Organization

This thesis is organized into five chapters, including this introductory chapter that presented the context and motivation for this doctoral research project.

Chapter 2 introduces the key concepts for understanding the problems addressed. An overview of methods and visualizations employed in this research is presented, with separate sections for underwater boat detection, acoustic event annotations and radial visualizations. Frameworks for acoustic ecology and visualization are also presented and discussed.

[Chapter 3](#) describes the audio recordings and dataset analysed, along with the equipments used to conduct the extraction and analysis of the audio data, and the necessary libraries employed in developing the Seecology framework. This chapter also present methods developed for attaining the results, namely the automatic detection of boats in underwater recordings, by means of a new spectral amplitude variation signature. An approach for the annotation of acoustic events based on the use of a Concentric Radviz visualization, and a description of the Seecology, a data visualization framework for acoustic ecology applications, containing each part included, such as custom feature extraction tools and the resulting directory and files structure.

In [Chapter 4](#), we present and discuss the main results obtained with the methods introduced in Chapter 3, namely the automatic detection of boats in underwater recordings, the approach for the annotation of acoustic events, and the Seecology Framework, with the resulting web based user interface developed.

In [Chapter 5](#), the results obtained during this doctoral research are presented, along with future work to be undertaken to advance the Seecology framework.

RELATED WORK

In this chapter, we discuss related work on three topics concerned with Soundscape data analysis. In Section 2.1, we present the main concepts for understanding this research. In Section 2.2 we present existing approaches on the literature for acoustic ecology and visualization frameworks. In Sections 2.3 and 2.4, we present respectively studies in acoustic events annotation with current methods and tools used to perform this task, including several libraries for audio feature extraction available in the literature. Finally in Section 2.5, methods for boat detection in underwater recordings are presented. These related contributions provided the basis for tackling the problems presented in Chapter 1, providing the necessary knowledge for the development of this research.

2.1 Basic Concepts

Before presenting the related work we introduce some basic techniques relevant to this work, such as the Fourier transform and the generation of spectrograms from audio.

The term "framework", can be interpreted in different ways. In the case of this work, with a focus on ecology and analysis of frequency variations in audio recordings, as well as the analysis of acoustic events, the term refers to a set of methods and procedures to assist researchers and specialists in acoustic ecology. The goal is to facilitate the execution of audio inspection and analysis tasks by means of a range of tools for data extraction and analysis.

Researchers in acoustic ecology need to analyze the different frequencies that make up an acoustic environment separately. This process of decomposing the audio in individual frequencies is carried out through the Fourier transform, an integral transform that expresses a function as a set of sinusoidal basis functions. The Fourier transform allows decomposing a temporal function expressing an audio signal and analyze each of the component frequencies.

Another important tool for audio analysis is the spectrogram, a visual representation

of frequency variations of a recording. Each column of the spectrogram is generated from a Fourier Transform obtained from a temporal sample of the audio. When computed on each audio segment, a two-dimensional matrix filled with amplitude values for each frequency is generated. Each matrix entry is converted to a color scale, according to its amplitude value, resulting in an image representing frequency variations of an audio over time. An example can be seen in Figure 1.

Figure 1 – Audio spectrogram depicting frequencies up to 22kHz (Y axis). It depicts two acoustic events in the recording: insect noise (from second 0 to second 35) and cicada noise (from second 37 to the end). The colors map signal intensity, usually in decibels (dB) at each frequency, from lower (blue) to higher (red) values.

2.2 Acoustic Ecology and Visualization Frameworks

Some authors proposed frameworks to deal with ecology recordings, [Pijanowski *et al.* \(2011b\)](#) for example proposed a conceptual integrative framework that describes how the dynamic soundscapes are created as a result of factors that directly affect the environment. Those factors may include the climate of the region, land transformations processes, biodiversity patterns, timing of life history events and human activities. According to the authors, the human presence transforms the landscape through the use of land or changes in occupation. These changes in turn, interact with biophysical features such as terrain and ground, to yield heterogeneity on the spatial structure of the landscape ([FARINA, 2006.](#)).

The conceptual framework presented in Figure 2 considers that human modifications introduced in the landscape change its structure, such as terrain and soils. Those changes influence the distribution and abundance of species and their interactions in multiple temporal and spatial scales. Land forms, like rivers and valleys, provide specific patterns, particularly those that continuously produce sounds, such as water or wind. The climate (arrow 3) helps controlling the distribution of species along with the occurrence of specific life-history events such as noisy

Figure 2 – Conceptual framework for Soundscape Ecology studies. (Extracted from Pijanowski *et al.* (2011b)).

insects like the cicadas. Geophonic sounds are also influenced by climate changes (arrow 4). The human presence on the environment produces sounds (anthrophony), that influence the soundscape. Biophony, geophony, and anthrophony (arrows 4 and 5) combine to create the soundscape of a specific area. Changes in the soundscape patterns influence natural processes (arrow 6), for example, animal vocalizations could be masked by human-generated noise, consequently altering the population or community dynamics such as predator-prey relationships (BARBER; CROOKS; FRISTRUP, 2010; PROULX; WALDINGER; KOPER, 2019).

As stated by Pijanowski *et al.* (2011b), although there are many tools and software packages available, most are not "soundscape-ready", meaning they are not ready to deal with the volume of data produced in soundscape studies, least of all perform custom data extraction for multiple features at once. Because of the large volume of data to be processed, researchers need advanced scripting capabilities and interfaces to enable the use of sophisticated pattern matching and databases. Also, according to them, *"a web-based Soundscape Information System is needed to enable a soundscape ecologist to input, archive, retrieve, visualize, listen to, analyze, manage and access raw and processed information based on large collections of audio recordings from automated sensors"* as shown in Figure 3, where they present the major features of a Soundscape Information Workbench. The proposed workbench does not determine which features to be retrieved from recordings or how to perform the extraction process.

The area of Music Information Retrieval provides contributions to the Soundscape Ecology research field. On Bellisario and Pijanowski (2019), the authors argue that MIR toolkits could provide a new generation of soundscape metrics and contribute new methods to manage natural resources. By using spectral timbral analysis, they were able to distinguish dominant soundscape components considering timbre as an acoustic property, with data that had to be previously manually annotated for each recording (BELLISARIO *et al.*, 2019). In order to

Figure 3 – Workbench for Soundscape Information Systems with the major features of a system that could collect, manage, analyze and present Soundscape Ecology data. (Source: [Pijanowski et al. \(2011b\)](#)).

classify audio features, [Bellisario et al. \(2019\)](#) used machine learning with a k-nearest neighbor approach. On their study, they found seventeen acoustic features capable of predicting a group of annotated labels composed of acoustic events from biophonic, geophonic and anthrophonic sources with good accuracy.

In the area of marine acoustics, a framework for assessing the effects caused by anthropogenic sounds in the marine mammals habitat has been developed, in order to supply the lack of existing means to estimate how the contribution of human activities may affect a specific species or population ([MOORE et al., 2012](#)). Although the effects of antropogenic sound on marine mammals have been the focus of much research over the last 40 years, [Moore et al. \(2012\)](#) state that quantifying that type of sound is a complex task, and the numerous field studies that try to measure the properties of that interaction are often ambiguous, given that results are highly dependent on context.

[Moore et al. \(2012\)](#) proposed a framework based on the acoustic habitats of marine mammals, to help not only researchers but also environmental agencies to supervise the effects of humans over an environment and their inhabitants. In US waters for example, three statutes provide the legal framework on the supervision of effects of anthropogenic sounds, but although care for the environment is demanded by legislation, the assessment of cumulative effects on anthropogenic from multiple sources over the habitat of marine mammals is a very difficult endeavor. According to [Moore et al. \(2012\)](#), there are no ready means to estimate the contribution

of one specific activity over a species or population, therefore, they proposed a framework where the sounds produced by multiple sources are mapped, along with information on their distribution, density, movement patterns and ecology of marine mammals in assess potential risks resulting from the introduction of sound-generated activities, providing means for assessing the "human footprint" on a local and a global scales.

In order to construct their proposed framework, the authors elaborate a three step cycle, which they called "Science foundation", "Acoustic-habitat framework" and "Research and evaluation". "Science foundation" is the framework is responsible for gathering information about the environment, habitats, marine mammals, their abundance and ecologically important areas. On the "Acoustic-habitat framework" step, researchers must overlay acoustic-habitat maps with the marine mammals distribution patterns, abundance numbers and their ecological importance, as well as identifying periods of concern and data gaps. The final step of the cycle, "Research and evaluation" deals with gaps and the research needed to close those gaps. Once identified the causes, preventive measures must be implemented to protect marine mammals, and to prevent any further exposure, regular reviews of progress need to be conducted, both to supervise and to adjust the necessary measures that were taken.

2.3 Acoustic Event Annotation

The investigation of feature spaces in multidimensional data is a highly relevant problem in many domain applications. A feature set is the set of features employed in representing the multidimensional data. Several classical visualization techniques have been employed to support feature investigation and the specific data mining task of feature selection, e.g., multidimensional data visualizations such as Parallel Coordinates ([INSELBERG; DIMSDALE, 1990](#)), Star Coordinates ([KANDOGAN, 2000](#)) and Radviz ([Hoffman *et al.*, 1997](#)) afford direct manipulation of the data items while allowing analysts to infer how distinct features contribute to visual patterns observed.

Annotated datasets are useful in several tasks, such as to describe the content of an ecological recording study, or as input to machine learning algorithms which require annotated data for training, validation, and testing. In soundscape ecology applications, researchers are interested in annotated acoustic events of interest, aiming the study of the interactions between those events, the analysis of soundscape frequencies and gaps due to the lack of an acoustic source, impacts due the introduction or removal of an species or changes in the landscape, among others interests.

In order to create annotated data from ecological recordings, researchers spend a considerable time and effort listening, classifying and manually annotating discernible acoustic events ([KRIJNDERS; NIESSEN; ANDRINGA, 2009](#); [KHOLGHI *et al.*, 2018](#)). Although some sounds can be instantly placed in a specific class such as birds or rain, others may be difficult,

considering that some frequencies are not easily perceived by the human ear. Audio spectrograms (Figure 1) allows the researchers to visualize acoustic events within the ecological recordings, without the need to listen to it.

There are tools specifically created to assist researchers perform manual annotations over large audio datasets (VILLANUEVA-RIVERA; PIJANOWSKI, 2012; CARTWRIGHT *et al.*, 2017; MELÉNDEZ-CATALÁN; MOLINA; GÓMEZ, 2017), directly on the audio waveform and spectrogram representation. The work by Villanueva-Rivera and Pijanowski (2012) represent an effort to archives, manage, retrieve, analyze, visualize, and listen to sound files, with the Pumilio¹ tool. The web interface used makes it easy to researchers to navigate and annotate large dataset collections, unlike previous tools available for the analysis of individual recordings at a time, such as Raven². Another ecological recording manual annotation tool was developed by Cartwright *et al.* (2017), while studying the effect of visualizations and complexity on crowdsourced audio annotations. The developed tool, CrowdCurio Audio-Annotator³, provided a simple interface where the dataset recordings are presented one by one to the user, which in turn performs the task of associating the acoustic events present in the recordings with the predefined labels, marking the acoustic events with the mouse, directly in the recording spectrogram visualization. The work of Meléndez-Catalán, Molina and Gómez (2017), produced the BMAT Annotation Tool⁴, which is an extension of the CrowdCurio tool, with some new resources, such as the possibility to switch the sound visualization between its waveform and its spectrogram.

Both the Pumilio, CrowdCurio and BMAT tools however, does not provide the user with any feature extraction methods other than the automatic generation of spectrograms. They are required to be installed on a web server with specific libraries, and while the Pumilio is designed mainly as a large datasets management system with built-in annotation tools, the CrowdCurio and BMAT tools approach the task from the perspective of crowdsourcing analysis, presenting analysts with one sequential recording at a time, for entering acoustic event annotations.

However, annotation of ecological recordings requires a great deal of manpower and analysis time, especially if we consider that ecology studies easily span months to years of recordings. In order to reduce labor requirements for data analysis and annotation tasks, efforts have been made to find automated ways to annotate acoustic events of interest in large datasets (OLIVEIRA *et al.*, 2015; GUYOT; PINQUIER, 2015; Heittola *et al.*, 2013; HEITTOLA *et al.*, 2011; KRIJNDERS; NIESSEN; ANDRINGA, 2009; WIMMER *et al.*, 2013; Truskinger *et al.*, 2014; TRUSKINGER, 2015; Stowell; Benetos; Gill, 2017; KHOLGHI *et al.*, 2018), preferably with reduced or none human interaction.

Automatic detection and classification of acoustic events of interest require that the desired audio classes are previously known, some authors have used previously annotated datasets

¹ <<http://lvillanueva.github.io/pumilio/>>

² <<http://ravensoundsoftware.com/>>

³ <<https://github.com/CrowdCurio/audio-annotator>>

⁴ <<https://github.com/BlaiMelendezCatalan/BAT>>

to train models to recognize specific audio events (OLIVEIRA *et al.*, 2015; GUYOT; PINQUIER, 2015; Podwinska *et al.*, 2019). In order to generate those annotated data automatically, a common approach, known as bag-of-features, is to extract features from audio recordings that could be employed to segregate data into a visualization or train models for acoustic event recognition.

The bag-of-features approach for detection of acoustic events has been used in various pattern recognition applications (OLIVEIRA *et al.*, 2015), and it is used for building codebooks of acoustic words from a recordings training set, which should be previously annotated with events organized into classes. The method proposed by Oliveira *et al.* (2015) extracts Mel and gammatone frequency cepstral coefficients, originated from psychoacoustic models, to generate a bag-of-features representation of the audio classes, used in the subsequent classification process of acoustic events such as steps, chairs, paper and keyboard noises, which were recorded in a smart room environment, allowing the creation of annotated datasets. Another work in acoustic event detection not related to ecology was presented by Guyot and Piquier (2015), where 61 types of sounds were trained to be recognized using Hidden Markov Model (HMM) with 3 states, and Mel-Frequency Cepstral Coefficients (MFCC) features, with the percentage of sounds detected in each scenery presented with a new type of visualization called samocharts, capable of representing the decibel level distribution of acoustic events of interest present on recordings.

Rather than trying to recognize single events, some authors (Heittola *et al.*, 2013; HEITTOLA *et al.*, 2011) presented an interesting method to detect overlapping sound events, applying non-negative matrix factorization (NMF) to the original recording to produce four audio streams. The separation process is not supervised, therefore the outcome of the factorization cannot be controlled. The resulting sound sources in each audio stream may represent the sum of one or more components, and each component may be originated from one or more sound sources. For each separated audio stream, a set of 16 mel-frequency cepstral coefficients (MFCCs), first and second time derivatives of the static coefficients are extracted in order to train Hidden Markov Models (HMMs) to detect events in each audio stream, making it possible to recognize overlapping acoustic events.

The work by Krijnders, Niessen and Andringa (2009) presents a semi-automated annotation tool, where possible regions containing acoustic events are segmented into the time-frequency-plane, with sounds likely to stem from a single source. The presence of a human annotator is still required to classify the selected regions, while the system learns to suggest possible annotations based on previous user input. Those suggestions are also presented for acceptance for the user, which according to the authors were able to attain the correct class in 75% of the cases. For each segmented acoustic event the system extracts a 37-dimensional bag-of-features representing properties related to the physics of the source, later used in a simple k-nearest-neighbor (kNN) classifier. Although this method presents a speed up in the annotation process, human factor in the equation still makes the method laborious. Similar methods were presented for big data analysis with collaborative annotation (WIMMER *et al.*, 2013; Trusk-

inger *et al.*, 2014; TRUSKINGER, 2015) where acoustic events of interest were annotated by a human-in-the-loop semi-automated online workbench. Although the approaches follow the same principles for classifying acoustic events, they provide the researcher with new methods and techniques to improve the annotation task. The workbench presented (WIMMER *et al.*, 2013; Truskinger *et al.*, 2014; TRUSKINGER, 2015) provides users with the capability to perform manual annotations of the recordings assisted by a few built-in call identification libraries, run several fully automated tools designed to find specific species, or even interact with the system to reduce repetitive annotation tasks.

Other authors also resort to manual annotation to train recognition algorithms, Stowell, Benetos and Gill (2017) for example, placed recorders in backpacks attached to birds, performing visual inspections on the audios by specialists in order to determine the sounds types and categories for each vocalization on the recordings. The detection and classification of events was performed by a probabilistic method, trained with a total of 8.4 hours of video-validated and 18.5 hours of human-coded sound recordings, along with their respective annotations. A similar approach was presented by Kholghi *et al.* (2018), with approximately 5-16 hours of manual annotation effort, and 12 features extracted from 1-minute audio files, to train a classifier, which was enough to automatically annotate 13 months of unlabelled audio data.

2.4 Radviz visualization and variations

Among the many techniques for visualizing multidimensional data available in the literature, Radviz (Hoffman *et al.*, 1997) can display similarity between samples of a dataset by placing features as dimensional anchors along a circle, also called dimensional axis, allowing to visually determine the formation of data clusters. Some contributions in the literature have developed variations of the traditional Radviz model, seeking a better way to present the formed clusters (ONO *et al.*, 2015; WANG *et al.*, 2018; CHENG; XU; MUELLER, 2017) or a relationship between their samples and the formed groups (CORREA; CARVALHO, 2016).

The basis of a Radviz visualization is presented in Figure 4, where two Radviz visualizations are shown (A and B). The Radviz in Figure 4-A present 8 samples distributed inside the main circle by the placement of four dimensional anchors (Features), named 1 to 4, alongside a dimensional axis. Below each sample, an array of feature values is presented, relative to dimensional anchors 1 to 4, which will be used to calculate the sum of vectors for each sample to determine their position on the visualization. The central sample of the visualization depicts two overlapped samples with distinct feature values, a common problem in this type of visualization handled by Ono *et al.* (2015) with a sigmoidal function to remap the feature values in realtime. In Figure 4-B, the vectors originating from the center of the visualization (Red) are related to the feature values of each sample, the normalized sum of those vectors will result in a new vector (Blue), equivalent to the position of the sample (Green) in the visualization. The sum of the

Figure 4 – Radviz visualization with 4 features placed as dimensional anchors along the dimensional axis. (A) Examples of samples with corresponding features array values are positioned inside the visualization. (B) A sample position calculated by the sum of vectors (Red lines), resulting in the vector that gives the sample position in the visualization (Blue line). The final sample position is represented by a green dot.

forces is calculated for each sample of the visualization on every change, such as rotating the dimensional axis or changing features. Samples with similar features will be placed together in the visualization, representing similar acoustic events.

In order to assist in the analysis of data classification, the work of [Correa and Carvalho \(2016\)](#) is a good example of using Radviz integrated with Parallel Coordinates visualization (Figure 5), named DualRadviz. The authors proposed a Radviz visualization with attributes distributed as dimensional anchors along a circle called a "dimension group" and a second concentric circle called a "class group", displaying the result of the probabilistic classifications performed on the samples. The user can transition between the projections of both Radviz circles in order to facilitate the understanding of how some particular sample was classified into a specific class. A user-made selection on Radviz samples will result in a filter selection on the same samples in parallel coordinate view, providing the user with a better method to inspect individual samples. Any changes in the dimensional anchors position inside the dimension group result in the same modifications in the Parallel Coordinates visualization and vice versa.

Another variant of Radviz, proposing several approaches to manipulate the features values by analyzing its histogram, in order to remap the values to modify the final data projection, was proposed by [Wang et al. \(2018\)](#). The work deals with a reorganization of the samples positions in the Radviz visualization through a "radial equalization", aiming at a value distribution along a histogram as uniform as possible (Figure 6). In this way, the values of each feature are stretched so that a better distribution of samples occurs over the projection space, avoiding the "clumping"

Figure 5 – DualRadviz overview with a Parallel Coordinates visualization (B), the DualRadviz visualization (C) with dimensional anchors placed in the central circle and sample classes distributed along the external circle. Moving the slider (A) will result in the transition between the feature and class projections, allowing to see the influence of each feature to a specific class. Source: Adapted from [Correa and Carvalho \(2016\)](#)

effect when a larger number of features are selected. The data transformation proposed in the work of [Wang et al. \(2018\)](#) may lead to the destruction of existing projection patterns. To solve this problem, the authors proposed a "radial specification" where the user can specify a custom spectrogram distribution, and a "radial movement", where a set of samples selected in the histogram can be moved to another position to produce better separation of the formed clusters.

Figure 6 – Radviz visualization before and after histogram equalization. Source: Adapted from [Wang et al. \(2018\)](#)

In another approach, [Cheng, Xu and Mueller \(2017\)](#) used a Radviz-based approach to reveal clusters and outliers with "specific distinguishing properties", applying a three-step iterative refinement process to Radviz visualization, aiming at reducing layout errors in what the authors describe as the three essential relationships (Sample to sample, attribute to attribute and sample to attribute), applying distance constraints to them. According to the authors, using the

proposed non-linear layout optimizations, they were able to achieve less ambiguous visualizations, with greater insight about the patterns within the data. Similar to the work by [Correa and Carvalho \(2016\)](#), the dimensional anchors on [Cheng, Xu and Mueller \(2017\)](#) can be manipulated independently to achieve better sample or cluster separation.

On the other hand, [Ono et al. \(2015\)](#) presented a variation of the traditional Radviz, namely Concentric Radviz (Figure 7), with concentric circles, named "dimensional groups", each composed of several dimensional anchors. One rotation of each dimensional group causes all dimensional anchors belonging to that group to be rotated as well. To avoid the effect of force cancellation, which is a common problem in a Radviz visualization, a user-customizable sigmoidal function has been applied to the values at the moment of force calculation, causing sample placement to have a tendency to shift towards the higher force anchor. With this toolkit, the authors have found a way to emphasize data of interest to users based on the characteristics of the anchor values. Tests performed with the proposed visualization and using annotated datasets already known from the literature, allowed the visualization of previously known groups.

Figure 7 – Concentric Radviz visualization. Source: Adapted from [Ono et al. \(2015\)](#)

Although the Radviz visualizations and variations have been presented as a good solution for visually clustering samples with similar properties, they are mostly employed for data exploration and analysis, not taking advantage of the cluster formation in order to produce annotated datasets.

The visualization interface proposed in this work is based on the concentric Radviz designed by [Ono et al. \(2015\)](#), in conjunction with a sigmoidal function and a custom non-categorical data extraction system. However, instead of the sample positioning calculation being

based on a force system, our algorithm performs a sum of vectors to quickly position the samples. The arrangement of the dimensional anchors is left to the user, and there is no automatic selection performed by any method as in the work of [Ankerst, Berchtold and Keim \(1998\)](#) where the final arrangement was given by applying an adaptation of the traveling salesman algorithm, by repositioning similarly behaving dimensional anchors next to each other. Or in [Reis, Santos and Oliveira \(2018\)](#), where the initial placement of the dimensional anchors is obtained using genetic algorithms.

2.5 Boat Detection

A significant increase in the number of large vessels navigating the oceans has motivated investigations on the effects of noise levels on marine life ([WITTEKIND; SCHUSTER, 2016](#); [MERCHANT *et al.*, 2014](#)), as large vessels emit acoustic signals in the low frequency spectrum, the same frequency bands used for communication by certain groups of animals, such as baleen whales ([WITTEKIND; SCHUSTER, 2016](#)). A recent study suggested motorboats cause a stress response in fish embryos ([JAIN-SCHLAEPFER *et al.*, 2018](#)), with an increase in their heart rate and even stronger effects, depending on the engine type, e.g., two or four strokes. Also, noise due to ships has also been found to disrupt the echolocation mechanism of endangered species such as killer whales ([VEIRS; VEIRS; WOOD, 2016](#)).

The potential impact of underwater noise on marine mammals is widely recognized. [Merchant *et al.* \(2014\)](#) characterized the natural and anthropogenic contributors to underwater noise, specifically in a conservation area with increased shipping activity associated with offshore energy developments. They monitored ship activity using *Automatic Identification System* (AIS), in which transponders send information about each ship present in the area, in addition to video surveillance and automated sound recorders to monitor noise levels. They correlated the data collected from those sources to investigate the relationship between broadband sound exposure and indicators proposed by the *EU Marine Strategy Framework Directive* to assess the effects of noise exposure over a locally protected bottlenose dolphin population. The authors determined that the bottlenose dolphin population is not under current danger due to the current noise levels they are exposed, however their study will serve as a baseline for future investigations on the effect of noise levels on the local marine life.

The combination of low-cost automated acoustic recorders and algorithms capable of detecting boats in long duration recordings can provide an effective solution to monitoring conservation areas for the presence of vessels. [Leal, Leal and Sanchez \(2015\)](#), for instance, introduced a method for boat classification based on Fourier transform and signal processing to identify the occurrence of vessels in the recordings.

Features extracted from the audio spectrum and fed to *Artificial Neural Networks* (ANNs) and *Support Vector Machines* (SVMs) have been successfully employed to classify ships accord-

ing to their type (LEAL; LEAL; SANCHEZ, 2015). A 10-dimensional feature array was created, given by the highest peak value in the original audio signal plus eight values corresponding to the energies of the first eight 500-Hz segments, and the central frequency of the smoothed signal. Leal, Leal and Sanchez (2015) argue that acoustic emissions by vessels are too complex to be analyzed directly from the raw form of the signal's discrete Fourier transform.

Averbuch *et al.* (2011) relied on wavelet packet coefficients to detect certain types of vessel in recordings with background noise. They derive acoustic signatures using a combination of a Linear Discriminant Analysis (LDA) classifier with Classification and Regression Trees (CART), plus an additional component to reduce false alarms. Automatic real-time detection requires a training step with a set of pre-registered signatures of interest, where the signatures are generated from "energy maps" derived from the blocks of "wavelet packet coefficients".

The method by Yan *et al.* (2017) uses a combination of resonance-based sparse signal decomposition (RSSD) and Hilbert marginal spectrum (HMS) analysis to recognize certain types of vessels, defined *a priori*. The authors use an RSSD decomposition of the original audio signal to separate it into high and low resonance signals, corresponding respectively to the target and background noise. Hilbert-Huang transform (HHT) is applied to extract characteristics from the purified audio and SVM is employed for signal classification. Gaussian white noise was added to the recordings in different proportions to perform the decomposition tests and obtain the boat signatures.

Both previous solutions demand transformations of the original signal, e.g., employing RSSD or waveletets, and are targeted at the problem of categorizing the vessel signatures against a predefined set of classes. The solution proposed in this paper to compute vessel signatures does not require complex transformations of the original signal, as it relies directly on its FFT frequency spectrum. Moreover, it is aimed at detecting the presence of acoustic events introduced by arbitrary types of vessels in the audio recordings, without performing classification.

A well-known method for vessel detection in audio recordings is the *Detection of Envelope Modulation On Noise* (DEMON) algorithm. By enhancing audio frequencies characteristic of vessel emissions it creates an acoustic signature that can be employed for classification (POLLARA *et al.*, 2017; POLLARA; SUTIN; SALLOUM, 2016; CHUNG *et al.*, 2011). Introduced over 50 years ago (TUTEUR, 1963), the DEMON signature has inspired several novel ship detection (Lowes *et al.*, 2019) and classification algorithms (CHUNG *et al.*, 2011).

A recent study compared four DEMON-type algorithms (SLAMNOIU *et al.*, 2016) regarding their ability to detect small ships and divers, as well as their robustness to acoustic noise. The same authors (SLAMNOIU *et al.*, 2016) introduced a variation of the classic algorithms that was shown to yield similar results with less computational effort, also arguing that their modified DEMON algorithm is more robust to noise. Kemper, Ponce and Telles (2019) obtained the boat engine RPM (Revolutions Per Minute) from the DEMON signature and wavelets packets transform.

Chung *et al.* (2011) described an approach for identification and classification of multiple ships in busy harbor conditions using the DEMON algorithm. Employing a system consisting of four hydrophones, the authors used cross-correlation to compute the delay between the corresponding recordings and could even estimate the relative positions of the boats. Chung *et al.* (2011) stated that ship detection and classification are governed by the propeller and by engine parameters such as number of blades and rotational speed. As such, the audio spectrum is one of the most reliable acoustic parameters to accomplish such tasks. A major limitation of approaches that rely on the standard DEMON algorithm is that human intervention is required to obtain good results. For instance, the user must select a suitable band pass filter to recognize a specific signature.

In order to improve the resolution of the resulting DEMON algorithm and increase the capability of detecting subtle details within the recordings, Hanson *et al.* (2008) used cyclostationary signal processing, exploiting the spectral redundancy inherent to the propeller signal. This enhancement enabled blind identification of the shaft speed and number of propeller blades, even in noisy signal conditions. This modification improved signal resolution and allowed identifying the presence of multiple vessels. Emission by snapping shrimps is, according to the authors, a major source of noise that can render methods based on the DEMON signature ineffective.

Several relevant contributions previously discussed relied on the DEMON algorithm and variations therein. Nonetheless, besides its sensitivity to certain types of noise, computing a DEMON signature from raw audio signals requires non-trivial levels of human intervention. The solution introduced in this paper adopts a different strategy to obtain an acoustic signature, which is simpler than the DEMON signature, practically eliminating the need of user intervention during the extraction and detection processes. Our tests, performed on audio recordings collected at marine ecological conservation areas with intense crustacean acoustic activity, indicate that the technique can successfully detect the presence of different types of boats and also identify their corresponding acoustic signatures.

2.6 Final Considerations

This chapter introduced some basic concepts in Section 2.1. Literature on four topics has been reviewed: Acoustic ecology and visualization frameworks (Section 2.2), acoustic event annotation (Section 2.3), Radviz visualization and variations (Section 2.4) and boat detection (Section 2.5). It is important to reflect on the contributions on each topic, their strenghts and difficulties, in order to reflect on the development of more effective methods and approaches for solving problems in the field of acoustic ecology. In addition, among the definitions in the literature on each topic, it is relevant to provide details on the approaches adopted in this doctoral research. The next chapter describes the materials and methods adopted in developing this thesis

to answer the research questions raised.

MATERIALS AND METHODS

In this chapter, we describe the instruments used to conduct this research, the computer configuration used for analysis and performance measurement (Section 3.1), during the feature extraction and boat detection stages. Both the terrestrial and underwater ecological recordings analysed and the datasets generated from these recordings are described (Section 3.2), as well as libraries (Section 3.3) and methods used to produce the necessary resources for the interactive visualizations (Section 3.4). Also, in Section 3.5, a visualization framework for acoustic ecology analysis named Seecology, is presented encompassing the methods and tools developed to support feature extraction and audio data manipulation, along with visualizations for exploring ecological recordings. A method has been developed for the automatic detection of vessel signatures in audio recordings with a novel spectral amplitude variation signature described in Section 3.6. Finally, an exploratory approach for the annotation of acoustic events based on an enhanced Radviz visualization named Concentric Radviz is presented in Section 3.7.

3.1 Equipment

As one of the goals of this research is to allow specialists to conduct analysis on ecological recordings without incurring in high computational costs, therefore, any solution should run on accessible computers. During this research experimental process, 2 different computers were used, a Dell Inspiron 15 notebook with a 6th generation Intel© Core™ i7 Quad – Core with 8GB memory and a NVIDIA© GTX970 GPU with 4GB of RAM, and a custom built desktop with an Intel© Core™ i7 – 7700 @ 3.60GHz × 4 with 16GB memory and a NVIDIA© GTX1060 GPU with 6GB of RAM. Although both computers were able to perform well the feature extraction, resources generation and data visualizations, all tasks submitted to performance measurements were conducted on the desktop.

Tests were conducted on the extraction and analysis of recording on computers with lower specifications, such as an Intel© i3 desktop with 8GB of RAM and on an Intel© i5 desktop.

The goal of these tests was to determine if the solution proposed in this research and made available to researchers, would be able to be used interactively, considering the time demanded to conduct the extraction of the features and necessary resources, employed in the visualizations. Although the proposed methods were able to functioning with these low spec computers, they were not used to conduct the majority of this research, therefore not included in the performance measeruments.

3.2 Ecological Recordings and Datasets

Two collections of ecological recordings have been used in the experiments and case studies conducted to validate the contribution of this research. The first one with terrestrial recordings was provided by Prof. Bryan Pijanowski (Purdue University), composed of 1246 audio files, these being stored in sets of four, the first audio with 10 minutes at the beginning of each hour and the others with one minute recording every 15 minutes of the hour, recorded from March 6 to March 19 in 2015 on a primary rain forest in Costa Rica (Recorded at $10^{\circ}25'54.0''N$, $84^{\circ}01'16.9''W$), one of the most biodiverse countries in the world, containing sounds from insects, amphibians, birds, rain, engines, etc. The dataset was recorded at a sampling rate of 22 kHz with 2 channels (Stereo), totaling 4054 minutes of audio recordings.

The second set of ecological recordings, provided by Prof. Linilson Padovese (Polytechnic School of USP), was used on the boat detection experiments and consists of underwater audio recordings collected at two sites located on the West coast of the State of São Paulo, namely the Laje de Santos Marine State Park (LSMSP, located at $24^{\circ}15'48''S$, $46^{\circ}12'00''W$) and the Xixová-Japuá State Park (XJSP, located at $24^{\circ}0'22''S$, $46^{\circ}23'29''W$). Both underwater sites exhibit boat traffic and are characterized by great marine biodiversity with consistent fish choruses at dusk and dawn. The LSMSP is 30km (19 miles) off the coast and serves as a protected area for reposition of fish shoals and safe reproduction of species. The XJSP conservation unit encompasses a marine area SW of the Santos Bay and an adjacent inland region of tropical forest. This is an area with strong human presence, close to a very busy port in Brazil ([SÁNCHEZ-GENDRIZ; PADOVESE, 2017](#); [SÁNCHEZ-GENDRIZ; PADOVESE, 2015](#)).

Both LMSP and XJSP ecological recording consists of 15-minute monaural audio files recorded at a sampling rate of 24 kHz. Recordings were obtained underwater with a custom autonomous hydrophone recorder developed at LACMAN (Laboratory of Acoustics and Environment of the Polytechnic School of the University of São Paulo)¹. In total the locations are composed of 2,675 recordings from LSMSP and 2,390 recordings from XJSP.

From the recordings, datasets containing audio features and ecological indices were calculated, using a time-based sample approach where for each recording, audio features and ecological recordings will be extracted for each sample window defined, resulting in several

¹ <http://lacmam.poli.usp.br>

values for each feature selected during extraction. The resulting values for each recording are stored in comma separated values (CSV) files with the same name of the original audio file and extension ".csv". This procedure with is explained in details in section 4.1.

To allow an easy loading of the CSV files produced during the extraction process, all of them are combined to a single CSV file, which will be used to speed up the access of individual feature values. If the user chooses to summarize this file, a new CSV will be created with each line corresponding to the mean values of each CSV relate to each recording.

3.3 External Libraries

In order to perform analysis over the ecological recordings, audio libraries were compared by number of available features, processing speed and operating system availability. Although some libraries had more features to be extracted, most of them did not applied to the area of acoustic ecology, therefore a solution with a considerable number of features and good speed performance was chosen. The Librosa Audio Library (MCFEE *et al.*, 2015) stood out in the ease of developing a custom features extractor with good processing speed.

Besides the existing features obtained directly form the Librosa library, other features called **Ecological Indices** were implemented on the extraction process, as there was no implementation of them in Python. The algorithm for those indices, listed in Section 3.5.1, were obtained directly from the published papers.

Interactive visualizations have been developed with the **D3.JS** library (BOSTOCK; OGIEVETSKY; HEER, 2011). The D3 is an open-source JavaScript library for manipulating documents based on data, using HTML, SVG and CSS that takes advantage of modern browsers capabilities, combining powerful visualization components and a data-driven approach to DOM (Document Object Model) manipulation. Although there are other solutions available to create interactive visualizations, the **D3.JS** library allows the development of custom built visualizations with fine tuning and custom behaviour tailored to the user specific needs.

Modern browsers prevent users to directly access local files as a security measure, however, Python provides means to create local servers on specific folders. Combining Python built in servers with the Flask framework² allows a centralized server in a specific path to extract features and perform data visualization analysis. Flask is a lightweight web server gateway interface (WSGI) web application framework that makes it easy to create Python applications that respond directly to HTML requests. This approach allows the feature extraction of large volumes of ecological recordings on a usual desktop computer.

² <<http://flask.palletsprojects.com/en/1.1.x/>>

3.4 Visualization Resources

As widely used tool in the field of acoustic ecology, the audio spectrogram visualization is a tool that express the recording frequency and amplitude variations through time. It is calculated by taking samples of the audio and performing a fourier transform on them, with the result being expressed with colors to represent the amplitude of the signal on each frequency. Several tools and libraries are available for specialists in acoustic ecology to produce spectrograms to collections of ecological recordings, as the Python language that provides a few libraries to the calculation of the spectrograms. The Librosa library used on this research is one of them, allowing the user to customize the resulting spectrograms in terms of color palette, time and frequency resolution.

However, the ready tools and libraries leave something to be desired in terms of processing speed. Therefore, to enhance processing speed, the spectrograms generation was performed with the short-time Fourier transform (STFT) alone, taking samples from recordings and saving the results as an image. With this method we were able to generate the spectrogram of a 10 minute recording in less than a second, allowing to produce the necessary resources to use in the interactive visualizations.

The same technique was used to produce daily spectrograms, looping through several consecutive ecological recordings instead of just one audio file. Each vertical pixel line of the resulting image represents the mean of the frequencies calculated by the STFT from each sample, whose length will depend on the resolution chosen by the user during the extraction process.

The original audio files are available for listening during visualizations, allowing the user to associate acoustic events to audio feature values. To reduce the loading time, the original uncompressed audio files are converted to the MP3 format using the FFMPEG library³, available to all operating systems. This conversion is performed only if the user choose to generate the spectrograms during the feature extraction and only if the corresponding MP3 file hasn't been created.

3.5 Seecology: Data Visualization Framework for Soundscape Ecology Applications

A major goal of this research project was to develop a framework to assist the analysis of acoustic ecology data incorporating interactive visualization methods. Achieving this goal required developing several utility resources, including a personalized feature extraction tool presented in Section 3.5.1. In that section we list the audio features available for analysis and a directory and data structure to store the feature vectors, which may be used in multiple visualizations. A method to handle local files in a web environment is described in Section 3.5.2.

³ FFMPEG Library: <<https://www.ffmpeg.org/>>

3.5.1 Feature Extraction Tool

Creating a data visualization requires a previous step of data extraction and treatment, being necessary to resort to appropriate libraries or tools. It is often required for the user to implement code in some language to perform the extraction, as existing tools and libraries do not provide user-friendly interfaces for this process. Current libraries to extract audio features from audio recordings provide ready functions to perform the extractions of each audio feature, however the data input on those functions are left to the user, requiring to split the audio for achieving a detailed perception of the acoustic changes that occur during the recording. Also the manipulation of the resulting data is a manual process, as audio extraction libraries do not provide this functionality.

To solve this problem, a feature extractor was implemented in the framework to facilitate the process of extracting and processing audio files and extract the underwater signatures of boats. Moreover, additionally the data resulting from an analysis is stored in a custom directory and data structure to provide means to create visualizations and perform data analysis. Thus, in addition to the visualizations produced by the Seecology framework, researchers can reuse the feature extraction results to perform other types of analysis or visualizations. Both the feature extraction tool and boat signature extractor presented here can be used as an independent tool for other purposes than data visualization.

Figure 8 shows the main screen of the feature extraction tool with the first 6 features from the 47 available. The first line applies to all the extraction process, allowing the user to select the sample size to be used, which are the length of each sample taken from the recording. A one minute audio file with a 10 seconds sample size will result in 10 samples for that audio, from which will be extracted multiple features for each sample. On the first line, the user can also determine the dB threshold and audio channel to be analyzed. Grouping operations refer to the ability to summarize the data from all samples in a file on a single line in the final file that concatenates all audios processed. The user can choose the normalization method from a list of methods available, removing outliers that may occur in those data and extract spectrograms and MP3 files to be used in the visualizations. The file format is a timestamp used to obtain the date and time of the recording directly from the filename, which can be filled with the assistance of a visual tool included.

Each feature enabled for extraction will result in at least one value per sample of the recording, where some features will produce multiple values, as in the case of MFCC (Mel Frequency Cepstral Coefficients) which originally results in 12 values per sample, although this number can be changed by the user.

Figure 9 presents the screen to configure and extract boat acoustic signatures and consecutive detection of their presence in ecological underwater recordings. On this screen the maximum frequency of analysis can be defined, originally set to $1kHz$ due to that being the

Feature Extraction Settings

☐ ALL

General Extraction Settings

Sample size (Seconds): 2 DB Threshold: 60 Audio channel to use: Left Grouping operation (Main file): None

Normalization operation (Main file): n4: Utilization with zero ... Remove outliers: False Extract spectrograms: True

Extract MP3: True Filename format: %Y.%m.%d_%H.%M.%S Open filename date/time formatting tool

Acoustic Complexity Index (ACI)

Min frequency (Hz): 0 Max frequency (Hz): 22050 Cluster size (Seconds): 1 FFT window size: 512

Acoustic Diversity Index (ADI)

Max frequency (Hz): 10000 Size of frequency bands (Hz): 1000 FFT window size: 512

Acoustic Evenness Index (AEI)

Max frequency (Hz): 10000 Size of frequency bands: 1000 FFT window size: 512

Bioacoustic Index (BIO)

Min frequency (Hz): 2000 Max frequency (Hz): 8000 FFT window size: 512

Mel-Frequency Cepstral Coefficients (MFCC)

Number of cepstra: 12 Min frequency (Hz): 0 Max frequency (Hz): 8000 Use MEL: False

Normalized Difference Soundscape Index (NDSI)

Min freq. Anthrophony (Hz): 1000 Max freq. Anthrophony (Hz): 2000 Min freq. Biophony (Hz): 2000 Max freq. Biophony (Hz): 11000

FFT window size: 512

Figure 8 – Feature Extraction Tool screen with some of the features available for analysis.

frequency range of boat sounds propagation, the audio channel and threshold to be used in the detection process, along with the same timestamp as the feature extraction to obtain time and date from the filename.

If the user intends to use the framework to view the results of this extraction, individual spectrograms should be generated, along with 24H spectrograms, that can be also have defined the maximum frequency visible in the resulting spectrogram images. Also, the 24H spectrograms can also have different resolutions from the 60s originally set. Should the user need to be able to see more details in the final 24H spectrogram image, this resolution can be set as low as 10 seconds, however it is necessary that the computer used in the process has sufficient memory for the task.

Boat Signature Extraction Tool

Extraction Settings

Maximum Frequency of Analysis (Hz): 1000

Signal Detection Threshold: 1.5 x Standard Deviation

Audio Channel: Left

Filename format: %Y.%m.%d_%H.%M.%S Open filename date/time formatting tool

Generate Individual Spectrograms: True Option needed for the visualization step.

Generate 24H Spectrograms: True Option needed for the visualization step (There can be no gap in file dates).

- Maximum frequency (24H Spectrogram): 1000 Leave 0 for a full frequency spectrogram (sample rate / 2)

- Sample window (24H Spectrogram): 60 seconds

Extraction Progress

Begin Extraction

Boat Signatures

Boat Spectrograms

24H Spectrograms

Overall Progress Summary:

Signatures extracted
Spectrograms generated
24H Spectrograms generated

Figure 9 – Boat Signature Extraction Tool screen with available settings and progress bar.

Feature	Description	Package	Feature	Description	Package
ACI	Acoustic Complexity Index	Villanueva-Rivera <i>et al.</i> (2011)	SpecQuartile25 SpecQuartile50 SpecQuartile75 SpecQuartileIQR	Spectral Quartiles	McFee <i>et al.</i> (2015)
ADI	Acoustic Diversity Index	Villanueva-Rivera <i>et al.</i> (2011)	SpecMode	Spectral Mode	McFee <i>et al.</i> (2015)
AEI	Acoustic Evenness Index	Villanueva-Rivera <i>et al.</i> (2011)	SpecSkewness	Spectral Skewness	McFee <i>et al.</i> (2015)
BIO	Bioacoustic Index	Boelman <i>et al.</i> (2007)	SpecKurtosis	Spectral Kurtosis	McFee <i>et al.</i> (2015)
MFCC_0 ... MFCC_11	Mel-Frequency Cepstral Coefficients	McFee <i>et al.</i> (2015)	SpecEntropy	Spectral Entropy	McFee <i>et al.</i> (2015)
NDSI NDSI_Anthro NDSI_Bio	Normalized Difference Soundscape Index	McFee <i>et al.</i> (2015)	SpecVariance	Spectral Variance	McFee <i>et al.</i> (2015)
RMS	Root Mean Square	McFee <i>et al.</i> (2015)	SpecChromaSTFT	Spectral Chroma STFT	McFee <i>et al.</i> (2015)
TH	Temporal Entropy	McFee <i>et al.</i> (2015)	SpecChromaCQT	Spectral Chroma CQT	McFee <i>et al.</i> (2015)
ZCR	Zero Crossing Rate	McFee <i>et al.</i> (2015)	SpecCentroid	Spectral Centroid	McFee <i>et al.</i> (2015)
SpecSpread	Spectral Spread	McFee <i>et al.</i> (2015)	SpecBandwidth	Spectral Bandwidth	McFee <i>et al.</i> (2015)
SpecMean	Spectral Mean	McFee <i>et al.</i> (2015)	SpecRolloff	Spectral Rolloff	McFee <i>et al.</i> (2015)
SpecMedian	Spectral Median	McFee <i>et al.</i> (2015)	SpecPolyFZero SpecPolyFLinear SpecPolyFQuadratic	Spectral Poly Features	McFee <i>et al.</i> (2015)
SpecSD	Spectral SD	McFee <i>et al.</i> (2015)	SPL	Sound Pressure Level	McFee <i>et al.</i> (2015)
SpecSEM	Spectral Standard Error Mean	McFee <i>et al.</i> (2015)	BoatDet BoatSignature	Boat Detection and Boat Signature	Reis, Padovese and Oliveira (2019)

Table 1 – Available features for the terrestrial and underwater extraction process and its origin package.

The data extraction process presented here is performed automatically, with a web interface for each type of recording, terrestrial and underwater. The extraction process for terrestrial and underwater recordings were kept separate in the framework, since they have different goals. As the audio feature extraction aimed to analyze the audio frequencies to create a feature vector for each sample, the underwater extraction goal is to detect the presence of boats in the recordings, as well as the corresponding collection of signature frequencies produced by each boat. Since the boat detection will be presented in Section 4.2, in this section only the terrestrial feature extraction methods will be contemplated.

For the terrestrial recordings, several features used in acoustic ecology and related areas can be extracted with an initial setup, simply by checking a checkbox to include that particular feature, that can be also configured independently. A complete list of the the 47 available features is presented in Table 1.

By selecting and configuring the features, the user can choose to perform analysis on smaller portions of the recording rather than using the entire file as usual. Given a one-minute recording (60 seconds) and a 5-second time window, a single audio file generates 12 samples, which is a more convenient representation of the acoustic events in the recording as far as capturing the variations they introduce in the acoustic features.

Each sample analyzed will be described as a feature vector, in which its components will depend on the user features selection. For each audio file, multiple feature vectors will be produced, with components depending on the user features selection. The resulting feature vectors are combined and stored in a text file (CSV), with additional columns identifying the name of the source file, from there date and time are obtained, and the initial second of the sample relative to the start time obtained by the source file name.

Complementing the ability to extract multiple features for each record sampling window, outlier removal and various data normalization methods have been added to allow researchers to use the audio feature extraction tool outside the framework. A mechanism for summarizing

the data obtained in each file is also provided to reduce the size of the final file size. For each recording, multiple samples will be analyzed, resulting in a CSV file with multiple entries for each recording. At the end of the extraction process, a single file will be produced containing every feature vector calculated, the size of which in bytes can be very large. To prevent this from happening, instead of adding every feature vector produced, a single feature vector with the mean of each recording multiple feature vectors will be added. This will decrease the resulting file size, however, will also reduce the number of samples presented in the visualizations to one sample per recording. However, by clicking in each sample, the user can access the original CSV file from which the mean feature vector was produced, allowing the analysis of the original feature values calculated for each sample.

Since the feature extraction process receives a directory as input containing multiple recordings, computing the date and time of each respective sample is done directly by the file name converted in a timestamp format, added by the starting time of the sample being analyzed. A web interface tool, shown in Figure 10, was added to the Seecology framework to assist in the creation of the timestamp.

Year	without century	Hour	00 - 23
Year	with century	Hour	01 - 12
Month	01 - 12	Hour	0 - 23
Month	Jan	Hour	0 - 12
Month	January	Minute	00 - 59
Day	Mon	Second	00 - 59
Day	Monday	AM/PM	
Day	01 - 31	am/pm	
Day	1 - 31	Ex 1:	2019.12.31_23.59.59 (%Y.%m.%d_%H.%M.%S)

Figure 10 – Timestamp mouting screen

As a result of the entire extraction and standardization process for future analysis, a directory structure on the source path is created in a "Extraction" subdirectory (Figure 11 - Feature Extraction). This new directory contains an "AudioFeatures" directory with CSV files relating to feature extractions from each audio file in the input directory, as well as the starting date, time, and second of each sample taken in that file. To speed up the process of loading audio into the preview internal player, each file is converted to .MP3 format and stored in the AudioMP3 directory along with a CSV file of the same name with the maximum frequency information, sample window size and total file duration in seconds. This last CSV file is used for the spectrogram visualization and stored in the "AudioSpectrogram" folder. In the Extraction

root directory a *parameters.json* file is created with the settings used in the extraction process, log files for possible errors starting with "PROCESSING_ERROR_", dataset information file and files containing all features extracted for each of the sample windows, unified in one place, these files containing the original, grouped and normalized data. The settings of files containing grouped and normalized data are user-defined at the time of extraction. After the extraction process is completed, a new file in CSV format is generated, grouping all the contents of the individual feature files produced.

Figure 11 – Directory structure containing the files with the feature vectors extracted, related to each audio sample, and visualization resources produced.

On the same "Extraction" path, a directory structure with corresponding files are create to accomodate the results of a boat detection process (Figure 11 - Boat Detection). Instead of audio features, the boat detection process, inside a "detection" subdirectory, produces 4 files for each recording processed with results for every minute inside the recording. The content of each files are a detection summary, the boat signatures detected, the mean spectrum and the variance signature (FAV signature). A subdirectory "BoatSpectrogram" is created to store the spectrogram and corresponding CSV file, exactly as the "AudioSpectrogram" folder of the feature extraction process. The reason for two separate directories is due to the user defined

maximum frequency set for analysis on the boat detection process. For each 24 hours of data, a spectrogram containing a summary of the frequencies can be produced for quick inspection of acoustic events occurring in daily underwater recordings, stored in the "24HSpectrogram" folder. In the Extraction root directory, a *BoatDetectionSummary.csv* file is created with the detection information of all samples analysed on the recordings, as well as log files for possible errors on the detection and 24H spectrogram generation process, starting with "PROCESSING_ERROR_". The "DetectionImage" folder is only filled with images for each minute analysed by user request, and it is usefull to debug the extraction process, visually testing the results of the sensitivity adjustment.

The advantage of the proposed feature extraction method is that a user can test alternative extraction configurations in order to determine which is the more effective for characterizing acoustic events of interest. Whether filtering a specific frequency range, changing sample window size settings, sensitivity adjustment, or combinations of various elements. The result can be immediately analyzed in the visualizations available in the framework, this process can be repeated until a satisfactory configuration of features produce a satisfactory result in characterizing acoustic events of interest. In addition, the subdivision of large recordings into 1 minute files may help the analist to visualize acoustic events of interest more easily than using the original recording.

3.5.2 Accessing Local Files In Web Environment

Current browsers prevent access to local files on the client side for security measures. As we intended to develop a web-based interface for the framework, it was necessary to find a solution to access local directories with the ecological recordings.

The problem was solved with a combined solution of a Python's native web server and the Flask library⁴. The Flask library is used to generate a webserver to provide the interface pages, and receive requests from directly from those pages without requiring an intermediate language to handle calls to Python function such as PHP, allowing direct access to all the Python language libraries and capabilities. Once the system is initialized, two web servers are instanced, one for the framework functions and interface and one to support access to local files. This configuration allows read and write access to local files.

⁴ <<http://flask.palletsprojects.com/en/1.1.x/>>

3.6 Automatic Detection of Vessel Signatures in Audio Recordings with Spectral Amplitude Variation Signature

A first step in automatic detection of boats, ships and other types of water crafts in underwater acoustic recordings is to distinguish their acoustic signatures from those due to natural elements such as fish choruses or crustacean acoustic activity. Audio signals emitted by vessels can be characterized by narrow stationary frequency signatures, as the engine, propellers and cavitation produce sounds that usually remain stable along the duration of a single recording.

Figure 12 – Logarithmic scale spectrograms of two underwater recordings. The red squares mark frequency peaks due to boats and the blue squares mark peaks due to fish choruses. Both recordings include background noise due to acoustic emission by crustaceans, that can be perceived as vertical lines in the spectrograms. (A) 15-minute audio spectrogram with two types of fish chorus and a boat signature characterized by three peaks; (B) 15-minute spectrogram with a single boat signature characterized by 10 peaks and heavy background noise due to crustacean emissions.

The audio spectrograms shown in Figure 12 illustrate occurrences of three acoustic events captured underwater, namely boats, indicated by the red squares, fish choruses, indicated by the blue squares and background noise from crustaceans, visually perceptible as vertical lines in both spectrograms. The spectral signatures of both the fish choruses and the boats are characterized by frequency peaks that persist along extensive time periods. Even though they may appear similar, the frequency peaks due to boats appear over narrow frequency bands, unlike the fish choruses, which are observed over wider bands. The distinction poses a challenging scenario for automatic boat detection, as boats and fish choruses can be easily mistaken, and boat events may be totally masked by strong background noise.

Our proposed method for automatic vessel detection in underwater ecological recordings considers certain unique features of the sounds emitted by vessels to discern their characteristic frequency peaks from those produced by other acoustic events. The algorithm considers as input the *dB* frequency spectrum of the audio signal segmented into one-minute audio samples. It initially computes a so-called **frequency amplitude variation** (FAV) signature from each one-minute segment of the audio spectrum.

The overall process, as well as the motivation for our choices in deriving the proposed algorithm are detailed next. Unlike the DEMON method, which considers the audio envelope of the signal, our algorithm relies on information obtained from the audio spectrogram matrix. The audio spectrogram is computed with the *Short-Time Fourier Transform* (STFT), using as input the audio signal and an FFT size, which was established as 11025, equal to the audio sampling rate, in order to yield a nominal frequency resolution of 1Hz while resulting in one FFT sample per second. The resulting spectrogram matrix thus has $F = (11025/2)$ frequencies and $S = (N \times 60)$ samples, or one sample for each second of the input audio, where N is the audio duration, in minutes. Our analyses have been conducted on 15-minute recordings.

Algorithm 3.6 describes the necessary steps to obtain the FAV signature and the corresponding boat detection signature.

Algorithm 1 – FAV signature computation

Input:

1	Sample [0..60][0..Rate/2]	▷ 1-Minute Spectrogram matrix
2	Rate	▷ Audio sampling rate
3	MaxFreq	▷ Maximum frequency of analysis
4	SDMultiplier	▷ SD threshold multiplier

Result:

4	FAV	▷ Frequency Amplitude Variation signature
5	BoatDetection	▷ Boat detection signature

```

6 Mean[0..Rate/2] ← frequencyMean(Sample)
7 SmoothMean[0..Rate/2] ← smooth(Mean, blackman, 8)
8 SD ← StandardDeviation(SmoothMean)
9 Mean ← Mean[0..MaxFreq]
10 SmoothMean ← SmoothMean[0..MaxFreq]
11 FAV[0..MaxFreq] ← 0
12 BoatDetection[0..MaxFreq] ← 0
13 for i ← 0 to MaxFreq do
14   variation ← SmoothMean[i] − SmoothMean[i − 1]
15   if variation > 0 then
16     | variation ← 0 ▷ Post frequency peak
17   end
18   FAV[i] ← |(variation)3|
19   if FAV[i] > (SD × SDMultiplier) then
20     | BoatDetection[i] ← 1
21   end
22 end
23 BoatDetection[0..20] ← 0
24 BoatDetection[(MaxFreq − 10)..MaxFreq] ← 0
25 return FAV, BoatDetection
```

As described, the spectrogram matrix covers N one-minute samples of the audio recording. The frequency amplitudes of each one-minute sample are averaged in order to obtain a single representative frequency amplitude. The result is a spectrum signature of the one-minute audio segment stored in an array of size F , where each entry corresponds to an average frequency amplitude (Algorithm 3.6, lines 6-7). Computing the average spectrum signature of one-minute audio samples, instead of the entire audio, avoids missing short duration events. It also allows precisely locating the events along the audio duration, which is useful for analysis and visualization purposes.

The spectrum signature is the input to the FAV signature computation, which is at the

core of the boat detection procedure. The size of the spectrum signature is equal to the maximum frequency component in the recording, as it has been obtained considering F , however, it can be set to a maximum frequency of analysis (Algorithm 3.6, lines 9-10). The spectrum signature is thus stored in an array in which each entry corresponds to a frequency window of 1Hz. Let $SmoothMean[0...F-1]$ denote the array storing the smoothed spectrum signature, with F values. It is possible to compute the differences between each pair of frequency amplitudes in the signature, applying Equation 3.1 to each element in this array (Algorithm 3.6, lines 14-16).

$$FAV_f = |(SmoothMean_f - SmoothMean_{f+1})^3| \quad (3.1)$$

Equation 3.1 yields the unsigned differences between consecutive frequency pairs of the input signal (Algorithm 3.6, line 18). The resulting differences, raised to power 3, are stored as elements of a new array FAV , of size $F-1$, which stores the FAV signature. The reason for raising the resulting differences to power 3 is to emphasize the frequency peaks due to boats against the peaks due to other events that also introduce frequency amplitude variations. Raising to other power values could yield a similar effect, but empirical investigations led us to this choice.

The FAV signature captures and emphasizes persistent variations in the frequency amplitudes contained in the input spectrum. Such variations can yield a single peak or multiple frequency peaks, resulting in different patterns characteristic of certain acoustic events, including boats. Our goal is thus to identify the peak patterns characteristic of acoustic emissions by boats.

A threshold of $\alpha = 1.5 \times \text{Standard Deviation}$ of the original spectrum has been empirically found suitable to identify the frequency peak patterns due to boats, even in noisy conditions. The algorithm applies this threshold to the FAV signature (Algorithm 3.6, lines 19-20) to yield a binary boat detection signature array, also of size $F-1$, applying Equation 3.2:

$$BoatDetection_f = h(FAV_f, \alpha) \quad (3.2)$$

where

$$h(FAV_f, \alpha) = \begin{cases} 1, & \text{if } FAV_f \geq \alpha \\ 0, & \text{otherwise} \end{cases} \quad (3.3)$$

$$\alpha = \left(\sqrt{\frac{1}{n} \sum_{i=1}^n (SmoothMean_i - \mu)^2} \right) \times 1.5 \quad (3.4)$$

$$\mu = \left| \frac{1}{n} \sum_{i=1}^n SmoothMean_i \right| \quad (3.5)$$

The result is a binary array *BoatDetection*, in which each entry is flagged to 1 whenever the corresponding entry in the FAV signature array is above the threshold ($SmoothMean_n > \alpha$), signalling this as a peak indicative of the presence of a boat, otherwise the corresponding array entry remains set as 0.

The acoustic events introduce sound amplitude variations in the audio signal spectrum, which can be observed in the form of characteristic patterns of peak amplitudes. The underlying rationale in computing the FAV signature is to identify the peak patterns due to boats and ignore those related to other acoustic events captured in the recordings. Our method identifies a characteristic signature of each one-minute audio sample. It is thus possible to distinguish the patterns due to boats from those originated from overlapping acoustic events. Once the frequency peaks have been identified in the FAV signature, a thresholding operation can be applied to obtain a binary array with the boat detection signature.

Figure 13 – In both charts the blue lines depict, respectively, the frequency spectra (up to 2.7kHz) of an input signal in dB (A) and in dB^2 (B); the green and red dotted lines depict the spectral signal smoothed by convolution with a size 11 blackman window, scaled respectively by $1.5\times$ and by $3\times$ its standard deviation. The vertical watermarks in red identify the frequency peaks due to boat emissions.

Since our algorithm measures the differences in amplitude between neighboring frequencies, it performs better with a standard spectrum in dB scale, in which the amplitude variations are smoother. Measuring this frequency variation along the audio spectrum will result in a

characteristic signature with amplitude peaks due to noticeable acoustic events. Operating on the *dB* scale audio spectrum (Figure 13-A) facilitates distinguishing the peaks due to boats from those introduced by other acoustic events.

Figure 14 illustrates two signatures computed with this approach for two one-minute samples from the recordings depicted in the spectrograms in Figure 12. The charts show the spectrum signature (blue line), the FAV signature (green line) and the boat detection signature (red vertical watermarks).

Figure 14 – Spectrum signature (blue line), FAV signature (green line) and binary boat detection signature (red marks) of (A) a recording that includes fish chorus and a boat; and (B) a recording that includes a small boat at minute 11. The horizontal dotted purple lines indicate the threshold defined to associate frequency peaks with acoustic emissions by boats.

The signatures in Figure 14-A refer to the first minute of the recording depicted in Figure 12-A, which includes a fish chorus overlapped with a ship or large boat, whereas those in Figure 14-B refer to minute 11 in the recording depicted in Figure 12-B, which includes a small boat and heavy noise due to crustacean acoustic emission. One observes in the FAV signature in

Figure 14-A that only frequency peaks due to boats stand out, whereas other acoustic events are largely ignored. The FAV and boat detection signatures depicted in Figure 14-B show that the algorithm detected the faint signal of the small boat even in the presence of heavy background noise.

Our method to process an input audio database consists of the following steps. Given a root directory path containing the audio files to be processed, first the signatures are computed for each individual recording applying Algorithm 3.6, i.e., the spectrum, the frequency amplitude variation (FAV) and the boat detection signatures. The resulting signatures for each file are stored in a corresponding CSV file that preserves the source file name. The algorithm also saves summary information on each boat detection signature, namely the number of peaks identified, the shortest distance between the peaks, and the lowest and highest amplitude peak values – this information may be useful in further vessel classification procedures. The CSV signature files are saved in a child folder of the root directory; the summary information is stored in a single CSV file saved in the root directory.

The process requires two parameters: the sampling rate of the audio recordings, which is known *a priori*, and a boat detection threshold. The default threshold, which yielded good results in the databases studied, has been set empirically to 1.5 times the standard deviation of the amplitude spectrum of the signal. Thus, any peak that surpasses this threshold is interpreted as a boat detection.

3.7 An Exploratory Approach For Annotating Acoustic Events

The identification of acoustic events is an important task in the field of acoustic ecology, where the correct characterization of a series of acoustic event of interest could express valuable information about an environment. Several methods and tools were proposed in the literature to assist researchers and specialists with the task of annotating the occurrence of acoustic events of interest. Many methods use techniques based on the manual analysis of ecological recordings using spectrograms, which can be effortly demanding since each recording needs to be analyzed individually, as this records may contain a large number of events occurring simultaneously (Truskinger *et al.*, 2014). Another problem with manual inspection of ecological recordings is the size of the task, limited by the human ability to precisely provide large-scale spatiotemporal observations (WIMMER *et al.*, 2012).

The method for annotating acoustic events of interest in ecological recordings proposed here, is based on a Concentric Radviz visualization (ONO *et al.*, 2015) in conjunction with the custom audio feature extraction included in the Seecology framework. As each sample appearing in the the visualization area corresponds to a single input in the CSV file, produced during the extraction process, by exploratory analysis and performing lasso selections of samples, users

can annotate sample sets belonging to acoustic events of interest. The method for calculating sample positions in our visualization was changed to a sum of vectors, as in the original Radviz, instead of using the force simulation implemented on the work of [Ono *et al.* \(2015\)](#) to reduce computational time, allowing a larger number of samples per visualization.

The analysis of acoustic events with similar characteristics, assisted by the Concentric Radviz visualization, is performed using the user's perception in the joint movement of the samples during the rotation of the dimensional groups. Although the samples are presented as a bidimensional image, rotating a dimensional axis causes a motion effect that simulates a three-dimensional rotation. Thus, groups of samples move at different speeds within the visualization, allowing the user to perceive acoustic events clusters hidden in the data (Figure 16).

As in the work of [Ono *et al.* \(2015\)](#), we use a sigmoidal function to obtain a better sample distribution in the visualization, while avoiding common issues of Radviz like samples with all features values equals to each other being positioned close in the visualization. The sigmoidal function enable an adjustment of the samples feature values, temporarily redistributing the samples in the visualization area and facilitating the recognition of samples of interest.

Figure 15 shows the effects of adjusting the feature values with the sigmoidal function on samples spatial placement, where in Figure 15-A the sigmoidal function default values for width and height are presented, resulting in the original samples placement when users add features as dimensional anchors to the dimensional axis. Figure 15-B shows the effect of changing the width and height respectively to 0.15 and 10, and Figure 15-C presents the same effect with width and height values respectively set to 0.65 and 10. The horizontal axis in the sigmoidal function visualization represent the original feature values from 0 to 1, which after been readjusted by the function will result in the values presented in the vertical axis. Two values in the horizontal axis, 0.3 and 0.7, were marked with a red dot and gray dotted lines to demonstrate the resulting values in each case, represented in the vertical axis. It is noticeable that new sample clusters are formed depending on the values of the sigmoidal function, even without rotating the dimensional axis, which will reveal hidden groups in the visualization.

Rotation of dimensional axes allows observing clusters of samples with similar behavior regarding a set of selected features. Besides the placement of the dimensional anchors in Radviz, the visualization admits two types of rotation, namely by dimensional axis or by dimensional anchor. A rotation on a dimensional axis rotates all dimensional anchors included in it. If multiple dimensional axis are being displayed in Concentric Radviz, the joint movement of samples will reveal similar groups. This behavior is due to the samples movement variance indicating a measure of dispersion relative to the selected features, allowing the user to explore data by rotating the dimensional axis, until he can identify clusters that visually stand out from other samples.

Figure 16 shows the effect of rotating a dimensional axis on the samples spacial placement in the visualization. A group of samples was selected to demonstrate the effect of a 90°

Figure 15 – Effect of the sigmoid function on sample spatial placement in the Concentric Radviz visualization. Original feature values are presented in the horizontal axis and resulting values after the use of the function are presented in the vertical axis. The values 0.3 and 0.7 were marked with a red dot and gray dotted lines to demonstrate the original and resulting values depending on the sigmoidal function values.

degrees rotation in one of the dimensional axis on samples position. The selected samples did not characterize any particular acoustic event, however it is noticeable that after rotating a dimensional axis, represented by a black arrow in Figure 16-C, D and E, it is possible to see that different acoustic events were overlapped, making their selection difficult. After rotation, it is possible to characterize four sample clusters from the selected samples that originally were not able to characterize individual acoustic events. The new sample placement allows them to be selected for analysis by lasso selection. Should the samples belong to the same acoustic event, the user can annotate them with a chosen nomenclature.

By clicking on a sample in the visualization, an audio player will open the respective recording from which the sample was extracted. This audio player, shown in Figure 17, contains the representation of the recording spectrogram aligned with the audio waveform, with a watermark highlighting the selected sample. Underneath the spectrogram, a line chart with the corresponding feature values extracted for each sample of the recording is displayed. The user can select any feature for analysis by choosing from a dropdown, containing all the audio features extracted for the dataset. Once a feature is chosen to appear in the line chart, clicking other samples will open the same selected feature.

The ability to hear the original audio from which a sample was taken, associated with

Figure 16 – Effect of the rotation of a dimensional axis over the dataset samples, represented with a black arrow originating from the original position of the first dimensional anchor. Concentric Radviz A presents the original sample placement, B with a clockwise rotation of 30° , C with a rotation of 60° and the final placement of the samples after a 90° clockwise rotation.

its spectrogram representation, is useful for identifying acoustic events. Being able to relate these acoustic events to their respective response in the feature vectors, provides the researcher with the means to choose new features that best suit the characterization of acoustic events of interest. Some acoustic events are difficult to see in the spectrogram due to their frequency ranges being predominantly low, mixing with the background noise of the recordings, such as airplanes and howling wolves. However by using our extraction method and the Concentric Radviz visualization, we were able to be characterized those events by using this method.

In Figure 17 two different one minute recordings are presented to demonstrate the usefulness of the small sample windows. For this example, the MFCC feature with 12 components was used, as this feature takes into account human perception for sensitivity at appropriate frequencies by converting the conventional frequency to Mel Scale, trying to model the way that the human hearing auditory system works. On both, samples were extracted from the recordings using a time window of 2 seconds, resulting in 30 samples for each one minute recording. Figure 17-A presents a recording with cicada noise and a very faint insect presence around 2.2kHz, with the 5th MFCC component (MFCC_4) being displayed in the line chart underneath. It is possible to perceive the effect of the cicada noise in the MFCC component, where the louder the noise, the higher the component value.

In Figure 17-B is presented a recording with insects and birds in the background with rain noise in varying intensities, with the 1th MFCC component (MFCC_0) being displayed in the line chart. During our tests, it was evident that rain noise is very well reflected in this component values. The variation of the 1th MFCC component is noticeable along with the increase in rain noise in the recording, also visible in the spectrogram.

A useful tool incorporated in the visualization is the ability to select samples lasso selection (Figure 18). When selected, the samples turn red and larger in diameter. The selected samples are presented in a list on the left, just under the spectrogram, with information about the original file name and sampling initial time within that file in seconds. The user can view the

Figure 17 – Audio player associated with a spectrogram for the full recording and a line chart to visualize feature values variation on the samples. The sample window on both player was set to 2 seconds during the extraction process. The sample selected on each player is represented as a vertical red watermark. On A, a recording with cicadas is presented, and in B a recording with the presence of rain in different intensities.

Figure 18 – Sample selection on the visualization and list of each recording and sample of the respective selection.

spectrogram in the player by clicking on the buttons to the right of the sample, namely **Small** and **Big**.

The **Small** button will load the audio player with the respective recording of the sample, as the **Big** button will result in the same audio player, however with a larger spectrogram and a window covering the main visualization. This larger spectrogram could be useful in some cases

as the magnified view can assist in identifying shorter acoustic events.

The selected samples can be analysed by the user in order to determine if they correspond to the same acoustic event, that if true, the user can insert a label with the name of the acoustic event. At this point the new annotation will be checked by traversing the previously registered so that there is no duplication. If a label with the same name already exists the samples are added to it, otherwise the new label is included and associated with the samples. It is worth noticing that a sample may have multiple associated labels, in the current implemented visualization, only the most recent one is displayed.

3.8 Final Considerations

This chapter introduced the materials and methods employed in developing the techniques and perform case studies in order to answer the research questions raised on Chapter 1. We presented the equipment used to develop and test methods on this thesis, as well as to perform case studies (Section 3.1), a description of the ecological recordings used in this research and resulting datasets obtained from them (Section 3.2), external libraries for audio, visualization and web applications used in the development of the Seecology framework (Section 3.3) and the methods and techniques for reducing computational costs in the generation of visualization resources (Section 3.4). Also, methods developed for the visualization framework (Section 3.5), detection of boats in underwater recordings (Section 3.6) and annotation of acoustic events (Section 3.7) were described in detail. Moreover, the next chapter presents the results and case studies conducted as part of this research.

RESULTS

This chapter is organized in three sections: In Section 4.1 the implementation of the visualization framework for acoustic ecology analysis named Seecology, is presented with the resulting user interface of the system presented in detail. The visualizations included in the Seecology framework are introduced briefly, and are further detailed in Sections 4.2 and 4.3.

A case study applying the method FAVs for automatic detection of vessel signatures in audio recordings, described in Section 3.6, is presented in Section 4.2, along with a comparison with the DEMON (Detection of Envelope Modulation On Noise) technique to demonstrate the effectiveness of the FAVS method in both speed and accuracy. In Section 4.3, results on the annotation of acoustic events are presented, applying the method described in Section 3.7, with a Concentric Radviz visualization approach.

Final considerations are presented in Section 4.4.

4.1 Seecology: Data Visualization Framework for Soundscape Ecology Applications

The implementation results, encompassing the proposed methods for a visualization framework for soundscape ecology applications (Discussed in Section 3.5.1 and Section 3.5.2), are described in this section. The user interface and available visualizations are described in Section 4.1.1, depicting a line chart to perform analysis on each extracted audio feature and detection of data outliers (Section 4.1.1.1), the Concentric Radviz for acoustic events annotation (Section 4.1.1.2) and the visualization of the data resulting from the boat detection method (Section 4.1.1.3).

4.1.1 User Interface and Visualizations

Using available libraries to extract features from audio typically requires some programming effort, particularly if it involves segmenting the audio files in smaller samples from which the features will be extracted. Concatenating the results into a ready-to-analyze format may also demand considerable time and human effort. The Seecology framework aims to ease the process for anyone interested in analyzing audio features without the burden of creating personalized code.

The Seecology framework delivers a simple user interface for acoustic ecology data analysis and visualization, using a custom web server to provide easy access to local files as input data. This approach allows for experimenting with multiple extraction settings to achieve one that is more suitable for characterizing acoustic events of interest. The interactive visualizations included in the framework provides means for researchers to carry out analyzes and annotations on the extracted data without the need for knowledge of an additional tool. Resulting data and resources from the feature extraction tools can be used by researchers outside the Seecology framework.

Figure 19 – Working directory selection screen.

To begin the extraction process, as soon as the framework is started, the user is presented with the Seecology logo and requested to select the working path containing the target ecological recordings (Figure 19).

Once the working path is selected, the user is presented four options, a tool to split audios into one-minute files, a generic audio feature extraction tool, a boat feature extraction tool and the available visualizations. For each presented option, the user can access areas with specific goals, such as feature extraction configurations, boat detection or visualizations of available data, in order to perform data analysis or create annotated datasets. Figure 20 provides an overview of the framework areas, accessible from the home screen.

The spectrogram is a tool, highly used by specialists to visually determine acoustic events of interest that may appear in the recordings, however depending on the length of the audio and

Figure 20 – Framework areas accessible from the home screen.

the width of the generated spectrogram, viewing short duration acoustic events can become a difficult task.

As the Seecology framework is capable of delivering feature extraction of samples up to a second in length, and the built-in visualizations provide an audio player for associating acoustic events to feature values, the user can choose to split the input ecological recordings into one-minute files so acoustic events can be more easily seen on the spectrogram.

To perform this task with the built-in audio splitter, the user is required to insert the audios filename format so the system can extract date and time information. To ease this step, a visual script (Figure 10) eliminates the need to learn the timestamp format. The user will be presented with a screen where each button pressed will add a part of the filename to the timestamp.

For example, given the file *"2019.12.11_17.20.50.wav"*, the user must determine the format of each segment, such as the year with four digits, month and day with two digits, and time in 24H format. A few symbols are also provided in the screen such as underline, comas and points, however the user can at anytime add a custom symbol directly to the screen where the timestamp format will be assembled. A real-time example of the filename will appear on the top of the screen for user validation, once the timestamp format is finished the user can copy it to the "Filename format" field in the previous screen (Figure 21) and start the audio split process by clicking in "Begin Split The Audio Files".

The framework will create a new subfolder called "MinuteFiles" and saves the new one-minute files there with the filename added the time correction to match the beginning of the

corresponding audio part in the new recordings. Also, the current working path is changed to this new folder, leaving the original recordings intact.

Figure 21 – Screen used to split audios into one-minute files.

The generic feature extraction and boat feature extraction processes included in the Seecology framework were both explained in Sections 4.1, 3.5.1 and 4.2. The extraction tools can be accessed through the home screen at any time by clicking on the "HOME" menu on the top left of every screen and selecting the corresponding area.

Although both extractions are applicable any type of ecological recording, the boat feature extraction tool (Section 4.1) is designed for a specific task, detection of the presence and corresponding signatures of boats in underwater recordings. The same is true for using underwater recordings with the audio feature extraction tool (Section 4.2), although audio features could be extracted, the nature of the underwater and terrestrial recordings differ from each other, and most features would not provide meaningful results. Therefore, it is up to the researcher to determine the correct feature extraction tool for his research.

The tools in this framework were structured in independent areas due to its goals. The split of ecological recordings in one-minute audio files is optional for the feature extraction process, however can be useful to produce more detailed spectrograms. Both feature extractors included are designed for different underwater and terrestrial problems, since the recordings present unsimilar characteristics, as the analysis tasks to be performed on them.

The visualizations were kept apart from the feature extraction process and designed to receive the same data as input, which facilitates the development of new visualizations. If the user finishes the feature extraction process and closes the framework, the resulting data is available for analysis once the framework is restarted.

There are three working visualizations currently available in the Seecology, a line chart (Section 4.1.1.1), a Concentric Radviz (Section 4.1.1.2) and the boat detection visualization (Section 4.1.1.3).

4.1.1.1 Line Chart

A common line chart visualization is available on the Seecology framework (Figure 22), providing means to observe the features behavior through recordings duration. By selecting a

feature, the values for all samples are presented, allowing the identification changes in the values, which can be analyzed by clicking the visualization at any point of the line chart. This action will open an audio player window (Figure 23), with the spectrogram representation of the original recording, along with a line chart of the current selected feature for all the samples taken from that recording. A red watermark over the spectrogram representation indicates the sample clicked by the user on the line chart.

Figure 22 – Line chart visualization of recordings from 2015-03-06 depicting the comparison of three features extracted (MFCC_0, MFCC_9 and RMS).

This visualization allows the user to save the current state of the line chart for comparison with other features previously extracted. This line chart exploration allows the user to listen to specific events that could result in variations on the features, and the ability to produce static representations could be used to create a set of features that would form a good combination in the task of segregating specific acoustic events of interest.

4.1.1.2 Concentric Radviz

The Concentric Radviz visualization, presented in Section 4.2, was designed with the goal to enhance the hability to segregate samples with similar acoustic events, and consequently, similar feature values. Through data exploration using features distributed along concentric circles, the user can easily experiment with multiple feature combinations in order to achieve a better segregation of a specific acoustic event.

To simplify the feature selection process, a drag and drop user interface was implemented, allowing the user to manually drag features from the left collumn (Feature List) into five dimensional axis (Layers), which will appear in the main visualization area as they are loaded with a minimum number of features, being three for the first dimensional axis and two for the others.

Figure 23 – Modal window with an audio player containing a spectrogram representation of the original recording for the sample clicked on the line chart visualization..

Requiring a minimum of three features for the first dimensional axis, ensures better dispersion of the samples in the visualization area, while only two features would display the samples over a straight line between the two dimensional anchors placed at opposite sides of a dimensional axis.

Figure 24 – Main screen of the Radviz Concentric visualization with feature selection on the left tab.

Figure 24 shows the main Concentric Radviz visualization screen with an extracted dataset and some samples already annotated. The samples are distributed in the visualization using three dimensional axis, each composed with features as dimensional anchors. In Figure 24 both the first and second dimensional axis have three dimensional anchors, which results in a

33.3° between each. The third dimensional axis, however, has only two dimensional anchors, resulting in a 180° between the two of them.

Each of the five dimensional axis available are represented with a different color, as well as each dimensional anchor belonging to that axis. By clicking and dragging the mouse on any of the dimensional anchors, it is possible to rotate the dimensional axis and recalculate the position of each sample of the visualization. Exploratively, rotation of different dimensional axes containing different features may result in better segregation of samples that have similar characteristics for the selected features.

Figure 25 – Tabs for Feature selection (A), labels and filters (B), and sigmoidal function (C) to remap the sample values in the visualization.

When initializing the visualization, the user is presented with the first tab listing the available features, selected during extraction, along with colored boxes corresponding to layers (Figure 24), or dimensional axis, in the Concentric Radviz. Three more tabs are available to help user explore, analyse and annotate datasets. The second layer (Figure 25-A) presents the sample visualization tool, composed of the player with line chart to listen and visually explore the spectrogram of the recordings, and a list of samples that have been selected on the radviz by circling them with the mouse. Each sample selected by the user is presented in the list with the filename from which the sample was extracted, the position in seconds of the start of the sample within the original recording, two buttons, namely "Small" and "Big" that open the sample respectively on the player above the list and on the modal window with a larger version of the same player with line chart depicting the feature values. The features depicted in the line

chart can be chosen with a drop down menu on top of the spectrogram, and the current sample is represented in the spectrogram by a red watermark covering the entire sample duration.

After an analysis of the selected samples, the user can select which ones represent the acoustic event of interest, in order to create a label with a chosen name, that can be inserted in the text area on the bottom of the tab (Figure 25-A).

On the third tab (Figure 25-B) the user can chose to enable or remove previously created labels, which are represented in the visualization samples with different colors. Whenever the user select a checkbox on the "Enabled" collumn, the corresponding samples in the visualization will apper with the color designed for its label. By clicking on the "Remove label" button, a window will appear asking if the user is sure, case positive the label is removed.

The fourth tab (Figure 25-C) contains the sigmoidal function controls with the function of remaping the feature values for a possible better segmentation of the samples in the visualization. By changing the center and width of the sigmoidal function, the samples will be repositioned within the Concentric Radviz visualization.

With this interactive visualization, researchers can quickly create annotated datasets to perform studies with machine learning or acoustic event classification for example.

4.1.1.3 Boat Detection

The algorithm for boat signatures detection in underwater recordings presented in Section 4.1 is implemented in this proposed framework with a custom extractor (Section 3.5.1), requiring a minimal user configuration as input to perform the detection process, including the boat signatures (FAV) of the detected boats. Besides the user interface provided for the extraction of boat signatures and consequent detection of boat presence in the recordings, the visualization included in the Seecology framework allows the user to quickly verify the presence of boats through daily spectrograms and individual analysis of the recording files.

Figure 26 – Boat detection visualization.

Figure 26 shows the visualization for inspection of the presence of boats, containing a daily spectrogram on the right produced during the extraction process with the mean values of a user selected sample length, from 10 to 120 seconds of audio signal in the recordings. Just below the daily spectrogram, a line chart depicting the number of boat signature peaks detected in every one minute sample is presented to inform the presence of boats in those recordings.

If the user select part of the daily spectrogram by dragging the mouse on the daily spectrogram (Right), a list containing the audio files corresponding to the date and time selection will be presented on the bottom-left of the visualization. By selecting an audio file from this list, a spectrogram of the selected file will be presented above the list, on the top-left region of the visualization. Similar to the daily spectrogram, these spectrogram representing each recording will also presents a stepped line chart, with the number of boat signature peaks detected from each one minute sample of that recording. At any time the user can select another day for analysis by choosing from the dropdown menu on the top-left of the visualization, namely "24H Spectrogram List".

Since underwater recordings do not have a wide range of acoustic events, such as terrestrial recordings, and the frequencies analyzed range from 0 to 1000 Hz with long-duration acoustic events, the player added in other Seecology visualizations has not been included.

4.2 Automatic Detection of Vessel Signatures in Audio Recordings with Spectral Amplitude Variation Signature

The method has been validated on audio recordings collected at two sites located on the West coast of the State of São Paulo, namely the Laje de Santos Marine State Park (LSMSP, located at $24^{\circ}15'48''S$, $46^{\circ}12'00''W$) and the Xixová-Japuí State Park (XJSP, located at $24^{\circ}0'22''S$, $46^{\circ}23'29''W$). Both sites exhibit boat traffic and are characterized by great marine biodiversity. The LSMSP is 30km (19 miles) off the coast and serves as a protected area for reposition of fish shoals and safe reproduction of species. The XJSP conservation unit encompasses a marine area SW of the Santos Bay and an adjacent inland region of tropical forest. This is an area with strong human presence, close to a very busy port in Brazil (SÁNCHEZ-GENDRIZ; PADOVESE, 2017; SÁNCHEZ-GENDRIZ; PADOVESE, 2015).

The datasets consist of 15-minute monaural audio files recorded at a sampling rate of 24 kHz. Recordings were obtained underwater with a custom autonomous hydrophone recorder developed at LACMAN (Laboratory of Acoustics and Environment of the Polytechnic School of the University of São Paulo)¹.

¹ <http://lacmam.poli.usp.br>

In total, we analyzed 2,675 recordings from LSMSP and 2,390 recordings from XJSP. As general observations, we noticed that the sound levels of the recordings taken at XJSP are, in average, 10dB higher than those taken at LSMSP. This is possibly because XJSP is closer to the coast and has intense boat traffic, whereas LSMSP is located in a quieter area farther away from the coast. We also observed in some XJSP audio files a recording failure in the low frequencies up to 100Hz. Both databases include many recordings with multiple occurrences of diverse boats and ships, under different conditions. Our goal was to assess the effectiveness of the proposed algorithm in automatically detecting the presence of ships and boats on recordings collected from February 1 to February 28, 2015 (LSMSP site) and from February 4 to February 28, 2015 (XJSP site).

Figure 27 – 24-hour spectrogram (frequency band from 0 to 2KHz) of the audio recordings taken at LSMSP in February 17, 2015. The recordings captured three types of fish chorus (A, B and C), as well as long and short duration sound emissions by boats and ships (D). The line chart at the bottom shows the detections obtained with the proposed method minute-by-minute and the number of peaks that characterize the corresponding boat signature.

The spectrograms in Figures 27 and 28 depict, respectively, recordings taken at two arbitrary days, namely February 17 at LSMP and February 10 at XJSP. These particular days have been selected as illustrative examples of the multiple acoustic events that can be found in both databases. The figures also illustrate the outcome of the proposed boat detection algorithm. Detections are displayed in the form of a line graph at the bottom and temporally aligned with the corresponding audio spectrogram.

The red line indicates the boat detections at their exact location in time along the duration of the audios, and also informs the number of frequency amplitude peaks associated with the boat event, as detected by the algorithm. In the examples, the number of peaks varied from 0 (no boat detected) up to 5 peaks in the recordings taken at the LSMSP site (Figure 27); or from 0 up

Figure 28 – 24-hour spectrogram (frequency band from 0 to 2KHz) of audio recordings taken at XJSP in February 10, 2015. The recordings captured three types of fish choruses (A, B and C) and sound emissions by multiple ships (D) appearing as long and short duration acoustic events. The detection line chart shown under the spectrogram reveals a concentration of short duration boat events in the period from around 9:00am to 11:00am; their corresponding signatures are characterized by up to 13 frequency peaks.

to 13 peaks in the recordings taken at the XJSP site (Figure 28).

The boat detections identified have been confirmed by listening to the corresponding audio files. In all cases, a manual inspection confirmed the automatic detection has been successful. In a few and very specific cases false negatives occurred, with the algorithm failing to detect a boat that can be identified in the audio; specific cases are discussed later in this section.

The boat detection line for the recordings from LSMSP (Figure 27) shows that the signatures of the boats detected in the early morning are characterized by having one to three peaks and are not contiguous, whereas the signatures of boats detected from 9:00 am onwards are characterized by more peaks which are also very close – these usually correspond to a series of short duration acoustic events.

Observing the recordings from XJSP (Figure 28), one notices that longer duration acoustic events associated with boats are frequent during certain morning and evening periods. The boat detection line chart reveals multiple short duration boat events occurring frequently from nearly 9:00am until dusk. Short-term boat events have also been detected in the early morning, before 9:00am, and during the evening, after 6:00pm. After 6:30pm, when an overlapping fish chorus event is visible in the spectrogram, multiple boat events have been detected with signatures characterized by higher numbers of peaks. From 10:00pm it is possible to observe a long duration boat event that persists until midnight.

Acoustic events typical of the site, such as fish choruses, can mask some of the peaks characteristic of a boat signature, as frequency peaks of both events overlap. They can still be distinguished by further inspecting the differences in the amplitudes of the neighboring frequencies. Nonetheless, the overlap may cause boat peaks at these frequencies to go undetected. In these cases, successful detection depends on a proper choice of the detection threshold defined relative to the standard deviation of the spectrum signature, which is the single setting of the automatic detection process that may require user adjustment.

Figure 29 – 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at XJSP in February 4, 2015 at 05h:39m:52s. A boat has been detected in the interval from minutes 1 to 9; the corresponding boat detection line chart is shown at the bottom.

Figure 29 shows an example in which a long duration boat event has been detected. The boat detection line chart appears as a line with a constant number of peaks over a range of frequencies, which persists during nearly the entire duration of the recording. The detection line chart shows two to three peaks along minutes 1 to 9. The peaks overlap with the frequencies of a simultaneous fish chorus event.

The audio of the spectrogram in Figure 30 includes an acoustic event occurring from minutes 7 to 12, approximately. Listening to the audio one observes the sound of (apparently) a motor boat that gradually approaches the hydrophone and then its engine is turned off. The boat signature detected has a varying number of peaks: an initial detection indicates 3 peaks,

Figure 30 – 15-minute spectrogram of a recording taken at XJSP in February 4, 2015 at 21h:12m:44s (frequency band from 0 to 1KHz). A boat occurrence has been detected in the interval from minutes 7 to 12; the corresponding boat detection line chart is shown at the bottom.

followed by a short period with no detection; the boat is again detected with a signature of 8 peaks, going up to 12 peaks in minute 11, then in minute 12 only a single peak is detected, when the engine is powered off. The varying number of peaks is due to the non uniform properties of the sound emission by the boat. Its signal amplitude increases as it approaches the hydrophone and fades away when the engine is suddenly turned off. The algorithm successfully detected the boat, despite the unfavourable conditions and overlapping acoustic noise from a fish chorus.

Another difficult case for the algorithm is illustrated in Figure 31, in which the audio depicts a short duration and non-uniform sound emission by a boat, which is masked by extensive background noise due to crustacean acoustic activity. Nonetheless, the algorithm managed to achieve a partial recognition, as indicated by the detection line graph, in minutes 5 and 6. Only one of the peaks of the boat signature has been identified, due to the short interval between the acoustic events and the overlapping noise.

A particular case in which the algorithm failed to detect the boat signature is shown in Figure 32, which depicts an audio that includes a boat with non-constant acceleration. In this case, the frequencies of the spectral signature do not exceed the default threshold that indicates a positive boat detection. The acoustic event occurred in minute 12 of the recording, going

Figure 31 – 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at XJSP in February 4, 2015 at 10h:40m:46s, in which a boat has been detected during minutes 5 and 6; the corresponding boat detection line chart is shown at the bottom.

undetected due to the insufficient duration of the peaks in a single frequency.

Many boat occurrences have been detected in the recordings taken at the LSMP site, even though boat visitation is controlled or even prohibited at certain periods. Interestingly, from February 10 to 28 long-term acoustic emissions by boats are detected often in the mornings, mostly boat signatures characterized by up to two peaks. These emissions are low amplitude signals and the boats are apparently located distant from the hydrophone. Yet the algorithm successfully detected such occurrences, even though the recorded signal often includes multiple acoustic events in addition to the boats.

Other cases confirm the effectiveness of the proposed algorithm. Consider, for example, the audio illustrated in the spectrogram in Figure 33. The algorithm detected a boat signature characterized by a single frequency peak in this recording. This was a particularly challenging case, because the boat signal was faint and heavily masked by a loud noise due to crustacean acoustic activity.

A second example is illustrated in Figure 34, which depicts an audio in which the algorithm detected a boat signature characterized by 35 to 59 peaks along the duration of the

Figure 32 – 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at XJSP in February 4, 2015 at 10h:10m:42s. Listening to the audio it is possible to identify the sound of a boat at minute 12, which the algorithm failed to detect, as observed in the corresponding detection line chart at the bottom.

recording. Although we cannot estimate the distance of the boat to the hydrophone, its signal is sufficiently loud for its signature to be properly detected even above the 1KHz frequency band.

Yet another challenging example are cases of multiple boats captured in the same recording, possibly as overlapping acoustic events. The audio spectrogram in Figure 35 shows a boat event that persists along the duration of the recording, however a second boat event occurs in minutes 4 to 6. This is difficult to handle due to the overlap of both boat signatures. The signatures computed for the audio samples depicting minutes 2 to 5 are shown in Figure 35 to illustrate how multiple events interfere in the detection process. In the signatures of minutes 2 and 3 one observes the complete and the partial signatures of the persistent boat; in minute 4 the transient boat event partly masks the previous signature, introducing additional peaks. The emission by the transient boat in minute 5 is sufficient to mask the signature of the persistent boat, as it adds noise in the frequency band from 200Hz to 2kHz. The consequence is that only the signature of the transient boat is detected.

Figure 33 – 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at LSMP in February 11, 2015 at 06h:27m:57s. The low amplitude signal from a boat has been detected in the intervals from minutes 0 to 3 and from 8 to 11, even though it is masked by crustacean noise, as observed in the boat detection line chart shown at the bottom.

4.2.1 Comparison: FAV vs DEMON

A DEMON implementation based on square-law demodulation by Pollara, Sutin and Salloum (2016) available online² has been used as baseline for a comparison with the proposed FAV signature, regarding detection accuracy and processing times. The code and audio samples used in the analysis and comparisons can be downloaded from Github (REIS, 2019).

We conducted the comparison on 96 audio files selected from our databases (Xixová and Japuí State Parks), corresponding to 24 hours of audio analysed at a one-minute sample resolution, totaling 1440 one-minute samples representative of typical scenarios found in underwater ecological recordings. We conducted a manual annotation process and identified that from the 1440 samples considered, 706 do not include boat events, whereas the remaining 734 include diverse observable boat occurrences, varying from very clear to very faint signals. The samples depict distinct situations, including non-uniform vessel emissions near the hydrophone and other difficult scenarios characterized by overlapping with crustacean acoustic emissions or fish choruses.

² <<https://github.com/lxpollara/pyDEMON>>

Figure 34 – 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at LSMSP in February 11, 2015 at 15h:29m:54s. The signature of a boat has been detected along the duration of the recording, characterized by 35 up to 59 frequency peaks, as observed in the corresponding detection line chart.

The DEMON signature computation requires the definition of a target frequency window for each detection scenario. Inadequate choice of this initial input parameter can render the detection task unfeasible. In order to ensure a proper comparison, the input samples were processed similarly for both methods, setting the higher frequency to 900Hz and the low frequency to 400Hz as the default parameters for the DEMON signature calculations. Each one-minute sample was smoothed with a size 11 Blackman window convolution, and the corresponding standard deviations were extracted.

In both methods, a detection occurs whenever the corresponding signature exceeds the corresponding detection threshold, where each excess will correspond to a frequency peak in the boat detection signature. We kept the default detection threshold for the FAV signature, i.e., $1.5 \times \text{standard deviation}$. Setting the detection threshold for DEMON requires a different approach, however, as the base signals vary in amplitude due to interference of noise. Thus, we set an initial threshold as in FAV ($1.5 \times \text{standard deviation}$), then obtained a basis amplitude for each frequency by smoothing the DEMON signature with a size 64 Blackman window convolution. The basis amplitudes thus obtained were then added to the initial threshold.

Figure 35 – 15-minute spectrogram of a recording taken at LSMSP in February 4, 2015 at 04h:48m:55s, where overlapping sound emissions from two boats can be observed. The charts show the signatures computed of the audio samples corresponding to minutes 2, 3, 4 and 5, displaying the input spectrum signature (blue lines) with the corresponding boat detections highlighted by the red vertical lines.

Under normal conditions these threshold settings enabled both methods to correctly detect the presence of boats and create the corresponding boat signatures. An example for a particular one minute sample is illustrated in Figure 36, where a boat with two frequency peaks can be observed, which has been correctly detected in both FAV and DEMON signatures. However, results from the methods differ in more challenging conditions. Detection results for

a noisy sample are illustrated in Figure 37, where FAV correctly identified a boat with eight frequency peaks. The DEMON signature, however, detected only one of the peaks and also incurred in a false positive detection of a frequency peak which is not due to a boat.

Figure 36 – FAV and DEMON signatures obtained for a one-minute audio sample that includes crustacean acoustic activity overlapped with a persistent acoustic emission by a boat (spectrogram on the right). The dotted gray lines indicate the detection threshold ($1.5 \times \text{Standard Deviation}$). Both algorithms correctly detected the boat signature, with two frequency peaks. The red watermarks indicate the peaks detected in the FAV signature (purple line); the black dots indicate the detections in the DEMON signature (green line).

Considering all 1440 samples, we computed the detection accuracy and the correctness of the number of peaks identified in the signatures obtained from both methods. Detection accuracy is reported as the numbers and percentages of correct and incorrect boat detections. The signature accuracy is measured as the number of detected frequency peaks, relative to the true number of peaks. We report the average percentages for each method, over all signatures extracted.

Table 2 – Detection accuracy of FAV and DEMON algorithms on a database of 1440 one minute audio samples.

Confusion matrix FAV			
		Reality	
		NO Boat	Boat
Algorithm	NO Boat	706	50
	Boat	0	684
# Samples		706	734

Confusion matrix DEMON			
		Reality	
		No Boat	Boat
Algorithm	NO Boat	622	328
	Boat	84	406
# Samples		706	734

Figure 37 – FAV and DEMON signatures obtained for a one-minute audio sample that includes crustacean acoustic activity overlapped with a persistent acoustic emission by a boat (spectrogram shown on the right). The dotted gray lines indicate the detection threshold ($1.5 \times \text{Standard Deviation}$). In this case the boat signature is characterized by eight frequency peaks. The red watermarks indicate the boat peaks correctly detected in the FAV signature (purple line); the black dots indicate the detections in the DEMON signature (green line). The DEMON algorithm identified a single peak due to the boat, the other peak is actually a false-positive, not related to the boat.

Table 3 – Boat detection accuracy and signature accuracy results obtained with the FAV and DEMON implementations, as well as false positive and false negative detection rates.

	FAV	DEMON
Detection accuracy	93.19%	55.31%
Signature accuracy	80.38%	29.21%
% of false positives	0.00%	10.63%
% of false negatives	20.42%	45.14%

Table 2 shows two confusion matrices displaying the detection accuracy results for the FAV and the DEMON implementations, respectively. The columns labeled **Reality** inform the number of boat occurrences according to the manual annotation, the rows labeled **Algorithm** show the results obtained with FAV or DEMON. Table 3 summarizes the accuracy rates and also the false positive and false negative rates.

FAV correctly detected the presence of boats in 684 out of the 734 samples with boat occurrences, an accuracy rate of 93%. It missed 50 occurrences, and incurred in no false positives. DEMON correctly detected 406 boat occurrences, an accuracy rate of 55%. It missed 328 boat events, and incorrectly detected 84 acoustic events as boats. The **false-negative** rates in our analysis of DEMON could be disputed, since the manual annotation identified some very faint boat signals. Possibly different calibrations of DEMON would be required for it to perform

better on the database as a whole, as the conditions describing boat events vary considerably. On the other, FAV has shown very robust performance with the settings employed, regardless of the varied conditions.

For the 734 samples including boat occurrences we also computed the correctness of the number of frequency peaks identified, as a measure of the quality of the signatures computed. For instance, given a boat signature with 5 frequency peaks, if a method computes the signature with 3 peaks, its resulting signature accuracy is 60%. As informed in Table 3 (second row), the average signature accuracy of the FAV method over all samples was 80%, whereas for the DEMON method it was 29%.

Table 4 – Times (in seconds) to compute the boat signatures of a 15 minute audio recording and of a one minute sample extracted from the same audio. Time measures were obtained from five executions of each (DEMON/FAV) implementation on a *Intel® Core™ i7 – 7700 @ 3.60GHz* × 4 with 16GB memory.

Time (Seconds)	15 minute audio		One minute audio	
	DEMON	FAV	DEMON	FAV
	6.672	0.031	0.413	0.006
	6.760	0.029	0.421	0.006
	6.743	0.030	0.416	0.006
	6.790	0.027	0.431	0.006
	6.606	0.029	0.409	0.006
Avg	6.743	0.029	0.416	0.006

Table 4 shows the times, in seconds, to extract the DEMON and FAV signatures for a 15 minute audio recording and for a one minute sample of the same recording, considering five executions of the algorithm implementations. In the 15 minute sampling we observe a speed up of nearly 230, as the average execution times drop from 6.743 seconds to compute the DEMON signature to 0.029 seconds to compute the FAV signature. In the one minute sample the DEMON signature has been computed in 0.416 seconds in average, and the FAV signature in 0.006 seconds, a speed up of nearly 70.

These empirical investigations provide evidence on the improved performance of the FAV algorithm over the DEMON implementation, regarding both computational cost and accuracy, the later measured in terms of detection correctness and signature quality. Moreover, unlike DEMON, FAV requires no specific settings or prior knowledge of the boat frequency signatures. It is a self-contained solution that relies solely on the input signal of the audio spectrum. A single user setting is required to establish the detection threshold, where the default setting as $1.5 \times \text{Standard Deviation}$ of the original spectrum proved adequate in all cases appearing in the databases considered.

4.3 An Exploratory Approach For Annotating Acoustic Events

A set of recordings containing acoustic events of interest was selected for a case study, totaling 37 one-minute files with 44100Hz sampling rate, several features (12 MFCC components, RMS, SPL, NDSI with anthrophony and biophony components) were extracted using the default system settings, without any user interaction to filter frequencies or any changes in the default settings of each feature. The feature vectors were extracted using a sample window of 2 seconds, so that acoustic events and frequency variations through the recordings could be better represented.

For specialists in ecology, the production of annotated datasets is an important task that requires time and resources. Having annotated datasets can help answer questions about the condition and behavior of an acoustic environment, in addition to serving as input for machine learning algorithms, which could be used to generate recognition models for acoustic events of interest. To determine the effectiveness of the method in the dataset annotation task, feature sets were inserted as dimensional anchors in different dimensional axis of the Concentric Radviz, allowing overlapping clusters to be visualized by rotating the samples. These overlapping clusters may belong to different classes of acoustic events or the same acoustic event at different levels.

While rotating the dimensional axis, we can observe the joint movement of samples clusters, revealing similar groups. A selection of different features will result in different clusters, which can be better defined according to the feature extraction process. In other words, given that at the moment of extraction the user defines settings that favor the detection of a particular acoustic event, the use of this feature in Concentric Radviz may result in better segmentation of samples corresponding that particular event.

Figure 38 presents a Concentric Radviz configuration, with components 0 to 3 of the MFCC feature in the first layer, and components 5 to 6 of the same feature in the second layer. By rotating the layers it was possible to identify several clusters, which were selected and verified using the embedded spectrogram viewer, then a label was created for the specified samples.

During the exploratory analysis of the samples in the data set, 3 recordings with a high presence of insect noise were grouped in the visualization side by side, yet the samples related to each of the files were visibly segregated (Figure 39-A). This is an interesting result which could allow researchers to characterize the same acoustic event with different amplitudes. Observing the details of the samples, it was possible to verify that the insect noise in each of the audios, although very similar, presents subtle variations in both the amplitudes of each frequency as in the dominant frequency of the samples.

The same can be observed in the samples characterized by rain noise, which could be separated by rotating the layers until 3 very distinct clusters could be identified, as can be seen in

Figure 38 – Acoustic events annotated during the exploratory analysis of the dataset, using 7 of the 12 components of the MFCC feature, distributed as dimensional anchors along two dimensional axis, depicted in "Layers". The samples were annotated for 8 acoustic events of interest, depicted in "Sample labels". Some acoustic events of the same group were divided into two groups, as in the case of rain and heavy rain.

Figure 39 – Experiments in segregating variations of the same acoustic event (A and B), and the resulting segregation response of adding more layers to the visualization (C).

Figure 39-B. In this visualization it is also possible to observe a single sample separated from the others, between the red and green samples, which despite being part of the rain noise class, was not effectively displayed next to the others. By observing the visualization of the spectrogram it was possible to verify that some of the features were calculated with diverging values from others, which could lead to an investigation on how this feature is calculated in order to validate

its effectiveness.

The more layers overlapping, the greater the possibilities for segmentation, based on the rotation of these layers and configuration of the sigmoid function. A segmentation using 3 layers with sigmoid adjustment can be seen in Figure 39-C, showing several annotated labels (Same labels from Figure 38) with segmentation even between acoustic events of the same type.

4.4 Final Considerations

In this chapter we presented the results in the development of a visualization framework for applications in acoustic ecology. Section 4.1 introduced the Seecology framework, with methods for extracting audio features with a sample-based approach, performed by a customized audio features extractor, as well as methods for the generation of resources for use in visualizations and analysis of ecological recordings. Section 4.2 described a case study using an approach based on the Concentric Radviz visualization for annotating datasets generated from the custom audio features extractor, part of the Seecology framework. Finally, a new method for detecting and characterizing the presence of boats in underwater recordings, presented in Section 4.3, was tested with real underwater recordings to demonstrate its effectiveness, with a corresponding visualization to assist in the analysis of the resulting data. The method was also compared to other solutions from the literature, proving to be more accurate, demanding less computational resources.

CONCLUDING REMARKS

In this chapter, the final considerations of this doctoral thesis are presented. Section 5.1 takes up the established objectives and presents the main results and contributions for each. Section 5.2 shows the publications generated and, finally, in Section 5.3 presents the limitations and future work.

5.1 Conclusions and Contributions

The analysis and visualization of large amounts of ecological recordings, along with resource generation, constitute a major challenge in Soundscape Ecology. Researchers require means for the extraction of audio features, producing datasets that represent acoustic events and frequency variations through the duration of each recording, allowing the analysis of acoustic events effects in multiple audio features. The characterization of acoustic events present on ecological is a powerful tool for the analysis of acoustic environments, assisting in the identifications of relevant acoustic events, assessing their impact and identification of changes in environmental behavior, whether originating from Biophony, Geophony and Anthropony. Annotated datasets produced from those recordings can be a resource and time consuming task. Methods for the characterization and annotation of acoustic events could assist researchers through this process.

On this research, we aimed to solve problems associated with the extraction of audio features to help researchers in the selection of meaningful features, representing the subtle variations within the recordings, and visualization methods for the exploration and analysis of the produced datasets, along with the annotation of acoustic events of interest, which is an important asset to other approaches, such as training machine learning algorithms to produce models for the detection of specific acoustic events of interest. Also, a novel method was proposed for the identification and characterization of boats in underwater recordings, which could assist ecologists and authorities to secure protected marine areas.

Each result achieved on this research was obtained by studying and developing methods and solutions to answer the research questions presented in Chapter 1, as follows:

- *Q1 - Can we formalize a visualization framework for soundscape ecology data that is robust enough to represent the wide range of information characterizing soundscapes, without incurring in the high computational processing costs of current approaches for audio data analysis?*

The Seecology framework, encompassing custom tools for feature extraction and dataset annotation, creation of visualization resources, and exploratory visualizations provides an useful resource for researchers and specialists to assist in the conduction of their studies. Whether by using the Seecology framework to exploratory analysis or by using the resulting datasets as input to other research outside the framework.

The methods for exploring acoustic ecology recordings of the Seecology framework combine multiple visualizations with custom feature extractors to help guide user in a subspace feature investigation process. We applied this approach to data from acoustic ecology ecology studies, in order to identify and characterize acoustic events of interest, as illustrated in the case studies for boat detection (Section 4.2) and the annotation of acoustic events (Section 4.3). Many other possibilities are afforded by the system, due to the effective integration of multiple functionalities to support audio analysis, e.g., spectrograms, Concentric Radviz visualization, feature temporal behavior and audio playback.

In order to represent the wide range of information that characterizes soundscapes, a large collection of features was included in the Seecology framework, developed during this research. The available features were gathered from several works published in the literature, and as some of the audio libraries available already implement those, a comparison of the number of features in each and the processing speed was determinant to select the best candidate. Some features, such as the ecological indices, were not in these audio libraries, and needed to be implemented.

The processing speed of audio features and visualization resources was improved in the Seecology framework by developing custom extractors instead of using the ones in the available libraries. The audio features extraction tool included in the Seecology framework incorporate an automatic sample strategy, in which the user can select a time-based sample window in seconds, revealing nuances in the audio features throughout the recordings duration, instead of one feature per file. Also, the selection of multiple features and resulting CSV files make it easy for researchers to produce data for their studies. The creation of assets for the visualizations and analysis, such as single and long period spectrograms, were also developed without ready-to-go libraries, which speed up the process, reducing the computational costs as well as allowing for a higher customization of final results.

The web interface allows for specialists without knowledge in programming to extract a large set of features from ecological recordings, with custom annotated data through the use of interactive visualizations.

- *Q2: How can acoustic visualizations fueled by audio features support the identification of similar acoustic events of interest, so there specialists can explore the universe of acoustic events contained in ecological recordings, analyzing the relationship between acoustic events and audio features?*

The Concentric Radviz visualization employed to segment and annotate samples from ecological recordings, with the combined use of the custom feature extractor included in the Seecology framework, has demonstrated an efficient way to annotate datasets by means of exploratory visualization as a decision-making tool. The resulting datasets and respective annotated data can be used in other scenarios that requires such resources as an input method as, for example, neural networks training or machine learning methods.

Tests conducted on the selected recordings, extracted with the custom audio feature extractor, were able to characterize several clusters with distinct acoustic events, each of which was annotated as an acoustic event of interest. Considering feature extraction targeted at specific frequency ranges, with adjustments to each feature's settings, specialists could improve the characterization of other acoustic events that were hidden in the data. By exploratory analysis using the Concentric Radviz visualization, choosing different features to be placed onto the dimensional axis, these new datasets could lead to insights about the recordings content.

The presented use of the Concentric Radviz visualization evidenced to be a good solution for the annotation of acoustic events by means of an exploratory approach. The method is also useful for outlier identification and incorrect feature extraction, using the spectrograms and audio player as a guide. It was evident that extraction in small sample windows proved effective in distinguishing frequency variations throughout recordings associated with the spectrogram visualization and built-in audio player.

- *Q3: Is it possible to automatically detected the presence of boats and other types of vessels in noisy underwater recordings?*

The novel boat detection method developed in this research proved to be effective and capable of automatically detecting not only the presence of boats in underwater recordings, but the corresponding signatures from passing boats. Results from experiments conducted on underwater recordings collected in February 2015 at two sites in Brazil provide empirical evidence on its effectiveness. The examples discussed in Section 4.2 and Subsection 4.2.1 show that the method can successfully detect the presence of boats even in difficult conditions, such as faint signals or multiple overlapping acoustic events. Manual inspection of all the recordings

considered in the experiments (Subsection 4.2.1) revealed no false positives and a small rate of missed detections in very specific situations.

Depending on the sensitivity level, chosen by the user during the extraction process, these small rate of missed detections can even be eliminated, however, to the expense of the detection of low level noise boats. In any case, the method is useful, along with its resulting data visualization, to assert the presence of boats in a location during periods of time.

The algorithm can recognize boats under constant acceleration, where the resulting sound signal caused by cavitation generates characteristic peaks at short frequency intervals. The failure cases were of boats with erratic acceleration behavior. This situation shall be addressed in further work, possibly with an analysis of the amplitude variation in time, combined with frequency information and machine learning methods to recognize specific patterns.

Besides detecting emissions by boats and ships, the FAV signature is also susceptible to sound emission by airgun arrays. The reason is that such devices produce acoustic energy with frequency amplitudes mostly under 100Hz (HERMANNSEN *et al.*, 2015; LANDRO; AMUNDSEN; BARKER, 2011), the same band of the boat emissions. For boat detection purposes, if an acoustic signal emitted by an airgun array is identified one can assume the consequent detection of a boat signal, since airgun arrays are towed by boats.

In comparison to the classic DEMON method, which also relies on analysis of the acoustic signal for boat detection, our solution provides an alternative that is easy to implement and does not require extensive processing power or complex configuration of input parameters. DEMON, on the other hand, requires human intervention to define the best filters for each type of target signal and is computationally more expensive.

In the case studies carried out, a surprising number of vessels were detected in the conservation areas throughout the analyzed period, which disturbs the marine life of those locations.

5.2 Publications

- Paper Published in International Conference:

Clausius Duque Gonçalves Reis, Thalisson Nobre Santos and Maria Cristina Ferreira de Oliveira, **A Visualization Framework for Feature Investigation in Soundscape Recordings**, 2018 22nd International Conference on Information Visualisation (IV), Fisciano, 2018, pp. 490-497. DOI: 10.1109/iV.2018.00091

- Paper Published in International Journal:

Clausius Duque Gonçalves Reis, Linilson Rodrigues Padovese and Maria Cristina Ferreira de Oliveira. **Automatic detection of vessel signatures in audio recordings with**

spectral amplitude variation signature. *Methods in Ecology and Evolution*. 2019; 10: 1501– 1516. DOI: 10.1111/2041-210X.13245

5.3 Limitations and Future Work

As future work, one of our goals is to conduct additional case studies with participation of domain experts and considering other soundscape databases. This will allow additional validation of the system and its adjustment to meet the specific needs of the user in the tasks necessary for conducting research. Several aspects of the system deserve further investigation, e.g., how to ensure the best possible clustering of features, how to improve and generalize the labeling functionalities, how to integrate additional data mining resources and how to improve speed in larger dataset visualizations.

Certain aspects of the framework could be improved by user validation to enhance detection of acoustic events of interest or the selection of initial settings that best suit an audio type. Machine learning techniques could be applied to the Seecology framework, and make use of the custom audio feature extractor and resulting data to detect, classify and predict patterns.

More visualizations will be added to the user interface, helpful for specialists in acoustic ecology, in order to better use the facilities implemented in the Seecology framework, such as the custom extractor and the long period spectrograms seen in the boat detection visualization.

We intend to add the automatic detection of new acoustic events of interest such as certain types of fish chorus and means to reload a settings history for tracking the latest analysis.

BIBLIOGRAPHY

Ankerst, M.; Berchtold, S.; Keim, D. A. Similarity clustering of dimensions for an enhanced visualization of multidimensional data. In: **Proceedings IEEE Symposium on Information Visualization (Cat. No.98TB100258)**. [S.l.: s.n.], 1998. p. 52–60. Citation on page 44.

AVERBUCH, A.; ZHELUDEV, V.; NEITTAANMÄKI, P.; WARTIAINEN, P.; HUOMAN, K.; JANSON, K. Acoustic detection and classification of river boats. **Applied Acoustics**, v. 72, n. 1, p. 22 – 34, 2011. ISSN 0003-682X. Citation on page 45.

BARBER, J. R.; CROOKS, K. R.; FRISTRUP, K. M. The costs of chronic noise exposure for terrestrial organisms. **Trends in Ecology & Evolution**, v. 25, n. 3, p. 180–189, March 2010. ISSN 01695347. Available: <<http://dx.doi.org/10.1016/j.tree.2009.08.002>>. Citation on page 35.

BARCHIESI, D.; GIANNOULIS, D. D.; STOWELL, D.; PLUMBLEY, M. D. Acoustic Scene Classification: Classifying environments from the sounds they produce. **IEEE Signal Processing Magazine**, v. 32, n. 3, p. 16–34, 2015. ISSN 10535888. Citation on page 29.

BELLISARIO, K. M.; BROADHEAD, T.; SAVAGE, D.; ZHAO, Z.; OMRANI, H.; ZHANG, S.; SPRINGER, J.; PIJANOWSKI, B. C. Contributions of mir to soundscape ecology. part 3: Tagging and classifying audio features using a multi-labeling k-nearest neighbor approach. **Ecological Informatics**, v. 51, p. 103 – 111, 2019. ISSN 1574-9541. Available: <<http://www.sciencedirect.com/science/article/pii/S1574954118301924>>. Citation on page 36.

BELLISARIO, K. M.; PIJANOWSKI, B. C. Contributions of mir to soundscape ecology. part i: Potential methodological synergies. **Ecological Informatics**, v. 51, p. 96 – 102, 2019. ISSN 1574-9541. Available: <<http://www.sciencedirect.com/science/article/pii/S1574954118301912>>. Citation on page 35.

BELLISARIO, K. M.; VANSCHAIK, J.; ZHAO, Z.; GASC, A.; OMRANI, H.; PIJANOWSKI, B. C. Contributions of mir to soundscape ecology. part 2: Spectral timbral analysis for discriminating soundscape components. **Ecological Informatics**, v. 51, p. 1 – 14, 2019. ISSN 1574-9541. Available: <<http://www.sciencedirect.com/science/article/pii/S1574954118301900>>. Citation on page 35.

BOELMAN, N. T.; ASNER, G. P.; HART, P. J.; MARTIN, R. E. Multi-trophic invasion resistance in hawaii: Bioacoustics, field surveys, and airborne remote sensing. **Ecological Applications**, v. 17, n. 8, p. 2137–2144, 2007. Citation on page 55.

BOGDANOV, D.; WACK, N.; GÓMEZ, E.; GULATI, S.; HERRERA, P.; MAYOR, O.; ROMA, G.; SALAMON, J.; ZAPATA, J. R.; SERRA, X. Essentia: an audio analysis library for music information retrieval. In: **International Society for Music Information Retrieval Conference (ISMIR'13)**. Curitiba, Brazil: [s.n.], 2013. p. 493–498. Citation on page 28.

BOSTOCK, M.; OGIEVETSKY, V.; HEER, J. D3 data-driven documents. **IEEE Transactions on Visualization and Computer Graphics**, IEEE Educational Activities Department,

Piscataway, NJ, USA, v. 17, n. 12, p. 2301–2309, Dec. 2011. ISSN 1077-2626. Available: <https://doi.org/10.1109/TVCG.2011.185>. Citation on page 51.

BOTTELDOOREN, D.; COENSEL, B.; MEUR, T. D. The temporal structure of the urban soundscape. **Journal of Sound and Vibration** **292**(1-2), p. 105–123, 2004. Citation on page 27.

BROWN, A.; GARG, S.; MONTGOMERY, J. Scalable preprocessing of high volume environmental acoustic data for bioacoustic monitoring. **PLOS ONE**, Public Library of Science, v. 13, n. 8, p. 1–24, 08 2018. Citation on page 29.

BULLOCK, J. LibXtract: A Lightweight Library for Audio Feature Extraction. In: **Proceedings of the International Computer Music Conference**. Copenhagen, Denmark: ICMA, 2007. Citation on page 28.

CARTWRIGHT, M.; SEALS, A.; SALAMON, J.; WILLIAMS, A.; MIKLOSKA, S.; MACCONNELL, D.; LAW, E.; BELLO, J.; NOV, O. Seeing sound: Investigating the effects of visualizations and complexity on crowdsourced audio annotations. **Proceedings of the ACM on Human Computer Interaction**, v. 1, 11 2017. Citation on page 38.

CHENG, S.; XU, W.; MUELLER, K. Radviz deluxe: An attribute-aware display for multivariate data. **Processes**, v. 5, n. 4, 11 2017. Citations on pages 40, 42, and 43.

CHUNG, K. W.; SUTIN, A.; SEDUNOV, A.; BRUNO, M. DEMON acoustic ship signature measurements in an urban harbor. **Advances in Acoustics and Vibration**, v. 2011, n. i, 2011. Available: <http://doi.org/10.1155/2011/952798>. Citations on pages 30, 45, and 46.

CORREA, I.; CARVALHO, A. Dualradviz: Preserving context between classification evaluation and data exploration with radviz. In: **2016 5th Brazilian Conference on Intelligent Systems (BRACIS)**. Los Alamitos, CA, USA: IEEE Computer Society, 2016. p. 241–246. Available: <https://doi.ieeecomputersociety.org/10.1109/BRACIS.2016.052>. Citations on pages 15, 40, 41, 42, and 43.

DEPRAETERE, M.; PAVOINE, S.; JIGUET, F.; GASC, A.; DUVAIL, S.; SUEUR, J. Monitoring animal diversity using acoustic indices: Implementation in a temperate woodland. **Ecological Indicators**, v. 13, n. 1, p. 46–54, 2012. ISSN 1470160X. Citation on page 28.

FARINA, A. **Principles and methods in landscape ecology**. Dordrecht :: Springer Netherlands,, 2006. (Landscape Series,). Available: <http://dx.doi.org/10.1007/978-1-4020-5535-5>. Citation on page 34.

FARINA, A.; PIERETTI, N. Sonic environment and vegetation structure: A methodological approach for a soundscape analysis of a Mediterranean maqui. **Ecological Informatics**, v. 21, p. 120–132, 2014. ISSN 15749541. Citation on page 28.

FARINA, A.; PIERETTI, N.; PICCIOLI, L. The soundscape methodology for long-term bird monitoring: A mediterranean europe case-study. **Ecological Informatics**, v. 6, n. 6, p. 354 – 363, 2011. ISSN 1574-9541. Available: <http://www.sciencedirect.com/science/article/pii/S1574954111000586>. Citation on page 28.

FIGUEIRA, L.; TELLA, J. L.; CAMARGO, U. M.; FERRAZ, G. Autonomous sound monitoring shows higher use of Amazon old growth than secondary forest by parrots. **Biological Conservation**, Elsevier Ltd, v. 184, p. 27–35, 2015. ISSN 00063207. Available: <http://dx.doi.org/10.1016/j.biocon.2014.12.020>. Citation on page 28.

FROMMOLT, K. H.; TAUCHERT, K. H. Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird. **Ecological Informatics**, Elsevier B.V., v. 21, p. 4–12, 2014. ISSN 15749541. Available: <<http://dx.doi.org/10.1016/j.ecoinf.2013.12.009>>. Citation on page 28.

FULLER, S.; AXEL, A. C.; TUCKER, D.; GAGE, S. H. Connecting soundscape to landscape: Which acoustic index best describes landscape configuration? **Ecological Indicators**, v. 58, n. July, p. 207–215, 2015. ISSN 1470160X. Available: <<http://linkinghub.elsevier.com/retrieve/pii/S1470160X15003179>>. Citation on page 28.

GUYOT, P.; PINQUIER, J. Browsing soundscapes. **Proceedings of the 1st International Conference on Technologies for Music Notation and Representation TENOR-2015**, p. 1–8, 2015. Citations on pages 38 and 39.

HANSON, D.; ANTONI, J.; BROWN, G.; EMSLIE, R. Cyclostationarity for Passive Underwater Detection of Propeller Craft: A Development of DEMON Processing. **Proceedings of Acoustics**, November, p. 1–6, 2008. Citations on pages 30 and 46.

HEITTOLA, T.; MESAROS, A.; VIRTANEN, T.; ERONEN, A. Sound event detection in multisource environments using source separation. In: . [S.l.: s.n.], 2011. Citations on pages 38 and 39.

Heittola, T.; Mesaros, A.; Virtanen, T.; Gabbouj, M. Supervised model training for overlapping sound events based on unsupervised source separation. In: **2013 IEEE International Conference on Acoustics, Speech and Signal Processing**. [S.l.: s.n.], 2013. p. 8677–8681. Citations on pages 38 and 39.

HERMANNSEN, L.; TOUGAARD, J.; BEEDHOLM, K.; NABE-NIELSEN, J. Characteristics and Propagation of Airgun Pulses in Shallow Water with Implications for Effects on Small Marine Mammals. **PLOS ONE**, July, p. 1–17, 2015. Available: <<https://doi.org/10.1371/journal.pone.0133436>>. Citation on page 98.

Hoffman, P.; Grinstein, G.; Marx, K.; Grosse, I.; Stanley, E. Dna visual and analytic data mining. In: **Proceedings. Visualization '97 (Cat. No. 97CB36155)**. [S.l.: s.n.], 1997. p. 437–441. Citations on pages 37 and 40.

INSELBERG, A.; DIMSDALE, B. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In: **Proceedings of the 1st Conference on Visualization '90**. Los Alamitos, CA, USA: IEEE Computer Society Press, 1990. (VIS '90), p. 361–378. ISBN 0-8186-2083-8. Available: <<http://dl.acm.org/citation.cfm?id=949531.949588>>. Citation on page 37.

JAIN-SCHLAEPFER, S.; FAKAN, E.; RUMMER, J. L.; SIMPSON, S. D.; MCCORMICK, M. I.; COOKE, S. Impact of motorboats on fish embryos depends on engine type. **Conserv Physiol**, v. 6, n. 61, p. 1410–1093, 2018. Available: <<https://doi.org/10.1093/conphys/coy014>>. Citations on pages 30 and 44.

KANDOGAN, E. Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. In: **In Proceedings of the IEEE Information Visualization Symposium, Late Breaking Hot Topics**. [S.l.: s.n.], 2000. p. 9–12. Citation on page 37.

KEMPER, G.; PONCE, D.; TELLES, J. An algorithm to obtain boat engine rpm from passive sonar signals based on demon processing and wavelets packets transform. **Journal of Electrical Engineering and Technology**, v. 14, 2019. ISSN 2093-7423. Available: <<https://doi.org/10.1007/s42835-019-00260-4>>. Citation on page 45.

KHOLGHI, M.; PHILLIPS, Y.; TOWSEY, M.; SITBON, L.; ROE, P. Active learning for classifying long-duration audio recordings of the environment. **Methods in Ecology and Evolution**, v. 9, n. 9, p. 1948–1958, 2018. Available: <<https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13042>>. Citations on pages 29, 37, 38, and 40.

KRIJNDERS, D.; NIESSEN, M. E.; ANDRINGA, T. C. Annotating soundscapes. In: **38th International Congress and Exposition on Noise Control Engineering 2009, INTER-NOISE 2009**. Ottawa, Canada: [s.n.], 2009. v. 6. Citations on pages 37, 38, and 39.

LAILOLO, P. The emerging significance of bioacoustics in animal species conservation. **Biological Conservation**, Elsevier Ltd, v. 143, n. 7, p. 1635–1645, 2010. ISSN 00063207. Available: <<http://dx.doi.org/10.1016/j.biocon.2010.03.025>>. Citation on page 28.

LANDRO, M.; AMUNDSEN, L.; BARKER, D. High-frequency signals from air-gun arrays. **Geophysics**, v. 76 NO. 4, n. July-August 2011, p. Q19–Q27, 2011. Available: <<http://dx.doi.org/10.1190/1.3590215>>. Citation on page 98.

LEAL, N.; LEAL, E.; SANCHEZ, G. Marine vessel recognition by acoustic signature. **ARPN Journal of Engineering and Applied Sciences**, v. 10, n. 20, p. 9633–9639, 2015. Citations on pages 30, 44, and 45.

Lowes, G. J.; Neasham, J. A.; Burnett, R.; Tsimenidis, C. C. Low energy, passive acoustic sensing for wireless underwater monitoring networks. In: **OCEANS 2019 MTS/IEEE SEATTLE**. [S.l.: s.n.], 2019. p. 1–9. Citation on page 45.

YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software. Citation on page 28.

MCFEE Brian; RAFFEL Colin; LIANG Dawen; ELLIS Daniel P.; MCVICAR Matt; BATTENBERG Eric; NIETO Oriol. *librosa: Audio and Music Signal Analysis in Python*. In: HUFF Kathryn; BERGSTRA James (Ed.). **Proceedings of the 14th Python in Science Conference**. [S.l.: s.n.], 2015. p. 18 – 24. Citations on pages 28, 51, and 55.

MELÉNDEZ-CATALÁN, B.; MOLINA, E.; GÓMEZ, E. Bat: An open-source, web-based audio events annotation tool. In: **Web Audio Conference**. [s.n.], 2017. Available: <<http://eecs.qmul.ac.uk/~keno/17.pdf>>. Citation on page 38.

MERCHANT, N. D.; PIROTTA, E.; BARTON, T. R.; THOMPSON, P. M. Monitoring ship noise to assess the impact of coastal developments on marine mammals. **Marine Pollution Bulletin**, v. 78, n. 1-2, p. 85–95, 2014. Available: <<https://doi.org/10.1016/j.marpolbul.2013.10.058>>. Citation on page 44.

MOFFAT, D.; RONAN, D.; REISS, J. D. An evaluation of audio feature extraction toolboxes. In: DAFX-15. **Proc. 18th International Conference on Digital Audio Effects (DAFx-15)**. [S.l.], 2015. Citation on page 28.

MOORE, S. E.; REEVES, R. R.; SOUTHALL, B. L.; RAGEN, T. J.; SUYDAM, R. S.; CLARK, C. W. A New Framework for Assessing the Effects of Anthropogenic Sound on Marine Mammals in a Rapidly Changing Arctic. **BioScience**, v. 62, n. 3, p. 289–295, 2012. ISSN 00063568. Available: <<http://bioscience.oxfordjournals.org/cgi/doi/10.1525/bio.2012.62.3.10>>. Citations on pages 28 and 36.

OLIVEIRA, A. G. de; VENTURA, T. M.; GANCHEV, T. D.; FIGUEIREDO, J. M. de; JAHN, O.; MARQUES, M. I.; SCHUCHMANN, K.-L. Bird acoustic activity detection based on morphological filtering of the spectrogram. **Applied Acoustics**, v. 98, p. 34 – 42, 2015. ISSN 0003-682X. Available: <<http://www.sciencedirect.com/science/article/pii/S0003682X15001309>>. Citations on pages 38 and 39.

ONO, J. H. P.; SIKANSI, F.; CORRÊA, D. C.; PAULOVICH, F. V.; PAIVA, A.; NONATO, L. G. Concentric radviz: Visual exploration of multi-task classification. In: **SIBGRAPI**. [S.l.]: IEEE Computer Society, 2015. p. 165–172. ISBN 978-1-4673-7962-5. Citations on pages 15, 29, 40, 43, 65, and 66.

PETROSSIAN, G. A. Preventing illegal, unreported and unregulated (iuu) fishing: A situational approach. **Biological Conservation**, v. 189, p. 39 – 48, 2015. ISSN 0006-3207. Detecting and Understanding Non-compliance with Conservation Rules. Available: <<http://www.sciencedirect.com/science/article/pii/S0006320714003140>>. Citation on page 30.

PIERETTI, N.; FARINA, A.; MORRI, D. A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). **Ecological Indicators**, Elsevier Ltd, v. 11, n. 3, p. 868–873, 2011. ISSN 1470160X. Available: <<http://dx.doi.org/10.1016/j.ecolind.2010.11.005>>. Citation on page 28.

PIJANOWSKI, B. C.; FARINA, A.; GAGE, S. H.; DUMYAHN, S. L.; KRAUSE, B. L. What is soundscape ecology? An introduction and overview of an emerging new science. **Landscape Ecology**, p. 1213–1232, 2011. Citations on pages 27, 28, 29, and 31.

PIJANOWSKI, B. C.; VILLANUEVA-RIVERA, L. J.; DUMYAHN, S. L.; FARINA, A.; KRAUSE, B. L.; NAPOLETANO, B. M.; GAGE, S. H.; PIERETTI, N. Soundscape Ecology: The Science of Sound in the Landscape. **BioScience**, v. 61, n. 3, p. 203–216, 2011. ISSN 00063568. Citations on pages 15, 27, 29, 34, 35, and 36.

Podwinska, Z.; Sobieraj, I.; Fazenda, B. M.; Davies, W. J.; Plumbley, M. D. Acoustic event detection from weakly labeled data using auditory salience. In: **ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.: s.n.], 2019. p. 41–45. Citation on page 39.

POLLARA, A. S.; LIGNAN, G.; BOULANGE, L.; SUTIN, A.; SALLOUM, H. Specifics of demon acoustic signatures for large and small boats. **The Journal of the Acoustical Society of America**, v. 141, n. 5, p. 3991–3991, 2017. Available: <<https://doi.org/10.1121/1.4989137>>. Citations on pages 30 and 45.

POLLARA, A. S.; SUTIN, A.; SALLOUM, H. Improvement of the Detection of Envelope Modulation on Noise (DEMON) and its application to small boats. **OCEANS 2016 MTS/IEEE Monterey, OCE 2016**, p. 1–10, 2016. Available: <<http://doi.org/10.1109/OCEANS.2016.7761197>>. Citations on pages 45 and 86.

PROULX, R.; WALDINGER, J.; KOPER, N. Anthropogenic landscape changes and their impacts on terrestrial and freshwater soundscapes. **Current Landscape Ecology Reports**, v. 4, 2019. ISSN 2364-494X. Available: <<https://doi.org/10.1007/s40823-019-00038-4>>. Citation on page 35.

RAIMBAULT, M.; DUBOIS, D. Urban soundscapes: experiences and knowledge. **Cities** 22(5), p. 339–350, 2005. Citation on page 27.

REIS, C. D. G. **FAV-Signature**. 2019. <<https://github.com/clausiusreis/FAV-Signature>>. [Online; accessed 24-March-2019]. Available: <<https://doi.org/10.5281/zenodo.3252202>>. Citation on page 86.

REIS, C. D. G.; PADOVESE, L. R.; OLIVEIRA, M. C. F. de. Automatic detection of vessel signatures in audio recordings with spectral amplitude variation signature. **Methods in Ecology and Evolution**, v. 10, n. 9, p. 1501–1516, 2019. Available: <<https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13245>>. Citation on page 55.

REIS, C. D. G.; SANTOS, T. N.; OLIVEIRA, M. C. F. de. A visualization framework for feature investigation in soundscape recordings. In: **2018 22nd International Conference Information Visualisation (IV)**. [S.l.: s.n.], 2018. p. 490–497. ISSN 2375-0138. Citations on pages 29 and 44.

SÁNCHEZ-GENDRIZ, I.; PADOVESE, L. R. Underwater soundscape of marine protected areas in the coastal of Sao Paulo, Brazil. **Marine Pollution Bulletin**, Elsevier Ltd, v. 105, p. 65–72, 2015. Available: <<http://doi.org/10.1016/j.marpolbul.2016.02.055>>. Citations on pages 28, 50, and 79.

_____. Temporal and spectral patterns of fish choruses in two protected areas in southern Atlantic. **Ecological Informatics**, v. 38, p. 31–38, 2017. Available: <<https://doi.org/10.1016/j.ecoinf.2017.01.003>>. Citations on pages 50 and 79.

SLAMNOIU, G.; RADU, O.; ROSCA, V.; PASCU, C.; DAMIAN, R.; SURDU, G.; CURCA, E.; RADULESCU, A. DEMON-type algorithms for determination of hydro-acoustic signatures of surface ships and of divers. **IOP Conference Series: Materials Science and Engineering**, v. 145, n. 8, 2016. Available: <<https://doi.org/10.1088/1757-899X/145/8/082013>>. Citations on pages 30 and 45.

Stowell, D.; Benetos, E.; Gill, L. F. On-bird sound recordings: Automatic acoustic recognition of activities and contexts. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 25, n. 6, p. 1193–1206, June 2017. Citations on pages 38 and 40.

STRÖM, C. Rapid biodiversity assessment of a Neotropical rainforest using soundscape recordings. n. June, p. 24, 2013. Citation on page 28.

SWANSON, F. J.; KRATZ, T. K.; CAINE, N.; WOODMANSEE, R. G. Landform effects on ecosystem patterns and processes. **BioScience** 38(2), p. 92–98, 1988. Citation on page 27.

TOWSEY, M.; ZHANG, L. False-colour spectrograms of long duration acoustic recordings. **Report**, 2014. Available: <<http://eprints.qut.edu.au/66786/>>. Citation on page 29.

TOWSEY, M.; ZHANG, L.; COTTMAN-FIELDS, M.; WIMMER, J.; ZHANG, J.; ROE, P. Visualization of long-duration acoustic recordings of the environment. **Procedia Computer Science**, Elsevier Masson SAS, v. 29, n. Krause 2008, p. 703–712, 2014. ISSN 18770509. Available: <<http://dx.doi.org/10.1016/j.procs.2014.05.063>>. Citation on page 27.

Truskinger, A.; Cottman-Fields, M.; Eichinski, P.; Towsey, M.; Roe, P. Practical analysis of big acoustic sensor data for environmental monitoring. In: **2014 IEEE Fourth International Conference on Big Data and Cloud Computing**. [S.l.: s.n.], 2014. p. 91–98. Citations on pages 29, 38, 39, 40, and 65.

TRUSKINGER, A.; TOWSEY, M.; ROE, P. Decision support for the efficient annotation of bioacoustic events. **Ecological Informatics**, v. 25, p. 14 – 21, 2015. ISSN 1574-9541. Citation on page 29.

TRUSKINGER, A. M. **Semi-automated annotation of environmental acoustic recordings**. Phd Thesis (PhD Thesis) — Queensland University of Technology, 2015. Available: <<https://eprints.qut.edu.au/84543/>>. Citations on pages 29, 38, 39, and 40.

TUTEUR, F. B. **Detection of wide-band signals modulated by a low-frequency sinusoid**. Appendix A-4, New Haven, Conn, USA: Yale University, 1963. Citation on page 45.

VEIRS, S.; VEIRS, V.; WOOD, J. D. Ship noise extends to frequencies used for echolocation by endangered killer whales. **PeerJ**, v. 4, p. e1657, 2016. Available: <<https://doi.org/10.7717/peerj.1657>>. Citations on pages 30 and 44.

VILLANUEVA-RIVERA, L. J.; PIJANOWSKI, B. C. Pumilio: A web-based management system for ecological recordings. **The Bulletin of the Ecological Society of America**, v. 93, n. 1, p. 71–81, 2012. Available: <<https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/0012-9623-93.1.71>>. Citations on pages 28 and 38.

VILLANUEVA-RIVERA, L. J.; PIJANOWSKI, B. C.; DOUCETTE, J.; PEKIN, B. A primer of acoustic analysis for landscape ecologists. **Landscape Ecology**, v. 26, n. 9, p. 1233, Jul 2011. Citation on page 55.

WANG, Y.; ZHANG, Q.; LIN, F.; GOH, C. K.; SEAH, H. S. Polarviz: a discriminating visualization and visual analytics tool for high-dimensional data. **The Visual Computer**, 05 2018. Citations on pages 15, 40, 41, and 42.

WIMMER, J.; TOWSEY, M.; PLANITZ, B.; WILLIAMSON, I.; ROE, P. Analysing environmental acoustic data through collaboration and automation. **Future Generation Computer Systems**, v. 29, n. 2, p. 560–568, 2012. ISSN 0167-739X. Citations on pages 29 and 65.

_____. Analysing environmental acoustic data through collaboration and automation. **Future Generation Computer Systems**, v. 29, n. 2, p. 560 – 568, 2013. ISSN 0167-739X. Special section: Recent advances in e-Science. Available: <<http://www.sciencedirect.com/science/article/pii/S0167739X1200057X>>. Citations on pages 38, 39, and 40.

WITTEKIND, D.; SCHUSTER, M. Propeller cavitation noise and background noise in the sea. **Ocean Engineering**, v. 120, p. 116–121, 2016. Available: <<http://dx.doi.org/10.1016/j.oceaneng.2015.12.060>>. Citations on pages 30 and 44.

YAN, J.; SUN, H.; CHENG, E.; KUAI, X.; ZHANG, X. Ship radiated noise recognition using resonance-based sparse signal decomposition. **Shock and Vibration**, v. 2017, 2017. Citation on page 45.

GLOSSARY

AIS: Automatic Identification System.

ANN: Artificial Neural Network.

CART: Classification And Regression Trees.

CSS: Cascading Style Sheets.

CSV: Comma Sepparated Values.

DEMON: Detection of Envelope Modulation On Noise.

FAV: Frequency Amplitude Variation.

FAVs: Frequency Amplitude Variation Signature.

FFMPEG: Fast Foward MPEG.

FFT: Fast Fourier Transform.

HHT: Hilbert Huang Transform.

HMM: Hidden Markov Model.

HMS: Hilbert Marginal Spectrum.

HTML: Hypertext Markup Language.

kNN: k-Nearest Neighbor.

LACMAN: Laboratory of Acoustics And Environment of the Polythecnic School of the University os São Paulo.

LDA: Linear Discriminant Analysis.

LSMSP: Laje de Santos Marine State Park.

MFCC: Mel-Frequency Cepstral Coefficients.

MP3: MPEG Audio Layer-3.

MPEG: Moving Picture Expert Group.

NDSI: Normalized-Difference Soundscape Index.

NMF/NNMF: Non-Negative Matrix Factorization.

Radviz: Radial Visualization.

RMS: Root Mean Square.

RSSD: Ressonance-Based Signal Sparse Decomposition.

SPL: Sound Pressure Level.

STFT: Short-Time Fourier Transform.

SVG: Scalable Vector Graphics.

SVM: Support Vector Machines.

WSGI: Web Server Gateway Interface.

WWW/Web: World Wide Web.

XJSP: Xixová-Japuí State Park.