

RECONHECIMENTO DE AVES DE NOMES ONOMATOPÉICOS UTILIZANDO ÁRVORE DE DECISÃO

Célio S. de Brito Junior¹, Paulo César M. Machado¹ e Rodrigo P. Lemos¹

¹Escola de Engenharia Elétrica e de Computação – Universidade Federal de Goiás UFG
Caixa Postal 131, 74001-970 - Goiânia – GO – Brasil

celiojunior01@gmail.com, pcmmachado@gmail.com, rodrigo.p.lemos@gmail.com

Abstract. *This work proposes a methodology for automatic determination of characteristics, classification and recognition of birds with onomatopoeic names. It is based on the analysis of the spectrograms of the sounds of these birds. Signal filtering and segmentation are made automatically. The recognition of the birds follows from a decision tree, built according to the more intense frequency and entropy of the signal.*

Resumo. *Este trabalho propõe uma metodologia de determinação automática de características, classificação e reconhecimento de aves de nomes onomatopéicos baseada na análise dos espectrogramas das vocalizações dessas aves. A filtragem e a segmentação dos sinais são feitas automaticamente. O reconhecimento das aves é feito a partir de uma árvore de decisão, construída com base na frequência mais intensa e na entropia do sinal.*

1. Introdução

O reconhecimento de padrões é definido como a área de pesquisa que tem por objetivo a classificação de objetos (padrões) em diversas categorias ou classes, com aplicações nas mais diversas áreas, tais como: reconhecimento de fala, de escrita e de faces, identificação de retina, análise de eletrocardiogramas, radiografias, tomografias, etc [Duda, Hart and Stork 2001], [Theodoridis and Koutrombas 2009].

Atualmente, uma aplicação que vem sendo objeto de estudos nessa área é a monitoração e o reconhecimento automático de aves a partir de suas vocalizações [Brandes 2008],[Kwan *et al.* 2004]. Com a crescente devastação de seus *habitats*, alguns animais e aves estão em processo de extinção. Hoje já é possível encontrar nas cidades animais e aves que só se encontravam nos campos, devido à perda de seu *habitat* natural. Com o intuito de preservação, um sistema de monitoramento e reconhecimento de aves pode ser utilizado para verificação das espécies existentes em uma região e para monitorar aquelas ameaçadas de extinção [Fagerlund 2004],[Fagerlund 2007].

Neste trabalho o foco são as aves conhecidas como onomatopéicas, ou seja, aves que tem o nome de acordo com suas vocalizações, como o Bem-te-vi, Fogo-apagou, Saracura-três-potes, Jaó e Tiziu, que vivem em regiões de mata rasteira e densa. Isto faz com que algumas sejam mais ouvidas do que vistas (Jaó e Saracura-três-potes), tornando difícil o seu estudo, ou são aves que tiveram que se adaptar ao grande avanço da urbanização (Bem-te-vi, Tiziu e Fogo-apagou) [Sick 1984]. Essas aves foram

escolhidas por se tratarem de aves típicas da fauna brasileira e produzirem sons bem característicos.

2. Metodologia

A metodologia proposta pode ser decomposta em fases sucessivas iniciando pela fase de pré-processamento, que consiste na normalização, filtragem, detecção de frequência e segmentação. Em seguida, escolhe-se a melhor parte segmentada do som e, por fim, realiza-se o treinamento do sistema de reconhecimento e o reconhecimento em si.

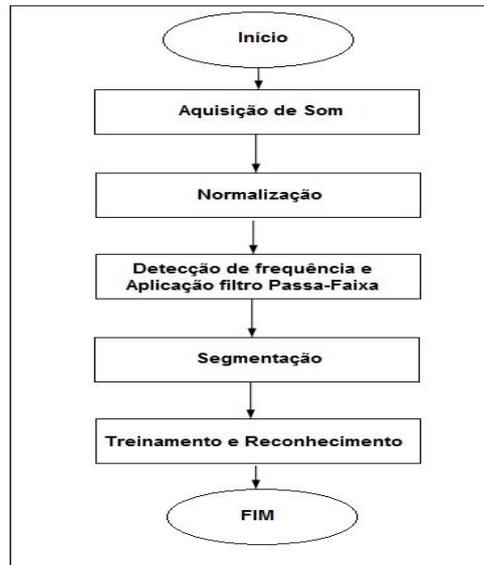


Figura 1. Diagrama de blocos do sistema desenvolvido, mostrando a sequência adotada para o reconhecimento de espécies de aves.

2.1. Normalização e Amostragem

Os sons utilizados neste trabalho foram coletados através de buscas em banco de dados disponíveis na internet [Xeno-Canto 2011], [WikiAves 2011] e, como foram gravados em condições diferentes, possuem diferentes níveis de intensidade. Portanto, o primeiro passo no tratamento desses sons é a extração de seu valor médio, conforme equação (1), e a sua normalização em relação ao seu valor máximo, conforme equação (2).

$$x(k) = x_{orig}(k) - média(x_{orig}) \quad (1)$$

$$x_{norm}(k) = \frac{x(k)}{\max_i(|x(i)|)} \quad (2)$$

onde $x(k)$ é o vetor que representa o canto do pássaro após a extração do valor médio, $x_{orig}(k)$ é o vetor que representa o canto original, ou seja, o que foi gravado, $média(x_{orig})$ é a média do vetor $x_{orig}(k)$, $x_{norm}(k)$ é o vetor x normalizado e $\max_i(|x(i)|)$ é o valor máximo do vetor $x(k)$. Dessa maneira todos os sinais terão valor médio zero e

todos os valores dos sinais estarão na mesma faixa de amplitude, entre -1 e 1 [Ricke, Povinelli and Johnson 2005].

Em seguida é definida a taxa de amostragem. Sabe-se que a intensidade máxima da frequência de vocalização das aves é em torno de 12000 Hz, portanto escolheu-se a taxa de amostragem de 25000 Hz para garantir uma margem de segurança para a filtragem.

2.2. Redução do Ruído

Para reduzir os ruídos provocados pelo ambiente, tais como o vento, a chuva ou outros animais presente no sinal, promoveu-se a filtragem do sinal com um filtro do tipo passa faixa. O filtro escolhido para utilização neste trabalho foi o Chebyshev Invertido, por possuir faixa de passagem plana, permitindo uma melhor captura do sinal desejado, e ter faixa de transição curta, exigindo menor ordem para o filtro [Rabiner and Gold 1975].

A frequência mínima e máxima de cada sinal, que são usadas como parâmetros do filtro, são detectadas automaticamente antes da aplicação do filtro. O método, descrito em [Conceição, Machado e Lemos 2011], utiliza uma varredura na matriz de PSD (Densidade Espectral de Potência) a partir do cálculo do periodograma [Stoica and Moses 2005]. Além da frequência mínima e máxima, também é calculada automaticamente a frequência mais intensa de cada sinal, que será armazenada como característica da ave.

2.3. Processo de segmentação

Neste trabalho foi utilizado um processo de segmentação, baseado na energia e no centróide do sinal, com o objetivo de separar as regiões de som das regiões de silêncio e, assim, coletar o tempo de duração do canto da ave.

Nesse processo, primeiramente o sinal é dividido em janelas de 50 ms e, para cada uma delas, a energia do sinal $E(i)$ é calculada pela equação (3):

$$E(i) = 1/N \sum_{n=1}^N |x_i(n)|^2 \quad (3)$$

onde $x[n]$ é o sinal e N é a dimensão do vetor $x[n]$.

Nos casos em que a energia do sinal é maior do que a do ruído este procedimento é bem sucedido, entretanto, quando a energia do ruído de fundo supera a do sinal, não se consegue boa segmentação. Para melhorar o desempenho da segmentação, calculam-se também os centróides espectrais (centro de gravidade do espectro do áudio) ao longo do sinal [Saunders 1996]. Este recurso é utilizado para trabalhar com características geométricas do espectro [Giannakopoulos 2009]. Seu cálculo é feito pela equação (4).

$$C(i) = \frac{\sum_{k=1}^N (k+1)X_i(k)}{\sum_{k=1}^N X_i(k)} \quad (4)$$

Onde $C(i)$ é o centróide da janela i , N é a dimensão do vetor $x[n]$ e $X_i(k)$ é a Transformada Discreta de Fourier (FFT), para cada segmento de tempo, definida em [Oppenheim and Schaffer (1999)].

2.4. Treinamento e Reconhecimento

Após o pré-processamento vêm a fases de treinamento e reconhecimento. No pré-processamento é encontrado o valor da frequência mais intensa e essa característica, juntamente com a entropia do sinal, são os parâmetros para reconhecimento que serão utilizados nesse trabalho. Essas duas características foram escolhidas pelo fato de ambas possuírem valores bem distintos entre as aves estudadas, diferentemente de outras características que foram consideradas, tais como a energia, o centróide e o fluxo espectral do sinal [Giannakopoulos 2009].

A entropia é um recurso a ser utilizado se o sinal de áudio apresentar mudanças bruscas no nível de energia. Inicialmente, as janelas i de tamanho N são subdivididas em sub-janelas j de tamanho K . Para cada sub-janela j , a energia e_j^2 é normalizada de acordo com a equação (5) [Chou, Lee and Wni 2007]:

$$e_j^2 = E_{sub-janela_j} / E_{janela_i} \quad (5)$$

Em seguida, a entropia da janela i é calculada utilizando-se a equação (6) [Petridis, Giannakopoulos and Perantonis 2010]:

$$Z(i) = -\sum_{j=1}^k e_j^2 \log_2(e_j^2) \quad (6)$$

Na fase de treinamento um determinado número de aves de cada espécie em estudo é escolhido. A frequência mais intensa e a entropia dessas aves são utilizadas para construir uma árvore de decisão [Li *et al.* 2001]. Na fase de reconhecimento a frequência mais intensa e a entropia das aves a serem reconhecidas são comparadas com os valores da árvore de decisão, obtendo-se assim a espécie mais provável a que pertence cada ave.

3. Resultados

A Fig. 2 mostra a energia normalizada, o centróide normalizado e o sinal após a filtragem. Pode-se observar para esse sinal que tanto a energia como o centróide são muito maiores na região do canto do que na de silêncio, o que torna possível realizar a segmentação do sinal separando essas duas regiões.

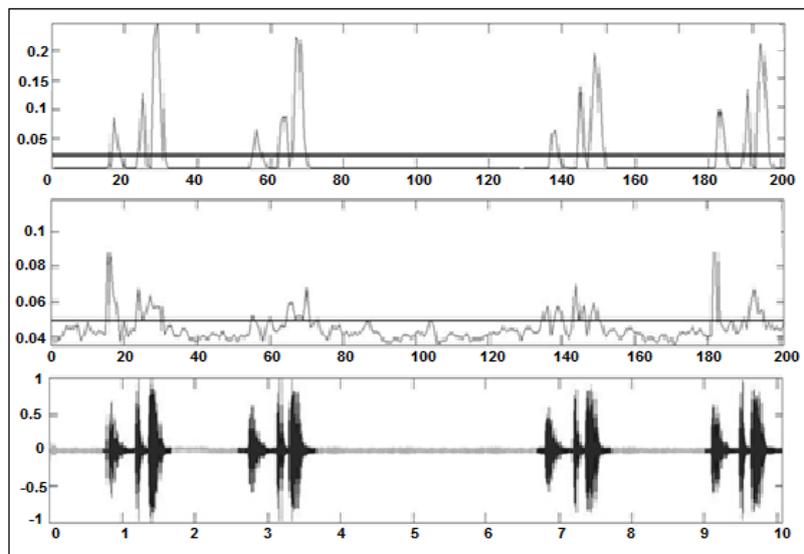


Figura 2. Energia do sinal, centróide e sinal filtrado.

A Fig. 3 mostra o sinal filtrado e os sinais resultantes da segmentação. Observa-se que o método utilizado foi eficiente em encontrar as quatro regiões em que há a vocalização da ave.

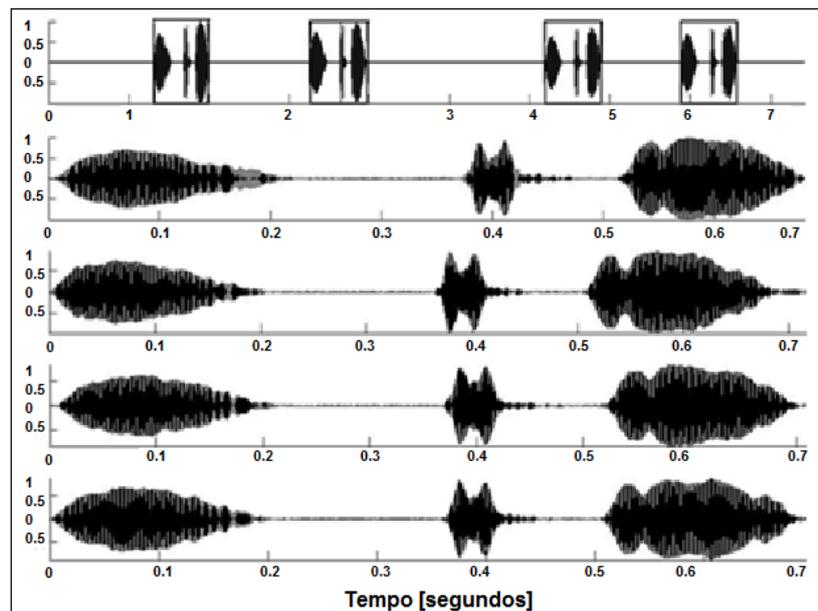


Figura 3. Regiões de som e silêncio detectadas automaticamente, mostrando quatro vocalizações da ave.

Após a segmentação, na fase de treinamento, foram calculadas a frequência mais intensa e a entropia dos sinais dos sons de 10 aves de cada espécie. As Figuras 4 e 5 mostram, respectivamente, as regiões de entropia e as regiões de frequências dos sons das espécies em estudo.

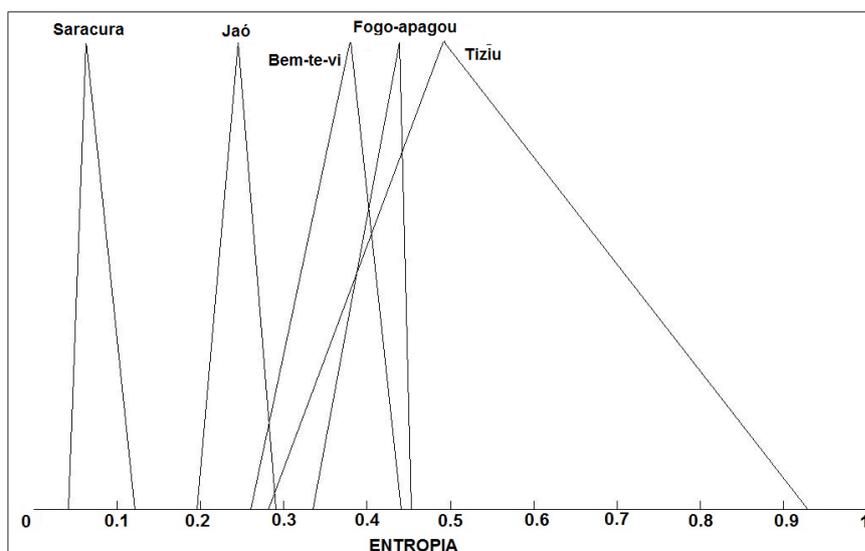


Figura 4. Regiões das entropias para as aves analisadas. As bases representam, respectivamente, as entropias mínimas e máximas e os picos representam os valores modais dos histogramas de cada espécie.

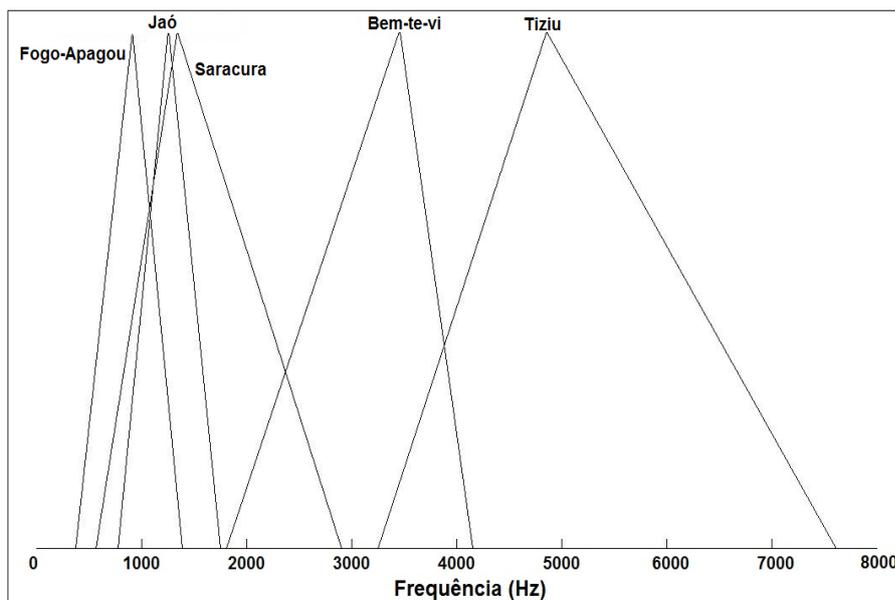


Figura 5. Regiões das frequências para as aves analisadas. As bases representam, respectivamente, as frequências mínimas e máximas e os picos representam as frequências mais intensas para cada ave.

Observa-se nas Figs. 4 e 5 que esses valores são bem homogêneos e distintos entre os grupos a serem reconhecidos, justificando assim a escolha dessas grandezas como características a serem utilizadas no reconhecimento automático dessas aves.

A Fig. 6 mostra a estrutura da árvore de decisão utilizada no reconhecimento, construída utilizando como parâmetros a frequência mais intensa e a entropia do sinal.

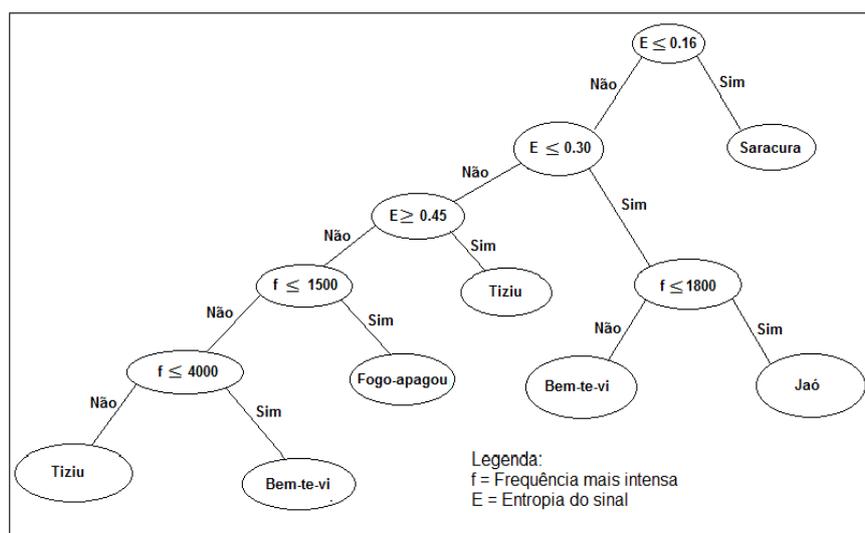


Figura 6. Estrutura da árvore de decisão utilizada na criação do programa de reconhecimento.

Na fase de reconhecimento foram utilizados os sons de 16 Saracuras, 17 Jaós, 16 Tizius, 74 Bem-te-vis e 42 Fogo-apagou. Os resultados do reconhecimento utilizando a árvore de decisão construída são mostrados na Tabela 1.

Pode se observar na tabela que o reconhecimento foi de 100% para as espécies Saracura, Jaó e Fogo-apagou, enquanto a menor taxa de reconhecimento foi de 87,5 % para o Tiziu. Além disso, 95% do total de aves foi reconhecido corretamente, enquanto o sistema confundiu alguns indivíduos das espécies Bem-te-vi com Tiziu e Fogo-apagou.

Tabela 1. Resultados obtidos na fase de reconhecimento.

	Saracura	Jaó	Tiziu	Bem-te-vi	Fogo – apagou
Saracura	16				
Jaó		17			
Tiziu			14	2	
Bem-te-vi			4	68	2
Fogo – apagou					42

4. Conclusão

Neste trabalho foi proposta uma metodologia para a realização do reconhecimento de aves de nomes onomatopéicos a partir de suas vocalizações. Foi utilizado um sistema de detecção automática da faixa de frequência para efeito de filtragem automática dos sinais, a segmentação do sinal em regiões de som e silêncio a partir da energia e centróide do sinal e, finalmente, o reconhecimento utilizando uma árvore de decisão construída a partir da entropia do sinal e da frequência mais intensa do canto.

O experimento foi desenvolvido utilizando como amostra sinais oriundos de diversas fontes, disponíveis em alguns sites especializados em aves, contendo gravações muito ruins. No entanto as técnicas apresentadas mostraram um bom desempenho para a maioria dos sinais testados, como mostram os resultados. As características utilizadas para o reconhecimento (entropia e frequência de maior intensidade) se mostraram satisfatórias, já que resultou em 95 % das aves reconhecidas corretamente.

Como continuidade deste trabalho pretende-se utilizar técnicas como redes neurais e cadeias ocultas de Markov, para construir um sistema de reconhecimento automático de qualquer ave a partir da vocalização.

Referências

- Brandes, T. S. (2008) “Automated Sound Recording and Analysis Techniques for Birds Surveys and Conservation”, In: Bird Conservation International p.163-173.
- Chou, C. H., Lee, C. H. and Wni, H. (2007) “Bird Species Recognition by Comparing the HMMs of the Syllables”, IEEE, p. 1-4.
- Conceição, P. F., Machado P. C. M. e Lemos, R. P. (2011) “Sistema Automático para Segmentação de Características de Sinais Sonoros de Aves do Cerrado Brasileiro a Partir da Análise Tempo Frequência” In: XXIX simpósio brasileiro de telecomunicações - SBrT, Curitiba, PR.
- Duda, R. O., Hart, P. E. and Stork, D. G. (2001) ”Pattern Classification”, 2nd edition, New York, Wiley.
- Fagerlund, S. (2004) “Automatic Recognition of Bird Species by Their Sounds”, Master Thesis, Dept. of Electrical and Communications Engineering, Helsinki University of Technology.
- Fagerlund, S. (2007) “Bird Species Recognition Using Support Vector Machines”, EURASIP Journal on Advances in Signal Processing, p. 1-11.
- Giannakopoulos, T. (2009) “Study and Application of Acoustic information for the detection of harmful content, and fusion with visual information”, Ph.D. dissertation, Dept. of Informatics and Telecommunications, University of Athens, Greece.
- Kwan, C., Mei, G., Zhao, X., Ren, Z., Xu, R., Stanford, V., Rochet, C., Aube, J. and Ho, K.C. (2004) “Bird Classification Algorithms: Theory and Experimental Results”, ICASSP2004, p. V289-V292.
- Li, D., Sethi, I.K., Dimitrova, N., McGee T. (2001) “Classification of general audio data for content-based retrieval” In: Pattern recognition Letters 22, pp533-544.
- Oppenheim, A. V. and Schaffer, R. W. (1999) “Discrete-Time Signal Processing” – 2nd edition, New Jersey, Prentice Hall.
- Petridis, S., Giannakopoulos, T. and Perantonis, S. (2010) “A Multi-Class Method for Detecting Audio Events in News Broadcasts”, Lecture Notes in Computer Science, 6040, p. 399-404.
- Rabiner, L.R., and Gold, B. (1975) “Theory and Application of Digital Signal Processing”, Englewood Cliffs, Prentice-Hall.

-
- Ricke, A. D., Povinelli, R. J. and Johnson, M. T. (2005) “Automatic Segmentation of Heart Sound Signals Using Hidden Markov Models”, *Computers in Cardiology*, p. 953–956.
- Saunders, J. (1996) “Real-Time Discrimination of Broadcast Speech/Music”, *Proceedings of the Acoustics, Speech, and Signal Processing* (ICASSP96), p. 993-996.
- Sick, H. (1984) “Ornitologia Brasileira - Uma Introdução”. Brasília, Ed. Universidade de Brasília.
- Stoica, P. and Moses, R. (2005) “Introduction to Spectral Analysis”, New Jersey, Prentice Hall.
- Theodoridis, S. and Koutrombas, K. (2009) “Pattern Recognition”, 4th edition, San Diego, Elsevier.
- “WikiAves”, <http://www.wikiaves.com.br>. Acesso em 07/06/2011.
- “Xeno-Canto”, <http://www.xeno-canto.org/>. Acesso em 07/06/2011.

