



Departamento de Engenharia Informática
Faculdade de Ciências e Tecnologia
Universidade de Coimbra

Sistemas de Classificação Automática em Géneros Musicais

Ricardo Manuel da Silva Malheiro
Licenciado em Matemática / Ramo de Sistemas e Métodos de
Computação Gráfica

Coimbra, 2003



Departamento de Engenharia Informática
Faculdade de Ciências e Tecnologia
Universidade de Coimbra

Sistemas de Classificação Automática em Géneros Musicais

Dissertação submetida para obtenção do grau de Mestre em
Engenharia Informática

Ricardo Manuel da Silva Malheiro
Licenciado em Matemática / Ramo de Sistemas e Métodos de
Computação Gráfica

Coimbra, 2003

Dissertação realizada sob a orientação do

Professor Doutor António José Mendes

Professor Auxiliar do
Departamento de Engenharia Informática da
Faculdade de Ciências e Tecnologia da
Universidade de Coimbra

À Paula e à Marta

A G R A D E C I M E N T O S

Gostaria de agradecer em primeiro lugar ao Professor Doutor António José Mendes na qualidade de orientador científico pela oportunidade concedida para a realização deste trabalho, bem como pela motivação e disponibilidade demonstradas.

Ao Mestre Rui Pedro Paiva os meus agradecimentos especiais por todas as discussões estimulantes, sugestões, observações prestadas, bem como pela amizade demonstrada.

Finalmente, gostaria de agradecer à minha família em especial à Paula, porque sem ela tudo isto não teria sido possível.

RESUMO

Como resultado da massificação do computador, do aumento generalizado da largura de banda disponível e da universalização da Internet, a indústria da distribuição electrónica de música teve um enorme crescimento nos últimos anos. Esse crescimento está também relacionado com a facilidade com que à velocidade de um clique se pode aceder a bases de dados de música de grandes dimensões. Essas bases de dados têm de estar sempre actualizadas com toda a música que é produzida diariamente e têm de estar organizadas de acordo com as taxonomias definidas para poder responder da melhor maneira às pesquisas dos utilizadores.

A catalogação de peças musicais com base nas taxonomias utilizadas, é um processo cada vez mais difícil de realizar de uma forma manual, devido a questões de tempo e de eficiência de quem as faz. Surgiu portanto a necessidade da utilização do computador para a criação de sistemas de classificação automáticos.

Este tipo de sistemas envolve tarefas como a extracção de características de cada música e o desenvolvimento de classificadores que utilizem as características extraídas.

Quanto à extracção de características, utiliza-se neste trabalho o *zcr*, *loudness*, centróide, largura de banda e uniformidade. Estas características são estatisticamente manipuladas fazendo um total de 40 características para cada música.

Em seguida são utilizados três classificadores: KNN, GMM e MLP. A classificação consistiu em três problemas, todos relacionados com a música clássica. No primeiro pretendeu-se discriminar entre música para flauta, piano e violino. No segundo problema pretendeu-se distinguir música coral de ópera. Finalmente no terceiro classificou-se num dos 5 géneros musicais anteriores.

Após a comparação dos resultados dos classificadores, chegou-se à conclusão que o MLP originou os melhores resultados em todas as tarefas de classificação, conseguindo percentagens de músicas bem classificadas de 85%, 90% e 76% respectivamente para os primeiro, segundo e terceiro problema de classificação.

Este classificador foi então utilizado para fazer uma aproximação a um sistema de classificação automático de géneros musicais. Neste sistema, cada música foi representada por dez extractos escolhidos de igual forma para todas as músicas. Cada música foi classificada no género musical mais representado pelos seus extractos.

A B S T R A C T

The massification of computer use, and the availability of Internet access with increasing available bandwidth, created conditions to the tremendous growth that the industry of electronic music delivery has experienced in the last few years. This growth is also related to the ease that, at the speed of one click, one can access huge music databases. Those databases must be permanently updated with all the music that is produced every day and must also be organized according to the defined taxonomies, so that they give the best possible answer to user's queries.

The process of music labeling according to the used taxonomies is more and more difficult to carry out manually due to time necessary and the subjectivity of the task. Therefore, it becomes necessary to use the computer as a tool for automatic classification.

Classification systems involve tasks such as the extraction of features from each musical piece and the development of classifiers that use the extracted features.

Regarding feature extraction, in this work we use the *zcr*, *loudness*, centroid, bandwidth and uniformity. These features are statistically manipulated, making a total of 40 features for each piece of music.

Then, three classifiers were used: KNN, GMM and MLP. The classification consisted on three problems, all of them related to classical music. In the first one, the goal was to discriminate between music for flute, piano and violin. In the second problem, we aimed to separate choral music from opera. Finally, in the third problem, the classification was conducted using the five referred genres.

After comparing the results obtained with the three classifiers, we came to the conclusion that MLP originated the best results in all the classification tasks, achieving accuracies of 85%, 90% and 76% for the first, second and third problems, respectively.

This classifier was then used for approximating an automatic music genre classification system. In this system, each musical piece was represented by ten extracts, chosen in the same manner for all the pieces. Each piece was then classified according to the extract's most represented musical genres.

ÍNDICE GERAL

Agradecimentos	vii
Resumo	ix
Abstract.....	xi
Índice Geral	xiii
Índice de Figuras	xvii
Índice de Tabelas	xix
Simbologia.....	xxi
CAPÍTULO 1 INTRODUÇÃO	1
1.1. Motivação e Enquadramento	1
1.2. Abordagens	4
1.3. Contribuições da Dissertação.....	5
1.4. Organização da Dissertação.....	7
CAPÍTULO 2 ESTADO DE ARTE	9
CAPÍTULO 3 EXTRACÇÃO DE CARACTERÍSTICAS.....	13
3.1. Motivação	13
3.2. Formato de gravação das músicas.....	15
3.3. Extractos musicais: Escolha e duração.....	15

3.4. Conversões: Tempo – Frequência	16
3.5. Características	21
3.5.1. <i>Características base</i>	23
3.5.2. <i>Características intermédias</i>	27
3.5.3. <i>Características finais</i>	27
3.6. Normalização	29
CAPÍTULO 4 CLASSIFICAÇÃO	31
4.1. Tarefas de Classificação	31
4.1.1. <i>Primeira Tarefa de Classificação</i>	33
4.1.2. <i>Segunda Tarefa de Classificação</i>	34
4.1.3. <i>Terceira Tarefa de Classificação</i>	35
4.2. K-Vizinhos mais Próximos	36
4.3. Modelos de Misturas Gaussianas	39
4.3.1. <i>Equações Fundamentais</i>	39
4.3.2. <i>Processo de Treino</i>	40
4.3.3. <i>Processo de Validação</i>	43
4.4. Redes Neurais Artificiais	43
4.4.1. <i>Topologias</i>	44
4.4.2. <i>Componentes e Processamento de Dados</i>	46
4.4.3. <i>Treino</i>	49
4.4.4. <i>Validação</i>	56
CAPÍTULO 5 RESULTADOS EXPERIMENTAIS.....	57
5.1. Pressupostos Iniciais	57
5.2. K-Vizinhos mais Próximos	59
5.2.1. <i>Variantes do Algoritmo e Regras de Classificação</i>	59
5.2.2. <i>Primeira Classificação: Três Géneros Musicais</i>	60
5.2.3. <i>Segunda Classificação: Dois Géneros Musicais</i>	61
5.2.4. <i>Terceira Classificação: Cinco Géneros Musicais</i>	63
5.3. Modelos de Misturas Gaussianas	65
5.3.1. <i>Princípios Gerais</i>	65
5.3.2. <i>Primeira Classificação: Três Géneros Musicais</i>	66
5.3.3. <i>Segunda Classificação: Dois Géneros Musicais</i>	67
5.3.4. <i>Terceira Classificação: Cinco Géneros Musicais</i>	68
5.4. Redes Perceptrão Multicamada	69
5.4.1. <i>Princípios Gerais</i>	69
5.4.2. <i>Regras de Classificação</i>	71
5.4.3. <i>Primeira Classificação: Três Géneros Musicais</i>	72
5.4.4. <i>Segunda Classificação: Dois Géneros Musicais</i>	75
5.4.5. <i>Terceira Classificação: Cinco Géneros Musicais</i>	77
5.5. Comparação entre Classificadores	80
5.6. Protótipo	81
5.6.1. <i>Extracção das Peças Musicais</i>	82
5.6.2. <i>Regras de Classificação</i>	83

5.6.3. *Primeira Tarefa de Classificação* 84
 5.6.4. *Segunda Tarefa de Classificação* 89
 5.6.5. *Terceira Tarefa de Classificação* 93

CAPÍTULO 6 CONCLUSÕES.....101

6.1. Conclusões Gerais 102
 6.2. Perspectivas Futuras 103

ANEXOS105

Anexo 1. Músicas 105

BIBLIOGRAFIA113

LISTA DE FIGURAS

Figura 3.1. Transformada de Fourier de dois sinais: um estacionário e outro não estacionário, com frequências de 10, 50, 100 e 200 Hz.....	18
Figura 3.2. Resultado da aplicação de duas janelas de Hanning contíguas a um sinal ..	20
Figura 4.1. Taxonomia utilizada.....	32
Figura 4.2. Primeira tarefa de classificação. Taxonomia.....	33
Figura 4.3. Segunda tarefa de classificação. Taxonomia.....	34
Figura 4.4. Terceira tarefa de classificação. Taxonomia	35
Figura 4.5. Padrões de treino representados no espaço de características X	37
Figura 4.6. Cálculo dos vizinhos mais próximos para um exemplo de teste	37
Figura 4.7. Rede neuronal com ligações para a frente.....	44
Figura 4.8. Rede neuronal recorrente.....	45
Figura 4.9. Relação entre dois neurónios p e j	46
Figura 4.10. Rede MLP. Processamento de informação efectuado por cada neurónio ..	48
Figura 4.11. Rede MLP. Treino supervisionado	51

LISTA DE TABELAS

Tabela 3.1. Assinatura de cada extracto musical.....	28
Tabela 5.1. Matriz de confusão da música instrumental: KNN(1).....	60
Tabela 5.2. Matriz de confusão da música instrumental: KNN(3).....	61
Tabela 5.3. Matriz de confusão da música instrumental: KNN(5).....	61
Tabela 5.4. Matriz de confusão da música vocal: KNN(1).....	62
Tabela 5.5. Matriz de confusão da música vocal: KNN(3).....	62
Tabela 5.6. Matriz de confusão da música vocal: KNN(5).....	62
Tabela 5.7. Matriz de confusão da música vocal e instrumental: KNN(1).....	63
Tabela 5.8. Matriz de confusão da música vocal e instrumental: KNN(3).....	64
Tabela 5.9. Matriz de confusão da música vocal e instrumental: KNN(5).....	64
Tabela 5.10. Matriz de confusão da música instrumental: GMM	67
Tabela 5.11. Matriz de confusão da música vocal: GMM.....	68
Tabela 5.12. Matriz de confusão das músicas instrumental e vocal: GMM	68
Tabela 5.13. Matriz de confusão das camadas para a primeira classificação: MLP	73
Tabela 5.14. Matriz de confusão da música instrumental, RCP1: MLP	73
Tabela 5.15. Matriz de confusão da música instrumental, RCP2: MLP.....	73
Tabela 5.16. Matriz de confusão das camadas para a segunda classificação: MLP.....	75
Tabela 5.17. Matriz de confusão da música vocal, RCP1: MLP.....	75
Tabela 5.18. Matriz de confusão da música vocal, RCP2: MLP	76
Tabela 5.19. Matriz de confusão das camadas para a terceira classificação – MLP	77
Tabela 5.20. Matriz de confusão das músicas instrumental e vocal, RCP1: MLP.....	78
Tabela 5.21. Matriz de confusão das músicas instrumental e vocal, RCP1: MLP.....	78
Tabela 5.22. Matriz de confusão das músicas instrumental e vocal, RCP1 (2): MLP ..	79
Tabela 5.23. Resultados finais dos classificadores para as três tarefas de classificação.....	80

Tabela 5.24. Critérios de extracção das peças musicais	83
Tabela 5.25. Resultados gerais de validação em 3 classes.....	85
Tabela 5.26. Matriz de confusão do protótipo para a primeira classificação: RCP1	88
Tabela 5.27. Matriz de confusão do protótipo para a primeira classificação: RCP2	88
Tabela 5.28. Resultados gerais de validação em 2 classes.....	90
Tabela 5.29. Matriz de confusão do protótipo para a segunda classificação: RCP1	92
Tabela 5.30. Matriz de confusão do protótipo para a segunda classificação: RCP2	92
Tabela 5.31. Resultados gerais de validação em 5 classes.....	94
Tabela 5.32. Matriz de confusão do protótipo para a terceira classificação: RCP1	98
Tabela 5.33. Matriz de confusão do protótipo para a terceira classificação: RCP2	99

SIMBOLOGIA

Abreviaturas

ANN	Artificial Neural Network
CDA	Canonical Discriminant Analysis
DFT	Discrete Fourier Transform
EM	Expectation-Maximization
EIS	Engineering of Intelligent Systems
ESANN	European Symposium on Artificial Neural Networks
<i>fdp</i>	função de densidade de probabilidade
FFT	Fast Fourier Transform
FFNN	FeedForward Neural Networks
GMM	Gaussian Mixture Models
KNN	<i>K</i> -Nearest Neighbors
LSP-VQ	Linear Spectral Pairs – Vector Quantization
LVQ	Learning Vector Quantization
MLP	Multilayer Perceptron
PCA	Principal Component Analysis

PNN	Probabilistic Neural Network
PPCA	Probabilistic Principal Component Analysis
QDA	Quadratic Discriminant Analysis
RBF	Radial Basis Function
RCP1	Regras de cálculo de percentagens 1
RCP2	Regras de cálculo de percentagens 2
SOM	Self-Organizing Map
SVM	Support Vector Machines
RN	Redes Neurais
STFT	Short Time Fourier Transform
TF	Transformada de Fourier

Símbolos

Extracção de Características

$L(r)$	<i>loudness</i> da janela r
$x(n)$	amplitude da n -ésima amostra
N	número de amostras de cada janela
$Z(r)$	número de intersecções com o eixo das abcissas na janela r
$sgn(x(n))$	sinal da amplitude da n -ésima amostra
$C(r)$	centróide da janela r

$M_r(k)$	magnitude da transformada de Fourier na janela r no índice de frequências k
$B(r)$	largura de banda da janela r
$U(r)$	uniformidade da janela r
$m_i(c_j)$	valor da característica j para a música i
c_j^{min}	valor mínimo da característica j
c_j^{max}	valor máximo da característica j

***K*-vizinhos mais próximos**

K	número de vizinhos a considerar
X	espaço de características dos padrões de treino
T	conjunto de vectores de características de teste
N	dimensão do espaço de características X
n	dimensão do conjunto de teste T

Modelos de Misturas Gaussianas

v	número de <i>clusters</i>
$p_i(x)$	função densidade de probabilidade para o <i>cluster</i> i
w_i	pesos da combinação linear relativamente ao <i>cluster</i> i
R_i	matriz de covariância do <i>cluster</i> i
\mathbf{m}_i	vector média do <i>cluster</i> i

\mathbf{q}	todos os parâmetros do modelo de misturas gaussianas
x	vector de características
d	dimensão de cada vector de características
X	conjunto dos vectores de características de treino
T	dimensão do conjunto X
c	número de iterações do algoritmo
$\tilde{\mathbf{q}}$	valor actualizado para os parâmetros da rede \mathbf{q}
$\tilde{\mathbf{m}}$	valor actualizado de \mathbf{m}
\tilde{R}	valor actualizado de R
\tilde{w}	valor actualizado de w

Redes Neurais

W	matriz de pesos
w_{ij}	peso da ligação entre o neurónio i (camada k) e o neurónio j (camada $k+1$)
b_i	termo de polarização do neurónio i
I	matriz de entradas na rede
$i_p^{(s)}$	p -ésima entrada da rede relativamente ao padrão s
$y_p^{(s)}$	p -ésima saída da rede relativamente ao padrão s
$\hat{y}_p^{(s)}$	p -ésima saída desejada para a rede, relativamente ao padrão s

f_i	função de activação do neurónio i
E	erro total
$E^{(s)}$	erro da rede relativamente ao padrão s
g	velocidade de aprendizagem
R	número de entradas da rede
n	número de saídas da rede
N	número de padrões de treino
x	vector de parâmetros da rede
$H(x)$	matriz Hessiana do vector de parâmetros da rede x
$J(x)$	matriz Jacobiana do vector de parâmetros da rede x
$g(x)$	gradiente do vector de parâmetros da rede x
D	matriz identidade

Capítulo 1

INTRODUÇÃO

“A música exprime a mais alta filosofia numa linguagem que a razão não compreende.”

Schopenhauer

1.1 MOTIVAÇÃO E ENQUADRAMENTO

A classificação de música tem-se tornado para o Homem cada vez mais importante, à medida que aumentam a quantidade de música disponível e as necessidades de catalogação. Essa catalogação pode tomar as mais diversas formas: por género musical, por artista, por época, por nacionalidade, por instrumento(s), por tipo de voz (e.g., feminina), por contexto em que se insere (e.g., música pertencente à banda sonora de um filme), etc.

Desde sempre a classificação tem sido feita manualmente, o que acarreta alguns problemas óbvios, como por exemplo o tempo que é necessário para concluir essa operação caso os dados a organizar sejam volumosos. Outro problema é o grau de subjectividade associado a cada classificação, já que os resultados desta dependem da pessoa que classifica, do seu conhecimento musical e da sua experiência.

Nos últimos anos, devido a vários factores como por exemplo, a evolução rápida dos computadores tanto a nível de software como hardware, o aumento gradual e generalizado de largura de banda disponível, a universalização cada vez maior da Internet e a criação de novos formatos de compressão de música mais eficientes (e.g., MP3), a Internet tornou-se rapidamente um mercado apetecível para um determinado conjunto de serviços que tiram partido desses factores para aumentar a satisfação e a comodidade dos utilizadores.

Estes serviços, como por exemplo, os de compra e venda de música (e.g., sítios como *AllMusicGuide* – www.allmusic.com, ou *CDNOW* – www.cdnw.com) necessitam, para se tornarem interessantes para os utilizadores, de bases de dados de música sempre actualizadas e motores de pesquisa em tempo real eficientes e rápidos que respondam sempre que possível com sucesso às pesquisas desses utilizadores.

Para se conseguir o sucesso nas pesquisas a essas bases de dados, é necessário organizá-las segundo taxonomias que vão de encontro às necessidades dos utilizadores. Uma das taxonomias mais comuns na classificação de música consiste na hierarquização de géneros musicais (e.g., música clássica dividida nos géneros vocal e instrumental. Por sua vez, o género vocal dividido em ópera, coral e o género instrumental dividido em música para flauta, música para piano e música para violino). Assim sendo, cada nova música a inserir na base de dados, deverá ser previamente classificada num dos géneros da taxonomia em causa, tarefa esta por vezes bastante subjectiva. Isto fará, como é óbvio, que o processo de actualização das bases de dados se torne ainda mais moroso e complexo. Esta consequência, juntamente com o facto de todos os dias serem adicionadas milhares de novas músicas nas bases de dados, faz com que os métodos manuais de classificação sejam ineficazes na resposta a essas necessidades. Assim, surge a inevitabilidade da utilização do próprio computador para esse tipo de tarefas, através de sistemas de classificação automática.

Uma outra dificuldade a ultrapassar, prende-se com a falta de uniformidade das taxonomias utilizadas pelos diferentes fornecedores de serviço. Diferentes serviços utilizam classes de tipos distintos, bem como diferentes especializações. Uma aproximação a este problema é conduzida por [Pachet & Cazaly, 2000], onde se procura definir regras uniformes para a descrição de taxonomias.

A classificação de música, não é a única causa para a criação de sistemas de classificação automática. Áreas como reconhecimento de voz ou em geral reconhecimento de padrões de sinal na Engenharia, na Medicina ou em outras ciências foram igualmente causadoras dessa evolução.

A investigação nesta área tem-se multiplicado, na tentativa de fazer com que esse tipo de sistemas, que estão ainda numa fase relativamente inicial, se tornem cada vez mais eficientes e fiáveis.

O objectivo desta dissertação é a classificação de música em subgéneros da música clássica¹. São considerados 3 problemas de classificação. No primeiro pretende-se a discriminação em 3 subgéneros instrumentais: música para flauta, música para piano e música para violino. No segundo, o objectivo é a distinção de música vocal: música coral e ópera. Finalmente o terceiro problema de classificação consiste na separação num dos 5 subgéneros musicais anteriores.

Escolheu-se, ao contrário de alguns investigadores que optaram por géneros bastante díspares (e.g., disco, música clássica, *jazz*, *rock*, etc), um conjunto de géneros musicais bastante similares divididos em vários problemas de classificação: discriminação de música instrumental, de música vocal e de música clássica no seu todo. Pretendeu-se criar classificadores especializados, mais focalizados numa única classe, ao invés de classificadores genéricos que abarcassem classes mais díspares.

Optou-se pela música clássica por várias razões. Primeiro por não haver ainda muitos estudos específicos apenas sobre este tipo de música. Algumas excepções são [Tzanetakis & Cook, 2002], que subdivide música clássica em 4 subgéneros (Capítulo 2), e alguns autores que fazem análises tímbricas de sons separados de instrumentos não englobados em ambientes polifónicos [Agostini et al., 2003]. Em segundo pela necessidade crescente de investigação neste campo específico, já que existem na Internet cada vez mais sítios especializados ou com componentes especializadas na

¹ Mais correctamente dever-se-ia chamar música erudita, já que música clássica representa uma época da música erudita. Optou-se no entanto pela denominação de música clássica, pelo facto de ser mais intuitivo e de uso geral.

música clássica. Em terceiro, por que não dizê-lo, por uma questão de gosto pessoal do autor por este tipo de música.

1.2 ABORDAGENS

O objectivo deste estudo não é construir um sistema que utilize trechos de música de longa duração, mas sim de pequena duração e significativos para cada um dos géneros musicais. A ideia é imitar de alguma maneira, a forma como os seres humanos classificam a música, i.e. conseguir classificar utilizando pequenos segmentos de música [Perrot & Gjerdigen, 1999].

Vão ser extraídas características de cada extracto musical. Essas características foram escolhidas de forma a privilegiar a análise do timbre e do *pitch*² do sinal, o que, de acordo com os objectivos de classificação presentes (distinção entre timbre e *pitch* de instrumentos e voz) parece ser o ideal. As características base são, centróide, largura de banda, uniformidade, *loudness*³, *zcr*⁴ e as diferenças entre duas janelas consecutivas das 5 características anteriores. Nos géneros musicais em questão, o ritmo não é uma característica tão importante, como tal, não foram extraídas características importantes para a análise rítmica.

Neste trabalho são utilizados 3 tipos de classificadores: *K*-Vizinhos mais Próximos⁵ (KNN); Modelos de Misturas Gaussianas⁶ (GMM) e Redes Neurais Artificiais (ANN), mais concretamente Redes Percepção Multicamada⁷ (MLP).

² *Pitch* corresponde à percepção que o ouvido humano tem da frequência do sinal.

³ *Loudness* corresponde à percepção que o ouvido humano tem da intensidade do som.

⁴ Em terminologia Inglesa: Zero Crossing Rate. Esta característica mede a taxa de intersecções com o eixo das abcissas por unidade de tempo, enquanto neste estudo o que se mede é o número total de intersecções com o eixo das abcissas nos 6 s que compõem o sinal. Apesar desta discrepância de conceitos, optou-se por chamar à característica calculada, *zcr*.

⁵ Em terminologia Inglesa: *K*-Nearest Neighbors - KNN.

O objectivo de utilização de mais do que um classificador é precisamente para, partindo dos mesmos pressupostos iniciais, poder comparar resultados e assim ter uma noção do comportamento de cada um deles, num sistema de classificação deste tipo. Assim sendo, o KNN é um classificador estatístico muito simples de implementar que não necessita de treino. Por outro lado, além de ser mais lento na classificação, também exige mais memória para armazenar simultaneamente todos os vectores de características. O GMM, é um classificador probabilístico fácil de implementar e computacionalmente eficiente, utilizado principalmente, até há uns anos, em reconhecimento de voz. Em relação à rede MLP, é um modelo determinístico capaz de aproximar funções complexas.

Toda a programação utilizada neste trabalho foi produzida em *Matlab*⁸, já que esta ferramenta contém algumas funções nativas para manipular sinais áudio e *toolboxes* específicas para alguns dos classificadores utilizados. Não obstante as vantagens óbvias do *Matlab*, pretende-se migrar o código para uma linguagem mais universal como o C++, por uma questão de eficiência computacional.

Foi ainda utilizada para extração de segmentos de áudio uma excelente aplicação chamada *Cool Edit Pro*⁹.

1.3 CONTRIBUIÇÕES DA DISSERTAÇÃO

Do estudo, análise e desenvolvimentos efectuados ao longo deste trabalho, resultaram um conjunto de contribuições científicas, algumas delas com algum carácter original.

⁶ Em terminologia Inglesa: Gaussian Mixture Models - GMM.

⁷ Em terminologia Inglesa: Multilayer Perceptron - MLP.

⁸ Matlab – Versão 6.1.0.450 Release 12.1, The Mathworks, Inc.

⁹ Cool Edit Pro – Versão 2.00 [2095.0] , Syntrillium Software Corporation

Tentou-se sempre ao longo deste trabalho, utilizar técnicas conhecidas da literatura de uma forma combinada, com o intuito de apresentar novas abordagens que pudessem maximizar a resolução dos objectivos específicos deste estudo.

Em relação aos géneros musicais considerados, utilizou-se um conjunto de subgéneros da música clássica, bastante similares entre si. O objectivo foi tentar especializar um sistema de classificação, para desde logo poder atingir o mais possível “a folha da árvore” (e.g., música para piano). Outros autores fizeram estudos semelhantes, mas quase sempre para géneros muito diferentes entre si, em que por consequência classificavam de uma forma muito genérica (e.g., *jazz*, música clássica). Além disso, não há praticamente estudos específicos sobre música clássica, pelo menos com géneros tão similares. Esta diferente abordagem pode ser considerada uma contribuição original.

Quanto às características escolhidas foi utilizada uma abordagem que resultou na combinação de perspectivas de vários autores, tendo sempre em atenção o tipo de classificação que se pretendeu efectuar. Assim, foram escolhidas características reconhecidamente importantes para a análise de timbre e *pitch* do sinal.

Em relação à classificação propriamente dita, pretendeu-se comparar metodologias, utilizando para tal três classificadores bastante diferentes entre si, que utilizaram os mesmos dados e cujos resultados penderam claramente em todos os problemas de classificação para as redes MLP, mostrando a sua superioridade em relação a métodos estatísticos e probabilísticos, nestes problemas com grande similaridade entre as classes. Além disso os resultados alcançados, comparativamente com os obtidos por outros autores, podem ser considerados muito promissores.

Em termos de critérios de classificação de música para as redes MLP, foram criados dois conjuntos de regras (regras de cálculo de percentagens 1 e regras de cálculo de percentagens 2) que em conjunto fornecem uma ideia clara das influências de cada género musical numa determinada música, permitindo desta forma tirar conclusões mais precisas. Os critérios de classificação podem ser considerados contribuições originais.

O presente trabalho está condensado em dois artigos: um aceite para publicação na conferência *Engineering of Intelligent Systems* (EIS 2004), onde é definida uma aproximação a um sistema de classificação automática de subgéneros da

música clássica, desde a identificação dos problemas de classificação, passando pela extracção de características até à classificação utilizando redes neuronais; outro, submetido ao *European Symposium on Artificial Neural Networks* (ESANN 2004), onde é efectuado um estudo comparativo entre as várias metodologias de classificação descritas.

1.4 ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação está organizada em seis capítulos que estão dispostos de uma forma independente e sequencial em relação ao trabalho de experimentação realizado. O objectivo é torná-la tão coerente quanto possível.

O capítulo 1 corresponde a esta introdução, onde se discutem as motivações que estão na origem deste trabalho, as abordagens seguidas e as contribuições do trabalho.

No capítulo 2 será apresentado um resumo sobre o estado de arte actual na área da classificação automática de música.

Os capítulos 3 e 4 expõem cronologicamente os aspectos técnicos que estão por detrás da construção de um sistema de classificação automático. No capítulo 3 irá ser explicado o porquê da extracção de características das músicas. Irão ser explanados ainda alguns aspectos sobre as peças de música utilizadas neste trabalho, como a sua escolha, tipo e tamanho. Serão ainda apresentadas as características extraídas, bem como o seu cálculo. No capítulo 4 e após estar definida e calculada a representação de cada música (ficheiro onde estão definidas as características extraídas), são explicados os problemas de classificação em análise neste trabalho. Em seguida são descritos ao pormenor os 3 classificadores utilizados, KNN, GMM e MLP. Esta análise dos classificadores, apesar de ser bastante teórica é sempre que possível direccionada para os objectivos deste trabalho, nomeadamente através de exemplos.

O capítulo 5 irá descrever a forma como as experiências de investigação foram realizadas, nomeadamente, a forma como os classificadores foram configurados e as premissas que serviram de base a essas experiências. São apresentados os resultados de

toda a parte experimental e tiradas conclusões. No final do capítulo, com o classificador mais eficiente é construído um protótipo que se pretende seja uma aproximação a um sistema de classificação automático do mundo real.

No capítulo 6 são apresentadas as conclusões fundamentais deste trabalho e são identificadas algumas direcções a seguir no que concerne a trabalho futuro.

A dissertação contém ainda um anexo que contém a listagem de todas as músicas utilizadas no trabalho. A dissertação termina com a bibliografia.

Capítulo 2

ESTADO DE ARTE

Vai ser apresentado neste capítulo um resumo do trabalho produzido por vários investigadores na área do reconhecimento automático de géneros musicais. Não ser privilegiados aqueles trabalhos, cujo resultado final tenha sido considerado de alguma forma uma referência para a realização deste trabalho.

De notar que a problemática da discriminação de música em géneros é relativamente recente, ao contrário da classificação de voz ou da discriminação entre voz e música. Como tal, muito do conhecimento do reconhecimento em géneros, nomeadamente algumas das características utilizadas, foi herdado desses tipos de classificação.

Assim, vão ser descritos em seguida alguns trabalhos de investigação cujo objectivo é construir um sistema de classificação automático em géneros musicais, desde a definição dos problemas em questão, passando pela extracção de características, até à classificação propriamente dita.

George Tzanetakis e Perry Cook em [Tzanetakis & Cook, 2002] classificam música em 10 géneros musicais, nomeadamente música clássica, *country*, disco, *hip-hop*, *jazz*, *rock*, *blues*, *reggae*, *pop* e metal. Especializam ainda os classificadores utilizados em dois tipos de música: música *jazz* e música clássica. No caso do *jazz*,

consideram 6 subgêneros; *bigband*, *cool*, fusão, piano, quartetos e *swing* e no caso da música clássica 4 subgêneros; música coral, orquestra, piano e quarteto de cordas. As características tímbricas utilizadas são: centróide, *rolloff*, *flux*, MFCC e *zcr*. Estas características são calculadas em janelas de análise de pequena duração (23ms). Posteriormente são calculadas as médias e variâncias das características anteriores em intervalos de tempo de 1s. É calculada ainda a característica *low-energy* nos mesmos intervalos de 1s. As características rítmicas usadas são calculadas a partir do histograma de batidas da música e são A0, A1, RA, P1, P2 e SUM. As características do conteúdo de *pitch* são baseadas em técnicas de detecção múltipla de *pitch* e chamam-se FA0, UP0, FP0, IPO1 e SUM. Tanto as características rítmicas como as de conteúdo de *pitch* são calculadas em relação à música toda. No sentido de avaliar a importância das características foram utilizados dois classificadores: GMM e KNN. Os resultados de classificação alcançados foram de 61% para os 10 gêneros e 82,25% para a música clássica. Outro estudo com algumas variantes, dos mesmos autores juntamente com Georg Essl, foi publicado em [Tzanetakis et al., 2001 (1)].

Seth Golub [Golub, 2000] classifica música em sete gêneros bastante diferentes: *a cappella*, celta, clássica, eletrônica, jazz, latina e *pop-rock*. As características utilizadas são, *loudness*, centróide, largura de banda e uniformidade, bem como outras características estatísticas obtidas a partir delas. Foram utilizados três classificadores: Modelo linear generalizado (GLM)¹⁰, MLP e KNN. Os melhores resultados de classificação conseguidos foram de 67%.

Karin Kosina [Kosina, 2002] classifica em apenas três gêneros muito diferentes: metal, dança e clássica. Utiliza um conjunto de características que engloba MFCC, *zcr*, energia e *beat*. Foi conseguida uma taxa de sucesso na classificação de 88% para o classificador utilizado, KNN.

Hagan Soltau e colegas [Soltau et al., 1998] classificam música em quatro gêneros musicais, *rock*, *pop*, *techno* e clássica. Utilizam HMM e *explicit time modelling with neural networks* (ETM-NN). Os melhores resultados foram de 86,1%.

¹⁰ Em terminologia Inglesa: Generalized Linear Model - GLM

David Pye [Pye, 2000] utiliza MFCC e GMM para classificar música em seis géneros: *blues*, *easy listening*, clássica, ópera, dança e *rock*. É conseguida uma percentagem de sucesso de 92%.

Relacionado também com esta dissertação, no que concerne ao reconhecimento tímbrico de instrumentos, existe um estudo publicado por Agostini, Longari, e Pollastri [Agostini et al., 2003] que pretende avaliar um conjunto de características espectrais com o objectivo de classificar sons monofónicos de 27 instrumentos. O objectivo é identificar tons de instrumentos unicamente pelo seu timbre. Foram utilizados quatro classificadores: Análise canónica discriminante (CDA)¹¹, análise quadrática discriminante (QDA)¹², máquinas de suporte a vectores (SVM)¹³ e KNN. O melhor resultado de classificação para os 27 instrumentos foi de 92,8%.

Keith Martin [Martin, 1998] e [Martin & Kim, 1998] estuda igualmente o problema da identificação de instrumentos. Ele propõe um conjunto de características relacionadas com as propriedades físicas dos instrumentos com o objectivo de os identificar num ambiente polifónico. Outros estudos ainda sobre identificação de instrumentos foram publicados em [Fraser & Fujinaga, 1999].

É ainda utilizado por alguns autores o método de *clustering* conhecido por *Self-Organizing Map* (SOM) [Kohonen, 1989] que permite facilmente expressar graficamente os resultados da classificação. O SOM é de facto um tipo de rede neuronal. Um dos autores que o utiliza é Elias Pampalk [Pampalk, 2001], num treino não supervisionado para agrupar músicas pela sua similaridade. O resultado da classificação é um mapa de “ilhas” no qual cada ilha contém peças de música similares entre si. Frühwirth e Rauber [Frühwirth & Rauber, 2001] utilizam SOM’s para organizar colecções de música de acordo com o seu género e as suas características. São usadas características espectrais. A segmentação é efectuada em duas etapas: primeiro agrupa segmentos de música de acordo com a sua similaridade, e depois agrupa as

¹¹ Em terminologia Inglesa: Canonical Discriminant Analysis - CDA

¹² Em terminologia Inglesa: Quadratic Discriminant Analysis - QDA

¹³ Em terminologia Inglesa: Support Vector Machines - SVM

composições de acordo com a similaridade dos segmentos. Outro estudo com algumas variantes, dos mesmos autores, foi publicado em [Rauber & Frühwirth, 2001].

Existem ainda estudos não relacionados especificamente com a classificação de géneros musicais, mas com a discriminação de voz, música e outros sons (ambientais). Algumas das abordagens desses trabalhos são úteis para a classificação em géneros musicais.

Lu, Jiang e Zhang [Lu et al., 2001] propõem um conjunto de características com o objectivo de classificar sinais áudio em voz, música, sons ambientais e silêncio. Esta classificação é hierárquica, i.e., primeiro classifica-se em sinais com voz ou sem voz e em seguida classifica-se os sinais sem voz em música, sons ambientais e silêncio. Foram utilizados como classificadores o KNN e o *Linear Spectral Pairs – Vector Quantization* (LSP-VQ). Foi conseguida uma taxa de sucesso de 98,03% na discriminação voz, música. Variantes deste estudo podem ser analisadas em [Zhang & Kuo, 1998] e [Lu & Hankinson, 2001].

Liu e Wan [Liu & Wan, 2000] classificam áudio em voz, música e sons ambientais. São utilizados quatro classificadores: redes neuronais, KNN, GMM e rede neuronal probabilística (PNN)¹⁴. É conseguida uma taxa de sucesso de 93,7%.

¹⁴ Em terminologia Inglesa: Probabilistic Neural Network - PNN

Capítulo 3

EXTRACÇÃO DE CARACTERÍSTICAS

Neste capítulo irá ser explicado o porquê da extracção de características. Irão ser explanados ainda alguns aspectos sobre as peças de música utilizadas neste trabalho, como a sua escolha, tipo e tamanho. Em seguida, serão descritas as características extraídas, o seu cálculo e normalização de dados.

3.1. MOTIVAÇÃO

Para classificar, através do computador, música em géneros musicais ou em outro tipo de taxonomia, torna-se necessário definir uma representação para a música de forma a tornar viável essa mesma classificação. Assim, apenas 1 minuto de uma música extraída de um CD (44,1 khz, 16 bits, estéreo), ocupa em disco cerca de 10 MB ($44100 \cdot 2 \cdot 60 \cdot 2$ bytes). Pode-se optar pela redução da qualidade da música digitalizada para valores em que os objectivos da classificação não sejam todavia muito penalizados, e.g., o mesmo minuto a 22050 hz, em mono ocupa agora cerca de 2,5 MB. Mesmo que se considere que os parâmetros de digitalização anteriores são aceitáveis a nível da percepção das características fundamentais da música, 2,5 MB/min. continua a ser um

valor elevado, tendo ainda por cima em conta que a classificação (capítulo 4) pressupõe um processo de treino. Este processo de treino implica um grande número de peças musicais simultaneamente em memória, o que facilmente esgotaria os recursos computacionais disponíveis. Mais decisivo para não se estar ainda na situação ideal, é o facto de um classificador que no processo de treino tenha um número de entradas muitas vezes superior (2,5 milhões) ao número de músicas de treino não ser capaz de aproximar a função alvo com um mínimo de exactidão [Bishop, 1995].

Mostrou-se que o tamanho da representação de cada peça musical é muito importante para a viabilidade da classificação, mas não é o único factor. Mesmo que se consiga a redução do tamanho para valores aceitáveis, não se conseguirá produzir conhecimento com esta nova representação, já que consiste em informação crua. É necessário extrair características baseadas em propriedades do sinal a partir das quais se possam detectar padrões em relação aos vários géneros musicais.

Assim, a base de qualquer sistema automático de classificação de áudio em geral é a extracção de características, ou seja cada sinal áudio a classificar deverá ser representado por um vector de características. O tipo de características a extrair depende do tipo de classificação a efectuar.

Deste modo, no nosso caso, para cada peça musical foram extraídas 40 características que em conjunto definem a assinatura de cada música. Essa assinatura fica armazenada num ficheiro de dados com menos do que 1 KB.

A classificação automática de música, nomeadamente de géneros musicais, é relativamente recente comparado com a classificação de áudio em geral, cuja investigação desde há algum tempo investe no reconhecimento de voz e na discriminação entre voz e música. São reconhecidos normalmente na classificação automática de música três conjuntos de características importantes, conforme a função a que são destinadas: análise do timbre, análise do conteúdo rítmico e análise do conteúdo a nível de *pitch*.

3.2. FORMATO DE GRAVAÇÃO DAS MÚSICAS

As músicas foram todas extraídas a partir de CD's por uma questão de qualidade dos dados finais. Foram gravadas com uma frequência de amostragem de 22050 Hz, com 16 bits de quantização e com formato monoaural (mistura em partes iguais num só canal dos dois canais estéreo). Apesar da diminuição da frequência de amostragem para metade e da redução do número de canais para um, é um formato com uma qualidade muito aceitável no que diz respeito à percepção que o Homem tem da música¹⁵ e principalmente em relação à qualidade das características extraídas de cada música, como se poderá ver através dos resultados de classificação (capítulo 4).

Os extractos de música utilizados neste estudo foram armazenadas em ficheiros *wav*, já que o programa responsável pela extracção de características foi feito em *Matlab*, que dispõe de um conjunto de funções nativas para manipular esse tipo de ficheiros.

A aplicação utilizada para fazer a extracção dos segmentos de 6 segundos, foi o *Cool Edit Pro*.

3.3. EXTRACTOS MUSICAIS: ESCOLHA E DURAÇÃO

Qualquer tarefa de classificação pressupõe um processo de treino, cujo objectivo é definir um determinado conjunto de parâmetros que permitam modelar cada um dos géneros musicais a classificar. O objectivo é, ao tentar validar uma nova música, o classificador atribuir o género musical correcto à peça em questão. Com certeza, se isso acontecer é porque no conjunto dos exemplos de treino existem músicas típicas ou características de cada um dos géneros musicais e, conseqüentemente, o classificador “aprendeu” da melhor forma possível a identificar cada um dos géneros musicais.

¹⁵ Uma frequência de amostragem de 22050Hz é suficiente segundo o teorema de Nyquist para representar as frequências audíveis de sons provenientes da fala e da música.

Portanto para cada género musical deve ser seleccionado para treino um conjunto de músicas o mais abrangente possível para esse género musical. É claro que a qualidade das amostras escolhidas, no que toca ao seu conteúdo informativo, influencia o sucesso desse processo, assim como o número de músicas utilizadas para treino.

Neste trabalho recolheu-se um conjunto de extractos musicais com 6 segundos de duração. O objectivo deste estudo não é construir um sistema que utilize trechos de música de longa duração, mas sim de pequena duração e significativos para cada um dos géneros musicais. A ideia é imitar de alguma maneira, a forma como os seres humanos classificam a música [Perrot & Gjerdigen, 1999], i.e. conseguir classificar utilizando pequenos segmentos de música e usando apenas características extraídas directamente da análise de superfície feita ao sinal.

3.4. CONVERSÕES: TEMPO - FREQUÊNCIA

A representação habitual de cada música capturada é no domínio do tempo, i.e., cada amostra do seu sinal é um par ordenado com o tempo no eixo das abcissas e a amplitude do sinal no eixo das ordenadas. A resolução de cada uma das componentes depende do formato de gravação. Assim, como já foi referido, neste estudo a frequência de amostragem é de 22050 Hz e a quantização é de 16 bits. As capturas são feitas em formato monoaural, ou seja numa única pista.

O que distingue na maior parte das vezes um género musical de outro tem a ver com factores como o timbre, as frequências fundamentais ou o ritmo. Para distinguir estes factores é normalmente essencial um conhecimento profundo sobre o conteúdo de frequência do sinal. Este tipo de informações só são conhecidas se o sinal for convertido do domínio do tempo para o domínio da frequência. Nesta nova representação, o eixo das abcissas representa as frequências em Hz em que o sinal é composto e as ordenadas representam as magnitudes relativas a essas frequências, ou seja, as intensidades dessas frequências no sinal. As duas representações são perfeitamente equivalentes e portanto podem ser convertidas de uma para outra sem perda de informação.

A transformada de Fourier¹⁶ (TF) é precisamente o meio mais utilizado para converter um sinal do domínio do tempo para o domínio da frequência. Esta transformação é reversível.

Existem determinados conceitos inerentes à utilização da TF que importa conhecer. Um desses conceitos é que, segundo o teorema de Nyquist, a taxa de amostragem do sinal, i.e. o número de amostras por segundo tem de ser pelo menos duas vezes maior que a maior frequência presente no sinal. Por exemplo num sinal com frequência de amostragem igual a 22050 Hz, a frequência mais alta presente nesse sinal pode ir até 11025 Hz. Outro conceito é que não há informação sobre a altura no sinal em que ocorrem as frequências, e.g. sabe-se que existe uma frequência de 850 Hz com determinada magnitude, mas não se sabe em que altura no sinal é que essa frequência ocorre. Ora no caso em estudo, tal situação inviabilizaria qualquer associação entre um instante temporal da música e a sua representação espectral.

Considerem-se como exemplo dois sinais: um estacionário (um sinal em que todas as frequências presentes, existem em todos os momentos do sinal) com a duração de 1s e com as frequências de 10 Hz, 50 Hz, 100 Hz e 200 Hz e outro não estacionário (as frequências presentes podem existir em momentos diferentes do sinal) com a mesma duração e com as frequências anteriores presentes isoladamente em cada quarto de segundo de sinal. As suas representações, através da TF seriam muito semelhantes, como se verifica pela figura 3.1, mormente os sinais serem diferentes. Em consequência disso, a TF só deve ser aplicada em sinais estacionários.

¹⁶ Jean Baptiste Fourier (1768-1830)

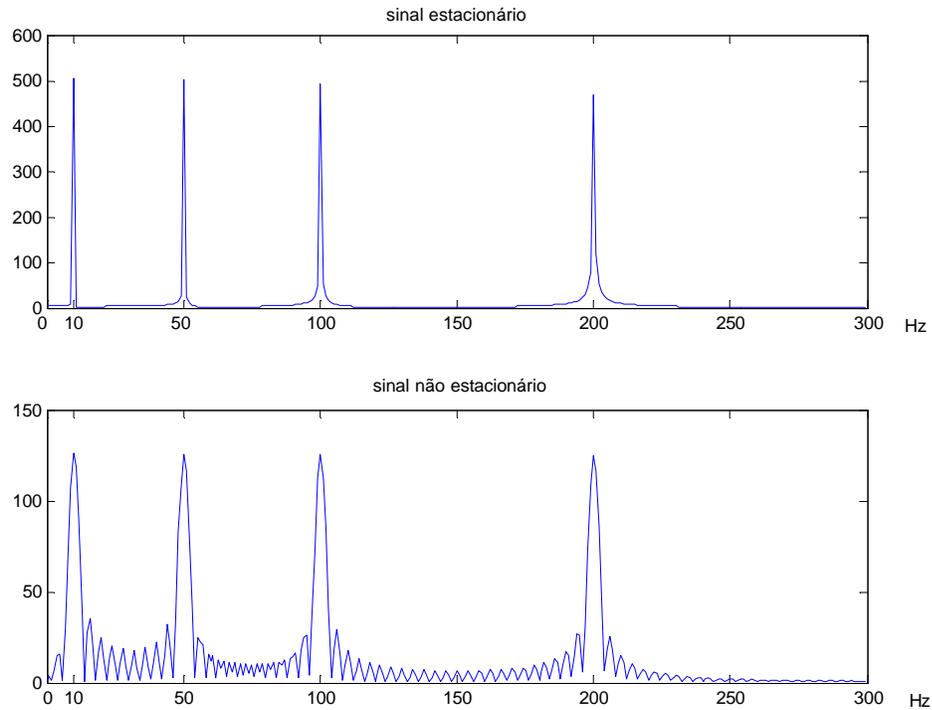


Figura 3.1. Transformada de Fourier de dois sinais: um estacionário e outro não estacionário, com frequências de 10, 50, 100 e 200 Hz.

Na maioria das aplicações da vida corrente que utilizam processamento de sinais, tanto na Engenharia como em qualquer outra área, é fundamental aliar na mesma representação informações temporais e espectrais. Por exemplo, em processos de diagnóstico médico como os electrocardiogramas ou electroencefalogramas, a informação espectral pode servir para diagnosticar doenças, quando algo de anormal é detectado a nível da informação temporal.

Uma das técnicas mais utilizadas para unir as duas representações numa só, é a chamada Transformada de Fourier para Pequenos Segmentos (STFT)¹⁷ [Polikar, 2003].

¹⁷ Em terminologia Inglesa: Short-Time Fourier Transform - STFT

Como se viu, a Transformada de Fourier não deve ser aplicada em sinais não estacionários. Mas se for possível dividir o sinal em sinais mais pequenos que por sua vez sejam estacionários, então o problema ficará resolvido e poder-se-á aplicar a TF a cada segmento. A esta técnica chama-se STFT.

Uma das formas de implementar a STFT, consiste em escolher um tamanho fixo para os segmentos de sinal em que estes sejam estacionários, multiplicar cada um desses segmentos por uma função janela deslizante, com o mesmo tamanho dos segmentos e achar a TF desses produtos. A partir desse momento passa-se a ter informação temporal/espectral do sinal, já que tem-se as componentes espectrais dos segmentos de sinal que estão definidos temporalmente no sinal original. Interessa saber que se o tamanho da janela for menor, aumenta-se a resolução temporal e diminui-se a espectral, se for maior, aumenta-se a resolução espectral e diminui-se a temporal. É preciso arranjar um compromisso de qualidade, de acordo com as necessidades, para o tamanho da janela. Para se perceber melhor, supondo que para a mesma frequência de amostragem, e.g. 44100 Hz, se tem numa primeira situação janelas com 512 amostras e numa segunda situação com 1024 amostras, a largura das janelas é respectivamente para a primeira e segunda situações 11,6 ms e 23,22 ms. As resoluções de frequência são também respectivamente 86,13 Hz e 43,06 Hz. Como se vê aumentando a largura da janela, aumenta-se a resolução espectral (86,13 para 43,06 Hz), por outro lado, diminui-se a resolução temporal (11,6 para 23,22 ms).

Ao utilizar a STFT, a escolha da janela a utilizar é fundamental para diminuir o esbatimento espectral¹⁸ que se gera com a sua aplicação. Assim, janelas como as de Hanning ou Hamming têm um menor esbatimento espectral que a janela rectangular, no entanto perdem em resolução. Foi utilizada neste trabalho a janela de Hanning que é um bom compromisso entre resolução e esbatimento espectral. No entanto a aplicação desta janela vai provocar uma redução da amplitude das amostras do centro para as extremidades da janela. Deste modo, há necessidade de sobrepor janelas consecutivas com o objectivo de amostras que pertençam a duas janelas contíguas, terem as suas amplitudes atenuadas em relação a cada uma das janelas, de forma a que somando essas duas amplitudes, o resultado seja a amplitude original referente a essa amostra. A figura

¹⁸ Em terminologia Inglesa: Spectral leakage

3.2 mostra precisamente a perda de informação entre janelas quando se multiplica pelo sinal janelas de Hanning sem sobreposição. Tipicamente, são utilizadas taxas de sobreposição de 50% entre janelas contíguas, por forma a reduzir a perda de informação na fronteira entre janelas.

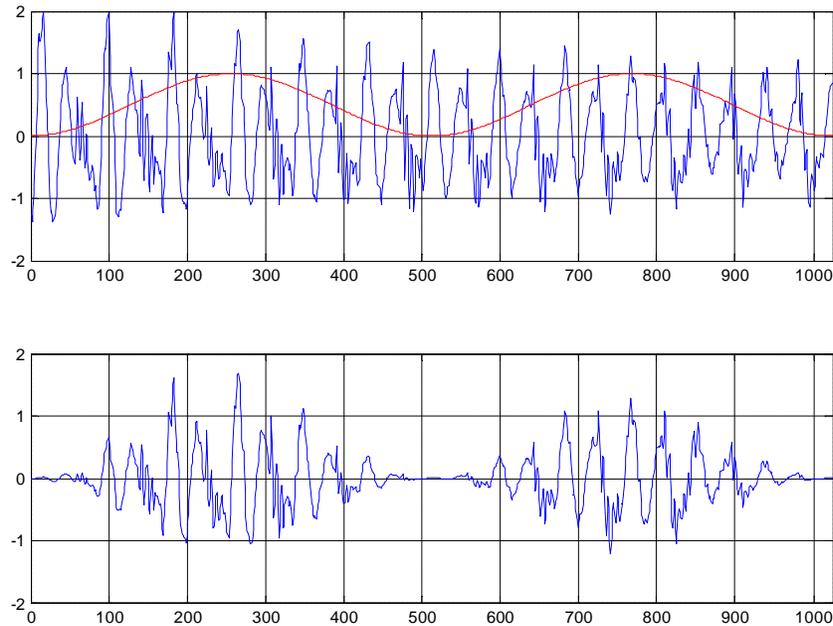


Figura 3.2. Resultado da aplicação de duas janelas de Hanning contíguas a um sinal.

Resta dizer que, neste trabalho quando se fala em Transformada de Fourier, nos referimos à Transformada Discreta de Fourier (DFT)¹⁹ [Smith, 1997], aplicada em sinais periódicos e discretos, não fosse o seu objectivo de utilização o computador. A DFT só pode, aliás, ser utilizada em sinais infinitos, logo, para a sua utilização no computador, supõe-se que o sinal a processar se prolonga para a esquerda e para a direita infinitamente, sendo esses prolongamentos duplicações repetidas do próprio sinal. O sinal infinito é portanto periódico com período igual ao do segmento de sinal original.

¹⁹ Em terminologia Inglesa: Discrete Fourier Transform - DFT)

Em 1965, J.W Cooley e J.W Tukey [Cooley & Tukey, 1965] criaram um algoritmo²⁰ que torna muito mais eficiente e mesmo praticável a utilização da DFT num computador. Este algoritmo chama-se Transformada Rápida de Fourier (FFT)²¹ [Smith, 1997]. Esse aumento de eficiência é de tal ponto significativo que se pode fazer a analogia entre andar a pé e passar a andar num avião a jacto [Smith, 1997]. Basta dizer que, num computador o tempo de execução da DFT é de ordem N^2 e o do FFT de ordem $N \cdot \log(N)$, em que N é o número de amostras do sinal.

Deste ponto em diante, todas as referências à utilização de uma transformada de Fourier, referem-se à utilização da FFT.

3.5. CARACTERÍSTICAS

Foram extraídas para cada peça de música um total de 40 características. Elas compõem na sua totalidade a assinatura dessa peça.

De acordo com os tipos de classificação a efectuar, baseados essencialmente na discriminação entre instrumentos e na discriminação entre voz e parte instrumental (Capítulo 4) e tendo como base trabalhos efectuados por outros autores como Golub [Golub, 2000] e Tzanetakis [Tzanetakis & Cook, 2002] as características extraídas de cada peça musical foram escolhidas de forma a privilegiar a análise do timbre e do *pitch* do sinal. Não foram utilizadas características rítmicas já que não pareceram relevantes para os problema de classificação em questão.

O processo de extracção de características começa para cada uma das músicas representadas, por dividir o sinal de 6s em janelas de 23,22 ms com 50% de sobreposição entre duas janelas consecutivas. Se não houvesse sobreposição, iria haver perdas de informação nos limites entre janelas contíguas. Em seguida o sinal em cada uma das janelas é multiplicado pela função de Hanning, o que faz com que as

²⁰ A técnica já tinha sido criada por Karl Friedrich Gauss (1777-1855), mas faltava-lhe a ferramenta, o computador, para tornar possível a sua utilização

²¹ Em terminologia Inglesa: Fast Fourier Transform - FFT

amplitudes das amostras de sinal de cada janela, sejam atenuadas a partir das amostras centrais para as extremas. A amostra central não é atenuada (multiplica por 1) e a mais extrema é totalmente atenuada (multiplica por 0).

Neste ponto, são extraídas directamente do sinal, para cada janela, o *loudness* e o *zcr*. São portanto extraídas no domínio do tempo.

O comprimento específico de cada janela de 23,22 ms, foi escolhido por forma ao número de amostras em cada janela ser potência de 2 ($512 = 2^9$), o que é fundamental para otimizar a eficiência da FFT [Smith, 1997].

Após a aplicação da FFT a cada janela, são extraídas três características espectrais: o centróide, a largura de banda e a uniformidade. A partir destas cinco características base são calculadas por processos estatísticos todas as 40 características que irão representar cada peça de música.

Nos 6 s de cada peça musical, existem portanto 132300 (6×22050) amostras. Sabendo que cada janela tem 23,22 ms e estas têm 50% de sobreposição, existem no total 512 amostras ($512 \times \text{período de amostragem do sinal} = 0,02322 \text{ s}$) por janela de um total de 515 janelas ($((132300 - 512/2) / (512/2) = 515)$). Em cada janela, a resolução a nível de frequência²² é de 43,06 Hz, i.e., no domínio da frequência de uma amostra para a seguinte, existe um salto de 43,06 Hz.

As características são calculadas em três etapas. Primeiro em relação a cada janela (características base). Depois em intervalos de 2 s de sinal (características intermédias) e finalmente em relação a todo o sinal (características finais). Cada um destes três passos, a começar no segundo, é calculado utilizando as características calculadas no passo anterior.

A música clássica, objecto deste estudo, caracteriza-se em geral por variações acentuadas nas características base, descritas seguidamente, ao longo do tempo. Por

²² Resolução de frequência = (frequência de amostragem) / (número de amostras) [Smith, 1997].

isso, pensa-se que as manipulações estatísticas em relação a essas características poderão influir na obtenção de bons resultados.

Vai ser descrito em seguida todo o processo para o cálculo das características tanto temporais como espectrais.

3.5.1. CARACTERÍSTICAS BASE

Ao pretender-se extrair características da música para poderem ser utilizadas num computador para efeitos de algum tipo de classificação, o objectivo é, como é óbvio, facilitar a vida ao Homem, poupando-lhe trabalho em tarefas repetitivas e mesmo difíceis de realizar caso o conjunto de dados a processar seja muito grande. É claro que o grande objectivo é conseguir resultados de classificação pelo menos tão bons como os resultados alcançados pelo ser humano. Para tal é necessário pautar os critérios de classificação do computador pelos critérios de classificação do Homem, já que estes é que definem o que é um resultado correcto. Esta colagem de critérios começa logo pela extracção de características, já que se sabe que a percepção do ser humano em relação às frequências não é linear, mas sim à escala \log_2 [Golub, 2000]. Assim, a todas as características baseadas na frequência do sinal, é aplicado \log_2 . Já em relação à característica *loudness* (a percepção que o ser humano tem da intensidade do som), dentro da mesma frequência, a nossa percepção é igualmente logarítmica [Golub, 2000].

São calculadas no total 10 características base. Dividem-se em três grupos: características extraídas no domínio do tempo, características extraídas no domínio da frequência e características extraídas a partir de duas janelas consecutivas.

Domínio do Tempo

Foram utilizadas duas características extraídas no domínio do tempo. O *loudness* e o número de intersecções com o eixo das abcissas ou seja do tempo (*zcr*).

Loudness

É uma característica perceptual que tenta captar a percepção que o ouvido humano tem da intensidade do som. A informação extraída do sinal sonoro que vai servir de base ao cálculo do *loudness*, é a amplitude.

Assim, o *loudness*, i.e., a percepção da amplitude pode ser representada pela equação seguinte (3.1):

$$L(r) = \log_2 \left(1 + \frac{1}{N} \sum_{n=1}^N |x(n)| \right) \quad (3.1)$$

onde L representa o *loudness*, r o número da janela actual, N é o número de amostras em cada janela, n é o número da amostra actual na janela actual e finalmente $x(n)$ representa a amplitude da n -ésima amostra na janela actual.

De notar que a amplitude pode ser influenciada pelo nível da gravação efectuada. Assim, para que o *loudness* possa ser considerado para a diferenciação de músicas, é ideal que estas partam dos mesmos pressupostos no que concerne à sua captação. Apesar de não haver garantias totais disso, a medida a tomar foi só considerar para este estudo músicas extraídas a partir de CD.

ZCR

Esta característica mede simplesmente o número de vezes que o sinal sonoro atravessa o eixo das abcissas (tempo).

Esta característica pode ser representada por (3.2):

$$Z(r) = \frac{1}{2} \sum_{n=1}^N |\operatorname{sgn}(x(n)) - \operatorname{sgn}(x(n-1))| \quad (3.2)$$

Na expressão anterior, $Z(r)$ representa o número de intersecções com o eixo das abcissas que existem na janela r e $\operatorname{sgn}(x(n))$ representa o sinal da amplitude da n -ésima amostra da janela r .

Esta é uma medida do conteúdo de frequência do sinal. É muitas vezes usada em problemas de discriminação entre música e voz e para determinar a quantidade de ruído de um sinal [Tzanetakis & Cook, 2002].

Domínio da Frequência

As características espectrais utilizadas, calculadas no domínio da frequência, foram o centróide, a largura de banda e a uniformidade.

Como foi visto na secção 2.4, o processo que culmina no cálculo das características espectrais anteriores, começa pela conversão do sinal para o domínio da frequência, utilizando para tal a STFT.

Centróide

Esta característica espectral pode ser definida como a média pesada das magnitudes das frequências. É também um indicador do “brilho” do sinal [Wold et al., 1996]. Assim, valores altos para esta característica indicam um sinal com maior brilho e frequências globalmente mais altas para esse sinal. Wold explica claramente este conceito de brilho com uma experiência: se ao emitir um som, se puser a mão à frente da boca, vai-se diminuir o brilho e o *loudness* do som.

Normalmente o centróide reflecte-se na voz por valores mais baixos e na música por valores mais altos, portanto é considerada uma característica fundamental para a discriminação entre voz e música.

Esta característica pode ser representada pela seguinte equação (3.3):

$$C(r) = \frac{1}{N} \frac{\sum_{k=1}^N M_r(k) \cdot \log_2 k}{\sum_{k=1}^N M_r(k)} \quad (3.3)$$

onde $C(r)$ representa o valor do centróide na janela r e $M_r(k)$ representa a magnitude da transformada de Fourier na janela r e no índice de frequências k .

Largura de Banda

A definição desta característica espectral pode ser dada como a média pesada dos desvios padrões das bandas de frequência, ou muito simplesmente como desvio padrão da frequência. Se esta característica tiver um valor baixo, isso significa que as frequências do sinal concentram-se todas perto do centróide, i.e., há uma gama mais estreita de frequências no sinal.

Para se perceber melhor, uma sinusóide seno tem largura de banda igual a zero, enquanto um ruído tem normalmente uma largura de banda elevada.

A equação (3.4) permite calcular esta característica:

$$B(r) = \sqrt{\frac{\sum_{k=1}^N (C(r) - \log_2 k)^2 M_r(k)}{\sum_{k=1}^N M_r(k)}} \quad (3.4)$$

onde $B(r)$ representa a largura de banda da janela r , $C(r)$ é como verificamos atrás o centróide dessa mesma janela.

Uniformidade

A última característica espectral calculada foi a uniformidade. Mede a similaridade entre as magnitudes das bandas de frequência presentes no sinal.

Esta característica é fundamental para discriminar entre sinais com magnitudes muito altas para um reduzido número de frequências e sinais em que os valores das magnitudes são parecidos para a grande maioria das frequências. No caso extremo uma sinusóide tem uniformidade igual a zero, enquanto um sinal em que o ruído existente seja claro, deverá ter um valor elevado para esta característica.

Esta característica pode ser representada pela equação que se segue (3.5):

$$U(r) = - \sum_{k=1}^N \frac{M_r(k)}{\sum_{k=1}^N M_r(k)} \cdot \log_N \frac{M_r(k)}{\sum_{k=1}^N M_r(k)} \quad (3.5)$$

onde $U(r)$ representa a uniformidade da janela r .

Primeiras Diferenças

São ainda calculadas as diferenças dos valores das cinco características anteriores entre janelas consecutivas e.g. $L(r) - L(r-1)$ para o caso do *loudness* em que r representa a janela corrente.

O que se pretende com o cálculo das primeiras diferenças é ter uma ideia da variação no tempo ou da trajectória das características fundamentais.

Estas novas cinco características juntamente com as cinco calculadas anteriormente para cada janela constituem na totalidade as características base utilizadas neste trabalho.

3.5.2. CARACTERÍSTICAS INTERMÉDIAS

Para cada uma das 10 características base, são calculadas, de 2 em 2 segundos, as características intermédias. São constituídas pelas médias e pelos desvios padrões dos valores de cada característica base em todas as janelas em cada intervalo de 2 segundos. Como os extractos de música neste trabalho têm 6 segundos, isso quer dizer que estas características são calculadas 3 vezes para cada uma das características base. Portanto existem 20 (2x10) características intermédias para cada intervalo de 2s de sinal.

3.5.3. CARACTERÍSTICAS FINAIS

Calculando as médias e desvios padrões das características intermédias, chegamos às características que constituem no seu conjunto a representação de cada

extracto musical. Essa representação é também chamada de assinatura do extracto musical. A assinatura é portanto constituída por 40 características (2x2x10).

Na tabela seguinte (Tabela 3.1) estão discriminadas as 40 características utilizadas neste trabalho.

1. média(média(zcr))	21. média(média(zcrdif))
2. desviop(média(zcr))	22. desviop(média(zcrdif))
3. média(desviop(zcr))	23. média(desviop(zcrdif))
4. desviop(desviop(zcr))	24. desviop(desviop(zcrdif))
5. média(média(loudness))	25. média(média(loudnessdif))
6. desviop(média(loudness))	26. desviop(média(loudnessdif))
7. média(desviop(loudness))	27. média(desviop(loudnessdif))
8. desviop(desviop(loudness))	28. desviop(desviop(loudnessdif))
9. média(média(centróide))	29. média(média(centróidedif))
10. desviop(média(centróide))	30. desviop(média(centróidedif))
11. média(desviop(centróide))	31. média(desviop(centróidedif))
12. desviop(desviop(centróide))	32. desviop(desviop(centróidedif))
13. média(média(larguradebanda))	33. média(média(larguradebandadif))
14. desviop(média(larguradebanda))	34. desviop(média(larguradebandadif))
15. média(desviop(larguradebanda))	35. média(desviop(larguradebandadif))
16. desviop(desviop(larguradebanda))	36. desviop(desviop(larguradebandadif))
17. média(média(uniformidade))	37. média(média(uniformidadedif))
18. desviop(média(uniformidade))	38. desviop(média(uniformidadedif))
19. média(desviop(uniformidade))	39. média(desviop(uniformidadedif))
20. desviop(desviop(uniformidade))	40. desviop(desviop(uniformidadedif))

Tabela 3.1. Assinatura de cada extracto musical.

Exemplificando para a 10^a característica, parte-se do cálculo do centróide para cada uma das 515 janelas, para em seguida se calcular as médias desses centróides em intervalos de 2 segundos (cerca de 171 janelas). Finalmente calcula-se o desvio padrão das 3 médias calculadas anteriormente. Já na 34^a característica, a única diferença é que em vez de se partir do cálculo do centróide para cada janela, parte-se do cálculo da diferença de largura de banda entre janelas consecutivas.

3.6. NORMALIZAÇÃO DOS DADOS

As diversas características ao serem calculadas podem resultar em valores de ordem de grandeza diferentes umas das outras. Os vários classificadores, nomeadamente as redes neuronais e os K -vizinhos mais próximos, são sensíveis à escala das características principalmente de umas em relação às outras [Golub, 2000].

É fundamental que os diferentes classificadores utilizem os mesmos dados para os resultados comparativos terem sentido. Para tal foi definida para cada característica uma normalização uniforme (3.6):

$$m_i(c_j) = \frac{(m_i(c_j) - c_j^{min})}{(c_j^{max} - c_j^{min})} \quad (3.6)$$

onde $m_i(c_j)$ representa o valor da característica j para a música i . c_j^{min} e c_j^{max} designam respectivamente os valores mínimo e máximo da característica j para todas as músicas em estudo.

Capítulo 4

CLASSIFICAÇÃO

Após a identificação da forma como se pretende classificar as músicas (taxonomias utilizadas) é necessário definir e extrair as características de cada música, aplicar o classificador e analisar os resultados. Foram utilizados neste estudo três classificadores, para assim determinar o melhor para cada problema de classificação

Este capítulo começa precisamente por definir quais os problemas de classificação que são objectivo deste estudo. Em seguida, cada um dos classificadores utilizados, *K*-vizinhos mais próximos, modelos de misturas gaussianas e Redes Neurais Artificiais (Percepção Multicamada), são analisados, sendo essa análise direccionada para os objectivos deste trabalho.

4.1. TAREFAS DE CLASSIFICAÇÃO

A classificação de sinais de áudio e em particular de sinais musicais, é uma área de investigação relativamente recente e está de alguma maneira associada ao progresso e às novas necessidades da Internet. A ideia será construir um sistema que consiga dizer, para qualquer sequência musical, qual a categoria em que ela se insere (e.g. *rock*, clássica). O que a investigação tem mostrado é que existem características da

música que podem ser relevantes para distinguir um determinado género musical, mas completamente insignificante para outros, isto é, a extracção de características a efectuar deve levar em conta o problema de classificação em questão.

Alguns investigadores têm optado por considerar problemas de classificação em géneros musicais que abrangem uma grande diversidade de estilos [Tzanetakis & Cook, 2002]. Neste trabalho, a ideia não foi tentar construir um sistema global, mas sim um subsistema que permitisse assim classificar subgéneros musicais com grande similaridade.

A taxonomia utilizada neste trabalho foi a ilustrada na Figura 4.1.

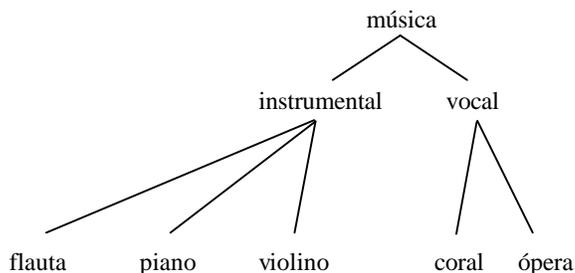


Figura 4.1. Taxonomia utilizada

Consideram-se três problemas de classificação a partir da taxonomia anterior. O primeiro consiste em classificar as peças de música em três classes de música instrumental: peças para flauta, piano e violino. No segundo, o objectivo é classificar em duas classes: música coral e ópera. Finalmente, no terceiro problema, consideram-se os cinco géneros musicais anteriores.

A taxonomia anterior dá a ideia que iria ser feita uma classificação hierárquica, primeiro em música instrumental ou vocal e em seguida, mediante a resposta, numa das subárvores. De facto, utiliza-se a taxonomia desta forma por uma questão de clareza, de informação sobre o tipo de músicas escolhidas e já agora segundo uma perspectiva de trabalho futuro. No entanto, a implementação prática conduzida, não é neste momento hierárquica.

4.1.1. Primeira tarefa de classificação

É mostrada em seguida (Figura 4.2) a taxonomia utilizada nesta primeira tarefa de classificação.

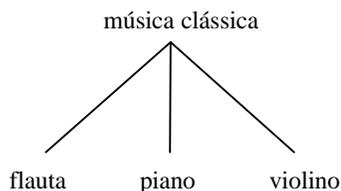


Figura 4.2. Primeira tarefa de classificação. Taxonomia

O primeiro problema de classificação, aborda a organização de peças musicais, segundo uma perspectiva do instrumento musical predominante. Assim, uma peça musical pertence a cada uma das classes flauta, piano e violino se essa peça for um solo de um desses instrumentos (e.g. um ou mais violinos) ou tendo orquestra, esses instrumentos forem predominantes. Esta forma de classificação foi sempre utilizada no âmbito do mundo real da música clássica, já que os próprios compositores escreveram muitas das suas obras especificamente para um ou vários instrumentos musicais. Algumas obras famosas que ilustram o tipo de classificação anterior são, *Piano Concerto n.º 1* de Beethoven, *Concerto for Flute, Harp and Orchestra in C.K. 299* de Mozart ou *Sonata No.2 in A minor, BWV 1003 for Solo Violin* de J.S. Bach. Além do mais, esta taxonomia, de distinção de instrumentos musicais é das mais intuitivas e preferidas para o ser humano que se interessa por música clássica. A provar, está o facto da maioria dos *sites* especializados em música, como por exemplo *AllMusicGuide* (<http://www.allmusic.com/>), ou *CDNOW* (<http://www.cdnw.com/>), a utilizarem largamente.

Os extractos de música escolhidos para este estudo tentaram ser o mais abrangentes possível para cada um dos subgéneros musicais em questão: composições a solo ou com orquestra, com o instrumento musical principal interpretado de formas diversas (e.g. um piano tocado de uma forma muito enérgica *Allegro con brio* de *Piano Concerto N.º2 in B Flat Major, Op 19* de Beethoven ao contrário do tipo de composição a que o associamos mais facilmente – *Für Elise* de Beethoven).

4.1.2. Segunda tarefa de classificação

Vamos visualizar em seguida (Figura 4.3) a taxonomia utilizada neste segundo problema de classificação.

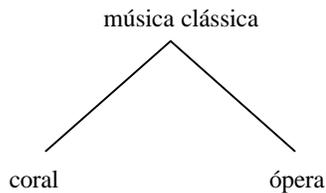


Figura 4.3 – Segunda tarefa de classificação. Taxonomia

No segundo problema, o objectivo é classificar em subgéneros da música clássica vocal – música coral e ópera. Antes de mais nada, importa definir o que se entende por cada um destes estilos, já que nomeadamente a ópera é de facto uma representação teatral em que os actores cantam poemas líricos com o acompanhamento de uma orquestra. Há uma característica que a distingue normalmente, que tem a ver com o tipo de solistas normalmente utilizados (tenor²³, soprano²⁴, mezzo-soprano²⁵). A forma como numa ópera eles cantam (e.g., *tremolo*, *vibrato*), faz com que normalmente o ser humano consiga distinguir com uma percentagem razoável de sucesso a ópera da música coral. Por sua vez, este último estilo é normalmente caracterizado por várias vozes em coro, sem *vibrato* nem *tremolo*.

O que acontece muitas vezes, é a mesma peça musical poder ser classificada de várias maneiras, dependendo do trecho em análise; por exemplo uma parte não cantada

²³ Tenor é a segunda voz masculina mais alta. Enrico Caruso, Ben Heppner e Luciano Pavarotti são três famosos tenores.

²⁴ Soprano é a voz feminina mais alta. Joan Sutherland e Maria Callas são duas famosas sopranos.

²⁵ Mezzo-soprano é a voz feminina entre a soprano e a alto. Cecilia Bartoli, Marilyn Horne e Anne Sofie Von Otter são três famosas mezzo-sopranos.

de uma determinada ópera pode ser interpretada como música para violino, se este for predominante, ou música coral numa parte mais calma da ópera.

Alguns dos trechos escolhidos neste estudo, não deixam margem para dúvidas sobre a classe a que pertencem. Por exemplo interpretações de Maria Callas, Luciano Pavarotti, Plácido Domingo ou Montserrat Caballé são claramente associadas a ópera. Já Cantos Gregorianos, *Requiem's*, ou *Missas* são normalmente relacionadas com música coral. Outras no entanto demonstram bastante ambiguidade como por exemplo a introdução (*Fortuna Imperatrix Mundi*) de *Carmina Burana* de Carl Orff.

4.1.3. Terceira tarefa de classificação

Neste tarefa de classificação, consideraram-se os cinco estilos musicais em simultâneo (Figura 4.4).

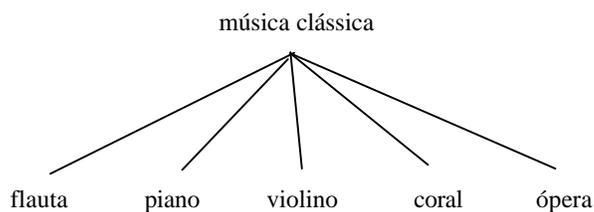


Figura 4.4. Terceira tarefa de classificação. Taxonomia

Tentou-se ver até que ponto, utilizando as mesmas características dos problemas anteriores, se conseguem bons resultados para uma classificação em que a similaridade entre os géneros musicais não é tão óbvia e em que mesmo o objectivo da classificação não é tão claro. Enquanto no primeiro problema de classificação o objectivo é a análise do instrumento, no segundo é a análise da voz. Devido à mistura, neste classificador de géneros vocais e instrumentais pode perfeitamente existir porventura uma música coral em que, por exemplo, a parte instrumental do piano seja predominante, o que poderá provocar que a música seja classificada como piano. Por outro lado, analisando ao pormenor os resultados obtidos, poder-se-á chegar à conclusão, que os valores obtidos para a classe coral também foram elevados. Nesse

caso, poder-se-ia concluir que o problema talvez não estivesse nos resultados, mas sim na própria taxonomia utilizada.

De qualquer forma, mesmo com algumas reticências, decidiu-se analisar este problema, através dos vários classificadores em estudo.

4.2. *K-VIZINHOS MAIS PRÓXIMOS*

O primeiro classificador a analisar é o algoritmo dos K -vizinhos mais próximos²⁶, um classificador estatístico não paramétrico de implementação muito simples. Os dados (vectors de características) estão divididos em dois conjuntos: o conjunto dos padrões de treino e o conjunto de teste ou validação. Sabe-se à partida qual a classe a que pertencem os dados do primeiro conjunto, consistindo a classificação em calcular as distâncias de cada vector de características de teste a todos os padrões de treino. São determinados os K vectors de características de treino mais próximos de cada vector de teste. A classificação da instância de teste representada pelo vector de características em análise é por conseguinte efectuada de acordo com a classe mais representada no conjunto dos K vizinhos mais próximos.

De seguida é exemplificada graficamente a utilização do algoritmo, com $K=3$. Assume-se que se pretende catalogar um conjunto de instâncias de determinado tipo em duas classes A e B e que cada instância é representada por duas características de forma a possibilitar a representação das mesmas num espaço de características de duas dimensões. Os elementos da classe A são representados por um círculo branco, enquanto os da classe B são representados por um círculo preto.

A figura 4.5 apresenta os elementos das duas classes, representados no espaço de características X , que pertencem ao conjunto de treino. São 20 elementos no total.

²⁶ em terminologia Inglesa: K -Nearest Neighbors

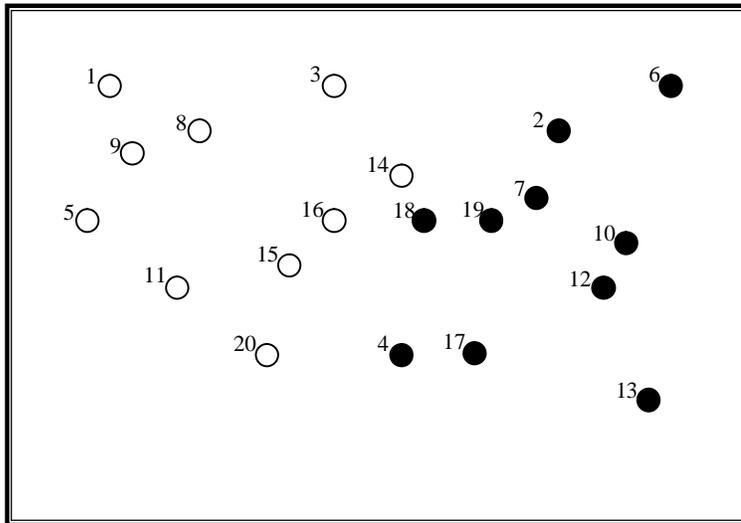


Figura 4.5. Padrões de treino representados no Espaço de Características X

A figura 4.6 representa o processo de teste para um determinado elemento (é representado na figura por um quadrado). São calculadas as distâncias euclidianas do elemento de teste a todos os elementos do conjunto de treino.

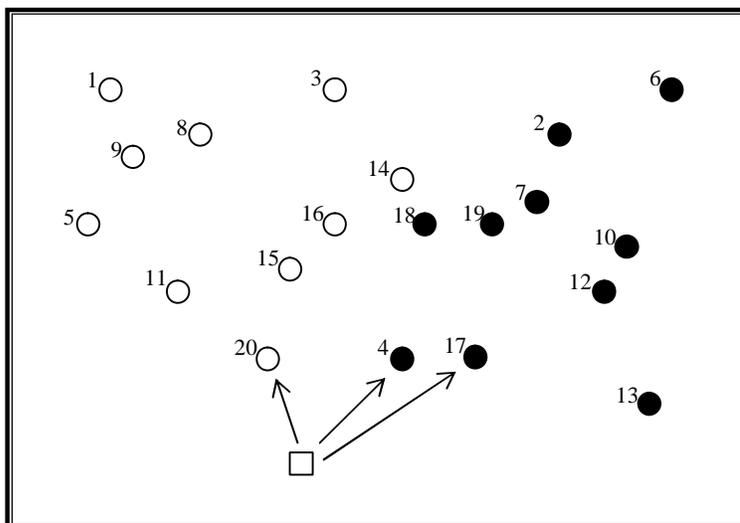


Figura 4.6. Cálculo dos vizinhos mais próximos de um Exemplo de teste

Como K é igual a 3, só são considerados os três elementos mais próximos do espaço de características. São eles, a contar do mais próximo, o elemento 20 (classe A), o elemento 4 (classe B) e o elemento 17 (classe B). Existem dois elementos da classe B

para um da classe A . O elemento de teste é portanto classificado na classe B , apesar do mais próximo pertencer à classe A .

Pode acontecer que para um dado exemplo de utilização deste algoritmo, as classes da vizinhança estejam igualmente representadas (e.g., existem 3 classes representadas no espaço de características e $K=3$. Os 3 vizinhos mais próximos de um dado exemplo de teste são um elemento de cada uma das 3 classes). Nesses casos, será seleccionada a classe a que corresponde o exemplo mais próximo.

O método dos K -vizinhos mais próximos é sintetizado no Algoritmo 4.1.

Seja X um conjunto de vectores de características de dimensão N . São conhecidas as classes a que pertencem.
Seja T um conjunto de vectores de características de teste com dimensão n .

Dado K (número de vizinhos a considerar para o algoritmo).

Para cada vector de características p de T :

- Calcular a distância Euclidiana entre p e cada um dos elementos de X . Guardar os valores calculados num vector de distâncias, (4.9).

$$d(p, x) = \sqrt{\sum_{r=1}^n (p_r - x_r)^2} \quad (4.1)$$

$$p \in T, x \in X$$

- Calcular os K vizinhos mais próximos i.e. as menores K distâncias do item anterior.
- Classificar p de acordo com a classe mais representada nos K vizinhos mais próximos.

Algoritmo 4.1. K -vizinhos mais próximos

4.3. MODELOS DE MISTURAS GAUSSIANAS

Na área de reconhecimento estatístico de padrões, o problema da classificação consiste na estimativa da função densidade de probabilidade (*fdp*) para os vectores de características de cada classe ou *cluster*. Os modelos de misturas gaussianas (GMM), também conhecidos por redes neuronais probabilísticas [Streit & Luginbuhl, 1994], são uma técnica para estimar essas funções, já que modelam cada classe como uma combinação linear (mistura) de várias funções de densidade de probabilidade.

O objectivo da utilização dos GMM é maximizar uma função de verosimilhança. Esta função dá-nos uma medida da forma como a *fdp* se ajusta ao conjunto de dados. Cada *fdp* é caracterizada por um conjunto de parâmetros e portanto a ideia deste algoritmo é estimar esses parâmetros de forma a que o ajustamento da *fdp* correspondente aos dados de treino, seja o melhor possível. Para o cálculo desses parâmetros utiliza-se um algoritmo iterativo muito conhecido: o algoritmo *Expectation-Maximization* (EM). De notar, como se viu anteriormente, que nos GMM a *fdp* é de facto uma mistura de *fdp*'s.

Este algoritmo garante que a função de verosimilhança é monótona não decrescente e que converge pelo menos para um máximo local.

Em seguida vão ser explicitadas algumas das equações principais utilizadas pelo classificador. Serão ainda descritas, em secções diferentes, a forma de funcionamento do algoritmo no que concerne ao treino e à validação dos dados.

4.3.1. Equações Fundamentais

As equações para o cálculo da combinação linear das *fdp*'s e para o cálculo de cada uma das *fdp* são apresentadas a seguir, em 4.2 e 4.4 respectivamente:

$$p(x \setminus \mathbf{q}) = \sum_{i=1}^v w_i p_i(x) \quad (4.2)$$

onde x é um vector de características de dimensão d . O inteiro v é o número de densidades gaussianas, classes ou *clusters* no espaço de características. Os w_i representam os pesos da combinação linear. Obviamente que a soma dos pesos das classes em cada vector de características é igual a 1, ou seja (4.3):

$$\sum_{i=1}^v w_i = 1 \quad (4.3)$$

A função de densidade de probabilidade para cada um dos *clusters* é representada por $p_i(x)$. De notar que $p_i(x)$ é de facto a mesma coisa que escrever $p(x \setminus \mathbf{m}_i, R_i)$. É uma questão de escolha de sintaxe.

$$p_i(x) = \frac{1}{(2\mathbf{p})^{d/2} |R_i|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mathbf{m}_i)^T R_i^{-1} (x - \mathbf{m}_i) \right\} \quad (4.4)$$

Em (4.4), \mathbf{m}_i e R_i representam respectivamente a média e a matriz de covariância do *cluster* i que fica completamente caracterizado com esses dois parâmetros. Finalmente, (4.5):

$$\mathbf{q} = \{w_i, \mathbf{m}_i, R_i\} \\ i = 1, \dots, v \quad (4.5)$$

\mathbf{q} representa todos os parâmetros do modelo de misturas gaussianas.

4.3.2. Processo de Treino

O treino deste classificador é efectuado com recurso ao algoritmo iterativo EM. Este algoritmo é composto por dois passos que se repetem ciclicamente até se atingir a convergência. Em cada ciclo ou iteração, os parâmetros do classificador serão ajustados de forma a maximizar-se uma função de verosimilhança. Como se pode verificar através do Algoritmo 4.2, os parâmetros das componentes gaussianas (\mathbf{m}, R) têm de ser inicializados. Para tal utiliza-se normalmente o algoritmo *k-means* [Duda et al., 2000].

Existem muitas variantes para este algoritmo. A utilizada nesta dissertação faz parte integrante da *toolbox Netlab*.

Vão ser descritos em seguida todos os passos do algoritmo EM (Algoritmo 4.2):

Sendo conhecido o número de classes do modelo v

Seja $X = \{x_1, \dots, x_T\}$ um conjunto de vectores de características de treino com dimensão T . Sabe-se a que classe pertence cada um dos elementos deste conjunto.

Sendo conhecido o número de iterações c a considerar para o algoritmo EM

- Inicialização dos parâmetros do modelo, \mathbf{m}, R, w , para cada classe
 - As médias de cada classe \mathbf{m} , são calculadas normalmente através do algoritmo *k-means*.
 - É definido o tipo de matriz de covariância do modelo (esférica, diagonal, ...). É inicializada.
 - Os pesos w são inicializados. Normalmente valores iguais para classes igualmente representadas.
- A função de verosimilhança dos GMM pode ser representada da seguinte forma (4.6):

$$p(X \setminus \mathbf{q}) = \prod_{t=1}^T p(x_t \setminus \mathbf{q}) \quad (4.6)$$

O que se pretende é partindo do conjunto de parâmetros \mathbf{q} estimar $\tilde{\mathbf{q}}$ tal que

$$p(X \setminus \tilde{\mathbf{q}}) \geq p(X \setminus \mathbf{q})$$

- Processo de treino

Algoritmo EM

- Passo E:

Calcular as probabilidades a posteriori do exemplo actual em cada classe, utilizando para tal os parâmetros do modelo calculados na iteração anterior. O que se pretende é determinar o peso que cada componente do modelo tem em cada um dos exemplos de treino.

As probabilidades para a mistura i são representadas por (4.7):

$$p(i \setminus x_t, \mathbf{q}) = \frac{w_i p_i(x)}{\sum_{k=1}^v w_k p_k(x)} \quad (4.7)$$

- Passo M:

Ajustar os parâmetros \mathbf{m}, R, w de cada classe de forma a maximizar a função de verosimilhança. Como se vê nas fórmulas, os parâmetros são ajustados considerando os pesos que cada componente gaussiana tem em cada um dos exemplos de treino.

$\tilde{\mathbf{m}}, \tilde{R}, \tilde{w}$ correspondem aos valores \mathbf{m}, R, w actualizados e que servirão de ponto de partida para o cálculo do passo E da iteração seguinte (4.8), (4.9), (4.10).

$$\tilde{w}_i = \frac{1}{T} \sum_{t=1}^T p(i \setminus x_t, \mathbf{q}) \quad (4.8)$$

$$\tilde{\mathbf{m}}_i = \frac{\sum_{t=1}^T p(i \setminus x_t, \mathbf{q}) x_t}{\sum_{t=1}^T p(i \setminus x_t, \mathbf{q})} \quad (4.9)$$

$$\tilde{R}_i = \frac{\sum_{t=1}^T p(i \setminus x_t, \mathbf{q}) x_t x_t^T}{\sum_{t=1}^T p(i \setminus x_t, \mathbf{q})} - \tilde{\mathbf{m}}_i \tilde{\mathbf{m}}_i^T \quad (4.10)$$

Os passos EM repetem-se até que as diferenças do logaritmo da função de verosimilhança entre duas iterações consecutivas sejam suficientemente pequenas.

O algoritmo pode ainda parar, se o número de iterações definido inicialmente, for atingido. É preferível como é obvio o primeiro critério de paragem.

Algoritmo 4.2. Processo de treino dos GMM

É preciso ter atenção à inicialização dos parâmetros do modelo neste processo de treino. De facto diferentes inicializações produzirão resultados diferentes. É necessário portanto utilizar a repetição para escolher a melhor configuração.

Mais detalhes sobre o algoritmo EM podem ser consultadas em [Dempster et al., 1977].

4.3.3. Processo de Validação

Nesta fase, são calculadas muito simplesmente as probabilidades de cada classe em cada vector de características de validação. A classificação para cada vector de validação corresponderá à classe cuja probabilidade seja maior.

4.4. REDES NEURONAIS ARTIFICIAIS

As redes neuronais artificiais foram criadas com o intuito de emular o funcionamento do cérebro humano. Pretende-se, à imagem do cérebro, que as redes neuronais artificiais tenham capacidade de aprendizagem, de adaptação e de generalização. Não obstante o ainda pouco conhecimento sobre determinados mecanismos que o cérebro utiliza e as limitações dos próprios computadores, sabe-se que para um largo espectro de problemas, as redes neuronais artificiais conseguem aproximar uma solução com resultados muito satisfatórios.

São aplicadas em áreas tão diversas como reconhecimento de voz, robótica, investigação médica, telecomunicações, marketing, análise de investimentos, reconhecimento automático de géneros musicais, jogos, etc.

Não obstante o termo redes neuronais (RN) por princípio se referir às redes neuronais biológicas do nosso cérebro, daqui para a frente neste trabalho e por uma questão de simplificação, os termos redes neuronais e redes neuronais artificiais serão utilizados indiferentemente para referir redes neuronais no âmbito do computador. As redes neuronais biológicas serão referenciadas explicitamente por esse termo.

As redes neuronais podem ser classificadas quanto à sua arquitectura e quanto ao seu tipo. Essa classificação é escolhida e utilizada de acordo com o problema que se pretende resolver.

Os vários tipos de classificação e arquitecturas referidas no parágrafo anterior serão abordados na secção seguinte.

4.4.1. Topologias

No âmbito das redes neuronais artificiais, existem duas arquitecturas distintas. Estas estão relacionadas com a forma como as ligações entre os neurónios são efectuadas. As redes neuronais, mediante a sua arquitectura ou topologia podem ser classificadas em *redes com ligações para a frente*²⁷ (FFNN) ou *redes recorrentes*.

Nas redes com ligações para a frente, os neurónios recebem informações de todos os neurónios da camada anterior e enviam para todos os neurónios da camada seguinte, desde a camada de entrada até à camada de saída (Figura 4.7).

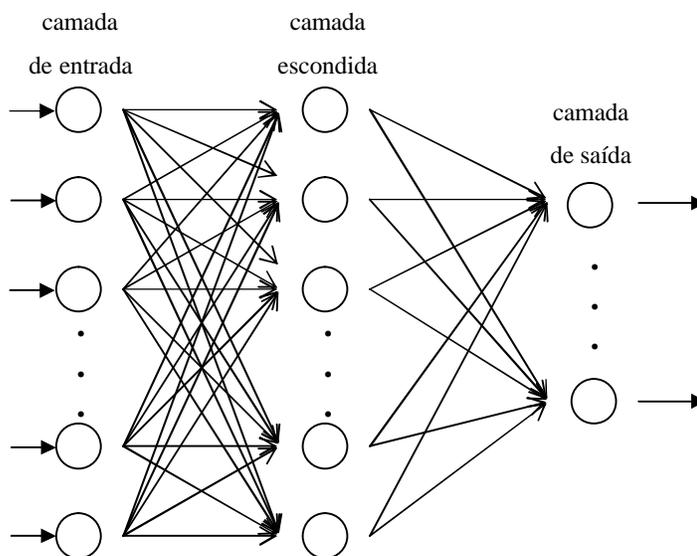


Figura 4.7. Rede neuronal com ligações para a frente

Como exemplos de redes deste tipo, encontramos as redes MLP²⁸ (Perceptrão Multicamada) [Rumelhart & McClelland, 1986], as redes RBF (redes com função de base radial) [Broomhead & Lowe, 1988] e as redes LVQ²⁹ [Kohonen, 1989].

²⁷ em terminologia inglesa: Feedforward Neural Networks

²⁸ em terminologia inglesa: Multilayer Perceptron

Nas redes recorrentes (Figura 4.8), os neurónios têm ligações bidireccionais i.e. recebem e enviam informação dos neurónios da camada seguinte ou anterior ou da própria camada. Outra característica associada a este tipo de redes é a de terem memória dinâmica, isto é, as saídas são influenciadas não só pelas entradas dessa iteração, mas também pelas entradas e saídas passadas.

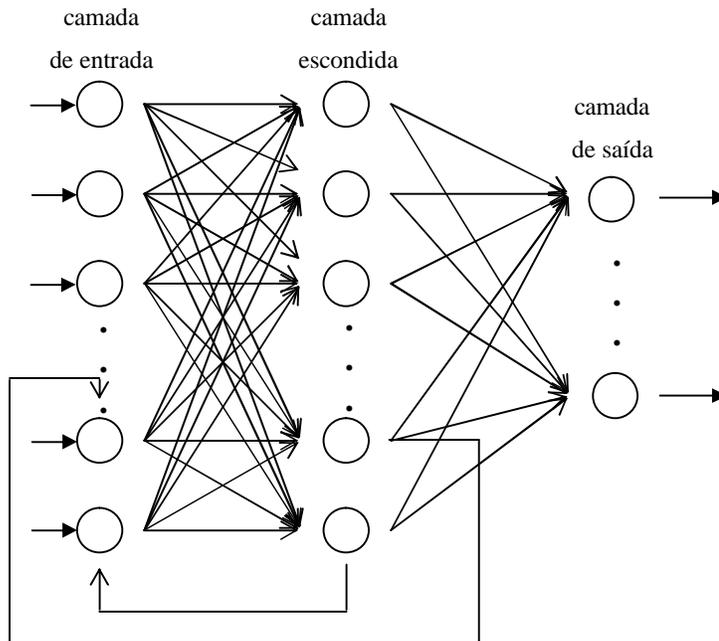


Figura 4.8. Rede neuronal recorrente

Como exemplos de redes recorrentes temos as redes de Elman [Elman, 1990] e as redes de Hopfield [Hopfield, 1982].

Deste ponto em diante, irá ser dada especial importância às redes com ligações para a frente, especificamente às redes Perceptrão Multicamada, já que o nosso trabalho apenas incidiu sobre esse tipo de redes. Assim, todas as considerações futuras, serão referentes a este tipo de redes, apesar de algumas delas poderem ser válidas para outras redes com ligações para a frente ou para as redes recorrentes.

²⁹ em terminologia inglesa: Learning Vector Quantization

4.4.2. Componentes e Processamento de Dados

Como se verificou anteriormente, as redes neuronais nasceram do pressuposto que seria possível em determinados problemas “imitar” o funcionamento do nosso cérebro, utilizando para tal estruturas de dados e processos que fossem uma analogia das estruturas e do funcionamento do cérebro humano (e.g. neurónios, sinapses).

Ao longo deste trabalho, alguns destes termos serão utilizados no âmbito das redes neuronais artificiais, como por exemplo, o termo neurónio³⁰.

A estrutura fundamental numa rede neuronal é portanto o neurónio. Cada neurónio é estimulado ou seja, recebe sinais dos neurónios vizinhos (ver secção anterior), enviando sinais após processamento, para outros neurónios. Este processo de comunicação entre os neurónios assemelha-se muito ao que se passa com os neurónios biológicos e as sinapses.

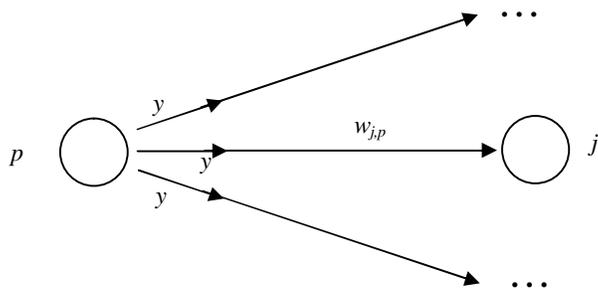


Figura 4.9. Relação entre dois neurónios p e j .

Através da figura 4.9, pode-se visualizar que a relação entre dois neurónios p e j depende do sinal enviado pelo neurónio p (y) para os neurónios vizinhos e do peso da ligação sináptica entre p e j (w_{jp}).

³⁰ neurónio ou simplesmente nó na terminologia das redes neuronais artificiais.

No cérebro humano, o raciocínio assenta na organização dos neurónios biológicos, no seu número e nas ligações sinápticas. O mesmo se passa nas redes neuronais artificiais.

A capacidade de uma rede assenta em grande medida na quantidade de neurónios presentes que deve ser em número ideal, na sua organização dentro da rede e na relação entre eles.

A camada (Figura 4.7) é outra estrutura importante de uma rede neuronal artificial. Como se verificou na secção anterior, uma determinada camada, engloba sempre um conjunto de neurónios com a característica comum de se situarem ao mesmo nível na estrutura da rede.

Pode-se visualizar na figura 4.10, para um dado neurónio de uma rede neuronal, o processo de produção de um sinal a partir de sinais enviados pelos neurónios vizinhos e dos pesos que influenciam esse neurónio.

Como se verifica na figura, uma rede com ligações para a frente é constituída tipicamente por uma camada de entrada, que corresponde aos dados que entram na rede, uma ou mais camadas escondidas cujos neurónios recebem os dados produzidos pelos neurónios da camada de entrada ou das camadas escondidas anteriores, e uma camada de saída, cujos neurónios recebem dados da camada escondida anterior e que correspondem à saída da rede.

Pode-se ter um número variável de camadas escondidas, de acordo com o problema em questão. De qualquer forma, será concluído mais à frente (secção 5.4.1) que com as configurações utilizadas ao longo deste trabalho, basta uma camada escondida para aproximar qualquer função com uma precisão arbitrária.

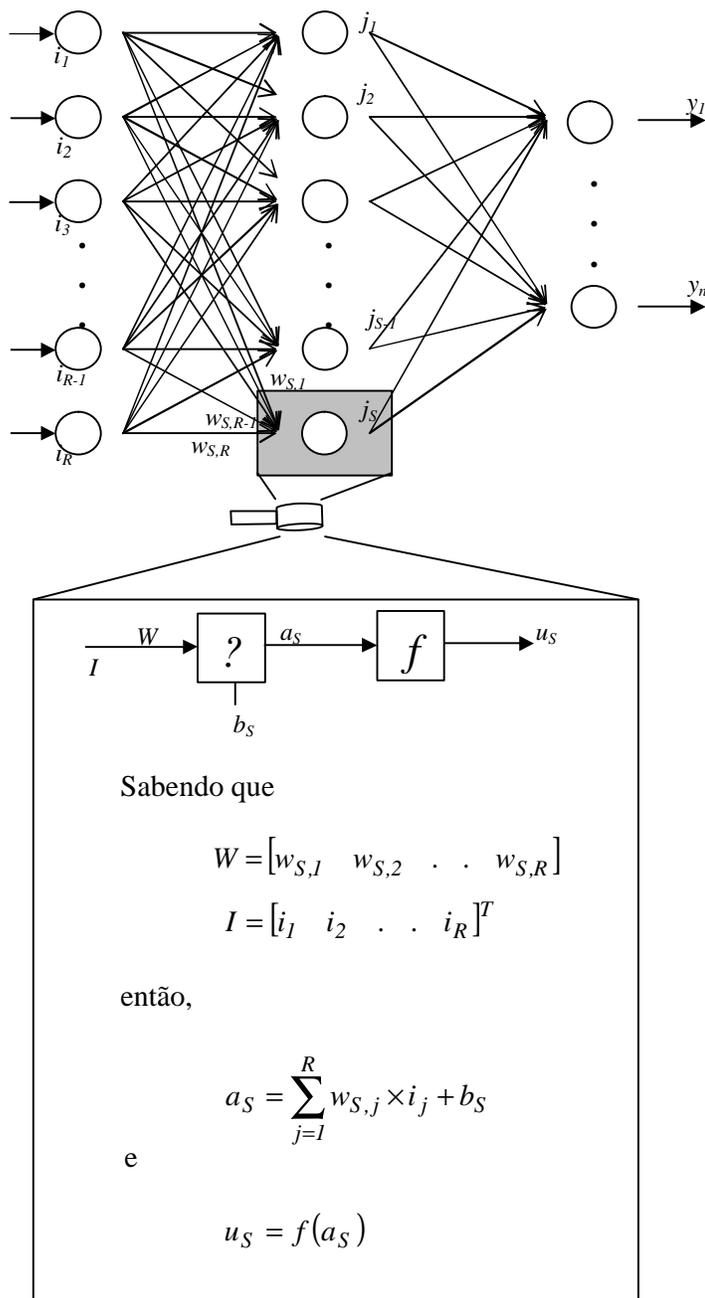


Figura 4.10. Rede MLP. Processamento de informação efectuado por cada neurónio

Os resultados de saída da rede neuronal, dependem dos dados de entrada, dos valores iniciais dos parâmetros da rede e da relação entre os próprios neurónios. Essa

relação, como se visualiza por exemplo na figura anterior para o s -ésimo neurónio da camada escondida é representada pelo produto da matriz de pesos que incide nesse neurónio, W , (e.g. o elemento $w_{S,R}$ da matriz, corresponde ao sinal propagado pelo neurónio R da camada de entrada para o neurónio S da camada escondida) pelos valores de entrada na rede, I , ao qual se adiciona normalmente um viés³¹ (b_S) associado a cada neurónio. A esse resultado finalmente aplica-se uma função de activação (f_s) de acordo com o problema em questão. Da aplicação da função de activação resultará um valor que será propagado para os neurónios da camada seguinte.

Existem várias funções de activação que são usadas conforme o tipo de rede neuronal que se está a utilizar, o intervalo de compreensão dos resultados que se pretende e obviamente do problema em questão. As mais conhecidas são a função linear e a sigmoideal. Esta última será utilizada no âmbito deste trabalho com as redes MLP (mais concretamente é utilizada a função do *Matlab*, *logsig*) e as suas características principais são ter o domínio \Re e, como contradomínio, o intervalo $[0,1]$. É ainda diferenciável, o que permite a sua utilização em redes cujo treino utilize a técnica de retropropagação³² do erro. Esta função de activação é representada pela seguinte equação (4.11):

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4.11)$$

4.4.3. Treino

O grande objectivo de uma rede neuronal é validar correctamente os dados de entrada, i.e. produzir as saídas mais adequadas para os exemplos que são introduzidos na rede. Para tal é necessário que todos os parâmetros da rede sejam devidamente configurados. Entre esses parâmetros incluem-se os pesos sinápticos e os parâmetros

³¹ Em terminologia Inglesa: bias - Termo de polarização. É opcional

³² Em terminologia Inglesa: Backpropagation

opcionais de viés. Para ajustar esses pesos para que a rede produza os melhores resultados de validação possíveis é necessário que a rede seja devidamente treinada.

O treino pode ser supervisionado ou não supervisionado.

No treino não supervisionado, não utilizado no nosso estudo de redes neuronais, os valores de saída são calculados autonomamente pela própria rede em resposta aos exemplos de entrada, i.e., espera-se que a rede encontre padrões semelhantes nos dados de entrada e os segmente de acordo com isso. Os parâmetros da rede são alterados de acordo com essa segmentação.

O treino supervisionado segue o paradigma “com tutor”, i.e., o ajuste dos parâmetros é efectuado com base num conjunto de exemplos de treino previamente catalogados. No processo de treino, a rede irá ajustar os seus parâmetros de forma a que, no final, os dados de entrada sejam correctamente mapeados nos dados de saída. Portanto este tipo de treino consiste em introduzir na rede exemplos de pares entrada/saída, nos quais a saída deve corresponder ao valor esperado para a entrada. Concretizando, para o nosso contexto de classificação musical, um par poderia ser a coluna t da matriz de entrada $*$ (Figura 4.11) que corresponde às 40 características extraídas da t -ésima peça musical de treino e a coluna t da matriz de saída $**$ (Figura 4.11) correspondente à saída pretendida para essa mesma peça. Se a peça t for de piano e tivéssemos definido o neurónio y_2 para representar as músicas dessa classe, então a coluna t de $**$ teria 1 na posição $(2,t)$ e zero nas restantes.

Esta figura representa precisamente a aplicação das redes neuronais ao primeiro problema de classificação deste trabalho (secção 4.1.1).

O treino pode ainda ser de dois tipos, quanto à sua forma de decurso: treino incremental ou adaptativo, no qual os parâmetros da rede são actualizados após a entrada na mesma de cada um dos exemplos de treino ou treino *batch* no qual os parâmetros da rede só são actualizados após todos os exemplos de treino serem processados pela rede.

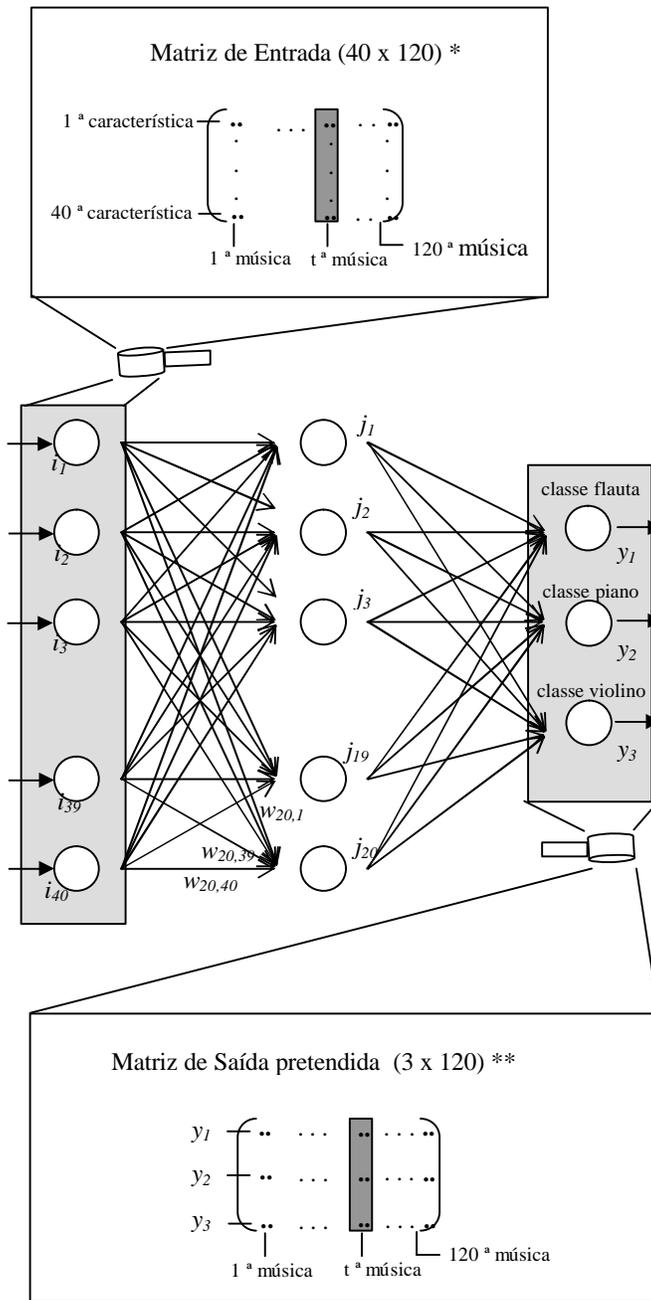


Figura 4.11. Rede MLP. Treino supervisionado

Neste estudo foi usado sempre o segundo tipo de treino. Como tal os pesos e os viés são actualizados no final de cada iteração, i.e., após a rede ter calculado a saída correspondente aos vectores de características de todas as músicas de entrada e portanto ter uma visão global de todo o problema para tentar minimizar da melhor forma a

função de desempenho ou erro. É óbvio que neste caso, a ordem de entrada das músicas na rede é irrelevante, ao contrário do treino incremental.

O modo de treino *batch* realiza-se por um número de iterações ou épocas definidas inicialmente. Como se viu anteriormente, os valores de saída da rede (e.g., y_1, y_2, y_3) são calculados em cada iteração, de acordo com os valores I de entrada na rede e com os parâmetros da rede (W e b). Em seguida, os valores de saída da rede são comparados com os valores de saída fornecidos nos exemplos. Através dessa comparação, é calculada uma função de erro ou função de desempenho e em seguida os parâmetros da rede (W e b) são ajustados de forma a minimizar essa função. Esses ajustamentos são efectuados primeiro em relação aos parâmetros da camada anterior à de saída e assim por diante até à camada de entrada. Esta técnica de ajustar iterativamente os parâmetros da rede, de camada em camada da frente para trás, na direcção da descida da negativa do gradiente da função de desempenho, chama-se retropropagação do erro [Rumelhart & McClelland 86]. Depois de ajustados todos os parâmetros, termina essa iteração e poderá começar a seguinte repetindo-se todo o processo.

O método da descida do gradiente, desde o cálculo do erro em cada iteração até ao ajuste dos pesos vai ser descrito através das equações seguintes. Assim, a função erro a minimizar expressa-se pela equação (4.12), onde o erro total E é igual à soma quadrática dos erros $E^{(s)}$ (4.13) determinados para cada um dos exemplos de treino.

$$E = \sum_{s=1}^N E^{(s)} \quad (4.12)$$

$$E^{(s)} = \frac{1}{2} \sum_{p=1}^n \left(y_p^{(s)} - \hat{y}_p^{(s)} \right)^2 \quad (4.13)$$

onde $y_p^{(s)} \in \mathfrak{R}$ e $\hat{y}_p^{(s)} \in \mathfrak{R}$ representam respectivamente a activação e a saída desejada para os neurónios de saída p relativamente aos exemplos de treino s . N e n designam respectivamente o número de exemplos de treino fornecidos à rede e o número de neurónios da camada de saída.

Em seguida os pesos são alterados progressivamente no sentido da diminuição do erro quadrático (4.14):

$$Dw_{pj} = -g \frac{\partial E^{(s)}}{\partial w_{pj}} \quad (4.14)$$

onde g representa a velocidade de aprendizagem, definida inicialmente na configuração da rede. Este parâmetro determina o passo da rede na descida da superfície de erro.

O processo de treino continua até, se possível, o erro estar abaixo de um limiar pré-definido, o número de iterações ter atingido um máximo definido na configuração ou ainda ser atingido um tempo limite também pré-definido. O critério ideal de paragem é obviamente o primeiro, já que é o único que garante que os parâmetros da rede determinados no final reflectem uma aprendizagem da rede em relação a cada uma das classes de classificação, i.e. os parâmetros calculados fazem com que a rede aplicada aos exemplos de treino resulte em valores de saída bastante próximos dos valores correctos fornecidos também nos exemplos.

Existem variantes ao método da descida do gradiente, descrito atrás, baseadas em técnicas de optimização, que fazem aumentar a velocidade de convergência da rede. Uma dessas variantes chama-se método do gradiente conjugado [Demuth & Beale, 2001].

Em muitos casos nenhum dos métodos anteriores é eficiente. Neste trabalho foram feitas algumas experiências, utilizando como algoritmo de treino, o método da descida do gradiente nas suas várias variantes. Os resultados não foram considerados aceitáveis, com certeza também devido à natureza dos próprios dados. A convergência foi sempre muito lenta e nem mesmo definindo números de iterações muito elevadas, ou manipulando a variável que controla a velocidade de aprendizagem, se conseguiu minimizar a função de erro para valores considerados aceitáveis.

Assim sendo, foi utilizado o algoritmo de treino Levenberg-Marquardt [Hagan & Menhaj, 1994]. Este algoritmo é adequado para treinar redes que tenham até algumas centenas de pesos e é 10 a 100 mais rápido [Demuth & Beale, 2001] que os anteriores. Em alguns casos consegue mesmo convergir em situações em que os anteriores não conseguem [Hagan & Menhaj, 1994].

Este algoritmo revelou-se adequado para os problemas que se pretendia resolver, já que os resultados de convergência foram satisfatórios, como se poderá verificar no capítulo 5.

Será apresentado de seguida o modo de funcionamento do algoritmo.

Algoritmo de Levenberg-Marquardt

Este algoritmo é uma aproximação ao método de Newton e, como ele, foi pensado no sentido de ser uma alternativa aos métodos do gradiente conjugado e de permitir convergências mais rápidas.

O algoritmo de Levenberg-Marquardt difere do método de Newton por não ser necessário cálculo directo da matriz Hessiana (segundas derivadas da função de desempenho em relação aos parâmetros da rede), que é computacionalmente muito difícil de calcular para redes FFNN. Pelo contrário, esta pode ser aproximada, caso a função de erro seja a soma dos quadrados, pelo cálculo da matriz Jacobiana (derivadas de 1ª ordem da função de erro em relação aos parâmetros da rede) e pelo produto pela sua transposta da seguinte forma, (4.15):

$$H(x) = J^T(x)J(x) \quad (4.15)$$

em que H representa a matriz Hessiana, J a matriz Jacobiana e $x = \{x_1, \dots, x_m\}$ representa o vector de parâmetros da rede.

Já o gradiente pode ser expresso da seguinte forma, (4.16):

$$g(x) = J^T(x)E(x) \quad (4.16)$$

em que g é o gradiente, e E é o vector erro calculado entre o valor desejado e o valor de facto calculado para a saída.

O algoritmo de Levenberg-Marquardt é iterativo, tal como o método da descida do gradiente, e é igual a este para cada iteração, até à altura em que é calculado o erro (soma dos quadrados dos erros) dos N exemplos de treino, entre a saída efectiva da rede e a saída desejada.

No passo seguinte é calculada a matriz Jacobiana, (4.17):

$$J(x) = \begin{bmatrix} \frac{\partial E_1(x)}{\partial x_1} & \frac{\partial E_1(x)}{\partial x_2} & \dots & \frac{\partial E_1(x)}{\partial x_m} \\ \frac{\partial E_2(x)}{\partial x_1} & \frac{\partial E_2(x)}{\partial x_2} & \dots & \frac{\partial E_2(x)}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial E_N(x)}{\partial x_1} & \frac{\partial E_N(x)}{\partial x_2} & \dots & \frac{\partial E_N(x)}{\partial x_m} \end{bmatrix} \quad (4.17)$$

É calculada em seguida, a seguinte fórmula com base no método de Newton, (4.18):

$$Dx = -[J^T(x)J(x) + \mu D]^{-1} J^T(x)E(x) \quad (4.18)$$

Desta forma, obtêm-se Dx , o incremento dos parâmetros da rede entre a iteração actual e a iteração seguinte. Na equação anterior, D é a matriz identidade e μ é um escalar inicializado (no início do processo de treino) normalmente com um valor da ordem dos 0,001.

Após o ajuste dos pesos (somar x a Dx), é necessário recalculer a função de erro. Há agora duas situações possíveis:

Na primeira, se esta última soma dos quadrados dos erros for mais pequena que a calculada para a iteração anterior, então, deve-se actualizar μ multiplicando-o por um factor de decremento (parâmetro definido no início do treino - normalmente 0,1). Em seguida, inicia-se uma nova iteração com os pesos e viés já actualizados.

Na segunda, se esta soma dos quadrados dos erros for maior que a calculada na iteração anterior, então deve-se actualizar μ multiplicando-o por um factor de incremento (parâmetro definido no início do treino - normalmente 10) e voltar a calcular Dx pela equação (4.8). Este processo repete-se até estarmos na situação do primeiro caso.

O treino pára, analogamente ao método da descida do gradiente, quando o algoritmo converge, i.e., a função de desempenho fica abaixo de um limiar fornecido. Pode parar ainda se μ atingir um valor máximo definido inicialmente, ou forem

atingidos limites no treino a nível de tempo ou iterações (parâmetros definidos inicialmente).

De notar que todos os valores de parâmetros referidos por omissão, correspondem aos valores por omissão para esses parâmetros da *toolbox* de redes neuronais do *Matlab*.

A grande limitação do algoritmo de Levenberg-Marquardt é o facto de guardar muita informação em memória durante a execução (principalmente a matriz Jacobiana em cada iteração). Como tal, para redes de maior dimensão pode-se ter, através da modificação de um parâmetro inicial, de optar por reduzir o número de linhas da Jacobiana que será guardado em memória a cada momento. Obviamente que isto provoca uma perda de desempenho da rede. Por tudo isso, existe um tamanho para a rede a partir do qual as vantagens do método deixam de existir em termos de eficiência computacional e em termos de resultado. Em situações em que a rede seja realmente muito grande, poderemos ter forçosamente de optar por outro método.

Uma análise mais detalhada deste algoritmo é apresentada em [Hagan & Menhaj, 1994] e [Demuth & Beale, 2001].

4.4.4. Validação

Só é possível avaliar o processo de treino, após aplicar à rede os exemplos de validação e verificar os resultados. Os pesos e viés iniciais são determinantes para os resultados obtidos no processo treino/validação, por isso é fundamental repetir várias vezes esse processo, com inicializações diferentes dos pesos, para escolher o melhor sistema.

Capítulo 5

RESULTADOS EXPERIMENTAIS

Neste capítulo serão apresentados os resultados obtidos na fase de experimentação para cada um dos três classificadores utilizados: KNN, GMM e MLP.

Irá ser efectuada uma comparação entre os classificadores, através dos resultados alcançados por estes para cada um dos três problemas de classificação.

Finalmente irá ser feita uma validação centrada na música que pretende ser uma aproximação a um sistema real de classificação automática de géneros musicais. Esta validação será feita, utilizando apenas o classificador com melhores resultados, para cada uma das tarefas de classificação.

5.1. PRESSUPOSTOS INICIAIS

Como foi referido no capítulo anterior, o objectivo deste estudo é o de classificar extractos de música em géneros musicais, utilizando para tal vários classificadores: KNN, GMM e MLP.

Todos eles utilizaram as mesmas premissas iniciais. Assim para cada um dos classificadores:

- O conjunto de extractos musicais utilizados na sua totalidade é o mesmo. Existem 300 peças de música, 60 de cada um dos 5 géneros musicais em estudo: flauta, piano, violino, coral e ópera.
- O conjunto de problemas em análise é o mesmo. O primeiro problema consiste em classificar música em três géneros musicais: música para flauta, música para piano e música para violino. O segundo em classificar música, em música coral e ópera. Finalmente, no terceiro problema pretende-se classificar música num dos cinco géneros musicais anteriores.
- A assinatura de cada música é igual, i.e., o conjunto de características extraídas de cada música é o mesmo, independentemente do problema de classificação em questão. Foram extraídas 40 características de cada peça de música.
- As peças musicais utilizadas são divididas em dois conjuntos. No primeiro conjunto, chamado de treino, são incluídas 2/3 das peças e no segundo conjunto, chamado de validação ou de teste, são incluídas 1/3 das peças. Cada um dos conjuntos contém um número igual de peças musicais por género. Exemplificando para o primeiro problema de classificação, existem no total 180 peças musicais, 60 de cada um dos 3 géneros musicais em análise. Logo o conjunto de treino inclui 120 peças, 40 de cada género e o conjunto de validação inclui 60 peças, 20 de cada género.

A razão da igualdade de critérios para os classificadores está relacionada com o facto de se pretender uma comparação entre os mesmos. Para tal eles devem, por uma questão de coerência, partir em igualdade de circunstâncias.

O classificador que obtiver melhores resultados será utilizado para um teste final (criação de uma aproximação a um sistema de classificação automática) que irá validar 1000 extractos musicais.

O sistema (hardware e software) utilizado em toda a fase de experimentação, desde a captação dos extractos musicais até à classificação propriamente dita foi o mesmo.

Assim, a nível de hardware, foi utilizado um computador portátil *HP Omnibook xe4500* com processador *Intel Pentium 4 Mobile* ® 1,6Ghz e 256 Mb de memória *RAM*. A nível de software foi utilizado: Sistema operativo *Windows XP Professional* ® e programas *Matlab* ® e *CoolEditPro* ®.

5.2. K-VIZINHOS MAIS PRÓXIMOS

Vão ser expostos em seguida alguns aspectos sobre a forma como o KNN foi utilizado neste estudo. Serão descritos em seguida, os resultados alcançados para as três tarefas de classificação e as conclusões.

5.2.1. Variantes do Algoritmo e Regras de Classificação

Foram efectuadas três configurações do algoritmo KNN para o estudo dos três problemas de classificação. Essas configurações estão relacionadas com o número de vizinhos K considerados. Assim, estudou-se cada um dos problemas com K igual a 1, a 3 e a 5. Estes valores para K são típicos neste tipo de problemas.

O algoritmo diz que uma dada música terá uma determinada classificação num género musical, de acordo com o género mais representado nos K vizinhos mais próximos. Exemplificando para $K=3$, supondo que se considera a classificação de extractos de música em dois estilos: coral e ópera. Para uma determinada música de teste, as 3 músicas de treino mais próximas são, partindo da mais próxima, coral, ópera, ópera, então a música em questão seria classificada como ópera.

Foi estabelecida uma regra de classificação, além da existente no próprio algoritmo. Essa regra pode eventualmente ser aplicada nos problemas de classificação em que haja mais do que 2 classes, ou seja no 1º problema de classificação (3 classes) e no 3º problema de classificação (5 classes). O algoritmo classifica uma música na classe

mais representada entre os K vizinhos mais próximos, mas pode acontecer duas classes ou mais serem igualmente representadas. Nesse caso a música será classificada entre as músicas das classes mais representadas, de acordo com a mais próxima. Exemplificando para o 3º problema de classificação com $K=5$, supondo que para uma peça musical de teste em análise, as músicas “mais próximas” pertencem às classes flauta, coral, piano, coral e piano (ordenadas a partir da mais próxima), então essa peça será classificada na classe coral.

Vão ser analisados em seguida os três problemas de classificação. Essa análise é feita, para cada um, através das matrizes de confusão obtidas, considerando $K=\{1,3,5\}$.

5.2.2. Primeira Classificação: Três Gêneros Musicais

Este problema em estudo pretende classificar música em três classes: flauta, piano e violino. Partindo de um total de 180 peças musicais, consideraram-se 2/3 (120 - 40 de cada) como padrões de treino e consideraram-se 1/3 (60 - 20 de cada) como músicas de validação.

Para $K=1$, (Tabela 5.1).

KNN(1) 68,3%	flauta	piano	violino
flauta	65	5	30
piano	20	80	10
violino	15	15	60

Tabela 5.1. Matriz de confusão da música instrumental: KNN(1).

A matriz de confusão mostra, considerando $K=1$, que foram bem classificadas 65% das músicas de flauta, 80% das músicas de piano e 60% das músicas de violino. No total este classificador obteve 68,3% de taxa de sucesso.

Para $K=3$, (Tabela 5.2).

KNN(3) 80%	flauta	piano	violino
flauta	80	10	10
piano	15	80	10
violino	5	10	80

Tabela 5.2. Matriz de confusão da música instrumental: KNN(3).

Através da matriz anterior pode-se verificar que a taxa de sucesso na classificação foi de 80%. Esta percentagem foi obtida através da média das taxas de sucesso por instrumento: 80% para flauta, piano e violino.

Para $K=5$, (Tabela 5.3).

KNN(5) 80%	flauta	piano	violino
flauta	80	15	15
piano	15	80	5
violino	5	5	80

Tabela 5.3. Matriz de confusão da música instrumental: KNN(5).

Considerando K igual a 5, a taxa global e as taxas parcelares de sucesso foram iguais às obtidas com K igual a 3, ou seja, 80%.

5.2.3. Segunda Classificação: Dois Géneros Musicais

Neste segundo problema, o objectivo é catalogar músicas numa das seguintes classes: música coral e ópera. Foram utilizadas no total 120 músicas, 60 de música coral e 60 de ópera. De cada um desses conjuntos de 60 músicas, foram utilizadas 2/3 como padrões de treino (40 de cada classe) e 1/3 como músicas de validação (20 de cada classe).

Para $K=1$, (Tabela 5.4).

KNN(1) 82,5%	coral	ópera
coral	85	20
ópera	15	80

Tabela 5.4. Matriz de confusão da música vocal: KNN(1).

Foi obtida uma taxa global de músicas correctamente classificadas de 82,5%, que corresponde a 85% de músicas corais correctamente classificadas e 80% de músicas de ópera igualmente bem classificadas.

Para $K=3$, (Tabela 5.5).

KNN(3) 85%	coral	ópera
coral	100	30
ópera	0	70

Tabela 5.5. Matriz de confusão da música vocal: KNN(3).

Neste caso a taxa de músicas bem classificadas foi de 85% em que todas as músicas corais foram bem classificadas e apenas 70% das músicas de ópera o foram.

Para $K=5$, (Tabela 5.6).

KNN(5) 82,5%	coral	ópera
coral	100	35
ópera	0	65

Tabela 5.6. Matriz de confusão da música vocal: KNN(5).

Considerando os 5 vizinhos mais próximos de cada música de teste em análise, a taxa de sucesso foi de 100% para a música coral e de 65% para a música de ópera. A média dos dois valores, 82,5%, corresponde à taxa de sucesso global do classificador.

O melhor resultado foi obtido com $K=3$ (85%).

5.2.4. Terceira Classificação: Cinco Géneros Musicais

Neste problema pretende-se classificar música em cinco subgéneros da música clássica; música para flauta, música para piano, música para violino, música coral e música de ópera.

Para o estudo foram utilizadas um total de 300 peças musicais (60 de cada um dos cinco géneros musicais). Destas, $2/3$ (200, 40 de cada) foram, como nos problemas anteriores, utilizadas como padrões de treino e $1/3$ (100, 20 de cada) utilizadas para validação.

Seguem-se as matrizes de confusão para as três variantes do algoritmo.

Para $K=1$, (Tabela 5.7).

KNN(1) 59%	flauta	piano	violino	coral	ópera
flauta	65	5	25	25	5
piano	20	70	0	20	5
violino	15	15	40	0	20
coral	0	5	20	50	0
ópera	0	5	15	5	70

Tabela 5.7. Matriz de confusão da música vocal e instrumental: KNN(1).

Como se pode verificar pela matriz de confusão anterior, a taxa de peças musicais bem classificadas para os géneros flauta, piano, violino, coral e ópera foi respectivamente de 65%, 70%, 40%, 50% e 70%. A taxa global de sucesso foi de apenas 59%.

Para $K=3$, (Tabela 5.8).

KNN(3) 61%	flauta	piano	violino	coral	ópera
flauta	70	10	25	20	5
piano	20	70	0	20	5
violino	10	15	40	0	25
coral	0	0	15	60	0
ópera	0	5	20	0	65

Tabela 5.8. Matriz de confusão da música vocal e instrumental: KNN(3).

Considerando três vizinhos mais próximos da música a classificar, as percentagens de músicas bem classificadas foram de 70% para flauta e piano, 40% para violino, 60% para música coral e 65% para ópera. No geral foi atingida uma taxa de 61%.

Para $K=5$, (Tabela 5.9).

KNN(5) 67%	flauta	piano	violino	coral	ópera
flauta	80	10	10	20	10
piano	15	75	0	15	0
violino	5	5	50	0	10
coral	0	5	20	65	15
ópera	0	5	20	0	65

Tabela 5.9. Matriz de confusão da música vocal e instrumental: KNN(5).

Considerando $K=5$, foram conseguidos os melhores resultados para este problema de classificação. Assim, a taxa geral de sucesso foi de 67%. Por géneros, as percentagens de músicas bem classificadas foram de 80% para flauta, 75% para piano, 50% para violino, 65% para música coral e para ópera.

Os resultados da aplicação do KNN a este problema de classificação são insatisfatórios, principalmente para a classe violino. De notar que a “confusão” de classes acontece mais para os casos em que as duas classes são uma vocal e uma

instrumental. Por exemplo 20% das músicas de violino foram classificadas em coral e outros 20% em ópera. 20% das de coral foram classificadas em flauta mas também acontece entre classes vocais e instrumentais. Por exemplo 15% das músicas de flauta foram classificadas em piano.

Apesar de tudo, ficou claro que aumentando neste problema o número de vizinhos, melhoraram os resultados, com certeza devido à complexidade do problema em que os géneros “se tocam”.

A descida clara na percentagem de sucesso deste problema de classificação leva a concluir que provavelmente as características escolhidas são boas para identificar géneros similares vocais e instrumentais, mas não são boas para discriminar entre voz e instrumentos.

5.3. MODELOS DE MISTURAS GAUSSIANAS

Vão ser explanados nesta secção alguns aspectos sobre a forma como o GMM foi utilizado neste estudo. Serão descritos ainda, os resultados alcançados para as três tarefas de classificação e as respectivas conclusões.

5.3.1. Princípios Gerais

São expostos em seguida alguns aspectos referentes à estrutura dos classificadores utilizados, bem como às ferramentas utilizadas neste estudo.

- O número de *clusters* para o algoritmo é escolhido de acordo com o número de classes de cada uma das tarefas de classificação.
- Os centros de cada *cluster* têm que ser inicializados. Essas inicializações são feitas através do algoritmo *k-means clustering* (função do *Netlab*) com 10 iterações.

- A matriz de covariância associada a cada *cluster* pode ser de vários tipos: esférica, diagonal e análise probabilística dos componentes principais³³ (PPCA). Para cada problema de classificação é escolhida a que obtém melhores resultados.
- A ferramenta que se utiliza durante todo o estudo com este classificador, é o *Matlab* ®, mais concretamente, a *toolbox* externa *Netlab* (<http://www.ncrg.aston.ac.uk/netlab/>).
- As matrizes de covariância são inicializadas, independentemente da tarefa de classificação, com todos os seus valores iguais a 1. Esta inicialização corresponde aos valores por omissão do *Netlab*.
- O número de iterações escolhido é suficientemente elevado para garantir que esse nunca seja o critério de paragem do algoritmo.

A configuração do classificador escolhida, para cada uma das três tarefas de classificação, foi a que obteve melhores resultados de validação, após um número razoável de repetições do processo de treino.

5.3.2. Primeira Classificação: Três Géneros Musicais

O objectivo nesta primeira tarefa de classificação consiste em classificar música em três classes: peças para flauta, piano e violino.

Como foi referido na secção 5.1, é utilizado um conjunto de treino composto por 120 peças musicais e um conjunto de validação com 60 peças musicais. Os três subgéneros musicais, flauta, piano e violino, estão igualmente representados em cada um dos conjuntos.

³³ Em terminologia Inglesa: Probabilistic Principal Component Analysis – PPCA.

Vão ser mostrados em seguida os melhores resultados alcançados na distinção entre as três classes, (Tabela 5.10). A matriz de covariância utilizada nesta classificação foi a PPCA.

GMM 75%	flauta	piano	violino
flauta	65	0	10
piano	25	90	20
violino	10	10	70

Tabela 5.10. Matriz de confusão da música instrumental: GMM.

A matriz de confusão mostra, que foram bem classificadas 65% das músicas de flauta, 90% das músicas de piano e 70% das músicas de violino. No total este classificador obteve 75% de taxa de sucesso.

O classificador conseguiu separar da melhor forma a classe piano, o mesmo não acontecendo para as classes flauta e violino.

5.3.3. Segunda Classificação: Dois Géneros Musicais

Nesta segunda tarefa de classificação pretende-se classificar música em duas classes: coral e ópera.

Os conjuntos de treino e validação utilizados são compostos respectivamente por 80 e 40 peças musicais. Cada um desses conjuntos, contém uma representação em igual número, dos dois géneros musicais em estudo.

Na tabela seguinte, podem ser visualizados os melhores resultados para este segundo problema de classificação (Tabela 5.11).

GMM 85%	coral	ópera
coral	80	10
ópera	20	90

Tabela 5.11. Matriz de confusão da música vocal: GMM.

A matriz de covariância utilizada foi a esférica. Em relação aos resultados de classificação, foram conseguidos percentagens de peças musicais bem classificadas de 80% para música coral e 90% para ópera.

5.3.4. Terceira Classificação: Cinco Géneros Musicais

Pretende-se nesta tarefa de classificação distinguir música clássica em cinco géneros musicais: flauta, piano, violino, coral e ópera.

Os conjuntos de treino e validação utilizados são compostos respectivamente por 200 e 100 peças musicais. Cada um desses conjuntos, contém uma representação em igual número, dos cinco géneros musicais em estudo.

Vão ser apresentados em seguida os melhores resultados alcançados na distinção entre as cinco classes, (Tabela 5.12). A matriz de covariância utilizada nesta classificação foi a diagonal.

GMM 53%	flauta	piano	violino	coral	ópera
flauta	55	15	10	30	15
piano	0	70	0	0	0
violino	10	5	60	10	40
coral	30	10	10	50	15
ópera	5	0	20	10	30

Tabela 5.12. Matriz de confusão da música vocal e instrumental: GMM

Como se pode ver através da tabela anterior, foram obtidas como percentagens de músicas correctamente classificadas, 55% para flauta, 70% para piano, 60% para violino, 50% para coral e 30% para ópera.

Os resultados da aplicação do GMM a este problema de classificação são insatisfatórios, principalmente para as classes ópera, coral e flauta. De notar que a “confusão” de classes acontece essencialmente para os casos em que as duas classes são uma vocal e uma instrumental. Por exemplo 30% das músicas de flauta foram classificadas em coral, 30% das de coral foram classificadas em flauta e 40% das músicas de ópera foram classificadas em violino.

A observação anterior, juntamente com a evidência dos resultados nos 1º e 2º problemas de classificação, leva a concluir que o GMM conseguiu limitar razoavelmente bem as classes vocais e conseguiu o mesmo para as instrumentais. Já para a mistura de classes dos dois tipos os resultados foram deficientes. Provavelmente as características escolhidas são boas para identificar géneros similares vocais e instrumentais, mas não são boas para discriminar entre voz e instrumentos, tal como se concluiu para o KNN.

5.4. REDES PERCEPTRÃO MULTICAMADA

Vão ser expostas em seguida as premissas iniciais que nortearam esta classificação. Serão descritos ainda os resultados alcançados e as principais conclusões.

5.4.1. Princípios Gerais

As redes MLP utilizaram todas a mesma configuração no que concerne à sua estrutura:

- Utilização de uma camada escondida.
- Camada de entrada com 40 neurónios, correspondentes às 40 características extraídas de cada peça musical.

- Camada de saída com: três neurónios para o primeiro problema (correspondentes às classes flauta, piano e violino); dois neurónios para o segundo problema (correspondentes às classes coral e ópera); cinco neurónios para o terceiro problema (correspondentes às classes flauta, piano, violino, coral e ópera).
- Teste para cada um dos problemas de classificação, de vários conjuntos de neurónios (10, 15, 20, 25, 30) para a camada escondida. Estes testes consistem no processo completo de treino validação, repetido 20 vezes para cada conjunto de neurónios, inicializando os parâmetros da rede em cada repetição. Para cada um dos três problemas escolhe-se em seguida a configuração (número de neurónios na camada escondida e pesos após o treino) com melhores resultados de validação.

Foram escolhidos aqueles valores possíveis para o número de neurónios da camada escondida, baseado numa regra definida em [Sarle, 2001] que diz que o número de neurónios da camada escondida deve ser igual a ((número de neurónios de entrada + número de neurónios de saída) * (2/3)). Foram testados outros valores abaixo e acima para escolher o melhor.

Existiu também um conjunto de pressupostos comuns em relação ao processo de treino e validação:

- Utilização para treino da técnica da retropropagação do erro com o algoritmo de *Levenberg-Marquardt*.
- Treino do tipo *batch*.
- Utilização da função de activação *logsig*, tanto na camada escondida como na camada de saída da rede. O uso desta função faz com que os valores de saída da rede estejam compreendidos no intervalo [0,1]. Para cada neurónio de saída, valores próximos de 1 revelam proximidade entre a música de teste e o género musical associado a esse neurónio, enquanto valores próximos de 0 revelam o contrário.

Em todo o trabalho foi sempre utilizada uma única camada escondida na estrutura da rede. A razão para tal é que está provado que utilizando o algoritmo de

retropropagação do erro com uma função de activação não linear, (e.g., *logsig*), uma camada escondida é suficiente para aproximar qualquer função com uma precisão arbitrária [Funahashi, 1989].

5.4.2. Regras de Classificação

As classificações foram efectuadas segundo duas perspectivas, que serão identificadas deste ponto em diante, por regra de cálculo de percentagem 1 (RCP1) e regra de cálculo de percentagem 2 (RCP2).

Em seguida vão ser descritas as duas perspectivas anteriores.

Regras de cálculo de percentagens 1:

Considera-se que uma música de um determinado género musical é bem classificada, se o valor mais alto de saída da rede pertence a esse género e esse valor é maior ou igual a 0,7. Quando uma música é bem classificada, é-o sem margem para dúvidas.

Quando todos os valores de saída da rede são inferiores a 0,7, a música é considerada sem classificação. O valor mais alto não é suficientemente alto para evitar ambiguidades na classificação.

É ainda calculado, para permitir distinguir melhor os resultados obtidos para as várias classes, um campo representado por $\{gn_2 \leq 0,2\}$. O objectivo deste campo, é mostrar de entre as músicas bem classificadas, aquelas cuja distância ao 2º género com valor mais alto, é inferior ou igual a 0,2. (e.g., para uma música bem classificada, o valor mais alto de saída da rede é 0,8. O 2º valor mais alto é 0,65. Esta situação está no âmbito de aplicação desta regra). Uma peça musical nesta situação é apenas predominantemente de um dado género musical.

Se nenhuma das regras anteriores for aplicada, conclui-se que a música em questão foi mal classificada

Regras de cálculo de percentagens 2:

Nesta perspectiva, considera-se que uma música de um dado género é bem classificada, se o valor de saída da rede para esse género for o mais alto, independentemente da ordem de grandeza.

Se a regra anterior não for aplicada, então é porque a musica foi mal classificada.

Considera-se ainda um campo representado por $\{gn_2 \geq 0,7\}$. Este campo contabiliza de entre as músicas mal classificadas, aquelas cujo 2º valor mais alto de saída da rede corresponde ao género correcto e além disso esse valor é maior ou igual a 0,7. (e.g., uma música obtém 0,96 para uma saída da rede que não corresponde ao género correcto e 0,75 para a saída que corresponde ao género correcto, então essa música será contabilizada no âmbito desta regra). Esta regra permite detectar as músicas que embora mal classificadas, têm valores suficientemente altos para o género a que pertencem.

5.4.3. Primeira Classificação: Três Géneros Musicais

As peças musicais foram classificadas em três classes: flauta, piano e violino. Foram utilizadas para treino 120 peças (40 de cada classe) e 60 para validação (20 de cada classe). Foi obtida uma percentagem de peças bem classificadas de 85% para RCP2 e 83,3% para RCP1.

Como se verificou anteriormente, foram testadas várias estruturas diferentes para as redes neuronais. As diferenças têm unicamente a ver com o número de neurónios da camada escondida, já que o número de neurónios da camada de entrada é 40 e da camada de saída é 3. Pretende-se portanto, através da experiência e da repetição (com inicializações dos parâmetros da rede em cada repetição), descobrir qual a rede (estrutura e parâmetros) que obtém melhores resultados de validação.

Assim, a matriz de confusão das camadas, para esta primeira tarefa de classificação é (Tabela 5.13):

	RCP1	RCP2
10	81,6	81,6
15	83,3	83,3
20	83,3	85
25	80	81,6
30	80	83,3

Tabela 5.13. Matriz de confusão das camadas para a primeira classificação: MLP.

Como se verifica através da tabela anterior, o melhor resultado de validação foi para uma rede com 20 neurónios na camada escondida (40x20x3). Consegue 83,3% e 85% respectivamente para a primeira e segunda regra de cálculo de percentagens.

As tabelas 5.14 e 5.15 mostram respectivamente para RCP1 e RCP2, os resultados para o melhor conjunto treino/validação:

RCP1 83,3%	flauta	piano	violino
flauta	85	10	5
piano	5	80	10
violino	5	10	85
s/classificação	5	0	0
gn 2 <= 0,2	0	0	0

Tabela 5.14. Matriz de confusão da música instrumental, RCP1: MLP.

RCP2 85%	flauta	piano	violino
flauta	90	10	5
piano	5	80	10
violino	5	10	85
gn 2 >= 0,7	0	0	10

Tabela 5.15. Matriz de confusão da música instrumental, RCP2: MLP.

As colunas da matriz representam o género a que pertence a música e as linhas o resultado da validação. Assim por exemplo, para RCP1 (Tabela 5.14), 80% das músicas de piano foram bem classificadas, 5% das músicas de violino foram classificadas como flauta e 5% das músicas de flauta não obtiveram classificação.

Analisando RCP1 (Tabela 5.14), conclui-se que foram obtidas como percentagem de músicas bem classificadas, 85% para flauta, 80% para piano e 85% para violino. Em relação por exemplo às peças de flauta, 5% foram mal classificadas como piano, 5% foram classificadas como violino e 5% não obtiveram classificação (verdadeiros negativos). Verifica-se também que 10% das peças para piano e 5% das peças para violino são erradamente classificadas como sendo peças para flauta (falsos positivos). Infere-se ainda que a distância entre o valor do género correcto e o segundo valor mais alto é sempre superior a 0,2 (“ $gn_2 < 0,2$ ” = 0%), logo as músicas bem classificadas, são-o sem margem para dúvidas.

Em relação aos dados referentes a RCP2 (Tabela 5.15), a percentagem de músicas correctamente classificadas foi de 90% para flauta, 80% para piano e 85% para violino. Conclui-se ainda que 10% das músicas de violino que foram mal classificadas, obtiveram para a classe violino o segundo valor mais alto e esse valor é superior a 0,7 (“ $gn_2 > 0,7$ ” = 10%).

Analisando os erros de classificação, reparou-se que aconteceram essencialmente em músicas nas quais os instrumentos são tocados de uma forma pouco usual para esses mesmos instrumentos. Com certeza, esse tipo de músicas tem uma menor representação nos exemplos de treino da rede. Por exemplo, há 2 músicas de violino que obtiveram valores superiores a 0,7 para a classe violino, mas foram ultrapassados pelos valores da classe piano – 2 extractos musicais de *Bach* e *Mozart*. Esses extractos têm em comum, o facto de os instrumentos principais (violinos) serem tocados de uma forma muito lenta e terem poucas variações de amplitude. Essas características são típicas das músicas para piano. Os valores bastante altos para a classe violino explicam-se por, apesar de em diversos aspectos não serem músicas típicas de violino, as características tímbricas extraídas de cada uma dessas músicas detectou a presença desses instrumentos.

5.4.4. Segunda Classificação: Dois Géneros Musicais

As peças musicais foram classificadas em duas classes, música coral e ópera. Foram utilizadas para treino 80 músicas (40 de cada classe) e 40 para validação (20 de cada classe). Foi obtida uma percentagem de músicas bem classificadas de 90% (RCP1 e RCP2).

Os melhores resultados de classificação foram obtidos com 25 neurónios na camada escondida (40x25x2), como se pode verificar através da Tabela 5.16.

	RCP1	RCP2
10	87,5	87,5
15	87,5	87,5
20	87,5	87,5
25	90	90
30	87,5	87,5

Tabela 5.16. Matriz de confusão das camadas para a segunda classificação: MLP.

Foram obtidos 90% tanto para RCP1 como para RCP2.

As tabelas 5.17 e 5.18 resumem ao pormenor a classificação para o melhor conjunto treino/validação deste problema.

RCP1 90%	coral	ópera
coral	90	10
ópera	10	90
s/classificação	0	0
gn 2 <= 0,2	0	0

Tabela 5.17. Matriz de confusão da música vocal, RCP1: MLP.

RCP2 90%	coral	ópera
coral	90	10
ópera	10	90
gn 2 \geq 0,7	0	0

Tabela 5.18. Matriz de confusão da música vocal, RCP2: MLP.

Analisando RCP1 (Tabela 5.17), conclui-se que foram obtidas como percentagem de músicas bem classificadas, 90% para música coral e 90% para ópera.

De notar que as percentagens anteriores mostram total ausência de ambiguidade nesta classificação, já que as distâncias entre as músicas correctamente classificadas e o segundo valor mais alto são superiores a 0,2 (“gn 2 \leq 0,2” = 0%).

Quanto a RCP2 (Tabela 5.18), os resultados obtidos são os mesmos de RCP1, i.e., 90% de músicas bem classificadas, tanto para música coral como para ópera.

São quatro as peças musicais mal classificadas, duas de ópera e duas corais. Em relação às músicas corais que são erradamente classificadas como peças de ópera, uma delas tem parte instrumental, ao contrário da maioria das músicas de treino dessa classe que são *a cappella*. Essa música tem ainda a sobressair uma voz feminina cuja prestação pode facilmente, para a média dos humanos ser confundida com ópera. A outra música coral tem várias vezes cujas prestações atingem altas frequências em todo o excerto. Quanto aos excertos de ópera mal classificados, tirando o facto de serem um pouco atípicos em relação a essa classe, já que são partes bastante calmas de ópera e com pausas, sendo mesmo uma delas *a cappella* (como a maioria das peças corais), não se encontram razões claras para o erro na classificação. A única conclusão que se pode tirar talvez, é que as características extraídas das peças musicais sejam bastante boas para os casos bem comportados, sendo necessário a inclusão de novas características e/ou a eliminação de características redundantes para os casos mais atípicos.

5.4.5. Terceira Classificação: Cinco Géneros Musicais

O objectivo deste problema é o de classificar em cinco classes: flauta, piano, violino, coral e ópera. As peças musicais utilizadas para treino foram 200 (40 de cada classe) e para validação 100 (20 de cada classe). Foi obtida uma percentagem de sucesso de 76% (RCP2) na classificação.

A matriz de confusão de camadas para este problema de classificação ilustra-se em seguida, (Tabela 5.19).

	RCP1	RCP2
10	63	64
15	64	70
20	64	70
25	64	69
30	62	76

Tabela 5.19. Matriz de confusão das camadas para a terceira classificação: MLP.

Analisando a tabela anterior, conclui-se que os melhores resultados de classificação para RCP1, com 64%, foram obtidos igualmente com 15, 20 e 25 neurónios na camada escondida. No que concerne a RCP2, o melhor resultado foi de 76% com 30 neurónios na camada escondida (40x30x5).

Das três redes com melhor resultado para RCP1, considera-se para estudo mais aprofundado a rede com 20 neurónios na camada escondida, já que foi a rede cujas classificações foram mais categóricas, i.e., tiveram uma percentagem média do campo “ $gn_2 \leq 0,2$ ” inferior às restantes.

A Tabela 5.20 apresenta os resultados de RCP1 para a rede escolhida.

RCP1 64%	flauta	piano	violino	coral	ópera
flauta	65	15	5	0	10
piano	10	65	0	10	0
violino	0	10	70	10	0
coral	15	0	5	50	0
ópera	0	0	5	15	70
s/classificação	10	10	15	15	20
gn 2 <= 0,2	0	10	20	5	15

Tabela 5.20. Matriz de confusão das músicas instrumental e vocal, RCP1: MLP.

Como se pode ver através da tabela anterior, foram obtidas como percentagens de músicas correctamente classificadas, 65% para flauta, 65% para piano, 70% para violino, 50% para coral e 70% para ópera. A percentagem geral de sucesso foi de 64%.

Os resultados de validação para RCP2 são ilustrados na tabela seguinte (Tabela 5.21).

RCP2 76%	flauta	piano	violino	coral	ópera
flauta	75	20	0	10	10
piano	5	65	0	15	5
violino	0	5	85	0	0
coral	10	5	10	75	5
ópera	10	5	5	0	80
gn 2 >= 0,7	0	5	5	0	0

Tabela 5.21. Matriz de confusão das músicas instrumental e vocal, RCP2: MLP.

Segundo RCP2, a percentagem de sucesso na classificação foi de 75% para flauta, 65% para piano, 85% para violino, 75% para coral e 80% para ópera. Nota-se aqui que tendo em conta a junção de géneros musicais tão distintos, uma classificação global de 76% pode ser considerada bastante razoável.

A análise das duas tabelas anteriores parece indicar que a aprendizagem das características de cada género musical foi superior nesta última rede (Tabela 5.21). Para tentar comprovar isso, vai ser apresentada em seguida a classificação RCP1 correspondente à rede com 30 neurónios na camada escondida (Tabela 5.22).

RCP1 62%	flauta	piano	violino	coral	ópera
flauta	65	20	0	5	0
piano	0	65	0	0	0
violino	0	5	70	0	0
coral	5	5	5	50	5
ópera	0	0	5	0	60
s/classificação	30	5	20	45	35
gn 2 <= 0,2	10	0	0	15	5

Tabela 5.22. Matriz de confusão das músicas instrumental e vocal, RCP1 (2): MLP.

Como se verifica na tabela anterior, as percentagens de músicas bem classificadas são muito parecidas com as da rede melhor para RCP1 (Tabela 5.20) e as classificações globais são muito próximas: 64% e 62%.

O que se nota na Tabela 5.22 em relação à Tabela 5.20 é uma muito menor percentagem média de músicas mal classificadas: 11% contra 22%. Em contrapartida na tabela 5.22 há uma percentagem média muito maior de músicas que não obtiveram qualquer classificação: 27% contra 14%. A juntar a isto, ainda o facto de o campo $\{gn\ 2 \leq 0,2\}$ ter tido um valor médio mais baixo em 5.22 que em 5.20: 6% contra 10%. Estes resultados no seu conjunto, mostram que o classificador considerado em 5.22 aprendeu com maior precisão as características fundamentais de cada género musical, já que tendo uma percentagem de sucesso muito parecido com o classificador de 5.20 tem muito menos músicas mal classificadas, preferindo antes não lhes atribuir qualquer classificação. Além do mais há uma percentagem menor de músicas que apesar de bem classificadas o foram com alguma ambiguidade, como se pode atestar na diferença de percentagens de $\{gn\ 2 \leq 0,2\}$: 6% contra 10%. É evidente ainda que para um sistema real de classificação seria sempre preferível, apesar de tudo uma não classificação que uma má classificação.

No caso deste terceiro problema de classificação em cinco géneros musicais esperava-se sem dúvida resultados menos precisos do que nos primeiro e segundo problemas, já que nesses, os géneros além de serem em menor número são mais parecidos e têm sempre algo em comum (instrumentos apenas e voz). As respectivas redes são treinadas fundamentalmente no sentido da distinção do timbre e do *pitch* dos instrumentos principais e da voz.

No terceiro problema misturam-se géneros instrumentais com géneros vocais, numa rede que não é treinada explicitamente para distinguir música instrumental de música vocal. Depois de ouvidos novamente alguns dos extractos de música mal classificados, concluiu-se que muitos deles foram confundidos com outros géneros pela forma como eram interpretados.

5.5. COMPARAÇÃO ENTRE CLASSIFICADORES

Pretende-se, após a apresentação de todos os resultados experimentais, escolher, para cada uma das tarefas de classificação, o classificador mais eficiente, ou seja, aquele que alcançou as melhores taxas de sucesso.

Vão ser sumariados em seguida, através da Tabela 5.23, os resultados de todos os classificadores.

	1ª T.C:	2ª T.C.	3ª T.C.
KNN(1)	68,3	82,5	59
KNN(3)	80	85	61
KNN(5)	80	82,5	67
GMM	75	85	53
MLP	85	90	76

Tabela 5.23. Resultados finais dos classificadores para as três tarefas de classificação.

Os melhores classificadores foram as redes MLP, independentemente da tarefa de classificação em análise. Tal conclusão seria de algum modo esperado, uma vez que as redes MLP permitem obter funções de decisão fortemente não lineares. Pelo

contrário os resultados alcançados pelo GMM são os piores de todos os classificadores, à excepção do KNN com $K=1$.

Por tarefas de classificação, nota-se que independentemente do classificador, os resultados alcançados são bastante satisfatórios para as 1ª e 2ª tarefas de classificação, já que se situam numa gama de valores entre 75% para o GMM (1ª tarefa de classificação) e 90% para as redes MLP (2ª tarefa de classificação). As redes MLP aprenderam, através do treino, a identificar bem as características da música instrumental e da música vocal em análise. O GMM conseguiu separar bem as classes, principalmente as classes vocais. O resultados do KNN foram muito parecidos considerando K igual a 3 ou 5, piorando muito com $K=1$.

Na 3ª tarefa de classificação, os resultados do KNN e do GMM desiludiram, principalmente este último. Esse desempenho pode ter ficado a dever-se à complexidade do problema que não permitiu definir fronteiras claras entre as várias classes. Quanto às redes MLP, os resultados podem-se considerar muito satisfatórios para a complexidade de géneros em análise.

O classificador escolhido para o teste final (secção 5.6) será portanto o MLP.

5.6. PROTÓTIPO

Nos capítulos e secções anteriores descreveu-se todo o processo desde a captura dos extractos musicais, passando pela extracção de características das peças capturadas e terminando na sua classificação em géneros musicais.

Nesta secção, pretende-se fazer uma aproximação a um sistema real de classificação automática de géneros musicais. Este sistema irá tentar validar correctamente um conjunto de 100 músicas, 20 de cada uma das classes, flauta, piano, violino, coral e ópera. De cada música serão extraídos 10 extractos de 6s, escolhidos de igual forma para todas as músicas. Cada música será classificada no género musical mais representado entre os seus 10 extractos.

Para cada um dos três problemas de classificação, o classificador mais eficiente foi o MLP, como se verifica através da (Tabela 5.23). Os classificadores a utilizar serão

portanto os que obtiveram 85%, 90% e 76% respectivamente para as primeira, segunda e terceira tarefas de classificação.

O estudo irá ser validado da mesma forma que na secção 5.2, para as redes MLP, i.e., vão ser consideradas as duas regras de classificação RCP1 e RCP2.

De notar que as 10 músicas a validar são as correspondentes ao conjunto de validação inicial. Só que nesse, para cada música foi extraído um extracto teoricamente “bem comportado”, aqui são extraídos 10 extractos de uma forma perfeitamente automática, em que o critério de escolha dos extractos está relacionado apenas com a sua posição relativa na música.

A listagem de todas as músicas encontra-se no Anexo 1.

5.6.1. Extração das Peças Musicais

Ao extrair os 10 extractos de cada música, pretende-se que estas sejam uma representação de toda a música e não de uma parte apenas. Como as músicas podem ter tamanhos muito variados, decidiu-se considerar três tipos de extração, cuja escolha depende precisamente da duração de cada música.

Consideraram-se como critérios de extração os seguintes intervalos de duração: menos de 2 minutos, entre 2 e 3 minutos e mais de 3 minutos.

Podem ser visualizados na tabela seguinte os três tipos de extração efectuados, (Tabela 5.24).

Critérios de Extracção	Duração da Música			
	< 2 m	³ 2m e < 3m	³ 3m	
Peças Musicais	1º	3 - 9	10 - 16	30 - 36
	2º	9 - 15	22 - 28	46 - 52
	3º	15 - 21	34 - 40	1:02 - 1:08
	4º	21 - 27	46 - 52	1:18 - 1:24
	5º	27 - 33	58 - 1:04	1:34 - 1:40
	6º	33 - 39	1:10 - 1:16	1:50 - 1:56
	7º	39 - 45	1:22 - 1:28	2:06 - 2:12
	8º	45 - 51	1:34 - 1:40	2:22 - 2:28
	9º	51 - 57	1:46 - 1:52	2:38 - 2:44
	10º	57 - 1:03	1:58 - 2:04	2:54 - 3:00

Tabela 5.24. Critérios de extracção das peças musicais.

Como de verifica pela tabela anterior, mediante o tamanho da música em análise, são escolhidos 10 momentos da música para se efectuarem capturas de 6 segundos. Exemplificando para uma música de 5 minutos, o 7º extracto corresponderá a um intervalo dessa música entre os momentos 2:06 minutos e 2:12 minutos.

5.6.2. Regras de Classificação

As regras básicas de classificação continuam a ser as aplicadas na secção 5.2, ou sejam, RCP1 e RCP2.

No entanto foram definidas mais algumas regras relacionadas especificamente com a validação das músicas baseada na validação dos seus extractos.

Para RCP1:

Uma música é classificada no género musical mais representado pelos seus extractos. Se existirem mais extractos sem classificação do que classificados num determinado género musical, então a música será considerada sem classificação.

Se dois ou mais géneros estiverem igualmente representados e não existir outro género com maior representação, então a música será classificada nesses géneros.

Se existir igual representação, de um ou mais géneros e de extractos sem classificação e não existir outro género com maior representação, então a música será classificada nesses géneros.

Para RCP2:

Uma música é classificada no género musical mais representado pelos seus extractos.

Se existir igual representação de dois ou mais géneros, a música pertencerá ao género ou géneros que em RCP1 tenham tido maior representação. Caso se mantenham igualmente representados então a música será classificada nesses géneros, (e.g., Em RCP1, coral=4, ópera=5, s/classificação=1. Em RCP2, coral=ópera=5. A música é classificada em ópera, segundo RCP1 e RCP2).

5.6.3. Primeira Tarefa de Classificação

Neste primeiro problema pretende-se classificar um total de 60 músicas, 20 de cada um dos estilos: flauta, piano e violino.

Cada música é representada por 10 extractos e a sua classificação resulta no género musical mais representado pelos seus extractos.

A tabela seguinte mostra ao pormenor, utilizando RCP1 e RCP2, a classificação de cada música, de acordo com as classificações de cada um dos seus extractos (Tabela 5.25).

	RCP1					RCP2			
	flauta	piano	violino	s/Clasf	CF RCP1	flauta	piano	violino	CF RCP2
FLAUTA									
1	10	-	-	-	flauta	10	-	-	flauta
2	7	1	2	-	flauta	7	1	2	flauta
3	10	-	-	-	flauta	10	-	-	flauta
4	5	1	4	-	flauta	5	1	4	flauta
5	6	-	2	2	flauta	8	-	2	flauta
6	5	1	3	1	flauta	6	1	3	flauta
7	3	2	5	-	violino	3	2	5	violino
8	6	2	2	-	flauta	6	2	2	flauta
9	7	-	3	-	flauta	7	-	3	flauta
10	2	2	5	1	violino	2	3	5	violino
11	2	7	-	1	piano	3	7	-	piano
12	5	1	4	-	flauta	5	1	4	flauta
13	3	1	6	-	violino	3	1	6	violino
14	7	1	2	-	flauta	7	1	2	flauta
15	9	1	-	-	flauta	9	1	-	flauta
16	5	-	4	1	flauta	5	-	5	flauta
17	9	1	-	-	flauta	9	1	-	flauta
18	6	-	3	1	flauta	6	-	4	flauta
19	5	2	3	-	flauta	5	2	3	flauta
20	4	-	6	-	violino	4	-	6	violino
PIANO									
21	-	10	-	-	piano	-	10	-	piano
22	1	3	5	1	violino	2	3	5	violino
23	4	4	2	-	flauta / piano	4	4	2	flauta / piano
24	1	6	3	-	piano	1	6	3	piano
25	3	7	-	-	piano	3	7	-	piano

26	-	9	-	1	piano	1	9	-	piano
27	2	1	7	-	violino	2	1	7	violino
28	1	8	1	-	piano	1	8	1	piano
29	3	4	2	1	piano	3	4	3	piano
30	4	2	2	2	flauta	5	2	3	flauta
31	1	2	7	-	violino	1	2	7	violino
32	1	3	6	-	violino	1	3	6	violino
33	2	4	4	-	piano / violino	2	4	4	piano / violino
34	1	9	-	-	piano	1	9	-	piano
35	1	6	3	-	piano	1	6	3	piano
36	3	-	6	1	violino	3	1	6	violino
37	1	2	6	1	violino	1	2	7	violino
38	1	9	-	-	piano	1	9	-	piano
39	1	9	-	-	piano	1	9	-	piano
40	1	9	-	-	piano	1	9	-	piano
VIOLINO									
41	-	-	10	-	violino	-	-	10	violino
42	2	-	8	-	violino	2	-	8	violino
43	-	-	10	-	violino	-	-	10	violino
44	-	-	9	1	violino	-	-	10	violino
45	-	-	10	-	violino	-	-	10	violino
46	1	-	9	-	violino	1	-	9	violino
47	3	-	6	1	violino	3	-	7	violino
48	-	-	10	-	violino	-	-	10	violino
49	2	-	7	1	violino	3	-	7	violino
50	2	-	8	-	violino	2	-	8	violino
51	-	-	9	1	violino	-	-	10	violino
52	1	-	9	-	violino	1	-	9	violino
53	-	-	10	-	violino	-	-	10	violino
54	3	1	5	1	violino	4	1	5	violino
55	-	-	10	-	violino	-	-	10	violino

56	-	-	10	-	violino	-	-	10	violino
57	-	1	9	-	violino	-	1	9	violino
58	-	-	10	-	violino	-	-	10	violino
59	2	-	8	-	violino	2	-	8	violino
60	-	-	9	1	violino	-	1	9	violino

Tabela 5.25. Resultados gerais de validação em 3 classes.

Analisando a tabela anterior, verifica-se por exemplo que segundo RCP1, para a 18ª música de flauta: 6 extractos foram validados como flauta, 3 extractos como violino e 1 extracto sem classificação. Segundo RCP2: 6 extractos como flauta e 4 como violino. Em ambas as perspectivas, a música foi classificada correctamente em flauta.

A listagem de todas as músicas encontra-se no Anexo 1. Assim a 18ª música referida anteriormente é *Presto de Concerto in G major (Wq. 169/H.445) de Carl Bach*.

Existem duas músicas de piano classificadas simultaneamente em 2 géneros musicais. São elas, a 23ª em flauta e piano e a 33ª em piano e violino.

Não houve nenhuma música que fosse considerada sem classificação.

Vão ser sintetizados nas tabelas seguintes os resultados alcançados pelas 60 músicas de validação, (Tabela 5.26) e (Tabela 5.27).

RCP1 78%	flauta	piano	violino
flauta	75	9	-
piano	5	59,1	-
violino	20	31,9	100
s/classificação	-	-	-

Tabela 5.26. Matriz de confusão do protótipo para a primeira classificação: RCP1.

RCP2 78%	flauta	piano	violino
flauta	75	9	-
piano	5	59,1	-
violino	20	31,9	100

Tabela 5.27. Matriz de confusão do protótipo para a primeira classificação: RCP2.

A percentagem de músicas bem classificadas foi de 78%, independentemente das perspectivas de classificação. Por género, foram bem classificadas, 75% das músicas de flauta, 59,1% das de piano e 100% das de violino. Não houve nenhuma música sem classificação.

De notar os resultados excelentes para a classe violino que obteve 100%. Todas as músicas foram categoricamente classificadas, o que mostra que a rede aprendeu a identificar correctamente as características do instrumento violino, nomeadamente o seu timbre.

As classificações de piano desiludiram um pouco com apenas 59,1% de músicas correctamente classificadas. 31,9% foram classificadas erradamente na classe violino. Não se encontrou uma justificação razoável para este facto.

De qualquer forma estes resultados são considerados promissores, já que os resultados globais alcançados com esta extracção cega de peças (78%) ficou bastante perto da extracção de exemplos “bem comportados” (85%).

5.6.4. Segunda Tarefa de Classificação

O objectivo deste problema é a classificação de música num de dois géneros musicais: coral e ópera.

Vão ser classificados 400 extractos musicais, pertencentes a um total de 40 músicas. Dessas, 20 são corais e 20 são óperas.

O género musical mais representado pelos extractos de uma determinada música é o género em que essa música é classificada.

A tabela seguinte mostra ao pormenor, utilizando RCP1 e RCP2, a classificação de cada música, de acordo com as classificações de cada um dos seus extractos (Tabela 5.28).

	RCP1				RCP2		
	coral	ópera	s/Clasf	C.F. RCP1	coral	ópera	C.F. RCP2
CORAL							
61	9	1	-	coral	9	1	coral
62	5	5	-	coral / ópera	5	5	coral / ópera
63	9	1	-	coral	9	1	coral
64	3	7	-	ópera	3	7	ópera
65	3	7	-	ópera	3	7	ópera
66	9	1	-	coral	9	1	coral
67	7	3	-	coral	7	3	coral
68	6	4	-	coral	6	4	coral
69	10	-	-	coral	10	-	coral
70	10	-	-	coral	10	-	coral
71	8	2	-	coral	8	2	coral
72	7	3	-	coral	7	3	coral
73	10	-	-	coral	10	-	coral
74	10	-	-	coral	10	-	coral
75	9	1	-	coral	9	1	coral
76	7	3	-	coral	7	3	coral
77	9	1	-	coral	9	1	coral
78	10	-	-	coral	10	-	coral
79	9	1	-	coral	9	1	coral
80	5	5	-	coral / ópera	5	5	coral / ópera
ÓPERA							
81	3	7	-	ópera	3	7	ópera
82	2	8	-	ópera	2	8	ópera
83	5	5	-	coral / ópera	5	5	coral / ópera
84	6	4	-	coral	6	4	coral

85	4	6	-	ópera	4	6	ópera
86	1	9	-	ópera	1	9	ópera
87	6	4	-	coral	6	4	coral
88	4	6	-	ópera	4	6	ópera
89	4	5	1	ópera	5	5	ópera
90	3	7	-	ópera	3	7	ópera
91	7	3	-	coral	7	3	coral
92	2	8	-	ópera	2	8	ópera
93	6	4	-	coral	6	4	coral
94	1	9	-	ópera	1	9	ópera
95	5	5	-	coral / ópera	5	5	coral / ópera
96	2	8	-	ópera	2	8	ópera
97	3	7	-	ópera	3	7	ópera
98	6	4	-	coral	6	4	coral
99	4	6	-	ópera	4	6	ópera
100	5	5	-	coral / ópera	5	5	coral / ópera

Tabela 5.28. Resultados gerais de validação em 2 classes.

Pode-se verificar pela tabela anterior que existem cinco músicas que foram classificadas simultaneamente em coral e ópera. São elas a 62^a, 80^a, 83^a, 95^a e 100^a. As restantes músicas foram classificadas sem ambiguidade num dos dois géneros musicais. De notar que em 400 extractos, apenas um foi considerado sem classificação. Esse extracto pertence à 89^a música – *Allí está, riyendo* da ópera *De la vida breve* de *Manuel de Falla*.

Os resultados da análise da tabela anterior vão ser sintetizados em seguida nas tabelas (5.29) para RCP1 e (5.30) para RCP2.

RCP1 73,5%	coral	ópera
coral	81,8	34,8
ópera	18,2	65,2
s/classificação	-	-

Tabela 5.29. Matriz de confusão do protótipo para a segunda classificação: RCP1.

RCP2 73,5%	coral	ópera
coral	81,8	34,8
ópera	18,2	65,2

Tabela 5.30. Matriz de confusão do protótipo para a segunda classificação: RCP2.

A percentagem de músicas bem classificadas neste segundo problema de classificação foi de 73,5% tanto para RCP1 como para RCP2. Esta percentagem resulta da média das percentagens por género, que foram de 81,8% para coral e de 65,2% para ópera.

Como se pode visualizar na (Tabela 5.29), nenhuma música foi considerada sem classificação.

Estes resultados surpreenderam um pouco pela negativa, já que houve uma grande descida na percentagem global de sucesso entre a classificação com exemplos

“bem comportados” e a classificação cega. Essa descida foi de 90% para 73,5% e está fundamentalmente relacionada com a percentagem grande de músicas de ópera que foram erradamente classificadas na classe coral (34,8%). Analisando alguns desses casos, constatou-se que algumas partes da maioria das óperas têm grandes parecenças com a música coral principalmente nas partes mais calmas. Na classificação que obteve melhores resultados, foram utilizados essencialmente exemplos típicos de ópera.

5.6.5. Terceira Tarefa de Classificação

Neste último problema pretende-se fazer a classificação de 100 músicas. Essa classificação depende, para cada música, da classificação dos seus 10 extractos. O género mais representado nos extractos será considerado o género da música.

A tabela seguinte mostra ao pormenor, utilizando RCP1 e RCP2, a classificação de cada música, de acordo com as classificações de cada um dos seus extractos (Tabela 5.31).

	RCP1							RCP2					
	flauta	piano	violino	coral	ópera	s/Clasf	CF PCR1	flauta	piano	violino	coral	ópera	CF PCR2
FLAUTA													
1	6	1	-	1	-	2	flauta	6	1	1	1	1	flauta
2	6	-	1	1	1	1	flauta	6	-	2	1	1	flauta
3	9	-	-	-	-	1	flauta	10	-	-	-	-	flauta
4	3	1	3	1	1	1	flauta / violino	3	1	4	1	1	violino
5	3	-	-	2	-	5	s/c	3	-	2	2	3	flauta
6	4	-	2	-	-	4	flauta	4	1	2	1	2	flauta
7	-	-	2	1	1	6	s/c	-	-	4	5	1	coral
8	5	-	-	-	1	4	flauta	5	1	2	1	1	flauta
9	6	1	1	-	-	2	flauta	6	2	1	1	-	flauta
10	1	1	1	1	1	5	s/c	1	1	1	2	5	ópera
11	1	5	-	1	1	2	piano	1	7	-	1	1	piano
12	6	1	-	1	1	1	flauta	6	1	-	1	2	flauta
13	4	-	2	-	2	2	flauta	4	-	2	-	4	flauta
14	3	2	2	-	-	3	flauta	4	2	3	-	1	flauta
15	8	1	-	-	1	-	flauta	8	1	-	-	1	flauta
16	3	-	1	2	2	2	flauta	3	-	3	2	2	flauta
17	6	-	-	-	-	4	flauta	9	1	-	-	-	flauta
18	1	-	3	3	-	3	violino / coral	2	-	4	4	-	violino / coral
19	3	1	4	-	-	2	violino	3	1	5	-	1	violino
20	3	-	1	1	1	4	s/c	4	1	1	1	3	flauta
PIANO													
21	-	9	-	-	-	1	piano	-	10	-	-	-	piano
22	-	2	2	1	3	2	ópera	-	3	3	1	3	ópera
23	-	3	1	3	3	-	piano / coral /	-	3	1	3	3	piano / coral /

							ópera						ópera
24	1	3	-	1	1	4	s/c	1	3	2	3	1	piano
25	1	6	-	-	-	3	piano	3	6	-	1	-	piano
26	1	6	-	-	1	2	piano	1	7	-	1	1	piano
27	-	-	1	2	4	3	ópera	-	1	1	4	4	ópera
28	-	6	-	2	1	1	piano	-	7	-	2	1	piano
29	2	3	-	3	2	-	piano / coral	2	3	-	3	2	piano / coral
30	3	1	-	-	1	5	s/c	5	1	1	2	1	flauta
31	-	2	2	3	-	3	coral	-	2	2	5	1	coral
32	2	2	2	2	-	2	flauta / piano / violino / coral	2	2	3	3	-	violino / coral
33	1	1	1	3	1	3	coral	1	2	1	5	1	coral
34	-	8	-	1	-	1	piano	1	8	-	1	-	piano
35	1	2	-	-	-	7	s/c	2	4	-	3	1	piano
36	1	-	2	2	-	5	s/c	1	-	4	5	-	coral
37	1	1	3	2	-	3	violino	1	2	3	4	-	coral
38	1	7	-	-	-	2	piano	1	9	-	-	-	piano
39	1	7	-	-	-	2	piano	2	8	-	-	-	piano
40	1	7	-	-	-	2	piano	1	9	-	-	-	piano
VIOLINO													
41	-	-	6	-	2	2	violino	-	-	7	1	2	violino
42	-	-	6	2	-	2	violino	-	-	7	3	-	violino
43	-	-	7	-	1	2	violino	-	-	7	-	3	violino
44	-	-	7	1	2	-	violino	-	-	7	1	2	violino
45	-	-	7	-	2	1	violino	-	-	8	-	2	violino
46	-	-	4	-	3	3	violino	-	-	6	-	4	violino
47	1	-	6	-	1	2	violino	2	-	7	-	1	violino
48	-	-	9	-	-	1	violino	-	-	10	-	-	violino
49	-	-	5	1	1	3	violino	-	-	7	1	2	violino

50	-	1	4	-	-	5	s/c	-	1	4	2	3	violino
51	-	-	7	-	-	3	violino	-	-	9	-	1	violino
52	1	-	7	-	-	2	violino	3	-	7	-	-	violino
53	-	-	3	1	1	5	s/c	-	-	7	1	2	violino
54	-	1	2	3	1	3	coral	-	2	4	3	1	violino
55	-	-	6	1	1	2	violino	-	-	7	1	2	violino
56	-	-	8	1	-	1	violino	-	-	9	1	-	violino
57	-	1	7	-	1	1	violino	-	1	8	-	1	violino
58	-	-	5	1	2	2	violino	-	-	6	1	3	violino
59	-	-	8	-	1	1	violino	-	-	9	-	1	violino
60	1	-	7	1	-	1	violino	1	-	7	2	-	violino
CORAL													
61	-	1	1	5	1	2	coral	-	1	1	7	1	coral
62	-	-	-	5	1	4	coral	2	-	-	7	1	coral
63	1	1	-	3	-	5	s/c	1	3	-	5	1	coral
64	-	-	-	3	6	1	ópera	-	-	-	4	6	ópera
65	1	-	2	-	3	4	s/c	1	1	3	-	5	ópera
66	-	1	-	5	2	2	coral	1	1	-	6	2	coral
67	-	3	-	3	2	2	piano / coral	-	3	1	4	2	coral
68	-	-	-	4	2	4	coral	2	-	-	5	3	coral
69	2	3	-	4	-	1	coral	3	3	-	4	-	coral
70	-	-	-	6	-	4	coral	-	-	-	10	-	coral
71	1	1	-	4	1	3	coral	2	1	-	5	2	coral
72	4	3	-	1	-	2	flauta	5	3	-	2	-	flauta
73	-	3	-	6	-	1	coral	-	3	-	7	-	coral
74	1	1	-	8	-	-	coral	1	1	-	8	-	coral
75	5	-	-	5	-	-	flauta / coral	5	-	-	5	-	flauta / coral
76	3	1	2	2	-	2	flauta	3	1	3	3	-	flauta
77	-	-	1	7	1	1	coral	-	-	2	7	1	coral
78	1	-	1	-	-	8	s/c	1	4	1	3	1	piano

79	-	1	-	5	1	3	coral	1	1	-	7	1	coral
80	5	-	-	4	-	1	flauta	5	-	-	4	1	flauta
ÓPERA													
81	1	-	-	-	4	5	s/c	3	-	1	1	5	ópera
82	2	-	-	-	4	4	ópera	3	-	-	1	6	ópera
83	3	1	-	-	2	4	s/c	3	2	-	2	3	flauta
84	-	-	2	4	2	2	coral	-	-	2	6	2	coral
85	-	-	3	-	1	6	s/c	1	-	3	1	5	ópera
86	1	-	1	1	4	3	ópera	1	-	2	2	5	ópera
87	-	-	1	4	1	4	coral	1	-	2	4	3	coral
88	1	1	2	3	2	1	coral	1	1	2	3	3	coral
89	-	1	-	2	3	4	s/c	1	1	-	4	4	ópera
90	1	-	1	2	4	2	ópera	1	-	1	2	6	ópera
91	-	-	2	7	-	1	coral	-	-	3	7	-	coral
92	-	-	1	2	7	-	ópera	-	-	1	2	7	ópera
93	-	-	4	-	2	4	violino	-	-	4	2	4	violino
94	-	-	2	2	4	2	ópera	-	-	2	3	5	ópera
95	2	-	1	2	4	1	ópera	2	-	2	2	4	ópera
96	-	-	3	1	4	2	ópera	-	-	5	1	4	violino
97	1	-	3	1	3	2	violino / ópera	1	1	4	1	3	violino
98	3	-	-	4	1	2	coral	4	-	-	5	1	coral
99	1	-	1	3	3	2	coral / ópera	1	-	2	3	4	ópera
100	-	1	1	3	2	3	coral	1	1	2	4	2	coral

Tabela 5.31. Resultados gerais de validação em 5 classes.

Pode-se visualizar pela tabela anterior que segundo RCP1, 17 músicas foram consideradas sem classificação. A maioria das restantes músicas foi classificada sem ambiguidade excepto um pequeno número que foi catalogada em mais do que um género, existindo mesmo por exemplo uma música classificada em 4 géneros: flauta, piano, violino e coral. A música em questão é a 32^a, *Allegro con brio – Piano Concerto N°2 in B Flat Major, Op. 19* de *Beethoven*. Esta peça tem 13:59 minutos e através da audição dos extractos verifica-se que apenas 2 mostram claramente o género correcto (piano). Além disso a música tem 13:59 minutos e só se considera para os extractos a música até aos 3 minutos. Uma solução óbvia para problemas destes parece estar em aumentar o número de extractos e fazer com que representem uma área maior da música original.

Os resultados da análise da tabela anterior vão ser sintetizados em seguida nas tabelas (5.32) para RCP1 e (5.33) para RCP2.

RCP1 57,3%	flauta	piano	violino	coral	ópera
flauta	59,2	3,9	0	18,2	0
piano	4,5	42,3	0	4,5	0
violino	13,6	7,7	85	0	9,1
coral	4,5	19,2	5	59,2	31,8
ópera	0	11,5	0	4,5	40,9
s/classificação	18,2	15,4	10	13,6	18,2

Tabela 5.32. Matriz de confusão do protótipo para a terceira classificação: RCP1.

Verifica-se pela tabela anterior que a percentagem de sucesso de músicas bem classificadas foi respectivamente para flauta, piano, violino, coral e ópera de 59,2%, 42,3%, 85%, 59,2% e 40,9%. A taxa geral de sucesso foi de 57,3%. A percentagem de músicas sem classificação foi de 15%.

RCP2 66,7%	flauta	piano	violino	coral	ópera
flauta	66,7	4,2	0	19,1	5
piano	4,8	50	0	4,7	0
violino	14,2	4,2	100	0	15
coral	9,5	29,1	0	66,7	30
ópera	4,8	12,5	0	9,5	50

Tabela 5.33. Matriz de confusão do protótipo para a terceira classificação: RCP2.

Através da tabela anterior, verifica-se que a percentagem de músicas bem classificadas foi de 66,7% para flauta, 50% para piano, 100% para violino, 66,7% para coral e 50% para ópera. A percentagem geral de sucesso foi de 66,7%.

De notar antes de mais a classificação conseguida pela classe violino. Segundo RCP2 todas as suas músicas foram correctamente classificadas. Em RCP1, existiram apenas 3 músicas que não foram classificadas em violino: duas sem classificação e uma classificada em coral. Conclui-se que o classificador aprendeu da melhor forma a identificar as características do instrumento violino.

Em oposição à classe violino, as classes piano e ópera obtiveram classificações que desiludiram. Através da inspecção à Tabela 5.33 repara-se que 29,1% das músicas de piano foram classificadas em coral e 30% das de ópera foram classificadas em coral. Esta confusão entre ópera e coral já tinha sido detectada no segundo problema de classificação (Secção 5.6.4). Por análise de alguns casos, detectou-se que partes atípicas da ópera, muito lentas e com interpretações parecidas com a música coral, são classificadas facilmente em coral. Isto está relacionado com o facto dos classificadores terem sido treinados essencialmente com casos típicos de ópera. Além disso a ópera não mantém normalmente durante toda a sua duração as características que a identificam facilmente. Uma solução para aumentar a fiabilidade da classificação pode passar numa primeira fase por aumentar o número de extractos de cada música. Outra possível solução será treinar a rede com um conjunto maior de exemplos de treino que contenha mais casos atípicos.

Quanto ao facto da confusão entre piano e coral, não se encontrou uma justificação razoável. No entanto há parecenças entre os casos típicos dos dois estilos no que toca à sonoridade. Ambos são bastante calmos.

Pode-se considerar que os resultados alcançados por esta abordagem a um sistema real de classificação são no mínimo interessantes numa perspectiva de melhoria futura, já que foi obtida uma taxa de sucesso de 66,7% relativamente próxima da taxa que tinha sido obtida para casos mais típicos de classificação (76%).

Capítulo 6

CONCLUSÕES

O objectivo principal deste estudo foi o de classificar música clássica em subgéneros. Para tal foram seleccionadas músicas de cinco subgéneros, das quais foram extraídas peças de 6s. Em seguida foram extraídas 40 características de cada uma das peças formando assim a sua assinatura. Finalmente foram utilizados três classificadores: KNN, GMM e MLP para efectuar a classificação.

Após a comparação dos resultados dos três classificadores, chegou-se à conclusão que o MLP originou os melhores resultados em todas as tarefas de classificação.

Este classificador foi então utilizado para fazer uma aproximação a um sistema de classificação automático de géneros musicais. Neste sistema, cada música foi representada por dez extractos escolhidos de igual forma para todas as músicas. Cada música foi classificada no género musical mais representado pelos seus extractos.

Vão ser expostas nas secções seguintes, as principais conclusões deste trabalho, bem como perspectivas para trabalho futuro

6.1 – CONCLUSÕES GERAIS

De uma forma geral, concluiu-se que este trabalho atingiu resultados muito promissores e que definiu mais um contributo para o estado de arte da área de investigação em reconhecimento automático de géneros musicais.

Baseado nos objectivos de classificação referidos e tendo em conta os resultados alcançados em trabalhos similares, foi dada especial importância a características relevantes para a análise do timbre e do *pitch* do sinal. Os resultados alcançados apontam para uma escolha acertada. No entanto, será importante no futuro fazer uma análise detalhada do espaço de características, no sentido de detectar e eliminar características que sejam redundantes, bem como definir e utilizar outras que possam ser importantes na discriminação dos casos mais atípicos.

Apresentou-se ainda neste trabalho uma metodologia e um estudo comparativo para a classificação dos géneros musicais considerados. Pode-se considerar que os resultados foram satisfatórios, com qualquer um dos três classificadores, para a discriminação de música instrumental e vocal. Já no problema mais complexo de discriminação em 5 géneros musicais, os resultados foram menos precisos, principalmente para os classificadores KNN e GMM, que não conseguiram resultados tão satisfatórios.

Analisando directamente os resultados dos classificadores, as redes neuronais (MLP) foram sempre superiores em todos os problemas de classificação ao KNN e ao GMM, apresentando resultados promissores em todas as situações. Neste último problema pretende-se utilizar um classificador hierárquico, que classifique primeiro em música instrumental e vocal e só depois num género.

Em relação ao protótipo de classificação, os resultados superaram as expectativas, porque ficaram bastante próximos dos resultados com exemplos de teste “bem comportados”, o que prova que o classificador foi treinado com exemplos bastante abrangentes de cada género musical. Estes resultados são essencialmente uma motivação para prosseguir na investigação nesta área.

6.2 – PERSPECTIVAS FUTURAS

À medida que se evolui na realização de um trabalho de investigação, vão surgindo normalmente novas ideias a seguir e perspectivas de evolução futura.

No trabalho prático que está na base desta dissertação aconteceu precisamente isso. Por conseguinte, o trabalho futuro a realizar passa por algumas ideias que são consequência lógica de tudo o que foi feito até agora e que já foram referidas na secção anterior, e outras que não sendo tão óbvias, buscam novas perspectivas a seguir.

Assim, as ideias a concretizar futuramente são:

- Efectuar a classificação hierárquica da taxonomia realizada neste estudo, i.e., classificar em música vocal ou instrumental, e em seguida implementar, mediante o resultado, as classificações da primeira ou da segunda tarefas de classificação.

- Detectar e eliminar características que sejam redundantes. Para tal, poderão ser usadas técnicas já muito conhecidas como a análise dos componentes principais (PCA)³⁴.

- Fazer a expansão da árvore em profundidade, com novas classificações intermédias, e em largura, com mais subgéneros da música clássica.

- Utilizar mais características, nomeadamente rítmicas mas não só, de forma a melhorar a visão sobre os problemas deste estudo e de forma a poderem ser considerados novos géneros que à partida se sabe terem uma componente rítmica forte associada (e.g., valsa).

- Utilizar outros classificadores que permitam aumentar a eficiência na classificação.

- Considerar outros géneros além da música clássica (e.g., *jazz*, *rock*).

³⁴ Em terminologia Inglesa: Principal Component Analysis - PCA

- Migrar o código efectuado em Matlab para uma linguagem universal como o C++. Esse código deverá ainda permitir a manipulação directa de ficheiros MP3, ao contrário dos WAV utilizados até agora. Juntar a isso um interface gráfico para a manipulação de todo o programa.

ANEXOS

ANEXO 1. MÚSICAS

As colunas da tabela seguinte apresentam da esquerda para a direita a numeração das músicas de treino e de validação, a classe a que pertencem (F-Flauta; P-Piano; V-Violino; C-Coral e O-Ópera), o nome das músicas e o seu compositor.

Nº	Classe	Nome	Compositor
Músicas de Treino			
1	F	Poco adagio - Sonata in A minor for solo flute	C. Bach
2	F	Largo - Concerto in G major	C. Bach
3	F	Allegro - Sonata in A minor for solo flute	C. Bach
4	F	Allegro - Concerto in D minor	C. Bach
5	F	Allegro (2) - Sonata in A minor for solo flute	C. Bach
6	F	Adagio - Concerto in B flat major	C. Bach
7	F	Allegro assai - Concerto in A minor	C. Bach
8	F	Andante - Concerto in A minor	C. Bach
9	F	Allegro assai (2) - Concerto in A minor	C. Bach
10	F	Allegretto - Concerto in B flat major	C. Bach
11	F	Allegro assai - Concerto in B flat major	C. Bach
12	F	Adagio - in D/D-dur/en ré majeur, K.285	Mozart
13	F	Allegretto - in D/D-dur/en ré majeur, K.285	Mozart
14	F	Allegro - in C/C-dur/en ut majeur K.Anh.171/K.285b	Mozart
15	F	Andantino - in C/C-dur/en ut majeur K.Anh.171/K.285b	Mozart
16	F	Variation V - in C/C-dur/en ut majeur K.Anh.171/K.285b	Mozart
17	F	Variation II - in A/A-dur/en la majeur, K.298	Mozart
18	F	Variation IV - in A/A-dur/en la majeur, K.298	Mozart
19	F	Rondeau - in A/A-dur/en la majeur, K.298	Mozart
20	F	Andantino - Concerto for Flute, Harp and Orchestra in C,K.299	Mozart
21	F	Rondo - Flute concerto in G,K.313	Mozart

22	F	Rondo - Concerto for Flute, Harp and Orchestra in C, K.299	Mozart
23	F	Variation III - in C/C-dur/en ut majeur K.Anh.171/K.285b	Mozart
24	F	Adagio non troppo - Flute concerto in G, K.313	Mozart
25	F	Allegro - Concerto for Flute and Orchestra N°2 in D major	F. Devienne
26	F	Adagio - Concerto for Flute and Orchestra N°2 in D major	F. Devienne
27	F	Allegretto - Concerto for Flute and Orchestra N°2 in D major	F. Devienne
28	F	Romance - Concerto for Flute and Orchestra N°4 in G major	F. Devienne
29	F	(1ª) - Concerto for recorder, strings and Harpsichord in C major, RV 443	Vivaldi
30	F	(2ª) - Concerto for recorder, strings and Harpsichord in C minor, RV 441	Vivaldi
31	F	(3ª) - Concerto for recorder, strings and Harpsichord in C major, RV 443	Vivaldi
32	F	(3ª) - Concerto for recorder, strings and Harpsichord in C minor, RV 441	Vivaldi
33	F	Largo - Concerto para flauta-doce, flauta-travessa, cordas e baixo contínuo em mi menor	G. Telemann
34	F	Allegro - Concerto para flauta-doce, flauta-travessa, cordas e baixo contínuo em mi menor	G. Telemann
35	F	Largo (2) - Concerto para flauta-doce, flauta-travessa, cordas e baixo contínuo em mi menor	G. Telemann
36	F	Presto - Concerto para flauta-doce, flauta-travessa, cordas e baixo contínuo em mi menor	G. Telemann
37	F	Allegro - Concerto in A major	C. Bach
38	F	Allegro assai - Concerto in A major	C. Bach
39	F	Allegro - Concerto in D minor	C. Bach
40	F	Allegro di molto - Concerto in D minor	C. Bach
41	P	Träumerai op. 15/7	Schumann
42	P	Für Elise	Beethoven
43	P	Sonho de amor n°3	F. Liszt
44	P	Moonlight piano sonata No.14 in C minor op.27	Beethoven
45	P	Allegro - Concerto for Piano and Orchestra N°11, in F major, K413	Mozart
46	P	Larghetto - Concerto for Piano and Orchestra N°11, in F major, K413	Mozart
47	P	Tempo di minuetto - Concerto for Piano and Orchestra N°11, in F major, K.413	Mozart
48	P	Allegro con spirito - Sonata for 2 Pianos, in D major, K.448	Mozart
49	P	Allegro con spirito (2) - Sonata for 2 Pianos, in D major, K.448	Mozart
50	P	Andante - Sonata for 2 Pianos, in D major, K.448	Mozart
51	P	Allegro molto - Sonata for 2 Pianos, in D major, K.448	Mozart
52	P	Allegro vivace - Concerto for Piano and Orchestra N°14, in E flat major, K.449	Mozart
53	P	Allegro ma non troppo - Concerto for Piano and Orchestra N°14, in E flat major, K.449	Mozart
54	P	Allegro - Piano concerto No.20 in D minor, K.466	Mozart
55	P	Allegro (2) - Piano concerto No.20 in D minor, K.466	Mozart
56	P	Romanze - Piano concerto No.20 in D minor, K.466	Mozart
57	P	Allegro maestoso - Piano concerto No.21 in C, K.467	Mozart
58	P	Andante - Piano concerto No.21 in C, K.467	Mozart
59	P	Allegro con brio - Piano concerto No.1 in C, op.15	Beethoven
60	P	Largo - Piano concerto No.1 in C, op.15	Beethoven
61	P	Rondo: Allegro scherzando - Piano concerto No.1 in C, op.15	Beethoven
62	P	Allegro con brio - Piano concerto No.2 in B flat major, op.19	Beethoven
63	P	Adagio - Piano concerto No.2 in B flat major, op.19	Beethoven
64	P	Rondo: Molto allegro - Piano concerto No.2 in B flat major, op.19	Beethoven
65	P	Rondo: Molto allegro - Piano concerto No.2 in B flat major, op.19	Beethoven
66	P	Allegro - Piano concerto No.5 in E flat major, op.73 "Emperor"	Beethoven
67	P	Adagio um poco mosso (2) - Piano concerto No.5 in E flat major, op.73 "Emperor"	Beethoven
68	P	Rondo: Allegro - Piano concerto No.5 in E flat major, op.73 "Emperor"	Beethoven
69	P	Largo - Piano concerto No.3 in C minor, op.37	Beethoven
70	P	Allegro con brio - Piano concerto No.3 in C minor, op.37	Beethoven
71	P	Andante con moto - Piano concerto No.4 in G major, op.58	Beethoven
72	P	Rondo: Allegro - Piano concerto No.3 in C minor, op.37	Beethoven
73	P	Rondo: Allegro (2) - Piano concerto No.3 in C minor, op.37	Beethoven

74	P	Allegro moderato - Piano concerto No.4 in G major, op.58	Beethoven
75	P	Rondo: Vivace - Piano concerto No.4 in G major, op.58	Beethoven
76	P	Prelude 04	Chopin
77	P	Prelude 07	Chopin
78	P	Prelude 20	Chopin
79	P	Prelude 02	Chopin
80	P	Prelude 04 (2)	Chopin
81	V	No.1 in E/E-dur/en mi majeur	Pagannini
82	V	No.2 in B minor/h-moll/en si mineur	Pagannini
83	V	No.4 in C minor/c-moll/en ut mineur	Pagannini
84	V	No.5 in A minor/a-moll/en la mineur	Pagannini
85	V	No.7 in A minor/a-moll/en la mineur	Pagannini
86	V	No.8 in E flat/Es-dur/en mi bémol majeur	Pagannini
87	V	No.10 in G minor/g-moll/en sol mineur	Pagannini
88	V	No.11 in C/C-dur/en ut majeur	Pagannini
89	V	No.13 in B flat/B-dur/en si bémol majeur	Pagannini
90	V	No.14 in E flat/Es-dur/en mi bémol majeur	Pagannini
91	V	No.16 in G minor/g-moll/en sol mineur	Pagannini
92	V	No.17 in E flat/Es-dur/en mi bémol majeur	Pagannini
93	V	No.19 in E flat/Es-dur/en mi bémol majeur	Pagannini
94	V	No.20 in D/D-dur/en re majeur	Pagannini
95	V	Adagio - Sonata No.1 in G minor, BWV 1001	J. Bach
96	V	Fuga: Allegro - Sonata No.1 in G minor, BWV 1001	J. Bach
97	V	Allemanda - Partita No.1 in B minor, BWV 1002	J. Bach
98	V	Double - Partita No.1 in B minor, BWV 1002	J. Bach
99	V	Double (2) - Partita No.1 in B minor, BWV 1002	J. Bach
100	V	Grave - Sonata No.2 in A minor, BWV 1003	J. Bach
101	V	Allegro - Sonata No.2 in A minor, BWV 1003	J. Bach
102	V	Allemanda - Partita No.2 in D minor, BWV 1004	J. Bach
103	V	Giga- Partita No.2 in D minor, BWV 1004	J. Bach
104	V	Ciaconna - Partita No.2 in D minor, BWV 1004	J. Bach
105	V	Allegro assai - Sonata No.3 in C major, BWV 1005	J. Bach
106	V	Preludio - Partita No.3 in E major, BWV 1006	J. Bach
107	V	Allegro - Violin concerto No.4 in D, K.218	Mozart
108	V	Andante cantabile - Violin concerto No.4 in D, K.218	Mozart
109	V	Allegro aperto - Violin concerto No.5 in A, K.219	Mozart
110	V	Adagio - Violin concerto No.5 in A, K.219	Mozart
111	V	Adagio (2) - Violin concerto No.5 in A, K.219	Mozart
112	V	Rondo concertante in B flat, K.269	Mozart
113	V	Adagio in E, K.261	Mozart
114	V	Allegro - Concerto in Mi maggiore RV269 "La primavera"	Vivaldi
115	V	Danza pastorale: allegro - Concerto in Mi maggiore RV269 "La primavera"	Vivaldi
116	V	Allegro non molto - Concerto in sol minore RV315 "L'estate"	Vivaldi
117	V	Allegro - Concerto in Fa maggiore RV293 "L'autunno"	Vivaldi
118	V	Allegro (2) - Concerto in Fa maggiore RV293 "L'autunno"	Vivaldi
119	V	Remance para violino op.50 reprise	Beethoven
120	V	Loure - Partita No.3 in E major, BWV 1006	J. Bach
121	C	Agnus - Missa Pange Lingua	J. Desprez
122	C	Alleluia - Missa Pange Lingua	J. Desprez
123	C	Hymne "Pange Lingua" - Missa Pange Lingua	J. Desprez
124	C	Sanctus - Missa Pange Lingua	J. Desprez
125	C	Venite, Benedicti - In Dulci Jubilo	Ambrosian Chant
126	C	Haec dies. Confitemini - In Dulci Jubilo	Ambrosian Chant
127	C	Videns dominus sororem - In Dulci Jubilo	Ambrosian Chant
128	C	Crastina die - Coros Amadores da Região de Coimbra	CARC
129	C	In monte olivetti - Coros Amadores da Região de Coimbra	CARC

130	C	Auprè de toi - Coro dos professores de Coimbra	Coros
131	C	Boina, boina - Coral de Letras da Universidade de Coimbra	Coros
132	C	Meditação - Grupo vocal Ad libitum	Coros
133	C	In memoriam - Orfeon Académico de Coimbra	Coros
134	C	Kyrie eleison - Coro dos pequenos cantores de Coimbra	Coros
135	C	Não choro por me deixares - Coro misto da Universidade de Coimbra	Coros
136	C	Balada para Coimbra - Antigos orfeonistas da Universidade de Coimbra	Coros
137	C	Introitus: in medio ecclesie - Mass for St. Anthony of Padua	G. Du Fay
138	C	Kyrie - Mass for St. Anthony of Padua	G. Du Fay
139	C	Graduale: os iusti - Mass for St. Anthony of Padua	G. Du Fay
140	C	Alleluia: Antoni compare inclite - Mass for St. Anthony of Padua	G. Du Fay
141	C	Communio: domine, quinque talenta - Mass for St. Anthony of Padua	G. Du Fay
142	C	Hymnus: veni creator spiritus - Mass for St. Anthony of Padua	G. Du Fay
143	C	Pie Jesu - Requiem, op.48	G. Fauré
144	C	Agnus dei - Requiem, op.48	G. Fauré
145	C	Ave Maria - D 839	Schubert
146	C	Kyrie - Mass in C minor, K.427	Mozart
147	C	Psalm 42	Mendelssohn
148	C	Agnus dei	S. Barber
149	C	Agnus dei (2)	S. Barber
150	C	Ave Maria – Vespers, op.37	Rachmaninov
151	C	Tantum ergo, op.10 No.2	M. Duruflé
152	C	Laudamus te - Gloria	F. Poulenc
153	C	Beatus vir à 6 - Selva morale e spirituale	C. Monteverdi
154	C	Spem in alium, motet à 40	T. Tallis
155	C	Jesus, Bleibet meine Freude - Cantata BWV 147	J. Bach
156	C	Jauchzet, frohlocket, auf, preiset die tage	J. Bach
157	C	Euntes ibant et flebant - In convertendo	J. Rameau
158	C	Fortuna Emperatrix Mundi	C. Orff
159	C	Allegro assai - Sinfonia No.9 en re menor, op.125, “Coral”	Beethoven
160	C	Andante: In Terra Pax – Gloria RV.589	Vivaldi
161	O	Signore, ascolta - Turandot	G. Puccini
162	O	Come per me sereno - La sonnambula	V. Bellini
163	O	Poveri fiori - Adriana Lecouvreur	F. Cilea
164	O	La mamma morta - Andrea chenier	U. Giordano
165	O	O mio babbino caro - Gianni schicchi	G. Puccini
166	O	Qual fiamma ave anel guardo - I Pagliacci	Leoncavallo
167	O	Ebben ne andro lontana - La wally	A. Catalani
168	O	Al vostri giochi – Amleto	A. Thomas
169	O	Granada	A. Lara
170	O	Por al humo - D. Francisquita	A. Vives
171	O	Tu, tu, amore? Tu? – Manon lescaut	G. Puccini
172	O	O incantesimo! Parla! Parla! - Mefistofele	Boito
173	O	Core' ngrato - La fanciulla del west	G. Puccini
174	O	Ay, ay, ay - La fanciulla del west	G. Puccini
175	O	Cruda sorte! Amor tiranno - L' Italiana in Argeri	Rossini
176	O	Poema en forma de canciones. Los dos miedos	J. Turina
177	O	Tus ojillos negros	M. Falla
178	O	Canción de Cuna	Montsalvatge
179	O	Un giorno di regno	Verdi
180	O	I Lombardi	Verdi
181	O	I Due Foscari	Verdi
182	O	Giovanna D'Arco	Verdi
183	O	Attila	Verdi
184	O	Macbeth	Verdi
185	O	Jérusalem	Verdi
186	O	Il Corsaro	Verdi
187	O	Stiffelio	Verdi
188	O	Rigoletto	Verdi
189	O	La Traviatta	Verdi

190	O	La Traviatta (2)	Verdi
191	O	Les Vêpres Siciliennes	Verdi
192	O	Aroldo	Verdi
193	O	La Forza del Destino	Verdi
194	O	Don Carlos	Verdi
195	O	Falstaff	Verdi
196	O	Rigoletto (2)	Verdi
197	O	Una voce poco fa – II Barbiere di Siviglia	Rossini
198	O	La Traviatta (4)	Verdi
199	O	Spargi d’amor pianto – Lucia di Lammermoor	G. Donizetti
200	O	Casta Diva – Norma	V. Bellini

Músicas de Validação

1	F	Poco adagio - Sonata in A minor for solo flute	C. Bach
2	F	Allegro - Sonata in A minor for solo flute	C. Bach
3	F	Allegro (2) - Sonata in A minor for solo flute	C. Bach
4	F	Allegro assai - Concerto in A minor	C. Bach
5	F	Adagio - Concerto in B flat major	C. Bach
6	F	Allegro – in D/D-dur/en ré majeur, K.285	Mozart
7	F	Tempo di Menuetto - in G/G-dur/en sol majeur, K.285a	Mozart
8	F	Variation I - in A/A-dur/en la majeur, K.298	Mozart
9	F	Adagio - Concerto for flute and orchestra N°8 in G major	F. Devienne
10	F	Allegro – Concerto for Flute, Harp and Orchestra in C,K.299	Mozart
11	F	Andante – For Flute and Orchestra in C,K.315	Mozart
12	F	Allegro maestoso - Flute concerto in G,K.313	Mozart
13	F	Allegro - Concerto for Flute and Orchestra N°4 in G major	F. Devienne
14	F	Rondo: Moderato - Concerto for Flute and Orchestra N°4 in G major	F. Devienne
15	F	(2º) - Concerto for recorder, strings and Harpsichord in C major, RV 443	A. Vivaldi
16	F	(1º) - Concerto for recorder, strings and Harpsichord in C minor, RV 441	A. Vivaldi
17	F	Largo – Concerto para flauta-doce, flauta-travessa, cordas e baixo contínuo em mi menor	G. Telemann
18	F	Presto - Concerto in G major	C. Bach
19	F	Allegro di molto - Concerto in G major	C. Bach
20	F	Un poco andante - Concerto in D minor	C. Bach
21	F	Nocturno op.9/2	Chopin
22	P	Andantino - Concerto for Piano and Orchestra N°14, in E flat major, K.449	Mozart
23	P	Larghetto (2) - Concerto for Piano and Orchestra N°11, in F major, K413	Mozart
24	P	Tempo di minuetto (2) - Concerto for Piano and Orchestra N°11, in F major, K.413	Mozart
25	P	Andante (2) - Sonata for 2 Pianos, in D major, K.448	Mozart
26	P	Allegro molto (2) - Sonata for 2 Pianos, in D major, K.448	Mozart
27	P	Allegro vivace (2) - Concerto for Piano and Orchestra N°14, in E flat major, K.449	Mozart
28	P	Rondo - Piano concerto No.20 in D minor, K.466	Mozart
29	P	Allegro vivace assai - Piano concerto No.21 in C, K.467	Mozart
30	P	Allegro con brio (2) - Piano concerto No.1 in C, op.15	Beethoven
31	P	Rondo in B flat Wo06 - Piano concerto No.1 in C, op.15	Beethoven
32	P	Allegro con brio - Piano concerto No.2 in B flat major, op.19	Beethoven
33	P	Allegro - Piano concerto No.5 in E flat major, op.73 “Emperor”	Beethoven
34	P	Adagio um poco mosso - Piano concerto No.5 in E flat major, op.73	Beethoven
35	P	Rondo: Allegro (2) - Piano concerto No.5 in E flat major, op.73	Beethoven
36	P	Allegro con brio - Piano concerto No.3 in C minor, op.37	Beethoven
37	P	Allegro moderato - Piano concerto No.4 in G major, op.58	Beethoven
38	P	Prelude 06	Chopin
39	P	Minute Waltz	Chopin
40	P	Prelude 06	Chopin
41	V	No.3 in E minor/e-moll/en mi mineur	Pagannini

42	V	No.6 in G minor/g-moll/en sol mineur	Pagannini
43	V	No.9 in E/E-dur/en mi majeur	Pagannini
44	V	No.12 in A flat/As-dur/en la bémol majeur	Pagannini
45	V	No.15 in E minor/e-moll/en mi mineur	Pagannini
46	V	No.18 in C/C-dur/en ut majeur	Pagannini
47	V	Siciliana - Sonata No.1 in G minor, BWV 1001	J. Bach
48	V	Correnta - Partita No.1 in B minor, BWV 1002	J. Bach
49	V	Fuga - Sonata No.2 in A minor, BWV 1003	J. Bach
50	V	Corrente - Partita No.2 in D minor, BWV 1004	J. Bach
51	V	Fuga - Sonata No.3 in C major, BWV 1005	J. Bach
52	V	Minuett II - Partita No.3 in E major, BWV 1006	J. Bach
53	V	Rondeau - Violin concerto No.4 in D, K.218	Mozart
54	V	Rondeau - Violin concerto No.5 in A, K.219	Mozart
55	V	Rondo in C, K.373	Mozart
56	V	Largo e pianissimo sempre - Concerto in Mi maggiore RV269 "La primavera"	Vivaldi
57	V	Adagio presto - Concerto in sol minore RV315 "L'estate"	Vivaldi
58	V	Allegro non molto - Concerto in Fa minor RV297 "L'inverno"	Vivaldi
59	V	Gavotte en rondeau - Partita No.3 in E major, BWV 1006	J. Bach
60	V	Gigue - Partita No.3 in E major, BWV 1006	J. Bach
61	C	Credo - Missa Pange Lingua	J. Desprez
62	C	Kyrie eleison - In Dulci Jubilo	Ambrosian Chant
63	C	Levavi oculos meos - In Dulci Jubilo	Ambrosian Chant
64	C	Romance a 4 à Rainha Santa Isabel - Coros Amadores da Região de C	CARC
65	C	Magnificat tertii toni - Coros Amadores da Região de Coimbra	G. Du Fay
66	C	Oh meu menino jesus - Coro D. Pedro de Cristo	G. Du Fay
67	C	Canção de embalar - Grupo coral de Santa Cruz	G. Du Fay
68	C	Chanson d'amitié - Coro dos professores de Coimbra	G. Du Fay
69	C	Gloria - Mass for St. Anthony of Padua	G. Du Fay
70	C	Offertorium: veritas mea - Mass for St. Anthony of Padua	G. Du Fay
71	C	Sanctus - Requiem, op.48	G. Fauré
72	C	Agnus dei - Requiem	J. Catoire
73	C	Ave verum corpus - Motet K.618	Mozart
74	C	Ein Deutsches Requiem	J. Brahms
75	C	Ubi caritas, op.10 No.1	M. Duruflé
76	C	Miserere à 9	G. Allegri
77	C	Hallelujah - Messiah	G. Handel
78	C	Omnes gentes, motet à 16	G. Gabrieli
79	C	Organum: Alleluia V.video celos apertos	F.Anon
80	C	Antiphonae - In Dulci Jubilo	Ambrosian Chant
81	O	Siciliana - I vespri siciliani	Verdi
82	O	Ombra leggera - Il perdono di ploermel	J. Meyerbeer
83	O	Pace pace mio dio - La forza del destino	Verdi
84	O	Non mi lasciara, o speme - Guilherme Tell	Rossini
85	O	Jeveux vivre dans le rêve - Romeu e Julieta	C. Gounod
86	O	Ch'ella mi creda libero i lontano - La fanciulla del west	G. Puccini
87	O	Piangerò la sorte mia - Júlio César	Haendel
88	O	Nacqui all' affanno... Non più mesta - La cenerentola	Rossini
89	O	Allí está, riyendo - De la vida breve	M. Falla
90	O	Ernani	Verdi
91	O	Alzira	Verdi
92	O	I Masnadieri	Verdi
93	O	Luisa Miller	Verdi
94	O	Il Trovatore	Verdi
95	O	La Traviatta (3)	Verdi
96	O	Un Ballo in Maschera	Verdi
97	O	Aida	Verdi

98	O	Un bel di vedremo - Madame Butterfly	G. Puccini
99	O	Si, mi chiamano mimi – La Bohème	G. Puccini
100	O	Vissi d'arte - Tosca	G. Puccini

BIBLIOGRAFIA

[Agostini et al., 2003]

Agostini, G., Longari, M. & Pollastri, E., 2003, “Musical instrument timbres classification with spectral features”, *EURASIP Journal on Applied Signal Processing*, 2003:1, 1-11.

[Alsabti et al., 1998]

Alsabti, K., Ranka, S. & Singh, V., 1998, “An efficient K-Means clustering algorithm”, *11th International Parallel Processing Symposium - IPPS 1998*.

[Berenzweig & Ellis, 2001]

Berenzweig, A. & Ellis, D., 2001, “Locating singing voice segments within music signals”, *Proceedings of Workshop on Applications of Signal Processing to Audio and Acoustics WASPAA 2001*, pp. 119-123, Mohonk, NY.

[Bishop, 1995]

Bishop, C.M., 1995, “*Neural Networks for Pattern Recognition*”, Oxford University Press.

[Bradley et al., 1998]

Bradley, P., Fayyad, U. & Reina, C., 1998, "Scaling clustering algorithms to large databases", *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining - AAAI 1998*, pp. 9-15, NY.

[Broomhead & Lowe, 1988]

Broomhead D.S. & Lowe D., 1988, "Multivariable function interpolation and adaptativo networks", *Complex Systems*, Vol. 2, pp 321-355.

[Cano, 1998]

Cano, P., 1998, "Fundamental frequency estimation in the SMS analysis", *Proceedings of the Cost G-6 Workshop on Digital Audio Effects - DAFX 1998*, Barcelona.

[Cooley & Tukey, 1965]

Cooley, J.W. & Tukey, J.W., 1965, "An Algorithm for the machine calculation of complex Fourier Series", *Mathematics Computation*, Vol.19, pp 297-301.

[Dempster et al., 1997]

Dempster, A., Laird, N. & Rubin, D., 1977, "Maximum likelihood from incomplete data via EM algorithm", *Journal of the Royal Statistical Society, Series B*, 39:138.

[Demuth & Beale, 2001]

Demuth, H. & Beale, M., 2001, "*Neural Network Toolbox User's Guide*", version 4, Mathworks.

[Duda et al., 2000]

Duda, R., Hart, P. & Stork, D., 2000, "*Pattern Classification*", Wiley & Sons, NY.

[Ellis, 1987]

Ellis, D., 1987, "*A perceptual representation of audio*", MSc Thesis, MIT.

[Elman, 1990]

Elman, J. L., 1990, "Finding structure in time," *Cognitive Science*, vol. 14, pp. 179-211.

[Foote, 1997 (1)]

Foote, J., 1997, "A similarity measure for automatic audio classification", *Proceedings of American Association for Artificial Intelligence, AAAI 1997*.

[Foote, 1997 (2)]

Foote, J., 1997, "Content-based retrieval of music and audio", *Proceedings of the International Society for Optical Engineering, SPIE 1997*, vol. 3229, pp.138-147.

[Foote, 1998]

Foote, J., 1998, "An overview of audio information retrieval", *ACM Multimedia Systems*, vol. 7, pp. 2-11.

[Foote, 1999]

Foote, J., 1999, "*Methods for the automatic analysis of music and audio*", Xerox Park Technical Report FXPAL –TR-99-038.

[Foote, 2000]

Foote, J., 2000, "Arthur: Retrieving orchestral music by long-term structure", *1st International Symposium on Music Information Retrieval - ISMIR 2000*.

[Fraley & Raftery, 1998]

Fraley, C. & Raftery A., 1998, "*How many clusters? Which clustering method? Answers via model-based cluster analysis*", Technical Report No 329, Department of Statistics, University of Washington.

[Fraser & Fujinaga, 1999]

Fraser, A. & Fujinaga, I., 1999, "Toward real-time recognition of acoustic musical instruments", *Proceedings of the International Computer Music Conference - ICMC 1999*.

[Frühwirth & Rauber, 2001]

Frühwirth, M. & Rauber, A., 2001, "Self-organizing maps for content-based music clustering", *Proceedings of 12th Workshop Italian Workshop on Neural Nets - WIRN 2001*.

[Fujinaga, 1998]

Fujinaga, I., 1998, "Machine recognition of timbre using steady-state tone of acoustic musical instruments", *Proceedings of the International Computer Music Conference - ICMC 1998*.

[Funahashi, 1989]

Funahashi, K., 1989, "On the approximate realization of continuous mappings by neural networks", *Neural Networks*, Vol. 2, pp. 183-192.

[Gerhard, 1997]

Gerhard, D., 1997, "Computer music analysis", Technical Report CMPT TR 97-13 Simon Fraser University School of Computing Science.

[Gerhard, 2000 (1)]

Gerhard, D., 2000, "Audio signal classification", Ph.D. Depth Paper, Simon Fraser University.

[Gerhard, 2000 (2)]

Gerhard, D., 2000, "Audio signal classification: An overview", *Canadian Artificial Intelligence - CAI 2000*, Ann Grbavec (ed.), *Canadian Soc. for Computational Studies of Intelligence*. 45: 4-6.

[Gibbs, 1997]

Gibbs M., 1997, "Bayesian gaussian processes for regression and classification", PhD thesis, Cambridge University.

[Golub, 2000]

Golub, S., 2000, "Classifying Recorded Music", MSc Thesis, University of Edinburgh.

[Li & Khokhar, 2000]

Li, G. & Khokhar, A., 2000, "Content-based indexing and retrieval of audio data using wavelets", *Proceedings of IEEE International Conference on Multimedia and Expo (II) - ICME 2*.

[Hagan & Menhaj, 1994]

Hagan, M. & Menhaj, M., 1994, "Training Feedforward Networks with the Marquardt Algorithm", *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989-993.

[Hagan et al., 1996]

Hagan, M., Demuth, H. & Beale, M., 1996, "*Neural network design*", PWS Publishing, Boston, MA, 1996.

[Han et al., 1998]

Han, K., Pank, Y., Jeon, S., Lee, G. & Ha, Y., 1998, "Genre classification system on TV sound signals based on a spectrogram analysis", *IEEE Transactions on Consumer Electronics*, 55(1):33-42.

[Haykin, 1994]

Haykin S., 1994, "*Neural Networks: A Comprehensive Foundation*", Macmillan College Publishing.

[Heittola & Klapuri, 2002]

Heittola, T. & Klapuri, A., 2002, "Locating segments with drums in music signals", *Proceedings of 3rd International Conference on Music Information Retrieval - ISMIR 2002*.

[Hopfield, 1982]

Hopfield, J. J., 1982, "Neural networks and physical systems with emergent collective computational abilities", *National Academy of Sciences*, Vol. 79, pp. 2554-2558.

[Ichihashi et al., 2000]

Ichihashi, H., Honda, K. & Tani, N., 2000, "Gaussian mixture PDF approximation and Fuzzy c-Means clustering with entropy regularization", *Proceedings of the 4th Asian Fuzzy Systems Symposium - AFSS 2000*.

[Jensenius, 2002]

Jensenius, A., 2002, "*How do we recognize a song in one second?*", PhD Thesis, University of Oslo.

[Kohonen, 1989]

Kohonen T., 1989, "*Self-Organization and Associative Memory*", 3rd edition, Springer-Verlag, Berlin.

[Kosina, 2002]

Kosina, K., 2002, "*Music Genre Recognition*", MSc Thesis, Hagenberg.

[Lefèvre et al., 2002]

Lefèvre, S., Maillard, B. & Vincent, N., 2002, "A two level classifier process for audio segmentation", *RFAI Publication: IAPR 2002*, pp.891-894.

[Li, 2000]

Li, S., 2000, "Content-based classification and retrieval of audio using the nearest feature line method", *Proceedings of IEEE International Conference on Multimedia and Expo (II) - ICME 2*.

[Li et al., 1989]

Li, J., A. N. Michel, & W. Porod, 1989, "Analysis and synthesis of a class of neural networks: linear systems operating on a closed hypercube", *IEEE Transactions on Circuits and Systems*, vol. 36, no. 11, pp. 1405-1422.

[Liu & Wan, 2001]

Liu, M. & Wan, C., 2001, "A study on content-based classification and retrieval of audio database", *Proceedings of International Database Engineering and Applications Symposium - IDEAS 2001*.

[Liu et al., 2002]

Liu, M., Chang, E. & Dai, B., 2002, "Hierarchical gaussian mixture model for speaker verification", *Proceedings of 7th International Conference on Spoken Language Processing - ICSLP 2002*, pp. 1353-1356.

[Logan & Salomon, 2001]

Logan, B. & Salomon, A., 2001, "A music similarity function based on signal analysis", *Proceedings of IEEE International Conference on Multimedia and Expo - ICME 2001*.

[Lu & Hankinson, 2001]

Lu, L. & Hankinson, T., 2001, "A technique towards automatic audio classification and retrieval", *Proceedings of ACM Multimedia - ACM MM 2001*.

[Lu et al., 2001]

Lu, L., Jiang, H. & Zhang H., 2001, "A robust audio classification and segmentation method", *Proceedings of ACM Multimedia - ACM MM 2001*.

[Marques & Moreno, 1999]

Marques, J. & Moreno, P., 1999, "A study of musical instrument classification using Gaussian Mixture Models and Support Vector Machines", Technical Report, CRL 99/4, Cambridge Research Laboratory.

[Martin, 1998]

Martin, K., 1998, "Toward Automatic Sound Source Recognition: Identifying Musical Instruments", *NATO Computational Hearing Advanced Study Institute*, Il Ciocco, Italy.

[Martin, 1999]

Martin, K., 1999, "Sound-source recognition: A theory and computational model", PhD. Thesis, MIT.

[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 - ASA 1998*.

[Martin et al., 1998]

Martin, K. D., Scheirer, E. D. & Vercoe, B. L., 1998, "Musical content analysis through models of audition", *ACM Multimedia Workshop on Content-Based Processing of Music*.

[Matityaho & Furst, 1995]

Matityaho, B. & Furst, M., 1995, "Neural network based model for classification of music type", *Electrical and Electronic Engineers in Israel*, pp. 1-5.

[Modha & Spangler, 2002]

Modha, D. & Spangler, S., 2002, "Feature weighting in K-Means clustering", *Machine Learning*, vol. 47, 2002.

[Oppenheim & Willsky, 1997]

Oppenheim, A. & Willsky, A., 1997, "Signals and Systems", 2nd Edition, Prentice-Hall.

[Pachet & Cazaly, 2000]

Pachet, F. & Cazaly, D., 2000, "A taxonomy of musical genres", *Proceedings of Content-Based Multimedia Information Access - RIA O 2000*.

[Pachet et al., 2000]

Pachet, F., Roy, P. & Cazaly, D., 2000, "A combinatorial approach to content-based music selection", *IEEE Multimedia*, vol. 7(1), pp. 44-51.

[Paiva, 1999]

Paiva, R. P., 1999, "*Identificação Neuro-Difusa – Aspectos de Interpretabilidade*", MSc Thesis, Universidade de Coimbra, em Português.

[Perrot e Gjerdigen, 1999]

Perrot D. & Gjerdigen R., 1999, "Scanning the dial: An exploration of factors in identification of musical style", *Society for Music Perception and Cognition*, pp. 88, 1999.

[Pfeiffer, 1998]

Pfeiffer, S., 1998, "*The importance of perceptive adaptation of sound features in audio content processing*", Technical Report, Universitat Mannheim.

[Pfeiffer et al., 1996]

Pfeiffer, S., Fischer, S. & Effelsberg, W., 1996, "*Automatic audio content analysis*", Technical Report, Universitat Mannheim.

[Polikar, 2003]

Polikar, R., 2003, "*The Wavelet Tutorial*", <http://engineering.rowan.edu/~polikar/WAVELETS/WTtutorial.html>, disponível em Julho de 2003.

[Pye, 2000]

Pye, D., 2000, "Content-based methods for the management of digital music", *IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP 2000*, pp. 2437-2440.

[Rauber & Frühwirth, 2001]

Rauber, A. & Frühwirth, M., 2001, "Automatically analyzing and organizing music archives", *Proceedings of European Conference on Research and Advanced Technology for Digital Libraries - ECDL 2001*.

[Rossignol et al., 1998]

Rossignol, S., Rodet, X., Soumagne, J., Collette, J. & Depalle, P., 1998 "Features extraction and temporal segmentation of acoustic signals", *Proceedings of International Computer Music Conference - ICMC 1998*.

[Sarle, 2001]

Sarle W. (maintainer), 2001, "Neural Nets FAQ", <ftp://ftp.sas.com/pub/neural/FAQ3.html>.

[Scheirer, 1996]

Scheirer, E., 1996, "Bregman's Chimerae: Music perception as auditory scene analysis", *Proceedings of 4th International Conference Music Perception and Cognition - ICMPC 1996*.

[Scheirer, 1998]

Scheirer, E., 1998, "A Music perception systems", A proposal for a PhD dissertation, MIT.

[Scheirer, 2000]

Scheirer, E., 2000, "Music-listening systems", PhD Thesis, MIT.

[Scheirer & Slaney, 1997]

Scheirer, E. & Slaney, M., 1997, "Construction and evaluation of a robust multifeature speech/music discriminator", *Proceedings of IEEE International Conference Acoustics Speech and Signal Processing - ICASS 1997*.

[Smith, 1997]

Smith, S, 1997, "*The Scientist and Engineer's Guide to Digital Signal Processing*", California Technical Publishing.

[Soltau et al., 1998]

Soltau, H., Schultz, T., Westphal, M. & Waibel, A., 1998, "Recognition of music types", *IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP 1998*, pp. 1137-1140.

[Streit & Luginbuhl, 1994]

Streit R.L. & Luginbuhl T.E, 1994, "Maximum likelihood training of probabilistic neural networks", *IEEE Transactions on Neural Networks*, vol. 5, pp. 764-783.

[Subramanya et al., 1999]

Subramanya, S., Youssef, A., Narahari, B. & Simha, R., 1999, "Automated classification of audio data and retrieval based on audio classes", *Proceedings of International Conference on Computers and their Applications*.

[Tzanetakis, 2002]

Tzanetakis G., 2002, "*Manipulation, analysis and retrieval systems for audio signals*", PhD Thesis, University of Princeton.

[Tzanetakis & Cook, 2000 (1)]

Tzanetakis G. & Cook P., 2000, "Audio information retrieval tools", *Proceedings of International Symposium on Music Information Retrieval - ISMIR 2000*.

[Tzanetakis & Cook, 2000 (2)]

Tzanetakis G. & Cook P., 2000, "MARSYAS: A framework for audio analysis", *Organised Sound 4(3)*, Cambridge University Press.

[Tzanetakis & Cook, 2000 (3)]

Tzanetakis G. & Cook P., 2000, "Sound analysis using mpeg compressed audio", *Proceedings of International Conference Audio, Speech and Signal Processing - ICASSP 2000*.

[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.

[Tzanetakis et al., 2001 (1)]

Tzanetakis, G., Essl, G. & Cook, P., 2001, "Automatic Musical Genre Classification of Audio Signals", *Proceedings of International Symposium on Music Information Retrieval - ISMIR 2001*.

[Tzanetakis et al., 2001 (2)]

Tzanetakis, G., Essl, G. & Cook, P., 2001, "Audio analysis using the Discrete Wavelet Transform", *Proceedings of WSES International Conference Acoustics and Music: Theory and Applications - AMTA 2001*.

[Whitman & Smaragdis, 2002]

Whitman, B. & Smaragdis, P., 2002, "Combining musical and cultural features for intelligent style detection", *Proceedings of International Symposium on Music Information Retrieval - ISMIR 2002*.

[Williams, 2002]

Williams, C., 2002, "Probabilistic modelling and reasoning: Assignment 2 mixture models, PCA and ICA", Technical Report, University of Edinburgh.

[Welsh et al., 1999]

Welsh, M., Borisov, N., Hill, J., Behren, R. & Woo, A., 1999, "Querying large collections of music for similarity", Technical Report, University of California.

[Wold et al., 1996]

Wold, E., Blum, T., Keislar, T. & Wheaton, J., 1996, "Content-based classification, search and retrieval of audio", *IEEE Multimedia*, Vol. 3, n° 2.

[Wold et al., 1999]

Wold, E., Blum, T., Keislar, T. & Wheaton, J., 1999, "Classification, search and retrieval of audio", *CRC Handbook of Multimedia Computing 1999*.

[Xu et al., 2003]

Xu, C., Maddage, C., Shao, X., Cao, F. & Tian, Q., 2003, "Musical genre classification using support vector machines", *IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP 2003*.

[Yang, 2001]

Yang, C., 2001, "Music database retrieval based on spectral similarity", Technical Report 2001-14, Stanford University.

[Zhang & Kuo, 1998]

Zhang, T. & Kuo, C., 1998, "Content-based classification and retrieval of audio", *Proceedings of the International Society for Optical Engineering - SPIE 1998*.