

Abordagem para Extração Automática de Características para Reconhecimento de Aves Tinamidae Baseada em Análise Espectral

Paulo Francisco da CONCEIÇÃO
Escola de Engenharia Elétrica e de Computação
pfrancisco43@gmail.com

Paulo César Miranda MACHADO
Escola de Engenharia Elétrica e de Computação
pcmmachado@gmail.com

Palavras-chave: Periodograma, segmentação, reconhecimento automático, análise espectral.

1 Introdução

Tem havido recentemente grande interesse nas pesquisas sobre avaliação computacional dos sons de animais para diversos fins, como identificação comportamental e detecção de espécies (KWAN et al., 2004), (CORREIO, 2010), (SELIN; TURUNEN; TANTTU, 2007). Com relação às aves é possível a catalogação e até mesmo a descoberta de novas aves, tendo em vista que com análise da vocalização é possível identificar características únicas para cada espécie. Com a catalogação e monitoração de espécies é possível contribuir com dados estatísticos que podem auxiliar na preservação das mesmas.

A família Tinamidae é composta por 23 espécies, das quais mais de 90% vivem no Brasil em ambientes como o cerrado. São aves tipicamente terrestres de aspecto galináceo e apresentam cores miméticas no colorido da plumagem, em adaptação aos seus hábitos terrestres. Esta família de aves, além de sofrerem com a degradação do ambiente em que vivem, ainda sofrem com caçadores, pois a carne destas aves é considerada saborosa. Assim, por se tratarem de aves tipicamente brasileiras e produzirem som bem característicos, a vocalização destas aves foi escolhida para testar a metodologia proposta por este trabalho (SAVE... , 2010).

A metodologia proposta por este trabalho consiste na análise automática do espectrograma do sinal do canto da ave visando a identificação das frequências mínima, máxima e mais intensa e também a determinação da duração do canto da ave. Estas características servem como dados de entrada para um sistema de reconhecimento automático da espécie.

2 Materiais e Métodos

Primeiramente foi feita a coleta e escolha dos áudios a serem usados para testar a abordagem proposta. Estes sinais são oriundos de bancos de dados disponíveis na internet (WIKIAVES... , 2011). Em seguida foi definida a taxa de amostragem. Para isso, primeiramente foi verificada qual a faixa de frequência da vocalização dos tinamídeos em geral, constatando que produzem sons entre 300 Hz e 4000 Hz, aproximadamente. Portanto a taxa de amostragem adotada como padrão foi cerca de 9000 Hz, garantindo assim uma margem de segurança para a filtragem.

Também foi tratado de forma automática se no sinal contém interferência de 60 Hz, que ocorre quando o equipamento utilizado na aquisição do som está conectado à rede elétrica, que é de 60 Hz no caso do Brasil. Esta interferência causa distúrbios ao sinal, tanto em frequências bem próximas de 60 Hz como na maioria de seus harmônicos.

Com este tratamento inicial o sinal está pronto para ser analisado, de forma automática, no intuito de buscar as características de frequências mínima, máxima e mais intensa. Para isso o sinal foi janelado em instantes de tempo e, para cada segmento, foi calculada a transformada rápida de Fourier (FFT), definida em (OPPENHEIM; SCHAFER, 1999) como

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi\frac{k}{N}n}, k = 0, 1, \dots, N - 1. \quad (1)$$

Então, para cada janela da matriz obtida com cálculo de FFT, é calculada a PSD, que segundo (STOICA; MOSES, 2005), pode ser obtida de forma simplificada via periodograma, calculando

$$\hat{\phi}_p(w) = \frac{1}{N} \left| \sum_{t=1}^N y(t)e^{-iwt} \right|^2. \quad (2)$$

Após a geração da matriz contendo a PSD para cada segmento, é adotada a seguinte lógica para encontrar as frequências: primeiramente encontra-se o valor máximo da matriz e a linha em que este valor ocorre, em seguida, a partir deste ponto, é feita uma análise de cada linha acima e abaixo, calculando o valor máximo da linha, este valor tende a diminuir cada vez mais. Assim, quando o valor máximo encontrado na linha diminui cerca de 70% do valor máximo da matriz, a busca é encerrada e verificadas as frequências nos instantes de parada. Com esta metodologia é estimada a frequência mais intensa, onde a PSD máxima foi encontrado, a frequência mínima, no instante de parada abaixo do valor máximo e a frequência máxima, no instante de parada acima do valor máximo.

Com as frequências mínima e máxima detectadas é aplicado um filtro do tipo passa banda, deixando permanecer no sinal apenas componentes de frequência exis-

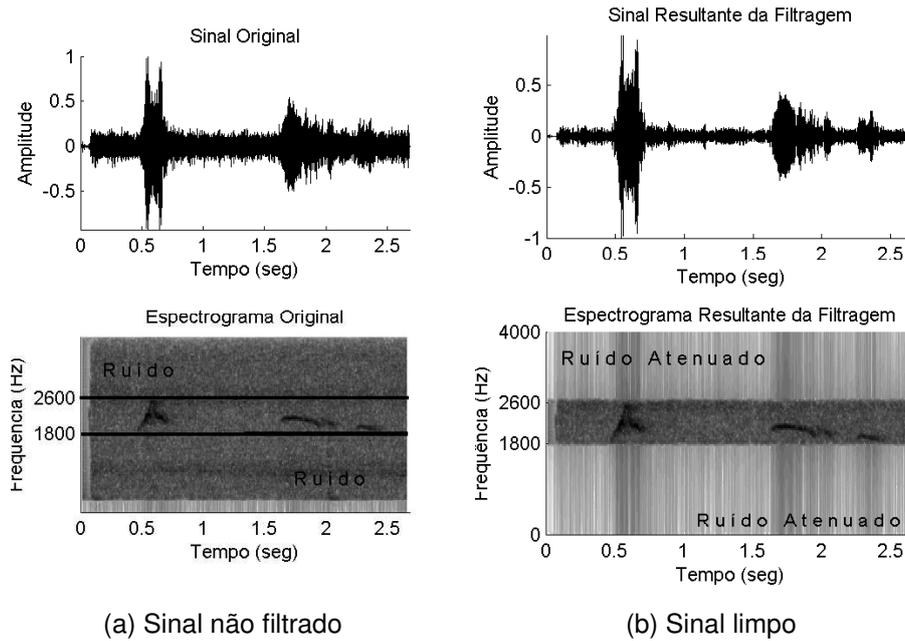


Figura 1: Sinal e espectro de uma Perdiz, evidenciando em (a) o nível de ruído do sinal e a demarcação da análise visual da faixa de frequência fundamental do canto da ave e (b) - Sinal limpo com a aplicação do filtro *Chebyshev* Invertido.

tentes entre os valores mínimo e máximo detectados. O filtro escolhido para utilização neste trabalho foi o *Chebyshev* Invertido, por possuir banda de passagem plana e banda de transição curta, exigindo baixa ordem (INGLE; PROAKIS, 1997).

A figura 1 mostra o resultado da abordagem de detecção automática das frequências, evidenciando que a busca automática traz resultados bem satisfatórios em relação à análise visual do espectrograma.

Para identificar as regiões de som e calcular o tempo de duração de um canto é feita uma nova análise na matriz de PSD, em que é analisado cada segmento de tempo e verificado qual o seu valor máximo, criando um vetor com marcações das regiões de som e silêncio, definido como

$$y[i] = \begin{cases} 1, & \max(P(:, i)) \geq \text{limiar} \\ 0, & \max(P(:, i)) < \text{limiar} \end{cases} \quad (3)$$

em que 0 representa silêncio e 1 representa som, P é a matriz de PSD. Com o vetor y gerado, este é analisado para verificar se as regiões marcadas como silêncio e som estão de acordo com limites pré-definidos. Para este trabalho foi definido um tempo superior a 2 segundos de pausa para ser considerado silêncio e 0.5 segundo de som para ser considerado canto (KWAN et al., 2004).

Tabela 1: Resumo de dados coletados de algumas espécies.

Ave		FMín	FMáx	FInt	MédC
Perdiz	μ	1730,30	2988,65	2048,94	1,95
	σ	57,36	69,49	130,81	0,39
Inhambu	μ	1291,71	1561,23	1424,07	4,64
	σ	78,81	32,84	32,61	0,38
Jaó	μ	1063,74	1386,74	1216,36	1,53
	σ	35,51	29,83	66,87	0,15
Codorna	μ	2346,05	2701,34	2493,02	0,5
	σ	115,85	61,65	102,18	0,08
Tururim	μ	1701,12	2183,47	1931,6	1,47
	σ	111,97	126,10	112,50	0,16
Azulona	μ	1055,13	1309,22	1184,43	0,75
	σ	49,48	48,72	50,15	0,14
Macuco	μ	1198,32	1445,95	1317,62	1,75
	σ	76,59	88,15	89,93	0,20

3 Resultados e Discussão

Analisando a tabela 1 é possível avaliar os resultados da abordagem proposta por este trabalho. A tabela mostra dados coletados de algumas espécies de aves da família dos tinamídeos, mostrando a média de cada grupo de dados e o desvio padrão. Embora tenham sido realizados testes com a maioria das espécies dos tinamídeos, por razões de espaço não são mostrados todos os resultados. as seguintes abreviações são usadas na tabela: FMín (Frequência Mínima, em Hz), FMáx (Frequência Máxima, em Hz), FInt (Frequência mais forte, em Hz), MédC (Média do tempo das regiões de canto, em segundos), μ (Média do grupo) e σ (Desvio padrão do grupo).

Nota-se na tabela 1 que a detecção automática de frequências e regiões de som tem estimativas aceitáveis, com desvios padrões baixos em relação à grandeza dos valores. Com relação aos tempos de som, nota-se um desvio padrão menor que 0,4 segundos para todos os casos, o que também pode ser considerado um valor baixo. Portanto, é possível concluir que os valores detectados automaticamente são distintos para espécies diferentes e bem homogêneos entre indivíduos da mesma espécie, servindo assim como características que auxiliam na classificação automática da espécie.

4 Conclusões

Neste trabalho foi proposta uma metodologia para determinação de características de sinais sonoros de aves da família dos tinamídeos, tratando desde a padronização dos sinais de entrada e aplicação de filtros até a detecção automática da faixa

de frequência e regiões de sons, tomando como ponto de partida a análise Tempo-Frequência, calculada através do espectrograma do sinal.

O experimento foi realizado com sinais oriundos de diversas fontes, contendo gravações muito ruins, o que se tornou um agravante para a metodologia, em alguns casos, não funcionar como esperado. No entanto, a técnica mostrou um bom desempenho para a maioria dos sinais testados. A metodologia mostrou ter um melhor desempenho quando o sinal não tem um nível de ruído alto, mas mesmo nesses casos os resultados podem ser considerados satisfatórios.

Como continuidade deste trabalho pretende-se investigar técnicas de eliminação de ruído remanescente o que, conseqüentemente, melhorará o desempenho da metodologia proposta. Além disso, pretende-se dar continuidade à fase de determinação de características do sinal, agora com técnicas mais elaboradas e conceituadas, para que então possa ser concluído o sistema capaz de reconhecer automaticamente as aves pela vocalização.

Referências

CORREIO. Bioacústica identifica novas aves na amazônia. *Artigo Eletrônico Correio do Brasil*, 2010. Disponível em: <<http://correiodobrasil.com.br/bioacustica-identifica-novas-aves-na-amazonia/152369/>>. Acesso em: 28/02/2011.

INGLE, V. K.; PROAKIS, J. *Digital Signal Processing - Using MATLAB V.4*. Boston: PWS Publishing Company, 1997.

KWAN, C. et al. Bird classification algorithms: Theory and experimental results. *ICASSP2004*, v. 289-292, 2004.

OPPENHEIM, A. V.; SCHAFER, R. W. *Discrete-Time Signal Processing*. second. New Jersey: Prentice Hall, 1999. ISBN 0-13-754920-2.

SAVE, Sociedade para Conservação das Aves do Brasil - Save Brasil. 2010. Disponível em: <<http://www.savebrasil.org.br/?q=content/por-que-aves>>. Acesso em: 27/02/2010.

SELIN, A.; TURUNEN, J.; TANTTU, J. T. Wavelets in recognition of bird sounds. *EURASIP Journal on Applied Signal Processing*, n. 2007, p. 1–9, 2007.

STOICA, P.; MOSES, R. *Introduction to Spectral Analysis*. NJ: Prentice Hall, 2005.

WIKIAVES, a Enciclopédia das Aves do Brasil. 2011. Disponível em: <<http://www.wikiaves.com.br/>>. Acesso em: 08/04/2011.