

UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE ENGENHARIA ELÉTRICA, MECÂNICA E DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E DE
COMPUTAÇÃO

**RECONHECIMENTO AUTOMÁTICO DE AVES DE NOMES
ONOMATOPÉICOS UTILIZANDO ÁRVORE DE DECISÃO**

CÉLIO SEIXO DE BRITO JUNIOR

ORIENTADOR: PROF. DR. PAULO CÉSAR MIRANDA MACHADO

GOIÂNIA
2013

**Dados Internacionais de Catalogação na Publicação na (CIP)
GPT/BC/UFG**

862r Brito Junior , Célio Seixo de.
Reconhecimento Automático de Aves de Nomes
Onomatopeicos Utilizando Árvore de Decisão
[manuscrito] / Célio Seixo de Brito Junior. - 2013.
130 f. : il.

Orientador: Prof. Dr. Paulo César Miranda Machado
Dissertação (Mestrado) – Universidade Federal de Goiás,
Escola de Engenharia Elétrica, Mecânica e de Computação,
2013.

Bibliografia.

Inclui lista de figuras, abreviaturas e tabelas.

Apêndices.

1. Aves – Classificação 2. Aves – Arvore de decisão I.
Título.

CDU: 636.5

RECONHECIMENTO AUTOMÁTICO DE AVES DE NOMES ONOMATOPÉICOS UTILIZANDO ÁRVORE DE DECISÃO

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Computação da Escola de Engenharia Elétrica, Mecânica e de Computação da Universidade Federal de Goiás para obtenção do grau de Mestre em Engenharia Elétrica.

Área de Concentração: Engenharia Elétrica.
Orientador: Prof. Dr. Paulo César Miranda Machado

GOIÂNIA

2013

CÉLIO SEIXO DE BRITO JUNIOR

**RECONHECIMENTO AUTOMÁTICO DE AVES DE NOMES
ONOMATOPÉICOS UTILIZANDO ÁRVORE DE DECISÃO**

Dissertação definida no programa de Pós-Graduação da Engenharia Elétrica, Mecânica e de Computação da Universidade Federal de Goiás como requisito parcial para obtenção de título de Mestre no Programa de Mestrado em Engenharia Elétrica e de computação, aprovada em 30 de Abril de 2013, pela Banca Examinadora constituída pelos professores:

Paulo César Miranda Machado, Ph.D

Escola de Engenharia, Mecânica e de Computação – UFG
Presidente da Banca.

Anamaria Achtschin Ferreira, Dr^a.

Ciências Biológicas - UEG

Rodrigo Pinto Lemos, Dr.

Escola de Engenharia, Mecânica e de Computação – UFG

Flávio Henrique Teles Vieira, Dr.

Escola de Engenharia, Mecânica e de Computação – UFG

AGRADECIMENTOS

Durantes esses dois anos em que foi desenvolvido o meu trabalho de mestrado, recebi ajuda de muitas pessoas. Alguns me auxiliaram academicamente e outros me dando estrutura física e moral. Gostaria de agradecer a cada um deles que de uma forma ou de outra me auxiliou durante o desenvolvimento deste trabalho, e de antemão peço desculpas caso eu me esqueça de citar o nome de alguém.

Primeiramente a Deus, pela vida e sabedoria que me foi concedida.

Agradeço à minha mãe, Maria do Socorro Carvalho Brito, por todo o cuidado, carinho, dedicação e principalmente o amor que tens por mim. Ela é responsável por toda a estrutura física e moral que hoje possuo.

Agradeço a minha irmã Célia Cristyna Carvalho Brito por ser simplesmente a melhor irmã do mundo.

Agradeço a minha namorada, Laylla Silva Ramalho, pelo apoio, dedicação, cuidado e amor. Por acreditar em mim e no meu trabalho, e além de tudo, me fazer muito feliz.

Ao meu orientador prof. Dr. Paulo César Miranda Machado, pela paciência ao me ensinar, dedicação ao meu trabalho, e por confiar a mim este projeto. Além de tudo agradeço por me ensinar a ser um pesquisador e amar essa profissão. Ainda tenho muito a aprender.

Ao prof. Dr. Rodrigo Pinto Lemos pela paciência ao me ensinar e retirar dúvidas referentes ao projeto, pela dedicação que teve com o meu trabalho e por todas as contribuições no meu amadurecimento enquanto pesquisador.

Aos meus colegas do grupo que de uma forma ou outra contribuíram no desenvolvimento do meu trabalho. Em especial ao Paulo Francisco da Conceição pela ajuda com os programas de pré-processamento e pelo desenvolvimento de toda a base inicial deste projeto.

E aos meus grandes amigos Didi (vulgo André), Marcão, Daniel, Ícaro e Rodrigo, pelo companheirismo e pela paciência que me auxiliou muito para concluir esse trabalho.

E a todos os meus amigos e familiares.

E dedico este trabalho a memória de meu pai, Célio Seixo de Brito, que faleceu durante o desenvolvimento deste trabalho e não pode presenciar a conclusão de tal conquista em minha vida.

RESUMO

O presente trabalho propõe uma metodologia de determinação automática de características, classificação e reconhecimento de aves de nomes onomatopéicos baseada na análise dos espectrogramas das vocalizações dessas aves. A filtragem e a segmentação dos sinais são feitas automaticamente. A segmentação, que é a separação da vocalização da ave em regiões de som e silêncio, é realizada com as características centroide e energia dos sinais. A filtragem é feita com filtros gerados utilizando transformada wavelet e filtros Chebyshev. O reconhecimento das aves é feito a partir de uma árvore de decisão, construída automaticamente com base nas seguintes características: entropia, ZCR, centróide, energia, fluxo espectral, frequência rolloff, frequência mínima, frequência máxima e frequência de maior intensidade do espectro.

Palavras-chaves: Reconhecimento automático, Processamento de sinais, Transformada wavelet, Árvore de decisão.

ABSTRACT

This work proposes a methodology for automatically determining characteristics, classification and recognition of birds from onomatopoeic names based on analysis of spectrograms of vocalizations of these birds. The filtering and segmentation of the signals are made automatically. Segmentation, which is the separation of the bird vocalization in regions of sound and silence, is performed with the centroid and energy characteristics of the signals. The filtering is made using filters generated using wavelet transform and Chebyshev filters. The recognition of birds is made from a decision tree, constructed automatically based on the following features: entropy, ZCR, centroid, energy, spectral flux, rolloff frequency, minimum frequency, maximum frequency and the frequency of greater intensity of the spectrum.

Keywords: Automatic recognition, Signal processing, Wavelet transform, Decision tree.

SUMÁRIO

LISTA DE FIGUAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

1	INTRODUÇÃO	18
2	AS AVES	22
2.1	A IMPORTÂNCIA DAS AVES NO MEIO AMBIENTE.....	22
2.2	A VOCALIZAÇÃO DAS AVES COMO MEIO DE RECONHECIMENTO DAS ESPÉCIES.....	24
2.3	ESTRUTURA BIOLÓGICA DE PRODUÇÃO DOS SONS DAS AVES	25
2.4	A FUNÇÃO DO CANTO DAS AVES.....	27
2.5	AVES ESTUDADAS NESTE TRABALHO.....	28
3	METODOLOGIA	37
3.1	DOMÍNIO DO TEMPO.....	39
3.2	DOMÍNIO DA FREQUÊNCIA.....	40
3.3	PADRONIZAÇÃO DOS SINAIS.....	42
3.4	ARQUIVOS WAV	42
3.5	NORMALIZAÇÃO E TAXA DE AMOSTRAGEM	43
3.6	FILTRAGEM.....	44
3.6.1	WAVELETS	45
3.6.2	WAVELET MÃE.....	47

3.6.3	TRANSFORMADA CONTÍNUA <i>WAVELET</i> (CWT).....	48
3.6.4	TRANSFORMADA <i>WAVELET</i> DISCRETA (DWT)	49
3.6.5	BANCO DE FILTROS E TRANSFORMADA <i>WAVELET</i>	50
3.6.6	REMOÇÃO DE RUÍDOS COM WAVELETS (DENOISING)	54
3.6.7	A ESCOLHA DE LIMIAR PARA A REMOÇÃO DE RUÍDOS.....	58
3.7	CARACTERÍSTICAS PARA O RECONHECIMENTO	59
3.8	PROCESSAMENTO PARA EXTRAÇÃO DE CARACTERÍSTICAS DE	
	ÁUDIO.....	61
3.8.1	CARACTERÍSTICAS DO DOMÍNIO DO TEMPO	63
3.8.2	CARACTERÍSTICAS DO DOMÍNIO DA FREQUÊNCIA	67
3.9	SEGMENTAÇÃO	69
3.10	MÉTODO DE RECONHECIMENTO	75
4	RESULTADOS E DISCUSSÃO	79
4.1	PRÉ-PROCESSAMENTO.....	79
4.2	RESULTADOS DA FILTRAGEM.....	80
4.3	RESULTADOS DA SEGMENTAÇÃO	83
4.4	DETECÇÃO AUTOMÁTICA DE FREQUÊNCIA	86
4.5	ANÁLISE DAS CARACTERÍSTICAS UTILIZADAS	89
4.6	FASE DE TREINAMENTO	98
4.7	RECONHECIMENTO	115
5	CONCLUSÃO E TRABALHOS FUTUROS	118

6	BIBLIOGRAFIA	120
7	ANEXOS	125

LISTA DE FIGURAS

Figura 1 - Espectrograma do som de um Fogo – apagou.	25
Figura 2 - Estrutura do sistema de produção de vocalização das aves (SELIN, 2007)..	26
Figura 3 - Diagrama de blocos do sistema desenvolvido, mostrando a sequência adotada para o reconhecimento de espécies de aves. Adaptado de (SELIN, 2007).	38
Figura 4- Sinal no domínio do tempo, canto de um fogo-apagou.	40
Figura 5 - Sinal no domínio da frequência, canto de um Fogo-apagou.	42
Figura 6 - Forma de onda de uma senóide e de uma Wavelet (MISITI et al. , 2006). ...	45
Figura 7 – Função Haar. Fonte: Matlab R2006a.	47
Figura 8 – Funções Daubechies. Fonte: Matlab R2006a.....	48
Figura 9 – Processo de Filtragem do sinal (MISITI et al., 2006).	51
Figura 10 – Processo de decomposição do sinal. Adaptado de (MISITI et al., 2006). ..	51
Figura 11 – Processo de diminuição do número de amostras utilizando o operador downsampling . Adaptado de (MISITI et al., 2006).....	52
Figura 12 – Técnica de decomposição AMR.	53
Figura 13 – Árvore gráfica de uma AMR (MISITI et al., 2006).	53
Figura 14 - O Processo da esquerda ilustra uma decomposição (análise) e da direita uma reconstrução (síntese). Adaptado de (MISITI et al., 2006).	54
Figura 15 – processo de remoção de ruídos denoising.	55
Figura 16 – (a) Sinal original, (b) hard-thresholding e (c) soft-thresholding.	57
Figura 17 – Sinal com ruído e sinal após o processo denoising (Matlab, 2006).	58
Figura 18 – Processo de janelamento para um sinal de áudio (GIANNAKOPOULOS, 2009).	63
Figura 19 - Energia do som de um Fogo-apagou.	65
Figura 20 – Energia do som de um Peixe-frito.....	65

Figura 21– Centróide do som de um Fogo-apagou.	68
Figura 22– Centróide do som de um Peixe-frito.	68
Figura 23 – Segmetação do som de um Peixe-frito, sem pré-filtragem.	70
Figura 24 - Segmentação do som filtrado de um Peixe-frito.....	70
Figura 25 - Segmentação do som de um Peixe-frito utilizando centróide e energia.	71
Figura 26 – Segmentação do som filtrado de um Peixe-frito utilizando centróide e energia.	72
Figura 27 – Segmentação com janela e passo de 20 ms.....	74
Figura 28 – Segmentação do canto de um Fogo-apagou. Energia do sinal, Centroides do sinal e sinal.	74
Figura 29 – Modelo Modelo de Árvore de Decisão (GAMA, 2004).	77
Figura 30 – Sinal padronizado do som de um Fogo-apagou.	80
Figura 31 – Sinal e espectrograma do som de um Fogo-apagou.....	81
Figura 32 – Sinal e espectrograma do som de um Fogo-apagou após a aplicação do método (a).....	82
Figura 33 – Sinal e espectrograma do som de um Fogo-apagou após a aplicação do método (b).	82
Figura 34 – Detecção visual da duração de canto de um Fogo-apagou.	83
Figura 35 – Detecção automática da duração de canto de um Fogo-apagou.	84
Figura 36 – Energia do sinal, Centroides do sinal e sinal do canto de um Fogo-apagou.	85
Figura 37 - Sinal segmentado do canto de um Fogo-apagou.	86
Figura 38 – Centroides, sinal e espectrograma do canto de um Jaó.	91
Figura 39 - Centroides, sinal e espectrograma do canto de um Teque-teque.	91
Figura 40 - Energia, sinal e espectrograma do som de um Peixe-frito.....	92
Figura 41 - energia, sinal e espectrograma do som de um Acauã.	92

Figura 42 - Entropia, sinal e espectrograma do som de um Peixe-frito.	93
Figura 43 - Entropia, sinal e espectrograma do som de um Fogo-apagou.	93
Figura 44 – Fluxo Espectral, sinal e espectrograma do som de um Jaó.	94
Figura 45 – Fluxo Espectral, sinal e espectrograma do som de um Tiziu.	95
Figura 46 – Frequência <i>Rolloff</i> , sinal e espectrograma do som de um Peixe-frito.	95
Figura 47- Frequência <i>Rolloff</i> , sinal e espectrograma do som de um Teque-teque.	96
Figura 48 - ZCR, sinal e espectrograma do som de um Peixe-frito.	97
Figura 49 - ZCR, sinal e espectrograma do som de um Teque-teque.	97
Figura 50 – Regiões de Entropia dos sons das doze espécies de aves em estudo.	100
Figura 51 – Regiões de Centróide dos sons das doze espécies de aves em estudo.	100
Figura 52 – Regiões de ZCR dos sons das doze espécies de aves em estudo.	101
Figura 53 – Regiões de frequência <i>Rolloff</i> dos sons das doze espécies de aves em estudo.	101
Figura 54 – Regiões de Fluxo Espectral dos sons das doze espécies de aves em estudo.	102
Figura 55 – Regiões de Energia dos sons das doze espécies de aves em estudo.	102
Figura 56 – Regiões de Frequência Mínima dos sons das doze espécies de aves em estudo.	103
Figura 57 – Regiões de Frequência Máxima dos sons das doze espécies de aves em estudo.	103
Figura 58 – Regiões de Frequência mais Intensa dos sons das doze espécies de aves em estudo.	104
Figura 59 - Regiões de ZCR dos sons das aves na segunda etapa.	105
Figura 60 – Regiões de Frequência mais Intensa dos sons das aves na segunda etapa.	105
Figura 61 – Regiões de Frequência máxima dos sons das aves na segunda etapa.	106

Figura 62 – Regiões de ZCR dos sons das aves na terceira etapa.	106
Figura 63 – Regiões de Fluxo espectral dos sons das aves na terceira etapa.	107
Figura 64 – Regiões de Entropia dos sons das aves na quarta etapa.	108
Figura 65 – Regiões de Centróide dos sons do Bem-te-vi e do Neinei.	109
Figura 66 – Regiões de ZCR dos sons do Bem-te-vi e do Neinei.	109
Figura 67 – Regiões de Frequência máxima dos sons do Bem-te-vi e do Neinei.	110
Figura 68 – Regiões de Frequência mais intensa dos sons do Bem-te-vi e do Neinei.	110
Figura 69 - Árvore de decisão gerada automaticamente pelo método proposto.	114
Figura 70 – Sequência para o reconhecimento de um sinal de entrada.	115

LISTA DE TABELAS

Tabela 1– <i>Espécies Utilizadas neste trabalho.</i>	29
Tabela 2- <i>ZCR, Frequência principal (FP), mínima (Fmím), máxima (Fmáx) e mais intensa (FI).</i>	66
Tabela 3– <i>Distribuição geográfica das aves estudadas.</i>	79
Tabela 4– <i>Comparação entre inspeção visual e detecção automática</i>	84
Tabela 5 – <i>Tabela com os espectrogramas das aves.</i>	86
Tabela 6– <i>Faixa de frequência e frequência mais intensa de todas as espécies.</i>	89
Tabela 7– <i>Limites das características.</i>	98
Tabela 8– <i>Limites das características</i>	98
Tabela 9– <i>Limites das características.</i>	99
Tabela 10- <i>Intervalos e valores de PNS para as características do Bem-te-vi e Neinei.</i>	111
Tabela 11– <i>Treinamento: características para o reconhecimento.</i>	111
Tabela 12– <i>Tabela de treinamento das espécies estudadas.</i>	113
Tabela 13– <i>Resultados obtidos na classificação das espécies.</i>	116
Tabela 14– <i>Matriz de confusão dos sinais testados</i>	117
Tabela 15 – <i>Características extraídas dos sons do Acauã utilizados na fase de treinamento.</i>	125
Tabela 16 - <i>Características extraídas dos sons do Bem-te-vi utilizados na fase de treinamento.</i>	125
Tabela 17 - <i>Características extraídas dos sons do Fim-fim utilizados na fase de treinamento.</i>	126
Tabela 18 - <i>Características extraídas dos sons do Fogo-apagou utilizados na fase de treinamento.</i>	126

Tabela 19 - <i>Características extraídas dos sons do Irerê utilizados na fase de</i> <i>treinamento.</i>	127
Tabela 20 - <i>Características extraídas dos sons do Jaó utilizados na fase de treinamento.</i>	127
Tabela 21 - <i>Características extraídas dos sons do Neinei utilizados na fase de</i> <i>treinamento.</i>	128
Tabela 22 - <i>Características extraídas dos sons do Peixe-frito utilizados na fase de</i> <i>treinamento.</i>	128
Tabela 23 - <i>Características extraídas dos sons do Pitiguari utilizados na fase de</i> <i>treinamento.</i>	129
Tabela 24 - <i>Características extraídas dos sons do Saracura utilizados na fase de</i> <i>treinamento.</i>	129
Tabela 25 - <i>Características extraídas dos sons do Teque-teque utilizados na fase de</i> <i>treinamento.</i>	130
Tabela 26 - <i>Características extraídas dos sons do Tiziu utilizados na fase de</i> <i>treinamento.</i>	130

LISTA DE ABREVIATURAS

A	Sinal de Aproximação
AMR	Análise de Multiressolução
CART	Classification and Regession Trees
Cent	Centroide
CWT	(Transformada contínua <i>wavelet</i>)
D	Sinal de Detalhe
dB	Decibel
DSP	<i>Digital Signal Processing</i> (Processamento Digital de Sinal) <i>Discret-Time Fourier Transform</i> (Transformada de Fourier para
DTFT	Sistemas Discretos)
DWT	<i>Discret Wavelet Transform</i> (Transformada <i>Wavelet</i> Discreta)
EE	Entropia
FE	Fluxo Espectral
FI	Frequência mais Intensa
Fmax	Frequência Máxima
Fmin	Frequência Mínima
FFT	<i>Fast Fourier Transform</i> (Transformada Rápida de Fourier) <i>Inverse Discret Wavelet Transform</i> (Transformada Inversa <i>Wavelet</i>
IDWT	Discreta)
<i>mp3</i>	MPEG 1 Layer-3
MATLAB	<i>MaTrix Laboratory</i>
PNS	Porcentagem de não Superposição
PSD	<i>Power Spectral Density</i> (Densidade Espectral de Potência)
RNA	Redes Neurais Artificiais
s	Segundos
TF	<i>Fourier Transform</i> (Transformada de Fourier)
<i>wav</i>	WAVE Form audio format
ZCR	(Taxa de Cruzamento em Zero)

CAPÍTULO 1

1 INTRODUÇÃO

A presente dissertação insere-se na área de reconhecimento de padrões que é a área de pesquisa que tem por objetivo a classificação de objetos (padrões) em diversas categorias ou classes, com aplicações nas mais diversas áreas, tais como: reconhecimento de fala, de escrita e de faces, identificação de retina, análise de eletrocardiogramas, radiografias, tomografias, etc. (DUDA, 2001), (PATRICK, 1972) e (THEODORIDIS; KOUTROUMBAS, 2009).

Na literatura, as técnicas de reconhecimento de padrões têm sido utilizadas para a detecção de diferentes doenças em diagnósticos médicos (SIQUEIRA, 2010), (ARGOUD, 2001) e (GARCIA, 2001) e em desenvolvimento de dispositivos de segurança que utilizam da fala, da escrita ou da retina para liberação de acesso (NOGUEIRA *et al.*, 2006). Na indústria, retirando o homem da tomada de decisão, deixando para as máquinas essa escolha. E, principalmente, no reconhecimento de aves, utilizando como característica para reconhecimento sua forma (cor, tamanho ou plumagem) (NOBRE, 2012) ou os sons emitidos por eles, como os cantos ou os chamados dos pássaros (CONCEIÇÃO, 2012).

A avifauna é fundamental para o estudo da qualidade dos ambientes, pois as aves são excelentes indicadores dessa qualidade (SAVE, 2010). Assim, o monitoramento de aves é uma atividade que pode auxiliar no controle das espécies e sua conservação. Uma das formas mais prática para realizar a identificação de uma ave é através do som emitido por esta, já que de modo geral, cada espécie de ave produz um som diferente. De acordo com Sick (1985), um das melhores maneiras para o monitoramento de uma espécie é através de seu som, pois em uma mata densa muitas vezes não é possível visualizar a ave. Então, devido a uma série de fatores, a vocalização é uma das melhores características para ser utilizada no reconhecimento de aves. Alguns desses fatores são (SICK, 1985):

- Em ambientes fechados como no interior das florestas, a vocalização assume maior importância, ante a dificuldade de visualização entre os indivíduos de uma mesma espécie;

- No campo, apesar da importância da vocalização ser menor, ela não deixa de ter certo valor porque nas áreas abertas com gramíneas as aves também ficam de certa forma "cobertas" pela ramagem rasteira e as árvores de grande porte são poucas, daí a necessidade do som para a comunicação;
- Por fim temos aves que possuem hábitos noturnos. Para essas aves fica muito mais difícil um estudo de reconhecimento se não for através da vocalização, já que as mesmas são mais difíceis de serem visualizadas.

Recentemente tem aumentado o número de pesquisas com o intuito da classificação de aves:

Nelson (1989) estudou as características que poderiam ser úteis para reconhecimento de aves das espécies *pusilla* e *passerina*, do gênero *Spizella*.

Anderson *et al.* (1996) utilizaram modelos ocultos de Markov para desenvolver um reconhecimento automático das aves *Taeniopygia guttata* e *Passerina cyanea*.

McIlraith *et al.* (1997) realizaram o reconhecimento de aves utilizando redes neurais e como característica de reconhecimento parâmetros espectrais e temporais da vocalização das aves.

Härmä (2003) e Härmä *et al* (2004) propuseram um reconhecimento utilizando as sílabas das vocalizações das aves e de suas harmônicas.

Selin (2007) também utilizou redes neurais para realizar o reconhecimento, mas a inovação foi a utilização de *wavelets* para gerar as características de reconhecimento.

Nobre (2012) mostra em seu trabalho técnicas para extrair características de imagens para serem utilizadas no reconhecimento de aves a partir de imagens digitais.

Conceição (2012), em sua pesquisa, sugere um reconhecimento automático de 12 espécies de aves da família Tinamidae utilizando distância euclidiana e, como características de reconhecimento, o tempo e a faixas de frequência do canto da ave.

Esses trabalhos mostram que técnicas de classificação e reconhecimento das aves podem contribuir como indicado em Correio (2010) para a possibilidade de catalogação e até mesmo a descoberta de novas espécies, tendo em vista que com análise da vocalização é possível identificar características únicas para cada espécie. Com a catalogação e monitoração de espécies, é possível contribuir com dados

estatísticos para auxiliar na tomada de decisão quanto a conservação e preservação das mesmas.

Assim, nesta dissertação, o interesse principal está na utilização de técnicas de reconhecimento de padrões para a classificação de aves utilizando a sua vocalização para o reconhecimento. Para tal, utilizam-se técnicas de pré-processamento para trabalhar e extrair características das vocalizações das aves. Foram tratadas aqui questões fundamentais que podem auxiliar biólogos a catalogar espécies de aves de uma determinada região com o mínimo esforço possível. Com o aprimoramento da metodologia, bastará o pesquisador deixar o gravador ligado e será emitida uma lista com todas as aves cujo canto fora gravado.

Normalmente pesquisadores menos experientes não possuem a capacidade de reconhecer as aves somente pelo canto (que é definido como “reconhecimento de ouvido”), prática que somente é obtida com o tempo e muita dedicação no trabalho (ANDRADE, 1995). A motivação deste trabalho é auxiliar a esses observadores e os pesquisadores que não têm prática, em um momento inicial, e os que têm prática, diminuindo o esforço de análise dos sons coletados.

As aves escolhidas para objeto de estudo deste trabalho são as de nomes populares onomatopéicos, por possuírem um canto bem característico. Essas aves estão distribuídas em várias regiões do nosso país e até mesmo em outros países (SICK, 1985), sendo as técnicas aqui desenvolvidas aplicáveis a qualquer espécie, deste que sejam ajustadas para as novas espécies a serem analisadas. A vocalização foi escolhida como meio de reconhecimento por ser o meio mais evidente de reconhecimento dessas aves, já que muitas vezes o pesquisador em campo não consegue visualizar as aves, mas somente escutá-las (SICK, 1985).

Os objetivos principais foram:

- (i) Realizar todo o pré-processamento de maneira automática, normalizando, filtrando, segmentando e retirando os parâmetros para o reconhecimento;
- (ii) O desenvolvimento de uma rotina para o reconhecimento das espécies escolhidas.

O primeiro objetivo visou o tratamento do som, para que assim fosse possível a obtenção dos parâmetros para o reconhecimento e o segundo objetivo foi a utilização desses parâmetros para o reconhecimento, mostrando que eles podem ser utilizados para esse fim.

A presente dissertação apresenta a seguinte estrutura:

O Capítulo 2 apresenta a importância das aves e um breve estudo de suas vocalizações. O Capítulo 3 descreve as técnicas utilizadas no pré-processamento dos sinais (normalização, filtragem e segmentação dos sons), a extração dos parâmetros de reconhecimento e o método desenvolvido para, utilizando-se desses parâmetros, realizar o reconhecimento. O Capítulo 4 apresenta os resultados obtidos no trabalho, as discussões desses resultados e perspectivas para outras pesquisas. Finalmente, o Capítulo 5 apresenta as conclusões.

CAPÍTULO 2

2 AS AVES

As aves sempre fascinaram o homem, devido às suas colorações e principalmente devido ao seu canto (MARQUES, 2009). O canto das aves tem sido uma importante fonte de inspiração para muitos compositores e escritores.

Nos dias atuais tem crescido muito o interesse das pessoas pela observação das aves (*birdwatching*). No Brasil essa prática vem gerando muitos investimentos para o ecoturismo. Esses observadores são capazes de reconhecer, pelo menos, algumas espécies mais comuns pelo seu som, enquanto que um observador experiente pode reconhecer centenas de espécies só pelo seu som.

2.1 A IMPORTÂNCIA DAS AVES NO MEIO AMBIENTE

Ao todo no mundo são catalogadas aproximadamente 10000 espécies de aves que compõem um vasto universo de formas, cores e cantos (BIRDLIFE, 2011). Cerca de 18,5% dessas espécies de aves ocorrem no Brasil, entre espécies residentes e visitantes, correspondendo a mais da metade das espécies de aves registradas na América do Sul (ANDRADE, 1995). Deste total, 145 são espécies migratórias que passam anualmente no país, sendo que 89 são visitantes do Hemisfério Norte e 56 do Hemisfério Sul (SICK, 1985). O mesmo autor afirma que do total de espécies da América do Sul, 117 são endêmicas do Brasil e 115 estão ameaçadas de extinção, segundo a lista oficial do IBAMA, publicada em 1989.

No meio ambiente as aves são excelentes indicadores da qualidade de nosso ambiente (SAVE, 2012), pois são detectores de mudanças na saúde e condições do ecossistema. Uma situação desfavorável ou não para as aves normalmente reflete a mesma situação para os outros animais e plantas, inclusive para o homem.

As aves são, portanto, essenciais para compreender os desafios ambientais e auxiliar nas mudanças de hábitos necessárias para a continuidade da vida.

As aves são excelentes indicadores de qualidade ambiental (SAVE, 2012), pois são:

- Grupo bastante estudado pela ciência;
- Amplamente presentes: cerca de 10000 espécies de aves no mundo e 1800 espécies no Brasil;
- Facilmente amostradas - por sua voz, cores, forma - através de técnicas simples e de baixo custo;
- Podem ser monitoradas ao longo do tempo, permitindo uma melhor compreensão sobre as transformações ambientais.

Através da observação da avifauna de um local (espécies presentes ou ausentes e sua abundância) é possível conhecer suas características, seu estado de conservação e a qualidade dos serviços ambientais prestados, tais como qualidade do ar, regulação do clima e qualidade da água. Ou seja, de um modo, geral estudar o comportamento e os tipos de aves presentes em algumas regiões é fundamental para compreender o desenvolvimento do meio ambiente daquele local (MORRISON, 1986; ALMEIDA; ALMEIDA, 1998; ALMEIDA, 1981; ALEGRINI, 1997).

2.2 A VOCALIZAÇÃO DAS AVES COMO MEIO DE RECONHECIMENTO DAS ESPÉCIES

Existem várias maneiras de as aves se comunicarem e o som é apenas uma delas. Por exemplo, as sinalizações visuais também são importantes para as aves, mas têm várias restrições. A luz fraca ou o escuro faz da sinalização visual um meio difícil de comunicação e, além disso, a informação visual precisa de proximidade dos parceiros. Portanto o som é um método melhor para a comunicação, porque com o som é possível transmitir rapidamente e eficientemente grandes quantidades de informação a distâncias maiores.

Som é uma maneira de se comunicar com outros indivíduos da mesma espécie e também entre espécies diferentes. Os sons são produzidos apenas quando necessário, e assim, todos os sons têm algum significado.

As aves apresentam vários tipos de sons, e dentre eles temos os cantos (que são sons mais complexos) e as chamadas (que são mais simples e com finalidades mais objetivas dentro dessa comunicação animal) (KREBS; KROODSMA, 1980).

Existem vários tipos de chamadas como, por exemplo, chamadas de acasalamento, de alarme ou fuga. Assim, cada tipo de chamada pode ser associado com um contexto específico.

O canto é uma série de notas, geralmente de mais de um tipo, emitidas em sucessão e relacionadas entre si formando uma sequência de sons bem reconhecidos. Tem características complexas de ritmos e de modulação, estando basicamente sob controle dos hormônios sexuais. Relacionam-se com a época da reprodução, bem como para estabelecimento e defesa do território e ainda com a manutenção do par (macho e fêmea). A "chamada" é principalmente constituída por notas monossilábicas ou

dissilábicas, nunca em número superior a quatro ou cinco notas, e aparece na coordenação de comportamento do jovem, do bando e da família durante as atividades de manutenção, como alimentação, migração, bando e resposta a predadores (KREBS; KROODSMA, 1980).

As “chamadas” segundo (CHEN; MAHER, 2006) são geralmente sequências compactas de frases, enquanto canções são vocalizações longas e complexas. As frases são divididas em sílabas e as sílabas em elementos. Esses termos são definidos como:

- Elementos – As unidades elementares sonoras em vocalizações das aves;
- Sílabas – Conjunto de um ou mais elementos;
- Frase – Grupo de sílabas.

A Figura 1 a seguir, mostra o espectrograma do som de um Fogo – apagou mostrando as notas. Essa vocalização é uma espécie de chamada.

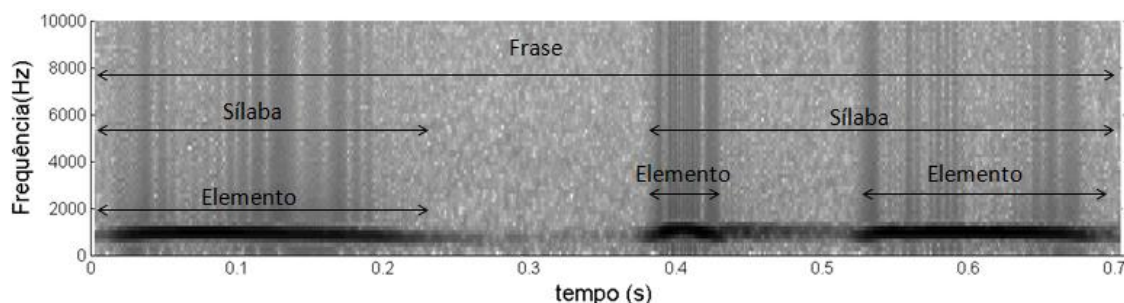


Figura 1 - Espectrograma do som de um Fogo – apagou.

Neste trabalho será utilizada somente uma frase do som de cada ave para o seu reconhecimento.

2.3 ESTRUTURA BIOLÓGICA DE PRODUÇÃO DOS SONS DAS AVES

O sistema de comunicação de som das aves é dividido em duas partes: emissão e recepção. Uma ave emite um sinal pelo bico, usando um órgão de produção de som

especial, a siringe, que propaga o som no ar (SELIN, 2007). Outra ave utiliza a audição e percepção na recepção do sinal.

As aves usam seus pulmões, brônquios, siringe, traquéia e bico no fenômeno de produção do som (KING; MCLELLAND, 1989).

A laringe está localizada na extremidade superior da traquéia. Ao contrário dos seres humanos, os pássaros não precisam da laringe na produção do som, mas a laringe é necessária para proteger o sistema respiratório, por exemplo, durante a alimentação (BRACKENBURY, 1989).

O órgão vocal, chamado siringe, é o principal órgão na produção de cantos e chamadas (KING; MCLELLAND, 1989) e (BRACKENBURY, 1989). Ele está localizado no fundo do peito da ave. A Figura 2 ilustra como a siringe e os demais órgãos responsáveis no fenômeno de vocalização estão localizados nas aves. Observe-se que a siringe localiza-se na extremidade terminal da traquéia e porções iniciais dos dois brônquios primários.

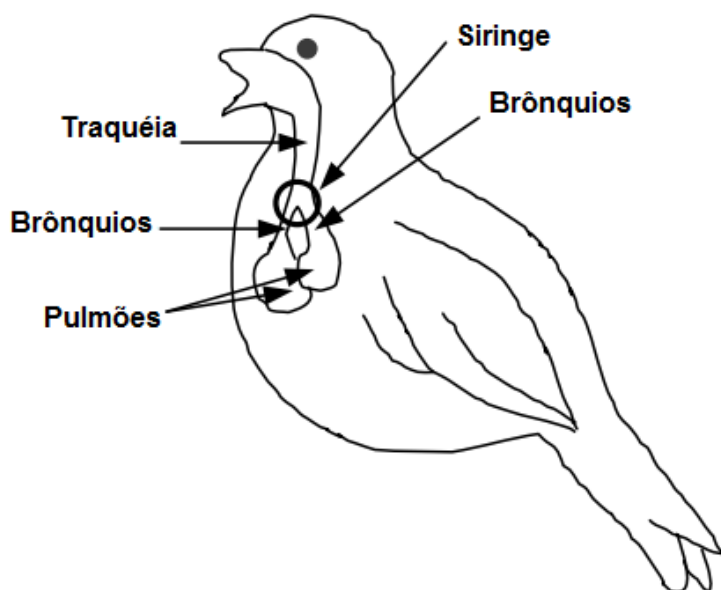


Figura 2 - Estrutura do sistema de produção de vocalização das aves (SELIN, 2007).

Uma parte controlável independentemente da siringe reside em cada brônquio, de modo que essas duas fontes de som são usadas para gerar o som final. Algumas aves podem cantar duas notas ou frases diferentes ao mesmo tempo. No entanto, um lado do sistema parece dominar o outro (SELIN, 2007).

Entre a siringe e o bico existe um canal vocal, que é usado para modificar o som final. O trato traqueal atua como um filtro seletivo, o qual pode ser ajustado em uma variedade de maneiras (SELIN, 2007).

O alongamento e a retração do pescoço e o alongar e encurtar do trato vocal pode afetar a forma do som final. A qualidade do som pode ser ainda influenciada pela constrição da laringe, por músculos da garganta ou pela estrutura e os movimentos do bico (KING; MCLELLAND, 1989), ou seja, umas séries de fatores próprios das aves podem modificar o som final da vocalização.

Segundo (MARQUES, 2009), além das próprias influências da ave, o meio em que a ave vive pode modificar de maneira sutil as características de sua vocalização. Por exemplo, na floresta as aves podem emitir o seu canto em uma sonoridade mais baixa já que o número de ruídos deste ambiente será menor que os encontrados no meio urbano onde a ave deve aumentar a intensidade sonora da sua vocalização.

2.4 A FUNÇÃO DO CANTO DAS AVES

Nas aves, o sistema de comunicação a longa distância é realizado exclusivamente pelo canto, assumindo assim, relações espaciais e sociais entre os indivíduos. O canto foi definido funcionalmente como: o sinal de comunicação sonora cuja função biológica primordial é o reconhecimento específico (VIELLIARD, 1987). Em diversas espécies de aves apenas o macho canta, portanto, além do reconhecimento

específico, nos machos outras funções são atribuídas ao canto, dentre elas destacamos (MARQUES, 2008):

- (1) Atração da fêmea;
- (2) Sinais de alerta;
- (4) Sinais sobre o temperamento e habilidades da ave;
- (4) Territorialidade (KROODSMA; BYERS, 1991; BAPTISTA; GAUNT, 1994).

Portanto, é fundamental a ampliação dos conhecimentos sobre a bioacústica da fauna brasileira, pois, é um mecanismo de comunicação que sofre influências diversas, que afetam diretamente a biologia destas aves (sociobiologia, reprodução, territorialismo, ecologia, etc.). Enfim, o entendimento da bioacústica é uma poderosa ferramenta para a compreensão dessas aves, que terá reflexo direto na conservação das mesmas.




2.5 AVES ESTUDADAS NESTE TRABALHO

Os cantos das aves podem ser ouvidos de maneira mais ou menos diferentes, de acordo com o indivíduo que escuta e com a região em que se encontra. Alguns são puramente onomatopéicos, como por exemplo o Fim-fim (*Euphonia chlorotica*) também conhecido como vim-vim (Maranhão e Piauí), vem-vem (Natal/RN) e vi-vi (interior de São Paulo). Outros são expressões humanizadas, como por exemplo o Pitiguari (*Cyclarhis gujanensis*), também conhecido com gente-de-fora-vem, segundo a sonoridade de seu canto.

Apesar de haver uma grande quantidade de aves de nomes onomatopéicos, as espécies escolhidas para estudo se deu por possuírem uma grande quantidade de sons nos bancos de dados disponíveis na internet (XENO, 2012) e (WIKIAVES, 2012).

Abaixo, a Tabela 1, apresenta as aves escolhidas para serem utilizadas no presente trabalho, coordenando suas imagens a seus nomes científico e comum. As imagens das aves foram obtidas da seguinte fonte: Wikiaves (2012).

Tabela 1– *Espécies Utilizadas neste trabalho.*

<p><i>Nome Científico</i> <i>Nome Comum</i></p>	<p><i>Imagem</i></p>
<p><i>Crypturellus undulatus</i> Jaó</p>	
<p><i>Dendrocygna viduata</i> Irerê</p>	
<p><i>Herpetotheres cachinnans</i> Acauã</p>	

Aramides cajanea
Saracura-três-potes



Columbina squammata
Fogo-apagou



Dromococcyx pavoninus
Peixe-frito-pavonino


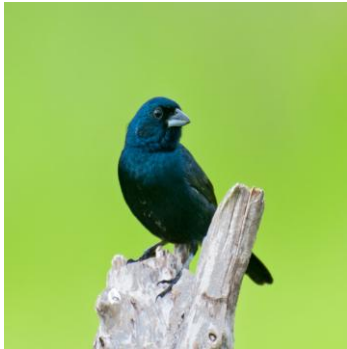




Pitangus sulphuratus
Bem-te-vi



Megarynchus pitanguá
Neinei



<p><i>Cyclarhis gujanensis</i> Pitiguari</p>	 A photograph of a Pitiguari bird (Cyclarhis gujanensis) perched on a branch. It has a grey head, a bright yellow throat, and a brown body.
<p><i>Volatinia jacarina</i> Tiziu</p>	 A photograph of a Tiziu bird (Volatinia jacarina) perched on a branch. It has a dark blue body and a yellow throat.
<p><i>Todirostrum poliocephalum</i> Teque-teque</p>	 A photograph of a Teque-teque bird (Todirostrum poliocephalum) perched on a branch. It has a yellow body and a dark head.
<p><i>Euphonia chlorotica</i> Fim-fim</p>	 A photograph of a Fim-fim bird (Euphonia chlorotica) perched on a branch. It has a yellow body and a dark head.

Abaixo segue descrição de cada ave estudada. Todas as características referentes às aves descritas abaixo foram retiradas de (SICK,1985).

- **Jaó (*Crypturellus undulatus*):** ave da ordem Tinamiformes, da família Tinamidae. Seu nome tem origem do seu canto, pois parece que a ave está dizendo “Eu sou Jaó”,

tendo maior ênfase justamente na parte referente ao seu nome. Também é conhecido como Macucauá e Sururina na região Norte do Brasil. Nome em inglês: *Undulated Tinamou*. Mede cerca de 31 centímetros de comprimento e tem o formato do corpo que lembra a uma galinha. Uma característica interessante é que, embora possam voar, não possuem a quilha óssea do osso do peito que possibilita o voo contínuo. Presente em todas as regiões brasileiras e também na Venezuela, Peru, Guiana e Colômbia. Habita a mata de várzea, galeria, capoeira, matas secas, raias e cerrados, que são regiões abundantes em alimentos para essa ave que se alimenta de pequenos frutos, sementes e insetos.

- **Irerê (*Dendrocygna viduata*):** O Irerê é uma ave anseriforme da família Anatidae. Também conhecido por paturi, marrecão (Rio Grande do Sul), siriri, marrecá-viúva e marrecá-piadeira. Seu nome irerê é devido a sua vocalização, pois quando emite um som parece estar pronunciando “Irerê”. Nome em inglês: *White-faced whistling-duck*. Tem aproximadamente 45 centímetros de altura. É encontrada em todo o território brasileiro e na Argentina. É uma ave totalmente migradora que se alimenta de pequenas sementes, larvas de inseto e pequenos crustáceos.

- **Acauã (*Herpetotheres cachinnans*):** ave pertencente à ordem dos Falconiformes, da família Falconidae. Seu canto dá origem ao seu nome “acauã” e é repetido seguidamente durante alguns segundos. Conhecido também como acanã, acuã, cauã, macaguã, macauã, nacauã, uacanã, gavião-cauã e gavião-couã. Nome em inglês: *Laughing Falcon*. Mede cerca de 47 centímetros de comprimento. Presente em todo o Brasil e também do México à Argentina. Comum em bordas de florestas, capoeiras, florestas de galeria, campos com árvores e cerrados. Alimenta-se de lagartos e cobras, tendo preferência por morcegos.

- **Saracura-três-potes (*Aramides cajanea*):** é uma ave pertencente à ordem Gruiformes e família Rallidae. Possui um canto bem característico e peculiar, já que ela consegue pronunciar de forma bem marcante o nome “Três-potes”. Normalmente participa de duetos ou coro com indivíduos mais distantes. Nome em inglês: *Gray-necked Wood-rail*. É uma ave que em geral é mais escutada do que vista, pois vivem no chão de áreas alagadas com vegetação densa, manguezais, margens de rios e lagoas. Mede cerca de 39 centímetros e são consideradas saracuras de grande porte. Sua coloração permite uma excelente camuflagem juntamente com as vegetações locais. Pode ser encontrada em todos os estados do Brasil. São onívoras, alimentando-se de capim, sementes, larvas de insetos, pequenas cobras, peixes e crustáceos.

- **Fogo-apagou (*Columbina squammata*):** Ave da ordem Columbiformes e família Columbidae. O seu nome é praticamente a tradução de sua vocalização, porém ela é conhecida por vários outros nomes, tais como rolinha-carijó, fogo-pagô, rolapedrês, Felix-cafofo (Paraíba), paruru e galinha-de-Deus. Já nos estados do Rio Grande do Norte, ela é conhecida como rolinha-cascavel, pois sua coloração lembra as escamas da cobra cascavel e também pelo fato de emitir um forte som com as asas, lembrando o chocalho de uma cascavel. Nome em inglês: *Scaled Dove*. Mede cerca de 19,5 centímetros. Sua coloração e tamanho auxiliam e muito para que essa ave possa se camuflar entre os galhos das árvores. Presente nas regiões Nordeste, Centro-oeste e nos estados de São Paulo, Paraná e Rio Grande do Sul. Encontrado também na Guiana Francesa, Venezuela, Paraguai e Argentina. Alimentam-se de pequenos insetos e de frutos.

- **Peixe-frito-pavonino (*Dromococcyx pavoninus*):** é uma ave Cuculiforme da família Cuculidae. Nome em inglês: *Pavonine Cuckoo*. Mede cerca de 28,5 centímetros. São aves de hábitos solitários, tanto que colocam seus ovos nos ninhos de outras espécies

principalmente de alguns tiranídeos. Vivem em florestas tropicais perene. Alimentam-se de pequenos insetos e lagartos. Estão distribuídos por quase toda a América Central e do Sul. No Brasil é encontrado na Amazônia, Mato-grosso, Goiás, Rio de Janeiro, São Paulo e Paraná.

- **Bem-te-vi (*Pitangus sulphuratus*):** é uma ave Passeriforme da família dos Tiranídeos. É conhecido também como bem-te-vi-de-coroa e bem-te-vi-verdadeiro. Estão distribuídos em todo o território brasileiro, sendo encontradas em matas, cidades, regiões próximas de água, plantações e pastagens. É uma ave popular em outros países, onde também são reconhecidos por nomes onomatopéicos, como *Kiskadee* em inglês, *qu'est ce* em francês e *bichofêo* em espanhol (Argentina). Mede cerca de 22,5 centímetros. Possui coloração que é comum a outros tiranídeos, como o Neinei, o bentivizinho-do-brejo, o bentivizinho-de-penhacho-vermelho e o bentivizinho-de-asa-ferruginea o que diferencia um do outro é o tamanho da ave, da cauda e do bico. Possuem a impressionante característica de se adaptar a qualquer meio, descobre sempre novas fontes de alimentos. Alimentam-se de insetos, de roedores, ovos de outras aves, lagartos, girinos, pequenos peixes e frutas. Uma característica interessante dessa ave é que elas são agressivas, ameaçando até gaviões e urubus quando esses se aproximam de seus territórios.

- **Neinei (*Megarynchus pitangú*):** assim como o bem-te-vi são da ordem Passeriforme da família dos Tiranídeos. À primeira vista muito parecido com o Bem-te-vi, dele se distinguindo pelo bico nitidamente maior e chato, o que motivou o nome do gênero, e principalmente pela voz, claramente diferente, que lhe valeu o nome onomatopéico. É conhecido também como Bem-te-vi-gameleiro (Minas Gerais) e Bem-te-vi-pato (Espírito Santo) Costuma-se encontrar esta espécie em florestas, paisagens abertas com árvores esparsas e cerrados. Nome em inglês: *Boat-billed Flycatcher*. Mede cerca de

21,5 centímetros e possuem praticamente a mesma alimentação que o Bem-te-vis, porém por possuir um bico mais longo consegue alimentar-se de insetos e frutos maiores. Está distribuído em todo o território brasileiro.

- **Pitiguari (*Cyclarhis gujanensis*):** é uma ave passeriforme da família dos Vireonidae. Conhecido também popularmente como gente-de-fora-vem, segundo a sonoridade do seu canto. Nome em inglês: *Rufous-browed Peppershrike*. Mede aproximadamente 18 centímetros. Presente em todo o Brasil, exceto em pequenas áreas da Amazônia. Alimentam-se de insetos, larvas e pequenos frutos. Costumam matar os insetos batendo-os contra as árvores.

- **Tiziu (*Volatinia jacarina*):** é uma ave passeriforme da família Emberizidae. Conhecido também como tizirro, saltador, veludinho, papa-arroz, bate-estaca (RJ), serrador, serra-serra e alfaiate. Nome em inglês: *Blue-black Grassquit*. Mede cerca de 12 centímetros. Está distribuído em todos os países da América do Sul, do México ao Panamá, sendo que no Brasil está presente em todos os estados. Alimentam-se principalmente de semente de gramíneas e pequenos insetos.

- **Teque-teque (*Todirostrum poliocephalum*):** é uma ave passeriforme da família Rhynchocyclidae. Também conhecido como sebinho, caga-sebo, ferreirinho, ferreirinho-teque-teque, joão-de-cristo, marrequinha, papa-sebo, patinho e sebinho-teque-teque. Nome em inglês: *Yellow-lored Tody-Flycatcher*. É uma ave pequena, medindo aproximadamente 8,8 centímetros. Ocorre na Mata Atlântica litorânea, Mata Atlântica de encosta, Mata Atlântica de altitude, na restinga arbórea e em matas mesófilas, ou seja, se encontra praticamente na Região sul, Paraná, Santa Catarina e parte da Bahia. Alimentam-se de pequenos insetos, invertebrados e frutos pequenos.

- **Fim-fim (*Euphonia chlorotica*):** é uma passeriforme da família Fringillidae. Também conhecido como vim-vim (Maranhão e Piauí), fi-fi-verdadeiro, vem-vem (Natal/RN),

vi-vi e puvi (interior de São Paulo). No Nordeste recebe também as denominações de vem-vem, gaturamo-fifi ou guriatã. Nome em inglês: *Purple-throated Euphonia*. Mede aproximadamente 10 centímetros. É encontrado na mata rala, no cerrado, na caatinga, nos cocais e nas matas serranas. Está presente em todos os estados brasileiros e em vários outros países, como Venezuela, Guianas e Peru. Alimentam-se de frutos, o que tornam essas aves excelentes dispersoras de sementes.

CAPÍTULO 3

3 METODOLOGIA

Nesta parte do trabalho serão apresentadas as técnicas utilizadas no pré-processamento para o tratamento do sinal, mostrando também as fases de treinamento e reconhecimento das aves.

Realizar o pré-processamento é fundamental para a garantia de um bom reconhecimento, pois os sons, coletados na internet, foram obtidos de várias formas e os mesmos podem possuir ruídos que prejudiquem o reconhecimento das aves. Esses ruídos são oriundos de várias fontes tais como: chuva, ventos, outros animais e quaisquer outros ruídos que não sejam oriundos da vocalização da ave (HAYES, 2006).

O trabalho foi executado de acordo uma sequência de ações que vai desde a aquisição do sinal até a fase final, o reconhecimento das aves. Essa sequência é descrita a seguir:

- Em um primeiro momento o sinal é adquirido em fontes como (XENO, 2012) e (WIKIAVES, 2012);
- Então o sinal passa por uma padronização, sendo essa padronização a conversão do sinal (para formato *wav*), a sua normalização e sua amostragem;
- É realizada uma filtragem para a eliminação dos ruídos indesejados;
- Em seguida, para obtermos somente um chamado da ave, é realizada a segmentação, separando regiões de sons e silêncios;
- São então extraídas do sinal as características a serem utilizadas para o reconhecimento;

- De posse das características é feito uma análise desses parâmetros, fase conhecida como treinamento;
- Após o treinamento é construída a árvore de decisão;
- Finalmente é feito o reconhecimento utilizando a árvore de decisão.

A Figura 3 mostra a sequência adotada neste trabalho, desde aquisição do sinal até a fase final o reconhecimento das aves:

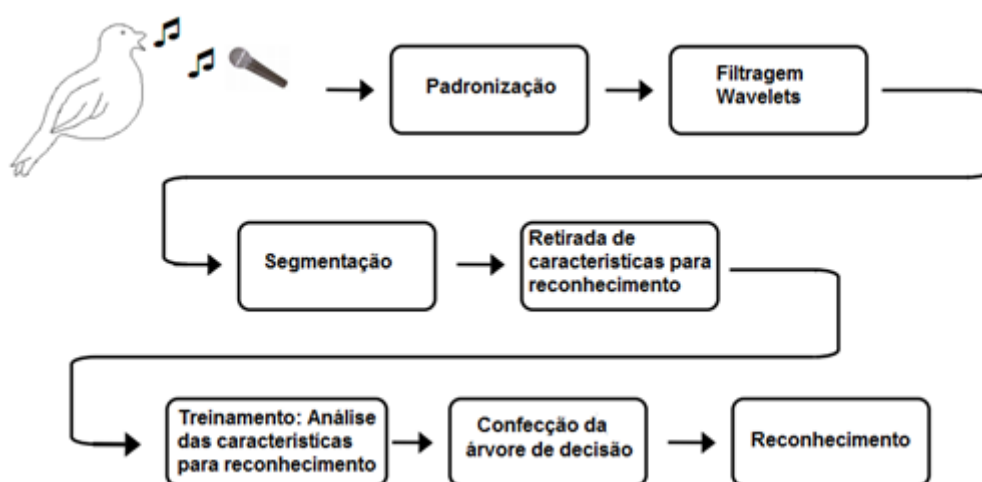


Figura 3 - Diagrama de blocos do sistema desenvolvido, mostrando a sequência adotada para o reconhecimento de espécies de aves. Adaptado de (SELIN, 2007).

O sinal (chamado ou canto da ave) é um sinal analógico e de modo geral, em um sinal analógico, os valores são processados ou medidos com variáveis indefinidas por se tratar de um sinal contínuo que varia em função do tempo.

Para que o sinal possa ser processado de forma digital o mesmo deve ser coletado em forma de amostras discretas (descontínuas) no tempo, que visam reconstruir o sinal original no formato digital (OPPENHEIM; SCHAFER, 1999).

O sinal analógico é amostrado e transformado em sinal digital, e é então normalmente guardado em um vetor. Este vetor é indexado e cada ponto é uma função

de uma variável inteira n (que representa o tempo) e a função é chamada $x(n)$, dada pela equação (1). Os valores de x são então sequenciados para gerar o vetor coluna x (OPPENHEIM; SCHAFER, 1999):

$$x = [x(0), x(1), x(2) \dots x(N - 1)]' \quad (1)$$

Obtido o vetor x , pode-se realizar o pré-processamento fazendo-se a análise no domínio do tempo e no domínio da frequência.

3.1 DOMÍNIO DO TEMPO

No domínio do tempo o sinal é representado pelo vetor x distribuído ao longo do tempo t (segundos), podendo sofrer ao longo do tempo variação na amplitude (HAYKIN; VEEN, 2001). A figura 4 mostra o canto de um Fogo-apagou no domínio do tempo, onde o sinal varia sua amplitude no intervalo $[-1 \ 1]$ e possui duração de aproximadamente 10 s. No domínio do tempo serão extraídas do sinal algumas características que serão utilizadas para o reconhecimento das aves.

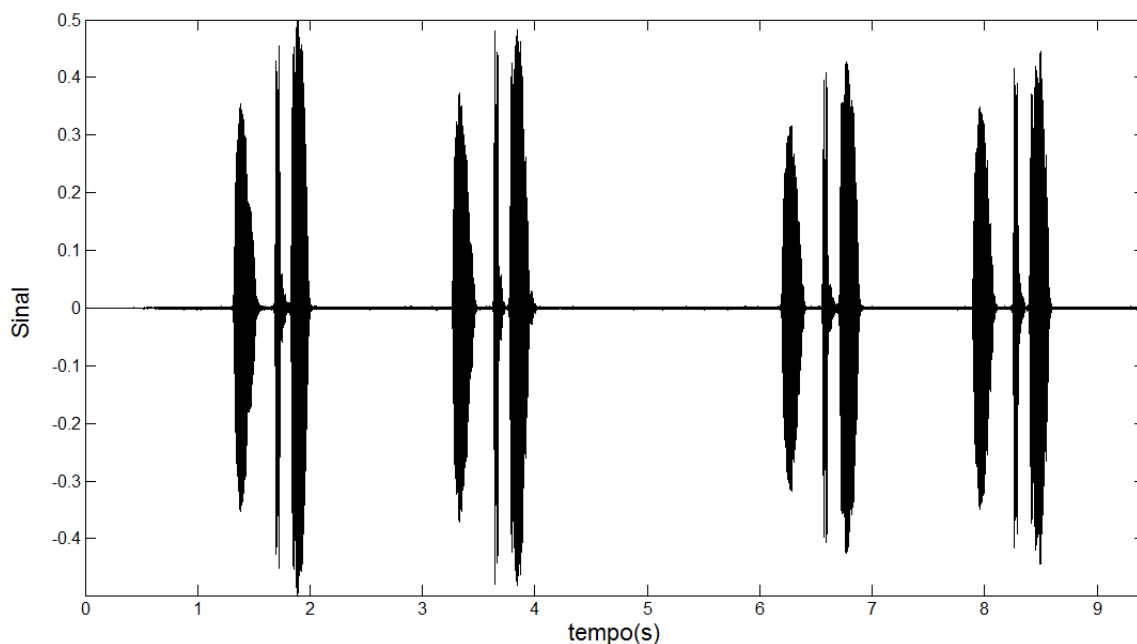


Figura 4- Sinal no domínio do tempo, canto de um fogo-apagou.

3.2 DOMÍNIO DA FREQUÊNCIA

No domínio da frequência o sinal é representado pelo seu espectro, sendo possível realizar diversas operações como a aplicação de filtros e a extração de características relacionadas às frequências (OPPENHEIM; SCHAFER, 1999). Essas características, juntas com as extraídas no domínio do tempo, serão de grande valia para o reconhecimento das aves.

A representação do sinal no domínio da frequência pode ser obtida a partir da transformada de Fourier para sistemas discretos (DTFT), que é aplicada em sinais digitais e discretos.

A transformada de Fourier do sinal x é calculada por (2) conforme apresentado em (OPPENHEIM; SCHAFER, 1999):

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-i\omega n} \quad (2)$$

onde $X(k)$ é o espectro em frequência do sinal $x(n)$, obtido pela decomposição de $x(n)$ para cada instante de tempo n em uma base exponencial $e^{-i\omega n}$, sendo N o número total de pontos.

Para o cálculo computacional da Transformada de Fourier, foi usada no decorrer desta pesquisa, a Transformada Rápida de Fourier (FFT), definida por (OPPENHEIM; SCHAFER, 1999), como (3):

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{nk} \quad (3)$$

Sendo: $W_N^{nk} = e^{-i2\pi/N}$.

A figura 5 mostra o espectro de magnitude do sinal da figura 4 representado no domínio da frequência. Pode-se observar que o sinal está contido entre 600 e 1200 Hz com uma magnitude máxima de aproximadamente 120 dB.

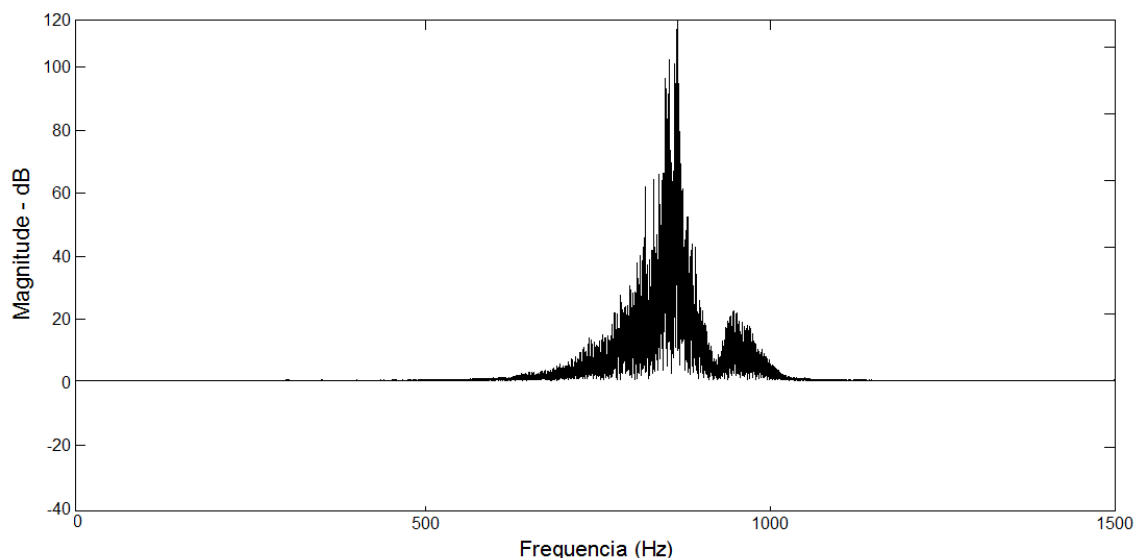


Figura 5 - Sinal no domínio da frequência, canto de um Fogo-apagou.

3.3 PADRONIZAÇÃO DOS SINAIS

Para este trabalho, a coleta do áudio não foi realizada diretamente no *habitat* da ave, mas foram retirados de bancos de dados disponíveis na Internet (XENO, 2012) e (WIKIAVES, 2012). Como os sinais coletados são oriundos de fontes diferentes e de formas de gravações distintas, com taxas de amostragem, equipamentos de gravação, ambiente e formato de arquivos variados, é necessário padronizá-los (HAYES, 2006).

3.4 ARQUIVOS WAV

O software utilizado para trabalhar com a análise dos sons e para a realização do reconhecimento foi o MATLAB, que só permite trabalhar com os arquivos de sons no formato wav (MATLAB, 2010). Portanto, o primeiro passo é converter todos os arquivos de formatos diferentes em formato wav.

A maioria dos arquivos retirados de (XENO, 2012) e (WIKIAVES, 2012) são do formato mp3. Este formato de arquivo é um método de compressão de áudio com

perdas (WIKIPEDIA, 2012), porém essa compressão consiste em retirar do áudio tudo aquilo que o ouvido humano normalmente não conseguiria perceber, devido a fenômenos de mascaramento de sons e de limitações da audição humana.

Para alguns especialistas, a grande desvantagem do mp3 em relação ao wav continua sendo a qualidade do som, que embora passe quase despercebida pela grande maioria dos usuários, faz grande diferença (CEO, 2012). Portanto, ao converter estes arquivos para o formato mp3 podem ocorrer perdas na qualidade do áudio, pois alguns componentes de frequência são descartados para tornar o arquivo menor (AURELIO, 2004). Esses componentes de frequência, uma vez descartados, não podem ser recuperados, mesmo que o arquivo seja novamente convertido para o formato wav (BRANDENBURG, 2009).

As perdas que ocorrem no formato mp3 não são recuperadas quando o arquivo é convertido para wav, contudo espera-se que essa perda não influencie nos resultados da pesquisa.

3.5 NORMALIZAÇÃO E TAXA DE AMOSTRAGEM

Como os sons foram gravados em condições diferentes, possuem diferentes níveis de intensidade. Para que fiquem na mesma faixa de trabalho, o primeiro passo no tratamento desses sons é a extração de seu valor médio, conforme equação (4), e a sua normalização em relação ao seu valor máximo, conforme equação (5).

$$x(k) = x_{orig}(k) - média(x_{orig}) \quad (4)$$

$$x_{norm}(k) = \frac{x(k)}{\max(|x(i)|)} \quad (5)$$

onde $x(k)$ é o vetor que representa o canto do pássaro após a extração do valor médio, $x_{orig}(k)$ é o vetor que representa o canto original, ou seja, o que foi gravado, $média(x_{orig})$ é a média do vetor $x_{orig}(k)$, $x_{norm}(k)$ é o vetor x normalizado, $\max(|x(i)|)$ é o valor máximo do vetor $x(k)$ e $k = 1, 2, 3, \dots, k$ onde k é o número de amostras do sinal. Dessa maneira todos os sinais terão valor médio zero, o que facilita a análise do espectrograma, e todos os valores dos sinais estarão na mesma faixa de amplitude, entre -1 e 1 (SELIN *et al.*, 2007; KWAN *et al.*, 2004).

Em seguida define-se a taxa de amostragem. Sabe-se que máxima da frequência de vocalização das aves é em torno de 12000 Hz podendo os seus harmônicos chegar a 24000 Hz. Escolheu-se, portanto, a taxa de amostragem de 25000 Hz, de acordo com o teorema de Nyquist (OPPENHEIM; SCHAFER, 1999), para garantir uma margem de segurança para a extração de características como as frequências mais intensa, mínima e máxima (SMITH, 1999).

3.6 FILTRAGEM

A filtragem é um processo essencial para que assim o som possa ser utilizado de forma segura no processo de reconhecimento (HAYES, 2006). Muitas vezes a vocalização da ave gravada contém sons ou ruídos indesejáveis que podem ter sido ocasionados por outros animais, vento, chuva ou sons do ambiente onde a ave vive. Então, para eliminá-los o máximo possível, utilizou-se o sistema de filtragem, que no caso deste trabalho utiliza como base a teoria das *wavelets* num processo conhecido como *denoising*.

3.6.1 WAVELETS

A teoria *wavelet* é uma ferramenta matemática nova e extremamente poderosa (MISITI *et al.*, 2006). Foi desenvolvida e proposta pela primeira vez por Jean Morlet (1975) e pela equipe de Alex Grossman, na França (1988).

Wavelet significa ao pé da letra “ondinha” e recebeu este nome devido às suas características de serem oscilantes, de rápido decaimento e de energia concentrada. Atualmente muitas áreas da ciência vêm utilizando a transformada *Wavelet* em seus trabalhos, mas a área de maior destaque em sua utilização é a de processamento digital de sinais (DSP) (MORETTIN, 1999).

Uma das características das wavelets é o fato de sua forma de onda ter duração limitada e decair rapidamente para zero. (MISITI *et al.*, 2006). Se a forma de onda das *wavelets* forem comparadas com a forma de onda de uma senóide (Figura 6), pode-se observar que existem diferenças importantes entre elas. A primeira delas é que as senóides têm duração infinita, ou seja, elas vão de menos infinito até mais infinito. Outra diferença é que as senóides são formas suaves e previsíveis enquanto as *wavelets* tendem a ser irregulares. Formalmente, diz-se que as senóides têm suporte infinito, enquanto as *wavelets* são funções de suporte compacto. Por isso, elas se prestam à detecção de fenômenos localizados no tempo.



Figura 6 - Forma de onda de uma senóide e de uma Wavelet (MISITI *et al.* , 2006).

A transformada de Fourier (TF) consiste na idéia de que um sinal pode ser representado como a soma infinita de senos e cossenos, ou seja, o sinal pode ser decomposto em um conjunto de várias ondas senóides de frequências distintas

(HAYES, 2006). Já a transformada *wavelet* trabalha de forma semelhante, porém ao invés de decompor o sinal em senóides ela secciona o sinal em versões transladadas e escalonadas de uma determinada *wavelet* padrão, as chamadas *wavelets-mãe*.

A grande desvantagem da transformada de Fourier é que na passagem do sinal para o domínio da frequência são perdidas as informações referentes ao domínio do tempo (HAYES, 2006). No entanto a novidade em relação à análise de Fourier é que a análise *wavelet* não é feita segundo a frequência, mas sim segundo a escala. Os algoritmos usados no domínio *wavelet* processam dados em diferentes escalas e resoluções, permitindo que sejam vistos tanto o comportamento global quanto os detalhes de um sinal (GRAPS, 1995). Como a TF trabalha com janelamento, o que torna um problema para essa transformada é o fato de que se a janela for relativamente grande ela irá comportar diferentes frequências, mas não irá distinguir os trechos específicos de cada uma. Já para janelas estreitas ou curtas, melhora-se a resolução temporal, mas prejudica-se a resolução em frequência, pois não se consegue distinguir frequências baixas ou pequenas variações (HAYES, 2006).

Para compreender de uma maneira clara a transformada *wavelet*, tem-se que ter em mente que na prática vários sinais, assim como os trabalhados nesta pesquisa, vêm representados no domínio do tempo e por uma determinada amplitude. Quando é aplicada a transformada de Fourier para sistemas discretos (DTFT) ou a transformada *wavelet* para discreta (DWT) o intuito é obter a representação de um sinal no domínio da frequência, e realiza-se esta transformação de domínios porque muitas vezes a informação que se precisa não pode ser vista no domínio do tempo. Em muitos casos a parte mais importante da informação do sinal está “escondida” nas suas frequências, ou seja, o domínio da frequência pode dizer muitas coisas sobre aquele sinal.

3.6.2 WAVELET MÃE

A *wavelet* mãe $\Psi(t)$ representa a função que irá decompor o sinal. Existe uma infinidade de possibilidades para essa função. Dentre a variedade existente, foram escolhidas para serem utilizadas neste trabalho somente duas famílias: Haar e a daubechies. Elas foram escolhidas por se tratarem de funções mães extremamente conhecidas e utilizadas em inúmeros estudos nas mais variadas aplicações, tendo sido aplicada na filtragem de sons de aves por Selin (2006).

As características dessas famílias são as seguintes:

- Haar: a função *wavelet* Haar é a mais simples *wavelet*. Ela é representada pela equação (6) a seguir.

$$\Psi(t) = \begin{cases} 1, se\ 0 \leq t < \frac{1}{2} \\ -1, se\ \frac{1}{2} \leq t < 1 \\ 0, caso\ contrário \end{cases} \quad (6)$$

A figura 7 mostra a função Haar, que é descontínua e representa uma função de degrau 1. Como estas funções são simples, elas possuem aplicações limitadas (SOARES, 2009).

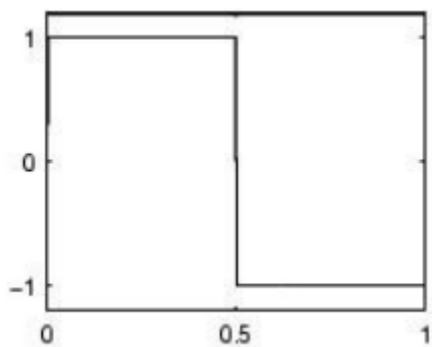


Figura 7 – Função Haar. Fonte: Matlab R2006a.

- Daubechies: Essa *wavelet* foi desenvolvida por Ingrid Daubechies, uma das maiores pesquisadoras no desenvolvimento e aplicações das *wavelets* (DAUBECHIES, 1992).

Possui como característica o fato de não ter uma forma analítica, exceto a db1, que é a função de Haar.

As *wavelets* ortogonais de Daubechies, “dbN”, são perfeitamente compactas no tempo, mas no domínio da frequência, tem um alto grau de superposição espectral entre as escalas. Sua maior vantagem é ser ortogonal, o que significa que um erro no sinal de entrada não cresce com a transformação e a estabilidade numérica computacional é assegurada (DAUBECHIES, 1997). Por outro lado, não possuem fase linear. Como características, essas *wavelets* apresentam suporte compacto e normalmente não apresentam suavidade em seu comportamento (DAUBECHIES, 1992). A Figura 8 mostra as Daubechies de ordem 2 a 10, que foram estudadas neste trabalho.

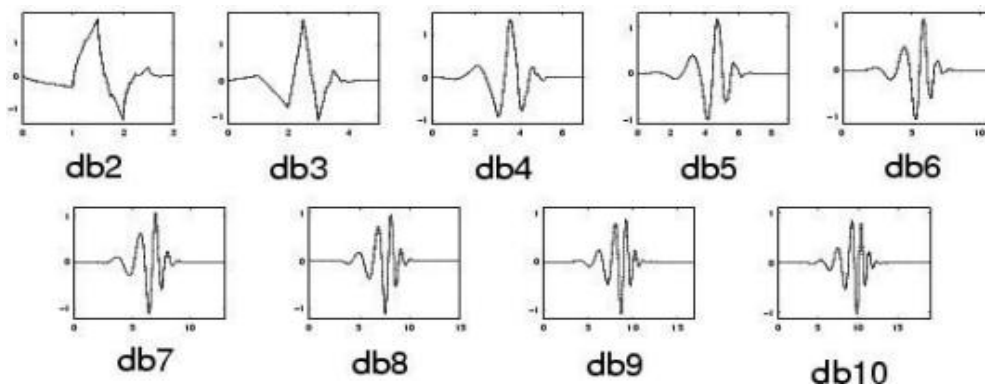


Figura 8 – Funções Daubechies. Fonte: Matlab R2006a

3.6.3 TRANSFORMADA CONTÍNUA WAVELET (CWT)

Sendo $\Psi(t)$ uma *wavelet* mãe e $f(t)$ o sinal em questão, temos que a CWT em relação à *wavelet* mãe escolhida será representado pela equação (7):

$$W(a,b) = \int_{-\infty}^{\infty} f(t)\Psi_{a,b}(t)dt \quad (7)$$

Onde a função $\Psi_{a,b}(t)$ é uma translação e dilatação da *wavelet* mãe escolhida. Essa função $\Psi_{a,b}(t)$ tem a seguinte característica:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad (8)$$

Sendo a o parâmetro responsável pela dilatação e b o responsável pela translação. Ambos são valores reais e a é diferente de zero.

Observe que o fator a possibilita realizar na função *wavelet* uma “dilatação” (se $a > 1$) ou uma “compressão” (se $a < 1$) da função *wavelet* mãe (SOARES *et al.*, 2005). Essa modificação da *wavelet* mãe é definida segundo Daubechies (1992), como uma *wavelet* filha.

Na transformada contínua *wavelet* (CWT) a *wavelet* mãe sempre deve ser definida, pois a representação do sinal no domínio *wavelet* sempre é feita em relação à *wavelet* mãe, ou seja, as considerações estabelecidas para uma transformada são totalmente inválidas para outra transformada, se as *wavelets* mãe utilizadas não forem as mesmas (LIMA, 2004). Portanto se for comparar resultados de sinais, obviamente eles têm que ser trabalhados com as mesmas *wavelets* mãe.

3.6.4 TRANSFORMADA WAVELET DISCRETA (DWT)

Normalmente a DWT é utilizada para as aplicações práticas enquanto a CWT é utilizada mais nas aplicações teóricas para a obtenção e compreensão das características matemáticas das funções *wavelets*.

Segundo Hernandes e Weis (1996) uma maneira de se discretizar a equação referente ao CWT é adotar $a = a_0^m$ e $b = na_0^m b_0$ com m e n pertencentes aos inteiros e $a_0 > 1$ e $b_0 \neq 0$.

Então tem-se a equação 9:

$$W(m, n) = \frac{1}{\sqrt{|a_0^m|}} \int_{-\infty}^{\infty} f(t) \Psi \left(\frac{t - na_0^m b_0}{a_0^m} \right) dt \quad (9)$$

De acordo com (Soares, 2009), e a partir da análise da equação (9), pode ser feito as seguintes observações:

- 1) a transformada *wavelet* discreta é definida apenas para valores de escalas positivos, $a_0 > 1$;
- 2) o passo da translação é proporcional à escala ($b = na_0^m b_0$);
- 3) a transformada *wavelet* discreta produz um conjunto finito de coeficientes *wavelet* ($DWT^f(m, n)$);
- 4) o processamento é realizado sobre tempo contínuo.

3.6.5 BANCO DE FILTOS E TRANSFORMADA WAVELET

Daubechies (1992) diz em seu trabalho que a transformada *wavelet* discreta é basicamente um processo de filtragem.

Na Análise *Wavelet* Discreta o sinal é dividido em aproximações e detalhes, onde as aproximações são os componentes de baixa frequência do sinal e os detalhes são os componentes de alta frequência do sinal (MISITI *et al.*, 2006).

Para inúmeros sinais, o conteúdo a baixa frequência é a parte mais importante. É isto que dá ao sinal a sua identidade. O conteúdo de alta frequência, por outro lado, mostra a característica do sinal. Considerando o canto de algumas aves, se removermos os conteúdos de alta frequência, o canto soa diferente, mas para outras continua normal. O problema é que as aves estudadas neste trabalho possuem faixas de frequências variáveis. O processo de filtragem é mostrado de forma simplificada na Figura 9, considerando-se apenas o primeiro nível de filtragem.

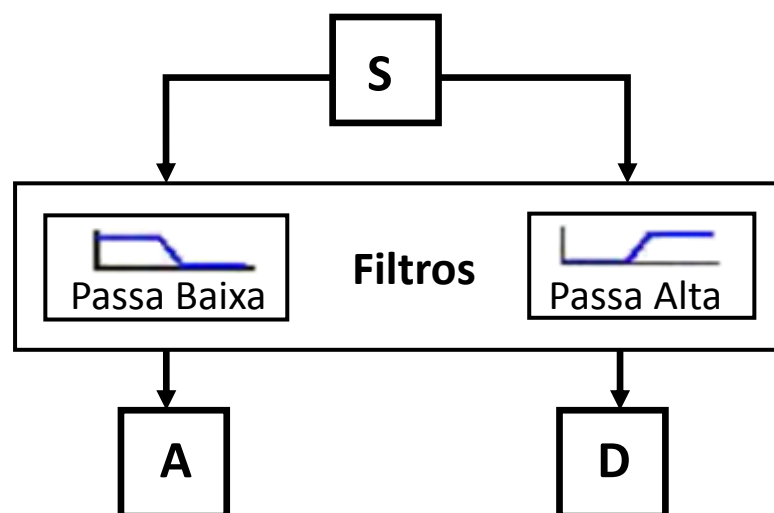


Figura 9 – Processo de Filtragem do sinal (MISITI et al., 2006).

O sinal original S passa através de dois filtros complementares, o filtro Passa Baixa e o filtro Passa Alta, que fornecem como saída os sinais de Aproximação (A) que possui as baixas frequências e o de Detalhe (D) que possui as altas frequências, respectivamente. Porém, se for utilizado esse processo em um sinal digital real, a quantidade de dados será duplicada em relação aos dados iniciais, devido à conclusão (MISITI *et al.*, 2006). Na figura 10 observa-se este problema um sinal de 1000 amostras que é decomposto nos sinais D e A , cada um de 1000 amostras, totalizando então 2000 amostras.

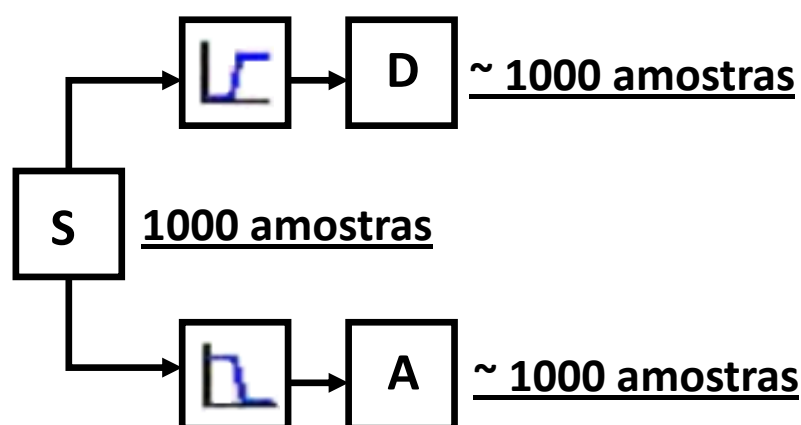


Figura 10 – Processo de decomposição do sinal. Adaptado de (MISITI et al., 2006).

Para evitar este problema, deve-se introduzir um operador para diminuir o número de amostras, conhecido como *downsampling*, onde são considerados os dados

de forma alternada, ou seja, considera-se o primeiro dado, rejeita-se o segundo, e assim por diante (BURRUS, 1998). A Figura 11 ilustra o processo de filtragem com a utilização do operador *downsampling*.

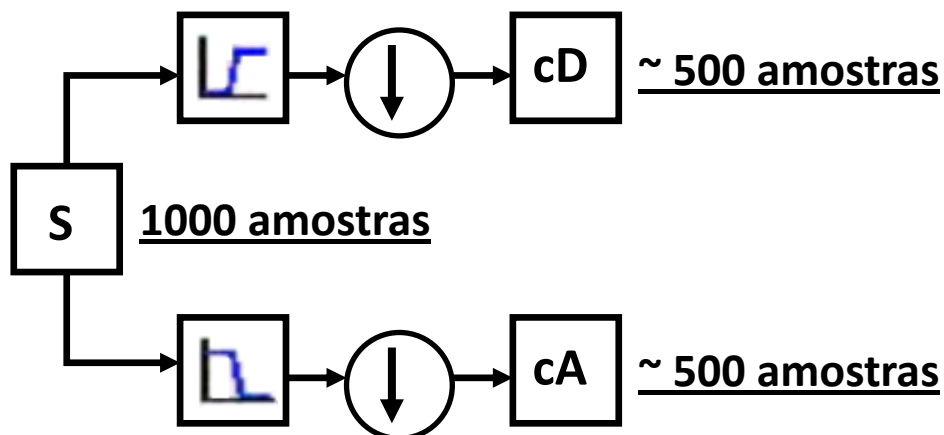


Figura 11 – Processo de diminuição do número de amostras utilizando o operador downsampling . Adaptado de (MISITI et al., 2006).

O processo da figura 11, que inclui a diminuição de amostras, produz os coeficientes cD e cA referentes a Transformada *Wavelet* Discreta.

O processo de decomposição pode ser iterativo, fazendo com que sucessivas aproximações sejam decompostas em novas aproximações e detalhes. Isto é chamado de “árvore de decomposição *Wavelet*” ou Análise de Multiresolução (AMR). Como o processo de análise é iterativo, em teoria ele pode ser continuado de forma indefinida. Na realidade, a decomposição pode ser efetuada até que os detalhes individuais consistam numa amostra simples ou pixel (no caso de imagens) (MISITI *et al.*, 2006). A técnica consiste na decomposição do sinal a ser analisado em dois outros sinais um de aproximação e outro de detalhe. A cada iteração, o sinal de aproximação (proveniente do filtro passa baixa) é novamente decomposto resultando em dois outros novos sinais, detalhado e atenuado, em níveis de frequência diferentes, fornecendo informações no domínio da frequência e do tempo (DAUBECHIES, 1992). Esta técnica de análise é

ilustrada na Figura 12, onde o sinal $f(x)$ é submetido ao processo de decomposição em AMR, em três níveis.

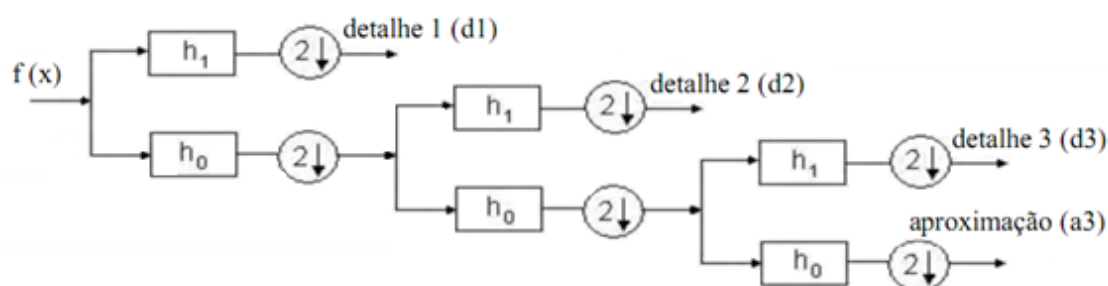


Figura 12 – Técnica de decomposição AMR.

A figura 13 mostra o processo de decomposição graficamente para três níveis.

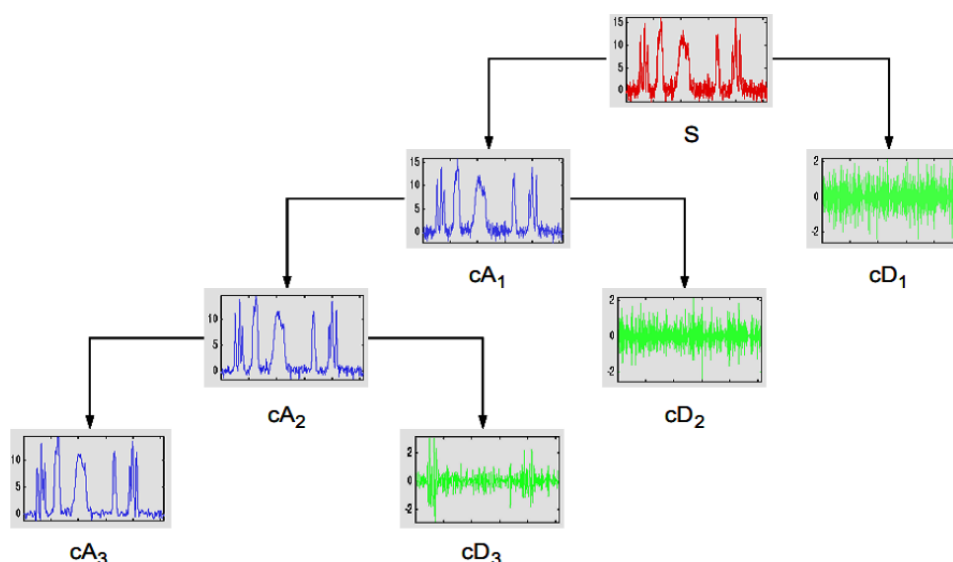


Figura 13 – Árvore gráfica de uma AMR (MISITI et al., 2006).

O sinal pode ser recuperado utilizando a transformada *wavelet* discreta inversa (IDWT – *Inverse Discrete Wavelet Transform*). Os componentes podem ser combinados, no sinal original, sem perdas de informações referentes ao sinal original. Para sintetizar um sinal, este é reconstruído a partir dos coeficientes *wavelet* (HE, 2006).

A transformada inversa de *Wavelet* (síntese) consiste em percorrer o caminho inverso na seguinte sequência: primeiramente, realiza-se a sobre-amostragem dos vetores de coeficientes, intercalando um valor de zero entre os coeficientes. Em

seguida, aplicam-se os filtros passa baixas L' e passa-altas H' aos coeficientes cAn e cDn respectivamente (onde n é o nível do detalhamento da transformada). Finalmente, somam-se os vetores resultantes. A Figura 14 exemplifica a transformada *wavelet* direta e a transformada inversa. Ambas feita para um nível de detalhamento igual a 2. Desta maneira, $cA2$ passa por um filtro passa-baixas L' e $cD2$ passa por um passa-altas H' e em seguida são somados. O resultado cria $cA1$ que passará novamente por um L' enquanto o $cD1$ está passando por um H' . Em seguida eles são somados e o resultado é o vetor original s . Os filtros de síntese só serão iguais ao de análise se a família *wavelet* utilizada for ortonormal (DINIZ et al., 2004).

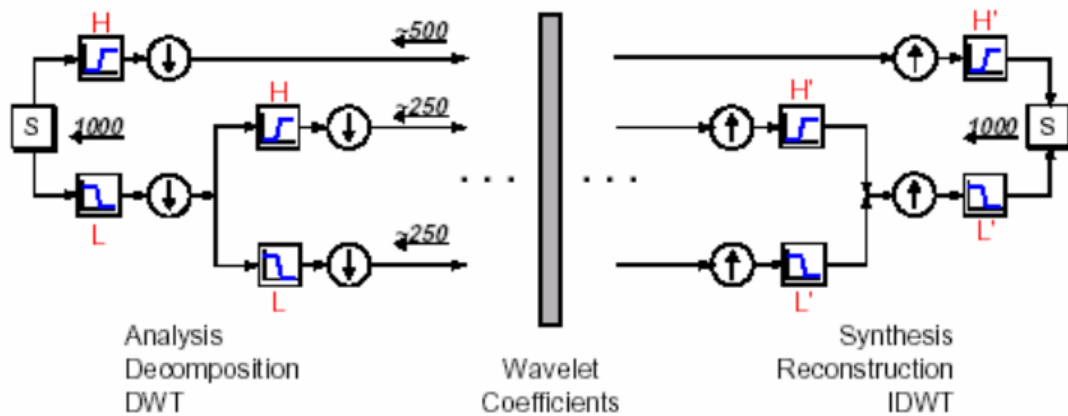


Figura 14 - O Processo da esquerda ilustra uma decomposição (análise) e da direita uma reconstrução (síntese). Adaptado de (MISITI et al., 2006).

3.6.6 REMOÇÃO DE RUÍDOS COM WAVELETS (DENOISING)

O sinal $x(n)$ às vezes é alterado com a adição de um ruído $r(n)$ gerando assim o sinal corrompido $y(n)$.

$$y(n) = x(n) + r(n) \quad (10)$$

A idéia da *wavelet denoising* é tentar transformar o sinal $y(n)$ o mais próximo possível de $x(n)$ ou pelo menos em um sinal similar $x'(n)$. A figura 15 mostra esse processo.

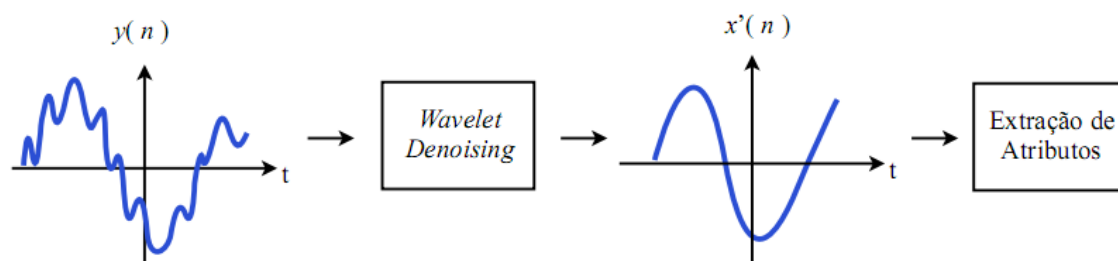


Figura 15 – processo de remoção de ruídos denoising.

Realizar a extração de ruído no domínio do tempo é praticamente inviável, pois nesse domínio é difícil diferenciar o que é sinal do que é ruído. Portanto, a filtragem é realizada no domínio da frequência. Um das maneiras de realizar a filtragem seria aplicar a transformada discreta de Fourier no sinal, como ela é linear tem-se:

$$Y(k) = X(k) + R(k) \quad (11)$$

Tipicamente $R(k)$ ocupa uma faixa larga que se sobrepõem à faixa estreita de $X(k)$, então somente a aplicação de um filtro de faixa é suficiente.

Mas infelizmente há casos em que os espectros do sinal e do ruído estão sobrepostos, tornando a separação sinal/ruído por filtragem difícil, já que pode ocorrer a retirada de parte do sinal e não somente do ruído. Portanto é necessária outra operação que consiga realizar a separação do sinal/ruído sem retirar as características do sinal. Para isso, Donoho (1995) propôs a utilização da transformada *wavelet* para a remoção de ruídos, técnica conhecida como *denoising*. Esta técnica é aplicada ao vetor de coeficientes de detalhes resultantes da transformada *Wavelet*. O procedimento consiste em escolher quais desses coeficientes serão mantidos para preservar a informação do sinal, retirando, então, os coeficientes relacionados ao ruído (DONOHO, 1995). Existem duas formas básicas de fazer a escolha dos coeficientes. A primeira é a denominada *hard*

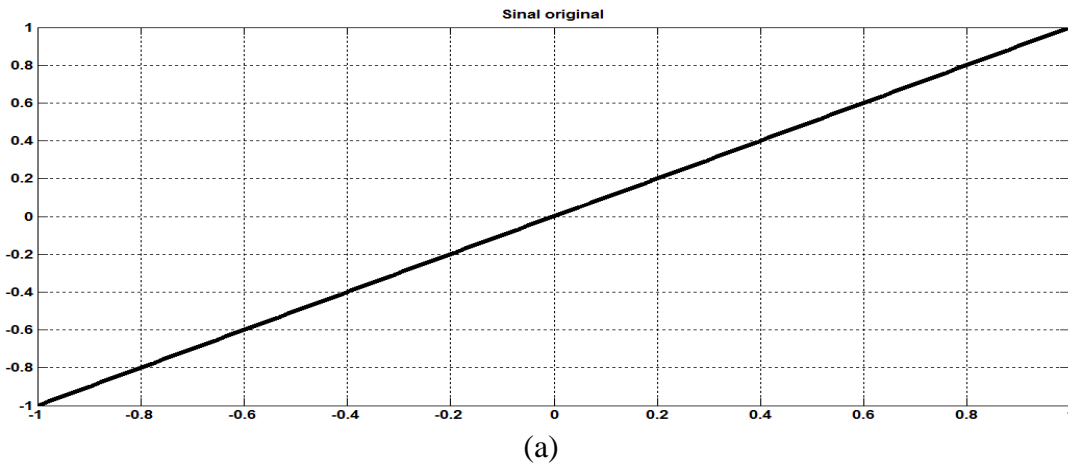
thresholding e a segunda, de *soft thresholding*. Para o *hard thresholding* os coeficientes do detalhe cD_{jk} cujos módulos estão abaixo do limiar são zerados.

$$c'D_{jk} = \begin{cases} cD_{jk}, & |cD_{jk}| \geq \lambda \\ 0, & \text{caso contrário} \end{cases} \quad (12)$$

Já para o *soft thresholding*, que é considerado uma extensão do *hard thresholding*, além de se zerar os coeficientes cujos módulos estão abaixo do limiar também subtrai-se o valor do limiar dos coeficientes com módulos acima do limiar, mantendo o seu sinal, mas atenuando-os.

$$c'D_{jk} = \begin{cases} \text{sgn}(cD_{jk})(|cD_{jk}| - \lambda), & |cD_{jk}| \geq \lambda \\ 0, & \text{caso contrário} \end{cases} \quad (13)$$

A Figura 16 ilustra a remoção de coeficientes através do *hard thresholding*, e através do *soft thresholding*. Os coeficientes dos detalhes estão normalizados e o limiar $\lambda=0.4$.



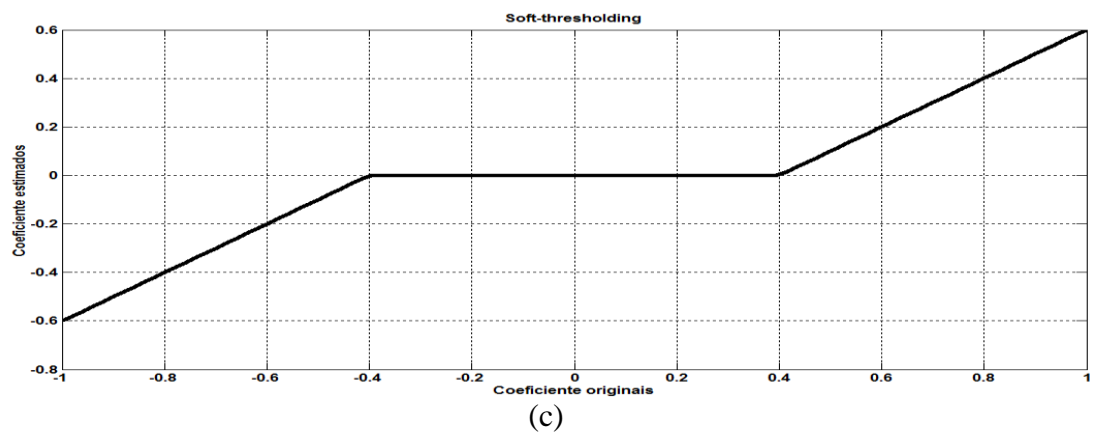
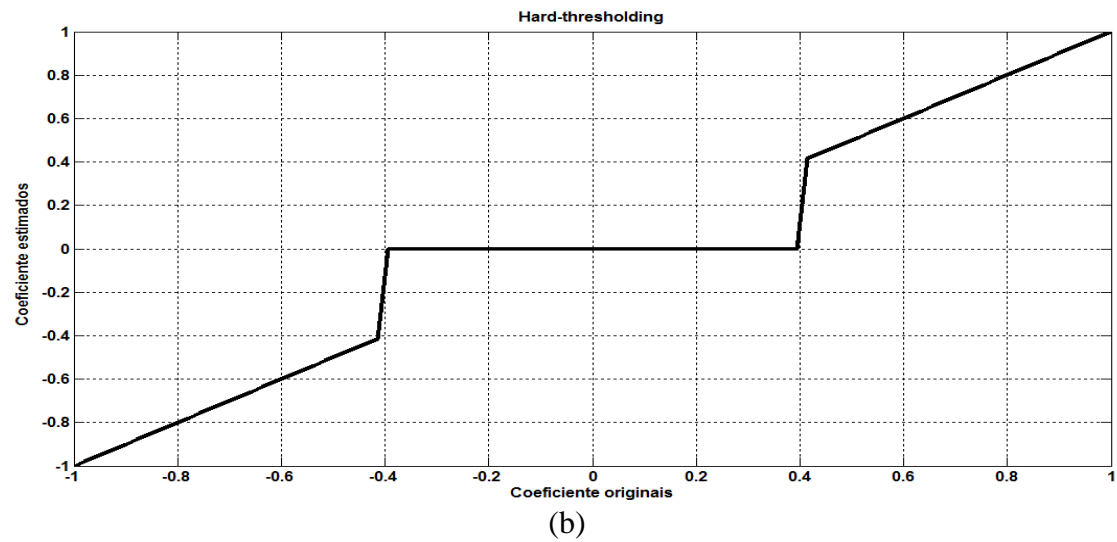


Figura 16 – (a) Sinal original, (b) hard-thresholding e (c) soft-thresholding.

A figura 17 mostra um sinal ruidoso e o mesmo sinal com os ruídos eliminados após o processo de *denoising*.

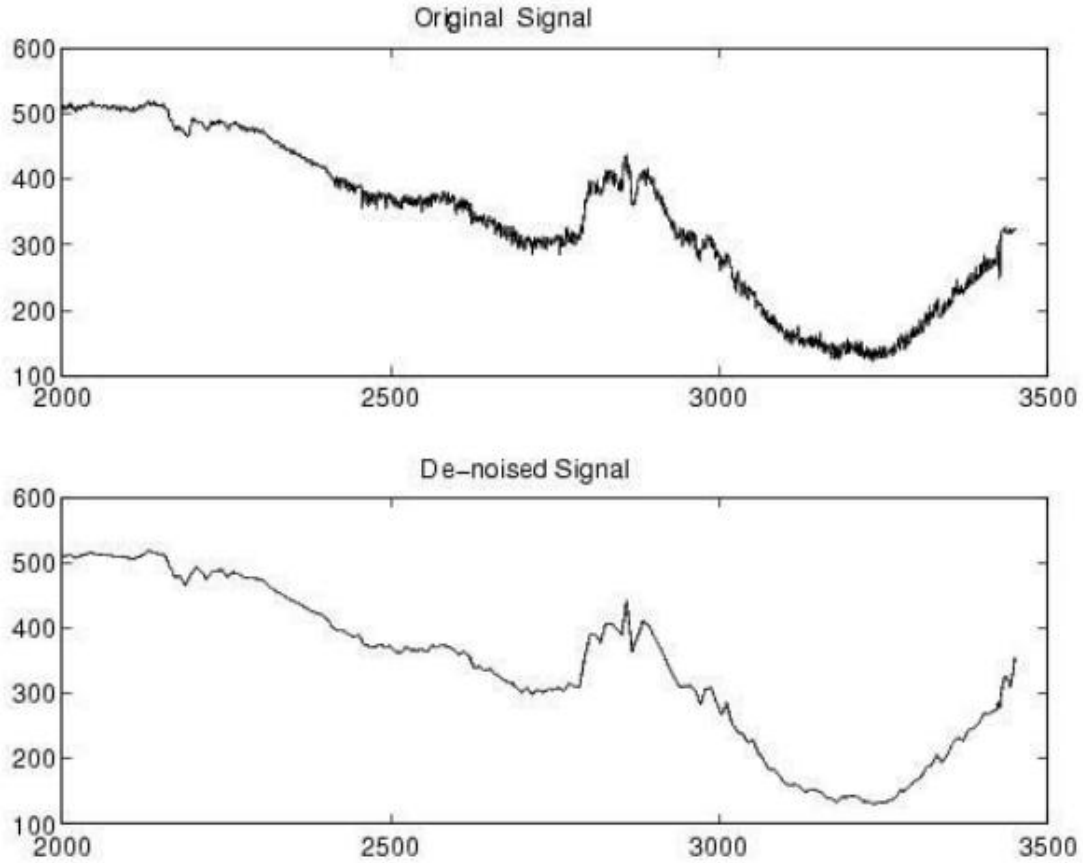


Figura 17 – Sinal com ruído e sinal após o processo denoising (Matlab, 2006).

3.6.7 A ESCOLHA DE LIMIAR PARA A REMOÇÃO DE RUÍDOS

A escolha do limiar é feita baseado numa generalização do método SURE (*Stein's Unbiased Estimate of Risk*) feito por Donoho (1995). Este método considera o termo $r(t)$ como sendo um ruído gaussiano de média zero e variância unitária. Nessa situação, podemos estimar o limiar como:

$$\lambda = \sigma \sqrt{2 \log(n)} \quad (14)$$

Onde n é o número de amostras do sinal, $\sigma = MAD/0.6745$, MAD representa a mediana dos valores absolutos dos coeficientes do detalhe da transformada *wavelet* no seu nível de maior detalhe:

$$MAD = \text{mediana}(cD_{1,1}) \quad (15)$$

onde $cD_{1,1}$ é o detalhe no nível 1, que contem as altas frequências do sinal. Espera-se que os coeficientes associados estejam mais presentes nesse nível.

Então a remoção de ruídos com *wavelet* pode ser sintetizada nos seguintes passos:

- Faz-se a Transformada *Wavelet*;
- Escolhe-se um limiar λ para remoção de ruído a ser calculado a partir dos coeficientes do vetor de detalhes;
- Os coeficientes que representariam o ruído são zerados e os coeficientes que não representa os ruídos são ajustados, de acordo com o *soft thresholding* ou *hard-thresholding*;
- Faz-se a Transformada Inversa de *Wavelet*.

3.7 CARACTERÍSTICAS PARA O RECONHECIMENTO

Extração de características, como em qualquer problema de reconhecimento de padrões, é uma etapa muito importante para a análise de áudio e tarefas de processamento. Neste capítulo algumas características importantes de áudio, que podem ser usadas para métodos de classificação e segmentação (GIANNAKOPOULOS, 2009), são apresentados.

Tazanetakis (2002) propõe em seu trabalho a utilização das características taxa de cruzamento em zero (ZCR), frequência *Rolloff*, fluxo espectral e centróide para realizar a classificação de discursos (diferenciando por sexo, masculino ou feminino) e de estilos musicais (Rock, jazz, Country, etc.). Giannakopoulos (2009) utilizou em seu trabalho, além das mesmas características, a energia e a entropia, para avaliar discurso, chuva e música, observando que estas características variam para cada uma delas e que

podem ser utilizadas para a classificação. Conceição (2012) utilizou faixa de frequência (frequência mínima e máxima do canto) e frequência mais intensa (definido como a frequência onde a Densidade Espectral de Potência (PSD) máxima foi encontrada) para o reconhecimento de aves da família Tinamidae. Nesse trabalho foi utilizado o método do vizinho mais próximo para realizar o reconhecimento. Em todos esses trabalhos os resultados foram consideráveis, mostrando a eficiência dessas características para resolver os problemas propostos. Isso motivou a utilização dessas características neste trabalho.

Neste trabalho serão, portanto, estudadas as características taxa de cruzamento em zero (ZCR), frequência *Rolloff*, fluxo espectral, centróide, energia, entropia e frequência máxima, mínima e mais intensa para serem utilizadas na classificação de espécies de aves, verificando a eficiência das mesmas no reconhecimento.

As características utilizadas são extraídas do sinal no domínio do tempo ou no domínio da frequência, sendo:

- Domínio do tempo:
 - 1) Energia;
 - 2) Entropia;
 - 3) Taxa de cruzamento em zero.
- Domínio da frequência:
 - 1) Centróide;
 - 2) Frequência *Rolloff*;
 - 3) Fluxo espectral;
 - 4) Faixa de frequência e frequência mais intensa.

3.8 PROCESSAMENTO PARA EXTRAÇÃO DE CARACTERÍSTICAS DE ÁUDIO

Seja $x(n)$, $n = 1, 2, \dots, L$, a amostra de sinal de áudio e L o comprimento do sinal. Para o cálculo das características como taxa de cruzamento em zero (ZCR), frequência *Rolloff*, fluxo espectral, centróide, energia e entropia do sinal de áudio $x(n)$, é necessário adotar uma técnica de processamento que trabalhe com o sinal em um curto intervalo de tempo (*short-term*) devido ao fato do sinal de áudio não ser estacionário, ou seja, as suas propriedades variam com o tempo (HAYES, 2006) e (THEODORIDIS; KOUTROUMBAS, 2009).

Para resolver tal problema será utilizada a técnica de janelamento (*windowing*): o sinal de áudio é dividido em superposição (*overlapping* ou *non-overlapping*) de pequenos intervalos (curto prazo), as chamadas janelas ou quadros (*frames*) e o cálculo da característica é executado para cada *frame*. Assim, durante o intervalo de tempo de um quadro, o sinal pode ser considerado como estacionário, o que permite o cálculo dos valores das características.

Seja $w(n)$ uma janela de N amostras. A janela mais simples para se utilizar é a janela retangular que é descrita de acordo com a equação (16):

$$w(n) = \begin{cases} 1, & 0 \leq n \leq N - 1 \\ 0, & \text{caso contrário} \end{cases} \quad (16)$$

O processo de janelamento do sinal original é equivalente à multiplicação do sinal pela janela $w(n)$ deslocada em n . Portanto a amostra do i -ésimo quadro é descrito como:

$$x(n') = x(n)w(n - m_i) \quad (17)$$

onde m_i é o deslocamento da janela no i -ésimo quadro. O valor de m_i depende do tamanho da janela e do tamanho do passo. Portanto, deve-se considerar que o tamanho da janela deve ser grande o de maneira a ter dados suficientes para o cálculo das características e pequeno o suficiente de forma a garantir a estacionariedade do sinal (OPPENHEIM; SCHAFER, 1999) e (HAYES, 2006).

Na literatura os tamanhos das janelas geralmente variam entre 10-50 ms, enquanto o passo está associado ao nível de superposição. Neste trabalho adotou-se janelas de 50 ms para a segmentação dos sinais e 20 ms para a determinação de características sem superposição, isto é, adotou-se o passo igual ao tamanho da janela. Na segmentação adotou-se maiores janelas de tempo devido ao sinal original ter maior duração, chegando a mais de minuto. Para a extração do valor das características foi utilizado 20 ms, pois neste caso o som é somente um segmento do canto (“chamada”) da ave, tendo durações pequenas e com esse valor foi possível adquirir um valor confiável para as características, uma vez que o sinal se mantém estacionário para este tamanho de janela.

A figura 18 mostra o janelamento de um sinal com uma janela de 200 amostras e passo de 100 amostras, com *overlapping* de 50%.

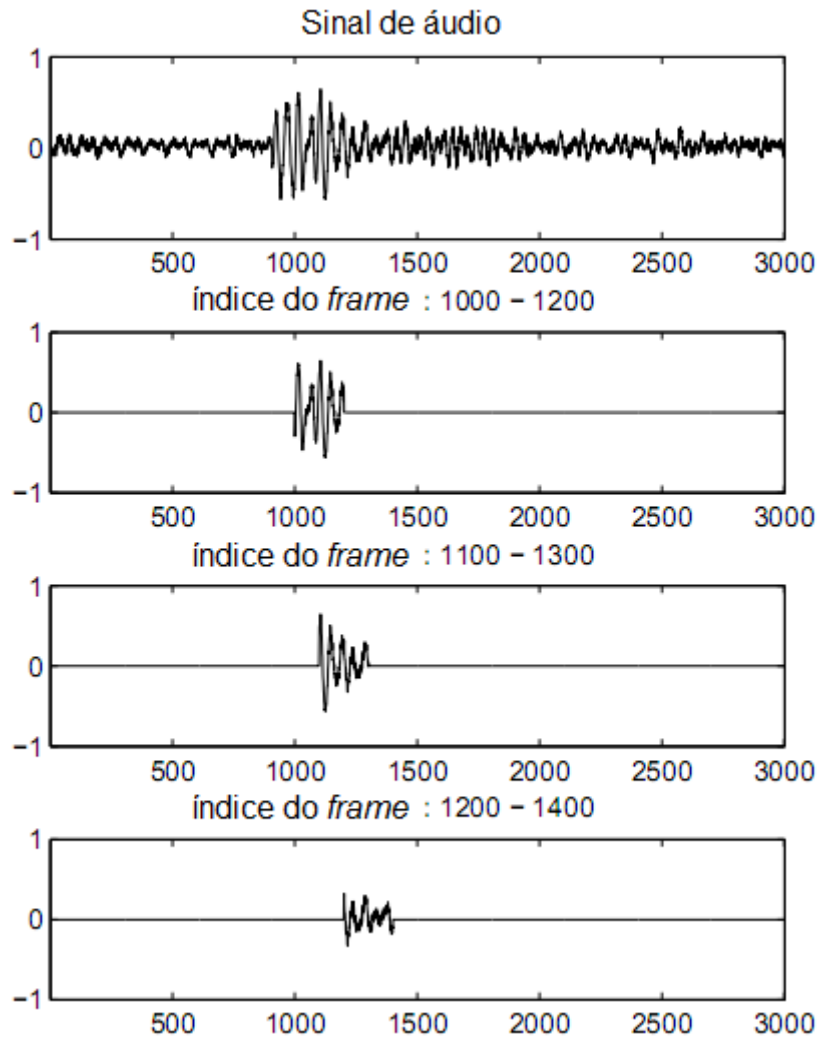


Figura 18 – Processo de janelamento para um sinal de áudio (GIANNAKOPOULOS, 2009).

O número de quadros a ser utilizado é definido pela equação (18):

$$M = \left\lceil \frac{L - S}{N} \right\rceil + 1 \quad (18)$$

onde N é o número de amostra da janela, S o número de amostra do passo e L o tamanho do sinal.

3.8.1 CARACTERÍSTICAS DO DOMÍNIO DO TEMPO

Algumas características utilizadas no reconhecimento são extraídas do domínio do tempo, ou seja, elas são retiradas diretamente das amostras do sinal. Geralmente são

representações simples das mudanças de sinal de energia (GIANNAKOPOULOS, 2009).

3.8.1.1 ENERGIA

Sendo $x_i(n)$ a amostra do sinal de áudio com $n = 1, 2, \dots, N$, N é o tamanho da janela, a energia média E do sinal é calculada para cada quadro de acordo com a equação (19).

$$E = \frac{1}{N} \sum_{n=1}^N |x_i(n)|^2 \quad (19)$$

Esta característica simples pode ser usada para a detecção de períodos de silêncio em sinais de áudio e também para a discriminação entre as classes de áudio. Observa-se na equação que a energia depende somente do sinal e de seu tamanho.

Na figura 19 observa-se que a energia é maior na região de canto e nula na região de silêncio. Entretanto na figura 20 a energia também é maior na região de canto, porém na região de silêncio ela não ficou nula. Isso se deve ao fato do sinal estar ruidoso, mas como a amplitude do ruído é menor que a do canto, a energia é maior no local do canto.

Pode ainda ser observado nas figuras 19 e 20 que como a energia se difere em valores da região de canto para a região de silêncio, essa característica é útil para a segmentação do sinal. Contudo se o sinal for muito ruidoso, de forma que a energia seja maior fora da região de canto, esta técnica não será vantajosa, por isso a grande necessidade da filtragem do sinal.

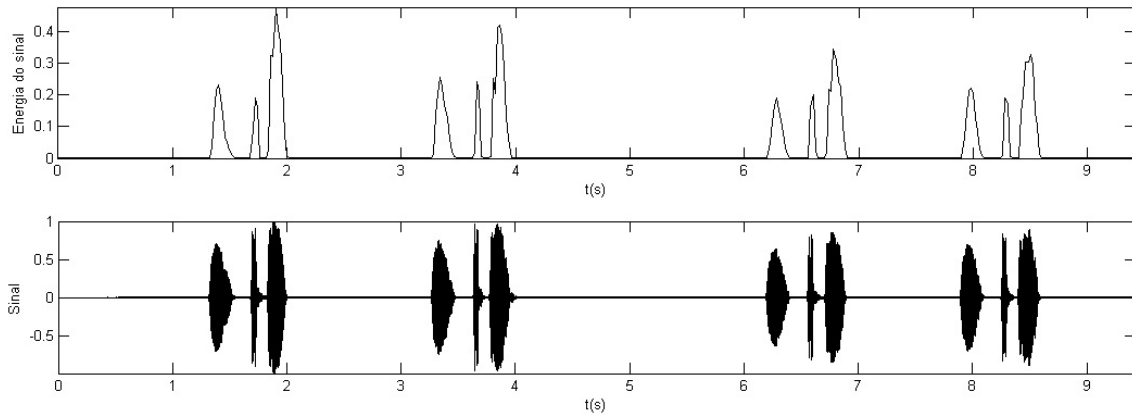


Figura 19 - Energia do som de um Fogo-apagou.

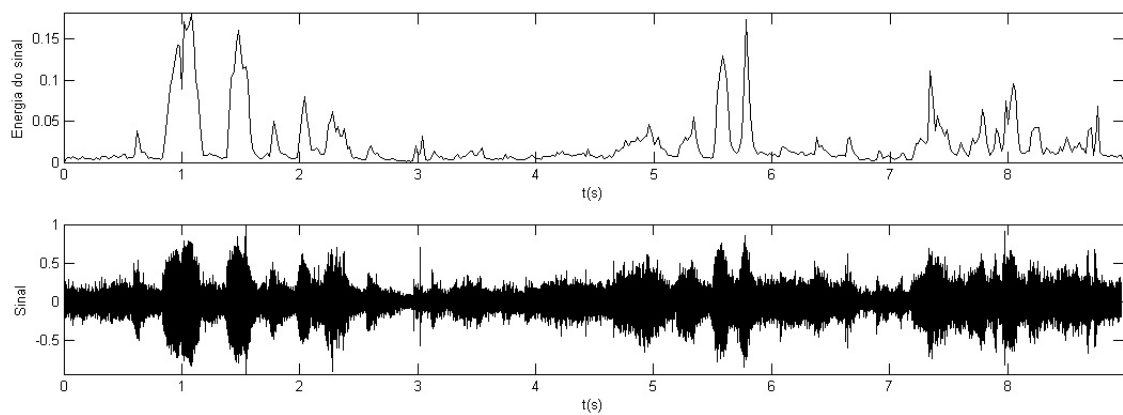


Figura 20 – Energia do som de um Peixe-frito.

3.8.1.2 TAXA DE CRUZAMENTO EM ZERO (ZCR)

Consiste em um método de análise no domínio do tempo, que corresponde à medida da quantidade de mudanças da amplitude do sinal de áudio de positivo para negativo e vice-versa.

O ZCR é definido pelo número de cruzamentos de zero no domínio do tempo em uma janela de processamento, dividido pelo número de amostras na janela, como apresentado na equação (20), onde N é o número total de amostras em uma janela de processamento do sinal $x_i(n)$.

$$ZCR = \frac{1}{2N} \sum_{n=1}^N |sgn[x_i(n)] - sgn[x_i(n-1)]| \quad (20)$$

onde $sgn\{.\}$ é a função sinal:

$$sgn[x(n)] = \begin{cases} 1, x(n) \geq 0 \\ -1, x(n) < 0 \end{cases} \quad (21)$$

A taxa de cruzamento de zero está correlacionada ao conteúdo de frequência de um sinal. O valor do ZCR multiplicado pela metade da frequência de amostragem ($f_s/2=22050$) nos fornece a frequência principal (FP) do som que seria a frequência mais intensa calculada por Conceição (2012).

A Tabela 2 nos mostra o valor médio do ZCR, frequência principal ($FP = 22050 * ZCR$), a faixa de frequência e a frequência mais intensa para os cantos de todas as aves estudadas neste trabalho. É importante salientar que a FP possui valor bem próximo de FI para todas as espécies e que quanto maior a frequência (FI) do sinal maior o ZCR.

Tabela 2- ZCR, Frequência principal (FP), mínima (Fmín), máxima (Fmáx) e mais intensa (FI).

Aves	ZCR	FP	Fmín	Fmáx	FI
Acauã	0.0707	1559	1033	2153	1550
Bem-te-vi	0.1345	2966	2326	4048	3198
Fim-fim	0.1734	3823	3273	4306	3810
Fogo-apagou	0.0569	1255	430	1378	859
Irerê	0.1804	3978	2240	7150	3633
Jaó	0.0596	1314	775	1722	1275
Neinei	0.2042	4503	3531	5426	4635
Peixe-frito	0.1024	2258	1895	2842	2296
Pitiguari	0.1099	2423	2584	3876	3390
Saracura	0.0663	1462	689	2153	1334
Teque-teque	0.302	6659	4187	7062	6329
Tiziu	0.2848	6280	5857	9365	6365

3.8.1.3 ENTROPIA

A entropia é um recurso a ser utilizado se o sinal de áudio apresentar mudanças bruscas no nível de energia. Inicialmente, as janelas i de tamanho N são subdivididas em sub-janelas j de tamanho K . Para cada sub-janela j , a energia e_j é normalizada de acordo com a equação (22):

$$e_j^2 = E_{sub-janela_j} / E_{janela_i} \quad (22)$$

onde $E_{sub-janela_j}$ é a entropia da sub-janela j e E_{janela_i} é a entropia da janela principal.

Em seguida, a entropia da janela i é calculada utilizando-se a equação (23) (GIANNAKOPOULOS, 2009):

$$Z(i) = - \sum_{j=1}^k e_j^2 \log_2(e_j^2) \quad (23)$$

A entropia de um quadro de áudio é mais baixa se houver alterações bruscas presentes nesse quadro de áudio. Portanto, ela pode ser usada para a discriminação de mudanças abruptas de energia, por exemplo, nos cantos das aves.

3.8.2 CARACTERÍSTICAS DO DOMÍNIO DA FREQUÊNCIA

No domínio da frequência as características são obtidas do espectro do sinal, que é produzido pela transformada de Fourier.

3.8.2.1 CENTROIDE

Os centroides espectrais (centro de gravidade do espectro do áudio) são calculados ao longo do sinal. Este recurso é utilizado para trabalhar com características

geométricas do espectro (GIANNAKOPOULOS, 2009). Seu cálculo é feito pela equação (24).

$$C(i) = \frac{\sum_{k=1}^N k \cdot X_i(k)}{\sum_{k=1}^N X_i(k)} \quad (24)$$

onde $X_i(k)$, $k=1,...,N$ é a FFT do i -ésimo quadro e N é o comprimento do quadro.

Esta característica, juntamente com a energia, pode ser utilizada para a segmentação do sinal. As figuras 21 e 22 mostram o centroide para o som de um Fogo-apagou e de um Peixe-frito, respectivamente. Observa-se na figura 22 que se o sinal não tiver nenhum ruído o centroide do canto será maior que o da região de silêncio, porém se tiver algum ruído o centroide sofrerá grandes variações, não se distinguindo as duas regiões. Isso pode ser justificado pelo fato do ruído possuir as mais variadas frequências, e como o centroide é calculado utilizando o espectro essas frequências interferem consideravelmente no comportamento do mesmo.

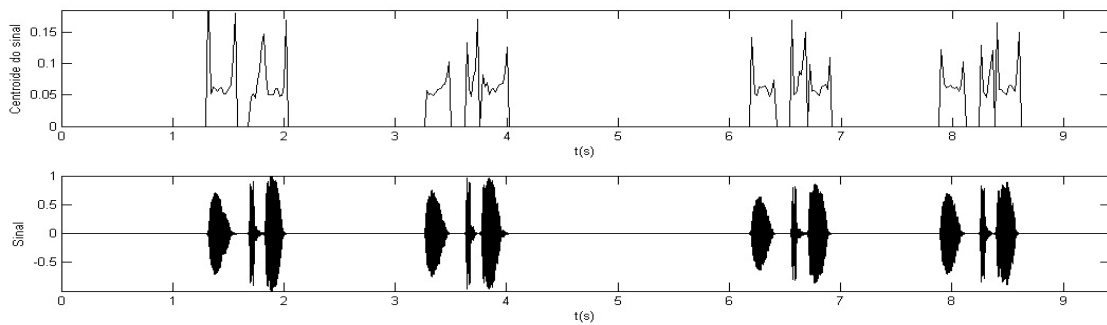


Figura 21– Centróide do som de um Fogo-apagou.

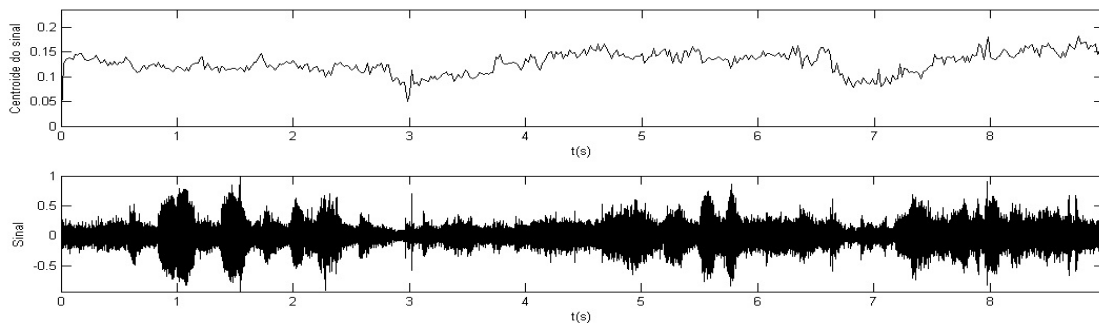


Figura 22– Centroide do som de um Peixe-frito.

3.8.2.2 FLUXO ESPECTRAL

Esta característica é uma medida da variação do espectro local entre quadros sucessivos. É definida como o quadrado da diferença entre a magnitude normalizada do espectro de dois quadros sucessivos, como mostra a equação (25):

$$Fl_{(i,i-1)} = \sum_{k=1}^N (EN_i(k) - EN_{i-1}(k))^2 \quad (25)$$

onde $EN_i(k)$ e $EN_{i-1}(k)$ são as magnitudes normalizadas da Transformada de Fourier no quadro atual i e no quadro anterior $i-1$, respectivamente.

3.8.2.3 FREQUÊNCIA ROLLOFF

Frequência *Rolloff* (R_t) é definido como a frequência abaixo da qual determinada percentagem (geralmente 90%) de distribuição da magnitude do espectro é concentrada. Conforme mostra a equação 26 a seguir:

$$\sum_{k=1}^{R_t} X_i(k) = C \sum_{k=1}^N X_i(k) \quad (26)$$

onde C é a percentagem adotada. Neste trabalho utiliza-se $C = 90\%$.

3.9 SEGMENTAÇÃO

A segmentação é necessária para realizar a separação entre a região de som da região de silêncio, para que assim as características do reconhecimento sejam extraídas somente do canto. Ela também é útil para a detecção da duração do tempo de canto. Neste trabalho foi utilizado um método proposto por Giannakopoulos (2009) que utiliza

as características energia e centroide para realizar a segmentação. Esse método foi comparado com o método utilizado por Conceição (2012) que ele utiliza a varredura da matriz Densidade Espectral de Potência (PSD) para realizar a segmentação. Observou-se nessa comparação que o método aqui utilizado se mostrou mais eficiente para a segmentação. As figuras 23 e 24 mostram a segmentação utilizando o método da varredura da matriz de PSD.

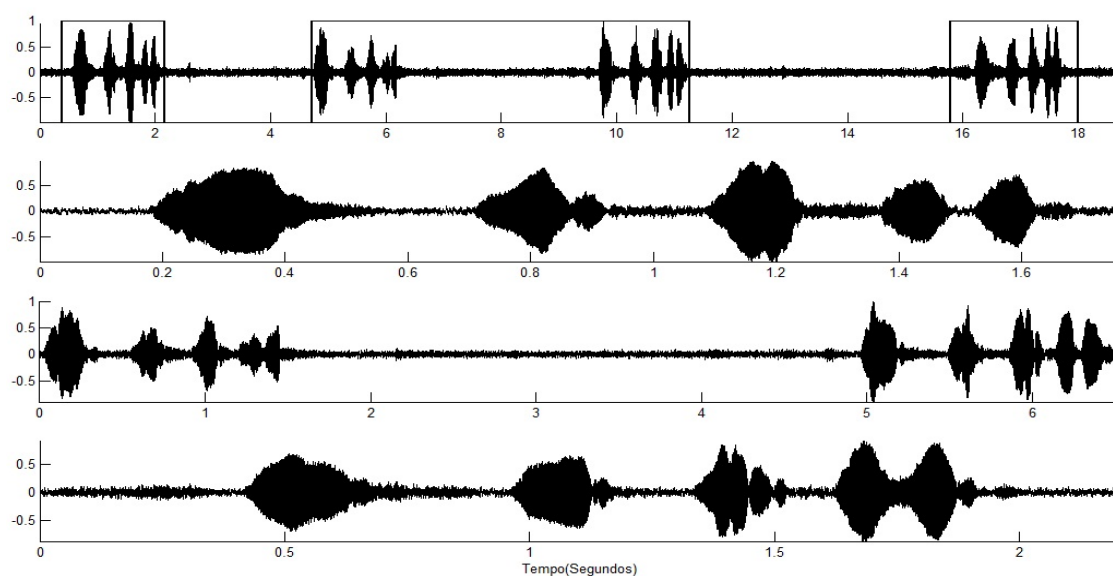


Figura 23 – Segmentação do som de um Peixe-frito, sem pré-filtragem.

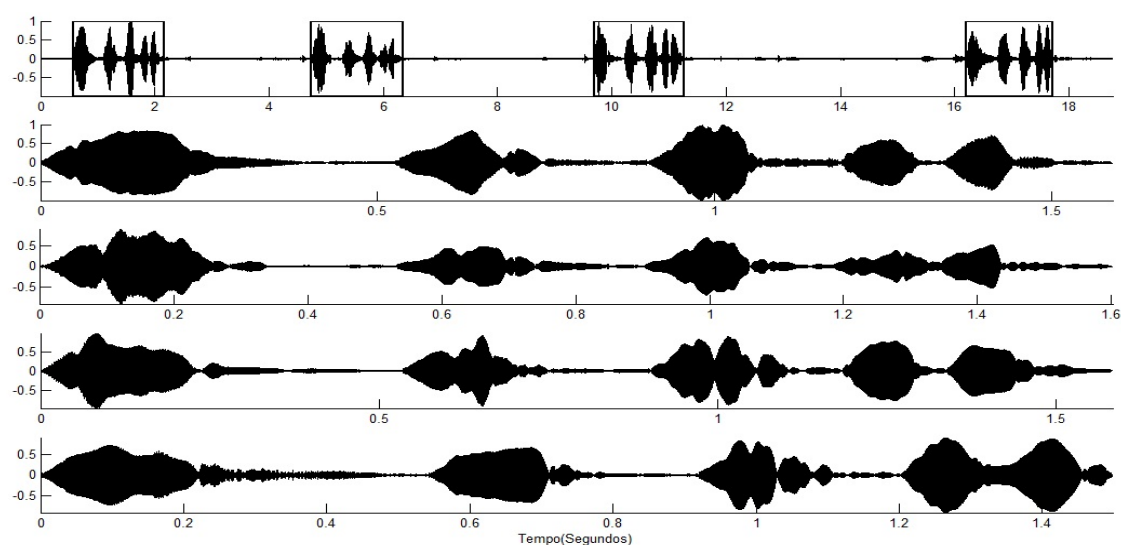


Figura 24 - Segmentação do som filtrado de um Peixe-frito.

Ao analisar as figuras 23 e 24 observa-se que se o sinal não estiver bem filtrado a segmentação não será realizada corretamente uma vez que a varredura da matriz de PSD se utiliza da análise da frequência para realizar essa segmentação. Portanto, se o sinal não estiver com o mínimo de ruído possível a segmentação não será realizada corretamente.

As figuras 25 e 26 mostram o resultado da segmentação utilizando o método do centróide e da energia. Pode ser observado que mesmo que o sinal não esteja completamente filtrado (Fig. 25) a segmentação é realizada de maneira correta.

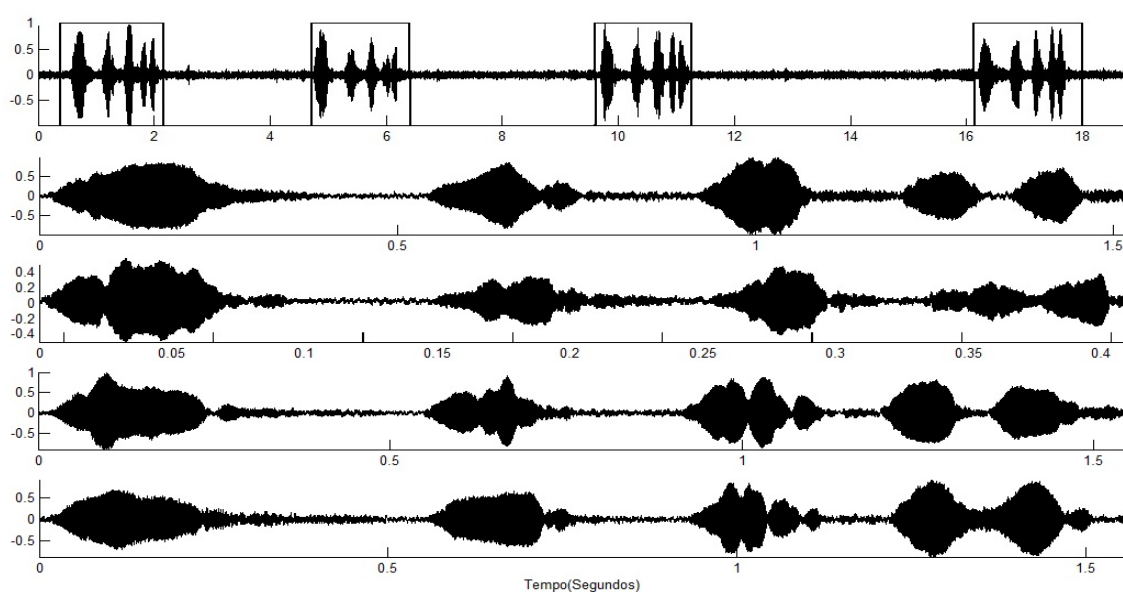


Figura 25 - Segmentação do som de um Peixe-frito utilizando centróide e energia.

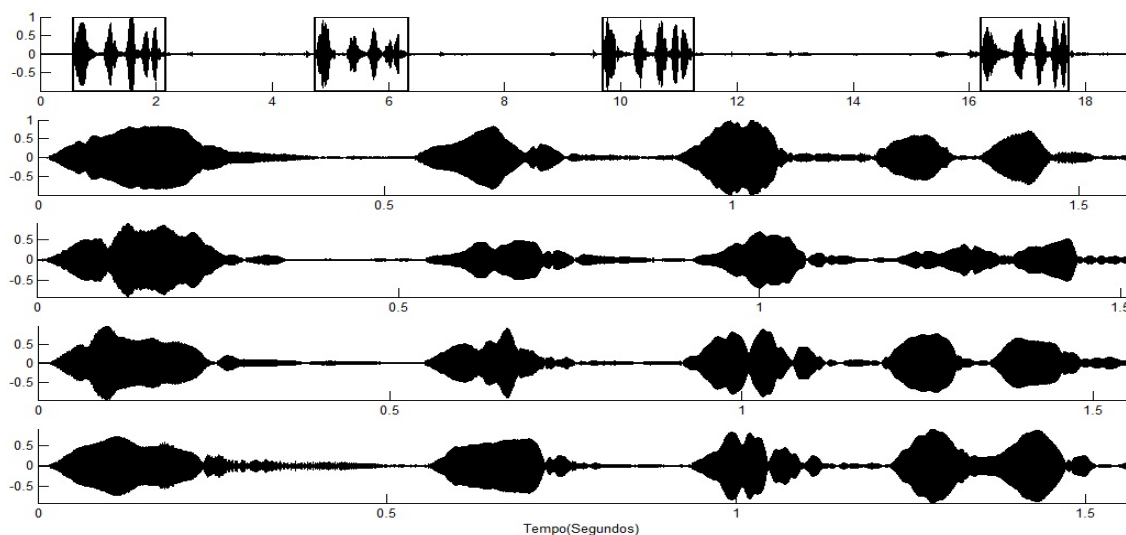


Figura 26 – Segmentação do som filtrado de um Peixe-frito utilizando centróide e energia.

No caso da segmentação do sinal filtrado a segmentação continua a mesma (Fig. 26). Portanto a segmentação utilizando a energia e o centroide é mais eficiente neste caso, pois com ela não há a necessidade do sinal estar bem filtrado.

O método utilizando a energia e o centroide é realizado utilizando os seguintes passos (GIANNAKOPOULOS, 2009):

As características energia e centroide são extraídas do sinal, conforme (seção 3.9.1.1 e 3.9.2.1). A janela e o passo utilizados foram de 50 ms, pois a segmentação com 20 ms forneceu erros de segmentação, conforme mostra a figura 27.

- 1) Para cada característica são estimados os limites que definem a região de canto e de silêncio.
- 2) As sequências das características são calculadas e as mesmas são comparadas com os limiares propostos. Para este fim, o seguinte processo é realizado, para cada sequência de característica:
 - a) Calcula-se o histograma de cada característica;
 - b) Detectam-se os valores de máximos do histograma de cada característica;

- c) Sendo M_1 e M_2 as posições do primeiro e segundo valores máximos locais, respectivamente, calcula-se o limiar de cada característica de acordo com a equação (27), onde W é um valor escolhido pelo usuário e tem como objetivo melhorar o valor de limiar, pois grandes valores de W desloca o limiar para mais próximo de M_1 . Neste trabalho foram realizados vários testes e o melhor valor encontrado para W foi 5.

$$T = \frac{WM_1 + M_2}{W + 1} \quad (27)$$

Após encontrar o limiar as seguintes análises são realizadas:

$$y(i) = \begin{cases} 1, & E \geq T \\ 0, & E \leq T \end{cases} \quad (28)$$

$$x(i) = \begin{cases} 1, & C \geq T \\ 0, & C \leq T \end{cases} \quad (29)$$

onde E é a sequência referente à energia e C é a sequência referente ao centróide, em que 0 e 1 representam que a característica tem valor menor ou maior que o limiar, respectivamente. Em seguida $y(i)$ e $x(i)$ são comparados.

$$z(i) = \begin{cases} 1, & y(i) \text{ e } x(i) = 1 \\ 0, & \text{Caso contrário} \end{cases} \quad (30)$$

Nesse caso 1 representa região de som e 0 região de silêncio.

Observa-se que esse método é eficiente, pois em um sinal a energia tem que ser maior na região de canto que na de silêncio e se isso não ocorrer é devido ao sinal estar ruidoso. O centróide, porém, é quase sempre maior na região de canto que na região de silêncio. Isso só não irá acontecer se o ruído prevalecer muito sobre o sinal e neste caso esse sinal é eliminado do banco de dados.

A figura 27 mostra um sinal segmentado com janela e passo de 20 ms em que a segmentação não ocorreu da forma correta. Nessa figura pode-se observar que a

segmentação do sinal ocorreu dentro do canto da ave, pois tem somente dois cantos e a segmentação forneceu 6 valores, isto é cada canto foi segmentado em 3 partes.

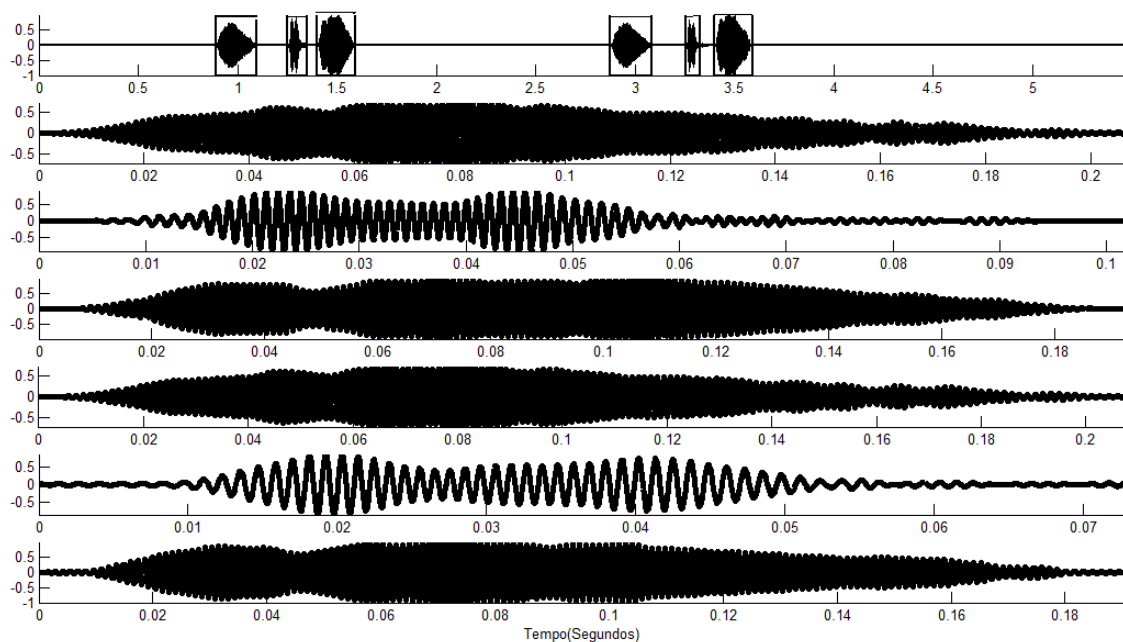


Figura 27 – Segmentação com janela e passo de 20 ms.

A figura 28 mostra o resultado da segmentação utilizando janela e passo de 50 ms. Nela observa-se que na região de canto a energia e o centroide são maiores do que na região de silêncio e que nessa região seus valores são maiores que o valor limiar.

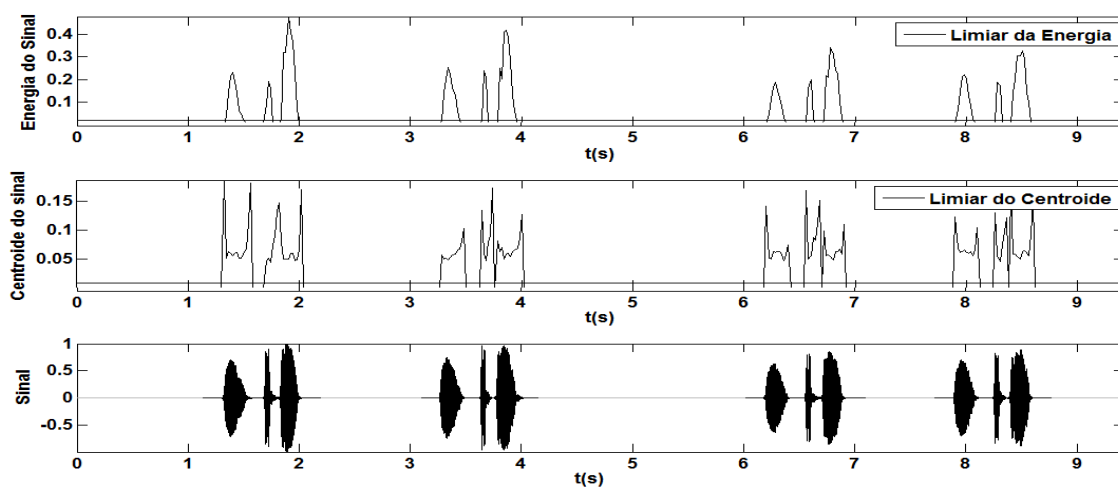


Figura 28 – Segmentação do canto de um Fogo-apagou. Energia do sinal, Centroide do sinal e sinal.

3.10 MÉTODO DE RECONHECIMENTO

Neste trabalho foi utilizado como método de reconhecimento a árvore de decisão. Árvores de decisão são modelos estatísticos que utilizam um treinamento supervisionado para a classificação e previsão de dados. Em outras palavras, em sua construção é utilizado um conjunto de treinamento formado por entradas e saídas. Desta maneira elas podem ser utilizadas para dar ao agente a capacidade de aprender e assim poder tomar decisões.

A ideia de aprendizado é que os dados (características de reconhecimento) sejam analisados, para após essa análise a árvore seja confeccionada. O aprendizado ocorre na medida em que o agente observa suas interações com o mundo e seu processo interno de tomada de decisões. Aprendizado de árvores de decisão é um exemplo de aprendizado indutivo: cria uma hipótese baseada em instâncias particulares que geram conclusões gerais.

A árvore de decisão é uma estrutura muito usada na implementação de sistemas especialistas e em problemas de classificação. Elas tomam como entrada uma situação descrita por um conjunto de atributos e retorna uma decisão, que é o valor predito para o valor de entrada.

As árvores de decisão constituem uma técnica muito poderosa e amplamente utilizada em problemas de classificação. Uma das razões para que esta técnica seja bastante utilizada é o fato do conhecimento adquirido ser representado por meio de regras. Essas regras podem ser expressas em linguagem natural, facilitando o entendimento por parte das pessoas. Em termos computacionais as árvores de decisões são simples para serem geradas e relativamente rápidas no processo de classificação.

As vantagens de utilizar são:

- É um método de decisão simples, o que torna sua interpretação bem fácil;
- Tem facilidade de interpretar e trabalhar com qualquer tipo de informação (Real, nominal, ordinal e etc.);
- Escolha automática dos atributos (características) mais relevantes em cada caso. Atributos mais relevantes aparecem mais acima na árvore;
- Separa de forma simples e rápida as classes, mostrando quais características são mais relevantes para o reconhecimento da espécie.

Como qualquer método as árvores de decisão também possuem suas desvantagens que são:

- Sensibilidade a pequenas perturbações no conjunto de treino, acabam gerando redes muito diferentes, ou seja, uma escolha adequada para o treinamento é de suma importância;
- Uma grande quantidade de dados pode gerar uma alta superposição de dados não permitindo uma separação das espécies na fase de treinamento.

A figura 29 mostra graficamente o modelo de uma árvore de decisão. Nele podemos observar a raiz que seria a entrada da árvore e os nós de decisões onde ocorre a classificação e as folhas que são o resultado da classificação.

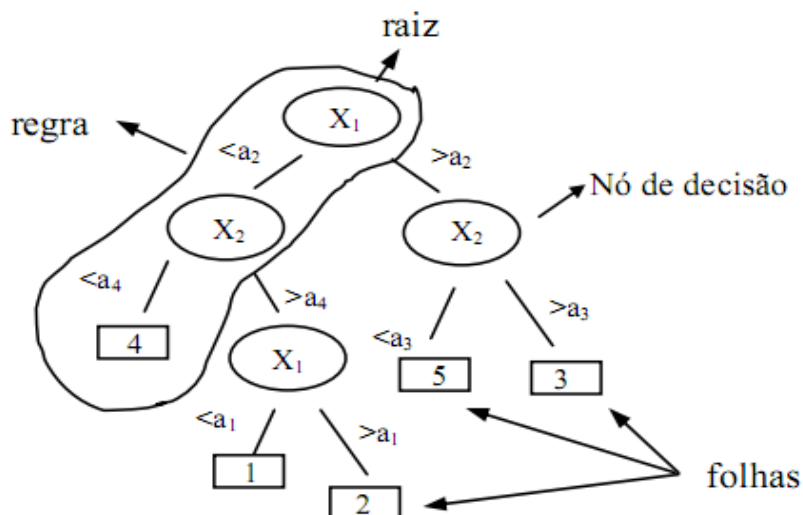


Figura 29 – Modelo Modelo de Árvore de Decisão (GAMA, 2004).

O algoritmo utilizado para a construção da árvore foi uma adaptação do método de CART (*Classification and Regression trees*). O algoritmo CART (*Classification and Regression Trees*) foi proposto em BREIMAN et al. (1984) e consiste de uma técnica não-paramétrica que induz tanto árvores de classificação quanto árvores de regressão, dependendo se o atributo é nominal (classificação) ou contínuo (regressão).

- Dentre as principais virtudes do CART está a grande capacidade de pesquisa de relações entre os dados, mesmo quando elas não são evidentes, bem como a produção de resultados sob a forma de árvores de decisão de grande simplicidade e legibilidade (FONSECA, 1994).
- As árvores geradas pelo algoritmo CART são sempre binárias, as quais podem ser percorridas da sua raiz até as folhas respondendo apenas a questões simples do tipo “sim” ou “não”.

No caso deste trabalho para a geração da árvore temos um processo de treinamento. Esse processo consiste na análise interativa dos valores das

características (Entropia, Energia, Centroide, Fluxo espectral, *Rolloff*, ZCR, frequência máxima, frequência mínima e frequência mais intensa). No treinamento foram utilizadas 10 aves de cada espécie para gerar a árvore de decisão. Após a geração interativa da árvore foi realizado o reconhecimento.

CAPÍTULO 4

4 RESULTADOS E DISCUSSÃO

Nesta parte do trabalho serão apresentados os resultados obtidos aplicando a metodologia proposta, ou seja, os passos de todas as fases necessárias para o reconhecimento: o pré-processamento (normalização, filtragem, segmentação e extração das características utilizada no reconhecimento), treinamento e por fim o reconhecimento. O reconhecimento foi feito utilizando 575 sinais referentes às 12 espécies escolhidas. A tabela 3 mostra a distribuição dessas aves pelo território brasileiro e por outros países da América.

Tabela 3– Distribuição geográfica das aves estudadas.

Ave/Região	Centro-oeste	Norte	Nordeste	Sudeste	Sul	Outros Países	Total
CRYUND	18	4	0	2	0	17	41
DENVID	1	0	2	10	9	0	22
HERCAC	2	2	6	23	6	0	39
ARACAJ	11	3	6	30	11	6	67
COLSQU	32	3	2	19	0	30	86
DROPAV	3	3	6	8	13	0	33
PITSUL	3	0	0	25	20	34	82
MEGPIT	4	4	7	26	12	0	53
CYCGUJ	4	1	1	11	9	0	26
VOLJAC	7	6	1	6	4	4	28
TODPOL	0	0	1	23	6	0	30
EUPCHO	6	3	13	27	19	0	68
Total	91	29	45	210	109	91	575

4.1 PRÉ-PROCESSAMENTO

A primeira fase do pré-processamento é a normalização e a extração do valor médio do sinal, que têm como intuito padronizar todos os sinais coletados, já

que os mesmos foram gravados em condições diferentes, ou seja, alguns têm o som gravado próximo e outros distantes dos gravadores. A normalização mantém a amplitude do sinal entre -1 e 1 e a extração do valor médio faz com que o sinal tenha média zero.

A figura 30 mostra um som já normalizado, com amplitude entre -1 e 1.

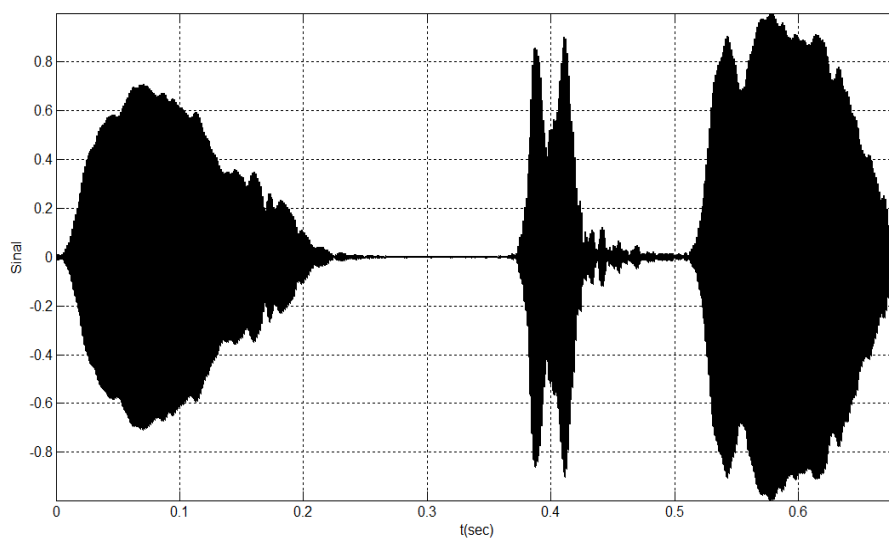


Figura 30 – Sinal padronizado do som de um Fogo-apagou.

4.2 RESULTADOS DA FILTRAGEM

Como foi discutido no capítulo referente à filtragem, as funções utilizadas na confecção dos filtros foram a Haar, e as Daubechies, com níveis de decomposição do sinal entre 1 e 6. Após várias filtrações utilizando diferentes níveis e funções mães observou-se que os filtros de funções mães mais baixas que db10 não conseguiram filtrar de maneira eficiente, pois não retiravam o ruído de forma significativa e ainda causavam grande distorção no som final. Os melhores resultados foram obtidos com a função mãe db10 com 6 níveis de decomposição, mesmo resultado obtido por Selin

(2007). O limiar utilizado foi o soft-thresholding, pois o hard-thresholding acabou eliminando mais som que ruído.

Um problema recorrente da *denoising* é o surgimento de harmônicos no som processado. Para eliminar tal problema utilizou-se um filtro passa-faixa do tipo chebyshev invertido (CONCEIÇÃO, 2012). As ordens de aplicação dos filtros estudadas foram as seguintes:

- a) Filtro chebyshev, filtro wavelet e filtro chebyshev novamente;
- b) Filtro wavelet e filtro chebyshev.

A figura 31 mostra o som de um Fogo-apagou, sinal original e seu espectrograma, utilizado para a análise do processo de filtragem.

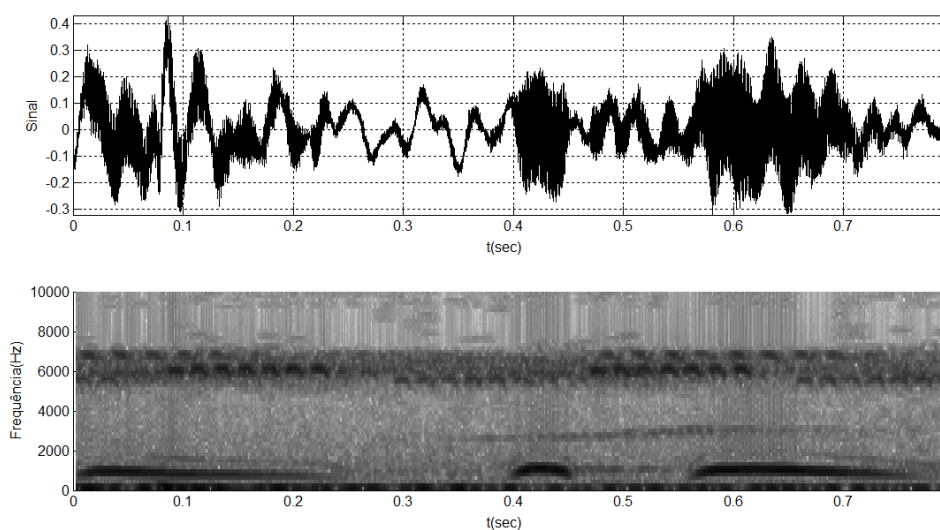


Figura 31 – Sinal e espectrograma do som de um Fogo-apagou.

Aplicando o método (a): duas filtragem com filtro chebyshev e uma filtragem com transformada wavelet, obteve-se o sinal filtrado mostrado na figura 32.

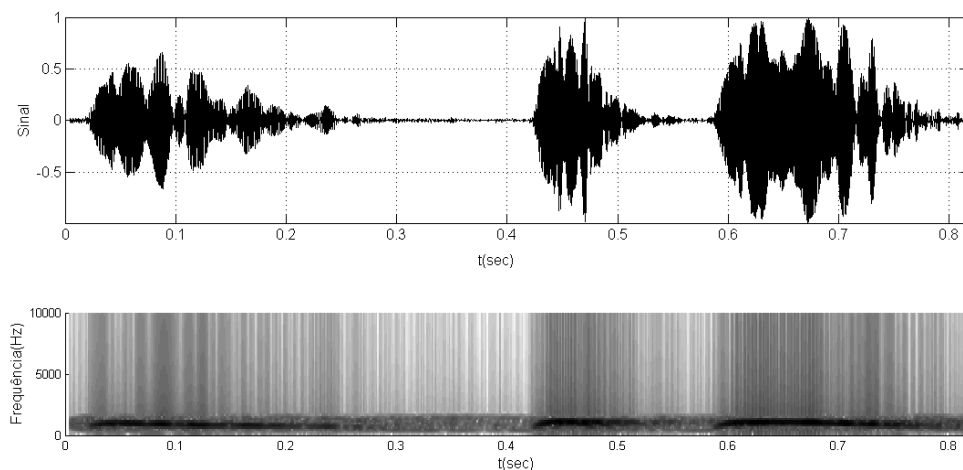


Figura 32 – Sinal e espectrograma do som de um Fogo-apagou após a aplicação do método (a).

A figura 33 mostra o resultado da aplicação do método (b): uma filtragem com transformada wavelet e outra com filtro chebyshev.

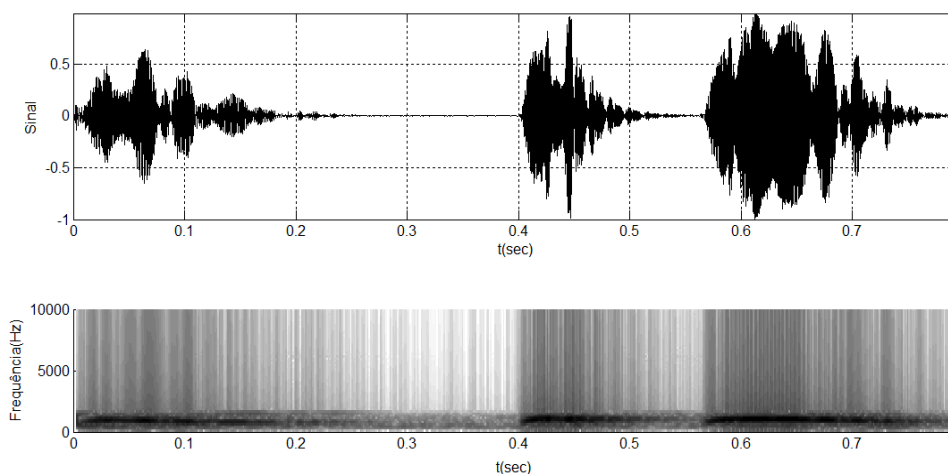


Figura 33 – Sinal e espectrograma do som de um Fogo-apagou após a aplicação do método (b).

Analisando as figuras 32 e 33 pode-se concluir que a aplicação do método (b), utilizando o filtro wavelet primeiramente e depois um chebyshev, obteve melhor resultado, pois retirou mais ruídos que na aplicação do método (a), que utiliza um filtro chebyshev antes e outro depois da filtragem wavelet. O tempo gasto no método (b) também é menor, já que utiliza uma filtragem a menos que o método

(a). Outro ponto importante é que em ambos os processos os filtros chebyshev tiveram a mesma ordem e os mesmos valores de faixa de frequência esses valores foram determinados conforme em (CONCEIÇÃO, 2012).

4.3 RESULTADOS DA SEGMENTAÇÃO

Um processo importante no pré-processamento é a segmentação, pois com ela pode-se separar a região de canto da região de silêncio e desta maneira separar apenas uma vocalização da ave. Este processo é feito de forma automática, conforme descrito na seção 3.11. O tempo de duração do canto da ave pode também ser encontrado por inspeção visual, conforme figura 34, que mostra o canto de um Fogo-apagou. Nota-se que o canto inicia em aproximadamente 0,25 segundos e termina em aproximadamente 1 segundo durando, portanto, cerca de 0,75 segundos.

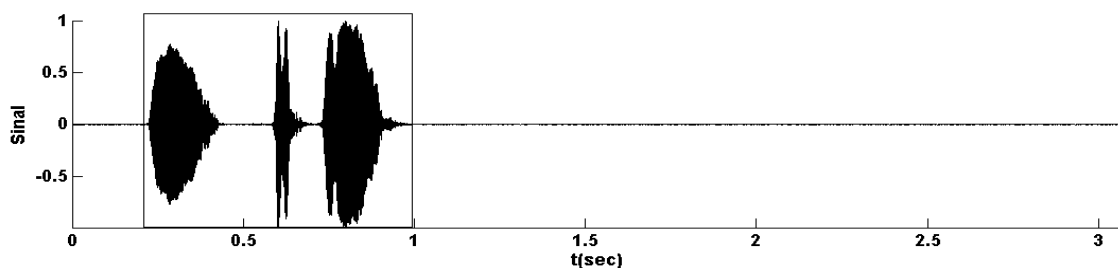


Figura 34 – Detecção visual da duração de canto de um Fogo-apagou.

A figura 35 mostra o mesmo sinal segmentado de forma automática. Foi detectado que o canto se inicia em 0,226 segundos e termina em 0,9422 segundos, durando cerca de 0,7162 segundos, conforme mostra a parte inferior da figura, isolando e ampliando apenas a região de som.

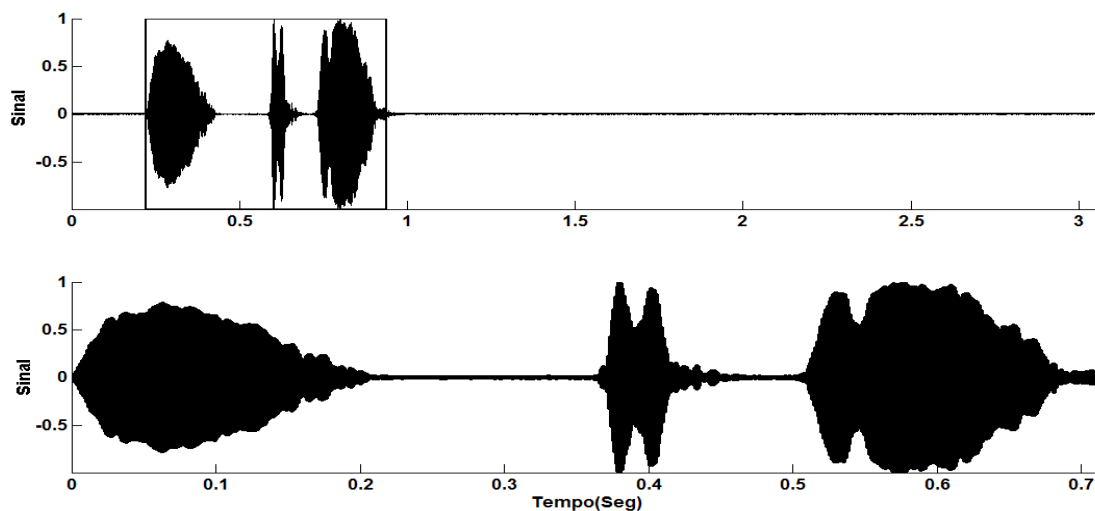


Figura 35 – Detecção automática da duração de canto de um Fogo-apagou.

A tabela 4 mostra a comparação de todas as espécies estudadas. É possível constatar que a detecção automática para o tempo de canto não apresenta grande diferença em relação à análise visual. Notou-se que, mesmo na análise visual, o tempo de canto apresenta uma variação mais acentuada, em alguns casos a mesma ave apresenta tempos de canto diferentes. Por esse motivo a duração de canto não foi utilizada como característica de reconhecimento, já que aves como a Saracura e Teque-teque possuem durações variáveis, para o seu canto.

Tabela 4– Comparação entre inspeção visual e detecção automática

Espécie	Inspeção visual (s)	Detecção automática (s)	Diferença (%)
Acauã	0,8	0,8597	7,46
Bem-te-vi	0,9	0,8661	3,77
Fim-fim	0,8	0,8203	2,54
Fogo-apagou	0,7	0,7162	2,31
Irerê	0,7	0,701	0,14
Jaó	1,4	1,388	0,86
Neinei	1	1,016	1,60
Peixe-frito	1,6	1,4992	6,30
Pitiguari	1,1	1,0344	5,96
Saracura	9	8,8648	1,50
Teque-teque	5	4,5916	8,17
Tiziu	0,7	0,5867	16,19

Observa-se que o método de segmentação mostrou-se eficiente, pois comparado com a análise visual o mesmo apresentou como maior erro de diferença o valor de 16,19%, um valor baixo já que análise visual não é precisa.

A figura 36 mostra a energia, o centróide e o sinal após a filtragem. Pode-se observar para esse sinal que tanto a energia como o centróide são muito maiores na região do canto do que na de silêncio, o que torna possível realizar a segmentação do sinal separando essas duas regiões.

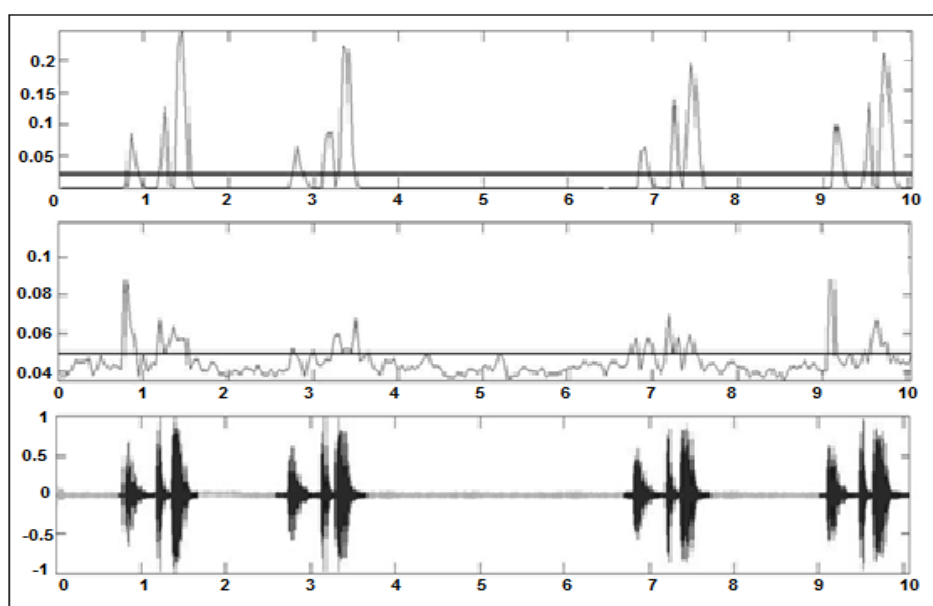


Figura 36 – Energia do sinal, Centroide do sinal e sinal do canto de um Fogo-apagou.

A figura 37 mostra o sinal filtrado e os sinais resultantes da segmentação. Observa-se que o método utilizado mostrou-se eficiente em encontrar as quatro regiões em que há a vocalização da ave.

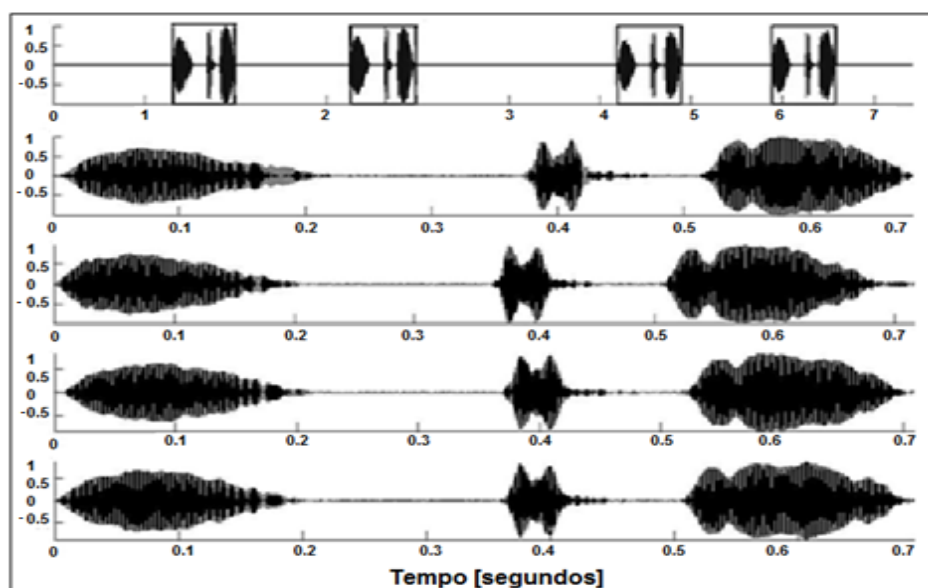


Figura 37 - Sinal segmentado do canto de um Fogo-apagou.

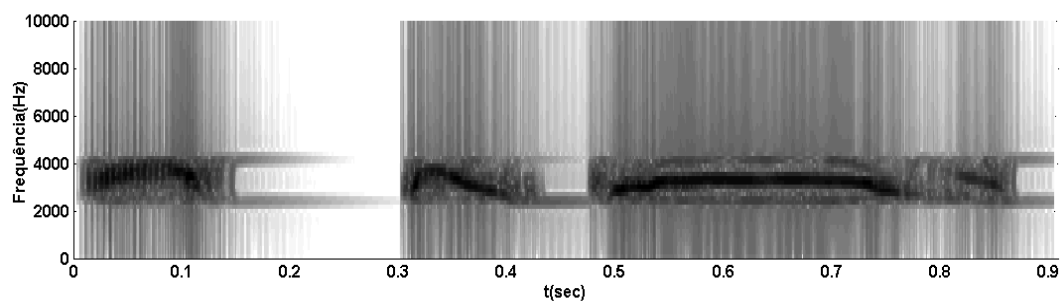
4.4 DETECÇÃO AUTOMÁTICA DE FREQUÊNCIA

A detecção automática de frequência foi realizada como proposto por Conceição (2012). Em seu trabalho o método proposto mostrou-se eficiente e as características calculadas (frequências máximas, mínimas e mais intensas) mostraram-se úteis no reconhecimento de espécies. A tabela 5 mostra o espectrograma de cada espécie estudada neste trabalho, o que permite a análise da faixa de frequência de cada espécie.

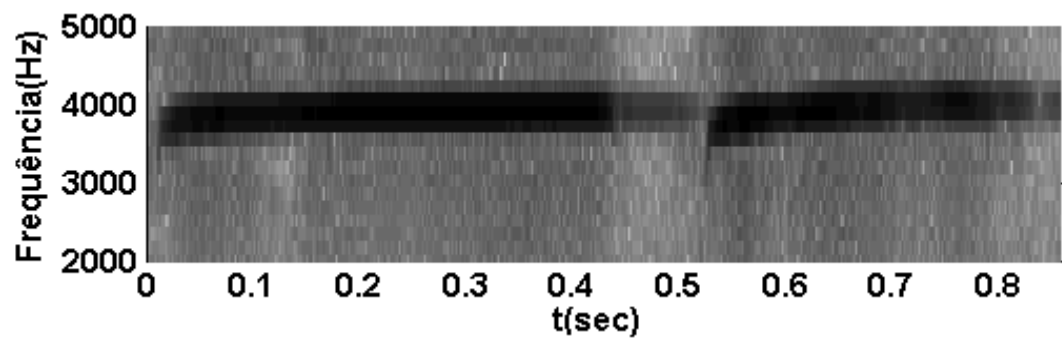
Tabela 5 – Tabela com os espectrogramas das aves.

Espécie	Espectrograma
Acauã	

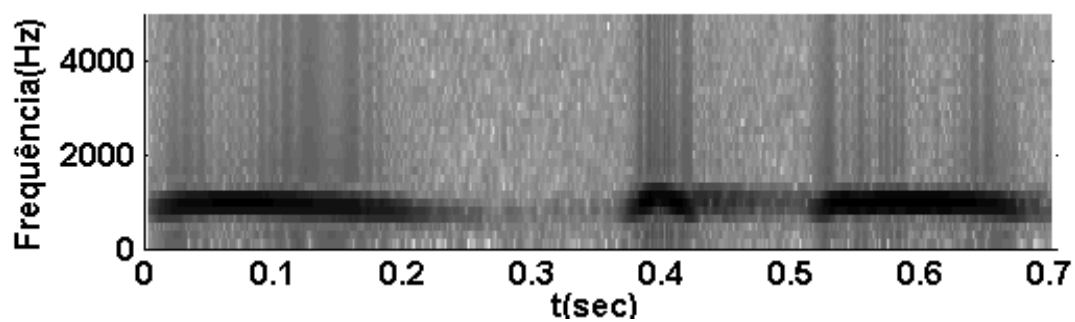
Bem-te-vi



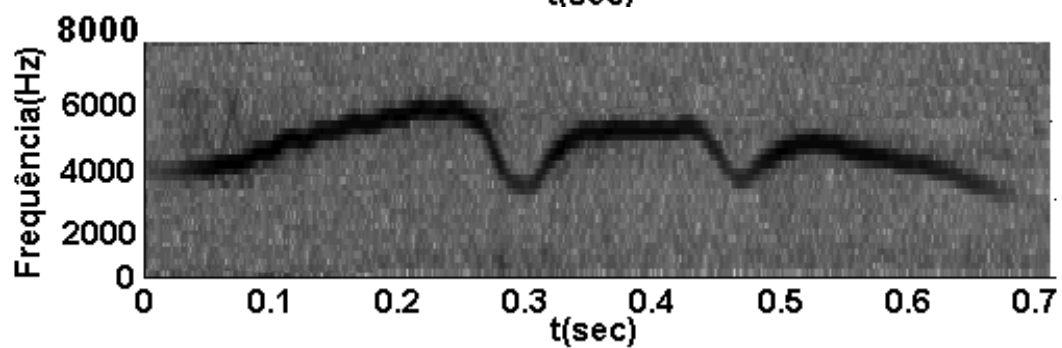
Fim-fim



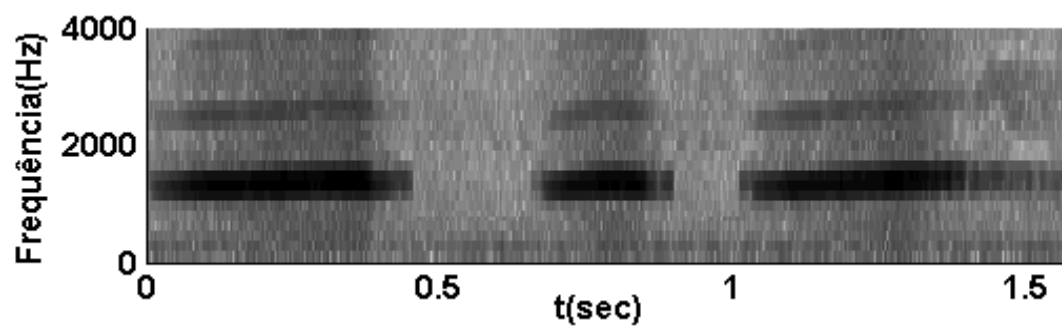
Fogo-apagou



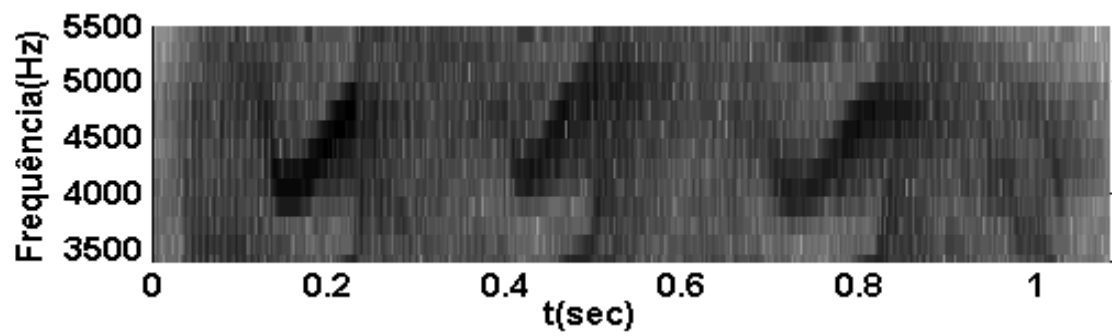
Irerê



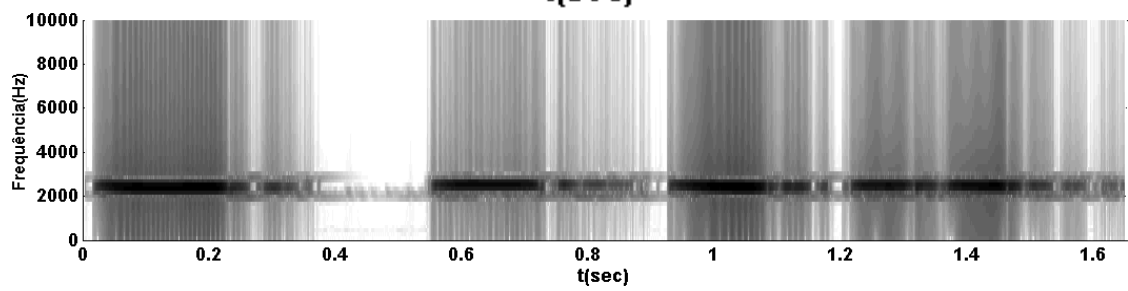
Jaó



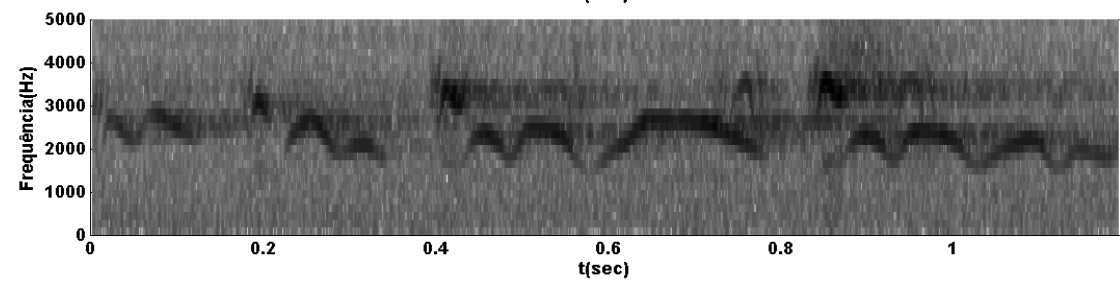
Neinei



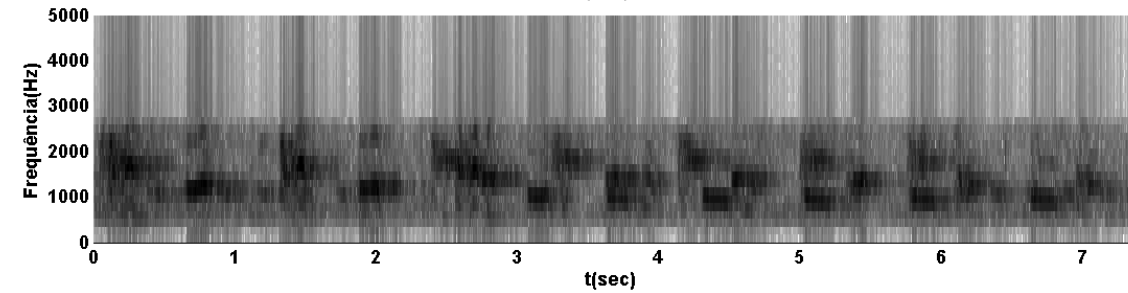
Peixe-frito



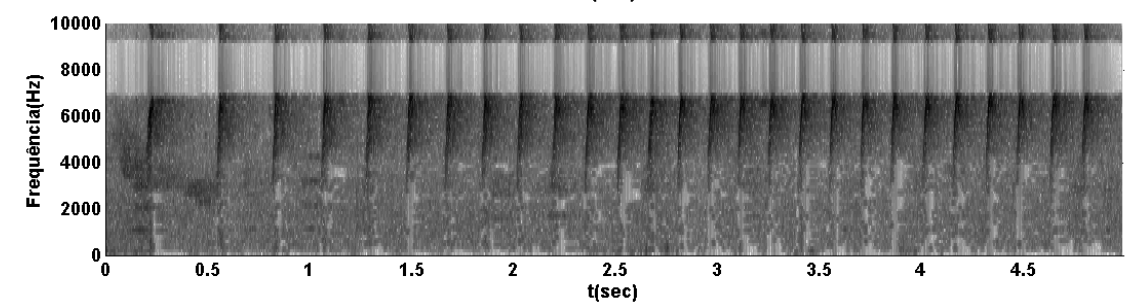
Pitiguari



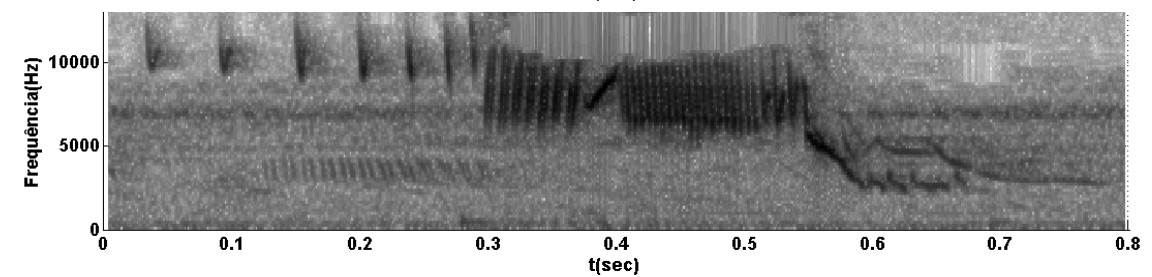
Saracura



Teque-teque



Tiziu



A tabela 6 mostra os valores da frequência máxima (Fmax), frequência mínima (Fmin) e frequência mais intensa (FI), de todas as espécies em estudo. A análise do espectrograma de cada ave e sua faixa de frequência mostra a adequação do método, pois se obteve valores similares à análise visual do espectrograma.

Tabela 6– *Faixa de frequência e frequência mais intensa de todas as espécies.*

Aves	Fmín	Fmáx	FI
Acauã	1033	2153	1550
Bem-te-vi	2326	4048	3198
Fim-fim	3273	4306	3810
Fogo-apagou	430	1378	859
Irerê	2240	7150	3633
Jaó	775	1722	1275
Neinei	3531	5426	4635
Peixe-frito	1895	2842	2296
Pitiguari	2584	3876	3390
Saracura	689	2153	1334
Teque-teque	4187	7062	6329
Tiziu	5857	9365	6365

Pode-se concluir das tabelas que a detecção automática é satisfatória, mesmo nos casos em que a diferença entre a análise visual e a automática seja aparentemente grande, por alguns fatores: na inspeção visual os valores verificados são aproximados, sendo que uma pequena variação dá uma grande diferença em Hz.

4.5 ANÁLISE DAS CARACTERÍSTICAS UTILIZADAS

As figuras 38 a 49 mostram os gráficos das características usadas neste trabalho para reconhecimento das espécies. Para cada característica é mostrado um gráfico do som de uma espécie em que essa característica se destaca e um gráfico do som de uma espécie que apresenta valor máximo para essa característica, para fins de comparação.

As figuras 38 e 39 mostram os gráficos do Centróide do som de um Jaó e de um Teque-teque, respectivamente. Na figura 38 pode-se observar que o valor do Centróide do som do Jaó é aproximadamente constante na região de som e maior que o valor médio do sinal. Isso se deve ao fato de a frequência do som do Jaó ser aproximadamente constante, ocupando uma faixa estreita de frequência, com a frequência mais intensa do som localizada nessa região. Como o sinal está filtrado pode ser observado que o centróide é maior na região de canto e nulo ou praticamente nulo na região de silêncio. Na figura 39 pode-se observar que o valor médio do Centróide do som do Teque-teque é maior que o valor médio do Centróide do som do Jaó, devido ao som do Teque-teque possuir frequências muito maiores que o do Jaó. Assim, sons com maiores frequências possuem Centróides com maiores valores.

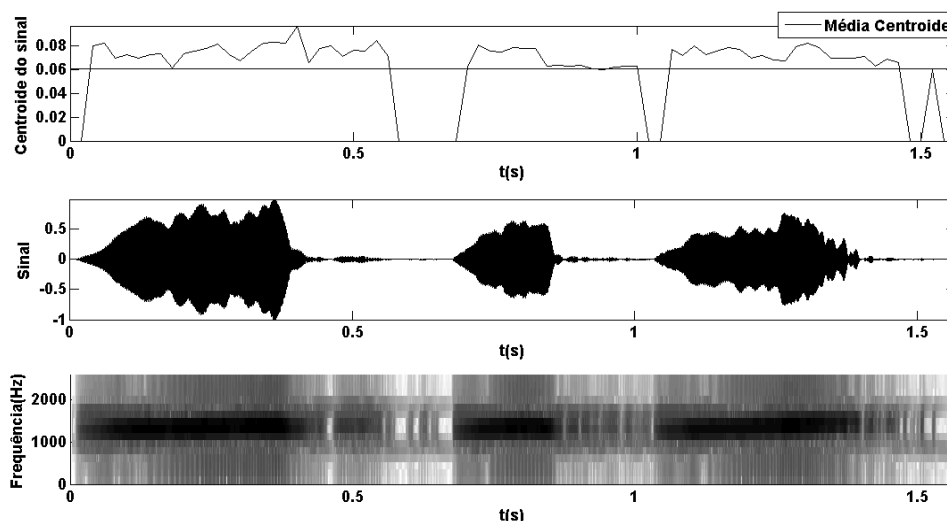


Figura 38 – Centroide, sinal e espectrograma do canto de um Jaó.

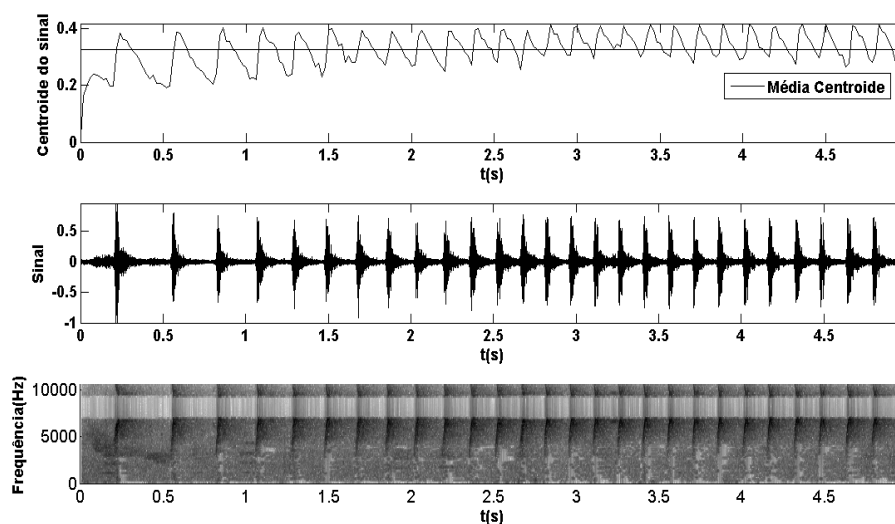


Figura 39 - Centroide, sinal e espectrograma do canto de um Teque-teque.

As figuras 40 e 41 mostram a Energia do sinal do canto de um Peixe-frito e de um Acauã, respectivamente. Nessas figuras pode-se observar que o valor da Energia é proporcional à magnitude do sinal e que a Energia assume valores acima da média do sinal nas regiões de sons enquanto nas regiões de silêncio seu valor é nulo ou quase nulo. Observa-se também que sinais que “ocupam maiores áreas”, como o som do Acauã (Fig. 41), tem Energia média maior que os que “ocupam menores área”, como o som do Peixe-frito (Fig. 40).

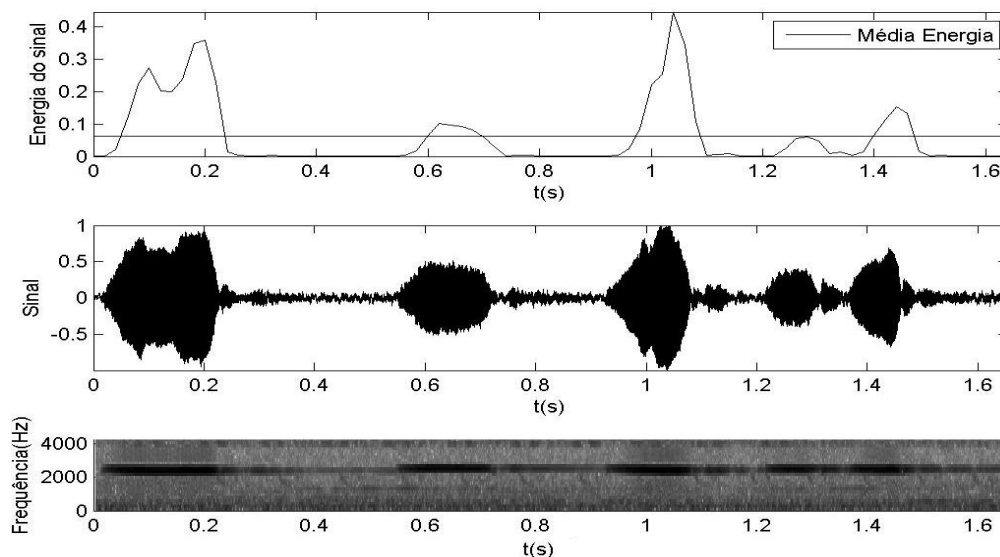


Figura 40 - Energia, sinal e espectrograma do som de um Peixe-frito.

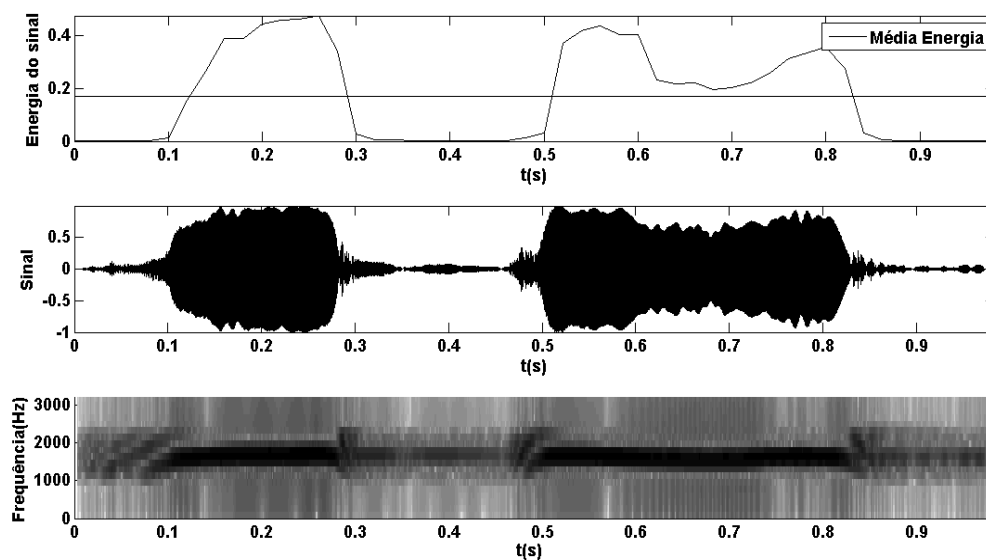


Figura 41 - energia, sinal e espectrograma do som de um Acauã.

As figuras 42 e 43 mostram a Entropia do canto de um Peixe-frito e de um Fogo-apagou, respectivamente. Como a energia é normalizada no valor 1 (um), o cálculo da Entropia tende a valores maiores para os valores de energia menores que 0,5 e para valores menores para valores de energia acima de 0,5, como pode-se ver na equação da Entropia, equação (23) e comparando-se as figuras 40 e 42.

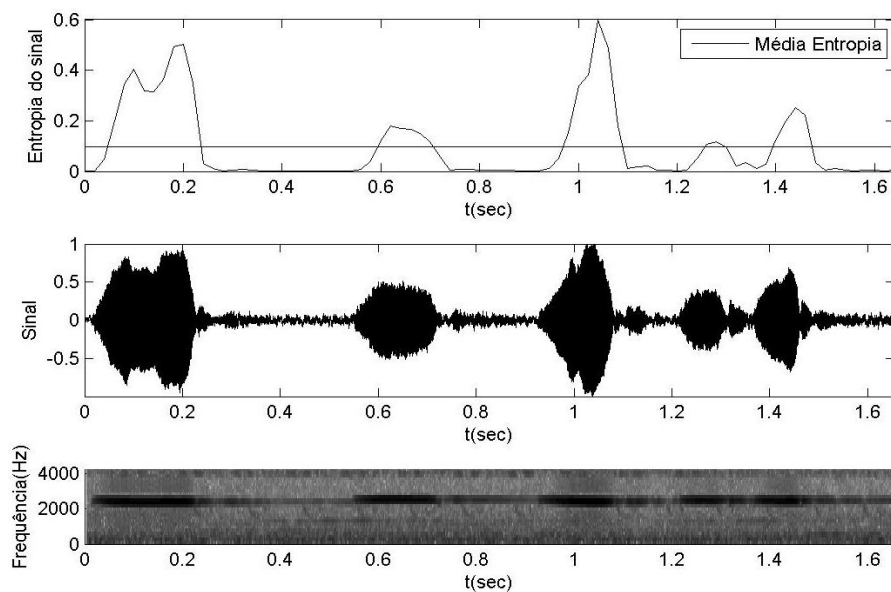


Figura 42 - Entropia, sinal e espectrograma do som de um Peixe-frito.

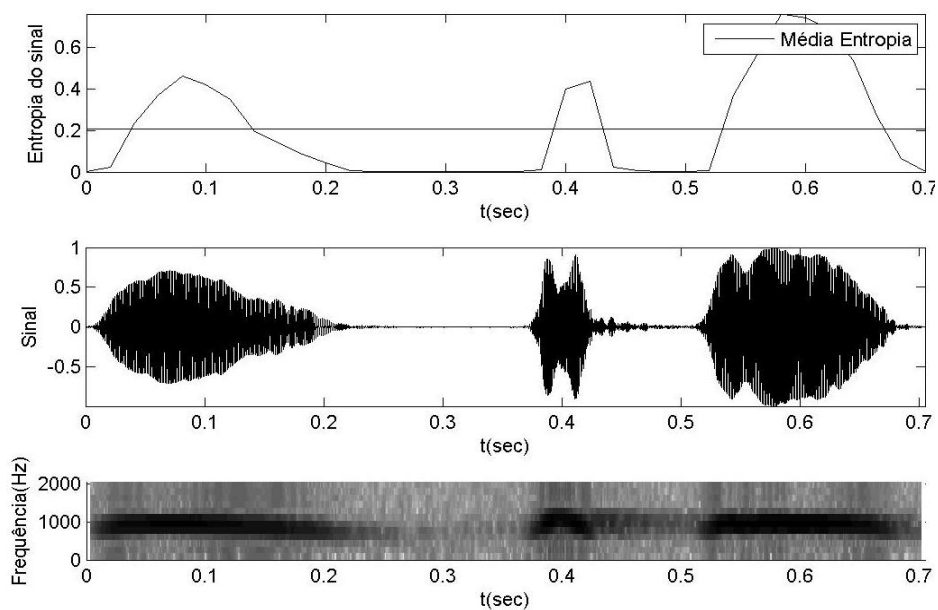


Figura 43 - Entropia, sinal e espectrograma do som de um Fogo-apagou.

O gráfico da Entropia do sinal é semelhante ao da Energia, com os valores abaixo de 0,5 aumentados e os valores acima de 0,5 diminuídos. Essa característica é uma medida de mudanças abruptas no nível de energia de um sinal. Assim, sinais que apresentam maiores variações de energia, como o som de um Fogo-apagou (Fig. 43)

apresentam Entropia média maiores que os sinais com menores variações de energia, como o som do Peixe-frito (Fig. 42).

As figuras 44 e 45 mostram o Fluxo Espectral do canto de uma Jaó e de um Tiziu, respectivamente. Na figura 44, referente ao Jaó, pode-se notar que nas regiões de som o valor do Fluxo Espectral é aproximadamente constante e menor que a média do sinal, devido à frequência do som do Jaó ser aproximadamente constante, ocupando uma faixa estreita de frequência. Observa-se também nessa figura picos maiores que a média no início e no final das regiões de sons, mostrando que nesse caso essa característica pode ser utilizada para a segmentação do sinal.

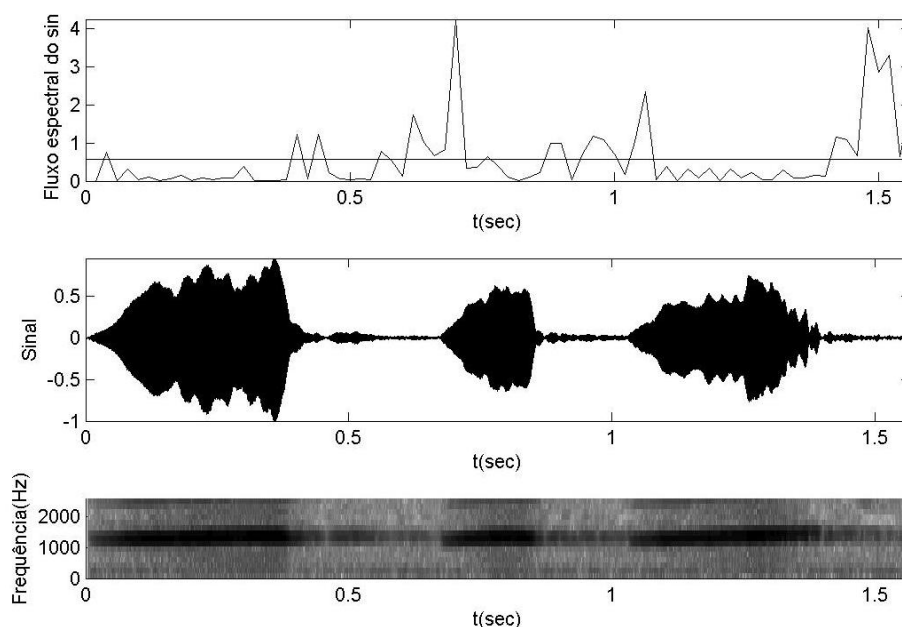


Figura 44 – Fluxo Espectral, sinal e espectrograma do som de um Jaó.

Na figura 45, referente ao som do Tiziu, observam-se altos valores do Fluxo Espectral, devido às grandes variações de frequências de seu som, que ocupa uma larga faixa de frequência, muito maiores que a frequência do Jaó. Assim, o som do Tiziu apresenta valor médio do Fluxo Espectral muito maior que o do Jaó.

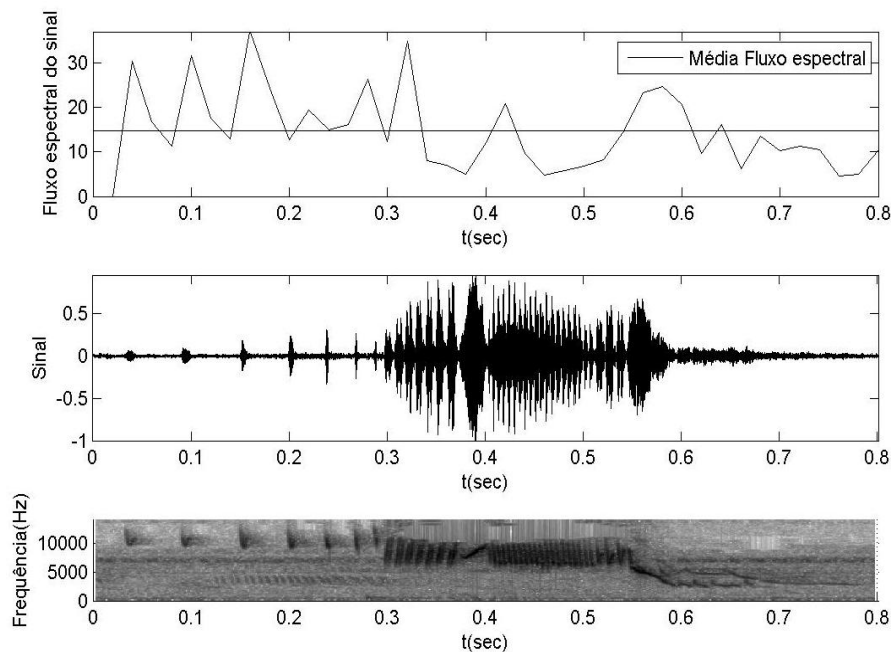


Figura 45 – Fluxo Espectral, sinal e espectrograma do som de um Tiziu.

As figuras 46 e 47 mostram a característica frequência *Rolloff* do canto de um Peixe-frito e de um Teque-teque, respectivamente. A figura 46, referente ao Peixe-frito, mostra que a frequência *Rolloff* apresenta valores menores nas regiões de som, porque a maior parte da energia está concentrada nessas regiões.

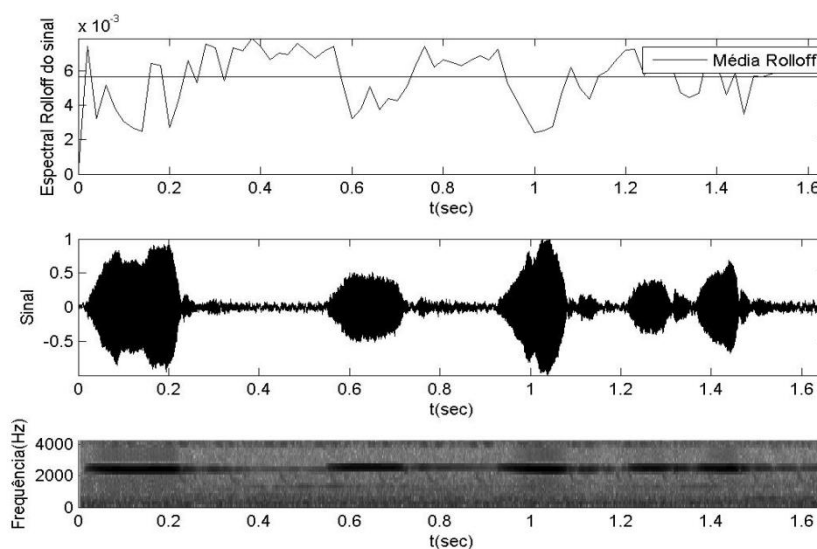


Figura 46 – Frequência *Rolloff*, sinal e espectrograma do som de um Peixe-frito.

A figura 47, referente ao som do Teque-teque, que contém frequências maiores que do Peixe-frito, mostra altos valores da frequência *Rolloff*. Assim, o

som do Teque-teque apresenta valor médio da frequência *Rolloff* muito maior que o do Peixe-frito.

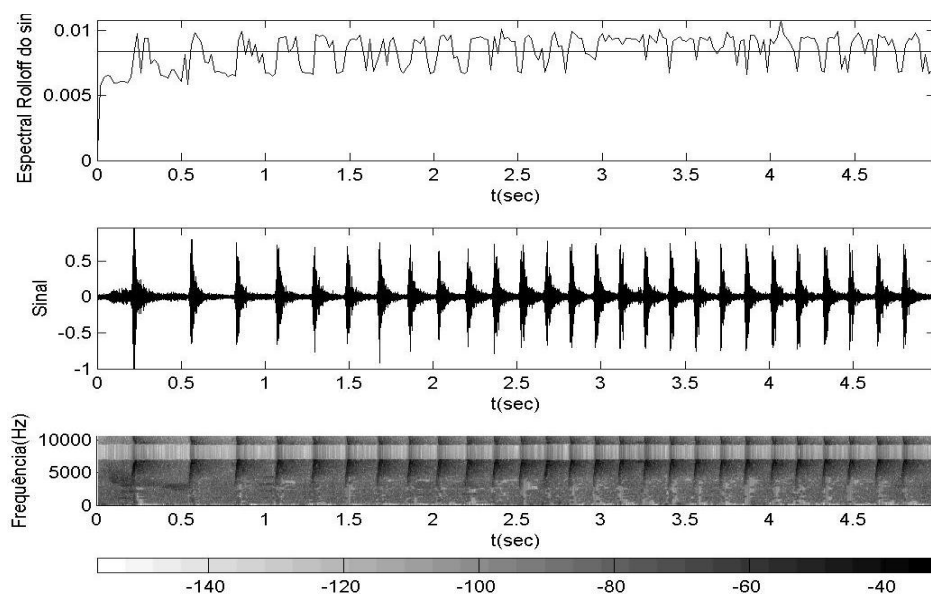


Figura 47- Frequência *Rolloff*, sinal e espectrograma do som de um Teque-teque.

As figuras 48 e 49 mostram, respectivamente, a característica ZCR do som de um Peixe-frito e de um Teque-teque. A figura 48 mostra que o valor do ZCR é aproximadamente constante no sinal, onde a frequência do som é aproximadamente constante. Como o ZCR é normalizado por $f_s/2=22050$ Hz, onde f_s é a frequência de amostragem, a frequência principal do som pode ser obtida pelo produto de ZCR por $f_s/2$, no caso aproximadamente $0.1 \times 22050 = 2205$ Hz. Na figura 49, referente ao som do Teque-teque, que contém frequências maiores que do Peixe-frito, observam-se valores maiores do ZCR. Assim, o som do Teque-teque (Fig. 49) apresenta valor médio do ZCR muito maior que o do Peixe-frito (Fig. 48).

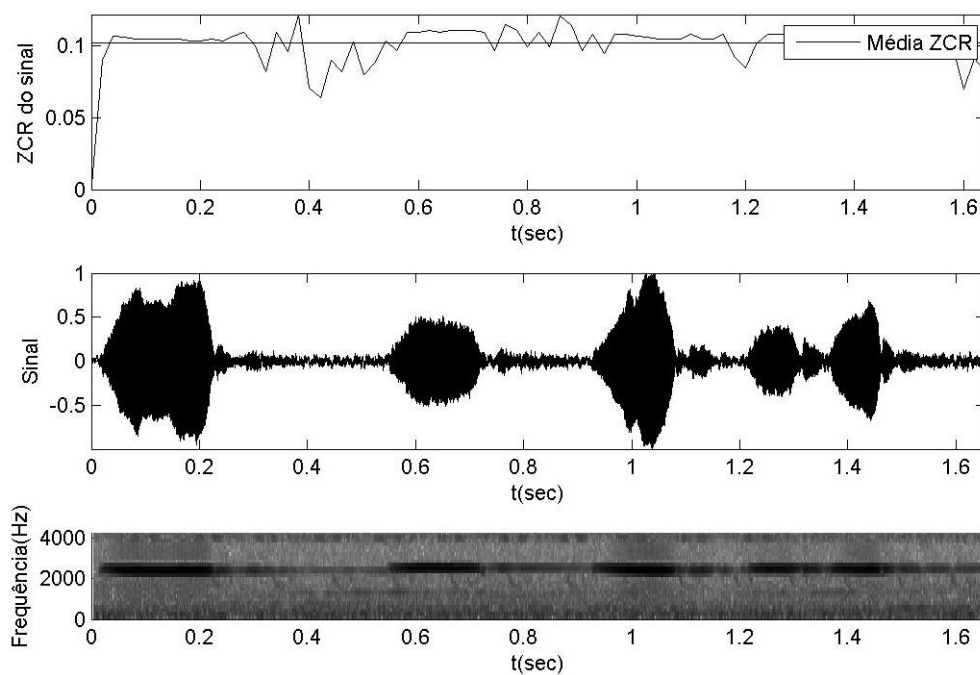


Figura 48 - ZCR, sinal e espectrograma do som de um Peixe-frito.

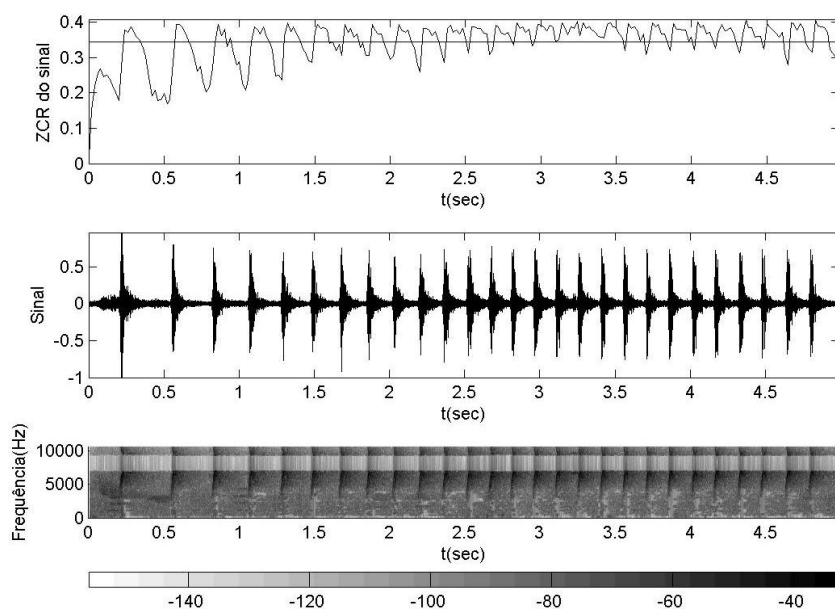


Figura 49 - ZCR, sinal e espectrograma do som de um Teque-teque.

4.6 FASE DE TREINAMENTO

Na fase de treinamento foram utilizadas 10 aves de cada espécie para a geração automática da árvore de decisão. As frequências mínimas, máximas e mais intensas de cada vocalização foram calculadas automaticamente utilizando o método proposto por Conceição (2012). Foram também calculados os valores médios das características (Entropia, Energia, Centroide, Fluxo Espectral, ZCR e Frequência *Rolloff*) para cada vocalização. Foi analisada ainda a razão entre o valor máximo e a média das características para ser utilizado no reconhecimento, mas esse parâmetro não forneceu bons resultados, pois não permitiu a separação de muitas espécies, já que houve um grande número de superposições nos gráficos. As tabelas 7, 8 e 9 mostram os valores dos limites de cada característica para cada espécie.

Tabela 7– Limites das características.

Ave	Neinei		Bem-te-vi		Tiziu		Jaó	
Entropia	0,163	0,235	0,138	0,222	0,266	0,405	0,086	0,121
ZCR	0,143	0,225	0,120	0,163	0,294	0,356	0,054	0,057
<i>Rolloff</i>	0,003	0,004	0,003	0,003	0,004	0,005	0,003	0,004
Centroide	0,147	0,242	0,107	0,168	0,297	0,352	0,062	0,075
Fluxo espec.	5,356	9,945	3,957	7,548	7,027	15,194	0,216	0,704
Energia	0,018	0,052	0,021	0,084	0,029	0,067	0,038	0,123
Fmin	1120	3015	1034	2326	3273	7235	689	775
Fmax	4048	8872	3618	5082	5599	11456	1637	1723
FI	2994	5215	2490	3374	6789	8667	1197	1275

Tabela 8– Limites das características

Ave	Fogo-apagou		Peixe-frito		Teque-teque		Irerê	
Entropia	0,164	0,212	0,094	0,116	0,039	0,175	0,236	0,306
ZCR	0,035	0,042	0,096	0,192	0,216	0,271	0,151	0,173
<i>Rolloff</i>	0,003	0,004	0,004	0,004	0,004	0,005	0,003	0,004
Centroide	0,048	0,056	0,118	0,229	0,234	0,282	0,160	0,176
Fluxo espec.	1,647	2,903	0,624	4,721	6,748	12,128	3,748	7,011
Energia	0,048	0,119	0,029	0,165	0,010	0,034	0,059	0,110
Fmin	258	431	1292	1895	2929	4048	2326	3000
Fmax	1378	1464	2412	2929	6632	8441	4307	5254
FI	841	1016	1835	2522	5878	6754	2865	3878

Tabela 9– *Limites das características.*

Ave	Fim-fim		Saracura		Acauã		Pitiguari	
Entropia	0,163	0,336	0,010	0,034	0,119	0,172	0,102	0,160
ZCR	0,160	0,218	0,051	0,064	0,064	0,093	0,095	0,118
<i>Rolloff</i>	0,004	0,006	0,003	0,003	0,003	0,004	0,003	0,004
Centroide	0,156	0,226	0,061	0,075	0,075	0,155	0,103	0,129
Fluxo espec.	0,546	1,650	2,116	3,716	1,008	3,195	3,854	6,759
Energia	0,064	0,180	0,007	0,047	0,040	0,126	0,033	0,061
Fmin	1938	3273	345	861	258	1240	1464	1809
Fmax	3188	4307	1723	2498	1981	2412	2929	3704
FI	2611	3832	1124	1639	1469	1571	2196	3024

As figuras 50 a 58 mostram as regiões de cada característica e os cruzamentos entre as regiões de cada uma das doze espécies de aves em estudo. Nos gráficos de algumas características observa-se, em um primeiro momento, que algumas espécies de aves têm suas regiões bem definidas, pois não há superposição de sua região nessa característica com as regiões das características de outras aves.

Ao analisar os gráficos observa-se que o gráfico da Entropia (Fig. 50) mostra que a Saracura possui uma região bem definida, ou seja, não há cruzamentos de sua região de Entropia com a região de outras espécies, então essa característica pode ser utilizada para classificar a Saracura. O Fogo-apagou pode ser reconhecido pelo Centroide (Fig. 51), pelo ZCR (Fig. 52), pela frequência máxima (Fig. 57) ou pela frequência mais intensa (Fig. 58). O Tiziu pelo Centroide (Fig. 51) e pela frequência mais intensa (Fig. 58). A frequência mais intensa (Fig. 58) permite também reconhecer o Teque-teque e a frequência máxima (Fig. 57) permite reconhecer o Jaó. Num primeiro instante pode-se, portanto, extrair as características necessárias para o reconhecimento de cinco espécies de aves (Saracura, Fogo-apagou, Tiziu, Teque-teque e Jaó).

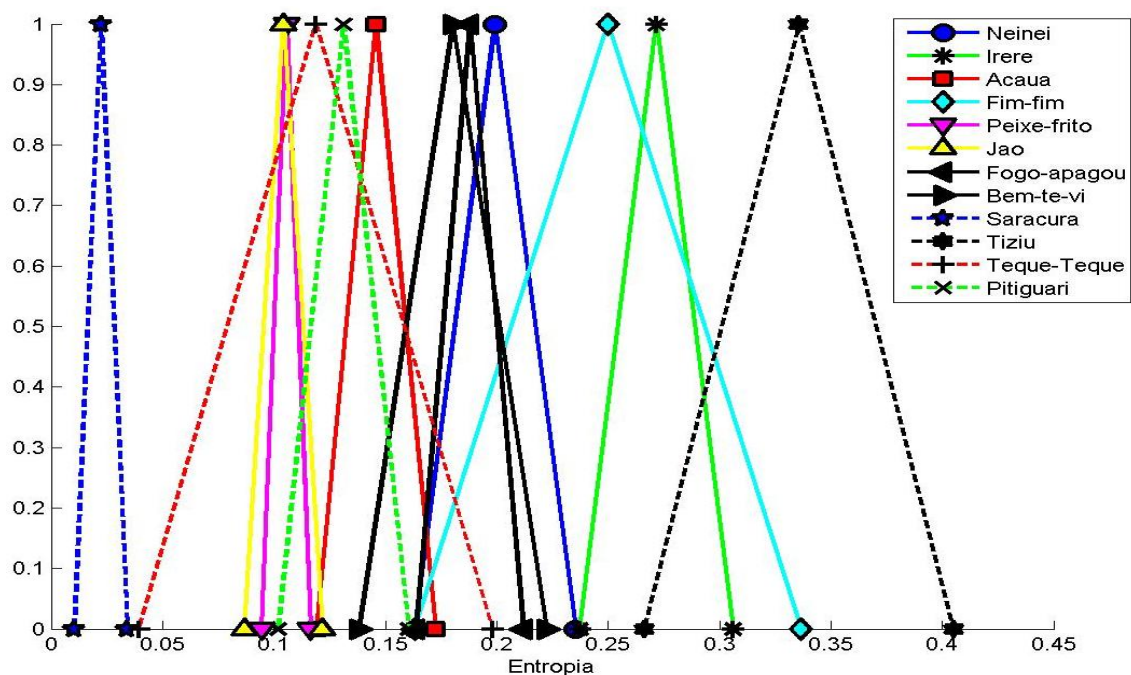


Figura 50 – Regiões de Entropia dos sons das doze espécies de aves em estudo.

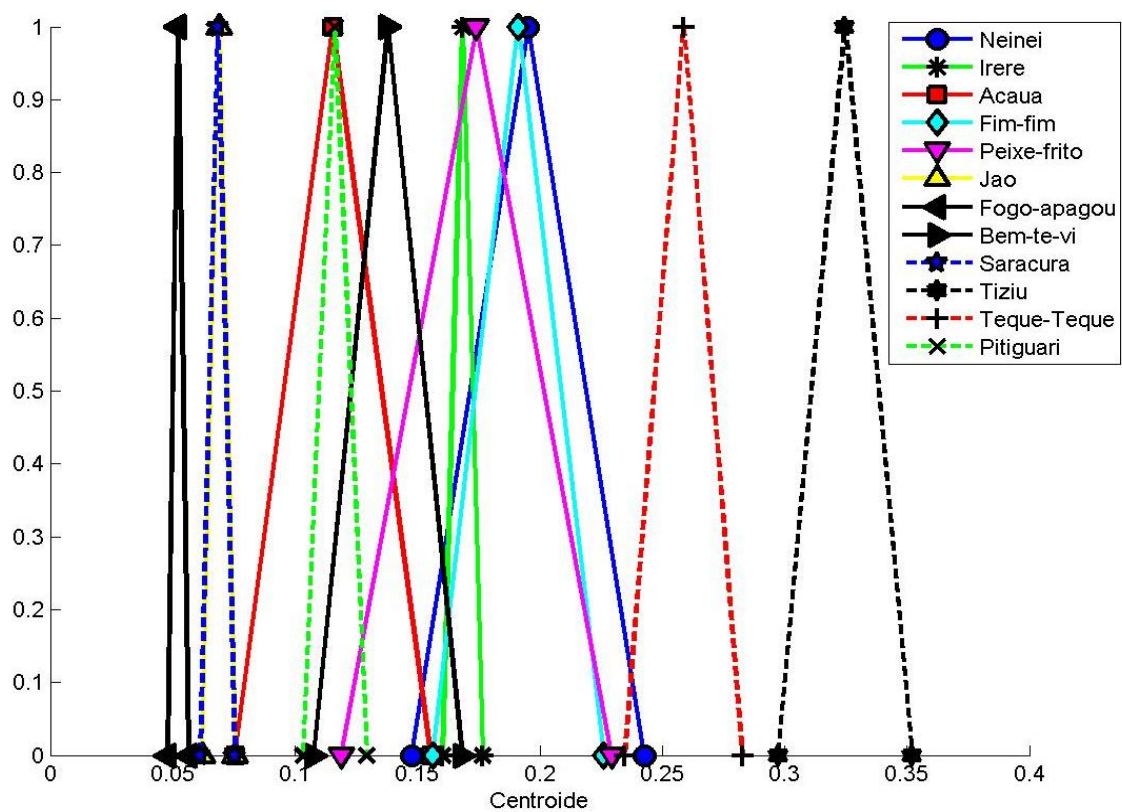


Figura 51 – Regiões de Centroide dos sons das doze espécies de aves em estudo.

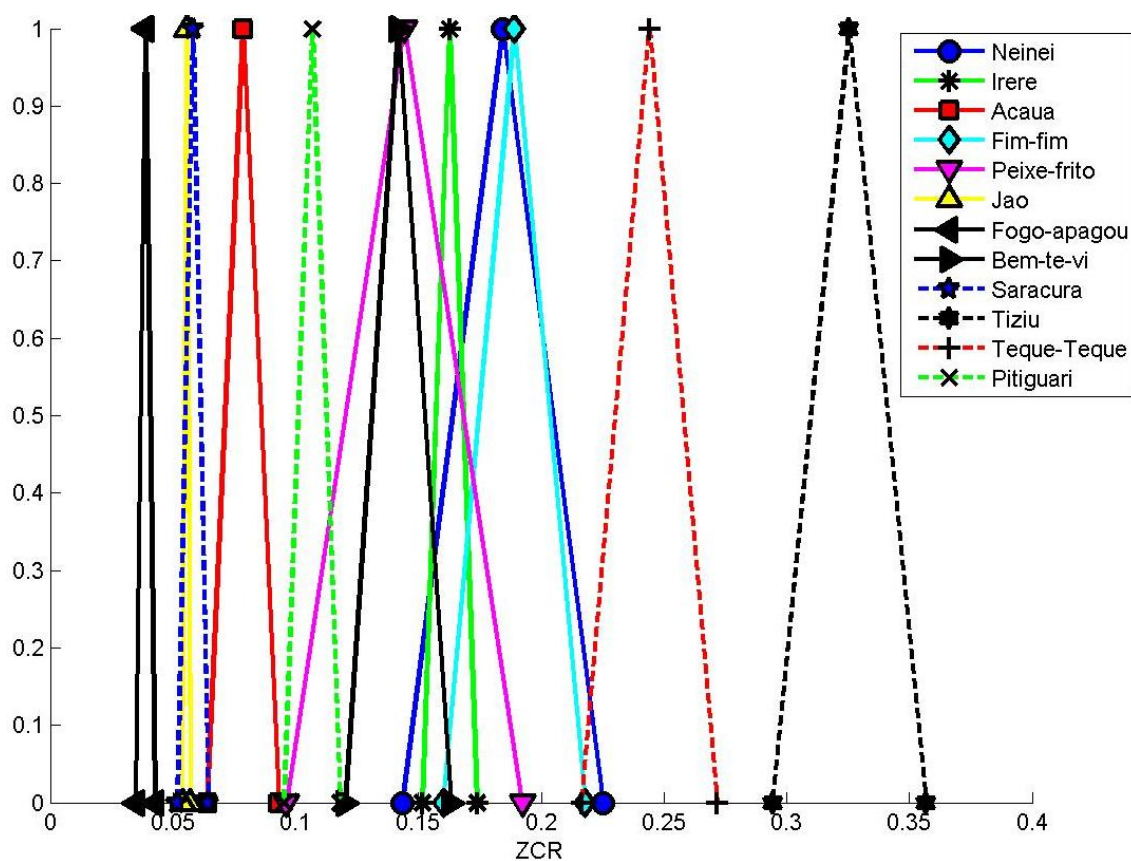


Figura 52 – Regiões de ZCR dos sons das doze espécies de aves em estudo.

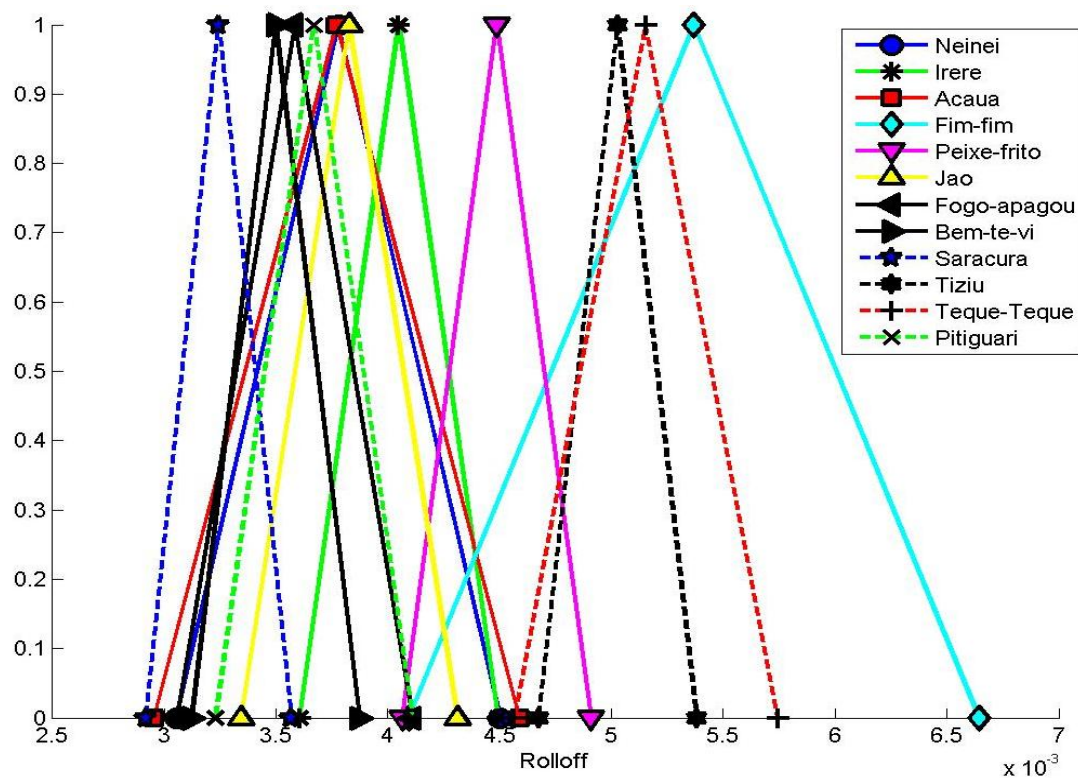


Figura 53 – Regiões de frequência *Rolloff* dos sons das doze espécies de aves em estudo.

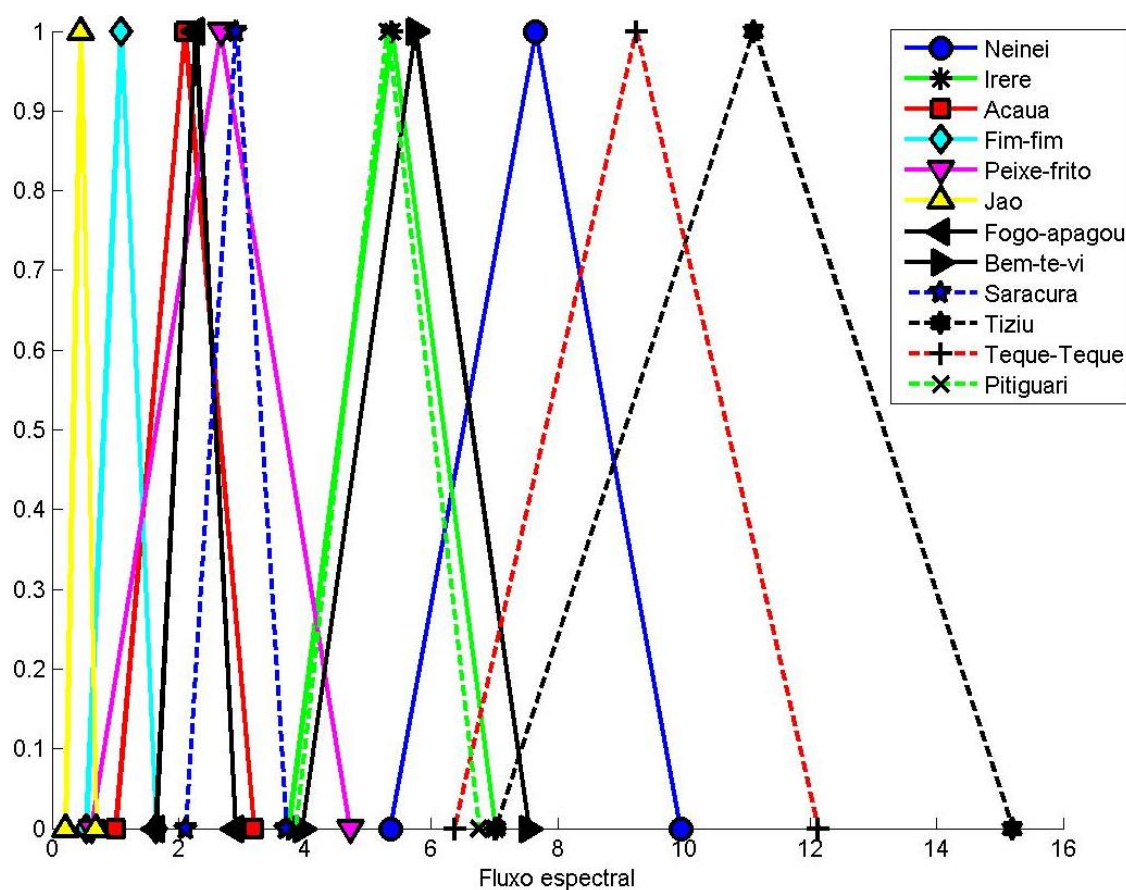


Figura 54 – Regiões de Fluxo Espectral dos sons das doze espécies de aves em estudo.

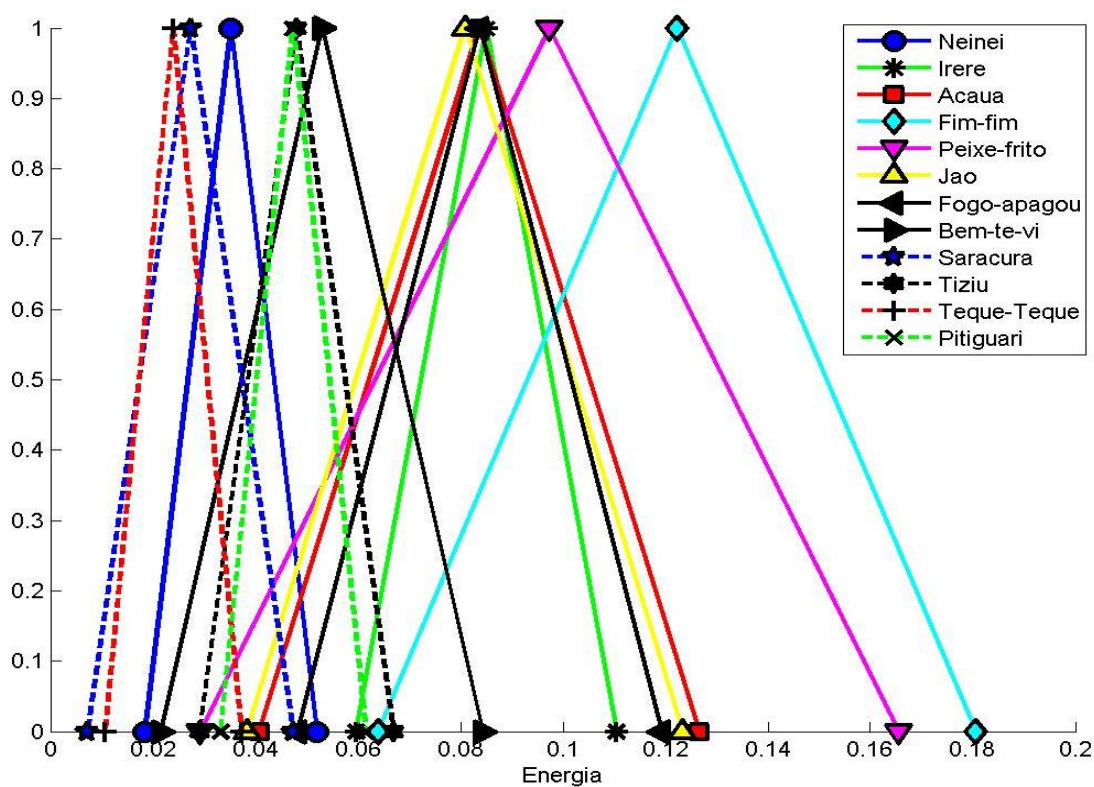


Figura 55 – Regiões de Energia dos sons das doze espécies de aves em estudo.

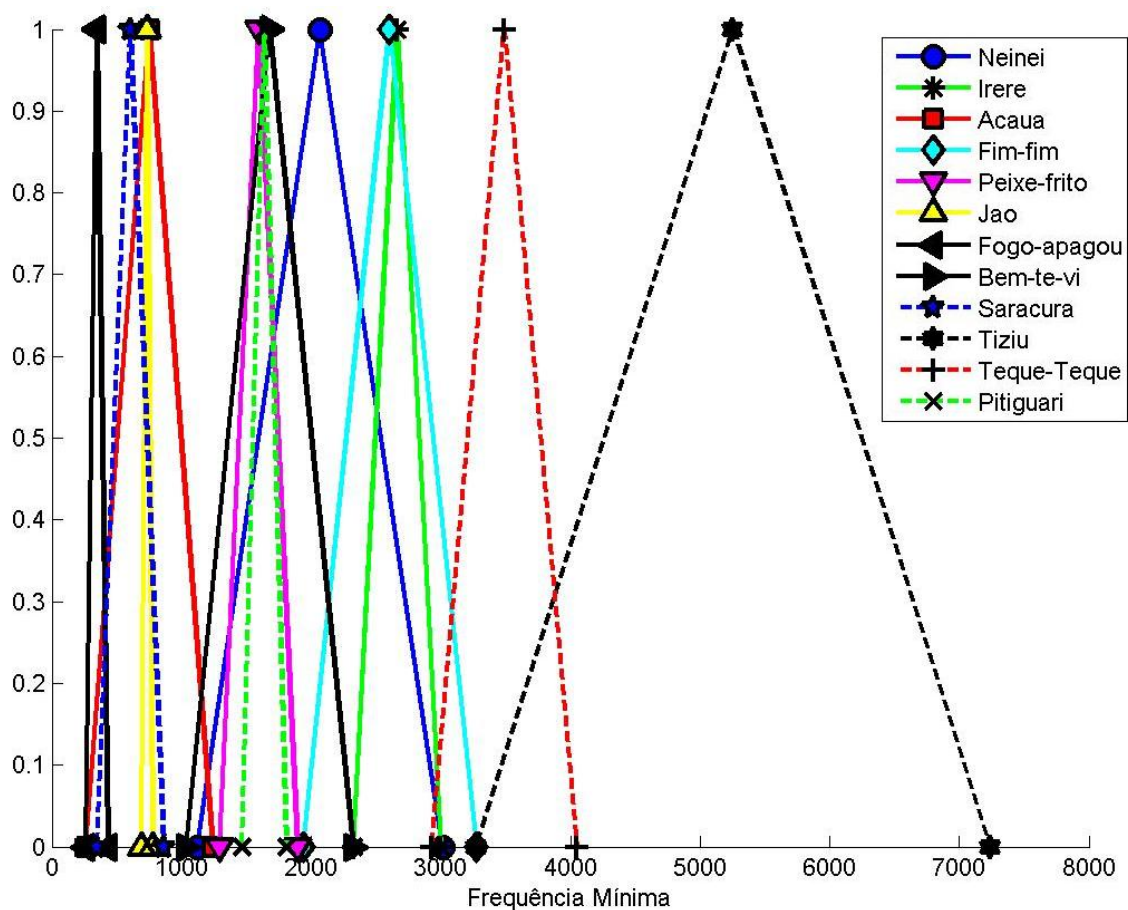


Figura 56 – Regiões de Frequência Mínima dos sons das doze espécies de aves em estudo.

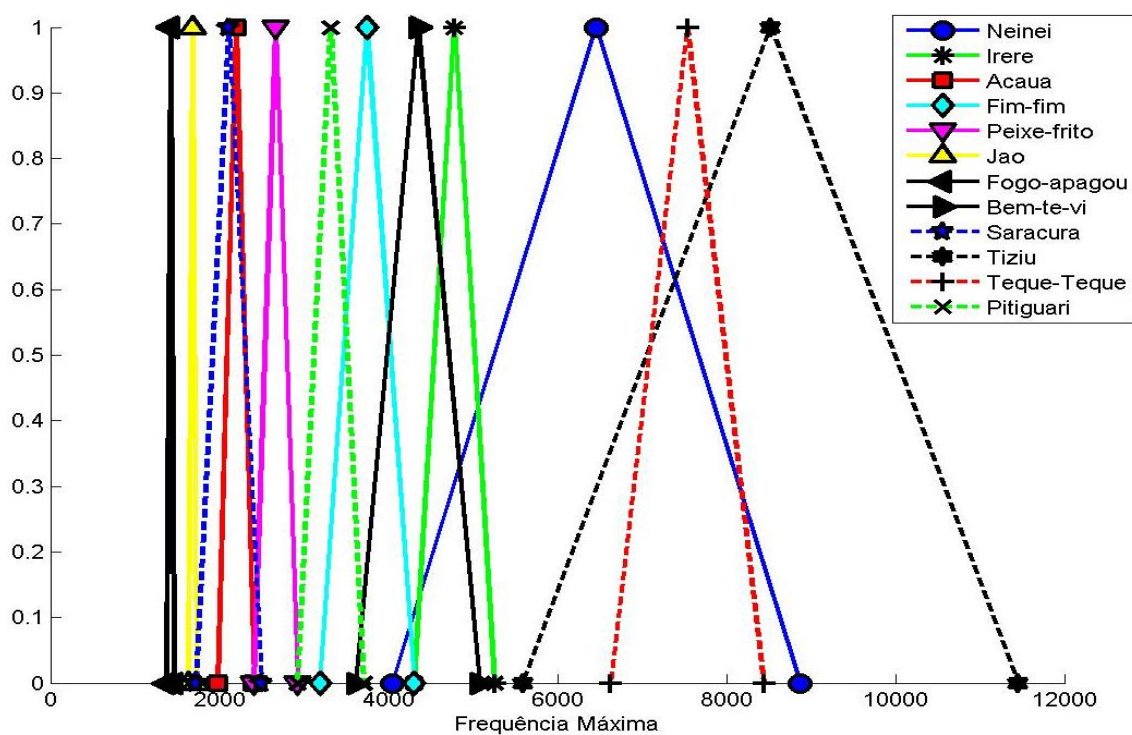


Figura 57 – Regiões de Frequência Máxima dos sons das doze espécies de aves em estudo.

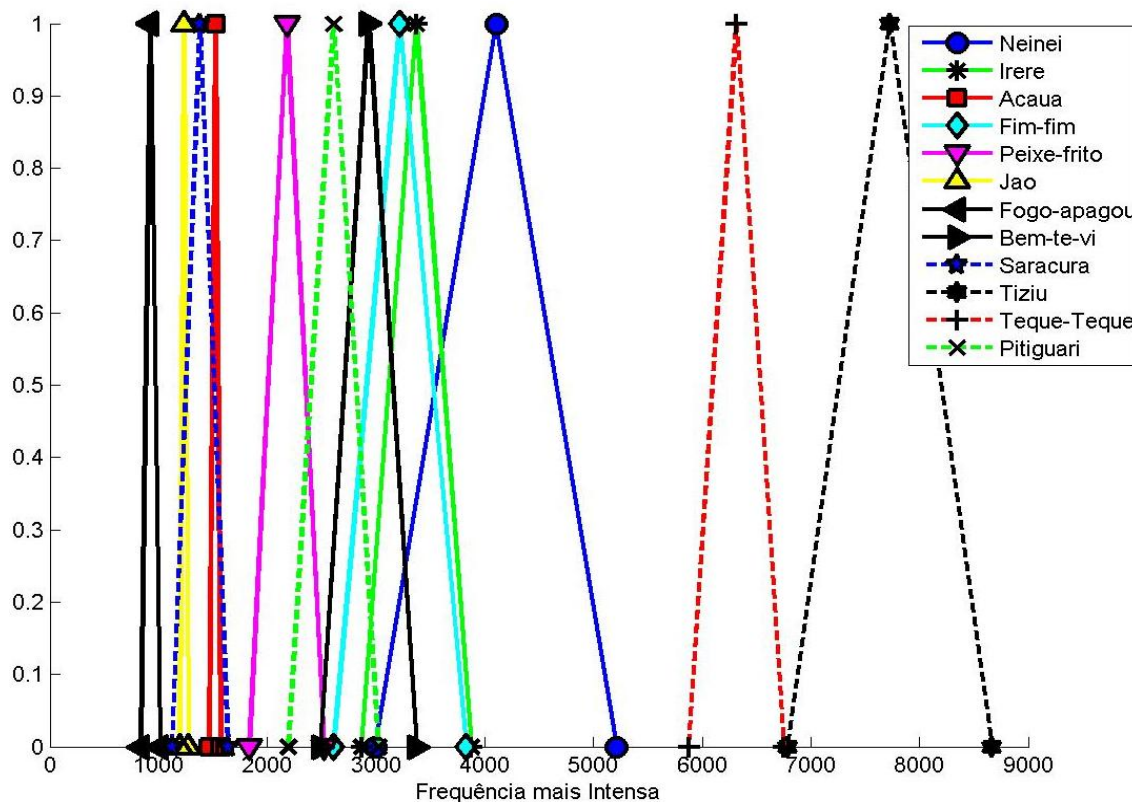


Figura 58 – Regiões de Frequência mais Intensa dos sons das doze espécies de aves em estudo.

Retira-se então do banco de dados, nesta primeira etapa, as espécies com características que permitem seu reconhecimento, neste trabalho Jaó, Saracura, Fogo-apagou, Tiziu e Teque-teque, restando ainda sete espécies. Repete-se novamente o processo e, numa segunda etapa, verifica-se que mais duas espécies podem ser reconhecidas: Acauã, pelo ZCR, Frequência máxima e mais Intensa e Peixe-frito pela Frequência máxima, como mostrado nas figuras de 59 a 61.

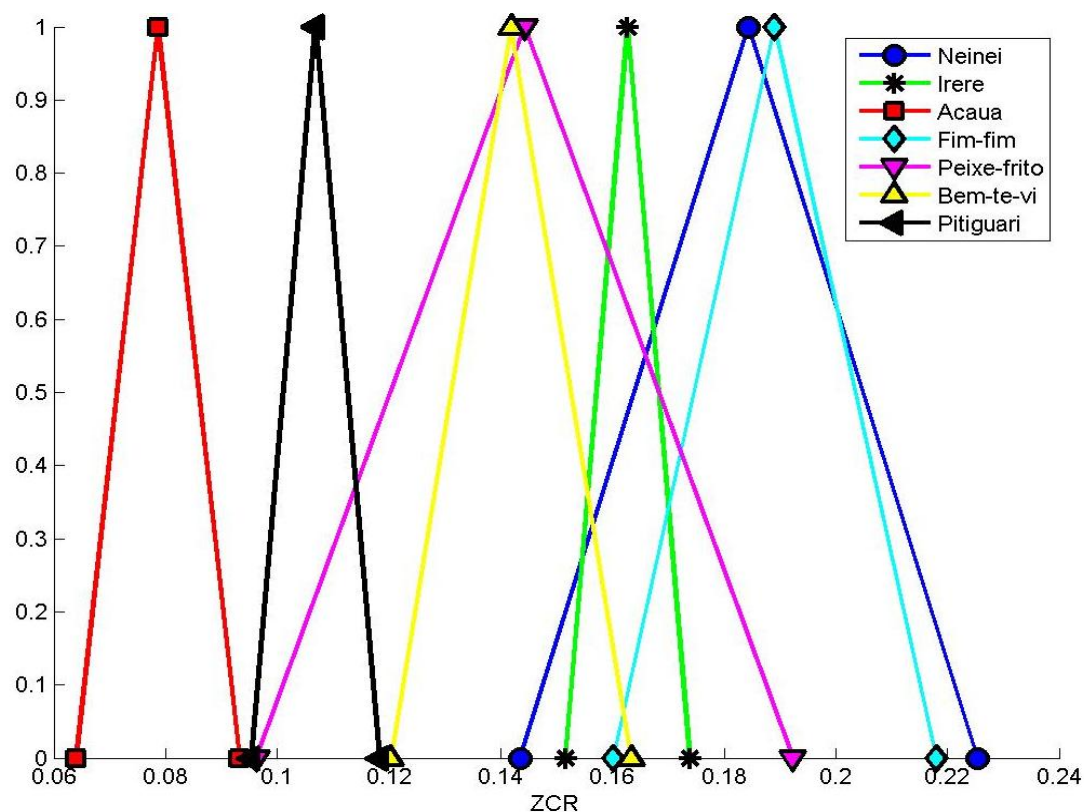


Figura 59 - Regiões de ZCR dos sons das aves na segunda etapa.

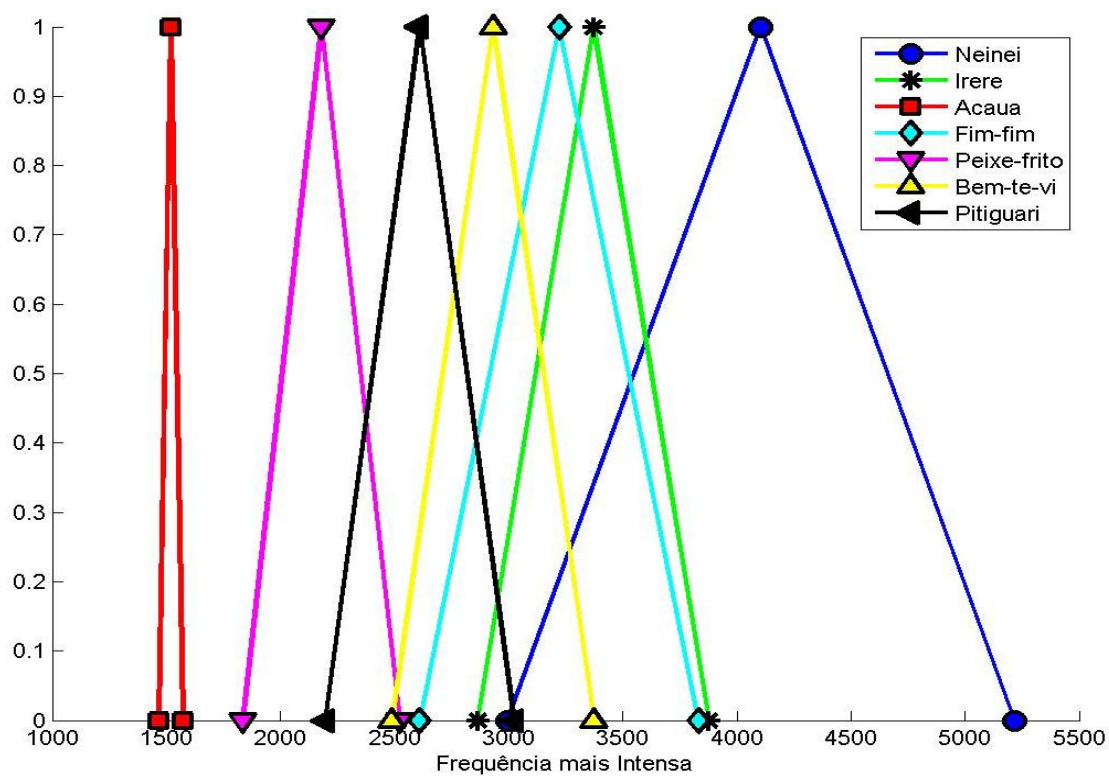


Figura 60 - Regiões de Frequência mais Intensa dos sons das aves na segunda etapa.

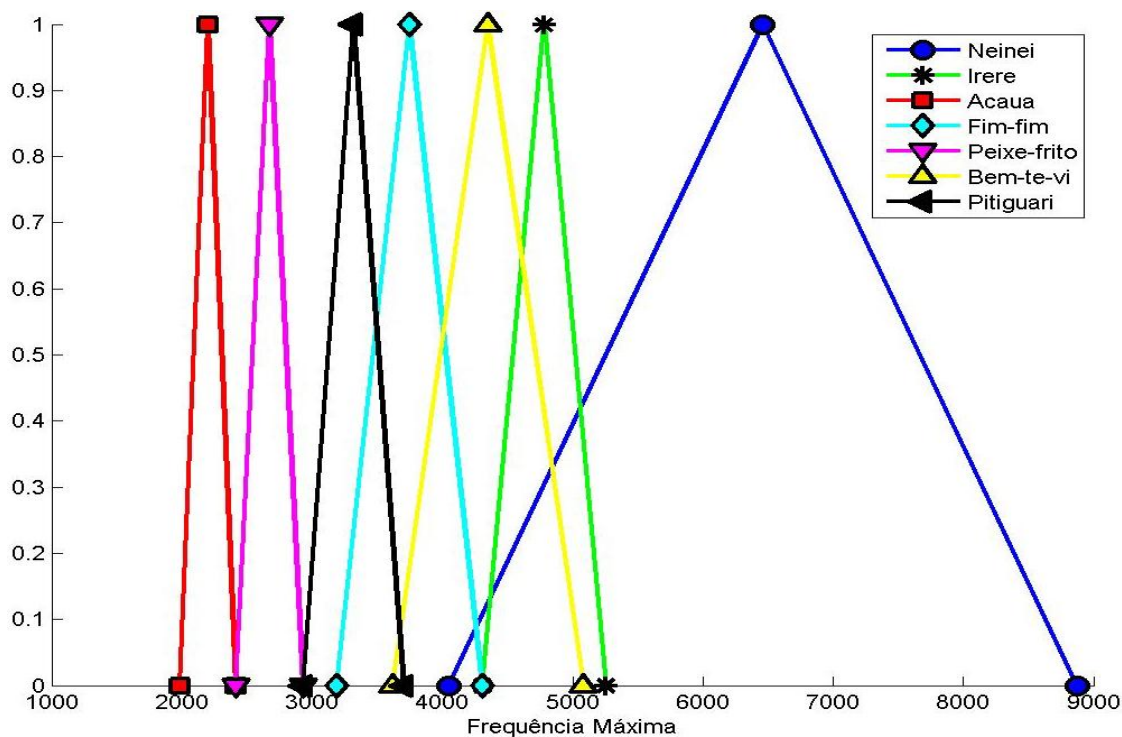


Figura 61 – Regiões de Frequência máxima dos sons das aves na segunda etapa.

Repetindo novamente o processo, numa terceira etapa, verifica-se que mais duas espécies podem ser reconhecidas: Pitiguari, pelo ZCR e o Fim-fim, pelo Fluxo Espectral, como mostra as figuras 62 e 63.

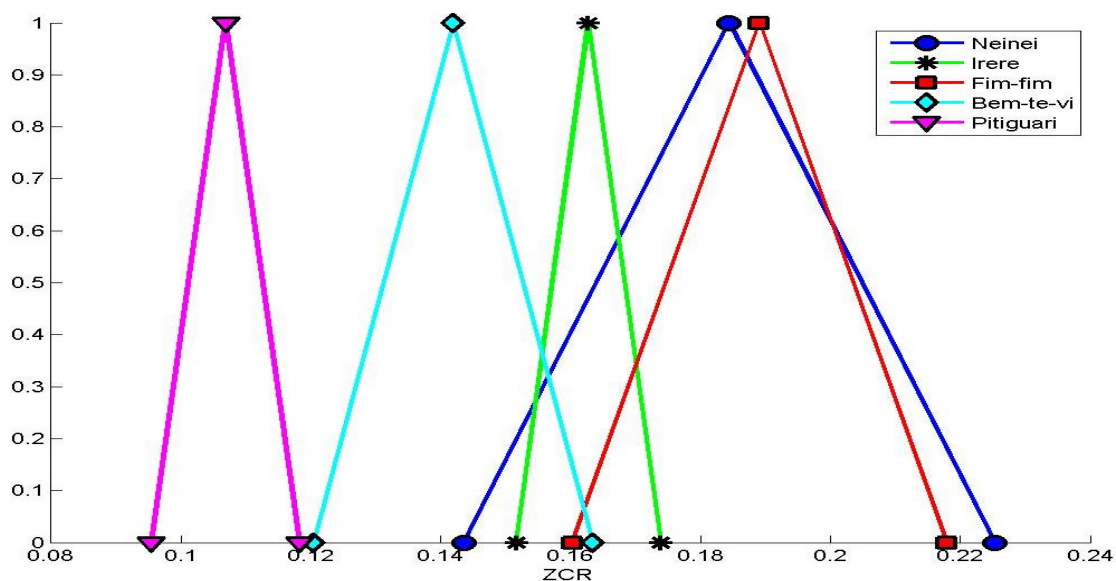


Figura 62 – Regiões de ZCR dos sons das aves na terceira etapa.

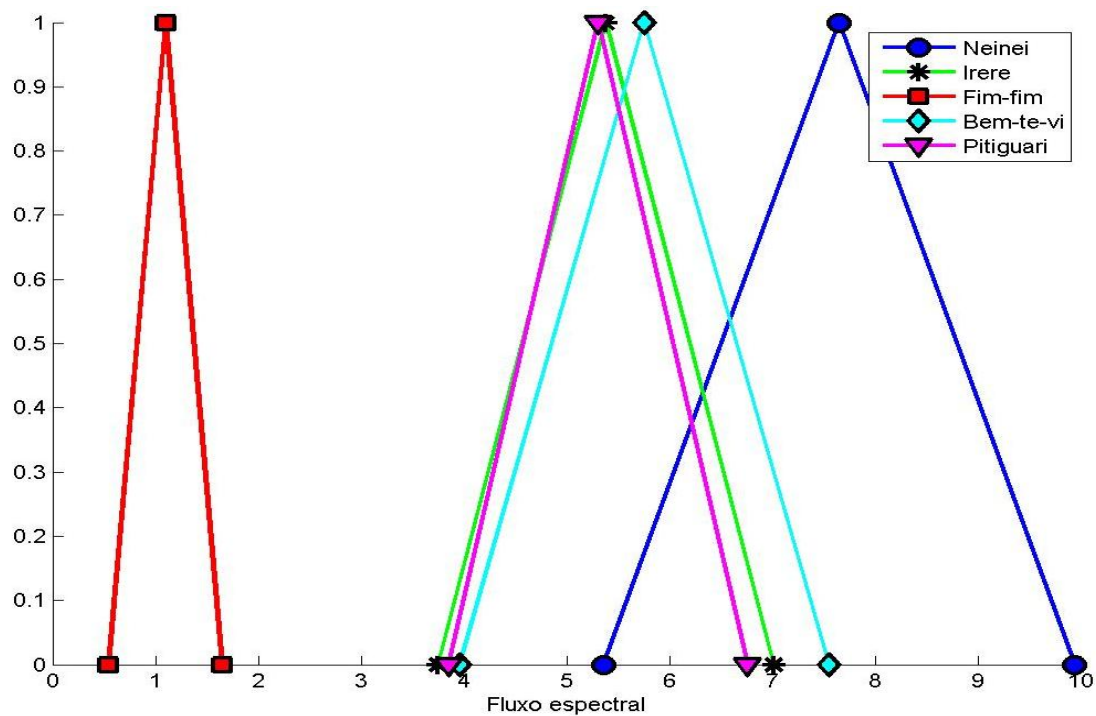


Figura 63 – Regiões de Fluxo espectral dos sons das aves na terceira etapa.

Prosseguindo com o processo, restam somente três espécies de aves e, numa quarta etapa, verifica-se que mais uma espécie pode ser reconhecida: Irerê, pela Entropia (Figura 64), restando então o Bem-te-vi e o Neinei. Para essas duas espécies de aves todas as características apresentam cruzamentos em seus gráficos, mostrando que não é possível uma separação simples das características dessas aves.

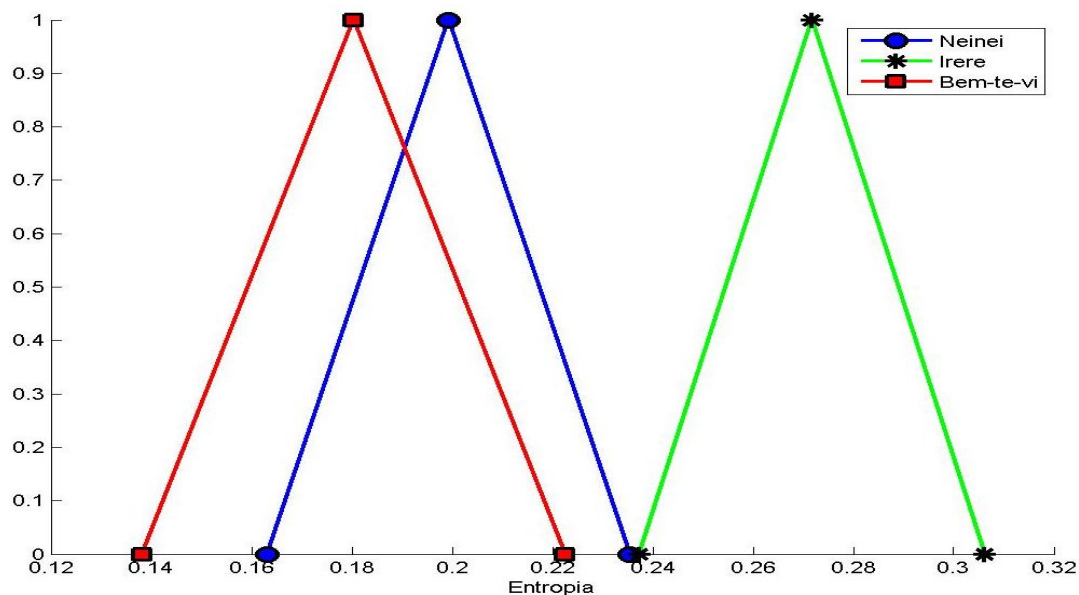


Figura 64 – Regiões de Entropia dos sons das aves na quarta etapa.

Para saber quais características utilizar para reconhecer as aves que não foram separadas, Bem-te-vi e o Neinei, foi analisada a porcentagem de não superposição (PNS) de cada característica. Essa porcentagem representa a porção daquela característica que é capaz de reconhecer somente aquela ave e foi definido da seguinte maneira: intervalo sem superposição dividido pelo intervalo da ave. As características que apresentaram maior PNS (conforme Tabela 10) são mostradas nas figuras 65 a 68.

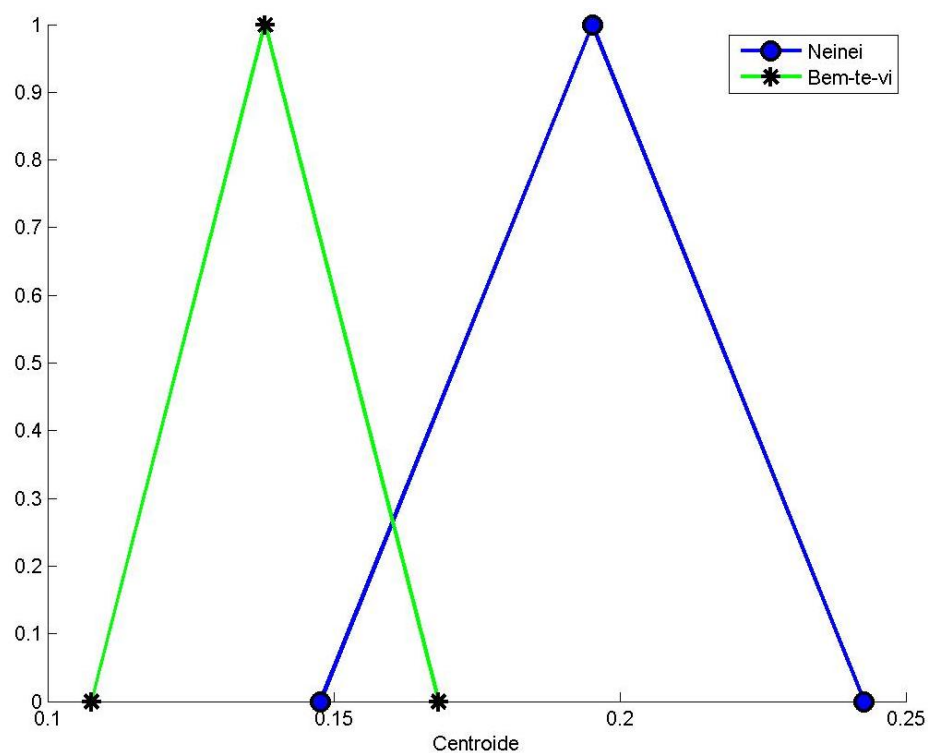


Figura 65 – Regiões de Centroide dos sons do Bem-te-vi e do Neinei.

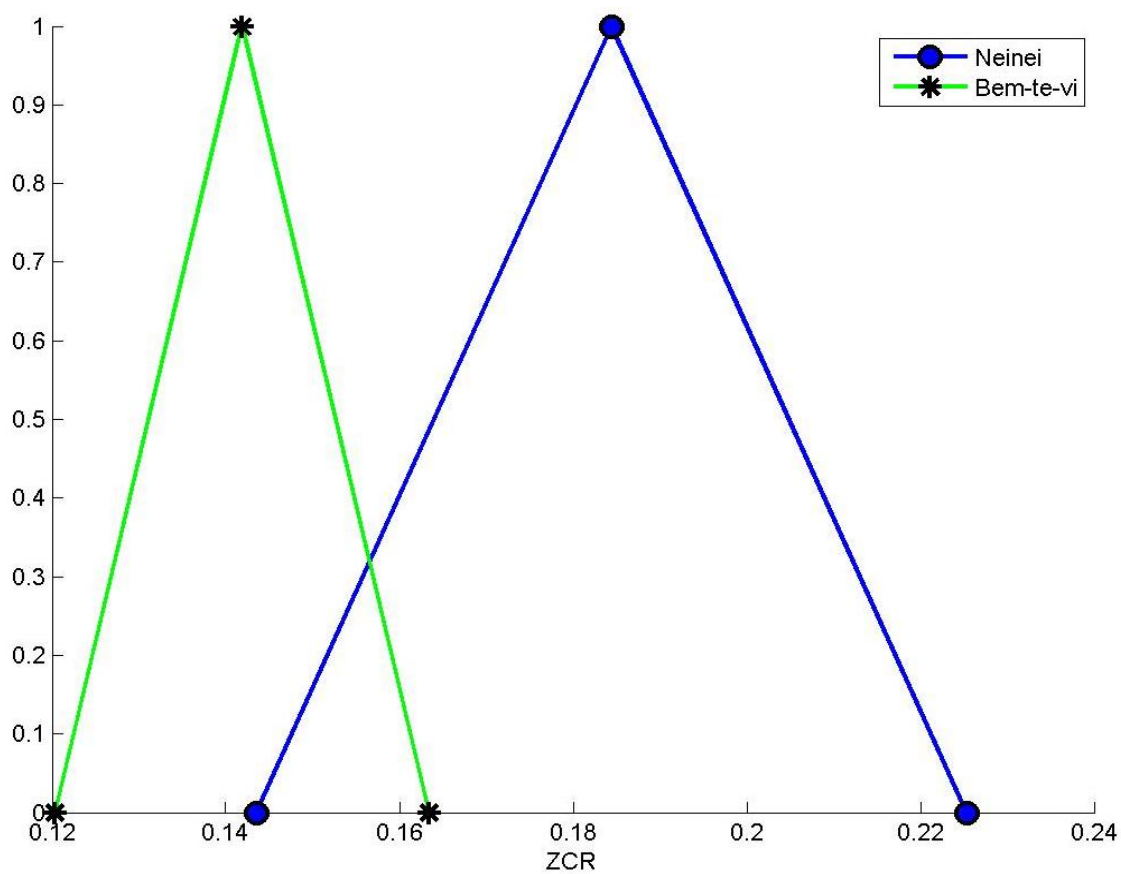


Figura 66 – Regiões de ZCR dos sons do Bem-te-vi e do Neinei.

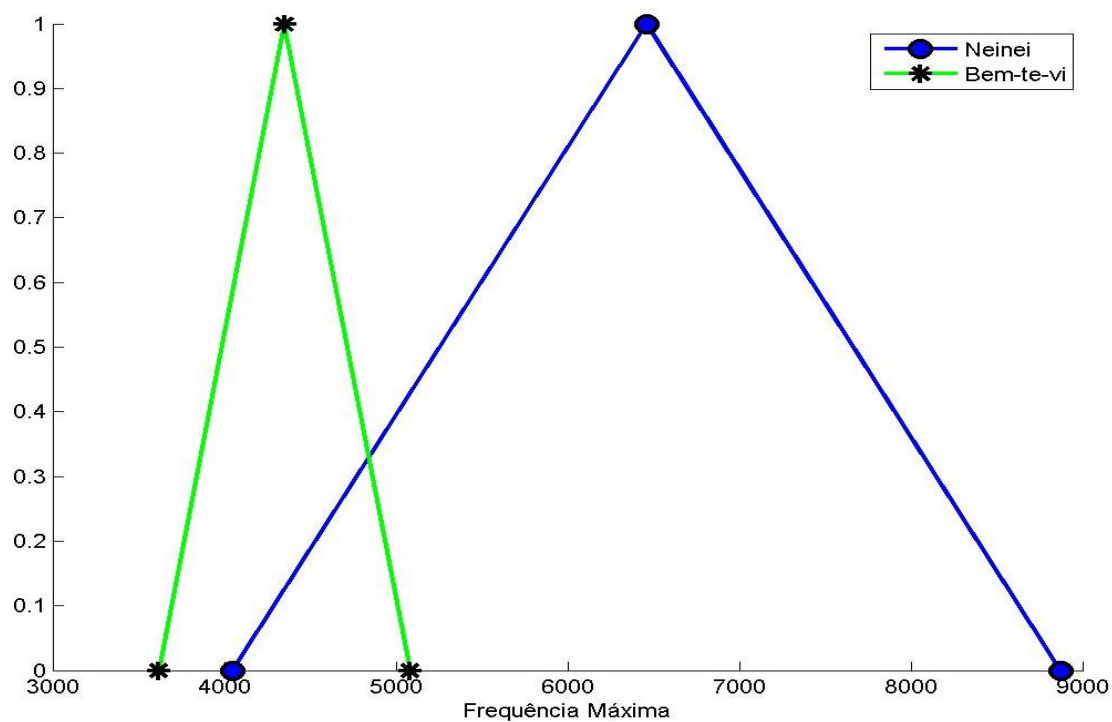


Figura 67 – Regiões de Frequência máxima dos sons do Bem-te-vi e do Neinei.

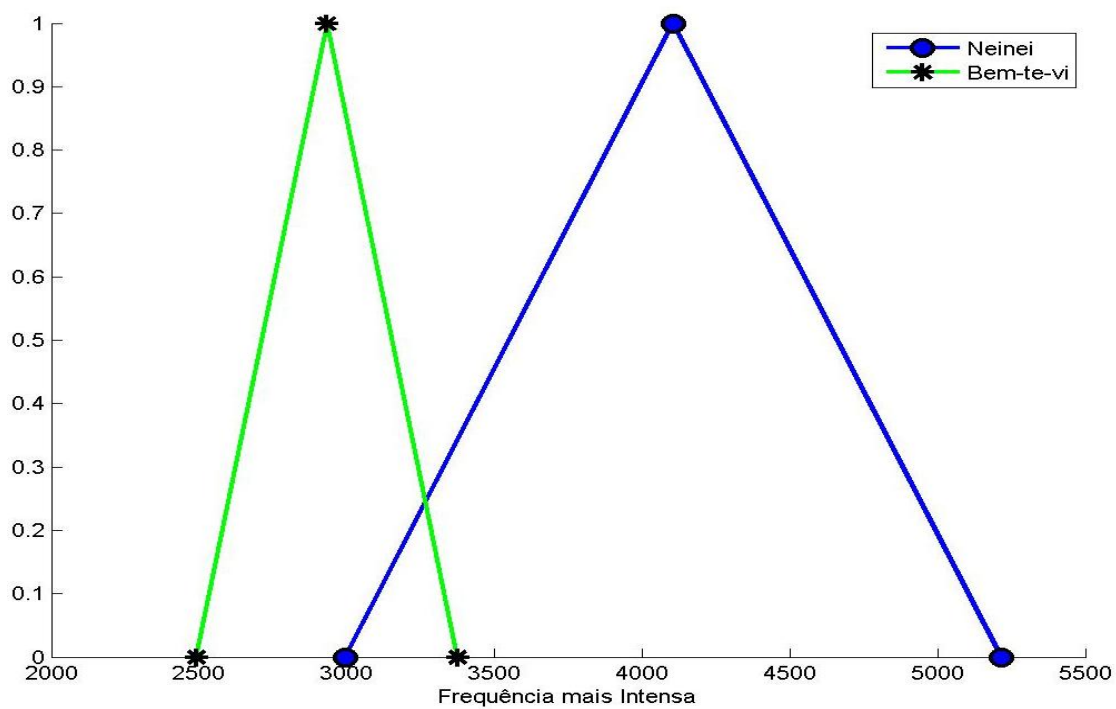


Figura 68 – Regiões de Frequência mais intensa dos sons do Bem-te-vi e do Neinei.

A Tabela 10 mostra os valores dos intervalos e do PNS de cada característica para o Bem-te-vi e o Neinei.

Tabela 10- Intervalos e valores de PNS para as características do Bem-te-vi e Neinei.

Aves	Neinei		Bem-te-vi		PNS Bem-te-vi	PNS Neinei
Centroide	0,148	0,243	0,108	0,168	66,01	78,32
Entropia	0,163	0,235	0,138	0,223	29,50	17,70
Energia	0,0184	0,052	0,022	0,0844	51,59	9,52
<i>Rolloff</i>	0,00305	0,0045	0,0031	0,0039	0	41,38
ZCR	0,144	0,225	0,120	0,163	53,72	75,67
Fluxo espec.	5,356	9,945	3,957	7,549	38,95	52,22
Freq. Max	4048	8871	3617	5081	29,44	78,58
Freq. Min	1119	3014	1033	2325	6,66	36,36
Freq. mais Int.	2994	5214	2490	3374	57,01	82,88

Foram então utilizadas para o reconhecimento do Bem-te-vi e do Neinei as características Centróide, ZCR e Frequência mais Intensa. A característica Frequência máxima foi utilizada somente para o Neinei, pois o mesmo possui um PNS alto.

Dessa maneira temos:

Tabela 11– Treinamento: características para o reconhecimento.

Aves	Características
Saracura	Entropia
Fogo-apagou	ZCR, Centróide, Freq. máxima e mais intensa
Jaó	Freq. Maxima
Tiziu	Freq. mais intensa, Centróide e ZCR
Teque-teque	Freq. mais intensa
Acauã	ZCR, Freq. máxima e mais intensa
Peixe-frito	Freq. Maxima
Pitiguari	ZCR
Fim-fim	Fluxo spectral
Irerê	Entropia
Bem-te-vi	ZCR, Centróide, Freq. mais intensa
Neinei	ZCR, Centróide, Freq. máxima e mais intensa

É importante ressaltar que os valores base para originar a tabela 12 foram detectados automaticamente e apresentam-se bem homogêneos dentro de cada espécie e distintos em relação às outras espécies diferentes. Nela pode ser observado que a maioria das características utilizadas para o reconhecimento possui valores baixos de coeficiente de variação, e aqueles que têm valores elevados estão com os valores bem separados dos demais, por exemplo, a entropia da Saracura, que tem variação de

entropia de 30,18%, mas sua entropia está em uma região bem distinta das demais espécies.

Na tabela 12 tem-se: média (μ), desvio padrão (σ), coeficiente de variação (Cv em %), fluxo espectral (FE), frequência máxima (Fmax), frequência mínima (Fmin) e frequência mais intensa (FI).

Tabela 12– *Tabela de treinamento das espécies estudadas.*

Espécie		Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
Acauã	μ	0,14	0,09	0,08	0,0042	0,11	2,20	925,3	2004,1	1502,9
	σ	0,02	0,04	0,01	0,0013	0,03	0,84	109,4	120,66	50,71
	Cv	13,9	41,2	12,4	31,8	25,2	38,1	11,82	6,02	3,37
Bem-te-vi	μ	0,18	0,04	0,14	0,0035	0,14	5,54	2041,4	4186,1	3058,8
	σ	0,03	0,02	0,01	0,0002	0,02	1,10	383,2	430,09	326,37
	Cv	15,0	50,7	10,2	5,50	11,5	19,8	18,77	10,27	10,67
Fim-fim	μ	0,23	0,12	0,17	0,0048	0,18	1,10	2973,5	4012,2	3570,5
	σ	0,06	0,03	0,02	0,0007	0,02	0,36	307,9	395,96	351,70
	Cv	24,4	29,5	9,99	15,0	10,1	32,8	10,35	9,87	9,85
Fogo-apagou	μ	0,18	0,08	0,04	0,0036	0,053	2,28	396,2	1421,2	912,51
	σ	0,02	0,02	0,0026	0,0004	0,003	0,40	60,22	45,40	58,40
	Cv	8,32	28,5	6,71	10,4	5,06	17,7	15,20	3,19	6,40
Irerê	μ	0,26	0,09	0,16	0,0040	0,17	6,15	2495,8	4606,1	3506,1
	σ	0,03	0,02	0,01	0,0003	0,01	1,18	243,58	325,19	374,38
	Cv	9,66	22,7	4,72	7,43	3,38	19,1	9,76	7,06	10,68
Jaó	μ	0,10	0,09	0,06	0,0038	0,07	0,38	766,58	1696,8	1217,8
	σ	0,01	0,03	0,0009	0,0003	0,0005	0,13	27,24	41,61	24,42
	Cv	9,91	31,4	1,69	8,09	6,49	33,1	3,55	2,45	2,01
Neinei	μ	0,19	0,03	0,18	0,00036	0,19	7,20	2319,4	5855,5	4279,6
	σ	0,03	0,01	0,02	0,0004	0,03	1,44	484,59	1194,4	642,90
	Cv	14,6	37,6	13,2	11,3	14,7	20,0	20,89	20,40	15,02
Peixe-frito	μ	0,11	0,09	0,12	0,0046	0,15	1,67	1602,1	2687,3	2140,1
	σ	0,01	0,05	0,03	0,0003	0,03	1,25	203,83	180,67	255,69
	Cv	6,71	54,0	22,3	6,34	21,3	74,8	12,72	6,72	11,95
Pitiguari	μ	0,13	0,04	0,11	0,0036	0,12	5,02	1653,8	3428,1	2610,6
	σ	0,02	0,01	0,01	0,0003	0,01	0,93	88,96	225,34	284,21
	Cv	17,5	20,0	7,05	8,61	6,89	18,6	5,38	6,57	10,89
Saracura	μ	0,02	0,02	0,06	0,0034	0,07	2,95	628,77	2230,8	1322,7
	σ	0,01	0,01	0,004	0,0002	0,01	0,61	146,68	238,32	190,54
	Cv	30,2	74,3	6,35	6,16	7,74	20,8	23,33	10,68	14,41
Teque-teque	μ	0,12	0,02	0,25	0,0052	0,26	8,35	3781,2	7114,6	6252,8
	σ	0,06	0,01	0,02	0,0005	0,02	1,58	420,89	470,37	287,37
	Cv	54,0	40,0	8,12	8,79	6,29	18,9	11,13	6,61	4,60
Tiziu	μ	0,35	0,05	0,32	0,005	0,32	10,6	4780,4	8690,8	7889,4
	σ	0,05	0,01	0,02	0,0002	0,02	2,10	1646,4	1724,3	603,50
	Cv	14,1	27,7	5,83	4,86	4,99	19,8	34,44	19,84	7,65

Na figura 69 temos que:

- Cent – Centroide;
- ZCR – ZCR;
- EE – Entropia;
- FE – Fluxo espectral;
- Fmax – Frequência máxima do canto;
- Fi – Frequência mais intensa.

É interessante notar que a energia (Fig. 33), a frequência *Rolloff* (Fig. 31) e a frequência mínima (Fig. 34) dos sinais não foram adequadas como características para classificação das espécies em estudo, pois suas regiões apresentaram muitas superposições em todas as etapas de treinamento realizadas, porém podem ser características úteis para o trabalho com outras espécies.

4.7 RECONHECIMENTO

Após o pré-processamento e o treinamento é realizado o reconhecimento, fase que utiliza a árvore de decisão gerada automaticamente no processo de treinamento. A figura 70 ilustra o processo.

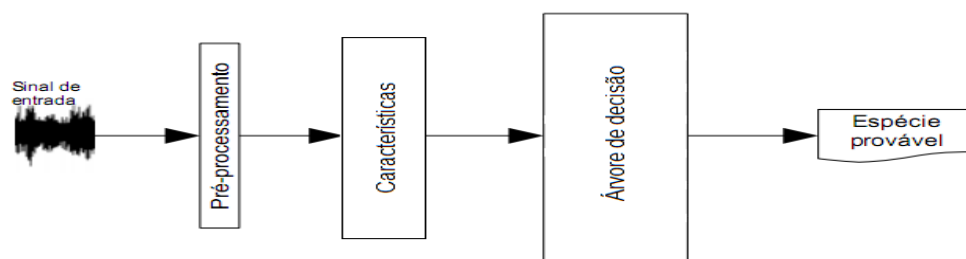


Figura 70 – Sequência para o reconhecimento de um sinal de entrada.

Pode ser observado que o sinal de entrada (som da ave do qual se deseja reconhecer a espécie) passa primeiramente pelo pré-processamento (padronização, filtragem e segmentação) e depois pela extração de características, que consiste em calcular os valores de entropia, ZCR, energia, etc. Por fim as características obtidas são comparadas com as regiões da árvore de decisão fornecendo uma classificação provável da espécie de ave.

Os resultados obtidos com essa técnica são mostrados nas tabelas 13 e 14.

Tabela 13– *Resultados obtidos na classificação das espécies.*

Aves	Quantidade Testada	Acerto (%)
Jaó	31	100
Irerê	12	83,3
Acauã	29	100
Saracura	57	87,7
Fogo-apagou	76	100
Peixe-frito	23	100
Bem-te-vi	72	81,9
Neinei	43	90,7
Pitiguari	16	100
Tiziu	18	100
Teque-teque	20	100
Fim-fim	58	100
Total	455	94,28

Observa-se pela tabela 13 que o resultado foi satisfatório, uma vez que 94,28% das espécies testadas foram reconhecidas corretamente. Tivemos como menor índice de acerto o Bem-te-vi (PITSUL) com 81,9%. Isso se justifica pelo fato do PITSUL ter suas regiões de características bem próximas das regiões das aves com a qual ela foi confundida, ou seja, o Neinei.

Na tabela 14 pode ser observado mais detalhadamente o resultado para cada tentativa de classificação, evidenciando assim os falsos positivos e falsos

negativos. Vale salientar que para as espécies com menor porcentagem de reconhecimento a classificação equivocada foi realizada para uma espécie que possui regiões de características próximas às regiões da espécie correta. Por outro lado, a maioria das espécies teve uma porcentagem ótima de acerto.

Tabela 14— *Matriz de confusão dos sinais testados*

[illegible]

5 CONCLUSÃO E TRABALHOS FUTUROS

Numa primeira fase, o pré-processamento do sinal, os sons das aves em estudo passaram por um processo de filtragem, utilizando transformada *wavelet* e filtro chebyshev, e por um processo de segmentação que utiliza a energia e o centroide do sinal. Ainda nessa fase foram extraídas as características do sinal (entropia, energia, ZCR, espectral *rolloff*, fluxo espectral, centroide, frequência máxima, mínima e mais intensa). Observou-se que essas características são mais representáveis para sinais com menos ruídos, portanto nos casos em que a filtragem não foi muito eficiente o som não atingiu a qualidade necessária para ser utilizado nesta pesquisa. Mesmo assim as características extraídas dos sinais foram consideradas eficientes por haver permitido um bom índice de reconhecimento das espécies em estudo. Outra característica positiva é que não foram utilizadas todas as características para o reconhecimento, o que diminui o processamento.

Na etapa seguinte, a fase de treinamento, foi gerada automaticamente uma árvore de decisão com as características extraídas de 10 aves de cada espécie. As características energia e frequência mínima não se mostraram adequadas para a classificação das espécies de aves em estudo.

A próxima etapa, a fase de reconhecimento, utilizou a árvore de decisão obtida na etapa anterior. A metodologia se mostrou eficaz, já que permitiu uma classificação correta das espécies de 94,28% das aves estudadas. As menores porcentagens de acerto ocorreram para espécies de aves em que os valores calculados de suas características estavam levemente fora dos valores das regiões

obtidas na fase de treinamento. Isso mostra que na fase de treinamento as aves utilizadas não foram suficientes para fornecer regiões de características mais abrangentes, realmente representando uma porcentagem maior de aves dessa espécie.

O uso da árvore de decisão mostra-se uma técnica eficiente quando o número de espécies a ser reconhecida é pequeno, já que um maior número de espécies causa uma maior sobreposição de regiões nos gráficos das características, tornando muito difícil a construção automática da árvore.

Como sugestão para trabalhos futuros, recomenda-se utilizar outras técnicas de reconhecimento, como por exemplo, as Redes Neurais Artificiais (RNA) ou cadeia de Markov, pois estas podem ser usadas de forma mais generalizada. Outra sugestão seria a determinação de mais características do sinal de entrada e a otimização destas características, analisando qual melhor se encaixa em determinada técnica de reconhecimento.

Como maior contribuição espera-se que este trabalho incentive diversas outras pesquisas relacionadas ao reconhecimento de aves e que, de alguma forma, possa auxiliar na monitoração e preservação de aves, pois o número de espécies em ameaça de extinção aumenta a cada ano.

6 BIBLIOGRAFIA

ANDERSON, S. E., DAVE, A. S. & MARGOLIASH, D. Template-based automatic recognition of birdsong syllables from continuous recordings, *J. Acoust. Soc. Am.* v. 100, n. 2, p. 1209–1219, 1996.

ANDRADE, M. A. Atração e alimentação de aves silvestres. Belo Horizonte: Littera Maciel, 1995.

ARGOUD, F. I. M.. Tese de Doutorado. Contribuição à Automatização da Detecção e Análise de Eventos Epileptiforme. Florianópolis: UFSC, 2001.

AURELIO, D. Desvendando o MP3. São Paulo: Digerati Books, 2004. ISBN 85-89535-35-5.

BAPTISTA, F. L.; GAUNT, S. L. L. Advances in studies of avian sound communication. *The Condor* 96, p. 817-830, 1994.

BIRDLIFE. Birdlife International – working together for birds and people. Birdlife International, 2011. Disponível em: < <http://www.birdlife.org/index.html>> Acesso em: 20/08/2012.

BRACKENBURY, J. H. Functions of the syrinx and the control of sound production, in ‘(King & McLelland 1989)’, ch. 4, p. 193-220., 1989

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A., & STONE, C. J. (1984). Classification and Regression Trees. Wadsworth.

CEO. CEO – Gravando vozes de aves. CEO, 2011. Disponível em: <<http://www.ceo.org.br/observ/bioac.htm>>. Acesso em: 12/08/2012.

CHEN, Z.; MAHER, R. C. Semi-automatic classification of bird vocalizations using spectral peak tracks. *J. Acoust. Soc. Am.* v. 120, n. 5, p. 2974-2984, 2006.

CONCEIÇÃO, P. F. Dissertação de Mestrado: Reconhecimento Automático de Aves da Família Tinamidae Através da Vocalização. Goiânia, 2012.

DAUBECHIES, I. Ten lectures on wavelets. Philadelphia: SIAM Books, 1992.

- DAUBECHIES, I. The wavelet transform, time frequency localization and signal analysis. IEEE Trans. on Information Theory, New York, v. 36, p. 961-1005, 1997.
- DINIZ, P. S. R., SILVA, E.A.B., NETTO, S. L. (2004). Processamento Digital de Sinais: Projeto e Análise de Sistemas. Editora Bookman. Porto Alegre. (p. 419-432)
- DONOHU, D. L., (1995) De-noising by soft thresholding, IEEE Trans. Inform. Theory, vol. 41, pp. 613–627.
- DUDA, R.O.; HART, P.E.; STORK, D.G. Pattern Classification. Segunda edição. USA: Wiley-Interscience, 2001.
- FONSECA, J. (1994). Indução de árvores de decisão. Tese de Mestrado, Lisboa.
- FONSECA, M. S. Dissertação de Mestrado. Combinando Imagem e Som para Detecção de Transições em Vídeos Digitais. Niterói: UFF, 2006.
- GAMA, João Manuel Portela da. Combining classification algorithms. 1999. 195p. Tese (Doutorado) – Departamento de Ciência de Computadores, Universidade do Porto, Porto.
- GARCIA, E. V. Dissertação de Mestrado: Instrumentação para a monitoração das alterações eletrocardiográficas decorrentes da hipoglicemia. Florianópolis: UFSC, 2001.
- GIANNAKOPOULOS, T. D. PhD Dissertation. Study and application of acoustic information for the detection of harmful content, and fusion with visual information. Atenas: 2009.
- GRAPS, A. An Introduction to wavelets. IEEE Comput. Sci. and Eng., v. 2 n. 2, 1995. Disponível em: <<http://www.cis.udel.edu/~amer/CISC651/IEEEwavelet.pdf>>. Acesso em: 20/01/2013.
- GRAPS, A. Na introduction to wavelets. IEEE, 1995.
- HÄRMÄ, A. & SOMERVUO, P. Classification of the harmonic structure in bird vocalization, in ‘IEEE Int. Conf. Acoust. Speech and Signal Processing’, ICASSP2004, pp. V701-V704, 2004.
- HÄRMÄ, A. Automatic identification of bird species based on sinusoidal modeling of syllables, in ‘IEEE Int. Conf. Acoust. Speech and Signal Processing’, p. v545-v548, 2003.

- HAYES, M. H. Processamento Digital de Sinais. São Paulo: Bookman, 2006.
- HAYKIN, S.; VEEN, B. V. Sinais e Sistemas. Porto alegre: Bookman, 2001.
- HE, F., SUN, Y.E., (2006). Glial Cells More Than Support Cells, International Journal of Biochemistry and Cells.
- HERNANDES, E.; WEISS, G. A first course on wavelet. Boca Raton: CRC Press, 1996.
- KING, A. S.; MCLELLAND, J. eds. Form and Function in Birds, Vol. 4, Academic Press, 1989.
- KREBS, J. R.; KROODSMA, D. E. Repertoires and geographical variation in BIRD song, Adv. Study Behav. v. 11, p. 143-177, 1980.
- KROODSMA, D. E; BYERS, B. E. The function of birds song. Am. Zool. v. 31, p. 318-328, 1991.
- KWAN, C.; MEI, G.; ZHAO, X.; REN, Z.; XU, R.; STANFORD, V.; ROCHET, C.; AUBE, J.; HO, K. Bird classification algorithms: Theory and experimental results. ICASSP2004, p. V289-V292, 2004.
- LIMA, P. C. Wavelets: teoria, algoritmos e aplicações. Belo Horizonte: Departamento de Matemática-ICEX – UFMG, 2004.
- MARQUES, A. B. Tese de doutorado. Avaliação do canto do Trinca-Ferro em Relação ao Processo de Domesticação das Aves Canoras. Campos dos Goytacazes: UENF, 2009.
- MCLLRAITH, A. L. & CARD, H. C. Birdsong recognition using backpropagation and multivariate statistics, IEEE Trans. Signal Processing. v. 45, n. 11, p. 2740–2748, 1997.
- MISITI, M., MISITI, Y., OPPENHEIM, G., POGGI, J.M..(2006).Wavelet Toolbox for use with Matlab. The Math Works Inc. Natick, Massachusetts.
- MORETTIN, P. A. Ondas e ondaletas: da análise de fourier à análise de ondaletas. 2ed. São Paulo:Edusp, 1999.
- NELSON, D. A. The importance of invariant and distinctive features in species recognition of bird song, Condor 91, p. 120–130, 1989.

- NOBRE, F. S. Dissertação de Mestrado: Estudo Comparativo de Métodos de Segmentação em Imagens Digitais de Aves. Goiânia: UFG 2012.
- NOGUEIRA, A; AZEVEDO, J; BAPTISTA, V; SIQUEIRA, S. Um Overview Sobre Reconhecimento de Padrões. Rio de Janeiro: Resende, 2006.
- OPPENHEIM, A. V; SCHAFER, R. W. Discrete-Time signal Processing. Second. New Jersey: Prentice Hall, 1999. ISBN 0-13-754920-2.
- PATRICK, E. A. Fundamentals of Pattern Recognition. Primeira edição. USA: Prentice-Hall, 1972. São Paulo: Edusp, 1999.
- SAVE. Sociedade para Conservação das Aves do Brasil – Save Brasil, 2012. Disponível em: <<http://www.savebrasil.org.br/?q=content/por-que-aves>>. Acesso em: 11/09/2012.
- SELIN, A.; TURUNEN, J.; TANTTU, J. T. Wavelets in recognition of birds sounds. EURASIP Journal on Applied Signal Processing, v. 2007 p. 1-9, 2007.
- SICK, H. Ornitologia Brasileira, uma introdução; Brasília, Ed. Univ. de Brasília, v. I e II, 1985.
- SIQUEIRA, M. L. Tese de Doutorado. Reconhecimento automático de padrões em imagens Ecocardiográficas. Porto Alegre: UFRGS, 2010.
- SMITH, S. W. The Scientist and Engineer's Guide to Digital Processing. San Diego: California Technical Publishing, 1999. ISBN 0-9660176-6-8.
- SOARES, W. C. Tese de Doutorado. Um Método Não-Limiar para Redução de Ruído em sinais de Voz no Domínio Wavelet. Ilha Solteira: UNESP, 2009.
- SOARES, W. C.; DUARTE, M. A. Q.; VILLARREAL, F.; VIEIRA FILHO, J. Wavelets: teoria e aplicações. In: BRAZILIAN CONFERENCE ON DYNAMICS, CONTROL AND THEIR APPLICATIONS, DINCON'2005, 6, 2005.
- THEODORIDIS, S.; KOUTROUMBAS, K., Pattern Recognition, 4nd Ed., Elsevier, San Diego, 2009.
- TZANETAKIS, GEORGE. Musical Genre Classification of áudio Signals. IEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL.10. NO.5, July 2002.
- VIELLIARD, J. Uso da bioacústica na observação das aves. II Encontro Nac. Anilhad. Aves, Rio de Janeiro, p. 98-121, 1987.

WIKIAVES. Wikiaves, a Enciclopédia das aves do Brasil. Wikiaves, 2012. Disponível em: <<http://www.wikiaves.com.br/>>. Acesso em: 20/01/2013.

WIKIPEDIA. MP3. Wikipedia, 2012. Disponível em: <<http://pt.wikipedia.org/wiki/MP3>>. Acesso em: 12/08/2012.

XENO. Xeno-canto – sharing bird from around the world. Xeno-canto, 2012. Disponível em: <<http://www.xeno-canto.org/>>. Acesso em: 20/01/2013.

7 ANEXOS

Em anexo estão as tabelas com os valores de cada característica, para cada espécie, utilizada no treinamento, nas tabelas temos: média, desvio padrão, coeficiente de variação (em %), fluxo espectral (FE), frequência máxima (Fmax), frequência mínima (Fmin) e frequência mais intensa (FI).

Tabela 15 – *Características extraídas dos sons do Acauã utilizados na fase de treinamento.*

Ave	Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
ACAUABA	0.1339	0.0730	0.0933	0.0042	0.1442	3.0102	1034	2153	1533
ACAUACE	0.1191	0.1222	0.0647	0.0033	0.0783	2.4924	947	1981	1485
ACAUAMG9	0.1434	0.0623	0.0848	0.0040	0.1382	3.1959	689	1895	1496
ACAUAMG	0.1234	0.0767	0.0770	0.0037	0.0995	2.2293	775	1981	1471
ACAUARJ	0.1561	0.1033	0.0713	0.0041	0.1130	1.0082	947	1981	1571
ACAUASC	0.1437	0.0698	0.0863	0.0046	0.1554	3.0887	947	1981	1512
ACAUASP3	0.1726	0.0590	0.0717	0.0033	0.0855	1.8066	947	2153	1547
ACAUASP37	0.1100	0.1576	0.0788	0.0077	0.1043	0.8201	984	1781	1398
ACAUASP	0.1467	0.0408	0.0640	0.0030	0.0754	1.7722	947	1981	1469
ACAUATO	0.1608	0.1265	0.0724	0.0040	0.1132	2.5686	1034	2153	1547
Média	0.1410	0.0891	0.0764	0.0042	0.1107	2.1992	925	2004	1503
Desvio	0.0196	0.0367	0.0095	0.0013	0.0279	0.8381	109	121	51
Cof. Variação	13.91	41.19	12.44	31.83	25.17	38.11	11.82	6.02	3.37

Tabela 16 - *Características extraídas dos sons do Bem-te-vi utilizados na fase de treinamento.*

Ave	Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
BEM-TE-VIARG1	0.2226	0.0309	0.1341	0.0035	0.1391	6.2862	1809	4393	3274
BEM-TE-VICOL2	0.1700	0.0216	0.1229	0.0031	0.1323	5.0483	1723	3962	2490
BEM-TE-VIGO	0.1822	0.0284	0.1467	0.0039	0.1375	5.3185	2326	4307	3366
BEM-TE-VIMG2	0.1382	0.0358	0.1204	0.0034	0.1077	5.0848	2067	3618	2671
BEM-TE-VIPR	0.2176	0.0698	0.1633	0.0035	0.1683	6.6965	2498	5082	3374
BEM-TE-VIRJ1	0.1488	0.0264	0.1510	0.0036	0.1508	4.2644	2153	4221	3099
BEM-TE-VIRS2	0.1898	0.0844	0.1362	0.0035	0.1306	7.5487	1981	4393	3277
BEM-TE-VISP2	0.1778	0.0277	0.1464	0.0036	0.1407	5.2272	2326	4307	3366
BEM-TE-VISP3	0.1861	0.0350	0.1286	0.0035	0.1329	3.9570	2326	3618	2854
BEM-TE-VISP5	0.1623	0.0513	0.1251	0.0034	0.1265	5.9415	1206	3962	2816
Média	0.1795	0.0411	0.1374	0.0035	0.1366	5.5373	2041	4186	3059
Desvio	0.0269	0.0208	0.0140	0.0002	0.0158	1.0964	383	430	326
Cof. Variação	14.97	50.65	10.19	5.50	11.54	19.80	18.77	10.27	10.67

Tabela 17 - Características extraídas dos sons do Fim-fim utilizados na fase de treinamento.

Ave	Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
FIMFIMBA	0.2237	0.0914	0.1601	0.0043	0.1562	1.3037	2929	4048	3587
FIMFIMCE	0.1997	0.1260	0.1732	0.0050	0.1911	0.5467	3273	4307	3811
FIMFIMMG18	0.1632	0.0831	0.1639	0.0045	0.1755	0.5819	3015	4048	3633
FIMFIMMG	0.2212	0.1073	0.1629	0.0052	0.1744	1.2414	2188	2938	2611
FIMFIMMS	0.2016	0.1805	0.1702	0.0044	0.1823	0.7603	3187	4221	3778
FIMFIMRS10	0.2010	0.0640	0.1727	0.0046	0.1828	1.1550	3273	4307	3832
FIMFIMRS7	0.1824	0.1276	0.1648	0.0047	0.1760	1.1431	3015	4134	3662
FIMFIMSC	0.2467	0.1058	0.1633	0.0045	0.1735	1.1604	2929	4134	3641
FIMFIMSP1	0.3139	0.1557	0.2180	0.0066	0.2262	1.6500	3000	3938	3526
FIMFIMSP8	0.3368	0.1216	0.1619	0.0041	0.1713	1.4144	2929	4048	3625
Média	0.2290	0.1163	0.1711	0.0048	0.1809	1.0957	2974	4012	3571
Desvio	0.0559	0.0343	0.0171	0.0007	0.0183	0.3592	308	396	352
Cof. Variação	24.41	29.49	9.99	14.97	10.11	32.78	10.35	9.87	9.85

Tabela 18 - Características extraídas dos sons do Fogo-apagou utilizados na fase de treinamento.

Ave	Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
FOGO-APAGOUARG	0.2122	0.1110	0.0364	0.0040	0.0517	1.6471	431	1378	860
FOGO-APAGOU DFA	0.1739	0.0736	0.0415	0.0033	0.0552	2.1993	431	1464	984
FOGO-APAGOU GO	0.1890	0.0617	0.0371	0.0031	0.0539	2.2270	345	1378	914
FOGO-APAGOU GO2	0.1640	0.0483	0.0416	0.0034	0.0550	2.6095	431	1464	862
FOGO-APAGOU MG	0.1890	0.0617	0.0370	0.0031	0.0539	2.2265	345	1378	914
FOGO-APAGOU MG3	0.1907	0.0702	0.0350	0.0039	0.0501	2.6009	258	1378	841
FOGO-APAGOU MG4	0.1809	0.1193	0.0392	0.0039	0.0549	2.0710	431	1464	916
FOGO-APAGOU MT	0.1649	0.0881	0.0367	0.0041	0.0480	1.7255	431	1378	862
FOGO-APAGOU MT2	0.1696	0.0936	0.0426	0.0037	0.0564	2.9031	431	1464	957
FOGO-APAGOU VEN1	0.1953	0.1044	0.0383	0.0038	0.0555	2.6054	431	1464	1016
Média	0.1829	0.0832	0.0386	0.0036	0.0535	2.2815	396	1421	913
Desvio	0.0152	0.0237	0.0026	0.0004	0.0027	0.4041	60	45	58
Cof. Variação	8.32	28.51	6.71	10.41	5.06	17.71	15.20	3.19	6.40

Tabela 19 - Características extraídas dos sons do Irerê utilizados na fase de treinamento.

Ave	Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
IRERECE_2	0.3061	0.1104	0.1639	0.0036	0.1690	7.0089	2498	4737	3781
IREREDF_3	0.2858	0.0687	0.1739	0.0042	0.1767	6.2995	2326	5254	3393
IREREPR4_5	0.2372	0.0807	0.1561	0.0038	0.1681	7.0112	3000	4594	3878
IREREPR8_6	0.2483	0.0823	0.1596	0.0045	0.1677	3.7484	2326	4307	3838
IRERERS_7	0.2617	0.1049	0.1543	0.0038	0.1603	5.5413	2326	4479	2865
IRERESC_8	0.2677	0.1058	0.1559	0.0041	0.1646	6.6899	2412	4307	3579
IRERESP8_9	0.2384	0.0597	0.1517	0.0042	0.1605	6.7576	2584	4565	3210
Média	0.2636	0.0875	0.1593	0.0040	0.1667	6.1510	2496	4606	3506
Desvio	0.0255	0.0199	0.0075	0.0003	0.0056	1.1762	244	325	374
Cof. Variação	9.66	22.70	4.72	7.43	3.38	19.12	9.76	7.06	10.68

Tabela 20 - Características extraídas dos sons do Jaó utilizados na fase de treinamento.

Ave	Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
JAOMAM	0.1045	0.1171	0.0559	0.0039	0.0738	0.3861	689	1723	1207
JAOBOL1	0.1034	0.0807	0.0561	0.0040	0.0722	0.7046	775	1723	1199
JAOBOL3	0.0931	0.0579	0.0571	0.0041	0.0757	0.3978	775	1723	1248
JAOGO	0.1001	0.1015	0.0547	0.0037	0.0740	0.2166	775	1637	1197
JAOMGA	0.1219	0.0976	0.0570	0.0043	0.0741	0.3385	775	1723	1275
JAOMT2	0.1155	0.0680	0.0540	0.0039	0.0746	0.3277	775	1637	1215
JAOMT3	0.0866	0.1133	0.0555	0.0036	0.0694	0.3069	775	1637	1205
JAOMT5	0.1004	0.0384	0.0559	0.0034	0.0695	0.3871	775	1723	1213
JAOMT7	0.0974	0.0824	0.0561	0.0033	0.0644	0.3614	775	1723	1210
JAOPER1	0.1019	0.1234	0.0561	0.0038	0.0621	0.3720	775	1723	1210
Média	0.1025	0.0880	0.0558	0.0038	0.0710	0.3799	767	1697	1218
Desvio	0.0102	0.0276	0.0009	0.0003	0.0046	0.1259	27	42	24
Cof. Variação	9.91	31.35	1.69	8.09	6.49	33.14	3.55	2.45	2.01

Tabela 21 - Características extraídas dos sons do Neinei utilizados na fase de treinamento.

Ave	Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
NEINEICE1	0.1737	0.0184	0.1892	0.0039	0.1867	7.6417	2153	5685	4228
NEINEIMG1	0.2299	0.0337	0.1639	0.0033	0.1598	8.9481	1120	5857	3447
NEINEIMG	0.2276	0.0520	0.1982	0.0037	0.1963	7.0373	2584	5599	4681
NEINEIRS1	0.1688	0.0241	0.1667	0.0031	0.1704	5.4344	2250	5531	4544
NEINEIRS	0.1631	0.0364	0.2144	0.0038	0.2126	6.4322	2498	6115	5215
NEINEISC10	0.1838	0.0349	0.1782	0.0034	0.1739	7.2955	2438	5531	4723
NEINEISC	0.2354	0.0512	0.2253	0.0045	0.2427	9.9451	2412	8872	4121
NEINEISP25	0.2035	0.0205	0.1436	0.0036	0.1477	5.3560	2239	4048	2994
NEINEISP8	0.1851	0.0292	0.1887	0.0037	0.1984	6.3652	3015	5771	4379
NEINEISP24	0.1658	0.0207	0.1768	0.0033	0.1786	7.5467	2486	5546	4465
Média	0.1937	0.0321	0.1845	0.0036	0.1867	7.2002	2319	5856	4280
Desvio	0.0283	0.0121	0.0243	0.0004	0.0274	1.4433	485	1194	643
Cof. Variação	14.62	37.64	13.16	11.26	14.70	20.04	20.89	20.40	15.02

Tabela 22 - Características extraídas dos sons do Peixe-frito utilizados na fase de treinamento.

Ave	Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
PEIXE-FRITOMT	0.1165	0.0494	0.1922	0.0048	0.2294	4.7211	1378	2412	1835
PEIXE-FRITOMT1	0.1037	0.1644	0.1199	0.0048	0.1599	1.9139	1464	2584	2035
PEIXE-FRITOPR	0.1153	0.1465	0.1131	0.0048	0.1575	2.0644	1464	2584	2005
PEIXE-FRITOPR1	0.1091	0.1655	0.1214	0.0046	0.1597	2.5616	1464	2584	1957
PEIXE-FRITOPR2	0.1047	0.1076	0.0963	0.0047	0.1469	0.9433	1464	2584	1930
PEIXE-FRITOPR3	0.1052	0.0948	0.1196	0.0049	0.1668	1.3400	1464	2584	1930
PEIXE-FRITOSP	0.0945	0.0491	0.1049	0.0045	0.1355	0.7493	1809	2842	2388
PEIXE-FRITOSP1	0.1000	0.0292	0.1092	0.0043	0.1229	1.0846	1723	2929	2522
PEIXE-FRITOSP2	0.1077	0.0578	0.1115	0.0041	0.1234	0.7416	1895	2929	2374
PEIXE-FRITOSP3	0.0977	0.0721	0.1069	0.0043	0.1188	0.6244	1895	2842	2425
Média	0.1054	0.0937	0.1195	0.0046	0.1521	1.6744	1602	2687	2140
Desvio	0.0071	0.0505	0.0267	0.0003	0.0324	1.2517	204	181	256
Cof. Variação	6.71	53.96	22.34	6.34	21.29	74.76	12.72	6.72	11.95

Tabela 23 - Características extraídas dos sons do Pitiguari utilizados na fase de treinamento.

Ave	Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
PITIGUARIAM	0.1515	0.0527	0.1043	0.0034	0.1151	5.8198	1550	3445	2390
PITIGUARIDF	0.1544	0.0375	0.1129	0.0038	0.1213	5.8873	1637	3704	2867
PITIGUARIGO1	0.1606	0.0331	0.1148	0.0041	0.1152	4.2384	1637	3618	2196
PITIGUARIPR	0.1044	0.0411	0.1148	0.0035	0.1240	3.8547	1723	3359	2725
PITIGUARIPR4	0.1150	0.0538	0.1154	0.0038	0.1241	4.9622	1723	3531	3024
PITIGUARISC	0.1074	0.0398	0.1162	0.0032	0.1242	4.3861	1809	3445	2946
PITIGUARISP	0.1020	0.0385	0.1153	0.0034	0.1250	4.3753	1637	3531	2638
PITIGUARISP2	0.1244	0.0401	0.1006	0.0034	0.1090	6.7590	1550	3187	2434
PITIGUARISP5	0.1495	0.0615	0.1183	0.0040	0.1294	5.5457	1723	3531	2611
PITIGUARISP7	0.1247	0.0464	0.0954	0.0034	0.1032	4.4183	1550	2929	2275
Média	0.1294	0.0445	0.1108	0.0036	0.1191	5.0247	1654	3428	2611
Desvio	0.0226	0.0089	0.0078	0.0003	0.0082	0.9336	89	225	284
Cof. Variação	17.48	20.03	7.05	8.61	6.89	18.58	5.38	6.57	10.89

Tabela 24 - Características extraídas dos sons do Saracura utilizados na fase de treinamento.

Ave	Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
SARACURAAL	0.0278	0.0167	0.0563	0.0033	0.0686	2.4704	517	2498	1132
SARACURAAP	0.0302	0.0139	0.0605	0.0036	0.0738	2.1162	689	2153	1374
SARACURAGO1	0.0201	0.0073	0.0559	0.0035	0.0614	3.0367	689	1981	1124
SARACURAMG1	0.0324	0.0105	0.0630	0.0034	0.0754	2.2943	861	1723	1312
SARACURAMG	0.0342	0.0474	0.0570	0.0035	0.0646	3.1102	431	2239	1474
SARACURAPER	0.0203	0.0119	0.0595	0.0034	0.0670	3.4217	517	2412	1639
SARACURARG	0.0104	0.0098	0.0519	0.0029	0.0620	3.7161	431	2239	1590
SARACURARJ	0.0202	0.0099	0.0615	0.0032	0.0732	3.5425	775	2239	1153
SARACURARS	0.0200	0.0110	0.0643	0.0033	0.0617	3.5238	689	2498	1186
SARACURAVEN	0.0239	0.0184	0.0606	0.0036	0.0682	2.2344	689	2326	1242
Média	0.0239	0.0157	0.0591	0.0034	0.0676	2.9466	629	2231	1323
Desvio	0.0072	0.0116	0.0038	0.0002	0.0052	0.6137	147	238	191
Cof. Variação	30.18	74.29	6.35	6.16	7.74	20.83	23.33	10.68	14.41

Tabela 25 - Características extraídas dos sons do Teque-teque utilizados na fase de treinamento.

Ave	Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
TEQUE-TEQUEMG	0.1310	0.0318	0.2557	0.0046	0.2521	8.3999	3876	6891	6341
TEQUE-TEQUEMG1	0.1751	0.0344	0.2317	0.0054	0.2460	6.7425	4134	6546	6053
TEQUE-TEQUERJ	0.1726	0.0211	0.2163	0.0053	0.2348	7.8360	4134	6977	5878
TEQUE-TEQUERJ1	0.0557	0.0210	0.2490	0.0047	0.2488	9.0212	3790	7407	6754
TEQUE-TEQUERJ2	0.0393	0.0107	0.2719	0.0057	0.2827	8.0673	3015	7838	6328
TEQUE-TEQUERJ3	0.1355	0.0329	0.2571	0.0046	0.2534	8.5029	3876	6891	6341
TEQUE-TEQUERJ4	0.0399	0.0109	0.2717	0.0057	0.2824	8.8778	3101	7838	6328
TEQUE-TEQUESP	0.1894	0.0372	0.2348	0.0053	0.2503	6.4326	4134	6546	6053
TEQUE-TEQUESP1	0.1981	0.0242	0.2245	0.0054	0.2416	7.5110	4134	6891	5878
TEQUE-TEQUESP4	0.0606	0.0172	0.2675	0.0048	0.2634	12.1277	3618	7321	6576
Média	0.1197	0.0241	0.2480	0.0052	0.2555	8.3519	3781	7115	6253
Desvio	0.0647	0.0097	0.0201	0.0005	0.0161	1.5753	421	470	287
Cof. Variação	54.02	40.01	8.12	8.79	6.29	18.86	11.13	6.61	4.60

Tabela 26 - Características extraídas dos sons do Tiziu utilizados na fase de treinamento.

Ave	Entropia	Energia	ZCR	Rolloff	Centroide	FE	Fmin	Fmax	FI
TIZIUAM1	0.2663	0.0670	0.3390	0.0051	0.3085	15.1945	3273	9733	7322
TIZIUAM2	0.3972	0.0522	0.2970	0.0048	0.2972	11.5431	3790	8958	6789
TIZIUAM3	0.2867	0.0371	0.3259	0.0054	0.3358	9.0806	7752	8958	8368
TIZIUAM	0.4051	0.0540	0.3122	0.0047	0.3170	10.9712	6115	9130	8273
TIZIUBA	0.3536	0.0291	0.3146	0.0048	0.3192	7.0277	3962	7580	8069
TIZIUGO2	0.3876	0.0396	0.3254	0.0049	0.3247	10.8758	4134	9475	8667
TIZIUGO3	0.3717	0.0502	0.3132	0.0051	0.3168	11.3023	3359	6374	7489
TIZIUMT2	0.3115	0.0490	0.3259	0.0050	0.3268	9.8270	4565	11456	7516
TIZIURJ1	0.3097	0.0297	0.3568	0.0053	0.3522	10.8976	3618	5599	7856
TIZIUSP1	0.3635	0.0638	0.2942	0.0047	0.3034	9.4511	7235	9647	8546
Média	0.3453	0.0472	0.3204	0.0050	0.3202	10.6171	4780	8691	7889
Desvio	0.0486	0.0131	0.0187	0.0002	0.0160	2.1023	1646	1724	603
Cof. Variação	14.07	27.74	5.83	4.86	4.99	19.80	34.44	19.84	7.65