

# Sistema Automático para Segmentação e Determinação de Características de Sinais Sonoros de Aves do Cerrado Brasileiro a Partir da Análise Tempo-Frequência

Paulo Francisco da Conceição, Paulo César Miranda Machado, Rodrigo Pinto Lemos

**Resumo**—Este trabalho propõe uma metodologia de determinação automática de característica de classificação de aves Tinamidae do cerrado brasileiro a partir da análise do espectrograma do canto dessas aves. A partir da identificação da frequência mais forte e de seu intervalo de variação pode-se ajustar automaticamente filtros digitais para redução de ruído no sinal, estimar as faixas de frequência de vocalização e efetuar a segmentação do sinal em intervalos de canto e de silêncio. As médias e desvios padrões das características extraídas mostraram-se distintas para as diversas espécies estudadas, estabelecendo parâmetros para uma etapa posterior de classificação e reconhecimento dessas aves.

**Palavras-Chave**—Espectrograma, classificação automática, filtro digital, segmentação, vocalização de aves.

**Abstract**—This paper proposes a methodology for automatic feature determination of birds of the Tinamidae family from the Brazilian Cerrado environment using spectrogram analysis of the bird sounds. This methodology proposes the identification of the stronger frequency and tracking its variation through time in order to estimate the frequency bands of vocalization. This allows automatic setting of digital filters for noise reduction and the segmentation of the sounds in intervals of singing and silence. Means and standard deviations of the extracted features showed to be different among the species, establishing parameters for bird recognition and classification in the future.

**Keywords**—Spectrogram, automatic classification, digital filters, segmentation, bird songs.

## I. 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 [1-8]. 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 [9].

A avifauna do cerrado, especificamente, merece grande atenção neste sentido, pois tende a diminuir rapidamente, uma vez que seus ambientes naturais estão cada vez mais fragmentados e são destruídos de forma rápida, sendo substituídos pela

agricultura e pecuária. Ainda assim, o cerrado ocupa 25% do território brasileiro, entretanto, apenas 2,2% da área do cerrado é legalmente protegida, o que contribui mais ainda para a destruição deste ambiente. O cerrado é um dos biomas mais ricos da avifauna brasileira, com cerca de 837 espécies, das quais 30 são endêmicas e 14 estão ameaçadas [10].

A família dos tinamídeos (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 tratar de aves tipicamente brasileiras e produzirem som bem característicos, a vocalização destas aves foi escolhida para testar a metodologia proposta por este trabalho [11]. Portanto, todas as aves citadas nos exemplos e testes são da família tinamidae.

Um sistema completo de reconhecimento e classificação de aves, de forma automática, está sendo desenvolvido em consequência das pesquisas relacionadas à vocalização de aves da família dos tinamídeos e técnicas de Processamento Digital de Sinais. Entretanto, este trabalho aqui apresentado visa mostrar os resultados parciais do desenvolvimento deste sistema, destacando o processo automático adotado para detecção da faixa de frequência do canto da ave e segmentação das regiões de som e silêncio através da análise Tempo-Frequência.

## II. PRÉ-PROCESSAMENTO

A figura 1 traz o diagrama do sistema desenvolvido até então, mostrando a sequência adotada para determinar as características de frequência máxima, frequência mínima, frequência mais forte e segmentação das regiões de som e silêncio.

Quando o sinal de áudio é coletado, vários fatores fazem com que ele seja corrompido, dentre os quais: interferência da rede elétrica, ruído, barulhos e sons do ambiente de coleta. A etapa de pré-processamento visa basicamente tratar o sinal e prepará-lo para determinação de características, buscando eliminar os possíveis distúrbios presentes no sinal [12,13].

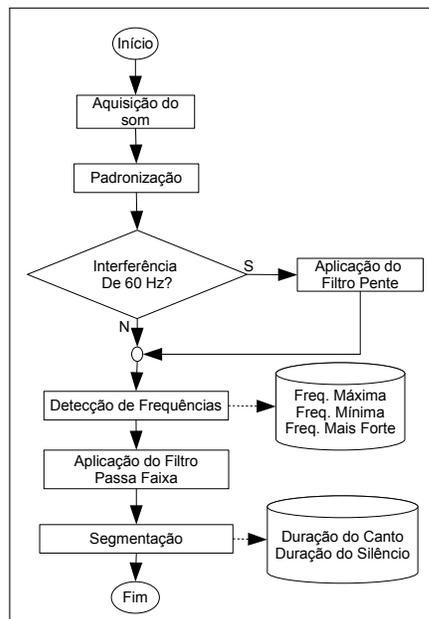


Fig. 1. Diagrama de blocos do sistema desenvolvido, mostrando a sequência adotada para a determinação de características.

### A. Padronização dos Sinais

A fase de coleta do áudio não foi realizada diretamente no *habitat* da ave, mas sim através de buscas em bancos de dados disponíveis na internet [14,15], assim foi possível dar maior atenção nos aspectos diretamente ligados a Processamento Digital de Sinais. Tendo em vista que os sinais analisados são oriundos de várias fontes de gravações diferentes, com taxas de amostragem distintas e equipamentos também diversos, se faz necessário padronizar os sinais adquiridos.

Primeiramente os sinais de áudio foram normalizados em relação ao seu valor máximo de tal forma que ocupem a mesma faixa de amplitude, entre -1 e 1 [5,6]. Também o valor médio do sinal foi calculado e extraído do sinal para facilitar a análise do espectrograma.

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.

### B. Detectando e Eliminando Interferência de 60 Hz

A interferência de 60 Hz 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.

Tendo em vista que esta interferência pode coexistir juntamente com a faixa de frequência do sinal, é necessário aplicar um filtro pente, que pode ser projetado para atuar de tal forma que elimine/atenuate as frequências em torno de 60 Hz e todos os seus harmônicos. Para avaliar a necessidade da aplicação

desta filtragem, é verificada qual a componente de frequência de maior força/intensidade presente no espectro e, quando esta ocorre em torno de 60 Hz, é aplicado o filtro pente.

A figura 2 mostra o sinal de uma Perdiz e o seu espectro. É notado que a componente mais evidente no espectro está em torno de 60 Hz, evidenciando assim a interferência da rede elétrica. Pode-se notar ainda que há várias componentes de frequência que são mais evidentes no espectro, estes são os harmônicos de 60 Hz. Este tipo de interferência é tão evidente, com magnitudes tão elevadas, que o espectro do canto da ave praticamente não pode ser observado.

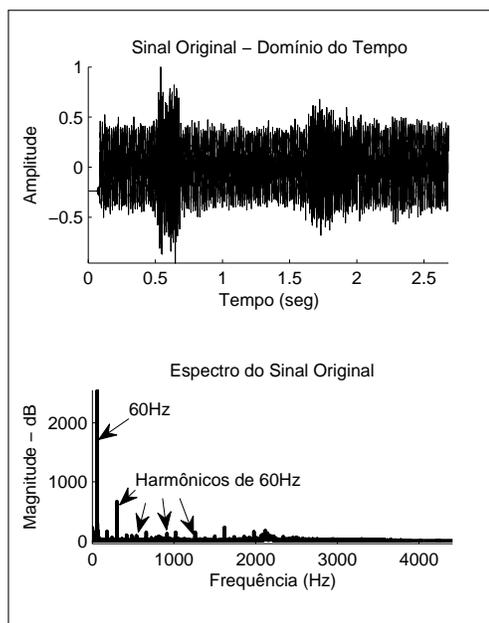


Fig. 2. Sinal e espectro de uma Perdiz, evidenciando a interferência de 60 Hz e seus harmônicos.

A figura 3 mostra o sinal e o espectro da Perdiz após a aplicação do filtro pente. É nítido como o sinal agora está limpo e seu espectro livre dos picos de 60 Hz e dos seus harmônicos.

### C. Detecção Automática de Frequências

As detecções das frequências mínima e máxima devem ocorrer antes da aplicação de um filtro passa faixa, pois as frequências mínima e máxima são usadas como parâmetros do filtro, que será aplicado de forma automática. Além disso, os dados de frequência máxima, mínima e mais forte coletados são também armazenados como características da ave.

Realizar a estimativa da faixa de frequência visualmente é relativamente simples, bastando visualizar o espectrograma do sinal e verificar o valor aproximado das frequências no instante do som. No entanto, como o objetivo do sistema é funcionar de modo automático, buscou-se uma maneira de encontrar a faixa de frequência através de uma rotina que, automaticamente, analisa os valores do espectrograma do sinal. Os valores do espectrograma são dispostos em uma matriz, de tal forma que é possível verificar a densidade espectral da frequência nas linhas e em relação ao tempo nas colunas.

Ao calcular o espectrograma é preciso definir alguns parâmetros que influenciam na quantidade de pontos existentes

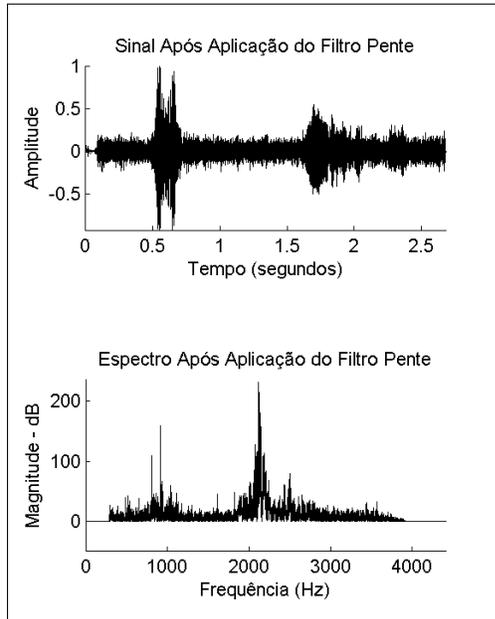


Fig. 3. Sinal e espectro de uma Perdiz, evidenciando a clareza do sinal após a aplicação do filtro pente, que eliminou a interferência de 60 Hz e seus harmônicos.

na matriz de Tempo-Frequência. Para este trabalho o sinal foi janelado com janelas do tipo *Hamming* de tamanho 256, sobreposição de 128 e o tamanho para o cálculo de FFT de 512, resultando em uma quantidade de pontos suficiente para resultados bem aproximados [6-8].

Para calcular os valores da Densidade Espectral de Potência (PSD) é preciso primeiro calcular a Transformada Rápida de Fourier (FFT), para cada segmento de tempo [16]. Então, para cada janela da matriz obtida com cálculo de FFT, é calculada a PSD, que segundo [17], pode ser obtida de forma simplificada via periodograma.

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.

Entretanto, para que esta técnica funcione bem, o som da ave analisada deve conter a maior intensidade presente no sinal, ou seja, nem o ruído, nem outros sons podem ter um volume maior que o da ave em análise.

O valor de limiar adotado foi 70%, escolhido depois de vários testes e análise de resultados, sendo o valor que apresentou um melhor desempenho para a maioria dos sinais analisados. Entretanto, para alguns casos, dependendo do nível

de ruído no sinal, este valor pode ser ajustado, para um melhor desempenho. A subseção seguinte exemplificará a aplicação da filtragem com as frequências detectadas automaticamente.

#### D. Aplicação do Filtro Passa Faixa

Com as frequências mínima e máxima detectadas é aplicado um filtro do tipo passa banda, deixando permanecer no sinal apenas componentes de frequência existentes 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 [16].

Para exemplificar o processo de filtragem automática, serão analisadas as figuras 4 e 5. A figura 4 mostra o mesmo sinal mostrado na figura 3, de uma Perdiz, agora com seu espectrograma servindo como sinal original para a filtragem. A parte mais escura do espectrograma é a frequência fundamental do canto da ave que, visualmente, conforme demarcado, está entre 1800 Hz e 2600 Hz, aproximadamente. É possível notar ainda que o sinal tem um alto nível de ruído.

Ao aplicar a detecção automática nos valores base do espectrograma da figura 4, foram detectadas como frequência mínima, 1808 Hz, máxima, 2590 Hz e frequência mais forte, 2136 Hz. Os valores detectados são bem próximos dos estimados pela análise visual.

A figura 5 mostra o sinal e o espectrograma resultantes do processo de filtragem. É possível perceber inicialmente que o sinal está mais limpo. No espectrograma nota-se que, fora da faixa de frequência do canto da ave, o ruído foi bem atenuado. Percebe-se também que a faixa de frequência não atenuada contempla, de forma bem aproximada, a faixa de frequência fundamental do canto da ave.

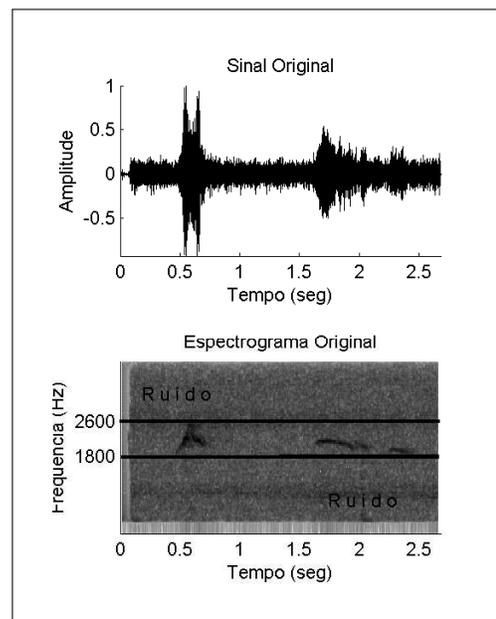


Fig. 4. Sinal e espectrograma de uma Perdiz, destacando o nível de ruído e demarcada a frequência fundamental da ave, através de análise visual.

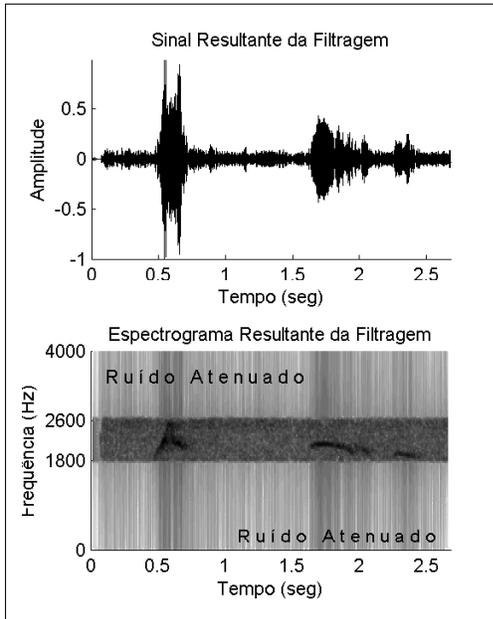


Fig. 5. Sinal e espectrograma resultantes do processo de filtragem, evidenciando uma maior clareza do sinal e atenuação do ruído, sem distorção na frequência fundamental do canto da ave.

### III. ESTIMATIVA DO TEMPO DE CANTO

Com base no espectrograma resultante da filtragem os valores da PSD Tempo-Frequência são novamente analisados, porém, com objetivo de realizar a segmentação das regiões de som e silêncio. Primeiramente é encontrado o valor da PSD relacionado à frequência mínima detectada e este valor é usado então como limiar. A partir de então 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

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

em que 0 representa silêncio e 1 representa som,  $P$  é a matriz de PSD e a notação  $P(:, i)$  indica que estão sendo analisadas todas as linhas de uma determinada coluna  $i$  da matriz  $P$ . 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.

A figura 6 demonstra o resultado da técnica para o som de uma Perdiz com duas regiões de som. Percebe-se que neste caso a segmentação funcionou com estimativa bem precisa das regiões de som, formando duas partes distintas de igual tamanho.

### IV. ANÁLISE DE RESULTADOS

Para análise de alguns resultados, obtidos com a aplicação da metodologia proposta, são apresentadas as tabelas I a III com 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

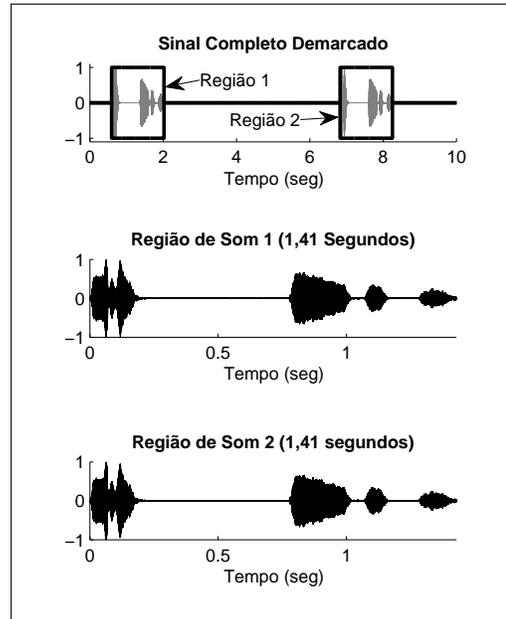


Fig. 6. Exemplo da detecção de regiões de som para o canto de uma Perdiz, sendo detectadas duas regiões de som de igual tamanho.

testes com todas as espécies dos tinamídeos, por razões de espaço não são mostrados todos os resultados. Em seguida, na tabela IV, é apresentado um resumo dos dados coletados para algumas espécies. Nas tabelas serão usadas as seguintes abreviações: 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),  $\mu$  (Média do grupo) e  $\sigma$  (Desvio padrão do grupo).

TABELA I  
DADOS COLETADOS DE *Crypturellus undulatus* (JAÓ).

	FMín	FMáx	FInt	MédC
Jaó1	1033,59	1360,90	1193,32	1,60
Jaó2	1033,59	1378,13	1132,41	1,40
Jaó3	1102,50	1378,13	1262,31	1,71
Jaó4	1085,27	1429,80	1277,40	1,41
$\mu$	1063,74	1386,74	1216,36	1,53
$\sigma$	35,51	29,83	66,87	0,15

TABELA II  
DADOS COLETADOS DE *Crypturellus strigulosus* (INHAMBU RELÓGIO).

	FMín	FMáx	FInt	MédC
Inhambu1	1187,50	1593,75	1380,59	4,13
Inhambu2	1360,90	1584,84	1459,58	4,69
Inhambu3	1274,77	1533,16	1422,93	5,04
Inhambu4	1343,67	1533,16	1433,17	4,69
$\mu$	1291,71	1561,23	1424,07	4,64
$\sigma$	78,81	32,84	32,61	0,38

Nota-se nas tabelas I a III que a detecção automática de frequências e regiões de som tem estimativas aceitáveis, com desvio padrão baixo em relação à grandeza dos valores. O mesmo acontece para os testes não exibidos aqui. Em alguns casos raros porém, foram detectadas discrepâncias,

TABELA III  
DADOS COLETADOS DE *Rhynchotus rufescens* (PERDIZ).

	FMín	FMáx	FInt	MédC
Perdiz1	1670,98	3049,10	1927,96	2,38
Perdiz2	1718,75	3031,25	2088,49	2,00
Perdiz3	1722,66	2894,06	1963,53	1,43
Perdiz4	1808,79	2980,20	2215,78	1,99
$\mu$	1730,30	2988,65	2048,94	1,95
$\sigma$	57,36	69,49	130,81	0,39

TABELA IV  
RESUMO DE DADOS COLETADOS DE ALGUMAS ESPÉCIES.

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

como constatado da tabela III em que o desvio padrão está relativamente alto, para a frequência mais forte, isso ocorre por se tratar de sinais com alto nível de ruído. 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.

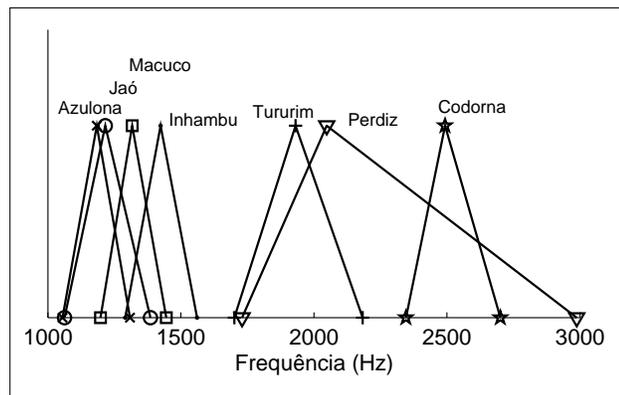


Fig. 7. Regiões das frequências para as aves analisadas. As bases representam, respectivamente, as frequências mínima e máxima e os picos representam as frequências mais intensas para cada espécie

A tabela IV apresenta apenas as médias e desvios padrões para 7 das espécies. É possível com esta tabela ratificar que os valores detectados automaticamente são bem homogêneos e distintos entre os grupos, conforme pode ser notado graficamente pela figura 7, servindo como características que auxiliam a classificação automática da espécie.

## V. 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 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

- [1] Correio do Brasil, "Bioacústica Identifica Novas Aves na Amazônia", 2011, <http://correiodobrasil.com.br/bioacustica-identifica-novas-aves-na-amazonia/152369/>.
- [2] T. S. Brandes, "Automated Sound Recording and Analysis Techniques for Bird Surveys and Conservation", Bird Conservation International, pp. S163-S173, 2008.
- [3] C.-H. Lee, C.-C. Han and C.-C. Chuang, "Automatic Classification of Bird Species From Their Sounds Using Two-Dimensional Cepstral Coefficients", IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, n. 8, pp. 1541-1550, 2008.
- [4] J. C. Brown and P. Smaragdis, "Hidden Markov and Gaussian Mixture Models for Automatic Call Classification", J. Acoust. Soc. Am. Vol. 125, n. 6, pp. EL221-EL224, 2009.
- [5] A. Selin, J. Turunen, and J. T. Tanntu, "Wavelets in Recognition of Bird Sounds", EURASIP Journal on Applied Signal Processing, vol. 2007, pp. 1-9, 2007.
- [6] K. Kwan, G. Mei, X. Zhao, Z. Ren, R. Xu, V. Stanford, C. Rochet, and J. Aube, K.C. Ho, "Bird Classification Algorithms: Theory and Experimental Results", ICASSP2004, pp. V289-V292, 2004.
- [7] S. Fagerlund, "Bird Species Recognition Using Support Vector Machines", EURASIP Journal on Advances in Signal Processing, pp. 1-11, 2007.
- [8] C.-H. Chou, C.-H. Lee and H.-W. Ni, "Bird Species Recognition by Comparing the HMMs of the Syllables", IEEE, pp. 1-4, 2007.
- [9] J. T. Descourtiz, "História Natural das Aves do Brasil – Notáveis por Sua Plumagem, Canto e Hábitos", Belo Horizonte, Itatiaia, 1983.
- [10] R. T. Pinheiro, "Conservação de Aves do Cerrado: Conhecimento e Perspectivas - O Caso do Tocantins", Anais do VIII Congresso de Ecologia do Brasil, 23 a 28 de Setembro de 2007, Caxambu - MG.
- [11] H. Sick, "Ornitologia Brasileira - Uma Introdução". Brasília: Ed. Universidade de Brasília, 1984.
- [12] D. H. Kil; F. B. Shin, "Pattern Recognition and Prediction with Applications to Signal Characterization", AIP Press, American Institute of Physics, 1996.
- [13] J. W. Picone, "Signal Modeling Techniques in Speech Recognition", Proceedings of The IEEE, vol. 81, no. 9, september 1993, 1215-1247.
- [14] "Cornell Lab of Ornithology", <http://macaulaylibrary.org/index.do>. Último acesso 08/04/2011.
- [15] "WikiAves", <http://www.wikiaves.com.br/>. Último acesso 08/04/2011.
- [16] A. V. Oppenheim; R. W. Schaffer, "Discrete-Time Signal Processing – Second Edition", New Jersey, Prentice Hall, 1999.
- [17] P. Stoica; R. Moses, "Introduction to Spectral Analysis", Prentice Hall, 2005.