



MARINHA DO BRASIL
INSTITUTO DE ESTUDOS DO MAR ALMIRANTE PAULO MOREIRA
PROGRAMA DE PÓS-GRADUAÇÃO EM ACÚSTICA SUBMARINA

CARACTERÍSTICAS ESPECTRAIS DINÂMICAS EXTRAÍDAS DE SINAIS
ACÚSTICOS SUBMARINOS PARA A CLASSIFICAÇÃO DE EMBARCAÇÕES
UTILIZANDO ALGORITMOS DE APRENDIZAGEM SUPERVISIONADA.

Murillo de Brito Santos

Dissertação de Mestrado apresentada ao
Programa de Pós-Graduação em Acústica
Submarina, PPGAS, do Instituto de Estudos
do Mar Almirante Paulo Moreira, como parte
dos requisitos necessários à obtenção do título
de Mestre em Acústica Submarina.

Orientador: Rogério de Moraes Calazan

Arraial do Cabo
Junho de 2024

CARACTERÍSTICAS ESPECTRAIS DINÂMICAS EXTRAÍDAS DE SINAIS
ACÚSTICOS SUBMARINOS PARA A CLASSIFICAÇÃO DE EMBARCAÇÕES
UTILIZANDO ALGORITMOS DE APRENDIZAGEM SUPERVISIONADA.

Murillo de Brito Santos

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA
DE PÓS-GRADUAÇÃO EM ACÚSTICA SUBMARINA DO INSTITUTO DE
ESTUDOS DO MAR ALMIRANTE PAULO MOREIRA COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ACÚSTICA SUBMARINA.

Orientador: Prof. Rogério de Moraes Calazan, Ph.D.

Aprovada por: Prof. Vicente Barroso Junior, D.Sc
Prof. Renato Peres Vio, Ph.D.
Prof. Fábio Contrera Xavier, D.Sc.
Prof. Fabrício de Abreu Bozzi, Ph.D.

ARRAIAL DO CABO, RJ – BRASIL
JUNHO DE 2024

de Brito Santos, Murillo

Características Espectrais Dinâmicas extraídas de sinais acústicos submarinos para a classificação de embarcações utilizando algoritmos de aprendizagem supervisionada./Murillo de Brito Santos. – Arraial do Cabo: IEAPM, 2024.

XV, 89 p.: il.; 29,7cm.

Orientador: Rogério de Moraes Calazan

Dissertação (mestrado) – IEAPM/Programa de Pós-Graduação em Acústica Submarina, 2024.

Referências Bibliográficas: p. 77 – 83.

1. Características Espectrais Dinâmicas. 2. Classificação de embarcações. 3. Extração de características. I. de Moraes Calazan, Rogério. II. Instituto de Estudos do Mar Almirante Paulo Moreira, Programa de Pós-Graduação em Acústica Submarina. III. Título.

*À minha amada esposa, Lívia,
por sua infinita paciência,
compreensão e apoio
incondicional durante toda a
jornada do mestrado. Agradeço
por ter acreditado em mim, por
ter compartilhado meus sonhos e
anseios de carreira, mesmo nos
momentos mais desafiadores,
especialmente durante a gravidez
e nos primeiros meses da vida da
nossa filha. Sua presença
constante e amor incondicional
foram a base fundamental para o
meu sucesso.*

Agradecimentos

Primeiramente agradeço a Deus pelo dom da vida e por me permitir concluir mais um desafio e a minha família pelo apoio incondicional ao longo desses 2 anos.

Em segundo lugar, gostaria de manifestar meu mais sincero agradecimento ao meu orientador, Comandante Moraes por sua orientação, sabedoria e incentivo constante ao longo do desenvolvimento deste trabalho. Agradeço a oportunidade de ter aprendido com sua experiência e conhecimento, que foram essenciais para o meu crescimento profissional e acadêmico. Sua dedicação e compromisso com o meu aprendizado foram fundamentais para a realização deste trabalho.

Em seguida, gostaria de agradecer aos meus queridos colegas de turma, Wendel, Richard, Marco, Bárbara e Yasmim, por proporcionarem um ambiente de aprendizado e companheirismo durante o mestrado. As nossas trocas de ideias, os momentos de estudo em conjunto e a amizade que se construiu ao longo do curso nunca será esquecido. Agradeço por todo o apoio, incentivo e por tornarem essa etapa da minha vida ainda mais especial.

Ao coordenador do curso, CF Renato, pela sua dedicação e empenho em tornar o curso de mestrado em acústica submarina um programa de excelência. Agradeço por acreditar no potencial do curso e por contribuir para a formação de profissionais qualificados e comprometidos com o desenvolvimento da área.

Aos docentes Fábio Contreras, CMG Barroso, CF Guarino, CMG Simões e aos demais professores do PPGAS, meu profundo agradecimento por suas valiosas contribuições para a minha formação, pelas suas aulas inspiradoras, orientações valiosas e incentivos constante.

Não poderia deixar de agradecer aos profissionais e amigos do Departamento de Acústica, CC Backer, CC Lima, CC Amilcar, CC Bozzi e Ten Tavares, sempre solícitos e comprometidos em ajudar, além dos bons momentos de risadas.

A todos os que me acompanharam e me apoiaram nesta jornada, meu mais profundo agradecimento

Resumo da Dissertação apresentada ao IEAPM como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

CARACTERÍSTICAS ESPECTRAIS DINÂMICAS EXTRAÍDAS DE SINAIS ACÚSTICOS SUBMARINOS PARA A CLASSIFICAÇÃO DE EMBARCAÇÕES UTILIZANDO ALGORITMOS DE APRENDIZAGEM SUPERVISIONADA.

Murillo de Brito Santos

Junho/2024

Orientador: Rogério de Moraes Calazan

Programa: Acústica Submarina

A classificação de embarcações desempenha um papel fundamental no monitoramento do tráfego marítimo e na identificação de fontes sonoras no ambiente submarino sejam elas de origem antropogênicas ou biológicas. No entanto, classificar uma embarcação com base apenas em seu som irradiado é desafiador devido à complexidade do oceano como guia de onda submarino. Este trabalho visa desenvolver um algoritmo de extração de características robusto e eficaz para capturar aspectos dinâmicos dos sinais submarinos ao longo do tempo. Além disso, avalia-se um sistema de classificação composto por diferentes algoritmos de aprendizado de máquina. O algoritmo proposto extrai características espectrais dinâmicas (Spectral Dynamic Features - SDF), combinando-as com coeficientes Mel cepstrais (Mel Frequency Cepstral Coefficients) para formar um vetor de características combinado. O desempenho do método proposto é avaliado utilizando dois conjuntos de dados contendo sons de embarcações, o primeiro coletado na costa noroeste da Espanha e o segundo coletado nas proximidades do município de Arraial do Cabo – RJ. Adicionalmente, de forma a ampliar a análise do método proposto, os dois conjuntos de dados contêm diferentes embarcações e condições ambientais. Os resultados demonstram que o método proposto alcança uma precisão de classificação superior aos métodos tradicionais baseados em características estáticas.

Abstract of Dissertation presented to IEAPM as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

SPECTRAL DYNAMIC FEATURES EXTRACTED FROM AUDIO SIGNALS
FOR MARITIME VESSEL CLASSIFICATION USING SUPERVISED
LEARNING ALGORITHMS.

Murillo de Brito Santos

June/2024

Advisor: Rogério de Moraes Calazan

Department: Underwater Acoustics

The classification of vessels plays a fundamental role in monitoring maritime traffic and identifying sound sources in the underwater environment, whether they are of anthropogenic or biological origin. However, classifying a vessel based solely on its emitted sound is challenging due to the complexity of the ocean as an underwater waveguide. This work aims to develop a robust and effective feature extraction algorithm to capture dynamic aspects of underwater signals over time. Additionally, a classification system composed of different machine learning algorithms is evaluated. The proposed algorithm extracts dynamic spectral features (Spectral Dynamic Features - SDF), combining them with Mel Frequency Cepstral Coefficients (MFCCs) to form a combined feature vector. The performance of the proposed method is evaluated using two datasets containing vessel sounds, the first collected off the northwest coast of Spain and the second collected near the municipality of Arraial do Cabo – RJ. Furthermore, to broaden the analysis of the proposed method, the two datasets contain different vessels and environmental conditions. The results demonstrate that the proposed method achieves a classification accuracy superior to traditional methods based on static features.

Sumário

Lista de Figuras	x
Lista de Tabelas	xiii
Lista de Abreviaturas	xiv
1 Introdução	1
1.1 Apresentação do problema	2
1.2 Motivação	3
1.3 Objetivos	6
1.4 Publicação	6
1.5 Organização	7
2 Referencial teórico	8
2.1 Acústica submarina	8
2.1.1 Influência do meio na propagação da onda sonora	9
2.1.2 Considerações sobre o nível de pressão sonora (Sound Pressure Level - SPL) na tarefa de classificação	11
2.1.3 Considerações para a extração de características	11
2.1.4 Características do ruído emitido por embarcações	12
2.2 Processamento digital de sinais para extração de características	13
2.2.1 Segmentação	14
2.2.2 Pré-processamento	15
2.2.3 Análise Espectral	17
2.2.4 Extração de características	18
2.3 O Cepstrum	19
2.4 Coeficientes Mel Cepstrais	21
2.5 Aprendizagem de máquina	24
2.5.1 Terminologias e introdução a conceitos básicos de <i>ML</i>	25
2.5.2 Algoritmos de classificação supervisionado	28
2.5.3 Método de validação cruzada "deixar-um-fora" (<i>Leave-One-Out</i>)	32

3	Metodologia	35
3.1	Arquitetura do algoritmo de extração de características proposto . . .	35
3.2	Análise espectral proposta	36
3.3	Extração do vetor de característica	39
3.3.1	Extração do vetor de características MFCC	39
3.3.2	Extração do Vetor de características espectrais dinâmicas (SDF)	40
3.3.3	Vetor de características Combinado	43
4	Conjunto de dados	44
4.1	Banco de dados <i>ShipsEar</i>	44
4.2	Experimento na Enseada dos Anjos - Arraial do Cabo-RJ	48
5	Resultados e Discussão	51
5.1	Considerações sobre os conjuntos de dados	51
5.2	Resultados da classificação de embarcações marítimas utilizando o conjunto de dados <i>ShipsEar</i>	52
5.2.1	Resultados da classificação pelo método de validação cruzada <i>Leave-One-Out</i>	52
5.2.2	Resultados da classificação de todo um arquivo pelo método de validação cruzada <i>Leave-One-Out</i>	56
5.2.3	Resultados da classificação pelo método de validação <i>holdout</i> simples	57
5.2.4	Análise do tamanho dos segmentos e distribuição dos vetores de características no espaço 2D	59
5.3	Classificação das embarcações marítimas utilizando dados coletados na Enseada dos Anjos - Arraial do Cabo- RJ	62
5.4	Análise de robustez dos extratores de características	66
5.4.1	Avaliação da robustez pela adição de ruído gaussiano branco (<i>AWGN</i>)	68
5.4.2	Avaliação da robustez com a adição de ruído coletado na Enseada dos Anjos- Arraial do Cabo - RJ	71
6	Conclusões e sugestões para trabalhos futuros	75
	Referências Bibliográficas	77
	Código em Python para os Extratores de características	84

Lista de Figuras

2.1	Diagrama de blocos do processamento digital de sinais para extração de características	14
2.2	Representação gráfica ilustrando as etapas do pré-processamento e o Fluxograma simplificado das etapas do Pré-processamento	16
2.3	Transformação do sinal do domínio do tempo para o domínio Cepstral	20
2.4	Banco de Filtros Mel	21
2.5	Exemplo de um banco de filtros Mel e o efeito da aplicação do filtro no espectro de potência	23
2.6	Fluxograma do processo de obtenção dos MFCCs	24
2.7	Exemplo de modelo de aprendizado supervisionado	26
2.8	Distribuição de um conjunto de dados de entrada previamente rotulados em um espaço 2-D	27
2.9	Criação de um classificador utilizando Aprendizado de Máquina (ML) Supervisionado	27
2.10	Estrutura geral do algoritmo de Máquina de Vetor de Suporte (SVM, <i>Support Vector Machine</i>)	28
2.11	Transformação de um problema não linearmente separável em um problema linearmente separável utilizando a função <i>kernel</i> Φ	29
2.12	Fluxograma de classificação feito pelo KNN	30
2.13	Simplificação do Algoritmo Floresta Aleatória	32
2.14	Método de validação cruzada "deixar-um-fora" (<i>Leave-One-Out</i>)	33
3.1	Arquitetura do algoritmo de extração de características proposto utilizando a combinação do SDF com MFCCs.	36
3.2	Processo de obtenção do Vetor de características combinado	37
3.3	Banco de filtros triangulares projetado	39
3.4	Detalhamento da etapa Mel contendo a representação matricial do processo de extração do vetor de característica MFCC proposto	41
3.5	Aplicação da <i>FFT</i> nas colunas da Matriz E	41

4.1	Interface do site http://atlanttic.uvigo.es/underwaternoise , banco de dados de teste	45
4.2	Espectrogramas de várias amostras no conjunto de dados Shipsear.	47
4.3	Configuração vertical do experimento na Enseada dos Anjos	48
4.4	Área do experimento na Enseada dos Anjos, Arraial do Cabo	49
4.5	Hidrofone desenvolvido no IEAPM com material de baixo custo.	49
5.1	Matriz de confusão gerada a partir do conjunto de dados da Tabela 4.2, utilizando o algoritmo de classificação <i>SVM</i> , tendo como entrada dois tipos de extratores de características. a) Vetor de características combinado ($MFCC^\dagger + SDF$) e b) Vetor de Características Espectrais Dinâmicas(<i>SDF</i>).	56
5.2	Matriz de confusão considerando todo o arquivo(6 segmentos)	57
5.3	Efeito da variação do tamanho do segmento sobre a acurácia do classificador <i>SVM</i>	60
5.4	Representação gráfica da distribuição em 2-D dos vetores de características extraídos por 3 métodos distintos. O vetor de característica é reduzido a 2 dimensões representados como componentes principais X_1 e X_2 . a) vetor $MFCC^\dagger + SDF$, b) vetor com os Coeficientes espectrais de potência (<i>PSC</i>) e c) vetor de características <i>LPC</i>	61
5.5	Matriz de confusão obtida utilizando como entrada para o algoritmo <i>SVM</i> , os vetores de características (a) $MFCC^\dagger$ e (b) $MFCC^\dagger + SDF$	64
5.6	Matriz de confusão obtida utilizando como entrada para o algoritmo <i>SVM</i> , os vetores de características (a) <i>LPC</i> e (b) $LPC + SDF$	65
5.7	Distribuição bidimensional das características principais do vetor <i>LPC</i> com rótulos previamente definidos.	65
5.8	Distribuição bidimensional das características principais do vetor $LPC + SDF$ com rótulos previamente definidos	66
5.9	Representação do calculo da distância entre vetores de características $ds_{SDF/MFCC^\dagger+SDF/MFCC/MFCC^\dagger}$	68
5.10	Espectrogramas do sinal original da embarcação (Banco de Dados "ShipsEar") e do sinal original após adição de ruído branco gaussiano a uma $SNR = 17dB$ e $SNR = 23 dB$	70
5.11	Dispersão do espaço de características (D_s) para diferentes níveis de <i>SNR</i> devido a adição de <i>AWGN</i> utilizando os extratores de características $MFCC^\dagger + SDF$, <i>SDF</i> , $MFCC^\dagger$ e <i>MFCC</i>	70

5.12	Espectrogramas do sinal de teste original da embarcação (Conjunto de Dados "ShipsEar"), do sinal de ruído gravado na Enseada dos Anjos e do sinal original da embarcação após adição de ruído gravada na Enseada a uma $SNR = 7\text{dB}$ e $SNR = 20\text{ dB}$	72
5.13	Dispersão do espaço de características (D_s) para diferentes níveis de SNR devido a adição do ruído coletado na Enseada dos Anjos utilizando os extratores de características $MFCC^\dagger + SDF$, SDF , $MFCC^\dagger$ e $MFCC$	73

Lista de Tabelas

3.1	Configuração do banco de filtros Mel Projetado	38
4.1	Subdivisões das classes	46
4.2	Identificação (ID) no banco de dados ShipsEar, dos arquivos .wav utilizados nos experimentos	46
4.3	Subdivisões das classes - Experimento realizado na Enseada dos Anjos	50
5.1	Comparação da porcentagem de classificações corretas entre o vetor de características proposto e outros 12 tipos de vetores de características para 4 tipos de classificadores, utilizando dados do conjunto de dados " <i>ShipsEar</i> ".	54
5.2	Comparação da porcentagem de classificações corretas entre o vetor de características proposto e outros 12 tipos de vetores de características para 4 tipos de classificadores, utilizando dados do conjunto de dados " <i>ShipsEar</i> ".	58
5.3	Comparação da porcentagem de classificações corretas, para cada classe, entre o vetor de características proposto e outros 12 tipos de vetores de características, classificados pelo SVM, utilizando dados do conjunto de dados " <i>ShipsEar</i> ".	59
5.4	Comparação da porcentagem de classificações corretas entre o vetor de características proposto e outros 12 tipos de vetores de características para 4 tipos de classificadores, com dados obtidos na enseada dos Anjos.	63
5.5	Resultado da classificação, pelo algoritmo SVM, dos 12 segmentos do sinal de teste com adição de ruído.	74

Lista de Abreviaturas

AWGN	ruído gaussiano branco aditivo (<i>Additive White Gaussian Noise</i>), p. 68
CC	coeficientes cepstrais (<i>cepstral coefficient</i>), p. 20
CD	componentes dinâmicos, p. 42
DEMON	<i>Demodulation of Envelope Modulation On Noise</i> , p. 5
DFT	transformada Discreta de Fourier (<i>Discrete Fourier transform</i>), p. 17
FFT	Transformada Rápida de Fourier (<i>Fast Fourier transform</i>), p. 4
FT	transformada de Fourier (<i>Fourier Transform</i>), p. 17
GFCC	Coeficientes Cepstrais de Frequência Gammatone, <i>Gammatone Frequency Cepstral Coefficients</i> , p. 5
IEAPM	Instituto de Estudos do Mar Almirante Paulo Moreira, p. 48
IFT	Transformada Inversa de Fourier (<i>Inverse Fourier Transform</i>), p. 19
IPqM	Instituto de Pesquisas da Marinha, p. 48
LOFAR	<i>Low-Frequency Analysis and Recording</i> , p. 5
LOOCV	validação cruzada "deixar-um-fora" (<i>Leave-One-Out Cross-Validation</i>), p. 32
LPCC	coeficientes cepstrais de predição linear (<i>linear prediction cepstral coefficients</i>), p. 4
LPC	codificação preditiva linear (<i>Linear Predictive Coding</i>), p. 4

MFCC	coeficientes cepstrais de frequência Mel (<i>Mel-frequency cepstral coefficients</i>), p. 4
ML	aprendizagem de máquina (<i>machine learning</i>), p. 1
PSC	coeficientes espectrais de potência (<i>power spectral coefficients</i>), p. 18
RMS	<i>Root Mean Square</i> , p. 4
SDF	Características Espectrais Dinâmicas, <i>Spectral Dynamic Features</i> , p. 4
SNR	relação sinal-ruído (<i>Signal-to-Noise Ratio</i>), p. 5
STFT	transformada de Fourier de Tempo Curto (<i>Short-Time Fourier transform</i>), p. 18
ZCR	taxa de cruzamento zero (<i>zero-crossing rate</i>), p. 4
fs	frequência de amostragem, p. 14
k-NN	K-vizinhos mais próximo, <i>k-Nearest Neighbors</i> , p. 5

Capítulo 1

Introdução

A classificação de um sinal sonoro submarino é um processo de atribuir rótulos ou categorias a dados acústicos, ou seja, representações digitais de sons captados no ambiente submarino, com base em padrões e características extraídos desses sinais [1]. O objetivo é treinar um modelo a reconhecer automaticamente a classe à qual cada sinal pertence [2], complementando e otimizando o trabalho de classificação manual realizado por operadores de sonares. Nesse contexto, a detecção e classificação de embarcações com base em seus padrões acústicos desempenham um papel fundamental no monitoramento do tráfego marítimo, além de contribuir para a identificação de fontes sonoras em sistemas de monitoramento ambiental submarino [3–6]. No entanto, alcançar precisão na classificação de embarcações pela análise do seu ruído emitido, trata-se de uma tarefa desafiadora principalmente devido a complexidade do meio de propagação acústico submarino [7]. Variações no ambiente submarino, como temperatura, salinidade e pressão, podem afetar a propagação dos sinais acústicos. Além disso, o oceano apresenta ruídos ambientais diversos, provenientes de correntes oceânicas, atividades biológicas e eventos geofísicos [8, 9], tornando a distinção entre os padrões acústicos das embarcações mais complexa.

Superar os desafios da classificação de sons submarinos exige uma combinação de técnicas avançadas de processamento de sinal e algoritmos de aprendizado de máquina (ML, *machine learning*). O processamento de sinal extrai características relevantes dos sons, como energia em diferentes faixas de frequência, harmônicos e padrões temporais, servindo como base para a classificação. Já o ML treina modelos para classificar automaticamente esses sons. Algoritmos de ML supervisionados aprendem a partir de um conjunto de dados rotulados, onde cada som possui sua classe¹ definida. O modelo aprende a identificar padrões nos dados e associá-los às

¹Em um contexto de ML, uma “classe” refere-se a uma categoria ou rótulo atribuído a um conjunto específico de dados. No contexto da classificação de sons, cada som pode ser associado a uma classe que representa o tipo de som que é. Por exemplo, em um conjunto de dados de sons de embarcações, as classes podem incluir “embarcação pesqueira”, “navio mercante”, “lanchas”, entre outros.

classes de sons correspondentes. Essa integração cria um sistema autônomo para analisar, interpretar e classificar os sons submarinos. Este trabalho busca aplicar tais técnicas para desenvolver um sistema robusto e eficaz para a classificação dos sons irradiados por embarcações.

A Seção 1.1 deste capítulo apresenta os problemas que envolvem a classificação de sons submarinos. Em seguida, a Seção 1.2 explana sobre a motivação que impulsionou este trabalho. Posteriormente, na Seção 1.3, são enumerados os objetivos que norteiam este estudo. Na Seção 1.4, são listadas as publicações resultantes desta dissertação, enquanto a Seção 1.5 discute a organização detalhada deste trabalho.

1.1 Apresentação do problema

A classificação de sons submarinos pode ser uma tarefa complexa e desafiadora. A propagação do som na água é um processo heterogêneo no espaço e tempo, sujeito a transformações intrínsecas às características do meio onde se propaga. As ondas sonoras no oceano estão sujeitas a atenuações resultante dos fenômenos de espalhamento e absorção na água, a variação de sua velocidade de propagação, a ocorrência de refração, manifestada pelo desvio na direção de propagação devido às variações na velocidade do som, e à reflexão nas interfaces distintas. Adicionalmente, um sinal emitido sofre distorções devido à heterogeneidade do meio, como variações de frequência ocasionadas pelo efeito de Doppler, e a presença de outras fontes sonoras no oceano, como correntes oceânicas e vocalizações de animais marinhos, que pode, por si só, mascarar o sinal acústico de interesse [10]. A esses fatores somam-se a sobreposição de múltiplos espectros devido a presença de múltiplas fontes, as variações nas condições ambientais e a sobreposição de características, tornando a distinção entre diferentes tipos de fontes acústicas uma tarefa desafiadora [7].

A complexa classificação de sons submarinos se torna ainda mais desafiadora quando se busca discernir entre os diferentes tipos de embarcações. Cada classe de embarcação emite padrões acústicos distintos e dentro de uma mesma classe ainda pode haver padrões diferentes a depender da configuração do maquinário, regime de operação e as condições do meio de propagação [1]. Outro fator limitante é a falta de conjuntos de dados abrangentes e diversificados. A baixa quantidade e qualidade desses registros acústicos compromete a capacidade de treinar algoritmos de classificação de forma robusta, minando a eficácia de sistemas destinados à identificação precisa de embarcações marítimas [4].

No contexto da classificação de sons submarinos, o “ruído” se refere a qualquer som indesejado ou irrelevante que interfere na detecção e identificação do “sinal de interesse”. Esse ruído pode ser gerado por diversas fontes, como correntes oceânicas, vocalizações de animais marinhos, reverberação do fundo do mar e até mesmo o

próprio maquinário da embarcação. O sinal de interesse, por outro lado, se refere ao som emitido por uma embarcação que se deseja classificar. Este sinal geralmente apresenta características acústicas específicas que o distinguem de outros tipos de ruído.

É importante salientar que, em algumas seções da dissertação, o termo “sinal” pode ser utilizado de forma mais ampla para se referir ao som emitido por embarcações em geral, incluindo tanto o sinal de interesse quanto o ruído indesejado. Essa utilização mais ampla visa facilitar a discussão e a compreensão dos conceitos, mas não compromete a distinção clara entre os dois tipos de som quando necessário.

Assim, o sinal emitido por uma embarcação, captado por um hidrofone, gera um conjunto de dados² com grande volume. O processamento direto desse conjunto para classificação utilizando algoritmos de ML pode ser inviável devido à grande quantidade de dados envolvidos. Para superar essa limitação, é comum evitar a alimentação completa do sinal ao algoritmo classificador. Em vez disso, utiliza-se uma representação reduzida dos sinais, conhecida como características (*features*) [11–13].

As características representam informações essenciais do sinal de áudio, resumindo suas propriedades relevantes para a classificação. O conjunto de características extraídas de um sinal forma um vetor de características, que é utilizado como entrada para o algoritmo classificador [14].

A escolha da metodologia de extração de características e do algoritmo classificador adequado, são cruciais para o desempenho da classificação [15–17]. Encontrar um método eficiente que extraia aspectos relevantes do sinal emitido pela embarcação, preservando a informação necessária para a classificação, é um grande desafio [5]. As características extraídas devem ser concisas, preservando as propriedades chave do sinal, permitindo uma redução substancial do tamanho dos dados sem comprometer a precisão da classificação [11].

Essa simplificação do sinal visa otimizar a temporalidade dos algoritmos de ML, tornando-os mais adequados para aplicações em tempo “quase” real. Em resumo, a extração de características atua como um processo de redução de dimensionalidade, transformando o sinal de áudio em uma representação mais compacta e eficiente para os algoritmos de ML. [6, 18].

1.2 Motivação

Tradicionalmente, a maioria dos extratores de características utilizados no contexto de classificação de embarcações, são os mesmos comumente usados na fala e pro-

²O termo "dados" utilizado nesta dissertação refere-se ao conjunto de amostras dos sinais acústicos interpretados computacionalmente e convertidos em vetores de entrada para algoritmos de machine learning.

cessamento de música como, por exemplo, a taxa de cruzamento zero (ZCR, *zero-crossing rate*), Coeficientes Cepstrais de Frequência Mel (MFCC, *Mel-Frequency Cepstral Coefficients*), raiz média quadrada de energia (RMS energy, *Root Mean Square Energy*), centroide espectral (*spectral centroid*), (*Chroma Feature*), fluxo espectral (*spectral flux*), codificação preditiva linear (LPC, *Linear Predictive Coding*) e coeficientes cepstrais de predição linear (LPCC, *linear prediction cepstral coefficients*) detalhados em SHARMA *et al.* [18].

No contexto da classificação e extração de características do ruído de embarcações marítimas, uma gama diversificada de metodologias de extração, empregando diversas técnicas de processamento de sinais, foi desenvolvida para enfrentar os desafios associados ao ambiente marinho. Por exemplo, XIN-XIN *et al.* [19] propuseram um algoritmo para extração de características de emissões sonoras submarinas baseado na Wavelet Packet Transform. Este método extrai características de energia em cada banda de frequência através de uma decomposição em quatro estágios do pacote wavelet do sinal, resultando em vetores de características de baixa dimensão.

KARBASI *et al.* [15] apresentaram um método de extração de características adequado para reconhecimento de sons aéreos que explora mudanças espectrais ao longo do tempo. Seu método divide segmentos de áudio de 4 segundos em vários quadros e extrai um vetor de Características Espectrais Dinâmicas (SDF, *Spectral Dynamic Features*) que representa o segmento inteiro. Este vetor é obtido aplicando um banco de filtros Mel ao espectro de frequência de cada quadro e calculando a Transformada Rápida de Fourier (FFT, *Fast Fourier transform*) em todos os quadros para cada saída de filtro. Embora esta abordagem capturasse variações de recursos e melhorasse a acurácia em alguns casos, o uso de segmentos longos dificulta aplicações em tempo real.

DAS *et al.* [7] utilizou uma abordagem cepstral para extração de características para mitigar a distorção do sinal causada pelo canal de propagação submarino raso. Os resultados demonstraram que o método de elevação cepstral reduziu significativamente os efeitos da distorção multipercusso. Além disso, os autores observaram que, no domínio cepstral, o sinal irradiado de uma embarcação marítima ocupa amplamente os índices cepstrais mais baixos, enquanto o canal submarino ocupa os índices mais altos, de modo que, para várias condições de alcance e profundidade, os dois podem ser separados. Essa separação facilita o processo de classificação, por permitir uma deconvolução adequada do sinal irradiado pela embarcação e a resposta ao impulso do canal variável no tempo.

LIU *et al.* [5] propuseram um método de extração de características baseado no espectro de linha do sinal de ruído do navio. Este método envolve análise de espectro de potência, seguida de aproximação a um espectro contínuo e separação de componentes espectrais. A forma de onda é então corrigida e tanto o número

quanto a intensidade do espectro da linha são incorporados como elementos do vetor de características. Embora alcance uma taxa de precisão superior a 80%, este método limita-se a classificar apenas duas classes de navios.

CHEN e XU [20] analisa os espectros de baixa frequência (LOFAR, *Low-Frequency Analysis and Recording*) e detecção de modulação de envelope no ruído (DEMON, *Demodulation of Envelope Modulation On Noise*) para extrair as características dos sinais acústicos submarinos. Embora os autores tenham desenvolvido uma nova metodologia, os resultados utilizam apenas dados modelados e não houve validação com dados experimentais.

TONG *et al.* [21] e LIAN *et al.* [22], por sua vez, introduziram abordagens para obter vetores de características por meio da extração dos MFCCs e dos Coeficientes Cepstrais de Frequência Gammatone (GFCC, *Gammatone Frequency Cepstral Coefficients*). Tais metodologias implementam um banco de filtros Mel e um banco de filtros Gammatone, respectivamente, como parte do processo de extração. TONG *et al.* [21] utilizam os MFCCs extraídos de três tipos de alvos submarinos medidos em ambientes reais, como entrada para o algoritmo de classificação K-vizinhos mais próximos (k-NN, *k-Nearest Neighbors*). Os autores realizaram experimentos que evidenciam um aumento na taxa de classificações corretas à medida que a ordem dos MFCCs (número de componentes do vetor de característica) é incrementada. Entretanto, esse incremento está associado ao aumento tanto no número de dados quanto na complexidade computacional do processo. LIAN *et al.* [22] extrai o vetor de características implementando modificações aos GFCCs originais para adequar ao ambiente submarino. Os experimentos são conduzidos com amostras coletadas de navios com uma relação sinal-ruído (SNR, *Signal-to-Noise Ratio*) variável, e mostrou que a acurácia do algoritmo decai com o aumento do ruído de fundo.

Quando se trata de caracterizar um sinal de áudio submarino deve-se considerar a natureza dinâmica e variável do sinal. Esta variação é resultado da combinação de múltiplos espectros de curta duração, cada um com diferentes características, seguindo algum padrão no tempo [7]. A maioria dos métodos de extração apresentados concentram-se nas características temporais e, ou, espectrais de um sinal dentro de um único quadro (*frame*), da ordem de milissegundos, onde cada *frame* é processado separadamente para extrair um vetor de características, deixando de oferecer informações sobre as variações das características entre *frames* consecutivos [15].

Este trabalho propõe um algoritmo modificado a partir de KARBASI *et al.* [15] para extrair as Características Espectrais Dinâmicas (vetor SDF) em segmentos de áudio mais extensos (da ordem de segundos), capturando os padrões da variação temporal das características entre *frames* consecutivos. A principal motivação para utilização do vetor SDF reside na exploração da variabilidade temporal. Como o

ruído emitido por embarcações não é estático, apresentando variação ao longo do tempo, as características extraídas em um instante de tempo (um único *frame*), não necessariamente serão as mesmas extraídas no instante de tempo seguinte. Dessa forma a representação de todo um sinal por um vetor de características extraído de um único *frame* não reflete com precisão a natureza dinâmica do ruído das embarcações.

A extração do vetor SDF em segmentos de áudio da orde de segundos permite a captura dessa variabilidade temporal, fornecendo uma visão mais completa e abrangente do comportamento do sinal. Ao considerar um intervalo temporal mais extenso, o algoritmo proposto mitiga a influência de eventos de curta duração, como ruídos transitórios ou sons de outras fontes, que podem distorcer a caracterização geral do ruído de embarcações.

Além disso, busca-se melhorar a acurácia frente aos métodos tradicionais abordados, fornecendo como entrada para os algoritmos de classificação um vetor de características, combinando o vetor SDF, que extrai características dinâmicas, com o vetor formado pelos coeficientes Mel cepstrais, que capturam características estáticas. Por fim, a partir da adaptação de um algoritmo já consolidado na acústica área, este estudo analisa sua utilização para classificação de embarcações, o que será explorado em detalhes na Seção 2.3.

1.3 Objetivos

O objetivo principal desta pesquisa é desenvolver e implementar um algoritmo extrator de características, capaz de capturar aspectos dinâmicos oriundos da variação temporal dos sinais submarinos de interesse e avaliar a acurácia da classificação utilizando diferentes extratores de características e um banco de dados composto por diferentes embarcações. Os objetivos específicos incluem:

- Investigar e adaptar métodos de extração de características espectrais dinâmicas para o contexto de classificação acústica submarina;
- Avaliar a eficácia de diferentes técnicas de aprendizado de máquina na classificação de sinais acústicos submarinos;
- Validar a metodologia proposta utilizando dados experimentais, considerando diferentes tipos de embarcações e condições ambientais.

1.4 Publicação

Uma publicação foi produzida durante a elaboração desta dissertação:

- *Improved Spectral Dynamic Features Extracted from Audio Data for Classification of Marine Vessels*: aceito para publicação na revista *Intelligent Marine Technology and Systems*.

1.5 Organização

Este trabalho está organizado em seis capítulos. O Capítulo 2 engloba uma revisão de conceitos fundamentais no campo da Acústica Submarina, essenciais para a compreensão deste estudo. Além disso, são apresentados conceitos relacionados ao processamento de sinais digitais, com ênfase na extração de características no domínio cepstral, seguido por uma revisão bibliográfica dos métodos de extração de características mais comumente encontrados na literatura, os quais se alinham com a metodologia proposta neste trabalho.

O Capítulo 3 descreve detalhadamente a metodologia utilizada para a extração das características espectrais dinâmicas e a formação do vetor de características combinado. No Capítulo 4, são descritos e organizados os conjuntos de dados empregados para validar este estudo.

O Capítulo 5 se dedica à apresentação dos resultados obtidos e à discussão relacionada a esses resultados. Finalmente, as conclusões do estudo e sugestões para pesquisas futuras são abordadas no Capítulo 6.

Capítulo 2

Referencial teórico

A classificação de sons submarinos emitidos por embarcações através da extração de características espectrais dinâmicas é um campo de pesquisa promissor com potencial para superar os métodos tradicionais. Para explorar esse potencial, este capítulo aprofunda os fundamentos da acústica submarina, fornecendo a base teórica necessária para o desenvolvimento da metodologia proposta. A compreensão dos princípios da propagação do som no oceano e das técnicas de processamento de sinais é fundamental para a extração eficaz de características que permitam a identificação precisa de diferentes tipos de embarcações. Além disso é discutido nas Seções 2.3 e 2.4, a teoria acerca do domínio cepstral e a obtenção dos coeficientes Mel Cepstrais, que compõe o processo de obtenção do vetor SDF. Por fim, a Seção 2.5 discute o uso de algoritmos de aprendizagem de máquina, em particular aqueles baseados em aprendizado supervisionado, para aplicação em sistemas de classificação de alvos acústicos.

2.1 Acústica submarina

A acústica submarina é um ramo da física que se dedica ao estudo da propagação do som na água e à interação das ondas sonoras com o guia de ondas submarino. A compreensão dos fenômenos acústicos submarinos é fundamental para diversas aplicações, como a comunicação subaquática, a localização de objetos submersos, o mapeamento do fundo do mar e o monitoramento ambiental.

No contexto deste trabalho, a acústica submarina assume um papel crucial na classificação de embarcações com base em seus padrões acústicos. As características do guia de ondas submarino, como temperatura, salinidade, pressão, presença de obstáculos, bem como fenômenos acústicos associados, influenciam significativamente a propagação do som e, conseqüentemente, os padrões acústicos captados por sensores subaquáticos.

2.1.1 Influência do meio na propagação da onda sonora

Ao contrário do ambiente aéreo, onde as condições tendem a ser mais homogêneas, o oceano apresenta características heterogêneas e variáveis em diferentes regiões. Essa heterogeneidade, combinada com os fenômenos acústicos mencionados na Seção 1.1, pode causar distorção e modificação do sinal emitido por uma embarcação antes de atingir o receptor.

Diversos fatores contribuem para essas variações na propagação do som na água, incluindo as condições da superfície da água, o relevo e tipo do fundo do mar, a composição da água e a profundidade. A rugosidade da superfície da água afeta a reflexão e a dispersão das ondas sonoras, enquanto o relevo do fundo do mar pode interferir na propagação do som, causando fenômenos de refração. Além disso, a velocidade do som na água é influenciada pela salinidade, temperatura e densidade da água, que variam de acordo com as condições locais [23, 24].

A velocidade média de propagação do som no oceano é aproximadamente 1500 m/s . No entanto, não se trata de um valor fixo estando sujeito a pequenas variações, condicionadas a estação do ano, hora do dia, profundidade local, posição geográfica, proximidade de rios e geleiras [25]. Existem mais de uma expressão matemática, estabelecidas empiricamente, que relacionam a velocidade de propagação do som no oceano (ou celeridade) com a pressão (variável com a profundidade), temperatura e salinidade [9, 23, 25, 26]. A simplificação dessa dependência proposta MEDWIN e CLAY [25] é dada pela Equação 2.1 :

$$c = 1449,2 + 4.6T - 0.055T^2 + 0.00029T^3 + (1.34 - 0.010T)(S - 35) + 0.016D \quad (2.1)$$

onde:

c = velocidade (m/s)

T = temperatura ($^{\circ}\text{C}$)

S = salinidade (partes por mil ‰)

D = profundidade (m)

Entre os fenômenos acústicos que desempenham um papel fundamental na propagação do som na água, destacam-se a refração, a reflexão e a atenuação. Esses fenômenos estabelecem uma variedade de caminhos percorridos pela onda sonora desde a fonte até o receptor.

Essa multiplicidade de caminhos provoca a dispersão temporal do sinal, ou seja, a chegada do sinal no receptor em diferentes instantes de tempo, distorcendo sua forma original. Enquanto no ar, o multicaminho geralmente não é um fator significativo devido à homogeneidade relativa do meio, na água, pode causar distorções que

confundem algoritmos de classificação [8, 23].

A refração ocorre devido à curvatura das ondas sonoras causada pela variação da velocidade do som com a profundidade, enquanto a reflexão envolve a mudança de direção da onda incidente, com o regresso total ou parcial da energia sonora, na fronteira entre dois meios com propriedades diferentes, como o fundo do mar ou a superfície da água. Por fim, a atenuação volumétrica refere-se à perda gradual da energia sonora ao longo da propagação, causada pela absorção e dispersão pelas moléculas da água. Esses fenômenos combinados contribuem para a complexidade da propagação do som na água e têm um impacto direto na detecção e classificação de fontes sonoras submarinas.

A heterogeneidade do guia de ondas submarino e os fenômenos acústicos afetam a classificação de embarcações de diversas maneiras:

- **Distorção do sinal:** O sinal emitido por uma embarcação pode ser distorcido e modificado durante a propagação, dificultando sua identificação.
- **Variabilidade das características acústicas:** As características acústicas de uma mesma embarcação podem variar em diferentes locais e condições, dificultando a criação de um modelo de classificação único.
- **Presença de ruído ambiental:** O ruído ambiental presente no oceano, como o gerado por correntes marítimas e vida marinha, pode mascarar os sons emitidos pelas embarcações.

A frequência do sinal acústico também desempenha um papel importante na propagação subaquática. A atenuação causada pela absorção na água varia de acordo com a frequência. Frequências baixas geralmente sofrem menor atenuação e propagam-se por maiores distâncias do que frequências altas.

Essa relação entre frequência e atenuação pode ser descrita por equações matemáticas empíricas, como a Equação 2.2, apresentada em JENSEN *et al.* [8]:

$$\alpha' \approx 3.3 \times 10^{-3} + \frac{0.11f^2}{1 + f^2} + \frac{44f^2}{4100 + f^2} + 3.0 \times 10^{-4}f^2 \quad (dB/km) \quad (2.2)$$

onde:

α é o coeficiente de atenuação em dB/km;

f é a frequência do sinal em kHz.

Além disso, a análise espectral do sinal, obtida por técnicas como a FFT, permite observar a distribuição de energia do sinal em diferentes frequências. O conhecimento das bandas de frequência, principalmente aquelas em que o sinal tem concentrada a

maior parte de sua energia, favorece por exemplo, o projeto de filtros de sinais que compõe alguns extratores de características.

É importante considerar também a geometria do guia de onda submarino. Em muitos casos, esses guias de onda assumem uma forma cilíndrica ou circular. Essa geometria não é apenas uma característica física, mas também influencia significativamente a propagação do som na água. Ao considerar o meio de propagação como um guias de onda com forma cilíndrica, por exemplo, as ondas sonoras podem ser direcionadas e guiadas ao longo do eixo do cilindro. Isso resulta em propriedades de propagação distintas em comparação com meios mais abertos. A forma cilíndrica do guia de onda pode impactar a reflexão e a dispersão das ondas sonoras, modificando sua trajetória e comportamento ao longo da propagação.

2.1.2 Considerações sobre o nível de pressão sonora (Sound Pressure Level - SPL) na tarefa de classificação

O Nível de Pressão Sonora (SPL), é uma medida que quantifica a intensidade do som em relação a um nível de referência. Ele é matematicamente definido como o logaritmo da razão entre o valor RMS da pressão sonora, medida em Pascal (Pa), e uma pressão de referência em Pascal, definido pela seguinte Equação 2.3 [24].

$$SPL = 20 \times \log_{10}\left(\frac{P_{rms}}{P_{ref}}\right) \quad (2.3)$$

Onde o SPL é expresso em dB *re* P_{ref} . Em acústica submarina o valor típico de referência é de 1 micropascal (μPa) [8, 24].

O SPL é um indicador da intensidade do som, mas não fornece informações sobre a frequência ou a forma da onda sonora. Devido à atenuação volumétrica, o SPL medido pelo receptor geralmente não corresponde ao SPL emitido pela fonte sonora. Níveis de SPL mais altos indicam sons mais intensos, enquanto níveis mais baixos indicam sons mais fracos. Na classificação de embarcações, o SPL é utilizado para avaliar a intensidade do sinal captado pelos sensores. O conhecimento do SPL permite ajustar os parâmetros do sistema de classificação para otimizar o desempenho, especialmente em ambientes com altos níveis de ruído ambiental.

2.1.3 Considerações para a extração de características

Considerando os desafios impostos pelas características do guia de ondas submarino, a extração de características para a classificação de embarcações deve levar em consideração os seguintes aspectos:

- **Robustez à distorção:** As características devem ser robustas à distorção do sinal, de forma a preservar as informações relevantes para a classificação.
- **Adaptabilidade às condições variáveis:** As características devem ser adaptáveis às diferentes condições do guia de ondas submarino, permitindo a classificação precisa em diversos cenários.
- **Discriminação de ruído:** As características devem ser capazes de discriminar o ruído ambiental dos sons emitidos pelas embarcações.
- **Análise espectral:** A análise da distribuição de energia do sinal em diferentes frequências (informação espectral) obtida por técnicas como a FFT deve ser explorada para auxiliar na extração de características discriminativas.

Ao desenvolver um método de extração de características que atenda a esses requisitos, é possível superar muito dos desafios impostos pela heterogeneidade do guia de ondas submarino e alcançar uma melhora na performance de classificação de embarcações.

2.1.4 Características do ruído emitido por embarcações

A assinatura acústica de uma embarcação é uma combinação complexa de sons gerados por diversas fontes, cada uma com suas características e dependências. As principais fontes de ruído em embarcações podem ser categorizadas em três grupos principais[1]:

- Hélice;
- Máquinas; e
- Processos Hidrodinâmicos.

A hélice é uma das principais fontes de ruído, especialmente em altas velocidades. O giro da hélice causa o ruído de cavitação, um fenômeno que ocorre quando a pressão da água ao redor das pás da hélice cai abaixo da pressão de vapor, formando bolhas que colapsam rapidamente, gerando sons de alta intensidade. A frequência do ruído da hélice está diretamente relacionada à velocidade de rotação da hélice, e o espectro do ruído é predominantemente de banda larga, com componentes em diversas frequências [27].

As máquinas a bordo, como motores, geradores e bombas, também contribuem significativamente para o ruído emitido pela embarcação. As vibrações mecânicas dessas máquinas são transmitidas para o casco e, em seguida, para a água, gerando sons de banda estreita em frequências características de cada máquina. O nível e

a frequência do ruído das máquinas variam de acordo com a carga e o estado de operação das máquinas[1, 28].

Os processos hidrodinâmicos também contribuem para o ruído irradiado pelas embarcações. Entre eles, destaca-se o fluxo turbulento da água ao redor do casco e das superestruturas, gerando ruído de banda larga, especialmente em velocidades elevadas. A cavitação, que também pode se manifestar em outras partes do casco, como bordas afiadas ou protuberâncias, produzindo sons de alta intensidade em frequências específicas. Por fim, a entrada e saída de ar nos compartimentos do navio, especialmente em mares agitados, geram ruído de ventilação.

É importante ressaltar que a intensidade e a frequência desses sons variam de acordo com diversos fatores, como velocidade da embarcação, condições do mar, estado de operação e condições do casco e das máquinas. Por exemplo, o ruído gerado pelo gerador diesel de serviço de um navio pode criar uma série de harmônicos com amplitudes e frequências independentes da velocidade da embarcação. Já as hélices, especialmente em altas velocidades, geram cavitação, que produz um ruído de espectro contínuo.

A distribuição espectral do ruído emitido por embarcações é complexa e mesmo um componente discreto (ruído de banda estreita) gerado por uma única fonte de ruído é irregular e variável[1]. Embarcações do mesmo modelo podem ter assinaturas acústicas similares, mas não idênticas, devido a variações em suas condições operacionais, características do casco, hélice e outros fatores [28].

Uma abordagem comum para identificar os ruídos submarinos de embarcações em movimento é analisar seus espectros. Essa análise permite isolar componentes discretos nos espectros associados ao funcionamento de mecanismos e equipamentos a bordo, juntamente com o espectro de banda larga que reflete o trabalho da hélice em cavitação, o fluxo turbulento em tubulações e ventiladores[1]. Essas informações espectrais podem ser utilizadas como características discriminativas no processo de classificação de embarcações.

2.2 Processamento digital de sinais para extração de características

Conforme ilustrado no diagrama de blocos da Figura 2.1, o processamento para extração de características analisados neste trabalho pode ser dividido em quatro etapas : segmentação, pré-processamento, análise espectral e extração de características. Tais etapas estão descritas nas Seções 2.2.1, 2.2.2, 2.2.3 e 2.2.4 respectivamente.

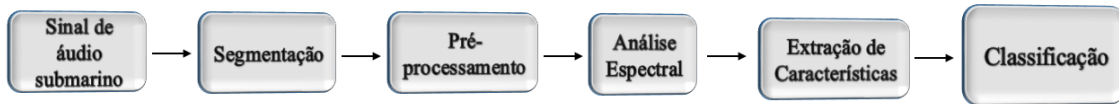


Figura 2.1: Diagrama de blocos do processamento digital de sinais para extração de características. Um sinal de áudio digitalizado captado por hidrofones passa por etapas de processamento até a extração dos vetores de características que serão submetidos a algoritmos de classificação.

2.2.1 Segmentação

A captação de emissões acústicas por embarcações gera arquivos de áudio com durações variáveis, desde segundos até horas, conforme o contexto da gravação. É possível extrair um único vetor de características, a partir de cada arquivo, porém de forma a padronizar e aumentar o conjunto de vetores de características, é realizada uma etapa de segmentação, subdividindo cada arquivo de áudio em intervalos menores. Cada segmento resultará na criação de um vetor de características, assim um mesmo arquivo pode gerar k vetores de características, sendo k o número de segmentos.

A entrada da segmentação é um arquivo de áudio digitalizado, representado por $x[l]$, sendo $l = 1, 2, \dots, L$, onde L é o número total de amostras. Este arquivo digitalizado é obtido através da conversão de um sinal acústico captado por um hidrofone em um conjunto de valores numéricos. A frequência de amostragem fs determina a resolução temporal do sinal digitalizado, definido pela máxima frequência que será utilizada no processamento posterior.

A segmentação consiste na subdivisão do arquivo de áudio $x[l]$ em segmentos menores, denominados $s[n]$, onde $n=1,2, \dots, Ns$. O número de amostras de cada segmento Ns é definido pelo produto entre a duração do segmento desejado em segundos (t) e da frequência de amostragem, ou seja, $Ns = t \times fs$. Cada segmento s representa uma porção específica do arquivo de áudio original, com duração fixa e características próprias.

Aumentar o número de vetores de características pode aprimorar o treinamento de modelos de aprendizado de máquina, desde que se mantenha um equilíbrio entre a riqueza de informações e a simplicidade do modelo.

Essa abordagem de segmentação é essencial para lidar com a variabilidade nas configurações de uma mesma embarcação e a heterogeneidade do ruído emitido em diferentes intervalos de tempo. Ao segmentar o sinal, padroniza-se o tamanho dos vetores de entrada, facilitando assim a comparação e análise de diferentes gravações [29]. Além disso, ao aumentar o número de vetores de características disponíveis para análise, pode-se melhorar o treinamento de modelos de aprendizado de má-

quina, desde que se mantenha um equilíbrio entre a riqueza de informações e a simplicidade do modelo. Isso ocorre porque a diversidade de dados aumenta, gerando um número maior de exemplos e diminuindo a possibilidade de *overfitting*¹. Dessa forma, torna-se menos provável que o modelo memorize padrões específicos dos dados de treinamento e, em vez disso, generalize para novos dados.

A segmentação também permite capturar nuances temporais nas emissões acústicas, o que pode ser crucial para identificar falhas ou anomalias na operação da embarcação. Ademais, ao focarmos em segmentos específicos do sinal acústico que contêm informações relevantes, podemos reduzir o impacto de ruídos de fundo.

2.2.2 Pré-processamento

O objetivo principal da etapa de pré-processamento é preparar o sinal para a análise espectral subsequente, aprimorando a relação sinal-ruído (SNR) e mitigando interferências irrelevantes [6]. Um tratamento adequado do conjunto de dados durante essa etapa envolve a aplicação de técnicas como remoção de ruídos, normalização, tratamento de valores ausentes e seleção de características relevantes. Essas técnicas visam garantir que os dados estejam em um formato adequado e de qualidade para serem utilizados pelos algoritmos de aprendizado de máquina. A qualidade do pré-processamento exerce uma influência significativa no desempenho do processo de classificação. Quando os dados são tratados adequadamente, removendo-se ruídos e destacando características relevantes, os modelos de ML têm uma base mais sólida para aprender e realizar classificações precisas. Isso, por sua vez, contribui para a extração de características mais precisas e robustas, levando a um melhor desempenho na classificação. Além disso, prepara o sinal para a transição do domínio do tempo para o domínio da frequência, facilitando a aplicação de técnicas de análise espectral e modelagem.

As etapas de pré-processamento abrangem três etapas principais conforme apresentado na Figura 2.2.

Cada segmento $s[n]$ é pré-processado como se segue:

- **Pré-ênfase** - A pré-ênfase refere-se a um procedimento de filtragem que enfatiza as frequências mais altas, compensando a atenuação natural dessas frequências durante a propagação do som na água, e aumenta a SNR. Esse processo visa criar um espectro de frequências mais balanceado, tornando o sinal mais propício para análises espectrais subsequentes [30].

¹O sobreajuste (no inglês, *overfitting*) refere-se a um problema em que um modelo se ajusta muito bem aos dados de treinamento, mas não generaliza bem para novos dados ou dados de teste. Em outras palavras, o modelo aprende não apenas os padrões subjacentes nos dados de treinamento, mas também o ruído e as variações aleatórias presentes nesses dados.

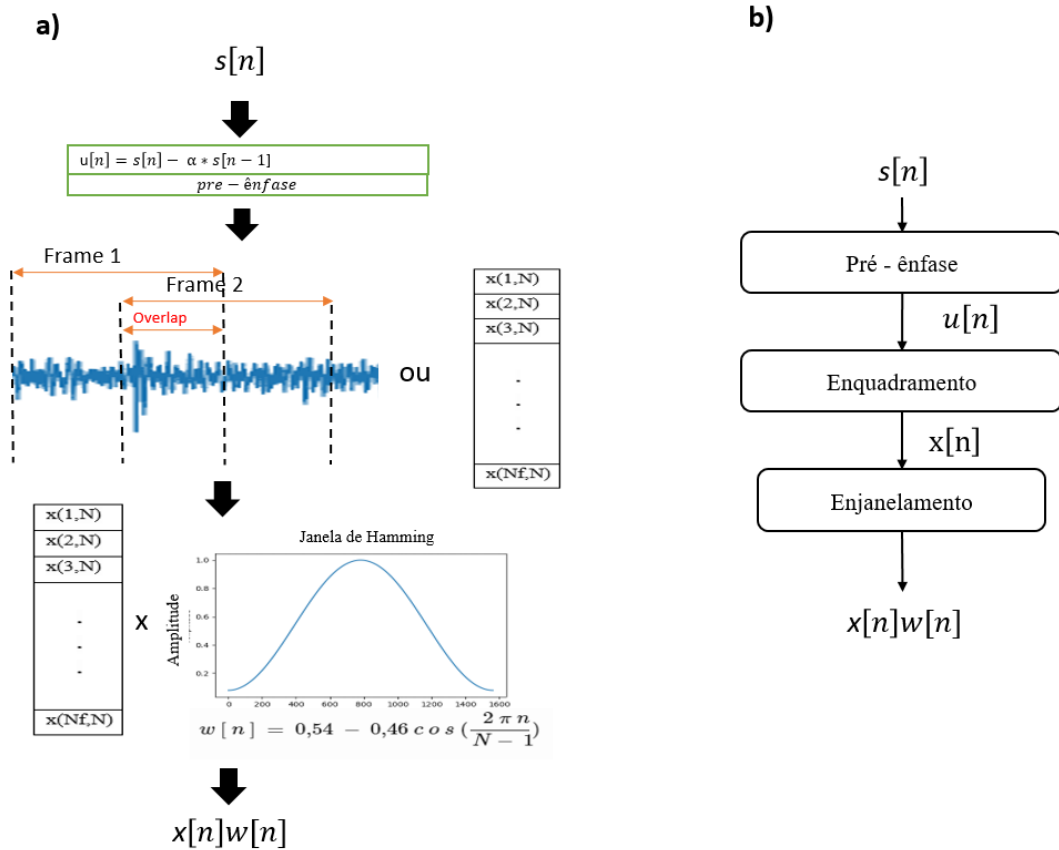


Figura 2.2: a) Representação gráfica ilustrando as etapas do pré-processamento. b) Fluxograma simplificado das etapas do Pré-processamento. Inicia-se com um segmento do sinal de áudio tendo como saída o segmento enjanelado.

O segmento $s[n]$ é pré-enfatizado por meio da aplicação de um filtro digital de 1.^a ordem cuja função de transferência do filtro é dada por :

$$H(z) = 1 - \alpha \cdot z^{-1}, \quad 0.9 < \alpha < 1 \quad (2.4)$$

Onde α é o coeficiente de pré-ênfase, que controla a intensidade da amplificação das altas frequências. Valores mais elevados de α indicam um realce mais acentuado [31].

O sinal pré-enfatizado é representado na Figura 2.2 por $u[n]$ resultado da equação: $u[n] = s[n] - \alpha \cdot s[n-1]$.

- **Enquadramento** - No processo de enquadramento, o segmento pré-enfatizado $u[n]$ é dividido em quadros menores, chamados de *frames*. Mesmo em um segmento, da ordem de segundos, o sinal varia lentamente. Dividir o segmento em *frames*, com tamanhos típicos de 20-40 ms [32] e com uma taxa de sobreposição de 50% do *frame* posterior sobre o *frame* anterior [33, 34], significa aproximar o sinal a um processo quase estacionário antes da aplicação da Transformada

de Fourier (FT, *Fourier Transform*) [30]. Isso garante que as características do sinal sejam capturadas de forma contínua ao longo do tempo.

O conjunto de *frames* extraídos de um segmento pré-enfatizado $u[n]$, é dado por: $\{x_i[n], 1 \leq i \leq NF, 1 \leq n \leq N\}$, sendo NF o número de *frames* e N o número de amostras do *frame*. A Figura 2.2 trás também uma representação matricial desse conjunto de *frames* onde cada linha da matriz corresponde ao conjunto de amostras de cada *frame*.

- **Enjanelamento** - Com intuito de minimizar as descontinuidades entre o início e o fim de cada *frame* causado pelo enquadramento [31, 33], cada *frame* é multiplicado por uma função janela. O tipo de janela mais utilizado no contexto de reconhecimento de sinais é a janela de *Hamming*, definida como [34]:

$$w[n] = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (2.5)$$

A escolha dos parâmetros que se ajustam ao conjunto de dados em análise é fundamental para um bom desempenho de um sistema de classificação. Por exemplo, um valor muito baixo de α pode não compensar suficientemente a atenuação das altas frequências, enquanto um valor muito alto pode amplificar o ruído. O tamanho e a sobreposição dos *frames* também influenciam a qualidade do processamento. *Frames* muito curtos podem não capturar características transitórias importantes, enquanto *frames* muito longos podem violar a suposição de estacionariedade. A sobreposição de *frames* garante continuidade na análise e evita perda de informação.

A Figura 2.2 sintetiza as etapas envolvidas no pré-processamento, onde um segmento de sinal é pré-enfatizado, enquadrado e enjanelado tendo como saída $x[n]w[n]$ que na sequência será submetida a uma análise espectral.

2.2.3 Análise Espectral

A análise espectral é uma técnica utilizada para extrair características e consiste em examinar a distribuição de energia de um sinal em relação à frequência, revelando informações significativas sobre as componentes espectrais presentes no sinal. A estimativa do espectro é realizada analisando pequenos intervalos de tempo (*Short-Time intervals*). Esses intervalos (*frames*) são dimensionados de forma a garantir estacionariedade, como abordado na Seção 2.2.2, ao mesmo tempo em que possuem tamanho suficiente para abranger pelo menos um ciclo da menor frequência de interesse [35].

A estimativa do espectro para cada *frame* é obtida por meio da aplicação repetida da transformada Discreta de Fourier (DFT, *Discrete Fourier transform*) em cada um deles. Esse procedimento é conhecido como transformada de Fourier de Tempo Curto

(STFT, *Short-Time Fourier transform*). O resultado da *STFT* são os coeficientes espectrais, que podem ser utilizados como vetor de características e submetidos aos algoritmos de classificação e, neste caso específico, o bloco da Figura 2.1 referente ao processo de extração de características deixa de existir tornando-se parte integrante da análise espectral. Esses coeficientes são números complexos com informação de magnitude e fase [31].

Os coeficientes espectrais de potência (PSC, *power spectral coefficients*), também são utilizados como vetores de características extraídas de um sinal, calculados tomando o valor absoluto da *STFT* e elevando-o ao quadrado. Assim, os coeficientes espectrais de potência para cada *frame* são resultados da Equação 2.6:

$$p[k] = \frac{\left| \sum_{n=1}^N x[n]w[n]e^{-j\frac{2\pi nk}{N}} \right|^2}{N}; \quad 1 \leq k \leq N \quad (2.6)$$

Alguns autores, ao abordarem a metodologia de extratores clássicos, consideram a submissão dos coeficientes espectrais de frequência ou os coeficientes espectrais de potência a um banco de filtros [21, 22, 31, 32, 36, 37] como um processo dentro da análise espectral. Nesse processo, cada *frame* de entrada passa por um banco de filtros passa-faixa, nos quais as frequências centrais abrangem as frequências de interesse. Isso resulta em uma saída em que a componente espectral relativa à frequência central é realçada [35]. As escalas de frequências centrais, são tradicionalmente inspiradas na percepção do ouvido humano, como exemplo : a escala Mel[21, 31, 37] e a escala Bark[38]. O tipo de filtro, o formato, a escolha das frequências centrais e a faixa de interesse, variam de acordo com conjunto de dados que se deseja classificar.

2.2.4 Extração de características

A extração de características se baseia em extrair dos sinais acústico informações relevantes que permitam distinguir as diferentes classes de embarcações. Considerando a complexa natureza dos sinais submarinos, a escolha da metodologia de extração de características é fundamental para o sucesso do sistema de classificação.

Na seção anterior, foi mencionado que o resultado da análise espectral pode ser utilizado como vetor de características, eliminando a necessidade de outros processos de extração. No entanto, a literatura apresenta metodologias de extração de características, como as abordadas na Seção 1.2, que utilizam etapas adicionais de processamento com diferentes técnicas de processamento de sinal. Essas técnicas, representadas na Figura 2.1 pelo bloco "Extração de Características", visam superar desafios específicos do problema de classificação de embarcações mencionados por DAS *et al.* [7], tais como a variabilidade nos sinais emitidos por embarcações, a heterogeneidade do canal de propagação, que induz flutuações aleatórias em fase,

amplitude e frequência, além do ruído ambiente no receptor.

Outras abordagens adotam processos de extração de características que não utilizam a análise espectral. Elas incluem a análise temporal, que se concentra nas características do sinal ao longo do tempo, como a energia total, a energia média e a variância. Além disso, há a análise de tempo-frequência, que combina a análise espectral com a análise temporal. Isso permite a extração de características que capturam tanto a informação em diferentes faixas de frequência quanto a variação do sinal ao longo do tempo.

Por fim, o resultado da extração de característica é um vetor ou um conjunto de vetores de características que representa um sinal de áudio digitalizado com um grande número de amostras em um vetor de dimensão inferior e que será utilizado como entrada para os algoritmos de classificação. Essa redução de dimensionalidade possibilita uma redução do tempo de processamento computacional e favorece a interpretabilidade do conjunto de dados pelos modelos de classificação baseados em ML [11].

2.3 O Cepstrum

O cepstrum, anagrama da palavra spectrum (espectro), é obtido aplicando a FT ao logaritmo do espectro de potência de um sinal. Essa operação matemática revela informações sobre mudanças lentas e estruturas harmônicas no espectro, sendo valiosa para análise de sinais periódicos. É amplamente utilizado em processamento de fala [31, 36], recentemente aplicado na detecção de eco-localização de mamíferos marinhos [39, 40] e na análise de assinaturas acústicas de explosões subaquáticas [41].

A Figura 2.3 exibe as etapas da transformação de um sinal do domínio do tempo para o domínio cepstral. A primeira etapa envolve o cálculo do módulo ao quadrado da FT de um sinal no domínio do tempo, resultando no espectro de potência. Este espectro destaca as componentes espectrais de potência e suas respectivas amplitudes.

Em seguida, ocorre a etapa de aplicação do logaritmo, onde o espectro de potência é convertido para decibéis (dB) por meio do logaritmo das amplitudes das componentes espectrais de potência. A aplicação do logaritmo destaca as diferenças de amplitude em uma ampla faixa de frequências.

Por fim, o logaritmo da magnitude do espectro de potência é submetido a uma Transformada Inversa de Fourier (IFT, *Inverse Fourier Transform*), chamada de Transformada Cepstral. Essa transformada mapeia o espectro de frequência de volta para o domínio do tempo, mas agora o tempo é representado pelas "quefrecys", que estão relacionadas à frequência fundamental do som. A transformada cepstral

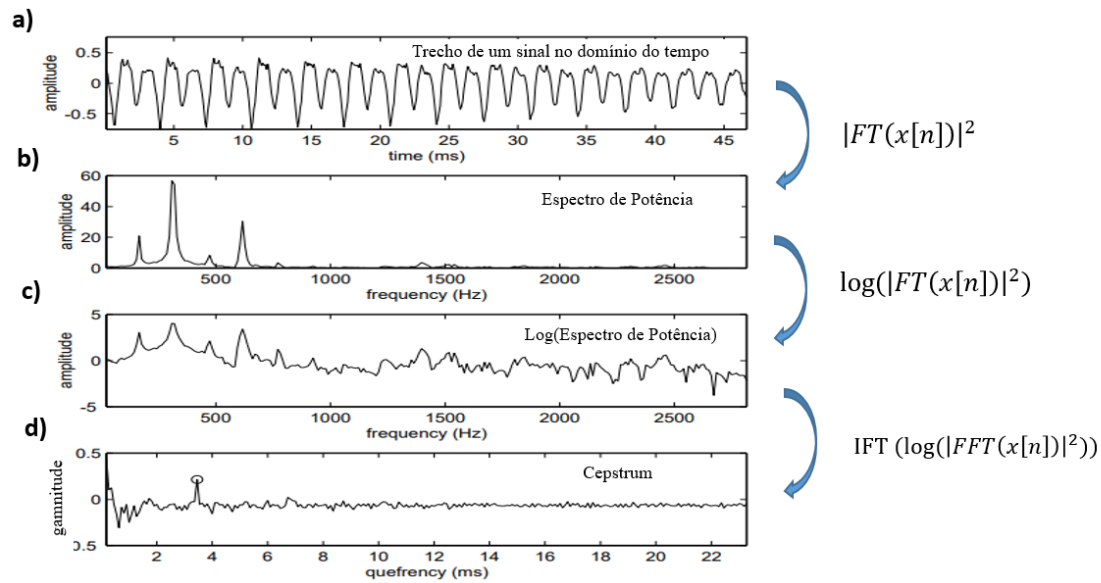


Figura 2.3: Transformação do sinal do domínio do tempo para o domínio Cepstral. a) Trecho de sinal no domínio do tempo. b) Espectro de potência obtido ao calculando o módulo ao quadrada da FT do sinal. c) Espectro de potência logaritimizado. d) Cepstrum de potência, com o pico representando a frequência fundamental do trecho do sinal avaliado. Fonte: adaptado de GERHARD [42]

captura informações sobre as características espectrais do sinal e as representa em coeficientes cepstrais (CC , *cepstral coefficient*). Esses coeficientes são essenciais para a análise e extração de padrões temporais e periódicos no sinal original.

O cepstrum revela como as frequências do som se distribuem ao longo do tempo. Picos proeminentes representam as frequências fundamentais e seus harmônicos do sinal original. A posição desses picos no tempo (quefrency) está relacionada à frequência fundamental do sinal e a magnitude dos picos indica a força relativa dos harmônicos [43]. Ou seja, as quefrencys fornecem informações sobre o intervalo de tempo entre os eventos de frequência no sinal. No caso da voz humana, por exemplo, o cepstrum mostrará picos em frequências que correspondem às vogais e consoantes. No caso de um motor, o cepstrum mostrará picos em frequências relacionadas à rotação do motor e outros componentes mecânicos [44].

Embora o cepstrum seja empregado na concepção de extratores de características, suas aplicações vão além disso. Ele também pode ser usado para detectar atrasos de tempo entre sinais diretos e refletidos (multipercursos) possibilitando estimativas precisas de alcance e velocidade das embarcações [45, 46], identificar características específicas dos motores das embarcações e até mesmo classificar o tamanho das embarcações pela análise dos picos [44].

2.4 Coeficientes Mel Cepstrais

Os MFCCs são uma representação cepstral do espectro de potência de um sinal com a incorporação de um banco de filtros Mel. O processo de obtenção dos *MFCCs* difere-se dos *CC* pela aplicação de um banco de filtros digitais ao espectro real do sinal, antes da aplicação da função logarítmica.

Os coeficientes espectrais de potência obtidos da Equação 2.6, são submetidos a um conjunto de filtros passa-banda com envelopes triangulares, cujos centros são espaçados de acordo com a escala Mel, chamados de banco de filtros Mel e ilustrados na Figura 2.4.

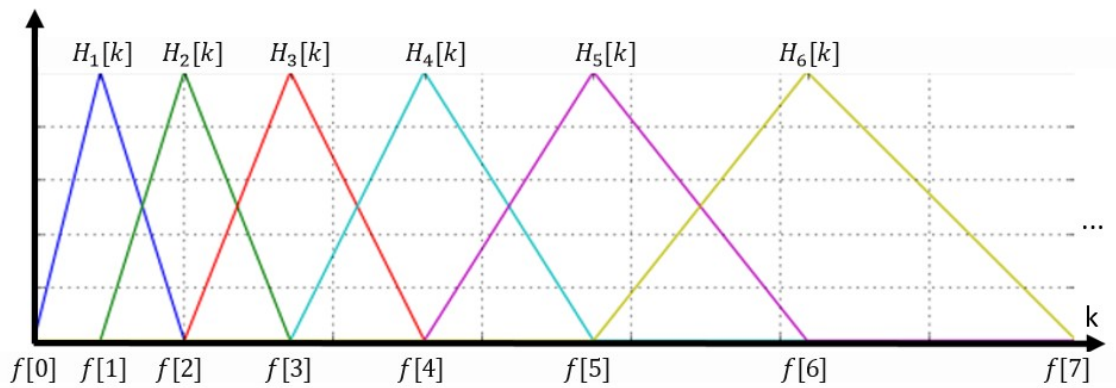


Figura 2.4: Representação de um banco de filtros Mel. f são as frequências inicial, central e final de cada filtro, determinando a faixa de frequências analisadas pelo filtro. H são as funções de transferência que definem os coeficientes de cada filtro. A saída de cada filtro é um valor que representa a energia presente na banda de frequência do filtro.

O filtro mel converte o espectro de potência linear em um espectro de potência não-linear na escala de frequência Mel [21]. A escala Mel é uma escala psicoacústica que explora a relação de percepção da frequência fundamental entre dois tons simulando a percepção do ouvido humano de um tom puro. Abaixo de 1 kHz, a escala Mel mantém uma relação aproximadamente linear com a frequência em Hz, proporcionando uma resolução mais alta para frequências mais baixas, e, acima de 1 kHz, adota um espaçamento logarítmico tendo uma menor resolução nas frequências mais altas [30]. Esta é uma característica desejada no contexto de classificação de embarcações, uma vez que, o ruído irradiado por embarcações é mais significativo em frequências abaixo de 1 kHz [1, 4, 47]. O mapeamento entre a escala de frequência real, em Hz, e a escala de frequências mel, está expresso na equação (2.7).

$$f_{mel} = 2595 \log \left(1 + \frac{f_{hz}}{700} \right) \quad (2.7)$$

Aqui, f_{mel} representa a frequência em Mel e f_{hz} denota a frequência em Hz.

Esta equação converte essencialmente frequências lineares para a escala Mel usando uma função logarítmica. O fator constante 2595 dimensiona a conversão de modo que as frequências mais baixas sejam mapeadas para valores de Mel mais baixos e vice-versa.

O banco de filtros Mel é comumente implementado digitalmente como se segue: O primeiro passo consiste na escolha dos limites do conjunto de filtros, geralmente entre zero Hz e a frequência de Nyquits ($f_{Nyquist} = fs/2$). Os limites são convertidos para escala mel pela Equação 2.7. Em seguida define-se a quantidade m de filtros desejados, que serão gerados uniformemente espaçados entre eles ao longo dos limites na escala Mel [48]. Para um conjunto de 26 filtros Mel, por exemplo, o intervalo compreendido entre os limites é dividido em 28 pontos, sendo que as frequências Mel centrais de cada filtro corresponde a metade do intervalo entre dois pontos consecutivos. Por fim as frequências Mel são convertidas de volta para frequências em Hertz.

O calculo do banco de filtros convencional obedece a Equação 2.8 a seguir:

$$H_m(k) = \begin{cases} 0 & \text{se } k < f(m-1) \\ -\frac{k-f(m-1)}{f(m)-f(m-1)} & \text{se } f(m-1) \leq k \leq f(m) \\ -\frac{f(m+1)-k}{f(m+1)-f(m)} & \text{se } f(m) \leq k \leq f(m+1) \\ 0 & \text{se } k > f(m+1) \end{cases} \quad (2.8)$$

onde $H(k)$ define as inclinações de cada filtro, m é o número do filtro e $f()$ é a lista de frequências Mel espaçadas, calculadas na Equação 2.7. O filtro m começa quando o filtro $m-1$ atinge seu pico, e o filtro m atinge seu pico à medida que o filtro $m-1$ retorna a zero [48].

Os bancos de filtros Mel são utilizados para calcular a quantidade de energia que existente em várias regiões de frequência. As energias de saída desses filtros, podem ser interpretadas como subamostragens do espectro, onde o primeiro filtro indica a quantidade de energia próxima a 0 Hertz, sendo este filtro inicial muito estreito. A Figura 2.5 exhibe um banco composto por 26 filtros Mel convencionais, o espectro de potência de um sinal qualquer, bem como o efeito da aplicação dos filtro sobre o espectro de potência.

As energias de saída do banco de filtros Mel representam o espectro mel do espectro de potência $p[k]$, calculado multiplicando o espectro de potência por cada um dos filtros triangulares, utilizando a equação (2.9):

$$e[m] = \sum_{k=1}^N p[k]H_m[k]; \quad 1 \leq m \leq M \quad (2.9)$$

O resultado é a média do espectro em torno da frequência central calculado por

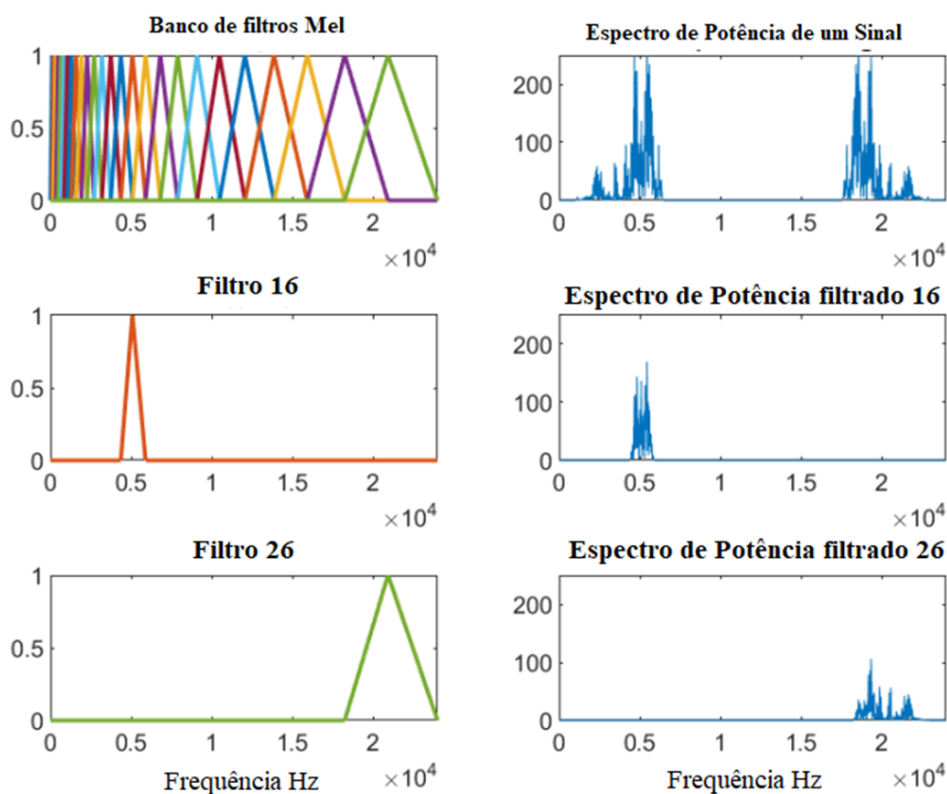


Figura 2.5: Exemplo de um banco de filtros Mel e o efeito da aplicação do filtro no espectro de potência. O espaçamento nas bandas de frequências mais baixas destaca a ênfase dada aos coeficientes extraídos nessa faixa. Fonte : adaptado de SURANYI e NGUYEN [48]

cada filtro. M é o número de filtros Mel e N o número de amostras do *frame*.

O processo de obtenção dos MFCCs envolve, inicialmente, submeter o resultado obtido pela Equação 2.9, que representa a energia de saída do banco de filtros Mel, a uma operação logarítmica, o que ajuda a comprimir a informação e a destacar diferenças significativas em várias faixas de frequência. Em seguida, realiza-se a transformação para o domínio cepstral por meio da aplicação da Transformada Cepstral. A forma mais comum de transformada cepstral é a Transformada Discreta do Cosseno (DCT, *Discrete Cosine Transform*), que descreve um sinal original como uma combinação linear de componentes de cosseno com diferentes frequências e amplitudes.

DAVIS e MERMELSTEIN [37], considerado um pioneiro no estudo dos MFCCs, justifica a escolha da DCT com base em experimentos que evidenciaram a capacidade dessa transformada em concentrar a maior parte das informações relevantes nos termos iniciais. Esse aspecto proporciona uma representação mais compacta dos dados espectrais, resultando em ganhos de eficiência computacional e economia de espaço de armazenamento. Outra vantagem significativa é a capacidade da DCT em descorrelacionar os coeficientes, ou seja, torná-los estatisticamente independentes

entre si. Isso implica que os coeficientes resultantes da DCT carregam informações discriminativas únicas, contribuindo para a redução da redundância nos dados. Além disso, a DCT oferece a possibilidade de reduzir a dimensão do vetor de características MFCC ao descartar termos de ordem superior. Essa prática é respaldada pelo fato de que as características mais distintivas geralmente estão contidas nos primeiros termos da DCT [19]. O resultado da aplicação da DCT são os Coeficientes Cepstrais de Frequência Mel (MFCCs), expressos por:

$$c[n] = \sum_{m=1}^M \log(e[m]) \cos\left(\frac{\pi(m-0.5)n}{M}\right); \quad n = 1, 2, \dots, p. \quad (2.10)$$

Onde $c[n]$ são os Coeficientes Mel Cepstrais e p o número de pontos da DCT e também o número de MFCCs extraído de cada *frame*.

As fases do processo de extração do vetor de características MFCC é apresentado na Figura 2.6 e inicia pelo pré-processamento do vetor de entrada $s[n]$, análise espectral pelo cálculo do quadrado do módulo da STFT, seguido da passagem por um banco de filtros Mel e concluído aplicando a DCT sobre o logaritmo das energias de saída dos filtros Mel.

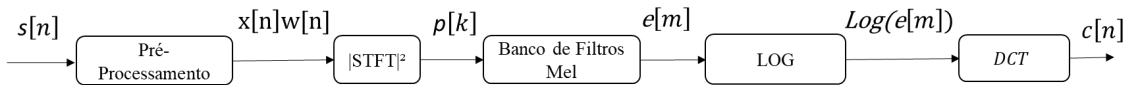


Figura 2.6: Fluxograma do processo de obtenção dos MFCCs. Inicia-se com o segmento do sinal, tendo como resultado os MFCCs extraídos de um *frame*

2.5 Aprendizagem de máquina

A identificação manual de alvos submarinos a partir de suas características requer um esforço humano significativo, o que impõe limitações no atendimento às demandas práticas [49]. Nos últimos anos, o surgimento dos métodos de ML forneceu suporte teórico, direcionando o foco dos pesquisadores na área de acústica submarina para o desenvolvimento de abordagens automatizadas para classificação de alvos submarinos, utilizando tais métodos [11].

O *ML* é um conjunto de técnicas de inteligência artificial, cujo intuito é fazer com que as máquinas (computadores) aprendam sozinhas usando dados fornecidos e, assim, consigam realizar previsões com uma dada incerteza. Uma perspectiva alternativa para compreender as técnicas de *ML* é entendê-las como abordagens de análise de dados que procuram automatizar a formulação de modelos analíticos. Esses algoritmos têm a habilidade de aprender de maneira iterativa a partir dos

dados, permitindo que os computadores descubram “*insights*” sem a necessidade de programação explícita sobre onde ou o quê procurar nas informações [50], estabelecendo uma regra discriminativa automática que possa prever a classe mais provável de uma nova observação. No entanto, uma clara limitação das técnicas baseadas em *ML* é que eles são orientados por dados e, portanto, exigem grandes quantidades de dados para teste e treinamento [11]. Além disso o desempenho do sistema é determinado pela qualidade dos dados. Como tal, é importante construir bases de dados consolidadas com métricas bem definidas que garantam a confiabilidade dos dados [12].

Atualmente, os métodos de *ML* têm mostrado melhor desempenho do que os métodos tradicionais [51]. A interseção entre o campo da acústica submarina e o *ML* oferece uma abordagem inovadora para compreender e explorar fenômenos sonoros, proporcionando avanços substanciais em áreas como reconhecimento e classificação automática de embarcações [52, 53], detecção e classificação de espécies marinhas com base em seus sons característicos [54, 55], localização de fonte acústica submarina [56], separação de fonte de canal único [57] e assim por diante. Estudos apontam para a integração do método de extração de características de sinais acústicos do ambiente submarino com um modelo de *ML*, utilizando as características do domínio cepstral do sinal como entrada para os algoritmos de *ML* [5, 17], e, amplamente abordado neste trabalho.

2.5.1 Terminologias e introdução a conceitos básicos de *ML*

Nesta subseção são apresentadas algumas terminologias e introduzidos conceitos necessários para o entendimento e aplicação das técnicas de *ML* na tarefa de classificação. O primeiro deles refere-se a constituição e divisão dos dados, categorizados como conjunto de treinamento e conjunto de teste/validação. O conjunto de treinamento é usado para criar um modelo² adequado para o aprendizado de máquina, onde a acurácia do modelo é aumentada por meio de um processo repetitivo de atualização do modelo. O modelo é então validado por meio do conjunto de testes [12].

Existem duas categorias principais de aprendizado de máquina: aprendizado supervisionado e aprendizado não supervisionado. A distinção fundamental entre esses dois tipos reside no fato de que o aprendizado supervisionado utiliza uma verdade básica, ou seja, já conhecemos previamente os valores de saída (rótulos) para nosso conjunto de dados de entrada (*dataset*). O conjunto de dados de entrada neste caso, é apresentado na forma de pares ordenados (entrada — saída desejada). Portanto, o objetivo do aprendizado supervisionado é desenvolver uma função que, a

²Modelo refere-se a uma representação matemática ou computacional de um sistema ou processo que é aprendido a partir dos dados de treinamento. Esse modelo é usado para fazer previsões ou tomar decisões sobre novos dados que não foram vistos durante o treinamento.

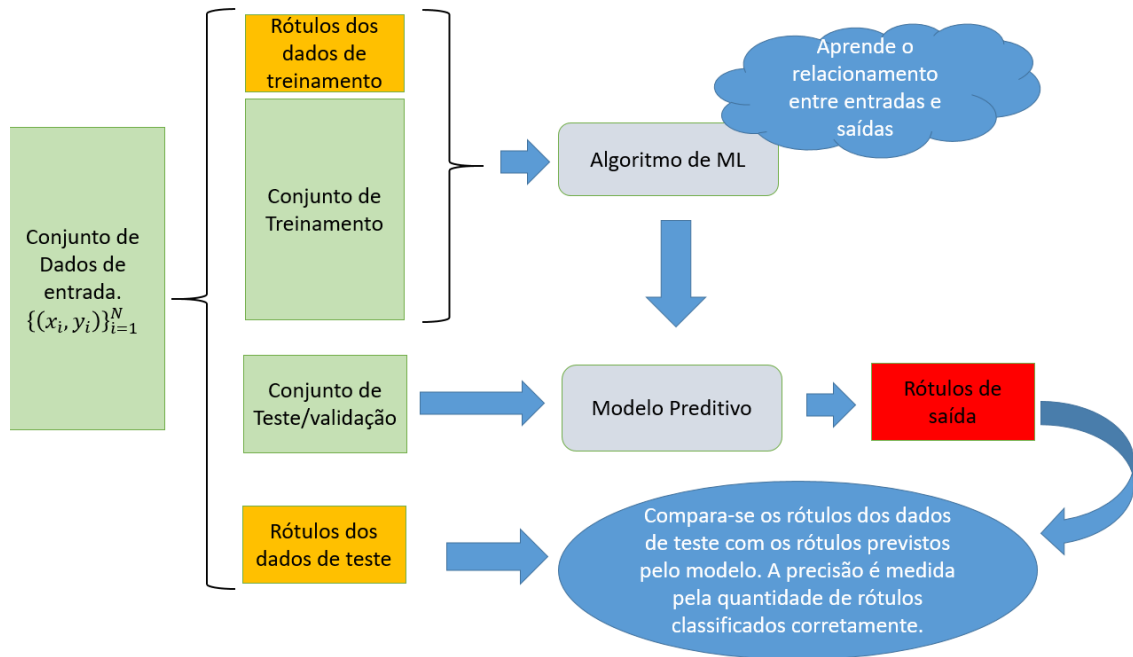


Figura 2.7: Exemplo de modelo de aprendizado supervisionado. O conjunto de dados é dividido em conjunto de treinamento e conjunto de teste/validação. Também é fornecido os rótulos dos dados de treinamento para construção do modelo. O modelo recebe os dados de teste/validação e faz previsões dos rótulos de saída. A acurácia da classificação é estimada comparando os rótulos previstos (classificação mais provável) pelo modelo com os rótulos reais

partir de exemplos de entrada e saídas desejadas (rotuladas previamente), estabeleça uma relação entre entrada esperada e saída observável em um novo conjunto de dados de entrada. Em contrapartida, o aprendizado não supervisionado não possui saídas rotuladas, e a tarefa é descobrir novas estruturas ou características que se destaquem dentro dos dados [11, 58]. No modelo supervisionado o conjunto de dados de entrada é mapeado em N pares de entradas/saídas $\{(x_i, y_i)\}_{i=1}^N$ onde [12]:

- x são as características (também chamado de atributos e no inglês *features*), que são as variáveis de entrada que caracterizam o sinal e que serão utilizados na determinação das saídas. x pode ter uma estrutura complexa, como uma imagem ou um sinal no tempo [12]. A dimensão de x varia conforme o número de características extraídas.
- y são os rótulos, que são as classes e também o que se espera como predição do modelo.

A Figura 2.7 ilustra um exemplo generalizado de um modelo de aprendizagem supervisionado onde o conjunto de dados previamente rotulado é dividido em conjunto de treinamento, utilizado para construção do modelo, e conjunto de teste,

necessário para estimar a acurácia da classificação. Os rótulos de saída do conjunto de teste, previstos pelo modelo, são comparados com os rótulos previamente estabelecidos. Pode ser observado também a Figura 2.8, que traz a distribuição de dois vetores de características, reduzidos a 2 características principais (X1 e X2) em um plano bidimensional. Neste gráfico, cada ponto corresponde a um vetor de características (X1,X2) extraído de um sinal e as cores são os rótulos/classe desses sinais.

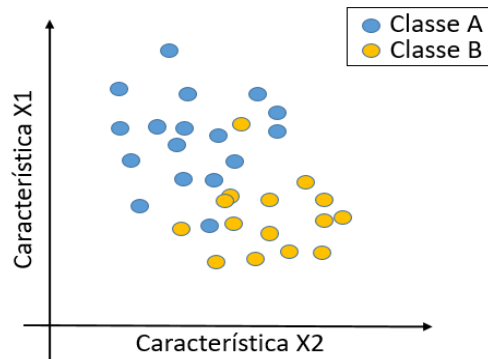


Figura 2.8: Distribuição de um conjunto de dados de entrada previamente rotulados em um espaço 2-D. Cada círculo corresponde a um sinal distinto, reduzido a duas características (X1 e X2) que os definem. As cores são os rótulos, sendo o círculo azul os sinais da classe A e o círculo amarelo os sinais da classe B.

A Figura 2.9 apresenta os conceitos referentes a criação de um classificador de forma simplificada. Neste trabalho o conjunto de dados de entrada fornecido a máquina para criação do modelo são os conjuntos de características (vetor de característica) de n dimensões extraídos de um sinal de áudio submarino, equivalente a uma linha de x_i ; atrelado a essa linha um rótulo/classe referência y_i , representa os tipos de embarcação/ruído ambiente a qual o sinal pertence.

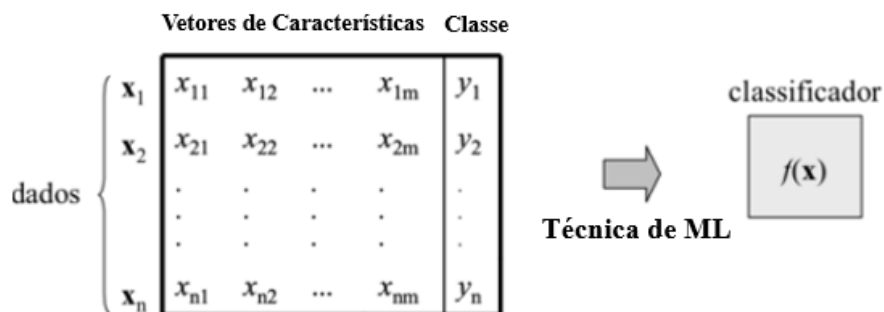


Figura 2.9: Criação de um classificador utilizando Aprendizado de Máquina (ML) Supervisionado. A matriz de dados são as entradas para criação do modelo (classificador). Fonte: adaptado de JÚNIOR *et al.* [59]

2.5.2 Algoritmos de classificação supervisionado

Nesta subseção são descritos, dentre os diversos algoritmos de aprendizado supervisionado disponíveis, alguns dos mais comuns. São eles: *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)*, *Random Forest (RF)* e *Decision Tree (DT)*.

Máquina de Vetor de Suporte (SVM, *Support Vector Machine*.)

O algoritmo conhecido como Máquina de Vetor de Suporte (SVM), uma técnica de aprendizado supervisionado, é capaz de lidar com problemas de classificação binária ou multiclasse. Na abordagem multiclasse, cada classe é tratada individualmente com um SVM dedicado. Seu princípio fundamental está enraizado na teoria de aprendizagem estatística, visando a classificação de conjuntos de dados em espaços multidimensionais por meio da criação de um hiperplano de separação (superfície de decisão). Esse hiperplano é posicionado de forma a agrupar os dados com características semelhantes do mesmo lado conforme ilustrado na Figura 2.10. O algoritmo busca identificar os pontos mais próximos às margens de cada classe, chamados de vetores de suporte, e calcula a distância da margem. A otimização do SVM busca maximizar essa distância para encontrar o hiperplano ideal que separa as classes de forma mais eficaz. Para isso, utiliza uma variedade de funções núcleos (*kernel*), como linear, Gaussiana ou função de base radial (RBF, *radial basis function*), polinomial e sigmoide [60].

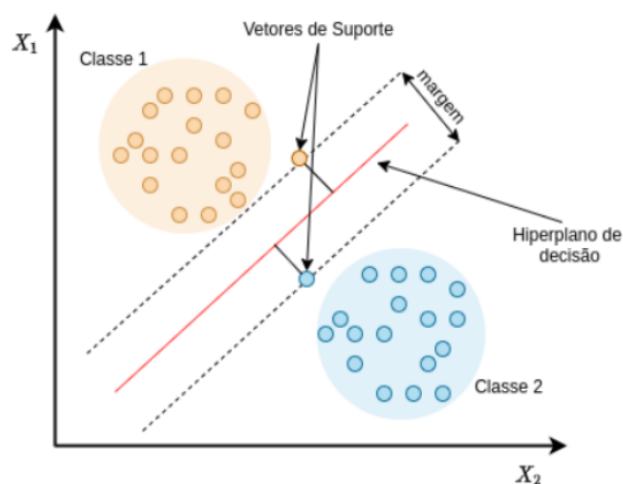


Figura 2.10: Estrutura geral do algoritmo de Máquina de Vetor de Suporte (SVM, *Support Vector Machine*). A estrutura geral do algoritmo de Máquina de Vetor de Suporte (SVM, *Support Vector Machine*) envolve a busca por amostras de dados chamadas vetores de suporte, que delimitam o hiperplano de separação entre as classes. A seleção desses vetores de suporte é realizada com o objetivo de maximizar a margem de separação entre as classes. Fonte : BARBOSA *et al.* [60]

A escolha do núcleo ou *kernel* possibilita a construção de diferentes SVMs. O

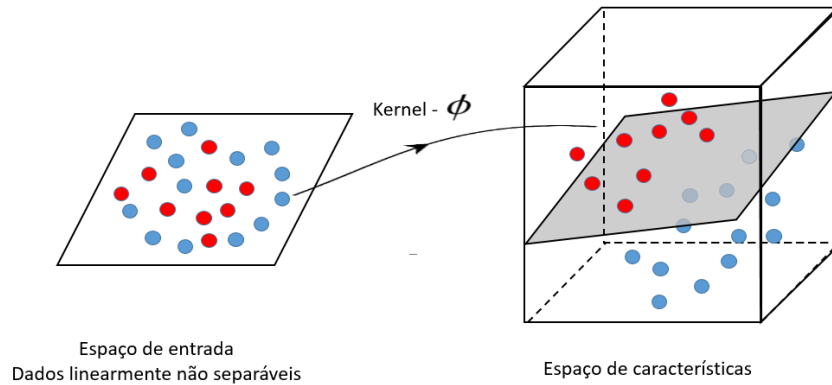


Figura 2.11: Transformação de um problema não linearmente separável em um problema linearmente separável utilizando a função *kernel* Φ .

kernel tem por finalidade transformar o espaço de dados tornando um problema não linearmente separável em linearmente separável. Esse processo envolve o mapeamento dos dados de entrada em um espaço de dimensão superior, onde é mais provável que sejam linearmente separáveis, permitindo assim a construção de um hiperplano de separação. A Figura 2.11 mostra a aplicação da função *kernel* Φ no espaço de entrada preenchido por dados linearmente não separáveis, elevando os dados a uma dimensão superior onde é possível traçar um hiperplano de separação.

K vizinhos mais próximos (KNN, *K-Nearest Neighbors*)

O KNN é um algoritmo simples e intuitivo que classifica um ponto de dados com base nas classes de seus vizinhos mais próximos no espaço de características. A ideia principal por trás do KNN é que pontos de dados semelhantes tendem a pertencer à mesma classe. O valor de K , que representa o número de vizinhos considerados, é um hiperparâmetro crucial que afeta o desempenho do algoritmo.

O k -vizinhos mais próximos insere-se no grupo da aprendizagem baseada em instâncias³, em que o conjunto de treino é simplesmente armazenado e o processo de classificação passa por comparar um novo conjunto de dados (conjunto de teste) com os dados de treino devidamente classificados, não há a construção de um modelo, como acontece para outros classificadores [61].

O algoritmo KNN compara instâncias calculando a distância entre vetores ou dados no espaço euclidiano. A ideia principal por trás do KNN é classificar um novo ponto de dados com base na classe predominante dos k pontos de dados mais próximos a ele no espaço de características como pode ser observado na Figura 2.12.

Diversas métricas estão disponíveis para calcular a distância entre pontos nesse

³A terminologia "instância" em ML refere-se a um exemplo individual de dados usado para treinar, validar ou testar um modelo. Em outras palavras, uma instância é uma observação específica que contém atributos ou características relevantes para o problema em questão.

espaço, incluindo Euclidiana, Hamming, Manhattan, Minkowski, Chebyshev, entre outras. O método mais comum é o cálculo da distância euclidiana dada pela Equação 2.11, que calcula a distância entre duas instâncias P_i e Q_j no espaço n-dimensional.

$$d(P_i, Q_j) = \sqrt{\sum_{k=1}^n (P_{ik} - Q_{jk})^2}, \quad (2.11)$$

onde P_{ik} e Q_{jk} para $k = 1, \dots, n$ são respectivamente, o vetor com as n características do dado a ser classificado e o vetor que contém as informações referentes aos dados existentes (dados de treino).

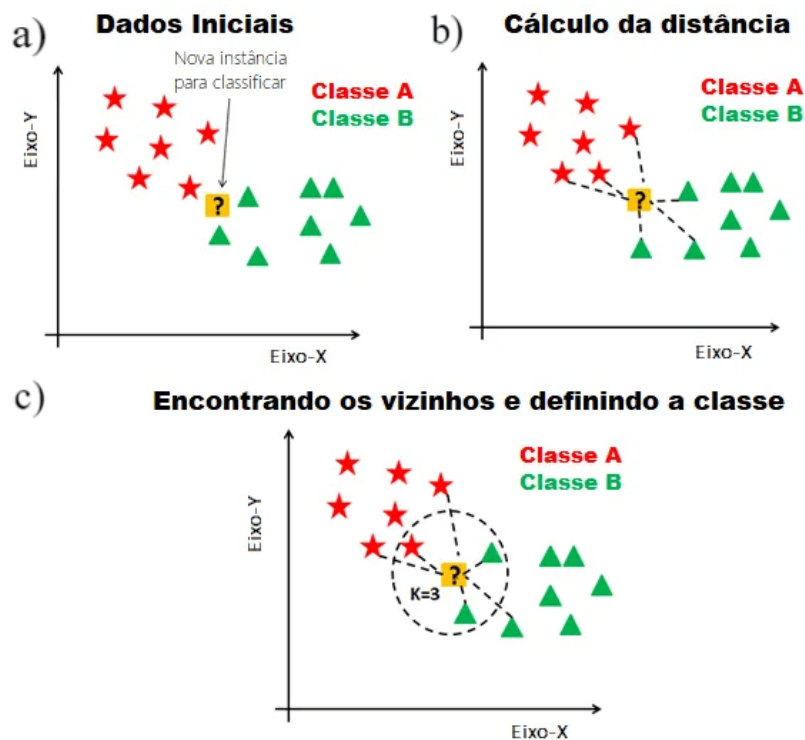


Figura 2.12: Fluxograma de classificação feito pelo KNN. O primeiro gráfico a) refere-se a todos dados iniciais, incluindo o que se deseja classificar, dispostos no espaço. O gráfico b) mostra o cálculo da distância euclidiana entre o novo dado e os dados já classificados. O gráfico c) mostra que para um número estipulado de 3 vizinhos aqueles que estão mais próximos são os da classe B, logo o novo dado será agrupado nesta classe. Fonte: REMIGIO [62]

Na Figura 2.12, inicialmente temos todos os dados dispostos no espaço euclidiano, incluindo o dado que se deseja classificar. O primeiro passo é definir o valor de k , que representa o número de vizinhos mais próximos a serem considerados para determinar a classe do dado (instância) que se deseja classificar. Em seguida, o algoritmo realiza o cálculo da distância euclidiana entre o dado que se deseja classificar e todos os outros dados já classificados onde a classe já é conhecida. Com todas as distâncias calculadas, os K elementos mais próximos, com menor distância, são selecionados e suas respectivas classes são analisadas, conforme representado no gráfico

da Figura 2.12 c). Por fim, a classe que obteve maior representatividade entre os K vizinhos mais próximos será usada para definir a classe do novo dado [62].

Árvore de Decisão (DT, *Decision Tree*)

As árvores de decisão são um dos algoritmos de classificação mais simples e interpretáveis. Elas dividem repetidamente o conjunto de dados em subconjuntos menores com base nas características mais significativas, procurando otimizar algum critério de pureza em cada divisão (como o ganho de informação e entropia). As decisões são representadas por um fluxograma de ramificação, estrutura semelhante a uma árvore, compostas por nós, ramos (ou ligações) e folhas (ou nós terminais) o que facilita a interpretação do modelo [62].

Assim como um fluxograma, a árvore de decisão estabelece nós (decision nodes) que se relacionam entre si por uma hierarquia. Os nós são as estruturas que armazenam as informações a respeito dos dados, onde são armazenados os testes e suas determinadas respostas em relação a uma característica. Existe o nó-raiz (maior nível hierárquico), que é o mais importante, e os nós-folha (leaf nodes), que são os resultados finais. No contexto de ML, o nó raiz é uma das características da base de dados e o nó-folha é a classe ou rótulo que será gerado como resposta. De modo geral, dado um vetor de característica, o nó raiz recebe a característica principal e com base em critérios de divisão como ganho de informação e entropia, por exemplo, as demais características vão se alocando em cada nível da árvore, buscando formar uma base de dados cada vez mais homogênea.

As variáveis, ganho de informação e entropia, dizem respeito à desorganização e falta de uniformidade nos dados. Quanto mais alta a entropia, mais caóticos e misturados estão os dados. Quanto menor a entropia, mais uniforme e homogênea está a base. Uma vez construída, a árvore de decisão pode ser usada para fazer previsões para novos exemplos, seguindo os caminhos da árvore com base nos valores de atributo desses exemplos.

Floresta Aleatória (RF, *Random Forest*)

O *Random Forest* é um algoritmo de conjunto que combina várias árvores de decisão individuais para criar um modelo mais robusto e geral. Cada árvore de decisão é treinada em uma amostra aleatória do conjunto de dados (chamada de amostra bootstrap) e, durante a previsão, a classe é determinada pela maioria das votações das árvores individuais [63].

O princípio de funcionamento pode ser resumido nas seguintes etapas:

- **Treinamento:** Cada árvore de decisão é treinada em um subconjunto aleatório do conjunto de dados original (amostra *bootstrap*).

- **Previsão:** Durante a fase de previsão, uma nova amostra é passado por todas as árvores de decisão no conjunto.
- **Agregação:** Para uma tarefa de classificação, uma votação majoritária – ou seja, a variável categórica mais frequente – produzirá a classe prevista.

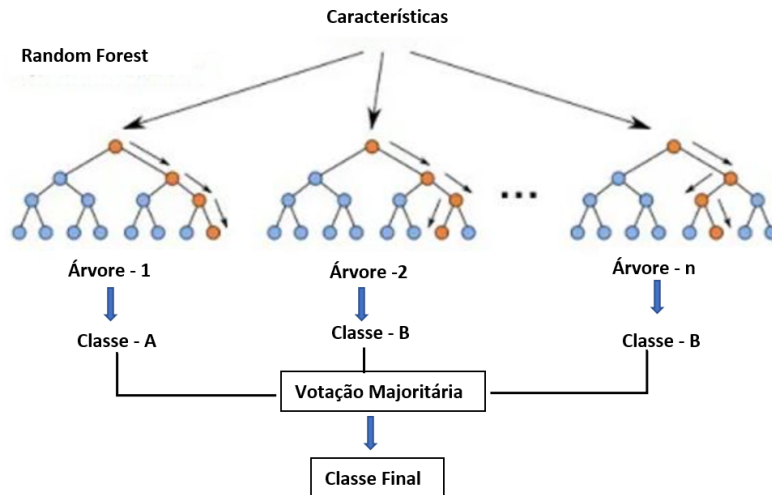


Figura 2.13: Simplificação do Algoritmo Floresta Aleatória. Este algoritmo é uma composição de árvores de decisão. Cada árvore recebe uma parte dos elementos do vetor de características, enquanto as demais características são alocadas em cada nível dessa árvore. O último nó representa a classe prevista. Assim, a decisão final do algoritmo é determinada pela classe prevista em um maior número de árvores de decisão. Fonte: adaptado de KOEHRSEN [64]

2.5.3 Método de validação cruzada "deixar-um-fora" (*Leave-One-Out*)

Para avaliar a eficácia desses algoritmos, é essencial realizar uma validação do modelo. Como já mencionado, um mecanismo comum de validação é a divisão do conjunto de dados em um conjunto de treinamento e um conjunto de teste/validação. O conjunto de treinamento é usado para treinar o modelo, enquanto o conjunto de teste é usado para avaliar a performance do modelo em dados não vistos. A utilização de mecanismos de validação cruzada também permite o ajuste dos parâmetros do modelo, como o valor de K no KNN ou o número de árvores no Random Forest, de forma recursiva, ou seja, tais parâmetros são alterados e a taxa de classificação correta avaliada. O ajuste ótimo desses parâmetros é fundamental para evitar o sobreajuste (*overfitting*).

A validação cruzada "deixar-um-fora" (LOOCV, *Leave-One-Out Cross-Validation*) é uma técnica especializada de validação cruzada que utiliza cada

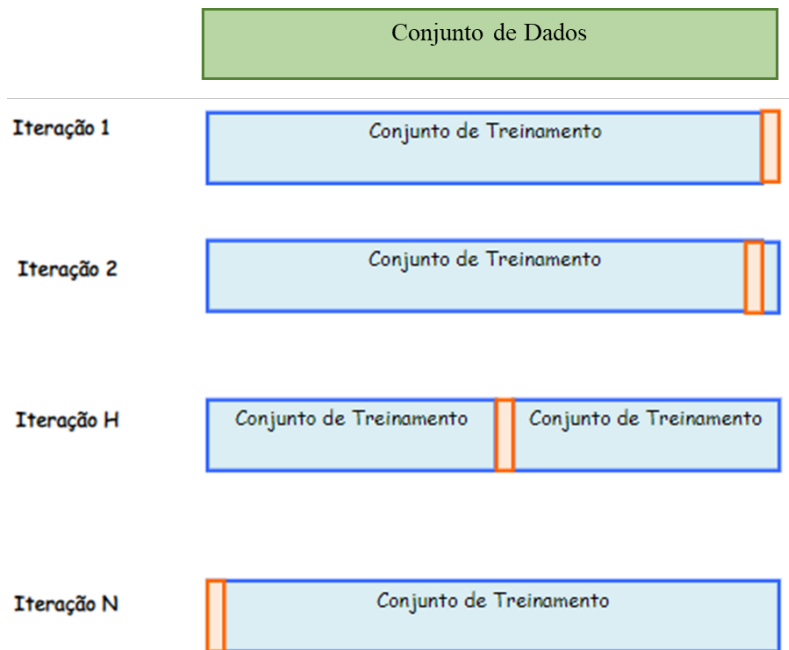


Figura 2.14: Método de validação cruzada "deixar-um-fora" (*Leave-One-Out*). Cada amostra de dado a ser testada é retirada do conjunto uma a uma enquanto as demais são utilizadas para treinar o modelo. Em laranja o dados testado na respectiva iteração, e N é o número total de amostras (vetor de característica) do conjunto de dados.

exemplo de dados como conjunto de teste uma vez, enquanto os demais exemplos são utilizados como conjunto de treinamento [65]. Trata-se de um processo iterativo, em que a cada iteração retira-se do conjunto de dados o dado a ser validado e os demais são utilizados para treinar o modelo. Na iteração seguinte esse dado é reintegrado ao conjunto de dados e retira-se outro dado para ser validado, havendo um novo treinamento do modelo com os dados restantes. Esse processo é repetido até que todo o conjunto de dados seja testado, como pode ser visto na Figura 2.14

A utilização da LOOCV, permite que cada linha de dados tenha a oportunidade de ser incluída no conjunto de teste, garantindo assim uma avaliação abrangente e representativa do desempenho do modelo. Isso resulta em uma estimativa mais robusta e menos tendenciosa, uma vez que todos os exemplos são utilizados tanto para treinamento quanto para teste [66]. O LOOCV é particularmente útil quando há uma quantidade limitada de dados disponíveis, pois maximiza a utilização do conjunto de dados existente. No entanto, para conjuntos de dados muito grandes, com dezenas ou centenas de milhares de exemplos, o LOOCV pode não ser apropriado devido ao alto custo computacional envolvido na construção de vários modelos.

Uma das métricas de desempenho de um classificador é a acurácia medida pela quantidade de previsões corretas em relação ao total de amostras do conjunto de dados, conforme a Equação 2.12:

$$\text{Acurácia} = \frac{\text{Número de previsões corretas}}{\text{Total de previsões}} \quad (2.12)$$

Nesta equação, o número de previsões corretas representa o total de amostras classificadas corretamente pelo modelo, enquanto o total de amostras é o número total de exemplos no conjunto de dados. A acurácia fornece uma medida geral da habilidade do classificador em fazer previsões corretas sobre o conjunto de dados.

Além da acurácia, outra métrica importante é a precisão, que mede a proporção de verdadeiros positivos⁴ em relação ao total de exemplos classificados como positivos pelo modelo. Em outras palavras, a precisão indica a habilidade do classificador em evitar fazer falsas previsões positivas. A fórmula para calcular a precisão é dada por:

$$\text{Precisão} = \frac{\textit{Verdadeiros Positivos}}{\textit{Verdadeiros Positivos} + \textit{Falsos Positivos}} \quad (2.13)$$

Onde os verdadeiros positivos representam o número de exemplos positivos corretamente classificados pelo modelo, e os falsos positivos representam o número de exemplos negativos erroneamente classificados como positivos. A precisão é especialmente útil quando a minimização de falsos positivos é fundamental para a tarefa em questão.

⁴o termo “positivo” refere-se à classe ou categoria que o modelo está tentando identificar ou prever. Por exemplo, em um problema de classificação de navios quanto ao tipo, a classe "positiva" pode ser representada pelos navios do tipo cargueiro que desejamos identificar.

Por outro lado, o termo "negativo" refere-se à classe oposta àquela que está sendo identificada. No exemplo anterior, os navios que não são cargueiros seriam considerados exemplos "negativos"

Capítulo 3

Metodologia

Este capítulo descreve a metodologia de extração do vetor de característica combinado, empregado como entrada para os algoritmos de classificação. O algoritmo de extração e de classificação proposto neste trabalho foi desenvolvido utilizando a linguagem de programação Python¹.

Na Seção 3.1 é apresentada a visão macro da arquitetura do algoritmo de extração de características proposto. A Seção 3.2 aborda as etapas utilizadas na análise espectral e a Seção 3.3 apresenta as etapas envolvidas no processo de extração de características.

3.1 Arquitetura do algoritmo de extração de características proposto

Uma versão simplificada da arquitetura proposta em nosso estudo para o processamento, análise espectral e extração de características dos dados acústicos submarinos é exibida no fluxograma apresentado na Figura 3.1.

O sinal de áudio submarino é segmentado e pré-processado conforme detalhado nas Seções 2.2.1 e 2.2.2. O tamanho do segmento adotado foi de quatro segundos, justificado experimentalmente na Seção 5.2.4 e os parâmetros de pré-processamento foram selecionados conforme a configuração do banco de dados utilizados nas simulações, sendo detalhados no Capítulo 4.

A metodologia propõe uma análise espectral em duas etapas englobando a obtenção do espectro de potência e a filtragem do espectro por um banco de filtros Mel e a etapa de extração de características compreendendo a extração dos MFCCs, a extração do vetor SDF e a combinação desses dois vetores.

A classificação do sinal de áudio é realizada submetendo o vetor de característica combinado, obtido seguindo a metodologia proposta, à análise de um algoritmo de

¹<https://www.python.org/downloads/>

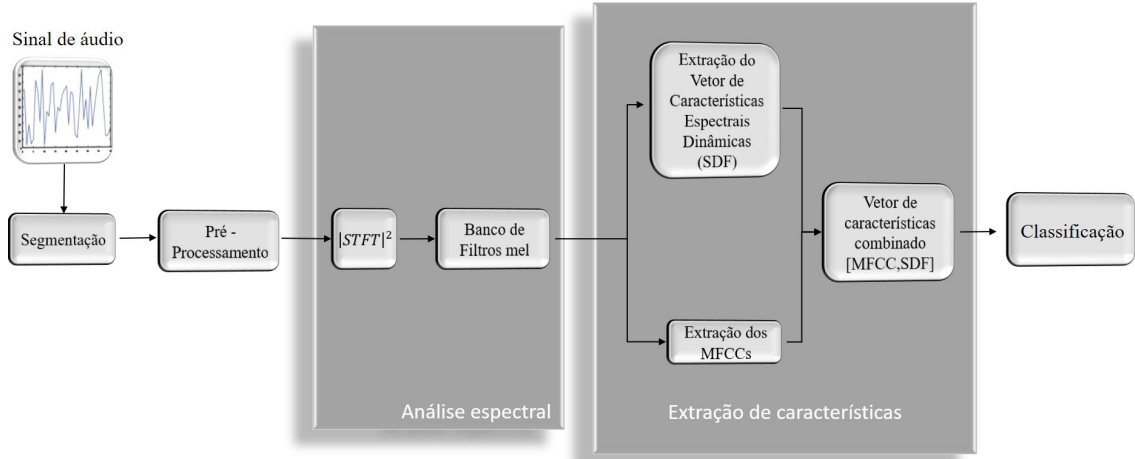


Figura 3.1: Arquitetura do algoritmo de extração de características proposto utilizando a combinação do SDF com MFCCs. A entrada é um sinal de áudio e o resultado é a classificação dos sinais presentes no arquivo de áudio.

classificação.

O diagrama apresentado na Figura 3.2 detalha todo o processo de obtenção do vetor de características combinado desde a entrada do sinal de áudio. Esse diagrama sintetiza as etapas da metodologia proposta e servirá como referência ao longo das seções subsequentes.

3.2 Análise espectral proposta

A análise espectral proposta consiste em duas etapas distintas. A primeira etapa envolve a obtenção dos coeficientes espectrais de potência, conforme detalhado na Seção 2.2.3. Para isso, aplica-se a $|STFT|^2$ ao conjunto de *frames* janelados resultante do pré-processamento, conforme a Equação 2.6 e exibido na etapa de Pré-processamento da Figura 3.2. O resultado é a Matriz P (3.1), onde cada linha representa os coeficientes espectrais de potência de um *frame* em um segmento do sinal. N representa o número de amostras por *frame* e NF o número total de *frames*.

$$P = \begin{bmatrix} p_1[1] & \cdots & p_1[k] & \cdots & p_1[N] \\ \vdots & \ddots & \vdots & & \\ p_i[1] & \cdots & p_i[k] & \cdots & p_i[N] \\ \vdots & & \vdots & \ddots & \\ p_{NF}[1] & \cdots & p_{NF}[k] & \cdots & p_{NF}[N] \end{bmatrix}; \quad 1 \leq i \leq NF, \quad 1 \leq k \leq N \quad (3.1)$$

Na segunda etapa, a Matriz P é submetida a um banco de filtros Mel projetado no contexto da classificação de embarcações marítima. O banco de filtros Mel

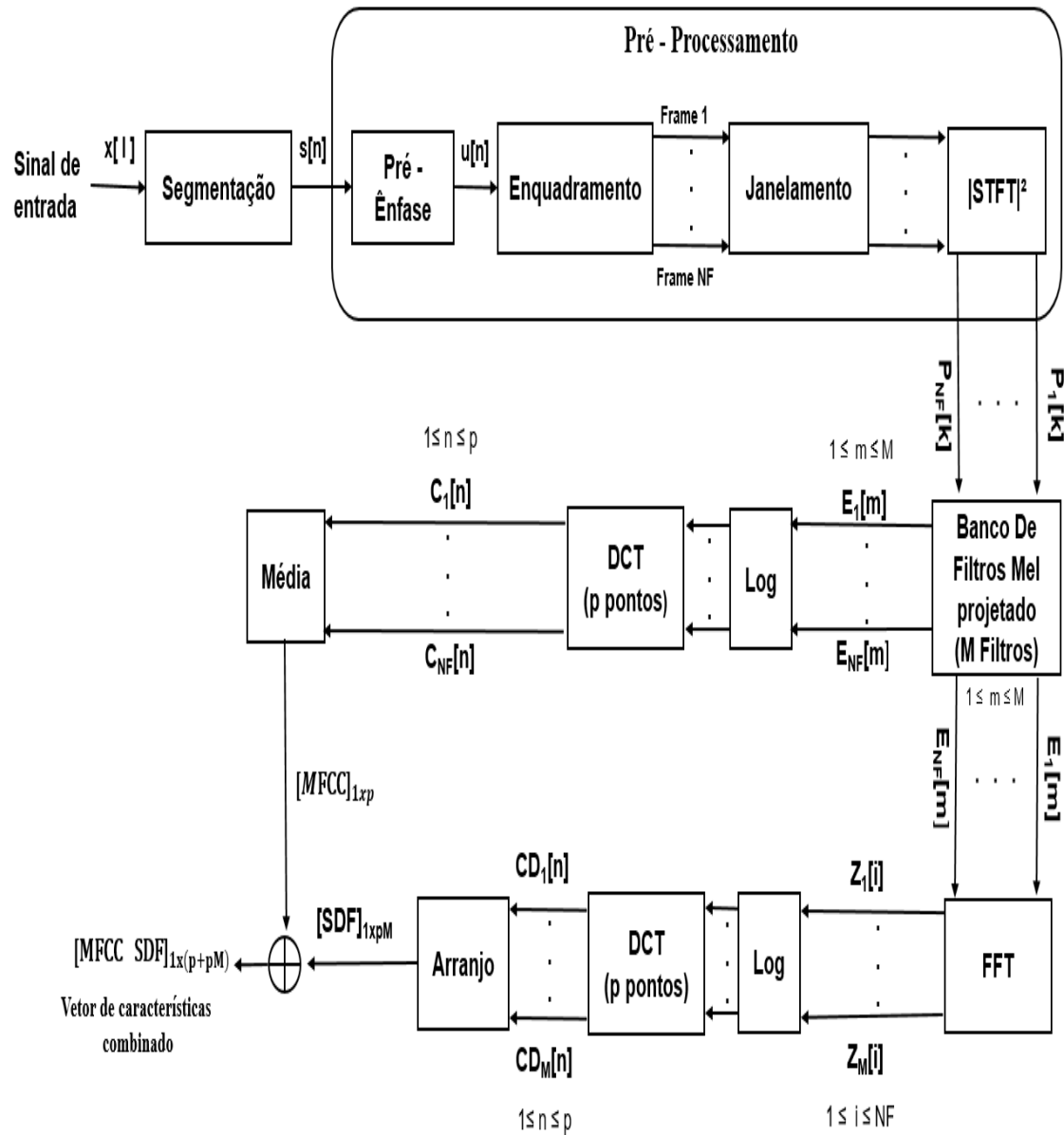


Figura 3.2: Processo de obtenção do Vetor de características combinado. Sinal de entrada é segmentado em trechos menores, pré-processado e extraído os coeficientes espectrais de potência. A matriz de coeficientes é submetida ao banco de filtros Mels projetado, cujas saídas são utilizadas para extração dos MFCCs e do vetor SDF. Por fim, ambos são combinados. Cada segmento origina um vetor de características combinado

convencional delineado na Seção 2.4, obedece uma padronização em que o formato e os limites de cada filtro são definidos pela Equação 2.8, variando a configuração de acordo com o número de filtros selecionados e a fs adotada. Neste trabalho, o banco de filtros Mel é projetado empiricamente levando em consideração a faixa de frequência de interesse, que será detalhado a seguir.

A literatura destaca que, os filtros Mel são geralmente configurados no domínio da frequência, com as frequências centrais distribuídas uniformemente ao longo do eixo Mel [30, 36]. A escolha do número de filtros, das frequências centrais e da largura de banda desempenha um papel crucial na determinação da cobertura do banco de filtros Mel. Esses parâmetros foram escolhidos empiricamente, variando manualmente o número, a frequência central e a largura de banda de cada filtro de modo a abranger a faixa de frequências de maior interesse e que se adequasse as características distintas do conjunto de dados utilizado nas simulações.

As particularidades do contexto de classificação de embarcações marítimas são respaldadas pelos trabalhos de MCKENNA *et al.* [47] e ZAK [1], que destacam a concentração do ruído irradiado por embarcações, predominantemente em frequências abaixo de 1 kHz. Portanto, com o objetivo de obter uma distribuição de bandas de frequências que favoreça alta resolução na faixa de baixas frequências e baixa resolução nas frequências mais altas [31], foi projetado experimentalmente um conjunto de 42 filtros triangulares passa-banda.

Esses filtros estão limitados a uma faixa de frequência de 0 a 7 kHz, sendo que metade deles está centrada em frequências abaixo de 1 kHz para atender ao requisito de alta resolução nessa faixa. A utilização de filtros com frequências acima de 7 kHz não apresentou ganhos significativos na taxa de classificação, pois os alvos de interesse (embarcações e ruído ambiente) concentram suas emissões abaixo dessa faixa. A configuração final do banco de filtros está disposta na Tabela 3.1.

Tabela 3.1: Configuração do banco de filtros Mel projetado. Foram projetados 42 filtros triangulares, com frequência central e largura de banda variáveis

Filtros Triangulares	Frequência Central	Largura de Banda
1 a 10	25, 50, 75, 100, ..., 250 Hz	50 Hz
11 a 17	300, 350, 400, ..., 600 Hz	100 Hz
18 e 19	675, 750 Hz	150 Hz
20 a 26	850, 950, 1050, ..., 1450 Hz	200 Hz
27 a 30	1600, 1750, ..., 2050 Hz	300 Hz
31 a 33	2300, 2500 e 2800 Hz	400 Hz
34 a 37	3100, 3400, ..., 4000 Hz	600 Hz
38 a 42	4500, 5000, ..., 6500 Hz	1000 Hz

A saída do banco de filtros Mel é a matriz E (3.2), também apresentada na Figura 3.2, cujas linhas representam a energia de saída de m filtros para um conjunto de

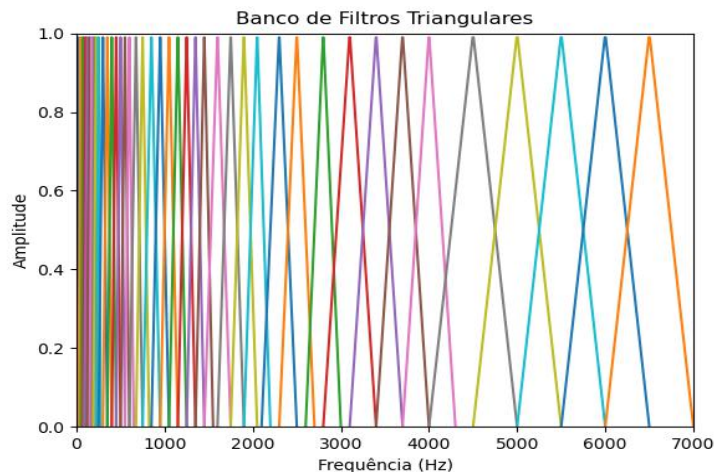


Figura 3.3: Banco de filtros triangulares descritos a Tabela 3.1.

frames. Observe que a mesma matriz E é utilizada como entrada para o cálculo do MFCC e do SDF, etapas que serão detalhadas nas Seções 3.3.1 e 3.3.2.

$$E = \begin{bmatrix} e_1[1] & \cdots & e_1[m] & \cdots & e_1[M] \\ \vdots & \ddots & \vdots & & \\ e_i[1] & \cdots & e_i[m] & \cdots & e_i[M] \\ \vdots & & \vdots & \ddots & \\ e_{NF}[1] & \cdots & e_{NF}[m] & \cdots & e_{NF}[M] \end{bmatrix}; \quad 1 \leq i \leq NF, 1 \leq m \leq M, \quad (3.2)$$

sendo M o número de filtros e NF o número de *frames*.

3.3 Extração do vetor de característica

Conforme apresentado na Figura 3.1 o vetor de característica proposto é formado pela combinação do vetor de característica resultante da extração dos MFCCs, com o vetor de características espectrais dinâmicas. As Seções 3.3.1, 3.3.2 e 3.3.3 descrevem o processo de extração desses vetores.

3.3.1 Extração do vetor de características MFCC

A extração do vetor de características MFCC é realizada seguindo as etapas apresentadas na Seção 2.4 com o detalhamento apresentado na Figura 3.2. A matriz E , apresentada na Equação 3.2, que exibe as saídas do banco de de filtros projetado, é submetida a uma transformação logarítmica seguida da aplicação da DCT. O resultado é a matriz C :

$$C = \begin{bmatrix} c_1[1] & \cdots & c_1[n] & \cdots & c_1[p] \\ \vdots & \ddots & \vdots & & \\ c_i[1] & \cdots & c_i[n] & \cdots & c_i[p] \\ \vdots & & \vdots & \ddots & \\ c_{NF}[1] & \cdots & c_{NF}[n] & \cdots & c_{NF}[p] \end{bmatrix}; \quad 1 \leq i \leq NF, 1 \leq n \leq p, \quad (3.3)$$

onde cada linha refere-se aos MFCCs obtidos para um determinado *frame*, p o números de pontos escolhidos para a DCT e também o número de MFCCs extraído de um *frame*.

Por fim, depois de extrair o vetor de característica MFCC para cada *frame*, combinamos todos os vetores de características de um segmento de áudio. A combinação é realizada calculando a média das colunas da matriz C , ou seja, a média da energia de cada filtro ao longo de todos os *frames*:

$$mean_{MFCC} = [\mu_1 \cdots \mu_i \cdots \mu_p]; \quad 1 \leq i \leq p, \quad (3.4)$$

onde μ_i representa o valor médio da i^{th} coluna da Matriz C .

A Figura 3.4 apresenta uma descrição detalhada do processo de obtenção do vetor de características MFCC a partir das saídas do banco de filtros Mel projetado. A matriz E passa por etapas de aplicação do logaritmo e transformação cepstral, realizada pela aplicação da DCT, resultando na matriz C . Finalmente, o vetor de características MFCC de um segmento é obtido calculando-se a média das colunas da matriz C .

3.3.2 Extração do Vetor de características espectrais dinâmicas (SDF)

O vetor SDF baseia-se na proposta de KARBASI *et al.* [15], no qual os autores abordam a tarefa de classificação de sons ambientais na acústica aérea. Essa abordagem foi adaptada para atender ao contexto específico da classificação de sons submarinos.

Conforme visto na Seção 1.2, caracterizar um sinal de áudio submarino é uma tarefa complexa, muito em função da variação temporal das características espectrais, resultado da combinação de múltiplos espectros de curta duração. Dado que os métodos tradicionais de extração geralmente se concentram nas características espectrais de um sinal em um único *frame* e, visando capturar informações ao longo de *frames* sucessivos, é proposto um algoritmo para extrair informações dinâmicas utilizando a energia de saída do banco de filtros Mel projetado. Esta abordagem aproveita a continuidade temporal do sinal, explora a não uniformidade e captura

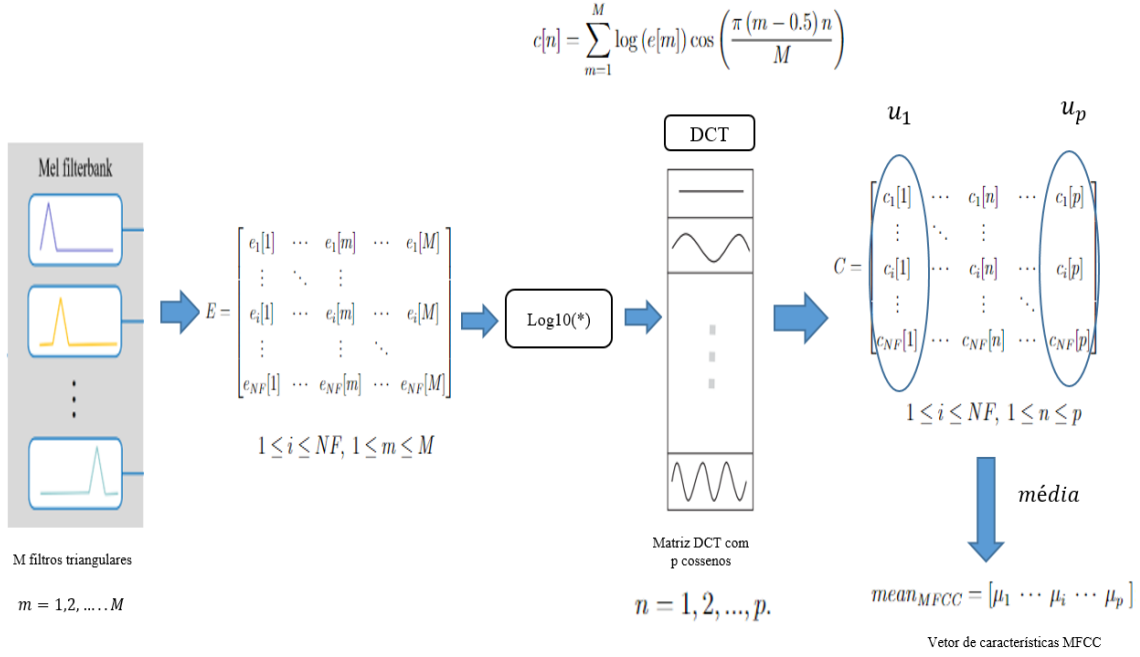


Figura 3.4: Detalhamento da etapa Mel contendo a representação matricial do processo de extração do vetor de característica MFCC proposto. As saídas do banco de filtros Mel projetado, representadas pela matriz E, passa pelas etapas de aplicação do logaritmo e transformação cepstral. O vetor de característica MFCC de um segmento é obtido pela média das colunas da matriz C resultante.

como as características mudam entre *frames* sucessivos. Ao estender a janela de análise, o método fornece uma visão mais abrangente do comportamento das características ao longo do tempo, contribuindo para melhorar a acurácia da classificação.

A extração dessa tendência de variação temporal das características é realizada por meio da aplicação de uma *FFT* nas colunas da matriz E, exibido na Figura 3.5.

$$\begin{array}{c}
 \mathbf{FFT} \\
 E = \begin{bmatrix} e_1[1] & \cdots & e_1[m] & \cdots & e_1[M] \\ \vdots & \ddots & \vdots & & \vdots \\ e_i[1] & \cdots & e_i[m] & \cdots & e_i[M] \\ \vdots & & \vdots & \ddots & \vdots \\ e_{NF}[1] & \cdots & e_{NF}[m] & \cdots & e_{NF}[M] \end{bmatrix}
 \end{array}$$

Figura 3.5: Aplicação da *FFT* nas colunas da Matriz E. Procedimento para extrair a tendência de variação da saída de um filtro em uma banda de frequência.

Cada coluna dessa matriz representa a energia média de saída de um determinado filtro em todos os *frames*. Por exemplo, os elementos da primeira coluna são as

energias de saída do filtro de largura de banda de 50Hz, e cuja frequência central é de 25Hz extraído em cada *frame*, visto na Tabela 3.1. Em outras palavras, aplicar uma *FFT* na coluna da Matriz E, busca-se capturar as informações da variação da energia de saída em uma banda de frequência específica.

Seja $w_m[i]$ os vetores extraídos da coluna da matriz E, $v_m[i] = \{e_i[m], 1 \leq i \leq NF, 1 \leq m \leq M\}$, Os resultados da *FFT* aplicada as colunas de E são agrupados na matriz Z como segue:

$$Z = \begin{bmatrix} \mathcal{F}\{v_1[i]\} \\ \vdots \\ \mathcal{F}\{v_m[i]\} \\ \vdots \\ \mathcal{F}\{v_M[i]\} \end{bmatrix} \triangleq \begin{bmatrix} z_1[1] & \cdots & z_1[i] & \cdots & z_1[NF] \\ \vdots & \ddots & \vdots & & \\ z_m[1] & \cdots & z_m[i] & \cdots & z_m[NF] \\ \vdots & & \vdots & \ddots & \\ z_M[1] & \cdots & z_M[i] & \cdots & z_M[NF] \end{bmatrix}; \quad 1 \leq i \leq NF, 1 \leq m \leq M \quad (3.5)$$

Assim, o resultado da *FFT* aplicada nas colunas são agora dispostas como linhas na Matriz Z.

Para concluir o processo de obtenção do vetor *SDF* para um segmento do sinal, são realizadas duas etapas adicionais com o objetivo de reduzir a redundância dos dados e a dimensionalidade. De maneira análoga ao cálculo dos *MFCCs*, aplicamos o logaritmo à matriz Z para equalizar o espectro. Em seguida, aplicamos a *DCT* para extrair os parâmetros cepstrais das energias de saída dos filtros. O resultado é uma matriz de componentes dinâmicos (*CD*) que após um rearranjo, suas linhas são concatenadas para dar origem ao vetor *SDF*, conforme apresentado na Figura 3.2.

$$CD = \begin{bmatrix} cd_1[1] & \cdots & cd_1[n] & \cdots & cd_1[p] \\ \vdots & \ddots & \vdots & & \\ cd_m[1] & \cdots & cd_m[n] & \cdots & cd_m[p] \\ \vdots & & \vdots & \ddots & \\ cd_M[1] & \cdots & cd_M[n] & \cdots & cd_M[p] \end{bmatrix}; \quad 1 \leq m \leq M, 1 \leq n \leq p \quad (3.6)$$

onde p é o número de pontos da *DCT*.

Note que a matriz Z é de alta dimensionalidade, o que dificulta a convergência dos algoritmos de classificação. Ao utilizar a *DCT* sobre o logaritmo da Matriz Z, conseguimos reduzir a dimensionalidade do vetor *SDF*, limitando o número de colunas ao número de pontos da *DCT*. Isso otimiza o processo de classificação, e, como observado anteriormente, os primeiros coeficientes da *DCT* concentram a maior parte das características discriminatórias, resultando em uma redução na

dimensão final do vetor sem afetar significativamente a taxa de acerto.

3.3.3 Vetor de características Combinado

O vetor de características combinado ([MFCC SDF]) é construído pela concatenação do vetor de características MFCC, tradicionalmente utilizado para capturar informações estáticas do sinal acústico, com o vetor de características SDF, que extrai informações dinâmicas. Tanto a dimensão do MFCC quanto a dimensão do SDF varia conforme o número de filtros adotados e também a escolha do número de pontos da DCT. A dimensão final do vetor de característica combinado é a soma das dimensões dos vetores MFCC e SDF. Essa fusão de informações complementares, características estáticas e características dinâmicas, enriquece a representação do sinal, fornecendo ao classificador uma visão mais completa das características sonoras das embarcações.

A adição de novas dimensões, que contêm informações dinâmicas, aos MFCCs melhora o desempenho do classificador como será demonstrado nos resultados da Seção 5.2. Além disso, a metodologia proposta não limitou-se à combinação de MFCCs e SDFs. A combinação de outros extratores clássicos com o vetor SDF também foi implementada e os resultados comparados.

Por fim, o vetor de característica combinado, que representa um conjunto de características compactas e informativas do sinal acústico, serve como entrada para os algoritmos de classificação. Estes, por sua vez, utilizam essas características para determinar a classe a qual cada embarcação pertence.

Capítulo 4

Conjunto de dados

Este capítulo detalha o conjunto de dados e sua organização. Foram utilizados dois conjuntos de dados experimentais coletados em diferentes locais para validar a eficácia do algoritmo de extração do vetor de características combinado. A Seção 4.1 descreve o banco de dados "*ShipsEar*" composto de sinais de ruído irradiado por embarcações e sinais de ruído ambiente gravados na costa noroeste da Espanha. Na Seção 4.2 é apresentado o conjunto de dados formado por gravações de ruído irradiado de embarcações e ruído ambiente coletados da Enseada dos Anjos localizada no município de Arraial do Cabo - RJ.

4.1 Banco de dados *ShipsEar*

Os resultados apresentados nas Seções 5.2 e 5.4 utilizaram o banco de dados *ShipsEar* para testar a acurácia do algoritmo proposto. Trata-se de um conjunto de gravações de sons submarinos, apresentado na literatura em SANTOS-DOMÍNGUEZ *et al.* [3]. O banco de dados é composto de gravações de 90 embarcações de 11 tipos diferentes e também do ruído ambiente natural em diferentes pontos da costa noroeste da Espanha, coletados durante os anos de 2012 e 2013, totalizando cerca de 3 horas de gravações. As gravações foram realizadas com os gravadores acústicos digitais autônomos Hyd SR-1 da MarSensing Lda, que possuem um hidrofone com sensibilidade de $-193,5$ dB re 1 V/ 1 μ Pa e uma resposta de frequência plana de 1 Hz a 28 kHz.

O *ShipsEar* também fornece informações das condições ambientais, identificação exclusiva para cada registro (ID), nome da embarcação, tipo da embarcação (Draga, barco de pesca, lanchas, transatlântico, embarcação de passageiros, embarcação de piloto, RO-RO, barco a vela, traineira, rebocador), imagens das embarcações, dados de localização (posição GPS), horário e data de aquisição, ganho e profundidade do hidrofone, profundidade do canal, distância aproximada entre a embarcação registrada e a vertical do hidrofone, Velocidade do vento (km/h) medida in-situ, dados

atmosféricos e oceanográficos, duração dos registros e notas com observações sobre cada gravação. A Figura 4.1 exibe um exemplo de interface gráfica, disponíveis em <http://atlanttic.uvigo.es/underwaternoise>.





Id	Name	Type	Pic	Localization	Date	H_G_D	Video	Chanel depth	Wind	AIS link	Distance	Atmospheric and oceanographic data
33	"Duda" (going out)	Motorboat		 <p>Red lighthouse at the port</p>	2013-07-19 12:38:00	(1,2)(2,32)(2,5,3,5)	SI	5.8	13	Not available	<50	Mean air temperature (°C) 23.4 Maximum air temperature (°C) 31 Minimum air temperature (°C) 17.4 Mean relative humidity (%) 68 Maximum relative humidity (%) 100 Minimum relative humidity (%) 44 Dew temperature (°C) 16.5 Cold hours (7 °C base) (h) 0 Wind speed (km/h) 11.81 Streak (km/h) 36.94 Streak direction (Degrees) 2 Main wind direction (Degrees) 45 Rain (L/m2) 0
80	Adriostuy	Dredger		 <p>Punta Langosteira</p>	2012-10-04 14:00:00	(3),(16),(5)	No	20	Not available	AIS data	350 m	Considerable swell. Moderate / low wave

Figura 4.1: Interface do site <http://atlanttic.uvigo.es/underwaternoise>, banco de dados de teste. O site contém informações detalhadas de todas as gravações que constituem o banco de dados.

A quantidade de gravações para cada tipo de embarcação não é uniforme. Para otimizar os testes e prover um balanceamento entre as classes, foram selecionados 53 arquivos de áudio do banco de dados, organizando-os em quatro categorias/classes distintas, limitando-os a uma duração de 24 segundos a partir de recortes, de trechos com melhor SNR, da gravação original.

Objetivando aumentar o espaço amostral e a variabilidade de vetores de características para uma mesma embarcação ou ruído ambiente, uma vez que o ruído emitido não é uniforme ao longo de toda gravação, todos os arquivos de 24 segundos foram divididos em segmentos com duração de 4 segundos, permitindo a extração de um vetor de características combinado para cada segmento. A seleção de segmentos com duração de 4 segundos foi feita após condução de simulações, nas quais, diferentes durações foram experimentadas, e a eficácia da classificação foi avaliada.

A quantidade de gravações bem como porte e tipo de embarcação foi levado em consideração para separação dos dados em classes distintas. O agrupamento correto dos arquivos que apresentam vetor de características semelhantes, influencia na taxa de acerto do algoritmo. Optou-se por subdividir os 53 arquivos em 4 classes distintas, agrupando em uma mesma classe, quando necessário, embarcações que apresentaram

semelhanças no vetor de características extraído. Dessa forma, foi promovido um balanceamento na quantidade de amostra de cada classe. A configuração final com as subdivisões das classes é apresentada na Tabela 4.1, semelhante ao apresentado em SANTOS-DOMÍNGUEZ *et al.* [3] e para cada classe foi atribuído, manualmente, um rótulo que a identificava.

A duração corresponde ao somatório de todos os arquivos de áudio pertencente a classe e o número de segmentos representa a quantidade de vetores de características para cada classe, uma vez que é extraído um vetor a cada segmento de 4 segundos. A Tabela 4.2, exibe a identificação (ID) dos arquivos .wav (Waveform Audio File Format) no conjunto de dados ShipsEar.

Tabela 4.1: Subdivisões das classes para formar o conjunto de dados dos experimentos.

Classe	Tipo de Embarcação/Ruído	Nº de arquivos de áudio	Duração(s)*
A	Ruído Ambiente	10	240s (60)
B	Lanchas / Embarcações de Piloto	15	360s (90)
C	Transatlântico/Mercantes	11	264s (66)
D	Embarcação de passageiros	17	408s (102)
TOTAL		53	1272s (318)

* Os valores entre parênteses representam o número de segmentos para cada classe (vetores de características).

Tabela 4.2: Identificação (ID) no banco de dados ShipsEar, dos arquivos .wav utilizados nos experimentos.

Classe	Tipo de Embarcação/Ruído	ID
A	Ruído Ambiente	81, 82, 83, 84, 85, 86, 87, 88, 89, 90
B	Lanchas / Embarcações de Piloto	21, 26, 27, 29, 30, 33, 39, 45, 50, 51, 52, 70, 72, 77, 79
C	Transatlântico/Mercantes	16, 18, 19, 20, 22, 24, 25, 58, 69, 71, 78
D	Embarcação de passageiros	7, 8, 9, 10, 11, 12, 14, 17, 32, 34, 35, 36, 38, 40, 41, 42, 43

A Figura 4.2 exibe os espectrogramas de varias amostras do conjunto de dados ShipsEar, mostrando claramente uma variabilidade entre as classes abordadas. Adicionalmente, pode ser observado que as embarcações diferentes possuem modos vibracionais semelhantes, como ruído de cavitação da hélice, e ruído de modulação de ritmo causado pelo movimento do pistão diesel. Essa semelhança pode levar a pontos em comum entre diferentes classes de embarcações [67]. Observa-se nos

espectrogramas por exemplo, tons discretos em frequências relativamente próximas entre o Transatlântico 24 e a Lancha 72, o que não é observado no espectro referente ao ruído ambiente.

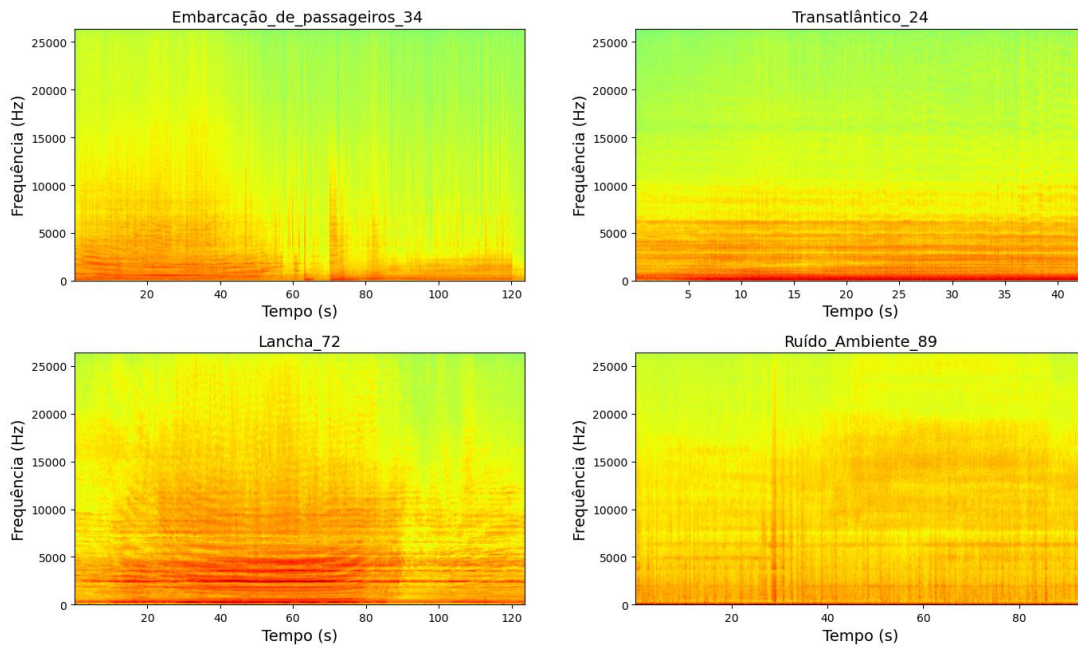


Figura 4.2: Espectrogramas de várias amostras no conjunto de dados Shipsear. Gravações da Embarcação_de_passageiros_34 chegando no porto, Transatlântico_24 desacelerando antes da manobra de atracação, Lancha_72 passando próximo aos hidrofones com possível interferência no final da gravação e amostras de Ruído_ambiente_89 natural.

4.2 Experimento na Enseada dos Anjos - Arraial do Cabo-RJ

Nesta seção, apresenta-se um conjunto de dados de gravações de ruído ambiente e de embarcações que transitaram pela Enseada do Anjos, no município de Arraial do Cabo - RJ, nos dias 11, 13, 14 e 15 de outubro de 2023.

Para gerar esse conjunto de dados foi conduzido um experimento que obedeceu a configuração apresentada na Figura 4.3, onde, um hidrofone posicionado a 8 metros de profundidade foi utilizado para capturar emissões sonoras emitidas por embarcações que trafegavam a uma distância horizontal média de 400 metros do local do hidrofone, a uma profundidade de 10 metros.

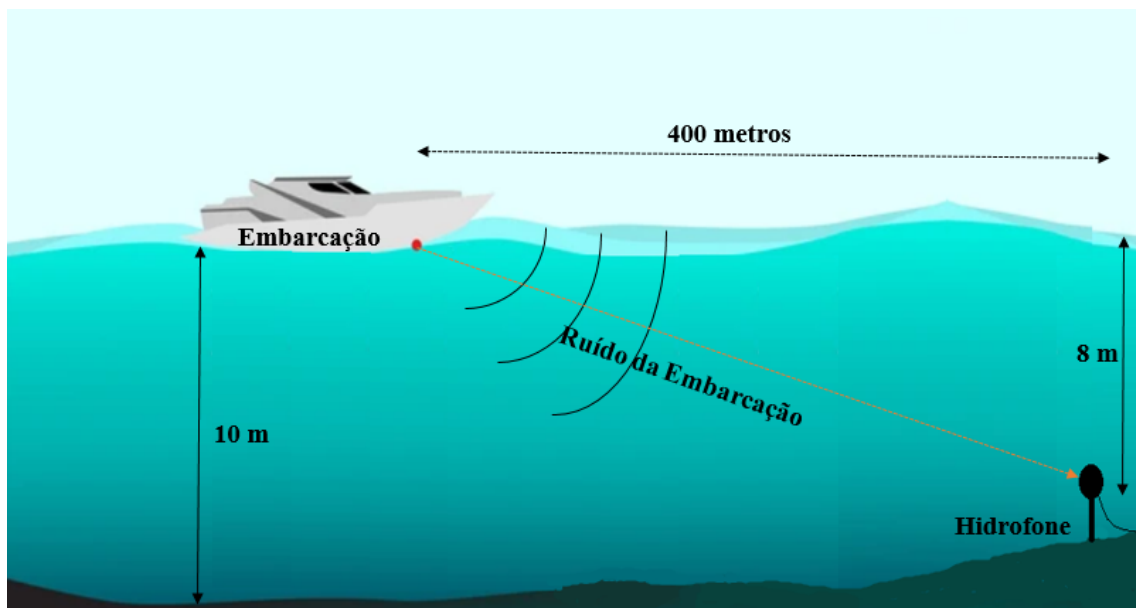


Figura 4.3: Configuração vertical do experimento na Enseada dos Anjos. Embarcação emitindo ruído a uma distância média de 400 metros do sensor (hidrofone). O hidrofone está localizado a uma profundidade de 8 metros e as embarcações trafegam em uma área com profundidade de 10 metros.

A Figura 4.4 ilustra a área do experimento, onde pode ser observado a posição do hidrofone e uma área de trânsito de embarcações representada por um retângulo amarelo um retângulo amarelo.

O hidrofone utilizado no experimento, foi construído no IEAPM com materiais de baixo custo e calibrado pelo Instituto de Pesquisas da Marinha (IPqM) para uma faixa de frequência entre 800Hz e 25kHz .

Durante a realização do experimento, além das gravações, imagens também foram capturadas simultaneamente e as gravações foram realizadas em modo contínuo ao longo dos dias mencionados, com uma taxa de amostragem (f_s) de 50kHz . Contudo, optou-se por gerar arquivos .wav a cada 2 minutos de gravação, para evitar que uma



Figura 4.4: Área do experimento na Enseada dos Anjos, Arraial do Cabo, contendo a posição do hidrofone e da área de transito com maior fluxo de embarcações. A distância aproximada entre o hidrofone e a área de transito é de 400m.



Figura 4.5: Hidrofone desenvolvido no IEAPM com material de baixo custo.

falha na geração de um arquivo compromete-se todo o experimento.

Todos os arquivos de áudio e respectivas imagens foram analisadas previamente, com o objetivo de estabelecer um banco de dados balanceado para o algoritmo. No entanto, devido ao fluxo intenso de embarcações, foi selecionado manualmente arquivos de gravações em que o sinal da embarcação desejada não sofresse interferência do sinal de outras embarcações. A análise manual possibilitou a separação em três grandes grupos, composto por ruído ambiente, embarcação de médio porte (transporte de passageiros) e embarcações de pequeno porte (embarcações de pesca), exposto na Tabela 4.3.

As gravações de ruído ambiente selecionadas ocorreram em dias e horários distintos totalizando 580 segundos de gravações. Os arquivos selecionados de gravações

Tabela 4.3: Subdivisões das classes - Experimento realizado na Enseada dos Anjos

Classe	Tipo de Embarcação/Ruído	Nº de arquivos de áudio	Duração(s)*
A	Ruído Ambiente	5	580s (145)
B	Embarcação de médio porte	9	616s (154)
C	Embarcação de pequeno porte	5	364s (91)
TOTAL		19	1560s (390)

* Os valores entre parênteses representam o número de segmentos para cada classe (vetores de características).

de embarcações de médio porte, totalizaram 616 segundos, referente a 9 embarcações distintas, e, os arquivos das embarcações de pequeno porte, resultaram em 370 segundos oriundos de 5 embarcações distintas. Assim como realizado com arquivos do banco de dados “*ShipsEar*”, todos os arquivos de áudio selecionados foram divididos em segmentos de 4 segundos originando um vetor de característica por segmento, como mostrado na quarta coluna da Tabela 4.3.

Capítulo 5

Resultados e Discussão

Neste capítulo, são analisados e discutidos os resultados obtidos na classificação, utilizando diferentes extratores de características e dois conjuntos de dados distintos. Para isso, é utilizada a acurácia como métricas de avaliação, que mede a proporção de exemplos corretamente classificados em relação ao total de exemplos avaliados [65], definida pela Equação 2.12.

Nas Seções 5.2 e 5.3, o desempenho foi medido sujeitando o vetor de características combinado como fonte de dados de entrada para quatro métodos de classificação supervisionado; *Support Vector Machine* (SVM), *k-Nearest Neighbor* (k-NN), *Random Forests* (RF) e *Decision Tree* (DT) [11, 13, 68]. Além disso, foi avaliada sua performance frente a outros vetores de características disponíveis na literatura: *MFCC*, *Power Spectral Coefficients* (PSC), *Linear Prediction Coefficients* (LPC), *Linear Prediction Cepstral Coefficients* (LPCC) e *Zero-crossing rate* (ZCR) [13, 18].

A Seção 5.1 apresenta algumas considerações sobre os conjuntos de dados utilizados nos experimentos. Na Seção 5.2 e 5.4 os experimentos são conduzidos utilizando o banco de dados *ShipsEar* enquanto na Seção 5.3 o conjunto de dados é proveniente de gravações de passagem de embarcações coletadas na Enseada dos Anjos, descrito na Seção 4.2

Além disso, Seção 5.4 avalia a dispersão no espaço de características de vetores de características extraídos de um mesmo arquivo de gravação porém com adição de ruídos em diferentes *SNRs*. A métrica de avaliação de desempenho foi baseado no calculo entre a distância da característica extraída do arquivo com o som original e a característica extraída do arquivo com o som original mais a adição de ruído.

5.1 Considerações sobre os conjuntos de dados

A utilização de dois conjuntos de dados distintos, coletados em locais com diferentes características ambientais, possibilita a análise do comportamento do algoritmo proposto em diferentes condições, fornecendo uma visão abrangente da ro-

bustez e generalização do algoritmo em diferentes contextos. O conjunto de dados “*ShipsEar*” oferece uma ampla variedade de condições submarinas e tipos de embarcações, abrangendo diversos ambientes marinhos. Por outro lado, o experimento realizado na Enseada dos Anjos, em Arraial do Cabo - RJ, oferece *insights* específicos sobre as características do ruído ambiente e das embarcações presentes nessa região costeira.

A diversidade desses conjuntos de dados complementares proporciona um cenário mais completo para avaliar a eficácia do algoritmo de extração do vetor de características combinado. Ao incorporar informações de locais com características ambientais distintas, nosso estudo busca garantir uma validação abrangente e robusta do método proposto. Essa abordagem visa aprimorar a acurácia e confiabilidade do algoritmo, preparando-o para uma variedade de cenários de aplicação real.

5.2 Resultados da classificação de embarcações marítimas utilizando o conjunto de dados *ShipsEar*

Conforme visto na tabela 4.1, cada arquivo foi dividido em 6 segmentos de 4s, totalizando um conjunto de dados de 318 vetores de características que serão utilizados para treinamento e teste. Foi mantida a taxa de amostragem original de 52.734 *Hz* dos sinais de áudio das embarcações e do ruído ambiente. Na etapa de pré-processamento, adotamos um coeficiente de pré-ênfase de 0,97, uma janela retangular de duração de 40 *ms* com taxa de sobreposição de 50% e aplicada a janela de *hamming* sobre cada *frame*.

5.2.1 Resultados da classificação pelo método de validação cruzada *Leave-One-Out*

Inicialmente foi avaliado o percentual de previsões corretas quando submetido o vetor de características combinado, desenvolvido neste trabalho, aos classificadores, SVM, KNN, RF e DT. Nessa primeira análise a divisão do conjunto de dados foi estabelecida para atender o método de validação cruzada *leave-one-out* [65], no qual removemos todos os segmentos de um mesmo arquivo de áudio do conjunto de treinamento e classificamos cada um desses segmentos individualmente. Isto foi realizado com intuito de tornar os resultados menos tendenciosos, uma vez que, retirando apenas um vetor de característica de um segmento de uma determinada gravação do conjunto de treinamento os outros 5 vetores de características seriam

utilizados para treinar o modelo, e por serem semelhantes, dado que são objetos de uma mesma gravação, induziria a classificação correta pelo algoritmo [66, 69]. Em outras palavras, a cada teste retiramos os 6 vetores correspondentes a um mesmo arquivo, utilizando os outros 312 vetores como dados de treinamento e classificamos cada um dos 6 individualmente, estabelecendo uma proporção de 312:6 entre dados de treinamento e teste a cada iteração. Como foram testados 53 arquivos, um por vez, ocorreram 53 iterações (vide Tabela 4.1).

A título de comparação, foram calculados 42 *MFCCs* de duas formas distintas. A primeira obedece a distribuição convencional do banco de filtros Mel, conforme visto em [5, 30, 36], e designado simplesmente como *MFCC*. A segunda abordagem extrai os *MFCCs* usando a configuração de filtros Mel detalhada na Seção 3.3 e é designada como *MFCC*[†], sendo também parte do vetor de características combinado. O vetor *SDF* foi calculado com uma transformada discreta do cosseno (*DCT*) de 4 pontos ($p = 4$), limitando a matriz CD, a 4 colunas. Como o número de filtros (M) é igual a 42, a dimensão final do vetor *SDF* foi de $4 \times 42 = 168$. Assim, o vetor de características combinado *MFCC*[†] + *SDF* é composto por $42 + 168 = 210$ componentes.

Os parâmetros dos classificadores foram ajustados para maximizar o desempenho no conjunto de dados utilizado. Para o *SVM*, foi selecionado experimentalmente o *kernel* função de base radial (RBF) dentre os quatro testados (linear, polinomial, função de base radial ou gaussiano, sigmoid). Do classificador *K-NN* foram selecionados os 5 vizinhos mais próximos e 100 árvores foram estimadas referente ao algoritmo *RF* e para a *DT* as características não foram alocadas aleatoriamente em cada nó, ao invés disso, foi adotado o critério de melhor divisão. A Tabela 5.1 trás os resultados das taxas de classificação para cada vetor de característica referente aos quatro tipos de classificadores.

A tabela mostra que entre os 4 classificadores o *SVM* foi o que obteve o melhores resultados para a maior parte dos extratores de características, com exceção dos *PSCs*, das componentes do *ZCR* e dos *LPCCs*. Esse fato pode ser atribuído à natureza menos dispersa no espaço n-dimensional do conjunto de características extraídas do domínio cepstral. Essa peculiaridade facilita o traçado de hiperplanos para estabelecer as divisões de classe, contribuindo para o desempenho superior do *SVM* em relação aos outros classificadores testados. Além disso, comparando os resultados das taxas de classificação obtidos pelos 4 algoritmos de classificação alimentado com 13 vetores de características distintos, observa-se que o vetor de característica *MFCC*[†] + *SDF* atingiu melhores resultados em sua grande maioria, tendo uma acurácia de 85.5% para o classificador *SVM*, evidenciando também, a questão da natureza menos dispersiva dos vetores de características extraídos pelo algoritmo proposto. Utilizando como métrica de desempenho a precisão, o vetor

Tabela 5.1: Comparação da porcentagem de classificações corretas entre o vetor de características proposto e outros 12 tipos de vetores de características para 4 tipos de classificadores, utilizando dados do conjunto de dados "*ShipsEar*".

Vetor de Características	Taxa de Classificação			
	SVM	K-NN	RF	DT
<i>SDF</i> (168)	77.5%	74.5%	61.3%	64.5%
<i>MFCC</i> [†] (42)	78%	64%	60%	58%
<i>MFCC</i> (42)	71%	61%	62%	58.4%
<i>LPC</i> (12)	62%	53%	48%	49%
<i>LPCC</i> (12)	60%	65%	57.4%	57.2%
<i>PSC</i> (960)	45%	66%	71%	71%
<i>ZCR</i> (376)	37%	45%	40%	37%
<i>MFCC</i> [†] + <i>SDF</i> (210)	85.5%	77%	72%	70%
<i>MFCC</i> + <i>SDF</i> (210)	84%	75%	76%	65%
<i>LPC</i> + <i>SDF</i> (180)	80.5%	72.8%	70.7%	70%
<i>LPCC</i> + <i>SDF</i> (180)	81.8%	75%	69.6%	70%
<i>PSC</i> + <i>SDF</i> (1128)	75%	74%	71%	70%
<i>ZCR</i> + <i>SDF</i> (544)	78.5%	74%	66%	65%

*Os valores entre parênteses significam a dimensão do vetor de características.

de característica *MFCC*[†] + *SDF* associado ao classificador *SVM* obteve o melhor resultado alcançando um percentual de 87% de acerto.

A combinação dos extratores de características, *MFCC*[†], *MFCC*, *LPC*, *LPCC* e *ZCR* com o vetor *SDF*, elevou o percentual de classificações corretas em todos os casos. Essa melhora ocorre devido a adição de características estaticamente descorrelacionadas as características desses extratores, oriundas do vetor *SDF*.

O vetor de características espectrais dinâmicas (*SDF*) proposto por [15], foi inicialmente idealizado para prover o reconhecimento de diferentes sons ambientais da acústica aérea sujeito a interferência de varias fontes. Aqui, o mesmo, não apresentou resultados tão satisfatórios no que diz respeito a separação entre as classes de embarcações, ainda assim seu desempenho superou o de extratores tradicionais como *MFCC*, *LPC*, *LPCC* e *ZCR*. No entanto na identificação do ruído ambiente o vetor *SDF* apresentou uma acurácia de 98,33% conforme pode ser visto na matriz da Figura 5.1 b), que mostra a matriz de confusão resultante da classificação pelo *SVM* do vetor de características *SDF*.

Isso mostra que, o algoritmo desenvolvido teve boa aderência à classificação do ruído ambiental da acústica submarina, semelhante aos bons resultados alcançados pelo algoritmo de KARBASI *et al.* [15] na classificação do ruído ambiental na acústica aérea. A maneira encontrada para contornar o problema da separação das classes de embarcação e elevar o percentual de classificações corretas, adequando-se ao contexto do conjunto de dados existente, foi agrupar os *MFCC*[†] aos *SDF*s,

criando o vetor de características combinado. Essa estratégia mostrou-se assertiva, uma vez que, os resultados experimentais apontaram para uma maior taxa de acerto dos classificadores utilizando o vetor de características combinado entre todas as características investigadas na Tabela 5.1, igual a 85.5%.

Foi observado também, um aumento na taxa de acertos em 14.5% quando comparada a classificação, pelo algoritmo *SVM*, do vetor de características combinado em comparação com o vetor formado apenas pelos *MFCCs* clássicos.

É fundamental destacar a importância de um projeto individualizado do banco de filtros mel de modo a atender as peculiaridades dos diferentes tipos de embarcações que se deseja classificar, haja visto que, o *MFCC*[†] teve um aumento de 7% em relação aos *MFCCs* convencionais classificados por *SVM*.

A acurácia do reconhecimento para cada classe, utilizando como vetores de características o *MFCC*[†] + *SDF* e o *SDF*, é apresentada na Figura 5.1, na forma da matriz de confusão¹, tendo como classificador o *SVM (Kernel RBF)*. Optou-se por utilizar o *SVM* para ilustrar esses resultados, pois, entre os classificadores avaliados, foi o que obteve o maior número de classificações corretas. A diagonal principal da matriz de confusão contém o número de previsões corretas para cada classe, em outras palavras, representa a contagem de segmentos em que a classe verdadeira é igual a classe prevista. Exemplificando, tomando como referência a primeira matriz de confusão da Figura 5.1, e analisando a 2 linha de resultados dessa matriz (classe B), tem-se um total de 90 segmentos previamente rotulados como classe B, no entanto o modelo previu 74 segmentos corretamente e classificou de forma errada 2 segmentos como classe C e 14 segmentos como classe D.

Observa-se na Figura 5.1 (a) que, para a classe A (ruído ambiente), composta por 10 arquivos contendo um total de 60 segmentos, o algoritmo obteve uma taxa de acerto próxima de 100%. Isso sugere que o método é adequado para atividades de monitoramento, onde o objetivo é detectar a presença ou ausência de embarcações sem a necessidade de distingui-las individualmente. Por outro lado, a classe C (Transatlânticos/Navios Mercantes) apresentou os piores resultados, em comparação com as outras classes de embarcações (Classe B e Classe D), tendo classificado corretamente 48 segmentos de um total de 66 segmentos, correspondendo a 72,72%. Esse resultado pode ser atribuído a menor quantidade de arquivos de áudio disponível para treinamento em comparação com as outras classes de embarcações. Isso pode ter afetado a capacidade do algoritmo de aprender e generalizar adequadamente os padrões dessa classe específica e diferencia-la das demais classes de embarcações,

¹A matriz de confusão é uma tabela que compara as previsões de um modelo de classificação com as verdadeiras classes dos dados de teste. As linhas representam as classes reais dos dados, enquanto as colunas representam as classes previstas pelo modelo. Cada célula na matriz mostra quantos exemplos pertencentes à classe correspondente foram corretamente ou incorretamente classificados pelo modelo.

a)

		CLASSE PREVISTA			
		CLASSE A	CLASSE B	CLASSE C	CLASSE D
CLASSE REAL	CLASSE A	59 (98.33%)	0 (0%)	0 (0%)	1 (1,66%)
	CLASSE B	0 (0%)	74 (82.22%)	2 (2,22%)	14(15,55%)
	CLASSE C	0(0%)	9 (13.63%)	48 (72.72%)	9 (13.63%)
	CLASSE D	0 (0%)	8 (7.84%)	5 (4,90%)	89 (87.25%)

b)

		CLASSE PREVISTA			
		CLASSE A	CLASSE B	CLASSE C	CLASSE D
CLASSE REAL	CLASSE A	59 (98.33%)	0 (0%)	0 (0%)	1 (1,66%)
	CLASSE B	0 (0%)	61 (67.77%)	10 (11,11%)	19 (21,11%)
	CLASSE C	0(0%)	9 (13.63%)	45 (68.18%)	12 (18.18%)
	CLASSE D	0 (0%)	20 (19.6%)	5 (4.90%)	77 (75.49%)

Figura 5.1: Matriz de confusão gerada a partir do conjunto de dados da Tabela 4.2, utilizando o algoritmo de classificação *SVM*, tendo como entrada dois tipos de extratores de características. Cada célula contém o número e a porcentagem de segmentos de áudio de classificações previstas para as classes representadas nas colunas. A diagonal principal da matriz de confusão contém o número de previsões corretas para cada classe. a) Vetor de características combinado ($MFCC^\dagger + SDF$) e b) Vetor de Características Espectrais Dinâmicas(SDF).

haja visto que, 18 segmentos foram incorretamente classificados como classe B e classe D.

5.2.2 Resultados da classificação de todo um arquivo pelo método de validação cruzada *Leave-One-Out*

Também foi avaliada a classificação final de todo o arquivo de áudio(24s) extraíndo um único vetor, por arquivo de áudio, através do cálculo da média dos 6 vetores de características combinado extraído de cada um dos 6 segmentos presentes em um arquivo. Embora essa abordagem funcione bem para sinais relativamente homogêneos, ela não é apropriada quando o sinal contém regiões muito distintas [69]. O método de validação também foi o LOOCV, retirando um arquivo inteiro do conjunto de dados e extraíndo um único vetor de características para ser testado. Sendo o conjunto de dados formado por 53 arquivos, a proporção do conjunto de dados entre treinamento e teste foi de 52:1 a cada iteração. Os resultados estão ilustrados na matriz de confusão da Figura 5.2. Os 10 arquivos de áudio da classe A, ruído

ambiente, foram classificados corretamente obtendo 100% de acertos, assim como os 15 arquivos da classe B. A classe C, transatlânticos/ Mercantes, assim como na classificação por segmentos teve o pior desempenho com 8 arquivos identificados corretamente.

		CLASSE PREVISTA			
		CLASSE A	CLASSE B	CLASSE C	CLASSE D
CLASSE REAL	CLASSE A	10 (100%)	0 (0%)	0 (0%)	0 (0%)
	CLASSE B	0 (0%)	15 (100%)	0 (0%)	0 (0%)
	CLASSE C	0(0%)	1 (9,09%)	8 (72.72%)	2(18.18%)
	CLASSE D	0 (0%)	1 (8.82%)	1 (8.82%)	15(88.2%)

Figura 5.2: Matriz de confusão considerando todo o arquivo(6 segmentos). Calculado o vetor médio $MFCC^{\dagger}+SDF$ dos 6 segmentos e submetido a classificação do algoritmo SVM . Os 10 arquivos de gravação do ruído ambiente foram corretamente identificados.

5.2.3 Resultados da classificação pelo método de validação *holdout* simples

Embora o método de validação cruzada LOOCV seja considerada uma técnica mais robusta, pois testa todos os exemplos de um conjunto de dados um por vez, e se adequa bem a um conjunto de dados pequenos, ela é computacionalmente mais custosa. Além disso, aplicar essa técnica pode promover um desbalanceamento na proporção entre o conjunto de treinamento e o conjunto de teste aumentando o risco de *overfitting*, onde o modelo se ajusta excessivamente aos dados de treinamento mas perde a capacidade de identificar novos dados [11].

Na análise de desempenho realizada na Seção 5.2.1, em que foi classificado os segmentos de cada arquivo individualmente, tinha-se uma proporção entre dados de treinamento e teste de 312:6 a cada iteração. Com objetivo de mitigar um possível *overfitting*, foi realizada uma nova análise dessa vez adotando o método de validação *holdout* simples [66]. Esse método consiste em dividir o conjunto de dados disponível em parte para treinamento e outra parte para teste. A divisão mais comum é de 75% dos dados para treinar o modelo e 25% para testar a acurácia do modelo. Também pode ser utilizadas proporções como 80/20 ou 60/40. Optou-se por uma proporção de 75/25; assim, retirou-se aleatoriamente do conjunto de dados, 12 arquivos para teste, totalizando $6 \times 12 = 72$ segmentos, e, conseqüentemente 72 vetores de características. Os outros 41 arquivos (246 segmentos) formaram o conjunto de treinamento, resultando em uma proporção final de 246:72. Além disso

teve-se o cuidado de não promover um grande desbalanceamento entre as classes na formação do conjunto de teste, retirando um maior número de arquivos da classe com a maior quantidade de arquivos. Retirou-se 4 arquivos da classe D, 3 da classe B, 3 da classe A e 2 da classe C.

A Tabela 5.2 trás os resultados das classificações adotando o método de validação *holdoult* simples. Os parâmetros dos classificadores e tipos de extratores são os mesmos que foram utilizados na Seção 5.2.1.

Tabela 5.2: Comparação da porcentagem de classificações corretas entre o vetor de características proposto e outros 12 tipos de vetores de características para 4 tipos de classificadores, utilizando dados do conjunto de dados "*ShipsEar*".

Vetor de Características	Taxa de Classificação			
	SVM	K-NN	RF	DT
<i>SDF</i> (168)	84%	85%	74.6%	57.3%
<i>MFCC</i> [†] (42)	86.5%	67.7%	54%	52%
<i>MFCC</i> (42)	81.2%	59%	50%	65%
<i>LPC</i> (12)	60%	45%	40%	54.5%
<i>LPCC</i> (12)	69%	72.5%	57%	57%
<i>PSC</i> (960)	47%	66%	75%	71%
<i>ZCR</i> (376)	48%	52%	43%	47%
<i>MFCC</i> [†] + <i>SDF</i> (210)	94%	86%	80.5%	82.5%
<i>MFCC</i> + <i>SDF</i> (210)	90%	85%	90%	88%
<i>LPC</i> + <i>SDF</i> (180)	91%	88%	80%	84%
<i>LPCC</i> + <i>SDF</i> (180)	91.5%	86%	85%	85%
<i>PSC</i> + <i>SDF</i> (1128)	50%	48%	45%	44%
<i>ZCR</i> + <i>SDF</i> (544)	78%	76%	45%	48%

*Os valores entre parênteses significam a dimensão do vetor de características.

Observa-se pelos resultados da Tabela 5.2, que mesmo adotando um método de validação diferente, o vetor *MFCC*[†] + *SDF* continuou tendo o melhor resultado quando submetido ao classificador *SVM* . A combinação da maioria dos extratores com *SDF* elevou consideravelmente o percentual de acertos quando comparado com os resultados dos extratores isoladamente, exceto quando combinado com o vetor *PSC*.

A tabela 5.3 apresenta o percentual de acertos por classe quando submetidos os 13 vetores de características ao classificador *SVM* . As linhas dessa tabela são simplesmente as diagonais principais das matrizes de confusão de cada extrator classificado pelo SVM.

Os resultados obtidos para classe A, denotam a vantagem de associar os extratores de características avaliados, ao extrator *SDF*, onde essa combinação elevou a taxa de acertos a 100%, demonstrando a capacidade do extrator *SDF* de promover a separação do ruído ambiental do ruído irradiado por embarcações.

Tabela 5.3: Comparação da porcentagem de classificações corretas, para cada classe, entre o vetor de características proposto e outros 12 tipos de vetores de características, classificados pelo SVM, utilizando dados do conjunto de dados "ShipsEar".

Vetor de Características	Taxa de Classificação			
	Classe A	Classe B	Classe C	Classe D
<i>SDF</i> (168)	100%	88.9%	91.7%	55%
<i>MFCC</i> [†] (42)	85%	100%	75%	87%
<i>MFCC</i> (42)	88.9%	94.4%	50%	91.7%
<i>LPC</i> (12)	94.4%	44.4%	66.7%	29.2%
<i>LPCC</i> (12)	33.3%	100%	100%	41.7%
<i>PSC</i> (960)	0%	38.9%	33.3%	91.7%
<i>ZCR</i> (376)	33.4%	88.9%	0%	70.8%
<i>MFCC</i> [†] + <i>SDF</i> (210)	100%	94.4%	100%	79.2%
<i>MFCC</i> + <i>SDF</i> (210)	100%	94.4%	83.3%	83.3%
<i>LPC</i> + <i>SDF</i> (180)	100%	94.4%	91.7%	79.2%
<i>LPCC</i> + <i>SDF</i> (180)	100%	88.8%	100%	79.2%
<i>PSC</i> + <i>SDF</i> (1128)	100%	27%	66.66%	0%
<i>ZCR</i> + <i>SDF</i> (544)	100%	44.4%	100%	62.5%

*Os valores entre parênteses significam a dimensão do vetor de características.

5.2.4 Análise do tamanho dos segmentos e distribuição dos vetores de características no espaço 2D

A escolha de segmentos de 4 segundos foi ajustado empiricamente variando-se o tamanho do segmento desde 50 milissegundos até segmentos de 6 segundos. Conforme exibido na Figura 5.3, observou-se um aumento da acurácia a medida que o tamanho do segmento é aumentado, atingindo a melhor performance em segmentos de 4 segundos, utilizando o algoritmo de classificação SVM, tendo como entrada o vetor de características combinado. Segmentos muito pequenos da mesma ordem dos *frames* obtiveram os piores resultados provavelmente por não conter amostras suficientes para extraírem características advindas da variabilidade temporal do sinal.

A Figura 5.4 apresenta a distribuição de vetores de características no plano bidimensional quando reduzimos a dimensão desse vetor para 2 componentes principais, utilizando três métodos de extração de características. Observando o gráfico da Figura 5.4 a), é possível perceber como os vetores de características extraído do ruído ambiente, na cor verde, estão alocados em um grupo disjunto dos vetores extraído de embarcações, no espaço bidimensional, o que não ocorre nos resultados b) e c). Assim, é possível observar como os componentes principais do vetor de características *MFCC*[†] + *SDF* estão menos dispersas, ou seja, as classes estão mais agrupadas, do que as componentes dos outros vetores de características analisados. Esse resultado pode ser explicado pela natureza menos dispersiva das características cepstrais [7],

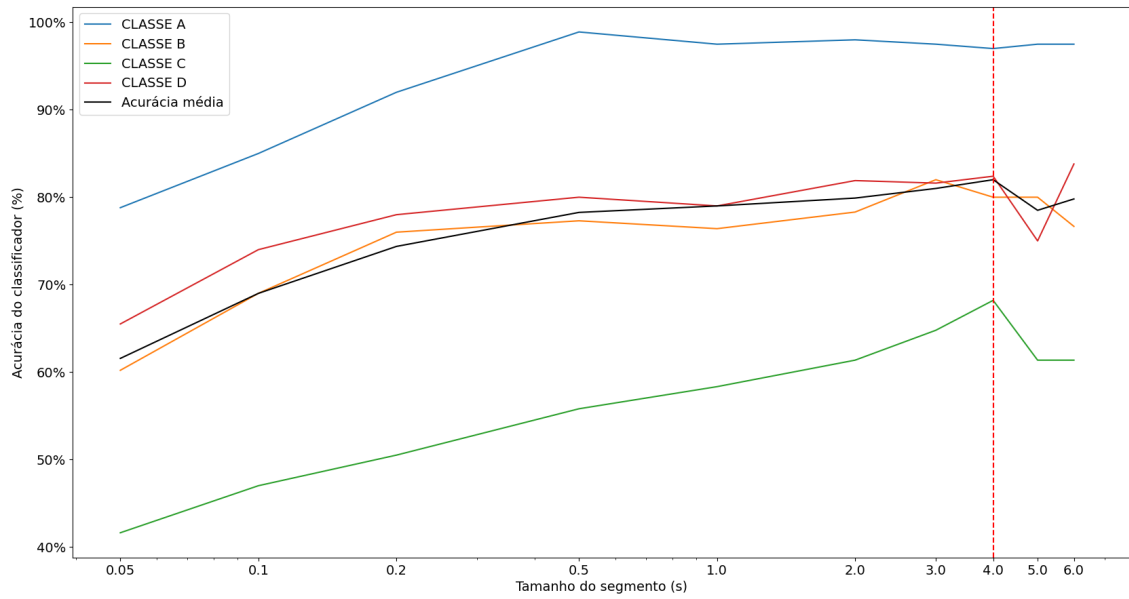
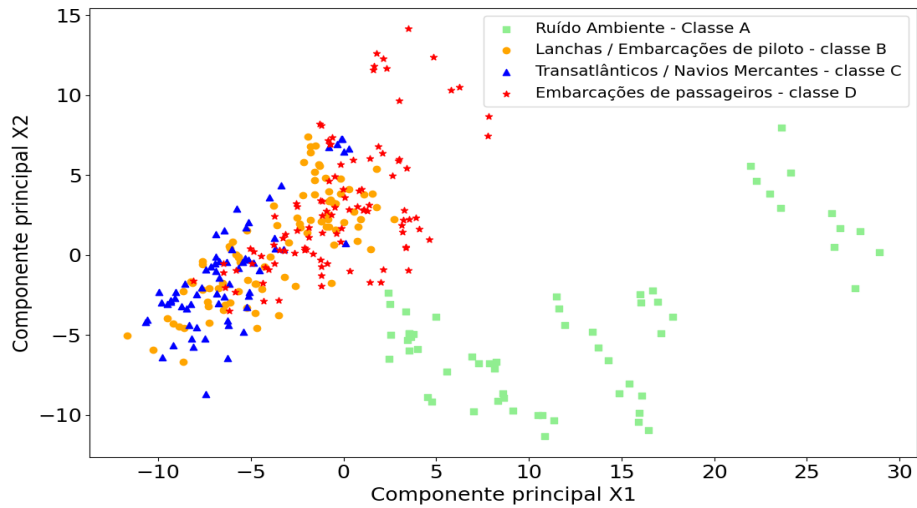
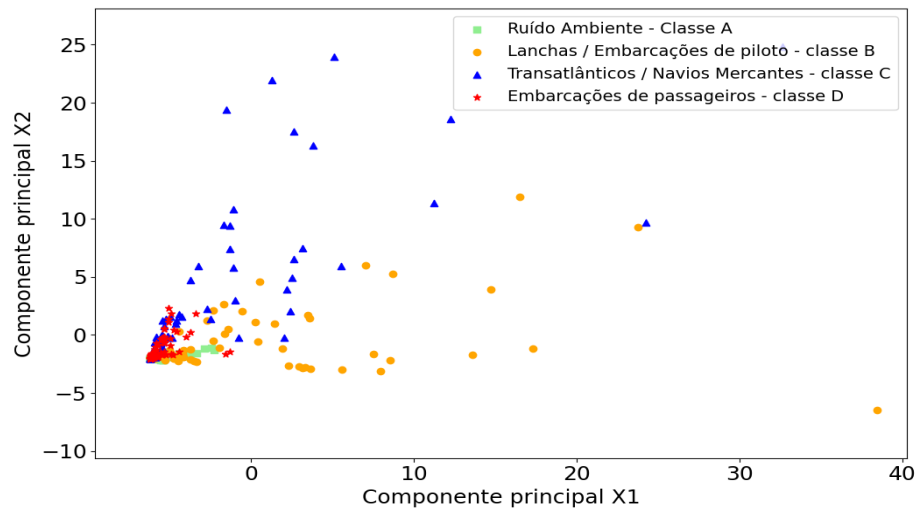


Figura 5.3: Efeito da variação do tamanho do segmento sobre a acurácia do classificador SVM. Em preto a acurácia média das 4 classes. Segmentos de 4 segundos tiveram melhor desempenho para o contexto de dados utilizados

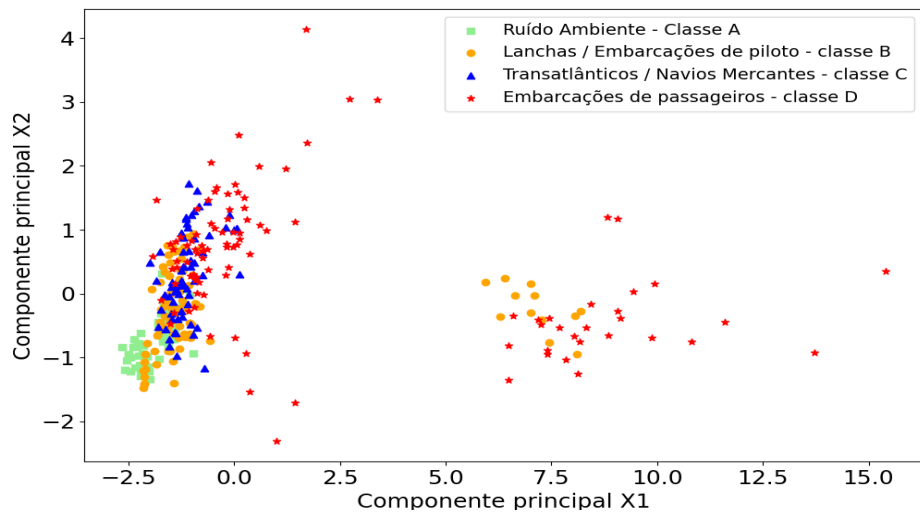
o que permite uma classificação mais eficiente dos algoritmos de ML.



(a) $MFCC^\dagger + SDF$



(b) Coeficientes espectrais de potência (PSC)



(c) LPC

Figura 5.4: Representação gráfica da distribuição em 2-D dos vetores de características extraídos por 3 métodos distintos. O vetor de característica é reduzido a 2 dimensões representados como componentes principais X1 e X2. a) vetor $MFCC^\dagger + SDF$, b) vetor com os Coeficientes espectrais de potência (PSC) e c) vetor de características LPC .

5.3 Classificação das embarcações marítimas utilizando dados coletados na Enseada dos Anjos - Arraial do Cabo- RJ

Nesta seção foi avaliado o desempenho de 4 classificadores utilizando como entrada 13 vetores de características distintos, similar ao realizado na Seção 5.2, porém utilizando o conjunto de dados de gravações de passagens de embarcações pela Enseada dos Anjos descrito na Seção 4.2.

Os sinais de áudio das embarcações e do ruído ambiente foram amostrados na frequência original de gravação de 50kHz. Na etapa de pré-processamento, foi adotado os mesmos parâmetros da Seção 5.2.

A avaliação da acurácia foi realizada adotando o critério de validação cruzada *leave-one-out*, devido ao número reduzido de arquivos de áudio, onde se testa todos os segmentos de um arquivo de áudio por vez. Para tanto, foi retirado do conjunto de dados apresentado na Tabela 4.3, composto por 19 arquivos e 390 segmentos, um arquivo por vez e os demais restantes formariam o conjunto de treinamento. Como cada arquivo tinha uma duração distinta, a proporção dos dados entre treinamento e teste variava a cada simulação. Por exemplo, em uma simulação em que se retirou do conjunto de dados um arquivo da classe A para teste, com duração de 116 segundos e 29 segmentos a proporção ficou 361:29 (treinamento:teste). Em outra simulação em que foi retirado um arquivo de 88 segundos e 22 segmentos a proporção ficou 378:22.

A Tabela 5.4 trás os resultados da classificação para cada vetor de característica referente aos quatro tipo de classificadores. Os parâmetros dos classificadores foram os mesmos utilizados na Seção 5.2.

A análise dos resultados indica que, neste caso, o vetor de características $MFCC^\dagger$ obteve melhores resultados em termos de desempenho quando comparado ao vetor MFCC tradicional. No entanto, ao combinar o vetor de características $MFCC^\dagger$ com o vetor SDF, não foi observada uma melhoria adicional nos resultados. Embora o vetor de características combinado tenha demonstrado um alto índice de acurácia (92%), as características espectrais dinâmicas não contribuíram de forma significativa para uma maior distinção entre os sinais, além das características já presentes no vetor $MFCC^\dagger$.

A pouca variabilidade do sinal dentro de uma mesma classe foi determinante para o resultado, haja visto que, os dados foram coletados em um mesmo local em dias com condições ambientais semelhantes, e com pouca alteração do regime de máquina das embarcações amostradas. Como mencionado anteriormente, a capacidade discriminativa do vetor SDF depende da variação espectral entre os sinais ao

Tabela 5.4: Comparação da porcentagem de classificações corretas entre o vetor de características proposto e outros 12 tipos de vetores de características para 4 tipos de classificadores, com dados obtidos na enseada dos Anjos.

Vetor de Características	Taxa de Classificação			
	SVM	K-NN	RF	DT
<i>SDF</i> (168)	77.7%	67.4%	84%	83%
<i>MFCC</i> [†] (42)	92%	82%	83%	82%
<i>MFCC</i> (42)	86%	81%	73%	75%
<i>LPC</i> (12)	35%	34%	41%	42.5%
<i>LPCC</i> (12)	37%	38%	36%	34%
<i>PSC</i> (960)	35%	33%	78%	75%
<i>ZCR</i> (376)	30%	32%	34%	33%
<i>MFCC</i> [†] + <i>SDF</i> (210)	92%	83%	82.5%	82%
<i>MFCC</i> + <i>SDF</i> (210)	89%	78.5%	77%	76%
<i>LPC</i> + <i>SDF</i> (180)	82%	67%	82%	81%
<i>LPCC</i> + <i>SDF</i> (180)	75%	67%	82.5%	82.5%
<i>PSC</i> + <i>SDF</i> (1128)	65%	50%	82%	81%
<i>ZCR</i> + <i>SDF</i> (544)	63%	54%	82%	79%

*Os valores entre parênteses significam a dimensão do vetor de características.

longo de *frames* sucessivos.

Uma outra consideração importante é o número limitado de classes analisadas. A separação em apenas 3 classes pode ter influenciado positivamente o desempenho do vetor *MFCC*[†], tornando-se desnecessária a inclusão do vetor *SDF* para aprimorar os resultados. A Figura 5.5 trás a matriz de confusão da classificação do conjunto de dados utilizando o vetor de características *MFCC*[†] e o vetor de características combinado. Os resultados obtidos para cada classe indicaram que não ocorreu melhora no percentual de acertos com a adição do vetor *SDF*.

Também cabe destacar que a seleção adequada dos filtros Mel resultando no vetor *MFCC*[†], proporcionou um aumento de 6 pontos percentuais na acurácia em comparação ao vetor composto pelos *MFCC*s tradicionais, classificados pelo algoritmo *SVM*, tendo aumentado também nos outros classificadores.

Os demais extratores de características, *LPC*, *LPCC*, *PSC* e *ZCR*, não tiveram bom desempenho, quando utilizados isoladamente como entrada para os classificadores, com exceção dos *PSC* classificados por *RF* que alcançaram uma acurácia de 78%. No entanto, a combinação desses extratores com o vetor *SDF* elevou o percentual de acerto em todas as análises. A Figura 5.6, por exemplo, trás a representação da matriz de confusão da classificação do conjunto de dados pelo extrator de característica *LPC* em sua forma isolada e em combinação com o vetor *SDF*, ambos classificados pelo classificador *SVM*. Observa-se que o *LPC*, tem um desempenho inferior a 65% para Classe A, isto significa que o mesmo tem uma boa capacidade

a)

		CLASSE PREVISTA		
		CLASSE A	CLASSE B	CLASSE C
CLASSE REAL	CLASSE A	141 (97.24%)	4 (2.75%)	0 (0%)
	CLASSE B	5 (3.6%)	130 (93.52%)	4 (2.87%)
	CLASSE C	0 (0%)	12 (13.18%)	79 (86.81%)

b)

		CLASSE PREVISTA		
		CLASSE A	CLASSE B	CLASSE C
CLASSE REAL	CLASSE A	141 (97.25%)	3 (2.1%)	1 (0.7%)
	CLASSE B	8 (5.75%)	130 (93.52%)	1 (0.71%)
	CLASSE C	2 (2.2%)	10 (11%)	79 (86.81%)

Figura 5.5: Matriz de confusão obtida utilizando como entrada para o algoritmo *SVM* , os vetores de características (a) $MFCC^\dagger$ e (b) $MFCC^\dagger + SDF$

de discriminação entre o ruído ambiental e as demais classes de embarcações. A adição do *SDF* a esse extrator elevou a acurácia da classe A para 97.5%, promovendo melhoria também nas demais classes.

As Figuras 5.7 e 5.8 mostram respectivamente a distribuição das 2 componentes principais do vetor *LPC* e do vetor $LPC + SDF$. Observa-se na Figura 5.8 as características estão mais bem agrupadas enquanto que na Figura 5.7 as características estão mais dispersas. Esse agrupamento ocasionado pela adição do vetor *SDF*, favorece a separação de classes.

a)

		CLASSE PREVISTA		
		CLASSE A	CLASSE B	CLASSE C
CLASSE REAL	CLASSE A	93 (64.13%)	34 (23.45%)	18 (12.41%)
	CLASSE B	43 (30.9%)	38 (27.34%)	58 (41.72%)
	CLASSE C	39 (42.8%)	45 (49.45%)	7 (7.69%)

b)

		CLASSE PREVISTA		
		CLASSE A	CLASSE B	CLASSE C
CLASSE REAL	CLASSE A	141 (97.25%)	3 (2.1%)	1 (0.7%)
	CLASSE B	14 (10%)	117 (84.17%)	8 (5.75%)
	CLASSE C	22 (24.17%)	22 (24.17%)	47 (51.64%)

Figura 5.6: Matriz de confusão obtida utilizando como entrada para o algoritmo *SVM*, os vetores de características (a) *LPC* e (b) *LPC + SDF*. A combinação com o vetor *SDF* elevou a acurácia em todas as classes.

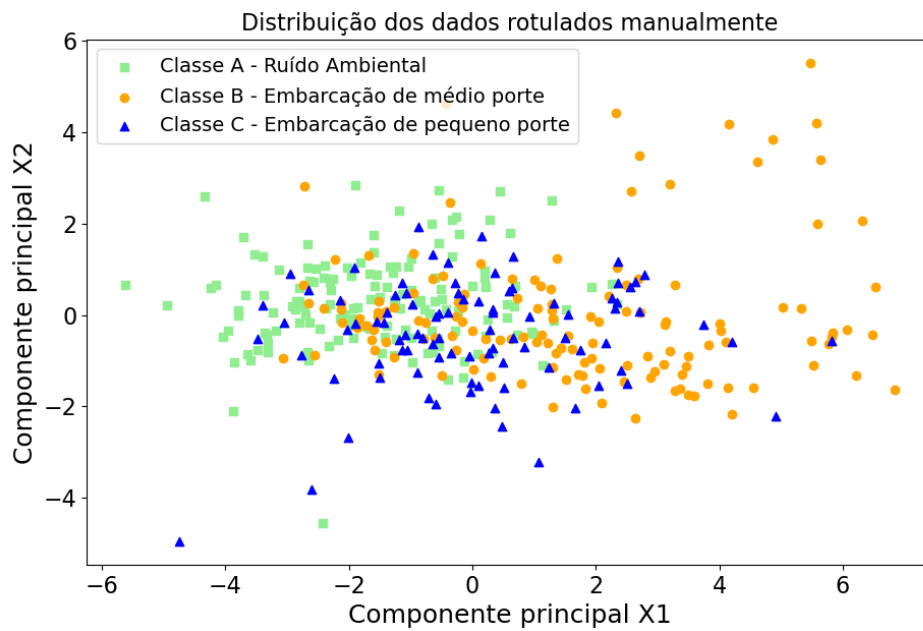


Figura 5.7: Distribuição bidimensional das características principais do vetor *LPC* com rótulos previamente definidos.

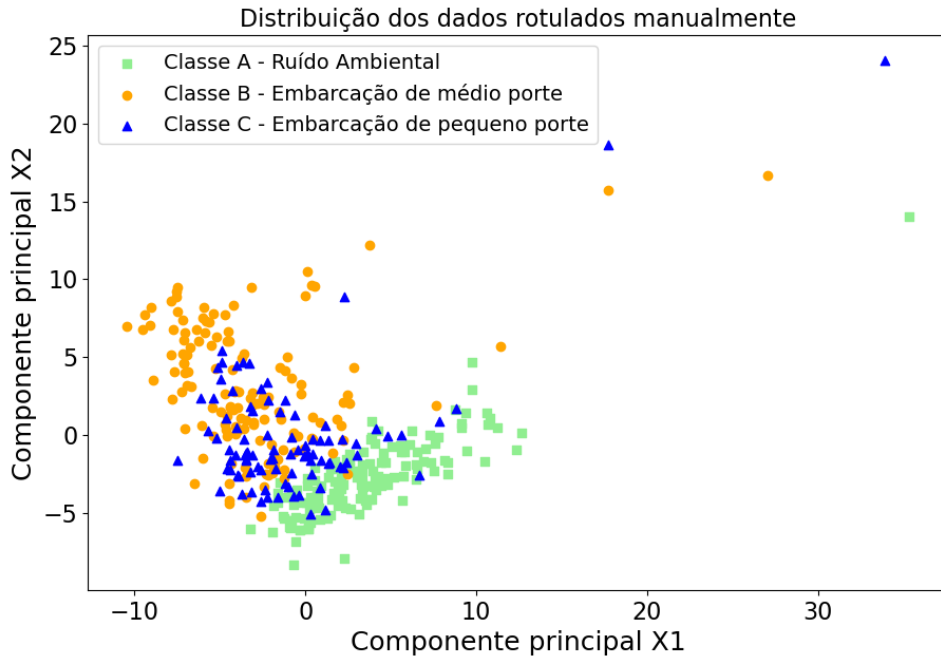


Figura 5.8: Distribuição bidimensional das características principais do vetor $LPC + SDF$ com rótulos previamente definido

5.4 Análise de robustez dos extratores de características

Nesta seção a robustez do extrator de característica proposto e a acurácia na classificação foi avaliada com base na metodologia apresentada no trabalho de mestrado de ALMEIDA *et al.* [35], que faz uma análise da dispersão das características quando adicionamos ruído a um sinal. Para tanto, foi retirado do conjunto de dados *ShipsEar*, descrito na Seção 4.1, 1 arquivo de gravação de ruído irradiado por uma embarcação classe D (Embarcação de passageiros). Esse arquivo deixou de fazer parte do conjunto de dados de treinamento do algoritmo. Assim dos 53 arquivos, 52 são mantidos para geração do conjunto de treinamento, e o arquivo retirado será testado n vezes, onde, a cada teste esse sinal é modificado pela adição de ruído.

Aqui, o arquivo foi subdividido em 12 segmentos de 2 segundos a serem classificados. O segmento de sinal desse arquivo de teste foi expresso como s_k , sendo $k=1,2,3,\dots,12$. s_k são os segmentos do sinal de teste original captado pelos hidrofo-nes e disponibilizados no conjunto de dados *ShipsEar*. Trata-se do sinal irradiado pela embarcação somado ao ruído ambiente⁴ local.

A análise do efeito dispersivo das características no espaço de características se dá a partir da adição de ruído ao sinal de teste original. Cabe destacar que, essa adição ocorreu somente no arquivo de teste, não ocorrendo para os demais 52 arquivos que compunham a base de treinamento. Em se tratando de um algoritmo

supervisionado, a adiç o do ru do a todos os arquivos, inclusive aos arquivos de treinamento, ocasionaria no treinamento do algoritmo com o sinal ruidoso, o que poderia simplificar a classifica o e n o refletir de forma precisa a influ ncia real da adi o de ru do.

Foi extra do desse arquivo de teste um vetor de caracter sticas por segmento do sinal original, livre de ru do adicional, para quatro tipos de extratores, representados por, x_{SDF} , $x_{MFCC^\dagger+SDF}$, x_{MFCC} e x_{MFCC^\dagger} . Foram extra dos tamb m, novos vetores de caracter sticas, xm_{SDF} , $xm_{MFCC^\dagger+SDF}$, xm_{MFCC} e xm_{MFCC^\dagger} , a partir do sinal original com adi o de ru do.

O c culo da distancia euclidiana entre o vetores de caracter sticas, x_{SDF} , $x_{MFCC^\dagger+SDF}$, x_{MFCC} , x_{MFCC^\dagger} do sinal original e os vetores de caracter sticas, xm_{SDF} , $xm_{MFCC^\dagger+SDF}$, xm_{MFCC} , xm_{MFCC^\dagger} do sinal com perturba o, normalizado pelo vetores de caracter sticas do sinal original, foi utilizada como m trica para avaliar a dispers o no espa o de caracter sticas, denominada aqui como ds . Esta medida, representada por $ds_{SDF/MFCC^\dagger+SDF/MFCC/MFCC^\dagger}$,   definida pela Equa o 5.1.

$$ds_{SDF/MFCC^\dagger+SDF/MFCC/MFCC^\dagger} = \frac{\sqrt{\sum_{i=1}^n (x_{i_{SDF/MFCC^\dagger+SDF/MFCC/MFCC^\dagger}} - xm_{i_{SDF/MFCC^\dagger+SDF/MFCC/MFCC^\dagger}})^2}}{\sqrt{\sum_{i=1}^n (x_{i_{SDF/MFCC^\dagger+SDF/MFCC/MFCC^\dagger}})^2}} \quad (5.1)$$

Sendo n o n mero de dimens es do vetor de caracter sticas.

A Figura 5.9   uma representa o do c culo da dist ncia entre esses vetores de caracter sticas. Em azul o sinal de teste original sem ru do e em marrom claro esse mesmo sinal com adi o de ru do. Na sequencia s o extra dos 4 tipos de vetores de caracter sticas para ambos sinais. E por fim observa-se a dispers o desses vetores no espa o de caracter stica e a dist ncia entre eles.

Para cada segmento foi calculado um $ds(k)_{SDF}$, $ds(k)_{MFCC^\dagger+SDF}$, $ds(k)_{MFCC}$ e $ds(k)_{MFCC^\dagger}$, sendo k o segmento em analise, e, por fim, calculada as m dias $Ds(k)_{SDF}$, $Ds(k)_{MFCC^\dagger+SDF}$, $Ds(k)_{MFCC}$ e $Ds(k)_{MFCC^\dagger}$ referente a cada vetor de caracter sticas ao longo dos 6 segmentos.

Os valores $Ds(k)_{SDF}$, $Ds(k)_{MFCC^\dagger+SDF}$, $Ds(k)_{MFCC}$ e $Ds(k)_{MFCC^\dagger}$ representam as varia es ocorridas no espa o de caracter stica ocasionada pela adi o de ru do (e.g. perturba o) ao sinal original. A robustez do extrator esta condicionada a menor varia o poss vel das caracter sticas extra das, ou seja, quanto menor o Ds mais pr ximas est o as caracter sticas do extra das do sinal original das caracter sticas extra das do sinal perturbado [35].

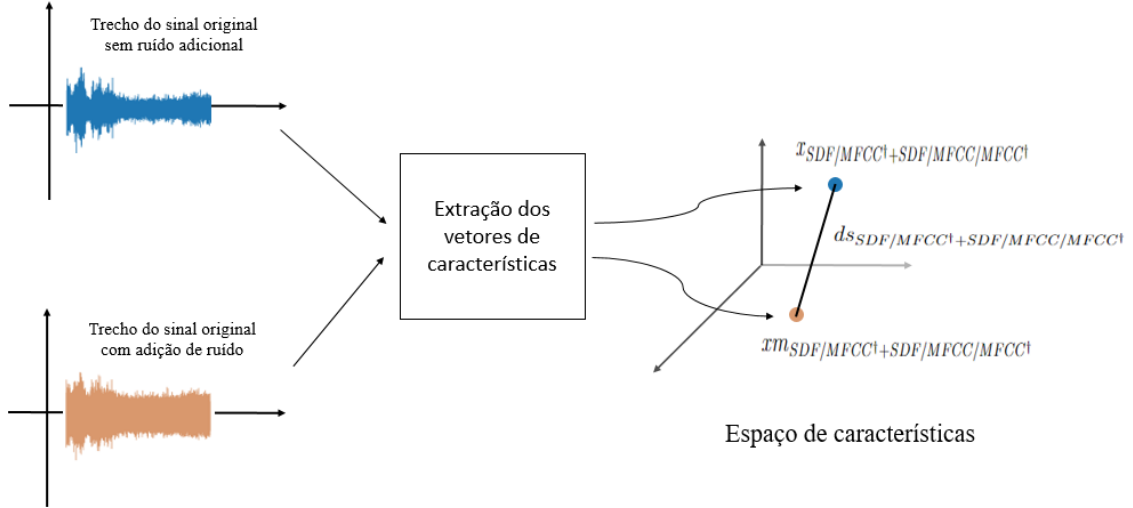


Figura 5.9: Representação do cálculo da distância entre vetores de características $d_{SDF/MFCC^{\dagger}+SDF/MFCC/MFCC^{\dagger}}$. Fonte: adaptado de ALMEIDA *et al.* [35]

5.4.1 Avaliação da robustez pela adição de ruído gaussiano branco (AWGN)

O primeiro teste envolveu a adição de um ruído gaussiano branco aditivo (AWGN, *Additive White Gaussian Noise*) aos segmentos s_k do sinal de teste original. Os segmentos com ruído adicional, $sr_k(m)$, foram gerados através da soma de $s_k(m)$ com um ruído branco $r(m)$ de média zero, resultando na seguinte equação: $sr_k(m) = s_k(m) + r(m)$, sendo $m = 1, 2, 3, \dots, M$ o número de amostras por segmento. Para simular a influência do ruído na extração do vetor de características proposto, diferentes *SNRs* foram calculadas subtraindo o nível de pressão sonora do ruído branco aditivo do nível de pressão sonora do sinal de teste original referentes a $1\mu Pa$ conforme Equação 5.2.

$$SNR_{dB} = SPL_{s_k} - SPL_r \quad (5.2)$$

O valor médio de SPL_{s_k} calculado para o arquivo contendo o sinal de teste original a ser validado, foi de $81 dB re 1 \mu Pa$, para uma faixa de frequência compreendida entre 0 e $fs/2$, sendo fs igual a $52.734 Hz$. Esse valor, conforme já mencionado, representa o nível de pressão sonora do sinal de teste original captado pelos hidrofonos. No entanto, mesmo que esse sinal não correspondesse apenas ao ruído irradiado pela embarcação e sim ao ruído irradiado pela embarcação mais o ruído ambiente local, a utilização do mesmo como arquivo de teste, não acarretou prejuízos ao experimento, uma vez que, a comparação dos vetores de características ocorre entre o vetor extraído deste sinal de teste original e o vetor extraído do mesmo sinal após adição do ruído branco.

Para simular as diferentes $SNRs$, a variância do ruído branco foi ajustada para alcançar o SPL_r que correspondesse a SNR desejada, segundo a Equação 5.3².

$$\sigma_{\text{ruído branco}}^2 = \frac{\sigma_s^2}{10^{\frac{SNR_{\text{desejada}}}{10}}} \quad (5.3)$$

Onde:

$\sigma_{\text{ruído}}^2$ representa a variância do ruído branco.

σ_s representa a variância do sinal original utilizado para teste.

SNR_{desejada} é a relação sinal-ruído desejada.

O sinal de ruído branco que atende a SNR desejada, foi sintetizado multiplicando um conjunto de N amostras geradas de modo aleatório com o mesmo tamanho do sinal de teste original, obedecendo uma distribuição normal, de média zero e variância calculada pela Equação 5.3.

Dois exemplos de configurações testadas são mostradas nos gráficos da segunda e terceira linha da Figura 5.10, onde foi adicionado ruído branco com SPL_r de $64\text{ dB re } 1\ \mu\text{Pa}$ e $74\text{ dB re } 1\ \mu\text{Pa}$, para atingir as $SNRs$ desejadas de $81 - 64 = 17\text{ dB}$ e $81 - 74 = 7\text{ dB}$, respectivamente.

As medidas $Ds(k)_{SDF}$, $Ds(k)_{MFCC^\dagger+SDF}$, $Ds(k)_{MFCC}$ e $Ds(k)_{MFCC^\dagger}$, foram então estimadas para diferentes valores de SNR , variando entre -15 dB e 30 dB , sendo realizadas um total de 100 simulações de Monte Carlo para cada configuração de SNR desejada. Os resultados do experimento estão ilustrados na Figura 5.11.

O gráfico da Figura 5.11 revela uma diminuição na distância entre as características extraídas do sinal original e as características extraídas do sinal com ruído à medida que aumentamos a SNR . Isso indica que, à medida que se diminui a intensidade do ruído introduzido, o novo vetor de características extraído do sinal com ruído se aproxima do vetor de características extraído do sinal original, mantendo a acurácia do classificador próxima à acurácia prévia à adição de ruído. Além disso, o método de extração do vetor de características combinado ($MFCC^\dagger+SDF$) demonstrou maior robustez em comparação ao método de extração dos $MFCCs$ tradicionais, evidenciado pelo menor valor de Ds . Também observou-se que o método de extração do vetor de características SDF apresentou a menor distância entre as características extraídas do sinal original e as características extraídas do sinal ruidoso para $SNRs$ inferiores a 12 dB . Por fim, nota-se que a combinação do SDF

²A Equação 5.3 é derivada da definição da relação sinal-ruído (SNR) em termos de potências. A SNR é definida como: $SNR = \frac{\text{Potência do Sinal}}{\text{Potência do Ruído}}$. A variância é uma medida da dispersão dos valores em torno da média de um conjunto de dados. Para um sinal de entrada x , a variância é dada por: $\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$, sendo N é o número de amostras no sinal, x_i é o valor da amostra i , \bar{x} é a média das amostras. Ao assumir que a média do ruído é zero (o que é comum em muitos casos de ruído branco), podemos simplificar a expressão para a potência do ruído: $\text{Potência do Ruído} = \frac{1}{N} \sum_{i=1}^N x_i^2 = \text{Variância do Sinal}$

Portanto, a variância do ruído (e consequentemente a variância do sinal) pode ser calculada como a potência do sinal dividida pela SNR desejada: $\text{Variância do Ruído} = \frac{\text{Potência do Sinal}}{SNR_{\text{desejada}}}$

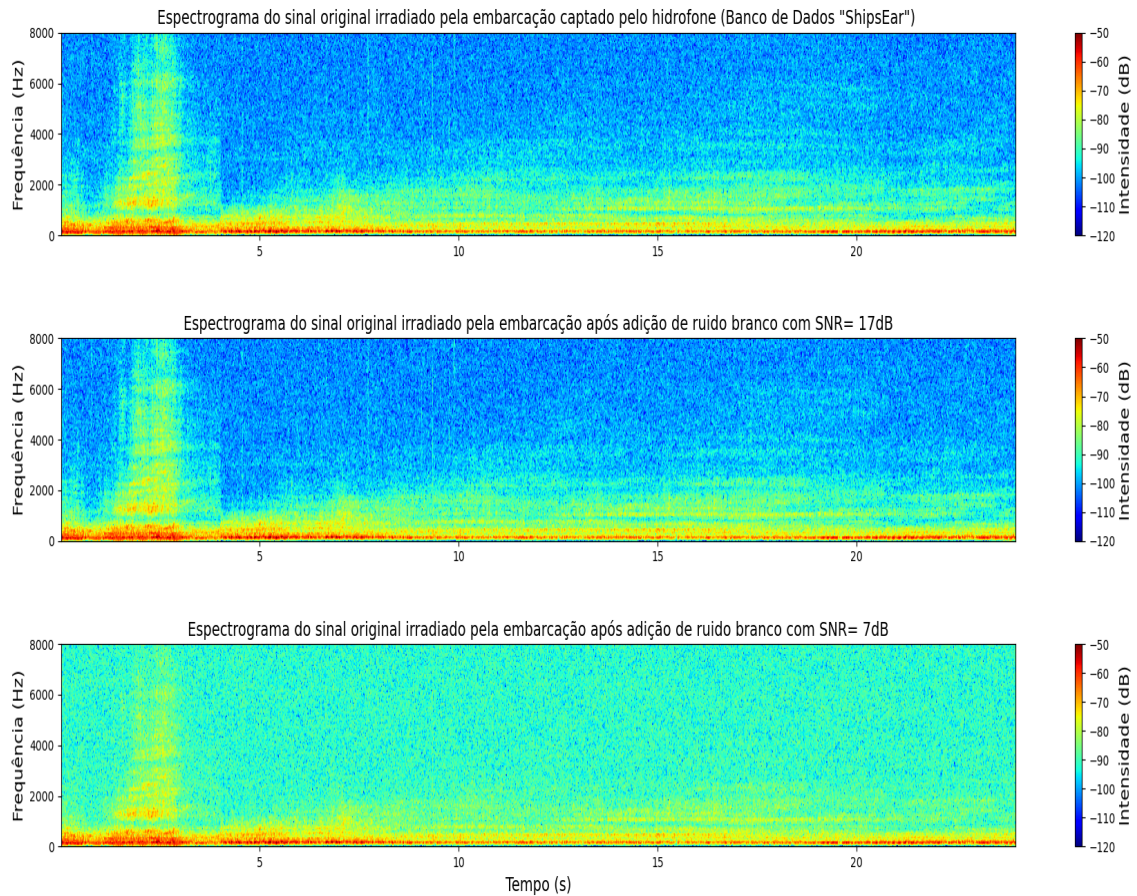


Figura 5.10: Espectrogramas do sinal original e do sinal original após adição de ruído branco gaussiano a uma SNR = 17dB e SNR= 7 dB

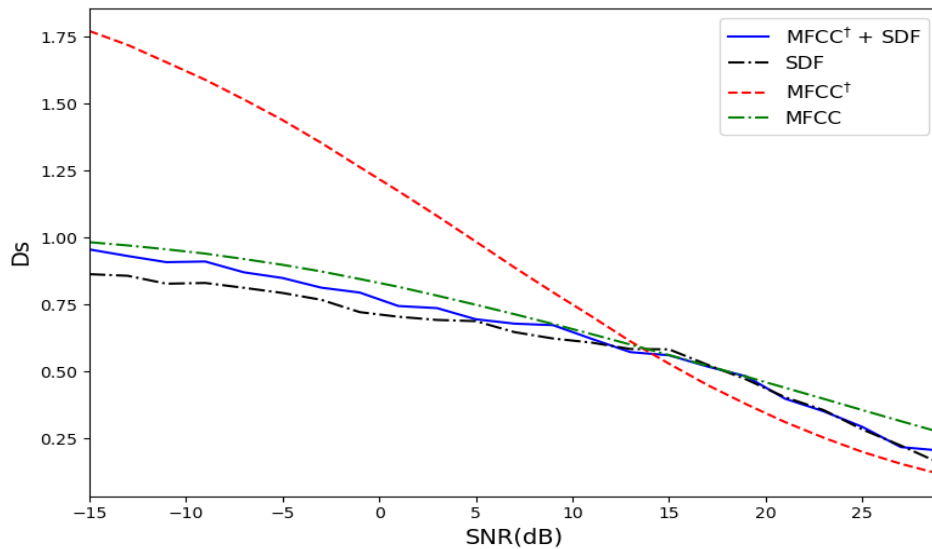


Figura 5.11: Dispersão do espaço de características (D_s) para diferentes níveis de SNR devido a adição de $AWGN$ utilizando os extratores de características $MFCC^\dagger + SDF$, SDF , $MFCC^\dagger$ e $MFCC$.

ao vetor $MFCC^\dagger$ melhora a robustez do extrator em situações de baixo SNR .

5.4.2 Avaliação da robustez com a adição de ruído coletado na Enseada dos Anjos- Arraial do Cabo - RJ

Além do ruído branco sintetizado, também foram empregadas nos experimentos envolvendo distorção aditiva, amostras de ruído real coletadas na Enseada dos Anjos, Arraial do Cabo-RJ. Aos segmentos do sinal original s_k , cujo valor médio do SPL_{s_k} calculado, foi de $81\text{ dB re } 1\ \mu\text{Pa}$, foi adicionado inicialmente um sinal oriundo de uma gravação de ruído coletado na Enseada do Anjos às 06 horas do dia 15 de outubro de 2023, com total de amostras igual ao número de amostras do sinal original, com a mesma fs e com SPL calculado de 74 dB , para uma faixa de frequência compreendida entre 0 e $fs/2$. Utilizando a equação 5.2, onde neste caso SPL_r refere-se ao ruído coletado na Enseada dos Anjos teve-se inicialmente uma SNR de $81\text{ dB} - 74\text{ dB} = 7\text{ dB}$. Na demais simulações a SNR foi ajustada para alcançar a SNR desejada mediante a aplicação de um fator de ajuste nas amostras do sinal do ruído gravado na Enseada, calculado conforme equações abaixo:

$$\text{potência_embarcação} = \frac{1}{N} \sum_{i=0}^{N-1} x_i^2 \quad (5.4)$$

$$\text{potência_ruído_desejada} = \frac{\text{potência_emb}}{10^{(SNR_desejada/10)}} \quad (5.5)$$

$$\text{potência_ruído_Enseada_dos_Anjos} = \frac{1}{M} \sum_{j=0}^{M-1} y_j^2 \quad (5.6)$$

$$\text{fator_ajuste} = \sqrt{\frac{\text{potência_ruído_desejada}}{\text{potência_ruído_Enseada_dos_Anjos}}} \quad (5.7)$$

Onde x_i representa as amostras do sinal original, y_i representa as amostras do ruído gravado na Enseada dos Anjos, e N e M referem-se ao número de amostras em cada sinal, respectivamente. A Figura 5.12 apresenta os espectrogramas da combinação do sinal original com o ruído.

Os resultados da análise estão ilustrados na Figura 5.13, e apontam para um comportamento menos dispersivo do vetor de características SDF e consequentemente, também, do vetor de características combinado nos cenários com menores SNR , quando avaliado a distância para as características do sinal original. Isso significa que as características extraídas por esse método são mais estáveis em condições ambientais adversas onde o nível de ruído é elevado, tornando-o um extrator mais robusto. Além disso, o método de extração do vetor de características SDF e

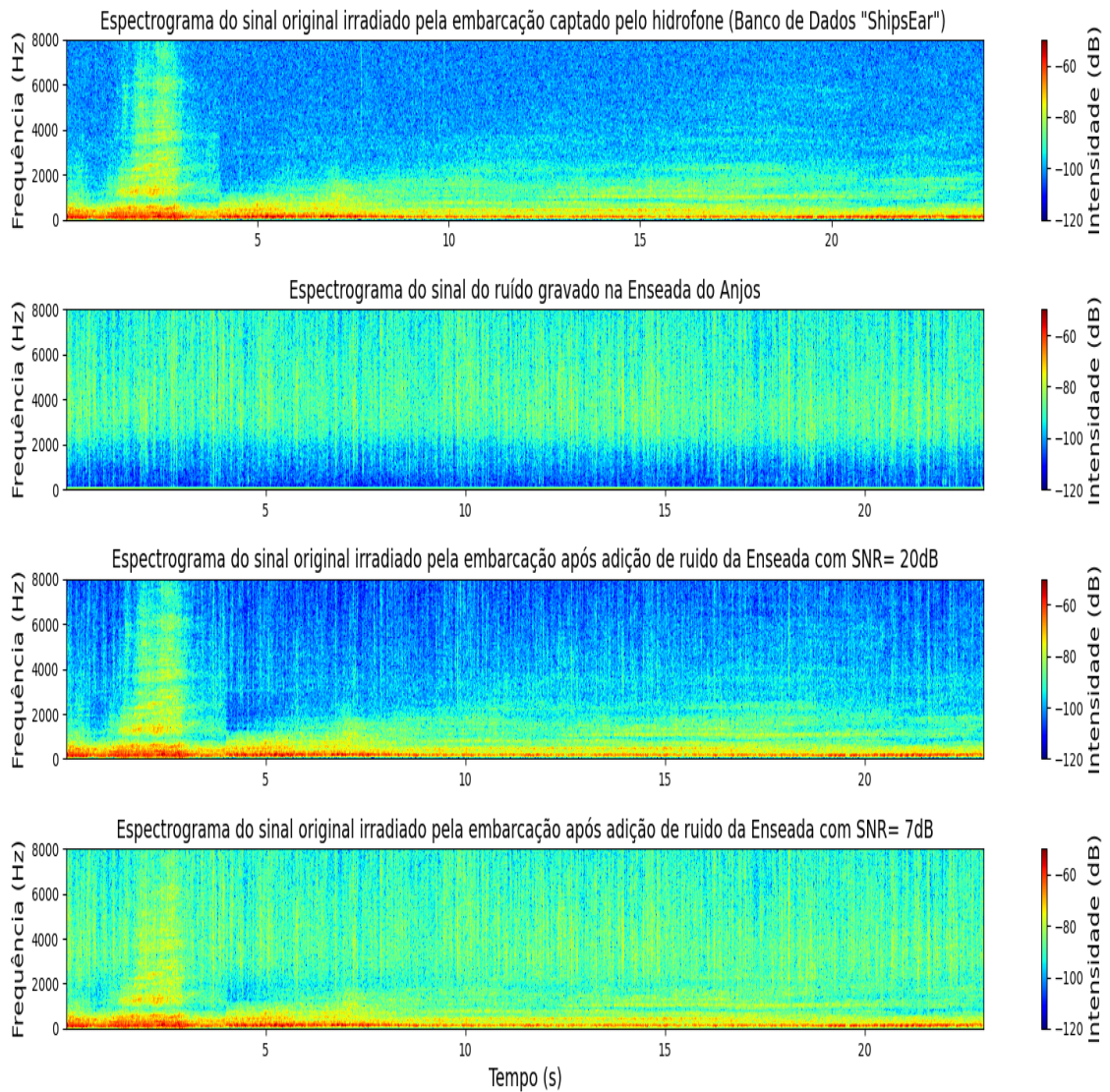


Figura 5.12: Espectrogramas do sinal de teste original da embarcação (Banco de Dados "ShipsEar"), do sinal de ruído gravado na Enseada dos Anjos e do sinal de teste original da embarcação após adição de ruído gravada na Enseada a uma SNR = 20dB e SNR= 7dB

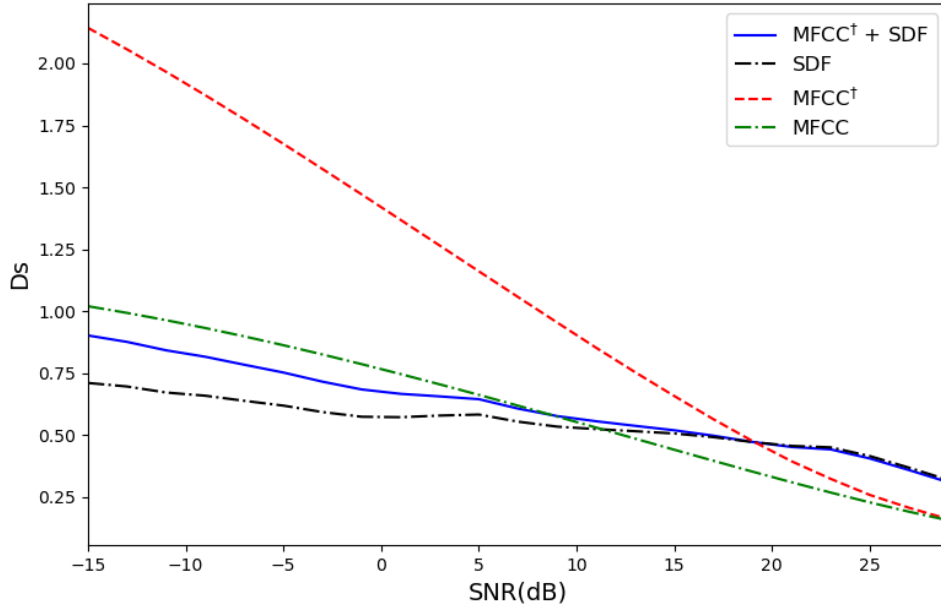


Figura 5.13: Dispersão do espaço de características (D_s) para diferentes níveis de SNR devido a adição do ruído coletado na Enseada dos Anjos utilizando os extratores de características $MFCC^\dagger + SDF$, SDF , $MFCC^\dagger$ e $MFCC$.

do vetor de características combinado ($MFCC^\dagger + SDF$) apresentou valores de D_s menores que os extratores $MFCC$ e ($MFCC^\dagger$ em situações de baixa SNR).

Os resultados sugerem que a robustez do extrator por si só não garante a acurácia na classificação. Embora as características extraídas por um determinado extrator possam apresentar pouca variação quando perturbadas, sua distribuição no espaço de características pode ser tão dispersa que dificulta o agrupamento e a classificação. Assim, a eficácia na categorização não é meramente uma função da estabilidade das características, mas também da forma como essas características se distribuem no espaço.

Também foi avaliado o percentual de classificações corretas dos 12 segmentos do arquivo de teste para cada uma das configurações de SNR , utilizando o extrator de características proposto, apresentado na Tabela 5.5, com adição de ruído $AWGN$ e ruído proveniente da Enseada do Anjos. Os sinais de teste originais tiveram 100% dos segmentos classificados corretamente pelo classificador SVM . No entanto, os dados indicam que, a adição de ruído as gravações de teste afetou o percentual de classificações corretas, e, como esperado, o desempenho do classificador é deteriorado com a diminuição da SNR . Observou-se também que, dentre os extratores de características avaliados, o vetor de característica combinado obteve o melhor desempenho mesmo em condições de baixa SNR .

Tabela 5.5: Resultado da classificação, pelo algoritmo SVM, dos 12 segmentos do sinal de teste com adição de ruído.

Extrator de Características	$MFCC^\dagger + SDF$	$MFCC^\dagger$	$MFCC$	SDF
Sinal original - "limpo"	100 %	91 %	91 %	91 %
AWGN (SNR -1 dB)	66,6 %	66,6 %	0 %	16,6 %
AWGN (SNR 3 dB)	91,6 %	83,3 %	16,6 %	16,6 %
AWGN (SNR 7 dB)	100 %	91,6 %	42 %	42 %
AWGN (SNR 21 dB)	100 %	91,6 %	75 %	75 %
Ruído da Enseada (SNR 7 dB)	50 %	33,3 %	50 %	50 %
Ruído da Enseada (SNR 10 dB)	75 %	58,3 %	75 %	66 %
Ruído da Enseada (SNR 13 dB)	91,6 %	84 %	91,6 %	66 %
Ruído da Enseada (SNR 21 dB)	100 %	91,6 %	91,6 %	75 %

Capítulo 6

Conclusões e sugestões para trabalhos futuros

Esta dissertação apresentou uma metodologia de extração de características baseada em cepstrum para a classificação de embarcações marítimas. Dois conjuntos de dados experimentais, compostos por sinais de ruído irradiado por embarcações, coletados na costa da Espanha e gravados da Enseada dos Anjos, Arraial do Cabo - RJ, foram utilizados para validar a eficácia do algoritmo de extração do vetor de características combinado.

Quatro classificadores foram avaliados: SVM, RF, KNN e DT. O SVM apresentou o melhor desempenho na maioria dos casos, com exceção de algumas simulações com os extratores ZCR, PSC e LPCC no conjunto ShipsEar e ZCR, PSC, LPCC, LPC no conjunto Arraial do Cabo. Para o caso específico em que a entrada para os classificadores foi o vetor de características combinado ($MFCC^\dagger + SDF$), o SVM superou os demais classificadores em pelo menos 8% no ShipsEar e 9% em Arraial do Cabo. A superioridade do SVM se deve à sua capacidade de lidar com problemas de alta dimensionalidade e dados não lineares.

No que tange à acurácia dos extratores de características, o $MFCC^\dagger + SDF$ também se mostrou mais preciso do que todos os outros extratores, com exceção das simulações para o conjunto "ShipsEar" envolvendo os classificadores RF e DT, cujos resultados apontaram para um desempenho superior do extrator MFCC + SDF. No entanto os melhores resultados entre todos os cenários avaliados, foram observados com o classificador SVM tendo como entrada o vetor de características $MFCC^\dagger + SDF$. Essa combinação superou os demais classificadores em ambos os conjuntos de dados, alcançando 94% de acurácia no ShipsEar e 92% no conjunto de dados de Arraial.

A combinação do SDF com os $MFCC^\dagger$ aprimorou a representação das características acústicas das embarcações do conjunto de dados ShipsEar, uma vez que, combinou em um único vetor, um conjunto de características estáticas e dinâmica

tendo uma acurácia de 94% enquanto que o $MFCC^\dagger$ em sua forma isolada, obteve uma acurácia de 86,5%. Porém não foi observado a mesma melhoria para o conjunto de Arraial do Cabo, fato este possivelmente relacionado a pouca variação do regime de máquinas das embarcações coletadas, bem como uma menor variabilidade de tipos de embarcações. Ainda, levando em consideração a separação entre classes, o vetor SDF e a combinação dele com os demais extratores, mostrou-se eficiente em separar as embarcações do ruído ambiental, alcançando 100% de classificações corretas de sinais de ruído ambiental.

Quanto à robustez, o vetor de características combinado e o SDF individualmente apresentaram menor dispersão das características no espaço de características após a adição de ruído ao sinal de teste, demonstrando maior robustez à degradação do sinal. Essa robustez foi confirmada pela análise com ruído real coletado na Enseada dos Anjos, onde a medida de dispersão foi menor para estes extratores com SNRs inferiores a 10 dB. No entanto, é importante ressaltar que a robustez não garante necessariamente uma melhoria na taxa de classificações corretas após a adição de ruído, pois isso também depende da distribuição dos dados no espaço de características.

Partindo da metodologia desenvolvida nessa pesquisa, uma possível direção para trabalhos futuros seria a incorporação de modelagem da propagação acústica para teste de robustez. Isso pode ser feito utilizando softwares de simulação acústica para modelar a propagação do som das embarcações em diferentes ambientes, como águas rasas e águas profundas, variando parâmetros como distância entre a fonte e o receptor, profundidade do guia de ondas submarino, perfil de velocidade do som e tipo de fundo.

Outra sugestão para trabalhos futuros, seria investigar outros métodos de análise para avaliar a robustez da metodologia, uma vez que o método de dispersão das características no espaço de características utilizado neste estudo não correlacionou diretamente a robustez com a acurácia da classificação. Isso pode incluir, por exemplo, uma análise de sensibilidade para identificar quais parâmetros dos extratores de características têm maior impacto na robustez do sistema de classificação, mediante a variação dos parâmetros de configuração dos extratores e avaliação do efeito resultante na acurácia da classificação sob diferentes condições. Por fim sugere-se o uso de aprendizado não supervisionados para descobrir padrões, agrupar instâncias semelhantes e detectar anomalias utilizando o extrator de características combinado.

Referências Bibliográficas

- [1] ZAK, A. “Ships classification basing on acoustic signatures”, *WSEAS Transactions on Signal Processing*, v. 4, 04 2008.
- [2] HAN, J., PEI, J., TONG, H. *Data mining: concepts and techniques*. Morgan kaufmann, 2022.
- [3] SANTOS-DOMÍNGUEZ, D., TORRES-GUIJARRO, S., CARDENAL-LÓPEZ, A., et al. “ShipsEar: An underwater vessel noise database”, *Applied Acoustics*, v. 113, pp. 64–69, 2016.
- [4] YAO, Q., WANG, Y., YANG, Y. “Underwater Acoustic Target Recognition Based on Data Augmentation and Residual CNN”, *Electronics*, v. 12, n. 5, pp. 1206, 2023.
- [5] LIU, J., HE, Y., LIU, Z., et al. “Underwater Target Recognition Based on Line Spectrum and Support Vector Machine”. In: *Proceedings of the 2014 International Conference on Mechatronics, Control and Electronic Engineering*, pp. 79–84. Atlantis Press, 2014/03. doi: 10.2991/mce-14.2014.17.
- [6] NIU, H., LI, X., ZHANG, Y., et al. “Advances and applications of machine learning in underwater acoustics”, *Intelligent Marine Technology and Systems*, v. 1, n. 1, pp. 8, 2023.
- [7] DAS, A., KUMAR, A., BAHL, R. “Marine vessel classification based on passive sonar data: The cepstrum-based approach”. 2013.
- [8] JENSEN, F. B., KUPERMAN, W. A., PORTER, M. B., et al. *Computational ocean acoustics*, v. 2011. Springer, 2011.
- [9] URICK, R. “Principles of underwater sound”, *McGraw-Hill google schola*, v. 2, pp. 2760–2766, 1983.
- [10] LURTON, X. *An introduction to underwater acoustics: principles and applications*. Springer Science & Business Media, 2002.

- [11] BIANCO, M. J., GERSTOFT, P., TRAER, J., et al. “Machine learning in acoustics: Theory and applications”, *The Journal of the Acoustical Society of America*, v. 146, n. 5, pp. 3590–3628, 2019.
- [12] YANG, H., LEE, K., CHOO, Y., et al. “Underwater acoustic research trends with machine learning: general background”, *Journal of Ocean Engineering and Technology*, v. 34, n. 2, pp. 147–154, 2020.
- [13] SABARA, R., SOARES, C., ZABEL, F., et al. “Automatic Acoustic Target Detection and Classification off the Coast of Portugal”. In: *Global Oceans 2020: Singapore – U.S. Gulf Coast*, pp. 1–9, 2020. doi: 10.1109/IEEECONF38699.2020.9389067.
- [14] AZIMI-SADJADI, M., YAO, D., HUANG, Q., et al. “Underwater target classification using wavelet packets and neural networks”, *IEEE Transactions on Neural Networks*, v. 11, n. 3, pp. 784–794, 2000. doi: 10.1109/72.846748.
- [15] KARBASI, M., AHADI, S. M., BAHMANIAN, M. “Environmental sound classification using spectral dynamic features”. In: *2011 8th International Conference on Information, Communications & Signal Processing*, pp. 1–5, 2011.
- [16] ZHANG, Y., JIAO, L., SONGHUA, H. “An efficient method of target classification”. In: *ICSP '98. 1998 Fourth International Conference on Signal Processing (Cat. No.98TH8344)*, v. 2, 1998. doi: 10.1109/ICOSP.1998.770828.
- [17] ZHANG, Q., DA, L., ZHANG, Y., et al. “Integrated neural networks based on feature fusion for underwater target recognition”, *Applied Acoustics*, v. 182, pp. 108261, 2021. doi: <https://doi.org/10.1016/j.apacoust.2021.108261>.
- [18] SHARMA, G., UMAPATHY, K., KRISHNAN, S. “Trends in audio signal feature extraction methods”, *Applied Acoustics*, v. 158, pp. 107020, 2020.
- [19] XIN-XIN, L., SHI-E, Y., MING, Y. “Feature extraction from underwater signals using wavelet packet transform”. In: *2008 International Conference on Neural Networks and Signal Processing*, pp. 400–405, 2008. doi: 10.1109/ICNNSP.2008.4590381.
- [20] CHEN, Y., XU, X. “The research of underwater target recognition method based on deep learning”. In: *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 1–5, 2017. doi: 10.1109/ICSPCC.2017.8242464.

- [21] TONG, Y., ZHANG, X., GE, Y. “Classification and Recognition of Underwater Target Based on MFCC Feature Extraction”. In: *2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 1–4, 2020. doi: 10.1109/ICSPCC50002.2020.9259457.
- [22] LIAN, Z., XU, K., WAN, J., et al. “Underwater acoustic target classification based on modified GFCC features”. In: *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 258–262, 2017. doi: 10.1109/IAEAC.2017.8054017.
- [23] ETTER, P. C. *Underwater acoustic modeling and simulation*. CRC press, 2018.
- [24] KINSLER, L. E., FREY, A. R., COPPENS, A. B., et al. *Fundamentals of acoustics*. John wiley & sons, 2000.
- [25] MEDWIN, H., CLAY, C. S. *Fundamentals of acoustical oceanography*. Academic press, 1997.
- [26] MACKENZIE, K. V. “Nine-term equation for sound speed in the oceans”, *The Journal of the Acoustical Society of America*, v. 70, n. 3, pp. 807–812, 1981.
- [27] GAGGERO, S., GAGGERO, T., RIZZUTO, E., et al. “Ship propeller side effects: pressure pulses and radiated noise”, *Noise Mapping*, v. 3, n. 1, 2016.
- [28] MITSON, R. *Underwater noise of research vessels: review and recommendations*. ICES Cooperative Research Reports (CRR), 1995.
- [29] HE, L., SHEN, X.-H., ZHANG, M.-H., et al. “Segmentation method for ship-radiated noise using the generalized likelihood ratio test on an ordinal pattern distribution”, *Entropy*, v. 22, n. 4, pp. 374, 2020.
- [30] RAO, K. S., MANJUNATH, K. E. *Speech recognition using articulatory and excitation source features*. Springer, 2017.
- [31] MAJEED, S. A., HUSAIN, H., SAMAD, S. A., et al. “Mel Frequency Cepstral Coefficients (MFCC) feature extraction enhancement in the application of speech recognition: a comparison study”, *Journal of Theoretical & Applied Information Technology*, v. 79, n. 1, 2015.
- [32] WINURSITO, A., HIDAYAT, R., BEJO, A. “Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition”. In: *2018 International Conference on Information and Communications Technology (ICOIACT)*, pp. 379–383. IEEE, 2018.

- [33] TRANG, H., LOC, T. H., NAM, H. B. H. “Proposed combination of PCA and MFCC feature extraction in speech recognition system”. In: *2014 international conference on advanced technologies for communications (ATC 2014)*, pp. 697–702. IEEE, 2014.
- [34] RABINER, L. R., JUANG, B.-H. *Fundamentals of speech recognition*. Tsinghua University Press, 1999.
- [35] ALMEIDA, C. R., OTHERS. “Extratores de características acústicas inspirados no sistema periférico auditivo”, 2014.
- [36] DIXIT, A., VIDWANS, A., SHARMA, P. “Improved MFCC and LPC algorithm for bundelkhandi isolated digit speech recognition”. In: *2016 international conference on electrical, electronics, and optimization techniques (ICEEOT)*, pp. 3755–3759. IEEE, 2016.
- [37] DAVIS, S., MERMELSTEIN, P. “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences”, *IEEE transactions on acoustics, speech, and signal processing*, v. 28, n. 4, pp. 357–366, 1980.
- [38] ZWICKER, E. “Subdivision of the audible frequency range into critical bands (Frequenzgruppen)”, *The Journal of the Acoustical Society of America*, v. 33, n. 2, pp. 248–248, 1961.
- [39] BAGGENSTOSS, P. M., KURTH, F. “Comparing shift-autocorrelation with cepstrum for detection of burst pulses in impulsive noise”, *The Journal of the Acoustical Society of America*, v. 136, n. 4, pp. 1574–1582, 2014.
- [40] TELONI, V., ZIMMER, W., WAHLBERG, M., et al. “127 consistent acoustic size estimation of sperm whales using clicks recorded from unknown aspects”, *J. Cetacean Res. Manage*, v. 9, n. 2, pp. 127–136, 2007.
- [41] PRIOR, M. K., BROWN, D. J. “Estimation of depth and yield of underwater explosions from first and second bubble-oscillation periods”, *IEEE Journal of Oceanic Engineering*, v. 35, n. 1, pp. 103–112, 2010.
- [42] GERHARD, D. B. “Computationally measurable temporal differences between speech and song”, *Burnaby, BC, Canadá*, v. 207, 2003.
- [43] BOGERT, B., OSSANNA, J. “The heuristics of cepstrum analysis of a stationary complex echoed Gaussian signal in stationary Gaussian noise”, *IEEE Transactions on Information Theory*, v. 12, n. 3, pp. 373–380, 1966.

- [44] POLLARA, A., SUTIN, A., SALLOUM, H. “Passive acoustic methods of small boat detection, tracking and classification”. In: *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, pp. 1–6. IEEE, 2017.
- [45] WU, G. “Estimation of range and depth of a submerged moving object by using noise cepstrum”, *Journal of Sound and Vibration*, v. 245, n. 5, pp. 835–843, 2001.
- [46] FERGUSON, B. G., LO, K. W., THURAISSINGHAM, R. A. “Sensor position estimation and source ranging in a shallow water environment”, *IEEE Journal of Oceanic Engineering*, v. 30, n. 2, pp. 327–337, 2005.
- [47] MCKENNA, M. F., ROSS, D., WIGGINS, S. M., et al. “Underwater radiated noise from modern commercial ships”, *The Journal of the Acoustical Society of America*, v. 131, n. 1, pp. 92–103, 2012.
- [48] SURANYI, B., NGUYEN, B. “Cepstral coefficient feature extraction for active sonar classification”. In: *Proceedings of ACOUSTICS*, v. 10, 2019.
- [49] XIE, Y., REN, J., XU, J. “Underwater-art: Expanding information perspectives with text templates for underwater acoustic target recognition”, *The Journal of the Acoustical Society of America*, v. 152, n. 5, pp. 2641–2651, 2022.
- [50] FATHY, A. “College Level Neural Nets [I] - Basic Nets: Math & Practice!” <https://www.udemy.com/course/deep-learning-neural-nets-with-math-derivations-part-1/>. Acessado em 25 de Novembro de 2023.
- [51] LUO, X., CHEN, L., ZHOU, H., et al. “A Survey of Underwater Acoustic Target Recognition Methods Based on Machine Learning”, *Journal of Marine Science and Engineering*, v. 11, n. 2, pp. 384, 2023.
- [52] YUAN, F., KE, X., CHENG, E. “Joint representation and recognition for ship-radiated noise based on multimodal deep learning”, *Journal of Marine Science and Engineering*, v. 7, n. 11, pp. 380, 2019.
- [53] CHOI, J., CHOO, Y., LEE, K. “Acoustic classification of surface and underwater vessels in the ocean using supervised machine learning”, *Sensors*, v. 19, n. 16, 2019.

- [54] SHIU, Y., PALMER, K. J., ROCH, M. A., et al. “Deep neural networks for automated detection of marine mammal species”, *Scientific Reports*, v. 10, n. 1, pp. 1–12, 2020.
- [55] WANG, D., ZHANG, L., LU, Z., et al. “Large-Scale Whale Call Classification Using Deep Convolutional Neural Network Architectures”, *2018 IEEE International Conference on Signal Processing, Communications and Computing, ICSPCC 2018*, 2018.
- [56] WANG, Y., PENG, H. “Underwater acoustic source localization using generalized regression neural network”, *The Journal of the Acoustical Society of America*, v. 143, n. 4, pp. 2321–2331, 2018.
- [57] CHEN, J., LIU, C., XIE, J., et al. “Time–Frequency Mask-Aware Bidirectional LSTM: A Deep Learning Approach for Underwater Acoustic Signal Separation”, *Sensors*, v. 22, n. 15, pp. 5598, 2022.
- [58] WITTEN, I. H., FRANK, E. “Data mining: practical machine learning tools and techniques with Java implementations”, *Acm Sigmod Record*, v. 31, n. 1, pp. 76–77, 2002.
- [59] JÚNIOR, J. M. O., LELES, M. C. R., VALE-CARDOSO, A. S., et al. “Operação de um manipulador por meio da detecção de gestos baseada em aprendizagem de máquina”. In: *Congresso Brasileiro de Automática-CBA*, v. 1, 2019.
- [60] BARBOSA, G. N., BEZERRA, G. M. G., DE MEDEIROS, D. S., et al. “Segurança em Redes 5G: Oportunidades e Desafios em Detecção de Anomalias e Predição de Tráfego baseadas em Aprendizado de Máquina”, *Sociedade Brasileira de Computação*, 2021.
- [61] DE PÁDUA SANTOS, C. I. R. *Classificação de navios aplicando algoritmos de aprendizagem supervisionada (data mining)*. Dissertação para obtenção do grau de mestre em ciências militares navais, na especialidade de marinha, Escola Naval, Marinha Portuguesa, 2015.
- [62] REMIGIO, M. “Aprendizagem Baseada em Instâncias — KNN”. <https://medium.com/@msremigio/aprendizagem-baseada-em-inst%C3%A2ncias-knn-7e2c6f0778bc>. Acessado em 10 de janeiro de 2024.
- [63] BREIMAN, L. “Random forests”, *Machine learning*, v. 45, pp. 5–32, 2001.

- [64] KOEHRSEN, W. “Random Forest: Simple Explanation”. 2 2023. Disponível em: <<https://williamkoehrsen.medium.com/random-forest-simple-explanation-377895a60d2d>>.
- [65] JORDAN, M. I., MITCHELL, T. M. “Machine learning: Trends, perspectives, and prospects”, *Science*, v. 349, n. 6245, pp. 255–260, 2015.
- [66] KOHAVI, R., OTHERS. “A study of cross-validation and bootstrap for accuracy estimation and model selection”. In: *Ijcai*, v. 14, pp. 1137–1145. Montreal, Canada, 1995.
- [67] XIE, Y., REN, J., XU, J. “Unraveling complex data diversity in underwater acoustic target recognition through convolution-based mixture of experts”, *Expert Systems with Applications*, p. 123431, 2024.
- [68] SARAVANAN, R., SUJATHA, P. “A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification”. In: *2018 Second international conference on intelligent computing and control systems (ICICCS)*, pp. 945–949, 2018.
- [69] TZANETAKIS, G., COOK, P. “Musical genre classification of audio signals”, *IEEE Transactions on speech and audio processing*, v. 10, n. 5, pp. 293–302, 2002.

Apêndice

Código em Python para os Extratores de características

L:\1-TESE MESTRADO\EXPERIMENTO2\codigos\extract_Exp2.py

```
1
2 import numpy as np
3 import scipy.fftpack
4 from scipy.fftpack import dct
5 from sklearn.decomposition import IncrementalPCA, PCA
6 import pickle as pkl
7 import librosa
8 import os
9 import matplotlib.pyplot as plt
10 from scipy.signal import lfilter
11
12 plt.close('all')
13
14 def create_custom_mel_filterbank(n_fft, sr, filter_params):
15     """
16     Cria um banco de filtros mel personalizados.
17
18     Args:
19         n_fft (int): Tamanho da Transformada de Fourier Rápida (FFT).
20         sr (int): Taxa de amostragem do sinal de áudio.
21         filter_params (list): Lista de tuplas definindo frequências centrais
22         e larguras de banda para cada filtro mel.
23
24     Returns:
25         numpy.ndarray: Matriz representando o banco de filtros mel.
26     """
27     n_filters = len(filter_params)
28     filterbank = np.zeros((n_filters, int(np.floor(n_fft / 2 + 1))))
29
30     for m, (freq_center, bandwidth) in enumerate(filter_params):
31         # Calcula frequências centrais e larguras de banda em escala mel
32         mel_center = 2595 * np.log10(1 + freq_center / 700)
33         mel_bandwidth = 2595 * np.log10(1 + (freq_center + bandwidth) / 700) - mel_center
34
35         # Converte frequências mel para escala linear
36         hz_center = 700 * (10 ** (mel_center / 2595) - 1)
37         hz_bandwidth = 700 * (10 ** (mel_bandwidth / 2595) - 1)
38
39         # Calcula índices dos bins do filtro na FFT
40         bin_idxs = np.floor((n_fft + 1) * np.array([hz_center - hz_bandwidth / 2, hz_center,
41                                                     hz_center + hz_bandwidth / 2]) / sr)
42         .astype(int)
43
44         # Preenche os pesos do filtro de acordo com a distância dos bins da frequência central
45         for k in range(bin_idxs[0], bin_idxs[2] + 1):
46             if k <= bin_idxs[1]:
47                 filterbank[m, k] = (k - bin_idxs[0]) / (bin_idxs[1] - bin_idxs[0])
48             else:
49                 filterbank[m, k] = (bin_idxs[2] - k) / (bin_idxs[2] - bin_idxs[1])
50
51     return filterbank
52
53 #Definição dos parâmetros dos filtros MEL (Largura de Banda e Frequência central)
54 #especificado na TABELA 3.1
55 filter_params = [
56     (20, 50), # Filtro 1: Frequência central = 20 Hz, Largura de banda = 50 Hz
```

```

56 # ... (outros filtros omitidos )
57 (6500, 1000), # Filtro 42: Frequência central = 6500 Hz, Largura de banda = 1000 Hz
58 ]
59
60
61 ### PROCESSAMENTO DAS PASTAS CONTENDO OS ARQUIVOS .wav #####
62
63 """
64 Um diretório principal é composto por pastas contendo os arquivos de .wav de cada
65 classe (pasta para cada pasta)
66 além disso, esse diretório também contém uma pasta com os códigos e outra pasta com o
67 resultado da simulação ( Matriz de vetores
68 de características para cada classe)
69 O código itera por cada pasta no diretório principal, extraíndo características dos
70 arquivos de áudio (.wav)
71 e salvando os resultados em arquivos pickle.
72 """
73
74 diretorio = "L:/1-TESE MESTRADO/EXPERIMENTO02" # diretório principal onde estão armazenadas
75 as pastas contendo os arquivos de áudio (.wav) das embarcações.
76 pastas = os.listdir(diretorio)
77 pastas = [pasta for pasta in pastas if pasta not in ['simulacao', 'codigos']]
78
79 a = 0 # Iniciar o vetor de rotulagem de classes
80 b = 0
81 intervalo_extracao = 4 # Duração do intervalo para extração das características
82
83
84 for pasta in pastas:
85     print(f"Processando pasta {pasta}")
86     os.chdir(os.path.join(diretorio, pasta)) # Mudar para o diretório da pasta atual
87     arquivos = os.listdir() # Lista todos os arquivos no diretório atual
88     nome_pasta = str(pasta)
89     dados = []
90
91     # Filtrar apenas arquivos .wav
92     for arquivo in arquivos:
93         if arquivo.endswith(".wav") or arquivo.endswith(".WAV"):
94             dados.append(arquivo)
95
96     features_da_class = [] # Lista para armazenar as características de todos os arquivos
97     na pasta
98     nomes_arquivos = []
99     for file in dados:
100         # Carregar o arquivo de áudio
101         X, sr = librosa.load(file, sr=50000, mono=False)
102
103         # Segmentação do arquivo e extração de features (MFCC,SDF,LPC, psc, etc.)
104         n_amostras_intervalo = int(intervalo_extracao * sr) # Tamanho do segmento para
105         análise
106         total_amostras = np.size(X) # Duração do sinal (em amostras)
107         num_segmentos = int(np.floor(total_amostras / n_amostras_intervalo)) \
108             if n_amostras_intervalo <= total_amostras else 1
109
110         features_do_arquivo = []
111         arquivo_nome = os.path.splitext(file)[0] # Nome do arquivo sem extensão
112
113         for j in range(num_segmentos):
114             # Define o intervalo do sinal para análise no segmento atual
115             amostra_inicial = n_amostras_intervalo * j
116             amostra_final = amostra_inicial + n_amostras_intervalo
117             signal = X[amostra_inicial:amostra_final]

```

```

111
112     # Extração dos vetores de características (LPC,LPCC, MFCC tradicional,PSC )
113
114     ## lpc e lpcc
115     lpc = librosa.lpc(signal, order=12, axis=0) # Coeficientes LPC
116     def lpc_to_lpcc(lpc, order_lpcc):
117         return -lfilter([1] + lpc[1:], [1], np.arange(1, order_lpcc + 1))
118     lpcc = lpc_to_lpcc(lpc, 12) # LPCCs
119     ### MFCCs tradicionais
120     mfccs = librosa.feature.mfcc(y=signal, sr=sr, n_mfcc=42).T # extração do MFCCs
convencionais (42 filtros)
121     mean_mfccs = np.mean(mfccs, axis=0) # Média dos MFCCs
122     ####PSC
123     stft = np.abs(np.fft.fft(signal)) # STFT
124     power_spectrum = np.abs(stft) ** 2 # espectro de potência
125     PSC= np.mean(power_spectrum, axis=0)
126     #zcr
127     zcr = librosa.feature.zero_crossing_rate(signal) # Taxa de cruzamento zero
128     zcr = zcr.reshape(-1) # Reshape
129
130
131     # Pré-ênfase para realçar altas frequências
132     pre_emphasis = 0.97
133     emphasized_signal = np.append(signal[0], signal[1:] - pre_emphasis * signal[:-1])
134
135     # Enquadramento do sinal em frames com salto (overlap)
136     frame_size = 40e-3 # Tamanho do frame (duração em segundos)
137     hop = (1 / 2) * (40e-3) # Salto entre frames (overlap 50%)
138     frame_length, hop_length = frame_size * sr, hop * sr
139     signal_length = len(emphasized_signal)
140     frame_length = int(round(frame_length))
141     hop_length = int(round(hop_length))
142     num_frames = int(np.ceil(float(np.abs(signal_length - frame_length)) / hop_length))
143     pad_signal_length = num_frames * hop_length + frame_length
144     z = np.zeros((pad_signal_length - signal_length))
145     pad_signal = np.append(emphasized_signal, z) # Acoplamento de zeros para frames
completos
146
147     # Criação de índices para extração de frames
148     indices = np.tile(np.arange(0, frame_length), (num_frames, 1)) + np.tile(
149         np.arange(0, num_frames * hop_length, hop_length), (frame_length, 1)).T
150     frames = pad_signal[indices.astype(np.int32, copy=False)]
151
152     # Janelamento (função Hanning) para reduzir vazamento espectral
153     frames *= np.hamming(frame_length)
154
155     # Realiza a Transformada de Fourier Rápida (FFT)
156     NFFT = frame_length
157     mag_frames = np.abs(np.fft.rfft(frames, NFFT)) # Magnitude do espectro
158     pow_frames = ((1.0 / NFFT) * ((mag_frames) ** 2)) # Espectro de potência
159
160     # Filtros Mel
161
162     ### BAnco de filtros Mel projetado ####
163     filterbank = create_custom_mel_filterbank(NFFT, sr, filter_params) # Criação do
banco de filtros Mel
164
165     # Aplicação dos filtros Mel
166     filter_banks = np.dot(pow_frames, filterbank.T)
167     filter_banks = np.where(filter_banks == 0, np.info(float).eps, filter_banks) #
Estabilidade numérica

```

```

168     filter_output = filter_banks # Saída do filtro
169
170     # Conversão do espectro de potência em dB
171     filter_banks = 20 * np.log10(filter_banks) # dB (vrf)
172
173     # Cálculo de MFCCs utilizando o banco de filtros projetado
174     num_ceps = len(filter_params) # Número de coeficientes MFCCs
175     mfcc = dct(filter_banks, type=2, axis=1, norm='ortho')[:, 0 : (num_ceps)] # MFCCs
176     MFCC = np.mean(mfcc, axis=0) # vetor MFCC para o banco de filtros projetado
177
178
179     #####
180     ##### Modificação para extrair features dinâmicas #####
181     #####
182
183     sdf = np.zeros((np.size(filter_output, axis=1), np.size(filter_output, axis=0)))
184     for ii in range(np.size(filter_output, axis=1)):
185         y = filter_output[:, ii] # Coluna do espectro de potência
186         fft = np.abs(scipy.fftpack.fft(y)) # FFT da coluna
187         pow_y = ((1.0 / len(y)) * ((fft) ** 2)) # Espectro de potência da coluna
188         sdf[ii, :] = fft
189
190     sdf = 20 * np.log10(sdf) # Conversão para dB
191     n_p = 4 # Número de pontos para DCT
192     sdf = dct(sdf, type=2, axis=1, norm='ortho')[:, 0 : (n_p)] # DCT do espectro
dinâmico
193
194     n_point=4
195     filter_output=dct(filter_output, type=2, axis=1, norm='ortho')[:, 1 : (n_point + 1)]
196
197     SDF=sdf.flatten() # vetor de recursos espectrais dinâmicos(Vetor SDF)
198
199     # Combinação dos vetores de características
200     """
201     A próxima linha define qual vetor de características ou combinação deles será
extraído e colocados
202     em uma matriz que posteriormente será submetida ao classificador
203     """
204     ext_features = np.hstack([MFCC, SDF]) # Vetor de caracterísca que será submetido ao
classificador
205
206     features_do_arquivo.append(ext_features)
207
208     features_do_arquivo = np.c_[features_do_arquivo] # Concatenação dos vetores de
características de todos os segmentos
209     new_column = np.array([arquivo_nome] * features_do_arquivo.shape[0]) # Criar um array
NumPy com a nova string repetida para cada linha
210     features_do_arquivo = np.concatenate((features_do_arquivo, new_column[:, np.newaxis]),
axis=1) # Adiciona nome do arquivo como última coluna
211     features_da_class.append(features_do_arquivo)
212
213     features_da_class = np.vstack(features_da_class) # Concatena os vetores de
características de todos arquivos da pasta
214
215     # Gera o vetor de classe
216     size = features_da_class.shape[0]
217     classe = np.zeros(size)
218     for i in np.arange(size):
219         classe[i] = a # Define a classe (valor atribuído manualmente)
220     a = a + 1 # Incrementa o valor da classe para a próxima pasta
221
222     # Concatena features e classe, inverte a ordem das duas últimas colunas

```

```
223 resultado = np.c_[features_da_class, classe]
224 resultado[:, -2], resultado[:, -1] = resultado[:, -1], resultado[:, -2].copy()
225
226 # Salva os dados em um arquivo pickle
227 os.chdir("L:\\1-TESE MESTRADO\\EXPERIMENTO2\\simulacao")
228 with open(str(pasta) + '.pkl', 'wb') as i:
229     pickle.dump(resultado, i)
230
231
```