Erica Moura Pereira

Estudos Sobre uma Ferramenta de Classificação Musical

Dissertação de Mestrado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do título de Mestre em Engenharia Elétrica. Área de concentração: Engenharia Elétrica

Orientador: Amauri Lopes

Co-orientador: Jayme Barbedo

Campinas, SP

2009

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

Pereira, Erica Moura
P414e Estudos Sobre uma Ferramenta de Classificação Musical
/Erica Moura Pereira. – Campinas, SP: [s.n.], 2009.

Orientadores: Amauri Lopes, Jayme Barbedo.
Dissertação de Mestrado - Universidade Estadual de Campinas,
Faculdade de Engenharia Elétrica e de Computação.

1. Taxonomia. 2. Processamento de Sinais.
I. Lopes, Amauri. II. Barbedo, Jayme. III. Universidade Estadual de Campinas.
Faculdade de Engenharia Elétrica e de Computação. IV. Título

Título em Inglês: A Study on a Classifier of Musical Signals
Palavras-chave em Inglês: Taxonomy, Signal Processing
Área de concentração: Telecomunicações e Telemática
Titulação: Mestre em Engenharia Elétrica
Banca Examinadora: Jônatas Manzolli, Fúrio Damiani
Data da defesa: 04/03/2009
Programa de Pós Graduação: Engenharia Elétrica

COMISSAO JULGADORA - TESE DE MESTRADO

Candidata: Erica Moura Pereira

Data da Defesa: 4 de março de 2009

Título da Tese: "Estudos Sobre uma Ferramenta de Classificação Musical"

Prof. Dr. Amauri Lopes (Presidente): _____

Prof. Dr. Jônatas Manzolli: _____

Prof. Dr. Furio Damiani: _____

Resumo

Esta dissertação apresenta alguns estudos sobre um classificador automático para sinais musicais proposto na literatura, o qual se distingue pelas soluções inovadoras e pelo desempenho. Os objetivos foram as análises da viabilidade de ampliação da aplicabilidade do classificador em taxonomias distintas daquela usada pelos autores e da possibilidade de redução do esforço computacional para treinamento. Foi desenvolvido um novo programa de computador para a implementação do classificador e o mesmo foi validado usando a base de dados original. Os estudos sobre aplicabilidade usaram uma nova taxonomia, menor que a original, e mostram um desempenho similar ao original e que, portanto, o classificador pode se aplicado a taxonomias distintas daquela original. O novo processo de treinamento proposto permitiu reduzir o esforço computacional e ainda manter o desempenho original do classificador.

Palavras-chave: Classificação Automática de Sinais Musicais, Taxonomia de Gêneros Musicais, Processamento Digital de Sinais de Áudio.

Abstract

This work presents the results of some investigations on an automatic musical signal classifier proposed in the literature and notable by its solutions and performance. The objectives were to analyze the viability of extending the applicability of the classifier over new taxonomies, distinct from that proposed by the authors, and to reduce the computational effort demanded by the training process. A new computer implementation of the classifier was developed and validated using the original database. The investigations on the applicability used a new taxonomy, smaller than the original, and showed a performance similar to the original and that the classifier can be applied to taxonomies different from the original. The new training process allowed a reduced computational effort and assured the original performance level.

Keywords: Automatic Musical Signal Classifier, Musical Genre Taxonomy, Digital Audio Signal Processing.

Aos meus pais

Agradecimentos

Ao meu orientador e co-orientador, Profs. Amauri Lopes e Jayme Barbedo, sou grata pela orientação.

Aos meus pais por todo amor e apoio durantes estes anos.

Aos meus familiares pelo carinho e amor.

Aos meus queridos amigos de Campinas Marcelo Ribeiro, Pablo Dalben, Ricardo Freitas, Rosseane Barata, Tiago Lacerda, Tiago Massaruto pelos momentos de distração que passamos juntos.

Aos professores Adolfo Maia, Fúrio Damiani, Jônatas Manzolli e Renato Baldini Filho por participarem da banca.

A todos aqueles que me ajudaram com a finalização deste trabalho.

À CAPES, pelo apoio financeiro.

Sumário

Capítulo 1

Introdução

Nas últimas décadas, a relação entre o ser humano e o áudio mudou radicalmente. O desenvolvimento dos codecs perceptuais tornou possível o armazenamento de grandes conjuntos de música com a exigência de relativamente pouca memória [Pohlmann, 2000]. Ao mesmo tempo, os dispositivos de armazenamento evoluíram e a Internet tornou-se disponível no mundo todo, possibilitando a troca de dados e informações rapidamente. Como resultado destas evoluções e inovações tecnológicas, os conjuntos de áudio digital também evoluíram produzindo um crescimento contínuo das bases de dados. Paralelamente, a indústria da distribuição eletrônica de música vem crescendo velozmente.

Neste contexto, o desenvolvimento de ferramentas capazes de manipular este vasto conjunto de áudio digital de maneira simples e rápida tornou-se essencial, devendo facilitar a consulta, a procura, a recuperação e o tratamento destes dados. Uma das ferramentas mais importantes é a Classificação Automática de Sinais Musicais em Gêneros, uma vez que permite, em princípio, indexar e localizar automaticamente conjuntos de áudio sem a interferência humana. A classificação automática se baseia na extração de parâmetros capazes de fornecer o máximo possível de informação a respeito dos sinais de áudio e no uso destes parâmetros para classificar os sinais musicais de acordo com um critério pré-estabelecido.

Esta é uma área de pesquisa ampla com várias lacunas a serem preenchidas, o que representa uma ótima oportunidade para o desenvolvimento de novas tecnologias, relevantes não apenas no meio acadêmico, mas também na área comercial [Wang & Jo, 2006], [Büchler, 2002], [V. Peltonen & Sorsa, 2002]. Diversas aplicações podem se beneficiar diretamente com avanços do

desempenho da classificação automática. Por exemplo, a criação de aparelhos auditivos seletivos, a transcrição automática de música ou texto, a remoção de ruído, o reconhecimento de voz e locutor, a identificação de conteúdo de mídia, a equalização automática de sinais musicais de acordo com o conteúdo, a busca rápida em bancos de dados, entre outras. Muitas dessas aplicações possuem uma demanda imediata [V. Peltonen & Sorsa, 2002], [Büchler, 2002].

Apresentaremos a seguir um breve resumo das principais publicações sobre a temática, usando como critérios de seleção não apenas os estudos bibliográficos realizados no contexto desta dissertação, mas também as opiniões de alguns autores. O primeiro trabalho a ser destacado é [T. Lambrou & Linney, 1998] onde a classificação dos sinais é feita em três gêneros musicais através da extração de alguns parâmetros nos domínios temporal e da transformada *wavelet*. O próximo trabalho, [H. Deshpande & Nam, 2001], também utiliza três gêneros musicais, porém a classificação é feita através de uma análise gráfica de espectrogramas. O terceiro trabalho, [Logan, 2000], realiza um estudo sobre a adaptação dos coeficientes Mel cepstrais (MFCCs) para realizar a classificação dos sinais musicais. A dissertação de [Golub, 2000] realiza a classificação dos sinais musicais em sete gêneros bastante diferentes, utilizando quatro parâmetros, assim como outras características estatísticas obtidas a partir destes parâmetros. Foram utilizados três classificadores: Modelo Linear Generalizado (GLM), K vizinhos mais próximos (K-NN) e a rede Perceptron Multicamadas (MLP). Em [Liu & Wan, 2001] os sinais de áudio são classificados em voz, música e outros sons utilizando quatro classificadores: redes neurais, K vizinhos mais próximos, Modelo de Mistura Gaussiana (GMM) e rede neural probabilística (PNN). A dissertação de [Kosina, 2002] classifica as músicas em três gêneros, utilizando o classificador K vizinhos mais próximos.

Os trabalhos [Martin, 1999] e [Martin & Kim, 1998] abordam o problema de identificação de instrumentos. Eles propõem um conjunto de características relacionadas com as propriedades físico-acústicas dos instrumentos com o objetivo de identificá-los em ambiente polifônico. Outro estudo sobre a identificação de instrumentos foi publicado por [Fraser & Fujinaga, 1999].

O trabalho a seguir é um dos mais relevantes e foi publicado em 2002 por George Tzanetakis e Perry Cook [Tzanetakis & Cook, 2002]. Este trabalho emprega três conjuntos de parâmetros representando a textura timbral, o conteúdo rítmico e o pitch. Os sinais de áudio são classificados em 10 gêneros usando classificadores estatísticos para o reconhecimento de padrões e o índice médio de

acerto foi de 60%.

Este último trabalho foi usado por muitos pesquisadores como base para o desenvolvimento de outros estudos. Como um exemplo, [Agostini et al., 2003] propuseram uma estratégia de classificação baseada no instrumento musical sendo tocado.

Já existem eventos especializados na temática em questão, como por exemplo, o *International Symposium on Music Information Retrieval* (ISMIR), realizado desde o ano de 2000. Durante a 6ª Conferência do ISMIR em 2005 aconteceu o primeiro *Music Information Retrieval Evaluation eXchange*. Nesta competição, os participantes deveriam realizar a classificação de sinais musicais em 10 gêneros diferentes. Diversos algoritmos de classificação foram propostos e seus índices médios de acerto foram entre 60% e 82% [MIREX Contest Results, 2005].

Tais eventos impulsionaram significativamente as publicações nesta temática. Neste contexto, [K.West & Cox, 2004] apresentaram uma nova ferramenta de classificação baseada na construção não-supervisionada de árvores de decisão. Este estudo também apresenta testes com diversos parâmetros e procedimentos para a classificação dos sinais musicais. Em [F. Gouyon & Widmer, 2004] foi realizada uma pesquisa sobre a eficiência de parâmetros ligados à característica rítmica. [O. Hellmuth & Wistorf, 2004] realizaram um estudo comparando alguns parâmetros básicos com uma proposta de classificação particular, a qual é baseada na semelhança entre um sinal e alguns sinais de referência. A tese [Pampalk, 2006] apresenta um amplo estudo sobre os modelos de classificação dos sinais de áudio. Um método de classificação musical utilizando padrões rítmicos foi proposto em [S. Dixon & Widmer, 2004]. Um estudo a respeito de abordagens acústicas e subjetivas para o cálculo da semelhança entre músicas foi feito por [A. Berenzweig & Whitman, 2004]. [C. McKay & Fujinaga, 2005] propõem uma estrutura para otimizar a classificação dos sinais em gêneros musicais. Em [S. Lippens & Tzanetakis, 2004] é apresentado um estudo comparativo entre os desempenhos de seres humanos e de estratégias automáticas de classificação. Outro estudo comparativo foi realizado por [T. Pohle & Widmer, 2005], o qual compara o desempenho de vários parâmetros comumente usados para a classificação de sinais musicais. A máquina de vetores suporte foi aplicada em [C. Xu & Shao, 2005] para realizar a classificação hierárquica dos sinais musicais.

Finalmente, em [Barbedo & Lopes, 2006] os autores propõem uma nova estratégia de classifi-

cação, a qual será abordada nesta dissertação objetivando um estudo de algumas de suas características. Apresentamos a seguir uma visão geral e simplificada deste método e, em seguida, uma descrição sucinta dos estudos realizados nesta dissertação. O Capítulo 4 apresentará esta estratégia com maiores detalhes.

A taxonomia utilizada por [Barbedo & Lopes, 2006] possui quatro camadas hierárquicas, com três gêneros na primeira camada e 29 gêneros na camada mais refinada. O classificador apresenta três aspectos importantes. O primeiro é a classificação de baixo para cima ao longo da estrutura taxonômica, isto é, um sinal musical é inicialmente classificado na camada mais refinada e a sua classificação nas camadas superiores é conseqüência desta primeira classificação. O segundo aspecto é o uso de 29 gêneros nesta camada mais baixa de modo a dispor de uma classificação bem mais refinada do que aquela pretendida nos gêneros-alvo, mais acima. Por fim, o terceiro aspecto é um processo de classificação baseado na comparação entre pares de gêneros. Isto significa que são consideradas todas as possíveis combinações de dois gêneros da camada mais refinada e que a classificação inicial é feita a cada par de gêneros, definindo um gênero vencedor a cada par. A classificação final na camada mais refinada é dada pelo gênero com o maior número de vitórias na classificação por pares.

Prosseguindo com a descrição, abordaremos inicialmente o processo de treinamento e posteriormente o processo de classificação. O processo de treinamento se inicia com a divisão do sinal musical em quadros. Em seguida, quatro parâmetros são extraídos de cada quadro. Os quadros são agrupados em segmentos de análise de um segundo. Para cada segmento e cada parâmetro são calculadas três medidas estatísticas: média, variância e prevalência do pico principal (os detalhes desta última medida estatística são apresentados no Capítulo 4), denominadas pelos autores de *summary features*. As 12 *summary features* são agrupadas no vetor de *summary features*, o qual, então, caracteriza o segmento de análise do sinal. Este vetor é calculado para cada segmento de análise de cada sinal de cada gênero da base de dados destinada ao treinamento.

Durante o treinamento, usamos os vetores de *summary features* para escolhermos, a cada possível par de gêneros, seis vetores de referência, três para cada gênero de um par. Tais vetores de referência são aqueles que propiciam a melhor separação entre os gêneros de um par. Sua escolha é feita através de um processo de comparação, usando a distância Euclidiana, entre seis vetores candidatos a vetores de referência, pré-escolhidos, e todos os demais vetores de *summary features* de ambos os gêneros.

Após o cálculo da distância, cada vetor de *summary features* é rotulado com o gênero do vetor candidato mais próximo. Em seguida, trocamos os seis vetores candidatos e repetimos o processo de classificação. O conjunto de seis vetores candidatos que propiciar o maior número de classificações corretas é definido como o conjunto de vetores de referência para o par de gêneros considerado. Este processo é repetido para todas as combinações de pares de gêneros da camada mais refinada.

Após o treinamento, inicia-se o processo de classificação. Para uma dada música a ser classificada, primeiramente é feito o cálculo do vetor de *summary features* para cada segmento de análise. Em seguida, para um dado segmento, é feito um processo de comparação, através do cálculo da distância Euclidiana, entre o conjunto de vetores de referência de um dado par de gêneros e o vetor de *summary features* do segmento de análise, a fim classificar o segmento em termos dos gêneros do par.

Este processo é repetido para as demais combinações de pares de gêneros e para todos os segmentos de análise do sinal musical, com o intuito de definir o gênero vencedor ao longo de todos os segmentos de análise.

Este processo classificatório e os três aspectos antes citados configuram um classificador distinto e com um índice de acerto de 80%, nível este notável se comparado ao estado da arte na época da publicação.

Estas características motivaram um novo estudo desta ferramenta com o objetivo de enriquecer o conhecimento de suas potencialidades. A seguir apresentaremos os detalhes deste novo estudo, o qual é o objeto desta dissertação.

1. **Estrutura Taxonômica**:

   Foi comentado anteriormente que o método proposto por [Barbedo & Lopes, 2006] emprega 29 gêneros na camada mais refinada, propiciando uma distinção refinada entre gêneros. A classificação é feita inicialmente nesta camada, explorando este grau de distinção. Entretanto, existem situações onde se demanda taxonomias menores, ou seja, taxonomias com menor número de camadas e também com menor número de gêneros na sua camada mais refinada.

   Este fato nos motivou a realizar uma investigação para verificar se o classificador manterá o seu desempenho quando for aplicado a uma taxonomia reduzida, com um número menor de gêneros e de camadas.

2. **Escolha dos Vetores Candidatos a Vetores de Referência**:

No processo de escolha dos vetores de referência para cada par de gêneros, os autores não usaram todos os vetores de *summary features* disponíveis a cada par. Ao contrário, usam apenas uma parcela destes vetores, cujos elementos são denominados de vetores candidatos a vetores de referência. Este conjunto de vetores candidatos é determinado calculando-se inicialmente a média e a variância dos vetores de *summary features* de cada gênero de um par. São considerados vetores candidatos apenas os vetores que estão dentro de uma região com variância pré-fixada. Os vetores de *summary features* que estiverem fora deste grupo são descartados.

Visando simplificar este processo de definição dos vetores candidatos e reduzir o esforço computacional de treinamento, investigamos uma nova forma de seleção dos vetores candidatos a vetores de referência. Propomos usar um número pré-definido de vetores de *summary features*, escolhidos aleatoriamente. Adicionalmente, pretendemos definir o menor número que assegure o desempenho original do método.

3. **Análise da Robustez dos Vetores de Referência**:

   Realizamos uma análise da robustez dos vetores de referência gerados segundo o novo procedimento proposto. Para isto escolhemos alguns casos críticos de classificação de músicas e realizamos experimentos gerando novos vetores de referência e analisamos o desempenho resultante.

Podemos observar, então, que o trabalho aqui proposto não visa gerar um novo método classificatório. Seu objetivo é analisar o grau de flexibilidade de aplicação da ferramenta proposta por [Barbedo & Lopes, 2006] e também analisar procedimentos alternativos de treinamento que simplifiquem este processo e ofereçam redução de esforço computacional sem comprometimento do desempenho original.

Para a realização destes estudos foi desenvolvido um novo programa Matlab®, o qual foi validado através de testes. A base de dados utilizada foi a mesma adotada em [Barbedo & Lopes, 2006].

Esta dissertação está organizada da seguinte forma. O Capítulo 2 apresenta a importância da taxonomia para a classificação automática de áudio, os problemas encontrados para a definição dos

níveis hierárquicos, alguns critérios para a elaboração de uma estrutura taxonômica e como é realizada a classificação através da mesma. O Capítulo 3 apresenta alguns parâmetros que normalmente são extraídos dos sinais de áudio no contexto de sistemas automáticos de classificação e o Capítulo 4 descreve os detalhes do funcionamento da ferramenta de classificação proposta por [Barbedo & Lopes, 2006]. No Capítulo 5 apresentamos os estudos realizados e as respectivas conclusões. Finalmente, o Capítulo 6 apresenta as conclusões gerais e propostas para trabalhos futuros.

Capítulo 2

Taxonomia

> The value of knowledge is largely tied to the way in which that knowledge is organized. If you can't find it, it is not likely to be of much use to you.
>
> Marc Rapport

Conforme já foi dito no Capítulo 1 o desenvolvimento tecnológico das últimas décadas prodiziu um crescimento explosivo nas bases de dados musicais. Em vista disso, foi preciso organizar estas bases de dados, tornando necessário o desenvolvimento de taxonomias apropriadas.

As taxonomias utilizadas para a classificação de áudio são desenvolvidas para diversas aplicações. No entanto, as que mais nos interessam aqui são aquelas desenvolvidas para a classificação automática dos sinais musicais.

As estruturas taxonômicas podem ser simples ou complexas, ou seja, podem conter diversos níveis hierárquicos, onde os níveis mais baixos representam as classificações mais específicas e refinadas. Por exemplo, na biologia, a classificação dos seres vivos ocorre de acordo com uma hierarquia, que começa pelos reinos, os quais se subdividem em filos, os filos em classes, e assim por diante. O mesmo ocorre nas taxonomias desenvolvidas para os sinais de áudio. Geralmente, a classificação destes sinais começa pela diferenciação entre voz, música e ruído.

Nas próximas seções deste capítulo veremos a importância da utilização de uma taxonomia no

contexto da classificação automática de áudio, quais os problemas encontrados para a definição dos níveis hierárquicos, qual a melhor maneira de elaborar uma estrutura taxonômica e como é realizada a classificação através da mesma.

2.1 Importância da Taxonomia

Existem diversos motivos para desenvolvermos uma taxonomia para sinais musicais. O mais importante no contexto deste trabalho é o fato de que uma estrutura taxonômica é parte fundamental de sistemas de classificação automática de sinias musicais, sistemas esses que, por sua vez, contam com diversas aplicações importantes.

Como segundo motivo, o desenvolvimento de uma taxonomia e, em particular, padronizada, representaria uma ferramenta de grande utilidade para a realização de pesquisas e consultas de maneira organizada, particularmente para usuários da Internet. Da mesma forma, representaria uma ferramenta importante para a divulgação e comercialização de produtos musicais através da Internet.

Por fim, deve-se ressaltar que as estruturas taxonômicas mostram as relações de dependência entre diferentes tipos de sinais musicais, como, por exemplo, entre os gêneros musicais. Nos sistemas de classificação, a definição e uso adequados dessas estruturas permitem alcançar uma maior eficiência durante o processo de aprendizagem e classificação.

2.2 Problemas para a definição da taxonomia

Ainda não há um consenso para a definição de uma taxonomia padrão para sinais musicais. Isto ocorre devido a dois problemas básicos.

O primeiro problema está relacionado à construção da taxonomia, a qual é construída por e dirigida para usuários humanos, dependendo de vários fatores que serão discutidos a seguir. Como conseqüência, a diferença entre os gêneros de todas as taxonomias disponíveis é baseada principalmente em critérios individuais, ou seja, essas diferenças podem mudar de pessoa para pessoa.

Em [Barbedo & Lopes, 2006] os autores mostram que os critérios para definir os gêneros são influenciados por origem geográfica e pela época em que foram criados, pelos requisitos técnicos

exigidos, pela influência de críticos musicais através das diversas mídias, dentre outros. Os critérios podem atender aos requisitos subjetivos de usuários particulares, mas são inadequados aos objetivos de um sistema taxonômico padrão e também de sistemas de classificação automática de sinais musicais.

Como conseqüência desta característica, há uma falta de padronização na nomenclatura para os gêneros musicais e, portanto, para as taxonomias, como pode ser visto em [Pachet & Cazaly, 2000], [Julien & Pachet, 2003], [Basili et al., 2004], [Fabbri, 1999], [Fabbri & Chambers, 1982]. O grau de arbitrariedade e inconsistência da classificação musical em gêneros é comentado em [Pachet & Cazaly, 2000], onde os autores comparam três taxonomias distintas presentes na rede mundial de computadores: **mp3.com**, a qual apresentava 430 gêneros musicais na época do estudo, **allmusic.com**, com 513 gêneros e **amazon.com**, a qual possuía a maior taxonomia com 719 gêneros. Neste estudo, os autores observaram que não existia um consenso na nomenclatura utilizada por estes sites e que apenas 70 palavras eram comuns dentre as taxonomias utilizadas. E também observaram que as estruturas não possuiam definições comuns para termos amplamente utilizados, como *Pop* e *Rock*, ou seja, estes não representavam exatamente o mesmo grupo de músicas. Por fim, os autores notaram ainda que a localização dos níveis hierárquicos também diferia de uma taxonomia para outra. Embora tenham havido esforços para uniformizar os critérios, a situação na verdade piorou, pois o número de *websites* que oferecem músicas teve um crescimento significativo e cada um adota seus próprios critérios.

Outro exemplo da falta de consenso na nomenclatura de gêneros é o caso da música clássica (ou erudita). Existem diversas contestações a respeito da terminologia mais adequada, uma vez que, para alguns músicos, o termo "música clássica" deveria ser destinado apenas à música erudita produzida durante o período da história da música designado como era clássica ou classicismo, enquanto outros não concordam com essa imposição cronológica. Para completar a confusão envolvendo o termo "clássico", é comum que tal termo seja usado para indicar qualquer música (de qualquer gênero) cujo impacto se estendeu para muito além do período em que foi pela primeira vez introduzida ao público.

O segundo problema para a definição de uma taxonomia para sinais de áudio diz respeito ao grau de subjetividade que uma taxonomia possui. Mesmo que se conseguisse padronizar a nomenclatura de gêneros musicais, ainda assim não haveria garantia de que se pudesse atender às expectativas

da maioria dos usuários em termos de classificação de suas preferências musicais, dado que tais expectativas são sempre subjetivas e, mais importante, variam de pessoa para pessoa. Com isto dois usuários poderiam classificar uma mesma peça musical em gêneros distintos em uma taxonomia padrão.

A isto deve ser acrescentado o fato de que os gêneros podem variar com o tempo, fato este cada vez mais comum em função da intensa relação que se observa entre os compositores, produtores e usuários. Por fim, devemos mencionar que vários gêneros musicais, especialmente os modernos, resultam de várias tendências culturais e, portanto, incluem elementos de vários outros gêneros. Por exemplo, alguns estilos de *Jazz* incorporam elementos de outros gêneros como o *Fusion* (*Jazz* + *Rock*); também, algumas canções recentes de *Reggae* possuem fortes elementos de *Rap*; como último exemplo, várias canções de *Rock* incorporam elementos sonoros gerados por sintetizadores.

Como apontado em [Barbedo & Lopes, 2006], se mesmo as principais estruturas taxonômicas apresentam tantas inconsistências entre elas, não podemos esperar qualquer grau de interoperabilidade semântica entre os diferentes gêneros destas taxonomias [Julien & Pachet, 2003]. E para que qualquer proposta tenha significado, é necessário que seja usada uma linguagem comum para ambos, o usuário e o propositor de uma taxonomia. Apesar destas dificuldades, pesquisas estão sendo realizadas para o desenvolvimento de taxonomias cuidadosamente projetadas [Pachet & Cazaly, 2000], [Julien & Pachet, 2003].

É neste contexto que a taxonomia proposta em [Barbedo & Lopes, 2006] foi desenvolvida, usando gêneros e nomenclaturas que estão amplamente presentes nas principais taxonomias em uso e que, portanto, proporcionam maior probabilidade de serem facilmente identificadas pela maioria dos usuários.

2.3 Critérios para a elaboração de uma taxonomia

A primeira questão a ser considerada para a elaboração de uma taxonomia é a sua finalidade, pois é a partir dela que serão definidos os critérios a serem seguidos. Estes, por sua vez, definirão as classes de sinais ou peças musicais, gêneros, subgêneros e suas relações de interdependência.

As taxonomias destinadas a sistemas automáticos de classificação de sinais de áudio, em geral,

classificam os sinais de áudio em voz, música e ruído, em um primeiro nível. Posteriormente, as peças musicais são classificadas uma a uma, sendo indexadas a gêneros. Ou seja, o critério principal é a análise de certas características sonoras das músicas, características estas capazes de classificá-las em um mesmo gênero, ou então, em gêneros distintos. Este processo dá origem aos próprios gêneros a serem considerados na taxonomia. Este critério difere de outros como a classificação por artistas ou álbuns, como ocorre na indústria da música, ou então, o critério histórico, o geográfico ou o comercial. Por exemplo, as músicas distintas de um mesmo país seriam agrupadas em uma mesma classe caso o critério fosse o geográfico. Porém, podem pertencer a diferentes gêneros; e, portanto, no contexto da classificação automática de sinais se áudio, deveriam ser agrupadas em gêneros distintos.

Outro assunto que deve ser analisado é a forma como os diferentes gêneros se ramificam em subgêneros. Considerando uma hierarquia baseada em uma estrutura em árvore, esta variação se aplicará à profundidade e à largura dos vários ramos da árvore. Por exemplo, alguns gêneros podem conter muitos subgêneros, como a música eletrônica que possui o *Techno*, o *Jungle*, o *Rave*, o *Trance*, e outros. Já outros gêneros como o *Reggae* tendem a ter menos ramificações.

Os níveis hierárquicos devem ser constituídos levando-se em conta a possibilidade de se adicionar novos gêneros e subgêneros, considerando que a composição musical está sempre em evolução. Isto permitirá que futuras modificações sejam facilmente inseridas na estrutura taxonômica.

A Figura 2.1 apresenta um exemplo ilustrativo de uma taxonomia. A classificação de um sinal de áudio, nesta estrutura pode ocorrer de duas maneiras:

1. *De cima para baixo*: classifica o sinal a partir do nível mais elevado até os níveis mais refinados.

2. *De baixo para cima*: classifica do nível mais refinado até o mais elevado.

Na literatura sobre classificação dos gêneros musicais, o mais comum é a direção *de cima para baixo*.

Como exemplo de uso destas direções de classificação, se fossemos classificar a música clássica de Luciano Pavarotti *O Sole Mio* baseados na primeira forma, a classificação começaria com a distinção entre Instrumental ou Vocal. Como há elementos vocais nesta música, a classe escolhida seria a Vocal. O próximo passo seria classificá-la em Ópera ou Coral. Sabendo que esta música possui características de Ópera, isto é, um diálogo cantado de forma dramática, o gênero que melhor rep-

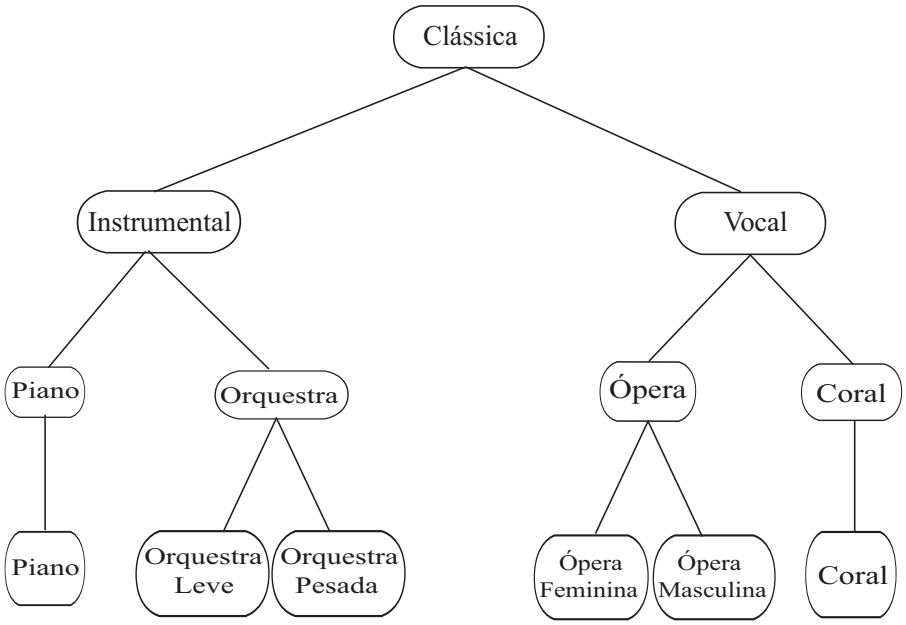

Figura 2.1: Estrutura Taxonômica.

resentaria esta música seria Ópera. O último passo seria a distinção entre Ópera Masculina e Ópera Feminina, o que levaria, finalmente, ao gênero Ópera Masculina.

Por outro lado, se a classificação for baseada na segunda forma, de baixo para cima, o primeiro passo da classificação seria definir qual gênero possui características mais parecidas com as da música analisada. Neste caso o gênero que melhor representa a música é Ópera Masculina. A próxima classificação será o gênero Ópera, e então Vocal.

Embora esta direção de classificação de baixo para cima não seja a mais usual, ela é adotada em [Barbedo & Lopes, 2006]. O Capítulo 4 apresentará as motivações para tal escolha.

Por fim, o processo de classificação automática de sinais musicais se baseia na extração de parâmetros dos sinais de áudio, os quais devem fornecer informação a respeito do sinal, de modo a permitir a sua caracterização segundo uma dada taxonomia. Neste contexto, os diversos gêneros da taxonomia escolhida devem ser caracterizados através de valores típicos de um conjunto de parâmetros. Tais valores típicos são determinados através de um processo de treinamento. Este conjunto de valores típicos será então usado, de acordo com determinada estratégia, no processo de classificação dos sinais musicais. As estratégias propostas na literatura comparam os parâmetros do sinal que desejamos classificar com os valores típicos que caracterizam os gêneros. O Capítulo 3 apresenta uma

lista de parâmetros comumente utilizados.

Capítulo 3

Extração dos Parâmetros

A primeira etapa da classificação dos sinais de áudio é a extração de parâmetros, os quais devem, idealmente, retratar aquelas características que são responsáveis pela possibilidade de se classificar os sinais musicais nos seus respectivos gêneros. Também, idealmente desejamos minimizar o número de parâmetros necessários para uma representação adequada destas características.

Estas características idealizadas ainda motivam pesquisas e a proposição de novos parâmetros. Não obstante, constatamos uma tendência na literatura para o uso daqueles parâmetros já consagrados, parte dos quais são apresentados neste capítulo. Observaremos que cada um destes parâmetros representa uma característica diferente do sinal musical, ou seja, cada parâmetro fornece uma informação diferente sobre a qualidade acústica do sinal [Theodoridis & Koutroumbas, 2006]. É também importante observar que esses parâmetros têm pouca capacidade discriminativa quanto aplicados individualmente a sinais musicais. Eles só funcionam adequadamente se combinados com outros parâmetros.

Para a extração de parâmetros, o sinal musical é dividido em pequenos trechos denominados quadros. Na literatura encontramos quadros cuja duração varia desde dezenas até centenas de milissegundos. Neste trabalho adotamos quadros com duração de 21,3 ms, taxa de amostragem de 48.000 amostras por segundo, de modo que cada quadro é composto por 1024 amostras, em concordância com as especificações adotadas em [Barbedo & Lopes, 2006].

A seguir apresentamos alguns dos parâmetros mais usados na extração de informação dos sinais de áudio: centróide espectral, largura de faixa, fluxo espectral, freqüência fundamental, taxa de cruza-

mentos por zero, coeficientes cepstrais, *roll-off*, sonoridade e proporção de quadros com pouca energia.

3.1 Centróide Espectral

O centróide espectral é definido como o centro de massa do espectro de energia de cada quadro [Theodoridis & Koutroumbas, 2006] e é calculado através da expressão:

$$ce(i) = \frac{\sum_{k=1}^{K} k \cdot |X_i(k)|^2}{\sum_{k=1}^{K} |X_i(k)|^2}, \tag{3.1}$$

onde $X_i(k)$ representa as componentes da Transformada Discreta de Fourier (TDF) do quadro i e K é metade do número de componentes espectrais adotado para a TDF. Usamos este valor para k, dado que os sinais musicais assumem valores reais e, neste caso, as componentes da TDF apresentam simetria.

Este parâmetro está relacionado ao "brilho" do sinal, ou seja, à importância relativa das componentes espectrais nas altas freqüências de áudio. Considerando a gama dos valores deste parâmetro, os maiores valores indicam texturas mais "brilhantes", ou seja, com uma maior quantidade de energia nas componentes de alta freqüência. Comparando com os sinais de voz, os sinais musicais tendem a apresentar valores maiores para o centróide espectral.

Este parâmetro é utilizado com êxito em muitos trabalhos como [Tzanetakis et al., 2002], [Agostini et al., 2003], [Li, 2000], [Cataltepe et al., 2007], [Burred & Lerch, 2003], [Lu & Hankinson, 1998] e [Saunders, 1996].

3.2 Largura de Faixa

É uma medida da largura de faixa de freqüência de um sinal em Hertz e é dada pela expressão:

$$bw_i = \sqrt{\frac{\sum_{k=1}^{K} [(ce_i - k)^2 \cdot |X_i(k)|^2]}{\sum_{k=1}^{K} |X_i(k)|^2}}, \tag{3.2}$$

onde ce_i é o centróide espectral (expressão 3.1).

Os sinais musicais tendem a apresentar uma largura de faixa de freqüências maior que aquela dos sinais de voz [Lu & Hankinson, 1998]. A utilização deste parâmetro pode ser vista em [Lu & Hankinson, 1998], [Agostini et al., 2003] e [McKinney, 2003].

3.3 Fluxo Espectral

O fluxo espectral é definido como a diferença quadrática entre o logaritmo da magnitude das distribuições espectrais de quadros sucessivos. O objetivo deste parâmetro é caracterizar quão rápidas são as mudanças no espectro do sinal ao longo dos quadros, sendo calculado através de:

$$fe_i = \sum_{k=1}^{K} \{\log_{10}[X_i(k)] - \log_{10}[X_{i-1}(k)]\}^2. \tag{3.3}$$

Comparando os sinais musicais e os de voz, estes últimos normalmente apresentam maiores valores para o fluxo espectral, pois apresentam seqüências de fonemas surdos e sonoros. A utilização deste parâmetro pode ser vista em [Tzanetakis et al., 2002], [Cataltepe et al., 2007], [Barbedo & Lopes, 2006], [Julien & Pachet, 2003], [Burred & Lerch, 2003], [Lu & Zhang, 2002] e [Li & Ogihara, 2006].

3.4 Freqüência Fundamental

A freqüência fundamental é um dos principais atributos auditivos dos sons harmônicos, ou seja, daqueles cujos sinais apresentam periodicidades em curto prazo. Esta periodicidade produz a sensação sonora designada por "*pitch*" [Barbedo & Lopes, 2004]. Este parâmetro foi usado em [Tzanetakis & Cook, 2002], [Wold et al., 1996], [Lu & Hankinson, 1998].

Nas aplicações práticas, normalmente é necessário estimar várias freqüências fundamentais simultâneas e isto exige algoritmos relativamente sofisticados. Na literatura, existem diversos algoritmos especializados para a estimação da freqüência fundamental.

3.5 Taxa de Cruzamento por Zero

Este parâmetro fornece o número de cruzamentos por zero do sinal no domínio do tempo. Sinais de música tendem a apresentar uma taxa de cruzamentos por zero diretamente relacionada às freqüências fundamentais dos instrumentos que estão sendo tocados e normalmente os valores não variam muito ao longo do sinal. No caso de sinais de voz, os valores variam intensamente de quadro para quadro: fonemas sonoros tendem a gerar poucos cruzamentos por zero, enquanto trechos de fonemas fricativos e trechos de silêncio normalmente resultam em valores elevados. Trechos de silêncio resultam em taxa de cruzamentos por zero elevada porque este é, na prática, constituído por ruído branco.

Este parâmetro é calculado como:

$$Z_i = \frac{1}{2N} \sum_{n=0}^{N-1} |\text{sgn}(x_i[n]) - \text{sgn}(x_i[n-1])|, \tag{3.4}$$

onde N é o número de amostras, $x_i[n]$ são as amostras do sinal temporal para o i-ésimo quadro e a função sgn vale 1 ou -1 para os argumentos positivos e negativos, respectivamente.

Este parâmetro foi aplicado em muitos trabalhos, como [Saunders, 1996], [Wold et al., 1996], [Scheier & Slaney, 1997], [Tzanetakis & Cook, 1999], [Carey et al., 1999], [El-Maleh et al., 2000], [Zhang & Kuo, 2001], [Lu & Zhang, 2002], [Tzanetakis et al., 2002], [McKinney, 2003], [Panagiotakis & Tziritas, 2005] e [Bergstra et al., 2006].

3.6 Coeficientes Cepstrais

As medidas cepstrais fornecem informação a respeito do formato do espectro. O termo cepstro é usado para designar a transformada de Fourier inversa do logaritmo do espectro de potência de um sinal. Em muitos casos, os próprios coeficientes cepstrais são usados como parâmetros, mas é comum a criação de variáveis derivadas, como o cepstro residual, o qual é obtido comparando-se a versão original do cepstro do sinal com uma versão suavizada.

O cálculo para obtenção dos coeficientes cepstrais é muito extenso. Por isso optamos por não apresentá-lo neste trabalho. Porém, mais informações a respeito deste

parâmetro podem ser encontradas em [Davis & Mermelstein, 1980], [J. R. Deller & Proakis, 1993] e [Theodoridis & Koutroumbas, 2006].

Este parâmetro foi utilizado em [Scheier & Slaney, 1997], [Foote, 1997], [Carey et al., 1999], [Pye, 2000], [Li, 2000], [Tzanetakis et al., 2002], [Burred & Lerch, 2003], [McKinney, 2003], [Guo & Li, 2003] e [Bergstra et al., 2006].

3.7 *Roll-Off*

Este parâmetro determina a freqüência R_i abaixo da qual se concentra 85% da área do espectro de magnitude. Como variantes, existem autores que usam o valor de 90% ou 95% da área do espectro de magnitude ou do espectro de potência. Este parâmetro é calculado conforme:

$$\sum_{k=1}^{R_i} |X_i(k)| = 0,85 \cdot \sum_{k=1}^{K} |X_i(k)|. \tag{3.5}$$

Os sinais de música tendem a apresentar valores maiores para este parâmetro que os sinais de voz. Isto porque os sons percussivos, os transientes de ataque, etc, têm mais energia nas altas freqüências que os sinais de voz. O *roll-off* foi utilizado em [Scheier & Slaney, 1997], [Tzanetakis et al., 2002], [Burred & Lerch, 2003], [McKinney, 2003], [Julien & Pachet, 2003], [Junior et al., 2005] e [Bergstra et al., 2006].

3.8 Sonoridade

Este parâmetro pretende reproduzir a percepção humana em relação à sonoridade de um sinal sonoro. O primeiro passo para o seu cálculo é modelar a resposta em freqüência dos ouvidos médio e externo. Uma das possibilidades para esta resposta é [Thiede, 1999]:

$$W(k) = -0,6 \cdot 3,64 \cdot f(k)^{-0,8} - 6,5 \cdot e^{-0,6 \cdot (f(k)-3,3)^2} + 10^{-3} \cdot f(k)^{3,6}, \tag{3.6}$$

onde a freqüência $f(k)$ é medida em kHz, dada por $f(k) = k \cdot d$, e d é diferença entre duas linhas espectrais consecutivas em kHz.

$W(k)$ é uma função de ponderação que atenua ou enfatiza as componentes espectrais de acordo com o comportamento auditivo. Desta forma, a sonoridade de um quadro i é calculada como:

$$ld_i = \sum_{k=1}^{K} |X_i(k)|^2 \cdot 10^{W(k)/20}. \tag{3.7}$$

Este parâmetro é uma medida objetiva de grande utilidade na identificação dos gêneros musicais e foi utilizado em [Wold et al., 1996], [McKinney, 2003] e [Burred & Lerch, 2003].

3.9 Proporção de Quadros com Pouca Potência

Este parâmetro mede a proporção de quadros cuja potência é menor que metade da potência ao longo de um trecho de 1 segundo do sinal [Scheier & Slaney, 1997]. A potência de um trecho de sinal com M amostras no tempo é dada por:

$$E = \frac{1}{M} \sum_{m=0}^{M-1} |x_i(m)|^2 . \tag{3.8}$$

Normalmente, os sinais de voz possuem uma proporção maior de quadros de baixa potência que sinais de música, pois nestes sinais há mais quadros de silêncio. Logo, este parâmetro possui valores maiores para os sinais de voz, dentro da sua gama de valores. É muito utilizado na diferenciação de sinais musicais e voz e foi usado em [Scheier & Slaney, 1997], [Tzanetakis & Cook, 2002] e [Saunders, 1996].

Capítulo 4

Ferramenta de Classificação

A estratégia de classificação desenvolvida por [Barbedo & Lopes, 2006] é apresentada em detalhes neste capítulo, visando dar suporte à descrição dos estudos realizados sobre esta ferramenta e que são apresentados no Capítulo 5.

Pretendemos uma descrição objetiva da ferramenta, evitando analisá-la ou avaliá-la, o que pode ser feito pelo leitor interessado diretamente em [Barbedo & Lopes, 2006]. Esta postura combina com o objetivo de oferecer apenas o suporte adequado aos estudos que são o objeto desta dissertação.

A taxonomia utilizada por [Barbedo & Lopes, 2006] possui 4 camadas hierárquicas, com três gêneros na primeira camada e 29 gêneros musicais na camada mais refinada da estrutura, como pode ser visto na Figura 4.1. A descrição detalhada de todos os gêneros pode ser encontrada em [Barbedo & Lopes, 2006].

Durante o desenvolvimento da taxonomia os autores procuraram utilizar um grande número de gêneros, mantendo ao mesmo tempo a consistência da estrutura e usando uma nomenclatura que é facilmente identificável pelos usuários e que se encontra na maioria das taxonomias. Desta forma, minimizam-se alguns problemas comumente encontrados nas estruturas hierárquicas presentes na literatura, como a falta de padronização na nomenclatura dos gêneros musicais e o grau de subjetividade das estruturas taxonômicas, como visto no Capítulo 2. Adicionalmente, a ampla gama de gêneros na camada mais baixa é muito importante para a eficácia do classificado, conforme será visto mais adiante.

Assim como a taxonomia, a base de dados também foi construída cuidadosamente pelos autores.

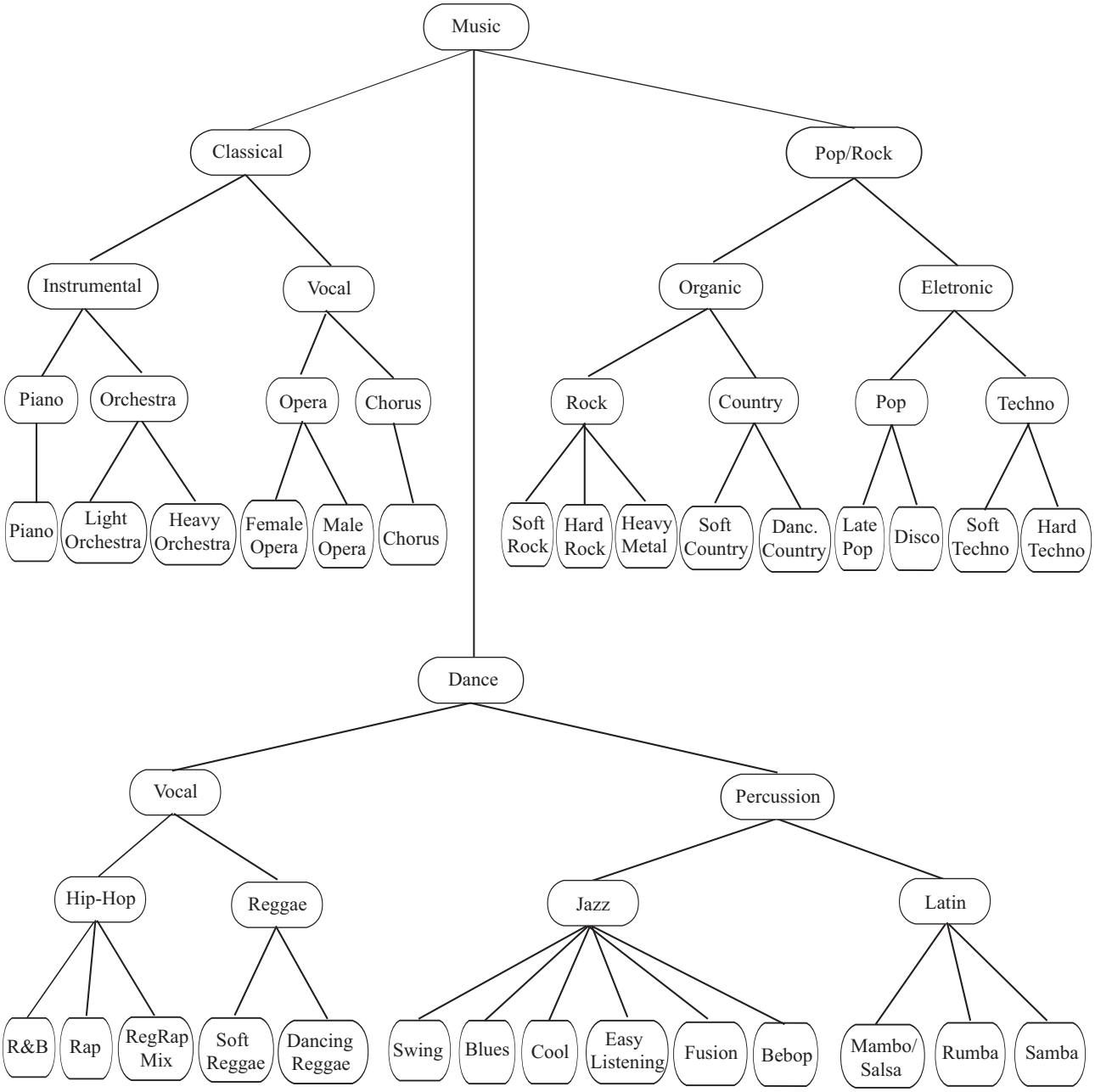

Figura 4.1: Estrutura Taxonômica.

A base de dados completa é composta por 2266 extratos musicais, os quais representam mais de 20 horas de áudio. Cada gênero é representado por pelo menos 40 sinais. Cada sinal tem duração de 32 segundos. A escolha dos 32 segundos foi feita pelos autores através de um programa que buscava os 32 segundos do sinail musical com maior energia. Os sinais foram amostrados a 48 kHz e quantizados com 16 bits. O material de áudio foi extraído de CD (compact disk), rádio da Internet e de arquivos codificados (MP3, WMA, OGG, AAC). A base de dados foi dividida em 580 arquivos destinados ao treinamento e 1686 arquivos destinados à avaliação da ferramenta.

É importante enfatizar que os autores tomaram algumas precauções adicionais para não polarizar os resultados de avaliação da ferramenta. Primeiro, a base de dados foi construída usando-se apenas uma música de cada artista, evitando assim que músicas de um mesmo artista estejam presentes nas bases de treinamento e de teste. Além disso, cada um dos 2266 extratos foi retirado de músicas distintas, o que assegura que a base de treinamento e a teste são completamente diferentes.

Como será visto nas próximas seções, esta ferramenta de classificação apresenta mais dois aspectos importantes, que a diferenciam das demais presentes na literatura. O primeiro é a classificação de baixo para cima ao longo da estrutura taxonômica, isto é, um sinal musical é inicialmente classificado na camada mais refinada e a sua classificação nas camadas superiores é conseqüência desta primeira classificação. A vantagem de se utilizar esta estratégia de classificação é que as diferenças entre os gêneros na camada mais refinada podem ser exploradas de maneira mais eficiente, dada uma ampla gama de gêneros aí presente.

O segundo aspecto é um processo de classificação baseado na comparação entre pares de gêneros. Isto significa que são consideradas todas as possíveis combinações de dois gêneros da camada mais refinada e que a classificação inicial é feita a cada par de gêneros, definindo um gênero vencedor a cada par. A classificação final na camada mais refinada é dada pelo gênero com o maior número de vitórias na classificação por pares.

A partir deste ponto até o final desta dissertação, os gêneros musicais presentes na primeira camada da taxonomia serão chamados de gêneros-alvo e aqueles da camada mais refinada, de gêneros.

A presença dos 29 gêneros na camada mais refinada da taxonomia implica em um grande esforço computacional durante o processo de treinamento. Contudo, a utilização destas duas estratégias - processo de comparação entre os pares de gêneros e a divisão refinada dos gêneros-alvo - mostrou-se

bem sucedida, como será visto ao longo deste capítulo.

Nas próximas seções apresentaremos a descrição desta ferramenta de classificação e os resultados obtidos pelos autores. Em seguida, usando as características descritas, contextualizaremos a proposta de estudo sobre esta ferramenta.

4.1 Ferramenta de Classificação: Detalhes

Inicialmente apresentaremos uma visão geral dos processos de treinamento e de classificação e nas próximas seções serão apresentados os detalhes desses processos.

Antes do processo de treinamento, o sinal musical é dividido em quadros. Em seguida, 4 parâmetros são extraídos individualmente para cada quadro. Tais parâmetros são agrupados em segmentos de análise de 1 segundo. Para cada segmento são calculadas três medidas estatísticas dos parâmetros extraídos: média, variância e prevalência do pico principal, denominadas pelos autores de *summary features* e que compõem o vetor de *summary features*, o qual caracteriza o segmento de análise do sinal.

A prevalência do pico principal é calculada para cada um dos quatro parâmetros de acordo com:

$$P_{ft} = \frac{\max\left[ft\left(i,j\right)\right]}{\left(1/I\right)\sum_{i=1}^{I} ft\left(i,j\right)}, \tag{4.1}$$

onde $ft(i,j)$ é o valor de um parâmetro ft no quadro i do segmento de análise j e I é o número de quadros dentro de cada segmento de análise. Esta *summary feature* objetiva inferir o valor relativo do valor máximo de cada parâmetro em relação à média dos valores do parâmetro.

Durante o processo de treinamento, são calculados os vetores de *summary features* para todos os segmentos dos sinais de cada gênero da base de treinamento. Os vetores de *summary features* são usados para a escolha dos vetores de referência de cada par de gêneros da camada mais refinada. Os vetores de referência são aqueles que propiciam a melhor separação entre os gêneros de um par, segundo a distância Euclidiana. São escolhidos seis vetores de referência a cada par, três para cada gênero, e o processo é realizado para todos os pares possíveis.

Após o treinamento, inicia-se o processo de classificação. Para uma dada música a ser classificada, calcula-se inicialmente o vetor de *summary features* para cada segmento de análise. Em seguida, a

cada segmento, é feito um processo de comparação, através do cálculo da distância Euclidiana, entre o conjunto de vetores de referência de um dado par de gêneros e o vetor de *summary features* a fim de determinar um gênero vencedor para o segmento sob análise.

Este processo é repetido para as demais combinações dos gêneros, tomados dois a dois, e para todos os segmentos de análise do sinal musical, com o intuito de definir o gênero vencedor ao longo de todos os segmentos de análise.

Nas próximas seções apresentaremos os detalhes de cada passo do funcionamento desta ferramenta de classificação.

4.1.1 Extração dos Parâmetros

Antes de iniciar a extração dos parâmetros, o sinal de áudio é dividido em quadros, utilizando uma janela de Hamming de 21,3 milissegundos, com sobreposição de 50% dos quadros consecutivos.

A extração dos parâmetros é feita individualmente para cada quadro. Os parâmetros utilizados pelos autores foram: *roll-off*, sonoridade, largura de faixa e fluxo espectral.

Os parâmetros extraídos para cada quadro são agrupados em segmentos de análise de um segundo. Posteriormente, para cada parâmetro de cada segmento de análise são calculadas três medidas estatísticas: média, variância e prevalência do pico principal, denominadas pelos autores de *summary features*.

Com base nas *summary features*, forma-se o vetor de *summary features*, o qual caracteriza o segmento de análise. Desta forma, cada vetor possui 12 *summary features*, como mostrado na Figura 4.2.

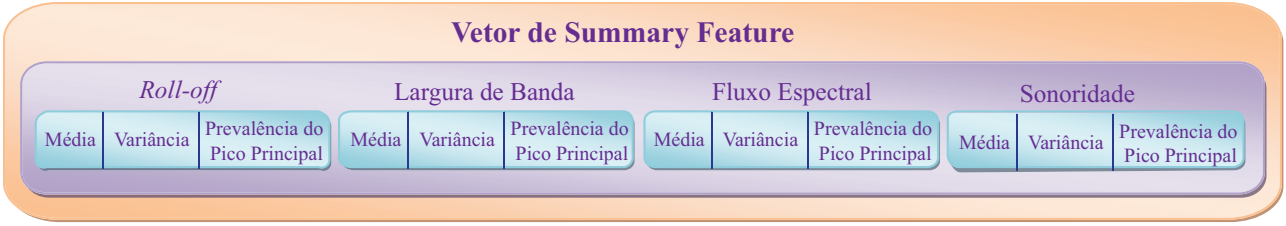

Figura 4.2: Vetor de Summary Feature.

É importante salientar que estas três etapas: processo de extração dos parâmetros, formação dos segmentos de análise e obtenção dos vetores de summary features ocorrem da mesma forma tanto

para o processo de treinamento como para o processo de classificação.

4.1.2 Processo de Treinamento

Como dito anteriormente, o treinamento objetiva a definição dos vetores de referência para casa possível par de gêneros da camada mais refinada.

Antes de explicar como é feito o treinamento, é importante ressaltar que, como cada música tem 32 segundos de duração, de cada sinal musical serão extraídos 32 vetores de *summary features* referentes a cada segmento de análise de 1 segundo do sinal. Os vetores de *summary features* dos sinais musicais que foram destinados ao treinamento são chamados de vetores de treinamento.

O processo de treinamento começa pela determinação do conjunto de vetores candidatos a vetores de referência para cada gênero. Este é um subconjunto dos vetores de treinamento formado pelos vetores que se situam no interior de uma região cuja fronteira corresponde a um certo grau de variância ao redor da média dos vetores de treinamento do gênero considerado. Os vetores candidatos são usados na determinação dos vetores de referência de cada par de gêneros, conforme descrito a seguir.

Após a determinação dos vetores candidatos, a cada par de gêneros é feito um processo de seleção entre os vetores candidatos para a determinação de seis vetores de referência, sendo três de cada gênero do par. Por exemplo, considere os gêneros A e B. O processo de comparação é feito calculando-se a distância Euclidiana entre seis vetores candidatos previamente escolhidos e todos os demais vetores de treinamento de ambos os gêneros. Após o cálculo da distância, cada vetor de treinamento é rotulado com o gênero do vetor candidato mais próximo. A seguir, é feita a troca dos seis vetores de candidatos, repete-se o cálculo da distância e a determinação do rótulo dos vetores de treinamento do gênero vencedor. Ao final, o conjunto dos seis vetores candidatos que apresentou o maior número de classificações corretas é considerado o conjunto de vetores de referência para o par de gêneros A e B.

O processo de comparação entre pares e determinação dos vetores de referência é repetido para todas as possíveis combinações dos gêneros dois a dois. Assim, ao final do processo, cada possível par de gêneros da camada mais refinada é caracterizado por seis vetores de referência, três para cada gênero do par.

4.1.3 Processo de Classificação

O processo de classificação começa pela extração do vetor de *summary features* referente a cada segmento de análise de 1 segundo da música que se pretende classificar.

Em seguida é feito um processo de comparação entre pares de gêneros. Por exemplo, suponha que desejamos classificar um sinal musical de 5 segundos, como mostrado na Figura 4.3. Para o 1^o segmento de análise é feito o processo de comparação entre o seu vetor de *summary features* e os vetores de referência correspondente a cada par de gêneros, através do cálculo da distância Euclidiana, Figura 4.3.

Para uma melhor compreensão, vamos considerar inicialmente apenas os vetores de referência do par de gêneros *A* e *B*. Escolhemos o vetor de referência do gênero *A* ou *B*, que estiver mais próximo do vetor de *summary features* do segmento de análise. Neste caso, o vetor de referência mais próximo é aquele sombreado na Figura 4.3 (a), o qual pertence ao gênero *A*. Portanto, o rótulo deste vetor de referência é considerado o gênero vencedor para este par de gêneros.

Este processo é repetido para as demais combinações dos pares de gêneros (ver Figura 4.3 (a)). Desta forma, para cada processo de comparação entre pares haverá um gênero vencedor. Após este processo de comparação, é determinado o número de vezes que cada gênero venceu no primeiro segmento de análise, Figura 4.3 (b). O gênero que obteve o maior número de vitórias é o gênero vencedor do segmento. No exemplo, foi o gênero *D*.

Os processos de comparação e determinação do gênero vencedor do segmento de análise são repetidos para todos os demais segmentos do sinal musical a ser classificado (Figura 4.3 (c)).

Após a classificação de cada segmento de análise do sinal, determina-se quantas vezes cada gênero venceu ao longo dos segmentos de análise. Por fim, o gênero musical do sinal é determinado pelo gênero que obteve o maior número de vitórias ao longo de todos os segmentos de análise. Para este exemplo o gênero vencedor ao longo de todos os segmentos foi o *D* (Figura 4.3 (d)).

De acordo com o procedimento anterior, o sinal musical é primeiramente classificado na camada mais refinada da estrutura taxonômica. A classificação nos gêneros nas camadas superiores é uma conseqüência desta primeira classificação. Por exemplo, na Figura 4.1 uma música é classificada primeiramente como sendo do gênero *Rap*. Subindo para as camadas superiores podemos constatar que ela pertence ao gênero *Hip-Hop*, acima do qual se encontra o gênero *Vocal* e, por fim, o gênero-

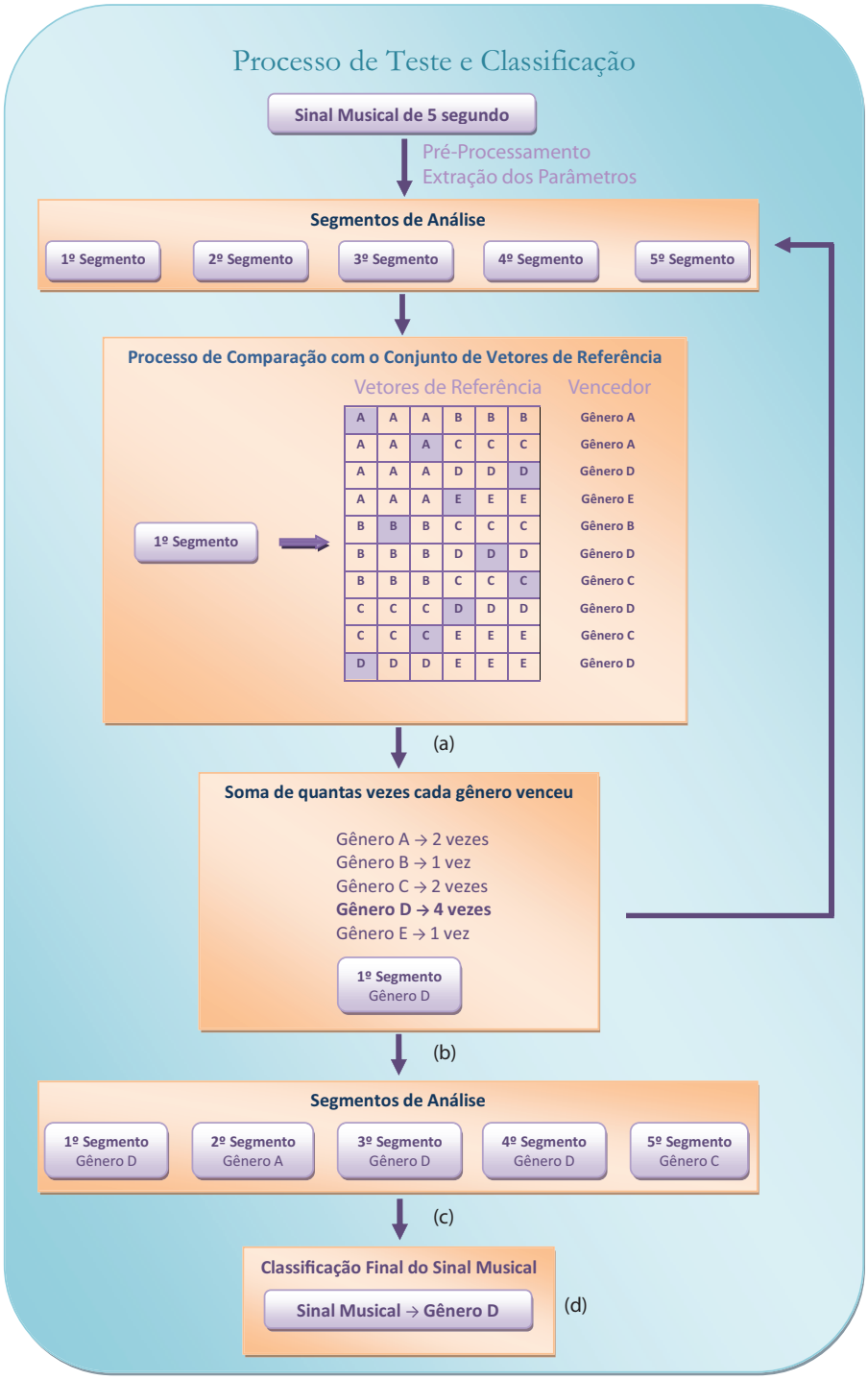

Figura 4.3: Processo de Treinamento.

alvo desta música é o *Dance*.

4.1.4 Resultados

Nesta seção serão apresentados os resultados obtidos por esta ferramenta de classificação ao ser aplicada na parcela da base de dados reservada para a avaliação do método.

Na Figura 4.4 pode ser vista a estrutura taxonômica com os índices de acerto para cada gênero ao longo das camadas, até os gêneros-alvo *Classical*, *Pop/Rock* e *Dance*, na primeira camada. A Tabela 4.1 apresenta a média de acerto em cada camada.

Camadas	Acertos
1^a Camada	87%
2^a Camada	80%
3^a Camada	72%
4^a Camada	61%

Tabela 4.1: Índice de Acerto em cada camada.

Analisando os resultados podemos afirmar que a divisão refinada dos três gêneros-alvo em 29 gêneros na camada mais refinada e o processo de comparação entre os pares de gêneros se mostrou bem sucedida, uma vez que propiciou um índice médio de acerto para a primeira camada de 87%. Este índice representa um desempenho notável no contexto do estado da arte a nível internacional.

4.1.5 Considerações sobre a proposta

Conforme evidenciado nas descrições anteriores, o sucesso do classificador em questão se assenta em duas de suas características inéditas. A primeira é a taxonomia com ampla gama de gêneros na camada mais baixa e a classificação inicial nesta camada, explorando a diferenciação refinada entre seus gêneros. A segunda é a classificação por pares de gêneros desta camada, usando o conceito de vetores de referência.

Se, por um lado, estas características são notáveis pelo excelente desempenho que propiciam, por outro lado, suscitam algumas questões que demandam estudos adicionais.

Em primeiro lugar, a necessidade de uma ampla gama de gêneros na camada mais baixa implica em estruturas taxonômicas grandes e, aparentemente, invalida a aplicação da ferramenta em taxono-

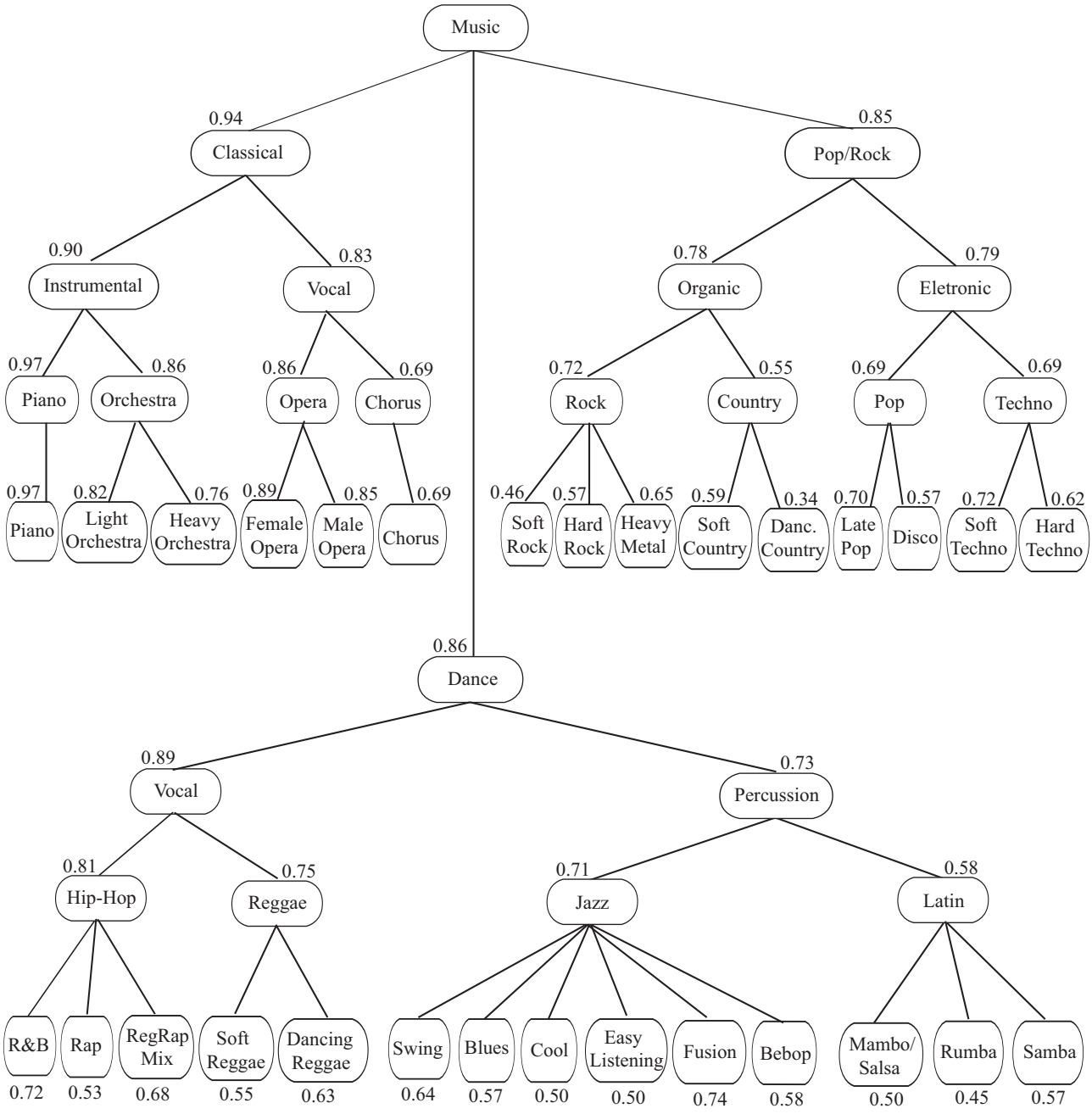

Figura 4.4: Estrutura taxonômica com os índices de acerto.

mias menores, com número reduzido de gêneros na camada mais baixa. No entanto, existem situações em que se pode desejar estruturas menores, como, por exemplo, para atender usuários com preferências ou interesses mais restritos ou mais especializados. Em tais casos, o uso de estruturas menores torna a classificação mais objetiva e pode levar a uma redução significativa do esforço computacional para o treinamento, ao permitir uma redução da base de dados de treinamento e do número de pares possíveis. Torna-se então necessário avaliar o desempenho do classificador após a sua adaptação para a operação com outros tipos de taxonomia.

Como segunda questão a merecer um estudo adicional, no contexto do processo de treinamento, temos o esforço computacional demandado pela definição dos vetores de referência a cada par de gêneros. Esta definição exige, em primeiro lugar e a cada gênero, a escolha do subconjunto dos vetores candidatos a vetores de referência. Tal escolha exige o cálculo da média e variância dos vetores de *summary features* do gênero e a seleção dentre estes últimos, daquele cuja variância é menor que um valor pré-estabelecido. Em segundo lugar e a cada par de gêneros, é necessário considerar todas as possíveis combinações de seis vetores candidatos e para cada combinação fazer a classificação de todos os vetores de *summary feature* do par. Portanto, é evidente que o número de vetores candidatos é crítico para a determinação do esforço computacional necessário para todas estas operações. Desta forma, é relevante pesquisar qual o número mínimo de vetores candidatos que assegure o desempenho apresentado pela proposta original. Adicionalmente, é também interessante modificar o processo de seleção dos vetores candidatos, objetivando reduzir o volume de cálculo necessário e, simultaneamente, assegurar o uso de um número pré-definido de candidatos, independente das características estatísticas dos vetores de *summary features*.

O principal objetivo desta dissertação é o desenvolvimento de estudos abordando as duas questões anteriores e este é o objetivo do próximo capítulo.

Capítulo 5

Estudos sobre a Ferramenta de Classificação

Abordaremos aqui as duas questões apresentadas ao final do Capítulo 4, relativas ao classificador proposto em [Barbedo & Lopes, 2006]. Uma das questões explicita a necessidade de avaliarmos o desempenho do classificador quando aplicado a estruturas taxonômicas menores que a original, particularmente quanto ao número de gêneros na camada mais baixa. A segunda questão propõe a investigação de um novo procedimento para a definição dos vetores de referência para cada par de gêneros da camada mais refinada da taxonomia. O objetivo é simplificar o processo de escolha dos vetores candidatos a vetores de referência e definir um número de candidatos que assegure o desempenho original do classificador. O sucesso nestes tópicos assegurará uma redução do esforço computacional demandado pelo processo de treinamento do classificador. Iniciamos este capítulo abordando a questão taxonômica e, em seguida, apresentamos os estudos relativos à segunda questão.

Após a concretização destes últimos estudos, constatamos que o índice de acerto para alguns gêneros na camada mais baixa foi inferior àqueles típicos para os demais gêneros. Este fato motivou a realização de novos ensaios para investigar se a causa destes desempenhos inferiores era a qualidade dos respectivos vetores de referência. Os resultados destes ensaios permitiram conclusões sobre a robustez do novo procedimento para escolha dos vetores de referência.

5.1 Estrutura Taxonômica

A nova taxonomia foi escolhida seguindo alguns critérios. Em primeiro lugar, optamos por usar parte da taxonomia original, mantendo a nomenclatura. Isto permitiu manter as definições de árvores e gêneros apresentados em [Barbedo & Lopes, 2006]. Mais importante que isto, podemos usar a base de dados original, o que representa um procedimento seguro tendo em conta as comparações de desempenho pretendidas. Após esta decisão, resta apenas selecionar parcelas da taxonomia original e separar as parcelas correspondentes da base de dados, adequadamente à nova taxonomia.

Optamos por uma estrutura com duas camadas, três gêneros-alvo na camada superior e oito gêneros na camada mais refinada. A escolha dos gêneros foi feita segundo os seguintes critérios. Os gêneros-alvo devem conter um gênero com características bem definidas, como, por exemplo, o *Classical*, de modo a oferecer condições favoráveis para um bom desempenho do classificador. Por outro lado, os dois outros gêneros-alvo devem, ao contrário, apresentar características semelhantes, de modo a criar dificuldades para o classificador. Para estes escolhemos o *Techno* e o *Rock*. A estrutura em árvore a partir destes três gêneros-alvo foi composta com partes da estrutura original, conforme especificado na Figura 5.1.

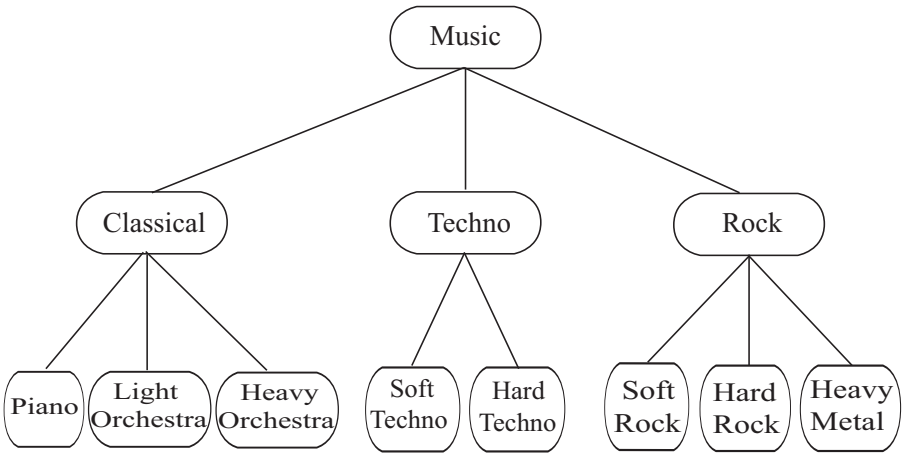

Figura 5.1: Estrutura Taxonômica.

Definida a nova taxonomia, o próximo passo seria o treinamento do classificador original e a análise de seu desempenho nesta nova condição. Entretanto, visando uma economia de esforços, optamos por realizar este procedimento em conjunto com as investigações acerca de um novo processo de seleção de vetores candidatos a vetores de referência. Embora este procedimento acarrete a mis-

tura dos efeitos de duas transformações e, portanto, o risco de falsas conclusões, os resultados das próximas seções mostram que a opção é viável para o problema em questão. Sendo assim, os resultados e conclusões sobre a aplicação do classificador em uma taxonomia reduzida serão apresentados no final do capítulo.

5.2 Novo Procedimento de Escolha dos Vetores Candidatos

Iniciamos recordando que o processo de treinamento começa com o cálculo dos vetores de *summary features* para cada gênero da camada mais refinada. Estes vetores são aqui denominados vetores de treinamento. O próximo passo é a seleção, a cada gênero, de um subconjunto de vetores, de treinamento, constituindo assim os vetores candidatos a vetores de referência. O critério adotado em [Barbedo & Lopes, 2006] para tal seleção é a escolha dos vetores com variância menor que um valor pré-definido, tendo como referência a média dos vetores de treinamento do gênero.

Ressaltamos dois aspectos deste procedimento. O primeiro é o esforço computacional necessário para a seleção e o segundo é a impossibilidade de conhecermos a priori o número de candidatos que serão escolhidos, uma vez que tal número depende das características estatísticas dos vetores de treinamento de cada gênero.

Prosseguindo, recordamos que os vetores de referência, a cada par de gêneros, são escolhidos testando-se todas as combinações de seis vetores candidatos, três de cada gênero. Como não conhecemos a priori o número de candidatos, também não conhecemos o número de combinações a serem testadas. Não podemos, então, prever o esforço computacional.

A nossa proposta de procedimento alternativo para a seleção de candidatos e, portanto, de combinações, é simples: selecionamos aleatoriamente um número pré-definido de combinações dos vetores de treinamento, a cada par de gêneros. Pretendemos ainda definir um número mínimo de combinações que assegure o desempenho original do classificador.

Passamos agora a explicar este proposta em seus detalhes. Para auxiliar tal explicação, vamos considerar 2 gêneros, A e B. Para cada gênero geramos uma matriz composta por todos os vetores de treinamento, como pode ser visto na Figura 5.2 (a). A matriz $\mathbf{A}$ tem dimensões $I \times J$, onde I é o número de linhas, referente ao número de vetores de treinamento do gênero A, e J é o número de

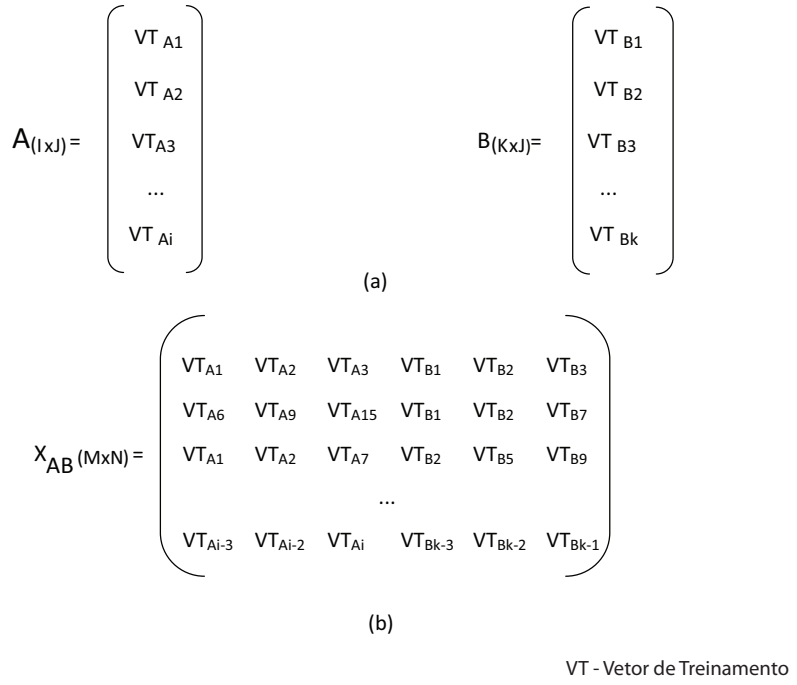

Figura 5.2: Formação da Matriz Aleatória.

colunas, cada qual composta pelas 12 *summary features* de cada vetor de treinamento. A matriz $\mathbf{B}$ é similar à matriz $\mathbf{A}$, embora o número K de linhas seja, em geral, diferente de I.

A seleção aleatória das combinações dos vetores de treinamento dos gêneros A e B é resumida na matriz $\mathbf{X}_{AB}$, Figura 5.2 (b). Esta matriz tem M linhas, onde M é o número pré-definido de combinações dos vetores de treinamento para o par A , B. Cada linha contém uma combinação escolhida aleatoriamente. Logo, é composta por seis vetores de treinamento, três de cada gênero do par A, B. Os seis vetores são escolhidos de forma aleatória usando o comando *rand* do Matlab®, o qual gera números aleatórios obedecendo a uma distribuição uniforme entre 1 e o número total de vetores de treinamento de cada gênero. Estes números aleatórios sorteiam os vetores de treinamento de cada gênero do par.

A próxima seção apresenta a descrição dos experimentos realizados para o novo procedimento de escolha dos vetores candidato.

5.2.1 Descrição dos Experimentos Realizados

Nos ensaios que realizamos com esta proposta, usamos, a cada par de gêneros, um número de combinações entre 5.000 e 150.000, limites estes, definidos pelos resultados dos experimentos. Para contextualizar tais limites, vamos avaliar o número típico de vetores de treinamento de cada gênero e o número de combinações possíveis a cada par de gêneros.

A base de dados usada nos experimentos foi a mesma desenvolvida por [Barbedo & Lopes, 2006]. Desta forma, cada gênero da última camada tem pelo menos 40 músicas típicas, com duração de 32 segundos cada. A base de dados foi dividida em $1/3$ para treinamento e $2/3$ para teste. Sendo assim, cada gênero possui pelo menos 416 vetores de treinamento ($40 \cdot (1/3)$ músicas$\times 32$ segundos$=416$ vetores) e 864 vetores de teste ($40 \cdot (2/3)$ músicas$\times 32$ segundos$=864$ vetores). O número mínimo de 416 vetores de treinamento produz um total de $1,40 \cdot 10^{14}$ ($(C_3^{416})^2$) combinações possíveis a cada par de gêneros.

Nos testes para avaliação da eficácia da seleção aleatória de combinações, usamos inicialmente 5.000 combinações, ou seja, foram geradas matrizes X com 5.000 linhas, a cada par de gêneros. Em seguida, realizamos o treinamento e a classificação. Este processo foi repetido seis vezes, cada qual com novas matrizes X para todos os pares.

O índice médio de acerto para o gênero-alvo *Classical* foi de 98%, o qual é semelhante àquele alcançado pelo classificador original, conforme pode ser observado na Figura 4.4. Entretanto, os índices médios de acerto para os outros dois gêneros-alvo, *Rock* e *Techno*, foram 32% e 36%, respectivamente. Estes índices são muito inferiores àqueles apresentados pelo classificador original.

Estes desempenhos ruins motivaram a realização de novos experimentos usando números maiores de combinações.

Realizamos 12 experimentos distintos, todos com 10.000 combinações e obtivemos resultados semelhantes ao caso anterior. Com 20.000 combinações, conseguimos índices máximos de acerto de 43% e 54% para os gêneros-alvo *Rock* e *Techno*, respectivamente. Para o *Classical* conseguimos o mesmo valor médio anterior de 98%. Apesar da melhora, os resultados ainda são inferiores àqueles oferecidos pelo classificador original.

Os índices de acerto para o *Rock* e *Techno* somente alcançaram níveis semelhantes àqueles da Figura 4.4 quando empregamos 50.000 combinações. Neste caso foram realizados 19 experimentos

e os índices médios para o *Classical*, *Rock* e *Techno* foram de 98%, 80% e 84%, respectivamente.

Estes últimos índices não se alteraram em vários outros experimentos com números de combinações maiores que 50.000, até o valor máximo testado de 150.000 combinações.

A próxima seção apresenta os detalhes dos resultados com 50.000 combinações e algumas conclusões preliminares.

5.2.2 Resultados e Conclusões Preliminares

Para 50.000 combinações e a taxonomia da Figura 5.1, o tempo médio de treinamento da ferramenta foi de 8 horas, usando um programa no ambiente Matlab® sendo executado em um computador com um processador Core 2 Duo, 1 GB de memória RAM e com o Windows XP®.

Com relação aos gêneros presentes na camada mais refinada o melhor resultado obtido, em porcentagem, pode ser visto na Tabela 5.1, a qual apresenta a matriz confusão associada aos experimentos.

Gêneros	Piano	Light Orchestra	Heavy Orchestra	Soft Rock	Hard Rock	Heavy Metal	Soft Techno	Hard Techno
Piano	0,90	0,10						
Light Orchestra	0,08	0,77	0,15					
Heavy Orchestra	0,14	0,27	0,55	0,04				
Soft Rock			0,02	0,65	0,13	0,05	0,13	0,02
Hard Rock				0,11	0,36	0,28	0,10	0,15
Heavy Metal				0,05	0,24	0,58	0,05	0,08
Soft Techno				0,06		0,06	0,44	0,44
Hard Techno				0,06	0,03	0,06	0,19	0,66

Tabela 5.1: Melhor índice de acerto para a camada mais refinada com 50.000 combinações.

A primeira coluna nos mostra os gêneros e a primeira linha mostra os gêneros estimados durante a classificação. Considerando a primeira linha da Tabela 5.1 (*Piano*) como exemplo, podemos observar que 90% das músicas do gênero *Piano* foram classificadas corretamente e 10% foram classificadas

como *Light Orchestra*. Portanto, a diagonal principal da Tabela 5.1 apresenta as classificações corretas, e todos os valores que estão fora da diagonal principal representam os erros de classificação.

A Tabela 5.2 mostra a matriz de confusão para os gêneros-alvo, com as mesmas características da Tabela 5.1. Os resultados da Tabela 5.2 permitem calcular um índice médio de acerto para os gêneros-alvo, o qual atinge 88% e corresponde aos melhores resultados obtidos para 50.000 combinações. Comparando este índice com aquele da Tabela 4.1, para a primeira camada, constatamos que o sistema sob teste neste capítulo atingiu o mesmo desempenho que o classificador original proposto por [Barbedo & Lopes, 2006].

Gêneros	Classical	Rock	Techno
Classical	0,98	0,02	
Rock	0,01	0,81	0,18
Techno		0,14	0,86

Tabela 5.2: Melhor índice de acerto para os gêneros-alvo, com 50.000 combinações.

É interessante agora comparar estes resultados ótimos com os piores obtidos nas mesmas condições e que estão mostrados na Tabela 5.3. Como a diferença entre os resultados afins nas duas tabelas não é significativa, concluímos que os índices médios de acerto com 50.000 combinações mencionados no final da seção 5.2.1 (98% (*Classical*), 80% (*Rock*) e 84% (*Techno*)) são representativos para o desempenho do classificador após as duas alterações realizadas: redução de taxonomia e seleção aleatória de combinações de vetores de treinamento com 50.000 combinações.

Gêneros	Classical	Rock	Techno
Classical	0,97	0,03	
Rock	0,05	0,77	0,18
Techno		0,19	0,81

Tabela 5.3: Pior índice de acerto para os gêneros-alvo, com 50.000 combinações.

O fato de que este desempenho médio com 50.000 combinações é similar àquele desempenho original do classificador permite uma conclusão importante no contexto deste capítulo. Podemos considerar cada uma das duas alterações realizadas de forma individual e afirmar que o desempenho original é alcançado quando cada uma delas é aplicada isoladamente. Esta conclusão justifica a decisão tomada no início deste capítulo, quando optamos por analisar o desempenho resultante aplicando as duas alterações simultaneamente.

As análises anteriores permitem explicar algumas conclusões preliminares, antes que aprofundemos a análise dos resultados já apresentados.

Em primeiro lugar podemos afirmar que o classificador proposto em [Barbedo & Lopes, 2006] pode ser aplicado a taxonomias menores que a original até a envergadura daquela em uso neste capítulo.

Em segundo lugar, o procedimento de seleção de combinações proposto não reduz o desempenho do classificador original. Associada a esta conclusão, podemos afirmar que o número de combinações aleatórias a serem testadas deve ser da ordem de 50.000. Números inferiores a este podem acarretar redução de desempenho em alguns gêneros e números superiores a este não só não propiciam melhores resultados, como aumentam desnecessariamente o esforço computacional. Concluímos relembrando que não é possível comparar este número de combinações com aqueles empregados na versão original do classificador. Também, o número total de combinações possíveis a cada par de gêneros é cerca de 140 trilhões.

Continuando a analisar o número de combinações necessárias, concluímos que podemos usar um número menor que 50.000 combinações para o treinamento dos gêneros associados ao gênero-alvo *Classical*. Isto porque, os resultados finais da Tabela 5.1 já haviam sido alcançados com apenas 5.000 combinações. Este fato ocorre, pois estes gêneros possuem características próprias e bem definidas, variando relativamente pouco de uma música para outra.

O mesmo não acontece na diferenciação entre o *Rock* e o *Techno*, pois estes gêneros-alvo e suas derivações possuem algumas características relativamente semelhantes, as quais dificultam a classificação quando usamos os quatro parâmetros como elementos de representação e diferenciação destas características. Como conseqüência, foram necessárias 50.000 combinações para encontrarmos vetores de referência robustos para a separação entre tais gêneros. Estas características comuns entre o *Rock* e o *Techno*, dificultando a sua diferenciação, motivaram a escolha destes dois gêneros para comporem a taxonomia reduzida, conforme já mencionado no início deste capítulo.

As facilidades e as dificuldades antes analisadas para a separação entre os gêneros-alvo são percebidas com clareza ao analisarmos os resultados da Tabela 5.2. Constatamos percentuais de 2% de erro nas classificações associadas ao *Classical* e tais erros são em direção apenas do *Rock*. O mesmo acontece com o *Rock* e o *Techno* em direção ao *Classical*. Portanto, os elementos objetivos de difer-

enciação entre os gêneros-alvo do classificador são robustos para a separação bidirecional entre os dois blocos *Classical* e *Rock/Techno*.

O mesmo não ocorre entre os gêneros-alvo *Rock* e *Techno*, uma vez que ocorrem erros de 14% e 18%, corroborando as observações anteriores sobre a relativa dificuldade de discernimento entre estes dois gêneros-alvo.

Continuando esta análise com os resultados da Tabela 5.1, podemos focar agora nos erros entre os gêneros de uma família associada a um gênero-alvo. Tais erros não são tão graves no contexto do classificador proposto em [Barbedo & Lopes, 2006], uma vez que o foco é o desempenho em termos de gêneros-alvo e os erros dentro de uma mesma família não produzem erros na camada mais alta. Apesar disto, é interessante analisar os erros internos nas famílias.

Começando com o *Classical*, encontramos percentuais de erros de 10%, 23% e 41%. Para o *Rock* temos 18%, 29% e 39%, e para o *Techno*, 19% e 44%. Observamos, assim, um comportamento relativamente uniforme para todas as famílias. Entretanto, destacamos os índices de acerto dos gêneros *Hard Rock* e *Soft Techno*, os quais são inferiores aos valores correspondentes para os demais gêneros, conforme mostram os números da diagonal principal da matriz de confusão. Estes índices destacados motivaram uma pesquisa adicional para averiguarmos se as discrepâncias para o *Hard Rock* e *Soft Techno* são devidas à qualidade dos respectivos vetores de referência selecionados pelo novo processo de seleção de combinações.

A próxima seção apresenta a descrição destes experimentos adicionais na tentativa de otimizar os vetores de referência dos pares de gêneros onde ocorre a participação do *Hard Rock* ou do *Soft Techno*.

5.3 Otimização dos Vetores de Referência

Para facilitar a explicação dos procedimentos adotados neste processo de otimização, vamos inicialmente considerar um exemplo ilustrativo artificial envolvendo a classificação de um total de três gêneros, A, B e C, da camada mais refinada. Consideramos inicialmente o processo de geração dos vetores de referência e os resultados de classificação conforme já descritos neste capítulo. Propositalmente, criamos resultados onde o índice de acerto para o gênero B foi inferior aos demais. Criada

então a situação similar à da Tabela 5.1, prosseguimos explicando o processo de geração de novos vetores de referência para os pares de gêneros onde existe a participação de B.

Primeiro, realizamos o processo de treinamento com a definição de três matrizes $\mathbf{X}$, uma para cada par de gêneros, conforme mostra a Figura 5.3 (a). Após a escolha dos vetores de referência para cada par de gêneros, obtemos a matriz de referência $\mathbf{M}_{Ref}$, Figura 5.3 (b). Por fim, é realizado o processo de classificação, resultando nos índices de acerto mostrado na Figura 5.3 (c), onde conforme já mencionado, optamos por um índice de acerto inferior para o gênero B.

A situação gerada no exemplo motiva a otimização do conjunto de vetores de referência dos pares de gêneros onde há a participação de B. Os novos conjuntos serão obtidos através da geração das duas novas matrizes aleatórias $\mathbf{X}_{AB}$ e $\mathbf{X}_{BC}$ mostradas na Figura 5.3(d), cada uma destas matrizes contém um conjunto de vetores candidatos diferente daquele inicialmente mostrado na Figura 5.3 (a).

Os novos conjuntos de vetores de referência geram a nova matriz $\mathbf{M}_{Ref}$ mostrada na Figura 5.3 (e). Em seguida, o processo de classificação é repetido gerando os resultados hipotéticos mostrados na Figura 5.3 (f).

Como conclusão deste exemplo ilustrativo, o processo de otimização dos vetores de referência produziu uma melhora no índice de acerto para *B*, mas uma piora para nos índices de acerto dos outros dois gêneros. Este comportamento será observado nos experimentos a serem descritos na próxima seção e pode ser explicado no contexto deste exemplo artificial. A degradação dos índices de acerto dos gêneros A e C após o processo de otimização para melhorar o índice de B, pode ser entendida lembrando inicialmente que alteramos os vetores de referência de todos os pares onde há a participação de B - no nosso caso, dos pares AB e BC. Tal alteração afeta as decisões classificatórias para cada gênero de um par. Por exemplo, tomando o par AB, escolhemos um novo conjunto de vetores de referência para este par, o qual melhora o índice de acerto geral para B. Logo, este novo conjunto de vetores de referência pode provocar uma polarização no contexto de AB com privilégio para B, de modo que uma música que antes era classificada, corretamente, como A no contexto de AB, pode agora ser classificada, erradamente, como B.

Este mesmo mecanismo pode ocorrer para BC.

Este fenômeno de degradação após a otimização dos vetores de referência, será observado nos experimentos a serem descritos na próxima seção. Entretanto, o grau de complexidade dos mecanismos

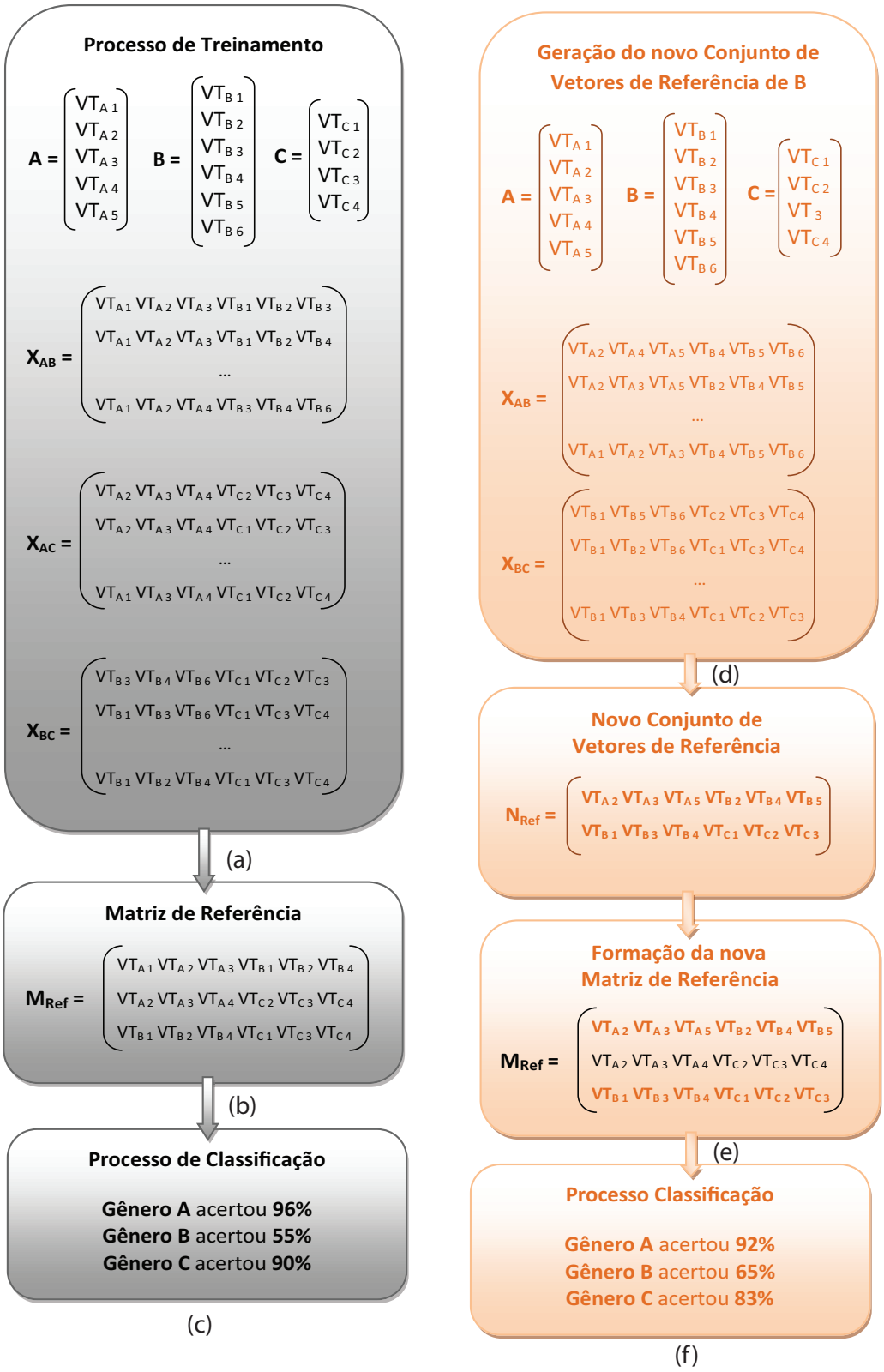

Figura 5.3: Otimização dos vetores de treinamento.

envolvidos é muito maior devido ao maior número de gêneros e, portanto de pares de gêneros.

As idéias básicas apresentadas no contexto deste exemplo artificial norteiam os experimentos para a otimização dos vetores de referência dos gêneros *Soft Techno*, *Hard Rock* e outros, os quais fazem parte da próxima seção.

5.3.1 Resultados da Otimização dos vetores de Referência

Antes de apresentarmos os resultados, cabe ressaltar que usamos números de combinações maiores que 50.000. A conclusão já anunciada de que este número é suficiente para a reprodução dos resultados originais em [Barbedo & Lopes, 2006], se refere ao desempenho em termos de gêneros-alvo. Entretanto, constatamos que para os testes específicos de otimização foi necessário considerar até 150.000 combinações. Mais adiante voltamos a este assunto para uma conclusão definitiva.

Na otimização referente ao gênero *Soft Techno*, foram gerados seis novos conjuntos de vetores de referência para 50.000, 70.000 e 150.000 combinações, totalizando 18 experimentos. Os resultados de todos os experimentos foram similares, com índices de acerto ao redor de 48%. Portanto, verificamos que houve apenas uma pequena melhora de 4%. Adicionalmente, observamos que os novos conjuntos de vetores de referência também provocaram uma modificação no índice de acerto de outros gêneros, como o *Hard Rock* e *Soft Rock*, os quais pioraram em média 5%.

Para o caso do *Hard Rock* foram realizados 14 experimentos, com sete novos conjuntos de vetores para 50.000 combinações e sete para 90.000. Ao contrário do que ocorreu no caso do *Soft Techno*, não conseguimos melhorar o índice de acerto para o gênero *Hard Rock*. Estes dois testes sugerem que o processo de otimização dos vetores de referência não produz resultados interessantes, principalmente tendo em conta o esforço computacional demandado por este processo. Portanto, temos uma indicação preliminar que tal procedimento não deve se adotado quando da eventual tentativa de aplicação do classificador em novas situações em termos de taxonomia.

Não obstante desta conclusão preliminar, prosseguimos com testes semelhantes para outros gêneros. Apresentamos aqui apenas os melhores resultados conseguidos, os quais englobam aqueles conseguidos para o *Soft Techno*, já apresentados, mais outros referentes aos gêneros *Soft Rock*, *Light Orchestra* e *Heavy Orchestra*. Assim, após a obtenção dos melhores vetores de referência para os gêneros citados, compusemos uma nova matriz de referência, M_{Ref}, usando todos estes novos ve-

tores de referência. Em seguida realizamos um novo processo classificatório, obtendo os resultados da Tabela 5.4.

Gêneros	Piano	Light Orchestra	Heavy Orchestra	Soft Rock	Hard Rock	Heavy Metal	Soft Techno	Hard Techno
Piano	1							
Light Orchestra	0,15	0,85						
Heavy Orchestra	0,05	0,27	0,68					
Soft Rock			0,01	0,84	0,04	0,04	0,07	
Hard Rock				0,15	0,32	0,25	0,17	0,11
Heavy Metal				0,08	0,18	0,58	0,08	0,08
Soft Techno				0,06		0,06	0,50	0,38
Hard Techno				0,03	0,03	0,09	0,25	0,60

Tabela 5.4: Melhor resultado obtido para a otimização dos vetores de refrência.

Devemos agora comparar os resultados da Tabela 5.4 com aqueles da Tabela 5.1. Constatamos que o processo de otimização melhorou os índices de acerto dos gêneros *Piano*, *Light Orchestra*, *Heavy Orchestra*, *Soft Rock* e *Soft Techno*. Por outro lado, o mesmo processo piorou os índices do gênero *Hard Rock* e *Hard Techno*, e manteve inalterado o índice do gênero *Heavy Metal*.

Analisando detalhadamente os casos em que os índices de acerto melhoraram, constatamos que estes ganhos se deveram preponderantemente à redução de erros entre gêneros dentro de uma mesma família, definida pelo gênero-alvo correspondente. Dizendo de outra forma, não obtivemos redução significativa de erros entre gêneros-alvo, particularmente para os casos *Rock* e *Techno*. Estes fatos podem ser facilmente comprovados através da Tabela 5.5, a qual apresenta os resultados classificatórios em termos de gêneros-alvo, tendo como base os resultados da Tabela 5.4. Ao compararmos os resultados das Tabelas 5.2 e 5.5, constatamos que os percentuais de erros para os gêneros-alvo *Rock* e *Techno* praticamente não se alteraram. O caso do *Classical* não é relevante neste contexto pois seus erros para os outros gêneros-alvo são pouco significativos.

Estas constatações atestam que o processo de otimização dos vetores de referência produz resultados limitados. Em primeiro lugar, não se consegue melhorar o índice de acerto de todos os gêneros

Gêneros	Classical	Rock	Techno
Classical	1		
Rock		0,83	0,17
Techno		0,14	0,86

Tabela 5.4: Melhor resultado obtido para os gêneros-alvo.

e, nos casos em que ocorre melhoria, há uma correspondente redução do índice de acerto de outros gêneros. Em segundo lugar, quando se consegue melhorias, estas ocorrem pela redução dos erros intra-família, mantendo praticamente inalterados os índices de erros entre gêneros-alvo. Por fim, o esforço computacional exigido por tal processo de otimização é exageradamente elevado frente aos benefícios proporcionados.

Como conclusão final sobre o processo de otimização dos vetores de referência, afirmamos que o mesmo é desnecessário, uma vez que o objetivo do classificador em questão é a classificação nos gêneros-alvo, a qual não se beneficia de reduções de índices de erros intra-família.

Como conseqüência desta conclusão e de uma comparação entre os resultados para os gêneros *Classical*, *Pop/Rock* e *Dance* da Figura 4.4 e os resultados da Tabela 5.5, podemos afirmar que o novo processo de escolha de combinação aqui proposto mantém o desempenho original do classificador, conforme proposto em [Barbedo & Lopes, 2006]. Adicionalmente, no contexto das especificações adotadas neste trabalho, é suficiente usar 50.000 combinações de vetores de treinamento a cada par de gêneros, uma vez que os vetores de referência resultantes deste processo de seleção aleatória de combinações são robustos para assegurar o nível original de desempenho do classificador, e não são passíveis de otimização. Para mitigar a possibilidade de um resultado atipicamente ruim, sugerimos a realização de dois ou três experimentos com 50.000 combinações.

Podemos analisar agora as razões que impossibilitaram a redução do nível de 15% de erros entre os gêneros-alvo *Rock* e *Techno*. Conforme já mencionado na seção 5.1, estes gêneros apresentam similaridade que dificultam sua diferenciação através das comparações entre vetores de *summary features* e os vetores de referência. Como os vetores de referência são escolhidos entre os vetores de *summary features* gerados no processo de treinamento, concluímos que todo o processo classificatório se assenta sobre as características dos vetores de *summary features*. Logo, temos uma primeira constatação, indicando que estes últimos vetores não dispõem de elementos capazes de diferenciar

plenamente os sinais musicais pertencentes aos dois gêneros-alvo.

Podemos aprofundar mais esta análise lembrando que os vetores de *summary features* são compostos por medidas baseadas nos parâmetros extraídos dos sinais musicais. Logo, concluímos que, na verdade, os parâmetros escolhidos e a forma como são usados pelo classificador, não são suficientes para uma plena individualização dos dois gêneros-alvo.

Estas considerações sugerem a realização de novas investigações visando gerar novos mecanismos que possibilitam o classificador atingir um melhor desempenho para os gêneros-alvo: *Rock* e *Techno*, bem como para os outros gêneros com características similares entre si.

Estas novas investigações poderiam seguir em duas direções. Em uma primeira, poderíamos testar os efeitos de aumento no número de vetores de referência, na expectativa de que o aumento do grau de liberdade de ajustes permitirá que alguns vetores se especializem na diferenciação entre os gêneros-alvo naquelas situações mais difíceis que induzem ao erro. Caso esta proposição se mostre ineficaz, poderíamos propor que o classificador use um número variável de vetores de referência a cada par de gêneros, de acordo com as dificuldades classificatórias.

Uma segunda linha investigatória poderia buscar a definição de parâmetros aptos a evidenciar as diferenças entre os gêneros-alvo. A construção destes parâmetros passa pela identificação dos exemplares musicais que provocam erros de classificação e pelo estudo detalhado das características musicais.

Por fim, para completar o estudo desenvolvido nesta dissertação, resta apresentar as conclusões alcançadas a cerca do desempenho do classificador quando aplicado a taxonomias menores que a original.

A Figura 5.4 mostra o resultado obtido com a nova taxonomia. Os resultados da segunda camada foram extraídos da Tabela 5.1 e os resultados para a primeira camada são aqueles mostrados na Tabela 5.2. O índice médio de acerto para a camada mais refinada foi de 61% e para a primeira camada, 88%. Em contraste, os resultados obtidos por [Barbedo & Lopes, 2006] foram 61% para a quarta camada e 87% para a primeira camada, conforme mostrado na Tabela 4.1. A comparação destes resultados mostra que os resultados são semelhantes, o que permite esboçar algumas conclusões. Em primeiro lugar, o classificador proposto apresenta bom desempenho mesmo quando aplicado a taxonomia com menor número de gêneros na camada mais refinada. Como conseqüência, este

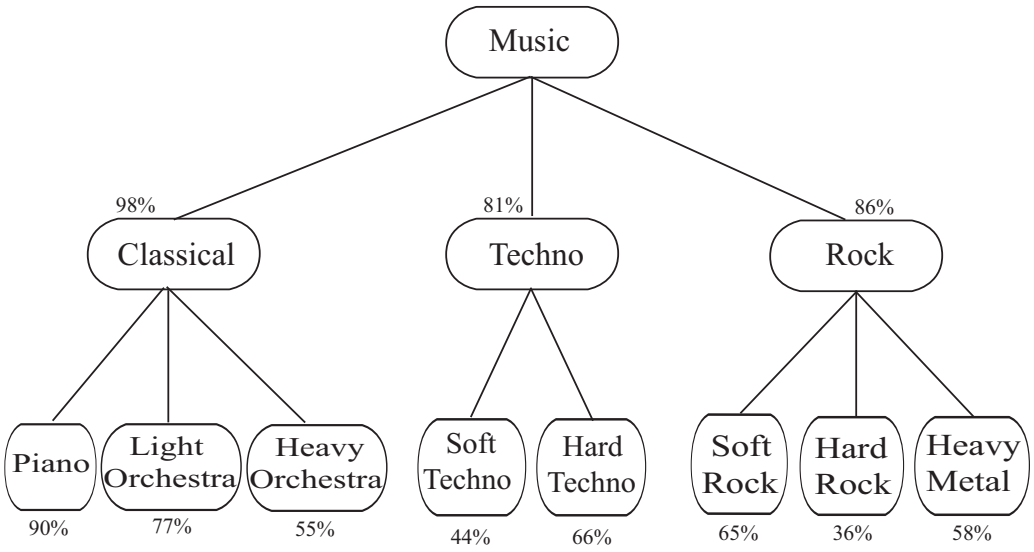

Figura 5.4: Índice de acerto obtido para todos os gêneros.

estudo sugere que o classificador apresenta um maior grau de flexibilidade de aplicação que aquele indicado pelos autores, indicando que ele também pode ser aplicado em estruturas menores, como, por exemplo, em taxonomias focadas em estilos musicais particulares, definidos por eventuais usuários.

Capítulo 6

Conclusão

Esta dissertação apresentou alguns estudos sobre as características do Classificador Automático de Sinais Musicais proposto em [Barbedo & Lopes, 2006]. Os objetivos foram as análises da viabilidade de ampliação da aplicabilidade do classificador em taxonomias distintas daquela usada pelos autores e da possibilidade de redução do esforço computacional para treinamento. Para estes estudos foi desenvolvida uma nova implementação em Matlab® do classificador, a qual foi validada usando a base de dados original.

Iniciando pela aplicabilidade, uma das características notáveis do classificador é o uso de uma taxonomia com 29 gêneros na camada inferior e a exploração dessa diversidade de gêneros para obter precisão de classificação. Neste contexto, o objetivo do nosso estudo foi verificar se o classificador mantém o desempenho original quando aplicado a taxonomias menores, particularmente em termos de gêneros na camada mais baixa. Propusemos uma taxonomia com duas camadas, três gêneros-alvo na primeira camada e oito gêneros na camada inferior. Um dos gêneros-alvo apresenta características bem definidas e oferece condições favoráveis para um bom desempenho do classificador. Por outro lado, os dois outros gêneros-alvo, ao contrário, apresentam características similares e foram escolhidos propositadamente para criar dificuldades ao classificador. Todos os elementos desta taxonomia foram extraídos da taxonomia original, o que permitiu usar a base de dados original para treinamento e avaliação de desempenho. Como resultado, demonstramos que o classificador atinge desempenho similar àquele obtido com a taxonomia original. Concluímos então que o classificador pode se aplicado a taxonomias menores que aquela proposta pelos autores, até a dimensão daquela usada neste

trabalho. Isto amplia a aplicabilidade de modo a atender demandas particulares dos mais variados usuários .

O segundo objetivo visou reduzir o esforço computacional para o treinamento. A proposta original usa vetores de referência como instrumento de classificação de sinais musicais a cada possível par de gêneros da camada mais baixa,. Os vetores de referência são escolhidos dentro de um conjunto de vetores candidatos. A definição do conjunto de candidatos demanda esforço computacional e não assegura o conhecimento a priori do número de candidatos selecionados. Propusemos uma alteração do processo original de definição, optando por um número pré-definido de candidatos a serem selecionados e fazendo a seleção de forma aleatória. Os testes de desempenho mostraram que o novo processo não altera o desempenho original do classificador e, adicionalmente, forneceram um valor indicativo para o número pré-definido de candidatos, de modo a assegurar este desempenho. Os testes também revelaram que a seleção proposta leva à definição de vetores de referência robustos, os quais dispensam processos de otimização.

Como subproduto dos estudos e testes realizados, constatamos que o classificador, tanto na versão original como naquela modificada pelas alterações aqui propostas, apresenta um nível de erro ao redor de 15% entre os dois gêneros-alvo similares e que este comportamento deve se repetir para quaisquer pares de gêneros-alvo com similaridades. Propusemos explicações sobre este comportamento, as quais apontaram a necessidade de novas investigações visando gerar mecanismos que robusteçam o classificador quanto à sua capacidade de distinguir gêneros-alvo similares.

Estas novas investigações podem seguir em duas direções. Em uma primeira, devemos testar os efeitos do aumento do número de vetores de referência a cada par de gêneros da camada taxonômica mais baixa, na expectativa de que o correspondente aumento de grau de liberdade de ajuste permita que alguns vetores se especializem na diferenciação entre os gêneros-alvo naquelas situações mais difíceis que induzem ao erro. Caso esta tentativa tenha sucesso, podemos propor o classificador com um número variável de vetores de referência a cada par de gêneros da camada mais baixa, conforme as dificuldades classificatórias.

A segunda linha investigatória deverá focar nas características dos parâmetros que são extraídos dos sinais a serem classificados. Terá como objetivo encontrar novos parâmetros aptos a evidenciar as diferenças entre os gêneros-alvo. A construção destes parâmetros passa pela identificação dos

exemplares musicais que provocam erros de classificação e pelo estudo de suas características, as quais poderão indicar as propriedades a serem exibidas pelos novos parâmetros.

Referências Bibliográficas

[A. Berenzweig & Whitman, 2004] A. Berenzweig, D. Ellis, B. L. & Whitman, B. (2004). A large-scale evaluation of acoustic and subjective music-similarity measures. *Proceedings of the 5th International Conference on Music Information Retrieval*, 28(2), 63 – 76. http://alumni.media.mit.edu/ bwhitman/ismir03-sim.pdf (acessado em 14.04.09).

[Agostini et al., 2003] Agostini, G., Longari, M., & Pollastri, E. (2003). Musical instrument timbres classification with spectral features. *EURASIP Journal on Applied Signal Processing*, 2003, 5 – 14. doi: 10.1155/S1110865703210118.

[Barbedo & Lopes, 2004] Barbedo, J. G. A. & Lopes, A. (2004). Estado da arte da classificação de sinais de Áudio. *VIII Convenção Nacional da Sociedade de Engenharia de Áudio*, 2004, 1–6.

[Barbedo & Lopes, 2006] Barbedo, J. G. A. & Lopes, A. (2006). Automatic genre classification of musical signals. *EURASIP Journal on Advances in Signal Processing*, 2007, 12 pages. doi: 10.1155/2007/64960.

[Basili et al., 2004] Basili, R., Serafini, A., & Stellato, A. (2004). Classification of musical genre: A machine learning approach. *Proceedings of the 5th Internetional Conference on Musical Information Retrieval*, 2004, 4 pages. http://ismir2004.ismir.net/proceedings/p092-page-505-paper239.pdf (acessado em 14.04.09).

[Büchler, 2002] Büchler, M. C. (2002). *Algorithms for Sound Classification in Hearing Instruments*. PhD thesis, Zurich Switzerland. http://www.forschungsportal.ch/unizh/p1395.htm (acessado em 14.04.09).

[Bergstra et al., 2006] Bergstra, J., Casagrande, N., Erhan, D., Eck, D., & Kegl, B. (2006). Aggregate features and adaboost for music classification. *Machine Learning*, 65(2 - 3), 473 – 484. doi: 10.1007/s10994-006-919-7.

[Burred & Lerch, 2003] Burred, J. J. & Lerch, A. (2003). A hierarchical approach to automatic musical genre classification. *Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03)*, 2003, 308 – 311. http://www.elec.qmul.ac.uk/dafx03/proceedings/pdfs/dafx06.pdf (acessado em 14.04.09).

[C. McKay & Fujinaga, 2005] C. McKay, R. Fiebrink, D. M. B. L. & Fujinaga, I. (2005). Ace: a framework for optimizing music classification. *Proceedings of the 6th International Conference on Music Information Retrieval, London, UK.*, 2005, 42 – 49.

[C. Xu & Shao, 2005] C. Xu, N. C. M. & Shao, X. (2005). Automatic music classification and summarization. *IEEE Transactions on Speech and Audio Processing*, 13(3), 441 – 450. doi: 10.1109/TSA.2004.840939.

[Carey et al., 1999] Carey, M. J., Parris, E., & Lloyd-Thomas, H. (1999). A comparison of features for speech, music discrimination. *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, 1, 149 – 152. doi: 10.1109/ICASSP.1999.758084.

[Cataltepe et al., 2007] Cataltepe, Z., Yaslan, Y., & Sonmez, A. (2007). Music genre classification using midi and audio features. *EURASIP Journal on Advances in Signal Processing*, 2007, 8. doi: 10.1155/2007/36409.

[Davis & Mermelstein, 1980] Davis, S. B. & 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.

[El-Maleh et al., 2000] El-Maleh, K., Klein, M., Petrucci, G., & Kabal, P. (2000). Speech/music discrimination for multimedia applications. *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 4, 2445–2448. doi: 10.1109/ICASSP.2000.859336.

[F. Gouyon & Widmer, 2004] F. Gouyon, S. Dixon, E. P. & Widmer, G. (2004). Evaluating rhythmic descriptors for musical genre classification. *Proceedings of the 25th International AES Conference*, 2004, 6 – 2. http://mtg.upf.edu/files/publications/AES25-GouyonDixonPampalkWidmer.pdf (acessado em 14.04.09).

[Fabbri, 1999] Fabbri, F. (1999). Browsing music spaces: Categories and the musical mind. *IASPM*, 1999, 1–14. http://www.mediamusicstudies.net/tagg/xpdfs/ffabbri9907us.pdf (acessado em 14.04.09).

[Fabbri & Chambers, 1982] Fabbri, F. & Chambers, I. (1982). What kind of music? *Popular Music*, 2, 131–143. doi: 10.1017/S0261143000001264.

[Foote, 1997] Foote, J. (1997). Content-based retrieval of music and audio. *C.-C. J. Kuo et al., editor, Multimedia Storage and Archiving Systems II, Proc. of SPIE*, 3229, 138 –147. http://www.rotorbrain.com/foote/papers/spie97.pdf (acessado em 14.04.09).

[Fraser & Fujinaga, 1999] Fraser, A. & Fujinaga, I. (1999). Toward real-time recognition of acoustic musical instruments. In *Proceedings of the International Computer Music Conference* (pp. 175 – 177).

[Golub, 2000] Golub, S. (2000). Classifying recorded music. Master's thesis, University of Edinburgh.

[Guo & Li, 2003] Guo, D. & Li, S. (2003). Content-based audio classification and retrieval by support vector machines. *IEEE Trans. on Neural Networks*, 14, 209–215. doi: 10.1109/TNN.2002.806626.

[H. Deshpande & Nam, 2001] H. Deshpande, R. S. & Nam, U. (2001). Classification of musical signals in the visual domain. *Proceedings of the COST G-6 Conference on Digital Audio Effects*, 2001, 1 – 4.

[J. R. Deller & Proakis, 1993] J. R. Deller, J. H. L. H. & Proakis, J. G. (1993). *Discrete-Time Processing of Speech Signals*. Macmiller.

[Julien & Pachet, 2003] Julien, J.-. & Pachet, F. (2003). Representing musical genre: A state of the art. *Journal of New Music Research*, 32(1), 83–93. doi: 10.1076/jnmr.32.1.83.16801.

[Junior et al., 2005] Junior, C. N. S., Kaestner, C. A. A., & Koerich, A. L. (2005). Classificação automática de gêneros musicais utilizando métodos de bagging e boosting. In *10º Simpósio Brasileiro de Computação Musical*. http://www.cefala.org/sbcm2005/papers/12438.pdf (acessado em 14.04.09).

[Kosina, 2002] Kosina, K. (2002). Music genre recognition. Master's thesis, Hagenberg. http://kyrah.net/mugrat/mugrat.pdf (acessado em 14.04.09).

[K.West & Cox, 2004] K.West & Cox, S. (2004). Features and classifiers for the automatic classification of musical audio signals. *Proceedings of the 5th International Conference on Music Information Retrieval, Barcelona, Spain.*, 2004, 1 – 6. http://ismir2004.ismir.net/proceedings/p096-page-531-paper115.pdf (acessado em 14.04.09).

[Li, 2000] Li, S. Z. (2000). Content-based classification and retrieval of audio using the nearest feature line method. *IEEE Transactions on Speech and Audio Processing*, 8, 619 – 625. doi: 10.1109/89.861383.

[Li & Ogihara, 2006] Li, T. & Ogihara, M. (2006). Toward intelligent music information retrieval. *IEEE Transactions on Multimedia*, 8(3), 564 – 574. doi: 10.1109/TMM.2006.870730.

[Liu & Wan, 2001] Liu, M. & Wan, C. (2001). A study on content-based classification and retrieval of audio database. *Proceedings of International Database Enginnering and Applications Symposium - IDEAS*, 2001, 339 – 345. doi: 10.1109/IDEAS.2001.938102.

[Logan, 2000] Logan, B. (2000). Mel-frequency cepstral coefficients for music modeling. *Proceedings of the International Conference on Music Information Retrieval*, 2000, 1–13. http://ciir.cs.umass.edu/music2000/papers/logan-abs.pdf (acessado em 14.04.09).

[Lu & Hankinson, 1998] Lu, G. J. & Hankinson, T. (1998). A technique towards automatic audio classification and retrieval. *Proc. IEEE Intl. Conf. on Signal Processing*, 2, 1142–1145. doi: 10.1109/ICOSP.1998.770818.

[Lu & Zhang, 2002] Lu, L. & Zhang, H. (2002). Content analysis for audio classification and segmentation. *IEEE Trans. Speech Audio Process.*, 10(5), 504 – 516. doi: 10.1109/TSA.2002.804546.

[Martin, 1999] Martin, K. (1999). Toward automatic sound source recognition: Identifying musical instruments. *NATO Computational Hearing Advanced Study Institute, Il Ciocco, Italy.*, 1999, 1 – 6. http://sound.media.mit.edu/Papers/kdm-comhear98.pdf (acessado em 14.04.09).

[Martin & Kim, 1998] Martin, K. & Kim, Y. (1998). Musical instrument identification: A pattern-recognition approach. *Proceedings of the 136th meeting of the Acoustical Society of America*, 104, 1 – 12. http://sound.media.mit.edu/Papers/kdm-asa98.pdf (acessado em 14.04.09).

[McKinney, 2003] McKinney, M. F. (2003). Features for audio and music classification. *Proceedings of ISMIR*, 2003, 1 – 8. http://jhir.library.jhu.edu/handle/1774.2/22 (acessado em 14.04.09).

[MIREX Contest Results, 2005] MIREX Contest Results (2005). Mirex contest results - audio genre classification. http://www.music-ir.org/evaluation/mirex-results/audio-genre/index.html (acessado em 14.04.09).

[O. Hellmuth & Wistorf, 2004] O. Hellmuth, E. Allamanche, J. H. T. K. N. L. & Wistorf, R. (2004). Music genre estimation from low level audio features. *Proceedings of the 25th International AES Conference*, 2004, 3–6.

[Pachet & Cazaly, 2000] Pachet, F. & Cazaly, D. (2000). A taxonomy of musical genres. *Content-Based Multimedia Information Access Conference (RIAO)*, 2000, 1 – 8. http://www.csl.sony.fr/downloads/papers/2000/pachet-riao2000.pdf (acessado em 14.04.09).

[Pampalk, 2006] Pampalk, E. (2006). *Computational models of music similarity and their application to music information retrieval*. PhD thesis, Vienna University of Technology, Vienna, Austria. http://www.ofai.at/ elias.pampalk/publications/pampalk06thesis.pdf (acessado em 14.04.09).

[Panagiotakis & Tziritas, 2005] Panagiotakis, C. & Tziritas, G. (2005). A speech/music discriminator based on rms and zero-crossings. *IEEE Transaction on Multimedia*, 7(1), 155 – 166. doi: 10.1109/TMM.2004.840604.

[Pohlmann, 2000] Pohlmann, K. C. (2000). *Principles of Digital Audio*. Mc. Graw - Hill, 4 edition.

[Pye, 2000] Pye, D. (2000). Content-based methods for the management of digital music. *In IEEE International Conference on Acoustics, Speech and Signal Processing*, 6, 2437 – 2440. doi: 10.1109/ICASSP.2000.859334.

[S. Dixon & Widmer, 2004] S. Dixon, F. G. & Widmer, G. (2004). Towards characterisation of music via rhythmic patterns. *Proceedings of the 5th International Conference on Music Information Retrieval*, 2004, 1 – 8. http://ismir2004.ismir.net/proceedings/p093-page-509-paper165.pdf (acessado em 14.04.09).

[S. Lippens & Tzanetakis, 2004] S. Lippens, J. P. Martens, T. D. M. & Tzanetakis, G. (2004). A comparison of human and automatic musical genre classification. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 4, iv–233 – iv–236. doi: 10.1109/ICASSP.2004.1326806.

[Saunders, 1996] Saunders, J. (1996). Real-time discrimination of broadcast speech/music. *Proceedings of the IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing*, 2, 993 – 996. doi: 10.1109/ICASSP.1996.543290.

[Scheier & Slaney, 1997] Scheier, E. & Slaney, M. (1997). Construction and evaluation of a robust multifeature speech/music discriminator. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2, 21 – 24. doi: 10.1109/ICASSP.1997.596192.

[T. Lambrou & Linney, 1998] T. Lambrou, P. Kudumakis, R. S. M. S. & Linney, A. (1998). Classification of audio signals using statistical features on time and wavelet transform domains. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 6, 3621–3624. doi: 10.1109/ICASSP.1998.679665.

[T. Pohle & Widmer, 2005] T. Pohle, E. P. & Widmer, G. (2005). Evaluation of frequently used audio features for classification of music into perceptual categories. *Proceedings of the 4th International Workshop on Content-Based Multimedia Indexing, Riga, Latvia.*, 2005, 1 – 8. http://www.cp.jku.at/research/papers/Pohle-CBMI-2005.pdf (acessado em 14.04.09).

[Theodoridis & Koutroumbas, 2006] Theodoridis, S. & Koutroumbas, K. (2006). *Pattern Recognition*. Academic Press, third edition edition.

[Thiede, 1999] Thiede, T. V. (1999). *Perceptual audio quality assessment using a non-linear filter bank*. PhD thesis, Technical University of Berlin, Berlin. http://www.mp3-tech.org/programmer/docs/scriptHF021.pdf (acessado em 14.04.09).

[Tzanetakis & Cook, 1999] Tzanetakis, G. & Cook, P. (1999). A framework for audio analysis based on classication and temporal segmentation. *In Proc.25th Euromicro Conference. Workshop on Music Technology and Audio Processing*, 2, 61 – 67. doi: 10.1109/EURMIC.1999.794763.

[Tzanetakis & Cook, 2002] Tzanetakis, G. & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5), 293 – 302. doi: 10.1109/TSA.2002.800560.

[Tzanetakis et al., 2002] Tzanetakis, G., Essl, G., & Cook, P. (2002). Automatic musical genre classification of audio signals. *In Proc. Int. Symposium on Music Inform. Retriev. (ISMIR)*, 10(5), 205 – 210. http://ismir2001.ismir.net/pdf/tzanetakis.pdf (acessado em 14.04.09).

[V. Peltonen & Sorsa, 2002] V. Peltonen, J. Tuomi, A. K. J. H. & Sorsa, T. (2002). Computational auditory scene recognition. In *Proceedings of ICASSP 02*, volume 2 (pp. 1941 – 1944). Florida, USA.

[Wang & Jo, 2006] Wang, J. & Jo, C. (2006). Performance of gaussian mixture models as a classifiers for pathological voice. In P. Warren & C. I. Watson (Eds.), *Procedings of 11th Australian International Conference on Speech Science & Technology* (pp. 165 – 169). New Zealand. http://www.assta.org/sst/2006/sst2006-84.pdf (acessado em 14.04.09).

[Wold et al., 1996] Wold, E., Blum, T., Keislar, D., & Wheaton, J. (1996). Content-based classification, search, and retrieval of audio. In *IEEE Multimedia Magazine*, volume 3 (pp. 27 – 36). doi: 10.1109/93.556537.

[Zhang & Kuo, 2001] Zhang, T. & Kuo, C. C. J. (2001). Audio content analysis for online audiovisual data segmentation and classification. *IEEE Transactions on Speech and Audio Processing*, 9, 441 – 457. doi: 10.1109/89.917689.