UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO

DEPARTAMENTO DE ENGENHARIA DE COMPUTAÇÃO E

AUTOMAÇÃO INDUSTRIAL

CONTROLE DE SÍNTESE SONORA POR ANALOGIA ACÚSTICA E SEMÂNTICA APLICANDO COMPUTAÇÃO BIO-INSPIRADA

César Rennó Costa

Orientador: Fernando José Von Zuben
(DCA - FEEC - UNICAMP)

Co-orientador: Jônatas Manzolli
(NICS - 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.

Campinas - São Paulo - Brasil

Dezembro de 2007

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

Rennó Costa, César
C823c Controle de Síntese Sonora por Analogia Acústica
e Semântica aplicando Computação Bio-Inspirada
César Rennó Costa. – Campinas, SP: [s.n.], 2007.

Orientadores: Fernando José Von Zuben, Jônatas Manzolli.
Dissertação de Mestrado - Universidade Estadual de Campinas,
Faculdade de Engenharia Elétrica e de Computação.

1. Processamento de sinais. 2. Ciência cognitiva
3. Redes neurais (Computação). 4. Computação evolutiva.
5. Música por computador. 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:	Control of Sound Syhtesis by Acoustic and Semantic Analogy using Bio-Inspired Computing
Palavras-chave em Inglês:	Signal processing, Cognitive science, Neural networks (Computer), Evolutionary computation, Computer music
Área de concentração:	Engenharia de Computação
Titulação:	Mestre em Engenharia Elétrica
Banca Examinadora:	Amauri Lopez, Artemis Moroni, e Mauricio Ferreira Magalhães
Data da defesa:	17/12/2007
Programa de Pós Graduação:	Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE MESTRADO

Candidato: Cesár Rennó Costa

Data da Defesa: 17 de dezembro de 2007

Título da Tese: "Controle de Síntese Sonora por Analogia Acústica e Semântica Aplicando Computação Bio-Inspirada"

Prof. Dr. Fernando José Von Zuben (Matr. 263958): Fernando José Von Zuben

Dra. Artemis Maria Francelin Sanchez Moroni:

Prof. Dr. Rafael Santos Mendes:

Prof. Dr. Amauri Lopes:

Secretário(a) Edson Vicente Sanches Filho: Edson V. Sanches Filho

Coordenador de PG Prof. Dr. Michel Daoud Yacoub:

Resumo

Este trabalho sugere novos paradigmas de controle de mecanismos de síntese sonora. Utilizando conceitos das ciências cognitivas, o processo gerativo é modelado como um sistema de conversões entre representações, da atuação subjetiva do usuário, passando pela descritiva e culminando no material sonoro. A partir do estudo da analogia descritiva, engendra-se a analogia acústica, representação por amostras sonoras, e a analogia semântica, representação por linguagem. Aplicadas à arquitetura modelada, essas analogias permitem que o processo de síntese sonora tenha um caráter mais intuitivo. São apresentadas duas implementações práticas, sendo que técnicas de computação bio-inspirada fornecem o maquinário computacional para a realização do mapeamento entre representações e controle do processo de síntese.

Palavras-chave: Síntese Sonora, Representação por Analogia, Computação Bio-Inspirada.

Abstract

This work suggests novel control paradigms of sound synthesis mechanisms. Applying cognitive science concepts, the generative process is modeled as a system of conversions throughout representations: from user's insight, through descriptive, to the sound material. From descriptive analogy studies, the acoustic analogy (representation through sound) and the semantic analogy (representation through language) are engendered. Applied to the modeled architecture, these analogies allow the synthesis process to have a more intuitive nature. Two practical implementations are presented. Bio-inspired computing provides the computational machinery used to map different representations and to control the synthesis process.

Keywords: Sound Synthesis, Representation by Analogy, Bio-Inspired Computation.

à liberdade de dedicar a quem quer que seja
e à grata responsabilidade de dedicar à minha mãe.

O Artista Inconfessável

Fazer o que seja é inútil.
Não fazer nada é inútil.
Mas entre fazer e não fazer
mais vale o inútil do fazer.
Mas não, fazer para esquecer
que é inútil: nunca o esquecer.
Mas fazer o inútil sabendo
que ele é inútil, e bem sabendo
que é inútil e que seu sentido
não será sequer pressentido,
fazer: porque ele é mais difícil
do que não fazer, e dificil-
mente se poderá dizer
com mais desdém, ou então dizer
mais direto ao leitor Ninguém
que o feito o foi para ninguém.

João Cabral de Melo Neto

Agradecimentos

A auto-organização rege muitos dos caminhos trilhados pela natureza. Está presente na capacidade do simples controlar o caótico e fazer com que deste surja ordem, padrão e sentido. No núcleo de seu engenho, encontram-se os conceitos de diversidade, interação e regulação. É a regulação da entropia por meio da interação de elementos diversos. O resultado, na forma da ordem, padrão e sentido, apesar de parecer ser o que tinha que ser, atemporalmente, se torna evidente para nós num estralar infinitesimal: da inexistência à percepção de sempre ter sido verdade. É aquela idéia que surge, surpreende, e se torna óbvia. Remete-nos ao: "porque não pensei nisso antes?".

Ser criativo é ter novas idéias que sejam úteis (ou inúteis das que mais valem), mas na acepção de que útil é aquilo que nos faça perceber ordem, padrão e sentido. São assim criativos os processos auto-organizados ao fazê-las surgir. É, assim, criativa a natureza.

Vejo o trabalho que culminou nesta dissertação como um tanto auto-organizado, o que me permite dizer: natural e criativo. Seu desenvolvimento foi fortemente alimentado de diversidade, interação e regulação, mas sob uma estrita noção do que é ordem, padrão e sentido, ou no caso, objetivo. E atribuo a esta característica grande responsabilidade pelo seu sucesso.

Agradeço assim a todas as minhas fontes:

À diversidade de idéias que me foram apresentadas. De forma mais marcante, ao levantar da poeira, o endoidecer acadêmico que me permitiu meu co-orientador, Professor Doutor Jônatas Manzolli.

À interação com diversas áreas, possível pela pluralidade de direções de meus professores. Seja na FEEC, no NICS ou em toda a UNICAMP. Incluo aqui os membros da banca, Professor Doutor Rafael Mendes, Professor Doutor Amauri Lopes e Doutora Artemis Moroni. A estes, agradeço em especial a paciência com a extensão do texto, o envolvimento e os comentários.

À lucidez de meu orientador, professor Fernando Von Zuben, na incansável tarefa de regular estas idéias e guiar-me em sua organização e materialização.

À minha família em tudo. Diversa em carinho. Interativa por estar sempre presente. Reguladora, por não permitir me desviar e por ser meu porto seguro.

À diversidade de más idéias e falta de juízo de meus colegas na Coconut Square Garden, seus anexo(a)s, da -70^0 Lyrics e do grupo GATO. À interação destas idéias com Minas Gerais, só reguladas pela paciência e amabilidade de tiiinha e maninha.

Ao clã Furlanete e aos nossos momentos de Café, Tabaco e Música Esquisita.

Ao coach Fábio Furlanete, pela co-autoria do Capítulo 6 e co-desenvolvimento do Acusmata e todos os esclarecimentos e questionamentos.

Ao amigo André Luvizotto, porque louco por louco ele também é.

À FAPESP, pelo financiamento, por acreditar na diversidade e por incentivar o desenvolvimento do Brasil.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Sons

Os sons se encontram na mídia anexa, organizados em pastas relativas a cada capítulo e subpastas relativas aos experimentos individuais.

Capítulo 3 : Experimento 1	-	**\CAP3\EXP1**
5.9 (a) e 5.10	-	EXP1_REF_MMJB.wav
5.9 (b) e 5.11	-	EXP1_REF_LICK.wav
5.9 (c) e 5.12	-	EXP1_REF_FALA.wav
5.13 (a) e 5.14	-	EXP1_dFALA_sFALA.wav
5.13 (b) e 5.15	-	EXP1_dFALA_sLICK.wav
5.13 (c) e 5.16	-	EXP1_dFALA_sMMJB.wav
5.17 (a) e 5.18	-	EXP1_dLICK_sFALA.wav
5.17 (b) e 5.19	-	EXP1_dLICK_sLICK.wav
5.17 (c) e 5.20	-	EXP1_dLICK_sMMJB.wav
5.21 (a) e 5.22	-	EXP1_dMMJB_sFALA.wav
5.21 (b) e 5.23	-	EXP1_dMMJB_sLICK.wav
5.21 (c) e 5.24	-	EXP1_dMMJB_sMMJB.wav
5.25 (a) e 5.26	-	EXP1_dFALA_sALL.wav
5.25 (b) e 5.27	-	EXP1_dLICK_sALL.wav
5.25 (c) e 5.28	-	EXP1_dMMJB_sALL.wav
Capítulo 3 : Experimento 2	-	**\CAP3\EXP2**
5.29	-	EXP2_SENO_ENV.wav
5.30	-	EXP2_SENO_REF.wav
5.31	-	EXP2_SPITCH.wav
5.32	-	EXP2_CPITCH.wav
Capítulo 3 : Experimento 3	-	**\CAP3\EXP3**
5.33	-	EXP3_512.wav
5.34	-	EXP3_1024.wav
5.35	-	EXP3_2048.wav
5.36	-	EXP3_4096.wav
5.37	-	EXP3_8192.wav

Capítulo 3 : Experimento 4 - **\CAP3\EXP4**
5.38 (a) - EXP4_REF_CLAR_C4.wav
5.38 (b) - EXP4_SYNTH_C4.wav
5.39 (a) - EXP4_REF_CLAR_C5.wav
5.39 (b) - EXP4_SYNTH_C5.wav
5.40 (a) - EXP4_REF_CLAR_C6.wav
5.40 (b) - EXP4_SYNTH_C6.wav
5.41 (a) - EXP4_SYNTH_440Hz.wav
5.41 (b) - EXP4_SYNTH_1000Hz.wav
5.42 (a) e 5.44 (a) - EXP4_REF_LUCK.wav
5.42 (b) e 5.44 (b) - EXP4_SYNTH_C_LUCK.wav
5.44 (c) - EXP4_SYNTH_CV_LUCK.wav
5.43 (a) - EXP4_REF_CLAR_C6.wav
5.43 (b) - EXP4_SYNTH_C6.wav
5.43 (c) - EXP4_SYNTH_CV_C6.wav

Capítulo 4 : Experimento 1 - **\CAP4\EXPC**
6.8 (a) e 6.6 (a) - CLARINETA_C4.wav
6.8 (b) - EXPC_REF_SIN_C4.wav
6.8 (c) - EXPC_STH_SIN_C4.wav
6.9 (a) e 6.6 (b) - CLARINETA_C5.wav
6.9 (b) - EXPC_REF_SIN_C5.wav
6.9 (c) - EXPC_STH_SIN_C5.wav
6.10 (a) e 6.6 (c) - CLARINETA_C6.wav
6.10 (b) - EXPC_REF_SIN_C6.wav
6.10 (c) - EXPC_STH_SIN_C6.wav

Capítulo 4 : Experimento 2 - **\CAP4\EXPS**
6.14 - EXPS_CRESC_BRILHO.wav
6.15 - EXPS_CRESC_CALOR.wav
6.16 - EXPS_CRESC_RUIDESA.wav
6.17 - EXPS_CRESC_INHARM.wav
6.18 - EXPS_CRESC_DEFINI.wav

Amostras de Referência - **\REFERENCIA**

Glossário

altura	-	freqüência fundamental de um som.
ataque	-	porção inicial de um tom.
codec	-	representação codificada de áudio.
síntese sonora	-	processo de produção de um som.
envelope	-	contorno de energia de um sinal.
figura semântica	-	termo adjetivo de alguma qualidade subjetiva.
frequência fundamental	-	menor frequência de ressonância de um sinal.
grânulo, grão	-	amostra de som de curta duração.
harmônico	-	freqüência múltipla inteira da freqüência fundamental.
janela	-	perfil de energia.
janelamento	-	aplicação de uma janela sobre uma amostra sonora.
loudness	-	intensidade percebida de um som.
material sonoro	-	amostra sonora gerada por síntese sonora.
objetivo sonoro	-	expectativa sonora do músico.
pitch	-	altura percebida de um som harmônico.
som harmônico	-	sinal sonoro com presença de harmônicos.
sonoridade	-	ver Seção 2.1.
sustentação	-	porção estável de um tom.
timbre	-	ver Seção 2.1.
tom	-	som com aplicação musical.

Lista de Abreviações e Siglas

AG	-	Algoritmos Genéticos.
$aiNet$	-	Artificial Immune Network.
ASA	-	Associação Americana de Padrões (American Standards Association).
BMU	-	Best Matching Unit.
CE	-	Computação Evolutiva.
DCT	-	Discrete Cosine Transform (Transformada Discreta do Coseno).
DFT	-	Discrete Fourier Transform (Transformada Discreta de Fourier).
FFT	-	Fast Fourier Transform (Transformada Rápida de Fourier).
FM	-	Frequency Modulation (Modulação em Freqüência.
HPS	-	Harmonic Product Spectrum Algorithm.
ISO	-	International Organization for Standardization.
LFC	-	Linear Frequency Coefficients.
LPC	-	Linear Predictive Coding.
MAO	-	Mapas Auto-Organizáveis.
$MFCC$	-	Mel-Frequency Cepstral Coefficients.
MLP	-	MultiLayer Perceptron (Perceptron de Múltiplas Camadas).
PCM	-	Pulse Code Modulation (Modulação por Códigos de Pulso).
RNA	-	Redes Neurais Artificiais.
$SAMT$	-	Semantics And digital Media Technologies.
$SGBP$	-	Síntese Gerativa Baseada em População.
SIA	-	Sistemas Imunológicos Artificiais.
SMS	-	Spectral Modeling Synthesis.
$STFT$	-	Short-Time Fourier Transform (Transformada de Fourier de Curta Duração).
TSS	-	Trajetória Semântica Sonora.
$UNICAMP$	-	Universidade Estadual de Campinas.

Trabalhos Publicados Pelo Autor

1. C. R. COSTA, J. MANZOLLI & F. J. VON ZUBEN "Population-Based Generative Synthesis: A Real-Time Texture Synthesizer based on Real-World Sound Streams" In *Proceeding of IX Audio Engineer Society (AES) Brazil Congress*, São Paulo, Brazil. 2006.

2. C. R. COSTA, F. FURLANETE, J. MANZOLLI & F. J. VON ZUBEN "Using Sound Streams as a Control Paradigm for Texture Synthesis" In *Proceeding of XVI Congress of Associação Nacional de Pesquisa e Pós-Graduação em Música (ANPPOM)*, Brasilia, Brazil. 2006.

3. C. R. COSTA, F. FURLANETE, J. MANZOLLI & F. J. VON ZUBEN "Sound synthesis based on semantic descriptors" In *Proceedings of the 11th Brazilian Symposium on Computer Music (SBCM)*, São Paulo, Brazil. 2007.

Não relacionadas diretamente à dissertação:

1. M. CAETANO, C. R. COSTA, J. MANZOLLI & F. J. VON ZUBEN "Self-organizing topological timbral design methodology using a Kohonen neural network." In *Proceedings of the 10th Brazilian Symposium on Computer Music* (SBCM), Belo Horizonte, Brazil. 2005.

2. A. L. LUVIZOTTO, & C. R. COSTA "Context Sensitive Harmonic Processor" In *Proceeding of 11th Brazilian Symposium on Computer Music* (SBCM), 2007, São Paulo, Brazil. 2007.

3. A. L. LUVIZOTTO, & C. R. COSTA "Harmonic Equalizer" In *Proceeding of X Audio Engineer Society (AES) Brazil Congress*, São Paulo, Brazil. 2007.

Capítulo 1

Introdução

Em princípio, um computador pode produzir qualquer som.
Seu potencial só é limitado pela imaginação do compositor.

John R. Pierce

Esta frase de John R. Pierce (1992) exemplifica a exaltação inicial com o surgimento da computação musical. O primeiro som sintetizado por computador foi criado por Max Mathews e um grupo de pesquisadores, que incluía John R. Pierce, em 1957 no laboratório de Acústica e Psicologia da Bell AT&T em Murray Hill, New Jersey. Foi utilizado um programa chamado Music I, executado em um IBM 7040 (Meyers, 2002). Pioneiro na área, Mathews (1963) enfatizou o poder de síntese do computador: "Gerar sons a partir de números é uma forma completamente geral de sintetizar som, porque a banda e faixa dinâmica de audição são limitadas e, portanto, qualquer som que percebemos pode ser gerado desta forma"[1]. Pierce (1965) publicou: "Como um instrumento musical, o computador tem ilimitado potencial de gerar sons. Ele pode, de fato, produzir números representando qualquer som concebível ou audível. Coisas maravilhosas poderiam sair desta caixa se pelo menos nós soubéssemos como evocá-las."[2] referenciando o problema capital da produção

[1] "*...generating sounds from numbers is a completely general way to synthesize sound because the bandwidth and dynamic range of hearing are bounded and therefore any sound we perceive may be generated in this way.*"

[2] "*As a musical instrument, the computer has unlimited potentialities for uttering sound. It can, in fact, produce strings*

musical por computador: o paradigma de controle.

A teórica capacidade do computador de produzir qualquer som implica na necessidade de se lidar com um extenso espaço de possibilidades. Do ponto de vista do estudo aqui desenvolvido, percebemos que o otimismo inicial, retratado na idéia de Pierce (1992), sofreu um processo de evolução regido por dois vetores: numa direção, a busca por melhorias algorítmicas através de modelagem matemática e soluções de processamento de sinal; na outra direção, a comunidade de usuários, buscando, através de uma transgressão criativa, expandir os domínios e a qualidade sonora dos métodos de síntese. Portanto, o problema aqui estudado parte do princípio de que o que torna um mecanismo de síntese eficiente não é somente a capacidade de produzir qualquer som imaginável, mas a capacidade de permitir ao usuário, de forma consciente, explorar e interagir com este espaço para encontrar o material sonoro que satisfaça o seu imaginário.

A partir de Pierce e Mathews, a pesquisa histórica mostrou que a estrutura da representação sonora é baseada na própria arquitetura de geração do sinal, e cabe ao usuário aprender e adaptar-se a ela. Assim, existe uma grande distância entre o modelo utilizado no controle do processamento computacional e a visão perceptiva do compositor.

Em Teoria de Controle, a eficiência do sistema é medida pela minimização do erro entre a entrada e a resposta esperada. Na visão desenvolvida nesta pesquisa, o controle do processo de síntese sonora é eficiente quando há uma similaridade entre o resultado sonoro e a expectativa do usuário. Mas não é trivial representar de forma computacional e satisfatória a gama de variáveis que descrevem esse entorno. Quando um sistema envolve decisões que passam pelo viés da percepção do usuário, a entrada não pode ser totalmente descrita por um conjunto paramétrico fechado. Há necessidade de expandir-se a noção de representação do sistema, de forma a incluir variáveis que sejam suscetíveis a um espectro que envolve desde variações ambientais a, possivelmente, noções estéticas. Em geral, é importante que o desenvolvimento de tais sistemas de engenharia englobe a natureza humana dos processos envolvidos.

---

of numbers representing any conceivable or hearable sound. Wonderful things would come out of that box if only we knew how to evoke them."

Como já apresentado, o foco do estudo aqui reportado é avançar na direção de formas de representação para mecanismos de síntese que permitam ao usuário descrever os seus objetivos de forma amigável. Assim o usuário poderá, com auxílio do computador, explorar o espaço de sonoridades com maior facilidade, levando-o a um controle mais intuitivo e eficiente.

Para a realização deste ideal, foi exercida em pesquisa uma exploração interdisciplinar, envolvendo: computação musical, que estabelece o domínio do problema (i.e., síntese sonora); ciências cognitivas, que, através dos conceitos de representação, analogia e intencionalidade, fornecem suporte teórico para o entendimento da interação entre som e percepção, e estabelecimento de novos paradigmas de controle; e processamento digital de sinais e computação natural, que são a infra-estrutura computacional que permite a implementação dos dois modelos propostos.

Nesta dissertação, do Capítulo 2 ao 4, faz-se um levantamento temático e teórico das facetas que envolvem o problema aqui estudado. O Capítulo 2 discute o cerne do problema, ao introduzir os conceitos de sonoridade e timbre e realizar uma breve revisão histórica da importância desses na Teoria Musical e no surgimento das bases da computação musical, Seção 2.1, e apresenta o processo de síntese e de manipulação sonora pelo computador, Seção 2.2. O Capítulo 3 se refere ao processo de modelagem relacionado às técnicas de síntese. Intitulado "Representação Sonora e Analogia", introduz os modelos de representação utilizados em síntese sonora, Seção 3.1, e apresenta a aplicação do conceito de analogia aplicado no desenvolvimento dos paradigmas representacionais estudados nesta dissertação, Seção 3.2. O Capítulo 4 apresenta o conjunto de métodos computacionais que inspiraram e possibilitaram a aplicação prática dos paradigmas tratados. Apresenta-se uma gama de processos meta-heurísticos que envolve: redes neurais artificiais, computação evolutiva e sistemas imunológicos artificiais.

As representações aqui estudadas, que representam a contribuição original deste trabalho, são apresentadas nos Capítulo 5, "Representação por analogia acústica", e Capítulo 6, "Representação por analogia semântica".

Por fim, o Capítulo 7 conclui a dissertação apresentando uma síntese dos problemas estudados, a discussão de seu alcance e as projeções para extensões futuras da pesquisa.

Capítulo 2

Sonoridade e Síntese Sonora

2.1 Sonoridade e timbre

Os termos sonoridade e timbre dizem respeito a uma qualidade do som que lhe permite atribuir identidade. Timbre é definido pela ASA[1] como "o atributo da sensação nos termos pelo qual um ouvinte pode julgar que dois sons com mesmo *loudness* e *pitch* são dissimilares"[2]. Ou simplesmente "o atributo que permite ao ouvinte identificar o instrumento que produz o tom."[3] (Risset & Mathews, 1969). "É o correlato subjetivo de todas as propriedades do som que não influenciam diretamente *pitch* e *loudness*."[4] (Houtsma, 1997).

Schaeffer (1966) diferencia os dois termos de estudo que introduzem a questão principal deste trabalho. Ele define sonoridade como o atributo qualitativo e distintivo de um som, o que lhe garante identidade sob a ótica perceptiva do ouvinte: o que Schaeffer chama de "Escuta". Em contrapartida, na visão schaefferiana, timbre se refere às propriedades físicas do fenômeno sonoro, denominadas "propriedades materiais ou concretas". Na sua obra "*Traité des Objets Musicaux*", ele faz esta distinção, pois, enquanto sonoridade carrega a abstração do ouvinte, o termo timbre define

[1] American Standards Association (1960). *American Standard Acoustical Terminology*. New York. Definição 12.9, Timbre, p.45

[2] "*...that attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar.*"

[3] "*...the attribute that enables the listener to identify the instrument producing the tone.*"

[4] "*Timbre is the subjective correlate of all those sound properties that do no directly influence pitch or loudness.*"

de forma única a fonte sonora. Portanto, sonoridade para Schaeffer inclui e estende a qualificação que o ouvinte dá ao timbre ao tratar o som como uma entidade autônoma. O termo sonoridade possui várias acepções na literatura musical e em todo o escopo desta dissertação será utilizado o sentido apresentado por Schaeffer. Um dos paradigmas abordados nesta pesquisa ("Representação por analogia semântica", Capítulo 6), teve como objetivo criar um espaço exploratório da sonoridade via um conjunto de "figuras semânticas", definidas na Seção 6.1. Ao desenvolver este modelo, implementado no programa "Acusmata", buscou-se uma conexão com os atributos qualitativos que compõem a sonoridade, como destacou Schaeffer.

Os primeiro estudos da relação entre o sinal sonoro e a sensação causada são reportados no fim do século XIX, realizados por Hermann Helmholtz. Helmholtz (1877) criou modelos de ressonância e tentou descrever o timbre de instrumentos a partir dos componentes espectrais entendidos como relações de freqüências harmônicas de acordo com a Série de Fourier. Porém, dada a complexidade do sinal sonoro e a dificuldade de analisar o conteúdo espectral com maior precisão, modelos mais avançados só surgiram com o uso do computador, envolvendo o próprio Mathews (Risset & Mathews, 1969) e o laboratório Bell AT&T (Tenney, 1965). A criação de métodos de síntese seguiu paralela ao desenvolvimento de processos de análise. O trabalho de Risset & Wessel (1982) foi pioneiro ao demonstrar que é essencial a interação entre síntese e análise para que haja coerência no processo de validação do sistema de geração de sons.

O que este processo histórico mostrou é que, devido à sua capacidade de manipulação de dados, o computador pode atuar como uma ferramenta extremamente útil no tratamento de síntese sonora. Na música, a possibilidade de criar um som e trabalhar diretamente com a sonoridade é um avanço que amplia o vocabulário para além dos limites estabelecidos no contexto dos instrumentos musicais tradicionais. Anteriormente ao uso do computador, os instrumentos musicais eram tomados como suporte para o processo de composição timbrística. Esse suporte condicionava o resultado obtido pelo músico a um conjunto de timbres previamente definidos e a outros aspectos como harmonia, melodia e ritmo. A sonoridade e o timbre são elementos essenciais da música. Porém, até o início do século XX, foram considerados coadjuvantes no contexto da música ocidental. Apesar dos

grandes avanços apresentados nos tratados de orquestração, havia uma dificuldade essencial de manipular diretamente o conteúdo espectral. Por isso, o acesso a mecanismos que permitam uma livre exploração do espaço de sonoridades sempre foi um anseio dos músicos.

2.1.1 Contexto histórico da sonoridade na música

A música se desenvolveu calcada em um conjunto discreto de timbres de instrumentos acústicos. Conjunto este limitado pelas possibilidades mecânicas e cuja sintonia fina dependia de conhecimento técnico de luthieria de seus construtores e mantenedores. A dificuldade de manipulação da sonoridade a colocava como um aspecto secundário. No que diz respeito ao material sonoro, os avanços decorrentes do surgimento da computação musical vieram ao encontro da idéia, já disseminada no começo do Século XX, de que a música poderia se desenvolver a partir de uma base mais ampla que a Música Tonal. Esta postura teve suas bases em meados do Século XIX com as obras de Chopin, Wagner e Lizt, que ampliaram grandemente a exploração da sonoridade musical, principalmente do piano. Além deles, e de forma muito inventiva, a expansão do espaço de sonoridade se tornou possível pela mente criativa de compositores do início do século XX como Stravinsky, Debussy e Ravel. A saída foi criar novas sonoridades com a mistura dos sons dos instrumentos existentes, o que exigiu um profundo conhecimento do comportamento timbrístico e da natureza da interação entre eles. Esta forma de síntese foi chamada de Síntese Instrumental (Fineberg, 1999).

Os avanços do uso da eletrônica como forma de representação analógica do sinal sonoro, iniciados por Graham Bell, somados aos anseios por novas sonoridades, inspirou a criação de novos instrumentos musicais. Em 1906, Thaddeus Cahill inventou o *Telharmonium*, que com mais de 200 toneladas, produzia sons de várias freqüências e intensidades, alternando a corrente em dínamos que eram as fontes de oscilação e geração de sinal. Trata-se do primeiro dispositivo de síntese aditiva (Weidenaar, 1995). Em 1920, Lev Theremin inventou o *Téremin* (Theremin, 1927), instrumento baseado em osciladores radiofônicos projetado para ser tocado sem contato físico, utilizando-se da proximidade das mãos do executante de duas antenas. Tratava-se de um circuito RLC primitivo, em que uma antena controlava a variação da freqüência e a outra a variação de intensidade sonora.

Um dos mais notáveis trabalhos musicais de exploração da sonoridade, em uma noção ampla, foi desenvolvida na obra do compositor Edgar Varèse. Entre outros aspectos importantes, ele incorporou elementos não tradicionais (ou até o momento não musicáveis) às suas obras na Década de 30. Como exemplo, o uso de uma sirene junto com instrumentos de percussão e o naipe de metais na obra "Amériques". Ele ainda utilizou o *Téremin* na obra "Ecuatorial" de 1934. Com inspiração na física quântica, Varèse estabeleceu uma nova noção de sonoridade se abstraindo dos parâmetros tradicionais da música. Ele ainda engendrou novos conceitos para explicar a evolução sonora na música, utilizando modelos de nuvens sonoras baseados na Cinética dos Gases (Chou, 1979). Varèse influenciou compositores como Boulez e Stockhausen, tornando-se presente na música pop ao interagir com nomes como Frank Zappa e Pink Floyd. Os avanços artísticos de Varèse enunciavam a necessidade dos novos músicos da época transporem o universo dos elementos tradicionalmente estabelecidos na música, e, para isso, criou-se a demanda por uma ferramenta versátil o suficiente, como viria a ser o computador.

Proposto como uma ferramenta de manipulação de sonoridade, o computador supre esta demanda criativa da composição musical, que surge a partir da segunda metade do Século XX. Ele abre novos caminhos, erradica antigas limitações e propõe novos problemas. Ele torna acessível o processo de design sonoro[5], i.e., do projeto arquitetônico de novas sonoridades por procedimentos de síntese ou de manipulação de timbres já existentes com base em processamento computacional. Estabeleceu-se um processo de co-evolução em que as tendências da música são influenciadas pelos modelos sonoros computacionais, e a evolução musical cria a demanda por novos modelos e soluções.

O advento do computador como ferramenta de produção sonora abriu a possibilidade de se obter sonoridades novas, antes não imagináveis e sem qualquer associação a uma fonte real. Foi possível ainda manipular timbres existentes de formas inusitadas. Permitiu-se, assim, experimentar de forma real a diferenciação dos conceitos de sonoridade e timbre, como imaginou Schaeffer.

[5] O termo design sonoro surgiu no teatro e, posteriormente, ganhou destaque na produção sonora para o cinema. Mais recentemente, tem sido utilizada para designar processos gerais de criação de sons, englobando a criação de material sonoro para as áreas mais gerais, incluindo a produção musical.

2.1.2 Computador como ferramenta de manipulação de sonoridade

O computador digital é uma ferramenta universal de processamento de dados capaz de tratar qualquer tipo de informação representável em uma base computável, i.e., uma base de estrutura com definição formal e de natureza numérica ou simbólica. A discriminação desta representação é exatamente o ponto de maior tribulação no processo de design sonoro. A especificação da representação visa não só portar a informação para o domínio computacional, mas também é essencial para enfatizar as nuances que permitem otimizar o processamento para a obtenção dos resultados de síntese.

Um problema pode ser computável, por possuir uma representação e um algoritmo de solução em passos finitos, e não ser tratável pela inexistência de recursos computacionais para a execução do processamento em tempo prático. Mathews et al. (1969) foi pioneiro ao descrever os problemas da computação de sinais musicais: "Os dois problemas fundamentais em síntese sonora são: (1) a vasta quantidade de dados necessária para especificar a função de pressão - conseqüentemente a necessidade de programas muito rápidos - e (2) a necessidade de uma linguagem simples e poderosa para descrever uma complexa seqüência de sons"[6]. Serra (1997) comenta que, dos problemas citados por Mathews, o primeiro já foi resolvido pelo desenvolvimento tecnológico, enquanto o problema de representação não possui ainda uma solução satisfatória: "É impossível descrever sons se for necessário definir cada um dos números que representam a onda acústica. Devemos ser capazes de descrever sons através de menos números ou a partir de sons gravados"[7].

É de fato possível, e aplicável a muitos problemas de processamento sonoro, a representação do som por uma grande quantidade de amostras. A Modulação por Códigos de Pulso (PCM) é uma representação quantizada e codificada que discretiza a variação temporal da onda de pressão

[6] "*The two fundamental problems in sound synthesis are (1) the vast amount of data needed to specify a pressure function - hence the necessity of a very fast program - and (2) the need for a simple and powerful language in which to describe a complex sequence of sounds.*"

[7] "*...it is impossible to describe sounds if it is necessary to define each and every one of the numbers which represent an acoustic wave. We must be able to describe sounds from less numbers or start from recorded sounds.*"

que define um sinal sonoro. A PCM é computável e tratável com as ferramentas tradicionais de processamento digital de sinais. Um exemplo são os filtros digitais, que são ferramentas úteis para manipulação sonora. Todavia, é difícil descrever como esses filtros efetivamente modificam o som sem uma compreensão da relação entre a entrada, os parâmetros do filtro e o resultado sonoro esperado.

Na computação de sinais musicais, há um agravante quando se considera o processo de criação sonora: a diferença essencial entre o que é percebido como som e o sinal sonoro. A cognição do sinal sonoro envolve uma série de fatores ambientais, perceptivos e psicológicos. Assim, para uma computação eficiente em um processo de síntese sonora, deve-se obter uma representação computável, tratável e que seja coerente com a representação cognitiva. Definir esta representação computacional é ainda mais complicado pelo fato de inexistir um modelo formal e consensual de como se realiza a cognição sonora. De fato, não são totalmente conhecidas as acepções perceptuais do sinal sonoro, ou quais de seus detalhes evocam certas qualidades perceptivas.

2.2 Síntese sonora

2.2.1 Mecanismo tradicional de síntese sonora

Serra (1997) diz que os músicos esperam de um mecanismo de síntese (1) a capacidade de criar qualquer som imaginável e (2) a capacidade de manipular qualquer som existente de qualquer forma concebível. Estes objetivos são limitados tanto pela incapacidade do ser humano de imaginar todos os sons possíveis quanto por restrições tecnológicas. Serra (1997) cita que, no processo de desenvolvimento de uma técnica de síntese sonora, devem ser respeitadas algumas diretrizes:

- **Qualidade sonora:** relacionada à riqueza do som, um som de qualidade seria um som próximo a sons naturais, enquanto um som pobre seria um som facilmente percebido como sintético e simples.
- **Flexibilidade:** capacidade do método de obter variações do material sonoro com a variação de

seus parâmetros.

- **Generalidade:** a possibilidade de uma técnica gerar qualquer timbre.
- **Demanda computacional:** quantidade de processamento computacional necessária para a geração do material sonoro.

Smith III (1991) cria uma taxonomia que organiza as técnicas de síntese sonora em algoritmos abstratos, modelagem espectral, modelagem física e processamento de amostras reais. A Tabela 2.1 traz uma listagem de técnicas de síntese organizadas de acordo com a forma com que Smith desenvolveu sua taxonomia.

Processamento de Gravação	**Modelagem Espectral**	**Modelagem Física**	**Algoritmos Abstratos**
Concrète	Wavetable F	Cordas de Ruiz	VCO,VCA,VCF
Wavetable T	Aditiva	Karplus-Strong Ext.	Some Music V
Sampling	Vocoder Fase	Waveguide	FM Original
Vector	PARSHL	Modal	FM Feedback
Granular	Seno+Ruído	Cordis-Anima	Waveshaping
Comp. Principais T	Comp. Principais F	Mosaic	Distorção de Fase
Wavelet T	Chant		Karplus-Strong
	VOSIM		
	Risset FM Metais		
	Chowing FM Voz		
	Subtrativa		
	LPC		
	FFT Inversa		
	Clusters Lineares Xenakis		

Tab. 2.1: Taxonomia das técnicas de síntese digital. T = domínio do tempo. F = domínio da freqüência. Extraído de (Smith III, 1991)

A Figura 2.1 faz uma representação gráfica de um modelo abstrato de um mecanismo tradicional de síntese sonora. Cada sintetizador possui um espaço paramétrico específico, formado pelas variações possíveis de suas entradas. De forma geral, o sintetizador trabalha convertendo uma trajetória no espaço paramétrico em material sonoro.

Fig. 2.1: Esquemático de um sintetizador tradicional

A Síntese Aditiva é um exemplo de técnica tradicional, em que o material sonoro é gerado a partir de um conjunto de osciladores harmônicos, tendo como base uma freqüência fundamental dada. Cada oscilador é modulado em amplitude e em fase por envelopes específicos. Em um exemplo de uma implementação digital, a Equação 2.1 relaciona a seqüência de saída com os parâmetros de entrada. A Figura 2.2 mostra como a síntese aditiva se enquadra no modelo abstrato da Figura 2.1. O espaço paramétrico é definido pelas possibilidades de variação da freqüência fundamental (f_0), e vetores de amplitude das componentes senoidais e cossenoidais (a_k e b_k).

Fig. 2.2: Síntese aditiva: conjunto de k_{max} osciladores harmônicos da freqüência fundamental f_0 modulados pelos arranjos de envelopes a_k e b_k. De acordo com a Equação 2.1.

$$s[n] = \frac{1}{2}a_0[n] + \sum_{k=1}^{k_{max}} a_k[n]cos\left(\frac{2\pi f_0}{F_s}kn\right) - b_k[n]sin\left(\frac{2\pi f_0}{F_s}kn\right) \tag{2.1}$$

2.2.2 Modelo do processo de síntese sonora

Neste trabalho, o processo de síntese sonora foi relacionado à busca do usuário por um som capaz de suprir expectativas estéticas aqui denominadas de "objetivo sonoro". A Figura 2.3 faz uma ilustração deste ponto de vista, explorado na pesquisa. Primeiramente, um objetivo sonoro é concebido na forma de uma idéia, o compositor se utiliza de um mecanismo computacional de síntese sonora para obter uma amostra sonora física e audível, referenciada por "material sonoro". Portanto, a noção fundamental do nosso estudo baseia-se na capacidade que o sistema terá de satisfazer, ou auxiliar, a analogia construtiva do usuário. Este problema será tratado na Seção 3.2 do Capítulo 3.

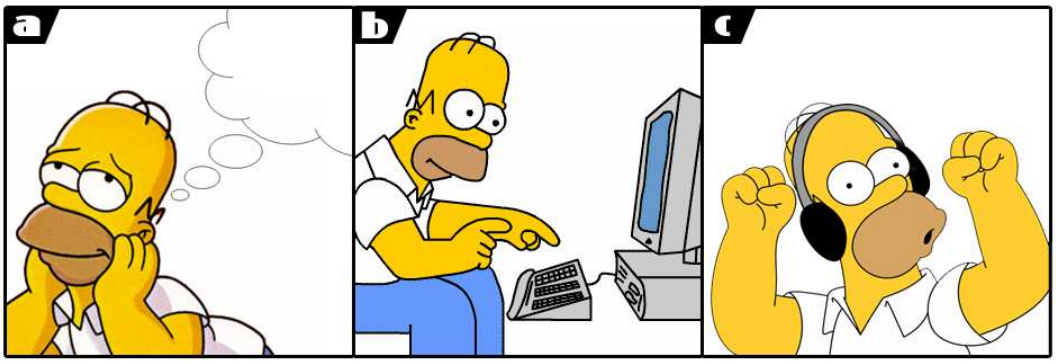


Fig. 2.3: Ilustração do processo de síntese sonora: a) Objetivo Sonoro b) Processo de Síntese c) Material Sonoro

Ao modelar a idéia apresentada no parágrafo anterior, é possível imaginar, inicialmente, um fluxo informacional sem retroalimentação. O objetivo sonoro do usuário interfere no processo de síntese do material sonoro; entretanto, o objetivo não sofre interferência do material. Este modelo inicial é incapaz de representar de forma verossímil a complexidade do processo de design sonoro, por considerar o objetivo sonoro estático e bem definido no imaginário do usuário. Muitas vezes, o objetivo sonoro é nebuloso e incerto e amadurece com a interação do usuário com o sistema de produção sonora. Levando em conta que existe um processo de tentativa e erro, a própria representação computacional de algum método de síntese influencia a formação do objetivo sonoro via retroalimentação, como menciona Furlanete (2000). Ainda assim, apesar de ser uma simplificação que considera o objetivo sonoro estático, o modelo inicial é suficiente para ilustrar os mecanismos envolvidos no processo de síntese sonora.

Pode-se também relacionar este processo criativo com a noção de auto-organização, como menciona Manzolli (1996): "Na medida que o músico vai desenvolvendo suas idéias musicais, elas ganham forma e corpo e passam a ter vida própria..." e "Vemos o processo de criação musical como um eterno aprender, surpreender e surpreender-se. É da exploração das peculiaridades do som que uma nova composição nasce.". Desta forma, o estudo desenvolvido durante esta pesquisa voltou-se a ampliar a capacidade de um processo computácional de síntese sonora de fornecer ferramentas para que o usuário criativo pudesse explorar cada vez mais as peculiaridades do som.

No continuar da modelagem, associando a visão objetivo-material com o modelo de sintetizadores da Figura 2.1, é possível identificar alguns processos internos do procedimento de síntese sonora. A Figura 2.4 apresenta um diagrama funcional deste novo modelo. Primeiramente, o compositor idealiza um objetivo sonoro (considerado aqui completo e estável). Em seguida, ele realiza um processo de tradução de seu objetivo em parâmetros do sintetizador. A partir destes parâmetros, o sintetizador produz o material sonoro. Por fim, o usuário realiza uma comparação subjetiva entre o material sonoro e o objetivo. Se o resultado sonoro for insatisfatório, o usuário terá que repetir o processo. Porém, o erro subjetivo obtido pela comparação interfere no processo de tradução no sentido de minimizá-lo. Neste sistema, é necessário que o próprio usuário se adapte aos parâmetros do sistema, de forma a aprender como obter os resultados desejados. A capacidade do usuário de transferir para os parâmetros do sistema seu objetivo sonoro é chamado, neste trabalho, de "tradução" e a possibilidade de modificar os mecanismos que permitem esta tranferência é chamado de "controle".

A tradução é um processo cognitivo de mapeamento das qualidades sonoras esperadas em parâmetros computacionais. Portanto, ela é dependente dos conhecimentos do compositor sobre o espaço paramétrico do método e de como o sintetizador o converte em material sonoro. Este mapeamento é complexo, porque o espaço paramétrico dos processos computacionais é, em geral, abstrato e pouco representativo em relação às qualidades subjetivas do som. Há aqui um exemplo de discrepância entre a noção de controle de timbre e de sonoridade, como já foi apontado anteri-

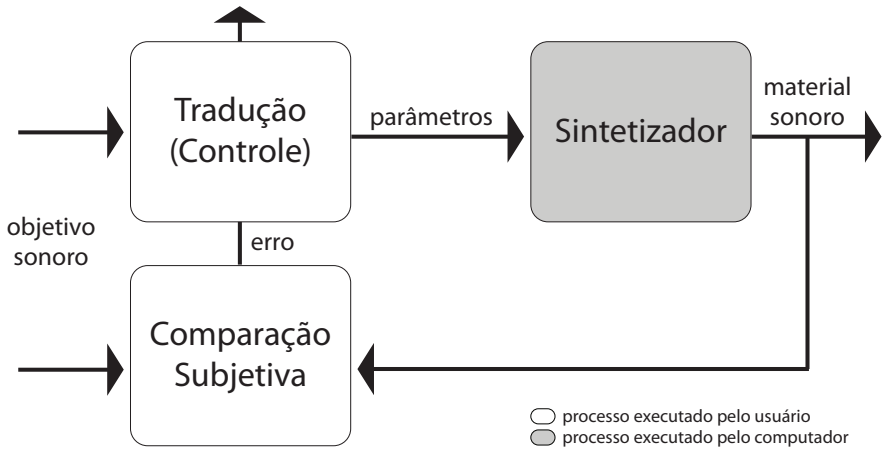


Fig. 2.4: Modelo abstrato do processo de síntese sonora. A tradução é realizada pelo compositor na geração de uma trajetória paramétrica a partir de um objetivo sonoro.

ormente ao citar o trabalho de Schaeffer. Sem um conhecimento profundo do estímulo sonoro e do mecanismo de síntese, é impossível imaginar qualquer relação entre eles, e, portanto, estabelecer uma política eficiente de controle. Esta política de controle consiste na modificação do mapeamento entre o objetivo e os parâmetros do sintetizador, de forma a minimizar a medida subjetiva de erro entre o material gerado e o objetivo. O único processo computacional envolvido é o do próprio sintetizador. Este produz material sonoro a partir de um conjunto de parâmetros, geralmente, de forma determinística e sem retroalimentação. O sintetizador atua como a planta a ser controlada.

Os atributos do modelo, apresentados no parágrafo anterior, permitem visualizá-lo como uma otimização cujo controle e estimativa de eficiência são subjetivos e realizados pelo usuário. Tal característica impossibilita a aplicação de automação na busca pelo objetivo, pela impossibilidade de estabelecer uma medida computacional de erro entre o objetivo e o material. Para este fim, são necessárias uma estratégia capaz de portar o objetivo sonoro para o domínio computacional, de forma a ser inteligível aos algoritmos abstratos, e um avaliador da diferença entre o objetivo e o material.

A Figura 2.5 apresenta o diagrama funcional deste caso. As funções de tradução, controle e avaliação são realizadas pelo computador. O laço de retroalimentação permite a busca automática pelos parâmetros que imprimam a menor diferença entre o objetivo e o material sonoro. O usuário é

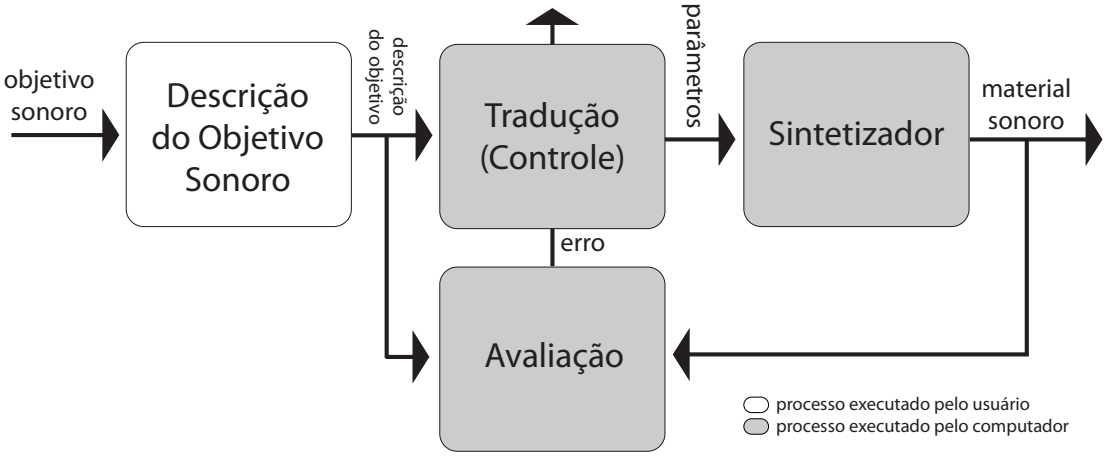


Fig. 2.5: Modelo abstrato do processo de síntese sonora por descrição de objetivo. A tradução, o controle e a avaliação são realizadas pelo computador. O usuário realiza descrição do objetivo em uma estrutura de dados inteligível ao computador.

incumbido de descrever seus objetivos sonoros em algum meio concreto intermediário.

A arquitetura de um sistema com a disposição da Figura 2.5 requer um entendimento de como representar o objetivo sonoro e passa, necessariamente, pelo estudo de como o som é entendido pelo usuário. Deve-se estabelecer uma estrutura de dados que possa ser traduzida em parâmetros de síntese e que seja intuitiva ao usuário, no ato de descrição da sonoridade esperada. Neste sentido, o Capítulo 3 aborda os conceitos de representação e analogia, e como estes podem ser aplicados na síntese do modelo proposto.

Capítulo 3

Representação Sonora e Analogia

3.1 Modelos de representação sonora

O termo "representação" possui várias acepções dependentes da área do conhecimento no qual é aplicado. Para Vickery (1986), citando Goethe, pode-se apresentar como a noção de representação a expressão "*ato de colocar algo no lugar de*". David Marr (1982) a define como: "um sistema formal para tornar explícito certas entidades ou tipos de informação, junto com uma especificação de como o sistema faz isso"[1]. A representação pode ser tomada como uma forma de informação estruturada e definida em algum meio que sirva de base à descrição de algum objeto, relação ou idéia. O meio é o domínio da representação, tendo como exemplo o domínio matemático, numérico ou simbólico. Nesse domínio, representar é formular expressões formais. Em outro domínio, a representação é realizada também a partir do uso da linguagem natural, ao atribuir qualidades do mundo real a objetos. A descrição segue no sentido de tornar a informação inteligível a um processo atuante sobre o meio da representação.

Quanto às formas de representação, para Vickery existem dois tipos: as primárias, que estão vinculadas à expressão dos resultados do pensamento do sujeito; e as secundárias, que é o resultado da interação da primária com mecanismos físicos, como a escrita, e, por extensão, em uma codificação

[1] *"a formal system for making explicit certain entities or types of information, together with a specification of how the system does this."*

computacional. Ampliando esta noção ao trato da cognição do fenômeno sonoro, a representação primária se refere à idéia do som e à sensação da sonoridade. A representação secundária se refere aos mecanismos formais de descrição sonora que permitem o processamento computacional. Neste trabalho, as representações primária e secundária são referenciadas, respectivamente, por "representação cognitiva" e "representação computacional".

Assim, neste trabalho, representação sonora é considerada sob dois aspectos: (1) "representações sonoras computacionais", Seção 3.1.1, se referem à representação secundária utilizada no processamento de síntese sonora digital; e (2) modelagem da representação primária do usuário, que visa elucidar o seu objetivo sonoro, aqui denominada de "representação cognitiva do som", Seção 3.1.2. De acordo com o que foi apresentado no capítulo anterior, no caso (1), o estudo visa o entendimento dos atributos da sonoridade que podem ser representados computacionalmente e satisfazem o objetivo sonoro do usuário e, no caso (2), busca-se a modelagem da sonoridade.

3.1.1 Representações sonoras computacionais

A representação computacional é uma forma de descrever o som (ou sinal sonoro) que permite a sua manipulação por processos algorítmicos. Ela é dependente do meio computacional e determina o espaço paramétrico. São necessários não só em processos de síntese, mas também em qualquer processo computacional que envolva som.

A representação computacional mais direta e próxima do sinal sonoro é a Modulação por Códigos de Pulso (PCM), desenvolvida em 1939 por Alec H. Reeves (Jayant & Noll, 1984). É uma representação quantizada e codificada da discretização no tempo da função de pressão que define um sinal sonoro, ilustrada na Figura 3.1. Ela foi criada para permitir a transmissão de áudio em canais de transmissão analógicos. Ela é pura no sentido de não levar em consideração fatores cognitivos e ser uma representação fiel do sinal discretizado. A representação ponto a ponto do PCM é eficiente, por permitir aplicar todas as técnicas de processamento digital de sinais no sinal de áudio. Representações codificadas, como o PCM, que permitam codificação e decodificação, também são

chamadas de *Codec*s.

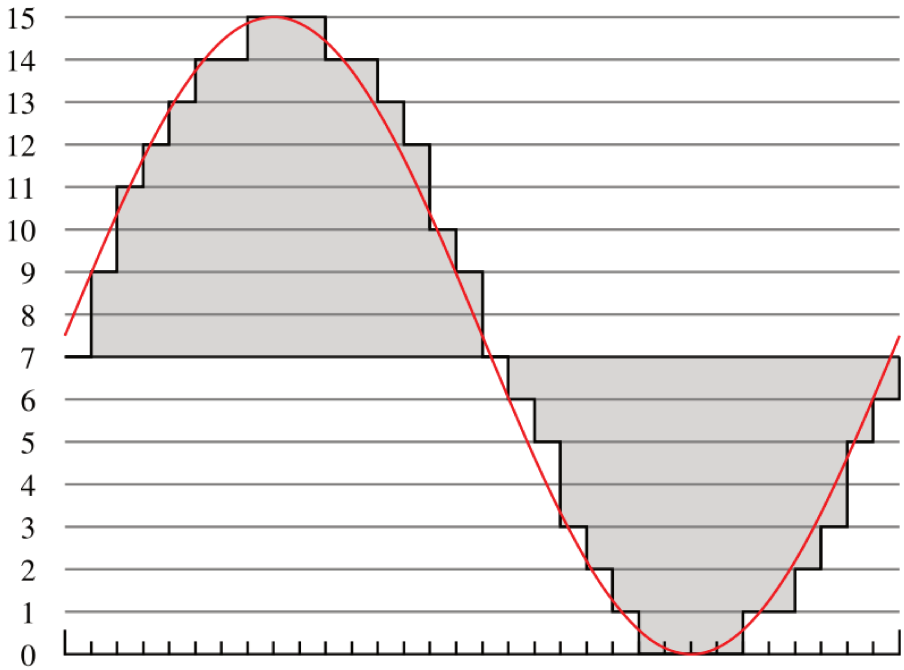


Fig. 3.1: Exemplo de quantização e codificação de 4 bits em Modulação por Códigos de Pulso.

Outra representação tradicional é na forma de espectro de freqüências, obtidos pela Transformada de Fourier, Equação 3.1. Trata-se de uma representação completa, no sentido de que o sinal original pode ser reconstruído de forma plena a partir da representação utilizando a Transformada Inversa. Na equação 3.2, $X(f)$ representa o espectro de freqüências, sendo f uma dada freqüência medida em Hertz; $x(t)$ é o sinal no tempo t medido em segundos.

$$X(f) = \int_{-\infty}^{\infty} x(t)\, e^{-i2\pi ft}\, dt \tag{3.1}$$

$$x(t) = \int_{-\infty}^{\infty} X(f)\, e^{i2\pi ft}\, df \tag{3.2}$$

Em aplicações computacionais, tanto o sinal quanto a transformada são discretos, com uma dada taxa de amostragem. Utiliza-se então a Transformada Discreta de Fourier (DFT), Equações 3.3 e 3.4. Aplica-se um algoritmo eficaz para o cálculo desta transformada, denominado Transformada Rápida de Fourier (FFT). Esta representação pode ser obtida diretamente do PCM com a aplicação do FFT.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \qquad k = 0, \ldots, N-1 \tag{3.3}$$

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N} kn} \qquad n = 0, \ldots, N-1. \tag{3.4}$$

Existe ainda uma variação denominada Transformada de Fourier de Curta Duração (STFT), que inclui um janelamento do sinal original e a aplicação da transformada discreta sobre as amostras de sinal janelado. Uma explicação destas representações e de outras derivadas, como Transformada Discreta do Coseno (DCT), são discutidas em Oppenheim & Schafer (1999).

No mesmo sentido instantâneo da STFT, a transformada *Wavelets* incorpora o janelamento ao núcleo de transformação, permitindo adaptar o tamanho da janela a determinadas freqüências, otimizando a resolução tempo-freqüencial.

A essência do PCM e das representações espectrais que visam descrever o sinal da forma mais direta também é o maior limitante para aplicações reais. A informação sonora nestas representações demanda muito espaço, seja no armazenamento ou na transmissão. Existem codecs com codificação mais compacta, sendo divididos em duas classes: com e sem perda de qualidade.

Os codecs sem perda de qualidade de áudio permitem que um dado em sua representação seja recuperado integralmente em PCM. Em geral, os codecs aplicam codificação baseada em análise de entropia, como a codificação Huffman (Huffman, 1952).

Os codecs com perda de qualidade levam em consideração aspectos perceptuais no objetivo de obter um compromisso entre uma representação econômica e a capacidade de reconstruir o sinal. Algumas variações da codificação do PCM, como o μ-Law e A-Law, utilizam uma codificação logarítmica numa tentativa de diminuir o efeito de quantização na compressão de dados de fala (Rabiner & Schafer, 1978). Derivada da representação espectral, MFCC é uma representação baseada em um modelo auditivo e, geralmente, é aplicado em reconhecimento de fala (Mermelstein, 1976). LFC é uma derivação mais simples da MFCC. Nesta categoria, se enquadra a grande maioria dos codecs de áudio atuais, como MP3 e OGG Vorbis (Moffitt, 2001).

Nos mecanismos tradicionais de síntese, a representação utilizada é estipulada pelo próprio mecanismo de síntese. Por exemplo, síntese FM (Chowning, 1973) tem o sinal de saída representado pelos parâmetros do sinal definido pela Equação 3.5. A representação de fato é determinada por dois parâmetros: o índice de modulação, que é proporcional à amplitude do modulador, e inversamente proporcional à freqüência do modulador, e a freqüência fundamental. Outro exemplo é a supracitada síntese aditiva, Figura 2.2.

$$X(t) = a_1 cos((w_1 + a_2 cos(w_2 t))t) \tag{3.5}$$

Na síntese granular (Roads, 1988), o material sonoro é composto por amostras sonoras muito rápidas (entre 10 e 100 milisegundos) arranjadas no tempo. Existe a necessidade de definir as amostras sonoras e a ordenação delas na composição do material final. A representação de livros de Xenakis (1971) descreve cada amostra com uma abordagem espectral e estabelece um aparecimento serial com uma freqüência de amostragem granular definida.

Dada a necessidade de alta densidade de grãos para a produção de texturas complexas, o próprio Xenakis e outros autores criaram uma série de heurísticas para controlar a geração granular, podendo

estas serem determinísticas ou estocásticas (Truax, 1996). Uma abordagem comum é o controle de natureza estocástica a partir de funções de probabilidade e máscaras de tendência, que controlam desde a densidade granular até a composição espectral das micro-porções sonoras (Truax, 1988). Xenakis utilizava operações booleanas sobre sua representação seqüencial de quadros. Ele criou a idéia de *ataxy* que reflete o nível de mudanças entre quadros subseqüentes. Ele utilizava cadeias de Markov para controlar o fluxo de diferentes eventos (Xenakis, 1971). Outros métodos realizam controle do posicionamento granular, como, por exemplo, síncrono à altura (De Poli & Piccialli, 1991). Outro método realiza formações periódicas para obter um fenômeno formante (Rodet, 1984). Roads (1997) propõe um controle determinístico baseado em notação de notas, enquanto Di Scipio (1994) utiliza funções não-lineares no controle de fluxos granulares. Em todos os casos, a representação computacional é baseada na heurística gerativa dos grãos, seja na forma de máscaras de tendência, seja na forma de uma transcrição em notação musical.

Em síntese sonora por modelagem física, o material sonoro é gerado a partir de modelos que simulam a produção do som no meio real, através de equacionamento e algoritmos. A parametrização da sonoridade é feita em termos da descrição do meio físico, i.e., dos materiais envolvidos no instrumento e na forma como o usuário interage com ele. Um exemplo é a síntese de cordas de Karplus-Strong (Karplus & Strong, 1983) apresentada na Figura 3.2. A excitação representa a forma de ataque à corda. No algoritmo original, é utilizado ruído branco, mas pode incluir qualquer sinal limitado em freqüência. A excitação é repassada à saída e realimentada com um atraso de L amostras. A saída do atraso é filtrada, em geral, por um filtro passa-baixa de primeira ordem. Este sinal é retornado à linha de saída e o atraso é realimentado. A altura do material gerado é determinado pelo atraso somado da linha de atraso e do filtro (Moore, 1990). Neste caso, a parametrização está na atribuição da excitação, na determinação do filtro e na especificação do atraso. Uma extensão deste modelo é a *Digital Waveguide Synthesis* (Smith III, 1987) que teve aplicação comercial pela Yamaha.

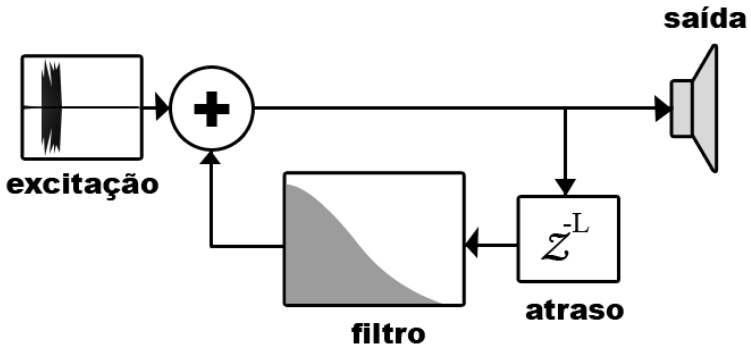


Fig. 3.2: Esquemático da síntese de cordas de Karplus-Strong.

3.1.2 Representação cognitiva do som

Na fenomenologia de Husserl (1969) e Heidegger (1962), o que o homem percebe como realidade, seu estado de consciência ou *noesis*, é uma mera interpretação intencional do universo exterior criada a partir de estímulos provenientes dos sentidos, de sua experiência e suas reflexões sobre como elas se relacionam. As experiências, por sua vez, se atualizam a partir desta realidade criada, i.e., pelos estímulos, pelas experiências anteriores e pelas relações entre elas. Para que indivíduos diferentes criem uma realidade semelhante, eles devem ter estruturas sensitivas, experiências e mecanismos de introspecção totalmente iguais. Este é um dos pontos principais de Husserl: a consciência sobre as coisas é sempre inacabada, i.e., são possíveis diferentes consciências sobre o mesmo objeto. Isso inspirou filósofos existencialistas, como Jean-Paul Sartre (1957), que diz que a natureza da própria consciência é infinita e qualquer tentativa de descrição ou entendimento é fadada à forma de ansiedade, incompletude.

Tratando o som como um fenômeno da consciência, os estudos de Husserl e Heidegger implicam que qualquer consciência sobre o fenômeno sonoro é inacabada, assim como a consciência da própria consciência do som, seguindo Sartre. O que isso quer dizer é que, segundo a fenomenologia e o existencialismo, são possíveis diferentes representações cognitivas do som e não é possível obter um modelo completo desta representação. Schaeffer foi influenciado pela visão fenomenológica de Husserl. Ao postular a noção de sonoridade, explorada nesta dissertação e apresentada anteriormente

no Capítulo 2, Seção 2.1, ele deixou latente que existe uma experiência sonora que perpassa os domínios do mundo físico, de forma tal que indivíduos diferentes têm experiências sonoras próximas, mas dessemelhantes. Portanto, pode-se interpretar que o conceito de "escuta" *schaefferiano* é um reflexo da experiência do indivíduo no domínio sonoro e a mesma se atualiza a partir de realidades ou estímulos sonoros criados durante a sua existência.

No decorrer deste texto, utiliza-se o termo "percepção sonora" para o processo de aquisição do estímulo sonoro bruto, através dos órgãos dos sentidos, e conversão para um conjunto de estímulos neurais. O termo "cognição sonora" se refere aos processos que ocorrem no cérebro e que moldam a noção do mundo exterior e sobre seus próprios processos informacionais, que incluem a própria percepção sonora. Chalmers et al. (1991), do grupo de Douglas Richard Hofstadter, discorre sobre a impossibilidade de dissociar a cognição da percepção em uma crítica à teoria das estruturas compartilhadas, modelo de inteligência artificial da época. Chalmers cita que Immanuel Kant dividia o processo de percepção, em faculdade da sensibilidade, a recepção da informação bruta; e faculdade do entendimento, a organização desta informação em uma experiência coerente e significativa do mundo. Chalmers utiliza o termo percepção de alto-nível para o entendimento e o termo percepção de baixo-nível para o processo sensorial. No campo de estudo da percepção do fenômeno sonoro, podemos destacar a psico-acústica como a ciência que estuda a percepção de baixo-nível, neste trabalho denominada de "percepção sonora". A noção de percepção de alto-nível relaciona o estímulo de ordem psico-acústica com as experiências do indivíduo, neste trabalho denominada de "cognição sonora".

A construção de modelos relacionados à cognição sonora é alvo de um amplo estudo, como trabalho recente de Koelsch et al. (2004) e Huron (2006). O que torna essa tarefa árdua é o fato de que a cognição sonora não é limitada pela natureza física do estímulo. O fenômeno sonoro relaciona-se com outras experiências do indivíduo, que estão imersas no contexto sociocultural no qual o estímulo é gerado. Em um exemplo, o som de um latido é normalmente associado à figura de um cachorro. Outro exemplo é o fato da marca da cerveja influenciar a percepção do gosto, como menciona Allison & Uhl (1964). Devido a essa complexidade, é necessário que, no processo de

modelagem de síntese sonora, seja estabelecido um ponto de vista relacionado ao processo cognitivo, de forma a contemplar no espaço de atributos, ou parâmetros, as noções de relevância do usuário.

Ao considerar os sons para uma aplicação musical, estabelece-se um ponto de vista. Neste caso, os modelos de construção da realidade atuam em um espaço de atributos com três componentes: a intensidade subjetiva, (*loudness*), que é principalmente influenciada pela energia do sinal sonoro; a altura, que é determinada pela distribuição espectral; e a sonoridade dentro de um campo de exploração timbrística.

Alguns estímulos possuem qualidades que são percebidas de forma semelhante por diferentes indivíduos. Este fenômeno é facilitado por atributos que possuem a sua percepção melhor definida e por mecanismos de percepção semelhantes. Isso permite o estabelecimento de padrões perceptivos de estímulos, como por exemplo, as cores, que numa gama de nuances de tonalidade é invariante. A percepção causada pelo estímulo de luz, restrito a uma dada faixa espectral, cria uma sensação que é reconhecida como sendo a mesma à maioria das pessoas e que permite uma associação semântica: azul, verde, vermelho, etc. Em geral, é possível realizar um mapeamento entre propriedades físicas do estímulo e o tipo de padrão sensorial. As alterações dos mecanismos sensoriais, como a degeneração pela idade, implicam na modificação da percepção de padrões estáveis e invariantes na experiência individual. Por exemplo, a preferência por sabores complexos, como baunilha e morango, tende a ser acentuado em pessoas mais idosas (Plattig et al., 1980). No caso das cores, disfunções, como o daltonismo, levam o indivíduo a perder a capacidade de perceber algumas tonalidades e, eventualmente, confundir a associação semântica entre o verde e o vermelho. No caso do som, são padrões perceptuais relativamente invariantes: a altura e a intensidade subjetiva. A sensibilização do ouvido para a percepção de intensidade diminui com o avanço da idade do indivíduo. Neste caso, apesar de invariante para a maioria da população, a sensibilização diminui, pois os cílios cocleares se degeneram com o impacto mecânico de ondas acústicas na cóclea.

Na percepção de intensidade sonora subjetiva, existe um mapeamento entre uma escala perceptiva padrão e os atributos físicos de intensidade (dB) e freqüência (Hz), ilustrado na curva de

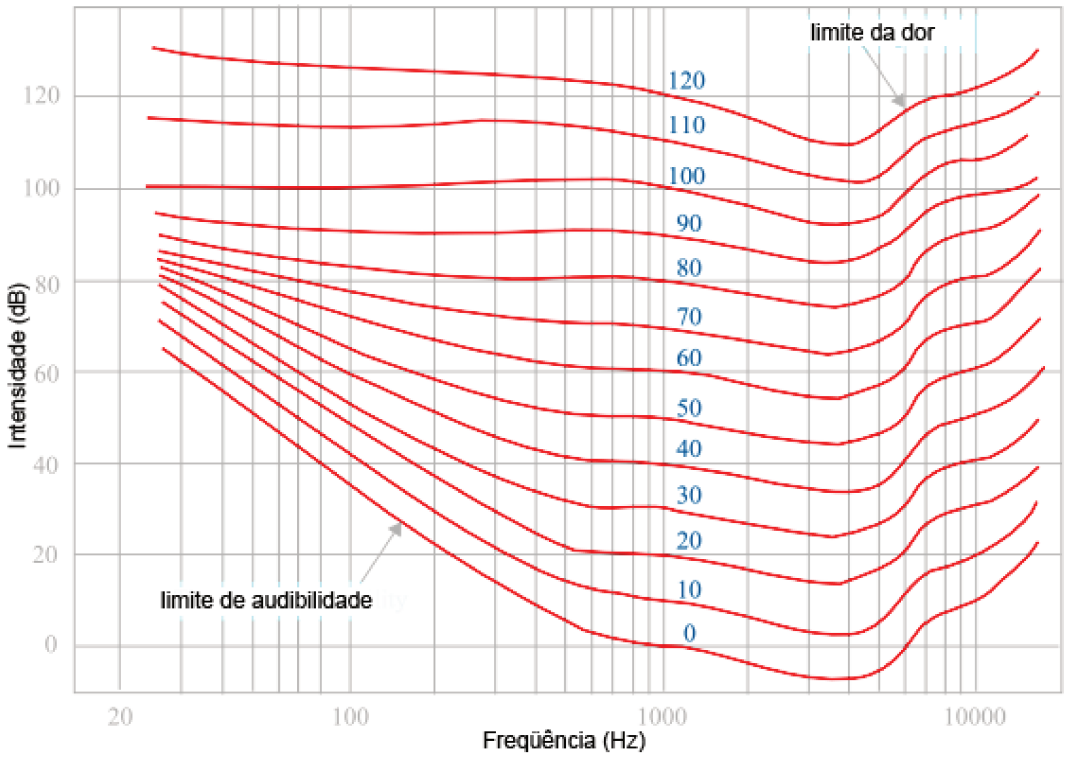


Fig. 3.3: Curva de Fletcher & Munson (1933). Mapeamento de intensidade e freqüência do sinal na intensidade percebida. Linhas correspondem a valores que são percebidos como de mesma intensidade subjetiva. A unidade subjetiva é chamada de *phon* e coincide com a intensidade física em dB em 1kHz.

Fletcher & Munson (1933), Figura 3.3. Este mapeamento foi obtido por análise estatística de testes subjetivos com uma população de indivíduos. As medidas mais modernas resultaram em uma curva padronizada pela "International Organization for Standardization" (ISO). A altura de sinais musicais, *pitch*, também é percebida em forma de uma escala unidimensional. A percepção da altura é influenciada pela distribuição do espectro de freqüência, de forma a estar sujeita não apenas à fundamental do espectro, mas, também, à relação de harmonicidade das componentes espectrais.

A criação de um mapeamento de características físicas em uma escala de timbres é complexa, uma vez que o ser humano não realiza esta associação em uma escala unidimensional, e de fato inexiste um padrão perceptual estabelecido de quais e quantos seriam os eixos multidimensionais que descreveriam tal escala. Um som pode ser mais ou menos intenso ou mais ou menos alto que outro, mas não é possível realizar uma comparação neste nível levando em consideração o timbre.

Outros sentidos possuem atributos subjetivos multidimensionais; é o caso do paladar: um gosto é associado a algum alimento pela composição dos padrões perceptivos amargo, azedo, salgado e doce. Neste caso, apesar de multidimensional, o mapeamento é simplificado pelo conhecimento dos eixos do espaço sensorial. Além disso, a percepção de um padrão particular, como o amargo, é fortemente associada ao estímulo das papilas gustativas com sensibilidade específica, e é possível realizar uma associação da composição química do alimento com a sensação causada. Neste caso, a percepção dos padrões é independente, formando uma base mínima para uma cognição de mais alto nível, i.e., identificação de um gosto de um dado alimento preparado por um "*chef*" francês.

Na percepção de timbre, não existe um vocabulário específico. É comum o empréstimo de termos semânticos derivados de outros domínios sensoriais para descrever o fenômeno sonoro: da visão (brilhante, opaco), do tato (áspero); ou a associação com outros timbres: metálico ou amadeirado. A associação de atributos físicos do sinal sonoro com a percepção destes padrões é ainda incipiente. E ainda, certos atributos modificam a percepção de múltiplos padrões. Kurigawa et al. (1978) discute a dificuldade de rotular estas dimensões para classificar o estímulo timbrístico.

As tentativas mais comuns são de contornar a inexistência de um espaço definido e envolvem a criação de um espaço artificial e o mapeamento das amostras sonoras a partir de informações comparativas entre as amostras. Grey (1975) realizou um estudo comparativo de escuta em um grupo, medindo distância subjetiva entre timbres de diferentes instrumentos, aplicando a mesma nota e intensidade e realizando, por técnica multivariável, a alocação nos eixos de qualidades físicas. Destes resultados, ele dispôs as amostras em um espaço tridimensional, de forma a respeitar ao máximo a medida unidimensional, Figura 3.4. De Poli & Prandoni (1997) utilizam um modelo de análise acústica associado com redes neurais para o estabelecimento de grupos. A principal aplicação é a de caracterização automática de timbres e identificação de fontes. Numa tentativa de estabelecer um mecanismo de validação dos experimentos realizados durante a pesquisa de mestrado, no Capítulo 5, Seção 5.3.8, e no Capítulo 6, Seção 6.3.3, são apresentados resultados relacionados aos dois testes psico-acústicos realizados, solicitando aos sujeitos que realizem comparações sonoras seguindo a metodologia de Grey. No primeiro experimento, é avaliada a similaridade entre duas amostras

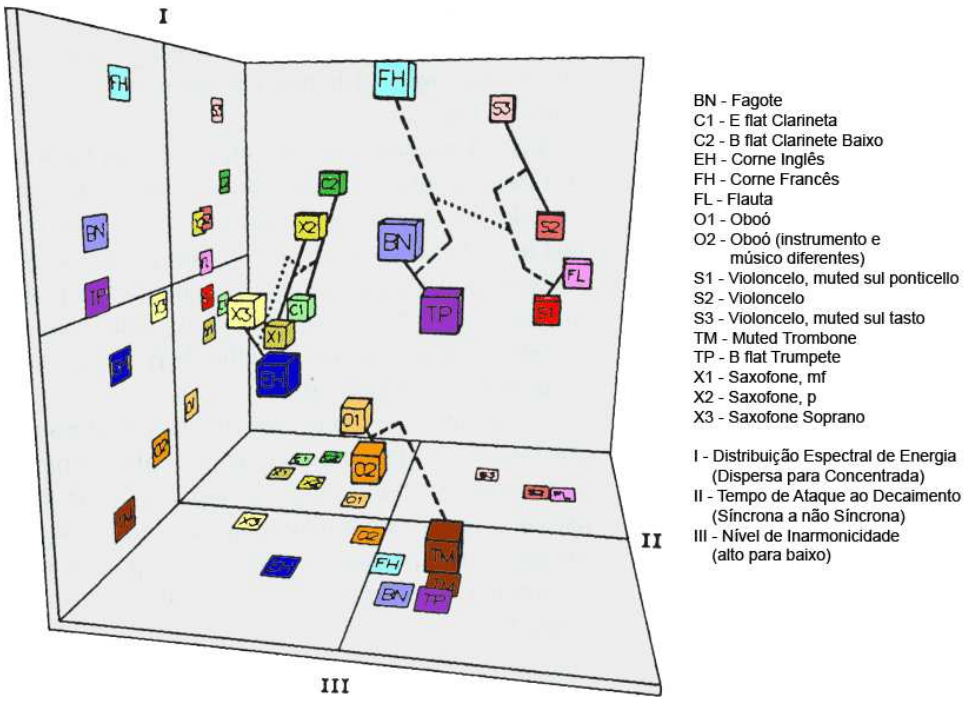


Fig. 3.4: Espaço multi-dimensional de Grey (1975).

sonoras, com análise estatística de acertos. No segundo experimento, é solicitado aos indivíduos que façam a associação entre cinco classes de sonoridades e uma seqüência de estímulos, com análise da correlação por faixas.

Outra parte importante do mapeamento é a identificação das características físicas do som que são realmente relevantes, o que propicia a definição do espaço a ser mapeado nos eixos subjetivos. Nos estudos com trompete de Risset (1960), foi mostrado que variações temporais das componentes espectrais podem ser cruciais para essa identificação. Ele verificou a importância do tempo inicial e final relativo das parciais na percepção do timbre de um instrumento musical. Grey (1975) verificou que alguns aspectos podem ser decisivos para alguns sons e desprezíveis na caracterização de outros. Mertens (1975) descreve que o timbre de alguns instrumentos, como a voz, pode ser definido por acúmulo de energia em parciais fixas, denominadas de freqüências formantes. Em Brown et al. (2001) é apresentado um método de identificação timbral a partir de características das sonoridades. Hajda et al. (1997) faz um estudo sobre as características sonoras importantes na percepção de timbre. Houve tentativas de estabelecer padrões de descrição semântica utilizando-se de morfologias

e tipologias sonoras. Estas foram realizadas nos anos 50 pelo "*Group d'Essay*" (Schaeffer, 1966; Smalley, 1986).

3.1.3 Modelo representacional do processo de síntese sonora

Ao revisar o modelo do processo tradicional de síntese sonora, apresentado na Figura 2.4 da Seção 2.2.2, sob a ótica representacional é possível explicar o processo como uma evolução da informação entre representações, como mostrado na Figura 3.5. O ponto de partida é a idéia do objetivo sonoro, a expectativa do usuário. Trata-se, portanto, de um representação cognitiva. A tradução é a conversão da informação do objetivo sonoro nos parâmetros do sintetizador. Estes parâmetros constituem uma estrutura de dados formal que é especificada pelo mecanismo de síntese, o que define representação computacional. O sintetizador produz o material sonoro a partir destes parâmetros, sendo o material sonoro uma amostragem da função de pressão da onda sonora, com definição formal e especificação.

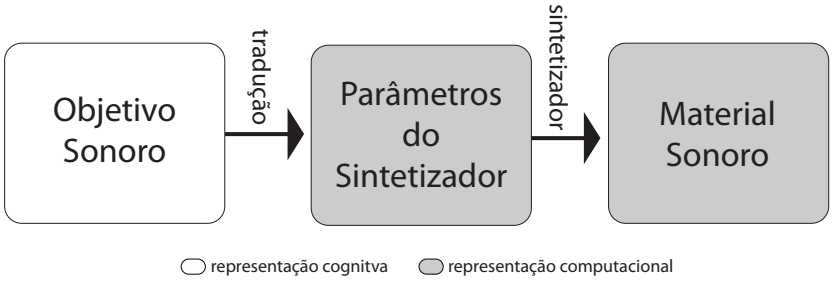


Fig. 3.5: Visão representacional no processo tradicional de síntese sonora.

A Figura 3.6 apresenta o resultado da análise representacional aplicada ao modelo proposto de síntese por descrição do objetivo sonoro, apresentado na Figura 2.5 da Seção 2.2.2. A diferença deste modelo e do anterior é a presença de uma representação computacional intermediária, localizada entre o objetivo sonoro e os parâmetros do sintetizador. Esta representação, denominada "Descrição do objetivo sonoro", é obtida por um processo de descrição do objetivo sonoro realizado pelo usuário. A tradução para os parâmetros do sintetizador é realizada como um processo computacional.

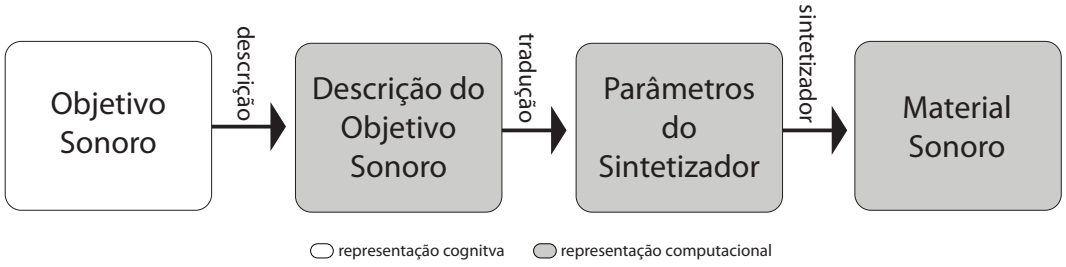


Fig. 3.6: Visão representacional no processo descritivo de síntese sonora.

A criação de um mecanismo que corresponda ao modelo da Figura 3.6 exige o desenvolvimento de uma representação computacional de descrição. Para tal, é importante o entendimento do processo de transição entre representações, mais especificamente, neste caso, a compreensão de como se dá a descrição do som pelo usuário. Neste sentido, a Seção 3.2 aborda o conceito de analogia e como ele se relaciona com representações.

3.2 Analogia e representação

O estudo de processos que se utilizam de analogias é relacionado com várias áreas do conhecimento. Foge ao escopo desta dissertação um estudo mais aprofundado deste mecanismo de vital importância para os processos de linguagem natural. Todavia, neste trabalho, fazemos uso do termo analogia como um mecanismo utilizado pelo usuário ao relacionar o objetivo sonoro com uma representação computacional.

Em engenharia, analogia é o fundamento da modelagem de sistemas. Dois processos são análogos quando se equivalem, em algum aspecto, apesar de pertencerem a paradigmas distintos. A própria concepção do termo "processamento analógico" se refere à possibilidade de realizar processamento em domínios distintos, análogos a um princípio inicial. Por exemplo, os computadores analógicos utilizam sistemas elétricos, mecânicos ou hidráulicos, análogos a algum processo real não computável por restrições estruturais, temporais ou de alguma outra natureza. Entre as décadas de 40 e 60, computadores eletrônicos analógicos eram utilizados em simulação de componentes

aeroespaciais, por serem mais rápidos e de mais fácil programação que os computadores digitais da época (Truitt & Rogers, 1960). Existem registros de computação analógica datados de 100 anos antes de Cristo: ao mecanismo *Antikythera* é creditada a funcionalidade de calcular posições astronômicas (Wright, 2004).

Decorrente do processo histórico da computação analógica e a necessidade de diferenciá-la do uso de codificação binária, o termo analógico foi atribuído aos sinais contínuos, contrastando com os sinais quantizados. Em sistemas de áudio, o termo analógico é utilizado para designar dispositivos de armazenamento e reprodução não digitais. Aplica-se esse conceito pelo fato da inscrição mecânica em discos de vinil ser uma analogia ao sinal sonoro. Conceitualmente é errado afirmar que uma representação digital não é analógica, já que a diferença entre uma gravação em PCM de outra em fita magnética é apenas a quantificação e codificação, e não o que ambas representam.

Estes processos determinam representações computacionais que são análogas aos sinais físicos. Contudo, o estudo da analogia também se relaciona com a pesquisa do processo cognitivo humano. De fato, é uma das áreas de maior convergência interdisciplinar no contexto das ciências cognitivas. Há fortes conexões entre os trabalhos de computação e psicologia, como Falkenhainer (1988), Holyoak & Thagard (1989), Halford (1992) e Keane et al. (1994).

Existe também uma discussão a respeito da importância da analogia no processo cognitivo e da forma como se desenvolve o processo de representação. Hofstadter (2001) diz que a analogia é a base da cognição, pois no processo cognitivo são estabelecidas representações do estímulo externo percebido. Ele faz uma crítica aos modelos de representação programados e diz que a representação é extraída de forma natural do estímulo e não o estímulo que é condicionado a uma forma de representação. Hofstadter diz que percepção e analogia são a mesma coisa. Forbus et al. (1998) critica esta visão dizendo que se trata de uma simples metáfora, que os processos cognitivos são muito complexos para serem simplificados pela simples noção de analogia. Em síntese, Hofstadter defende que a representação deve ser adaptativa enquanto Forbus defende uma representação inata. Morrison & Dietrich (1995) fazem uma crítica às querelas entre Hofstadter e Forbus dizendo que

ambos estão tratando de aspectos diferentes do estudo da analogia, e que um modelo mais efetivo consideraria tanto a adaptação como a pré-estruturação.

Gentner (1989) discute que o mapeamento é o processo geral que está no núcleo da analogia. Neste sentido, a noção de analogia inclui representações distintas e o mapeamento entre elas. Este mapeamento indica a relação entre os dados de duas representações. Segundo Gentner, no contexto da cognição, a analogia inclui o mundo real, a representação mental e sua intencionalidade. Nesta visão, é possível incluir os modelos de Hofstadter e de Forbus. A discordância é sobre a formulação da representação e não sobre os processos de mapeamento.

Utilizando-se do modelo de Gentner (1989), neste trabalho consideramos que, na representação computacional de sinais sonoros, a analogia inclui o sinal sonoro original, a representação computacional e a relação entre elas. Neste caso, o mapeamento está sujeito às imperfeições dos mecanismos de captura, como as distorções de microfones, ou da própria natureza da representação, como a quantização do PCM ou a resolução das trilhas de um disco de vinil, por exemplo. Para Holyoak & Thagard (1995), em um contexto de representações mentais, analogia "comprime cada situação do mundo real em uma caricatura pequena e congelada dela mesma, contendo precisamente seu núcleo e algo mais."[2]. O ponto é que, na analogia, o mapeamento não precisa ser perfeito e nem a representação precisa cobrir todos os aspectos da outra parte.

3.2.1 Analogia no processo de síntese sonora

Neste trabalho, a analogia é vista como o meio pelo qual o usuário se relaciona com a representação computacional. De forma restrita, no estudo que se desenvolve, parte-se do princípio de que qualquer representação, quando utilizada em um processo de síntese, torna-se análoga a uma representação cognitiva se o processo de modelagem for capaz de conciliar as discrepâncias entre o modelo de processamento e a idéia de sonoridade do usuário. Ainda de forma mais restrita, para

[2] *"to hand shrink each real-world situation into a tiny, frozen caricature of itself, containing precisely its core and little else"*

tornar possível um modelo, o que diferencia e qualifica duas representações computacionais é a facilidade com que o usuário estabelece analogias sonoras que se coadunam com o objetivo sonoro. Assim, a arquitetura da representação de descrição deve buscar a maior facilidade de estabelecimento de analogia.

A partir do princípio de que analogia é o mecanismo pelo qual se realiza a tradução entre a representação cognitiva e a computacional, o modelo representacional de síntese apresentado na Figura 3.5 se mostra como um caso específico do modelo proposto da Figura 3.6, em que a representação de descrição possui um comportamento físico idêntico aos parâmetros do sintetizador. Neste caso, a descrição do sinal sonoro se dá por meio dos parâmetros do sistema. O estabelecimento de analogia aqui é complexo e depende da capacidade do usuário de relacionar o comportamento dinâmico dos parâmetros com as qualidades da sonoridade desejada.

A proposta desta dissertação é a dissociação da representação descritiva dos parâmetros do sintetizador, com base em uma arquitetura de representação que permita o fácil estabelecimento da analogia entre o sinal sonoro e o meio de representação. Assim, são duas as tarefas para a formação do sistema proposto: (1) a arquitetura da representação descritiva; e (2) o mapeamento desta representação nos parâmetros do sintetizador.

No Capítulo 4, apresentam-se ferramentas computacionais capazes de lidar com a natureza do problema de mapeamento. Como descrito anteriormente, trata-se de um processo exploratório, que não presume a existência de uma solução ótima global e sim de um espaço de soluções para serem utilizadas de forma criativa pelo usuário. Neste sentido são úteis métodos de Computação Natural, apresentados no Capítulo 4, como: computação evolutiva, redes neurais artificiais e sistemas imunológicos artificiais.

No Capítulo 5, apresenta-se o primeiro modelo no qual busca-se uma aplicação da relação entre parâmetros de síntese e sonoridade. Utiliza-se um conceito de analogia acústica em que amostras sonoras são utilizadas como representação computacional e como referência de controle

de um processo de otimização. Em uma implementação formalizada e validada por meio de testes subjetivos, utiliza-se um som para carregar informação da evolução dinâmica global do sinal e um conjunto de sons (pequenos grãos sonoros) para representar a diversidade de uma população de sonoridades.

No Capítulo 6, faz-se uma outra relação entre analogia e objetivo sonoro. Neste caso, utiliza-se um conjunto de "figura semânticas" que carregam a noção de sonoridade. Em uma implementação chamada "Acusmata", redes neurais são utilizadas para mapear o espaço paramétrico instantâneo de um mecanismo de síntese aditiva, acoplado à síntese subtrativa, em um espaço regido por eixos relacionados às figuras semânticas.

No primeiro modelo, a sonoridade de um som é relacionada de forma direta com as amostras pertencentes a uma população, enquanto, no segundo, a sonoridade é relacionada de forma indireta, utilizando-se linguagem natural para construir as analogias entre o controle e a representação computacional. O Apêndice A traz um tutorial ilustrativo, simplificado e didático dos tipos de analogia empregados no trabalho.

Capítulo 4

Métodos Computacionais

Este capítulo apresenta os métodos computacionais aplicados nos modelos de síntese desta dissertação. São técnicas que se incluem no campo da computação natural por possuírem inspiração em conceitos e analogias provenientes da biologia. Uma assimilação parcial dos paradigmas apresentados nos próximos capítulos é possível sem o entendimento desses métodos. Contudo, os resultados obtidos certamente estão sustentados em atributos matemáticos e computacionais apresentados por estas técnicas. Assim, este capítulo apresenta o núcleo de processamento empregado neste trabalho.

4.1 Computação natural e inspiração biológica

Computação natural é definida por de Castro (2007) como uma terminologia introduzida para englobar três classes de métodos: (1) aqueles baseados na natureza para o desenvolvimento de novas técnicas computacionais de resolução de problemas; (2) aqueles que se baseiam no uso do computador para simular e reproduzir artificialmente fenômenos naturais; e (3) aqueles que aplicam materiais naturais para realizar computação. As técnicas principais são: redes neurais, computação evolutiva, inteligência coletiva, sistemas imunológicos artificiais, geometria fractal, vida artificial, computação por DNA e computação quântica.

A motivação principal das técnicas incluídas na classe (1) é a de prover meios de solução

ou alternativas aos mecanismos tradicionais de solução de problemas que não possuam uma resolução satisfatória, como por exemplo programação linear, não-linear e dinâmica (de Castro & Von Zuben, 2005). Estas soluções são comumente denotadas como heurísticas ou meta-heurísticas (Kochenberger & Glover, 2003). Neste capítulo, são abordadas três destas técnicas: redes neurais artificiais, inspiradas no sistema nervoso dos vertebrados; computação evolutiva, baseada na teoria de seleção natural de Darwin; e sistemas imunológicos artificiais, inspirados nos mecanismos de defesa de organismos superiores. No contexto deste trabalho, os mapas auto-organizáveis, Seção 4.2.3, algoritmos genéticos, Seção 4.4, e sistemas imunológicos artificiais, Seção 4.3, são utilizados como mecanismos de "quantização vetorial", Seção 5.3.5, no Capítulo 5. "Rede neural perceptron de múltiplas camadas"[1], Seção 4.2.1, é aplicada no Capítulo 6 na tarefa de aproximação de uma função de mapeamento.

4.2 Redes neurais artificiais

Os trabalhos envolvendo redes neurais artificiais (RNAs) se motivam pelo fato do cérebro humano processar informações de forma diferente dos paradigmas tradicionais de computação. O sistema nervoso, que inclui o cérebro, é capaz de realizar tarefas, tais como reconhecimento de padrões e controle motor de forma rápida através de um processamento altamente complexo, não-linear e paralelo (Haykin, 2001). Outro fator relevante é que o sistema nervoso está em constante modificação e atualização, sendo capaz de realizar aprendizado e organizar informação na forma de conhecimento.

Dentre as funcionalidades do cérebro se incluem: adaptabilidade por intermédio de aprendizado; comportamento sensível ao contexto; tolerância a erro; capacidade de operar com conhecimento parcial; grande capacidade de memória; capacidade de processamento paralelo e em tempo real. O cérebro é capaz de perceber regularidades no meio e gerar abstrações que capturam a estrutura destas regularidades, possibilitando a predição de observações futuras e o planejamento de ações visando o atendimento de múltiplos objetivos. Haykin (2001) cita que o uso de redes neurais

[1] *"Multilayer Perceptron."*

possibilita, dentre outras, as seguintes propriedades: mapeamento multidimensional de entrada-saída, através de aprendizado supervisionado, com aplicação em aproximação de funções multivariáveis e classificação de padrões; não-linearidade, sendo capaz de reproduzir não-linearidades arbitrárias em um mapeamento; adaptabilidade, por ser possível modificar sua composição frente a novos estímulos externos ou ao contexto; tolerância a falhas, de modo que a danificação de um componente da rede geralmente exerce pouca influência sobre o desempenho global.

O ponto de partida para o estudo das RNAs está associado à primeira modelagem matemática do neurônio, unidade básica de processamento do cérebro, por McCulloch & Pitts (1943). No cérebro, os neurônios são interconectados em redes através de ligações chamadas sinapses. No modelo da Figura 4.1, o neurônio é constituído por N entradas ($e_1, e_2, e_3, \ldots, e_N$) e seus respectivos pesos sinápticos ($p_1, p_2, p_3, \ldots, p_N$), um bias (b), um nível de ativação interna (v), uma função de ativação ($\varphi(\cdot)$) e uma saída (s). A saída é calculada a partir das Equações 4.1 e 4.2.

$$v = \sum_{i=1}^{N} e_i p_i + b \tag{4.1}$$

$$s = \varphi(v) \tag{4.2}$$

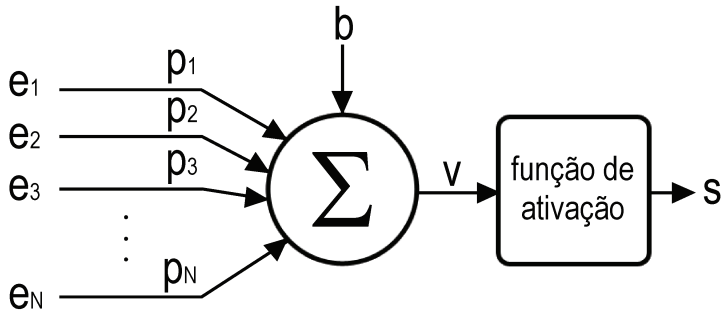


Fig. 4.1: Modelo de neurônio de McCulloch-Pitts.

A função de ativação faz a conversão entre o acumulado ponderado da entrada, denominado campo local induzido no neurônio (v na Equação 4.1), e a saída. McCulloch e Pitts utilizavam uma função limiar (também conhecida como função de Heaviside) definida na Equação 4.3 e com gráfico na Figura 4.2. Esta função retorna 0 para ativações internas negativas, e 1 para ativações internas nula ou positivas:

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases} \tag{4.3}$$

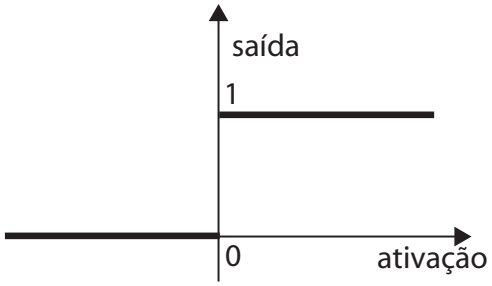


Fig. 4.2: Função de ativação limiar ou de *Heaviside*

Outra função é a linear por partes, Equação 4.4 e Figura 4.3, que cria uma transição linear entre os valores 0 e 1:

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq +\frac{1}{2} \\ v + \frac{1}{2} & \text{se } +\frac{1}{2} > v > -\frac{1}{2} \\ 0 & \text{se } v \leq -\frac{1}{2} \end{cases} \tag{4.4}$$

Visando suavizar ainda mais a transição entre os dois níveis de saturação da saída, é geralmente adotada a função sigmóide, que transita de 0 a 1 em uma curva em formato de s. Um exemplo é a função logística, Equação 4.5 e Figura 4.4, onde o parâmetro a é normalmente tomado com valor unitário. Estas funções emitem saída limitada ao intervalo $[0, 1]$, mas é comum a aplicação do intervalo $[-1, 1]$, com poucas modificações nas funções de ativação. No caso de função sigmóide, um exemplo que produz saída no intervalo $[-1, 1]$ é a função tangente hiperbólica ($\tanh$).

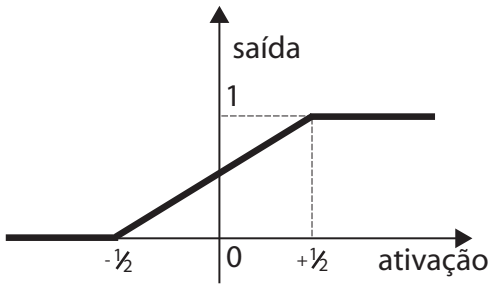


Fig. 4.3: Função de ativação linear por partes.

$$\varphi(v) = \frac{1}{1 + \exp(-av)} \tag{4.5}$$

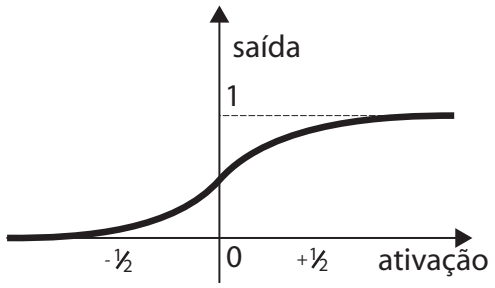


Fig. 4.4: Função de ativação sigmóide com $a = 1.2$.

Uma rede neural é uma organização de vários neurônios interconectados. A forma como a rede é estruturada é chamada arquitetura de rede. Existem basicamente dois tipos de arquitetura: alimentadas adiante e recorrentes. Nas arquiteturas alimentadas adiante, todas as ligações sinápticas ocorrem de um neurônio a outro de forma que nem sua saída nem a de nenhum adjacente sejam ligadas em sua entrada, i.e., não há retroalimentação. Redes recorrentes são caracterizadas pela presença de laços de retroalimentação.

As redes alimentadas adiante são, em geral, organizadas em camadas, podendo ser de camada única ou de múltiplas camadas. A Figura 4.5 apresenta um exemplo de rede alimentada adiante de uma camada. A rede, organizada em uma única camada, faz um mapeamento direto entre o conjunto

de dados de entrada e a saída. Este exemplo possui 4 entradas e 4 neurônios em sua única camada. A Figura 4.6 apresenta uma rede alimentada adiante de múltiplas camadas, tendo 4 entradas, 5 neurônios na camada intermediária e 2 na camada de saída. Esta arquitetura de rede é conhecida como *MultiLayer Perceptron* (MLP). A Figura 4.7 apresenta uma rede recorrente com uma única camada de neurônios.

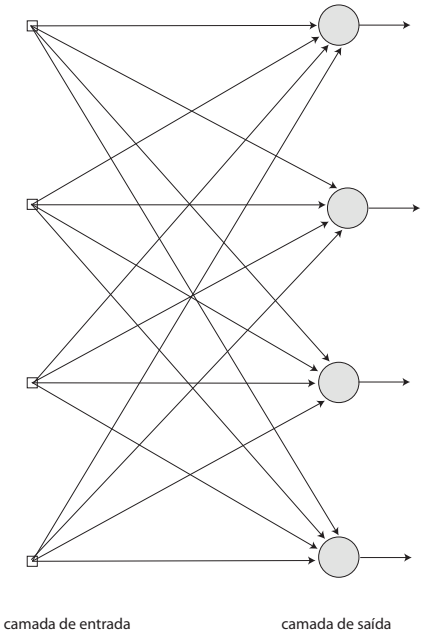


Fig. 4.5: Rede neural alimentada adiante de camada única.

Além das supracitadas arquiteturas tradicionais, existem estratégias para o desenvolvimento construtivo de arquiteturas de rede, como, por exemplo, "aprendizado por busca de projeção"[2] (Hwang et al., 1994) e também propostas que envolvem poda[3] de componentes (Reed, 1993). Programação genética pode ser utilizada no desenvolvimento automático de propostas de arquiteturas de RNAs (Fogel, 1993; Yao & Liu, 1997). Em Angeline et al. (1994), algoritmos genéticos são utilizados para o projeto automático de redes recorrentes. Existem ferramentas para a síntese de arquiteturas customizadas, como o IQR (Bernardet et al., 2002), que permite o desenvolvimento de

[2] *Projection Pursuit Learning*
[3] *pruning*

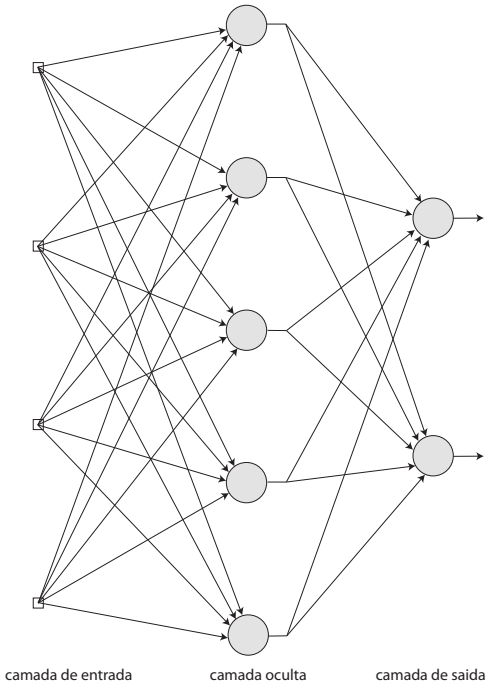


Fig. 4.6: Rede neural alimentada adiante de múltiplas camadas.

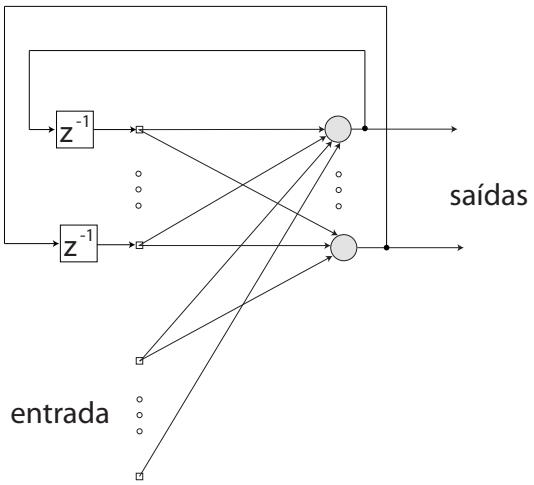


Fig. 4.7: Rede neural recorrente.

arquiteturas complexas para aplicações em tempo real.

Uma vez definida a arquitetura, o passo seguinte é proceder com o aprendizado a partir de estí-

mulos externos. O aprendizado consiste em estabelecer os pesos sinápticos, no caso de arquiteturas fixas, ou estabelecer a própria arquitetura e seus pesos sinápticos, no caso em que a topologia é ajustável. Haykin (2001) cita cinco mecanismos de aprendizado: por correção de erro, baseado em memória, hebbiano, competitivo e de Boltzmann. Estes aprendizados podem ser classificados em supervisionado ou não-supervisionado. O treinamento supervisionado fornece como estímulo externo a entrada e a saída desejada e o algoritmo de treinamento utiliza o erro entre a saída obtida e a desejada como controle. Nos algoritmos não-supervisionados, a base é a auto-organização. Auto-organização se dá a partir de estímulos externos e regras de interação local, de modo que a rede neural seja capaz de assimilar propriedades estatísticas presentes nos estímulos externos, na forma de padrões de organização.

4.2.1 Rede alimentada adiante de múltiplas camadas

As redes neurais do tipo perceptron alimentadas adiante e de múltiplas camadas, ou MLP[4], são aplicadas basicamente para mapeamentos entrada-saída a partir de treinamento supervisionado. As aplicações principais são a aproximação de funções multidimensionais, a partir de dados amostrados, e classificação de padrões.

O procedimento de aprendizagem consiste em apresentar entradas à rede e modificar os pesos das conexões sinápticas de forma a minimizar o erro entre as correspondentes saídas obtidas e as desejadas. Estes procedimentos se diferenciam basicamente na estratégia de apresentação de dados e na heurística de alteração dos pesos.

Com relação à apresentação dos dados, ela pode ocorrer por batelada, em que todos os dados são apresentados e apenas após isso os pesos são alterados, ou em seqüência, sendo os pesos alterados a cada amostra apresentada. Em ambos os casos, varios ciclos de apresentação de todas as amostras são necessários para concluir o processo de treinamento.

[4] *"Multilayer Perceptron."*

Já relativo às técnicas de modificação dos pesos, a informação utilizada é o erro entre a saída obtida e a desejada frente a uma entrada específica e suas derivadas em relação aos pesos. Assim, o treinamento se assemelha a um processo de otimização não-linear e irrestrita (Von Zuben, 1996).

de Castro (1998) apresenta uma taxonomia para os mecanismos de aprendizado para redes neurais artificiais, os dividindo em quatro grupos: métodos sem diferenciação, métodos de primeira ordem, métodos de segunda ordem e métodos empíricos. Métodos sem diferenciação utilizam a métrica de erro de forma direta, como o caso de algoritmos genéticos (van Rooji et al., 1996) e *simulated annealing* (Boese & Kahng, 1993). Os métodos de primeira ordem utilizam a primeira derivada do erro. Neste grupo, inclui-se o método de retropropagação[5] associado ao gradiente descendente, mais tradicional e historicamente relevante (Rumelhart et al., 1986). Os métodos de segunda ordem utilizam a derivada segunda do erro, constituindo o grupo de maior aplicação. O último grupo constitui as técnicas com pouca formalização e geralmente não são considerados em teoria de otimização clássica.

4.2.2 Treinamento supervisionado e validação

A validação interfere sobre o momento de parada do treinamento supervisionado. Uma vez que normalmente o grupo de amostras de treinamento não representa de forma plena o mapeamento a ser aproximado, o excesso de treinamento pode causar um mapeamento viciado a estes dados. A Figura 4.9 ilustra o problema do sobre-treinamento. As bolas representam as amostras utilizadas no treinamento, enquanto as estrelas são amostras que pertencem à função a ser mapeada, mas que não são utilizadas durante o treinamento. O treinamento tendencioso acerta todos os valores do conjunto de bolas, mas possui grande distância das estrelas. Isso implica em uma generalização pobre. O segundo mapeamento possui certo erro em relação às bolas, porém generaliza melhor, ou seja, apresenta um bom comportamento junto às estrelas, as quais não foram utilizadas durante o treinamento.

Uma estratégia para implementar a validação requer a separação dos dados de entrada em um grupo de treinamento e outro de validação. A rede é treinada com o primeiro grupo e o treinamento é

[5] *backpropagation*

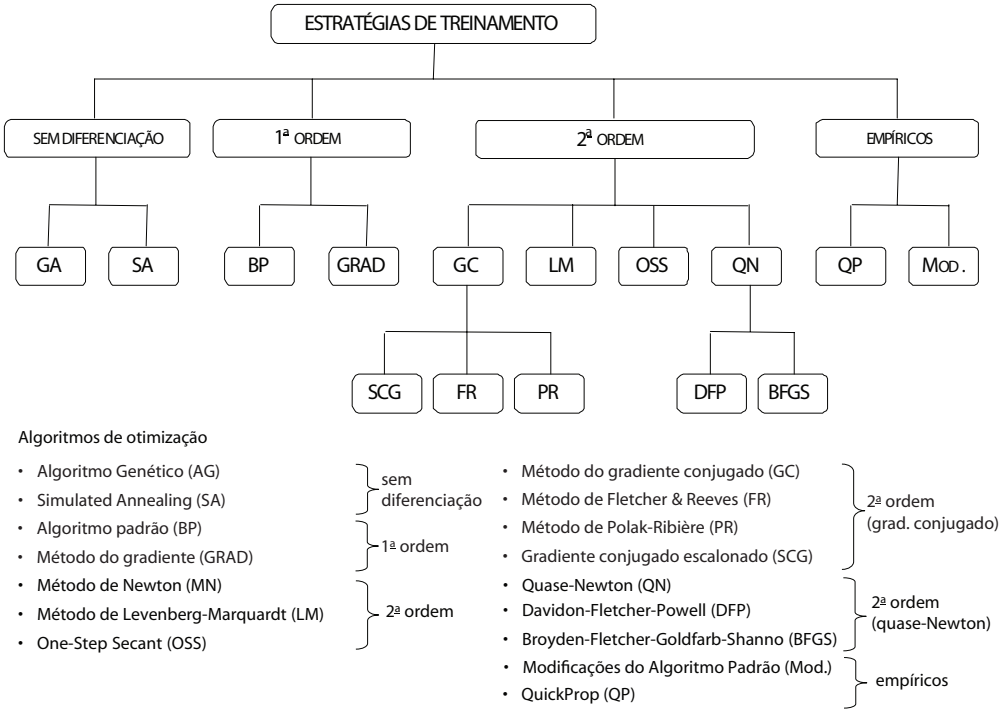


Fig. 4.8: Topologia de métodos de aprendizado para MLP (de Castro, 1998).

interrompido levando em consideração uma medida de generalização relacionada ao segundo grupo. Um dos problemas desta abordagem é a diminuição do número de amostras para o treinamento. Validação cruzada consiste em dividir os dados de forma que todos possam ser utilizados no treinamento e na validação, num processo de reamostragem dos conjuntos de treinamento e de validação. Goutte (1997) discute o ganho de performance pelo uso de validação cruzada. Uma das técnicas é chamada de *k-fold*, em que o conjunto de dados é dividido em k grupos: a cada iteração, é separado um grupo para a validação enquanto os outros são utilizados para o treinamento. Outro método é chamado de *leave-one-out*: consiste em remover uma amostra e realizar um treinamento, em seguida, calcula-se o erro relativo a esta amostra. Realiza-se o mesmo para todas as amostras e calcula-se o erro médio. Esta e outras técnicas que criam restrições de suavidade para aproximações de mapeamentos de entrada-saída a partir de um conjunto finito de amostras são conhecidas como regularização (Tikhonov, 1963).

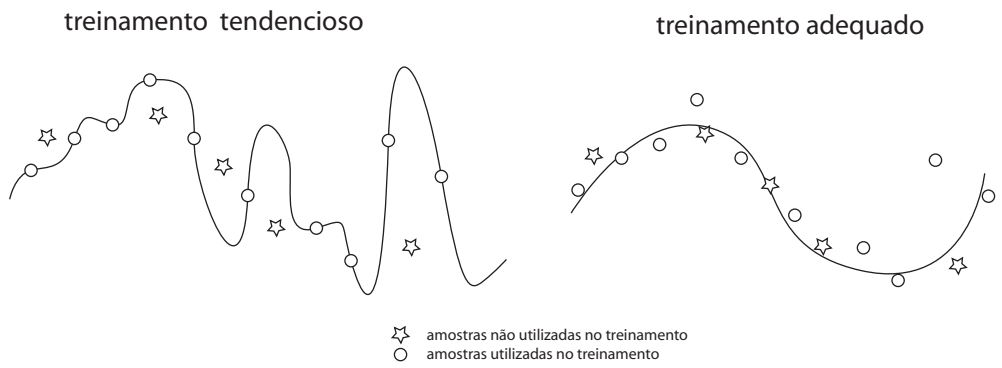


Fig. 4.9: Ilustração de um mapeamento tendencioso por excesso de treinamento e um mapeamento suave.

4.2.3 Mapas auto-organizáveis

Um dos princípios dos caminhos sensoriais do cérebro é que o posicionamento de neurônios é ordenado e comumente reflete algumas características físicas dos estímulos sentidos (Kandel & Schwartz, 1985). Por exemplo, em cada nível do caminho auditivo, células nervosas e fibras são arranjadas anatomicamente em relação à freqüência que elicia a maior resposta em cada neurônio. Esta organização se estende até o córtex auditivo (Moller, 1983). Apesar de grande parte da organização de baixo nível ser pré-definida geneticamente, parte da organização em níveis mais elevados é criada durante o aprendizado, por algoritmos que promovem auto-organização. Kohonen (2000) apresenta estes algoritmos, que produzem o que ele chama de mapas auto-organizáveis, os quais apresentam algum grau de equivalência com aqueles que ocorrem no cérebro.

Os mapas auto-organizáveis (MAOs) de Kohonen associam dados de alta-dimensionalidade em uma grade de baixa dimensão, caracterizando um quantizador vetorial (Lippmann, 1987). Nós de saída são interconectados localmente. Baseado nos princípios organizacionais dos neurônios, nós topologicamente próximos são sensíveis a estímulos fisicamente similares. Conseqüentemente, os nós de saída são ordenados naturalmente, sem interferência externa, em um processo chamado de aprendizado não-supervisionado. Depois de repetidas apresentações do conjunto de entrada, as posições dos nós de saída irão especificar clusters ou centros de vetores que amostram o espaço de entrada, de forma que a função densidade de probabilidade dos centros de vetores tende a se aproximar da função densidade de probabilidade dos vetores de entrada (Kohonen, 1984b).

A arquitetura de um mapa, ilustrada na Figura 4.10 em uma configuração bidimensional, é um conjunto de nós ou neurônios m_i com um conjunto de pesos sinápticos $[m_{i,1} \dots m_{i,D}]$ associado, de mesma dimensão que um vetor de dados de entrada $[v_1 \dots v_D]$. Entre os nós, existe uma relação de vizinhança, ver Figura 4.11. A ordem de distância entre um nó e outro é o número mínimo de nós no menor caminho entre eles.

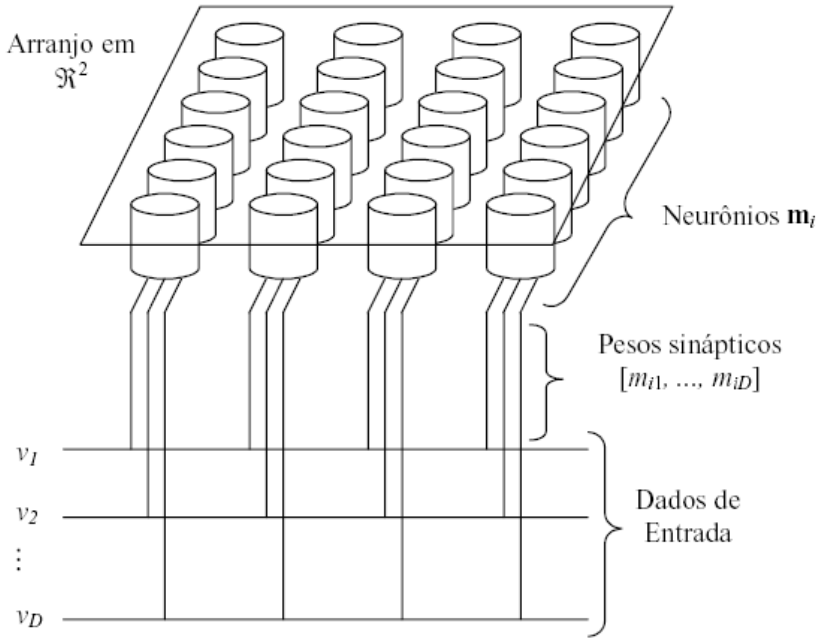
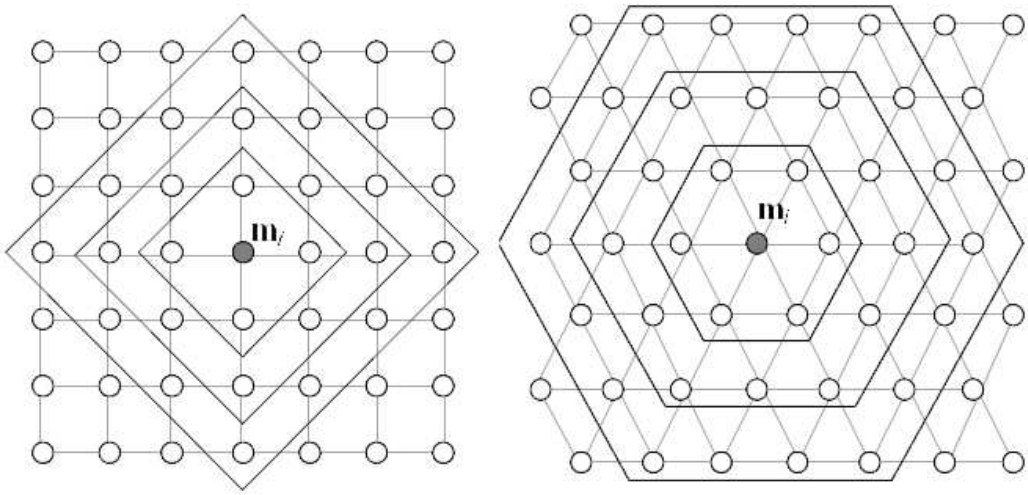


Fig. 4.10: Ilustração da arquitetura de um MAO bidimensional. Extraído, com permissão do autor, de Zuchini (2003).

O processo de treinamento, ou seja, o ajuste dos pesos sinápticos, se dá pela apresentação repetida do conjunto de dados de entrada. A atualização dos pesos sinápticos pode ocorrer na apresentação de cada dado novo ou em batelada, após a apresentação de todo o conjunto, evento denominado época de treinamento. Cada época é composta por três fases: Fase competitiva, Fase Cooperativa e Fase de Aprendizado. A Figura 4.12 apresenta um fluxograma do algoritmo.

Na fase competitiva, o nó com menor distância ao dado de entrada é escolhido para aprender, ele é denominado *best matching unit* (BMU). Na fase cooperativa, os nós vizinhos àquele rotulado como



(a) **(b)**

Fig. 4.11: Diferentes relações de vizinhança em arranjos bidimensionais:(a) retangular; (b) hexagonal. Conexões claras entre nós definem nós vizinhos. Linhas escuras definem a ordem de vizinhança em relação ao nó m_i. Extraido, com permissão do autor, de Zuchini (2003).

BMU também terão seus pesos sinápticos ajustados, mas a uma taxa inferior, proporcional ao inverso da distância no mapa. Por exemplo, essa taxa de ajuste pode obedecer a uma função de vizinhança que é monotonicamente decrescente com o distanciamento da BMU (Kohonen, 1984b). A última fase é a do aprendizado. Os pesos dos nós são atualizados de acordo com a Equação 4.6.

$$\omega_j[n+1] = \omega_j[n] + \eta[n] \times h_{j,i(x)}[n] \times (x - \omega_j[n]) \tag{4.6}$$

A cada época n, o peso ω de cada nó j é modificado na direção do vetor de dados de entrada x. O grau de aprendizado é dado pelo produto da taxa de aprendizado global η, relativa à época, e da função de vizinhança h da BMU $i(x)$, considerando o nó sendo atualizado. A taxa de aprendizado global e a dispersão da função de vizinhança decaem com o tempo (Gomes & Von Zuben, 2003). Esta política garante dois estágios diferentes no desenvolvimento do mapa: um começo com taxas de

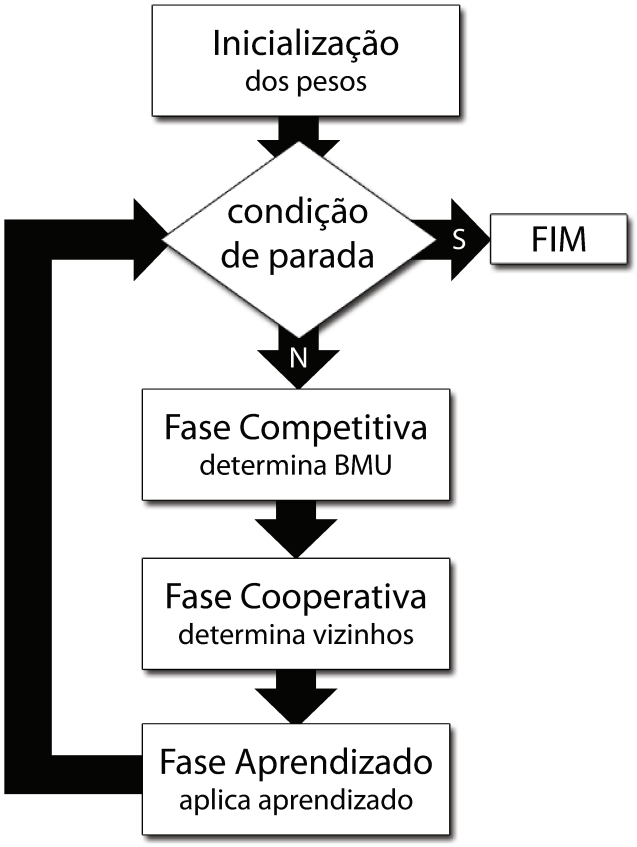
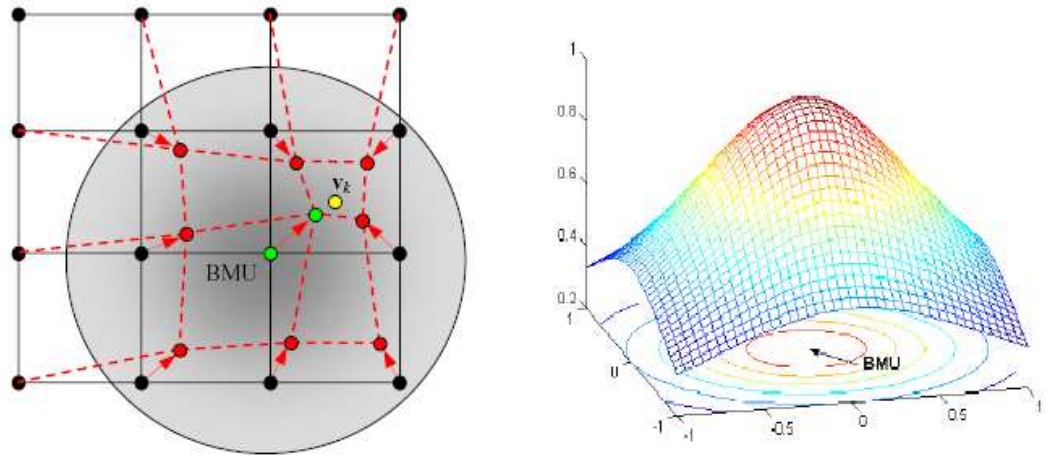


Fig. 4.12: Fluxograma do algoritmo de aprendizado do MAO.

aprendizado mais elevadas e função de aprendizado mais abrangente, implicando num aprendizado menos rebuscado e sem tendência à convergência; um final com sintonia fina, taxa de aprendizado mais baixa e vizinhança mais localizada, visando a convergência do ajuste.

A Figura 4.13a ilustra graficamente, em um cenário bidimensional, o processo de aprendizado vetorial, considerando uma amostra V_k de entrada. A BMU se aproxima mais do vetor de entrada V_k, com os vizinhos se aproximando em um grau menor, indicado pela função de vizinhança da Figura 4.13b.



(a) **(b)**

Fig. 4.13: (a) Ilustração do processo de aprendizado vetorial, (b) BMU e curva de grau de aprendizado. Extraído, com autorização do autor, de Zuchini (2003).

4.3 Sistemas imunológicos artificiais

Sistemas Imunológicos Artificiais (SIAs) são procedimentos adaptativos inspirados pelos sistemas imunológicos biológicos e voltados para a solução de problemas computacionais (de Castro & Timmis, 2002). Esta definição está coerente com aquela de Dasgupta (1999), que apresentou os SIAs como "uma composição de metodologias inteligentes, inspiradas por sistemas imunológicos naturais, para a resolução de problemas do mundo real"[6]. A aplicação que pode ser vislumbrada de forma mais direta é a de sistemas de segurança em redes de computadores, baseada em aspectos da Teoria de Rede Imunológica (Jerne, 1974).

A função básica do sistema imunológico biológico é proteger o organismo da ação de antígenos, os quais podem ser definidos como agentes causadores de doenças. O sistema atua identificando um antígeno e sinalizando quimicamente para a ação de mecanismos internos de combate. Esta identificação é realizada a partir da comparação de padrões protéicos e a sinalização é realizada por um tipo de proteína denominada anticorpo, produzido pelas células B. Anticorpos se ligam ao antígeno, atraindo células especiais para a sua eliminação, denominadas macrófagos.

[6] *"a composition of intelligent methodologies, inspired by the natural immune system for the resolution of real world problems"*

O problema maior da identificação por antígenos é que se estima que existam 10^{16} variedades de padrões possíveis. Assim, é impossível manter uma estrutura de identificação individual. É necessário, então, que cada anticorpo seja capaz de se ligar a um grupo de antígenos, em um processo denominado identificação parcial, e que existam mecanismos de maturação de afinidade, permitindo a produção de anticorpos com maior capacidade de se ligar a cada antígeno.

Para realizar este processo com sucesso, o sistema deve possuir três capacidades básicas:

- **Diversidade:** de criar anticorpos para qualquer tipo de antígeno;
- **Memória:** de lembrar quais anticorpos obtiveram sucesso com quais antígenos;
- **Tolerância:** de diferenciar antígenos de elementos do próprio organismo e/ou que não causam doenças.

A essência do mapeamento antígeno-anticorpo está na população de células de identificação de antígenos e produção de anticorpos. A capacidade de diversidade é possível pelo fato destas células serem capazes de gerar proteínas não baseadas na sua estrutura de DNA. A geração de células que identificam um dado antígeno é possível devido à seleção clonal (Burnet, 1959). Neste processo, na presença de um antígeno, as células de identificação são clonadas, com pequenas variações no antígeno a ser identificado. As células capazes de identificá-los com maior eficiência obtêm prioridade em posteriores clonagens. Estas células têm vida curta e por isso atuam de forma específica em uma invasão. A memória é obtida a partir de uma população de células imaturas de vida longa. Estas células são os pontos iniciais do processo de clonagem citado anteriormente. Na presença de um antígeno novo, elas são geradas ocasionalmente a partir de células que sejam sucessivamente clonadas em um processo denominado "Maturação de Afinidade" (Nossal, 1993). Estes dois processos são a base do modelo de representação dos sistemas imunológicos.

A capacidade dos sistemas imunológicos naturais de reconhecer uma ampla gama de antígenos com um número reduzido de anticorpos inspira uma metáfora computacional em que uma população de anticorpos digitais representa um grupo mais extenso de antígenos digitais. Esse mecanismo de

obtenção de protótipos caracteriza um processo de quantização vetorial. Na metáfora computacional, os protótipos representativos correspondem aos anticorpos, enquanto que os dados a serem representados correspondem aos antígenos, sendo que o formalismo matemático está vinculado à concepção de um "espaço de formas"[7] (Perelsen & Oster, 1979).

A rede imunológica artificial (aiNet[8]) (de Castro & Von Zuben, 2001) é uma aplicação computacional de identificação de grupos (clusterização) do paradigma de representação dos sistemas imunológicos. O reconhecimento entre os anticorpos e os antígenos é devido a um processo competitivo de afinidade. A medida da afinidade é a norma da distância específica do espaço de formas em que se efetua a metáfora, em geral, euclidiano. Em seu processo competitivo, redes de anticorpos são semelhantes a mapas auto-organizáveis, exceto pela falta de vizinhança local para guiar as interações de anticorpos. Outra característica particular da aiNet é que o tamanho da população é auto-regulado: positivamente, por clonagem de seus elementos, e negativamente, por supressão. No treinamento, a competição entre os anticorpos implica a ativação da rede e ploriferação de células bem sucedidas, além da eliminação das que falham.

A Figura 4.14 apresenta um fluxograma da aiNet. O algoritmo é episódico, sendo interrompido por alguma condição de parada, sendo que o número máximo de épocas e a satisfação de alguma especificação de qualidade são exemplos. Na iteração, é realizado um ciclo para cada padrão: primeiramente, é calculada a afinidade entre cada protótipo da população e o padrão; a partir da população, são gerados clones, sendo que a quantidade de clones gerados por protótipo é proporcional ao seu grau de afinidade; a estes clones, é aplicada mutação (variação estocástica de posição) de grau inversamente proporcional à afinidade; a afinidade entre esses clones e o padrão é calculada; realiza-se a apoptose, processo no qual os clones com pouca afinidade (abaixo de um limiar de morte natural σ_d) são eliminados; é calculada a afinidade entre cada clone e os outros clones; realiza-se a supressão clonal, processo no qual clones com alta afinidade com outro clone (acima de um limiar de supressão σ_s) são eliminados; os clones resultantes são incorporados à população de protótipos. Uma vez aplicado o ciclo a todos os protótipos, é calculada a inter-afinidade dos protótipos, seguido

[7] *shape space*
[8] *Artificial Immune Network*

do processo de supressão de rede (similar à supressão clonal, utilizando o mesmo limiar).

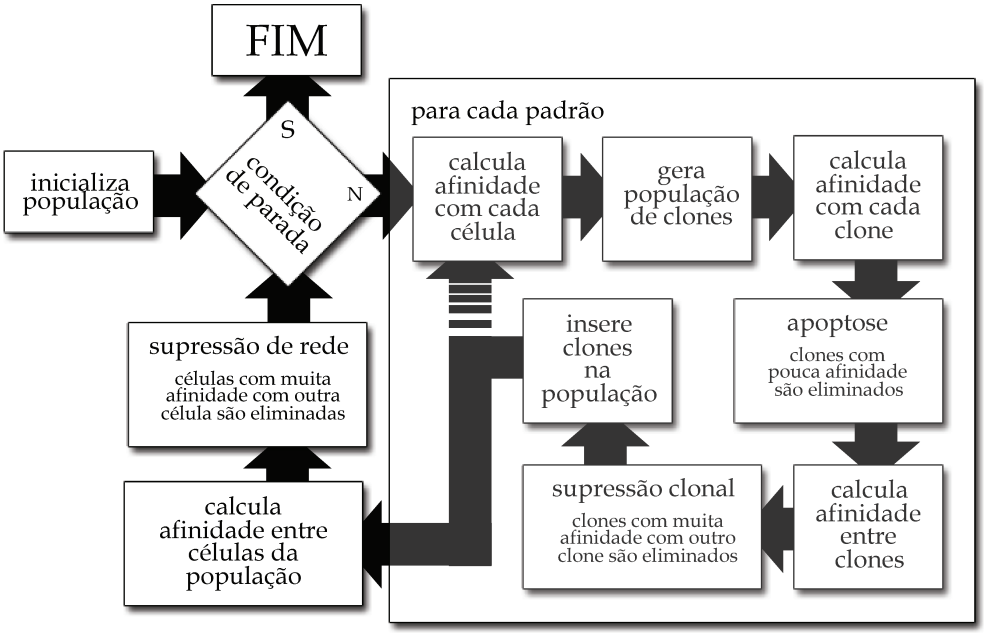


Fig. 4.14: Fluxograma da aiNet.

O limiar de supressão fornece um controle sobre o grau de especificidade da rede. Um número menor de protótipos, causados por um limiar baixo, implica em uma representação mais geral. Um limiar mais alto cria uma representação com um número maior de protótipos, mas que tende a ser mais precisa.

Outra aplicação de sistemas imunológicos, não explorada no contexto deste trabalho, é a chamada "imunologia computacional"[9] (Forrest et al., 1996), que se preocupa com outros paradigmas além da representação em si. Forrest define que um sistema deste tipo deve ter alguns componentes básicos: uma definição estável de ser; prevenção ou detecção e posterior eliminação de atividades externas perigosas (infecções); memória de infecções anteriores; métodos de detecção de novas infecções; meios de proteger o próprio sistema imunológico. Isto pode ser aplicado em sistemas operacionais e redes de computadores, tomando quatro tipos de arquiteturas (Forrest et al., 1997): proteção de dados estáticos; proteção de processos ativos em computador único; proteção de rede de computadores fixa; proteção de rede de computadores mutável. Um exemplo é LISYS (Forrest & Hofmeyr, 1999),

[9] *Computer Immunology*

sistema de detecção de intrusos em redes de computadores.

4.4 Algoritmos Genéticos

Algoritmos Genéticos (AG) representam um paradigma da computação evolutiva que consiste em um conjunto de técnicas computacionais baseadas na teoria de evolução de Darwin e no princípio de sobrevivência do mais apto. Dada uma população de indivíduos cujas propriedades físicas são codificadas em uma lista de atributos, operadores genéticos simples, como mutação, crossover e seleção, são repetidamente aplicados para produzir novas gerações. O grau de adaptação (*fitness*) de cada indivíduo na população é dado por uma função-objetivo. Os operadores genéticos promovem uma exploração paralela do espaço de busca com concentração de indivíduos nas regiões mais promissoras, i.e., regiões em que as amostras obtêm os valores de fitness mais elevados. Isso ocorre porque indivíduos com maiores índices de adaptação são favorecidos nas fases de reprodução, tendo uma maior probabilidade de propagar o seu material genético para gerações futuras.

A Figura 4.15 apresenta o fluxograma de um algoritmo genético padrão, propostos e descritos por Holland (1975) e analisados por Goldberg (1987). Inicia-se em uma época t, criando uma população $P(t)$ de um tamanho fixo e avaliando seus indivíduos. Em seguida, entra-se no ciclo incremental $t = t+1$, que cessa quando a condição de parada for satisfeita. Primeiramente são escolhidos, através de algum método de seleção estocástico seguindo uma taxa de reprodução definida, indivíduos para sofrerem mutação e recombinação, criando uma população temporária $C(t)$. Após a aplicação dos operadores genéticos, os novos indivíduos são avaliados. Por último, um processo de seleção recompõe a população $P(t)$ a partir dos elementos da população anterior $P(t-1)$ e da temporária $C(t)$, utilizando alguma heurística de seleção.

Mutação é um operador genético no qual um indivíduo é modificado para gerar outro indivíduo diferente, em geral modificando aleatoriamente algum valor da codificação. Recombinação, ou *crossover*, é um operador genético no qual uma dupla de indivíduos é utilizada para gerar um novo

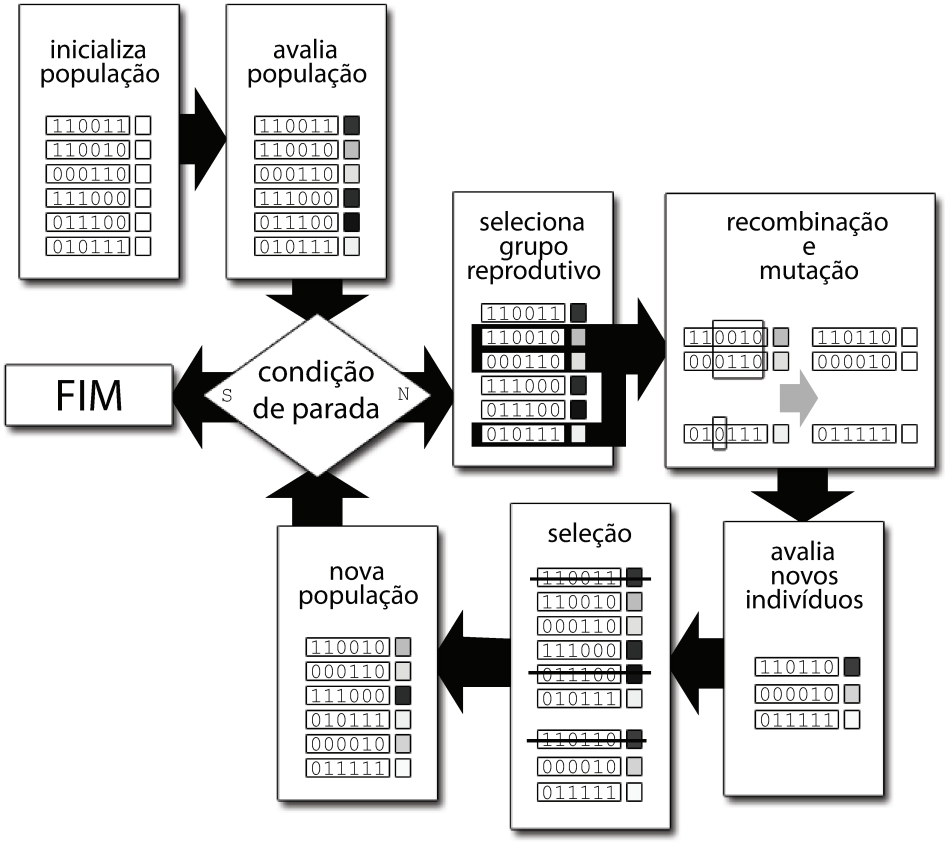


Fig. 4.15: Fluxograma de um algoritmo genético.

indivíduo. O funcionamento mais comum, chamado "recombinação de um ponto", estabelece um ponto aleatório na codificação e cria um novo indivíduo pela combinação da parte anterior ao ponto do primeiro pai com a porção posterior do segundo pai. Outra técnica é denominada "recombinação uniforme", em que o valor de cada posição do novo indivíduo é escolhida por uma decisão estocástica de distribuição uniforme do valor da mesma posição nos pais. Na aplicação utilizada neste trabalho, no Capítulo 5, o operador de recombinação é uma combinação convexa, com índice escolhido randomicamente, dos pais. Este operador é muito utilizado quando a codificação dos indivíduos emprega números reais (Michalewicz, 2004).

4.5 Aplicação em Síntese Sonora

A aplicação destes métodos em síntese sonora pode não ser clara à primeira vista. Meta-heurísticas são geralmente aplicadas em tarefas de busca e exploração de espaços de alta dimensionalidade e comportamento não-linear, mas cuja principal característica é o formalismo de sua definição. Caso este da aproximação de funções por redes neurais e otimização combinatória por algoritmos genéticos. Uma das principais características é a capacidade destes métodos de lidar com aprendizado a partir de dados, muitas vezes repletos de incompletude de amostragem, incerteza e inconsistência.

Como discutido na Seção 2.2.2 do Capítulo 2, o procedimento de síntese pode ser modelado como um processo de busca em um espaço de sons. Processo esse marcado pela interferência humana, fonte potencial de incompletude e inconsistência de definição do espaço e o que caracteriza ainda o fato do processo de busca modificar o próprio espaço, sendo explorado. Assim, existe um casamento entre as motivações de uso dos métodos bio-inspirados e as necessidades para o controle do processo de síntese sonora.

As técnicas apresentadas já foram utilizadas em mecanismos de síntese. Mapas auto-organizáveis são aplicados em contextos voltados para agrupamento e extração de informações em conjuntos extensos de dados, como em aplicações envolvendo áudio, principalmente na busca de taxonomias de timbre (De Poli & Tonella, 1993; Cosi et al., 1994b,a; Feiten & Gunzel, 1994; Damiani et al., 1995; De Poli & Prandoni, 1997; Loureiro et al., 2004). Kohonen (1984a), em uma das suas primeiras aplicações, apresenta como seus mapas auto-organizáveis atuam como quantizadores vetoriais em uma aplicação de reconhecimento de fala. Caetano et al. (2005a) apresenta uma proposta para montar um espaço topológico de controle de síntese sonora a partir de amostras utilizando um mapa auto-organizável.

O algoritmo da aiNet já foi utilizado em aplicações sonoras de síntese de timbres por Caetano et al. (2005b). Algoritmos genéticos foram utilizados em um contexto de síntese sonora e composição por, entre outros, Horner (1995), Yuen & Horner (1997),Fornari et al. (2001), Caetano et al. (2005c),

Santarosa et al. (2006) e Miranda & Biles (2007). Outros métodos incluem autômatos celulares (Miranda, 1995) e fractais (Polotti & Evangelista, 2001).

Capítulo 5

Representação por analogia acústica

Representação por analogia acústica consiste no uso de amostras sonoras como paradigma de descrição de objetivos. Este tipo de representação já é utilizado para busca em coleções de mídias em um método chamado "Busca por similaridade de conteúdo". Este é um bom ponto de partida para a aplicação em uma metodologia de síntese sonora.

5.1 Busca por similaridade de conteúdo

Mecanismos de busca em uma base de dados apresentam uma demanda similar àquela do controle de síntese sonora: a descrição dos objetivos. Logo, servem de um bom ponto de partida na busca de novos paradigmas de controle para síntese sonora.

O modelo de busca em bases de dados tradicional é baseado no uso de palavras-chave textuais. A cada objeto da coleção são atribuídas palavras-chave e a busca é realizada no domínio que as compreende. Modelos mais complexos de bases de dados, como o modelo relacional (Codd, 1970), permitiram a busca por descrição de atributos. No modelo relacional, cada objeto possui uma série de entradas, ou tuplas, em tabelas de atributos. Estas entradas definem qualitativamente o dado. A busca é feita pela restrição de atributos, sendo que os dados que se enquadram nestas restrições são retornados ao usuário.

Em coleções multimídia, meta-dados são utilizados como atributos. Meta-dados são dados que se referem ao conteúdo de uma mídia. Eles são obtidos por métodos de rotulação, automáticos ou manuais. MPEG-7 (Martínez, 2004) é um padrão descritivo para mídias projetado para otimizar a aplicação de técnicas de busca por conteúdo.

Busca por similaridade de conteúdo consiste em realizar busca em coleções de alguma mídia especifica pela similaridade qualitativa com algum objeto de referência da mesma mídia. Difere dos mecanismos tradicionais de busca em uma base de dados por não atuar com uma descrição de palavras-chave ou atributos, mas com a extração automática desta descrição de um objeto de referência. Existem aplicações em vídeo (Smith & Chang, 1997), modelos 3D (Funkhouser et al., 2003), imagens (Wang et al., 2001)(Barnard & Forsyth, 2001) e áudio (Cano, 2007). Os modelos são ilustrados na Figura 5.1

A representação por amostras de mídia na busca por similaridade de conteúdo é perceptualmente amigável ao usuário, por permitir uma associação direta dos parâmetros de busca com o resultado esperado. Isso permite que o usuário manipule, de forma natural e consciente, os operadores do mecanismo de busca. Assim, do ponto de vista do usuário, a representação por amostras é uma representação associativa, já que pela percepção é possível realizar de forma direta o mapeamento da representação para a sensação que ela representa. Devido à natureza associativa, esta representação se mostra bem adequada para o computador, por ser bem definida e computável.

O que difere a representação nos dois pontos de vista é a forma com que ela é utilizada para explorar o espaço de mídias. Um fator vital para qualquer estratégia exploratória é a capacidade de medir os passos, i.e., a existência de uma métrica que permita mensurar a distância entre dois pontos do espaço de busca ou exploração. Para o usuário, esta métrica é subjetiva, resultado de um processo cognitivo. Por outro lado, no computador é necessária a adoção de um formalismo matemático, levando a uma métrica objetiva.

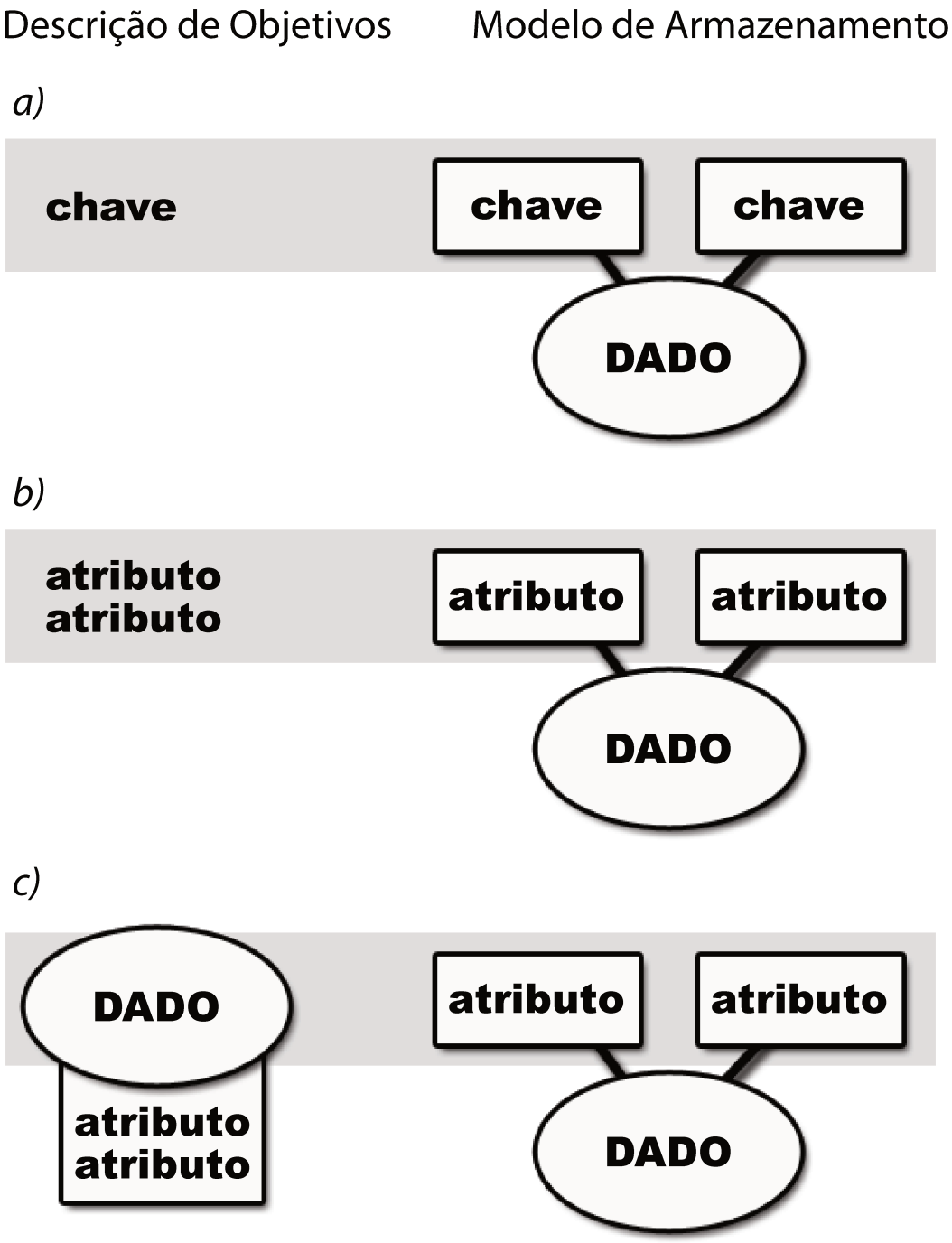


Fig. 5.1: Modelos de busca em uma base de dados: (a) Busca por palavra-chave; (b) Busca por descrição de atributos; e (c) Busca por similaridade de conteúdo.

A representação computacional por amostras sonoras é uma representação de alta dimensionalidade, o que implica em um espaço de busca extenso e de difícil exploração. Contudo, considerando o processo cognitivo, este espaço pode ser reduzido, sem um prejuízo extenso a sua completude, por uma inversão de representação. Serra (1997) disse que "a maioria das ondas não possui interesse musical e muitos sons são fisicamente distintos, mas perceptualmente iguais. Portanto, não é necessário gerar todas as ondas possíveis e a busca é por grupos reduzidos de técnicas de síntese e controle que permitirão explorar todo o espaço tímbrico de interesse musical"[1]. Na busca por similaridade, portanto, a representação por amostra é convertida em uma representação de atributos qualitativos, o que permite aproveitar a infra-estrutura das coleções de mídia existentes.

Este paradigma pode ser aplicado à síntese sonora, no que chamamos de síntese sonora baseada em conteúdo.

5.2 Síntese baseada em conteúdo

O paradigma de síntese por conteúdo consiste no uso de objetos sonoros como modelos representativos do objetivo sonoro. Um esquemático é apresentado na Figura 5.2. Amostras sonoras são apresentadas na entrada. O sintetizador as utiliza como parâmetros de controle em um mecanismo de síntese que retorna na saída um material sonoro original.

Um primeiro modelo de síntese que utiliza amostras sonoras como mecanismo de entrada é chamado "síntese por *matching*"[2]. Neste método, uma única amostra sonora é apresentada como referência. Alguma ferramenta de otimização é aplicada sobre os parâmetros de um sintetizador de forma a achar os valores que minimizem a distância do material sonoro obtido com a amostra de referência. Em Horner et al. (1993), algoritmos genéticos foram utilizados sobre um sintetizador FM. Em Horner et al. (1995) e Wun et al. (2004), a técnica foi aplicada sobre outros algoritmos de

[1] "*...most waves are not musically interesting, and many sounds which are physically different are perceptually equal. Therefore, it is not necessary to generate all possible waves and the aim is to find a reduced group of synthesis and control techniques which will allow us to explore the whole timbre space of musical interest.*"

[2] *Matching Synthesis*

Fig. 5.2: Esquemático da síntese sonora por conteúdo.

síntese.

Síntese por concatenação (Schwarz, 2005) também pode utilizar amostras sonoras como entrada. Neste modelo de síntese, pequenas amostras de som são armazenadas em uma base de dados. O material sonoro é gerado pela concatenação destes pequenos pedaços de som. A política de escolha da seqüência de sons pode ser controlada por uma amostra sonora: ou em uma tentativa de resíntese ou a partir de alguma heurística. A aplicação principal é em síntese de voz. A transmissão de índices de um dicionário de porções de som pode ser mais econômica que a transmissão do sinal em si.

A limitação das sínteses por *matching* e concatenação é que elas permitem apenas a reconstrução de um som já existente ou o controle de variação da geração de novos sons, não se mostrando como uma forma de representação para novos objetivos sonoros. Contudo, estes modelos de síntese inspiram uma metáfora: o paradigma pode ser reduzido a um sistema de otimização multi-objetivo. Isto remete à questão: como otimizar os parâmetros de um sintetizador de forma a maximizar o grau de similaridade do material sintetizado (ou minimizar o grau de dissimilaridade) a um grupo de amostras sonoras de entrada ?

Em problemas multi-objetivos, normalmente não existe uma solução ótima, mas sim um conjunto de soluções não-dominadas. Soluções não-dominadas são aquelas que obedecem ao princípio de otimalidade de Pareto (1906). Para toda solução não-dominada ou Pareto-ótima, não há como melhorar algum dos objetivos sem que algum outro objetivo seja piorado. Existem várias técnicas de solução. A mais direta é o uso de heurísticas para a obtenção de um máximo virtual, o que reduz o

problema para um sistema de otimização de objetivo único. Um exemplo é a minimização da soma ponderada das funções, apresentada na Equação 5.1. No contexto de síntese, a função f_i indica o grau de similaridade do som sintetizado para com a amostra i. Para cada amostra é dado um grau de ponderação α_i. O som ótimo é aquele que maximiza essa soma.

$$\sum_{i=1}^{n} \alpha_i f_i(x), \qquad \alpha_i \geq 0, i = 1, 2, \ldots, n \tag{5.1}$$

Outra estratégia é chamada de programação multinível (Candler & Norton, 1976). Os objetivos são ordenados por ordem de importância. O procedimento é o de achar o conjunto de pontos com o melhor valor possível para o objetivo de maior importância, e atuar de forma recursiva sobre este conjunto com os outros objetivos de menor importância.

Visando povoar a fronteira de Pareto, ou seja, visando localizar soluções não-dominadas, as técnicas que vêm apresentando desempenho bem satisfatório são as meta-heurísticas populacionais. Um exemplo são os algoritmos evolutivos multi-objetivos (Deb, 1999), que possuem variações (Schaffer, 1985; Hajela & Lin, 1992; Fonseca & Fleming, 1993; Horn et al., 1994; Srinivas & Deb, 1995; Zitzler et al., 2000). Outro exemplo é a otimização por nuvens de partículas (Parsopoulos & Vrahatis, 2002; Hu & Eberhart, 2002).

Independente da técnica de solução, sistemas de otimização necessitam de uma métrica. No caso da síntese sonora, a métrica é uma forma analítica de medir a similaridade entre dois sons. O problema apresentado novamente conduz à questão: como realizar a comparação de sons em uma representação computacional ?

A síntese por conteúdo sofre com a alta dimensionalidade referente à representação por amostras sonoras, de forma análoga à busca por similaridade de conteúdo. A diferença é que, no processo de busca, as qualidades são otimizadas para o armazenamento em coleções, enquanto no processo de

síntese elas devem ser otimizadas para o controle do mecanismo gerativo de áudio.

Assim, dada uma representação qualitativa com uma métrica e uma estratégia de otimização multi-objetivo, é possível construir um sintetizador por conteúdo.

5.2.1 Considerações de construção

A Figura 5.3 apresenta a arquitetura de um sintetizador por conteúdo. As amostras de áudio são convertidas em uma representação qualitativa pelos extratores de qualidades. O núcleo é a parte principal do sintetizador. Ele é um sistema que otimiza os parâmetros do mecanismo de síntese para que o material sonoro represente, de forma mais adequada, as amostras de entrada.

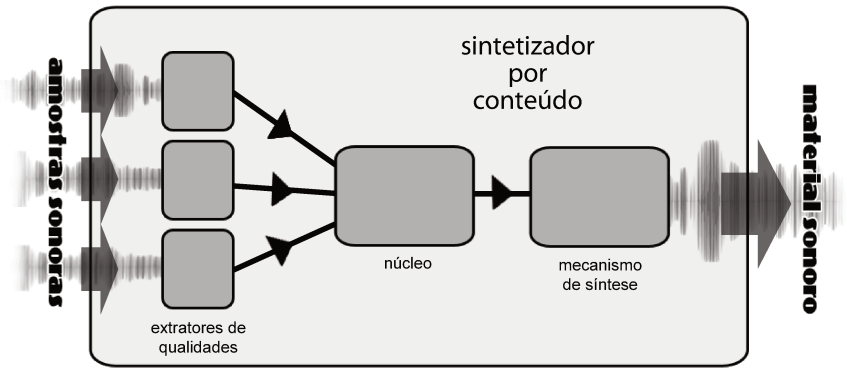


Fig. 5.3: Arquitetura de um sintetizador baseado em conteúdo.

A função dos extratores de qualidades é converter um sinal de entrada em uma representação qualitativa. Há duas abordagens possíveis para a escolha desta representação: baseada em um modelo de representação cognitiva; baseada na representação computacional do mecanismo de síntese.

Na primeira opção, deve-se escolher algum modelo cognitivo ou conjunto de qualidades que descrevam a amostra sonora. Neste modelo de representação, uma métrica simples é capaz de criar uma tipologia espacial próxima da subjetiva. Entretanto, após a determinação das qualidades a serem

apresentadas no resultado final, deve-se traduzi-las para o espaço paramétrico do mecanismo de síntese. Conjunto de qualidades aplicadas em processos de identificação de instrumentos fornecem um caminho (Mierswa & Morik, 2005). Esta técnica se assemelha ao uso de linguagem inata em síntese semântica, descrito na Sessão 6.2 do próximo capítulo. Verfaille & Arfib (2001) propõem o uso de um espaço semântico intermediário em um processo de controle de síntese. A síntese concatenativa no modelo CATERPILLAR (Schwarz, 2004) utiliza o espaço qualitativo semântico para a realização da heurística de seqüenciamento.

Na segunda opção, ao se basear na representação do mecanismo de síntese, evita-se a necessidade de tradução para uma representação além da paramétrica e a da amostra sonora. Por outro lado, uma métrica simples no espaço dos parâmetros não permite uma associação coerente com a representação cognitiva. Isso implica que a busca se realiza no espaço paramétrico e não em um espaço simbólico baseado num modelo cognitivo. Além disso, é necessário que seja possível migrar da representação paramétrica para uma amostra sonora (síntese) e da amostra sonora para uma representação paramétrica (análise). Assim, é preferível o uso de codecs de áudio, como OGG Vorbis ou LPC, ao uso de transformações matemáticas como modulação de freqüência.

O paradigma de síntese é independente do mecanismo de síntese. Porém, a sua escolha influencia a performance do sintetizador, pois uma tradução ineficiente de uma representação qualitativa, ou um algoritmo ineficiente atuando sobre um espaço paramétrico muito acidentado, torna o processo de otimização penoso.

O núcleo é o responsável pela otimização. A escolha da meta-heurística de otimização multi-objetivo é o que determina a forma como as amostras de entrada influenciam a saída. Existem duas abordagens diferentes. A primeira é a de se valorizar o que há de comum em todas as amostras. Assim, a saída tende a ser um consenso das entradas. A segunda é a de enfatizar o que cada amostra tem de diferente, sendo possível identificar nuances particulares no resultado final.

5.3 Síntese gerativa baseada em população

Síntese Gerativa baseada em População (SGBP) (Costa et al., 2006a,b) é um sintetizador que aplica os conceitos de síntese por conteúdo para a geração de texturas sonoras. O processo de síntese de texturas sonoras via síntese granular tem como motivação a dificuldade de se obter um controle gerativo que seja intuitivo e simples em virtude da vasta quantidade de eventos a serem controlados. SGBP controla um mecanismo de síntese granular utilizando um modelo populacional de representação interna de sonoridade, obtido com a aplicação de meta-heurísticas de quantização vetorial sobre amostras sonoras reais.

É possível dizer que o SGBP é um tipo de síntese por concatenação, mas que difere de sistemas tradicionais por desenvolver uma coleção de porções de sons que tentam explicitar nuances de uma sonoridade, e não serem gerais e permitirem resíntese. Há certa semelhança com a arquitetura CATERPILLAR (Schwarz, 2004) em algumas formas de uso, como a resíntese de sons com nuances sonoras de outras amostras. Contudo, há diferenças fundamentais nos mecanismos internos, i.e., no mecanismo de obtenção da coleção e na heurística de seqüenciamento da saída. Outro mecanismo de síntese com certa semelhança é apresentado por Kobayashi (2003). Nele, frames sonoros são clusterizados, e não extraídos de uma coleção, e depois utilizados em um mecanismo de síntese por concatenação.

5.3.1 Síntese granular

Síntese Granular (Roads, 1988) surgiu nos anos 70 como um novo paradigma de geração sonora fundamentado na exploração das limitações estudadas por Gabor (1946). Baseia-se na idéia de que qualquer som complexo pode ser composto pela subseqüente exposição de microporções sonoras de composição espectral estática, de duração abaixo da variação mínima perceptível. Estas porções são metaforicamente chamadas de grãos ou *microsons* (Roads, 1996, 2001).

Gabor realizou estudos sobre as limitações da percepção sonora em relação a variações de

freqüência e intensidade. Ele verificou que existe uma resolução temporal limite relacionada à freqüência, i.e, variações de altura ou intensidade só são observadas em estímulos com uma duração mínima que é dependente da freqüência sonora.

A Figura 5.4 realiza uma analogia com a percepção visual espacial. Um quarto de disco é apresentado em duas diferentes resoluções espaciais. Na esquerda, a resolução apresentada é menor que a resolução limite da percepção: os elementos que formam a imagem são diferenciáveis, expondo a descontinuidade. Na direita, a resolução apresentada é superior ao limite perceptivo e a imagem, apesar de discreta, é percebida como um disco contínuo. Qualquer aumento de resolução acima do limite é imperceptível ao ser humano. A percepção visual também possui uma resolução temporal limite. Esta limitação possibilita que a variação seqüencial de imagens estáticas no cinema seja percebida como um movimento contínuo.

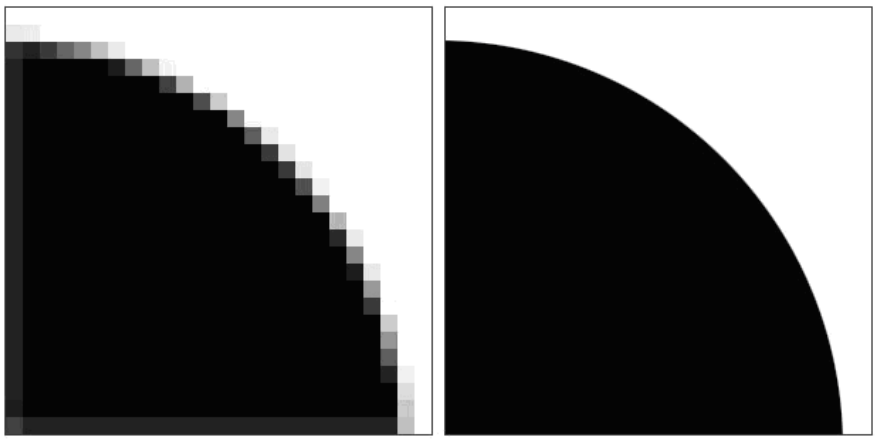


Fig. 5.4: Efeito da resolução sobre a percepção visual.

Xenakis (1971) escreveu que sons complexos podem ser reproduzidos pela exibição a uma taxa regular de um livro de quadros, ou *screens*. Ele define um quadro como um som de baixa duração com composição espectral estática e bem definida. A Figura 5.5 ilustra um livro de quadros.

O método mais tradicional de síntese granular é o modo sem sincronia (Roads, 1991), em que os grão são distribuídos de forma irregular no tempo e na frequência, geralmente controlados por

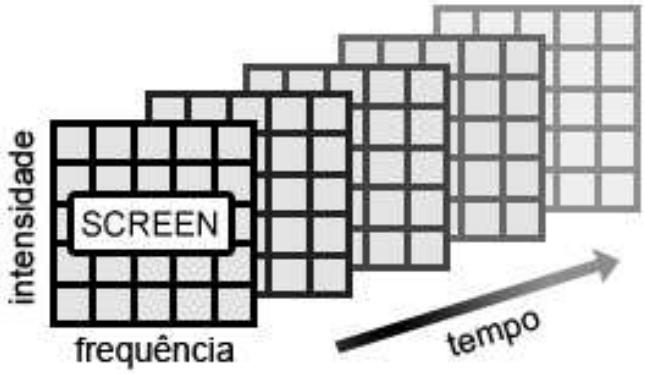


Fig. 5.5: Livro de quadros de Xenakis.

máscaras de tendência. Os grãos pode ser formas sonoras simples, tons FM (Waschka & Ferreira, 1988) e amostras reais (Jones & Parks, 1988; Lippe, 1993). Alguns métodos posteriores de síntese granular trabalham com sincronismo de fase, para evitar aberrações espectrais nas transições dos quadros. Outros realizam sincronismo de altura, reestruturando o quadro em multi-níveis espectrais, com diferentes taxas de exposição. Ver Roads (1996) para uma especificação mais detalhada. Truax (1988) discute a implementação de síntese granular em aplicações de tempo real.

Apesar da grande capacidade de geração sonora ser a principal qualidade da síntese granular, ela implica em uma alta dificuldade de controle do processo de síntese, devido à elevada diversidade de grãos e formas de arranjo. Por isso, são utilizados mecanismos de geração granular, em geral com métodos estocásticos, que tentam obter resultados complexos sem a necessidade de manipulação direta dos quadros espectrais. Xenakis utiliza os livros de quadros como uma base, aplicando uma metodologia estocástica na conversão de eventos discretos. Existe uma série de abordagens de controle gerativo, em geral desenvolvidas no início dos anos 90: Miranda (1995), Bowcott (1989) e Orton et al. (1991) utilizam autômatos celulares; Miranda & Maia (2005) utiliza um modelo baseado em matrizes markovianas e controle nebuloso; Fujinaga & Vantomme (1994) utilizam algoritmos genéticos; Bowcott (1990) utiliza modelagem populacional; Hamman (1991) utiliza funções não-lineares; e Nagashima (1992) utiliza redes neurais.

A grande dificuldade dos métodos de controle do processo gerativo da síntese granular é que a relação entre o material sonoro obtido e a variação paramétrica necessária para obtê-lo é, em geral, complexa e pouco intuitiva. Uma abordagem comum é tentar remapear o controle do mecanismo gerativo para alguma metáfora com associações no mundo real, como dinâmicas ecológicas (Keller & Truax, 1998; Gadd & Fels, 2002).

5.3.2 Textura sonora

Uma importante aplicação de síntese granular é a síntese de texturas sonoras, seja na criação de cenários sonoros (*soundscapes*) ou de objetos sonoros. Schaeffer (1966) define textura sonora como um fluxo sonoro com sonoridade perceptível que expressa algum tipo de variação temporal-espectral e que não pode ser definido como um evento pontual ou objeto. O barulho da chuva, de uma cachoeira ou de um ar condicionado desregulado são exemplos reais de texturas sonoras. Além destes, existem texturas no contexto da *l'écoute réduite* (Schaeffer, 1966) que não podem ser associadas a alguma fonte sonora. No contexto de síntese granular, uma textura sonora pode ser vista como um fluxo de grãos.

Na composição da sonoridade de uma textura sonora, variações instantâneas localizadas exercem maior influência sobre a percepção do que tendências globais. Considerando o som de chuva como exemplo, variações de intensidade de uma duração mais elevada que poucos segundos são interpretadas como variação de força da chuva, mas ainda sendo percebido como som de chuva. Contudo, se a dinâmica das micro-variações de intensidade for modificada, o timbre é descaracterizado. Para um mecanismo de síntese, isso implica que o controle de sonoridade deve priorizar variações locais. Outro exemplo é o som de uma nota tocada em um instrumento acústico como uma guitarra: trata-se de um evento sonoro e não de uma textura. Neste caso, o comportamento espectral frente à excitação da corda e o tempo que esta leva até o repouso tem influência sobre a percepção do timbre, i.e., existe uma evolução com um início e um fim.

5.3.3 Síntese de textura sonora baseada em conteúdo

A síntese baseada em conteúdo objetiva amenizar as limitações da síntese granular: mecanismo de controle pouco intuitivo e dificuldade de representação de objetivos. A aplicação de síntese baseada em conteúdo na obtenção de texturas sonoras consiste em gerar um fluxo sonoro contido no contexto definido pelas amostras de referência, i.e., cujas características de sonoridade sejam semelhantes às das amostras sonoras apresentadas como referência.

Um mecanismo de síntese de textura sonora deve permitir dois tipos de controle por parte do usuário: controle de sonoridade e controle de dinâmica. De forma pictórica, no caso da síntese do "som de chuva", o controle de sonoridade seria a escolha da sonoridade como "som de chuva". Enquanto que o controle de dinâmica seria uma forma de gerenciar como esta chuva se comporta no tempo. Em um modelo de síntese baseada em conteúdo, a sonoridade é definida por um grupo de amostras sonoras que caracterizem a sonoridade desejada.

No SGBP, o controle de sonoridade é exercido por um fluxo sonoro denominado Referência de Sonoridade. O controle de dinâmica é exercido por outro fluxo sonoro chamado Controle de Dinâmica. Em seu funcionamento, o mecanismo interno do SGBP estabelece uma representação interna das características sonoras da Referência de Sonoridade.

5.3.4 Arquitetura do SGBP

O mecanismo interno do SGBP utiliza como representação qualitativa os quadros de Xenakis, chamada base granular. Neste texto, quando o sinal está na representação temporal utilizada na entrada, é dito estar na base de sinal. Os fluxos de entrada são convertidos para uma seqüência de quadros através do uso de uma janela deslizante. Em seguida, aplica-se a Transformada Discreta de Fourier pelo algoritmo de FFT para a obtenção do espectro. Por último, informações sobre o contexto de aparecimento deste grão são adicionadas. A Figura 5.6 ilustra as fases do sinal. O processo de conversão da base de sinal para a base granular é dito granulador, enquanto o processo contrário é chamado de misturador. O processamento é realizado em base granular, i.e., em um nível de discretização mais elevado que a amostragem na base de sinal.

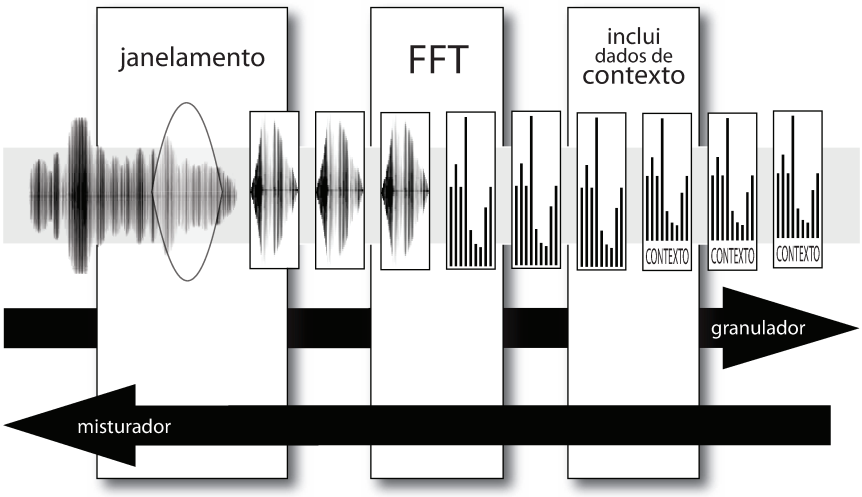


Fig. 5.6: Processo de conversão entre base de sinal e base granular. O sinal é inicialmente janelado, em seguida aplica-se o algoritmo de FFT para a obtenção do espectro. Ao espectro, são adicionadas as informações de contexto.

A Figura 5.7 ilustra a arquitetura do SGBP. A referência sonora de entrada é realizada por um fluxo sonoro contínuo, denominado Referência de Sonoridade. Um granulador é utilizado para a conversão em base granular. O mecanismo interno do sintetizador realiza um aprendizado de características sonoras relevantes da referência e os armazena em uma estrutura computacional denominada Estrutura Representativa. Mecanismos de quantização vetorial são utilizados como ferramenta de aprendizado, criando uma população de protótipos que amostre o contexto.

No processo de síntese, o SGBP prevê uma segunda entrada denominada Controle de Dinâmica, também em base granular, de onde o sintetizador extrai informações de evolução de energia (associada ao envelope dinâmico) e algum dado espectral. O Gerador de Saída gera material com estas informações e as qualidades sonoras descritas na Estrutura Representativa na forma de uma seqüência granular. Por fim, o misturador converte para a base de sinal.

Fig. 5.7: Arquitetura do SGBP.

5.3.5 Processo de obtenção de protótipos

O mecanismo de representação é baseado em um modelo de síntese por dicionário ou *codebook*. É um modelo comumente utilizado por codecs de áudio como MP3, OGG Vorbis e vocoders. Um dicionário é um conjunto de protótipos de amostras sonoras. O processo de síntese consiste em criar um sinal sonoro a partir do agrupamento seqüencial destas amostras.

No SGBP, os protótipos são compostos por uma distribuição espectral e uma estrutura que representa o contexto granular. Contexto granular são os estados do sinal sonoro no aparecimento de um grão. A estrutura de contexto possui informações como freqüência fundamental, energia, histórico breve de grãos que precederam a aparição do grão em análise e posição relativa à última mudança de freqüência fundamental.

O procedimento de obtenção de protótipos é chamado Quantização Vetorial e consiste basicamente em determinar um conjunto de protótipos que representem as características de um determinado grupo. Quantização vetorial é comumente aplicada em processos de compressão de

dados com perda de informação, onde o sinal é substituído por índices em um conjunto limitado de representações. A Figura 5.8 descreve graficamente um processo de quantização vetorial bidimensional. Os círculos cinza são as amostras de entrada pertinentes a algum contexto, que é representado pelos círculos pretos. Obviamente, o cenário bidimensional deve ser interpretado somente como uma visão pictórica do que acontece na prática, com os círculos cinzas e pretos residindo em dimensões muito mais elevadas.

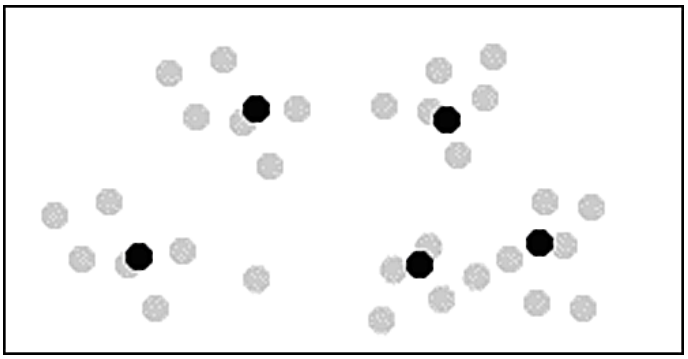


Fig. 5.8: Visão pictórica da quantização vetorial. Em cinza: amostras sonoras do contexto; Em preto: protótipos representativos.

No SGBP, os círculos pretos correspondem à população de protótipos que passam por um processo de auto-organização visando atingir uma configuração equivalente à configuração espacial final apresentada na Figura 5.8. Observe que os círculos pretos são organizados de forma a capturar os mais relevantes aspectos das amostras de entrada. Eles são chamados protótipos representativos porque podem ser interpretados como uma representação concisa das amostras de entrada, geralmente expressando uma explicação consensual da variabilidade local na vizinhança das amostras de entrada.

Como processos de auto-organização, o SGBP aplica três meta-heurísticas bio-inspiradas apresentadas no capítulo 4: mapas auto-organizáveis (MAOs), Seção 4.2.3; sistemas imunológicos artificiais (SIAs), Seção 4.3; e algoritmos genéticos (AG), Seção 4.4. Com MAOs e SIAs a aplicação populacional é direta. Com MAO, utiliza-se um arranjo bidimensional com dimensão definida como parâmetro do sistema. O vetor de entrada varia de acordo com a representação usada pelo grão: sinal, espectro ou espectro com histórico de energia. Com os algoritmos genéticos, o *fitness* é proporcional à similaridade do indivíduo com as amostras da seqüência granular da referência de sonoridade.

Contrário às aplicações tradicionais, no SGBP o interesse é em toda a população e não somente nos melhores indivíduos da população. É possível observar que uma população, em uma dada geração, atua como um mecanismo de quantização vetorial.

Todas essas meta-heurísticas exigem alguma métrica de semelhança entre os grãos. Caetano et al. (2005c) sugere as vantagens da comparação no domínio do tempo, citando Smalley (1986): "espectro é percebido pelo tempo e tempo é percebido como movimento espectral"[3]. No SGBP, realizam-se comparações utilizando um cálculo híbrido envolvendo a métrica euclidiana da distribuição espectral e uma comparação de um histórico recente dos grãos que antecedem o grão do instante do cálculo. A idéia é não só considerar a existência de um grão em si, no contexto, mas também os eventos que acionaram o seu aparecimento.

5.3.6 Mecanismo gerativo

O Mecanismo Gerativo é o responsável por sintetizar a seqüência granular que será convertida no fluxo de saída. Ele recebe como entrada o fluxo de controle em base granular e tem acesso à população de protótipos. O procedimento gerativo é simples: é estabelecida uma métrica que envolve uma combinação ponderada por um fator α de uma distância entre as distribuições espectrais e outra ponderada por β entre o contexto granular da saída gerada e do protótipo. O protótipo escolhido é aquele com menor valor produzido pela métrica. Este procedimento é repetido para cada grão a ser inserido na seqüência de saída. A escolha dos valores de α e β pode indicar uma maior influência do contexto ou da distribuição espectral do controle de dinâmica sobre o material sonoro de saída.

5.3.7 Implementação

A implementação do protótipo do SGBP foi realizada em linguagem MATLAB. A motivação da escolha foi a facilidade de reuso de algoritmos bio-inspirados, desenvolvidos pelo próprio grupo e presentes em toolboxes de outros grupos, e a disponibilidade de ferramentas de processamento de sinais. O foco foi no desenvolvimento da arquitetura e na configuração dos algoritmos, não

[3] "*...spectrum is perceived through time and time is perceived as spectral motion.*"

sendo objetivada a capacidade de atuar em tempo real. A interface sonora da implementação é em fluxos sonoros em codificação PCM em 16 bits, encapsulados no formato de arquivo de áudio WAVE.

O protótipo possui dois módulos: módulo de aprendizado de contexto e módulo de síntese. O módulo de aprendizado de contexto foi implementado nas três meta-heurísticas apresentadas, sendo que a versão de mapas auto-organizáveis utilizou os algoritmos do toolbox da Universidade de Tecnologia de Helsinki (Oy et al., 2000). Nas outras duas meta-heurísticas, foram utilizados códigos próprios do grupo de pesquisa. O módulo de aprendizado recebe como entrada os parâmetros do algoritmo, da representação e o fluxo sonoro, emitindo como saída uma população de protótipos. O módulo de síntese recebe como entrada a população de protótipos e o fluxo de controle dinâmico. Para cada grão da seqüência de controle, é determinado um protótipo da população, criando uma seqüência de saída. A saída é o fluxo sintetizado. O sistema não tem dependência quanto à taxa de amostragem. Contudo, a população de grânulos possui uma taxa específica que deve ser respeitada na síntese e as taxas de amostragem do fluxo de referência de sonoridade e de controle de dinâmica devem ser as mesmas. Esta é uma limitação da implementação e não do método.

5.3.8 Experimentos

Foram planejados quatro experimentos para avaliar o método, com os seguintes objetivos específicos:

- **Experimento 1:** Verificar se há transferência de sonoridade e dinâmica para o fluxo de saída;
- **Experimento 2:** Verificar comportamento de acompanhamento espectral;
- **Experimento 3:** Verificar a influência da variação paramétrica na performance;
- **Experimento 4:** Ressíntese e análise do som de clarineta.

Os resultados obtidos são apresentados na forma de espectrograma: uma ilustração gráfica da evolução da energia do espectro de freqüência instantâneo no tempo. O espectro instantâneo é obtido a partir do janelamento do sinal e da aplicação da transformada de Fourier de curta duração pelo

algoritmo FFT. As figuras são representadas, em seu eixo vertical, com a freqüência em uma escala linear na unidade Hertz e, no eixo horizontal, o tempo em segundos.

O espectrograma é a ferramenta padrão para a visualização de amostras sonoras, sendo especialmente adequado no contexto dos experimentos seguintes, dada a própria natureza da representação interna utilizada pela implementação do SGBP: o espectro instantâneo.

De forma alternativa, as amostras sonoras são disponibilizadas em mídia anexa.

Experimento 1

No primeiro experimento, foram utilizadas três amostras sonoras (Figura 5.9) com características particulares individuais: som sintético, com notas concorrentes (acordes) tocadas com dinâmica e tempo bem definidos, ilustrado na Figura 5.10; fraseado de guitarra elétrica, com notas harmonicamente bem definidas, tocadas individualmente no tempo, mas com freqüência fundamental com forte variação, ilustrado na Figura 5.11; e frase "Teste um. Testando. Testando.", proferida por voz masculina que apresenta espectro fortemente ruidoso com concentração de energia até 1kHz e na faixa entre 1,5kHz e 2kHz, ilustrada na Figura 5.12. A amostra de voz apresenta forte concentração de energia na faixa até 10 kHz e entre 20 e 25 kHz. As outras duas amostras têm a energia praticamente confinada em um limite superior a 5 kHz. O experimento consistiu em apresentar ao sistema estas amostras como referência de sonoridade e controle de dinâmica, para posterior análise comparativa do resultado com as entradas.

Foram realizados quatro grupos de execução distintos. Nos três primeiros grupos, o controle dinâmico foi mantido fixo, variando a referência de sonoridade. Os resultados com o som de voz são mostrados nas Figuras 5.13 a 5.16, com o som de guitarra nas Figuras 5.17 a 5.20 e com o som sintético nas Figuras 5.21 a 5.24. Umas das execuções de cada grupo utiliza a mesma amostra para as duas entradas do sistema e, conseqüentemente, podem ser confrontadas com a amostra original, chamada de execução mimética. As outras utilizam amostras diferentes entre si para as entradas.

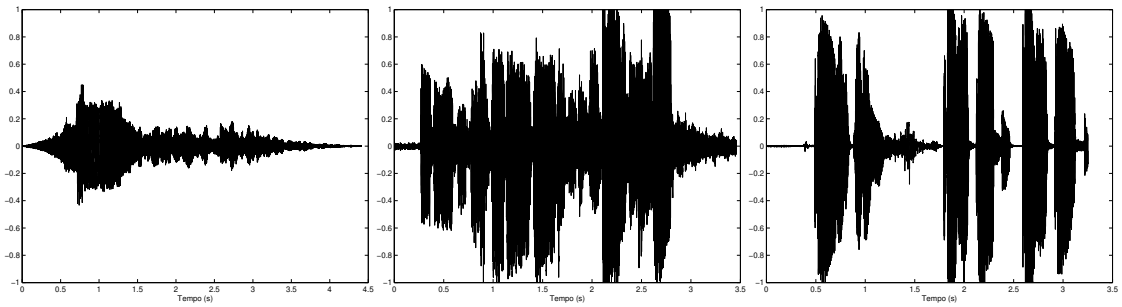


Fig. 5.9: Envelope de energia das amostras utilizadas no experimento 1: (a) sinal sintetizado com componentes harmônicos; (b) frase melódica de guitarra; (c) voz masculina.

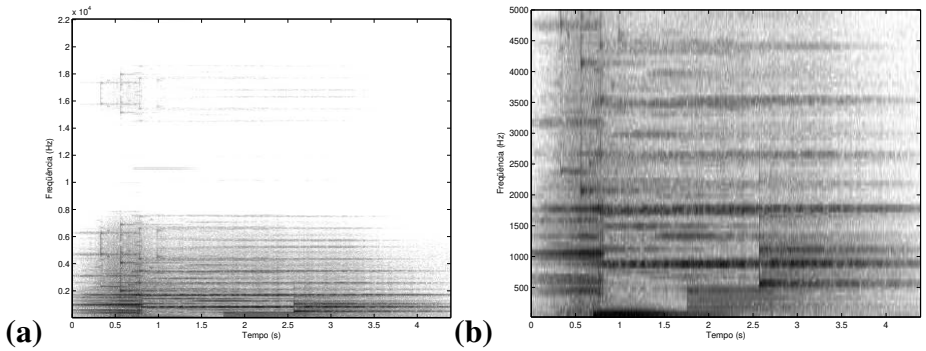


Fig. 5.10: (a) Espectrograma do sinal sintético harmônico. (b) Zoom limitado em 5 kHz.

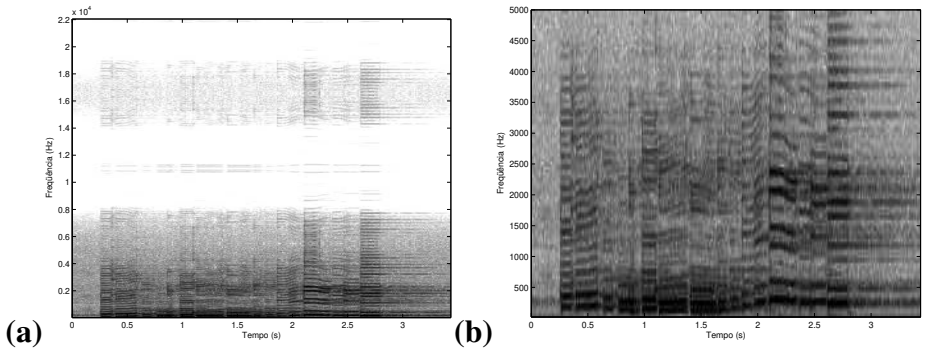


Fig. 5.11: (a) Espectrograma do sinal de guitarra. (b) Zoom limitado em 5 kHz.

No quarto grupo de execução todas as amostras foram utilizadas como referência de sonoridade, variando o controle de dinâmica, apresentadas nas Figuras 5.25 a 5.28.

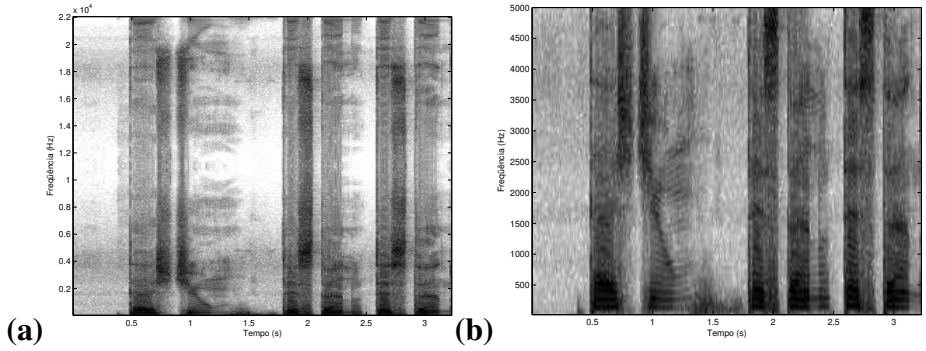


Fig. 5.12: (a) Espectrograma do sinal de voz. (b) Zoom limitado em 5 kHz.

No primeiro grupo de execuções, o sinal de voz foi utilizado como controle de dinâmica. Todos os materiais obtidos renderam um comportamento dinâmico semelhante ao controle dinâmico utilizado, com aparente perda de energia, Figura 5.13. Quando utilizado o som de voz como referência de sonoridade, Figura 5.14, o espectro resultante apresentou certa semelhança ao original: fortemente ruidoso em pontos de alta energia e faixas de concentração de energia bem definidas, apesar de menos contrastante que o original (ver franjas visíveis no espectro limitado da Figura 5.12 (b), e não presentes no material obtido). Importante observar que a frase se manteve inteligível, com som ligeiramente mais grave e forte presença de artefatos ruidosos de alta freqüência, dando um aspecto artificial ao som. Nas outras duas execuções, Figura 5.15 e 5.16, o material sonoro se limitou a 5 kHz, mantendo uma faixa de concentração antes dos 500 Hz no caso da guitarra e aproximadamente 900 Hz no caso do som sintético. Essas faixas de concentração estão presentes nas amostras originais, Figuras 5.10 e 5.11.

No segundo grupo de execuções, o sinal de guitarra foi utilizado como controle de dinâmica. O comportamento dinâmico da saída, Figura 5.17, é semelhante à referência com aparente perda de energia. Em (c), esta perda é menor. Quanto ao comportamento espectral, na execução mimética, Figura 5.19, a distribuição energética é semelhante à original, contudo, há perda de definição das linhas harmônicas. Por isso, o material obtido é mais ruidoso que o original, sendo impossível distinguir a linha melódica. Com a referência de voz, Figura 5.18, e sintético, Figura 5.20, o comportamento da concentração de enegia seguiu as características das referências.

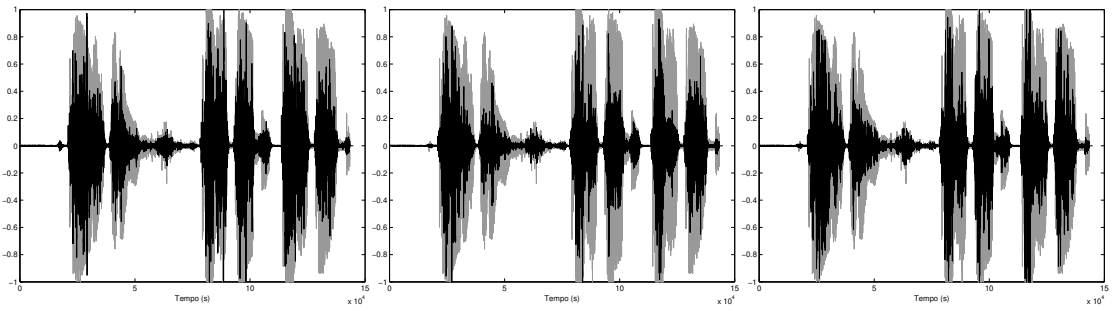


Fig. 5.13: Experimento 1, Amostra de Fala como controle de dinâmica (claro). Envelope de energia do material sonoro (escuro) obtido com diferentes referências de sonoridade: (a) voz; (b) guitarra; (c) som sintético.

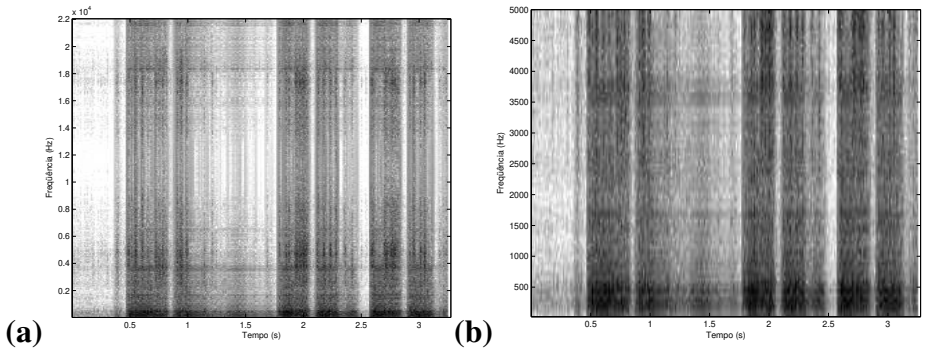


Fig. 5.14: (a) Espectrograma do material sonoro obtido com fala (controle de dinâmica) e fala (referência de sonoridade). (b) Zoom limitado em 5 kHz.

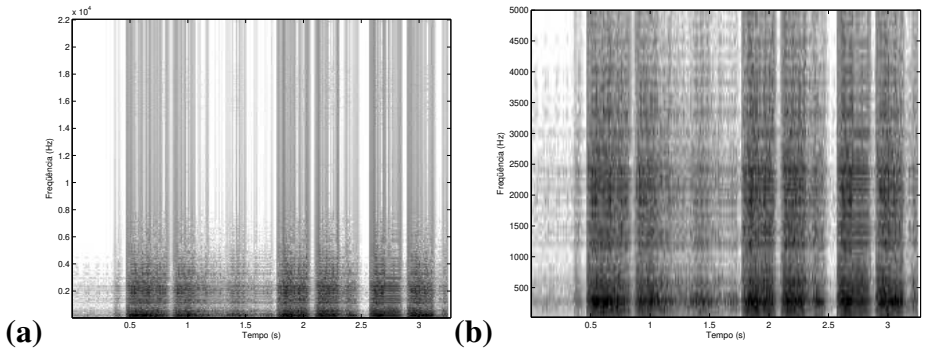


Fig. 5.15: (a) Espectrograma do material sonoro obtido com fala (controle de dinâmica) e guitarra (referência de sonoridade). (b) Zoom limitado em 5 kHz.

No terceiro grupo de execuções, o som sintético foi utilizado como controle dinâmico. O comportamento dinâmico foi bem coerente com as referências, Figura 5.21. Na execução mimética,

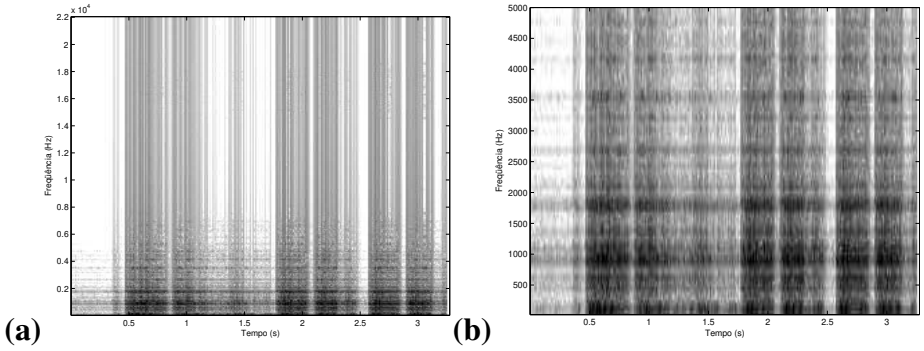


Fig. 5.16: (a) Espectrograma do material sonoro obtido com fala (controle de dinâmica) e som sintético (referência de sonoridade). (b) Zoom limitado em 5 kHz.

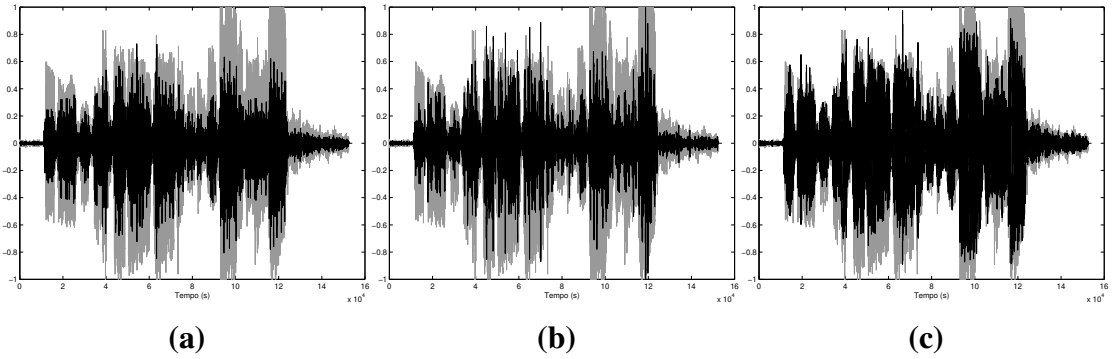


Fig. 5.17: Experimento 1, Amostra de guitarra como controle de dinâmica (claro). Envelope de energia do material sonoro (escuro) obtido com diferentes referências de sonoridade: (a) voz; (b) guitarra; (c) som sintético.

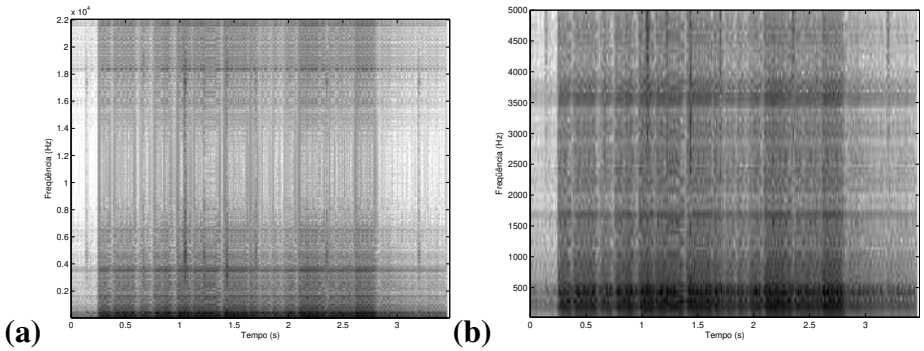


Fig. 5.18: (a) Espectrograma do material sonoro obtido com guitarra (controle de dinâmica) e fala (referência de sonoridade). (b) Zoom limitado em 5 kHz.

Figura 5.24, o comportamento espectral é parecido com o som original, com faixas harmônicas bem definidas. Mas estas faixas estão presentes em todo o som, não existindo a pluralidade de faixas

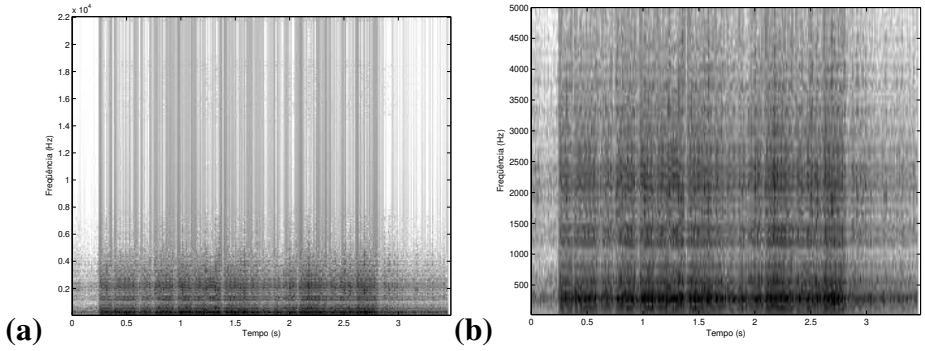


Fig. 5.19: (a) Espectrograma do material sonoro obtido com guitarra (controle de dinâmica) e guitarra (referência de sonoridade). (b) Zoom limitado em 5 kHz.

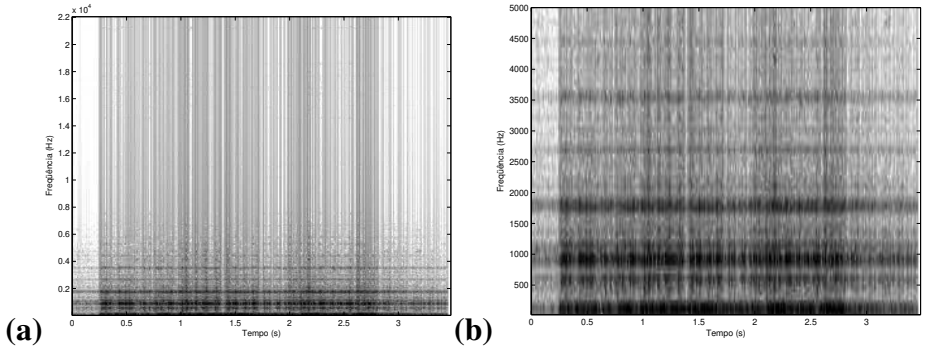
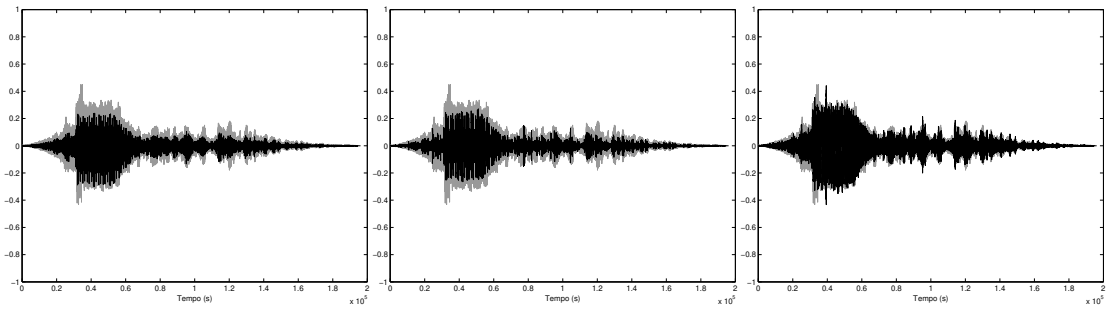


Fig. 5.20: (a) Espectrograma do material sonoro obtido com guitarra (controle de dinâmica) e som sintético (referência de sonoridade). (b) Zoom limitado em 5 kHz.

harmônicas do som original. O mesmo ocorre com referência ao som de guitarra, Figura 5.23. A referência de voz obteve a coerência da distribuição energética, Figura 5.22.

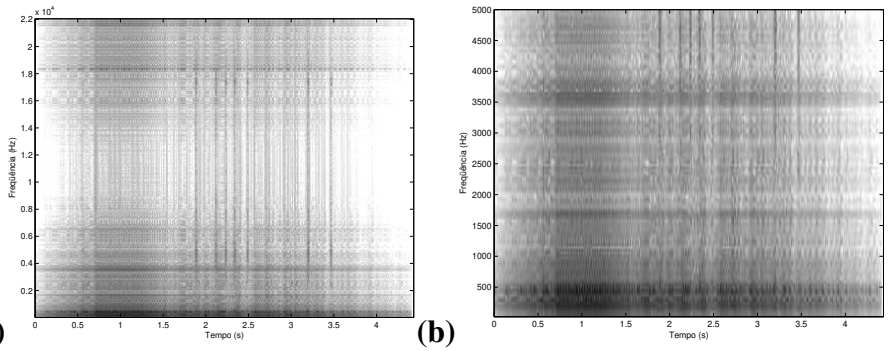
No quarto grupo de execução, todas as amostras foram utilizadas como referência de sonoridade. Houve coerência no comportamento dinâmico nas três execuções, Figura 5.25. Utilizando o controle pela voz, Figura 5.26, o som obtido possui forte característica do som de voz original, sendo possível identificar a frase dita. Nas duas outras execuções, Figuras 5.27 e 5.28, a identificação com o som original não é direta. Isso se deve à incapacidade de se generalizar componentes harmônicos. Contudo, a distribuição espectral dos dois sons é semelhante.

Em um passo seguinte, os materiais obtidos no primeiro experimento foram apresentados a



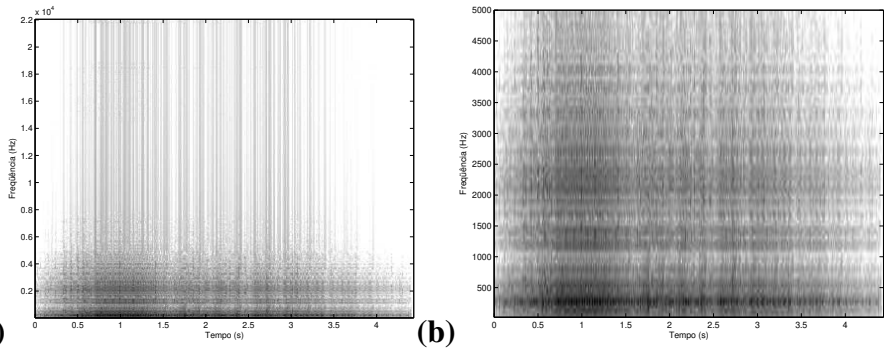
(a) **(b)** **(c)**

Fig. 5.21: Experimento 1, Amostra de guitarra como controle de dinâmica (claro). Envelope de energia do material sonoro (escuro) obtido com diferentes referências de sonoridade: (a) voz; (b) guitarra; (c) som sintético.



(a) **(b)**

Fig. 5.22: (a) Espectrograma do material sonoro obtido com som sintético (controle de dinâmica) e fala (referência de sonoridade). (b) Zoom limitado em 5 kHz.



(a) **(b)**

Fig. 5.23: (a) Espectrograma do material sonoro obtido com som sintético (controle de dinâmica) e guitarra (referência de sonoridade). (b) Zoom limitado em 5 kHz.

sujeitos com treinamento musical. Aos objetos da pesquisa, foi pedido que fosse mensurada a semelhança entre o material obtido e as amostras de entrada. A análise da pesquisa subjetiva é apresentada

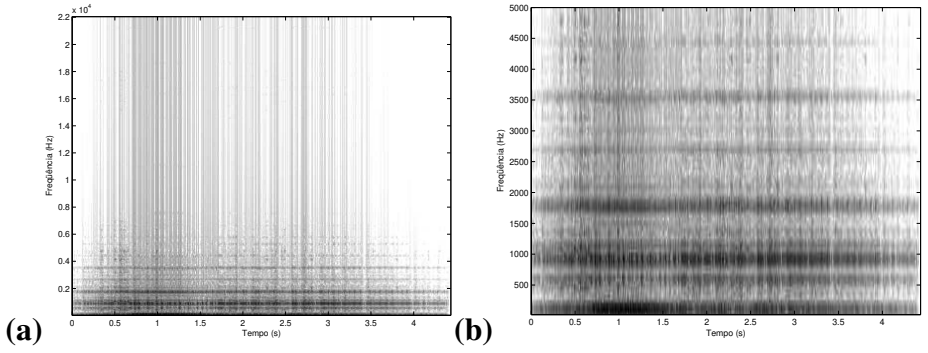


Fig. 5.24: (a) Espectrograma do material sonoro obtido com som sintético (controle de dinâmica) e som sintético (referência de sonoridade). (b) Zoom limitado em 5 kHz.

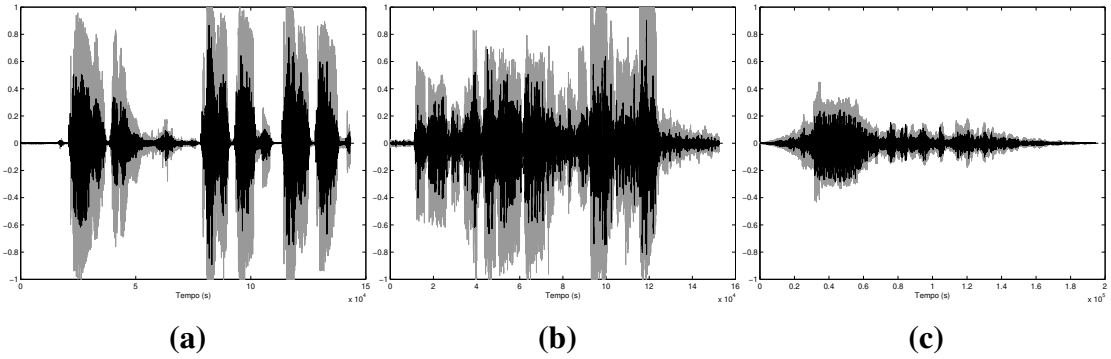


Fig. 5.25: Experimento 1, todos os sons como referência de sonoridade (claro). Envelope de energia do material sonoro (escuro) obtido com diferentes controles de dinâmica: (a) voz; (b) guitarra; (c) som sintético.

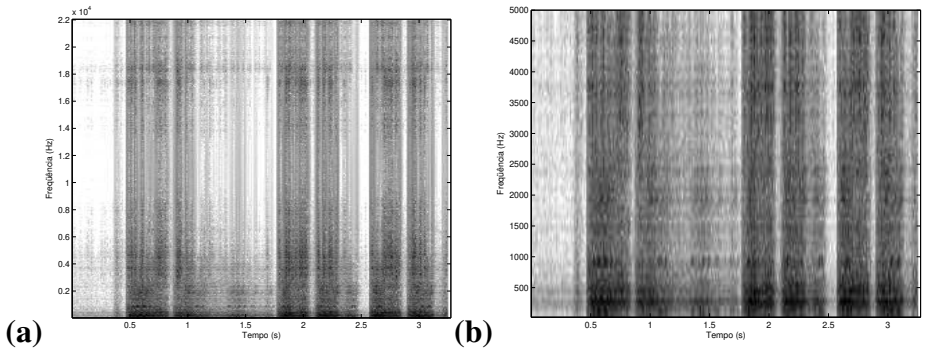


Fig. 5.26: (a) Espectrograma do material sonoro obtido com todos os sons (referência de sonoridade) e fala (controle de dinâmica). (b) Zoom limitado em 5 kHz.

na Tabela 5.1. A Tabela 5.2 traz as médias nos valores subjetivos referentes à comparação entre o material sonoro e as entradas. A comparação do material com o controle de dinâmica e com a

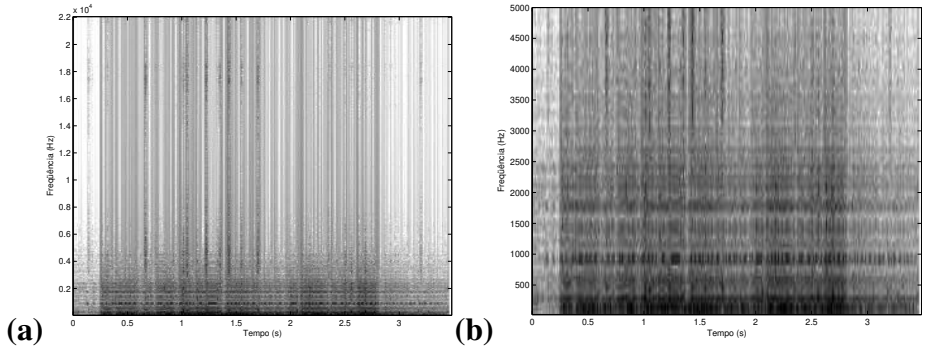


Fig. 5.27: (a) Espectrograma do material sonoro obtido com todos os sons (referência de sonoridade) e guitarra (controle de dinâmica). (b) Zoom limitado em 5 kHz.

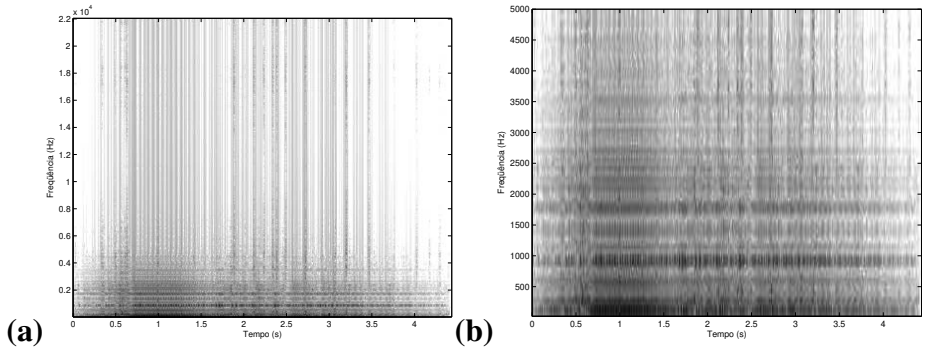


Fig. 5.28: (a) Espectrograma do material sonoro obtido com todos os sons (referência de sonoridade) e som sintético (controle de dinâmica). (b) Zoom limitado em 5 kHz.

referência de sonoridade implicou em valores acima da média do valor obtido comparando todas as amostras. Este é um indício de que o material sonoro é mais fortemente identificável com as amostras que foram utilizadas como referência para a sua síntese.

Deste primeiro experimento, é possível tirar algumas conclusões sobre o funcionamento do sistema. Há forte dificuldade de generalizar figuras harmônicas bem definidas, só sendo possível quando há casamento de altura entre a referência de sonoridade e o controle de dinâmica, ou quando a referência de sonoridade não possui uma diversidade de alturas bem definidas no espectro, caso da primeira execução do terceiro grupo. O som de guitarra, apesar de apresentar figuras harmônicas bem definidas, possui glissanos, i.e., variação contínua de altura, que na discretização cria um conjunto de indivíduos com alturas diferentes. Na fase de treinamento, a forma comparativa utilizada se torna

Tab. 5.1: Resultado da consulta de similaridade subjetiva entre material e amostras sonoras. $0 \rightarrow$ ausência de semelhança e $10 \rightarrow$ fortíssima semelhança. RSo = Referência de Sonoridade e CDi = Controle de Dinâmica

CDi	**RSo**	Voz	Guitarra	Sintético
Voz	Voz	8.8 ± 0.4	0.4 ± 0.2	1.2 ± 0.7
	Guitarra	6.4 ± 0.9	0.8 ± 0.3	1.2 ± 0.3
	Sintético	6.2 ± 1.2	1.8 ± 0.5	3.4 ± 0.2
	Todos	8.0 ± 0.7	1.8 ± 0.3	2.4 ± 0.9
Guitarra	Voz	3.4 ± 0.9	5.0 ± 1.0	2.4 ± 0.7
	Guitarra	2.2 ± 0.8	4.0 ± 1.3	2.8 ± 0.7
	Sintético	2.0 ± 1.1	2.8 ± 1.0	4.2 ± 1.9
	Todos	5.8 ± 1.1	1.2 ± 0.3	0.6 ± 0.4
Sintético	Voz	2.6 ± 0.9	3.0 ± 0.7	3.6 ± 0.9
	Guitarra	3.0 ± 1.1	2.6 ± 0.8	4.8 ± 0.8
	Sintético	2.6 ± 0.9	3.4 ± 0.4	5.6 ± 1.2
	Todos	2.2 ± 0.8	1.8 ± 0.8	6.0 ± 1.1

Tab. 5.2: Média dos valores subjetivos. RSo = Referência de Sonoridade e CDi = Controle de Dinâmica.

	Média
CDi	5.2 ± 0.3
RSo	3.9 ± 0.4
Não CDi ou RSo	2.1 ± 0.2
Todas amostras	3.3 ± 0.1

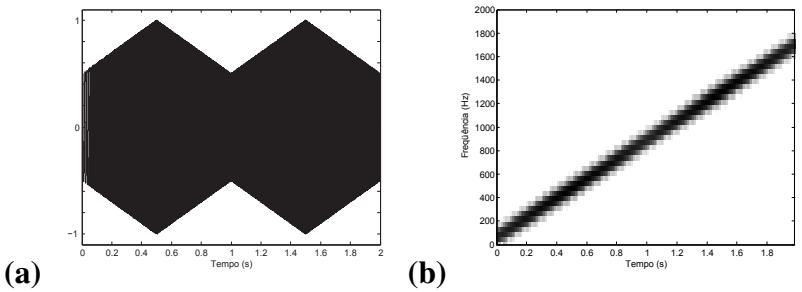
limitada com a abordagem populacional, uma vez que é necessário um indivíduo para representar cada altura. Por isso, quando utilizado como referência de sonoridade, o som de guitarra não gerou menos figuras harmônicas bem definidas que o som sintético. O som sintético apresenta figuras horizontais no espectro que não variam verticalmente, gerando uma variação menor de indivíduos e facilitando a tarefa de treinamento. Esta limitação é causada pela representação utilizada, i.e., atuação direta sobre o espectro.

Contudo, a transferência dinâmica para a saída é evidente e há coerência entre as distribuições de energia da referência de sonoridade e do materia sonoro obtido. No caso da amostra de voz, cuja característica harmônica é menos relevante na determinação da sonoridade, os resultados obtidos são mais próximos da amostra de referência, sendo possível manter a frase proferida.

Experimento 2

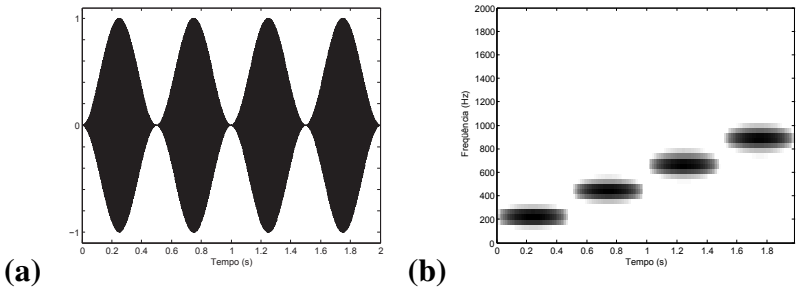
O segundo experimento foi concebido de forma a permitir uma melhor compreensão de como o sintetizador atua em algumas circunstâncias específicas, principalmente quanto ao problema da altura observado no primeiro experimento. Tentou-se verificar um comportamento denominado "Acompanhamento Espectral": ao sistema foi apresentado como referência de sonoridade sons puramente senoidais com altura bem definida (220 Hz, 440 Hz, 660 Hz e 880 Hz), Figura 5.31, e como controle de dinâmica um som com dinâmica bem definida e freqüência linearmente crescente (de 110 Hz a 1760 Hz), Figura 5.30. O resultado é apresentado na Figura 5.31.

Os resultados expressam o comportamento esperado e ajudam a explicar os fenômenos observados no primeiro experimento. Ao sistema, foram apresentados quatro níveis de altura diferentes. Na representação atual, só são aprendidos estes quatro níveis. Há uma incapacidade de gerar amostras com alturas diferentes, no máximo composições das alturas apresentadas. Ao se utilizar um controle de dinâmica de espectro crescente linearmente, a saída gerada se limita aos níveis de altura apresentados e aprendidos. Se fossem apresentados níveis de altura em maior quantidade que o número de indivíduos da população, então o sistema seria incapaz de representar mais níveis.



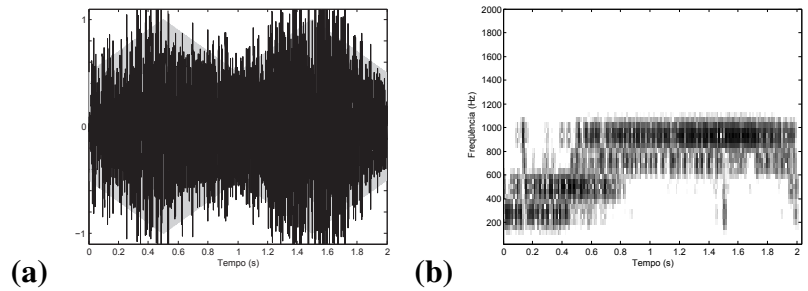
(a) **(b)**

Fig. 5.29: Sinal senoidal de freqüência crescente linear, utilizado como controle de dinâmica, no experimento de verificação de acompanhamento espectral. (a) sinal (b) sonograma.



(a) **(b)**

Fig. 5.30: Sinal utilizado como referência de sonoridade no experimento de verificação de acompanhamento espectral. Fluxo senoidal com quatro freqüências bem definidas. (a) sinal (b) sonograma.



(a) **(b)**

Fig. 5.31: Resultado do experimento de acompanhamento espectral com janela de 1024 pontos após 200 épocas de treinamento com população de 64 protótipos. (a) sinal (b) sonograma.

Uma representação dos indivíduos que abstraísse a altura poderia amenizar esta limitação. Na representação por comparação espectral, o sistema se comporta bem com amostras menos complexas, como o som de fala. Uma forma alternativa de representação é utilizar um sistema de detecção de

freqüência fundamental e realizar as comparações em uma freqüência normalizada. A informação de altura seria incluída na estrutura descritiva de contexto.

O novo modelo foi concebido com o uso de uma ferramenta de detecção de freqüência fundamental, no caso o HPS (Noll, 1969). Cada grânulo é normalizado antes do processo de quantização vetorial. Na recomposição, é determinada a freqüência fundamental do grânulo a ser produzido, seguido da modulação do protótipo para esta freqüência.

A Figura 5.32 apresenta o resultado deste experimento com o modelo com normalização de freqüência fundamental. Ao contrário do modelo anterior, Figura 5.31, o resultado é capaz de generalizar a curva referência em todas as suas freqüências, enquanto no anterior ela representava a partir de freqüências encontradas no conjunto de dados de entrada.

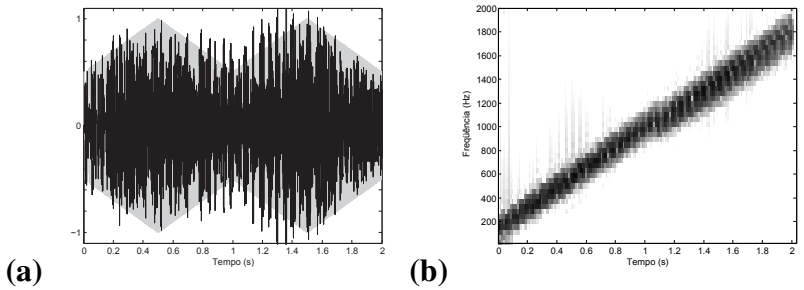


Fig. 5.32: Resultado do experimento de acompanhamento espectral com janela de 4096 pontos e modelo com normalização de freqüência fundamental. (a) sinal (b) sonograma.

Experimento 3

No terceiro experimento, foram aproveitadas as amostras do segundo experimento e utilizado o modelo sem normalização de freqüência. O procedimento foi o mesmo, mas variando o tamanho da janela que caracteriza um grânulo. Os resultados são apresentados nas Figuras 5.33 a 5.37. Variou-se em projeção geométrica de fator duplo a janela na faixa 512 pontos a 8192 pontos (11ms a 92ms em 44 kHz de freqüência de amostragem). Os resultados mostram que grânulos menores implicam

em uma menor definição de freqüência com maior vazamento espectral, enquanto permitem uma dinâmica mais fiel ao controle apresentado.

Para sons cujos transientes atuam com forte influência sobre a percepção de sonoridade, o sistema perde performance de representação com janelas mais extensas. Contudo, para sons cuja característica harmônica é mais influente na percepção, janelas muito curtas diminuem a definição das figuras harmônicas presentes.

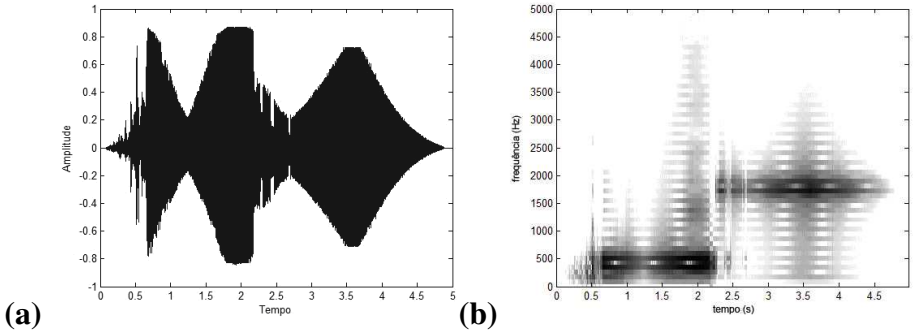


Fig. 5.33: Experimento 3: (a) envelope e (b) espectrograma com tamanho granular 512.

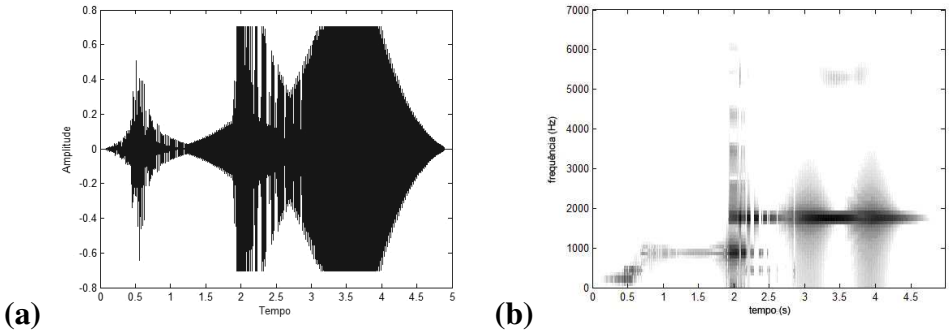


Fig. 5.34: Experimento 3: (a) envelope e (b) espectrograma com tamanho granular 1024.

Experimento 4

O experimento 4 objetivou a ressíntese de um instrumento acústico. Foi utilizado o modelo de representação interna utilizando normalização de freqüência fundamental. Foi escolhida a clarineta

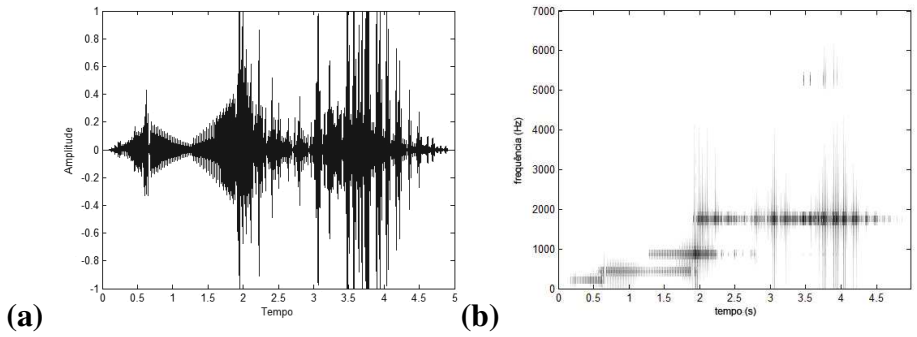


Fig. 5.35: Experimento 3: (a) envelope e (b) espectrograma com tamanho granular 2048.

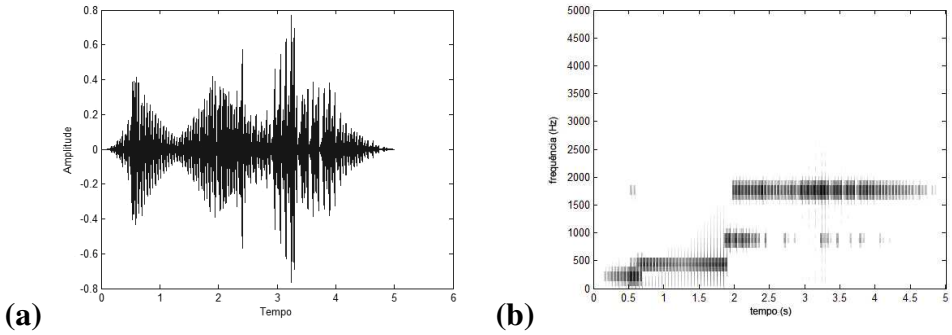


Fig. 5.36: Experimento 3: (a) envelope e (b) espectrograma com tamanho granular 4096.

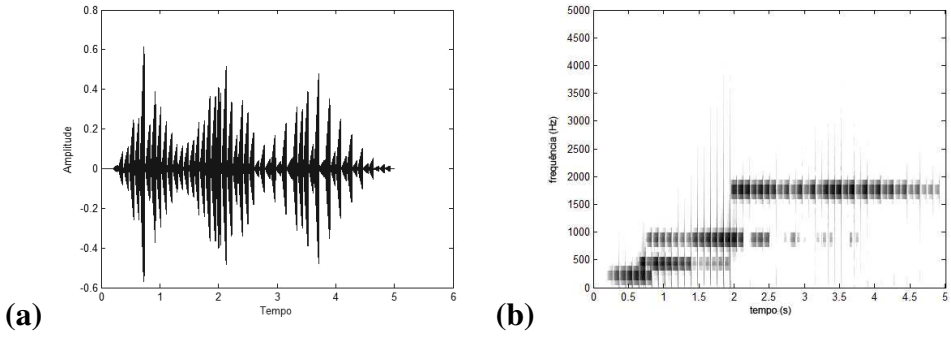


Fig. 5.37: Experimento 3: (a) envelope e (b) espectrograma com tamanho granular 8192.

pela simplicidade do seu comportamento harmônico, permitindo uma fácil visualização gráfica, por espectrograma, dos resultados. Utilizou-se quantização vetorial por mapas auto-organizáveis com 100 neurônios, treinada por 100 épocas. Foi utilizada janela de 1024 com 512 pontos de salto entre os grânulos. Incorporou-se a cada protótipo as informações de freqüência fundamental, energia,

histórico de 2 amostras e posição relativa ao início. Foram apresentadas notas listadas na Tabela 5.3.

Tab. 5.3: Notas de clarineta utilizadas para o treinamento no Experimento 4

Nota	Freqüência Fundamental
Ré3 (D3)	146,83
Ré#3 (D#3)	155,56
Mi3 (E3)	164,81
Lá#3 (A#3)	233,08
Dó4 (C4)	261,63
Ré4 (D4)	293,66
Lá#4 (A#4)	466,16
Dó5 (C5)	523,25
Ré5 (D5)	587,33
Lá#5 (A#5)	932,33
Dó6 (C6)	1046,50
Ré6 (D6)	1174,66

Na primeira execução, foram sintetizadas notas incluídas nas treinadas, no caso: Dó4, Figura 5.38; Dó5, Figura 5.39; e Dó6, Figura 5.40. Em uma execução seguinte, o som da clarineta foi novamente ressintetizado, mas em freqüências não presentes durante o treinamento. Nesse processo, uma amostra de clarineta em outra freqüência, foi utilizada como controle de dinâmica, tendo o determinador de freqüência fundamental alterado para retornar um valor específico e não o real da amostra de controle. A Figura 5.41 apresenta os resultados para freqüências de 400Hz e 1000Hz.

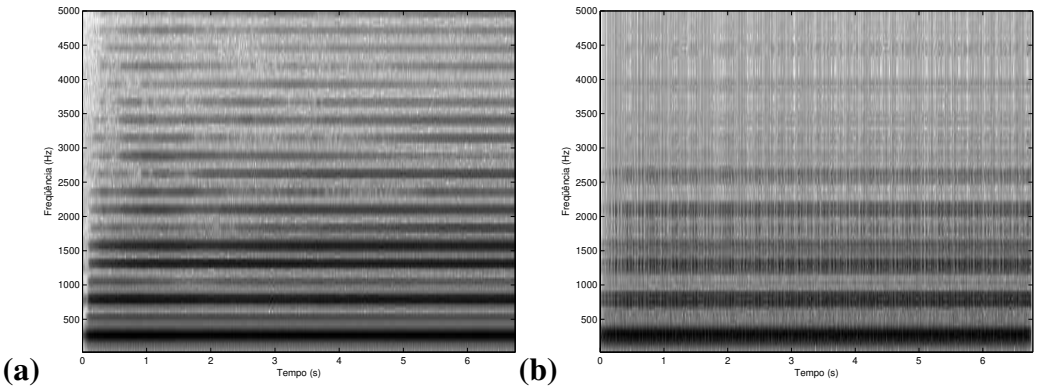


Fig. 5.38: Ressíntese do som da clarineta. (a) sinal original em Dó4 (C4) (freqüência fundamental = 261,63 Hz) (b) resultado da ressíntese pelo SGBP.

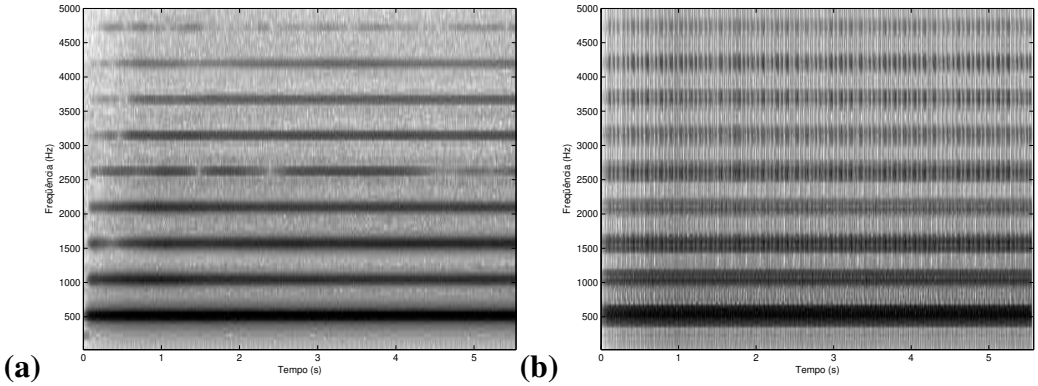


Fig. 5.39: Ressíntese do som da clarineta. (a) sinal original em Dó5 (freqüência fundamental = 523,25 Hz) (b) resultado da ressíntese pelo SGBP.

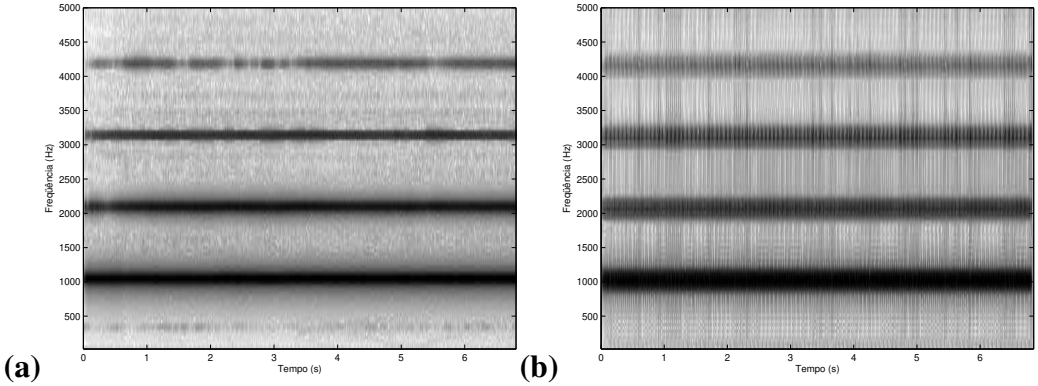


Fig. 5.40: Ressíntese do som da clarineta. (a) sinal original em Dó6 (freqüência fundamental = 1046,50 Hz) (b) resultado da ressíntese pelo SGBP.

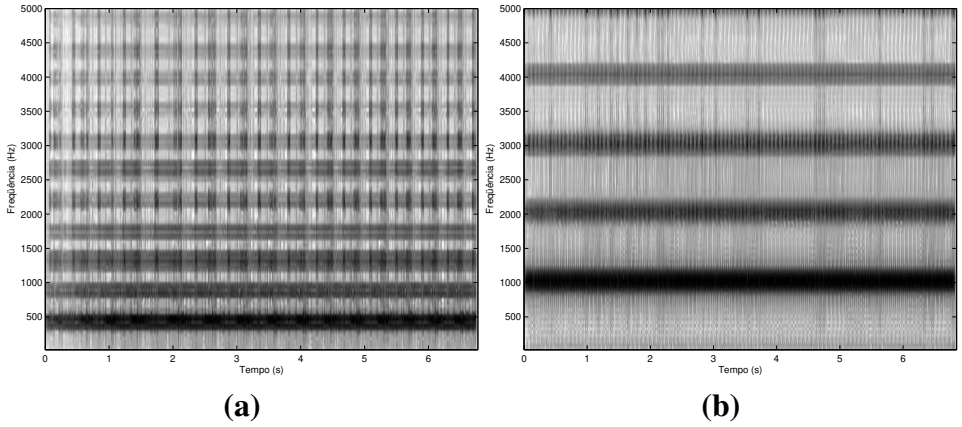


Fig. 5.41: Ressíntese do som da clarineta em outras freqüências: (a) 440 Hz e (b) 1000 Hz.

Os resultados obtidos lembram fortemente os sons utilizados para gerá-los. O comportamento harmônico foi sintetizado com bastante semelhança ao original. No caso dos sons de freqüência fundamental não utilizada no treinamento, o comportamento obtido foi o esperado. Contudo, nas amostras foi possível observar a presença de pequenos artefatos de freqüência (que também podem ser ouvidos nas amostras sonoras) com certa periodicidade.

Estes experimentos testaram uma sonoridade controlada por um sinal com comportamento semelhante ao treinado. A característica harmônica é muito forte nesses sinais. Para testar o comportamento da sonoridade em um sinal menos harmônico foi utilizado um controle dinâmico com um sinal de fala, Figura 5.42. A fala é caracterizada por ser composta por momentos de ruído e outros com forte característica harmônica. Assim, o som sintetizado é capaz de lembrar a sonoridade original em momentos de som de fala harmônica, como entre os segundos 3 e 6 da Figura 5.42. Porém, em momentos puramente ruidosos, como próximo aos 9 segundos, tenta-se aproximar um sinal ruidoso por um protótipo harmônico, não permitindo semelhança com nenhuma das amostras.

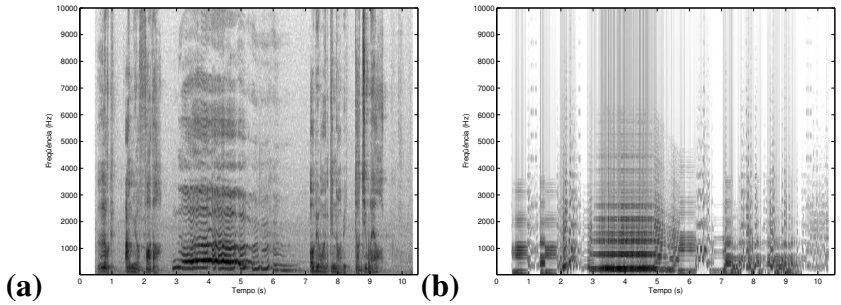
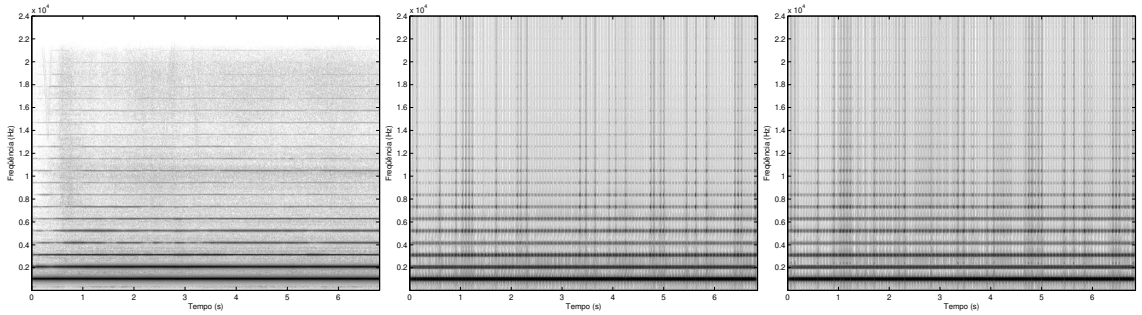


Fig. 5.42: Voz controlando clarineta: (a) som original de voz e (b) som sintetizado.

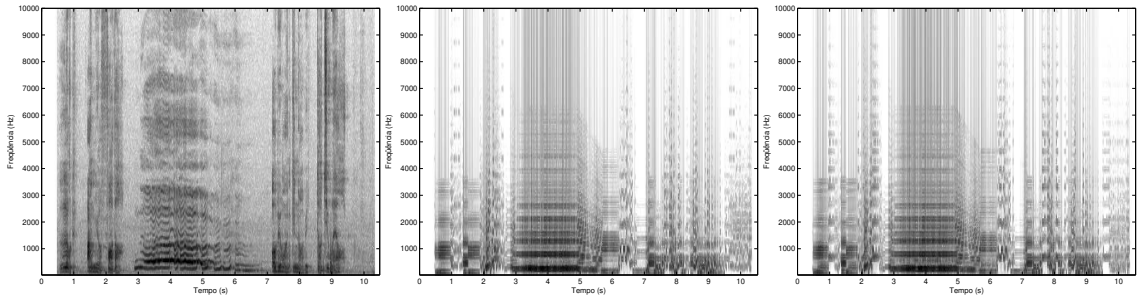
Em outra execução do experimento, o material de fala foi adicionado ao grupo de treinamento. A mesma configuração da execução anterior foi aplicada. A Figura 5.43 apresenta o som controlado pela clarineta em Dó6 (C6), nesta execução, frente ao gerado sem a parte de voz no grupo de treinamento. A Figura 5.44 apresenta o som controlado pelo sinal de voz nas duas execuções. O resultado obtido neste caso apresenta boa compreensão nas partes de natureza harmônica, como na região próxima ao segundo 5. Contudo, não é possível a compreensão da frase ditada como no caso do experimento

1, Figura 5.14. Possivelmente, a causa foi a política de tratamento da freqüência fundamental, que implicou em modulações em sinais puramente ruidosos. Assim, para o tratamento dos dois casos, o ideal seria que o mecanismo de contexto considerasse o fato de um grânulo ser ruidoso ou harmônico.



(a) **(b)** **(c)**

Fig. 5.43: Diferentes entradas de sonoridade controladas pela (a) clarineta Dó6 (C6): (b) apenas som de clarineta e (c) som de clarineta e voz.



(a) **(b)** **(c)**

Fig. 5.44: Diferentes entradas de sonoridade controladas pelo (a) som de voz: (b) apenas som de clarineta e (c) som de clarineta e voz.

Com relação à manutenção de energia, a Figura 5.45 apresenta o sinal obtido e o original no controle pela clarineta Dó6 e a Figura 5.46 pela voz. O envelope de energia aparenta ser mantido satisfatoriamente em todos os experimentos.

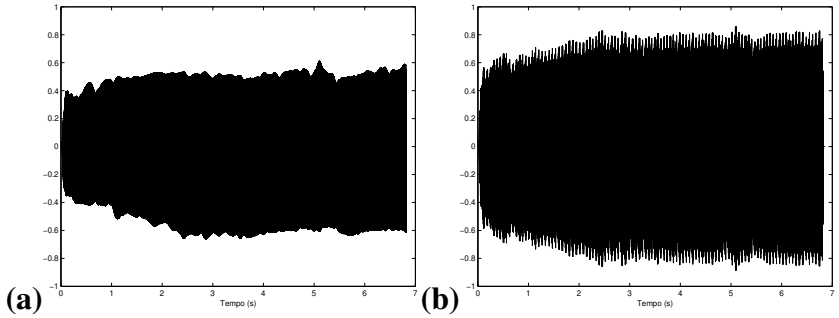


Fig. 5.45: Manutenção de energia: (a) clarineta em Dó6 (C6) e (b) som sintetizado.

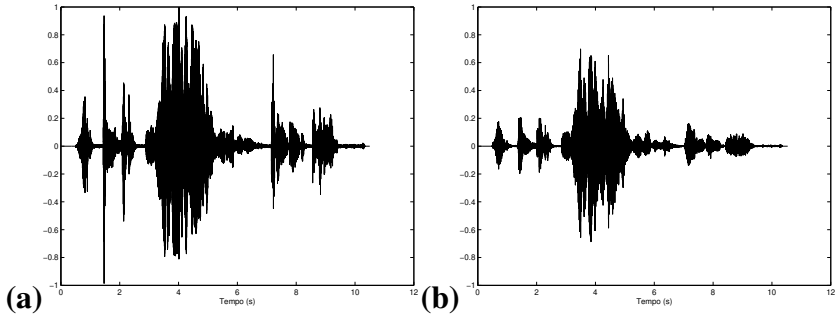


Fig. 5.46: Manutenção de energia: (a) voz e (b) som sintetizado.

Discussão

O SGBP se enquadra de forma satisfatória como um mecanismo de síntese sonora por conteúdo, aplicando o conceito de representação sonora por analogia acústica. A versão implementada foi bem sucedida na tarefa de sintetizar material sonoro com características das amostras descritivas, como pode ser atestado pelo leitor nos quatro experimentos ou através da pesquisa subjetiva empregada no primeiro experimento.

O modelo utilizado aplicou uma estratégia de manipulação direta sobre o espectro, com o auxílio de dados ambientais encapsulados em uma estrutura descritiva de contexto. A atuação direta sobre o espectro é uma atividade penosa, devido à complexidade dinâmica do som e à alta dimensionalidade da representação espectral (da mesma ordem dimensional do sinal puro, se considerada a fase). Desta forma, cabe ao mecanismo de aprendizado generalizar estruturas naturais do som, como freqüência fundamental, através de auto-organização. A escolha da representação por espetro já é uma tentativa

de organizar os dados em uma base mais natural para a percepção. O uso da estrutura de contexto e normalização de freqüência fundamental são mecanismos que amenizam este problema, por agregarem pré-organização ao processamento. Contudo, a abstração de várias outras nuances físicas, que implicam em qualidades cognitivas associadas ao som, são passadas como responsabilidade do mecanismo de processamento.

O uso de estruturação no espectro não é uma tarefa simples. A estruturação da freqüência fundamental, no Experimento 4, apresentou bons resultados com amostras fundamentalmente harmônicas, no caso, o som de clarineta, resolvendo o problema de acompanhamento espectral discutido no Experimento 2. Porém, com amostras de natureza ruidosa, como, por exemplo, as amostras de voz, a estruturação implicou em uma maior distorção dos resultados.

Assim, o modelo apresentado foi suficiente para ilustrar o paradigma de síntese. Porém, para aplicação em situações reais, em síntese sonora, são necessárias algumas melhorias, desde a implementação em uma plataforma mais amigável e integrável, com softwares de produção musical, até o estabelecimento de uma representação mais estruturada e condensada, que permita a aplicação mais eficiente de processamento computacional.

Capítulo 6

Representação por analogia semântica

A base da representação semântica é o estabelecimento de um conjunto de ícones que contenham um significado, e que possam ser organizados em uma linguagem descritiva. O ponto crucial deste processo é a atribuição destes significados, a determinação da fronteira de uma sensação ou fenômeno que permita discriminar conceitos.

Na representação qualitativa de sons, dois fatores primordiais a serem considerado são: a distribuição espectral e a escala temporal. Com relação ao espectro, obtém relevância não só a sua composição, seja em elementos harmônicos ou ruidosos, mas também a forma como ele evolui no tempo. De fato, é impossível dissociar o fator espectral da noção temporal, pois são correlacionados em suas próprias definições. Segundo Smalley (1986), "espectro é percebido pelo tempo e tempo é percebido como movimento espectral"[1]. O princípio da incerteza de Heisenberg (1927) delimita a observação do fenômeno: é impossível ter uma noção instantânea de freqüência. Gabor (1946), na instauração do seu "quanta sonoro", relaciona estes limites de resolução de tempo e de freqüência com a percepção humana do som. Com isso, Gabor estipula uma base mínima de informação acústica, a qual podem ser atribuídos valores qualitativos. A Figura 6.1 ilustra graficamente o quanta de Gabor. O ser humano necessita ser exposto a uma variação de freqüência (Δf) mínima e por um tempo de exposição (Δt) mínimo para a discriminação entre duas freqüências.

[1] *"spectrum is perceived through time and time is perceived as spectral motion"*

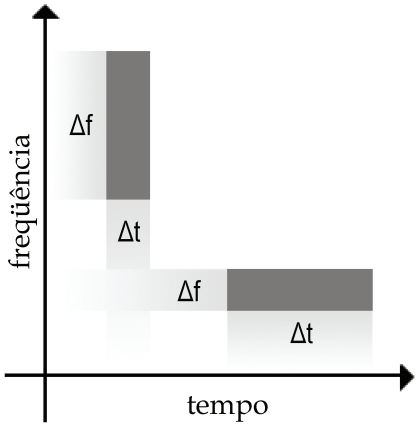


Fig. 6.1: Quanta sonoro de Gabor, representado nos retângulos escuros. Resolução em freqüencia (Δf), dependente da resolução temporal (Δt), estabelece limites de percepção sonora.

Na questão da escala temporal, o "quanta sonoro" de Gabor é o nível perceptível mais instantâneo. Porém, o som pode ser analisado em outros níveis: do granular, na escala instantânea, até o nível dos eventos sonoros. Para cada nível, é possível estabelecer o seu próprio conjunto de descritores qualitativos. Contudo, estes níveis não são independentes. A evolução de uma qualidade em uma dada escala influencia a emergência de outras qualidades em uma escala mais externa. Por exemplo, uma rápida evolução da energia de grãos (escala granular) determina um tipo específico de ataque (escala gestual). Um ataque rápido, seguido de um decaimento lento e uma pequena sustentação, define a sonoridade de um evento sonoro individual (escala de eventos). Assim, um som ou um evento sonoro pode ser visto como uma trajetória no espaço de qualidades instantâneas. Esta é a base da morfologia e tipologia sonora. A Figura 6.2 faz uma ilustração deste processo.

6.1 Linguagens descritivas

O contexto descritivo exerce forte influência no processo de concepção de uma linguagem. Entende-se como contexto descritivo o meio de aplicação de material sonoro que demanda algum tipo de mecanismo descritivo. O contexto melhor desenvolvido é o da música. O próprio formalismo da teoria musical tradicional se enquadra como uma linguagem descritiva, estabelecendo esta como a base em que se desenvolve a expressão artística. Como exemplo, uma partitura é capaz de descrever

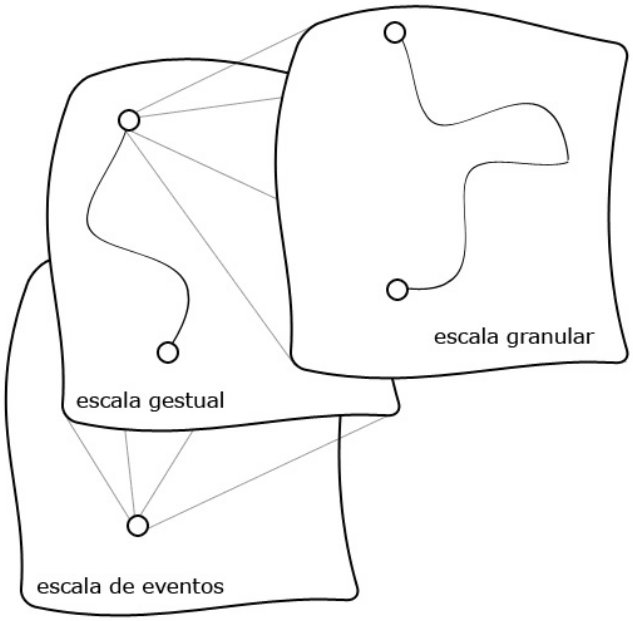


Fig. 6.2: Múltiplas escalas de percepção sonora.

o instrumento, a altura, a duração e a intensidade de uma nota musical e, na música tradicional, a obra musical é descrita a partir destes elementos.

O final do Século XIX e a primeira metade do Século XX viram surgir, na música de concerto, uma série de experiências com o uso da sonoridade como material musical. O efeito desse alargamento do vocabulário musical, sobre o discurso musical e sobre a percepção dos ouvintes começou a ser estudado e aplicado de forma sistemática a partir do início da Música Eletroacústica. Cabe destaque para o movimento da Música Concreta, iniciada por Pierre Schaeffer em 1948, e para a publicação do Tratado dos Objetos Musicais, em 1966, também por Schaeffer. Um dos pontos importantes do tratado é a idéia de que, na música tradicional, a construção se dá da linguagem para o material. O compositor parte de estruturas abstratas pré-estabelecidas pela linguagem, como escalas e acordes, para, a partir deles, gerar o material sonoro. Schaeffer sugere que, ao invés disso, a música deveria ter a escuta do material sonoro como ponto de partida. Somente a partir da escuta de um conjunto específico de sons e da seleção de suas características mais relevantes e que a linguagem poderia ser estruturada. Nos termos de Emmerson (1986), ao invés da música ser uma linguagem abstrata, que seria imposta ao material sonoro, ela seria abstraída desse material através da escuta.

Esse processo de abstração envolve a criação, mesmo que inconsciente, de uma tipologia sonora. O ouvinte deve estabelecer famílias de sons, de acordo com suas características morfológicas, para separá-los da paisagem sonora a que pertencem. Essas características morfológicas e sua articulação entre as famílias de sons estabelecidas pela tipologia formam, segundo Schaeffer, a base da construção da linguagem musical. Como a escuta e, consequentemente, a criação da tipologia e a identificação das morfologias são subjetivas, a cada escuta um discurso musical diferente é criado pelo ouvinte, mesmo que a partir do mesmo sinal acústico. Apesar da fluidez e da subjetividade atribuídas ao processo de escuta por Schaeffer, ele apresenta em seu tratado um exemplo de tipo-morfologia sonora que serviu de guia para muitos músicos e professores de música, na abordagem desse vocabulário alargado. A partir deste primeiro exemplo, vários outros apareceram nas obras de músicos e teóricos como Smalley, Wishart e Bayle. Exemplos que, sempre reafirmando a subjetividade das tipo-morfologias que emergem de cada escuta, buscaram novas formas de representar a experiência do material sonoro não articulado e o processo de organizá-lo em um discurso musical.

Como exemplo ilustrativo do discurso musical, temos o trabalho de Smalley (1997). Na composição de sua linguagem, ele divide a evolução do som musical em três fases: início, continuação e cessação. Cada fase pode ser qualificada de acordo com os adjetivos apresentados na Tabela 6.1. Esta taxonomia é um exemplo de linguagem descritiva do sinal musical, partindo da ótica do músico.

Tab. 6.1: Termos utilizados por Smalley (1997) para descrever as fases do som musical.

Inícios		Continuações		Cessações	
Termo	Tradução	Termo	Tradução	Termo	Tradução
departure	partida	passage	passagem	arrival	chegada
emergence	emergência	transition	transição	disappearance	desaparecimento
anacrusis	anacrusis	prolongation	prolongação	closure	encerramento
attack	ataque	maintenance	manutenção	release	liberação
upbeat	ascendente	statement	declaração	resolution	resolução
downbeat	descendente			plane	planificação

Outro contexto relevante é o da descrição para fins computacionais. O objetivo, nesse caso, é

determinar um domínio que permita a manipulação algorítmica do material sonoro, o que exige uma representação matemática rígida. Este contexto extrapola o da música e, em geral, não cria linguagens tão completas e complexas com relação às possibilidades de uso artístico. As aplicações são mais específicas e podem incluir uma intersecção com uma descrição musical, como, por exemplo, na representação de timbres.

6.1.1 Padrões descritivos

A representação semântica tem sido estudada para a descrição de material sonoro para fins computacionais. Entre outras iniciativas, existe um workshop internacional denominado "Learning the Semantics of Audio Signals" (Cano et al., 2006), que significa "Aprendendo a semântica de sinais sonoros", incluído na conferência internacional em semânticas e tecnologias de mídias digitais (SAMT). Um pressuposto importante é de que uma representação ideal seria capaz de descrever a sensação sonora desejada em sua completude e sem redundâncias, assim, permitindo descrever um som com o número mais reduzido possível de dados. As possibilidades incluem a aplicação de novos paradigmas de indexação, mecanismos de busca em coleções, transmissão e armazenamento de mídias.

A tendência atual mais forte é a dos esquemas descritivos padrão como o MPEG-7 (Manjunath et al., 2002), que tenta estabelecer um padrão de figuras semânticas, e outros como o CLAM (Amatrian et al., 2002), que apenas estabelecem o esquema descritivo, deixando em aberto as figuras semânticas a serem utilizadas. Estes padrões descritivos formam uma base para o estabelecimento de uma linguagem. Contudo, esta linguagem evolui com a incorporação de novas figuras semânticas, diferentemente de Smalley e Schaeffer, que já partem da necessidade de instituir um conjunto de semânticas e gramáticas definidas para a representação em música. Esta diferença se deve à distinção entre os objetivos propostos para cada linguagem.

As figuras semânticas são denominadas descritores e podem se referenciar as qualidades perceptuais ou físicas do sinal sonoro. Alguns estudos, como Lakatos et al. (2000), Rocchesso (2001) e

Cook & Lakatos (2003), investigaram a capacidade do ser humano de perceber estes atributos físicos em fontes sonoras. Ottaviani & Rocchesso (2004) apresentaram um estudo que indica que ouvintes diferentes podem perceber facilmente os mesmos atributos. Assim, é possível criar um conjunto de figuras semânticas padrão.

Descritores do MPEG-7

No esquema do MPEG-7, apresentado em Martínez (2004), são instituídos dois grupos de descritores: de baixo e de alto nível. Os descritores de baixo-nível, listados na Tabela 6.2, são divididos em 7 grupos: básico, espectral básico, parâmetros do sinal, timbre (temporal e espectral), base espectral e silêncio. Os grupos básico e espectral básico fornecem descritores da forma de onda e do seu espectro. Os parâmetros de sinal se aplicam à distinção entre sinais periódicos e sinais não periódicos. O grupo de base espectral permite a descrição por uma projeção do espectro em um espaço de menor dimensão. Os grupos de timbre englobam descritores relacionados a qualidades do som que formam a base para a caracterização de um timbre.

Os descritores de alto nível tentam explicar o material sonoro de acordo com a sua aplicação. Um exemplo é o descritor de melodia que cria um contorno de uma melodia monofônica do sinal representado para fins de busca por similaridade melódica, como, por exemplo, busca por *humming* (Ghias et al., 1995). Outro exemplo é o descritor de texto, que lista o texto dito em um arquivo de áudio, permitindo, por exemplo, a indexação de uma base de discursos.

O descritor de alto nível mais relevante no contexto deste trabalho é o descritor de timbre de instrumento musical. São permitidos quatro grupos de instrumentos, sendo dois bem definidos: instrumentos harmônicos e instrumentos percursivos. Um instrumento harmônico é definido a partir dos descritores de timbre espectral somado ao tempo de ataque. Um instrumento percursivo é definido a partir dos descritores de timbre temporal somado ao centróide espectral.

A concepção de que noções de alto nível são compostas a partir de descrições de baixo nível

Tab. 6.2: Descritores de baixo-nível do MPEG-7

Tipo	Descritor
Básico	Mínimo-máximo da forma de onda Energia
Espectral básico	Envelope do espectro Centróide do espectro Espaçamento do espectro Nivelamento do espectro
Parâmetros do sinal	Harmonicidade Freqüência fundamental
Timbre (temporal)	Tempo de Ataque (Log) Centróide Temporal
Timbre (espectro)	Centróide do espectro harmônico Desvio do espectro harmônico Espaçamento do espectro harmônico Variação do espectro harmônico Centróide do espectro
Base espectral	Base Projeção
Silêncio	Silêncio

não é exclusivo ao MPEG-7 (ou mesmo original deste). Na literatura, existe uma série de outros descritores de baixo nível e associações destes com qualidades cognitivas. Por exemplo, o centróide espectral, utilizado no MPEG-7, é geralmente associado à escala de brilho, seja de forma direta ou ajustado pela freqüência fundamental (Schubert & Wolfe, 2006).

Descritores do CATERPILAR

Em Schwarz (2004), é realizada uma listagem de descritores de baixo nível utilizadas no sistema de síntese concatenativa CATERPILAR. Schwarz os divide em quatro grupos básicos de descritores (Tabela 6.3): de sinal; espectrais; harmônicos; e perceptivos.

Os descritores de sinal correspondem a descritores obtidos diretamente de ferramentas de processamento digital de sinais. Os descritores de energia são obtidos a partir da média quadrática do sinal, em escala linear e/ou logarítmica, e tendo sua derivada calculada pela diferença de valor

Tab. 6.3: Descritores de baixo nível utilizados em Schwarz (2004)

Tipo	**Descritor**
Sinal	Frequência fundamental (f_0)
	Derivada de f_0 (Δf_0)
	Energia
	Energia logaritmica
	Derivada de energia logarítmica
	Taxa de cruzamento em zero
	Auto-correlação de primeira ordem
Espectral	Centróide espectral
	Inclinação espectral
	Espaçamento espectral
	Assimetrial espectral
Harmônico	Razão de energia harmônica
	Tristimulus $(1, 1)$
	Tristimulus $(2, 4)$
	Tristimulus $(5, ...)$
	Paridade harmônica
	Desvio harmônico
Perceptivo	*Loudness*
	Sharpness
	Largura Timbral

entre dois pontos instantâneos consecutivos. A freqüência fundamental (f_0) é obtida pelo método de picos espectrais (Rodet, 1997). A taxa de cruzamentos em zero corresponde ao número de vezes que há inversão de sinal (positivo/negativo). A autocorrelação de primeira ordem é obtida pela Equação 6.1, em que N representa o número de amostras e s_i o valor da amostra no momento i:

$$AR1 = \frac{1}{N} \sum_{i=1}^{N-1} s_i \cdot s_{i+1} \tag{6.1}$$

Os descritores espectrais são extraídos do espectro de magnitude de uma transformada discreta de Fourier, obtido pela aplicação do algoritmo FFT. O centróide espectral é o centro de gravidade do espectro de freqüência, obtido a partir da Equação 6.2, em que: N é o número de amostras (e conseqüentemente o número de amostras de freqüências no espectro) e a_i e f_i são, respectivamente,

a energia e a freqüência da amostra de freqüência i:

$$centroide\_espectral = \frac{\sum_{i=1}^{N} a_i f_i}{\sum_{i=1}^{N} a_i} \tag{6.2}$$

Inclinação espectral, ou "tilt", corresponde ao decaimento de energia na escala de freqüências do espectro. Ela é calculada a partir de um procedimento de regressão linear (Rodet & Tisserand, 2001) e conversão de unidades (Rodet & Schwarz, 2000). Espaçamento espectral corresponde ao desvio padrão das amostras (segundo momento centrado). Assimetria espectral corresponde à obliqüidade das amostras (terceiro momento centrado).

Descritores harmônicos correspondem à modelagem do sinal a partir de síntese aditiva e ruído (Risset & Mathews, 1969; Serra & Smith, 1990), i.e., decomposição em componentes harmônicos. Estes descritores apenas são aplicáveis a sons modeláveis desta forma, sendo que a medida de razão de energia harmônica (em relação à energia de ruído) informa sobre a aplicabilidade dos descritores. Esta medida é obtida pela razão da energia do sinal sintetizado puramente harmônico pela energia total do sinal.

Paridade harmônica é a razão entre a energia dos harmônicos pares e a energia de todos os harmônicos. *Tristimulus* (X, Y) (Jehan, 1997) é a razão entre a somatória da energia dos componentes harmônicos compreendidos entre o X-ésimo e o Y-ésimo harmônicos e a energia total do sinal harmônico. São utilizadas por Schwarz 3 faixas: $(1, 1)$, relação de energia entre a fundamental e o sinal harmônico; $(2, 4)$, segundo, terceiro e quarto harmônicos; e $(5, ...)$, todos os harmônicos a partir do quinto. Desvio harmônico é uma soma, ponderada pela amplitude A_i de cada parcial i, da diferença entre a freqüência efetiva f_i e a freqüência ideal if_0, expressa na Equação 6.3:

$$desvio\_harmonico = \frac{\sum_{i=1}^{num\_parciais} A_i \frac{f_i - if0}{f_0}}{\sum_{i=1}^{num\_parciais} A_i} \tag{6.3}$$

Os descritores perceptivos são extraídos de estudos de psico-acústica (Zwicker, 1982; Moore, 2003; Cook, 1999). O primeiro é uma medida de *loudness* baseada no modelo de bandas críticas chamadas "bandas Bark" (Moore et al., 1997). Neste modelo, através de filtragem é estipulado um valor de *loudness* para cada banda individual que, por fim, são somados para compor o valor global. A medida segue a Equação 6.4, em que é designado o *loudness* específico de uma banda i (de um total de K bandas) por L_i:

$$loudness = \sum_{i=1}^{K} L_i \tag{6.4}$$

Outro descritor perceptivo é o de *sharpness*[2], que é relativo ao conteúdo de alta freqüência. Ele é calculado de forma semelhante ao valor de *loudness*, porém, com uma ponderação mais elevada pelas altas freqüências, definido pela função $g(i)$ na Equação 6.5:

$$g(i) = \begin{cases} 1 & \text{se } i < 15 \\ 0.066e^{0.171i} & \text{se } i \geq 15 \end{cases}, \tag{6.5}$$

$$sharpness = 0.11 \frac{\sum_{i=1}^{K} ig(i)L_i}{loudness}. \tag{6.6}$$

O terceiro descritor perceptivo é chamado de Largura Timbral[3]. Trata-se de uma medida da faixa dinâmica do *loudness* entre as várias bandas, calculada pela Equação 6.8, como o fator entre o *loudness* específico máximo, L_{max} da Equação 6.7, e o *loudness* global, na forma:

[2] grosseiramente traduzido como "cortante"
[3] *Timbral Width*

$$L_{max} = \max_{1 \leq i \leq K} L_i \tag{6.7}$$

$$largura\_timbral = \left(\frac{loudness - L_{max}}{loudness} \right) \tag{6.8}$$

6.2 Síntese sonora semântica

Síntese sonora semântica consiste em gerar material sonoro descrito a partir de uma representação semântica. Como linguagem, entende-se um conjunto de figuras semânticas que sejam associadas a alguma qualidade ou relação, a qual podem ser associados valores.

As considerações fundamentais são: como definir uma linguagem e como estabelecer uma relação de controle entre os valores das figuras semânticas e os mecanismos de geração sonora.

6.2.1 Mapeamento paramétrico

Mapeamento paramétrico é um método de portar o espaço paramétrico de algum mecanismo para um espaço artificial. As motivações principais para este procedimento são as restrições operacionais que a natureza do espaço provoca em dispositivos de controle. No caso de controle humano, é uma forma de sobrepassar procedimentos de controle pouco intuitivos.

No caso dos sintetizadores de som, existem dois alentos principais: controle instrumental ou gestual e síntese sonora.

Com o controle instrumental, o objetivo é estabelecer uma interface física para o sintetizador que

se assemelhe a um instrumento. Desta forma, estabelece-se um conjunto de gestos e digitações em que são mapeados os parâmetros de síntese. Wanderley et al. (1998) nomeia este procedimento de mapeamento de composição de instrumentos. Exemplos são Garnett & Goudeseune (1999), Hunt & Kirk (2000), Verfaille et al. (2006) e Carrillo (2006). Lee & Wessel (1992) comentam que a principal motivação é que mecanismos de síntese, em geral, possuem um número de parâmetros impossível de ser controlado de forma explícita por seres humanos.

Em síntese sonora, o objetivo é permitir uma exploração mais consciente do espaço de possibilidades sonoras. A política é a de determinar um conjunto semântico que defina as qualidades subjetivas de controle e estabelecer um mapeamento para o espaço paramétrico. Neste caso, alguns parâmetros possuem uma relação direta e intuitiva com alguma qualidade perceptiva. É o caso da freqüência de um oscilador e a sensação de altura. Mas em geral, em mecanismos de síntese mais sofisticados, o efeito sobre a percepção de uma qualidade mais complexa, como a sonoridade, de um parâmetro é imprevisível e fortemente dependente da configuração de outros parâmetros. Assim, um mapeamento do espaço paramétrico para um espaço que possa ser explorado mais intuitivamente tende a ser fortemente acidentado e não-linear.

Nos dois casos de mapeamento, a tendência é que exista um número inferior de qualidades subjetivas comparadas aos parâmetros de síntese. Garnett & Goudeseune (1999) discutem que, como normalmente a dimensão do espaço paramétrico é superior à do espaço de controle, há a necessidade de uma técnica de interpolação de alta dimensionalidade. Garnett & Goudeseune (1999) utilizam um método de interpolação geométrica encontrado em Choi et al. (1998). Outras técnicas incluem o uso de redes neurais (Fels & Hinton, 1998) e interpolação pelo vizinho mais próximo[4] (Bencina, 2005).

6.2.2 Construção de linguagem

A definição do conjunto semântico pode seguir duas linhas: o estabelecimento de linguagem inata ou adaptativa. Como linguagem inata, entende-se o uso de expressões semânticas determinadas

[4] *nearest neighbor interpolation*

por atributos percebidos de forma comum, enquanto a linguagem adaptativa é definida pelo próprio usuário.

A escolha da linha a se seguir remete à discussão de Hofstadter e Forbus (Seção 3.2) a respeito da origem da representação e as considerações relativas a flexibilidade e emprego em aplicações reais. Hofstadter defende que representações inatas não são flexíveis o suficiente para terem o poder de descrição do ser humano, enquanto Forbus diz que as representações adaptativas são genéricas demais, não sendo capazes de estabelecer modelos funcionais em aplicações reais.

A grande diferença do tipo de linguagem se apresenta no processo de mapeamento. O uso de uma linguagem inata possibilita a aplicação de mapeamentos inatos, i.e., a partir de observações e definições formais de que características físicas estimulam uma dada sensação. Um exemplo é definir uma relação linear entre altura percebida e a freqüência de um oscilador. Isto não é possível no caso de uma linguagem adaptativa. As relações que sustentam o mapeamento devem ser extraídas da análise de dados reais, os quais estão vinculados à avaliação subjetiva.

Na linha da linguagem inata, Matt Hoffman, do grupo de Perry Cook em Stanford, criou um modelo de síntese denominado *Feature-Based Sound Synthesis* (Hoffman & Cook, 2006). Este sintetizador aplica um mapeamento paramétrico sobre um grupo semântico pré-definido, que corresponde às "propriedades acústicas de baixo nível que ouvintes humanos usam para deduzir os atributos físicos mais complexos da fonte sonora"[5]. O que Hoffman propõe é que as qualidades relevantes são comuns a ouvintes distintos e que se deve desenvolver uma representação qualitativa semântica a partir destas qualidades. Sua arquitetura utiliza três estruturas: Analisador de qualidades, sintetizador paramétrico e métrica de distância. O funcionamento do sistema consiste em manipular os parâmetros de síntese de forma a minimizar a distância entre as qualidades desejadas e as qualidades extraídas do material sonoro pelo analisador de qualidades.

Em um modelo em tempo-real, Hoffman & Cook (2007) utilizam como objetos semânticos:

[5] *low-level acoustical properties that human listeners use to deduce the more complex physical attributes of a sound's source*

brilho, associado ao valor do centróide espectral; harmonicidade, com uma métrica da força com que a altura é definida[6]; e os cinco primeiros coeficicentes mel-cepstrais, que são extraídos da voz do interprete. Desta forma, no seu modelo executa-se um controle gestual e semântico. Da forma semelhante, Verfaille & Arfib (2001) utilizam uma linguagem inata como camada intermediária em uma composição de instrumento. MUSART (Joseph & Lodha, 2002) e *the Sonification Sandbox* (Walker & Cothran, 2003) são sistemas que realizam mapeamento utilizando parâmetros derivados de teoria musical, com certas variações na flexibilidade dos parâmetros.

Nestes modelos, o mapeamento é estabelecido pela relação das características físicas com a qualidade semântica, incorporada no sistema de análise qualitativa. Em outra política de mapeamento, Johnson & Gounaropoulos (2006) utilizam um subconjunto semântico fixo definido em Etherington & Punch (1994), Tabela 6.4, i.e., qualidades normalmente inatas. Porém, seu mapeamento é determinado por treinamento estatístico sobre dados subjetivos. Assim, o estabelecimento da linguagem é inato, mas o mapeamento é adaptativo.

Tab. 6.4: Figuras semânticas utilizadas em Johnson & Gounaropoulos (2006), retiradas de Etherington & Punch (1994)

Figura semântica
brilhante (bright)
quente (warm)
áspero (harsh)
batido (hit)
pinçado (plucked)
constante (constant)
denso (thick)
metálico (metallic)
amadeirado (woody)

No uso de semântica adaptativa, o mapeamento entre a linguagem e os parâmetros é realizado a partir de dados reais, obtidos com base em uma análise subjetiva. Este tipo de mapeamento se baseia em um descritor não computacional subjetivo, i.e., o compositor. Por este motivo, ele está sujeito a contradições e inconsistências causadas por algum evento contextual. Um exemplo é um compositor

[6] *"a measure of how strongly pitched the sound is"*

escutar uma amostra por mais de uma vez e associar qualidades diferentes a ela, por algum motivo subjetivo, como, por exemplo, o humor. Por outro lado, este evento contextual abre um leque de possibilidades de mapeamento que seria capaz de fornecer uma ferramenta de síntese adaptada às condições do humor do compositor.

A crítica levantada por Hoffman & Cook (2006) a este tipo de síntese é a necessidade de se realizar um mapeamento diferente a cada vez que se estipular um novo mecanismo de síntese. Isso é verdade, uma vez que o mapeamento é dependente do espaço paramétrico do sintetizador. Por outro lado, se o mecanismo de síntese permitir análise, i.e., obtenção automática dos parâmetros a partir da amostra sonora, como, por exemplo, no uso de codecs, a associação das qualidades adaptativas e as amostras no conjunto de treinamento não precisa ser reconstruída. Assim, o fardo é a simples realização do treinamento, tarefa minimamente custosa relativo ao processo de associação qualitativa subjetiva.

6.3 Acusmata

Acusmata (Costa et al., 2007) é um sintetizador que aplica os conceitos de síntese sonora semântica. Considerando os níveis qualitativos da Figura 6.2, ele atua no nível instantâneo ou granular, associando adjetivos quantificados a micro-porções de som. A linguagem utilizada é adaptativa e totalmente definida pelo usuário. Seu processo de síntese consiste em: determinar um dicionário de adjetivos; estabelecer um conjunto de amostras sonoras de treinamento; e realizar uma associação entre os adjetivos e as amostras. Uma rede neural é treinada para realizar o mapeamento destes adjetivos com o material sonoro a ser sintetizado por um mecanismo gerativo. Uma vez estabelecido o mapeamento, o processo de síntese é realizado pela manipulação destes adjetivos.

O uso de uma semântica adaptativa permite que seja criado um mapeamento individual a cada compositor, possibilitando inclusões artísticas que não são modeláveis por um dicionário inato. Outra possibilidade é o uso de semânticas de controle variadas, até mesmo não humanas, como a

saída de algum sistema computacional ou qualquer tipo de sinal genérico. Esta aplicação permitiria a comunicação entre sistemas computacionais iterativos por um processo de sonificação.

Acusmata utiliza um paradigma de "trajetórias sonoras" em que o som é representado por uma trajetória em um espaço qualitativo. A primeira parte do processo de representação é a definição do dicionário semântico. Este dicionário é um conjunto de figuras semânticas qualitativas, i.e., adjetivos. Dada a escala utilizada em Acusmata, a relação de um adjetivo com o som é instantânea: ele qualifica um ponto definido no tempo do som. A associação do adjetivo com a porção sonora é quantificável, podendo assumir valores dentro de um intervalo definido. Assim, cada adjetivo estabelece um eixo em um espaço qualitativo. Cada ponto instantâneo de uma amostra sonora é, então, relacionado a um ponto nesse espaço qualitativo.

Uma vez que o som é formado pelo seqüenciamento no tempo dessas micro-porções sonoras instantâneas adjetivadas, uma amostra sonora corresponde a um grupo de pontos do espaço qualitativo seqüenciado no tempo, formando uma trajetória. Desta forma, um som pode ser descrito como uma trajetória no espaço de qualidades instantâneas, uma trajetória sonora. A Figura 6.3 faz uma ilustração de uma trajetória sonora tridimensional. Na esquerda, são apresentados três eixos qualitativos (que possuem uma figura semântica associada) com um valor evoluindo no tempo. Na direita é representado o espaço qualitativo tridimensional. A trajetória é definida pela evolução dos valores nos eixos. Esta trajetória é uma analogia ao som.

6.3.1 Procedimento de Síntese

Vamos considerar a coleção de sons $S = \{s_1, s_2, s_3, \ldots, s_N\}$, com N amostras, e um conjunto de palavras-chave $W = \{w_1, w_2, w_3, \ldots, w_M\}$, com M palavras pré-definidas. Para cada som em S, define-se uma Trajetória Semântica Sonora (TSS) T_k, com $k = 1 \ldots N$, como segue na Equação 6.9: Cada par $(t_j, G_j(n_j))$, para $j = 1 \ldots P$, é fornecido pelo usuário e são denominados "momentos". t_j é uma janela de tempo pré-definida e $G_j(n_j)$ uma marca de valores.

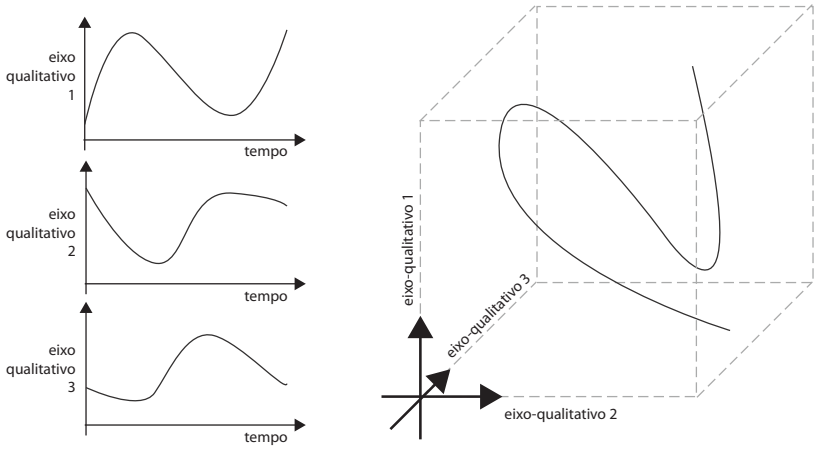


Fig. 6.3: Trajetória no espaço qualitativo. Parametrização dos eixos qualitativos à esquerda. Trajetória em um espaço tridimensional à direita.

$$T_k = (t_1, G_1(n_1)), (t_2, G_2(n_2)), \ldots, (t_P, G_P(n_P)) \tag{6.9}$$

A trajetória expressa a avaliação de como uma janela de tempo (t_j) de um som é relacionada com a palavra-chave de um dicionário W. Assim, do ponto de vista do usuário, o processo de concepção da linguagem consiste em três passos:

- **Passo 1:** Comece com um conjunto de amostras, defina um dicionário de palavras-chave, crie uma TSS para cada som. As TSSs vão ser o modelo para que a base de dados do sistema associe valores das palavras-chave com os fragmentos de som, definindo um espaço com K trajetórias semânticas $T = T_1, T_2, \ldots, T_K$.

- **Passo 2:** Alimente a base de dados do sistema com $T = T_1, T_2, \ldots, T_K$. A trajetória correspondente a cada arquivo de som anotado é automaticamente armazenada na base de dados.

- **Passo 3:** Use o sistema para criar um novo momento sonoro $(t_j, G_j(n_j))$. Este processo de síntese de um novo fragmento de som, assim como sua interpolação em uma nova trajetória é

tratada pelo sistema.

Este modelo de descrição por TSSs é independente da escala descritiva. Podem ser aplicadas trajetórias semânticas associadas a outras trajetórias semânticas que se associam aos fragmentos de áudio. Na aplicação do Acusmata, o mapeamento é realizado com fragmentos em escala granular.

6.3.2 Arquitetura

A arquitetura básica do Acusmata é a de um sintetizador semântico que atua com mapeamento, conforme ilustrado na Figura 6.4. Ela possui um mecanismo de síntese que gera material sonoro a partir de um conjunto paramétrico. O tradutor realiza a conversão de uma trajetória semântica numa trajetória no espaço paramétrico do sintetizador interno, processo denominado tradução.

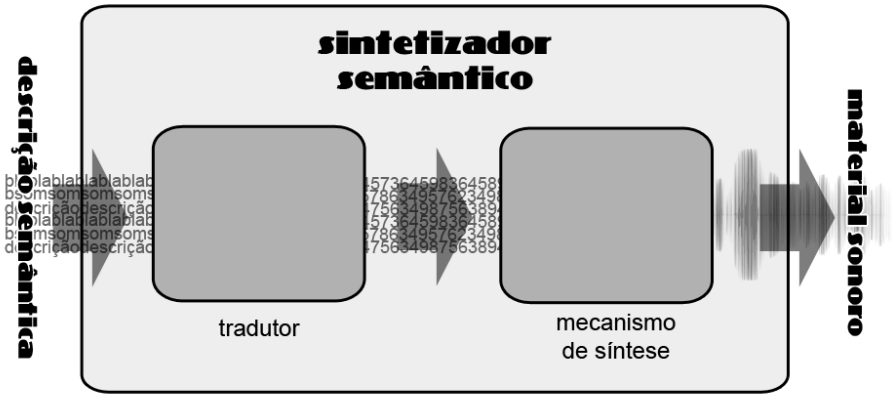


Fig. 6.4: Arquitetura do Acusmata.

O processo de tradução é instantâneo, realizado pelo tradutor. A Figura 6.5 faz uma ilustração do processo. Em um dado instante, o ponto qualitativo, definido pelo vetor de valores $[a\ b\ c\ d\ e]$, é introduzido ao tradutor, no caso uma rede neural, que o converte no vetor de parâmetros com valores $[\alpha\ \beta\ \gamma]$.

O mecanismo de síntese é responsável pela geração do material sonoro. A sua arquitetura define o espaço paramétrico. Existem duas classes de mecanismos de síntese: de uma e de duas vias. De

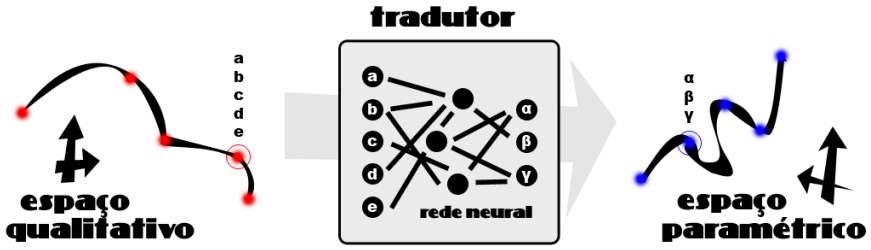


Fig. 6.5: Processo de tradução. Um ponto instantâneo do espaço qualitativo é mapeado em um ponto do espaço paramétrico.

uma via são aqueles cuja transição entre o espaço paramétrico e a amostra sonora é rápida, mas o inverso é altamente custoso. Os mecanismos de duas vias são aqueles que permitem a transição contrária, i.e., síntese e análise. Dentre os métodos de uma via, é incluída a maioria dos sintetizadores tradicionais, como síntese FM e granular. Nos métodos de duas vias, entram os codecs de áudio e métodos espectrais, como o SMS.

A classe do método de síntese implica em estratégias diferentes de treinamento. O mapeamento do tradutor é entre o espaço qualitativo e o espaço paramétrico. Porém, para o usuário, o mapeamento se realiza entre o espaço qualitativo e a amostra sonora. A transição entre o espaço paramétrico e a amostra deve ser transparente ao usuário. No processo de treinamento, a associação é entre valores adjetivos e amostras sonoras, que deve ser convertido para a associação entre valores adjetivos e parâmetros de síntese. Com o uso de mecanismos de duas vias, este processo é direto, podendo o usuário utilizar qualquer amostra que deseje. Contudo, no caso dos mecanismos de uma via, o procedimento é estipular parâmetros de síntese, gerar o material sonoro e, em seguida, proceder com a associação com os adjetivos.

Outra consideração importante é relativa à qualidade das associações. Verificar se um dado adjetivo não exerce influência sobre um dado parâmetro pode permitir a redução dimensional do mapeamento. É o caso de associar dois valores adjetivos distintos a sons com um mesmo valor de um dado parâmetro. É necessário ainda verificar a existência de inconsistências, na forma de contradições. É o caso de associar dois valores adjetivos distintos a um mesmo conjunto de valores de parâmetro.

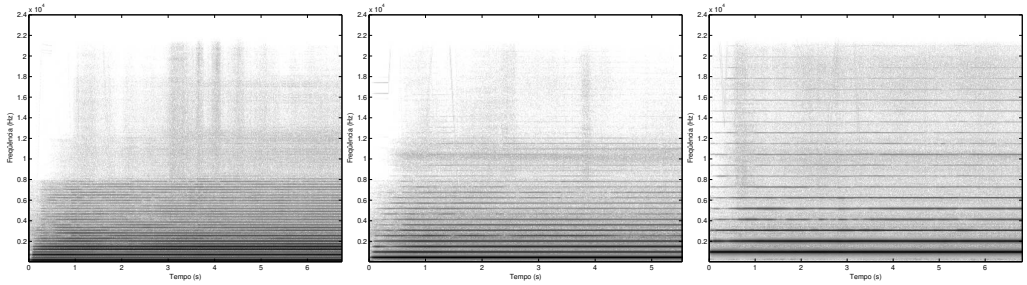
6.3.3 Experimentos

Foram planejados dois experimentos para a implementação:

- **Experimento 1:** Ressíntese de uma amostra de clarineta a partir de uma descrição instantânea objetiva;
- **Experimento 2:** Síntese de um conjunto semântico a partir de descrições de um usuário e verificação da capacidade de síntese;

Experimento 1

O primeiro experimento objetivou a ressíntese de sons de clarineta. A clarineta produz um som fortemente harmônico composto por parciais ímpares e uma componente ruidosa que se estende por todo o espectro, associada ao sopro do intérprete. A Figura 6.6 apresenta o espectrograma de amostras de clarineta.

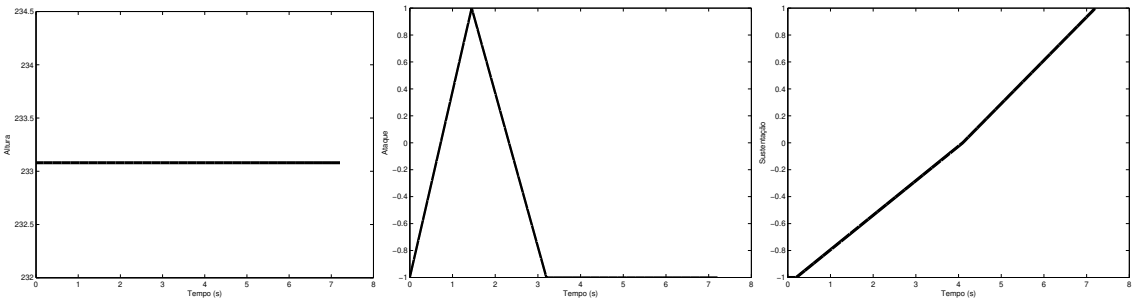


(a) **(b)** **(c)**

Fig. 6.6: Espectrograma de amostras de clarineta: (a) Dó4, (b) Dó5 e (c) Dó6.

O procedimento consiste em associar valores semânticos, que descrevam tanto a posição quanto a altura da nota emitida, a cada porção instantânea das amostras de clarineta. O dicionário semântico é composto por três figuras: altura, ataque, sustentação. A altura é definida diretamente pela freqüência

fundamental da amostra apresentada. O ataque e a sustentação são especificados subjetivamente. O ataque é associado à posição inicial da amostra e sustentação à porção estacionária do sinal. A Figura 6.7 apresenta um exemplo de uma trajetória semântica, seguindo o dicionário proposto, relativo à amostra da Figura 6.6.



(a) **(b)** **(c)**

Fig. 6.7: Exemplo de uma trajetória semântica, parametrizada no tempo, aplicando o dicionário: (a) altura, (b) ataque e (c) sustentação.

No experimento, a representação granular utilizada foi a do espectro de energia do sinal, limitada à metade do espectro (redundante, segundo propriedade da simetria), com normalização de freqüência fundamental. Os grânulos são obtidos diretamente do sinal original através de um processo de janelamento. Utiliza-se janela de Hanning. Este sinal é modulado, modificando a freqüência fundamental do sinal a uma freqüência padrão. O grânulo obtido é submetido ao algoritmo FFT para obtenção da transformada de Fourier de curta duração. Do espectro obtido, calcula-se a energia (módulo do valor complexo). São desprezados os valores de espectro acima do valor de Nyquist (i.e., a metade superior do espectro) por serem redundantes, uma vez que o sinal original é real.

Na composição do grupo de treinamento, é amostrado um valor da trajetória semântica para cada grão extraído da amostra. O grupo de treinamento é então formado por grânulos na representação espectral normalizada e amostras da trajetória semântica.

Uma vez treinada a rede neural, o material sonoro pode ser sintetizado a partir de novas trajetórias semânticas. O processo de geração consiste em amostrar a trajetória semântica na freqüência de

produção de grânulos. A partir de cada ponto semântico, é obtido o espectro normalizado do grânulo de saída. Deste espectro, o sinal é reconstruído. Para a reconstrução a partir da representação, foram utilizadas duas técnicas: a modulação de amplitude de um banco de osciladores nas freqüências das componentes de Fourier; a síntese de grânulos com fase linearmente crescente em freqüência e interligados por um processo de *overlap-and-add*.

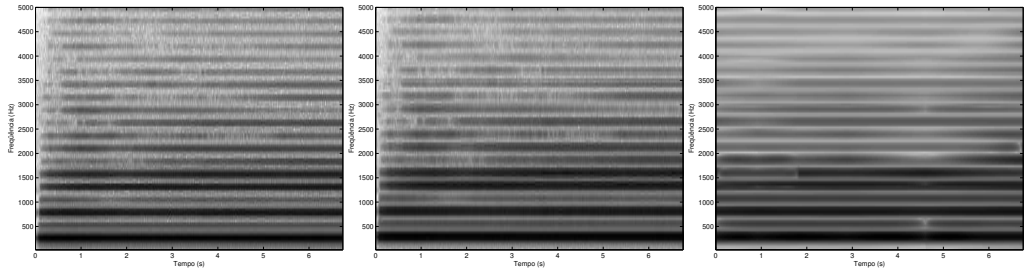
A representação utilizada não permite uma reconstrução perfeita do sinal e, de fato, a supressão da fase implicou uma distorção perceptível. Por isso, o objetivo foi obter sons que se aproximassem do som ressintetizado a partir da representação do som original.

Neste experimento, foram utilizadas as 12 amostras com as alturas apresentadas na Tabela 6.5. São as mesmas amostras utilizadas no quarto experimento do capítulo anterior (Seção 5.3.8). As amostras contêm, em média, 4 segundos, o que levou a um conjunto de treinamento de 6776 grânulos. As trajetórias semânticas, relativas às figuras de ataque e sustentação, foram definidas subjetivamente pelo autor.

Tab. 6.5: Notas de clarineta utilizadas para o treinamento no Experimento 1

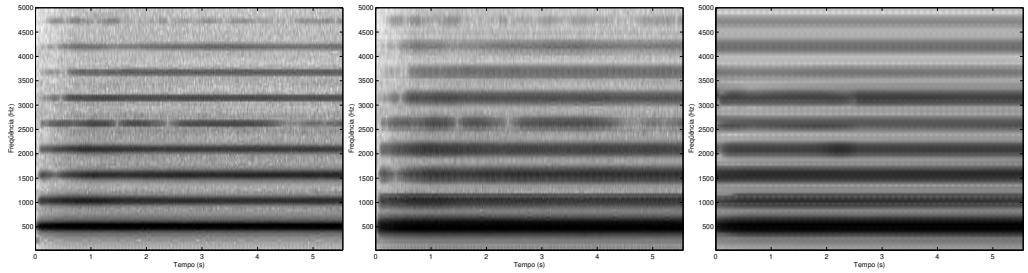
Nota	**Freqüência Fundamental**
Ré3 (D3)	146,83
Ré#3 (D#3)	155,56
Mi3 (E3)	164,81
Lá#3 (A#3)	233,08
Dó4 (C4)	261,63
Ré4 (D4)	293,66
Lá#4 (A#4)	466,16
Dó5 (C5)	523,25
Ré5 (D5)	587,33
Lá#5 (A#5)	932,33
Dó6 (C6)	1046,50
Ré6 (D6)	1174,66

Em uma primeira execução, foram utilizadas como referência para a síntese as trajetórias sonoras de treinamento. A intenção foi comparar o material sonoro obtido com a amostra sonora utilizada como referência. Figuras 6.8, 6.9 e 6.10 ilustram este processo.



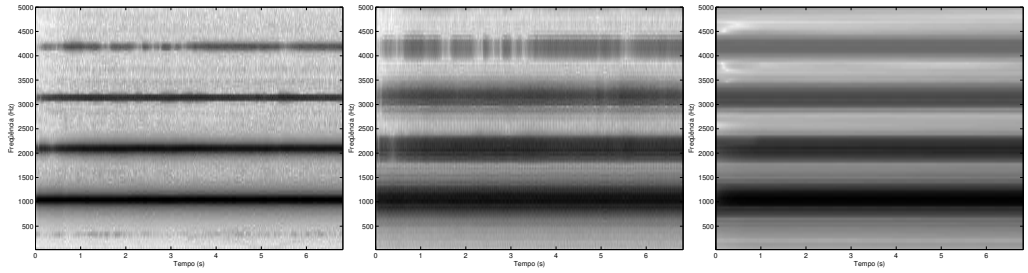
(a) **(b)** **(c)**

Fig. 6.8: Ressíntese do som da clarineta. (a) sinal original em Dó4 (C4) (freqüência fundamental = 261,63 Hz), (b) após supressão de fase e (c) resultado da ressíntese pelo SGBP.



(a) **(b)** **(c)**

Fig. 6.9: Ressíntese do som da clarineta. (a) sinal original em Dó5 (freqüência fundamental = 523,25 Hz), (b) após supressão de fase e (c) resultado da ressíntese pelo SGBP.



(a) **(b)** **(c)**

Fig. 6.10: Ressíntese do som da clarineta. (a) sinal original em Dó6 (freqüência fundamental = 1046,50 Hz), (b) após supressão de fase e (c) resultado da ressíntese pelo SGBP.

Experimento 2

No segundo experimento, objetivou-se a verificação subjetiva da capacidade de síntese baseada em uma linguagem descritiva. Para tal, são percorridas três etapas:

1. **Formação da linguagem:** desenvolvimento do dicionário semântico. O procedimento padrão é apresentar ao sujeito do teste amostras geradas pelo mecanismo de síntese e dar-lhe liberdade para instituir suas próprias figuras semânticas.

2. **Aprendizado computacional:** primeiramente, ao sujeito são apresentadas amostras sonoras geradas a partir de um conjunto paramétrico conhecido. O sujeito é responsável por qualificar estas amostras, seguindo o dicionário desenvolvido na etapa anterior. O conjunto de associações é apresentado para o mecanismo de aprendizado computacional.

3. **Teste subjetivo:** são sintetizadas amostras sonoras a partir de trajetórias semânticas conhecidas (mas ignoradas pelo sujeito). O sujeito é incumbido de qualificá-las, de maneira semelhante ao processo da etapa anterior. A trajetória de estímulo é, então, comparada com a trajetória obtida.

Para este experimento, foi concebido um mecanismo de síntese composto por: 1 sintetizador aditivo com 3 parâmetros; 1 banco de filtros aplicado sobre ruído branco com 3 parâmetros; 1 parâmetro de mistura; e 1 filtro global com 4 parâmetros. A Tabela 6.6 lista os seus parâmetros e a Figura 6.11 apresenta o esquemático do mecanismo. Estes parâmetros são instantâneos, podendo gerar material sonoro a partir de uma trajetória paramétrica. O som gerado foi em formato PCM de 16 bits com taxa de amostragem de 44 kHz.

Como ouvinte, foi escolhido um profissional com treinamento musical, graduado em música popular pela UNICAMP. Foram apresentadas amostras sonoras geradas com conjuntos paramétricos aleatórios. O ouvinte definiu um dicionário composto por cinco figuras semânticas: brilho, calor, ruidesa, inarmonicidade e definição.

Na etapa de treinamento, foi avaliado subjetivamente um conjunto de 70 amostras sonoras. Cada amostra sonora, de duração de dois segundos, foi sintetizada a partir de um conjunto de valores

Tab. 6.6: Parâmetros do mecanismo de síntese do Experimento 2.

Sintetizador	Parâmetro	Descrição
Global	Harmoniciade	Relação de energia entre a síntese aditiva e subtrativa.
	Equalizador 1	Nível da banda de equalização (<703Hz).
	Equalizador 2	Nível da banda de equalização (703Hz,1880Hz).
	Equalizador 3	Nível da banda de equalização (1880Hz,3000Hz).
	Equalizador 4	Nível da banda de equalização (>3000Hz).
Subtrativa	Filtro 1	Freqüência central do primeiro filtro.
	Filtro 2	Freqüência central do segundo filtro.
	Filtro 3	Freqüência central do terceiro filtro.
Aditivo	Inclinação Espectral	Decaimento de energia por oitava.
	Paridade Harmônica	Relação de energia harmônicos pares/ímpares.
	Distorção Espectral	Deslocamento randômico de cada harmônico.

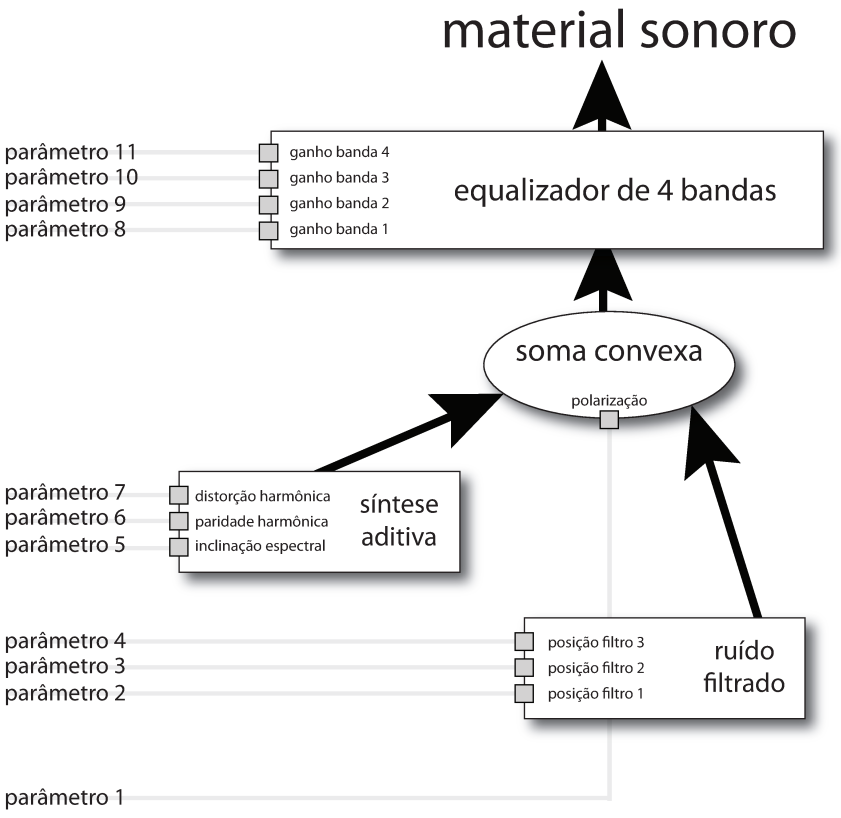


Fig. 6.11: Esquemático do mecanismo de síntese utilizado no Experimento 2.

paramétricos individuais e constantes durante a amostra.

A Figura 6.12 apresenta a correlação entre as figuras semânticas e os parâmetros do sintetizador. Foi utilizado o método de correlação por rank de Spearman. Correlação por rank permite avaliar se duas variáveis possuem alguma relação de crescimento monotônica e a direção desta relação. Assim, é possível realizar uma análise da influência de uma dada qualidade semântica sobre um parâmetro de síntese. Contudo, esta análise é limitada, uma vez que são consideradas apenas interações individuais das figuras semânticas com os parâmetros de síntese.

Na análise da Figura 6.12, alguns fatos interessantes se destacam. O brilho obteve forte correlação direta com a banda de alta freqüência do equalizador e com a distorção harmônica. Estes dois parâmetros estão associados ao aumento de componentes de alta freqüência. Calor esta associado com a inclinação espectral com a equalização de freqüências médias e com a atenuação de altas e baixas freqüências. Ruidesa e inarmonicidade têm uma correlação com os parâmetros muito parecida. Uma explicação é que, devido à realização de soma convexa entre o sinal ruidoso e o harmônico, a existência de sinal harmônico implica em pouco sinal ruidoso, e vice-versa. A definição é fortemente ligada ao grau de harmonização e à inclinação harmônica.

Outro mecanismo de análise interessante é o histograma dos valores semânticos. No histograma, divide-se a faixa dinâmica em 10 subgrupos igualmente espaçados. O gráfico representa a freqüência (normalizada em porcentagem) de ocorrência de dados pertinentes a cada grupo. A Figura 6.13 apresenta o histograma para cada figura semântica. O histograma fornece informações sobre a amostragem do espaço qualitativo. Um mapeamento homogêneo corresponderia a um histograma uniforme. No caso, por exemplo, brilho obteve poucos valores nas faixas 1, 2 e 10. Por este motivo, houve um povoamento pobre desta região do espaço no treinamento, o que pode levar a uma imprecisão por falta de amostragem. Em geral, o mapeamento foi bem amostrado nas raias médias.

Neste experimento, foi utilizada uma rede MLP, com treinamento supervisionado de segunda ordem, via método do gradiente conjugado e escalonado (Moller, 1993; Pearlmutter, 1994). Após o treinamento, a rede neural faz o mapeamento de um conjunto de valores semânticos para um conjunto de valores paramétricos.

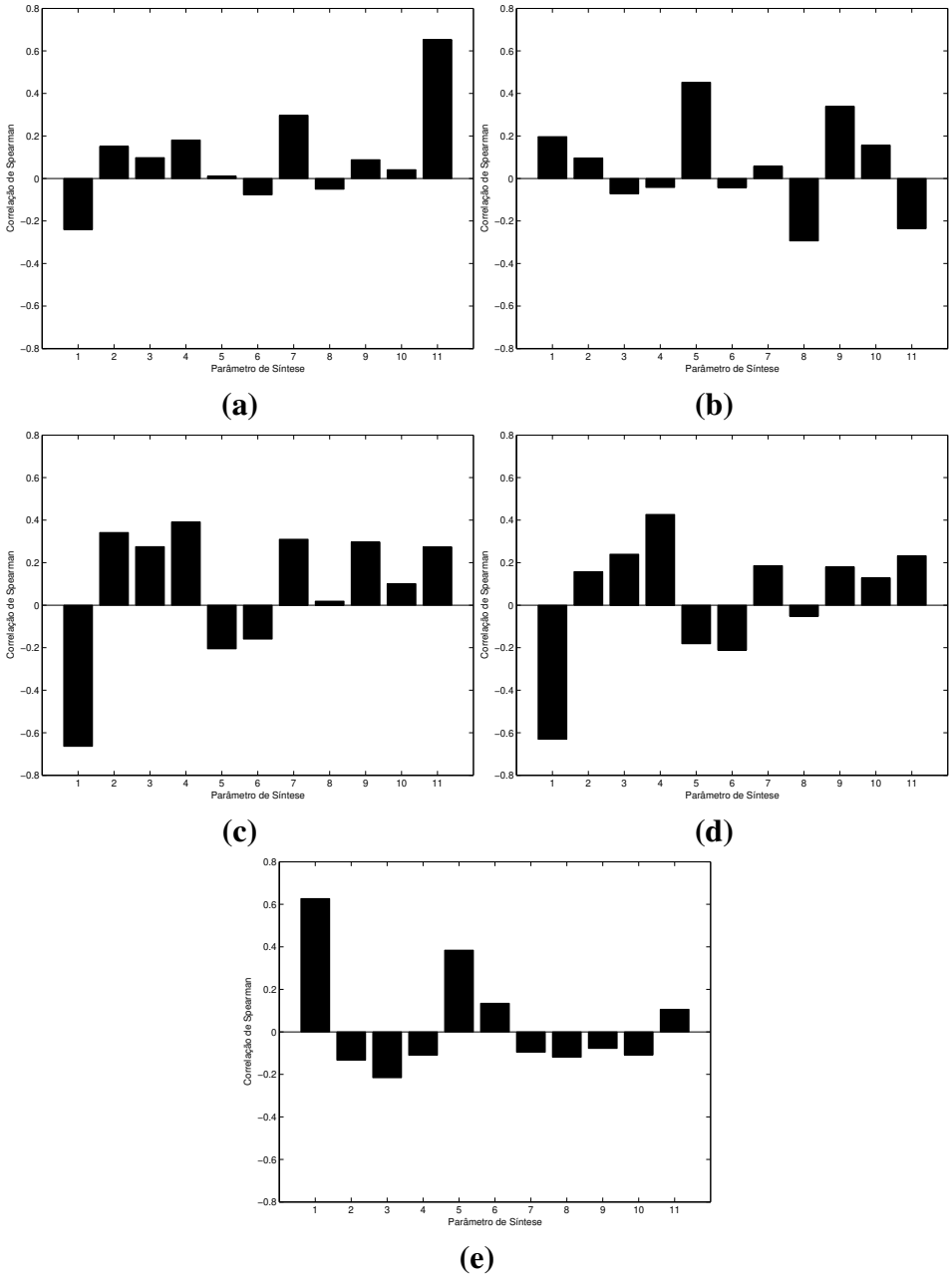


Fig. 6.12: Correlação entre os valores semânticos e os parâmetros do mecanismo de síntese. (a) brilho, (b) calor, (c) ruideza, (d) inarmonicidade e (e) definição.

O primeiro experimento teste foi o de gerar material sonoro a partir de trajetórias sonoras com alguma característica conhecida. As Figuras 6.14 até 6.18 representam o material sonoro obtido a partir de trajetórias que mantêm constante em 5 o valor de todas as figuras semânticas, variando

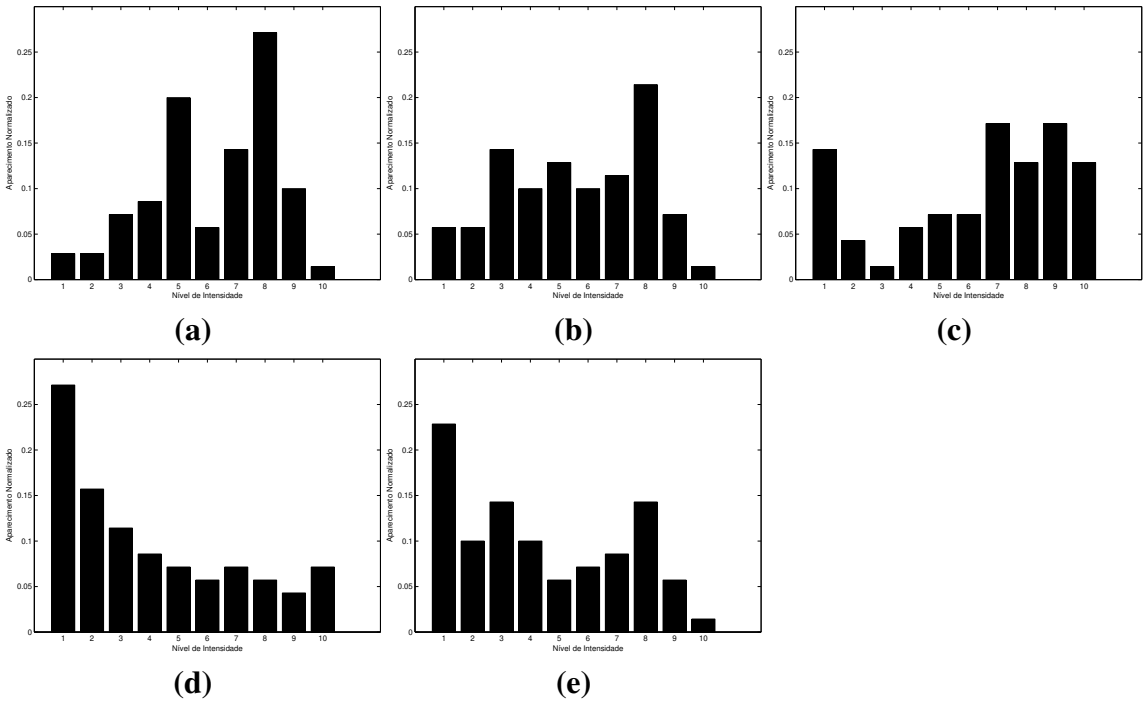


Fig. 6.13: Histograma dos valores semânticos. (a) brilho, (b) calor, (c) ruidesa, (d) inarmonicidade e (e) definição.

apenas uma variável entre 0 e 10, de forma monotonicamente crescente. Ao usuário, foi pedido que fosse identificada qual qualidade semântica foi variada. As amostras de brilho e calor foram identificadas corretamente. As amostras de inarmonicidade e ruidesa foram identificadas pelo sujeito com dúvida, mas limitando-se corretamente às duas amostras. A amostra de definição não foi identificada.

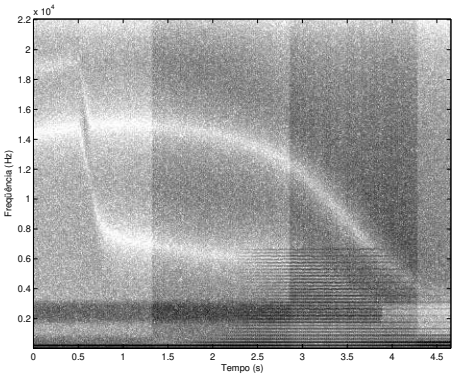


Fig. 6.14: Amostra sonora com variação crescente de brilho (de 0 a 10).

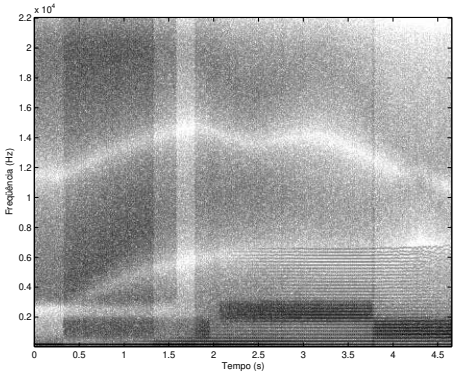


Fig. 6.15: Amostra sonora com variação crescente de calor (de 0 a 10).

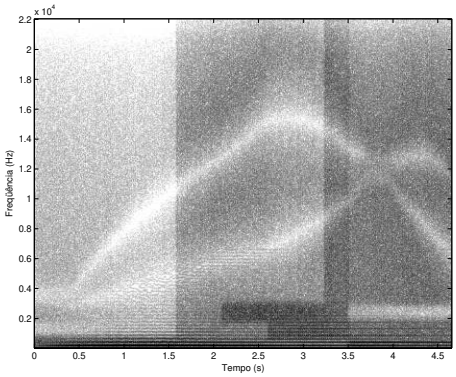


Fig. 6.16: Amostra sonora com variação crescente de ruidesa (de 0 a 10).

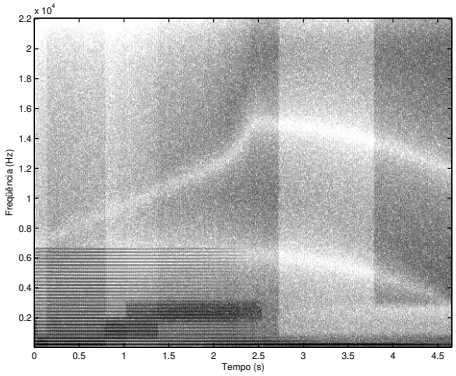


Fig. 6.17: Amostra sonora com variação crescente de inarmonicidade (de 0 a 10).

É possível identificar algum comportamento notável nos espectrogramas, em uma análise por inspeção. Na variação crescente de brilho, Figura 6.14, houve um crescimento progressivo: das componentes harmônicas, da distorção harmônica (visível no leve tremular das linhas harmônicas)

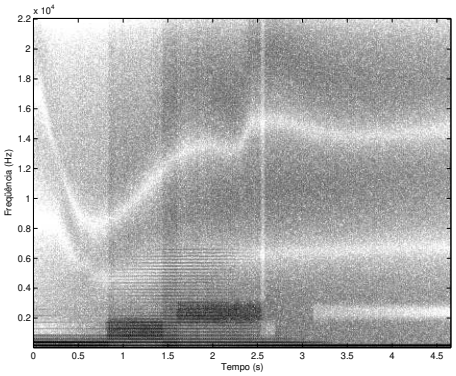


Fig. 6.18: Amostra sonora com variação crescente de definição (de 0 a 10).

e da energia na banda de alta freqüência. Após o quarto segundo, a uma forte diminuição da energia da banda de equalização mais baixa. No som com crescimento do calor, Figura 6.15, há um crescimento das componentes harmônicas e uma concentração das faixas de filtragem do ruído nas freqüências médias. No som de crescimento da ruidesa, Figura 6.16, há um aumento na energia geral do espectro, mas uma manutenção da parte harmônica. No aumento da inarmonicidade, Figura 6.17, há o desaparecimento das faixas harmônicas e a manutenção da energia geral da faixa ruidosa. Na definição, Figura 6.18, há um aumento das raias harmônicas até o segundo 2,5. Após isso, não há um comportamento identificável.

Outro fenômeno observável, tanto no espectrograma quanto nos sons, é uma descontinuidade de algumas propriedades. Isso é causado pela reduzida quantidade de amostras do mapeamento e, talvez, por excesso de treinamento da rede neural.

Em um segundo experimento, foram sintetizadas amostras a partir de um grupo randômico de valores semânticos. A um grupo de ouvintes: o objeto do treinamento, ouvinte X; ouvinte P, com treinamento musical, professor do Instituto de Artes da UNICAMP; e ouvinte E, sem treinamento musical, engenheiro de telecomunicações. A cada amostra gerada, foi pedido a cada usuário que atribuísse subjetivamente valores descritivos, seguindo o dicionário criado no instante inicial deste experimento, pelo ouvinte X. Em seguida, foi extraída a correlação por rank de Spearman, Figura 6.19.

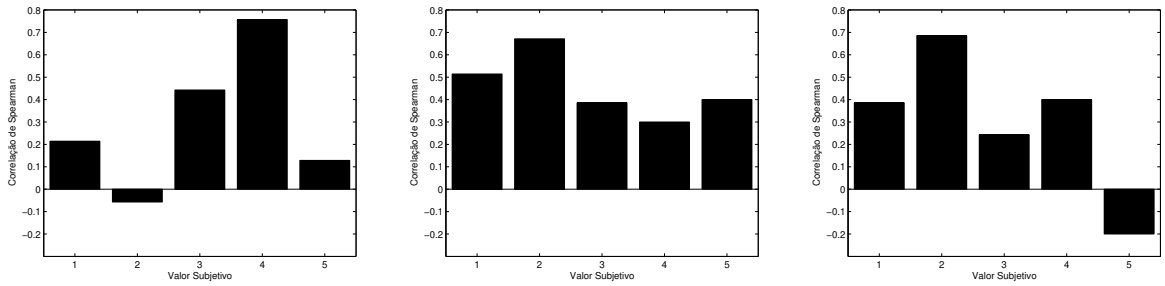


Fig. 6.19: Correlação entre os valores subjetivos e os utilizados para a síntese. 1-5 respectivamente: brilho, calor, ruidesa, inarmonicidade, definição. (a) ouvinte X, (b) ouvinte P, (c) ouvinte E.

É interessante observar que o ouvinte P obteve uma distribuição de correlação mais homogênea e com valores altos para todas as amostras. O ouvinte E obteve uma correlação levemente inferior, mas com um valor baixo para a variável semântica de definição. Isso é explicável pela falta de familiaridade com o termo, por ser um sujeito sem treinamento artístico. Outro fator relevante é a baixa correlação em alguns termos pelo próprio ouvinte X, foco do treinamento.

Discussão

Acusmata se enquadra como um exemplo de aplicação da representação por analogia semântica, por permitir a descrição do material sonoro a partir de um dicionário estabelecido pelo usuário. Os experimentos indicam a capacidade de controle sobre o material sonoro, seja em um mapeamento objetivo, no caso da ressíntese da clarineta, ou em um mapeamento puramente subjetivo, como no caso do experimento 2. Neste sentido, esta implementação de Acusmata foi bem sucedida, ilustrando de forma satisfatória o paradigma apresentado.

Os experimentos de fato podem ser aprimorados. A representação do mecanismo de síntese utilizada na ressíntese da clarineta não permite a reconstrução do sinal. Acusmata foi bem sucedido em atuar sobre a representação, mas devido à limitação desta, foi incapaz de gerar sons efetivamente semelhantes ao som da clarineta.

Ainda, no segundo experimento, são possíveis algumas melhorias no processo. Foram utilizadas poucas amostras, seja no treinamento, seja na validação. Isso não permitiu a aplicação efetiva de um método de validação cruzada no treinamento, o que limitou o controle sobre o excesso de treinamento. O número reduzido de amostras impediu que se criasse um mapeamento menos acidentado nas transições entre pontos notáveis. O fato de a avaliação subjetiva ser uma forma de medida imprecisa e ruidosa torna mais grave a escassez de dados. Ainda assim, os resultados mostraram o controle das variáveis semânticas sobre o material sonoro.

Assim, o modelo de representação proposto não se limita ao modelo de síntese ou representação semântica utilizado no Acusmata. Quanto ao modelo de síntese, Acusmata atua em nível instantâneo. O modelo de representação pode ser aplicado em múltiplas escalas perceptivas, como mostrado na Figura 6.2.

Quanto à representação semântica, o processamento semântico utilizado no Acusmata é puramente numérico. As figuras semânticas atuam como rótulos para as variáveis numéricas. Uma alternativa seria um processamento simbólico, com a aplicação de analisadores léxicos e sintáticos. Esta forma permitiria um controle puramente textual da síntese. Contudo, seriam necessárias considerações próprias e mais complexas com relação à construção da linguagem e ao mecanismo de aprendizado.

Capítulo 7

Conclusão

Neste trabalho, o processo de síntese sonora foi relacionado à busca do usuário por um som capaz de suprir expectativas estéticas aqui denominadas de "objetivo sonoro". Partimos da hipótese de que a eficiência do sistema não está somente na possibilidade de produzir qualquer som imaginável, mas na capacidade de permitir ao usuário, de forma consciente, explorar e interagir com o espaço sonoro e encontrar o som que satisfaça o seu imaginário. O som sintetizado por este processo é aqui denominado de "material sonoro".

Ao discutir a percepção sonora, Capítulo 2, considerou-se a noção de sonoridade, a qual foi definida por Schaeffer (1966) como o atributo qualitativo e distintivo de um som, o que lhe garante identidade sob a ótica perceptiva do ouvinte. Diferenciou-se sonoridade do conceito tradicional de timbre, que, entre outras definições, foi descrito por Risset & Mathews (1969) como "o atributo que permite ao ouvinte identificar o instrumento que produz o tom". Na visão *schaefferiana*, timbre se refere às propriedades físicas do fenômeno sonoro, denominadas "propriedades materiais ou concretas", enquanto que ele postula a experiência sonora relacionada à sonoridade através do fenômeno da "Escuta".

Como visto na Seção 2.1.1, dada a natureza da evolução dos sintetizadores, a exploração histórica dos métodos de síntese se estabeleceu através da adequação do usuário ao espaço de parâmetros. Desta forma, coube ao usuário fazer a conexão entre o controle e o conceito subjetivo de

sonoridade. O foco do estudo aqui reportado foi criar condições para avançar na direção de novas representações para a síntese que facilitassem a realização dessa conexão. Assim, seria possível, com o auxílio do computador, explorar o espaço de sonoridades com maior facilidade, proporcionando um controle mais intuitivo e eficiente. Esta foi a questão de controle, problema de fundo desta dissertação.

Um modelo de síntese que só permite a realimentação a partir da tentativa e erro do usuário emprega uma aplicação restrita de técnicas computacionais de busca e controle. Frente a este cenário, lançou-se nesta dissertação a idéia de analogia, que, a partir de Gentner (1989), é vista como o caminho para estabelecer representações computacionais do "objetivo sonoro". Isso fomentou uma modelagem computacional de uma métrica e a aplicação de retroalimentação no sistema, o que levou à aplicação de técnicas computacionais de busca e controle diferentes das historicamente empregadas em síntese.

Nesta dissertação, discutiu-se representação para definir melhor o problema. Considerou-se "representação cognitiva" a expressão do pensamento do usuário, definida como representação "primária" por Vickery (1986); e "representação computacional" como a interação da primária com meios físicos, também denominada "secundária" por este autor. Desta forma, estruturou-se um modelo de controle composto por processos cognitivos, os quais estão vinculados ao usuário (i.e. à tradução do objetivo sonoro nos parâmetros de síntese e à avaliação da diferença entre o material e o objetivo sonoro), e processos computacionais (i.e. a síntese paramétrica).

Este modelo, que inclui a inferência humana na definição do objetivo, foi apresentado como uma otimização cuja função-objetivo possuía uma definição nebulosa, uma métrica não-formal, alta dimensionalidade de dados e alto grau de incerteza. Para tal, não se presumiu a existência de um ótimo e recorreu-se a meta-heurísticas para a elaboração do espaço de busca. Foram as técnicas computacionais bio-inspiradas, apresentadas no Capítulo 4, as ferramentas adequadas de busca, mapeamento e controle utilizadas na solução do problema.

Esta dissertação propôs pesquisar duas analogias descritivas: "Analogia acústica", a associação

da expectativa do usuário com amostras sonoras diversas; e "Analogia semântica", a associação de qualidades do "objetivo sonoro" relacionadas ao dicionário de adjetivos.

No uso da analogia acústica, Capítulo 5, o modelo de controle proposto realizou a minimização da distância entre o material sonoro e as amostras de referência, o que pode ser descrito como uma otimização multi-objetivo. Foi desenvolvido um protótipo denominado "Síntese Gerativa Baseada em População", ou SGBP, Seção 5.3, que implementou uma representação computacional baseada em analogia acústica no controle de um sintetizador granular na geração de texturas sonoras. A SGBP utilizou o conceito de "quantização vetorial" para criar um "dicionário de grânulos sonoros" representativos de um grupo de amostras de referência, denominadas "referência de sonoridade". Para esta tarefa são aplicados: mapas auto-organizáveis, computação evolutiva e sistemas imunológicos artificiais. A SGBP gera material sonoro pela justaposição dos elementos do dicionário, escolhidos a partir de uma segunda referência denominada "controle de dinâmica". Nos experimentos, foi possível: (1) avaliar a preservação de características da sonoridade do conjunto de amostras de referência no material sonoro e (2) estudar o comportamento de síntese diante de amostras com nuances notáveis.

Para (1), foram utilizadas: uma amostra de voz, uma linha melódica de guitarra e um som harmônico sintético. Na análise subjetiva dos resultados, tendo como referência o valor médio de semelhança de uma amostra frente a amostras ordinárias, o material sonoro apresentou um acréscimo de 147% na semelhança subjetiva com o controle de dinâmica e um acréscimo de 85% na semelhança subjetiva com a referência de sonoridade, como mostrado na Tabela 5.2. A inspeção de sonogramas permitiu ainda a visualização de padrões das referências no material sintetizado, como as figuras da Seção 5.3.8. Para (2), foram utilizadas amostras senoidais e diversas amostras executadas por uma clarineta. A principal questão levantada na análise foi a necessidade (ou não) de se considerar características inatas do som (no caso, a altura) na concepção da representação.

No uso da analogia semântica, Capítulo 6, o modelo proposto vislumbrou a criação de uma linguagem descritiva para sonoridade. O dicionário desta linguagem deve possuir adjetivos que

são relacionados às qualidades de um determinado objetivo sonoro. Mostra-se como este modelo pode ser empregado de duas formas: via uma linguagem inata, relacionando, por expressões matemáticas, atributos físico a qualidades subjetivas, como no trabalho de Hoffman & Cook (2006); ou via linguagem adaptativa, específica de um indivíduo, gerada por um mapeamento baseado em amostragem, caso do protótipo desenvolvido nesta dissertação, chamado de "Acusmata", Seção 6.3. O Acusmata foi utilizado para controlar mecanismos de síntese com parâmetros instantâneos, realizando o mapeamento do espaço paramétrico para um espaço alternativo cujos eixos dimensionais foram rotulados pelos adjetivos do dicionário. Este espaço alternativo foi denominado de "espaço qualitativo". O mapeamento foi feito com o uso de redes neurais artificiais. O processo de síntese consistiu em estipular uma trajetória semântica, convertida numa trajetória de parâmetros de controle do sintetizador. Foram realizados dois experimentos: (1) treinamento de dicionário de adjetivos com características bem definidas e (2) treinamento de um dicionário pessoal.

Para (1), foram utilizadas amostras de clarineta e um sintetizador baseado em DFT. O mapeamento qualitativo considerou nos eixos a posição temporal (sob os adjetivos relacionados ao ataque e sustentação) e a freqüência fundamental instantânea. A análise por inspeção mostrou a capacidade de generalização para freqüências não treinadas e a semelhança com o sinal original em casos de mimetização. Para (2), foi utilizado um sintetizador modular de 11 parâmetros, relacionados com síntese aditiva, síntese subtrativa e filtragem. O ouvinte estabeleceu um dicionário de 5 adjetivos (brilho, calor, ruidesa, inarmonicidade e definição). O mapeamento foi realizado a partir de uma análise subjetiva de amostras com parâmetros aleatórios. Na avaliação do mapeamento, amostras sintetizadas a partir de um ponto conhecido (mas cego ao ouvinte) no espaço qualitativo foram alvo de uma nova avaliação subjetiva. A análise do estímulo qualitativo e do valor medido se deu por "correlação por rank", e mostraram uma coerência na maioria dos adjetivos testados.

É importante observar que o SGBP e o Acusmata não exaurem as possibilidades dos paradigmas apresentados. De fato, as duas implementações foram propostas com o intuito de ilustração dos mesmos. Em ambos os casos, verificou-se a preservação de características da referência no material sintetizado (pela análise subjetiva e por inspeção). Atesta-se assim a capacidade da representação

computacional de se relacionar com o objetivo sonoro. É possível aprofundar os testes psico-acústicos com mais amostras e ouvintes. Todavia, os resultados obtidos já foram suficientes para apontar indícios de um bom funcionamento do sistema.

Com relação a extensões futuras, são possíveis diferentes modelos de controle em ambos os paradigmas. A cada novo modelo haverá uma gama de heurísticas aplicáveis. É possível pensar como exemplo, na analogia acústica, a aplicação de algoritmos genéticos sobre um sintetizador paramétrico, se for possível definir uma métrica entre a representação computacional do objetivo e o material sonoro. Em outro exemplo, na analogia semântica, a utilização de inferência em redes semânticas pode permitir gerar um conjunto paramétrico de síntese a partir de uma descrição de alto nível do objetivo.

É necessário ainda considerar a importância do próprio mecanismo de síntese sobre a qualidade do controle. Um sintetizador baseado em DFT com FFT pode ter facilmente milhares de coordenadas no espaço de parâmetros, o que torna absurdamente custoso o processo de busca. Contudo, sintetizadores com poucos parâmetros, como síntese FM, podem ter um comportamento demasiadamente não-linear, o que dificultaria, por exemplo, o mapeamento em um espaço alternativo no caso do Acusmata. Uma extensão interessante seria a aplicação de Wavelets, em especial o modelo de permutação de coeficientes Luvizotto (2007).

Em última instância, projeta-se a implementação dos modelos em software comercial (ou de proposta profissional) de produção musical. Assim, seria possível uma real avaliação do potencial dos paradigmas, mas implicaria em preocupações como o desenvolvimento de interface e desempenho em tempo real.

Em relação ao impacto do trabalho, é importante destacar a característica interdisciplinar da pesquisa. Interdisciplinaridade é um processo decorrente do afrouxamento das fronteiras entre várias áreas do conhecimento. Este fenômeno ocorre em vários níveis, de forma que ciências básicas, associadas por intersecções e normas de interação, organizem uma super-disciplina. Esta, por

sua vez, em seu processo de solidificação serve como base ao mesmo processo que lhe formou, organizando áreas de um nível ainda mais elevado e complexo. A temática central deste trabalho engloba questões relacionadas a três áreas essencialmente multidisciplinares, que obtiveram um estado de maturidade após um forte impulso de pesquisa e desenvolvimento na segunda metade do Século XX: "Controle de Síntese Sonora (Computação Musical) por Analogia Acústica e Semântica (Ciências Cognitivas) aplicando Computação Bio-Inspirada (Computação Natural)". A computação musical estabelece o domínio do problema a ser tratado: síntese sonora. As ciências cognitivas moldam a essência da resolução do problema: paradigma de representação. Por fim, a computação natural permite uma implementação prática da resolução do problema.

A contribuição direta, e mais evidente, deste trabalho está voltada para a computação musical com a proposta de novos paradigmas de síntese. O trabalho se mostra em sintonia com alguns trabalhos de centros importantes, como o grupo de Xavier Rodet na Universidade de Paris (no trabalho de Schwarz (2005)) e o grupo de Perry Cook na Univesidade de Princeton (no trabalho de Hoffman & Cook (2006)). Na análise da bibliografia, verifica-se que a maioria dos trabalhos que buscam inovações no paradigma de controle do processo de síntese data de menos de uma década. Como exemplo, pode-se citar alguns trabalhos como Lee & Wessel (1992), Wanderley et al. (1998), Garnett & Goudeseune (1999), Hunt & Kirk (2000), Verfaille & Arfib (2001), Joseph & Lodha (2002), Walker & Cothran (2003), Terasawa et al. (2005) e Johnson & Gounaropoulos (2006). Atesta-se, assim, a contemporaneidade do trabalho e o posicionamento em uma linha de vanguarda em sua área.

Envolvidas com engenharia de controle, ciências cognitivas e computação bio-inspirada estão as contribuições menos evidentes do trabalho. Quando um sistema envolve decisões que passam pelo viés da percepção do usuário, há a necessidade de expandir a noção de representação do sistema de forma a incluir variáveis que sejam suscetíveis a um espectro que envolve desde variações ambientais a, possivelmente, noções estéticas. De forma geral, é importante que o desenvolvimento de tais sistemas de engenharia englobe a natureza humana dos processos envolvidos. Desta forma, o trabalho envolve a aplicação de conceitos epistemológicos na modelagem de sistemas de forma a

reduzi-los a problemas de controle. Esta pesquisa pode, assim, contribuir na proposta de soluções junto a outros problemas que envolvam a participação humana em seu processo.

Por fim, na visão do autor, é o próprio caráter inerente da percepção humana presente no processo de síntese que incita a discussão central deste trabalho. Noções dependentes da formação intelectual de um indivíduo, tais como as estéticas, são complexas e carentes de métodos de modelagem. Voltaire (1924) exemplifica complexidade inerente a estas em um adágio ao citar que, aos olhos de um sapo, uma figura grotesca ao ser humano pode ser bela: Pergunte a um sapo o que é beleza, o "to kalon"? Ele vai responder a você que é a sua mulher sapo com seus olhos esbugalhados saltando de sua cabeça pequena, uma boca larga e chata, uma barriga amarela e umas costas marrom [1].

[1] adaptação ao português: *"Demandez à un crapaud ce que c'est que la beauté, le grand beau, le to kalon. Il vous répondra que c'est sa crapaude avec deux gros yeux ronds sortant de sa petite tête, une gueule large et plate, un ventre jaune, un dos brun."*

Referências Bibliográficas

Allison, R. I. & Uhl, K. P. (1964). Influence of beer brand identification on taste perception. *Journal of Marketing Research*, 1(3), 36–39.

Amatrian, X., Arum, P., & Ramírez, M. (2002). Clam, yet another library for audio and music processing ? In *Proceedings of the 2002 Conference on Object Oriented Programming, Systems and Applications (OOPSLA 2002)* Seattle, USA: ACM.

Angeline, P. J., Saunders, G. M., & Pollack, J. B. (1994). An evolutionary algorithm that constructs recurrent neural networks. *IEEE Transactions on Neural Networks*, 5(1), 54–65.

Barnard, K. & Forsyth, D. (2001). Exploiting image semantics for picture libraries. In *The First ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 469).

Bencina, R. (2005). The metasurface - applying natural neighbor interpolation to two-to-many. In *Proceedings of the Conference on New Interfaces for Musical Expression (NIME05)*.

Bernardet, U., Blanchard, M., & Verschure, P. (2002). Iqr: a distributed system for real-time real-world neuronal simulation. *Neurocomputing*, 44-46, 1043–1048.

Boese, K. D. & Kahng, A. B. (1993). Simulated annealing of neural networks: the 'cooling' strategy reconsidered. In *Proceedings of IEEE International Symposium on Circuits and Systems* (pp. 2572–2575).

Bowcott, P. (1989). Cellular automation as a means of high level compositional control of granular synthesis. In *Proceedings of the 1989 International Computer Music Conference* (pp. 55–57).

Bowcott, P. (1990). High level control of granular synthesis using the concepts of inheritance and social interaction. In *Proceedings of the 1990 International Computer Music Conference* (pp. 50–52).

Brown, J. C., Houix, O., & Mcadams, S. (2001). Feature dependence in the automatic identification of musical woodwind instruments. *Journal of Acoustic Society of America*, 109(3), 1064–1072.

Burnet, F. M. (1959). *The Clonal Selection Theory of Acquired Immunity*. Cambridge University Press.

Caetano, M., Costa, C. R., Manzolli, J., & Von Zuben, F. J. (2005a). Self-organizing topological timbral design methodology using a kohonen neural network. In *Proceedings of the 10th Brazilian Symposium on Computer Music (SBCM)* (pp. 94–105). Belo Horizonte, Brazil.

Caetano, M., Manzolli, J., & Von Zuben, F. J. (2005b). Application of an artificial immune system in a compositional timbre design technique. In *Proceedings of ICARIS 2005* Alberta, Canada.

Caetano, M., Manzolli, J., & Von Zuben, F. J. (2005c). Interactive control of evolution applied to sound synthesis. In *Proceedings of the 18th International Florida Artificial Intelligence Research Society (FLAIRS)* Clearwater Beach, EUA.

Candler, W. & Norton, R. (1976). *Multilevel Programming*. Technical report, DRC, World Bank, Washington, D.C.

Cano, P. (2007). *Content-Based Audio Search from Fingerprinting to Semantic Audio Retrieval*. PhD thesis, UPF, Barcelona.

Cano, P., Nürnberger, A., Stober, S., & Tzanetakis, G., Eds. (2006). *Learning the Semantics of Audio Signals*, Athens, Greece.

Carrillo, A. A. P. (2006). *Gesture based synthesis of bowed string instruments*. PhD thesis, Universitat Pompeu Fabra, Barcelona.

Chalmers, D. J., French, R., & Hofstadter, D. R. (1991). *High-Level Perception, Representation, and Analogy*. Technical Report 49, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Indiana.

Choi, I., Bargar, R., & Goudeseune, C. (1998). A manifold interface for a high dimensional control space. In *Proceedings of the International Computer Music Conference*.

Chou, W.-C. (1979). Ionisation: The function of timbre in its formal and temporal organization. In S. V. Solkema (Ed.), *The New Worlds of Edgard Varèse: A Symposium*. Brooklyn: Institute for Studies in American Music.

Chowning, J. (1973). The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society*, 21(7), 526–534.

Codd, E. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6), 377–387.

Cook, P. R., Ed. (1999). *Music, Cognition and Computerized Sound: an Introduction to Psychoacoustics.* Cambridge, Massachusetts, USA: MIT Press.

Cook, P. R. & Lakatos, S. (2003). Using dsp-based parametric synthesis models to study human perception. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* Mohonk, NY, USA.

Cosi, P., De Poli, G., & Lauzzana, G. (1994a). Auditory modelling and self-organizing neural networks for timbre classification. *Journal of New Music Research*, 23, 71–98.

Cosi, P., De Poli, G., & Lauzzana, G. (1994b). Timbre classification by nn and auditory modeling. In M. Marinaro & P. G. Morasso (Eds.), *Proceedings of ICANN'94, International*.

Costa, C. R., Furlanete, F., Manzolli, J., & Von Zuben, F. J. (2007). Sound synthesis based on semantic descriptors. In *Proceedings of the 11th Brazilian Symposium on Computer Music (SBCM)* São Paulo, Brazil.

Costa, C. R., Manzolli, J., & Von Zuben, F. J. (2006a). Population-based generative synthesis: A real-time texture synthesizer based on real-world sound streams. In *Proceeding of IX AES Brazil Congress* São Paulo, Brazil.

Costa, C. R., Manzolli, J., & Von Zuben, F. J. (2006b). Using sound streams as a control paradigm for texture synthesis. In *Proceeding of XVI ANPPOM Congress* Brasilia, Brazil.

Damiani, F., Pérez, M. J., & Fornari, J. E. (1995). Reconhecimento de timbres musicais através da rede neural auto-organizável de kohonen. In *XV Congresso da Sociedade Brasileira de Computação, II Simpósio Brasileiro de Computação e Música* (pp. 107–113).

Dasgupta, D. (1999). *Artificial Immune Systems and their Applications*. Springer-Verlag.

de Castro, L. N. (1998). Análise e síntese de estratégias de aprendizado para redes neurais artificiais. Master's thesis, FEEC/UNICAMP.

de Castro, L. N. (2007). Fundamentals of natural computing: an overview. *Physics of Life Reviews*, 4, 1–36.

de Castro, L. N. & Timmis, J. I. (2002). *Artificial Immune Systems: A New Computational Intelligence Approach*. London: Springer-Verlag.

de Castro, L. N. & Von Zuben, F. J. (2001). ainet: An artificial immune network for data analysis. In H. Abbas, R. Sarker, & C. Newton (Eds.), *Data Mining: A Heuristic Approach*. Idea Group Publishing.

de Castro, L. N. & Von Zuben, F. J., Eds. (2005). *Recent Developments in Biologically Inspired Computing*. Idea Group Publishing.

De Poli, G. & Piccialli, A. (1991). Pitch-synchronous granular synthesis. In G. D. Poli, A. Piccialli, & C. Roads (Eds.), *Representations of Musical Signals*. Cambridge, MA: MIT Press.

De Poli, G. & Prandoni, P. (1997). Sonological models for timbre characterization. *Journal of New Music Research*, 26, 170–197.

De Poli, G. & Tonella, P. (1993). Self-organizing neural networks and grey's timbre space. In *Proceedings of the 1993 International Computer Music Conference*.

Deb, K. (1999). Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary Computation*, 7(3), 205–230.

Di Scipio, A. (1994). Micro-time sonic design and timbre formation. *Contemporary Music Review*, 10(2), 135–148.

Emmerson, S. (1986). *The language of Electroacoustic Music*. Macmillan, Basingstoke & London.

Etherington, R. & Punch, B. (1994). Seawave: A system for musical timbre description. *Computer Music Journal*, 18(1), 30–39.

Falkenhainer, B. (1988). *Learning from physical analogies : a study in analogy and the explanation process*. PhD thesis, University of Illinois at Urbana-Champaign.

Feiten, B. & Gunzel, S. (1994). Automatic indexing of a sound database using selforganizing neural nets. *Computer Music Journal*, 18(3), 53–65.

Fels, S. & Hinton, G. E. (1998). Glove-talk ii: A neural network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Transactions on Neural Networks*, 9(1), 205–212.

Fineberg, J. N. (1999). *Sculpting Sound: An introduction to the Spectral Movement - its ideas, techniques and music*. PhD thesis, Columbia University.

Fletcher, H. & Munson, W. A. (1933). Loudness of a complex tone, its definition, measurement and calculation. *The Journal of the Acoustical Society of America*, 5(1), 65.

Fogel, D. B. (1993). Using evolutionary programming to create neural networks that are capable of playing tic-tac-toe. In *International Joint Conference on Neural Networks (IJCNN 93)*.

Fonseca, C. M. & Fleming, P. J. (1993). Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. In *Proceedings of the Fifth International Conference on Genetic Algorithms* (pp. 416–423). San Mateo, CA, USA.

Forbus, K. D., Gentner, D., Markman, A. B., & Ferguson, R. W. (1998). Analogy just looks like high-level perception: why a domain-general approach to analogical mapping is right. *Journal of Experimental and Theoretical Artificial Intelligence (JETAI)*, 10(2), 231–257.

Fornari, J. E., Manzolli, J., & Maia, A. J. (2001). The evolutionary sound synthesis method. In *Proceedings of the ninth ACM international conference on Multimedia*.

Forrest, S. & Hofmeyr, S. A. (1999). Architecture for an artificial immune system. *Evolutionary Computation*, 7(1), 45–68.

Forrest, S., Hofmeyr, S. A., & Somayaji, A. (1996). Computer immunology. *Comunications of the ACM*.

Forrest, S., Hofmeyr, S. A., & Somayaji, A. (1997). Principles of a computer immune system. In *Proceedings of New Security Paradigms Workshop*.

Fujinaga, I. & Vantomme, J. (1994). Genetic algorithms as a method for granular synthesis regulation. In *Proceedings of the 1994 International Computer Music Conference* (pp. 138–141).

Funkhouser, T., Min, P., Kazhdan, M., Chen, J., Halderman, A., Dobkin, D., & Jacobs, D. (2003). A search engine for 3d models. *ACM Transactions on Graphics*, 22(1), 83–105.

Furlanete, F. (2000). Escritura sismográfica. Master's thesis, PUC-SP.

Gabor, D. (1946). Acoustical quanta and the theory of hearing. *Nature*, 159, 591–594.

Gadd, A. & Fels, S. (2002). Metamuse: Metaphors for expressive instruments. In *Proceedings of 2nd International Conference on New Interfaces for Musical Expression (NIME02)* (pp. 143–148).

Garnett, G. E. & Goudeseune, C. (1999). Performance factors in control of high-dimensional spaces. In *Proceedings of the International Computer Music Conference* (pp. 268–271). San Francisco, CA, USA.

Gentner, D. (1989). The mechanisms of analogical learning. In S. . Vosniadou & A. . Ortony (Eds.), *Similarity and Analogical Reasoning* (pp. 199–241). New York : Cambridge University Press.

Ghias, A., Logan, J., Chamberlin, D., & Smith, B. C. (1995). Query by humming – musical information retrieval in an audio database. In *Proceedings of ACM Multimedia*.

Goldberg, D. E. (1987). Simple genetic algorithms and the minimal, deceptive problem. In D. L. (Ed.), *Genetic Algorithms and Simulated Annealing* (pp. 74–88). San Mateo, CA: Morgan Kaufmann.

Gomes, L. C. T. & Von Zuben, F. J. (2003). Multiple criteria optimization based on unsupervised learning and fuzzy inference applied to the vehicle routing problem. *Journal of Intelligent & Fuzzy Systems*, 13(2-4), 143–154.

Goutte, C. (1997). Note on free luches and cross-validation. *Neural Computation*, 9, 1211–1215.

Grey, J. M. (1975). *An Exploration of Musical Timbre*. PhD thesis, Stanford University.

Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harshberger, M. L. (1997). Methodological issues in timbre research. In I. Deliege & J. Sloboda (Eds.), *Perception and Cognition of Music*. East Essex, UK.

Hajela, P. & Lin, C.-Y. (1992). Genetic search strategies in multicriterion optimal design. *Structural Optimization*, 4, 99–107.

Halford, G. S. (1992). Analogical reasoning and conceptual complexity in cognitive development. *Human Development*, 35(4), 193–217.

Hamman, M. (1991). Mapping complex systems using granular synthesis. In *Proceedings of the 1991 International Computer Music Conference* (pp. 475–478).

Haykin, S. (2001). *Redes neurais: princípio e prática*. Bookman, 2 edition. trad. Paulo Martins Engel.

Heidegger, M. (1962). *Being and Time*. London: SCM Press. trans. by John MacQuarrie and Edward Robinson.

Heisenberg, W. (1927). Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. *Zeitschrift für Physik*, 43, 172–198.

Helmholtz, H. L. F. (1877). *On the Sensations of Tone: As a Physiological Basis for the Theory of Music*. Dover Publications, INC., New York.

Hoffman, M. & Cook, P. R. (2006). Feature-based synthesis for sonification and psychoacoustic research. In *Proceedings of the 12th International Conference on Auditory Display* London, UK.

Hoffman, M. & Cook, P. R. (2007). Real-time feature-based synthesis for live musical performance. In *Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME '07)*.

Hofstadter, D. R. (2001). Analogy as the core of cognition. In D. Gentner, K. Holyoak, & B. Kokinov (Eds.), *The Analogical Mind: Perspectives from Cognitive Science*. Cambridge, MA: The MIT Press/Bradford Book.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

Holyoak, K. . J. & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13(3), 295–355.

Holyoak, K. . J. & Thagard, P. (1995). *Mental Leaps . Analogy in Creative Thought*. Cambridge, MA : MIT Press.

Horn, J., Nafpliotis, N., & Goldberg, D. E. (1994). A niched pareto genetic algorithm for multiobjective optimization. In *Proceedings of the First IEEE Conference on Evolutionary Computation*, volume 1 (pp. 82–87). Piscataway, NJ, USA: IEEE Service Center.

Horner, A. (1995). Envelope matching with genetic algorithms. *Journal of New Music Research*, 24(4), 318–341.

Horner, A., Beauchamp, J., & Haken, L. (1993). Machine tongues xvi: Genetic algorithms and their application to fm matching synthesis. *Computer Music Journal*, 17(3), 17–29.

Horner, A., Cheung, N., & Beauchamp, J. (1995). Genetic algorithm optimization of additive synthesis envelope breakpoints and group synthesis parameters. In *Proceedings of the International Computer Music Conference* (pp. 215–222).

Houtsma, A. J. M. (1997). Pitch and timbre: Definition, meaning and use. *Journal of New Music Research*, 26, 104–115.

Hu, X. & Eberhart, R. (2002). Multiobjective optimization using dynamic neighborhood particle swarm optimization. In *Proceedings of IEE Congress on Evolutionary Computation (CEC2002),*.

Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. In *Proceedings of the Institute of Radio Engineers*, volume 40 (9) (pp. 1098–1101).

Hunt, A. & Kirk, R. (2000). Mapping strategies for musical performance. In M. Wanderley & M. Battier (Eds.), *Trends in Gestural Control of Music* (pp. 231–258). Institut de Recherche et Coordination Acoustique Musique - Centre Pompidou,.

Huron, D. (2006). *Sweet anticipation: the psychology of expectation*. Cambridge - MIT Press.

Husserl, E. (1969). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie*. Springer.

Hwang, J.-N., Lay, S.-R., Maechler, M., Martin, R. D., & Schimert, J. (1994). Regression modeling in back-propagation and projection pursuit learning. *IEEE Transactions on Neural Networks*, 5(3), 342–353.

Jayant, N. S. & Noll, P. (1984). *Digital Coding of Waveforms*. Prentice Hall.

Jehan, T. (1997). Musical signal parameter estimation. Master's thesis, IFSIC, Universitité de Rennes, France and Center for New Music and Audio Technologies (CNMAT), University of California, Berkley, USA.

Jerne, N. K. (1974). Towards a network theory of the immune system. *Ann. Immunol. (Inst. Pasteur)*, (pp. 373–389).

Johnson, C. & Gounaropoulos, A. (2006). Timbre interfaces using adjectives and adverbs. In *Proceedings of the 2006 Internacional Conference on New Interfaces for Musical Expression (NIME06)* Paris, France.

Jones, D. L. & Parks, T. (1988). Generation and combination of grains for music synthesis. *Computer Music Journal*, 12(2), 27–34.

Joseph, A. & Lodha, S. K. (2002). Musart: Musical audio transfer function real-time toolkit. In *Proceedings of the International Conference on Auditory Display* (pp. Kyoto, Japan).

Kandel, E. R. & Schwartz, J. H. (1985). *Principles of Neural Science*. Elsevier.

Karplus, K. & Strong, A. (1983). Digital synthesis of plucked string and drum timbres. *Computer Music Journal*, 7(2), 43–55.

Keane, M. T., Ledgeway, T., & Duff, S. (1994). Constraints on analogical mapping : a comparison of three models. *Cognitive Science*, 18, 387–438.

Keller, D. & Truax, B. (1998). Ecologically-based granular synthesis. In *ICMC 1998* Ann Arbor, Michigan.

Kobayashi, R. (2003). Sound clustering synthesis using spectral data. In *Proceedings of the International Computer Music Conference (ICMC '03)*.

Kochenberger, G. A. & Glover, F. (2003). *Handbook of metaheuristics*. Kluwer Academic Publishers.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., & Gunter, T. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, 7(3), 302–307.

Kohonen, T. (1984a). Phonotopic maps: Insightful representation of phonological features for speech representation. In *Proceedings IEEE 7th Internacional Conference on Pattern Recognition* Montreal, Canadá.

Kohonen, T. (1984b). *Self-Organization And Associative Memory*. Springer-Verlag.

Kohonen, T. (2000). *Self-Organizing Maps*. Springer-Verlag.

Kurigawa, M., Yahiro, H., & Kashiwagi, S. (1978). Seven attributes in tone quality evaluation. *Journal of Acoustic Society (Japan)*, 34, 493–500.

Lakatos, S., Cook, P. R., & Scavone, G. (2000). Selective attention to the parameters of a physically informed sonic model. *Acoustics Research Letters Online*.

Lee, M. & Wessel, D. L. (1992). Connectionist models for real-time control of synthesis and compositional algorithms. In *Proceedings of the International Computer Music Conference* (pp. 277–280).

Lippe, C. (1993). A musical application of real-time granular sampling using the ircam signal processing workstation. In *Proceedings of the 1993 International Computer Music Conference* (pp. 190–193).

Lippmann, R. P. (1987). An intoduction to computing with neural nets. *IEEE ASSP Magazine*, 4(2), 4– 22.

Loureiro, M. A., de Paula, H. B., & Yehia, H. C. (2004). Timbre classification of a single musical instrument. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)* Barcelona, Espanha.

Luvizotto, A. L. (2007). Modelos de representação de sinais musicais via transformada wavelets. Master's thesis, UNICAMP.

Manjunath, B., Salembier, P., & Sikora, T., Eds. (2002). *Introduction to MPEG-7: Multimedia Content Description Interface*. Wiley & Sons.

Manzolli, J. (1996). *Auto-Organização*, volume 18 of *Coleção CLE*, chapter Auto-Organização: Um Paradigma Composicional, (pp. 417–435). Centro de Epistimologia e Lógica. Drebrun, M. and Gonzales, M. E. Q. and Pessoa Jr., O. (Eds).

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman.

Martínez, J. M. (2004). *MPEG-7 Overview (Version 10)*. Jtc1/sc29/wg11, ISO/IEC, Palma de Mallorca, Spain.

Mathews, M. V. (1963). The digital computer as a musical instrument. *Science*, 142(11), 553–557.

Mathews, M. V., Miller, J. E., Moore, F. R., Pierce, J. R., & Risset, J. C. (1969). *The Technology of Computer Music*. Cambridge, Mass.: MIT Press.

McCulloch, W. & Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.

Mermelstein, P. (1976). Distance measures for speech recognition, psychological and instrumental. In C. H. Chen (Ed.), *Pattern Recognition and Artificial Intelligence* (pp. 374–388). New York: Academic.

Mertens, P. H. (1975). *Die Schumannschen Klangfarbengeretze und ihre Bedeutung für die Übertragung von Sprache und Muzik*. Frankfurt, Germany: Verlag Erwin Bochinski.

Meyers, O. C. (2002). *The Life and Times of Max V. Mathews*. Technical report, McGill University.

Michalewicz, Zbigniew, F. D. B. (2004). *How to Solve It: Modern Heuristics*.

Mierswa, I. & Morik, K. (2005). Automatic feature extraction for classifying audio data. *Machine Learning Journal*, 58, 127–149.

Miranda, E. R. (1995). Granular synthesis of sounds by means of a cellular automaton. *Leonardo*, 28(4), 297–300.

Miranda, E. R. & Biles, J. A., Eds. (2007). *Evolutionary Computer Music*. Springer.

Miranda, E. R. & Maia, A. J. (2005). Granular synthesis of sounds through markov chains with fuzzy control. In *Proceedings of the International Computer Music Conference 2005* Barcelona, Spain.

Moffitt, J. (2001). Ogg vorbis-open, free audio-set your media free. *Linux J.*, 2001(81es), 9.

Moller, A. R. (1983). *Auditory Physiology*. Academic Press.

Moller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4), 525–533.

Moore, B. C. J. (2003). *An Introduction to the Pscychology of Hearing*. Elsevier.

Moore, B. C. J., Glasberg, B. R., & Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, 45(4), 224–240.

Moore, F. R. (1990). *Elements of Computer Music*. Upper Saddle River: Prentice Hall.

Morrison, C. T. & Dietrich, E. (1995). Structure-mapping vs. high-level perception: The mistaken fight over the explanation of analogy. In *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 678–681).

Nagashima, Y. (1992). Real-time control system for 'pseudo granulation'. In *Proceedings of the 1992 International Computer Music Conference* (pp. 404–405).

Noll, M. (1969). Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate. In *Proceedings of the Symposium on Computer Processing Communications* (pp. 779–797).

Nossal, G. J. V. (1993). The molecular and cellular basis of affinity maturation in the antibody response. *Cell*, 68, 1–2.

Oppenheim, A. V. & Schafer, R. W. (1999). *Discrete-Time Signal Processing*. Prentice Hall Signal Processing Series, 2nd edition edition.

Orton, R., Hunt, A., & Kirk, R. (1991). Genetic algorithms and computer-assisted music composition. In *Proceedings of the 1991 International Computer Music Conference* (pp. 416–418).

Ottaviani, L. & Rocchesso, D. (2004). Auditory perception of 3d size: experiments with synthetic resonators. *ACM Transactions on Applied Perceptions*, 1(2), 118–129.

Oy, L., Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. (2000). *SOM Toolbox for Matlab 5*. Technical report, Helsinki University of Technology. Report A57.

Pareto, V. (1906). *Manuale d'Economia Politico*. Milan.

Parsopoulos, K. E. & Vrahatis, M. N. (2002). Particle swarm optimization method in multiobjective problems. In *Proceedings of the ACM 2002 Symposium on Applied Computing (SAC2002)*.

Pearlmutter, B. A. (1994). Fast exact multiplication by the hessian. *Neural Computation*, 6(1), 147–160.

Perelsen, A. S. & Oster, G. F. (1979). Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-nonself discrimination. *Journal of Theoretical Biology*, 81, 645–670.

Pierce, J. R. (1965). Portrait of the computer as a young artist. *Playboy*, June, 120–184.

Pierce, J. R. (1992). *The Science of Musical Sound*. New York: Freeman.

Plattig, K. H., Kobal, G., & Thumfart, W. (1980). The chemical senses of smell and taste in the course of life - changes of smell and taste perception. *Z Gerontol Germany*, 13(2), 149–157.

Polotti, P. & Evangelista, G. (2001). Fractal additive synthesis via harmonic-band wavelets. *Computer Music Journal*, 25(3), 22–37.

Rabiner, L. R. & Schafer, R. W. (1978). *Digital Processing of Speech Signals*. Prentice Hall.

Reed, R. (1993). Pruning algorithms - a survey. *IEEE Transactions on Neural Networks*, 4(5), 740–747.

Risset, J.-C. (1960). *Computer Study of Trumpet Tones*. Technical report, Bell Telephone Laboratories. File MM-66-122-2.

Risset, J.-C. & Mathews, M. V. (1969). Analysis of musical instrument tones. *Physics Today*, 22(2), 26.

Risset, J.-C. & Wessel, D. L. (1982). Exploration of timbre by analysis and synthesis. In D. Deutsch (Ed.), *Psychology of Music*. San Diego, California: Academic Press.

Roads, C. (1988). Introduction to granular synthesis. *Computer Music Journal*, 12(2).

Roads, C. (1991). Asynchronous granular synthesis. In *Representations of Musical Signals* (pp. 143–186). Cambridge, MA: MIT Press.

Roads, C. (1996). *Computer Music Tutorial*. Cambridge, MA: MIT Press.

Roads, C. (1997). Sound transformation by convolution. In C. Roads, S. Pope, A. Piccialli, & G. D. Poli (Eds.), *Musical Signal Processing* (pp. 411–438). Lisse: Swets & Zeitlinger.

Roads, C. (2001). *Microsound*. Cambridge, MA: MIT Press.

Rocchesso, D. (2001). Acoustic cues for 3-d shape information. In *Proceedings of the 2001 International Conference on Auditory Display* Espoo, Finland.

Rodet, X. (1984). Time-domain formant wave-function synthesis. *Computer Music Journal*, 8(3), 9–14.

Rodet, X. (1997). Musical sound signals analysis/synthesis: Sinusoidal+residual and elementary waveform models. In *Proceedings of the IEEE Time-Frequency and Time-Scale Workshop*.

Rodet, X. & Schwarz, D. (2000). Spectral envelopes and additive+residual analysis-synthesis. In J. Beauchamp (Ed.), *The Sound of Music*. unpublished.

Rodet, X. & Tisserand, P. (2001). *ECRINS: Calcul des descripteurs bas niveaux.* Technical report, IRCAM - Centre Pompidou, Paris, France.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representation by back-propagation. *Nature*, 323, 533–536.

Santarosa, R., Moroni, A., & Manzolli, J. (2006). Layered genetical algorithms evolving into musical accompaniment generation. In *EvoWorkshops 2006* (pp. 722–726).

Sartre, J.-P. (1957). *The Transcendence of the Ego: An Existentialist Theory of Consciousness*. Noonday Press.

Schaeffer, P. (1966). *Traité des Objets Musicaux.* Editions Du Seuil.

Schaffer, J. D. (1985). Multiple objective optimization with vector evaluated genetic algorithm. In *Proceedings of the First International Conference on Genetic Algorithms* (pp. 99–100).

Schubert, E. & Wolfe, J. (2006). Does timbral brightness scale with frequency and spectral centroid? *ACTA Acustica united with Acustica*, 92, 820–825.

Schwarz, D. (2004). *Data-Driven Concatenative Sound Synthesis*. PhD thesis, Ircam.

Schwarz, D. (2005). Current research in concatenative synthesis. In *Proceedings of the International Computer Music Conferece* Barcelona, Spain.

Serra, X. (1997). Current perspectives in the digital synthesis of musical sounds. *Formats*, 1.

Serra, X. & Smith, J. (1990). Spectral modeling synthesis: a sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4), 12–24.

Smalley, D. (1986). Spectro-morphology and structuring processes. In S. Emmerson (Ed.), *The Language of Electroacoustic Music* (pp. 61–93). London: Macmillan.

Smalley, D. (1997). Spectromorphology: explaining sound-shapes. *Organised Sound*, 2(2), 107–126.

Smith, J. R. & Chang, S.-F. (1997). An image and video search engine for the world-wide web. In *Proceedings of IS&T/SPIE Symposium on Electronic Imaging: Science and Technology (EI'97) - Storage and Retrieval for Image and Video Databases V* San Jose, CA.

Smith III, J. O. (1987). *Music Applications of Digital Waveguides*. Technical report, CCRMA.

Smith III, J. O. (1991). Viewpoints on the history of digital synthesis. In I. C. M. Association (Ed.), *Proceedings of the 1991 International Computer Music Conference* San Francisco, CA.

Srinivas, N. & Deb, K. (1995). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, 2(3), 221–248.

Tenney, J. (1965). The physical correlates of timbre. *Gravesaner Bläter*, 26, 106–109.

Terasawa, H., Slaney, M., & Berger, J. (2005). Perceptual distance in timbre space. In *Proceedings of the International Conference on Auditory Display* Limerick, Ireland.

Theremin, L. (1927). Mein ziel. *Berliner Tageblatt*, 1.

Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, 4, 1035–1038.

Truax, B. (1988). Real-time granular synthesis with a digital signal processor. *Computer Music Journal*, 12(2), 14–26.

Truax, B. (1996). Soundscape, acoustic communication and environmental sound composition. *Contemporary Music Review*, 15(1), 47–63.

Truitt, T. D. & Rogers, A. E. (1960). *Basics of Analog Computers*. John F. Rider Publisher Inc.

van Rooji, A. J. F., Jain, L. C., & Johnson, R. P. (1996). *Neural network training using genetic algorithms*. World Scientific.

Verfaille, V. & Arfib, D. (2001). A-dafx: Adaptive digital audio effects. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx-01)*.

Verfaille, V., Boissinot, J., Depalle, P., & Wanderley, M. M. (2006). Ssynth: a real time additive synthesizer with flexible control. In *Proceedings of the International Computer Music Conference (ICMC '06)*.

Vickery, D. C. (1986). Knowledge representation: a brief review. *Journal of Documentation*, 42(3), 145–159.

Voltaire (1924). *The Philosophical Dictionary*. New York: Knopf.

Von Zuben, F. J. (1996). *Modelos Paramétricos e Não-Paramétricos de Redes Neurais Artificiais e Aplicações*. PhD thesis, FEEC/UNICAMP.

Walker, B. N. & Cothran, J. T. (2003). Sonification sandbox: A graphical toolkit for auditory graphs. In *Proceedings of the International Conference on Auditory Display* Boston, MA.

Wanderley, M. M., Schnell, N., & Rovan, J. B. (1998). Escher - modeling and performing composed instruments in real-time. In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics (SMC '98)*.

Wang, J. Z., Li, J., & Wiederhold, G. (2001). Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9), 947–963.

Waschka, R. I. & Ferreira, T. (1988). Rapid event deployment in a midi environment. *Interface*, 17, 211–222.

Weidenaar, R. (1995). *Magic Music from the Telharmonium*. Tha Scarecrow Press.

Wright, M. T. (2004). The scholar, the mechanic and the antikythera mechanism. *Bulletin of the Scientific Instrument Society*, 80, 4–11.

Wun, S., Horner, A., & Ayers, L. (2004). A comparison between local search and genetic algorithm methods for wavetable matching. In *Proceedings of the International Computer Music Conference* (pp. 386–389). Miami, USA.

Xenakis, I. (1971). *Formalized Music*. Indiana University Press.

Yao, X. & Liu, Y. (1997). A new evolutionary system for evolving artificial neural networks. *IEEE Transactions on Neural Networks*, 8(3), 694–713.

Yuen, J. & Horner, A. (1997). Hybrid sampling-wavetable synthesis with genetic algorithms. *Journal of the Audio Engineering Society*, 45(5), 316–330.

Zitzler, E., Deb, K., & Thiele, L. (2000). Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8(2), 173–195.

Zuchini, M. H. (2003). Aplicações de mapas auto-organizáveis em mineração de dados e recuperação de informação. Master's thesis, Faculdade de Engenharia Elétrica e de Computação (FEEC/Unicamp).

Zwicker, E. (1982). *Psychoakustik*. Berlin, Germany: Springer Verlag.

Índice Remissivo de Autores

Apêndice A

Tutorial sobre analogia acústica e semântica

O processo de comunicação, dada sua natureza complexa, tem como uma de suas necessidades básicas a utilização de mecanismos descritivos eficazes. De forma intuitiva, pode-se postular que a comunicação implica na possibilidade de externar uma idéia através de um mecanismo de representação capaz de ser decodificado de forma eficaz pelo receptor. Este modelo já foi estudado de forma ampla e, nesta dissertação, é utilizado apenas como um meio de contextualizar o problema.

Considere o caso em que um indivíduo pretende descrever o objeto da Figura A.1 para um segundo indivíduo. Ele procederá de duas formas: através da comparação com um outro objeto que estimule uma sensação semelhante ao original, como, por exemplo, apresentando o objeto da Figura A.2; ou utilizando uma linguagem descritiva, como, por exemplo, o português na frase "carro azul-marinho".

Neste caso, na primeira descrição não há necessidade de conhecimento por parte do emissor ou do receptor. A comparação é direta. No caso do uso de linguagem, os indivíduos precisam de um processo de aprendizado em que a linguagem é construída a partir de outros estímulos causados por outros objetos.

Definidos os mecanismos descritivos, o passo seguinte é estabelecer representações computacionais baseadas neles.

Fig. A.1: Objeto a ser descrito.

Fig. A.2: Objeto semelhante.

Considere o cenário hipotético em que um indivíduo pretende descrever a cena visual da Figura A.3. Por analogia, é possível descrever a cena de forma visual, com o uso de objetos ou outras cenas que tenham relação com a cena sendo descrita, ou de forma indireta a partir de uma linguagem, referenciada por outros objetos.

Na descrição por analogia visual, a cena é representada por ícones com características presentes na cena a ser descrita. Esse tipo de representação não permite uma descrição perfeita do objeto, a não ser pelo uso de uma imagem perfeitamente igual à original. Porém, é possível caracterizar certos conceitos presentes na imagem com um determinado nível de profundidade. No exemplo da cena da Figura A.3, a Figura A.4 é uma descrição visual em um nível em que se estabelece o conceito de

Fig. A.3: Cena a ser descrita.

imagem de um carro, um grupo de pessoas e um ambiente arborizado. Contudo, não há detalhamento sobre o modelo do carro, as pessoas que compõem o grupo, horário e data da fotografia, etc. O fardo deste tipo de descrição é que, para se aproximar de forma exata de uma cena, pode-se exigir uma grande quantidade de outras figuras dependendo de sua complexidade. Em contra partida, é possível descrever uma mídia a partir de outras de mesma natureza, o que permite um discernimento subjetivo direto por parte do interpretante.

Representação acústica é a aplicação deste paradigma no contexto de representação sonora. Neste modelo, são utilizadas amostras sonoras como meio de descrever outras amostras sonoras. No caso de representação para o processo de síntese sonora, o objetivo sonoro não é conhecido de forma detalhada e precisa, como na imagem do exemplo visual da Figura A.3. Por isso, a limitação descritiva do modelo não é tão representativa para a aplicação com sinais acústicos. A dificuldade que se apresenta para o método é de como escolher as amostras sonoras dentro de uma base de dados sonora, que em geral são extensas e carentes de mecanismos eficientes de exploração.

Descrição por analogia semântica é um paradigma de representação que utiliza uma linguagem como canal intermediário por onde se realiza a analogia. A linguagem é estabelecida a partir de um

Fig. A.4: Descrição por analogia visual: objetos diferentes mas que apresentam os mesmos conceitos da cena original.

conjunto de elementos semânticos cujo significado é obtido por referência a objetos do mundo real. Assim, essa é construída antes da descrição, em um processo de aprendizado.

A linguagem é empregada na descrição das características do objeto, que de forma indireta é relacionada com os objetos utilizados na instituição da linguagem. No caso da cena da Figura A.3, a Figura A.5 apresenta o processo de criação da linguagem. São estabelecidos, a partir de uma coleção de imagens, os conceitos de "carro", "grupo de pessoas", "praça arborizada" e "na frente". A partir disso, é possível descrever a cena como "Grupo de pessoas na frente de um carro na frente de uma praça arborizada". Neste caso, a aplicação sonora é análoga à visual.

O que determina a capacidade de descrever detalhes neste método é a complexidade de linguagem, que tende a ser proporcional ao tamanho da coleção de mídias. Porém, como a referência é indireta, o usuário não precisa indicar os objetos que atuam na analogia, mas apenas utilizar os ícones semânticos que sintetizam os conceitos por eles representados. Existe uma maior inércia no processo de estabelecimento da linguagem do que no procedimento descritivo em si. Mas como uma linguagem pode ser aproveitada para a síntese de várias mídias diferentes, sua criação não precisa ser

Fig. A.5: Descrição por analogia semântica: definição de termos a partir de analogia.

repetida.