

Universidade Estadual de Campinas (Unicamp)

Faculdade de Engenharia Elétrica e de Computação (FEEC)

Departamento de Engenharia de Computação e Automação Industrial (DCA)

Laboratório de Bioinformática e Computação Bio-Inspirada (LBiC)

Síntese Sonora Auto-Organizável Através da Aplicação de Algoritmos Bio-Inspirados

Autor: Marcelo Freitas Caetano

**Orientador: Prof. Dr. Fernando José Von Zuben
LBiC/DCA/FEEC/Unicamp**

**Co-Orientador: Prof. Dr. Jônatas Manzolli
NICS/Unicamp | IA/Unicamp**

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

Banca Examinadora

Fernando José Von Zuben, Dr. DCA/FEEC/Unicamp
Rafael Santos Mendes, PhD.DCA/FEEC/Unicamp
Romis Ribeiro de Faissol Attux, Dr.DECOM/FEEC/Unicamp
Artemis Maria Francelin Sanchez Moroni, Dra.Cenpra/SP

Abril/2006

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

C116s	Caetano, Marcelo Freitas Síntese sonora auto-organizável através da aplicação de algoritmos bio-inspirados / Marcelo Freitas Caetano. -- Campinas, SP: [s.n.], 2006. Orientadores: Fernando José Von Zuben, Jônatas Manzolli Dissertação (Mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação. 1. Música por computador. 2. Música eletrônica. 3. Inteligência artificial. 4. Aprendizado do computador. 5. Sistemas auto-organizadores. 6. Redes neurais (Computação). I. Von Zuben, Fernando José. II. Manzolli, Jônatas. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.
-------	--

Título em Inglês: Self-organizing sound synthesis by means of the application of bio-inspired algorithms

Palavras-chave em Inglês: Computer music, Sound synthesis, Self-organization, Bio-inspired algorithms, Artificial neural networks, Artificial immune systems

Área de concentração: Engenharia de Computação

Titulação: Mestre em Engenharia Elétrica

Banca examinadora: Rafael Santos Mendes, Romis Ribeiro de Faissol Attux, Artemis Maria Francelin Sanchez Moroni

Data da defesa: 20/04/2006

RESUMO

Não há limitações teóricas para o uso do computador como fonte de sons musicais. O computador digital permite a produção de qualquer som concebível *dada a seqüência correta de números* (amostras digitais). No entanto, produzir uma dada seqüência de números que corresponda a um som musical que possua determinadas características perceptivas desejadas é uma tarefa de difícil resolução. Grande parte dos métodos e sistemas de síntese sonora digital utiliza modelos e/ou incorpora técnicas que não levam em conta a natureza dinâmica dos sons musicais ou que não foram originalmente desenvolvidas para manipulação musical. Neste trabalho, é apresentada uma abordagem populacional para síntese sonora no domínio temporal. Foi estudado um espaço sonoro e um conjunto de atratores, isto é, um conjunto de formas de onda com qualidades sonoras desejadas e definidas *a priori*, e foi possível obter sons que possuem características associadas a um ou mais atratores, representando variantes dos mesmos. Este método de síntese de sons musicais pode ser interpretado como um processo de busca no espaço vetorial que contém todas as possibilidades sonoras decorrentes da representação adotada, e tem por objetivo a criação de formas de onda digitalizadas com características emergentes e potencial para serem utilizadas em diversas aplicações musicais. Os resultados representam variantes e/ou possuem intersecções das características próprias dos atratores, responsáveis por indicar as regiões de interesse do espaço de busca. A proposta de pesquisa envolveu a utilização de algoritmos bio-inspirados – os quais expressam propriedades de sistemas auto-organizados e adaptativos – como definidores de processos de geração e estruturação dos elementos sonoros, entendidos aqui como problemas de otimização. A auto-organização e os mecanismos de manutenção de diversidade e de adaptação, intrínsecos aos sistemas bio-inspirados, fundamentam a proposta no sentido de viabilizarem a emergência temporal de estruturas estáveis sem um elemento organizador externo.

ABSTRACT

There are no theoretical limitations to the use of the computer as a source of musical sounds. The digital computer allows for the production of any conceivable sound *given the correct sequence of numbers* (digital samples). Nevertheless, producing the correct sequence of numbers that correspond to a musical sound expressing predefined perceptual characteristics is a very difficult task. Most sound synthesis methods and systems utilize models and/or incorporate techniques which do not take into account the dynamic nature of musical sounds or were not originally developed for the manipulation of musical tones. In this work we are proposing a populational sound synthesis approach in the time domain. A soundspace and a set of attractors, i.e. waveforms containing *a priori* desired features or qualities, and a population of agents communicating by means of local interaction were studied, and it was possible to attain sounds which share some qualities from more than one of the attractors, resulting exclusively from low-level rules followed by these agents. This sound synthesis method can be regarded as a search in the vector space that contains all the possible sounds resulting from the adopted representation, and its objective is to synthesize digital waveforms that possess emergent properties and the potential to be used in musical applications. The resulting sounds are variants or hybrids that share some of the intrinsic features of the attractors, which are responsible for indicating the regions of interest in the search space. This proposal involved the use of bio-inspired algorithms, which express features of adaptive, self-organizing systems, as definers of generating and structuring processes of sound elements, regarded herein as optimization processes. Self-organization and diversity maintenance and adaptation mechanisms, intrinsic to bio-inspired systems, lay the foundations of this proposal so as to make viable the temporal emergence of stable structures without an external organizing element.

A
todos que estiveram comigo ao longo do caminho.

V

"Music is sounds, sounds around us whether we're in or out of concert halls"
"Música é sons, sons à nossa volta, quer estejamos dentro ou fora das salas de concerto"
John Cage

Agradecimentos

Ao Prof. Jônatas Manzolli, amigo e mentor, que já me ensinou muito e parece nunca se cansar de me ensinar a cada dia, responsável por eu ter seguido esta linha de pesquisa, por seu constante incentivo, orientação e apoio;

Ao incansável Prof. Fernando José Von Zuben, pela dedicação sem limites à pesquisa e à ciência, por todo o conhecimento que já adquiri com ele, pela enorme confiança que depositou em mim e pelo apoio incondicional que me deu sempre;

Ao Prof. Rafael Santos Mendes, pelo seu interesse e questionamento inestimáveis;

À Artemis Maria Francelin Sanchez Moroni, que me acompanha desde a Iniciação Científica;

Ao Prof. Leandro Nunes de Castro, por ter gentilmente cedido o material para a redação do capítulo referente a Sistemas Imunológicos Artificiais;

A Mariana Lúcio Lyra, pelo incentivo e ajuda nas horas mais difíceis, sem os quais muitas de minhas conquistas não teriam sido possíveis e, principalmente, pela revisão minuciosa e valiosa que melhorou muito a qualidade do texto final;

A Taísa de Oliveira Amendola Sanches, pela paciência e compreensão, e pelo apoio sem o qual eu não teria conseguido;

A todos do LBiC, pela amizade e carinho com que me receberam, em especial:

A George Barreto Pereira Bezerra, pelas discussões acaloradas e iluminadas;

A Helder Knidel, pela preciosa ajuda que sempre ofereceu incondicionalmente;

A Tiago Venturieri Barra, que sempre esteve comigo quando eu precisei;

A Renato Reder Cazangi, pela inestimável amizade e pela comunidade no Orkut;

A Eurípedes Pinheiro dos Santos, por todos os livros com que me presenteou, fonte inestimável de conhecimento;

A Wilfredo Jaime Puma Villanueva que sempre se lembrou de mim nas suas viagens de volta pra casa;

A todos do NICS, pela maneira calorosa com que me acolheram rapidamente de volta ao grupo, apesar da minha grande distância e ausência, em especial:

A André Luiz Luvizotto, Fábio Furlanete, César Renno Costa e ao Prof. Adolfo Maia Jr.

Às pessoas que participaram dos experimentos de avaliação perceptiva: Helder Knidel, Tiago Venturieri Barra, Eurípedes Pinheiro dos Santos, Guilherme Palermo Coelho, George Barreto Pereira Bezerra, Wilfredo Jaime Puma Villanueva, Renan Cipriano Moioli;
E à Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) pelo apoio financeiro (processo no. 03/11122-8).

ÍNDICE

I) INTRODUÇÃO	1
I.1 ANTES DE 1945.....	2
I.2 <i>MUSIQUE CONCRÈTE</i>	6
I.3 <i>ELEKTRONISCHE MUSIK</i>	9
I.4 MÚSICA COMPUTACIONAL	17
I.5 POSICIONAMENTO DA PESQUISA, OBJETIVOS E METODOLOGIA.....	20
I.6 ORGANIZAÇÃO DO TEXTO	23
II) FUNDAMENTOS TEÓRICOS.....	24
II.1 MODELO SONORO.....	24
II.2 PERCEPÇÃO PSICOACÚSTICA	26
II.2.i) Altura	28
II.2.ii) Percepção de Intensidade	33
II.2.iii) Relações Temporais	35
II.2.iv) Teoria Clássica de Timbre	38
II.2.v) Estudos Modernos de Timbre	44
II.2.vi) Sons musicais complexos.....	47
II.2.vii) Timbre Revisitado	49
II.3 SÍNTESE SONORA.....	55
II.3.i) Síntese Aditiva.....	57
II.3.ii) Síntese Subtrativa	59
II.3.iii) Síntese FM	61
II.3.iv) <i>Waveshaping</i>	68
II.3.v) Síntese Granular	74
III) MOTIVAÇÃO.....	78
III.1 FORMALIZAÇÃO TEÓRICA	78
III.1.i) Representação Sonora por Computador	78
III.1.ii) Representação Vetorial.....	84
III.1.iii) Interpretação Vetorial das Formas de Onda	85
III.2 RETICULADO VS CONTÍNUO SONORO	86
III.3 ABORDAGEM ADOTADA NESTE TRABALHO.....	88
III.3.i) Otimização	89
III.3.ii) Codificação e Mapeamento	90
III.3.iii) Espaço Vetorial de Busca	91
III.3.iv) Caracterização da Região do Espaço Escolhida	93
III.4 POR QUE BIO-INSPIRAÇÃO?	102
III.4.i) Auto-Organização em Sistemas Biológicos	103
IV) SÍNTESE SONORA POR SISTEMAS IMUNOLÓGICOS ARTIFICIAIS.....	110
IV.1 INTRODUÇÃO.....	110
IV.1.i) Modelo Simplificado do Sistema Imunológico	112
IV.1.ii) Sistemas Imunológicos Artificiais	115
IV.1.iii) Princípio da Seleção Clonal e Maturação de Afinidade	115
IV.1.iv) Teoria da Rede Imunológica	117

IV.1.v)	aiNet	118
IV.1.vi)	Algoritmo da aiNet.....	119
IV.1.vii)	Parâmetros.....	121
IV.2	MODELO E OBJETIVOS	122
IV.3	EXPERIMENTOS E RESULTADOS.....	123
IV.4	DISCUSSÃO.....	131
V)	SÍNTESE SONORA POR REDES NEURAIIS ARTIFICIAIS	134
V.1	INTRODUÇÃO.....	134
V.2	MAPA AUTO-ORGANIZÁVEL DE KOHONEN	136
V.2.i)	Formação do Mapa Topológico.....	138
V.2.ii)	Inicialização do Mapa.....	138
V.2.iii)	O Processo Competitivo.....	138
V.2.iv)	O Processo Cooperativo	139
V.2.v)	O Processo Adaptativo.....	140
V.2.vi)	Algoritmo de Ajustes de Peso	142
V.2.vii)	Parâmetros.....	142
V.2.viii)	Problema do Caixeiro Viajante:.....	143
V.3	MODELO	144
V.4	EXPERIMENTOS E RESULTADOS	145
V.5	DISCUSSÃO.....	155
VI)	CONCLUSÃO	160
VII)	REFERÊNCIAS BIBLIOGRÁFICAS	166

ÍNDICE DE FIGURAS

FIGURA II.1.1: REPRESENTAÇÃO DE UMA FONTE SONORA E DA ONDA SONORA RESULTANTE.	25
FIGURA II.1.2: REPRESENTAÇÃO DE UMA ONDA SONORA COMO UMA FORMA DE ONDA.....	26
FIGURA II.2.1: OUVIDO HUMANO REPRESENTANDO OS DISPOSITIVOS EMPREGADOS NO PROCESSAMENTO E RESPECTIVA CONVERSÃO DA ONDA SONORA EM ESTÍMULO SONORO.	27
FIGURA II.2.2: DECOMPOSIÇÃO DE FOURIER DE UMA ONDA QUADRADA COMO UMA SOMA DE OSCILAÇÕES SENOIDAIS HARMONICAMENTE RELACIONADAS.	29
FIGURA II.2.3: PERIODICIDADE DE UMA FORMA DE ONDA.	30
FIGURA II.2.4: REPRESENTAÇÃO DO FENÔMENO DE BATIMENTO.	31
FIGURA II.2.5: BANDA CRÍTICA DE DISCRIMINAÇÃO DE ALTURA.	32
FIGURA II.2.6: MÍNIMA DIFERENÇA DISCRIMINÁVEL PARA NÍVEL DE PRESSÃO SONORA.	33
FIGURA II.2.7: CURVA DE FLETCHER-MUNSON.	34
FIGURA II.2.8: PERCEPÇÃO DE ALTURA DE UMA SEQUÊNCIA DE TONS.....	36
FIGURA II.2.9: PERCEPÇÃO DE SIMETRIA DE UMA SEQUÊNCIA DE TONS.	37
FIGURA II.2.10: RELAÇÕES TEMPORAIS.....	38
FIGURA II.2.11: MODELO SIMPLIFICADO DE UM TOM MUSICAL DE HELMHOLTZ.	39
FIGURA II.2.12: ENVELOPE ESPECTRAL DO ESPECTRO DE FOURIER DE UMA FORMA DE ONDA.....	41
FIGURA II.2.13: PICOS DE ENERGIA EM REGIÕES DE FREQUÊNCIA ABSOLUTA, OU FORMANTES.....	42
FIGURA II.2.14: ILUSTRAÇÃO DE UMA FONTE EXCITATÓRIA E UMA CAVIDADE ACÚSTICA RESSONANTE.	43
FIGURA II.2.15: ENVELOPE TEMPORAL DE UMA FORMA DE ONDA DE UM TOM MUSICAL.....	44
FIGURA II.2.16: ESPECTRO DINÂMICO DE UM TOM MUSICAL.	45
FIGURA II.2.17: ESCALAMENTO MULTIDIMENSIONAL DE GREY DE 14 INSTRUMENTOS MÚSICAIS.	48
FIGURA II.2.18: VISUALIZAÇÃO DA COMPLEXIDADE DE UM TOM MUSICAL NOS DOMÍNIOS DO TEMPO E DA FREQUÊNCIA.	49
FIGURA II.3.1: EXEMPLO DE SÍNTESE ADITIVA.	58
FIGURA II.3.2: ONDA QUADRADA E SEU RESPECTIVO ESPECTRO.....	59
FIGURA II.3.3: SÍNTESE SUBTRATIVA NOS DOMÍNIOS DO TEMPO E FREQUÊNCIA.	60
FIGURA II.3.4: EXEMPLO DE SÍNTESE FM.....	62

FIGURA II.3.5: ESPECTRO DO RESULTADO DO EXEMPLO DE SÍNTESE FM.....	63
FIGURA II.3.6: ESPECTRO DINÂMICO DO RESULTADO DO EXEMPLO DE SÍNTESE FM.	63
FIGURA II.3.7: DEPENDÊNCIA DAS FUNÇÕES DE BESSEL DO PRIMEIRO TIPO COM O ÍNDICE DE MODULAÇÃO.	65
FIGURA II.3.8: EXEMPLO DE ESPECTRO DE SÍNTESE FM EM FUNÇÃO DE FC, FM E B.	67
FIGURA II.3.9: EXEMPLO DE UMA FUNÇÃO WAVESHAPING.	70
FIGURA II.3.10: FORMA DE ONDA E ESPECTRO DE UMA OSCILAÇÃO SENOIDAL.	72
FIGURA II.3.11: FORMA DE ONDA E ESPECTRO DO RESULTADO DA APLICAÇÃO DA FUNÇÃO WAVESHAPING DA EQUAÇÃO 2.13.	73
FIGURA II.3.12: QUANTUM SONORO DE ACORDO COM O MODELO DE GABOR.....	75
FIGURA II.3.13: RESULTADO DO EXEMPLO ILUSTRATIVO DE SÍNTESE GRANULAR NO DOMÍNIO DO TEMPO.	76
FIGURA II.3.14: RESULTADO DO EXEMPLO ILUSTRATIVO DE SÍNTESE GRANULAR NO DOMÍNIO DA FREQUÊNCIA.....	77
FIGURA III.1.1: AMOSTRAGEM DE UM SINAL ANALÓGICO.	81
FIGURA III.1.2: AMOSTRAGEM DE UM SINAL ANALÓGICO.	82
FIGURA III.1.3: REPRESENTAÇÃO BINÁRIA.....	83
FIGURA III.1.4: REPRESENTAÇÃO VETORIAL.	85
FIGURA III.2.1: REPRESENTAÇÃO ESQUEMÁTICA DE TRÊS VARIÁVEIS MUSICAIS.	88
FIGURA III.3.1: VISUALIZAÇÃO DA CODIFICAÇÃO DAS AMOSTRAS SONORAS, DO ESPAÇO RESULTANTE E DO MAPEAMENTO PARA O ESPAÇO SONORO ONDE SE DÁ A MANIPULAÇÃO DOS PARÂMETROS MUSICAIS RELEVANTES.	91
FIGURA III.3.2: ESQUEMA DA REPRESENTAÇÃO DA MATRIZ DO ESPECTRO DINÂMICO.	93
FIGURA III.3.3: FORMA DE ONDA E ESPECTRO DINÂMICO DO SOM ATRATOR 1: SAX ALTO.	94
FIGURA III.3.4: FORMA DE ONDA E ESPECTRO DINÂMICO DO SOM ATRATOR 2: BAIXO ELÉTRICO.	94
FIGURA III.3.5: : FORMA DE ONDA E ESPECTRO DINÂMICO DO SOM ATRATOR 3: VIOLÃO DEDILHADO.....	95
FIGURA III.3.6: : FORMA DE ONDA E ESPECTRO DINÂMICO DO SOM ATRATOR 4: ACORDE DE PIANO.	95
FIGURA III.3.7: : FORMA DE ONDA E ESPECTRO DINÂMICO DO SOM ATRATOR 5: GAITA.....	96
FIGURA III.3.8: : FORMA DE ONDA E ESPECTRO DINÂMICO DO SOM ATRATOR 6: VOZ.....	96
FIGURA III.3.9: FORMA DE ONDA E ESPECTRO DINÂMICO DO SOM ATRATOR 7: APITO.....	97

FIGURA III.3.10: REPRESENTAÇÃO DA DISTÂNCIA ESPACIAL AOS PARES DE TRÊS PONTOS NO PLANO.	98
FIGURA III.4.1: EXEMPLOS DOS DIVERSOS PADRÕES BIOLÓGICOS CUJA ESTRUTURAÇÃO PODE SER EXPLICADA ATRAVÉS DE PROCESSOS AUTO-ORGANIZÁVEIS. RETIRADA DE CAMAZINI <i>ET AL.</i> [11].	104
FIGURA IV.1.1: ILUSTRAÇÃO DO RECONHECIMENTO DE UM ANTICORPO POR UM ANTÍGENO.	113
FIGURA IV.1.2: MODELO SIMPLIFICADO DO FUNCIONAMENTO DO SISTEMA IMUNOLÓGICO.	114
FIGURA IV.1.3: MODELO SIMPLIFICADO DA REDE IMUNOLÓGICA.	117
FIGURA IV.2.1: REPRESENTAÇÃO PICTÓRICA DA CAPACIDADE DE EXTRAÇÃO DE CARACTERÍSTICAS (QUANTIZAÇÃO VETORIAL) DA AI ^{NET}	123
FIGURA IV.3.1: FORMA DE ONDA E ESPECTRO DINÂMICO DO ANTICORPO 1	125
FIGURA IV.3.2: FORMA DE ONDA E ESPECTRO DINÂMICO DO ANTICORPO 2	125
FIGURA IV.3.3: FORMA DE ONDA E ESPECTRO DINÂMICO DO ANTICORPO 3	125
FIGURA IV.3.4: FORMA DE ONDA E ESPECTRO DINÂMICO DO ANTICORPO 4	126
FIGURA IV.3.5: FORMA DE ONDA E ESPECTRO DINÂMICO DO ANTICORPO 5	126
FIGURA IV.3.6: FORMA DE ONDA E ESPECTRO DINÂMICO DO ANTICORPO 1 DEPOIS DE FILTRADO.	127
FIGURA IV.3.7: FORMA DE ONDA E ESPECTRO DINÂMICO DO ANTICORPO 2 DEPOIS DE FILTRADO.	127
FIGURA IV.3.8: FORMA DE ONDA E ESPECTRO DINÂMICO DO ANTICORPO 3 DEPOIS DE FILTRADO.	128
FIGURA IV.3.9: FORMA DE ONDA E ESPECTRO DINÂMICO DO ANTICORPO 4 DEPOIS DE FILTRADO.	128
FIGURA IV.3.10: FORMA DE ONDA E ESPECTRO DINÂMICO DO ANTICORPO 5 DEPOIS DE FILTRADO.	128
FIGURA V.2.1: IMAGENS TRIDIMENSIONAIS DE RESSONÂNCIA MAGNÉTICA DA ATIVIDADE CEREBRAL.....	137
FIGURA V.2.2: ESQUEMA DE UMA REDE DE KOHONEN UNIDIMENSIONAL.....	137
FIGURA V.2.3: FUNÇÃO DE VIZINHANÇA GAUSSIANA.	140
FIGURA V.2.4: ADAPTAÇÃO DO VETOR DE PESO DO NEURÔNIO VENCEDOR.	141
FIGURA V.4.1: REPRESENTAÇÃO DA APLICAÇÃO DE MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN AO PROBLEMA DE SÍNTESE SONORA DIGITAL DESENVOLVIDO COMO UMA SOLUÇÃO DO PROBLEMA DO CAIXEIRO VIAJANTE.	146
FIGURA V.4.2: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 1.....	147
FIGURA V.4.3: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 2.....	148
FIGURA V.4.4: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 3.....	148
FIGURA V.4.5: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 4.....	148

FIGURA V.4.6: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 5.....	149
FIGURA V.4.7: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 6.....	149
FIGURA V.4.8: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 7.....	149
FIGURA V.4.9: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 8.....	150
FIGURA V.4.10: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 9.....	150
FIGURA V.4.11: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 10.....	150
FIGURA V.4.12: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 11.....	151
FIGURA V.4.13: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 12.....	151
FIGURA V.4.14: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 13.....	151
FIGURA V.4.15: FORMA DE ONDA E ESPECTRO DINÂMICO DO NEURÔNIO 14.....	152

ÍNDICE DE TABELAS

TABELA III.3.1: DESCRIÇÃO DOS SONS UTILIZADOS COMO ATRADORES.....	94
TABELA III.3.2: DISTÂNCIAS DOIS A DOIS ENTRE OS PONTOS MOSTRADOS NA FIGURA III.3.10.	98
TABELA III.3.3: PONTO MAIS PRÓXIMO DE CADA UM DOS PONTOS APRESENTADOS NA FIGURA III.3.10. ...	99
TABELA III.3.4: DISTÂNCIAS ENTRE OS ATRADORES CALCULADAS PELA MÉTRICA EUCLIDIANA.....	99
TABELA III.3.5: ATRADORES MAIS PRÓXIMOS DE CADA ATRATOR DE ACORDO COM A MÉTRICA EUCLIDIANA.	99
TABELA III.3.6: DISTÂNCIAS ENTRE OS ATRADORES CALCULADAS PELA MÉTRICA TEMPO-FREQÜENCIAL.	100
TABELA III.3.7: ATRADORES MAIS PRÓXIMOS DE CADA ATRATOR DE ACORDO COM A MÉTRICA TEMPO- FREQÜENCIAL.	100
TABELA III.3.8: SIMILARIDADES SUBJETIVAS ENTRE OS ATRADORES ESTIMADAS PELOS INDIVÍDUOS.....	101
TABELA III.3.9: AGRUPAMENTO DAS SIMILARIDADES SUBJETIVAS ENTRE OS ATRADORES COMO ESTIMADO PELOS INDIVÍDUOS. X SIGNIFICA NÃO SIMILARIDADE.....	101
TABELA IV.1.1: PARÂMETROS DE ENTRADA DA aiNet.	121
TABELA IV.3.1: VALORES DOS PARÂMETROS DA aiNet UTILIZADOS NO EXPERIMENTO.	124
TABELA IV.3.2: AGENTES MAIS PRÓXIMOS DE CADA ATRATOR DE ACORDO COM O CRUZAMENTO VISUAL. X INDICA NÃO SIMILARIDADE.	127
TABELA IV.3.3: DISTÂNCIAS ENTRE OS ANTÍGENOS E OS ANTICORPOS CALCULADAS PELA MÉTRICA EUCLIDIANA.....	129
TABELA IV.3.4: AGENTES MAIS PRÓXIMOS DE CADA ATRATOR DE ACORDO COM A DISTÂNCIA EUCLIDIANA.	130
TABELA IV.3.5: DISTÂNCIAS ENTRE OS ANTÍGENOS E OS ANTICORPOS CALCULADAS PELA MÉTRICA TEMPO- FREQÜENCIAL.	130
TABELA IV.3.6: AGENTES MAIS PRÓXIMOS DE CADA ATRATOR DE ACORDO COM A DISTÂNCIA TEMPO- FREQÜENCIAL.	130
TABELA IV.3.7: SIMILARIDADES SUBJETIVAS ENTRE OS ATRADORES ESTIMADAS PELOS INDIVÍDUOS.	131
TABELA IV.3.8: AGENTES MAIS PRÓXIMOS DE CADA ATRATOR DE ACORDO COM A AVALIAÇÃO PERCEPTIVA DOS INDIVÍDUOS.	131
TABELA V.2.1: PARÂMETROS DE ENTRADA DO MAPA AUTO-ORGANIZÁVEL.	143

TABELA V.4.1: PARÂMETROS DO SOM.....	147
TABELA V.4.2: AGENTES MAIS PRÓXIMOS DE CADA ATRATOR DE ACORDO COM O CRUZAMENTO VISUAL.	152
TABELA V.4.3: DISTÂNCIA TEMPORAL (EUCLIDIANA) ENTRE-OS ATRADORES E OS AGENTES PARA SOM	153
TABELA V.4.4: AGENTES MAIS PRÓXIMOS DE CADA ATRATOR DE ACORDO COM A DISTÂNCIA TEMPORAL.	153
TABELA V.4.5: DISTÂNCIA TEMPO-FREQÜENCIAL ENTRE-OS ATRADORES E OS AGENTES PARA SOM. ...	154
TABELA V.4.6: AGENTES MAIS PRÓXIMOS DE CADA ATRATOR DE ACORDO COM A DISTÂNCIA TEMPO- FREQÜENCIAL.	154
TABELA V.4.7: SIMILARIDADES SUBJETIVAS ENTRE OS AGENTES E OS ATRADORES DE ACORDO COM AS RESPOSTAS DOS INDIVÍDUOS.	155
TABELA V.4.8: AGENTES MAIS PRÓXIMOS DE CADA ATRATOR DE ACORDO COM A AVALIAÇÃO PERCEPTIVA.	155

I) INTRODUÇÃO

Toda cultura humana desenvolveu – paralelamente à língua – pelo menos uma forma de arte visual e música. Pouco pode ser dito das verdadeiras origens da música pois suas raízes são envoltas na mais remota antigüidade humana. “A música é, talvez, mais uma característica definidora dos seres humanos que uma fase passageira na evolução da inteligência” [55].

O significado da música tem sido debatido acaloradamente em vários graus através da história registrada. Todos estes debates, por mais iluminados, falharam em capturar a essência da música pela simples razão de que os atributos musicais não podem ser expressos em palavras.

A incapacidade de se traduzir seu significado em palavras cedeu à música tanto respeito quanto um certo grau de temor. Entre os gregos antigos, o grande filósofo e poeta Platão repetidamente revelava sua inquietação em relação à música. Parece-nos, hoje, que Platão estava entre os primeiros e mais significativos a cair na armadilha de confundir conhecimento com linguagem. Para um músico “descobrir o que um dado ritmo e afinação significam” está dentre as mais fáceis - não as mais difíceis - tarefas, desde que, certamente, não peçamos a ele para traduzi-los em palavras. Invariavelmente, essas palavras, não importa o quão habilmente empregadas, serão empobrecidas quando comparadas com a ampla extensão do conhecimento humano [55]. Confundir conhecimento que pode ser expresso em palavras com todo o conhecimento é o mesmo que confundir o mapa com o território. Todo conhecimento tem uma forma de expressão própria, mas a expressão tem muitas formas, e música é só uma delas. Em ciência moderna, por exemplo, a matemática é comumente usada para investigar e expressar conceitos fundamentais sobre a natureza. O poder da matemática reside no fato de que captura-se a essência de certos tipos de relações de uma maneira que revela mais que esconde para qualquer um fluente em seus princípios. No entanto, seria na maioria dos casos difícil, se não impossível, fazer descobertas matemáticas através apenas do uso das palavras.

Se se tentasse a impossível tarefa de definir com palavras o que a música é realmente, poder-se-ia sugerir que a música estimula um sentido para o qual nossas mentes são o órgão primário (se não o único): nosso senso de tempo. Pelo fato de música

ser uma arte temporal, seu estudo inclui necessariamente um método de capturar, representar e interpretar informação de sucessivos momentos no tempo. Tal avanço tecnológico ocorreu por volta de um século atrás com o advento de meios de se gravar sons. A gravação sonora não só transformou drasticamente o estudo da música, mas também os métodos pelos quais a música é feita de maneiras fundamentais. O uso de sons gravados na música foi um precursor do desenvolvimento da música concreta (*musique concrète*) na França por volta de 1945, a qual era baseada na justaposição de gravações de sons “concretos” da natureza, em oposição aos sons “abstratos” dos instrumentos musicais tradicionais.

I.1 Antes de 1945

Instrumentos musicais elétricos e eletrônicos vêm agraciando o horizonte musical desde a virada do século XX. Data de 1897 a primeira patente de um dispositivo de geração elétrica de som. A invenção é um marco tão importante para a música eletrônica quanto o bastante celebrado fonógrafo, patenteado por Edison e Berliner cerca de vinte anos antes. O Dynamophone, nome da invenção cuja primeira apresentação pública foi em 1906, era essencialmente um dínamo elétrico modificado para produzir corrente alternada com diferentes frequências na faixa audível. A máquina pesava em torno de 200 toneladas e tinha 20 metros de altura, assumindo as proporções de um gerador de energia elétrica [50]. O Dynamophone oferecia recursos de produção sonora novos e flexíveis com um grau que não foi igualado por projetos subsequentes por algum tempo. Thaddeus Cahil, o dono da patente, via sua invenção não somente como um substituto para instrumentos de teclado mas como uma ferramenta para exploração de uma vasta gama de sons, sendo possível produzir qualquer timbre desejado somente através dos componentes elétricos da máquina, variando-se a qualidade dos sons produzidos pela alteração das intensidade relativas das frequências harmônicas associadas com uma nota. O alto custo, aliado à descoberta de que a invenção interferia em chamadas telefônicas, levou o projeto visionário à falência logo antes de a primeira grande guerra irromper na Europa.

Avanços no recém-estabelecido campo da eletrônica, no entanto, estavam pavimentando o caminho para outras abordagens mais baratas e compactas para a geração de sons sintéticos. Próximo do fim da guerra, com a indústria já estabelecida, vários engenheiros puderam investigar a possibilidade de se usar a nova tecnologia para

a construção de instrumentos musicais eletrônicos. A motivação primária por trás desses projetos era o desejo de se criar adições e alternativas ao conjunto de possibilidades orquestrais, na esperança de que os compositores seriam persuadidos a fornecer um repertório adequado. Os dispositivos que surgiram eram, portanto, pensados para satisfazer as idéias tradicionais de composição musical. Alguns, aliás, eram nada mais que instrumentos acústicos modificados que usavam captadores para utilizar vibrações produzidas naturalmente e modificá-las eletronicamente. O exemplo mais conhecido é a guitarra elétrica. A maior parte deles usava algum meio de geração eletrônica. Temos como exemplo o Etherphone ou Thèremin inventado pelo russo Leon Thèremin em 1920. Usando-se este instrumento, era possível controlar a intensidade e a altura do som gerados por osciladores quase-senoidais através do movimento das mãos próximo de duas antenas [5], [28]. A maioria usava um teclado como interface de produção musical e um pedal ou alavanca para controle de volume. Instrumentos eletrônicos deste tipo floresceram brevemente durante o período entre guerras e, apesar das contribuições de alguns compositores, apenas um repertório limitado surgiu. Esta possibilidade de programar a produção de sons antecipou dispositivos como os sintetizadores e, de maneira mais geral, o uso de computadores na produção sonora, aproximadamente um quarto de século mais tarde.

O potencial desses instrumentos como substitutos dos instrumentos acústicos tradicionais foi rapidamente reconhecido e explorado no campo da música popular. No entanto, essas aplicações contribuíram muito pouco para uma apreciação do potencial artístico deste novo meio de produção sonora. Era, portanto, inevitável que as primeiras incursões neste mundo desconhecido seriam guiadas pela prática instrumental tradicional. Alguns pioneiros estavam ansiosos para explorar as possibilidades de uma esfera sonora ampliada e menos restrita. Uma das primeiras tentativas de se utilizar técnicas não tradicionais de geração sonora como parte de uma arte comunicativa surgiu das atividades dos membros do movimento Futurista, que pregava a rejeição dos princípios musicais e métodos de ensino tradicionais em nome da expressão livre, inspirada pela natureza em todas as suas manifestações. O *Manifesto Técnico da Música Futurista* sugeria que os compositores deveriam dominar todos os elementos dinâmicos e expressivos técnicos de instrumentação, e trabalhar a orquestra como um universo sonoro em constante mobilidade, integrada por suas partes constituintes. Além disso, o documento considerava que o trabalho dos compositores futuristas deveria refletir todas as forças da natureza domadas pelo homem através de suas descobertas científicas

contínuas, por exemplo, o som musical de multidões, de grandes fábricas, de trens, grandes navios transatlânticos, automóveis, aviões. No manifesto futurista “*The Art of Noises*” lê-se: o som musical é muito limitado qualitativamente quanto a timbre. As orquestras mais complexas se reduzem a quatro ou cinco classes de instrumentos. Deve-se romper este estreito círculo de sons musicais puros e conquistar uma infinita variedade de sons [50]. As manifestações práticas da proposta envolviam a construção de instrumentos musicais especialmente projetados para tal fim. O movimento futurista não foi bem sucedido em sua tentativa de produzir uma revolução na música, mas seu desvio das relações tradicionalmente aceitas entre a ciência acústica e a arte da produção de sons musicais foi profético.

Um compositor em especial, Pierre Varèse, estava disposto a explorar novos conceitos de expressão musical, tornando-se um pioneiro na estética necessária para a aceitação das técnicas eletrônicas de processamento sonoro em composição musical com sua música instrumental. Na década de 1921-1930 Varèse continuou sua busca por novas texturas sonoras sem ajuda de facilidades técnicas. Seu trabalho com recursos instrumentais naturais foi, também, inovador porque utilizava procedimentos com características do processamento eletrônico de sons. Ele estava interessado em particular no uso de instrumentos como blocos componentes de massas sonoras com qualidades, densidades e volumes variados, em contraste com seus papéis tradicionais como fonte de contraponto musical linear. Com o auxílio de efeitos como sirenes, chicotes, rugidos de leão e sinos, ele lutou para desenvolver uma estética musical que integrasse os sons naturais do meio ambiente com as fontes mais tradicionais de expressão musical. Suas tentativas de analisar sons musicais em termos músico-acústicos, usando características identificadoras como melodia, ritmo, timbre, harmonia, dinâmica e intensidade, antecipou o tipo de metodologia que seria aplicado por muitos outros na procura por uma morfologia da música eletrônica [50], [72].

A primeira década do século XX presenciou o nascimento do gramofone de 78 rotações por minuto, e a década seguinte do mesmo século o desenvolvimento de sistemas elétricos de gravação, tornando disponível não só uma técnica para guardar informação sonora, mas também para trazer a efeito certas alterações durante sua reprodução. Percebeu-se que mudanças na velocidade de reprodução de gravações alteram não somente a altura, mas também qualidades acústicas intrínsecas do material. No entanto, ainda não havia chegado o momento em que procedimentos de composição

musical que se estendessem além da tradicional orquestra fossem aceitos. Já se discutiam as implicações artísticas do uso da tecnologia eletrônica em composições musicais, estreitando os laços entre cientistas e artistas. Sistemas de gravação magnética existem desde 1898. A qualidade de reprodução dos primeiros sistemas era muito ruim, mas o advento da fita plástica magnética no fim da segunda guerra revolucionou as técnicas de gravação de estúdio, tornando possível ao compositor ter contato direto com o material sonoro que manipula. Os compositores de então, assim como ainda hoje, encaravam os problemas de especificação, particularmente buscando equacionar o mundo subjetivo do músico criativo com as características altamente objetivas da nova tecnologia. No fim da década de trinta os avanços científicos produziram as teorias básicas para o projeto de técnicas de síntese sonora, e o músico criativo e informado estava em posição de prever as possíveis linhas de desenvolvimento. O compositor John Cage, em muitos aspectos pertencente a uma escola estética diferente, apoiou o potencial uso da eletrônica digital em composição musical. Contudo, os princípios artísticos estabelecidos estavam muito além dos meios práticos para realizá-los. O nascimento subsequente dos estúdios de música eletrônica se deu em um clima em que os problemas encontrados ao se relacionar a tecnologia à linguagem musical já tinham sido identificados, se não solucionados. O interesse no meio composicional havia se expandido para o uso de técnicas eletrônicas de apresentação ao vivo, e o palco estava montado para o surgimento dos primeiros estúdios equipados apropriadamente, e o conseqüente surgimento de suas escolas de composição associadas.

A ressurreição das artes após a segunda guerra se deu em ambiente mais favorável ao desenvolvimento da música eletrônica. Os rápidos avanços da tecnologia como resultado da guerra, um crescimento vertiginoso no interesse por novas técnicas sonoras e um clima econômico expansionista forneceram incentivos suficientes para as instituições darem suporte à empreitada. Na Europa, a iniciativa foi tomada por duas redes de radiodifusão, *Radiodiffusion Télévision Française* (RTF) em Paris e *Westdeutscher Rundfunk* (WDR) em Colônia. Ambas estabeleceram estúdios de importância considerável, a princípio perseguindo objetivos radicalmente diferentes. O grupo de Paris se dedicou ao desenvolvimento de seu estilo próprio, denominado *musique concrète*, enquanto que o grupo de Colônia abraçou a causa da *elektronische Musik*.

I.2 *Musique Concrète*

Falar de um grupo quando se consideram os primeiros anos do estúdio de Paris seria impreciso, pois a iniciativa e liderança do projeto partiram de um pioneiro, Pierre Schaeffer. No curso de sua pesquisa, a atenção de Schaeffer se voltou para o uso de técnicas de gravação como meio de isolar eventos sonoros acústicos e, em 1948, ele começou a considerar como esse material poderia ser usado como base de composições. As investigações preliminares de Schaeffer, inspiradas até certo ponto nos Futuristas, se davam no âmbito da exploração das propriedades de sons percussivos. Durante os primeiros meses de 1948, ele estudou o efeito de se percutir instrumentos diferentes de maneiras diferentes. Isso o levou a observar que qualquer evento musical é caracterizado não somente pelo timbre e sustentação do som, mas também pela natureza do ataque e decaimento (ver seção II.2.vi). Tendo concluído um estudo superficial do ataque, sustentação e decaimento de eventos sonoros isolados, além dos efeitos de tocar gravações de trás para frente, Schaeffer voltou sua atenção para a tarefa de resíntese. Schaeffer percebeu rapidamente que sérios problemas de associação foram criados usando-se fontes que retinham uma porção significativa de suas características identificáveis após o processamento. Em uma tentativa de superar essa dificuldade, ele se voltou a fontes de sons musicais mais convencionais, investigando os efeitos de se tocar gravações em diferentes velocidades. Isto levou à descoberta de que tais alterações afetam não somente a altura e duração total dos eventos individuais, mas também seu envelope (seção II.2.v). Tal interdependência tornava impossível variar um desses fatores sem afetar os outros. Seus primeiros estudos, realizados em meados de 1948, resultaram insatisfatórios devido à falta de diálogo coerente entre as diferentes áreas de material sonoro, criando a impressão de peças desconexas sobrepostas. Esta descoberta dos problemas de se integrar fontes distintas foi importante por identificar um grande desafio para compositores de música eletrônica.

Em 1949, a RTF nomeou o compositor Pierre Henry co-pesquisador e o engenheiro de som Jacques Poullin, que já havia expressado interesse no trabalho de Schaeffer, técnico de estúdio. As principais preocupações de Schaeffer nesta época eram os possíveis paralelos que podem ser traçados entre os processos de composição convencional e concreto. Isto levou à identificação de duas abordagens distintas. Por um lado, um compositor pode escolher começar desenvolvendo um conceito claro das estruturas sonoras que quer atingir. Tal panorama requer racionalização e modificação

em termos do equipamento disponível, levando, no caso do trabalho concreto, a um conjunto preciso de rotinas de estúdio que devem ser executadas. Por outro lado um compositor pode desejar começar com uma seleção de potenciais fontes sonoras, oferecendo uma gama de características com as quais vai trabalhar, construindo os elementos da composição a partir dos resultados das investigações. Essas distinções se provaram importantes não apenas para Schaeffer, mas para o desenvolvimento da música eletrônica em geral, pois destacaram dificuldades procedurais importantes encontradas ao se relacionar o mundo subjetivo da criatividade musical ao mundo objetivo e científico do estúdio. É necessário, portanto, para se trabalhar de acordo com a primeira abordagem, uma linguagem capaz de traduzir a variedade de idéias musicais em procedimentos de estúdio equivalentes. A segunda abordagem, em contraste, requer um diálogo menos complexo entre o compositor e o sistema, dependente das características funcionais oferecidas pelos dispositivos e, no caso de material concreto, das características das fontes escolhidas. Pode-se perseguir, portanto, as idéias clássicas de orquestra e partitura com dispositivos eletrônicos fazendo as vezes dos instrumentos tradicionais onde apropriado.

Os primeiros trabalhos de Schaeffer tinham, em sua maioria, procedido da idéia geral do resultado desejado para uma tentativa de realizá-lo através da seleção de material e processos apropriados. No entanto, ele também realizou experimentos com a abordagem inversa, estudar as características intrínsecas da música instrumental e aplicar procedimentos concretos apropriados para produzir uma nova música. Schaeffer fez uso maciço de técnicas de transposição de altura, levadas a efeito tocando-se a gravação-fonte com diferentes velocidades. Os resultados musicais eram bastante insatisfatórios, não sem causa. A primeira impressão era de repetição monótona com pouco senso de direção ou forma. Essas dificuldades o fizeram analisar mais de perto a natureza dos sons, levando a uma definição primária de um objeto sonoro, um evento sonoro básico, isolado de seu contexto original e examinado através de suas características inatas, sem levar em conta aspectos temporais. Ele afirmou que a abstração de tais eventos das fontes sonoras naturais, visando fornecer componentes para a regeneração de material musical, necessitava de processos compatíveis com os princípios do serialismo pós Webern. Essa afirmação seria, mais tarde, ferozmente refutada pela escola alemã de *elektronische Musik* [50]. Schaeffer tentou estabelecer por que seus procedimentos de transformação falhavam em remover significativamente muitas das características marcantes das fontes sonoras. Ele concluiu que técnicas como

tocar as gravações de trás para a frente, ou com diferentes velocidades, aliadas ao uso de montagens elementares, não produziam nada de essencialmente novo. O uso de instrumentos musicais, hábitos musicais e estruturas musicais tinham condicionado a maneira na qual ele tinha conduzido seus processos de análise e resíntese. Parecia-lhe apropriado retornar ao ponto de partida, o mundo dos ruídos como fonte básica de informação sonora.

Durante os primeiros estágios de formulação, Schaeffer encontrou considerável dificuldade ao selecionar fontes apropriadas de material. Duas linhas de desenvolvimento estavam proeminentes em sua mente na época: a extensão de possibilidades de fontes instrumentais por meio de novas facilidades técnicas; e o desenvolvimento de seus princípios de objetos sonoros e suas regras de composição. Depois de muito considerar, ele concluiu que era necessário formalizar uma estruturação para os objetos sonoros, classificando-os em hierarquias de tessitura, timbre, ritmo e densidade. Adicionou-se também, mais tarde, mais uma dimensão à *musique concrète* através da possibilidade de projeção espacial do material com quatro ou cinco alto-falantes distribuídos. O uso de um alto-falante no teto possibilitou a criação de ilusões de movimento vertical e horizontal, adicionando uma dimensão extra na difusão sonora. Isso levou Schaeffer a formular a idéia de uma *orchestre concrète* baseado na observação de que certos sons continuariam a reter certas características qualquer que fosse o grau de transformação que sofressem, dentro dos limites das técnicas aplicadas e da percepção do ouvido humano. A persistência dessas características fez com que esses elementos fossem tratados como pseudo-instrumentos, notados na partitura de maneira similar com a utilizada para instrumentos convencionais.

Foi durante 1951 que os desentendimentos entre os proponentes da *musique concrète* e *elektronische Musik* começaram a se ampliar. A “*Symphonie pour um homme Seul*” de Schaeffer e Henry, tocada em rádios de Colônia, Hamburgo, Baden Baden e Munique, foi recebida com hostilidade por aqueles que adotaram o lado alemão. A “*Internationale Ferienkurse für neue Musik*” assumiu a controvérsia organizando um simpósio sobre tecnologia sonora e música. Os franceses e os alemães discordaram violentamente e os suíços criticaram ambos por descrever seu trabalho como música. Schaeffer, em particular, criticava os conceitos da *elektronische Musik* por não fornecer uma solução ou caminho para lidar com os problemas de comunicação associados com a música contemporânea.

Em 1966 a estruturação de Schaeffer dos objetos sonoros estava formalizada suficientemente bem para ser consolidada em uma obra formidável, *“Traité des objets musicaux”* [68]. Durante o renascimento das artes após a Segunda Guerra, Paris havia se tornado cada vez mais importante como ponto focal de desenvolvimentos da nova música. Compositores que trabalharam no estúdio naquela época incluíam Jean Barraqué, Pierre Boulez, Michel Philippot e Hermann Scherchen. Também visitaram o estúdio Olivier Messiaen, Darius Milhaud e Karlheinz Stockhausen, dentre outros. Nenhum desses compositores, no entanto, sustentou mais que uma conexão passageira com Schaeffer, desencorajados provavelmente pela inevitável falta de refinamento na época. A breve associação de Stockhausen com o estúdio aconteceu durante seu ano de estudos com Messiaen no Conservatório de Paris. Goeyvaerts tinha chamado sua atenção para as possibilidades de síntese do oscilador senoidal e o equipamento de gravação do *“Club d’Essai”* o atraiu por razões distintamente contrárias às de Schaeffer. Próximo do fim da década, os princípios da *musique concrète* se tornaram mais canônicos e as técnicas eletrônicas de geração e processamento de som foram ganhando aceitação gradual. Fontes sonoras naturais, no entanto, permaneceram o interesse primário de muitos compositores. Em 1958, o compositor Iannis Xenakis produziu seu primeiro trabalho em um estúdio grande, *Diamorphoses*, no qual aplicava princípios de organização matemática ao processo de composição com sons concretos. No mesmo período, Schaeffer começou uma grande reorganização tanto de seus métodos de composição quanto do estúdio em si, e o termo *musique concrète* foi substituído pela descrição muito mais universal *expériences musicales*. Esta abordagem refletia a filosofia do estúdio de Colônia. A diversificação de idéias e técnicas estava então levando alguns a lugares remotos e inexplorados. Xenakis, por exemplo, já tendo trabalhado tanto com sons de origem natural quanto eletrônica, começou a fazer experimentos com computadores, primeiramente como máquinas processadoras de dados, e então como fonte de sons sintetizados. No início de 1960 o trabalho inspirado pelo estúdio de Schaeffer estava se mesclando com outras linhas de desenvolvimento para criar um meio de rápida expansão da música eletrônica. A idade do pioneirismo isolado estava acabando.

I.3 *Elektronische Musik*

A *elektronische Musik*, diferentemente da *musique concrète*, começou não como resultado dos esforços de um indivíduo, mas como resultado da colaboração entre várias

partes interessadas tanto no aspecto técnico quanto no musical. Durante 1948, Meyer-Eppler foi visitado por Homer Dudley, que trouxe consigo uma máquina desenvolvida recentemente, o Vocoder, que podia funcionar como analisador de fala ou falante artificial. A síntese de sons era conseguida processando-se representações de timbres selecionados e passando-se essas características para um banco de filtros analíticos para produzir espectros audíveis. Apesar de a fidelidade da máquina ser limitada, pois seu propósito básico era processar fala e não música, Meyer-Eppler ficou impressionado. Durante 1949, ele usou gravações de sons produzidos pelo instrumento para ilustrar uma palestra sobre produção eletrônica de sons. Por uma feliz coincidência, Robert Beyer, da rádio North West German, estava na platéia e já tinha mostrado interesse nos usos musicais da tecnologia eletrônica. Beyer já tinha publicado um artigo em 1928 que incluía uma discussão sobre o uso de instrumentos eletrônicos em composição musical. Os dois cientistas acharam que o interesse geral no assunto merecia exposição pública e ambos deram palestras sobre 'O Mundo Sonoro da Música Eletrônica' em 1950. Beyer se concentrou nos princípios de projeto aplicados na construção de instrumentos musicais eletrônicos e Meyer-Eppler falou sobre o estado da pesquisa no campo de síntese de fala. O compositor Herbert Eimert expressou entusiasmo e os três começaram uma associação informal com o objetivo de avançar o desenvolvimento da *elektronische Musik*.

Em 1950, Harald Bode entregou um de seus Melochords, um instrumento eletrônico, a Meyer-Eppler em Bonn. O instrumento foi usado para preparar uma série de estudos simples sobre a produção de sons eletrônicos criados por camadas de gravações de tons selecionados. Os resultados dessas investigações preliminares foram apresentados por Meyer-Eppler em uma palestra intitulada 'As possibilidades da Produção Eletrônica de Sons'. Beyer contribuiu com um artigo sobre 'Música e Tecnologia' e Eimert discutiu 'Música na Fronteira'. Schaeffer estava presente e essa confrontação pública forneceu ímpeto para as diferenças crescentes entre as filosofias francesa e alemã de música eletrônica.

Em 1951, o diretor técnico da Rádio Colônia concordou em fundar um estúdio de música eletrônica para perseguir os processos sugeridos por Meyer-Eppler e compor diretamente em fita. O trabalho no estúdio começou naquele ano, mas demorou dois anos para que o estúdio estivesse pronto. Naquele mesmo ano, Meyer-Eppler deu outra palestra sobre 'Novos Métodos de Geração de Tons' para uma platéia de aproximadamente mil pessoas em um encontro de tecnologistas em Bonn. Durante a

primeira metade de 1952, o compositor Bruno Maderna produziu uma peça em associação com Meyer-Eppler no instituto de fonética e a mesma foi apresentada para uma platéia que incluía, dentre outros, Goeyvaerts e Stockhausen, que viriam, mais tarde, se envolver com composição de *elektronische Musik* em Colônia. Vista em retrospecto, a composição de Maderna para flauta, percussão e um alto-falante reproduzindo uma fita com material gerado eletronicamente é estilisticamente interessante. Esta integração de sons naturais e eletrônicos não estava de acordo nem com os princípios da *musique concrète* nem da *elektronische Musik*, antecipando uma época em que o processo de composição musical deixaria de ser sujeito a tais restrições rígidas.

Beyer e Eimert compuseram sua primeira peça inteiramente eletrônica enquanto o estúdio ainda estava em construção. Próximo do fim de 1952, o estúdio ficou parcialmente pronto, permitindo que o trabalho fosse transferido para este ambiente mais apropriado. Neste mesmo ano, ironicamente para Schaeffer, as primeiras peças completas de *elektronische Musik* estrearam em Paris. Foi nesta época que Stockhausen começou sua longa e influente associação com o estúdio de Colônia. Suas peças fornecem um ponto focal útil na avaliação das características musicais dessa era, mas primeiro é necessário levar em conta o clima técnico e filosófico em que foram produzidas. O início da década de 1950 já havia sido identificado como uma época caracterizada por poderosas influências na busca por novos horizontes em composição musical. Um desejo crescente de possuir controle sobre todos os aspectos da composição musical levou a um grande interesse nas possibilidades da síntese eletrônica, pois tal domínio eliminava não somente o processo intermediário de performance, mas também a necessidade de aceitar as características inatas de fontes sonoras naturais. A aquisição de tal poder, no entanto, precipitou uma grave crise no movimento em direção ao determinismo total, pois revelou-se extremamente difícil criar alternativas culturalmente aceitáveis para essas características essenciais da música tradicional. A *musique concrète*, com todas as suas limitações e problemas, reteve um grau tranquilizante de familiaridade em virtude de sua derivação de recursos do mundo sonoro natural. Os primeiros defensores da *elektronische Musik* não apenas se restringiam a meios inteiramente sintéticos de produção sonora, mas também não dissociaram seu trabalho das qualidades imitativas dos instrumentos musicais eletrônicos. Um interesse continuado em pesquisa fonética sob a supervisão de Meyer-Eppler, no entanto, exerceu uma grande influência na busca por um serialismo eletrônico puro, trabalhando inteiramente com princípios de síntese sonora. Uma cativação inicial pelas sutilezas da síntese de formantes foi levada à integração de

sons naturais de voz com fontes eletrônicas, e então para outros materiais produzidos naturalmente. Tal desenvolvimento quase não pôde ser reconciliado com os objetivos expressos durante os primeiros anos. Eimert já tinha chamado atenção para uma grande dificuldade, uma questão ainda não resolvida quanto ao fato de a música eletrônica, sendo a mesma uma fonte universal de todos os sons, possuir ou não uma força coerente para sustentar formas correspondentes à tonalidade, um sistema de timbres auto-sustentável.

O desejo de exercer controle total sobre o processo de especificação timbral levou, após alguma deliberação, à seleção do oscilador senoidal como a fonte mais adequada de material eletrônico. O princípio de análise de funções de Fourier permite expressar qualquer forma de onda em função de componentes senoidais de diferentes frequências, amplitudes e fases (ver seção II.2.i); e a chave para completar flexibilidade no processo reverso de síntese residia na provisão de equipamento capaz de gerar cada um desses componentes separadamente. Inicialmente, o estúdio de Colônia estava equipado com apenas um oscilador senoidal. O projeto original do estúdio, entretanto, oferecia três outras possibilidades de fontes eletrônicas. Duas delas eram instrumento eletrônicos, o *Melochord* e o *Electronic Monochord*. Ambos os instrumentos eram equipados com um par de teclados, cada um com uma saída monofônica independente. O *Melochord* gerava alturas derivadas da escala com doze notas do temperamento igual, com alavancas de transposição de oitavas expandindo a faixa de três oitavas para sete. Um pedal era utilizado para controlar o nível de volume e um envelope de ataque e decaimento gerado eletronicamente podia ser acionado para cada nota. O sistema de teclado do *Monochord* era mais flexível, sendo possível variar tanto o tamanho do intervalo quanto o compasso geral, usando controles separados. O volume também era controlado por um pedal, mas ao invés de um dispositivo gerador de envelope, ambos os teclados podiam ser configurados para produzir uma resposta dinâmica variando de acordo com a pressão exercida em cada tecla. O oscilador do teclado produzia ondas dente de serra, triangular e quadrada. Em seus estados puros, essas três ondas são bastante decepcionantes, pois produzem timbres que não possuem qualidades distintas. Apesar de matematicamente diferentes, a regularidade de sua série harmônica torna difícil distinguir perceptivamente entre essas fontes. Suas características só se tornam significantes quando as mesmas são sujeitas a processamento eletrônico. A técnica mais útil envolve a aplicação de filtragem para salientar certos componentes harmônicos e atenuar outras. Tanto o *Monochord* quanto o *Melochord* possuíam tais meios, mas sua aplicação era restrita a um

“alisamento” dos harmônicos superiores associados com o gerador de onda dente de serra, ao invés de estruturação do timbre. O resultado era uma consistência de qualidade apropriada a um instrumento de performance musical, mas altamente restritivo quando aplicado aos fins composicionais da *elektronische Musik*.

Tentativas como esta de fornecer métodos simplificados de criação de sons sintéticos levam muito rapidamente a um sacrifício de flexibilidade. Em Colônia, durante o começo da década de 1951-1960, a única solução aceitável para este dilema parecia ser a rejeição de qualquer técnica que impusesse restrições indesejáveis em favor de métodos de síntese mais básicos, não importando o quão laboriosa sua aplicação.

Outra fonte de material composicional disponível em Colônia era um gerador de ruído branco. Ruído branco consiste de uma sucessão aleatória de valores distribuídos segundo uma distribuição de probabilidades (uniforme ou gaussiana) e contendo todas as frequências do espectro audível, soando como um chiado. O ruído branco pode ser visto como a antítese da oscilação senoidal “pura” (ver seção II.3.ii). Como fonte de material, o ruído branco é de extremo interesse, pois é possível isolar certas bandas de frequência através de filtragem. É também possível manipular o resultado de modo a atingir desvios interessantes da progressão gradual de sons com altura em sons ruidosos. Essa técnica se tornou uma característica importante da *elektronische Musik* por fornecer uma base útil para criação de um contínuo entre ruído e tons puros.

O termo “mistura de notas” era reservado para combinações de tons senoidais cujas frequências não eram relacionadas harmonicamente com a fundamental. Esta tentativa de distinguir entre espectros harmônicos e inarmônicos não foi totalmente bem sucedida, pois as sutilezas existentes entre esses dois tipos de timbre têm uma zona de transição bastante larga. Stockhausen considerou o uso de todas essas fontes no estúdio de Colônia antes de selecionar o gerador senoidal como única fonte para algumas obras.

Processamento sonoro com velocidades variadas tinha sido uma ferramenta poderosa para o compositor de música eletrônica, e seu potencial era apreciado tanto pela escola da *elektronische Musik* quanto, como já mencionado anteriormente, pela *musique concrète*. Na esfera serialmente orientada da primeira, a própria consistência do processo em impor mudanças proporcionais em todas as características gravadas recebeu mais atenção. Em contraste, era o próprio grau de transformação que apresentava o maior desafio a Schaeffer. Dobrar a velocidade de reprodução não apenas reduz pela metade a duração dos eventos gravados, mas também reduz as variações de

amplitude e frequência em menor escala, tais como vibrato e o ataque e decaimento. Como Schaeffer descobriria logo, tocar gravações de sons naturais em várias velocidades podia produzir efeitos interessantes. O ataque “afiado” de uma nota de trompete, por exemplo, é percebido como uma sucessão ruidosa e irregular de elementos quando tocada a uma fração da velocidade de gravação original. A permanência surge da impossibilidade de se manipular tais características separadamente de outros aspectos do som. As transformações aparentes resultam de reações subjetivas a processos determinados cientificamente que afetam o material como um todo.

Em contraste ao processamento sonoro com velocidades variadas, as técnicas de síntese eletrônica fornecem ao compositor a opção de evitar essas relações fixas entre envelope, altura e duração; ao menos durante os estágios de preparação do material. Se uma forma dinâmica é aplicada a uma fonte eletrônica antes do processamento em velocidade variada, a situação não difere daquela na qual se aplica o envelope em uma fonte natural. Se, no entanto, uma fonte que oferece um espectro estável é sujeita primeiro à variação de velocidade e depois à forma dinâmica, qualquer forma de envelope temporal pode ser atingida para o material transposto. Estes exemplos breves de técnicas de *elektronische Musik* ilustram o tipo de processo envolvido em se relacionar princípios contemporâneos de estruturalismo com os equipamentos disponíveis no início para síntese eletrônica em Colônia.

O estúdio de Colônia era equipado com um filtro passa faixa de frequência central variável, oferecendo os modos de operação passa baixa e passa alta como opção. Este equipamento tinha aplicação limitada, pois somente uma operação simples de forma podia ser efetuada por vez. Ruído branco parecia ser a fonte mais versátil para este tipo de equipamento, devido a sua densidade espectral uniforme. Um banco de filtros afinado em oitavas musicais ou divisões das mesmas não divide o espectro em bandas iguais de frequência, pois altura é uma escala logarítmica, e não linear. A largura de banda absoluta de cada filtro aumenta em proporção no banco, dobrando a cada oitava. O uso de ruído branco, fornecendo energia constante por unidade de divisão de frequência, fará com que a saída de qualquer característica selecionada com o banco de filtros tenha uma tendência a salientar a região de altas frequências, que conterà proporcionalmente mais energia. Para superar essa dificuldade, muitos estúdios adquiriram geradores de “ruído rosa” como fonte alternativa. Ruído rosa é parecido com ruído branco, no entanto a amplitude decai a uma taxa constante inversamente proporcional à frequência (ou 3 dB

por oitava). Ou seja, as baixas frequências contêm mais energia que as altas. Ruído rosa é o resultado de ruído branco processado com um filtro passa baixa com uma queda de 3 dB por oitava começando das baixas frequências. O resultado é ruído que contém a mesma quantidade de energia por oitava. Esta característica do ruído rosa reproduz com maior fidelidade o espectro dos sons aos quais estamos expostos no dia-a-dia.

Uma outra área importante no processamento de sons refere-se às técnicas de modulação, duas das quais, modulação em amplitude e em frequência (também chamada modulação FM, descrita em maiores detalhes na seção II.3.iii), foram usadas no projeto dos primeiros estúdios como o de Colônia. Um tipo de modulador, conhecido como modulador em anel, foi especialmente utilizado no início e ainda o é nos dias de hoje. O produto de sua modulação é conhecido como tons de soma e diferença dos sinais de entrada. Um som com as frequências 100 e 200 Hz, por exemplo, modulado com um tom de 25 Hz, produzirá um espectro na saída contendo as frequências 75, 125, 175 e 225 Hz. A natureza matemática do processo resulta em uma gama bem distinta de timbres. Os produtos obtidos serão inarmônicos a não ser que a fonte dos tons seja relacionada em razões específicas de números inteiros. A habilidade de transformar espectros harmônicos em inarmônicos fornece uma técnica de estúdio útil. A utilidade da técnica, no entanto, é restringida pela sua própria natureza, pois as razões são determinações fixas. Pode-se variar as fontes, mas não o modo como interagem. Este processo é particularmente útil na produção de sons de sino, passível de modificação adicional por dispositivos como filtros. Com respeito a outras técnicas de modulação disponíveis em Colônia, a base para um tipo de modulador de amplitude existia no dispositivo de envelope anexado ao *Melochord*. O estúdio incorporou outro equipamento de modulação de amplitude que permitia que o timbre de um som variasse dinamicamente via o banco de filtros de oitava. Outro dispositivo disponível no instrumento alterava o perfil da forma de onda de maneira dinâmica.

Já foi feita referência ao uso de reverberação e eco por Stockhausen na preparação de algumas de suas peças. No início, em Colônia, os equipamentos para tal tipo de técnicas de realce consistia de uma sala com superfícies altamente refletoras equipada com um alto falante para reproduzir os sons e um microfone para captar os sons refletidos.

Devido ao extremismo com que alguns adotaram o dogma do uso unicamente de técnicas totalmente especificáveis de síntese, muitos dos primeiros trabalhos a emergir de

Colônia carregam este estigma. Reverberação fornece uma ferramenta valiosa para criar calor e profundidade no mundo clínico dos sons eletrônicos, e o desejo por tal amenização da impessoalidade dos resultados era mais forte que o compromisso com tais princípios deveras rígidos.

A projeção espacial de sons era um aspecto que recebeu cada vez mais atenção conforme o estúdio se desenvolveu. Além dos dispositivos versáteis de monitoramento dentro do estúdio, um sistema multi-canaís de alto falantes foi instalado no principal estúdio de gravação para recitais em salas de concerto. A possibilidade de alterações dinâmicas das localizações dos sons trouxe uma nova dimensão ao processo composicional, pois até aquele momento idéias de se utilizar o espaço em performance musical tinha quase sempre sido associada a distribuições estáticas dos instrumentos. Os engenheiros em Colônia logo perceberam o desejo por quatro alto-falantes e a decisão de localizar o quarto alto falante no teto constituiu uma notável introdução de movimento vertical na localização de imagens sonoras.

Os problemas práticos encontrados no uso de um único gerador senoidal como fonte exclusiva de material eletrônico levaram a considerável frustração, especialmente entre os jovens compositores em Colônia, menos inclinados a seguir as doutrinas rígidas de Eimert. A reavaliação do potencial de outras fontes sonoras gerou o maior desafio ao estúdio. Parecia claro para a nova escola que não havia razões *a priori* para lançar mão de princípios ao embarcar em uma composição eletrônica. Claramente, a diversidade de idéias composicionais preveniu a possibilidade de uma estética unificada para a música eletrônica. Havia uma tendência, no entanto, a adotar procedimentos hierárquicos ao se preparar e estruturar material sonoro, de tal maneira que os processos composicionais se tornaram cada vez mais ligados com a manipulação de complexos pré-definidos. Em música tradicional, os instrumentos são, de certa maneira, tratados como tais elementos, pois oferecem uma gama específica de características funcionais determinadas no momento de sua construção [50]. As doutrinas de Eimert devem ser elogiadas por seu desafio ao pressuposto de que o instrumento fornece a única base filosófica para composição. A mudança, então, se voltou a um ambiente de trabalho mais flexível no estúdio, onde os compositores eram livres para definir seus próprios termos de referência, escolhendo livremente dentre uma gama expandida de recursos básicos.

No início da década de 1961-1970 vários compositores tinham contribuído para o conteúdo artístico do estúdio de Colônia. Os títulos desses trabalhos dão uma indicação

do forte interesse em procedimentos ordenados serialmente em síntese sonora. O desejo de adotar abordagens mais flexíveis para o processo composicional, no entanto, motivou o movimento a se afastar do dogma rígido dos anos anteriores. Mudanças de perspectiva mais radicais, no entanto, preconizaram o momento da mudança na diretoria do estúdio de Colônia em 1963, quando Stockhausen substituiu Eimert. Foi iniciada, então, uma longa reforma no estúdio, incorporando equipamento de geração e processamento de sons em uma sala, e em outra um largo espectro de dispositivos de gravação e manipulação sonora, permitindo que a fonte de material gravado fosse o meio-ambiente. Em termos práticos, as técnicas da *musique concrète* e da *elektronische Musik* se tornaram formalmente reconhecidas como facetas complementares da música eletrônica.

I.4 Música Computacional¹

De acordo com Trevor Wishart [78] em seu livro “*On Sonic Art*”, desde o último quarto do século XX, “o divisor de águas na mudança de visão do que constitui música está mais intimamente ligado com a invenção da gravação musical, e posterior processamento e síntese, do que qualquer desenvolvimento na linguagem musical propriamente dita. Estes últimos desenvolvimentos expandiram amplamente nosso conhecimento da natureza dos sons e nossa percepção dos mesmos e contradisseram muitas das preconcepções do século XIX a respeito da natureza da percepção de altura e de sua relação com timbre (vide seções II.2.i), II.2.v), II.2.vi) e II.2.vii) para uma discussão detalhada). A tecnologia computacional nos oferece um controle dos parâmetros internos dos sons musicais extremamente detalhado, parcialmente realizando o sonho original dos primórdios da música eletrônica: a capacidade de esculpir todos os aspectos do som. No entanto, esta capacidade também acarreta uma mudança de paradigma sobre o que constituem sons musicais e, principalmente, o que não admite tal adjetivação. Não se pode mais estabelecer limites rígidos”. Wishart, por fim, conclui que, “no futuro, talvez se crie uma nova terminologia para os usuários de tais técnicas – projetistas sonoros ou engenheiros sonoros ao invés de compositor – uma vez que o termo compositor está fortemente associado à organização de notas no papel”.

¹ Do termo em Inglês *Computer Music*. Há muita controvérsia a respeito da utilização do termo Música Computacional em Português, sendo Computação Musical defendido por alguns como mais apropriado. O autor deste trabalho adotou a utilização do termo Música Computacional como padrão no texto.

O termo Música Computacional abrange uma grande variedade de atividades composicionais, variando da geração de partituras com notação convencional, com dados calculados pelo computador, à síntese digital de sons, prontos para conversão em sinais de áudio via conversor analógico-digital, amplificador e alto falante [50].

As primeiras tentativas de utilizar o computador como ferramenta de síntese sonora datam da década de 1951-1960. Nessa época o “*Bell Telephone Laboratories*” estava interessado nas possibilidades de se transmitir conversas telefônicas de forma digitalizada, através da conversão da informação analógica em padrões equivalentes de amostras numéricas de um lado da linha e realizando-se o processo inverso do outro. Foi nesse clima de investigação que um dos engenheiros da Bell começou a explorar o uso do computador como meio de se calcular e gerar amostras sonoras. A introdução dos circuitos transistorizados no fim da mesma década alterou radicalmente o panorama da tecnologia para computador, e no fim da década máquinas com maior capacidade de processamento e memória estavam aparecendo no mercado. Neste momento vários músicos e compositores mostravam interesse na utilização do computador como ferramenta de processamento de sinais sonoros. Com o advento dos circuitos digitais, fornecendo meios de se incorporar vários transistores miniaturizados em uma única pastilha, uma terceira geração de computadores surgiu.

Os princípios básicos do processamento digital de sinais já haviam sido esboçados, e a síntese sonora digital por computador já consistia dos seguintes passos: programação do computador para calcular as amostras numéricas correspondentes às características de variação de pressão em função do tempo da forma de onda; conversão das amostras em passos de tensão via conversor analógico-digital e, finalmente, filtragem e amplificação do sinal que é reconvertido em variações temporais de pressão atmosférica, ou seja, onda sonora (seção II.1). A fidelidade do resultado depende tanto da taxa com a qual as amostras são produzidas – a taxa de amostragem (seção III.1.i) – quanto da precisão numérica dos valores que representam as amostras nos diferentes instantes de tempo (seção III.1.i). A novidade introduzida no uso do computador digital para síntese sonora é a escala e flexibilidade dos processos de síntese à disposição do compositor, particularmente a habilidade de especificar sons refinados e precisamente ordenados através de uma sintaxe musicalmente ordenada, baseada nos conceitos musicais tradicionais. Do ponto de vista artístico, o potencial criativo dessas técnicas de síntese digital ainda deve ser explorado. As razões para esta latência são numerosas, sem contar

as dificuldades fundamentais experimentadas por muitos compositores ao trabalhar criativamente com o computador, um sistema que (ainda) demanda a especificação de idéias através de um sistema de codificação de caracteres alfanuméricos que oferece muito pouca oportunidade para experimentação interativa. Do ponto de vista do compositor, esta vasta gama de características provê grande flexibilidade na manipulação de muitos aspectos de tarefas de síntese e manipulação de qualidades sonoras.

O que afinal é tão significativamente diferente a respeito da música computacional? Os computadores permitem a investigação de assuntos puramente musicais de uma maneira que permite observações mais precisas que as disponíveis a “ouvido nu”. O computador representa um avanço tecnológico que tem uma aplicação clara no estudo da música de uma maneira que não era possível no passado. Os computadores permitem experimentação sonora precisa e passível de ser repetida. De fato, os músicos hoje podem projetar sons de acordo com as suas necessidades musicais ao invés de contar com um número reduzido de instrumentos musicais tradicionais. Além disso, os computadores agem como um instrumento de apoio à mente, estendendo os princípios da composição musical para a consideração simultânea de muitos mais níveis de organização do que é praticável sem esse apoio. Os computadores também ampliam as capacidades dos músicos de controlar a produção de sons durante apresentações ao vivo [55].

Música computacional é um dos campos de estudo mais interdisciplinares existentes, porque inclui aspectos significativos de arte, ciência e tecnologia [55]. Qual desses é enfatizado depende muito dos objetivos de cada um. Enquanto o objetivo for produzir música, música computacional é uma arte; se o objetivo é entender música como uma atividade humana, música computacional é definitivamente mais uma ciência. De um ponto de vista musical, os computadores transformam abstrações em percepções concretas. Os computadores são de interesse para os músicos, portanto, não por si próprios, mas porque eles permitem estabelecer uma ligação consistente entre imaginação musical e realidade.

I.5 Posicionamento da Pesquisa, Objetivos e Metodologia

De maneira ingenuamente simples, música computacional é a arte de fazer música com computadores digitais [50]. Uma vez que os computadores não têm uma função fixa no processo de fazer música, seu papel pode variar bastante. A maioria dos programas de música computacional se encaixa em três categorias: algoritmos para geração de sons (síntese sonora), programas que assistem o músico na tarefa de composição com sons sintetizados e/ou sons de instrumentos musicais tradicionais, e programas que permitem a apresentação da composição usando sons sintetizados. Aqui trataremos da primeira classe de programas. Estamos lidando com o problema da síntese sonora. Síntese sonora pode ser definida como a geração de um sinal musical que cause uma determinada sensação acústica [20]. Isto é, que contenha determinadas características desejadas de modo a permitir sua aplicação musical.

As perspectivas da música eletrônica e computacional são objeto de grande especulação devido à crescente popularização e interesse no uso do computador como ferramenta processadora de dados e à enorme difusão de arquivos de música e programas de manipulação sonora. A situação é tão volátil que em poucos meses surgem mudanças de impacto significativo. A necessidade de um entendimento mais profundo da natureza do som e das complexidades de sua especificação e percepção em termos musicais fornecem uma base sólida para os sistemas a serem desenvolvidos e/ou em desenvolvimento. Nos primeiros anos da música eletrônica, a escolha limitada e características restritas do equipamento do estúdio representavam as maiores restrições no seu uso musical e criativo. O campo da música computacional, no entanto, apenas recentemente começou a produzir resultados composicionais que vão de encontro ao nível mais modesto de expectativa dessa tecnologia tão robusta e flexível. Até hoje, a natureza altamente especializada da pesquisa nessa esfera levou inevitavelmente à predominância de cientistas ao invés de músicos nos grupos de pesquisa, afetando não somente a qualidade musical dos trabalhos, mas também o curso das melhorias nos sistemas. A equação foi rebalanceada, mas será necessário muito tempo para que a tendência atual de projeto de sistemas musicais seja favorável aos músicos e compositores, os usuários mais especializados de tal tecnologia.

Grande parte dos métodos e sistemas de síntese sonora digital utiliza modelos e/ou incorpora técnicas que não levam em conta a natureza dinâmica dos sons musicais ou que não foram originalmente desenvolvidas para manipulação musical. A síntese aditiva (seção II.3.i) é um exemplo de método que se baseia em um modelo bastante restrito do som musical, o modelo de Helmholtz [77]. Em seu trabalho mais célebre, Helmholtz concluiu que a descrição espectral de um som possui a correlação mais direta com seu timbre. Como resultado, quase toda técnica tradicional de síntese desenvolvida se destina à produção de um sinal com um conteúdo espectral específico, ao invés de uma forma de onda específica. Em contraste, a proposta deste trabalho é a manipulação de sinais que correspondam diretamente à forma de onda, isto é, sinais que podem possuir toda a complexidade inerente dos sons musicais. Mesmo que os métodos de síntese baseados em decomposição espectral sejam ampliados para dar conta (incorporar) de fenômenos de natureza dinâmica, esses revelam-se pouco práticos e de aplicação limitada em variadas situações, pois a complexidade do som é construída empilhando e, se necessário, variando os componentes individuais do som até que o resultado final seja atingido. Por exemplo, pode-se começar com um conjunto de harmônicos senoidais e, progressivamente, injetar aperiodicidade no som através da variação da frequência e amplitude dos harmônicos. Para Xenakis, esta abordagem baseada na análise de Fourier não basta para recriar a complexidade inerente do som. Ele prefere adotar uma abordagem global na qual a síntese sonora é realizada no domínio temporal somente, sem lançar mão de decomposição espectral. Ao invés de começar com um som periódico e modificá-lo (podendo incluir variações aleatórias), ele começa “de um conceito de desordem e, então, introduz meios para se aumentar ou diminuir essa desordem” [79]. Em outras palavras, Xenakis propõe começar com um som aperiódico, um sinal aleatório no qual diferentes graus de regularidade são injetados.

A abordagem que este trabalho segue é baseada na proposta por Xenakis. Estamos propondo uma abordagem populacional no domínio do tempo para a geração de sons musicais tendo a auto-organização como paradigma gerativo. A emergência de padrões complexos derivados do processo auto-organizável pode dar origem a sons que contenham as características desejadas, permitindo, assim, sua aplicação musical.

Dado um conjunto de atratores² (sons musicais contendo qualidades sonoras desejadas *a priori*) e uma população de indivíduos se comunicando por meio de interação local, é possível obter sons musicais que contenham qualidades sonoras advindas de mais de um dos atratores, representando variantes dos mesmos. Esta proposta envolve a aplicação de algoritmos bio-inspirados como portadores das características relevantes dos sistemas auto-organizáveis desejáveis nos resultados sonoros de qualquer processo de síntese que se proponha à geração de sons com aplicação musical. O fenômeno da auto-organização pode ser resumido como um processo no qual um padrão global emerge de inúmeras interações de componentes relativamente simples de um sistema, usando somente informação local, sem referência ao padrão global. A essência da auto-organização reside na emergência de estrutura e organização que não são impostas de fora do sistema³.

Em atividades de composição e improvisação musicais, uma vez formalizado o espaço composicional e definidos os elementos estruturais, a navegação pelo espaço composicional em computador conduz à composição de material sonoro com tendência a apresentar algum tipo de estruturação e forma. Neste trabalho, o processo de síntese sonora é interpretado como uma busca no espaço vetorial que contém todas as possibilidades sonoras decorrentes da representação adotada, e o objetivo é definido como a criação de formas de onda digitalizadas com características emergentes que tem o potencial de serem utilizadas em diversas aplicações musicais. A concepção da caracterização de uma região de interesse do espaço de busca pelos atratores e de um processo de busca que não contemple a interferência de um elemento organizador externo, e que pode resultar em padrões variantes e intersecções das características

² É importante esclarecer para o leitor que o termo atrator é definido na teoria de sistemas dinâmicos como sendo um conjunto de pontos (ou estados) para os quais o sistema converge em regime permanente, após o transiente. Neste trabalho, o termo atrator está sendo usado como o conjunto condicionante de pontos para o qual o sistema é atraído.

³ A esta altura, convém chamar a atenção para o fato de que os atratores, da maneira como foram definidos, podem ser interpretados como externos ao sistema, uma vez que é o usuário quem os define *a priori*. Sob essa mesma perspectiva, pode-se clamar que é o usuário quem inicia o processo, aqui chamado de auto-organizável, quando roda o programa. Está fora do escopo deste texto discutir a Teoria da Auto-Organização a fundo. Encarando-se o sistema como as interações de seus componentes depois de iniciado o processo, pode-se interpretar os atratores como sendo parte do sistema, e a necessidade de o usuário iniciar o processo como irrelevante, pois não faz parte do processo levado em consideração. É importante perceber que são os mecanismos presentes em processos auto-organizáveis e os padrões resultantes desses processos que são o foco de interesse aqui. Os algoritmos utilizados possuem características intrínsecas de sistemas auto-organizáveis. Não se está clamando que os sistemas e processos são legitimamente auto-organizados.

intrínsecas de cada um dos atratores, se apresenta como principal foco de interesse e maior motivação deste trabalho.

I.6 Organização do Texto

O capítulo II) trata dos fundamentos teóricos do problema, apresentando brevemente o modelo sonoro considerado, aspectos perceptivos das características do modelo, fundamentais para as aplicações musicais que podem decorrer do método, e aborda algumas das técnicas tradicionais de síntese sonora com o intuito de compará-las com a presente proposta. O capítulo III) apresenta a motivação por trás da abordagem de busca populacional em um espaço vetorial apresentada neste trabalho através do ponto de vista da otimização de funções, e explora o potencial da proposta através da inspiração biológica que encontra paralelo em sistemas simples que, no entanto, apresentam a emergência de padrões complexos. O capítulo IV) apresenta a aplicação de sistemas imunológicos artificiais em síntese sonora sob a luz da fundamentação teórica e da motivação apresentadas. A inspiração biológica e motivação, bem como a formalização do modelo, precedem os experimentos, resultados e a sua discussão. Em seguida, o capítulo V) discorre sobre a aplicação de redes neurais artificiais em síntese sonora. Novamente, a inspiração e o funcionamento do método são discutidos, seguidos da formalização do modelo. Por fim, os experimentos e resultados são apresentados, seguidos da discussão. Finalmente, o capítulo VI) discute as conclusões a respeito do método, compara-o com os métodos tradicionais apresentados anteriormente e apresenta as perspectivas futuras.

II) Fundamentos Teóricos

Torna-se necessário apresentar alguns fundamentos a respeito das formas de onda, de sons musicais e das características perceptivas do ouvido humano, sem os quais se tornaria obscura e vaga a apresentação e fundamentação das idéias que sustentam o método a ser descrito no decorrer do texto.

II.1 Modelo sonoro

Primeiramente, é importante ressaltar que, no decorrer do texto, o termo tom será usado para se referir a um som musical. Antes de começar a detalhar características sonoras físicas e psicoacústicas, torna-se crucial fazer distinção entre som e onda sonora. Neste texto, onda sonora trata do fenômeno físico, isto é, a propagação de uma perturbação no ar na forma de uma onda. Som, por sua vez, trata do fenômeno perceptivo. Acústica é o estudo da física e transmissão das ondas sonoras. Trata de conceitos tais como frequência, amplitude, tempo, espectro e localização sonora. Psicoacústica é o estudo da maneira como os seres humanos percebem sons. A preocupação aqui é com a resposta subjetiva à onda sonora em termos de altura (*pitch*), intensidade percebida (*loudness*), duração, timbre e localização aparente.

A onda sonora é uma onda mecânica, isto é, um padrão de perturbações que necessita de um meio material para se propagar. A onda sonora é produzida por uma fonte vibratória. As vibrações perturbam as moléculas de ar que estão próximas da fonte empurrando e puxando-as alternadamente em sincronia com as vibrações da fonte. Portanto, a energia sonora produz regiões no ar em que a pressão é menor que a média (rarefações) e regiões em que a pressão é maior que a média (compressões). Essas regiões de rarefação e compressão no ar se propagam na forma de uma onda sonora. A Figura II.1.1 ilustra a fonte vibratória e as regiões de compressão e rarefação se propagando no ar na forma de uma onda mecânica longitudinal. Diferentemente das ondas eletromagnéticas, que são transversais, a onda sonora é longitudinal. Isto é, a direção de propagação é a mesma da perturbação. Quando a onda sonora atinge uma superfície, faz com que a mesma vibre em simpatia com a onda sonora. Desta maneira,

energia acústica é entregue da fonte para o receptor pela onda sonora, a qual retém o padrão de vibração característica da fonte.

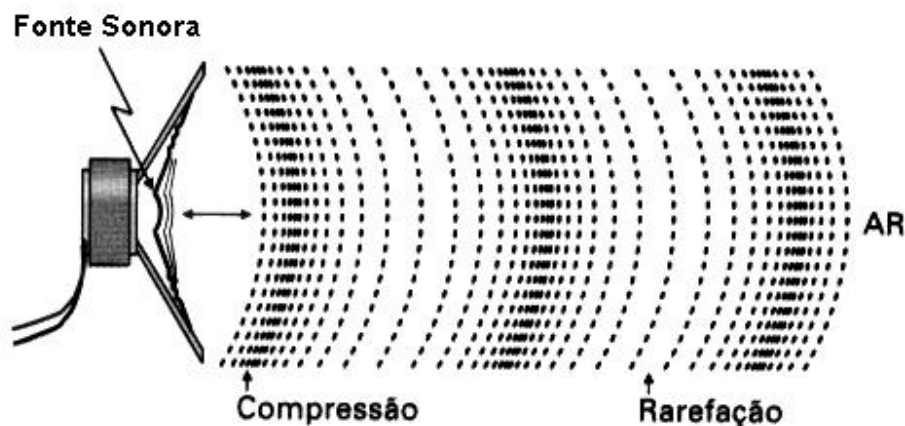


Figura II.1.1: Representação de uma fonte sonora e da onda sonora resultante.

Note-se que o padrão vibratório da fonte provoca regiões de pressão mais alta que a média (compressão) e de pressão mais baixa que a média (rarefação) se propagando no ar na forma de uma onda mecânica longitudinal.

O padrão de variação de pressão ao longo do tempo produzido por uma fonte sonora é conhecido como forma de onda. A Figura II.1.2 ilustra uma forma de onda periódica de uma oscilação senoidal. Quando a forma de onda está na região de pressão negativa (abaixo do eixo horizontal) há rarefação, caso contrário há compressão. Características fundamentais das formas de onda estão representadas na Figura II.1.2. Primeiramente, deve-se notar que a forma de onda apresenta um padrão que se repete, por isso é chamada de periódica. A menor unidade do padrão que não apresenta repetição é um ciclo. O tempo necessário para completar um ciclo é chamado de período. A taxa na qual os ciclos de uma forma de onda periódica se repetem é chamada frequência, normalmente medida em Hertz, ou ciclos por segundo. A faixa de frequências que o ouvido humano alcança varia aproximadamente de 20 Hz a 20 kHz. A frequência apresenta correlação com a percepção de altura (*pitch*, do original em inglês), de modo que sons percebidos como mais altos, isto é, mais agudos, têm maior frequência. A seção II.2.i) se dedica à complexa relação entre a frequência de uma onda sonora e a correspondente percepção de altura.

Amplitude é o quanto a pressão do ar varia em função dos ciclos de compressão/rarefação das moléculas de ar. A amplitude é uma medida da máxima pressão atmosférica causada pela perturbação que originou a forma de onda. A Figura

II.1.2 ilustra a amplitude da forma de onda mostrada. É um indicativo da quantidade de energia acústica de uma onda sonora e é um fator crucial na percepção de intensidade (*loudness*, do original em inglês). Certamente, uma caracterização mais significativa da sensação auditiva associada à amplitude é a intensidade de uma onda sonora. Intensidade é a medida da potência sonora que entra em contato com uma certa área, tal como o tímpano. Potência é uma medida da taxa com que a energia é entregue por uma fonte por unidade de tempo. A faixa de intensidades que o ouvido humano percebe varia de 10^{-12} W/m^2 , o limiar de audibilidade, até 1 W/m^2 , o limiar da dor.

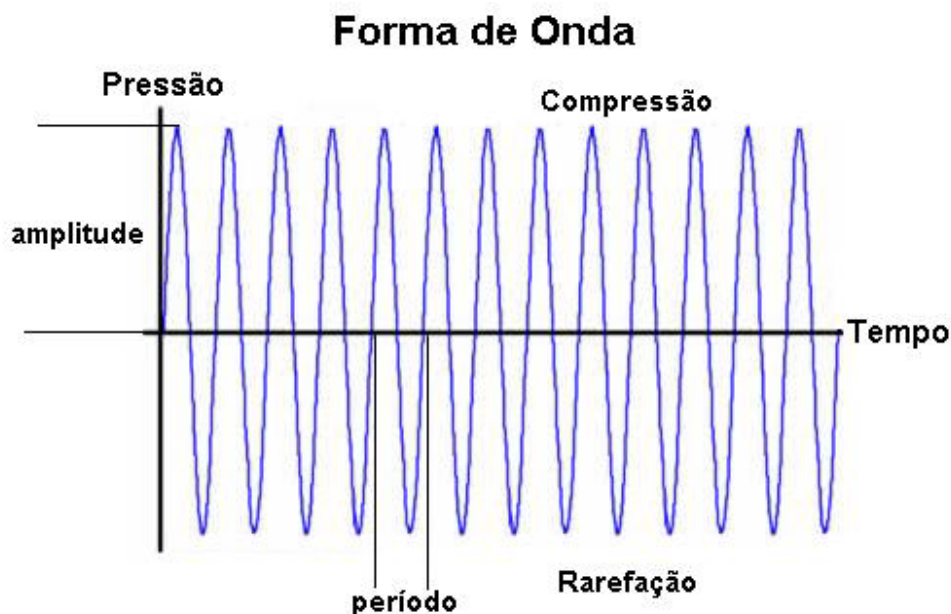


Figura II.1.2: Representação de uma onda sonora como uma forma de onda.

Uma forma de onda é uma medida da variação da pressão atmosférica ao longo do tempo. As regiões de compressão e rarefação estão mostradas, bem como algumas características físicas fundamentais para se caracterizar a forma de onda, tais como o período e a amplitude do padrão vibratório.

A forma do ciclo de uma forma de onda periódica tem um grande efeito na maneira como a mesma soa, tal como explorado na próxima seção.

II.2 Percepção Psicoacústica

A onda sonora é recebida pelo ouvido e processada pelo sistema auditivo, que se estende da orelha até o córtex cerebral. O ouvido faz o papel de transdutor, convertendo

o movimento mecânico dos padrões vibratórios da onda sonora em impulsos nervosos. A Figura II.2.1 ilustra as partes importantes do ouvido.

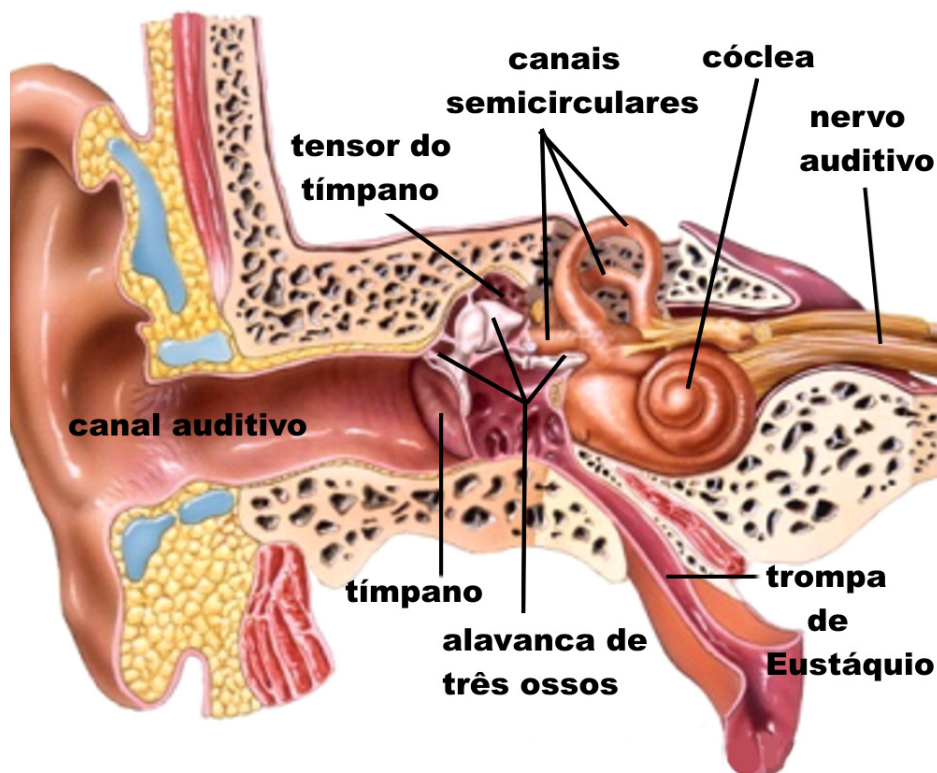


Figura II.2.1: Ouvido humano representando os dispositivos empregados no processamento e respectiva conversão da onda sonora em estímulo sonoro.

A onda sonora é captada pela orelha e viaja pelo canal auditivo na forma de vibração mecânica até o tímpano, o qual transmite as vibrações para os ossículos, uma cadeia de três ossos, o martelo, a bigorna e o estribo. O estribo transmite as vibrações para uma membrana chamada de janela oval, fazendo com que uma onda se propague pelo fluido na cóclea, a qual, por sua vez, estimula os cílios na membrana basilar.

Quando a onda sonora atinge o ouvido, ela viaja pelo conduto auditivo até o tímpano. A onda sonora transmite seu padrão vibratório característico para o tímpano, o qual transmite as vibrações para os ossículos, uma cadeia de três ossos, o martelo, a bigorna e o estribo. O estribo transmite as vibrações para uma membrana chamada de janela oval, fazendo com que uma onda se propague pelo fluido na cóclea, a qual, por sua vez, estimula os cílios na membrana basilar. Neste ponto a conversão do estímulo mecânico em elétrico está completa. Nervos ligados a esses cílios transmitem o estímulo para o córtex cerebral, constituindo a percepção de estímulo sonoro.

Quando vibrações senoidais com frequências dentro do limiar de audibilidade chegam ao ouvido interno, elas excitam as terminações nervosas da membrana basilar em regiões distintas, com localizações proporcionais à frequência do estímulo. A sensação de intensidade (*loudness*) está primariamente ligada à quantidade de energia acústica recebida pelo ouvido.

II.2.i) Altura

Percepção de altura é a resposta subjetiva à frequência. É a principal sensação causada pelas terminações nervosas na membrana basilar. A faixa aproximada de frequências audíveis é entre 20 Hz e 20 kHz. As frequências entre 20 Hz e 4 kHz constituem a região de maior acuidade perceptiva e sensibilidade a mudança de frequência. Esta região ocupa dois terços da membrana basilar, as altas frequências restantes equivalem ao terço restante [20]. Dentro da região de maior acuidade, a discriminação de frequência para senóides é proporcionalmente pior para baixas frequências e se torna cada vez mais difícil conforme a duração diminui. É independente da amplitude.

Para um ouvinte detectar uma mudança de frequência de um tom senoidal simples (contendo apenas uma frequência), a mudança deve ser maior que a “mínima diferença discriminável⁴” (MDD) em frequência. É postulado por alguns que a MDD em frequência corresponde a uma separação mínima espacial entre pontos de estímulos na membrana basilar [20]. O valor exato da MDD, assim como tantas medidas psicoacústicas, varia de ouvinte para ouvinte e de acordo com o método de medição. Para o ouvinte médio, o valor da MDD para senóides é 3% em 100 Hz e 0,5% em 2000 Hz [67].

Há uma relação não linear entre a frequência do estímulo e a percepção de altura. O estímulo na membrana basilar ocorre em pontos quase exatamente proporcionais ao logaritmo da frequência na base 2. Isto é, pelo menos para altas frequências, ao dobrar a frequência percorre-se uma distância aproximadamente constante entre os pontos de estímulo na membrana basilar. Portanto, os ouvintes comparam tons baseados no intervalo musical que os separa – isto é, baseado na razão das frequências – ao invés de na diferença entre suas frequências. Por exemplo, o intervalo correspondente à oitava corresponde à razão 2:1. Este intervalo é percebido como o mesmo independente da região onde ocorre no contínuo de frequências.

⁴ “Just noticeable difference” (JND) do original em inglês.

Na prática, os ouvintes recebem ondas sonoras cuja forma de onda é muito mais complexa que a onda senoidal simples mostrada na Figura II.1.2. Quando um som composto de múltiplas senóides chega ao ouvido, o fluido da cóclea responde com um movimento que é uma combinação dos padrões de vibração dos componentes de frequência. Neste caso, a altura ou as alturas percebidas depende(m) não apenas dos valores das frequências que constituem a onda, mas também das relações entre as mesmas.

Uma relação possível entre as frequências é a relação harmônica, isto é, as relações entre as frequências podem ser expressas por números inteiros. Neste caso, mesmo que cada frequência estimule uma localização diferente na membrana basilar, a percepção será de uma única altura correspondente à frequência fundamental, como representado na Figura II.2.2.

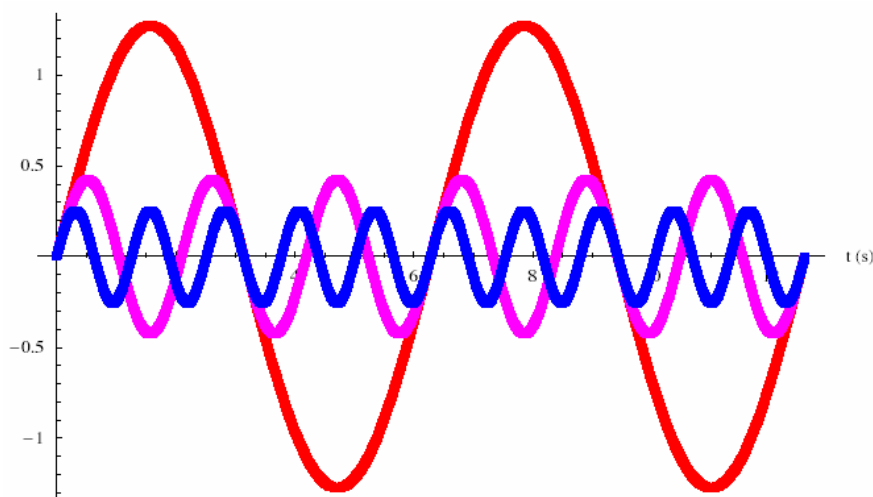


Figura II.2.2: Decomposição de Fourier de uma onda quadrada como uma soma de oscilações senoidais harmonicamente relacionadas.

Apenas a fundamental e as duas primeiras harmônicas são apresentadas.

Note-se que, na Figura II.2.2, as frequências das oscilações senoidais se relacionam por múltiplos inteiros. A menor frequência presente no sinal, representada pela onda de maior amplitude, é chamada de frequência fundamental. As outras frequências presentes no sinal são chamadas de harmônicos, cada um correspondendo a um múltiplo inteiro da frequência da fundamental. Quando as múltiplas frequências estão harmonicamente relacionadas, a forma de onda resultante é periódica, como mostrado na Figura II.2.3. Pensa-se que periodicidade é um fator primordial na percepção de uma

altura definida. Na região de aproximadamente 150 Hz a 500 Hz, há fusão dos harmônicos na sensação de uma altura definida mesmo quando a fundamental está ausente [77]. Uma teoria postula que nosso mecanismo de escuta determina o padrão de repetição da onda sonora no tempo e determina o período. O ouvinte aceita a frequência correspondente ao período medido como sendo a fundamental. Neste caso, mesmo com a fundamental faltando, o padrão de repetição levaria à medição da fundamental através da determinação do período de repetição, isto é, as parciais restantes continuam a formar um padrão de repetição na frequência do maior divisor comum, a fundamental.

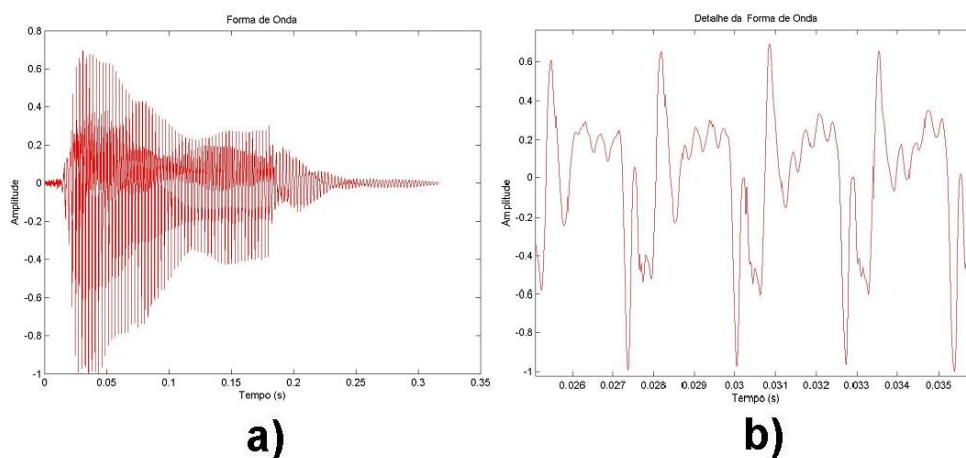


Figura II.2.3: Periodicidade de uma forma de onda.

Tons que evocam a sensação de altura, como o de trompete mostrado na parte a) da figura, possuem espectros quase harmônicos, o que é equivalente a possuir uma forma de onda aproximadamente periódica, como salientado na parte b), a qual corresponde a um pequeno intervalo da forma de onda da parte a).

Também pode haver a situação em que dois tons senoidais que não são harmonicamente relacionados chegam ao ouvido com a mesma amplitude. Se a razão entre as frequências dos dois tons for suficientemente grande, dois tons distintos serão percebidos. Se a razão entre as frequências é muito pequena, o ouvinte percebe uma única altura correspondente à média das duas frequências. No entanto, tons com frequências muito próximas produzem uma sensação de batimento a uma razão que é a diferença entre as frequências dos dois tons originais. Como exemplo, considere-se a composição resultante de duas oscilações senoidais com frequências, respectivamente, de 1440 e 1442 Hz. Devido à interferência construtiva e destrutiva entre as ondas, forma-se um padrão de batimento a uma razão de 2 Hz, a diferença entre as frequências.

Batimento é ouvido como modulação em amplitude, ou seja, uma mudança periódica na amplitude do tom resultante, como mostrado na Figura II.2.4.

Um fenômeno parecido com o batimento entre dois tons fundidos ocorre quando dois tons são afinados próximo de uma oitava. O batimento ocorre a uma razão correspondente à diferença entre a frequência do tom perto da oitava e a frequência exatamente uma oitava acima do tom mais baixo. Este fenômeno, chamado batimento de consonâncias desafinadas, desaparece acima de 1500 Hz [55].

Batimento entre os tons pode ocorrer, apesar de muito mais fraco, quando a diferença entre os dois tons ocorre próximo ou no intervalo da quinta pura (3:2 em razão de frequência) ou da quarta pura (4:3 em razão de frequência). Este fenômeno é usado para afinar instrumentos musicais.

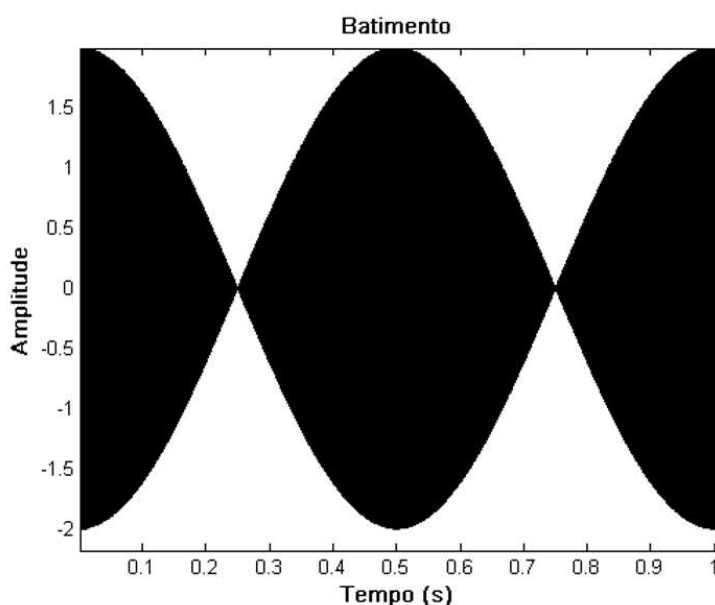


Figura II.2.4: Representação do fenômeno de batimento.

O fenômeno de batimento ocorre quando dois sons com frequência próximas interferem um no outro. O batimento é percebido como modulação em amplitude a uma taxa correspondente à diferença entre as frequências dos sons. A figura mostra a composição resultante de duas oscilações senoidais com frequências, respectivamente, de 1440 e 1442 Hz. A frequência da oscilação resultante é de 2 Hz.

A inclusão de frequências parciais inarmônicas em uma onda sonora garante que, na maioria das circunstâncias, serão ouvidos batimentos. Batimento pode, por sua vez, ajudar o tom a soar mais vivo.

Dois tons puros causam a sensação de batimento até que a diferença entre suas frequências exceda aproximadamente 10 a 15 Hz. Dentro da região de batimento existe o limite de discriminação, um ponto em que a separação de frequência de dois tons é grande o suficiente para que o ouvinte perceba dois tons diferentes. Para separações um pouco maiores que a região de batimento, há uma região chamada de aspereza tonal. Esta região é associada com a banda crítica. A banda crítica é uma medida da habilidade do ouvido para discriminar tons adjacentes. Por exemplo, quando o ouvido recebe dois tons muito pouco espaçados, haverá uma sobreposição dos neurônios estimulados na membrana basilar. Isto causa a sensação para estes dois sons de uma complexidade maior do que no caso de dois outros com uma maior separação. A diferença de frequência para a qual a sensação muda abruptamente é conhecida como banda crítica.

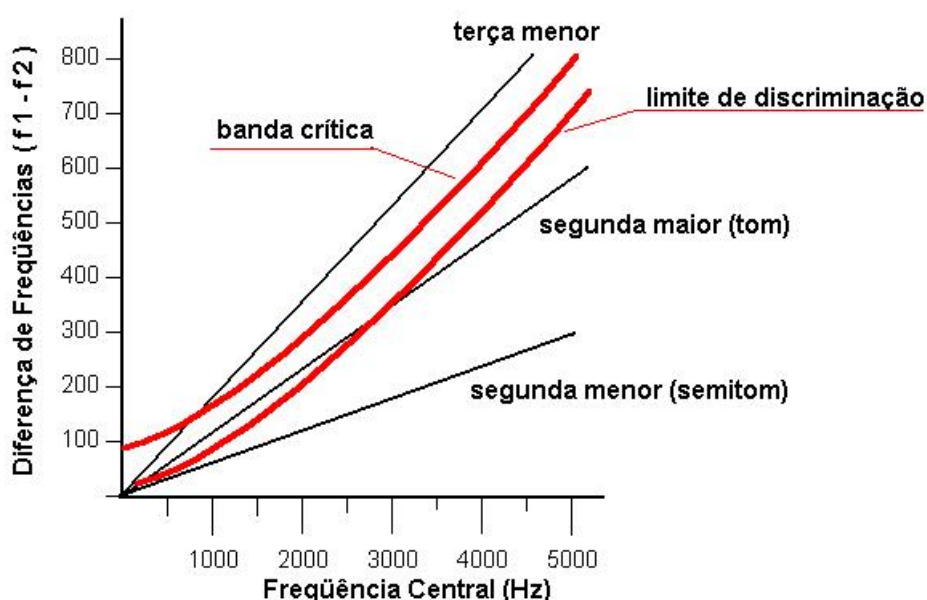


Figura II.2.5: Banda Crítica de discriminação de altura.

Estão mostradas as retas correspondentes aos intervalos musicais terça maior, segunda maior e menor. A figura mostra também a curva correspondente à banda crítica e ao limiar de discriminação entre dois tons, um ponto em que a separação de frequência de dois tons é grande o suficiente para que o ouvinte perceba dois tons diferentes. Para separações um pouco maiores que a região de batimento, há uma região chamada de aspereza tonal. Esta região é associada com a banda crítica.

A largura da banda crítica, representada no eixo das ordenadas pela diferença entre as frequências dos tons, varia com a frequência central do tom, como mostra a Figura II.2.5. É uma grande porcentagem da frequência média de dois tons em baixas

freqüências, e uma pequena porcentagem para tons em altas freqüências. O fato de que a banda crítica é maior em baixas freqüências ajuda a explicar o uso comum de intervalos musicais maiores no registro baixo na maioria das músicas. A banda crítica tem um papel crucial na percepção de amplitude e timbre também.

II.2.ii) Percepção de Intensidade

A intensidade percebida de um som é uma medida da resposta subjetiva à amplitude. A intensidade percebida de um som é fortemente influenciada por sua freqüência, bem como por sua composição espectral. Portanto, a menor mudança de amplitude detectável – a Mínima Diferença Discriminável (MDD) de amplitude – depende tanto do conteúdo espectral quanto da amplitude do tom. Em geral, nas faixas de freqüência e amplitude musicalmente relevantes, a MDD em amplitude para tons senoidais é entre 0,2 e 0,4 dB [67]. A Figura II.2.6 mostra a comparação das MDDs para tons senoidais em 70 Hz, 200 Hz e 1000 Hz.

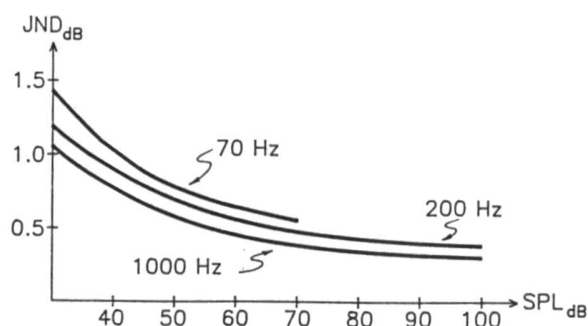


Figura II.2.6: Mínima Diferença Discriminável para Nível de Pressão Sonora.

A figura mostra a dependência da Mínima Diferença Discriminável (JND ou *Just Noticeable Difference*, em inglês) em amplitude com o nível de pressão sonora (SPL em inglês) e com a freqüência. Retirada de Moore [55].

A Figura II.2.7 mostra as curvas de “Fletcher-Munson” [23]. As linhas representam os níveis de amplitude para os quais os tons senoidais de freqüências diferentes são percebidos como tendo a mesma intensidade. Por exemplo, a Figura II.2.7 mostra que próximo do limiar de audibilidade, para que dois tons puros (senoidais) de 100 Hz e 1000 Hz soem igualmente altos, a amplitude do tom mais baixo (grave) deve ser aumentada quase 40 dB. A figura mostra que o ouvido é mais sensível ao som na faixa de 250 Hz a aproximadamente 3500 Hz e que abaixo de 60 Hz e acima de 10 kHz a perda de sensibilidade é considerável.

Quando é apresentado a um ouvinte mais de um tom, a intensidade percebida varia com as relações de frequência e amplitude dos tons. Isto é um resultado do mascaramento: a redução à sensibilidade a amplitude devido à fadiga dos neurônios na membrana basilar. Portanto, na presença de um tom mais intenso, um menos intenso pode não ser ouvido de todo. O limiar de mascaramento é definido como o nível de amplitude para o qual o tom menos intenso desaparece. Ele depende bastante da diferença das frequências dos tons, da amplitude do tom mais intenso, e, em menor escala, da frequência do tom mais intenso. O mascaramento é particularmente efetivo quando os tons ocupam a mesma banda crítica. Em muitos casos, um tom que tem uma amplitude de apenas 15 a 20 dB menor que um mais intenso não é percebido.

Quando os tons são próximos o suficiente em frequência para causar batimento, o limiar de mascaramento é levemente reduzido. Para dois tons intensos fora da mesma banda crítica, o tom de frequência mais baixa mascara o tom mais agudo mais facilmente. Tons de baixa intensidade fora da mesma banda crítica geralmente não mascaram um ao outro [2].

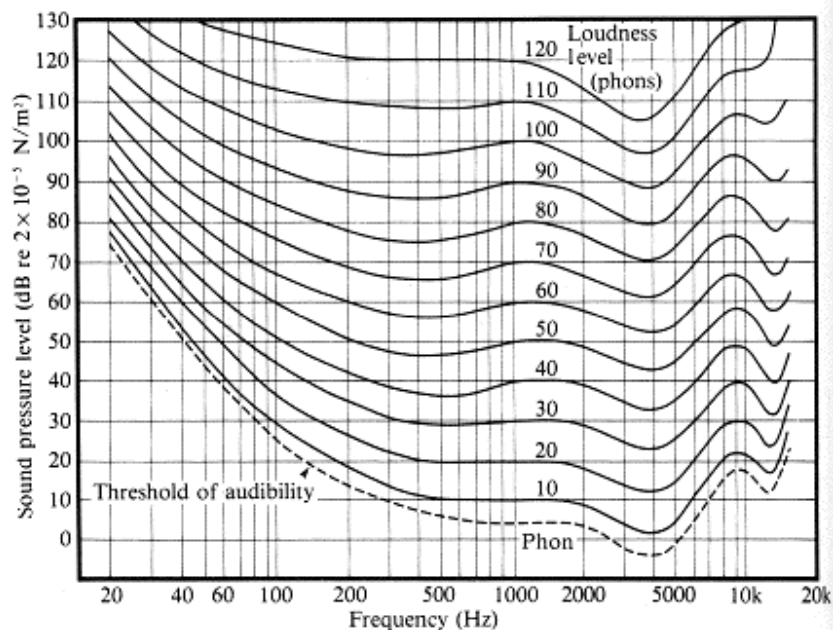


Figura II.2.7: Curva de Fletcher-Munson.

As linhas representam os níveis de amplitude para os quais os tons senoidais de frequências diferentes são percebidos como tendo a mesma intensidade.

Mascaramento não ocorre apenas para tons simultâneos, mas também para tons sucessivos. Em muitas circunstâncias, a fadiga do neurônio causada por um tom anterior intenso pode afetar nossa percepção de um tom subsequente menos intenso.

Quando um tom puro (senoidal) é recebido com intensidade suficiente, o ouvido cria “harmônicos aurais” não presentes no tom. Por exemplo, um tom de 1 kHz intenso pode produzir sensação em 2 kHz, 3 kHz, etc. Além disso, um par de tons (com frequências f_1 e f_2 , com $f_2 > f_1$ e $2f_1 > f_2$) fazem com que o sistema perceptivo adicione outras sensações tonais, combinação de tons, que são combinações das frequências presentes. A sensação adicional mais forte é um tom com a diferença das duas frequências ($f_2 - f_1$). Para tons complexos, ocorrem sensações nas diferenças das frequências das fundamentais e em todos seus harmônicos. Os produtos mais notáveis são definidos pelas relações $2f_1 - f_2$ e $3f_1 - f_2$. A sensação é mais fraca para harmônicos superiores. Por exemplo, as sensações adicionais mais significativas produzidas por dois tons complexos intensos com 150 Hz e 190 Hz, respectivamente, são ouvidas em 40 Hz ($f_2 - f_1$), 110 Hz ($2f_1 - f_2$), 230 Hz ($2f_2 - f_1$), 70 Hz ($3f_1 - 2f_2$) e 270 Hz ($3f_2 - 2f_1$) [20].

II.2.iii) Relações Temporais

O tempo é um limitante fundamental na habilidade de percepção de altura. Quando um tom é soado, é necessário algum tempo para que o ouvinte desenvolva a sensação de altura. A duração deste tempo depende da frequência do tom. Para que uma altura seja percebida, um ouvinte deve receber uma certa quantidade mínima de ciclos de um tom. Portanto demora mais tempo para perceber a altura de um tom de frequência mais baixa porque o tom apresenta um período maior. Por exemplo, um tom deve durar pelo menos 40 ms em 100 Hz, enquanto que em 1000 Hz o tom deve durar apenas 13 ms para ser percebido [2].

Quando um ouvinte ouve uma seqüência de tons, há algumas maneiras distintas que a mesma pode ser percebida. Estas diferentes maneiras de percepção têm conseqüências composicionais importantes. A maneira que uma seqüência é percebida depende de alguns fatores: a taxa na qual os tons são soados, o intervalo de altura entre os tons, diferenças de amplitude e diferenças timbrais entre os tons, dentre outras. O determinante mais importante é o contexto composicional no qual estão inseridos. Apesar de os resultados dos testes científicos, tais como os apresentados a seguir, serem úteis para determinar princípios perceptivos gerais, os resultados podem variar com a experiência e expectativas dos ouvintes. Por exemplo, um tema ou motivo em um

trabalho, uma vez bem estabelecido, pode ser apresentado de maneiras que normalmente se pensa que tal tema resultaria melodicamente incompreensível, e ainda assim ser compreendido. Os trabalhos eletrônicos de Milton Babbitt contêm exemplos deste efeito [55].

Nos exemplos que se seguem, serão consideradas seqüências tocadas em um único instrumento com pouca variação de dinâmica. Isto permitirá que se mostre mais claramente o efeito tanto da taxa na qual a seqüência está sendo tocada, quanto do intervalo de altura entre os membros da seqüência. As três maneiras básicas de se perceber uma seqüência são [20]:

- Como uma única linha;
- Como dividido em múltiplas linhas;
- Fundido em um único som.

No primeiro caso, em que a seqüência é ouvida como uma única linha, diz-se que a seqüência evoca coerência temporal. O segundo caso, em que o ouvinte particiona a seqüência em duas ou mais linhas distintas é conhecido como fissão. O último caso, em que a seqüência é tocada tão rapidamente que a mesma forma uma única entidade audível, é conhecido como fusão.

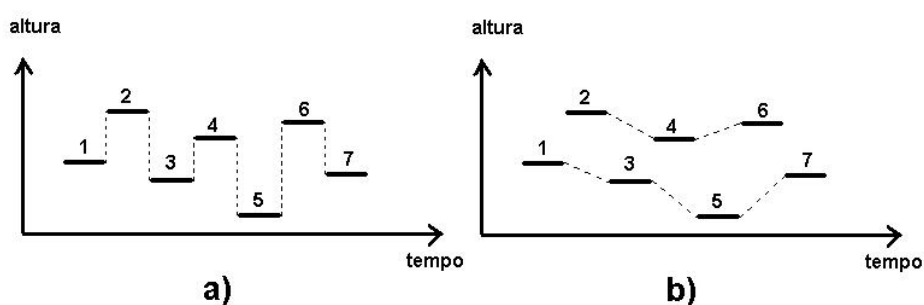


Figura II.2.8: Percepção de altura de uma seqüência de tons.

A parte a) mostra a seqüência intercalada a uma taxa de 5 a 6 tons por segundo percebida como temporalmente coerente. Quando a mesma seqüência é tocada a uma taxa de 10 a 12 tons por segundo, como mostra a parte b), pode ocorrer a percepção de fissão.

Considere-se uma linha melódica com sete tons onde os tons pares e os tons ímpares estão em registros distintos. Como mostrado na Figura II.2.8 a), a uma taxa de 5 ou 6 tons por segundo, um ouvinte ouviria essa seqüência como uma sucessão coerente temporalmente. A uma taxa rítmica mais alta – 10 a 12 tons por segundo, como mostrado

na Figura II.2.8 b) – o grupo de tons mais agudos se junta formando uma linha separada dos tons mais graves. A uma taxa rítmica intermediária, por volta de 7 ou 8 tons por segundo, pode-se direcionar a atenção para qualquer uma das três percepções: uma sucessão de tons graves e agudos se alternando, uma linha de tons agudos somente, ou somente uma linha de tons graves. A uma taxa rítmica ainda maior (maior que 20 tons por segundo), o ouvinte perceberá a linha como a fusão em uma sonoridade complexa.

O tamanho do intervalo entre tons adjacentes na seqüência também afeta a velocidade com a qual a sensação de coerência dá lugar à de fissão. No caso de um trinado, quando o intervalo separando os dois tons alternados é pequeno, próximo de meio passo, a seqüência evoca coerência temporal até que a taxa de alternância é acelerada a mais de aproximadamente 20 tons por segundo. Ouve-se, então, a fusão dos dois tons em um único. No entanto, com uma seqüência de tons que consiste de intervalos musicais grandes na faixa de uma 12^a, mesmo a velocidades moderadas o ouvinte consegue separar os tons em secessões por registro.

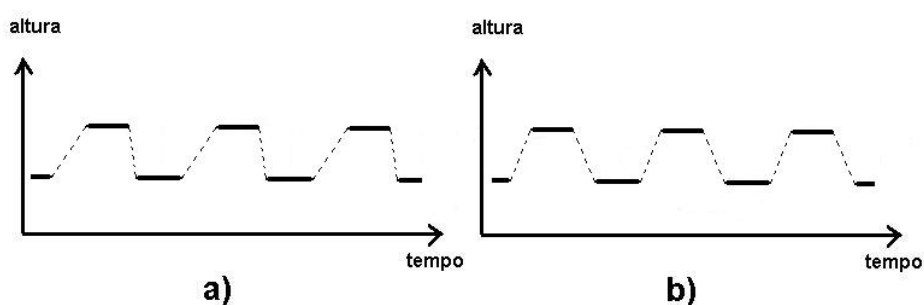


Figura II.2.9: Percepção de simetria de uma seqüência de tons.

A altas velocidades e intervalos melódicos grandes, seqüências de tons com ritmos desiguais, como mostrado na parte a), são percebidas como simétricas, como a seqüência mostrada em b).

A alternância de dois tons a altas taxas também pode produzir outros efeitos. Considere-se a rápida alternância de dois tons separados por um intervalo grande. Mesmo quando os tons mais agudos não estão equidistantes no tempo entre os mais graves, os tons agudo e grave soarão igualmente espaçados. A Figura II.2.9 ilustra esta situação. Este efeito se perde para intervalos de altura pequenos.

Quando uma seqüência rápida de dois tons é tocada com um pequeno intervalo, um passo ou meio passo, a ordem dos dois tons é facilmente percebida. No entanto, quando os intervalos são aumentados bastante, a situação se complica. Para alguns ouvintes, os

tons soam simultâneos, enquanto que para outros, a ordem dos dois tons não é clara. A altas velocidades, onde os tons pulam entre registros por grandes intervalos, o ouvinte perde o efeito de velocidade constante. A passagem parecerá consistir de subfrases procedendo a taxas diferentes.

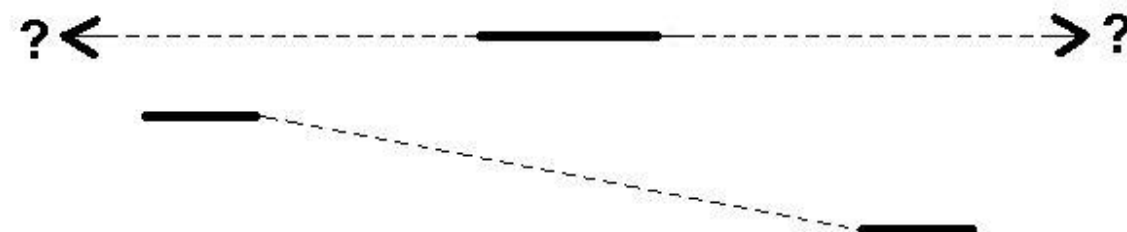


Figura II.2.10: Relações temporais.

A grandes intervalos e velocidades rápidas, pode ser difícil discernir onde, na seqüência de três tons, a nota aguda começa.

O efeito pode ser ainda mais ambíguo em uma seqüência de três tons. A Figura II.2.10 ilustra seqüências muito rápidas de três tons nas quais um largo intervalo de registros separa um tom dos outros dois. Mesmo que o ouvinte perceba os dois tons no mesmo registro como pertencendo ao mesmo grupo, pode não ser possível determinar a ordem temporal do outro tom.

II.2.iv) Teoria Clássica de Timbre

O conceito de timbre está relacionado à resposta perceptiva subjetiva a qualidades do tom que permitem a identificação de sua fonte. Timbre é muito mais difícil de caracterizar do que intensidade percebida ou altura devido ao fato de ser um fenômeno bastante diverso. Não há uma escala unidimensional – tal como grave/agudo para altura – postulada para timbre, pois parece não existir um par simples de opostos entre os quais pode-se estabelecer uma escala. Devido ao fato de timbre ter tantas facetas, técnicas computacionais de escalamento multidimensional [29], [30] constituíram o primeiro grande progresso em uma descrição quantitativa de timbre desde o trabalho pioneiro de Hermann von Helmholtz no século XIX.

Hermann von Helmholtz criou a base para os estudos modernos de timbre em seu livro *“On the Sensations of Tone”* [77]. Este trabalho contém uma inestimável quantidade de conceitos e fundamentos necessários ao estudo de timbre. Helmholtz caracterizou os tons como consistindo de uma forma de onda envolta em um envelope de amplitude que consiste de três partes: o ataque, o estado estacionário e o decaimento, como mostrado

na Figura II.2.11. Durante o ataque do tom, a amplitude cresce de zero até o seu pico. Durante o estado estacionário, a amplitude é idealmente constante. Durante o decaimento, o som esmaece. Ele concluiu que os sons que evocam uma sensação de altura possuem formas de onda periódicas e descreveu-as como possuindo formas fixas e invariantes no tempo. Helmholtz também estabeleceu que a natureza da forma de onda tem grande impacto no timbre percebido de um som.

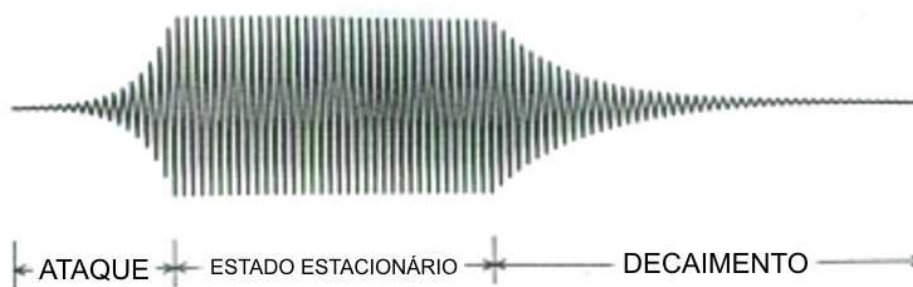


Figura II.2.11: Modelo simplificado de um tom musical de Helmholtz.

Os três segmentos que, de acordo com o modelo, caracterizam o tom são mostrados. Deve-se notar que o ataque, o estado estacionário e o decaimento se referem à forma de onda. Adaptada de Dodge [20].

Para determinar quais características melhor se correlacionam com timbre, ele lançou mão do trabalho de Jean Baptiste Fourier, que demonstrou que qualquer forma de onda periódica pode ser expressa como a soma de uma ou mais senóides. Como exemplo, a Figura II.2.2 mostra os três primeiros harmônicos da decomposição de Fourier de uma onda quadrada como uma soma de oscilações senoidais harmonicamente relacionadas. Somando-se as três formas de onda ponto a ponto, obtêm-se uma aproximação de uma onda quadrada de período igual ao da fundamental, isto é, da senóide de menor frequência.

Fourier mostrou que toda forma de onda periódica é composta por um conjunto único de senóides cujas frequências estão relacionadas harmonicamente. Portanto, qualquer forma de onda periódica também pode ser descrita em termos de seus componentes senoidais. Cada componente senoidal é caracterizado por três parâmetros: frequência, amplitude e fase relativa à fundamental. Os dois primeiros parâmetros têm papel fundamental na percepção de timbre. No entanto, as relações de fase entre as senóides têm apenas efeito mínimo. Dá-se o nome de espectro à descrição de uma forma de onda através de seus componentes senoidais. Os componentes espectrais são às vezes referidos como parciais da forma de onda. No caso de o espectro ser harmônico,

pode-se referir aos componentes como parciais harmônicos, ou simplesmente harmônicos.

Helmholtz concluiu que a descrição espectral de um som possui a correlação mais direta com seu timbre [77]. Como resultado, quase toda técnica tradicional de síntese desenvolvida se destina à produção de um sinal com um conteúdo espectral específico, ao invés de uma forma de onda específica (a proposta deste trabalho é a manipulação de sinais que correspondam diretamente à forma de onda). Por exemplo, a descrição qualitativa de brilhante para timbre caracteriza espectros que têm uma grande quantidade de energia em altas frequências. O espectro da maioria dos instrumentos da família dos metais exibe esta tendência. Sons com grandes quantidades de energia em harmônicos altos soam como um zumbido. Um espectro com pouca ou nenhuma energia nos harmônicos pares caracteriza o timbre específico que é produzido pela clarineta em seu registro baixo. A maioria dos sons percussivos possui espectros que não estão nem próximos de serem harmônicos. Por exemplo, o som ressonante de um sino é o resultado de seu espectro altamente inarmônico. Adicionalmente, pesquisas mostraram que muitos instrumentos acústicos que evocam a sensação de altura exibem espectros levemente inarmônicos, ou seja, as parciais superiores são levemente desafinadas de harmônicos exatos. Isto causa a sensação de batimento no tom, contribuindo para a vivacidade do som [55].

O envelope espectral de um som é um dos fatores mais importantes na determinação do timbre [20]. O envelope espectral evidencia o padrão de distribuição de energia no espectro. A Figura II.2.12 ilustra um espectro e seu envelope correspondente, contornado pela linha pontilhada.

Examinando-se os envelopes espectrais das formas de onda mais frequentemente encontradas, tanto em síntese digital como em sons de instrumentos acústicos, percebe-se que elas são limitadas em banda. Isto é, há uma frequência acima da qual os tons não possuem uma quantidade significativa de energia acústica. É claro que o ouvido também é um receptor limitado em banda, pois pode converter energia apenas dentro de uma certa faixa de frequências. A largura de banda de um som é a largura da região de frequências na qual os componentes significativos de um som complexo residem – uma dentre muitas maneiras possíveis e úteis de se classificar um espectro. Espectros são usualmente caracterizados pela inclinação com a qual a energia no espectro diminui com a frequência, isto é, a inclinação do envelope espectral. Max Mathews e J. R. Pierce

observaram que “instrumentos musicais normais tendem a produzir um espectro que diminui mais rápido que 6 dB/oitava, mas não tanto quanto 12 dB/oitava” [53].

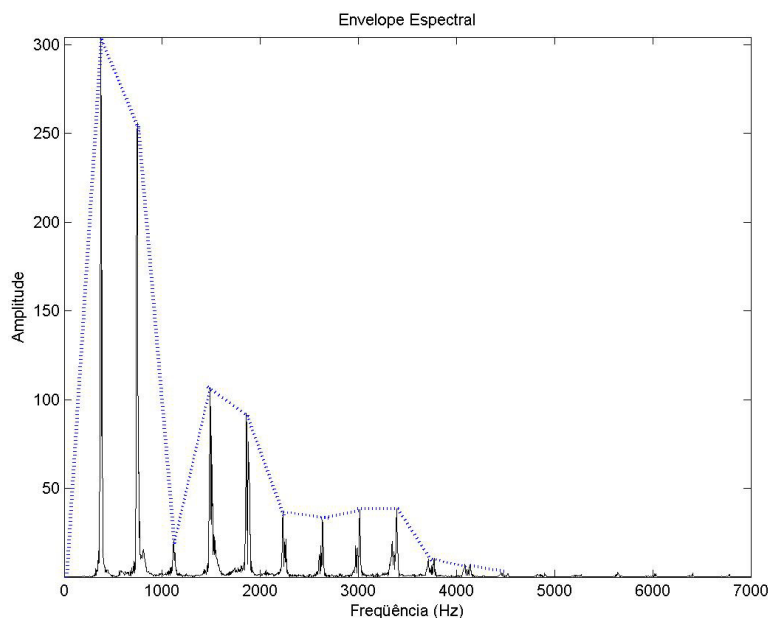


Figura II.2.12: Envelope espectral do espectro de Fourier de uma forma de onda.

A figura ilustra o padrão de distribuição da energia no espectro de uma forma de onda através do envelope espectral, evidenciado pelo pontilhado conectando os picos das parciais que constituem o espectro.

Para sintetizar sons, necessita-se de uma compreensão mais detalhada das relações entre freqüência fundamental, espectro e timbre. Se dois tons com o mesmo envelope de amplitude e aproximadamente a mesma freqüência fundamental têm relações idênticas entre suas componentes espectrais, seus timbres soarão similares [20]. No entanto, se dois tons com freqüências fundamentais diferentes têm a mesma distribuição de energia entre suas parciais, normalmente não se afirmará que esses tons possuem o mesmo timbre. Nossa percepção de similaridade timbral é fortemente baseada na presença de energia espectral em bandas de freqüência absoluta. Dois tons com freqüências fundamentais diferentes que julga-se ter o mesmo timbre não apresentarão as mesmas relações entre as amplitudes dos componentes espectrais. Ao invés disso, seus envelopes espectrais exibirão picos de energia nas mesmas regiões de freqüência. Dadas outras características parecidas, esses tons serão julgados como tendo timbres parecidos. Os picos espectrais em regiões de freqüência absoluta são chamados formantes. A Figura II.2.13, a seguir, ilustra o conceito.

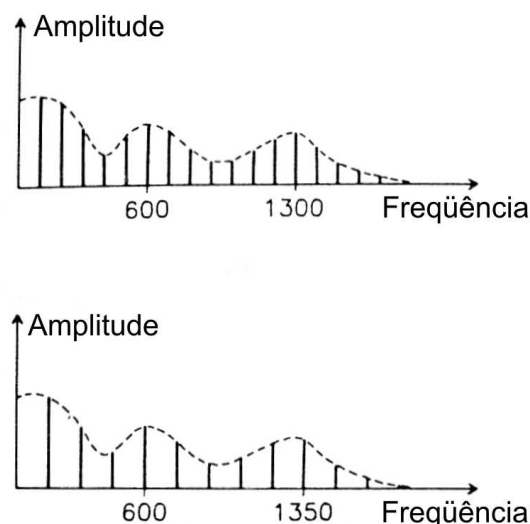


Figura II.2.13: Picos de energia em regiões de frequência absoluta, ou formantes.

A figura ilustra os espectros de Fourier de dois tons de alturas (notas) diferentes provenientes da mesma fonte sonora. Deve-se notar que as relações entre as frequências dos parciais não é o que é preservado, mas sim a presença de energia em certas regiões de frequência absoluta (em Hz), isto é, os formantes; no caso da figura, há um pico em baixas frequências, outro próximo de 600 Hz e um último em torno de 1300 Hz em ambos os espectros. Retirada de Moore [55].

A voz humana e a maioria dos instrumentos acústicos exibem formantes em seus espectros. Seus sons característicos resultam de um sistema que consiste de uma fonte de excitação, tal como as cordas vocais ou uma palheta vibrando, e um sistema ressonante, tal como o trato vocal ou o corpo de um instrumento. A cavidade ressonante faz com que as amplitudes de parciais em certas regiões de frequência sejam realçadas, enquanto que outras sejam atenuadas. A Figura II.2.14 ilustra a situação. Isto provoca picos nas regiões de frequência característica da cavidade ressonante, relacionadas ao tamanho, formato e material da cavidade. Os mesmos picos estarão presentes, com maior ou menor predominância, em todos os tons do instrumento, independentemente da frequência fundamental da fonte. Isto explica porque, em muitos instrumentos, tons diferentes produzidos no mesmo instrumento possuem uma qualidade/coloração similar. A técnica de síntese por modelamento físico segue esta idéia através do uso de equações diferenciais acopladas para modelar a resposta em frequência de cada uma das diferentes partes de um instrumento musical.



Figura II.2.14: Ilustração de uma fonte excitatória e uma cavidade acústica ressonante.

A presença de formantes nos sons de instrumentos musicais e na voz humana pode ser explicada através do modelo da fonte excitatória e da cavidade ressonante que modela o espectro, aumentando certas faixas de frequência, correspondentes aos modos naturais de vibração, e atenuando outras.

Picos formantes possuem uma relação com as bandas críticas de percepção de frequência. Harmônicos adjacentes acima da quinta aproximadamente estão na mesma banda crítica; portanto, o ouvinte pode julgar apenas a energia acústica presente em uma região geral do contínuo de frequência. As amplitudes dos harmônicos individuais dentro de uma mesma banda crítica se combinam para produzir a percepção de uma quantidade total de energia presente naquela banda. Portanto, quando é apresentado um espectro rico a um ouvinte, esse será, normalmente, incapaz de detectar a presença de um harmônico alto (não a fundamental) específico; apenas de notar sua contribuição para a quantidade total de energia encontrada em uma banda crítica específica.

Como mencionado anteriormente, cada componente em um espectro harmônico é caracterizado não somente por uma amplitude e uma frequência, mas também por uma fase medida em relação à fundamental. Por exemplo, em um espectro composto por componentes co-senoidais com a mesma fase, quando a fundamental atinge seu pico, todos os harmônicos também estão no pico. Alterando-se a fase de um dos harmônicos, tomando-se uma senóide ao invés de co-senóide, por exemplo, esse harmônico em específico estará no zero quando todos os outros estiverem em seu máximo. Esta mudança de fase afetará o formato da forma de onda. No entanto, Helmholtz observou que mudando-se a fase das parciais obtém-se, via de regra, um efeito mínimo na qualidade/coloração percebida do tom, ainda que o formato da forma de onda seja radicalmente diferente [77]. Estudos usando computadores para gerar formas de onda com espectros de fase arbitrária confirmaram que mudanças na fase dos componentes

produzem uma pequena mudança timbral (em comparação com o resultado de mudanças nas relações de amplitudes entre as parciais). A maior diferença em timbre é entre tons complexos contendo só senos ou só co-senos e aqueles contendo senos e co-senos alternadamente [20].

II.2.v) Estudos Modernos de Timbre

O modelo de Helmholtz do som musical, uma forma de onda fixa envolta em um envelope, representa o trabalho mais significativo na pesquisa em acústica musical no século XIX. Desde então, os pesquisadores determinaram um modelo mais preciso do som musical. Primeiramente, estudos recentes permitiram identificar quatro regiões distintas no curso da evolução temporal de uma forma de onda típica de um instrumento musical, em contraste com o modelo simplificado de Helmholtz do envelope espectral com três regiões distintas mostrado na Figura II.2.11. Pode-se identificar o ataque, o decaimento, a sustentação e o relaxamento, mostrados na Figura II.2.15, como o envelope temporal de grande parte dos tons musicais.

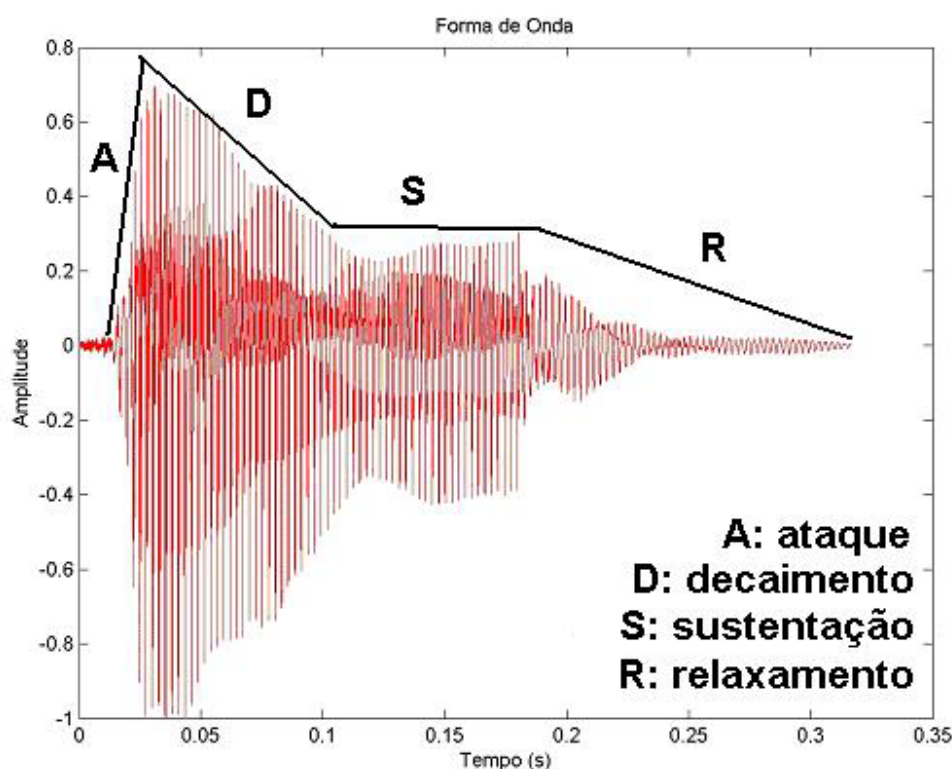


Figura II.2.15: Envelope temporal de uma forma de onda de um tom musical.

A gravação digital permitiu ao pesquisador moderno mostrar que a forma de onda e, portanto, o espectro, pode mudar drasticamente durante o curso de um tom. Uma técnica computacional que permite obter o espectro de Fourier de um sinal digital é a chamada Transformada Discreta de Fourier (TDF). Em seu *“Computer Study of Trumpet Tones”* [64], Jean Claude Risset utilizou um algoritmo baseado na TDF para obter informação sobre a evolução espectral ao longo do curso do tom – isto é, no tempo. De forma simplificada, ao invés de aplicar a transformada de Fourier ao estado estacionário do tom, o tom foi analisado através de curtas TDFs sucessivas ao longo de seu curso, possibilitando visualizar o som como uma sucessão de espectros de curta duração. A Figura II.2.16 ilustra a progressão temporal das amplitudes das parciais de um tom, aqui denominada de espectro dinâmico.

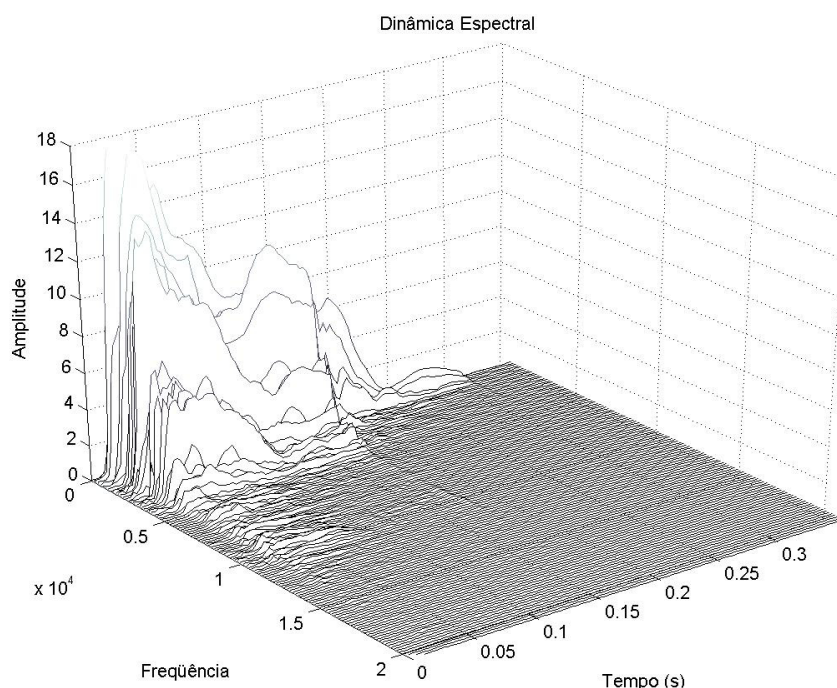


Figura II.2.16: Espectro dinâmico de um tom musical.

O espectro dinâmico mostra a progressão temporal das parciais ao longo do curso da forma de onda. Note-se que não há um único par de parciais com o mesmo envelope temporal.

Como exemplificado na Figura II.2.16, Risset descobriu que cada parcial do tom possui um envelope temporal de amplitude único. Isto claramente contrasta com o modelo de Helmholtz, no qual os envelopes de todos os parciais têm o mesmo formato.

Risset tirou as seguintes conclusões de seus estudos: o espectro de um tom de trompete é próximo de harmônico; os harmônicos mais altos tornam-se mais ricos conforme a intensidade aumenta (dinâmica); há uma flutuação em frequência que é rápida, tem amplitude pequena e é quase aleatória; harmônicos sucessivamente mais altos têm subidas mais lentas até a amplitude máxima durante o ataque; e há um pico formante em aproximadamente 1.500 Hz.

Em estudos seguintes, Risset observou que a evolução temporal do espectro do trompete tem papel fundamental na percepção do timbre característico do instrumento. Outros pesquisadores, seguindo os passos de Risset, fizeram uso sistemático de análise por computador para classificar a evolução espectral dos tons de uma variedade de instrumentos. James Moorer e John Grey publicaram análises por computador mostrando a evolução temporal do espectro de um tom de violino [56], de clarineta, de oboé [57] e de trompete [58]. Em adição à progressão de amplitude, as análises determinaram a variação em frequência de cada parcial do tom. Descobriu-se que as parciais são raramente harmônicas. A frequência de cada parcial flutua ao longo do curso do tom, e essa flutuação pode ser bastante errática durante o ataque do tom. Resíntese do tom sem as flutuações em frequência produziram uma mudança perceptível na característica do tom.

A progressão espectral mostrada na Figura II.2.16 contém uma grande quantidade de detalhes – pequenas flutuações na amplitude de um dado harmônico. Podem estes dados ser simplificados conservando a característica do tom analisado? Para responder, Grey [29], [30] aproximou as variações no envelope temporal de cada parcial por segmentos de reta, utilizando aproximadamente oito segmentos por envelope. Usando esses dados simplificados, ele foi capaz de sintetizar tons de trompete que foram julgados virtualmente idênticos aos originais. Portanto, essas pequenas variações no envelope não contribuem significativamente à percepção de timbre.

É importante ter em mente que re-síntese usando os dados originais (ou aproximações desses) só é efetiva dentro de uma faixa limitada de frequência. Por exemplo, um tom baseado nos dados, mas uma oitava acima, normalmente não causará a mesma sensação timbral. Similarmente, mudanças na duração normalmente criam um novo timbre. Além disso, padrões de evolução espectral diferem notavelmente para diferentes dinâmicas, isto é, intensidades diferentes [20].

Quando se apresenta um conjunto de componentes espectrais a um ouvinte, esse pode ou não fundi-los perceptualmente em um único som. Um dos fatores determinantes é a chamada assincronia de ataque do espectro [30], que se refere às diferenças nos tempos de entrada dos componentes espectrais. Grey e Moorer determinaram que a assincronia de ataque é tipicamente da faixa de 20 ms [30]. As flutuações em frequência das várias parciais também são normalmente necessárias para que as parciais se fundam perceptualmente em um único tom [20].

Timbre é aparentemente um fenômeno multidimensional [20], [62], [27]. A percepção timbral envolve correlacionar um número de fatores do tom, incluindo a natureza do ataque, o conteúdo harmônico e a afinação dos parciais. Até certo ponto, a amplitude, altura e aspectos temporais contribuem conjuntamente para a caracterização timbral. Diferentes pesquisadores já sugeriram conjuntos de dimensões independentes (descorrelacionadas) para classificação timbral. John Grey estudou timbre através da gravação de 14 tons de diferentes instrumentos normalizados em altura, dinâmica (intensidade) e duração [29]. Isto é, os únicos graus de liberdade possíveis são aqueles relacionados à percepção timbral dos tons de acordo com a definição da ANSI [1]. A seguir, ele apresentou todos os possíveis pares para ouvintes, pedindo que eles estimassem perceptualmente o quão relacionados os tons aparentam ser. Foi então aplicada uma técnica de escalamento multidimensional [44] no espaço tridimensional nos dados obtidos das avaliações subjetivas dos ouvintes, resultando em um mapeamento tridimensional do espaço timbral. O resultado é mostrado na Figura II.2.17. Deve-se ressaltar que é possível usar outros números de dimensões. O espaço timbral é um mapeamento da correlação dos tons. Tons que aparecem próximos no espaço timbral indicam que os ouvintes julgaram possuir uma proximidade em termos timbrais.

II.2.vi) Sons musicais complexos

Muito já se falou a respeito de sons musicais complexos, tons puros, sons naturais, sem, no entanto, uma definição formal e uma diferenciação entre esses. Primeiramente, sons (ou tons) naturais são aqueles que podem ser achados na natureza, ou seja, são produzidos por fontes acústicas. Em contrapartida, todo som que é produzido artificialmente (por fonte não acústica), por meios analógicos ou digitais, independentemente de suas qualidades, é denominado artificial ou sintetizado. A complexidade do tom é diretamente proporcional ao número (e natureza) das parciais em seu espectro. Tons senoidais puros, isto é, sem flutuação em amplitude ou em frequência,

são chamados sons puros. Quanto maior o número e mais errática a flutuação, tanto em amplitude quanto em frequência, mais complexo o tom.

Em última instância é a percepção e julgamento individual do ouvinte que definirá a complexidade de um tom. No entanto, um espectro dinâmico em que as parciais possuem envelopes temporais únicos com flutuações de pequena amplitude e desvios de frequência de natureza estocástica são características de tons musicais complexos o bastante para que haja a percepção de fusão das parciais, caracterizando, assim, o objeto de estudo deste trabalho.

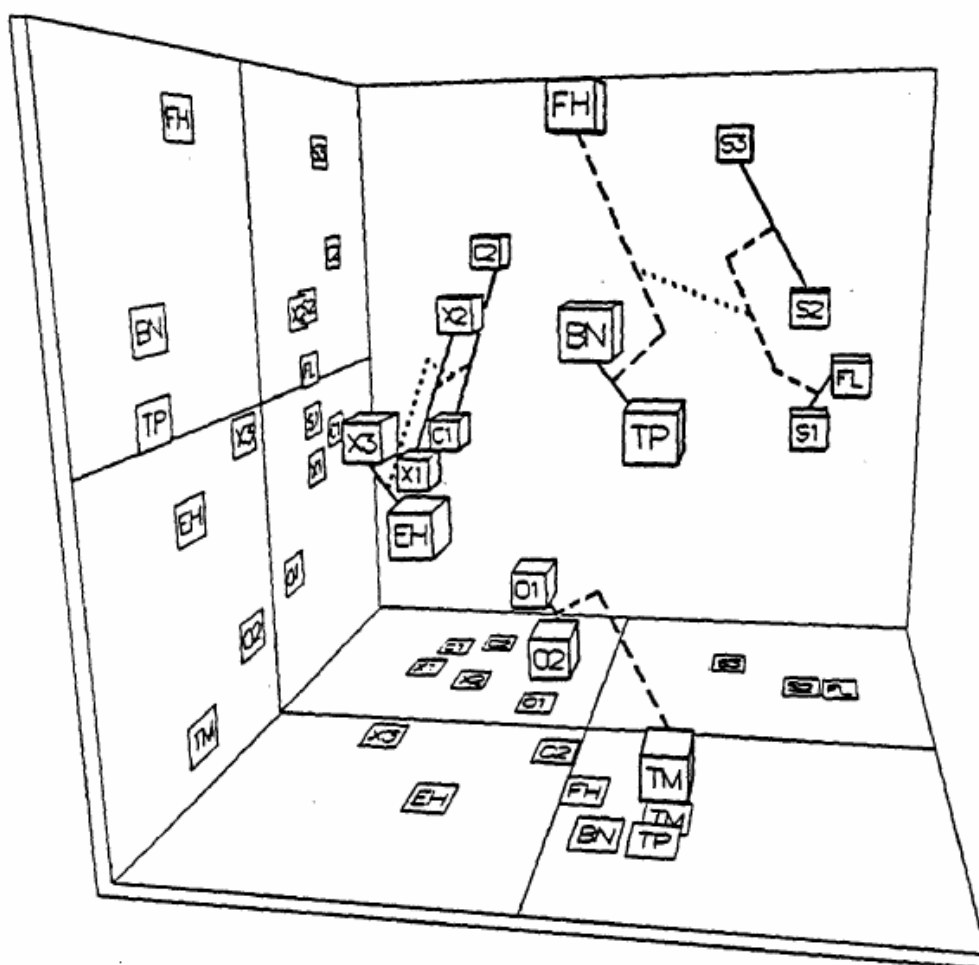


Figura II.2.17: Escalamento multidimensional de Grey de 14 instrumentos musicais.
A figura mostra a configuração tridimensional resultante da aplicação de uma técnica de escalamento multidimensional. Note-se as projeções bidimensionais nos planos correspondentes à parede e chão. Retirada de Grey [29].

Sons musicais são notavelmente complexos e difíceis de gerar devido a todas as características detalhadas nas seções anteriores, ressaltadas nos domínios temporal e frequencial na Figura II.2.18. Os resultados relatados a respeito da natureza das propriedades tempo-freqüenciais das formas de onda que caracterizam timbre oferecem informações essenciais para a síntese de sons com propósitos musicais através de computador. Poucos sons naturais assumem a forma do modelo clássico da forma de onda invariante dentro de um envelope temporal fixo. Nossos ouvidos são altamente seletivos e freqüentemente rejeitam sons que são muito matematicamente perfeitos e estáveis [9]. Conseqüentemente, não é uma tarefa trivial sintetizar sons musicais complexos, independentemente do método empregado.

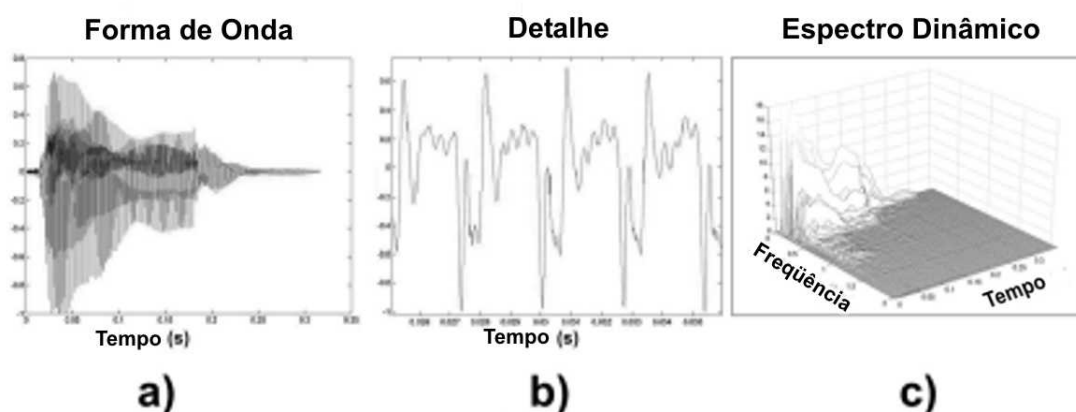


Figura II.2.18: Visualização da complexidade de um tom musical nos domínios do tempo e da freqüência.

Na parte a) vê-se a forma de onda e um detalhe da mesma na parte b). Note-se a periodicidade, característica de espectros harmônicos (sons com altura definida), e a flutuação irregular característica de tons complexos. A parte c) mostra o espectro dinâmico do tom, isto é, o espectro de Fourier instantâneo em momentos sucessivos do curso do tom. É importante destacar que cada freqüência parcial possui um envelope temporal único ao longo do curso do tom.

II.2.vii) Timbre Revisitado

O modelo clássico de Helmholtz de tom musical, de 1885, indica que a amplitude da vibração determina a intensidade percebida, e o período da vibração a altura. Helmholtz concluiu que a qualidade tonal (timbre) não pode, portanto, depender de nenhum dos dois [77]. A única hipótese possível para Helmholtz é que a qualidade tonal deve depender da maneira com que o movimento acontece dentro do período de cada vibração. No entanto,

Helmholtz se pergunta até que ponto as diferenças na qualidade musical podem ser reduzidas à combinação de diferentes tons parciais com diferentes intensidades em tons musicais distintos. De acordo com ele, em sua época, havia uma inclinação geral para creditar qualidade a todas as possíveis peculiaridades dos tons musicais que não fossem evidentemente devidas a intensidade e altura. Helmholtz declara que isto está correto, pois qualidade tonal era apenas vista como uma concepção negativa [77].

Fletcher declara, em 1934, que, apesar de ser consenso que intensidade, altura e timbre estão diretamente correlacionados à intensidade sonora, frequência fundamental e estrutura das parciais superiores, seus experimentos mostram que não há uma relação simples um para um. Sua sugestão de definição de timbre é: “aquela característica da percepção que permite que o ouvinte reconheça o tipo de instrumento musical produzindo o tom” [22]. É interessante notar que sua definição, apesar de vaga, já leva em conta a percepção do ouvinte como fator definidor de timbre. Plomp [61] afirma que o ouvido não é tão insensível a variações de fase como foi sugerido anteriormente e conclui que o efeito das variações de fase no timbre de um tom complexo é pequeno quando comparado com a variação das amplitudes do espectro.

Se timbre é aquela característica da percepção que permite ao ouvinte reconhecer o tipo de instrumento musical que está produzindo o tom, como pode haver variações timbrísticas em um instrumento devido a, por exemplo, variações de intensidade? As variações existem, o termo precisa ser refinado, como faz Seashore. Seashore conjectura que “a qualidade tonal tem dois aspectos fundamentais, a saber: timbre, que é a presença simultânea ou fusão da fundamental e seus parciais superiores em um dado momento; e sonância, a presença sucessiva ou fusão de variações timbrísticas, de altura e intensidade em um tom como um todo” [70]. De acordo com Seashore, timbre é aquela característica de um tom que depende de sua estrutura harmônica tal qual modificada pela frequência absoluta e intensidade total [70]. Seashore afirma ainda que, fisicamente, o timbre de um tom é a seção transversal da qualidade tonal no momento representado pela duração de uma vibração no som [70]. Ao definir sonância, Seashore traça um paralelo com a experiência de fusão que remete a movimento contínuo provocada pela rápida sucessão de imagens estáticas ao assistir a um filme. Seashore afirma que o mesmo princípio de fusão existe na percepção auditiva. Ao ouvir tons musicais, as ondas sucessivas vêm tão rápido que não podem ser ouvidas individualmente; isto é, não se pode separar o timbre, a altura ou a intensidade de uma onda e da próxima ouvida. O

resultado é que elas se fundem. Nesse contexto, timbre corresponde a uma única imagem instantânea e sonância à progressão de imagens sucessivas [70].

Em 1960, o Instituto Nacional Americano de Padrões (*American National Standards Institute* – ANSI) publica uma definição padrão de timbre que passa a ser amplamente aceita e utilizada: “Timbre é aquele atributo da percepção auditiva através do qual um ouvinte pode julgar se dois sons similarmente apresentados e que tenham a mesma intensidade e altura são distintos [1].” Apesar desta definição permitir a interpretação de que, aparentemente, timbre, frequência e intensidade sejam parâmetros independentes de um tom, alguns pesquisadores levantaram pontos a respeito da possível dependência perceptiva entre esses parâmetros que devem ser mencionados. Fletcher [22] declara que mudanças em intensidade ou altura sem mudanças na estrutura das parciais superiores produzem mudanças timbrais, concluindo que timbre depende, em princípio, das estruturas das parciais superiores, mas que grandes mudanças na intensidade e frequência também produzem mudanças no timbre. Esta foi, talvez, a primeira vez que se sugeriu que timbre, altura e intensidade não estão totalmente descorrelacionados. Licklider [48] declara que tanto a intensidade quanto a altura de um tom complexo são influenciadas até um certo ponto pela estrutura de parciais superiores. Licklider conclui que não se deve usar a noção mal definida de que timbre está relacionado com a distribuição ou padrão de alturas e intensidade na sensação tonal. Levarie e Levy [46], por outro lado, ao comentar sobre a relação entre intensidade e timbre, afirmam que, conforme a intensidade de um tom aumenta, parciais superiores que podem não ter estado presentes no tom anteriormente são estimuladas, e conforme a intensidade diminui, alguns parciais são amortecidos e esvaecem antes de outros. Eles concluem que qualquer mudança de pressão ou posição traz consigo uma nova constelação de parciais; e se uma nova altura ou um novo nível de intensidade é o propósito primário de tal mudança, timbre pode ser justificadamente tido como um atributo que caracteriza as outras propriedades de um tom [46]. Plomp [61] afirma que a dependência entre timbre e frequência implicaria que tons simples, por exemplo, tons senoidais, também são caracterizados por um timbre específico. De acordo com Plomp, tons senoidais graves soam suaves enquanto que os agudos são cortantes [61]. Plomp ainda afirma que a conclusão de que sons de diferentes alturas soam mais similares se seus envelopes espectrais correspondem não significa que não existam diferenças perceptivas além de altura [61]. De acordo com ele, o espaçamento dos parciais, determinados pela frequência da fundamental, é responsável pela dissimilaridade timbral de sons com alturas diferentes

mas envelopes espectrais similares. Plomp conjectura sobre o efeito no timbre da variação da frequência da fundamental do estímulo. Ele afirma que dependerá de o timbre ser determinado pelo envelope espectral relativo à fundamental ou pela posição de frequência absoluta do envelope espectral, sem relação com a frequência fundamental. Plomp afirma categoricamente que uma condição necessária para similaridade timbral de sons com diferentes frequências fundamentais é o envelope espectral ser parecido em termos de frequência absoluta, apesar desta similaridade ser necessariamente menor para diferenças maiores na frequência fundamental [61].

Slawson [71] dispensa o uso do termo timbre uma vez que, de acordo com ele, a definição é inadequada, apesar de amplamente aceita. Ele propõe a utilização do termo coloração tonal, e afirma que uma teoria da coloração tonal deveria ser capaz de especificar como preservar a cor sob mudanças de intensidade, altura e duração do som. Além do mais, Slawson declara que se a coloração tonal é um fenômeno complexo constituído de vários aspectos ou dimensões, deve-se ser capaz de mostrar como variá-los independentemente. Slawson admite, no entanto, que não há consenso sobre o que constitui um elemento de coloração tonal. Deve-se associar coloração tonal com um instrumento musical específico, por exemplo, um violino Stradivarius? Ou o som de todos os instrumentos do mesmo tipo deve ser tomado como elemento base da coloração tonal? Quando diz-se que um violino, qualquer violino, possui uma coloração tonal diferente do oboé, está-se usando o termo desta última maneira [71].

Rasch e Plomp [62] afirmaram que os sons não podem ser ordenados em uma única escala com respeito a timbre. Timbre é um atributo multidimensional da percepção de sons. Erickson afirma que timbre é um estímulo claramente multidimensional e não pode ser correlacionado com nenhuma dimensão física simples [21]. Muito esforço de pesquisa foi devotado a correlacionar dimensões timbrais perceptivas com características físicas dos tons. Gabrielson [26] afirma que uma grande dificuldade é achar rótulos verbais apropriados para cada uma das dimensões perceptivas. Conjectura também se há designações verbais adequadas na linguagem para timbre. Por fim, Gabrielson apresenta algumas sugestões de sua própria experiência de designações para algumas dimensões timbrais como sendo: Agudo X Suave, Brilhante X Opaco, Claro X Ruidoso e Compacto X Difuso. É interessante notar que todos os atributos são originalmente usados para se referir a aspectos perceptivos visuais ou táteis. Schouten [69], por outro lado, afirma que timbre pode ser expresso em termos de cinco parâmetros, a saber: a faixa entre o caráter

tonal e ruidoso; o envelope espectral; o envelope temporal em termos de ataque, sustentação e relaxamento; a mudança tanto do envelope espectral quanto da frequência fundamental; e o prefixo, um atraso em um som um tanto diferente da vibração que persiste.

Grey conjectura que “timbre deve se referir às características do tom que servem para identificar que um som se origina de um instrumento específico ou família de instrumentos, por exemplo, que é um oboé, ou talvez algum instrumento de palheta dupla, ou talvez um instrumento de sopro da família das madeiras” [29]. Roederer [67] alerta que a percepção timbral é somente o primeiro estágio da operação de reconhecimento da fonte sonora – no caso de sons musicais, identificação do instrumento. Deste ponto de vista, a percepção da qualidade tonal é o mecanismo através do qual se extrai informação do sinal auditivo de maneira a torná-lo apropriado para armazenamento na memória com um rótulo de identificação adequado e comparação com informação sonora previamente armazenada e identificada. A primeira operação envolve aprendizado ou condicionamento, enquanto a segunda representa a resposta condicionada ao padrão aprendido. Pode-se exemplificar o primeiro mecanismo imaginando-se uma criança aprendendo a reconhecer um dado instrumento musical. Apresenta-se repetidamente uma melodia tocada naquele dado instrumento e diz-se o nome do instrumento. O cérebro da criança extrai a informação adequada da sucessão de estímulos auditivos, rotula a informação com o nome do instrumento e a guarda na memória para referência futura. Um exemplo do segundo mecanismo seria quando a criança ouve o mesmo instrumento em uma ocasião diferente depois do aprendizado, seu cérebro compara a informação extraída do sinal auditivo (isto é, seu timbre, de acordo com esta definição) com a previamente armazenada e, caso haja identificação positiva, produz o nome do instrumento como resposta. Por outro lado, quando se ouve um som novo, tal como uma série de tons criados com um sintetizador eletrônico, nosso sistema de extração de informação processará o sinal auditivo e tentará identificar a fonte sonora com a informação armazenada. No caso de identificação negativa, um novo rótulo será criado para esta inesperada classe nova de qualidades sonoras. No caso de identificação parcial, a resposta da avaliação perceptiva humana costuma ser ‘parecido com uma clarineta’ ou ‘tal qual um trombone latindo’ [67].

Recentemente, há uma tendência a generalizar a noção de timbre para além dos instrumentos acústicos orquestrais tradicionais [43]. Krumhansl aponta para alguns

problemas na definição de timbre como residindo na complexidade das medições acústicas, que normalmente consistem de tomar a forma das distribuições espectrais de energia e envelopes de amplitude. De acordo com Krumhansl, no entanto, estas descrições são tão complexas que é difícil isolar características que distinguem timbres [43]. Krumhansl também chama atenção para a suposta independência de timbre das outras dimensões do som musical. Pode-se mesmo assumir que as diferenças em distribuição de energia espectral são totalmente descorrelacionadas dos mecanismos de percepção de altura? Bregman [6], ao comentar sobre a definição da ANSI [1], atenta para o fato de que não é, na verdade, uma definição, pois implicaria na impossibilidade de decidir se certos sons possuem a qualidade de timbre. Para que a definição possa ser aplicada, os dois sons apresentados devem ter a mesma altura, mas há sons que não possuem altura definida, como o som de uma pá penetrando uma pilha de cascalho. Bregman imediatamente reconhece que este fato se constitui em um problema, pois, ou deve-se afirmar que apenas sons com altura definida podem possuir a qualidade de timbre, o que significa que não se pode discutir o timbre de instrumentos percussivos, por exemplo, ou há algo terrivelmente errado com a definição [6]. Bregman, no entanto, aponta que parte do problema deve residir no fato de que altura e intensidade são controláveis de maneira direta e consistente em instrumentos musicais, mas timbre não é. Os métodos de variação timbrística variam de instrumento para instrumento, e quando se acha uma característica sonora que pode ser obtida em diferentes instrumentos, como vibrato, a mesma tende a ser rotulada separadamente de timbre [6]. Bregman afirma que, dentre os requerimentos para um vocabulário adequado para timbre, estão rótulos para dimensões salientes que deveriam agir em maneiras psicologicamente simples e deveriam ter correspondentes físicos diretos [6]. Handel [33] usa o termo timbre para se referir às qualidades perceptivas de objetos e eventos sonoros; ou seja, como soam. Handel afirma que, tradicionalmente, timbre é pensado como relacionado a uma propriedade acústica mensurável, de modo que as notas de um instrumento ou cada som falado por uma voz poderiam ser caracterizados por um único valor daquela propriedade. No entanto, devido à natureza dinâmica e interativa da produção sonora, há muitas propriedades acústicas estáveis e variáveis no tempo. De acordo com Handel, é improvável que uma propriedade ou combinação das mesmas determine timbre unicamente. Handel afirma que a sensação de timbre advém das propriedades emergentes e interativas dos padrões de vibração acústica e conjectura que uma possibilidade é que timbre seja percebido em termos das ações necessárias para gerar o

evento sonoro e a percepção das invariâncias neste processo de produção nos permitiria ouvir o mesmo objeto apesar das grandes diferenças no sinal acústico. De acordo com Handel, outra possibilidade seria que timbre é percebido simplesmente em termos das propriedades acústicas, e que a conexão entre as propriedades acústicas e o objeto é apreendida através da experiência. Sob esta ótica, as propriedades acústicas são usadas para identificar que evento teria maior chance de ter produzido aquele som. Hajda et al. [32] afirmam que timbre não é algo tangível, e sim uma abstração, um atributo de um tom musical que é abstraído da entidade que chamamos tom musical. De acordo com eles, timbre nem é o único atributo do tom conectado com qualidade sonora: considere-se volume e densidade. Eles afirmam, também, que apesar disso, timbre possui uma ordem perceptiva; na verdade, de acordo com eles, como atributo multidimensional, possui muitas, geralmente relacionadas a características físicas comuns a grupos de sons. Há, por fim, a interpretação de que timbre é uma propriedade emergente da interação da produção do tom com a percepção do ouvinte, um agrupamento do conjunto acústico influenciado pelo contexto acústico e pelo aprendizado e experiência prévia do ouvinte [6]. Neste trabalho, assim como para Bregman [6], timbre é tido como um fenômeno perceptivo complexo e subjetivo.

II.3 Síntese Sonora

Esta seção destina-se a apresentar alguns métodos clássicos de síntese sonora e definir um critério de avaliação de síntese sonora baseado nas qualidades desejáveis que esses critérios devem possuir. Pretende-se, dessa forma, apresentar a essência e relevância da nossa proposta, comparando-a e contrastando-a com os métodos clássicos apresentados e com o resultado das suas avaliações de acordo com as características apontadas como relevantes no critério de síntese sonora adotado.

Síntese sonora pode ser definida como a geração de um sinal que cria uma desejada sensação acústica [20]. Em música computacional, o termo instrumento se refere a um algoritmo que realiza um evento musical. O algoritmo (instrumento) calcula os valores das amostras de um sinal de áudio usando os parâmetros de entrada do programa. Por exemplo, um instrumento projetado para gerar um único tom simples pode ter como parâmetros a duração, a frequência e a amplitude do tom.

Torna-se importante ressaltar que as características de um método de síntese sonora consideradas relevantes neste trabalho são controlabilidade, flexibilidade,

adaptabilidade e simplicidade. Controlabilidade se refere à capacidade do usuário de prever a mudança que uma certa alteração nos parâmetros de entrada causa na resposta, ou seja, o usuário deve ser capaz de conceber qual configuração de parâmetros gerará um certo resultado desejado. Os métodos tradicionais requerem treinamento até que se desenvolva uma noção intuitiva do resultado psicoacústico decorrente de uma certa configuração de parâmetros. Risset [63], ao trabalhar com métodos tradicionais, comentou que os parâmetros são geralmente pouco intuitivos, tornando difícil a tarefa do usuário de prever o resultado perceptivo correspondente a uma dada configuração de parâmetros. Flexibilidade diz respeito à possibilidade de se produzir resultados substancialmente diferentes, isto é, sons com características suficientemente distintas, com a mesma técnica, através apenas da alteração da configuração dos parâmetros. Adaptabilidade é uma característica extremamente relevante porque está presente invariavelmente em todos os instrumentos musicais acústicos. A adaptabilidade de um instrumento é sua capacidade de se adaptar às variações do ambiente que o cerca. Isto é, mudanças de temperatura, umidade, do próprio músico tocando o instrumento, além de outros fatores que não podem ser controlados e que influenciam na sua sonoridade. Sabe-se que um mesmo instrumento, sob exatamente as mesmas condições não produz exatamente o mesmo som quando tocado duas vezes devido a pequenas variações estocásticas intrínsecas ao processo acústico de produção sonora. Por exemplo, variações de pressão no bocal de um instrumento de sopro, de velocidade do ar, dentre outros fatores, fariam com que duas notas nunca soassem exatamente iguais. Essas pequenas variações são esperadas pelo ouvido quando processando informação sonora, sendo parte do caráter único e vivo dos instrumentos musicais acústicos. Por fim, simplicidade é importante por razões óbvias de parcimônia (navalha de Occam [25]) e de capacidade de processamento computacional. Simplicidade está diretamente ligada, neste caso, à possibilidade de utilização do método em tempo real, ou seja, processamento instantâneo ao invés de acessar dados previamente armazenados em memória.

As técnicas tradicionais de síntese sonora apresentam limitações, principalmente devido ao fato de não levarem em consideração a natureza dinâmica e/ou subjetiva da música, usando processos determinísticos ou não especificamente desenvolvidos para a manipulação de sons musicais [10]. Em especial, as técnicas de síntese tradicionais explicadas a seguir possuem algumas limitações devido à sua abordagem e outras características individuais a serem exploradas adiante. Tanto as técnicas tradicionais

quanto a que será proposta serão analisadas de acordo com o critério de controlabilidade, flexibilidade, adaptabilidade e simplicidade para avaliação de uma técnica de síntese sonora adotado neste trabalho como contendo as características mais relevantes.

II.3.i) Síntese Aditiva

A técnica de síntese aditiva é baseada no modelo de Helmholtz de som musical. A composição espectral da forma de onda é determinada pela composição espectral do tom, obtido através da análise de Fourier. Isto é, a oscilação senoidal é utilizada como base para descrição do conteúdo de frequências do som. O espectro vai sendo construído através da adição dos parciais: geralmente utiliza-se o conteúdo do estado estacionário do tom. Pode-se adaptar o modelo com a aplicação de envelopes temporais diferentes a cada um dos parciais, como é mostrado na Figura II.3.1. A figura mostra um tom constituído da adição de 22 parciais com frequências aleatoriamente distribuídas entre 20 Hz e 5 kHz com posterior aplicação de envelope.

O som produzido por este instrumento difere do som natural em dois aspectos importantes. Primeiramente, as amplitudes de todos os componentes espectrais são variadas igualmente pelo envelope, de maneira que as amplitudes dos componentes não mudam umas em relação às outras durante o curso do tom. Portanto, falta ao som evolução temporal independente dos harmônicos, uma característica importante do som natural. Em segundo lugar, todos os componentes espectrais são harmônicos exatos da frequência fundamental, não as parciais ligeiramente desafinadas que normalmente ocorrem em sons acústicos.

Cada componente do som pode ser representado pela sua própria amplitude e frequência. A síntese de um tom baseada neste modelo requer um oscilador senoidal separado para cada parcial, com o envelope de amplitude apropriado aplicado a cada uma delas. As saídas dos osciladores são adicionadas para obter o resultado final.

A síntese aditiva possibilita grande flexibilidade nos tipos de som que podem ser gerados. Dado um número suficiente de osciladores, qualquer conjunto de componentes espectrais pode ser sintetizado e virtualmente qualquer som pode ser gerado, implicando em um alto grau de controlabilidade de acordo com o critério estabelecido anteriormente. Pode-se, ainda, melhorar a adaptabilidade da técnica atenuando os problemas mencionados acima, isto é, adicionando-se uma leve flutuação em frequência nas parciais

e desafinando-as ligeiramente para obter sons que causem a sensação de fusão dos parciais, como os sons naturais complexos.

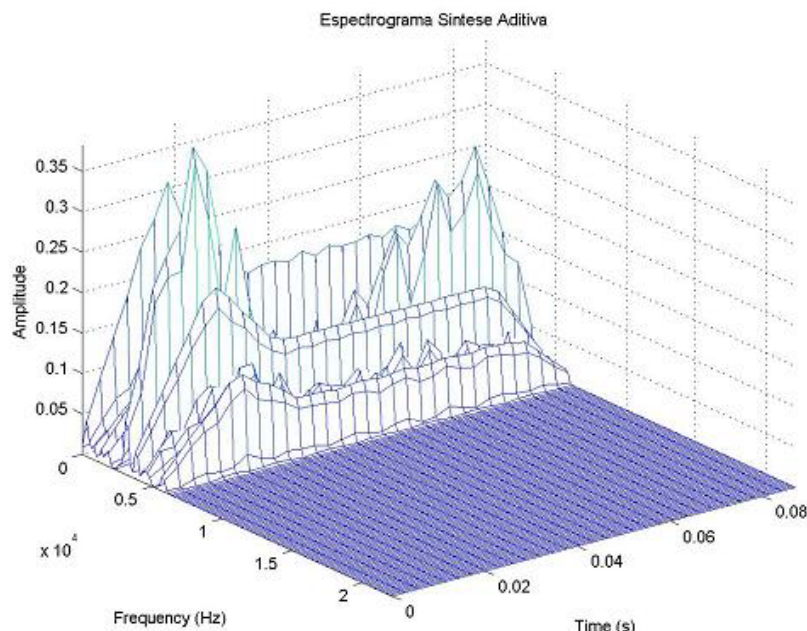


Figura II.3.1: Exemplo de síntese aditiva.

A figura mostra um tom construído com 22 parciais com frequências aleatórias distribuídas entre 20 Hz e 5 kHz. Como as frequências não estão relacionadas, o tom é inarmônico. Foi aplicado um envelope em cada parcial previamente.

A síntese aditiva produz sons de alta qualidade, mas requer uma quantidade relativamente grande de dados para descrever o som, além de um grande número de osciladores dependendo do número de parciais a ser gerado, resultando em pouca simplicidade. Uma complicação adicional surge porque um dado conjunto de funções é normalmente útil apenas para uma faixa limitada de alturas e intensidades. Se um conjunto é determinado por análise de um tom natural em uma determinada altura, então esse mesmo conjunto produzirá a qualidade sonora da fonte original em apenas um pequeno intervalo em torno daquele ponto. Quaisquer formantes presentes no espectro migrarão juntamente com a frequência. Portanto, muito da similaridade timbral entre os tons será perdida em alturas diferentes. Além disso, as funções são muito dependentes da dinâmica, de maneira que um conjunto de funções obtido para *mezzo forte* produzirá um *pianissimo* pouco realista. Para se contemplar na totalidade os benefícios da síntese aditiva, é necessário ter uma grande biblioteca de funções.

II.3.ii) Síntese Subtrativa

A síntese subtrativa cria tons musicais a partir de fontes complexas esculpindo porções selecionadas do espectro original. Em síntese subtrativa, uma fonte com espectro amplo, tal como ruído branco ou um pulso estreito, serve de matéria prima para a formação do tom por filtragem.

Qualquer som pode ser usado como fonte para síntese subtrativa. A título de ilustração, será apresentado um caso em que uma onda quadrada é submetida a filtragem. O filtro utilizado rejeita frequências acima de um limite de aproximadamente 5 kHz. Primeiramente a Figura II.3.2 a) mostra uma forma de onda quadrada, isto é, uma forma de onda que se mantém constante ou em seu valor máximo, ou em seu valor mínimo, sem apresentar valores intermediários. O espectro da onda quadrada pode ser visualizado na parte b) da Figura II.3.2. Deveria ser possível notar que, devido às transições abruptas no tempo, o espectro da onda quadrada se estende indefinidamente na frequência. No entanto, a maior frequência representável sem sobreposição espectral, de acordo com o teorema de Nyquist, é metade da frequência de amostragem (44.100 amostras/segundo, na figura), o limitante superior mostrado na figura.

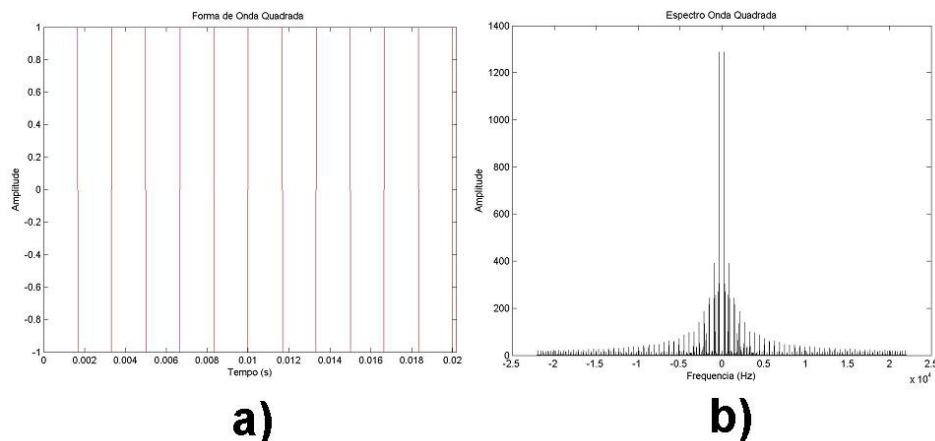


Figura II.3.2: Onda quadrada e seu respectivo espectro.

A figura mostra uma onda quadrada no tempo e na frequência. A parte a) mostra a onda quadrada no tempo e a parte b) mostra seu espectro de Fourier. Note-se que o conteúdo em frequência do espectro deveria ser ilimitado devido às transições abruptas no tempo.

A seguir, é mostrada na figura Figura II.3.3 a) e b), respectivamente, a forma de onda e o espectro resultante após a filtragem por um filtro passa baixa digital com frequência de corte em 5 kHz. Deve-se notar o efeito da filtragem não somente na forma

de onda – domínio do tempo – mas também na composição espectral resultante – domínio da frequência.

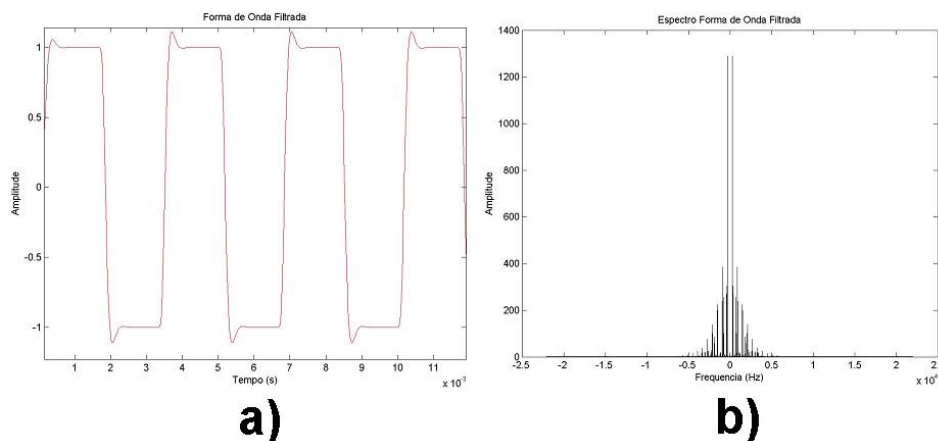


Figura II.3.3: Síntese subtrativa nos domínios do tempo e frequência.

A figura mostra a mesma onda quadrada depois de filtrada. O filtro utilizado foi um passa-baixas com frequência de corte em 5 kHz. O espectro mostrado na parte b) confirma que o espectro é agora limitado. A este fato corresponde no tempo a deformação da onda quadrada, evidenciada na parte a) da figura.

Pode-se obter resultados diferentes utilizando-se diferentes tipos de filtros e combinando-os. A adaptabilidade dos parâmetros dos filtros dita a adaptabilidade do método, tornando possível a variação dinâmica de certas características dos resultados.

A síntese subtrativa permite um certo grau de controlabilidade, podendo-se atingir resultados que possuem certas características espectrais gerais desejadas, tal como energia em uma certa banda de frequências. A flexibilidade do método, de acordo com o critério adotado anteriormente, é baixa, pois depende muito da fonte utilizada e das características dos filtros aplicados à mesma. É um método que permite um certo grau de simplicidade, pois necessita-se apenas de fontes sonoras com espectros complexos e filtros adaptáveis o suficiente para se obter resultados musicalmente interessantes, tais como os obtidos por Charles Dodge em seu trabalho intitulado *Changes*, no qual há três elementos texturais – linhas de contraponto, sons percussivos e acordes arranjados irregularmente. Paul Griffiths [31] escreveu a respeito de *Changes* que cada vez que um acorde novo aparece os timbres das linhas são alterados. É como se com cada novo acorde um filtro de cor diferente fosse colocado à frente do contraponto. A mudança no timbre é atingida mudando-se a frequência central do banco de filtros através do qual o pulso é processado.

II.3.iii) Síntese FM

Modulação em Frequência (FM) pode ser interpretada como a alteração ou distorção da frequência de um oscilador de acordo com a amplitude de um sinal modulante. Um exemplo musical da ocorrência de tal fenômeno é a técnica conhecida como *vibrato*, em que há variação da frequência de uma nota com uma dada amplitude. Quando a taxa de variação é sub-sônica (a frequência é menor que 20 Hz ou 20 ciclos por segundo) e a amplitude do *vibrato* é menor que um semi-tom, o resultado tem uma variação perceptivamente lenta em torno da frequência da fundamental. No entanto, quando a frequência do *vibrato* está na faixa de áudio (de 20 Hz a 20 kHz) e sua amplitude alcança valores maiores, a modulação FM pode ser usada para gerar uma ampla gama de timbres que podem ser controlados facilmente através de apenas um parâmetro, como será evidenciado a seguir.

A Figura II.3.4, a seguir, exemplifica a técnica de modulação FM. É mostrada uma oscilação senoidal de 1000 Hz, chamada portadora. A frequência da mesma será modulada por uma oscilação senoidal de 100 Hz, chamada de modulante ou moduladora. O sinal modulado resultante apresenta uma variação na frequência de acordo com a amplitude da moduladora.

A Figura II.3.5, a seguir, mostra o espectro do sinal modulado. É importante notar que o espectro da portadora corresponde a apenas uma raia na frequência 1000 Hz. O espectro apresentado na Figura II.3.5 mostra a inclusão de muitos componentes de frequência devido à modulação. Note-se também que a frequência original da portadora permanece no sinal. A Figura II.3.6 apresenta o espectro dinâmico do resultado. Note-se que há variação dinâmica do conteúdo espectral ao longo do tempo. Em termos musicais, este enriquecimento no espectro através de uma técnica relativamente simples, se devidamente explorado, pode ser bastante útil. O pioneiro na aplicação de modulação FM em síntese sonora foi John Chowning [12] e uma breve análise de sua utilização da técnica para fins composicionais será feita a tempo.

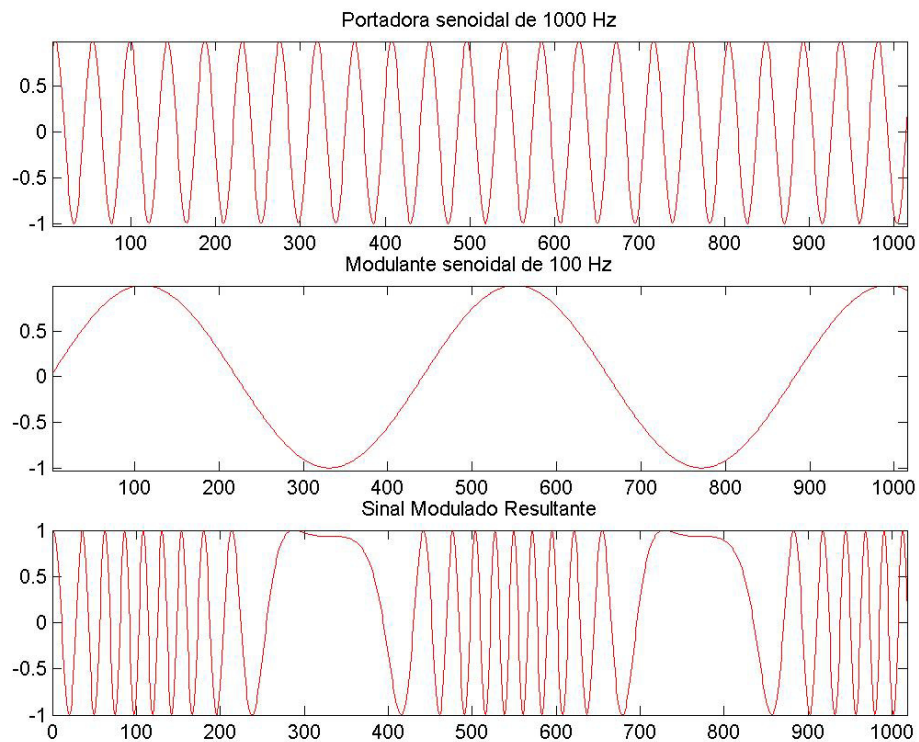


Figura II.3.4: Exemplo de síntese FM.

É mostrada uma portadora senoidal de 1000 Hz e uma moduladora também senoidal de 100 Hz. O sinal modulado resultante mostra que sua frequência varia a uma taxa de 100 Hz. Os valores mínimo e máximo da excursão de frequências depende da amplitude do sinal modulante.

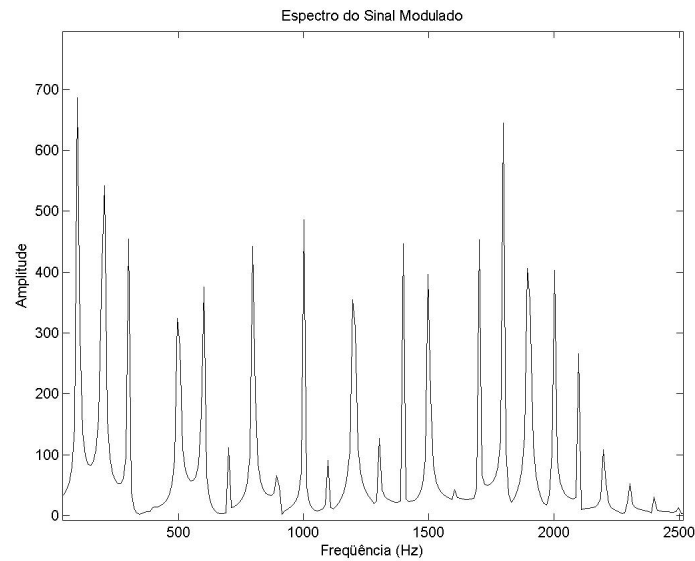


Figura II.3.5: Espectro do resultado do exemplo de síntese FM.

Este é o conteúdo espectral do sinal modulado mostrado na Figura II.3.4. Note-se a inclusão de muitas frequências, além da frequência da portadora.

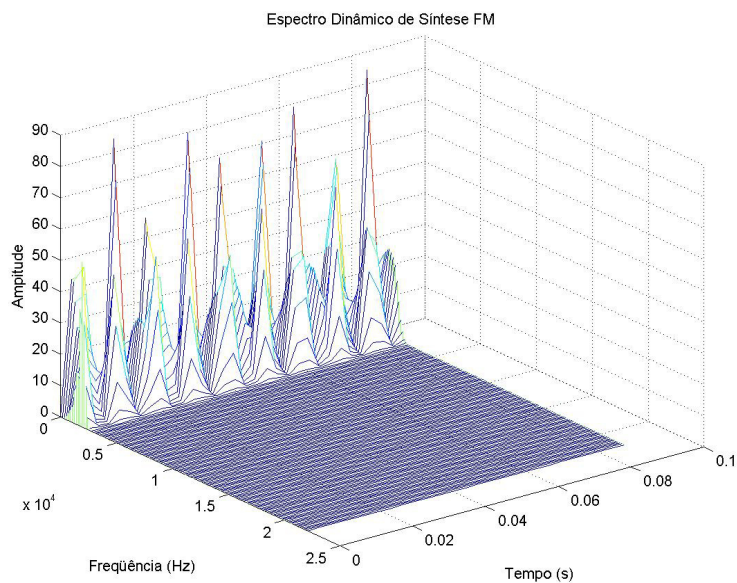


Figura II.3.6: Espectro dinâmico do resultado do exemplo de síntese FM.

A figura mostra a mesma informação mostrada na Figura II.3.5 anterior com a adição da informação temporal. Deve-se notar que o conteúdo adicional de frequência não é constante ao longo do tempo, variando a uma taxa de 100 Hz. Isto pode ser visualmente confirmado por inspeção da figura. Note-se que um ciclo leva 0,01 segundo para se completar.

Um sinal FM pode ser descrito matematicamente pela equação 2.1 a seguir

$$s(t) = A_c \cos[2\pi f_c t + \beta \sin(2\pi f_m t)] \quad (2.1)$$

onde A_c é a amplitude do sinal modulado $s(t)$. A frequência da portadora é f_c , e a frequência do sinal modulador é f_m . O índice de modulação β é definido pela equação 2.2, cujos termos são definidos a seguir. Quando ocorre modulação, o sinal modulador (na fórmula, assumiu-se que o sinal modulador é uma oscilação senoidal) leva a frequência da portadora acima e abaixo da frequência f_c da portadora. Esta mudança da frequência da portadora é proporcional à amplitude do sinal modulador.

O pico de desvio de frequência Δf_c é definido como o desvio máximo da frequência da portadora. Este desvio é dado pela amplitude da moduladora. O valor máximo instantâneo da frequência da portadora é $f_c + \Delta f_c$ e o valor mínimo é $f_c - \Delta f_c$. É possível que o pico de desvio de frequência seja maior que a frequência da portadora, resultando valores instantâneos de frequência do sinal modulado negativos. Isto é equivalente à inversão de fase, isto é, a uma oscilação de mesma frequência em módulo e fase 180° em relação às outras presentes no sinal. O índice de modulação é definido como o pico de desvio de frequência em relação à frequência da moduladora, mostrado na equação 2.2.

$$\beta = \frac{\Delta f_c}{f_m} \quad (2.2)$$

A distribuição de energia entre os componentes espectrais depende em parte do pico de desvio de frequência Δf_c produzido pela moduladora. Quando $\Delta f_c = 0$, não ocorre modulação e toda a energia do sinal reside na portadora. Aumentando-se o desvio, fornece-se energia aos outros componentes de frequência, chamados bandas laterais. Quanto maior o desvio, maior a distribuição de energia entre as bandas laterais e maior o número das mesmas. Portanto, o valor do desvio pode ser usado como controle do espectro do sinal FM. Este é o parâmetro único de controle referido anteriormente. É possível calcular as frequências das bandas laterais através da equação 2.3 a seguir, matematicamente equivalente à equação 2.1, usada para definir a modulação FM.

$$s(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \cos[2\pi(f_c + n f_m)t] \quad (2.3)$$

Na equação 2.3, $J_n(\beta)$ é a amplitude de cada componente de banda lateral. J_n representa a função de Bessel do primeiro tipo de ordem n . Note-se que a amplitude de cada componente de banda lateral depende do índice de modulação. A Figura II.3.7 mostra as funções de Bessel do primeiro tipo até ordem 8 em função do índice de modulação. A figura mostra que, para se obter bandas laterais com amplitudes significativas, é necessário um valor alto do índice de modulação.

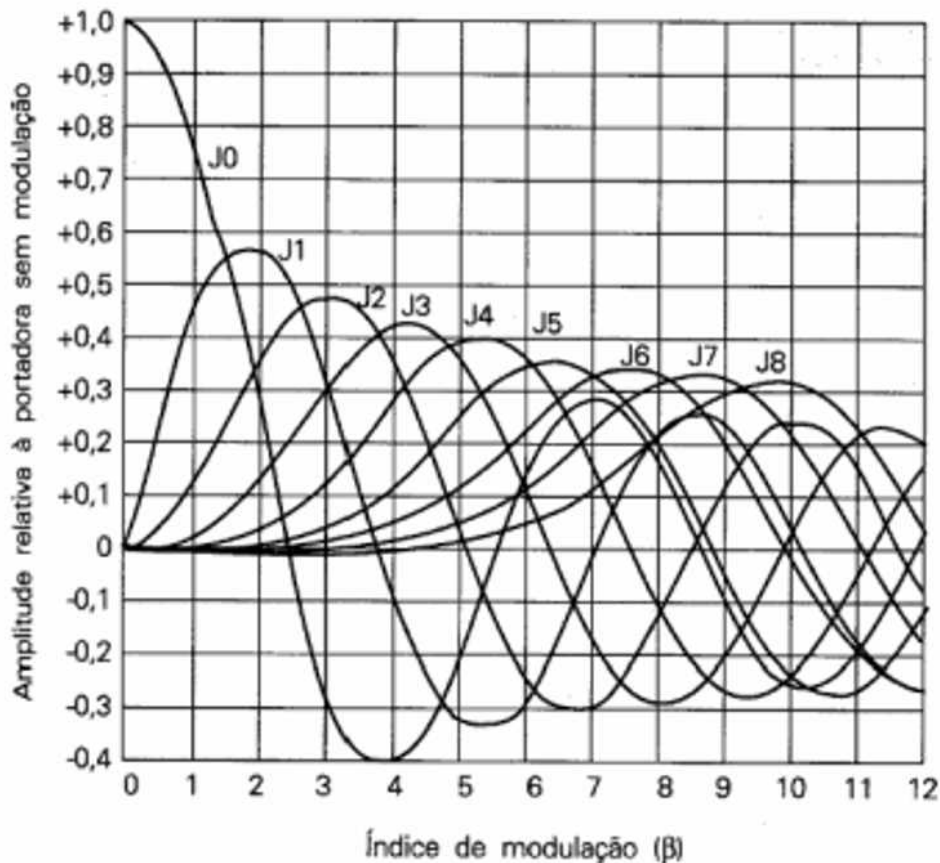


Figura II.3.7: Dependência das funções de Bessel do primeiro tipo com o índice de modulação.

A figura mostra as funções de Bessel do primeiro tipo da primeira até a oitava ordem em função do índice de modulação. A amplitude das bandas laterais é dada pela amplitude das funções de Bessel dependendo do índice de modulação.

Torna-se, portanto, fácil prever os valores das frequências das bandas laterais presentes no sinal modulado, bem como de suas respectivas amplitudes, sabendo-se as frequências da portadora, da moduladora e o índice de modulação. As frequências das bandas laterais são apresentadas a seguir.

$$\begin{aligned}
&f_c + f_m; f_c + 2f_m; f_c + 3f_m \dots \\
&f_c - f_m; f_c - 2f_m; f_c - 3f_m \dots
\end{aligned} \tag{2.4}$$

Muitas propriedades úteis do sinal modulado podem ser extraídas da razão entre a frequência da portadora e da moduladora, expressa como a menor fração inteira, como mostrado na equação 2.5 a seguir. N_1 e N_2 são inteiros sem fatores comuns.

$$\frac{f_c}{f_m} = \frac{N_1}{N_2} \tag{2.5}$$

A frequência fundamental do som resultante será f_0 , como expressa pela equação 2.6.

$$f_0 = \frac{f_c}{N_1} = \frac{f_m}{N_2} \tag{2.6}$$

No caso de $N_2 = 1$, o espectro resultante contém todos os harmônicos de f_0 . Se $N_2 = M$, onde M é um inteiro maior que um, então todos os harmônicos de ordem M estarão ausentes. Por exemplo, se $N_2 = 2$, o espectro não terá harmônicos pares. Quando $N_2 = 1$ ou $N_2 = 2$, os componentes de frequência negativa espelhados em torno do zero coincidirão com os componentes positivos e terão de ser combinados com esses, isto é, suas amplitudes são subtraídas desses componentes devido à inversão de fase. A Figura II.3.8 mostra um exemplo em que $N_1 = 1$, $N_2 = 2$ e $\beta=3$.

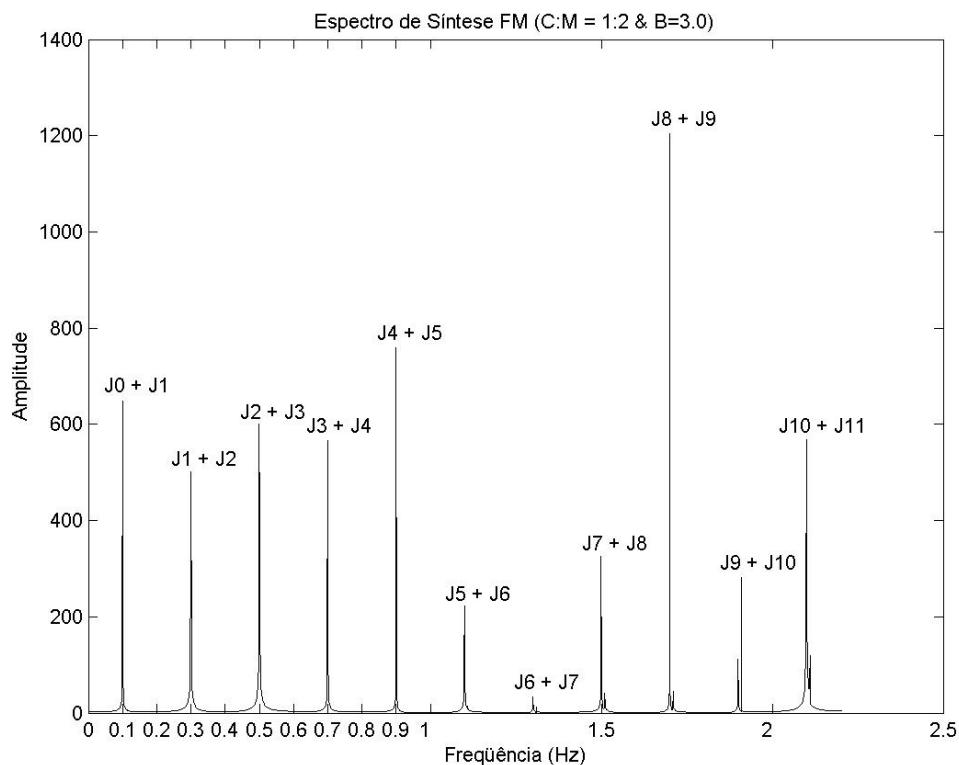


Figura II.3.8: Exemplo de espectro de síntese FM em função de f_c , f_m e β .

Amplitudes das bandas laterais de frequência em função das funções de Bessel, de acordo com a Figura II.3.7 e com os valores do índice de modulação e da razão entre a frequência da portadora e da moduladora.

Por fim, deve-se notar que, quando a frequência da portadora ou da moduladora é irracional, o espectro resultante é inarmônico [20].

As composições de John Chowning representam destilações de sua extensa pesquisa em técnicas computacionais para síntese de som. Sua perícia em acústica e psicoacústica tem papel fundamental na formulação de suas peças, bem como seu interesse em expressar fenômenos físicos matematicamente. No entanto, artisticamente, cada uma de suas composições é altamente unificada em torno de uma técnica ou relação.

A síntese FM permite uma flexibilidade desejável ao possibilitar que se construam conteúdos espectrais harmônicos ou inarmônicos bastante complexos a partir apenas da especificação de uns poucos parâmetros. Podem ser utilizados vários tipos de ondas como portadora, não sendo necessário se restringir a portadoras senoidais. Alguns

exemplos são portadora quadrada, triangular, dente de serra, dentre outras possibilidades. Pode-se enriquecer o método através da aplicação de diferentes envelopes temporais para se obter efeitos diversos. A adaptabilidade do método também é razoável, podendo-se obter espectros dinâmicos através da variação de apenas um parâmetro: o coeficiente de modulação.

A controlabilidade, no entanto, deixa um pouco a desejar, pois apesar de se poder prever o aumento do número de harmônicos, com conseqüente aumento de complexidade no resultado sonoro, através apenas da variação de um parâmetro (coeficiente de modulação), o verdadeiro resultado psicoacústico não é intuitivo. Não se pode projetar um som que se tenha em mente. Uma vantagem é a grande simplicidade do método, que requer basicamente dois osciladores para a obtenção de espectros bastante ricos.

II.3.iv) *Waveshaping*

Waveshaping é uma técnica de síntese sonora não-linear, ou de distorção, proposta por Marc Lebrun [45]. Como síntese FM, *waveshaping* é mais eficiente computacionalmente que síntese aditiva na geração de sons complexos. Diferentemente de síntese FM, *waveshaping* permite a geração de espectros de banda limitada. O espectro produzido por *waveshaping* muda com a amplitude do som. Devido ao fato de esta mudança corresponder às características dos espectros de instrumentos acústicos, a síntese *waveshaping* se tornou uma ferramenta bastante útil na produção de tons que se parecem com os dos instrumentos acústicos tradicionais.

Waveshaping é a distorção da amplitude de um som para produzir uma alteração em sua forma de onda [20], [45]. Um exemplo simples deste tipo de distorção pode ser ouvido nos estalos que ocorrem quando um amplificador de áudio é sobrecarregado. A introdução de distorção controlada cuidadosamente em um sinal pode resultar em uma vasta gama de timbres musicalmente interessantes. Assim como FM, *waveshaping* fornece controle contínuo do espectro por meio de um índice, tornando espectros dinâmicos possíveis através da variação temporal deste índice [45].

Basicamente, um sintetizador que utiliza *waveshaping* é um processador não-linear que altera o perfil da forma de onda que passa pelo mesmo. Em um processador linear, tal como um amplificador ideal, uma mudança na amplitude do sinal de entrada produz uma mudança igual no sinal de saída em todas as frequências de seu espectro. Portanto,

o perfil da forma de onda não é mudado, de maneira que, enquanto a amplitude de cada componente espectral dobra, a energia relativa entre os componentes é preservada. Em um processador não linear, a relação entre a entrada e a saída depende tanto da amplitude do sinal de entrada quanto da natureza da não linearidade. Portanto, um aumento na amplitude vai resultar em uma mudança no perfil da forma de onda de saída. Quando a forma de onda muda, seu espectro também muda, fato que geralmente corresponde a um aumento no número e intensidade dos harmônicos, assim como a uma alteração na energia relativa entre os componentes do espectro.

Um processador não linear que realiza *waveshaping* é caracterizado pela sua função de transferência, que relaciona a amplitude do sinal de entrada com o de saída. Portanto, para um determinado valor de amplitude da entrada, pode-se determinar a amplitude correspondente do sinal de saída através do gráfico da função. A Figura II.3.9 a seguir exemplifica uma tal função. O eixo rotulado x corresponde ao sinal de entrada e o eixo $F(x)$ corresponde à amplitude da saída. O perfil da forma de onda, e, portanto, seu espectro, varia com a amplitude da entrada. Deve-se notar que a quantidade de distorção depende do nível do sinal de entrada: o espectro se torna mais rico quando o nível é aumentado [45]. Note-se que esta função “achata” sinais com amplitudes baixas e “estica” sinais com amplitudes altas. Pode-se desenvolver uma certa intuição a respeito do resultado audível de uma função pelo seu gráfico. Um processador cuja função é uma reta não introduz distorção. Qualquer função diferente de uma reta introduz alguma distorção. Quanto maior a mudança na inclinação, maior o número e intensidade dos harmônicos adicionados. Uma função de transferência com descontinuidades sempre produz um número infinito de harmônicos e, portanto, sobreposição espectral (*aliasing*) [20].

Para evitar sobreposição espectral, a função de transferência é expressa algebricamente em forma de polinômio, mostrado na equação 2.7. Neste ponto, é importante lembrar que os polinômios formam uma base, assim como as funções seno e co-seno. Isto é, pode-se expressar qualquer função contínua em termos de polinômios. Tal procedimento matemático é chamado expansão em série de Taylor.

$$F(x) = a_N x^N + \dots + a_2 x^2 + a_1 x^1 + a_0 \quad (2.7)$$

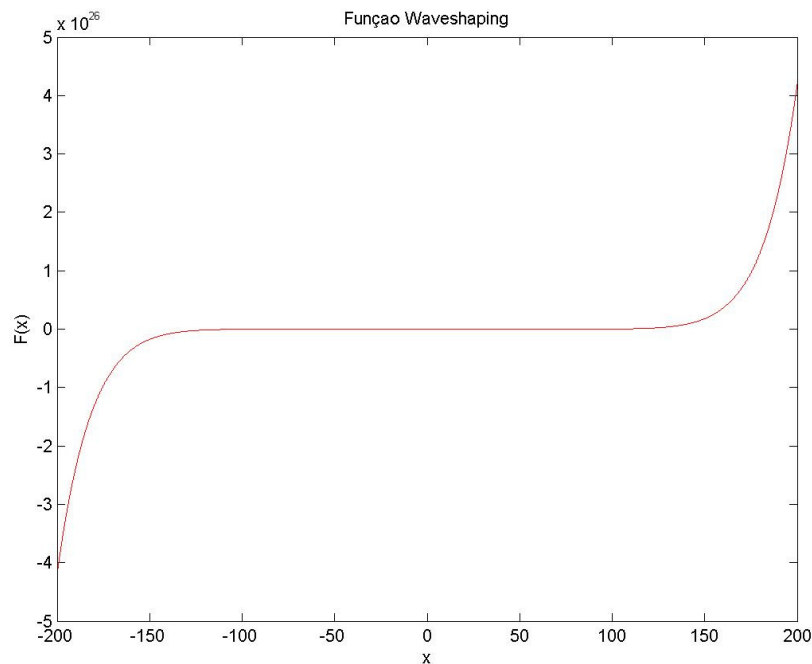


Figura II.3.9: Exemplo de uma função wavelshaping.

O eixo rotulado x corresponde ao sinal de entrada e o eixo $F(x)$ corresponde à amplitude da saída.

Quando se utilizam sinais senoidais na entrada do processador não linear (correspondente ao argumento da função), uma função de ordem N não produz harmônicos de ordem superior a N . No entanto, o termo de ordem N não produz apenas o harmônico de ordem N . Por exemplo, $\cos^2(\theta) = \frac{\cos(2\theta) + 1}{2}$. Nota-se que, além da introdução do segundo harmônico, há também um termo indesejável responsável pela introdução de um nível constante. Portanto, torna-se necessário compensar este efeito indesejável na função utilizada. É possível especificar um conjunto de funções que gere apenas os harmônicos desejados, onde cada termo é responsável apenas por um determinado harmônico para entradas senoidais. O termo geral é dado pela equação 2.8 abaixo [45]. Pode-se, a partir da mesma, construir cada uma das funções responsáveis por harmônicos específicos, como segue representado no conjunto de equações 2.9. Estas equações são conhecidas como polinômios de Chebyshev e estão representadas até a ordem sete. Isto é, o polinômio de Chebyshev de ordem N gera somente o N -ésimo harmônico quando a entrada (argumento) é senoidal.

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \quad (2.8)$$

É importante observar que, utilizando-se o polinômio de ordem dois, obtém-se a compensação desejada para o exemplo anterior usando $\cos^2(\theta)$. Isto é, multiplicando-se o termo $\cos^2(\theta)$ por dois e subtraindo-se um obtém-se apenas o termo desejado, $\cos(2\theta)$.

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x \\ T_4(x) &= 8x^4 - 8x^2 + 1 \\ T_5(x) &= 16x^5 - 20x^3 + 5x \\ T_6(x) &= 32x^6 - 48x^4 + 18x^2 - 1 \\ T_7(x) &= 64x^7 - 112x^5 + 56x^3 - 7x \end{aligned} \quad (2.9)$$

A função de transferência pode agora ser projetada para copiar exatamente a parte constante de um espectro, correspondente à sustentação de um tom (vide seção II.2.v) Figura II.2.15). Expressa-se a função de transferência como uma combinação linear (somatória) dos polinômios de Chebyshev. Cada polinômio, responsável por um dado harmônico, pode ser multiplicado por um coeficiente que representa a energia do harmônico no espectro, como mostra a equação 2.10.

$$F(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n T_n(x) \quad (2.10)$$

Foi extensivamente destacado que a distorção pode depender da amplitude do sinal de entrada. Considere-se agora, portanto, que a mesma pode ser controlada. A intensidade da entrada pode ser representada por uma constante multiplicativa B , chamada índice de distorção [45]. A função agora é expressa pela equação 2.11.

$$F(Bx) = a_N B^N x^N + \dots + a_2 B^2 x^2 + a_1 B^1 x^1 + a_0 \quad (2.11)$$

Pode-se, portanto, controlar a quantidade de distorção introduzida por uma dada função controlando-se o valor de B . Para espectros que variam dinamicamente no tempo,

deve-se utilizar um fator multiplicativo que dependa do tempo, denotado $B(t)$. A escolha da função de *waveshaping* agora pode ser expressa como na equação 2.12.

$$F(B(t)x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n(B(t))T_n(x) \quad (2.12)$$

Deve-se notar, no entanto, que o índice de distorção B não produz mudanças iguais nos harmônicos, pois é elevado a uma potência correspondente à ordem do harmônico [45]. Os coeficientes dependem do índice de distorção. Dodge [20] afirma que é raro polinômios que produzam mudanças suaves e que forneçam sons que não sejam metálicos, correspondendo normalmente a um espectro com bastante energia nos harmônicos superiores.

Jean Claude Risset trabalhou nos anos 1960 com *waveshaping*, projetando funções que produziam espectros com determinadas características e utilizando os resultados para composição [63]. Finalmente, será considerado um exemplo de instrumento, isto é, de função *waveshaping* que produz sons metálicos, mostrada na equação 2.13. A função é a mesma já mostrada na Figura II.3.9. A entrada utilizada foi uma oscilação senoidal simples de frequência 440 Hz, correspondente à nota A (lá), mostrada na Figura II.3.10.

$$F(x) = 20,48x^{11} - 43,52x^9 + 12,8x^8 + 40,32x^7 - 16x^6 - 17,44x^5 + 7,2x^4 + 3,6x^3 - 1,8x^2 - 1,27x - 0,3 \quad (2.13)$$

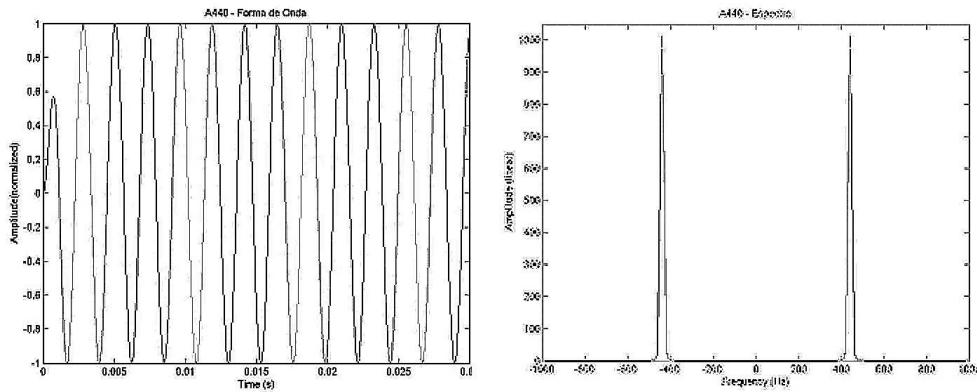


Figura II.3.10: Forma de onda e espectro de uma oscilação senoidal.

É apresentada uma oscilação senoidal simples em 440 Hz, correspondente a um lá, e seu espectro com apenas uma raia. Apesar da largura aparente da raia, devida exclusivamente à resolução da plotagem, a raia é apenas um pulso em 440 Hz.

A Figura II.3.11, a seguir, mostra o resultado tanto no tempo quanto na frequência. Note-se que a função utilizada não possui a décima potência da entrada, isto é, o coeficiente do termo x^{10} é zero. Isto produz, como esperado, a falta do harmônico de ordem dez (4400 Hz) no espectro resultante.

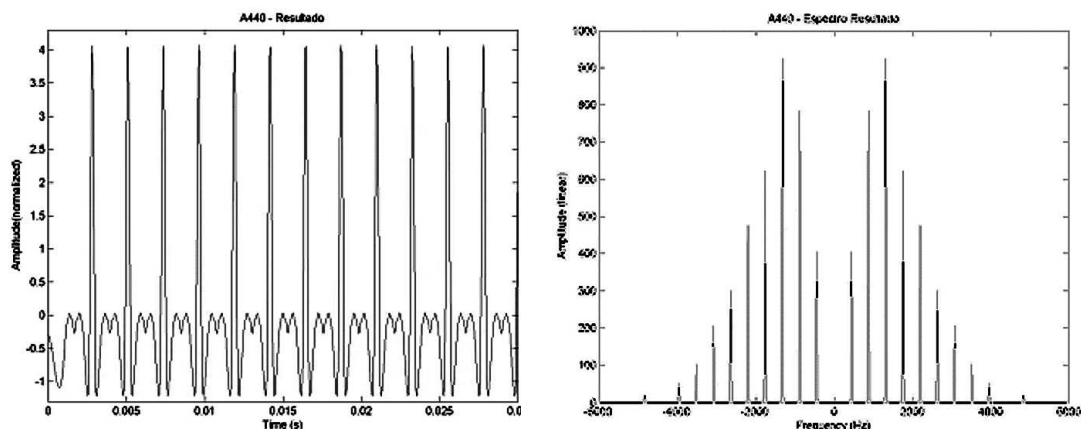


Figura II.3.11: Forma de onda e espectro do resultado da aplicação da função waveshaping da equação 2.13.

A forma de onda resultante é mostrada juntamente com o espectro resultante da aplicação da técnica. Chama-se atenção para o fato de que o décimo harmônico não está presente devido ao coeficiente do termo de ordem 10 do polinômio (equação 2.13) ser nulo.

Waveshaping permite ao compositor desfrutar de uma grande flexibilidade ao possibilitar que se construa o conteúdo espectral do estado estacionário de qualquer som harmônico a partir apenas da especificação da função de transferência. O método também apresenta alta adaptabilidade, podendo-se obter espectros dinâmicos através da variação de apenas um parâmetro, o coeficiente de distorção. Marc Le Brun mostrou que, em certas condições, *waveshaping* pode produzir os mesmos efeitos que FM, sendo uma generalização da técnica de distorção de sinais para a qual FM seria um caso especial [45].

A controlabilidade da técnica é aceitável, pois se pode prever o aumento do número de harmônicos, com conseqüente aumento de complexidade no resultado sonoro, através apenas da variação de um parâmetro, o coeficiente de distorção. No entanto, ainda não é possível prever com precisão os resultados psicoacústicos decorrentes da escolha de uma dada função e um índice de distorção, a não ser que o usuário da técnica tenha desenvolvido uma certa intuição através do uso constante da mesma. O método apresenta uma simplicidade bastante grande após a especificação da função de

transferência. Há, no entanto, uma grande limitação: somente espectros harmônicos podem ser gerados.

II.3.v) Síntese Granular

A síntese granular envolve a geração de milhares de eventos sonoros muito curtos, chamados grãos, para formar eventos acústicos mais longos [66]. Pode ser traçado um paralelo entre esta técnica e síntese aditiva, uma vez que os grãos são combinados linearmente (somados) para compor os sons resultantes. Densidades de grãos da ordem de centenas de grãos por segundo podem ser necessárias para criar sons complexos que evoluem dinamicamente no tempo.

De acordo com o modelo de Gabor [26], um grão é um *quantum* acústico, isto é, o evento sonoro fundamental portador de informação. O quantum sonoro foi caracterizado por Gabor como um sinal sonoro com duração tipicamente da ordem de milissegundos e envelope temporal quase-gaussiano, como mostra a Figura II.3.12, a seguir. De acordo com Gabor, qualquer som poderia ser descrito através de uma representação granular. Esta conjectura foi verificada por Bastiaans [3]. Iannis Xenakis [79] foi o primeiro a formalizar uma teoria composicional para grãos sonoros. Ele descreveu uma aproximação possível para o modelo de Gabor no contexto da implementação de síntese analógica e sugeriu que formas de onda granulares poderiam ser apropriadamente geradas pelo computador digital. Xenakis afirmou que “todo som é uma integração de grãos, de partículas sônicas elementares, de *quanta* sonoros [79]”. Curtis Roads realizou experimentos para implementar síntese granular usando o computador digital e relata que o potencial musical ficou claro desde o início [66]. Roads utilizou sons produzidos pelo método em várias composições.

Barry Truax [74] estende as possibilidades composicionais do método propondo que o conteúdo espectral dos grãos poderia ser uma oscilação senoidal de frequência definida, um espectro obtido através de modulação FM ou amostras de sons gravados de diferentes fontes, tal como sons da natureza ou de instrumentos acústicos. Pode-se também utilizar grãos compostos [66], obtendo-os através da mistura de grãos provenientes das fontes descritas anteriormente, bem como de outras possíveis.

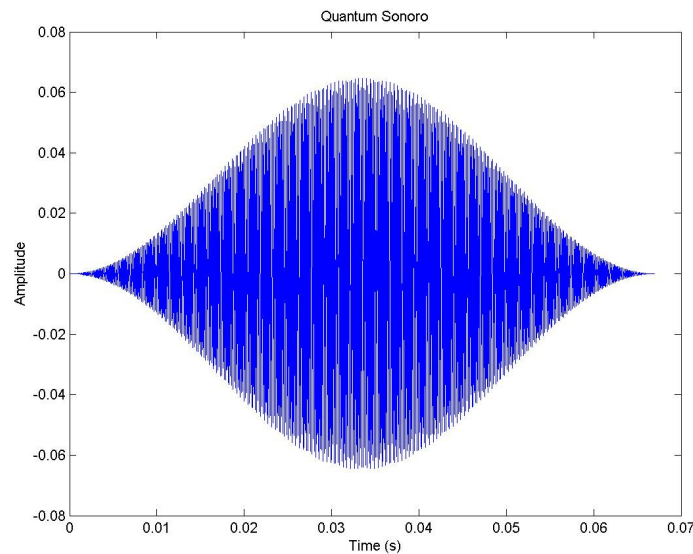


Figura II.3.12: Quantum sonoro de acordo com o modelo de Gabor.

O modelo propõe que um quantum sonoro é um evento de duração tipicamente da ordem de milisegundos e envelope temporal quase-gaussiano.

Vários modos de organização de alto nível dos grãos foram propostos. Xenakis [79] propôs o uso de planos de amplitude e frequência sobre os quais os grãos são espalhados, batizados por ele mesmo de telas (*screens*). Cada uma dessas telas pode ser tratada como um quadro de um filme, com a evolução temporal da textura sonora criada a partir da troca destas telas seqüencialmente seguindo algum critério. Roads [66] propôs um modelo organizacional baseado em eventos. Cada evento é caracterizado por doze parâmetros:

1. Tempo inicial;
2. Duração;
3. Forma de onda inicial;
4. Inclinação da forma de onda (a taxa de transição de uma senóide para um pulso de banda limitada);
5. Frequência central inicial;
6. Inclinação de frequência;
7. Largura de banda;
8. Inclinação da largura de banda;

9. Densidade inicial de grãos;
10. Inclinação da densidade de grãos;
11. Amplitude inicial;
12. Inclinação da amplitude.

Dentro de um evento, os grãos são espalhados aleatoriamente de acordo com a densidade inicial dos grãos e com a inclinação da densidade. Dependendo dos valores dos parâmetros especificados para os eventos, esses podem ser plotados em um gráfico de frequência versus tempo. Uma vez que os grãos são organizados em eventos, é possível variar os parâmetros para criar texturas uniformes, ou misturar grãos com parâmetros diferentes para criar nuvens de sons com espectros que evoluem no tempo [66].

Aqui está representado um exemplo ilustrativo gerado com duzentos *quanta* de durações de 20 a 250 milissegundos, com conteúdo senoidal de frequência variando entre 20 e 10kHz, distribuídos ao longo da duração de um tom. O envelope temporal mostrado na Figura II.3.13 foi posteriormente aplicado ao resultado, também mostrado à direita da Figura II.3.13.

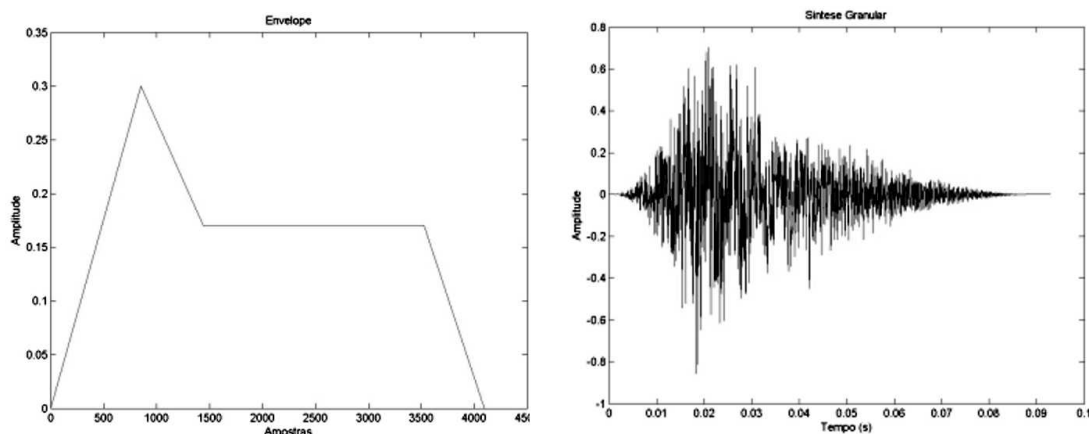


Figura II.3.13: Resultado do exemplo ilustrativo de síntese granular no domínio do tempo. À esquerda é mostrado o envelope temporal aplicado posteriormente à geração sonora. À direita a forma de onda resultante é mostrada. A mesma é composta de 200 quanta senoidais com frequências entre 20 e 10 kHz de durações variando entre 20 e 250 ms.

A seguir, o resultado é mostrado no domínio da frequência na Figura II.3.14. À esquerda tem-se o espectro de Fourier e à direita a progressão temporal de cada uma das parciais de frequência é mostrada no espectro dinâmico. Ressalta-se grande densidade de raiais

espectrais decorrente de o conteúdo em frequência não se restringir apenas a alguns valores discretos, mas poder assumir um grande número de valores devido ao grande número de grãos utilizados na composição deste resultado.

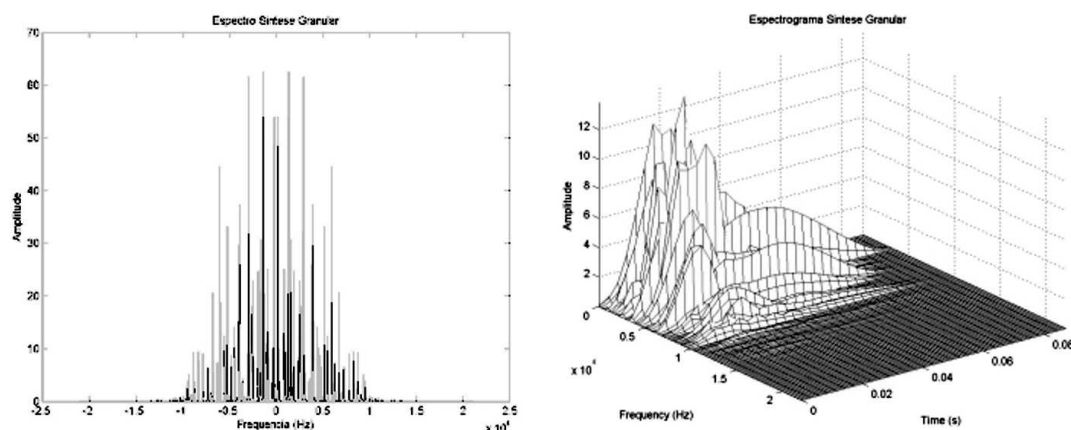


Figura II.3.14: Resultado do exemplo ilustrativo de síntese granular no domínio da frequência.

À esquerda é mostrado o espectro de Fourier do resultado. À direita tem-se a progressão temporal de cada uma das parciais de frequência mostrada no espectro dinâmico.

A síntese granular oferece uma vasta gama de cores e texturas diferentes à palheta de sons disponíveis ao compositor de música computacional, mostrando-se altamente flexível e adaptável de acordo com os critérios adotados neste trabalho. O método deixa um pouco a desejar, no entanto, em termos de simplicidade, sendo necessário grande esforço computacional para o processamento de um grande número de grãos em paralelo, necessários para a construção de resultados que atinjam os níveis de complexidade requeridos por certas aplicações.

Por fim, a síntese granular apresenta um nível mínimo de controlabilidade, permitindo ao compositor apenas controlar certos parâmetros que necessitam ser estabelecidos *a priori*. A previsibilidade das qualidades psicoacústicas do resultado final também é muito baixa e dependente dos parâmetros de controle estabelecidos. Torna-se, por exemplo, muito difícil a obtenção de um sinal com conteúdo espectral específico ou que tenha características perceptivas específicas.

III) MOTIVAÇÃO

Música computacional é um campo em constante crescimento, em parte porque permite ao compositor grande flexibilidade na manipulação sonora ao procurar pelo resultado desejado. Verifica-se, ainda hoje, a carência de métodos que possuam um nível desejável de controlabilidade, flexibilidade, adaptabilidade e simplicidade. Este trabalho apresenta um método de síntese sonora que permite a geração de sons que possuem características definidas *a priori*, presentes em um ou mais atratores. Para se compreender a técnica, é necessário compreender primeiramente a codificação utilizada, da qual decorre o espaço de busca sonoro. Para tal, serão brevemente apresentados conceitos relativos à representação de sinais analógicos por computador, à representação vetorial e, finalmente, à interpretação vetorial dos sinais digitais como advinda desta representação.

Adianta-se que todos os arquivos de som estão no formato *wave*. O formato de arquivo *wave* guarda áudio digital. A própria forma de onda é digitalizada e codificada apropriadamente. Várias resoluções de bits, taxas de amostragem e canais de áudio são possíveis (ver seção III.1.i). Este formato é bastante popular, sendo usado em programas profissionais de áudio para processamento digital de áudio e CDs. Os arquivos *wave* são tipicamente grandes devido a sua alta qualidade.

III.1 Formalização Teórica

Esta seção se dedica a apresentar a codificação utilizada e como a mesma possibilita que o espaço de busca resultante aproxime o conceito de contínuo sonoro, formalizado na seção III.2.

III.1.i) Representação Sonora por Computador

Os sinais analógicos são matematicamente descritos por uma função real ou complexa de variável independente real. Isto é, o argumento da função, correspondente aos valores normalmente associados ao eixo das abscissas, assume valores reais. A variável independente será normalmente identificada com a grandeza tempo. O fato de a variável independente ser real, e poder, portanto, assumir um contínuo de valores, é

freqüentemente sublinhado pela expressão sinal de tempo contínuo. Esses sinais desempenham freqüentemente o papel de modelos de certas variáveis físicas (corrente elétrica, diferença de potencial, pressão atmosférica em um dado ponto do espaço) em função do tempo. No caso de ondas sonoras (fenômeno físico) a representação contínua é a forma de onda e a representação digital usada é o arquivo *wave*. O processo de digitalização será explicado a seguir.

Os sinais discretos são, fundamentalmente, conjuntos de números que se podem enumerar, como $\{...h_{-2}, h_{-1}; h_0; h_1; h_2;...\}$. Isto é, a variável independente só pode assumir certos valores, não sendo, portanto, real, mas inteira. É importante salientar que a variável dependente de um sinal discreto pode assumir valores reais. Isto é, os valores da função propriamente dita, normalmente associados ao eixo das ordenadas, também podem assumir valores reais. Os sinais discretos resultam muitas vezes da amostragem dos sinais analógicos. Os sinais discretos têm grande importância prática, uma vez que qualquer operação que se pretenda realizar utilizando computador sobre um dado sinal analógico só é possível após a sua conversão num conjunto de números, isto é, num sinal digital. Como ficará claro mais adiante, um sinal digital é o resultado da fase final do processo de digitalização, a codificação. Por hora, basta lembrar que um sinal digital resulta da representação binária de um sinal discreto. Lembremo-nos que os computadores operam sobre representações binárias finitas, (números, num certo sentido), e não sobre funções. Por isso, para que o computador possa ser usado, é necessário reduzir a função (ou sinal) sobre o qual se quer operar a uma seqüência de números. É este o propósito do processo de digitalização, que será descrito de modo mais detalhado posteriormente. A digitalização de um sinal analógico consiste de quatro etapas: filtragem, amostragem, quantização e codificação. A filtragem e os motivos para tal são explicados nos parágrafos a seguir. As etapas de amostragem e quantização serão ilustradas na Figura III.1.2. Após a amostragem, o sinal analógico já é discreto, isto é, só possui valores diferentes de zero em certos instantes regulares de tempo. Finalmente, o processo de codificação é ilustrado na Figura III.1.3.

De acordo com o Teorema de Nyquist, a quantidade de amostras por unidade de tempo de um sinal, chamada taxa ou freqüência de amostragem, deve ser maior que o dobro da maior freqüência contida no sinal a ser amostrado, para que possa ser reproduzido integralmente sem erro de sobreposição espectral (*aliasing*). A metade da mínima freqüência de amostragem é chamada freqüência de Nyquist e corresponde ao

limite máximo de frequência do sinal que pode ser reproduzido. Como não é possível garantir que o sinal não contenha frequências acima deste limite (distorções, interferências, ruídos, etc), é necessário filtrar o sinal com um filtro passa baixa com frequência de corte igual a, ou menor que, a frequência de Nyquist. Esta etapa é chamada filtragem.

A Figura III.1.1 mostra um sinal senoidal sendo amostrado com taxas próximas ao limite. Em cima, amostragem com frequência maior que duas vezes a do sinal, há amostras suficientes para que o sinal possa ser reproduzido sem erro de sobreposição espectral. No meio, a taxa de amostragem é igual a duas vezes a frequência do sinal : não é possível a sua reprodução pois o sinal PAM (*Pulse Amplitude Modulation*, sinal discreto resultado do processo de amostragem mostrado na Figura III.1.2) vale zero. Se houvesse defasamento dos pontos de amostragem, haveria sinal PAM, porém com amplitude errada, a não ser que por coincidência os pontos caíssem nos picos da senóide. Deste fato decorre a necessidade da taxa de amostragem ser *maior que o dobro* da maior frequência do sinal. Em baixo, a frequência de amostragem é menor que o dobro da frequência do sinal, a quantidade de amostras é insuficiente e o sinal reproduzido estará errado. Este erro é causado pelo fenômeno de sobreposição espectral (*aliasing*).

Como o sinal analógico é contínuo no tempo e em amplitude, contém uma infinidade de valores. E como o meio de comunicação tem banda limitada, somos obrigados a transmitir apenas uma certa quantidade de amostras deste sinal, como enunciado anteriormente no Teorema de Nyquist. É obvio que, quanto maior a frequência de amostragem, mais fácil será reproduzir o sinal, mas haverá desperdício de banda ocupada sem nenhuma melhoria na qualidade. A Figura III.1.2, a seguir, ilustra o princípio da amostragem. A operação de amostragem consiste, essencialmente, em tomar os valores $f(kT)$ que um sinal f assume em determinados instantes kT do tempo, $k = 0, 1, \dots$, designados instantes de amostragem. Ao número T , que determina o intervalo de tempo que decorre entre duas amostras consecutivas, chama-se período de amostragem. O recíproco de T , $f_s = \frac{1}{T}$ chama-se frequência ou taxa de amostragem. Se T for expresso em segundos, f_s vem em Hertz (Hz). Também é costume utilizar a grandeza $\omega_s = 2\pi f_s = \frac{2\pi}{T}$, a que nos referiremos também como frequência de amostragem, e que se exprime em radianos por segundo (rad/s).

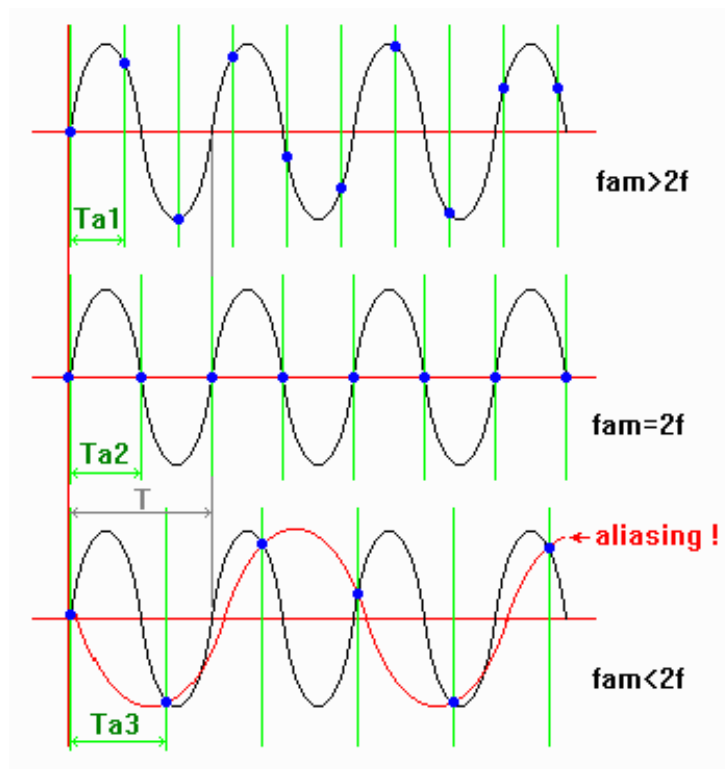


Figura III.1.1: Amostragem de um sinal analógico.

Esta figura ilustra a sobreamostragem e a subamostragem. A sobreamostragem corresponde a amostrar o sinal analógico a taxas acima da frequência de Nyquist. O sinal analógico pode ser recuperado. A subamostragem corresponde a taxas de amostragem menores do que postula o teorema de Nyquist, produzindo sobreposição espectral (*aliasing*) e impedindo a recuperação do sinal original.

Agora que temos o sinal analógico amostrado (discreto), na forma de amostras ou pulsos PAM (*Pulse Amplitude Modulation*) ainda analógicos, precisamos quantificar (ou quantizar) esta infinidade de valores possíveis em outros que possam ser representados por uma quantidade finita de bits, para obter um sinal digital. Esta conversão é feita por um circuito chamado conversor analógico-digital.

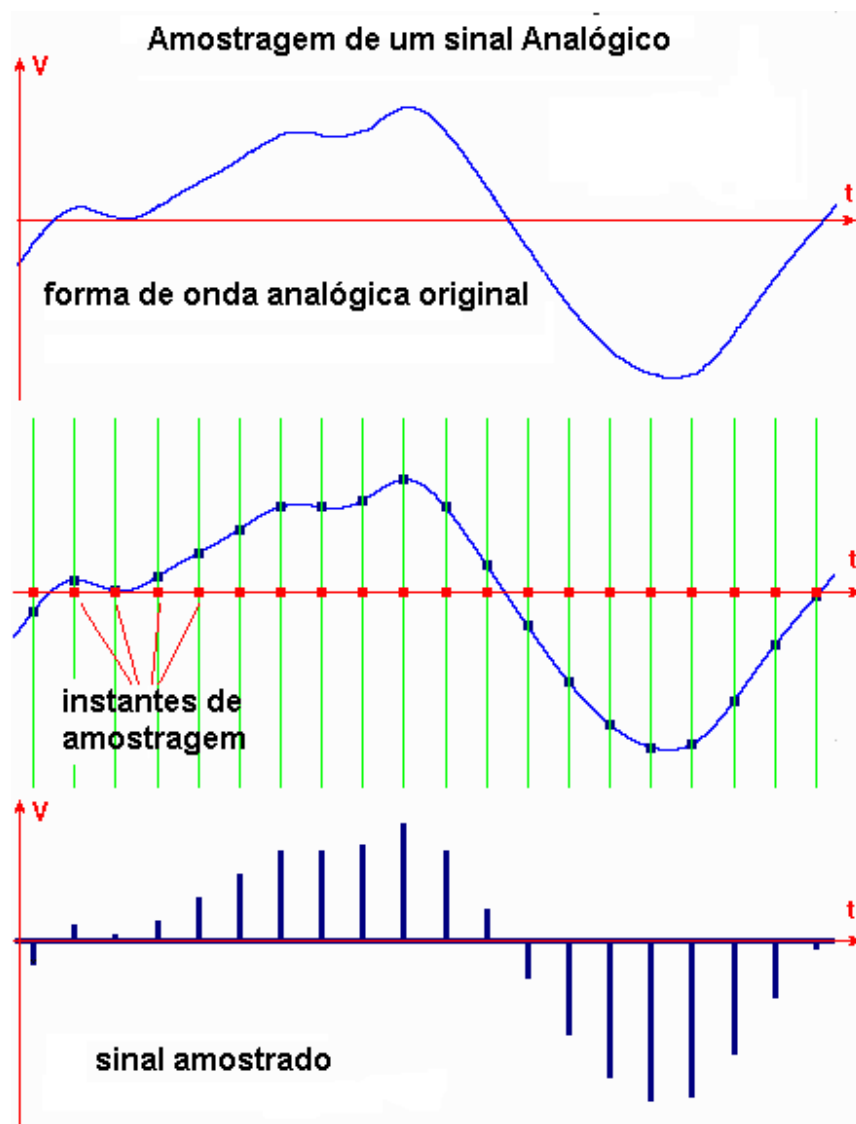


Figura III.1.2: Amostragem de um sinal analógico.

Vê-se a forma de onda analógica original e os valores correspondentes da mesma nos instantes de amostragem. Finalmente, a figura mostra o sinal amostrado. O sinal amostrado é um sinal discreto composto por pulsos PAM.

Cada amostra ou pulso PAM é transformada em uma quantidade predefinida de bits. Como cada bit pode assumir dois valores lógicos, verdadeiro ou falso (nível alto e nível baixo de tensão), pode-se representar, respectivamente, dois números com um bit, quatro com dois, oito com três, e assim sucessivamente, como ilustrado na Figura III.1.3. É possível representar 2^n números com n bits. Portanto com 8 bits é possível representar 256 valores diferentes, de 0 a 255, por exemplo. Para facilitar, vamos supor que os pulsos

PAM são limitados entre 0 e 255 Volts. Um pulso qualquer pode ter como valor real 147,39 V, mas terá de ser quantizado como tendo 147 V ou 148 V, pois não é possível representar 147,39 com 8 bits. O valor quantizado (para mais ou para menos) depende dos valores dos níveis de decisão no projeto do conversor analógico-digital. Teremos então um erro, no caso de -0,39 V ou +0,61 V, respectivamente, chamado erro de quantização. Esta falta ou excesso no valor do sinal provoca o surgimento de um sinal chamado ruído de quantização, que é inversamente proporcional ao número de bits utilizados na quantização. Isto é, quanto maior o número de bits, maior o número de níveis e, conseqüentemente, menor o espaçamento entre os níveis.

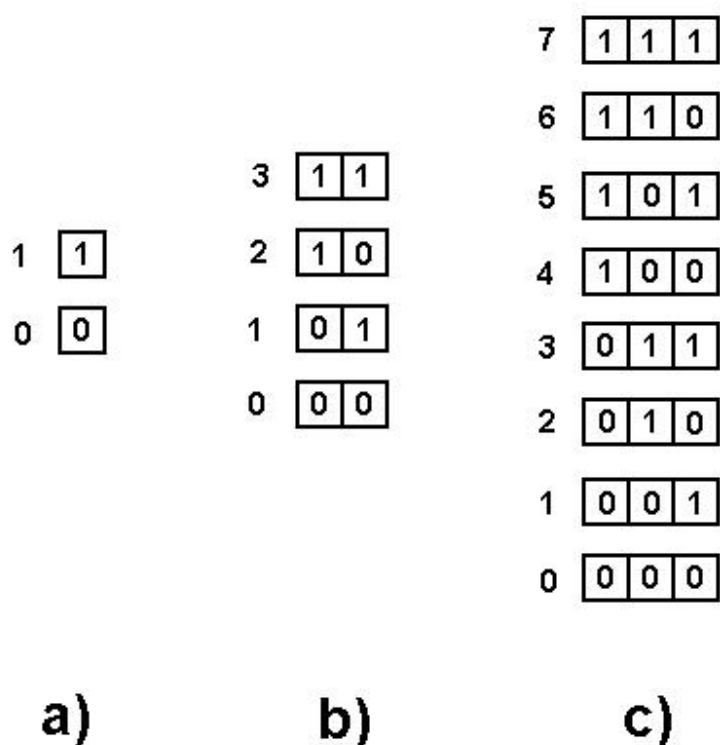


Figura III.1.3: Representação binária.

A figura mostra a representação binária de uma série de números inteiros. Cada uma dessas sequências pode ser associada a um dos valores escolhidos na quantização, caracterizando a etapa de codificação.

Após a digitalização, o sinal analógico é representado como uma seqüência de palavras binárias (grupo de n bits). Cada palavra representa uma aproximação da amplitude do sinal em um dado instante de amostragem, ou seja, uma amostra discreta

do sinal analógico original. O sinal digital pode, então, ser finalmente interpretado como uma sequência de números codificados de forma binária.

III.1.ii) Representação Vetorial

Existem dois tipos de grandezas na natureza, as grandezas escalares e as vetoriais. Grandezas escalares são aquelas que podem ser representadas através de um número e uma unidade de medida apenas, tal como temperatura. As grandezas vetoriais, por sua vez, são aquelas que necessitam de mais informação para sua completa descrição. Deslocamento é um exemplo de grandeza vetorial. É necessário mais do que um número associado à distância que se percorreu (com sua respectiva unidade de medida) durante o deslocamento para especificá-lo completamente. É necessário especificar também a direção e o sentido em que ocorreu o deslocamento para uma descrição completa. Utiliza-se um vetor para representar grandezas vetoriais. Um vetor pode ser interpretado como uma reta com um certo comprimento (seu módulo) e uma direção e sentido específicos. É comum representar vetores como setas, como mostrado na Figura III.1.4 a seguir. Na parte a) tem-se o vetor representado pela seta, sua origem é a extremidade sem a ponta e dela partem os eixos coordenados. Os eixos coordenados são ortogonais (perpendiculares) entre si e formam uma base vetorial, isto é, pode-se projetar o caminho representado pelo vetor original como um caminho percorrido pelos eixos e chegar no mesmo ponto do espaço, como representado na parte b) da Figura III.1.4. A base formada pelos eixos coordenados é ortogonal, fato que implica que cada um dos componentes vetoriais resultantes da projeção nos eixos da base é totalmente independente dos outros, ou seja, um deles não pode ser expresso em termos dos outros. Resulta que, dada uma base, um vetor pode ser representado especificando-se o comprimento de sua projeção em cada uma das dimensões, chamadas de coordenadas vetoriais e correspondentes aos pontos na Figura III.1.4 c).

Matematicamente, é possível trabalhar com vetores de qualquer dimensão, mesmo não sendo possível visualizá-los.

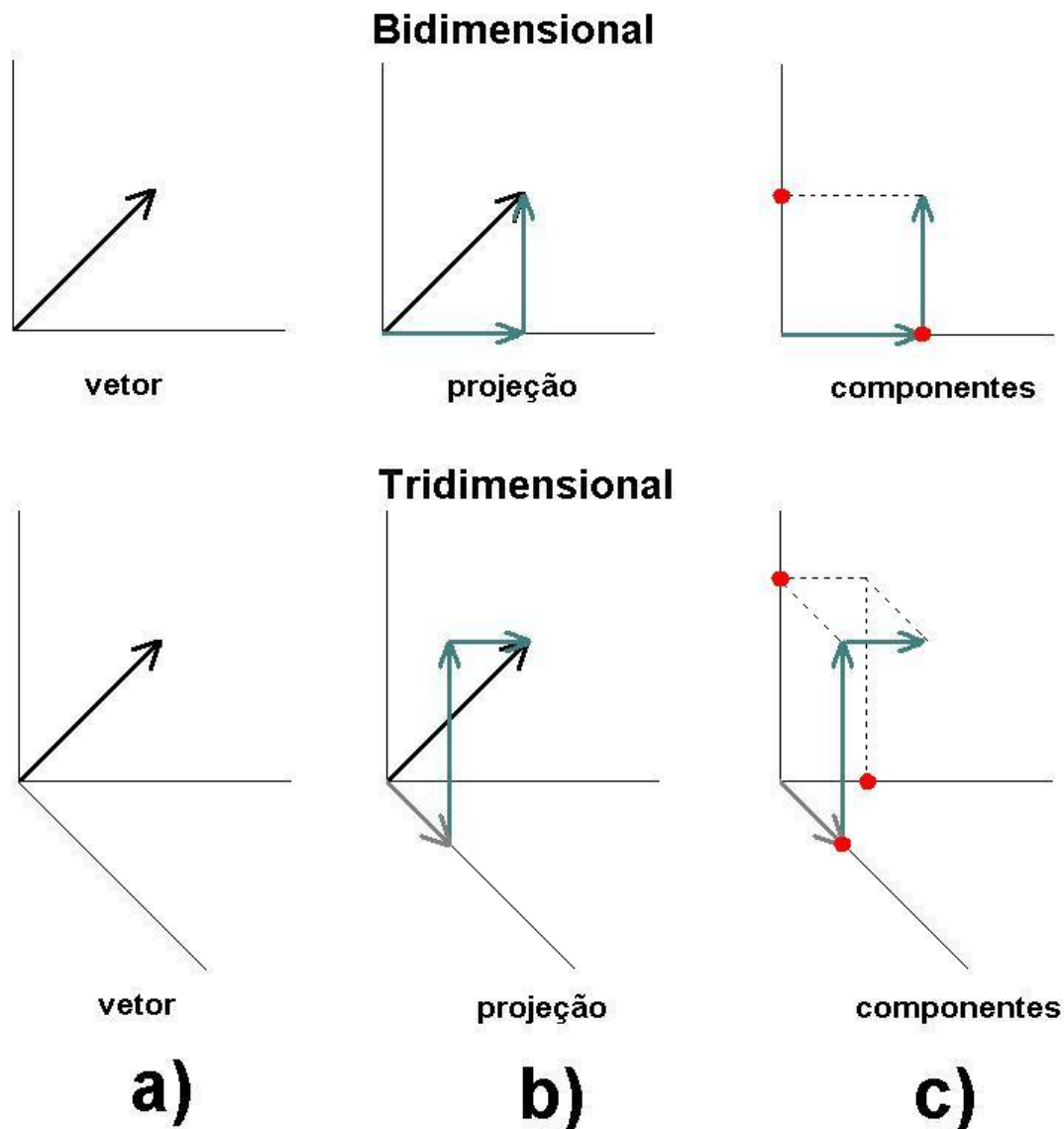


Figura III.1.4: Representação vetorial.

A figura ilustra a representação vetorial pelos seus componentes projetados em uma base ortogonal. A visualização em duas e três dimensões é possível e ambos os processos de projeção e representação pelos componentes é mostrado.

III.1.iii) Interpretação Vetorial das Formas de Onda

Como os sinais digitais são conjuntos de números, é possível identificar um dado sinal, composto por N números, por um vetor de N componentes, ou seja, de dimensão N . A afirmação recíproca também é verdadeira, isto é, qualquer vetor de dimensão N determina um sinal digital de N números. Decorre, portanto, que N amostras de uma

forma de onda –independentemente de sua frequência de amostragem – podem sempre ser tratadas e interpretadas como um vetor de dimensão N . Definindo-se a métrica utilizada para medir distâncias entre pontos neste espaço, a distância entre duas formas de onda, como codificadas neste espaço, está também definida. É importante notar que um vetor em um espaço qualquer pode sempre ser especificado a partir da origem, sendo, portanto, necessário especificar apenas o ponto onde termina. Advém, portanto, que para especificar um vetor em um espaço gerado por uma base ortogonal de qualquer dimensão, necessita-se, apenas, especificar um ponto neste espaço através de suas coordenadas em cada uma das dimensões. Naturalmente, se o sinal consistir numa infinidade de números, não poderá ser interpretado como um vetor de dimensão N finita, por maior que seja N . É útil, mesmo assim, falar de tais sinais como vetores de dimensão infinita, pelas possibilidades de interpretação geométrica que daí advêm.

III.2 Reticulado vs Contínuo Sonoro

Esta seção tem por objetivo primordial apresentar o conceito de contínuo sonoro. Mais adiante, na seção III.3.ii), é mostrado como a utilização do computador digital, aliada à representação temporal e manipulação vetorial das formas de onda, leva a um espaço de busca que corresponde a uma aproximação bastante fiel do conceito de contínuo, através de sua amostragem. A noção de contínuo é melhor compreendida em termos musicais quando contrastada com as possibilidades decorrentes dos instrumentos e notação musicais tradicionais. Para tal, é preciso estudar as características do sistema musical ocidental tradicional. Primeiramente, na música ocidental tradicional, os parâmetros musicais manipuláveis são:

- Frequência: notas musicais, intervalos;
- Tempo: duração, ritmo;
- Intensidade: dinâmica;
- Timbre: instrumentos musicais.

Como ilustrado pela Figura III.2.1 a), os parâmetros musicais são tradicionalmente discretos. Sob este prisma, composição musical significa, primariamente, organização de um conjunto finito de notas, ritmos, timbres e diferentes intensidades usando notação somativa (compasso) e durações fixas [78]. Este trabalho não está limitado ao reticulado musical tradicional representado na Figura III.2.1 a). A utilização do computador como

instrumento musical permite que não se fique limitado a um conjunto tão restrito de possibilidades, pois é possível explorar o contínuo sonoro e o conceito de transformações [78]. O espaço contínuo se encontra representado na Figura III.2.1 b). Uma trajetória neste espaço representa uma transformação gradual em uma ou mais dimensões representadas na figura. Neste espaço, a frequência é uma variável real (medida em Hertz, por exemplo), assim como o tempo, que pode ser medido em segundos. O timbre está representado como unidimensional, por simplicidade. Tradicionalmente, timbre também é uma variável sonora discreta, pois cada instrumento musical caracteriza um determinado timbre e não há intermediários entre dois ou mais instrumentos. Na representação contínua, no entanto, é possível obter sons com características advindas de mais de um instrumento através do simples posicionamento intermediário a dois ou mais instrumentos nas dimensões representativas das qualidades sonoras. Essas características não precisam ser necessariamente advindas apenas de uma ou mais dimensões timbrais, podendo-se cruzar outras características advindas de outras dimensões no resultado, dependendo do posicionamento no espaço euclidiano (vide Figura III.3.1 e interpretação na seção III.3.ii). Deve-se ressaltar também que as dimensões sonoras apresentadas não são necessariamente descorrelacionadas, ou seja, não formam uma base ortogonal. É sabido que mudanças de dinâmica em um mesmo instrumento musical acústico, por exemplo, acarretam mudanças timbrais, que podem ser visualizadas comparando-se os espectros dinâmicos. Sons com mais energia tendem a apresentar um maior número de parciais.

A possibilidade do posicionamento intermediário entre dois ou mais sons na representação contínua é a motivação por trás da proposta desta dissertação. O método proposto consiste em representar uma certa região do espaço de busca através dos atratores e utilizar a auto-organização como definidora dos processos que posicionarão os elementos que realizam a busca em determinadas posições do espaço, em relação aos atratores, de modo a obter como resultado sons com características advindas de mais de um atrator. Estes elementos que realizam a busca são denominados agentes neste trabalho. É importante notar que o espaço contínuo contém o discreto, isto é, qualquer som que possa ser representado por um ponto no espaço discreto também faz parte do espaço contínuo. O espaço contínuo é uma extensão do discreto. A esta altura é conveniente definir um processo como qualquer agente ou atividade que transforma informação de uma forma para outra [55].

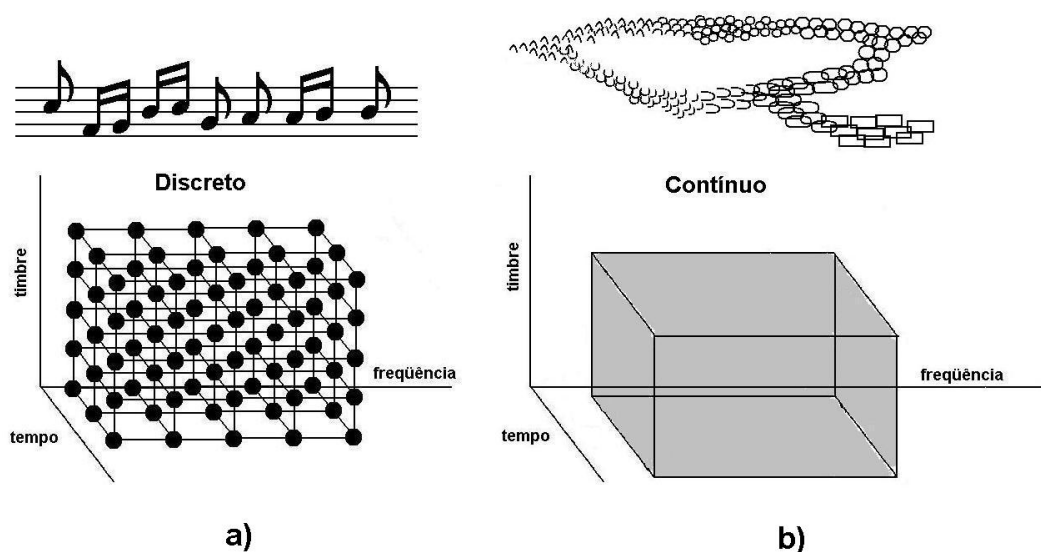


Figura III.2.1: Representação esquemática de três variáveis musicais.

A parte a) mostra a notação tradicional discretizada, representada pelo pentagrama acima da figura. A parte b) representa a extensão contínua do espaço, com todos os pontos possíveis. Trajetórias no espaço contínuo correspondem a transformações em uma ou mais dimensões, como ilustra a figura acima do gráfico representando o espaço contínuo.

III.3 Abordagem Adotada Neste Trabalho

A proposta deste trabalho é estudar extensões de uma técnica pioneira de síntese sonora digital a partir de algoritmos genéticos (AG), paradigma da computação evolutiva, denominada de “Método de Síntese Sonora Evolutiva” – ESSynth [51], [52]. A essência inovadora da metodologia proposta reside na representação temporal dos indivíduos, isto é, das formas de onda que compõe o conjunto população manipulado pelo algoritmo. A codificação é a própria forma de onda, em vez de seus parâmetros fixos, tal como na representação MIDI, em que se manipulam altura, duração e dinâmica de forma discreta, tal como no reticulado da Figura III.2.1 a). No protocolo MIDI, o timbre é dado por um programa pré-estabelecido que leva a fazer uma escolha de instrumentos com estrutura sonora determinada e não pode ser manipulado. O método proposto nesta dissertação trata de um processo inédito e criativo por utilizar ferramentas de resolução de problemas de otimização na obtenção de um ótimo subjetivo – um som artificial musicalmente interessante. As aplicações de sistemas inteligentes e bio-inspirados em síntese sonora apresentadas geraram publicações [8], [9], [10]. Ressalta-se que um dos objetivos dos processos utilizados é a diversidade dos resultados para subsidiar sua utilização criativa,

permitindo ao compositor liberdade no processo de composição. A próxima seção trata brevemente do conceito de otimização, importante para se compreender a aplicação dos algoritmos na obtenção de um ótimo subjetivo.

III.3.i) Otimização

Otimização é um tópico essencial em matemática aplicada, ciência da computação, IA, pesquisa operacional e outras áreas relacionadas. Fora dessas comunidades científicas, o significado de otimização é bastante impreciso: significa apenas fazer melhor. No contexto da ciência da computação, no entanto, otimização é o processo de se buscar a melhor solução possível, uma vez definido o espaço de busca, ou seja, uma vez formalizado matematicamente o problema de otimização [13]. Um problema de otimização é um problema para o qual existem diferentes soluções possíveis e uma noção clara de qualidade da solução. Isto é, as diferentes soluções candidatas podem ser comparadas e contrastadas.

Muitos problemas de otimização são fundamentalmente difíceis. Essencialmente, um problema difícil é aquele para o qual não se pode garantir achar a melhor solução em tempo razoável. Na prática, no entanto, é possível achar boas soluções para tais problemas através do uso de métodos aproximados. Um método aproximado é um algoritmo usado para buscar a solução de um problema de otimização para o qual não existam métodos exatos tratáveis. Os métodos aproximados rodam muito rápido, mas não garantem encontrar a solução ótima, ou mesmo convergência em alguns casos, ao contrário dos métodos exatos. Um exemplo comum de um problema para o qual não existe um método de resolução conhecido que garanta encontrar a melhor solução em tempo razoável é o problema do caixeiro viajante [13]. O problema consiste em, dispondo da distância entre os pares de pontos, achar a rota mais curta que passe por cada um dos pontos de um dado conjunto apenas uma vez e retorne ao ponto de partida. A seção V.2.viii) volta a tratar deste mesmo problema sob uma óptica inteiramente nova, a qual será explicada em tempo. Problemas de difícil resolução como este têm motivado um grande número de idéias para algoritmos aproximados. Uma classe de métodos aproximados é a composta pelos métodos de busca estocástica iterativa [13]. Tais métodos estão baseados na idéia de gerar soluções candidatas e testar sua qualidade. Se esta solução gerada não atende a algum critério de qualidade, o processo é repetido. Caso contrário, uma solução satisfatoriamente boa foi achada. É importante notar que a definição da qualidade da solução é feita pelo usuário, e satisfatoriamente boa não

significa a melhor solução, e nem mesmo implica no conhecimento do quão distante se está da solução ótima.

Um bom método aproximado gera novas soluções a partir de informação a respeito da qualidade dos candidatos anteriores para guiar a busca por candidatos melhores. Os métodos aproximados mais comuns utilizam dois mecanismos básicos: busca local e busca populacional. Na busca local, a solução atual é mantida e variações da mesma, chamadas soluções vizinhas, são exploradas para achar soluções de melhor qualidade. Na busca populacional, a noção de uma única solução atual é substituída pela manutenção de uma população de soluções candidatas. Novas soluções são geradas selecionando-se indivíduos dessa população para serem cruzados e darem origem a uma prole. Ao final do processo obtém-se uma nova geração de indivíduos. A procura por novas técnicas de otimização levou alguns pesquisadores a se voltarem para a natureza e para a maneira como resolve seus problemas como fonte de inspiração. Surgem os algoritmos bio-inspirados [16], [19], que são explorados nesse trabalho quanto ao seu potencial na música computacional.

III.3.ii) Codificação e Mapeamento

A Figura III.3.1 ilustra a codificação utilizada neste trabalho e o espaço (vetorial) de busca resultante da métrica utilizada, a distância euclidiana. Na parte a) tem-se N amostras de uma forma de onda digitalizada a uma taxa SR amostras/segundo. Cada forma de onda pode ser interpretada como um vetor no espaço de dimensão N . A parte b) da figura representa o espaço euclidiano bidimensional (correspondente a $N=2$, tomado apenas para efeito de visualização) e a distribuição de um conjunto de sons interpretados como vetores neste espaço. É neste espaço que os sons são manipulados pelos algoritmos. Há um mapeamento para o espaço sonoro das variáveis de interesse, pois cada forma de onda amostrada possui uma duração, uma altura definida (no caso de possuir espectro harmônico), um timbre e foi tocada a uma dada dinâmica. O processo composicional se dá neste espaço. Isto é, posições no espaço euclidiano de busca correspondem a posições no espaço composicional contínuo. Estas posições podem apresentar uma ou mais qualidades advindas de diferentes sons utilizados como atratores, ou seja, podem ser intermediárias a dois ou mais sons ou representar uma variante de um único som, por não corresponder exatamente a mesma posição do espaço. Smalley [72] declarou que a informação contida no espectro de frequência não pode ser separada da informação no domínio do tempo uma vez que “espectro é percebido através do tempo e tempo é

percebido como movimento espectral”. Portanto, ao especificar as formas de onda objetivo (atratores), o usuário também está especificando o conteúdo espectral e, por sua vez, as qualidades sonoras dos tons.

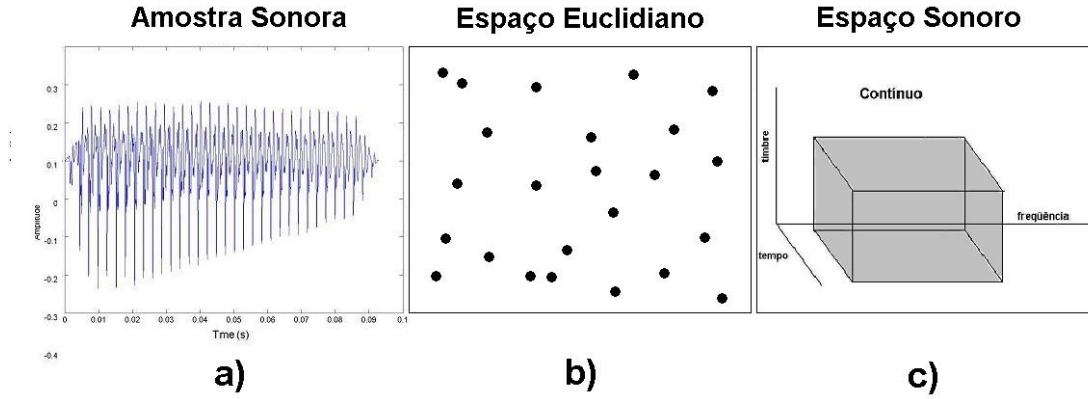


Figura III.3.1: Visualização da codificação das amostras sonoras, do espaço resultante e do mapeamento para o espaço sonoro onde se dá a manipulação dos parâmetros musicais relevantes.

Na parte a) tem-se N amostras de uma forma de onda amostrada a uma taxa SR amostras/segundo. Cada forma de onda pode ser interpretada como um vetor no espaço de dimensão N . A parte b) da figura representa o espaço euclidiano bidimensional e a distribuição de um conjunto de sons interpretados como vetores neste espaço. É neste espaço que os sons são manipulados pelos algoritmos. Há um mapeamento para o espaço sonoro da parte c) das variáveis de interesse, pois cada forma de onda amostrada possui uma duração, uma altura definida (sons harmônicos), um timbre e foi tocada a uma dada dinâmica. O processo composicional se dá neste espaço ilustrado na parte c).

III.3.iii) Espaço Vetorial de Busca

Dado um espaço vetorial com o conceito definido de métrica, é possível especificar uma norma (distância) entre dois (ou mais) vetores neste espaço. A norma mais comum é a euclidiana, apresentada na equação 3.1, onde $\mathbf{A}$ e $\mathbf{B}$ são dois vetores de dimensão N , n é o índice de uma dada dimensão e a e b representam as coordenadas dos vetores nas respectivas dimensões.

$$d(\mathbf{A}, \mathbf{B}) = \sqrt{\sum_{n=1}^N (a_n - b_n)^2} \quad (3.1)$$

Esta métrica, no entanto, não reflete necessariamente as características perceptivas de proximidade entre dois sons. Não há nenhuma razão especial para notas diferentes de um mesmo instrumento, por exemplo, estarem próximas no espaço euclidiano porque suas formas de onda podem ser bastante diferentes, retendo apenas algumas características em comum. Com o intuito de se avaliar automaticamente os resultados, foi desenvolvida uma métrica de distância baseada no espectro dinâmico das formas de onda. Intuiu-se que, levando-se em conta as características dinâmicas da evolução das parciais ao longo do curso do tom, as suas características perceptivas de distância seriam melhor refletidas.

A Figura III.3.2, a seguir, ilustra a matriz que contém o resultado do cálculo do espectro dinâmico de uma forma de onda. O eixo X representa o domínio temporal pelo índice j , o eixo Y representa o domínio da frequência pelo índice i . Cada coluna branca é, portanto, a evolução temporal de uma parcial de frequência, como $f1$, por exemplo. As colunas em cinza são espectros instantâneos calculados em instantes de tempo ao longo da evolução temporal do tom, como no instante $t1$, mostrado na figura. O encontro entre linha e coluna, na figura, representado pelo quadrado em preto, fornece o valor da amplitude de uma dada parcial em um certo instante de tempo.

A equação 3.2 define a métrica espectro-temporal desenvolvida, Δ , onde **A** e **B** representam duas matrizes de espectro dinâmico de dimensão $F \times T$, sendo F o número total de valores de frequência e T o número total de valores de instantes de tempo, i e j são os índices, respectivamente, da frequência e do tempo de um dado valor **a(i,j)**.

$$\Delta(A, B) = \sqrt{\frac{1}{FT} \sum_i^F \sum_j^T (|a(i, j)| - |b(i, j)|)^2} \quad (3.2)$$

O espaço de busca de todos os algoritmos aplicados ao problema de síntese sonora apresentado aqui será o espaço euclidiano. No entanto, é interessante notar que a aproximação no espaço euclidiano de busca corresponde à aproximação também nesta métrica. A explicação é simples, quando dois sons se aproximam no tempo (aproximação medida pela métrica euclidiana), esses sons também se aproximam no conteúdo espectral (aproximação medida pela métrica tempo-freqüencial), uma vez que tanto o tempo quanto a frequência são apenas maneiras diferentes de se extrair informação sobre o som. Pode-se evocar a integração entre os domínios do tempo e da frequência na percepção de sons musicais, como proposta por Smalley [72] para fundamentar a

afirmação. Smalley afirma que o “espectro é percebido através do tempo e que o tempo é percebido como progressão espectral”.

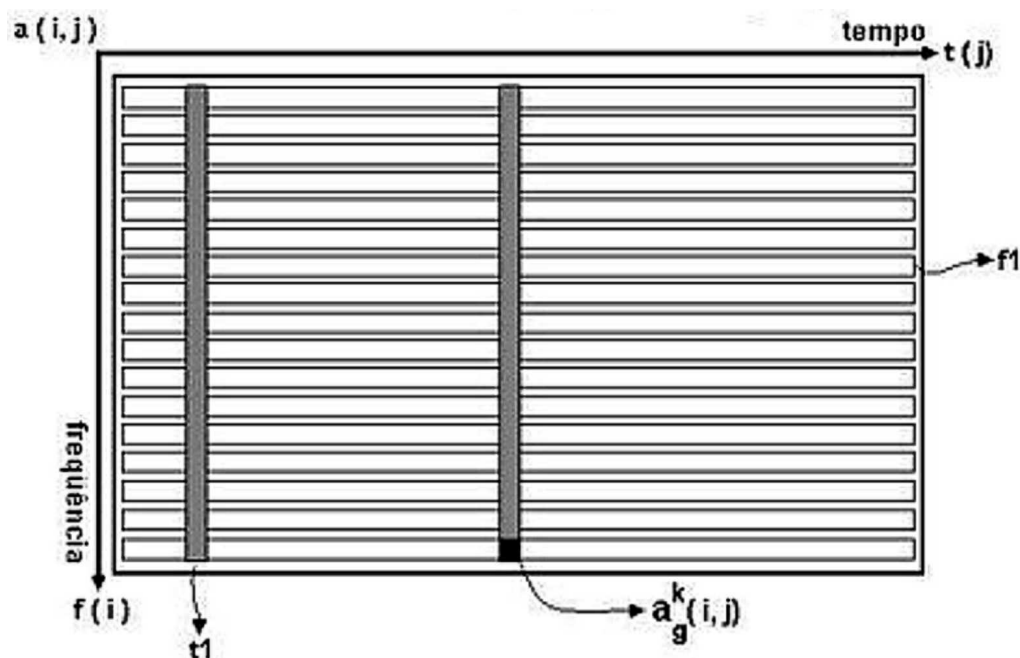


Figura III.3.2: Esquema da representação da matriz do espectro dinâmico.

O eixo X representa o domínio temporal pelo índice j , o eixo Y representa o domínio da frequência pelo índice i . Cada linha branca é, portanto, a evolução temporal de uma parcial de frequência, como $f1$, por exemplo. As colunas em cinza são espectros instantâneos calculados em instantes de tempo ao longo da evolução temporal do tom, como no instante $t1$, mostrado na figura. O encontro entre linha e coluna na figura, representado pelo quadrado em preto, fornece o valor da amplitude de uma dada parcial em um certo instante de tempo.

A próxima seção apresenta o espaço como caracterizado pelos atratores utilizados neste trabalho. Isto é, quais são os atratores e como ficam as distâncias entre eles, tanto segundo a métrica euclidiana quanto segundo a métrica tempo-freqüencial.

III.3.iv) Caracterização da Região do Espaço Escolhida

A Tabela III.3.1 apresenta os sete sons escolhidos como atratores. Estes sons em particular foram escolhidos por terem sido considerados perceptivamente como sendo suficientemente distintos entre si para permitir a avaliação de cruzamentos de características nos resultados do processo de síntese.

Da Figura III.3.3 até a Figura III.3.8, a seguir, são mostradas as formas de onda e os espectros dinâmicos correspondentes dos sons escolhidos como atratores.

Tabela III.3.1: Descrição dos sons utilizados como atratores.

Som 1	Som 2	Som 3	Som 4	Som 5	Som 6	Som 7
Sax Alto	Baixo Elétrico	Violão dedilhado	Acorde de Piano	Gaita	Voz	Apito

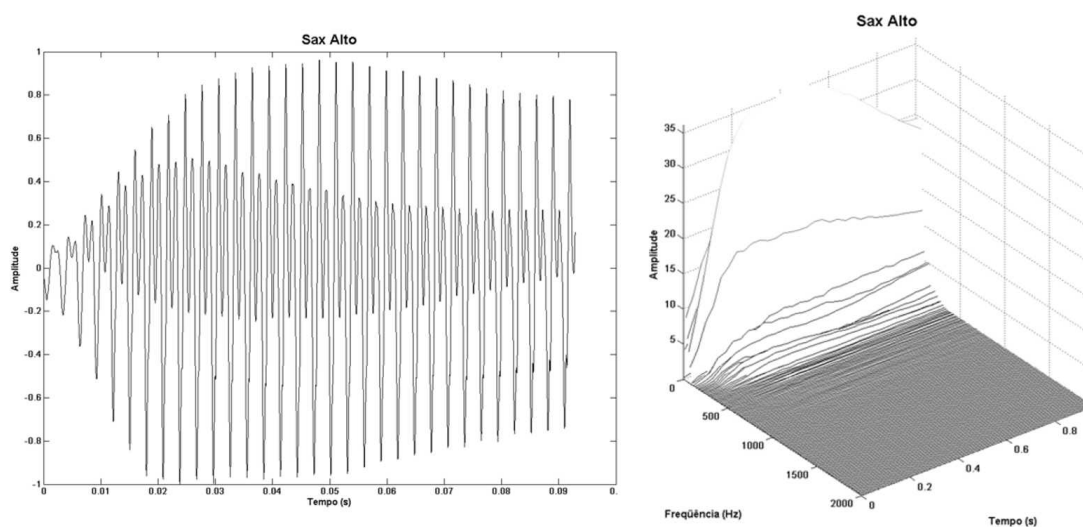


Figura III.3.3: Forma de onda e espectro dinâmico do som atrator 1: sax alto.

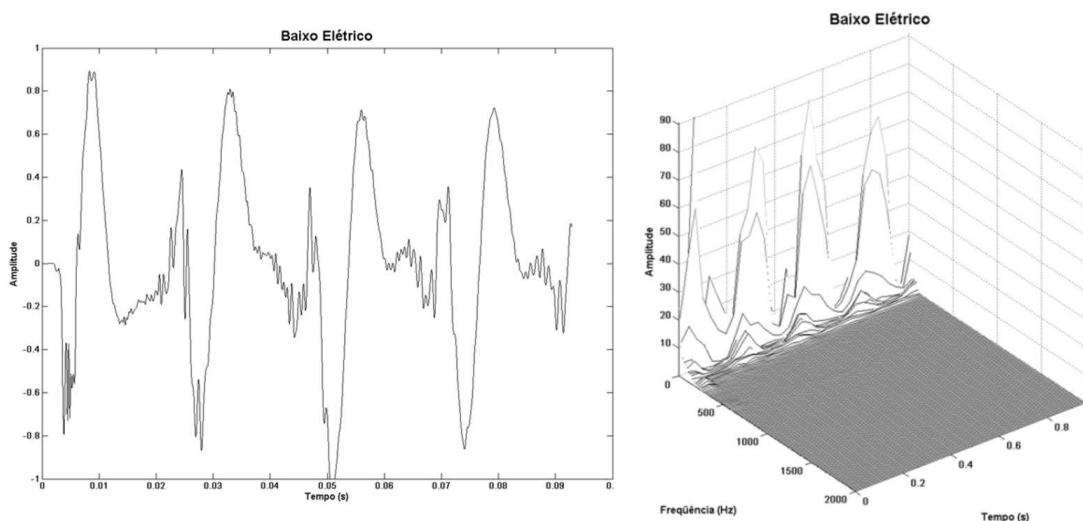


Figura III.3.4: Forma de onda e espectro dinâmico do som atrator 2: baixo elétrico.

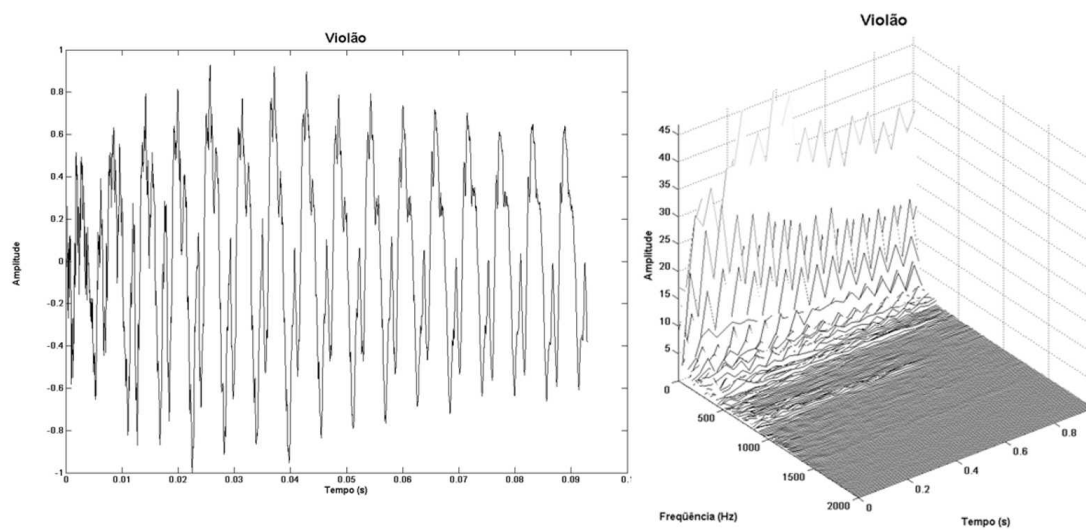


Figura III.3.5: : Forma de onda e espectro dinâmico do som atrator 3: violão dedilhado.

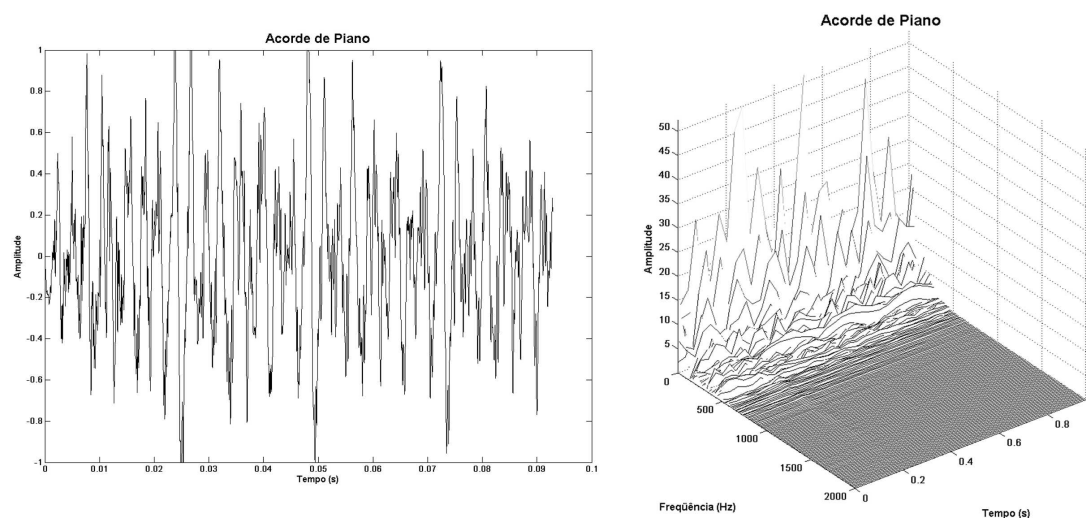


Figura III.3.6: : Forma de onda e espectro dinâmico do som atrator 4: acorde de piano.

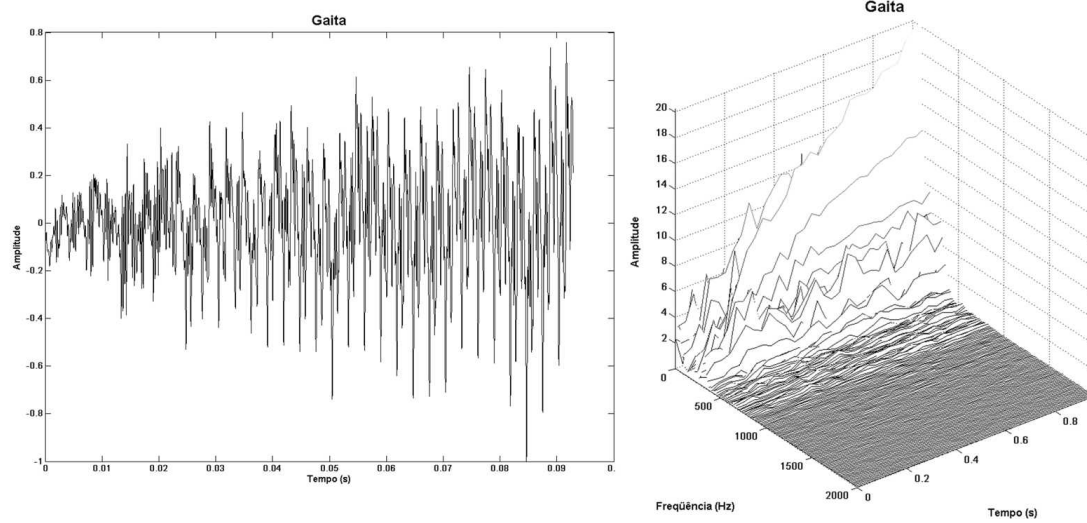


Figura III.3.7: : Forma de onda e espectro dinâmico do som atrator 5: gaita.

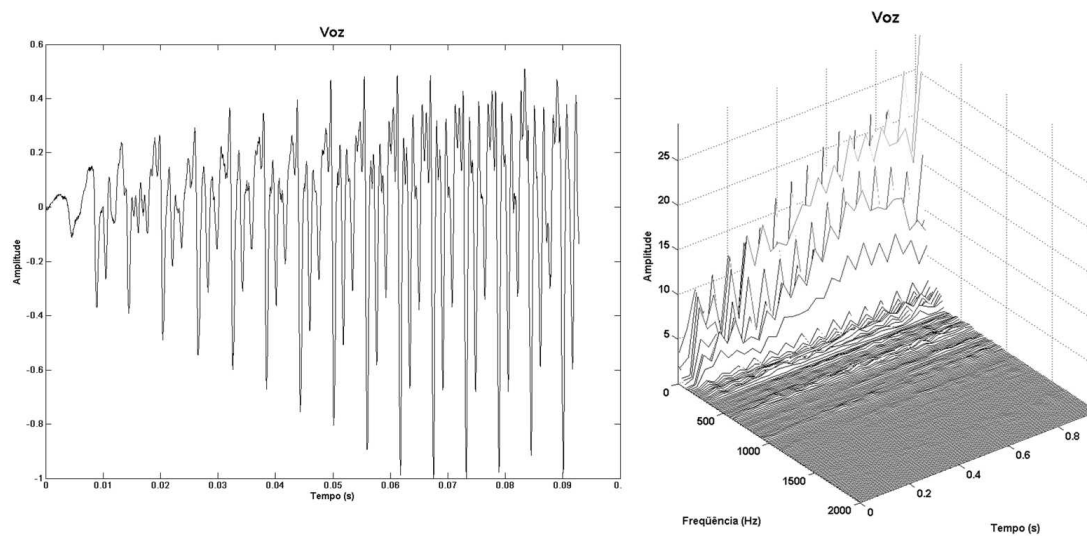


Figura III.3.8: : Forma de onda e espectro dinâmico do som atrator 6: voz.

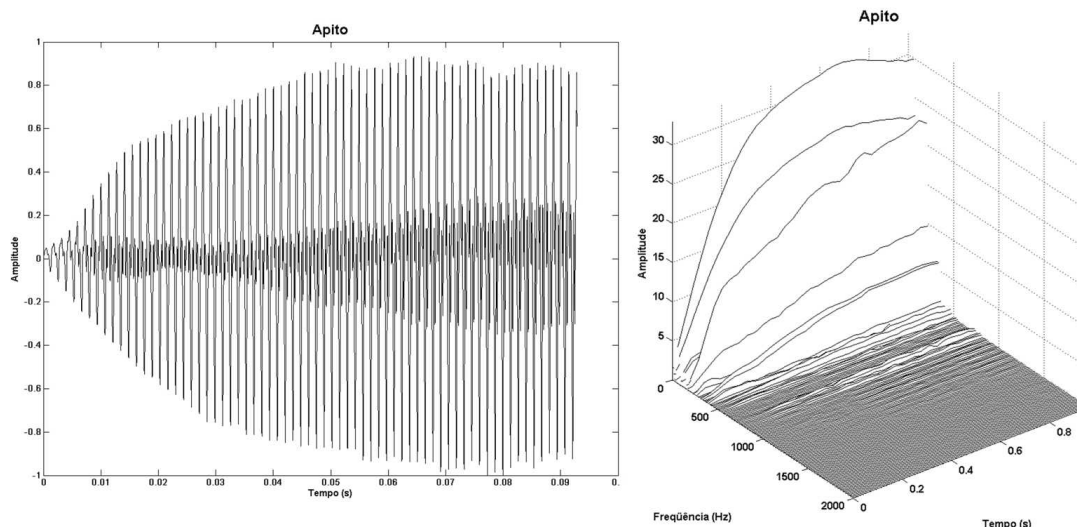


Figura III.3.9: Forma de onda e espectro dinâmico do som atrator 7: apito.

Para uma completa caracterização da região do espaço definida por estes sons escolhidos, é necessário apresentar as distâncias entre eles. As distâncias serão apresentadas na forma de tabelas simétricas em relação à diagonal principal, isto é, todos os sons nas linhas e nas colunas, como pode ser verificado a seguir. A análise desses dados deve ser feita em duas etapas, devido ao fato de as distâncias serem apresentadas aos pares. Isto significa que a informação referente à distância relativa de um som para todos os outros (como se esperaria de uma análise visual) não está contida nas tabelas de distância. A Figura III.3.10 apresenta um exemplo ilustrativo de 3 pontos no plano e suas respectivas distâncias aos pares. A Tabela III.3.2 e a Tabela III.3.3 apresentam essencialmente a mesma informação que as tabelas de distâncias aos pares apresentadas para os sons atratores, como a Tabela III.3.4, e as tabelas de distância mínima, como a Tabela III.3.5. Este exemplo ilustra porque não é suficiente analisar uma diagonal de uma tabela de distâncias aos pares para se extrair informação a respeito da proximidade dos pontos (sons atratores). Isto é, devem-se cruzar as informações de distância mínima como exemplificado adiante para a Tabela III.3.2 e a Tabela III.3.3. Esse exemplo ilustrativo é considerado importante para a compreensão da extração da informação de proximidade das tabelas de distâncias aos pares e distâncias mínimas para os sons atratores porque esses sons estão representados em um espaço vetorial de alta dimensionalidade que não admite visualização direta.

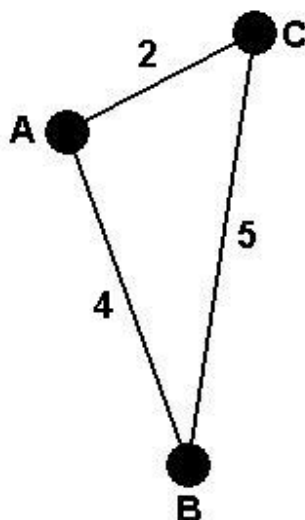


Figura III.3.10: Representação da distância espacial aos pares de três pontos no plano.

Examinando-se a figura, é fácil inferir apenas visualmente que os pontos A e C estão próximos um do outro, enquanto que o ponto B está distante dos dois. No entanto, quando se possui apenas a informação das distâncias aos pares, como apresentado na Tabela III.3.2, deve-se fazer uma análise mais cuidadosa. A análise consiste de se extrair a distância mínima ao longo das linhas de uma tabela de distâncias aos pares como a Tabela III.3.2, marcadas em **negrito**, e as distâncias mínimas ao longo das colunas, marcadas em *itálico*. Os pontos que forem marcados tanto nas linhas como nas colunas se encontram próximos no espaço. Caso contrário não se pode afirmar nada. No exemplo da Figura III.3.10, o ponto mais próximo de A é C e o ponto mais próximo de C é A, fazendo com que esta distância fique marcada em **negrito** e em *itálico* na Tabela III.3.2. No entanto, o ponto mais próximo de B é A, mas o ponto mais próximo de A não é B. Nesse caso em particular, pode-se concluir que A e C estão próximos e B está distante dos dois.

Tabela III.3.2: Distâncias dois a dois entre os pontos mostrados na Figura III.3.10.

	A	B	C
A	0	3	<i>1</i>
B	3	0	5
C	<i>1</i>	5	0

A mesma informação pode ser extraída de uma tabela de distâncias mínimas como a Tabela III.3.3 a seguir. O ponto mais próximo de A é C e vice versa, portanto A e C estão próximos. Não se pode afirmar o mesmo de B.

Tabela III.3.3: Ponto mais próximo de cada um dos pontos apresentados na figura III.3.10.

A	B	C
C	A	A

Após esse exemplo ilustrativo, a análise das posições relativas dos atratores fica mais clara a partir dos dados de distâncias aos pares. As distâncias entre os atratores calculadas pela métrica euclidiana são apresentadas na Tabela III.3.4. Ao se analisar os dados da Tabela III.3.4, as distâncias mínimas ao longo das linhas foram marcadas em negrito e ao longo das colunas foram marcadas em itálico.

Tabela III.3.4: Distâncias entre os atratores calculadas pela métrica euclidiana.

	Som 1	Som 2	Som 3	Som 4	Som 5	Som 6	Som 7
Som 1	0	37.4232	30.6717	29.0806	22.0268	33.3855	29.8475
Som 2	37.4232	0	27.7606	37.1827	29.7050	30.1695	36.3239
Som 3	30.6717	27.7606	0	35.6427	29.7071	28.5751	35.0325
Som 4	29.0806	37.1827	35.6427	0	35.5446	36.8538	27.0144
Som 5	22.0268	29.7050	29.7071	35.5446	0	30.0083	36.7353
Som 6	33.3855	30.1695	28.5751	36.8538	30.0083	0	37.6538
Som 7	29.8475	36.3239	35.0325	27.0144	36.7353	37.6538	0

A Tabela III.3.5 apresenta as distâncias mínimas individualmente, marcadas em negrito na Tabela III.3.4.

Tabela III.3.5: Atratores mais próximos de cada atrator de acordo com a métrica euclidiana.

Som 1	Som 2	Som 3	Som 4	Som 5	Som 6	Som 7
5	3	2	7	1	3	4

As distâncias entre os atratores calculadas pela métrica tempo-freqüencial são apresentadas na Tabela III.3.6. Ao se analisar os dados da Tabela III.3.6, as distâncias mínimas ao longo das linhas foram marcadas em negrito e ao longo das colunas foram marcadas em itálico.

Tabela III.3.6: Distâncias entre os atratores calculadas pela métrica tempo-freqüencial.

	Som 1	Som 2	Som 3	Som 4	Som 5	Som 6	Som 7
Som 1	0	192.6836	179.5933	172.5182	137.8076	124.9343	170.3028
Som 2	192.6836	0	193.2999	188.1357	157.6999	164.8004	189.0845
Som 3	179.5933	193.2999	0	189.0845	149.7575	147.8129	182.2963
Som 4	172.5182	188.1357	189.0845	0	133.9777	143.5312	159.7946
Som 5	137.8076	157.6999	149.7575	133.9777	0	105.4125	139.3060
Som 6	124.9343	164.8004	147.8129	143.5312	105.4125	0	146.5382
Som 7	170.3028	189.0845	182.2963	159.7946	139.3060	146.5382	0

A Tabela III.3.7 apresenta as distâncias mínimas individualmente, marcadas em negrito na Tabela III.3.6.

Tabela III.3.7: Atratores mais próximos de cada atrator de acordo com a métrica tempo-freqüencial.

Som 1	Som 2	Som 3	Som 4	Som 5	Som 6	Som 7
6	5	6	5	6	5	5

Note-se que, como esperado, ambas as métricas resultam em distância zero de um som para si próprio. Deve-se também notar que, de acordo com a análise de proximidade dos sons atratores, a métrica euclidiana (temporal) indica que o som atrator 1 está próximo do som atrator 5, o som atrator 2 está próximo do som atrator 3 e o som atrator 4 está próximo do som atrator 7. A métrica tempo-freqüencial indica que o som atrator 6 está próximo do som atrator 5. A mesma análise pode ser repetida para as outras distâncias, levando à conclusão de que as métricas não são equivalentes, isto é, que proximidades de acordo com a métrica euclidiana não são necessariamente refletidas como proximidades na métrica tempo-freqüencial. Resta comparar as métricas com uma medida subjetiva perceptiva das distâncias entre os sons. Isto se faz necessário porque a análise do cruzamento de características presentes em mais de um atrator nos resultados se faz de forma perceptiva, isto é, ouvindo-se os atratores e os resultados.

A proximidade subjetiva (perceptiva) entre os atratores é apresentada na Tabela III.3.8 a seguir, como relatada por 5 indivíduos que aferiram a similaridade entre os sons. Os sons foram apresentados aos pares e pediu-se que os indivíduos aferissem a proximidade perceptiva entre eles em uma escala que varia de 1 a 5. O valor 1 corresponde a pouca proximidade, ou grande distância (distinção) entre os sons e o valor

5 corresponde a uma proximidade grande, refletida por uma distância perceptiva pequena entre os sons. Note-se que, agora, está-se medindo similaridade subjetiva, ao invés de distância. Pode-se inferir, no entanto, que ambas são complementares, isto é, uma grande similaridade pressupõe uma pequena distância subjetiva. Destaca-se que os 5 indivíduos não têm treinamento musical, no entanto, possuem treinamento científico suficiente para compreender a tarefa que lhes foi solicitada, isto é, a avaliação subjetiva da similaridade entre pares de sons em uma escala de 1 a 5. A Tabela III.3.8 apresenta a média e o desvio padrão da aferição subjetiva dos indivíduos. É importante, nesse ponto, ressaltar que, devido ao pequeno número de aferições (apenas 5 indivíduos), os erros são bastante significativos. Assim como, devido ao fato de que a escala só possui 5 divisões, os resultados das aferições são bastante susceptíveis a ruído de quantização, isto é, o resultado de uma aferição pode não expressar com precisão suficiente a percepção subjetiva do indivíduo aferidor. No entanto, considerou-se que, para os propósitos desse trabalho, a mera confirmação dos resultados expressos pelas métricas matemáticas de distância, os altos valores de erro e ruído não prejudicam os resultados.

Finalmente, devido à natureza estatística e ao alto valor de erro da avaliação perceptiva, é possível destacar mais de um som em uma linha como representando estatisticamente o mesmo valor.

Tabela III.3.8: Similaridades subjetivas entre os atratores estimadas pelos indivíduos.

	Som 1	Som 2	Som 3	Som 4	Som 5	Som 6	Som 7
Som 1	5,0±0,0	1,2±0,45	2,4±1,14	2,4±1,14	3,6±0,55	1,8±0,84	3,4±0,89
Som 2	1,2±0,45	5,0±0,0	3,0±0,71	1,8±0,84	1,4±0,55	2,2±0,84	1,4±0,55
Som 3	2,4±1,14	3,0±0,71	5,0±0,0	3,2±1,30	3,2±1,30	1,4±0,55	1,6±0,89
Som 4	2,4±1,14	1,8±0,84	3,2±1,30	5,0±0,0	2,4±0,55	1,2±0,45	2,2±0,84
Som 5	3,6±0,55	1,4±0,55	3,2±1,30	2,4±0,55	5,0±0,0	1,4±0,55	2,6±0,55
Som 6	1,8±0,84	2,2±0,84	1,4±0,55	1,2±0,45	1,4±0,55	5,0±0,0	2,0±1,0
Som 7	3,4±0,89	1,4±0,55	1,6±0,89	2,2±0,84	2,6±0,55	2,0±1,0	5,0±0,0

Tabela III.3.9: Agrupamento das similaridades subjetivas entre os atratores como estimado pelos indivíduos. X significa não similaridade.

Som 1	Som 2	Som 3	Som 4	Som 5	Som 6	Som 7
5,7	3	2,4,5	3	1,3	X	1

De acordo com a avaliação perceptiva, os sons atratores 1, 5 e 7 estão próximos e os sons atratores 2,3, 4 e 5 formam outro grupo. É importante notar que o som atrator 5 pertence a ambos os grupos, indicando um elo entre os grupos. Comparando-se, finalmente, a avaliação subjetiva destes sons com as medidas de distância, é possível afirmar que nenhuma das duas métricas reflete totalmente a avaliação perceptiva, apenas parcialmente. No entanto, algumas das relações passíveis de ser extraídas da Tabela III.3.5 e da Tabela III.3.7 podem ser verificadas nas tabelas relativas a ambas as métricas matemáticas. Pode-se concluir, assim, que cada métrica reflete uma certa particularidade da avaliação subjetiva.

III.4 Por que Bio-Inspiração?

Os algoritmos bio-inspirados possuem inúmeras características desejadas em tarefas de exploração de espaços de busca complexos. São inspirados na maneira como de sistemas biológicos resolvem problemas complexos de maneira adaptativa em ambientes dinâmicos, isto é, cujas condições variam no tempo. Algumas dessas características relevantes são:

- Auto-organização;
- Emergência de padrões;
- Aprendizado;
- Adaptação;
- Manutenção de diversidade;
- Preservação topológica.

Em particular, cada algoritmo bio-inspirado representa uma maneira diferente de explorar o espaço de busca devido às diferenças intrínsecas entre eles. Ao longo das iterações, a população de agentes flui de maneira distinta pelo espaço, dependendo da dinâmica dos processos envolvidos. É justamente esse fluxo populacional, juntamente com um subconjunto das características acima que cada algoritmo possui, que justifica a aplicação de mais de um algoritmo ao mesmo problema. Pretende-se estudar, portanto, o potencial musical da aplicação de diferentes estratégias de exploração do espaço composicional.

III.4.i) Auto-Organização em Sistemas Biológicos

Padrões e estruturas parecidas aparecem em todo o mundo natural. Muitos desses padrões e estruturas compartilham o mecanismo intrínseco através do qual os padrões emergem através de auto-organização. O fenômeno da auto-organização pode ser descrito como um processo no qual um padrão global emerge a partir de numerosas interações locais de componentes de um sistema [11]. Sendo assim, as regras especificando as interações dos componentes do sistema são executadas usando apenas informação local, sem referência ao padrão global. A essência da auto-organização está no surgimento de estruturas (formas restritas) e organização sem que estas sejam impostas externamente ao sistema [11]. Em sistemas auto-organizados, emergem propriedades ausentes em seus componentes e que caracterizam estruturas não estabelecidas a priori [76].

Os primeiros estudos de auto-organização tratavam de sistemas físicos e químicos (Figura III.4.1 a) [11]. Quando nos voltamos para os sistemas biológicos, em particular sistemas celulares envolvidos em processos de desenvolvimento, a auto-organização foi invocada para explicar, por exemplo, padrões de pele e superfície em mamíferos (Figura III.4.1 b), conchas, asas de borboletas e de besouros (Figura III.4.1 c), peixes, répteis e anfíbios. Além desses exemplos envolvendo animais, a formação de padrões auto-organizados freqüentemente se revela em plantas e fungos: os padrões divisórios de *slime molds* e os movimentos de amebas individuais; a ornamentação na superfície de grãos de pólen e cogumelos; o crescimento de líquens; e o arranjo espiral de partes de flores em margaridas e pinhas (Figura III.4.1 d). Entre os insetos sociais, a auto-organização mostrou fazer parte no comportamento construtivo, tomada de decisão, sincronização de atividades e formação de trilhas. Muito da complexidade de estruturas auto-organizadas em biologia surge porque as regras que governam as interações dos componentes dos sistemas biológicos evoluíram pela seleção natural [11]. O processo de evolução gerou uma enorme diversidade de comportamentos e interações fisiológicas, complementando a diversidade de interações possíveis em sistemas químicos e físicos. Os padrões emergentes e processos de tomada de decisão desses sistemas são normalmente não-intuitivos, devido ao grande número de interações não-lineares envolvidas.

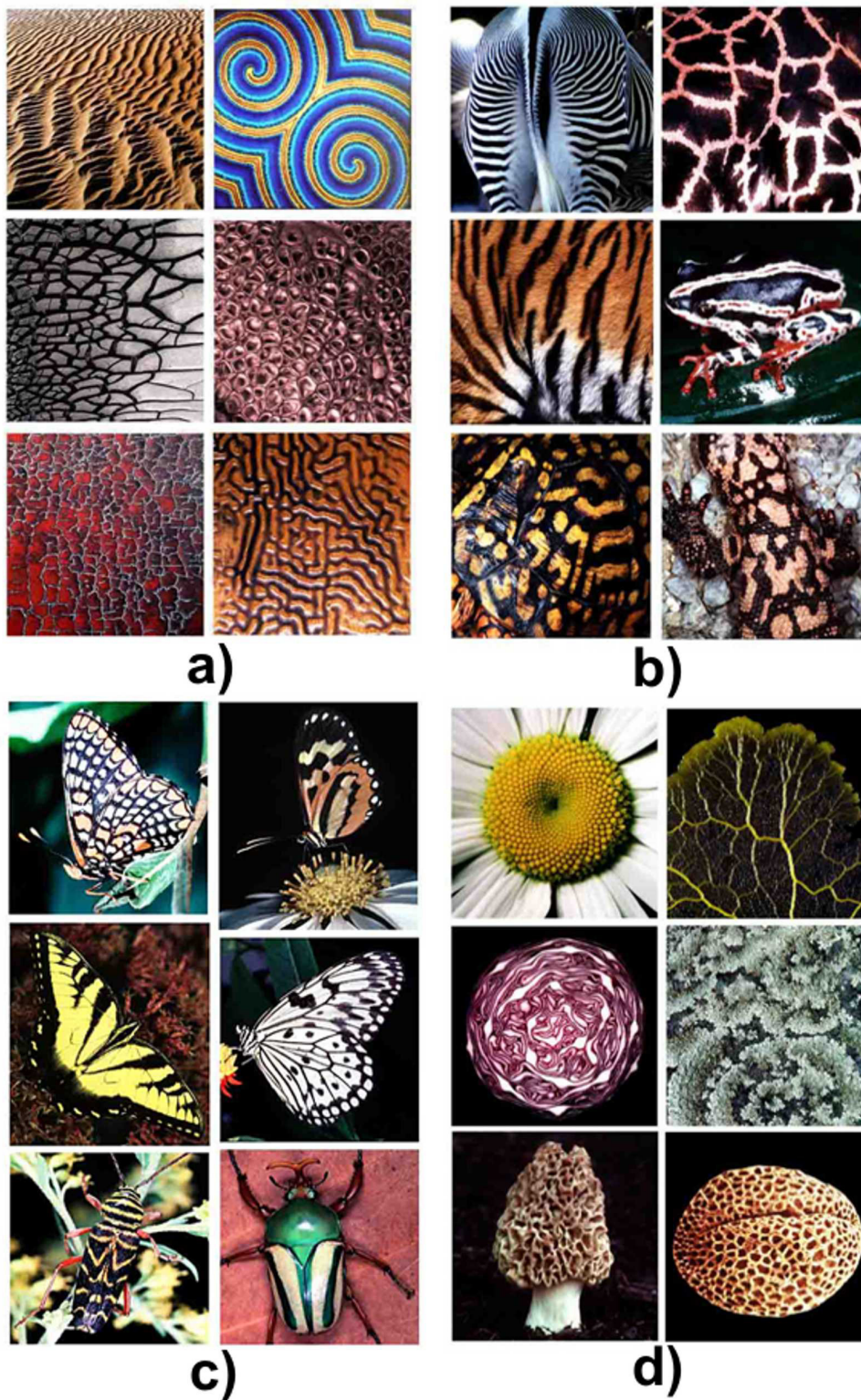


Figura III.4.1: Exemplos dos diversos padrões biológicos cuja estruturação pode ser explicada através de processos auto-organizáveis. Retirada de Camazini *et al.* [11].

Por essa razão, modelos matemáticos e simulações computacionais provêm técnicas úteis no estudo de sistemas e processos auto-organizados e na exploração das consequências das inúmeras interações de sub-unidades componentes dos sistemas. Muitas dessas simulações são baseadas em agentes, isto é, cada sub-unidade individual é monitorada durante a simulação e seu comportamento no tempo é determinado através de interações locais com outras sub-unidades e com informações do ambiente. Este é precisamente o caso deste trabalho, em que serão realizadas simulações onde cada forma de onda caracteriza um agente de uma população realizando a busca no espaço. Os agentes interagem localmente e o conjunto resultante, ao invés de indivíduos isolados, representa aquela região do espaço sonoro em particular, caracterizado pelos atratores (sons definidos *a priori* pelo usuário/compositor).

Os sistemas auto-organizáveis possuem características altamente relevantes para o problema em questão, isto é, o posicionamento dos agentes no espaço contínuo, influenciado pela posição dos atratores e seguindo algum critério de otimização sem elemento organizador externo, guiada apenas pelos mecanismos de sistemas auto-organizáveis visando a emergência de padrões sonoros complexos. Tais características são as listadas a seguir:

- Sistemas auto-organizáveis são dinâmicos;
- Sistemas auto-organizáveis exibem propriedades emergentes;
- Sistemas auto-organizáveis realizam auto-ajuste de parâmetros internos;
- Auto-organização pode promover padrões estáveis;
- Padrões complexos podem emergir de regras simples em sistemas auto-organizáveis.

Primeiramente, deve-se entender que, neste texto, padrão significa um arranjo específico de amostras sonoras que compõe uma forma de onda. O termo complexidade também deve ser discutido. Aqui, padrões complexos representam sons que possam ter aplicações musicais. A complexidade de um som neste trabalho está diretamente ligada à quantidade de informação sonora que a sua forma de onda possui. Uma oscilação senoidal simples não possui muita informação, apenas uma frequência constante no tempo. No outro extremo do espectro, ruído branco, correspondente neste caso a um arranjo aleatório das amostras sonoras, também não apresenta muita informação para ser processada pelo ouvido. Neste caso, complexidade está ligada ao fato de que os

indivíduos usam regras comportamentais relativamente simples para gerar estruturas que são mais complexas que os componentes e processos envolvidos. Complexidade não requer componentes complicados nem inúmeras regras complicadas.

Sistemas auto-organizáveis são tipicamente constituídos de um grande número de componentes ou eventos relativamente simples interagindo localmente. Em particular, as subunidades em sistemas biológicos adquirem informação sobre as propriedades locais do sistema e se comportam de acordo com regras imbuídas em seus códigos genéticos, que estão sujeitos à seleção natural de características. Isto é, cada elemento do sistema troca informação apenas com elementos que se encontram fisicamente próximos, ou que, de alguma maneira, modificaram aquela região específica do ambiente onde se encontra o elemento levado em consideração [11]. A informação pode fluir dentro de grupos via dois mecanismos distintos: sinais e pistas [11]. Sinais são estímulos usados para troca de informação que tomaram forma através da seleção natural, enquanto que pistas são estímulos que passam informação acidentalmente. As interações em sistemas biológicos auto-organizáveis são baseadas em ambos, sinais e pistas, diferindo apenas na maneira como a informação é trocada pelos agentes. Pode-se trocar informações diretamente com os vizinhos ou com as pistas deixadas no ambiente por outros agentes. Um exemplo biológico de troca de informação local entre vizinhos é como os peixes coordenam seus movimentos em um cardume. Em cardumes com milhares de peixes, é inconcebível que haja um líder monitorando as posições individuais de cada membro ou que cada peixe seja capaz de monitorar os movimento de um líder que não seja um vizinho próximo. O conceito de que indivíduos podem operar efetivamente com informação adquirida somente de seus vizinhos próximos se aplica não somente a peixes em um cardume, mas a muitos sistemas auto-organizados [11].

Informação adquirida diretamente dos vizinhos é apenas uma fonte de informação usada pelos organismos em sistemas auto-organizáveis. Em situações em que muitos indivíduos contribuem coletivamente para um esforço comum, tal como cupins construindo um cupinzeiro, o estímulo fornecido pela estrutura emergente pode ser uma rica fonte de informação para os indivíduos. Em outras palavras, a informação do ambiente e do trabalho em progresso pode guiar a atividade futura. Conforme a estrutura do cupinzeiro se desenvolve, o estado da construção fornece continuamente informação nova para os construtores. No estudo de insetos sociais, o termo *estigmergia* foi usado para descrever esta atividade construtiva recursiva. Durante o trabalho stigmergético, é o produto do

trabalho já realizado, ao invés da comunicação entre vizinhos, que induz os insetos a continuar trabalhando. Durante os estágios iniciais da atividade estigmergética, flutuações aleatórias e heterogeneias fortuitas podem ser amplificadas pela realimentação positiva para criar estruturas.

Há duas formas básicas de interação em sistemas auto-organizáveis: realimentação positiva e negativa [11]. A realimentação positiva tem a característica de amplificar uma mudança ocorrida no sistema, tal como uma bola de neve. Em contraste com a realimentação negativa, a realimentação positiva promove mudanças em um sistema. Dadas condições iniciais aleatórias, o mecanismo de realimentação positiva é responsável por favorecer a emergência de padrões através da amplificação de flutuações aleatórias [11]. A natureza amplificadora da realimentação positiva significa que a mesma tem o potencial de produzir explosões destrutivas em qualquer processo de que faça parte. Há, portanto, a necessidade de um mecanismo capaz de conter a explosão em qualquer processo em que a realimentação positiva atue. É aí que a realimentação negativa tem papel crítico, garantindo a inibição necessária para contrabalançar a amplificação e estabilizar o processo, permitindo a formação de um padrão estável [11]. Um exemplo importante de mecanismo de realimentação negativa é o processo biológico conhecido como homeostase. Homeostase é a propriedade de um sistema de regular seu estado interno para manter uma condição estável e constante através de múltiplos ajustes do equilíbrio dinâmico do sistema, controlado por mecanismos de regulação interna [47]. Os níveis de açúcar no sangue, por exemplo, são um exemplo de regulação homeostática no organismo. Bem como o controle da temperatura corporal. Realimentação positiva que se auto-alimenta acoplada com realimentação negativa fornece um mecanismo poderoso para a criação de estruturas e padrões em muitos sistemas físicos e biológicos envolvendo um grande número de componentes.

Mostrou-se que sistemas auto-organizáveis possuem estruturas emergentes. Emergência se refere a um processo através do qual um sistema de subunidades que interagem adquire propriedades qualitativas novas que não podem ser entendidas simplesmente adicionando-se as contribuições individuais. Na terminologia de sistemas dinâmicos estes padrões emergentes são chamados atratores. Dadas certas condições iniciais e valores dos parâmetros, um atrator é um estado para o qual o sistema converge no tempo. Existem parâmetros intrínsecos aos sistemas e parâmetros externos que

surgem do ambiente ou de limitações físicas. A distinção é irrelevante, no entanto, para o processo de formação de padrões. Ambos os tipos de parâmetros podem ser ajustados com conseqüências surpreendentes em sistemas auto-organizáveis. Os parâmetros internos dos sistemas biológicos são ajustados pela seleção natural, enquanto que o meio ambiente está em constante mudança. Estas propriedades dos sistemas auto-organizáveis fornecem respostas flexíveis e adaptativas às condições variáveis do meio ambiente. Pequenos ajustes nos parâmetros individuais em um grupo podem acarretar em grandes mudanças nas propriedades coletivas do grupo, resultando em uma vasta gama possível de respostas e permitindo ao grupo trocar de comportamento em resposta a diferentes estímulos.

A maioria dos sistemas auto-organizáveis é bastante robusta, significando que são estáveis em uma gama variável de valores dos parâmetros. Dado que a seleção natural favorece alguns padrões, espera-se que a maioria dos sistemas auto-organizáveis resista a perturbações e opere com grande estabilidade em um único regime. Neste caso, o padrão exibe a propriedade de auto-reparação. Em outras palavras, o padrão é um atrator do sistema. O conceito de auto-organização alerta para a possibilidade de que padrões extremamente diferentes podem resultar do mesmo mecanismo operando em diferentes ajustes de parâmetros.

A maioria dos sistemas auto-organizáveis é altamente complexa e usa poucas regras simples. Interações não lineares de um grande número de indivíduos podem levar a padrões surpreendentemente complexos no nível do grupo. Os biólogos tentam entender como a quantidade de informação guardada nos genes é muito menor que a quantidade de informação necessária para descrever a estrutura de um indivíduo. Sabe-se que é impossível cada detalhe de um organismo estar explicitamente codificado no genoma. O mistério pode ser resolvido se se notar que os genes não precisam carregar informação a respeito da estrutura, mas apenas carregar um conjunto de regras que geram esta informação [11]. Na música computacional, especialmente em síntese sonora, o mesmo pode ser dito a respeito da complexidade da forma de onda de um som musical. Utilizando-se a auto-organização como paradigma em síntese sonora, não é mais necessário descrever ou modelar exaustivamente todas as minúcias das formas de onda através de um sem número de parâmetros, como nos métodos de síntese sonora tradicionais apresentados na seção II.3. Ao realizar a busca no espaço vetorial de alta dimensionalidade que representa diretamente as formas de onda, a emergência de

padrões complexos neste espaço, por conta da estratégia auto-organizável, é diretamente utilizada como o resultado sonoro desejado. Um dos objetivos principais dessa pesquisa é verificar o potencial musical da utilização de algoritmos de otimização bio-inspirados em música computacional. Os algoritmos a serem considerados utilizam a auto-organização como paradigma; isto é, exploram características de sistemas auto-organizáveis, modeladas nos algoritmos bio-inspirados, para síntese sonora. Está-se utilizando a capacidade de aprendizado, de adaptação e de manutenção de diversidade que caracterizam os padrões emergentes de processos auto-organizáveis como recursos para o posicionamento dos sons sintetizados no espaço de busca contínuo digitalizado, isto é, discretizado e quantizado, nas regiões de interesse indicadas pelos atratores. Decorre das características dos algoritmos utilizados que o posicionamento será, na grande maioria das vezes, intermediário em relação aos atratores, resultando em variantes desses ou híbridos que mesclam características perceptivas freqüenciais, timbrais ou outras possíveis.

IV) SÍNTESE SONORA POR SISTEMAS IMUNOLÓGICOS ARTIFICIAIS

Este capítulo se destina à apresentação da primeira implementação das idéias desenvolvidas no texto, a utilização de um algoritmo de agrupamento de dados inspirado em modelos do sistema imunológico biológico em síntese sonora digital. Primeiramente serão apresentados alguns conceitos a respeito do sistema imunológico natural e um modelo simplificado de seu funcionamento. Em seguida, o algoritmo utilizado, inspirado em algumas teorias imunológicas, é apresentado e sua aplicação em síntese sonora é formalizada. Por fim, os resultados do experimento realizado são apresentados e discutidos sob a luz dos conceitos apresentados anteriormente e do critério de avaliação de um método de síntese sonora.

IV.1 Introdução

O sistema imunológico dos vertebrados é um conjunto complexo de células, moléculas e órgãos distribuídos com o importante papel de manter um estado de equilíbrio dinâmico interno no organismo. Sua complexidade já foi comparada à do cérebro em muitos aspectos: os sistemas imunológicos são capazes de reconhecer sinais internos e externos; controlar ações de componentes imunológicos; influenciar o comportamento de outros sistemas, tal como os sistemas nervoso e endócrino; e, o mais importante, adaptar-se visando maximizar a eficácia no combate a agentes causadores de doenças, normalmente associado a etapas de reconhecimento de padrões [16].

Todos os seres vivos possuem a habilidade de apresentar resistência a agentes causadores de doenças, ou patógenos. Patógenos incluem vírus, bactérias, fungos e parasitas. A natureza da resistência varia de uma espécie para outra e em função da complexidade do organismo. Os mamíferos desenvolveram um sistema imunológico altamente sofisticado que age em conjunto com diversos outros sistemas, tais como o nervoso e endócrino, em prol da manutenção da vida [16]. O principal papel do sistema imunológico é proteger o organismo de infecções causadas pelos patógenos [37], [73].

A defesa deve ocorrer em vários níveis e abranger o corpo todo. Como consequência, vários níveis de mecanismos de defesa e barreiras evoluíram como forma de oferecer proteção. O sistema imunológico pode ser dividido em sistema imunológico inato e sistema imunológico adaptativo, este último composto de diversos conjuntos de células, moléculas e órgãos que trabalham em conjunto para proteger o organismo. O sistema imunológico inato é uma primeira linha de defesa extremamente importante contra diversos tipos de patógenos e também é crucial para a regulação do sistema imunológico. As células do sistema inato são capazes de reconhecer padrões moleculares genéricos que estão presentes nos antígenos. Uma vez que um patógeno é reconhecido por uma célula do sistema imunológico, a mesma sinaliza através de mensageiros químicos para que outras células, inclusive do sistema adaptativo, comecem a combater o patógeno. Desta maneira, o sistema inato é responsável pelos co-estímulos, sem os quais, para a maior parte dos patógenos, o sistema adaptativo não é capaz de agir. No entanto, nem todos os patógenos podem ser reconhecidos pelo sistema inato, sendo deixados ao encargo do sistema adaptativo. O sistema adaptativo possui algumas características bastante relevantes do ponto de vista computacional, tais como adaptação a padrões moleculares de exposições recentes e geração e manutenção de memória. Leva algum tempo para que o sistema adaptativo desenvolva uma resposta imunológica a um dado patógeno, então o sistema inato tenta eliminar o patógeno até que o sistema adaptativo entre efetivamente em ação. Uma vez que o sistema adaptativo esteja preparado para agir, este é capaz de se adaptar a um dado patógeno e criar padrões moleculares específicos para combater a mesma infecção no futuro, ou uma parecida. O princípio da vacinação está arraigado a esta capacidade de aprendizado e memória do sistema imunológico. Inoculando-se indivíduos saudáveis com amostras enfraquecidas de agentes causadores de doenças, é possível fazer com que o sistema imunológico desenvolva conjuntos de estruturas moleculares específicas para combater aquele patógeno sem sujeitar o organismo aos sintomas da doença.

Por fim, há teorias que sugerem que o sistema imunológico é um sistema dinâmico cujas células e moléculas são capazes de interagir entre si. Sob este prisma, os patógenos são responsáveis por modificações na estrutura dinâmica do sistema [38]. A perspectiva mais tradicional prega que o sistema imunológico é composto por conjuntos discretos de células e moléculas que são ativadas apenas por patógenos [36].

IV.1.i) Modelo Simplificado do Sistema Imunológico

Uma característica marcante do sistema imunológico é a distributividade pelo organismo. Há células, moléculas e órgãos por todo o organismo. Os órgãos que compõem o sistema imunológico têm duas funções básicas: os órgãos linfóides primários são responsáveis pela produção, crescimento, desenvolvimento e maturação dos linfócitos ou glóbulos brancos, células do sangue especializadas que possuem receptores específicos para reconhecer patógenos; e os órgãos linfóides secundários são onde os linfócitos interagem com os patógenos. Os principais tipos de linfócitos são os linfócitos B e os linfócitos T. Ambos possuem moléculas receptoras em sua superfície que são capazes de reconhecer padrões moleculares específicos presentes nos patógenos. Elas também são capazes de se multiplicar através de um processo chamado clonagem, e os linfócitos B são capazes de sofrer variação genética durante a reprodução. Portanto, essas células podem ter tanto seu número quanto sua estrutura molecular variando no tempo.

Os dois órgãos linfóides primários - a medula óssea e o timo - têm funções importantes na produção e maturação das células imunológicas. Todas as células do sangue são geradas na medula óssea, e os linfócitos B também são maturados, ou seja, se tornam prontas para agir como células imunológicas na medula óssea. Os linfócitos T são gerados na medula óssea, mas migram para o timo, onde vão se tornar células ativas do sistema imunológico. Acredita-se também que o timo é o local onde os linfócitos T aprendem a reconhecer as células próprias do organismo, de acordo com o modelo de seleção negativa dos linfócitos T.

O primeiro passo para promover uma resposta imunológica contra infecção é o reconhecimento de um patógeno. As porções dos patógenos que promovem uma resposta imunológica são chamadas antígenos. Os antígenos possuem moléculas em sua superfície chamadas epítomos, os quais podem ser reconhecidos pelas moléculas e células do sistema imunológico. A Figura IV.1.1 ilustra linfócitos B e seus receptores (anticorpos) reconhecendo os epítomos de um dado antígeno. Os linfócitos B são monoespecíficos, significando que possuem apenas um tipo de receptor, enquanto os antígenos podem apresentar vários tipos de epítomos, isto é, podem ser reconhecidos por vários tipos diferentes de anticorpos.

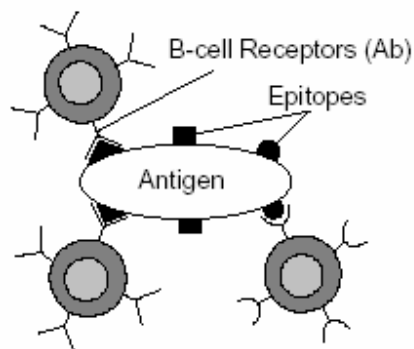


Figura IV.1.1: Ilustração do reconhecimento de um anticorpo por um antígeno.

A figura mostra o antígeno com seus epítomos sendo reconhecidos pelos receptores dos linfócitos B, chamados anticorpos. É importante notar que cada linfócito B produz anticorpos que reconhecem apenas um tipo de epítomo, isto é, são monoespecíficos. Retirado de de Castro & Timmis [17].

Para que o receptor de uma célula imunológica reconheça um antígeno, o epítomo deve apresentar complementaridade de forma com a molécula receptora. A molécula receptora é a parte do anticorpo responsável por reconhecer um antígeno através da complementaridade de forma com o epítomo, e é chamada paratopo. A complementaridade não precisa ser total para que haja reconhecimento. A força e especificidade da interação entre antígeno e anticorpo é chamada afinidade. O paratopo também é chamado região variável devido a sua capacidade de alterar sua forma de modo a atingir uma melhor complementaridade com um dado antígeno. No entanto, apenas reconhecimento não é suficiente para causar uma resposta imunológica, deve haver uma ligação entre o paratopo e o epítomo para que haja sinalização para que outros componentes do sistema imunológico entrem em ação. O complexo formado pelo antígeno coberto por anticorpos sinaliza para outras células chamadas fagócitos para que estas destruam o complexo e, portanto, o antígeno.

A Figura IV.1.2 a seguir mostra um modelo simplificado do funcionamento do sistema imunológico a fim de facilitar o entendimento dos mecanismos e teorias imunológicas a serem explicadas adiante e exploradas computacionalmente no algoritmo aiNet [18], utilizado neste trabalho.

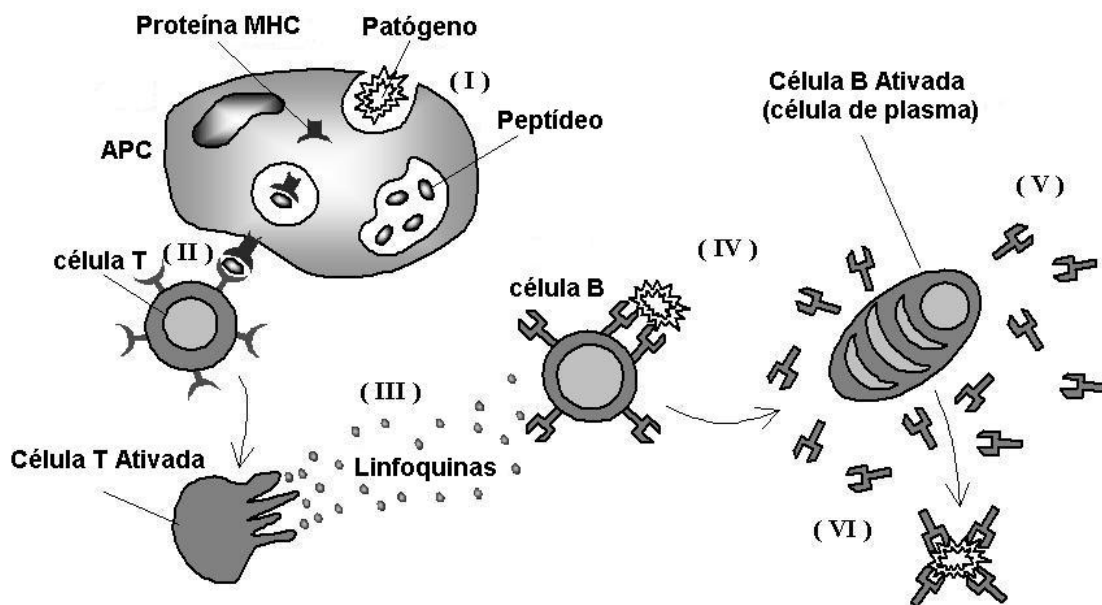


Figura IV.1.2: Modelo simplificado do funcionamento do sistema imunológico.

É mostrado um esquema simplificado das etapas desde o reconhecimento de um antígeno pelas células apresentadoras de antígeno, até a produção das células de memória. Retirado de de Castro & Timmis [17].

Na parte (I) da Figura IV.1.2, vêem-se as células apresentadoras de antígeno (APC), células especializadas, tais como os fagócitos, que circulam pelo organismo ingerindo e digerindo antígenos. Os antígenos são fragmentados em peptídeos antigênicos [59]. Partes desses peptídeos se ligam em moléculas chamadas MHC, e todo esse complexo é apresentado na superfície das células APC (II). Os linfócitos T carregam em sua superfície receptores que os permitem reconhecer diferentes complexos MHC-peptídeo. Uma vez ativados pelo reconhecimento de um complexo MHC-peptídeo, os linfócitos T se dividem e secretam sinais químicos que estimulam outros componentes do sistema imunológico a entrar em ação (III). Em contraste com os linfócitos T, os linfócitos B têm receptores com a habilidade de reconhecer partes dos antígenos quando livres em solução (IV). Os receptores na superfície dos linfócitos B são ativados por um único antígeno específico. Quando os receptores dos linfócitos B recebem um sinal, o linfócito B é ativado e se prolifera, diferenciando-se em células de plasma que secretam anticorpos a altas taxas (V). Esses anticorpos liberados são usados para neutralizar os patógenos (VI), levando à sua destruição. Alguns desses linfócitos B e T ativados se diferenciam em células de memória que permanecem circulando pelo corpo por longos períodos de

tempo, garantindo proteção contra futuras infecções do mesmo antígeno que causou a resposta ou de outro muito parecido.

IV.1.ii) Sistemas Imunológicos Artificiais

Depois do sucesso dos Algoritmos Genéticos (AGs) [35], um outro paradigma de computação bio-inspirada vem sendo explorado: os Sistemas Imunológicos Artificiais (SIAs). Os SIAs são sistemas adaptativos inspirados no sistema imunológico natural e estão sendo amplamente utilizados em problemas de otimização, busca, reconhecimento de padrões, segurança de redes, entre outros [17]. Dasgupta os define como “uma composição de metodologias inteligentes inspiradas no sistema imunológico natural aplicadas à resolução de problemas [14]”. Diferentemente dos algoritmos evolutivos, que se fundamentam em uma idéia central, o princípio da sobrevivência do mais apto, os sistemas imunológicos artificiais foram propostos levando-se em conta várias características, processos e modelos do sistema imunológico. Do ponto de vista computacional, a resposta imunológica adaptativa é rica em idéias e princípios passíveis de serem modelados e utilizados como algoritmos com características semelhantes às do sistema imunológico natural. A quantidade de aplicações e algoritmos presentes na literatura é vasta. No entanto, algumas idéias centrais têm sido amplamente exploradas, tais como seleção clonal [7], [60], maturação de afinidade [59], seleção negativa [24], teoria da rede imunológica [38], discriminação próprio/não próprio [36] e teoria do perigo [54]. A seguir, duas dessas teorias que inspiraram o modelo de sistema imunológico artificial utilizado neste trabalho são abordadas em detalhe, a saber: o princípio da seleção clonal e a maturação de afinidade; e a teoria da rede imunológica.

IV.1.iii) Princípio da Seleção Clonal e Maturação de Afinidade

Várias teorias foram propostas para tentar explicar como o sistema imunológico é capaz de lidar com os antígenos. Burnet e Talmage propuseram, em meados do século XVIII, a teoria da seleção clonal [7], amplamente aceita como uma explicação de como o sistema imunológico adaptativo responde a infecções por patógenos. Juntamente com a teoria da maturação de afinidade dos anticorpos [59], a seleção clonal é a base da resposta adaptativa, e é muito usada nos modelos de sistemas imunológicos artificiais.

A teoria da seleção clonal afirma que o sistema imunológico é composto por um conjunto de células e moléculas discretas que permanecem em repouso até que um agente patogênico invada o organismo. Após a invasão, um subconjunto das células

imunológicas, linfócitos B ou T, é capaz de reconhecer o antígeno invasor e se ligar a ele. Este processo de reconhecimento estimula as células imunológicas capazes de reconhecer o antígeno a começar a se multiplicar por mitose. Essa multiplicação se refere à reprodução celular e também é conhecida como clonagem. Então, um subconjunto de células (clones), todas capazes de reconhecer um dado antígeno, é gerado. Os linfócitos B com maior afinidade com o antígeno são selecionados para se proliferar (reproduzir).

Como em todo processo de reprodução celular, o processo de divisão dos linfócitos B está sujeito à mutação. Uma característica do sistema imunológico é que a mutação é inversamente proporcional à afinidade do anticorpo com o antígeno ao qual este está ligado. Os linfócitos B com maior afinidade com o antígeno são selecionados para se proliferar. Isto é, um mecanismo de seleção permite que as células mutadas com afinidades maiores que as que as originaram sejam retidas em um repertório de células de memória. Este processo é chamado maturação de afinidade [59].

Os princípios da seleção clonal e maturação de afinidade sugerem que o sistema imunológico se torna mais adaptado à estimulação provocada pelos antígenos. Pode-se considerar que os eventos de reconhecimento ocorrem em um espaço de formas onde cada eixo corresponde a uma medida físico-química característica de uma forma molecular [18]. Há dois graus de plasticidade que podem ser observados neste espaço durante o reconhecimento: a plasticidade estrutural e a paramétrica. As células imunológicas que reconheceram e se ligaram a um antígeno serão selecionadas para se proliferar, e aquelas células que não são selecionadas morrem e são eliminadas do repertório imunológico. A maturação de afinidade leva à adaptação paramétrica na estrutura molecular dos anticorpos e, portanto, contribui para a adaptabilidade no sistema imunológico e para o aprendizado de padrões moleculares (formas) presentes nos patógenos.

Quando um antígeno invade o organismo, uns poucos anticorpos capazes de reconhecer e se ligar a este antígeno serão selecionados para se proliferar. É preciso algum tempo para que estas células imunológicas comecem a se reproduzir e produzam uma resposta, chamada resposta primária. No caso de uma segunda exposição a este mesmo antígeno (ou a uma variação desse mesmo antígeno), ocorre uma resposta mais rápida e eficiente, chamada resposta secundária. A resposta secundária é mais rápida porque os anticorpos são imuno-específicos, isto é, são capazes de reconhecer apenas um antígeno ou uma variação dele. Isto é conhecido como resposta cruzada, e é

importante para a criação de um modelo do espaço de formas representado pelos antígenos. Cada célula de memória é responsável por mapear uma certa região deste espaço. Portanto, adaptação e aprendizado no sistema imunológico envolvem a variação do tamanho da população e afinidade de células específicas que foram bem sucedidas na tarefa de reconhecer e se ligar aos patógenos.

IV.1.iv) Teoria da Rede Imunológica

A teoria da rede imunológica foi formalizada por Jerne em meados da década de 1971-1980 [38]. Em contraste com a teoria da seleção clonal, que propõe um sistema imunológico composto de células e moléculas discretas em repouso e estimuladas por antígenos, a teoria da rede imunológica sugere que todas as células e moléculas imunológicas são capazes de reconhecer umas às outras [38]. Portanto, não é necessário estimulação externa para que haja uma dinâmica em progresso no sistema imunológico.

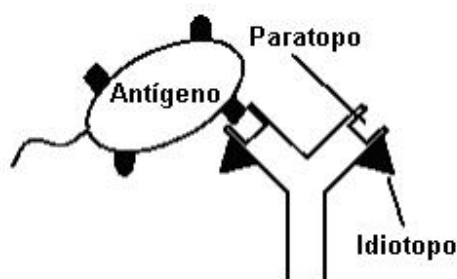


Figura IV.1.3: Modelo simplificado da rede imunológica.

Ilustração da capacidade dos anticorpos, receptores dos linfócitos B, reconhecerem os epítopos dos antígenos e ter seus idiotopos reconhecidos pelos paratopos de outros anticorpos. Adaptado de de Castro [18].

Jerne propôs que os receptores de uma célula imunológica possuem padrões moleculares que permitem a esta célula reconhecer e ser reconhecida por outras células do sistema imunológico, como ilustrado na Figura IV.1.3. A região que pode ser reconhecida em um antígeno é chamada epítipo. Em um anticorpo, o paratopo é capaz de reconhecer os epítopos e os paratopos presentes nos anticorpos, chamados idiotopos. Esta visão sugere um sistema imunológico dinâmico com comportamento governado pelos padrões de interação das células e moléculas imunológicas, e cujo estado muda de acordo com as perturbações causadas por antígenos. Como neste modelo as células imunológicas são monoespecíficas, o reconhecimento de uma outra célula imunológica pode ser visto como correspondendo ao reconhecimento de seu receptor. Quando o

receptor de uma célula reconhece o receptor de outra célula, a célula que reconheceu é estimulada e a reconhecida é suprimida. O efeito global dos estímulos e supressões varia de um modelo para outro. No modelo proposto por Varela e Coutinho [75], três conceitos importantes são salientados: estrutura, dinâmica e metadinâmica.

A estrutura da rede imunológica é o padrão de conectividade entre cada célula do sistema imunológico e todas as outras, isto é, como os componentes estão ligados. As redes imunológicas não apresentam estruturas definidas *a priori*, mas tendem a se adaptar ao espaço de formas determinado pelos antígenos aos quais é exposta. A dinâmica da rede imunológica corresponde a como a rede de células, moléculas e conexões varia no tempo. Esta é uma das principais características que permitem que as redes imunológicas se adaptem ao meio como consequência das interações dos componentes da rede com o ambiente. Por fim, a metadinâmica da rede pode ser definida como a constante geração de novas células imunológicas e morte de células não estimuladas. As células novas substituem as que não fazem parte da resposta imunológica. Portanto, a metadinâmica do sistema imunológico está relacionada à sua capacidade de ter sua estrutura variável no tempo.

Em geral, a estrutura da rede imunológica pode ser representada como mostrado na equação 4.1:

$$TVP = \Xi - \Psi + \Phi \quad (4.1)$$

onde TVP representa a taxa de variação da população, Ξ representa a entrada de novas células, Ψ representa a morte das células não estimuladas e Φ representa a reprodução das células estimuladas.

IV.1.v) aiNet

Há inúmeras implementações de sucesso de SIAs, vide [17]. Em um SIA, um problema com solução desconhecida é tratado como antígeno, enquanto que potenciais soluções são modeladas como anticorpos. Particularmente, o algoritmo imunológico aiNet [18], do inglês “*artificial immune network*”, foi desenvolvido para resolver problemas de agrupamento de dados e apresenta as seguintes propriedades:

- Tamanho da população dinamicamente ajustável;
- Manutenção de diversidade;

- Auto-organização;
- Distributividade;
- Definição automática do número de protótipos;
- Não assume nenhuma forma específica para o agrupamento.

Essa ferramenta foi aplicada em um método de síntese sonora [9] fazendo uso das características descritas acima. Os sons produzidos pelo método foram chamados sons imunológicos. Basicamente, foram gerados sons-anticorpo para reconhecer um conjunto fixo e predefinido de sons-antígeno, produzindo assim variantes com as características desejadas. O algoritmo aiNet forneceu manutenção de diversidade e um número adaptável de sons-anticorpo resultantes (células de memória), de modo que o resultado estético pretendido foi conseguido corretamente, evitando a definição formal das qualidades sonoras. A principal funcionalidade aqui é a auto-organização. Os mecanismos de realimentação positiva, realimentação negativa e interação local são responsáveis pela posição final relativa dos sons-anticorpo (agentes) em relação aos sons-antígeno (atratores).

Nas próximas seções, serão explicados o algoritmo e todos os parâmetros do programa, internos e de entrada. Então, uma breve análise dos parâmetros será apresentada.

IV.1.vi) Algoritmo da aiNet

O algoritmo original [18], utilizado neste trabalho, é apresentado a seguir. Alguns comentários a respeito de etapas do algoritmo seguem para esclarecer o funcionamento do método e elucidar etapas cruciais.

O algoritmo de aprendizado da aiNet se destina a construir um conjunto de células de memória que reconhece e representa a organização estrutural dos dados de entrada. Quanto mais específicos forem os anticorpos, menos parcimoniosa será a rede (baixa taxa de compressão de dados), enquanto que quanto menos específicos forem os anticorpos, mais parcimoniosa será a rede com relação ao número final de anticorpos (taxa de compressão de dados melhorada). Isto implica em anticorpos mapeando uma maior região do espaço. Como será visto adiante, o limiar de supressão controla o nível de especificidade dos anticorpos, a precisão de agrupamento e a plasticidade da rede.

1. A cada iteração faça:
2. Para cada antígeno faça:
3. Determine sua afinidade usando a métrica euclidiana,
4. Selecione um subconjunto composto dos n anticorpos com maior afinidade,
5. Clone-os proporcionalmente à afinidade, quanto maior a afinidade, maior o número de clones, criando o conjunto C de clones,
6. Aplique o processo de maturação de afinidade como segue:

$$C^* = C + \alpha(Ag - C)$$
, onde α é inversamente proporcional à afinidade antigênica,
7. Determine a nova afinidade entre C^* e os antígenos usando a métrica euclidiana,
8. Selecione $\zeta\%$ dos anticorpos com maior afinidade em C^* e inclua-os na matriz de memória M ,
9. *Apoptose*: Elimine da matriz de memória todos os clones cuja afinidade antigênica for menor que um limiar σ_d ,
10. Determine a afinidade entre as células de memória,
11. *Supressão Clonal*: Elimine todos os clones cuja afinidade com as outras células de memória for maior que um limiar σ_s ,
12. Concatene a matriz total de memória com a matriz clonal resultante,
13. Determine a afinidade entre todos os anticorpos de memória,
14. *Supressão da Rede*: elimine todos os anticorpos da matriz de memória cuja afinidade seja maior que o limiar σ_s ,
15. Construa a nova matriz de anticorpos,
16. Teste o critério de parada.

Algoritmo 1: Algoritmo simplificado da aiNet.

Alguns passos são fundamentais para o entendimento do funcionamento do método. Os passos 5 e 6 correspondem ao princípio da seleção clonal e maturação de afinidade. O processo de apoptose correspondente ao passo 9 garante que um nível mínimo de desempenho para os elementos da rede imunológica artificial. Existem dois processos de supressão importantes, respectivamente a supressão clonal e a supressão da rede. A supressão clonal elimina os clones que não mapeiam regiões diferentes do

espaço que aquelas já representadas na matriz de memória. A supressão da rede tem a mesma finalidade, em uma etapa diferente do processo, afetando a plasticidade final da rede.

IV.1.vii) Parâmetros

Os parâmetros do programa ajustáveis pelo usuário são apresentados a seguir na Tabela IV.1.1. Além desses parâmetros, existem alguns parâmetros internos que não são manipuláveis. Estes serão apresentados brevemente a seguir apenas para um completo entendimento do funcionamento do método.

Tabela IV.1.1: Parâmetros de entrada da aiNet.

Parâmetro	Descrição
L	Número de amostras por indivíduo
SF	Taxa de amostragem
D	Número de antígenos
N	Número inicial de anticorpos
σ_s	Limiar de supressão
n	Número de células com maior afinidade tomadas para cada antígeno
CM	Número multiplicador de clone
gen	Número máximo de gerações
ζ	Porcentagem de clones para serem selecionados
σ_d	Distância mínima entre anticorpo e antígeno

- N_c : Número de clones gerados;
- M_i : taxa de aprendizado (hipermutação). Valor padrão: 4.0;
- nR : Número de anticorpos re-selecionados.

Os parâmetros número de amostras L e taxa de amostragem SF são específicos da aplicação em síntese sonora e definem a duração de cada som. Os parâmetros restantes são válidos para qualquer aplicação. O número de sons-antígeno (sons atratores) interfere diretamente no resultado final, pois esta é a caracterização da região do espaço escolhida. É importante perceber que, dentre todos os parâmetros, dois têm papel fundamental na controlabilidade do método, a saber: o limiar de supressão σ_s e a distância mínima entre som-anticorpo e som-antígeno σ_d . A distância mínima controla a proximidade entre os sons-anticorpo (sons agentes) e os sons-antígeno (sons atratores), isto é, o quão próximos no espaço de busca o resultado final estará dos sons que representa. O limiar de supressão é responsável pela plasticidade e tamanho da rede

final, isto é, o número de sons-anticorpo (ou células de memória, os sons agentes) e o tamanho da região que representam.

IV.2 Modelo e Objetivos

Os grupos resultantes da aiNet representam a imagem interna responsável pelo mapeamento de grupos no espaço original (mostrado na Figura IV.2.1 a) em grupos na rede de células de memória (sons-anticorpo) resultante (Figura IV.2.1 b). As células de memória tendem a representar características comuns presentes nos dados de um mesmo grupo. Imagine-se um conjunto de sons como antígenos e sua imagem interna, como suas variantes [64]. Pode-se interpretar a configuração final da rede como uma espécie de transformação imunológica aplicada aos sons-antígeno que resulta em variações desses com grau adaptável de complementaridade, dependendo de um número reduzido de parâmetros, resultando em um cruzamento de características dos antígenos, uma vez que um som-anticorpo que não seja específico demais pode reconhecer mais de um som-antígeno, representando, assim, características advindas de múltiplas amostras. É extremamente importante notar que, quando um som-anticorpo está representando mais de um som-antígeno, esse som-anticorpo se encontra em um ponto intermediário do espaço, apresentando características advindas de mais de um antígeno sendo mapeado por ele. Tal fenômeno leva à intersecção de características sonoras dos elementos, efeito representado na Figura IV.2.1 c) e de crucial importância no contexto deste trabalho. É importante ressaltar que tais qualidades podem ser advindas de mais de um parâmetro musical, isto é, pode-se identificar misturas de frequências, de qualidades timbrais, de intensidades, ou até mesmo de mais de uma ao mesmo tempo. Desta maneira, as formas de onda que fazem o papel de sons-antígeno podem ser interpretadas como o repertório ao qual o sistema está exposto e as qualidades sonoras associadas podem ser herdadas pela resposta gerada.

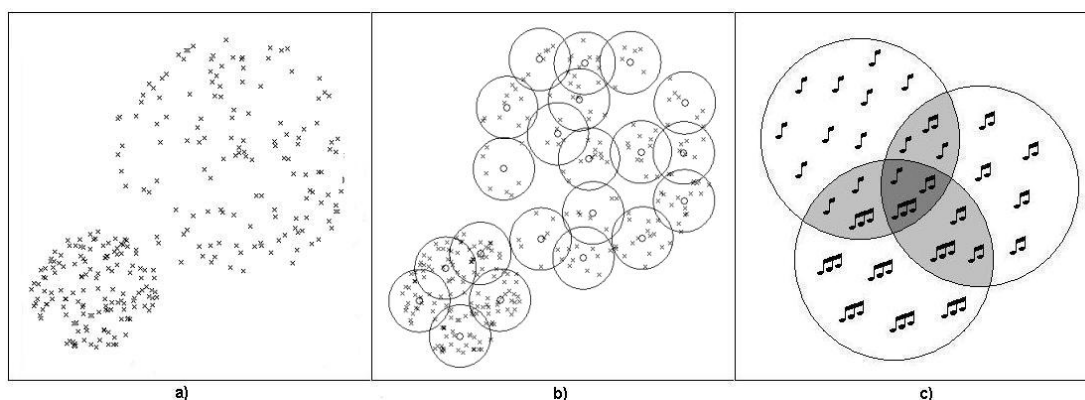


Figura IV.2.1: Representação pictórica da capacidade de extração de características (quantização vetorial) da aiNet.

A parte a) mostra os dados originais (que podem ser interpretados como sons-antígeno), a parte b) mostra os protótipos (células de memória) resultantes responsáveis pela representação dos dados originais. A parte c) ilustra qualidades sonoras que três classes de tons possuem em comum.

No modelo da aiNet os sons atratores são chamados de sons-antígeno e os sons agentes são os sons-anticorpo, também chamados de células de memória. O método consiste em apresentar um grupo de sons-antígeno aos sons-anticorpo, que têm a tarefa de reconhecê-los e mapeá-los, isto é, se posicionar em regiões do espaço próximas a um ou mais sons-antígeno (sons atratores). O experimento realizado, apresentado a seguir, consiste da apresentação dos sons que foram definidos na seção III.3.iv) como sons atratores (sons-antígeno) à rede imunológica aiNet, que retornará um conjunto de sons-anticorpo (sons agentes) que representam os sons-antígeno aos quais foram expostos. O propósito do experimento é mostrar a eficácia do método em produzir resultados que representam variações dos sons atratores (sons-antígeno) e, possivelmente, apresentam características advindas de mais de um desses sons atratores por representarem um ponto intermediário no espaço entre dois ou mais desses sons.

IV.3 Experimentos e Resultados

Nesta seção, o experimento realizado será apresentado e os resultados (formas de onda e espectros dinâmicos) serão mostrados, acompanhados das distâncias de acordo com as métricas matemáticas euclidiana (temporal) e tempo-frequencial, além de uma análise subjetiva de similaridade perceptiva similar à apresentada para caracterizar os atratores. Será feita uma análise comparativa levando-se em conta os critérios de

flexibilidade, controle, adaptabilidade e simplicidade, adotados como relevantes em um método de síntese sonora. Finalmente algumas conseqüências da representação vetorial adotada, juntamente com o aspecto temporal e a abordagem populacional serão explorados.

A Tabela IV.3.1 mostra os valores dos parâmetros utilizados no experimento. É importante destacar que os sons-anticorpo foram inicializados como ruído branco. A taxa de amostragem SF é 44100 para que todas as freqüências audíveis (até 20 kHz) possam ser representadas e não se perca informação espectral. O número de amostras L foi definido como 4096 para que, com o valor de SF , a duração dos sons fosse aproximadamente 0,1 segundo, suficiente para que os resultados pudessem ser avaliados perceptivamente. No entanto, é importante salientar que L é a dimensão do espaço vetorial de busca, portanto, definida a taxa de amostragem, quanto maior o número de amostras L , maior a dimensionalidade do espaço de busca e, conseqüentemente, maior a dificuldade de se achar boas soluções. Os parâmetros considerados importantes para um método de síntese sonora, o limiar de supressão σ_s e a distância mínima entre som-anticorpo e som-antígeno σ_d tiveram os valores apresentados escolhidos para que o número final de células de memória (sons-anticorpo) da rede fosse menor que o número de sons-antígeno apresentados. Conseqüentemente, espera-se que os sons-anticorpo resultantes se posicionem em pontos intermediários do espaço e apresentem características advindas de mais de um som-antígeno. Por fim, estudou-se a variação dos parâmetros restantes e concluiu-se que esses não afetam significativamente os resultados.

Tabela IV.3.1: Valores dos parâmetros da aiNet utilizados no experimento.

L	SF	D	N	σ_s	n	CM	gen	ζ	σ_d
4096	44100	7	10	0.5	1	7	500	70%	0.001

A seguir, são apresentadas tanto as formas de onda quanto os espectros dinâmicos dos 5 sons resultantes, isto é, as células de memória (ou sons-anticorpo).

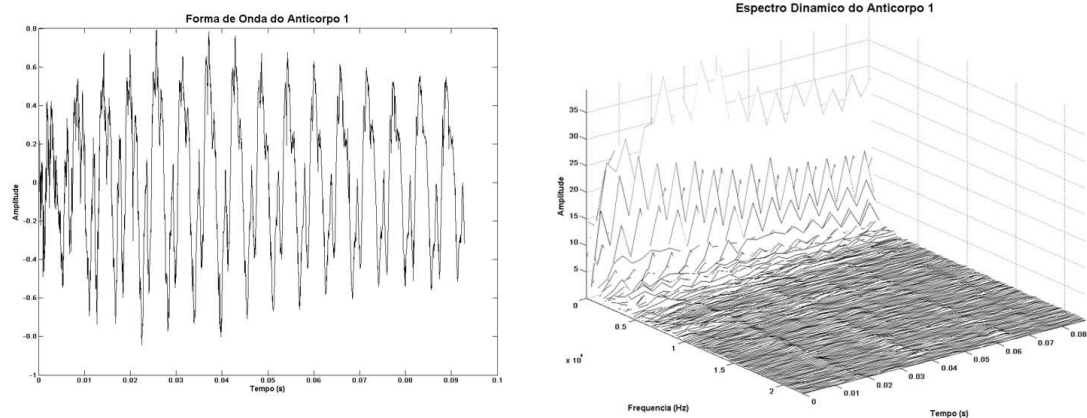


Figura IV.3.1: Forma de onda e Espectro Dinâmico do Anticorpo 1

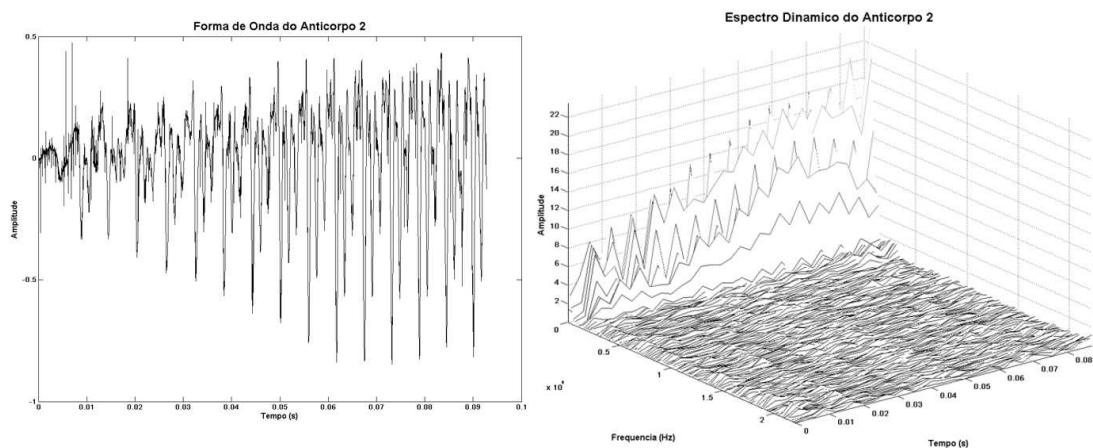


Figura IV.3.2: Forma de onda e Espectro Dinâmico do Anticorpo 2

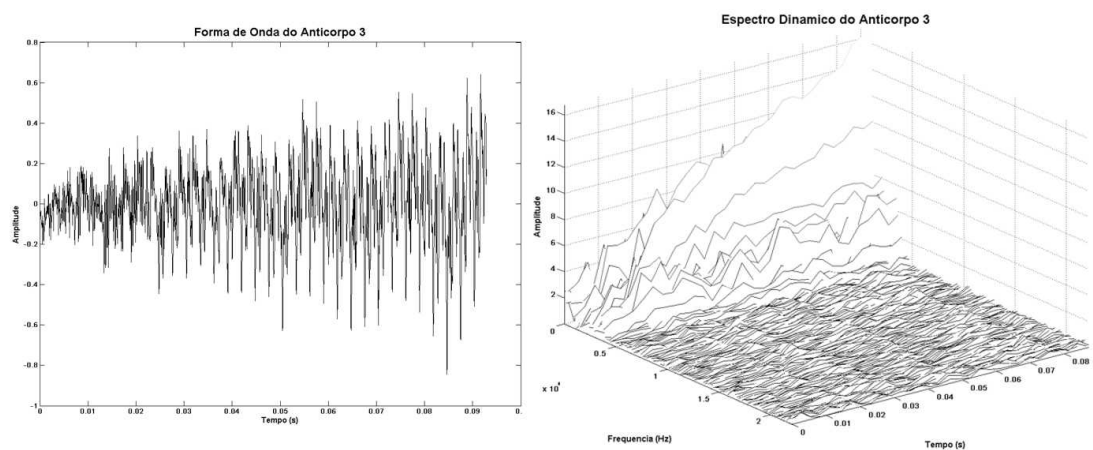


Figura IV.3.3: Forma de onda e Espectro Dinâmico do Anticorpo 3

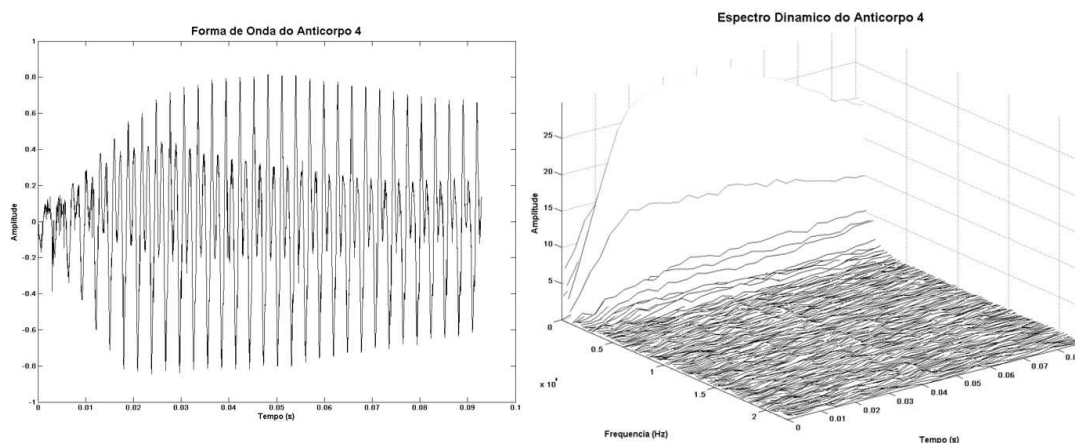


Figura IV.3.4: Forma de onda e Espectro Dinâmico do Anticorpo 4

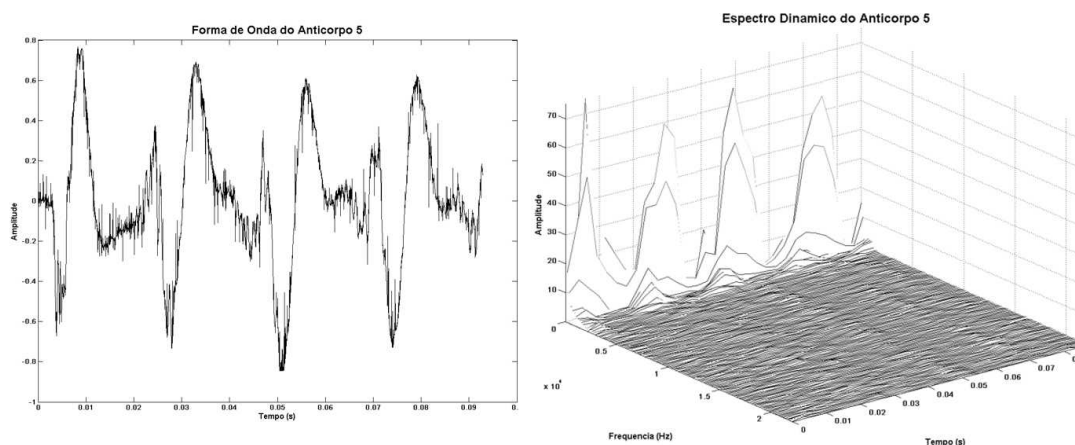


Figura IV.3.5: Forma de onda e Espectro Dinâmico do Anticorpo 5

Primeiramente, é essencial perceber que todos os espectros dinâmicos apresentam variações individuais dos envelopes dos parciais ao longo do tempo, como desejado em sons complexos para aplicações musicais. Uma breve comparação visual com os sons atratores apresentados na seção III.3.iv) já transparece a grande similaridade visual existente entre a forma de onda e o espectro dinâmico de cada um dos resultados (sons-anticorpo) e pelo menos um dos sons atratores (sons-antígeno). A Tabela IV.3.2 apresenta a resultado do cruzamento visual dos sons atratores mostrados da Figura III.3.3 à Figura III.3.9 com os resultados apresentados da Figura IV.3.1 à Figura IV.3.5. Um X indica que não há similaridade.

Os sons obtidos estão bastante ruidosos e foram filtrados com um filtro passa-baixa com frequência de corte em 5 kHz para avaliação perceptiva. A avaliação visual das formas de onda e dos espectros dinâmicos não fica tão prejudicada pela presença de ruído, bem como as medidas de distância temporal e espectro temporal. No entanto, para os propósitos de aplicações musicais destacados neste trabalho, a avaliação perceptiva é fundamental e foi decisiva na decisão de filtrar os sons resultantes.

Tabela IV.3.2: Agentes mais próximos de cada atrator de acordo com o cruzamento visual. X indica não similaridade.

Atrator	1	2	3	4	5	6	7
Agente relacionado	4	5	1	X	3	2	X

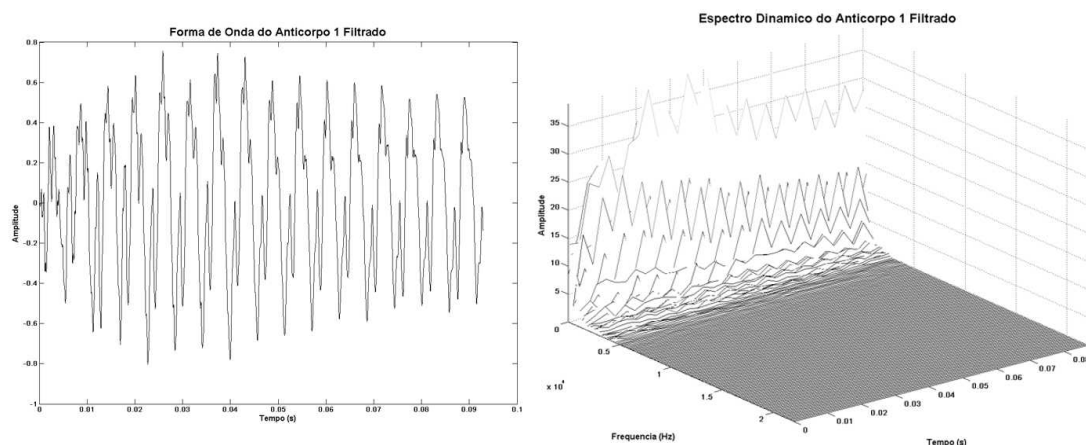


Figura IV.3.6: Forma de onda e Espectro Dinâmico do Anticorpo 1 depois de filtrado.

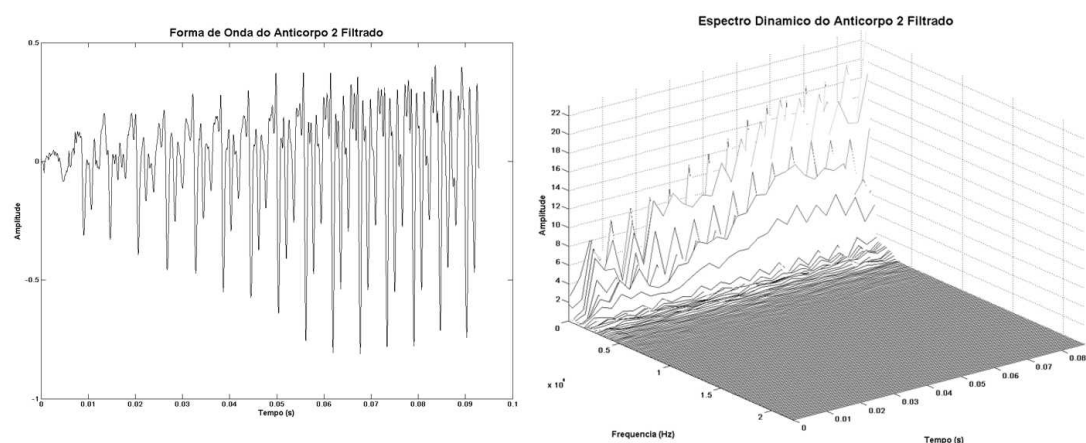


Figura IV.3.7: Forma de onda e Espectro Dinâmico do Anticorpo 2 depois de filtrado.

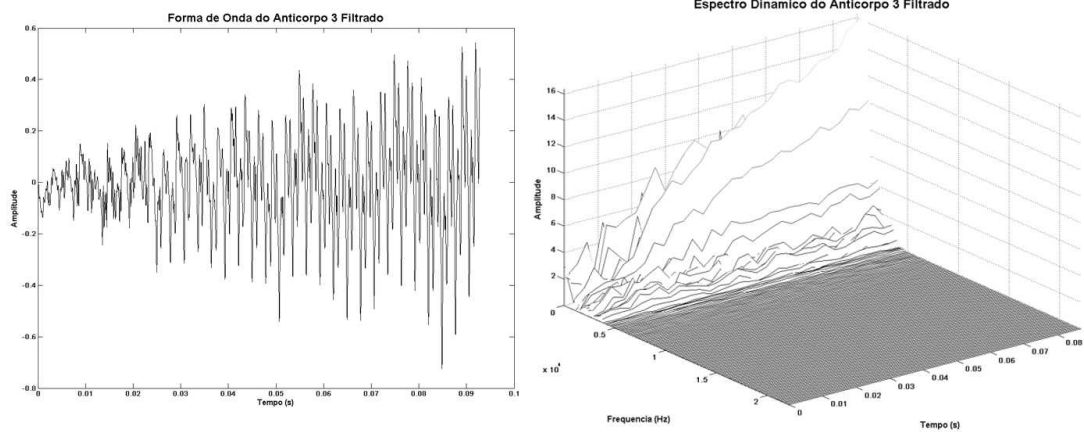


Figura IV.3.8: Forma de onda e Espectro Dinâmico do Anticorpo 3 depois de filtrado.

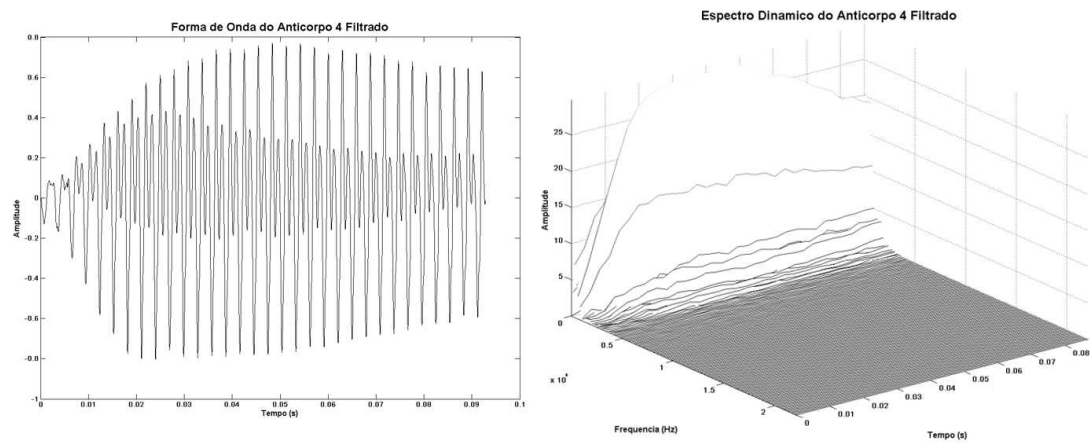


Figura IV.3.9: Forma de onda e Espectro Dinâmico do Anticorpo 4 depois de filtrado.

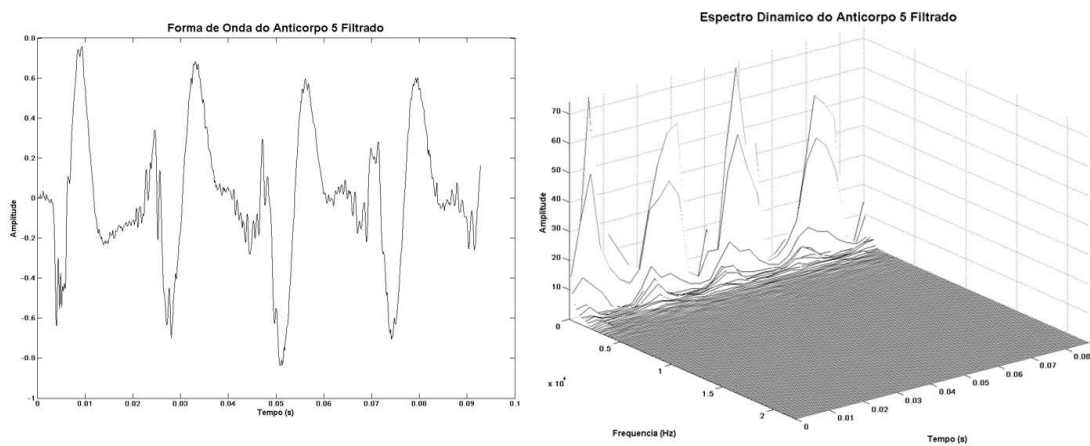


Figura IV.3.10: Forma de onda e Espectro Dinâmico do Anticorpo 5 depois de filtrado.

Pode-se destacar que, após a filtragem, tanto as formas de onda quanto os espectros dinâmicos dos resultados (sons-anticorpo) são tanto visualmente quanto perceptivamente menos ruidosos que os obtidos originalmente.

A seguir, serão apresentadas as distâncias de acordo com as métricas matemáticas euclidiana (temporal), na Tabela IV.3.3, tempo-freqüencial, na Tabela IV.3.5, e a avaliação subjetiva de cinco indivíduos das similaridades perceptivas entre os sons-agente (sons-anticorpo ou células de memória) e os sons atratores (sons-antígeno) na Tabela IV.3.7. Destaca-se, novamente, que os 5 indivíduos não têm treinamento musical, no entanto, possuem treinamento científico suficiente para compreender a tarefa que lhes foi solicitada, isto é, a avaliação subjetiva da similaridade entre pares de sons em uma escala de 1 a 5.

As distâncias entre os sons-anticorpo (sons agentes) e os sons-antígeno (sons atratores) calculadas pela métrica euclidiana são apresentadas na Tabela IV.3.3. A distância mínima ao longo das linhas da Tabela IV.3.3 e da Tabela IV.3.5 estão marcadas em negrito. Estas distâncias, quando lidas ao longo das colunas da tabela, indicam qual som-antígeno (som atrator) está sendo representado por qual som-anticorpo (som agente).

Tabela IV.3.3: Distâncias entre os antígenos e os anticorpos calculadas pela métrica Euclidiana.

	Atrator 1	Atrator 2	Atrator 3	Atrator 4	Atrator 5	Atrator 6	Atrator 7
Agente 1	35.3156	32.6502	11.7453	33.0380	25.3689	25.8602	33.8414
Agente 2	26.1344	28.2853	28.5805	28.4176	18.8594	12.4926	28.9943
Agente 3	29.6642	26.7901	29.4714	26.1620	14.6724	20.6806	27.3606
Agente 4	18.1232	34.1210	34.0686	33.0494	25.0156	25.4369	33.5635
Agente 5	34.4504	6.2457	32.7158	32.1525	25.0173	26.2750	32.5658

A Tabela IV.3.4 apresenta as distâncias mínimas individualmente, marcadas em negrito na Tabela IV.3.3. Como o número de sons-anticorpo (agentes) resultantes é menor que o número de sons-antígeno, há sons-antígeno (atratores) que não apresentam um som-anticorpo (agente) com distância mínima. Estes estão marcados com um X na Tabela IV.3.4, na Tabela IV.3.6 e na Tabela IV.3.8.

Tabela IV.3.4: Agentes mais próximos de cada atrator de acordo com a distância Euclidiana.

Atrator	1	2	3	4	5	6	7
Agente relacionado	4	5	1	X	3	2	X

As distâncias entre os sons-antígeno (atratores) e os sons-antígeno (atratores), calculadas pela métrica tempo-freqüencial, são apresentadas na tabela 3.5.

Tabela IV.3.5: Distâncias entre os antígenos e os anticorpos calculadas pela métrica tempo-freqüencial.

	Atrator 1	Atrator 2	Atrator 3	Atrator 4	Atrator 5	Atrator 6	Atrator 7
Agente 1	169.6819	182.1929	53.5078	166.3705	129.7763	131.8675	168.5284
Agente 2	120.0205	158.1027	140.3136	136.5618	89.2327	60.6355	138.9397
Agente 3	135.7996	150.5621	146.6392	126.2434	69.4460	101.0775	130.4524
Agente 4	82.9139	178.9715	164.2600	153.7620	114.6551	118.5242	154.7692
Agente 5	173.2680	33.9078	174.9691	169.5481	135.8244	143.5056	169.8316

A Tabela IV.3.6 apresenta as distâncias mínimas individualmente, marcadas em negrito na Tabela IV.3.5.

Tabela IV.3.6: Agentes mais próximos de cada atrator de acordo com a distância tempo-freqüencial.

Atrator	1	2	3	4	5	6	7
Agente relacionado	4	5	1	X	3	2	X

As similaridades entre os sons-antígeno (atratores) e os sons-antígeno (atratores), estimadas pelos 5 indivíduos são apresentadas na Tabela IV.3.7. A tabela apresenta a média e o desvio padrão das aferições. É importante ressaltar que, devido à natureza estatística e subjetiva desta medida de similaridade, diferentemente das métricas matemáticas, nesta tabela é possível destacar mais de um som-antígeno (som atrator) para cada som-anticorpo (som agente). Isto corresponde perceptivamente à noção de que um som-anticorpo (som agente) resultante do processo de síntese possui características advindas de mais de um som-antígeno (som atrator).

Tabela IV.3.7: Similaridades subjetivas entre os atratores estimadas pelos indivíduos.

	Atrator 1	Atrator 2	Atrator 3	Atrator 4	Atrator 5	Atrator 6	Atrator 7
Agente 1	3,2±1,09	2,0±0,00	4,2±0,84	1,8±0,45	1,6±0,55	2,4±0,55	1,6±0,89
Agente 2	1,6±0,89	1,4±0,55	1,2±0,45	1,0±0,00	1,2±0,45	4,8±0,45	1,0±0,00
Agente 3	3,2±0,84	1,0±0,00	1,8±0,45	2,2±1,30	3,8±0,45	2,0±0,71	2,0±0,00
Agente 4	4,6±0,55	1,4±0,55	2,0±0,70	2,0±1,0	2,2±0,45	1,8±1,30	2,0±0,00
Agente 5	1,2±0,45	4,8±0,45	1,8±0,84	1,2±0,45	1,2±0,45	1,2±0,45	1,0±0,00

A Tabela IV.3.8 apresenta as similaridades máximas individualmente, marcadas em negrito na Tabela IV.3.7.

Tabela IV.3.8: Agentes mais próximos de cada atrator de acordo com a avaliação perceptiva dos indivíduos.

Atrator	1	2	3	4	5	6	7
Agente relacionado	1,3,4	5	1	X	3	2	X

IV.4 Discussão

Os resultados destas tabelas são de mais fácil interpretação quando comparados com a interpretação cruzada da caracterização dos sons atratores na seção III.3.iv). Primeiramente, deve-se notar que, diferentemente da análise anterior, agora busca-se a similaridade entre os sons atratores e os sons agentes. Portanto, basta analisar a Tabela IV.3.4, a Tabela IV.3.6 e a Tabela IV.3.8. Estas tabelas apresentam o resultado, tanto no domínio temporal quanto tempo-freqüencial, a respeito de qual som-anticorpo (som agente) se encontra mais próximo de qual som-antígeno (som atrator). Este mesmo resultado, mostrado na Tabela IV.3.2 e na Tabela IV.3.3, foi atingido através da análise visual e cruzamento da Figura III.3.3 à Figura III.3.9, onde os atratores são apresentados, e da Figura IV.3.6 à Figura IV.3.10, onde os resultados deste experimento são apresentados.

A Tabela IV.3.4 e a Tabela IV.3.6 mostram qual som-anticorpo (som agente) se aproximou de qual som-antígeno (som atrator), respectivamente, nos domínios temporal e tempo-freqüencial. Deve-se notar que os sons-anticorpo correspondem a variantes dos sons-antígeno que estão representando porque a distância entre eles não é zero, ou seja,

são formas de onda diferentes. Por fim, é importante ressaltar que ambas as métricas resultaram na mesma análise, ou seja, a Tabela IV.3.4 e a Tabela IV.3.6 são iguais.

A Tabela IV.3.7 contém os resultados da Tabela IV.3.3 e da Tabela IV.3.5, sugerindo que, perceptivamente, outras relações de similaridade podem ser estabelecidas. Analisando-se a Tabela IV.3.8 e cruzando-se seu resultado com a Tabela IV.3.4 e a Tabela IV.3.6, pode-se inferir que ambas as métricas são bastante capazes de capturar a essência perceptiva dos resultados quando se trata de variantes dos sons-antígeno (sons atratores). Portanto, é importante notar que nenhuma das métricas forneceu resultados discrepantes com o perceptivo. Deve-se chamar a atenção também para o fato de que, na Tabela IV.3.3, quando um som-anticorpo (som agente) apresenta similaridade perceptiva com mais de um som-antígeno (som atrator), isto significa que o som-anticorpo está representando todos esses sons-antígeno. Isto é, este som-anticorpo apresenta características cruzadas dos sons-antígeno que está representando. Estas características podem ser devidas a similaridades perceptivas de altura, intensidade, timbre ou combinações das mesmas. Também é importante notar que estes cruzamentos são muito difíceis de detectar visualmente. Uma nova análise da Tabela IV.3.7, agora sob uma perspectiva diferente, permite avaliar se, quando um som-antígeno representa mais de um som-anticorpo, isso é consistente com o agrupamento dos sons atratores (sons-antígeno) realizado na seção III.3.iv). Os sons atratores próximos no espaço vetorial de busca original (métrica de distância euclidiana) eram os sons atratores 1 e 5, 2 e 3 e 4 e 7. Portanto, dever-se-ia esperar que um som-anticorpo que representasse mais de um som-antígeno refletisse essa estrutura dos grupos. Uma análise da proximidade dos resultados de acordo com a distância euclidiana mostrada na Tabela IV.3.7 não reflete necessariamente essa estrutura, sugerindo outras relações. No entanto, como foi verificado anteriormente na seção III.3.iv), a distância euclidiana não reflete a similaridade perceptiva linearmente. Isto é, apenas alguns aspectos da similaridade perceptiva são refletidos na proximidade no espaço euclidiano.

Em relação às características de um método de síntese consideradas relevantes, pode-se afirmar que o método apresenta uma controlabilidade alta. Isto é, os resultados apresentam características presentes nos atratores, tornando bastante fácil prever qual será o resultado em termos perceptivos da variação dos parâmetros do programa. Em especial, é possível obter resultados diferentes através da alteração de apenas um parâmetro, o limiar de supressão σ_s . Caetano *et al.* [9] mostram que é possível controlar a

especificidade dos resultados (o número resultante de anticorpos) através apenas da alteração do limiar de supressão. O método apresenta uma flexibilidade alta. Basta alterar os atratores que suas características são passadas aos resultados. Deve-se notar que uma característica fundamentalmente vantajosa é a adaptabilidade intrínseca ao método. A auto-organização faz com que os resultados apresentem variações cada vez que o método é utilizado. Também é possível atingir variações quando mais de um som-anticorpo (agente) representa o mesmo som-antígeno (atrator).

Por fim, deve-se chamar a atenção para a simplicidade relativa do método proposto. O algoritmo é complexo e de difícil implementação, mas de fácil utilização, bastando especificar certos parâmetros. Uma vez implementado, o algoritmo apresenta muitas vantagens em relação aos métodos tradicionais, como robustez, manutenção de diversidade, estratégia populacional que permite que mais de um som seja obtido em uma única rodada, capacidade de sintetizar qualquer tipo de som, harmônico ou inarmônico, sem nenhum esforço adicional devido à estratégia de busca vetorial no domínio do tempo. Isto é, o resultado final já é a seqüência de amostras sonoras que pode ser convertida em som sem necessidade de processamento, a menos eventualmente de filtragem dependendo do valor dos parâmetros. Evita-se, dessa maneira, a parametrização das variáveis musicais, comum em grande parte dos métodos tradicionais de síntese sonora, que leva, invariavelmente, a uma simplificação excessiva do modelo do tom musical. Por outro lado, uma descrição exaustiva das qualidades sonoras, como o modelo de síntese granular de Roads [66], permite a obtenção de uma vasta gama de sons, sem, entretanto, apontar como se pode obter um resultado perceptivo específico desejado. Isto é, a proposta desse trabalho permite que se obtenha um resultado desejado em termos perceptivos através da escolha apropriada dos atratores e ajuste dos parâmetros.

Ainda assim, verificou-se a necessidade de simplificação, especialmente no que tange o número de parâmetros, alguns pouco intuitivos. Pensa-se que muitos dos parâmetros de entrada poderiam ser parâmetros internos do programa. Há claramente a necessidade de estudos adicionais voltados à simplificação do método, tal como desenvolver uma interface gráfica que possibilite ao usuário ter contato mais intuitivo com as variáveis relevantes e desenvolver suas idéias de forma criativa sem, no entanto, a necessidade de um conhecimento profundo de acústica, psicoacústica ou técnicas de programação ou qualquer outra ferramenta de engenharia.

V) SÍNTESE SONORA POR REDES NEURAIS ARTIFICIAIS

Este capítulo apresenta a implementação das idéias desenvolvidas no texto através da utilização de um algoritmo de redes neurais artificiais aplicado à resolução do problema do caixeiro viajante (ver seção III.3.i) e seção V.2.viii) em síntese sonora digital. Primeiramente serão apresentados alguns conceitos de redes neurais e aprendizado. Em seguida, um modelo simplificado do funcionamento do algoritmo do Mapa Auto-Organizável (MAO) de Kohonen é apresentado, e sua aplicação em síntese sonora é formalizada como uma proposta de resolução do problema do caixeiro viajante no espaço vetorial sonoro. Por fim, os resultados do experimento realizado são apresentados e discutidos sob a luz dos conceitos apresentados anteriormente e do critério de avaliação de um método de síntese sonora, especialmente no que tange as potenciais aplicações musicais.

V.1 Introdução

A pesquisa em redes neurais artificiais (RNAs) foi motivada desde seu início pelo reconhecimento de que o cérebro humano processa informação de uma maneira totalmente diferente do computador digital convencional. O cérebro é um computador complexo, não-linear e paralelo [34]. Ele possui a capacidade de organizar seus elementos estruturais, chamados neurônios, de maneira a realizar tarefas complexas, como reconhecimento de padrões, percepção e controle motor, muitas vezes mais rápido e preciso que o computador mais rápido disponível hoje [34]. Considere-se como exemplo a visão humana, uma tarefa de processamento de informação. O sistema visual é um sistema cognitivo. É função do sistema visual fornecer uma representação do meio à nossa volta e, mais importante, fornecer a informação necessária para possibilitar interação com o meio ambiente [34]. Mais especificamente, o cérebro realiza rotineiramente tarefas de reconhecimento perceptivo – por exemplo, reconhecer um rosto familiar em um ambiente desconhecido – com facilidade, enquanto computadores convencionais são menos eficazes em tarefas menos complexas.

Um outro exemplo é o sonar de um morcego. O sonar é um sistema ativo de eco-localização [15]. Além de fornecer informação sobre a distância do alvo, o sonar dos morcegos também processa informação de velocidade relativa, tamanho, azimuth e elevação do alvo [34]. Os cálculos complexos necessários para extrair toda essa informação do eco do alvo ocorrem em um cérebro do tamanho de uma ameixa. De fato, um morcego eco-localizador pode perseguir e capturar seu alvo com uma taxa de acerto longe de ser atingida por dispositivos computacionais. Como o cérebro de um morcego é capaz de fazer isso? No nascimento, um cérebro tem uma estrutura flexível e a capacidade de construir suas próprias regras através do que, em geral, refere-se como experiência. De fato, experiência é construída com o tempo, com a maior parte do aprendizado do cérebro humano se dando durante os dois primeiros anos desde o nascimento; mas o desenvolvimento continua até bem depois desse estágio. Plasticidade permite ao sistema nervoso em desenvolvimento adaptar-se ao meio ambiente que o cerca, isto é, criar um modelo do meio. Assim como plasticidade parece ser essencial para o funcionamento dos neurônios como unidades de processamento de informação no cérebro humano, o mesmo se aplica às redes neurais artificiais e seus neurônios artificiais. Em termos gerais, uma rede neural artificial é um modelo da maneira como o cérebro realiza certas tarefas [34]. Este modelo é implementado com circuitos eletrônicos ou simulado em computador.

Este capítulo se destina à apresentação da implementação do método de síntese sonora utilizando o mapa auto-organizável unidimensional de Kohonen (MAO), um dos paradigmas de redes neurais artificiais. Uma característica importante das redes neurais é a chamada capacidade de aprendizado, isto é, adaptação de sua estrutura ao ambiente. Os MAOs utilizam uma estratégia de aprendizado não-supervisionada. O propósito do aprendizado não supervisionado é descobrir padrões e características importantes nos dados de entrada (informações do ambiente) sem o auxílio de um mentor. Isto é, quando estimulados com um conjunto de sons atratores, os neurônios do mapa se acomodam em posições no espaço vetorial de modo a representar estes sons de entrada e caracterizar a região do espaço onde tais sons se encontram. O MAO é, portanto, caracterizado pela formação de um mapeamento topológico do ambiente (espaço de entrada) no qual as localizações espaciais dos neurônios correspondem a características intrínsecas dos sons de entrada [34]. Portanto, os sons resultantes do método serão posicionados em regiões do espaço que tendem a maximizar as similaridades presentes em um certo conjunto de

sons, fazendo com que os resultados tendam a apresentar características advindas de mais de um dos sons de entrada.

Para demonstrar todo o potencial do método aqui proposto, primeiramente serão apresentados alguns conceitos a respeito do MAO e um modelo simplificado de seu funcionamento. Em seguida, o algoritmo utilizado, inspirado na capacidade de aprendizado e adaptação das redes neurais biológicas, é apresentado e sua aplicação em síntese sonora é formalizada. Por fim, os resultados do experimento realizado são apresentados e discutidos sob a luz dos conceitos apresentados anteriormente e do critério de avaliação proposto de um método de síntese sonora.

V.2 Mapa Auto-Organizável de Kohonen

A rede neural de Kohonen é uma grade de neurônios que constitui um sistema auto-organizável, também conhecida como mapa auto-organizável. Ela foi inicialmente proposta por Kohonen [40], [39], [41], [42] e é baseada em observações do comportamento cerebral, que reage a estímulos correlatos em regiões próximas no cérebro. A Figura V.2.1 ilustra esta propriedade da rede neural biológica, mostrando que neurônios próximos são ativados quando realizamos cada uma das tarefas descritas.

O principal objetivo do mapa auto-organizável é transformar um padrão de sinal incidente n -dimensional em um mapa discreto m -dimensional (normalmente uni ou bidimensional) e realizar esta transformação adaptativamente de uma maneira topologicamente ordenada. Desta forma, é criado um mapa topográfico dos padrões de entrada, sendo possível abstrair características intrínsecas contidas nestes padrões de acordo com as suas localizações espaciais [34]. A principal fonte de inspiração é a ilustrada na Figura V.2.1, dando origem à propriedade da rede de Kohonen de mapear vetores que se encontram próximos no espaço original, os quais constituem os estímulos de entrada da rede de Kohonen, em uma mesma região do mapa.

No caso unidimensional, a grade é formada por uma sequência de neurônios conectados lateralmente, como mostra a Figura V.2.2. Cada neurônio é conectado a, no máximo, dois outros neurônios. Existem dois tipos de grade: aquelas que possuem extremos (i.e., dois neurônios nas extremidades da rede ligados a apenas um neurônio vizinho) e as cíclicas (neurônios ligados na forma de um anel).

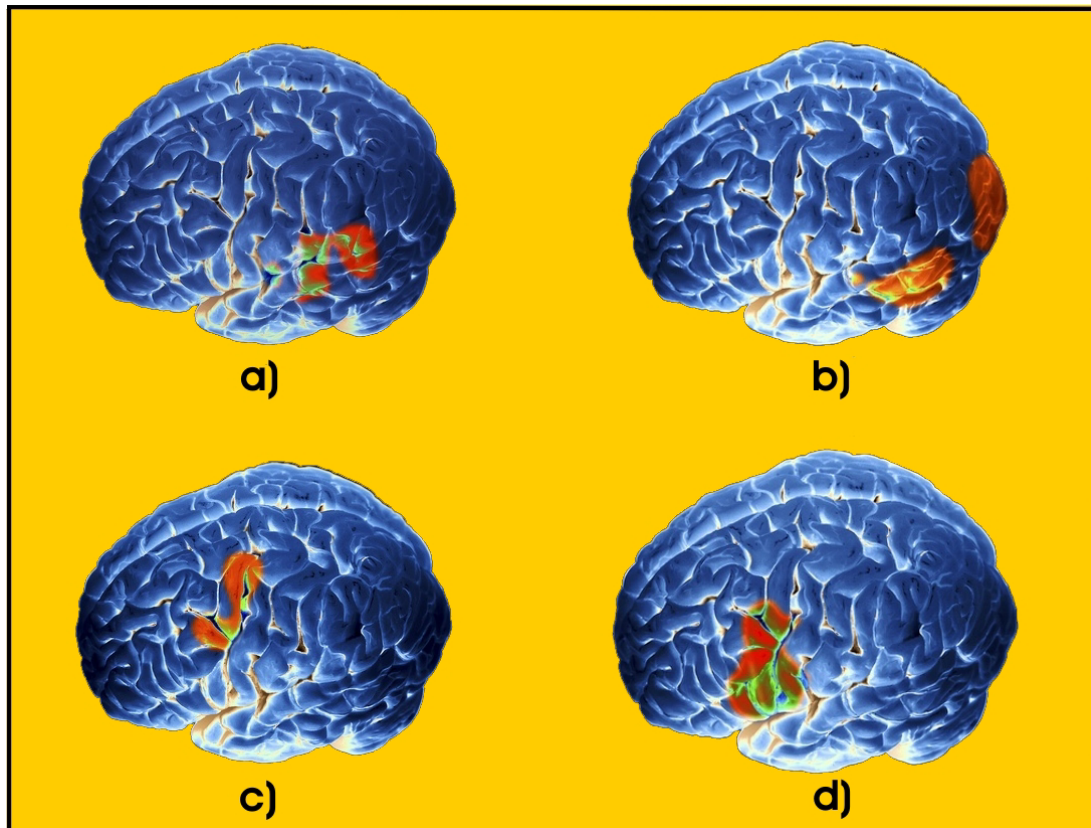


Figura V.2.1: Imagens tridimensionais de ressonância magnética da atividade cerebral.
A região em vermelho mostra a área do cérebro que está ativa durante cada uma das tarefas descritas a seguir: parte a) ouvindo palavras, parte b) lendo palavras, parte c) falando palavras e parte d) pensando em palavras.

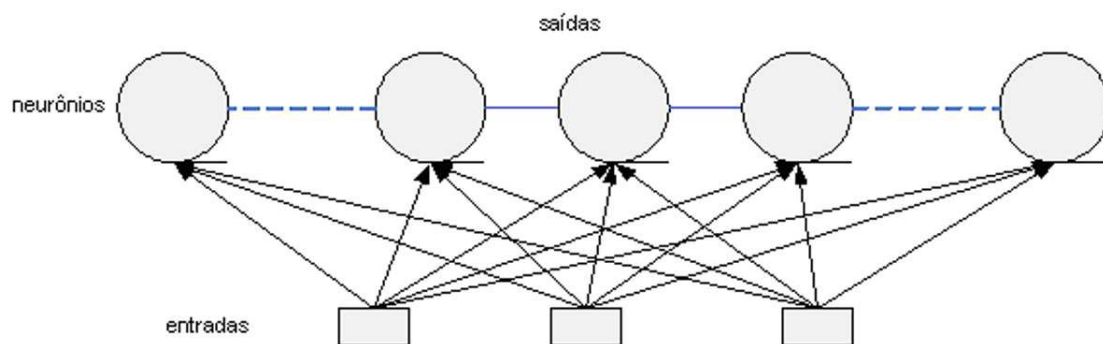


Figura V.2.2: Esquema de uma rede de Kohonen unidimensional.
Os neurônios, correspondentes à saída da rede, no caso unidimensional, são ligados a apenas um vizinho de cada lado, formando um arranjo em anel. Os pesos sinápticos correspondem à importância relativa de cada uma das conexões das entradas aos neurônios.

Existem provas específicas sobre a garantia de convergência de Mapas Auto-Organizáveis (MAOs) unidimensionais [40]. Foi ainda provado que, no caso unidimensional, os pesos dos neurônios sempre ficarão ordenados após a convergência. MAOs são a estratégia mais utilizada em redes neurais artificiais (RNAs) para treinamento não supervisionado. Durante o processo de treinamento, os neurônios tendem a representar propriedades estatísticas dos dados de entrada, preservando ao máximo as relações topológicas do espaço de entrada.

V.2.i) Formação do Mapa Topológico

O mapa topológico é formado através de um processo de aprendizado na rede. Durante o processo de treinamento, os neurônios tendem a representar propriedades estatísticas dos dados de entrada, incluindo a preservação de relações topológicas do espaço de entrada. Este treinamento é não-supervisionado, por isso o mapa gerado emerge dos dados de treinamento. Esta é a fase auto-organizável, quando as entradas são apresentadas à rede, que se auto-organiza formando o mapa. Existem três processos principais nesta fase: o competitivo, o cooperativo e o adaptativo. A fase de formação termina quando a dinâmica do sistema converge para um estado estável, isto é, quando não houver modificações significativas no mapa com a apresentação das entradas.

V.2.ii) Inicialização do Mapa

Os algoritmos de convergência de redes auto-organizáveis são, em geral, altamente sensíveis à condição inicial dos neurônios. A inicialização ideal da rede é específica do problema e da arquitetura da rede a ser utilizada, e deve ser considerada no projeto de soluções com uso de MAOs. Normalmente, a inicialização é aleatória, de maneira a não impor nenhuma ordem a priori sobre o mapa [34].

V.2.iii) O Processo Competitivo.

O processo competitivo é um mecanismo de seleção baseado em uma medida de adequação (*fitness*). Dada uma entrada para o sistema, cada agente (neurônio) calcula a sua adequação e a repassa para os outros agentes. Então, o agente que possui um valor

maior do que qualquer outro recebido é declarado vencedor [34]. Em um SOM, o valor da adequação é uma métrica do erro entre a entrada (estímulo ou atrator) e o neurônio. Como consequência, o neurônio vencedor é aquele que possui o menor erro, considera-se então o inverso do erro como adequação a ser maximizada. Esta métrica pode ser um produto escalar (somente no caso de pesos normalizados) ou norma euclidiana.

Uma vez obtido o neurônio vencedor, este tem o seu vetor de pesos modificado com a finalidade de aumentar a sua adequação (i.e., diminuir o erro) à entrada apresentada.

V.2.iv) O Processo Cooperativo

Esta etapa é inspirada em um processo biológico no qual um neurônio ativado tende a afetar o estado dos neurônios próximos de sua vizinhança topológica. Em particular, um neurônio ativado tende a excitar neurônios em sua vizinhança imediata mais do que aqueles mais distantes [34]. No caso do MAO, não só o vencedor tem sua adequação evoluída; os neurônios que cooperaram para a vitória também são contemplados com a possibilidade de aprender. São considerados neurônios cooperativos aqueles vizinhos ao vencedor: quanto mais próximos do vencedor, mais cooperativos eles são [49], [65]. É definida então uma função de vizinhança que indica o grau de cooperatividade de um determinado neurônio em função de sua distância ao vencedor. O grau de aprendizado do neurônio em relação à entrada é proporcional ao grau de cooperatividade. Um exemplo comum de função de vizinhança que satisfaz as condições do modelo descrito é a função gaussiana, descrita na equação 5.1 e mostrada na Figura V.2.3.

$$\gamma_{j,i(x)} = \exp\left(-\frac{d_{j,i(x)}^2}{2\sigma^2}\right) \quad (5.1)$$

onde $i(x)$ é o índice seqüencial no neurônio vencedor para o estímulo de entrada x , j é o índice do neurônio vizinho, $d_{j,i}$ representa a distância lateral entre o neurônio vencedor e o vizinho j , e σ é o desvio padrão, isto é, a abertura da gaussiana.

Uma característica ímpar do SOM é que a vizinhança diminui com o tempo de modo a permitir a convergência do mapa. Satisfaz-se facilmente este requerimento diminuindo-se o valor de σ com as etapas sucessivas de treinamento. Uma das escolhas mais comuns é o decaimento exponencial da abertura da função de vizinhança, como descrito pela equação 5.2.

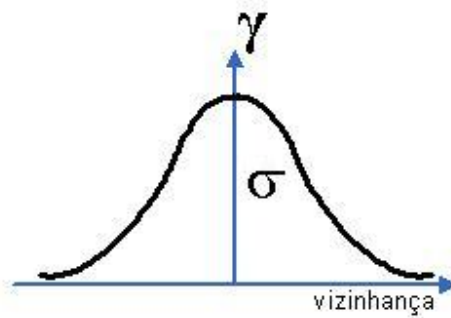


Figura V.2.3: Função de vizinhança gaussiana.

É mostrada a função de vizinhança gaussiana e sua abertura σ variável ao longo das iterações.

$$\sigma(n) = \sigma_0 \exp\left(-\frac{n}{\tau_1}\right) \quad (5.2)$$

Aqui, σ_0 é o valor inicial, τ_1 é a constante de tempo (ou de decaimento) e n é a iteração (passo no tempo discreto).

V.2.v) O Processo Adaptativo

No processo adaptativo, o neurônio vencedor tem os seus pesos sinápticos alterados no sentido de se aproximar da entrada. De acordo com a regra de aprendizado de Hebb, um peso sináptico é aumentado com a ocorrência de atividade sináptica [34]. O ajuste é feito com base na seguinte expressão, aplicada a todos os neurônios da rede que estão dentro da vizinhança do neurônio vencedor:

$$\omega_j(n+1) = \omega_j(n) + \eta(n) \times \gamma_{j,i(x)} \times (x - \omega_j(n)) \quad (5.3)$$

sendo j o neurônio vencedor, $\omega_j(n)$ o vetor de pesos do neurônio j , x o estímulo de entrada e $\eta(n)$ a taxa de adaptação (ou taxa de aprendizado) e $\gamma_{j,i(x)}$ a função de vizinhança.

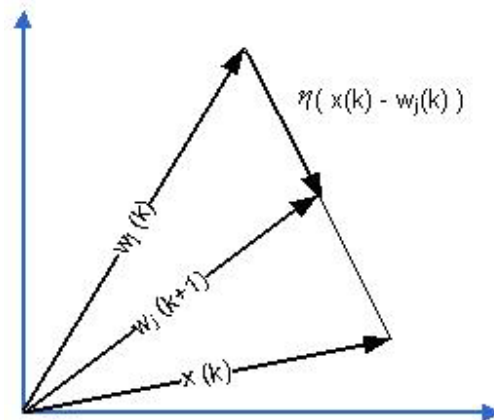


Figura V.2.4: Adaptação do vetor de peso do neurônio vencedor.

A adaptação do vetor de pesos sinápticos do neurônio vencedor se dá como mostra a figura que ilustra a equação 4.3 em termos vetoriais.

A equação 5.3 tem o efeito de mover o vetor de pesos sinápticos do neurônio vencedor em direção ao estímulo que o ativou (dado de entrada), juntamente com seus vizinhos, diminuindo progressivamente obedecendo a função de vizinhança. Graficamente, é possível verificar que o vetor de pesos após a adaptação se aproximou da entrada, e a taxa de adaptação indica o grau dessa aproximação, como mostrado na Figura V.2.4. A taxa de adaptação também varia com o passar das interações, fazendo com que a influência da entrada sobre os pesos diminua gradativamente com o passar do processo de acordo com a equação 5.4.

$$\eta(n) = \eta_0 \exp\left(-\frac{n}{\tau_2}\right) \quad (5.4)$$

Aqui, η_0 é a taxa de aprendizado inicial e τ_2 é a constante de tempo (decaimento) do aprendizado.

Nas próximas seções, serão explicados o algoritmo e todos os parâmetros (internos e de entrada) do programa implementado. Então uma breve análise dos parâmetros será apresentada para o problema do caixeiro viajante. O problema do caixeiro viajante é uma aplicação bastante difundida de MAOs e de grande relevância pois, neste problema, os dados estão distribuídos no plano, facilitando a visualização do processo de auto-organização. A diferença é que a instância do problema apresentada aqui tem sons ao invés de cidades.

V.2.vi) Algoritmo de Ajustes de Peso

O algoritmo do mapa auto-organizável [34] é apresentado a seguir. As etapas competitiva, cooperativa e adaptativa estão explicitadas no algoritmo.

1. *Inicialização*: Inicialize os pesos sinápticos de cada neurônio aleatoriamente ou de acordo com algum critério de interesse,
2. A cada iteração faça:
3. Para cada estímulo faça:
4. *Competição*: Calcule a distância de todos os neurônios a este estímulo usando a métrica euclidiana,
5. Determine o neurônio vencedor, isto é, aquele que apresentar menor distância entre seu vetor de pesos e o estímulo de entrada,
6. *Cooperação e Adaptação*: Ajuste o peso sináptico de todos os neurônios de acordo com a fórmula: $\omega_j(n+1) = \omega_j(n) + \eta(n) \times h_{j,i(x)} \times (x - \omega_j(n))$, onde $\eta(n)$ é a taxa de aprendizado e $h_{j,i(x)}$ é a função de vizinhança centrada no neurônio vencedor, ambas variam com n , o número de iterações,
7. Teste o critério de parada.

Algoritmo 2: Algoritmo simplificado do mapa auto-organizável de Kohonen.

Diferentemente do algoritmo da aiNet apresentado na seção IV.1.vi), este é mais simples e de fácil implementação, necessitando apenas algumas linhas de código. As fases de competição, relativa ao passo 6, e de cooperação e adaptação, ambas realizadas na implementação da equação 5.3, mostrada no passo 5, são as etapas de aprendizado do mapa. A formação do mapa se dá com a repetição dos passos 4 a 6.

V.2.vii) Parâmetros

Os parâmetros manipuláveis do programa implementado se encontram detalhados a seguir.

Tabela V.2.1: Parâmetros de entrada do mapa auto-organizável.

Parâmetro	Descrição
L	Número de amostras por indivíduo
SF	Taxa de amostragem
G	Número de estímulos
N	Número de neurônios (pesos)
Lr	Taxa de aprendizado inicial
$Radius$	Raio inicial
τ_2	Constante de tempo de decaimento exponencial da taxa de aprendizado
τ_1	Constante de tempo de decaimento exponencial do raio
$Epoch$	Número de épocas

Os parâmetros número de amostras sonoras digitais por indivíduo L e taxa de amostragem SF são específicos da aplicação em síntese sonora e definem a duração de cada som. Os parâmetros restantes são válidos para qualquer aplicação. O número de estímulos (atratores) interfere diretamente no resultado final, pois está diretamente relacionada à caracterização da região do espaço escolhida. Diferentemente da aiNet, a implementação do mapa auto organizável utilizada não possui ajuste dinâmico do número de neurônios. Isto é, escolhe-se o número de neurônios que vão compor o mapa até o fim do treinamento. Alguns fenômenos bastante interessantes decorrem da utilização criativa desta característica do método, que serão exploradas na seção V.3. O número de épocas é a quantidade de vezes que os dados são apresentados à rede para treinamento. Os parâmetros essenciais são as constantes de tempo τ_1 e τ_2 das equações 5.2 e 5.4 e o raio inicial. A abertura (desvio padrão) da vizinhança gaussiana, juntamente com a taxa de decaimento da vizinhança, também podem ter uma influência interessante no resultado, controlando a rapidez de convergência do método. Esta é uma variável de controle mais sutil, no entanto. A constante de tempo da taxa de aprendizado e a taxa de aprendizado inicial definem a plasticidade final da rede e a representação do espaço original pela grade de neurônios.

V.2.viii) Problema do Caixeiro Viajante:

Este é o problema que foi apresentado brevemente na seção III.3.i), que trata de otimização. No entanto, o problema será formulado novamente e sua resolução por mapas auto-organizáveis unidimensionais será explicada. O problema consiste da obtenção de um caminho mínimo no problema clássico do caixeiro viajante. Dado um número de cidades a ser visitado, pede-se a trajetória mínima que passe uma única vez

em cada cidade e retorne ao ponto de partida. A proposta de solução para o problema do caixeiro viajante por mapas auto-organizáveis é obtida com uma rede unidimensional cíclica, isto é, com um mapa unidimensional em forma de anel em que o número de neurônios normalmente é o mesmo que o de cidades. Após a convergência, cada neurônio representará uma cidade, e o caminho é descrito pelas vizinhanças dos neurônios. A resolução de problemas de caixeiro viajante é a principal aplicação das redes de Kohonen unidimensionais. Apesar de não apresentar os melhores desempenhos da literatura, em instâncias do caixeiro viajante com muitas cidades, este método é interessante por conseguir resultados satisfatórios com um custo computacional baixo. Pode ser usado também como estado inicial de outros métodos mais eficientes. A sua principal vantagem sobre outros modelos é a sua versatilidade e fácil implementação. Por este motivo, os algoritmos de MAO têm sido aplicados a variações do problema do caixeiro viajante original. Por exemplo, o método pode ser usado também no problema de caixeiros viajantes múltiplos e concorrentes, com o uso de múltiplas grades [13]. Neste caso, vamos nos ater ao problema clássico. Ressalta-se novamente que a instância do problema apresentada aqui utiliza sons ao invés de cidades.

V.3 Modelo

Um método de síntese sonora utilizando o mapa auto-organizável unidimensional de Kohonen foi desenvolvido a partir das idéias iniciais apresentadas na introdução teórica e motivação deste texto. No modelo do MAO os atratores são chamados de estímulos e os agentes são os pesos sinápticos dos neurônios artificiais, chamados de neurônios por simplicidade. A mesma codificação utilizada no método imunológico descrito anteriormente foi utilizada, resultando no mesmo espaço de busca; apenas a técnica utilizada para se atingir o resultado final é outra. A justificativa é simples: diferentes paradigmas bio-inspirados exploram diferentes características dos problemas por lançarem mão de diferentes recursos encontrados em sistemas naturais, tais como: aprendizado, memória, adaptação, manutenção de diversidade, autonomia, auto-organização, complexidade, etc. Este estudo, em particular, fez uso das características de aprendizado não-supervisionado, adaptação, preservação de topologia e auto-organização dos mapas auto-organizáveis para a geração de material sonoro com determinadas características definidas a priori. Este método permite ao usuário usufruir de algum grau de subjetividade através da escolha do número de neurônios da rede de acordo com determinados critérios estéticos. O usuário é capaz de obter soluções

candidatas que respeitam certas exigências musicais através do uso de um conjunto de formas de onda (estímulos) como exemplos das qualidades sonoras desejadas (Figura V.4.1 a). Ao invés de tentar descrever as qualidades exaustivamente com parâmetros numéricos ou através de qualquer outra ferramenta lingüística, um conjunto de formas de onda é usado para representar as qualidades sonoras desejadas no resultado.

Não há um consenso a respeito da organização ou classificação dos tons no espaço de formas composicional. Devido à característica auto-organizável e à preservação topológica da rede de Kohonen, é possível propor ordenamentos sonoros respeitando a topologia do espaço original. Isto é, sons que se encontram próximos no espaço de formas composicional estarão associados a neurônios vizinhos no mapa resultante (Figura V.4.1 b).

Há alguns fenômenos interessantes neste processo. Primeiramente, um estímulo pode ativar mais de um neurônio, resultando em um posicionamento tal dos neurônios no espaço cuja representação final resulte em uma variação dos estímulos de entrada (ampliações na Figura V.4.1 b). Há também a possibilidade de mais de um estímulo ativar um mesmo neurônio, forçando-o a representar ambos e, conseqüentemente, conter qualidades sonoras mistas, advindas de mais de um estímulo (Figura V.4.1 c).

Finalmente, o arranjo unidimensional resultante (Figura V.4.1 b) corresponde a uma ordenação cíclica dos estímulos que pode ser utilizada como uma espécie de macro-transformação sonora. Seguindo a ordem resultante da aplicação do método, pode-se evidenciar o caráter híbrido (cruzado) de certos neurônios, bem como as variantes representadas por outros, dependendo apenas de sua posição relativa no espaço e da ordenação proposta.

V.4 Experimentos e Resultados

Nesta seção, o experimento realizado será explicado e os resultados (formas de onda e espectros dinâmicos) serão mostrados, acompanhados das distâncias obtidas de acordo com as métricas matemáticas euclidiana (temporal) e tempo-freqüencial, além de uma análise subjetiva da similaridade perceptiva entre os sons-neurais (atratores) e os resultados. O propósito do experimento é mostrar a eficácia do método em produzir resultados que representam variações dos sons atratores (sons-estímulo) e possivelmente apresentam características advindas de mais desses sons atratores por representarem um ponto intermediário no espaço entre dois ou mais desses sons. Será

feita uma análise comparativa levando-se em conta os critérios de flexibilidade, controlabilidade, adaptabilidade e simplicidade adotados como relevantes em um método de síntese sonora. Finalmente algumas conseqüências da representação vetorial adotada, juntamente com o aspecto temporal e a abordagem populacional serão explorados.

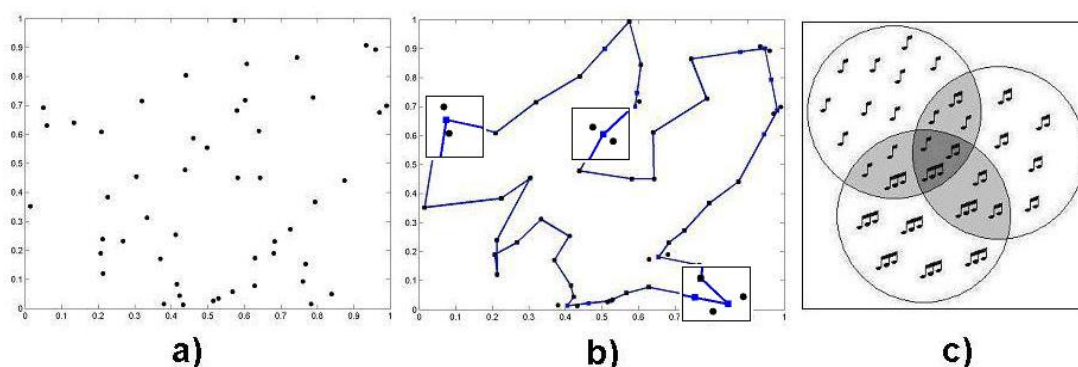


Figura V.4.1: Representação da aplicação de Mapas Auto-Organizáveis de Kohonen ao problema de síntese sonora digital desenvolvido como uma solução do problema do caixeiro viajante.

A parte (a) mostra a distribuição dos sons-estímulo no espaço original; a parte (b) mostra o arranjo unidimensional resultante e algumas ampliações locais; e a parte (c) ilustra as qualidades sonoras comuns a três classes distintas de sons.

Os parâmetros utilizados nos experimentos estão mostrados a seguir na tabela 4.2. A rede utilizada foi composta de 14 neurônios em anel. As relações de vizinhança são expressas pelos rótulos dos neurônios. Isto é, o neurônio 1 é vizinho do 2 e do 14, e assim respectivamente. Os sons atratores utilizados são os mostrados da Figura III.3.3 à Figura III.3.8. Foram usados 14 sons-neurais para que o número de sons-neurais seja maior que o número de sons-estímulo (cidades, no caso do problema do caixeiro viajante) e, portanto, alguns sons-neurais sejam forçados a não vencer nunca na fase competitiva e não representar nenhum som-estímulo específico, tendo apenas seus pesos sinápticos ajustados na fase cooperativa através das relações de vizinhança. Conseqüentemente, espera-se que esses sons-neurais se posicionem em pontos intermediários a dois sons-estímulo (cidades), devidamente representados por outros sons-neurais (possivelmente seus vizinhos), e apresentem características advindas de mais de um som-estímulo. Ressalta-se, mais uma vez, que essas qualidades podem ser advindas de variados parâmetros musicais, tais como freqüência, qualidades timbrais ou outras. O raio inicial da vizinhança é 3 – isto é, cada neurônio interage com até o terceiro vizinho da esquerda e

da direita - para que cada som-neural interaja com os outros sons-neurais vizinhos de modo a maximizar a cooperação entre eles e possibilitar aos sons-neurais que nunca vencem na fase competitiva que tenham seus pesos sinápticos alterados. Os pesos sinápticos dos sons-neurais foram inicializados como ruído branco, o que significa que foram distribuídos aleatoriamente pelo espaço. Se o raio da vizinhança fosse menor, esses sons-neurais que não vencem permaneceriam nesses pontos, ou seja, seriam ruído. A taxa de amostragem SF é 44100 para que todas as frequências audíveis (até 20 kHz) possam ser representadas e não se perca informação espectral. O número de amostras L foi definido como 4096 para que, com o valor de SF , a duração dos sons fosse aproximadamente 0,1 segundo, suficiente para que os resultados pudessem ser avaliados perceptivamente. No entanto, é importante salientar que L é a dimensão do espaço vetorial de busca, portanto, definida a taxa de amostragem, quanto maior o número de amostras L , maior a dimensionalidade do espaço de busca e, conseqüentemente, maior a dificuldade de se achar boas soluções. A taxa inicial de aprendizado e as duas constantes de tempo foram determinadas empiricamente a partir do número de épocas.

Tabela V.4.1: Parâmetros do SOM

L	SF	G	N	Lr	$epoch$	$Radius$	$\tau_2 r$	$\tau_1 r$
4096	44100	7	14	0.2	50	3	150	150

A seguir, são apresentadas tanto as formas de onda quanto os espectros dinâmicos dos 14 sons resultantes, isto é, os sons-neurais.

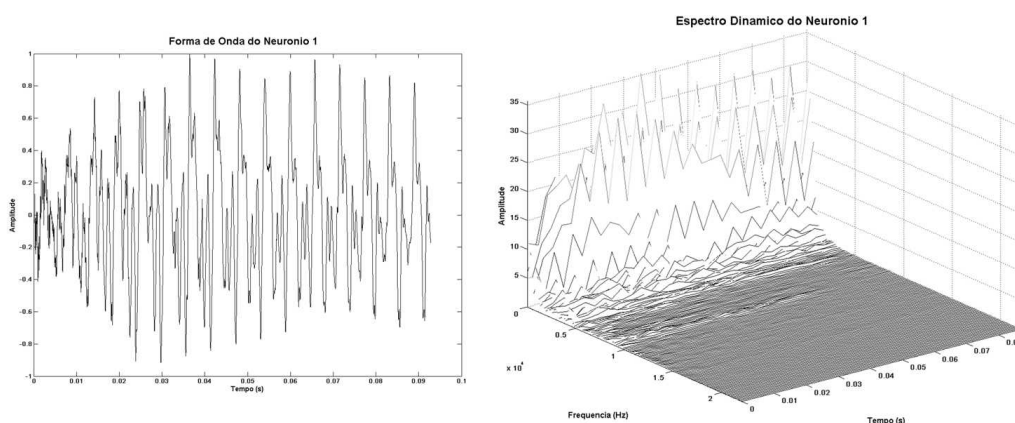


Figura V.4.2: Forma de onda e Espectro Dinâmico do Neurônio 1.

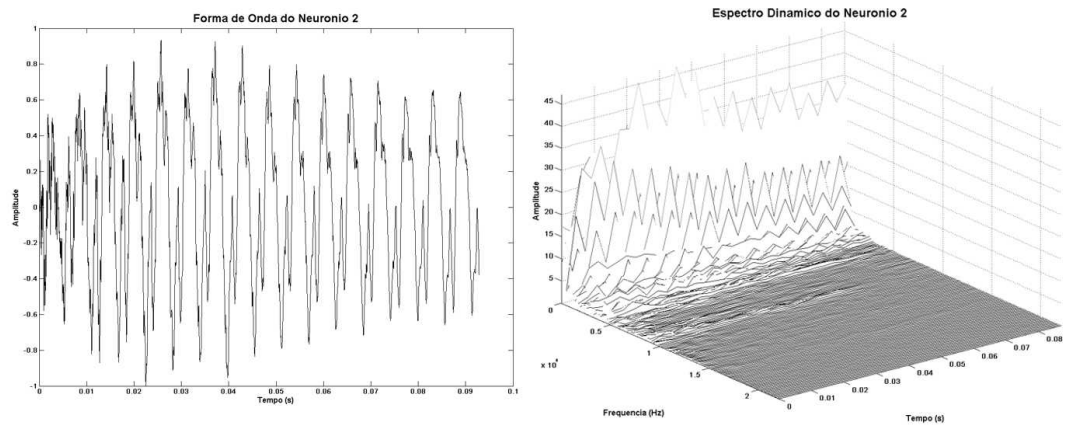


Figura V.4.3: Forma de onda e Espectro Dinâmico do Neurônio 2.

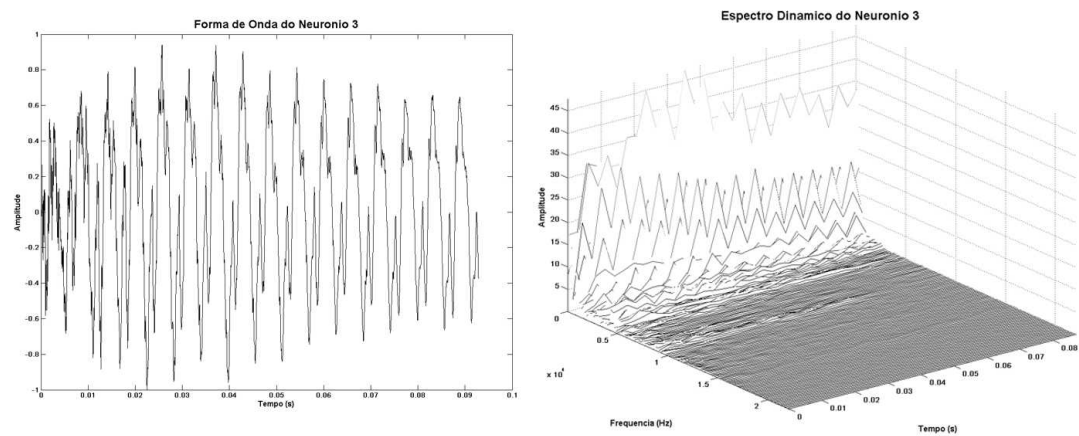


Figura V.4.4: Forma de onda e Espectro Dinâmico do Neurônio 3.

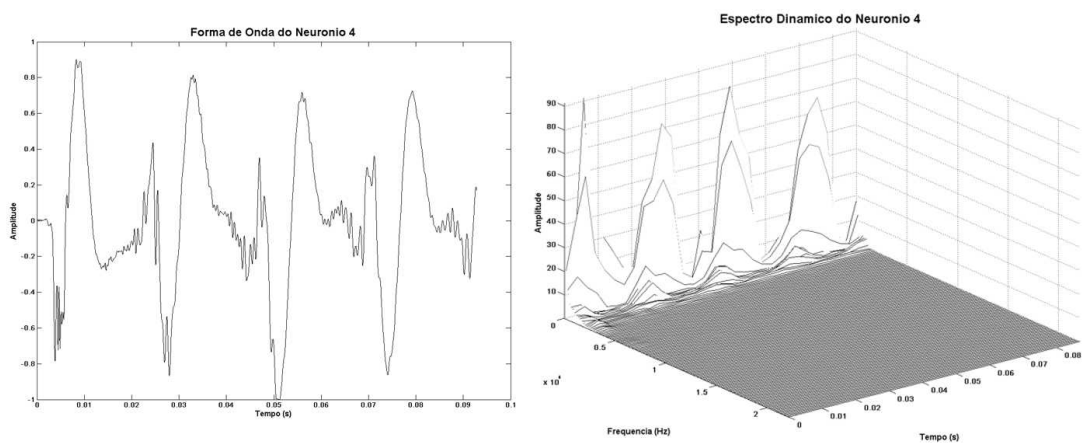


Figura V.4.5: Forma de onda e Espectro Dinâmico do Neurônio 4.

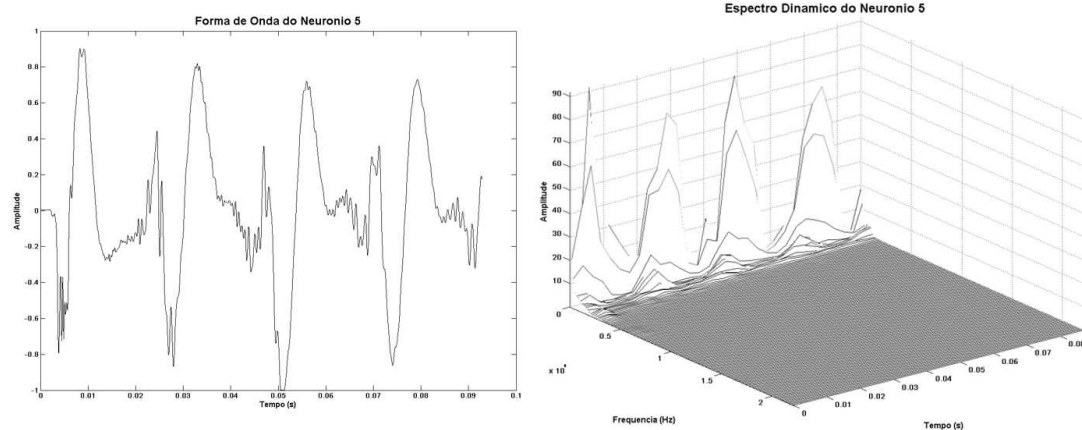


Figura V.4.6: Forma de onda e Espectro Dinâmico do Neurônio 5.

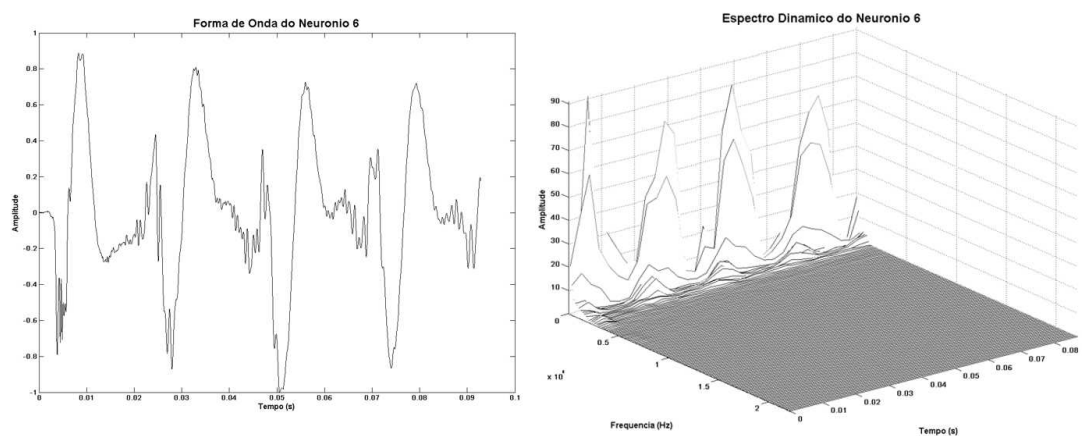


Figura V.4.7: Forma de onda e Espectro Dinâmico do Neurônio 6.

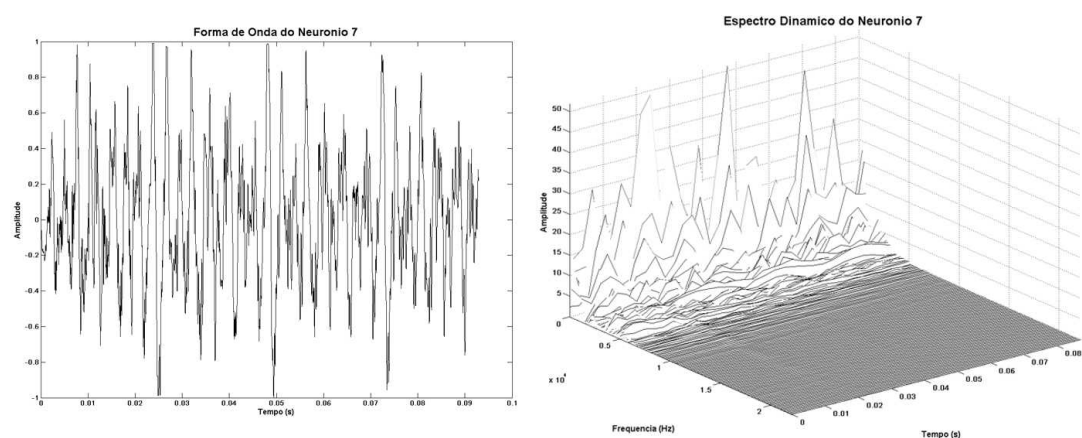


Figura V.4.8: Forma de onda e Espectro Dinâmico do Neurônio 7.

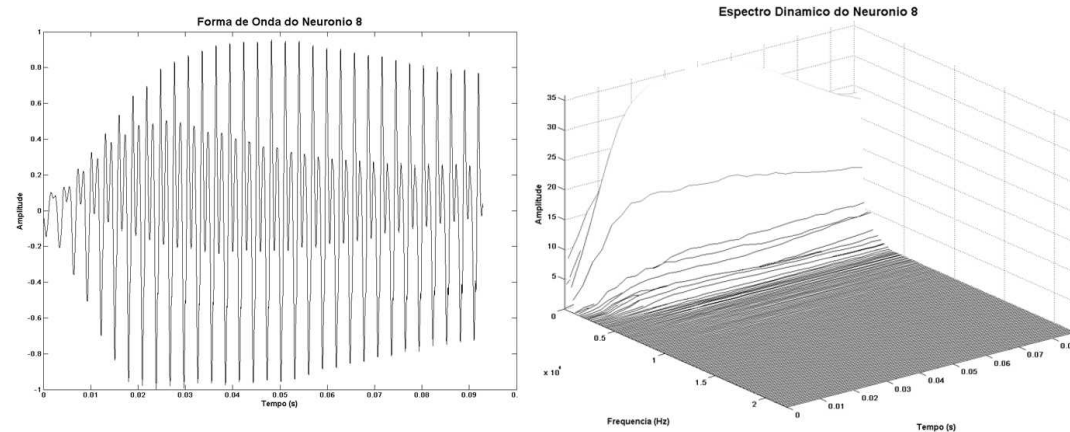


Figura V.4.9: Forma de onda e Espectro Dinâmico do Neurônio 8.

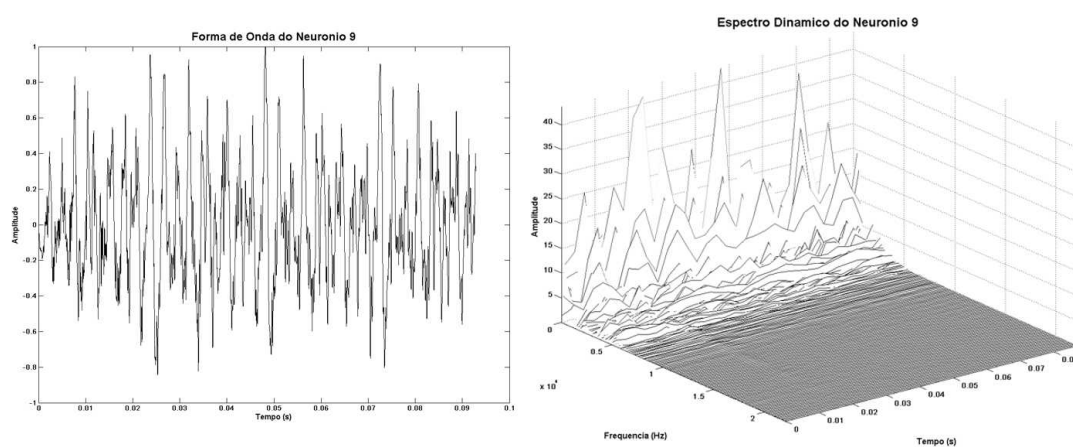


Figura V.4.10: Forma de onda e Espectro Dinâmico do Neurônio 9.

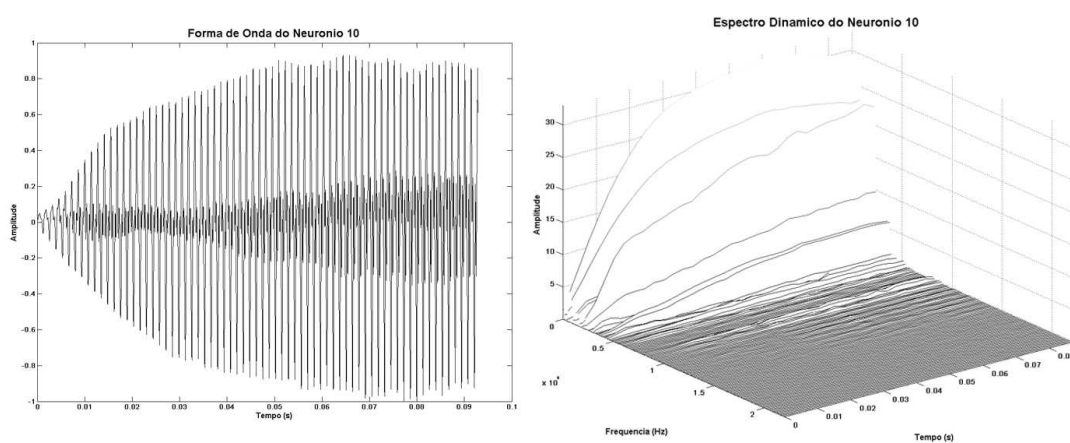


Figura V.4.11: Forma de onda e Espectro Dinâmico do Neurônio 10.

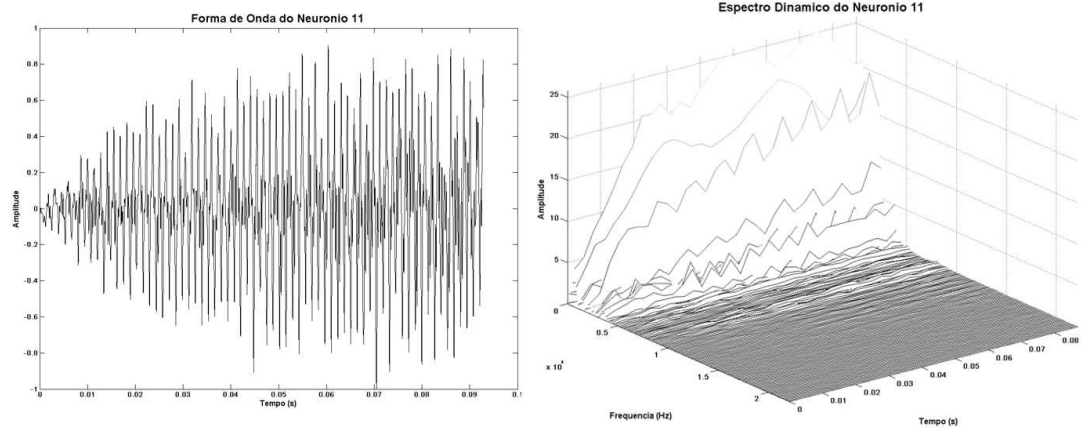


Figura V.4.12: Forma de onda e Espectro Dinâmico do Neurônio 11.

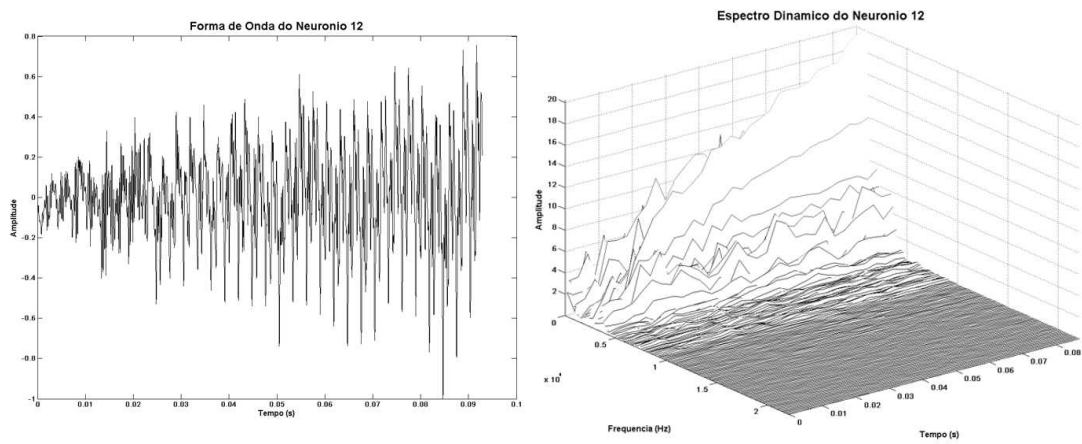


Figura V.4.13: Forma de onda e Espectro Dinâmico do Neurônio 12.

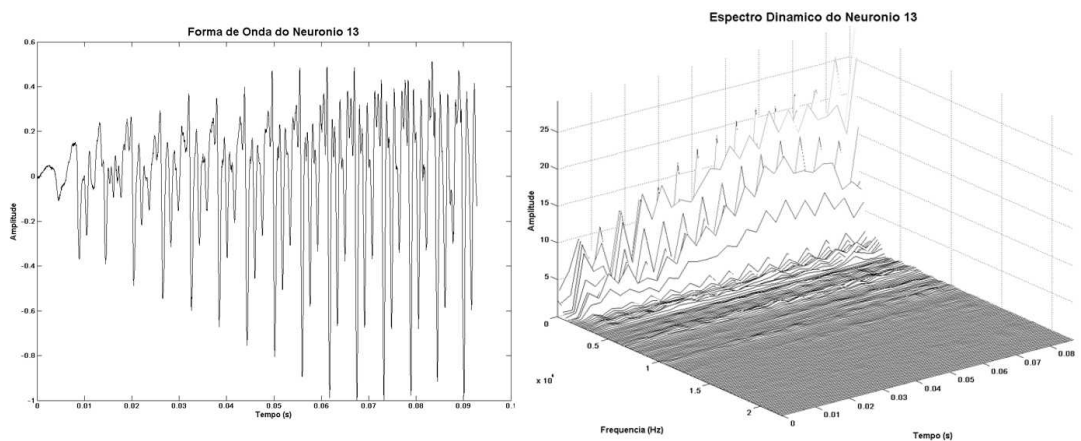


Figura V.4.14: Forma de onda e Espectro Dinâmico do Neurônio 13.

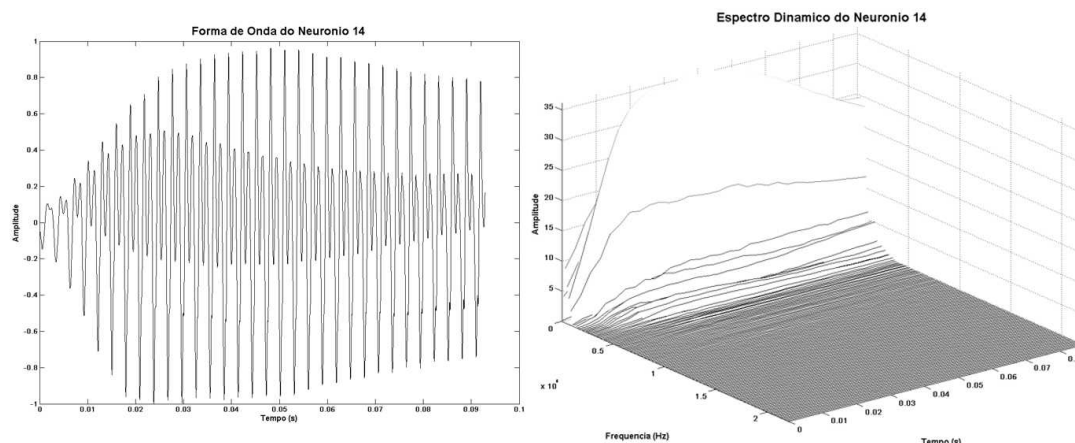


Figura V.4.15: Forma de onda e Espectro Dinâmico do Neurônio 14.

Primeiramente, é essencial perceber que todos os espectros dinâmicos apresentam variações individuais dos envelopes dos parciais ao longo do tempo, como desejado em sons complexos para aplicações musicais. Uma breve comparação visual com os sons atratores apresentados na seção III.3.iv) já transparece a grande similaridade visual existente entre a forma de onda e o espectro dinâmico de cada um dos resultados (sons-neurais) e pelo menos um dos sons atratores (sons-estímulo). A Tabela V.4.2 apresenta a resultado do cruzamento visual dos sons atratores mostrados da Figura III.3.3 à Figura III.3.9 com os resultados apresentados da Figura V.4.2 à Figura V.4.15.

Tabela V.4.2: Agentes mais próximos de cada atrator de acordo com o cruzamento visual.

Atrator	1	2	3	4	5	6	7
Agente relacionado	8,14	4,5,6	1,2,3	7,9	12	13	10

A seguir serão apresentadas as distâncias de acordo com as métricas matemáticas temporal, na Tabela V.4.3, tempo-freqüencial, na Tabela V.4.5, e a avaliação subjetiva de cinco indivíduos das similaridades perceptivas entre os sons-neurais (sons agentes) e os sons-estímulo (sons atratores) na Tabela V.4.7. Os indivíduos usados nessa avaliação não são os mesmos que os da avaliação perceptiva anterior. Destaca-se, novamente, que os 5 indivíduos não possuem treinamento musical, mas possuem treinamento científico para compreender a tarefa que lhes foi solicitada, isto é, a avaliação subjetiva da similaridade entre pares de sons em uma escala de 1 a 5.

A Tabela V.4.3 mostra a medida de distância entre os sons-neurais (sons agentes) e os sons-estímulo (sons atratores) segundo a métrica temporal (euclidiana). A distância mínima ao longo das linhas da Tabela V.4.3 e da Tabela V.4.5 estão marcadas em negrito. Estas distâncias, quando lidas ao longo das colunas da tabela, indicam qual som-estímulo (som atrator) está sendo representado por qual som-neural (som agente).

Tabela V.4.3: Distância temporal (euclidiana) entre-os atratores e os agentes para SOM

	Atrator 1	Atrator 2	Atrator 3	Atrator 4	Atrator 5	Atrator 6	Atrator 7
Agente 1	21.7086	33.8899	16.9087	34.2074	27.0109	24.5645	34.8182
Agente 2	37.7066	35.6005	0.1962	36.3815	30.2357	29.7689	37.2401
Agente 3	37.9338	35.0004	1.2359	36.6201	30.4813	30.0053	37.4889
Agente 4	37.3797	0.4077	35.3123	34.9960	28.5233	29.6436	35.6108
Agente 5	37.5051	0.3304	35.6231	35.1198	28.6764	29.8063	35.7292
Agente 6	37.3999	0.5957	35.5161	34.6522	28.5271	29.6811	35.5908
Agente 7	36.5072	34.4135	36.0604	0.6289	27.4507	29.5460	33.1584
Agente 8	2.4780	37.2758	37.4001	36.5677	29.6523	26.9786	36.6662
Agente 9	34.6632	32.9521	34.3713	6.1268	25.1880	27.4059	27.2811
Agente 10	36.8411	35.6289	37.1697	33.3723	29.0646	30.6558	0.0504
Agente 11	33.1993	31.9755	33.5492	29.9397	20.1583	26.3801	9.6224
Agente 12	29.9931	28.5583	30.1542	27.7446	0.1265	22.0013	29.0639
Agente 13	27.0126	29.7187	29.7139	29.8595	22.0395	0.0789	30.6837
Agente 14	0.0079	37.4260	37.6561	36.7382	30.0115	27.0159	36.8567

A seguir, a Tabela V.4.4 apresenta as distâncias mínimas individualmente, marcadas em negrito na Tabela V.4.3, como descrito no parágrafo anterior. Deve-se notar que, no caso em que o número de sons agentes (sons-neurais) é maior que o número de sons atratores (sons-estímulo), é possível ter mais de um som-neural representando o mesmo som-estímulo.

Tabela V.4.4: Agentes mais próximos de cada atrator de acordo com a distância temporal.

Atrator	1	2	3	4	5	6	7
Agente relacionado	8,14	4,5,6	1,2,3	7,9	12	13	10,11

A Tabela V.4.5 mostra a avaliação de distância entre os neurônios e os estímulos segundo a métrica tempo-freqüencial.

Tabela V.4.5: Distância tempo-frequencial entre-os atratores e os agentes para SOM.

	Atrator 1	Atrator 2	Atrator 3	Atrator 4	Atrator 5	Atrator 6	Atrator 7
Agente 1	34.7606	140.7773	37.6053	117.6477	34.4166	33.3402	44.7520
Agente 2	66.3417	144.5577	7.1025	123.9408	53.3262	53.3984	60.8768
Agente 3	65.7337	141.9019	8.3761	123.6183	52.4717	52.5984	60.1636
Agente 4	127.6824	18.9327	129.2990	165.1742	119.4274	120.5729	123.7715
Agente 5	129.9995	16.4692	131.9691	166.9731	121.8598	123.0056	126.1364
Agente 6	127.6733	19.0678	129.7479	163.7705	119.3749	120.5557	123.6851
Agente 7	107.3686	167.9962	113.6441	15.6206	97.3370	99.4349	99.5542
Agente 8	6.7819	143.9433	68.7541	119.4962	38.6160	36.9926	47.6565
Agente 9	70.1802	149.4947	80.6534	61.0166	55.0873	58.3408	53.5189
Agente 10	50.4559	141.3996	65.7026	113.7396	28.8807	33.9405	3.4725
Agente 11	45.2253	139.1221	61.4320	112.2853	15.1187	25.3820	17.0501
Agente 12	43.9312	138.1801	60.1144	112.5711	1.4608	22.8388	31.5043
Agente 13	41.3641	138.9457	59.6131	114.1093	21.3422	2.5079	35.2750
Agente 14	5.1640	144.0565	69.1358	119.7276	39.2329	37.2416	48.0991

A seguir, a Tabela V.4.6 apresenta as distâncias mínimas individualmente, marcadas em negrito na Tabela V.4.5.

Tabela V.4.6: Agentes mais próximos de cada atrator de acordo com a distância tempo-frequencial.

Atrator	1	2	3	4	5	6	7
Agente relacionado	8,14	4,5,6	1,2,3	7,9	12	13	10,11

As similaridades entre os sons-estímulo (sons atratores) e os sons-neurais (sons atratores), estimadas pelos 5 indivíduos, são apresentadas na Tabela V.4.7. A tabela apresenta a média e o desvio padrão das aferições. É importante ressaltar que, devido à natureza estatística e subjetiva desta medida de similaridade, diferentemente das métricas matemáticas, nesta tabela é possível destacar mais de um som-estímulo (som atrator) para cada som-neural (som agente). Isto corresponde perceptivamente à noção de que um som-neural (som agente) resultante do processo de síntese possui características advindas de mais de um som-estímulo (som atrator).

Tabela V.4.7: Similaridades subjetivas entre os agentes e os atratores de acordo com as respostas dos indivíduos.

	Atrator 1	Atrator 2	Atrator 3	Atrator 4	Atrator 5	Atrator 6	Atrator 7
Agente 1	3.0±1.58	1.6±0.89	3.8±0.45	2.2±0.84	1.8±0.84	2.4±1.34	1.6±0.89
Agente 2	2.0±1.41	1.4±0.89	4.8±0.45	2.0±0.71	1.6±0.89	1.6±0.89	1.4±0.89
Agente 3	2.0±1.41	1.6±0.89	4.6±0.55	2.0±0.71	1.2±0.45	1.8±0.84	1.4±0.89
Agente 4	1.2±0.45	4.2±1.30	1.8±0.84	1.4±0.55	1.0±0.00	1.0±0.00	1.2±0.44
Agente 5	1.2±0.45	4.8±0.45	1.8±0.84	1.2±0.45	1.0±0.00	1.6±0.89	1.2±0.44
Agente 6	1.2±0.45	4.8±0.45	1.6±0.89	1.2±0.45	1.0±0.00	1.2±0.45	1.2±0.44
Agente 7	1.2±0.45	1.0±0.00	1.4±0.55	5.0±0.00	1.6±0.89	1.0±0.00	1.2±0.44
Agente 8	4.8±0.44	1.2±0.45	1.2±0.45	1.0±0.00	1.8±0.84	1.4±0.89	2.0±0.71
Agente 9	1.2±0.45	1.0±0.00	1.6±0.89	4.6±0.55	1.4±0.55	1.0±0.00	1.4±0.89
Agente 10	2.0±1.00	1.2±0.45	1.0±0.00	1.0±0.00	1.8±0.84	1.2±0.45	5.0±0.00
Agente 11	1.4±0.55	1.2±0.45	1.2±0.45	1.0±0.00	3.0±2.00	1.2±0.45	4.0±0.71
Agente 12	1.6±0.89	1.0±0.00	1.8±0.84	1.0±0.00	5.0±0.00	1.0±0.00	1.6±0.55
Agente 13	1.4±0.89	1.4±0.55	1.4±0.55	1.0±0.00	1.2±0.45	5.0±0.00	1.2±0.45
Agente 14	4.4±0.89	1.4±0.89	1.6±0.89	1.2±0.45	1.8±1.30	1.8±1.30	1.6±0.89

A seguir, a Tabela V.4.8 apresenta as similaridades máximas individualmente, marcadas em negrito na Tabela V.4.7.

Tabela V.4.8: Agentes mais próximos de cada atrator de acordo com a avaliação perceptiva.

Atrator	1	2	3	4	5	6	7
Agente relacionado	8,14	4,5,6	1,2,3	7,9	11,12	13	10,11

V.5 Discussão

Uma breve inspeção visual da Figura V.4.2, da Figura V.4.3 e da Figura V.4.4 revela que, apesar de estarem representando o mesmo som-estímulo, esses 3 sons-neurais não são exatamente iguais, mas variações do som-estímulo que estão representando. Novamente, através da inspeção visual, pode-se claramente perceber que a Figura V.4.12 apresenta um som-neural que é intermediário entre os apresentados na Figura V.4.11 e na Figura V.4.13, exemplificando o caso em que um som-neural é posicionado no espaço entre dois sons-estímulo (ou duas cidades no problema do caixeiro viajante). É interessante notar que, comparando-se a Tabela V.4.2, a Tabela V.4.4, a Tabela V.4.6 e a Tabela V.4.8, conclui-se que, apesar de pouco prática, a inspeção visual confirma as similaridades entre os sons-neurais resultantes e os sons-estímulo apresentados. Exceto para o caso particular do som-neural 11 mostrado na Figura V.4.12. Visualmente é difícil decidir se esse está mais próximo do mesmo som-estímulo (no caso o som-estímulo 7)

que o som-neural 10, mostrado na Figura V.4.11, ou do mesmo som-estímulo (no caso o som-estímulo 5) que o som-neural 12, mostrado na Figura V.4.13. Tanto a avaliação perceptiva, quanto ambas as métricas matemáticas indicam que o som-neural 11 está mais próximo do som-estímulo 7. Entretanto, devido ao grande valor do erro associado à similaridade do som-agente 11 (som-neural) com o som-atrator 5 (som-estímulo), pode-se inferir que a similaridade é suficiente para se associá-los. É interessante notar que, segundo as distâncias da métrica tempo-freqüencial apresentadas na Tabela V.4.5, a impressão deixada pela inspeção visual de que o som-neural 10 está próximo do som-estímulo 5 e do som-estímulo 7 é confirmada. No entanto, a métrica euclidiana não confirmou este fato. Por fim, é interessante notar que, neste caso, apesar de os sons atratores (sons-estímulo) não terem sido considerados próximos no espaço original (métrica euclidiana) na seção III.3.iv), a avaliação perceptiva agrupou os sons atratores 5, 7 e 1. Pode-se, portanto, confirmar que as características perceptivas de similaridades entre ambos são detectadas facilmente nos resultados.

Os resultados da Tabela V.4.3, da Tabela V.4.5 e da Tabela V.4.7 são de mais fácil interpretação quando comparados com a interpretação cruzada da caracterização dos atratores na seção III.3.iv). Primeiramente, deve-se notar que, diferentemente da análise anterior, agora busca-se a similaridade entre os sons atratores (sons-estímulo) e os sons agentes (sons-neurais). Portanto, basta analisar a Tabela V.4.4, a Tabela V.4.6 e a Tabela V.4.8. Estas tabelas apresentam o resultado tanto no domínio temporal quanto tempo-freqüencial a respeito de qual agente se encontra mais próximo de qual atrator.

A Tabela V.4.4 e a Tabela V.4.6 mostram qual som-neural (som agente) se aproximou de qual som-estímulo (som atrator), respectivamente, nos domínios temporal e tempo-freqüencial, enquanto que a Tabela V.4.7 apresenta a mesma análise em termos de similaridades perceptivas. É importante, neste ponto, lembrar que as relações de vizinhança da rede são expressas pelos índices dos sons-neurais. Isto é, o som-neural 2 é vizinho do som-neural 1 e do som-neural 3. Deve-se notar que os sons-neurais representam variantes dos sons-estímulo que estão representando porque a distância entre eles não é zero, ou seja, são formas de onda diferentes. Deve-se chamar a atenção também para o fato de que, quando sons-neurais vizinhos apresentam distância relativa pequena para o mesmo som-estímulo, isto significa que os sons-neurais estão representando variações desse. No entanto, é mais importante notar que existem diferenças de similaridade (ou de distância) entre estas representações. Isto é, quando

temos um ou mais sons-neurais representando um som-estímulo com grande similaridade, seu vizinho apresenta similaridade média, e o próximo vizinho apresenta uma grande similaridade com outro estímulo, o som-neural do meio, que não está representando nenhum som-estímulo em particular. O mesmo só foi posicionado devido às relações de vizinhança e características do aprendizado da rede, e assim apresenta características cruzadas dos sons-estímulo representados por seus dois vizinhos imediatos. Pode-se exemplificar este caso com os sons agentes (sons-neurais) de 1 a 6 consultando-se a Tabela V.4.3. Os sons agentes de 1 a 3 representam o som atrator 3, enquanto que os de 4 a 6 representam o som atrator 2. Todos estes são vizinhos de acordo com a relação de vizinhança em anel. A similaridade do som agente 1 com o som atrator 3 é menor que a dos agentes 2 e 3. O som agente 1 é vizinho do som agente 14, que está representando o som atrator 1. Devido à diferença de similaridade entre os sons agentes 1 a 3, e pelo fato de o som agente 1 estar entre os sons agentes 2 e 14, que representam diferentes atratores, pode-se concluir que o som agente 1 possui características cruzadas dos sons atratores 3 e 1. Verificando-se a avaliação perceptiva na Tabela V.4.7, é possível verificar que, apesar da similaridade do som agente 1 com o som atrator 1 ter média 3, o erro associado é bastante grande, permitindo-se concluir que existe uma similaridade perceptiva do som agente 1 com os sons atratores 1 e 3. O mesmo vale para os sons agentes 3 e 4, que apresentam uma similaridade menor com os sons atratores 3 e 2, respectivamente, e são vizinhos imediatos dos sons agentes 5 e 2, que apresentam alta similaridade com os sons atratores 3 e 2 que estão representando. Também é essencial notar que estes cruzamentos são difíceis de detectar visualmente. A avaliação perceptiva é crucial por se tratar de sons.

Como já foi destacado anteriormente, todos os sons agentes (sons-neurais) nesta aplicação são vizinhos de dois outros sons agentes (sons-neurais), de acordo com os rótulos numéricos deles. Essa relação de vizinhança possibilita proximidade no espaço vetorial, permitindo que dois ou mais sons neste espaço compartilhem características temporais, freqüenciais ou timbrais. A relação de vizinhança em anel, no entanto, sugere uma ordenação a priori dos sons agentes (sons-neurais). Dessa ordenação podem advir aplicações musicais interessantes quando se ouvem os sons resultantes seguindo-a. Fenômenos perceptivos como mascaramento, ou outros equivalentes aos descritos na seção II.2.iii), devido a cruzamentos perceptivos em dimensões timbrais, por exemplo, poderiam advir dessa ordenação intrínseca ao método de síntese sonora por MAO.

Por fim, é importante ressaltar que ambas as métricas matemáticas resultaram na mesma análise da avaliação subjetiva, ou seja, as tabelas retratam o mesmo cenário. Analisando-se a Tabela V.4.8 e cruzando-se seu resultado com a Tabela V.4.4 e a Tabela V.4.6, pode-se inferir que ambas as métricas são bastante capazes de capturar a essência perceptiva dos resultados quando se trata de variantes dos sons atratores (sons-estímulo). Neste caso, perceptivamente, os cruzamentos de características apontados pelos indivíduos se revelaram em ambas as métricas. Estes resultados foram publicados por Caetano et al. [8] e comparados com o agrupamento realizado através do MAO bi-dimensional, que não foi capaz de capturar as mesmas relações.

Em relação às características de um método de síntese sonora levadas em consideração, pode-se afirmar que o método apresenta uma controlabilidade alta. Isto é, os resultados apresentam características presentes nos sons atratores, tornando bastante fácil prever qual será o resultado em termos perceptivos da variação dos parâmetros do programa. Em especial, é possível obter resultados diferentes através da alteração de apenas dois parâmetros: a taxa de aprendizado inicial e sua constante de decaimento τ_2 . O método apresenta uma flexibilidade alta. Basta alterar os sons atratores que suas características são passadas aos resultados. Deve-se notar que uma característica fundamentalmente vantajosa é a adaptabilidade intrínseca ao método. A auto-organização faz com que os resultados apresentem variações cada vez que o método é utilizado. Também é possível atingir variações quando mais de um som-neural (som agente) representa o mesmo som-estímulo (som atrator). Uma vez que o usuário definiu os sons atratores (sons-estímulo) e o número de sons agentes (sons-neurais) que realizarão a busca, a emergência de padrões sonoros complexos advindos da estratégia de aprendizado não supervisionado está intrinsecamente ligada à subjetividade do usuário, responsável pela definição de parâmetros que afetam fundamentalmente os resultados. Isto significa que é possível obter resultados perceptivos desejados através da definição dos atratores apropriados.

Por fim, deve-se chamar a atenção à simplicidade do método proposto. O algoritmo é de fácil utilização, bastando especificar certos parâmetros. Uma vez implementado, o algoritmo apresenta muitas vantagens em relação aos métodos tradicionais, como robustez, preservação topológica, estratégia populacional que permite que mais de um som seja obtido em uma única rodada, capacidade de sintetizar qualquer tipo de som, harmônico ou inarmônico, sem nenhum esforço adicional devido à estratégia de busca

vetorial no domínio do tempo. Isto é, o resultado final já é a seqüência de amostras sonoras que pode ser convertida em som sem necessidade de processamento.

Ainda assim, verificou-se a necessidade de simplificação, especialmente no que tange o número de parâmetros, alguns pouco intuitivos. Há a necessidade de estudos adicionais voltados à simplificação do método, tal como desenvolver uma interface gráfica que possibilite ao usuário ter contato mais intuitivo com as variáveis relevantes e desenvolver suas idéias de forma criativa sem, no entanto, a necessidade de um conhecimento profundo de acústica, psicoacústica ou técnicas de programação ou quaisquer ferramentas de engenharia. Os resultados preliminares indicaram um alto grau de flexibilidade e robustez do método, capaz de gerar variantes das formas de onda de entrada, bem como híbridos com características claramente advindas de mais de um estímulo de entrada. Pode-se estender a aplicação para esta proposta seguindo-se a trajetória de cada som-neural durante o aprendizado dos sons-estímulo para revelar a transformação sonora em todas as etapas do aprendizado da rede.

VI) CONCLUSÃO

A alta flexibilidade permitida pela utilização do computador digital na síntese de som acarreta na necessidade de se utilizar uma técnica que torne possível explorar essa grande flexibilidade fornecendo controlabilidade, adaptabilidade e simplicidade. Os métodos tradicionais de síntese sonora não foram desenvolvidos especificamente para o computador digital. Portanto, não lançam mão de todos os recursos disponíveis de processamento de sinais digitais que o computador permite. Em especial, alguns deles foram propostos ainda para utilizar circuitos analógicos ou técnicas limitadas de processamento de sinais, como filtragem, devido às limitações dos recursos da época. Em última instância, os métodos tradicionais podem ser implementados com circuitos analógicos, enquanto que o método apresentado aqui não permite isso.

Hoje em dia, com o crescimento exponencial da capacidade de processamento e armazenagem de informação do computador digital, não se está mais limitado por ele, ao contrário, está-se limitado pela infinidade de possíveis resultados que se podem atingir. Em outras palavras, pode-se obter tantos sons diferentes por meio do computador digital que surgiu a necessidade de métodos para guiar o usuário na busca pelo resultado. Uma grande vantagem da utilização do computador digital na síntese e processamento de sons é a fácil obtenção de resultados com características dinâmicas e complexidade suficiente para que não sejam rejeitados pelo ouvido humano. Ainda assim, é necessário achar uma maneira de especificar essas características desejadas e, principalmente, utilizar um processo que permita atingir essa complexidade inerente dos sons musicais como resultado da aplicação desse processo.

Este trabalho propõe uma técnica de síntese sonora digital que utiliza uma estratégia populacional na obtenção de sons com qualidades definidas *a priori* pelo usuário. Essas qualidades são especificadas por um conjunto de atratores, formas de onda escolhidas pelo usuário para caracterizar uma certa região do espaço de busca. O espaço de busca dos agentes é o espaço de todas as formas de onda passíveis de representação pela codificação utilizada, interpretadas como vetores de alta dimensionalidade. Os agentes realizam a busca no espaço definido pelos atratores de acordo com estratégias auto-organizáveis. As características relevantes são a realimentação positiva e negativa e as interações locais dos agentes do sistema, levando

à emergência de padrões complexos e não especificados pelo usuário. Isto é, os sons resultantes emergem da interação dos agentes, apresentando características advindas dos atratores e representando variações desses atratores. Foram utilizados dois algoritmos bio-inspirados com características auto-organizáveis na implementação do método proposto, a saber, um sistema imunológico artificial e uma rede neural artificial, mais especificamente, o mapa auto-organizável de Kohonen.

Foi também proposta uma métrica de distância entre os sons que leva em conta tanto características temporais quanto frequenciais. A métrica consiste basicamente em se medir a diferença entre os espectros dinâmicos de duas formas de onda. A mesma apresentou resultados consistentes com a avaliação perceptiva de similaridade por indivíduos que realizaram uma análise subjetiva dos resultados.

Os resultados preliminares indicaram um alto grau de flexibilidade e robustez do método, capaz de gerar variantes das formas de onda de entrada, bem como híbridos com características claramente advindas de mais de um atrator. Basta alterar os atratores que suas características são passadas aos resultados. Também é importante ressaltar a alta controlabilidade do método. É bastante alta a previsibilidade das características dos resultados. Novamente, é possível prever com facilidade qual será o efeito da variação dos parâmetros de ambas as implementações do método.

O método apresenta adaptabilidade intrínseca. A auto-organização faz com que os resultados apresentem variações cada vez que o método é utilizado. Também é possível atingir variações quando mais de um agente representa o mesmo atrator. Por fim, é um método de simplicidade relativa. Uma vez implementados, os algoritmos são de fácil utilização, bastando especificar certos parâmetros. O algoritmo de sistemas imunológicos artificiais utilizado (aiNet), no entanto, poderia ser simplificado e ainda assim manter as características relevantes. A implementação utilizada tem muitos parâmetros, alguns pouco intuitivos. Já existem versões mais parcimoniosas deste algoritmo [4].

O corpo e o escopo da dissertação foram vinculados a uma abordagem interdisciplinar para que se pudesse colocar a dimensão do problema estudado. Foram apresentados aspectos históricos e conceituais que ampliam a discussão sobre a natureza dos sinais digitais utilizados para aplicações musicais. Neste sentido, o contexto e a visão do domínio das possíveis aplicações foi ampliado, pois houve uma preocupação de apontar para direções que entrelaçam a discussão de processos musicais tradicionais com visões contemporâneas.

O trabalho interagiu com a difícil tarefa de conciliar questões de ordem qualitativa com modelos mais pragmáticos do processo de produção de sons com características sonoras desejáveis. O método, ao utilizar um conjunto de sons condicionantes (atratores) como mecanismo de indiciamento do processo de busca por um “som ideal”, tenta conciliar a visão estética do artista sônico no nível material onde ele atua: o som. Especialmente quando o material gerado no processo de síntese é aplicado em composição musical ou na produção de obras audiovisuais contemporâneas. Desta forma, em oposição aos métodos já conhecidos de síntese sonora, onde é necessário realizar um mapeamento entre o domínio perceptivo do músico para um conjunto de parâmetros, o método pesquisado apresenta uma solução original: utilizar amostras sonoras como parâmetro de controle. Esta estratégia inovadora é associada a métodos bio-inspirados que trazem no seu escopo a capacidade desejável de se trabalhar com a diversidade que o processo de síntese sonora digital gera.

Finalmente, a dissertação explora a noção de controle através de estruturas vetoriais, utilizando-se de um espaço de grande de dimensão, pois cada som é um vetor no espaço de soluções. Neste ponto, ao trabalhar com a alta dimensionalidade do espaço no sentido de varrer todas as possíveis variações sonoras de um sinal sonoro digitalizado, o trabalho mostra a eficiência de dois métodos bio-inspirados nessa tarefa. Compara os resultados com um processo de avaliação que incorpora variáveis psicoacústicas, dado que seria necessário verificar se o método se mostrou eficiente no domínio da escuta. Os resultados de ambos os processos bio-inspirados mostraram-se capazes e eficientes neste processo avaliatório.

Diferentemente dos métodos tradicionais de síntese sonora, a proposta permite ao usuário incorporar sua subjetividade ao possibilitar que se obtenham resultados que contenham características perceptivas presentes nos sons atratores escolhidos. Os atratores determinam as qualidades sonoras presentes nos resultados sem a necessidade de uma descrição exaustiva ou simplificação excessiva decorrente da parametrização das qualidades sonoras que sacrifique a complexidade sonora dos resultados; especialmente no que tange percepção timbral, para a qual não se possui um consenso nem na definição nem na modelagem. A abordagem vetorial possibilita a obtenção de, virtualmente, qualquer som - harmônico ou inarmônico, simples ou complexo, “concreto” ou “eletrônico”, “artificial” ou “natural” – sem estar restrito a um modelo matemático parametrizado do som musical. Através da utilização da própria forma de onda e das suas

qualidades perceptivas é possível obter os resultados desejados, sendo a complexidade emergente intrínseca ao método e dependente da escolha dos atratores. Pretende-se, com isso, possibilitar ao usuário a obtenção de sons com características perceptivas definidas *a priori* pelo próprio usuário; seria possível obter o som correspondente, por exemplo, à percepção de um “trompete agüado” ou uma voz de “soprano chuvosa” através da escolha dos atratores convenientes e do algoritmo bio-inspirado mais apropriado.

Deve-se enfatizar uma diferença importante nos resultados obtidos com cada algoritmo bio-inspirado. A aiNet foi originalmente proposta como uma ferramenta de agrupamento de dados através do mapeamento da região do espaço ocupada pelos dados. O mapa auto-organizável unidimensional, originalmente utilizado para propor soluções para o problema do caixeiro-viajante, isto é, o caminho mínimo passando por um conjunto de cidades, foi adaptado, substituindo-se as cidades por sons. Isso gera a tendência de se obter resultados sonoros diferentes dependendo do método utilizado. A aiNet tende a gerar variantes sonoras decorrentes do mapeamento da região de interesse do espaço, as relações de vizinhança do MAO são o principal mecanismo responsável pelas características encontradas nos resultados. Isso significa que diferentes algoritmos bio-inspirados exploram diferentes características do problema e realizam a busca de maneiras distintas, resultando em diferentes propostas que solucionam o problema de modos diversos.

Construiu-se nessa conjunção, síntese sonora digital e métodos bio-inspirados, o ganho conceitual e tecnológico que essa dissertação representa. O posicionamento desse trabalho quanto à necessidade de métodos que gerem sons com as características desejadas é tratar o problema como uma busca no espaço vetorial representado pelos sons atratores. Nesse sentido, as características de aprendizado, manutenção de diversidade e emergência de complexidade dos algoritmos bio-inspirados foi essencial na obtenção de resultados que atenderam as exigências impostas pela natureza subjetiva e perceptiva do problema.

Por fim, as perspectivas futuras são muitas e relacionadas a diferentes aspectos da proposta. Primeiramente, a utilização de outras vizinhanças no MAO, tal como vizinhança bi-dimensional em grade, pode subsidiar uma gama nova de resultados com características distintas das apresentadas aqui. Pensa-se que é importante aprofundar a pesquisa a respeito do espaço vetorial gerado e das suas características sonoras. A

utilização de outras métricas de distância não apenas na avaliação dos resultados, mas no processo de busca em si poderia gerar resultados com características relevantes para o domínio do problema e, principalmente, na esfera das aplicações musicais. O mesmo pode ser dito a respeito da utilização de outros algoritmos bio-inspirados, especialmente os baseados no paradigma da inteligência coletiva. Algoritmos de formiga como *ant colony optimization* (ACO) e enxames de partículas como *particle swarm optimization* (PSO) podem gerar resultados com grande potencial para aplicações musicais. A utilização de uma interface gráfica com o usuário, bem como permitir que esse controle parâmetros musicais tradicionais como frequência, intensidade, e duração dos sons, deve facilitar a utilização criativa do método proposto e a busca estética sonora. Por outro lado, transformações sonoras podem ser atingidas como trajetórias no espaço contínuo. Seguir a trajetória de um ou mais agentes ao longo das iterações até a convergência equivale a obter uma transformação sonora gradual do som utilizado para inicializar o agente (por exemplo, ruído branco) no som resultante. Pode-se pensar na utilização do método em apresentações ao vivo onde o músico controla certos parâmetros dos métodos de acordo com critérios estéticos musicais e utiliza atratores dinâmicos, isto é, os atratores são re-posicionados no espaço ou sofrem uma transformação sonora equivalente a trajetórias que redirecionam a busca pelos agentes.

VII) REFERÊNCIAS BIBLIOGRÁFICAS

- [1] American National Standards Institute (1960). USA Standard Acoustical Terminology (Including Mechanical Shock and Vibration) SI.1-1960 (R1976). New York: American National Standards Institute.
- [2] Backus, J. The Acoustical Foundations of Music, 2nd Ed. New York: Norton, 1977.
- [3] Bastiaans M. Gabor's Expansion of a signal into Gaussian elementary signals: Proceedings of the IEEE. 68: 538-539, 1980.
- [4] Bezerra, G.B., Barra, T.V., de Castro, L.N., Von Zuben, F.J. (2005), Adaptive Radius Immune Algorithm for Data Clustering, C. Jacob, M. L. Pilat, P. J. Bentley, J. Timmis, eds., ICARIS 2005: 4th International Conference on Artificial Immune Systems, Lecture Notes in Computer Science, Banff, Canada. Springer.
- [5] Boston, T. R. 1989. "Clara Rockmore: The art of the Thèremen", Computer Music Journal 13 (1) p.61-63.
- [6] Bregman, A.S. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound. Cambridge, Massachusetts: MIT Press.
- [7] Burnet, F.M. The Clonal Selection Theory of Acquired Immunity, Cambridge University Press, 1959.
- [8] Caetano, M., Manzolli, J. and Von Zuben, F. J. (2005 c) "Topological Self-Organizing Timbre Design Methodology Using a Kohonen's Neural Network". 10 Simpósio Brasileiro de Computação e Música, Belo Horizonte, Brazil.
- [9] Caetano, M., Manzolli, J., Von Zuben, F. J. Application of an Artificial Immune System in a Compositional Timbre Design Technique. Proceedings of ICARIS 2005, Alberta, Canada. In Press, 2005.
- [10] Caetano, M., Manzolli, J., Von Zuben, F. J. Interactive Control of Evolution Applied to Sound Synthesis. Proceedings of the 18th International Florida Artificial Intelligence Research Society (FLAIRS), Clearwater Beach, EUA, 2005.

- [11] Camazine, S., Deneubourg, J.-L., Franks, N. R., Sneyd, J., Theraulaz, G., Bonabeau, E. (2001) *Self-Organization in Biological Systems*. Princeton University Press.
- [12] Chowning, J. The Synthesis of Complex Audio Spectra by Means of Frequency Modulation. *Journal of the Audio Engineering Society*, 21 (7), 526-534, 1973.
- [13] Corne, D. Dorigo, M. and Glover, F. *New Ideas In Optimization*, McGraw Hill, 1999. ISBN 007 709506 5.
- [14] Dasgupta, D. *Artificial Immune Systems and their Applications*, Springer-Verlag. (ed.), 1999.
- [15] Dawkins, R. 1986 *The Blind Watchmaker*, Penguin Books.
- [16] de Castro, L. N. *Fundamentals of Natural Computing: Basic Concepts, Algorithms, and Applications*, CRC Press LLC, 2006.
- [17] de Castro, L. N.; Timmis, J. "Artificial Immune Systems: A New Computational Intelligence Paradigm". Springer-Verlag, 2002.
- [18] de Castro, L.N and Von Zuben, F. aiNet: An Artificial Immune Network for Data Analysis, in *Data Mining: A Heuristic Approach*. Abbas, H, Sarker, R and Newton, C (Eds). Idea Group Publishing, 2001.
- [19] de Castro, L.N and Von Zuben, *Recent Developments in Biologically Inspired Computing*, Idea Group Publishing, 2004.
- [20] Dodge, C. Jerse, T. A. *Computer Music: Synthesis, composition and performance*. Schirmer Books, Macmillan, New York, 1985. ISBN 0-02-873100-X.
- [21] Erickson; R. (1975). *Sound Structure in Music*: Berkeley: University of California Press.
- [22] Fletcher, H. (1934). Loudness, pitch and the timbre of musical tones and their relation to the intensity, the frequency and the overtone structure. *Journal of the Acoustical Society of America* 6, 59-69.
- [23] Fletcher, H., and W. A. Munson 1933. "Loudness, its Definition, Measurement and Calculation." *Journal of Acoustical Society of America*.
- [24] Forrest, S., Perelson A. S., Allen, L. and Cherukuri, R. Self-Nonself Discrimination in a Computer, *Proceedings of the 1994 IEEE Symposium on*

- Research in Security and Privacy, Los Alamitos, CA , IEEE Computer Society Press pp.202-212, 1994,
- [25] Forster, M. Key Concepts in Model Selection: Performance and Generalizability. *Journal of Mathematical Psychology* 44, 205-231, 2000.
 - [26] Gabor, D. Acoustical Quanta and the Theory of Hearing. *Nature* 159 (1044), 591-594, 1947.
 - [27] Gabrielson, A. "Music Psychology – a Survey of Problems and Current Research Activities". *Proceedings of the Royal Swedish Academy of Music*, 1981.
 - [28] Glinski, A., 1992. "The Thèremin and the Emergence of Electronic Music". *Dissertação de doutorado*, New York University, USA.
 - [29] Grey, J. M. An Exploration of Musical Timbre, *Doctoral dissertation*, Stanford Univ, 1975.
 - [30] Grey, J. M., and Moorer, J. A., "Perceptual Evaluations of Synthesized Musical Instrument Tones". *Journ. Ac. Soc. Am.*, 62, 2, pp 454-462, 1977.
 - [31] Griffiths, P. A Guide to Electronic Music. London: Thames and Hudson, 1979.
 - [32] Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harshberger, M. L. (Eds.) 1997. *Methodological Issues in Timbre Research*. Psychology Press, London.
 - [33] Handel, S. (1995). Timbre perception and auditory object identification, *Hearing*, B.C.J. Moore (Ed.). New York, NY: Academic Press.
 - [34] Haykin, S., *Redes Neurais: Princípios e Práticas*, segunda edição, Bookman, 2001.
 - [35] Holland, John H. *Adaptation in Natural and Artificial Systems*. Ann Arbor University of Michigan Press, 1975.
 - [36] Janeway Jr, C. A. The Immune System Evolved to Discriminate Infectious Self from Noninfectious Self. *Immunology Today*, 13 (1), pp. 11-16, 1992.
 - [37] Janeway Jr, C. A., Travers, P., Walport, M. and Capra, J. D. *Immunobiology: The immune system in Health and Disease*, 4th Ed., Garland publishing, 1999.
 - [38] Jerne, N. K. "Towards a Network Theory of the Immune System", *Ann. Immunol. (Inst. Pasteur)*, pp. 373-389, 1974.

- [39] Kohonen, T. "Self-Organization And Associative Memory", 3rd. Edition, Springer, 1989 (1st. Edition, 1984).
- [40] Kohonen, T. "Self-Organized Formation Of Topologically Correct Feature Maps", *Biological Cybernetics*, 43:59-69, 1982.
- [41] Kohonen, T. "Self-Organizing Maps", 2nd. Edition, Springer, 1997.
- [42] Kohonen, T. "The Self-Organizing Map", *Proceedings Of The Ieee*, 78:1464-1480, 1990.
- [43] Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In: *Structure and perception of electroacoustic sound and music*. New York: Excerpta Medica, 43-54.
- [44] Kruskal, JB, and Wish, M., *Multidimensional Scaling*, Sage University Paper series on Quantitative Application in the Social Sciences, Beverly Hills. CA, 1978.
- [45] LeBrun, M. Digital Waveshaping Synthesis. *Journal of the Audio Engineering Society*, 27(4), 250-266, 1979.
- [46] Levarie, Sigmund, and E. Levy, *Tone- A Study in Musical Acoustics*. Kent, Ohio: Kent State University Press, 1968.
- [47] Lewin, R. *Complexity - Life at the Edge of Chaos*, Macmillan: New York, 1992.
- [48] Licklider, J. C. R. (1951). "A duplex theory of pitch perception." *Experientia* 7: 128-133.
- [49] Lo, Z.-P., Yu, Y., Bavarian, B., "Analysis of the convergence properties of topology preserving neural networks." *IEEE Transactions on Neural Networks*, vol. 4, pp.207-220, 1993.
- [50] Manning, P. *Electronic and Computer Music*. Clarendon Press, Oxford, 1987. ISBN 0-19-311918-8.
- [51] Manzolli, J, A. Maia Jr., J.E. Fornari & F. Damiani, "The Evolutionary Sound Synthesis Method", 2001.
- [52] Manzolli, J, A. Maia Jr., J.E. Fornari & F. Damiani, "Waveform Synthesis using Evolutionary Computation", *Anais do VII Simpósio Brasileiro de Computação e Música*, Recife, UFPB. 2001.

- [53] Mathews, M. and Pierce, J. "Harmonic and Non-Harmonic Partial". IRCAM Rapports, 28, 1980.
- [54] Matzinger, P. The Danger Model: A renewed Sense of Self. *Science*, 296, pp. 301-305, 2002.
- [55] Moore, F. R., "Elements of Computer Music": Univ. of California, San Diego, Prentice Hall, Englewood Cliffs, New Jersey 07632, ISBN 0-13-252552-6.
- [56] Moorer, J. A. and Grey, J. M. Lexicon of Analyzed Tones (Part 1: A Violin Tone). *Computer Music Journal*, 1 (2), 1977, pp. 39-45.
- [57] Moorer, J. A. and Grey, J. M. Lexicon of Analyzed Tones (Part 2: Clarinet and Oboe Tones). *Computer Music Journal*, 1 (3), 1977, pp. 12-29.
- [58] Moorer, J. A. and Grey, J. M. Lexicon of Analyzed Tones (Part 3: The Trumpet). *Computer Music Journal*, 2 (2), 1977, pp. 23-31.
- [59] Nossal, G.J.V. The Molecular and Cellular Basis of Affinity Maturation in the Antibody Response, *Cell*, 68, pp. 1-2, 1993.
- [60] Perelson, A.S. & Oster, G.F. Theoretical Studies of Clonal Selection: Minimal Antibody Repertoire Size and Reliability of Self-Nonself Discrimination. *J. Theoret. Biol.*, vol. 81, pp. 645-670, 1979.
- [61] Plomp, R.: Aspects of tone sensation. London: Academic Press, 1976.
- [62] Rasch, R.A. & Plomp, R.: The Listener and the Acoustic Environment. In Deutsch, D. (Hrsg.): *The psychology of music*. London: Academic Press, 1983.
- [63] Risset, J. C. An introductory Catalogue of Computer Synthesized Sounds. Murray-Hill, N. J.: Bell Telephone Laboratories, 1969.
- [64] Risset, J. C. Computer Study of Trumpet Tones. Murray Hill, N.J.: Bell Telephone Laboratories, 1966.
- [65] Ritter, H., Martinetz, T., and Schulten, K., *Neural Computation and Self-Organizing Maps: An Introduction*. Addison-Wesley, Reading, MA. 1992.
- [66] Roads, C. Introduction to Granular Synthesis. *Computer Music Journal*, 12(2) Summer 1988.
- [67] Roederer, J.G. (1975). *Introduction to the physics and psychophysics of music*. New York: Springer-Verlag.

- [68] Schaeffer, P. 1966. *Traité des objects musicaux*. Paris: Editions du Seuil.
- [69] Schouten, J. F. The perception of timbre. In: Report of the 6th International Congress on Acoustics. Tokio, 1968.
- [70] Seashore, Carl E. *Psychology of Music*, McGraw-Hill, 1938.
- [71] Slawson, W. (1985). *Sound Color*. Berkeley: University of California Press.
- [72] Smalley, D. "Spectro-morphology and Structuring Processes". In *The Language of Electroacoustic Music*, 61-93. London: Macmillan. 1990.
- [73] Tizard, I. R. *Immunology: An Introduction*, 4th Ed., Saunders College Publishing, 1995.
- [74] Truax, B. Real-Time Granular Synthesis with a Digital Signal Processor. *Computer music Journal*, 12(2), Summer, 1988.
- [75] Varela, F. J. and Coutinho, A. Second Generation Immune Networks. *Immunology Today*, vol. 12 (5), 159-166, 1991.
- [76] Von Foerster, H. , "On Self-Organizing Systems and Their Environments." In: *Self-Organizing Systems*, M. C. Yovits und S. Cameron (Hg.), Pergamon Press, London, pp. 31–50, 1960.
- [77] Von Helmholtz, H. *On the Sensations of Tone*. London, Longman, 1885.
- [78] Wishart, T. "On Sonic Art". Simon Emerson: Harwood Academic Publishers, ISBN 3-7186-5847-X, 1998.
- [79] Xenakis, I. *Formalized Music*. Bloomington/London: Indiana University Press, 1970. ISBN 0-253-32378-9.