

Classificação de Anuros Baseado em Vocalizações para Monitoramento Ambiental Pervasivo

Juan Gabriel Colonna¹, Eduardo Freire Nakamura^{1,2}, Eulanda Miranda dos Santos¹

¹Departamento de Ciência da Computação (DCC)
Universidade Federal do Amazonas (UFAM)

²Núcleo de Estudos e Pesquisas em Computação (NEPComp)
Fundação Centro de Análise, Pesquisa e Inovação Tecnológica (FUCAPI)

{juancolonna, nakamura, emsantos}@dcc.ufam.edu.br

Abstract. *In this work, we developed an automatic system for classifying anura (frogs and toads) based on their vocalizations. Each bioacoustic signal has been segmented on the first stage into smaller units called “syllables”, followed by a preprocessing step, composed by a pre-emphasis filter and a Hamming window that prepares the signal for feature extraction. We used Mel-fourier Cepstral Coefficients (MFCCs) to represent the acoustic signals and two classifiers were evaluated: kNN and SVM. In our experiments, we have achieved a classification rate of 98.97%, which shows that the MFCC, usually used in speech recognition, can be used for the recognizing anura species as well. The anuran recognition rate was improved in 16.09%, using SVM and MFCCs, compared with results found in literature.*

Resumo. *Neste trabalho desenvolvemos um sistema automático de classificação de anuros (rãs e sapos) baseado em suas vocalizações (coachar). Cada sinal bioacústico foi primeiramente segmentado em unidades menores chamadas “sílabas”, posteriormente uma etapa de pré-processamento, composta por um filtro de pré-ênfase e uma janela de Hamming, prepara o sinal para a extração das características. Foram usados os Coeficientes Mel (MFCCs) para representar os sinais bioacústicos e dois classificadores foram avaliados: kNN e SVM. Em nossos experimentos, alcançamos uma taxa de acerto de 98,97%, que demonstra que MFCC, geralmente utilizada em reconhecimento de fala, pode ser usada para o problema de reconhecimento de anuros. A taxa de reconhecimento de anuros foi melhorada em 16,09%, usando SVM e os MFCCs, comparada com os resultados encontrados na literatura.*

1. Introdução

Os anfíbios, mais precisamente os anuros, relacionam-se intimamente com o ecossistema, sendo afetados de forma direta pelas mudanças ambientais [Collins and Storfer, 2003]. Segundo Carey et al. [2001] existe uma relação direta entre as mudanças climáticas e a mortalidade em populações de anfíbios, de maneira que estes podem ser utilizados como indicadores de estresse ecológico precoce. Assim, um sistema de classificação de espécies

e estimação de população de anuros representa uma ferramenta de grande importância para a área de Ecologia e Biologia.

Geralmente, a classificação de anuros usando as vocalizações é realizada manualmente mediante o uso de um analisador de espectro e um computador [Riede, 1993]. O resultado da classificação depende da experiência da pessoa que realiza a tarefa, tornando-se lenta e sujeita a erros. A combinação do sistema de classificação automático com uma Rede de Sensores Sem Fio (RSSF) [Akyildiz et al., 2002], auxilia esta tarefa.

Resultados prévios [Taylor et al., 1996; Huang et al., 2009], usando técnicas de aprendizagem de máquina [Marsland, 2009], mostram que é possível reconhecer anuros mediante a aplicação das técnicas: C4.5 [Quinlan, 1993], *k-Nearest Neighbor* (kNN) [Cover and Hart, 1967] e *Support Vector Machine* (SVM) [Boser et al., 1992].

Neste trabalho, são usados os *Mel-Fourier Cepstral Coefficient* (MFCCs) [Davis and Mermelstein, 1980], geralmente usados em reconhecimento de fala, para representar os sinais bioacústicos emitidos por anuros. Estes coeficientes são comparados com as características implementadas por Huang et al. [2009], e com os classificadores kNN e SVM para melhorar a taxa de classificação (Seção 4).

Usando os MFCCs, é possível reduzir a quantidade de informação necessária, em comparação à transmissão das amostras completas de um áudio, para classificar os sinais bioacústicos. Esta vantagem permite integrar os MFCCs junto com uma RSSF para realizar a classificação só transmitindo o valor dos coeficientes pela rede, aumentando a vida útil da RSSF, poupando energia e memória. A coleta dos áudios e a classificação pode ser feita como é mostrado na figura 1.

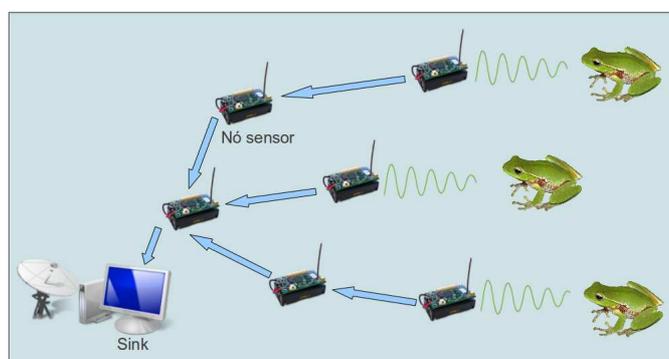


Figura 1. Coleta de áudios e classificação usando uma RSSF.

A principal contribuição deste trabalho é um método de classificação para anuros que apresenta uma melhora de 16,09% na taxa de reconhecimento de anuros, em relação aos métodos da literatura. A taxa de acerto foi comparada com o resultado obtido pela aplicação da técnica de Huang et al. [2009] com o mesmo classificador. Portanto, o foco do trabalho está na eficácia do método proposto, em relação à tarefa de classificação. Assim, a rede de sensores foi simplificada para apenas coletar o áudio dos animais.

O restante deste trabalho está organizado como segue. A seção 2 apresenta uma síntese dos trabalhos relacionados. Na seção 3 é explicado o método de reconhecimento proposto. Na seção 4 encontram-se os experimentos e resultados obtidos. Finalmente, na seção 5 são analisados os resultados e discutidas extensões futuras deste trabalho.

2. Trabalhos relacionados e fundamentos

Os anuros produzem vocalizações intensas e agudas, as quais possuem características de excitação na frequência mais concentradas, com uma maior quantidade de harmônicos. Existe similaridade entre o modelo de geração da fala humana, descrita por Deller et al. [1993], e o modelo de geração da vocalização de um anuro, ambos formados por cavidades ressonantes. Além disso, [Bee and Micheyl, 2008] destacam as similaridades entre a forma de percepção dos sons emitidos por animais e por humanos.

A primeira proposta para monitoramento de anuros usando uma RSSF foi apresentada por Hu et al. [2005], cujo interesse principal era identificar o anuro *Cane-Toad*, conhecido como *Bufo marinus*, no Parque Nacional Kakadu, Austrália. A técnica de classificação usada foi desenvolvida por Taylor et al. [1996]. A característica usada pelo classificador foi o nível de energia dos pixels do espectrograma. As informações de cada pixel e seus vizinhos eram classificadas por uma árvore de decisão C4.5. O classificador mostrou um desempenho de 50% de classificação correta em condições onde ocorre ruído proveniente de outras espécies. Neste trabalho o áudio capturado pelos nós sensores da rede era transmitido até o nó central, que era o encarregado de efetuar a classificação. O trabalho permitiu concluir que 90% da área do parque Kakadu é habitada pelo *Bufo marinus*, permitindo identificar as zonas de concentração de suas populações.

Uma abordagem diferente de reconhecimento de anuros mediante a aplicação das técnicas kNN e SVM foi proposta por Huang et al. [2009]. Neste trabalho foram utilizadas cinco espécies de anuros diferentes. Cada sinal bioacústico foi segmentado em unidades menores chamadas sílabas (seção 4.1), e de cada sílaba três características foram extraídas: *Spectral Centroid* (S), *Signal Bandwidth* (B) e *Zero-crossing Rate* (R). O *Spectral Centroid* - S é considerado o ponto central de espectro de frequências e se relaciona com o tom da vocalização. Sua definição se assemelha ao do centro de massa, representando a distribuição de potência sobre a frequência, e definido como:

$$S = \frac{\sum_{n=0}^M n |X_n|^2}{\sum_{n=0}^M |X_n|^2}, \quad (1)$$

onde X_n é a DFT (*Transformada Discreta de Fourier*) da sílaba, M é a metade positiva do espectro de X_n e n é a frequência.

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

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

onde X_n é a DFT da sílaba, M é metade positiva do espectro de X_n e S é o *Spectral Centroid*.

A taxa de cruzamento por zero (*Zero-crossing Rate* - R) indica quantas vezes o sinal passou de um valor positivo para um negativo, ou vice-versa. Esta característica

proporciona uma estimativa do comprimento da sílaba e definida como:

$$T = \frac{1}{2} \sum_{n=0}^{M-1} |tsgn(x_n) - tsgn(x_{n+1})|, \quad (3)$$

$$tsgn(x_n) = \begin{cases} +1 & x_n \geq \eta \\ -1 & x_n \leq -\eta \end{cases}, \quad (4)$$

onde x_n é o valor de amplitude da amostra e η é o limiar de amplitude.

Os coeficientes Mel MFCCs foram usados anteriormente por Clemins [2005] para classificar elefantes e baleias. O objetivo de seu trabalho era desenvolver um sistema de reconhecimento animal, que permita identificar a maior quantidade de animais possível.

3. Método proposto

A arquitetura do sistema proposto para a identificação de anuros é composta por três blocos principais: pré-processamento, extração das características e classificação, onde cada bloco contém blocos menores que realizam tarefas específicas. A figura 2(a) mostra como está formado o sistema.

O primeiro bloco do sistema, que realiza o pré-processamento, é formado por três blocos menores: segmentação, pré-ênfase e janelamento, figura 2(b). Esta etapa é a encarregada de acondicionar o sinal para a posterior extração de características (figura 2(c)).

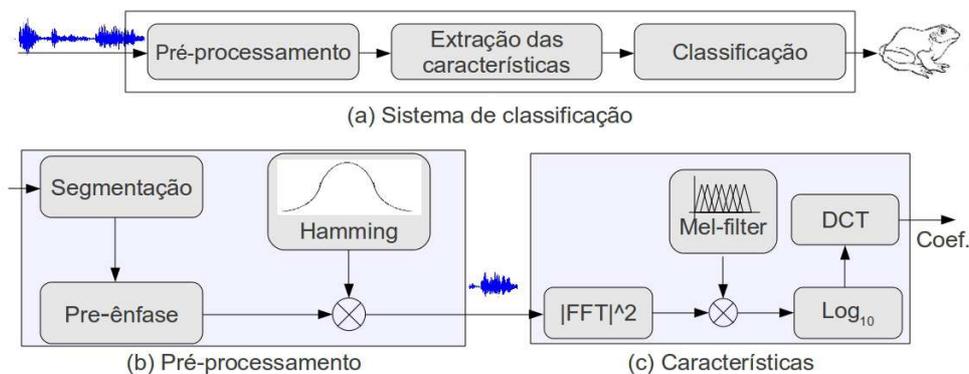


Figura 2. Sistema de classificação.

Começando pela segmentação, cada sinal é dividido em unidades menores chamadas sílabas (seção 4.1). As sílabas, dependendo da espécie, possuem frequências fundamentais (*pitch*) superiores a 1,0 kHz. Em algumas espécies, como o *Bufo ornatus*, as frequências começam em 2,5 kHz.

Dada a característica de geração, as frequências elevadas tendem a possuir magnitudes menores, distorção ou saturação. Para compensar este efeito, um filtro de pré-ênfase amplifica as altas frequências. Para realizar a pré-ênfase, o sinal digitalizado $s_{[n]}$, é modificado mediante a aplicação do filtro:

$$s_{[n]} = s_{[n]} - a s_{[n-1]}, \quad (5)$$

onde normalmente em sistemas de comunicação $a = 0,9$.

Cada sílaba pode ser considerada um sinal aperiódico. O algoritmo de cálculo da transformada de Fourier (FFT) produz uma aproximação do espectro com um número finito de harmônicos. Para diminuir o efeito desta aproximação, as sílabas são multiplicadas por uma janela tipo *Hamming*, onde as amplitudes dos componentes são alteradas pela aplicação da seguinte equação:

$$w_{[n]} = 0,54 - 0,46 \cos \left(2\pi \frac{s_{[n]}}{N} \right), 0 \leq s_{[n]} \leq N, \quad (6)$$

onde N é o comprimento da janela e $s_{[n]}$ são as amostras do sinal.

Após o pré-processamento, cada sinal ingressa no bloco de extração de características, figura 2(c). A característica usada para representar os sinais são os coeficientes MFCCs [Davis and Mermelstein, 1980]. A idéia básica desta técnica é fazer uma análise espectral baseada num banco de filtros triangulares espaçados logaritmicamente na frequência, figura 3(a).

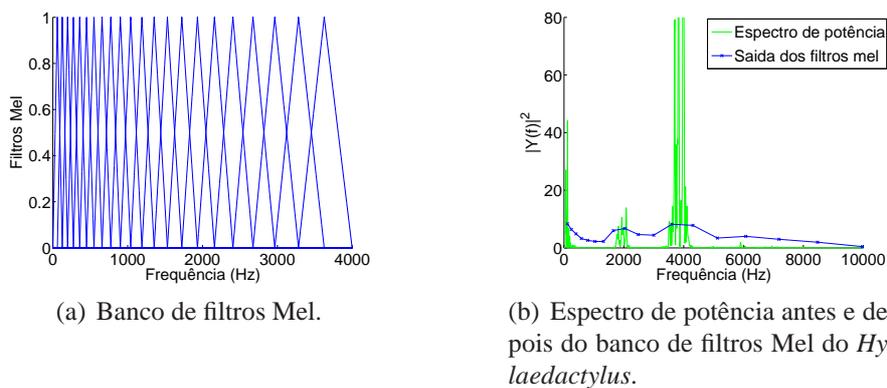


Figura 3. Banco de filtros Mel e espectro de frequências.

Esta característica representa melhor os sinais da fala que as características que usam uma escala de frequência linear, 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 figura 3(b) mostra o espectro de frequências de uma sílaba da espécie *Hylaedactylus* e a redução de informação após aplicar o banco de filtros Mel. O espaçamento entre filtros é definido na escala Mel por:

$$f_{mel} = 1127 * \ln \left(1 + \frac{f_{Hz}}{700} \right). \quad (7)$$

A aplicação dos filtros sobre o espectro do sinal produz um valor $M_{[r]}$ para cada filtro. Após obter os valores $M_{[r]}$, é feito o logaritmo e aplicada a transformada discreta do cosseno (DCT), para obter os coeficientes Mel $mfcc_{[m]}$ [Rabiner and Schafer, 2007]:

$$mfcc_{[m]} = \frac{1}{R} \sum_{r=1}^R \log (M_{[r]}) \times \cos \left(\frac{2\pi}{R} \left(r + \frac{1}{2} \right) m \right), \quad (8)$$

onde m é o número de coeficiente, R é a quantidade de filtros.

O bloco de classificação, figura 4, identifica a espécie de anuro baseando-se nos valores dos coeficientes Mel. Os métodos implementados dentro deste bloco são: *k-Nearest Neighbors* (kNN) e *Support Vector Machine* (SVM) [Marsland, 2009]. Posteriormente na seção 4.2 são comparados os resultados dos classificadores.

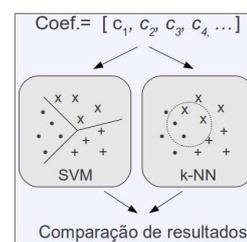


Figura 4. Bloco de classificação.

4. Avaliação

A avaliação do sistema foi realizada sobre uma base de dados com quatro espécies: *Hylaedactylus*, *Rhinella granulosa*, *Adenomera andreae* e *Ameerega trivittata*. A métrica de avaliação usada é a taxa de classificação correta T_c , definida como a razão entre a quantidade de mostras classificadas corretamente e o total de amostras.

4.1. Metodologia

As amostras de áudio foram coletadas nas regiões próximas a Manaus, durante um intervalo de dois meses e complementados com áudios de Haddad [2005]; Márquez et al. [2002]. Cada vocalização foi armazenada em formato *wav* com uma frequência de amostragem máxima de 44,1 kHz e 16 bits por amostra, o que permite analisar sinais bioacústicos de até 22,05 kHz.

Um exemplo típico de vocalização do anuro *Adenomera andreae* é mostrado na figura 5. Existem dois métodos para segmentar o sinal em sílabas: (1) o método descrito por Huang et al. [2009] e (2) o desenvolvido por Harma [2003]. Neste trabalho é usado primeiro método, pois proporciona um comprimento uniforme para todas as sílabas.

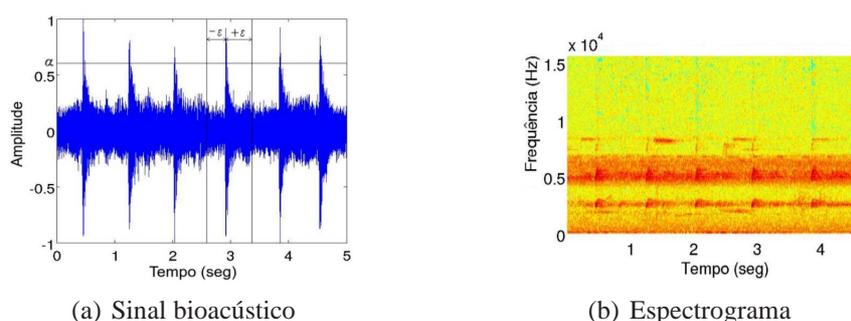


Figura 5. Vocalização da espécie *Adenomera andreae*.

A segmentação em sílabas é realizada pelo seguinte algoritmo simplificado:

1. Buscar o valor de máximo de amplitude do sinal $S(t)$;
2. Se $S(t)$ for menor que o limiar α , ir ao passo 5, senão, ir ao passo 3;
3. Extrair ε milissegundos a direita e a esquerda do valor máximo, $S_{(t-\varepsilon)} \leq$ sílaba $\leq S_{(t+\varepsilon)}$, formando a sílaba;
4. Extrair as características da sílaba e apagar os valores do sinal original S no intervalo $t \pm \varepsilon$, voltar ao passo 1;
5. Terminar a segmentação.

A variável ϵ , que determina o comprimento da sílaba, é definida com o valor de 200 ms, o que garante a inclusão da sílaba inteira, das cinco espécies, e mais uma margem. No mesmo processo de segmentação, a variável α , do limiar de amplitude, foi definida com os valores seguintes: 40% ($\alpha = 0,4$), 50% ($\alpha = 0,5$) e 60% ($\alpha = 0,6$) da amplitude máxima. Os valores de α maiores que 60% não apresentaram melhoras significativas na taxa de classificação. Com valores menores que 40% foram extraídas sílabas com uma relação sinal-ruído baixa.

Cada sílaba passa pela etapa de pré-processamento e extração das características para formar a base de treinamento e teste dos classificadores. Do processo de segmentação são obtidas três bases, uma para cada valor da variável α . A tabela 1 mostra a quantidade de indivíduos de cada espécie e a quantidade total de sílabas usadas como base de treino e teste para os classificadores.

Espécie	Indivíduos	Sílabas		
		$\alpha = 0,4$	$\alpha = 0,5$	$\alpha = 0,6$
<i>Rhinella granulosa</i>	8	1147	1103	964
<i>Hylaedactylus</i>	3	3354	3172	2898
<i>Adenomera andreae</i>	7	348	241	202
<i>Ameerega trivittata</i>	5	614	545	328
Total	23	5463	5061	4392

Tabela 1. Quantidade de sílabas nas bases usadas nos experimentos.

Cada base é formada pelos vetores de características. Neste caso, cada vetor é composto por quinze valores de características. A primeira característica usada possui doze coeficientes Mel (MFCCs), que são obtidos pela aplicação de vinte e dois filtros triangulares, distribuídos logaritmicamente entre 0 e 22,05 kHz. As últimas três características são: *Spectral Centroid* (S), *Signal Bandwidth* (B) e *Zero-crossing Rate* (R), usadas por Huang et al. [2009] descritas na seção 2.

O espectro de frequências de cada sílaba é obtido através da FFT com 2048 pontos. A figura 6 exhibe o espectro de frequências de uma sílaba da espécie *Hylaedactylus*.

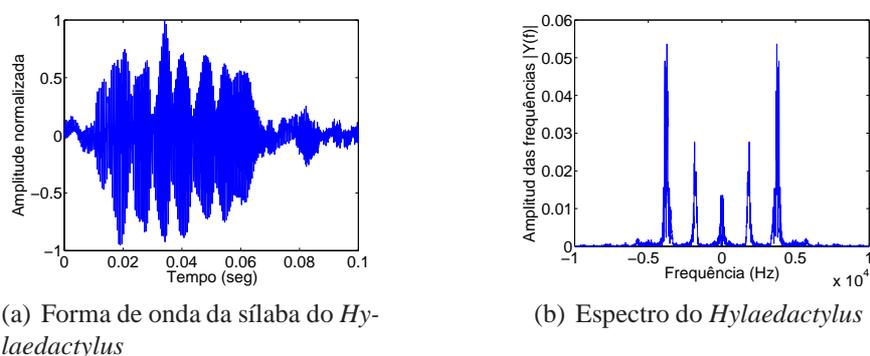


Figura 6. Forma de onda e espectro da vocalização do anuro *Hylaedactylus*.

Para a implementação dos classificadores foi usado o software Weka [Hall et al., 2009]. Estes classificadores foram configurados da forma seguinte: kNN com valores $k = 5$ e $k = 10$ e SVM com Kernel polinomial de grau um e dois ($p = 1$ e $p = 2$). Os

experimentos foram realizados usando validação cruzada com dez conjuntos ($folds = 10$). Os valores definidos para os parâmetros foram calculados empiricamente para maximizar a taxa de classificação.

4.2. Resultados

Os resultados da classificação realizada com kNN, onde k é igual a cinco e dez, são apresentados na tabela 2. A tabela mostra a relação entre a taxa de classificação e a variável de amplitude da segmentação em sílabas α . A coluna esquerda, indica a combinação de característica testada, e no centro, o valor indica a taxa de classificação correta. Nas tabelas a seguir, S significa *Spectral Centroid*, B, *Signal Bandwidth*, R, *Zero-crossing Rate*, e RSB as três características em conjunto (método de Huang et al. [2009]).

Características	kNN $k = 5$			kNN $k = 10$		
	$\alpha = 0,4$	$\alpha = 0,5$	$\alpha = 0,6$	$\alpha = 0,4$	$\alpha = 0,5$	$\alpha = 0,6$
MFCCs	98,90%	98,97%	98,97%	98,90%	98,85%	98,90%
RSB	90,40%	91,70%	93,76%	90,75%	91,87%	93,28%
R e MFCCs	98,95%	98,99%	98,99%	99,13%	98,97%	99,02%
S e MFCCs	99,17%	99,26%	99,18%	99,08%	99,15%	99,22%
B e MFCCs	99,19%	99,22%	99,15%	99,17%	99,15%	99,22%
RSB e MFCCs	99,28%	99,24%	99,27%	99,30%	99,26%	99,24%

Tabela 2. Taxa de classificação de kNN em relação a α .

A tabela 2 mostra que com o classificador kNN a maior taxa de classificação correta foi obtida usando $k = 5$, $\alpha = 0,6$ e a combinação de RSB. A maior diferença entre a característica MFCCs e RSB foi 8,5% para $k = 5$ e 8,15% para $k = 10$, ambos na base com $\alpha = 0,4$. Este último resultado indica que os MFCCs possuem um desempenho melhor que os RSB em sinais com ruído. A característica combinada com os MFCCs que produziu maior impacto no resultado foi S.

O mesmo procedimento de combinação de características descrito anteriormente foi utilizado com o classificador SVM. A tabela 3 mostra os resultados das classificações com SVM em relação à variável α , com Kernel polinomial de grau um e dois.

Características	SVM $p=1$			SVM $p=2$		
	$\alpha = 0,4$	$\alpha = 0,5$	$\alpha = 0,6$	$\alpha = 0,4$	$\alpha = 0,5$	$\alpha = 0,6$
MFCCs	97,10%	97,35%	97,60%	97,74%	97,98%	98,26%
RSB	81,54%	86,36%	88,02%	82,53%	85,55%	87,24%
R e MFCCs	97,36%	97,72%	97,65%	97,98%	98,24%	98,31%
S e MFCCs	97,36%	97,47%	97,79%	98,09%	98,39%	98,58%
B e MFCCs	97,08%	97,47%	97,54%	97,98%	98,39%	98,13%
RSB e MFCCs	97,45%	97,86%	97,90%	98,62%	98,69%	98,56%

Tabela 3. Taxa de classificação de SVM em relação a α .

Neste caso, usando o classificador SVM e os MFCCs, o melhor resultado foi uma taxa de classificação correta de 98,26% com $p = 2$, $\alpha = 0,5$. O resultado total maior foi 98,69% com $p = 2$, $\alpha = 0,5$ e a combinação de todas as características. Este resultado superou em 0,71% o resultado obtido com os MFCCs e 13,14% com RSB, na mesma base. Neste caso, pode ser observado que a combinação dos MFCCs e os

RSB, não apresentou uma diferença substancial no resultado. Novamente a característica combinada que produz o maior impacto foi S .

Para mostrar os erros de classificação, montamos uma matriz de confusão que 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 que foi confundida. A tabela 4 apresenta as matrizes para os casos kNN ($k = 5$) e SVM ($p = 2$) que apresentaram taxa de classificação melhor usando os MFCCs e $\alpha = 0,6$. Pode-se observar que as espécies mais confundidas pelo classificador foram: *Rhinella granulosa* com *Hylaedactylus* e *Adenomera andre* com *Ameerega trivittata*. Isto ocorre pela proximidade entre as características sonoras entre estas espécies.

Espécie	kNN, $k = 10$				SVM, $p = 2$			
	a	b	c	d	a	b	c	d
$a = Adenomera andre$	179	20	0	3	166	34	0	2
$b = Ameerega trivittata$	5	320	1	2	15	308	1	4
$c = Hylaedactylus$	0	2	952	10	0	5	949	10
$d = Rhinella granulosa$	0	1	1	2896	1	1	3	2893

Tabela 4. Matriz de confusão para kNN e SVM (melhor caso usando MFCCs).

5. Comentários finais

Neste trabalho, propusemos e avaliamos um sistema de classificação automática de anuros, usando MFCCs para a representação dos sinais bioacústicos emitidos por eles. O uso desta característica foi comparado com características desenvolvidas anteriormente no trabalho realizado por Huang et al. [2009]. Com um resultado de 98,97% de acerto, podemos concluir que os MFCCs, usados para representar sinais de fala, podem ser adaptados ao problema de reconhecimento de anuros. Observamos ainda que a combinação de todas as características não produz uma diferença substancial na taxa de acerto. Os classificadores kNN e SVM apresentaram bom desempenho. A melhor combinação foi: o classificador kNN com todas as características. O aumento do limiar α produz um aumento na taxa de classificação usando os RSB, indicando uma dependência desta característica com o nível de ruído, não sendo assim com os MFCCs. A taxa de reconhecimento de anuros foi melhorada em 16,09% usando SVM comparada o estado-da-arte (método de Huang et al. [2009]).

Futuramente, mais espécies serão agregadas com objetivo de determinar a capacidade de diferenciação dos MFCCs. O classificador será implementado em uma RSSF para facilitar o monitoramento das variações nas populações de anuros. Também consideramos a implementação de uma etapa de filtragem de ruídos no método, visando melhorar o resultado da classificação em cenários de floresta.

Agradecimentos

Agradecemos ao Professor Marcelo Gordo (UFAM) pela colaboração na coleta e análise dos áudios; agradecemos também aos projetos: ANURA (PRONEX e LACCIR) financiado pela FAPEAM, CNPq e Microsoft; e SAUIM (CNPq, processo 55.4087/2006-5). Agradecemos também à CAPES pelo suporte através de bolsa de mestrado.

Referências

- Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cayirci, E. (2002). Wireless sensor networks: a survey. *Computer Networks*, 38(4):393–422.
- Bee, M. A. and 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.
- Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory.*, pages 144–152, New York, USA. ACM.
- Cai, J., Ee, D., Pham, B., Roe, P., and Zhang, J. (2007). Sensor Network for the Monitoring of Ecosystem: Bird Species Recognition. In *3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pages 293–298. IEEE.
- Carey, C., Heyer, W. R., Wilkinson, J., a. Alford, R., Arntzen, J. W., Halliday, T., Hungerford, L., Lips, K. R., Middleton, E. M., a. Orchard, S., and a. Stanley Rand (2001). Amphibian Declines and Environmental Change: Use of Remote-Sensing Data to Identify Environmental Correlates. *Conservation Biology*, 15(4):903–913.
- Clemins, P. J. (2005). *Automatic classification of animal vocalizations*. PhD thesis, Marquette University, Wisconsin.
- Collins, J. P. and Storfer, A. (2003). Global amphibian declines: sorting the hypotheses. *Diversity and Distributions*, 9:89–98.
- Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *Transactions on Information Theory, IEEE*, 13(1):21–27.
- Davis, S. and 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.
- Deller, J., Hansen, J., and Proakis, J. (1993). *Discrete-time processing of speech signals*, volume 1. IEEE.
- Haddad, C. (2005). Guia Sonoro dos Anfíbios Anuros da Mata Atlântica.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA Data Mining Software: An Update.
- Harma, A. (2003). Automatic identification of bird species based on sinusoidal modeling of syllables. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 545–548. IEEE.
- Hu, W., Tran, V. N., Bulusu, N., Chou, C. T., Jha, S., and Taylor, A. (2005). The design and evaluation of a hybrid sensor network for cane-toad monitoring. In *Fourth International Symposium on Information Processing in Sensor Networks.*, pages 503–508.
- Huang, C.-J., Yang, Y.-J., Yang, D.-X., and Chen, Y.-J. (2009). Frog classification using machine learning techniques. *Expert Systems with Applications*, 36(2):3737–3743.
- Márquez, R., Riva, I., Matheu, B., and Matheu, E. (2002). Sounds of Frogs and Toads of Bolivia.
- Marsland, S. (2009). *Machine Learning: an algorithmic perspective.*, volume 1. CRC Press, Palmerston North, New Zealand.
- Quinlan, J. R. (1993). *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Rabiner, L. R. and Schafer, R. W. (2007). Introduction to Digital Speech Processing. *Foundations and Trends in Signal Processing*, 1:1–194.
- Riede, K. (1993). Monitoring biodiversity: analysis of Amazonian rainforest sounds. *Ambio*, 22(8):546–548.
- Taylor, A., Watson, G., Grigg, G., and McCallum, H. (1996). Monitoring Frog Communities: An Application of Machine Learning. *Proceedings of the 8th Innovative Applications of Artificial Conference*.