# Reconhecimento automático de gêneros musicais utilizando classificadores baseados em múltiplas características

Luiz Eduardo Romano<sup>1</sup> André Gustavo Adami<sup>1</sup>

Resumo: O reconhecimento de gêneros musicais é um importante problema de pesquisa que tem recebido muita atenção de pesquisadores e profissionais de música. Os benefícios dessa área podem ser aplicados em diversas tarefas, como na organização de bases de áudios digitais, na construção de novos mecanismos de buscas e recomendações de músicas. Este artigo propõe um método para reconhecimento automático de gêneros musicais baseado na variação da arquitetura padrão dos sistemas de reconhecimento de gêneros musicais, envolvendo técnicas como extração por segmentos, sistemas multiexpert e combinação de classificadores (paramétricos e não paramétricos). O reconhecedor foi avaliado em uma base de gêneros musicais própria extraída de um website de recomendações musicais e na base GTZAN, com dez gêneros musicais cada, onde obteve uma taxa de reconhecimento de 46,2% e 75,1%, respectivamente.

*Palavras-chave:* Processamento de sinais. Reconhecimento de gêneros musicais. Reconhecimento estatístico de padrões.

**Abstract:** Automatic music genre recognition is an important research topic that has received much attention from researchers and digital music professionals. The benefits of such field can be applied to various tasks, such as in the organization of digital audio databases, in building new search engines and music recommendations. This paper proposes a method to automatically recognize musical genres based on a variation of the standard architecture of music genre recognize systems that combines extraction per segment, multiexpert systems and ensemble of classifiers (parametric and non-parametric). The recognizer was evaluated on an experimental built from a music recommendation service website and on the GTZAN music genre databases, with 10 musical genres each, achieving a recognition rate of 46.2% and 75.1%, respectively.

Keywords: Music genre recognition. Signal processing. Statistical pattern recognition.

## 1 Introdução

Nos últimos anos, por meio da expansão da internet e do desenvolvimento da informação e das tecnologias multimídia, a música digital cresceu consideravelmente. Esse tipo de informação está se tornando cada vez mais disponível às pessoas a partir de diversas fontes de mídia [1]. Essa expansão da música digital trouxe à tona a necessidade de desenvolvimento de ferramentas capazes de extrair informações relevantes do áudio para gerenciar essas coleções [2]. Muitas informações podem ser extraídas de uma música, inclusive o seu gênero, que é provavelmente o mais popular descritor de conteúdo de músicas [3]. Isso se deve, principalmente, ao fato de que o gênero permite agrupar músicas parecidas e de que as pessoas só se interessam por certos tipos de música.

Um classificador automático de gêneros musicais possui diversas aplicações. Permite a organização de grandes coleções de músicas, como em bibliotecas, na internet, em rádios, entre outros [4]. Esse tipo de

http://dx.doi.org/10.5335/rbca.2015.4281

<sup>&</sup>lt;sup>1</sup> Curso de Bacharelado em Ciência da Computação, UCS, Rua Francisco Getúlio Vargas, 1130 - Caxias do Sul (RS) - Brasil. {ler.luizeduardo@gmail.com, agadami@ucs.br}

ferramenta pode trazer outros benefícios, como o auxílio na criação de bases de dados de músicas, criação de listas de reprodução por gênero, além de fornecer novos mecanismos de buscas e recomendações de músicas [5].

Neste trabalho, propomos a criação de um sistema de reconhecimento de gêneros musicais, no qual informações do conteúdo do áudio são extraídas em segmentos da música. Em cada segmento analisado há um conjunto de classificadores, que são posteriormente combinados para a escolha do gênero da canção. Para verificar a eficácia do sistema construído, este foi testado em duas bases de músicas: em uma base experimental, construída pelos autores deste trabalho, e na base GTZAN Genre Collection², comumente utilizada em trabalhos de reconhecimento de gêneros musicais.

Este texto está dividido em cinco seções. A Seção 2 aborda os princípios e trabalhos correlatos de reconhecimento de gêneros. A Seção 3 apresenta o método proposto para reconhecimento de gêneros musicais. A Seção 4 mostra os experimentos realizados e resultados obtidos. Por fim, a Seção 5 apresenta as conclusões deste trabalho.

# 2 Reconhecimento de gêneros musicais

O reconhecimento de gêneros musicais faz parte da tarefa da Recuperação de Informações Musicais (Music Information Retrieval – MIR). A MIR é uma emergente área de estudo direcionada para atender às necessidades dos usuários de músicas, cobrindo diferentes aspectos relacionados a gerenciamento, fácil acesso e utilização de músicas [6]. O objetivo dos sistemas de reconhecimento de gêneros musicais é categorizar um sinal de áudio de gênero desconhecido em um gênero musical previamente conhecido a partir das características desse áudio. Para isso, é necessário proceder a uma extração de características relevantes do som e fazer uso dessas informações, para que, mediante uma análise computacional, seja possível identificar a que gênero musical pertence o sinal analisado [5,7].

Existem propriedades e métodos comuns para se construir um sistema de reconhecimento de gêneros musicais. Esse processo de construção se divide, basicamente, em quatro partes: (1) coleta da base de dados, (2) extração de características, (3) aprendizado de máquina e (4) avaliação dos resultados obtidos [8]. A extração de características se preocupa em obter dados da entrada, enquanto que o aprendizado de máquina procura encontrar combinações e padrões com base nos dados obtidos na extração [5].

## 2.1 Trabalhos correlatos

O desempenho humano em classificar gêneros musicais foi analisado no trabalho de Gjerdingen e Perrott [9]. Nesse trabalho, um grupo de estudantes analisou e classificou diversas músicas de dez gêneros musicais diferentes. A taxa de acerto foi de aproximadamente 70% para trechos de 3 segundos, e de aproximadamente 44% para trechos de música de até 250 milissegundos.

Na análise de classificação automática de gêneros musicais usando computadores, um dos primeiros trabalhos mais importantes publicados na área foi o de Tzanetakis e Cook [10]. Os autores propuseram um sistema baseado em características como timbre, ritmo e pitch que obteve uma taxa de acerto de aproximadamente 61% para dez gêneros. No mesmo ano, Kosina desenvolveu em [5] um sistema de classificação de gêneros musicais, chamado MUGRAT, obtendo aproximadamente 88% de reconhecimento para três gêneros musicais. Silla et al. [11] realizaram uma análise diferente na qual, em vez de extraírem informações de toda a música, dividiram o áudio em três partes (começo, meio e fim, com trechos de 30 segundos cada), utilizando classificadores diferentes para cada parte e combinando os resultados. Nesse trabalho, foi alcançada uma taxa média de acerto 3% maior do que o melhor resultado obtido individualmente (55,15% sem a combinação contra 58,07%, utilizando três segmentos, para dez gêneros musicais). Paradzinets et al. [12] desenvolveram um sistema no qual as características do áudio foram extraídas e agrupadas em três grupos (acústicas, rítmicas e timbre), tendo cada grupo sido analisado por um classificador diferente, que retornava as probabilidades de a música pertencer a cada gênero. Nesse trabalho também foi utilizado o conceito de sistemas multiexpert, em que a saída de cada classificador foi utilizada como um novo dado de entrada para um novo classificador, que combinava os resultados e definia o gênero da música. Por meio dessa combinação, os autores constataram um aumento de 12% no reconhecimento em comparação ao desempenho do melhor classificador individual (54,6% sem a combinação e 66,7% utilizando grupos de características, para seis gêneros musicais).

Revista Brasileira de Computação Aplicada (ISSN 2176-6649), Passo Fundo, v. 7, n. 1, p. 85-99, abr. 2015 86

<sup>&</sup>lt;sup>2</sup> http://marsyas.info/download/data\_sets/

Alternativamente, diversas formas de construção de um sistema de reconhecimento de gêneros musicais foram apresentadas. Lampropoulos *et al.* [13] separaram o sinal do áudio em várias partes (representando instrumentos de corda, sopro e percussão) e extraíram características de cada uma dessas partes. Costa *et al.* [14] propuseram um método que, em vez de extrair os dados diretamente do sinal do áudio, levou a que os dados fossem extraídos de uma imagem do sinal, utilizando processamento de imagem. Ariyaratne e Zhang [15] desenvolveram um modelo de sistema hierárquico, no qual os gêneros foram divididos em grupos menores e organizados de forma hierárquica. Os dados do áudio, então, foram processados, e o classificador decidia para qual grupo o dado era mais similar, e esse processo se repetia para todos os subgrupos dentro do grupo escolhido, até o classificador decidir o gênero final.

# 3 Sistema proposto

Diferentemente da maioria dos sistemas de reconhecimento de gêneros apresentados, que dividem basicamente o seu funcionamento em extração de características e classificação, a ideia do sistema proposto é de utilizar diferentes técnicas em um único modelo de sistema para melhorar o reconhecimento de gêneros. Dentre as técnicas utilizadas, pode-se citar a extração por segmentos, sistemas multiexpert e combinação de classificadores.

A arquitetura do sistema proposto procura fazer uma mescla dos modelos de extração por segmentos com o sistema multiexpert. Essa arquitetura se divide em quatro etapas, conforme mostrado na Figura 1: segmentação do áudio, extração de características, classificação por grupo e definição do gênero.

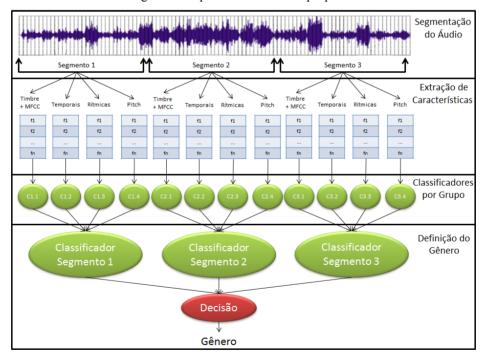


Figura 1: Arquitetura do sistema proposto.

## 3.1 Segmentação do áudio

Nessa etapa, a música é dividida em três segmentos: início, meio e fim da canção. A análise de toda uma música representa um custo de processamento elevado e um grande volume de dados abrangidos e envolve o processamento de informações repetidas, já que a maioria das músicas possui certa estrutura musical que se repete. Portanto, a análise de segmentos evita todo o processamento da música, mas mantendo a qualidade dos dados extraídos. É realizada a análise de três segmentos em pontos diferentes da música, para que, no caso de algum segmento não representar corretamente o gênero da música, os outros compensem aquele segmento. A análise de segmentos é uma técnica comumente utilizada em trabalhos de reconhecimento de gêneros [11,16,17].

#### 3.2 Extração de características

Nessa etapa, para cada segmento, características são extraídas e organizadas em grupos, registrando-se que cada um apresenta um classificador próprio, que realiza o processamento somente daquelas características. A ideia da classificação por grupo de características é de que algumas características, pelo tipo de informação que representam, são mais discriminatórias para certos gêneros do que para outros. Além disso, o agrupamento reduz a dimensionalidade dos dados para a classificação, reduzindo a complexidade dos modelos gerados [12,18,19]. No sistema proposto, é realizado o agrupamento das características em quatro grupos: timbre e coeficientes mel-cepstrais, temporais, rítmicas e pitch.

#### 3.2.1 Relacionadas ao timbre e aos coeficientes mel-cepstrais

Essas características incluem informações de baixo nível, que são analisadas em pequenas janelas de áudio. As características analisadas para esse grupo são:

 Spectral Centroid: é o ponto balanceado do espectro, uma medida associada, frequentemente, com a noção do brilho espectral, sendo calculada pela equação

$$C_{t} = \frac{\sum_{n=1}^{N} X_{t}[n] * n}{\sum_{n=1}^{N} X_{t}[n]}$$
(1)

onde  $X_t$  [n] é o valor da transformada de Fourier no quadro t e na faixa de frequência n. É uma característica utilizada para discriminar instrumentos musicais.

• Spectral Rolloff: é o ponto  $R_t$ , onde a frequência que está abaixo de alguma porcentagem  $P_o$  (geralmente 85%) da energia do espectro reside. O ponto  $R_t$  é estimado de acordo com a equação

$$\sum_{n=1}^{R_t} X_t[n] = P_o * \sum_{n=1}^{N} X_t[n], \qquad 0 \le P_o \le 1$$
 (2)

onde  $X_t[n]$  é a magnitude da transformada de Fourier no quadro t e na faixa de frequência n. O spectral rolloff é outra forma de medir a frequência média do espectro e tem objetivos similares ao spectral centroid.

• Spectral Flux: é a diferença entre duas amplitudes normalizadas de sucessivas distribuições espectrais, utilizada para calcular a quantidade de mudanças no espectro no decorrer do tempo. É definida pela equação

$$F_{t} = \sum_{n=1}^{N} (X_{t}[n] - X_{t-1}[n])^{2}$$
(3)

onde  $X_t[n]$  e  $X_{t-1}[n]$  representam magnitudes do espectrograma de frequência n nos tempos t e t-1. É uma característica que indica as variações na música (se é uma música constante, ou se apresenta mudanças bruscas).

• **Zero-Crossing Rate:** esta taxa é calculada contando o número de vezes que a onda do áudio cruza o eixo x para cada unidade de tempo. É estimada por meio da equação

$$Z_{t} = \frac{1}{2} \sum_{n=1}^{N} \left| sign(x[n]) - sign(x[n-1]) \right|$$

$$\tag{4}$$

onde a função sign() retorna o valor 0 (zero) para valores negativos e o valor 1 (um) para valores positivos, e x[n] denota o domínio do sinal no tempo t. Essa característica é sensível aos vocais e instrumentos de percussão, onde valores altos indicam maior presença desses sons.

• Coeficientes Cepstrais da Frequência Mel (Mel Frequency Cepstral Coefficients – MFCC): proposto originalmente em [20], são uma representação compacta do espectro do sinal do áudio, com o objetivo de aproximar a distribuição não linear da largura de banda da audição humana com frequência. Para o cálculo do MFCC, inicialmente o áudio é dividido em janelas, e para cada uma é aplicada a transformada de Fourier discreta no sinal. Em seguida, é definido um conjunto de filtros, em que, geralmente, filtros lineares são aplicados nas frequências baixas e filtros logarítmicos nas frequências altas. Após, é calculado o logaritmo da magnitude do espectro, e os valores das faixas pré-determinadas são convertidos em novos valores, utilizando o conjunto de filtros. Em seguida, é aplicada uma transformação de cosseno discreta nas saídas dos filtros. Os coeficientes são obtidos diretamente da saída da transformação.

# 3.2.2 Temporais

As características temporais tem como objetivo capturar a evolução temporal do sinal. Nesse grupo, é formado um vetor de características por segmento, utilizando principalmente momentos estatísticos. As características que constituem esse grupo são a média e a variância das características relacionadas ao timbre (spectral centroid, spectral rolloff, spectral flux, zero-crossing rate) e MFCC, além do low energy. O low energy representa a porcentagem de janelas que têm menos energia do que a energia média de todas as janelas. Músicas com partes silenciosas ou muito calmas terão um low energy maior, por exemplo.

#### 3.2.3 Rítmicas

Esse grupo é constituído por informações relacionadas à batida e à estrutura rítmica de uma música. Para o cálculo dessas informações, foi utilizado o método proposto por Tzanetakis e Cook [10], em que é gerado um histograma da batida. Inicialmente, o sinal é analisado em janelas geralmente grandes (aproximadamente 4 segundos), para capturar uma estrutura rítmica completa. Para cada janela, o sinal é decomposto em um número de bandas de frequência utilizando uma transformada discreta, geralmente a Transformada Discreta Wavelet. Após, para cada banda, a amplitude no domínio do tempo é extraída realizando retificação de onda completa, filtragem passa-baixa, diminuição da taxa de amostragem e remoção de médias. Em seguida, as bandas são somadas, reconstruindo um novo sinal com as bandas modificadas. Após, é aplicada uma autocorrelação que resulta nas periodicidades da batida. Depois, o resultado da autocorrelação é reforçado para reduzir o efeito de inteiros múltiplos das periodicidades. Ao final dos cálculos, tem-se as forças das batidas em várias velocidades na janela de áudio analisada. Em seguida, é feita a junção das forças das batidas de cada janela para formar o histograma. Esse grupo produz um vetor de características de vinte dimensões estimado a partir de diversas informações estatísticas do histograma: soma e variância do histograma, amplitudes, períodos e larguras dos dois maiores picos, distância entre os dois maiores picos, quantidade total de picos, quantidade de picos com amplitude acima da média, distância média entre os picos, média das amplitudes, quantidade de posições sem picos, diferença da amplitude entre os dois maiores picos, quantidade de grupos fechados (sequência contínua de períodos com picos), amplitude relativa dos dois maiores picos em relação à soma do histograma, diferença relativa entre as amplitudes e distâncias dos dois maiores picos.

#### 3.2.4 Pitch

Pitch é a percepção da frequência fundamental de uma nota musical, onde, dado um determinado trecho de música polifônica (múltiplas notas tocadas simultaneamente), esse é utilizado para determinar qual nota é a que recebe mais destaque e é percebida por uma pessoa. Para o uso do pitch neste trabalho, foi utilizada a montagem de um histograma do pitch, que faz uso de um algoritmo de detecção de múltiplos pitch desenvolvido por Tolonen e Karjalainen [21]. Primeiro, o sinal é dividido em pequenas janelas, geralmente de 100 ms ou menos. Para cada janela, o sinal é filtrado para remover a correlação do sinal. Em seguida, o sinal é decomposto em duas bandas de frequência (abaixo e acima de 1kHz), e envelopes de amplitude são extraídos para a banda alta, aplicando retificação de meia onda e filtragem passa-baixa. Após, é calculada a detecção de periodicidade de cada banda por meio de uma transformada de Fourier discreta, compressão da magnitude da representação espectral, e uma transformada inversa. As bandas, então, são unidas, e, em seguida, a autocorrelação das bandas somadas é estimada. Finalmente, o sinal é reforçado para reduzir o efeito de múltiplos inteiros no pico das frequências (método similar ao utilizado na construção do histograma rítmico). Assim, para cada janela, tem-se a força do pitch nas mais diversas frequências. Após, as forças de cada janela são unidas para formar o histograma. Semelhante ao grupo das características rítmicas, as mesmas características são extraídas do histograma de pitch.

#### 3.3 Classificadores por grupo

Seguindo a ideia da organização das características em grupos, nesta etapa, para cada um dos quatro grupos de cada segmento, há um classificador próprio (representado, na Figura 1, pelos classificadores C1.1 a C1.4 para o primeiro segmento, C2.1 a C2.4 para o segundo, e C3.1 a C3.4), que analisa somente as características daquele grupo. Nessa fase, foi testado o uso do classificador paramétrico Gaussian Mixture Model (GMM) e do não paramétrico Support Vector Machines (SVM) [22]. O SVM é restrito para a classificação de duas classes, portanto, para torná-lo apto a resolver problemas com mais de duas classes, foi utilizada a abordagem do "um contra todos". Nessa abordagem, para x classes, são feitas x classificações, e em cada classificação uma classe diferente é comparada com todas as outras. No GMM, para cada classificador, há um GMM para cada classe, modelando os dados pertinentes ao gênero.

#### 3.3.1 Supervetor

Comumente utilizado em trabalhos de reconhecimento de imagens e de locutor [23] [24], cada gênero musical é modelado através de um SVM estimado a partir de supervetores extraídos de um GMM. Essa técnica baseia-se na ideia de que o supervetor é um mapeamento entre o sinal de áudio e um vetor de grande dimensão [24]. Além disso, como cada áudio ou segmento produz um único supervetor, a redução na quantidade de amostras permite a possibilidade da estimação do SVM, já que a análise convencional de sinais produz muitos vetores que tornariam essa estimação proibitiva computacionalmente.

O supervetor é obtido por meio da concatenação das médias dos componentes gaussianos do modelo. Na implementação deste trabalho, supervetores foram utilizados somente para o grupo de características relacionadas ao timbre e MFCC, da seguinte forma: para cada música, em cada segmento, os dados relacionados ao timbre e MFCC foram extraídos normalmente. Após, esses dados foram passados para o GMM, que modela aqueles dados daquela música naquele segmento. Em seguida, as médias dos componentes das misturas foram concatenadas, tornando-se as novas características que foram passadas para o SVM, conforme representado na equação

$$\boldsymbol{x} = [\boldsymbol{m}_1, \boldsymbol{m}_2, \dots, \boldsymbol{m}_N] \tag{5}$$

onde x são os novos dados utilizados no SVM,  $m_1$ ,  $m_2$ ,...,  $m_N$  são as médias dos componentes do GMM e N=MC, para M igual ao número de componentes gaussianas e C, a quantidade de características. Assim, para cada música analisada, em cada segmento, tem-se uma amostra de características.

#### 3.3.2 Saída dos classificadores

A saída dos classificadores dessa fase consiste em probabilidades de o vetor de entrada pertencer a cada um dos gêneros envolvidos. Logo, para x gêneros analisados, a saída de cada classificador é um vetor de x dimensões. Os vetores de cada grupo são normalizados entre zero e um (para padronizar a faixa de valores das probabilidades, independentemente do tipo de classificador utilizado) e, após, unidos, formando um vetor de 4x dimensões, que é passado para o classificador principal do segmento.

Para cada classificador, foi utilizada uma abordagem diferente para capturar a probabilidade de as amostras pertencerem a cada um dos gêneros. Para o GMM, foi calculada a função de densidade de probabilidade de cada classe. Diferentemente da abordagem do GMM, como o SVM não é um classificador que trabalha diretamente com probabilidades, as chances de uma amostra pertencer a uma determinada classe foram estimadas por meio da distância da amostra para o hiperplano separador. Como para o SVM foi utilizada a abordagem do "um contra todos", em um grupo de classificadores há um SVM para cada gênero (comparado com todos os demais). Cada classificador utiliza o mesmo conjunto de dados, então, a partir disso, dada uma amostra, sua distância é calculada em cada um dos classificadores. Quanto mais distante a amostra do hiperplano separador, maior a chance de aquela amostra pertencer ao gênero analisado naquele classificador. O uso da distância da amostra para o hiperplano é uma das maneiras de estimar probabilidades para o SVM, e foi originalmente proposto por Madevska-Bogdanova *et al.* [25].

## 3.4 Definição do gênero

Na última etapa do sistema proposto, é realizada a fusão dos classificadores para a escolha final do sistema. Inicialmente, os classificadores por grupo de cada segmento são combinados e, posteriormente, os

classificadores principais de cada segmento são combinados por meio de regras de decisão, resultando na escolha final do sistema para a atribuição do gênero ao áudio analisado.

Na primeira parte dessa etapa, os classificadores de cada um dos três segmentos recebem as saídas dos quatro classificadores de cada grupo de características do segmento (probabilidades de as amostras pertencerem a cada um dos gêneros). Assim como nos classificadores por grupo, o GMM e o SVM foram testados como classificadores principais de cada segmento.

Na segunda parte dessa etapa, os classificadores principais de cada um dos três segmentos também retornam novas probabilidades (calculadas conforme o tipo do classificador, como nos classificadores por grupo). As probabilidades de cada um dos três classificadores são combinadas e é definido o gênero da música. Foram testadas quatro técnicas de combinação de probabilidades, que são comumente utilizadas nos trabalhos de reconhecimento de gêneros musicais para agregar resultados de vários classificadores [11] [17] [26]:

- Maior valor: são analisadas todas as probabilidades, e a maior, independentemente do segmento de origem, é a escolhida. Em caso de empate, é escolhido um gênero aleatoriamente dentre os gêneros empatados.
- Regra da soma: as probabilidades de cada gênero oriundas de cada segmento são somadas, e o gênero
  com maior valor é escolhido. Em caso de empate, é escolhido um gênero aleatoriamente dentre os
  gêneros empatados.
- **Regra do produto:** as probabilidades de cada gênero são multiplicadas, e o gênero com o maior valor é escolhido. Em caso de empate, é escolhido um gênero aleatoriamente dentre os gêneros empatados.
- Peso do classificador: nesse método, é atribuído um peso para cada classificador, calculado durante a fase de treinamento, refletindo a taxa de reconhecimento deste. Na fase de testes, as probabilidades dos classificadores são multiplicadas pelo respectivo peso do classificador, resultando em novas probabilidades. Após, podem ser aplicadas as três regras anteriores nas novas probabilidades. Em caso de empate, é escolhido, dentre os gêneros empatados, o gênero que obteve maior valor no classificador com melhor taxa de acerto. Essa técnica permite ajustar a influência de cada classificador individual na escolha final do gênero, permitindo que um classificador com melhor taxa de acerto tenha maior influência na decisão do que um classificador que obteve desempenho inferior.

## 4 Experimentos

Os experimentos foram realizados em duas bases de dados. Uma das bases, chamada de experimental, reúne 5 mil músicas adquiridas do site Last.fm³, divididas em dez gêneros musicais (clássica, country, eletrônica, hip hop, jazz, latinas, metal, pop, reggae e rock), com quinhentas músicas cada. As músicas são completas, podendo ser instrumentais ou não, e se encontram no formato MP3, com frequência de amostragem de 44.100 Hz. A outra base utilizada, denominada de GTZAN, foi criada e utilizada por Tzanetakis e Cook em [10]. Essa base reúne mil músicas, divididas em dez gêneros (blues, clássica, country, disco, hip hop, jazz, metal, pop, reggae e rock) com cem músicas cada. Os arquivos dessa base estão em formato AU, com frequência de 22.050 Hz, e cada música tem 30 segundos de duração (segmento retirado do meio da música). Essa base foi escolhida por ser uma das mais conhecidas e utilizadas em diversos trabalhos de reconhecimento de gêneros musicais, permitindo, assim, comparar os resultados obtidos com outros métodos publicados.

A segmentação das músicas foi definida conforme a base de dados. Para a base experimental, o tamanho do segmento é de 30 segundos, totalizando 90 segundos analisados por música. Os 20 segundos iniciais e finais da música são descartados, pois esses trechos geralmente apresentam partes silenciosas e sem todos os instrumentos, podendo não representar corretamente o gênero. O segmento inicial é analisado logo após os 20 segundos iniciais. Para o segmento intermediário, é calculado o ponto que representa a metade da música, e os 15 segundos anteriores e posteriores a esse ponto são analisados. O segmento final é extraído do trecho de 30 segundos anterior aos 20 segundos finais da música. Caso a duração da canção seja inferior a 130 segundos, os segmentos são parcialmente sobrepostos, garantindo a análise do mesmo volume de dados em todas as músicas. Para a base GTZAN, que apresenta músicas de apenas 30 segundos de duração, toda a canção é analisada, com segmentos de 10 segundos cada, sem sobreposição.

<sup>&</sup>lt;sup>3</sup> www.last.fm

Para o cálculo das características relacionadas ao timbre e MFCC, foram analisadas janelas com duração de 100 ms com deslocamento de 50 ms. Para o spectral centroid, spectral rolloff e spectral flux, foi estimada a transformada de Fourier do sinal com amplitude normalizada entre 0 e 1. Para o MFCC, foram utilizados treze coeficientes. Ao final, esse grupo forma um vetor de dezessete características (treze do MFCC e quatro de informações relacionadas ao timbre). O grupo das características temporais totaliza um vetor com 35 dimensões.

Para o cálculo das características rítmicas, optou-se por janelas de 4 segundos com deslocamento de 2 segundos. Os demais parâmetros foram definidos conforme os valores definidos por Tzanetakis e Cook [10]. Na construção do histograma, foi realizada a soma das forças das batidas de todas as janelas. No total, vinte informações foram extraídas do histograma, conforme descrito na Seção 3.2.3.

Para o cálculo das características de pitch, optou-se por janelas de 100 ms de duração, com deslocamento de 100 ms. Os mesmos parâmetros propostos por Tolonen e Karjalainen [21] foram utilizados para a extração de pitch. A construção do histograma foi feita somando-se as forças do pitch nas janelas. No total, foram extraídas as mesmas vinte informações que foram obtidas do histograma rítmico.

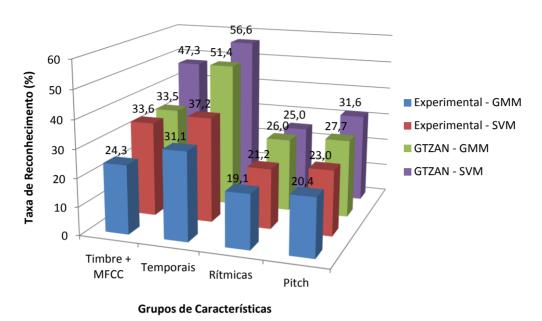
Para a definição dos parâmetros dos classificadores (GMM e SVM), foi utilizada 3-fold cross-validation nos dados de treinamento de ambas as bases para determinar a combinação de parâmetros que produzisse o melhor resultado. Para o GMM, os parâmetros testados para a construção dos modelos foram o número de componentes gaussianas e o tipo da matriz de covariância (completa ou diagonal). Os melhores parâmetros variaram para cada base: matriz de covariância completa e duas componentes gaussianas para a base GTZAN e matriz de covariância diagonal e 32 componentes gaussianas para a base experimental. Para o SVM, os parâmetros testados foram a função do kernel (linear, polinomial, quadrático, radial basis function – RBF – e multilayer perceptron – MLP), a penalidade de erro (onde foram testados os seguintes valores: 0.01, 0.1, 1 e 5) e parâmetros do kernel, quando necessário (ordem do algoritmo polinomial e valor sigma para o RBF). A melhor configuração, para ambas as bases, foi o kernel RBF, com penalidade de erro igual a 5 e valor sigma igual a 15. Os parâmetros do GMM utilizados na construção dos supervetores foram matriz de covariância completa e duas componentes gaussianas.

#### 4.1 Resultados

Para a realização dos treinamentos e testes em cada uma das bases de dados, foi utilizado 10-fold cross-validation. O teste binomial para diferenças de proporção apresentado em [27] é usado para testar se as diferenças são estatisticamente significantes. Quando não especificado, o nível de significância  $\alpha$  é igual a 0,05.

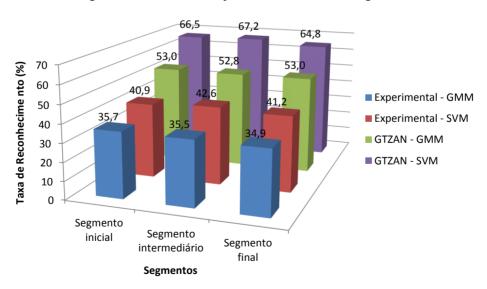
A Figura 2 mostra a taxa de acerto de cada grupo de características, quando utilizadas individualmente, em cada base de dados e classificadores. É possível perceber diferenças de qualidade entre os grupos de características. Os resultados mostram que as informações estáticas e dinâmicas (temporais) do timbre e MFCC foram as mais discriminatórias, produzindo os melhores resultados individuais. O grupo de informações temporais (média e variância do timbre e MFCC, além do low energy) obteve os melhores desempenhos individuais, com uma taxa de 37,2% na base experimental e 56,6% na base GTZAN. Em seguida, as características de timbre e MFCC em cada janela obtiveram os segundos melhores resultados, com 33,6% na base experimental e 47,3% na base GTZAN. Por fim, as características de pitch e rítmicas obtiveram as piores taxas de reconhecimento quando utilizadas individualmente (melhores resultados de 23% na base experimental e 31,6% na base GTZAN para o pitch; e taxas de 21,2% na base experimental e 26% na base GTZAN para o grupo de informações rítmicas). Os resultados obtidos corroboram os apresentados em [10,28,29], onde as diferenças encontradas foram similares.

Figura 2: Resultados da etapa de classificadores por grupos de características.



A Figura 3 mostra a taxa de reconhecimento do classificador principal do segmento, em cada base de dados e classificadores. Tanto com o uso do GMM como do SVM, em ambas as bases, as diferenças entre as taxas de reconhecimento para diferentes segmentos não foram estatisticamente significantes (35,7%, 35,5% e 34,9% para os segmentos inicial, intermediário e final, respectivamente, com o uso do GMM; e 40,9%, 42,6% e 41,2%, utilizando o SVM). Assim, não é possível afirmar qual segmento apresenta o melhor conteúdo musical para a definição do gênero da música. Isto é, todos contêm informações relevantes para a definição do gênero musical.

Figura 3: Resultados da etapa de classificadores do segmento.



A Figura 4 mostra a taxa de acerto com o uso de cada uma das regras de decisão aplicadas, definindo a taxa de reconhecimento final do sistema. A regra do maior valor para o classificador GMM apresentou resultados estatisticamente inferiores aos demais (35,7% contra 40,2% da regra da soma e do produto, na base experimental; e 53% contra 61,5% da regra da soma e 61% da regra do produto, na base GTZAN). Não houve

diferença estatisticamente significante entre os desempenhos das regras da soma e a do produto em nenhuma das bases. Utilizando SVM, os resultados foram um pouco diferentes. Em ambas as bases de dados, a regra do produto mostrou resultados estatisticamente inferiores em comparação às outras regras (39,1% contra 44,3% da regra do maior valor e 46% da regra da soma, na base experimental; e 61,1% contra 71,4% da regra do maior valor e 73,7% da regra da soma, na base GTZAN). Entre a regra da soma e a do maior valor, tanto na base experimental como na GTZAN, as diferenças entre elas não foram estatisticamente significantes. Por fim, o uso das regras de decisão com alteração das probabilidades pelo desempenho dos classificadores mostrou poucas diferenças. Em todas as regras de decisão aplicadas, em ambas as bases e classificadores, as diferenças encontradas com a alteração não foram estatisticamente significantes, não sendo possível afirmar se essa técnica melhora ou piora a taxa de reconhecimento.

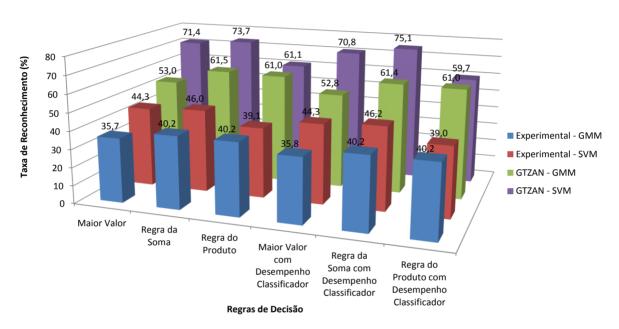
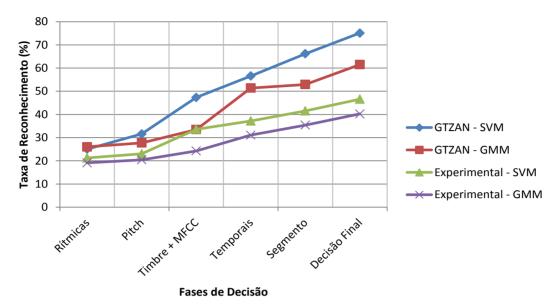


Figura 4: Resultados das regras de decisão.

Com a obtenção dos resultados em todas as etapas do sistema, percebe-se que o uso de um classificador apropriado para o reconhecimento de gêneros faz uma grande diferença. Nos testes realizados, o contraste entre as taxas de reconhecimento finais obtidas com o uso do GMM e do SVM foi grande. Em ambas as bases de dados, o SVM alcançou resultados estatisticamente superiores em relação ao GMM (46,2% contra 40,2% na base experimental, e 75,1% contra 61,5% na base GTZAN), como mostra a Figura 4.

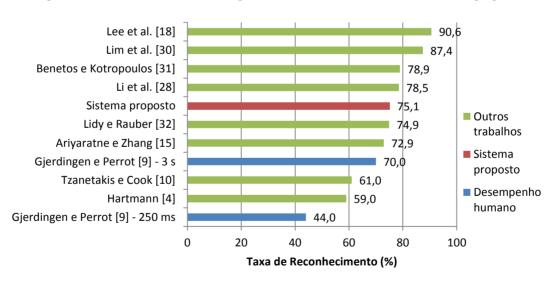
Um fato a ser considerado sobre o sistema proposto foi a evolução da taxa de reconhecimento ao longo de todas as etapas do sistema. O sistema proposto possui a divisão da classificação em três etapas principais: classificadores por grupo (separado por grupo de características), classificadores por segmento (com a junção dos classificadores por grupo de cada segmento) e decisão final (com a análise das probabilidades do classificador principal de cada um dos três segmentos, por meio das regras de decisão para a escolha final do sistema). Em todas as bases, em ambos os classificadores, o sistema mostrou evolução na taxa de reconhecimento em cada etapa, conforme mostrado na Figura 5. Isso evidencia a informação complementar que existe entre as características e os segmentos no reconhecimento dos gêneros.

Figura 5: Evolução da taxa de reconhecimento ao longo das etapas do sistema.



A Figura 6 mostra os melhores resultados obtidos em outros trabalhos de reconhecimento de gêneros musicais, utilizando todas as músicas da base de dados GTZAN, com 10-fold cross-validation. A Figura 6 também inclui o desempenho humano em reconhecer gêneros musicais, conforme reportado no trabalho de Gjerdingen e Perrot [9]. Todos os resultados apresentados representam a melhor taxa média de reconhecimento em dez gêneros musicais.

Figura 6: Taxa de reconhecimento de gêneros de trabalhos correlatos e do sistema proposto.



A fim de analisar o reconhecimento do sistema nos diferentes gêneros, foi utilizada uma matriz de confusão para verificar que tipos de erro o sistema produz. A Tabela 1 e a Tabela 2 mostram a matriz de confusão do sistema baseado no SVM para as bases GTZAN e experimental, respectivamente, nos seguintes gêneros: BLU – blues, CLA – clássica, COU – country, ELE – eletrônica, HIP – hip hop, JAZ – jazz, LAT – latinas, MET – metal, POP – pop, REG – reggae, ROC – rock.

Tabela 1: Matriz de confusão da base GTZAN utilizando o classificador SVM.

	BLU	CLA	COU	ELE	HIP	JAZ	MET	POP	REG	ROC
BLU	72	0	8	2	0	3	5	0	3	7
CLA	2	89	2	0	0	2	0	0	0	5
COU	9	0	65	6	0	0	3	3	4	10
ELE	3	1	5	73	2	0	3	7	1	5
HIP	2	0	1	4	74	0	3	4	12	0
JAZ	4	6	2	0	1	82	1	2	0	2
MET	1	0	0	6	0	0	87	0	0	6
POP	0	1	5	4	5	1	1	71	2	10
REG	5	1	6	2	7	1	0	6	69	3
ROC	4	1	6	8	1	3	2	0	6	69

Tabela 2: Matriz de confusão da base experimental utilizando o classificador SVM.

	CLA	COU	ELE	HIP	JAZ	LAT	MET	POP	REG	ROC
CLA	348	30	23	2	52	15	9	12	3	6
COU	16	226	15	11	40	40	9	57	35	51
ELE	41	19	215	33	27	29	16	65	31	24
HIP	0	20	41	282	17	30	5	20	81	4
JAZ	54	35	26	20	214	73	5	29	17	27
LAT	32	72	39	37	61	108	12	51	62	26
MET	3	11	11	0	3	2	387	8	1	74
POP	17	67	70	18	27	40	12	139	30	80
REG	5	31	40	77	18	41	12	29	208	39
ROC	7	72	21	5	15	12	71	63	32	202

Percebe-se que alguns gêneros são naturalmente mais distinguíveis do que outros, como é o caso da música clássica e de metal, que apresentaram as maiores taxas de reconhecimento em ambas as bases. De fato, esses são dois gêneros com características bem peculiares. A música clássica apresenta batidas leves e instrumentos que dificilmente aparecem em outros gêneros, como piano, violino e órgão; enquanto que o Metal apresenta ritmos mais rápidos e uma sonoridade mais grave, o que não é encontrado em outros gêneros [5]. O mesmo não acontece, por exemplo, com os gêneros pop e rock, os quais apresentaram baixas taxas de reconhecimento em ambas as bases.

As matrizes de confusão mostram que o sistema esteve no caminho correto para a classificação das músicas, e grande parte das confusões foram entre gêneros que de fato são muito parecidos. A música clássica, por exemplo, apresenta certas semelhanças com o jazz (ambas revelam ritmos mais lentos e fazem uso de tonalidades mais altas) [10], e grande parte das confusões da música clássica deu-se com o jazz, e vice-versa. O rock, por ter sido originado da combinação de diversos gêneros como blues e country, guarda forte similaridade instrumental e rítmica com esses gêneros [26], havendo bastante confusão entre eles. Portanto, nota-se que o sistema consegue encontrar características diferenciais entre os gêneros, mas é afetado pelo fato de que algumas informações são comuns entre as classes. Essas confusões entre os gêneros foram similares às encontradas em [10,26,31].

## 5 Conclusões

O sistema proposto neste trabalho está focado na modificação da arquitetura padrão dos sistemas de reconhecimento de gêneros, dividindo a tarefa de identificar um gênero em várias etapas. Inicialmente, as músicas são analisadas em três segmentos (início, meio e fim da música). Para cada segmento, características são extraídas e unidas em quatro grupos principais, conforme o tipo dessas características. Classificadores são utilizados para cada um desses grupos, e, em seguida, as probabilidades desses classificadores são combinadas para classificadores em cada segmento. Após, as probabilidades de cada gênero em cada segmento são analisadas, e é decidido o gênero da música. O sistema obteve uma taxa de reconhecimento para dez gêneros de 46,2% e 75,1% em uma base experimental e na GTZAN, respectivamente. O resultado do sistema na base

GTZAN é superior ao desempenho obtido por Tzanetakis e Cook [10], o qual obteve uma taxa de reconhecimento de 61% nas mesmas condições de treinamento e teste.

Ao longo do desenvolvimento deste trabalho, várias informações foram obtidas quanto ao reconhecimento de gêneros musicais:

- A divisão da música em segmentos mostra-se uma boa forma de analisar o gênero de uma música. A
  análise de um segmento não necessariamente aumenta a taxa de reconhecimento, mas evita o
  processamento completo de uma música sem diminuir o reconhecimento.
- A separação de características por grupos com base no seu tipo ajuda no reconhecimento. Conforme reportado, a análise da junção das saídas dos classificadores de cada grupo mostrou resultados superiores em relação aos resultados individuais de cada grupo.
- Todas as características utilizadas auxiliaram no reconhecimento dos gêneros. Porém, algumas dessas informações obtiveram melhores resultados, como as informações de timbre e MFCC, especialmente quando utilizadas de forma temporal. As informações rítmicas e de pitch apresentaram resultados inferiores.
- Na classificação, a escolha do algoritmo tem um papel importante no reconhecimento do sistema. Neste trabalho, o SVM apresentou melhores taxas de reconhecimento do que o GMM.
- A combinação de classificadores aumenta consideravelmente a taxa de acerto. A análise de probabilidades de vários classificadores e a decisão a partir desses valores contribuem para um melhor reconhecimento. Nas regras de combinação utilizadas, a regra da soma com alteração de probabilidades pelo desempenho dos classificadores se mostrou a melhor alternativa para unir as informações de diferentes classificadores.

Para futuros trabalhos, será continuado o uso da arquitetura proposta, com algumas modificações. O SVM continuará a ser utilizado, e poderá ser feito uso de outro classificador não paramétrico. As características permanecerão sendo usadas em grupos, porém novas informações do áudio poderão ser utilizadas. Novos algoritmos de detecção rítmica e de pitch serão analisados para verificar se provêm informações mais precisas sobre o conteúdo musical processado. Também será analisado o uso de algoritmos de seleção de características, com o objetivo de reduzir a quantidade de informações utilizadas sem prejudicar o reconhecimento.

#### Referências

- [1] FU, Z. et al. A Survey of Audio-Based Music Classification and Annotation. *IEEE Transactions on Multimedia*, v. 13, n. 2, p. 303-319, 2011. ISSN 1520-9210. Disponível em: <a href="http://dx.doi.org/10.1109/TMM.2010.2098858">http://dx.doi.org/10.1109/TMM.2010.2098858</a>>. Acesso em: 14 set. 2014.
- [2] PANAGAKIS, I.; BENETOS, E.; KOTROPOULOS, C. Music Genre Classification: A Multilinear Approach. In: INTERNATIONAL SYMPOSIUM MUSIC INFORMATION RETRIEVAL. Philadelphia: ISMIR, 2008. p. 583-588. Disponível em: <a href="http://openaccess.city.ac.uk/2109/">http://openaccess.city.ac.uk/2109/</a>. Acesso em: 14 set. 2014.
- [3] AUCOUNTURIER, J. J.; PACHET, F. Representing Musical Genre: A State of the Art. *Journal of New Music Research*, v. 32, n. 1, p. 83-93, 2003. ISSN 0929-8215. Disponível em: <a href="http://www.csl.sony.fr/downloads/papers/uploads/pachet-02c.pdf">http://www.csl.sony.fr/downloads/papers/uploads/pachet-02c.pdf</a>>. Acesso em: 14 set. 2014.
- [4] HARTMANN, A. *Testing a Spectral-Based Feature Set for Audio Genre Classification*. Tese (Mestrado em Humanidades) Music Department, University of Jyväskylä, 2011.
- [5] KOSINA, K. *Music Genre Recognition*. Tese (Trabalho de Conclusão em Mídia e Tecnologia) Technical College of Hagenberg, Austria, Hagenberg, 2002.
- [6] ORIO, N. Music Retrieval: A Tutorial and Review. *Foundations and Trends in Information Retrieval*, v. 1, n. 1, p. 1-90, 2006. Disponível em: <a href="http://dx.doi.org/10.1561/1500000002">http://dx.doi.org/10.1561/1500000002</a>>. Acesso em: 14 set. 2014.
- [7] BARREIRA, L. F. M. *Unsupervised Automatic Music Genre Classification*. Tese (Mestrado em Engenharia Informática) Universidade Nova de Lisboa, Lisboa, 2010.

- [8] TERMENS, E. G. Audio Content Processing for Automatic Music Genre Classification: Descriptors, Databases, and Classifiers. Tese (Doutorado em Ciência da Computação e Comunicação Digital) Universitat Pompeu Fabra, Barcelona, 2009.
- [9] GJERDINGEN, R. O.; PERROTT, D. Scanning the Dial: The Rapid Recognition of Music Genres. *Journal of New Music Research*, v. 37, n. 2, p. 93-100, 2008. Disponível em: <a href="http://dx.doi.org/10.1080/09298210802479268">http://dx.doi.org/10.1080/09298210802479268</a>>. Acesso em: 14 set. 2014.
- [10] TZANETAKIS, G.; COOK, P. Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing*, v. 10, n. 5, p. 293-302, 2002. ISSN 1063-6676. Disponível em: <a href="http://dx.doi.org/10.1109/TSA.2002.800560">http://dx.doi.org/10.1109/TSA.2002.800560</a>>. Acesso em: 14 set. 2014.
- [11] SILLA, C. N.; KAESTNER, C. A. A.; KOERICH, A. L. Automatic Music Genre Classification Using Ensemble of Classifiers. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS. Montreal: ICSMC, 2007. p. 1687-1692. Disponível em: <a href="http://dx.doi.org/10.1109/ICSMC.2007.4414136">http://dx.doi.org/10.1109/ICSMC.2007.4414136</a>>. Acesso em: 14 set. 2014.
- [12] PARADZINETS, A.; HARB, H.; CHEN, L. *MultiExpert System For Automatic Music Genre Classification*, France, 2009. Laboratoire d'InfoRmatique en Image et Systèmes d'information Ecole Centrale de Lyon. Disponível em: <a href="http://liris.cnrs.fr/publis/?id=4224">http://liris.cnrs.fr/publis/?id=4224</a>. Acesso em: 14 set. 2014.
- [13] LAMPROPOULOS, A., S.; LAMPROPOULOU, P. S.; TSIHRINTZIS, G. A. Musical Genre Classification Enhanced by Improved Source Separation Techniques. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL. London: ISMIR, 2005. p. 576-581. Disponível em: <a href="http://ismir2005.ismir.net/proceedings/1079.pdf">http://ismir2005.ismir.net/proceedings/1079.pdf</a>>. Acesso em: 14 set. 2014.
- [14] COSTA, Y. M. G. et al. Music genre recognition using spectrograms. In: INTERNATIONAL CONFERENCE ON SYSTEMS, SIGNALS AND IMAGE PROCESSING. Sarajevo: IWSSIP, 2011. p. 1-4. ISSN 2157-8672. Disponível em: <a href="http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5977391">http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5977391</a>. Acesso em: 14 set. 2014.
- [15] ARIYARATNE, H. B.; ZHANG, D. A Novel Automatic Hierarchical Approach to Music Genre Classification. In: IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO WORKSHOPS. Melbourne: ICMEW, 2012. p. 564-569. Disponível em: <a href="http://dx.doi.org/10.1109/ICMEW.2012.104">http://dx.doi.org/10.1109/ICMEW.2012.104</a>. Acesso em: 14 set. 2014.
- [16] COSTA, C. H. L.; VALLE, J. D.; KOERICH, A. L.; Automatic classification of audio data. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS. ICSMC, 2004, p. 562-567, v. 1. ISSN 1062-922X. Disponível em: <a href="http://dx.doi.org/10.1109/ICSMC.2004.1398359">http://dx.doi.org/10.1109/ICSMC.2004.1398359</a>>. Acesso em: 14 set. 2014.
- [17] COSTA, Y. M. G. et al. Music genre classification using LBP textural features. *Signal Processing*, Elsevier, v. 92, n. 11, p. 2723-2737, 2012. Disponível em: <a href="http://dx.doi.org/10.1016/j.sigpro.2012.04.023">http://dx.doi.org/10.1016/j.sigpro.2012.04.023</a>. Acesso em: 14 set. 2014.
- [18] LEE, C. H. et al. Automatic Music Genre Classification Based on Modulation Spectral Analysis of Spectral and Cepstral Features. *IEEE Transactions on Multimedia*, v. 11, n. 4, p. 670-682, 2009. ISSN 1520-9210. Disponível em: <a href="http://dx.doi.org/10.1109/TMM.2009.2017635">http://dx.doi.org/10.1109/TMM.2009.2017635</a>>. Acesso em: 14 set. 2014.
- [19] BALTI, H.; FRIGUI, H. Feature Mapping and Fusion for Music Genre Classification. In: IEEE 11<sup>th</sup> INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND APPLICATIONS. Boca Raton: ICMLA, 2012, p. 306-310. Disponível em: <a href="http://dx.doi.org/10.1109/ICMLA.2012.59">http://dx.doi.org/10.1109/ICMLA.2012.59</a>. Acesso em: 14 set. 2014.

- [20] DAVIS, S. B.; MERMELSTEIN, P. Comparison of Parametric Representation for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. 28, n. 4, p. 357-366, 1980. ISSN 0096-3518. Disponível em: <a href="http://dx.doi.org/10.1109/TASSP.1980.1163420">http://dx.doi.org/10.1109/TASSP.1980.1163420</a>>. Acesso em: 14 set. 2014.
- [21] TOLONEN, T.; KARJALAINEN, M. A computationally efficient multipitch analysis model. *IEEE Transactions on Speech and Audio Processing*, v. 8, n. 6, p. 708-716, 2000. ISSN 1063-6676. Disponível em: <a href="http://dx.doi.org/10.1109/89.876309">http://dx.doi.org/10.1109/89.876309</a>>. Acesso em: 14 set. 2014.
- [22] THEODORIDIS, S.; KOUTROUMBAS, K. Pattern Recognition. Greece: Elsevier Academic Press, Second Edition, 2003.
- [23] ZHOU, X. et al. Image Classification Using Super-Vector Coding of Local Image Descriptors. In: 11<sup>th</sup> EUROPEAN CONFERENCE ON COMPUTER VISION. Heraklion: ECCV-10, 2010, p. 141-154. ISSN 0302-9743. Disponível em: <a href="http://dx.doi.org/10.1007/978-3-642-15555-0">http://dx.doi.org/10.1007/978-3-642-15555-0</a> 11>. Acesso em: 14 set. 2014.
- [24] CAMPBELL, W.M.; STURIM, D.E.; REYNOLDS, D.A, Support vector machines using GMM supervectors for speaker verification. *IEEE Signal Processing Letters*, v. 13, n. 5, p.308-311, 2006. ISSN 1070-9908. Disponível em: <a href="http://dx.doi.org/10.1109/LSP.2006.870086">http://dx.doi.org/10.1109/LSP.2006.870086</a>>. Acesso em: 14 set. 2014.
- [25] MADEVSKA-BOGDANOVA, A.; NIKOLIK, D.; CURFS, L. Probabilistic SVM outputs for pattern recognition using analytical geometry. *Elsevier Neurocomputing*, v. 62, p. 293-303, 2004. Disponível em: <a href="http://dx.doi.org/10.1016/j.neucom.2003.03.002">http://dx.doi.org/10.1016/j.neucom.2003.03.002</a>>. Acesso em: 14 set. 2014.
- [26] CHATHURANGA, Y. M. D.; JAYARATNE, K. L. Automatic Music Genre Classification of Audio Signals with Machine Learning Approaches. *GSTF Journal on Computing*, v. 3, n. 2, 2013. Online ISSN 2010-2283. Disponível em: <a href="http://dx.doi.org/10.7603/s40601-013-0014-0">http://dx.doi.org/10.7603/s40601-013-0014-0</a>. Acesso em: 14 set. 2014.
- [27] GILLICK, L.; COX, S. J. Some statistical issues in the comparison of speech recognition algorithms. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING. Glasgow: ICASSP-89, 1989, p. 532-535, v. 1. ISSN 1520-6149. Disponível em: <a href="http://dx.doi.org/10.1109/ICASSP.1989.266481">http://dx.doi.org/10.1109/ICASSP.1989.266481</a>>. Acesso em: 14 set. 2014.
- [28] LI, T.; OGIHARA, M.; LI, Q. A comparative study on content-based music genre classification. In: CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL. Toronto: SIGIR '03, 2003, p. 282-289. Disponível em: <a href="http://dx.doi.org/10.1145/860435.860487">http://dx.doi.org/10.1145/860435.860487</a>>. Acesso em: 14 set. 2014.
- [29] LI, T.; OGIHARA, M. Toward intelligent music information retrieval. *IEEE Transactions on Multimedia*, v. 8, n. 3, p. 564-574, 2006. ISSN 1520-9210. Disponível em: <a href="http://dx.doi.org/10.1109/TMM.2006.870730">http://dx.doi.org/10.1109/TMM.2006.870730</a>>. Acesso em: 14 set. 2014.
- [30] LIM, S. C. et al. Music-genre classification system based on spectro-temporal features and feature selection. *IEEE Transactions on Consumer Electronics*, v. 58, p. 1262-1268, 2012. ISSN 0098-3063. Disponível em: <a href="http://dx.doi.org/10.1109/TCE.2012.6414994">http://dx.doi.org/10.1109/TCE.2012.6414994</a>. Acesso em: 14 set. 2014.
- [31] BENETOS, E.; KOTROPOULOS, C. Non-Negative Tensor Factorization Applied to Music Genre Classification. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 18, n. 8, p. 1955-1967, 2010. ISSN 1558-7916. Disponível em: <a href="http://dx.doi.org/10.1109/TASL.2010.2040784">http://dx.doi.org/10.1109/TASL.2010.2040784</a>. Acesso em: 14 set. 2014.
- [32] LIDY, T.; RAUBER, A. Evaluation of Feature Extractors and Psycho-Acoustic Transformations for Music Genre Classification. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL. London: ISMIR, 2005, p. 34-41. Disponível em: <a href="http://ismir2005.ismir.net/proceedings/1033.pdf">http://ismir2005.ismir.net/proceedings/1033.pdf</a>>. Acesso em: 21 jan. 2015.