UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Uma estratégia para análise visual de Paisagens Acústicas com base em seleção de características discriminantes

Fábio Felix Dias

Dissertação de Mestrado 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: _____

Fábio Felix Dias

Uma estratégia para análise visual de Paisagens Acústicas com base em seleção de características discriminantes

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

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientadora: Profa. Dra. Rosane Minghim

USP – São Carlos
Junho de 2018

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

D541e Dias, Fábio Felix
 Uma estratégia para análise visual de Paisagens
Acústicas com base em seleção de características
discriminantes / Fábio Felix Dias; orientadora
Rosane Minghim. -- São Carlos, 2018.
 133 p.

 Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2018.

 1. Visualização de Informação. 2. Paisagens
Acústicas. I. Minghim, Rosane, orient. II. Título.

Fábio Felix Dias

**A Soundscape visual analysis strategy based on
discriminate feature selection**

Master dissertation submitted to the Institute of
Mathematics and Computer Sciences – ICMC-USP,
in partial fulfillment of the requirements for the
degree of the Master Program in Computer Science
and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and
Computational Mathematics

Advisor: Profa. Dra. Rosane Minghim

USP – São Carlos
June 2018

Este trabalho é dedicado à todos que sofrem angustiados pela falta de esperança no amanhã.

AGRADECIMENTOS

Inicialmente, e principalmente, aos meus pais que sempre acreditaram e investiram na minha formação mesmo com todas as dificuldades imprimidas por uma renda apertada. Aos colegas do laboratório e do instituto que ajudaram direta ou indiretamente nesta pesquisa. Ao Prof. Dr. Hélio Pedrini (Unicamp) pela ajuda nos trabalhos com imagem e aprendizagem de máquina. À Profa. Dra. Rosane Minghim que orientou o desenvolvimento desta idéia. Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento durante esse mestrado. Por fim, ao Nosso Senhor Jesus Cristo, que ampara, concede ânimo e coragem para enfrentar as diversas dificuldades do caminho, através da interseção de sua mãe Maria.

“Et Verbum caro factum est...”
(Jo 1, 14)

RESUMO

DIAS, F. F. **Uma estratégia para análise visual de Paisagens Acústicas com base em seleção de características discriminantes**. 2018. 133 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2018.

O crescimento do volume de dados ocasionado pelo desenvolvimento tecnológico atual, tem sido fortemente evocado como premissa para a utilização de técnicas que auxiliem a exploração, análise e entendimento desses dados. Um conjunto dessas técnicas está compreendido na área de Visualização de Dados, que proporciona maneiras visuais de identificar padrões e tendências, além de extrair características obscuras dos dados. Tais abordagens podem ser aplicadas a qualquer problema que culmine na análise de dados. Um desses problemas é a utilização do som como ferramenta para descrever as características de uma paisagem, área denominada análise de Paisagens Acústicas. Nesta pesquisa de mestrado é apresentada uma abordagem visual para análise de Paisagens Acústicas. Essa abordagem consiste em duas etapas, sendo que na primeira são aplicadas técnicas visuais (Projeção Multidimensional t-SNE) e numéricas (Coeficiente de Silhueta) para avaliar quais conjuntos de características melhor descrevem uma Paisagem Acústica específica. A segunda etapa utiliza técnicas de visualização para analisar diferenças globais e características específicas de paisagens terrestres e aquáticas. Para isso foram utilizados o Mapa de Calor, as Coordenadas Paralelas e a xHiPP, uma extensão da projeção HiPP. A xHiPP busca melhorar a HiPP para alcançar melhor capacidade de análise e flexibilidade de aplicação. Com a aplicação das etapas apresentadas foi possível encontrar evidências de que o *Mel-frequency Cepstrum Coefficients* formam um conjunto de atributos eficaz para representação e segregação de Paisagens Acústicas. Também foi possível verificar que as técnicas de visualização empregadas na análise são capazes de destacar atributos semelhantes dos áudios, facilitando a análise, permitindo que o usuário dê enfoque às características relevantes do ambiente, no lugar de analisar áudios individuais para extrair informações.

Palavras-chave: visualização, características, espectrograma, projeção.

ABSTRACT

DIAS, F. F. **A Soundscape visual analysis strategy based on discriminate feature selection.** 2018. 133 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2018.

The growth of data volume caused by current technological development has been strongly evoked as a premise for use of techniques that help exploration, analysis, and understanding of data. A subset of these techniques is yielded by the field of Data Visualization, which provides visual manners to identify patterns and trends, as well as the extraction of hidden data features. Such approaches can be applied to problems that aim at data analysis with a strong exploratory component. One such problem is the use of sound as a tool to describe environmental landscapes, named ecological Soundscapes. A visual approach to analysis Soundscapes is presented in this master research. The approach contains two steps, and the first step applies visual (tSNE Multidimensional Projection) and numeric (Silhouette Coefficient) techniques to evaluate attributes groups that better describe a specific Soundscape. The second step employs visual techniques to analysis global differences and specific features of the terrestrial and underwater environment. To achieve these goals, the research used Heatmap, Parallel Coordinates, and xHiPP, an extension of HiPP projection. The xHiPP enhanced HiPP to improve its analytical capabilities and flexibility. The presented steps were able to show evidence of the Mel-frequency Cepstrum Coefficients is an effective attribute collection to represent and segregate Soundscapes. As well, visual techniques employed in the analysis are capable to highlight similar audio features, making exploration easy, allowing users to focus relevant environmental attributes, instead of analyzing individual audios to extraction some information.

Keywords: visualization, features, spectrogram, projection.

LISTA DE ILUSTRAÇÕES

Figura 1 – Exemplo de Espectrograma.	37
Figura 2 – Envelope e Transformada e Hilbert.	38
Figura 3 – Divisões dos eixos de tempo e frequência para cálculo do ACI.	42
Figura 4 – Espectro bioacústico (intensidade x frequência)	43
Figura 5 – Exemplo de Matrizes de Coocorrência.	46
Figura 6 – Exemplo do processo do descritor LBP.	47
Figura 7 – Transformada e Fourier de uma imagem de textura	48
Figura 8 – Esquema do descritor WLD	49
Figura 9 – Ilustração de arquitetura de um Autoencoder.	51
Figura 10 – Exemplo de projeção t-SNE	54
Figura 11 – Exemplo de particionamento, projeção e interação com a HiPP.	56
Figura 12 – Exemplo de Coordenadas Paralelas utilizando a base de dados Íris.	57
Figura 13 – Exemplo de Mapa de calor	58
Figura 14 – Representação de arquivo de áudio no REAL.	65
Figura 15 – Variação mensal do PSD.	66
Figura 16 – Mapa de calor apresentando a variação do NDSI por horários do dia.	67
Figura 17 – Espectrogramas dos índices.	68
Figura 18 – Combinação dos índices para gerar a imagem de falsa cor.	68
Figura 19 – Espectrograma de um período de 24 horas	69
Figura 20 – Espectrograma mensal.	69
Figura 21 – Sumário Acústico Estendido.	70
Figura 22 – Distribuição dos neurônios SOM	71
Figura 23 – Agrupamentos dos minutos de um espectrograma de falsa cor.	71
Figura 24 – Valores do SPL	72
Figura 25 – Sumarização utilizando média de espectrograma e SPL	73
Figura 26 – Sumarização utilizando mediana de espectrograma e SPL	74
Figura 27 – Visualização <i>SPL-gram</i>	75
Figura 28 – Passos para análise dos dados acústicos.	77
Figura 29 – xHiPP de arquivos de áudio.	82
Figura 30 – Coordenadas Paralelas de características de áudio	83
Figura 31 – Boxplot de características de áudio	83
Figura 32 – Mapa de Calor do índice H	84
Figura 33 – Decomposição de Série Temporal do índice H	84

Figura 34 – Variação da ordem das etapas da xHiPP	88
Figura 35 – Variação do algoritmo de agrupamento da xHiPP	89
Figura 36 – Variação do algoritmo de projeção da xHiPP	90
Figura 37 – Variação do agrupamento, da projeção e da ordem das etapas da HiPP	91
Figura 38 – Apresentação de características X Projeções Multidimensionais	94
Figura 39 – Projeção t-SNE de 4.340 áudios utilizando 27 índices e medidas acústicas como características. As cores dos pontos representam as áreas de coleta dos dados.	96
Figura 40 – Projeção t-SNE de cada uma das bases de dados separadamente, utilizando 27 índices e medidas acústicas como características	97
Figura 41 – Projeção t-SNE das bases utilizando as características acústicas divididas em: Índices Acústicos, MFCC e Outras Medidas Acústicas.	97
Figura 42 – Projeção t-SNE das bases separadas utilizando conjuntos distintos de características acústicas.	97
Figura 43 – Projeção t-SNE de 4.340 áudios utilizando 600 valores de descritores de textura e cor.	98
Figura 44 – Projeção t-SNE de cada uma das bases de dados separadamente, utilizando 600 valores de características geradas pelos descritores de textura e cor. O grupo destacado na Laje de Santos apresenta áudios com pouco ou nenhum som.	99
Figura 45 – Projeção t-SNE das bases utilizando conjuntos distintos de características de imagem.	99
Figura 46 – Projeção t-SNE das bases separadas utilizando conjuntos distintos de características de imagem. Os grupos destacados na Laje de Santos apresentam áudios com pouco ou nenhum som.	100
Figura 47 – Teste com descritor de textura WLD	101
Figura 48 – Exemplo do pré-processamento de um espectrograma	101
Figura 49 – Arquitetura de Autoencoder utilizada.	102
Figura 50 – Reconstrução das partes do espectrograma apresentado na Figura 48. Cada imagem está rotulada entre s01 e s20.	102
Figura 51 – Características extraídas pelo Autoencoder de partes de um espectrograma	103
Figura 52 – Espectrogramas para testes com Autoencoder	104
Figura 53 – Características extraídas pelo Autoencoder de 10 espectrogramas	104
Figura 54 – Características extraídas pelo Autoencoder de 173 imagens de teste	105
Figura 55 – Coordenadas Paralelas de índices e medidas acústicas de Costa Rica	108
Figura 56 – Mapa de Calor do SPL da área CostaRica1	109
Figura 57 – xHiPP dos dados de Costa Rica utilizando MFCC como características, rotulados pelas áreas de coleta	110

Figura 58 – xHiPP dos dados de Costa Rica utilizando MFCC como características, rotulados pelo período do dia	111
Figura 59 – xHiPP dos dados de Ilhéus utilizando MFCC como características	111
Figura 60 – xHiPP dos dados da Laje de Santos utilizando Índices Acústicos como características	112
Figura 61 – Separação de Fontes de Áudio	127
Figura 62 – Segmentação de imagem de espectrograma	128
Figura 63 – Valores de silhueta para técnicas de seleção automática de características acústicas	130
Figura 64 – Projeção t-SNE dos dados com as medidas acústicas selecionadas automaticamente.	130
Figura 65 – Valores de silhueta para técnicas de seleção automática de descritores de imagem	131
Figura 66 – Projeção t-SNE dos dados com os descritores de imagem selecionados automaticamente.	131
Figura 67 – Valores de silhueta para técnicas de seleção automática de 23.040 características geradas pelo Autoencoder	132
Figura 68 – Valores de silhueta para técnicas de seleção automática de 160 características sumarizadas do Autoencoder	132
Figura 69 – Projeção t-SNE dos dados com as medidas extraídas com o Autoencoder selecionadas automaticamente.	133
Figura 70 – Projeção t-SNE dos dados com a sumarização das medidas extraídas com o Autoencoder selecionadas automaticamente.	133

LISTA DE QUADROS

Quadro 1 – Lista de artigos que trabalham com Índices Acústicos.	62
Quadro 2 – Lista de bases de dados utilizadas para testes com a xHiPP.	88

LISTA DE TABELAS

Tabela 1 – Valores de silhueta para características normalizadas X projeções (5 classes)	93
Tabela 2 – Valores de silhueta para características normalizadas X projeções (4 classes)	95
Tabela 3 – Valores de silhueta para dados com conjuntos distintos de características . .	106
Tabela 4 – Valores de silhueta do espaço original utilizando medidas acústicas e descri- tores de imagem	125
Tabela 5 – Valores de silhueta para características X projeções	126
Tabela 6 – Valores de silhueta dos dados considerando as características acústicas divi- didas em 3 grupos, índices, MFCC e outras métricas.	126
Tabela 7 – Valores de silhueta dos dados considerando os descritores de imagem dividi- dos em 2 grupos, textura e cor	126

LISTA DE ABREVIATURAS E SIGLAS

D_{cf}	<i>Cumulative frequency dissimilarity</i>
H_f	Entropia Espectral
H_t	Entropia Temporal
H'	Índice de Shannon
ACC	Auto Correlograma de Cores - <i>Auto Color Correlation</i>
ACI	Índice de Complexidade Acústica - <i>Acoustic Complexity Index</i>
ADI	Índice de Diversidade Acústica - <i>Acoustic Diversity Index</i>
AEI	Índice de Uniformidade Acústica - <i>Acoustic Evenness Index</i>
AR	Índice de Riqueza Acústica - <i>Acoustic Richness</i>
BIC	Classificação de Pixel Borda/Interior - <i>Border/Interior Pixel Classification</i>
BIO	Índice Bioacústico - <i>Bioacoustic Index</i>
CCV	Vetor de Coerência de Cor - <i>Color Coherence Vector</i>
CVR	Índice de Cobertura Acústica - <i>Acoustic Cover Index</i>
D	Índice de Dissimilaridade Acústica - <i>Acoustic Dissimilarity</i>
DBSCAN	<i>Density-based spatial clustering of applications with noise</i>
DCT	Transformada Discreta de Cosseno - <i>Discrete Cosine Transform</i>
DNN	Rede Neural Profunda - <i>Deep Neural Network</i>
EASY	<i>Extended Acoustic Summary</i>
EM	<i>Expectation Maximization</i>
FLAC	<i>Free Lossless Audio Codec</i>
G	Coeficiente de Gini-Simpson
GCH	Histograma Global de Cor - <i>Global Color Histogram</i>
GFCC	<i>Gammatone-frequency Cepstrum Coefficients</i>
GLCM	Matriz de Coocorrência - <i>Gray Level Co-occurrence Matrix</i>
GMM	<i>Gaussian Mixture Model</i>
H	Índice de Entropia Acústica - <i>Acoustic Entropy Index</i>
HiPP	<i>Hierarchical Point Placement Strategy</i>
InfoVis	Visualização da Informação - <i>Information Visualization</i>
kNN	<i>k-Nearest Neighbor</i>
LAMP	<i>Local Affine Multidimensional Projeciton</i>
LBP	<i>Local Binary Patterns</i>

LDA	<i>Latent Dirichlet Allocation</i>
LoMAX	<i>Local maximum</i>
LSP	<i>Least Square Projection</i>
MaxAT	<i>Maximal Area Triangulation</i>
MDS	<i>Multidimensional Scaling</i>
MFCC	Coeficientes Mel-Cepstrais - <i>Mel-frequency Cepstrum Coefficients</i>
MSP	<i>Mean Square Pressure</i>
NDSI	Índice de Diferença Normalizada - <i>Normalized Difference Soundscape Index</i>
NMF	Fatoração de Matriz Não-negativa - <i>Non-Negative Matrix Factorization</i>
NP	<i>Number of Peaks</i>
PCA	Análise de Componentes Principais - <i>Principal Component Analysis</i>
PLMP	<i>Part-Linear Multidimensional Projection</i>
PSD	Densidade Espectral - <i>Power Spectral Density</i>
REAL	<i>Remote Environmental Assessment Laboratory</i>
RMS	Valor Quadrático Médio - <i>Root Mean Square</i>
SAI	<i>Stabilized Auditory Image</i>
SIF	<i>Spectrograma Image Features</i>
SNR	<i>Signal-to-noise ratio</i>
SOM	<i>Self-organizing Map</i>
SPL	Nível de Pressão Sonora - <i>Sound Pressure Level</i>
STFT	Transformada de Fourier de Tempo-Curto - <i>Short-Time Fourier Transform</i>
SVM	<i>Support Vector Machine</i>
t-SNE	<i>t-Stochastic Neighbor Embedding</i>
WAV	<i>WAVEform audio format</i>
WLD	<i>Weber Local Descriptor</i>
xHiPP	HiPP Estendida - <i>eXtended HiPP</i>

SUMÁRIO

1	INTRODUÇÃO	29
1.1	Objetivo e Contribuições	32
1.2	Organização do Texto	32
2	CONCEITOS FUNDAMENTAIS	35
2.1	Som e suas características físicas	35
2.2	Análise de Sinais e Áudio	36
2.2.1	<i>Transformada de Fourier</i>	36
2.2.2	<i>Transformada de Hilbert</i>	37
2.2.3	<i>Densidade Espectral</i>	38
2.2.4	<i>Nível de Pressão Sonora</i>	38
2.2.5	<i>Variação do Sinal</i>	39
2.2.6	<i>Coeficientes Mel-Cepstrais</i>	39
2.3	Índices Acústicos	40
2.3.1	<i>Índice de Shannon</i>	40
2.3.2	<i>Coeficiente de Gini-Simpson</i>	41
2.3.3	<i>Índice de Diversidade Acústica</i>	41
2.3.4	<i>Índice de Uniformidade Acústica</i>	41
2.3.5	<i>Índice de Complexidade Acústica</i>	41
2.3.6	<i>Índice Bioacústico</i>	43
2.3.7	<i>Índice de Entropia Acústica</i>	43
2.3.8	<i>Índice de Diferença Normalizada</i>	44
2.3.9	<i>Índice de Riqueza Acústica</i>	44
2.3.10	<i>Outros índices e medidas acústicas</i>	44
2.4	Segmentação	45
2.5	Representação e Descrição de Imagens	46
2.5.1	Descritores de Textura	46
2.5.1.1	<i>Matriz de Coocorrência</i>	46
2.5.1.2	<i>Local Binary Patterns</i>	47
2.5.1.3	<i>Descritor Espectral</i>	48
2.5.1.4	<i>Weber Local Descriptor</i>	48
2.5.2	Descritores de Cor	49
2.5.2.1	<i>Histograma Global de Cor</i>	49

2.5.2.2	<i>Auto Correlograma de Cores</i>	50
2.5.2.3	<i>Vetor de Coerência de Cor</i>	50
2.5.2.4	<i>Classificação de Pixel Borda/Interior</i>	50
2.6	Autoencoder	50
2.7	Técnicas de Visualização de Dados	52
2.7.1	<i>Projeções Multidimensionais</i>	53
2.7.2	<i>Avaliação de Projeções Multidimensionais</i>	54
2.7.3	<i>Hierarchical Point Placement Strategy</i>	55
2.7.4	<i>Coordenadas Paralelas</i>	57
2.7.5	<i>Mapa de Calor</i>	57
2.8	Considerações sobre os conceitos apresentados	57
3	TRABALHOS RELACIONADOS	59
3.1	Trabalhos que utilizam Índices Acústicos	59
3.2	Análises de Paisagens Acústicas que utilizam outras Medidas Acústicas	61
3.3	Trabalhos que segmentam o Áudio	63
3.4	Trabalhos que segmentam a Imagem do Espectrograma	64
3.5	Indexação de Dados Acústicos	65
3.6	Trabalhos que implementam avaliação do ambiente	65
3.7	Visualização de gravações de longa duração	67
3.8	Considerações sobre os trabalhos apresentados	73
4	ANÁLISE DE CARACTERÍSTICAS E VISUALIZAÇÃO DE PAISAGENS ACÚSTICAS	77
4.1	Extração de Características	78
4.1.1	<i>Características de Áudio</i>	78
4.1.2	<i>Características de Imagem</i>	78
4.2	Avaliação das Características	78
4.3	Análise Visual	80
4.3.1	HiPP Estendida	80
4.3.1.1	<i>Modificações Realizadas</i>	81
4.3.1.2	<i>Aplicação</i>	81
4.3.2	<i>Coordenadas Paralelas</i>	82
4.3.3	<i>Mapa de Calor</i>	83
4.4	Combinação das Etapas de Análise	85
4.5	Detalhes de Implementação	86
5	EXPERIMENTOS E RESULTADOS	87
5.1	HiPP Estendida	87

5.1.1	Resultados	87
5.1.2	Discussão	90
5.2	Exploração de Paisagens Acústicas	92
5.2.1	Bases de Dados	92
5.2.2	Extração e Avaliação de Características	93
5.2.2.1	<i>Segmentação de Imagens</i>	95
5.2.2.2	<i>Índices e Medidas Acústicas</i>	95
5.2.2.3	<i>Descritores de Imagem</i>	98
5.2.2.4	<i>Características de Imagem extraídas por Autoencoder</i>	100
5.2.2.5	<i>Comparação e discussão sobre as características extraídas</i>	104
5.2.3	Análise Visual	107
5.2.3.1	<i>Coordenadas Paralelas</i>	107
5.2.3.2	<i>Mapa de Calor</i>	107
5.2.3.3	<i>HiPP estendida</i>	107
5.2.3.4	<i>Discussão</i>	109
6	CONCLUSÕES	113
6.1	Contribuições	113
6.2	Limitações	114
6.3	Trabalhos Futuros	115
REFERÊNCIAS		117
APÊNDICE A	SILHUETA DE CONJUNTOS	125
APÊNDICE B	SEGMENTAÇÃO DE ESPECTROGRAMA	127
APÊNDICE C	RESULTADOS DE SELEÇÃO AUTOMÁTICA DE ATRIBUTOS	129
C.1	Índices e Metricas Acústicas	129
C.2	Descritores de Imagem	131
C.3	Características de Imagem extraídas por Autoencoder	132

INTRODUÇÃO

O som é uma propriedade perpétua e dinâmica de um determinado ambiente e está intimamente ligado à caracterização desse local (PIJANOWSKI *et al.*, 2011b). Os padrões acústicos dessa área são sensíveis às suas características (temperatura, umidade, ruído etc.) e qualquer mudança nesse meio refletem no seu comportamento acústico (KRAUSE, 2016). Desse modo, o som pode ser utilizado como um indicador para mensurar a saúde e biodiversidade de um ambiente (SERVICK, 2014).

De acordo com Pijanowski *et al.* (2011a), a expressão *Soundscape* (definida em português como Paisagem Sonora ou Paisagem Acústica) é utilizada para descrever a relação entre uma paisagem e os seus sons, e algumas das primeiras citações do termo podem ser encontradas nos trabalhos de Southworth (1969) e Schafer (1977).

Uma Paisagem Acústica é, portanto, o conjunto de todos os sons que emanam de um ambiente, definindo um padrão acústico único no tempo e no espaço (PIJANOWSKI *et al.*, 2011b). Os sons da Paisagem Acústica são categorizados como sons de Biofonia (biophony), Geofonia (geophony) e Antrofonia (anthrophony). Krause (1987) apresenta Biofonia como o conjunto de todos os sons biológicos produzidos por organismos vivos e Geofonia como os sons não biológicos ou geofísicos criados por eventos como chuva, vento, trovão, movimentação da terra etc. Pijanowski *et al.* (2011b) estenderam essa taxonomia adicionando o conceito de Antrofonia, que se refere aos sons gerados pelo ser humano (antropológicos), como, por exemplo, vozes, ruídos de veículos de transporte, indústrias, ar-condicionado, entre outros. O termo Tecnofonia (technophony) também é encontrado em alguns trabalhos para designar apenas os sons criados por máquinas em geral (GAGE, 2008).

O estudo das Paisagens Acústicas procura descrever a diversidade de sons e, através deles, entender as alterações ambientais causadas pelo crescente tráfego de aviões e embarcações, pelas construções, pela chegada de novas espécies ou pelo efeito das mudanças climáticas (SERVICK, 2014). Vale ressaltar que esse tipo de estudo é diferente de áreas como a Ecologia Acústica

(*Acoustic Ecology*) e a Bioacústica (*Bioacustics*). A primeira é uma complementação da ecologia tradicional, focada no ser humano e nos sons de um ambiente incluindo música e *design* acústico (SCHAFER, 1977), enquanto a última se concentra nos sons produzidos por uma determinada espécie ou na comparação entre espécies (FLETCHER, 2007). Assim, essas três áreas possuem escopos distintos. Enquanto a Bioacústica está ligada à análise de uma ou duas espécies animais, a Ecologia Acústica foca na espécie humana e o ambiente acústico ao seu redor. Por outro lado, a Paisagem Acústica é o conjunto de todos os sons e suas inter-relações em um determinado ambiente.

Alguns estudos nessa área de Paisagens Acústicas obtiveram os seguintes resultados: determinação da influência da urbanização na vida animal (JOO; GAGE; KASTEN, 2011), estimativa da diversidade de pássaros em determinadas áreas florestais (PIERETTI; FARINA; MORRI, 2011), mensuração da diversidade animal em áreas alagadas (DEPRAETERE *et al.*, 2012) e quantificação da biodiversidade em regiões marinhas (PARKS; MIKSIS-OLDS; DENES, 2014). Além desses, outros trabalhos (cf. Capítulo 3) têm descrito resultados de análise desse tipo de dados, confirmando a utilidade da pesquisa em Paisagens Acústicas.

A Análise de Paisagens Acústicas deve considerar as características dos meios onde o som se propaga, porque existem diferenças entre a propagação do som em ambientes terrestres e aquáticos (cf. Seção 2.1). Normalmente sistemas marinhos são dominados por sons abióticos (geofonia) (PARKS; MIKSIS-OLDS; DENES, 2014). Outra questão apresentada por Parks, Miksis-Olds e Denes (2014) diz respeito à perda de energia acústica em meios aquáticos, que é menor que no ar. Dessa maneira, ondas sonoras alcançam distâncias maiores em ambientes subaquáticos, fazendo com que sons biológicos (biofonia) de uma região precisem competir com fontes acústicas de regiões distantes da posição em análise.

Segundo Neal (2012), um problema comum para a análise e classificação de áudio ambiental é o fato do processo ser manual e, portanto, trabalhoso, pois necessita de observadores treinados e experientes para o trabalho de coleta e análise. Além disso, o mesmo autor menciona que o fato do observador precisar mover-se fisicamente entre os sítios de coleta, limita a quantidade de coletas, dificulta a logística e eleva os custos do processo. Vale ressaltar também, que a presença humana possivelmente gera distúrbios no ambiente analisado (SERVICK, 2014).

A crescente quantidade de áudios analisados pelos especialistas também é um desafio. Por causa do constante melhoramento dos aparelhos de captura e gravação de sons e devido ao barateamento das mídias para armazenamento de dados, o volume cresce rapidamente. Por exemplo, o laboratório mencionado por Servick (2014), em um espaço de 5 anos, gerou cerca de 85 *terabytes*, algo em torno de 100.000 horas de gravações; na mesma linha o laboratório mencionado em Towsey *et al.* (2014b), em 6 anos, gerou 28 *terabytes* de dados; e no trabalho de Sankupellay *et al.* (2015) são citadas 259.000 gravações, durante 7 anos, completando aproximadamente 35 *terabytes*. Assim sendo, os especialistas necessitam tanto de estratégias de coleta que possam diminuir as incontáveis horas gastas dentro de sítios quanto de técnicas para

análise mais eficientes que possam agilizar a aquisição de conhecimento desses dados.

Por ser a visão um sentido chave no entendimento de uma informação, a utilização de técnicas de visualização (cf. Seção 2.7) são úteis para dar suporte à análise de dados. Essas técnicas comunicam informações através de representações gráficas geradas por computador, apresentando visões alternativas dos dados e auxiliando a descrição de suas estruturas, padrões e anomalias (WARD; GRINSTEIN; KEIM, 2015). Sendo assim, é provável que essas técnicas, aplicadas à análise de Paisagens Acústicas, possam auxiliar na tomada de decisões referentes ao controle de espécies, no monitoramento de áreas de preservação ambiental, e na análise da qualidade do ambiente e de sua diversidade animal. Além disso, podem, dar apoio à busca de respostas sobre questões específicas referentes à influência de sons antropológicos no ambiente, sobre alterações devido ao clima ou à poluição, à migração de espécies etc.

No contexto de análise e compreensão de dados acústicos, algumas abordagens foram propostas para sumarização, visualização e compreensão desses dados. Por exemplo, algumas delas envolvem a utilização de Índices Acústicos, que são funções matemáticas que avaliam aspectos da biodiversidade a partir dos sons (SUEUR *et al.*, 2014), a combinação deles para gerar visualizações de imagens de falsa cor capazes de revelar o conteúdo acústico de um ambiente e facilitar a navegação pelas gravações (TOWSEY *et al.*, 2014b) e a Segmentação de Áudio, capaz de identificar grupos de sons significativos no domínio do tempo ou da frequência (cf. Seção 3.3).

A Segmentação de Áudio pode ser realizada tanto no domínio do tempo quanto no domínio da frequência a depender das necessidades da aplicação. Nos casos de análise de Paisagens Acústicas, onde vários sons de diferentes amplitudes e frequências são emitidos no mesmo instante de tempo, a utilização do domínio da frequência mostra-se mais apropriada (NEAL, 2012). Alguns trabalhos que utilizam esse tipo de estratégia preferem fazer a segmentação da imagem do espectrograma, no lugar de utilizar os valores complexos gerados pela Transformada de Fourier (cf. Seção 2.2.1) como pode ser visto na Seção 3.4.

Pode-se notar a variedade de abordagens e resultados provenientes do estudo de Paisagens Acústicas. No entanto, não se observa um estudo que defina sistematicamente uma melhor caracterização dos dados acústicos, com a finalidade de distinguir ambientes naturais ou refletir suas alterações em um longo espaço de tempo.

A maior parte dos trabalhos (cf. Capítulo 2 e 3) está focada em apresentar como uma determinada medida ou conjunto de medidas são capazes de descrever um ambiente específico, mas não apresentam caminhos para identificar como essas características podem se adequar aos atributos e necessidades de outros ambientes.

1.1 Objetivo e Contribuições

No contexto apresentado acima, o presente trabalho de mestrado tem como finalidade definir uma estratégia de exploração e análise de Paisagens Acústicas, capaz de evidenciar características dos dados que revelem semelhanças e diferenças entre ambientes naturais distintos. A partir disso, as principais questões de pesquisa investigadas foram a identificação de atributos que segregassem ambientes naturais e uma forma visual adequada de investigá-los e avaliá-los.

Durante a pesquisa foi necessário investigar quais conjuntos de características melhor se aplicam à descrição de uma Paisagem Acústica específica (índices e medidas acústicas e descritores de imagem de espectrograma) e quais ferramentas de visualização poderiam auxiliar o processo de análise desses dados.

A busca por estratégias visuais se justifica pela falta de entendimento pleno de um conjunto de atributos que permitam diferenciar Paisagens Acústicas. A avaliação atual dessas características ainda é exploratória e sujeita à interpretação do usuário. Além disso, tais estratégias visuais têm se mostrado úteis na tarefa de avaliar conjuntos de atributos para domínios variados como coleções de textos, imagens, séries temporais etc. A busca por um conjunto de características que, ao descrever uma Paisagem Acústica, permita destacar distinção entre tipos de ambientes, se justifica pela necessidade de processar e filtrar a grande quantidade de dados disponíveis, até mesmo correspondentes a anos de gravação. No futuro um processo automático poderá ser utilizado para essa tarefa.

Como resultado dessa pesquisa, as contribuições que podem ser destacadas são: *i*) definição de uma abordagem visual e numérica para escolha e validação de características relevantes para representação de dados acústicos; *ii*) a aplicação de técnicas visuais para exploração dos dados representados pelas características e apoio na sua validação; e *iii*) a implementação, avaliação e aplicação da xHiPP, uma extensão da projeção multidimensional HiPP, na exploração desses dados.

Os resultados demonstram a pertinência do uso de Índices e Medidas Acústicas (cf. Capítulo 2) extraídos do espectro (domínio da frequência) e do sinal original (domínio do tempo). Essas estratégias se mostraram mais capazes de segregar dados acústicos, do que estratégias de extração de atributos sobre a imagem do espectrograma. Adicionalmente, os resultados destacam a importância da visualização de espaços de atributos como maneira de apoiar a validação da escolha de características adequadas para a descrição de Paisagens Acústicas.

1.2 Organização do Texto

Este trabalho está organizado como segue. No Capítulo 2, são apresentados conceitos fundamentais de processamento de sinais e áudio, necessários para o entendimento das características utilizadas para análise de Paisagens Acústicas. Além desses conceitos, são apresentadas,

também, várias técnicas utilizadas para representação dos áudios. Também são apresentadas noções sobre descrição de imagens, Aprendizagem de Máquinas e Visualização de Dados. Os trabalhos relacionados são apresentados no Capítulo 3, com foco em trabalhos que aplicaram técnicas para descrição, análise, visualização e exploração de Paisagens Acústicas. No Capítulo 4, é descrita a abordagem utilizada para análise de dados acústicos, além das modificações realizadas na projeção HiPP para sua utilização na análise de dados. O Capítulo 5 apresenta experimentos, resultados obtidos e discussões sobre os resultados dos experimentos. O Capítulo 6 discute os resultados e tece observações sobre este trabalho e suas ramificações.

CONCEITOS FUNDAMENTAIS

Para áreas como Visualização de Dados, Mineração de Dados e Aprendizagem de Máquina, a extração de características visa obter um conjunto de valores que possa representar um conjunto de dados de maneira relevante e com pouca ou nenhuma redundância. Então, um item de um determinado domínio (áudio, texto, imagem etc.) pode ser representado como um vetor numérico, conhecido como vetor de características, de atributos, de variáveis ou de dimensões.

Para cada domínio de dados, existem técnicas distintas para extração de suas características. Para imagens, existem descritores como os apresentados na Seção 2.5. Para texto, as características podem ser baseadas nas relações de frequência de termos dentro de cada texto e dentro do conjunto de textos analisado. Para áudio, podem ser utilizados os índices e medidas descritos em algumas das próximas seções deste capítulo.

Assim sendo, neste capítulo são apresentadas características de áudio, conceitos utilizados na Análise de Sinais e de Dados Acústicos. Adicionalmente, são apresentadas definições de Segmentação, Descrição de Imagens, Aprendizagem de Máquina e Visualização de Dados que serão utilizadas no projeto. Esses conceitos são considerados essenciais para o desenvolvimento da pesquisa descrita neste trabalho de mestrado.

2.1 Som e suas características físicas

O som é uma onda mecânica e longitudinal que pode se propagar através de meios físicos como sólidos, líquidos ou gases (HALLIDAY; RESNICK; WALKER, 2009). Sua velocidade depende da frequência e do comprimento da onda. Além disso, a velocidade do som depende do meio onde ele se propaga. Por exemplo, a velocidade do som na água é maior que no ar (HALLIDAY; RESNICK; WALKER, 2009). No ar, a uma temperatura de 20 °C, ela chega aos 343 m/s. Na água, à mesma temperatura, a 1.482 m/s. E no mar, com um grau de salinidade em

torno de 3,5, a velocidade de propagação do som alcança 1.522 m/s.

Além da frequência, que informa se o som é grave ou agudo, do comprimento de onda e da sua velocidade, podem ser consideradas mais algumas características do som que influenciam sua análise, como: a Amplitude da onda (deslocamento máximo para qualquer um dos lados do estado de equilíbrio da onda); a Intensidade Sonora (taxa média de transmissão de energia por unidade de área); Nível Sonoro (relação entre a intensidade de um som analisado e o limite inferior de intensidade suportado pelo ser humano); e o Formato da onda (que define o timbre do som, o que diferencia sons de mesma frequência emitidos por fontes distintas) (HALLIDAY; RESNICK; WALKER, 2009).

2.2 Análise de Sinais e Áudio

Nesta seção são descritas ferramentas utilizadas na Análise de Sinais e de Áudio que são necessárias para o entendimento dos métodos utilizados para análise de Paisagens Acústicas. São apresentadas apenas as intuições básicas de cada uma sem muito aprofundamento teórico ou matemático. Maiores explicações podem ser encontradas nas referências apresentadas.

2.2.1 Transformada de Fourier

O matemático Jean Baptiste Fourier descreveu que qualquer função periódica pode ser expressa por uma soma de senos ou cossenos com diferentes frequências, cada uma multiplicada por um coeficiente diferente (Série de Fourier). Funções não periódicas podem ser descritas como uma integral de senos ou cossenos multiplicados por alguma função de ponderação (Transformada de Fourier) (Equação 2.1) (GONZALEZ; WOODS, 2010). Assim, essa transformação decompõe a função original em funções mais simples de analisar. Uma função expressa pela Transformada de Fourier pode ser reconstruída por meio do processo inverso (Equação 2.2) sem nenhuma perda.

$$\mathfrak{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t) \exp^{-i2\pi ut} dt = F(u) \quad (2.1)$$

$$f(t) = \mathfrak{F}^{-1}\{F(u)\} = \int_{-\infty}^{\infty} F(u) \exp^{i2\pi ut} du \quad (2.2)$$

Com base nessa transformada é criado o espectrograma que representa a relação entre frequências (eixo x) e intensidade do sinal (eixo y). A partir dessa idéia, pode-se deduzir a forma discreta da transformada e estendê-la para comportar duas dimensões (GONZALEZ; WOODS, 2010). Essas abordagens tanto podem ser aplicadas ao processamento de sinais quanto ao de imagens, por exemplo. Com base no conceito da transformada para duas variáveis, é criado um

espectrograma que representa a relação das frequências (eixo y), com o tempo (eixo x) e com a intensidade (cores). Vale ressaltar que não existe um padrão para a faixa de cores utilizadas no espectrograma. A Figura 1 apresenta uma dessas visualizações. Ela representa os 10 s iniciais de uma gravação da música *Die Schnitzelbank*¹, apresentando frequências entre 0 e 12 kHz e intensidades entre 90 e 220 dB.

Figura 1 – Exemplo de Espectrograma.

Fonte: Zalkow (2013).

2.2.2 Transformada de Hilbert

Uma função complexa $z(x) = f(x) + ig(x)$, onde a parte real é uma determinada função analisada e a imaginária é a Transformada de Hilbert (Equação 2.3), é chamada de Função Analítica (POULARIKAS, 1999). A transformada não decompõe a função nem modifica o seu domínio e sua inversa pode ser obtida pela Equação 2.4.

$$g(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(u)}{x-u} du \quad (2.3)$$

$$f(x) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(u)}{x-u} du \quad (2.4)$$

A Função Analítica possui algumas características importantes como: é infinitamente diferenciável, sua amostragem pode ser modificada sem alterar $f(x)$ e sua modulação e demodulação são facilitadas (POULARIKAS, 1999). Sendo assim, um sinal pode ser manipulado a partir do seu envelope de amplitudes definido como $|z(x)|$. A Figura 2 apresenta a relação entre a função, sua transformada e o envelope gerado.

¹ Canção alemã gravada na década de 1920 pelo *Manhattan Quartett*.

Figura 2 – Envelope e Transformada e Hilbert.

Fonte: MacCallum (2015).

2.2.3 Densidade Espectral

A Densidade Espectral - *Power Spectral Density* (PSD) estima a distribuição de energia como uma função da frequência, baseada em um conjunto de amostras de um determinado processo (MADISETTI, 2009). Com tal estimativa é possível obter informações que possam ser utilizadas para refinar um modelo, para predição ou filtragem de um sinal. Segundo Madisetti (2009) a energia total de um sinal $f(t)$ pode ser encontrada a partir de uma integração da Transformada de Fourier desse sinal (Equação 2.5).

$$\varepsilon = \int_0^1 |\mathfrak{F}\{f\}|^2 df \quad (2.5)$$

Essa medida pode ser estimada através do método proposto por Welch (1967). Como resultado, pode ser obtido um vetor de valores cujas posições são as respectivas frequências, ou uma matriz que relaciona frequência (linhas), tempo (colunas) e os valores obtidos (células).

2.2.4 Nível de Pressão Sonora

O Nível de Pressão Sonora - *Sound Pressure Level* (SPL), descrito na Equação 2.6, é a representação em decibel do *Mean Square Pressure* (MSP) tendo como base uma pressão de referência ($p_{ref} = 1\mu Pa$ em meios aquáticos e $p_{ref} = 20\mu Pa$ em meios terrestres) (SANCHE-ZÁGENDRIZ; PADOVESE, 2015). Na Equação 2.7 os valores de p referem-se à pressão sonora em um determinado instante n que pode ser substituída pelos valores do PSD como verificado

nos artigos citados anteriormente.

$$SPL = 10 \log_{10} \frac{MSP}{p_{ref}^2} \quad (2.6)$$

$$MSP = \frac{1}{N} \sum_{n=1}^N p(n)^2 \quad (2.7)$$

2.2.5 Variação do Sinal

O nível de variação de um sinal pode ser medido por *Roughness* (RAMSAY, 2006) ou por *Rugosity* (MEZQUIDA; MARTÍNEZ, 2009), definidas, respectivamente, pelas Equações 2.8 e 2.9. Valores baixos dessas medidas podem indicar um sinal puro com pouca variação, enquanto que valores altos podem destacar sinais com algum tipo de ruído.

$$Roughness = \sum \left(\frac{d^2 f(x)}{dx^2} \right)^2 \quad (2.8)$$

$$Rugosity = \sqrt{\frac{\sum_{i=1}^n (x_{i+1} - x_i)^2}{n}} \quad (2.9)$$

2.2.6 Coeficientes Mel-Cepstrais

Os Coeficientes Mel-Cepstrais - *Mel-frequency Cepstrum Coefficients* (MFCC) são um conjunto de características normalmente utilizadas para reconhecimento de fala (LUTTER, 2014). Com a finalidade de gerar características para o áudio, o MFCC procura modelar tanto a produção da fala quanto sua percepção. Para isso, um dos principais passos é utilizar a escala Mel, que procura modelar a audição humana. Esse mapeamento ocorre de tal maneira, que entre 0 e 1kHz a função possua comportamento aproximadamente linear e acima disso logarítmico, da mesma maneira que a audição humana funciona (Aalborg University, 2004). Uma formulação comumente utilizada para converter entre Hertz e Mel é

$$F_{mel} = \frac{1000}{\log_{10}(2)} \times \log_{10} \left(1 + \frac{F_{hertz}}{1000} \right). \quad (2.10)$$

Segundo Lutter (2014), os coeficientes do MFCC podem ser gerados seguindo os seguintes passos:

1. Aplicação da Transformada de Fourier sobre o sinal e geração do *Power Spectrum* (cf. Seção 2.2.3);
2. Mapeamento do *Power Spectrum* para a escala Mel através da Equação 2.10;

3. Aplicação do logaritmo nos valores de cada uma das frequências Mel geradas anteriormente;
4. Aplicação da Transformada Discreta de Cosseno - Discrete Cosine Transform (DCT);
5. Tomar a n termos da DCT como os coeficientes MFCC.

Existem variações desse processo, como utilizar primeira e segunda derivada dos valores retornados pela DCT, ou uma média da normalização dos valores etc.

2.3 Índices Acústicos

Os Índices Acústicos são funções matemáticas que sumarizam aspectos de um áudio, como elementos estruturais e distribuição de energia acústica e de informação (TOWSEY *et al.*, 2014a). Muitos deles foram desenvolvidos com a finalidade de quantificar riqueza, uniformidade, regularidade, abundância de espécies etc. (SUEUR *et al.*, 2014).

2.3.1 Índice de Shannon

Popular na literatura sobre ecologia para descrever a abundância de espécies em uma determinada área, o Índice de Shannon (H') possui fundamentação na Teoria da Informação, sendo proposto inicialmente (Equação 2.11) para quantificar a entropia de *strings* em textos (SHANNON, 1948). O H' apresenta bons resultados em locais onde a atividade animal é alta e o ruído (vento, chuva, atividade humana) é baixo (SUEUR *et al.*, 2014).

$$H' = - \sum_{i=1}^S p_i \log_a p_i, \quad (2.11)$$

tal que S é a quantidade de bandas de frequência utilizadas, a base a é o Número de Euler (e) e p_i é a proporção de som que a banda i possui. Por vezes a Equação 2.11 é apresentada dividida pelo seu valor máximo ($\ln S$) como na Equação 2.12.

$$H' = \frac{- \sum_{i=1}^S p_i \ln p_i}{\ln S} \quad (2.12)$$

Ao transpor a idéia de entropia utilizada com *strings* no trabalho original para a análise acústica, o resultado do índice se aproxima do valor 1 quando são grandes a quantidade de bandas de frequência utilizadas e a regularidade entre elas (valores de p_i semelhantes para a maioria das bandas i).

2.3.2 Coeficiente de Gini-Simpson

O Coeficiente de Gini-Simpson (G) proposto por Gini (1912) é utilizado para medir a desigualdade entre distribuições de frequência, como níveis de renda. Posteriormente essa medida foi adaptado por Simpson (1949) para ser utilizada em estudos biológicos sendo formulada como

$$G = 1 - \sum_{i=1}^S p_i^2 \quad (2.13)$$

Os valores de S e p_i têm o mesmo significado da Equação 2.12. Quanto mais homogêneo e regular é o áudio analisado, mais próximo do valor 1 é o valor do índice.

2.3.3 Índice de Diversidade Acústica

Para o cálculo do Índice de Diversidade Acústica - *Acoustic Diversity Index* (ADI) o espectrograma do áudio é dividido em b faixas. Sobre cada faixa é aplicado um limiar l e as frequências acima dele são retornadas. O valor do índice é obtido a partir do módulo do resultado da Equação 2.11 aplicada a essas frequências (PEKIN *et al.*, 2012) como apresentado na Equação 2.14.

$$ADI = \left| \sum_{i=1}^S p_i \ln p_i \right| \quad (2.14)$$

Esse índice mede a quantidade de som nas bandas de frequência, indicando que diferentes faixas estão ocupadas (PEKIN *et al.*, 2012). Quanto maior o valor retornado, mais bandas de frequência distintas estão ocupadas e maior quantidade de espécies emitindo algum som.

2.3.4 Índice de Uniformidade Acústica

O Índice de Uniformidade Acústica - *Acoustic Evenness Index* (AEI) executa o mesmo processo realizado no cálculo do ADI, mas em vez de aplicar *Shannon* para obter o resultado, aplica o Coeficiente de Gini (VILLANUEVA-RIVERA *et al.*, 2011). Por isso, o índice busca apresentar a homogeneidade entre as faixas de frequência.

2.3.5 Índice de Complexidade Acústica

O Índice de Complexidade Acústica - *Acoustic Complexity Index* (ACI) produz uma quantificação direta da complexidade dos sons, por meio do cálculo da variação das intensidades de frequência registradas em uma gravação (PIERETTI; FARINA; MORRI, 2011). Baseia-se na idéia de que sons biológicos possuem maior variabilidade e sons antropológicos menor variação. Para o seu cálculo, como ilustrado na Figura 3, o eixo do tempo (t) é dividido em pedaços j que

são particionados em n unidades Δt_k . O eixo da frequência ($I\Delta f_l$) é dividido em q bandas Δf_l . Assim, $j = \sum_{k=1}^n \Delta t_k$ e $\Delta f = \sum_{l=1}^q \Delta f_l$.

Figura 3 – Divisões dos eixos de tempo e frequência para cálculo do ACI.

Fonte: Pieretti, Farina e Morri (2011).

Inicialmente, partindo da Figura 3, em uma banda l de frequência, calculam-se as diferenças de intensidade entre todas as frequências adjacentes pertencentes ao conjunto de tempo j (Equações 2.15 e 2.16).

$$d_k = |I_k - I_{k+1}| \quad (2.15)$$

$$D = \sum_{k=1}^n d_k \quad (2.16)$$

Esse resultado é dividido pela soma das intensidades de frequência dentro dessa janela (l, j)

$$ACI_j = \frac{D}{\sum_{k=1}^n I_k} \quad (2.17)$$

Agora, para uma determinada banda l , obtemos o seu valor de ACI somando todos os valores obtidos com a Equação 2.17 para cada um dos instantes j dessa banda de frequência

$$ACI_{\Delta f_l} = \sum_{j=1}^m ACI_j \quad (2.18)$$

Por fim, todos os valores de cada banda de frequência são somados retornando o valor total do índice

$$ACI_{tot} = \sum_{l=1}^q ACI_{\Delta f_l} \quad (2.19)$$

Os resultados do ACI são cumulativos e quanto maior o tamanho do arquivo de áudio analisado, maior o valor desse cálculo. Essa situação pode gerar problemas na comparação do ACI com outros índices. Uma estratégia apresentada na documentação do pacote *SoundEcology* da linguagem R é dividir o valor total do índice pela quantidade de minutos do arquivo de áudio analisado.

2.3.6 Índice Bioacústico

O Índice Bioacústico - *Bioacoustic Index* (BIO), proposto por Boelman *et al.* (2007), inicialmente mapeia a relação entre frequências e intensidades do áudio (Figura 4). Depois, são selecionadas apenas as intensidades acima de um limiar pré-definido. Por fim, as áreas sob as curvas remanescentes são calculadas e somadas, resultando no valor do índice. Então, avaliando o nível sonoro e a quantidade de bandas de frequência utilizadas, os autores conseguiram mensurar a abundância de aves.

Figura 4 – Espectro bioacústico (intensidade x frequência). Valores obtidos ao amanhecer entre 7h e 9h (a) e no meio do dia entre 12h e 14h (b).

Fonte: Boelman *et al.* (2007).

2.3.7 Índice de Entropia Acústica

O resultado do Índice de Entropia Acústica - *Acoustic Entropy Index* (H) é obtido pela multiplicação da Entropia Espectral (H_f) pela Entropia Temporal (H_t). Sueur *et al.* (2008) descreve que o primeiro índice é calculado a partir da aplicação de *Shannon* sobre o espectrograma gerado pela Transformada de Fourier, buscando a concentração de energia sobre as frequências. O mesmo trabalho aborda que o segundo índice é calculado aplicando-se também *Shannon*, mas agora sobre o envelope de amplitudes criado a partir da Transformada de Hilbert, procurando visualizar a energia acústica sobre as amplitudes. O H mede a concentração de energia e quanto mais alto seu valor, existe uma maior quantidade de frequências ou maior amplitude dos sinais sendo utilizada.

2.3.8 Índice de Diferença Normalizada

A partir da proporção entre sons antropológicos e biológicos (Equação 2.20), o Índice de Diferença Normalizada - *Normalized Difference Soundscape Index* (NDSI) procura estimar o nível de distúrbio causado pelos sons antropofônicos em uma paisagem (KASTEN *et al.*, 2012).

$$NDSI = \frac{\beta - \alpha}{\beta + \alpha}, \quad (2.20)$$

tal que β e α são os valores de PSD calculados para faixas de frequência de biofonia e antropofonia, respectivamente². Esse índice varia entre -1 e 1, sendo que os valores mais baixos indicam uma forte presença de sons antropológicos, os valores mais altos de sons biológicos e valores próximos a 0 indicam presença semelhante desses dois tipos de som. Kasten *et al.* (2012) informam que em alguns casos, a depender do tipo de vocalização dos animais na gravação, os valores de NDSI também podem ser baixos mesmo existindo pouca emissão de sons antropológicos.

2.3.9 Índice de Riqueza Acústica

O Índice de Riqueza Acústica - *Acoustic Richness* (AR) retorna valores que refletem a riqueza de espécies encontradas em uma série de áudios analisados. É calculado a partir do índice M (Equação 2.21) e da Entropia Temporal H_t (DEPRAETERE *et al.*, 2012).

$$M = \text{median}(|z(x)|) \times 2^{1-\text{depth}} \quad (2.21)$$

O índice M calcula a mediana do envelope de amplitude do sinal, ponderada por uma potência de 2. O valor de *depth* corresponde a profundidade da digitalização que Depraetere *et al.* (2012) utilizam como 16 *bits*. Os valores desse índice, depois de normalizados, ficam entre 0 e 1, sendo que quanto mais alto o resultado, maior a riqueza animal encontrada.

$$\mathbf{AR} = \frac{\text{rank}(\mathbf{h}) \times \text{rank}(\mathbf{m})}{n^2} \quad (2.22)$$

Na Equação 2.22, $\mathbf{h}$ e $\mathbf{m}$ são vetores com os valores de H_t e M , respectivamente, para cada um dos arquivos de áudio analisados, e n é a quantidade de arquivos. A função *rank* é a mesma da linguagem R que retorna uma ordenação dos valores do vetor passado para ela. Assim, o resultado $\mathbf{AR}$ do índice é um vetor, dado pela multiplicação item a item dos vetores envolvidos, dividida pelo quadrado de n .

2.3.10 Outros índices e medidas acústicas

Várias outras funções são propostos na literatura com a finalidade de extrair maior grau de informação sobre as Paisagens Acústicas. Pode ser citado, por exemplo, o trabalho de Sueur

² Sons antropofônicos normalmente estão entre 1 *kHz* e 2 *kHz* e os biofônicos entre 2 *kHz* e 8 *kHz*.

et al. (2014), que além dos índices já discutidos, apresenta uma revisão de índices, como: os que medem algum parâmetro L do nível de som (ponderados por diferentes pesos de frequência e tempo), utilizados para avaliar paisagens urbanas; o *Number of Peaks* (NP), que contabiliza os picos de frequência, assumindo que quanto maior a quantidade de espécies maior a quantidade de picos de frequência; ou o Índice de Dissimilaridade Acústica - *Acoustic Dissimilarity* (D), que busca comparar comunidades de animais. Além disso, o artigo citado possui uma boa revisão e referências para os índices.

Towsey *et al.* (2014b), além de alguns dos índices já apresentados, utilizam o Índice de Cobertura Acústica - *Acoustic Cover Index* (CVR) que calcula a proporção de som existente dentro das bandas de frequência.

Já Towsey *et al.* (2014a) definem e utilizam, além dos índices de entropia, medida como: o *Background noise*, que é calculado sobre o envelope de amplitude, a fim de determinar o grau de ruído (em *dB*) da gravação; o *Signal-to-noise ratio* (SNR) que apresenta a diferença entre os máximos valores do envelope de amplitude e o *Background noise*; o *Acoustic activity* que calcula a fração de *frames* dentro de um segmento de 1 *minuto* de áudio, onde a amplitude do envelope é maior que 3 *dB* acima do nível do *Background noise*; ou o *Mid band activity* que considera a região média do espectrograma (entre 482 *Hz* e 3.5 *kHz*) onde a amplitude está acima de um limiar, que por sua vez está acima do *Background noise*.

Trabalhos que fazem análises sobre a eficácia da aplicação dos índices são apresentados no Capítulo 3.

2.4 Segmentação

A segmentação pode ser definida como o processo pelo qual uma imagem (ou áudio) é dividida em regiões ou objetos que a compõem (GONZALEZ; WOODS, 2010). A segmentação simplifica ou transforma um objeto em uma representação mais significativa ou de fácil análise. O nível de detalhes de uma segmentação depende do que está sendo tratado. O Capítulo 3 apresenta tanto trabalhos que utilizaram a segmentação do sinal de áudio quanto os que realizaram segmentação da imagem do seu espectrograma.

Tomando como exemplo imagens, a segmentação pode se basear em propriedades como descontinuidade e similaridade, segundo Gonzalez e Woods (2010). No primeiro tipo, uma imagem é dividida a partir de suas mudanças de intensidade, como as mudanças por causa das bordas. No outro tipo, as abordagens dividem uma imagem em regiões que possuem algum atributo em comum, como cor. Algumas técnicas podem ser elencadas, como os algoritmos para **detecção de bordas**, como a **limiarização** que separa a imagem em duas regiões, o processamento **morfológico** que aplica a Teoria dos Conjuntos para extrair componentes de uma imagem etc.

2.5 Representação e Descrição de Imagens

Os descritores de imagens são algoritmos que extraem vetores de características de uma imagem, podendo ser classificados com base no tipo de informação que extraem, como cor, textura, forma etc. (PENATTI *et al.*, 2009). Alguns exemplos desses descritores são apresentados nas próximas seções.

2.5.1 Descritores de Textura

Mesmo a textura sendo uma característica importante para descrever uma imagem, é difícil defini-la. Entretanto, de maneira simples, pode-se dizer que ela é um conjunto de variações de intensidade que formam padrões e repetições (PENATTI *et al.*, 2009). A seguir são apresentados alguns descritores que podem ser utilizados para representar as texturas.

2.5.1.1 Matriz de Coocorrência

A Matriz de Coocorrência - *Gray Level Co-occurrence Matrix* (GLCM) (HARALICK; SHANMUGAM; DINSTEN, 1973) é um método estatístico popular para descrição de textura. Essa técnica busca quantificar relações de intensidade entre os pixels de uma imagem, em diferentes direções. Para criar essa matriz, segundo Gonzalez e Woods (2010) pode-se inicialmente quantizar os níveis de intensidade da imagem. Por exemplo, em uma imagem de 256 níveis de cor, os níveis de 0 a 31 podem ser representados pelo valor 0; os níveis de 32 a 63, representados por 1 etc. Após essa etapa, cria-se as matrizes propriamente ditas, uma para cada direção analisada, como 0° , 45° , 90° e 135° . A Figura 5 apresenta um exemplo com uma imagem 4×4 , com 4 níveis de cor. Na matriz que representa as transições de 0° , a posição (0, 0) representa as transições na horizontal de 0 para 0 e assim por diante. Na matriz que representa transições de 90° , a posição (3, 2) representa as transições verticais de 3 para 2 etc.

Portanto, para cada matriz criada são geradas medidas como: correlação, entropia,

Figura 5 – Exemplo de Matrizes de Coocorrência.

Fonte: Elaborada pelo autor.

momento angular, contraste etc. Essas medidas são tomadas como os descritores da textura, podendo ser um conjunto de medidas para cada matriz (quantidade de medidas x 4) ou cada medida sendo a média das matrizes (ex.: o valor da entropia seria a média das entropias das 4 matrizes).

2.5.1.2 Local Binary Patterns

O *Local Binary Patterns* (LBP) (OJALA; PIETIKAINEN; MAENPAA, 2002) extrai informações referentes a variação de intensidade entre os pixels. Com essa finalidade, ele define uma janela centrada no pixel a ser analisado, com raio R e quantidade de pixels P . Essa janela é deslizada pela imagem para avaliar cada pixel e seus respectivos vizinhos. Tomando como exemplo uma janela com $R = 1$ e $P = 4$, a Figura 6 apresenta o processo realizado pelo LBP. Com a finalidade de ser invariante à iluminação, é utilizada a diferença entre os vizinhos e o pixel central. Para garantir que o processo seja invariante à escala, as diferenças são binarizadas. As diferenças maiores ou iguais a zero são consideradas 1, as demais como 0. A invariância à rotação é obtida, tomando-se o conjunto de todos os deslocamentos de bits à direita e escolhendo o menor valor dentre eles. Para melhorar a invariância à rotação, notou-se que para $R = 1$ e $P = 8$ alguns padrões ocorriam até 90% das vezes. Esses padrões, por possuírem poucas transições entre 0 e 1, são chamados de uniformes. Assim sendo, Ojala, Pietikainen e Maenpaa (2002) definiram uma medida de uniformidade. Se essa medida for menor ou igual a 2, o padrão é representado com a sua quantidade de *bits* 1, do contrário com $P + 1$. A partir dos $P + 1$ valores que a imagem é representada, é criado um histograma desses valores, que é o vetor de características de textura LBP.

Figura 6 – Exemplo do processo do descritor LBP.

Fonte: Elaborada pelo autor.

2.5.1.3 Descritor Espectral

Como mencionado na Seção 2.2.1, a Transformada de Fourier pode ser aplicada ao processamento de imagens. Uma das aplicações possíveis é a descrição de padrões de textura que podem ser vistos como áreas de alta energia no espectro (GONZALEZ; WOODS, 2010). As características podem ser interpretadas em termos de coordenadas polares dentro da imagem. Dessa maneira, para um raio r qualquer, uma descrição global pode ser obtida a partir de $S(r) = \sum_{\theta=0}^{\pi} S_{\theta}(r)$, onde S é uma função do espectro. Se o ângulo for fixado, a função se torna $S(\theta) = \sum_{r=1}^R S_r(\theta)$.

A Figura 7 apresenta uma textura, sua transformada, e: uma área circular destacada para um raio r_1 e uma área retangular para um ângulo θ_1 qualquer. Os valores de espectro são somados como apresentado anteriormente e constituem o par de valores $S(r_1)$ e $S(\theta_1)$ para as coordenadas em questão. Segundo Gonzalez e Woods (2010), variando essas coordenadas é possível obter um conjunto de características para uma imagem completa ou para uma região específica. Além disso, podem ser utilizadas medidas estatísticas para representar os valores obtidos. Pode-se, também, utilizar apenas os valores de $S(r)$ ou de $S(\theta)$ com alguma perda de informação.

Figura 7 – Transformada e Fourier de uma imagem de textura. Na segunda linha estão a transformada, e as áreas correspondentes a exemplos de coordenadas de raio e ângulo para criação dos descritores de textura.

Fonte: Adaptada de Abyoto, Wirdjosoedirdjo e Watanabe (1998).

2.5.1.4 Weber Local Descriptor

O *Weber Local Descriptor* (WLD) (CHEN *et al.*, 2010) é baseado na lei de Weber que trata sobre a percepção humana de estímulos físicos. O WLD cria um histograma que representa a relação entre excitação diferencial ξ (diferença de intensidade entre cada pixel e sua vizinhança) e a orientação θ dos padrões (baseada no gradiente) calculados para cada um dos pixels das imagens. Seguindo o esquema da Figura 8, as orientações calculadas são quantizadas em T orientações dominantes. Com os valores de excitação e orientação, é criado um histograma 2D

Figura 8 – Esquema da criação do histograma do descritor de textura WLD.

Fonte: Elaborada pelo autor.

que representa a excitação (linhas) e as orientações dominantes (colunas). Cada coluna da matriz é transformada em um histograma 1D que representa os valores de excitação para cada uma das orientações. Esses histogramas são divididos em M segmentos. Agora, uma segunda matriz é criada com esses histogramas, onde cada coluna é um dos histogramas anteriores e cada linha é um dos M segmentos criados. Todas as M linhas dessa nova matriz são concatenadas formando um novo e único histograma. Por fim, esse histograma é quantizado em S valores, gerando assim o descritor final.

2.5.2 Descritores de Cor

A cor é uma característica importante para a visão humana identificar uma imagem (PENATTI *et al.*, 2009). Logo, a cor é uma propriedade constantemente utilizada na literatura de processamento de imagens. A seguir são apresentados alguns descritores que podem ser utilizados para representar cores.

2.5.2.1 Histograma Global de Cor

O Histograma Global de Cor - *Global Color Histogram* (GCH) (SWAIN; BALLARD, 1991) é um descritor popular na literatura de Processamento de Imagens e simples de ser implementado. Inicialmente as cores são quantizadas em, por exemplo, 64 valores. A partir dessa quantização, é criado um histograma dessas cores quantizadas. Os valores desse histograma correspondem a descritor das cores da imagem.

2.5.2.2 Auto Correlograma de Cores

O Auto Correlograma de Cores - *Auto Color Correlation* (ACC) (HUANG *et al.*, 1997), procura descobrir a probabilidade de encontrar dois pixels de uma mesma cor a uma determinada distância um do outro. Como em outros descritores, inicialmente os valores de cores são quantizados em m valores (64 por exemplo). Depois dessa quantização, para cada pixel p , é calculada a probabilidade de existir um outro pixel com o mesmo valor a uma determinada distância d . São utilizados vários valores de distâncias com a finalidade de gerar um vetor de características. No artigo original são utilizadas 1, 3, 5, e 7. Por causa disso, o vetor de características possui quantidade de posições igual à multiplicação da quantidade de distâncias pela quantidade de valores da quantização.

2.5.2.3 Vetor de Coerência de Cor

O Vetor de Coerência de Cor - *Color Coherence Vector* (CCV) (PASS; ZABIH; MILLER, 1997), cria um histograma e classifica os pixels em coerentes e incoerentes de acordo com um limiar L pré-definido. Inicialmente a imagem é suavizada e quantizada em, por exemplo, 64 valores. O histograma possuirá 128 posições, sendo 64 para os valores coerentes e 64 para os incoerentes. Após essa etapa, são encontrados componentes conexos para cada uma das cores da imagem quantizada. Para uma determinar cor C , um componente conexo é avaliado. Se a quantidade de pixels desse componente for maior que o limiar L , essa quantidade é contabilizada na posição de C , na parte do histograma referente aos valores coerentes. Do contrário, quantidade menor que L , a quantidade é contabilizada como incoerente. Esse processo é repetido para cada componente encontrado para a cor C e para todas as cores quantizadas da imagem.

2.5.2.4 Classificação de Pixel Borda/Interior

O descritor Classificação de Pixel Borda/Interior - *Border/Interior Pixel Classification* (BIC) (STEHLLING; NASCIMENTO; FALCÃO, 2002), gera um histograma que caracteriza os pixels da imagem em de borda ou de interior. Para isso, a imagem é quantizada em 64 valores. O histograma final, possui 128 posições (metade para borda, metade para interior). Para cada pixel da imagem, sua vizinhança é analisada, vizinhos-8 (todos os 8 pixels ao redor do pixel analisado) ou vizinhos-4 (os pixels exatamente acima, abaixo, da esquerda e da direita). Se todos os vizinhos do pixel possuem a mesma cor que ele, esse pixel é contabilizado na região do histograma destinada a interior, do contrário na região de borda.

2.6 Autoencoder

Para compreender o conceito de *Autoencoder* é preciso iniciar pela idéia de Aprendizagem de Máquina (*Machine Learning*). Nos dias atuais, a Aprendizagem de Máquina é utilizada para melhorar ou automatizar atividades como pesquisas na internet, reconhecimento de fala,

carros autônomos etc. Ela pode ser caracterizada como um processo automático pelo qual padrões significativos são extraídos de conjuntos de dados (SHALEV-SHWARTZ; BEN-DAVID, 2014). Também, a Aprendizagem de Máquina é uma subárea da Inteligência Artificial que busca ensinar para máquinas como aprender sem a necessidade de ser explicitamente programada para isso (GULLI; PAL, 2017). Existem vários tipos de aprendizado, como supervisionado, não supervisionado, semi- supervisionado etc. Algumas técnicas de aprendizagem exemplificam esses tipos são o *k-Nearest Neighbor* (kNN), *Support Vector Machine* (SVM), Redes Neurais, para aprendizado supervisionado; *K-means*, *Density-based spatial clustering of applications with noise* (DBSCAN), para aprendizado não supervisionado; etc. Dentro das técnicas de aprendizado supervisionado, a *Deep Learning* é, nos dias atuais, largamente utilizada. Ela é um conjunto de metodologias que utiliza Redes Neurais com uma grande quantidade de camadas (dezenas ou centenas delas) (GULLI; PAL, 2017).

Sendo assim, dentro das abordagens de *Deep Learning*, um *Autoencoder* busca criar uma representação que seja o mais próxima possível dos dados de entrada (PONTI *et al.*, 2017). Na verdade o foco principal não está na saída da rede, mas sim na representação criada pelas camadas intermediárias dela, segundo Ponti *et al.* (2017). Isso ocorre porque se a rede consegue reconstruir os dados de entrada com alguma precisão, suas camadas intermediárias são capazes de extrair características significantes desses dados (GULLI; PAL, 2017). Para isso, um *Autoencoder* é formado por duas partes (Figura 9). Uma que cria uma representação compacta dos dados de entrada (*encoder*) e a outra, que a partir dessa representação, reconstrói de forma aproximada os dados originais (*decoder*) (GULLI; PAL, 2017). Ainda segundo o autor anterior, o *Autoencoder* é capaz de encontrar correlações entre os dados, se elas existirem, e aprender uma representação reduzida delas, algo similar às técnicas de redução de dimensionalidade como Análise de Componentes Principais - *Principal Component Analysis* (PCA) (PEARSON, 1901; HOTELLING, 1933). Mesmo sendo uma Rede Neural, o *Autoencoder* é uma abordagem não

Figura 9 – Ilustração de arquitetura de um Autoencoder.

Fonte: Adaptada de Dertat (2017).

supervisionada, porque não considera os rótulos dos dados na criação do modelo (PONTI *et al.*, 2017).

A modelagem de uma arquitetura de Autoencoder como exemplificada na Figura 9 passa por passos para definir, por exemplo, quais tipos de camadas serão utilizadas. Existem vários tipos de camadas como: *convolutional* que possui um conjunto de filtros (matrizes de pesos $k \times k$) que são aplicados sobre uma entrada, normalmente uma imagem. Os valores desses filtros são justamente os parâmetros aprendidos pelo Autoencoder; *pooling* é um conjunto de camadas utilizadas para reduzir o tamanho de uma entrada com base em uma determinada proporção. Uma dos tipos mais utilizados é o *maxpooling* que, para cada janela consecutiva $n \times n$ do dado original, representa essa janela através do maior valor contido nela; *upsampling* utilizada para reescalar uma entrada para uma resolução maior a uma proporção $n \times n$. Por exemplo, pode repetir cada linha da entrada n vezes e as colunas n vezes; *flatten* utilizada para “nivelar” os dados, como, por exemplo, transformar uma matriz de entrada em um vetor; etc.

Outra decisão a ser tomada, diz respeito às funções de ativação que serão utilizadas em cada camada. Essas funções têm a propriedade de definir um mapeamento para os valores dos filtros. Algumas funções utilizadas são a função sigmoide, a função SeLU (Equação 2.23) etc.

$$selu(x) = \lambda \begin{cases} x, & \text{se } x > 0 \\ \alpha(e^x - 1), & \text{se } x \leq 0 \end{cases} \quad (2.23)$$

tal que α e λ são constantes pré-definidas³. Também é necessária a definição de uma função de perda ou função de custo para medir o quão bom ou ruim é o modelo que está sendo gerado. Essa função de perda deve ser minimizada através de alguma técnica de otimização. Existem várias funções de perda que podem ser utilizadas como *Mean Squared Error*, *Mean Squared Logarithmic Error* etc. E para otimização pode ser utilizado o método *Stochastic Gradient Descent*, dentre outros.

Outras questões a serem ponderadas são: a quantidade de amostras nos conjuntos de treinamento (utilizado para gerar o modelo), validação (utilizado para estimar o quão bom é o modelo que está sendo gerado) e teste (utilizado para examinar o modelo); valor do *batch* que representa a quantidade de amostras apresentadas à rede por ciclo; e a quantidade de épocas de treinamento, que especifica a quantidade de vezes que o modelo processa toda a base de treinamento.

2.7 Técnicas de Visualização de Dados

A área de Visualização da Informação - *Information Visualization* (InfoVis) procura aplicar representações visuais para compreender valores de atributos e relações entre os dados,

³ Na biblioteca Keras do Python são definidas como $\alpha \approx 1.67$ e $\lambda \approx 1.05$.

com o objetivo de obter conhecimento sobre eles (CHI, 2002; TELEA, 2014). Os tipos de dados utilizados não possuem características físicas como formato e cor, assim sua representação é abstrata e as técnicas de visualização buscam criar uma representação visual desses dados para os usuários (TELEA, 2014). Tais técnicas estão focadas na representação de dados de alta dimensionalidade (grande quantidade de atributos).

As técnicas de visualização para dados com múltiplas dimensões podem ser divididas, segundo Ward, Grinstein e Keim (2015), em três grupos: baseadas em *i*) pontos, *ii*) linhas e *iii*) regiões. Para o primeiro grupo, os registros são mapeados de um espaço n -dimensional para um espaço visual 2-dimensional (2D) ou 3-dimensional (3D) e representados nas respectivas posições por alguma entidade gráfica, como ponto, triângulo, retângulo, entre outras. Dessa maneira, conseguem comunicar o valor de alguma dimensão de um objeto ou o próprio objeto. Como exemplos dessas técnicas temos o Gráfico de Dispersão (*Scatterplot*) e as Projeções Multidimensionais (Seção 2.7.1). No segundo grupo, os pontos também representam um objeto em particular ou uma de suas dimensões e, além disso, podem ser conectados por linhas ou curvas. Assim, pode-se verificar as relações entre os dados e identificar padrões e tendências. Gráficos de Linha e Coordenadas Paralelas (Seção 2.7.4) são alguns exemplos desse grupo. O último grupo utiliza características de polígonos, como forma, cor, tamanho etc. para representar valores dos atributos. Gráficos de Barras, Histogramas e Mapas de Calor (Seção 2.7.5) são alguns exemplos dessas visualizações.

Antes da aplicação das técnicas de visualização, em algumas situações é necessária alguma etapa de pré-processamento. Nessa etapa os dados podem ser resumidos com medidas estatísticas, dados redundantes podem ser removidos, normalização ou padronização dos valores pode ser realizada etc. As técnicas de normalização consistem em modificar a faixa de valores das características para poder compará-las umas com as outras. Uma normalização comumente utilizada é a *Min-max*, que mapeia todos os valores de um atributo para o intervalo $[0, 1]$, aplicando $y_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$, sendo x_i cada valor do atributo e x_{min} e x_{max} os valores mínimo e máximo do atributo que está sendo normalizado.

2.7.1 Projeções Multidimensionais

Além de reduzir as dimensões dos dados de n para duas ou três dimensões e representá-los por alguma entidade gráfica como pontos, as técnicas de Projeção Multidimensional têm por objetivo manter as relações de vizinhança existentes entre os dados (Figura 10). Mesmo assim, nem sempre esse objetivo é alcançado, devido à quantidade de atributos do espaço original (WARD; GRINSTEIN; KEIM, 2015).

Com o propósito de posicionar os pontos em espaços visuais preservando propriedades relevantes do espaço original de atributos, várias técnicas de projeção foram propostas, como: *Multidimensional Scaling* (MDS) (KRUSKAL; WISH, 1978), *Force scheme* (TEJADA; MINGHIM; NONATO, 2003), *t-Stochastic Neighbor Embedding* (t-SNE) (MAATEN; HINTON,

Figura 10 – Técnica de projeção t-SNE aplicada a uma base de dados de imagens de dígitos manuscritos. Dígitos de 0 a 9.

Fonte: Maaten e Hinton (2008).

2008), *Least Square Projection* (LSP) (PAULOVICH *et al.*, 2008), *Part-Linear Multidimensional Projection* (PLMP) (PAULOVICH; SILVA; NONATO, 2010), *Local Affine Multidimensional Projection* (LAMP) (JOIA *et al.*, 2011) e assim por diante. Técnicas de redução de dimensionalidade como PCA, também podem ser utilizadas com a finalidade de projeção. Por causa de suas particularidades, a eficácia e eficiência dessas técnicas depende das características do conjunto de dados que está sendo projetado.

2.7.2 Avaliação de Projeções Multidimensionais

Para avaliar os resultados de uma projeção podem ser utilizadas medidas como o *Stress* que é formulada como a Equação 2.24. Essa medida mede a perda de informação durante a projeção (KRUSKAL, 1964). Possui valores no intervalo [0, 1], onde 0 é a menor perda e 1 a maior.

$$stress = \sqrt{\frac{\sum_{i < j} (\delta(\mathbf{y}_i, \mathbf{y}_j) - d(\mathbf{x}_i, \mathbf{x}_j))^2}{\sum_{i < j} \delta(\mathbf{y}_i, \mathbf{y}_j)^2}}, \text{ sendo } \begin{cases} \mathbf{x} \text{ o vetor no espaço original,} \\ \mathbf{y} \text{ o vetor no espaço projetado,} \\ d \text{ a função de similaridade do espaço original,} \\ \delta \text{ a função de similaridade do espaço projetado.} \end{cases} \quad (2.24)$$

Também pode ser aplicado o Coeficiente de Silhueta, que normalmente é empregado

para avaliar agrupamentos e mede a coesão e a separação entre grupos, podendo variar entre $[-1, 1]$ (TAN; STEINBACH; KUMAR, 2005). Essa medida pode ser obtida para cada item da base de dados com a Equação 2.28, e pode também representar toda a base através da média dos valores de silhueta das instâncias. O valor de silhueta é obtido pelo relacionando dos valores de coesão (Equações 2.25) e de separação (Equações 2.26 e 2.27) referentes a cada item dos dados.

$$a(\mathbf{x}_i) = \frac{1}{N_A - 1} \sum_{\substack{i \neq j \\ x_j \in A}} d(\mathbf{x}_i, \mathbf{x}_j), \text{ sendo } \begin{cases} A \text{ o grupo de } \mathbf{x}_i, \\ N_A \text{ a quantidade de itens do grupo } A, \\ d \text{ a função de similaridade.} \end{cases} \quad (2.25)$$

$$D(\mathbf{x}_i, C) = \frac{1}{N_C} \sum_{\substack{x_j \in C \\ C \neq A}} d(\mathbf{x}_i, \mathbf{x}_j) \quad (2.26)$$

$$b(\mathbf{x}_i) = \min_{\forall C \neq A} \{D(\mathbf{x}_i, C)\} \quad (2.27)$$

$$s(\mathbf{x}_i) = \frac{b(\mathbf{x}_i) - a(\mathbf{x}_i)}{\max\{a(\mathbf{x}_i), b(\mathbf{x}_i)\}} \quad (2.28)$$

Quanto mais próximos estão os grupos, o valor médio da silhueta tende para 0; quanto mais separados o valor tende para 1; e quanto mais os grupos estão sobrepostos para -1. A quantidade de grupos influencia o valor da média da silhueta. Quanto maior a quantidade de grupos, mesmo que eles estejam bem separados, os valores de silhueta ficam próximos de 0. Os resultados da silhueta podem ser classificados quanto a estrutura de agrupamento em: nenhuma estrutura $[-1; 0,26[$, estrutura fraca $[0,26; 0,51[$, estrutura razoável $[0,51; 0,71[$ e estrutura forte $[0,71; 1]$ (KAUFMAN; ROUSSEEUW, 2009).

Outras medidas utilizadas para avaliar projeções são o *Neighborhood Hit* e o *Neighborhood Preservation*, que medem, respectivamente, a proporção de vizinhos de um ponto no espaço projetado que pertencem a mesma classe (PAULOVICH *et al.*, 2008), e a proporção de vizinhos no espaço original que continuam vizinhos no espaço projetado (PAULOVICH; MINGHIM, 2008). Ambos variam entre $[0, 1]$ e quanto maior o valor, melhor a projeção.

Além dessas avaliações, também é possível verificar em um gráfico a relação entre as distâncias no espaço original e no espaço projetado, além da comparação de histogramas das distâncias nos dois espaços.

2.7.3 Hierarchical Point Placement Strategy

Como exemplo mais específico de projeção, a *Hierarchical Point Placement Strategy* (HiPP) busca manter características de agrupamento e separação dos dados (PAULOVICH;

MINGHIM, 2008). Segundo os autores, a estrutura gerada por essa projeção garante capacidade de exploração em vários níveis de detalhe, o que melhora a capacidade de análise e possibilita a utilização de grandes conjuntos de dados.

Para alcançar esses objetivos, a HiPP emprega basicamente três passos: *a)* criação de uma hierarquia dos dados, *b)* projeção dessa hierarquia e *c)* espalhamento e remoção parcial ou total de sobreposição dos dados. Os autores destacam que a complexidade computacional da HiPP é $O(n\sqrt{n})$.

Inicialmente uma árvore com apenas um nó e todas as instâncias de dados como seus filhos é criada. Posteriormente, esse nível de filhos é dividido em k grupos utilizando o algoritmo de agrupamento *bisecting k-means*. Para cada um dos grupos gerados é criado um nó filho do nó anterior. Cada um dos itens dos grupos é associado ao respectivo nó do grupo. Esse mesmo processo de particionamento é aplicado a cada um dos nós, recursivamente, até que uma quantidade mínima de filhos por nó seja alcançada. A quantidade de grupos que um nó deve ser dividido é estipulada através da regra da raiz, ou seja, $k = \sqrt{n}$, sendo n a quantidade de filhos do nó que está sendo dividido. Enquanto que a quantidade mínima de itens é estipulada pela mesma regra, tomando n como a quantidade total de instâncias de dados da árvore. O segundo passo é executar uma projeção de cada um dos nós da árvore utilizando a projeção LSP. Por fim, um algoritmo de espalhamento é aplicado para remover sobreposições entre nós. Os nós são movidos na direção do vetor que eles formam, mas em sentidos contrários. A Figura 11 apresenta um exemplo do particionamento, projeção e interação com os nós da HiPP. Inicialmente apenas os nós mais externos são apresentados e conforme o usuário interage com eles os demais níveis dentro do nó explorado são apresentados até que as folhas da árvore sejam alcançadas. Para exploração textual, os autores da HiPP implementaram uma abordagem de extração de tópicos dos textos baseada na covariância dos termos. Os grupos são coloridos a partir dos seus tópicos e esses tópicos também são apresentados para cada grupo.

Figura 11 – Exemplo de particionamento, projeção e interação com a HiPP.

Fonte: Adaptada de Paulovich e Minghim (2008).

2.7.4 Coordenadas Paralelas

Nas visualizações com Coordenadas Paralelas, os eixos (coordenadas, características, atributos, dimensões etc.) são paralelos em vez de ortogonais, possuindo um espaçamento uniforme entre eles (Figura 12). Uma instância é representada como uma curva que corta cada eixo em uma posição proporcional ao seu valor para aquela dimensão (WARD; GRINSTEIN; KEIM, 2015; TELEA, 2014). É possível perceber relações entre as dimensões e conjuntos de linhas que possuem comportamentos parecidos. Curvas com mesmo comportamento podem ser interpretados como grupos de itens similares. O usuário pode mover os eixos para ajudar na análise. Mesmo assim, a área de visualização disponível limita a quantidade de atributos que podem ser utilizados e a representação por linhas rapidamente causa confusão visual.

Figura 12 – Exemplo de Coordenadas Paralelas utilizando a base de dados Íris.

Fonte: Adaptada de Xie (2010).

2.7.5 Mapa de Calor

O Mapa de Calor (*Heatmap*) é normalmente utilizado para apresentar valores de tabelas como cores (Figura 13). Assim, os valores das células são mapeados em um espaço normalizado de cores, sendo apresentado como uma região colorida. Com essa visualização é possível verificar a variação e tendências dos valores. Pode-se utilizar diferentes mapas de cores e os usuários podem expandir ou comprimir esses intervalos para enfatizar ou atenuar alguma faixa de valores (WARD; GRINSTEIN; KEIM, 2015). Essa representação de cores é limitada à percepção humana, o que pode dificultar a visualização de faixas ou quantidades de valores grandes.

2.8 Considerações sobre os conceitos apresentados

Neste capítulo, a escolha de conceitos e técnicas apresentados se deu relativa à sua relevância para a pesquisa desenvolvida. As medidas acústicas são representações usuais do espaço de Paisagens Acústicas que são levadas em consideração em uma grande variedade de

Figura 13 – Exemplo de Mapa de calor. Linhas representado datas e colunas horários.

Fonte: Elaborada pelo autor.

aplicações. A segmentação é uma ferramenta necessária para a localização de eventos. As formas de descrição de imagens e a utilização de técnicas de Aprendizagem de Máquina são alternativas de representação das paisagens. Enquanto isso, os métodos de visualização são necessários para apoiar a intervenção do usuário na análise de ambientes através das paisagens. Em especial, a característica exploratória de muitas dessas análises requer métodos visuais para identificação de padrões de interesse.

TRABALHOS RELACIONADOS

Neste capítulo é apresentado um levantamento de trabalhos que fazem análise de Paisagens Acústicas. Alguns deles apresentam resultados obtidos com os Índices Acústicos. Outros artigos apresentam problemas relacionados à sua utilização na análise de paisagens. Também são apresentados artigos que efetuam a segmentação do áudio ou da imagem do espectrograma com a mesma finalidade. Por fim, são identificados trabalhos de indexação e visualização de dados acústicos. E algumas considerações sobre os artigos citados são apresentadas ao final do capítulo.

3.1 Trabalhos que utilizam Índices Acústicos

Algumas questões podem ser levantadas sobre a utilização dos índices para extração de informação sobre as Paisagens Acústicas. Primeiramente, alguns índices, como já visto, utilizam a ideia de entropia para determinar a diversidade de espécies. Mas entropia, na verdade, apresenta a incerteza de identificar uma espécie em uma amostra, e não a quantidade total de espécies nessa amostra (JOST, 2006). O artigo citado apresenta como a diversidade deveria realmente ser calculada a partir desses índices. Um dos exemplos apresentados é baseado no Índice de Shannon e calculado como $Div = \exp(H')$. Dessa maneira, segundo Jost (2006), o resultado possui um conjunto uniforme de propriedades matemáticas que melhor representam o conceito de diversidade.

Também Gasc *et al.* (2013) procuraram avaliar teoricamente se Índices Acústicos podem ser utilizados para avaliar a diversidade filogenética (dispersão das espécies em uma árvore de filogenia) e a diversidade funcional (medida por características fisiológicas, morfológicas, ecológicas, comportamentais, entre outras). Utilizando a correlação entre os índices, puderam concluir que eles conseguem refletir as diversidades filogenéticas e funcionais de uma determinada paisagem. Mesmo assim, afirmam que são necessários mais testes em cenários reais para complementar essas conclusões.

Enquanto isso, [Lellouch et al. \(2014\)](#) procuraram comparar índices de dissimilaridade, a fim de medir mudanças temporais para compreender a dinâmica de comunidades de animais. Utilizaram índices já abordados na literatura (*Correlation-based dissimilarity*, *Symmetric Kullback-Leibler divergence*, *The integral of pointwise difference* e *Kolmogorov-Smirnov distance*) e criaram o *Cumulative frequency dissimilarity* (D_{cf}) para efetuar seus experimentos. Com base em análises estatísticas como a correlação entre os índices, ponderaram que tais índices não podem ser utilizados para representar a diversidade. Isso ocorre porque eles não proveem detalhes sobre as mudanças globais ocorridas na composição das espécies, mas podem ser úteis para medir a dinâmica acústica dos ambientes.

[Parks, Miksis-Olds e Denes \(2014\)](#) procuraram utilizar o índice H para caracterizar a riqueza de espécies em ambientes marinhos. Devido à interferência de ruído natural e antropológico, esse índice não ajudou nas avaliações da biodiversidade. Seus resultados melhoraram depois da remoção de ruídos, aumentando a capacidade de avaliação da biodiversidade em ambientes marinhos.

Trabalhos como o de [Fuller et al. \(2015\)](#) buscam relacionar os índices discutidos em seções anteriores e encontrar significados mais precisos para seus valores. Como exemplo, o trabalho citado apresenta que NDSI e H estão mais ligados a características da paisagem, condições ecológicas e riqueza de espécies de pássaros. Além dessas análises, os autores propõem que sejam realizadas mais análises em regiões de condições climáticas distintas para confirmar a influência do ambiente nos padrões de Paisagem Acústica.

Enquanto isso, a pesquisa de [Gasc et al. \(2015\)](#) procurou descrever situações diversas para verificar a verdadeira capacidade dos índices (ACI, H, H_t , H_f , M, AR e NP) em descrever a diversidade e a riqueza de espécies em paisagens terrestres. Os autores simularam conjuntos de pássaros, a partir de sons naturais, o que proporcionou melhor controle sobre os parâmetros referentes às fontes sonoras. Com isso, puderam constatar quais dessas medidas acústicas são melhores para descrever a biodiversidade e quais são menos ou mais suscetíveis a variações como ruído, adição ou remoção de novos sons, aumento ou diminuição da duração dos sons dentro dos áudios etc. Como exemplos dos resultados, pode ser destacado que: o AR e o NP apresentaram menor sensibilidade ao ruído de fundo. Para os demais índices, uma remoção preliminar de ruído ou a aplicação de um limiar sobre as amplitudes do áudio podem melhorar os resultados; o índice H_t pôde identificar melhor a diversidade animal, mesmo removendo, adicionando e modificando sons existentes nos áudios. Portanto, baseado na análise apresentada, é necessário cuidado na interpretação dos índices de diversidade, porque eles podem ser influenciados por várias características sonoras além da diversidade e regularidade dos sons biológicos existentes nos áudios ([GASC et al., 2015](#)).

Outro ponto a ser abordado diz respeito à real eficácia dos índices como ferramenta de monitoramento ecológico. Como os índices trabalham no domínio do tempo ou no domínio da frequência, eles podem não refletir completamente a informação estrutural da paisagem

(ELDRIDGE *et al.*, 2016). Por exemplo, um aumento dos valores de energia apresentado por um índice como o H pode estar ligado a grande intensidade e frequência do canto de certos indivíduos, mas pode simplesmente acontecer porque um indivíduo está próximo ao microfone. Além disso, Eldridge *et al.* (2016) descrevem que os índices são influenciados por altas taxas de ruídos temporários ou permanentes. Eles procuraram apresentar uma abordagem que represente melhor a dinâmica tempo-frequência do som, sendo menos suscetível, por exemplo, às variações de distância dos sensores. Mesmo que os índices sejam utilizados em trabalhos de análise de áudio, ainda não existe uma compreensão clara de quais aspectos da biodiversidade eles podem representar (ELDRIDGE *et al.*, 2016).

Por fim, existe uma grande aplicação dos índices na medição da diversidade de espécies. Em áreas tropicais os nichos acústicos das espécies são estreitos e a complexidade do som está fortemente ligada com a diversidade animal (KRAUSE, 2016). Entretanto, em áreas temperadas, isso deve ser encarado com cuidado porque espécies individuais frequentemente estão alocadas em uma vasta região do espectrograma, segundo Krause (2016). Assim sendo, é relevante interpretar os índices de forma adequada ao ambiente que eles estão sendo aplicados.

O Quadro 1 apresenta um resumo de artigos e os índices que eles utilizam. São apresentados os índices mais utilizados na literatura que foi lida até o momento.

3.2 Análises de Paisagens Acústicas que utilizam outras Medidas Acústicas

Mesmo com a grande quantidade e utilização de Índices Acústicos, existem trabalhos que não os utilizam em suas análises, provavelmente por causa de algumas de suas limitações. Freeman *et al.* (2014) utilizam o espectrograma, os níveis de densidade de pressão espectral¹ e imagens obtidas por câmeras remotas para analisar e monitorar a conexão entre sistemas de recifes de coral. Eles conseguiram identificar que no período diurno os padrões de áudio são mais claros. Da mesma maneira identificaram que o aumento da atividade dos invertebrados durante a noite aumenta os níveis de densidade de pressão espectral. Por fim, perceberam que os crustáceos parecem ser o grupo dominante nas gravações das regiões estudadas.

Um estudo de Sánchez-Gendriz e Padovese (2016), realizado no Parque Marinho Estadual Laje de Santos e no Parque Estadual Xixová-Japuú, ambos no Brasil, utilizou algumas medidas como PSD, MSP e SPL. Igualmente utilizaram uma representação do espectrograma baseada no PSD, além de representações estatísticas como percentis e *boxplot*. Com essas ferramentas de análise, perceberam que a Paisagem Acústica das áreas é dominada por eventos biológicos, principalmente sons de peixes. Também constataram que a atividade no Porto de Santos não afetou, pelo menos no período analisado, as paisagens das áreas em questão.

¹ Obtido a partir do escalonamento da voltagem para a pressão de entrada.

Outros trabalhos utilizam abordagens baseadas em imagens e seu processamento. Algumas referências de trabalhos que utilizam esse tipo de técnica estão na Seção 3.4.

Quadro 1 – Lista de artigos que trabalham com Índices Acústicos.

Artigo	Objetivo	Índice
Boelman <i>et al.</i> (2007)	Detectar quanto a invasão de plantas em uma região do Havaí afeta a abundância e a composição da comunidade de aves da região.	BIO
Sueur <i>et al.</i> (2008)	Propor novos conceitos e métodos para descrever a biodiversidade terrestre .	H
Pieretti, Farina e Morri (2011)	Avaliar a adequação do índice para identificar a variabilidade de sons bióticos terrestres .	ACI
Villanueva-Rivera <i>et al.</i> (2011)	Apresentar maneiras de analisar arquivos de áudio terrestre utilizando análise espectral.	AEI
Depraetere <i>et al.</i> (2012)	Testar a validade do índice para medição da biodiversidade terrestre .	AR
Kasten <i>et al.</i> (2012)	Desenvolver sistema que automatize o processo para catalogar dados de sensores acústicos terrestres .	NDSI
Pekin <i>et al.</i> (2012)	Determinar a relação entre diversidade acústica e medidas da estrutura de florestas .	ADI
Parks, Miksis-Olds e Denes (2014)	Avaliar a aplicação do índice em ambientes marinhos .	H
Towsey <i>et al.</i> (2014a)	Analisar o problema de estimar a riqueza de espécies de aves .	ACI, H, NDSI, AR
Fuller <i>et al.</i> (2015)	Verificar se os índices retornam resultados que possam ser comparados em diversos níveis de fragmentação do ambiente terrestre .	AEI, ACI, BIO, H, NDSI
Gasc <i>et al.</i> (2015)	Prover recomendações sobre aplicação dos índices na avaliação da diversidade de espécies terrestres .	Hf, Ht, H, ACI, M, AR, NP
Eldridge <i>et al.</i> (2016)	Sugerir que os índices são limitados quanto a sua habilidade de retratar as características estruturais do ambiente terrestre .	ADI, AEI, ACI, BIO, H, NDSI

Fonte: Elaborada pelo autor.

3.3 Trabalhos que segmentam o Áudio

Existem abordagens que podem ser utilizadas para segmentar áudio como a proposta por [Wichern et al. \(2007\)](#), que criaram um *framework* Bayesiano para a segmentação do áudio. Essa segmentação utilizou um conjunto de características que mapeia alguma especificidade do som como: Sonoridade (*Loudness*) e Esparsidade Temporal (*Temporal Sparsity*) baseados no cálculo do Valor Quadrático Médio - *Root Mean Square* (RMS) da janela dos dados em decibéis; *Spectral Centroid* e *Harmonicity* baseados em cálculos sobre a Transformada de Fourier de Tempo-Curto - *Short-Time Fourier Transform* (STFT); e o *Transient Index* baseado no MFCC. O *framework* apresentado consegue descrever rápidas modificações no conjunto de características utilizado. Segundo os autores, os testes descreveram a utilidade dessa abordagem probabilística em extrair eventos sonoros dos ambientes. Além do mais, eles propuseram a utilização do algoritmo *Expectation Maximization* (EM) para automatizar o aprendizado dos parâmetros que especificam o modelo proposto.

Uma abordagem apresentada por [Plinge, Grzeszick e Fink \(2014\)](#), visa classificar eventos acústicos baseado na técnica de *Bag-of-Features*. O conjunto de características utilizado é formado pelo MFCC e o *Gammatone-frequency Cepstrum Coefficients* (GFCC), modificação do MFCC que utiliza uma Distribuição Gama. Esse conjunto é quantizado através de uma *Gaussian Mixture Model* (GMM) e posteriormente classificado por um *Multinomial Maximum Likelihood Classifier*, que busca maximizar a probabilidade de um conjunto de características pertencer a uma determinada classe. Os testes apresentados pelos autores mostram combinações das características com essa técnica de classificação, com o SVM e com kNN. Segundo os resultados e avaliações relatados no artigo, a técnica apresentada mostra-se competitiva quando aplicada a bases de dados com diferentes graus de dificuldade.

No artigo de [McLoughlin et al. \(2015\)](#), foi avaliado o desempenho de um *framework* de classificação de áudio ruidoso. Os autores utilizaram dois tipos de características: *Spectrogram Image Features* (SIF), que extraem distribuições estatísticas de blocos da imagem do espectrograma do áudio; e *Stabilized Auditory Image* (SAI), que procura modelar os efeitos sonoros na fisiologia do ouvido humano. Os autores utilizaram o SVM e Rede Neural Profunda - *Deep Neural Network* (DNN). Segundo eles, a partir das análises de desempenho apresentadas, a associação daqueles conjuntos de características com esses classificadores mostrou obter bons resultados mesmo em sons com ruído. Os melhores resultados foram obtidos pela Rede Neural.

O trabalho de [Thorogood, Fan e Pasquier \(2015\)](#) apresentou um método para segmentação de áudio utilizando o SVM. Inicialmente eles criaram uma base para validar a classificação a partir de áudios rotulados por humanos. Um conjunto de características baseado em [Peeters \(2004\)](#) foi extraído e sobre ele aplicado um método de redução de dimensionalidades disponível no Weka².

² Programa para Mineração de Dados e Aprendizagem de Máquina, disponível no endereço <<http://www.cs.waikato.ac.nz/ml/weka/>>.

Também, os autores testaram a segmentação com vários tamanhos de janela temporal (tamanho de partes dos arquivos de áudio analisadas). O conjunto de características associado com a variação dessas janelas foi utilizado para treinar o SVM, que obteve um bom funcionamento para todos os tamanhos de janelas utilizados no artigo.

3.4 Trabalhos que segmentam a Imagem do Espectrograma

Para segmentar imagens dos espectrogramas do áudio, existem abordagens como as propostas nos trabalhos de [Neal et al. \(2011\)](#), [Briggs et al. \(2012\)](#). Eles aplicaram uma fase de pré-processamento, onde são aplicados filtros para remoção de ruído do espectrograma. Vetores de características que representam partes do espectrograma foram gerados e essas características são utilizadas para treinar um classificador *Random Forest Classifier* que particiona a imagem em sons de pássaros e sons de fundo. Os dois trabalhos mostraram acurácia de mais de 90% nessa classificação.

O trabalho de [Dong et al. \(2015\)](#) apresentou um processo que define um vetor de características reduzido com a finalidade de agilizar o processo de busca e recuperação de áudio. Após uma etapa de remoção de ruídos, eles detectaram picos nas imagens dos espectrogramas. Esses picos representam os componentes (sílabas) do canto dos pássaros. Para isso eles aplicaram máscaras que detectam as direções dos picos (0°, 45°, 90° e 135°). Os picos foram marcados com cores que representam suas direções. Uma lista de eventos foi criada analisando-se os picos de cada direção (cor) separadamente. Eventos com áreas abaixo de um certo limiar definido pelos autores são desconsiderados. A partir dos eventos segmentados o processo cria e reduz o conjunto de características necessário para uma consulta por similaridade nos arquivos de áudio. Segundo os autores os resultados demonstram que o conjunto de características gerado alcança melhores resultados que os métodos propostos até a publicação do artigo e que as dimensões desse conjunto de características é bastante reduzida.

No artigo de [Xie et al. \(2015\)](#) foi utilizada uma abordagem de segmentação que primeiro aplica um filtro de *Wiener*³ para remover granulação (chuviscos) do fundo da imagem, uma remoção de ruídos sobre cada faixa de frequências e um filtro Gaussiano para suavização. No próximo passo é realizada uma binarização do espectrograma utilizando o método de *Otsu*⁴. Depois disso, operações morfológicas e de detecção de eventos são aplicadas na imagem. Por fim, dois limiares são aplicados para remover pequenos eventos e eventos unidos incorretamente. Com base nessa segmentação os autores extraíram um conjunto de características, detectaram e extraíram características dos picos e as utilizaram para classificação com o SVM. Os autores

³ Filtro utilizado para remoção de ruídos de imagem.

⁴ Algoritmo de limiarização de imagens em tons de cinza, utilizado para separar o fundo da frente da imagem.

alcançaram uma acurácia de 80%, segundo os resultados apresentados.

3.5 Indexação de Dados Acústicos

Algumas abordagens de indexação utilizadas em áudio podem ser elencadas, como a descrita por Wichern *et al.* (2010), que sobre a segmentação apresentada em Wichern *et al.* (2007), aplicam uma abordagem de indexação baseada em Modelos Escondidos de Markov. A partir desses índices são realizadas consultas por verossimilhança na base de dados de áudio. Ainda, para tornar o processo eficiente, é utilizada uma estratégia de indexação baseada em agrupamento, aplicando uma modificação do algoritmo *Spectral Clustering*. Esse algoritmo agrupa os dados a partir dos seus autovetores. Com os testes realizados, os autores concluíram que as técnicas utilizadas são eficazes tanto para gravações sem ruído quanto com ruídos.

3.6 Trabalhos que implementam avaliação do ambiente

O *Remote Environmental Assessment Laboratory* (REAL)⁵ é uma biblioteca que possibilita acesso e análise de dados acústicos (KASTEN *et al.*, 2012). Ela possui mecanismos que possibilitam seleção, extração e análise desses dados. Para isso, possui módulos que cobrem desde a leitura, gravação e processamento dos dados acústicos, até módulos de Aprendizagem de Máquina. A biblioteca cria imagens e medidas do som com a finalidade de auxiliar na identificação de padrões nas Paisagens Acústicas. A Figura 14 apresenta um conjunto de representações gráficas geradas. A partir de um dado arquivo de áudio, representado no canto superior esquerdo

Figura 14 – Representação de arquivo de áudio com espectrogramas, valores de PSD e NDSI, sumarização estatística etc.

Fonte: Adaptada de Gage (2008).

⁵ <<http://lib.real.msu.edu>>.

da imagem, são criados oscilogramas, espectrogramas (2D e 3D), uma tabela com a distribuição do PSD por faixa de frequência de 1 kHz de largura e gráficos que representam essa distribuição. Nessa tabela, para a faixa de frequências mais baixa, por conter apenas ruídos, foi atribuído o valor -1. Pode-se notar que as faixas entre 3-4 e 4-5 kHz possuem os maiores valores de energia. Igualmente, são calculados os valores do NDSI como é mostrado na tabela do canto inferior direito. Além do mais, estatísticas sobre a variação do PSD são calculadas. Uma abordagem que utiliza Mapa de Calor para visualizar a variação temporal do NDSI é exemplificada na próxima seção.

O trabalho de Gage e Axel (2014) emprega o mesmo sistema REAL para examinar a dinâmica da Paisagem Acústica em uma ilha nos *Twin Lakes*, Michigan, Estados Unidos. Para isso eles fizeram análises sobre o PSD e o NDSI.

Com o PSD eles apresentaram visualizações da variação por faixa de frequência, variação diária e mensal. Na Figura 15, analisando a variação mensal, os autores perceberam que a energia nas bandas de frequência analisadas fica mais uniforme quando os dias são mais longos (mês de Julho, no ápice do verão). Ou seja, a atividade acústica é igualmente intensa em todas as bandas de frequência.

Figura 15 – Variação mensal do PSD.

Fonte: Adaptada de Gage e Axel (2014).

A Figura 16 apresenta um Mapa de Calor que relaciona o dia Juliano⁶ (eixo horizontal), horário (eixo vertical) e os valores do NDSI mapeados como tons de cinza, sendo os valores baixos mais claros. Nessa figura, é possível perceber que os valores do índice são maiores durante o período noturno e diminuem de intensidade durante o diurno. Isso ocorre porque, durante a noite, a intensidade de sons biológicos é maior, enquanto que durante o dia, os sons antropológicos dominam a paisagem.

Outra ferramenta utilizada para avaliação de áudios ambientais, o *Ecosounds* é um sistema *web* que funciona como um repositório de áudios ambientais⁷. Essa página facilita o

⁶ Mapeia o Calendário Gregoriano de forma sequencial iniciando em uma data qualquer no passado.

⁷ <<https://www.ecosounds.org/>>

Figura 16 – Mapa de calor apresentando a variação do NDSI por horários do dia.

Fonte: Adaptada de Gage e Axel (2014).

gerenciamento, acesso, visualização e análise de dados acústicos. A manutenção do sistema é dada pelo *QUT Ecoacoustics Research Group*, que trabalha com tecnologias de detecção acústica aplicadas a monitoramento ambiental⁸. As análises realizadas na plataforma incluem detecção de eventos, geração e visualização de Índices Acústicos etc.

3.7 Visualização de gravações de longa duração

Como foi citado na Introdução, Towsey *et al.* (2014b) criaram uma abordagem de visualização de dados acústicos baseada em Imagens de Falsa Cor. Essas imagens apresentam cores que diferem das cores reais da imagem, com a finalidade de salientar características que não são facilmente detectadas na imagem original. Os testes apresentados foram aplicados em áudios de uma área da periferia de Bursbane, Austrália.

Após uma etapa de pré-processamento, foram calculados para cada 1 minuto de áudio os valores de Índices Acústicos. Com esses valores são criados espectrogramas que representam as intensidades de cada um deles (Figura 17). Cada espectrograma gerado é mapeado para uma cor do padrão RGB⁹ e por último são combinados para gerar a imagem de falsa cor (Figura 18).

Para aumentar a capacidade de visualização dos dados, os índices utilizados devem possuir uma baixa correlação, porque cada um é capaz de identificar atividades distintas no sinal. Os autores utilizaram o ACI, o H_t e o CVR.

Na Figura 19, que representa um dia completo de gravações, é possível verificar uma intensificação da atividade acústica por volta das 5 h e um declínio aproximadamente às 18 h.

⁸ <<http://research.ecosounds.org/>>

⁹ Nomenclatura atribuída ao sistema de cores vermelho (*red*), verde (*green*) e azul (*blue*).

Figura 17 – Espectrogramas dos índices.

Fonte: [Towsey et al. \(2014b\)](#).

Com base nas faixas de frequência, nas intensidades dos índices e nos respectivos áudios é possível identificar os insetos que apresentam algum sinal sonoro durante o período noturno. Também foi possível identificar a presença de espécies de aves durante o dia.

Pode-se, igualmente, combinar os índices em várias granularidades temporais (semanal, mensal, anual etc.), através da combinação de suas médias diárias. A Figura 20 apresenta um intervalo de meses onde é possível perceber que o período de atividade sonora diminui gradualmente até os meses de Junho e Julho e começa a aumentar novamente após esses meses. Isso ocorre por causa do inverno, onde o período do dia com luz solar é menor. Outra questão é a intensidade crescente das cores (valores dos índices) que ocorre por causa dos períodos de migração, reprodução e alimentação das espécies da região.

Para facilitar a navegação temporal, [Towsey et al. \(2014b\)](#) apresentaram o *Extended Acoustic Summary* (EASY). O exemplo dessa abordagem na Figura 21 mostra que o EASY não é um espectrograma, porque não apresenta as frequências, mas sim a relação entre os meses e horários. Mesmo assim, ainda é possível verificar tendências, como a diminuição do período de atividade sonora e a maior intensidade de atividade nos meses de Agosto e Setembro, principalmente entre 10 h e 17 h.

Figura 18 – Combinação dos índices para gerar a imagem de falsa cor.

Fonte: [Towsey et al. \(2014b\)](#).

Figura 19 – Espectrograma de um período de 24 horas. Horários no eixo horizontal e no vertical as faixas de frequência.

Espectrograma de Falsa cor

Fonte: [Towsey et al. \(2014b\)](#).

Figura 20 – Espectrograma mensal.

Fonte: Adaptada de [Towsey et al. \(2014b\)](#).

Utilizando a mesma idéia de imagens de falsa cor, [Sankupellay et al. \(2015\)](#) propuseram uma abordagem para comparar quantitativamente ambientes distintos. Eles coletaram áudios em 6 áreas próximas a Bursbane, Austrália, em um total de 20 dias de gravações. De cada uma das áreas foram selecionados 2 dias consecutivos, obtendo assim 12 arquivos de áudio de 24 h cada (na verdade os últimos 5 minutos de cada dia não estão presentes nos arquivos por serem

Figura 21 – Sumário Acústico Estendido.

Fonte: Adaptada de [Towsey et al. \(2014b\)](#).

justamente os períodos nos quais os áudios estavam sendo persistidos).

A partir desses áudios foram geradas imagens de falsa cor utilizando os índices ACI, ENT ($1 - H_t$) e EVN (que contabiliza os eventos acústicos). Adicionalmente, nove índices e medidas acústicas (*Background noise*, *Average signal-to-noise ratio*, *Acoustic event count*, *High-frequency coverage*, *Mid-frequency coverage*, *Low-frequency coverage*, ACI, H_t e H_s), foram utilizados para representar cada um dos minutos dos áudios.

Também foram utilizadas as técnicas de Aprendizagem de Máquina, *Self-organizing Map* (SOM) ([KOHONEN, 1982](#)), *Fuzzy C-Mean clustering* ([BEZDEK, 1981](#)) e *Hierarchical Clustering* ([ROKACH; MAIMON, 2005](#)) para agrupar os áudios, utilizando os índices citados como características.

Inicialmente foram apresentados para uma Rede Neural SOM 10x10 todos os minutos dos arquivos de áudio utilizados (com um vetor de 9 características cada). Sobre os vetores de pesos gerados pelos neurônios da SOM, o algoritmo *Fuzzy C-Mean clustering* foi aplicado para gerar 27 grupos (identificados por letras na Figura 22a). Para melhor visualizar o que foi aprendido pela rede SOM, cada um dos vetores de índices da base de dados foi atribuído ao neurônio com maior atividade. Dessa maneira, a Figura 22b) apresenta a distribuição dos instantes de tempo referentes a esses neurônios. Além disso, foi mensurada a quantidade de minutos que está associada a cada um dos neurônios (Figura 22c). Com essas visões, os autores perceberam os neurônios que possuíam maiores ligações com os minutos dos áudios e que os horários noturnos tenderam a ser mapiados nas áreas superiores da rede.

Com a finalidade de validar os grupos gerados, foi preparado o espectrograma de falsa cor de uma das áreas selecionadas. A partir dele, os minutos associados aos grupos identificados pelo SOM associado ao *Fuzzy C-Mean clustering* são dispostos lado a lado e identificados com letras, como na Figura 23.

Por fim, para averiguar se os padrões encontrados podem ser utilizados para comparações entre áreas, os autores consideraram cada um dos grupos como estados acústicos. Então, eles

Figura 22 – Distribuição dos neurônios SOM. a) Grupos gerados pelo *Fuzzy C-Mean clustering* são identificados por letras, b) minuto médio associado a cada neurônio e c) quantidade de minutos associados a cada um dos neurônios.

a) Mapa de nós por grupo										
	1	2	3	4	5	6	7	8	9	10
1	G	R	R	L	Y	Y	S	S	S	H
2	G	R	R	L	Y	Y	S	S	S	H
3	K	R	R	L	Y	Y	M	S	S	H
4	K	I	I	L	N	N	N	M	M	M
5	J	J	J	X	T	T	C	C	C	C
6	Z	Z	Z	X	T	T	T	C	C	D
7	A	Z	Z	O	O	O	W	D	D	D
8	A	A	A	Q	Q	Q	α	U	U	P
9	A	F	F	V	V	V	α	U	U	P
10	F	E	E	E	V	V	B	B	B	P

b) Horário										
	1	2	3	4	5	6	7	8	9	10
1	16:40	16:40	10:00	18:30	16:50	-	4:20	4:20	4:00	-
2	15:10	-	11:30	21:00	-	23:20	2:30	3:50	-	-
3	15:20	16:20	16:00	17:50	20:20	23:20	3:10	4:30	-	-
4	15:10	17:00	19:50	19:00	20:10	21:40	0:50	8:30	-	8:10
5	12:40	14:50	15:00	14:10	16:40	13:30	1:30	10:00	-	13:40
6	11:00	6:20	11:40	14:00	16:20	13:50	12:10	-	-	11:20
7	10:40	8:30	7:40	13:00	11:00	14:00	11:40	11:00	12:20	11:30
8	8:30	12:50	8:40	6:50	12:30	11:50	11:20	11:20	-	12:00
9	8:20	12:20	8:30	6:20	6:10	12:50	11:00	-	10:50	10:10
10	10:30	13:00	8:00	8:50	6:00	13:00	10:50	9:40	10:30	-

c) Quantidade de minutos atribuída a cada nó										
	1	2	3	4	5	6	7	8	9	10
1	11	8	6	25	1	-	3	10	3	-
2	10	-	3	4	-	2	202	30	-	-
3	15	4	11	8	6	2	3	6	-	-
4	31	8	21	26	100	121	46	4	-	2
5	2	2	5	40	10	17	4	2	-	4
6	6	1	5	19	13	9	5	-	-	10
7	13	18	37	65	22	17	6	5	2	10
8	5	12	10	23	5	6	15	4	-	6
9	7	22	13	47	15	3	14	-	3	4
10	16	6	7	18	65	7	16	3	2	-

Fonte: Adaptada de [Sankupellay et al. \(2015\)](#).

criaram um histograma para cada um dos doze arquivos de áudio utilizados. Ao comparar dois desses arquivos eles conseguiram perceber, por exemplo, diferenças nos padrões de sons noturnos e semelhanças nos padrões do período da manhã. Ainda, para determinar se os histogramas possuem informação significativa, um algoritmo de classificação hierárquica foi aplicado sobre eles. Sendo assim, conseguiram perceber que existem mais diferenças entre as áreas de onde os arquivos foram coletados do que entre áudios consecutivos.

Outro trabalho que busca explorar visualmente grandes quantidades de áudios foi proposto por [Sánchez-Gendriz e Padovese \(2017\)](#). Nesse trabalho, os autores utilizaram como base o espectrograma médio e o SPL para explorar visualmente dados acústicos. Com essas ferramentas os autores propuseram: uma visualização nomeada *SPL-gram*; duas abordagens para detecção automática de coros de animais no amanhecer e no anoitecer, nomeadas como *Maximal Area*

Figura 23 – Agrupamentos dos minutos de um espectrograma de falsa cor.

Fonte: [Sankupellay et al. \(2015\)](#).

Figura 24 – Valores do SPL nas bandas de frequência a) baixa (entre 0.01 e 2 kHz) b) e alta (entre 2 e 12 kHz).

Fonte: Adaptada de Sánchez-Gendriz e Padovese (2017).

Triangulation (MaxAT) e *Local maximum* (LoMAX); e a exploração da existência de relações entre os valores do SPL e as fases da lua.

Para calcular o SPL as faixas de frequência foram divididas em baixas (entre 0.01 e 2 kHz) e altas (entre 2 e 12kHz) frequências. Foram calculados além de espectrogramas diários, os valores diários do SPL e medidas de centralidade como média e mediana foram utilizadas para sumarização desses valores. Além do mais, um filtro foi aplicado sobre o SPL, suavizando a curva da medida, com a finalidade de eliminar valores produzidos por sons temporários como o de embarcações.

Nas análises, o artigo relata que nas baixas frequências (referentes a coro de peixes) não foram encontradas relações entre os valores do SPL e as fases da lua (Figura 24).

Sobre as técnicas de detecção propostas pelos autores, o MaxAT apresentou melhor comportamento quando aplicado sobre a curva do SPL calculada para as baixas frequências. Enquanto isso, para altas frequências obteve desempenho semelhante ao do LoMAX. Segundo os autores, nas baixas frequências existe uma predominância de sons de peixes e a variabilidade desses sons é maior, o que impõe mais restrições à detecção. Essas condições associadas à vulnerabilidade em detectar falsos positivos, fazem com que o LoMAX não obtenha bons resultados em baixas frequências. Portanto, o MaxAT é melhor por apresentar boa performance na detecção de tendências na curva do SPL tanto em baixas, quanto em altas frequências.

Com relação à sumarização dos dados, ao comparar as Figuras 25 e 26, geradas sobre dados de um mesmo período, é possível verificar a diferença entre a sumarização utilizando a

Figura 25 – Sumarização utilizando a média a) do espectrograma b) do SPL nas baixas frequências c) e do SPL nas altas frequências.

Fonte: Adaptada de [Sánchez-Gendríz e Padovese \(2017\)](#).

média e a mediana. Em uma determinada parte do período, uma embarcação de pesquisa esteve presente próxima à área de coleta. A energia do som dessa embarcação é predominante na Figura 25, mascarando os demais padrões acústicos, o que não ocorre na Figura 26. Assim sendo, por ser menos influenciada por valores extremos, é preferível utilizar a mediana para sumarizar os padrões de atividades acústicas.

Enfim, para criação do *SPL-gram*, cada curva diária do SPL (tanto baixa quanto alta frequência) foi convertida para uma faixa de cores variando entre azul e vermelho. Essa representação é disposta verticalmente em uma imagem onde as colunas representam os dias e as linhas representam os horários (Figura 27). Além disso, linhas horizontais representam os horários do nascer e pôr do sol, e pontos em torno delas destacam a atividade acústica detectada pelo MaxAT. Segundo [Sánchez-Gendríz e Padovese \(2017\)](#), com esse tipo de representação foi possível: verificar o ruído de embarcação (tarja vertical vermelha entre os dias 10 e 30 de Maio) que é predominante na Figura 25; a atividade acústica do amanhecer e anoitecer e a sua relação com o posicionamento do sol; e revelar a relação entre ao aumento nos valores do SPL próximo do nascer e pôr do sol, e próximo da atividade detectada pelo MaxAT.

3.8 Considerações sobre os trabalhos apresentados

Sobre as referências citadas neste capítulo é possível apresentar algumas colocações. Na Seção 3.1, pode-se perceber que mesmo com a vasta utilização dos índices na análise de Paisagens Acústicas, existe uma variedade de controvérsias sobre sua utilização. Alguns trabalhos, como

Figura 26 – Sumarização utilizando a mediana a) do espectrograma b) do SPL nas baixas frequências c) e do SPL nas altas frequências.

Fonte: Adaptada de [Sánchez-Gendriz e Padovese \(2017\)](#).

os da Seção 3.2, nem mesmo utilizam essas características. Pelo visto é necessária, para cada estudo realizado, uma avaliação de quais índices melhor descrevem um ambiente específico, como eles se comportam e qual a melhor interpretação para ele nessas condições.

Os trabalhos sobre segmentação das Seções 3.3 e 3.4 apresentam bons resultados para extração de objetos significativos dos áudios. Algumas das abordagens da Seção 3.3, mesmo apresentando bons resultados, não parecem ser eficientes computacionalmente. Também, até o momento, não foram encontrados trabalhos comparativos entre os dois caminhos de segmentação (sinal e imagem). Assim, não fica claro se uma abordagem é melhor que a outra ou se a junção das duas no processo de segmentação pode melhorar os resultados. Mesmo assim, os trabalhos apresentam que existem vantagens na utilização dessas abordagens.

Sobre a abordagem da Seção 3.5 é preciso identificar a eficiência do processo e verificar quanto os resultados com e sem ruído de fundo são satisfatórios. Também é necessário definir o que é ruído para a análise, porque um determinado padrão que é considerado como ruído em um domínio pode não ser em outro. Em uma região tropical, por exemplo, a chuva pode ser um ruído a ser desconsiderado quando o interesse da análise é a biodiversidade do local. Em um cenário subaquático, o interesse pode estar sobre a detecção de sons de motores de barcos, podendo ser descartados os sons de animais etc.

Os artigos sobre visualização de dados acústicos apresentados na Seção 3.6 que utilizam a biblioteca REAL possuem problemas similares. Os processos realizados por ela não realizam nenhum grau de segmentação do áudio para separar os diversos tipos de sons existentes neles.

Figura 27 – Visualização *SPL-gram*. Nas imagens a) e b), os valores do SPL foram mapeados para uma escala de cores entre o azul e o vermelho. Cada coluna da imagem representa um dos dias do período e cada linha o horário. As linhas pontilhadas informam os horários do nascer e por do sol. Os pontos próximos as linhas indicam a detecção de atividade realizada pelo MaxAT. A imagem a) apresenta as baixas frequências e a imagem b) as altas.

Fonte: Adaptada de [Sánchez-Gendriz e Padovese \(2017\)](#).

Enquanto isso, no primeiro artigo discutido na Seção 3.7, a normalização dos valores dos índices pode gerar perda de dados. Informações relevantes podem não ser percebidas pelo usuário por causa das limitações humanas para perceber tonalidades diferentes, mas de coloração muito próxima. Além disso, a quantidade de canais dos padrões de cores limita a quantidade de índices que podem ser associados. Essas visualizações são inteiramente dependentes dos índices escolhidos, das suas suposições básicas e das suas dificuldades quanto à representação dos dados.

Sobre o segundo artigo mencionado na mesma seção, os algoritmos de Aprendizagem de Máquina utilizados não são justificados, além da não apreciação do seu funcionamento em conjuntos de dados maiores para verificar a eficiência do processo. Por fim, o último artigo descrito nessa Seção 3.7, mesmo apresentando justificativas para isso, não procura comparar os resultados obtidos com o SPL com o de outras medidas utilizadas na avaliação de Paisagens Acústicas. Outras medidas para detecção de atividades durante o amanhecer e o anoitecer podem ser utilizadas, como, por exemplo, algumas das descritas no Capítulo 2. Essas características podem conseguir verificar a associação entre a atividade sonora do ambiente e as fases da lua, por exemplo.

ANÁLISE DE CARACTERÍSTICAS E VISUALIZAÇÃO DE PAISAGENS ACÚSTICAS

Neste capítulo são apresentadas as etapas propostas para uma exploração e análise de Paisagens Acústicas (Figura 28). Inicialmente é apresentada uma etapa de extração de características, utilizando além de Índices e Medidas Acústicas, as imagens dos espectrogramas dos áudios. Depois disso é necessário avaliar quais dessas características melhor representam as paisagens analisadas. Nessa etapa são utilizadas abordagens visuais (Projeção Multidimensional) e numéricas (Coeficiente de Silhueta) para verificar a capacidade que os atributos têm de agregar dados semelhantes e segregar dados distintos. Por fim, um processo de análise visual é discutido, apresentando tipos de técnicas de Visualização de Dados que podem ser empregadas e informações que podem ser extraídas com elas, além de uma comparação com abordagens estatísticas convencionais. Dentre as técnicas visuais, é apresentada a HiPP Estendida - *eXtended HiPP* (xHiPP), um conjunto de adições na projeção multidimensional HiPP (cf. Seção 2.7.3) que pode ser utilizada para explorar conjuntos de dados, como de áudios, textos, imagens dentro outros.

Figura 28 – Passos para análise dos dados acústicos.

Fonte: Elaborada pelo autor.

4.1 Extração de Características

Como pode ser visto no Capítulo 2, a quantidade de características para representar um domínio como áudio é extensa. Por causa disso, torna-se necessário definir quais delas melhor representam um determinado conjunto. Estratégias válidas para uma definição inicial são: escolher características comumente utilizadas na literatura; selecionar características que possuam pouca utilização, mas bons resultados onde são aplicadas; escolher características que possuam boa aplicação em áreas correlacionadas etc. Seguindo essas estratégias, as próximas seções apresentam as características que foram utilizadas durante esta pesquisa.

4.1.1 Características de Áudio

Para descrever os áudios foram utilizados os Índices Acústicos ACI, ADI, AEI, AR, BIO, H, Hf, Ht, M, NDSI, por serem comumente utilizados para descrever Paisagens Acústicas. Outras Medidas Acústicas para análise de sinais como SPL, NP, *Roughness*, *Rugosity* e RMS também foram utilizadas nos testes. Além disso, foram extraídos os coeficientes do MFCC que são empregados normalmente no reconhecimento de fala. As definições e aplicações de todas essas características podem ser verificadas nos Capítulos 2 e 3.

4.1.2 Características de Imagem

Além da utilização dos índices e medidas já mencionados, foi estabelecido nesta pesquisa a utilização de características das imagens dos espectrogramas dos áudios como alternativa para descrição de Paisagens Acústicas. A utilização de imagens pode ser proveitosa, pois limita o espaço de análise e pode proporcionar uma interação visual com o usuário. Foram utilizadas três abordagens para descrição de imagens, sendo que na primeira, as imagens foram segmentadas para segregar eventos dentro dos espectrogramas e encontrar características que pudessem distingui-los. Na segunda abordagem, foram extraídas características globais de cor e textura das imagens, como GLCM, LBP, Fourier para textura, GCH, ACC, CCV, BIC (cf. Seção 2.5). Por fim, foi utilizado o Autoencoder (cf. Seção 2.6) com a finalidade de criar um modelo capaz de extrair atributos das imagens sem a necessidade prévia de especificar quais características devem ser obtidas.

4.2 Avaliação das Características

Antes mesmo de iniciar o processo de avaliação da qualidade das características utilizadas, é importante ressaltar a relevância das perguntas que são “feitas” para os dados representados por elas. A qualidade das características extraídas depende de qual informação o usuário busca nos dados. Em alguns cenários, perguntas específicas podem existir, como: determinada espécie de pássaro pode ser encontrada no ambiente? Quanto sons de motores influenciam uma área

específica? Existem sons que não deveriam estar presentes na região analisada? Quando sons de embarcações ou de espécies invasoras passaram a ser percebidos? Etc. Quando essas perguntas são claras, a definição de quais características utilizar possui uma avaliação precisa, porque basta escolher os atributos que destacam os dados necessários para responder às questões. Entretanto, nem sempre essas perguntas estão claras, e até mesmo quando elas estão, outras perguntas podem surgir durante a análise, revelando novas necessidades. Essa falta de definição do principal foco da análise dos dados reforça o caráter exploratório e a necessidade da aplicação de abordagens visuais de análise como descrito no capítulo introdutório.

Por causa disso, nesta pesquisa, no lugar de explorar atributos com o intuito de evidenciar questões específicas e internas de cada região, buscou-se primeiro analisar as características a partir da maneira que elas diferenciam os ambientes. Sendo assim, foram utilizados dados de 5 áreas de coleta. Duas dessas áreas são terrestres e de uma mesma região. As outras 3 são subaquáticas, onde 2 são de uma mesma região. Nessa situação eram evidentes as diferenças entre as bases de dados terrestre e aquáticas. Entretanto, existia apenas uma suposição sobre a diferença entre as duas áreas de coleta terrestres e entre as duas subaquáticas que pertencem à mesma região.

Os valores de silhueta foram calculados para cada conjunto de características depois que elas foram extraídas. Para cálculo desses valores, foram considerados como rótulos dos dados suas áreas de coleta. A partir dos valores de silhueta foi possível medir a capacidade das características de segregar os dados.

Também foram aplicadas para inspecionar visualmente os dados as projeções MDS, ForceScheme, LSP, t-SNE, PLMP, LAMP, além do PCA, todas elas citadas anteriormente na Seção 2.7.1. A partir dessas projeções foi possível verificar a qualidade das características. Se para áreas que são distintas a projeção não consegue representar a separação entre esses dados, existe uma possibilidade das características utilizadas não serem as melhores para descrever as diferenças entre os dados.

Para gerar o *layout* dessas projeções também foram consideradas as áreas de coleta como rótulos para os áudios. É necessário avaliar qual projeção melhor se adequa aos dados. Essa avaliação pode ser apenas visual, mas ela é subjetiva e por vezes complexa, porque duas ou mais projeções podem apresentar visões semelhantes dos dados, ou estarem sendo aplicadas a uma quantidade tão grande de dados que seja difícil destacar diferenças entre essas projeções. Para sanar isso, foram calculados valores de silhueta para os dados gerados pelas projeções e esses valores foram comparados com os valores de silhueta gerados no espaço original das características.

Por fim, foram realizados alguns testes com técnicas de Seleção Automática de Atributos para verificar a qualidade da abordagem proposta nesta pesquisa frente um método automático de seleção/análise de características. Foram utilizados os métodos de seleção *Correlation based*,

Information Gain e *Relief-F* disponíveis no programa de mineração de dados Weka¹ e citados em trabalhos como o de Jović, Brkić e Bogunović (2015). A partir dos conjuntos de características gerados anteriormente, esses métodos foram aplicados e com as características selecionadas por eles foram gerados valores de silhueta e projeções para que fosse possível a comparação com os resultados já obtidos.

4.3 Análise Visual

Depois que as características foram avaliadas e o melhor conjunto delas para descrever os dados utilizados foi encontrado, foram empregadas técnicas de visualização discutidas na Seção 2.7, que destacam as relações entre os atributos dos dados (Coordenada Paralela e Mapa de Calor) e que representam a disposição espacial dos dados (Projeção Multidimensional). As próximas seções apresentam como essas técnicas podem ser empregadas e uma descrição do trabalho realizado com a projeção xHiPP. É apresentado, também, um contraponto das técnicas que visualizam de atributos com ferramentas usuais de análise estatística para verificar como aquelas técnicas visuais podem complementar e melhorar as análises dessas abordagens estatísticas.

4.3.1 HiPP Estendida

Como já foi descrito, é possível utilizar Projeções Multidimensionais (cf. Seção 2.7.1) para explorar estruturas e relações dentro de conjuntos de dados descritos por suas características. Com o objetivo de utilizar uma abordagem multiescala no problema da descrição de Paisagens Acústicas, a técnica HiPP (cf. Seção 2.7.3) foi modificada, com a finalidade de melhorar sua capacidade de análise e flexibilidade de aplicação. Essa projeção foi utilizada por possuir boas condições para explorar dados em vários níveis de detalhe, facilitar a interação com os dados analisados e por conseguir organizar visualmente conjuntos de dados de tamanhos variáveis sob a mesma perspectiva (PAULOVICH; MINGHIM, 2008).

Essa técnica pode apresentar vários níveis de detalhamento, então é necessário apresentar para o usuário uma sumarização de cada nível. O artigo original apresenta testes como conjuntos de textos e utiliza extração de tópicos desses conjuntos para descrever o que existe em cada grupo apresentado, característica também existente na xHiPP. Entretanto, ainda não foi implementada uma maneira para sumarizar dados como imagens e áudios. Sendo assim, todos os testes reportados neste trabalho com a xHiPP apresentam somente a distribuição dos pontos e o seu agrupamento, não destacando os níveis de detalhamento. Nesta seção são descritas as alterações realizadas sobre a técnica original e como ela pode ser aplicada.

¹ <<https://www.cs.waikato.ac.nz/ml/weka/>>

4.3.1.1 Modificações Realizadas

A primeira modificação foi aplicada na estimativa do raio do nó raiz. Isso é necessário, porque na estimativa do raio de um nó não folha é utilizado o raio do seu pai. Vale lembrar que a estimativa do nó raiz não está presente no artigo original. Dessa maneira, os nós do primeiro nível da árvore necessitam do raio do nó raiz. Logo, no início do algoritmo de particionamento, quando todas as instâncias de dados são filhas da raiz, o centroide desses dados é calculado e o raio é a maior distância entre os demais itens do conjunto e esse centroide.

A segunda modificação foi realizada no processo de criação e particionamento da árvore. Em vez de utilizar somente um algoritmo de agrupamento (*Bisecting k-means*), pode-se agora escolher entre 3 técnicas de agrupamento: *k-means*, *k-medoid* ou hierárquico aglomerativo. A quantidade de grupos é definida por padrão como $\sqrt{n}$, mas pode ser modificada para melhor se adequar aos dados projetados.

A terceira modificação foi realizada no processo de projeção dos dados. Pode-se utilizar, para as folhas da árvore, além da LSP utilizada no artigo original, as projeções MDS, ForceScheme, t-SNE, PLMP, LAMP ou o PCA, todas essas citadas anteriormente na Seção 2.7.1. Para melhorar a eficiência do processo implementado, os nós internos (não folha) da árvore são projetados apenas com a ForceScheme.

A quarta modificação adicionou a possibilidade de inverter a ordem do processo utilizado pela HiPP. No lugar de primeiro particionar os dados gerando uma estrutura hierárquica e depois projetar essa estrutura, tornou-se possível, também, primeiro projetar os dados e depois gerar uma hierarquia a partir dos valores dessa projeção.

Por fim, assim como no trabalho original, tópicos dos grupos são apresentados, quando é realizada exploração de dados textuais. Para extração desses tópicos, aqui foi utilizado o *Latent Dirichlet Allocation* (LDA) (BLEI; NG; JORDAN, 2003). Essa característica foi estendida e, para imagens, os grupos podem apresentar a figura referente ao seu medóide. Além disso, os dados referentes a cada folha da árvore podem ser verificados, como textos, imagens, áudios e seus respectivos espectrogramas. Para todos os tipos de dados, cada nó é colorido com uma cor referente à sua classe, se ela existir, e os grupos possuem a cor predominante dentro deles.

Com essas modificações é possível utilizar um conjunto de combinações e identificar qual delas melhor representa cada conjunto de dados analisado. Assim como nas visualizações apresentadas posteriormente, a próxima seção apresenta um exemplo da aplicação dessa técnica para dados acústicos. Mais testes, resultados, avaliações e discussões sobre a xHiPP são apresentados no capítulo de resultados.

4.3.1.2 Aplicação

A Figura 29 apresenta um conjunto de dados acústicos terrestres, utilizando 12 coeficientes do MFCC como atributos desses dados. Com isso, é possível perceber por exemplo,

Figura 29 – xHiPP para visualizar áudios de uma base de dados de 3.041 arquivos. As cores roxo e cinza representam duas regiões de coleta dos dados. A linha azul tracejada está destacando a separação entre os dados. As setas azuis são exemplos de arquivos que estão em meio a dados de classe diferente da sua.

Fonte: Elaborada pelo autor.

que existe distinção entre as duas áreas de coleta representadas (cores roxo e cinza). É possível também identificar instâncias dos dados que mesmo pertencendo a uma determinada classe, podem estar em meio a itens de outra classe.

4.3.2 Coordenadas Paralelas

As Coordenadas Paralelas podem ser empregadas para compreender como as características dos dados estão distribuídas e como elas se relacionam. A Figura 30 apresenta os mesmos dados utilizados na Figura 29, mas descritos por outras características acústicas. Nela é possível identificar: a faixa de valores, valores atípicos (*outliers*), a proporcionalidade entre as características, características que podem ser úteis na classificação dos dados etc. Por exemplo, nos eixos BIO e SPL, retângulos destacam regiões onde existe separação entre grupos dos dados, o que pode indicar que essas características sejam úteis para descrever as diferenças entre esses grupos de dados. Também é possível fazer comparações entre regiões, horários distintos etc., a depender de como os dados forem rotulados.

Uma ferramenta estatística que também é utilizada para analisar o comportamento das características é o *Boxplot*². Na Figura 31b, é possível verificar as faixas de valores dos índices, valores extremos, se as medianas dos índices são parecidas etc. No entanto, como pode ser visto

² Ferramenta gráfica que representa a variação dos atributos através dos valores máximo e mínimo, 1º, 2º (mediana) e 3º quartil, além de identificar valores atípicos (*outliers*).

Figura 30 – Coordenadas Paralelas para visualizar características de áudios extraídas de uma base de dados de 3041 arquivos. As cores roxo e verde representam duas regiões de coleta dos dados. As setas próximas ao ACI identificam valores extremos e os retângulos nos eixos BIO e SPL transições entre as classes.

Fonte: Elaborada pelo autor.

inicialmente na Figura 31a, essa abordagem apresenta dificuldade na visualização, porque o ACI e o RMS possuem faixas de valores maiores do que as faixas dos demais índices. Dessa maneira, é necessário aplicar algum pré-processamento como a utilização de logaritmo apresentada na Figura 31b. É claro que essa estratégia não é válida para os valores negativos que podem ocorrer no NDSI e no SPL. Abordagens como normalizações e padronizações podem ser aplicadas nesses valores, mas podem modificar e mascarar alguma peculiaridade dos dados. Esse procedimento não é inicialmente necessário quando as Coordenadas Paralelas são utilizadas.

4.3.3 Mapa de Calor

O Mapa de Calor pode ser empregado para compreender como cada atributo se comporta. Diferente das Coordenadas Paralelas, que avaliam várias dimensões ao mesmo tempo, essa técnica pode ser aplicada na visualização de cada dimensão separadamente. O índice H representado em um dos eixos da Figura 30 foi representado na Figura 32 através do Mapa de Calor. Como

Figura 31 – Boxplot de características de áudio extraídas de uma base de dados de 3041 arquivos.

(a) Valores originais das características.

(b) Valores das características modificados pela expressão $10 \times \log_{10}(x)$.

Fonte: Elaborada pelo autor.

Figura 32 – Mapa de Calor dos valores do índice H apresentado como uma das dimensões da Figura 30. Os valores do índices são mapeados entre o amarelo (valores baixos) e o marrom (valores altos). As linhas da tabela representam dias e as colunas os horários de coleta. As marcações em azul destacam alguns padrões do índice H para esses dados.

Fonte: Elaborada pelo autor.

destacado nos retângulos azuis, é possível, com essa técnica, perceber que o índice é menos sensível aos sons compreendidos entre 18 h e meia-noite, apresentando valores baixos nessa faixa de horários. Também é possível identificar que durante a madrugada os valores aumentam gradualmente, enquanto que durante o dia existe oscilação dos valores.

A análise de cada dimensão pode ser realizada, também, através da apresentação de Séries Temporais. A Figura 33 apresenta a decomposição³ dos valores do índice H apresentado no Mapa de Calor anterior. Na primeira linha da imagem podem ser vistos os valores originais do índice. Na segunda, pode-se verificar a representação das tendências (de alta ou de baixa)

Figura 33 – Série temporal do índice H decomposta. Cada linha da imagem representa: os valores do índice; a tendência de deslocamento desses valores; os padrões sazonais que acontece em curtos ciclos de tempo; e um resíduo que não pode ser explicado pelas componentes anteriores.

Fonte: Elaborada pelo autor.

³ Utilizada a Decomposição Aditiva $Y(t) = T(t) + S(t) + a(t)$, que é uma das maneiras de decompor uma série. Sendo Y os valores originais, T a tendência, S a sazonalidade e a um componente estocástico.

desse índice durante o tempo analisado. Na terceira linha, existe uma avaliação cíclica dos valores, onde é possível perceber a repetição de um padrão em curtos espaços de tempo. Na última linha, é apresentada uma componente estocástica que não pode ser descrita nas componentes determinísticas anteriores. Entretanto, para utilização desse tipo de abordagem o usuário necessita de um maior conhecimento estatístico do que com a utilização do Mapa de Calor.

4.4 Combinação das Etapas de Análise

A partir das etapas propostas na Figura 28 e descritas durante as seções anteriores, foi possível identificar quais as melhores abordagens para analisar as bases de dados utilizadas. Na fase de avaliação das características, a primeira etapa a ser realizada foi a análise de qual técnica de Projeção Multidimensional melhor apresentava às características utilizadas. A projeção t-SNE obteve os melhores resultados de segregação visual, embora não tenha alcançado os melhores valores de silhueta.

No que diz respeito à avaliação da qualidade das características analisadas, todas foram capazes de diferenciar os dados terrestres dos dados aquáticos tanto visualmente (com a t-SNE) quanto numericamente. Além disso, os índices e medidas acústicas foram as características que melhor descreveram diferenças entre as áreas de coleta terrestre. Especificamente sobre a base de dados terrestres, os coeficientes do MFCC apresentaram os melhores resultados para segregar suas áreas de coleta. Entretanto, as áreas de coleta que pertencem à mesma região subaquática provavelmente não possuem distinção alguma, porque nenhuma das características foi capaz de destacar qualquer separação entre elas.

Para os testes com Seleção Automática de Atributos, os resultados não apresentaram grande impacto nos valores de silhueta e aparentemente não modificaram os resultados da t-SNE.

Por fim, na fase de análise visual, foi possível destacar as diferenças entre as bases de dados analisadas e tendências dentro desses dados a partir das técnicas de visualização empregadas. Foi possível, por exemplo, com a xHiPP verificar arquivos de áudio que possuem mesmas características formarem agrupamentos, com as Coordenadas Paralelas foi possível identificar relações entre as características de uma mesma área e compará-las com outras áreas, e com o Mapa de Calor foi possível perceber o comportamento de características específicas durante o tempo. Além do mais, é possível perceber o auxílio que essas técnicas visuais podem trazer para análise e, por causa disso, os testes apresentados no capítulo de resultados dão preferência para elas, no lugar da utilização das abordagens estatísticas utilizadas para comparação nas seções anteriores.

4.5 Detalhes de Implementação

Para as implementações utilizadas nesta pesquisa foram utilizadas as seguintes ferramentas. A implementação da xHiPP foi feita com a linguagem R⁴, empregando o pacote de projeções *mp*⁵, o pacote de programação paralela *doParallel*⁶ e o pacote *topicmodels*⁷ para extração de tópicos de texto. Além disso, a visualização dos resultados da projeção utilizam a biblioteca JavaScript D3.js⁸ associada com a biblioteca *Shiny*⁹ do R para prover uma interface *web*. Os passos de particionamento e projeção dos dados são realizados no R, enquanto que os passos de espalhamento, remoção de sobreposições e renderização são executados com o JavaScript.

A biblioteca D3.js também foi utilizada para criar as visualizações de Coordenadas Paralelas¹⁰ e de Mapa de Calor¹¹. Os códigos iniciais foram manipulados para se adequar as necessidades da pesquisa.

Para extração de características de áudio, foram utilizados do R os pacotes: *Seewave*¹², *Soundecology*¹³ e *tuneR*¹⁴. As projeções utilizadas para avaliação das características também utilizaram o pacote *mp*. A geração dos espectrogramas também foi realizada com o pacote *Seewave* do R. Para extração de características de imagens, a biblioteca OpenCV¹⁵ foi utilizada com as linguagens Python (descriptor WLD) e C++ (demais descritores). Para utilização de *DeepLearning*, foram úteis as bibliotecas Keras¹⁶ e *TensorFlow*¹⁷, ambas empregadas com a linguagem Python. As etapas de pré-processamento das imagens como recorte, divisão e redimensionamento também utilizaram a linguagem Python. Por fim, a extração de características de áudio e de imagens foi realizada com um processador Intel Core i7-2600, 3.40GHz, 8 núcleos e 16GB de memória RAM. O treinamento do modelo de Autoencoder foi executado em uma placa de vídeo NVIDIA TITAN X.

4 <<https://www.r-project.org/>>
 5 <<https://cran.r-project.org/web/packages/mp/index.html>>
 6 <<https://cran.r-project.org/web/packages/doParallel/index.html>>
 7 <<https://cran.r-project.org/web/packages/topicmodels/index.html>>
 8 <<https://d3js.org/>>
 9 <<https://shiny.rstudio.com/>>
 10 <<https://bl.ocks.org/syntagmatic/05a5b0897a48890133beb59c815bd953>>
 11 <<http://bl.ocks.org/ianychang/8119685>>
 12 <<http://rug.mnhn.fr/seewave/>>
 13 <<http://jvillanueva.github.io/soundecology/>>
 14 <<https://cran.r-project.org/web/packages/tuneR/index.html>>
 15 <<https://opencv.org/>>
 16 <<https://keras.io/>>
 17 <<https://www.tensorflow.org/>>

EXPERIMENTOS E RESULTADOS

Este capítulo apresenta os principais resultados dos testes realizados com as ferramentas propostas neste trabalho. São descritos experimentos, resultados e avaliação da xHiPP; uma avaliação das alternativas para extração de características de Paisagens Acústicas; e resultados de explorações visuais das paisagens a partir das características que foram melhor avaliadas pelo processo proposto, ou seja, os índices e as medidas acústicas.

5.1 HiPP Estendida

Com o objetivo de verificar o comportamento das combinações de algoritmos adicionadas à xHiPP e listados na Seção 4.3.1.1, foram utilizadas as bases descritas no Quadro 2. Também nesse quadro são apresentados os valores de silhueta para cada uma das bases, com a finalidade de identificar como elas se comportam originalmente. Para avaliar os resultados da projeção, foram utilizadas as medidas de *Stress*, Coeficiente de Silhueta, *Neighborhood Hit* e *Neighborhood Preservation*, todas descritas na Seção 2.7.2. Nos cálculos do *Neighborhood Hit* e *Neighborhood Preservation* foi fixado o tamanho da vizinhança em $\sqrt{n}$, o mesmo valor que é utilizado no artigo da HiPP (cf. Seção 2.7.3) para quantidade mínima de instâncias por grupo. Para facilitar a apresentação, em todas as imagens de resultados da xHiPP é apresentado apenas o nível das folhas da estrutura.

5.1.1 Resultados

No primeiro teste, apresentado na Figura 34, foi utilizado o conjunto de dados Íris, o qual possui duas classes com fronteiras bem definidas, mas próximas uma da outra, e uma terceira bem separada das demais. O valor de silhueta dessa base (0,58) pode ser considerado como razoável segundo a classificação de silhueta descrita na Seção 2.7.2. Ao modificar a ordem de execução da xHiPP (agrupamento-projeção para projeção-agrupamento) a separação original das classes é mantida, além dos valores das medidas serem melhores para a ordenação utilizada.

Quadro 2 – Lista de bases de dados utilizadas para testes com a xHiPP.

Nome	Itens	Atributos	Classes	Silhueta	Descrição
Íris (LICHMAN, 2013)	150	4	3	0,58	Gênero de plantas Íris.
Wine (LICHMAN, 2013)	178	13	3	0,20	Vinhos italianos de diferentes produções.
Banknote (LICHMAN, 2013)	1372	4	2	0,18	Imagens de notas de dólar originais e falsificadas.
Corel (LI; WANG, 2003)	1000	150	10	0,15	Imagens de classes variadas como: praia, ônibus, flores etc.

Fonte: Elaborada pelo autor.

O teste representado na Figura 35 foi executado na base de dados Wine, que possui duas classes sobrepostas e uma terceira classe com fronteira bem definida, próxima das anteriores. O valor de silhueta da base (0,20) destaca falta de estrutura de agrupamento segundo a classificação que está sendo utilizada. Nesse teste, os algoritmos de agrupamento foram variados para verificar o impacto de cada um nas medidas de avaliação. O teste com *k-means* apresenta a melhor resultado de silhueta, enquanto que o teste com hierárquico possui o melhor valor de *Neighborhood Preservation* e o teste com *k-medoid* possui os melhores valores de *stress* e *Neighborhood Hit*. Mesmo com essa variação e com diferenças nas posições das projeções, a separação entre as classes não é visualmente modificada.

No teste seguinte, apresentado na Figura 36, a xHiPP foi aplicada sobre a base de dados Banknote. As classes dessa base possuem fronteiras bem definidas e próximas, possuindo

Figura 34 – Testes utilizando a base de dados Íris, variando a ordem das etapas da xHiPP. As cores dos pontos representam as classes da base de dados.

(a) *k-means* + *ForceScheme*. Com *Stress* = 0,7061; *N. Preservation* = 0,5483; *N. Hit* = 0,8644 e *Silhueta* = 0,4707. (b) *ForceScheme* + *k-means*. Com *Stress* = 0,4113; *N. Preservation* = 0,7494; *N. Hit* = 0,9483 e *Silhueta* = 0,5562

Fonte: Elaborada pelo autor.

Figura 35 – Testes utilizando a base de dados Wine, variando o agrupamento e projetando com a *ForceScheme*. As cores dos pontos representam as classes da base de dados.

Fonte: Elaborada pelo autor.

um valor de silhueta (0,18) que identifica nenhuma estrutura. Nesse teste foram variados os algoritmos de projeção, sendo apresentados na imagem os 4 melhores resultados. O melhor valor de *stress* foi obtido pelo PCA, o melhor valor de *Neighborhood Preservation* pelo MDS e os melhores valores de *Neighborhood Hit* e silhueta pela t-SNE. Esses resultados ocorreram com poucas modificações visuais entre as projeções.

O último teste, representado na Figura 37, apresenta os resultados da aplicação da xHiPP nos dados da base Corel. Nessa base existem classes com fronteiras bem definidas, mas próximas umas das outras. O valor de silhueta (0,15) é influenciado mais pela quantidade de classes do que pela proximidade entre elas. Nos testes foram variados os algoritmos de projeção, agrupamento e a ordem de execução das etapas da xHiPP. Cada algoritmo de agrupamento foi combinado com as 7 projeções, totalizando 42 testes (21 para cada ordenação de etapas). Apenas as combinações com melhores resultados são apresentadas na Figura 37.

Para os testes que aplicaram agrupamento seguido de projeção (Figuras 37a a 37c), a combinação de *k-means* e PCA possui os melhores valores de *stress* e *Neighborhood Preservation*. Por outro lado, os melhores valores de *Neighborhood Hit* e silhueta foram obtidos pelo *k-medoid* com MDS. Para os testes que aplicaram projeção seguida de agrupamento (Figuras 37d a 37f), as melhores combinações foram alcançadas com a t-SNE, sendo que elas possuem os mesmos valores para as 4 medidas utilizadas. Por fim, os valores das medidas para projeção-agrupamento são melhores do que para agrupamento-projeção.

Figura 36 – Testes utilizando a base de dados Banknote, mantendo o agrupamento *k-means* e variando as técnicas de projeção. As cores dos pontos representam as classes da base de dados.

(a) *ForceScheme*. Com *Stress* = 0,9442; *N. Preservation* = 0,6261; *N. Hit* = 0,9168 e *Silhueta* = 0,1914. (b) *MDS*. Com *Stress* = 0,9371; *N. Preservation* = 0,6346; *N. Hit* = 0,909 e *Silhueta* = 0,1922.

(c) *PCA*. Com *Stress* = 0,9297; *N. Preservation* = 0,6182; *N. Hit* = 0,9003 e *Silhueta* = 0,1939. (d) *t-SNE*. Com *Stress* = 0,9391; *N. Preservation* = 0,6265; *N. Hit* = 0,9195 e *Silhueta* = 0,1945.

Fonte: Elaborada pelo autor.

5.1.2 Discussão

A flexibilidade de modificar a ordem de execução (agrupamento-projeção, projeção-agrupamento) dos passos da xHiPP permite melhorar a representação dos dados a depender da base utilizada. Essa melhoria pode ser constatada pelos valores das medidas da Figura 34. A possibilidade de associar o agrupamento com várias técnicas de projeção aumenta a eficácia dessas técnicas de visualização. Isso pode ser visto, por exemplo, com o PCA que apresenta uma representação visual inconsistente quando os dados possuem relações não-lineares entre os atributos, mas que em alguns testes alcançou bons resultados (Figura 37a). Além disso, aproveitar a segregação realizada pelo t-SNE, por exemplo, para incrementar os resultados de técnicas de agrupamento simples como o agrupamento hierárquico, aumenta a capacidade de segregação e exploração dos dados.

Figura 37 – Melhores resultados dos testes utilizando a base Corel, variando ordem das etapas de execução, algoritmos de agrupamento e projeção. As cores dos pontos representam as classes da base de dados.

(a) *k-means* + PCA. Com Stress = 0,9169; *N. Preservation* = 0,4408; *N. Hit* = 0,6792 e Silhueta = 0,1052.

(b) *k-medoid* + MDS. Com Stress = 0,9288; *N. Preservation* = 0,4265; *N. Hit* = 0,7001 e Silhueta = 0,1329.

(c) Hierárquico + *ForceScheme*. Com Stress = 0,9334; *N. Preservation* = 0,329; *N. Hit* = 0,493 e Silhueta = -0,0407.

(d) t-SNE + *k-means*. Com Stress = 0,9083; *N. Preservation* = 0,5541; *N. Hit* = 0,7762 e Silhueta = 0,3177.

(e) t-SNE + *k-medoid*. Com Stress = 0,9083; *N. Preservation* = 0,5541; *N. Hit* = 0,7762 e Silhueta = 0,3177.

(f) t-SNE + Hierárquico. Com Stress = 0,9083; *N. Preservation* = 0,5541; *N. Hit* = 0,7762 e Silhueta = 0,3177.

Fonte: Elaborada pelo autor.

Infelizmente, em casos onde a quantidade de instâncias ou a quantidade de atributos aumenta (mais de 1.000 para instâncias e mais de algumas dezenas para atributos), a eficiência da implementação realizada cai consideravelmente. Isso ocorre por causa da complexidade original da projeção $O(n\sqrt{n})$, como apresentado na Seção 2.7.3, mas também por causa das características da linguagem R, mesmo utilizando técnicas de paralelismo. Outro gargalo na implementação é a lentidão da biblioteca D3.js para renderizar (desenhar) quantidades de dados acima dos milhares.

5.2 Exploração de Paisagens Acústicas

Esta seção apresenta os experimentos realizados com extração de características de Paisagens Acústicas e sua avaliação visual e numérica. Também são apresentados resultados de visualizações dessas paisagens através de Coordenadas Paralelas, Mapa de Calor e da xHiPP.

5.2.1 Bases de Dados

Para os testes de extração de características e visualização, foram utilizados 4.340 arquivos de áudio distribuídos em dois grupos significantes:

1. **Terrestres:** dados cedidos pelo professor Bryan C. Pijanowski da Universidade de Purdue, Indiana, EUA. Esses dados foram coletados em duas regiões da Estação Biológica *La Selva* em Costa Rica. A primeira região (CostaRica1) é uma floresta primária próxima ao rio *Sarapiquí*. A segunda região (CostaRica2) é uma floresta secundária na mesma reserva, mais afastada das margens desse rio¹. Para cada hora do dia foram coletados 4 arquivos de áudio, o primeiro com 15 minutos e os demais com 1 minuto cada. No total foram coletados 3.061 arquivos, sendo 1.246 de CostaRica1 e 1.815 de CostaRica2. Em CostaRica1 a coleta foi realizada entre os dias 06 e 19 de março de 2015. Enquanto que em CostaRica2, entre os dias 06 e 20 de março, e 15 e 20 de abril de 2015. Todos os arquivos possuem o formato *Free Lossless Audio Codec* (FLAC). Esses áudios possuem sons de insetos, anfíbios, pássaros, chuva, motores etc.;
2. **Subaquáticos:** dados cedidos pelo professor Linilson R. Padovese da Escola Politécnica da USP.
 - Dados coletados em Ilhéus, no litoral sul da Bahia. Foram coletados arquivos contínuos de 15 minutos cada. No total foram coletados 480 arquivos entre os dias 03 e 04 setembro, e entre os dias 18 e 22 de setembro de 2014. A base fornecida está dividida em duas partes, uma com 200 arquivos (Ilheus1) e a outra com 280 (Ilheus2). Todos os arquivos possuem o formato *WAVEform audio format* (WAV). O som predominante nesses áudios é de baleias Jubarte e coro de peixes;
 - Dados coletados na Laje de Santos, litoral de São Paulo. Foram coletados arquivos contínuos de 15 minutos cada. No total, foram coletados 799 arquivos entre os dias 17 e 18 de março, e entre os dias 27 de março e 03 de abril de 2015. Todos os arquivos possuem o formato WAV. Essa base possui sons de peixes, crustáceos e embarcações.

Imagens dos espectrogramas (1366x768) foram geradas para cada arquivo de áudio. Tomando como exemplo CostaRica2 que possui 1.815, são necessários para gerar essas imagens

¹ Coordenadas GPS do sensor utilizado na floresta primária: 10.43167528 -84.02136972. GPS da secundária: 10.42254278 -84.01599944.

aproximadamente 30 segundos para um áudio de 1 minuto e 4 minutos para um arquivo de 10 minutos. Para essa área o processamento durou cerca de 2 dias. De cada imagem, foi extraída apenas a região útil (1110x680) para todos os testes com processamento de imagens (segmentação, extração de descritores e processamento com *DeepLearning*). Essa região útil desconsidera as legendas das imagens. Assim, nas próximas seções, todas as vezes que forem utilizadas as imagens dos espectrogramas, as dimensões delas são da área útil extraída.

5.2.2 Extração e Avaliação de Características

Como descrito na Seção 4.1, a extração de características relevantes e pertinentes é de grande importância para a análise de dados. Assim sendo, as próximas seções apresentam os testes com diferentes conjuntos de características apresentados no capítulo anterior.

Nos testes que utilizam Projeção Multidimensional, foi aplicada a normalização *Min-max* entre [0, 1] e a projeção t-SNE. Para essa escolha, foram analisados os valores de silhueta dos dados no espaço original e dos dados projetados. A silhueta dos dados originais é um referencial para identificar se uma normalização ou projeção está proporcionando bons resultados. As tabelas com resultados do espaço original (com e sem normalização) e projetado (sem normalização) estão presentes no Apêndice A.

A Tabela 1 apresenta os valores de silhueta para toda a base de dados. As colunas destacam os valores para o espaço original e os valores referentes às projeções utilizando cada conjunto de características. Em ambos os casos as características utilizadas foram normalizadas.

Tabela 1 – Valores de silhueta considerando Ilhéus como **2 grupos**. Valores para o espaço original dos dados e para os conjuntos de características associadas com Projeções Multidimensionais. Em ambos os casos as características foram normalizadas (*Min-max*). Os melhores valores para cada linha de características x projeção estão destacados. O Autoencoder *completo* possui todas as características geradas por ele (23.040 valores) e o *sumário* possui uma sumarização dessas características (160 valores).

	Original Norm.	Force	LAMP	LSP	MDS	PCA	t-SNE
<i>Índices e Métricas</i>	0,12	0,0545	0,08	0,1234	0,1313	0,1313	0,1214
<i>Descritores de imagem</i>	0,09	0,0866	0,0808	0,111	0,0965	0,0965	0,0451
<i>Autoencoder Completo</i>	0,02	0,0679	0,0615	0,0697	0,0691	0,0691	0,0192
<i>Autoencoder Sumário</i>	0,01	0,0764	0,0749	0,0746	0,0577	0,0577	-0,0172

Fonte: Dados da pesquisa.

Ao analisar esses dados é possível identificar qual combinação de atributos e projeção possui melhor valor de silhueta e consequentemente pode representar as segregações dos dados. Inicialmente é possível perceber que as características acústicas retornam os melhores valores de silhueta. Entretanto, as diferenças entre os valores não é significativa, o que gera a necessidade de analisar visualmente os resultados das projeções. Observou-se que, mesmo que a t-SNE não possua os melhores valores de silhueta, essa projeção consegue gerar melhor segregação visual dos dados (ex.: Figura 38) com os atributos normalizados do que as demais projeções analisadas. Isso provavelmente acontece por causa de distorções das distâncias no espaço projetado, dentro de grupos específicos. Mesmo que a t-SNE tenha sido escolhida para visualizar os dados por causa dessas questões apresentadas, cada projeção é capaz de destacar diferentes aspectos do conjunto original dos dados.

Figura 38 – Apresentação dos conjuntos de características normalizadas associadas com Projeções Multi-dimensionais. As cores representam as áreas de coleta, Costa Rica 1 e 2 (vermelho e preto), Ilhéus 1 e 2 (azul escuro e verde) e Laje de Santos (azul claro).

A Figura 38 apresenta visualmente as projeções dos dados da Tabela 1. Nas imagens é possível perceber a separação entre as áreas de Costa Rica, Ilhéus e Laje. Também pode ser percebido que as áreas CostaRica1 e CostaRica2 podem ser segregadas com vários níveis de precisão, principalmente quando as características acústicas são empregadas com a t-SNE.

Tanto nesses testes quanto nos próximos, é possível perceber que os dados rotulados como Ilheus1 e Ilheus2 não são distintos. Essa percepção foi confirmada pelos pesquisadores que cederam os dados. A partir disso, os valores de silhueta foram recalculados utilizando Ilhéus como apenas um rótulo. Os resultados utilizando 4 rótulos para os dados são descritos na Tabela 2 e destacam melhorias quando comparados com os anteriores. Com isso foram verificadas melhorias nos valores de silhueta e diferenças nos melhores resultados para cada projeção.

Tabela 2 – Valores de silhueta considerando Ilhéus como **um grupo**. Valores para o espaço original dos dados e para os conjuntos de características associadas com Projeções Multidimensionais. Em ambos os casos as características foram normalizadas (*Min-max*). Os melhores valores para cada linha de características x projeção estão destacados. O Autoencoder *completo* possui todas as características geradas por ele (23.040 valores) e o *sumário* possui uma sumarização dessas características (160 valores).

	Original Norm.	Force	LAMP	LSP	MDS	PCA	t-SNE
<i>Índices e Métricas</i>	0.20	0.1064	0.1571	0.2011	0.2181	0.2181	0.1953
<i>Descritores de Imagem</i>	0.10	0.1117	0.1341	0.1405	0.1305	0.1305	0.1103
<i>Autoencoder Completo</i>	0.08	0.0985	0.0895	0.0979	0.0965	0.0965	0.0631
<i>Autoencoder Sumário</i>	0.08	0.0998	0.0983	0.0982	0.0815	0.0815	0.0259

Fonte: Dados da pesquisa.

5.2.2.1 Segmentação de Imagens

A partir da abordagem definida na Seção 4.1, os testes para descrição dos espectrogramas foram iniciados com a segmentação dessas imagens, para destacar regiões delas como pode ser visto nas imagens do Apêndice B. Procurou-se aqui utilizar uma abordagem que fosse o mais genérica possível, diferente do que foi proposto por pesquisas como as descritas na Seção 3.4. Isso ocorreu porque as bases utilizadas para testes possuem uma grande diversidade de sons, principalmente as bases de sons terrestres. Para alcançar a segmentação apresentada, foram aplicadas a separação de componentes do áudio e técnicas clássicas para segmentação, citadas na Seção 2.4, como: aplicação de um limiar na imagem (valor = 70), utilização da operação abertura da Morfologia Matemática e detecção de componentes conexos.

Embora esse primeiro teste tenha alcançado resultados relevantes na localização de estruturas de interesse, o restante do trabalho envolveu experimentos dedicados a destacar características gerais do ambiente, no lugar de identificar atributos específicos dos áudios, conforme relatam os demais experimentos.

5.2.2.2 Índices e Medidas Acústicas

Em sequência aos testes com segmentação, com base nas questões discutidas na Seção 3.1 a respeito da eficácia dos Índices Acústicos, foram utilizados para os testes as seguintes características: ACI, ADI, AEI, AR, BIO, H, Hf, Ht, M, NDSI, SPL, RMS, *Roughness*, *Rugosity*, *Number of Peaks* e os 12 primeiros coeficientes do MFCC, descritas no Capítulo 2.

Para obtenção desses atributos, foram utilizados os parâmetros padrões definidos nas bibliotecas utilizadas, listadas na Seção 4.5. Foi necessário converter os arquivos no formato FLAC de Costa Rica para o formato WAV, para que as bibliotecas do R pudessem extrair as características acústicas. Os valores extraídos consideraram todo o tempo de gravação contido em cada arquivo de áudio. O resultado do MFCC é uma matriz onde as colunas representam cada um dos coeficientes calculados e as linhas seus componentes. Com isso, foram utilizadas as médias das colunas para representar cada coeficiente utilizado. Para extração de todas essas características foi necessário aproximadamente 1 minuto de processamento para cada minuto de áudio. Assim sendo, por exemplo, para processar a área CostaRica2 que possui 1.815 arquivos, algo em torno de 5.883 minutos, foram necessários aproximadamente 4 dias de processamento.

A Figura 39 apresenta o resultado da projeção dos dados descritos pelas características acústicas. É perceptível a separação entre as áreas CostaRica1 e CostaRica2, Ilhéus e a Laje de Santos. Mesmo assim, as áreas de Ilhéus1 e Ilhéus2 não ficam separadas da mesma maneira que ocorre com as de Costa Rica.

Figura 39 – Projeção t-SNE de 4.340 áudios utilizando 27 índices e medidas acústicas como características. As cores dos pontos representam as áreas de coleta dos dados.

Fonte: Elaborada pelo autor.

A Figura 40 apresenta as bases separadamente. As áreas CostaRica1 e CostaRica2 ficam novamente separadas uma da outra. A base de Ilhéus, mesmo possuindo dois conjuntos, não apresenta separação clara. A base da Laje de Santos apresenta formação de um grupo destacado em vermelho com pouco ou nenhum som nos arquivos de áudio.

Enquanto isso, as Figuras 41 e 42 apresentam testes nas bases de dados utilizando as características divididas em três grupos: Índices Acústicos (ACI, ADI, AEI, BIO, Hf, Ht, H, NDSI, M, AR), MFCC (12 coeficientes), e outras medidas (SPL, RMS, Roughness, Rugosity, Number of Peaks). Com isso, é possível perceber que o MFCC consegue distinguir as áreas CostaRica1 e CostaRica2. Também se pôde verificar que dados de Ilhéus e Laje de Santos ficam misturados quando estão sendo utilizadas as outras medidas. Na Figura 42, para a Laje de Santos,

grupos ficam destacados (não necessariamente contendo os mesmos itens) com qualquer um dos conjuntos de características. Esses grupos possuem áudios com pouco ou nenhum som.

Figura 40 – Projeção t-SNE de cada uma das bases de dados separadamente, utilizando 27 índices e medidas acústicas como características. O grupo destacado na Laje de Santos apresenta áudios com pouco ou nenhum som.

Fonte: Elaborada pelo autor.

Figura 41 – Projeção t-SNE das bases utilizando as características acústicas divididas em: Índices Acústicos, MFCC e Outras Medidas Acústicas.

Fonte: Elaborada pelo autor.

Figura 42 – Projeção t-SNE das bases separadas utilizando conjuntos distintos de características acústicas. Os grupos destacados na Laje de Santos apresentam áudios com pouco ou nenhum som.

Fonte: Elaborada pelo autor.

5.2.2.3 Descritores de Imagem

No lugar de segmentar a imagem, como no primeiro teste, foi utilizada uma abordagem global, extraindo características de toda a imagem do espectrograma. Foram utilizados inicialmente para os testes os seguintes descritores: GLCM, LBP, Fourier para textura, GCH, ACC, CCV, BIC, descritos na Seção 2.5. Os parâmetros dos algoritmos, listados na Seção 2.5, são descritos a seguir.

Para execução do LBP foi utilizado $P = 8$. Para obter os valores do descritor GLCM, foram calculados os valores de contraste, segundo momento angular, correlação e entropia para cada uma das matrizes de direção. Por fim, a média das estatísticas das matrizes foram utilizadas. O descritor de Fourier utilizou apenas os valores de $S(r)$. Foram considerados 10 valores para o raio r . Também foram utilizados os valores externos ao último raio $S(r_{10})$, posicionados nas extremidades de cada imagem (superior esquerdo e direito, e inferior esquerdo e direito).

Os testes com o descritor WLD utilizaram valores pares nos intervalos de 2 a 10 (parâmetros T e M), e de 2 a 20 (parâmetro S). Todas as 250 combinações desses valores foram analisadas tanto visualmente (Projeção Multidimensional) quanto numericamente (valor da silhueta). Apenas os resultados com os valores $T = 2, M = 2, S = 2$; $T = 2, M = 2, S = 4$ e $T = 8, M = 4, S = 4$ geraram visualizações e silhuetas distintas.

Para os descritores de cor, o ACC utilizou os valores de distância 1, 3, 5 e 7, descritos no artigo original. No CCV o limiar foi definido como 100. E no BIC foi utilizada a vizinhança-8.

Para extração dos descritores foi necessário aproximadamente 2 segundos de processamento para cada imagem de espectrograma. Assim sendo, por exemplo, para processar a área CostaRica2 que possui 1.815 arquivos, foi necessário aproximadamente 1 hora de processamento.

A Figura 43 apresenta o resultado da projeção dos dados com esses atributos. É possível perceber a separação entre Costa Rica, Ilhéus e Laje de Santos. Entretanto, a separação entre as duas áreas CostaRica1 e CostaRica2 não possui a mesma nitidez alcançada com as medidas acústicas.

Figura 43 – Projeção t-SNE de 4.340 áudios utilizando 600 valores de descritores de textura e cor.

Fonte: Elaborada pelo autor.

A Figura 44 apresenta cada uma das bases separadamente. Tanto CostaRica1 e CostaRica2 quanto Ilhéus1 e Ilhéus2 não apresentam separação. A base da Laje de Santos apresenta formação de um grupo destacado em vermelho com pouco ou nenhum som nos arquivos de áudio.

Figura 44 – Projeção t-SNE de cada uma das bases de dados separadamente, utilizando 600 valores de características geradas pelos descritores de textura e cor. O grupo destacado na Laje de Santos apresenta áudios com pouco ou nenhum som.

Fonte: Elaborada pelo autor.

As Figuras 45 e 46 apresentam testes utilizando as características divididas em dois grupos: textura (GLCM, LBP e Fourier) e cor (GCH, ACC, CCV e BIC). Com isso, é possível perceber a separação entre as áreas terrestres e aquáticas e perceber que os descritores de cor conseguem distinguir as áreas CostaRica1 e CostaRica2 melhor do que os de textura. Entretanto, as áreas Ilhéus1 e Ilhéus2 não são separadas por nenhum dos grupos de características. Na Figura 46, novamente um grupo dentro dos dados da Laje é destacado. Esse grupo também representa dados com pouco ou nenhuma intensidade sonora.

Por fim, foram adicionados testes com o descritor de textura WLD. Os testes executados com esse descritor foram realizados apenas em uma amostra de 173 espectrogramas (cf. Seção 5.2.2.4) para verificar se eles melhorariam a descrição das imagens, principalmente para os dados de Costa Rica. Para cada combinação de parâmetros a implementação do descritor demorou 27,62 segundos para executar, precisando no total de aproximadamente 14 dias para extrair todas as combinações das 173 imagens. O melhor resultado para esse descritor foi obtido com a combinação dele com os descritores LBP e ACC (Figura 47), com os valores de parâmetro $T = 2$,

Figura 45 – Projeção t-SNE das bases utilizando conjuntos distintos de características de imagem.

Fonte: Elaborada pelo autor.

$M = 2$, $S = 4$. Essa combinação projetou os dados terrestres e aquáticos separados. Entretanto, os dados de Costa Rica1 e Costa Rica2 não ficaram claramente segregados como ocorre com os descritores de cor.

Figura 46 – Projeção t-SNE das bases separadas utilizando conjuntos distintos de características de imagem. Os grupos destacados na Laje de Santos apresentam áudios com pouco ou nenhum som.

Fonte: Elaborada pelo autor.

5.2.2.4 Características de Imagem extraídas por Autoencoder

Como alternativa para escolher qual descritor utilizar, foi aplicado o Autoencoder para definir e extrair um conjunto significativo de características das imagens dos espectrogramas. A base de dados foi dividida em três grupos: treinamento (3.334 imagens), validação (833 imagens) e teste (173 imagens). As imagens para cada um dos conjuntos foram escolhidas de maneira aleatória, mas mantendo as proporções gerais dos dados. Em cada um dos três grupos existem aproximadamente 70% de imagens de Costa Rica, 11% de arquivos de Ilhéus e 18% da Laje de Santos.

Na etapa de **pré-processamento**, todas as imagens foram divididas em 20 subimagens de 222x170 (Figura 48) e depois reescaladas para 192x192. Após esse processo, a base de imagens possui: 66.680 imagens para treinamento, 16.660 para validação e 3.460 para teste.

Figura 47 – Projeção t-SNE de 173 imagens de teste utilizando os descritores WLD, LBP e ACC como características.

Fonte: Elaborada pelo autor.

No **processamento**, as imagens de treinamento e validação foram apresentadas para a estrutura de Autoencoder descrita na Figura 49. A função de perda utilizada foi a *Mean Squared Logarithmic Error* e o método de otimização para essa função foi o *Stochastic Gradient Descent* aplicando os parâmetros padrão das bibliotecas utilizadas (cf. Seção 4.5).

Os valores da função de perda, tanto para treinamento quanto para validação, iniciaram próximos de 0,01 e tenderam para 0 durante o processo. Foram utilizados, no processo de treinamento, valores de *batch* iguais a 160 tanto para o conjunto de treinamento quanto de validação, e 30 como valor de época. O treinamento do modelo do Autoencoder durou aproximadamente 1 hora e meia. O modelo gerado extrai 1.152 características para cada subimagem do espectrograma. Como cada espectrograma foi dividido em 20 subimagens, o espectrograma completo possui 23.040 características.

Figura 48 – Exemplo do pré-processamento de um espectrograma. O espectrograma (esquerda) é dividido em 20 imagens (direita).

Figura 49 – Arquitetura de Autoencoder utilizada.

Fonte: Elaborada pelo autor.

Na etapa de **pós-processamento**, foi verificado visualmente o quão próximo da entrada é da reconstrução realizada pelo Autoencoder. A Figura 50 apresenta a reconstrução (decodificação) da imagem, apresentada na Figura 48, realizada pelo Autoencoder. A reconstrução destaca os detalhes gerais das subimagens. As imagens rotuladas de s01 a s05 são as mais claras e as de s06 a s10 mais escuras que as anteriores. As imagens de s11 a s15 mantêm a informação na parte inferior, enquanto que as imagens de s16 a s20 na parte superior.

Vale ressaltar que as características de cada parte do espectrograma (1.152) podem ser sumarizadas com 8 medidas estatísticas: valor mínimo e máximo, média, desvio padrão, 1º, 2º e 3º quartil e a distância interquartil. Por causa disso, o espectrograma também foi representado por um sumário contendo 160 valores.

Figura 50 – Reconstrução das partes do espectrograma apresentado na Figura 48. Cada imagem está rotulada entre s01 e s20.

Após essas etapas, foram executados três testes com as características extraídas. O **primeiro teste** consistiu em verificar se as características geradas são capazes de diferenciar partes de uma mesma imagem. Para isso foram utilizadas as imagens da Figura 48 e o resultado de uma projeção *ForceScheme* dos dados gerados é apresentado na Figura 51. A *ForceScheme* foi aplicada aos primeiros testes, porque apresentou melhor separação dos dados quando comparada com a t-SNE. Apenas o último teste com Autoencoder foi executado com a t-SNE para manter a similaridade com os resultados obtidos com as demais técnicas de extração de características. É possível perceber com esse teste que cada uma das linhas de subimagens criam grupos bastante segregados.

Figura 51 – Projeção *ForceScheme* das partes (s01 até s20) do espectrograma da Figura 48. As imagens utilizam as 1.152 características extraídas (esquerda) e a sumarização dessas características (direita). Os grupos destacados em verde correspondem a cada uma das 4 linhas do espectrograma utilizado.

No **segundo teste** foi verificada a capacidade dessas características de separar espectrogramas distintos. Para isso, foram selecionadas visualmente 10 imagens que formam quatro grupos diferentes (Figura 52). Com as características extraídas, essas imagens foram projetadas com *ForceScheme* da forma apresentada na Figura 53. Os grupos destacam as imagens que são visualmente similares.

Por fim, no **terceiro teste** as imagens do conjunto de testes (173 imagens) foram projetadas com *t-SNE* conforme apresentado na Figura 54. Para aplicar o modelo sobre as imagens foram necessários cerca de 6 segundos de processamento. Se o mesmo processo for aplicado sobre todas as imagens (treinamento, validação e testes), 70.820 subimagens, a duração do processo é de 3 minutos. A imagem apresenta uma boa segregação para os dados da Laje, mas Costa Rica e Ilhéus não são claramente separados.

Figura 52 – Imagens nomeadas como CR08, CR11, CR10, CR09, CR1302, CR12, OP4, OP3, CR1306, CR06. Agrupadas por semelhanças visuais (3 primeiras linhas) ou por serem diferentes das demais (última linha).

Figura 53 – Projeção *ForceScheme* de 10 espectrogramas. As imagens utilizam as 23.040 características extraídas (esquerda) e a sumarização dessas características (direita). Os grupos destacados em verde correspondem às linhas da imagem 52, enquanto que as as imagens CR1306 e CR06 estão fora dos grupos por serem diferentes das demais.

5.2.2.5 Comparação e discussão sobre as características extraídas

Como pode ser observado nas seções anteriores, é possível segregar visualmente dados terrestres e dados subaquáticos, utilizando o conjunto correto de características. Isso ocorre porque áudios terrestres são mais ricos do que os dados subaquáticos (cf. Introdução), além de existirem diferenças nas propriedades do som em cada um desses ambientes (cf. Seção 2.1).

As características que melhor separaram as áreas CostaRica1 e CostaRica2 foram os

Figura 54 – Projeção t -SNE de 173 espectrogramas. As imagens apresentam a projeção utilizando 23.040 características extraídas (esquerda) e a sumarização dessas características (direita).

MFCC, que são empregados com frequência para reconhecimento de fala (cf. Seção 2.2.6), mas que nos testes executados mostraram-se eficazes para segregação de Paisagens Acústicas. Os descritores de cor, para as imagens dos espectrogramas, melhor separaram os dados de Costa Rica, mas não possuem melhor separação do que o MFCC. No final, as características geradas pelo Autoencoder não separaram os dados dessa região.

A Tabela 3 foi obtida tomando as imagens de teste utilizadas pelo Autoencoder, com todos os conjuntos de características utilizados (índices e medidas acústicas, descritores de imagem clássicos e gerados pelo Autoencoder). Para esses dados foram calculados os valores de silhueta dispostos na tabela. Os melhores valores foram encontrados quando os dois conjuntos Ilheus1 e Ilheus2 foram considerados como uma mesma classe. Isso ocorre provavelmente porque não existe diferença clara entre eles, algo já mencionado no início da descrição dos testes com índices. Os valores da tabela, para a amostra de testes (contendo 173 itens), seguem a mesma tendência apresentada nos testes anteriores, ou seja, índices e medidas acústicas têm melhor desempenho do que as demais características empregadas. Os melhores resultados na separação dos dados de Costa Rica foram obtidos utilizando os descritores de cor e principalmente com o MFCC. Por fim, a combinação de alguns descritores (WLD, LBP e ACC) proporcionou valor de silhueta próximo ao obtido com descritores de cor (4 classes) e próximo dos índices (5 classes), mas não possuem a mesma segregação visual.

Além dos testes apresentados acima, foram aplicadas técnicas de Seleção Automática de atributos cujos resultados estão descritos no Apêndice C. A avaliação das características selecionadas também foi realizada com o Coeficiente de Silhueta e com projeções. Mesmo os resultados de silhueta sendo ligeiramente melhores do que os apresentados durante esta seção,

Tabela 3 – Valores de silhueta para dados com conjuntos distintos de características. Os valores das características foram normalizados entre [0, 1]. Com **5 classes** os dados de Ilhéus são considerados como 2 grupos, como **4 classes** como 1 grupo.

Características	5 classes	4 classes
<i>Índices e medidas</i>	0,10	0,17
<i>Descritores</i>	0,06	0,07
<i>Autoencoder Completo</i>	-0,06	0,00
<i>MFCC</i>	0,12	0,21
<i>Descritores de Cor</i>	0,06	0,07
<i>WLD + LBP + ACC</i>	0,10	0,09

Fonte: Dados da pesquisa.

os resultados das projeções não apresentaram melhorias significativas.

Ressalta-se também a redução do espaço em disco que pode ser alcançada com a extração das características dos dados acústicos. A base de dados utilizada possui 70 *gigabytes* de áudios. Um arquivo contendo todas as características acústicas extraídas da base de dados possui aproximadamente 2 *megabytes*, com os descritores de imagem 6,8 *megabytes*, com as características geradas pelo Autoencoder 1,94 *gigabytes* (completo) e 14 *megabytes* (sumarizado).

No que diz respeito aos tempos de processamento reportados nos testes, as características acústicas alcançaram os piores resultados, necessitando de dias para o processamento. Os descritores de imagem precisam de cerca de 1 hora para extrair as características. Entretanto os testes com o WLD necessitou de uma dezena de dias para ser finalizado. O Autoencoder precisa de alguns minutos para extrair as características com um modelo gerado em 2 horas. É importante ressaltar que os testes com o Autoencoder foram executados em uma placa gráfica e os demais testes em um processador de propósito geral. Mesmo que os dados acústicos, que obtiveram os melhores resultados, empreguem uma quantidade significativa de tempo para o processamento, as abordagens aplicadas nesta pesquisa não visam um monitoramento em tempo real das paisagens. Assim sendo, o tempo de extração dos atributos não influencia a análise dos dados.

Com base nesses resultados, as análises apresentadas nas próximas seções foram realizadas utilizando todos os valores de índices e as medidas acústicas discutidos. Entretanto, mesmo que os testes com descrição de imagens não tenham sido melhores nem iguais aos resultados obtidos com aquelas características (índices e medidas), eles não apresentaram resultados completamente equivocados e possuem tempo de processamento inferior ao dos dados acústicos. Ainda é possível buscar outras abordagens para melhorar a utilização dessas características. Por exemplo, pode existir uma melhor combinação de descritores que descrevam as informações das imagens. Além disso, é possível ajustar a arquitetura do Autoencoder para encontrar uma melhor representação dos dados ou utilizar outros modelos de *DeepLearning*. Também é possível

aplicar outras técnicas de seleção de atributos tanto automáticas quanto visuais para encontrar as características que melhor descrevem os dados.

5.2.3 *Análise Visual*

A partir dos resultados obtidos na Seção 5.2.2, análises adicionais foram realizadas utilizando os Índices e as Medidas Acústicas para representar os dados. As próximas seções apresentam alguns exemplos de resultados de explorações e análises realizadas, seguindo a abordagem descrita na Seção 4.3.

5.2.3.1 *Coordenadas Paralelas*

As Coordenadas Paralelas foram utilizadas para analisar a disposição dos valores e as relações entre os índices e medidas acústicas das áreas de coleta CostaRica1 e CostaRica2, como apresentado na Figura 55. Foi possível verificar, por exemplo, que o menor valor do ACI corresponde a um arquivo de áudio com “falhas na gravação”. Verificou-se também, que o arquivo correspondente ao maior valor do AEI possui padrões provavelmente pertencentes a pássaros destacados em azul. Outra observação se refere ao comportamento dos índices de diversidade. ADI, H, ACI, BIO e AR, mesmo sendo utilizados para descrever a mesma propriedade do ambiente, possuem comportamentos distintos nesse cenário.

5.2.3.2 *Mapa de Calor*

A Figura 56 apresenta os valores do SPL para a área CostaRica1. Essa área foi utilizada por possuir mais amostras de áudio. O SPL apresentou sensibilidade aos áudios noturnos, como destacado em azul. Alguns exemplos de espectrogramas expressam: que os valores altos do SPL ocorrem tanto pela chuva quanto pela intensidade de sons contínuos de animais; e que valores baixos podem ocorrer quando existe maior riqueza de sons.

5.2.3.3 *HiPP estendida*

Nesta seção a xHiPP foi aplicada para descobrir padrões dentro dos dados. Na Figura 57, os dados de Costa Rica foram caracterizados utilizando o MFCC. Mesmo com a boa separação apresentada para as duas regiões, alguns pontos ainda ficam misturados. Os espectrogramas destacados na parte inferior esquerda da imagem apresentam um grupo que possui a maior parte de áudios com sons de chuva. Mesmo que algum outro som exista (como uma das imagens apresenta delimitado em azul), em geral esses áudios são similares. O grupo apontado do lado direito da projeção apresenta áudios do período diurno, com padrões sonoros de pássaros e insetos (destacados em azul). No topo, os espectrogramas representam áudios noturnos com padrões de sons de insetos.

Figura 55 – Coordenadas Paralelas de índices e medidas acústicas de Costa Rica. As cores das curvas representam as áreas CostaRica1 e CostaRica2. São destacados no retângulo tracejado eixos que quantificam diversidade. No eixo AEI é destacado um áudio com valor alto do índice cujo espectrograma possui padrões de pássaros destacados em azul. No eixo ACI é destacado um valor baixo referente ao áudio cujo espectrograma apresenta falhas da gravação.

Na Figura 58 são apresentados os mesmos dados da Figura 57, com as mesmas características, mas rotulados com o período do dia (diurno e noturno). Na parte inferior da imagem, um espectrograma noturno (19:30 h) foi destacado. O padrão apresentado para esse espectrograma, e para boa parte dos dados do mesmo grupo onde ele está, pode ser de insetos ou de anfíbios que aparecem durante a noite. O espectrograma apresentado do lado esquerdo da projeção representa um áudio diurno (14 h) com uma área destacada referente a um padrão de sons de cigarras. Enquanto isso, as duas imagens da direita (05:15 h e 11 h) destacam padrões de pássaros e outro padrão constante. Esse padrão constante (que pode ser de insetos ou anfíbios) ocorre durante todo o áudio em torno de 10 kHz.

A Figura 59 apresenta dados de Ilhéus, sem rotulação, possuindo MFCC como atributos. Se no lugar do MFCC forem utilizados os índices com valores normalizados (*Min-max*), os resultados assemelham-se a esses. No lado direito da imagem dois espectrogramas de grupos separados, mas próximos, são destacados. A maior parte das imagens desses dois grupos possuem o mesmo padrão de cor de peixes que está delimitado. Ao mesmo tempo, a imagem apresentada à esquerda da projeção identifica um padrão sonoro de crustáceos que é predominante no grupo desse áudio.

Por fim, a Figura 60 apresenta a projeção de dados da Laje de Santos, representados por

Figura 56 – Mapa de Calor do SPL da área Costa Rica1. Destaque, delimitado pelo retângulo azul, para os valores noturnos. Os espectrogramas apresentados no topo indicam respectivamente valor baixo do SPL para áudio com som predominante de animais, e valor alto do SPL para áudio com chuva predominante. O espectrograma da parte inferior destaca valor alto do SPL para som constante de cigarra.

Índices Acústicos normalizados com *Min-max*. Os espectrogramas do lado inferior direito da imagem destacam padrões sonoros de embarcações. A imagem posicionada no topo do lado direito apresenta padrões de crustáceos. As demais imagens possuem sons de peixes (esquerda inferior) e de peixes e crustáceos (esquerda superior).

5.2.3.4 Discussão

Existem diversos trabalhos na literatura que utilizam os Índices Acústicos, como é constatado na Seção 3.1. Entretanto é possível verificar que, como mencionado anteriormente, existe uma quantidade significativa de controvérsias sobre esses atributos. Isso leva alguns pesquisadores a não cogitarem a utilização desses índices (cf. Seção 3.2 e o trabalho de [Sánchez-Gendriz e Padovese \(2017\)](#) discutido na Seção 3.7). Mesmo assim, esses autores não buscam avaliar quais atributos poderiam se adaptar melhor às suas pesquisas. Assim sendo, uma abordagem válida é avaliar qual conjunto de características melhor descreve o ambiente em questão, efetuando comparações e análises dos resultados da utilização de cada tipo de atributo. Esse é um dos motivos para iniciar a abordagem apresentada nesta pesquisa de mestrado com a avaliação de um conjunto pertinente de características.

Um ponto a ser destacado sobre as visualizações da abordagem descrita acima, é que elas

Figura 57 – Dados de Costa Rica utilizando MFCC como características (sem normalização), projetados com a xHiPP (t-SNE + kmeans). As cores dos pontos identificam as regiões de coleta em Costa Rica. Alguns pontos são destacados apresentando seus respectivos espectrogramas e algumas áreas dessas imagens são delimitadas em azul para identificar padrões existentes.

não dependem, por exemplo, da quantidade de canais de cores para expressar um conjunto de dados, como fazem os trabalhos baseados na pesquisa de [Towsey et al. \(2014b\)](#), citados na Seção 3.7. Além disso, a utilização de projeções não força que o analista precise possuir conhecimento prévio sobre o significado de cada uma das características utilizadas.

Também é interessante perceber a necessidade de separação e filtragem de sinais, algo que não é realizado em trabalhos, como os descritos nas Seções 3.6 que utilizam a biblioteca REAL. Essa separação/filtragem pode ser realizada através da segmentação dos dados (cf. Seções 3.3 e 3.4), ou da Separação de Fontes. Essas são fases que ainda precisam ser adicionadas à abordagem da pesquisa desse mestrado. O trabalho com Separação de Fontes está sendo realizado em paralelo.

Com os exemplos apresentados durante as seções anteriores é possível visualizar o quanto as técnicas utilizadas são capazes de auxiliar na exploração e análise de dados acústicos. A partir de uma melhor escolha dos atributos dos áudios, as visualizações produzem um caminho eficiente para análise, facilitando a localização de padrões gerais e específicos dos dados. O agrupamento de informações, como as apresentadas pelo Mapa de Calor, facilita a verificação do comportamento de características específicas do som. O destaque de áudios com características em comum melhora a busca por sons particulares, permitindo que o analista perceba relações

Figura 58 – Dados de Costa Rica utilizando MFCC como características (sem normalização), projetados com a xHiPP (t-SNE + kmeans). As cores dos pontos identificam os períodos diurno (amarelo) e noturno (marrom). As áreas delimitadas em azul nos espectrogramas destacam padrões dos sons.

Figura 59 – Dados de Ilhéus utilizando MFCC como características (sem normalização), projetados com a xHiPP (t-SNE + kmeans). As áreas delimitadas em azul nos espectrogramas destacam padrões dos sons.

entre os áudios sem previamente escutar a base de dados completa. Utilizando as Coordenadas Paralelas, é possível comparar o comportamento das características de áreas diferentes, conseguindo identificar as peculiaridades de cada região analisada. A aplicação da xHiPP possibilita a

Figura 60 – Dados da Laje de Santos utilizando Índices Acústicos como características (normalização *Min-max*), projetados com a xHiPP (t-SNE + kmeans). As áreas delimitadas em azul nos espectrogramas destacam padrões dos sons.

identificação de padrões tanto gerais quanto locais dos dados através do agrupamento por suas similaridades. Isso possibilita ao usuário filtrar os dados e concentrar-se em padrões relevantes.

Claro que, mesmo possuindo pontos que aumentam a capacidade de análise, as técnicas apresentadas também possuem limitações. A quantidade de características analisadas nas Coordenadas Paralelas depende do espaço em tela, além de que, muitas características (por exemplo 30) podem causar confusão visual que atrapalhe o entendimento dos dados. Para os Mapas de Calor, se as variações entre os valores forem pequenas, a distinção entre os níveis das cores utilizadas pode ser difícil. Enquanto isso, a xHiPP introduz distorções nos dados, por precisar recorrer a operações que mapeiam e aproximam o espaço original para o espaço 2D, dificuldade enfrentada pelas técnicas de projeção.

Como já mencionado o tempo e processamento das características, principalmente das características acústicas, é alto (1 minuto de processamento para cada 1 minuto de áudio) na máquina utilizada para processamento. Entretanto, como as análises não são feitas em tempo real, isso não é um problema relevante no momento.

Apesar das limitações apresentadas, as técnicas aqui propostas, testadas e avaliadas, adicionaram à análise de Paisagens Acústicas a habilidade de separar e localizar padrões de forma eficaz. Padrões esses não observados anteriormente pelas abordagens existentes, como as descritas no Capítulo 3.

CONCLUSÕES

Nesta dissertação foi apresentada uma abordagem visual para exploração e análise de Paisagens Acústicas, capaz de destacar tanto semelhanças quanto diferenças entre ambientes naturais distintos. Para isso foram analisados conjuntos de características que podem ser empregadas para descrever essas paisagens, além de uma abordagem para avaliação desses atributos. Ainda, foram destacadas técnicas utilizáveis para investigação dos dados, a partir das características examinadas. Dentre essas técnicas, uma Projeção Multidimensional foi estendida para melhorar sua aplicação na investigação dos dados. A fim de avaliar a utilização do processo, foram utilizadas bases de dados específicas de sons terrestres e aquáticos descritas na Seção 5.2.1.

Os resultados e discussões apresentadas no Capítulo 5 reforçaram a aplicabilidade da proposta do Capítulo 4, mesmo ela possuindo limitações como as descritas posteriormente.

6.1 Contribuições

Retomando o que já foi exposto no capítulo introdutório, as contribuições desta pesquisa podem ser apresentadas como:

- Implementação, avaliação e aplicação da xHiPP, uma extensão da Projeção Multidimensional HiPP. A partir das modificações aplicadas é possível encontrar uma melhor representação dos dados, associando algoritmos diversos de agrupamento e projeção, além da possibilidade de alternar as etapas agrupamento-projeção e projeção-agrupamento. Essas melhorias associadas à capacidade da xHiPP de trabalhar com níveis variados dentro dos dados auxilia a filtragem de volumes grandes de dados, permitindo que o usuário trabalhe tanto com características globais quanto locais;
- Definição de um processo que aplica Projeções Multidimensionais e Coeficiente de Silhueta

para escolha e validação de características relevantes para representação de dados acústicos. Esse processo permite analisar qual conjunto de características melhor descreve uma paisagem. Com isso, é viável avaliar a utilização, não só de atributos extraídos dos áudios, mas de atributos de domínios correlacionados, como, por exemplo, as imagens dos espectrogramas do áudio;

- A obtenção de evidências de que o MFCC é um conjunto de atributos eficaz para representação e segregação de Paisagens Acústicas;
- Diminuição do espaço de armazenamento necessário a partir da extração das características utilizadas;
- Aplicação de técnicas visuais, como Mapa de Calor, Coordenadas Paralelas e xHiPP para exploração de dados acústicos. Essas técnicas são capazes de destacar atributos semelhantes dos áudios, facilitando a análise, permitindo que o usuário dê enfoque às características relevantes como: falhas de gravação, presença de embarcações, presença de padrões sonoros de espécies animais etc;
- Capacidade das visualizações empregadas de salientar tanto características globais quanto específicas dos dados. Mesmo que a abordagem tenha partido da idéia de evidenciar diferenças globais entre Paisagens Acústicas distintas, os resultados apontaram a capacidade de destacar características específicas dos áudios. A aplicação da xHiPP apresenta não somente as diferenças entre regiões de coletas distintas, mas também particularidades dos grupos apresentados. Essas especificidades permitem então explorar diversos graus de detalhamento, auxiliando a exploração dos dados.
- Elaboração dos seguintes artigos:
 - DIAS, F. F.; PEDRINI, H.; MINGHIM, R. A Visual Analysis Strategy for Soundscapes based on Discriminative Features. **Ecological Informatics**. (*submetido*)
 - DIAS, F. F.; MINGHIM, R. eXtend HiPP. (*sendo finalizado*)

6.2 Limitações

Mesmo que os resultados alcançados durante esta pesquisa sejam favoráveis, tornando a abordagem proposta efetiva para exploração e análise de dados de Paisagens Acústicas, algumas limitações precisam ser destacadas.

O sucesso da exploração dos dados é dependente das técnicas de extração de características, das técnicas de agrupamento e de visualização dos dados. A junção dessas técnicas por um lado melhora a capacidade de análise, mas por outro aumenta a complexidade do processo.

Outras questões, como as discutidas na Seção 5.2.3.4, podem ser levantadas também: a limitação das Coordenadas Paralelas para trabalhar com grandes quantidades de atributos (mais que 30 por exemplo); a dificuldade do Mapa de Calor em representar pequenas variações entre as cores (representação dos valores); e as distorções no espaço que podem ser geradas pela xHiPP.

Por fim, a eficiência do processo pode ser questionada. Analisando os tempos de processamento apresentados na Seção 5.2, fica evidente a lentidão principalmente da extração de características acústicas e do processamento de descritores como o WLD. Claro que também precisam ser consideradas as diferenças das máquinas utilizadas para os processamentos do Autoencoder e das demais características. A utilização da linguagem R para extração das características do áudio e para geração dos dados das projeções é questionável. Também é um gargalo a visualização utilizando a biblioteca D3.js. Nas versões atuais, ela possui dificuldade para renderizar quantidades de dados que ultrapassam os milhares.

6.3 Trabalhos Futuros

Como consequência das contribuições e das limitações deste trabalho, é possível listar trabalhos que poderão ser realizados.

- **Processamento de áudio:**

1. Separação de fontes, segmentação de áudio e técnicas de filtragem;
2. Aplicação de técnicas de Aprendizagem de Máquina, como o Autoencoder, redes *Long Short Term Memory*, redes *Adversarially Learned Inference* etc.;
3. Utilização de abordagens mais eficientes para extração das características;

- **Processamento de imagem:**

1. Melhoria do processo de segmentação;
2. Aprimoramento dos descritores de imagem utilizando abordagens locais e melhoria de parâmetros (ex.: parâmetros do WLD);
3. Utilização de outros descritores de imagem, como: *Histogram of Oriented Gradients*, descritores baseados nos filtros de Gabor etc.;
4. Melhoria da arquitetura do Autoencoder aplicada à extração de características;

- **Visualização de dados:**

1. Aplicação de outras visualizações que possam se adequar melhor às necessidades dos dados;
2. Desenvolvimento de ferramenta que integre as visualizações utilizadas;
3. Melhoria da descrição dos itens contidos em cada grupo da xHiPP;

4. Utilização de ferramentas que sejam mais eficientes do que a biblioteca D3.js;
- **Seleção de atributos:** Estudo e utilização de técnicas visuais ou automáticas para seleção de atributos;
 - **Classificação de dados:** Abordagem visual para identificação e classificação de eventos acústicos específicos;
 - **Descrição de ambiente:** Adição de domínios de atributos que possam enriquecer a descrição do ambiente, como dados meteorológicos;
 - **Busca por similaridade:** Criação de busca de áudios por similaridade, utilizando indexação dos seus atributos.

REFERÊNCIAS

- Aalborg University. **The mel frequency scale and coefficients**. 2004. Disponível em: <http://kom.aau.dk/group/04gr742/pdf/MFCC_worksheet.pdf>. Acesso em: 09/12/2017. Citado na página 39.
- ABYOTO, K. W.; WIRDJOSOEDIRDJO, S. J.; WATANABE, T. **Unsupervised Texture Segmentation Using Multiresolution Analysis for Feature Extraction**. 1998. Disponível em: <<http://www.iic.tuis.ac.jp/edoc/journal/ron/r2-1-2/r2-1-2f.html>>. Acesso em: 10/12/2017. Citado na página 48.
- BEZDEK, J. C. **Pattern recognition with fuzzy objective function algorithms**. [S.l.]: New York: Plenum Press, 1981. Citado na página 70.
- BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent dirichlet allocation. **Journal of machine Learning research**, v. 3, n. Jan, p. 993–1022, 2003. Citado na página 81.
- 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**, Wiley Online Library, v. 17, n. 8, p. 2137–2144, 2007. Citado nas páginas 43 e 62.
- BRIGGS, F.; LAKSHMINARAYANAN, B.; NEAL, L.; FERN, X. Z.; RAICH, R.; HADLEY, S. J. K.; HADLEY, A. S.; BETTS, M. G. Acoustic classification of multiple simultaneous bird species: a multi-instance multi-label approach. **The Journal of the Acoustical Society of America**, ASA, v. 131, n. 6, p. 4640–4650, 2012. Citado na página 64.
- CHEN, J.; SHAN, S.; HE, C.; ZHAO, G.; PIETIKAINEN, M.; CHEN, X.; GAO, W. Wld: A robust local image descriptor. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 32, n. 9, p. 1705–1720, 2010. Citado na página 48.
- CHI, E. H. **A framework for Visualization Information**. [S.l.]: Springer, 2002. Citado na página 53.
- 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**, Elsevier, v. 13, n. 1, p. 46–54, 2012. Citado nas páginas 30, 44 e 62.
- DERAT, A. . 2017. Disponível em: <<https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798>>. Acesso em: 14/12/2017. Citado na página 51.
- DONG, X.; TOWSEY, M.; ZHANG, J.; ROE, P. Compact features for birdcall retrieval from environmental acoustic recordings. In: IEEE. **Proceedings of the 2015 IEEE 15th International Conference on Data Mining Workshops**. [S.l.]: IEEE Computer Society, 2015. p. 1–6. Citado na página 64.
- ELDRIDGE, A.; CASEY, M.; MOSCOSO, P.; PECK, M. A new method for ecoacoustics? Toward the extraction and evaluation of ecologically-meaningful soundscape components using sparse coding methods. **PeerJ**, PeerJ Inc., v. 4, p. e2108, 2016. Citado nas páginas 61 e 62.

- FLETCHER, N. Animal Bioacoustics. In: **Springer Handbook of Acoustics**. [S.l.]: Springer, 2007. p. 473–491. Citado na página 30.
- FREEMAN, S. E.; ROHWER, F. L.; D’SPAIN, G. L.; FRIEDLANDER, A. M.; GREGG, A. K.; SANDIN, S. a.; BUCKINGHAM, M. J. The origins of ambient biological sound from coral reef ecosystems in the Line Islands archipelago. **The Journal of the Acoustical Society of America**, ASA, v. 135, n. 4, p. 1775–1788, 2014. Citado na página 61.
- FULLER, S.; AXEL, A. C.; TUCKER, D.; GAGE, S. H. Connecting soundscape to landscape: Which acoustic index best describes landscape configuration? **Ecological Indicators**, Elsevier, v. 58, n. February 2016, p. 207–215, 2015. Citado nas páginas 60 e 62.
- GAGE, S. H. **Welcome to REAL**. 2008. Disponível em: <http://www.real.msu.edu/projects/one_proj.php?proj=TEST&page=overview#Navigation>. Acesso em: 02/01/2017. Citado nas páginas 29 e 65.
- GAGE, S. H.; AXEL, A. C. Visualization of temporal change in soundscape power of a Michigan lake habitat over a 4-year period. **Ecological Informatics**, Elsevier, v. 21, p. 100–109, 2014. Citado nas páginas 66 e 67.
- GASC, A.; PAVOINE, S.; LELLOUCH, L.; GRANDCOLAS, P.; SUEUR, J. Acoustic indices for biodiversity assessments: Analyses of bias based on simulated bird assemblages and recommendations for field surveys. **Biological Conservation**, v. 191, p. 306–312, 2015. ISSN 00063207. Citado nas páginas 60 e 62.
- GASC, A.; SUEUR, J.; JIGUET, F.; DEVICTOR, V.; GRANDCOLAS, P.; BURROW, C.; DE-PRAETERE, M.; PAVOINE, S. Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? **Ecological Indicators**, Elsevier, v. 25, p. 279–287, 2013. Citado na página 59.
- GINI, C. **Variabilità e Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche**. [S.l.: s.n.], 1912. Citado na página 41.
- GONZALEZ, R. C.; WOODS, R. E. **Processamento digital de imagens. tradução: Cristina Yamagami e leonardo piamonte**. [S.l.]: Pearson Prentice Hall, São Paulo, 2010. Citado nas páginas 36, 45, 46 e 48.
- GULLI, A.; PAL, S. **Deep Learning with Keras**. [S.l.]: Packt Publishing, 2017. ISBN 9781787129030. Citado na página 51.
- HALLIDAY, D.; RESNICK, R.; WALKER, J. **Fundamentos de física, volume 2: gravitação, ondas e termodinâmica**. [S.l.]: Rio de Janeiro: LTC, 2009. Citado nas páginas 35 e 36.
- HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural features for image classification. **IEEE Transactions on systems, man, and cybernetics**, IEEE, n. 6, p. 610–621, 1973. Citado na página 46.
- HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of educational psychology**, Warwick & York, v. 24, n. 6, p. 417, 1933. Citado na página 51.
- HUANG, J.; KUMAR, S. R.; MITRA, M.; ZHU, W.-J.; ZABIH, R. Image indexing using color correlograms. In: IEEE. **Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on**. [S.l.], 1997. p. 762–768. Citado na página 50.

- JOIA, P.; COIMBRA, D.; CUMINATO, J. A.; PAULOVIČ, F. V.; NONATO, L. G. Local affine multidimensional projection. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 17, n. 12, p. 2563–2571, 2011. Citado na página 54.
- JOO, W.; GAGE, S. H.; KASTEN, E. P. Analysis and interpretation of variability in soundscapes along an urban-rural gradient. **Landscape and Urban Planning**, Elsevier, v. 103, n. 3, p. 259–276, 2011. Citado na página 30.
- JOST, L. Entropy and diversity. **Oikos**, Wiley Online Library, v. 113, n. 2, p. 363–375, 2006. Citado na página 59.
- JOVIĆ, A.; BRKIĆ, K.; BOGUNOVIĆ, N. A review of feature selection methods with applications. In: IEEE. **Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on**. [S.l.], 2015. p. 1200–1205. Citado na página 80.
- KASTEN, E. P.; GAGE, S. H.; FOX, J.; JOO, W. The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology. **Ecological Informatics**, Elsevier, v. 12, p. 50–67, 2012. Citado nas páginas 44, 62 e 65.
- KAUFMAN, L.; ROUSSEUW, P. J. **Finding groups in data: an introduction to cluster analysis**. [S.l.]: John Wiley & Sons, 2009. Citado na página 55.
- KOHONEN, T. Self-organized formation of topologically correct feature maps. **Biological cybernetics**, Springer, v. 43, n. 1, p. 59–69, 1982. Citado na página 70.
- KRAUSE, B. Bioacoustics, habitat ambience in ecological balance. **Whole earth review**, v. 57, p. 14–18, 1987. Citado na página 29.
- _____. Using ecoacoustic methods to survey the impacts of climate change on biodiversity. **Biological Conservation**, v. 195, n. January, p. 245–254, 2016. Citado nas páginas 29 e 61.
- KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. **Psychometrika**, Springer, v. 29, n. 1, p. 1–27, 1964. Citado na página 54.
- KRUSKAL, J. B.; WISH, M. **Multidimensional scaling**. [S.l.]: Sage, 1978. Citado na página 53.
- LELLOUCH, L.; PAVOINE, S.; JIGUET, F.; GLOTIN, H.; SUEUR, J. Monitoring temporal change of bird communities with dissimilarity acoustic indices. **Methods in Ecology and Evolution**, Wiley Online Library, v. 5, n. 6, p. 495–505, 2014. Citado na página 60.
- LI, J.; WANG, J. Z. Automatic linguistic indexing of pictures by a statistical modeling approach. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, v. 25, n. 9, p. 1075–1088, 2003. Citado na página 88.
- LICHMAN, M. **UCI Machine Learning Repository**. 2013. Disponível em: <<http://archive.ics.uci.edu/ml>>. Citado na página 88.
- LUTTER, M. **Mel-Frequency Cepstral Coefficients**. 2014. Disponível em: <<http://recognize-speech.com/feature-extraction/mfcc>>. Acesso em: 14/12/2016. Citado na página 39.
- MAATEN, L. v. d.; HINTON, G. Visualizing data using t-sne. **Journal of Machine Learning Research**, v. 9, n. Nov, p. 2579–2605, 2008. Citado na página 54.

- MACCALLUM, K. **Pros and cons of two popular Ultrasound signal processing techniques**. 2015. Disponível em: <http://starfishmedical.com/2015/02/19/pros-cons-two-popular-ultrasound-signal-processing-techniques/>. Acesso em: 01/11/2016. Citado na página 38.
- MADISETTI, V. K. **The Digital Signal Processing Handbook**. [S.l.]: CRC Press, 2009. Citado na página 38.
- MCLOUGHLIN, I.; ZHANG, H.; Zhipeng Xie; SONG, Y.; XIAO, W. Robust Sound Event Classification Using Deep Neural Networks. **Ieee/Acm Transactions on Audio, Speech, and Language Processing**, IEEE, v. 23, n. 3, p. 540–552, 2015. ISSN 2329-9290. Citado na página 63.
- MEZQUIDA, D. A.; MARTÍNEZ, J. L. Platform for bee-hives monitoring based on sound analysis. a perpetual warehouse for swarm apos; s daily activity. **Spanish Journal of Agricultural Research**, v. 7, n. 4, p. 824–828, 2009. Citado na página 39.
- NEAL, L. **Detection and Segmentation of Bird Song in Noisy Environments**. Dissertação (Mestrado) — Oregon State University, 2012. Citado nas páginas 30 e 31.
- NEAL, L.; BRIGGS, F.; RAICH, R.; FERN, X. Z. Time-frequency segmentation of bird song in noisy acoustic environments. In: IEEE. **ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings**. [S.l.], 2011. p. 2012–2015. Citado na página 64.
- OJALA, T.; PIETIKAINEN, M.; MAENPAA, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, v. 24, n. 7, p. 971–987, 2002. Citado na página 47.
- PARKS, S. E.; MIKSIS-OLDS, J. L.; DENES, S. L. Assessing marine ecosystem acoustic diversity across ocean basins. **Ecological Informatics**, Elsevier, v. 21, p. 81–88, 2014. Citado nas páginas 30, 60 e 62.
- PASS, G.; ZABIH, R.; MILLER, J. Comparing images using color coherence vectors. In: ACM. **Proceedings of the fourth ACM international conference on Multimedia**. [S.l.], 1997. p. 65–73. Citado na página 50.
- PAULOVICH, F. V.; MINGHIM, R. Hipp: A novel hierarchical point placement strategy and its application to the exploration of document collections. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 14, n. 6, p. 1229–1236, 2008. Citado nas páginas 55, 56 e 80.
- PAULOVICH, F. V.; NONATO, L. G.; MINGHIM, R.; LEVKOWITZ, H. Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 14, n. 3, p. 564–575, 2008. Citado nas páginas 54 e 55.
- PAULOVICH, F. V.; SILVA, C. T.; NONATO, L. G. Two-phase mapping for projecting massive data sets. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 16, n. 6, p. 1281–1290, 2010. Citado na página 54.
- PEARSON, K. On lines and planes of closest fit to systems of points in space. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, Taylor & Francis, v. 2, n. 11, p. 559–572, 1901. Citado na página 51.

PEETERS, G. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. **CUIDADO IST Project Report**, v. 54, n. 0, p. 1–25, 2004. ISSN nul. Citado na página [63](#).

PEKIN, B.; JUNG, J.; VILLANUEVA-RIVERA, L.; PIJANOWSKI, B.; AHUMADA, J. Modeling acoustic diversity using soundscape recordings and lidar-derived metrics of vertical forest structure in aneotropical rainforest. **Landscape Ecology**, Springer, v. 27, n. 10, p. 1513–1522, 2012. Citado nas páginas [41](#) e [62](#).

PENATTI, O. A. B. *et al.* **Estudo comparativo de descritores para recuperação de imagens por conteúdo na web**. Dissertação (Mestrado) — Universidade Estadual de Campinas - Unicamp, 2009. Citado nas páginas [46](#) e [49](#).

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, v. 11, n. 3, p. 868–873, 2011. Citado nas páginas [30](#), [41](#), [42](#) e [62](#).

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**, Springer, v. 26, n. 9, p. 1213–1232, 2011. ISSN 09212973. Citado na página [29](#).

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**, Oxford University Press, v. 61, n. 3, p. 23–216, 2011. ISSN 00063568. Citado na página [29](#).

PLINGE, A.; GRZESZICK, R.; FINK, G. A. A Bag-of-Features approach to acoustic event detection. **ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings**, p. 3704–3708, 2014. ISSN 15206149. Citado na página [63](#).

PONTI, M. A.; RIBEIRO, L. S. F.; NAZARE, T. S.; BUI, T.; COLLOMOSSE, J. Everything you wanted to know about Deep Learning for Computer Vision but were afraid to ask. In: BRAZILIAN COMPUTER SOCIETY - SBC. **SIBGRAPI - Conference on Graphics, Patterns and Images**. [S.l.], 2017. Citado nas páginas [51](#) e [52](#).

POULARIKAS, A. D. **The Handbook of Formulas and Tables for Signal Processing**. [S.l.]: CRC Press, 1999. Citado na página [37](#).

RAMSAY, J. O. **Functional data analysis**. [S.l.]: Wiley Online Library, 2006. Citado na página [39](#).

ROKACH, L.; MAIMON, O. **Data Mining and Knowledge Discovery Handbook**. [S.l.]: US: Springer, 2005. 321–352 p. Citado na página [70](#).

SÁNCHEZ-GENDRIZ, I.; PADOVESE, L. Underwater soundscape of marine protected areas in the south Brazilian coast. **Marine Pollution Bulletin**, Elsevier, v. 105, n. 1, p. 65–72, 2016. Citado na página [61](#).

SÁNCHEZ-GENDRIZ, I.; PADOVESE, L. R. A methodology for analyzing biological choruses from long-term passive acoustic monitoring in natural areas. **Ecological Informatics**, v. 41, p. 1–10, 2017. ISSN 15749541. Citado nas páginas [71](#), [72](#), [73](#), [74](#), [75](#) e [109](#).

- SANCHEZÁGENDRIZ, I.; PADOVESE, L. R. Underwater sound pressure levels in the Port of Santos: Representative sample of forty-five days. **2015 IEEE/OES Acoustics in Underwater Geosciences Symposium (RIO Acoustics)**, p. 1–4, 2015. Citado na página 38.
- SANKUPELLAY, M.; TOWSEY, M.; TRUSKINGER, A.; ROE, P. Visual fingerprints of the acoustic environment: The use of acoustic indices to characterise natural habitats. In: IEEE. **Big Data Visual Analytics (BDVA)**, 2015. [S.l.], 2015. p. 1–8. Citado nas páginas 30, 69 e 71.
- SCHAFER, M. R. **The Tuning of the World**. 1st. ed. [S.l.]: Random House Inc (T), 1977. Citado nas páginas 29 e 30.
- SERVICK, K. Eavesdropping on Ecosystems. **Science (New York, N.Y.)**, American Association for the Advancement of Science, v. 343, n. February, p. 834–837, 2014. ISSN 1095-9203. Citado nas páginas 29 e 30.
- SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding machine learning: From theory to algorithms**. [S.l.]: Cambridge university press, 2014. Citado na página 51.
- SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, v. 27, p. 379–423, 623–656, July 1948. Citado na página 40.
- SIMPSON, E. H. Measurement of diversity nature. **Nature**, 1949. Citado na página 41.
- SOUTHWORTH, M. F. **The Sonic Environment of Cities**. Sage Publications, v. 1, n. 1, p. 49, 1969. ISSN 0013-9165. Citado na página 29.
- STEHLING, R. O.; NASCIMENTO, M. A.; FALCÃO, A. X. A compact and efficient image retrieval approach based on border/interior pixel classification. In: ACM. **Proceedings of the eleventh international conference on Information and knowledge management**. [S.l.], 2002. p. 102–109. Citado na página 50.
- SUEUR, J.; FARINA, A.; GASC, A.; PIERETTI, N.; PAVOINE, S. Acoustic indices for biodiversity assessment and landscape investigation. **Acta Acustica united with Acustica**, S. Hirzel Verlag, v. 100, n. 4, p. 772–781, 2014. Citado nas páginas 31, 40 e 45.
- SUEUR, J.; PAVOINE, S.; HAMERLYNCK, O.; DUVAIL, S. Rapid acoustic survey for biodiversity appraisal. **PloS one**, Public Library of Science, v. 3, n. 12, p. e4065, 2008. Citado nas páginas 43 e 62.
- SWAIN, M. J.; BALLARD, D. H. Color indexing. **International journal of computer vision**, Springer, v. 7, n. 1, p. 11–32, 1991. Citado na página 49.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to data mining**. 1st. [S.l.]: Boston: Pearson Addison Wesley. xxi, 2005. Citado na página 55.
- TEJADA, E.; MINGHIM, R.; NONATO, L. G. On improved projection techniques to support visual exploration of multi-dimensional data sets. **Information Visualization**, SAGE Publications, v. 2, n. 4, p. 218–231, 2003. Citado na página 53.
- TELEA, A. C. **Data visualization: principles and practice**. 2. ed. [S.l.]: CRC Press, 2014. ISBN 978-1466585263. Citado nas páginas 53 e 57.
- THOROGOOD, M.; FAN, J.; PASQUIER, P. BF-Classifer: Background/Foreground Classification and Segmentation of Soundscape Recordings. **Proceedings of the Audio Mostly on Interaction with Sound**, p. 32, 2015. Citado na página 63.

- TOWSEY, M.; WIMMER, J.; WILLIAMSON, I.; ROE, P. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. **Ecological Informatics**, Elsevier, v. 21, p. 110–119, 2014. Citado nas páginas 40, 45 e 62.
- 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, v. 29, p. 703–712, 2014. Citado nas páginas 30, 31, 45, 67, 68, 69, 70 e 110.
- VILLANUEVA-RIVERA, L.; PIJANOWSKI, B.; DOUCETTE, j.; PEKIN, B. A primer of acoustic analysis for landscape ecologists. **Landscape Ecology**, Springer, v. 26, n. 9, p. 1233–1246, 2011. Citado nas páginas 41 e 62.
- WARD, M. O.; GRINSTEIN, G.; KEIM, D. **Interactive data visualizaton: foundations, techniques, and applications**. 2. ed. [S.l.]: CRC Press, 2015. 578 p. ISBN 978-1482257373. Citado nas páginas 31, 53 e 57.
- WELCH, P. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. **IEEE Transactions on audio and electroacoustics**, IEEE, v. 15, n. 2, p. 70–73, 1967. Citado na página 38.
- WICHERN, G.; THORNBURG, H.; MECHTLEY, B.; FINK, A.; TU, K.; SPANIAS, A. Robust multi-feature segmentation and indexing for natural sound environments. **CBMI'2007 - 2007 International Workshop on Content-Based Multimedia Indexing, Proceedings**, p. 69–76, 2007. Citado nas páginas 63 e 65.
- WICHERN, G.; XUE, J.; THORNBURG, H.; MECHTLEY, B.; SPANIAS, A. Segmentation, indexing, and retrieval for environmental and natural sounds. **IEEE Transactions on Audio, Speech and Language Processing**, IEEE, v. 18, n. 3, p. 688–707, 2010. Citado na página 65.
- XIE, J.; TOWSEY, M.; ZHANG, J.; DONG, X.; ROE, P. Application of image processing techniques for frog call classification. **2015 IEEE International Conference on Image Processing (ICIP)**, p. 4190–4194, 2015. Citado na página 64.
- XIE, Y. **Parallel coordinates plot**. 2010. Disponível em: <<https://github.com/ggobi/cranvas/wiki/Parallel-coordinates-plot>>. Acesso em: 13/12/2017. Citado na página 57.
- ZALKOW, F. **Create audio spectrograms with Python**. 2013. Disponível em: <<http://www.frank-zalkow.de/en/code-snippets/create-audio-spectrograms-with-python.html?i=1>>. Acesso em: 24/01/2017. Citado na página 37.

SILHUETA DE CONJUNTOS

Tabela 4 – Valores de silhueta dos conjuntos de características no espaço original utilizando características com valores originais e normalizados (*Min-max*).

	5 Classes		4 Classes	
Data	Original	Normalizado	Original	Normalizado
Índices e métricas acústicas				
<i>Costa Rica</i>	0,01	0,06	0,01	0,05
<i>Ilhéus</i>	0,04	0,03	–	–
<i>Costa Rica + Ilhéus + Laje</i>	-0,13	0,12	-0,07	0,20
Descritores de Imagem				
<i>Costa Rica</i>	0,07	0,05	0,07	0,05
<i>Ilhéus</i>	0,07	0,05	–	–
<i>Costa Rica + Ilhéus + Laje</i>	0,10	0,09	0,12	0,10
Autoencoder completo				
<i>Costa Rica</i>	0,03	0,03	0,03	0,03
<i>Ilhéus</i>	0,04	0,03	–	–
<i>Costa Rica + Ilhéus + Laje</i>	0,02	0,02	0,08	0,08
Autoencoder sumarizado				
<i>Costa Rica</i>	0,04	0,04	0,04	0,04
<i>Ilhéus</i>	0,04	0,04	–	–
<i>Costa Rica + Ilhéus + Laje</i>	0,01	0,01	0,09	0,08

Fonte: Dados da pesquisa.

Tabela 5 – Valores de silhueta dos conjuntos de características associados com Projeções Multidimensionais. Os melhores valores para cada linha de características estão destacados. O Autoencoder *completo* possui todas as características geradas por ele (23.040 valores) e o *sumarizado* possui uma sumarização dessas características (160 valores).

	Force	LAMP	LSP	MDS	PCA	t-SNE
<i>Índices e métricas acústicas</i>	-0.1412	-0.153	-0.1254	-0.1468	-0.1468	0.0264
<i>Descritores de Imagem</i>	0.0502	0.0805	0.0882	0.0485	0.0485	0.0518
<i>Autoencoder completo</i>	0.0608	0.0557	0.0690	0.0643	0.0643	0.0220
<i>Autoencoder sumarizado</i>	0.0705	0.0717	0.0736	0.0696	0.0696	-0.0237

Fonte: Dados da pesquisa.

Tabela 6 – Valores de silhueta dos dados considerando as características acústicas divididas em 3 grupos, índices, MFCC e outras métricas. As colunas para cada conjunto indicam os valores de silhueta para o espaço original, normalizado e projetado, respectivamente. Os valores consideram 5 classes de dados.

	Índices			MFCC			Outras Métricas		
	Orig.	Norm.	Proj.	Orig.	Norm.	Proj.	Orig.	Norm.	Proj.
<i>Costa Rica</i>	0.0002	0.0187	0.0162	0.0948	0.0681	0.2216	0.0099	0.0567	0.0412
<i>Ilhéus</i>	0.0373	0.0395	0.0510	0.0255	0.0240	0.0082	0.0395	0.0102	0.0143
<i>Costa Rica + Ilhéus + Laje</i>	0.1587	0.0981	0.0119	0.1842	0.1389	0.1121	-0.1905	0.0091	0.0330

Fonte: Dados da pesquisa.

Tabela 7 – Valores de silhueta dos dados considerando os descritores de imagem divididos em 2 grupos, textura e cor. As colunas para cada conjunto indicam os valores de silhueta para o espaço original, normalizado e projetado, respectivamente. Os valores consideram 5 classes de dados.

	Textura			Cor		
	Original	Norm.	Proj.	Original	Norm.	Proj.
<i>Costa Rica</i>	0.0282	0.0246	0.0254	0.0742	0.0699	0.1003
<i>Ilhéus</i>	0.0868	0.0456	0.0373	0.0685	0.0590	0.0613
<i>Costa Rica + Ilhéus + Laje</i>	0.1157	0.1036	0.0412	0.0939	0.0811	0.0602

Fonte: Dados da pesquisa.

SEGMENTAÇÃO DE ESPECTROGRAMA

Figura 61 – Na primeira linha, a imagem apresenta um espectrograma de áudio subaquático. Na segunda linha, são apresentados espectrogramas de dois componentes extraídos com Fatoração de Matriz Não-negativa - *Non-Negative Matrix Factorization* (NMF).

Figura 62 – A partir dos componentes apresentados na figura anterior, as linhas seguintes identificam: limiarização das imagens dos componentes, aplicação de Morfologia Matemática para remoção de ruído das imagens, detecção de componentes conexos e demarcação desses objetos conexos nas imagens originais dos componentes de áudio.

RESULTADOS DE SELEÇÃO AUTOMÁTICA DE ATRIBUTOS

Foi utilizado o programa de Mineração de Dados Weka 3.8 para selecionar um conjunto de características para comparação com as seleções que foram realizadas. Foram aplicados os algoritmos *Correlation*, *InfoGain* e *ReliefF*, disponíveis no Weka ((??)). A seleção foi realizada em cada grupo de características (índices e medidas acústicas, descritores de imagem e características de imagem geradas pelo Autoencoder) e avaliada através dos valores de silhueta e das projeções correspondentes.

Cada algoritmo de seleção retorna uma lista com a ordenação das características. Para cálculo da silhueta foram utilizados conjuntos incrementais das características (ex.: silhueta utilizando os 2 primeiros atributos de cada lista, os 3 primeiros, os 4 primeiros etc.).

C.1 Índices e Metricas Acústicas

A Figura 63 apresenta a comparação da silhueta para cada algoritmo de seleção citado anteriormente. Os 6 primeiros atributos retornados pelo *Correlation* geram o maior valor de silhueta (0,2526): ACI, ADI, BIO, H, MFCC_2, MFCC_4, MFCC_12. A Figura 64 apresenta a projeção t-SNE dos dados utilizando essas características.

Figura 63 – Valores de silhueta gerados utilizando cada uma das listas de atributos geradas pelas técnicas de seleção de atributos empregadas. Foram consideradas 3 ou 4 classes (dados de Ilhéus juntos ou separados) para os cálculos.

Figura 64 – Projeção t-SNE dos dados com as medidas acústicas selecionadas automaticamente.

C.2 Descritores de Imagem

A Figura 65 apresenta a comparação da silhueta para cada algoritmo de seleção citado anteriormente. Os 2 primeiros atributos retornados pelo Correlation geram o maior valor de silhueta (0,1638): Contraste da Matriz de Coocorrência, LBP9. A Figura 66 apresenta a projeção t-SNE dos dados utilizando essas características.

Figura 65 – Valores de silhueta gerados utilizando cada uma das listas de atributos geradas pelas técnicas de seleção de atributos empregadas. Foram consideradas 3 ou 4 classes (dados de Ilhéus juntos ou separados) para os cálculos.

Figura 66 – Projeção t-SNE dos dados com os descritores de imagem selecionados automaticamente.

C.3 Características de Imagem extraídas por Autoencoder

As Figuras 67 e 68 apresentam a comparação da silhueta para cada algoritmo de seleção citado anteriormente, considerando todos os atributos gerados pelo Autoencoder (23.040 valores) e a sua sumarização (160 valores). Para 23.040 atributos, os 48 primeiros atributos retornados pelo ReliefF geram o maior valor de silhueta (-0,0180). Enquanto isso, para 160 atributos, os 58 primeiros atributos retornados pelo ReliefF geram o maior valor de silhueta (-0,0634). As Figuras 69 e 70 apresentam as projeções t-SNE dos atributos originais e dos sumarizados, utilizando as respectivas características selecionadas.

Figura 67 – Valores de silhueta gerados utilizando cada uma das listas de atributos geradas pelas técnicas de seleção de atributos empregadas. Foram consideradas 3 ou 4 classes (dados de Ilhéus juntos ou separados) para os cálculos. Foram utilizadas as 23.040 características geradas pelo Autoencoder.

Figura 68 – Valores de silhueta gerados utilizando cada uma das listas de atributos geradas pelas técnicas de seleção de atributos empregadas. Foram consideradas 3 ou 4 classes (dados de Ilhéus juntos ou separados) para os cálculos. Foram utilizadas as 160 valores gerados pela sumarização das características obtidas pelo Autoencoder.

Figura 69 – Projeção t-SNE dos dados com as medidas extraídas com o Autoencoder selecionadas automaticamente.

Figura 70 – Projeção t-SNE dos dados com a sumarização das medidas extraídas com o Autoencoder selecionadas automaticamente.