

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Orquestração Musical Usando um Sistema Imunitário Artificial

José Miguel Lima de Abreu



FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Orientador: Rui Penha

Coorientador: Marcelo Caetano

27 de Outubro de 2015

A Dissertação intitulada

“Orquestração Musical Usando um Sistema Imunitário Artificial”

foi aprovada em provas realizadas em 16-10-2015

o júri

Aníbal João de Sousa Ferreira

Presidente Professor Doutor Aníbal João de Sousa Ferreira
Professor Associado do Departamento de Engenharia Eletrotécnica e de
Computadores da Faculdade de Engenharia da Universidade do Porto

Fernando Amílcar Bandeira Cardoso

Professor Doutor Fernando Amílcar Bandeira Cardoso
Professor Catedrático do Departamento de Engenharia Informática da Faculdade de
Ciências e Tecnologia da Universidade de Coimbra

Rui Luis Nogueira Penha

Professor Doutor Rui Luis Nogueira Penha
Professor Auxiliar Convidado do Departamento de Engenharia Informática da
Faculdade de Engenharia da Universidade do Porto

O autor declara que a presente dissertação (ou relatório de projeto) é da sua exclusiva autoria e foi escrita sem qualquer apoio externo não explicitamente autorizado. Os resultados, ideias, parágrafos, ou outros extratos tomados de ou inspirados em trabalhos de outros autores, e demais referências bibliográficas usadas, são corretamente citados.

José Miguel Lima de Abreu

Autor - José Miguel Lima de Abreu

Resumo

Neste trabalho pretendeu-se implementar e avaliar o uso de algoritmos inspirados por processos biológicos na resolução de problemas de orquestração musical. Neste contexto, define-se a orquestração musical assistida por computador como sendo um problema de otimização combinatoria em que o objetivo reside na combinação de registos tímbricos de instrumentos diferentes de maneira a que o resultado final se aproxime, ou mesmo se iguale, a um timbre designado pelo utilizador.

A resolução deste problema manualmente acarreta um elevado conhecimento musical pelo que, através do uso de ferramentas computacionais, pretende-se auxiliar o utilizador nesta tarefa de combinação de registos tímbricos. Implementações prévias para a resolução deste tipo de problemas apresentam algumas limitações no que toca à diversidade de soluções produzidas e à abordagem do timbre sonoro como fenómeno multidimensional.

Neste trabalho foi inicialmente desenvolvida uma implementação de um algoritmo genético que foi posteriormente aplicada a um modelo teórico do problema de orquestração. O mesmo modelo foi também aplicado a um sistema imunitário artificial para comparação e análise dos resultados obtidos e validação do modelo teórico desenvolvido. Devido ao uso de um sistema imunitário artificial, o sistema de orquestração proposto oferece várias soluções ótimas simultaneamente, colmatando um dos principais problemas nas abordagens prévias ao problema de orquestração.

Os resultados obtidos foram submetidos a uma avaliação perceptiva utilizando um teste de escuta. Os dados da avaliação permitiram concluir que o sistema desenvolvido apresenta um bom desempenho na orquestração de sons musicais, com algumas limitações. Os dados permitiram também concluir que as funções desenvolvidas para a classificação de aptidão das soluções refletem razoavelmente bem a similaridade entre sons. Resultados preliminares deste trabalho foram apresentados no 1º Congresso Doutoral em Engenharia realizado na Faculdade de Engenharia da Universidade do Porto [1].

Abstract

This work aimed to implement and evaluate the use of algorithms inspired by biological processes in musical orchestration problems. In this context computer-aided musical orchestration is defined as a combinatorial optimization problem in which the objective is to combine different timbral records of different instruments such that the end result approaches a designated timbral record.

The manual resolution of this problem requires a high musical knowledge. Computational tools pretend to aid the user in this task of combining timbral records. Previous implementations to the resolution of this kind of problems have some limitations regarding the diversity of the solution set and the approach to the timbre as a multidimensional phenomenon.

In this work it was initially developed an implementation of a genetic algorithm which was subsequently applied to a theoretical model of the orchestration problem. The same model was applied to an artificial immune system for comparison and analysis of the results and validation of the developed theoretical model. Due to the use of an artificial immune system the proposed orchestration system offers several optimum solutions simultaneously, addressing one of the main problems of the previous approaches to the orchestration problem.

The results obtained were subjected to a perceptual evaluation using a listening test. The data from the evaluation allowed to conclude that the system developed has, with some limitations, a good performance in the orchestration of musical sounds. The obtained data allowed also to conclude that the developed functions used to classify the aptitude of the solutions reflect reasonably well the similarity between sounds. Preliminary results of this work were presented at the 1st Doctoral Congress in Engineering held at Faculty of Engineering of University of Porto [1].

Agradecimentos

Com a finalização deste trabalho termino uma longa jornada iniciada 5 anos atrás quando me inscrevi pela primeira vez na Faculdade de Engenharia da Universidade do Porto no Mestrado de Engenharia Eletrotécnica e de Computadores. Nestes 5 anos expandi o meu conhecimento numa área que sempre gostei e para além disso adquiri novas amizades que irão permanecer ao longo da minha vida. Foram certamente os melhores anos que já vivi.

Em primeiro lugar, gostaria de agradecer aos meus pais por me terem facultado a possibilidade de estudar numa das melhores universidades deste país e por terem estado sempre ao meu lado, apoiando-me nos bons e nos maus momentos. Sei que os sacrifícios foram muitos e espero um dia poder retribuir. Sei também que sem eles certamente não estaria neste momento a escrever este texto. Um muito obrigado. Agradeço também aos restantes elementos da minha família por todo o apoio e incentivo incessável.

Em segundo lugar, agradeço aos meus colegas de casa no Porto pela amizade e pelos bons momentos que vivemos. Foram anos muito felizes onde foram criadas memórias que irão permanecer para toda a vida. Ao Franco, Hugo, Ivo e Meira um muito obrigado pela amizade, pelo apoio e pelos bons momentos que passamos todos juntos.

Em terceiro lugar, não poderia deixar de agradecer aos meus colegas de faculdade que partilharam esta jornada comigo. Obrigado Garcia, Guilherme, Iria, Pereira, Tó e Zulu pelos bons anos, desejo tudo de bom e muita sorte quer a nível profissional quer a nível pessoal. Espero que as vossas amizades me acompanhem pelo resto da minha vida.

Finalmente, um agradecimento especial a todos os docentes que acompanharam a minha jornada nesta faculdade. Um agradecimento especial também aos meus orientadores pela paciência e disponibilidade que tiveram ao longo da realização deste trabalho.

Obrigado a todos.

José Miguel Lima de Abreu

“Music expresses that which cannot be put into words and that which cannot remain silent.”

Victor Hugo

Conteúdo

1	Introdução	1
1.1	Contexto	1
1.2	Especificação	1
1.2.1	Orquestração Musical	2
1.2.2	Computação Inspirada por Processos Biológicos	4
1.3	Objetivos	5
1.4	Estrutura	5
2	Orquestração Musical	7
2.1	Definição	7
2.2	Estado da Arte	8
2.3	Problema da Mochila	9
2.4	Timbre	10
3	Computação Inspirada por Processos Biológicos	13
3.1	Computação Evolutiva	14
3.1.1	Algoritmos Genéticos	15
3.2	Sistemas Imunitários Artificiais	18
3.2.1	Seleção Clonal e Maturação de Afinidade	20
3.2.2	Seleção Negativa	20
3.2.3	Teoria do Perigo	20
3.2.4	aiNet	21
4	Formalização	23
4.1	Modelo Teórico da Orquestração Musical	23
4.2	Base de Dados Sonora	26
4.3	Descritores	26
4.3.1	Centróide Espectral	27
4.3.2	Espalhamento Espectral	29
4.3.3	Intensidade Sonora	29
4.3.4	Picos Espectrais	29
4.3.5	RMS	36
4.3.6	Energia	36
4.3.7	Frequência Fundamental	36
4.4	Funções de Combinação	36
4.4.1	Avaliação das Funções de Combinação	38
4.5	Funções de Avaliação	39

5	Implementação	43
5.1	Base de Dados Sonora	43
5.2	Representação	45
5.3	Pré-Processamento	46
5.4	Algoritmo Genético	46
5.4.1	Operadores Genéticos	48
5.5	Sistema Imunitário Artificial	50
5.5.1	aiNet e opt-aiNet	50
5.5.2	Implementação do Algoritmo opt-aiNet	52
5.6	Aplicação	53
5.6.1	Estrutura	54
5.6.2	Funcionamento	55
5.6.3	Interface	58
6	Avaliação e Resultados	61
6.1	Método	61
6.2	Estrutura	63
6.3	Resultados Obtidos	64
7	Conclusões e Trabalho Futuro	77
7.1	Conclusões	77
7.2	Trabalho Futuro	79
7.3	Agradecimentos	80
	Referências	81

Lista de Figuras

1.1	Orquestração musical assistida por computador: a tarefa de combinar registos tímbricos de diferentes instrumentos de forma a que o resultado final se aproxime a um timbre designado	3
2.1	Representação gráfica dos resultados obtidos por Grey em [2]. A avaliação das diferentes características dos tons permitem o seu mapeamento num espaço ortogonal. As ligações entre os diferentes pontos, correspondentes aos diferentes tons avaliados, representam a semelhança e agrupamento propostos pelos utilizadores	11
4.1	Descritores sonoros avaliados para cada som da base de dados	27
4.2	Forma de onda de um som e a evolução dos seus descritores com o tempo	28
4.3	Forma de onda de um som e a evolução dos descritores entre 250ms e 750ms	28
4.4	Espetro em decibéis e limite da gama de valores selecionados para o cálculo de picos espectrais a tracejado	30
4.5	Gama de valores selecionados do espectro apresentado na figura 4.4	31
4.6	Picos espectrais identificados	31
4.7	Espetro de um som objetivo	32
4.8	Espetro de um som selecionado para o processos de pesquisa. O som apresenta um pico espectral em comum com o som objetivo (500 Hz)	33
4.9	Espetro de um som não selecionado para o processo de pesquisa. O som não apresenta nenhum pico espectral em comum com o som objetivo	33
4.10	Espetro de um som objetivo	34
4.11	Espetro de um som selecionado para o processos de pesquisa	35
4.12	Contribuição do som da figura 4.11 tendo em conta o som objetivo da figura 4.10	35
4.13	Funções de combinação	37
4.14	Funções de avaliação	39
4.15	Espetro de um som, nota F6	41
4.16	Espetro de um som, nota C7	41
4.17	Combinação do espectro de sons, notas F6 e C7	41
5.1	Resultados do processo de segmentação dos ficheiros da base de dados RWC	44
5.2	Resultados do processo de identificação da frequência fundamental utilizando o SWIPE	45
5.3	Ilustração do funcionamento clássico do esquema de <i>roulette wheel</i>	49
5.4	Probabilidade de mutação segundo o valor de <i>fitness</i> normalizado, com γ igual a 1.2	53
5.5	Módulos da aplicação desenvolvida e suas dependências	54
5.6	Organização dos ficheiros na base de dados. Para cada ficheiro de som (.WAV) existe um ficheiro de dados anexo (.MAT) com as características desse som	56
5.7	Resposta em frequência dos filtros ERB	57

5.8	Inicialização do processo de orquestração	59
5.9	Interface gráfica desenvolvida onde são apresentados os resultados obtidos	59
5.10	Seleção dos diferentes resultados	60
6.1	Interface do teste de escuta desenvolvido: mensagem inicial	64
6.2	Interface do teste de escuta desenvolvido: escolha da similaridade e diversidade	65
6.3	Resultados do teste de escuta: diversidade perceptiva	68
6.4	Resultados do teste de escuta: som objetivo “Car Horn”	70
6.5	Resultados do teste de escuta: som objetivo “Synth”	70
6.6	Resultados do teste de escuta: som objetivo “Tbreed”	71
6.7	Resultados do teste de escuta: som objetivo “Ahh”	71
6.8	Resultados do teste de escuta: som objetivo “Wind Harp”	72
6.9	Resultados do teste de escuta: som objetivo “Didgeridoo”	72
6.10	Resultados do teste de escuta: som objetivo “Elephant”	73
6.11	Resultados do teste de escuta: som objetivo “Frog”	73
6.12	Resultados do teste de escuta: som objetivo “Girl Scream”	74
6.13	Resultados do teste de escuta: som objetivo “Gong”	74

Lista de Tabelas

4.1	Erro relativo entre características calculadas utilizando as funções de combinação e características medidas da combinação. Em parênteses o desvio padrão do erro relativo.	38
5.1	Exemplo de índices e nomes de ficheiros da base de dados	46
6.1	Valores de aptidão dos melhores sons orquestrados obtidos pelos diferentes algoritmos para 3 sons objetivo	62
6.2	Sons objetivo utilizados no teste de escuta	66
6.3	Avaliação da similaridade percetiva média das orquestrações submetidas ao teste de escuta	67
6.4	Valores médios de diversidade objetiva e desvios padrões dos sons utilizados no teste de escuta	69
6.5	Valores de aptidão dos sons utilizados no teste de escuta	75

Abreviaturas e Símbolos

AG	Algoritmo Genético
RMS	<i>Root Mean Square</i>
RWC	<i>Real World Computing</i>
SIA	Sistema Imunitário Artificial
SWIPE	<i>Sawtooth Waveform Inspired Pitch Estimator</i>
WAV	<i>Waveform Audio File Format</i>

Capítulo 1

Introdução

1.1 Contexto

O processo criativo e a capacidade de inovação são características inerentes de todos os seres humanos e uma das qualidades mais importantes da nossa espécie distinguindo-nos das máquinas e dos algoritmos que são, na sua maioria, regidos por regras e determinismo. Estas mesmas máquinas fazem atualmente parte do nosso quotidiano, logo, torna-se natural o surgimento de interesse em aplicações que facilitem o processo criativo e que, talvez num futuro próximo, substituam o fator humano na totalidade dotando efetivamente uma máquina de criatividade.

Igualmente importante, e difícil, à criação de originalidade é a capacidade de reconhecimento da mesma. Ambas partilham de aspetos que estão ligados a perceções sensoriais e que são dificilmente expressos em medidas objetivas dificultando uma formalização detalhada destes processos no domínio computacional. Assim, e apesar da constante evolução no mundo tecnológico e científico, a criação artística continua a ser uma área fortemente dominada pela criatividade humana existindo um interesse cada vez maior em explorar a ligação entre arte e ciência.

No caso da criação musical, a investigação decorrente em torno de computação e síntese sonora têm vindo a providenciar aos compositores, musicólogos e entusiastas ferramentas poderosas para a manipulação e geração de diversos sons musicais, desde simples tons ou notas até peças complexas e partituras. Com a evolução do poder computacional disponível para os utilizadores deste tipo de ferramentas, a interação com os mesmos têm vindo a tornar-se cada vez mais diminuta, podendo este ser apenas responsável por especificar os parâmetros iniciais e julgar o resultado final. No entanto, e apesar da elevada capacidade de processamento disponível atualmente, vários problemas surgem na implementação destes sistemas, especialmente no julgamento autónomo dos parâmetros estéticos dos resultados produzidos.

1.2 Especificação

Nesta área de investigação podem ser especificados dois domínios diferentes e independentes que diferem quanto à complexidade do resultado produzido e aos parâmetros iniciais fornecidos:

o domínio de geração musical e o domínio de manipulação musical. As técnicas presentes nestes domínios utilizam diferentes algoritmos para atingir os resultados pretendidos, residindo em cada uma delas um objetivo comum: produzir o maior número possível de resultados que, por um lado, sejam úteis ao utilizador e que, por outro lado, apresentem características únicas, numa tentativa parca de simular o processo criativo utilizando uma aplicação computacional.

O domínio de geração musical apresenta técnicas de síntese sonora e composição algorítmica. Ambas são utilizadas para a geração de novos sons, composições ou músicas e diferem essencialmente na complexidade do resultado produzido: Na síntese sonora o resultado é um conjunto de sons individuais, enquanto que, na composição algorítmica o resultado é uma peça musical ou partitura. Neste domínio são ainda incluídas técnicas de orquestração musical assistida por computador, que pode ser definida como a tarefa de combinar registos tímbricos de instrumentos diferentes, de maneira a que o resultado final se aproxime, ou mesmo se iguale, a um timbre designado pelo compositor [3].

Por outro lado, o domínio de manipulação musical inclui técnicas de *sound morphing* e *mashup* que são utilizadas na manipulação de sons ou músicas existentes e permitem a um determinado utilizador alterar um som ou música de maneira a obter uma nova variação dos mesmos, ou mesmo combinar vários sons num só, assim como combinar diferentes partes de músicas numa só. No caso de *sound morphing*, um ou mais sons iniciais são combinados dando origem a um novo som, enquanto que no *mashup* várias músicas são combinadas para produzir uma nova música.

1.2.1 Orquestração Musical

A orquestração relaciona-se com a técnica de compor peças musicais para uma orquestra musical. A prática desta técnica necessita que o compositor conheça as propriedades dos diferentes instrumentos que compõem a orquestra e as sonoridades que advêm de cada um. No contexto deste trabalho considera-se a orquestração musical como um problema combinatório em que se pretende organizar determinados instrumentos de forma a obter um registo tímbrico específico. Para tal, recorreu-se a ferramentas computacionais para auxiliar este processo de organização de forma a autonomizar o funcionamento do mesmo.

Nos problemas de orquestração musical assistida por computador pretende-se aproximar um determinado registo tímbrico de um som utilizando combinação de registos tímbricos de outros sons. O registo tímbrico que se pretende aproximar é especificado pelo compositor e os sons que podem ser utilizados na combinação pertencem ao leque de sons produzidos pelos instrumentos tipicamente presentes numa orquestra musical. A resolução deste problema manualmente acarreta um elevado conhecimento musical. Efetivamente um compositor com pouca experiência na resolução deste tipo de problemas não sabe, à partida, qual será o timbre resultante da combinação de um som com outro, e mesmo compositores com elevada experiência encontram por vezes dificuldades em obter um registo tímbrico específico. Para além da experiência necessária, a orquestração musical é também um processo iterativo, sendo que, na maior parte das vezes o compositor acrescenta sucessivamente instrumentos, de forma a obter um registo tímbrico uniforme e que seja

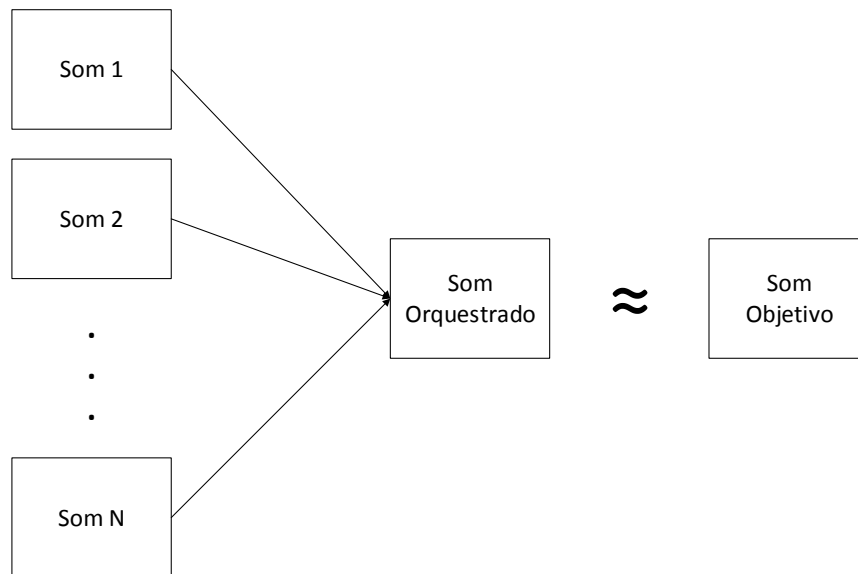


Figura 1.1: Orquestração musical assistida por computador: a tarefa de combinar registos tímbricos de diferentes instrumentos de forma a que o resultado final se aproxime a um timbre designado

próximo do objetivo. Na figura 1.1 apresenta-se um diagrama com os objetivos da orquestração musical assistida por computador.

A sonoridade de cada instrumento numa orquestra é única, apresentando características sonoras que variam conforme o instrumento e conforme o estilo de interpretação realizado pelo músico. A variedade de sons que cada instrumento consegue produzir é de tal forma alta, que se torna praticamente impossível a um compositor reter todas estas variações. É este o principal problema quando se pretende orquestrar um som manualmente. Por vezes, um determinado registo tímbrico até pode ser aproximado razoavelmente bem através de orquestração manual. No entanto, neste processo de aproximação uma grande parte das combinações possíveis não são consideradas podendo ser assim descartadas soluções com uma similaridade maior com o objetivo.

A orquestração musical com recurso a ferramentas computacionais pretende auxiliar o compositor nesta tarefa de combinação de registos tímbricos de forma a aproximar um registo tímbrico objetivo. Neste tipo de ferramentas o julgamento das combinações é feito autonomamente, sendo que o compositor apenas necessita de avaliar as soluções produzidas, colmatando-se o principal problema da orquestração manual através da pesquisa automática de soluções próximas do objetivo. Com este processo de pesquisa, e com o elevado poder computacional disponível atualmente, é possível percorrer eficientemente as diversas possibilidades existentes quando são considerados os diversos instrumentos de uma orquestra e as sonoridades criadas por cada um. Tais sistemas, apesar de serem eficientes no processo de pesquisa, não o são, por vezes, no processo de julgamento das soluções obtidas. Efetivamente várias considerações sobre o timbre como fenómeno sonoro multidimensional devem ser levadas em conta na especificação de funções que avaliem a similaridade entre sons.

1.2.2 Computação Inspirada por Processos Biológicos

As técnicas de criação musical com recurso a ferramentas computacionais apresentam em comum um vasto espaço de soluções possíveis que devem ser julgadas e selecionadas autonomamente pelo sistema, apresentando como resultado um conjunto restrito de soluções que tenham utilidade ao utilizador. Devido ao grande número de parâmetros que podem ser considerados na criação dos elementos musicais, algoritmos tradicionais de pesquisa não são particularmente eficientes na resolução destes problemas. Esta lacuna pode ser colmatada através do recurso a algoritmos de otimização ligados à área da inteligência computacional. Estes algoritmos oferecem um mecanismo que permite percorrer simultaneamente o espaço de soluções existente, oferecendo ao mesmo tempo uma rápida velocidade de convergência para valores ótimos. As soluções propostas por estes algoritmos são melhoradas a cada iteração, sendo que os resultados convergem para um conjunto de valores que maximizam ou minimizam determinadas funções de avaliação, responsáveis por quantificar a utilidade dos mesmos. Efetivamente a definição destas funções de avaliação é o problema de maior envergadura neste tipo de sistemas por não ser ainda claro como é que o julgamento dos parâmetros estéticos pode ser automatizado [4]. Uma descrição mais detalhada do funcionamento destes algoritmos é feita no capítulo 3.

A aplicação destes algoritmos no domínio da criação musical impulsionou uma abordagem informatizada ao processo de composição e síntese sonora. Mais recentemente, e utilizando modelos analíticos que descrevem diferentes elementos musicais, tornou-se possível a implementação de ferramentas automatizadas que assistem o processo de composição musical. Ao nível da síntese sonora foram propostas implementações que utilizam algoritmos genéticos [5] e sistemas imunitários artificiais [6]. Utilizando estes algoritmos, foram também propostas implementações de composição algorítmica recorrendo a algoritmos genéticos [7] [8], sistemas imunitários artificiais [9] e mapas auto organizáveis [10] [11] [12].

A variedade de ferramentas que podem ser criadas com estes algoritmos é praticamente ilimitada. Em [8] por exemplo, Biles propõem o GenJam (*genetic jammer*), uma aplicação de composição algorítmica interativa com capacidade de improvisar solos de *Jazz* recorrendo a um algoritmo genético. O GenJam é capaz de produzir solos autonomamente, assim como interagir com o compositor ou banda através da resposta a sons produzidos por estes. Outra aplicação de composição algorítmica recorrendo também a algoritmos genéticos é proposta no Vox Populi [7]. Nesta aplicação um número pré-definido de acordes é gerado e evoluído de forma a maximizar um critério de avaliação baseado em fatores físicos relevantes à perceção sonora. Os acordes gerados podem ser posteriormente utilizados pelo compositor numa progressão musical.

A variedade de algoritmos existentes que são inspirados por processos biológicos, o seu bom desempenho em problemas de otimização com um grande número de parâmetros e o elevado poder computacional disponível atualmente têm vindo a contribuir assim no desenvolvimento de ferramentas cada vez mais poderosas na resolução de problemas que eram previamente exclusivos do domínio humano e criativo. Com estas novas abordagens, a composição musical e a síntese sonora englobam uma perspetiva em que o espaço de combinações possíveis é praticamente ilimi-

tado, dependendo em última instância do compositor a mestria dos elementos produzidos por estas ferramentas na criação de novas e inovadoras composições artísticas. Por fim é importante realçar que, apesar da boa eficácia destes algoritmos em problemas de otimização multidimensional, uma correta formalização teórica das características do problema a abordar patenteia a fundação mais importante para a obtenção de soluções úteis ao utilizador.

1.3 Objetivos

Neste trabalho implementou-se e avaliou-se o uso de algoritmos inspirados por processos biológicos em problemas de orquestração musical. Inicialmente foi feita uma implementação de um algoritmo genético, tendo em conta as diferentes características sonoras que devem ser consideradas na avaliação de sons musicais. Foi posteriormente desenvolvido um modelo teórico, fundamentando o problema da orquestração numa perspetiva analítica. Este modelo foi aplicado num algoritmo genético de forma a demonstrar a sua validade, precedido de uma exploração e aplicação num sistema imunitário artificial inspirado na aiNet [13]. Após a aplicação deste modelo nos dois algoritmos propostos, foi feita uma comparação analítica e subjetiva dos resultados obtidos de forma a determinar as principais vantagens e possíveis limitações da utilização destes algoritmos e do modelo proposto. Uma introdução mais aprofundada aos conceitos de orquestração é feita no capítulo 2, onde são detalhados os principais desafios deste tipo de problema e são revistas as soluções já desenvolvidas. São também introduzidos alguns fundamentos teóricos que constam no modelo de orquestração desenvolvido.

Resultados preliminares do trabalho desenvolvido foram apresentados no 1º Congresso Doutoral em Engenharia realizado na Faculdade de Engenharia da Universidade do Porto. A participação envolveu a escrita de um artigo [1] e a realização de um poster cujo conteúdo refletiu o trabalho desenvolvido até à data de realização deste congresso.

1.4 Estrutura

Este documento contém mais 6 capítulos. No capítulo 2, é feita uma revisão aos conceitos de orquestração, apresentadas as implementações existentes e resumidas as vantagens e limitações de cada uma. No capítulo 3, são descritos os princípios da computação inspirada por processos biológicos e são introduzidos algoritmos baseados nesta linha de pesquisa. No capítulo 4, é descrito o modelo teórico desenvolvido para o problema de orquestração musical. No capítulo 5, apresenta-se a solução desenvolvida e a arquitetura do sistema de orquestração. No capítulo 6, são apresentados os resultados obtidos nos testes de escuta às soluções obtidas. Por fim, no capítulo 7, é feita a conclusão e descritos os principais aspetos que podem ser melhorados no sistema desenvolvido.

Capítulo 2

Orquestração Musical

Atualmente existe uma grande variedade de ferramentas computacionais que auxiliam compositores no processo de criação e exploração musical. A maior parte destas ferramentas providencia estruturas primitivas que são depois combinadas pelo compositor num nível crescente de complexidade até ser atingido o resultado final pretendido. Este processo, apesar de se basear em métodos computacionais para a criação das estruturas primitivas, apresenta uma grande componente manual, deixando à responsabilidade do compositor a combinação e integração das diferentes estruturas e dos seus registos tímbricos. A manipulação das estruturas primitivas por meios computacionais é uma área cada vez mais explorada e mais recentemente as evoluções em inteligência artificial, e em particular na área de computação evolutiva, oferecem mecanismos que permitem lidar com um grande número de parâmetros e soluções. Este tipo de algoritmos foi já aplicado com sucesso na resolução de problemas em áreas como engenharia, biologia e economia, e foi proposto em diversas implementações de criação musical, incluindo problemas de orquestração.

2.1 Definição

O objetivo da orquestração musical com recurso a ferramentas computacionais é o de auxiliar o utilizador na tarefa de combinar autonomamente os registos tímbricos de diferentes instrumentos de forma a que o resultado final se aproxime, ou mesmo se iguale, a um timbre designado pelo utilizador. A complexidade deste método pode ser melhor percebida tendo em conta que uma orquestra é composta por um grande número de instrumentos diferentes, com características únicas e sonoridades que dependem do estilo musical e do músico que as interpreta. Estes fatores, quando combinados, geram um espaço de sonoridades praticamente ilimitado que pode ainda ser estendido se considerarmos os fatores sonoros que advêm do posicionamento dos instrumentos na orquestra e os fenómenos acústicos do local físico em que a orquestra se encontra [3].

Na prática, apenas um reduzido número destas combinações são relevantes ao utilizador que pretende obter um registo sonoro específico. Nesses casos o problema de orquestração deve ter em

conta os timbres de cada instrumento e as combinações que deles advêm, num processo complexo e difícil de formalizar analiticamente. Para além disso, o timbre é um atributo sonoro que apresenta características multi-dimensionais que são impossíveis de quantificar numa escala linear como alto ou baixo, forte ou fraco, ou longo ou curto [14]. Esta impossibilidade acarreta uma complexidade acrescida no que toca à caracterização formal deste atributo nos problemas de criação musical.

2.2 Estado da Arte

A complexidade da orquestração pode ser dividida em três vertentes [3]: o grande espaço de combinações possíveis dos diferentes instrumentos da orquestra, as características multi dimensionais que o timbre apresenta, e a complexidade da evolução das características sonoras no tempo. Várias implementações foram propostas para a resolução deste problema que têm em conta estas vertentes, com algumas limitações. Em [15], é proposto um programa desenvolvido em Lisp (*SPORCH - Spectral ORCHEstration*) que utiliza uma base de dados de instrumentos e um método de pesquisa iterativo. As soluções são obtidas através de igualização com o conteúdo espectral do som objetivo e avaliadas através da distância Euclidiana entre os picos espectrais do som objetivo e da solução. Inicialmente o algoritmo tenta encontrar na base de dados um instrumento com os picos espectrais mais próximos do espectro objetivo. O espectro do melhor candidato é de seguida extraído do espectro do objetivo e o algoritmo continua a execução iterando sobre a diferença obtida. Por se basear num algoritmo iterativo este programa necessita de pouco poder computacional. No entanto, várias soluções são descartadas com este processo devido ao facto de apenas um instrumento ser comparado com o objetivo a cada iteração. Outro sistema, proposto em [16], utiliza também um algoritmo iterativo aproximando o resultado através de comparação com o envelope espectral do som objetivo. Como os picos de frequência não são considerados nesta implementação, os resultados propostos podem não igualar o som objetivo quanto à altura e intensidade sonora.

A última implementação a ser descrita utiliza uma classe particular de algoritmos pertencentes ao ramo de investigação de inteligência computacional. Denominados de algoritmos evolutivos, estes algoritmos são inspirados por processos biológicos e apresentam um bom desempenho em problemas complexos como a orquestração. Comparativamente aos algoritmos iterativos apresentados em [15] e [16], a utilização de algoritmos evolutivos providencia a um sistema de orquestração uma gama muito mais diversificada de resultados à custa de um uso de maior poder computacional.

Este sistema de orquestração pode ser encontrado em [3]. Nesta implementação é utilizado um algoritmo genético para a descoberta de soluções relevantes, tendo em conta restrições especificadas no início da execução pelo compositor. O processo de pesquisa do algoritmo genético utilizado mantém uma lista de indivíduos, codificados sobre a forma de um genoma binário, que representa as possíveis combinações de instrumentos. Tal como no processo biológico de evolução natural, estes indivíduos sofrem mutações no seu genoma e, através de cruzamento, originam novos indivíduos que partilham características com a geração anterior. Selecionando apenas os indivíduos mais aptos para cruzamento a população irá apresentar uma tendência de evolução para

regiões ótimas do espaço de pesquisa. Para além disso, a utilização de operações de cruzamento e mutação introduzem variações aleatórias nas soluções oferecendo uma diversidade muito maior nos resultados obtidos [17]. Esta implementação aborda a orquestração como um problema de otimização combinatória em que o objetivo é minimizar um conjunto de valores, calculados utilizando funções de avaliação que modelam o comportamento tímbrico e sonoro dos instrumentos. Através de um conjunto de restrições é possível limitar o espaço de pesquisa, especificando, por exemplo, o número máximo de instrumentos ou um conjunto de notas e acordes que a solução pode conter.

Nesta implementação [3], a evolução temporal do espectro e timbre não é abordada no processo de pesquisa. No entanto, são utilizados um conjunto de características calculadas para cada som. O cálculo do valor destas características é feito utilizando os modelos desenvolvidos em [18], [19] e [14], incluindo os modelos tímbricos multi-dimensionais propostos por estes. As características são posteriormente utilizadas nas funções de avaliação dos indivíduos produzidos pelo algoritmo genético.

2.3 Problema da Mochila

Ambos os métodos propostos em [15] e [16] utilizam algoritmos que necessitam de pouco poder computacional e apresentam resultados satisfatórios mas que são, numa certa medida, previsíveis. Efetivamente, nestas duas primeiras abordagens apresentadas o problema de orquestração é tratado utilizando uma otimização semelhante ao problema da mochila (*knapsack problem*) [20]. Este problema insere-se na classe da otimização combinatória e têm como objetivo calcular combinações que maximizam uma determinada função objetivo, tendo em conta restrições no número e tipo de elementos que constituem essas combinações. Num ponto de vista metafórico uma mochila, com capacidade limitada, deve ser preenchida com objetos de tamanhos diferentes de maneira a que o espaço ocupado pelos objetos não exceda a capacidade máxima da mochila e a que os objetos selecionados tenham a melhor utilidade possível ao portador da mochila. Mais formalmente, existem n objetos e um vetor de valores binários x_i ($i = 1, \dots, n$), tal que:

$$x_i = \begin{cases} 1 & \text{se objeto } i \text{ é selecionado} \\ 0 & \text{em caso contrário} \end{cases} \quad (2.1)$$

Então, se u_i for um valor que quantifica a utilidade do objeto i , e_i for o espaço ocupado pelo mesmo, e c a capacidade máxima da mochila, o objetivo é selecionar uma combinação que satisfaça a condição

$$\sum_{i=1}^n e_i x_i \leq c \quad (2.2)$$

e que maximize a função objetivo

$$\sum_{i=1}^n u_i x_i \quad (2.3)$$

Nas duas primeiras soluções previamente descritas ao problema de orquestração é utilizada uma abordagem semelhante ao problema da mochila, na medida em que inicialmente são escolhidos os instrumentos que apresentam uma maior similaridade com o objetivo, iterando e acrescentando instrumentos até ser atingido um resultado satisfatório. Devido ao uso de uma única função objetivo, as soluções produzidas por estes sistemas tendem a não considerar a complexidade do timbre como fenómeno multi dimensional que evolui com o tempo [3]. Efetivamente, o uso de parâmetros espectrais para otimizar os resultados não endereça na totalidade a complexidade do problema de orquestração por não incluir as características subjetivas despoletadas pelo fenómeno tímbrico.

2.4 Timbre

A abordagem à multi-dimensionalidade do timbre sonoro constitui uma das principais dificuldades nos problemas de criação musical devido à conjugação dos diversos fatores que o compõem. Em [21], Handel compara a percepção de fenómenos sonoros ao processo de deteção e reconhecimento facial. Assim, tal como uma face é composta por diversas características distintas, apenas a sua conjugação e organização espacial permitem o reconhecimento e identificação de uma determinada pessoa. Similarmente, a percepção de fenómenos sonoros depende das diferentes características que constituem esse fenómeno e a correspondente evolução temporal das mesmas. Estes fatores atribuem ao timbre qualidades que são dependentes tanto de fenómenos acústicos como de processos percecionais [21].

As técnicas encontradas na literatura para a abordagem da descrição qualitativa e analítica do timbre utilizam escalamento multi-dimensional para mapear certas características subjetivas num espaço ortogonal, sendo que diferenças entre sons podem ser calculadas através do cálculo da distância Euclidiana dos pontos correspondentes [22]. Grey [2] [23], investigou inicialmente a multi dimensionalidade do timbre utilizando dados subjetivos fornecidos por utilizadores humanos, responsáveis por avaliar a similaridade entre diferentes tons produzidos por técnicas de síntese aditiva. Os tons foram avaliados segundo a sua altura, tom e duração estimada e os dados obtidos foram correlacionados com o objetivo de estimar as características analíticas que influenciam a percepção do fenómeno tímbrico. As conclusões obtidas permitiram a construção de um modelo tri-dimensional do timbre e a determinação da relação entre cada dimensão. Neste modelo, a primeira dimensão corresponde à distribuição da energia espectral do tom, a segunda diz respeito à sincronização e altura dos harmónicos mais altos e a terceira é interpretada como a disposição temporal de sinais de alta frequência e baixa amplitude. A construção deste modelo tri-dimensional permitiu uma descrição inicial dos parâmetros analíticos que constituem a percepção dos fenómenos tímbricos a um utilizador humano. Na figura 2.1 pode ser visualizada uma representação gráfica dos resultados obtidos por Grey.

Estudos semelhantes forneceram um aprofundamento do modelo desenvolvido inicialmente por Grey. Em [24], Krumhansl propõem também um espaço com três dimensões relacionadas com *ataque* (momento no tempo em que o som é inicialmente ouvido), sincronismo entre os harmónicos e *brilho* (quantidade relativa de frequências elevadas presentes no espetro). Krimphoff [25],

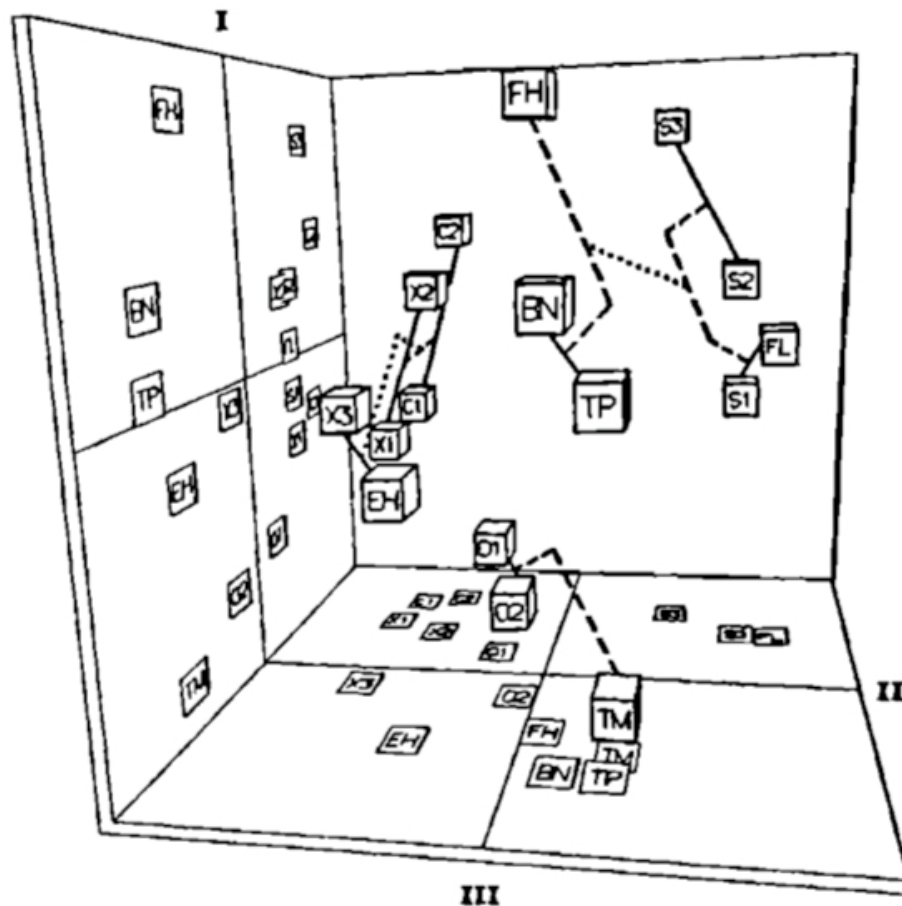


Figura 2.1: Representação gráfica dos resultados obtidos por Grey em [2]. A avaliação das diferentes características dos tons permitem o seu mapeamento num espaço ortogonal. As ligações entre os diferentes pontos, correspondentes aos diferentes tons avaliados, representam a semelhança e agrupamento propostos pelos utilizadores

estudou a relação entre as dimensões tímbricas, concluindo que o *brilho* de um som relaciona-se logaritmicamente com a distribuição espectral da sua energia e com a rapidez de *ataque* em relação ao tempo. Experiências semelhantes realizadas com sons sintetizados por McAdams [14], levaram à conclusão que as dimensões mais importantes neste espaço tri-dimensional são o logaritmo do tempo de *ataque*, a distribuição espectral e o seu correspondente grau de variação. Mais recentemente, e utilizando também técnicas de escalamento multi-dimensional, Caclin [26], concluiu que ouvintes humanos utilizam o tempo de *ataque*, a distribuição espectral e a resolução do número de harmónicos que compõem os sons para avaliar a similaridade entre estes.

Através da utilização destes modelos baseados no escalamento multi-dimensional, é possível avaliar as diferenças relativas entre sons [22]. Apesar de cada som ser essencialmente distinto e ocupar uma posição única neste espaço multi-dimensional, a exploração dos conceitos desenvolvidos permite interpretar as conexões inerentes que existem entre diferentes sons oferecendo

mecanismos analíticos para a análise de um processo que é essencialmente subjetivo. Para além das características que compõem o fenómeno tímbrico pertencentes a estes modelos, a descrição de um determinado som pode ser feita mais abstratamente, utilizando outros parâmetros para além dos já referidos. Entre os parâmetros utilizados encontram-se [27]:

- Amplitude da componente sinusoidal;
- Amplitude da componente residual;
- Forma do espetro da componente sinusoidal;
- Forma do espetro da componente residual;
- Distorção harmónica;
- Ruído;
- Centróide espectral;
- Inclinação espectral.

Cada um destes parâmetros pode ser medido segundo o seu valor instantâneo ou mediano e em relação ao seu valor absoluto ou segundo a variação do mesmo, constituindo assim quatro valores diferentes que podem ser utilizados para a caracterização dos aspetos de diferentes sons.

Capítulo 3

Computação Inspirada por Processos Biológicos

A computação inspirada por processos biológicos (ou *computação bio-inspirada*), é uma linha de pesquisa que utiliza modelos físicos e comportamentais de sistemas biológicos para o projeto e implementação de algoritmos capazes de solucionar problemas complexos. Os algoritmos apresentados neste capítulo pertencem a esta linha de pesquisa e foram já aplicados com sucesso na resolução de problemas em áreas como engenharia, arte, biologia e economia. Especialmente indicados para problemas de otimização, este tipo de algoritmos apresenta como vantagem mais significativa a possibilidade de resolver problemas apenas através da formalização matemática dos atributos que deverão estar presentes na solução e sem necessidade de especificar as operações que devem ser realizadas para atingir a mesma. Efetivamente, estes algoritmos abordam da mesma maneira a resolução de diversos problemas oferecendo um mecanismo robusto, flexível e passível de ser utilizado na resolução de problemas complexos para os quais outras técnicas conhecidas não são eficazes ou aplicáveis. É importante realçar que apesar da convergência para uma ou mais soluções ser garantida nestes sistemas, a convergência para a solução ótima não o é devido ao não determinismo como estas são obtidas. Formalizando, podemos afirmar que os algoritmos bio-inspirados modelam métodos estocásticos para a resolução de problemas complexos, garantindo a obtenção de uma ou mais soluções quase-ótimas.

Em todas as variações reside o conceito fundamental da existência de uma população constituída por um conjunto de indivíduos diferentes que representam as soluções possíveis do problema. Estes indivíduos são melhorados a cada iteração e a qualidade dos mesmos é avaliada através de uma determinada função de avaliação responsável por quantificar numericamente a semelhança das características desse indivíduo em relação à solução pretendida. A correta escolha das funções de avaliação permite então um acréscimo na qualidade das soluções obtidas e uma redução do número de iterações necessárias para atingir as mesmas, sendo esta a tarefa mais difícil na implementação deste tipo de sistemas por influenciar em larga escala o processo de pesquisa e

refinamento de soluções. Do ponto de vista criativo, estes algoritmos oferecem ferramentas poderosas na resolução de problemas com um grande espaço de soluções possíveis e apresentam resultados diversificados através do uso de incerteza no processo de otimização.

3.1 Computação Evolutiva

O ramo da computação evolutiva engloba um conjunto de métodos inspirados na teoria de evolução de Darwin tendo sido propostos pela primeira vez nos anos 50 ([28], [29], [30]). Nestas abordagens propostas o conceito de evolução é utilizado como meio de otimização das soluções, sendo que cada problema é resolvido através da evolução dos indivíduos de uma espécie por meios de reprodução e mutação. Apesar da facilidade de integração destes algoritmos na resolução de diversos tipos de problemas, existem três componentes importantes que devem ser definidos tendo em conta o problema a lidar: a representação dos indivíduos, a função de avaliação (*fitness*) e os operadores genéticos a utilizar na otimização.

A representação dos indivíduos é feita utilizando um vetor de valores que modelam as características presentes nesse indivíduo. Estes valores representam o material genético de cada indivíduo, ou seja o seu genótipo, manifestado sobre a forma de um fenótipo que corresponde aos valores das diferentes características que esse indivíduo possui. Dependendo do algoritmo a ser usado, esta representação pode ser feita através de um vetor de valores binários, valores reais ou máquinas de estado, e a sua correta escolha na resolução de um problema influencia tanto a qualidade da solução obtida como a velocidade com que o sistema converge para valores ótimos sendo também diretamente responsável pelo espaço de armazenamento ocupado pela população.

Após correta representação dos indivíduos é necessário determinar a qualidade de cada um, utilizando para isso funções de *fitness* que são responsáveis por quantificar numericamente a semelhança das características que cada indivíduo apresenta em relação a um determinado objetivo. Estas funções dependem em larga escala do problema que se pretende abordar e das características avaliadas, devendo ser implementadas tendo em conta a qualidade e resolução pretendidas nas soluções.

O último componente a definir é responsável pela evolução dos indivíduos da população constituindo um conjunto de operadores que atuam sobre o genótipo de cada um. Estes operadores simulam geralmente os mecanismos de evolução genética existentes na natureza e realizam operações de mutação, cruzamento e seleção para introduzir diversidade e aumentar a média da qualidade da população. O funcionamento intrínseco envolve parâmetros aleatórios incluindo probabilidades que afetam o número de mutações e o processo de cruzamento e seleção. A escolha dos operadores genéticos neste tipo de sistemas influencia tanto a velocidade com que as soluções evoluem como a diversidade dos indivíduos apresentados na solução.

Uma classe particular destes algoritmos, denominados de algoritmos genéticos, é apresentada na secção seguinte constituindo uma abordagem ao panorama da computação evolutiva que foi introduzida por Holland em 1975 [30]. Os algoritmos genéticos são provavelmente a técnica mais

disseminada na resolução de problemas complexos utilizando computação evolutiva devido aos resultados satisfatórios que apresentam perante um grande número de problemas.

3.1.1 Algoritmos Genéticos

Baseados no processo de seleção natural, estes algoritmos otimizam os indivíduos da população através de seleção, mutação e cruzamento das características de cada um, gerando a cada iteração uma nova população que partilha aspetos com a anterior. As características de cada indivíduo são codificadas geralmente sobre a forma de um genoma binário e idealmente apenas os melhores indivíduos de cada geração são designados para cruzamento, mantendo-se assim as características com uma melhor avaliação e introduzindo novas variações na população através de mutações. É esperável que, conforme o número de iterações aumente, a população evolua no sentido de prosperar o valor médio da avaliação dos indivíduos, imitando-se assim o processo de seleção natural em que os indivíduos mais aptos predominam na evolução de uma espécie [31].

A representação dos indivíduos da população pode ser feita usando, como já foi referido, um genoma binário em que as características de cada indivíduo são representadas sobre a forma de um vetor binário, ou utilizando um genoma com um vetor de valores reais. A escolha da representação a utilizar varia conforme o tipo de problema a resolver e deve ser tomada em conta na implementação destes sistemas. Um exemplo do fluxo de execução deste tipo de algoritmos pode ser visualizado no algoritmo 3.1.

Algoritmo 3.1: Algoritmo genético clássico

Entrada: $n_s, iter_n, p_c, p_m$

Saída: P

$P \leftarrow$ inicializa n_s indivíduos aleatoriamente;

enquanto número de iterações $< iter_n$ **faça**

 calcula valor de avaliação de cada indivíduo;

 seleção;

 cruzamento com probabilidade p_c ;

 mutação com probabilidade p_m ;

$P \leftarrow$ atualiza nova população;

fim

3.1.1.1 Avaliação

A avaliação de cada indivíduo é feita recorrendo a funções de avaliação que quantificam a aptidão desse indivíduo. Estas funções atribuem um valor numérico à aptidão que reflete a qualidade da solução representada por esse indivíduo, tomando em conta as suas características codificadas no genoma desse indivíduo.

3.1.1.2 Seleção

A geração de uma nova população deve ser feita tendo como objetivo a prosperidade do valor médio de aptidão dos indivíduos, para tal, a cada iteração são inicialmente escolhidos os indivíduos que farão parte desta tendo por base a população anterior sobre a qual é aplicada um operador de seleção. O operador de seleção é utilizado nestes algoritmos para determinar quais os indivíduos que irão constituir a próxima geração e geralmente baseia-se num esquema denominado de *roulette wheel*. Neste esquema é atribuída uma probabilidade de sobrevivência a cada indivíduo que é proporcional à sua avaliação em relação ao somatório da avaliação global da população em que este se insere, desta maneira indivíduos mais aptos apresentam uma probabilidade mais alta de continuarem a pertencer à população enquanto esta evolui. Devido ao não determinismo deste esquema são por vezes utilizadas variações que mantêm automaticamente um ou mais indivíduos ao longo das gerações, prevenindo-se assim a possibilidade de perder os indivíduos mais aptos de cada geração devido ao *roulette wheel*. Estas variações denominam-se de seleção elitista e constituem um conjunto de mecanismos utilizados no operador de seleção. Entre os mecanismos propostos, para além do *roulette wheel*, encontram-se:

Seleção por ranking Esquema semelhante ao *roulette wheel*, com a diferença de que em vez da probabilidade de seleção ser proporcional ao valor de avaliação, neste esquema é proporcional à posição em que o indivíduo se encontra na população (que está ordenada por valor de avaliação).

Seleção por diversidade São selecionados os N indivíduos mais diversos da população.

Seleção bi-classista São selecionados os Np_s melhores indivíduos e os $N(1 - p_s)$ piores indivíduos, com p_s arbitrário.

Seleção aleatória São selecionados aleatoriamente N indivíduos da população. Neste caso a seleção pode também ser elitista caso um ou mais dos melhores indivíduos sejam automaticamente selecionados.

Para além da aplicação no operador de seleção, estes mecanismos podem também ser utilizados na seleção dos indivíduos designados para os processos de mutação e cruzamento, assim como outras variações específicas para determinadas implementações encontradas na literatura destes algoritmos.

3.1.1.3 Cruzamento

Já na operação de cruzamento o objetivo é aumentar a qualidade dos indivíduos através do cruzamento das características das soluções consideradas mais aptas na iteração atual. O operador de cruzamento apresenta como principal objetivo o de subir os picos da superfície de *fitness* através da combinação das características dos indivíduos envolvidos. Geralmente são escolhidos um certo número de indivíduos e o genoma de cada um é recombinado com o genoma de um parceiro dando origem a um ou dois novos indivíduos que apresentam características em comum com os

seus descendentes. Devido a esta operação os melhores indivíduos contribuem, a cada iteração, para o aumento da média de aptidão da população correspondendo a um aumento na qualidade das soluções obtidas. Este operador apresenta diversas variações na sua implementação sendo que a utilizada mais frequentemente nestes algoritmos é a de cruzamento de um ponto. Nesta variação são selecionados dois indivíduos e a partir da combinação das características de cada um são originados dois novos indivíduos com um genoma constituído pela junção de segmentos do genoma dos indivíduos originais. A determinação dos segmentos é feita através da escolha aleatória de um ponto de corte originando quatro segmentos que são posteriormente combinados e integrados no genoma de dois novos indivíduos. Considere-se, por exemplo, os seguintes indivíduos selecionados para cruzamento e respetivos segmentos determinados:

$$x_i = \{0, 1, 1, 0, 1, 1, 1, 0, 0\}$$

$$x_j = \{1, 0, 1, 0, 0, 1, 0, 1, 1\}$$

Utilizando cruzamento de um ponto, os genomas dos novos indivíduos serão:

$$x'_i = \{1, 0, 1, 0, 1, 1, 1, 0, 0\}$$

$$x'_j = \{0, 1, 1, 0, 0, 1, 0, 1, 1\}$$

Tal como na mutação, o cruzamento do material genético deve respeitar a validade da representação no espaço de pesquisa, restringindo a posição em que o corte é efetuado de maneira a que uma determinada característica não apresente valores errados. Para além do cruzamento de um ponto, foram também propostos:

Cruzamento de dois pontos Neste tipo de cruzamento são escolhidas duas posições para o corte dos segmentos, originando três segmentos em cada indivíduo.

Cruzamento uniforme Os valores do genoma dos novos indivíduos são escolhidos de um dos dois genomas originais segundo uma determinada probabilidade. Neste caso a combinação de valores por oposição à utilização de segmentos originais produz indivíduos com uma maior diversidade de características devido à operação de cruzamento não depender da posição dos valores no genoma.

Em termos de desempenho não existe nenhuma diferença significativa entre as abordagens propostas. No entanto, algumas são particularmente eficientes numa determinada classe de problemas.

Sucintamente, a aplicação de um algoritmo genético na resolução de um problema complexo é um processo adaptativo que deve ter em conta tanto as características que constituem uma solução desejada como o processo que melhor se adequa para as obter. Para além dos mecanismos utilizados para a evolução da população, a avaliação das mesmas é a componente que mais contribui na eficácia destes algoritmos deduzindo-se assim a importância de uma correta implementação destas funções num sistema de computação evolutiva.

3.1.1.4 Mutação

Para aumentar a diversidade e qualidade dos indivíduos são usadas operações de mutação e cruzamento que operam sobre o genoma de cada um. Na operação de mutação são introduzidas alterações aleatórias no genoma originando modificações nas características que constituem cada indivíduo, estas alterações podem ser vantajosas ou não mas contribuem significativamente para o aumento do espaço de combinações abrangidas pelo algoritmo no processo de pesquisa. O operador de mutação é responsável pela exploração de novas regiões promissoras da superfície de *fitness*. Esta operação ocorre geralmente com uma probabilidade baixa e corresponde a uma inversão do valor de um gene num ponto aleatório do genoma do indivíduo. Assumindo como exemplo um indivíduo com o seguinte genoma binário:

$$x_i = \{0, 1, 1, 0, 1, 1, 1, 0, 0\}$$

Uma mutação possível seria:

$$x'_i = \{0, 1, 0, 0, 1, 1, 1, 0, 0\}$$

É importante referir que a operação de mutação, quando ocorre, deve ter em conta a validade do novo indivíduo, ou seja, o novo genoma originado da mutação não deve conter uma representação que seja inválida ao espaço de pesquisa. Este controlo pode ser efetuado quer através de restrições no operador de mutação, quer através de funções que validem posteriormente os novos indivíduos originados por esta operação. Devido à aleatoriedade deste operador garante-se a introdução de características inovadoras na população ao longo do processo de evolução.

3.2 Sistemas Imunitários Artificiais

O sistema imunitário biológico é constituído por um conjunto de células, moléculas e órgãos que atuam colaborativamente no processo de manutenção do equilíbrio dinâmico interno de um ser vivo [32]. A complexidade destes sistemas é patente na variedade das características que emergem deste, desde a capacidade de distinção de elementos externos ao sistema até ao processo de adaptação e aprendizagem por meios iterativos e evolutivos, incluindo também mecanismos de comunicação entre os vários componentes imunitários que despoletam a ativação ou inibição destes últimos. Esta complexidade inerente dos sistemas imunitários é necessária para a correta realização do seu objetivo primário: proteger o correspondente sistema biológico de substâncias infecciosas que podem prejudicar o normal funcionamento do mesmo [33].

Para a realização dessa proteção o sistema imunitário utiliza mecanismos de imunidade inata e imunidade adaptativa. A imunidade inata baseia o reconhecimento de elementos externos através de informação herdada geneticamente pelo indivíduo e apresenta uma resposta constante e imediata permitindo uma rápida deteção destes elementos. Este mecanismo não apresenta características adaptativas que permitam uma aprendizagem no reconhecimento de novos elementos

externos. No entanto, oferece um contributo significativo na tarefa de discriminação entre elementos pertencentes ou desconhecidos ao indivíduo.

Já na imunidade adaptativa é utilizado um processo iterativo que melhora sucessivamente a resposta a elementos externos através de evoluções nas células responsáveis pelo reconhecimento, permitindo a deteção e aprendizagem de novos elementos não abrangidos pelo mecanismo da imunidade inata. As células que constituem este mecanismo são denominadas de linfócitos e podem ser divididas em células do tipo B e tipo T. Ambas desenvolvem recetores nas suas membranas celulares que são capazes de identificar e de se combinar com moléculas que tenham uma forma ou características semelhantes à do recetor, esta semelhança corresponde ao grau de afinidade e faz parte integrante destes sistemas o objetivo de aumentar este valor através de adaptação e evolução dos recetores presentes nos linfócitos. A geração destes recetores é feita aleatoriamente sendo que o processo de seleção dos mesmos é realizado por um mecanismo de seleção clonal responsável por escolher os recetores que apresentem características particulares, passíveis de possuírem um elevado grau de afinidade com um determinado elemento [34]. Os elementos detetados pelos recetores presentes nos linfócitos variam quanto ao tipo de célula a que estes pertencem: no caso das células do tipo B os recetores destinam-se à identificação de elementos não pertencentes ao indivíduo, enquanto que, no caso das células do tipo T os recetores existentes são responsáveis pela identificação de elementos internos ao sistema imunitário especializados na interação com elementos desconhecidos. As moléculas que são identificadas pelos recetores destas células definem-se como substâncias antigénicas e cada uma apresenta diversos epítomos, correspondentes às regiões sobre as quais os recetores se combinam.

A identificação de antígenos pelas células de tipo B opera através da utilização de anticorpos. Os anticorpos são proteínas produzidas e segregadas por estas células em resposta a uma deteção de um elemento antigénico. Estas proteínas reconhecem e combinam-se com um determinado antigénio sendo usadas no sistema imunitário como meio de sinalização entre as diferentes partes responsáveis pela remoção do antigénio. Cada célula do tipo B produz anticorpos diferentes com afinidade destinada a um antigénio específico.

As células de tipo T apresentam uma função mais específica no sistema imunitário e são alvo de um processo constante de otimização e maturação consoante o encontro de novos antígenos. O funcionamento destas células colmata o processo despoletado quando os anticorpos se combinam com um determinado antigénio, induzindo a proliferação de novas células capazes de remover o antigénio detetado. As células de tipo T apresentam também características de memorização das respostas aos antígenos detetados permitindo uma evolução gradual na eficácia deste mecanismo.

A imunidade inata e imunidade adaptativa operam em conjunto no funcionamento do sistema imunitário num processo complexo e dependente de vários mecanismos que foram inicialmente formalizados em meados do século 20. Inicialmente proposta por Burnet [34], a teoria da seleção clonal é atualmente aceite como a melhor explicação de como o sistema imunitário responde à invasão de elementos externos. Em conjugação com a teoria da maturação de afinidade é possível definir os principais mecanismos necessários ao funcionamento de um sistema imunitário adaptável e evolutivo.

3.2.1 Seleção Clonal e Maturação de Afinidade

A teoria de seleção clonal proposta por Burnet [34] postula a ideia que, de entre as células pertencentes ao sistema imunitário, apenas aquelas com capacidade para reconhecer antígenos irão proliferar e as restantes serão suprimidas por um mecanismo de seleção. A seleção clonal atua sobre as células B e T, promovendo ou suprimindo a diferenciação de clones destas células segundo o grau de afinidade que apresentam perante antígenos [32]. Durante este processo de clonagem é possível a ocorrência de erros arbitrários, resultando em mutações que influenciam o valor de afinidade com o antígeno reconhecido. Algumas destas mutações originam células com um elevado grau de afinidade, proporcionando ao sistema imunitário uma resposta mais especializada para determinados antígenos. Segundo o mecanismo de seleção clonal, os clones que apresentam um elevado grau de afinidade para com um determinado antígeno serão selecionados e diferenciados em células de memória para referência futura enquanto os restantes serão suprimidos, promovendo-se assim apenas a maturação de células com utilidade ao sistema. Desta maneira, o sistema imunitário evolui o seu mecanismo de defesa iterativamente, adaptando-se a novos antígenos previamente desconhecidos assim que estes são detetados. Através de memorização, uma resposta mais rápida pode ser efetuada quando um antígeno já conhecido é detetado por este sistema.

Utilizando esta teoria, De Castro e Zuben propõem em [35] um algoritmo de aprendizagem e otimização baseado nas técnicas de seleção clonal denominado CLONALG. Inicialmente, CLONALG gera um certo número de anticorpos aleatórios, correspondendo cada um a uma solução possível do problema a otimizar. Através de seleção, clonagem e mutação os anticorpos existentes são iterativamente melhorados de forma a otimizar a população existente.

3.2.2 Seleção Negativa

A seleção negativa é utilizada no sistema imunitário para a deteção e supressão de células cujos recetores reconhecem erradamente elementos internos como antígenos. Este mecanismo inspirou o desenvolvimento de algoritmos aplicados principalmente na deteção de anomalias e padrões. Assim, tal como no sistema imunitário são suprimidas células com recetores que reconhecem elementos internos, nestes algoritmos as soluções que pertençam a um determinado conjunto são descartadas [36].

3.2.3 Teoria do Perigo

A Teoria do Perigo, proposta por Matzinger em [37], define a resposta imunitária como um processo reativo que é ativado quando é detetado dano nas células do indivíduo. A deteção de dano, e conseqüente ativação de mecanismos de resposta, é feita através da existência de sinais de auxílio que são segregados por células suprimidas devido a causas não naturais. As aplicações desta teoria incluem principalmente mecanismos de deteção de anomalias, como por exemplo vírus informáticos, transações eletrónicas forjadas ou falhas de *hardware* [38].

3.2.4 aiNet

A aiNet (*artificial immune network*) proposta por De Castro e Von Zuben em [13], é baseada num sistema imunitário artificial combinando o algoritmo CLONALG [35] com a teoria da rede imunitária [39]. Nesta abordagem, os diferentes anticorpos constituem uma rede imunitária em que a clonagem e seleção de cada um é feita proporcionalmente ao seu valor de aptidão e dependendo da interação desse anticorpo com a rede imunitária [40]. Nesta implementação é utilizado também um limiar do grau de afinidade para determinar as células a suprimir em cada iteração.

Os resultados do algoritmo criado permitem reduzir a redundância na descrição de estruturas de dados, incluindo a distribuição espacial e relações entre agrupamentos (*clusters*). A técnica de *clustering* é especialmente útil em tarefas de exploração de padrões incluindo segmentação de imagem, classificação automática de padrões e *data mining*. Uma descrição mais detalhada do funcionamento deste algoritmo é apresentada na secção 5.5.

Capítulo 4

Formalização

A complexidade do problema de orquestração musical está patenteada na variedade de interações multidisciplinares necessárias para uma correta formalização analítica deste. Desde a definição de uma função objetivo que capture as características perceptivas dos fenómenos sonoros, até à implementação de um algoritmo inspirado por processos biológicos que otimize o espaço combinatório existente, vários aspetos devem ser considerados, tanto numa perspetiva científica e pragmática como numa perspetiva artística e criativa.

A definição formal deste problema foi feita com base no trabalho desenvolvido por Carpentier em [18], [17] e [3], sendo descrita neste capítulo. Inicialmente é feita uma apresentação teórica do problema de orquestração e uma breve descrição da base de dados sonora utilizada, seguindo-se uma enumeração dos descritores utilizados na caracterização e distinção dos diferentes sons presentes na base de dados. Por fim são apresentadas funções de combinação e funções de avaliação de similaridade entre sons. As funções de combinação são responsáveis por calcular os novos valores dos descritores de uma determinada combinação de sons, já as funções de avaliação de similaridade são responsáveis por estimar a similaridade entre uma combinação e um som objetivo. Através da utilização deste conjunto de funções avaliam-se assim as soluções propostas pelos algoritmos de pesquisa.

4.1 Modelo Teórico da Orquestração Musical

O objetivo da orquestração musical com recurso a ferramentas computacionais é o de auxiliar o utilizador na tarefa de combinar autonomamente os registos tímbricos de diferentes instrumentos de forma a que o resultado final se aproxime, ou mesmo se iguale, a um timbre designado pelo utilizador. Nestes problemas pretende-se encontrar uma combinação de sons que, quando reproduzidos simultaneamente, se assemelhem o mais possível a um som objetivo pré-especificado. Os sons que podem ser utilizados para as combinações englobam as notas reproduzidas pelos diferentes instrumentos encontrados tipicamente numa orquestra musical. O problema de orquestração

torna-se desta maneira num processo de pesquisa por sons, representando as diferentes notas de diferentes instrumentos musicais, sons esses que quando combinados no mesmo instante de tempo resultem num timbre próximo do objetivo. No contexto deste trabalho assume-se que o timbre desejado não evolui com o tempo, apresentando assim um comportamento estático. Simplifica-se desta maneira significativamente o problema de orquestração uma vez que se descarta a componente temporal, não se considerando a evolução temporal dos sons. Tal simplificação foi necessária para tornar exequível a implementação do sistema proposto no espaço de tempo disponível.

As soluções existentes para este problema recorrem a uma base de dados contendo numerosos sons que representam as diferentes notas dos diferentes instrumentos tipicamente presentes numa orquestra. Estes sons são processados e posteriormente combinados pelo algoritmo de pesquisa, sendo que as combinações resultantes devem ser julgadas pelo algoritmo e apresentadas ao utilizador caso sejam relevantes. As combinações resultantes podem conter um ou mais sons de diferentes instrumentos e faz parte do objetivo destes sistemas a implementação de funções que avaliem o quão próximo do som pretendido uma determinada combinação se encontra. Considerando as diferentes combinações possíveis é esperado um espaço de pesquisa de tamanho elevado. Como exemplo, uma base de dados com 1000 sons e uma orquestração utilizando 4 sons diferentes cria um espaço de pesquisa com um número de combinações possíveis igual a:

$$\binom{1000}{4} = \frac{1000!}{4!(1000-4)!} = 4.14 \times 10^{10} \quad (4.1)$$

Ou seja, mais de 41 mil milhões de combinações diferentes. Este vasto espaço de pesquisa pode ainda ser exponencialmente aumentado se forem considerados os diferentes estilos de interpretação musical e as diferentes sonoridades que cada instrumento pode apresentar, originando um número de combinações praticamente ilimitado conforme a base de dados é aumentada. O objetivo dos algoritmos de pesquisa destes sistemas reside essencialmente em percorrer eficientemente o vasto espaço de pesquisa existente, de forma a encontrar rapidamente soluções ótimas e úteis ao utilizador.

Para a realização desta tarefa de uma forma autónoma é necessária, por um lado, uma forma de descrição analítica das principais características de cada som, e por outro uma definição de julgamento estético que permita a uma máquina avaliar a similaridade entre sons. A correta escolha do conjunto de características utilizadas na descrição de cada som permite a um sistema deste género lidar com certas dimensões do timbre utilizando valores analíticos, quantificando-se desta maneira determinadas características percetivas. Apesar de na literatura não existir concordância sobre quais as principais dimensões do timbre, existem dados que corroboram alguns dos descritores escolhidos como estando altamente relacionados com o fenómeno do timbre [17].

É necessária também a definição de funções de combinação que estimem as características conjuntas de uma determinada combinação de sons. Estas funções de combinação devem estimar as características partindo apenas do conhecimento do conjunto de sons que a compõem, permitindo ao algoritmo de pesquisa o cálculo das características de combinações utilizando apenas os sons individuais que integram as mesmas, sem haver a necessidade de combinar os sons indivi-

duais e calcular as suas correspondentes características conjuntas. Por fim, é imprescindível a definição de funções de avaliação de similaridade, funções estas que quantificam numericamente o quão próximo do som objetivo uma determinada combinação se encontra.

Num ponto de vista mais formal, a orquestração musical pretende aproximar um determinado som s_T , utilizando seleção e posterior combinação de sons existentes numa base de dados contendo N amostras de sons $(s_n)_{1 \leq n \leq N}$. O objetivo é encontrar um conjunto de sons C em $\{1, \dots, N\}$, tal que a combinação dos mesmos $(s_n)_{n \in C}$ seja o mais semelhante possível a s_T .

Cada som é caracterizado por K descritores que são responsáveis por quantificar analiticamente uma determinada característica percetiva desse som. Estes descritores são calculados para cada som existente na base de dados recorrendo a funções de estimação $\{F^1, \dots, F^K\}$.

Os valores destes descritores são também utilizados nas funções de combinação que são responsáveis por estimar o novo valor dos descritores perante uma combinação de diferentes sons. Estas funções de combinação recorrem aos diferentes valores dos descritores dos sons que compõem uma combinação para determinar analiticamente os valores dos descritores da correspondente combinação, auxiliando o funcionamento do algoritmo de pesquisa no que toca à seleção de soluções úteis ao utilizador.

A semelhança entre sons é caracterizada por funções de similaridade $\{D_T^1, \dots, D_T^K\}$, que quantificam numericamente a distância entre os valores dos diferentes descritores de cada som ou combinação de sons em relação aos valores dos descritores do som objetivo s_T . Os valores destas funções são utilizados como medida de aptidão dos indivíduos que compõem a população do algoritmo de pesquisa. A correta implementação das funções de similaridade e de combinação influencia assim em larga escala a qualidade das soluções obtidas neste tipo de sistemas [3].

A orquestração musical pode ser então definida como um problema de otimização combinatória em que o espaço de pesquisa S é avaliado através de funções de similaridade que quantificam a distância entre uma determinada combinação C e o som objetivo T , de forma a que:

$$\begin{cases} \min D_T^k(C) & k \in \{1, \dots, K\} \\ \text{com } C \in S \end{cases} \quad (4.2)$$

Ou seja, pretende-se encontrar uma combinação C pertencente ao espaço de pesquisa S que minimize o valor das funções de similaridade $\{D_T^1(C), \dots, D_T^K(C)\}$. As funções de similaridade implementam então uma métrica que permite a um sistema de orquestração comparar sons segundo valores analíticos, correspondentes aos valores destas funções. Modela-se assim as multi dimensões do timbre segundo um espaço euclidiano de forma a que sons mais díspares apresentem distâncias mais altas e conseqüentemente valores mais baixos de similaridade. Desta forma o problema de orquestração torna-se um problema de otimização combinatória em que se pretende minimizar as distâncias calculadas utilizando as funções de similaridade.

4.2 Base de Dados Sonora

A base de dados sonora utilizada nestes sistemas contém os diversos sons que podem ser utilizados no processo de orquestração musical. Estes sons incluem as diferentes notas e as diferentes dinâmicas produzidas por os instrumentos de uma orquestra. Neste trabalho são utilizadas 3 dinâmicas e 13 instrumentos na implementação da base de dados sonora. Os sons utilizados pertencem a uma coleção denominada de *RWC Music Database* [41], que contém, na sua forma original, 50 instrumentos interpretados por músicos profissionais. Estes sons foram selecionados e processados para a implementação do sistema de orquestração musical. Todos os sons apresentam a mesma frequência de amostragem e duração temporal variante, sendo que todos apresentam uma duração de pelo menos 1 segundo. A especificação dos instrumentos que compõem a base de dados utilizada e uma descrição mais detalhada do processamento efetuado sobre os sons da mesma será feita no capítulo 5.

4.3 Descritores

Cada som presente na base de dados é diferente e a sua combinação com outros sons pode ou não originar uma solução próxima da pretendida. Para distinguir cada som e saber qual a sua contribuição numa determinada combinação é necessário caracteriza-lo analiticamente, atribuindo valores numéricos a certos atributos desse som. Esta caracterização permite a avaliação da similaridade entre uma determinada combinação e o som objetivo. Uma lista de possíveis descritores que são utilizados na classificação de sons pode ser encontrada em [42].

No contexto deste projeto foram selecionados quatro descritores que são utilizados na caracterização dos sons presentes na base de dados. Os descritores selecionados foram:

1. Centróide espectral (ce)
2. Espalhamento espectral (ee)
3. Intensidade sonora (is)
4. Valores de frequência e amplitude dos principais picos espectrais (P)

Estes quatro descritores são utilizados no algoritmo de pesquisa recorrendo a funções de similaridade para avaliar a semelhança entre sons. Para além destes descritores foram também calculados para cada som o valor de RMS (*Root Mean Square*), o valor de energia e o valor da frequência fundamental desse som (f_0). Na figura 4.1 são enumerados os quatro descritores que são calculados para cada som.

O centróide espectral (ce) e o espalhamento espectral (ee) representam valores numéricos que caracterizam o espectro de cada som segundo o seu valor médio e desvio padrão. O cálculo destes valores é feito durante a parte em que o sinal se encontra estável, por simplicidade considerou-se que essa estabilidade se encontra entre os 250ms e 750ms do sinal. Para a obtenção dos valores referidos assumiu-se que o tempo de *ataque* de um som é inferior a 250ms e todos os sons têm

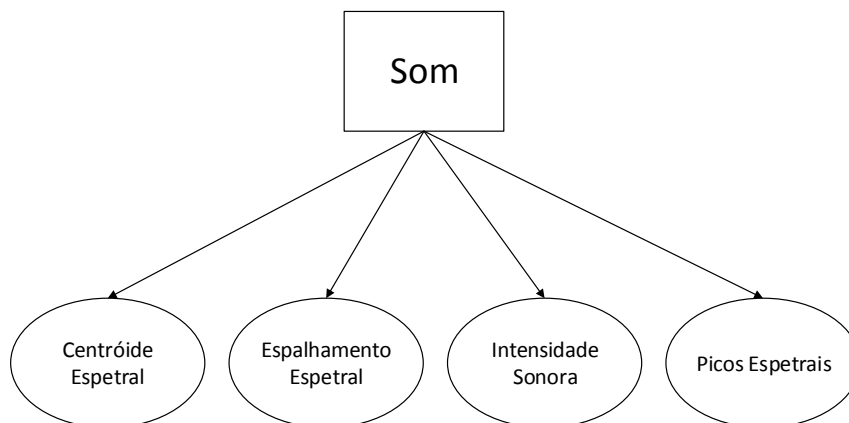


Figura 4.1: Descritores sonoros avaliados para cada som da base de dados

duração superior a 750ms. Na figura 4.2 é apresentada a forma de onda de um som e a evolução do valor normalizado de alguns descritores com o tempo. Os valores dos descritores considerados para a caracterização de um som correspondem à média dos valores na parte estável do sinal, apresentados na figura 4.3. O valor de RMS (*Root Mean Square*) é também utilizado como referência da energia média do sinal.

A intensidade sonora (*is*) caracteriza a intensidade de cada som. Este atributo permite a um ouvinte humano distinguir diversos sons segundo uma escala de fraco a forte e está principalmente relacionado com a pressão sonora, forma de onda do som e distribuição das frequências [43].

Já os picos espectrais (*P*) relacionam-se com as características espectrais de cada som e são representados por um vetor numérico contendo as frequências e amplitudes dos principais picos espectrais de cada som, ou seja o conjunto de frequências mais dominantes desse som, que são obtidas através da utilização de transformadas discretas de Fourier. Após ser feito o cálculo do espectro de cada som são selecionadas as frequências com maior amplitude, sendo que o valor de amplitude destas frequências é guardado para posteriormente tentar-se fazer uma igualização entre os picos espectrais dos sons de uma combinação e os picos espectrais do som objetivo.

O valor de energia do sinal é necessário para a correta implementação das funções de combinação, já o valor da frequência fundamental permite uma seleção mais cuidadosa dos sons que compõem o espaço de pesquisa, removendo deste espaço sons que apresentem frequências fundamentais mais baixas que a do som objetivo.

Apresenta-se a seguir a forma como é efetuado o cálculo do valor de cada característica referida.

4.3.1 Centróide Espetral

O centróide espectral (*ce*) caracteriza o espectro de um som segundo o seu centro de massa [42]. Neste trabalho, o seu cálculo é feito utilizando o quadrado do valor absoluto das amplitudes do

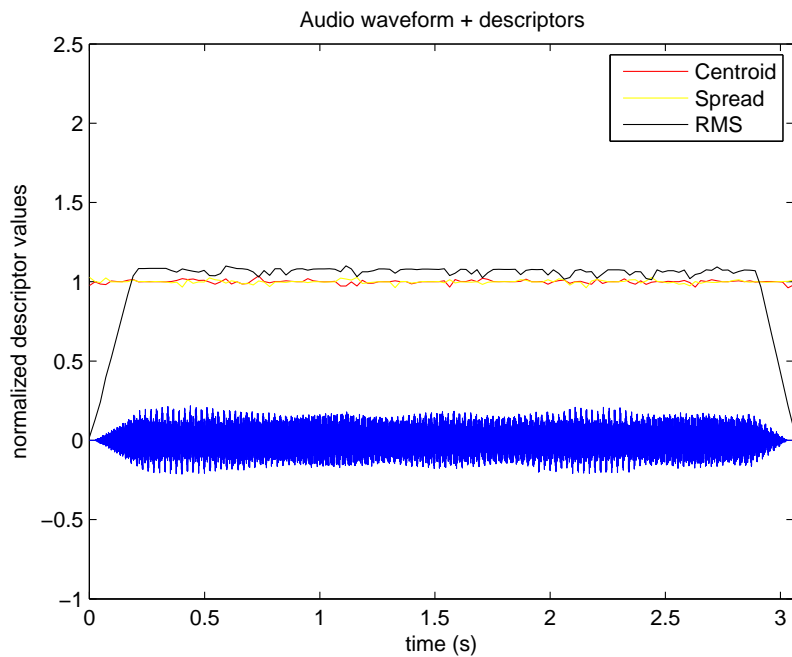


Figura 4.2: Forma de onda de um som e a evolução dos seus descritores com o tempo

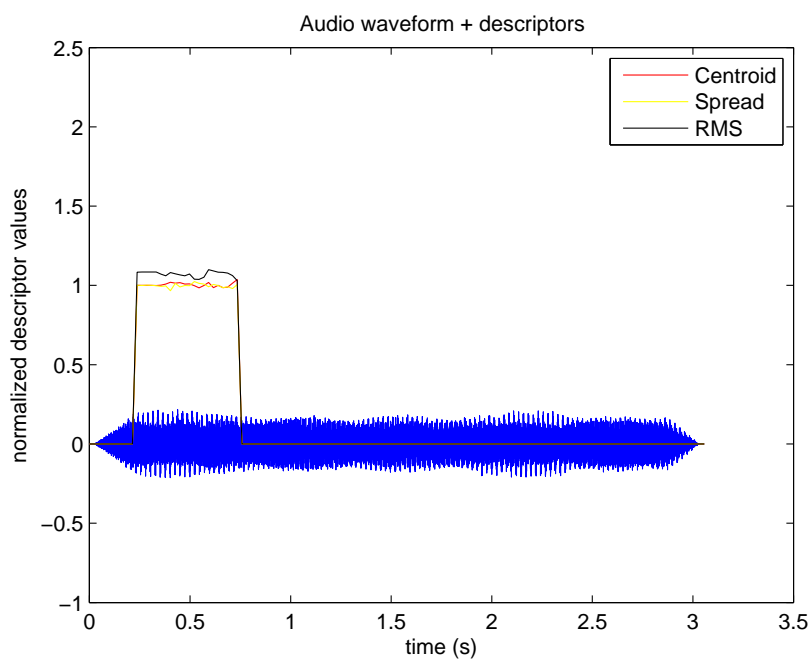


Figura 4.3: Forma de onda de um som e a evolução dos descritores entre 250ms e 750ms

espectro desse som. Estes valores são considerados como uma distribuição em que o valor de probabilidade de ocorrência de uma frequência corresponde ao valor da sua amplitude, normalizada segundo o somatório dos valores de amplitude desse espectro de energia. O centróide espectral é

então definido como o valor médio dessa distribuição.

Se f corresponder ao vetor de frequências do espectro de cada som, $p(f)$ à probabilidade de ocorrência de uma determinada frequência, e $a(f)$ ao valor de amplitude espectral da mesma, então o centróide espectral é calculado como:

$$ce = \mu = \int f p(f) df, \text{ com} \quad (4.3)$$

$$p(f) = \frac{|a(f)|^2}{\sum_f |a(f)|^2} \quad (4.4)$$

Este descritor é frequentemente associado ao *brilho* de um som estando relacionado, segundo McAdams, com uma das dimensões do timbre [14].

4.3.2 Espalhamento Espectral

O espalhamento espectral (ee) corresponde à dispersão do espectro segundo o seu valor central, sendo este último calculado utilizando a equação 4.3. O valor desta característica é definido como a variância da distribuição espectral [42]. No entanto, neste trabalho é utilizado a raiz quadrada desse valor de variância, correspondendo ao desvio padrão da distribuição espectral. O seu cálculo é feito com:

$$ee^2 = \sigma^2 = \int (f - \mu)^2 p(f) df \quad (4.5)$$

4.3.3 Intensidade Sonora

A intensidade sonora (is) está relacionada com o nível de pressão sonora, frequência e forma de onda de um som [43]. Neste trabalho o seu cálculo é feito considerando os valores de amplitude dos picos espectrais. Considerando um vetor p_e com os valores de amplitudes dos picos espectrais então o cálculo de intensidade sonora é feito através de:

$$is = 20 \log_{10} \left(\sum_i p_e(i) \right) \quad (4.6)$$

4.3.4 Picos Espectrais

A análise espectral de um som permite a identificação das contribuições relativas de cada frequência na forma final desse som. Os picos espectrais correspondem a frequências dominantes no espectro do som a ser analisado e são representados como máximos locais no espectro desse som. No contexto musical estes picos correspondem aos harmônicos gerados quando uma nota de um instrumento é reproduzida. Uma vez identificado cada pico, é possível caracterizar cada som segundo o seu conteúdo espectral. Para além disso, os picos encontrados na análise espectral de cada som permitem uma identificação da sua harmonia, destacando os valores que correspondem a uma determinada nota.

O cálculo do espectro é feito em decibéis sendo que para a análise dos picos espectrais é selecionada apenas uma gama de valores entre o máximo absoluto do espectro e -35 dB desse valor. Um exemplo do espectro de um som é apresentado na figura 4.4, com o limite da gama de valores selecionados assinalado na linha a tracejado. A correspondente gama de valores selecionados é apresentado na figura 4.5 e os picos espectrais identificados são apresentados na figura 4.6.

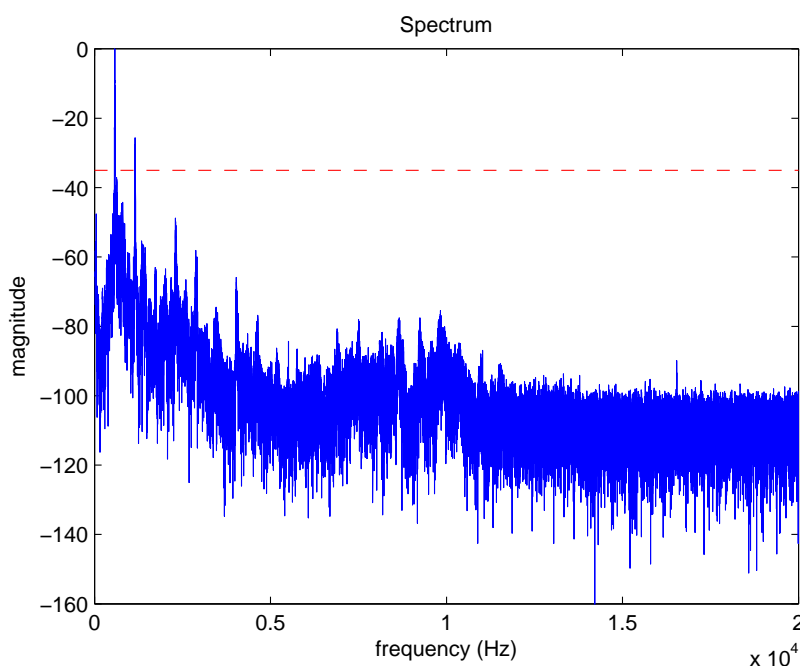


Figura 4.4: Espectro em decibéis e limite da gama de valores selecionados para o cálculo de picos espectrais a tracejado

Para cada gama de valores selecionados do espectro são analisados e identificados os picos que apresentam um valor de amplitude mais alto. A identificação de picos é feita através de comparação dos múltiplos valores do espectro correspondente. Um determinado máximo local é considerado pico se a diferença de amplitudes entre os valores mínimos que o antecedem e precedem forem mais altos que um valor de limiar. Este valor de limiar corresponde a 5% da distância entre o máximo e o mínimo absoluto da gama de valores selecionados do espectro. São identificados no máximo 25 picos espectrais em cada som. Uma identificação dos picos espectrais do espectro da figura 4.4 é apresentada na figura 4.6.

Estes picos são utilizados na seleção de sons que podem fazer parte da solução do problema de orquestração. Efetivamente no algoritmo desenvolvido são inicialmente identificados n_T picos espectrais do som objetivo e é de seguida feita uma pesquisa na base de dados por sons que apresentem pelo menos um pico em comum com o objetivo. Os sons que apresentem picos em comum são selecionados para fazerem parte do espaço de pesquisa. O critério de seleção baseia-se num valor relativo em relação à frequência do pico do som objetivo (2.5%). Se o som objetivo apresentar, por exemplo, um máximo espectral na frequência de 1000Hz, então são selecionados para o

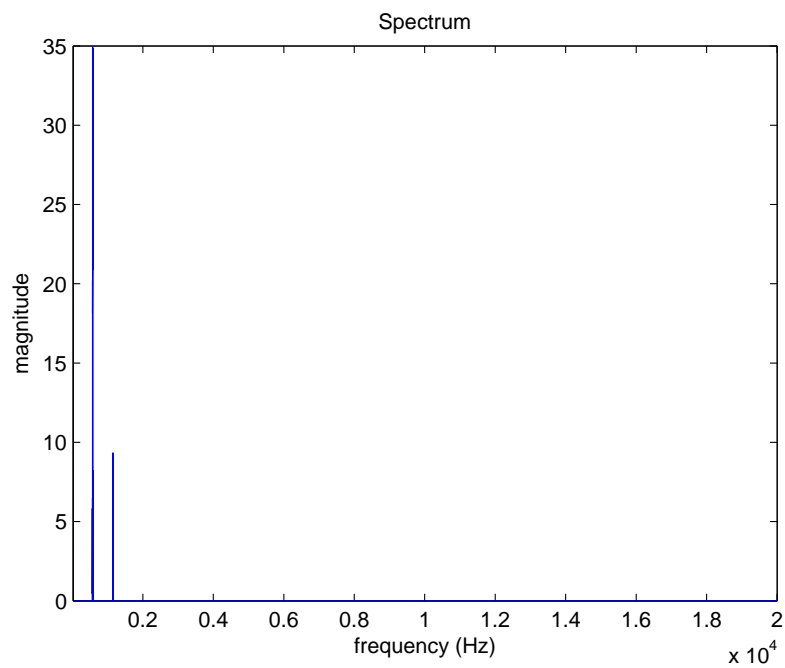


Figura 4.5: Gama de valores selecionados do espectro apresentado na figura 4.4

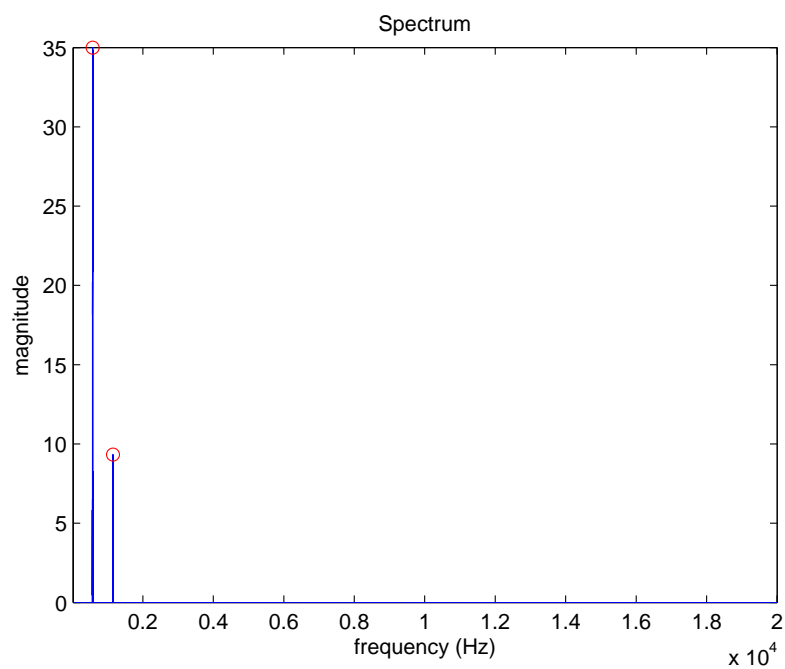


Figura 4.6: Picos espectrais identificados

processo de pesquisa sons que apresentem máximos espectrais entre 975 e 1025 Hz, aplicando-se este processo aos restantes picos presentes no som objetivo. Nas figuras 4.7, 4.8 e 4.9 pode ser observado um exemplo deste processo de seleção. O espectro de um som objetivo é apresentado na

figura 4.7. Este espectro apresenta picos espectrais nas frequências 500, 2500 e 4000 Hz. Na figura 4.8 é apresentado um espectro de um som selecionado para o processo de pesquisa. Este som foi selecionado por apresentar um pico espectral em comum com o som objetivo na frequência 500 Hz. Já na figura 4.9 é apresentado um espectro de um som que não foi selecionado para o processo de pesquisa por não apresentar picos espectrais em comum com o som objetivo.

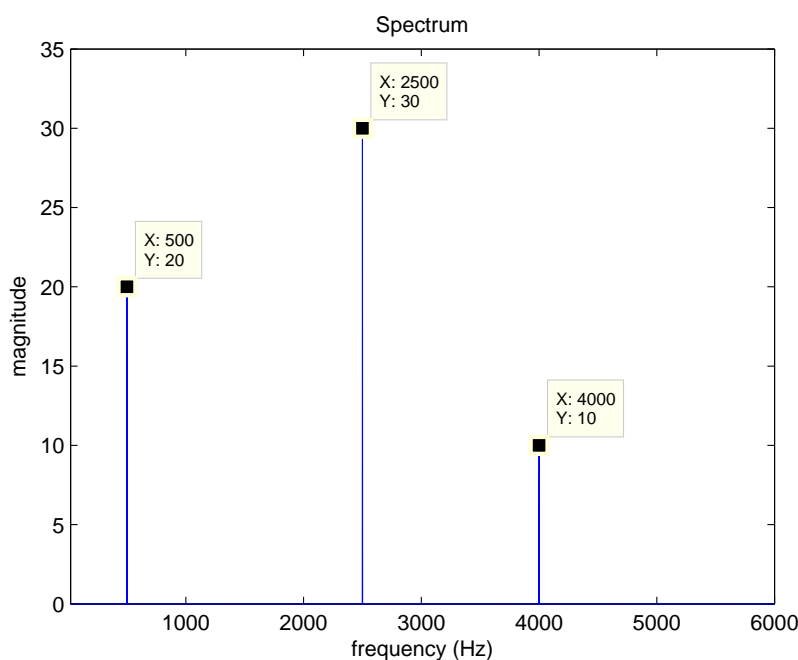


Figura 4.7: Espectro de um som objetivo

Após a seleção dos sons que apresentem picos espectrais em comum com o som objetivo é construído um vetor com a contribuição de cada som nos picos do som objetivo. Este vetor permite avaliar a similaridade entre o espectro de diferentes sons e o som objetivo através da comparação da posição em que os picos espectrais se encontram. Assim, para cada som selecionado para o processo de pesquisa é associado um vetor de tamanho n_T cujos valores representam a contribuição para cada pico, caso aplicável. Cada contribuição corresponde ao valor da amplitude do pico desse som que é comum ao pico do som objetivo, ou zero caso não exista um pico em comum.

Sejam $\{f_n^T\}$ as frequências dos picos do som objetivo e seja s_i um som da base de dados com picos espectrais $\{f_j^i\}$ e correspondentes amplitudes $\{a_j^i\}$, então o vetor de contribuições construído será:

$$P_{s_i}^T(n) = \begin{cases} a_j^i & \text{se, } (1 + \delta)^{-1} \leq f_j^i / f_n^T \leq 1 + \delta \\ 0 & \text{em caso contrário} \end{cases} \quad (4.7)$$

Em que δ é o valor relativo de seleção, igual a 2.5%. Este vetor de contribuições é construído para cada som da base de dados selecionado e reflete a semelhança espectral entre o espectro desse

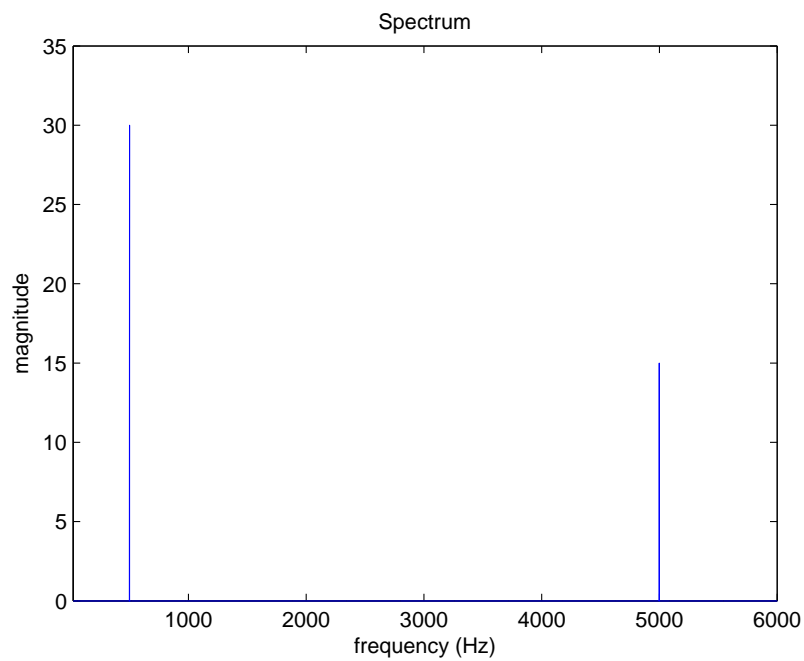


Figura 4.8: Espectro de um som selecionado para o processos de pesquisa. O som apresenta um pico espectral em comum com o som objetivo (500 Hz)

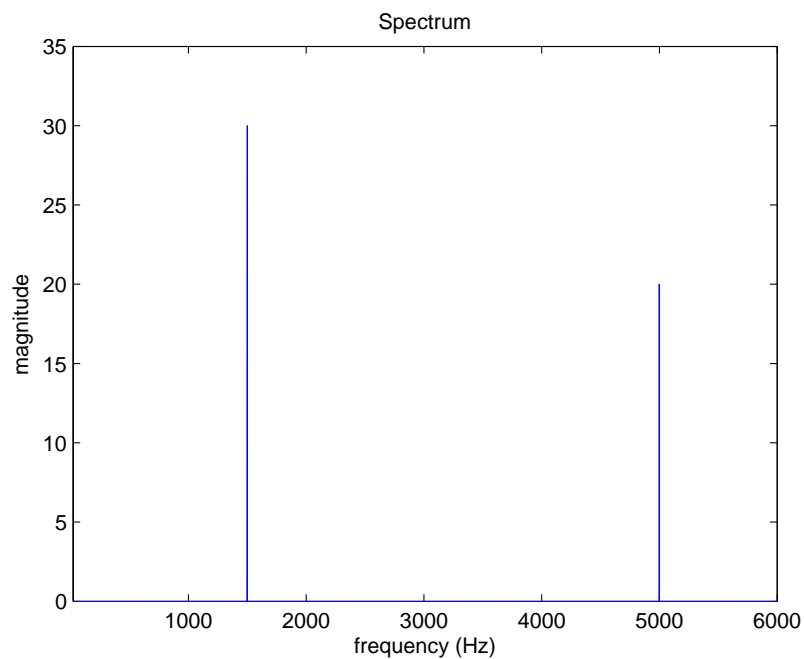


Figura 4.9: Espectro de um som não selecionado para o processo de pesquisa. O som não apresenta nenhum pico espectral em comum com o som objetivo

som e o espectro do som objetivo. Reflete também a contribuição que cada som providencia para uma determinada zona do espectro.

Como exemplo considere-se um som objetivo com picos espectrais nas frequências f^T e um som da base de dados com picos espectrais nas frequências f^i e correspondentes amplitudes a^i , tal que:

$$\begin{aligned} f^T &= \{500, 2000, 3500, 4000\} \\ f^i &= \{500, 1000, 3500, 4000, 5000\} \\ a^i &= \{30, 25, 15, 20, 10\} \end{aligned}$$

Então o vetor de contribuições construído será igual a:

$$P_{s_i}^T = \{30, 0, 15, 20\} \quad (4.8)$$

Este processo de construção do vetor de contribuições pode ser observado nas figuras 4.10, 4.11 e 4.12. Na figura 4.10 são apresentados os picos espectrais de um som objetivo. Os picos espectrais de um som selecionado para o processo de pesquisa são apresentados na figura 4.11 e o seu respectivo vetor de contribuições é apresentado na figura 4.12.

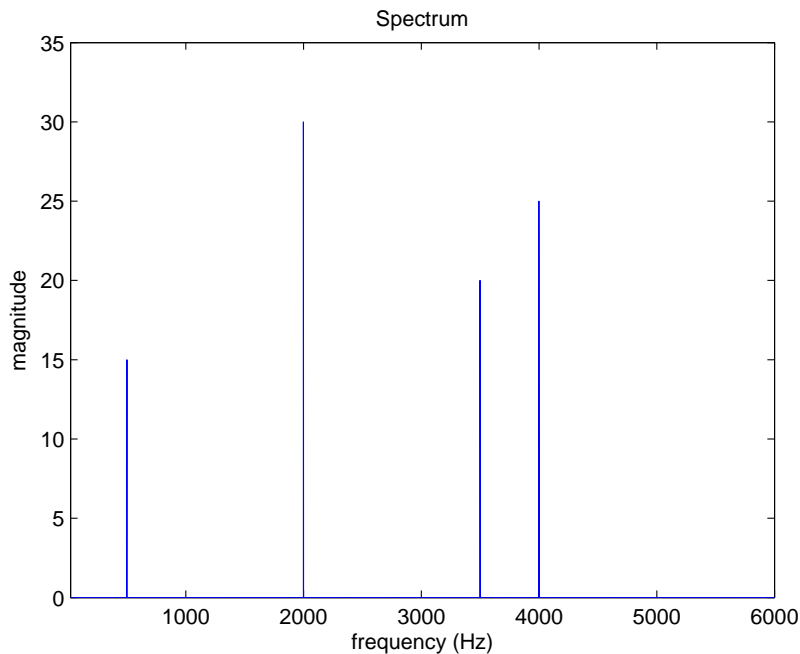


Figura 4.10: Espectro de um som objetivo

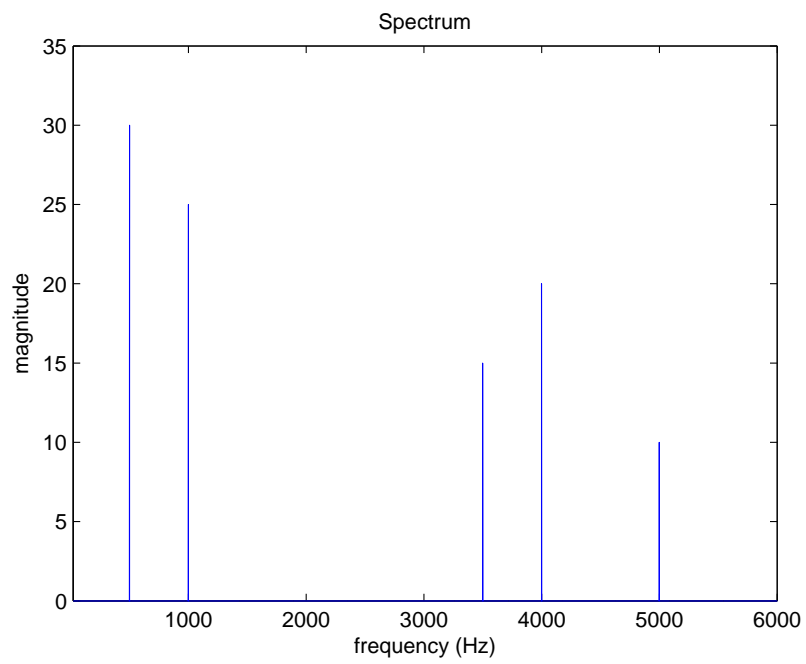


Figura 4.11: Espectro de um som selecionado para o processos de pesquisa

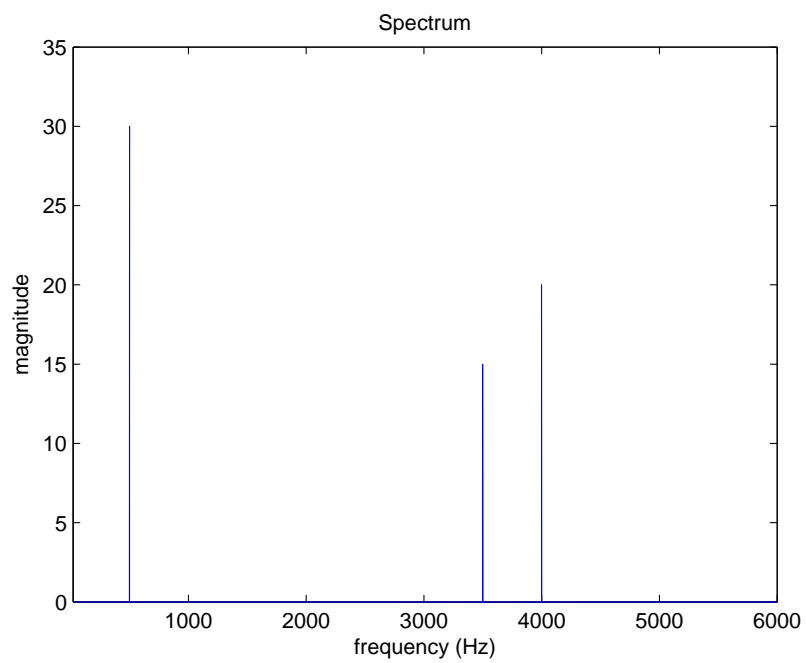


Figura 4.12: Contribuição do som da figura 4.11 tendo em conta o som objetivo da figura 4.10

4.3.5 RMS

O RMS (*Root Mean Square*), ou valor eficaz, representa a potência média de um sinal. Neste trabalho, o seu cálculo é feito no domínio das frequências, recorrendo ao teorema de Parseval, resultando em que:

$$RMS = \sqrt{\sum_m \left| \frac{X[m]}{N} \right|^2} \quad (4.9)$$

Sendo que $X[m]$ é a transformada discreta de Fourier do sinal a analisar e N o número de amostras que esse sinal contem.

4.3.6 Energia

O valor de energia de cada sinal $x[n]$ é utilizado na combinação de sons e é calculado utilizando:

$$E = \sum_n |x[n]|^2 \quad (4.10)$$

4.3.7 Frequência Fundamental

A estimação da frequência fundamental de cada som permite auxiliar o algoritmo de pesquisa a reduzir o espaço de pesquisa. Este processo de estimação foi feito utilizando o SWIPE [44]. O SWIPE (*Sawtooth Waveform Inspired Pitch Estimator*) é um algoritmo de estimação da frequência fundamental de sons que estima a frequência fundamental de um som como sendo a frequência fundamental de uma onda triangular, cujo espectro melhor se aproxima do espectro do sinal de entrada.

Após o valor da frequência fundamental ser calculado é possível diminuir o tamanho do espaço de pesquisa através da restrição de que os sons presentes neste espaço devem ter uma frequência fundamental mais alta que a frequência do primeiro pico do som objetivo. Reduz-se assim o número de combinações possíveis que o algoritmo de pesquisa deve analisar.

4.4 Funções de Combinação

A combinação de diferentes sons originam novas características que devem ser estimadas e comparadas com o som objetivo. Uma ilustração da tarefa a realizar pelas funções de combinação pode ser encontrada na figura 4.13: a partir de um conjunto de descritores de sons individuais conseguir encontrar o valor dos descritores do correspondente som orquestrado.

Assumindo que os registos tímbricos de cada som são combinados num único então o espectro resultante será uma soma das componentes espectrais de cada som que compõem essa combinação normalizados segundo a sua energia [17]. Partindo dessa suposição é possível deduzir as seguintes equações que correspondem aos estimadores dos valores das características do som orquestrado ce_o , ee_o e is_o , composto por l sons diferentes com valores de centróide espectral igual a ce_i , valores de espalhamento espectral igual a ee_i , valores de amplitude de picos espectrais iguais a p_{e_i} e energia

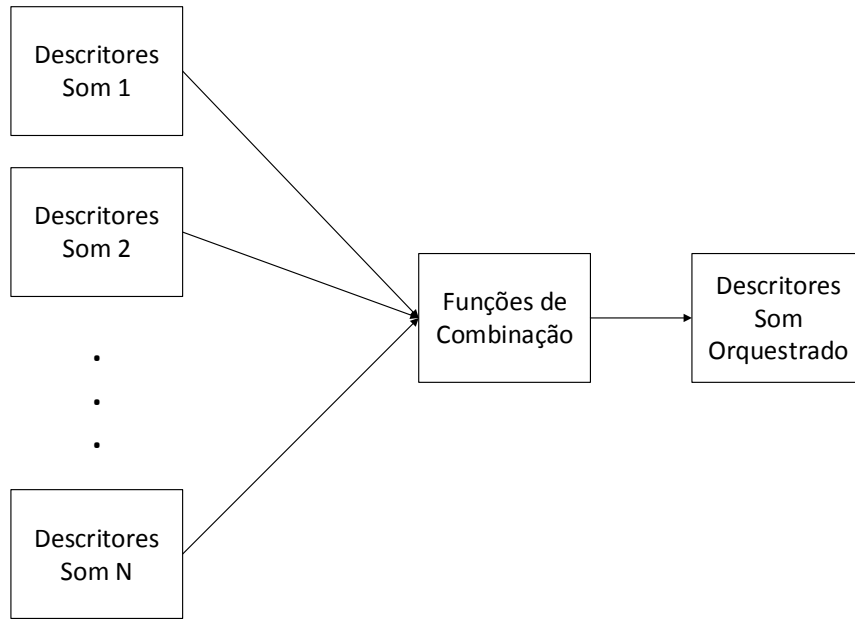


Figura 4.13: Funções de combinação

total igual a e_i :

$$ce_o = \frac{\sum_i^l e_i ce_i}{\sum_i^l e_i} \quad (4.11)$$

$$ee_o = \sqrt{\frac{\sum_i^l e_i (ce_i^2 + ee_i^2)}{\sum_i^l e_i} - ce_o^2} \quad (4.12)$$

$$is_o = 20 \log_{10} \left(\sum_i^l \frac{1}{l} \sum_j^l p_{e_i}(j) \right) \quad (4.13)$$

A estimação dos picos espectrais da combinação é feita através da comparação dos valores da contribuição de cada som nas frequências de interesse. Partindo dos vetores de contribuições dos diferentes sons é feito o cálculo do valor máximo de amplitude para cada frequência. O vetor de picos espectrais da combinação orquestrada é então:

$$P_o^T(n) = \left\{ \max_{i \in l} (P_{s_i}^T(1)), \max_{i \in l} (P_{s_i}^T(2)), \dots \right\} \quad (4.14)$$

Como o espectro é calculado em decibéis a utilização do máximo é uma boa aproximação do espectro combinado, no caso dos picos espectrais [17]. Um exemplo do resultado da combinação do espectro de dois sons pode ser observado nas figuras 4.15, 4.16 e 4.17.

4.4.1 Avaliação das Funções de Combinação

Para validar as funções de combinação descritas foram realizados testes em que um determinado número de sons foram combinados e as suas características medidas. Compararam-se as características medidas para cada combinação com as características calculadas utilizando as funções de combinação. Os resultados obtidos confirmam as funções de combinação como uma boa aproximação dos valores reais das características de uma combinação de sons.

Foram realizados testes para a combinação de 2, 3 e 4 sons diferentes, tendo sido cada um dos testes repetido 1000 vezes utilizando sons diferentes. Os valores de erros relativos são apresentados na tabela 4.1 e representam o quão bem as funções de combinação são uma aproximação do valor real dos descritores de um som combinado, ou seja, quanto mais baixo o valor de erro melhor é a aproximação. Assim, uma combinação de sons origina um novo conjunto de descritores e, a partir das funções de combinação, este novo conjunto de descritores pode ser calculado sabendo apenas à partida quais os sons que compõem essa combinação. Neste teste mediu-se a qualidade das funções de combinação através da comparação de valores reais dos descritores de uma combinação (medidos através da combinação manual dos sons e posterior cálculo dos descritores) e de valores estimados pelas funções de combinação (obtidos através da aplicação destas funções).

O valor de erro relativo para o centróide espectral, espalhamento espectral e intensidade sonora foi calculado da seguinte forma: admitindo por exemplo um valor de centróide espectral calculado utilizando as funções de combinação ce_c , e respetivo valor medido ce_m , então o erro relativo será:

$$e = 1 - \frac{\min(ce_c, ce_m)}{\max(ce_c, ce_m)} \quad (4.15)$$

Já no caso dos picos espectrais comparou-se o vetor de picos espectrais construído pelas funções de combinação e o vetor de picos espectrais medido. A comparação foi feita utilizando a distância cosinusoidal (*cosine similarity*) entre os dois vetores. Assim, e assumindo um vetor de picos espectrais construído pelas funções de combinação pe_c , e um vetor de picos espectrais medido pe_m , então o erro relativo será:

$$e = 1 - \cos(pe_c, pe_m) \quad (4.16)$$

	2 Sons	3 Sons	4 Sons
Centróide Espetral	7.9% (9.1%)	8.3% (8.6%)	7.8% (7.4%)
Espalhamento Espetral	10.2% (9.6%)	9.9% (8.8%)	10.0% (8.6%)
Intensidade Sonora	2.2% (1.8%)	2.4% (1.7%)	2.8% (1.8%)
Picos Espectrais	0.4% (1.9%)	0.2% (1.0%)	0.3% (1.0%)

Tabela 4.1: Erro relativo entre características calculadas utilizando as funções de combinação e características medidas da combinação. Em parênteses o desvio padrão do erro relativo.

Como pode ser observado, os valores de erro relativo obtidos são significativamente baixos, validando as funções de combinação como uma boa aproximação para o cálculo dos valores das características de uma combinação de sons.

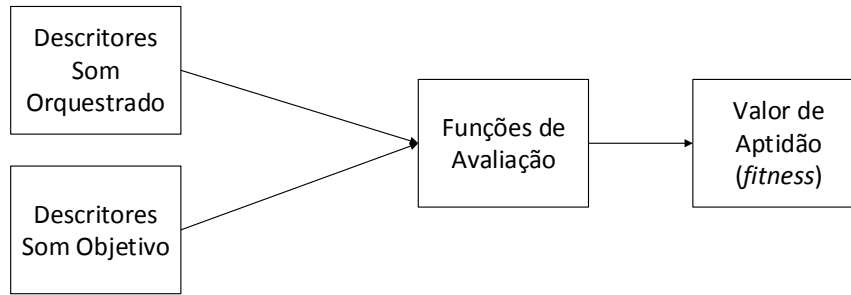


Figura 4.14: Funções de avaliação

4.5 Funções de Avaliação

Após serem definidas as características utilizadas para diferenciar cada som é necessário a definição de funções que atribuam uma distância numérica entre os diferentes sons. O objetivo das funções de avaliação é ilustrado na figura 4.14: a partir de um conjunto de valores de descritores de um som orquestrado e de um conjunto de valores de descritores de um som objetivo atribuir um valor real que quantifique a similaridade entre estes dois sons. No caso do centróide espectral, do espalhamento espectral e da intensidade sonora é utilizado o desvio absoluto, normalizado segundo o valor da característica do objetivo. Assim, se ce_o , ee_o e is_o forem as características da combinação orquestrada, ce_T , ee_T e is_T as características do som objetivo, então a distância entre estes dois sons é calculada usando:

$$D_T^{ce} = \frac{|ce_o - ce_T|}{ce_T} \quad (4.17)$$

$$D_T^{ee} = \frac{|ee_o - ee_T|}{ee_T} \quad (4.18)$$

$$D_T^{is} = \frac{|is_o - is_T|}{is_T} \quad (4.19)$$

Quanto aos picos espectrais é utilizada a distância cosinusoidal entre o vetor de picos espectrais do som objetivo e o vetor de picos espectrais do som orquestrado. A utilização desta distância permite atribuir um valor numérico à similaridade destes dois vetores, sendo que quanto menor for este valor, mais semelhantes serão os vetores. Esta distância é calculada com:

$$D_T^p = 1 - \cos(P_o^T, P_T^T) \quad (4.20)$$

Estes quatro valores quantificam um vetor multidimensional que avalia a distância de uma combinação ao som objetivo segundo os diferentes descritores escolhidos. Para a utilização dos algoritmos de pesquisa é necessário que estes valores sejam assimilados num único que representa a aptidão (*fitness*) dessa combinação. Esta combinação de valores é feita atribuindo pesos a cada descritor e somando os valores resultantes. Após refinamento sucessivo no valor dos pesos foram

escolhidos os seguintes valores:

$$\lambda = \{0.1, 0.1, 0.2, 0.6\} \quad (4.21)$$

Estes valores de pesos atribuem uma ponderação de 10% para o centróide espectral e espalhamento espectral, 20% para a intensidade sonora e 60% para a semelhança dos picos espectrais. O valor final de aptidão de uma combinação será então:

$$fitness = - \sum_i \lambda_i D_T^i \quad (4.22)$$

Como se pretende minimizar o valor das distâncias ao som objetivo e como os algoritmos de pesquisa utilizados maximizam uma determinada função objetivo torna-se necessário inverter o valor do sinal da aptidão de uma combinação, daí o valor de aptidão de uma combinação será sempre negativo. É este o valor utilizado pelo algoritmo de pesquisa para distinguir e avaliar os diferentes indivíduos presentes na população.

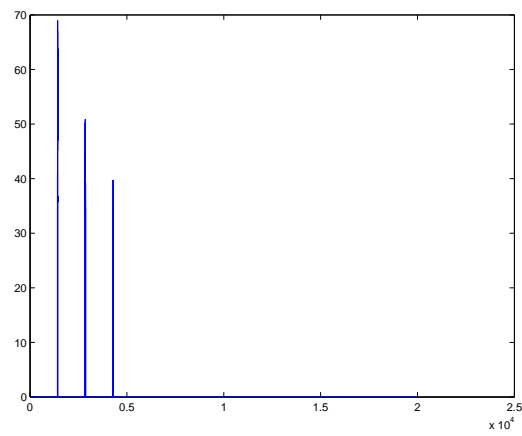


Figura 4.15: Espectro de um som, nota F6

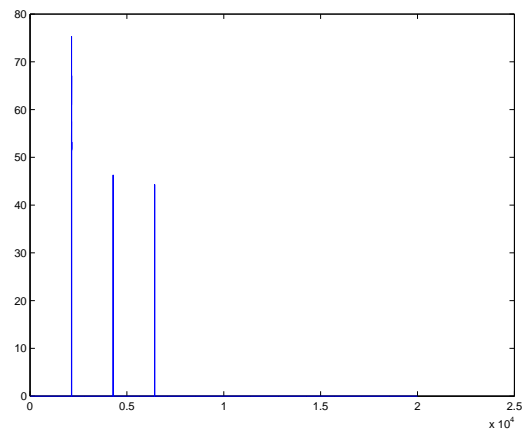


Figura 4.16: Espectro de um som, nota C7

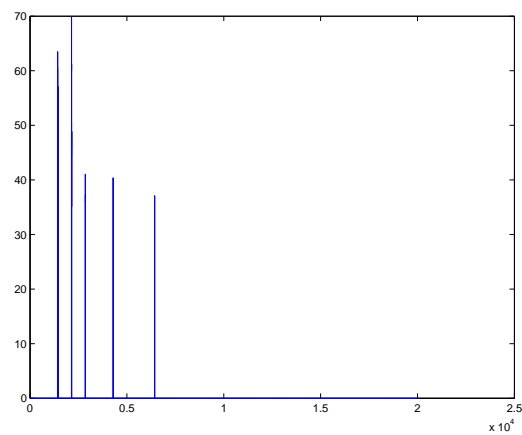


Figura 4.17: Combinação do espectro de sons, notas F6 e C7

Capítulo 5

Implementação

Neste capítulo apresenta-se a solução desenvolvida para a resolução do problema de orquestração musical. Inicialmente descreve-se a forma e conteúdo da base de dados sonora e as operações de pré-processamento que são realizadas antes do início de cada processo de pesquisa. De seguida descreve-se a arquitetura de um algoritmo genético e de um sistema imunitário artificial, desenvolvidos tendo em conta as restrições do problema de orquestração. Por fim, apresenta-se a arquitetura e funcionamento da aplicação desenvolvida.

5.1 Base de Dados Sonora

A base de dados sonora do sistema desenvolvido reside numa coleção de sons que representam as diferentes notas produzidas pelos diferentes instrumentos que compõem uma orquestra. Os sons presentes nesta coleção foram obtidos da *RWC Music Database* [41], tendo sido selecionados uma parte dos mesmos para a implementação e avaliação do sistema desenvolvido. Os instrumentos selecionados para fazerem parte da base de dados foram os seguintes:

- Violino
- Viola
- Violoncelo
- Contrabaixo
- Trombeta
- Trombone
- Tuba
- Trompa Francesa

- Oboé
- Trompa Inglesa
- Fagote
- Clarinete
- Flauta

Para além das diferentes notas produzidas por estes instrumentos, foram também utilizadas diferentes dinâmicas: *piano*, *mezzo forte* e *forte*.

Na sua forma original, os sons desta coleção contém num mesmo ficheiro todas as notas de um determinado instrumento musical. Foi por isso necessário realizar uma segmentação dos elementos desta coleção de forma a dividir as notas dos instrumentos em diferentes ficheiros. Um exemplo dos resultados desta segmentação são apresentados na figura 5.1. Cada ficheiro foi segmentado em notas individuais que foram posteriormente guardadas em diferentes ficheiros. Para além desta operação, foi necessário também estimar qual a frequência fundamental de cada som para diferenciar os ficheiros por notas musicais. Esta estimação foi feita utilizando o SWIPE [44], permitindo a caracterização de todos os elementos da base de dados segundo a sua nota musical. Um exemplo desta identificação pode ser visualizado na figura 5.2. Todos os ficheiros presentes na base de dados apresentam uma frequência de amostragem de 44100 Hz e estão guardados utilizando o formato de ficheiros .WAV.

A base de dados resultante contém 1439 ficheiros diferentes representando as notas produzidas pelos 13 instrumentos seleccionados e pelas 3 dinâmicas escolhidas. Para cada um destes ficheiros foram calculados e guardados os valores dos descritores apresentados na secção 4.3 para posterior uso pelo algoritmo de pesquisa.

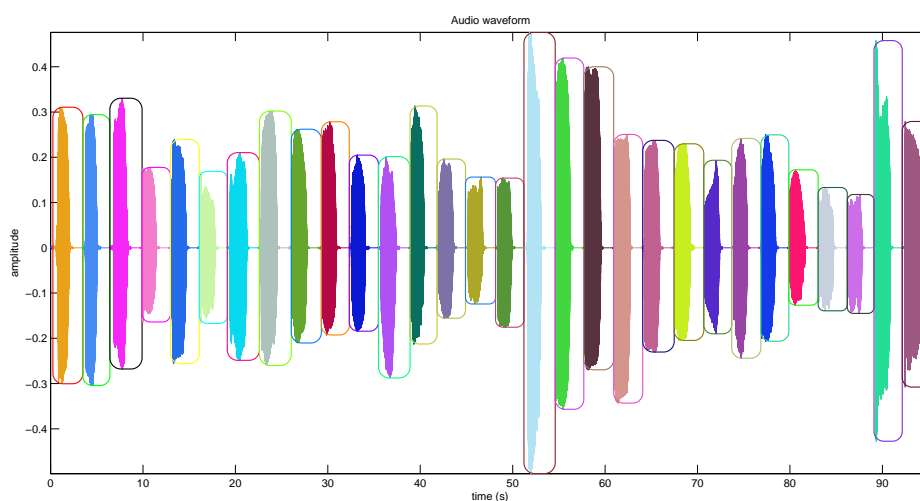


Figura 5.1: Resultados do processo de segmentação dos ficheiros da base de dados RWC

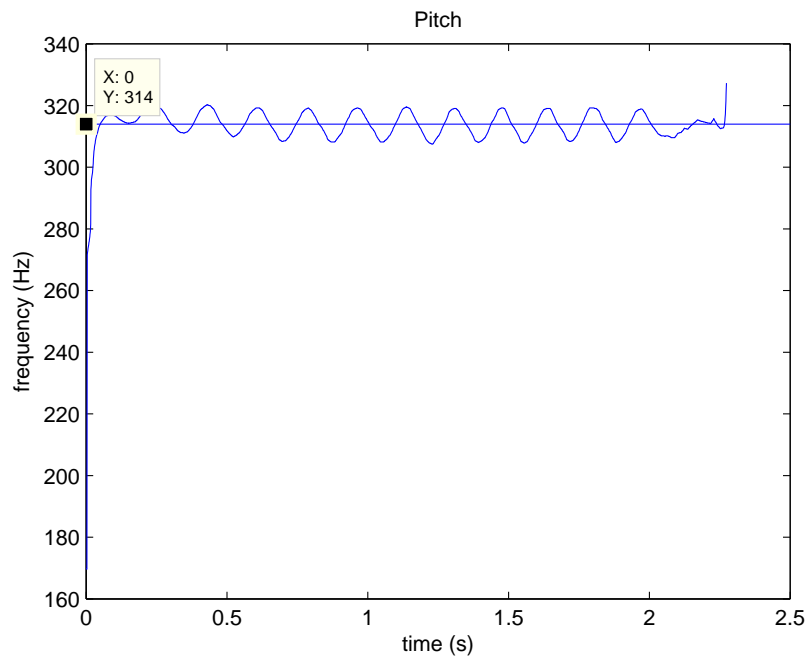


Figura 5.2: Resultados do processo de identificação da frequência fundamental utilizando o SWIPE

5.2 Representação

A representação dos indivíduos utilizados nos algoritmos de pesquisa é feita sob a forma de um vetor de índices com um tamanho limitado. Cada índice presente neste vetor é associado a uma determinada nota de um determinado instrumento e dinâmica, sendo um valor único para cada ficheiro da base de dados. Reserva-se o valor zero para o elemento nulo, ou seja o valor zero corresponde à não existência de um instrumento na posição em que este se encontra no vetor de índices.

Assim, cada solução é representada por um vetor de índices, correspondentes aos instrumentos e notas que compõem essa solução. A combinação desses índices origina a solução orquestrada, dependendo do utilizador a especificação do número de índices diferentes que podem fazer parte da solução. Este valor limite é especificado aquando a inicialização do algoritmo de pesquisa, limitando o tamanho do vetor de índices.

Na tabela 5.1 é fornecido um exemplo com os números de índice e os associados instrumentos, dinâmicas e notas musicais. Utilizando uma representação com 3 sons diferentes uma possível solução seria codificada da seguinte forma:

$$x_i = \{10, 1275, 1414\}$$

Nesta solução apresentada a combinação é composta por:

1. Uma nota C5, da dinâmica *forte*, do instrumento violino

2. Uma nota E5, da dinâmica *mezzo forte*, do instrumento clarinete
3. Uma nota G5, da dinâmica *piano*, do instrumento flauta

Índice	Ficheiro
...	...
10	/Violino/Forte/C5.wav
...	...
350	/Violoncelo/Mezzo/F5.wav
...	...
1275	/Clarinete/Mezzo/E5.wav
...	...
1414	/Flauta/Piano/G5.wav
...	...

Tabela 5.1: Exemplo de índices e nomes de ficheiros da base de dados

5.3 Pré-Processamento

Utilizar todos os sons disponíveis como soluções válidas para o problema de orquestração iria aumentar em grande escala a dimensão do espaço de pesquisa existente. Para reduzir o tamanho deste espaço de pesquisa é feito um pré-processamento sobre o som objetivo, extraindo-se as componentes espectrais mais importantes. Após ser feita a identificação do valor de frequência destas componentes a base de dados é filtrada, sendo selecionados apenas sons que apresentem alguma componente espectral em comum com o som objetivo, tal como já foi referido na secção 4.3.4.

Para além desta seleção por componente espectrais em comum, é feita uma posterior seleção em que apenas são selecionados sons com frequência fundamental mais alta que a do valor de frequência do primeiro pico do som objetivo. Esta seleção foi também já referida na secção 4.3.7.

São estes sons selecionados que constituem a lista de índices válidos que o algoritmo de pesquisa pode utilizar. Restringe-se assim por um lado o espaço de combinações possíveis e por outro o número de instrumentos que podem constituir a solução.

5.4 Algoritmo Genético

Foram desenvolvidas duas implementações diferentes deste algoritmo neste trabalho. Ambas as implementações foram especificadas tendo em conta as limitações do problema de orquestração e ambas utilizam as funções de avaliação e de similaridade introduzidas no capítulo 4. Estes algoritmos foram utilizados para a obtenção de soluções orquestradas através da combinação de sons presentes na base de dados construída.

A primeira implementação desenvolvida modela o comportamento clássico do algoritmo genético: uma população de indivíduos é inicializada aleatoriamente e evoluída segundo processos

de seleção, cruzamento e mutação. Nesta implementação não existe nenhuma restrição na constituição da população, podendo existir mais que um indivíduo com as mesmas características. Efetivamente o comportamento desta primeira implementação desenvolvida é a de convergir para uma única solução ótima. O algoritmo desta implementação é descrito no algoritmo 5.1.

Algoritmo 5.1: Algoritmo genético desenvolvido: Versão 1

Entrada: $n_s, iter_n, p_c, p_m$

Saída: P

inicializa n_s indivíduos aleatoriamente;

enquanto número de iterações $< iter_n$ **faça**

 calcula *fitness* de cada indivíduo;

 ordena população por *fitness* decrescente;

$P_e \leftarrow$ guarda os n_e melhores indivíduos (elitismo);

 seleciona n_s indivíduos por *roulette wheel*;

 efetua cruzamento uniforme nos n_s indivíduos selecionados com probabilidade p_c ;

$P_m \leftarrow$ efetua mutação nos indivíduos cruzados com probabilidade p_m ;

 atualiza nova população $P = [P_e; P_m]$;

fim

Já na segunda implementação desenvolvida existe um controlo sobre a população na medida em que cada indivíduo da população é único. A cada iteração é inicialmente feita uma filtragem por indivíduos repetidos, sendo eliminados os indivíduos duplicados. Desta maneira o algoritmo não converge para uma única solução ótima e, através do uso de elitismo, é possível a convergência para vários valores ótimos dependendo do número de indivíduos guardados pelo elitismo ao longo das iterações. Este comportamento simula parcamente uma das principais vantagens do sistema imunitário artificial: a possibilidade de convergência para vários valores ótimos. O algoritmo desta implementação é descrito no algoritmo 5.2.

Algoritmo 5.2: Algoritmo genético desenvolvido: Versão 2

Entrada: $n_s, iter_n, p_c, p_m$

Saída: P

inicializa n_s indivíduos aleatoriamente;

enquanto número de iterações $< iter_n$ **faça**

 elimina indivíduos duplicados;

 calcula *fitness* de cada indivíduo;

 ordena população por *fitness* decrescente;

$P_e \leftarrow$ guarda os n_e melhores indivíduos (elitismo);

 seleciona n_s indivíduos por *roulette wheel*;

 efetua cruzamento uniforme nos n_s indivíduos selecionados com probabilidade p_c ;

$P_m \leftarrow$ efetua mutação nos indivíduos cruzados com probabilidade p_m ;

 atualiza nova população $P = [P_e; P_m]$;

fim

Apesar destas diferenças, o funcionamento intrínseco dos dois algoritmos é o mesmo: A cada iteração é inicialmente calculado o valor de avaliação dos indivíduos da população utilizando as

funções de avaliação definidas no capítulo 4. Após este cálculo é feita uma seleção de indivíduos para cruzamento utilizando o esquema de *roulette wheel* descrito em [45]. Estes indivíduos são sujeitos a operações de cruzamento que, com uma determinada probabilidade, interpolam o valor dos índices de dois indivíduos, originando dois novos indivíduos. Após a operação de cruzamento ser realizada a população é alterada através de operações de mutação que, também com uma determinada probabilidade, alteram ou eliminam valores de índice do vetor de cada indivíduo. Os algoritmos permitem também a utilização de elitismo ao longo das gerações, guardando um certo número de indivíduos mais aptos conforme o resto da população é alterada.

5.4.1 Operadores Genéticos

Os operadores genéticos de seleção, cruzamento e mutação foram implementados nos algoritmos desenvolvidos. Através de comparação de resultados o método de funcionamento dos mesmos foi sucessivamente refinado, tendo em conta as restrições que o problema de orquestração coloca. Descrevem-se de seguida o funcionamento de cada um dos operadores genéticos implementados.

5.4.1.1 Seleção

A seleção dos indivíduos destinados a gerarem descendência é feita através do método de *roulette wheel* proposto em [45]. Neste método é escolhido um indivíduo aleatoriamente da população, que é selecionado para cruzamento com uma probabilidade de f_i/f_{max} , sendo f_i o valor de aptidão do indivíduo escolhido, e f_{max} o valor máximo de aptidão da população atual. Caso o indivíduo não seja selecionado o processo é repetido novamente escolhendo-se um novo indivíduo aleatório. Através de sucessivas seleções cria-se assim uma nova população com os indivíduos selecionados que irá apresentar uma constituição dominada por indivíduos com valores de aptidão mais altos. Ou seja, indivíduos mais aptos apresentam uma maior probabilidade de serem selecionados para cruzamento podendo um mesmo indivíduo ser mais que uma vez selecionado.

Na figura 5.3 ilustra-se o funcionamento do esquema clássico da *roulette wheel*. É atribuída uma probabilidade de seleção a cada indivíduo dependendo do seu valor de aptidão. É de seguida calculado um número aleatório entre 0 e 1, correspondente à posição da roleta a ser selecionada. Como indivíduos mais aptos ocupam um espaço maior na roleta então estes indivíduos irão apresentar uma tendência maior a serem selecionados.

5.4.1.2 Cruzamento

O operador de cruzamento implementado cruza dois indivíduos selecionados aleatoriamente com probabilidade p_c . O cruzamento é uniforme, sendo que os novos indivíduos contém aleatoriamente índices de um dos indivíduos pais. Ou seja, caso x_i e x_j sejam dois indivíduos selecionados para cruzamento com:

$$x_i = \{i_1, i_2, i_3, i_4, i_5\}$$

$$x_j = \{i_6, i_7, i_8, i_9, i_{10}\}$$

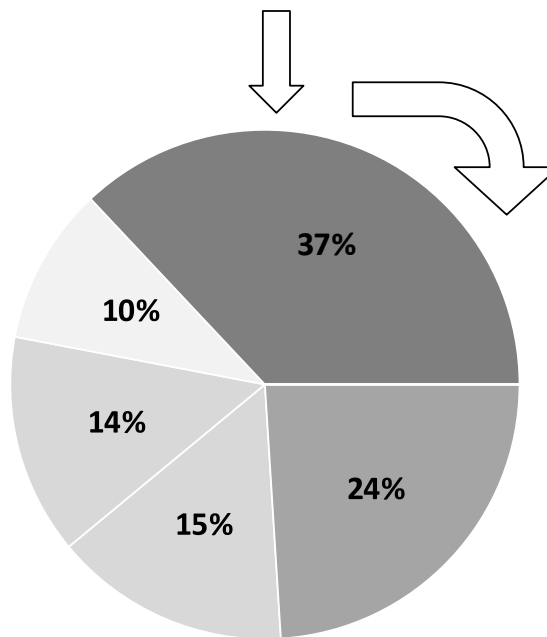


Figura 5.3: Ilustração do funcionamento clássico do esquema de *roulette wheel*

Então, dois possíveis descendentes serão:

$$x'_i = \{i_6, i_2, i_3, i_9, i_5\}$$

$$x'_j = \{i_1, i_7, i_8, i_4, i_{10}\}$$

5.4.1.3 Mutação

A mutação é feita com probabilidade p_m em posições aleatórias do vetor de índices, sendo o novo valor escolhido aleatoriamente como zero ou um índice válido. Ou seja caso uma determinada posição no vetor de índices tenha sido selecionada para mutação então o correspondente valor será, substituído por um valor de índice diferente ou substituído por o valor zero sendo a escolha do novo valor feita aleatoriamente. Caso x_i seja um indivíduo tal que:

$$x_i = \{i_1, i_2, i_3, i_4, i_5\}$$

Uma possível mutação seria:

$$x'_i = \{i_1, \underline{i_6}, i_3, i_4, i_5\}$$

5.5 Sistema Imunitário Artificial

Uma alternativa ao uso de algoritmos genéticos recai sobre a escolha de um sistema imunitário artificial. Estes últimos algoritmos baseiam-se no funcionamento do sistema imunitário dos vertebrados tendo sido inicialmente descritos na secção 3.2. As diferenças entre estes algoritmos e os algoritmos genéticos residem essencialmente nos operadores genéticos. Enquanto que no algoritmo genético os indivíduos sofrem mutação a uma taxa constante, no sistema imunitário esta taxa varia, sendo proporcional à aptidão de cada indivíduo. Assim, indivíduos mais aptos sofrem uma menor taxa de mutação enquanto que indivíduos pouco aptos sofrem uma maior taxa. Para além desta diferença, realça-se o facto de que nos sistemas imunitários artificiais não existe cruzamento entre indivíduos, realizando-se assim a otimização de indivíduos através de apenas mutação.

Em termos de resultados fornecidos por este algoritmo é importante destacar que, contrariamente aos algoritmos genéticos, o sistema imunitário artificial não converge para uma única solução. Pelo contrário, uma das principais vantagens do uso deste sistema recai sobre o facto de que várias soluções ótimas são exploradas simultaneamente, fornecendo ao utilizador uma diversidade maior no conjunto de soluções e colmatando assim uma das principais desvantagens dos algoritmos genéticos.

5.5.1 aiNet e opt-aiNet

A aiNet é um algoritmo que modela um sistema imunitário artificial e foi desenvolvida inicialmente para propósitos de compressão e agregação de dados [13]. Os autores deste algoritmo desenvolveram também uma versão, denominada de opt-aiNet, que apresenta como principal aplicação a resolução de problemas de otimização multi dimensionais [35]. Neste sistema, uma população de anticorpos (considerados como candidatos à solução) é evoluída através de processos de expansão clonal, mutação e posterior seleção. A opt-aiNet mantém ao longo da execução um conjunto de anticorpos que apresentam uma melhor aptidão em relação aos restantes anticorpos da população. As características mais marcantes da opt-aiNet são:

- Determinação automática do tamanho da população;
- Combinação de pesquisa local com pesquisa global;
- Definição de critérios de convergência;
- Capacidade de localizar e manter vários máximos locais.

Estas características distinguem o sistema imunitário artificial opt-aiNet do algoritmo genético que, no seu comportamento clássico, não é capaz de localizar vários máximos. Efetivamente esta é a característica mais importante nestes sistemas. Através da utilização de um algoritmo que forneça vários máximos locais, é possível a implementação de um sistema de orquestração capaz de fornecer várias soluções úteis ao utilizador.

Para compreender melhor o funcionamento do algoritmo opt-aiNet segue-se uma breve explicação da terminologia utilizada pelos autores deste algoritmo:

Célula Indivíduo da população. Cada célula corresponde a uma solução possível e é representada sob a forma de um vetor com valores reais mapeados num espaço euclidiano;

Fitness Aptidão de uma célula em relação a uma função objetivo que se pretende otimizar. O valor de *fitness* de cada célula corresponde ao valor da função objetivo recorrendo aos valores presentes na célula;

Afinidade Distância euclidiana entre duas células;

Clone Cópia idêntica de uma determinada célula. Os clones sofrem mutação para introduzir variação na população.

Algoritmo 5.3: Algoritmo opt-AiNet

Entrada: N_c, σ_s

Saída: C

inicializa um conjunto C de células aleatoriamente;

enquanto critério de convergência não é atingido **faça**

 calcula *fitness* de cada célula;

 gera N_c clones para cada célula;

 cada clone sofre mutação proporcional ao valor de *fitness* da célula original;

 calcula *fitness* de cada célula incluindo os novos clones;

 para cada célula selecione o clone mais apto e descartam-se os restantes;

 calcula valor médio de *fitness* da população;

se valor médio de *fitness* maior que o da iteração anterior **então**

 | continua;

senão

 | suprime células que apresentem uma afinidade mais baixa que σ_s ;

 | introduz uma percentagem d de células criadas aleatoriamente;

fim

fim

O algoritmo inicia a sua execução criando um conjunto de células com valores aleatórios. De seguida, é iniciado um ciclo em que são inicialmente calculados os valores de *fitness* de cada célula. Após este cálculo é iniciado um processo de clonagem das células da população. Estes clones sofrem uma mutação proporcional ao seu valor de *fitness*, sendo que o processo de mutação realiza-se segundo a seguinte expressão:

$$c' = c + \alpha N(0, 1) \quad (5.1)$$

$$\alpha = \left(\frac{1}{\beta} \right) \exp(-f^*) \quad (5.2)$$

Onde c' é uma célula c que sofreu mutação, $N(0, 1)$ é uma variável aleatória Gaussiana com média igual a zero e desvio padrão igual a um, β é um parâmetro que controla o decaimento da função

exponencial inversa (geralmente $\beta = 100$), e f^* é o valor de *fitness* do indivíduo normalizado segundo o intervalo $[0, 1]$.

Após serem feitas estas operações de clonagem e mutação sobre os clones, são calculados os novos valores de *fitness* de todas as células e seus respectivos clones que sofreram mutação. Para cada célula é selecionado o clone mais apto ou selecionada a célula original, caso nenhum clone seja mais apto que a célula original.

Posteriormente à seleção das células mais aptas é formada uma nova população contendo as células selecionadas. O valor médio de *fitness* desta nova população é calculado e, caso este valor seja diferente do valor obtido na iteração anterior então retoma-se a execução a partir do início do ciclo, caso contrário realiza-se uma operação de supressão na população.

A operação de supressão é realizada quando o valor médio de *fitness* da população não evolui e corresponde a uma eliminação de células que apresentem um valor de afinidade baixo, nomeadamente, são eliminadas células com afinidade mais baixa que a especificada pelo limiar de supressão σ_s . Ou seja, são eliminadas células que se encontram próximas no espaço de pesquisa, representando soluções semelhantes. As células restantes após esta operação ser efetuada denominam-se de células de memória. Após ser feita a supressão são introduzidas na população uma percentagem d de novas células criadas aleatoriamente. Estas operações permitem por um lado controlar o tamanho da população, e por outro manter os máximos locais com apenas uma célula evitando desperdício de recursos. Para além disso, através da introdução de novas células evita-se a estagnação do algoritmo em máximos locais e contribui significativamente para a exploração do espaço de pesquisa.

5.5.2 Implementação do Algoritmo opt-aiNet

Para a aplicação do algoritmo opt-aiNet no sistema de orquestração desenvolvido foi necessária a sua transposição para um algoritmo capaz de lidar com valores discretos na codificação das células. Para tal operação foi necessário alterar o funcionamento do processo de mutação uma vez que na implementação original este processo estava concebido para células com valores reais. O novo processo de mutação implementado utiliza a expressão da exponencial inversa para determinar a probabilidade de mutação que uma célula irá apresentar. Após este cálculo ser feito o vetor de índices que codifica a solução é percorrido e, para cada valor de índice, é calculado um número aleatório entre 0 e 1. Caso o número aleatório seja mais baixo que a probabilidade de mutação previamente calculada então esse valor de índice é alterado para um novo valor, sendo a escolha do novo valor feita aleatoriamente.

A função exponencial inversa utilizada é a seguinte:

$$p_m = \exp(-\gamma f^*) \quad (5.3)$$

Com γ geralmente igual a 1.2. Este valor de γ gera o gráfico apresentado na figura 5.4, correspondente à probabilidade de mutação segundo o valor de *fitness* normalizado.

Quanto à operação de supressão o funcionamento intrínseco é o mesmo que o do opt-aiNet, nomeadamente são calculados os valores de afinidade entre as diferentes células sendo suprimidas aquelas que apresentem um valor de afinidade inferior a σ_s . Os valores de afinidade no caso do problema de orquestração correspondem a um vetor com as distâncias relativas entre os valores dos diferentes descritores. Estas distâncias relativas são calculadas entre os valores obtidos através da aplicação das equações 4.11, 4.12 e 4.13.

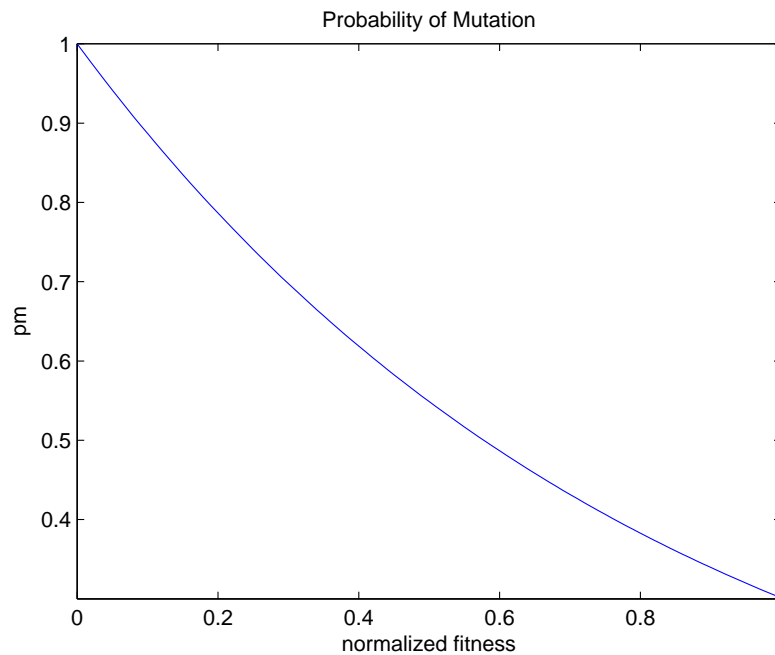


Figura 5.4: Probabilidade de mutação segundo o valor de *fitness* normalizado, com γ igual a 1.2

5.6 Aplicação

Apresenta-se nesta secção a aplicação desenvolvida para a resolução do problema de orquestração musical. Tendo em conta as várias restrições, e utilizando os algoritmos de pesquisa previamente descritos desenvolveu-se uma aplicação em MATLAB (R2014a, The MathWorks, Natick, Massachusetts, United States) com diferentes módulos, incluindo uma pequena interface gráfica para visualizar os resultados obtidos. Foi também utilizada uma *framework* de processamento de áudio digital, a *MIRToolbox* [46]. Esta ferramenta oferece várias funções de processamento de áudio, incluindo o cálculo do espectro e pesquisa automática por máximos locais nesse espectro. Os módulos desenvolvidos foram especificados tendo em conta a interoperabilidade entre os diferentes algoritmos de pesquisa propostos: o algoritmo genético e o sistema imunitário artificial.

A aplicação desenvolvida lida com o problema da orquestração musical como um problema de otimização combinatoria em que as diferentes combinações representam as diferentes orquestrações possíveis. As combinações são representadas por um vetor de valores inteiros representando

os diferentes instrumentos e notas musicais que contém essa combinação. Várias operações de pré-processamento são efetuadas sobre os sons presentes na base de dados de forma a otimizar o espaço de pesquisa existente e reduzir conseqüentemente o número de iterações necessárias para a obtenção de soluções ótimas. Os resultados obtidos pelo algoritmo de pesquisa são apresentados numa interface gráfica e guardados em ficheiros .WAV para o uso posterior pelo utilizador. Descreve-se de seguida a estrutura da aplicação desenvolvida.

5.6.1 Estrutura

Os diferentes módulos constituintes da aplicação desenvolvida são apresentados na figura 5.5.

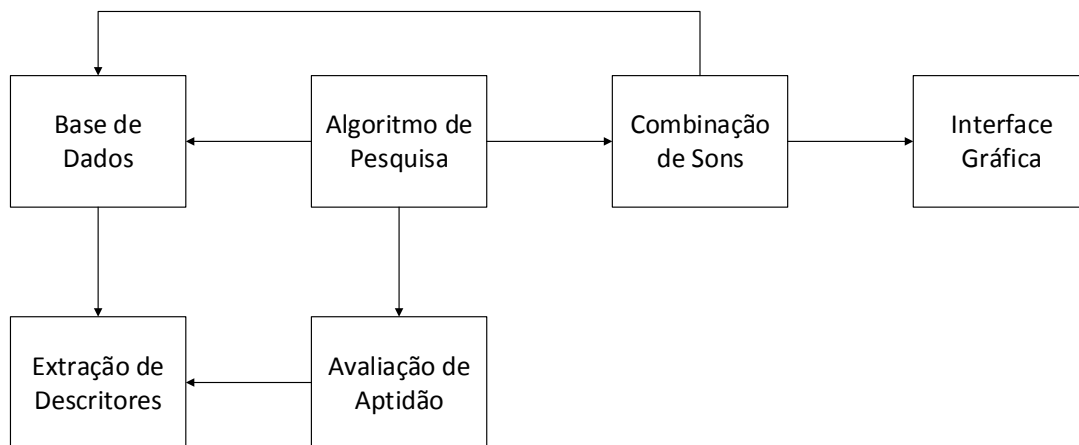


Figura 5.5: Módulos da aplicação desenvolvida e suas dependências

Inicialmente o sistema inicia a execução através da especificação, por parte do utilizador, da base de dados a utilizar, do som objetivo que se pretende orquestrar e do número máximo de sons que podem ser utilizados nessa orquestração. Após esta especificação inicial o sistema extrai as características do som objetivo descritas na secção 4.3. A partir das características do som objetivo são seleccionados, de entre os ficheiros presentes na base de dados, aqueles cuja relevância é maior para a resolução do problema de orquestração. Esta seleção é feita, primeiramente, sobre a frequência fundamental de cada som, descartando sons que apresentem frequência fundamental mais baixa que o valor de frequência do primeiro pico espectral do som objetivo, e em seguida, sobre o espectro de cada som, descartando sons que não apresentem picos espectrais em comum com o som objetivo. Estas operações de seleção e pré-processamento foram já descritas na secção 5.3.

Após serem seleccionados os sons que irão fazer parte do espaço de pesquisa o fluxo de execução é passado para o algoritmo de pesquisa que, através de sucessivas iterações, calcula o ou os indivíduos mais aptos correspondentes a soluções de maior relevância. A relevância de cada

indivíduo é avaliada segundo as funções de avaliação apresentadas na secção 4.5, sendo utilizadas pelo algoritmo de pesquisa para quantificar numericamente a qualidade de um determinado indivíduo.

Assim que o algoritmo de pesquisa termina a sua execução os indivíduos resultantes são combinados criando-se assim os sons orquestrados. Esta combinação é feita através da soma linear dos sons constituintes da solução sendo utilizado um *Phase Vocoder* [47] para a igualização da duração de cada som. A necessidade da utilização de um *Phase Vocoder* fundamenta-se pelo facto de que os sons da base de dados apresentam durações temporais diferentes sendo que, através da utilização deste tipo de algoritmos é possível alterar a duração de um som sem modificar a sua altura nem dinâmica, tornando-se assim especialmente útil para igualizar a duração temporal de cada som da base de dados. O resultado final é um conjunto de ficheiros de som representando a orquestração das diferentes soluções.

Terminando o fluxo de execução encontra-se uma pequena interface gráfica desenvolvida para a rápida visualização dos resultados obtidos. Esta interface apresenta os resultados obtidos, permite a reprodução dos sons resultantes e apresenta também o conjunto de instrumentos utilizados para a orquestração do resultado.

5.6.2 Funcionamento

Os principais módulos da aplicação desenvolvida e suas dependências encontram-se ilustrados na figura 5.5. O funcionamento dos diferentes módulos da aplicação desenvolvida é de seguida apresentado, assim como as funções mais importantes realizadas por cada um.

5.6.2.1 Base de Dados

Este módulo é responsável pelo processamento e seleção dos sons que irão fazer parte do espaço de pesquisa. Para tal apresenta funções que percorrem uma determinada lista de ficheiros, especificada pelo utilizador, seleccionando aqueles cuja relevância para o problema de orquestração é maior. Este módulo é também responsável por guardar as características extraídas de cada som pertencente à base de dados, assim, este cálculo de extração é necessário ser feito apenas uma vez. A forma como este módulo guarda estas características é feita através da utilização de ficheiros anexos a cada som contendo o valor dessas características. Dessa forma quando se pretende obter as características de um determinado som apenas é necessário ler o ficheiro anexo.

Por exemplo, quando um novo ficheiro de som (.WAV) é introduzido na base de dados o sistema calcula o valor das características desse ficheiro e guarda o correspondente resultado num ficheiro anexo (.MAT). Este ficheiro anexo será depois utilizado para ler rapidamente as características do ficheiro de som correspondente. Na figura 5.6 ilustra-se a organização dos ficheiros na base de dados.

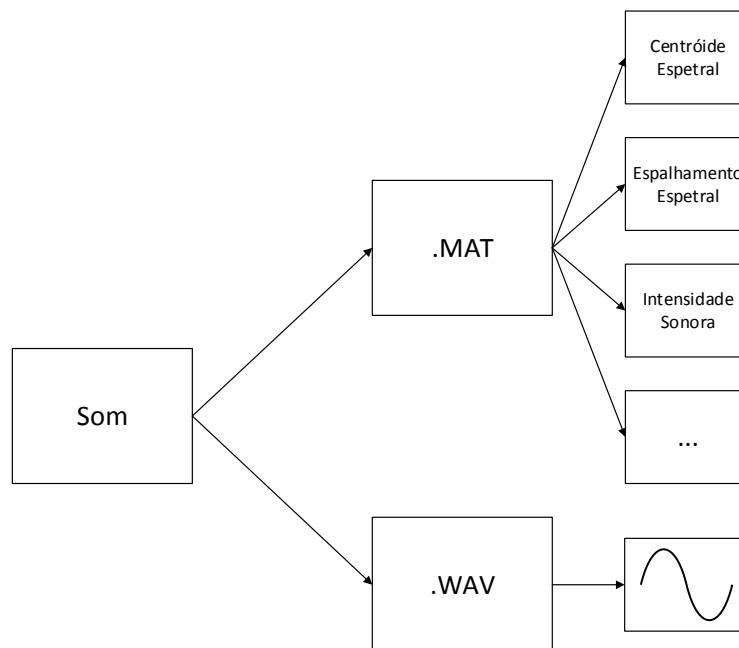


Figura 5.6: Organização dos ficheiros na base de dados. Para cada ficheiro de som (.WAV) existe um ficheiro de dados anexo (.MAT) com as características desse som

5.6.2.2 Extração de Descritores

A extração do valor dos descritores é feita recorrendo às expressões introduzidas na secção 4.3. Para o cálculo do espectro e correspondente picos espectrais são utilizadas duas funções pertencentes à *MIRToolbox*: a *mirspectrum* e *mirpeaks*. Estas funções calculam automaticamente o espectro de um som e identificam os picos espectrais mais proeminentes num determinado espectro, tendo sido escolhidas pela sua facilidade de uso e bons resultados.

Previamente ao cálculo dos descritores o sinal a analisar sofre uma filtragem nas frequências através de um filtro ERB (*Equivalent Rectangular Bandwidth*). Estes filtros providenciam uma aproximação aos filtros presentes no sistema auditivo humano. Para a implementação deste filtro neste módulo foram utilizadas as funções *MakeERBFilters* e *ERBFilterBank* pertencentes à *Auditory Toolbox* [48]. A resposta em frequência destes filtros pode ser observada na figura 5.7.

5.6.2.3 Algoritmo de Pesquisa

Como já foi descrito foram desenvolvidos dois algoritmos diferentes para a pesquisa de soluções úteis: o algoritmo genético e o sistema imunitário artificial. Estes algoritmos foram desenvolvidos de forma a apresentar interoperabilidade entre os mesmos, podendo um ser substituído pelo outro sem haver necessidade de uma alteração profunda do código. Assim ambos os algoritmos apresentam em comum os mesmos valores de entrada e os mesmos valores de saída: à entrada são fornecidos ao algoritmo uma lista de índices válidos para os indivíduos da população

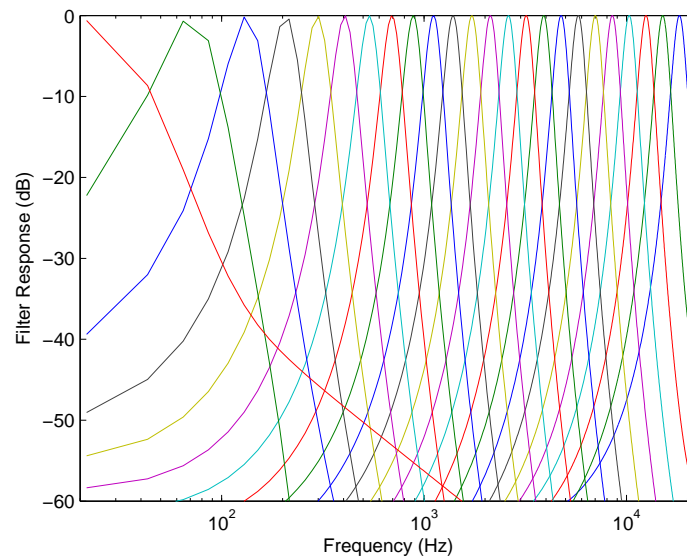


Figura 5.7: Resposta em frequência dos filtros ERB

assim como o número máximo de sons diferentes que uma solução pode conter; Já à saída são retornados uma lista de indivíduos considerados pelo algoritmo de pesquisa como os mais aptos e os seus correspondentes valores de aptidão. Toda a lógica interna do algoritmo de pesquisa é tratada independentemente dos restantes módulos do sistema, exceto no que toca à avaliação dos indivíduos. Nesse caso são utilizadas funções que recorrem a outros módulos do sistema.

5.6.2.4 Avaliação de Aptidão

A avaliação dos indivíduos na população é feita utilizando as funções de combinação e de avaliação apresentadas respetivamente nas secções 4.4 e 4.5. Inicialmente as funções de combinação estimam as características de uma determinada combinação a partir do vetor de índices representado pelo indivíduo. Após esta estimativa são calculadas as distâncias entre os valores das diferentes características e as características do som objetivo utilizando para tal as funções de avaliação. Finalmente, e a partir da equação 4.22 calcula-se o valor de aptidão de cada indivíduo. É este o valor utilizado pelos algoritmos de pesquisa para a avaliação das diferentes soluções.

5.6.2.5 Combinação de Sons

O módulo de combinação de sons efetua a combinação de dois ou mais sons, resultando num único som final. Esta combinação é feita segundo uma soma linear dos valores de cada som, sendo que a soma é ponderada segundo o número de elementos que irão compor o som final. Para compensar o facto de que alguns sons apresentam uma duração mais curta ou mais longa que outros foi utilizado um *Phase Vocoder* desenvolvido por Dan Ellis [47].

Este *Phase Vocoder* permite compensar a duração de certos sons, resultando em ficheiros que apresentam a mesma duração. O funcionamento deste algoritmo recorre à transformada de Fourier sendo que as posições no tempo das tramas da transformada são alteradas originando uma modificação da duração temporal do sinal original. Para além da alteração da posição temporal das tramas da transformada é necessária também uma compensação ao nível da fase para manter a coerência da mesma. No sistema desenvolvido o *Phase Vocoder* referido sincroniza as durações de cada combinação de sons utilizando como referência o valor médio da duração temporal desse conjunto de sons.

5.6.2.6 Interface Gráfica

A pequena interface gráfica desenvolvida serve essencialmente para uma avaliação rápida e intuitiva dos resultados produzidos. Para tal, são apresentados os espetros do som objetivo e dos diferentes sons orquestrados pelo algoritmo de pesquisa. São também apresentados os valores dos descritores dos sons orquestrados assim como a lista de instrumentos e notas que compõem a combinação. É possível reproduzir tanto o som objetivo como os sons orquestrados.

5.6.3 Interface

A interface com o utilizador é feita através de uma função que o utilizador deve executar no ambiente MATLAB. Nesta função deve ser especificada o caminho para a base de dados a utilizar, o caminho para o ficheiro contendo o som objetivo e o número máximo de instrumentos que podem fazer parte das combinações orquestradas. Um exemplo de execução pode ser observado na figura 5.8.

Após esta definição inicial de parâmetros o sistema inicia a sua execução, pesquisando autonomamente por combinações úteis ao utilizador. Efetivamente esta definição inicial de parâmetros é a única interação necessária para o sistema de orquestração funcionar. Quando o processo de pesquisa termina é iniciada uma interface gráfica que permite avaliar os resultados obtidos, esta interface pode ser observada na figura 5.9.

A interface gráfica apresentada na figura 5.9 permite a rápida avaliação dos resultados obtidos pelo processo de orquestração. No lado esquerdo desta interface são apresentados os principais picos espectrais do som objetivo, enquanto que no centro são apresentados os picos espectrais do som objetivo sobrepostos pelos picos espectrais do som orquestrado. Os dois conjuntos de picos distinguem-se pela cor: a vermelho os picos do som objetivo, a azul os picos do som orquestrado. Na posição inferior a este gráfico de picos existem dois botões que reproduzem respetivamente o som objetivo e o som orquestrado. No centro existe também uma caixa de seleção que permite selecionar entre os resultados obtidos como pode ser observado na figura 5.10. No lado direito da interface é apresentada uma tabela com o valor dos descritores tanto do som objetivo como do som orquestrado, assim como uma lista com os ficheiros que fazem parte da combinação orquestrada.

Esta interface apesar da elevada simplicidade permite a um utilizador verificar as soluções obtidas, através da visualização do espetro e do valor dos descritores.

```
>> orch('G:\RWC\Selected35', 'car-horn-trgt-norm.wav', 5)
6: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_21_G#4.wav
8: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_23_A#4.wav
9: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_24_B4.wav
10: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_25_C5.wav
11: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_26_C#5.wav
12: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_27_D5.wav
14: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_34_D#5.wav
15: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_35_E5.wav
16: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_36_F5.wav
18: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_38_G5.wav
19: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_39_G#5.wav
21: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_40_A5.wav
22: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_47_A#5.wav
23: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_48_B5.wav
24: G:\RWC\Selected35\151_Violin\Forte\151VNNOF_49_C6.wav
```

Figura 5.8: Inicialização do processo de orquestração

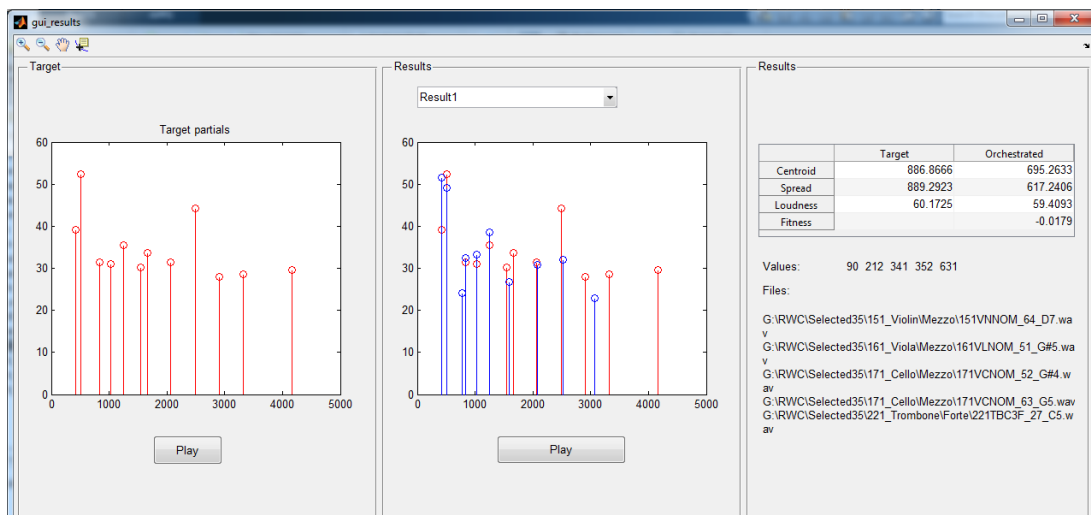


Figura 5.9: Interface gráfica desenvolvida onde são apresentados os resultados obtidos

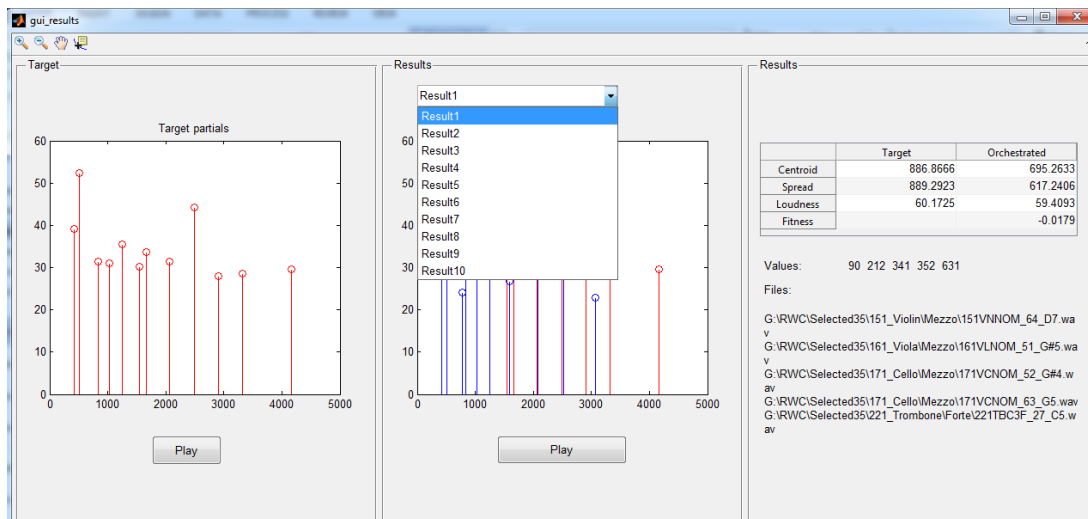


Figura 5.10: Seleção dos diferentes resultados

Capítulo 6

Avaliação e Resultados

Na fase final deste trabalho procedeu-se à comparação, avaliação e validação dos resultados obtidos recorrendo aos diferentes algoritmos de pesquisa desenvolvidos e utilizando o modelo teórico de orquestração proposto. Na fase de comparação pretendeu-se determinar as principais vantagens e desvantagens na utilização dos algoritmos de pesquisa desenvolvidos: quer o sistema imunitário artificial, proposto pela primeira vez em sistemas deste género, quer o algoritmo genético, que já foi proposto em implementações prévias de sistemas de orquestração. A fase de avaliação compreendeu avaliações subjetivas e objetivas à similaridade e diversidade das soluções obtidas. Finalmente procedeu-se à validação do novo algoritmo de pesquisa proposto nesta implementação: o sistema imunitário artificial.

6.1 Método

Os algoritmos de pesquisa desenvolvidos apresentam comportamentos diferentes no seu funcionamento. Como já foi descrito previamente, o algoritmo genético apresenta um comportamento de convergência para uma única solução ótima, enquanto que o sistema imunitário artificial apresenta convergência para várias soluções ótimas correspondentes a regiões diferentes do espaço de pesquisa. Esta diferença patenteia a principal vantagem da utilização de um sistema imunitário artificial num problema de otimização multidimensional como é o caso no problema proposto. A comparação entre estes dois algoritmos foi feita recorrendo às mesmas funções de *fitness*, à mesma base de dados sonora e utilizando os mesmos sons objetivo. Os resultados obtidos foram comparados em termos de valor de aptidão. Ou seja os diferentes algoritmos foram comparados apenas em termos de avaliação objetiva de similaridade.

A avaliação objetiva de similaridade compreendeu a comparação dos valores de aptidão obtidos nas diferentes orquestrações. Esta avaliação foi feita tanto para as soluções obtidas com o algoritmo genético como para as soluções obtidas com o sistema imunitário artificial.

	Som Objetivo		
	Car Horn	Synth	Wind Harp
Sistema Imunitário Artificial	-0.019032	-0.010264	-0.016615
Algoritmo Genético	-0.029667	-0.020635	-0.033281

Tabela 6.1: Valores de aptidão dos melhores sons orquestrados obtidos pelos diferentes algoritmos para 3 sons objetivo

Devido ao comportamento do algoritmo genético de convergência para uma solução única foram realizados diversos ciclos de execução tendo sido obtidas várias soluções diferentes enquanto que, utilizando o sistema imunitário artificial, apenas num ciclo de execução se obteve um máximo de 10 soluções diferentes. É importante realçar que as soluções obtidas pelo algoritmo genético apresentam um valor de aptidão mais baixo do que as obtidas pelo sistema imunitário artificial, deduzindo-se assim que as soluções obtidas pelo algoritmo genético serão teoricamente piores do que aquelas obtidas pelo sistema imunitário artificial. Uma comparação entre os valores de aptidão obtidos com o algoritmo genético e com o sistema imunitário artificial é apresentada na tabela 6.1. Na tabela referida disponibilizam-se os valores de aptidão obtidos para diferentes sons objetivo utilizando os dois algoritmos de pesquisa desenvolvidos. Os valores de aptidão de uma orquestração são sempre negativos e quanto mais próximos de zero estiverem mais similar com o objetivo a orquestração teoricamente será. Como pode ser observado pelos valores da tabela os resultados obtidos pelo sistema imunitário artificial apresentam valores de aptidão mais altos do que aqueles obtidos pelo algoritmo genético. Verifica-se assim que o novo algoritmo de pesquisa proposto apresenta para além de uma diversidade maior no conjunto de soluções, uma eficiência maior na obtenção das mesmas.

Para avaliar as soluções obtidas no que toca à similaridade percetiva foi desenvolvido um teste de escuta em que se pediu aos utilizadores uma quantificação da similaridade entre as soluções orquestradas e os sons objetivos. Este teste permitiu avaliar a qualidade das soluções obtidas pelo sistema desenvolvido. As soluções avaliadas foram obtidas utilizando apenas o sistema imunitário artificial como algoritmo de pesquisa. Este algoritmo retorna, dependendo do som objetivo, uma população com uma média de 50 indivíduos, tendo sido selecionados os 10 melhores para o teste de escuta desenvolvido. Todos os sons objetivo foram orquestrados utilizando 5 como o valor máximo de sons diferentes que podem compor uma orquestração. Quanto à similaridade objetiva esta foi avaliada utilizando os valores de aptidão das soluções recorrendo às funções de *fitness* descritas previamente.

Outro aspeto avaliado diz respeito à diversidade das soluções obtidas. Para tal considerou-se diversidade objetiva como o número de sons diferentes entre duas orquestrações diferentes. O cálculo da diversidade objetiva foi feito da seguinte forma: considerando duas orquestrações compostas por n sons que apresentem em comum k sons, então a diversidade objetiva será:

$$div = 1 - \frac{k}{n} \quad (6.1)$$

Desta forma a diversidade objetiva será um valor entre 0 e 1, correspondendo o valor 0 a valores baixos de diversidade (todos os sons em comum) e o valor 1 a valores altos de diversidade (nenhum som em comum). Como todos os sons orquestrados apresentam uma limitação de 5 sons diferentes então n será sempre igual a 5 para os sons objetivo utilizados. Para cada orquestração individual aplicou-se a equação 6.1, comparando-a com as restantes orquestrações do conjunto de 10 orquestrações retornadas pelo algoritmo de pesquisa. O valor final de diversidade objetiva de um conjunto de orquestrações corresponde ao valor médio da classificação da diversidade objetiva desse conjunto.

No teste de escuta desenvolvido pediu-se também aos utilizadores que avaliassem a diversidade percetiva entre os sons orquestrados. A diversidade percetiva compreende o quão diferentes entre si são um conjunto de 10 orquestrações para um dado som objetivo e depende do utilizador que realiza o teste de escuta a classificação desta diversidade. De tal forma, conseguiu-se avaliar se os diferentes sons orquestrados apresentam diversidade percetiva entre si.

Os resultados finais permitiram validar o novo algoritmo de pesquisa proposto: o sistema imunitário artificial. Este algoritmo para além de apresentar uma boa eficiência na obtenção de soluções apresenta também uma vantagem significativa em relação ao algoritmo genético no que toca à diversidade das soluções propostas. Esta diversidade patenteia a principal vantagem no sistema de orquestração desenvolvido comparativamente a implementações prévias.

6.2 Estrutura

Para o desenvolvimento do teste de escuta foi utilizada a *framework* BeagleJS [49]. Esta ferramenta oferece suporte nativo para testes semelhantes aos pretendidos, tendo sido adaptada para o teste de escuta desenvolvido. No teste desenvolvido é apresentado ao utilizador um som objetivo e um conjunto de sons orquestrados. O utilizador deve ouvir tanto o som objetivo como cada som orquestrado e avaliar a similaridade percetiva entre o som objetivo e os sons orquestrados. Esta avaliação é feita numa escala de “Muito Diferente” a “Muito Semelhante”, sendo a opção “Muito Diferente” destinada a sons completamente díspares e a opção “Muito Semelhante” destinada a sons praticamente iguais. O utilizador deve também quantificar a diversidade percetiva entre os sons orquestrados utilizando para isso a mesma escala com opções de “Muito Diferente” a “Muito Semelhante”, sendo a opção “Muito Diferente” destinada a sons muito diversos entre si e a opção “Muito Semelhante” destinada a sons muito semelhantes entre si. Para cada som objetivo foram selecionados um conjunto de 10 sons orquestrados, tendo sido utilizados 10 sons objetivos totalizando assim 100 sons orquestrados que foram avaliados segundo a sua similaridade percetiva com o objetivo e segundo a sua diversidade percetiva. Os sons orquestrados são apresentados ao utilizador de uma forma aleatória. Um exemplo da interface do teste de escuta desenvolvido pode ser visualizado na figura 6.1 e na figura 6.2.

O teste de escuta desenvolvido foi disponibilizado *online* numa página dedicada¹. Foram enviados pedidos de colaboração para diversas pessoas, tendo sido obtidos 18 resultados. Os resultados

¹<http://paginas.fe.up.pt/~ee10146/test/>

Musical Orchestration Using an Artificial Immune System

In this test the goal is to evaluate the **perceptual similarity** between a reference sound and orchestrations. The orchestrations are combinations of musical instrument sounds that aim to imitate the reference sound. Each time, you will be presented 1 reference sound and 10 orchestrations. For each orchestration you are asked to rate the **overall perceptual similarity** in relation to the reference sound using the following scale: Very Different; Different; Fairly Similar; Similar; Very Similar. The orchestrations are presented randomly and are not ordered by any **similarity** measure.

After rating the **perceptual similarity** in relation to the reference, you will be asked to evaluate the **overall diversity** of the orchestrations using the same scale as before. The **overall diversity** is your evaluation of how different the orchestrations are from one another.

The test shouldn't take more than 15 minutes. Please read carefully the following information, even though you are used to doing such tests!

Instructions:

- Each reference sound has 10 different orchestrations.
- Rate the **perceptual similarity** of the orchestrations only compared to the reference on top.
- Try to rate the **overall perceptual similarity** of an orchestration without focusing on specific perceptual features such as pitch or timbre.
- Rate the **overall diversity** of the 10 orchestrations with reference to one another.
- Please take the test in a quiet environment.
- We recommend the use of headphones.

Thank you for your time! If you want to find more information about this project click [here](#)

Available HTML5 browser features: WebAudioAPI, BlobAPI, WAV, Vorbis, MP3, AAC
This listening test has been created with [BeagleJS](#)

Figura 6.1: Interface do teste de escuta desenvolvido: mensagem inicial

obtidos foram processados utilizando o MATLAB e são apresentados na secção seguinte.

6.3 Resultados Obtidos

Apresentam-se de seguida os resultados do teste de escuta efetuado utilizando 10 sons objetivos com 10 orquestrações para cada som objetivo. A escala utilizada varia de 1 a 5, sendo o valor 1 atribuído à opção “Muito Diferente”, o valor 2 à opção “Diferente”, o valor 3 à opção “Razoavelmente Semelhante”, o valor 4 à opção “Semelhante” e o valor 5 à opção “Muito Semelhante”. Os dados recolhidos foram processados utilizando o MATLAB, e em cada figura apresentam-se os valores médios obtidos e os correspondentes desvios padrão assinalados na linha contínua. As orquestrações estão ordenadas por ordem decrescente de aptidão sendo que a primeira orquestração apresenta a melhor aptidão e a décima orquestração apresenta a pior aptidão.

Musical Orchestration Using an Artificial Immune System

Car Horn (1 of 10)

Reference	<input type="button" value="Play"/>	<input type="button" value="Stop"/>		Very Different		Different		Fairly Similar		Similar		Very Similar	
Test Item 1	<input type="button" value="Play"/>	<input type="button" value="Stop"/>		●		●		●		●		●	
Test Item 2	<input type="button" value="Play"/>	<input type="button" value="Stop"/>		●		●		●		●		●	
Test Item 3	<input type="button" value="Play"/>	<input type="button" value="Stop"/>		●		●		●		●		●	
Test Item 4	<input type="button" value="Play"/>	<input type="button" value="Stop"/>		●		●		●		●		●	
Test Item 5	<input type="button" value="Play"/>	<input type="button" value="Stop"/>		●		●		●		●		●	
Test Item 6	<input type="button" value="Play"/>	<input type="button" value="Stop"/>		●		●		●		●		●	
Test Item 7	<input type="button" value="Play"/>	<input type="button" value="Stop"/>		●		●		●		●		●	
Test Item 8	<input type="button" value="Play"/>	<input type="button" value="Stop"/>		●		●		●		●		●	
Test Item 9	<input type="button" value="Play"/>	<input type="button" value="Stop"/>		●		●		●		●		●	
Test Item 10	<input type="button" value="Play"/>	<input type="button" value="Stop"/>		●		●		●		●		●	
Overall Diversity				●		●		●		●		●	

00:00

Volume

Available HTML5 browser features: [WebAudioAPI](#), [BlobAPI](#), [WAV](#), [Vorbis](#), [MP3](#), [AAC](#)
This listening test has been created with [BeagleJS](#)

Figura 6.2: Interface do teste de escuta desenvolvido: escolha da similaridade e diversidade

Na base de dados sonora utilizada encontram-se presentes sons de instrumentos de cordas e sopro. Como cada som da base de dados corresponde a uma nota de um instrumento então o conteúdo espectral desses sons será bastante harmónico e a combinação de diferentes sons resultará num espectro composto por componentes harmónicas. Este facto permite supor que sons objetivo que apresentem um espectro com componentes harmónicas irão apresentar melhores resultados na orquestração do que sons que não apresentem essa harmonicidade por ser mais fácil encontrar combinações de instrumentos que assemelhem esse espectro. Como tal, e devido ao facto do sistema desenvolvido não considerar a evolução temporal dos sons objetivo, a escolha

Som Objetivo	Variação Temporal	Harmonicidade
Car Horn	Não	Baixa
Synth	Não	Alta
Tbreed	Não	Baixa
Ahh	Não	Alta
Wind Harp	Não	Alta
Didgeridoo	Não	Alta
Elephant	Sim	Baixa
Frog	Sim	Baixa
Girl Scream	Sim	Baixa
Gong	Sim	Baixa

Tabela 6.2: Sons objetivo utilizados no teste de escuta

dos sons objetivo a utilizar no teste de escuta foi feita tendo em conta a exploração destes dois aspetos. Assim, escolheu-se sons com bastante harmonicidade como por exemplo sons de instrumentos musicais (“Wind Harp” e “Didgeridoo”) e sons com pouca harmonicidade (“Car Horn”, “Elephant”, “Frog” e “Gong”). Foram também escolhidos sons que apresentam variação temporal e sons estáticos. Na tabela 6.2 apresentam-se os sons utilizados no teste de escuta acompanhados por uma pequena descrição em que se enumera quais os sons que apresentam variação temporal e o seu correspondente grau de harmonicidade.

Na tabela 6.3 apresentam-se os valores médios de classificação do conjunto de orquestrações e correspondentes referências das figuras contendo a classificação média de cada orquestração. A classificação média varia entre 1 a 5, sendo que quanto maior este valor melhor foi a classificação da similaridade percetiva. Ou seja valores mais altos de classificação média indicam orquestrações com uma similaridade percetiva mais alta. Como pode ser observado na tabela, a orquestração “Frog” (ver figura 6.11) apresenta os valores mais baixos de classificação média. A razão destes valores prende-se pelo facto de que o som objetivo utilizados nesta orquestração varia temporalmente, não apresentando por isso um comportamento estático. Como já foi referido previamente, o sistema de orquestração desenvolvido não considera a evolução temporal do timbre na orquestração de sons, devendo-se a esse facto os baixos valores de aptidão desta orquestração.

Apesar de o sistema desenvolvido não considerar a evolução temporal dos sons objetivo nos resultados obtidos verificam-se boas classificações para os sons “Girl Scream”, “Elephant” e “Gong”, sons estes que apresentam variação temporal. Este facto evidencia que a utilização de valores médios para a caracterização de cada som pode por vezes originar boas orquestrações, mesmo não considerando a evolução temporal. As orquestrações resultantes apresentam alguma semelhança com os sons objetivo, segundo os resultados obtidos.

O valor de aptidão das orquestrações utilizadas no teste de escuta são apresentados na tabela 6.5. Realça-se mais uma vez que os valores de aptidão de uma determinada orquestração são sempre negativos e quanto mais próximos de zero estiverem mais similar a orquestração teoricamente será com o som objetivo. Ou seja, teoricamente quanto maior o valor de aptidão melhor será a orquestração. O valor mais alto de aptidão corresponde à orquestração “Synth”. No entanto, este

Som Objetivo	Classificação Média	Figura
Car Horn	2.8889	6.4
Synth	1.8395	6.5
Tbreed	2.1667	6.6
Ahh	2.2901	6.7
Wind Harp	2.4320	6.8
Didgeridoo	2.2037	6.9
Elephant	2.4074	6.10
Frog	1.1914	6.11
Girl Scream	2.7716	6.12
Gong	2.4320	6.13

Tabela 6.3: Avaliação da similaridade perceptiva média das orquestrações submetidas ao teste de escuta

valor não reflete a maior similaridade perceptiva obtida no teste de escuta. Pelo contrário, a orquestração “Synth” apresenta dos valores mais baixos obtidos no teste. Na realidade, pode-se concluir que os valores de aptidão obtidos refletem de algum modo a similaridade entre sons. No entanto, não refletem na totalidade esta similaridade. Na verdade, excetuando a orquestração “Frog”, as restantes orquestrações apresentam valores de aptidão da mesma ordem de grandeza, concluindo-se que a resolução das funções de aptidão não é muito alta. Para corroborar este facto verifica-se que, nos resultados obtidos, apesar de as orquestrações estarem ordenadas por valor decrescente de aptidão, não se verifica uma variação nas classificações obtidas que comprovem que valores mais altos de aptidão correspondem a orquestrações melhores (ver figuras 6.4 a 6.13).

A orquestração com melhor classificação no teste de escuta foi a “Car Horn” (ver figura 6.4). A razão desta boa classificação prende-se pelo facto de que o som objetivo desta orquestração é constituído por duas séries harmónicas, correspondendo a duas notas musicais diferentes. De tal forma, torna-se fácil encontrar combinações de instrumentos que se assemelhem a este som objetivo. Tal como nas outras orquestrações verifica-se que o conjunto de orquestrações obtidas apesar de estar ordenado por aptidão, não está ordenado por similaridade perceptiva, segundo os resultados obtidos. Esta é mais uma razão que confirma que a resolução das funções de aptidão não é muito alta.

As restantes orquestrações apresentam valores semelhantes de similaridade perceptiva. Um dos participantes do teste de escuta referiu que algumas das orquestrações apresentam diferenças em termos de altura. Este comentário fornece mais uma justificação para a diversidade que o conjunto de orquestrações apresenta, existindo orquestrações que têm uma altura diferente da que o som objetivo. Outro comentário de um dos participantes referiu que, no geral, as orquestrações submetidas a teste não permitem reconhecer o som objetivo. Este comentário permite concluir que a subjetividade do teste de escuta desenvolvido não permite uma avaliação objetiva das orquestrações submetidas a avaliação. Efetivamente os desvios padrão dos dados obtidos são significativamente altos, deduzindo-se que não existe concordância entre os participantes quanto à classificação das orquestrações.

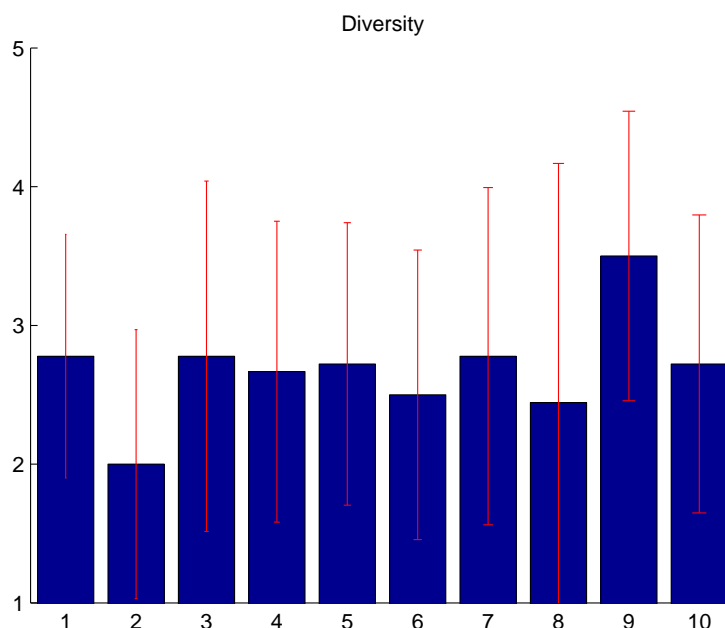


Figura 6.3: Resultados do teste de escuta: diversidade perceptiva

Os resultados da avaliação da diversidade perceptiva podem ser visualizados na figura 6.3. Nesta figura é utilizada uma escala de 1 a 5, correspondendo o valor 1 a valores altos de diversidade perceptiva e o valor 5 a valores baixos de diversidade perceptiva. O valor médio da classificação é próximo de 3, podemos por isso afirmar que os resultados obtidos no teste de escuta confirmam que o sistema produz orquestrações diversas entre si para um mesmo som objetivo. É de notar que as orquestrações foram obtidas através de uma única execução da aplicação, oferecendo assim ao utilizador um conjunto diverso de orquestrações que podem depois ser refinadas ou combinadas.

A diversidade objetiva dos sons avaliados no teste de escuta é apresentada na tabela 6.4. Nesta tabela os valores foram obtidos utilizando a equação 6.1 e representam numa escala de 0 a 1 quão diversos são os sons de cada conjunto de orquestrações. Valores altos de diversidade objetiva correspondem a diversidade alta. Os dados presentes na tabela corroboram os dados obtidos no teste de escuta, nomeadamente os valores de diversidade subjetiva. Na tabela referida pode-se verificar que a diversidade objetiva das orquestrações utilizadas no teste de escuta apresentam valores altos, correspondendo a uma diversidade alta de sons. Ou seja, o conjunto de sons orquestrados apresentam uma baixa inter-similaridade entre si, corroborando os dados de diversidade perceptiva obtidos no teste de escuta.

No geral, e após considerar os resultados obtidos no teste de escuta, podemos afirmar que as funções de aptidão implementadas não refletem na totalidade a similaridade entre sons mas são uma boa aproximação a esta similaridade, tal como esperado. Efetivamente, as funções de aptidão utilizadas foram desenvolvidas em implementações anteriores de um sistema de orquestração tendo sido apenas adaptadas para este trabalho sendo que a sua validade foi já verificada [18].

Som Objetivo	Diversidade	
	Média	Desvio Padrão
Car Horn	0.7822	0.1576
Synth	0.8444	0.1333
Tbreed	0.8178	0.1518
Ahh	0.8400	0.1380
Wind Harp	0.8000	0.1529
Didgeridoo	0.6356	0.1658
Elephant	0.8578	0.1315
Frog	0.7822	0.1518
Girl Scream	0.8489	0.1351
Gong	0.8222	0.1542

Tabela 6.4: Valores médios de diversidade objetiva e desvios padrões dos sons utilizados no teste de escuta

Considerando os valores de aptidão da tabela 6.5, verifica-se que a similaridade percetiva depende da ordem de grandeza do valor de aptidão, sendo que para valores da mesma ordem de grandeza nem sempre se verifica uma similaridade percetiva mais alta para valores de aptidão mais altos. Para além disso, verifica-se que vários fatores influenciam uma boa orquestração. Nomeadamente, o facto de haver variação no timbre do som objetivo e o conteúdo espectral do mesmo, influenciam a similaridade das orquestrações obtidas utilizando o sistema desenvolvido. Na realidade verifica-se que sons que apresentem variação tímbrica no tempo irão apresentar orquestrações piores e sons que apresentem um conteúdo espectral formado por componentes harmónicas irão apresentar orquestrações melhores. No que toca à diversidade podemos afirmar que a aplicação desenvolvida providencia soluções diversas e únicas, oferecendo a este sistema uma vantagem significativa em relação a implementações prévias.

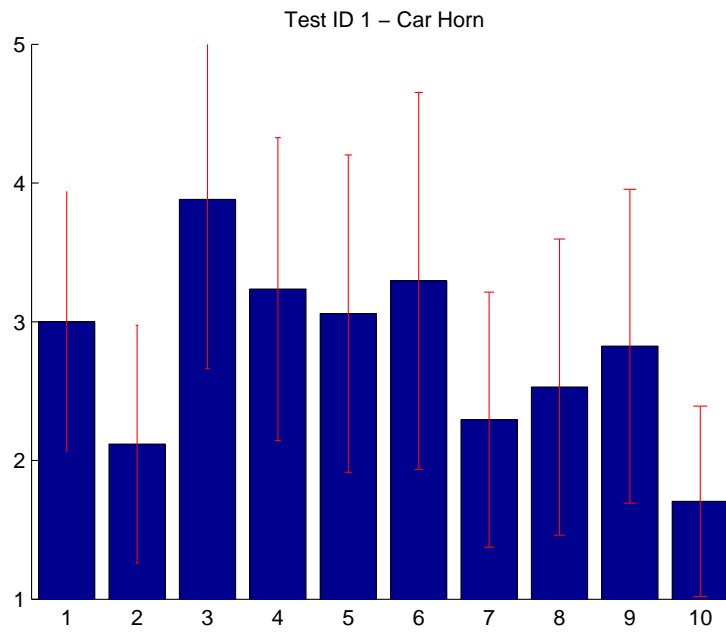


Figura 6.4: Resultados do teste de escuta: som objetivo “Car Horn”

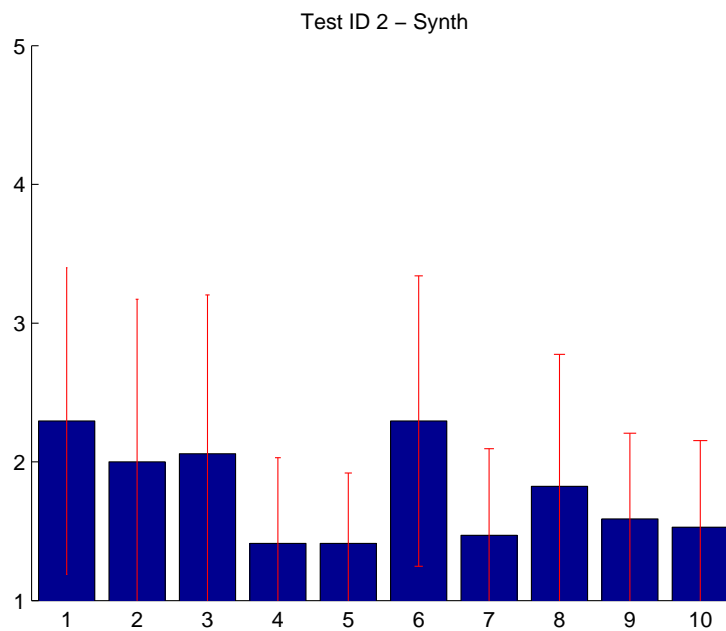


Figura 6.5: Resultados do teste de escuta: som objetivo “Synth”

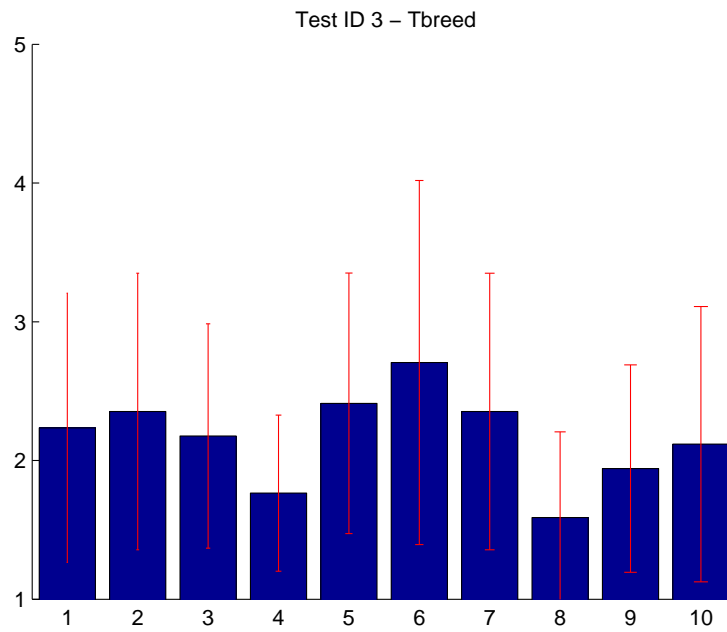


Figura 6.6: Resultados do teste de escuta: som objetivo “Tbreed”

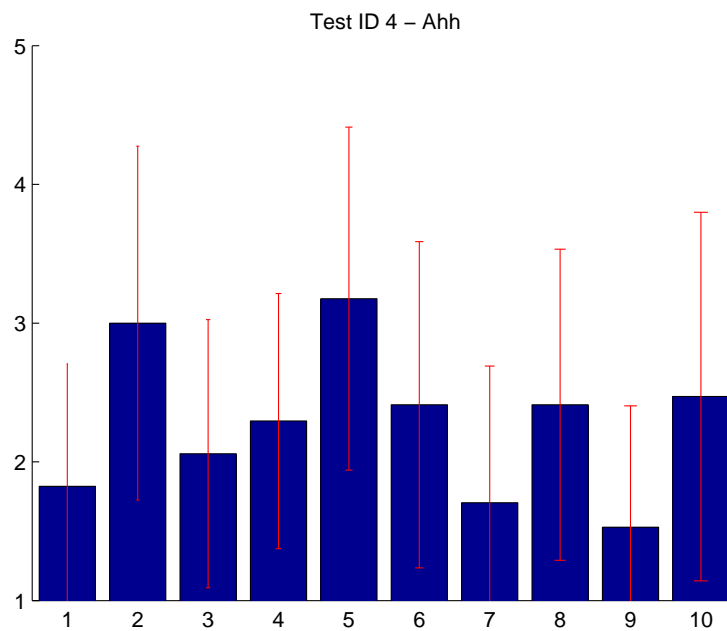


Figura 6.7: Resultados do teste de escuta: som objetivo “Ahh”

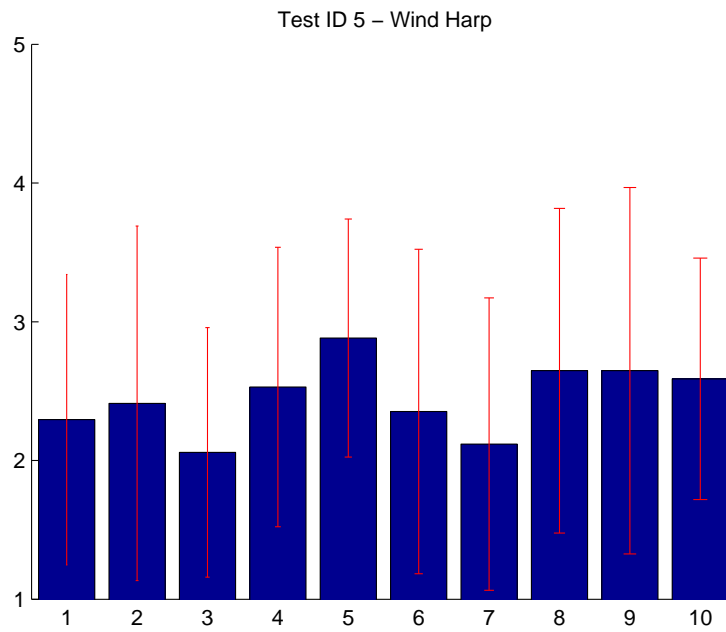


Figura 6.8: Resultados do teste de escuta: som objetivo “Wind Harp”

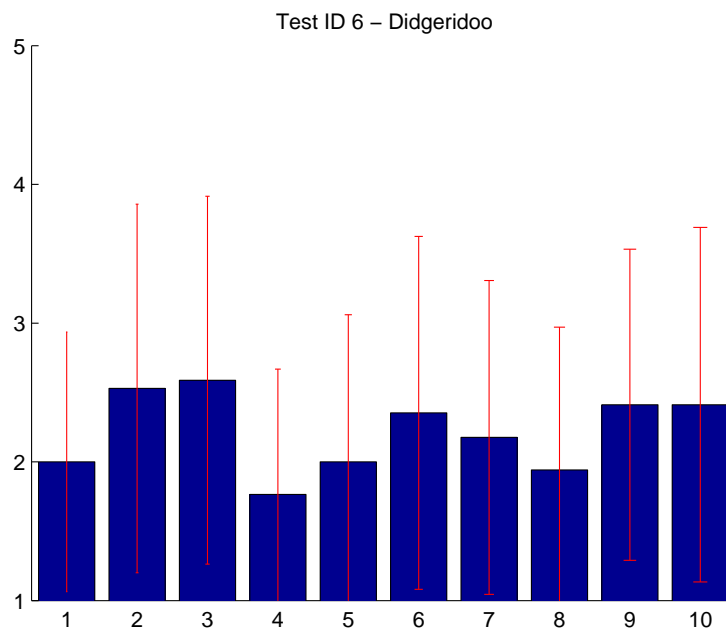


Figura 6.9: Resultados do teste de escuta: som objetivo “Didgeridoo”

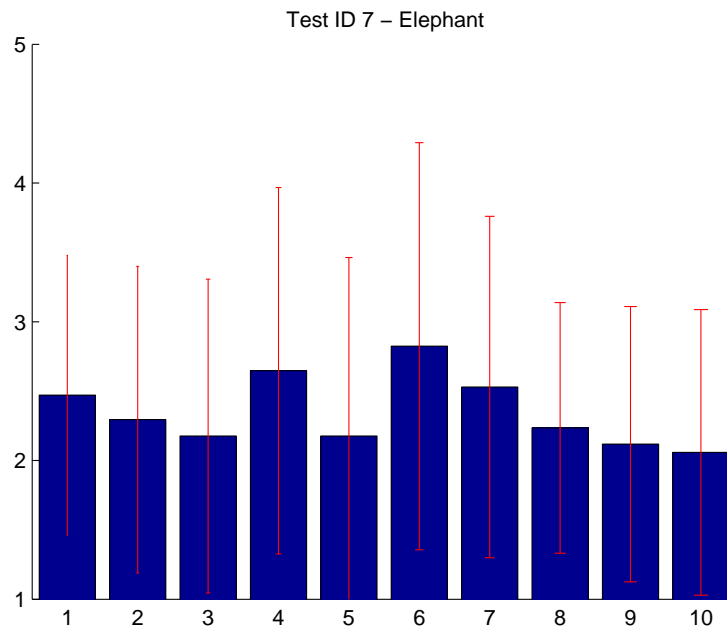


Figura 6.10: Resultados do teste de escuta: som objetivo “Elephant”

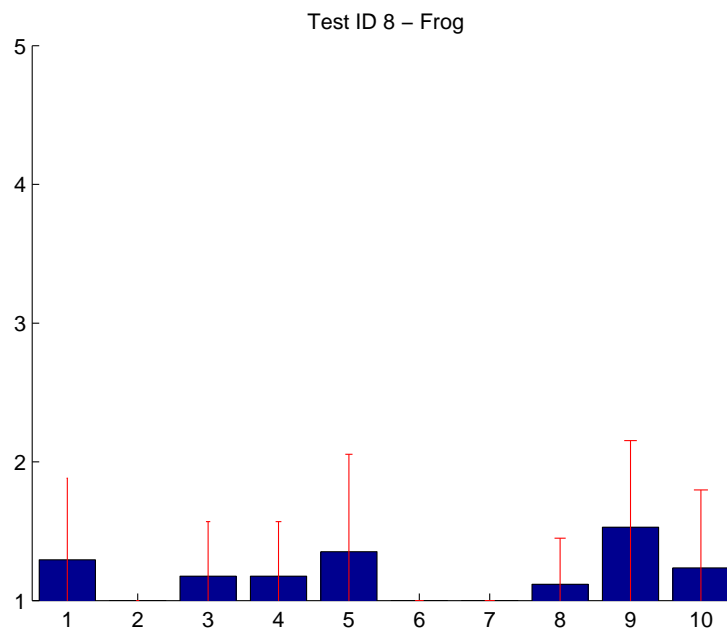


Figura 6.11: Resultados do teste de escuta: som objetivo “Frog”

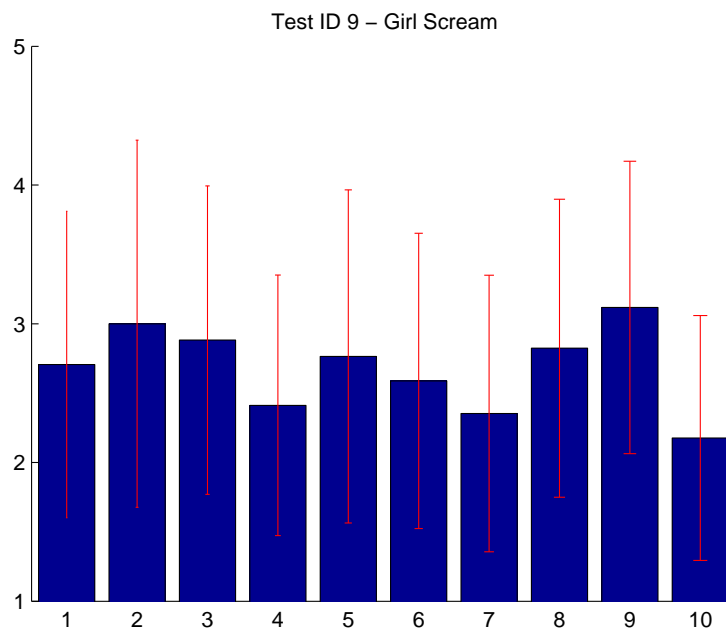


Figura 6.12: Resultados do teste de escuta: som objetivo “Girl Scream”

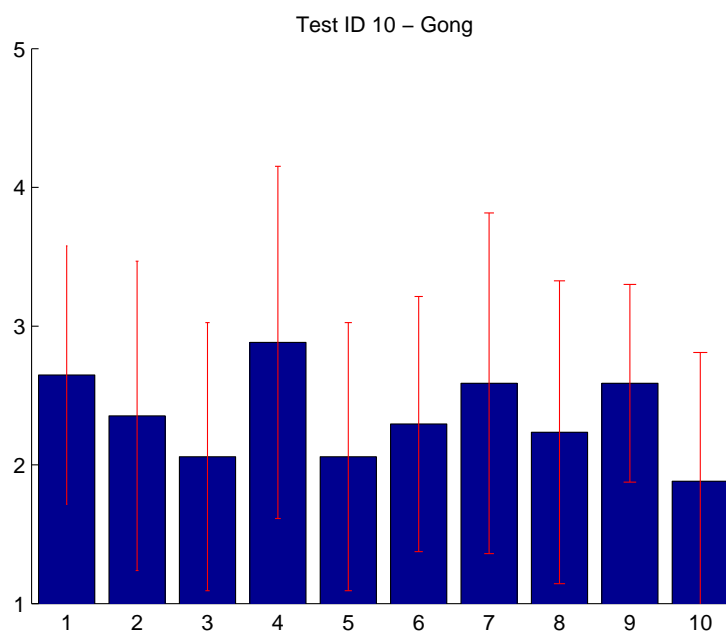


Figura 6.13: Resultados do teste de escuta: som objetivo “Gong”

Som Objetivo										
	Car Horn	Synth	Tbreed	Ahh	Wind Harp	Didgeridoo	Elephant	Frog	Girl Scream	Gong
1	-0.019032	-0.010264	-0.046056	-0.036213	-0.016615	-0.012187	-0.056997	-0.140671	-0.044604	-0.040410
2	-0.022489	-0.010788	-0.046735	-0.042767	-0.018130	-0.013198	-0.061252	-0.145754	-0.053563	-0.043728
3	-0.022833	-0.011109	-0.048572	-0.046089	-0.021532	-0.013334	-0.061455	-0.148071	-0.055773	-0.047353
4	-0.023788	-0.013245	-0.050426	-0.046302	-0.021670	-0.015652	-0.062501	-0.150639	-0.055896	-0.047757
5	-0.023828	-0.013481	-0.052122	-0.046696	-0.022255	-0.016122	-0.064109	-0.152719	-0.058683	-0.048497
6	-0.024953	-0.016818	-0.054270	-0.046843	-0.022936	-0.016403	-0.064173	-0.158426	-0.059136	-0.050172
7	-0.025485	-0.017298	-0.055639	-0.047404	-0.025648	-0.018349	-0.067149	-0.160608	-0.059827	-0.055160
8	-0.026717	-0.018254	-0.057451	-0.047629	-0.029635	-0.018985	-0.068409	-0.160731	-0.060011	-0.056150
9	-0.026875	-0.018349	-0.059322	-0.047819	-0.030277	-0.019328	-0.068552	-0.161910	-0.066271	-0.056473
10	-0.027186	-0.018865	-0.059383	-0.047865	-0.030322	-0.021939	-0.070081	-0.163506	-0.072653	-0.057448

Tabela 6.5: Valores de aptidão dos sons utilizados no teste de escuta

Capítulo 7

Conclusões e Trabalho Futuro

Apresentam-se de seguida as conclusões obtidas após a realização deste trabalho assim como os aspetos que devem ser melhorados futuramente para tornar o sistema desenvolvido mais eficiente.

7.1 Conclusões

A realização deste trabalho envolveu tanto aspetos musicais como aspetos tecnológicos que foram explorados e aplicados no sistema desenvolvido. Desde a definição de uma formalização para o problema de orquestração musical até à implementação de um algoritmo de pesquisa eficiente, várias vertentes tiveram de ser tomadas em conta para a construção de um sistema de orquestração eficiente. O sistema final preza praticamente pela completa autonomia com que as orquestrações são obtidas, autonomia esta que por vezes pode tornar-se numa desvantagem ao utilizador como será referido na secção seguinte. Para além deste facto, a utilização de um sistema imunitário artificial oferece uma possibilidade que sistemas de orquestração anteriores não oferecem: a possibilidade de obter diversas orquestrações simultaneamente para um mesmo som objetivo. Outra vantagem que este sistema apresenta é o facto de a aplicação ter sido desenvolvida utilizando módulos que são praticamente independentes entre si. Desta maneira torna-se fácil a implementação futura de, por exemplo, mais descritores para os sons ou um algoritmo de pesquisa diferente.

A primeira conclusão que pode ser obtida deste trabalho é o facto de que o timbre como fenómeno acústico apresenta características únicas e difíceis de formalizar analiticamente. Efetivamente, os descritores utilizados na descrição dos sons presentes na base de dados são suficientes para uma implementação básica de um sistema de orquestração mas não endereçam na totalidade o fenómeno tímbrico. Esta é uma área em que a investigação é decorrente, não existindo ainda concordância sobre os aspetos físicos que influenciam o fenómeno tímbrico.

Em segundo lugar, e como foi descrito no capítulo 6, vários aspetos influenciam a definição de uma boa orquestração. Como a evolução temporal do timbre não foi considerada na implementação do sistema desenvolvido é natural que sons que apresentem variação temporal originem orquestrações piores do que sons que não apresentam essa variação. Esta é uma das principais dificuldades na implementação deste tipo de sistemas, não tendo sido endereçada nem nesta implementação nem na implementação que serviu de base ao sistema desenvolvido [18]. Existe contudo uma implementação que já é capaz de lidar com a evolução temporal do timbre [50]. Para além desta limitação, sons que apresentem um conteúdo espectral formado por componentes harmónicas apresentam orquestrações com um grau de similaridade mais alto. Efetivamente, como a base de dados sonora é formada por notas de instrumentos musicais e os correspondentes sons apresentam componentes harmónicas, é natural que sons que apresentem essa mesma harmonicidade não sejam difíceis de orquestrar.

Finalmente, a utilização de um sistema imunitário artificial fornece ao sistema desenvolvido uma mais valia na diversidade e qualidade das soluções obtidas. Implementações prévias utilizaram um algoritmo genético para obter as soluções orquestradas. Este tipo de algoritmo foi também utilizado inicialmente neste trabalho tendo sido obtidos resultados com valores de aptidão mais baixos do que aqueles que foram obtidos através da utilização de um sistema imunitário artificial. Através de uma única execução do algoritmo é possível o retorno de várias soluções ótimas que se encontram próximas do som objetivo. Esta mais valia marca mais um ponto importante que destaca o sistema desenvolvido das implementações prévias. Os resultados obtidos por este algoritmo apresentaram boas classificações no teste de escuta desenvolvido, justificando o uso deste algoritmo em futuras implementações de sistemas de orquestração musical.

A escolha do algoritmo opt-aiNet como algoritmo de pesquisa prendeu-se pelo facto dos bons resultados que este apresenta em problemas multidimensionais. Este algoritmo foi significativamente alterado na implementação utilizada neste sistema, no entanto, o funcionamento intrínseco continuou o mesmo. Estas alterações foram realizadas tendo em conta as restrições do problema de orquestração e resultaram num algoritmo capaz de otimizar um espaço de pesquisa discreto que apresenta bons resultados, comparativamente ao algoritmo genético utilizado. Este algoritmo pode ser futuramente utilizado noutro tipo de problemas com restrições semelhantes ao problema da orquestração musical.

No geral, o sistema desenvolvido apresenta um bom desempenho na orquestração de sons musicais. As funções de *fitness* implementadas refletem eficientemente a similaridade entre orquestrações e os sons objetivo, tornando-se assim numa boa medida a utilizar nos algoritmos de pesquisa. Os descritores escolhidos não refletem na totalidade a descrição de cada som, mas são no entanto suficientes para uma orquestração razoável. Como descrito no capítulo 6 a similaridade percetiva das orquestrações, que foi obtida através de testes de escuta, apresenta bons valores para as orquestrações testadas. Isto é, os utilizadores consideraram que a maior parte das orquestrações refletem de alguma forma o som objetivo.

Os resultados deste trabalho, para além de terem sido disponibilizados na página web dedicada

ao tema¹, foram também parcialmente apresentados no 1º Congresso Doutoral em Engenharia realizado na Faculdade de Engenharia da Universidade do Porto. A participação envolveu a escrita de um artigo [1] e a realização de um poster cujo conteúdo refletiu o trabalho desenvolvido até à data de realização deste congresso.

7.2 Trabalho Futuro

O sistema de orquestração desenvolvido apresenta um bom desempenho na orquestração de sons musicais, no entanto, vários aspetos necessitam de ser melhorados para uma possível utilização desta ferramenta por músicos profissionais e compositores. Efetivamente, nesta fase o sistema desenvolvido apresenta diversas lacunas que impossibilitam um melhor aproveitamento das potencialidades da orquestração com recurso a ferramentas computacionais.

A principal lacuna do sistema desenvolvido é a consideração do timbre como fenómeno estático que não evolui com o tempo. Tal como foi descrito no capítulo 2, o timbre é um fenómeno multidimensional que evolui com o tempo, no entanto, para simplificar o desenvolvimento do sistema proposto foi necessário descartar a evolução temporal do timbre, utilizando-se apenas valores médios na caracterização de cada som. Esta consideração impossibilita ao sistema desenvolvido quer a orquestração de sons que apresentem uma variação temporal quer a utilização na base de dados de sons que também apresentem essa variação. Por exemplo, sons de instrumentos de percussão não podem ser utilizados na base de dados por apresentarem exatamente essa variação. A colmatação desta lacuna ofereceria ao utilizador um potencial muito maior por permitir a orquestração de praticamente qualquer som existente.

Outra lacuna existente no sistema é o facto de não haver nenhum controlo sobre o processo de orquestração. Nomeadamente, não existe nenhum mecanismo que permita ao utilizador especificar, por exemplo, quais os instrumentos que quer que façam parte no processo de orquestração, ou especificar uma determinada dinâmica que deva ser usada, ou até especificar um número máximo e mínimo de instrumentos que devem fazer parte da solução. Da forma como o sistema se encontra desenvolvido atualmente tais restrições não são possíveis. Efetivamente, o sistema desenvolvido preza pela completa autonomia no processo de orquestração, sendo que o utilizador apenas é capaz de especificar o número máximo de sons diferentes que podem fazer parte das soluções orquestradas. Assim, o controlo sobre o processo de orquestração é praticamente nulo da parte do utilizador. Futuramente, a adição de um mecanismo que permita controlar minimamente o processo de orquestração poderia aumentar em grande escala a utilidade do sistema, especialmente para um utilizador que pretenda obter um som orquestrado de entre um leque de restrições.

Outra possível melhoria seria a adição de mais descritores na caracterização de cada som, podendo esta adição resultar numa melhoria dos resultados do sistema. Inicialmente, no desenvolvimento deste sistema estavam a ser utilizados 3 descritores para a caracterização de cada som: centróide espectral, espalhamento espectral, picos espectrais. Com a adição de um quarto descritor (intensidade sonora), verificou-se uma melhoria significativa na qualidade dos resultados obtidos.

¹<http://paginas.fe.up.pt/~ee10146/#thesis>

Assim é expetável que a adição de mais descritores possa aumentar a qualidade dos resultados à custa de um maior custo computacional no cálculo da aptidão de cada solução.

No que toca aos resultados obtidos, uma comparação com soluções propostas por utilizadores humanos seria uma mais valia para a comparação da qualidade das soluções propostas pelo sistema. Neste teste seria pedido a um certo número de compositores que orquestrassem um som tendo como referência o som objetivo. De igual maneira a aplicação seria utilizada para orquestrar o mesmo som objetivo e as soluções obtidas quer pelos compositores quer pelo sistema seriam comparadas. Desta forma seria possível comparar a qualidade e utilidade das soluções propostas pelo sistema.

Finalmente, a criação de uma interface gráfica mais apelativa poderia aumentar a facilidade de interação com o sistema por parte dos utilizadores. Na forma atual, o utilizador necessita de correr um *script* MATLAB fornecendo a localização da base de dados, a localização do som objetivo e o número máximo de sons que podem estar presentes nas soluções. A substituição deste *script* por uma interface gráfica em que o utilizador escolheria os diversos parâmetros numa forma muito mais intuitiva melhoraria a usabilidade do sistema. A interface gráfica apresentada no final da execução (ver figura 5.9) pode também ser melhorada de forma a permitir ao utilizador, por exemplo, a edição manual dos índices das soluções ou a permitir a combinação de soluções.

7.3 Agradecimentos

O projeto *Media, Arts and Technologies* (MAT), NORTE-07-0124-FEDER-000061, é financiado pelo Programa Operacional Regional do Norte (ON.2 – O Novo Norte), sobre o Quadro de Referência Estratégica Nacional (QREN), através do Fundo Europeu de Desenvolvimento Regional (FEDER), e por fundos nacionais, através da agência de financiamento Portuguesa, Fundação para a Ciência e a Tecnologia (FCT).

Referências

- [1] José Abreu, Marcelo Caetano, e Rui Penha. Musical Orchestration Using Biological Principles. Em *1st Doctoral Congress in Engineering*, 2015.
- [2] John Grey. Multidimensional Perceptual Scaling of Musical Timbres. *The Journal of the Acoustical Society of America*, 61(5):1270–1277, 1977.
- [3] Grégoire Carpentier, Gérard Assayag, e Emmanuel Saint-James. Solving the Musical Orchestration Problem using Multiobjective Constrained Optimization with a Genetic Local Search Approach. *Journal of Heuristics*, 16(5):681–714, 2010.
- [4] Philip Galanter. The Problem with Evolutionary Art Is ... Em Cecilia Di Chio, Anthony Brabazon, Gianni Di Caro, Marc Ebner, Muddassar Farooq, Andreas Fink, Jörn Grahl, Gary Greenfield, Penousal Machado, Michael O’Neill, Ernesto Tarantino, e Neil Urquhart, editores, *Applications of Evolutionary Computation*, volume 6025 de *Lecture Notes in Computer Science*, páginas 321–330. Springer Berlin Heidelberg, 2010.
- [5] Marcelo Caetano, Jônatas Manzolli, e Fernando Von Zuben. Interactive Control of Evolution Applied to Sound Synthesis. Em *The Florida AI Research Society Conference*, páginas 51–56, 2005.
- [6] Marcelo Caetano, Jônatas Manzolli, e Fernando Von Zuben. *Application of an Artificial Immune System in a Compositional Timbre Design Technique*, volume 3627 de *Lecture Notes in Computer Science*, páginas 389–403. Springer Berlin Heidelberg, 2005.
- [7] Artemis Moroni, Jônatas Manzolli, Fernando Von Zuben, e Ricardo Gudwin. Vox Populi: An Interactive Evolutionary System for Algorithmic Music Composition. *Leonardo Music Journal*, 10:49–54, 2000.
- [8] John Biles. GenJam: A Genetic Algorithm for Generating Jazz Solos. Em *Proceedings of the International Computer Music Conference*, páginas 131–131. International Computer Music Association, 1994.
- [9] Maria Navarro, Marcelo Caetano, Gilberto Bernandes, Leandro Castro, e Juan Corchado. Automatic Generation of Chord Progressions with an Artificial Immune System. Em *Proceedings of EVOMUSART 2015*, 2015.
- [10] Otavio Carpinteiro. A Self-Organizing Map Model for Analysis of Musical Time Series. Em *Neural Networks, 1998. Proceedings. Vth Brazilian Symposium on*, páginas 140–145, Dec 1998.
- [11] Pedro de León e José Inesta. Musical Style Identification using Self-Organising Maps. Em *Web Delivering of Music, 2002. WEDELMUSIC 2002. Proceedings. Second International Conference on*, páginas 82–89. IEEE, 2002.

- [12] Markus Frühwirth e Andreas Rauber. *Self-Organizing Maps for Content-Based Music Clustering*, páginas 228–233. Springer, 2002.
- [13] Leandro Castro e Fernando Von Zuben. *aiNet: An Artificial Immune Network for Data Analysis*, chapter 12, páginas 231–259. Idea Group Publishing, 2001.
- [14] Stephen McAdams, Suzanne Winsberg, Sophie Donnadiou, Geert De Soete, e Jochen Krimphoff. Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes. *Psychological Research*, 58(3):177–192, 1995.
- [15] David Psenicka. SPORCH: An Algorithm for Orchestration Based on Spectral Analyses of Recorded Sounds. Em *Proceedings of International Computer Music Conference (ICMC)*, página 184, 2003.
- [16] Thomas Hummel. Simulation of Human Voice Timbre by Orchestration of Acoustic Music Instruments. Em *Proceedings of the International Computer Music Conference (ICMC)*, página 185, 2005.
- [17] Grégoire Carpentier, Damien Tardieu, Jonathan Harvey, Gérard Assayag, e Emmanuel Saint-James. Predicting Timbre Features of Instrument Sound Combinations: Application to Automatic Orchestration. *Journal of New Music Research*, 39(1):47–61, 2010.
- [18] Gregoire Carpentier. *Approche Computationnelle de L’Orchestration Musciale-Optimisation Multicritère sous Contraintes de Combinaisons Instrumentales dans de Grandes Banques de Sons*. Tese de doutoramento, Universite Pierre et Marie Curie-Paris VI, 2008.
- [19] Damien Tardieu e Xavier Rodet. An Instrument Timbre Model for Computer Aided Orchestration. Em *Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop*, páginas 347–350. IEEE, 2007.
- [20] Silvano Martello e Paolo Toth. *Knapsack Problems: Algorithms and Computer Implementations*. John Wiley & Sons, Inc., New York, NY, USA, 1990.
- [21] Stephen Handel. Timbre Perception and Auditory Object Identification. *Hearing*, páginas 425–461, 1995.
- [22] Marcelo Caetano e Xavier Rodet. Musical Instrument Sound Morphing Guided by Perceptually Motivated Features. *Audio, Speech, and Language Processing, IEEE Transactions on*, 21(8):1666–1675, Aug 2013.
- [23] John Grey e John Gordon. Perceptual Effects of Spectral Modifications on Musical Timbres. *The Journal of the Acoustical Society of America*, 63(5):1493–1500, 1978.
- [24] Carol Krumhansl. Why is Musical Timbre so Hard to Understand. *Structure and perception of electroacoustic sound and music*, 9:43–53, 1989.
- [25] Jochen Krimphoff, Stephen McAdams, e Suzanne Winsberg. Caractérisation du Timbre des Sons Complexes.II. Analyses Acoustiques et Quantification Psychophysique. *J. Phys. IV France*, 04(C5):625–628, 1994.
- [26] Anne Caclin, Stephen McAdams, Bennett Smith, e Suzanne Winsberg. Acoustic Correlates of Timbre Space Dimensions: A Confirmatory Study Using Synthetic Tones. *The Journal of the Acoustical Society of America*, 118(1):471–482, 2005.

- [27] Perfecto Herrera, Xavier Serra, e Geoffroy Peeters. Audio Descriptors and Descriptor Schemes in the Context of MPEG-7. *Proceedings of the 1999 ICMC, Beijing, China, 1999*.
- [28] Alex Fraser. Simulation of Genetic Systems by Automatic Digital Computers. II. Effects of Linkage on Rates of Advance Under Selection. *Australian Journal of Biological Science*, 10:492–499, 1957.
- [29] Hans Bremermann. *The Evolution of Intelligence: The Nervous System as a Model of Its Environment*. University of Washington, Department of Mathematics, 1958.
- [30] John Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, 1975.
- [31] Melanie Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, 1996.
- [32] Leandro De Castro e Fernando Von Zuben. Artificial Immune Systems: Part I - Basic Theory and Applications. *Universidade Estadual de Campinas, Tech. Rep*, 210, Dec 1999.
- [33] Frank Burnet. *Clonal Selection and After*, páginas 63–85. Marcel Dekker Inc, 1978.
- [34] Frank Burnet. *The Clonal Selection Theory of Acquired Immunity*. University Press Cambridge, 1959.
- [35] Leandro Castro e Jon Timmis. An Artificial Immune Network for Multimodal Function Optimization. Em *Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on*, volume 1, páginas 699–704, May 2002.
- [36] Gustav Nossal. Negative Selection of Lymphocytes. *Cell*, 76(2):229–239, 1994.
- [37] Polly Matzinger. Tolerance, Danger, and the Extended Family. *Annual review of immunology*, 12(1):991–1045, 1994.
- [38] Uwe Aickelin e Steve Cayzer. The Danger Theory and Its Application to Artificial Immune Systems. *arXiv preprint arXiv:0801.3549*, 2008.
- [39] Niels Jerne. Towards a Network Theory of the Immune System. Em *Annales d'immunologie*, volume 125, páginas 373–389, 1974.
- [40] Carlos Coello e Nareli Cortés. Solving Multiobjective Optimization Problems Using an Artificial Immune System. *Genetic Programming and Evolvable Machines*, 6(2):163–190, 2005.
- [41] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, e Ryuichi Oka. RWC Music Database: Popular, Classical and Jazz Music Databases. Em *ISMIR*, volume 2, páginas 287–288, 2002.
- [42] Geoffroy Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Relatório técnico, IRCAM, 2004.
- [43] Harry Olson. The Measurement of Loudness. *Audio Magazine*, páginas 18–22, 1972.
- [44] Arturo Camacho. *SWIPE: A Sawtooth Waveform Inspired Pitch Estimator for Speech and Music*. Tese de doutoramento, University of Florida, 2007.

- [45] Adam Lipowski e Dorota Lipowska. Roulette-wheel Selection via Stochastic Acceptance. *Physica A: Statistical Mechanics and its Applications*, 391(6):2193–2196, 2012.
- [46] Olivier Lartillot e Petri Toiviainen. A Matlab Toolbox for Musical Feature Extraction from Audio. Em *International Conference on Digital Audio Effects*, páginas 237–244, 2007.
- [47] Dan Ellis. A Phase Vocoder in Matlab. <http://labrosa.ee.columbia.edu/matlab/pvoc/>, 2002. Acedido: 13-07-2015.
- [48] Malcolm Slaney. Auditory Toolbox. Relatório técnico 45, Apple Computer, Inc., 1994.
- [49] Sebastian Kraft e Udo Zölzer. BeagleJS: HTML5 and JavaScript based Framework for the Subjective Evaluation of Audio Quality. Em *Linux Audio Conference*. Karlsruhe, Germany, 2014.
- [50] Philippe Esling, Grégoire Carpentier, e Carlos Agon. *Dynamic Musical Orchestration Using Genetic Algorithms and a Spectro-Temporal Description of Musical Instruments*, volume 6025 de *Lecture Notes in Computer Science*, páginas 371–380. Springer Berlin Heidelberg, 2010.