



UM SISTEMA PARA CLASSIFICAÇÃO AUTOMÁTICA DE GÊNEROS MUSICAIS

Vítor de Andrade Coutinho¹; Hélio Magalhães de Oliveira²

¹Estudante do Curso de Engenharia Eletrônica- CTG – UFPE; E-mail: vitor.andrade.coutinho@gmail.com,

²Docente/pesquisador do Depto de Eletrônica e Sistemas – CTG – UFPE. E-mail: hmo@ufpe.br.

Sumário: Este trabalho descreve a implementação de um aplicativo que realiza uma classificação automática de arquivos de música em gêneros musicais, utilizando a *plataforma MATLAB*[®]. O procedimento de classificação foi dividido em duas etapas, implementadas independentemente: A extração de características e a classificação propriamente dita. A extração consiste em identificar e separar informações relevantes do sinal de áudio, enquanto a classificação tem o objetivo de verificar quais combinações dessas informações correspondem a cada categoria. Neste trabalho, as características extraídas do sinal de áudio dizem respeito essencialmente ao timbre da música, o que se relaciona com os instrumentos presentes. A principal característica de timbre utilizada corresponde aos *Mel Frequency Cepstral Coefficients (MFCC)*. A classificação foi realizada através de redes neuronais artificiais. No caso de {clássica, MPB e Rock} os acertos médios foram 96%, e para {MPB, Forró, Frevo e Rock} obteve-se quase 80%.

Palavras-chave: matlab; mfcc; neural; música

INTRODUÇÃO

Com a larga distribuição de música por meios eletrônicos, as bibliotecas virtuais de áudio estão crescendo em grandes proporções. Esse fato sugere a necessidade da organização dos catálogos, através da classificação em gêneros musicais. O Gênero é uma característica musical com alto nível de abstração, o que torna o processo de classificação uma tarefa não trivial [1], [2].

Em 1997, Scheirer e Slaney [3] utilizaram reconhecimento de padrões estatísticos baseados em características espectrais para discriminar música e fala, assim como Martin o fez para a identificação de instrumentos. Em 2000, Scheirer explorou a extração de características psico-acústicas relacionadas à textura de música e o uso das mesmas para avaliação de similaridade [4]. Também em 2000, Logan propôs a utilização de *Mel Frequency Cepstral Coefficients*, usados até então no problema de reconhecimento de voz, como parâmetros para modelagem de música [5]. Aucouturier e Pachet propuseram computar a similaridade entre peças de músicas se baseando no timbre global das mesmas [6].

Esse trabalho descreve o projeto e o desenvolvimento de um sistema automático de classificação em gênero musical, através da extração de características *MFCC (Mel Frequency Cepstral Coefficients)* do sinal de áudio e subsequente classificação através de Redes Neuronais [7]. Ele envolve também o estudo de diferentes características relevantes além do *MFCC*, visando o melhor desempenho do sistema. O plano de fundo para o trabalho foi a Música Popular Brasileira (gêneros brasileiros), mas incluiu outros gêneros.



MATERIAIS E MÉTODOS

As características para a modelagem de música neste trabalho foram os *Mel Frequency Cepstral Coefficients (MFCC)*. *MFCCs* são frequentemente utilizados no problema de reconhecimento de fala por aparentemente extrair partes relevantes do espectro para a percepção humana e vem sendo utilizado também no reconhecimento de gênero musical [5]. As principais etapas na extração do *MFCC* são: Converter o sinal em janelas (*frames*); Cálculo da *DFT (Discrete Fourier Transform)* de cada *frame* [8]; Conversão para a escala Mel [9] através de Banco de filtros; Cálculo do Logaritmo; Cálculo da *DCT (Discrete Cosine Transform)* [8]. Ao final dessas etapas, um vetor de características *MFCC* é obtido. O banco de filtros é composto por doze (parâmetro ajustável) filtros triangulares, igualmente espaçados na escala mel [9]. A relação entre a escala de frequência em Hertz para mel é dada por

$$m = 1127 \log_e \left(1 + \frac{f}{700} \right), \quad (1)$$

em que m é a frequência em *mel* e f é expressa em *Hz*. As características descritas a seguir foram também utilizadas como informações adicionais a serem extraídas.

- **Taxa de Cruzamento por Zero:** Um “cruzamento por zero” ocorre quando o sinal digital $x(n)$ muda de sinal (algébrico). A taxa de cruzamento por zeros é a quantidade de vezes que isso ocorre, sendo uma medida do conteúdo de frequência do sinal e pode ser expressa por

$$Z = \frac{1}{2} \sum_{n=1}^{N-1} |\text{sgn}[x(n)] - \text{sgn}[x(n-1)]|; \quad (2)$$

- **Loudness:** Característica perceptual que tenta representar a intensidade sonora percebida pelo ouvido humano. É calculada de acordo com

$$L = \log_2 \left[1 + \frac{1}{N} \sum_{n=0}^{N-1} |x(n)| \right]; \quad (3)$$

- **Centroide Espectral:** Representa o “centro de massa” do espectro. Denota-se por $X(k)$, a DFT do sinal $x(n)$. Está associado com a noção perceptual de “brilho” do som e é expresso por

$$C = \frac{\sum_0^{N-1} f \cdot |X(k)|^2}{\sum_0^{N-1} |X(k)|^2}. \quad (4)$$

- **Frequência Rolloff:** representa uma medida do formato espectral. Corresponde a uma frequência R para a qual está concentrado $r\%$ da magnitude do espectro abaixo dela. Foi assumido neste projeto o valor de $r = 85\%$. R é tal que

$$\sum_0^{R-1} |X(k)|^2 = 0,85 \sum_0^{N-1} |X(k)|^2. \quad (5)$$

O treinamento da rede neuronal foi realizado com o auxílio do *Neural Network Pattern Recognition Toolbox* [10] do MATLAB[®], que gera uma *Feed-Forward Neural Network* com função de ativação *Sigmoide*.

RESULTADOS

Para construir um sistema que discrimine gêneros musicais, foi separado e organizado um conjunto de peças de gêneros musicais com timbres diferentes. Os gêneros utilizados foram o clássico, essencialmente com peças para piano executadas



por diferentes músicos, a MPB, basicamente contendo composições de samba e bossa nova, e o rock e heavy metal, com músicas de diferentes artistas e bandas internacionais. Após o treinamento de rede com mais de 30 arquivos para cada gênero, a matriz de confusão gerada após o treinamento é mostrada na Fig. 1. As classes 1, 2 e 3 na Fig. 1 representam Clássico, MPB e Rock, respectivamente.

	1	2	3	
1	8480 39.4%	205 1.0%	31 0.1%	97.3% 2.7%
2	29 0.1%	6683 31.1%	210 1.0%	96.5% 3.5%
3	1 0.0%	113 0.5%	5760 26.8%	98.1% 1.9%
	99.6% 0.4%	95.5% 4.5%	96.0% 4.0%	97.3% 2.7%
	1	2	3	

Figura 1: Matriz de Confusão gerada após treinamento da rede neuronal de classificação de gêneros.

A seguir, foi realizado teste para avaliar a capacidade de discriminação do sistema, utilizando músicas diferentes daquelas utilizadas para o treinamento da rede, mas atribuídas aos mesmos gêneros musicais. O resultado pode ser visto na matriz de confusão na Fig. 2, na qual a coluna vertical representa o gênero que a música de fato pertence e a linha horizontal representa o gênero apontado pelo sistema.

	Clássica	MPB	Rock	Total
Clássica	28	2	0	30
MPB	2	36	0	38
Rock	0	0	33	33

Figura 2: Matriz confusão do teste de validação do sistema de reconhecimento de gêneros musicais.

O mesmo método foi realizado utilizando gêneros musicais brasileiros. Os gêneros foram separados em MPB (Bossa Nova, Samba e Choro), Forró, Frevo e Rock Nacional. A matriz de confusão após o treinamento da rede apresentou erro de 21,8%.

DISCUSSÃO

A Figura 1 representa a matriz de confusão relacionada às porções (janelas) de músicas utilizadas para o treinamento da rede. O erro associado ao treinamento foi de $\varepsilon = 2,7\%$. A Figura 2 representa o teste da rede efetuado com músicas não utilizadas no treinamento da rede. Os erros para o gênero Clássico, MPB e Rock foram $\varepsilon_{Clássica} = 6,7\%$, $\varepsilon_{MPB} = 5,3\%$ e $\varepsilon_{Rock} = 0\%$, respectivamente. Por possuir guitarras distorcidas e percussão mais agressiva, o gênero Rock possui um timbre bastante característico, sendo mais difícil a ocorrência de erros de classificação. Já a MPB, por possuir algumas peças mais lentas com acompanhamento apenas de piano, é susceptível a ser confundida com o Gênero Clássico e vice-versa, já que algumas peças clássicas podem possuir uma textura se aproximando dos acordes de piano presentes na Bossa Nova. De forma geral, a taxa de erro razoavelmente baixa obtida é possível devido a considerável diferença existente entre os gêneros em termos de timbres. Em relação ao segundo sistema de classificação com gêneros exclusivamente brasileiros, observou-se um desempenho



bastante inferior em relação ao primeiro sistema, com erro total de 21,8% no treinamento da rede. Este resultado pode ser justificado pela grande diversidade instrumental entre os gêneros brasileiros, o que torna a modelagem através exclusivamente do timbre mais complicada. De fato, gêneros como frevo, forró, rock, samba e bossa nova possuem instrumentos mais frequentes, no entanto existe um frequente intercâmbio entre instrumentos de diferentes gêneros. Dentro dessa grande pluralidade na música brasileira, pode-se destacar que a estrutura rítmica dos gêneros brasileiros é bastante marcante e bem definida. Portanto, uma possível melhoria no desempenho para um sistema de classificação em gêneros musicais voltado para a música nacional pode ser alcançado através da extração de informações que caracterizem o ritmo, juntamente com a análise de timbre efetuada.

CONCLUSÕES

Pelos resultados obtidos no trabalho, verifica-se que a característica *Mel Frequency Cepstral Coefficient*, juntamente com outras informações que modelam o timbre global da música, possibilita a distinção entre gêneros musicais cujos timbres são suficientemente distintos. A viabilidade destes parâmetros na classificação automática foi confirmada, obtendo taxas de acerto aceitáveis em um sistema prospectivo. Para a diferenciação entre subgêneros da MPB (com texturas das músicas mais próximas), é necessário ir além da análise de timbre. Um possível exemplo é elucidar características que representem a estrutura rítmica da música, através das *Wavelets* [11]. Sugere-se para trabalhos futuros a inclusão de características rítmicas como parâmetros para classificação dos gêneros musicais juntamente com as características de timbre.

AGRADECIMENTOS

Às agências CNPq, pela bolsa de iniciação científica, e CAPES.

REFERÊNCIAS

- [1] N. Scaringella, G. Zoia, D. Mlynek,. Automatic genre classification of music content: a survey, Lausanne, Switzerland, 2005.
- [2] G. Tzanetakis, G. Essl, P. Cook. Automatic Musical Genre Classification of Audio Signals. IEEE Transactions on Speech and Audio Processing, vol. 10, July 2002.
- [3] E. Scheirer, M. Slaney, Construction and evaluation of a robust multifeature speech/music discriminator, Acoustics, Speech, and Signal Processing, ICASSP-97, vol.2, pp.1331–1334, 1997.
- [4] E. Scheirer, Music-Listening Systems, PhD thesis, MIT, 2000.
- [5] B. Logan. Mel Frequency Cepstral Coefficients for Music Modeling. Int. Symp. on Music Information Retrieval, 2000.
- [6] J-J. Aucouturier, F. Pachet. Finding songs that sound the same? Paris, France, 2000.
- [7] P. Scott. Music Classification using Neural Networks. Stanford University, 2001.
- [8] P.S.R. Diniz, E.A.B. da Silva, S.L. Netto, Digital Signal Processing: System Analysis and Design, Cambridge University Press, 632 p., 2002.
- [9] D. Niewiadomy, A. Pelikant, Implementation of MFCC vector generation in classification context, J. Applied Computer Sci., Vol. 16, N.2, pp.55-65, 2008.
- [10] <http://www.mathworks.com/products/neural-network/> acesso em 13/ago/2012.
- [11] H.M. de Oliveira, Análise de Sinais para Engenheiros: Uma abordagem via wavelets, Rio de Janeiro: Brasport, 2007.