PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

UMA ABORDAGEM PARA CLASSIFICAÇÃO DE ANUROS BASEADA EM VOCALIZAÇÕES

JUAN GABRIEL COLONNA

UMA ABORDAGEM PARA CLASSIFICAÇÃO DE ANUROS BASEADA EM VOCALIZAÇÕES

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal do Amazonas, Campus Universitário Senador Arthur Virgílio Filho, como requisito parcial para a obtenção do grau de Mestre em Informática.

Orientador: Eduardo Freire Nakamura
Co-orientador: Eulanda Miranda dos Santos

Manaus - AM
Março de 2011

Juan Gabriel Colonna
Uma Abordagem Para Classificação de Anuros Baseada em Vocalizações / Juan Gabriel Colonna. — Manaus - AM, 2011
xxi, 96 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal do Amazonas
Orientador: Eduardo Freire Nakamura
Co-
orientador:
Eulanda
Miranda
dos
Santos

1. Redes de Sensores Sem Fio. 2. Monitoramento Ambiental. 3. Classificação automática de anuros. 4. Aprendizagem de máquina. I. TÃŋtulo.

[Folha de Aprovação]

Quando a secretaria do Curso fornecer esta folha,
ela deve ser digitalizada e armazenada no disco em formato gráfico.

Se você estiver usando o `pdflatex`,
armazene o arquivo preferencialmente em formato PNG
(o formato JPEG é pior neste caso).

Se você estiver usando o `latex` (não o `pdflatex`),
terá que converter o arquivo gráfico para o formato EPS.

Em seguida, acrescente a opção `approval={`*nome do arquivo*`}`
ao comando `\ppgccufmg`.

Se a imagem da folha de aprovação precisar ser ajustada, use:
`approval=[`*ajuste*`][`*escala*`]{`*nome do arquivo*`}`
onde *ajuste* Ãľ uma distância para deslocar a imagem para baixo
e *escala* é um fator de escala para a imagem. Por exemplo:
`approval=[-2cm][0.9]{`*nome do arquivo*`}`
desloca a imagem 2cm para cima e a escala em 90%.

Agradecimentos

Simplesmente é muito difícil fazer uma lista com todas as pessoas que me ajudaram nestes dois anos. Estou muito grato pela oportunidade que tive de conhecer esta terra maravilhosa.

Agradeço principalmente a toda minha família, que sempre acreditou e torceu por mim. A Susi, minha mãe, que sempre esteve ao meu lado me apoiando em tudo, sendo meu exemplo na vida. Aos meus irmãos Eli e Nacho pelo carinho. Aos meus avós Yoli, Neli e Raúl pelo afeto. A meu pai Daniel pelo apoio. A minha tia e meus primos pela confiança em mim. Ao Marcelo pela a amizade e valores e à tia Peti.

Agradeço muito ao professor Eduardo por ter me recebido e ensinado a ser um pesquisador me dando o privilégio de aprender ao seu lado. À professora Eulanda pelo aprendizado, pela paciência e por acreditar neste trabalho. Ao professor Marco e à professora Rosiane pela troca de excelentes ideias que deram novos nortes ao trabalho. E em geral, a todos os professores do PPGI que realizam um ótimo trabalho mantendo uma excelente qualidade no ensino da UFAM. Ao professor Alejandro pela confiança que permitiu meu desenvolvimento. E ao professor Marcelo Gordo pela ajuda na parte biológica. Agradeço também à CAPES pelo suporte através de bolsa de mestrado.

A minha querida Jézika por todo o amor ao longo desse caminho e sua família pelo carinho e acolhida. A Luiz Leandro pela amizade e os ensinos de português. A meu colega Éfren pela amizade e companheirismo. A Clayton e Poly pela amizade e os longos papos compartilhados. A Antônio e João pela força. E a todos os colegas do laboratório e da turma pela amizade e por fazerem da UFAM um excelente lugar de trabalho tornando a jornada muito melhor!

Agradeço muito a todos os amigos da minha terra pela amizade que transcende o tempo e os limites geográficos. Agradeço também a todos cuja confiança, apoio, companheirismo e carinho, deram-me força para enfrentar as dificuldades da elaboração deste trabalho.

"La inspiración existe, solo tiene que encontrarte trabajando."
(Pablo Picasso)

Resumo

O monitoramento de animais selvagens é frequentemente usado por biólogos para coletar informações a cerca dos animais e seus habitats. Neste contexto, os sons produzidos pelos animais oferecem uma impressão biométrica que pode ser usada para classificar os animais em uma dada região. Assim, Redes de Sensores Sem Fio (RSSFs) representam uma importante alternativa para a classificação automática de animais, baseando-se nos sons por eles produzidos. Neste trabalho, apresentamos uma solução, que usa aprendizagem de máquina e processamento de sinais, para classificar animais selvagens com base em suas vocalizações. Como prova-de-conceito, aplicamos a solução para classificar anuros. O motivo é que anuros já são utilizados por biólogos como indicador precoce de estresse ecológico, oferecendo informação a cerca de ecossistemas terrestre e aquático. Além disso, a solução deve considerar as limitações de RSSFs, buscando reduzir a carga de comunicação para prolongar o tempo de vida da rede. Portanto, representamos os sinais acústicos por conjuntos de características. Esta representação nos permite identificar padrões específicos que descrevem cada uma das espécies monitoradas, reduzindo, assim, o volume de informação a ser trafegado na rede. A identificação destas características, ou combinações delas, é um ponto chave para aprimorar a relação custo-benefício da solução. Em nossa análise, primeiramente comparamos os conjuntos de características temporais e espectrais, usando a entropia como critério para geração das combinações. A seguir, reduzimos o conjunto de características, usando algoritmos genéticos. O framework proposto contém três passos: (i) pré-processamento, para preparar os sinais e extrair unidades chamadas sílabas; (ii) extração de características; e (iii) classificação, usando k-NN ou SVM. Os experimentos consistem de quatro estudos de casos, avaliando o efeito da quantização e o número de bits usados para representar o sinal quantizado. Isto nos permite concluir que, para os cenários avaliados, o conjunto de coeficientes Mel é o mais adequado para classificar vocalizações de anuros.

Palavras-chave: Redes de sensores sem fio, aprendizagem de máquina, monitoramento ambiental, classificação de anuros..

Abstract

Wildlife monitoring is often used by biologist to acquire information about animals and their habitat. In this context, animal sounds and vocalizations usually provide a specie fingerprint that is used for classifying the target species in a given site. For that matter, Wireless Sensor Networks (WSNs) represent an interesting option for automatically classifying animal species based on their vocalizations. In this work, we provide a solution that applies machine learning and signal processing techniques for classifying wildlife based on their vocalization. As a proof-of-concept, we choose anurans as the target animals. The reason is that anurans are already used by biologists as an early indicator of ecological stress, since they provide relevant information about terrestrian and aquatic ecosystems. Any solution must consider WSN limitations, trying to reduce the communication load to extend the network lifetime. Therefore, our solution represents the acoustic signals by a set of features. This representation allows us to identifiy specific signal patterns for each specie, reducing the amount of information necessary to classify it. Identifying such features, and/or combinations among them, is a key point to improve the solution benefit-cost ratio. As a consequence, we implemented and compared sets of existing features based on Fourier and Wavelet transforms. In our analysis, we first compare the sets of spectral and temporal characteristic, by using the entropy as a criterion for generating the combinations. Second, we reduce the set of features by using genetic algorithm. The proposed framework contains three steps: (i) the pre-processing to prepare the signals and perform the extraction of syllables, (ii) the extraction of features, and (iii) the species classification, using k-NN or SVM. Our experiments comprise four case studies, evaluating the effect of sampling frequency of the hardware and the number of bits used to represent each sample. This enable us to conclude that, in enviromental monitoring using WSNs, the set of Mel coefficients is the most appropriate for classifying anuran calls.

Keywords: Wireless Sensor Networks, Machine Learning, Enviromental Monitoring, Anuran Classification.

Lista de Figuras

Lista de Tabelas

Sumário

1

Introdução

Estudar condições ambientais relacionadas a mudanças climáticas, tais como efeito estufa e destruição da camada de ozônio, nos leva a uma discussão sobre a conservação do meio ambiente. O estudo de tais condições é de interesse social para manter a qualidade de vida e conservar as espécies [Guimarães, 2004].

Desde 1980 pesquisadores alertam a respeito de declínios dramáticos nas populações de anfíbios em todo o mundo. Vários fatores podem ser relacionados como causadores desse fenômeno, tais como: destruição ou modificação do habitat, exploração de recursos naturais, uso de pesticidas, poluição das águas e aumento da radiação. Tais causas, ainda em estudo, são difíceis de se compreender e colocam em risco a biodiversidade global.

Os anfíbios, particularmente os anuros, são usados por biólogos como indicadores para detectar estresse ecológico precoce [Collins & Storfer, 2003]. Pelo fato dos anuros estarem intimamente relacionados com o ecossistema, tornam-se vulneráveis e sensíveis às mudanças ambientais. Abordar o problema de monitoramento das populações a longo prazo permitiria mensurar a variação nas populações e correlacionar estas com informação relevante sobre o micro habitat [Carey et al., 2001].

O objetivo deste trabalho surge da necessidade de elaborar um método de classificação automática, que permita viabilizar uma proposta de monitoramento a longo prazo.

No restante deste capítulo encontra-se: o contexto de aplicação (seção 1.1), a definição do problema (seção 1.2), a motivação ambiental (seção 1.3), a justificativa e a abordagem escolhida (seções 1.4 e 1.5), a descrição dos cenários possíveis (seção 1.6), os objetivos propostos (seção 1.7), nossas contribuições (seção 1.8) e a organização da dissertação (seção 1.9).

1.1 Contexto

A tarefa de monitoramento de anuros implica no desafio de desenvolver um método que precise de menos intervenção humana comparada com as técnicas tradicionais. Em tais técnicas, o processo é realizado de forma manual, utilizando armadilhas para a coleta de amostras das espécies. A coleta manual pode requerer aproximadamente um ano, dependendo da quantidade de amostras necessárias, e da experiência da pessoa que realiza esta tarefa [Cechin & Martins, 2000].

Para melhorar a coleta e identificar os anuros existentes em uma determinada região podemos combinar técnicas de aprendizagem de máquina e Redes de Sensores Sem Fio (RSSF). Isto permitiria automatizar a amostragem das espécies, minimizando o impacto da intrusão causado no micro habitat, sendo também possível diminuir os tempos e os custos, garantindo resultados mais precisos [Taylor et al., 1996; Akyildiz et al., 2002; Marsland, 2009; Jain et al., 2000].

Entretanto, a capacidade dos anuros para emitir som pode ser explorada como característica para sua identificação e classificação [Huang et al., 2009; Vaca-Castaño & Rodriguez, 2010; Yen & Fu, 2002; Han et al., 2011; Colonna et al., 2011]. Através dessa identificação, podem ser inferidas informações sobre as populações que habitam uma determinada região, permitindo estabelecer uma relação com o meio ambiente.

Desta forma, podemos definir uma abordagem para o problema de classificação, na qual sejam extraídas características que representem de maneira única os sinais bioacústicos e identifiquem padrões dentro das formas de onda. Resultados prévios obtidos por diversos métodos, aplicados a diferentes animais, mostram que é possível utilizar técnicas de aprendizagem de máquina para tal fim [Huang et al., 2009; Clemins, 2005; Yen & Fu, 2002].

A lista de características encontradas na literatura, e implementadas por nós, inclui: o centroide (*Spectral Centroid* - S), largura de banda (*Signal Bandwidth* - B), taxa de cruzamento por zero (*Zero-crossing Rate* - ZC), coeficientes mel (*Mel Fourier Cepstral Coefficient* - *MFCCs*), energia (E), potência (Pw), o *Pitch* - P, a entropia de Shannon e a entropia de Rényi. Os classificadores utilizados para comparar os resultados foram: *k-Nearest Neighbor* (kNN) e *Support Vector Machine* (SVM) [Cover & Hart, 1967; Vapnik et al., 1992].

No contexto de monitoramento ambiental as RSSF são conhecidas pelo baixo custo dos nós, fato que permite espalhar grandes quantidades deles nas áreas desejadas. Entretanto, o baixo custo impõe restrições no hardware tais como: baixa capacidade de processamento e baterias com tempo de vida reduzido, sendo fundamental que a

abordagem de classificação esteja sujeita a essas restrições [Xing et al., 2005].

1.2 O problema de classificação

Embora os métodos de classificação existentes possuam resultados entre 60% e 100% de classificação correta, a maioria foram desenvolvidos de forma isolada pelo fato de que cada um é projetado para uma determinada espécie. Assim, cada experimento usa diferentes tipos de características e diferentes métodos de classificação, tornando o estudo comparativo extremamente difícil, carecendo da definição de um método padrão. Além disso, não foi realizada um análise comparando o impacto do custo destes em contextos de Redes de Sensores [Hu et al., 2005; Cai et al., 2007; Vaca-Castaño & Rodriguez, 2010; Han et al., 2011].

A carência de um método padrão nos motivou a (a) implementar os diversos métodos existentes na literatura, (b) compará-los e (c) definir uma abordagem para o problema de reconhecimento de anuros. Neste trabalho, abordamos o problema de classificação como uma forma de redução de informação, mediante a extração de características dos sinais bioacústicos. Priorizamos, em nossa abordagem, definir uma metodologia aplicável a uma Rede de Sensores Sem Fio e utilizamos a relação custo-beneficio como parâmetro de comparação entre os métodos implementados.

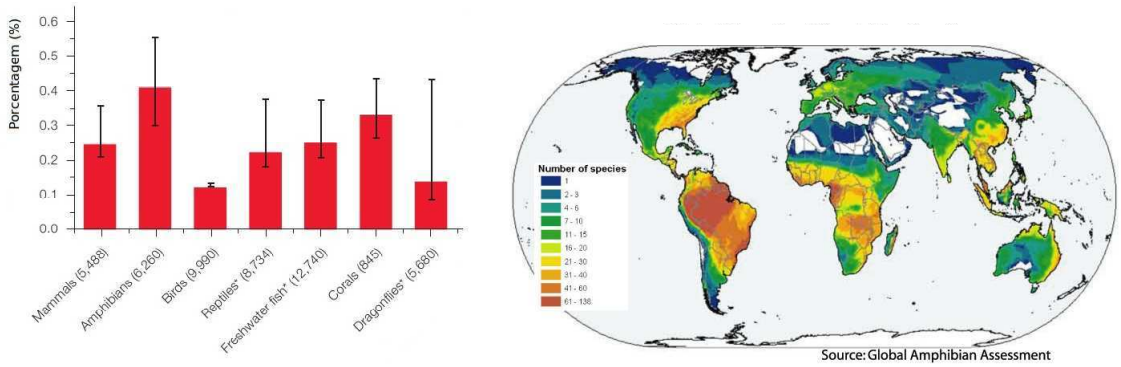
1.3 Motivação ambiental

Os desequilíbrios ecológicos, tais como mudanças no clima, desmatamento, aquecimento global, contaminação, impacto da urbanização e degradação da qualidade do habitat, são as principais causas de extinção das espécies animais e vegetais. A perda das espécies é um processo irreversível [Williams et al., 2003; Bernarde & Macedo, 2008; Mittermeier et al., 1998].

A União Internacional para a Conservação da Natureza elabora anualmente uma lista de espécies ameaçadas, conhecida como lista vermelha [IUCN, 2011]. Trata-se de um inventário sobre o estado das espécies em perigo. De acordo com Vié et al. [2009], no ano de 2008, foram listadas 44.838 espécies de animais, das quais 16.928 estão em perigo de extinção e 869 já foram extintas. Nela estão 6.260 espécies de anfíbios, sendo que 40% deles estão em perigo de extinção.

A Figura 1.1(a) mostra a quantidade de espécies listadas em cada categoria da lista vermelha em relação ao percentual da população que encontra-se em declínio. Aproximadamente 168 espécies de anfíbios já foram extintas, indicando que o número

de espécies em perigo continuará aumentando [Stuart et al., 2004]. Na figura 1.1(b) encontram-se identificadas a zonas de maior diversidade de anfíbios, sendo a região amazônica uma das mais propicias para viabilizar nosso estudo.



(a) Quantidade de espécies listadas pela UICN [Stuart et al., 2004].

(b) Distribuição mundial da biodiversidade de anfíbios.

Figura 1.1. Coleta de áudios e classificação usando uma RSSF.

A diminuição das populações de anfíbios pode ser explicada por seis diferentes hipóteses [Collins & Storfer, 2003]. As três primeiras hipóteses estão relacionadas à exploração abusiva e à mudança do uso da terra. Uma quarta hipótese é a mudança global, que inclui um aumento da radiação ultra-violeta e o aquecimento global. A quinta hipótese é a utilização crescente de pesticidas e outros químicos tóxicos. A sexta, são as doenças infecciosas emergentes.

Entender a interação dos fatores que causam isto é uma tarefa complexa e o simples entendimento das variações nas populações de anuros pode gerar um modelo que represente o decremento nas demais espécies animais. De acordo com Williams [2001] a relação entre a variação do habitat e a população das espécies permite:

- avaliar problemas ecológicos ainda em estágio inicial, e
- estabelecer estratégias de conservação da diversidade biológica.

Deste modo, desenvolver técnicas para monitorar e classificar anuros automaticamente permitiria viabilizar importantes estudos ambientais nas áreas biológicas e ecológicas.

1.4 Justificativa

A coleta dos sons da floresta e a análise destes é realizada normalmente de forma manual, mediante o uso de um analisador de espectro [Riede, 1993]. O processo de

reconhecimento de anuros mediante o som é realizado de forma similar, começando pela obtenção do som e fazendo-se uma primeira estimativa da espécie que se está ouvindo. Esta tarefa requer experiência, pois a visualização dos indivíduos nem sempre é possível. Os áudios são analisados com um software que realiza um espectrograma (ou sonograma), que é um gráfico comparativo de tempo e frequência do sinal.

A figura 1.2 ilustra um espectrograma da vocalização da espécie *Brachycephalus ephippium*. A intensidade de cores indica a energia das frequências de cada vocalização, com a o cor mais intensa correspondendo à maior energia. O eixo do tempo indica a duração da vocalização. Da figura 1.2 pode ser observada a regularidade das vocalizações no tempo.

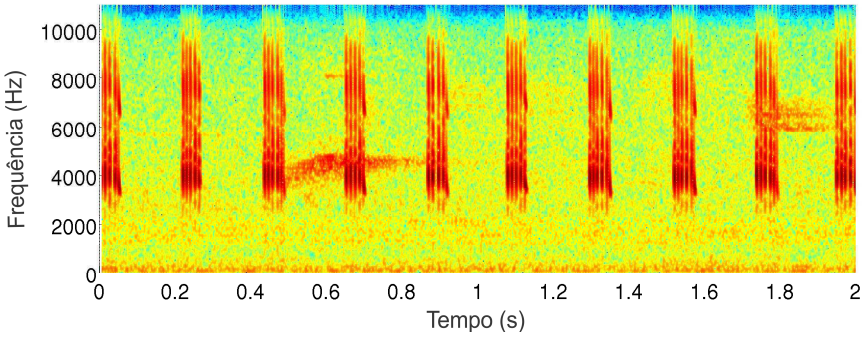


Figura 1.2. Espectrograma do *Brachycephalus ephippium*.

Em seguida, por comparação visual entre os espectrogramas adquiridos e um espectrograma padrão da mesma espécie, verifica-se o resultado da classificação. O processo completo depende quase totalmente da experiência da pessoa que executa a tarefa. A primeira parte depende da habilidade em reconhecer anuros a partir de suas vocalizações e a segunda da capacidade de comparação visual. Pela forma que esse processo é realizado, torna-se uma tarefa lenta e sujeita a erros.

Em nosso trabalho, pretendemos preencher as lacunas deixadas por trabalhos anteriores, propondo uma abordagem genérica para a identificação de espécies de anuros a partir das suas vocalizações. Para atingir este objetivo, baseamos nosso método em técnicas já existentes usadas em reconhecimento de fala humana [Campbell, 1997].

A complexidade do padrão da fala humana torna mais difícil a tarefa de reconhecimento de som, no qual pode-se usar um modelo como o apresentado na figura 1.3. Cada bloco que integra o modelo cumpre uma função específica, possuindo entrada, saída e parâmetros. A entrada e saída para cada bloco são as diferentes instâncias do sinal de áudio, antes e após o processamento. Em particular, cada bloco contém parâmetros a serem configurados, tais como constantes ou variáveis.

Este modelo genérico é usado como base para o sistema de reconhecimento de anuros. Desta forma, um dos objetivos é determinar quais blocos são necessários e quais parâmetros se ajustam melhor a nosso problema. Para adaptar o modelo a nosso problema, o classificador deverá escolher entre as diferentes classes, as quais se correspondem com as espécies de anuros.

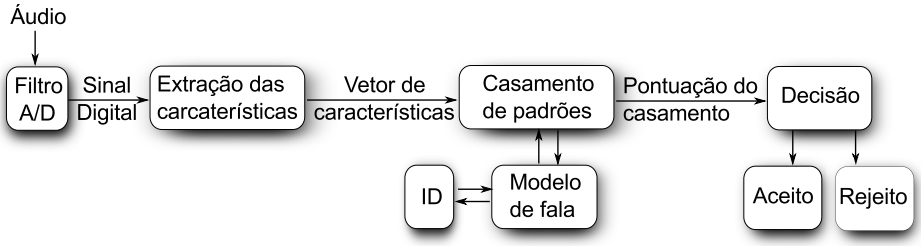


Figura 1.3. Sistema geral de reconhecimento de fala humana, figura adaptada de Campbell [1997].

Uma dificuldade adicional ao problema de reconhecimento é proveniente do cenário de aplicação. Este, trata-se de um ambiente adverso, contendo muitos tipos de ruídos, que dificultam ainda mais a tarefa de classificação. Os ruídos são provenientes de diferentes fontes, tais como a presença de outros anuros, outros animais, som da chuva, o vento nas folhas, entre outros.

1.5 Abordagem

Os métodos de classificação existentes baseiam-se na obtenção de características para representar os sinais bioacústicos. Em nosso trabalho escolhemos e implementamos três conjuntos de características principais: o primeiro baseado na transformada de Fourier, o segundo usando características temporais e o terceiro baseado na transformada Wavelet. Posteriormente combinamos estas características para formar novos conjuntos com o objetivo de identificar a combinação com a melhor relação custo-beneficio.

Para encontrar as características com melhor taxa de acerto, aplicamos uma metodologia de eliminação por etapas. Resumidamente, as etapas são:

1. Combinamos as características baseadas na transformada de Fourier (MFCCs, S e B) com as temporais (ZC, P, E e Pw), utilizando o critério de ganho de informação (*Information Gain - IG*), baseado na entropia, e escolhemos o subconjunto com melhor taxa de acerto;

2. Comparamos o subconjunto resultante da etapa (1) com as características extraídas da transformada Wavelet, para novamente escolher o melhor subconjunto; e

por fim

3. Utilizamos uma estratégia de otimização, baseada em algoritmo genético (GA), para encontrar as características ótimas dentro dos subconjuntos da etapa (2).

As três etapas descritas são realizadas fora de RSSF, como uma abordagem "*offline*". A partir da última (3) são obtidas características discriminantes o suficiente para que o classificador consiga diferenciar as espécies com a melhor taxa de acerto possível. Nesta etapa, também são descartadas aquelas caraterísticas que confundem a decisão. A abordagem de otimização é favorável para diminuir os custos de processamento nos nós sensores que realizam a captura dos áudios.

Nossa proposta de redução de informação implica em: "cada nó da rede deve processar e transmitir somente as características que foram obtidas na etapa de otimização (3)". Esta abordagem contribui com a redução dos custos de processamento e de transmissão, otimizando a vida útil das baterias e, por conseguinte, da rede.

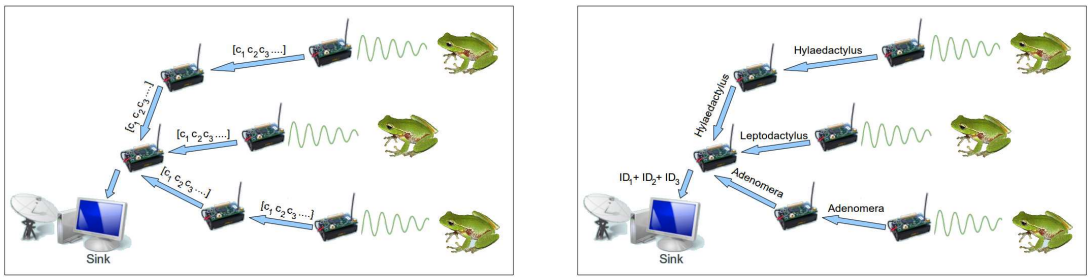
A redução dos custos, de processamento e de obtenção, está sujeita à quantidade de amostras obtidas em cada áudio. Assim, podemos reduzir ainda mais os custos diminuindo a frequência de amostragem no hardware (f_s [kHz]), representando cada sinal bioacústico com o menor número de amostras possíveis [Bulusu & Hu, 2008].

Nós correlacionamos os subconjuntos de características com f_s para caracterizarmos o impacto na taxa de classificação. Desta maneira, conseguimos correlacionar a taxa de acerto dos diferentes conjuntos de características com os custos de processamento e com a redução de f_s, o que permite simular cenários com diferentes tipos de hardware.

1.6 Cenários possíveis

No contexto de monitoramento ambiental de anuros com *RSSF* existem três cenários de transmissão possíveis, os quais são: (i) as amostras completas dos áudios viajam até o nó *sink*, (ii) só as características que o classificador precisa são transmitidas até o nó *sink* (figura 1.4(a)) ou (iii) a classificação é realizada no nó sensor (figura 1.4(b)).

Nestes contextos, é desejável diminuir a quantidade de informação transmitida, pois a operação de comunicação é a mais custosa em termos de consumo de energia, diminuindo consideravelmente a vida útil da rede. As operações de processamento são menos custosas. No entanto, uma carga excessiva destas é igualmente prejudicial. Dos três cenários, o melhor em relação ao custo de transmissão e processamento é o Cenário (ii), porque transmite menos pacotes que o Cenário (i) e processa menos informação que o Cenário (iii).



(a) Cenário (ii). (b) Cenário (iii).

Figura 1.4. Coleta de áudios e classificação usando uma RSSF.

O cenário escolhido possui uma desvantagem inerente: não é possível recuperar os áudios no receptor impossibilitando encontrar erros na execução da classificação. A tabela 1.1 resume as vantagens e desvantagem de cada cenário.

Cenários	**Vantagens**	**Desvantagem**
Cenário (i)	· Permite recuperar o áudio completo no nó *Sink*. · Baixo custo de processamento.	· Elevada quantidade de informação transmitida. · Requer elevada memória de armazenamento. · Elevado custo de energia.
Cenário (ii)	· Redução de informação: $R = 100 - \frac{12*100}{f_s*0,2} = 99,86\%$. · Diminui o custo de transmissão e energia. · Poupa memória no sensor.	· Não é possível recuperar o áudio. · Eleva o custo de processamento no nó sensor. · Confunde ruídos de outras espécies.
Cenário (iii)	· Redução de informação: $R = 100 - \frac{1*100}{f_s*0,2} = 99,98\%$. · Diminui o custo de transmissão e energia. · Poupa memória no nó sensor.	· Não é possível recuperar o áudio. · Eleva ainda mais o custo de processamento no nó sensor. · Confunde ruídos de outras espécies.

Tabela 1.1. Vantagem e desvantagem dos três cenários possíveis de monitoramento mediante RSSF, considerando $f_s = 44,1$kHz.

Resumidamente, os desafios e dificuldades apresentados na tabela 1.1 nos motivaram a encontrar uma técnica de redução de informação que possua menor consumo de energia para transmissão, permitindo melhorar a vida útil da rede. Na seção 2.8.1 encontra-se o consumo detalhado de cada operação realizada por um nó sensor MICA Weather Board [Mainwaring et al., 2002]

1.7 Objetivo Geral

O objetivo desta dissertação é classificar espécies de anuros, da Floresta Amazônica na região próxima a Manaus, usando as vocalizações emitidas por eles. Neste trabalho definimos as técnicas que melhor se adaptam ao monitoramento específico de anuros, combinando as características que melhor representam os sinais de áudio junto com técnicas de aprendizagem de máquina, para obter o melhor desempenho na classificação.

1.7.1 Objetivos Específicos

Os objetivos específicos são descritos abaixo:

1. Caracterizar e identificar os espectros de frequências das vocalizações de diferentes espécies de anuros estudadas.
2. Extrair e selecionar as características que melhor representam as vocalizações dos anuros.
3. Definir a técnica de classificação com melhor desempenho para classificar anuros.
4. Obter o conjunto mínimo de características que maximizem a taxa de acerto.
5. Identificar a mínima f_s que não altere o resultado da classificação.
6. Correlacionar o custo de processamento obtido de cada conjunto com a taxa de acerto.
7. Identificar uma ou mais espécies vocalizando ao mesmo tempo.

1.8 Nossas contribuições

As contribuições de nosso trabalho são:

1. Definimos um método de classificação automática de anuros, combinando as melhores características dos métodos existentes, como um sistema em blocos possível de ser aplicado em uma RSSF;
2. Demonstramos analiticamente o custo computacional e a ordem de complexidade da extração das características implementadas;

3. Realizamos uma análise comparativa entre as características em termos de ordem de complexidade, taxa de acerto e custo de processamento;

4. Diminuímos os custos de processamento e transmissão mantendo a porcentagem de classificação através do uso de algoritmo genético para reduzir os conjuntos de características;

5. Avaliamos o impacto na taxa de classificação da quantização e a diminuição da frequência de amostragem, realizando duas simulações com os subconjuntos de características retornados por GA;

6. Identificamos o melhor grupo de características para o problema de reconhecimento de anuros que maximizam a relação custo-beneficio; e

7. Definimos uma estratégia de reconhecimento de perfil de grupo para uma, duas ou três espécies.

Além do problema de classificação individual de cada espécie abordamos o problema de reconhecimento simulando uma situação real, na qual estariam presentes uma, duas ou três especies de anuros vocalizando ao mesmo tempo na mesma região. Para isto, utilizamos uma estratégia de reconhecimento de padrões de grupo ou "perfil de grupo", realizando todas as combinações possíveis das nove espécies utilizadas em nossa base de dados.

A partir de nossos experimentos concluímos que, somente com os coeficientes mel (12 MFCCs) é possível classificar com 97,60% de certeza, de forma individual, as nove espécies de anuros utilizadas em nossas bases. O resultado do GA mostrou que é possível reduzir o conjunto de coeficientes até 8 e manter uma taxa de acerto igual a 97,95%, o que implica em uma redução de 33% da informação necessária a ser transmitida, quando comparada com os 12 MFCCs. Da transformada Wavelet, concluímos que somente 4 das 9 características propostas são suficientes para manter a taxa de classificação e diminuir 55% da informação necessária para classificar.

Os resultados obtidos no desenvolvimento desta pesquisa foram publicadas no *III Simpósio Brasileiro de Computação Ubíqua e Pervasiva* (SBCUP) [Colonna et al., 2011] e no *International Joint Conference on Neural Networks* (IJCNN) [Colonna et al., 2012].

1.9 Organização da proposta

Esta dissertação está organizada da forma seguinte; no capítulo 2 são apresentados os fundamentos teóricos necessários para o entendimento dos métodos adotados; uma síntese dos trabalhos relacionados, expondo vantagens e desvantagens dos métodos existentes, é apresentada no capítulo 3; a abordagem proposta é descrita no capítulo 4; a avaliação dos resultados é mostrada nos capítulos 5 e 6; e por fim, no capítulo 7, são discutidas as conclusões e as possíveis extensões futuras deste trabalho.

2

Fundamentos

Neste capítulo são apresentados os conceitos prévios necessários para o entendimento e desenvolvimento do trabalho. O ferramental teórico é dividido em nove seções.

A seção 2.1 compreende um resumo sobre a geração das vocalizações e seus aspectos fundamentais, tais como os espectros de frequência e as formas da onda dos sinais. Nas seções 2.2 e 2.3 são definidas duas etapas correspondentes ao pré-processamento de nosso método, utilizadas para o acondicionamento das vocalizações. A seção 2.4 apresenta um resumo das transformadas de Fourier e Wavelet utilizadas para representar os sinais em espaços de dimensões diferentes. Após a transformação, são extraídas características que representem os sinas. Tais características e sua complexidade são estudadas na seção 2.5. O conjunto de características é utilizado para associar cada sinal com uma espécie de anuro. Isto é realizado mediante as técnicas de classificação descritas na seção 2.6. Os critérios para combinar características, assim como o teste utilizado para comparar os resultados, encontram-se na seção 2.7.

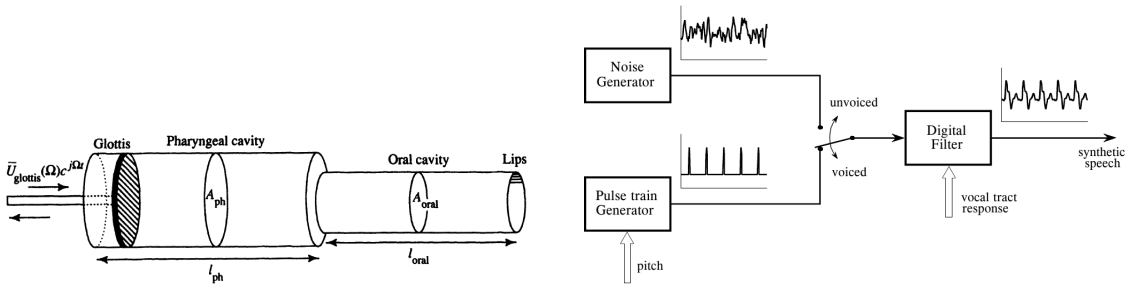
Na introdução foi contextualizado o problema de monitoramento com sensores de baixo custo. Por este motivo, na seção 2.8, são apresentadas as Redes de Sensores Sem Fio e as restrições de energia que possuem. Por fim, na seção 2.9, encontram-se as considerações finais deste capitulo.

2.1 Modelo de Vocalização

O modelo de vocalização mais amplamente estudado é o humano, que é usado neste trabalho como base para a discussão e comparação com os modelos de vocalizações de animais. Estudos recentes destacam as similaridades entre a forma de percepção dos sons emitidos por animais e por humanos [Bee & Micheyl, 2008]. Esses estudos

permitem supor que o método usado para reconhecer voz humana pode ser adaptado para reconhecer anuros.

O trato vocal humano, descrito na figura 2.1(a), é formado principalmente pela glote que puxa o ar formando pulsos de diferentes frequências dentro da faringe. Posterior à faringe, encontra-se a cavidade oral. As duas últimas são cavidades ressonantes, que funcionam como filtros para gerar os diferentes tons da fala [Deller et al., 1993]. A figura 2.1(b) mostra como pode ser modelado o filtro do trato vocal [Smith, 1999].



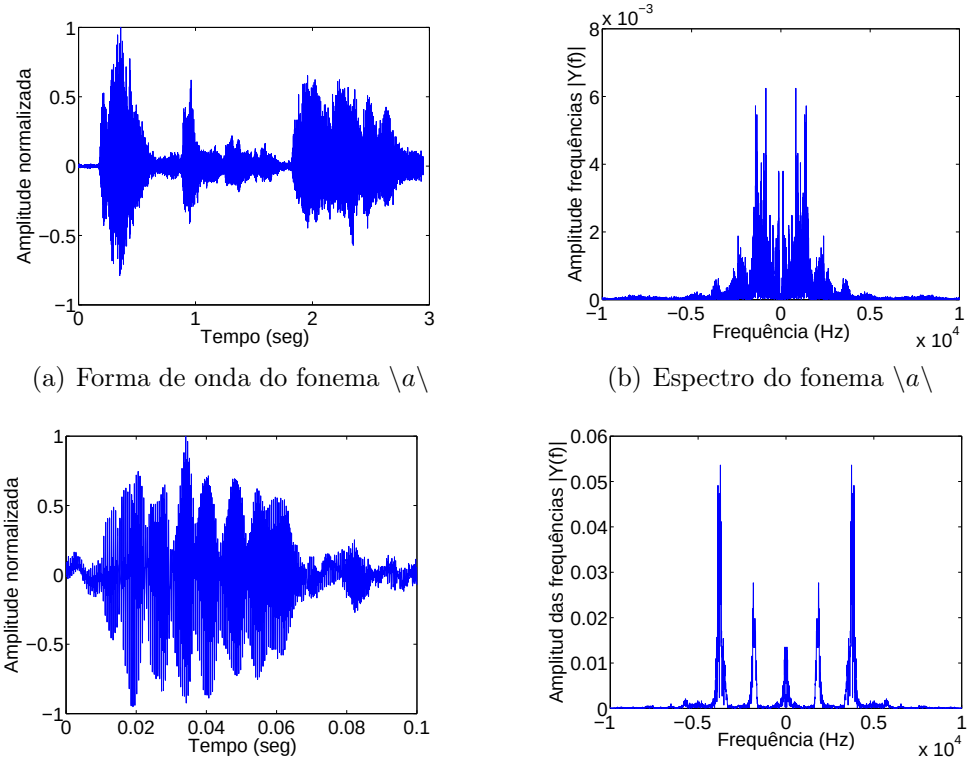
(a) (b)

Figura 2.1. (a) Modelo de fala humana, extraído de Deller et al. [1993] e (b) forma de produção da fala, extraído de Smith [1999].

O trato vocal modula um sinal portador de informação. Este sinal transporta a maior quantidade de informação. Geralmente os espectros dos sinais de fala são uniformes ao longo das frequências. Comparativamente, os sinais bioacústicos possuem uma característica de excitação na frequência mais concentrada. Isto produz espectros que variam dinamicamente, mas com uma maior quantidade de harmônicos [Clemins, 2005].

A maioria das espécies de anuros produz vocalizações intensas e agudas. Diferentemente de outros animais, eles conseguem fazer isso com as narinas e a boca fechada. O ar flui desde a laringe passando pela boca até o saco vocal. O saco vocal reforça cada vocalização funcionando como uma cavidade ressonante. Embora nem todas as espécies possuam saco vocal, a maioria tem uma cavidade ressonante que funciona como filtro para gerar diferentes tons e intensidades [Gerhardt, 1975; Martin & Gans, 1972; Purgue, 1997]. Pelas características citadas, pode-se observar que existe uma certa similaridade entre a geração da fala humana e a geração de uma vocalização de anuro. Na figura 2.2 é comparado o espectro de frequências do fonema vocal $\backslash a \backslash$ com o espectrograma de uma sílaba típica da espécie *Hylaedactylus*.

Os diferentes tons das vocalizações são captados e digitalizados pelos sensores de áudio dos nós. A captação e armazenamento são duas etapas que precisam transformar



(a) Forma de onda do fonema \a\

(b) Espectro do fonema \a\

(c) Forma de onda da sílaba do *Hylaedactylus*

(d) Espectro do *Hylaedactylus*

Figura 2.2. Comparação entre formas de onda e espectros, da voz humana e a vocalização do anuro *Hylaedactylus*.

o sinal analógico em digital. Esta transformação é explicada no próxima seção.

2.2 Quantização

Na prática, os hardwares de aquisição de sinal possuem um módulo conversor analógico-digital (*Analog-to-Digital Converter - ADC*). Este módulo discretiza o sinal analógico em níveis, representando cada valor como um número binário (figura 2.3).

Essa conversão produz um erro denominado ruído de quantização. Nos nós sensores MICA, utilizados por Mainwaring et al. [2002] a representação é realizada com palavras de 10 bits. Desta forma, a relação sinal-ruído de quantização pode ser calculada como:

$$\frac{S}{N_q} = 1.76 + 6.02n - 20\log\left(\frac{V_{max}}{V}\right), \tag{2.1}$$

em que n é o número de bits e V o valor do sinal em Volts, S e N_q as potências do sinal e do ruído de quantização respectivamente. Neste caso, considerando um sinal normali-

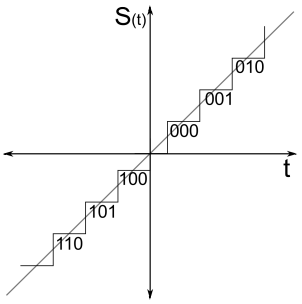


Figura 2.3. Quantização (ADC).

zado entre ±1, o ruído de quantização é aproximadamente 61,96 dB [Oppenheim et al., 1999].

O sinal quantizado e digitalizado é armazenado no memória dos nós durante um intervalo de tempo limitado. Esta janela de tempo é uma representação truncada do sinal original, possuindo uma representação em frequência aproximada. Para diminuir os efeitos não desejados desta aproximação, aplicamos uma janela de Hamming, explicada na próxima seção.

2.3 Janelamento

Estes sinais, da mesma forma que os sinais da fala humana, são processos estocásticos. Processos desta natureza possuem propriedades estatísticas que não mudam com o tempo. As operações de segmentação e extração das características requerem sinais estacionários.

Nós estamos interessados no conteúdo espectral do sinal bioacústico apenas em um período de tempo determinado. Como a duração deste período é um tempo finito, o resultado é um sinal aperiódico. Particularmente, o espectro de frequências de sinais aperiódicas contém infintos componentes harmônicos. O cálculo da transformada de Fourier (FFT), por exemplo, não produz uma saída com infinitos harmônicos, resultando em uma representação em frequência incompleta e com distorção nos extremos.

Nestes casos, a recuperação da informação, mediante a aplicação de transformada inversa de Fourier (IFFT), possui distorção conhecida como fenômeno de *Gibbs* [Oppenheim et al., 1999]. Para diminuir o impacto deste fenômeno, aplica-se uma janela do tipo *Hamming*. A aplicação desta janela, modifica as amplitudes do sinal, pela multiplicação de uma forma de onda como a representada na figura 2.4.

A ponderação do sinal pelo janelamento aprimora sua representação, conseguindo descrever os espectros de frequências com um grau de precisão maior. A obtenção dos

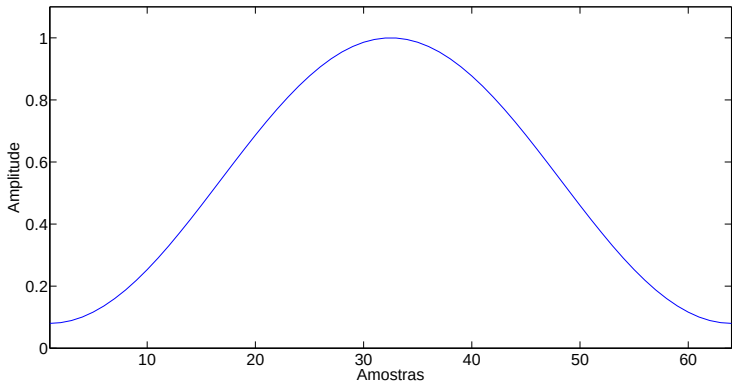


Figura 2.4. Janela de *Hamming*.

espectros dos sinais bioacústicos é realizada aplicando a Transformada de Fourier e Wavelet descritas na seção seguinte.

2.4 Transformadas

Para que a classificação de vocalizações seja realizada, primeiramente é necessário que características que representem os sinais bioacústicos sejam extraídas. Tais características podem ser obtidas no domínio do tempo ou da frequência. No domínio do tempo são calculadas diretamente, mas no domínio da frequência, é necessário realizar uma transformação prévia. Nesta seção são apresentadas duas transformadas: Fourier e Wavelet.

2.4.1 Transformada discreta de Fourier

Transformar cada vocalização no domínio da frequência permite extrair informação que represente de forma única e discriminante cada sinal bioacústico. A obtenção desta informação é realizada mediante a aplicação da transformada discreta de Fourier (DFT), dada pela equação seguinte:

$$X_k = \sum_{n=0}^{N-1} x_n e^{\frac{-i2\pi kn}{N}},\ k = 0, 1, \cdots, N-1. \tag{2.2}$$

Na equação 2.2, x_n é o vetor com as amostras de áudio das vocalizações, N é o número de pontos da transformada e X_k é o valor do componente em frequência no ponto $w = \frac{2\pi k}{N}$ do espectro. A relação entre a transformada contínua e discreta é dada por:

$$X_k = V(e^{iw})|_{w=\frac{2\pi k}{N}}, \tag{2.3}$$

em que X_k são amostras de $V(e^{iw})$ que formam uma progressão geométrica de inteiros com razão $\frac{2\pi}{N}$ [Oppenheim et al., 1999]. Da equação 2.2 observa-se que para obter N pontos da transformada de Fourier é preciso resolver o somatório com N operações, obtendo-se uma ordem de complexidade $O(N^2)$ [Duhamel & Vetterli, 1990].

Na década de 1960, Cooley & Tukey [1965] aplicaram a técnica de divisão e conquista para diminuir o custo computacional do cálculo da DFT e chamaram este cálculo de transformada rápida de Fourier (FFT). Na abordagem implementada, a quantidade de amostras desejadas N, do espectro de frequência, é dividida recursivamente em N_1 e N_2, subconjuntos de tamanhos similares. A divisão é realizada até que seja obtida a quantidade mínima de amostras indivisíveis. Por fim, são resolvidas N_2 transformações de tamanho N_1, para o primeiro conjunto, e N_1 transformações de tamanho N_2 para o segundo. Os resultados obtidos de cada subconjunto são combinados, obtendo-se uma complexidade de $\Theta(Nlog_2N)$.

A eficiência deste cálculo depende do valor escolhido para o parâmetro N, sendo que os melhores valores devem ser potências de 2. O software de cálculo numérico MATLAB utiliza para o cálculo da FFT a implementação $FFTW$ desenvolvida por Frigo & Johnson [1998], baseando-se na técnica de Cooley & Tukey [1965] com complexidade $\Theta(Nlog_2N)$. Em implementações práticas as limitações dos hardwares (nós sensores) impõem restrições na forma de cálculo da FFT, sendo necessário realizar o cálculo da FFT utilizando aritmética de ponto fixo, ocasionando um erro de arredondamento. A diferença entre a aproximação calculada e o valor matemático exato foi estudada por Welch [1969]. Uma alternativa de calculo é utilizar a transformada Wavelet descrita à continuação.

2.4.2 Transformada Wavelet

Além da transformação clássica na frequência, realizada com FFT, existe como alternativa a transformada Wavelet (WT). A diferença entre as transformadas são as funções de bases ortogonais, que podem ser diferentes a exponencias complexas, mais ainda, podem ser dilatadas e trasladadas no tempo [Graps, 1995]. A transformada Wavelet contínua (CWT) é definida como [Morettin, 1999]:

$$\gamma(s,\tau) = \int f(t)\Psi^*_{s,\tau}(t)dt, \tag{2.4}$$

em que $f(t)$ é o sinal, $\Psi^*_{s,\tau}(t)$ são as funções bases nas quais $f(t)$ é decomposto e * denota o conjugado complexo.

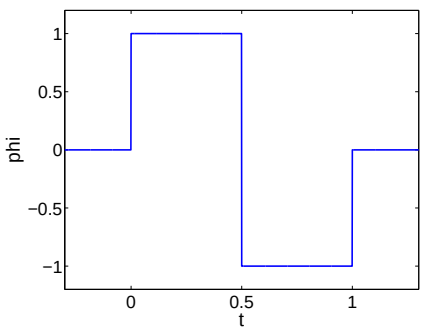
Os diferentes valores das variáveis s e τ, escalonamento e traslação respectivamente, fornecem as diferentes bases ortogonais gerando a família Wavelet. Formalmente, a função base (Ψ) ou protótipo, na qual é possível decompor o sinal (f), é conhecida como Wavelet "mãe" definida como:

$$\Psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\Psi\left(\frac{t-\tau}{s}\right), \tag{2.5}$$

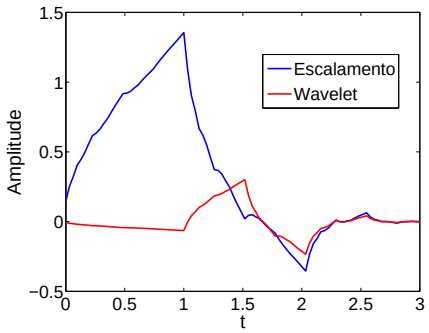
esta equação é interpretada como a aplicação de diversos filtros passa-banda sobre o sinal $f(t)$.

A transformada Wavelet nasceu em 1909 com a função mãe definida por Haar [1910] (figura 2.5(a)):

$$\Psi_{s,\tau}(t) = \begin{cases} +1 & 0 \leq t \leq \frac{1}{2} \\ -1 & \frac{1}{2} < t \leq 1 \\ 0 & \textit{outro caso} \end{cases} \tag{2.6}$$



(a) Função Wavelet Haar



(b) Função de escalamento e Wavelet Daubechies

Figura 2.5. Funções mãe da transformada Wavelet.

Posteriormente, Daubechies [1988] formalizou a família de funções *db* (figura 2.5(b)), constituindo as mais famosas e usadas atualmente. Além destas funções, existem diversas outras com propriedades diferentes, tais como biortogonal, coiflet, chapéu mexicano, Morlet e Meyer entre outras.

2.4.2.1 Transformada Wavelet discreta (*DWT*)

Em aplicações reais, os sinais são obtidos através de um hardware de amostragem, resultando em um conjunto de amostras de tempo discreto, tornando necessária uma

transformada de natureza discreta também. Consequentemente, as bases para representar os sinais devem ser vetores para possibilitar a decomposição do sinal ($f(t)$). Matematicamente, um sinal de tempo discreto $f(n)$ pode ser representado por dois conjuntos de coeficientes: os coeficientes de escala c_k (ou média) e os coeficientes de detalhes $d_{j,k}$ (ou diferença). Analiticamente isto é:

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \phi(t-k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} \psi(2^j t - k), \tag{2.7}$$

as funções $\phi(t)$ e $\psi(t)$ são conhecidas como funções "pai" e "mãe" respectivamente. É possível obter uma sequência de coeficientes isolando h e g das seguintes equações:

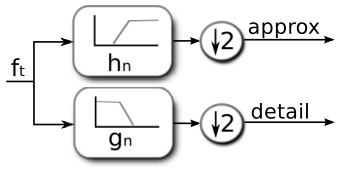
$$\psi(t) = \sqrt{2} \sum_{n \epsilon \mathbb{Z}} h_n \phi(2t - n) \tag{2.8}$$

$$\phi(t) = \sqrt{2} \sum_{n \epsilon \mathbb{Z}} g_n \phi(2t - n) \tag{2.9}$$

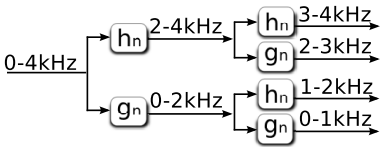
Tomando a transformada Z das duas equações anteriores, os valores h_n e g_n resultam coeficientes de filtros discretos. Desta forma, o sinal discreto pode ser convolucionado com dois filtros, um filtro passa-baixas (g_n), para obter os coeficientes Wavelet (ou de escala), e um filtro passa-altas (h_n), para obter os coeficientes de detalhes. Em outras palavras, para realizar a transformada DWT não é necessário escalonar ou trasladar a função Wavelet mãe, simplesmente pode-se filtrar sucessivamente o sinal de entrada. A seguir, são explicados os passos a serem realizados para a transformação.

2.4.2.2 Passos da DWT

Na prática, para que os valores da DWT sejam obtidos, são necessárias duas operações: a convolução do sinal com os filtros e um *downsample* (sub-amostragem) por dois. A figura 2.6(a) ilustra o procedimento inicial.



(a) Transformada discreta no primeiro nível.



(b) Árvore da transformada sucessiva ou *WPT*.

Figura 2.6. Transformada Wavelet.

Uma característica importante, que diferencia a DWT da FFT, é a possibilidade

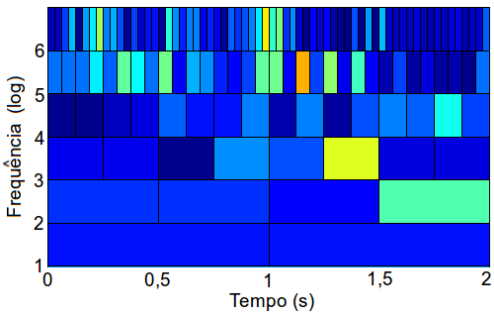
de aplicá-la recursivamente sobre o resultado obtido na transformação anterior. Em outras palavras, o resultado do filtro passa-baixas junto com o *downsampling* (sub-amostragem), pode ser convolucionado novamente com o filtro e outro *downsampling* realizado. Esta forma de recursão pode ser aplicada da mesma forma aos coeficientes de detalhes com o filtro passa-altas.

A figura 2.6(b) mostra como é realizado o procedimento descrito. A forma de árvore resultante, após a aplicação recursiva da transformada, é conhecida como *Wavelet Packet Transform - WPT*). Esta recursão pode ser aplicada até que um único coeficiente seja obtido no último nível.

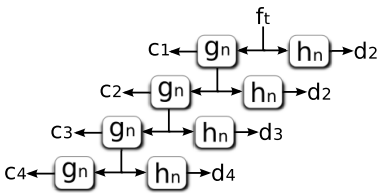
A possibilidade de decompor sinais em bases ortogonais que podem ser escaladas e trasladadas no tempo possibilita simultaneidade de representação no tempo e frequência. A representação em frequência/tempo, ou espectrograma, possui a limitação de resolução simultânea, em outras palavras, quando a resolução em frequência aumenta a resolução no tempo diminui e vice-versa [Rioul & Vetterli, 1991].

Mediante a representação pela aplicação da *DWT* do sinal, pode ser obtida uma representação similar ao espectrograma, mas com maior resolução tempo/frequência simultaneamente, chamada scalograma [Rioul & Vetterli, 1991]. O escalograma é a representação da energia dos coeficientes de cada nível (figura 2.7(a)).

A figura 2.7(a) exemplifica o aumento da resolução no tempo e na frequência após a aplicação da transformada. Em processamento de sinais de áudio, normalmente a transformada é aplicada sobre os coeficientes de escala para se obter maior resolução sobre as baixas frequências (figura 2.7(b)).



(a) Divisão no tempo e na frequência pela aplicação sucessiva da *DWT* ou *WPT*.



(b) Aplicação da *DWT* em processamento do sinais de áudio.

Figura 2.7. Escalograma (a) e árvore Wavelet unilateral (b).

Os passos descritos anteriormente podem ser resumidos nas seguintes operações matemáticas: transformada de *Haar* (equação 2.10) e transformada de *Daubechies* (equação 2.11) [Selesnick, 2007].

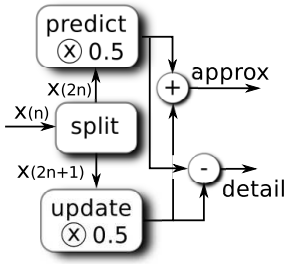
$$\begin{aligned} c(n) &= 0,5x(2n) + 0,5x(2n+1), \\ d(n) &= 0,5x(2n) - 0,5x(2n+1), \end{aligned} \tag{2.10}$$

$$\begin{aligned} c(n) &= h_0x(2n) + h_1x(2n+1) + h_2x(2n+2) + h_3x(2n+3), \\ d(n) &= h_3x(2n) - h_2x(2n+1) + h_1x(2n+2) - h_0x(2n+3), \end{aligned} \tag{2.11}$$

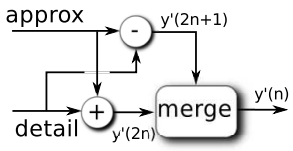
onde $x(n)$ são as amostras do sinal de entrada e h são os valores dos coeficientes dos filtros, que no caso *db* são valores:

$$h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}},\ h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}},\ h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}},\ h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}. \tag{2.12}$$

Além da possibilidade de aplicar as equações 2.10 e 2.11, as quais requerem armazenamento de no mínimo quatro amostras do sinal de entrada, existe uma forma de realizar o cálculo quase em tempo real denominado *Lifting Scheme*. Este esquema foi utilizado por Rein & Reisslein [2011], pelo baixo custo computacional, e está resumido na figura 2.8.



(a) Transformada Wavelet mediante esquema *lift*.



(b) Recuperação da informação mediante a transformada inversa.

Figura 2.8. Esquemas de baixo custo.

Os blocos *Predict* e *Update*, da figura 2.8(a), multiplicam as amostras do sinal de entrada pelos coeficientes dos filtros, enquanto o bloco *Split* divide as amostras em pares e ímpares e o bloco *Merge*, figura 2.8(b), recompõe o sinal.

2.4.2.3 Considerações finais da WT e o Escalograma

A transformada discreta Wavelet tem sido muito utilizada por possuir algumas vantagens comparada com a *DFT*. Como foi descrito na seção 2.4.1, o algoritmo da *FFT* possui uma ordem de complexidade computacional igual a $O(n\ \log_2 n)$, enquanto a *DWT* possui uma complexidade menor, $O(n)$.

Considerando as equações 2.10 ou o esquema *lift* (figura 2.8(a)), observa-se que para realizar a transformação são necessárias somente operações de adição e multiplicação, que podem ser consideradas de custo constante [Cormen et al., 2002]. Tais operações são aplicadas uma vez a cada amostra do sinal de entrada, consequentemente se o sinal possui n amostras, para que a transformada seja obtida são realizadas $O(n)$ operações.

A complexidade $O(n)$ é calculada levando-se em consideração que a DWT é realizada somente uma vez, ou seja, somente no primeiro nível. Já para a aplicação sucessiva da transformada ou WPT (figuras 2.6(b) e 2.7(a)) é necessário que a quantidade de níveis possíveis a serem transformados seja considerada. Na situação na qual a quantidade de amostras do sinal é potência de dois, a quantidade máxima de níveis de decomposição possíveis é $log_2 n$. Assim, para que os coeficientes de escala da WPT sejam obtidos, é necessária a aplicação da transformação $2^{\log_2 n} - 1 = n - 1$ vezes. A mesma quantidade de operações é necessária para a obtenção dos coeficientes de detalhes. Por fim, a totalidade de operações para obtenção da representação completa é $2(n-1)$ e a complexidade resultante é $O((2(n-1))n) = O(n^2)$.

A complexidade $O(n^2)$ da WPT é superior à $O(n \log n)$ da FFT, mas em nossa aplicação utilizamos a transformada DWT só no primeiro nível, com complexidade $O(n)$. Abordando-se o problema desta forma, torna-se evidente a vantagem de aplicar a transformada Wavelet.

A Wavelet possui ainda a vantagem adicional de decompor sinais não estacionários e pode ser aplicada em sinais de uma ou duas dimensões [Rein & Reisslein, 2011]. Devido a essa característica, a Wavelet tornou-se muito utilizada na compressão e reconhecimento de imagens [Kekre et al., 2010], filtragem de ruídos [Selesnick, 2007], extração de características em redes de sensores sem fio ($RSSF$) [Rein & Reisslein, 2011], aplicações em $FPGA$ (*Field Programmable Gate Array*) e identificação da fala [Deshpande & Holambe, 2010; Tan et al., 1996; Sarikaya et al., 1998; Wu & Lin, 2009].

Realizar a transformação de um sinal, seja esta a FFT ou Wavelet, é útil para obter uma representação em um espaço de dimensões diferentes. Estas representações diferentes permitem visualizar diferentes características contidas nos sinais. Além disso, a representação dos sinais em forma de características permite reduzir a quantidade de informação necessária para identifica-los. Na próxima seção são apresentados os cálculos para obter as características que representem diferentes informações após as transformações dos sinais.

2.5 Características

O desempenho do sistema de classificação depende diretamente das informações que descrevem os dados. Tais informações são as características extraídas do sinal. Características sensíveis ao ruído ou similares entre classes confundem o classificador, diminuindo o êxito da classificação. Idealmente, as características devem ser imparciais, não correlacionadas, e devem representar diferenças significativas ente as classes [Jain et al., 2000].

As características podem ser extraídas realizando-se ou não as transformações descritas nas seção 2.4. Nas próximas subseções apresentamos as características mais utilizadas pelos sistemas de reconhecimento de fala e pelos sistemas de reconhecimento de animais em geral, e implementadas por nós em nossos métodos.

2.5.1 Coeficientes Mel

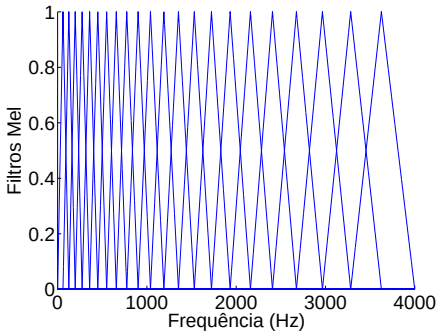
Atualmente os sistemas de reconhecimento de fala humana (*Automatic Speech Recognition - ASR*), baseiam seu funcionamento no uso de características tais como *Mel-frequency Cepstral Coefficients (MFCCs)* e os *PLP (Perceptual Linear Prediction*), para a correta representação dos sinais de fala [Rabiner & Schafer, 2007].

Os *MFCCs* foram originalmente desenvolvidos por Davis & Mermelstein [1980] sendo os recursos mais populares por causa de sua eficiência computacional, robustez aos ruídos e capacidade de capturar ressonâncias do trato vocal. Esta característica representa melhor os sinais da fala do que características que usam uma escala de frequência linear. A escala de frequência Mel é consistente com a forma de percepção humana da voz, servindo para sinais periódicos e aperiódicos. O uso destes coeficientes permite reduzir significativamente a quantidade de informação necessária para descrever o sinal, sem uma perda de informação relevante [Cai et al., 2007]. A ideia básica desta técnica é fazer uma análise espectral baseada num banco de filtros triangulares espaçados logaritmicamente na frequência, conforme figura 2.9(a).

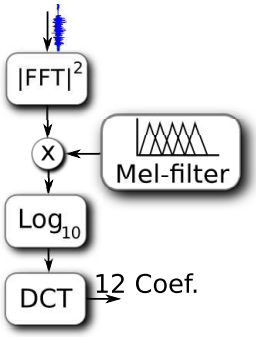
O espaçamento entre filtros é definido na escala Mel por [Cowling, 2003]:

$$f_{mel} = 1127 \ln \left(1 + \frac{f_{Hz}}{700}\right). \tag{2.13}$$

A aplicação do banco de filtros sobre o espectro do sinal produz um valor $M_{[r]}$ para cada filtro. Após a obtenção dos valores $M_{[r]}$, é feito o logaritmo e aplicada a transformada discreta do cosseno (DCT), para que os coeficientes Mel $mfcc_{[m]}$ sejam



(a) Banco de filtros Mel.



(b) Bloco de extração dos coeficientes Mel.

Figura 2.9. Extração das características utilizando os MFCCs.

obtidos [Rabiner & Schafer, 2007]:

$$mfcc_{[m]} = \frac{1}{R}\sum_{r=1}^{R}\log\left(M_{[r]}\right)\cos\left(\frac{2\pi}{R}\left(r+\frac{1}{2}\right)m\right), \tag{2.14}$$

em que m é o número de coeficientes e R é a quantidade de filtros.

A *DCT* é um método geralmente utilizado em processamento digital de sinais e de imagens pelo fato de proporcionar compressão de dados. Em outras palavras, o sinal original é representado em novos eixos perpendiculares. A função do cosseno é gerar novas bases ortogonais e atribuir um valor nestas bases para cada conjunto fornecido.

Considerando-se que operações de adição e multiplicação possuem custo constante, é necessário, para obter-se a saída dos filtros $M_{[r]}$, a realização de N multiplicações. Para obtenção de cada coeficiente $mfcc_{[m]}$, são necessárias mais R multiplicações. Por fim, para obter o custo dos $mfcc_{[m]}$ tem-se que somar $N\log(N)+N+mR$, resultando na ordem de complexidade $O(N\log(N))$.

2.5.2 Centro do espectro (ou centróide)

O centróide (*Spectral Centroid* - S) é considerado o centróide do espectro de frequências e está relacionado com o tom da vocalização [Schubert et al., 2004; Huang et al., 2009; Han et al., 2011]. Sua definição se assemelha ao do centro de massa e representa a distribuição de potência sobre a frequência. Este é definido como a largura de frequências da sílaba em torno do ponto central do espectro:

$$S = \frac{\sum_{n=0}^{M} n|X_n|^2}{\sum_{n=0}^{M}|X_n|^2}, \tag{2.15}$$

em que X_n é a DFT (Transformada Discreta de Fourier) da sílaba, M é metade positiva do espectro de X_n e n é a frequência. O custo de obtenção de S é aproximado a $2M$ mais a FFT com $N\log(N)$, ficando $2M + N\log(N)$ e na ordem de complexidade continua dominando a FFT com $O(N\log N)$.

2.5.3 Largura de Banda do Sinal

A largura de banda (*Signal Bandwidth* - B) inclui a ocupação da frequência fundamental e de seus principais harmônicos. Esta característica é calculada como a média dos pontos da DFT de cada sílaba [Huang et al., 2009]. É definida como:

$$B = \sqrt{\frac{\sum_{n=0}^{M}(n-S)^2|X_n|}{\sum_{n=0}^{M}|X_n|^2}}, \tag{2.16}$$

em que X_n é a DFT da sílaba, M é metade positiva do espectro de X_n e S é o centróide. Para obter o custo do B é necessário calcular primeiramente S. Por fim, o custo de B é $2M + 2M + N\log(N)$ e a complexidade $N\log(N)$.

2.5.4 Taxa de cruzamento por zero (*Zero-crossing Rate (ZC)*)

A taxa de cruzamento por zero (*Zero-crossing Rate* - ZC) indica quantas vezes o sinal da sílaba teve uma transição de um valor positivo a um valor negativo, ou na inversa. Esta característica foi implementada por Huang et al. [2009] e proporciona uma estimação do comprimento da sílaba. A distância entre cruzes por zero junto com a transformada Wavelet foi utilizada por Mallat [1991] para caracterizar o transiente do sinal. Esta característica, implementada com um limiar, visa eliminar os efeitos de cruzamento no eixo x produzidos pelo ruído aditivo e é definida como:

$$T = \frac{1}{2}\sum_{n=0}^{L-1}|tsgn(x_n) - tsgn(x_{n+1})|, \tag{2.17}$$

$$tsgn(x_n) = \begin{cases} +1 & x_n \geq \eta \\ -1 & x_n \leq -\eta \end{cases} \tag{2.18}$$

Em que x_n é o valor de amplitude da amostra. Esta característica pertence ao domínio do tempo, não sendo necessário o cálculo prévio da FFT. Se L é o comprimento do sinal, o custo está associado a L comparações e à ordem de complexidade $O(L)$.

2.5.5 Energia do sinal

A energia (E) de um sinal é uma característica de análise no domínio do tempo. Foi utilizada por Vaca-Castaño & Rodriguez [2010] no processo de segmentação dos sinas bioacústicos. Esta função é definida como:

$$E = \sum_{n} [x_{(n)} w_{(n)}]^2, \tag{2.19}$$

em que $x_{(n)}$ é o sinal e $w_{(n)}$ é uma janela de Hamming (seção 2.3). Sendo o comprimento do sinal L, esta característica possui um custo aproximado de L multiplicações e a complexidade resultante é $O(L)$.

2.5.6 Potência do sinal

De forma similar à energia (E) pode-se definir como característica para representar um sinal a potência (Pw) com custo L [Deller et al., 1993]:

$$Pw = \frac{1}{M} \sum_{n}^{M} |[x(n)]|^2, \tag{2.20}$$

em que $x_{(n)}$ é o sinal e M é o número total de amostras.

2.5.7 Pitch

O Pitch (P) é definido como a diferença entre os dois picos máximos da auto-correlação, figura 2.10 [Peeters, 2004; Plack et al., 2005]. Esta característica é considerada uma medida de similaridade dentro de uma forma de onda e relaciona-se com a frequência fundamental do sinal. Em outras palavras, é a frequência com a qual a forma de onde se repete calculada como:

$$R_{xx} = \frac{1}{L} \sum_{r=1}^{L} (x(n)x(n+r)), \tag{2.21}$$

$$Pitch = n_{Max_1(R_{xx})} - n_{Max_2(R_{xx})}, \tag{2.22}$$

onde L é a quantidade de amostras da vocalização. Neste cálculo, realizado no domínio do tempo, são necessárias $2L-1$ operações para obtenção de R_{xx} e posteriormente deve-se percorrer o vetor R_{xx} à procura da posição dos máximos Max_1 e Max_2. Por fim, o custo é $2L - 1 + L = 3L - 1$ e a ordem de complexidade $O(L)$.

Figura 2.10. Obtenção do Pitch.

2.5.8 Entropia

No trabalho realizado por Han et al. [2011] a entropia de Shannon (H) e a entropia de Rényi (H_α) são utilizadas como características do sinal para reforçar o resultado da classificação. A entropia de Shannon [1948] H indica o grau de previsibilidade de um sinal e é calculada como:

$$H(X) = E[I_{(p)}] = \sum_{i=1}^{n} p_i I_{(p)} = -\sum_{i=1}^{n} p_i \log_2(p_i). \tag{2.23}$$

Nesta equação, E indica o valor esperado, I é o peso da informação contida no sinal $X = \{x_1, x_2, x_3, \ldots, x_n\}$ e p_i é a probabilidade de cada valor acontecer.

A entropia de Rényi de ordem $\alpha \geq 0$ permite obter uma média de diferentes probabilidades e é definida como:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log_2 \sum_{i=1}^{n} p_i^\alpha, \tag{2.24}$$

em que p_i é a probabilidade de ocorrência de $X = \{x_1, x_2, x_3, \ldots, x_n\}$. No limite, quando $\alpha \to 1$, o resultado aproxima-se da entropia de Shannon. Para facilitar a comparação com o trabalho de Han et al. [2011], utilizamos como valor de parâmetro $\alpha = 3$.

Para realizar o cálculo destas entropias deve-se calcular p_i primeiro. A estimação deste parâmetro é realizada gerando-se o histograma dos valores de vetor $X = \{x_1, x_2, x_3, \ldots, x_n\}$. Este cálculo precisa separar os valores em i conjuntos para que a frequência relativa seja obtida. A separação dos valores é realizada por comparação, desta forma, se o comprimento do sinal for igual a L amostras, são necessárias L comparações. Para o cálculo completo do custo deve-se acrescentar i multiplicações provenientes do número de conjuntos do histograma, resultando em um custo total igual a $L + i$ e a ordem de complexidade $O(L)$.

2.5.9 Resumo das características implementadas

Resumidamente, a tabela 2.1 apresenta a complexidade computacional das características usadas. Nesta tabela as características encontram-se ordenadas em forma decrescente segundo o custo aproximado.

Características	**Ordem de complexidade**	**Custo computacional**
Pitch	$O(L)$	$3L-1$
B	$O(N\log(N))$	$2M+2M+N\log(N)$
12 *MFCC's*	$O(N\log(N))$	$N\log(N)+N+mR$
S	$O(N\log(N))$	$2M+N\log(N)$
H	$O(L)$	$L+i$
H_α	$O(L)$	$L+i$
ZC	$O(L)$	L
E	$O(L)$	L
Pw	$O(L)$	L

Tabela 2.1. Complexidade das características.

Além das características citadas anteriormente, existem outras caraterísticas que são usadas em aplicações de reconhecimento de fala, por exemplo os coeficientes LPC (*Linear Predictive Coding*). Tais coeficientes são úteis para representar sinais acústicos suaves e, segundo Trifa et al. [2008], não são eficientes para representar sinais com transições abruptas, como sinais de pássaros e anuros. Devido a essa razão, LPC não foram usados neste trabalho.

2.6 Técnicas de classificação

Nesta seção são descritos os fundamentos das técnicas de aprendizagem de máquina usadas para classificar os sinais bioacústicos. Considerando-se a transformada do sinal em vetores de características como uma nova representação deste, que somente será útil se for possível identificar padrões que não são evidentes nas amostras originais. A plicação das técnicas de classificação associam os conjuntos de características que formam cada vetor com um padrão, possibilitando desta forma, reconhecer anuros pelas similaridades das características.

Na literatura existem diversas técnicas de classificação tais como: Árvore de Decisão, Redes Neurais, *Naive Bayes*, *k-Nearest Neighbor (SVM)*, *Support Vector Machine (SVM)* entre outras. Cada uma destas técnicas possui vantagens e desvantagens, como assim cenários de aplicação mais propícios. Em Kotsiantis [2007] é realizada uma revisão de técnicas de classificação usando métricas como acurácia, velocidade de treino e velocidade de classificação entre outras.

A escolhas dos classificadores k-NN e SVM foi baseada na revisão dos trabalhos relacionados, apresentados no próximo 3, nos quais é destacado o desempenho superior destas técnicas de classificação.

2.6.1 k-Nearest Neighbors

Em reconhecimento de padrões, *k-Nearest Neighbors (k-NN)* é o método usado para classificar o objeto desconhecido X, baseando-se em um conjunto k de exemplos mais próximos, dentro de um espaço com n características. A decisão de classificar X como pertencente à categoria m depende unicamente dos valores das características e da classe a qual pertencem $(x_1, m), (x_2, m), \cdots, (x_n, m)$, sendo realizado para os k vizinhos mais próximos de X [Cover & Hart, 1967].

Mais precisamente, o método atribui o novo objeto X à categoria m, à qual pertence a maioria de seus k vizinhos. A figura 2.11 apresenta um exemplo com $k = 5$. A métrica de similaridade normalmente usada é a distância Euclidiana, que é calculada entre o vetor desconhecido e cada um dos vetores no espaço de treino como:

$$d = \sqrt{(x_1 - X_1)^2 + (x_2 - X_2)^2 + \cdots, (x_n - X_n)^2}, \tag{2.25}$$

em que x_n é o valor numérico de cada característica conhecida e X_n é o valor da mesma característica, mas da classe desconhecida.

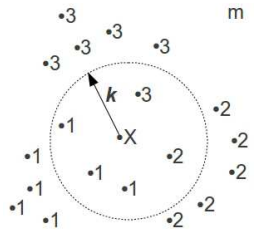


Figura 2.11. Exemplo de decisão *5-NN*.

O funcionamento de k-NN é descrito abaixo, de acordo com Theodoridis & Koutroumbas [2006]:

1. defina um valor para k, ou seja, a quantidade de vizinhos mais próximos;

2. calcule a distância da nova amostra a ser classificada a todas as amostras de treinamento;

3. identifique os k vizinhos mais próximos, independentemente do rótulo das classes;

4. conte o número de vizinhos mais próximos que pertencem a cada classe do problema;

5. classifique a nova amostra atribuindo-lhe a classe mais freqüente na vizinhança.

Este processo de classificação pode ser computacionalmente exaustivo quando o conjunto de treinamento possui muitos dados. Por este motivo, a desvantagem de k-NN é a complexidade computacional envolvida na obtenção dos k vizinhos mais próximos. Entretanto, uma vantagem é a independência quanto à distribuição de dados no espaço de características.

2.6.2 Support Vector Machines

Um método de classificação que normalmente apresenta alto grau de desempenho é *Support Vector Machines (SVMs)* [Kotsiantis, 2007]. Desenvolvido por Vapnik [1995], é um dos mais populares algoritmos de classificação. Este método transforma os vetores de características em um espaço de dimensões maiores, em que as classes podem ser separadas linearmente por hiperplanos [Marsland, 2009]. A separação entre classes determina o hiperplano ótimo, o qual maximiza a margem M, conforme figura 2.12.

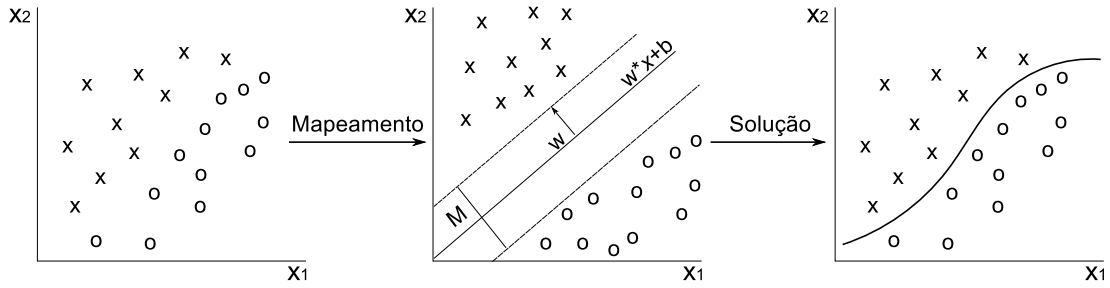


Figura 2.12. Hiperplano de separação ótima, mapeamento e solução com *SVM*.

Para o caso de duas classes (figura 2.12) o plano de separação é:

$$y = wx + b, \tag{2.26}$$

em que w (*support vector*) é o vetor perpendicular à linha que determina a separação entre classes e b é um escalar. A separação ótima é a que maximiza a distância M (figura 2.12) que é calculada como:

$$M = \frac{1}{2|w|} = \frac{1}{2\sqrt{ww}}. \tag{2.27}$$

As classes, as quais pertencem as amostras desconhecidas x, são determinadas pela aplicação da equação 2.26. Se o resultado é "positivo", a amostra é classificada como pertencente a x_1 e se o resultado é "negativo", pertence a x_2. No caso em que a quantidade de classes é maior que dois, são utilizadas estratégias como: um contra todos, todos contra todos, entre outras.

A transformação dos dados em um espaço m-dimensional, para tornar o problema linearmente separável, é realizada pela aplicação de uma função *Kernel*. Exitem diversas funções *Kernel*: polinomial, exponencial e tangencial entre outras, cujo objetivo é maximizar o resultado da classificação. A função *Kernel* modifica a equação do hiperplano de separação da seguinte forma Huang et al. [2009]:

$$y = sgn\left(\sum_{i=1}^{m} Kwx + b\right), \tag{2.28}$$

em que sgn é a função sinal.

Para permitir uma certa flexibilidade na separação das classes, SVM possui um parâmetro C de penalização (*soft margin*). Esse parâmetro controla a precisão da decisão, permitindo erros de classificação relaxando as margens rígidas. Em outras palavras C possibilita a criação de margens flexíveis. Quanto mais o valor de C aumenta, maior é a flexibilidade do modelo.

2.6.3 Considerações sobre as técnicas de classificação

Extensos trabalhos em reconhecimento de fala utilizam modelos escondidos de Markov (*HMM*) para classificar [Campbell, 1997]. *HMM* é um modelo estatístico no qual é assumido que o sistema a ser representado é um processo de Markov de parâmetros desconhecidos, com a capacidade de modelar series temporais. Na abordagem de diminuição de custos de nosso método optamos por não fragmentar as sílabas em unidades menores (*frames*) tornando-se desnecessário a utilização de *HMM*. Também foi descartada a alternativa de classificação baseada em árvore de decisão (*C4.5*), por possuir desempenho menor comparada com k-NN e SVM e pelo fato de mudar a forma da árvore para cada conjunto de características testado [Kotsiantis, 2007].

2.7 Ferramentas de avaliação de características

A manipulação das características permite aumentar a taxa de classificação ou eliminar características inúteis. As características são utilizadas em combinações. Assim,

para determinar as combinações, utilizamos dois critérios: o ganho de informação (IG) e algoritmo genético (GA). Para comparar o resultados provenientes das diferentes combinações e das diferentes técnicas de classificação, utilizamos o teste estatístico de *Wilcoxon*.

2.7.1 Ganho da informação

Das características descritas anteriormente, além do custo, é interessante calcular o nível de informação com o qual cada característica contribui para a tarefa de classificação. Isto possibilita a realização de um "ranking" baseado nesse nível de informação. Embora existam diversos tipos de obtenção de "ranking", neste trabalho é utilizada a técnica baseada em ganho da informação (*Information Gain - IG*).

Shannon [1948] definiu a entropia como quantidade real de informação numa mensagem codificada dentro de um alfabeto com distribuição de probabilidades $P(x_i)$, onde x_i é um simbolo do alfabeto. Em outras palavras, no contexto de classificação, a entropia é uma medida de incerteza para a tomada de decisões, que é calculada como:

$$H(X) = -\sum_i P(x_i)\log_2(P(x_i)). \tag{2.29}$$

A entropia condicional (também conhecida como informação mútua), é calculada da seguinte forma:

$$H(X/Y) = -\sum_j (P(y_j)\sum_i P(x_i/y_j)\log_2(P(x_i/y_j))), \tag{2.30}$$

em que $P(y_j)$ é a probabilidade a priori para todos os valores de Y e $P(x_i/y_j)$ a probabilidade de X dados os valores de Y. O ganho da informação (IG), que representa a quantidade de informação que o atributo Y fornece para a determinação da classe X, pode ser calculado como Leite et al. [2006]:

$$R = H(X) - H(X/Y). \tag{2.31}$$

2.7.2 Algoritmo genético

O algoritmo genético (*Genetic Algorithm - GA*), é uma meta-heurística utilizada para encontrar a combinação ótima de cada conjunto de características mediante um processo de *wrapper-based feature selection* (Deb [2001]; Raymer et al. [2000]). Cada combinação de características, que é avaliada pelo classificador, é representada como um

cromossomo. Após a avaliação, é realizado um ranking em ordem decrescente segundo o resultado da classificação. O conjunto de características com melhor desempenho passa para a próxima geração do algoritmo, para evoluir a solução. Inicialmente, cada população possui vinte cromossomos, que são sorteados de forma aleatória. Em aplicações de seleção de características, cada cromossomo é representado como um número binário, sendo que cada bit representa um gene contendo 1 (um) ou 0 (zero), que indica presença ou ausência da característica respectivamente, conforme figura 2.13.

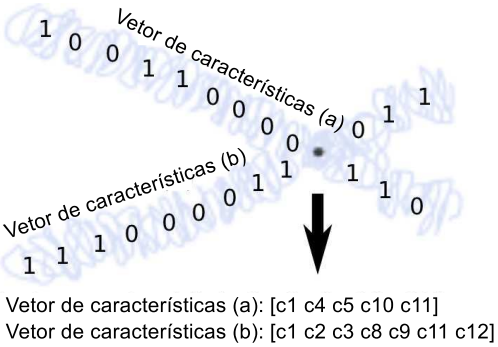


Figura 2.13. Correspondência entre cromossomo e o vetor de características.

Em cada geração, três operações são realizadas: elitismo, cruzamento e mutação. A operação de elitismo permite que os melhores cromossomos passem diretamente para a próxima geração. A operação de cruzamento é realizada sobre pares de cromossomos, com o objetivo de combinar as melhores características de ambos. Por cada operação de cruzamento são usados dois cromossomos e obtidos dois novos cromossomos. A operação de mutação é aplicada individualmente aos cromossomos restantes. Nesta operação, a quantidade de genes a serem alterados, assim como a presença ou ausência de tais genes, é realizada aleatoriamente.

2.7.3 Teste estatístico

O teste não paramétrico de *Wilcoxon* pode ser usado para comparar o resultado de duas classificações. Após ser obtido o resultado da classificação, é de interesse saber se o resultado obtido pode ser considerado empate. Este tipo de teste pode ser aplicado quando os dados são obtidos através do esquema de pareamento [Wilcoxon, 1945]. Os seguintes passos devem ser seguidos na sua construção:

1. Ordenar os valores das duas classificações por ordem crescente e atribuir o respectivo número de posição. No caso de existirem valores repetidos, a posição é calculada pela média aritmética das posições que receberiam se não fossem repetidos;

2. Designar D_1 e D_2 as médias das distribuições de frequências relativas dos resultados obtidos com validação cruzada;

3. Estabelecer a hipótese nula (H_0) e a hipótese alternativa (H_1);

4. Encontrar as zonas de rejeição e escolher a hipótese.

2.7.4 Amostragem estratificada

A amostragem estratificada é um processo de sub-amostragem de populações que podem ser separas em estratos. Cada estrato é um conjunto de indivíduos da população com características homogêneas. No procedimento de sub-amostragem seleciona-se uma percentual de amostras aleatoriamente de cada estrato. O número de amostras escolhidas de cada estrato deve ser proporcional ao seu número de elementos [Ross, 2003; McKay et al., 1979].

No processo de sub-amostragem, todos os elementos da população devem possuir igual probabilidade de serem escolhidos. A extração dos elementos de cada estrato tem que ser de forma aleatória e sem reposição. Para encontrar o número de elementos a serem extraídos de cada estrato deve-se definir a tolerância máxima de erro E_0. Se esta tolerância for de 5% então $E_0 = 0,05$.

Para o cálculo da quantidade de amostras necessárias a serem extraídas (n) primeiro deve-se realizar uma aproximação desta quantidade aplicando a equação seguinte:

$$n_0 = \frac{1}{E_0}, \tag{2.32}$$

e posteriormente:

$$n = \frac{Nn_0}{N + n_0}, \tag{2.33}$$

em que N representa o total de amostras da população. Para obter o valor n como percentagem aplicamos, $\frac{100n}{N}$. Esta fração indica quantas amostras devem ser escolhidas de cada estrato. O quantidade de amostras retiradas (s) podem ser calculadas como:

$$s = \frac{Ln}{N}, \tag{2.34}$$

em que L é o número de amostras em cada estrato. O objectivo de utilizar amostragem estratificada em nossos experimentos é conseguir representar um conjunto de amostras através de um conjunto de dimensões menores, possibilitando reduzir o número de simulações sem perder precisão no resultado.

2.8 As Redes de Sensores Sem Fio

Uma *RSSF* é um tipo especial de rede *ad-hoc* composta por dispositivos com recursos limitados, chamados nós sensores [Akyildiz et al., 2002]. Esses sensores são capazes de monitorar um ambiente, coletar dados, realizar processamento localmente e disseminar os dados coletados. O maior objetivo dessas redes é monitorar o ambiente para detectar e avaliar eventos de interesse. Este tipo de redes diferem de redes de computadores tradicionais em vários aspectos. Normalmente essas redes possuem um grande número de nós distribuídos, com restrições de processamento, transmissão e tempo de vida das baterias.

2.8.1 Redução de informação e consumo de energia

Representar cada vocalização com um número reduzido de coeficientes nos permite diminuir a quantidade de informação necessária a ser transmitida, diminuindo também a quantidade de pacotes necessários, aumentando a vida útil da rede.

Como foi descrito na seção 1.6 ha três possibilidades de transmissão: (i) as amostras completas dos áudios viajam até o nó *sink*, (ii) só as características que o classificador precisa são transmitidas até o *sink*, provocando uma redução de informação considerável, mas impossibilitando a recuperação dos áudios, ou (iii) a classificação é realizada no nó sensor, reduzindo mais a informação necessária a ser transmitida comparado com (ii), mas tendo como contrapartida o custo de processamento mais elevado dos três casos. Comparando os dois últimos casos, o cenário (ii) possui uma vantagem adicional, ou seja, a classificação é realizada no nó receptor que não possui restrições de energia ou de processamento.

Em uma *RSSF*, cada pacote de dados transmitido implica em um custo de energia relativamente alto quando comparado com outras tarefas. Um dos maiores desafios nestas redes é maximizar a quantidade de informação transmitida em cada pacote. Neste tipo de aplicação, a quantidade de informação coletada depende da necessidade, ou seja, depende do fenômeno físico que se deseja mensurar e do tipo de sensor utilizado. A tabela 2.2 extraída do trabalho realizado por Mainwaring et al. [2002], mostra o custo de energia da cada operação.

Em nossa aplicação, um sensor de áudio com uma taxa de amostragem de 8 kHz geraria 8000 $^{\text{amostras}}/_{\text{seg}}$, considerando que cada amostra é representada por um byte, seriam gerados 8000 $^{\text{bytes}}/_{\text{seg}}$, uma taxa de informação elevada para um nó sensor. Um pacote de um sensor MICA possui uma carga útil (*payload*) de 25 bytes, o qual requer 320 pacotes para transmitir as amostras correspondentes de cada segundo de áudio.

Operação	Custo (nAh)
Transmitir um pacote	20.000
Receber um pacote	8.000
Ler o ADC	0.011
Ler EEPROM	1.111
Escrever/Apagar EEPROM	83.333

Tabela 2.2. Energia requerida por diferentes operação do sensor MICA Weather Board. Tabela adaptada de Mainwaring et al. [2002]

Esta operação consumiria 6.4 mAh, implicando que em um segundo seria consumida a mesma quantidade de energia que, segundo Mainwaring et al. [2002], consumiam em um dia inteiro. Além disso, existe uma limitação de processamento, sendo quase impossível para um nó sensor transmitir 320 pacotes por segundo. Nos desenvolvemos uma técnica eficiente de redução de informação plausível de ser aplicada nas condições restritivas descritas.

2.9 Considerações finais

Neste capítulo foi apresentado um resumo dos conceitos teóricos necessários para entender o desenvolvimento de nosso trabalho. As diversas escolhas realizadas para definir os métodos baseiam-se nestes conceitos. O ferramental teórico estudado possibilista entender os trabalhos relacionados com nossa pesquisa (capítulo 3).

A tabela 2.1 possibilita relacionar a taxa de acerto, de cada conjunto de características, com o custo ou a complexidade computacional do método, podendo-se definir a relação custo-benefício. As ferramentas de avaliação permitem encontrar os conjuntos mínimos de características necessários para classificar com a máxima acurácia. Desta forma, reduzir os conjuntos de características impacta na quantidade de informação transmitida e no consumo de energia dos sensores (tabela 2.2). Além disso, o estudo da quantização e das frequências de amostragem viabiliza uma forma de simular cenários com diferentes tipo de hardware.

Nos próximos capítulos são apresentados os trabalhos relacionados, nosso método, os resultados da experimentação, os estudos de casos e as conclusões.

3

Trabalhos relacionados

Neste capítulo é apresentado um resumo das soluções recentes para os problemas de classificação de sinais bioacústicos e monitoramento ambiental. Os trabalhos revisados abordam a identificação de diversas espécies de animais e, em especial, anuros. Quanto à classificação, os métodos revisados baseiam-se na utilização de conjuntos diversos de características temporais e espectrais. A tarefa de monitoramento ambiental é abordada nos trabalhos que integram as técnicas de classificação com Redes de Sensores Sem Fio (RSSF). O monitoramento mediante utilização destas redes enfrenta alguns desafios, como perda de informação, devido à aquisição e transmissão dos dados, e redução na taxa de classificação correta, devido às condições de ruídos externos.

3.1 Estado da arte

Pesquisadores australianos estudaram uma solução para monitorar a espécie de anuro Cane-Toad, conhecido também como *Bufo marinus*, no Parque Nacional Kakadu, situado no Norte do Austrália [Shukla et al., 2004]. Fazendo uso de uma RSSF, os autores conseguiram monitorar uma extensa área geográfica. Esse trabalho aborda detalhes de implementação da RSSF tais como o hardware utilizado e a forma de espalhamento dos sensores pela área a ser monitorada. O objetivo principal de Shukla et al. [2004] era identificar as zonas habitadas pela espécie Cane-toad. Esta espécie de anuro foi introduzida na Austrália em 1935 e, desde então, espalhou-se ameaçando as espécies originais e tornando-se um forte predador [Price, 1996]. Por este motivo, o interesse dos pesquisadores era identificar as áreas de maior ação desta espécie, a quantidade de indivíduos existentes, o crescimento populacional e a taxa de migração para outras áreas. Os nós sensores utilizados foram denominados *PLEB*. Em cada *PLEB*,

foi implementada a técnica de classificação desenvolvida por Taylor et al. [1996], em que o espectrograma e a energia de seus *pixels* são usadas como características para identificar as diferentes espécies. Os autores destacaram que a maior dificuldade do classificador é a presença de ruídos, permitindo concluir que o 90% da área do parque Kakadu era habitada pelo anuro *Bufo marinus*, possibilitando identificar as zonas de concentração das populações.

Os desafios de implementação do trabalho de Shukla et al. [2004] foram estudados por Hu et al. [2005]. O cenário de aplicação e as limitações dos hardwares de baixo custo utilizados requiriu que Hu et al. [2005] desenvolvessem algoritmos que possibilitassem trabalhar com maiores frequências de amostragem, redução de ruídos e escalonamento de tarefas. A área do parque monitorada é de difícil acesso e a maior atividade da espécies ocorre na época de fortes chuvas, impossibilitando a intervenção humana. Por este motivo, a RSSF foi dividida em dois tipos de nós denominados pelos autores de *Stargate* e *Hybrid*. Os nós híbridos realizam a filtragem dos ruídos, a compressão dos áudios e a transmissão até o nó central *Stargate*, encarregado de efetuar a classificação. Embora os autores tenham aplicado uma técnica de compressão, para que a classificação seja realizada, é necessário transmitir um elevado número de amostras dos áudios. A tarefa de transmissão representa um custo elevado de energia, diminuindo a vida da rede.

Para minimizar o impacto deste problema, Bulusu & Hu [2008] implementaram uma técnica de amostragem parcial. Esta técnica transmite para o classificador uma taxa menor de amostras que as necessárias para reconstruir o sinal. As amostras transmitidas correspondem-se com a forma do envelope do sinal. Tais amostras, definidas aleatoriamente, reduzem a quantidade de informação transmitida, produzindo uma redução do custo de comunicação, e portanto do consumo de energia. Porém, ao diminuir a quantidade de amostras, o desempenho do classificador também diminui. A técnica de Bulusu & Hu [2008] é caracterizada por uma variável de sub-amostragem e um filtro de correlação que determina a espécie desconhecida. Dependendo do valor de sub-amostragem escolhido, aumentam os falsos positivos e falsos negativos detectados pelo sistema de classificação. Com valores iguais a cinco, o classificador proporciona uma taxa de 40% de falsos positivos na identificação do *Bufo marinus*. Com valores próximos a 200, a taxa de falsos positivos se aproxima a 70% para a espécie Cyclorana cryptotis. Embora o trabalho de Bulusu & Hu [2008] não apresente ótimo desempenho na classificação, o algoritmo desenvolvido possui um custo computacional reduzido que requer 2 KBytes de memória. Este método também estende a vida dos nós sensores e aumenta a capacidade de armazenamento.

No ano de 2006, uma tentativa de monitoramento de animais a longo prazo,

mais especificamente de anuros, foi realizada pela Universidade de Kentucky, pelo Museu Nacional de Ciências Naturais (Madrid, Espanha) e pela Universidade de Lisboa(Portugal). Preocupados com o estudo de diversos animais, esse grupo desenvolveu um hardware para capturar som de diferentes espécies, chamado *Amphibulator* [Cambron & Bowker, 2006]. O *Amphibulator* foi projetado para maximizar a capacidade de armazenamento de dados, minimizar o consumo da bateria, ser a prova d'água e para ser usado em ambientes hostis e de difícil acesso. O desenvolvimento deste hardware focou-se na capacidade de armazenamento e na durabilidade da bateria, portanto não transmite os dados coletados e não realiza classificação automática das espécies.

Já no ano de 2007, Cai et al. [2007] empregaram uma RSSF para monitorar o ecossistema de pássaros e o impacto da urbanização em Brisbane, Austrália. O trabalho teve como foco o uso de uma Rede Neural (*Artificial neural network - ANN)* para classificação. As características utilizadas foram os MFCCs, em que cada vetor de entrada para o classificador era formado por:

$$V = [MFCC_t, \Delta MFCC_t] \tag{3.1}$$

sendo que a característica dinâmica $\Delta MFCC_t$ é $MFCC_{t+1} - MFCC_{t-1}$. Esta característica diferencial serve para modelar as mudanças entre *frames* vizinhos. Os pássaros são uma espécie animal difícil de se monitorar, pois poucas vezes em seus trajetos transitam perto do nó sensor. Por este motivo, a razão sinal-ruído (S/N) de suas vocalizações diminui notavelmente. Para aumentar a taxa de classificação, Cai et al. [2007] utilizam um pré-processamento constituído de um filtro adaptativo que minimiza o erro quadrático médio (*Minimum Mean Square Error - MMSE*). A taxa de acerto deste método é 86%, sobre uma base de dados com quatorze espécies de pássaros. Entretanto, destaca-se que o principal problema do sistema é separar e detectar espécies que vocalizam ao mesmo tempo.

Recentemente o trabalho de monitoramento de Vaca-Castaño & Rodriguez [2010] concentrou-se na identificação de pássaros e anuros, combinando os MFCCs e k-NN. A técnica desenvolvida divide os sinas em *frames*, o que gera elevada quantidade de informação. Por este motivo, implementaram *PCA* (*Principal Component Analysis*) para reduzir as dimensões dos vetores de características. O método foi integrado com uma RSSF para obtenção dos sinais, na qual a topologia era formada por nós sensores que coletavam os áudios e transmitiam os vetores de características após aplicação de *PCA* até o nó central, implicando em um elevado custo de comunicação. O nó central (*Master sensor node - MSN*) possui maior capacidade de armazenamento e processamento para realizar a classificação. Esse método de monitoramento e classificação permitiu-

lhes classificar com um desempenho de 86,60% dez espécies de pássaros e com 91% vinte espécies de anuros. Porem, o maior desafio enfrentado por Vaca-Castaño & Rodriguez [2010] foi a grande quantidade de informação gerada a partir da sucessão de *frames*, sendo necessário aplicar um método de redução de dimensionalidade. Isto requer um custo de processamento adicional, aumentando o consumo de energia e diminuindo a vida útil da RSSF.

Geralmente os trabalhos de monitoramento contem duas etapas diferentes. A primeira etapa considera os parâmetros da RSSF tais como: topologia, custos e viabilidade de implementação, e uma segunda etapa que foca-se na técnica de classificação. A técnica de classificação demarca como será implementado o monitoramento na RSSF. Em outras palavras, esta define qual o processamento que cada nó deverá executar, quais características devem ser extraídas e transmitidas e como será realizada a classificação. Encontram-se na literatura diversos trabalhos de classificação de animais, nos quais o interesse principal é apenas o método de classificação, dispensando a possibilidade de aplicação sobre uma RSSF. Estes trabalhos utilizam bases de dados previamente armazenadas, misturando diferentes características e classificadores.

A abordagem de monitoramento de Shukla et al. [2004] utilizou a técnica de classificação desenvolvida por Taylor et al. [1996]. Nesta técnica, é realizado o espectrograma do sinal, mediante a aplicação da FFT. A informação da energia de cada pixel e dos pixels vizinhos do espectrograma foi utilizada como característica pelo classificador, mostrando um desempenho de 60% de classificação correta em condições de ruído decorrente da presença de outras espécies. Para classificar a energia dos pixels do espectrograma foi utilizada uma árvore de decisão [Quinlan, 1993]. Este processo foi dividido em três níveis de decisão usando uma quantidade de tempo diferente do espectro. Primeiramente, o espectro é divido em *frames* de 30 ms, posteriormente em *frames* de 300 ms e por fim em *frames* de 3 s. A técnica apresentada de divisão em intervalos de diferentes durações mostrou-se interessante para representar a característica de repetição periódica de vocalizações. Esta característica pode ser explorada como padrão que diferencia algumas espécies de anuros. Já a efetividade do classificador não superou 60% nos casos onde a mistura de seis diferentes espécies estavam presentes em uma mesma região.

Uma abordagem mais completa de reconhecimento de anuros, mediante técnicas de aprendizagem de máquina, foi a proposta por Huang et al. [2009]. Nesse trabalho utilizou-se uma base de dados com cinco espécies diferentes de anuros. Para isto, foi implementado um algoritmo que permite dividir um áudio, contendo uma vocalização completa, em unidades menores chamadas sílabas. De cada sílaba, três características foram extraídas: o *Spectral Centroid*, o *Signal Bandwidth* e o *Zero-Crossing Rate*.

Estas três características foram combinadas em um vetor, para ser classificado com a utilização de k-NN e SVM. A taxa de classificação correta varia em um intervalo de 83,87% até 100% com k-NN e 82,04% até 100% com SVM. Estes resultados devem-se ao fato de as características serem altamente discriminantes nessas cinco espécies. Outro fator que influencia no resultado é que os áudios utilizados eram livres de ruídos, diminuindo a probabilidade de confundir o classificador. Entretanto, a condição livre de ruído não é alcançável no cenário de floresta, onde outros animais estão presentes.

Dos trabalhos revisados, o mais recente é de Han et al. [2011]. Os autores implementaram uma abordagem similar ao trabalho de Huang et al. [2009], utilizando a mesma abordagem de segmentação em sílabas e o centróide como característica. Esta característica foi combinada com a entropia de Shanonn e de Rényi para reforçar os resultados da classificação. Os resultados obtidos utilizando k-NN e 9 espécies de anuros, variaram entre 83% e 100% de acerto.

Uma alternativa à transformada clássica de *Fourier* e à utilização de características temporais é usar a transformada *Wavelet* (seção 2.4.2). Esta transformada permite representar uma vocalização em tempo e frequência com maior resolução que a FFT. Esta transformada também possui a vantagem de ter um custo computacional menor. Estudos recentes em reconhecimento de fala começaram a usar a transformada Wavelet como alternativa à transformada de Fourier. Uma comparação entre Wavelet e os *MFCCs* foi realizado por Modic et al. [2003]. Esses autores destacaram a abrangência da transformada Wavelet sobre todo o espectro de frequências. Já no trabalho de Wu & Lin [2009] é combinada a escala de frequências Mel com a transformada Wavelet, aplicando-se os filtros Wavelet na escala Mel para reconhecimento de voz.

O trabalho desenvolvido por Yen & Fu [2002] utilizou a transformada *Wavelet discreta* para extrair informação relevante das vocalizações de anuros. Nesse trabalho foram utilizadas quatro espécies de anuros, classificadas como: dois de tipo 3, uma de tipo 2 e uma de tipo 1. As espécies de tipo 3 podem ser identificadas pelo valor do Pitch, após a aplicação de um filtro passa banda (*Infinite impulse response - IIR*). Para identificar as espécies tipo 2 e 1 é necessário aplicar os passos completos do método, ou seja, os quatro passos descritos a seguir:

1. aplicar o filtro passa banda,

2. aplicar o algoritmo de agrupamento (*Clustering*) sobre os valores do espectrograma,

3. realizar a transformada Wavelet para obter a energia de cada nó da árvore Wavelet (WPT),

4. reduzir a dimensionalidade dos vetores de características (*Fisher's criterion*).

O passo dois é realizado sobre um intervalo de tempo de 3,5 s, permitindo identificar padrões de repetição dentro do espectrograma. A finalidade deste passo é separar vocalizações de anuros de outros animais com padrões temporais diferentes. Posteriormente, é realizada a transformada Wavelet até que seja obtida uma estrutura de árvore completa, com os resultados dos coeficientes de escala e detalhes. De cada nó da árvore é obtido o valor de energia dos coeficientes para formar os vetores de características. A cada vetor de característica é aplicado um método para reduzir a dimensionalidade. A classificação, realizada por uma ANN do tipo MPL (*Multi-Layer Perceptron*) permite separar as espécies de tipo 1 e 2.

O resultado do método é comparado com diferentes conjuntos de características. Neste caso, são os LPC e os valores da TDFT (*Time-Dependent Fourier Transform*), que produziram desempenhos inferiores. Embora esta abordagem possua a vantagem de diminuir os falsos positivos e os falsos negativos pela aplicação da etapa de agrupamento, este método requer a realização de duas transformadas, um método para reduzir dimensões, um filtro, uma técnica de agrupamento e uma técnica de classificação, tornando o sistema pouco prático para uso em conjunto com uma RSSF.

Além de trabalhos de classificação de anuros, na literatura, existem diversos estudos nos quais o objetivo é classificar diferentes espécies animais, tais como: grilos, pássaros e elefantes entre outros, o que permite mostrar a diversidade de métodos existentes. Exemplos desses diferentes *frameworks* de classificação são os trabalhos de Mellinger & Clark [2000], Chesmore [2001], Lee et al. [2006], Vilches et al. [2006], Fagerlund [2007] e Trifa et al. [2008].

Mellinger & Clark [2000] desenvolveram um método para classificação automática de baleias (*Balaena mysticetus*). O método baseia-se na correlação do espectrograma desconhecido com um espectrograma padrão. Este método é comparado pelo autor com um filtro de correlação, um modelo oculto de Markov (*HMM*) e uma ANN. A correlação do espectrograma, realizada com 114 amostras de som, superou o filtro de correlação em 15% e o modelo *HMM* em 2%, mas não superou a técnica ANN. O maior problema enfrentado pelos autores foi a varição temporal das vocalizações, sendo difícil determinar o começo e o fim destas para realizar as correlações.

O trabalho realizado por Chesmore [2001] apresenta um sistema capaz de identificar 13 espécies diferentes de grilos e 10 espécies de pássaros. O sistema faz uso da técnica TDSC (*Time domain signal coding*) para extrair as características e de uma ANN para a classificação. A técnica TDSC utiliza a forma de onda do sinal, e portanto, pertence ao domínio do tempo. O sistema foi testado com 13 grilos, em condições sem

ruído, obtendo-se um resultado próximo a 99% quando a razão S/N foi igual a -40 dB, e próximo a 68% quando S/N foi -10 dB. Também foram testadas 25 espécies diferentes de insetos, incluindo gafanhotos, mas em condições com maior nível de ruído. Esta técnica é uma alternativa a uma análise espectral, possível de ser aplicada em tempo real, sem pré-processamento ou filtragem. Entretanto, a taxa de classificação varia amplamente com os valores do ruído.

Usando como característica a média dos coeficientes *MFCC* e a média dos coeficientes LPCC (*Linear Predication Cepstral Coefficients*), Lee et al. [2006] classificaram 420 espécies diferentes de pássaros. O método de extração de sílabas usada pelo autor foi proposto por Harma [2003]. Este método de extração é baseado no uso de um limiar de amplitude para determinar o começo e o fim da sílaba, ocasionando um problema de duração variável de cada sílaba. Dado que o processo de extração não gera um comprimento uniforme para todas as sílabas, o número de *frames* contidos nelas varia, gerando vetores de características de diferentes comprimentos. Para solucionar esse problema e obter uma quantidade padrão de características em cada vetor, é calculada a média das características de todos os *frames* pertencentes a cada sílaba. O resultado obtido por Lee et al. [2006] foi uma taxa de acerto que varia entre 29%, para a característica LPCC, e 87%, para a característica MFCC. Neste trabalho destacam-se dois problemas que afetam de forma direta a classificação: sílabas que não foram extraídas corretamente e ruído dos áudios que modificam os valores médios das características.

A tarefa de classificação de pássaros também foi objetivo de Vilches et al. [2006] mostrando que é possível utilizar outras técnicas de classificação, tais como árvore C4.5 e *Naive-Bayes*. Estas técnicas utilizam dados nominais, por este motivo tiveram que implementar um algoritmo de quantização de informação, trazendo como consequência um erro de quantização. Os resultados obtidos na classificação de três espécies foram 85,57% com *Naive-Bayes* e 98,39% com C4.5. Para atingir estes resultados, os autores implementaram um filtro de ruído passa-banda, com frequências de corte entre 517 Hz e 4200 Hz, para incrementar a taxa de classificação.

Outra técnica de classificação foi implementada por Fagerlund [2007], usando como características os MFCCs, os Δ-MFCCs e os $\Delta\Delta$-MFCCs em conjunto com um algoritmo de segmentação baseado no valor das frequências do espectrograma em *dB*. A taxa de classificação correta calculada com SVM para oito espécies diferentes de pássaros, resultou em aproximadamente 90%. Este resultado, juntamente com os já mencionados, indicam que é altamente possível usar técnicas de aprendizagem de máquina para classificar anuros.

A classificação de pássaros também foi o objetivo de Trifa et al. [2008], tendo como cenário uma floresta do México. Usando um modelo HMM, os autores classifi-

caram cinco espécies diferentes com uma taxa de acerto próxima a 95%. As características utilizadas foram os MFCCs e variações deles (MFCC-E, MFCC-D, MFCC-A), e os LPC. Os autores destacaram como principal problema a quantidade de amostras utilizadas para treinar o modelo HMM e a quantidade de estados desse modelo.

3.2 Sínteses dos trabalhos relacionados

A tabela 3.1 resume os aspectos fundamentais dos trabalhos relacionados neste capítulo, apresentando espécies, técnicas e resultados. Os trabalhos indicados por * foram utilizados como *baseline*, para comparar os resultados. Na sexta coluna encontra-se o intervalo de resultados obtidos pelos diferentes métodos, e a sétima coluna indica se o método foi implementado sobre uma RSSF.

3.3 Considerações finais

Os trabalhos atuais de análise de sinais bioacústicos incorporam métodos automáticos de classificação e extração de características. No entanto, não foi definido nenhum método comum (ou *framework*) para esta tarefa [Clemins, 2005]. Assim, cada experimento usa diferentes características e diferentes métodos de classificação, tornando o estudo comparativo extremamente complexo. A dificuldade para definir um método padrão é devida ao fato de cada espécie possuir características diferentes.

Foi demostrado por Nelson [1989] que a extração de características no domínio da frequência, tais como os MFCCs para a representação dos sinais bioacústicos, é mais eficiente no processo de extração e classificação da informação. Em contraste, Kogan & Margoliash [1998] asseguram que os *LPC* são apropriados para modelar qual-

Autor	**Animal**	**Características**	**Classificador**	**Resultados**	**RSSF**
Taylor et al. [1996]	Bufo marinus	Spectrograma	C4.5	60%	Não
Hu et al. [2005]	Bufo marinus	Spectrograma	C4.5	60%	Sim
Yen & Fu [2002]*	4 anuros	Wavelet Fisher's	MLP	71%	Não
Clemins [2005]	Elefantes	MFCCs PLP	HMM DTW	69% 73%	Não
Cai et al. [2007]	14 pássaros	MFCCs	ANN	81-86%	Sim
Huang et al. [2009]*	5 anuros	Centróide Largura de banda ZC	k-NN SVM	83-100% 82-100%	Não
Vaca-Castaño & Rodriguez [2010]*	10 pássaros	MFCCs	k-NN	86%	Sim
	20 anuros	PCA		91%	
Han et al. [2011]*	9 anuros	Centróide Entropia de Shannon Entropia de Rényi	k-NN	83-100%	Não

Tabela 3.1. Resumos dos trabalhos relacionados.

quer sinal, mas sem habilidade para representar transições abruptas ou não lineares, comumente presentes em sinais bioacústicos.

Embora os sistemas de classificação existentes possuam resultados satisfatórios, estes têm sido construídos e otimizados de formas diversas para cada problema. Em nossa proposta comparamos alguns destes métodos e sua aplicabilidade ao problema de classificação de anuros. Para este fim, desenvolvemos e adaptamos nosso próprio *framework* de classificação, escolhendo a combinação de melhores características, visando ser uma abordagem padrão de reconhecimento de anuros. O próximo capítulo irá discutir este método.

4

Abordagem experimental

Neste capítulo é descrito detalhadamente o processo que envolve a classificação de anuros, desde a obtenção dos sinais até o resultado da classificação, processo que é representado como um método dividido em etapas. A descrição geral do método em blocos fornece uma visão geral do funcionamento do sistema, uma descrição detalhada das tarefas que são executadas e da ordem dos passos para chegar ao resultado da classificação.

Nós trabalhamos o conceito de redução de informação. A redução é realizada transmitindo pela RSSF somente as características extraídas dos áudios que o classificador precisa. Esta abordagem possui a vantagem de transmitir menor quantidade de pacotes de informação, poupando energia. Porém, este contexto possui uma desvantagem inerente, os áudios não podem ser recuperados no nó *sink*. A figura 4 exemplifica o contexto.

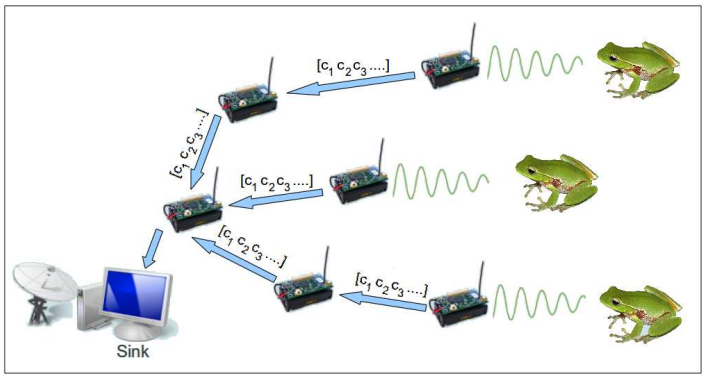


Figura 4.1. Exemplo de redução de informação, transmitindo somente as características até o classificador.

Baseados neste contexto de redução, nós conduzimos nosso trabalho objetivando definir a melhor combinação de características, tentando maximizar a relação custo-beneficio, o melhor classificador e o *framework* ótimo para classificação de anuros.

Na primeira etapa, coletamos as vocalizações dos anuros e as armazenamos para seu posterior processamento; em seguida, definimos o pré-processamento e o ajuste de seus parâmetros; depois, definimos os conjuntos de características e as suas combinações ótimas possíveis; e por fim, definimos como realizar a classificação, bem como o critério de avaliação.

4.1 Espécies e vocalizações

Primeiramente é necessário descrevermos algumas características das vocalizações das espécies com as quais é avaliado o método. As principais informações usadas pelo classificador são resultantes da análise espectral na frequência. A tabela 4.1 apresenta na segunda coluna a banda de frequência com energia superior a -60 dB e na terceira coluna, o valor médio e o desvio-padrão do *pitch* de todas as amostras da base.

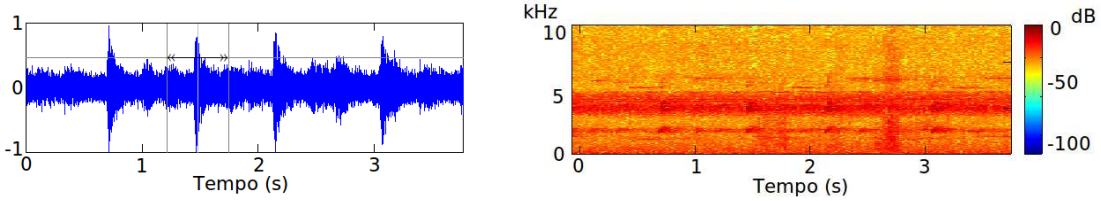
As espécies de anuros utilizadas em nossos experimentos são listadas na primeira coluna da tabela 4.1. Nesta lista as espécies (a), (d), (e), (f) e (g) foram coletadas no campus da Universidade Federal do Amazonas. Já as espécies (b), (c), (h) e (i) foram extraídas de gravações variadas de anuros da Mata Atlântica [Haddad, 2005], Bolívia [Márquez et al., 2002] e Guiana Francesa [Marty, 1999]. A utilização de especies provenientes de diferentes lugares, fornece generalidade ao método implementado. Tais amostras foram armazenadas em formato *wav* com frequência de amostragem máxima de 44,1 kHz e 32 bits por amostra, o que permite analisar sinais bioacústicos de até 22,05 kHz.

Espécie	**Banda de frequências (kHz)**	**Pitch ± Std (ms)**
(a) *Adenomera andreae*	2,10∼3,00 4,00∼6,92	0,511 ± 0,3
(b) *Ameerega trivittata*	2,00∼3,00 5,50∼7,00	0,220 ± 0,1
(c) *Hyla minuta*	1,50∼2,50 3,50∼5,00	0,404 ± 0,1
(d) *Hypsiboas cinerascens*	1,40∼1,80 3,00∼3,50	0,638 ± 0,1
(e) *Leptodactylus fuscus*	1,00∼3,50 6,50∼7,74	0,089 ± 0,1
(f) *Osteocephalus oophagus*	1,50∼3,00	0,914 ± 2,0
(g) *Rhinella granulosa*	1,70∼3,20	0,022 ± 0,2
(h) *Scinax ruber*	1,10∼4,15	0,043 ± 0,01
(i) *Hylaedactylus*	1,50∼2,50 3,50∼4,50	0,466 ± 0,1

Tabela 4.1. Características espectrais das espécies em nossa base.

Pode-se observar, na tabela 4.1, que algumas espécies, por exemplo *Leptodactylus*

fuscus, possuem mais de uma banda espectral de frequência com elevada energia na suas vocalizações. Outro detalhe importante é a existência de bandas de frequências sobrepostas entre diferentes espécies. Pelas características físicas, o sistema tem que ser capaz de obter informações sobre as vocalizações em um intervalo espectral entre 1,0 kHz e 7,7 kHz. Um exemplo deste fenômeno é ilustrado na figura 4.2(b). Nesta figura pode-se observar que ao longo do tempo permanecem duas bandas de frequências com maior energia, próximas de 2 kHz e 5 kHz.



(a) Forma de onda. (b) Espectrograma.

Figura 4.2. Vocalização da espécie *Adenomera andreae*.

Na floresta existem diversas espécies animais que emitem som. Tais sons são considerados ruídos para nosso sistema, dificultando a tarefa de classificação. Pelas características espectrais das espécies utilizadas, nas quais as bandas principais de frequência estão sobrepostas (4.1), torna-se difícil aplicarmos um filtro passa-banda sem que sejam suprimidas informações relevantes para a classificação.

Para resolver este desafio, optamos por utilizar áudios coletados em cenários reais, nos quais estão presentes ruídos diversos da floresta, bem como de outros animais ou mesmo outras espécies de anuros. Nos áudios utilizados, os valores dos ruídos ambientais variam aproximadamente entre -12 dB (ruído alto) e -30 dB (ruído baixo). Através da extração de características e do treino do classificador com áudios ruidosos obtemos duas vantagens:

- simular cenários reais, permitindo ao classificador identificar anuros em condições adversas; e
- economizar um filtro na etapa de pré-processamento, diminuindo o custo computacional do método.

4.2 Descrição do Método

A abordagem, ou método, utilizada para a identificação de anuros é composta por três etapas fundamentais: pré-processamento, extração das características e classificação.

Nesta abordagem, cada etapa constitui um conjunto de tarefas específicas. Na figura 4.3 são representadas as tarefas fundamentais de nossa abordagem. A composição de tais etapas em tarefas é detalhada a seguir.

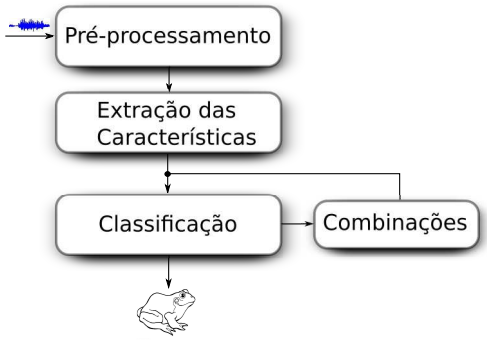


Figura 4.3. Sistema de classificação.

Na prática, as etapas de obtenção, combinação e classificação de características acontecem juntas, em outras palavras uma depende da outra. A entrada do classificador é uma possível combinação de características. Após avaliação do resultado é gerada uma nova combinação que ingressa no classificador. Este ciclo é repetido até que seja encontrado o conjunto ótimo de características.

4.2.1 Pré-processamento

Na primeira etapa, é realizado um pré-processamento que adéqua o sinal para a posterior geração das bases de dados para o classificador. Neste pré-processamento é realizada a segmentação dos sinais em unidades menores chamadas "sílabas", como foi descrito no trabalho de Huang et al. [2009], o pré-ênfase e o janelamento (figura 4.4).

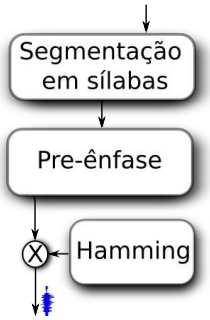


Figura 4.4. Etapa de Pré-processamento.

A etapa completa de pré-processamento começa pela segmentação, na qual são extraídas as sílabas. Este processamento de segmentação é descrito em detalhes na

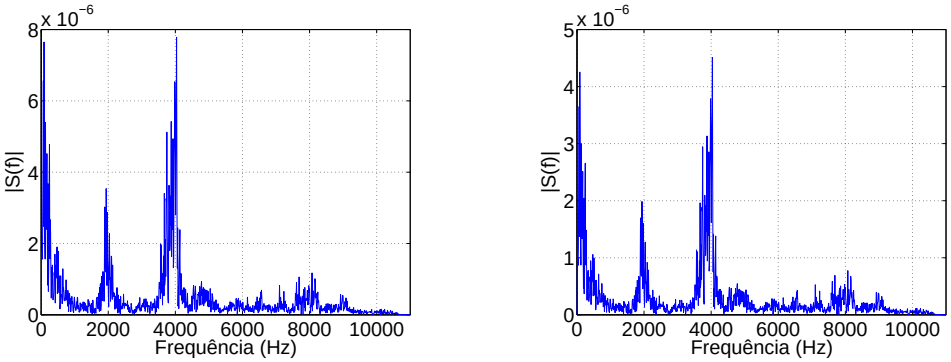
seção 4.2.1.1. Após serem obtidas as sílabas, estas ingressam no sistema como um vetor, no qual o valor de cada componente representa a amplitude do sinal de entrada. Esta é a representação digital de uma vocalização correspondente a um anuro. Sobre essa forma de representação é aplicado um filtro de pré-ênfase e uma janela de Hamming.

Dada a característica de geração dos sinais, as frequências elevadas tendem a possuir magnitudes menores, distorção ou saturação. Para compensar estes efeitos, um filtro de pré-ênfase amplifica as altas frequências, para que os detalhes destas frequências possam ser visualizados. As vocalizações dos anuros geralmente, dependendo da espécie, possuem frequências fundamentais superiores aos 1,0 kHz. Em algumas espécies, como o *Bufo ornatus*, as frequências das vocalizações começam em 2,5 kHz.

Para realizar a etapa de pré-ênfase, o sinal digitalizado $s_{[n]}$ é modificado mediante a aplicação do filtro:

$$s_{[n]} = s_{[n]} - a s_{[n-1]}. \tag{4.1}$$

Tipicamente a constante a é 0,9 em sistemas de comunicações. A escolha ótima para a é a que permita amplificar o efeito das altas frequências sem provocar distorção nas baixas frequências [Pribilova, 2003]. O efeito causado por este filtro pode ser considerado uma etapa de equalização do espectro. Na figura 4.5 é comparado o efeito da aplicação deste filtro. Observa-se na figura 4.5(b) como as baixas frequências são atenuadas e as altas são amplificadas.



(a) Sem pré-ênfase. (b) Com pré-ênfase.

Figura 4.5. Comparação de espectros após aplicação do filtro de pré-ênfase.

O processo de divisão das vocalizações em sílabas, gera um conjunto de sinais aperiódicos de comprimento menor, não necessariamente estacionários. Em trabalhos revisados e em nossa abordagem, as características são extraídas de cada sílaba [Huang et al., 2009; Vaca-Castaño & Rodriguez, 2010; Han et al., 2011]. Desta forma, elas podem não descrever adequadamente a variação temporal do sinal, principalmente

em sinais que apresentam características de variações temporais altamente dinâmicas [Campbell, 1997].

Posterior à aplicação do pré-ênfase, o sinal é multiplicado por uma janela do tipo *Hamming*. Esta janela é aplicada para diminuir o impacto do fenômeno de distorção, causado pela segmentação em sílabas. A multiplicação modifica as amplitudes dos componentes das sílabas pela aplicação da equação seguinte:

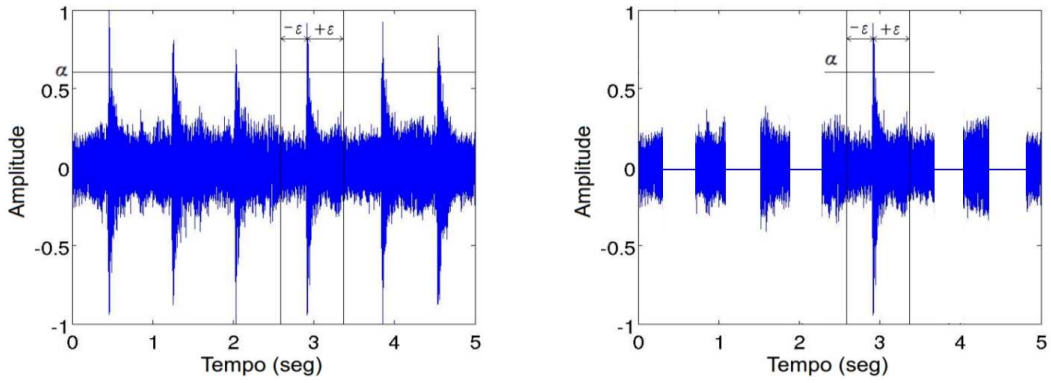
$$w_{(n)} = 0.54 - 0.46cos\left(2\pi\frac{n}{N}\right), 0 \leq n \leq N \tag{4.2}$$

em que N é o comprimento da janela e n é o número de amostra do sinal.

O resultado final do pré-processamento é um conjunto de sílabas acondicionadas para a extração ótima das características. A operação de extração de tais características é descrita na próxima etapa de nosso método, mas antes é descrito com detalhes o processo de segmentação utilizado.

4.2.1.1 Segmentação em sílabas

A segmentação é a ação de dividir o sinal em unidades menores, chamadas sílabas. Este processo é efetuado mediante a aplicação de um algoritmo iterativo. Na figura 4.6(a) é identificada uma sílaba dentro da vocalização da espécie *Adenomera andreae*, fato que inicia a segmentação.



(a) Começo da segmentação. (b) Última sílaba da segmentação.

Figura 4.6. Processo de segmentação na vocalização da espécie *Adenomera andreae*.

Existem dois métodos para segmentar o sinal (s_n) em sílabas: o método descrito por Huang et al. [2009] e o desenvolvido por Harma [2003]. Estes métodos, procuram o valor máximo de amplitude da amostra e extraem uma quantidade de componentes de $s_{[n]}$ à direita e à esquerda. O primeiro deles usa um comprimento fixo para cada

sílaba e o segundo baseia-se num limiar de amplitude. Neste trabalho optamos pela utilização do primeiro método, pois este proporciona um comprimento padrão para todas as sílabas. A segmentação em sílabas é realizada da forma seguinte:

S01. Buscar o valor máximo de amplitude do sinal s_n;

S02. se s_n for menor que o limiar α, ir ao passo S05, senão, ir ao passo S03;

S03. extrair ε milissegundos a direita e a esquerda do valor máximo, $s_{(n-\varepsilon)} \leq$ sílaba $\leq s_{(n+\varepsilon)}$, formando a sílaba;

S04. extrair as características da sílaba e apagar os valores do sinal original s_n no intervalo $n \pm \varepsilon$ (figura 4.6(b)), voltar ao passo S01;

S05. terminar a segmentação.

Na figura 4.6, a variável ε determina o comprimento da sílaba e é definida com um valor igual a 200 ms. Este valor garante a inclusão da sílaba inteira de cada classe utilizada mais uma margem de segurança. Neste processo de segmentação também deve ser definido o valor do limiar de amplitude α, de forma tal que produza a maximização no resultado da classificação. O valor a ser definido deve permitir extrair dos áudios a maior quantidade possível de sílabas mantendo a relação sinal-ruído baixa.

Para caracterizar a efeito da variável α, a segmentação foi testada com três valores possíveis: $\alpha = 0,4$, $\alpha = 0,5$ e $\alpha = 0,6$. No capítulo 5 é estudado o impacto destes valores na taxa de acerto no processo de classificação. A tabela 4.2 mostra a relação entre α e a quantidade de amostras extraídas de cada áudio armazenado, por espécie.

Espécies	Indivíduos	Sílabas		
		$\alpha = 0,4$	$\alpha = 0.5$	$\alpha = 0,6$
Adenomera andreae	8	686	528	442
Ameerega trivittata	5	673	572	339
Hyla minuta	11	300	261	225
Hypsiboas cinerascens	2	3364	3176	2898
Leptodactylus fuscus	4	315	252	233
Osteocephalus oophagus	4	130	103	84
Rhinella granulosa	3	1791	1684	1458
Scinax ruber	4	238	193	170
Hylaedactylus	8	358	326	249
Total	49	7855	7095	6098

Tabela 4.2. Número de sílabas resultantes do processo de segmentação utilizadas em nossas bases.

4.2.2 Extração das características e Geração das bases

Cada sílaba resultante da segmentação, é reduzida a uma representação na forma de vetor de características, conforme figura 4.7. Nesta etapa ocorre a redução de informação. Tais vetores podem ser compostos por uma, ou mais caraterísticas, e um rótulo que identifica a classe à qual pertence, definindo-se como classe cada espécie de anuro em particular.

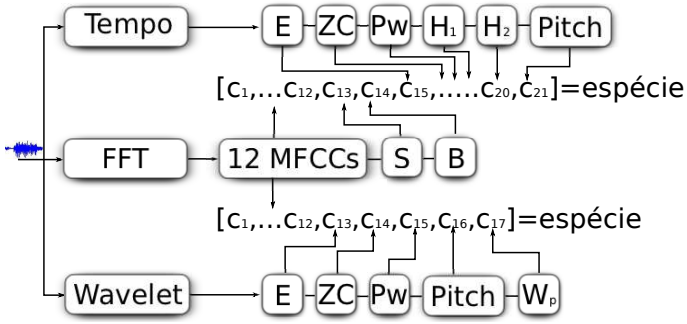


Figura 4.7. Obtenção do vetor de características.

Para facilitar a determinação do sub-conjunto ótimo de características, e diminuir a complexidade computacional devida à análise combinatória, em nossa abordagem dividimos a totalidade de características em dois conjuntos bases.

No primeiro conjunto base a ser testado incluímos todas as características temporais e as provenientes da transformada de Fourier. Nesta base incluímos as combinações de características utilizadas por Huang et al. [2009] e Han et al. [2011]. Neste caso, a forma dos vetores de características analisados pelos classificadores é a seguinte:

$$[coef_1, coef_2, coef_3, \cdots, coef_{12}, E, Pw, ZC, S, B, Pitch, H_1, H_2] -> especie \qquad (4.3)$$

em que $coef_x$ é o número de coeficiente mel, E e Pw são a energia e a potência do sinal respectivamente, ZC a taxa de cruzamento zero, S o centróide, B a largura de banda, H_1 e H_2 as entropias de Shannon e Rényi, respetivamente.

No segundo conjunto base comparamos os MFCCs com as características extraídas da transformada Wavelet, incluindo a Energia dos coeficientes de escala e detalhes utilizada por Yen & Fu [2002]. Neste caso, para a formação dos vetores de características, utilizamos as transformadas Wavelet *Haar* e *Daubechies* (Db), respectivamente, realizando o seguinte processamento:

1. Obter o Pitch (P) da sílaba;

2. aplicar a transformada Wavelet para obter os coef. de detalhes e de escala;

a) sobre os coef. de detalhes (details) e de escala (approximation) obter:

- energia (W_E^d),
- potência (W_{Pw}^d),
- diferença entre os dois máximos da autocorrelação (W_P^d), e
- taxa de cruzamento por zero dos coef. (W_{ZC}^d),

Por fim, o vetor de características resultante da transformada Wavelet, submetido ao classificador, é:

$$[P, W_E^d, W_{Pw}^d, W_P^d, W_{ZC}^d, W_E^a, W_{Pw}^a, W_P^a, W_{ZC}^a] -> especie \tag{4.4}$$

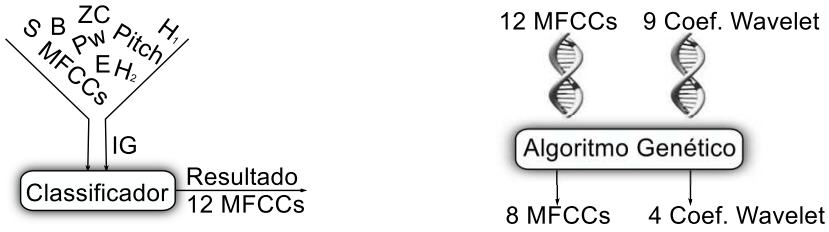
em que W é usada para diferenciar estas características das temporais, d e a indicam se foram calculados sobre os coeficientes de detalhes ou de escala respectivamente, e P o Pitch.

4.2.3 Determinação dos subconjuntos ótimos de características

Determinar a melhores características implica na tarefa de gerar todas as combinações possíveis. A abordagem combinatória possui custo computacional exponencial e nem sempre todas as combinações geradas produzem resultados interessantes. Para otimizar a tarefa de gerar as combinações e realizá-la de forma eficiente, utilizamos o valor de entropia para cada característica como critério de combinação, exemplificado na figura 4.8(a) (seção 2.7.1).

Os subconjuntos de combinações utilizadas na primeira etapa foram geradas de forma manual. Para isto, utilizando o valor de "ganho da informação (IG)" como guia para combinar as características. A métrica de avaliação usada para escolher as melhores características é a relação custo-beneficio. Nesta relação é ponderado o custo computacional de obtenção de cada subconjunto contra a taxa de classificação correta Tc, definida como a razão entre a quantidade de amostras classificadas corretamente e o total de amostras.

O mesmo critério de avaliação foi mantido para realizar a eliminação de características na segunda etapa. Nesta etapa não foram realizadas combinações de características devido ao fato de que os conjuntos MFCCs e Coef. Wavelet pertencem a transformadas distintas. Neste caso, em que não é possível realizar combinações, reduzimos cada conjunto para encontrar o número de coeficientes ótimos. Para realizar esta redução utilizamos uma abordagem baseada em algoritmos genéticos (seção 2.7.2), conforme figura 4.8(b).



(a) (b)

Figura 4.8. (a) Combinação de características na primeira base de dados mediante IG. (b) Otimização das características da segunda base de dados aplicando algoritmo genético (GA).

4.2.4 Normalização das características

Antes da implementação dos classificadores é necessário evitar que os valores das características dos vetores apresentem intervalos dinâmicos diferentes. Valores de características com diferenças significativas podem modificar o resultado da classificação. Este problema afeta principalmente o cálculo de distância Euclidiana entre os vetores, modificando, por exemplo, o desempenho do classificador *k-NN*.

A normalização dos vetores restringe os valores das características em um intervalo determinado. Geralmente, os valores são normalizados para obter-se média zero e variância um [Theodoridis et al., 2010]. Assumindo que cada característica possui N valores, a normalização é realizada pela aplicação de:

$$\hat{x}_i = \frac{x_i - \bar{x}}{\sigma},\ i = 1, 2, \ldots, N. \tag{4.5}$$

em que $\bar{x}$ e σ são os valores da média e do desvio padrão, respectivamente, da característica x, calculada para todas as classes. Por fim, $\hat{x}_i$ é o valor normalizado.

4.2.5 Classificação

Na etapa de classificação de nossa abordagem (figura 4.3) são utilizadas duas técnicas, k-NN e SVM (seção 2.6). A escolha destas duas técnicas deve-se ao fato de serem consideradas como as de melhor desempenho dentro da área de aprendizagem de máquina e por serem utilizadas na maioria dos trabalhos relacionados, fato que facilita a comparação de resultados.

O ajuste dos parâmetros dos modelos de classificação é realizado de forma empírica, com o objetivo de maximizar a taxa de acerto. A obtenção dos resultados (capítulo 5), permite estabelecer uma comparação entre SVM e k-NN e a escolha da técnica que melhor se adapte ao problema de reconhecimento de anuros.

4.3 Considerações finais sobre o método

Cada etapa do sistema possui parâmetros a serem configurados: (1) na etapa de pré-processamento deve-se escolher o valor de pré-ênfase e os valores das variáveis do processo de segmentação; (2) na extração das características foram definidas as estratégias de combinação e o critério de avaliação; (3) e por último os parâmetros dos classificadores. O desafio presente é encontrar a combinação de todos os valores descritos com o objetivo final de maximizar o resultado da classificação.

A figura 4.9 exemplifica, de forma sucinta, os caminhos escolhidos para a condução de nossos experimentos até chegarmos a definir o *framework* ótimo para a classificação de anuros. Como pode ser observado, a etapa de pré-processamento é compartilhada. Em outras palavras, o pré-processamento aplicado é o mesmo nos dois caminhos escolhidos em nossa metodologia.

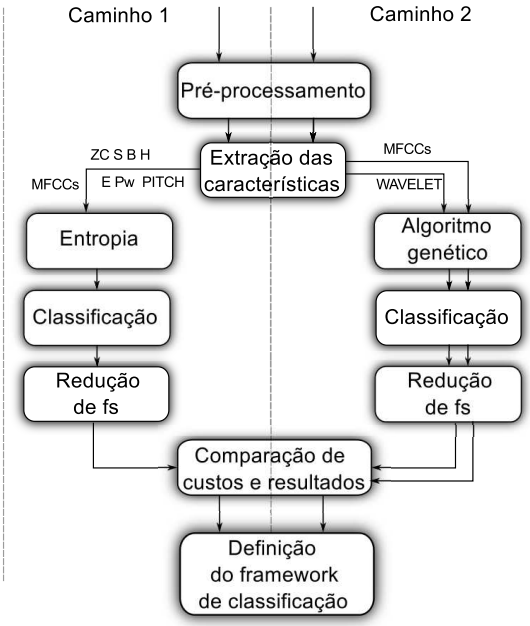


Figura 4.9. Metodologia adotada em nossa abordagem.

Após ser definido a melhor abordagem de reconhecimento de espécies individualmente, incluindo as características com melhor relação custo-beneficio, abordamos o problema de reconhecimento de grupo de anuros. A metodologia aplicada, e os resultados obtidos, são descritos no capítulo 6.4.

A utilização de áudios de espécies nativas de diferentes lugares, e diversos indivíduos de cada espécie, contribui para tornar a abordagem um método geral. Embora este não seja provado com diferentes animais, pelos indícios encontrados nos trabalhos relacionados, acreditamos que é possível sua aplicação em outros contextos. Em todos

nossos experimentos utilizamos o software de cálculo numérico MATLAB para realizar o pré-processamento e a extração das características e a ferramenta de classificação WEKA para obter os resultados dos classificadores. No restante desta dissertação, são apresentados os resultados das classificações de cada base avaliando: a taxa de acerto, o custo computacional e o efeito da redução de f_s.

5

Comparação entre características temporais e espectrais

Como foi descrito no capítulo 4, dividimos nossa abordagem em duas formas de seleção de características. Por este motivo, e para dar clareza à explicação, decidimos separar os resultados obtidos em dois capítulos. Neste capítulo, apresentamos os resultados correspondentes à primeira abordagem, os quais foram obtidos combinando as características: $MFCCs$, E, $Pitch$, Pw, H_1, H_2, ZC, S e B. O critério de seleção utilizado é o ganho da informação (IG) e os classificadores avaliados são k-NN e SVM.

5.1 Ganho da informação

No contexto de classificação, a entropia pode ser utilizada como uma medida de incerteza (seção 2.7.1). Em outras palavras, essa medida permite responder a pergunta: "Com quanta informação uma característica contribui para a determinação de uma classe?". A determinação da relevância das características é fundamental para maximizar a taxa de classificação.

É possível testar todas as combinações de características e avaliar o impacto na taxa de classificação, porém, a abordagem combinatória possui custo computacional exponencial, a qual aumenta com a quantidade de características. Ordenar as características segundo o nível de informação fornecido é útil como guia na escolha das combinações, uma vez que é possível estabelecermos uma ordem de relevância entre elas. A tabela 5.1 apresenta o ganho de informação para as bases de áudios com três níveis de ruído deferentes, $\alpha = 0,4$, $\alpha = 0,5$ e $\alpha = 0,6$.

$\alpha = 0,4$		$\alpha = 0,5$		$\alpha = 0,6$	
IG	**Característica**	**IG**	**Característica**	**IG**	**Característica**
1,28	S	1,33	S	1,35	S
1,27	Coef 4	1,30	Coef 4	1,32	Coef 4
1,23	Pitch	1,26	Pitch	1,24	Pitch
1,17	Coef 6	1,19	Coef 6	1,18	Coef 6
0,98	Coef 5	1,02	Coef 5	1,04	Coef 5
0,93	Coef 7	0,97	Coef 7	0,97	Coef 7
0,86	Coef 3	0,89	Coef 3	0,91	Coef 3
0,83	Coef 1	0,84	Coef 1	0,89	E
0,75	Coef 8	0,77	E	0,89	Pw
0,75	Coef 2	0,77	Pw	0,79	Coef 1
0,69	Coef 11	0,77	Coef 8	0,72	Coef 8
0,66	E	0,73	Coef 2	0,72	Coef 2
0,66	Pw	0,69	Coef 11	0,69	H_1
0,59	B	0,65	H_1	0,64	Coef 11
0,58	H_1	0,65	B	0,63	B
0,50	ZC	0,54	ZC	0,61	ZC
0,47	Coef 10	0,49	Coef 10	0,49	Coef 12
0,47	Coef 12	0,49	Coef 12	0,49	Coef 10
0,47	Coef 9	0,45	Coef 9	0,45	H_2
0,42	H_2	0,44	H_2	0,42	Coef 9

Tabela 5.1. IG das características: Pitch, centroide (S), largura de banda (B), taxa de cruzamento por zero (ZC), Potência (Pw), Energia (E), Coeficientes Mel (Coef), entropia de Shannon (H_1) e entropia de Rényi (H_2)

Podemos observar na tabela 5.1 as caraterísticas com maior contribuição para escolha entre as diferentes espécies. Elas são o centroide (S), os coeficientes 3, 4, 5, 6, 7 e o Pitch. Embora exista uma diferença na forma de cálculo da características S e Pitch, ambas relacionam-se diretamente com a frequência fundamental das vocalizações, indicando que essa é a informação mais relevante. Podemos observar também que nem todos os MFCCs contribuem com informações relevantes e que as características correspondentes ao domínio do tempo (E, Pw e ZC) são perturbadas pelo nível de ruído no ambiente.

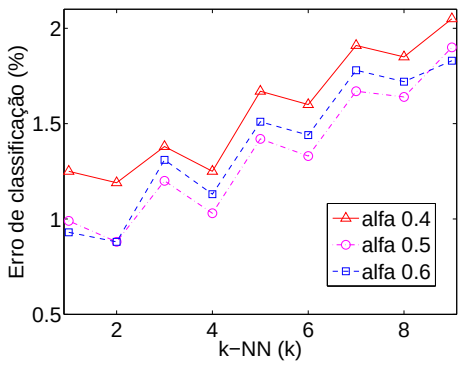
O IG ordena as características sem considerar as taxas de classificação. Porém, é importante verificarmos se os subconjuntos de características indicados como mais relevantes, de fato produzem elevada taxa de acerto. Logo, os classificadores precisam ser utilizados. Antes porém, é necessário otimizar os parâmetros destes.

5.2 Otimizar os parâmetros dos classificadores

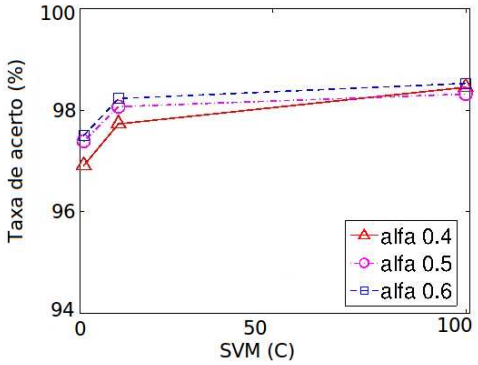
Em problemas de classificação, uma opção para mensurar o desempenho de um classificador é avaliando a taxa de erro. Desta forma, o classificador prediz a classe de cada instância. Se esta for correta, é contabilizada como sucesso, caso contrário, é contabilizada como erro. A taxa de erro é portanto a proporção de erros cometidos ao longo de um conjunto de instâncias, representando o desempenho global do classificador.

Os classificadores k-NN e SVM, descritos na seção 2.6, devem ser configurados com os valores dos parâmetros que minimizam o erro de classificação. No caso do k-NN, deve ser escolhida a quantidade de vizinhos com os quais será realizada a comparação, o valor k. Este valor depende das características utilizadas. Em outras palavras, se o conjunto de características muda, o valor k pode ou não mudar. Para sermos justos, calculamos este valor para cada conjunto de características avaliadas.

A figura 5.1(a) apresenta um exemplo do procedimento utilizado para encontrarmos o erro de classificação segundo o valor de k. Neste exemplo, pode-se observar que o valor $k = 2$ minimiza o erro nas três bases. Para as classificações utilizou-se validação cruzada com dez conjuntos ($fold = 10$).



(a) Verificação da taxa de erro em função do valor de k.



(b) Verificação da taxa de acerto em função do valor de C utilizando Kernel polinomial ($p = 2$).

Figura 5.1. Exemplificação da escolha empírica dos parâmetros ótimos dos classificadores. (a) Valor ótimo $k = 2$ e (b) valor $C = 100$

Para o classificador SVM foi escolhido o tipo de Kernel polinomial de grau dois ($p = 2$), para facilitar a comparação com o trabalho desenvolvido por Huang et al. [2009]. A constante de penalização, que maximiza a taxa classificação, foi encontrada empiricamente avaliando os valores: $C = 0,1$, $C = 10$ e $C = 100$ (seção 2.6). Um exemplo deste procedimento é apresentado na figura 5.1(b).

Nos exemplos apresentados na figura 5.1 a escolha dos parâmetros que maximizam o resultado da classificação seriam $k = 2$, no caso de k-NN, e $C = 100$, no caso de SVM. O procedimento de otimização descrito foi aplicado a cada conjunto de combinações de características avaliadas na próxima seção.

5.3 Resultados com k-NN

Os resultados da classificação realizada com k-NN são apresentados na tabela 5.2. Esta tabela mostra a relação entre a taxa de classificação e a variável de amplitude da segmentação em sílabas α. A primeira coluna indica a combinação de característica testada e as demais colunas indicam a taxa de classificação correta. A combinação das características encontra-se em ordem decrescente segundo os custos aproximados de cada combinação. Os números entre parênteses correspondem ao valor de k que minimiza o erro de classificação para cada determinado conjunto. As combinação entre características de diferentes conjuntos foram geradas manualmente, utilizando o critério de IG. Desta forma, foi realizado o cruzamento entre características com maior ganho de informação.

Características	**k-NN**		
	$\alpha = 0,4$	$\alpha = 0,5$	$\alpha = 0,6$
ZCSBEPitchPwH_1H_2-MFCCs	**99,35%**(1)*	**99,57%**(1)*	99,54%(1)*
ZCSBEPitchPw-MFCCs	99,27%(1)*	99,47%(1)*	99,45%(1)*
ZCSBEPitch-MFCCs	99,26%(1)*	99,52%(1)*	99,49%(1)*
ZCSBPitch-MFCCs	99,26%(1)*	99,53%(1)*	99,49%(1)*
ZCSBEPitch	93,07%(4)	95,19%(3)	96,08%(1)
ZCSBE-MFCCs	99,26%(1)*	99,54%(1)*	**99,55%**(2)*
ZCSB-MFCCs	99,26%(1)*	99,56%(1)*	99,26%(1)*
ZCSBE	92,47%(4)*	94,77%(3)	95,40%(1)
SH_1H_2	82,71%(9)	87,02%(11)	87,07%(7)
ZCSBPitch	90,69%(6)	93,51%(6)*	94,67%(3)
ZCEPPw	79,54%(11)	84,32%(11)	79,54%(11)
Pitch-MFCCs	99,22%(1)*	99,37%(1)*	99,22%(1)*
ZCSB	88,63%(7)	91,71%(5)	92,85%(3)
Pitch	74,18%(1)	77,26%(1)	74,18%(1)
ZC-MFCCs	99,24%(1)*	99,45%(1)*	99,24%(1)*
E-MFCCs	99,27%(1)*	99,49%(2)*	99,27%(1)*
B-MFCCs	99,33%(1)*	99,54%(1)*	99,33%(1)*
S-MFCCs	99,21%(1)*	99,42%(1)*	99,21%(1)*
MFCCs	99,19%(9)*	99,36%(2)*	99,19%(1)*

Tabela 5.2. Taxa de classificação com k-NN em relação a α, usando validação cruzada ($fold = 10$). Um * designa uma diferença estatística não significativa em relação ao resultado apresentado em negrito na mesma coluna.

Em todas as combinações os 12 MFCCs são utilizados como um conjunto inteiro, que não pode ser subdividido, para facilitar as comparações com os métodos existentes na literatura. A primeira linha da tabela 5.2 inclui todas as características implementadas. Esta combinação possui o maior custo de processamento e desempenho. Comparando as taxas de acerto entres as linhas identificamos que existe informação redundante entre características. Em outras palavras, se compararmos o resultado da combinação Pitch-MFCCs, S-MFCCs e MFCCs observamos que os resultados são valores muito próximos indicando que S, Pitch e os MFCCs capturam informações similares do espectro de frequências. Esta conclusão é coerente com o antecipado pelo cálculo

do IG.

Em negrito, destacamos em cada coluna da tabela 5.2 o melhor resultado. Na terceira coluna podemos comparar o resultado correspondente ao conjunto ZCSBE-MFCCs somante com os MFCCs, indicando uma diferença igual a 0,36%. Esta diferença, que não chegou a superar 1%, indica que a melhor escolha entre esses dois conjuntos de características são os MFCCs por possuirem um custo computacional menor. As características S, ZC, E e Pw individualmente possuem custo menor que os MFCCs. No entanto, é necessário combiná-las para obtermos um resultado satisfatório na classificação. Esta combinação eleva o custo computacional superando os MFCCs, mas possuindo um desempenho inferior.

Os conjuntos de características utilizadas por Huang et al. [2009] e Han et al. [2011], ZCSB e SH_1H_2 respectivamente, obtiveram resultados inferiores aos obtidos com os MFCCs com custo computacional mais elevado, portanto, não são a escolha mais adequada. Desta comparação, conseguimos responder a pergunta gerada em nossa revisão bibliográfica, sobre qual é o melhor conjunto de características existente, comparando custos e benefícios. Se comparamos os métodos ZCSB e SH_1H_2 observamos que os resultados obtidos por Huang et al. [2009] são levemente superiores.

Observando os resultados dos MFCCs, pode-se notar que a taxa de acerto aumenta levemente entre $\pm 0.17\%$ com a variação do limiar α, indicando baixa correlação com os níveis de ruído. Em outras palavras, os MFCC's apresentam menor correlação com a variação de α, fato que indica uma maior imunidade aos ruídos ambientais, em comparação com as características restantes.

Para mostrar os erros de classificação, montamos uma matriz de confusão, que indica quais são as espécies mais confundidas pelo classificador, o que ajuda a verificar de fato o que o classificador aprendeu em nosso experimento. Esta matriz indica quais espécies e quantas sílabas de cada uma foram confundidas. Cada linha é a espécie que originou a sílaba e a coluna é a espécie com a qual foi confundida.

As tabelas 5.3 e 5.4 apresentam as matrizes para o conjunto de características ZCSB e SH_1H_2 respetivamente, para o valor $\alpha = 0,5$. Pode-se observar que as espécies mais confundidas por estes métodos são *Hylaedactylus* com *Rhinella granulosa* e com *Ameerega trivittata*, devido à proximidade entre as características sonoras dessas espécies (4.1).

A figura 5.2 é uma alternativa de representação à matriz de confusão e à tabela de classificação. Esta representação fornece uma visão espacial da distribuição do resultado de classificação, incluindo erros e acertos. O eixo horizontal é o valor da classe correta e o eixo vertical indica a classe na qual cada vetor de características foi associado. Nesta representação quanto mais as características ficam concentradas na

Espécie	**k-NN,** $k = 5$								
	a	b	c	d	e	f	g	h	i
a	484	34	0	8	1	1	0	0	0
b	10	554	1	0	0	0	4	0	3
c	4	12	191	0	0	0	1	0	53
d	13	0	0	299	1	6	4	2	1
e	3	2	0	3	194	1	30	5	14
f	5	0	0	27	1	60	4	6	0
g	2	18	2	1	7	1	1580	5	**68**
h	1	7	0	9	8	9	55	95	9
i	1	3	6	5	7	1	**95**	8	3050

Tabela 5.3. Matriz de confusão do método de Huang et al. [2009].

Espécie	**k-NN,** $k = 11$								
	a	b	c	d	e	f	g	h	i
a	409	99	3	0	0	1	16	0	0
b	27	348	19	0	1	0	11	0	166
c	7	25	171	0	0	0	10	0	48
d	22	0	0	268	0	7	20	9	0
e	0	4	1	1	213	1	25	2	5
f	2	0	0	16	12	43	18	12	0
g	71	12	9	1 9	5	2	1511	6	49
h	8	4	1	17	19	7	79	51	7
i	1	40	5	6	3	1	105	1	3014

Tabela 5.4. Matriz de confusão do método de Han et al. [2011].

diagonal principal, menor é o erro da classificação.

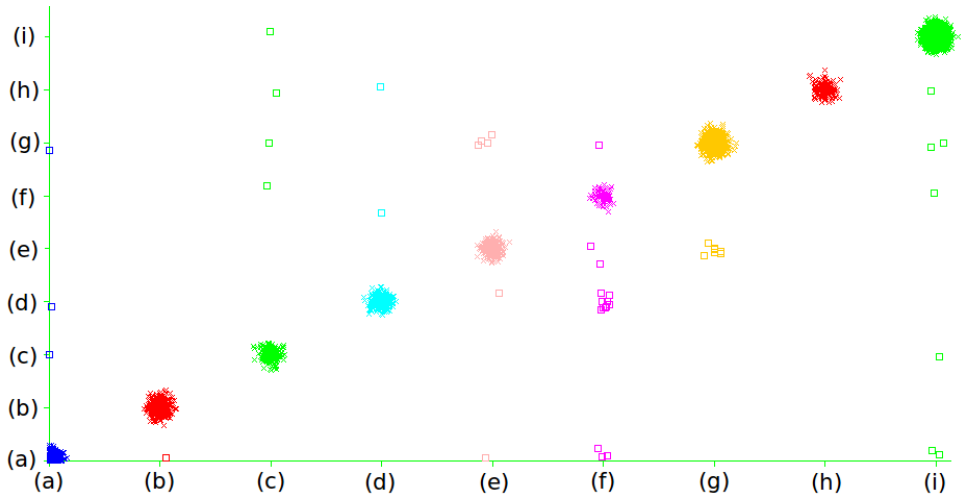


Figura 5.2. Representação espacial da classificação com os MFCCs. Espécies: (**a**) *Adenomera andreae*, (**b**) *Ameerega trivittata*, (**c**) *Hyla minuta*, (**d**) *Hypsiboas cinerascens*, (**e**) *Leptodactylus fuscus*, (**f**) *Osteocephalus oophagus*, (**g**) *Rhinella granulosa*, (**h**) *Scinax ruber* e (**i**) *Hylaedactylus*.

Como pode-se observar na tabela 5.2, a diferença entre os resultados máximos de alguns conjuntos de características é inferior a 1%. Nestes casos, é necessário saber se

essa variação do resultado é significativa, ou se pode ser considerada empate. Por este motivo, na próxima seção aplicamos o teste estatístico de Wilcoxon.

5.4 Verificação de empate

Nas situações em que a variação porcentual dos resultados entre conjuntos é comparativamente pequena, não é possível afirmar com certeza a superioridade de um método em relação aos demais. Pode-se observar nas tabelas 5.2 e 5.5 que não existe uma diferença significativa entre os maiores resultados das classificações, considerando que resultados superiores a 95% podem ser suficientes para nossa aplicação. Neste contexto é necessário determinar se existe diferença de desempenho entre conjuntos de características.

Para saber se existe uma diferença na classificação realizada com os diferentes conjuntos de características aplicamos o teste estatístico de *Wilcoxon* (seção 2.7.3), utilizando como dados os resultados de cada $fold$ na validação cruzada [Wilcoxon, 1945]. Este teste nos permite saber, com um nível de confiança igual a 95%, se os resultados nas classificações foram equivalentes.

Todos os resultados das classificações da tabela 5.2 foram comparados contra o conjunto que obteve o melhor desempenho em cada coluna. O simbolo * indica os resultados que, comparados com o melhor, aceitaram a hipótese H_0, em outras palavras, são os conjuntos que apresentaram empates nos resultados. Da tabela 5.2 podemos observar que o resultado do conjunto dos MFCCs empatou com o resultado do melhor conjunto, nas três bases. A mesma forma de aplicação do teste foi utilizada na seção seguinte para avaliar a possibilidade de empate entre os resultados obtidos com o classificador SVM.

5.5 Classificação com SVM

As mesmas combinações de características usadas na tabela 5.2 foram utilizadas com o classificador SVM. Como foi dito anteriormente, o objetivo de comparar os resultados de SVM com k-NN é tentar identificar se existe ou não diferença no desempenho de classificação de anuros. A tabela 5.5 mostra os resultados das classificações com SVM em relação à variável α, com Kernel polinomial de grau dois e $C = 100$. Os resultados foram obtidos usando *Cross-validation* com 10 *folds*.

Neste caso, o resultado melhor é 99,59% com $\alpha = 0.6$ e a combinação de todas as características. Novamente, pode-se observar que os MFCCs atingiram uma taxa

Características	**SVM, p=2,** $C = 100$		
	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$
ZCSBEPitchPwH_1H_2-MFCCs	99,26%*	**99,63%***	99,47%*
ZCSBEPitchPw-MFCCs	**99,31%***	99,52%*	**99,59%***
ZCSBEPitch-MFCCs	99,26%*	99,53%*	99,57%*
ZCSBEPitch	84,90%	91,06%	91,80%
ZCSBPitch-MFCCs	99,22%*	99,42%*	99,54%*
SH_1H_2	77,13%	78,85%	80,50%
ZCSBPitch	81,52%	84,14%	86,99%
ZCSBE-MFCCs	99,22%*	99,46%*	99,57%*
ZCSBE	81,96%	85,91%	87,14%
ZCSB	75,88%	80,01%	82,05%
Pitch-MFCCs	98,71%	99,11%	99,18%
ZCSB-MFCCs	99,23%*	99,42%*	99,57%*
ZC-MFCCs	98,82%	99,18%	99,29%
E-MFCCs	98,99%	99,39%*	99,36%*
ZCEPPw	65,58%	80,26%	82,76%
Pitch	58,43%	63,87%	67,36%
B-MFCCs	99,00%*	99,40%*	99,37%*
S-MFCCs	99,00%*	99,14%	99,26%*
MFCCs	98,66%	99,15%	99,14%

Tabela 5.5. Taxa de classificação de SVM em relação a α, usando *Cross-validation* ($fold = 10$). Um * designa uma diferença estatística não significativa em relação ao resultado apresentado em negrito na mesma coluna.

de classificação próxima à máxima, mas com custo computacional menor. O teste estatístico de Wilcoxon indicou empate entre as combinações que utilizam os MFCCs, S, B, E, Pitch e Pw, confirmando que existe informação desnecessária ou repetida.

O *Pitch*, sendo uma característica com alta entropia, não contribui com uma diferença substancial na classificação quando é combinado com os MFCCs. O mesmo acontece com o centroide, indicando que os MFCCs captam informação de todo o espectro de frequências.

Nas duas tabelas de classificação, os MFCCs obtiveram melhor desempenho com o valor $\alpha = 0,5$, utilizado no restante de nosso trabalho. Comparativamente, os resultados obtidos usando k-NN e SVM não apresentam maiores diferenças, entretanto a técnica SVM demanda um tempo de treinamento maior. Por este motivo continuamos nossos experimentos utilizando k-NN.

Como foi explicado anteriormente, no contexto de monitoramento utilizando RSSF, avaliar somente a taxa de acerto não é suficiente. Nestes casos, é fundamental correlacionar a taxa de acerto com o custo computacional de cada conjunto de características, para desta forma, identificar a melhor relação entre custos e benefícios. Por este motivo, na próxima seção simulamos quatro situações de diferentes hardwares.

5.6 Estudo de caso

Na prática, os nós sensores possuem um módulo conversor analógico-digital. Na maioria dos hardwares de baixo custo, as amostras são representadas com 8 bits e taxas de amostragem próximas de 8 kHz ou 4 kHz.

As classificações anteriores foram realizadas sobre uma base de áudios onde cada amostra foi representada com 32 bits a uma taxa de amostragem de 44,1 kHz. Para simular uma situação real quantizamos os áudios uniformemente em 256 níveis (8 bits) e diminuímos a frequência de amostragem em quatro (até 11 kHz), cinco (até 8 kHz) e oito vezes (até 5,5 kHz), produzindo uma diminuição na quantidade de informação adquirida pelos sensores de 75%, 81% e 87% respectivamente.

As amostras foram convertidas a números inteiros em uma escala entre [-128, 127], ocasionando consequentemente o denominado ruído de quantificação. A relação sinal-ruído de quantização, descrita na seção 2.2, resulta:

$$\frac{S}{N_q} = 1,76 + 6,02n - 20log\left(\frac{V_{max}}{V}\right) = 49,92dB \tag{5.1}$$

A tabela 5.6 apresenta a redução da taxa de acerto ocasionada pelo uso de 8 bits com 11 kHz, 8 kHz e 5,5 kHz, para representar algumas combinações da base $\alpha = 0,5$. Observa-se que, em um contexto real, usando sensores econômicos, a taxa de acerto é levemente diminuída, embora ainda seja uma taxa de acerto que pode ser considerada suficiente para nossa aplicação.

Características	**Classificação com kNN,** $k = 2$			
	32 bits 44,1 kHz	8 bits 11 kHz	8 bits 8 kHz	8 bits 5,5 kHz
ZCSBEPwPitch-MFCCs	99,47%	99,55%	98,83%	97,41%
ZCSB-MFCCs	99,56%	99,52%	98,90%	97,31%
ZCEPwPitch	84,32%	85,06%	81,04%	85,06%
SH_1H_2	87,02%	83,76%	86,11%	83,76%
ZCSB	91,71%	87,81%	89,03%	86,35%
MFCCs	99,36%	99,42%	98,51%	99,42%

Tabela 5.6. Comparação no resultado da classificação usando 32 bits com 44,1 kHz e 8 bits com 11 kHz, 8 kHz e 5,5 kHz por amostra.

É interessante observar na última linha da tabela 5.6 que a diminuição da informação quatro vezes por amostra não afetou significativamente o desempenho dos MFCCs, indicando que estes coeficientes são mais imunes ao ruído de quantização quando comparados às demais características. A figura 5.3 apresenta de forma gráfica a correlação entre f_s e o erro de classificação entre os métodos de Huang et al. [2009]; Han et al. [2011]; Vaca-Castaño & Rodriguez [2010] e o conjunto ZCSBEPwPitch-MFCCs.

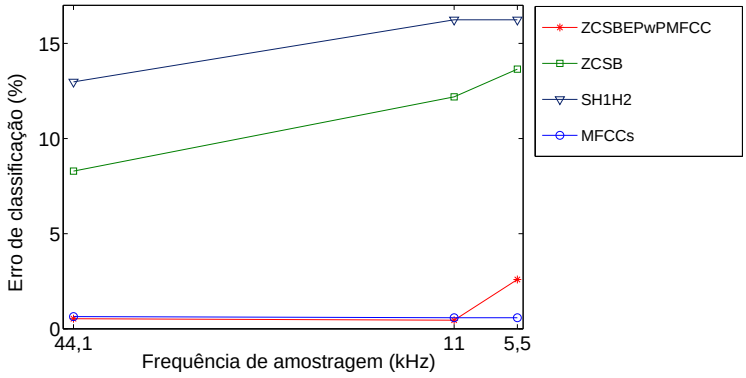


Figura 5.3. Correlação entre a diminuição de informação (f_s) e o erro de classificação.

Nesta figura é possível confirmar que o conjunto dos MFCCs é o menos afetado pela perda de informação devido ao efeito da quantificação. Além disso, o conjunto ZCSBEPwPitch-MFCCs, que combina os MFCCs sofre perda na classificação, demonstrando que as características ZCSBEPwPitch influenciam de forma negativa o classificador.

5.6.1 Correlação custo-beneficio

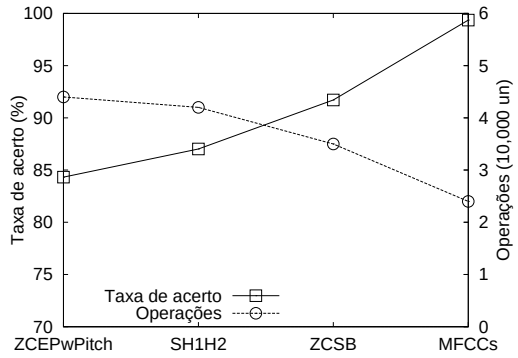
Na seção 2.5.9 foi realizada a análise matemática do custo de obtenção de cada característica. Foi mostrado que o custo dos MFCCs depende do número de pontos da FFT (N). No entanto, S e B dependem da quantidade de pontos N e da quantidade de amostras contidas em cada sílaba (L). Já as características temporais dependem exclusivamente das variações de L.

O custo dos MFCCs pode ser obtido como $Nlog_2(N) + N + mR$, lembrando que N é a quantidade de pontos da FFT, m a quantidade de coeficientes e R a quantidade de filtros. Considerando a obtenção de 12 coeficientes mediante a aplicação de 22 filtros e um espectro com 2048, o custo aproximado resulta em 24840 operações.

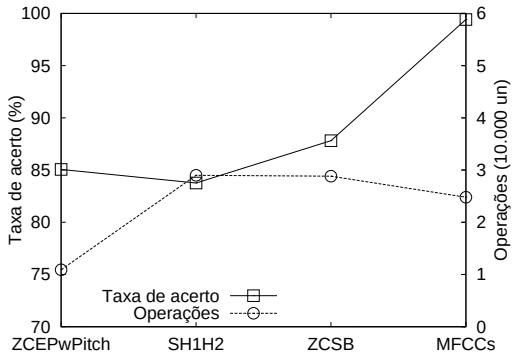
Para obter conjuntamente as características ZCSB o custo é $L + 2M + 2M + Nlog_2(N)$, sendo L o comprimento do sinal (200 mseg por sílaba). O comprimento L pode ser calculado como $L = fs * 0,200$, o que resulta em 8820 amostras para uma frequência igual a 44,1 kHz. O custo aproximado final resultante é 35444 operações. O custo de obtenção do conjunto SH_1H_2, considerando a mesma f_s, é calculado como: $2(L + i) + 2M + Nlog_2(N)$, resultando em um custo igual a 42236 operações. Para o conjunto de características temporais o custo é calculado como: $3L - 1 + L + L$ resultando 44099 operações.

Para comprovar a veracidade da análise analítica dos custos das características, realizamos testes de obtenção das características mensurando o tempo de obtenção de cada uma com a ferramenta de cálculo numérico MATLAB e um computador AMD Athlon 64X2 (figura 5.4).

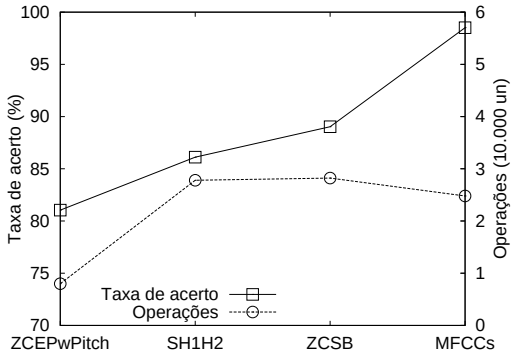
Para exemplificar melhor a relação entre o custo, a variação da quantidade de bits e a frequência de amostragem, montamos quatro gráficos, conforme figura 5.4. No eixo horizontal cada conjunto de características foi ordenado na mesma sequência da tabela 5.6. No eixo vertical esquerdo é representada a taxa de classificação para cada combinação e o eixo vertical direito representa custo, medido como o número aproximado de operações necessárias para obter cada combinação.



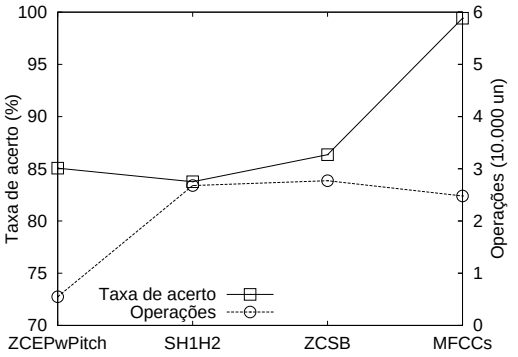
(a) 44,1kHz e 32bits.



(b) 11kHz e 8bits.



(c) 8kHz e 8bits.



(d) 5,5kHz e 8bits.

Figura 5.4. Relações custo-beneficio simulando cenários reais.

Estes quatro gráficos comparam as características dos métodos de Huang et al. [2009]; Han et al. [2011]; Vaca-Castaño & Rodriguez [2010] junto com as temporais. A melhor situação acontece quando o custo encontra-se sob a taxa de acerto. O caso ótimo resulta quando a diferença entre as curvas é a maior possível, maximizando a relação entre custo e acurácia (ou custo-beneficio). Nos quatro casos estudados, o conjunto de coeficientes Mel contribui com a maior taxa de classificação e custo constante,

comparado com ZCSB e SH_1H_2 que possuem um custo elevado e taxa de classificação significativamente menor que os MFCCs. O conjunto de características temporais (ZCEPwPitch) baseia-se na forma de onda do sinal, por isso o resultado da classificação não é muito afetado pelas variações de f_s, mas o custo diminui consideravelmente.

Das quatro figuras 5.4, pode-se observar que existe baixa correlação entre f_s e os MFCCs, tornando estes coeficientes independentes do hardware, o que constitui uma vantagem em situações práticas.

5.7 Considerações finais

A comparação dos conjuntos de características ZCSB e MFCCs, e avaliação da taxa de acerto do método utilizando quatro espécies foi publicado no Terceiro Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP) [Colonna et al., 2011]. Após esta publicação acrescentamos 5 novas espécies à nossa base e comparamos o método com um conjunto de características maiores (Pitch, Pw, E, H_1 e H_2) utilizados por outros autores.

Nesta primeira avaliação de resultados comparamos diversos conjuntos de características provenientes do tempo e da frequência usando a FFT. Depois de termos simulado quatro situações possíveis diminuindo f_s, como opção de redução de informação, concluímos que os MFCCs mantém a taxa de classificação e o custo independentemente do hardware. Além disso, os MFCCs mostraram-se ser mais imunes aos ruídos, por possuirem menor correlação com o valor de segmentação α e com o valor de quantização N_q.

Na próxima avaliação e comparação de métodos utilizamos estes coeficientes e os comparamos com características extraídas da transformada Wavelet (capítulo 6). Esta transformada possui caraterísticas de resolução em frequência e custo diferentes da transformada de Fourier e nos permite comparar com o método desenvolvido por Yen & Fu [2002]. No próximo capítulo abordamos também o problema de reconhecimento de varias espécies de anuros vocalizando ao mesmo tempo.

6

Comparação entre MFCC e Wavelet

Conforme apresentado no capitulo 5, os primeiros resultados foram obtidos comparando as características temporais com as espectrais. Esta comparação permitiu identificar qual dos métodos possui a melhor relação custo-beneficio, dentre os trabalhos de Huang et al. [2009]; Han et al. [2011]; Vaca-Castaño & Rodriguez [2010]. Esses experimentos indicaram que os MFCCs são a escolha adequada.

Nessa segunda série dos experimentos realizamos a comparação entre características extraídas da transformada Wavelet e os MFCCs provenientes da transformada de Fourier. Existem dois trabalhos que motivaram esta comparação, o trabalho de Rein & Reisslein [2011] que utiliza Wavelet em um contexto de RSSF, indicando como vantagem o custo menor comparada com a FFT, e por ter sido aplicada para identificar e classificar anuros no trabalho de Yen & Fu [2002].

Além das comparações de características, neste capítulo abordamos o problema de reconhecimento de grupos de espécies, ou "perfil de grupo". Para isto utilizamos os MFCCs e k-NN, combinando as nove espécies em grupos de dois e três anuros. O reconhecimento de grupo possibilita simular cenários reais nos quais poderiam estar presentes mais de uma espécie vocalizando no mesmo tempo (seção 6.4).

6.1 Metodologia utilizada para as comparações

A metodologia utilizada para conduzir esta segunda série de experimentos pode ser dividida em quatro etapas fundamentais:

1. extração das características, tanto baseadas na transformada Wavelet quanto os MFCCs, e geração das bases de dados para o classificador;

2. aplicação do algoritmo genético (GA) para selecionar os melhores subconjuntos de características para cada grupo, isto é, grupo baseado em Wavelet e MFCCs;

3. avaliação do impacto na classificação dos subconjuntos resultantes do GA; e

4. simulação de situações reais, avaliando o impacto da quantização e a frequência de amostragem na taxa de classificação.

No passo (1) é realizado o processamento prévio de forma semelhante ao descrito nos primeiros resultados, ou seja, aplicando uma etapa de segmentação, pré-ênfase e janelamento (seção 4.2.1). Já no passo (2), tentamos responder a seguinte pergunta: Quais são e quantos são os coeficientes necessários para obtermos um elevado desempenho de classificação correta? Desta forma, utilizamos GA para encontrar os subconjuntos ótimos. No passo (3), avaliamos o impacto desta diminuição e correlacionamos f_s com a relação custo-beneficio simulando situações reais (4).

Os conjuntos de características extraídas da transformada Wavelet, utilizadas nesta etapa, são a energia, a potência, a taxa se cruzamento por zero e a diferença entre os dois máximos da auto-correlação. Todas elas são extraídas dos coeficientes de detalhes e de escala. A este conjunto de caraterísticas foi acrescentado o valor do Pitch, com o objetivo de reforçar o resultado da classificação e para capturar informações sobre as frequências fundamentais das espécies. Na seção 4.2.2 foi descrito o método de obtenção de tais caraterísticas. Para obtermos os conjuntos aplicamos este método sobre duas funções Wavelet mãe: *Haar* e *Daubechies* (Db), de forma a comparar diferenças na taxa de classificação (seção 2.4.2).

6.2 Resultados com os conjuntos completos

Um passo anterior à seleção de subconjuntos de coeficientes é o ajuste dos parâmetros do método de classificação. Para tal fim, utilizamos k-NN. No capítulo anterior foi mostrado que k-NN não apresentou diferença significativa no resultado da classificação comparado com SVM. Antes de efetuarmos a classificação com k-NN, determinamos o valor de k que minimiza o erro de classificação para cada conjunto de características da mesma forma que foi descrita na seção 5.2.

Os resultados da classificação realizada com k-NN são apresentados na tabela 6.1. Esta tabela mostra a relação entre a taxa de classificação e a variável de amplitude da segmentação em sílabas α. A primeira coluna indica a combinação de características testadas e as demais colunas indicam a taxa de classificação correta. A combinação das características encontra-se em ordem decrescente segundo o custos aproximados de obtenção de cada combinação. O valor entre "()" indica o valor de k que minimiza o erro para as características dadas.

Características	**k-NN**		
	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$
Características Wavelet Transformada Daubechies	96,35%(3)	97,86%(1)	98,22%(1)
Características Wavelet Transformada Haar	96,70%(1)	97,90%(1)	98,38%(1)
MFCCs	99,19%(9)	99,36%(2)	99,19%(1)

Tabela 6.1. Taxa de classificação em relação a α, usando k-NN e validação cruzada $fold = 10$.

Na tabela de resultados (6.1) podemos notar que a função mãe *Haar* é melhor em taxa de classificação e em custo que a função *Daubechies*, e que os MFCCs superam ambos. As características baseadas na transformada Wavelet conseguem capturar informações similares às obtidas pelos MFCCs, mas com custo levemente maior. Pode-se notar também que a taxa de acerto com as características extraídas da transformada Wavelet aumenta levemente com a variação do limiar α, indicando correlação com os níveis de ruído do ambiente.

O teste *Wilcoxon* foi aplicado aos resultados mostrados na tabela 6.1. Foi usado nível de significância 95% sobre os dados obtidos no processo de validação cruzada (seção 2.7.3). Através deste teste estatístico, comparamos os 12 MFCCs de cada coluna com as classificações restantes e concluímos que, de fato, os coeficientes Mel possuem melhor desempenho.

A matriz de confusão 6.2 exemplifica quais são as espécies mais confundidas pelo classificador utilizando a transformada Wavelet *Haar* e $\alpha = 0.6$. Pode-se observar que as espécies *Rhinella granulosa* e *Scinax ruber* são as mais confundidas, por possuirem características sonoras que a transformada Wavelet não consegue diferenciar, não sendo assim com os MFFCs, que confundem mais as espécies *Rhinella granulosa* e *Leptodactylus fuscus*.

Espécie									k-NN
	a	b	c	d	e	f	g	h	i
a	433	3	0	0	0	1	2	0	0
b	3	258	0	0	0	0	2	2	3
c	2	3	202	0	0	0	6	0	10
d	0	0	0	247	0	2	0	0	0
e	1	1	1	0	226	1	1	2	0
f	1	0	0	1	0	80	0	2	0
g	1	1	1	0	0	0	1439	16	0
h	0	1	1	0	0	3	13	142	0
i	0	2	5	0	0	0	2	1	2888

Tabela 6.2. Matriz de confusão para k-NN usando *Haar*. Espécies: (**a**) *Adenomera andreae*, (**b**) *Ameerega trivittata*, (**c**) *Hyla minuta*, (**d**) *Hypsiboas cinerascens*, (**e**) *Leptodactylus fuscus*, (**f**) *Osteocephalus oophagus*, (**g**) *Rhinella granulosa*, (**h**) *Scinax ruber* e (**i**) *Hylaedactylus*.

Na seção seguinte são apresentados os resultados da otimização dos subconjuntos utilizando a abordagem genética para a seleção de características.

6.2.1 Seleção de características

A determinação da melhor combinação de características é fundamental para maximizarmos a taxa de classificação e minimizarmos os custos de processamento, através da remoção das características pouco relevantes. Como foi dito na seção 2.7.2, a otimização pode ser realizada avaliando o impacto na taxa de classificação de todas as possíveis combinações de caraterísticas. A abordagem combinatória, porém, possui custo computacional exponencial, fator que aumenta com a quantidade de características. Geralmente, a abordagem do problema de forma exaustiva é inviável em virtude do elevado custo de processamento. Além disso, características similares podem confundir o classificador, sendo desnecessário realizar a classificação para algumas combinações Deb [2001].

No contexto de nós sensores de baixo custo é interessante encontrarmos o conjunto mínimo de características necessárias para classificar as amostras. Em outras palavras, procuramos recombinar características para reduzir a dimensionalidade dos vetores de entrada, sendo nosso objetivo encontrar o subconjunto de características ótimas. Para encontrar esta combinação, utilizamos uma meta-heurística baseada em algoritmo genético (GA) Deb [2001]; Raymer et al. [2000]. Esta técnica foi aplicada sobre os dois conjuntos de características anteriores: os MFCCs e características provenientes da transformada Wavelet.

Para nosso algoritmo utilizamos primeiramente 20% da população para elitismo, 60% para cruzamento e 20% para mutação. Posteriormente, realizamos os experimentos com 50% da população para cruzamento, 40% para mutação e 10% para elitismo, com o objetivo de evitar convergência em máximos locais. Para as avaliações utilizamos o classificador k-NN, com validação cruzada ($fold = 10$). Foram rodadas 2000 gerações, sendo que a partir da ducentésima geração observamos convergência para a solução. O procedimento de otimização é exemplificado na figura 6.1.

A tabela 6.3 apresenta os conjuntos de características e os resultados das classificações obtidas antes e depois da busca genética, sobre a base com $\alpha = 0,5$. Na segunda coluna encontram-se os resultados obtidos com todos os coeficientes, isto é, antes da aplicação do GA. Na terceira e na quinta coluna estão os conjuntos de coeficientes encontrados durante a busca genética e na quarta e sexta colunas, os respectivos resultados de classificação.

Neste caso é interessante saber se existe empate entre as linhas da tabela 6.3. Pela

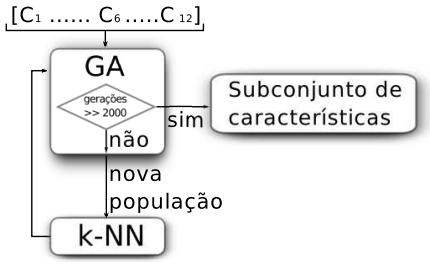


Figura 6.1. Processo de otimização genética.

Características	Classificação antes do GA	Cruzamento 50% Mutação 40%	Taxa de classificação	Cruzamento 60% Mutação 20%	Taxa de classificação
9 características utilizando Db	97,86%(1)	1,2,3,5	93,73%	1,2,3,4,5,6,8,9	96,83%
9 características utilizando *Haar*	97,90%(1)*	2,3,4,5,6,8,9	96,47%	1,2,3,4,5,6,7,8,9	97,90%*
12 MFCCs	99,36%(2)*	1,2,3,4,5,6,7,11	99,08%	1,2,3,4,5,6,7,8,9,11,12	99,33%*

Tabela 6.3. Resultados do algoritmo genético.

aplicação do teste de *Wilcoxon* podemos comparar os resultados do GA e identificar se as características eliminadas realmente não contribuem com informação relevante. Este teste conclui que somente no caso dos MFCCs, o decimo coeficiente pode ser dispensado do cálculo mantendo-se empate na classificação. Nos conjuntos restantes não existe evidência estatística suficiente para determinar empate, concluindo-se que todos os coeficientes contribuem com informação relevante.

Embora o teste estatístico não tenha determinado empate entre a segunda e quarta coluna, para nossa aplicação podemos considerar que a redução não afeta significativamente o resultado. Neste caso, o resultado do algoritmo genético diminuiu a quantidade necessária de coeficientes Mel em 4, passando de 12 para 8 características, quando foi utilizado 40% de probabilidade de mutação. Isto implicaria uma redução de informação que o nó sensor deverá transmitir de 33%.

O subconjunto de características usando a transformada *Haar* foi reduzido de 9 para 7, utilizando probabilidade de mutação 40%. Isto implica uma redução de 22% de informação que um nó deveria transmitir. Do vetor de características descrito na seção 4.2.2, foi removido a diferença entre os máximos da autocorrelação dos coeficientes de escala (W_P^a) e a potência dos coeficientes de detalhes (W_{Pw}^d). Após a remoção, a taxa classificação diminuiu 1,43%, e o teste estatístico não identificou empate.

No conjunto de características baseado na transformada *Daubechies* (Db) foram removidos 5 coeficientes, utilizando uma probabilidade de mutação de 40%, resultando em um novo vetor com: P, W_P^d, W_P^a, W_E^a. A taxa de classificação diminuiu 4,13%, com uma redução de informação final de 55%. Esta redução indica que as características

menos importantes são a potência e a taxa de cruzamento por zero dos dois conjuntos de coeficientes, não sendo assim com os valores da auto-correlação e o Pitch.

Podemos observar que na quinta coluna, utilizando somente 20% de probabilidade de mutação e os coeficientes *Haar*, a meta-heurística converge à solução que inclui todas as características. Note que, por ser uma abordagem aproximada, GA pode convergir em máximos locais, sem melhorar o desempenho na classificação.

Na próxima seção avaliamos o impacto da redução de informação, mediante a diminuição da frequência de amostragem (fs), sobre os conjuntos de coeficientes retornados pelo GA, e correlacionamos o desempenho com custo de processamento.

6.3 Estudo de caso

Da mesma forma que na seção 5.6, correlacionamos a quantização e a f_s com a taxa de acerto. Novamente, o objetivo desta abordagem é quantificar os custos e a dependência dos possíveis hardwares utilizados em situações reais. Em outras palavras, tentamos mensurar a relação custo-beneficio em tais situações. A tabela 6.4 mostra a relação entre a taxa de acerto obtida dos conjuntos de características, incluindo os encontrados com GA, a frequência de amostragem f_s e o número de bits de cada amostra dos áudios.

Características	**Classificação, k-NN**			
	32 bits 44,1kHz	8 bits 11kHz	8 bits 8kHz	8 bits 5,5kHz
9 Características Db	97,86%	92,71%	92,14%	88,27%
4 Características Db	93,73%	76,13%	84,20%	80,23%
9 Características Haar	97,90%	94,04%	91,92%	87,63%
7 Características Haar	96,47%	94,07%	91,60%	87,01%
12 MFCCs	99,36%	99,42%	98,51%	99,42%
11 MFCCs	99,33%	99,50%	98,53%	95,48%
8 MFCCs	99,08%	99,49%	98,66%	95,67%

Tabela 6.4. Comparação no resultado da classificação usando 32bits com 44,1kHz e 8bits com: 11kHz, 8kHz e 5,5kHz por amostra.

Nas duas últimas linhas da tabela 6.4 a diminuição de f_s, não afetou significativamente o desempenho dos 12 MFCCs, situação também observada nos resultados anteriores. Neste caso, o subconjunto com 8 MFCCs apresentou uma diminuição significativa só quando $f_s = 5,5kHz$, fato que indica que esse subconjunto não é tão independentes do hardware quanto os 12 MFCCs. Visando o objetivo de desenvolver um sistema genérico e abrangente, novamente os 12 MFCCs seriam uma escolha adequada.

As características baseadas nas transformadas *Haar* e *Daubechies* diminuem significativamente o desempenho em relação ao número de bits utilizados por amostra e à

frequência de amostragem. Além disso, por serem conjuntos de operações que incluem auto-correlação entre outras, o custo computacional é superior aos MFCCs.

O custo dos MFCCs pode ser obtido como $N\log(N)+N+mR$, lembrando que N é a quantidade de pontos da FFT, m a quantidade de coeficientes Mel e R a quantidade de filtros. Considerando $N = 2048$ pontos do espectro e 22 filtros, o custo aproximado de operações ao se utilizar 12, 11 ou 8 coeficientes resulta aproximadamente em 24840, 24818 ou 24752 operações, respectivamente.

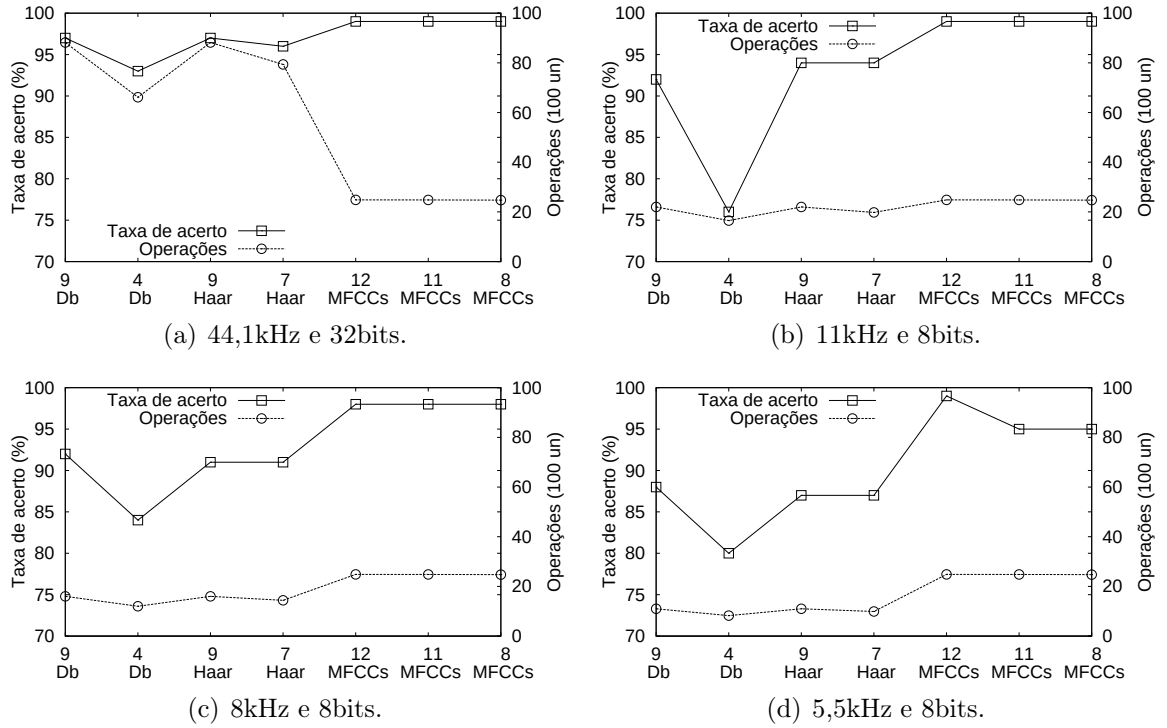
Na seção de fundamentação (2.4.2.3) foi explicado o custo da transformada Wavelet. Para obter o custo total dos conjuntos de características deve-se adicionar ao custo da transformada (L) o custo de cálculo de cada característica. Um exemplo de cálculo para a transformada *Haar* com os 9 coeficientes seria: $L+(3L-1)+2\frac{L}{2}+2\frac{L}{2}+2\frac{L}{2}+2(3\frac{L}{2}-1)$ em que (a) o primeiro termo corresponde à transformada; (b) o segundo ao Pitch; (c) o terceiro, quarto e quinto correspondem à energia, potência e taxa de cruzamento por zero dos coeficientes de escala e detalhes respectivos; e (d) o sexto termo corresponde com à diferença dos máximos da auto-correlação dos dois conjuntos de coeficientes. Finalmente resulta $10L-3$, para o caso dos conjuntos completos de coeficientes.

Para o cálculo de custo dos subconjuntos, deve-se eliminar da equação anterior os termos correspondentes às características não existentes, produto da otimização do GA. Para exemplificar as relações de custos e acurácia entre os diferentes conjuntos da tabela 6.4 montamos quatro gráficos correspondentes com as frequências utilizadas, conforme figura 6.2.

Como definimos na primeira parte dos resultados (5.6.1) consideramos a melhor combinação quando a diferença entre as curvas é a maior possível, maximizando a relação custo-beneficio. Nos quatro casos simulados os conjuntos de coeficientes Mel obtiveram desempenhos satisfatórios em relação ao custos de processamento. Observe-se que o custo a a classificação são mantidos aproximadamente constantes sem depender de f_s.

O custo das características baseadas na transformada Wavelet depende da quantidade de amostras do sinal. Na figura 6.2, observa-se a dependência com f_s. Em particular, quanto menor é a taxa de amostragem, menor é o custo. A taxa de classificação destas características é afetada pela diminuição de f_s, sendo que no pior cenário todas elas obtiveram resultados inferiores a 90%. Em outras palavras, a variabilidade da relação custo-beneficio é elevada em função de f_s.

Reforçando novamente os resultados obtidos no capítulo 5, podemos observar que os MFCCs maximizam a relação custo-beneficio. Na próxima seção, abordamos o problema de reconhecimento de grupos de espécies utilizando estes coeficientes.



(a) 44,1kHz e 32bits.

(b) 11kHz e 8bits.

(c) 8kHz e 8bits.

(d) 5,5kHz e 8bits.

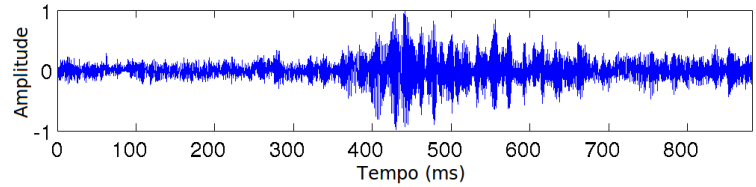
Figura 6.2. Relações custo-beneficio simulando cenários reais.

6.4 Reconhecimento de grupo

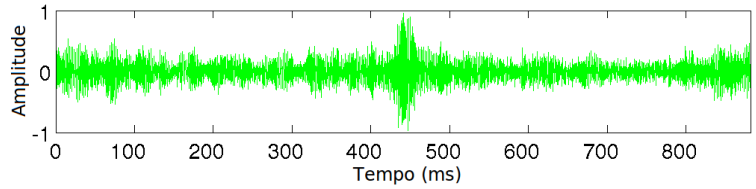
Nos resultados anteriores identificamos os MFCCs como característica mais adequada para classificar anuros. Este conjunto de coeficientes possui um ótimo desempenho tratando-se da identificação de espécies individualmente. Em uma situação real podem estar presentes mais de uma espécie em cada região vocalizando ao mesmo tempo. Quando isto acontece os sinais se misturam gerando novos padrões que chamamos "perfil de grupo".

Para simular estas situações combinamos e classificamos as sílabas de nossas bases de áudios. Desta forma, de cada combinação extraímos os MFCCs e os classificamos. Para isso formamos grupos de combinações de duas e três espécies. Os gráficos da figura 6.3 exemplificam o resultado do procedimento de combinação, utilizando a sílaba da espécie *Adenomera andreae* (figura 6.3(a)) e a sílaba da espécie *Ameerega trivittata* (figura 6.3(b)). Podemos observar que a combinação resulta em um novo padrão de forma de onda (figura 6.3(c)), o que muda o valor das características que a representam.

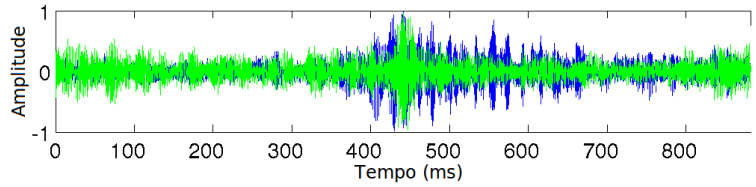
Utilizamos uma abordagem de amostragem estratificada para escolher as sílabas que foram combinadas (seção 2.7.4). Esta técnica possibilita reduzir os conjuntos de sílabas e diminuir o tempo de processamento da abordagem combinatória. Conside-



(a) Sílaba de *Adenomera andreae*.



(b) Sílaba de *Ameerega trivittata*.



(c) Novo padrão de onda.

Figura 6.3. Mistura das sílabas correspondentes às espécies *Adenomera andreae* e *Ameerega trivittata*.

rando um erro amostral tolerável de 5%, e aplicando a equação 2.33, podemos reduzir até 377 sílabas de um total de 7095. Na tabela 6.5 a primeira coluna é o total de sílabas de cada espécie, ou estrato, e a segunda coluna é quantidade de sílabas escolhidas após aplicação da amostragem estratificada, definindo-se as espécies como os estratos da população. Na obtenção do conjunto de sílabas combinadas em pares, cada sílaba de cada subconjunto de sub-amostras é combinada com todas as sílabas dos conjuntos restantes. Por exemplo, cada sílaba das 28 extraídas da espécie *Adenomera andreae* é combinada com todas as sílabas dos subconjuntos restantes, resultando em um base para o classificador com 51472 vetores de características. Um procedimento similar é aplicado para gerar o grupo de combinações de três espécies, resultando em uma base com 3416910 vetores de características.

Dado que originalmente possuímos nove especies, as combinações geradas resultam em 36 novas classes quando combinamos pares de espécies e 84 novas classes quando combinamos trios. O cálculo das classes é realizado mediante a aplicação da formula combinatória $C = \frac{N!}{K!(N-K)!}$, em que C representa o número possível de combinações, N o número total de espécies e K o valor de agrupação das espécies.

As amostras das 36 novas classes foram concatenadas com as 9 classes originais

Estratos	Total de sílabas	Amostragem
Adenomera andreae	528	28
Ameerega trivittata	572	31
Hyla minuta	261	14
Hypsiboas cinerascens	3176	169
Leptodactylus fuscus	252	13
Osteocephalus oophagus	103	5
Rhinella granulosa	1684	90
Scinax ruber	193	10
Hylaedactylus	326	17
Total	7095	377

Tabela 6.5. Numero de sílabas resultantes da amostragem estratificada.

para formar uma base de classificação com 45 classes. As amostras das 84 novas classes também foram concatenadas com as 9 classes originais formando uma segunda base de classificação com 93 classes.

Para identificar o perfil de grupo nas duas bases resultantes das combinações utilizamos k-NN com $k = 1$. O resultado da identificação entre uma e duas espécies foi de 77,74% e o resultado da identificação entre uma e três espécies foi de 22,57%. Destes resultados podemos concluir que a abordagem de perfil de grupo não é adequada quando trata-se de três espécies. No entanto, para duas espécies o resultado pode ser considerado bom dependendo dos requisitos e do contexto necessário de aplicação.

6.5 Considerações finais

Da comparação gráfica das figuras 6.2 podemos observar que não existe diferença significativa de custo entre os conjuntos com 12, 11 e 8 MFCCs. Comparando a taxa de acerto, podemos notar que no caso de serem implementados em um hardware altamente restritivo ($f_s = 5,5kHz$) seria melhor utilizar os 12 MFCCs. Devemos destacar que a otimização destes conjuntos ficou restringida as espécies utilizadas em nossas bases. Por este motivo, no desenvolvimento de um método geral indicamos como escolha adequada utilizar o conjunto completo dos MFCCs.

Os resultados das comparações entre os MFCC e Wavelet, assim como a otimização dos conjuntos de características, foram publicados no *International Joint Conference on Neural Networks* (IJCNN 2012) do *IEEE World Congress on Computational Intelligence* (IEEE WCCI 2012) [Colonna et al., 2012].

Da bordagem do problema de reconhecimento de grupo de espécies, como identificação de perfil de grupo, podemos concluir que tratando-se de duas espécies o método aplicado pode ser uma opção interessante, mais no caso de três não é recomendável. Este método de reconhecimento de duas espécies apresentou resultados de acerto melhor que o trabalho de Hu et al. [2005], no qual são classificadas as espécies indivi-

dualmente.

No próximo capítulo apresentamos todas as conclusões e contribuições obtidas ao longo desta pesquisa.

7

Conclusões

Neste trabalho abordamos o problema de classificação automática de anuros. O desafio de classificação automática é um subproblema, ou etapa, de uma abordagem maior de monitoramento ambiental mediante a utilização de RSSF. Por este motivo, conduzimos nossos experimentos para obtermos um método de classificação que maximize a relação entre o custo de operações, a quantidade de informação transmitida e a taxa de acerto, visando aumentar a vida útil dos nós sensores. As técnicas de monitoramento e classificação automáticas de animais possuem a vantagem de serem menos intrusivas comparadas às técnicas tradicionais, pelo fato de requererem menor intervenção humana, tornado-as menos sujeitas a erros de operação.

Em nossa abordagem aproveitamos a capacidade de vocalização dos anuros para obter características que facilitem sua identificação e classificação. O desenvolvimento desta pesquisa envolveu diversas etapas, desde a coleta dos áudios até a obtenção dos resultados, passando por um estudo minucioso do estado da arte e um desenvolvimento cuidadoso do método, comparando resultados e estratégias de reconhecimento de anuros.

Através da classificação de anuros, espera-se que possam ser inferidas informações sobre as populações que habitam uma determinada região, permitindo estabelecer uma relação com o meio ambiente.

7.1 Consideração finais

A representação dos sinais bioacústicos mediante um conjunto de características nos permitiu identificar padrões discriminantes dentro das formas de onda dos áudios. Além disso, este tipo de representação compacta possibilita reduzir a quantidade de informação necessária para realizar a classificação. Se considerarmos a utilização de um

hardware de amostragem com frequência igual a 44,1 kHz, a cada 200 ms são obtidos 8820 valores, mediante a representação com doze coeficientes, e a quantidade de informação transmitida é reduzida em 99,86%.

A procura do melhor conjunto de características nos levou a comparar quatro trabalhos existentes [Yen & Fu, 2002; Huang et al., 2009; Vaca-Castaño & Rodriguez, 2010; Han et al., 2011]. Desta comparação, assumimos que a melhor opção é utilizar os MFCCs por possuirem uma taxa de acerto superior a 99% utilizando k-NN. A partir da baixa relação obtida entre a taxa de classificação e as variáveis α e f_s, concluímos que este conjunto de características possui a maior imunidade aos ruídos, sejam estes ambientais ou de quantização. Além disso, estes coeficientes mantém o custo de processamento e a taxa de classificação aproximadamente constantes, independentemente da quantidade de bits utilizados para representar cada amostra, e da frequência de amostragem. Esta independência torna o método também independente do tipo de hardware.

A necessidade de redução dos custos de processamento nos motivou a otimizar o conjunto dos MFCCs mediante a aplicação de algoritmos genéticos. Neste caso, a plicação de GA nos permitiu identificar um subconjunto de oito coeficientes, com taxa de acerto superior a 98% em um intervalo de frequências de amostragem entre 44,1 kHz e 8 kHz. Esta otimização torna-se vantajosa em situações críticas, nas quais seja necessário diminuir o máximo possível os custos de processamento e transmissão. O conjunto de coeficientes obtidos da otimização mantem uma relação direta com as espécies utilizadas para treinar o classificador. Por este motivo, diminuir os conjunto dos MFCCs pode ocasionar uma perda de generalidade no método.

Após ter identificado o melhor conjunto de características e o pré-processamento necessário, conseguimos contribuir com um *framework* de classificação composto por três etapas fundamentais: (1) extração das sílabas, (2) aplicação da transformada de Fourier para obter os MFCCs e (3) classificação utilizando k-NN. Por fim, a estratégia de classificação definida torna o método possível de ser aplicado em situações práticas, presentes em diversos cenários. Logo, investigamos a possibilidade de classificação de grupos de espécies.

Além da definição da estratégia de classificação, ou *framework*, podemos destacar como contribuição adicional a metodologia utilizada para comparar os diferentes trabalhos existentes na literatura. Desta forma, podemos utilizar este trabalho como um guia prático possibilitando a inserção e comparação com métodos futuros de classificação.

A abordagem para identificar grupos de espécies, mediante a classificação de combinações de sílabas, mostrou-se mais apropriada para classificar duas espécies, resultando em uma taxa de acerto de 77%. No entanto, utilizando a combinação de três

espécies, a abordagem proposta não se mostrou satisfatória.

As contribuições obtidas no desenvolvimento desta pesquisa foram publicadas no *III Simpósio Brasileiro de Computação Ubíqua e Pervasiva* (SBCUP) [Colonna et al., 2011] e no *International Joint Conference on Neural Networks* (IJCNN) [Colonna et al., 2012].

7.2 Limitações do método

Abordar o problema de classificação como uma forma de redução de informação integrada com uma RSSF impossibilita a recuperação dos áudios no nó receptor. Desta forma, em uma situação real, torna-se difícil verificar se sons provenientes de outras espécies animais são confundidos pelo classificador. Esta dificuldade torna-se o principal entrave na busca por uma solução de monitoramento ambiental.

Além das limitações próprias do *framework* de redução de informação e classificação existem também restrições do hardware, tais como: reduzida capacidade de processamento ou impossibilidade de utilizar aritmética de ponto flutuante para realizar o cálculo da *FFT*. Tais limitações de implementação devem ser estudadas em detalhes utilizando os nós sensores disponíveis no mercado, no entanto, já existem implementações da *FFT* em outros tipos de hardware que utilizam aritmética de ponto fixo [Welch, 1969; Correa et al., 2012] .

7.3 Trabalhos futuros

O método de classificação desenvolvido começa pela etapa de segmentação dos áudios. Por este motivo, o resultado da classificação fica sujeito à extração correta das sílabas. Desta forma, uma possível extensão futura, para melhorar o resultado da classificação, seria melhorar o processo de segmentação. Esta melhoria ajudaria também em situações práticas, nas quais seja necessário identificar eventos esparsos em RSSF. Devido à possibilidade de integração com uma RSSF, como estratégia de monitoramento a longo prazo, um trabalho futuro seria avaliar a abrangência do *framework* de classificação, mediante o uso dos MFCCs e a utilização de áudios de outras espécies animais ou de um conjunto maior de anuros.

Futuramente, seria interessante abordar o problema de reconhecimento de grupo de espécies a partir de uma perspectiva diferente. Uma solução possível seria utilizar um classificador multi-nível para identificar se o áudio capturado pertence a um indivíduo ou a um conjunto de indivíduos, para posteriormente aplicar uma técnica

de separação de som. As técnicas de separação de áudio são frequentemente utilizadas em música com o objetivo de extrair o som de um instrumento específico [Wieczorkowska & Kolczyńska, 2008].

O método de classificação desenvolvido pode ser utilizado como métrica de avaliação de qualidade na recuperação de áudios comprimidos, fusão e agregação de dados em RSSF, detecção de eventos e classificação distribuída em RSSF e estimação das populações e a distribuição geográfica dos anuros. Atualmente, este método está sendo utilizado para mensurar a quantidade de compressão tolerável aplicando a técnica *compressive sensing*.

Referências Bibliográficas

Akyildiz, I. F.; Su, W.; Sankarasubramaniam, Y. & Cayirci, E. (2002). Wireless sensor networks: a survey. *Computer Networks*, 38(4):393--422.

Bee, M. A. & Micheyl, C. (2008). The cocktail party problem: what is it? how can it be solved? and why should animal behaviorists study it? *Journal of comparative psychology*, 122(3):235--51.

Bernarde, P. S. & Macedo, L. C. (2008). Impacto do desmatamento e formação de pastagens sobre a anurofauna de serapilheira em rondônia. *Iheringia. Série Zoologia*, 98(4):454--459.

Bulusu, N. & Hu, W. (2008). Lightweight acoustic classification for cane-toad monitoring. Em *42nd Asilomar Conference on Signals, Systems and Computers*, pp. 1601--1605. IEEE.

Cai, J.; Ee, D.; Pham, B.; Roe, P. & Zhang, J. (2007). Sensor network for the monitoring of ecosystem: Bird species recognition. Em *3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pp. 293--298. IEEE.

Cambron, M. E. & Bowker, R. G. (2006). An automated digital sound recording system: The amphibulator. Em *8th IEEE International Symposium on Multimedia (ISM).*, pp. 592--600.

Campbell, J. (1997). Speaker recognition: a tutorial. *Proceedings of the IEEE*, 85(9):1437--1462.

Carey, C.; Heyer, W. R.; Wilkinson, J.; Alford, R.; Arntzen, J. W.; Halliday, T.; Hungerford, L.; Lips, K. R.; Middleton, E. M.; Orchard, S. A. & Rand, A. S. (2001). Amphibian declines and environmental change: Use of remote-sensing data to identify environmental correlates. *Conservation Biology*, 15(4):903--913.

Cechin, S. Z. & Martins, M. (2000). Eficiência de armadilhas de queda (pitfall traps) em amostragens de anfíbios e répteis no brasil. *Revista Brasileira de Zoologia*, 17(3):729--740.

Chesmore, E. D. (2001). Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals. *Applied Acoustics*, 62(12):1359--1374.

Clemins, P. J. (2005). *Automatic classification of animal vocalizations*. PhD thesis, Marquette University ,Milwaukee, Wisconsin.

Collins, J. P. & Storfer, A. (2003). Global amphibian declines: sorting the hypotheses. *Diversity and Distributions*, 9:89--98.

Colonna, J. G.; dos Santos, E. M. & Nakamura, E. F. (2011). Classificação de anuros baseado em vocalizações para monitoramento ambiental pervasivo. *III Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP)*.

Colonna, J. G.; dos Santos, E. M. & Nakamura, E. F. (2012). Feature subset selection for automatic frog calls classiication applied to monitoring system based on sensor network. *International Joint Conference on Neural Networks (IJCNN)*.

Cooley, J. W. & Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):297--297.

Cormen, T. H.; Leiserson, C. E.; Rivest, R. L. & Stein, C. (2002). *Algoritmos - Teoria e Prática*.

Correa, I. S.; Freitas, L. C.; Klautau, A. & Albuquerque Costa, J. C. W. (2012). Vhdl implementation of a flexible and synthesizable fft processor. *Revista IEEE AmÃ©rica Latina*, 10(1).

Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification. *Transactions on Information Theory, IEEE*, 13(1):21--27.

Cowling, M. (2003). Comparison of techniques for environmental sound recognition. *Pattern Recognition Letters*, 24(15):2895--2907.

Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41(7):909--996.

Davis, S. & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4):357--366.

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, volume 16 edição.

Deller, J.; Hansen, J. & Proakis, J. (1993). *Discrete-time processing of speech signals*, volume 1. IEEE.

Deshpande, M. S. & Holambe, R. S. (2010). Speaker identification using admissible wavelet packet based decomposition. *World Academy of Science, Engineering and Technology*, 61:736--739.

Duhamel, P. & Vetterli, M. (1990). Fast fourier transforms: A tutorial review and a state of the art. *Signal Processing*, 19:259--299.

Fagerlund, S. (2007). Bird species recognition using support vector machines. *Journal on Advances in Signal Processing EURASIP*, 2007:1--9.

Frigo, M. & Johnson, S. G. (1998). Fftw: An adaptive software architecture for the fft. Em *International Conference on Acoustics, Speech and Signal Processing*, pp. 1381--1384. IEEE.

Gerhardt, H. C. (1975). Sound pressure levels and radiation patterns of the vocalizations of some north american frogs and toads. *Journal of Comparative Physiology ? A*, 102(1):1--12.

Graps, A. (1995). An introduction to wavelets. *IEEE Computational Science and Engineering*, 2(2):50--61.

Guimarães, M. (2004). *Educação Ambiental: No Consenso Um Embate?* Papirus.

Haar, A. (1910). Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331--371.

Haddad, C. (2005). Guia sonoro dos anfíbios anuros da mata atlântica.

Han, N. C.; Muniandy, S. V. & Dayou, J. (2011). Acoustic classification of australian anurans based on hybrid spectral-entropy approach. *Applied Acoustics*, 72(9):639--645.

Harma, A. (2003). Automatic identification of bird species based on sinusoidal modeling of syllables. Em *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 545--548. IEEE.

Hu, W.; Tran, V. N.; Bulusu, N.; Chou, C. T.; Jha, S. & Taylor, A. (2005). The design and evaluation of a hybrid sensor network for cane-toad monitoring. Em *4th International Symposium on Information Processing in Sensor Networks.*, pp. 503--508. IEEE.

Huang, C.; Yang, Y.; Yang, D. & Chen, Y. (2009). Frog classification using machine learning techniques. *Expert Systems with Applications*, 36(2):3737--3743.

IUCN (2011). The International Union for Conservation of Nature.

Jain, A. K.; Duin, R. P. W. & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE transactions on pattern analysis and machine intelligence*, 22(1):4--37.

Kekre, H. B.; Thepade, S. D.; Jain, J. & Agrawal, N. (2010). Iris recognition using texture features extracted from haarlet pyramid. *International Journal of Computer Applications*, 11(12):1--5.

Kogan, J. A. & Margoliash, D. (1998). Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden markov models: a comparative study. *The Journal of the Acoustical Society of America*, 103(4):2185--96.

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, 31:249--268.

Lee, C. H.; Lee, Y. K. & Huang, R. Z. (2006). Automatic recognition of bird songs using cepstral coefficients. *Journal of Information Technology and Applications*, 1(1):17--23.

Leite, D. S.; Helena, L. & Rino, M. (2006). A migração do supor para o weka: potencial e abordagens. Technical report.

Mainwaring, A.; Culler, D.; Polastre, J.; Szewczyk, R. & Anderson, J. (2002). Wireless sensor networks for habitat monitoring. *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications - WSNA '02*, p. 88.

Mallat, S. (1991). Zero-crossings of a wavelet transform. *IEEE Transactions on Information Theory*, 37(4):1019--1033.

Márquez, R.; Riva, I.; Matheu, B. & Matheu, E. (2002). Sounds of frogs and toads of bolivia.

Marsland, S. (2009). *Machine Learning: an algorithmic perspective.*, volume 1. CRC Press.

Martin, W. F. & Gans, C. (1972). Muscular control of the vocal tract during release signaling in the toad bufo valliceps. *Journal of morphology*, 137(1):1--27.

Marty, C.and Gaucher, P. (1999). Sound guide to the tailless amphibians of french guiana.

McKay, M. D.; Beckman, R. J. & Conover, W. J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2):239.

Mellinger, D. K. & Clark, C. W. (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. *The Journal of the Acoustical Society of America*, 107(6):3518 -- 3529.

Mittermeier, R.; Myers, N.; Thomsen, J. B.; Fonseca, G. A. B. & Olivieri, S. (1998). Biodiversity hotspots and major tropical wilderness areas: Approaches to setting conservation priorities. *Conservation Biology*, 12(3):516--520.

Modic, R.; Lindberg, B. & Petek, B. (2003). Comparative wavelet and mfcc speech recognition experiments on the slovenian and english speechdat2. Em *Proceedings of ISCA tutorial and research workshop on non-linear speech processing.*

Morettin, P. A. (1999). *Ondas e Ondaletas. Da Análise de Fourier à Análise de Ondaletas.* EdUSP, São Paulo.

Nelson, D. A. (1989). The importance of invariant and distinctive features in species recognition of bird song. *The Condor, Cooper Ornithological Society.*, 91(1):120--130.

Oppenheim, A. V.; Schafer, R. W. & Buck, J. R. (1999). *Discrete-time signal processing.* Prentice-Hall, Inc., 2nd edição.

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the cuidado project. Technical report, IRCAM.

Plack, C.; Oxenham, A. & Fay, R. (2005). *Pitch: Neural Coding and Perception.* Springer.

Pribilova, A. (2003). Preemphasis influence on harmonic speech model with autoregressive parameterization. *Radioengineering*, 12(3):32 -- 36.

Price, D. (1996). The toad that ate australia. *IEEE Expert*, 11(6):13--15.

Purgue, A. (1997). Sound radiation and post-glottal filtering in frogs: Vocal sac resonators revisited. *Acoustical Society of America in 134th Meeting Lay Language Papers.*

Quinlan, J. R. (1993). *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc.

Rabiner, L. & Schafer, R. (2007). Introduction to digital speech processing. *Foundations and Trends in Signal Processing*, 1:1--194.

Raymer, M.; Punch, W.; Goodman, E.; Kuhn, L. & Jain, A. (2000). Dimensionality reduction using genetic algorithms. *Evolutionary Computation, IEEE Transactions on*, 4(2):164 –171.

Rein, S. & Reisslein, M. (2011). Low-memory wavelet transforms for wireless sensor networks: A tutorial. *IEEE Communications Surveys & Tutorials*, 13(2):291--307.

Riede, K. (1993). Monitoring biodiversity: analysis of amazonian rainforest sounds. *Ambio*, 22(8):546--548.

Rioul, O. & Vetterli, M. (1991). Wavelets and signal processing. *IEEE Signal Processing Magazine*, 8(4):14--38.

Ross, S. (2003). *Introduction to Probability Models*. Academic Press.

Sarikaya, R.; Pellom, B. L. & Hansen, J. H. L. (1998). Wavelet packet transform features with application to speaker identification. Em *Proc. of IEEE Nordic Signal Processing Symp., Visgo*, pp. 81--84.

Schubert, E.; Wolfe, J. & Tarnopolsky, A. (2004). Spectral centroid and timbre in complex, multiple instrumental textures. Em *8th International Conference on Music Perception and Cognition*, pp. 654--657.

Selesnick, I. W. (2007). Wavelet transforms - a quick study. *Physics Today*, pp. 1--11.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379--423, 623--656.

Shukla, S.; Bulusu, N. & Jha, S. (2004). Cane-toad monitoring in kakadu national park using wireless sensor networks. Em *Proceedings of the Asia-Pacific Advanced Network (APAN)*.

Smith, S. W. (1999). Audio processing. Em *The Scientist and Engineer's Guide to Digital Signal Processing*, chapter 22 - Audio, pp. 351--372.

Stuart, S. N.; Chanson, J. S.; Cox, N. A.; Young, B. E.; Rodrigues, A. S. L.; Fischman, D. L. & Waller, R. W. (2004). Status and trends of amphibian declines and extinctions worldwide. *Science*, 306(5702):1783--1786.

Tan, B.; Spray, A. & Dermody, P. (1996). The use of wavelet transforms in phoneme recognition. Em *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, volume 4, pp. 2431--2434. IEEE.

Taylor, A.; Watson, G.; Grigg, G. & Hamish, M. (1996). Monitoring frog communities: an application of machine learning. Em *Proceedings of the eighth annual conference on Innovative applications of artificial intelligence*, number August, pp. 1564 -- 1569, Portland, Oregon. AAAI Press.

Theodoridis, S. & Koutroumbas, K. (2006). *Pattern recognition*. Elsevier/Academic Press.

Theodoridis, S.; Pikrakis, A.; Koutroumbas, K. & Cavouras, D. (2010). *Introduction to Pattern Recognition: a Matlab approach*. Elsevier Academic Press.

Trifa, V. M.; Kirschel, A. N. G.; Taylor, C. E. & Vallejo, E. E. (2008). Automated species recognition of antbirds in a mexican rainforest using hidden markov models. *The Journal of the Acoustical Society of America*, 123(4):2424--31.

Vaca-Castaño, G. & Rodriguez, D. (2010). Using syllabic mel cepstrum features and k-nearest neighbors to identify anurans and birds species. Em *Signal Processing Systems (SIPS), 2010 IEEE Workshop on*, pp. 466--471.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Statistics for engineering and information science. Springer.

Vapnik, V. N.; Boser, B. E. & Guyon, I. M. (1992). A training algorithm for optimal margin classifiers. Em *5th annual workshop on Computational learning theory.*, pp. 144--152. ACM.

Vié, J.; Hilton-Taylor, C. & Stuart, S. (2009). *Wildlife in a Changing World: An Analysis of the 2008 IUCN Red List of Threatened Species*. World Conservation Union.

Vilches, E.; Escobar, I. A.; Vallejo, E. E. & Taylor, C. E. (2006). Data mining applied to acoustic bird species recognition. Em *18th International Conference on Pattern Recognition, ICPR 2006.*

Welch, P. (1969). A fixed-point fast fourier transform error analysis. *IEEE Transactions on Audio and Electroacoustics*, 17(2).

Wieczorkowska, A. & Kolczyńska, E. (2008). Identification of dominating instrument in mixes of sounds of the same pitch. Em *Proceedings of the 17th international conference on Foundations of intelligent systems*, ISMIS'08, pp. 455--464, Berlin, Heidelberg. Springer-Verlag.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin. International Biometric Society*, 1(6):80--83.

Williams, S. (2001). Multiple determinants of australian tropical frog biodiversity. *Biological Conservation*, 98(1):1--10.

Williams, S. E.; Bolitho, E. E. & Fox, S. (2003). Climate change in australian tropical rainforests: an impending environmental catastrophe. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1527):1887--1892.

Wu, J. & Lin, B. (2009). Speaker identification using discrete wavelet packet transform technique with irregular decomposition. *Expert Systems with Applications*, 36(2):3136--3143.

Xing, G.; Wang, X.; Zhang, Y.; Lu, C.; Pless, R. & Gill, C. (2005). Integrated coverage and connectivity configuration for energy conservation in sensor networks. *ACM Transactions on Sensor Networks*, 1(1):36--72.

Yen, G. & Fu, Q. (2002). Automatic frog call monitoring system: a machine learning approach. Em *Proceedings of SPIE*, volume 4739, pp. 188--199. SPIE.