

UNIVERSIDADE FEDERAL DO RIO GRANDE
CENTRO DE CIÊNCIAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO
CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

Dissertação de Mestrado

**Ulises: Um Sistema Baseado em Agentes Para
Classificação de Timbres**

Eduardo Porto Teixeira

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande, como requisito parcial para a obtenção do grau de Mestre em Engenharia de Computação

Orientadora: Prof^a. Dr^a. Diana Francisca Adamatti

Rio Grande, 2016

Dedico este trabalho em memória ao professor, amigo, e idealizador da minha dissertação de mestrado Dr. Ulises Dardo Ferretti Martínez, exemplo de vida e grande profissional com quem fiz aulas de flauta doce e me instigou o apreço por outros instrumentos musicais. Aproveito este espaço para prestar esta singela homenagem e agradecer por toda a instrução e amizade que tivemos nesses dois anos de convivência que influenciaram muito a minha trajetória de vida. Sou grato por tudo, caro amigo!

AGRADECIMENTOS

Agradeço em primeiro lugar aos meus pais Jeanine Porto Teixeira e Maurício Mariano Teixeira pelo permanente apoio e auxílio nos períodos de felicidade e de tristeza, pois sem a ajuda deles nada disso seria possível. Agradeço a minha orientadora Dr^a. Diana Francisca Adamatti pelo empenho e dedicação em seu trabalho, pela disponibilidade e pelo incentivo de todas as minhas ideias, por todo o auxílio com o seu conhecimento, que, mesmo quando além de sua linha de conforto, se empenhou para aprimorar. Agradeço à FAPERGS pelo amparo à pesquisa, à FURG pela possibilidade de estudar nesta importante universidade, ao C3 e todo o seu corpo docente e de servidores. Por último mas não menos importante, agradeço à todos os colegas e amigos que me auxiliaram nessa caminhada em prol do conhecimento, onde trocamos experiências de forma mútua, nessa jornada utópica da evolução.

*If I were not a physicist, I would probably be a musician. I often think in music.
I live my daydreams in music. I see my life in terms of music.*

— ALBERT EINSTEIN

RESUMO

TEIXEIRA, Eduardo Porto. **Ulises: Um Sistema Baseado em Agentes Para Classificação de Timbres**. 2016. 81 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande, Rio Grande.

O campo de estudo da Computação Musical (CM) vem crescendo ao longo dos anos e cada vez existem mais conferências e pesquisadores especializados nesta grande área. A subárea de *Music Information Retrieval* (MIR), uma das principais linhas de pesquisa em CM, tem enfoque em obter informação a partir de dados sonoros. A questão mais essencial no que diz respeito à percepção humana em relação ao som é: quais são as características dos sons de instrumentos musicais para realizar o reconhecimento de suas fontes sonoras? Existem quatro principais dimensões em sons: altura, intensidade, duração e timbre. A quarta dimensão, o timbre, é a mais vaga e complexa das dimensões, uma propriedade multidimensional e de alto nível. O reconhecimento de timbres é uma área de grande interesse dentro de MIR, estando presente em diversos trabalhos encontrados no estado da arte em CM. Tratando-se de Sistemas Multiagente (SMA), o termo autônomo designa o fato de que os agentes têm uma existência própria, independente da existência de outros agentes, e estão aptos a tomarem decisões próprias, sem interferência externa. A tecnologia de agentes se torna particularmente adequada para aplicações musicais, devido à possibilidade de associar um agente computacional com o papel de um cantor ou instrumentista, como pode ser visto em trabalhos atuais na área de CM. Neste contexto, este trabalho propõe uma abordagem baseada em agentes para o reconhecimento de timbres, com enfoque na autonomia dos agentes ao modelo de classificação de timbres. Para isto, atribui-se um método de reconhecimento de timbres a diferentes agentes, onde cada agente é uma entidade especialista em um determinado timbre, característico de um instrumento específico, visando uma solução ao problema de reconhecimento de timbres de forma distribuída.

Palavras-chave: Sistema baseado em agentes, computação musical, recuperação de informação musical, timbre.

ABSTRACT

TEIXEIRA, Eduardo Porto. **Ulises: A Agent-Based System For Timbre Classification**. 2016. 81 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande, Rio Grande.

The Sound and Music Computing (SMC) field has grown over the years and every time there are more conferences and specialized researchers in this area. The sub-field of Music Information Retrieval (MIR), one of the main research fields on SMC has focused on getting information from sound data. The most critical issue with regard to the human perception of sound is: what are the qualities of musical instrument sounds to perform recognition of its sound sources. There are four main sound dimensions: pitch, loudness, duration and timbre. The fourth dimension, timbre, is the most vague and complex dimension, a complex and high-level multidimensional property. Recognition of timbres is an area of high interest within MIR, being present in several papers state of the art on SMC. About Multi-Agent Systems (MAS), the term autonomous refers to the fact that the agents have their own existence, regardless of the existence of other agents, and are able to take own decisions without outside interference. Agents technology is particularly suitable for musical applications because of the possibility of associating a computational agent with the role of a singer or instrumentalist as can be seen in works state of art in SMC area. In this context, this paper proposes a agent-based approach to timbre recognition, focusing on the parallelization of the classification model. For this, we assign a method of recognition of timbres to different agents, where each agent is a specialized entity in a particular timbre, characteristic of a specific instrument, seeking a distributed solution for solving the timbre recognition problem.

Keywords: agent-based system, sound and music computing, music and information retrieval, timbre.

LISTA DE FIGURAS

Figura 1	Características das Ondas [Imagem elaborada pelo autor].	18
Figura 2	Diferentes representações de uma entrada sonora. Acima, um sinal criado por um osciloscópio. Abaixo, a esquerda, seu respectivo espectro no domínio da frequência. Abaixo, a direita, a visualização de um sonograma [Imagem elaborada pelo autor].	22
Figura 3	Tempo de ataque, retirado de (LARTILLOT; TOIVIAINEN; EEROLA, 2014).	24
Figura 4	Inclinação de ataque, retirado de (LARTILLOT; TOIVIAINEN; EEROLA, 2014).	25
Figura 5	Salto de ataque, retirado de (LARTILLOT; TOIVIAINEN; EEROLA, 2014).	25
Figura 6	Troca de sinal ou zeros, retirado de (LARTILLOT; TOIVIAINEN; EEROLA, 2014).	26
Figura 7	Energia de alta frequência, retirado de (LARTILLOT; TOIVIAINEN; EEROLA, 2014).	26
Figura 8	Funcionamento do método MFCC, retirado de (LARTILLOT; TOIVIAINEN; EEROLA, 2014).	28
Figura 9	Fluxograma canônico de um sistema MIR (CASEY et al., 2008).	33
Figura 10	Formato do espectro de uma única nota sendo executada (BELLO et al., 2005).	35
Figura 11	Diagrama de blocos do sistema de reconhecimento de instrumentos implementado por (ERONEN et al., 2001).	37
Figura 12	Intersecção das grandes áreas da Computação Musical e Sistemas Multiagente [Imagem elaborada pelo autor].	40
Figura 13	Interface principal do CInBalada (SAMPAIO; RAMALHO; TEDESCO, 2008).	41
Figura 14	Exemplo de possíveis configurações do SMA proposto em (BOWN, 2009).	43
Figura 15	Fluxograma de funcionamento do sistema Ulises.	47
Figura 16	Ciclo de Vida do agente.	48
Figura 17	Fluxograma de funcionamento do ambiente.	49
Figura 18	Diagrama das ferramentas utilizadas na implementação do sistema Ulises.	51
Figura 19	Componentes da interface gráfica do sistema Ulises.	54
Figura 20	Visualização de Mundo do sistema após o <i>Setup</i>	55

Figura 21	Representação dos estados do agente especialista violão dentro da Visualização de Mundo.	56
Figura 22	a) espectrograma do arquivo “0-1.wav”. b) MFCC do arquivo “0-1.wav”. c) espectrograma do arquivo “0-2.wav”. d) MFCC do arquivo “0-2.wav”. e) espectrograma do arquivo “0-3.wav”. f) MFCC do arquivo “0-3.wav”.	59
Figura 23	Espectrograma e MFCC do arquivo “4-0.wav”.	60
Figura 24	Nas linhas 0-1, 0-2 e 0-3 estão os resultados do MFCC realizado no treinamento do agente violão, a quarta linha apresenta a média entre os resultados do treinamento, a linha 4-0 exhibe os valores extraídos do MFCC do arquivo a ser testado, e na última linha são exibidos os valores do módulo da diferença entre a média do especialista e o arquivo a ser testado.	61
Figura 25	Log do centro de comandos do NetLogo, após exemplo de execução do sistema Ulises.	65

LISTA DE TABELAS

Tabela 1	As Notas Musicais.	20
Tabela 2	Tabela de Notas e Suas Frequências Padrões em Hz.	20
Tabela 3	Exemplo de denominações dos sobretons.	22
Tabela 4	Exemplos de tarefas de MIR e suas especificidades (CASEY et al., 2008).	32
Tabela 5	Características de alto nível (difíceis de extrair) (CASEY et al., 2008).	34
Tabela 6	Definição das características dos timbres (LIU; LU; ZHANG, 2003).	38
Tabela 7	Tabela de referência dos identificadores de instrumentos.	53

LISTA DE ABREVIATURAS E SIGLAS

BDI	<i>Belief Desire Intention</i>
C3	Centro de Ciências Computacionais
CCP	<i>Concurrent Constraint Programming</i>
CM	Computação Musical
DCOP	<i>Distributed Constraint Optimization</i>
DCT	<i>Discrete Cosine Transform</i>
DFT	<i>Discrete Fourier Transform</i>
FAMA	Framework para Aplicações Multiagente
FAPERGS	Fundação de Amparo à Pesquisa do Rio Grande do Sul
FFT	<i>Fast Fourier Transform</i>
FIPA	<i>Foundation for Intelligent Physical Agents</i>
FURG	Universidade Federal do Rio Grande
IFFT	<i>Inverse Fast Fourier Transform</i>
k-NN	<i>k-Nearest Neighbor</i>
MAS	<i>Multi-agent System</i>
MFCC	<i>Mel Frequency Cepstral Coefficients</i>
MIDI	<i>Musical Instrument Digital Interface</i>
MIR	<i>Music Information Retrieval</i>
MPEG	<i>Moving Picture Experts Group</i>
RTCC	<i>Real-Time Concurrent Constraint Calculus</i>
SMA	Sistema Multiagente
SMC	<i>Sound and Music Computing</i>
SPEAR	<i>Sinusoidal Partial Editing Analysis and Resynthesis</i>
STFT	<i>Short-time Fourier Transform</i>

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Motivação	14
1.2	Problema	14
1.3	Justificativa	15
1.4	Objetivos	15
1.5	Organização do Trabalho	16
2	FUNDAMENTOS TEÓRICOS	17
2.1	Acústica	17
2.1.1	Características das Ondas	17
2.1.2	Fase de uma Onda	19
2.1.3	Características dos Sons	19
2.1.4	Harmônicas, Inarmônicas e Sobretom	21
2.1.5	As Representações dos Sons	22
2.1.6	Polifonia	23
2.2	Computação Musical	24
2.2.1	O Domínio Temporal	24
2.2.2	O Domínio da Frequência	26
2.3	Agentes e Sistemas Multiagente	28
2.3.1	Agentes	28
2.3.2	Sistemas Multiagente	29
3	TRABALHOS RELACIONADOS	31
3.1	Trabalhos Relacionados com CM e MIR	31
3.2	Trabalhos Relacionados com SMA	39
4	ULISES: UM SISTEMA BASEADO EM AGENTES PARA CLASSIFICAÇÃO DE TIMBRES	45
4.1	Origem do nome	45
4.2	Metodologia	45
4.2.1	Ciclo de Vida dos Agentes	47
4.2.2	Funcionamento do Ambiente	48
4.3	Ferramentas Utilizadas	49
4.3.1	Banco de Dados	51
4.4	Interface Gráfica	53
4.5	Implementação	56
4.5.1	O Método de Treinamento	58

4.5.2	O Método de Classificação	59
4.6	Exemplo de Execução	61
5	RESULTADOS	66
6	CONCLUSÃO	67
6.1	Objetivos Alcançados	67
6.2	Trabalhos Futuros	67
6.3	Aperfeiçoamentos	68
	REFERÊNCIAS	70
	ANEXO A CÓDIGO-FONTE NETLOGO - ULISES	76
	ANEXO B CÓDIGO-FONTE MATLAB - ULISES	80
	ANEXO C BANCO DE DADOS	81

1 INTRODUÇÃO

O presente trabalho abrange os campos de Sistemas Multiagente (SMA) e Computação Musical (CM), mais precisamente adotando uma abordagem multiagente para a solução de um problema presente no estado da arte em Computação Musical, dentro da subárea de recuperação de informação musical (*MIR - Music Information Retrieval*).

O campo de estudo da Computação Musical pode ser rastreado desde os anos 50, com alguns poucos compositores experimentais, juntamente com alguns engenheiros e cientistas que, independentemente e em diferentes partes do mundo, começaram a explorar o uso das novas tecnologias digitais para aplicações musicais. O termo Computação Musical (CM) faz referência ao termo em inglês *Sound and Music Computing* (SMC), que foi proposto pela primeira vez em meados dos anos 90 (Camurri; Poli; Rocchesso, 1995), e logo passou a ser largamente adotado.

Desde então, a Computação Musical vem crescendo ao longo dos anos e cada vez mais existem conferências e pesquisadores especializados nesta grande área. Entretanto, se comparada a outras áreas do conhecimento, a Computação Musical tem se desenvolvido a passos lento, por causa da grande dificuldade dos sistemas computacionais interpretarem a música em sua forma abstrata e subjetiva. Existem muitos sistemas capazes de realizar rotinas relativamente simples sobre ondas sonoras, e dentre as tarefas mais comuns estão a transcrição musical, acompanhamento sintetizado em tempo real, supressão de ruídos, equalização de sons, e uma vasta gama de softwares de edição de áudio.

Existem várias linhas de pesquisa em Computação Musical, e no Capítulo 2 são apresentados alguns dos principais grupos em que estas se classificam. Neste trabalho, o enfoque está primariamente no primeiro e segundo grupos: “processamento de sinais sonoros e música” e “compressão e modelagem de som e música”, dentro da subárea de MIR. Esta subárea de conhecimento trata de estudar métodos de extração de características do som, como o reconhecimento de *timbres*.

Segundo THOMAZ (2009), devido à possibilidade de associar o papel de um cantor ou instrumentista com um agente computacional, a tecnologia de agentes se torna particularmente adequada para aplicações musicais. Neste trabalho, se utilizou uma abordagem baseada em agentes para tirar proveito de características implícitas dos sistemas basea-

dos em agentes, como a escalabilidade, o processamento distribuído e a possibilidade de cada agente ser uma entidade computacional única, associada a músicos especialistas em determinados timbres.

1.1 Motivação

Comecei meus estudos em música no ano de 2005 na escola de Belas Artes “Heitor de Lemos” - EBAHL, com o curso de violão erudito, e em 2009 ingressei para o curso de Engenharia de Computação da Universidade Federal do Rio Grande - FURG. Não demorei para me apaixonar por estas duas grandes áreas do conhecimento que, quanto mais me aprofundava, mais se apresentavam relacionadas. Em um projeto de incentivo musical da FURG participei de aulas de flauta doce onde conheci o amigo e professor Ulises Dardo Ferretti Martínez, com quem costumava trocar ideias sobre o rumo da música e da computação, dentre elas a proposta inicial desta dissertação. Desta forma, este trabalho surgiu da vontade de conciliar estas duas grandes paixões, posteriormente sendo concluído sob a orientação da professora Diana Francisca Adamatti.

1.2 Problema

Como será apresentado no Capítulo 2 deste trabalho, a questão mais essencial no que diz respeito à percepção humana em relação ao som é: quais são as características dos sons de instrumentos musicais para realizar o reconhecimento de suas fontes sonoras?

Existem quatro principais dimensões em sons: altura, intensidade, duração e timbre. A quarta dimensão, o timbre, é a mais vaga e complexa das dimensões (ERONEN et al., 2001; CASEY et al., 2008). Até mesmo para a percepção humana, o reconhecimento de timbres é uma tarefa difícil, bem como a definição desta característica. O *American National Standards Institute* define timbre de uma maneira puramente excludente:

“...atributo de sensação em que um ouvinte pode julgar que dois sons com a mesma intensidade e altura são diferentes” (KLINGBEIL, 2009).

Em outros termos, o timbre é o que diferencia, por exemplo, o som de uma nota “lá” emitida por um violino do som desta mesma nota emitida por um piano, ambas com a mesma altura e frequência.

São muitos os trabalhos cujo tema é o reconhecimento e identificação de timbres, dentre os quais podem ser citados (HELMHOLTZ; ELLIS, 2009; STRONG, 1963; LUCE; CLARK JR, 1967; BENADE, 2012; NORDQVIST, 2004; KLAPURI, 2004; KITAHARA, 2007). Esta característica abstrata é de grande interesse no campo de MIR, e considerada um problema pertencente ao estado da arte (CASEY et al., 2008). O aprimoramento do reconhecimento de timbres, seja em relação à precisão ou à eficiência do resultado, é de grande valor científico. O foco principal deste trabalho é propor uma

abordagem multiagente para o problema do reconhecimento de timbres.

1.3 Justificativa

A tecnologia de Sistemas Multiagente é uma forma promissora para a performance musical interativa, em outros termos, execução colaborativa, onde agentes instrumentistas se unem na produção de uma performance musical, como visto em (SAMPAIO; TEDESCO; RAMALHO, 2005; SAMPAIO; RAMALHO; TEDESCO, 2008).

Em trabalhos recentes, essa tecnologia foi adaptada para resolver problemas específicos em um escopo musical limitado, como detecção de pulso, simulação de instrumentos ou acompanhamento automático. Sendo uma área bastante consolidada, adequada para solucionar problemas que exijam distribuição, seja de natureza lógica ou geográfica, e em que a complexidade do problema seja minimizada por esta abordagem, Sistemas Multiagente são úteis em várias subáreas da Computação Musical.

Segundo THOMAZ (2009), a tecnologia de agentes se torna particularmente adequada para aplicações musicais, devido à possibilidade de associar um agente computacional com o papel de um cantor ou instrumentista. Ele destaca algumas vantagens dessas associações, como mapear características como desempenho, percepção, adaptação e improvisação de um lado, e processos artificiais no outro. Além disso, é possível definir formas de inter-relação social entre os agentes, que traz esta tecnologia ainda mais perto de performance musical colaborativa.

Neste contexto, este trabalho explora as características de SMA para a solução do problema de reconhecimento de timbres, visando uma melhor qualidade nos resultados e iniciando os estudos em paralelização e distribuição dos métodos tradicionais de reconhecimento de timbres.

1.4 Objetivos

O objetivo principal deste trabalho é a elaboração de um sistema baseado em agentes para a solução do problema de reconhecimento de timbres, dentro da área de MIR.

Para execução do objetivo principal proposto, é necessária a execução dos seguintes objetivos específicos:

- Apresentar uma proposta de paralelização da extração de características ou do método de classificação para o reconhecimento de timbre, a fim de distribuir o processamento entre os agentes do sistema;
- Implementar uma arquitetura baseada em agentes classificadora de timbres;

1.5 Organização do Trabalho

Este trabalho está dividido em 6 capítulos. O Capítulo 2 apresenta a fundamentação teórica necessária para o desenvolvimento do mesmo, seja na área de sistemas multiagente ou computação musical. O Capítulo 3 descreve alguns dos principais trabalhos nas áreas de computação musical e sistemas multiagente que tenham temas relacionados. O Capítulo 4 apresenta o sistema desenvolvido, abordando a metodologia empregada, as ferramentas utilizadas, aspectos de implementação além de explicar as funcionalidades do sistema. O Capítulo 5 apresenta os resultados com o desenvolvimento do sistema. E por último, o Capítulo 6 descreve os objetivos alcançados e discute possibilidades de trabalhos futuros.

2 FUNDAMENTOS TEÓRICOS

Esse capítulo apresenta os principais conceitos sobre Computação Musical e Sistemas Multiagente, que são as duas principais áreas de conhecimento envolvidas nesse trabalho.

2.1 Acústica

O conceito mais básico por trás de qualquer área de estudo musical é a definição de som. Ele é produzido quando um objeto (a fonte sonora) vibra e faz o ar ao seu redor se mover (RUMSEY; MCCORMICK, 2012). Este efeito pode ser representado como uma esfera que pulsa regularmente, com centro na fonte sonora, e seu tamanho oscila levemente entre maior e menor que o normal. Assim que pulsa, o som irá comprimir e rarefazer o ar ao seu redor, resultando em uma série de compressões e rarefações viajando para longe da esfera, similar a uma versão tridimensional de uma pedra que cai sobre um lago. Se a pressão varia de acordo com um padrão repetitivo, diz-se que o som tem uma forma de onda periódica. Se não houver nenhum padrão discernível, é chamado de ruído. Entre esses dois extremos existe um vasto domínio de sons quase periódicos e quase ruidosos (ROADS, 1996).

Segundo LAPP (2003), as ondas podem ser classificadas em dois tipos básicos, dependendo de seu tipo de vibração. Imaginando uma mola presa ao chão e sacudindo-a para trás e para frente, tem-se o que é chamado de onda transversal. Uma onda transversal é aquela em que a forma vibra perpendicularmente à direção dos movimentos de energia. Se, em vez disso, a mola for empurrada para a frente e puxada para trás, tem-se uma onda de compressão. Ondas de compressão também são conhecidas como ondas longitudinais. Uma onda de compressão é aquela em que a forma vibra na mesma direção que o movimento da energia na onda, como é o caso das ondas sonoras.

2.1.1 Características das Ondas

Como representa a Figura 1, as ondas possuem algumas características básicas, que seguem leis físicas simples, são elas:

- **Crista:** o ponto mais alto da onda média ou maior distância positiva da posição de

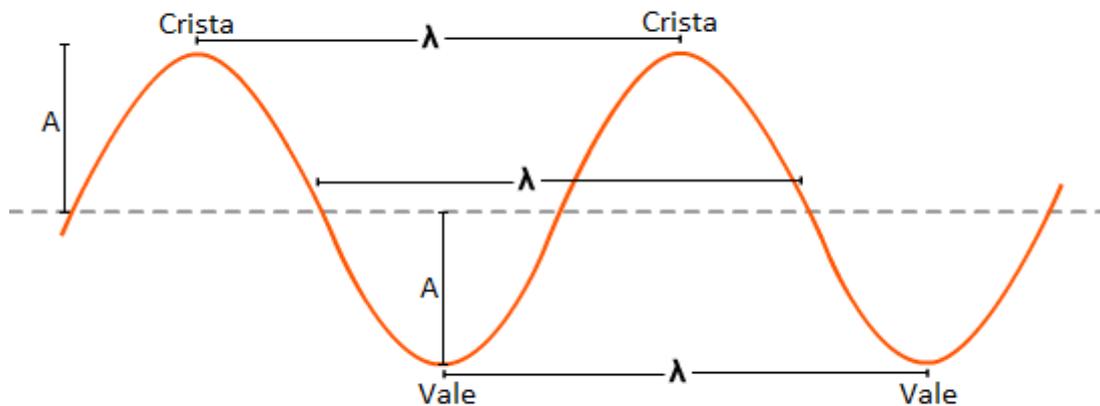


Figura 1: Características das Ondas [Imagem elaborada pelo autor].

repouso;

- Vale: o ponto mais baixo da onda média ou a maior distância negativa a partir da posição de repouso;
- Comprimento (λ): representa a distância entre duas cristas ou dois vales (ou dois pontos consecutivos) de uma onda que vibra em fase;
- Amplitude (A): a distância a partir da posição de repouso da crista ou do vale. A amplitude está relacionada com a energia da onda, quando a energia cresce, a amplitude também cresce. A energia é, na verdade, proporcional ao quadrado da amplitude. Uma onda com amplitude duas vezes maior tem, na verdade, quatro vezes mais energia, e uma com amplitude três vezes maior tem nove vezes mais energia;
- Frequência (f): é o número de comprimentos de onda que passam por um ponto, em um segundo. A frequência também é referida como “ondas por segundo”, “pulsos por segundo”, ou “ciclos por segundo”. A unidade do sistema internacional para a frequência é o Hertz (Hz). A altura do som é o mesmo que a frequência de uma onda sonora (LAPP, 2003). Sem perdas auditivas, o ouvido pode detectar frequências de onda entre 20Hz e 20.000Hz¹. No entanto, muitos idosos e soldados com experiência de combate ao vivo perdem a capacidade de ouvir frequências mais altas. Entretanto, a maior parte das conversas ocorre bem abaixo de 2.000Hz, devido a limitações físicas das cordas vocais;

¹Sons abaixo de 20Hz são classificadas como subsônicos; e sons com mais de 20.000Hz são denominados ultra-sons.

- Período (T): é o tempo necessário para o comprimento de uma onda inteira passar por um certo ponto. Seu valor é o inverso da frequência;
- Velocidade (v): é a razão do comprimento de onda pelo período, ou seja, $v = \lambda / T$.

2.1.2 Fase de uma Onda

O ponto de uma onda periódica no eixo dos “y”, ou o eixo da amplitude de partida é a sua fase inicial. Por exemplo, uma senoide típica começa no ponto de amplitude 0 e completa o seu ciclo em 0. Se deslocar o ponto de partida sobre o eixo horizontal em 2, ou seja, 90 graus, em seguida, a onda sinusoidal começa e termina no eixo de amplitude com o valor igual a 1.

Por convenção, esta onda é chamada de cossenoide. De fato, uma cossenoide é equivalente a uma senoide, que é deslocada de fase em 90 graus. Quando dois sinais começam no mesmo ponto, se diz que estão em fase ou alinhados. Isto contrasta com um sinal que é ligeiramente atrasado em relação a um outro sinal, no qual os dois sinais estão fora de fase (ROADS, 1996).

2.1.3 Características dos Sons

As ondas sonoras são ondas de compressão causadas por vibrações, mas a música de uma sinfonia varia consideravelmente do choro de um bebê ou do sussurro de um confidente.

Todas as ondas sonoras podem ser caracterizadas pela sua *velocidade*, por sua *altura*, pela sua *intensidade*, e pela sua qualidade sonora ou *timbre* (LAPP, 2003).

2.1.3.1 Altura

A altura é uma característica do som que faz referência a nossa percepção de agudos e graves. Fisicamente, sons agudos possuem maiores frequências, e sons graves menores. O ser humano é capaz de ouvir em um intervalo entre 20Hz e 20.000Hz. Baleias e golfinhos escutam frequências ainda maiores (LAPP, 2003).

Um conceito importante na música é a oitava - uma duplicação da frequência. Por exemplo, 40Hz é uma oitava acima de 20Hz. A audição humana é sensível ao longo de um intervalo de cerca de 10 oitavas (20Hz, 40Hz, 80Hz, 160Hz, 320Hz, 640Hz, 1280Hz, 2560Hz, 5120Hz, 10240Hz e 20480Hz) (LAPP, 2003).

Tratando-se de musicologia, a altura é representada por “notas”. Existe, na música moderna, sete notas musicais, como visto na Tabela 1. Elas se repetem ao longo da escala musical, podendo representar sons de diferentes frequências, dependendo da oitava em que se situam. O “A4”, por exemplo, é a nota Lá da quarta oitava, possuindo uma frequência de 440Hz, já o “A2” possui 110Hz.

Na Tabela 2, é apresentada a relação das notas musicais e seus “acidentes”

Tabela 1: As Notas Musicais.

C	D	E	F	G	A	B	C
Dó	Ré	Mi	Fá	Sol	Lá	Si	Dó

(frequências intermediárias entre as notas), com suas respectivas frequências em Hertz, presentes nas seis primeiras oitavas.

Tabela 2: Tabela de Notas e Suas Frequências Padrões em Hz.

Nota	1ª Oitava	2ª Oitava	3ª Oitava	4ª Oitava	5ª Oitava	6ª Oitava
A	55.00	110.00	220.00	440.00	880.00	1760.00
A \sharp /B \flat	58.27	116.54	233.08	466.16	932.33	1864.66
B/C \flat	61.74	123.47	246.94	493.88	987.77	1975.53
B \sharp /C	65.41	130.81	261.63	523.25	1046.50	2093.00
C \sharp /D \flat	69.30	138.59	277.18	554.37	1108.73	2217.46
D	73.42	146.83	293.66	587.33	1174.66	2349.32
D \sharp /E \flat	77.78	155.56	311.13	622.25	1244.51	2489.02
E	82.41	164.81	329.63	659.26	1318.51	2637.02
E \sharp /F	87.31	174.61	349.23	698.46	1396.91	2793.83
F \sharp /G \flat	92.50	185.00	369.99	739.99	1479.98	2959.96
G	98.00	196.00	392.00	783.99	1567.99	3135.96
G \sharp /A \flat	103.83	207.65	415.30	830.61	1661.22	3322.44

2.1.3.2 Intensidade

A intensidade do som está relacionada com a amplitude da onda de som. A maioria das pessoas tem algum reconhecimento da escala de decibéis (dB). Eles podem ser capazes de dizer que 0dB é o limiar de audição e de que o som na pista ao lado de um jato acelerando é de cerca de 140dB.

No entanto, a maioria das pessoas não percebe que a escala de decibéis é uma escala logarítmica. Isto significa que para cada aumento de 10dB, o aumento de intensidade de som é um fator de dez. Assim, indo de 60dB a 70dB é um aumento de dez vezes, e de 60dB a 80dB é um aumento de cem vezes. Desta forma pode-se ouvir intensidades sonoras com mais de 14 ordens de magnitude. Assim sendo, o jato de 140dB na pista tem um volume 10¹⁴ vezes maior do que o limiar, o que representa 100 trilhões. Os nossos ouvidos conseguem medir a intensidade com uma abrangência fantástica (LAPP, 2003).

2.1.3.3 Duração

A duração diz respeito ao tempo da onda sonora, o seu período, e o tempo que o som leva até cessar. É uma importante característica ao se estudar aspectos rítmicos dos sons.

2.1.3.4 *Timbre*

O timbre é a mais vaga e complexa das quatro dimensões dos sons (ERONEN et al., 2001). Considerável energia e esforços foram aplicados para promover o entendimento do timbre, uma das características mais abstratas da música.

É uma propriedade complexa e multidimensional (KLINGBEIL, 2009). Um trompete e um violino podem executar exatamente a mesma nota, com a mesma altura e intensidade, e se seus olhos estiverem fechados, a pessoa pode facilmente distinguir entre os dois. A diferença entre os sons está relacionada com a sua qualidade sonora ou timbre. A existência de frequências complementares, combinado com as frequências básicas, não alteram o tom básico, mas dá um “sabor” especial para o som que está sendo produzido. O timbre é a característica que dá a identidade para o som, e depende da fonte sonora que o produz (LAPP, 2003). Maiores informações sobre o timbre serão abordadas na seção 3.1.

2.1.4 **Harmônicas, Inarmônicas e Sobretom**

Além da frequência fundamental de um som, que representa o seu “tom”, ou a sua altura, podem haver muitas frequências presentes em uma forma de onda.

Uma representação no domínio da frequência, ou espectrograma, mostra as frequências principais contidas em um som. Os componentes individuais de frequência do espectro podem ser referidos como harmônicas e parciais. Frequências harmônicas são múltiplos inteiros da frequência fundamental (ROADS, 1996), e podem ser facilmente identificadas em um analisador de espectro. De modo mais geral, qualquer componente de frequência pode ser chamado de parcial. Se a parcial for um múltiplo inteiro de uma frequência fundamental, neste caso, pode ser denominada harmônica. Em outros termos, se a frequência fundamental de determinado sinal sonoro é 400Hz, uma frequência de 800Hz é uma harmônica, e uma frequência de 450Hz é uma parcial. Uma parcial que não é harmônica, é também usualmente chamada de parcial inarmônica, ou simplesmente inarmônica.

Um sobretom é qualquer frequência maior do que a frequência fundamental de um som. Usando o modelo de análise de Fourier, o fundamental e os sobretons juntos são chamados de parciais. Um sobretom é uma parcial (uma “onda parcial” ou “componente de frequência”) que pode ser uma harmônica parcial (uma harmônica), com exceção da fundamental, ou uma inarmônica parcial.

A frequência harmônica é um múltiplo inteiro da frequência fundamental, e frequência inarmônica é um múltiplo não inteiro de uma frequência fundamental.

Um exemplo de sobretons harmônicos (em uma situação ideal em que todas as parciais são harmônicas) é retratado na Tabela 3, com as suas respectivas denominações.

Tabela 3: Exemplo de denominações dos sobretons.

Frequência	Ordem	Nome 1	Nome 2	Nome 3
$1 \cdot f = 440 \text{ Hz}$	$n = 1$	tom fundamental	1ª harmônica	1ª parcial
$2 \cdot f = 880 \text{ Hz}$	$n = 2$	2º sobreton	2ª harmônica	2ª parcial
$3 \cdot f = 1320 \text{ Hz}$	$n = 3$	3º sobreton	3ª harmônica	3ª parcial
$4 \cdot f = 1760 \text{ Hz}$	$n = 4$	4º sobreton	4ª harmônica	4ª parcial

2.1.5 As Representações dos Sons

Existem diversas formas de representar o som visualmente, por este se tratar de uma onda longitudinal tridimensional. Entretanto, três formas são mais comuns: a primeira é particularmente útil para apresentar a forma de onda do som; a segunda, o espectro de frequência do som; e a terceira, a variação do espectro de frequência ao decorrer do tempo. Em outras palavras, a primeira representação trata do sinal no domínio do tempo, a segunda trata do domínio da frequência, e a terceira faz a representação mista dos dois domínios, como pode ser observado na Figura 2. Respectivamente, os equipamentos que geram estas representações são: o osciloscópio, o analisador de espectro e o sonograma.

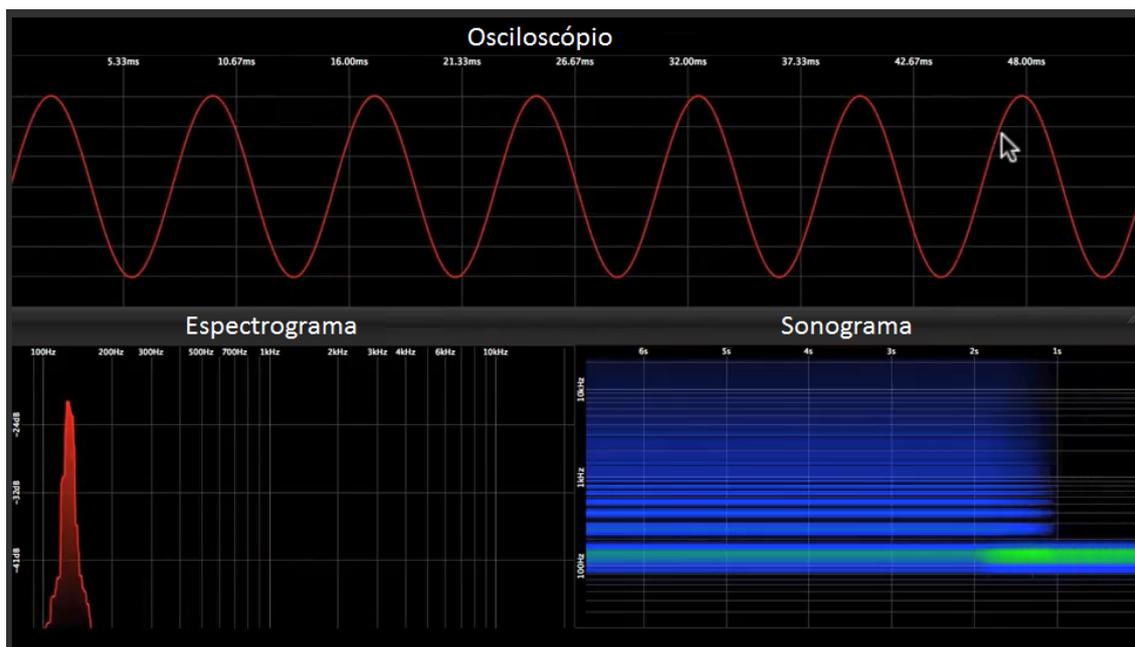


Figura 2: Diferentes representações de uma entrada sonora. Acima, um sinal criado por um osciloscópio. Abaixo, a esquerda, seu respectivo espectro no domínio da frequência. Abaixo, a direita, a visualização de um sonograma [Imagem elaborada pelo autor].

2.1.5.1 Osciloscópio

Um osciloscópio é usado para indicar a forma de onda de um som.

Ele aceita sinais sonoros em forma elétrica e exibe suas análises em tela. O osciloscópio exibe um ponto em movimento que varre horizontalmente um número de velo-

idades fixas da esquerda para a direita e cuja deflexão vertical é controlada pela tensão do sinal de som (positivo para cima, negativo para baixo). Deste modo, representa graficamente a forma de onda do som, uma vez que varia com o tempo. Muitos osciloscópios tem duas entradas e podem traçar duas formas de onda ao mesmo tempo. Isto pode ser particularmente útil para a comparação das fases relativas de dois sinais (RUMSEY; MCCORMICK, 2012).

2.1.5.2 *Analizador de Espectro*

O analisador de espectro funciona de diferentes formas, dependendo do método de análise de espectro.

Um analisador em tempo real mostra um espectro de linha em constante atualização, e mostra os componentes do sinal de entrada na escala horizontal, juntamente com as suas amplitudes na escala vertical de frequência (RUMSEY; MCCORMICK, 2012). Em um analisador de espectro é possível observar as harmônicas (ou parciais), que são características da fonte sonora utilizadas no reconhecimento de timbre.

No domínio da frequência é possível recuperar informação musical com os mais diversos objetivos.

2.1.5.3 *Sonograma*

Em um sonograma é representado um espectrograma que varia com o tempo.

Este tipo de representação é útil no reconhecimento de fala entre outras aplicações. Auxilia a visualizar as alterações nas frequências harmônicas em sons sendo executados ao longo do tempo. Ele representa três dimensões: a frequência (em Hertz, na vertical), o tempo (em segundos, na horizontal) e a intensidade (em decibéis, representado com diferentes colorações).

2.1.6 **Polifonia**

A polifonia, de uma forma geral, ocorre quando duas ou mais notas soam simultaneamente. Em outras palavras, quando duas ou mais notas musicais são executadas no mesmo intervalo de tempo tem-se a polifonia, também denominada de “acorde”. Assim sendo, acordes são, basicamente, conjuntos de notas executadas simultaneamente.

Dentro da classificação de timbres, que é o objetivo deste trabalho, o entendimento de sons sobrepostos é importante para compreender o funcionamento da classificação. Arquivos de áudio, em sua maioria, não distinguem quantas fontes sonoras estão produzindo sons, nem quais são estas fontes. Determinar quais são estas fontes é realizar a classificação de timbres, e esta classificação é realizada sobre sons monofônicos (apenas uma fonte sonora produz um único som no determinado instante de tempo). Para classificar sons polifônicos, o método mais comum consiste em separar os sons produzidos simultaneamente e logo após aplicar um método de classificação de sons monofônicos

para cada uma destas saídas (CASEY et al., 2008). É fácil de compreender o motivo da classificação de timbres funcionar apenas para sons monofônicos se pensarmos em um exemplo prático: imaginemos um piano e uma flauta executando ao mesmo tempo uma nota musical, seja ela a mesma nos dois instrumentos ou não. O resultado da classificação não deve ser um instrumento flauta-piano, e sim dois resultados, um instrumento flauta e um instrumento piano.

2.2 Computação Musical

2.2.1 O Domínio Temporal

Como visto anteriormente, o som pode ser analisado em diferentes domínios, entre os mais comuns estão o domínio temporal e o domínio da frequência. Se tratando de extração de características, os métodos de extração podem ser classificados nas principais dimensões musicais: dinâmica, rítmica, timbre, altura e tonalidade (LARTILLOT; TOIVAINEN; EEROLA, 2014). A seguir, serão descritas as características e os principais métodos de extração de características utilizados na classificação de timbres dentro do domínio temporal.

2.2.1.1 Tempo de ataque (*attack time*)

Uma forma simples de descrever a fase de ataque consiste em estimar sua duração. Este intervalo, em segundos, é calculado entre o ponto em que a envoltória começa a crescer (denominado “*onset*”²) até seu pico como é mostrado na Figura 3.

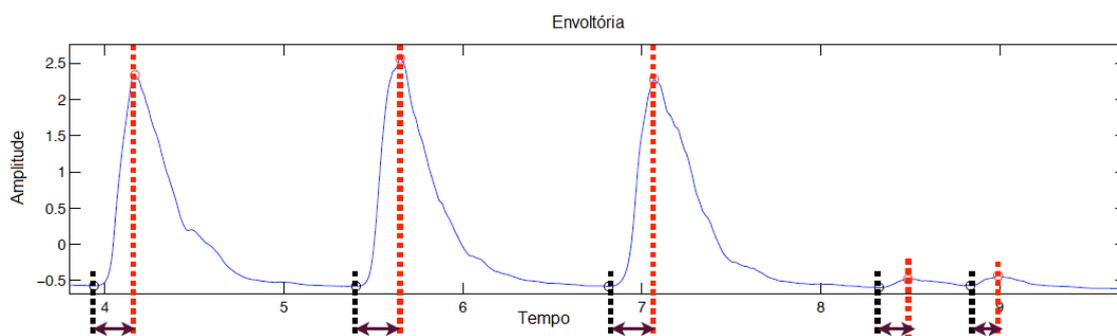


Figura 3: Tempo de ataque, retirado de (LARTILLOT; TOIVAINEN; EEROLA, 2014).

2.2.1.2 Inclinação de ataque (*attack slope*)

Outra descrição da fase de ataque está relacionada à sua inclinação média. Os valores são expressos na mesma escala do sinal original, mas normalizados em relação ao tempo (expresso em segundos). Algumas implementações calculam a inclinação como

²Onsets são pontos em que a envoltória sonora começa a crescer, sendo o ponto de início da fase de ataque.

uma razão entre a diferença de magnitude no início e no fim do período de ataque, e a diferença de tempo correspondente, como pode ser observado na Figura 4.

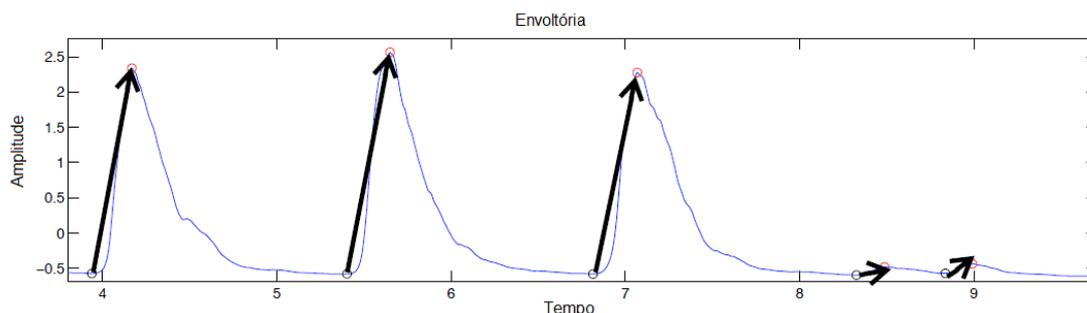


Figura 4: Inclinação de ataque, retirado de (LARTILLOT; TOIVIAINEN; EEROLA, 2014).

2.2.1.3 Salto de ataque (*attack leap*)

Outra maneira simples de descrever a fase de ataque, consiste em estimar a diferença de amplitude entre o início e o final da fase de ataque, como visto na Figura 5. Os valores são expressos na mesma escala que o sinal original.

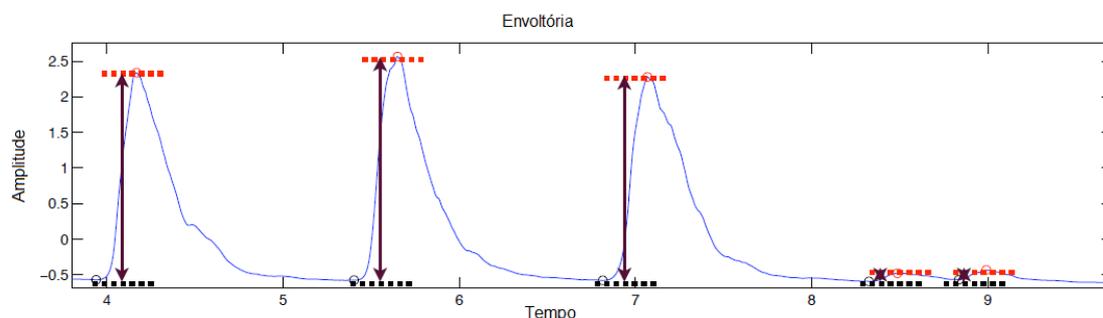


Figura 5: Salto de ataque, retirado de (LARTILLOT; TOIVIAINEN; EEROLA, 2014).

2.2.1.4 Decaimento (*decrease slope*)

Similar a inclinação de ataque, o decaimento é uma forma simples de descrever a fase de relaxamento (em inglês utiliza-se o termo *release*). Os valores são expressos na mesma escala do sinal original, mas normalizados em relação ao tempo (expresso em segundos).

2.2.1.5 Duração (*duration*)

A duração de uma nota também pode ser usada como característica na classificação de timbres. Ela consiste no intervalo de tempo em segundos entre o início da fase de ataque e o final da fase de relaxamento.

2.2.1.6 Zeros (zero cross)

Um indicador de ruidez (escalar que representa a quantidade de ruído) de um sinal pode ser calculado contando o número de vezes que o sinal atravessa o eixo horizontal, ou seja, troca de sinal. Também é chamado de “zeros” pois se considerar o sinal uma função da amplitude em relação ao tempo, as raízes da função seriam os pontos em que ocorre a troca de sinal, como pode ser observado na Figura 6.

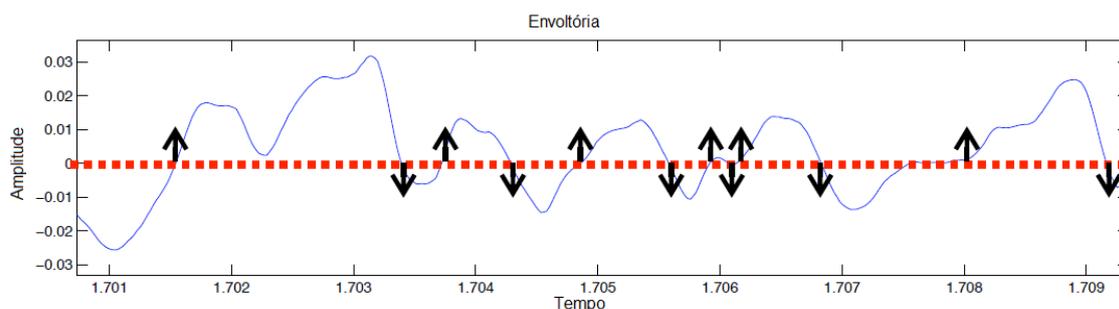


Figura 6: Troca de sinal ou zeros, retirado de (LARTILLOT; TOIVIAINEN; EEROLA, 2014).

2.2.2 O Domínio da Frequência

A seguir serão descritas as características e os principais métodos de extração de características utilizados na classificação de timbres dentro do domínio da frequência.

2.2.2.1 Energia de alta frequência (high-frequency energy)

Uma maneira de estimar a quantidade de alta frequência no sinal consiste em encontrar uma frequência de tal forma que uma certa fração do total de energia esteja contida abaixo dela. Esta razão é fixada por padrão em 0,85 segundo (TZANETAKIS; COOK, 2002). Outros autores propuseram o valor de 0,95 (POHLE; PAMPALK; WIDMER, 2005). Um exemplo é apresentado na Figura 7, adotando a razão de 0,85 proposta por (TZANETAKIS; COOK, 2002).

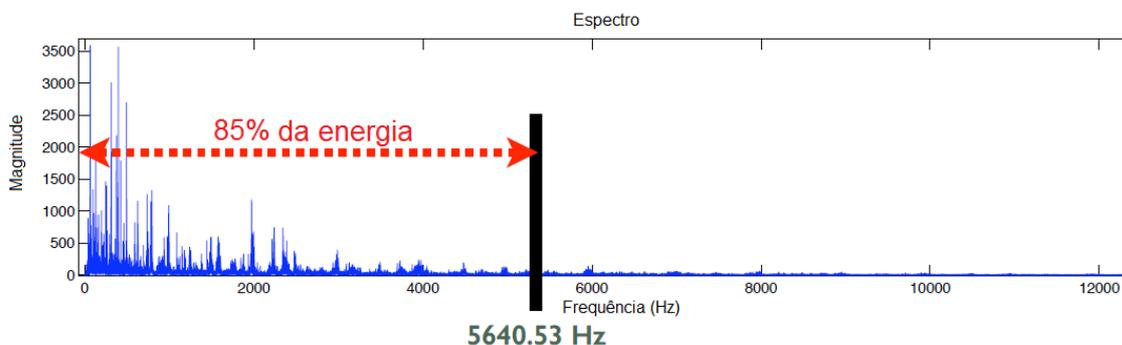


Figura 7: Energia de alta frequência, retirado de (LARTILLOT; TOIVIAINEN; EEROLA, 2014).

2.2.2.2 *Descritores estatísticos (statistical description of spectral distribution)*

A distribuição espectral pode ser descrita também por cálculos estatísticos: centroide, propagação, assimetria, curtose, achatamento, bem como pela entropia. São métodos utilizados em diversas áreas do conhecimento, úteis para definir características de distribuição e da forma dos sinais, estando presentes também em trabalhos de reconhecimento de timbres.

2.2.2.3 *MFCC (Mel-Frequency Cepstral Coefficients)*

O método de MFCC (DAVIS; MERMELSTEIN, 1980) se tornou uma das técnicas mais populares na extração de características em sistemas automáticos de reconhecimento de fala (ERONEN et al., 2001), pois oferece uma descrição da forma espectral do som bastante eficiente, posteriormente sendo amplamente utilizada no reconhecimento de timbres. Um *cepstro*³ é o resultado de aplicar a transformada inversa de Fourier (IFFT) ao logaritmo do espectro estimado de um sinal. O MFCC se baseia no procedimento de cepstro. Nele, as bandas de frequência estão posicionadas de forma logarítmica utilizando a escala de Mel⁴. A transformada de Fourier é substituída por uma transformada de cosseno discreta (DCT)⁵. Ela tem uma propriedade de “compactação de energia” eficiente: a maior parte da informação do sinal tende a se concentrar em alguns componentes do DCT de baixa frequência. É por isso que, por padrão, apenas os 13 primeiros componentes são devolvidos. Por convenção, o coeficiente zero indica simplesmente a energia média do sinal (LARTILLOT; TOIVIAINEN; EEROLA, 2014). O fluxo de execução do método de MFCC é demonstrado na Figura 8, onde o sinal de áudio passa por uma transformada de Fourier, logo após o resultado da transformada é selecionado com seus valores absolutos para uma função logarítmica, os valores são mapeados para a escala Mel e por último é realizada a transformada de cosseno discreta. Como resultado, o MFCC retorna as amplitudes do espectro resultante.

2.2.2.4 *Dissonância sensorial (sensory dissonance)*

Os autores (PLOMP; LEVELT, 1965) propuseram uma estimativa da dissonância sensorial, ou aspereza, relacionada ao fenômeno de ressonância sempre que pares de senóides estão próximos no domínio da frequência. Os autores propõem como resultado uma estimativa de aspereza, dependendo da proporção entre cada par de senóides. Esta estimativa também passou a ser utilizada em alguns trabalhos de classificação de timbres.

³A denominação “cepstro” vem da palavra “spectro” com as primeiras letras em ordem inversa, fazendo referência ao uso da transformada inversa de Fourier.

⁴A escala de Mel, nomeada por (STEVENS; VOLKMANN; NEWMAN, 1937), é uma escala de percepção de alturas julgada por ouvintes a ter espaçamentos iguais entre os sons. Esta escala de frequência se aproxima mais a resposta do sistema auditivo humano do que as faixas de frequências linearmente espaçadas.

⁵Uma transformada de cosseno discreta (DCT) é uma transformada de Fourier semelhante à transformada de Fourier discreta (DFT), mas utilizando apenas números reais.

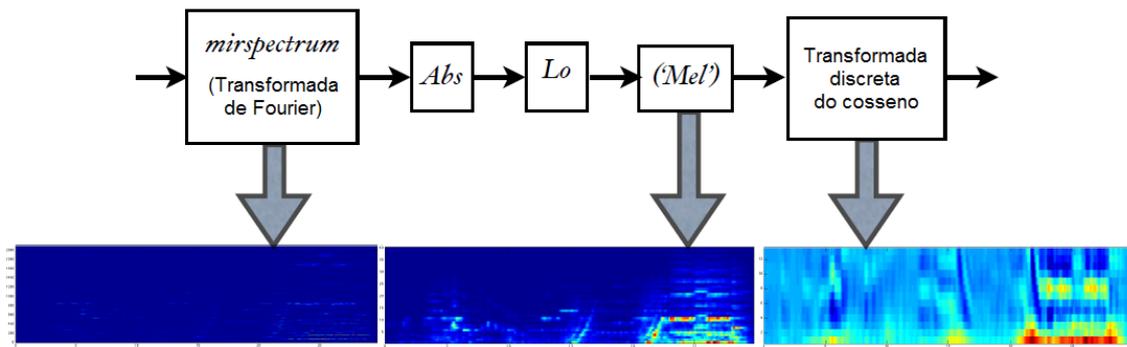


Figura 8: Funcionamento do método MFCC, retirado de (LARTILLOT; TOIVIAINEN; EEROLA, 2014).

2.2.2.5 Variabilidade de cristas espectrais (*Spectral peaks variability*)

A irregularidade de um espectro é o grau de variação dos picos sucessivos do espectro. Existem duas formas mais conhecidas para o cálculo da irregularidade: uma proposta por (JENSEN, 1999), vista na Equação 1, em que a irregularidade é o somatório do quadrado da diferença das amplitudes entre parciais adjacentes; e a outra proposta por (KRIMPHOFF; MCADAMS; WINSBERG, 1994), vista na Equação 2, em que a irregularidade é o somatório da amplitude atual menos a média da amplitude anterior com a atual e a próxima.

$$\left(\sum_{k=1}^N (a_k - a_{k+1})^2 \right) / \left(\sum_{k=1}^N a_k^2 \right) \quad (1)$$

$$\sum_{k=2}^{N-1} \left(a_k - \frac{a_{k-1} + a_k + a_{k+1}}{3} \right) \quad (2)$$

2.3 Agentes e Sistemas Multiagente

2.3.1 Agentes

Um agente é uma entidade computacional com um comportamento autônomo que lhe permite decidir suas próprias ações. A decisão de qual ação executar é determinada pelo agente, tendo em consideração as mudanças acontecidas no ambiente em que atua e o desejo de alcançar seus objetivos.

Usualmente, cada agente possui um conjunto de capacidades comportamentais que definem sua competência, um conjunto de objetivos, e a autonomia necessária para utilizar suas capacidades comportamentais a fim de alcançar seus objetivos (WEISS, 1999).

Na literatura, os agentes podem ser divididos de duas formas (WOOLDRIDGE, 2009):

- Reativos: podem desenvolver inteligência a partir de interações com seu ambiente, não necessitam de modelo preestabelecido. Por sua vez, estes podem ser classificados em agentes simples (respondem a percepções, interpretam a entrada, verificam regras correspondentes e agem), ou baseados em modelo (mantêm o estado interno para aspectos não percebidos, seus estados internos modelam o mundo, podem utilizar percepções passadas para embasar decisões futuras).
- Cognitivos: complexos, com mecanismos de tomada de decisões avançados, interações sofisticadas e com um objetivo fortemente estabelecido. Podem ser classificados em baseados em objetivos (procuram atingir objetivos e são mais flexíveis que os reativos), ou baseados na utilidade (tentam maximizar suas expectativas, ponderam probabilidades de sucesso em relação à importância de objetivos).

2.3.2 Sistemas Multiagente

Os Sistemas Multiagente (SMA) *Multi-Agent Systems* (MAS), formam uma subárea da Inteligência Artificial Distribuída e concentram-se no estudo de agentes em um universo multiagente.

Tratando-se de SMA, o termo autônomo designa o fato de que os agentes têm uma existência própria, independente da existência de outros agentes, e estão aptos a tomarem decisões próprias, sem interferência externa (WOOLDRIDGE, 2009).

A ideia principal em um sistema multiagente é que um comportamento global inteligente pode ser alcançado a partir da integração dos comportamentos individuais dos agentes. Em um SMA, não é necessário que cada agente seja individualmente inteligente para alcançar um comportamento global inteligente (ALVARES; SICHMAN, 1997).

O desenvolvimento de sistemas multiagente carrega consigo aspectos de complexidade relacionados à avaliação dos diferentes níveis e componentes desse sistema.

Um sistema multiagente pode ser composto por uma série de agentes heterogêneos, que apresentam variabilidade quanto à sua arquitetura interna, modelos utilizados para o seu desenvolvimento, linguagem de programação, de especificação e validação. Adicionalmente, contextos específicos de cada solução para com o ambiente para o qual foi projetado. Deste modo, mecanismos que permitam estabelecer métricas de avaliação para cada nível do desenvolvimento de um sistema multiagente, considerando dimensões como organização, comunicação entre agentes e os agentes em si são necessários (WOOLDRIDGE, 2009).

Os agentes, em um SMA, possuem algumas características importantes:

- Autonomia: os agentes são, ao menos, parcialmente independentes, autoconscientes, autônomos;
- Visão local: nenhum agente tem uma visão global completa do sistema, ou o sistema é muito complexo para um agente fazer uso prático deste conhecimento;

- Descentralização: não há designação de um agente controlador.

Os SMA podem ser auto-organizacionais, e os agentes podem assumir o controle do sistema, resultando em comportamentos complexos, mesmo se todos os agentes forem simples. Quando agentes podem compartilhar conhecimento usando alguma linguagem, dentro dos protocolos de comunicação do sistema, a abordagem pode levar a uma melhoria comum (DEVI; SRINIVAS; KRISHNA, 2012).

Muitas pesquisas na área de Sistemas Multiagente tem o foco no desenvolvimento e análise de sofisticadas resoluções de problemas e de controle em arquiteturas de inteligência artificial. Alguns exemplos de pesquisa em SMA incluem:

- Engenharia de software orientada a agentes;
- Crenças, desejos e intenções (arquitetura BDI);
- Cooperação e coordenação;
- Otimização de restrições distribuídas (DCOPs);
- Organização de agentes;
- Comunicação de agentes;
- Negociação entre agentes;
- Solução de problemas distribuídos;
- Aprendizagem multiagente;
- Resolução de problemas de comunidades científicas;
- Confiabilidade e tolerância a falhas;
- Robótica, sistemas multi-robóticos, clusters robóticos.

Desta forma, a utilização de um SMA se torna evidentemente útil para a solução de problemas paralelizáveis, em que cada unidade funcional é responsável por parte da solução.

Esse trabalho faz uso de agentes cognitivos, capazes de tomar uma decisão própria (no caso classificar um conjunto de características extraídas de um arquivo de áudio), e em conjunto com outros agentes e o ambiente solucionar um problema de grande complexidade, o reconhecimento de timbres.

3 TRABALHOS RELACIONADOS

Esse capítulo apresenta trabalhos relacionados a Computação Musical e a Sistemas Multiagente, buscando a inter-relação entre as duas áreas.

3.1 Trabalhos Relacionados com CM e MIR

O atual campo de pesquisa em Computação Musical pode ser agrupado em uma série de subcampos que se concentram em aspectos específicos da cadeia de comunicação de som e música (WIDMER et al., 2007), dentre eles:

- Processamento de sinais sonoros e música: Este subcampo enfoca técnicas de processamento de sinais de áudio para a análise, transformação e ressíntese de sons e sinais musicais;
- Compreensão e modelagem de som e música: Este subcampo enfoca na compreensão e modelagem de som e música usando abordagens computacionais. Aqui pode-se incluir musicologia computacional, recuperação de informação musical, e as abordagens mais computacionais da cognição musical;
- Interfaces para som e música: Este subcampo centra-se na concepção e implementação de interfaces de computador para som e música. Isso está basicamente relacionado com a interação homem-máquina;
- Criação assistida de som e música: Este subcampo se concentra no desenvolvimento de ferramentas computacionais para auxiliar projeto de som e composição musical. Aqui pode-se incluir campos tradicionais, como a composição algorítmica.

Dentre estes subcampos, é possível enumerar uma vasta gama de aplicações, como: instrumentos musicais digitais, produção musical, recuperação de informação musical, bibliotecas de sons e músicas digitais, sistemas multimídia interativos, interfaces auditivas, ação e percepção aumentadas.

A subárea de *Music Information Retrieval* (MIR), uma das principais linhas de pesquisa em CM, tem enfoque em obter informação a partir de dados sonoros. Em (CASEY

et al., 2008) são apresentadas as tarefas mais comuns e suas especificidades dentro de MIR, conforme a Tabela 4.

Tabela 4: Exemplos de tarefas de MIR e suas especificidades (CASEY et al., 2008).

Caso de Uso	Especificidade	Descrição
Identificação de música	Alta	Identificar um CD, prover metadados sobre uma faixa desconhecida, recuperação de informação musical: ex. shazam.com
Detecção de plágio	Alta	Identificar atribuições erradas de performances musicais, apropriação errada de propriedade intelectual
Monitoramento de direitos de cópia	Alta	Monitorar distribuição musical com infração de direitos ou <i>royalty</i>
Versões	Alta/Média	Remixes, gravações ao vivo vs. de estúdio, músicas cover. Usado para normalização de banco de dados e remoção de resultados duplicados
Melodia	Alta/Média	Procurar trabalhos contendo um fragmento melódico
Trabalho/Título Idêntico	Média	Recuperar performances de mesmo número <i>opus</i> ou título musical
Artista	Média	Procurar música por um artista específico
Se parece com	Média	Procurar músicas que se parecem com uma dada gravação
Alinhamento de Performance	Média	Mapear uma performance em outro independente de tempo e estrutura de repetição
Compositor	Média	Procurar trabalhos de um compositor
Recomendação	Média/Baixa	Procurar música que casa com o perfil do usuário
Humor	Baixa	Procurar música usando conceitos emocionais
Estilo/Gênero	Baixa	Procurar música que pertence a uma categoria genérica: Jazz, Funk, Voz feminina
Instrumentos	Baixa	Procurar trabalhos com a mesma instrumentação
Música Falada	Baixa	Segmentação de rádio, arquivos musicais catalogados

Segundo CASEY et al. (2008), os principais componentes de um sistema MIR são: a formação de consulta, extração da descrição, correspondência e, finalmente, a recuperação do documento de música, como é exibido na Figura 9. O trabalho aqui proposto se enquadra a esta estrutura, onde pode-se verificar: a etapa de consulta faz referência aos arquivos de áudio que deverão ser classificados, eles são a entrada do sistema; já a extração de características está dentro do pré-processamento dos arquivos de

áudio, utilizando extratores predefinidos; a correspondência e a recuperação são etapas realizadas pelos agentes, que classificam o conjunto pré-processado de características e entregam ao ambiente, que por sua vez apresentará o resultado da classificação, etapa final do sistema, bem como na estrutura presente em (CASEY et al., 2008).

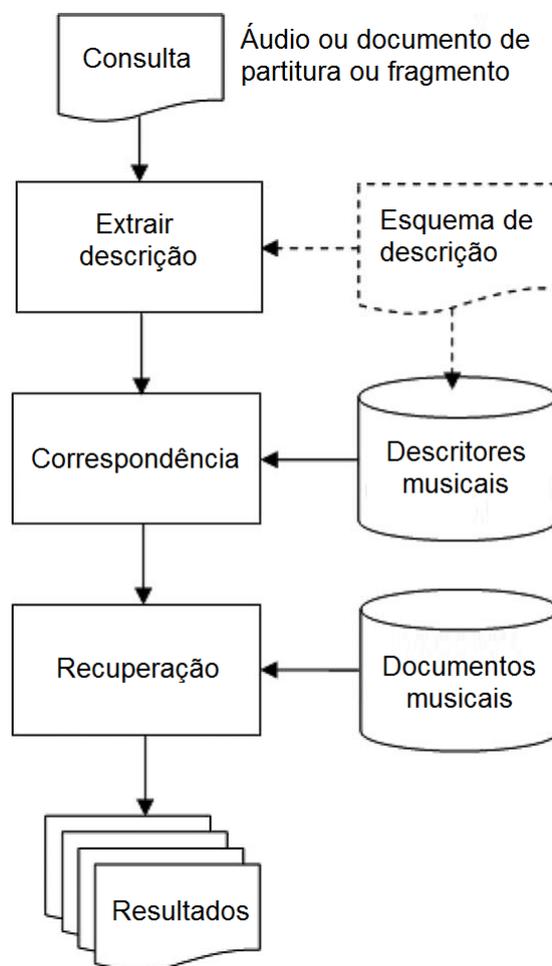


Figura 9: Fluxograma canônico de um sistema MIR (CASEY et al., 2008).

A questão mais essencial, no que diz respeito à percepção humana em relação ao som é: "quais são as qualidades dos sons de instrumentos musicais para realizar o reconhecimento de suas fontes sonoras?"

Segundo CASEY et al. (2008), o timbre é uma característica considerada de alto nível, e um subobjetivo de MIR, como apresentado na Tabela 5.

Como exemplo de trabalho encontrado no estado da arte em MIR, que trabalha com o timbre como característica principal, pode-se citar (DEVI; SRINIVAS; KRISHNA, 2012), que trata o timbre como um conjunto de características, como a envoltória sonora (a forma como o som se inicia, se mantém e termina ao longo do tempo) para a amplitude e a frequência; tempo de ataque (início de cada nota musical); caimento (em alguns instrumentos o som sofre um caimento após o ataque até se estabilizar); sustentação (corresponde ao tempo de duração da nota musical) e intensidade (DEVI; SRINIVAS;

Tabela 5: Características de alto nível (difíceis de extrair) (CASEY et al., 2008).

Descrição Alto-nível	Fonte de dados	Descrição da tarefa
Timbre	Áudio	Reconhecimento de Instrumentos Percussivo, Altura, Reconhecimento de conjunto
Melodia/Baixo	Áudio/Simbólico	Extração de linha melódica, extração de linha do baixo
Ritmo	Áudio	Detecção do ataque, identificação de andamento, alinhamento de andamento (barras), rastreio de batidas, rastreio de tempo, tempo médio
Altura	Áudio	Única frequência fundamental, múltipla frequência fundamental
Harmonia	Áudio/Simbólico	Extração de acordes, extração de baixo
Tom	Áudio/Simbólico	Rastreio de modulação, escrita melódica
Estrutura	Áudio/Simbólico	Extração de verso/coro, extração de repetições
Letras	Áudio	Detecção de vocais, identificação de letra, reconhecimento de palavras
Música oriental	Áudio	Sistemas de afinação micro tonais, conceitos canônicos orientais

KRISHNA, 2012; WIKIPÉDIA, 2015). Essas características podem ser observadas na Figura 10.

Outro estudo bastante interessante em relação ao timbre é o proposto em (SORAGHAN, 2014), que trata o timbre como uma característica multidimensional do som, assim como sugerido por (GREY, 1977), buscando uma representação visual tridimensional.

Percebe-se o quanto complexa é a extração desta característica abstrata, considerada de alto nível. Historicamente, os primeiros estudos datam dos anos cinquenta, onde pode-se citar o trabalho de (HELMHOLTZ; ELLIS, 1954), que evidenciou que as amplitudes relativas das parciais harmônicas de um som, muito mais que suas fases relativas, são determinantes primários do timbre. Este foi um dos primeiros estudos científicos que tratou o timbre como uma característica matemática, mais precisa e identificável.

Seguindo a linha do tempo, STRONG (1963) interpretou o espectro de vários instrumentos de orquestra e demonstrou que o oboé, clarinete e fagote são identificados primeiramente na base no seu espectro de magnitude.

Com estes trabalhos, começam a surgir um maior número de estudos que buscam determinadas características no formato espectral dos sons para o reconhecimento de timbre, como o trabalho proposto em (LUCY; CLARK JR, 1967), que demonstrou que a família de metais (instrumentos de sopro como trombone, saxofone, trompete, entre outros), é caracterizada por um único corte de frequência, e que esta característica se correlaciona fortemente com a identificação desta família de instrumentos. Posteriormente, (BENADE,

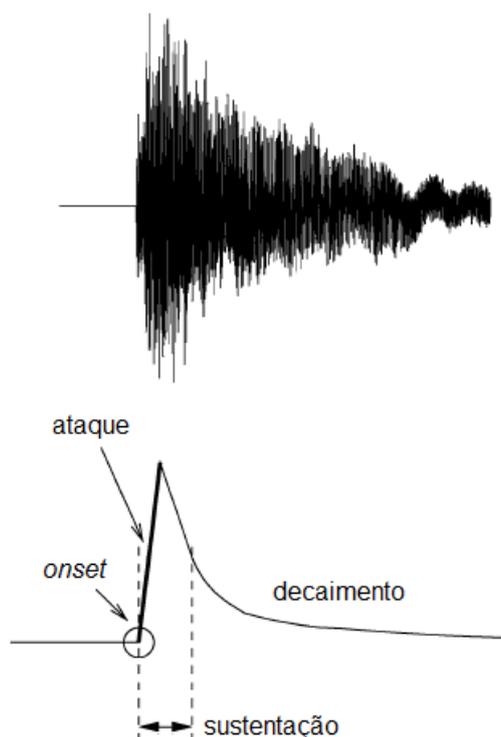


Figura 10: Formato do espectro de uma única nota sendo executada (BELLO et al., 2005).

2012) foi além, e mostrou que o corte na frequência é um dos principais determinantes do timbre nos instrumentos de sopro de um modo geral.

Outra característica que vem sendo fortemente correlacionada com o timbre é o centróide espectral, que em termos gerais pode ser definido como o “ponto de balanço” do espectro, se mostra diretamente ligado ao “brilho” do instrumento, uma dimensão primária e subjetiva do timbre, como verificado em (GREY, 1977; LICHTÉ, 1941; BISMARCK, 1974). Nos trabalhos (BEAUCHAMP, 1982a,b), o autor demonstra que o centróide varia em muitos instrumentos com a intensidade do som.

Ainda se tratando do estado da arte nesta área, destaca-se o trabalho (NEUMAN; OKPALA; BONEZZI, 2014), que analisa os efeitos da movimentação da fonte sonora em relação ao timbre produzido pela fonte, levando em conta as características do movimento, como: direção, velocidade e aceleração.

Uma linha de estudo muito atual e que apresentou relevantes publicações no último biênio, são os trabalhos que estudam a correlação entre o timbre e a emoção causada pela música. Dentre estes trabalhos pode-se destacar (CHAU; WU; HORNER, 2014; WU; HORNER; LEE, 2014), que provam a forte correlação entre o centróide do espectro (uma das principais características do timbre), com a emoção causada pela música, sendo esta uma linha de pesquisa destacada em (CASEY et al., 2008). Em (LIU; LU; ZHANG, 2003), é apresentado um estudo de extração de humor em dados acústicos musicais, utilizando características de timbre, intensidade e rítmicas.

No estudo realizado em (STRONG, 1963), foi demonstrado que, no reconhecimento de muitos instrumentos, a identificação ocorre principalmente pela envoltória temporal, em parte porque suas envoltórias espectrais não são únicas, lembrando que a definição de timbre é excludente, sendo este responsável por diferenciar sons que possuem mesma altura e intensidade (KLINGBEIL, 2009).

Neste contexto, é possível observar que o reconhecimento do timbre está fortemente ligado à forma do espectro sonoro, que usualmente é dividido em *onset*, ataque (*attack*), sustentação (*transient*) e decaimento (*decay*). Este formato é associado ao caso ideal de uma única nota sendo executada, como ilustrado na Figura 10 por (BELLO et al., 2005).

Outra característica que passa a ser analisada em relação à influência de alteração do timbre são os aspectos temporais do som, cuja importância foi reconhecida primeiramente por (RISSET; WESSEL, 1982). Em sequência a esta descoberta, pode-se citar o trabalho de (HANDEL, 1995), que apresenta a unificação da utilização das características temporais e espectrais no reconhecimento de timbre.

Já SCHOENBERG (1978) considerou o timbre como uma segunda dimensão do tom, e hoje sabe-se que o timbre pode ser considerado de alguma forma em uma característica multidimensional, como descrito primeiramente em (GREY, 1977), e inclusive pode ser representado visualmente, como foi apresentado no trabalho proposto por (SORAGHAN, 2014).

Um dos primeiros trabalhos completos com enfoque de reconhecimento de timbre, com a utilização de várias das características apresentadas, foi o proposto por (MARTIN; KIM, 1998), que considera que existem duas categorias de características para reconhecimento de timbre: temporal e espectral, ambas com grande importância no reconhecimento de timbre. Neste trabalho algumas características são utilizadas, como:

- **Altura:** sinais produzem uma estrutura identificável em relação à altura, num frame bidimensional, com defasagem no eixo horizontal e a frequência em relação à vertical, sulcos verticais indicam o período do sinal, e por inversão, a altura.
- **Envoltória espectral:** uma vez que a altura tenha sido detectada, a altura de crista vertical pode ser mensurada como uma função da frequência, para se obter uma estimativa da forma da envoltória espectral. O centroide espectral é simplesmente o centroide da envoltória espectral.
- **Intensidade:** a soma da energia na envoltória espectral aproxima a intensidade sonora instantânea do sinal. Acompanhar esta ao longo do tempo leva a medidas simples de modulação de amplitude, o que pode revelar o *tremolo* e, por correlação com modulações de frequência, ressonâncias. Como sugerido em (BEAUCHAMP, 1982a,b), a relação entre a intensidade e o centroide espectral pode ser uma importante correlação perceptual do timbre.

- Ataque assíncrono: ao rastrear a envoltória espectral ao longo do tempo, com estimativas de altura concorrentes, é possível medir as características de ataque de um tom harmônico musical.
- Inarmonia: os desvios harmônicos no sinal vão ser refletidos como desvios da estrutura vertical no gráfico de amplitude por tempo.

Ao todo, o trabalho de (MARTIN; KIM, 1998) utilizou trinta e uma características diferentes, a maioria sendo variações das citadas acima. Dentre as estratégias de reconhecimento adotadas, utilizou-se o método k-NN, hierárquico e não-hierárquico para identificação de timbres.

Posteriormente, ERONEN et al. (2001) apresenta um sistema de reconhecimento de instrumentos, utilizando conhecimento psicoacústico¹ nos algoritmos de extração de características, se valendo de alguns dos principais métodos, como é representado na Figura 11, um diagrama de blocos do sistema de reconhecimento de instrumentos por ele implementado.

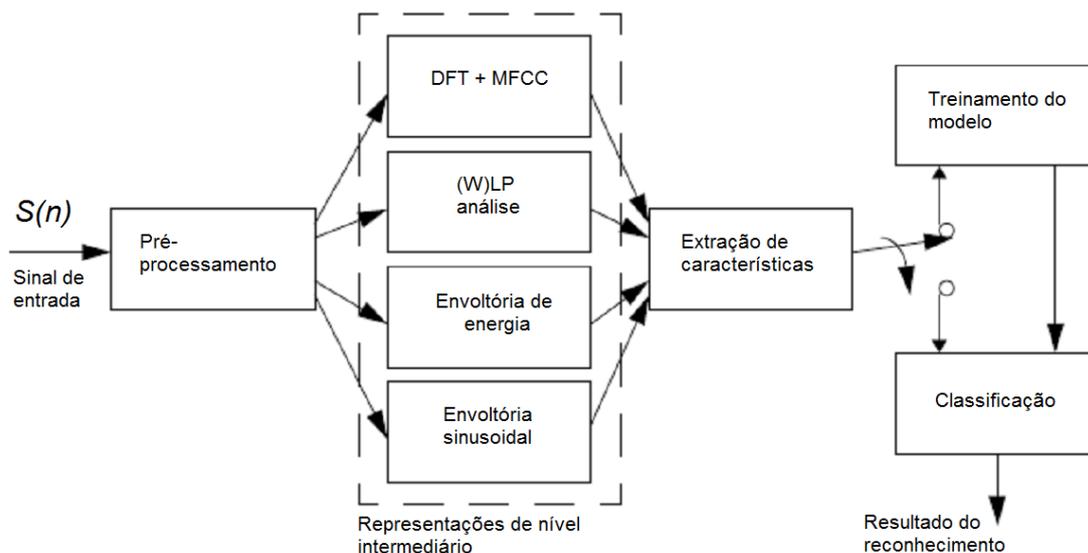


Figura 11: Diagrama de blocos do sistema de reconhecimento de instrumentos implementado por (ERONEN et al., 2001).

Como descrito por (ERONEN et al., 2001), *Formantes* são picos de energia em uma região do espectro sonoro. Os parciais que se encontram nessa região de ressonância são realçados. Elas representam a informação essencial para o reconhecimento de voz e fala, e também para reconhecimento de instrumentos musicais. Um recurso robusto para medir a informação de *Formantes*, ou a envoltória espectral suavizada, são os coeficientes

¹Psicoacústica é o estudo científico da percepção do som. Mais especificamente, é o ramo da ciência que estuda as respostas psicológicas e fisiológicas associadas ao som (incluindo a fala e a música).

cepstrais. Dentre as técnicas utilizadas no trabalho de (ERONEN et al., 2001), está *Mel-frequency Cepstral Coefficients* (MFCC) (DAVIS; MERMELSTEIN, 1980), que se tornou uma das técnicas mais populares na extração de características em sistemas automáticos de reconhecimento de fala (ERONEN et al., 2001).

Existem muitas classificações diferentes para as características dos timbres. Uma proposta com as principais definições destas características é apresentada por (TZANETAKIS; COOK, 2002), que também pode ser encontrada em (LIU; LU; ZHANG, 2003), conforme Tabela 6.

Tabela 6: Definição das características dos timbres (LIU; LU; ZHANG, 2003).

Característica		Definição
Características relativas a forma do espectro	Centroide	Média da transformada rápida de Fourier da amplitude do espectro.
	Largura de banda	Média da amplitude ponderada das diferenças entre os componentes do espectro e o centroide.
	<i>Roll-off</i>	95% da distribuição espectral.
	Fluxo espectral	Distância de duas normais da diferença de amplitude espectral <i>frame a frame</i> .
Características de contraste	Pico de sub-banda ²	Valor médio do entorno de uma pequena vizinhança dos valores de máxima amplitude dos componentes espectrais em cada sub-banda ² .
	Vale de sub-banda ²	Valor médio do entorno de uma pequena vizinhança dos valores de mínima amplitude dos componentes espectrais em cada sub-banda ² .
	Média de sub-banda ²	Amplitude média de todos os componentes espectrais em cada sub-banda ² .

No estudo proposto por (ZHANG; RAS, 2007), é demonstrado que existem muitas formas de categorizar as características de áudio. O autor apresenta uma classificação em descritores MPEG7 e não-MPEG7, um padrão de descrição de conteúdo multimídia, formalmente conhecido como *Multimedia content description interface*, e criado pelo grupo *Moving Picture Experts Group*, que provê a sigla MPEG (WATKINSON, 2004).

Existe uma grande quantidade de trabalhos que lidam com o timbre. No entanto, a maior parte centrou-se em amostras simples de som, monofônicas, dentro do contexto de reconhecimento de instrumentos (ERONEN et al., 2001; HERRERA-BOYER; PEETERS; DUBNOV, 2003; ZHANG; RAS, 2007). Ou seja, esses trabalhos tentam identificar se uma nota, por exemplo “A4”, está sendo executada por um trompete ou um clarinete (AUCOUTURIER; PACHET; SANDLER, 2005). Como grande contribuição, o trabalho de (AUCOUTURIER; PACHET; SANDLER, 2005) explora o reconhecimento de timbres

²Sub-banda é qualquer forma de transformação de codificação que divide um sinal para um certo número de bandas de frequência diferentes e codifica cada uma de forma independente.

polifônicos, ou seja, como modelar o som global de uma determinada música. Segundo o autor, misturas de distribuições gaussianas sobre mais de um espaço de coeficientes cepstrais (resultado da transformada inversa de Fourier sobre o espectro) são uma maneira eficiente para modelar as texturas encontradas em um determinado sinal de música. Tais modelos de timbre fornecem novas soluções para muitas questões tradicionalmente encontradas no processamento de sinais musicais e na recuperação de informação musical.

Os métodos de áudio de baixo nível normalmente tratam música polifônica em massa. Então, essa música com vários instrumentos tocando simultaneamente é representada como a soma da informação desses instrumentos separados. Grandes avanços foram feitos nos últimos anos em métodos de separação de fontes sonoras para extração de informações relativas aos instrumentos individualmente a partir de uma mistura polifônica. Esse tipo de processamento é um dos problemas mais desafiadores em processamento de áudio, mas os novos trabalhos nesta área parecem promissores (CASEY et al., 2008).

Percebe-se, pelos trabalhos analisados, que ainda estão sendo realizados estudos para aumentar a precisão no reconhecimento de timbres, bem como identificar novos descritores e aumentar o desempenho, eficiência e robustez desses sistemas. Também percebe-se que diferentes características são utilizadas no reconhecimento de timbres, como o *corte de frequência*, que é uma importante característica no reconhecimento de metais (LUCE; CLARK JR, 1967); a envoltória espectral determina diferentes famílias de instrumentos (STRONG, 1963). Como cada instrumento possui determinadas características de timbre que são identificadas por diferentes descritores, é possível indagar que uma solução distribuída e paralela seja adequada para reconhecimento polifônico, aumentando a eficiência na solução de problemas.

3.2 Trabalhos Relacionados com SMA

A tecnologia de Sistemas Multiagente é uma nova forma promissora para a performance musical interativa. Em trabalhos recentes, essa tecnologia foi adaptada para resolver problemas específicos em um escopo musical limitado, como detecção de pulso, simulação de instrumentos ou acompanhamento automático. Dentre deste contexto, existe alguns trabalhos na área de Sistemas Multiagente que solucionam problemas de Computação Musical, nas suas mais variadas possibilidades, conforme ilustra a Figura 12.

A inter-relação entre SMA e CM tem crescido nas últimas décadas. É possível inferir algumas causas desta consequência, como o fato da área de Sistemas Multiagente já ser bastante consolidada, sendo adequada para solucionar problemas que exijam distribuição, seja de natureza lógica ou geográfica, e em que a complexidade do problema seja minimizada por esta abordagem.

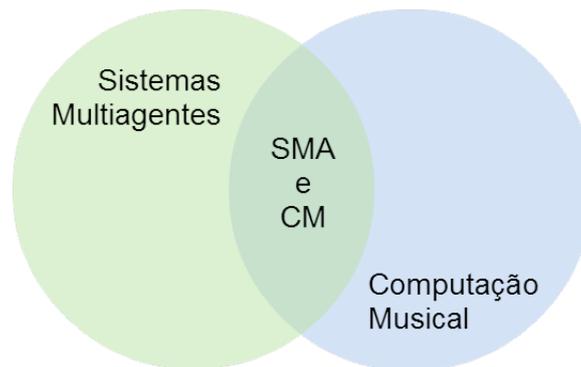


Figura 12: Intersecção das grandes áreas da Computação Musical e Sistemas Multiagente [Imagem elaborada pelo autor].

Segundo (THOMAZ, 2009), a tecnologia de agentes se torna particularmente adequada para aplicações musicais, devido à possibilidade de associar um agente computacional com o papel de um cantor ou instrumentista. O autor destaca algumas vantagens dessas associações, como mapear características como desempenho, percepção, adaptação e improvisação de um lado e processos artificiais no outro. Além disso, é possível definir formas de inter-relação social entre os agentes, que traz esta tecnologia ainda mais perto de performance musical colaborativa.

Segundo SAMPAIO; TEDESCO; RAMALHO (2005), na interseção entre as áreas de Inteligência Artificial e Computação Musical, vários sistemas foram criados aplicando o conceito de agentes inteligentes. Nesses sistemas, diversos aspectos da música são abordados, porém a ênfase maior costuma ser na geração de harmonia e/ou melodia. Poucos sistemas focam especificamente a parte rítmica, que é, entretanto, considerada fundamental para a “sensação” que a música causa em ouvintes humanos, além de ter grande importância cultural, estando presente nos estados mais primitivos da humanidade.

Um exemplo da interseção entre SMA e CM é o trabalho de (THOMAZ, 2009), que apresenta uma taxonomia de sistemas multiagente musicais, e uma implementação de um framework computacional que engloba obras de trabalhos anteriores e problemas de baixo nível como a sincronização em tempo real, a comunicação sonora e mobilidade espacial dos agentes. Ao usá-lo, o usuário pode desenvolver um sistema multiagente musical centrado principalmente em suas necessidades musicais, deixando a maior parte dos problemas técnicos para o framework. Utiliza a ferramenta JADE 3.6 para a criação do framework e tem o foco principalmente no ambiente e interações. Como continuidade deste trabalho, (THOMAZ, 2011a,b) discutem alguns detalhes de implementação e características de um framework implementado pelos autores, incluindo a troca de eventos entre agentes, movimentação dos agentes em um mundo virtual, simulação 3D de propagação do som realista, interface com outros sistemas e bibliotecas de processamento de áudio. Segundo THOMAZ (2009), embora a literatura sobre o uso de agentes em música seja bastante extensa, a maioria trata de problemas muito específicos.

A arquitetura MAMA (MURRAY-RUST; SMAILL; EDWARDS, 2006), oferece uma estrutura para projetar agentes musicais com o comportamento interativo baseado na teoria dos atos de fala (SEARLE, 1969), que se comunicam através de mensagens MIDI para executar uma peça musical. A Orquestra SWARM (D. BISIG; FLURY, 2008) é uma biblioteca que pode ser estendida pelo usuário que lida com grandes e complexas populações (enxames), que podem ser usados para controlar vários parâmetros musicais e de movimento simultaneamente (THOMAZ, 2011a).

Em contraponto aos trabalhos mais abrangentes, como os que visam a criação de um framework para utilização de sistemas multiagente musicais, existem trabalhos com enfoques mais específicos, como o de (SAMPAIO; TEDESCO; RAMALHO, 2005; SAMPAIO; RAMALHO; TEDESCO, 2008) (ver Figura 13), onde agentes percussionistas inteligentes participam de uma performance puramente rítmica. Cada agente conhece um repertório limitado de padrões rítmicos de um instrumento de percussão e precisa interagir com os demais agentes para escolher qual padrão tocar com o grupo de maneira que a performance seja satisfatória. Usuários humanos podem usar o sistema para testar diferentes formações e diferentes misturas de estilos, fazendo o sistema funcionar como um laboratório rítmico, que pode vir a ser útil para músicos compositores e arranjadores.

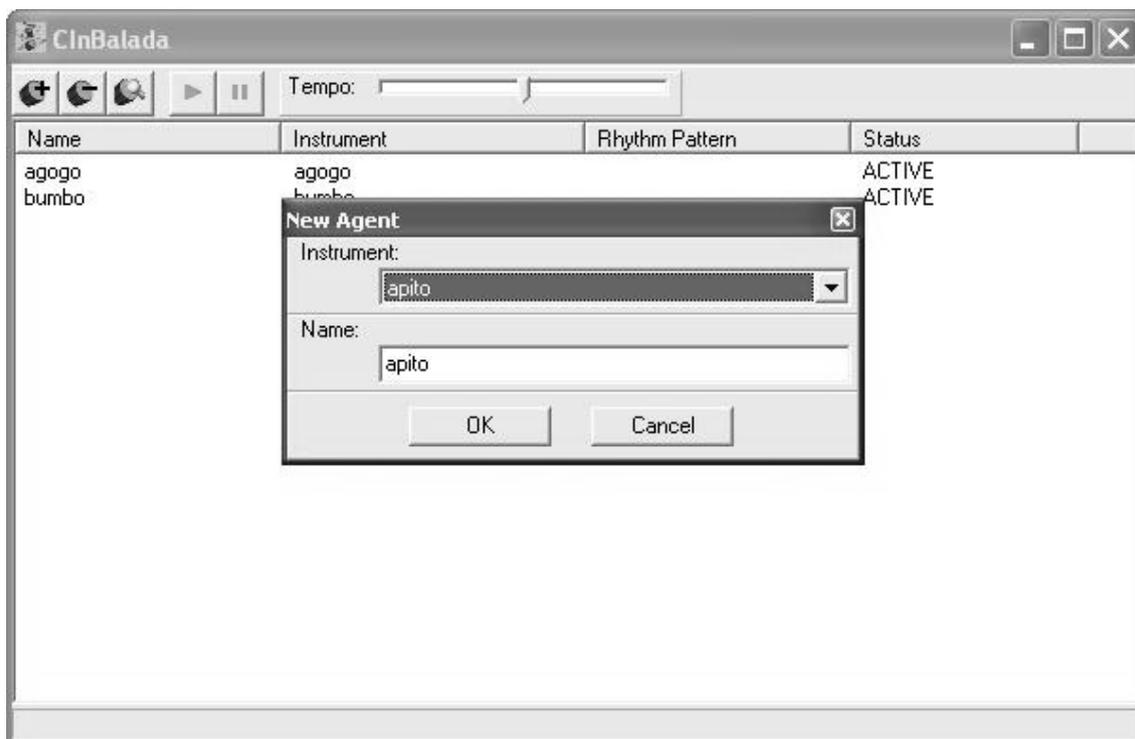


Figura 13: Interface principal do CInBalada (SAMPAIO; RAMALHO; TEDESCO, 2008).

O CInBalada (SAMPAIO; TEDESCO; RAMALHO, 2005) foi desenvolvido em C++ e possui três módulos centrais: uma biblioteca musical, a plataforma multiagente e os agentes do sistema. Para permitir a operação dos vários agentes do sistema foi imple-

mentada uma plataforma multiagente: o FAMA (Framework para Aplicações MultiAgente). O FAMA é inspirado no JADE (BELLIFEMINE; POGGI; RIMASSA, 1999), e foi implementado devido à carência de plataformas multiagente para C++. A plataforma oferece todos os serviços básicos recomendados pela FIPA (Foundation for Intelligent Physical Agents): gerenciamento do ciclo de vida dos agentes, serviços de localização de agentes (páginas amarelas e páginas brancas) e transporte mensagens. No entanto, a plataforma do FAMA não é distribuída, uma vez que distribuição não é um requisito do projeto CInBalada (SAMPAIO; TEDESCO; RAMALHO, 2005; SAMPAIO; RAMALHO; TEDESCO, 2008).

Um trabalho interessante que prova a grande variabilidade de aplicações que podem ser abordadas se valendo de sistemas multiagente e que resolve um problema bastante específico é apresentado por (PERCHY; SARRIA, 2009), que propõe um modelo de descrição de dissonâncias para uso computacional. Primeiramente, este trabalho trata de descrever a dissonância comparando-a com acordes não dissonantes. Em seguida, parte para uma modelagem e descrição computacional da representação desta dissonância, utilizando *Real-Time Concurrent Constraint Calculus*, que é uma extensão do CCP desenvolvida para especificar sistemas reativos com comportamento de tempo real. Realizaram a implementação de um interpretador de RTCC, e acreditam que um software ajudaria a melhor visualizar o comportamento dos sistemas, para ser possível ouvir os resultados de áudio dos modelos em tempo real, e provar propriedades destes modelos (PERCHY; SARRIA, 2009). Neste trabalho, o foco está no agente e na interação entre os agentes, diferentemente dos anteriores citados, pois estes se valiam de sistemas complexos, bem como a elaboração de frameworks para utilização de agentes. O trabalho de (PERCHY; SARRIA, 2009), por sua vez, utiliza agentes para solucionar um problema mais básico, utilizando as performativas de comunicação *ask* e *tell*.

Outro exemplo é o proposto em (BOWN, 2009). Se trata da música gerada computacionalmente, que usa os princípios utilizados em evolução e dinâmica de ecossistemas. Esses algoritmos são diferentes dos tradicionais algoritmos genéticos iterativos em vários aspectos. As possibilidades de uma abordagem deste tipo podem ser exploradas utilizando sistemas multiagente, que podem assumir diferentes configurações (ver Figura 14). Desta forma, este trabalho tem por objetivo criar um framework que possa utilizar agentes para gerar música fazendo uso dos princípios da evolução e dinâmica de ecossistemas.

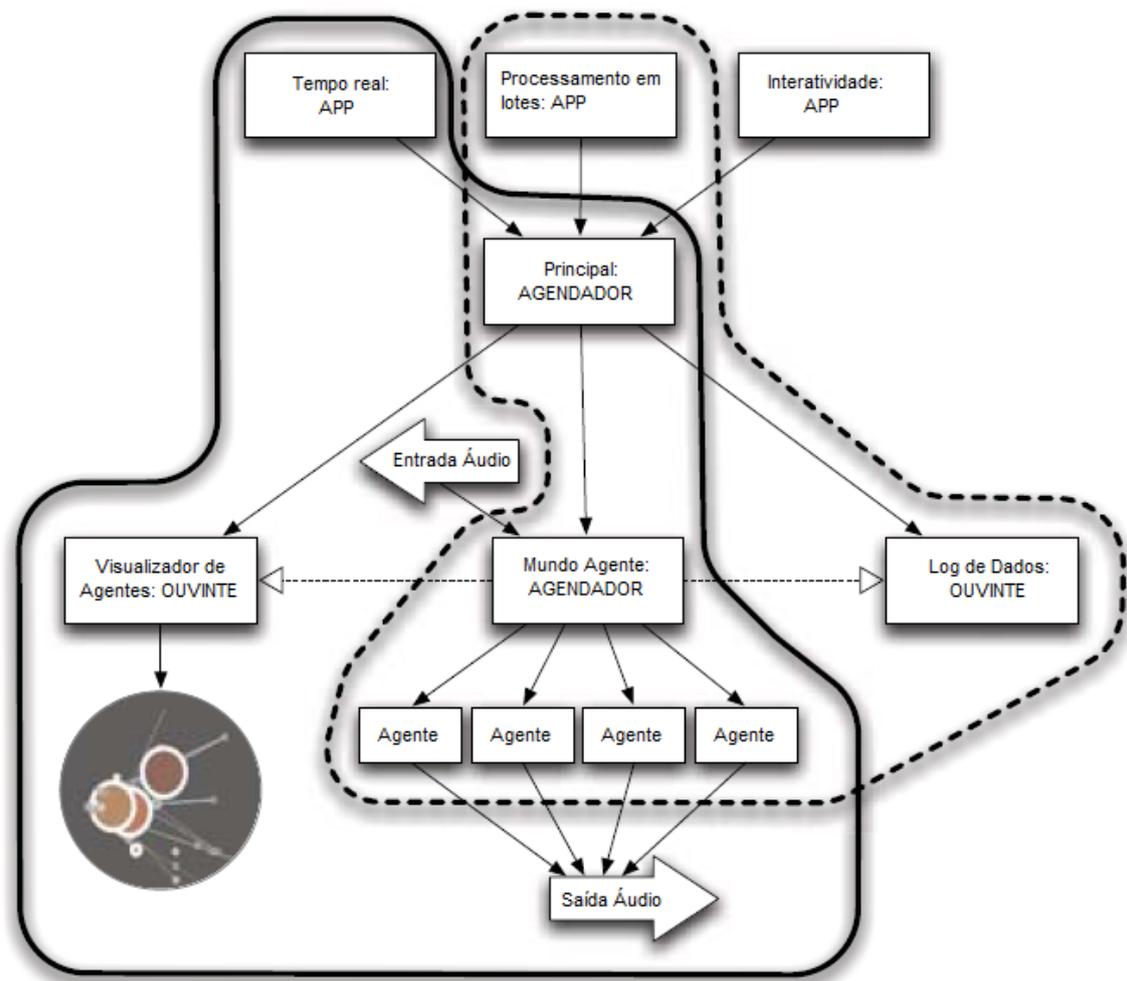


Figura 14: Exemplo de possíveis configurações do SMA proposto em (BOWN, 2009).

Resumindo, existe uma grande diversidade de trabalhos que adotam Sistemas Multiagente como ferramenta tanto se tratando do modelo ou para verificar algum resultado de forma empírica. Na área da Computação Musical isto não ocorre de forma diferente, sendo os Sistemas Multiagente úteis para modelar os mais diversos problemas, sejam eles de grande complexidade ou não.

Contudo, dentro do contexto apresentado, ainda são carentes trabalhos na área de MIR que utilizem abordagens distribuída e paralela de reconhecimento de timbre, utilizando a tecnologia de sistemas multiagente.

4 ULISES: UM SISTEMA BASEADO EM AGENTES PARA CLASSIFICAÇÃO DE TIMBRES

Esse capítulo apresenta o sistema Ulises, um sistema baseado em agentes para resolução do problema de classificação de timbres.

4.1 Origem do nome

A proposta inicial deste trabalho foi idealizada em conjunto com o prof. Dr. Ulises Dardo Ferretti Martínez, grande amigo e professor de música com quem compartilhei ótimos momentos e ideias. Este começou a ser desenvolvido sob sua coorientação, mas infelizmente ele veio a falecer antes de ver o trabalho concluído. Em memória a esta grande pessoa, faço esta singela homenagem.

4.2 Metodologia

Como visto anteriormente, o estudo do timbre é uma área de grande importância para a computação musical. Segundo (ERONEN et al., 2001), o problema de reconhecimento de timbre exige quatro etapas principais:

- Pré-processamento do sinal a ser analisado.
- Representação em nível intermediário deste sinal.
- Extração de características da fonte sonora.
- Método de classificação baseado nestas características.

Com intuito de paralelizar sistemas deste tipo, é possível imaginar duas formas de paralelização, a primeira diz respeito à representação em nível intermediário e extração de características, e a segunda diz respeito ao método de classificação. A partir disto, duas abordagens são possíveis, uma provavelmente visando maior eficiência do sistema e outra provavelmente visando maior qualidade dos resultados. Ao se paralelizar a etapa

de nível intermediário do sinal, cada entidade não terá conhecimento de todas as características que podem ser extraídas do timbre, que seriam representadas por outras unidades no nível intermediário. Entretanto, os cálculos mais demorados seriam paralelizados, assim aumentando a eficiência do sistema. Na segunda abordagem, em que o modelo de classificação é paralelizado, podem surgir resultados mais precisos, pelo fato que cada unidade computacional teria o conhecimento de todas as características mais importantes para o reconhecimento, e, a partir da interação com outros agentes, cada agente aperfeiçoaria sua classificação baseando-se nas características extraídas e o *feedback* de outros agentes.

Existe uma questão importante na paralelização destes sistemas, a polifonia. A polifonia diz respeito à execução de múltiplos sons simultaneamente, sejam de instrumentos distintos ou não, dentro de uma mesma janela temporal. O resultado desta simultaneidade, em nível de sinal, é uma sobreposição de ondas. O trabalho (AUCOUTURIER; PACHET; SANDLER, 2005) aborda uma metodologia de classificação para sons polifônicos, bem como em (KLINGBEIL, 2009). Os trabalhos atuais de reconhecimento de timbre com polifonia adotam uma mesma abordagem: o som sofre um pré-processamento, então este é discriminado se valendo de algum método, como misturas de distribuições gaussianas sobre mais de um espaço de coeficientes cepstrais, a fim de se obter um conjunto de sons separados e então classificá-los (AUCOUTURIER; PACHET; SANDLER, 2005). Assim sendo, se tratando de polifonia, poderia se desenvolver um método com diversas unidades computacionais diferentes, cada uma responsável pela detecção de um timbre, e estas unidades iriam inferir sobre este sinal polifônico reportando a presença ou não do timbre reconhecível por esta unidade.

Este trabalho tem como enfoque a paralelização do modelo de classificação, sendo um passo inicial nas abordagens baseadas em agentes para o problema de reconhecimento de timbres. O sistema desenvolvido é baseado em agentes cognitivos, que podem receber um conjunto de características do ambiente ou de outros agentes, descritos na Seção 4.5. Ele é flexível para solucionar o problema da classificação de timbres em sons polifônicos, caso seja adicionado um módulo que discrimine a entrada polifônica em sons monofônicos como é realizado na maioria das abordagens. De um modo geral, o sistema pode ser descrito nas seguintes etapas:

1. A partir do banco de dados (um conjunto de entradas de áudio), os arquivos são carregados e é realizado pré-processamento para normalização;
2. É realizada a extração de características utilizando um conjunto de métodos descritos na Seção 4.5;
3. O ambiente troca informação com os agentes que têm acesso ao conjunto de características extraídas dos sinais de entrada, que servem de parâmetro para duas etapas:

- (a) treinamento do agente;
 - (b) classificação do sinal de entrada;
4. Por fim o conjunto de entrada é classificado pelos agentes, e o resultado é mostrado na visualização do ambiente.

Um fluxograma desta abordagem pode ser visualizado na Figura 15. Esta estrutura se assemelha bastante com os principais componentes de um sistema de MIR descritos em (CASEY et al., 2008): formação de consulta, extração da descrição, correspondência e, finalmente, a recuperação do documento de música.

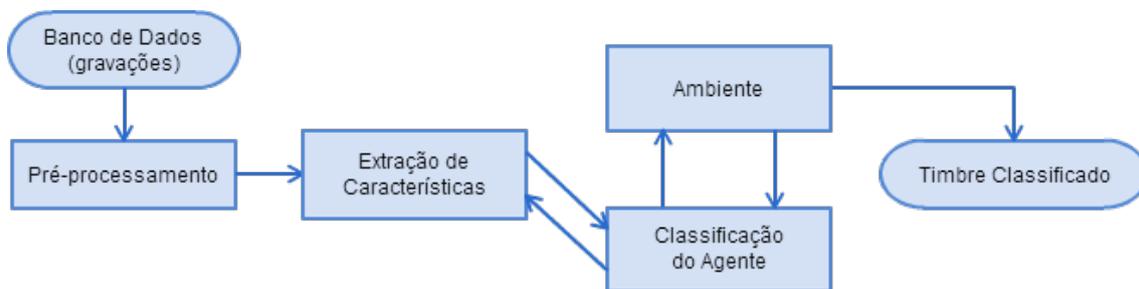


Figura 15: Fluxograma de funcionamento do sistema Ulises.

Dentre as vantagens esperadas em adotar um sistema baseado em agentes para a resolução do problema de classificação de timbres, pode-se citar:

- Maior escalabilidade do sistema, pois é possível adicionar um novo agente, responsável por classificar um novo instrumento, sem realizar um novo treinamento em todo o sistema.
- Paralelização da classificação, pois quando existir uma tarefa de classificação de muitas entradas, a classificação delas poderá ser distribuída entre vários agentes.
- Aperfeiçoamento dos resultados, já que cada agente será específico para cada timbre é possível que eles sejam melhores especialistas do que uma única entidade classificadora, ou que, dependendo da implementação, troquem informação entre si para obter melhores resultados.

4.2.1 Ciclo de Vida dos Agentes

Todos os agentes possuem o mesmo fluxo de execução. Eles foram implementados como é exemplificado pelo fluxograma da Figura 16. De uma forma resumida, logo após a criação do agente, a única opção possível é o seu treinamento ou então ele permanecerá no estado inicial. Após o treino, o agente está apto para que o ambiente ou outro agente atribua um arquivo para que ele efetue a classificação. Dependendo da saída da classificação, ele pode se considerar o especialista do instrumento analisado ou então deverá encaminhar a entrada para outro agente. Depois deste processo, ele reporta a saída

para o ambiente e volta para o estado de prontidão, se tornando apto para que o ambiente atribua uma nova entrada para classificação.

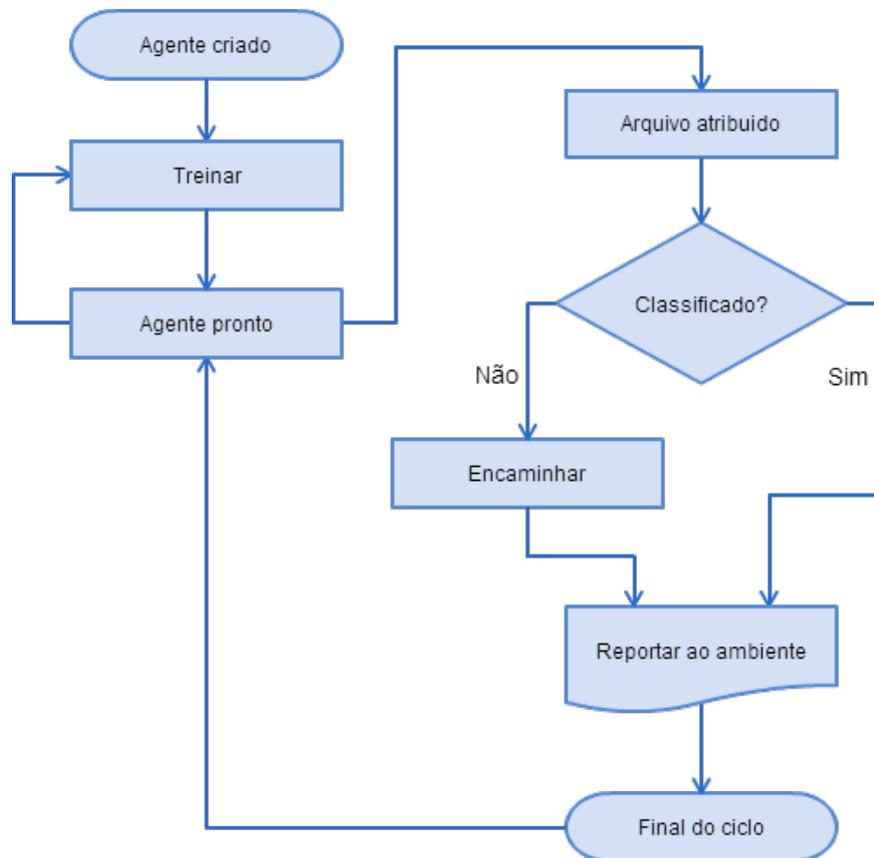


Figura 16: Ciclo de Vida do agente.

4.2.2 Funcionamento do Ambiente

O funcionamento do ambiente, demonstrado na Figura 17. Ele tem basicamente três funções: gerenciar a fila de solicitações de arquivos a serem classificados; atribuir as tarefas aos agentes; e exibir os resultados. Após a etapa de *setup*, o sistema está pronto para começar a operar, mas só apresentará alguma funcionalidade caso ao menos dois agentes estejam treinados. O ambiente gerencia a fila de solicitações, que é uma fila que contém identificadores dos arquivos que devem ser classificados.

Na execução de um passo, ele verifica se existem arquivos na fila a serem atribuídos aos agentes e se existem agentes disponíveis. Caso afirmativo, o ambiente atribui um arquivo para um agente, e, em caso negativo, ele solicita para um agente classificar a entrada a ele atribuída. O agente, por sua vez, realizará seu método de classificação, se o resultado obtido for maior que um *limiar empírico*, ele reportará ao ambiente que ele é o classificador correto da entrada; caso contrário, ele deverá encaminhar a entrada para outro agente que esteja disponível. O ambiente também é responsável por exibir toda a troca de informação do sistema como descrito na Seção 4.4.

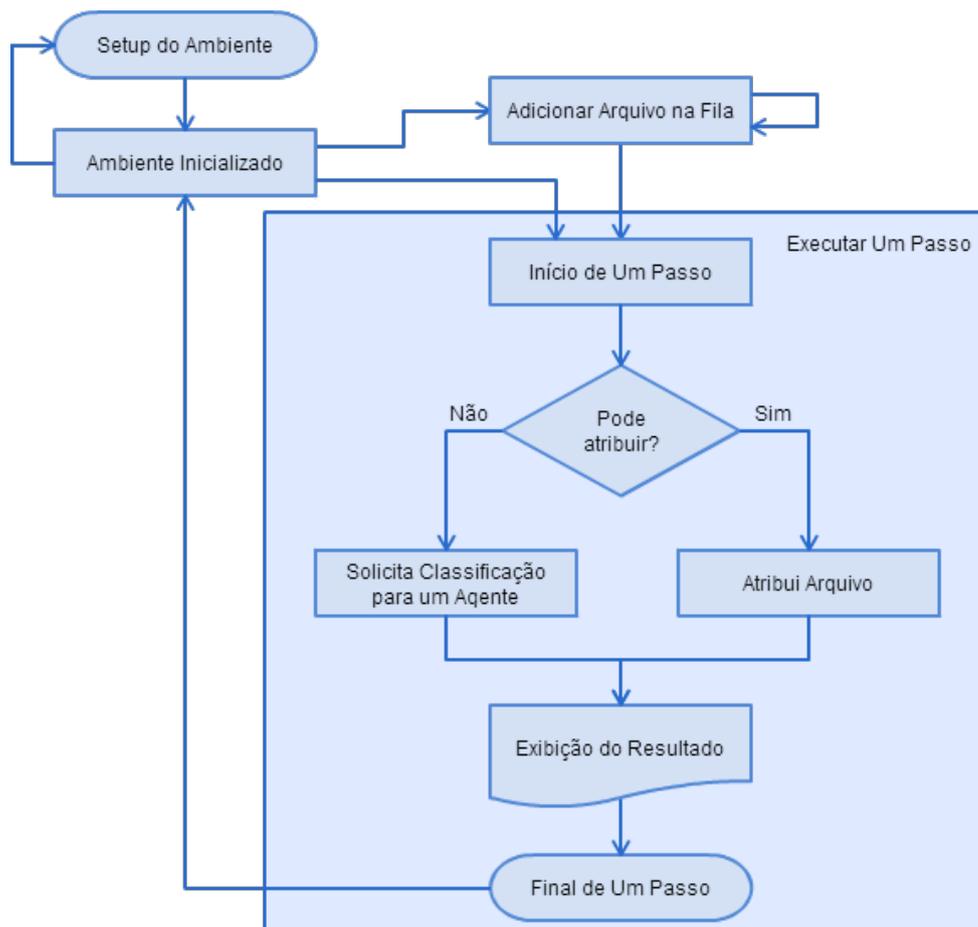


Figura 17: Fluxograma de funcionamento do ambiente.

4.3 Ferramentas Utilizadas

No desenvolvimento do sistema Ulises foram utilizadas quatro ferramentas principais, interligadas como é demonstrado no diagrama da Figura 18, são elas:

NetLogo¹ é um ambiente de modelagem programável para simular fenômenos naturais e sociais. Foi escrito por (WILENSKY, 1999) e está em desenvolvimento contínuo no *Center for Connected Learning and Computer-Based Modeling*. Ele é particularmente adequado para modelar sistemas complexos executados ao longo do tempo. Os modeladores podem dar instruções a centenas ou milhares de agentes que operam de forma independente. Isso torna possível explorar a conexão entre o comportamento individual dos agentes bem como os padrões que emergem de sua interação. O NetLogo é simples o suficiente para ser trabalhado em aula por estudantes e professores, e ainda robusto o suficiente para servir como uma ferramenta poderosa para os pesquisadores em diversos campos do conhecimento. Possui uma extensa documentação e um grande número de tutoriais. Ele também vem com uma vasta coleção de simulações pré-escritas que podem ser utilizados

¹<https://ccl.northwestern.edu/netlogo/>

e modificadas. Estas simulações abordam áreas de conteúdo nas ciências naturais e sociais, incluindo biologia e medicina, física e química, matemática e ciência da computação, além de economia e psicologia social (NETLOGO, 2016).

MatLab² trata-se de um software de alta performance voltado para o cálculo numérico. Ele integra análise numérica, cálculo com matrizes, processamento de sinais e construção de gráficos em um ambiente de fácil utilização, onde problemas e soluções são expressos matematicamente, com uma linguagem de programação voltada para a fácil manipulação de matrizes. Em outras palavras, é um sistema interativo cujo elemento básico de informação é uma matriz que não requer dimensionamento. Milhões de engenheiros e cientistas em todo o mundo usam o MatLab para analisar e projetar sistemas e produtos. Ele está presente em sistemas de segurança automotiva, naves interplanetárias, dispositivos de monitoramento de saúde, redes inteligentes de energia e redes de celulares. É usado para o aprendizado de máquina, processamento de sinais, processamento de imagem, visão computacional, comunicações, finanças computacionais, design de controle, robótica, e muito mais. A plataforma MatLab é otimizada para resolução de problemas de engenharia e pesquisa científica. A sua linguagem matricial é uma maneira bastante natural de expressar matemática computacional. Ele possui uma grande variedade de *toolboxes*, que são plugins com um conjunto de ferramentas específicas para determinadas áreas do conhecimento, como a recuperação de informação musical (MIRtoolbox). Todas ferramentas e características do MatLab são rigorosamente testadas e projetadas para trabalharem juntas, o que torna este software ainda mais versátil (MATHWORKS, 2016).

MatNet³ é uma extensão do NetLogo para utilização com o MatLab. Incluindo a extensão MatNet no código *Logo*, o NetLogo fará uma pausa enquanto o MatLab é iniciado. Uma janela do Matlab será aberta, e será possível de ver a saída a partir do servidor em execução no MatLab. O NetLogo só possui dois tipos de dados básicos (*Strings/Listas de Strings* e *Doubles/Listas de Doubles*) que são compatíveis com o Matlab. Desta forma, a extensão consegue utilizar apenas funções com esses tipos de dados, sejam eles parâmetros ou retornos. Todas as variáveis passadas para o MatLab, ou resultados de declarações “*eval*”, persistem no ambiente do MatLab e podem ser acessadas novamente. A recuperação de variáveis resultantes de operações nativas do MatLab dentro do NetLogo é uma tarefa simples. A dificuldade está na obtenção de valores de funções ou scripts próprios. A solução é o comando “*assignin*”, que cria variáveis na área de trabalho do MatLab, em vez de simplesmente retorná-las. Desta forma, é possível que o NetLogo recupere os

²<http://www.mathworks.com/products/matlab/>

³<http://github.com/mbi2gs/netlogo-matlab-extension/wiki>

resultados de funções próprias.

MIRtoolbox⁴ é uma *toolbox* do Matlab dedicada à extração de características musicais a partir de arquivos de áudio, incluindo rotinas de análise estatística, segmentação e clustering. MIRtoolbox integra uma sintaxe de fácil utilização que permite combinar facilmente operadores de baixo e de alto nível em fluxogramas complexos. O design modular da MIRtoolbox é guiado da seguinte forma: técnicas desenvolvidas para certos domínios de análise de música são transformados em operadores gerais que poderiam ser usados para diferentes propósitos analíticos. Cada método de extração de características pode aceitar como argumento um arquivo de áudio ou qualquer resultado preliminar de estágios intermediários da cadeia de operações. Também a mesma sintaxe pode ser usada para análises de arquivos de áudio individuais, lotes de arquivos, série de segmentos de áudio, sinais multicanais, etc. Para este efeito, os dados e os métodos da *toolbox* são organizados em uma arquitetura orientada a objetos (LARTILLOT; TOIVIAINEN, 2007).

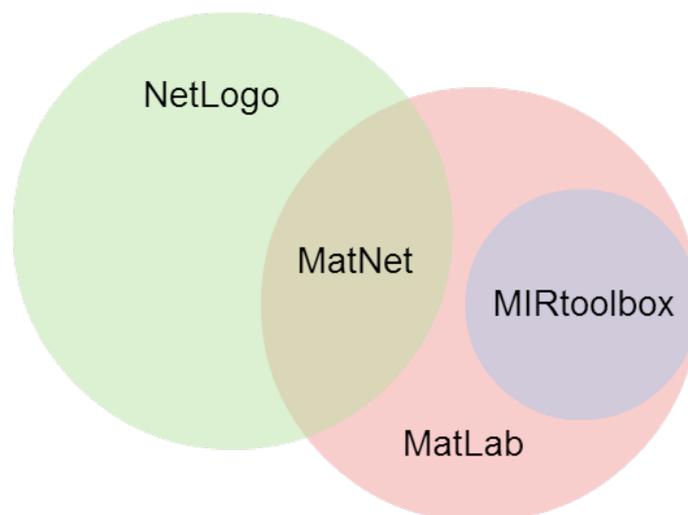


Figura 18: Diagrama das ferramentas utilizadas na implementação do sistema Ulises.

4.3.1 Banco de Dados

O banco de dados utilizado consiste em uma coleção de arquivos .wav com gravações de notas únicas de diversos instrumentos (Anexo C). As gravações utilizadas possuem diversas frequências fundamentais (alturas e notas musicais distintas) e não possuem ruídos ou sons ambiente. Os arquivos foram obtidos através da OLPC (*One laptop per child - free sound samples*⁵), que se encontra atualmente na sua segunda versão, contendo 8458 arquivos de áudio somando um volume maior que 8GB de dados. Entretanto, nem todos arquivos obtidos através da OLPC estão prontos para serem utilizados, pois esta biblioteca

⁴<http://www.mathworks.com/matlabcentral/fileexchange/24583-mirtoolbox>

⁵http://wiki.laptop.org/go/Free_sound_samples

contém uma grande diversidade de gravações, obtidas das mais variadas fontes. Existem gravações ruidosas, de exemplos de fala, de sons ambiente, entre muitos outros. Para o estudo da classificação de timbres, idealmente, os sons devem pertencer a instrumentos, ser gravados em ambientes isolados acusticamente, sem a presença de ruídos, e a execução deverá registrar apenas uma nota de cada vez, ou seja, sons monofônicos. Para este fim, foram selecionados arquivos desta biblioteca que se enquadrassem nestas características, das quais pode-se destacar:

- bancos da *Berklee College of Music*⁶
 - *The Berklee Sampling Archive - Volume 3*⁷
 - *The Berklee Sampling Archive - Volume 4*⁸
 - *The Berklee Sampling Archive - Volume 11*⁹
 - *The Berklee Sampling Archive - Volume 12*¹⁰
- bancos da *Open Path Music*¹¹
 - *The Open Path Music Collection - Volume 1*¹²
 - *The Open Path Music Collection - Volume 2*¹³
 - *The Open Path Music Collection - Volume 3*¹⁴
 - *The Open Path Music Collection - Volume 4*¹⁵
 - *The Open Path Music Collection - Volume 5*¹⁶
- bancos da *TamTam*¹⁷
 - *TamTam Drumkits*¹⁸
 - *TamTam Original Sample Set*¹⁹

Entre outros arquivos obtidos da comunidade da OLPC que se enquadraram nas características citadas.

⁶<http://www.berklee.edu/>

⁷<http://www.archive.org/details/Berklee44v3>

⁸<http://www.archive.org/details/Berklee44v4>

⁹<http://www.archive.org/details/Berklee44v11>

¹⁰<http://www.archive.org/details/Berklee44v12>

¹¹<http://www.openpathmusic.com/>

¹²<http://www.archive.org/details/OpenPathMusic44V1>

¹³<http://www.archive.org/details/OpenPathMusic44V2>

¹⁴<http://www.archive.org/details/OpenPathMusic44V3>

¹⁵<http://www.archive.org/details/OpenPathMusic44V4>

¹⁶<http://www.archive.org/details/OpenPathMusic44V5>

¹⁷<http://wiki.laptop.org/go/Tamtam>

¹⁸<http://www.archive.org/details/TamTamDrumKits>

¹⁹<http://www.archive.org/details/tamtam44old>

Por padrão, a nomeação dos arquivos no banco de dados segue a forma: “X-Y.wav”, onde X é o identificador do instrumento, conforme a Tabela 7, e Y é o identificador do arquivo.

Tabela 7: Tabela de referência dos identificadores de instrumentos.

0	=	violão
1	=	violino
2	=	piano
3	=	bumbo
4	=	chimbal
5	=	guitarra
6	=	saxofone
7	=	triângulo
8	=	flauta
9	=	prato

4.4 Interface Gráfica

A interface gráfica do sistema Ulises foi criada inteiramente no NetLogo e possui nove componentes, como podem ser vistos na Figura 19. Eles são:

1. Visualização de Mundo - é onde são exibidos os agentes especialistas e a troca de informação entre eles;
2. Botão de Setup - ele inicializa o sistema com as configurações iniciais, cria os agentes predefinidos e a fila de arquivos;
3. Seletor de Treino - ele serve para selecionar qual agente predefinido deverá ser treinado;
4. Botão de Treinamento - ele realiza o procedimento de treinamento para o agente selecionado no Seletor de Treino;
5. Fila - Exibe o estado atual da fila de solicitações das entradas a serem classificadas;
6. Seletor de Áudio - utilizado para selecionar quais entradas que deverão ser classificadas, respeitando a ordem de inserção.
7. Botão de Inserir Áudio - insere uma entrada na Fila de acordo com o áudio selecionado no Seletor de Áudio.
8. Botão Executar Um Passo - botão que executa o procedimento principal do sistema, dependendo do estado do ambiente este pode agir de três formas: atribuir uma entrada para um agente, solicitar a classificação de um agente e encaminhar entrada de um agente para o outro.

9. Botão Executar Sempre - similar ao botão Executar Um Passo, realizando os mesmos procedimentos, mas em loop infinito, parando apenas quando não existirem mais ações possíveis para a configuração do ambiente.

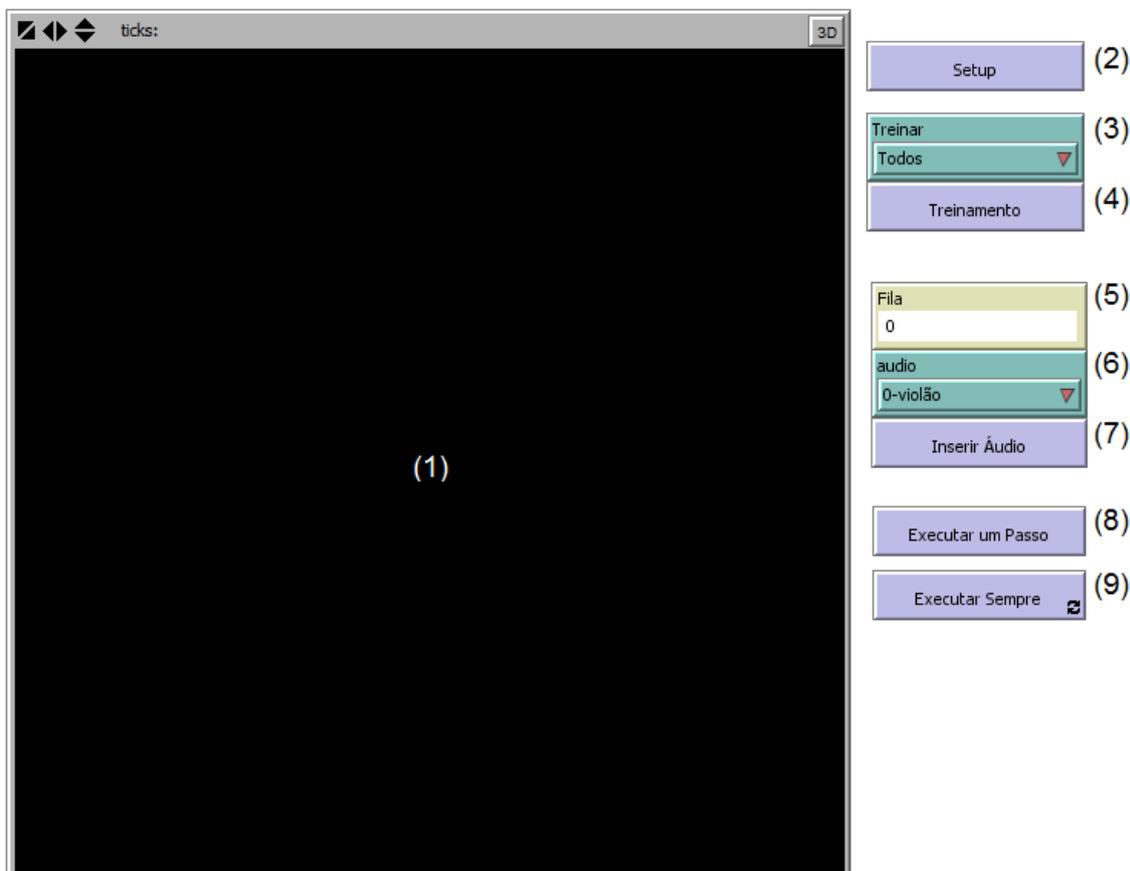


Figura 19: Componentes da interface gráfica do sistema Ulises.

Após executado o *Setup*, os agentes predefinidos são criados e exibidos na Visualização de Mundo, como na Figura 20. Para cada agente é atribuída uma figura, que representa o instrumento que ele é reconhecedor e um *label* indicando o nome do agente. Na fase atual do trabalho foram predefinidos 10 agentes especialistas: violão, violino, piano, bumbo, chimbau, guitarra, saxofone, triângulo, flauta e prato. Existe a possibilidade de inserir novos agentes predefinidos, basta uma pequena alteração no código de *Setup*, onde os agentes são criados. Maiores detalhes são abordados na Seção 4.5.

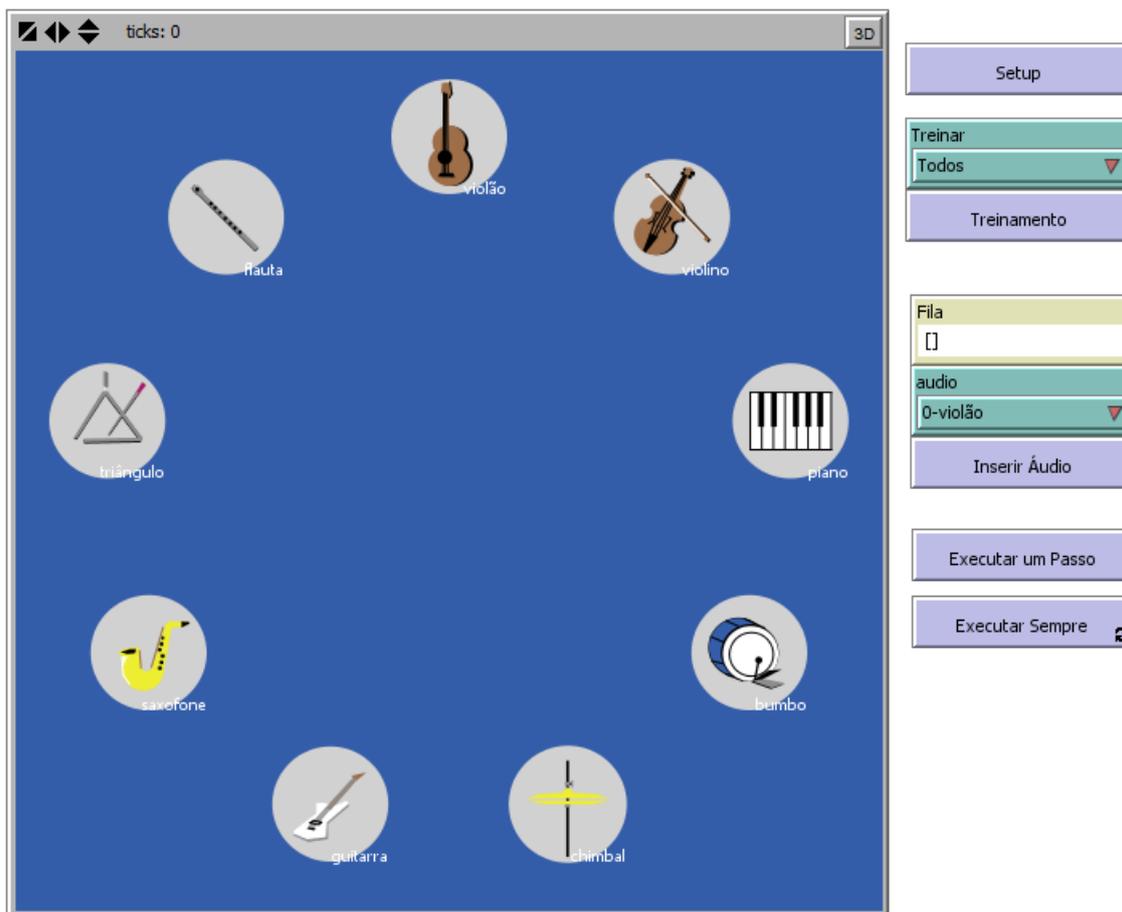


Figura 20: Visualização de Mundo do sistema após o *Setup*.

Cada agente possui um estado atual de execução, que no sistema foram representado por cores. Na Figura 21, estão representados todos os possíveis estados dos agentes. São eles:

1. Iniciado - cinza, este é o estado inicial de cada agente. Neste estado o agente não pode trabalhar, ele apenas é exibido no ambiente, mas permanece inativo.
2. Treinado - ciano, o agente atinge este estado após ter realizado ao menos um treinamento pela interface gráfica do sistema. Agora ele está pronto para trabalhar.
3. Atribuído - amarelo, quando o ambiente atribui uma entrada para o agente ele passa para o estado Atribuído. Também pode entrar neste estado caso receba uma entrada encaminhada por outro agente que teve como resposta do procedimento de classificação um valor abaixo do limiar estipulado empiricamente.
4. Classificado - verde, este estado dura apenas um único ciclo e serve de indicador de que o agente classificou corretamente a entrada atribuída. No próximo ciclo, ele volta para o estado Treinado.

5. Não Classificado - vermelho, este estado dura apenas um único ciclo e serve de indicador de que o agente não classificou corretamente a entrada atribuída. No próximo ciclo, ele volta para o estado Treinado e encaminha a sua entrada para um outro agente em estado Treinado.



Figura 21: Representação dos estados do agente especialista violão dentro da Visualização de Mundo.

4.5 Implementação

Como visto na Seção 4.3, foram utilizadas quatro ferramentas principais para a implementação do sistema Ulises. O código principal, presente no Anexo A, que implementa o sistema baseado em agentes e todas as funcionalidades do ambiente foi desenvolvido no NetLogo (WILENSKY, 1999), por ser uma ferramenta bem consolidada e adequada a pesquisa em sistemas multiagente. O processamento interno dos agentes para a classificação e treinamento foi implementado no MatLab (GUIDE, 1998), pela familiaridade com a ferramenta e seu grande desempenho em processamento de áudio. Para integrar estas duas ferramentas, utilizou-se uma extensão do NetLogo para trabalhar com o MatLab, denominada MatNet, que permite que procedimentos do MatLab sejam chamados pelo NetLogo e seu resultado recuperado dentro do ambiente do NetLogo. Por último, para realizar a manipulação do áudio, extração de características e utilizar métodos de classificação utilizou-se três *toolboxes* do MatLab: a MIRtoolbox (LARTILLOT; TOIVAINEN, 2007), utilizada na leitura de áudio e na extração de características musicais e as *toolboxes* oficiais de classificação e de processamento de sinais do MatLab, onde a primeira contém a implementação dos principais métodos de classificação e a segunda é requisito para utilização da MIRtoolbox. Para executar o sistema é preciso possuir todas estas ferramentas devidamente instaladas. Ele foi desenvolvido para o sistema operacional Windows, mas, apesar de não testado, é possível que funcione em sistemas Linux pois todas as ferramentas utilizadas possuem versões para esta plataforma.

A implementação do sistema Ulises segue o funcionamento descrito na Seção 4.2, tanto para o agente quanto o ambiente, desenvolvido em código Logo (Anexo A), se valendo da extensão MatNet para comunicação com o MatLab. Já os métodos de treinamento do agente e classificação foram desenvolvidos no MatLab, utilizando a *MIRtoolbox*. Ao todo são três funções implementadas no MatLab, que podem ser vistas no Anexo B, são elas:

Limpa função que exclui as variáveis do ambiente de trabalho do MatLab, é chamada no *setup* do NetLogo.

Treino esta função recebe dois parâmetros do NetLogo, o identificador do agente a ser treinado e o número de arquivos que ele usará no treinamento. Ela pode ser chamada no NetLogo individualmente para cada agente ou realizar o treinamento para todos os agentes do ambiente. Como resultado, ela cria uma matriz com os coeficientes do método de treinamento. Maiores detalhes são apresentados na Seção 4.5.1.

Classificação esta é a função principal no que diz respeito à funcionalidade do sistema. Ela recebe dois parâmetros: o identificador do agente que realizará a classificação e um identificador do arquivo de áudio correspondente. O retorno é o valor obtido com o método de classificação, caso seja maior que 0.6, considera o arquivo classificado, caso contrário ele encaminha para outro agente disponível. Maiores detalhes são apresentados na Seção 4.5.2.

Na atual fase de implementação, foram exploradas apenas as características do método de MFCC por sua ampla utilização na classificação de timbres como foi visto na Seção 3.1. Entretanto, existe uma vasta gama de métodos e características que poderiam ter sido abordadas, implementadas para todos os agentes ou para agentes específicos. Mais informações sobre outras abordagens são apresentadas na Seção 6.2.

4.5.1 O Método de Treinamento

O método de treinamento recebe três parâmetros do NetLogo, o identificador do agente a ser treinado, o nome do agente e o número de arquivos que ele usará no treinamento. Basicamente, esta função lê os arquivos de áudio do banco de dados, executa o método de MFCC se valendo da MIRtoolbox, e armazena 13 coeficientes para cada arquivo de áudio em uma variável global “base”, de dimensões $N \times 13 \times M$, onde N é o número de arquivos utilizados no treinamento e M diz respeito a uma terceira dimensão utilizada para cada agente especialista. Em outras palavras, se forem treinados 4 agentes utilizando 5 arquivos, a dimensão da matriz será $5 \times 13 \times 4$. Esta função pode ser chamada no NetLogo individualmente para cada agente ou realizar o treinamento para todos os agentes do ambiente.

Para exemplificação do procedimento de treinamento, vamos considerar a chamada da função de treino para o agente violão (associado ao índice 0 como visto na Tabela 7) com três arquivos para treino, como visto na chamada de função 3.

$$\text{treino}(0, \text{“violao”}, 3) \quad (3)$$

Por padrão, a função de treinamento considera os índices de arquivo de 1 até N , neste caso, onde $N = 3$, serão considerados os arquivos “0-1.wav”, “0-2.wav” e “0-3.wav”. Para cada uma dessas entradas o método de treinamento extrai o MFCC, resultando em 13 coeficientes para cada um dos arquivos, como pode ser visto na Figura 22.

Como resultado deste procedimento, a variável global “base”, matriz de dimensões $N \times 13 \times M$, assume, neste caso, as dimensões $3 \times 13 \times 1$. O valor de M incrementa toda vez que é feito o treinamento de um novo instrumento.

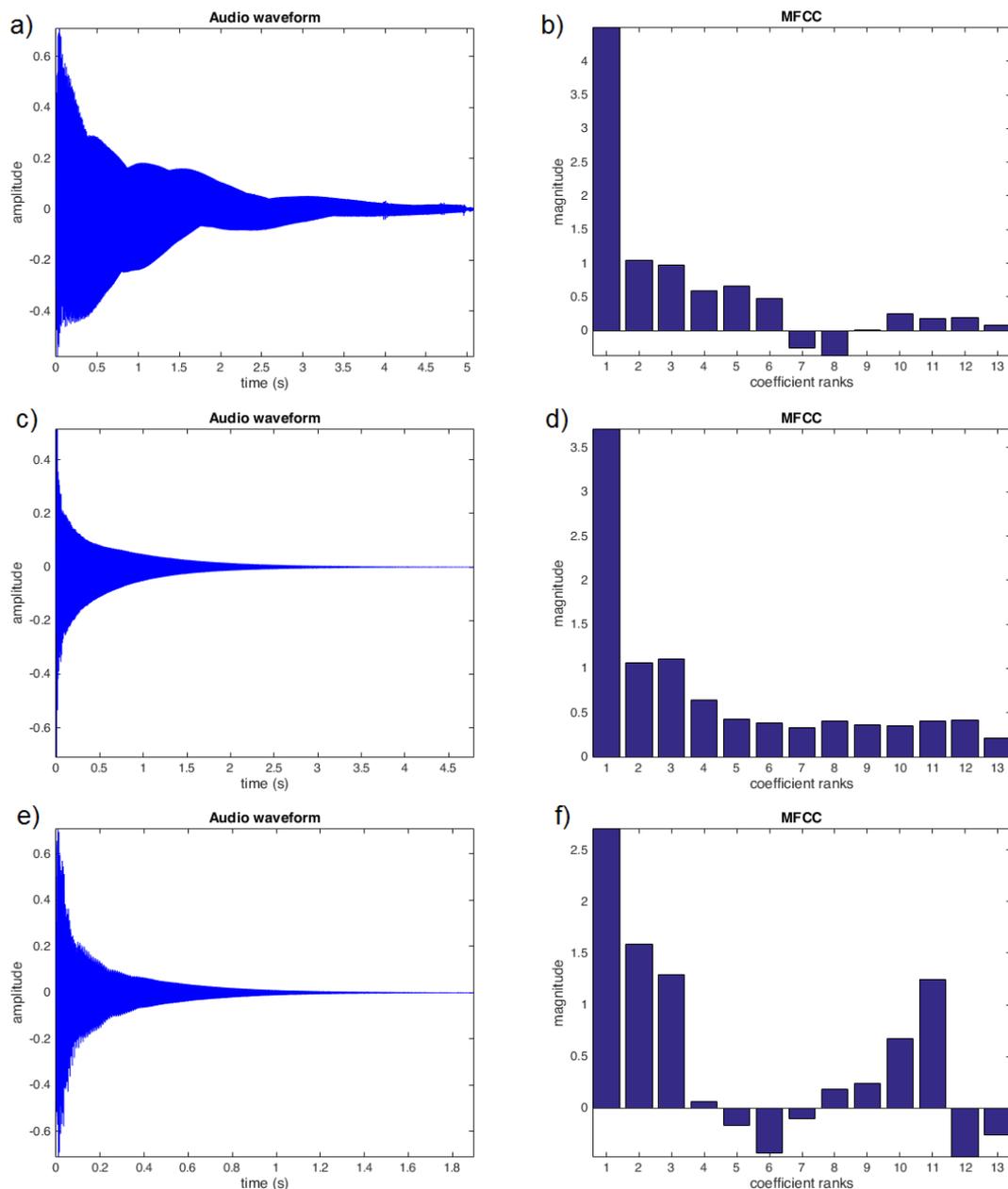


Figura 22: a) espectrograma do arquivo “0-1.wav”. b) MFCC do arquivo “0-1.wav”. c) espectrograma do arquivo “0-2.wav”. d) MFCC do arquivo “0-2.wav”. e) espectrograma do arquivo “0-3.wav”. f) MFCC do arquivo “0-3.wav”.

4.5.2 O Método de Classificação

O método de classificação recebe três parâmetros: o identificador do agente que realizará a classificação, o nome do agente especialista e um identificador do arquivo de áudio correspondente. O método de MFCC é aplicado para o áudio a ser analisado da mesma forma como é realizado no processo de treinamento, obtendo, também, 13 coeficientes. Posteriormente, é calculada a média entre o resultado do MFCC de todos os arquivos lidos no treinamento para ser utilizada como fonte de comparação com o arquivo que será

classificado. Logo após, é calculado o erro entre o arquivo e o resultado da média termo a termo, normalizados pela diferença entre o valor mais alto e o mais baixo do MFCC. Como resposta, a função retorna um valor acumulado do cálculo do erro normalizado, que dentro do NetLogo será testado: caso seja maior que 0.6, considera o arquivo classificado, caso contrário ele encaminha para outro agente disponível.

Para exemplificação do procedimento de classificação, vamos considerar a chamada da função de classificação para o agente violão (associado ao índice 0) com uma entrada de chimbau (associada ao índice 4) para ser testada, como visto na chamada de função 4.

$$\text{classificacao}(0, \text{"violao"}, 4) \quad (4)$$

A primeira etapa do procedimento de classificação consiste em extrair o MFCC do arquivo a ser testado. Por padrão, a entrada a ser comparada não faz parte do grupo de treinamento, sendo a esta atribuído o índice de arquivo zero. O espectrograma e o resultado do MFCC aplicado ao arquivo "4-0.wav", que contém a gravação para ser classificada do instrumento chimbau, são exibidos na Figura 23.

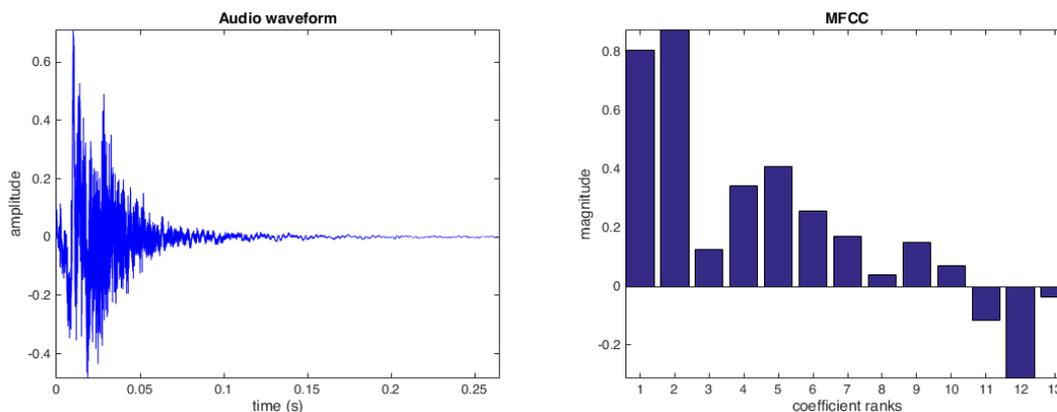


Figura 23: Espectrograma e MFCC do arquivo "4-0.wav".

A etapa seguinte consiste em calcular o módulo da diferença entre os valores da média dos coeficientes obtidos no treinamento com os valores do MFCC da entrada a ser classificada. Estes valores podem ser vistos na Figura 24.

Por último, as diferenças obtidas são somadas em uma variável acumuladora, representando o erro total obtido, neste caso 6,22. Este valor é normalizado em relação aos valores de máxima e mínima da média calculada, retornando apenas um único valor real.

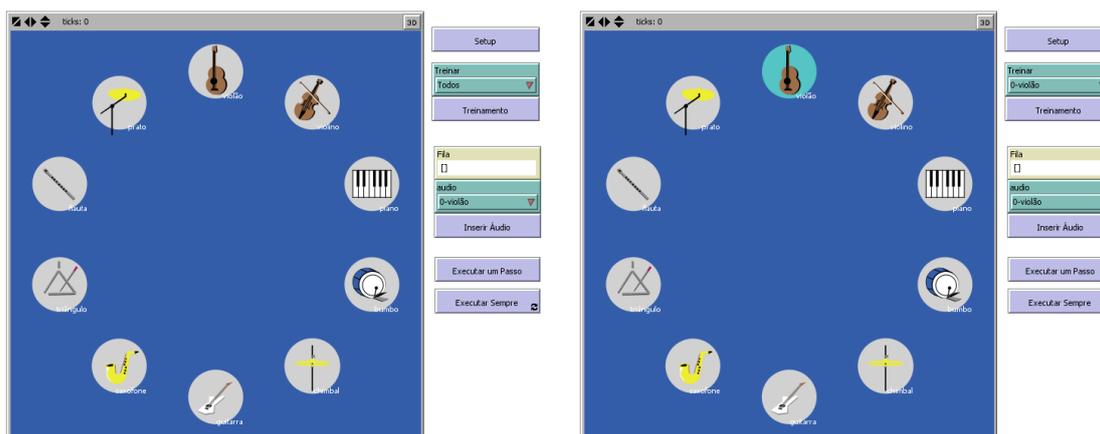
Finalmente ele é recuperado no NetLogo, caso seja maior que 0,6 considera-se a entrada corretamente classificada, caso contrário ela é encaminhada para um novo agente especialista, que repetirá este procedimento.

	1	2	3	4	5	6	7	8	9	10	11	12	13
0-1	4,49	1,04	0,97	0,59	0,65	0,47	-0,25	-0,36	0,00	0,25	0,18	0,19	0,07
0-2	3,70	1,06	1,10	0,63	0,42	0,38	0,33	0,40	0,36	0,34	0,40	0,41	0,21
0-3	2,69	1,59	1,29	0,06	-0,16	-0,43	-0,10	0,18	0,23	0,67	1,24	-0,46	-0,26
Média	3,63	1,23	1,12	0,43	0,30	0,13	-0,00	0,07	0,20	0,42	0,61	0,04	0,00
4-0	0,80	0,87	0,12	0,34	0,41	0,25	0,17	0,04	0,14	0,07	-0,11	-0,31	-0,03
Abs(Dif)	2,83	0,36	1,00	0,09	0,11	0,12	0,17	0,03	0,06	0,35	0,72	0,35	0,03

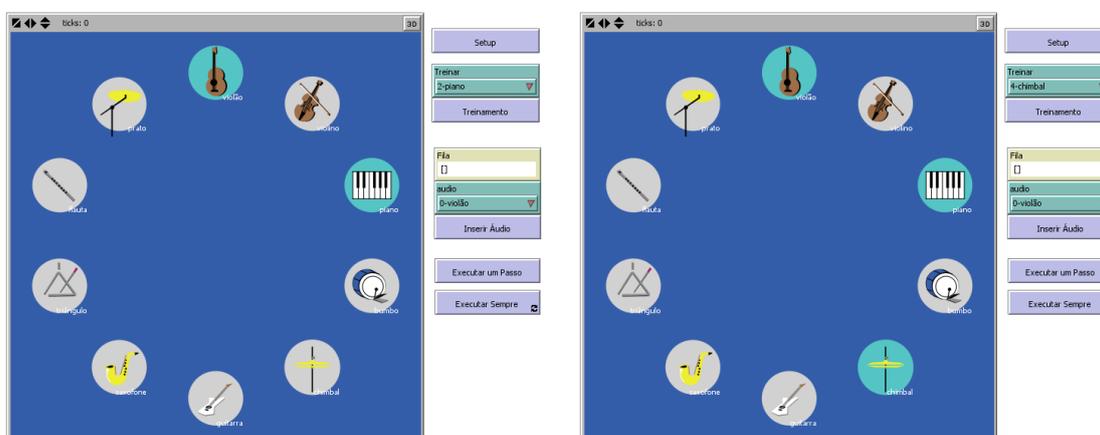
Figura 24: Nas linhas 0-1, 0-2 e 0-3 estão os resultados do MFCC realizado no treinamento do agente violão, a quarta linha apresenta a média entre os resultados do treinamento, a linha 4-0 exibe os valores extraídos do MFCC do arquivo a ser testado, e na última linha são exibidos os valores do módulo da diferença entre a média do especialista e o arquivo a ser testado.

4.6 Exemplo de Execução

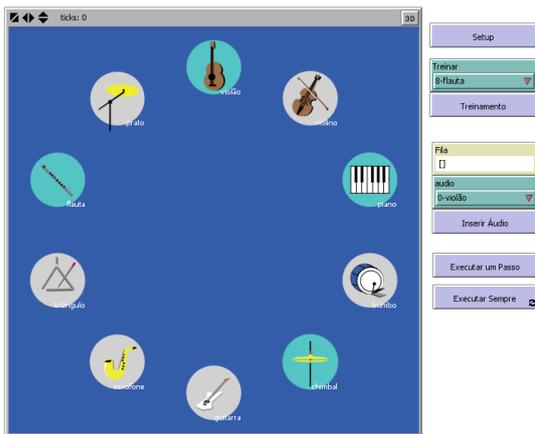
Nesta seção é apresentado um exemplo de execução passo a passo do sistema Ulises, descrevendo em detalhes todo o funcionamento do ambiente para um teste de caso.



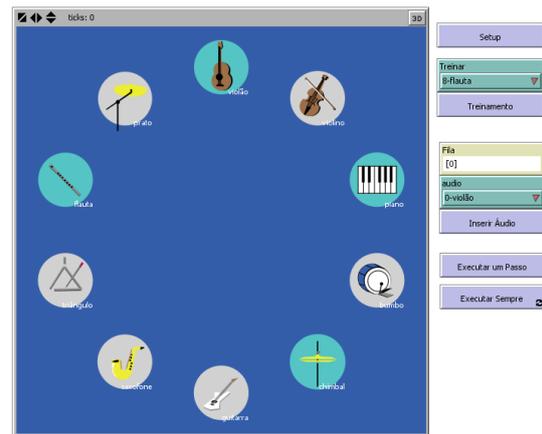
- 1) Ao clicar em *setup*, o sistema é iniciado, criando os agentes, inicializando a fila de treino e em seguida treinado e atualizando a visualização de mundo.
- 2) O agente violão é escolhido no seletor de treino e em seguida treinado.



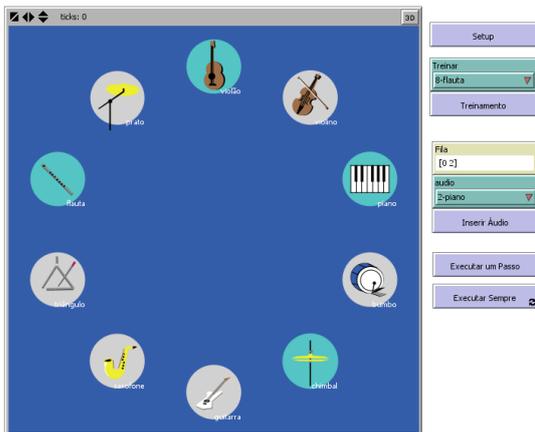
- 3) O agente piano é escolhido no seletor de treino e em seguida treinado.
- 4) O agente chimbau é escolhido no seletor de treino e em seguida treinado.



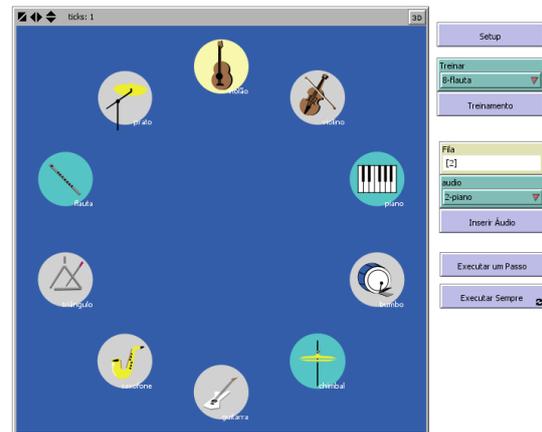
5) O agente flauta é escolhido no seletor de treino e em seguida treinado.



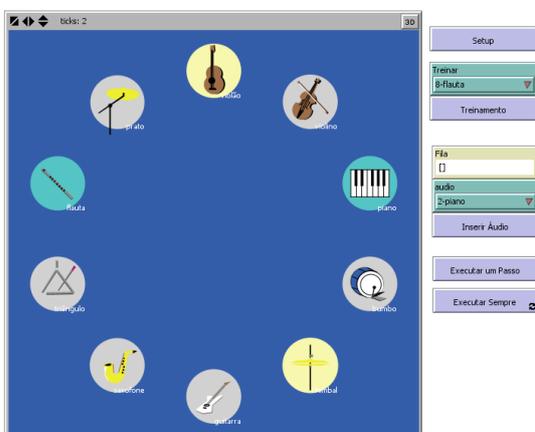
6) No seletor de áudio é escolhida a entrada correspondente ao violão e inserida na fila.



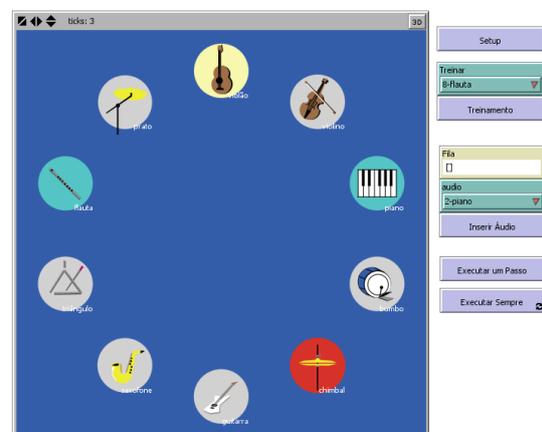
7) No seletor de áudio é escolhida a entrada correspondente ao piano e inserida na fila.



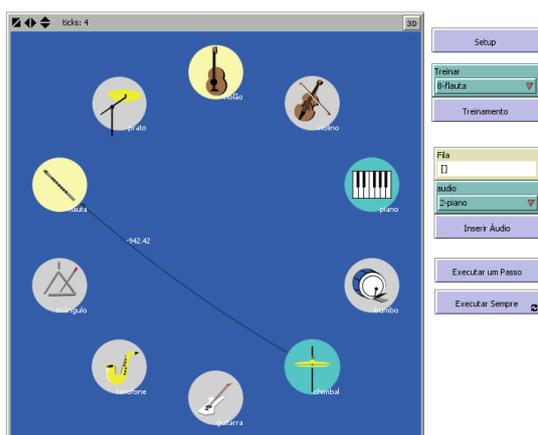
8) É executado um passo e o primeiro item da fila é atribuído ao agente violão.



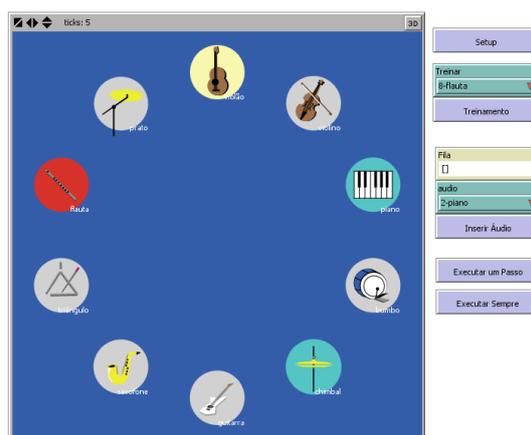
9) É executado um passo e o segundo item da fila inicial é atribuído ao agente chimbal.



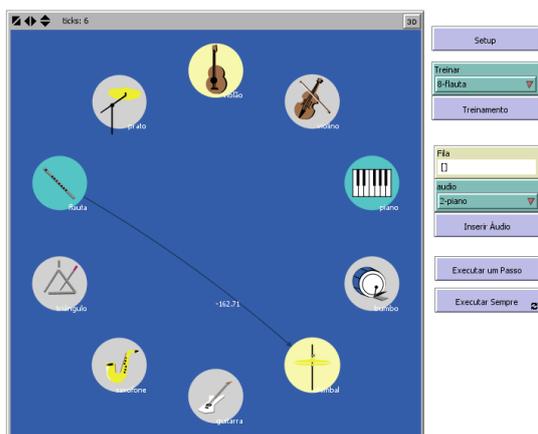
10) O agente chimbal faz um teste de classificação e conclui que ele não reconhece a entrada.



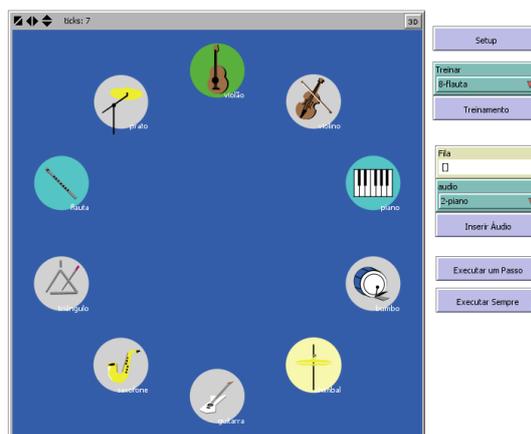
11) O agente chimbau encaminha a entrada que ele não classificou para o agente flauta.



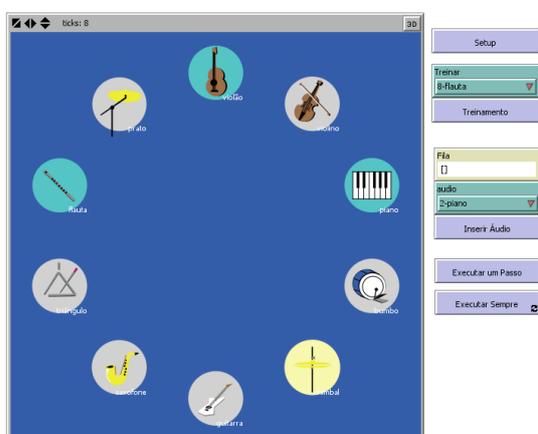
12) O agente flauta faz um teste de classificação e conclui que ele não reconhece a entrada.



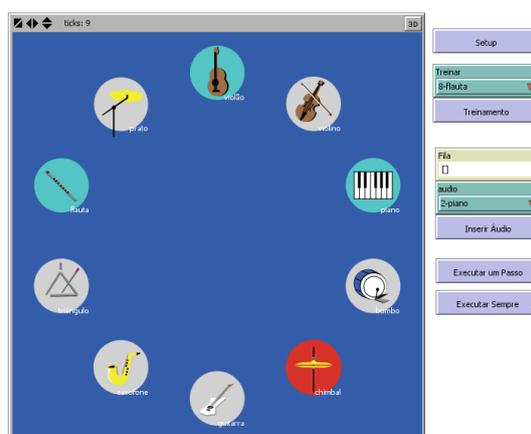
13) O agente flauta encaminha a entrada que ele não classificou para o agente chimbau.



14) O agente violão faz um teste de classificação e conclui que ele é o reconhecedor da entrada.



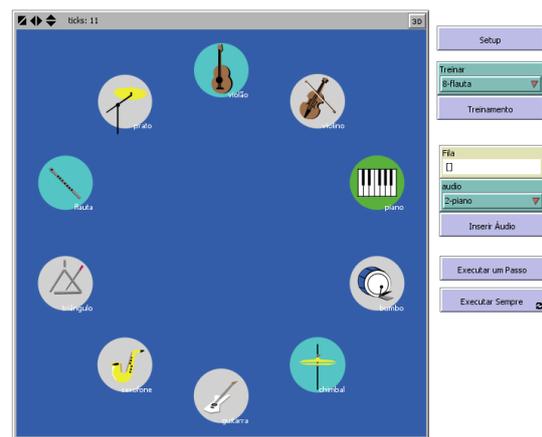
15) A entrada de índice zero, relativa ao instrumento violão foi corretamente classificada!



16) O agente chimbau faz um teste de classificação e conclui que ele não reconhece a entrada.



17) O agente chimbau encaminha a entrada que ele não classificou para o agente piano.



18) O agente piano faz um teste de classificação e conclui que ele é o reconhecedor da entrada.

Neste exemplo tem-se a correta classificação de duas entradas, mas é possível observar que o algoritmo não é ótimo. Na etapa 11, em que o agente chimbau não consegue classificar a entrada, ele encaminha para o agente flauta. Posteriormente, na etapa 13, o agente flauta reencaminha a entrada para o agente chimbau. Esta limitação ocorre devido o método de encaminhamento ser aleatório, onde o agente escolhe um outro que esteja disponível, independentemente se este já analisou a entrada. Soluções são discutidas na Seção 6.2.

Outra forma de visualizar a execução do sistema é a saída do centro de comandos do NetLogo, como observado na Figura 25.

```
(especialista 0): "Agente 0 treinado."  
(especialista 2): "Agente 2 treinado."  
(especialista 4): "Agente 4 treinado."  
(especialista 8): "Agente 8 treinado."  
(especialista 0): "Audio 0 atribuido ao agente 0"  
(especialista 4): "Audio 2 atribuido ao agente 4"  
(especialista 4): "Audio 2 não classificado pelo agente 4 com 9.42."  
(especialista 8): "Audio 2 encaminhado ao agente 8"  
(especialista 8): "Audio 2 não classificado pelo agente 8 com 1.63."  
(especialista 4): "Audio 2 encaminhado ao agente 4"  
(especialista 0): "Audio 0 classificado pelo agente 0 com acerto de 0.65."  
(especialista 4): "Audio 2 não classificado pelo agente 4 com 9.42."  
(especialista 2): "Audio 2 encaminhado ao agente 2"  
(especialista 2): "Audio 2 classificado pelo agente 2 com acerto de 0.76."
```

Figura 25: Log do centro de comandos do NetLogo, após exemplo de execução do sistema Ulises.

5 RESULTADOS

Com o desenvolvimento deste trabalho, foi possível implementar um sistema de classificação de timbres com ênfase na distribuição do modelo de classificação e do modelo de treinamento. Foram utilizadas as ferramentas NetLogo, MatLab, MatNet e MIR-toolbox para a implementação, onde foi possível validar o funcionamento da metodologia proposta. Para a maioria das execuções testadas, o sistema respondeu com grande quantidade de acertos, classificando corretamente timbres diversos, obtidos de diferentes bancos de dados, como descrito na Seção 4.3. O método de classificação implementado se provou bastante rápido, conforme a sua simplicidade, mas foram observados erros de classificação em instrumentos da mesma família, como o violão e a guitarra, possivelmente pela grande semelhança dos coeficientes de MFCC destes arquivos. Para a solução deste problema, uma provável solução seria utilizar um conjunto de características maior, que explorassem outros comportamentos no domínio do tempo e da frequência, como as características citadas na Seção 2.1.

Outra limitação diz respeito ao banco de dados, pois para alguns instrumentos, o número de arquivos utilizados foi baixo, o que ocasiona um treinamento que leva em consideração características de uma entrada que não necessariamente dizem respeito as características normais do timbre em questão. O banco foi formado por 80 arquivos que executam uma única nota, em diferentes alturas, por diversos instrumentos.

Os arquivos utilizados para treinamento podem ser vistos no Anexo C. Desta forma, para realizar o treinamento do agente chimbal com 3 arquivos, é chamada a função “treino(4,chimbal,3)”, que realizará o procedimento, conforme o Anexo B.

Outra limitação apresentada pelo sistema diz respeito ao método de encaminhamento dos agentes. Em alguns testes de execução, a mesma entrada chegou a passar quatro vezes pelo mesmo agente antes de alcançar o agente correto. Para solucionar este problema, poderia ser implementado um vetor associado a cada entrada, do tamanho do número de agentes no ambiente, servindo como um histórico por onde a entrada circulou, que deverá ser verificado antes de realizar uma atribuição.

6 CONCLUSÃO

Neste capítulo são apresentados os objetivos alcançados no desenvolvimento deste trabalho e posteriormente sugestões de trabalhos futuros.

6.1 Objetivos Alcançados

Com este trabalho se definiu e implementou um sistema baseado em agentes para a solução do problema de reconhecimento de timbres. Este trabalho representa o passo inicial em pesquisa de metodologias baseadas em agentes para este problema de grande complexidade musical, possibilitando novas pesquisas no campo da Computação Musical.

Abordagens deste tipo possuem características bastante relevantes, se tratando do seu uso computacional, como a escalabilidade: uma característica desejável em todo o sistema computacional, que indica sua capacidade de manipular uma porção crescente de trabalho de forma uniforme, ou estar preparado para crescer. A abordagem de agentes possibilita que novas entidades classificadoras sejam adicionadas sem a necessidade de refazer todo o sistema, ou toda a etapa de treinamento.

Outra característica importante é a paralelização: em ambientes que possuem mais de uma unidade de processamento, um código paralelizável normalmente apresenta uma vantagem em desempenho, pois, de um modo geral, a carga de trabalho que seria atribuída a apenas uma unidade é distribuída, tornando a sua execução mais rápida. Neste trabalho esta característica pode ser notada quando múltiplas entradas são adicionadas na fila. Elas serão atribuídas aos agentes treinados que poderão realizar os testes de classificação simultaneamente para $N-1$ entradas, onde N é o número de agentes disponíveis no sistema.

Desta forma, foi apresentado um método inovador para um problema complexo e de grande relevância científica, capaz de auxiliar profissionais da música e pesquisadores em geral.

6.2 Trabalhos Futuros

Este trabalho apresentou uma abordagem baseada em agentes para a solução do problema de classificação de timbres. Muitas são as possibilidades de trabalhos futuros,

tendo em vista que este se trata de um trabalho pioneiro em classificação de timbres de forma distribuída.

Uma possível aplicação que vem apresentando uma grande demanda por músicos das mais diferentes áreas é o aperfeiçoamento da transcrição musical¹. Muitos softwares realizam a transcrição musical com bastante precisão, mas ainda não apresentam resultados bons na identificação de instrumentos. O método mais utilizado para transcrição consiste numa simples divisão de alturas, baseando-se na frequência do som. Esta abordagem é bastante falha visto que existem instrumentos com registros amplos, variando de sons bem graves (baixas frequências) até sons bem agudos (altas frequências). Este é o caso por exemplo do piano. Um sistema de classificação de timbres ajudaria a separar, corretamente, as vozes dos instrumentos dentro de uma partitura, levando em conta suas características tímbricas e não apenas suas frequências. O sistema Ulises, aqui proposto, apresenta uma vantagem para uma abordagem como esta: por se tratar de um sistema distribuído, novos instrumentos não necessitariam de grandes atualizações de software, bastando adicionar um novo agente especialista quando necessário o reconhecimento de um instrumento até então não previsto.

Mais uma sugestão de trabalho futuro é o aperfeiçoamento dos métodos de classificação implementados. Foi utilizado apenas um método baseado na comparação com a média do MFCC entre várias entradas, mas existem diferentes características que auxiliam a aumentar a precisão da classificação. Um exemplo é a utilização de DTW (*Dynamic time warping*) (AL-NAYMAT; CHAWLA; TAHERI, 2009), uma técnica utilizada no reconhecimento de fala e que tem como diferencial comparar sequências temporais de forma independente de suas durações. Também pode-se citar o uso de UCR (RAKTHANMANON et al., 2012), que baseia-se no método de DTW, porém mostra-se mais eficiente em diversos aspectos.

Durante a elaboração do trabalho, uma questão foi levantada: seria mais eficiente possuir agentes com técnicas de classificação iguais ou que utilizassem características diferentes? Como visto na Seção 3.1, o *corte de frequência* é uma importante característica no reconhecimento de metais (LUCÉ; CLARK JR, 1967). Desta forma, visando o desempenho, um agente classificador de instrumentos de metais como o saxofone, poderia utilizar características de mais rápida extração. Estudar as possibilidades de sistemas com agentes mais heterogêneos também é uma proposta de trabalho futuro.

6.3 Aperfeiçoamentos

Se tratando do estado atual do software, muitos aperfeiçoamentos podem ser implementados. Como visto na Seção 4.6, uma limitação do sistema é que as entradas a serem classificadas ainda não possuem um histórico de agentes, o que ocasiona em

¹Transcrição musical é o ato de transcrever a música em uma forma documental, como uma partitura.

reclassificação pelos mesmos agentes. Este problema pode ser solucionado adicionando um vetor para cada entrada, que deverá ser verificado antes de realizar uma atribuição a fim de evitar processamento desnecessário. Ainda tratando da atribuição, foi visto no Capítulo 5 que instrumentos de uma mesma família possuem similaridade no resultado de classificação. Desta forma, um possível aperfeiçoamento seria a criação de um método de predição, baseado na resposta de classificação. Ele avaliaria qual dos agentes teria maior chance de ser o correto classificador, acertando o encaminhamento e poupando tempo de processamento.

Outra possibilidade seria a criação não supervisionada de agentes especialistas. Quando uma entrada circula pela classificação de todos os agentes do ambiente sem ser classificada, conjectura-se que se trata de um som novo, sem um agente especialista. Neste caso, se criaria um novo agente e usaria a própria entrada não classificada como arquivo de treinamento. Entretanto, apesar deste agente conseguir reconhecer corretamente um determinado timbre, ele não seria capaz de identificá-lo, pois a nomenclatura dos instrumentos é um conhecimento criado pelos humanos, o qual, até então, máquinas não saberiam adivinhar, necessitando de intervenção humana.

Por último, estudar o desempenho dos principais trabalhos de classificação de timbres e comparar a qualidade dos resultados com outros trabalhos relacionados, de forma a verificar se as características de um sistema baseado em agentes representam um ganho em relação a uma solução não distribuída, é uma tarefa importante de validação.

REFERÊNCIAS

AL-NAYMAT, G.; CHAWLA, S.; TAHERI, J. Sparsedtw: A novel approach to speed up dynamic time warping. In: EIGHTH AUSTRALASIAN DATA MINING CONFERENCE-VOLUME 101, 2009. **Proceedings...** [S.l.: s.n.], 2009. p.117–127.

ALVARES, L. O.; SICHMAN, J. S. Introdução aos Sistemas Multiagentes. In: JORNADA DE ATUALIZAÇÃO EM INFORMÁTICA. CONGRESSO DA SBC, 1997. **Anais...** [S.l.: s.n.], 1997. p.1–38.

AUCOUTURIER, J.-J.; PACHET, F.; SANDLER, M. The way it Sounds: timbre models for analysis and retrieval of music signals. **Multimedia, IEEE Transactions on**, [S.l.], v.7, n.6, p.1028–1035, 2005.

BEAUCHAMP, J. W. Data reduction and resynthesis of connected solo passages using frequency, amplitude, and “brightness” detection and the nonlinear synthesis technique. **The Journal of the Acoustical Society of America**, [S.l.], v.71, n.S1, p.S101–S101, 1982.

BEAUCHAMP, J. W. Synthesis by spectral amplitude and “Brightness” matching of analyzed musical instrument tones. **Journal of the Audio Engineering Society**, [S.l.], v.30, n.6, p.396–406, 1982.

BELLIFEMINE, F.; POGGI, A.; RIMASSA, G. JADE – A FIPA-compliant agent framework. In: INTERNATIONAL CONFERENCE AND EXHIBITION ON THE PRACTICAL APPLICATION OF INTELLIGENT AGENTS AND MULTI-AGENT TECHNOLOGY (PAAM’99), 4., 1999, London, UK. **Proceedings...** [S.l.: s.n.], 1999. p.97–108. <http://sharon.cselt.it/projects/jade/papers/PAAM.pdf>.

BELLO, J. P.; DAUDET, L.; ABDALLAH, S.; DUXBURY, C.; DAVIES, M.; SANDLER, M. B. A tutorial on onset detection in music signals. **Speech and Audio Processing, IEEE Transactions on**, [S.l.], v.13, n.5, p.1035–1047, 2005.

BENADE, A. H. **Fundamentals of Musical Acoustics**: Second. [S.l.]: Courier Corporation, 2012.

- BISMARCK, G. von. Timbre of steady sounds: A factorial investigation of its verbal attributes. **Acta Acustica united with Acustica**, [S.l.], v.30, n.3, p.146–159, 1974.
- BOWN, O. A Framework for Ecosystem-Based Generative Music. In: PROC OF 6TH SOUND AND MUSIC COMPUTING CONFERENCE (SMC2009), 2009. **Anais...** [S.l.: s.n.], 2009. p.195–200.
- Camurri, A.; Poli, G. D.; Rocchesso, D. A taxonomy for Sound and Music Computing. **Computer Music Journal**, [S.l.], v.19, p.4–5, 1995.
- CASEY, M.; VELTKAMP, R.; GOTO, M.; LEMAN, M.; RHODES, C.; SLANEY, M. et al. Content-based music information retrieval: Current directions and future challenges. **Proceedings of the IEEE**, [S.l.], v.96, n.4, p.668–696, 2008.
- CHAU, C.-j.; WU, B.; HORNER, A. Timbre Features and Music Emotion in Plucked String, Mallet Percussion, and Keyboard Tones. In: INT. COMP. MUSIC CONF.(ICMC), 40., 2014. **Proceedings...** [S.l.: s.n.], 2014. p.982–989.
- D. BISIG, M. N.; FLURY, J. Interactive Swarm Orchestra-a Generic Programming Environment for Swarm Based Computer Music. In: INTERNATIONAL COMPUTER MUSIC CONFERENCE, 2008. **Proceedings...** [S.l.: s.n.], 2008.
- DAVIS, S. B.; MERMELSTEIN, P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. **Acoustics, Speech and Signal Processing, IEEE Transactions on**, [S.l.], v.28, n.4, p.357–366, 1980.
- DEVI, J. S.; SRINIVAS, Y.; KRISHNA, N. M. A Study: Analysis of Music Features for Musical Instrument Recognition and Music Similarity Search. **IJCSI**, [S.l.], 2012.
- ERONEN, A. et al. Automatic musical instrument recognition. **Mémoire de DEA, Tampere University of Technology**, [S.l.], p.178, 2001.
- GREY, J. M. Multidimensional perceptual scaling of musical timbres. **The Journal of the Acoustical Society of America**, [S.l.], v.61, n.5, p.1270–1277, 1977.
- GUIDE, M. U. The mathworks. **Inc., Natick, MA**, [S.l.], v.5, p.333, 1998.
- HANDEL, S. Timbre perception and auditory object identification. **Hearing**, [S.l.], p.425–461, 1995.
- HELMHOLTZ, H. L.; ELLIS, A. J. **On the Sensations of Tone as a Physiological Basis for the Theory of Music**. [S.l.]: Cambridge University Press, 1954.
- HELMHOLTZ, H. L.; ELLIS, A. J. **On the Sensations of Tone as a Physiological Basis for the Theory of Music**. [S.l.]: Cambridge University Press, 2009.

HERRERA-BOYER, P.; PEETERS, G.; DUBNOV, S. Automatic classification of musical instrument sounds. **Journal of New Music Research**, [S.l.], v.32, n.1, p.3–21, 2003.

JENSEN, K. **Timbre models of musical sounds**. [S.l.]: Department of Computer Science, University of Copenhagen, 1999.

KITAHARA, T. Computational musical instrument recognition and its application to content-based music information retrieval. **Unpublished PhD Thesis, Kyoto University, Kyoto, Japan. Retrieved**, [S.l.], v.10, n.31, p.07, 2007.

KLAPURI, A. **Signal processing methods for the automatic transcription of music**. [S.l.]: Tampere University of Technology Finland, 2004.

KLINGBEIL, M. K. **Spectral Analysis, Editing, and Resynthesis: Methods and Applications**. 2009. Tese (Doutorado em Ciência da Computação) — Columbia University.

KRIMPHOFF, J.; MCADAMS, S.; WINSBERG, S. Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique. **Le Journal de Physique IV**, [S.l.], v.4, n.C5, p.C5–625, 1994.

LAPP, D. R. **The physics of music and musical instruments**. [S.l.]: Wright Center for Innovative Science Education, Tufts University, 2003.

LARTILLOT, O.; TOIVIAINEN, P. A Matlab toolbox for musical feature extraction from audio. In: INTERNATIONAL CONFERENCE ON DIGITAL AUDIO EFFECTS, 2007. **Anais...** [S.l.: s.n.], 2007. p.237–244.

LARTILLOT, O.; TOIVIAINEN, P.; EEROLA, T. **MIRtoolbox 1.6 User's Manual**. [S.l.: s.n.], 2014. v.1.6.

LICHTE, W. H. Attributes of complex tones. **Journal of Experimental Psychology**, [S.l.], v.28, n.6, p.455, 1941.

LIU, D.; LU, L.; ZHANG, H. Automatic mood detection from acoustic music data. In: ISMIR, 2003. **Anais...** [S.l.: s.n.], 2003.

LUCE, D.; CLARK JR, M. Physical Correlates of Brass-Instrument Tones. **The Journal of the Acoustical Society of America**, [S.l.], v.42, n.6, p.1232–1243, 1967.

MARTIN, K. D.; KIM, Y. E. 2pMU9. Musical instrument identification: A pattern-recognition approach. In: PRESENTED AT THE 136TH MEETING OF THE ACOUSTICAL SOCIETY OF AMERICA, 1998. **Anais...** [S.l.: s.n.], 1998.

MATHWORKS. **MatLab**.

MURRAY-RUST, D.; SMAILL, A.; EDWARDS, M. MAMA: An Architecture for Interactive Musical Agents. In: ECAI 2006, 17TH EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, AUGUST 29 - SEPTEMBER 1, 2006, RIVA DEL GARDA, ITALY, INCLUDING PRESTIGIOUS APPLICATIONS OF INTELLIGENT SYSTEMS (PAIS 2006), PROCEEDINGS, 2006. **Anais...** IOS Press, 2006. p.36–40. (Frontiers in Artificial Intelligence and Applications, v.141).

NETLOGO. **NetLogo Docs**. [S.l.]: NetLogo, 2016.

NEUMAN, I.; OKPALA, C.; BONEZZI, C. E. Mapping Motion to Timbre: Orientation, FM Synthesis and Spectral Filtering. In: INTERNATIONAL COMPUTER MUSIC CONFERENCE JOINT WITH SOUND AND MUSIC COMPUTING CONFERENCE, 2014. **Anais...** [S.l.: s.n.], 2014.

NORDQVIST, P. **Sound classification in hearing instruments**. 2004. Tese (Doutorado em Ciência da Computação) — KTH-S3.

PERCHY, S.; SARRIA, G. Dissonances: Brief Description and its Computational Representation in the RTCC Calculus. In: PROC OF 6TH SOUND AND MUSIC COMPUTING CONFERENCE (SMC2009), 2009. **Anais...** [S.l.: s.n.], 2009. p.53–58.

PLOMP, R.; LEVELT, W. J. Tonal consonance and critical bandwidth. **The journal of the Acoustical Society of America**, [S.l.], v.38, n.4, p.548–560, 1965.

POHLE, T.; PAMPALK, E.; WIDMER, G. Evaluation of frequently used audio features for classification of music into perceptual categories. In: FOURTH INTERNATIONAL WORKSHOP ON CONTENT-BASED MULTIMEDIA INDEXING, 2005. **Proceedings...** [S.l.: s.n.], 2005. v.162.

RAKTHANMANON, T.; CAMPANA, B.; MUEEN, A.; BATISTA, G.; WESTOVER, B.; ZHU, Q.; ZAKARIA, J.; KEOGH, E. Searching and mining trillions of time series subsequences under dynamic time warping. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 18., 2012. **Proceedings...** [S.l.: s.n.], 2012. p.262–270.

RISSET, J.-C.; WESSEL, D. L. Exploration of timbre by analysis and synthesis. **The psychology of music**, [S.l.], v.28, 1982.

ROADS, C. **The computer music tutorial**. [S.l.]: MIT press, 1996.

RUMSEY, F.; MCCORMICK, T. **Sound and recording: an introduction**. [S.l.]: CRC Press, 2012.

SAMPAIO, P. A.; RAMALHO, G.; TEDESCO, P. A. CinBalada: Multiagent Rhythm Factory. **J. Braz. Comp. Soc**, [S.l.], v.14, n.3, p.31–49, 2008.

SAMPAIO, P.; TEDESCO, P.; RAMALHO, G. Cinbalada: um laboratório multiagente de geração de ritmos de percussão. In: X BRAZILIAN SYMPOSIUM ON COMPUTER MUSIC, 2005. **Proceedings...** [S.l.: s.n.], 2005.

SCHOENBERG, A. **Theory of harmony**. [S.l.]: Univ of California Press, 1978.

SEARLE, J. R. **Speech acts**: An essay in the philosophy of language. [S.l.]: Cambridge university press, 1969. v.626.

SORAGHAN, S. Animating Timbre-A User Study. **The 2014 IEEE International Conference on Systems, Man, and Cybernetics**, [S.l.], 2014.

STEVENS, S. S.; VOLKMANN, J.; NEWMAN, E. B. A scale for the measurement of the psychological magnitude pitch. **The Journal of the Acoustical Society of America**, [S.l.], v.8, n.3, p.185–190, 1937.

STRONG, W. J. **Synthesis and recognition characteristics of wind instrument tones**. [S.l.]: Massachusetts Institute of Technology, 1963.

THOMAZ, L. F. A framework for implementing musical multiagent systems. **6th Sound and Music Computing Conference**, [S.l.], p.119–124, 2009.

THOMAZ, L. F. **Um arcabouço para construção de sistemas multiagentes musicais**. 2011. Tese de Doutorado — Universidade de São Paulo (USP).

THOMAZ, L. F. Ensemble: Implementing a Musical Multiagent System Framework. **8th Sound and Music Computing Conference**, [S.l.], p.198–205, 2011.

TZANETAKIS, G.; COOK, P. Musical genre classification of audio signals. **Speech and Audio Processing, IEEE transactions on**, [S.l.], v.10, n.5, p.293–302, 2002.

WATKINSON, J. **The MPEG Handbook**: MPEG-1, MPEG-2, MPEG-4. [S.l.]: Taylor & Francis, 2004.

WEISS, G. (Ed.). **Multiagent Systems**: A Modern Approach to Distributed Artificial Intelligence. Cambridge, MA, USA: MIT Press, 1999.

WIDMER, G.; ROCCHESO, D.; VÄLIMÄKI, V.; ERKUT, C.; GOUYON, F.; PRESSNITZER, D.; PENTTINEN, H.; POLOTTI, P.; VOLPE, G. Sound and music computing: Research trends and some key issues. **Journal of New Music Research**, [S.l.], v.36, n.3, p.169–184, 2007.

WIKIPÉDIA. **Timbre** — **Wikipédia, a enciclopédia livre**. [Online; accessed 26-agosto-2015].

WILENSKY, U. **NetLogo**. [S.l.]: Northwestern University, 1999.

WOOLDRIDGE, M. **An introduction to multiagent systems**. [S.l.]: John Wiley & Sons, 2009.

WU, B.; HORNER, A.; LEE, C. Musical Timbre and Emotion: The Identification of Salient Timbral Features in Sustained Musical Instrument Tones Equalized in Attack Time and Spectral Centroid. In: INT. COMP. MUSIC CONF.(ICMC), 40., 2014. **Proceedings...** [S.l.: s.n.], 2014. p.928–934.

ZHANG, X.; RAS, Z. W. Analysis of sound features for music timbre recognition. In: MULTIMEDIA AND UBIQUITOUS ENGINEERING, 2007. MUE'07. INTERNATIONAL CONFERENCE ON, 2007. **Anais...** [S.l.: s.n.], 2007. p.3–8.

ANEXO A CÓDIGO-FONTE NETLOGO - ULISES

```
;-----;
; Ulises: Sistema Baseado em Agentes Para Classificação de Timbres ;
;      Universidade Federal do Rio Grande - FURG ;
;
;      Eduardo Porto Teixeira, Diana Francisca Adamatti ;
;      {eduardoporteteixeira, dianaada}@gmail.com ;
;-----;

;Inclui a extensão do MatLab
extensions [ matlab ]

;Declaração das variáveis globais
globals [
  saída
  fila n_atribuidos
  n_arquivos
  encaminha
  classifica
  temp ]

;Declaração dos agentes
breed [ especialistas especialista ]
directed-link-breed [ comunicas comunica ]

;Variáveis dos agentes especialistas
especialistas-own [ estado id_arquivo ]

;-----;
; Colori os agentes conforme o estado ;
;-----;
to colorir
  if estado = "treinado"      [ set color 85 ]
  if estado = "iniciado"     [ set color 8 ]
  if estado = "atribuido"    [ set color 48 ]
  if estado = "classificado" [ set color 55 ]
  if estado = "nao_classificado" [ set color 15 ]
end
```

```

;-----;
; Setup
;-----;
to setup

;Limpar o ambiente
clear-all
matlab:eval "limpa()"

;Inicializa as variáveis globais
set encaminha false
set n_arquivos 3
set fila []

;Pinta o fundo do ambiente
ask patches [ set pcolor blue ]

;define os links curvos
set-default-shape comunicas "curved link"

;Cria agentes especialistas
create-ordered-especialistas 10 [
  set estado "iniciado"
  set size 6
  fd 18
  set id_arquivo 0
  colorir
]

;Personaliza os agentes especialistas
ask turtle 0 [ set shape "guitar"          set label "violão" ]
ask turtle 1 [ set shape "violin"         set label "violino" ]
ask turtle 2 [ set shape "piano"          set label "piano" ]
ask turtle 3 [ set shape "bass drum"      set label "bumbo" ]
ask turtle 4 [ set shape "hi-hat"        set label "chimbal" ]
ask turtle 5 [ set shape "electric guitar" set label "guitarra" ]
ask turtle 6 [ set shape "sax"           set label "saxofone" ]
ask turtle 7 [ set shape "triangle drum" set label "triângulo" ]
ask turtle 8 [ set shape "flute"         set label "flauta" ]
ask turtle 9 [ set shape "cymbal"        set label "prato" ]

;Reseta o contador de passos
reset-ticks
end

;-----;
; Insere novo arquivo na fila
;-----;
to inserir
  set fila lput read-from-string substring audio 0 1 fila
end

```

```

-----
; Realiza o treinamento
-----
to treinamento
  ifelse Treinar = "Todos" [
    ask especialistas [
      matlab:send-double "id" who
      matlab:send-string "especialista" label
      matlab:send-double "n_arquivos" n_arquivos
      matlab:eval "treino(id, especialista, n_arquivos)"
      show (word "Agente " who " treinado.")
      set estado "treinado"
      colorir
    ]
  ]
  [
    ask turtle read-from-string substring Treinar 0 1 [
      matlab:send-double "id" who
      matlab:send-string "especialista" label
      matlab:send-double "n_arquivos" n_arquivos
      matlab:eval "treino(id, especialista, n_arquivos)"
      show (word "Agente " who " treinado.")
      set estado "treinado"
      colorir
    ]
  ]
end

-----
; Realiza um passo
-----
to step
  ifelse encaminha = true
  [
    encaminhar
    set encaminha false
  ]
  [
    ask links [ die ]
    ifelse classifica = true
    [
      ask especialistas with [estado = "classificado"]
      [
        set estado "treinado"
        colorir
      ]
      set classifica false
    ]
    [
      ifelse length fila > 0 and ((count especialistas with ...
        ... [estado = "atribuido"]) + 1 < (count ...
        ... especialistas with [estado = "treinado"]))
      [
        atribuir
      ]
      [
        if (length fila > 0) or (count especialistas with ...
          ... [estado = "atribuido"] > 0)
          [
            classificar
          ]
        ]
      ]
    ]
  ]
  tick
end

```

```

-----
; Atribui um arquivo a um especialista
-----
to atribuir
  ask one-of especialistas with [ estado = "treinado" ] [
    set estado "atribuido"
    colorir
    set id_arquivo first fila
    set fila but-first fila
    show (word "Audio " id_arquivo " atribuido ao agente " who)
  ]
end

-----
; Realiza um teste de classificação
-----
to classificar
  ask one-of especialistas with [ estado = "atribuido" ] [
    matlab:send-double "id" who
    matlab:send-string "especialista" label
    matlab:send-double "entrada" id_arquivo
    matlab:eval "classificacao(id, especialista, entrada)"
    set saida matlab:get-double "saida_classificacao"
    ifelse saida > 0.6
      [
        set estado "classificado"
        colorir
        show (word "Audio " id_arquivo " classificado pelo agente ...
          ... " who " com " precision (saida * 100) 2 "%.")
        set classifica true
      ]
      [
        set estado "nao_classificado"
        colorir
        show (word "Audio " id_arquivo " não classificado pelo ...
          ... agente " who " com " precision (saida * 100) 2 "%.")
        set encaminha true
      ]
    ]
end

-----
; Encaminha o arquivo não classificado para outro agente
-----
to encaminhar
  ask one-of especialistas with [ estado = "nao_classificado" ] [
    set estado "treinado"
    colorir
    create-comunica-to one-of other especialistas with ...
      ... [estado = "treinado" ] [
        set color 103
        set label precision (saida * 100) 2
      ]
    set temp id_arquivo
    ask one-of out-link-neighbors with [estado = "treinado" ] [
      set estado "atribuido"
      set id_arquivo temp
      colorir
      show (word "Audio " id_arquivo " encaminhado ao ...
        ... agente " who) ]
    ]
end

```

ANEXO B CÓDIGO-FONTE MATLAB - ULISES

```
Editor - C:\Program Files\NetLogo 5.3.1\app\limpa.m
limpa.m x +
1 function [] = limpa()
2     clear all;
3 end

Editor - C:\Program Files\NetLogo 5.3.1\app\treino.m
treino.m x +
1 function [] = treino(agente_id, ~, n_arquivos)
2     global base
3     A = num2str(agente_id);
4     for i = 1:n_arquivos
5         base(i, :, agente_id+1) = mirgetdata(mirmfcc(miraudio ...
6             (strcat('banco\', A, '-', num2str(i), '.wav')), 'Rank', 1:13));
7     end
8     assignin('caller', 'banco', base);
9 end

Editor - C:\Program Files\NetLogo 5.3.1\app\classificacao.m
classificacao.m x +
1 function [] = classificacao(agente_id, ~, id_arquivo)
2     global base
3     FILE = num2str(id_arquivo);
4     MFCC = mirgetdata(mirmfcc(miraudio(strcat('banco\', FILE, '-', ...
5         num2str(0), '.wav')), 'Rank', 1:13));
6     TEMP = mean(base(:, :, agente_id+1));
7     TAM = max(TEMP) - min(TEMP);
8     ACC = 0;
9     for i = 1:13
10        ERRO = abs(MFCC(i) - TEMP(i)) / TAM;
11        ACC = ACC + ERRO;
12    end
13    RES = (1 - ACC / TAM);
14    assignin('caller', 'saida_classificacao', RES);
15 end
16
```

ANEXO C BANCO DE DADOS

