

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE ARTES
Mestrado em Música

**INTERAÇÃO TÍMBRICA NA
MÚSICA ELETROACÚSTICA MISTA**

JOSÉ IGNACIO DE CAMPOS JÚNIOR

CAMPINAS – 2005



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE ARTES
Mestrado em Música

**INTERAÇÃO TÍMBRICA NA
MÚSICA ELETROACÚSTICA MISTA**

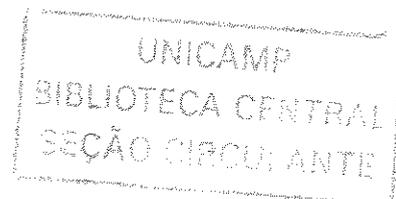
JOSÉ IGNACIO DE CAMPOS JÚNIOR

Este exemplar é a redação final da
Dissertação defendida pela Sr. José
Ignácio de Campos Junior e aprovado
pela Comissão Julgadora em 24/02/2005.


Profa. Dra. Denise Hortência Lopes
Garcia
-orientadora -

Dissertação apresentada ao Curso de
Mestrado em Música do Instituto de Artes
da Unicamp como requisito parcial para a
obtenção do grau de Mestre em Música
sob a orientação da Profa. Dra. Denise
Hortência Lopes Garcia.

CAMPINAS - 2005



UNIDADE	BC
Nº CHAMADA	J/UNICAMP
	C157i
V	EX
TOMBO BC/	64467
PROC.	16-86-05
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	11,00
DATA	24-A-05
Nº CPD	

acompanha 1 CD-ROM

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA CENTRAL DA UNICAMP

Bit. ID 361623

C157i

Campos Júnior, José Ignacio de.
Interação tímbrica na música eletroacústica mista /
José Ignacio de Campos Júnior. – Campinas, SP : [s.n.],
2005.

Orientador: Denise Hortência Lopes Garcia.
Dissertação (mestrado) - Universidade Estadual de
Campinas, Instituto de Artes.

1. Eletroacústica. 2. Música – Acústica e física.
3. Composição (Música). 4. Composição musical por
computador. 5. Psicoacústica. I. Garcia, Denise Hortência
Lopes. II. Universidade Estadual de Campinas. Instituto
de Artes. III. Título.

200511225

A meus pais José Ignácio e Eide

AGRADECIMENTOS

Gostaria de agradecer àquelas pessoas que de alguma forma ajudaram no meu desenvolvimento artístico e pessoal.

Frederico Grassano (in memoriam) – pela eterna amizade e dedicação inabalável à música do nosso tempo.

Alexandre Lunsqui – pela constante lição de consciência musical, pela amizade fraterna e pela ajuda atenta na revisão do texto.

Denise Garcia – agradeço à amizade terna e sincera, à sua orientação primorosa e à paciência e dedicação em me acalmar e tentar me mostrar que as coisas não são tão ruins quanto parecem.

Glória Cunha – por ter acreditado em mim e por ter me estimulado a dar início a esta longa jornada de compositor de música eletroacústica.

Fábio Gorodski – pela amizade e companheirismo desde nosso *début* na música eletroacústica e pelo apoio com material bibliográfico.

Bruno Ruviaro – pela estimulante troca de idéias musicais e pelo senso crítico aguçado.

Leo Aldrovandi – por compartilhar sonhos e esperanças de um meio musical mais rico e justo.

José Augusto Mannis – pela confiança, por compartilhar seu conhecimento de maneira sempre generosa e por ter sido meu guia no desbravamento da música do século XX.

Rodolfo Caesar – pela amizade, companheirismo, por ter me apresentado o Csound em 1995 no Festival de Inverno de Campos do Jordão e por ter sido meu exemplo como compositor de música eletroacústica.

Mikhail Malt – meu co-orientador não oficial que tanto me ajudou em orientação, idéias, materiais e no ensino de OpenMusic em meio ao seu tempo apertadíssimo e por sua amizade.

Andrew Gerzso – pela generosidade em me conceder uma entrevista pessoal e pela ajuda com material bibliográfico.

Fúrio Damiani – pela entrevista concedida e por ter me ajudado a descobrir o meu caminho.

Victor Lazzarini – pela ajuda especializada na programação do Csound.

E agradecimentos especiais a:

Meus pais – pessoas doces e amorosas que se desdoblaram por mim, sempre me incentivaram e acreditaram em mim incondicionalmente.

Patrícia Claro – à minha companheira maravilhosa, que ajudou tanto na realização deste trabalho que seria impossível detalhar, que me dá suporte e me suporta, minha musa inspiradora e incentivadora, meus maiores agradecimentos.

“Não serve a mesma para todos, cada um inventa a sua, a que lhe for própria, há quem leve a vida inteira a ler sem nunca ter conseguido ir mais além da leitura, ficam pegados à página, não percebem que as palavras são apenas pedras postas a atravessar a corrente de um rio, se estão ali é para que possamos chegar à outra margem, a outra margem é que importa... A não ser que esses tais rios não tenham duas margens, mas muitas, que cada pessoa que lê seja, ela, a sua própria margem, e que seja sua, e apenas sua, a margem a que terá que chegar...”

A Caverna, José Saramago

RESUMO

A relação da interação entre meios eletroacústicos e instrumentais na música eletroacústica mista pode ser problemática no que tange à diferença escritural e à divergência tímbrica. O presente projeto pretendeu então tratar da segunda questão (o timbre) e se propôs a encontrar uma solução poética à divergência tímbrica dos meios através de um estudo técnico detalhado, que resulta da concepção da obra eletroacústica mista cuja interação dos meios se circunscreve na delimitação do timbre no material eletrônico a partir dos próprios timbres dos instrumentos acústicos envolvidos na obra.

RÉSUMÉ

L'interaction des moyens électroacoustiques avec les instruments musicaux dans la musique electroacoustique mixte peut devenir problématique à ce que concerne la différence d'écriture e la divergence des timbres. Le présent projet est dirigé aux études sur la deuxième question (le timbre) et s'est mis aux efforts de trouver des solutions poétiques à la divergence des timbres des ces différents moyens, parmi des études techniques qui résultent de la conception de l'oeuvre électroacoustique mixte dont l'interaction entre des moyens se circonscrit au rapprochement des timbres pour le matériau électroacoustique aux timbres des instruments acoustiques sélectionnés pour l'oeuvre.

SUMÁRIO

1	Introdução	18
2	A Psicoacústica na Interação	22
2.1.	Processamento e Interpretação de Informações Sonoras e Musicais - Psicoacústica e Cognição	23
2.2.	A Questão da Interação	25
2.2.1.	Introdução	25
2.2.2.	A Conformação do Material	26
2.2.3.	A Interação Som instrumental – Som Eletrônico na Composição	29
2.3.	A Psicoacústica na Interação	30
2.3.1.	Psicoacústica, uma Introdução	30
2.3.2.	A Sensação de Identidade Tímbrica	31
2.3.3.	Breve Descrição dos Parâmetros Utilizados e de seus Empregos	35
3	A Análise Paramétrica do Som	41
3.1.	Introdução	42
3.2.	Digitalização	43
3.2.1.	Digitalização e suas Etapas	43
3.2.2.	Amostragem e a Física do Som	43
3.2.3.	Sistema Digital	44
3.2.4.	Taxa de Amostragem	45
3.2.5.	Teorema de Nyquist	47
3.2.6.	<i>Aliasing</i>	48
3.2.7.	Quantização	49
3.2.8.	O Sistema Binário	50
3.2.9.	O Erro de Quantização	51
3.2.10.	O Desafio da Fidelidade	51
3.3.	O Padrão FFT	52
3.3.1.	Introdução	52

3.3.2. Análise Paramétrica _____	54
3.3.3. Teorema de Fourier _____	55
3.3.4. <i>Windowing</i> e <i>Spectral Leakage</i> _____	58
3.3.5. Definição Espectral, Resolução Espectral e <i>Window Size</i> _____	62
3.3.5.1. Definição Espectral e sua Relação com a <i>Window Size</i> _____	62
3.3.5.2. Conseqüências da <i>Window Size</i> _____	63
3.3.5.3. Princípio da Incerteza _____	64
3.3.5.4. Resolução Espectral e sua Relação com a <i>Window Size</i> _____	65
3.3.6. FFT _____	66
3.3.6.1. A Relação DFT – FFT _____	66
3.3.6.2. Fator de Sobreposição (<i>Overlap</i>) _____	68
3.3.6.3. <i>Zero-padding</i> _____	71
3.4. Outros Métodos de Análise _____	71
3.4.1. Introdução _____	71
3.4.2. Métodos de Análise _____	73
3.4.2.1. <i>Phase Vocoder</i> _____	73
3.4.2.2. O Modelo Determinístico mais Estocástico _____	75
3.5. Enganos e Mal-Entendidos no AudioSculpt _____	77
3.5.1. Introdução _____	77
3.5.2. Mal-Entendidos _____	78
4 Simulação e Alteração de Parâmetros Sonoros para a Síntese _____	82
4.1. O OpenMusic _____	83
4.2. Criação e Manipulação de Dados para a Síntese Através do OpenMusic _____	85
4.2.1. Os Parâmetros Abordados _____	85
4.2.2. Reformatação e Conversão dos Dados _____	87
4.2.3. Quanto às Frequências _____	90
4.2.3.1. A Transformação de um Espectro Harmônico em um Espectro Inarmônico _____	90
4.2.3.2. A Transformação de um Espectro Inarmônico em um Espectro Harmônico _____	92

4.2.3.3. Compressão e Dilatação Progressiva Exponencial Bipolar Frequencial do Espectro a partir de um Âmbito Central Dado ____	94
4.2.3.4. Redistribuição Progressiva Exponencial Bipolar de Frequências sem Alteração do Âmbito Espectral _____	99
4.2.4. Quanto ao Tempo _____	100
4.2.4.1. Micro-Alterações do Tempo de Entrada de cada Parcial _____	100
4.2.4.2. Relação Variável de Duração entre os Segmentos do Envelope Dinâmico dos Parciais _____	102
4.2.5. Quanto às Amplitudes _____	109
4.2.5.1. Redistribuição da Energia Contida no Espectro _____	109
4.2.5.2. Criação de uma Relação Variável de Amplitude entre o Pico Máximo e a Amplitude de Sustentação do Som _____	114
4.2.6. Quanto às Fases _____	115
4.2.6.1. Em Fase _____	115
4.2.6.2. Fases Opostas entre Parciais Consecutivos _____	116
4.2.6.3. Fases Randômicas com Distribuição Uniforme _____	117
4.2.6.4. Distribuição Gaussiana _____	120
4.2.6.5. Distribuição Beta _____	122
4.2.7. Quanto ao Espaço _____	125
4.2.8. Resultado Final _____	128
5 A Síntese _____	130
5.1. Introdução _____	131
5.2. Técnicas de Síntese _____	133
5.2.1. O Modelo em Síntese _____	133
5.2.2. Classificações de Modelos e Técnicas Envolvidas _____	135
5.2.2.1. Tratamentos de Registro _____	135
5.2.2.2. Modelos Físicos _____	135
5.2.2.3. Algoritmos Abstratos _____	138
5.2.2.4. Modelos Espectrais _____	143
5.3. Idéias e Implementações de Síntese em Csound _____	147
5.3.1. Introdução _____	147

5.3.2. Implementações _____	148
5.3.2.1.Limiar de Audibilidade _____	149
5.3.2.2. <i>Jitter</i> _____	152
5.3.2.3. <i>Shimmer</i> _____	156
5.3.2.4.Distribuição Espacial _____	158
5.3.2.5.Senóides + Ruído _____	159
5.3.2.6.Hibridização _____	166
5.3.2.7.Criação do Arquivo de Análise de <i>Phase Vocoder</i> _____	169
5.3.2.8.Integração dos Processos para a Síntese Final _____	170
6 Conclusão _____	174
7 Apêndices _____	179
7.1. Apêndice A – Citações na Língua Original _____	180
7.2. Apêndice B – Relação de Exemplos Sonoros Contidos no CD _____	186
7.3. CD de Exemplos Sonoros	
8 Referências Bibliográficas _____	188

LISTA DE FIGURAS

Capítulo 3

Figura 3.1 - Representação de uma onda senoidal _____	44
Figura 3.2 - Curva representada por 32 pontos _____	45
Figura 3.3 - Curva representada por 2048 pontos _____	45
Figura 3.4 - Expressão do Teorema de Nyquist _____	47
Figura 3.5 - Representação de uma frequência a ser digitalizada _____	48
Figura 3.6 – Pontos representando a sequência de amostras _____	49
Figura 3.7 - Reconstrução de onda a partir dos valores da amostragem _____	49
Figura 3.8 - Forma de onda simples _____	55
Figura 3.9 - Forma de onda complexa _____	55
Figura 3.10 - Expressão da transformada de Fourier _____	56
Figura 3.11 - Expressão da DFT _____	57
Figura 3.12 - Exponencial complexo da DFT _____	57
Figura 3.13 - Senóide antes de ser segmentada _____	59
Figura 3.14 - Exemplo de senóide truncada pela segmentação _____	59
Figura 3.15 - Sinal senoidal antes de passar pelo processo de <i>windowing</i> _____	60
Figura 3.16 - Janela de análise von Hann _____	60
Figura 3.17 - Sinal resultante do processo de <i>windowing</i> com a janela von Hann _____	60
Figura 3.18 - Blackman-Harris _____	61
Figura 3.19 - Hamming _____	61
Figura 3.20 - Retangular _____	61
Figura 3.21 - Vale criado entre dois <i>frames windowed</i> _____	68
Figura 3.22 - Exemplo de sobreposição com fator 8 _____	70
Figura 3.23 – Sonograma com a <i>edit color palette</i> em primeiro plano _____	80
Figura 3.24 – Sonograma indicando espalhamento de frequência... _____	81

Capítulo 4

Figura 4.1 - Exemplo de <i>patch</i> de OpenMusic	84
Figura 4.2 - Lista de dados exportada pelo <i>partial tracking</i> do AudioSculpt	88
Figura 4.3 - <i>Patch list_cleaner</i> usado para reformatação e conversão de dados	89
Figura 4.4 - <i>Patch</i> para transformação de um espectro harmônico...	91
Figura 4.5 - <i>Patch</i> para transformação de um espectro inarmônico...	93
Figura 4.6 - <i>Patch</i> para cálculo de compressão e expansão frequencial	95
Figura 4.7 - <i>Loop</i> usado como separador de listas	96
Figura 4.8 - <i>Patch</i> para cálculo de dilatação de frequências em âmbito fixo	99
Figura 4.9 - <i>Patch</i> que implementa micro-variações randômicas...	101
Figura 4.10 - <i>Patch</i> que previne a geração de números negativos	102
Figura 4.11 - <i>Patch adsr-interno</i> que calcula as proporções...	106
Figura 4.12 - <i>Patch adsr-geral</i> que formata os dados de envelope...	107
Figura 4.13 - Notas correspondentes à frequência fundamental...	110
Figura 4.14 - <i>Patch</i> de redistribuição de energia do espectro	112
Figura 4.15 - Trecho de <i>patch</i> que implementa a geração de fases opostas	116
Figura 4.16 - Trecho de <i>patch</i> que implementa a geração de fases randômicas	118
Figura 4.17a, b e c - Formas de onda resultantes de distintas gerações...	120
Figura 4.18 - Curva padrão da distribuição gaussiana	121
Figura 4.19 - Expressão geral da distribuição gaussiana	121
Figura 4.20 - Expressão adaptada da distribuição gaussiana	121
Figura 4.21 - Forma adaptada da curva gaussiana	122
Figura 4.22 - Expressão para a distribuição beta padrão	123
Figura 4.23 - Expressão para a função beta	123
Figura 4.24 - Curvatura da distribuição beta...	123
Figura 4.25 - Segmento de <i>patch</i> que implementa a distribuição beta	124
Figura 4.26 - <i>Patch</i> criado para a geração de posições dos parciais	126
Figura 4.27 - <i>Patch</i> geral que dá o resultado final de todos os processos	128
Figura 4.28 - Excerto de <i>score</i> produzido pelo <i>patch</i> geral em OpenMusic	129

Capítulo 5

Figura 5.1 – As funções <i>linen</i> e <i>expon</i> são utilizadas...	132
Figura 5.2 – Exemplo de implementação de modelagem física de flauta	137
Figura 5.3 – Identidade trigonométrica que define o comportamento...	138
Figura 5.4 – Sonograma de um som gerado por FM	140
Figura 5.5 – Alteração de amplitude da onda portadora...	141
Figura 5.6 – Representação tridimensional da variação das amplitudes...	142
Figura 5.7 – Onda senoidal com amplitude próxima ao limite máximo...	143
Figura 5.8 – A mesma onda que a anterior distorcida por meio de amplificação	143
Figura 5.9 – Expressão matemática para ATH e a curva gerada por ela	150
Figura 5.10 – Excerto de <i>script</i> de Csound contendo a implementação de ATH	151
Figura 5.11 – Endereçamento de parciais para as três regiões do espectro	153
Figura 5.12 – Lista de valores gerados segundo os parâmetros de <i>jitter</i>	154
Figura 5.13 - Utilização do <i>opcode jitter2</i> com a composição...	155
Figura 5.14 – Utilização do <i>opcode jspline</i> com a composição...	156
Figura 5.15 – Implementação de todo o processo de <i>shimmer</i>	157
Figura 5.16 – Aplicação de distribuição espacial através do <i>opcode pan</i>	159
Figura 5.17 – Implementação de leitura e filtragem recursivas...	164
Figura 5.18 – Exemplo em Csound da integração de vários dos processos...	171
Figura 5.19 – Excerto de <i>score</i> de Csound correspondente à <i>orchestra</i> ...	173

Tabela

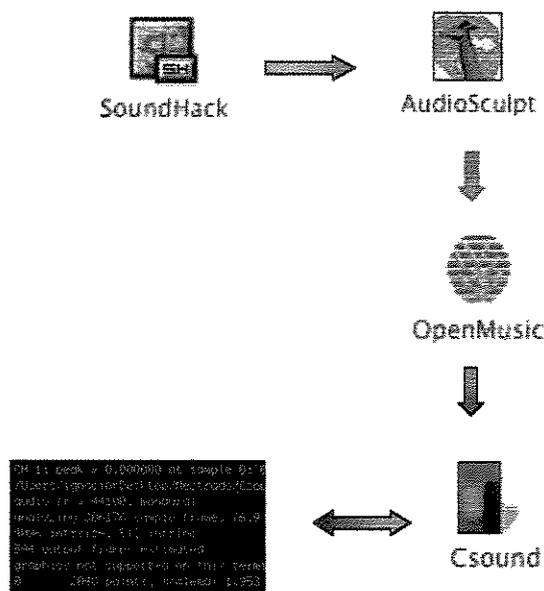
Tabela 4.1 – Dois excertos de uma análise de FFT de uma nota dó grave de piano	110
--	-----

1. Introdução

Quando se iniciou este trabalho, tinha-se em mente a investigação de uma escritura que fosse única para a música eletroacústica mista, que se pudesse, durante as primeiras fases de elaboração da composição, aplicar técnicas que levassem à atenuação da dicotomia domínio eletrônico – domínio instrumental que se apresenta em muitas das obras do gênero. Contudo, logo no início da pesquisa, ficou claro que esse era um problema grande demais e que poderia ser desdobrado em dois campos distintos, porém imbricados: a diferença escritural, ligada à estruturação e à organização dos eventos sonoros no espaço-tempo, e a divergência tímbrica, ligada à conformação particular do material próprio de cada domínio.

O presente projeto pretendeu então tratar da segunda questão, sob o recorte do tempo diferido, e se propôs a encontrar uma solução poética à divergência tímbrica dos meios através de um estudo técnico detalhado, que resulta da concepção da obra eletroacústica mista cuja interação dos meios se circunscreve na delimitação do timbre no material eletrônico a partir dos próprios timbres dos instrumentos acústicos envolvidos na obra. A proposta seria a de criar para o meio eletroacústico uma paleta tímbrica graduando desde timbres similares ao timbre original do instrumento até timbres perceptivelmente distantes através de variações controladas dos parâmetros sonoros resultantes da própria análise espectral do timbre original. Para isso, o estudo envolve três fases que encontram projeção prática permeadas pelos constructos psicoacústicos: a análise espectral paramétrica de um som instrumental seguida de manipulações e geração modelada dessas variáveis para futura concreção do material eletrônico final através da síntese, que resultará em timbres similares, variados, expandidos do timbre original. Para se encontrar as maneiras pelas quais são obtidas as sensações de similaridade e distanciamento tímbrico que validassem as manipulações de parâmetros propostas, contou-se com os estudos da psicoacústica. Assim, as decisões tomadas nas escolhas dos parâmetros de análise, nas variações propostas desses parâmetros e nos algoritmos de síntese derivam, em parte, de referências às respostas do nosso sistema auditivo, de seu funcionamento geral e, em menor grau, das estruturas cognitivas no processo de apreensão sonora. As fases de análise, programação das variações e síntese envolvem o uso de diferentes programas computacionais que são abordados no trabalho (AudioSculpt, OpenMusic e Csound) sem que se trate de questões básicas de

manipulação e sintaxe desses *softwares* evitando transformar o trabalho em um manual de utilização. O caminho proposto é largamente discutido em todas as suas fases: quais os tipos de análise espectral em uso, porquê o autor escolhe o tipo FFT, cujas qualidades e limitações são descritas; depois, cada programação de variações tímbricas é demonstrada, explicada e exemplificada sonoramente; por fim é abordada a questão da síntese com a apresentação dos diferentes tipos de síntese sonora e a justificativa da opção feita em favor da síntese aditiva.



Phase Vocoder Analysis

Esquema geral da ordem de utilização dos *softwares*.

Outros utilitários computacionais que dão suporte paralelamente aos principais processos estudados são explicados e brevemente analisados, como o SoundHack e o programa utilitário de análise de *phase vocoder* chamado PVANAL.

O trabalho de pesquisa se inicia com o capítulo sobre interação tímbrica com o aporte da psicoacústica que fornecerá o substrato para muitas das decisões tomadas nas outras partes do trabalho. Do capítulo seguinte em diante seguir-se-á um encadeamento de estudos extraído da minha práxis composicional: a análise sonora, examinada em A Análise Paramétrica do Som, a manipulação e geração de dados, estudadas em Simulação

e Alteração de Parâmetros Sonoros para a Síntese e, por fim, o processo de concreção final dos arquivos de áudio a partir dos dados gerados e/ou manipulados anteriormente, visto no capítulo A Síntese. Dentro da área de pesquisa em Processos Criativos este trabalho se circunscreve na pesquisa de materiais e meios instrumentais para a criação musical.

2. A Psicoacústica na Interação

2.1 Processamento e Interpretação de Informações Sonoras e Musicais - Psicoacústica e Cognição

Quando se ouve uma obra musical instrumental, qualquer que seja sua formação, tem-se uma situação composicional (ou um contexto composicional) definida pelo âmbito ou limites impostos pelo material sonoro, este escolhido dentro do limite imposto pelo instrumento, e pela organização desse material definida pelo compositor, ou seja, o ponto de vista do material e o ponto de vista da escritura musical. Esta dicotomia de perspectivas (material sonoro *versus* escritura instrumental) se justifica na estrutura do nosso sistema de processamento de informação musical no qual ela é fundamentada. Esse sistema é complexo, se estende por diferentes partes do sistema auditivo (englobando ouvido interno e cérebro) e inclui várias etapas. Essas etapas podem ser resumidas, grosso modo, em: transdução do som recebido e análises espectrais, processos de agrupamento auditivo primitivo, decisões de agrupamento, computação de atributos perceptivos, ativação de estruturas de conhecimento abstrato, processamento de estrutura de evento e representação mental de forma musical (MCADAMS, 1996, p. 253). A comunicação entre cada uma das etapas não se dá necessariamente em série, especialmente no que concerne à apreensão de abstrações e de grandes estruturas temporais, além do que essas primeiras interpretações da informação sonora criam expectativas que alteram o processamento das novas informações em várias etapas. “As expectativas dirigem e influenciam a ativação de estruturas de conhecimento que afetam o jeito que nós interpretamos informações sensoriais subsequentes.” (MCADAMS, 1996, p. 254) Dessa cadeia de processos, os quatro primeiros são orientados para a concreção do material em entidades claramente definidas e que são sustentadas pelos estudos psicoacústicos, enquanto os três processos ulteriores são direcionados para o nível da escritura musical, esta munida do aporte dos estudos cognitivos¹. Um problema que se encontra na interpretação de informações sonoras/musicais é que apenas parte do processo extrai suas respostas das nossas propriedades fisiológicas, ou seja, somente

¹ Não é possível fazer distinção nem separação rigorosas entre as etapas dessa cadeia de processos, mas as segmentações apresentadas são necessárias pela sua função didática.

alguns estágios do processo todo podem ser universalizados e assegurados como característicos do ser humano. Não importa quem seja a pessoa e em que situação se encontre, o(s) modelo(s) de percepção de *pitch*², a capacidade de fusão ou segregação de eventos sonoros, o limiar de audibilidade, os mascaramentos simultâneos e sequenciais, a largura de banda crítica, a percepção da localização espacial, o rastreamento de uma fundamental ausente, serão idênticos aos de qualquer outra pessoa e assim, quando estudados, podem ser generalizados³. Mas uma outra parte desse processo de interpretação de informações sonoras/musicais, aquela ligada à macro-organização de eventos e às abstrações, é individualizada, é dependente de fatores externos e não pode ser generalizada.

Considerando que a natureza e organização destes estágios são provavelmente similares através das culturas em termos de mecanismos de processamento perceptivo e cognitivo subjacentes envolvidos, os processos de mais alto nível além da computação de atributos perceptivos depende bastante fortemente da experiência e conhecimento acumulado que é necessariamente específico de cada cultura. (MCADAMS, 1996, p. 255)

E em relação às estruturas do conhecimento abstrato, em continuidade à idéia acima, temos que:

Compreensão auditiva e muitos aspectos da percepção auditiva são baseados em generalizações que têm sido aprendidos de experiências específicas. Estas generalizações incluem conhecimento implícito que resulta de uma aculturação dentro de um dado meio acústico e sociocultural. Elas são abstratas no sentido que elas não são específicas a um dado padrão musical que se ouve. (MCADAMS, 1996, p. 255)

Ao mesmo tempo que temos esses dois níveis de interpretação de informações, temos ainda dois modos de conhecimento interdependentes em jogo: o chamado

² *Pitch* é um termo que se refere à sensação de altura definida de um som (ou mais que uma altura simultaneamente), mas que não é sinônimo de frequência (pois frequência é apenas a descrição física de uma onda e não está necessariamente associada à altura percebida) nem de nota (pois nota está associada à representação simbólica que muitas vezes difere do que se percebe). O termo é ainda hoje difícil de se definir, pois não se conhece com precisão a maneira pela qual o percebemos e o que está envolvido em sua formação. Como o termo é complexo e amplamente conhecido, ele será usado em inglês para se evitar confusões ou extensas explicações. Para maiores explicações, ver CHEVEIGNÉ, 2004.

³ Na verdade, não existe nenhuma característica perceptiva realmente idêntica entre duas pessoas e os testes realizados, como são baseados nos julgamentos das pessoas testadas, são aproximativos, portanto, inexatos. O fato é que essas pequenas diferenças perceptivas não impedem que as características possam ser generalizadas.

conhecimento implícito e o conhecimento explícito. Estes agem sobre o 'mundo externo' criando um entendimento que não possui uma relação isomórfica com as informações de origem, isto é, não existe um sistema de correspondência um a um entre a matéria analisada e o resultado sensorial de saída (KENDALL, 2005). O nível explícito seria aquele comandado por regras das quais temos consciência, como regras gramaticais, de teorias da música, etc., enquanto que no nível implícito encontraria-se um conhecimento ao qual não temos acesso de maneira consciente (KENDALL, 2005). O rastreamento de uma frequência fundamental ausente, por exemplo, é um processo que se encontra no nível implícito da audição, ou seja, não consciente.

As propostas apresentadas neste trabalho se encontram no quadro definido pela manipulação do material sonoro sustentada pelos estudos da psicoacústica, como parte do sistema de processamento de informação musical, e do nível implícito de conhecimento, aquele do qual não temos consciência de atuação na fruição, mas que permite o manuseio consciente no processo de criação.

2.2 A Questão da Interação

2.2.1 Introdução

Segundo o dicionário Houaiss da língua portuguesa, interação, em sua primeira acepção, significa exercer ação mútua (com algo), afetando ou influenciando o desenvolvimento ou a condição um do outro, ou em outra acepção: ter comunicação, relacionar-se. Pois são esses os sentidos empregados aqui referentes à convergência, para o ponto comum da percepção, dos sons oriundos dos domínios instrumental e eletrônico.

O que pode haver de diferente entre a interação percebida entre sons puramente instrumentais e entre sons de naturezas distintas, como na música mista, para ela ser, neste último caso, tratada normalmente como um problema ou um desafio? No caso da música instrumental podem-se encontrar soluções nas técnicas de orquestração, que podem ser estudadas em manuais que se encontram aos montes na literatura musical ou

ainda nas técnicas de composição. E no caso da música mista? Poderíamos afirmar que o domínio eletrônico tem sua própria gramática e portanto não se beneficia dos aprendizados acerca do relacionamento dos sons na música instrumental, mas já as próprias unidades mínimas passíveis de manipulação de cada uma das linguagens têm caráter distinto. Por um lado temos as notas (ou outros sons não convencionais) tocadas pelos instrumentos que não podem ser decompostas em seus elementos constitutivos, pois são as notas (ou outros sons complexos) as unidades mínimas dentro desse domínio, e por outro lado temos o meio eletrônico capaz de gerar sons puros individualmente⁴. Então como relacionar os diferentes meios presentes na música mista? Não há muitas teorias desenvolvidas a esse respeito, mas a prática é corrente. A situação da música mista hoje assemelha-se àquela da música barroca que, em sua época, era praticada em diferentes lugares com diferentes instrumentos e estilos até Rameau, nas primeiras décadas do século XVIII, já no final do período barroco, definir a gramática geral praticada até então.

Dentre os muitos estudos possíveis no espaço da teoria da interação na música mista, o que é proposto neste trabalho é o estudo das interações no nível do material sonoro empregando técnicas de conformação do material eletrônico ao escopo do material instrumental.

2.2.2 A Conformação do Material

Na música mista podemos dividir dois grupos de materiais que se relacionam no decorrer da obra de maneira diversa daquela relação puramente instrumental. O grupo de materiais sonoros instrumentais possui uma maleabilidade de conformações de características espectro-temporais (ou espectro-morfológicas) ao mesmo tempo que possui uma coerência constitutiva, da qual nasce a idéia da família de sons. Uma particularidade comum aos sons produzidos por qualquer que seja o instrumento é a sua organização interna: um conjunto coerente de relações entre seus atributos físicos constitutivos que os identificam e os caracterizam. Assim, um som instrumental pode ser

⁴ O som puro, a senóide, só pode existir estritamente em conceito, pois as condições limitadas do sistema binário, para a sua representação nos *softwares*, dos conversores de sinal, dos amplificadores e, por fim, dos alto-falantes, deformam, mesmo que minimamente, a onda resultante tornando-a 'impura'.

visto como um agrupamento de componentes cujos atributos físicos, ordenados de maneira particular, se fundem formando uma entidade percebida como um evento único. Já em uma obra musical eletroacústica mista, constituída por dois domínios distintos – o instrumental ou vocal e o eletrônico, não temos a mesma situação. O material sonoro próprio do domínio eletrônico não possui um grupo de características que o restringem a um âmbito de possibilidades e que, portanto, não oferecem a esse material uma identidade no sentido que a tem um instrumento. Sabe-se que os materiais de ambos domínios não são diferentes em essência, mas suas ‘configurações’ são distintas. Horacio Vaggione, referindo-se à música mista escreve:

Minha posição com respeito a isso é que é interessante procurar pontes muito bem acabadas para poder tornar possível uma interação íntima entre o resultado sônico e o próprio processo de composição, trabalhando sobre o mesmo objetivo musical, isto é, assumindo o postulado de uma possível convergência ou de uma vetorização comum. (VAGGIONE, 1996, p. 105)

Diferentemente da posição de Vaggione, uma discussão freqüente entre os compositores é que o trabalho com os meios eletroacústicos leva a uma riqueza sonora incompatível com os sons instrumentais.

É discutido freqüentemente que trabalhar com meios eletroacústicos dirige-nos para uma riqueza sonora excessiva que torna problemática qualquer tentativa de integração em um processo que se constitui igualmente de uma parte de escrita instrumental. Assim, de um ponto de vista contrário à “música mista”, os objetos sonoros eletroacústicos são complexos demais e, portanto, não suficientemente neutros para serem usados integrados em larga escala às manipulações composicionais. É também discutido que esses sons complexos destroem, pelo fato de suas próprias singularidades, a uniformidade que é uma condição necessária para um bom funcionamento de um processo combinatorial. (VAGGIONE, 1996, p. 105)

Essas posturas céticas em relação à viabilidade do relacionamento do material eletrônico com o material instrumental não levam em conta todo o trabalho ‘sobre’ o material eletrônico, seja este gravado ou de origem sintética, que se pode realizar em estúdio. Praticamente todas as transformações dos sons gravados e todas as manipulações de dados para a síntese, ou seja, a maior parte do trabalho em estúdio, tem o intuito de adaptar ou buscar os sons mais adequados para as necessidades que se elevam do processo de composição. Seja o ‘se deparar’ com os sons resultantes de processos de

transformações não-lineares, seja a escultura minuciosamente determinada e pré-definida de um tratamento de sinal ou síntese, o que se faz é buscar a integração do material a um contexto que pode ser só eletrônico ou misto. O trabalho sobre os sons eletrônicos para emprego na música mista exige uma direcionalidade diferente do que se requereria em uma obra que não envolvesse instrumentos porque a configuração inerente do material que é próprio da música instrumental é específico, limitado em suas fronteiras e pode estar contido no espaço de atuação do meio eletrônico, dependendo apenas da fonte de som eletrônico utilizada. Já sobre a uniformidade como condição necessária para um bom funcionamento de um processo combinatorial, alegada na citação acima, parte-se do princípio que a relação instrumento – eletrônica tenha que ser combinatorial. Isso é um disparate, pois mesmo na música puramente instrumental escrita durante boa parte do século XX as relações já haviam extravasado o velho paradigma da combinatoriedade. Na música mista temos uma nova situação composicional que envolve materiais em diferentes configurações e que demanda ‘novas escrituras’ e portanto novas maneiras de relacionar os grupos de materiais. Assim, a combinatoriedade, para quem queira se valer dela, ainda pode ser uma possibilidade, mas não uma condição exigida.

A proposta base deste trabalho se situa e se define dentro do contexto apresentado acima: achar as pontes mais apropriadas, que tornem possível uma interação íntima entre o resultado sônico e o próprio processo de composição, através da adequação do material eletrônico à configuração do material instrumental. A maneira pela qual se dará essa adequação será especificamente através do trabalho de síntese. A síntese sonora foi escolhida como processo chave da criação do material eletrônico porque fornece as condições adequadas de manipulação dos parâmetros sonoros através da síntese aditiva. A criação de modelos por meio de análises sonoras permite alterar isoladamente os parâmetros derivados do material instrumental e conformá-los mais detalhadamente do que seria possível com a transformação de um som já pronto. Contudo, não se quer dizer com isso que as manipulações de um som gravado sejam menos ricas, interessantes ou possuam menor potencial de manipulação do que a síntese oferece, a síntese apenas se adequa mais às minhas propostas pessoais de trabalho específico no domínio da música mista.

2.2.3 A Interação Som instrumental – Som Eletrônico

A interação que existe entre som instrumental e som eletrônico assume dois significados. O mais óbvio é a própria interação física entre as ondas sonoras durante suas propagações, mas esse fenômeno, a interferência, se dá entre duas ou mais ondas sonoras quaisquer, mesmo não tendo origem em fontes distintas⁵. O outro significado de interação é aquele ligado à percepção dos sons. Aí a interação é ‘interna’, ela depende do funcionamento do nosso sistema auditivo, estudado pela psicoacústica, e do nosso processo de organização estrutural e temporal que leva a um ‘entendimento’ das informações, regido pelos processos cognitivos. Assim, esse relacionamento poderia ser visto como a percepção da ação mútua exercida entre os sons, tanto no nível do material (da sua constituição mais propriamente) quanto no nível da macro-organização estrutural deste, ou seja, no nível escritural. E em qual situação há interação entre dois sons? A resposta é sempre. Alguma interação sempre vai haver simplesmente porque buscamos por ela. Há uma tendência inata de buscar organização em qualquer agregado de elementos que levará à formação mental de uma determinada estrutura para a pessoa que os estiver apreendendo⁶. Se a organização dos elementos for construída de modo a torná-la clara para os nossos sentidos, ela será mais unívoca, ou seja, ela será percebida da mesma maneira por muito mais pessoas. Mesmo dois sons quaisquer que não tenham ‘sofrido’ a intenção de serem organizados de forma alguma apresentarão alguma relação. Essa relação pode não ser percebida como integrante de uma estrutura, mas causará uma sensação única determinada pela influência que ambos sons causam mutuamente na condição de serem percebidos. Essa ação mútua dos sons, desconsiderando suas relações sintáticas e organizações estruturais, situa-se no espaço de correlação de suas propriedades sônicas, mais precisamente, nos graus de ‘parentesco’ que assumem suas estruturas espectrais e comportamentais. Para se obter graus de parentesco entre os sons envolvidos na música mista, já que a idéia exposta anteriormente é a configuração do material eletrônico com o intuito de adequá-lo ao contexto fornecido pelo material

⁵ Um exemplo é a interferência de uma onda vinda diretamente de uma fonte sonora com a onda dessa mesma fonte produzida imediatamente antes e que foi refletida. Característico no fenômeno de reverberação.

⁶ Bregman (BREGMAN, 1990) trata profundamente desse tópico através de seu conceito de *Auditory Scene Analysis* (Análises do Cenário Auditivo).

instrumental, é preciso um modelo e técnicas de manipulação que criem as diferentes gradações de proximidade e distanciamento perceptivo necessárias para se criar uma família de sons não mais intra-material (como é o caso dos sons instrumentais), mas uma família de sons inter-materiais. O que se deseja criar, então, é um grupo de sonoridades cujas configurações se situem entre a configuração própria do material instrumental e qualquer outra situação organizacional possível dentro das possibilidades do material eletrônico que as técnicas de manipulação, criadas para esse fim, permitirem.

2.3 A Psicoacústica na Interação

2.3.1 Psicoacústica, uma Introdução

A psicoacústica, um ramo de estudo da acústica, se forma na estreita relação com outras áreas, como a psicologia, a neuropsicologia e uma parte da cognição. A abrangência do estudo da psicoacústica é grande, mas pode-se destacar generalizadamente duas idéias principais: o estudo do funcionamento da audição e das respostas comportamentais frente aos fenômenos sonoros. A psicoacústica, ao contrário da acústica ou de outros ramos da física, é fortemente calcada em dados e medidas subjetivas porque geralmente depende da resposta de um ser vivo. Assim, muitas medidas irão variar dependendo do indivíduo estudado, de sua idade, integridade física, origem, cultura, nacionalidade, etc. Se, por exemplo, tomarmos um grupo de indivíduos nascidos no Brasil e compararmos suas respostas em alguns testes psicoacústicos com as respostas de um outro grupo de indivíduos nascidos na Suécia, iremos obter dados conflitantes. Ou ainda, se fizermos uma determinada medida psicoacústica hoje, como o limiar de audibilidade, em um grupo de brasileiros e a compararmos à mesma medida tomada em 1930, certamente os dados serão outros. Portanto é essencial, em grande parte dos estudos, se trabalhar estatisticamente, com ‘amostragens’ de indivíduos, consciente do

recorte realizado. É claro que se pode chegar a valores mais ‘universais’ quando, por exemplo, trabalhamos com análises fisiológicas.

O que podemos concluir a partir dessa abordagem preliminar é que, sejam valores medidos, sejam escalas criadas ou agrupamento de sensações, estes não podem ser considerados absolutos ou definitivos. Também torna-se impossível criarmos escalas com gradações precisas quando tratamos com subjetividades, como a sensação de identidade de um timbre. Como poderíamos graduar as sensações de semelhança entre dois sons? No caso do timbre tudo pode ficar um pouco mais confuso porque ele é o resultado do entrelaçamento de uma série de parâmetros do som que, isolados, já são motivos de estudos há décadas ou séculos.

2.3.2 A Sensação de Identidade Tímbrica

Chamo de sensação de identidade tímbrica à sensação que temos quando uma série de padrões de reconhecimento são equivalentes em dois sons distintos. Esses padrões de reconhecimento podem ser pensados como grupos de parâmetros sonoros cujos comportamentos temporais possuem coerência entre si, isto é, uma conexão tal entre a maneira como se delineiam que não seria possível que acontecesse por acaso. Ainda não há nenhum resultado concreto sobre ‘como’ exatamente as alterações de cada parâmetro constituinte do timbre muda a sua percepção (JENSEN, 2001, p. 15), ou seja, não existem correlações precisas entre os inumeráveis dados físicos que descrevem um timbre e as alterações sensoriais que este suscita. Existem diversos estudos direcionados à criação de escalas de sensações de acordo com a variação de cada parâmetro do som ou ainda de acordo com a alteração conjunta de uma série de parâmetros, mas estes apenas apontam direções, possibilidades. Vários desses estudos tentam representar três parâmetros sonoros que mais influenciariam na identidade de um timbre em um gráfico tridimensional e associá-los a variações mínimas na percepção desse timbre. Podemos relacionar diversos parâmetros do som que colaboram em níveis diferentes para a percepção de mudança da sensação tímbrica. Comum a diversos trabalhos é a tentativa de

determinar um JND (*Just Noticeable Difference*⁷) relacionado ao timbre em escalas perceptivas, mas os resultados ainda são imprecisos e não generalizáveis. O modelo híbrido de Jensen e Marentakis (JENSEN, 2001) propõe como principais parâmetros considerados para a pesquisa em alteração de sensação tímbrica os seguintes atributos:

- envelope espectral
- as durações de ataque e extinção⁸ do som
- as amplitudes relativas de ataque e extinção do som
- a inarmonicidade
- variações randômicas de amplitude (*shimmer*)
- variações randômicas de frequência (*jitter*)

Esses parâmetros podem ainda ser desdobrados em outros parâmetros analisáveis, como o desvio padrão sustentado de irregularidade de frequência, largura de banda sustentada de irregularidade de amplitude, etc.

Como esse trabalho não visa um estudo particular extenso de acústica ou psicoacústica, mas sim uma pesquisa voltada para a criação composicional, não será levada em conta uma quantidade ampla de parâmetros possíveis na alteração da sensação de um timbre, mas apenas aqueles que indicarem (através do presente trabalho) algum interesse para a composição musical. A partir da sugestão de parâmetros do estudo de Jensen e do estudo de diversos outros trabalhos na área de psicoacústica (ASHIHARA, 2003; BREGMAN, 1990; DANIEL, 1997; JENSEN, 2002; LAGRANGE, 2001; PRESSNITZER, 1997; ZWICKER, 1999), foi selecionado um grupo próprio de parâmetros para alteração dentro do complexo tímbrico que leva em conta o contexto deste trabalho, ou seja, interesses musicais, composicionais, a limitação dos *softwares* sendo utilizados e a necessidade de se limitar a abrangência desta dissertação. Seguem abaixo os parâmetros adotados para investigação nesse estudo:

- relação frequencial entre os parciais (harmonicidade/inarmonicidade)

⁷ Um termo comum em psicoacústica usado para designar o mínimo valor necessário para se perceber alguma diferença ou modificação de um determinado estado.

⁸ Tradução para *release*.

- alteração da amplitude média dos parciais (redistribuição da energia do espectro)
- alteração da distribuição das frequências no espectro
- micro-variações randômicas dos tempos de início de cada parcial
- variações na relação temporal de ADSR⁹ dos parciais
- variações na relação de amplitude entre o ataque e a extinção dos parciais
- *jitter*¹⁰
- *shimmer*¹¹
- mudança de fase dos parciais
- quantidade relativa de dados não-determinísticos (transientes e ruídos) acrescentados à parte determinística (ressíntese senoidal de parciais estáveis)
- seleção de dados (filtro psicoacústico)

Esses parâmetros também podem ser agrupados pela sua função perceptiva, mas é muito difícil nomear e descrever as sensações causadas na alteração dos valores de cada um dos itens acima. Contudo, podemos derivar algumas idéias gerais:

Harmonicidade – determinada pela distribuição e organização das frequências no espectro sonoro.

Rugosidade – refere-se à sensação de aspereza de um som e pode ter origens diversas. Duas das causas levadas em conta neste trabalho são a relação de fase entre os parciais de um som e *shimmer*.

Clareza – refere-se à precisão temporal da posição de início de cada parcial (não confundir com a medida acústica chamada fator de clareza). Essa clareza pode ser diminuída com micro-variações randômicas dos tempos de início de cada parcial.

⁹ Abreviação, em inglês, dos quatro segmentos de um envelope dinâmico: *Attack*, *Decay*, *Sustain* e *Release*. O *attack* (ataque) é relativo ao tempo gasto na transição do início do som com amplitude zero ao pico inicial de amplitude. O *decay* (queda) é o tempo gasto na variação entre o primeiro pico de amplitude e a parte mais estável do som. *Sustain* (sustentação) é o tempo que perdura uma amplitude sustentada mais estável do som. *Release* (extinção) é o tempo de variação entre o fim da sustentação e a extinção completa do som. A palavra 'queda' não é normalmente associada com o significado a que o termo *decay* remete, pois este não costuma ser traduzido. Assim, para facilitar o entendimento e evitar possíveis mal-entendidos, continuará a ser usado o termos em inglês.

¹⁰ Palavra usada para designar rápidas variações randômicas que acontecem nas frequências dos parciais de um som.

¹¹ Tem o mesmo sentido de *jitter*, mas aplicada às variações de amplitude.

Identidade – determinada pela somatória de todos os fatores adotados em maior ou menor grau. A soma de cada um dos fatores em quantidades variáveis relaciona-se com a sensação de identidade de maneira muito complexa e pouco previsível pelo que se conhece até agora.

Poderíamos perguntar então quais seriam os parâmetros envolvidos na caracterização de semelhança e dessemelhança ou similaridade e dissimilaridade de timbre e como medi-los. Estamos falando de proximidade perceptiva e semelhança de timbres. Na verdade existe o conceito que descreve tais aproximações e distanciamentos chamados fusão e segregação, ou ainda coerência e fissão. Estes conceitos, estabelecidos por Bregman em suas extensas ‘análises do cenário auditivo’¹², referem-se à nossa capacidade de, analisadas as características individuais dos sons que nos atingem, “... decidir como agrupá-las para que cada grupo tenha sido derivado do mesmo evento ambiental.” (BREGMAN, 1990, p. 642) Chamamos de fusão ao agrupamento das características dos sons, na forma de seus parâmetros individuais, que percebemos como derivadas de um único evento. Quando o processo gera mais que um grupo, o denominamos segregação. É possível também que um único evento sonoro não seja percebido como tal e segregado em mais de um grupo. Isso depende das características dos parâmetros que o compõe. O contrário também pode ocorrer se dois eventos sonoros distintos possuírem tamanha similaridade entre si que possam ser agrupados como um único evento. Há casos ainda nos quais apenas uma parte dos parâmetros de um evento encontra correspondência suficientemente precisa com os parâmetros de outro evento para se confundirem e, como consequência, tornarem-se parcialmente fundidos. O quanto um evento se funde ou se segrega de outro é um dos pontos-chave da interação dos materiais de origem instrumental e eletrônica e é sobre essas premissas que se desenvolvem as técnicas de manipulação de dados propostas neste trabalho.

Efetivamente o que se pode fazer em relação ao método de trabalho é adotar um modelo de timbre instrumental e criar variações nos diversos parâmetros considerados a fim de alterar mais ou menos sua identidade. Dependendo do intuito composicional pode-se criar materiais que difiram do modelo em apenas um aspecto (envelope dinâmico ou

¹² *Auditory Scene Analysis*, título de seu livro que trata do assunto (ver Referências Bibliográficas).

alteração de fase dos parciais, por exemplo) de modo que sob um aspecto o som se distancie perceptivelmente do som original, mas sob outros ele continue proximamente relacionado. Um processo de alteração com interpolação de parâmetros pode criar variações contínuas da correspondência precisa à total dessemelhança em todos os aspectos do som (que os recursos informáticos permitirem) ou em parte deles. Aqui cabe ao compositor julgar a conveniência/necessidade de alteração de cada um dos aspectos do som, o quanto alterar e se os materiais apresentarão características fixas ou se haverá variação dinâmica das mesmas.

2.3.3 Breve Descrição dos Parâmetros Utilizados e de seus Empregos

O primeiro parâmetro considerado para o trabalho das alterações de timbre é a relação frequencial entre os parciais, se harmônica ou inarmônica. A questão da harmonicidade não é tão óbvia, pois possui sutilezas das quais muitas vezes não nos damos conta. O que consideramos ser exatamente 'harmonicidade'? É dito que os parciais que possuem razão frequencial de número inteiro em relação à frequência de um parcial mais grave são harmônicos deste, mas para um som composto ser considerado harmônico precisamos que todos seus parciais sejam harmônicos? E precisamos que a razão entre eles seja exatamente de número inteiro? O modelo da série harmônica, já conhecido por todos, não corresponde exatamente à realidade dos sons chamados 'naturais' (não artificiais). O modelo de série harmônica é tão somente um modelo e serve para representar, grosso modo, as relações encontradas nos sons harmônicos. Os sons que conhecemos como harmônicos não possuem razões exatas de números inteiros entre seus parciais (no caso, harmônicos), mas ligeiras diferenças ou desvios em relação ao modelo da série, como acontece com o som do piano cujos harmônicos a partir de um certo número se desviam para o agudo criando diferenças em frequência cada vez maiores conforme aumenta o número do harmônico. Pequenas diferenças entre as frequências dos parciais reais e as frequências projetadas a partir do modelo não são suficientes para tornar um som inarmônico. Mesmo um som que contenha uma quantidade razoável de parciais totalmente inarmônicos (com grandes desvios em relação

ao modelo) pode ser percebido como harmônico dependendo de uma série de fatores e contextos. Portanto, a alteração de harmonicidade proposta precisa estabelecer valores de razão e uma organização desses valores que não criem margens de dúvida quanto à harmonicidade ou inarmonicidade resultante. Um aprofundamento nessas questões pode ser acompanhado nos estudos de psicoacústica de Zwicker e Terhardt.

A dilatação e contração frequenciais do espectro de um som muda necessariamente o seu grau de harmonicidade, pois redefine os espaçamentos entre os parciais e cria um novo âmbito de frequências, mas a correspondência entre alteração de parâmetro e sensação resultante é bastante complexa. Mudanças sutis nos fatores de alteração do espectro não correspondem necessariamente a pequenas mudanças perceptivas. Pequenas mudanças iniciais dos parâmetros podem levar a uma sensação grande de afastamento do som original porque fogem do modelo aural conhecido. Se aplicarmos novamente os mesmos fatores de alteração, a sensação de mudança será muito menor que a inicial, pois, apesar dos desvios de frequência serem os mesmos, não se parte mais de uma situação de familiaridade com o conteúdo espectral e portanto a idéia de distanciamento é minimizada.

Ao contrário da alteração anterior, a redistribuição progressiva exponencial bipolar de frequências sem alteração do âmbito espectral pode manter o caráter harmônico ou inarmônico originais do som. O fato de ser possível estabelecer diferentes índices de alteração para cima e para baixo de um centro frequencial dado pode garantir a preservação de toda uma parte inalterada do espectro enquanto altera-se a outra. Dependendo de que lugar do espectro se define como centro, podem-se fazer ajustes mais finos da mudança sensorial resultante. Mas diferentemente do que seria o senso comum, a preservação do âmbito, e conseqüentemente da frequência fundamental, não mantém a mesma sensação de *pitch*¹³, ele é deslocado conforme se altera o fator de distribuição das frequências no espectro¹⁴. A quantidade de deslocamento de *pitch* em relação à mudança dos fatores de alteração e deslocamento do centro de divisão espectral precisam ser determinados experimentalmente devido à enorme quantidade de variáveis presentes.

¹³ Há uma discordância na literatura em relação à definição de *pitch*. Os psicólogos cognitivos tendem a achar que *pitch* é um atributo sensorial, e portanto subjetivo, enquanto Ernst Terhardt vê uma incapacidade em se definir rigorosamente o termo devido a uma série de problemas psicoacústicos (TERHARDT, 2004c).

¹⁴ Ver mais detalhes em ZWICKER, 1999.

A relação entre os tempos de início dos parciais no ataque de um som, segundo Jensen, é um dos fatores determinantes do reconhecimento do timbre, portanto se alterarmos essa relação estaremos alterando a nossa capacidade de identificação, distanciando ou nos aproximando do modelo escolhido. A aproximação perceptiva é mais complicada que o distanciamento porque necessita de um modelo com boa precisão temporal na análise, ou seja, com espaçamentos entre cada análise que não exceda muito os espaços de tempo médios entre o início de cada parcial. Veremos no próximo capítulo que quanto maior a precisão temporal de uma análise, menor será sua precisão freqüencial. Assim, se procurarmos obter um bom modelo de tempo de início de cada parcial, a leitura das freqüências desses mesmos parciais será deteriorada. Um bom modelo é essencial porque simplesmente não existe um padrão de espaçamentos de tempo dos *onsets* nos sons instrumentais. Sempre se pode estudar caso a caso e inferir alguma média de espaçamentos que pode ser aplicada à síntese, mas será um trabalho árduo de tentativa e erro. O que se pode conseguir com mais facilidade é o distanciamento perceptivo pela alteração de um modelo derivado de análise ou implantação de um modelo abstrato já pronto. Mais uma vez, dificilmente encontraremos uma correspondência entre a alteração dos valores dos parâmetros e o grau de mudança sensorial causada.

O envelope dinâmico de um som se aplica em dois diferentes níveis: no som como um todo e no parcial isolado, apesar de ambos envelopes possuírem características diferenciadas. Quando falamos em envelope dinâmico normalmente nos referimos ao som como um todo, mas na construção de um som através de síntese aditiva também precisamos definir o comportamento de amplitude dos parciais que o compõe. O envelope dos parciais não é passível de generalização, pois cada parcial, em cada situação, em cada som, comporta-se de maneira própria. O ideal neste caso seria a utilização das leituras de amplitude de cada *frame* de análise para descrever o comportamento de cada um dos parciais, mas esse procedimento mostrou-se inviável devido a limitações dos *softwares* empregados em processar tamanha quantidade de dados. Como os dados referentes à variação de amplitude de cada parcial são transformados em um único valor médio de amplitude, todas informações a respeito de seu comportamento se perdem e assim um envelope substituto precisa ser aplicado. Se

não for assim, o início e o fim dos parciais causarão ruído devido à variação abrupta de amplitude. O que se pode fazer dentro das possibilidades que se tem à mão é criar algoritmos que definam o uso de diferentes tipos de envelope ao modo de um envelope ADSR, apesar de não ser essa a forma encontrada em envelopes de parciais¹⁵. Uma escolha individual não é possível porque temos, por vezes, milhares de parciais aos quais se deve associar uma descrição de amplitude. A criação do algoritmo terá que levar em conta as características internas que podem ser inferidas do som. O problema que se ergue nesta situação é que a soma da energia sonora em cada momento não será a mesma do som original apesar da aplicação de um envelope global. Uma solução potencial seria a cópia do envelope do som original e sua aplicação ao resultado da síntese, o que seria feito em Csound com o *opcode balance*, mas dependendo da situação isso poderia provocar novas descontinuidades abruptas de amplitude e causar o aparecimento de ruídos. Vemos que no caso da síntese aditiva não basta aplicarmos um envelope global para termos a sensação correta de variação de amplitude, precisa-se tentar descrever o mais corretamente possível as variações de amplitude internas do som.

A redistribuição da energia contida no espectro leva a mudanças sensoriais da ‘qualidade’ do som no sentido aproximado àquelas que um equalizador propicia, mas de forma um pouco diferente e muito mais ‘fina’. Em primeiro lugar, estaremos alterando a amplitude de cada parcial e não de uma região inteira de frequência, o que dá uma precisão muito maior na alteração do espectro. Em segundo lugar, a amplitude geral do som é preservada, já que, ao se aumentar a amplitude de parciais com pouca energia, diminui-se a amplitude dos parciais com maior energia com a aplicação de várias proporcionalidades que serão vistas no capítulo 4. A fim de tornar audíveis frequências ‘escondidas’ no espectro, o que vim chamar desvelamento tímbrico, não basta só aumentar suas amplitudes, é preciso criar as condições para que isso aconteça. O que impede que uma determinada frequência seja percebida não é apenas sua baixa amplitude, mas sua relação de distância de amplitude e frequência com outros parciais. A esse ‘impedimento’ chamamos mascaramento. Um parcial proeminente o suficiente em

¹⁵ Quando nos referimos ao envelope de um parcial, estamos nos referindo ao seu comportamento dinâmico de amplitude e não ao modelo ADSR. Aqui é feita uma adaptação porque a descrição da variação de amplitude de um parcial não é segmentada em um número definido de partes, mas a aplicação de um envelope, nesta situação, só é possível através do modelo ADSR.

amplitude pode mascarar outros parciais (impedir que eles sejam percebidos) que estejam a uma determinada distância em frequência. Essa distância é variável e depende da região do espectro onde os parciais se encontram, se os parciais estão mais ou menos próximos e se os parciais mascarados se encontram acima ou abaixo da frequência do mascarador¹⁶. Por isso é preciso que se forneça os parâmetros adequados para a execução da redistribuição de energia do espectro, cujo algoritmo já potencializa a fuga ao mascaramento. Dessa forma é possível mudar a idéia de identidade tímbrica somente reconfigurando suas energias internas, sem mudança alguma das frequências constituintes do som.

Jitter e *shimmer* são as micro-variações de frequência e amplitude que os sons não artificiais possuem e que são parcialmente responsáveis pelo caráter de naturalidade que os sons apresentam, se bem que suas funções exatas diferem entre si.

“O shimmer (irregularidade na evolução de amplitude dos parciais) adiciona uma qualidade ao som que pode ser de vivacidade e de ruído acrescido. ... A inexistência de shimmer torna o som morto, não natural, e muito shimmer aumenta o ruído e a irregularidade do som.

...

O jitter (irregularidade nas frequências dos parciais) adiciona uma qualidade diferente daquela que o shimmer adiciona ao som. O jitter fornece mais variações randômicas de baixa frequência ao pitch (com baixa largura de banda) ou adiciona rugosidade (com alta largura de banda) ...” (JENSEN, 2001, p. 17)

Jitter e *shimmer* são constituídos por variações randômicas e caracterizados por suas frequências e amplitudes de variação, isto é, se suas variações acontecem mais rapidamente ou mais lentamente em um determinado tempo e se suas variações alteram pouco ou muito os valores finais da variação. Existem estudos que indicam as taxas médias de variação de frequência e amplitude do *jitter* e do *shimmer* e, baseados neles, podemos criar adaptações desses valores que levem a resultados mais interessantes em cada situação. *Jitter* e *shimmer* também não são padronizados dentro do mesmo som, seus comportamentos normalmente apresentam variações em diferentes regiões do espectro. As implementações desses efeitos propostas neste trabalho prevêem até três configurações distintas aplicadas as mesmo som.

¹⁶ Como o assunto é muito amplo para ser tratado aqui, darei apenas as idéias gerais que forem necessárias. Para saber mais, ver ZWICKER, 1999.

As fases das ondas são responsáveis, de forma direta, pela sensação de rugosidade dos sons (PRESSNITZER, 1997) e de forma indireta, pela reconstituição da forma de onda do som original. Existem ainda vários outros fatores que contribuem para a sensação de rugosidade (como visto acima) agindo de outra forma sobre o nosso sistema de processamento de informações sonoras, mas esta forma tem a vantagem de não envolver outros parâmetros sonoros, como frequência ou amplitude. Entretanto, como não foi possível se fazer a leitura de fases dos parciais analisados por limitação dos recursos informáticos utilizados, foi necessária a criação de modelos artificiais de fases. A ausência de um modelo de fases na síntese implicaria, mesmo assim, no uso de um determinado modelo, pois não se pode simplesmente desconsiderar as fases na síntese aditiva - a onda sonora precisa começar em algum ponto de sua representação. Quando não se fornecem dados referentes a fases na síntese aditiva, fica implícita a utilização de fase zero para todos os parciais e isto é um modelo, apesar de divergente do original. A relação de fases dos parciais de um som é bastante complexa e de forma alguma se aproximaria de um modelo no qual todas as fases fossem iguais. Com o intuito de criar várias sensações de rugosidade distintas, criando assim grupos de sons que divergem só neste aspecto, foram propostos cinco diferentes modelos gerais de fase, três dos quais podem fornecer um sem número de configurações ligeiramente modificadas. Dessa forma consegue-se trabalhar com uma sensação específica independentemente de se trabalhar com outras.

3. A Análise Paramétrica do Som

3.1 Introdução

A análise é uma peça chave no cumprimento do objetivo a que me proponho nessa pesquisa já que, para a criação de material sonoro sintético com função de ligação perceptiva entre os domínios instrumental e eletrônico, há necessariamente uma dependência da investigação estrutural e comportamental dos sons instrumentais empregados na composição mista, ou seja, a extração de um modelo. A necessidade de um capítulo inteiro dedicado à explicação dos processos envolvidos em uma análise paramétrica de um sinal sonoro é fundamental para o entendimento das escolhas que serão feitas dos tipos e parâmetros de análise e das razões pelas quais determinados processos de síntese são preferíveis ou mais adequados que outros. Algumas das etapas desta seqüência de processos, principalmente referentes à digitalização, não são passíveis de modificação, pois são dependentes das estruturas dos *hardwares*¹, mas nos levarão a uma melhor compreensão dos seus dados evitando que resultados não esperados sejam atribuídos à falhas ou imprecisões nas subseqüentes etapas da análise. Os assuntos abordados neste capítulo sobre análise paramétrica são referentes a uma parte mais teórica, essencial para a compreensão do processo de análise, e a uma outra parte ligada à prática, a análise propriamente dita, toda executada no *software* AudioSculpt². Portanto, a direção que este capítulo toma é grandemente conduzida pelos conceitos, algoritmos e procedimentos de análise próprios do *software* empregado. A versão utilizada é a 1.2B1. Apesar de existirem versões mais recentes com a mesma interface (1.4, 1.7, 1.9), a versão 1.2B1 é sabidamente a mais estável e funcional. Novas versões sob uma interface remodelada foram lançadas (2.0, 2.1, 2.2), mas estas ainda não possuem a implementação de uma análise fundamental para este trabalho que é o *partial tracking*, assim, a versão mais antiga (mas funcional) foi utilizada.

¹ Refiro-me aos conversores AD-DA (analógico-digital, digital-analógico) integrantes das placas ou sistemas de áudio e dos computadores que executam essa função, como os computadores Macintosh da Apple. A maior parte dos *hardwares* fornece poucas opções de escolha de taxa de amostragem e número de bits, pois estas estão confinadas à sua arquitetura e ao propósito com o qual eles foram projetados.

² O AudioSculpt é um *software* de análise e ressíntese desenvolvido pelo IRCAM que realiza transformação sonora através dos dados de análise de FFT e uma série de diferentes tipos de análise, incluindo, na maior parte dos casos, suas representações visuais.

3.2 Digitalização

3.2.1 Digitalização e suas Etapas

A digitalização é o processo de conversão de um sinal analógico em sinal digital, o que viabiliza a utilização dos sons em equipamentos com sistemas digitais como o computador. A digitalização não é um processo único, ela envolve uma seqüência de sub-processos com diferentes funções:

- Filtragem com filtro *low-pass*³
- Amostragem
- Quantização

Há ainda uma etapa anterior a essas que muitas vezes é considerada como parte do processo de digitalização porque dela depende a efetivação do processo que é a transformação do som em sinal analógico. Essa transformação consiste apenas na conversão da variação de pressão em variação de corrente elétrica executada pelos microfones. Tecnicamente, essa transformação não faz parte da digitalização.

3.2.2 Amostragem e a Física do Som

A amostragem (ou *sampleamento*, *aportuguesamento de sampling*), muitas vezes confundida com a própria digitalização, é apenas uma das etapas do processo de conversão de um sinal analógico em sinal digital, o que não minimiza sua importância. Para entendermos como essa conversão ocorre, será útil recorrer a alguns princípios fundamentais da física do som.

³ O filtro *low-pass* é um tipo de filtro usado para atenuação/corte de frequências agudas acima de uma determinada marca, chamada *cut-off frequency* ou frequência de corte.

O som é uma onda mecânica formada por um fluxo contínuo de compressão e descompressão do meio onde ela ocorre, no mais comum dos casos, o gasoso, ou seja, o ar. O aumento e diminuição da pressão do ar em relação à pressão atmosférica local acontece através do aumento e diminuição da concentração das moléculas dos diferentes gases constituintes do ar de forma contínua, sem saltos ou quebras, já que estamos falando de deslocamento de matéria e esse se dá no tempo. É essa variação contínua de pressão que costumamos representar nas formas de onda, como nos gráficos bidimensionais cartesianos com o eixo y representando a pressão e o eixo x representando o tempo no qual acontece a variação de pressão.

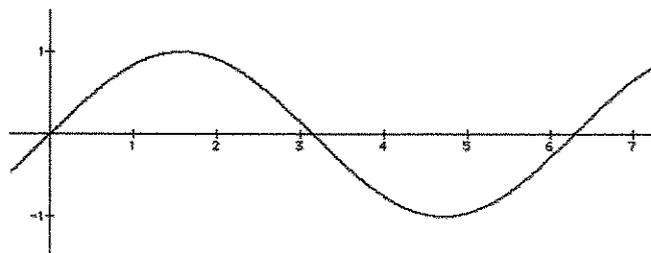


Figura3.1 Representação de uma onda senoidal.

Quando gravamos um som, a variação contínua de pressão do ar é convertida pelo microfone para uma também contínua variação de tensão elétrica de maneira análoga ao fenômeno contínuo de entrada e assim obtemos o sinal analógico (TENÓRIO, 2001, p. 41).

3.2.3 Sistema Digital

A digitalização é a transformação do sinal analógico em sinal digital e envolve várias etapas, mas o sistema digital tem princípios diferentes daqueles do sistema analógico. Enquanto o sinal analógico é um fluxo contínuo de informações, o sinal digital é uma seqüência de amostras contendo valores numéricos que descrevem o valor da tensão em cada ponto amostrado (no caso do som). Como o sistema digital no

computador funciona através do código binário, apenas a alternância entre zeros e uns pode ser registrada, impedindo assim a variação contínua de uma determinada grandeza. Tomemos como exemplo o desenho de um círculo desenhado no computador através de algum programa de imagem apropriado. Uma curva é uma variação contínua de pontos e não pode ser reproduzida como tal na tela de um computador, pois a tela é constituída por uma grade de pontos dispostos em linhas e colunas (chamados *pixels*) incapazes de representar uma variação contínua. O que se faz então é gerar pequenos segmentos de reta alternando as posições horizontal e vertical. Se os segmentos forem suficientemente pequenos, teremos a idéia que se trata de uma curva, com a qual poderemos simular nosso círculo.



Figura 3.2 e 3.3 A curva à esquerda foi representada por 32 pontos (ou alternâncias de segmentos) enquanto a curva à direita por 2048 pontos⁴.

Um círculo poderia também ser definido como um polígono de infinitos lados. No caso do som, os problemas envolvidos são bem maiores, pois há muito mais a ser representado.

3.2.4 Taxa de Amostragem

Podemos dizer que, de certa forma, o nosso sistema auditivo funciona de maneira análoga ao nosso sistema de visão. No cinema, para termos a sensação verossímil de movimento precisamos de uma seqüência mínima de aproximadamente 24 fotogramas

⁴ Evidentemente a resolução final da imagem depende da resolução do monitor e da placa de vídeo do computador que está sendo usado, ou seja, suas capacidades em representar tantos pontos em uma área limitada. Além disso temos a limitação de resolução da impressora usada, mas, de qualquer forma, pode-se facilmente entender a idéia por trás do uso dessas imagens.

sendo mostrados a cada segundo. Já na audição, precisamos de uma seqüência mínima de aproximadamente 44.100 amostras de som sendo tocadas a cada segundo para termos sensações realísticas da maior parte dos sons⁵. Se uma quantidade menor de amostras for tocada no mesmo espaço de tempo, não perderemos a sensação de “movimento” do som, a questão envolvida aqui é de outra natureza. Taxas menores de amostragem⁶ (quantidades de amostras por tempo, usualmente segundos) significam a incapacidade de representação de certas faixas de freqüência, normalmente parciais muito agudos de algum som. E daí podemos derivar uma correspondência entre taxa de amostragem e representação de freqüências: quanto menor a taxa de amostragem, menos freqüências agudas poderão ser representadas. A consequência disso resultaria em sons mais apagados, opacos, com menor definição do espectro e portanto menos verossímeis, além da geração de freqüências inexistentes devido a erros na amostragem. A verossimilhança dos sons não depende exclusivamente da taxa de amostragem, mas também da quantidade de *bits* que ‘representarão’ a amplitude do som. Esse assunto será abordado mais tarde em Quantização.

A especificação de um valor mínimo para uma taxa de amostragem adequada à nossa percepção leva em conta vários fatores. Um dos principais fatores é um dado psicoacústico que estabelece um âmbito de freqüências capaz de ser percebido como som audível pelo ouvido humano: 20 a 20.000 Hertz. Esse âmbito é variável pois, além de ter sido estabelecido a partir da média dos resultados obtidos com experiências feitas com uma amostragem de pessoas, ele varia com o passar da idade, com a quantidade de exposição sonora sofrida durante a vida, com características físicas pessoais, etc. Então, a partir desse dado, poderíamos nos dar por satisfeitos se freqüências até aproximadamente 20.000 Hertz puderem ser representadas no computador.

⁵ Alguns sons, por serem puros ou por terem um espectro limitado em freqüência, não necessitam de tantas amostras. Isso será explicado mais tarde.

⁶ A expressão em inglês, *sample rate*, é amplamente empregada no Brasil.

3.2.5 Teorema de Nyquist

Outro fator fundamental a ser levado em conta no estabelecimento de uma taxa de amostragem é um princípio chamado Teorema de Nyquist. Esse teorema estabelece que

“... a seqüência discreta de tempo de uma função contínua amostrada $\{V(t_n = n.T_s)\}$ contém informação suficiente para reproduzir exatamente a função $V = V(t)$ desde que a taxa de amostragem ($f_s = 1/T_s$) seja pelo menos duas vezes aquela da freqüência mais alta contida no sinal original $V(t)$.” (ROLAND-MIESZKOWSKI, 2003)

$$V(t) = \sum_{n=-\infty}^{+\infty} V[n] \frac{\sin[\pi(t - nT_s)]}{\pi(t - nT_s)}$$

Figura 3.4 Expressão do Teorema de Nyquist.

Onde:

$f_s = 1/T_s$ -> freqüência de amostragem

$V(t)$ -> valor do sinal (vtagem) em um tempo arbitrário t

$V[n] = V(nt_s)$ -> valor do sinal no tempo $t = nT_s$

Podemos expressar esse teorema de forma mais simples da seguinte maneira: temos que usar uma taxa de amostragem que seja pelo menos duas vezes maior que a freqüência que queremos representar. Se ouvimos, a princípio, freqüências até 20.000 Hz⁷, precisaremos de uma taxa de amostragem que seja pelo menos o dobro dessa freqüência, ou seja, 40.000 amostras por segundo. Devido a erros e imprecisões geradas quando tentamos representar uma freqüência próxima à metade da taxa de amostragem (chamada

⁷ Conseguimos perceber, a grosso modo, freqüências de até 20 KHz isoladamente, o que não implica que freqüências mais altas que essa não possam interferir na nossa percepção de timbres complexos. Algumas pesquisas demonstram que parciais acima de 22 KHz influenciam a percepção de espectros harmônicos complexos, mas de maneira sutil e em condições restritas (ASHIHARA, 2003, p. 545).

frequência de Nyquist), estabeleceu-se como padrão uma taxa de amostragem maior que 40.000 Hz e que ainda é o padrão na gravação de CDs – 44.100 Hz⁸.

3.2.6 Aliasing

Além de estarmos limitados a frequências que sejam no máximo metade da taxa de amostragem, se convertermos um sinal que possua frequências que ultrapassem esse valor, não apenas deixaremos de poder representá-las como também gerar-se-ão frequências inexistentes (como citado acima) devido a um erro chamado *aliasing*. O *aliasing* (também conhecido como *foldover*) é um erro de interpretação, ou melhor dizendo, de reconstrução de uma onda a partir dos valores de amplitude da onda obtidos em cada amostra. Como o computador é incapaz de acompanhar uma variação contínua, os valores de amplitude da onda digitalizada são medidos em intervalos regulares, através das amostras, e interpolados de tal forma que se possa ‘reconstruir’ a forma de onda original. No exemplo abaixo temos uma frequência de 7.000 Hz que está sendo amostrada na taxa de 8.000 amostras por segundo. Segundo o Teorema de Nyquist, só poderemos representar frequências até a metade da taxa de amostragem, nesse caso, 4.000 Hz. Podemos ver como resultado do erro a geração de uma determinada frequência muito mais grave que aquela sendo amostrada.

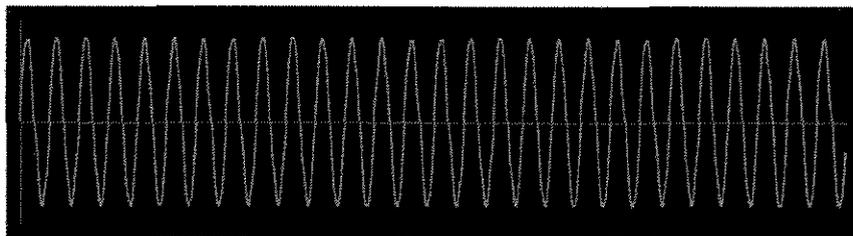


Figura 3.5 Representação de uma frequência a ser digitalizada.⁹

⁸ Um desses erros é um ‘efeito colateral’ do processo de filtragem inicial. O filtro *anti-aliasing*, usado para prevenir o sistema contra frequências que estejam acima da frequência de Nyquist, tem que ter uma ‘curva de corte’ muito acentuada – em torno de 90 dB por oitava – e dessa maneira acaba gerando uma distorção de fase do sinal que faz o âmbito de frequências se ampliar. A fim de evitar esses efeitos, métodos muito complicados e sofisticados como *oversampling* são empregados.

⁹ As figuras 3.5, 3.6 e 3.7 foram criadas em um *site* de exemplificação de *aliasing* por meio de um *applet* java em DIGITAL, 2004.

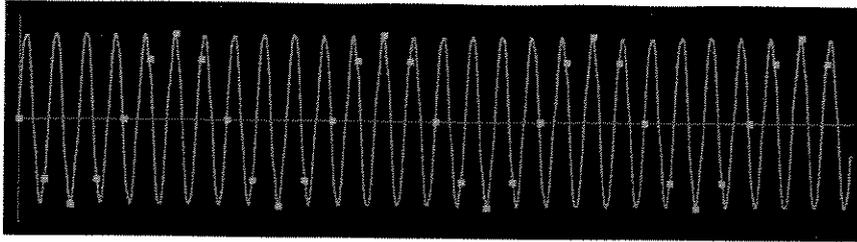


Figura 3.6 Os pontos verdes representam a seqüência de amostras a intervalos regulares no eixo x e seus valores individuais representados no eixo y.

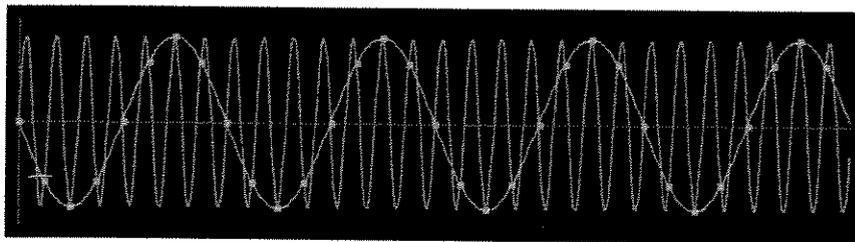


Figura 3.7 Reconstrução de onda a partir dos valores da amostragem.

A frequência gerada pelo *aliasing*, considerando a frequência original de 7.000 Hz e a taxa de amostragem de 8.000 Hz, é de 1.000 Hz. Para se calcular a frequência de *aliasing*, basta subtrair a frequência sendo amostrada da taxa de amostragem.

A fim de evitar esse erro, um filtro do tipo *low-pass* tem que ser usado antes do processo de amostragem para assegurar que nenhuma frequência acima da frequência de Nyquist seja erroneamente interpretada.

3.2.7 Quantização

A quantização é o processo de se estabelecer um valor referente à amplitude para cada amostra do som e é realizada por um aparelho conversor chamado ADC (*analog-to-digital converter*). Esse processo é importante não apenas pela fidelidade na descrição das corretas amplitudes do som original, mas também pela interferência no processo de análise através do ruído que pode ser adicionado ao som devido ao erro de quantização. O âmbito de números que podem ser usados pelo computador para descrever os valores de

amplitude do som depende do número de dígitos binários (*bits*) usados para se gravar cada número (ZICARELLI, 2004, p.18). Para se entender como ocorre a quantização e o erro derivado dela, precisamos entender o sistema binário.

3.2.8 O Sistema Binário

No sistema binário, qualquer número é descrito apenas com zeros e uns. Assim, os números 1, 2, 3, 4 e 5 são representados como 01, 10, 11, 100 e 101 e quanto maior o número no sistema decimal, mais dígitos precisaremos para representar no sistema binário, mas cada dígito a mais representa um aumento de potência de dois. O número 64 será convertido para 1000000 e o número 255 para 11111111. Assim, se eu tiver seis dígitos no sistema binário, eu poderei representar 2^6 números do sistema decimal, ou seja, 64 (64 números, significando 0 a 63) . Isso mostra que, para eu representar o número 64 (o sexagésimo quinto número iniciando do zero) precisarei de mais um dígito, como foi exposto acima. Se tivermos um *byte* (8 *bits*) para representar as amplitudes, teremos 2^8 números, ou 256. A quantidade padrão de *bits* estabelecida para uso no CD é de 16 *bits*, perfazendo 2^{16} números, ou 65.536.

A quantidade de números possíveis de serem representados no sistema binário equivale à máxima resolução de amplitude. Esta especificação de quantidade, em número de *bits* usados para representar uma amostra, é chamada *largura da amostra* ou *nível de quantização* de um sistema (*sample width* e *quantization level* respectivamente) (ROADS, 1996, p. 35). Se se tiver um valor de amplitude que se encontra entre dois números de representação, esse valor será aproximado para o número de representação mais próximo, alterando assim o valor original. Mas se a resolução for maior, ou seja, maior quantidade de números para representação, o espaço entre os dois valores será menor e o erro causado pela aproximação será minimizado. É aí que se encontra boa parte da lenda da maior qualidade do LP em relação ao CD. A precisão na representação da amplitude é grandemente responsável pela “fidelidade” das gravações e como no sistema analógico a representação da amplitude é contínua, podemos entender os motivos que fundamentam os defensores da qualidade do LP.

3.2.9 O Erro de Quantização

O erro de quantização é justamente a aproximação que o valor de amplitude lido na amostra sofre para se encaixar ao valor mais próximo definido pelo espaçamento entre os números de representação disponíveis. “Já que o erro de quantização para cada amostra é geralmente randômico ..., nós normalmente ouvimos o erro do efeito de quantização como ruído branco.” (ZICARELLI, 2004, p. 18).

O grande problema do erro de quantização quando relacionado com a análise paramétrica do som é que podemos interpretar erroneamente o ruído inserido pelo erro como parte constituinte dos sons analisados e, dessa forma, manter e manipular esses dados para uso em uma futura ressíntese.

3.2.10 O Desafio da Fidelidade

Com o grande desenvolvimento dos meios tecnológicos e a busca do mercado fonográfico pelo ‘som perfeito’, as empresas fabricantes de equipamentos de áudio tem investido enormemente nos aumentos das taxas de amostragem e na quantidade de *bits* disponíveis nos equipamentos de áudio profissional. Logo que surgiram os primeiros DATs (*Digital Audio Tape*), também surgiu uma nova taxa de amostragem de 48 KHz que se estabeleceu ao lado da taxa de 44.1 KHz do CD e perdurou até há poucos anos atrás quando as empresas resolveram investir no dobro da taxa mais alta usada até então – 96 KHz¹⁰. Enquanto muitos se questionavam se havia a necessidade de se usar tal taxa ou se era apenas uma jogada mercadológica, os estúdios adotaram o novo valor e um novo padrão começou a se estabelecer, pois alguns DVDs de áudio também passaram a usá-lo. Há pelo menos dois anos uma nova marca foi atingida e vários equipamentos de áudio profissionais já se utilizam da taxa de amostragem de 192 KHz.

Ao lado das novas taxas de amostragem, o número de *bits* utilizados na digitalização do som passou de 16 para 18 e então 20 e finalmente 24 *bits*. O aumento do

¹⁰ Há também a taxa de 88.2 KHz, mas essa foi pouco utilizada e não se tornou padrão, apesar de estar disponível em alguns equipamentos de áudio.

número de *bits* parece pouco em relação à taxa de amostragem, mas devemos lembrar que esse número se refere à potência de dois. A mudança de 16 *bits* para 24 *bits* representa, na verdade, um salto de 65.536 para 1.67772×10^7 números de representação. Enquanto a taxa de amostragem quadruplicou do último padrão até hoje, o número de bits se multiplicou por 256.

O processo de digitalização como um todo, apesar de essencial e gerar conseqüências diretas nos resultados das análises, não integra o processo de análise propriamente dito.

3.3 O Padrão FFT

3.3.1 Introdução

FFT – *Fast Fourier Transform* (transformada rápida de Fourier)¹¹ – é um nome bem conhecido no meio musical eletroacústico, mas mal compreendido. Deparamo-nos com a FFT em alguns *softwares* quando da execução de diversas tarefas: como quando precisamos realizar análises, sejam elas representações visuais, como o sonograma ou apenas visualizações dos dados de análise em formato texto, ou quando precisamos realizar algumas transformações de um som, como modificações da escala temporal (*time-stretching* e *time-shrinking*¹²) sem alteração de altura, transposições sem alteração temporal, filtragens, etc. Na maior parte dos casos, simplesmente aceitam-se os parâmetros sugeridos para a análise e em outros casos esses parâmetros simplesmente não são disponibilizados pelos *softwares*.

Uma das causas do desconhecimento do processo de FFT pela maior parte dos músicos é a sua origem. FFT não é um processo que nasceu no meio musical, foi desenvolvido nos campos da matemática e engenharia e ainda é dominante nessas áreas,

¹¹ Usarei a sigla em inglês, assim como em vários outros casos que apresentar-se-ão durante esse capítulo, por serem internacionalmente utilizadas e não possuem tradução consensual.

¹² Os nomes desses processos, assim como os de muitos outros no campo da música eletroacústica, são mais conhecidos em inglês e significam respectivamente estiramento temporal e compressão temporal.

além do que, não está diretamente relacionada com o mundo sonoro e muito menos com o musical. Uma outra causa, e acredito ser a mais importante, é a grande quantidade de conhecimento matemático envolvido e a falta de cursos regulares de música eletroacústica que possam minimizar o lento e árduo aprendizado autodidata. Com isso, dificulta-se enormemente a compreensão do processo por parte dos músicos que, pelo menos no Brasil, não têm boa formação matemática e nem bibliografia especializada à disposição.

A FFT é uma ferramenta poderosa integrante de um dos principais processos da análise sonora: a STFT – *short-time Fourier transform* ou *short-term Fourier transform* (transformada de ‘tempo curto’ de Fourier e transformada de curto termo de Fourier, respectivamente). Apesar do processo de análise que inclui a FFT não ser o único, ele é o mais empregado, o de mais fácil acesso e ainda um dos mais úteis no mundo musical desde que foi introduzido em 1965, pois fornecia descrições dinâmicas de frequência com suas correspondentes posições temporais e com gasto computacional muito menor que aquele que se tinha antes.

A idéia apresentada no título desse subcapítulo de que a FFT é um padrão baseia-se antes na sua ampla utilização (por motivos variados¹³) que na sua adequação para a análise de sons que não sejam estritamente harmônicos. Apesar de sua limitação, a FFT será usada pois fornecerá dados de grande importância na análise da maior parte dos sons instrumentais convencionais.

O entendimento do funcionamento dessa ferramenta é essencial, pois dele depende a escolha adequada dos parâmetros de análise e conseqüentemente a obtenção do melhor resultado para os fins previstos. Parâmetros de análise inadequados podem gerar resultados que não correspondem ao som original atrapalhando nossa compreensão do som quando o fim é a própria análise, e dando vazão a inúmeras distorções, quando do seu emprego em sínteses ou transformações do som através do meio analítico. Para começarmos a entender a FFT, necessitamos entender antes o que é e o que está envolvido em uma análise.

¹³ Para maiores explicações, ver MASRI, 97a.

3.3.2 Análise Paramétrica

Análise sonora é um conceito muito amplo e pode se referir a campos variados, como a análise da função musical de um som em uma determinada estrutura e a análise espectro-morfológica (SMALLEY, 1995, p. 107-126). Como conceitua Marie-Hélène Serra: “A análise sonora abarca todas as técnicas que dão descrições quantitativas das características do som.” (SERRA, Marie-Hélène, 1997, p. 32). A análise tratada aqui será referida como análise paramétrica, pois o que nos interessa são as descrições quantitativas medidas segundo parâmetros constitutivos, como frequência, amplitude, fase, energia de pico, duração, etc.

Depois de digitalizado um som, temos a reconstituição de sua forma de onda que será objeto de análise. A forma de onda é um tipo de representação do sinal na qual, em geral, apenas as características de variação de amplitude em função do tempo podem ser observadas. A partir do desenho da forma de onda não conseguimos intuir sua frequência fundamental, se se tratar de um som harmônico, e nem sua composição espectral. Podemos imaginar algumas raras exceções de sons, constituídos do agrupamento de apenas algumas poucas senóides¹⁴ organizadas harmonicamente e sem diferença de fase entre elas, a partir dos quais poderíamos nos aproximar de algum valor de frequência fundamental se levarmos em conta o tempo em que essas ondas acontecem. Isso seria possível porque os sons harmônicos são necessariamente periódicos, ou seja, suas ondas se repetem em tempos regulares. Mesmo assim, não teríamos nenhuma idéia de sua composição espectral. Já o som inarmônico, seja ele complexo, ruidoso ou ruído propriamente dito, possui ondas aperiódicas, as quais não contém nenhuma repetição regular de seu desenho e, dessa maneira, não contendo qualquer frequência fundamental.

¹⁴ Na verdade é uma questão de interpretação dizer que um som é constituído de senóides ou cossenóides. A verdadeira idéia, mas ainda simplificada, é que, qualquer som periódico mais complexo que uma senóide, segundo o Teorema de Fourier, ‘pode’ ser decomposto em uma soma de senóides com diferentes frequências, amplitudes e fases. O assunto será retomado com maior profundidade ainda neste capítulo.

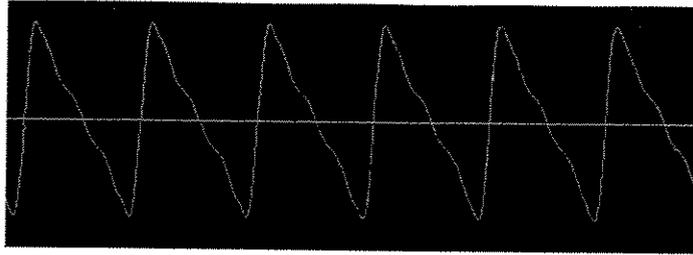


Figura 3.8 Forma de onda simples, claramente periódica, sintetizada a partir dos oito primeiros harmônicos.

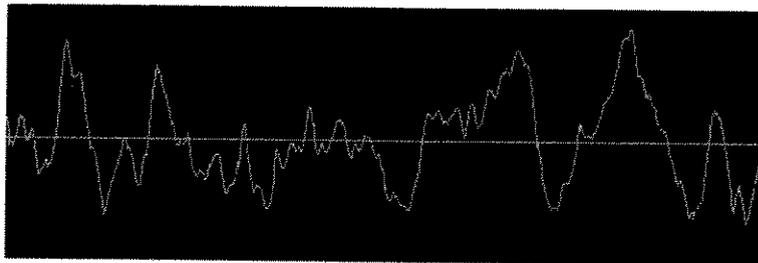


Figura 3.9 Forma de onda complexa, aperiódica e, portanto, sem estimativa de frequência.

Dizemos que esse sinal se encontra no ‘domínio temporal’ (*time domain*). A partir desse domínio unidimensional¹⁵ devemos realizar uma conversão tal que características frequenciais do sinal tornem-se disponíveis para análise. Com essa conversão chegamos ao chamado domínio tempo-frequência (*time-frequency domain*). Esse processo de conversão é chamado de TFR – *time-frequency representation* (representação tempo-frequência). Esse é o primeiro passo da STFT imediatamente seguido pelas estimativas de magnitude e fase segundo um complexo algoritmo chamado FFT. Esse algoritmo é derivado do Teorema de Fourier.

3.3.3 Teorema de Fourier

Em 1822, o matemático e físico francês Jean-Baptiste Joseph Fourier publicou um tratado no qual ele desenvolvia a teoria que “... complicadas vibrações poderiam ser

¹⁵ Chamado assim por representar apenas amplitude em função do tempo.

analisadas como uma somatória de sinais simples simultâneos. Em particular, Fourier provou que qualquer função periódica poderia ser representada como uma somatória infinita de termos seno e cosseno.” (ROADS, 1996, p. 545). Fourier não se referia ao sinal sonoro, mas as aplicações advindas de seu teorema passaram a ser cada vez mais extensas até que começaram a ser empregadas na análise de som.

O processo de obtenção da soma de senos e cossenos a partir de ondas mais complexas é chamado análise de Fourier ou transformada de Fourier. A descrição da transformada de Fourier é a seguinte:

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt$$

Figura 3.10 Expressão da transformada de Fourier.

Onde:

$X(\omega)$ é o espectro no domínio frequencial;

$x(t)$ é o sinal de entrada;

t é tempo;

j é o número complexo¹⁶;

ω é frequência.

Essa formulação tem como pressuposto que o sinal é infinito e contínuo no tempo e tem como resultado a somatória de um número infinito de senóides para representar o sinal periódico de entrada. Do jeito que se apresenta, essa transformada nos dá uma porção de problemas em vez de soluções, já que o sinal que desejamos analisar, dentro de nosso sistema digital, é finito, discreto (não contínuo), não pode ser representado por uma soma infinita e normalmente não é periódico ou simplesmente não é periódico por toda sua extensão. As soluções para a equação acima não descrevem ‘quando’ as frequências ocorrem e nem se elas variam no tempo. Assim, frequência e tempo são contínuos. Com

¹⁶ Também chamado de número imaginário com símbolo i , esse número é o resultado da raiz quadrada de -1 .

o intuito de adaptar o teorema, várias modificações foram implementadas. Para conseguirmos resultados no qual o tempo é discreto, (ou seja, nos sistemas digitais, já que não é possível a representação contínua), chegou-se à formulação da DTFT – *discrete time Fourier transform* (transformada de Fourier de tempo discreto) onde a transformada de Fourier original é aplicada a tempos discretos. Mas ainda temos frequências resultantes da análise que não variam no tempo e assim outras adaptações foram necessárias até chegarmos à formulação de um modelo adequado para usos mais específicos chamado DFT – *discrete Fourier transform* (transformada discreta de Fourier).

A DFT é computada com a seguinte equação:

$$S(k) = \sum_{m=0}^{M-1} s(m) e^{j\frac{2\pi}{M}km}, \quad k = 0, \dots, M-1,$$

Figura 3.11 Expressão da DFT.

Onde M é o número de amostras na sequência de entrada.

O espectro da DFT $S(k)$ é complexo e pode ser convertido em um espectro discreto de amplitude $|S(k)|$ e um espectro discreto de fase $\theta(k)$. $|S(k)|$ e $\theta(k)$ são as amplitudes e fases do exponencial complexo.

$$e^{j2\pi\frac{k}{M}} = \cos\left(2\pi\frac{k}{M}\right) + j\sin\left(2\pi\frac{k}{M}\right)$$

Figura 3.12 Exponencial complexo da DFT.

Onde frequência é kF_s/M Hz. (SERRA, Marie-Hélène, 1997, p. 37)

Podemos entender essas fórmulas do seguinte modo: na DFT, $s(m)$ é uma sequência discreta do sinal; onde m é o número de amostras que vai de 0 a $M-1$ e Σ é a somatória de todos as amostras da sequência. Vemos na segunda fórmula que o

exponencial complexo da DFT pode ser convertido em uma função trigonométrica real - uma soma balanceada de senos e cossenos, indicando valores de 'fase' do par magnitude-fase. Esse par refere-se à amplitude e frequências convertidas matematicamente. Os M pontos de amostragem corresponderão a 'canais' de frequência chamados *bins* que podem ser pensados como linhas individuais que traçarão as variações de frequência no tempo.

3.3.4 *Windowing*¹⁷ e *Spectral Leakage*¹⁸

Sabemos que uma das principais particularidades de um som não artificial é seu extremo dinamismo interno, responsável pela riqueza tímbrica e pela difícil reprodutibilidade de seu comportamento. Mesmo os sons mais estáveis e harmônicos possuem variações constantes, mas não regulares, em amplitude de seus parciais, desvios de frequência, aparecimento e desaparecimento de parciais e frequências transientes durante a evolução do som no tempo, deslocamento dos parciais em relação às formantes criando novos arranjos frequenciais no interior do espectro sonoro, etc. Vem daí um dos primeiros grandes interesses da análise sonora: entender a estrutura do som e relacioná-la com a nossa percepção.

A segmentação do sinal de entrada é fundamental quando se deseja saber 'quando' as variações de frequência de um som ocorrem. A cada segmento corresponde uma porção de tempo com novos dados de frequência e amplitude que nos informam como o espectro sonoro evolui. A cada porção segmentada de sinal de duração fixa chamamos *frame*¹⁹. O problema que surge com essa segmentação é que não há como estabelecer o tamanho do *frame* de acordo com o múltiplo²⁰ do período da onda²¹. Assim,

¹⁷ Utilizar-se-á a expressão em inglês por não haver uma tradução corrente desse termo.

¹⁸ Vazamento espectral.

¹⁹ Também não há uma tradução corrente para esse termo. Pode-se pensá-lo como uma moldura que delimita o tempo do sinal.

²⁰ O período de uma onda seria pequeno demais para se usar como referência para a segmentação. Assim, os múltiplos são os usuais.

²¹ Estamos considerando apenas os casos onde os sons são harmônicos, ou seja, com ondas periódicas, já que, de acordo com o Teorema de Fourier, apenas sinais harmônicos podem ser representados como soma de senóides.

o período seria truncado e a forma de onda resultante da segmentação, isto é, o novo desenho da onda com o final da amostra não sendo contínuo com o início da primeira amostra do segmento, seria interpretada como contendo parciais de alta frequência inexistentes no sinal original. A descontinuidade brusca gerada pelo corte da segmentação será convertida, juntamente com o restante do sinal, para o domínio freqüencial e entendida como resultado da soma de uma infinidade de novos parciais de alta frequência.

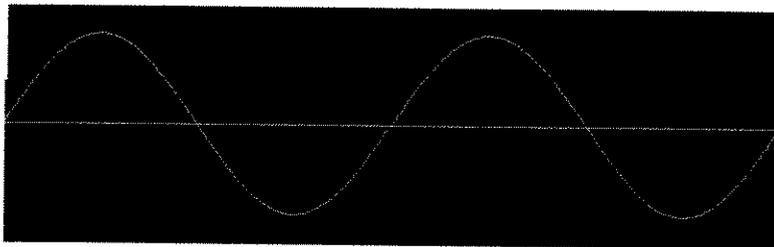


Figura 3.13 Senóide antes de ser segmentada.

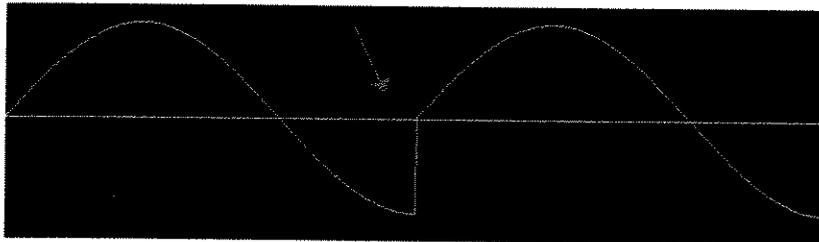


Figura 3.14 Exemplo de senóide truncada pela segmentação
(a seta indica onde ocorreu o corte).

Essas novas frequências, resultado do erro gerado pela incapacidade de se segmentar um sinal de acordo com os períodos de onda envolvidos, se encaixam no chamado ‘vazamento espectral’ (*spectral leakage*). A fim de minimizar o vazamento espectral, atenuações no início e no final do segmento selecionado são efetuadas por meio de ‘funções janela’ (*window function*), processo esse chamado *windowing*.

As *windows* ou janelas de análise são espécies de envelope desenhados para essa função. O significado da *windowing* é relativo a ‘colocar em janelas’, ou ainda, colocar um segmento de sinal sob a forma de uma determinada função com o intuito de atenuar

as discontinuidades geradas pelos recortes temporais. Esse processo é realizado multiplicando-se o segmento de sinal com uma determinada função janela ainda no domínio temporal. O resultado dessa multiplicação é então analisado pelo processo de DFT.

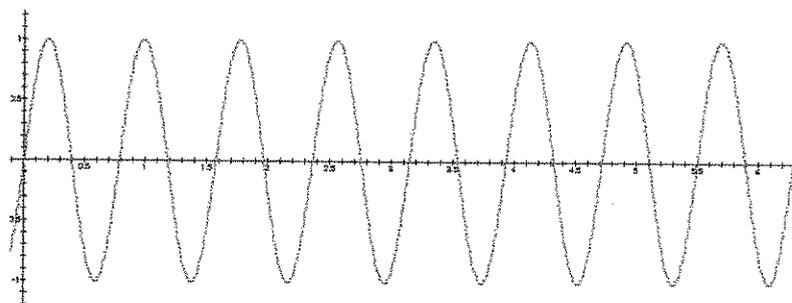


Figura 3.15 Sinal senoidal antes de passar pelo processo de *windowing*.

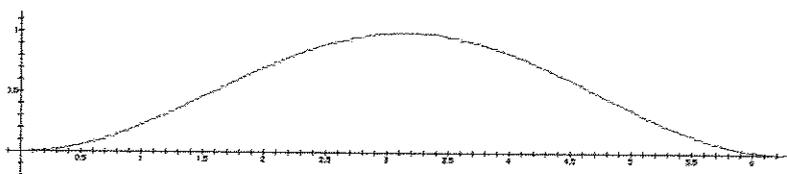


Figura 3.16 Janela de análise von Hann.²²

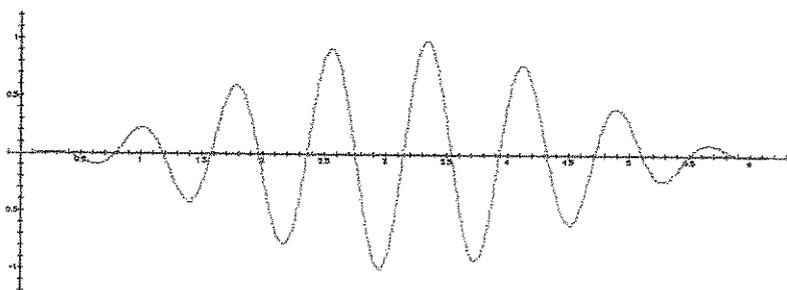


Figura 3.17 Sinal resultante do processo de *windowing* com a janela von Hann.

O tipo de janela padrão utilizado em sinais totalmente harmônicos é a janela retangular, cujo formato isola o segmento sem alterar o sinal de entrada (MASRI, 1997, p.174). Contudo, existem vários outros tipos de janela de análise que modelam o sinal de

²² A definição matemática da janela de análise von Hann a situa no âmbito $-\pi$ a π , mas para tornar o exemplo mais claro a janela foi deslocado em π radianos para que seu novo âmbito seja representado apenas na parte positiva do eixo x, ou seja, de 0 a 2π .

diferentes maneiras devido ao fato de que os sinais raramente são totalmente periódicos e suas segmentações precisam ser atenuadas em suas extremidades. Entre as janelas mais comuns estão a Hamming, a von Hann (também conhecida como Hanning), a Blackman, a Blackman-Harris, a Bartlett ou triangular, a Kaiser e a Gaussian. Podemos representar facilmente esses formatos através de suas respectivas funções. A von Hann da figura 3.16 tem a seguinte definição:

$$y = 0.5 + 0.5 \cos x$$

Estes são alguns dos exemplos de janelas de análise citados acima:

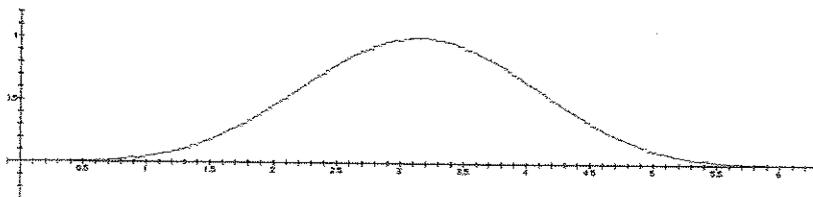


Figura 3.18 Blackman-Harris.

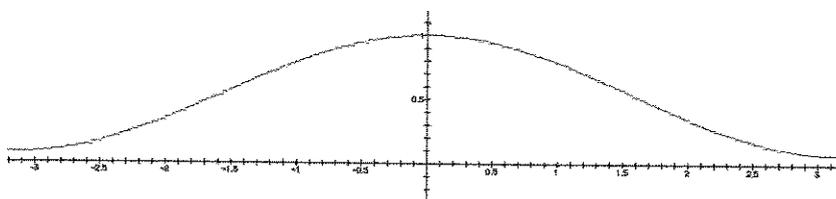


Figura 3.19 Hamming.

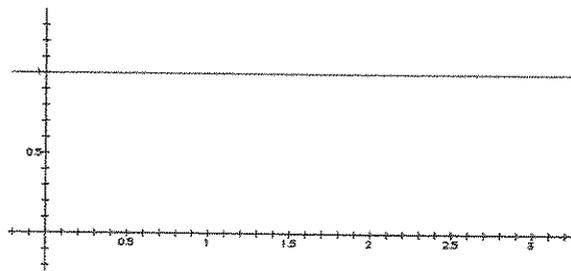


Figura 3.20 Retangular.

Quase a totalidade das janelas de análise possuem um lóbulo central (picos de amplitude da janela) e um decaimento suave em seus extremos, algumas delas inclusive com poucas diferenças visuais perceptíveis. A janela de análise Kaiser é particularmente muito interessante, pois pode ser moldada alterando-se uma série de parâmetros, tornando-a ajustável a diversos propósitos. Seu formato pode variar de uma janela retangular a formas que possuem lóbulos centrais muito estreitos, assemelhando-se a outras funções. Será visto mais adiante o motivo das diferenças no desenho dessas janelas.

3.3.5 Definição Espectral, Resolução Espectral e *Window Size*²³

3.3.5.1 Definição Espectral e sua Relação com a *Window Size*

A DFT se caracteriza por conseguir representar apenas os múltiplos inteiros da frequência fundamental de um segmento dado, ou seja, seus parciais harmônicos. Como dissemos anteriormente, o tamanho estabelecido para o *frame* terá uma consequência direta no tamanho do período de onda que poderá ser lido e portanto na frequência fundamental do espectro do *frame*. Dessa forma, teremos leituras possíveis apenas nas regiões do espectro que sejam múltiplos inteiros da frequência fundamental. As leituras da DFT serão realizadas com espaçamentos ‘constantes’ em frequência e não poderão ‘enxergar’ outras frequências que se encontrem entre os *bins*. O que podemos concluir dessa relação é que o tamanho da janela é diretamente proporcional à definição espectral. Isto é, quanto maior o tamanho da janela, maior o período de onda que se ajusta a ela e assim mais frequências poderão ser ‘enxergadas’. O espaçamento entre os *bins* serão mais estreitos devido aos passos menores que resultam dos múltiplos inteiros de uma frequência fundamental menor. “O raio F_s/M (taxa de amostragem dividida pelo tamanho da DFT) que separa os *bins* de frequência é chamado frequência de análise ou definição espectral da DFT. Ele dá o menor intervalo para a distinção entre dois parciais.” (SERRA, Marie-Hélène, 1997, p. 43). Como exemplo, se se tiver um som amostrado à

²³ Tamanho da janela de análise normalmente medido em número de amostras (*samples*).

taxa de 44.100 Hz com tamanho da janela de análise de 1024 amostras (*samples*), teremos uma definição espectral de aproximadamente 43,07 Hz. Ou seja, só obteremos da análise resultados de frequências que se encontrem espaçadas a cada 43,07 Hz.

3.3.5.2 Conseqüências da *Window Size*

Um espaçamento em frequências de 43,07 Hz não pode ser julgado como uma definição espectral boa ou ruim sem levarmos em conta o contexto: o conteúdo espectral do próprio som e o intuito da análise. Se se tiver um som com frequência fundamental alta, como por exemplo a nota dó mais grave da flauta - 261,63 Hz, só nos interessarão frequências espaçadas a cada 261,63 Hz, se pensarmos que seu espectro tem disposição harmônica dos parciais que são múltiplos inteiros da frequência fundamental. Com o intuito de obtermos uma análise harmônica do espectro, nossa definição espectral de 43,07 Hz pode ser considerada melhor do que o necessário. Mas se o que quisermos analisar do som da flauta forem formantes, que não necessariamente possuem características harmônicas, quereremos 'olhar entre' os espaços dos parciais e daí, quanto melhor for nossa definição espectral, mais chance teremos de realizar uma análise eficaz. Se tomarmos uma formante da flauta de 760 Hz e tentarmos encaixá-la na leitura frequencial a cada 43,07 Hz, os valores que se aproximam da frequência dessa formante situa-se entre o 17º e o 18º múltiplos, isto é, entre 732,19 e 775,26 Hz. Assim, com a definição espectral estabelecida, a formante de 760 Hz da flauta possivelmente seria 'invisível' à análise. Se considerarmos sons mais graves como exemplos, nos quais muito mais sons podem estar presentes em espaços frequenciais menores²⁴, a quantidade de inadequações seria ainda mais numerosa, pois dentro do espaço da definição espectral, ou seja, do menor intervalo para a distinção entre dois parciais, pode estar contida uma oitava justa inteira.

²⁴ Na parte grave do espectro sonoro, pequenas distâncias em frequência podem conter grandes intervalos musicais, pois a correspondência frequencial desses intervalos segue uma progressão geométrica. Um intervalo de uma oitava justa a partir da nota dó mais grave do violoncelo corresponde a 65,4 Hz enquanto que essa mesma medida não alcança sequer meio tom de uma nota aguda de uma flauta.

Outra consequência que o tamanho da janela de análise traz é a relação estabelecida entre imprecisão temporal e o tamanho das janelas. Dentro do espaço de tempo de um *frame*, nenhuma variação é computada, apenas uma média das variações de frequência é dada como resultado. Cada janela de análise permite a leitura de *bins* de frequência fixos dentro de cada *frame*, impedindo precisões de tempo da evolução espectral maiores do que o tamanho do próprio *frame*. No exemplo dado acima de um *frame* com tamanho de 1024 amostras, nossa precisão temporal seria de $1024/44.100$ segundos ou aproximadamente 23 milissegundos. Esse valor não pode ser considerado preciso se levarmos em conta que uma boa parte das informações necessárias para se reconhecer um timbre a partir de seu ataque se encontra dentro desse tempo. Mas *frames* grandes podem ser usados sem perda significativa de informação quando o som analisado for estável o suficiente para não apresentar grandes alterações frequenciais dentro desse espaço de tempo. A vantagem é a melhor definição espectral obtida com *frames* de maior duração.

De acordo com as duas proposições vistas até agora com respeito à dimensão de um *frame* (conseqüentemente de uma janela de análise), chegamos a um paradoxo. Conforme o tamanho do *frame* aumenta, melhoramos a definição espectral, pois tornamos os espaçamentos menores e, de maneira inversa, conforme aumentamos o tamanho do *frame*, pioramos a precisão temporal, fazendo com que percamos todas as alterações internas do espectro dentro do espaço definido de tempo.

3.3.5.3 Princípio da Incerteza

Podemos entender o paradoxo exposto acima de maneira análoga ao Princípio da Incerteza de Heisenberg. O Princípio da Incerteza, uma das pedras fundamentais da física quântica, estabelece que não se pode conhecer simultaneamente a posição e o momentum²⁵ de um elétron com precisão, mas em sua forma generalizada ele se aplica a cada par de variáveis conjugadas, como é o caso da precisão temporal versus precisão

²⁵ “Em física, momentum é uma quantidade física relacionada à velocidade e massa de um objeto.” (MOMENTUM, 2004).

frequencial. Pode-se obter resultados otimizados em cada domínio, mas não na mesma análise.

Considere a seguinte analogia: suponha que você tenha um sinal variante no tempo tal como uma onda sonora e você quer saber as exatas frequências em seu sinal em um exato momento no tempo. Isto é impossível: para poder determinar as frequências com precisão, você precisa amostrar o sinal por algum tempo e você, dessa maneira, perde precisão temporal. (Em outras palavras, um som não pode ter um tempo preciso, como em um pulso curto, e uma frequência precisa, como em um som contínuo puro). O tempo e a frequência de uma onda no tempo são análogos à posição e momentum de uma onda no espaço. (UNCERTAINTY Principle, 2004)

3.3.5.4 Resolução Espectral e sua Relação com a *Window Size*

Enquanto a definição espectral estipula a distância mínima entre frequências que podem ser lidas, a resolução espectral define a largura de cada *bin* permitindo que um parcial seja isolado de outros parciais vizinhos. A resolução espectral depende necessariamente do tamanho e da função da janela de análise. É através do conceito de resolução espectral que podemos entender a necessidade de diferentes formatos de janela, como von Hann, Hamming, Blackman, etc. As diferentes funções janela especificam larguras do lóbulo central e taxas de decaimento de amplitude diferenciadas na tentativa de criar um equilíbrio entre o isolamento de cada parcial e a diminuição de sua energia devido ao formato da janela pela qual ele foi multiplicado no domínio temporal. O equilíbrio a que se tenta chegar é dependente do tipo de sinal que se vai analisar e, mais uma vez, do intuito da análise. Com sons estritamente harmônicos, podemos usar formatos de janela com lóbulos mais largos porque não haveria a possibilidade de interferência de outros sons que se encontram nas proximidades. Em sons não estritamente harmônicos, se quisermos uma boa precisão na leitura de frequências harmônicas, devemos usar formatos de janela que possibilitem os seus isolamentos, a menos que a presença de sons não harmônicos seja desejável. Não estaríamos então falando de análise com o objetivo de retratar fielmente um som, mas sim de artifícios usados para criar modificações a partir do sinal original. Mas quaisquer que sejam os objetivos de uma análise, os tamanhos e formas de uma janela de análise sempre influenciarão o resultado.

Uma outra questão envolvida e que está diretamente relacionada ao formato da janela de análise é o vazamento espectral (*spectral leakage*) que ocorre, não mais pela descontinuidade gerada na onda pela segmentação do sinal, mas justamente pela minimização desse efeito com a aplicação de uma janela de análise. O que ocorre é um efeito colateral do uso das funções janela. A multiplicação de um sinal por uma janela de análise no domínio temporal leva a uma consequência no domínio frequencial: a convolução²⁶ de seus espectros (ROADS, 1997, p. 417). O resultado do processo de convolução dos espectros é o aparecimento de novas frequências responsáveis pelo vazamento espectral. A diferença em haver vazamentos espectrais por não se usar janelas de análise (o que gera descontinuidades na onda) e haver vazamentos espectrais como efeito colateral de usá-las é que o vazamento pelo uso da janela é muito menor se comparado com o vazamento pela não utilização de nenhuma janela.

3.3.6 FFT

3.3.6.1 A Relação DFT – FFT

O que vimos até esse momento foi uma seqüência de processos e recursos matemáticos usados para decompor um som em seus parâmetros constitutivos de amplitude e frequência a cada porção discreta de tempo. Reconstituindo a ordem dos processos e seus resultados temos:

²⁶ Convolução é um processo matemático que leva a amplas consequências no mundo musical por ser a fonte de inumeráveis processos de transformação de um sinal. A convolução envolve multiplicação de todos os segmentos de duas ondas e a soma de seus produtos. Para maiores explicações do processo ver ROADS, 1997.

- segmentação do sinal digital
 - segmentos discretos de tempo (*frames*)
- aplicação de janelas de análise aos *frames*
 - *frames* atenuados pelas janelas de análise
- processamento da DFT em cada *frame windowed*²⁷
 - dados de amplitude e frequência no tempo

Domínio Temporal

Domínio Freqüencial

A FFT (transformada rápida de Fourier), proposta em 1965, não é uma transformada distinta da DFT, mas sim uma implementação da própria DFT com uma adaptação matemática que faz com que essa seja computada somente em *frames* com duração de amostras em potências de dois. Enquanto a DFT pode segmentar o sinal em tamanhos quaisquer, a FFT pode fazê-lo apenas com tamanhos que sejam exatamente potências de dois, como 512 (2^9) amostras, 1024 (2^{10}), 2048 (2^{11}), etc. Mas por que se deveria limitar a escolha do tamanho do *frame* se se pode ajustar melhor seus tamanhos aos períodos de onda desejados sem essa limitação? A resposta não tem cunho sonoro/musical e pode, de início, parecer frustrante: custo computacional. “Cooley e Tukey (1965) demonstraram uma redundância na estrutura algébrica na computação da DFT que poderia ser explorada para acelerar o processo em diversas ordens de magnitude.” (ROADS, 1996, p. 1108) Com a utilização de tamanhos que sejam potência de dois números de amostras, alguns fatores tornam-se redundantes nas equações e podem ser omitidos. Isso faz com que se reduza enormemente a quantidade de cálculos necessários para a execução do processo. Com esse pequeno ‘truque’ matemático pode-se aumentar a eficiência do processo, em *frames* que tenham de 64 a 16.000 amostras, entre 10 e 1.100 vezes (ROADS, 1996, p. 1108). Por isso, toda referência feita anteriormente à DFT vale igualmente para a FFT, já que a transformada é a mesma. Mas a fim de se evitar a grande limitação aos tamanhos dos segmentos, assim como a perda de informações nos inícios e finais de cada *frame*

²⁷ *Windowed* poderia ser traduzido como ‘colocado em janelas’ ou ‘enjanelado’.

devido à atenuação causada pelas janelas de análise, outras técnicas foram criadas, as quais serão discutidas a seguir.

3.3.6.2 Fator de Sobreposição (*Overlap*)

O fator de sobreposição, recebendo os nomes de *overlap* ou *hop-size*, dependendo do ponto de vista, não tem relação apenas com a FFT, mas tem da mesma forma com a DFT, pois este é um recurso da *windowing*. Ele só aparece nesse momento do texto porque o seu cálculo está relacionado com o tamanho do *frame* e como ele, a partir daqui, será sempre tratado como potência de dois, o cálculo do fator de sobreposição será feito sempre como uma fração dessas potências.

Com a atenuação de amplitude causada pela utilização das janelas de análise, ainda no domínio temporal, uma parte das informações contidas no segmento selecionado é perdida e, caso seja de interesse a reconstituição posterior do sinal através de síntese, o sinal resultante conteria muito menos informações que o sinal original. A técnica usada para corrigir este problema é a sobreposição das janelas de análise. Em sua forma primária, sem sobreposição, um *frame* de, por exemplo, 1024 amostras sofreria *windowing* seguido do próximo *frame* de 1024 amostras que sofreria nova *windowing* e assim sucessivamente, criando regiões de ‘vale’ entre os *frames*.

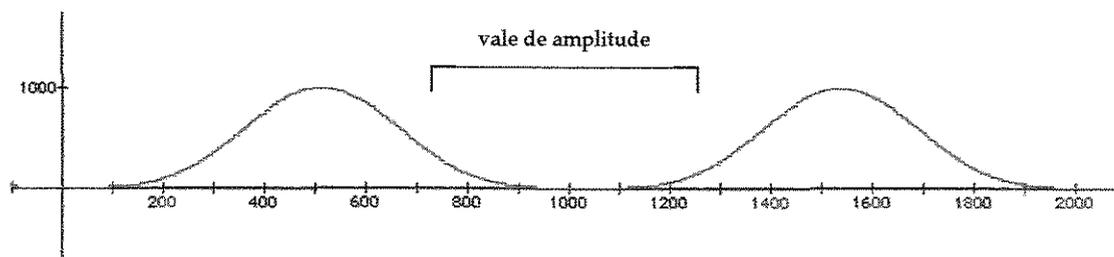


Figura 3.21 Vale criado entre dois *frames windowed*.

Em vez disso, aplicamos várias janelas de análise dentro do tamanho de um *frame* de modo que elas fiquem sobrepostas umas às outras e seu conteúdo possa ser interpolado. A quantidade de sobreposições dentro de um *frame* é chamado *hop-size* e o deslocamento em número de amostras que a janela sofre dentro do *frame* é chamado *overlap size* ou tamanho da sobreposição. O fator de sobreposição, independe do nome que receba, não pode ser arbitrário e depende da função janela utilizada. Os cálculos para essa determinação são complicados, mas as recomendações para as janelas von Hann, Hamming e Blackman-Harris são de no mínimo 4, e sugerido 8. No *software* AudioSculpt, por exemplo, o fator 8 é dado como padrão, apesar de permitir alterações. Assim, retomando o exemplo acima, um *frame* de 1024 amostras sofrendo *windowing* com sobreposição (*hop-size*) de 8, teria a aplicação de uma nova janela de análise a cada $1024/8$ amostras, ou seja, 128 amostras. Os resultados seriam interpolados e a perda de informação entre as distantes janelas seria minimizada. Além de evitar a perda de informações entre as janelas, estaríamos ao mesmo tempo aumentando a precisão temporal. Tomando o exemplo de uma janela de análise a cada 1024 amostras, teríamos uma precisão temporal (considerando uma taxa de amostragem de 44.100 Hz) de aproximadamente 23 milissegundos ($1024/44100$ segundos). Considerando um fator de sobreposição 8, isto é, um deslocamento a cada 128 amostras, nossa precisão aumentaria oito vezes mais – para aproximadamente 2,9 milissegundos.

As sobreposições são um excelente recurso para se evitar a perda de informações entre *frames*. Mas o que acontece com o primeiro e último deles? Não há sobreposição no início e no fim, apenas entre as janelas de análise, o que resulta em uma perda real nestas posições. Onde realmente vai importar haver perdas será no primeiro *frame*, pois é na região de ataque de um som (seu início) que se encontra a maior parte das informações relativas à sua reconhecibilidade. Deste modo, o uso de *frames* longos resulta em maior perda de informação e isso será relevante quando o som analisado tiver ataques de curta duração, como nos sons de caráter percussivo ou ‘atacado’. Um som típico de marimba concentra suas principais informações em seus primeiros 20 milissegundos. Se esse som for analisado com uma janela medindo 4096 amostras usando-se um fator de sobreposição 8, a primeira parte da janela de análise, que não possui sobreposição, isto é, seus primeiros 12 milissegundos, serão perdidos. Isso significa mais da metade das

informações de ataque do som só na parte da janela sem sobreposição. Mesmo com sobreposição, as informações não são necessariamente inteiramente reconstituídas – o segundo passo de deslocamento das janelas de análise possui apenas uma sobreposição e não consegue a reconstituição completa, o que significa mais 12 milissegundos de falha na análise do ataque.

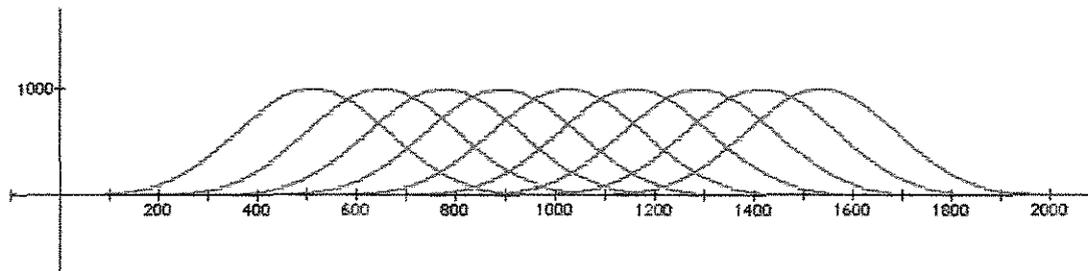


Figura 3.22 Exemplo de sobreposição com fator 8. Como o tamanho da janela de análise é de 1024 amostras, uma nova janela será adicionada a cada 128 amostras ($1024/8$) evitando, assim, perda significativa de informação.²⁸

Depois de um determinado número de amostras, a somatória das curvaturas das janelas permitirá uma muito boa reconstituição das amplitudes o que, depois do sinal convertido do domínio temporal para o freqüencial, resultará em menor vazamento espectral.

Outro efeito colateral evitado pela sobreposição de janelas é a modulação do sinal em amplitude. Em um sinal *windowed*, mas sem sobreposição, o resultado será uma variação contínua de amplitude (com o mesmo formato da função janela) gerando um sinal ondulante diferente do sinal real, caracterizando-se assim como uma *ring modulation*²⁹ (MIRANDA, 1998, p. 131). A *ring modulation*, nesse caso, alteraria drasticamente a composição freqüencial do sinal com adição de novos sons e supressão dos existentes.

²⁸ Na figura 3.22 estão representadas apenas as sobreposições das janelas de análise indicando seus deslocamentos umas em relação às outras, mas não a nova curvatura resultante da somatória das amplitudes individuais.

²⁹ *Ring modulation* (modulação em anel) é um tipo de modulação de amplitude por vezes caracterizado como método de síntese. Para maiores explicações, veja apêndice b.

3.3.6.3 Zero-padding³⁰

Antes do processo de *windowing*, se um segmento selecionado para análise for menor ou maior que um número correspondente à potência de dois, o agrupamento de amostras será rodeado por zeros até o tamanho do *frame* atingir o número exigido. Suponhamos um dado segmento com tamanho igual a 600 amostras. Como a próxima potência de dois é 1024, as amostras que faltam serão preenchidas com amostras de valor zero ao seu redor, isto é, 424 amostras divididas entre o início e o final do *frame* – 212 amostras com valor zero de cada lado do segmento selecionado inicialmente³¹.

Uma característica de destaque no processo de *zero-padding* é o aumento da definição de frequências do espectro. A fim de melhorarmos a detecção de parciais pela diminuição de seus espaçamentos, devemos aumentar o tamanho da janela prejudicando a resolução temporal. Mas se aumentamos o tamanho da janela através de *zero-padding* estaremos aumentando a definição espectral sem que tomemos mais amostras e, portanto, sem que diminuamos a resolução temporal. O tamanho do segmento selecionado será menor que o tamanho da janela de análise e assim manteremos um deslocamento temporal menor e mais preciso.

3.4 Outros Modelos de Análise

3.4.1 Introdução

A FFT, usada normalmente como integrante da STFT, pode ser o grande padrão de ferramenta de análise implementada nos *softwares* de música, mas de longe não é a única. Existe uma série de outras técnicas de análise paramétrica que são usadas em *softwares* muito específicos criados para estudos, pesquisas ou mesmo para aplicações

³⁰ O termo *zero-padding* também não possui uma tradução corrente. Uma possível tradução livre seria 'preenchimento com zero'.

³¹ Há variações acerca do posicionamento das amostras com valor zero, mas estudos mais aprofundados não serão necessários no contexto dessa pesquisa.

em outras áreas como os *softwares* de matemática Matlab e o Mathematica³². Estas outras técnicas de análise na maior parte das vezes ampliam o leque de aplicabilidade das análises a outros tipos de sons que não somente a aqueles caracteristicamente harmônicos. Há técnicas voltadas para o rastreamento de transientes, para sons inarmônicos, ruídos e formantes. Considerando os tipos de sons instrumentais utilizados na música contemporânea, ou seja, sons com características predominantemente não harmônicas, em sua maior parte advindos de técnicas de execução não convencionais, podemos afirmar que estas outras técnicas de análise seriam mais adequadas aos propósitos da composição tal qual fazemos hoje, no nosso tempo. A dificuldade em utilizá-las está calcada na pequeníssima disponibilidade de *softwares* que as utilizam, na falta de divulgação e, por vezes, no seu caráter experimental.

Softwares comerciais com ampla documentação e interface amigável, normalmente intuitivos, acabam por dominar o mercado e mascarar a existência daqueles outros sem intuito comercial e/ou não tão simples de se manusear. Uma característica de *softwares* que empregam interfaces altamente amigáveis com opções prontas de ação é a rigidez de sua estrutura, impedindo que alterações e escolhas possam ser feitas a fim de adaptar o processo a intuítos específicos. No geral, podemos dizer que quanto mais simples o manuseio de um *software*, mais superficial é a sua interação com o usuário. As opções pré-definidas disponibilizadas nos *softwares* possuem agrupamentos de dados ou idéias já incorporadas na ação, ou seja, essas opções induzem o usuário ao direcionamento, pois as escolhas já são pré-determinadas. De maneira inversa, um *software* que não possui ações pré-programadas dá ao usuário a liberdade de ‘compor’ suas ações de modo que essas satisfaçam suas necessidades mais específicas, mas ao custo de ter o árduo trabalho de programar todas as etapas constituintes inerentes à sua idéia musical e o conhecimento para fazê-lo.

Agrupamentos de parâmetros são apenas um exemplo da introdução de níveis de hierarquia em um sistema. Estrutura musical pode geralmente ser pensada como multinivelada e, portanto, um sistema de música em computador que pode endereçar diferentes níveis de estrutura musical será efetivamente uma ferramenta composicional poderosa. E ferramentas poderosas levam a imaginação criativa para novas direções. Entretanto, a introdução de cada nível de hierarquia implementa um modelo musical-acústico e ‘prejudica’ o sistema

³² *Softwares* desenvolvidos pelas empresas “The Mathworks” e “Wolfram Research” respectivamente.

com uma hipótese específica sobre como música é estruturada. Quanto mais alto o nível, maior o prejuízo. O nível mais baixo, por exemplo aquele de amostras de som, não é de maneira nenhuma restrito à música (já que ele pode descrever qualquer som), mas conforme um conceito de mais alto nível é introduzido, ele contribui para um modelo musical específico e requer mais conhecimento específico de música a ser implementado no sistema.

... sistemas de propósito geral, mostrando menos prejuízo porque eles incorporam menos conhecimento, são notoriamente mais difíceis e mais lentos para usar já que o conhecimento que lhes falta (sobre estrutura musical) tem que ser incorporado aos dados providos pelo usuário. (TRUAX, 1986, p. 157)

Das muitas outras técnicas de análise existentes, apenas algumas poucas serão estudadas aqui, pois o interesse em entendê-las está diretamente relacionado com a sua disponibilidade nos *softwares* escolhidos. Questões referentes à escolha dos *softwares* serão vistas mais à frente, no próximo subcapítulo.

3.4.2 Métodos de Análise

3.4.2.1 *Phase Vocoder*

O *Phase Vocoder* não é propriamente uma outra técnica de análise diferente da SFTF, mas um sistema de análise/síntese que emprega completamente a STFT. Por uma série de motivos que serão explicitados em breve, precisamos entender a função do *Phase Vocoder* e suas aplicações.

O *Phase Vocoder* engloba uma seqüência de processos que inclui a análise de um som e sua subsequente ressíntese³³ de tal forma que o resultado final, se não houver alteração dos dados em nenhum momento intermediário, será, por teoria, a reconstituição exata do som analisado. Mas onde estaria a vantagem em reconstituir um determinado som se se possui o som original? A vantagem encontra-se no que se pode conseguir 'entre' a análise e a ressíntese: transformação. A maior parte das transformações obtidas através do uso do *software* AudioSculpt utiliza esse sistema. O diferencial desse sistema é que, em vez de processarmos as formas de onda de um som para conseguirmos suas transformações, o que implica em uma porção de problemas adicionais, alteramos apenas

³³ É chamado ressíntese o processo de sintetizar um som a partir dos dados de análise executado dentro de um sistema análise/síntese.

os dados de análise, ou seja, números, e os resssintetizamos. Um exemplo comum da diferença entre um processo que se dá pela modificação da forma de onda e o outro pela alteração dos dados de análise é a modificação da escala temporal de um som, como o *time-stretching*. No primeiro caso, para alongarmos um som em duas vezes, por exemplo, como não se têm as informações a respeito das componentes espectrais e de seus tempos (tempo de ataque e duração), executa-se uma leitura mais lenta do som, nesse caso, duas vezes mais lenta. O som resultante tem duração duas vezes maior, mas leva consigo um inconveniente: passou a soar uma oitava justa abaixo do original. A leitura do som se tornando mais lenta é equivalente a estirar a forma de onda no tempo fazendo com que ela passe a ter uma frequência duas vezes menor resultando em um intervalo uma oitava abaixo. No segundo caso, no qual usamos o sistema análise/síntese, o que se altera são os dados de tempo extraídos na análise. Assim, a resssíntese é executada com os dados de frequência tais quais extraídos inicialmente e apenas com os dados temporais alterados, não implicando em nenhuma deturpação da ‘altura’ do som resultante em relação ao original. As filtragens também se beneficiam do processamento de dados. Enquanto filtros convencionais, analógicos ou digitais, cortam 20, 30 ou até 40 dB (decibel) por oitava, o filtro de *Phase Vocoder* do algoritmo usado no AudioSculpt pode realizar um ‘corte seco’³⁴ de até 106 dB.

A questão que se deve colocar em pauta aqui é por que usar um sistema de transformação se o interesse está no controle da análise para resssíntese. O sistema *Phase Vocoder* é eficiente na preservação dos dados espectrais do som analisado. É claro que não podemos nos esquecer de que a essência do *Phase Vocoder* é a STFT e, portanto, só responderá de maneira ideal a sons de natureza harmônica e quase harmônica. Se o som que queremos sintetizar possui características que podem ser alcançadas através do uso do *Phase Vocoder*, então não há motivo para não recorrermos a esse sistema.

Uma outra questão relacionada ao *Phase Vocoder* seria o ‘ponto de vista’ sob a qual a STFT é entendida. Há duas formas correntes de representação da STFT: a representação da *windowed DFT* e a representação por banco de filtros³⁵. Apesar de parecerem duas concepções diferentes, suas descrições matemáticas são equivalentes e

³⁴ Sem curva de atenuação, corte em ângulo reto.

³⁵ *Windowed DFT Representation* e *Filter Bank Representation* respectivamente.

portanto geram o mesmo resultado. Normalmente o *Phase Vocoder* é associado com o ponto de vista do banco de filtros e por isso podemos pensar, erroneamente, que se trata de um processo distinto. No escopo do presente trabalho bastar-nos-á ter em mente que o *Phase Vocoder* é um processo análogo à STFT e que se baseia na idéia de ‘banco de filtros’, apesar de haver correspondência matemática entre os processos.

3.4.2.2 O Modelo Determinístico mais Estocástico³⁶

Esse modelo proposto por Xavier Serra e Julius Smith é o desenvolvimento de um processo genérico chamado ‘*sines plus noise*’ (senóides mais ruído) que preserva a parte ruidosa de um som analisado no processo de recomposição do som utilizando diferentes métodos. As principais diferenças entre os métodos empregados está no modo como o ruído é preservado. No caso do modelo de Serra e Smith chamado determinístico mais estocástico, a parte ruidosa, assim como acontece com os parciais estáveis, passa por uma modelização espectral. A modelização dos parciais, a parte mais estável do som, é feita utilizando-se os princípios da FFT que, para isso, são bastante adequados. Já a parte ruidosa ou instável do som não pode ser analisada através da FFT pelos motivos que já foram longamente demonstrados nesse capítulo. Aqui a modelização ocorre fundamentada em outros princípios.

Na utilização do modelo determinístico mais estocástico vários estágios e alguns novos procedimentos são necessários:

- 1- Inicialmente, um sinal de áudio é *windowed* e transformado pela FFT em seu espectro, que irá fornecer magnitude e fase de cada segmento (*frame*);
- 2- A partir da detecção de picos no espectro, obtém-se o *pitch* do *frame*, ou seja, sua frequência fundamental, que irá realimentar os parâmetros de análise ‘ajustando’ o tamanho das janelas de análise ao período da frequência fundamental obtida. Esse processo é chamado de *pitch-synchronous analysis*³⁷. Dessa maneira é possível se fazer uma análise bastante precisa mesmo em sons que possuem grandes variações de frequência ou que sejam pseudo-harmônicos;

³⁶ *The deterministic plus stochastic model.*

³⁷ Não existe uma tradução corrente para esse nome, por isso costuma-se usar a expressão no original em inglês. Seu significado é aproximadamente ‘análise em sincronia com o *pitch*’.

- 3- Aqui um novo processo é utilizado visando a separação do que vai ser considerado ‘determinístico’ e o que vai ser considerado ‘estocástico’: a análise de ‘continuação de pico’. Essa separação é importante para que sejam evitados os efeitos colaterais da análise de FFT de sons não harmônicos: sons não existentes no som original incorporados ao resultado final por má interpretação do que deve ser considerado parcial (estável), transiente e ruído (instável) e sons inarmônicos (estáveis mas não ajustados às faixas dos canais de frequência ditados pelo tamanho da janela de análise). Através de um algoritmo criado com esse intuito (*peak continuation algorithm*) é possível fazer a distinção entre o determinístico e o estocástico e realizar a detecção de frequência, magnitude e fase do que será usado mais tarde como dados para a síntese aditiva;
- 4- Nesse estágio será obtida a parte ‘residual’ do som (que será chamada de estocástica): tudo o que não for determinístico, ou seja, analisável pela FFT. Essa operação se faz por subtração. A parte considerada determinística é sintetizada e é subtraída do sinal original no domínio temporal. O que sobra é o ‘ruído’;
- 5- Só agora será possível modelar o ruído para ser sintetizado mais tarde na recomposição do som. Evidentemente a modelização do ruído não acontece da mesma maneira e nem com a mesma ferramenta usada para os parciais. Aqui a FFT é (sabiamente) descartada.

A importância do ruído se faz presente em inúmeras situações: um som de flauta ou clarinete sem o som do vento que integra o que nós conhecemos dos sons desses instrumentos não é um som realista, pois só uma parte do que identificamos como flauta ou clarinete está presente; as consoantes da fala ou canto **são** o ruído; o ataque de um de piano sem os transientes iniciais gerados pela ‘pancada’ do martelo na corda torna o som praticamente irreconhecível, como Pierre Schaeffer já demonstrou na década de 60 no *Solfège de l’Objet Sonore* (SCHAEFFER, 1967). O ruído pode ser mais facilmente modelado que os parciais procurando-se picos de amplitude associados a regiões de frequência, mas também feito a cada *frame*. A diferença aqui estará no tamanho da janela empregada. Enquanto na análise de parciais é de interesse usar janelas de análise maiores para se obter melhor definição de frequência, na análise de ruído a

situação se inverte e janelas de tamanho menor tornam-se mais adequadas, pois a precisão temporal, nesse caso, é fundamental. Quem costuma trabalhar com processos que envolvam análise de FFT, como no *software* AudioSculpt, por exemplo, sabe que sons com tempos de ataque curtos, de caráter percussivo, são, na maioria dos casos, extremamente prejudicados, pois a definição do ataque se perde e com ele boa parte do interesse do som (dependendo dos parâmetros de análise que se usam). Assim, é importante que as janelas sejam pequenas o suficiente para conseguirmos retratar as pequenas variações de amplitude do sinal sonoro (nesse caso o resíduo).

Com isso, temos parciais e ruídos modelados adequadamente para uma posterior ressíntese. As modificações necessárias podem ser realizadas antes da ressíntese, nos próprios dados dos modelos espectrais, permitindo assim o controle da característica, da qualidade espectral do som resultante.

3.5 Enganos e Mal-Entendidos no AudioSculpt

3.5.1 Introdução

A utilização do *software* AudioSculpt por parte dos compositores (de música eletroacústica ou não) e de pesquisadores de alguma maneira ligados ao tratamento sonoro tem se ampliado muito nos últimos anos, mas nem todos entendem exatamente o limite de atuação do *software* e o que acontece durante a execução de seus processamentos. A documentação vaga e pouco esclarecedora fornecida com o *software* contribui para os mal-entendidos que proliferam com o seu uso. Alguns aspectos referentes à utilização do AudioSculpt que costumam gerar confusão serão abordados nesta parte e os princípios que já foram esclarecidos durante este capítulo serão brevemente revisitados.

3.5.2 Mal-Entendidos

Um dos mal-entendidos mais frequentes diz respeito ao ‘som de AudioSculpt’ aplicado ao som que é transformado, por exemplo, através de *time-stretching* (aumento do tamanho do som) sem alteração de *pitch*. O som resultante incorpora um caráter metálico, às vezes assemelhando-se a um *chorus* que inexistia antes da transformação. É a esse tipo de sonoridade que costumam chamar de ‘som de AudioSculpt’ ou ‘com cara de AudioSculpt’, normalmente associando-o a uma característica do *software*. O que acontece, na verdade, é que a ferramenta é inapropriada para o som que se propõe modificar. O AudioSculpt pode ser visto como a interface gráfica de um *software* de base chamado SVP (de *super vocoder de phase*) que se baseia totalmente nas análises de FFT. Como foi visto neste capítulo, o processo de análise através de FFT não consegue modelar sons ruidosos ou muito inarmônicos com eficiência, pois as variações de frequência de sons ruidosos ou de transientes acontecem muitas vezes dentro de um único *frame* de análise, que por sua vez fornecerá apenas uma única informação de frequência média dentro daquele *bin*. O que era antes um ruído que fazia parte e caracterizava o som original passa a ser entendido depois da análise como uma senóide estável dentro daquele *frame* e ressynetizado dessa forma. Ao contrário do que acontece em muitos outros *softwares* que implementam manipulações de sons, o AudioSculpt não trabalha com o arquivo sonoro propriamente dito no domínio temporal, mas com os dados numéricos da análise feita dele no domínio frequencial³⁸. É por isso que antes de cada processamento abre-se uma janela que permite editar os parâmetros de análise, porque a análise, via FFT, é a base do trabalho do *phase vocoder*, coração do AudioSculpt. O resultado da ressíntese senoidal de sons ruidosos ou muito inarmônicos, isto é, da aplicação de uma falsa estabilidade frequencial em vez da instabilidade do ruído, provoca esse efeito de sonoridade ‘com cara de AudioSculpt’. A única solução para se minimizar esse efeito é a correta configuração dos parâmetros de análise de acordo com as características do som a ser processado. De qualquer maneira, ruídos nunca serão modelados corretamente em

³⁸ Há algumas exceções de procedimentos mais comuns que não fazem uso do sistema de análise/ressíntese próprio do SVP, como edições e a transposição de *pitch* sem correção temporal.

análises baseadas em FFT (baseados nos algoritmos de implementação de FFT que são usados hoje).

Outros enganos ocorrem nas representações visuais das análises, ou seja, nos sonogramas. Os sonogramas podem ser base para processamentos do som, como os filtros executados sobre o sonograma (as ferramentas superfície retangular, superfície poligonal e 'canetas') e para outras análises, como é o caso do *partial tracking* (que gera os dados utilizados neste trabalho). Nestes casos, a escolha dos parâmetros de análise para a criação do sonograma é fundamental, pois serão os mesmos dados usados nos processamentos e análises subseqüentes que usem a representação visual como ponto de partida. Mas alguns parâmetros não são acessados através da janela de parâmetros de análise que se abre na requisição de um sonograma ou de um processamento. Esses parâmetros acessados através da *'edit color palette'* definem o âmbito do que se vê representado graficamente que, por incorporar o sistema WYSIWYG - What You See Is What You Get (o que você vê é o que você obtém), definem também o âmbito do que pode ser processado ou analisado. O âmbito padrão de representação é de 50 dB, -40 dB a +10 dB. Assim, análises e processos dependentes de representações visuais que usam o âmbito de amplitude padrão podem ser muito distorcidos e prejudicados, já que boa parte dos refinamentos e detalhes dos espectros sonoros de muitos sons se encontram abaixo do limite inferior padrão de -40 dB. Os sons resultantes podem ficar bastante alterados e as análises só mostrarão os dados referentes a essa estreita faixa de amplitude, por vezes indicando poucos parciais encontrados. Assim, conseqüências inesperadas e resultados inapropriados não significam necessariamente um problema com o *software*, mas possivelmente com seu uso.

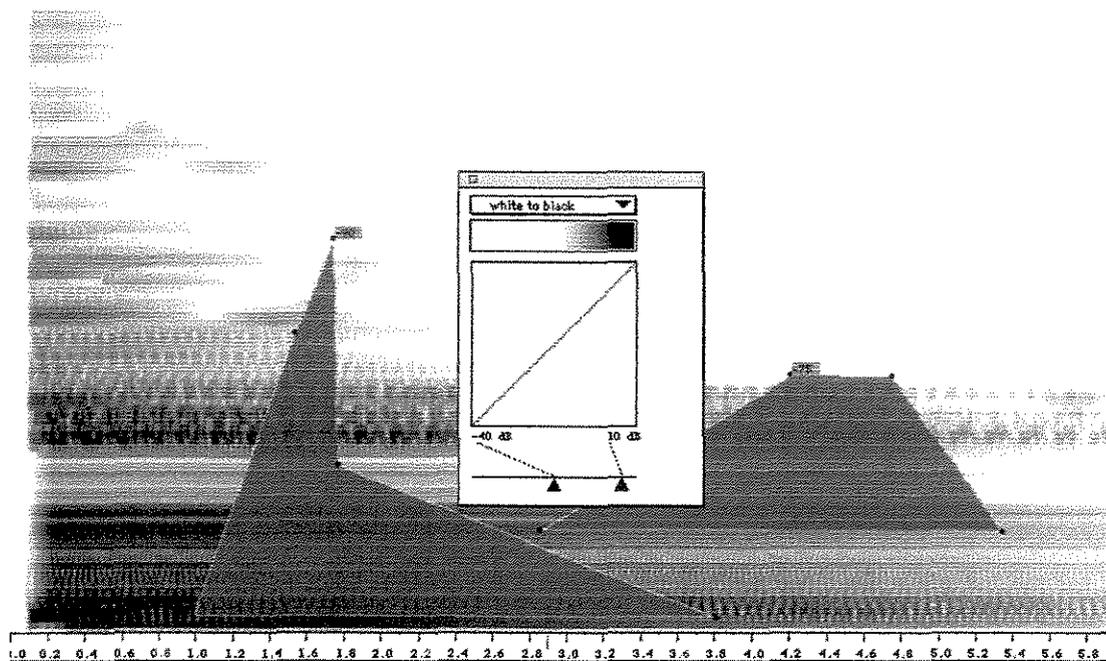


Figura 3.23 Sonograma com a *edit color palette* em primeiro plano.

Uma outra confusão comum referente à representação gráfica de uma análise é a distância que existe entre o que se vê, que é exatamente o que foi analisado, e o que se ouve ou se sabe a respeito do som. A questão agora é de interpretação dos dados exibidos. Na imagem abaixo, por exemplo, na qual se vê representada frequência em Hertz no plano vertical e tempo em segundos no plano horizontal, notam-se espalhamentos das faixas de frequência nos inícios e finais dos sons que, originalmente, são constantes em frequência e amplitude por todas as suas extensões.

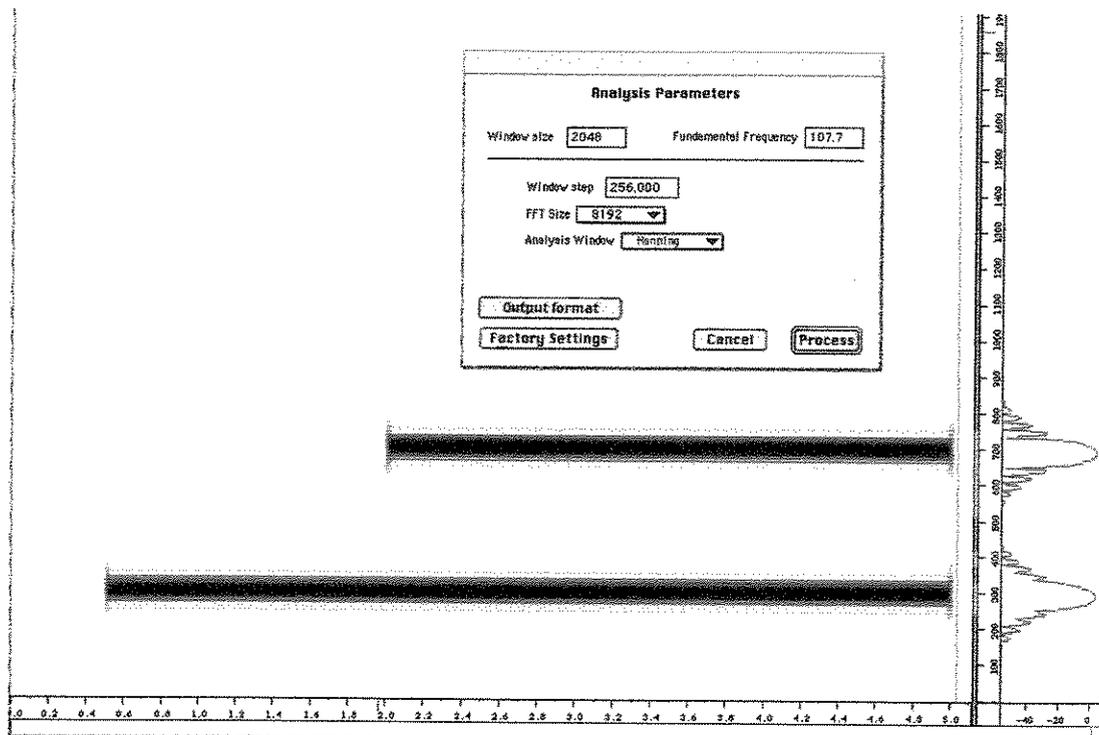


Figura 3.24 Sonograma indicando espalhamento de frequência nos incios e finais dos sons.

Ao contrário do que se poderia esperar, não se trata de erro de representação do AudioSculpt, mas de mais um efeito colateral do processo de análise por meio de FFT. As janelas de análise lidas do meio do som sofrem sobreposição (*overlap*) e podem, dessa forma, evitar descontinuidades nos seus recortes (*frames*), o que não acontece no início e fim do som analisado. Os incios e finais dos sons são justamente os incios e finais das sobreposições que não podem fornecer uma leitura média das informações porque não há nada antes do início e nada depois do final do som. O que acontece nesses pontos é uma mudança no formato da onda por conta da *windowing* e portanto indicação de um conteúdo espectral diferente. Esse é o fenômeno do 'vazamento espectral' que causa o espalhamento das frequências nesses pontos. Portanto, é preciso desconsiderar esses efeitos quando se tenta entender a composição do som, mas considerá-los quando for necessário usar suas análises, já que esses efeitos estarão presentes na forma de dados de novas frequências.

4. Simulação e Alteração de Parâmetros Sonoros para a Síntese

4.1 O OpenMusic

Alguns processos relacionados à criação de variações de um contínuo tímbrico acontecem muitas vezes em ambientes de programação ou de manipulação de dados anteriores ao processo da síntese. Esse ambiente tem que ser adequado às possibilidades variadas de processamento ou criação de grandes séries de dados que servirão mais tarde como parâmetros para a síntese sonora. Para esse fim utilizo um *software* desenvolvido no IRCAM chamado OpenMusic¹. OpenMusic é um ambiente gráfico de programação de auxílio à escritura musical (conhecido como *software* de *Ecriture Musicale Assistée par Ordinateur* – Escritura Musical Assistida por Computador) e faz parte da categoria de “... linguagens e ambientes destinados ao cálculo de eventos e estruturas musicais...” (MALT, 2000, p.190). É possível, neste *software*, construir elaborados esquemas visuais de programação para criação e/ou manipulação de dados chamados *patches*, nos quais é possível trabalhar com aleatoriedade de diversos tipos, expressões e funções matemáticas que vão da aritmética aos fractais e atratores das funções de caos, elaborações puramente musicais que são traduzidas como séries de números, etc. O OpenMusic mostra-se como um espaço aberto à tradução de idéias abstratas em funções, à pesquisa de relações estruturais por meio de séries de valores que as definem, enfim, aos processos de modelização. O OpenMusic é formado por um extenso número de objetos visuais chamados módulos e que são divididos, conforme suas aplicações, em funções e classes². Mas para algumas aplicações menos genéricas, são necessários módulos com funções específicas que não se encontram no *kernel*³ do programa. Com a intenção de prover o *software* com grupos de funções mais específicas em áreas diversas, são criados diferentes pacotes de módulos (*packages*) que são carregados (disponibilizados para uso dentro do programa) conforme a necessidade do usuário. Um desses pacotes, *om2csound*, é o responsável pela ponte criada entre o OpenMusic e o Csound: a formatação e exportação dos dados de maneira que estes sejam entendidos pelo Csound. Outro pacote

¹ O OpenMusic começou a ser desenvolvido no IRCAM em 1996 em continuidade ao *software* PatchWork.

² “Classes são modelos dos quais novos objetos podem ser criados.” (ASSAYAG, 1998, p.15) Classe é um conceito comum empregado em linguagens de programação.

³ O *kernel* refere-se ao corpo constitutivo do próprio programa.

4.2 Criação e Manipulação de Dados para a Síntese Através do OpenMusic

4.2.1 Os Parâmetros Abordados

Foram criados treze *patches* de OpenMusic para fornecer dados que serão interpretados pelo Csound como parâmetros para a síntese sonora⁴. Esses dados constituirão uma parte dos parâmetros necessários para a caracterização completa do som, enquanto a outra parte será resultante da leitura direta dos dados das análises do som e/ou serão fornecidos pelo próprio Csound. A determinação de quais dados virão de onde dependerá quase que exclusivamente das decisões tomadas quanto ao resultado desejado, como foi discutido no capítulo 2. Os *patches* não têm necessariamente uma função única, podendo mesmo abarcar mais que um parâmetro diferente.

Serão descritos agora cada um dos parâmetros trabalhados no OpenMusic, a lógica que envolveu suas construções e como elas foram desenvolvidas.

Os processos de simulação/criação de dados trabalhados especificamente no OpenMusic foram:

No campo frequencial:

- 1) transformação de um espectro harmônico em um espectro inarmônico
- 2) transformação de um espectro inarmônico em um espectro harmônico
- 3) dilatação frequencial do espectro a partir de um âmbito central dado
- 4) contração frequencial do espectro a partir de um âmbito central dado
- 5) redistribuição progressiva exponencial bipolar de frequências sem alteração do âmbito espectral

(Os itens 3 e 4 possuem a mesma lógica e são implementados pelo mesmo *patch* apesar de seus resultados criarem sensações distintas)

⁴ Alguns dos *patches* criados são *sub-patches* e têm apenas funções de conversão ou adaptação de dados para uso conjunto com outros *patches*.

No campo temporal

- 1) micro-alterações do tempo de entrada de cada parcial
- 2) relação variável de duração entre os segmentos do envelope dinâmico dos parciais (proporção interna do envelope)

No campo das amplitudes

- 1) redistribuição da energia contida no espectro
- 2) criação de uma relação variável de amplitude entre o pico máximo e a amplitude de sustentação do som

No que concerne às fases

* criação de posições de fase individuais para os parciais em cinco configurações

1. em fase (a mesma posição de fase para todos os parciais)
2. fases opostas entre parciais consecutivos
3. fases randômicas com distribuição uniforme
4. fases com distribuição gaussiana
5. fases com distribuição beta

No que concerne ao espaço

* distribuição de parciais no espaço estereofônico

Nenhuma dessas abordagens prevê a variação dinâmica dos dados, ao menos neste estágio, pois tecnicamente seria muito complicado e criaria problemas de limitação em relação ao tipo de variação. Se criarmos, por exemplo, um *patch* que crie interpolações entre duas diferentes configurações de fase, como distribuição gaussiana⁵ e randômica uniforme, fecharíamos esta única possibilidade em um *patch* feito para isso. O que quero

⁵ O conceito de cada distribuição será explicado mais tarde.

dizer é que, para cada possibilidade de interpolação entre duas diferentes configurações, teríamos que criar um novo *patch*. Só para as interpolações entre duas fases teríamos que criar vinte *patches*. Agora imaginemos que eu possa querer realizar variações entre três configurações diferentes ou com alguma repetição. Seria inviável. Eu poderia criar apenas uma ou outra interpolação, mas para que contexto musical ela seria útil? Daria para antever todas as necessidades? É evidente que não. E se interpolássemos duas configurações randômicas de fase para empregarmos em um som que dura 1.5 segundo e que possui rápida extinção? Essa interpolação seria percebida? Algumas possibilidades de interpolação poderão ser realizadas muito mais facilmente em Csound através da criação de controladores para mixagem entre sons que possuem diferentes configurações de quaisquer tipos de parâmetros. Pode-se executar diferentes sínteses utilizando parâmetros oriundos de diferentes configurações de fase e interpolá-los facilmente através de um *crossfade*⁶ criado em Csound.

4.2.2 Reformatação e Conversão dos Dados

Um *patch (list-cleaner)* foi construído especialmente para a reformatação dos dados de *partial tracking* importados do AudioSculpt e conversão para as unidades requeridas pelos outros *patches*. A lista de dados chega ao OpenMusic com a descrição dos parâmetros de cada parcial em ordem temporal de aparecimento. Os parâmetros são:

1. número total de parciais
2. número de pontos de leitura de cada parcial
3. tempo de início do parcial
4. frequência em Hertz do ponto inicial do parcial
5. amplitude em dB⁷ do ponto inicial do parcial
6. tempo de término do parcial

⁶ Forma de mixar dois sons seqüencialmente de maneira que o primeiro som diminua de intensidade enquanto o segundo som aumente. A intenção, normalmente, é que haja uma mudança gradual entre os dois sons.

⁷ O âmbito do dB usado no *partial tracking* é dependente dos limites mínimo e máximo determinados no AudioSculpt através da janela *edit color palette*.

7. frequência em Hertz do ponto final do parcial
8. amplitude em dB do ponto final do parcial

Exceto pelo primeiro dado, todos os outros se repetem para cada parcial presente na análise. Já que é feita a média de amplitude e frequência dos parciais pelo AudioSculpt, os dados referentes à frequência e amplitude iniciais e finais são idênticos e não precisarão ser lidos duas vezes.



PARTIALS		208	
POINTS		2	
0.093	101.584		-0.301
0.240	101.584		-0.301
POINTS		2	
0.093	130.027		-7.747
2.145	130.027		-7.747
POINTS		2	
0.093	162.789		-12.794
6.839	162.789		-12.794
POINTS		2	
0.093	195.399		-17.534
6.782	195.399		-17.534

Figura 4.2 Lista de dados exportada pelo *partial tracking* do AudioSculpt.

A duração dos parciais, não fornecida diretamente, é calculada pela diferença do tempo final do parcial com seu tempo de início. A reformatação requerida é a eliminação dos dados desnecessários e a separação de cada parâmetro, como frequência e amplitude, em listas individuais.

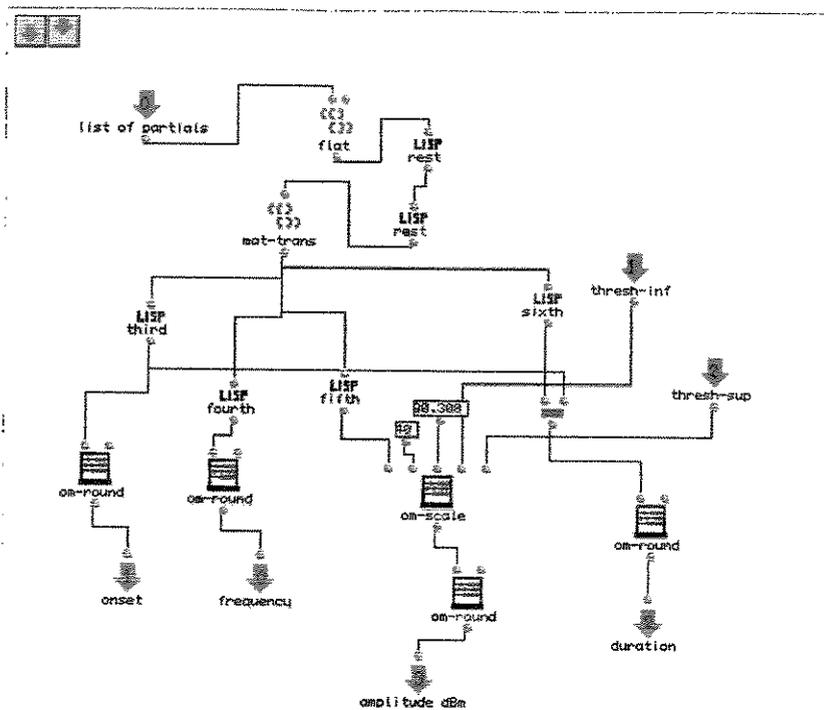


Figura 4.3 Patch *list_cleaner* usado para reformatação e conversão de dados.

As conversões incluem o cálculo da duração dos parciais e a conversão do âmbito em dB (especificado no AudioSculpt) para dBm⁸. O dBm é o tipo de dB possível de ser usado em Csound, que permite fácil conversão para a escala linear de amplitude. Na conversão das amplitudes em dB, utilizam-se os dados de limites de amplitude mínimo e máximo usados pelo AudioSculpt para análise e os limites para os novos valores. Foi assumido um valor mínimo de saída (dBm) de 40 dB e um valor máximo de 90.308 dB que equivalem, em escala linear, a 100 e 32764. Passados por este *patch*, os dados já estão prontos para serem usados na criação e modificação dos parâmetros sonoros usados para a síntese. O trabalho sobre os parâmetros, divididos em categorias, será visto a seguir.

⁸ O dBm utiliza como referência 1mW (miliWatt) em 600 ohms.

4.2.3 Quanto às Frequências

4.2.3.1 A Transformação de um Espectro Harmônico em um Espectro Inarmônico

A transformação de um espectro harmônico em um espectro inarmônico se deu através da construção de um *patch* dentro de um módulo *loop*. O *loop* é um módulo que permite a construção de uma lógica interna ao estilo de um *sup-patch*, ou seja, de um *patch* inserido dentro de outro⁹ e tem a vantagem de permitir recursividade usando outros módulos de uso exclusivo disponíveis dentro dele. Neste *patch* toma-se como princípio que o som de origem, cujos parâmetros se queira modificar, ou para o qual se queira criar, seja harmônico. A lógica consiste em relacionar cada frequência lida da análise a um harmônico e a partir da frequência virtual desse harmônico, criar desvios de até 22% em frequência acima ou abaixo da frequência original, o que corresponde a aproximadamente três semitons mais um quarto de tom (≈ 344 cents¹⁰) para cada lado. Com ‘frequência virtual’ refiro-me à frequência obtida a partir da multiplicação da frequência fundamental (dado extraído diretamente do AudioSculpt em um processo paralelo e inserido no *patch*) pelo número do harmônico correspondente (calculado dentro do *patch*). O porquê ela é virtual e não ‘real’ deve-se ao fato de que os parciais harmônicos de um som instrumental nunca correspondem exatamente à frequência calculada a partir da série harmônica. O piano, por exemplo, apresenta um desvio dos seus harmônicos para o agudo, quer dizer, os harmônicos do piano que se encontram logo acima dos primeiros têm frequências mais altas do que as calculadas a partir da série harmônica. Um exemplo com uma fundamental de 100 Hz é mostrado abaixo (este exemplo não corresponde exatamente aos desvios encontrados no piano, mas mostra como se dá o desvio para o agudo):

➤ série de frequências até o oitavo harmônico (em Hz)

(100 200 300 402 506 612 719 827)

⁹ A vantagem desse recurso é grande e inclui a economia de espaço, a organização, a criação de um banco de estruturas que podem ser prontamente usadas dentro de outros *patches*, etc.

¹⁰ Cent é uma unidade que corresponde a um centésimo de semitom.

Devido a essa característica do espectro dos instrumentos, realizei uma seleção de frequências que se encontrassem entre um semitom abaixo e um semitom acima da frequência virtual do harmônico. Assim, frequências que se encaixam nesse âmbito são direcionadas para o processo seguinte e as que não se encaixam simplesmente não são alteradas de forma alguma. Por quê? As frequências que não se encaixam nesse âmbito podem ser consideradas não harmônicas¹¹ e como o objetivo final deste processo é a obtenção de inarmonicidade, elas não precisam ser alteradas.

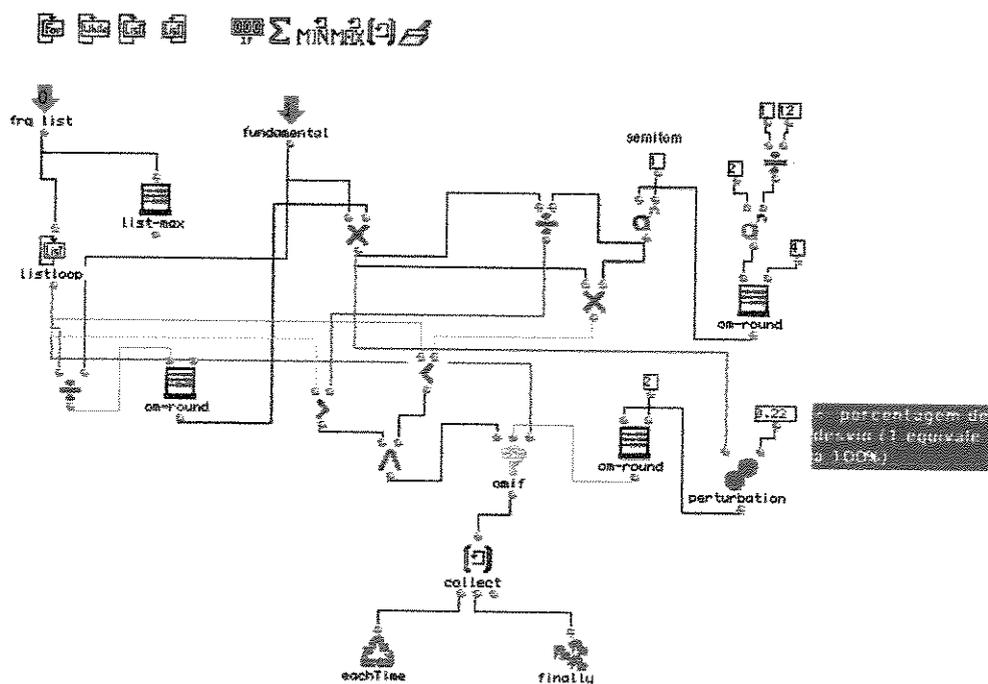


Figura 4.4 Patch para transformação de um espectro harmônico em um espectro inarmonico construído dentro do módulo *loop*.

¹¹ Na verdade isso é discutível. O que faz com que uma determinada frequência corresponda a um harmônico não é a igualdade dos valores, pois, como vimos no caso do piano, valores acima do esperado são considerados harmônicos também. Existe uma margem dentro da qual as frequências podem ser consideradas harmônicas, mesmo não havendo uma identidade entre elas. A discussão deste assunto vai além do escopo deste trabalho, assim, adotei um âmbito largo que abarca de um tom abaixo a um tom acima da frequência virtual do harmônico (visto em seguida).

Nesse próximo processo, as frequências que se encontram dentro do âmbito estipulado passam por um módulo de ‘perturbação’, que irá criar uma variação randômica dentro de uma porcentagem especificada. No caso, a porcentagem de 22% (0.22 no âmbito de 0 a 1) desencadeará a perda da relação harmônica entre os parciais. É claro que, como estamos usando um procedimento de distribuição randômica uniforme (empregado no módulo *perturbation*), alguns valores podem sofrer pequenas alterações e não se desviar muito de uma razão harmônica com a fundamental estimada, mas o que irá importar é que a probabilidade de se criarem relações harmônicas com os outros parciais é menor que se o desvio fosse grande para todos os parciais, o que poderia gerar novas relações harmônicas entre eles. Aqui está um exemplo do resultado da aplicação desse processo sobre as oito frequências listadas acima:

- série de frequências até o oitavo harmônico (em Hz) sem modificação

(100 200 300 402 506 612 719 827)

- série de frequências alteradas através do processo descrito

(100 227.99 360.08 458.12 506.32 601.64 626 802.93)



Som 4.1, 4.2 e 4.3¹² (faixa 1) Som de piano original, versão sintetizada do piano e a versão sintética ‘inarmonicizada’.

4.2.3.2 A Transformação de um Espectro Inarmônico em um Espectro Harmônico

Esta transformação segue princípios similares àqueles explicados acima, mas com algumas diferenças de abordagem.

¹² Os exemplos sonoros contidos neste capítulo são o resultado de processos que vão além dos domínios do OpenMusic, *software* esse que não é apto para realizar síntese, mas para poder mostrar o resultado sonoro concreto dos procedimentos estudados, está sendo utilizado o *software* Csound, que se encontra em uma etapa posterior na estrutura deste trabalho (estudado no próximo capítulo).

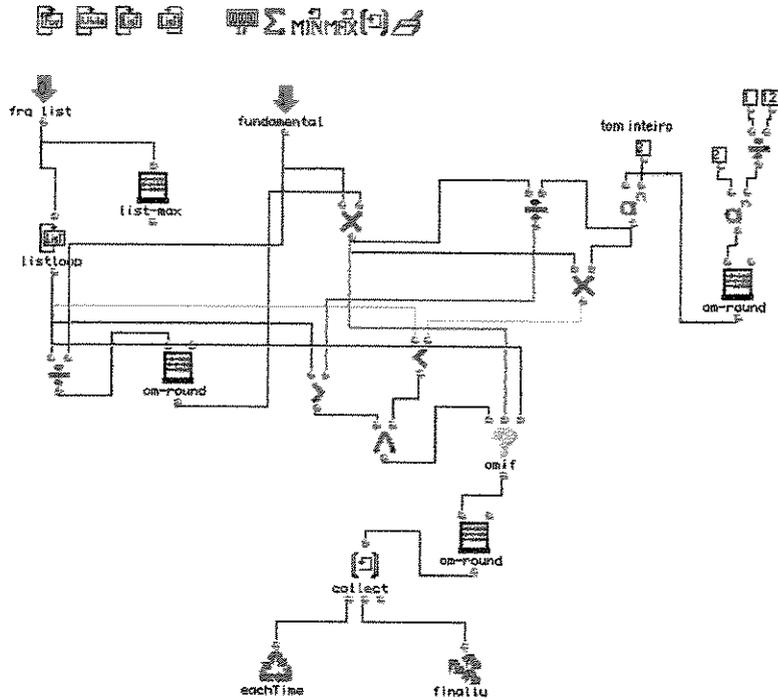


Figura 4.5 Patch para transformação de um espectro inarmônico em um espectro harmônico construído dentro do módulo *loop*.

A lista de frequências extraídas da análise é comparada à lista de frequências harmônicas virtuais, ou seja, aquelas obtidas pela multiplicação de uma frequência fundamental previamente fornecida pela razão, arredondada para número inteiro, entre a frequência original da análise e a frequência fundamental. O arredondamento é essencial, pois a simples razão entre as frequências fornecerá a razão real entre elas, seja esta harmônica ou inarmônica. A razão arredondada, fornecendo um número inteiro, assegura a harmonicidade da relação. Depois de comparadas as listas, se for verificado que as frequências das análises são pertencentes ao âmbito de um tom abaixo a um tom acima da frequência virtual, a frequência virtual é adotada no lugar da primeira, se não, não haverá alteração. Aqui ocorreu uma escolha de procedimento. Poderia-se ter criado um *patch* que transformasse todos os parciais inarmônicos em parciais harmônicos, mas as seguintes considerações foram feitas:

- 1- todas as frequências que se encontrassem dentro de um determinado âmbito seriam transformadas no mesmo harmônico e assim suas amplitudes seriam somadas e destacadas em relação a outras faixas de frequência nas quais isso não acontecesse, distorcendo a distribuição de energia do espectro.
- 2- um espectro resultante que só contivesse frequências puramente e exatamente harmônicas ficaria bastante pobre e poderia simplesmente ser substituído por uma síntese aditiva de harmônicos sem qualquer dependência das frequências originadas da análise.

Assim, resolvi transformar em harmônicos apenas os parciais que se encontrassem dentro do âmbito especificado. Mas há consequências. Se as frequências excluídas do âmbito forem em maior número, dependendo de qual região do espectro elas se encontram e dependendo ainda de qual a amplitude relativa e forma de distribuição desses parciais, o som resultante pode continuar inarmônico. Dessa maneira, esse procedimento de ‘harmonização’ de um espectro não funcionaria totalmente. Digo que não funcionaria totalmente porque, de qualquer forma, o espectro resultante estaria ‘mais harmônico que’ o espectro inicial e essa modificação seria percebida.



Som 4.4 (faixa 2) Som resultante do processo anterior (inarmônico) transformado de volta em som harmônico.

4.2.3.3 Compressão e Dilatação Progressiva Exponencial Bipolar Frequencial do Espectro a partir de um Âmbito Central Dado

A idéia contida neste *patch* é a compressão ou dilatação do espectro de frequências de maneira progressiva e exponencial com fatores de progressão independentes para o grave e para o agudo excluindo deste processo o âmbito especificado. Se for especificado um âmbito que vai de 500 Hz (nomeado no *patch* como *bound_inf*) a 732 Hz (nomeado como *bound_sup*), as frequências vindas da análise que se encontrarem nesse âmbito não sofrerão qualquer alteração. As alterações acontecerão

para além desse âmbito especificado segundo dois fatores fornecidos: o fator para modificações abaixo do âmbito (nomeado como *fator_inf*) e o fator para modificações acima do âmbito (nomeado como *fator_sup*). Os dois têm funcionamento similar e independente, podendo mesmo não haver alteração em um dos ‘lados’.

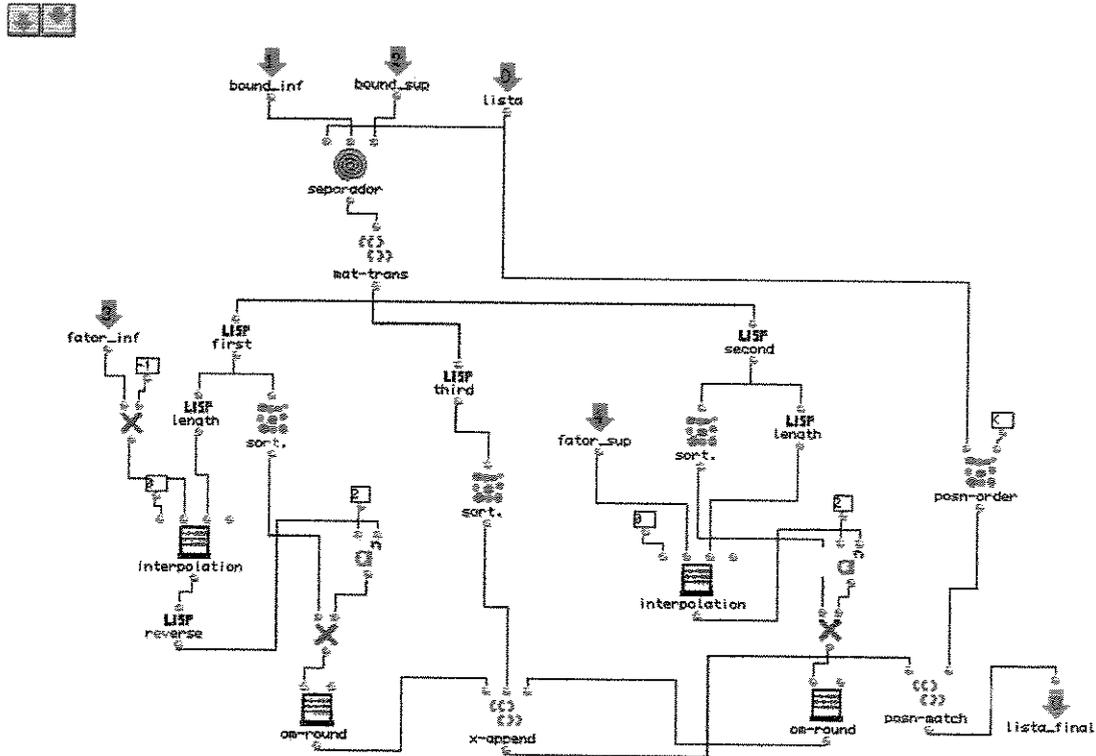


Figura 4.6 Patch para cálculo de compressão e expansão frequencial.

Já que a aplicação dos fatores de alteração progressiva é independente, é necessária a divisão da lista inicial de frequências em três partes distintas: uma lista com frequências abaixo do âmbito, uma lista com frequências acima do âmbito e uma lista com as frequências do próprio âmbito. Essa divisão é realizada por um pequeno *patch* criado dentro do módulo *loop*, aqui chamado de *separador* (representado pelo ‘caracol’ verde na parte superior do *patch*). Abaixo vemos seu interior:

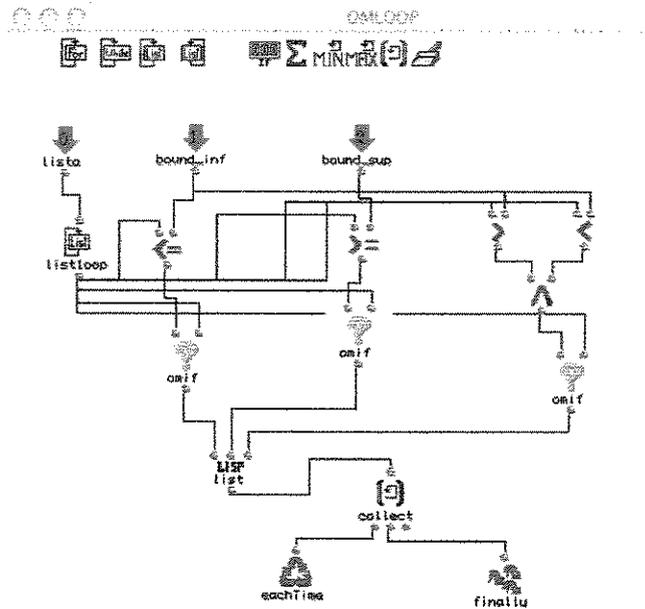


Figura 4.7 Loop usado como separador de listas.

As modificações sofridas pelas duas listas pela ação dos seus respectivos fatores serão progressivas e exponenciais. Progressivas porque, dado um determinado fator (com o valor zero representando a não alteração), será feita uma interpolação numérica entre zero e o fator dado com número de passos de interpolação igual ao tamanho de uma das listas específicas, criando assim uma alteração gradual. Como exemplo, se tivermos uma lista das frequências abaixo do âmbito, apresentada da seguinte forma (em Hz):

(100 200 250 350 400 450 550)

teremos sete passos de interpolação entre zero e o fator dado. Aqui temos que explicitar um pequeno truque para facilitar a compreensão da utilização dos fatores. A princípio, valores positivos correspondem a uma expansão do espectro com relação às frequências, pois seus valores aumentarão (isso será explicado em seguida), mas se temos a parte mais grave do espectro com seus valores aumentados, ou seja, tornando-se mais próximos do âmbito central, isso será visto como compressão e não expansão. Então, para termos expansão real, temos que ter fatores com valores positivos para a parte superior do

espectro e valores negativos para a parte inferior do espectro. Fica parecendo sem sentido para o usuário estipular dois fatores, um positivo e um negativo, para realizar um mesmo processo: a expansão espectral bipolar. Para facilitar, dentro do *patch* é executada uma multiplicação por -1 do fator inferior e então, qualquer valor positivo dado para a expansão frequencial da parte inferior será automaticamente adaptado para se obter os resultados esperados. Se se quiser uma expansão inferior com o fator 0.4 (transformados internamente em -0.4), a lista de fatores interpolados resultantes será:

$$(0 -0.067 -0.133 -0.2 -0.267 -0.333 -0.4)^{13}.$$

Dessa forma se dá a variação progressiva, mas há mais detalhes. O *patch* foi criado com a idéia de haver variação progressiva a partir do centro e se usarmos esses valores acima sobre cada uma das frequências da lista, na mesma ordem, a progressão será reversa. Assim, precisamos reverter a lista de fatores

$$(-0.4 -0.333 -0.267 -0.2 -0.133 -0.067 0)$$

para que haja uma correspondência direta entre a lista de frequências e a lista de fatores interpolados (já que um fator zero significa não alteração e queremos que a frequência mais próxima do âmbito não seja alterada).

E por que e para quê a alteração é exponencial? Para quê é uma questão um pouco mais simples. Nossa percepção de frequência é exponencial (a sensação de intervalos musicais iguais corresponde a uma variação exponencial de frequências) e qualquer alteração dessa forma estaria mais próxima da sensação de variação contínua de som. E o porquê pode ser respondido da seguinte forma: estou empregando esses fatores como expoentes da potência de 2. A lista de fatores interpolados, dessa forma, significa:

$$(2^{-0.4} 2^{-0.333} 2^{-0.267} 2^{-0.2} 2^{-0.133} 2^{-0.067} 2^0) = (0.758 0.794 0.831 0.871 0.912 0.955 1)$$

¹³ Foi usado um arredondamento de três casas decimais.

Esses são os números pelos quais se irá multiplicar os valores de frequência originais, resultando:

100	x	0.758	=	75.8
200	x	0.794	=	158.8
250	x	0.831	=	207.75
350	x	0.871	=	304.85
400	x	0.912	=	364.8
450	x	0.955	=	429.75
550	x	1	=	550

Ou seja, uma expansão progressiva e exponencial do espectro para o grave. O mesmo pode ser feito para a parte superior do espectro com fator diferente. Na verdade, pode-se ter uma expansão em um pólo e uma contração em outro, simultaneamente, deixando um âmbito central inalterado.



Som 4.5, 4.6 e 4.7 (faixa 3) Som de prato tocado com arco, a versão sintetizada e a versão com o espectro expandido.

Os fatores não são limitados em âmbito, mas números altos e baixos demais podem provocar ‘efeitos colaterais’. Expansões muito grandes podem provocar *aliasing*¹⁴ e contrações, utilizando valores negativos muito baixos, podem tornar o som ruidoso e com alta rugosidade. O *patch* foi criado de tal forma que o fator zero não altere os valores, o valor de 1 expanda as frequências para o seu dobro e -1 contraia pela sua metade.

¹⁴ Ver capítulo 3, seção 3.2.6.

4.2.3.4 Redistribuição Progressiva Exponencial Bipolar de Frequências sem Alteração do Âmbito Espectral

Neste caso, diferentemente do anterior, o âmbito de frequências do espectro original é preservado. Em alguns casos pode-se querer, por vários motivos, preservar os limites de frequência. Se se quer, por exemplo, redimensionar as razões entre os parciais de forma a aumentar ou diminuir progressivamente os espaçamentos entre eles, mas a frequência mais baixa do espectro está próxima do limiar de audibilidade, corremos o risco de ter infra-sons no espectro resultante, o que, no caso, não é desejado, já que esses não teriam propósito algum neste trabalho. Assim, há uma redistribuição das frequências do espectro, mas com um reescalonamento das novas frequências a fim de manter o âmbito fixo.

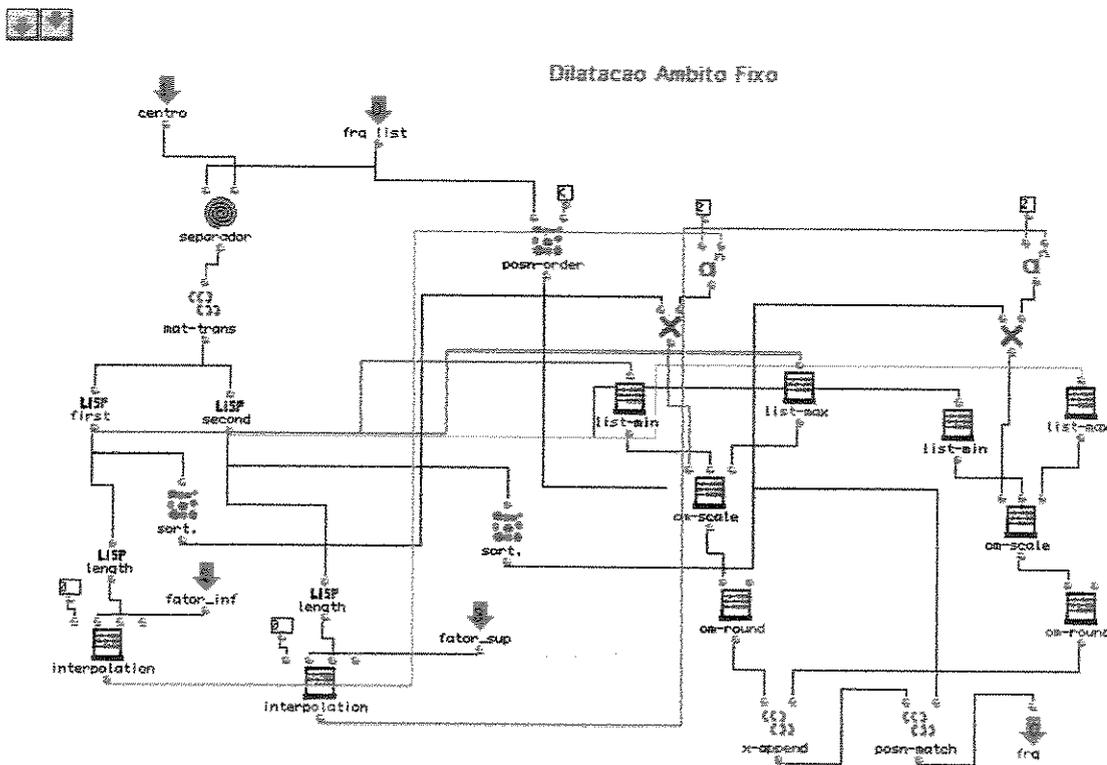


Figura 4.8 Patch para cálculo de dilatação de frequências em âmbito fixo.

As diferenças principais em relação ao *patch* anterior encontram-se no módulo *loop* chamado separador e no acréscimo do módulo *om-scale*. O separador é diferente do anterior, pois não especifica um âmbito, mas um centro, uma única frequência que será o pivô das alterações. Portanto, apenas duas listas são trabalhadas porque não precisa existir uma lista assegurando um âmbito inalterado. Todo o processo relativo aos fatores e interpolações que se seguem tem o mesmo princípio que o *patch* anterior. A diferença entre os *patches* volta a aparecer no uso de um módulo que irá reescalar todas as frequências resultantes de cada uma das listas para o âmbito de frequências original, ou seja, mantendo fixas as frequências limites, a mais alta e a mais baixa do espectro, enquanto preserva a nova distribuição dos parciais.



Som 4.8 (faixa 4) Som de piano com frequências redistribuídas.

Nota-se um deslocamento do *pitch* percebido sem, no entanto, ter-se alterado o âmbito. O único fator alterado é a distribuição das frequências explicada acima.

4.2.4 Quanto ao Tempo

4.2.4.1 Micro-Alterações do Tempo de Entrada de cada Parcial

Um dos fatores que contribuem na identificação e caracterização de um timbre é a relação temporal entre os parciais, se eles começam e finalizam juntos ou não (BREGMAN, 1990). Assim, se se deseja proceder ao distanciamento perceptivo além do que é criado pela margem de erro das análises, pode-se distanciar ainda mais as pequenas distâncias originais ou criar separações de parciais sincrônicos. Para isso foi criado um *patch* que executa micro-alterações no tempo de início de cada parcial de um som analisado.

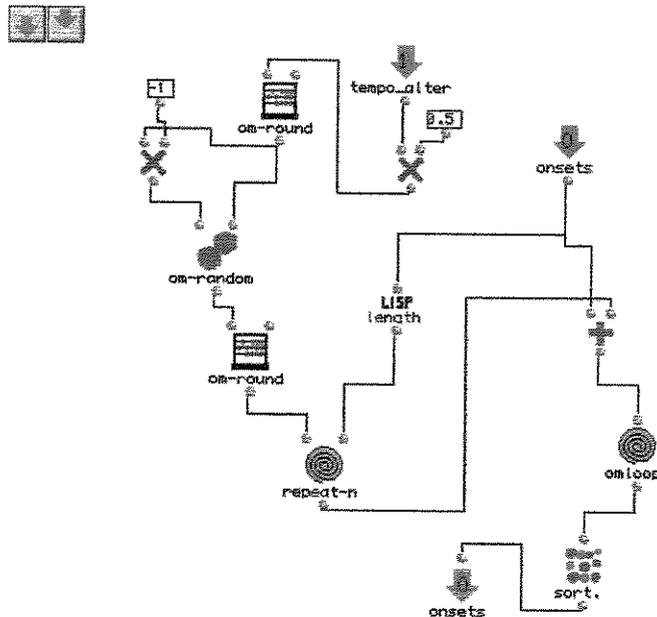


Figura 4.9 Patch que implementa micro-variações randômicas de tempo de início dos parciais.

Neste *patch*, uma lista de tempos de início (*onsets*) dos parciais é transformada através da soma com uma lista de valores de tempo variando de $-x/2$ ms a $x/2$ ms, sendo x o valor total de tempo para deslocamento fornecido pelo usuário. Essa lista de valores, com igual número de elementos em relação à lista original, é produzida através de geração randômica uniforme. Com isso, a lista resultante da soma terá uma variação de até x ms para mais ou para menos da lista inicial. Mas uma prevenção de erro precisou ser implementada, contida no módulo *loop* (*omloop*). O erro diz respeito aos valores negativos que poderiam ser obtidos, por exemplo, na soma de um tempo inicial zero com um valor randômico negativo. O valor negativo obtido na soma não seria aceito como dado temporal e causaria erro. Valores iniciais menores que x ms também poderiam gerar erro se os valores randômicos gerados fossem maiores que os primeiros. No módulo *loop* é verificado se cada dado resultante da soma é maior ou igual a zero. Se assim for, o valor é mantido, se não, o valor negativo é substituído por zero.

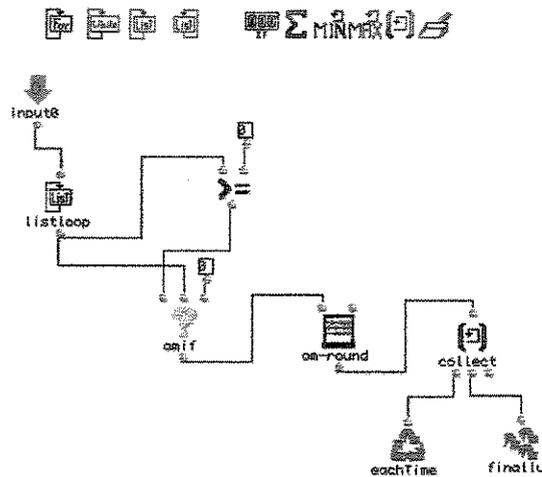


Figura 4.10 Patch que previne a geração de números negativos.



Som 4.9 e 4.10 (faixa 5) Som de piano resintetizado com variação randômica dos tempos de ataque e o mesmo som com os tempos de ataque originais.

No exemplo sonoro nota-se um ataque ainda com caráter percussivo no primeiro som (apesar da perda de definição devido à análise) e a descaracterização do ataque no segundo som.

4.2.4.2 Relação Variável de Duração entre os Segmentos do Envelope Dinâmico dos Parciais

A relação de duração entre cada segmento de um envelope dinâmico, também conhecida como ADSR (*Attack, Decay, Sustain e Release*) é muito importante em um som, pois define todo o seu comportamento dinâmico aplicando-lhe diferentes características perceptivas. Podemos deixar de identificar muitos sons se sua proporção temporal interna for modificada. É claro que há outros fatores em jogo quando nos referimos ao ADSR e à sua importância, como as proporções de amplitude entre cada uma das partes e a maneira como as amplitudes variam no tempo (se de maneira linear, exponencial ou de outra forma qualquer), mas a duração de cada uma dessas partes do

envelope certamente influem no nosso julgamento sensorial acerca do som. Os parâmetros não são percebidos nem julgados separadamente, eles fazem parte de um todo integrado que resulta em uma sensação sonora final, mas a alteração de cada parâmetro separadamente pode levar a diferenças mínimas na percepção de mudança do som e resultar em maior controle no estabelecimento de relações entre som sintetizado e o som do instrumento ele mesmo. Por isso, a proporção temporal do envelope e a relação de amplitude entre os segmentos de ataque e sustentação podem ser designadas separadamente, mas integrados no mesmo *patch* porque eles criam uma idéia única.

O *patch* divide-se em duas partes principais: o *patch* chamado de *adsr-geral* e um *subpatch*¹⁵ chamado *adsr-interno*. Grosso modo, podemos dizer que o trabalho de cálculo das proporções de tempo é realizado no *adsr-interno* e a formatação dos dados para serem entendidos pelo Csound se dá no *adsr-geral*. Mas para saber como criar um envelope, temos que saber como ele será empregado. Um envelope dinâmico não é um simples parâmetro que pode ser enviado para o programa de síntese de modo similar aos outros parâmetros, como frequência e amplitude, além do que, não existe uma única maneira de se criar e aplicar um envelope a um som dentro do Csound. A melhor maneira para se criar e implementar qualquer idéia é aquela que satisfaz a todas as necessidades da maneira mais eficiente possível e com a melhor qualidade. Bem, nada mais genérico e mais verdadeiro, mas é com essa abertura que temos que trabalhar. Precisaremos agora entender o funcionamento geral da criação do envelope que escolhi e por que o fiz.

Criar um envelope em Csound é criar uma variável que descreva numericamente comportamentos temporais e de amplitude. Essa variável pode ser criada de várias maneiras diferentes e em ‘lugares’ diferentes, com menor ou maior controle. Se queremos ou precisamos importar dados de outros programas, como é o caso, eles precisam estar localizados em uma parte do *script* chamada *score*¹⁶, que possibilita o fornecimento de dados individuais para cada evento sonoro através dos *parameter fields*¹⁷ (campos de

¹⁵ Um *subpatch* é um *patch* inserido dentro de outro com a adição de entradas e saídas para comunicação com o *patch* exterior.

¹⁶ Na verdade, existe também outro modo de importar dados em Csound (desconsiderando arquivos de som, pois para sons existem várias outras maneiras), que é através de arquivos texto aos quais pode-se fazer referência em uma *function table*, mas não é tão prático quanto a maneira exposta no texto.

¹⁷ Também conhecidos como *p-fields* ou simplesmente *p*.

parâmetros) ou de dados globais através das *function tables*¹⁸. Na busca de maior controle da síntese através do fornecimento (normalmente) individual e detalhado de parâmetros, espera-se que seja adotada a lógica dos *p-fields*, que através de códigos operacionais (*opcodes*¹⁹) situados em um outro ‘espaço’ do *script* – a *orchestra*, recriarão o envelope dinâmico de cada evento sonoro. A questão levantada aqui é por que criar envelopes individuais se não se conhece o comportamento dinâmico de cada parcial. Baseado em que estaríamos gerando dados de envelope? Não há modelo. O comportamento interno de tempo e amplitude dos parciais não segue um padrão pré-definido no qual podemos nos basear para, a partir do qual, criar variações que possam ser notoriamente de interesse perceptivo. E não haveria então como modelar esse comportamento nas análises? Sim, haveria, e aqui teremos que abrir um outro parêntese para explicar por que isso não foi feito.

Através do *partial tracking* do AudioSculpt conseguimos rastrear frequências e amplitudes a cada *window step*, que corresponde a aproximadamente 6 ms, considerando o tamanho da janela de análise igual a 2048 amostras, *hop size* igual a 8 com taxa de amostragem de 44100 Hz. Isso daria uma descrição bastante precisa, mas talvez precisa demais. A quantidade de dados envolvida normalmente é tão grande que impede o trabalho em OpenMusic em situações convencionais²⁰. O que exportamos das análises, então, é uma média dos valores de frequência e amplitude do parcial, perdendo assim os contornos que nos serviriam de modelo (ver capítulo 3 para maiores detalhes).

Voltemos à situação da criação de envelopes em Csound: podemos enviar dados em *p-fields* relativos à descrição de envelopes individuais ou enviar alguns poucos dados para a criação de *function tables* que dariam descrições globais de envelope. O primeiro caso demonstrou ser impraticável no contexto deste trabalho, por isso será adotada a

¹⁸ *Function tables* são espaços de memória alocados pelo Csound para delinear funções através de uma série de rotinas pré-programadas de geração de onda (*generator routines* ou *GENs*). Como cada função será delineada vai depender da rotina requisitada (*GEN*) e dos parâmetros fornecidos.

¹⁹ Ver capítulo 5 para maiores detalhes.

²⁰ Seria até possível se trabalhar com maior quantidade de dados, mas para isso seria necessário reprogramar o OpenMusic, em linguagem Lisp, para adaptar seu gerenciamento de memória. Mas, o domínio dessa linguagem de programação e a reprogramação do *software* exigiria tanto da grande maioria dos músicos que a via de acesso a essas possibilidades foi descartada por questões práticas.

descrição global²¹, que será rapidamente descrita agora. Na criação de uma *function table* que terá um uso global, tem-se que enviar os dados em forma de proporção entre as partes e não de uma duração absoluta, já que esta irá variar com cada parcial. Deixando de lado os detalhes mais técnicos acerca da criação de uma *function table*, decidiu-se empregar valores que indicam porcentagem na determinação das proporções internas do envelope (valores entre 0 e 100). Segue abaixo uma breve descrição dos *patches* de ADSR.

O *sub-patch adsr-interno* aceita três valores de porcentagem (as setas verdes nomeadas *pct-atq*, *pct-dec* e *pct-sus* na parte superior do *patch*), relativos aos tempos de ataque, *decay* e sustentação do som, que são fornecidos pelo usuário.

²¹ Global, aqui, assume o sentido de um único grupo de parâmetros que será utilizado por todos os eventos/parciais (ou quase todos) em Csound e não que será usado como envelope do som completo resultante.

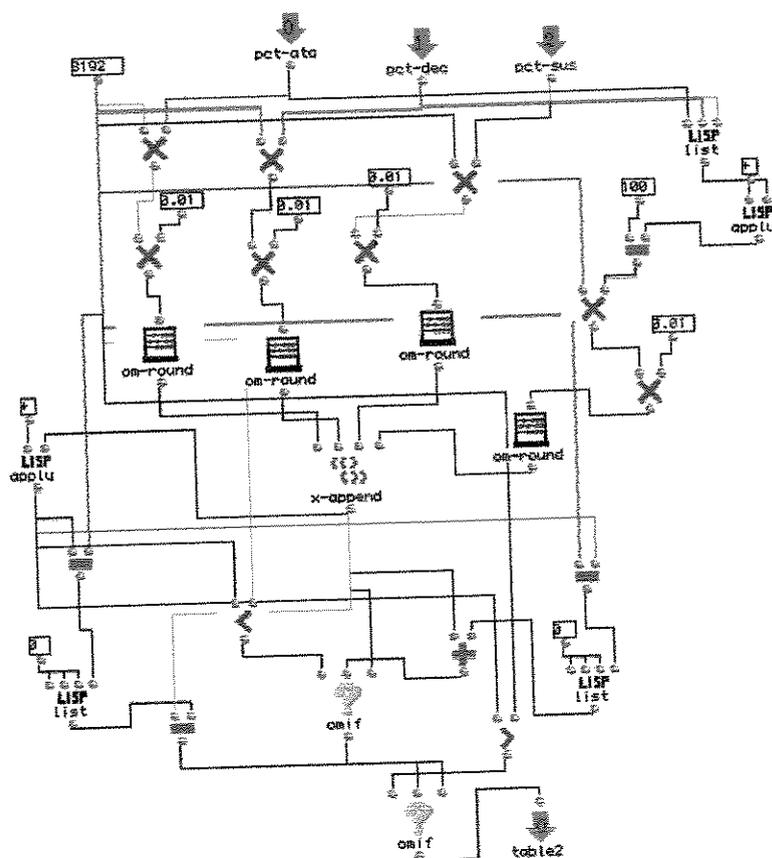


Figura 4.11 Patch *adsr-interno* que calcula as proporções temporais do envelope.

Se definirmos 2% para o tempo de ataque, 7% para o tempo de *decay* e 25% para o tempo de sustentação, obtemos o valor do tempo de extinção, que seria 100% menos a soma dos outros segmentos:

$$100 - (2 + 7 + 25) = 66\% \text{ para a extinção do som}$$

As porcentagens de tempo para cada parte do envelope têm que depender do tipo do som sendo recriado, pois as características referentes ao comportamento dos sons em amplitude e no tempo são drasticamente diferentes. Deve-se, portanto, adequar esses valores para cada tipo de som e intenção composicional. Depois de calculadas as

proporções segundo as porcentagens definidas, esses dados serão enviados para o *patch* externo, *adsr-geral*, para formatar os dados de forma apropriada.

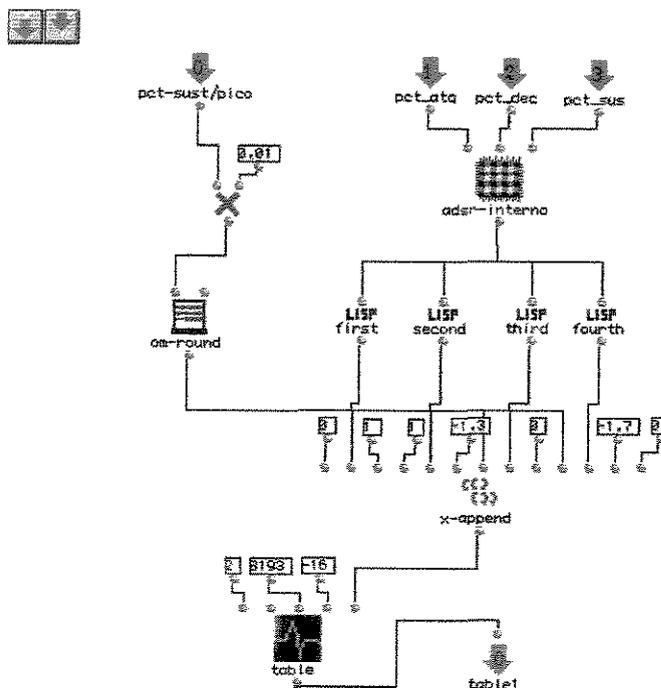


Figura 4.12 Patch *adsr-geral* que formata os dados de envelope para uma *function table* e adiciona informações de amplitude.

Um outro envelope desenhado para os parciais mostrou-se necessário durante os testes. De forma geral, podemos dizer que os parciais têm funções características de ataque e de permanência, ou seja, alguns que têm um pico de amplitude em seu início e uma queda acentuada de amplitude durante o restante do som e outros que têm um comportamento de amplitude mais homogêneo, com um grande período de sustentação²². Assim, um envelope padrão com características mais homogêneas foi implementado e

²² Esta definição é bem genérica e pode não corresponder aos comportamentos dos parciais de muitos sons instrumentais menos convencionais. Está-se tentando, no caso, obter uma descrição geral que possa ser implementada em primeira instância, a fim de que se entenda o processo e, com isso, mostre o caminho através do qual, em casos particulares, se possa definir comportamentos mais adequados e ou precisos.

será selecionado, conjuntamente com o outro envelope criado, para uso no Csound dependendo das condições apresentadas. Como esse envelope é de fácil criação (consiste de um simples módulo *table*) e poderá servir a propósitos variados, ele foi deixado fora dos *patches* específicos.

Até agora, no *patch adsr-interno*, definimos as proporções temporais internas do envelope deixando de lado as formas de variação de amplitude no tempo e as relações de amplitude entre as partes, que serão abordadas a partir deste ponto. As formas pelas quais as amplitudes de cada parte variam no tempo podem ser descritas por curvas desenhadas em um plano cartesiano representando essas variações, com o eixo x correspondendo ao eixo do tempo e o eixo y correspondendo ao eixo de amplitude. Para isso, precisamos escolher uma rotina de geração de onda (*GEN*) que permita determinar qualquer grau de curvatura, seja esta côncava ou convexa. A *GEN 16* permite designar uma curva distinta, com diferentes graus de curvatura, para cada segmento do envelope e será adotada como função geradora de envelope. A formatação da *function table*, portanto, deve incluir valores que designem os graus de curvatura desejados em cada segmento, que é o que é executado pelo módulo *x-append* (figura 4.10). O módulo *table* termina a formatação incluindo, entre outras coisas, o número da *GEN* com sinal negativo, indicando que os valores ali alocados não deverão ser normalizados²³.

Para estabelecer as relações de amplitude, apenas um dado é requerido: a porcentagem de amplitude da sustentação em relação ao pico. Como estamos trabalhando com proporções, consideramos o pico máximo 1. Já que o conceito de ataque é o tempo gasto entre o início do som em amplitude zero até o pico máximo, não é necessário se definir nenhum outro valor de amplitude até aqui. O *decay* vai estabelecer o tempo e a curvatura da variação da amplitude do pico até um valor estabelecido para o segmento de sustentação, onde a amplitude deve permanecer constante. Aqui entra o dado que se deve fornecer ao *patch*. A partir do valor de amplitude da sustentação, o que resta definir é o tempo gasto neste segmento até se iniciar a extinção. No conceito de extinção é demonstrado que nenhum outro valor de amplitude será necessário, pois é definido que neste segmento se dará a interpolação da amplitude estabelecida para a sustentação até o

²³ A normalização corresponde a um processo que irá reescalar todos os valores para o valor máximo de 1 (neste caso).

término do som com valor zero. Portanto, o único valor que importa é o que estabelece a relação da amplitude máxima com a amplitude da sustentação. Esse foi o último dado requerido neste *patch* para a criação de um envelope global em Csound.

4.2.5 Quanto às Amplitudes

4.2.5.1 Redistribuição da Energia Contida no Espectro

A maior parte dos timbres instrumentais contém regiões de acúmulo de energia em relação às outras partes do espectro. Isso se deve, muitas vezes, aos ressonadores dos próprios instrumentos que criam destaques de amplitude em certas regiões de frequência denominadas formantes. Outras causas para os picos de energia podem ser devido à presença de parciais harmônicos muito fortes. Em uma análise dirigida de parciais (como no *partial tracking* do AudioSculpt) ou em uma análise extensiva de FFT podemos notar indiretamente essas regiões através da observação de picos de amplitude em grupos de frequências²⁴.

²⁴ Uma maneira mais direta de observar as regiões formânticas seria através de uma análise de formantes e não através de uma análise de parciais, mas não seria útil dentro do escopo deste trabalho, já que o que é importante neste caso é apenas ter consciência dos conceitos envolvidos.

Frequência em Hz	Amplitude em dB
2858.045166	-33.643990
2883.127197	-32.746689
2895.357666	-7.413737
2906.393066	-0.515075
2918.384766	-4.393331
2843.637207	-23.050436
...	...
3620.799561	-33.696400
3644.832764	-28.815151
3661.353760	-27.204828
3609.220703	-17.417587
3619.496338	-9.254070
3631.100098	-11.752439
3643.834961	-29.290499
3689.509521	-32.706551



Tabela 4.1 Dois excertos de uma análise de FFT de uma nota dó grave de piano. As chaves indicam as regiões de pico.

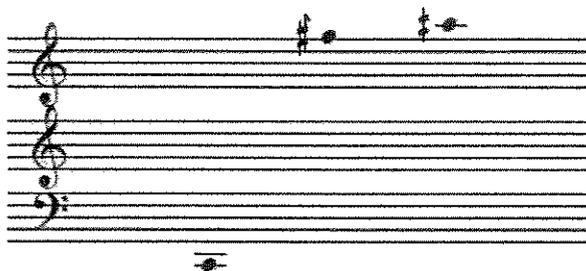


Figura 4.13 Notas correspondentes (na ordem de aparecimento) à frequência fundamental do som de piano e às frequências dos primeiro e segundo picos de amplitude (2906.39 Hz e 3619.5 Hz)²⁵ mostrados na tabela acima.

Ao lado das regiões de pico de amplitude, temos grandes regiões de frequência que não chegam a atingir a amplitude mínima suficiente para serem percebidas e há dois motivos

²⁵ A clave de sol superior representa as notas duas oitavas acima da clave de sol convencional, como se fosse a continuação da primeira clave em direção ao agudo. Quanto aos acidentes, o da nota fá é um sustenido aumentado de um oitavo de tom e o acidente da nota lá equivale a um quarto de tom acima da nota. As duas notas mais agudas são aproximações em oitavo de tom das frequências destacadas na tabela.

principais para isso. Um deles diz respeito ao nosso limiar de audibilidade²⁶ (que será visto em detalhes no capítulo 5), ou seja, essas frequências possuem amplitudes que se encontram abaixo do contorno de *equal loudness*²⁷ e não podem ser ouvidas não importa o contexto onde elas se insiram. Outro motivo é o mascaramento espectral (já discutido no capítulo 2) que impede a percepção, em condições muito específicas, de um componente do som pela proximidade frequencial de outros componentes espectrais com maior intensidade. Estes poderiam ser percebidos, mas a sua relação de amplitude e frequência com outros componentes os mascaram. É sobre este ponto especificamente que se situa o campo de atuação do *patch* de alteração de amplitude.

A idéia desenvolvida aqui é trazer à tona uma série de componentes do espectro que não poderiam ser percebidos antes devido ao mascaramento espectral. A esse processo chamei **desvelamento tímbrico**, pois revela uma parte do timbre antes encoberta (mascarada). Mas, antes de apenas elevar as amplitudes dos componentes imperceptíveis (ou menos imperceptíveis que outros), aumentando a energia sonora global, foi realizado um deslocamento ou redistribuição de determinadas ‘quantidades de amplitude’ com valores fornecidos pelo usuário. Dessa maneira, pode-se, em parte, manter o contorno de amplitude e a soma de energia do espectro do som original. O que poderá mudar, além da alteração de regiões frequenciais percebidas, é a sensação de intensidade resultante²⁸ devido a um grande número de fatores psicoacústicos em jogo.

²⁶ Ou *Audible Threshold of Hearing – ATH*.

²⁷ *Equal Loudness Contour* - curva que “... descreve a sensibilidade das características de frequência de nosso sistema auditivo.” (SUZUKI, 2004)

²⁸ *Loudness*.

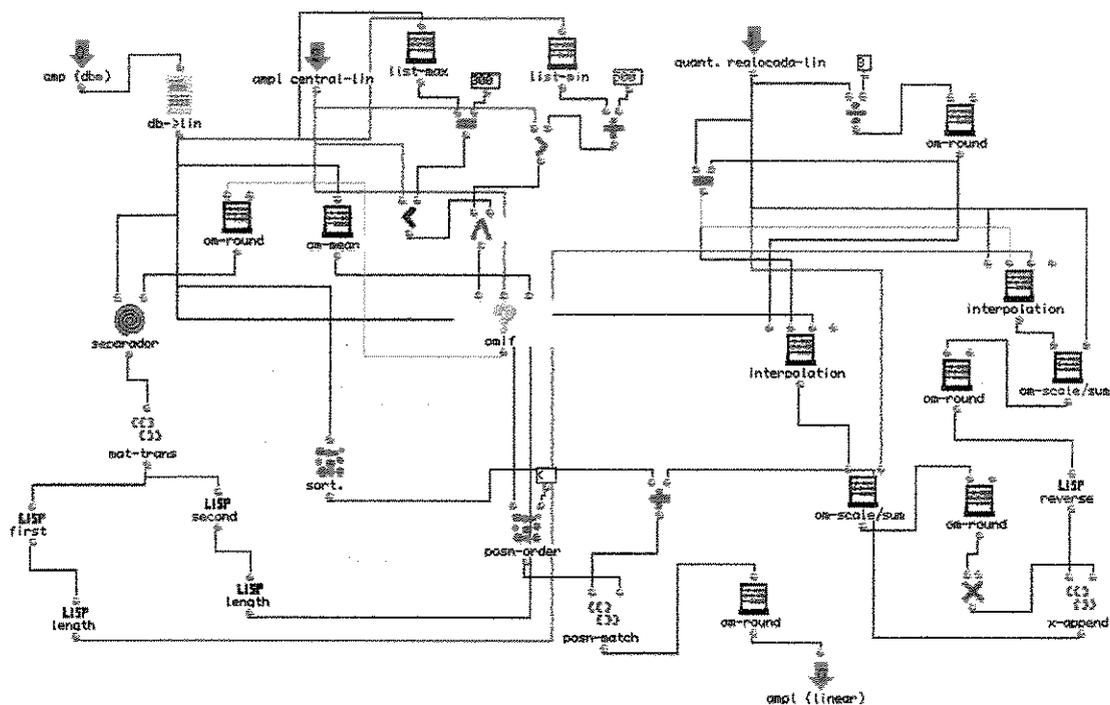


Figura 4.14 Patch de redistribuição de energia do espectro.

Como explicado acima, a redistribuição de energia do espectro foi implementada através da subtração de uma certa quantidade de amplitude dos parciais mais intensos e a sua realocação para os parciais menos intensos, mas não de forma linear. O objetivo foi executar a subtração progressiva de amplitude dependente da hierarquia de 'força' do parcial dentro do espectro. Dada uma lista de amplitudes de um som, tem-se que determinar uma amplitude de centro acima da qual os parciais terão suas amplitudes diminuídas e abaixo da qual os parciais terão suas amplitudes aumentadas. Se a amplitude de centro fornecida for muito próxima dos limites inferior e superior da lista analisada, uma amplitude correspondente à média aritmética de amplitude dos parciais do espectro será então usada no lugar daquela fornecida. Separada a lista inicial em duas outras – aquela do qual se retirará amplitude e aquela para a qual será adicionada amplitude – aplica-se a seguinte lógica: será subtraída duas vezes mais energia dos parciais que forem mais intensos do que dos menos intensos, com interpolação linear de subtração. Esse total de amplitude também será acrescentado à outra lista com relação de proporcionalidade:

os parciais menos intensos recebem duas vezes mais amplitude do que os mais intensos, com interpolação linear entre os valores, de modo que a soma total dos acréscimos corresponda ao valor estipulado para realocação. Tomemos um exemplo. Dada uma lista de amplitudes²⁹

(6000 100 300 10000 1000 3000 600)

com valor de centro **1015** e uma quantidade de realocação **100**, acontecem os seguintes passos:

1) colocá-la em ordem crescente

(100 300 600 1000 3000 6000 10000)

2) dividir em duas listas segundo o valor de centro

(100 300 600 1000) (3000 6000 10000)

3) calcular a quantidade a ser subtraída de cada item da segunda lista de forma que o valor retirado da maior amplitude seja duas vezes maior que o valor da menor amplitude

(2977 5965 9956)

(retirado **23** do primeiro item contra **44** do último – uma relação aproximada de $2/1$ ³⁰)

4) calcular a adição proporcional à outra lista

²⁹ Os valores de amplitude usados empregados neste *patch* se encontram na escala linear tal como usados em Csound. Devido ao fato de que estamos trabalhando com adições e subtrações de valores, a unidade dB não seria conveniente por se tratar de uma relação logarítmica. O âmbito da escala linear vai de 0 a 32767, pelo motivo que foi explicado no capítulo 3.

³⁰ A imprecisão acontece porque os cálculos de divisões incluem frações com muitos decimais e dízimas periódicas que precisam ser arredondadas.

(133 328 622 1017)

(o acréscimo de 33 na menor amplitude é praticamente o dobro do acréscimo de 17 no maior valor da lista)

5) reordenar a lista segundo a ordem original das amplitudes a fim de manter a correspondência em relação aos outros parâmetros do som

(5965 133 328 9956 1017 2977 622).

A soma de todas as amplitudes da lista original é de 21000, enquanto a soma da lista resultante é de 20998 – um erro aproximado de 0.01%. Podemos seguramente considerar esse desvio desprezível. Os caminhos traçados neste *patch* são os descritos acima (sem levarmos em conta os detalhes lógicos e de cálculo omitidos), mas o que fomentou a criação da estrutura lógica deste *patch* foi a intenção de destacar componentes do espectro que não são percebidos ou que não são dominantes na composição perceptiva do som³¹. Para isso, foi necessário o cálculo de realocação proporcional à intensidade de cada parcial dentro do espectro, de forma que os parciais com menor amplitude pudessem ser mais destacados do que os outros. Sem a proporcionalidade implementada, haveria a tendência de que os parciais mais fracos continuassem a não se destacar o suficiente para serem percebidos ou não serem percebidos de forma significativa.



Som 4.11, 4.12 e 4.13 (faixa 6) Som original de harmônico de piano, o mesmo som resintetizado e a versão com redistribuição da energia do espectro.

4.2.5.2 Criação de uma Relação Variável de Amplitude entre o Pico Máximo e a Amplitude de Sustentação do Som

Já visto no sub-capítulo 4.1.4.2, pois é parte integrante do *patch* de ADSR.

³¹ Para esse processo de redistribuição de energia ser efetivamente percebido, é necessário que a leitura de parciais no AudioSculpt alcance o limite inferior de -60 dB a -70 dB, determinados na janela *edit collar palette*.

4.2.6 Quanto às Fases

São propostas cinco diferentes configurações de fase para os parciais, cada uma delas com arranjos de valores bem diversificados. Por questões de dificuldade de gerenciamento de grandes quantidades de dados, similares àquelas dos envelopes dos parciais, não se têm as leituras de fase originais dos componentes espectrais, elas não seriam possíveis com análises de parciais via *partial tracking*, só seriam através de uma análise extensiva de FFT, o que elevaria enormemente a quantidade de dados a serem manipulados, inviabilizando o processo dentro dos recursos disponíveis para este trabalho. Cada uma das configurações será agora explicada.

4.2.6.1 Em Fase

Esta configuração é a mais simples, mas não necessariamente a mais próxima de uma situação real. As fases dos parciais seguem configurações muito particulares dependentes de uma infinidade de fatores que impedem a geração de uma modelização fiel, mas sabe-se que o posicionamento de fase igual para todos os parciais é inverossímil. Por que então se propõe o uso desse tipo de configuração? Como a modelização fiel é praticamente impossível, optou-se por buscar alterações de fases de modo que se obtivesse diferentes situações perceptivas com as quais o som real pudesse se relacionar, criando diferentes tipos de interações.

A expressão ‘em fase’ significa que uma onda possui a mesma posição de fase que outras ondas que ocorrem no mesmo tempo que ela. As unidades de fase podem variar entre graus, π radianos ou um determinado âmbito envolvendo um número absoluto e é este último caso que se enquadra com as necessidades da síntese que será usada (unidade requerida pelo *opcode* de Csound *oscil3*). As fases, neste caso, irão variar de 0 a 1, correspondendo aos 360 graus do círculo trigonométrico. Como todas as configurações de fase foram implementadas no mesmo *patch*, veremos alguns recortes relativos a cada configuração de fase. Esta primeira configuração é bastante simples e não precisa ser visualizada.

Para gerar valores de fase iguais para todos os parciais, basta contar quantos são os parciais e enviar uma lista de igual tamanho contendo zeros, que serão interpretados como valores de fase pelo *oscil3*. Pode-se levantar uma questão quanto ao significado real dessas fases em parciais que não se iniciam ao mesmo tempo e muitas vezes não se sobrepõe de maneira alguma. Iniciando-se sincronicamente ou não, as fases dos parciais podem interagir entre si e dar margem à criação de efeitos perceptivos, como rugosidades e alterações de amplitude (entre outras coisas, devido a cancelamento de fases) e portanto gerando alterações no resultado percebido, à exceção de parciais que eventualmente acontecessem isolados no tempo de qualquer outro parcial, não criando efeito de qualquer natureza³².

4.2.6.2 Fases Opostas entre Parciais Consecutivos

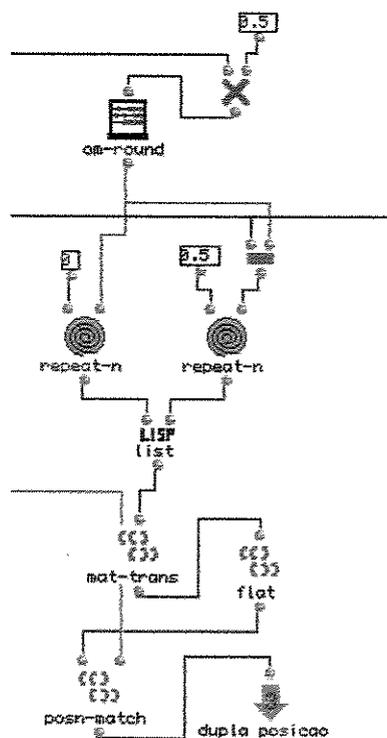


Figura 4.15 Trecho de *patch* que implementa a geração de fases opostas.

³² Nosso sistema auditivo não é sensível à diferença de fase de sons isolados.

Este esquema coloca os parciais em duas posições distintas: 0 e 0.5, ou seja, fases opostas. Aqui surgiria um problema: quais parciais teriam fase 0 e quais teriam fase 0.5? Poderíamos adotar várias referências: quanto à ordem de aparecimento dos parciais (assim teríamos fase 0 para o primeiro deles, fase 0.5 para o segundo, fase 0 para o terceiro e assim por diante, independente em que posição frequencial no espectro eles estariam), quanto à amplitude, etc. A referência adotada foi a ordem crescente de frequências, por isso é necessária a conexão da lista de frequências na entrada deste *patch*. Dessa entrada são analisados dois fatores: tamanho da lista (para determinar quantos números de valores de fase terão que ser gerados) e a posição de cada frequência dentro de uma ordem crescente. Se, por exemplo, tivermos uma lista

(300 500 100 200 400)

a lista de posições segundo a ordem crescente será

(2 4 0 1 3)³³

A partir dessa lista de posições podemos associar os valores de fase tais como gerados com a ordem crescente de frequências (o módulo leitor de posições (*posn-order*) e o medidor do tamanho da lista (*length*) estão situados em outra parte do *patch*, próximo da entrada dos dados). Quem executa o trabalho de reconstituição das posições é o módulo *posn-match*.

4.2.6.3 Fases Randômicas com Distribuição Uniforme

As três propostas de configuração de fase que exponho a seguir são baseadas em funções de distribuição estatísticas. Essas funções dão a uma variável a probabilidade de assumir um determinado valor dentro de um âmbito especificado, normalmente 0 a 1. Existem curvas que descrevem maiores ou menores probabilidades das variáveis

³³ As posições são numeradas de zero a $n-1$, sendo n o número de elementos na lista.

assumirem determinados valores dentro do âmbito. Neste nosso caso, a distribuição é chamada randômica uniforme porque distribui uniformemente valores randômicos entre 0 e 1^{34} .

(0.0155 0.1177 0.7659 0.5414 0.4755 0.2879 0.3051 0.1689 0.8119 0.7025 0.2556
0.1569 0.9527 0.7204 0.5448 0.6434 0.1193 0.1883 0.4707 0.5208)

Exemplo de números randômicos gerados através de distribuição uniforme.

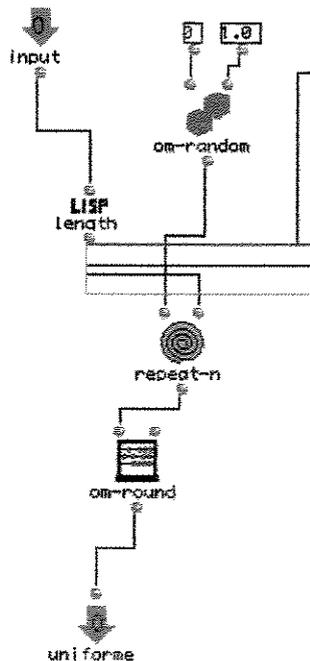


Figura 4.16 Trecho de *patch* que implementa a geração de fases randômicas uniformemente distribuídas.

A geração de números randômicos uniformemente distribuídos se dá através do módulo *om-random* que executará essa tarefa um número de vezes equivalente ao tamanho da lista de entrada. Arredondamentos fazem-se sempre necessários e são fixados em duas

³⁴ Os números randômicos gerados no computador não são verdadeiramente randômicos, são pseudo-randômicos, ou seja, são simulações de escolhas randômicas verdadeiras. “Diz-se que uma seqüência de números é randômica se o conhecimento da seqüência até aquele ponto não fornece nenhuma dica de qual será o próximo número na seqüência. Infelizmente, quando nós estamos usando computadores, nós somente podemos criar uma seqüência de números através de um algoritmo ou programa e esse é um processo determinístico. O próximo número pode ser extraído do programa e, assim, a seqüência não pode ser verdadeiramente randômica.” (BOULANGER, 2000, p. 330)

casas decimais para as fases (a precisão no OpenMusic é de 16 casas decimais). Aqui não vai importar nenhum tipo de correspondência com outros parâmetros, pois nenhuma situação alteraria o caráter randômico da lista de fases gerada. O intuito de uma distribuição randômica uniforme é não criar agrupamentos de dados de fase em nenhuma região do espectro ao mesmo tempo em que fornece uma maior variedade de relações de fase entre os parciais. Um problema que pode ocorrer é o diferente resultado da forma de onda resultante, podendo gerar algum desequilíbrio na sensação de intensidade final. Podemos ver abaixo três formas de ondas diferentes relativas ao som resultante da somatória de fases de três listas randômicas uniformemente distribuídas. O som é composto de 15 parciais harmônicos com frequência fundamental de 300 Hz e amplitude exponencialmente decrescente conforme a frequência aumenta³⁵, a única diferença de parâmetros são as fases usadas.

³⁵ Os parciais são sincrônicos e suas frequências exatas, em Hertz, são (300 600 897 1203 1509 1800 2091 2385 2718 3027 3300 3597 3873 4212 4470), com as correspondentes amplitudes lineares (3979 3418 2910 2454 2050 1694 1386 1125 908 733 599 501 439 406 398). As figuras representam 5 ms de cada som a partir de 1 segundo.

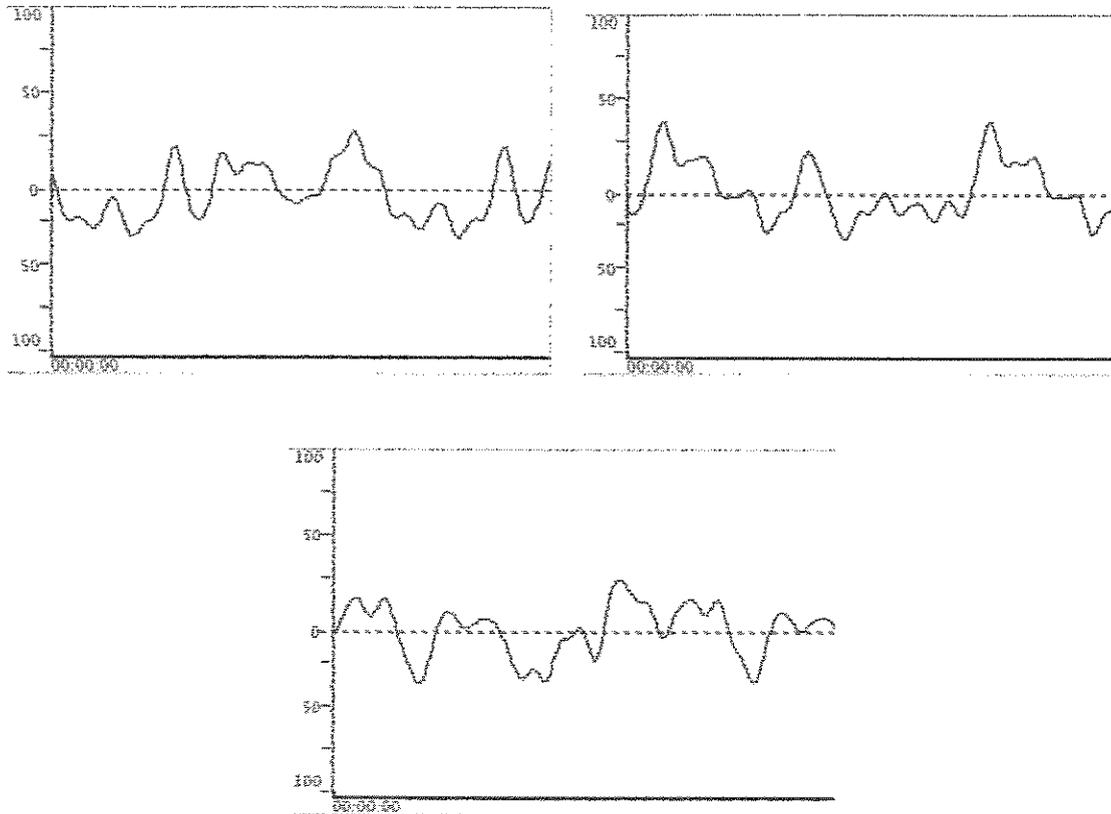


Figura 4.17a, b e c Formas de onda resultantes de distintas gerações de fases randômicas.

4.2.6.4 Fases com Distribuição Gaussiana

A distribuição gaussiana ou distribuição normal faz parte do grupo de distribuições chamadas *'bell shaped'* (com formato de sino) e são muito comuns em muitas aplicações diferentes, dentro e fora do mundo da música. Especificamente no domínio do som digital ela serve para a geração de ruídos, de envelopes de amplitude para a geração de grãos e para o controle de muitas situações musicais que precisem se valer da tendência dessa distribuição de concentrar os valores gerados em determinadas posições centrais, já que ela é simétrica.

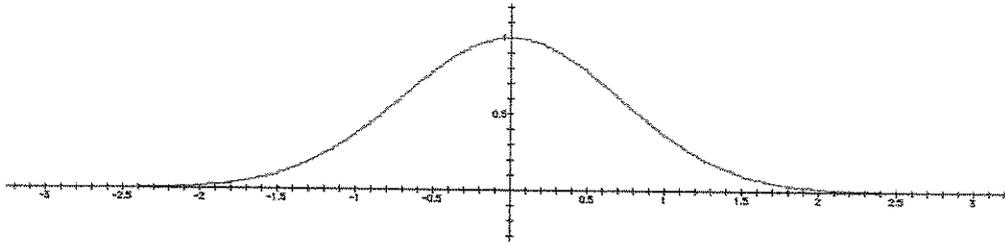


Figura 4.18 Curva padrão da distribuição gaussiana $(-\pi$ a $\pi)$.

A curva da distribuição gaussiana se apresenta originalmente da forma acima, segundo a expressão³⁶

$$\blacksquare y = e^{-x^2}$$

Figura 4.19 Expressão geral³⁷ da distribuição gaussiana³⁸.

mas para podermos visualizar como ficarão as fases em relação ao espectro frequencial, realizei a seguinte adaptação na expressão

$$\blacksquare x = e^{-(y-\pi)^2}$$

Figura 4.20 Expressão adaptada da distribuição gaussiana.

para gerar o seguinte formato:

³⁶ Esta é uma expressão geral. Outra expressão usada inclui as variáveis **média** e **desvio padrão**.

³⁷ Segundo GAUSS Curve, 2004.

³⁸ A letra e é uma constante matemática chamada número de Euler ou constante de Napier e tem o valor aproximado de 2.71828.

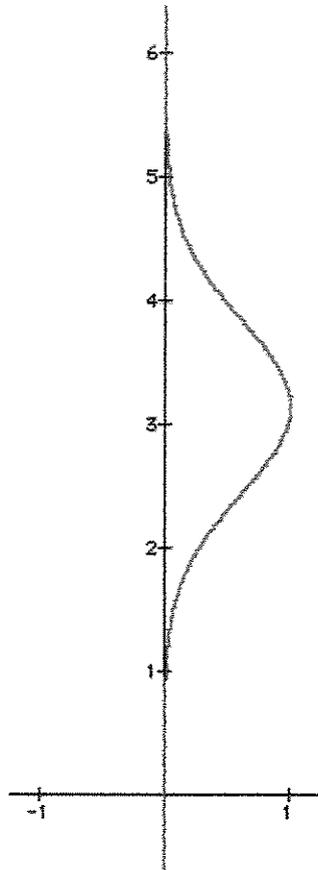


Figura 4.21 Forma adaptada da curva gaussiana. Eixo x representando as fases no âmbito 0 a 1 e eixo y representando frequências, sendo 2π equivalente ao âmbito frequencial máximo do som analisado.

Podemos interpretar a curva acima como deslocamentos de fase entre 0 e 1 no eixo x e frequências no eixo y (2π equivalendo ao âmbito frequencial dos parciais do som analisado). Cria-se, desse modo, uma tendência de endereçar fases zero e próximas de zero para as regiões mais graves e mais agudas do espectro.

4.2.6.5 Fases com Distribuição Beta

A distribuição beta pode assumir diferentes formatos, dependendo dos parâmetros fornecidos. A curva resultante pode ser dividida em duas partes: uma com valores no eixo x entre 0 e 0.5 e outra com valores entre 0.5 e 1 (no caso de uma distribuição beta

padrão). Cada uma destas partes pode assumir um formato diferente fornecendo diferentes parâmetros para as variáveis p e q . O tipo de curva que nos interessa é quando p e q são maiores que 0 e menores que 1, criando uma distribuição com características opostas à distribuição gaussiana.

$$f(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)} \quad 0 \leq x \leq 1; p, q > 0$$

Figura 4.22 Expressão para a distribuição beta padrão, na qual $B(p,q)$ designa a função beta³⁹.

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt$$

Figura 4.23 Expressão para a função beta.

Com as condições expostas acima, a maior probabilidade será de obtermos valores mais próximos de 0 e de 1, deixando ‘um vazio’ mais ou menos acentuado na região central.

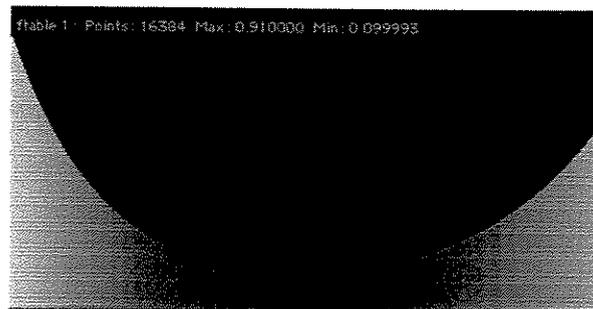


Figura 4.24 Curvatura da distribuição beta com parâmetros p e q diferentes em cada lado.

Assim como fizemos com a distribuição gaussiana, podemos imaginar o eixo x da curva acima como sendo o âmbito de frequência em questão em Hz e o eixo y representando os

³⁹ Segundo BETA, 2004.

valores de fase entre 0 e 1. Felizmente, para gerar valores de uma distribuição beta padrão em OpenMusic, não precisamos calcular nada. Basta fornecer os parâmetros para cada lado e gerar os valores.

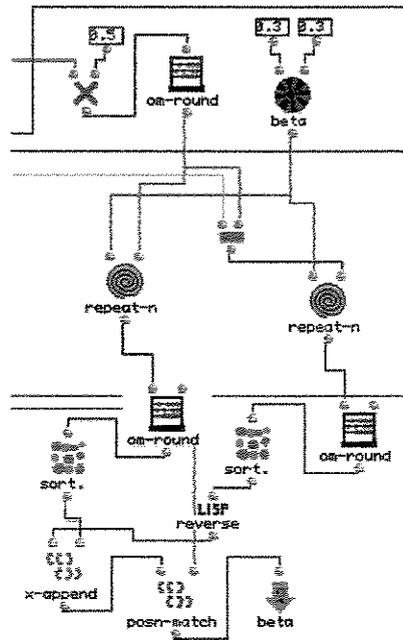


Figura 4.25 Segmento de *patch* que implementa a distribuição beta.

Na verdade, algumas adaptações precisam ser efetuadas para chegarmos a uma lista de valores de fase que esteja de acordo com as expectativas. O módulo *beta*, em OpenMusic, fornece um único dado por vez e por isso precisa ser executado tantas vezes quantos forem os elementos presentes na lista de entrada. A lista resultante precisa ser ordenada de tal maneira que a curva acima possa ser reproduzida. Como procedimento final, precisa-se criar uma ordem que corresponda à variação crescente de frequências da lista de entrada.



Som 4.14 (faixa 7) Som com alterações compassadas de fase (com *crossfade*) das seguintes configurações (na ordem de aparecimento): em fase, randômica uniforme, fases opostas, distribuição gaussiana e distribuição beta.

4.2.7 Quanto ao espaço

O tema da espacialidade na música eletroacústica é amplo, plurívoco e recorrente, mas não será abordado de acordo com a importância que lhe é devida, pois este mereceria um trabalho direcionado inteiramente ao assunto. Não só o tema em si seria extremamente extenso, mas também a sua relação com a nossa percepção seria um verdadeiro *tour de force*. A idéia apresentada aqui sobre o espaço é muito simples e diz respeito a uma única acepção, aquela referente ao espaço acústico. É a idéia da localização dos sons no espaço acústico alcançado pelas fontes sonoras, que pode ser fixa ou variável no tempo. Tecnicamente, a proposta está intimamente ligada ao conceito de síntese aditiva que formará a base para a implementação dessa idéia e, perceptivamente, a proposta é a fuga de uma escuta centralizada dos sons eletrônicos em direção a sensação de abertura estereofônica. Se produzimos, por exemplo, um único som eletrônico monofônico e não determinamos nenhuma localização específica para ele dentro do espaço das fontes sonoras que formam o par estereofônico, então esse som causará a sensação de estar vindo de uma fonte virtual posicionada equidistantemente entre as fontes sonoras reais. Sabe-se que há uma complexidade imensa no que diz respeito à dispersão de energia sonora dos instrumentos musicais no espaço e que, apesar de se conseguir mapear essa dispersão em condições cientificamente controladas, nenhuma função ou padrão de comportamento conseguem ser verificados e conseqüentemente reproduzidos. Portanto, uma correspondência precisa entre a dispersão de energia dos instrumentos e o posicionamento (ou a simulação de uma dispersão) de sons eletrônicos no espaço é, com o conhecimento que se tem no momento, impossível. De qualquer maneira, pode-se querer diminuir as distâncias entre as sensações causadas em ambos os casos. Sem buscar uma correspondência de sensações, mas ainda assim dando um passo relativamente significativo para suas aproximações, está a retirada do som eletrônico de

uma localização centralizada. Não estamos falando de trajetórias percorridas pelo som no espaço, nem de qualquer movimento, mas de localizações fixas distintas de ‘cada’ parcial sintetizado no processo da síntese aditiva, criando com isso a ‘abertura’ do som no espaço estereofônico e minimizando o disparate perceptivo de espacialidade entre os sons instrumental e eletrônico. A técnica utilizada é simples e consiste na geração de números que, em Csound, serão interpretados como posições no eixo x, associado com a lateralidade do som (esquerda – direita). Para cada parcial contido no espectro será associado um número entre 0 e 1, zero significando a posição mais à esquerda e 1 significando a posição mais à direita do eixo x.

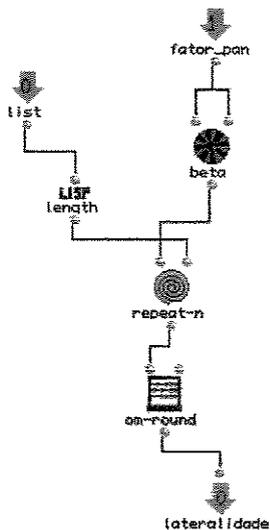


Figura 4.26 Patch criado para a geração de posições dos parciais.

Essas ‘posições’ podem ser geradas de muitas diferentes formas, mas nem todas criarão adequadamente a sensação de ‘abertura’ do som. Bem, que tipo de dados então forneceria de fato essa sensação? Se tencionamos que haja dispersão ‘perceptível’ dos sons pelo espaço estereofônico, devemos levar em conta a nossa capacidade em distinguir/perceber as posições de sons pelo espaço. É tido como norma que a nossa percepção de localização espacial de sons não é tão refinada como nossas percepções de altura e intensidade sonoras, mas alguns estudos têm mostrado que essa é uma verdade apenas

parcial. Segundo Léo Küpper, no âmbito de 20 a 20000 Hz, temos 1800 diferentes sensações de *pitch*, mas segundo seus cálculos, há 6000 pontos discrimináveis independentemente em uma esfera (KÜPPER, 1998, p. 298). Considerando uma distribuição estereofônica com fontes sonoras reais posicionadas em ângulo de 120 graus e equidistantes em relação a um ouvinte, teremos um amplo e preciso campo de audição onde poderemos perceber uma média de 63 posições espaciais distintas (KÜPPER, 1998, p. 290). Evidentemente será tanto mais difícil distinguir duas posições diferentes no espaço quanto mais próximas elas estiverem umas das outras. Uma distribuição uniforme de sons pelo espaço em questão faria com que as posições extremas (mais à esquerda e mais à direita) se destacassem das outras por serem mais facilmente percebidas, além de serem em menor quantidade que aquelas localizadas na grande região central. Com o intuito de equilibrar o grau de facilidade em perceber as localizações dos sons no espaço, utilizei uma distribuição randômica beta, que privilegia as extremidades de um âmbito dado. Não entrarei em detalhes acerca dessa distribuição porque, tanto a distribuição beta quanto o módulo que a executa já foram explicados acima. Contudo, será necessário ainda falar dos ‘fatores de panorâmica’ que precisam ser fornecidos ao módulo beta. O fator 1 corresponde à distribuição uniforme de dados e abaixo de 1 até 0 a acentuação da ‘lateralização’ da distribuição. Assim, quanto mais próximo de zero for esse fator, mais tendendo aos extremos será a distribuição. Deve-se atentar para o fato de que a tendência de distribuição em determinados pontos ou regiões não elimina a ocorrência de um evento fora do horizonte de probabilidades. O uso desses dados, referente ao posicionamento dos sons, será visto no próximo capítulo, sobre síntese, pois esse uso se dá na implementação de um algoritmo em Csound. Este *patch* foi criado para o fornecimento de parâmetros em uma configuração estereofônica, mas seria fácil ampliar o projeto para uma situação quadrifônica, bastaria que os dados de saída do *patch* fossem direcionados para duas entradas do módulo *instrument0*, a entrada já usada mais uma nova entrada ao lado da primeira. Ao contrário do que se possa pensar, mesmo extraindo duas séries de dados da mesma saída em uma única valoração do *patch*, os dados serão diferentes⁴⁰, mas respeitando os mesmos fatores de distribuição. Desejando-se fatores

⁴⁰ Se se quisesse que as duas séries de dados contivessem os mesmos valores, seria necessário empregar outra estrutura de geração de dados.

diferentes de distribuição para a lateralidade (esquerda – direita) e para a profundidade (frente – trás), tem-se que duplicar o *patch* e fornecer fatores individuais para o módulo beta. De qualquer forma, uma adaptação simples.

 Som 4.15 e 4.16 (faixa 8) Som sintetizado monofônico e o mesmo som com a distribuição estereofônica proposta.

4.2.8 Resultado Final

Como resultado final estrutural de todos esses processos, temos a geração de um arquivo texto para Csound no formato *score*⁴¹, onde são dispostos todos os dados produzidos em OpenMusic em uma disposição fixa.

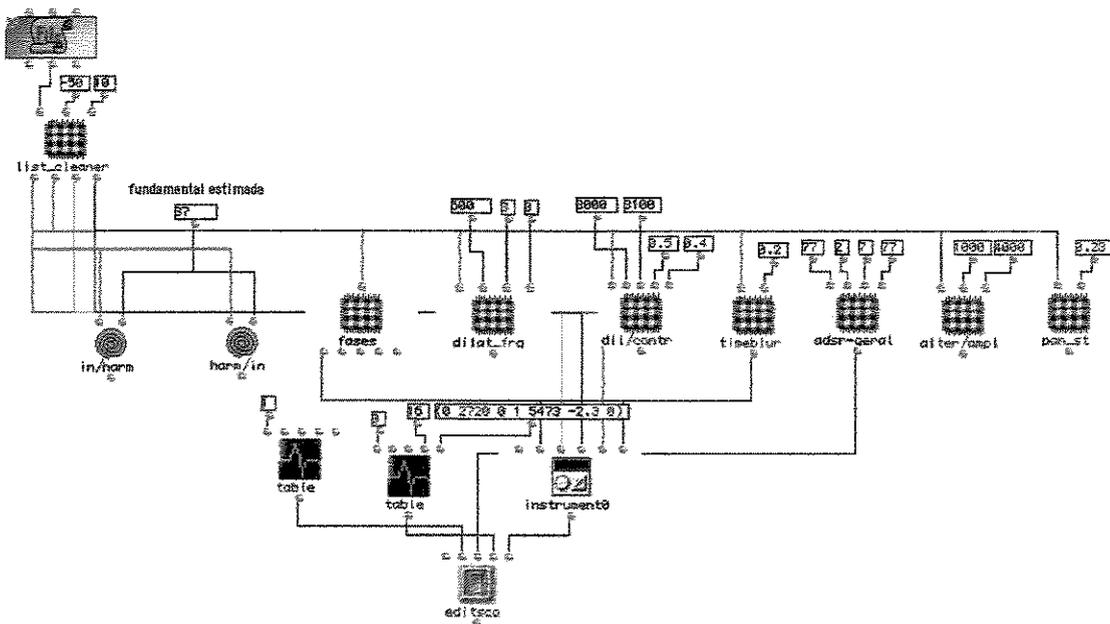


Figura 4.27 Patch geral que dá o resultado final de todos os processos.

⁴¹ Ver mais detalhes no capítulo 5.

O módulo *textfile*, que se encontra no canto superior esquerdo, é o responsável pela importação dos dados de análise e o módulo *editsco*, na parte de baixo do *patch*, cria o arquivo em formato *score*.

```
f1 0 16384 10 1
f2 0 8193 -16 0 164 1 182 -1.3000 0.5700 1147 0 0.5700 6799 -1.7000 0
f3 0 8193 16 0 2720 0 1 5473 -2.3 0

;p1 p2 p3 p4 p5 p6
i1 0.0930 0.1470 64.0310 101.5040 0.1900
i1 0.0930 2.0520 59.7120 130.0270 0.8700
i1 0.0930 6.7460 56.7840 162.7890 0.0300
i1 0.0930 6.6890 54.0350 195.3990 0.0100
i1 0.0930 6.7570 56.6970 228.2020 0.0900
i1 0.0930 6.7120 50.5620 293.6900 0.1500
i1 0.0930 6.7460 52.1080 326.5360 0.6300
i1 0.0930 6.5870 53.5690 359.4970 0.8500
i1 0.0930 6.6670 49.1960 392.5280 0.1700
i1 0.0930 6.7350 53.7330 425.9680 0.1100
i1 0.0930 6.6440 55.1990 459.1030 0.0500
i1 0.0930 6.7120 54.4230 492.4160 0.6000
```

Figura 4.28 Excerto de *score* produzido pelo *patch* geral em OpenMusic.

Mas não há um resultado final quanto aos dados contidos no *score*, pois são muitas as possibilidades de alteração de parâmetros e conexões possíveis entre os *patches*. Cada arquivo resultante pode conter um único parâmetro alterado ou todos simultaneamente. O resultado vai depender do compositor.

5. A Síntese

5.1 Introdução

A síntese sonora é uma atividade bastante abrangente que pode envolver dispositivos elétricos, eletrônicos, *softwares*, podendo ainda ser analógica ou digital. A sua história, dependendo do ponto de vista, pode remontar a mais de um século se tomarmos o Telharmonium de Thaddeus Cahill como referência, ou desde 1952 se pensarmos nos osciladores analógicos da Música Eletrônica Alemã com *Klangstudie I* de Herbert Eimert e Robert Beyer, 1957 se nos basearmos na síntese por computador que nasceu com Neuman Guttman na peça *The Siliver Scale*. Podemos ver que síntese é um termo amplo e que pode chegar a dimensões gigantescas se procurarmos assuntos relacionados. Um conceito de síntese que podemos assumir hoje seria, nas palavras de Depalle: “Segundo a aceção corrente, a síntese de sons é a atividade que consiste em criar novos sons a partir de dispositivos eletrônicos ou informáticos.” (DEPALLE, 1993, p. 122)

A síntese tratada neste capítulo é aquela da aceção que concerne à produção de novos sons a partir de dispositivos informáticos, mais propriamente via *software*. A história da síntese por meio de *softwares* já é extensa, apesar do relativamente pouco tempo de suas existências, e assistiu ao desenvolvimento e produção de uma grande quantidade de material relacionado - *softwares*, artigos, teorias, composições musicais, sons de naturezas diversas - e pessoas envolvidas - músicos, acústicos, engenheiros, matemáticos, etc.

Uma história exaustiva da síntese digital informática pode ser facilmente acessada via internet ou em muitos dos livros que tratam de ‘*computer music*’ (veja Referência Bibliográfica), mas uma breve idéia de seu desenvolvimento e de suas principais características serão úteis aqui.

O potencial de um *software* está diretamente relacionado com a capacidade da ‘máquina’ que lhe dá suporte (controlada por ele) e com os paradigmas adotados determinados pelo conhecimento da época. Assim não há maneiras de dissociarmos *software*, *hardware* e *modelo*. O avanço dessas três categorias evoluiu exponencialmente no tempo. Na década de 60 havia computadores enormes (poucos), altamente complexos

e dispendiosíssimos que eram programados para sintetizar poucas senóides simultâneas com pouquíssimo controle e qualidade sonora resultante baixíssima se a compararmos com o padrão de qualidade atual. Hoje temos *softwares* e computadores capazes de sintetizar milhares de sons, com controle de envelope, das mais variadas e complexas formas de onda em tempo real, mostrando um desenvolvimento vertiginoso da síntese. Os primeiros *softwares* criados para síntese sonora foram aqueles que ficaram conhecidos genericamente por ‘*Music N family*’ que começou com o Music I em 1957, desenvolvido por Max Mathews. Esse primeiro *software*, ainda programado diretamente em linguagem de máquina, só dispunha de uma única forma de onda: uma onda triangular. No ano seguinte nasceu o Music II que possuía a capacidade de sintetizar quatro vozes simultâneas. O auge dessa ‘família’ foi o Music V, criado em 1968, que foi amplamente distribuído e ajudou a popularizar a síntese por computador (ROADS, 1996). No meio desse caminho (no Music III, em 1960) surge um novo conceito: o da Unidade Geradora (UG) que se tornaria a lógica básica da grande maioria dos *softwares* utilizados ainda hoje. “UGs são módulos de processamento de sinal, como osciladores, filtros e amplificadores que podem ser interconectados para formar ‘instrumentos’ de síntese ou *patches* que geram sinais sonoros.” (ROADS, 1996, p. 89) Vemos abaixo uma pequena amostra da utilização de UGs em Csound:

```

...
kamp linen 10000, .02, 3, .1      ;produz um envelope linear do tipo ADSR
kfreq expon 440, 3, 880          ;produz uma variação exponencial
ionda = 1                        ;cria uma constante de valor 1

asynth oscili kamp, kfreq, ionda  ;cria uma oscilação com os dados
;anteriores
...

```

Figura 5.1 As funções *linen* e *expon* (em azul) são utilizadas dentro de outra função chamada *oscili*.

No exemplo acima podemos ver três UGs: *linen*, *expon* e *oscili*. Cada uma dessas unidades executa uma diferente função e podem ser interconectadas, como no caso da

unidade *oscili* que funciona a partir da utilização das variáveis *kamp* e *kfreq* criadas pelas outras unidades.

A partir dos modelos Music IV e Music V, muitos outros foram desenvolvidos, como Music 4BF, Music 360, Music 7, Music 11, Csound, Cmusic, Common Lisp Music e outros (ROADS, 1996, p. 90). Vários desses *softwares* foram passageiros (acabaram se transformando em outros ou simplesmente deixaram de ser desenvolvidos), mas alguns deles, como o Common Lisp Music e o Csound, continuam a ser desenvolvidos e utilizados ainda hoje por uma grande comunidade de músicos, matemáticos e engenheiros por todo o mundo. O Csound, por exemplo, além de um poderoso *software* de síntese, é gratuito e tem uma enorme quantidade de material relacionado à disposição (livros, artigos, tutoriais, cursos, fóruns de discussão, etc.), como bem mostram alguns *sites* na Internet, como www.csounds.com e www.csounds.com/ezine/.

Com a chegada de poderosos computadores pessoais, *softwares* constantemente desenvolvidos e melhorados por uma comunidade crescente de usuários e inúmeras novas pesquisas nos campos da acústica, psicoacústica, cognição, neuropsicologia, etc. as possibilidades na síntese tornaram-se enormes, aproximando-nos de uma situação na qual bastaria definir o que se quer e saber implementar as idéias.

5.2 Técnicas de Síntese

5.2.1 O Modelo em Síntese

O tópico abordado agora será o modelo, mencionado acima como complementar intimamente relacionado ao *hardware* e ao *software*. O modelo em síntese, que estabelece uma relação entre o pensamento escritural do compositor e a ação de criação de sons, é amplo e complexo, mas algumas questões podem nos impulsionar a estabelecer um conceito que nos sirva de guia. Que tipo de som se deseja criar? Qual o ponto de vista adotado, ou seja, o ponto de vista do som tal qual ele é produzido ou de como ele é

percebido? Quais objetivos musicais impulsionam o compositor a escolher este ou aquele tipo de síntese?

Em síntese de sons, como em todas as disciplinas que consistem em simular e extrapolar os fenômenos reais, a noção de modelo exerce um papel central e essencial.

Um modelo é uma abstração que tem por objetivo condensar de maneira formal a informação veiculada pelo som. Essa informação é representada por um grupo de valores que mudam no tempo, chamados parâmetros, que identificam o som no modelo escolhido. Mas um modelo não é somente um meio de representar os sons, é igualmente um meio de definir as características gerais do dispositivo de produção que, na prática, gera o som a partir de valores de parâmetros. (DEPALLE, 1993, p. 123)

Há vários diferentes modelos empregados em síntese (como será visto a seguir), mas cada um deles demanda a adoção de um paradigma particular aliado a pesquisas centradas no tema escolhido, tornando muito abrangente e complexa a integração de mais de um modelo simultaneamente. Desse modo, um único modelo foi adotado neste trabalho: o **modelo espectral**. Com o intuito de se entender o motivo pelo qual um modelo/técnicas foram adotados em detrimento de outros, serão explicados seus funcionamentos e suas principais idéias, por vezes com auxílio de exemplificações, de modo que possamos ter uma visão geral dos paradigmas e procedimentos envolvidos no domínio da síntese sonora. Uma visão geral é importante porque fornece as bases para comparação que, conseqüentemente, leva a um maior conhecimento da área através das relações de idéias e sons.

Existem várias classificações de uso corrente, pois sempre surgem novas técnicas (ou procedimentos de síntese dentro de cada modelo) e novas maneiras de se executar funções semelhantes, o que exige reagrupamentos dessas técnicas, além do que nem todas as técnicas de síntese consistem em processos tão unívocos que não gerem margem para diferentes interpretações. Uma das classificações mais pertinentes, de Julius Orion Smith III, possui quatro subdivisões (SMITH III, 1993, p. 85):

- Tratamento de Registro (gravação)
- Modelos Físicos
- Algoritmos Abstratos

- Modelos Espectrais

5.2.2 Classificações de Modelos e Técnicas Envolvidas

5.2.2.1 Tratamentos de Registro

Os **tratamentos de registro** são os processos que envolvem manipulações de sinais gravados, como a amostragem (*sampling*) que, a partir do uso manipulado ou não das amostras digitalizadas de som, se compõe um som que poderá manter, dependendo da manipulação realizada, as principais características perceptivas do som original. É esse o processo empregado em uma grande série de teclados sintetizadores que também são *samplers*¹. Outra técnica muito difundida é a síntese granular. Nesse tipo de síntese, amostras do som são fragmentadas em pequenas unidades sonoras chamadas grãos² as quais são manipuladas temporalmente, como a criação de distanciamento entre os grãos de modo que se formem espaços de silêncio ou aproximação, que poderá sobrepor os grãos uns sobre os outros, de modo regular ou randômico (este último mais característico), além das possíveis manipulações de altura e numerosas outras. Esta técnica pode gerar sons com características muito diferentes daquelas do som original, aplicando novos coloridos e alterando suas características morfológicas³.



Som 5.1 (faixa 9) Granulação de voz feminina falando a palavra 'raiva'.

5.2.2.2 Modelos Físicos

O **modelo físico** refere-se à criação de sons sob o ponto de vista da descrição física da fonte sonora, ou seja, da reconstrução do comportamento acústico dos

¹ Equipamentos que empregam o processo de amostragem de um sinal de áudio.

² Ver ROADS, 2001.

³ Morfologia, nesse caso, refere-se à configuração geral de forma e material de eventos sônicos conforme eles são percebidos através de sua evolução no tempo (MORPHOLOGY, 2004).

instrumentos (JEHAN, 2001, p. 18). O estudo das propriedades acústicas dos instrumentos, o que inclui o dispositivo que gera a energia sonora – excitador – e a parte que vibra sob o efeito da excitação moldando essa energia – o ressonador, envolve muita pesquisa e medidas acústicas. Aqui o nível de complexidade da modelização torna-se grande, pois cada um dos distintos excitador e ressonador terão descrições físicas diferentes, assim, as funções e expressões matemáticas que descrevem esses comportamentos terão que ser compostas individualmente. No caso de criarmos um modelo que representasse a geração de som em uma flauta, por exemplo, teríamos que descrever a pressão de ar exercida na flauta, efeitos da embocadura, o nível de ruído do sopro, o atraso do som causado pelo tamanho do tubo, taxa de vibrato, etc. Pesquisa da verossimilhança dos sons produzidos por diversos modelos físicos de flauta (realizada por mim durante este trabalho) mostrou que poucos dos modelos estudados são convincentes. As maiores vantagens desses modelos residem na produção de sons ruidosos e formantes geradas pelos ressonadores. Apresento aqui um relativamente bom modelo físico de flauta desenhado por Perry Cook, adaptado do exemplo do manual do Csound (VERCOE, 2003):

```

;orchestra
    instr      1902

aflute1  init      0
ifqc     =         cpspch(p5)
ipress   =         p6
ibreath  =         p7
ifeedbk1 =         p8
ifeedbk2 =         p9

; Configuração do fluxo
kenv1    linseg    0, .05, 1.1*ipress, .2, ipress, p3-.16, ipress, .02, 0
kenv2    linseg    0, .01, 1, p3-.02, 1, .01, 0
kenvibr  linseg    0, .5, 0, .5, 1, p3-1, 1 ; VIBRATO ENVELOPE

; Os valores têm que ser aproximadamente -1 a 1
aflow1   rand      kenv1
kvibr    oscil     .1*kenvibr, 5, 3

; ibreath pode ser usado para ajustar o nível de ruído
asum1    =         ibreath*aflow1 + kenv1 + kvibr
asum2    =         asum1 + aflute1*ifeedbk1

afqc     =         1/ifqc - asum1/20000 -9/sr + ifqc/12000000

; o delay da embocadura deve ser metade do delay do tubo
; ax
atemp1   delay     asum2, (1/ifqc-10/sr)/2
ax       delayr    1/ifqc/2
ax       deltapi   afqc/2 ; - asum1/ifqc/10 + 1/1000
ax       delayw    asum2

apoly    =         ax - ax*ax*ax
asum3    =         apoly + aflute1*ifeedbk2

avalue   tone      asum3, 2000

; O tamanho do tubo determina o pitch - tubo menor significa pitch mais alto
atemp2   delayr    1/ifqc
aflute1  deltapi   afqc
aflute1  delayw    avalue

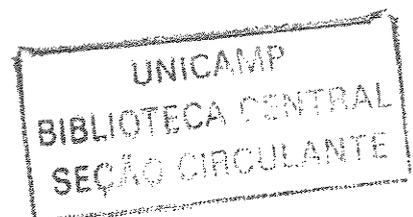
    out      avalue*p4*kenv2
    endin

;score
; SINE
f3 0 1024 10 1

t 0 500
; SLIDE FLUTE
; START DUR AMPLITUDE PITCH PRESSURE BREATH FEEDBK1 FEEDBK2
i 1902 0 10 5000 8.00 .9 .036 .4 .4
i 1902 + 10 . 8.03 .95 . . .
i 1902 . 10 . 8.07 .97 . . .
i 1902 . 10 . 8.08 .98 . . .
i 1902 . 17 . 7.11 .99 . . .
e

```

Figura 5.2 Exemplo de implementação de modelagem física de flauta.





Som 5.2 (faixa 10) Simulação de som de flauta por modelagem física segundo o modelo da figura 5.2.

A dificuldade na utilização dos modelos físicos é, fundamentalmente, a grande quantidade de conhecimento envolvido nas modelizações, uma dificuldade proibitiva para o músico sem formação na área de acústica ou engenharia. Dentre as mais conhecidas técnicas estão a síntese modal e o *waveguide*⁴ (ver ROADS, 1996).

5.2.2.3 Algoritmos Abstratos

Os **algoritmos abstratos**, como o próprio nome sugere, não possuem relação com eventos sonoros pré-existentes nem com constituições físicas de fontes sonoras. A sua origem ou é extramusical ou não se relaciona com o mundo acústico. Alguns exemplos conhecidos de algoritmos abstratos são as sínteses FM (*Frequency Modulation*), modulação em anel (*Ring Modulation*) e sínteses AM (*Amplitude Modulation*). Estes são exemplos de uma subcategoria de algoritmos abstratos chamada síntese por modulação⁵. Para se ter uma idéia do nível de abstração, tomemos a modulação em anel como exemplo. Podemos criar sons com essa técnica simplesmente multiplicando dois sinais, um pelo outro, no domínio temporal. O resultado obtido pode ser entendido através da seguinte identidade trigonométrica:

$$\cos (C) \cdot \cos (M) = 0.5 (\cos (C - M) + \cos (C + M))$$

Figura 5.3 Identidade trigonométrica que define o comportamento de uma modulação em anel.

⁴ Guia de onda.

⁵ Podemos entender modulação genericamente como variação de um sinal de acordo com aspectos de um outro sinal.

Podemos entender o exemplo acima como a multiplicação de dois sons puros (uma cossenóide é tão pura quanto uma senóide, mas tem um deslocamento de fase de 90° em relação a ela) de frequências C e M, resultando na soma de duas outras cossenóides, mas com frequências equivalentes à soma e à diferença das frequências iniciais. Se modularmos dois sons com frequências de 500 Hz e 100 Hz, obteremos como resultado a soma e a diferença dessas frequências, ou seja, 600 Hz e 400 Hz. No caso da modulação de sons não senoidais, o processo se dá para cada componente de frequência dos espectros envolvidos.



Som 5.3 e 5.4 (faixa 11) Som de vozes sem modificação e modulação em anel do mesmo som com um som senoidal com frequência variável de 350 a 320 Hz.

Já a síntese por modulação de frequência (FM) pode ir muito além em possibilidades de espectros sonoros resultantes. John Chowning, pioneiro da síntese por modulação de frequência, demonstrou que os resultados podem ser tão variados quanto sons quase senoidais, espectros harmônicos, inarmônicos ou ruidosos. A síntese FM historicamente assumiu um papel muito importante na integração dos sons instrumentais com os sons eletrônicos, apesar de seu caráter abstrato. No início da década de 70 surge uma corrente musical denominada música espectral que trouxe para o domínio da escritura instrumental, entre outros, modelos sonoros e de escritura como metáforas derivadas da prática de estúdio da música eletroacústica, como o *looping* com reinjeção⁶, a modulação em anel e a síntese FM. Em obras como *Time and Again* de Tristan Murail, por exemplo, são empregados sons sintetizados através de FM integrados à parte instrumental cujas notas foram derivadas do mesmo processo de síntese. A importância dessa técnica de síntese reside na fácil modelagem e integração que se pode obter de agregados ricos de sons que se fundem muito eficientemente entre si e com sons instrumentais. Seu uso pode ser facilmente integrado às técnicas de síntese aditiva com bons resultados e é por isso que vai ser dada a ela um pouco mais de atenção nesse capítulo.

⁶ Esta técnica remete à gravação de sons em fita que, em um círculo fechado, volta a ser gravada misturando os sons a cada vez que se repete o processo e, devido às imperfeições dos equipamentos e técnicas, sofre desgaste e gera novos sons (MURAIL, 2002, p. 115).

Na síntese FM os dois sinais envolvidos assumem diferentes funções: um deles é chamado onda portadora e o outro onda moduladora. Além das frequências dessas duas ondas, outro fator importante entra em jogo, que é o fator que vai ser o responsável pela maleabilidade da síntese: o índice de modulação. O índice nada mais é que a razão entre a amplitude do sinal da onda moduladora e sua frequência. Quanto maior o índice, mais rico e amplo se torna o espectro resultante. Para fácil entendimento da participação do índice de modulação na formação do espectro, podemos associar cada unidade de índice à criação de um par de componentes laterais igualmente espaçados em relação à onda portadora, com espaçamento igual à frequência da onda moduladora.

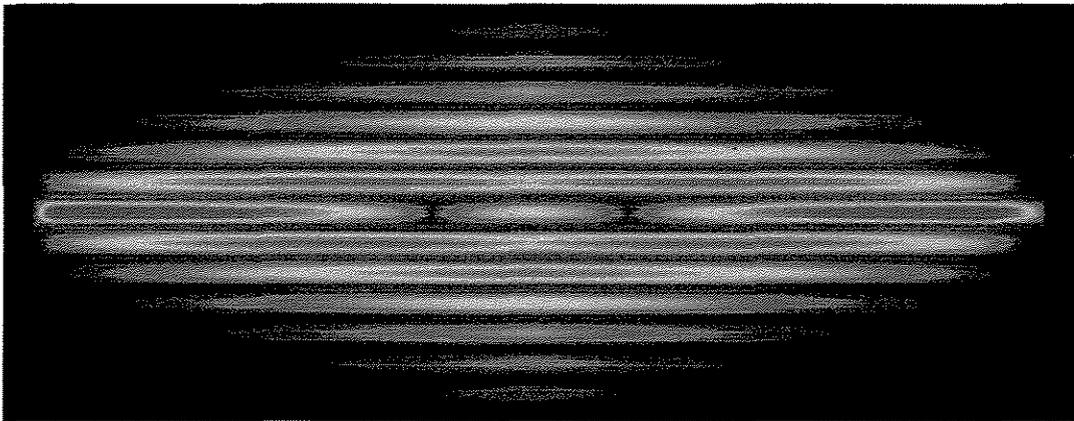


Figura 5.4 Sonograma de um som gerado por FM. O eixo x representa tempo, o eixo y, frequência e as cores representam as amplitudes – vermelho mais intenso, azul menos intenso.

Na figura acima pode-se ver uma frequência central mais avermelhada que corresponde à onda portadora e os pares de componentes ao seu redor gerados pelo aumento do índice de modulação que varia no tempo. O índice começa e termina em zero, tendo seu valor máximo no meio do som, onde o espectro está mais rico. Se estabelecermos que o índice será variável no tempo, então teremos espectros ‘dinâmicos’ como no exemplo acima, ou seja, que se alteram no tempo. Outro interesse é o fato de que quando se muda o índice no tempo, não estamos apenas alterando as componentes frequenciais do espectro, mas também a relação de amplitude dos parciais gerados, criando assim um espectro rico em

movimento (perceba a alteração de cores na figura 5.4). A variação de amplitude dos parciais segue uma função chamada função de Bessel.

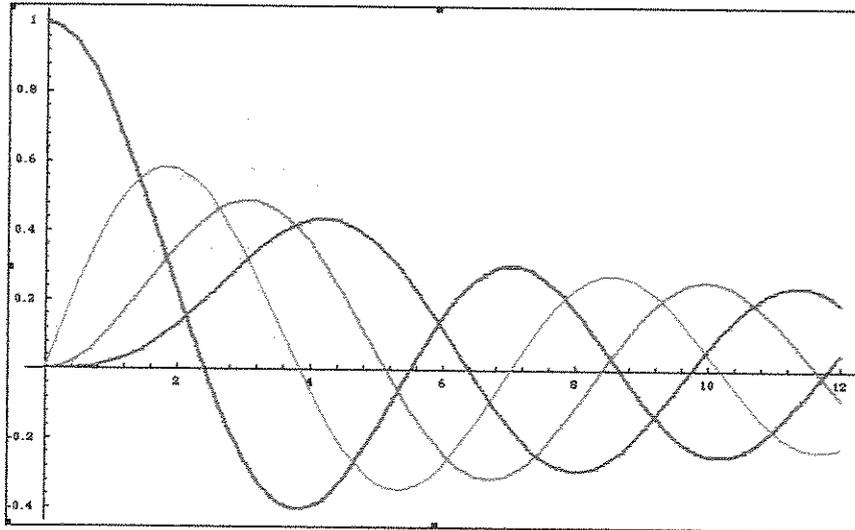


Figura 5.5 Alteração de amplitude da onda portadora (em vermelho) e dos componentes laterais até terceira ordem⁷ em uma FM com índice variando de 0 a 12 (eixo x – amplitude, eixo y – índice de modulação).

Esse é o processo básico de modulação de frequência, mas outros processos mais complexos tornaram-se mais usuais, quando há ondas portadoras e moduladoras em série, em paralelo e com realimentações⁸, criando algoritmos intrincados, como aqueles usados em sintetizadores, como o lendário DX7 da Yamaha.

⁷ De maneira simplificada, a cada índice de modulação, um par de componentes laterais simétricos em relação à onda portadora (*sidebands*) é gerado, correspondendo a uma 'ordem'.

⁸ Quando o resultado de um processo volta, de algum forma, para a entrada do algoritmo.

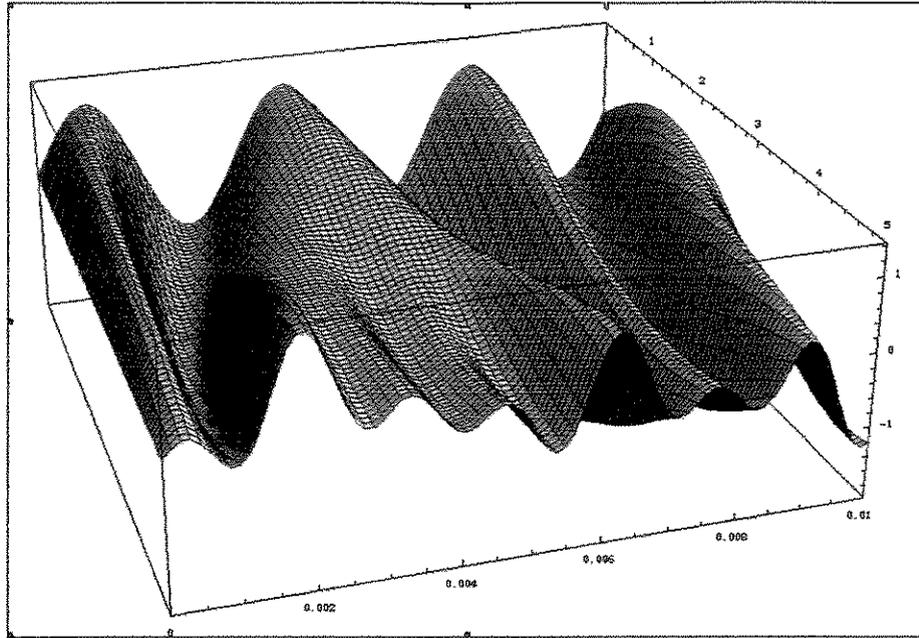


Figura 5.6 Representação tridimensional da variação das amplitudes da onda portadora e de componentes até terceira ordem em uma FM (portadora de 300 Hz e moduladora de 129 Hz) com índice variando de 0 a 5 (eixo x - tempo, eixo y – amplitude, eixo z – índice de modulação e frequência).



Som 5.5 (faixa 12) Som resultante de um encadeamento de processos básicos de FM.

Além das sínteses por modulação, há outras técnicas de algoritmos abstratos bastante difundidas, como a síntese estocástica, que faz uso das curvas de distribuição de probabilidades para gerar formas de onda complexas (tais como Poisson, Gaussiana, Uniforme, etc.), síntese através do uso de autômatas celulares (*Cellular Automata Lookup Table* – ver MIRANDA, 2002), a síntese por *waveshaping*, etc. A síntese por *waveshaping*, de uso comum há pelo menos duas décadas, procura alterar o formato de uma onda simples através de algum tipo de distorção (MIRANDA, 2002). Normalmente se utiliza uma ‘função de transferência’, conhecida como *waveshaper*⁹, para distorcer uma onda inicial. Um caso comum de *waveshaping*, porém nem sempre desejado, é o chamado *clipping*¹⁰, ou distorção da forma de onda pela ultrapassagem do limite máximo de amplitude que um sinal digital pode ter. Se sintetizarmos uma senóide com amplitude

⁹ ‘Modelador’ de onda.

¹⁰ *Clipping* significa corte e é o que acontece no contorno de uma onda quando sua amplitude ultrapassa o limite máximo de representação em um sistema digital.

próxima ao limite máximo e a amplificarmos, obteremos cortes na forma de onda, fazendo com que a senóide se pareça com uma onda quadrada e soando de acordo com o novo contorno (*shape*), ou seja, um som cheio de harmônicos no início e suavizando até voltar à forma, e portanto à sonoridade, de onda senoidal.

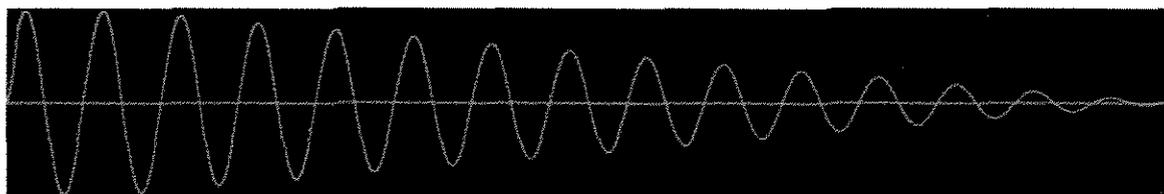


Figura 5.7 Onda senoidal com amplitude próxima ao limite máximo de representação.

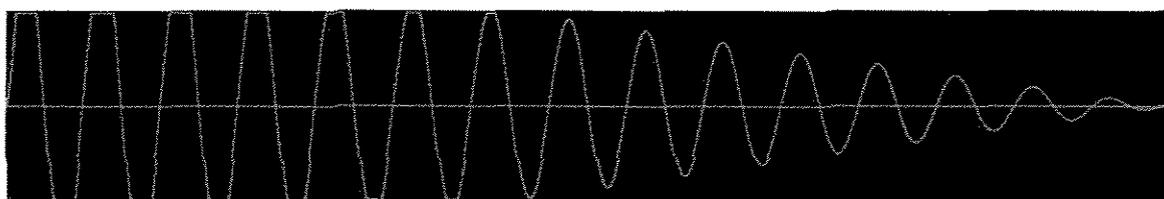


Figura 5.8 A mesma onda que a anterior distorcida por meio de amplificação.



Som 5.6 (faixa 13) Som senoidal distorcido através de *waveshaping*.

5.2.2.4 Modelos Espectrais

O **modelo espectral**, adotado no presente trabalho, é uma confluência de um conjunto de situações e contextos que levou muitos compositores a adotar esse conceito. A busca do timbre como elemento principal na composição em vez da ‘nota’, a consideração da psicoacústica e das ciências cognitivas como áreas mais respeitáveis e amadurecidas e a transposição da teoria de análise de Fourier para o campo sonoro levaram ao surgimento do paradigma do modelo espectral (MIRANDA, 2002, p. 49-50). Os modelos espectrais são baseados na percepção e não na descrição dos processos de produção do som ou em idéias extramusicais. “Modelos de espectro tentam parametrizar

um som na membrana basilar do ouvido, descartando qualquer informação que o ouvido pareça descartar no espectro.” (SERRA, Xavier, 1997, p. 91) Como foi visto no capítulo 3, a teoria de Fourier nos trouxe a idéia de que poderíamos compor qualquer som periódico pela sobreposição balanceada de ondas senoidais com diferentes frequências e fases através do processo inverso ao da análise. Considerando todas essas conjunções, podemos entender melhor o destaque deste modelo: primeiramente, as inúmeras novas descrições fisiológicas e neuropsicológicas levaram a um maior entendimento da percepção sonora; em segundo lugar a busca do trabalho com timbres nas composições; em terceiro lugar, a fundamentação matemática do processo de estruturação do som complexo que, juntos, formaram os princípios que justificam o sucesso do modelo.

Dentre as mais conhecidas técnicas do modelo espectral estão a síntese aditiva, a síntese subtrativa, o *phase vocoder* e a síntese de senóides + ruído (*sinusoidal plus noise synthesis*). A síntese aditiva se baseia em uma das idéias mais simples já empregadas: a adição. Esta técnica foi a primeira a ser usada nas sínteses por computador e é hoje uma das mais poderosas técnicas de modelagem espectral apesar de ter sido praticamente abandonada devido ao surgimento de outras técnicas até ser novamente retomada devido ao seu potencial ampliado que os novos contextos geraram. A síntese aditiva é fundamentada na sobreposição de sons simples, como senóides ou ondas triangulares (hoje em desuso nessa técnica), para a geração de sons mais complexos. As primeiras peças de síntese aplicavam essa técnica, como o *Studie I* de Stockhausen construído inteiramente sobre senóides (mas usando osciladores em vez de computador) ou *The Silver Scale* de Newman Guttman que usava apenas ondas triangulares por limitações da programação do sistema empregado no enorme computador IBM do Bell Laboratories. A limitação que se viu inicialmente (por volta da década de 60) era a extrema dificuldade em se chegar a sons mais complexos e interessantes com a adição de sons senoidais, um a um, com algum controle independente. A riqueza dos sons ditos naturais não estava somente na complexidade frequencial dos seus espectros, mas também na sua ‘dinâmica’, na combinação de micro-variações de frequência e amplitude de cada parcial e aí residiam as dificuldades. Estas diziam respeito à capacidade computacional limitada em criar as centenas de componentes com controladores de envelope que descrevessem seus percursos dinâmicos e frequenciais e principalmente à falta de um modelo que indicasse

os caminhos para a riqueza desejada. Esses problemas só puderam ser mais largamente resolvidos com a disseminação da análise paramétrica via FFT na década de 70¹¹ e com a ascensão dos computadores no início da década de 80. Os resultados musicais começaram então a aparecer. Um exemplo já clássico é a peça *Mortuos Plango, Vivos Voco* de Jonathan Harvey, de 1980, que se baseia nas análises de um som de sino para gerar a harmonia da peça através da síntese. Outro exemplo do uso de análise/ressíntese acontece de maneira metafórica, fora do domínio puramente eletrônico, na utilização que a Música Espectral fazia da idéia de síntese aditiva. A partir das análises de um determinado som, executa-se uma ‘síntese aditiva’ com instrumentos reais, cada um dos quais tocando ‘um parcial’, resultando na formação de um timbre através da técnica da adição¹². Esse é o caso de *Partiel* de Gerard Grisey, de 1979, para 18 músicos e *Désintégrations* de Tristan Murail, de 1982-83, para 17 instrumentos e *tape*.



Som 5.7 (faixa 14) Som resultante de síntese aditiva.

Da necessidade de obtenção de sons complexos mais facilmente surge a síntese subtrativa na qual, em vez de se ‘adicionar’ sons uns sobre os outros, subtraem-se sons ou faixas de sons de um espectro denso e já complexo, como o ruído branco. Dessa forma podia-se chegar a sons com espectros muito mais ricos apenas ‘esculpindo’ uma matéria-prima já complexa. Na verdade esse processo não se constitui necessariamente em uma síntese, mas sim em uma ‘filtragem’¹³. Normalmente o caracterizamos como síntese devido à geração artificial dos espectros complexos a serem filtrados (ruído branco, ruído

¹¹ A história da análise do som é uma verdadeira saga. Muitos tipos de análises já foram implementados, tanto analógicos quanto digitais, sob os mais variados pontos de vista e é complicado afirmar quando e qual algoritmo de análise mais influenciou o renascimento da síntese aditiva dentro do chamado processo análise/ressíntese. O que se pode dizer mais acertadamente é que a análise de espectro e o cálculo da STFT determinaram o início de sua ascensão entre o final da década de 70 e início da década de 80.

¹² Obviamente os instrumentos não colaboram com um único parcial, mas com toda sua riqueza tímbrica. A idéia é que os instrumentos colaborem com o *pitch* de destaque substituindo um parcial no processo de adição. O resultado notoriamente extrapola o que seria uma simples síntese aditiva, mas chega a sonoridades mais interessantes que não teriam sido alcançadas de outra forma.

¹³ Um filtro, em termos gerais, pode ser definido como um dispositivo que acentua ou atenua regiões de um espectro sonoro (ROADS, 1996, p. 185). Há muitos tipos diferentes com funcionamentos que podem ser bastante complexos.

rosa, *pulse wave*¹⁴, etc.). Parece mais difícil imaginar um resultado sonoro a partir da eliminação do som que ‘não’ se quer do que a adição do que se quer. Mas a síntese subtrativa pode ir além da eliminação de componentes particulares em um espectro complexo, ela permite também a eliminação de quase todo o espectro ‘deixando’ componentes particulares.



Som 5.8, 5.9 e 5.10 (faixa 15) Ruído branco e o mesmo ruído com *bandreject* e com *bandpass*. Percebe-se pouca diferença entre os dois primeiros.



Som 5.11 (faixa 16) Som obtido a partir de filtrações *bandpass* de ruído branco.

Essas diferenças constituem os diferentes tipos de filtrações. A eliminação de frequências ou faixas específicas de um som complexo se dá por meio de uma filtração chamada *bandreject* (rejeita banda), para a qual se define uma frequência de centro e um âmbito de frequência que a cerca. O filtro que define uma frequência ou uma faixa que vai restar da filtração é conhecido como *bandpass* (passa banda)¹⁵. Pode-se perguntar qual seria a vantagem de uma síntese subtrativa para a qual se configura uma fina faixa de frequência resultante em relação à uma síntese aditiva da mesma altura percebida. Duas diferenças principais são: 1- a síntese subtrativa nunca isolará uma única frequência, sempre resultará uma faixa, mesmo que estreita; 2- a amplitude resultante seguirá a variação de amplitude do som original relativa à faixa de frequência escolhida, enquanto que na síntese aditiva qualquer variação de amplitude precisa ser criada através de um controlador. Assim, dependendo do intuito da criação de um som, uma técnica pode ser mais vantajosa que outra.

O *phase vocoder* é o tipo de procedimento que levanta dúvidas sobre a sua natureza. Classificado por Smith III como técnica de modelagem espectral, muitos nem o

¹⁴ Em princípio, trata-se de uma simples amostra de som (*sample*) que conteria todos os parciais harmônicos infinitamente, por isso normalmente se usa o chamado *band-limited pulse* que tem seu espectro restrito à frequência de Nyquist (ZICARELLI, 2004).

¹⁵ Não vou entrar em detalhes sobre filtro e filtração, pois esse é um assunto muito extenso e não cabe como parêntese do assunto principal. Para ver mais detalhes, ver ROADS, 1996).

consideram uma técnica de síntese, mas sim de tratamento sonoro¹⁶. Aqui nos encontramos no limiar entre dois domínios: tratamento ou síntese. Se considerarmos o *phase vocoder* como síntese, todas as transformações de som executadas no AudioSculpt que exigem parâmetros de análise, como o *time stretch*¹⁷ (estiramento temporal) ou filtragens, terão que ser consideradas síntese, pois o *phase vocoder* é o processo base desse *software*¹⁸. Particularmente prefiro considerá-lo um processo de apoio ao tratamento sonoro e não síntese propriamente dita. Explicações desse processo e da síntese de senóides + ruído já foram vistas no capítulo 3; essa última será em parte revista no desenvolvimento desse capítulo.

5.3 Idéias e Implementações de Síntese em Csound

5.3.1 Introdução

O *software* Csound foi escolhido por *n* razões, a maior parte delas já explicadas no decurso deste trabalho, mas uma das razões que se destaca, que se já foi citada antes merece ser citada novamente, é a sua grande maleabilidade e abertura. Pode ser que não seja prático ou simples executar determinada operação em Csound, mas realmente poucas coisas não são possíveis de serem executadas. O Csound é conhecido como um *software* ‘difícil’ para músicos e muitos preferem passar ao largo dele, mas não sem justificativas. Sua ‘curva de aprendizado’ é bastante íngreme, exigindo, de modo geral, que se gaste muito tempo no aprendizado inicial antes de se realizar qualquer coisa sonoramente estimulante. Mas o que normalmente está por trás das justificativas é a opção de se trabalhar de forma intuitiva, dando preferência a *softwares* com interfaces gráficas para realizar suas transformações, ou o preconceito, não incomum no Brasil.

¹⁶ Transformação de som gravado.

¹⁷ Disponível no AudioSculpt como *Time Scale Modification*.

¹⁸ O algoritmo de *phase vocoder* implementado no AudioSculpt se chama SVP – *Super Vocodeur de Phase*.

O que não será feito aqui é ensinar a usar Csound. Eu poderia dizer que para isso se tem o manual, mas ele não cumpre bem essa tarefa e passa-se então a depender de tutoriais ou de aulas. O que será feito aqui são esclarecimentos a respeito da utilização de determinados *opcodes*¹⁹ e estruturas de programação que sejam fundamentais para o entendimento das implementações propostas neste trabalho.

5.3.2 Implementações

As implementações são de três categorias - implementação de modelos psicoacústicos, manipulação de dados para a síntese e técnicas de preparação e endereçamento de dados – que agrupam as seguintes propostas:

1- Modelos psicoacústicos

- limiar de audibilidade (ATH²⁰)
- *jitter*²¹
- *shimmer*²²
- distribuição espacial

2- Manipulação de dados para a síntese

- senóides + ruído
- hibridização

3- Técnicas de preparação e endereçamento de dados

- criação do arquivo de análise de *phase vocoder*
- integração dos processos para a síntese final

¹⁹ *Opcode* em Csound se refere a uma unidade geradora (UG), assim, todos os geradores de sinal, modificadores de sinal, conversores, processadores espectrais, etc. são *opcodes* nesse sentido. Segundo Rodolfo Caesar, “*Opcode* é o nome das sub-rotinas codificadas em C ou C++ estocadas na biblioteca do Csound.” (CAESAR, 2004)

²⁰ *Audible threshold of hearing*.

²¹ Micro-variações de frequência das componentes espectrais que caracterizam os sons não artificiais.

²² O mesmo que *jitter*, mas referente às amplitudes.

5.3.2.1 Limiar de Audibilidade

O limiar de audibilidade, ou ATH, como já discutido no capítulo sobre interação, refere-se a uma curva de amplitude versus frequência que representa uma ‘filtragem’ realizada pelo nosso sistema auditivo e que caracteriza nossa percepção quanto aos limites do que pode ser notado. Sua função no Csound será selecionar para síntese os parciais que serão efetivamente percebidos. Uma primeira vantagem é impedir que componentes não perceptíveis colaborem com a energia final do espectro, pois dessa forma pode-se realocar suas energias para o grupo de componentes percebidos, gerando um aumento relativo da sensação de volume. A expressão matemática para o cálculo de ATH é aproximada, pois uma expressão que representasse exatamente a curva de percepção, se possível, seria extremamente complexa. Como a própria curva é um resultado estatístico de testes de júri²³, não teria sentido buscar a perfeição da representação matemática. Para a implementação dessa curva em Csound, usou-se a expressão matemática proposta por Ernst Terhardt.

²³ Testes cujos resultados são obtidos através de respostas de um grupo de pessoas.

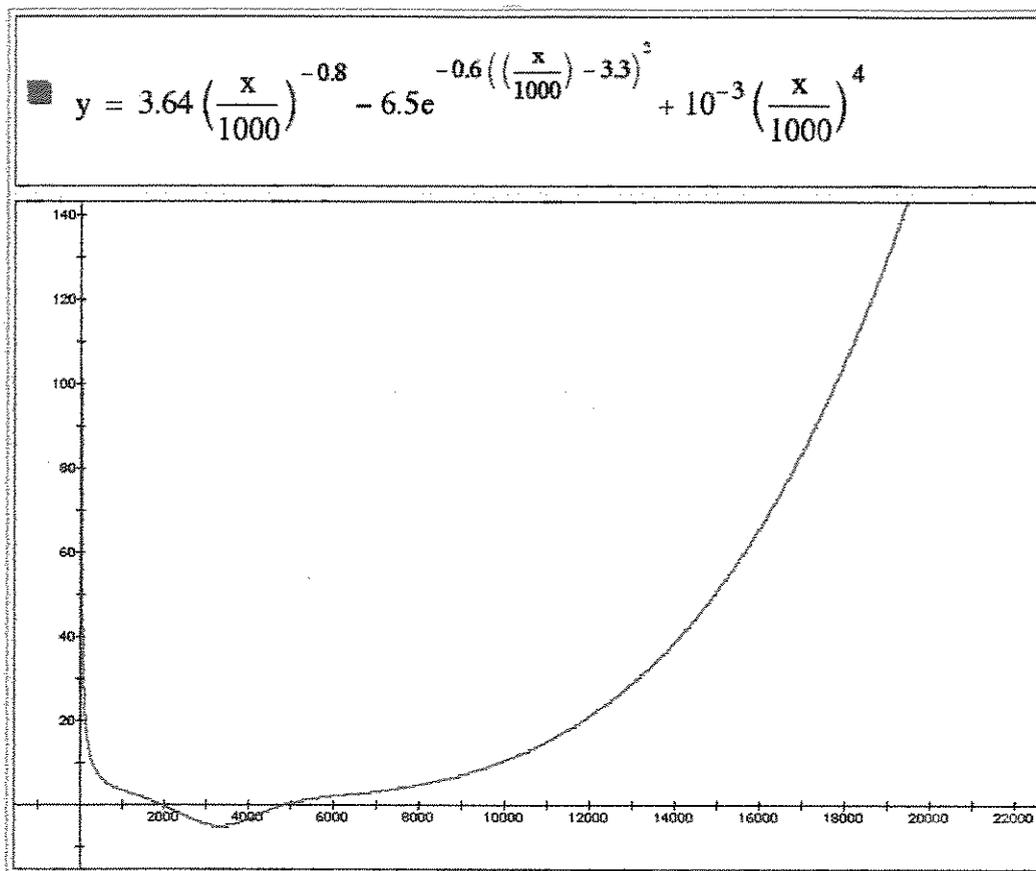


Figura 5.9 Expressão matemática para ATH e a curva gerada por ela. Eixo x representa frequência em Hertz e eixo y representa nível de pressão sonora em dB²⁴.

Nenhum som que tenha amplitude abaixo da linha mostrada na figura 5.9 será percebido isoladamente. A segunda vantagem da implementação dessa ‘filtragem psicoacústica’ é a maior nitidez do som resultante ligada ao sistema de alto-falantes na difusão. Quando um som é tocado através de um alto-falante, o que acontece neste é uma vibração resultante da soma de todos os seus sons constitutivos simultâneos. Quanto mais rico um espectro, maior será a complexidade da vibração e mais difícil será a perfeita reprodução através de um sistema de alto-falantes. É por isso que os melhores sistemas possuem duas ou três vias, ou seja, dois ou três alto-falantes com amplificação própria reproduzindo uma faixa específica de frequências sem interferência (ou com o mínimo de interferência) das outras

²⁴ Chamado dB SPL (*sound pressure level*).

faixas. A eliminação de componentes que não são percebidos pode, dessa maneira, aumentar a nitidez do som final pelo aumento da eficiência da reprodução.

A implementação se deu da seguinte forma:

```

iampdbm = dbamp(p4)           ;conversão para dBm
iampspl = (iampdb*120)/90.308 ;conversão para dB SPL
iamplin = p4                  ;amplitude linear
ifrq = p5                      ;constante de freq.

ithres = (3.64*(ifrq/1000)^(-0.8)) - (6.5*2.71828^(-0.6*((ifrq/1000)-3.3)^2)) + (10^(-3))*(ifrq/1000)^4

iampf = (iampspl < ithres + 4.99 ? 0 : iamplin) ;teste/decisão

```

Figura 5.10 Excerto de *script* de Csound contendo a implementação de ATH.

Considerando que os dados referentes a amplitude estejam em escala linear (as manipulações de amplitude em OpenMusic exportam os dados já convertidos para esse formato) será necessária uma conversão inicial para dBm, pois a expressão de ATH também fornece os resultados em escala logarítmica. Uma primeira conversão de amplitude linear para dBm é realizada pelo *opcode* *dbamp* que cria a constante *iampdbm*. O resultado é agora convertido para dB SPL (que é o tipo de dB empregado em ATH), gerando a constante *iampspl*, que será depois comparada ao resultado de ATH – a constante *ithresh*. Para ser calculado o ATH é necessário fornecer o valor da frequência, da qual se deseja saber se será percebida ou não (*ifrq*), à expressão exposta na figura 5.9 e realizar uma pequena adaptação que será explicada mais tarde. A princípio tem-se que assumir um pressuposto na associação de dBm com dB SPL. A unidade dBm, em Csound, atinge seu máximo com o valor aproximado de 90.308 porque é esse o equivalente ao máximo possível de representação no sistema digital da escala linear de amplitude (32767) e o mínimo de 0 que equivale a 1 na escala linear²⁵. Já para a unidade dB SPL não há um máximo, pois ela não se limita à representação do sistema digital e, portanto, pode-se sempre representar um som com um maior nível de pressão sonora. Normalmente assume-se um limite máximo de 120 dB que equivale, segundo a maior

²⁵ Segundo a conversão realizada pelo módulo de OpenMusic *db->lin*, parte integrante do pacote *om2csound*.

parte das tabelas consultadas, ao nosso limiar da dor. O pressuposto assumido é que o limiar da dor, 120 dB SPL, equivalerá à representação máxima de amplitude no sistema digital, 90.308 dBm. Para isso é necessário um mapeamento que é executado para se obter o novo valor, representado pela constante *iampspl* (figura 5.10). Voltemos ao ATH. Como pode ser visto na figura 5.9, uma parte da linha que traça o limiar de audibilidade cruza o eixo x e fornece números negativos de y naqueles pontos. Isso se deve ao fato que a referência adotada para a determinação do ponto zero do nível de pressão sonora²⁶ (o eixo x) é o limiar de audibilidade à 1000 Hz. Nessa frequência, a intensidade é de 10^{-12} W/m² ou pressão de 2×10^{-5} Newton/m². Contudo, há faixas de frequência, mais ou menos entre 1000 e 5000 Hz, com intensidade ainda menor que 10^{-12} W/m² que são possíveis de serem percebidas. Por isso a curva cruza o ponto zero de y no gráfico que representa a expressão de ATH. E que influência esse fato vai exercer na implementação que eu proponho? Não se pode considerar a comparação de um valor negativo em escalas que partem do zero. O que se fez foi determinar o mínimo atingido de y e somar esse valor à constante *ithresh* para que não tenhamos nenhum nível de pressão sonora menor que zero. Essa soma é realizada na comparação feita entre *iampspl* e *ithresh* para testar os valores de entrada. Se o valor de amplitude de entrada estiver abaixo do valor delineado pela curva de ATH + 4.99²⁷, então ele será igualado a zero, que é o mesmo que eliminá-lo. Se não for assim, o valor de amplitude de entrada é assumido e utilizado, mas em escala linear (*iamplin*), que é o padrão de utilização do Csound.

5.3.2.2 Jitter

O *jitter*, ou flutuações nas frequências das componentes de um som, foi implementado em Csound usando dois diferentes *opcodes*: *jitter2* e *jspline*. Cada *opcode* possui uma particularidade quanto à maneira que eles compõem as variações e como elas se dão no tempo. A escolha de dois *opcodes* com comportamentos diferentes para a implementação do efeito *jitter* se justifica nas diferentes características das flutuações de

²⁶ *Sound pressure level* ou SPL.

²⁷ 4.99 é o arredondamento para o valor mínimo absoluto encontrado para y, que somado ao *ithresh*, evitará valores negativos.

freqüência dos parciais em regiões distintas do espectro. Dependendo de qual região do espectro um determinado parcial faz parte, ele será encaminhado para um setor no *script*²⁸ que executará um algoritmo particular. O espectro foi dividido em três regiões: região aguda (*High*), com freqüências iguais ou maiores que 5000 Hz, região média (*Med*), com freqüências maiores que 900 e menores que 5000 Hz e região grave (*Low*), com freqüências iguais ou menores que 900 Hz. Para o endereçamento dos parciais para a região correta, foi usada a declaração condicional *if...goto* que, a partir de uma expressão lógica, direciona os dados para um determinado *label* (trecho do *script* que assume um nome de identificação). Segue abaixo a determinação das condicionais.

```
if (p5>=5000) goto High
if (p5>900 && p5<5000) goto Med
if (p5<=900) goto Low
```

Figura 5.11 Endereçamento de parciais para as três regiões do espectro.

Para a região aguda foi usado o opcode *jitter2*, que funciona como uma somatória de três seqüências de segmentos randômicos com variáveis próprias de amplitude e freqüência. A amplitude em *jitter* se refere à largura máxima da variação dos dados e a freqüência se refere à ‘velocidade’ de alteração dos dados. Tomemos um exemplo cujas variações referentes à amplitude tenham valor 1 (o que significa que a variação acontecerá entre -1 e 1) e cuja freqüência de variação seja de 4 Hz. Pedi uma leitura de resultados a cada 0.05 segundo, ou seja, 20 leituras por segundo. Dá para notar um novo número a cada 0.25 segundo (4 Hz) com interpolação linear entre eles (a cada leitura). A variação do número em si ficará no âmbito -1 a 1.

²⁸ *Scripts* são arquivos-texto que contém as programações em Csound que servirão de base para a obtenção dos resultados esperados, neste caso, sons.

tempo	valor	
0.00:	0.23661	; novo valor
0.05:	0.29773	
0.10:	0.36176	
0.15:	0.42579	
0.20:	0.48982	
0.25:	0.55385	; novo valor
0.30:	0.30869	
0.35:	0.04490	
0.40:	-0.21889	
0.45:	-0.48268	
0.50:	-0.74647	; novo valor
0.55:	-0.62881	
0.60:	-0.48323	
0.65:	-0.33766	
0.70:	-0.19208	
0.75:	-0.04650	; novo valor
0.80:	-0.10577	
0.85:	-0.18275	
0.90:	-0.25973	
0.95:	-0.33671	
1.00:	-0.41368	; novo valor
1.05:	-0.26080	
1.10:	-0.09329	
1.15:	0.07421	
1.20:	0.24172	
1.25:	0.40923	; novo valor
1.30:	0.27622	
1.35:	0.12510	
1.40:	-0.02602	
1.45:	-0.17714	
1.50:	-0.32826	; novo valor
1.55:	-0.39654	
1.60:	-0.45875	
1.65:	-0.52096	
1.70:	-0.58317	
1.75:	-0.64538	; novo valor
1.80:	-0.55890	
1.85:	-0.45955	
1.90:	-0.36021	
1.95:	-0.26086	
2.00:	-0.16152	; novo valor

Figura 5.12 Lista de valores gerados segundo os parâmetros de *jitter*.

Este foi um exemplo da geração de uma única seqüência randômica que será somada à frequência original de cada parcial. O *opcode jitter2* fornece três destas seqüências randômicas simultaneamente com parâmetros individualizados, sendo que se pode fornecer para cada parâmetro uma variação dinâmica em *krate*²⁹. A implementação do setor *High* é mostrada abaixo.

²⁹ *krate* é a taxa de variação de um controlador fixada no cabeçalho do *script* de Csound. Cada *script* pode conter uma taxa diferente conforme a necessidade. No *script* em questão foi determinada uma taxa de 11025 Hz, ou seja, 11025 novos dados por segundo.

```

High:
kampsum init 3 ;soma total de ampli.
kamp1 init .85 ;ampl. máxima da primeira seqüência
kamp2 init 1.15 ;ampl. máxima da segunda seqüência
kamp3 init 1 ;ampl. máxima da terceira seqüência
kfrq1 linseg 3,p3*.5,5,p3*.5,2.7 ;variação linear de freq. da primeira seq.
kfrq2 linseg 5,p3*.5,2.9,p3*.5,5.3 ;variação linear de freq. da segunda seq.
kfrq3 expseg 4,p3*.5,6.3,p3*.5,3.9 ;variação exponencial de freq. da terceira seq.

kjit jitter2 kampsum,kamp1,kfrq1,kamp2,kfrq2,kamp3,kfrq3

goto Shimmer

```

Figura 5.13 Utilização do *opcode jitter2* com a composição de seus parâmetros.

Foi escolhido um *opcode* que gera três seqüências randômicas porque os parciais mais agudos possuem maior taxa e complexidade de flutuação de frequência que os parciais mais graves. No caso das regiões média e grave, foi escolhido um *opcode* com outras características. O *opcode jspline* provê apenas uma seqüência randômica de números, mas com interpolações ‘curvas’ em vez de lineares. No exemplo criado, os parâmetros não variam no tempo, como realizado no exemplo anterior (*kfrq1*, *kfrq2* e *kfrq3*), mas há uma possibilidade extra fornecida pelo *jspline*: o estabelecimento de um valor mínimo e um valor máximo de frequência de geração dos novos dados, nomeados aqui *kfrqmin* e *kfrqmax*. Uma diferença, entretanto, foi estabelecida entre as regiões média e grave. A amplitude de variação e o âmbito da frequência de variação são menores na região grave, pois é característico das componentes com frequências baixas variações menores e mais lentas.

```

Med:
kamp init 2
kfrqmin init 2.5
kfrqmax init 4.5

kjit jspline kamp,kfrqmin,kfrqmax

goto Shimmer          ;endereça os dados resultantes para o próximo passo

Low:
kamp init 1.5
kfrqmin init 1.7
kfrqmax init 3.2

kjit jspline kamp,kfrqmin,kfrqmax

goto Shimmer          ;endereça os dados resultantes para o próximo passo

```

Figura 5.14 Utilização do *opcode jspline* com a composição de seus parâmetros nas regiões média e grave.

Apesar da aparente complexidade na composição das flutuações de frequência dos parciais, a proposta exposta aqui de implementação de *jitter* é escassa de detalhes que criariam uma modelização mais fidedigna. Um único parcial real pode apresentar diferentes taxas e larguras de variação dependendo do ‘momento’ em que ele se encontra dentro de seu envelope dinâmico. As características de *jitter* durante o ataque do som diferem daquelas da sustentação e da extinção do mesmo som. Os meios para se criar um modelo rico em detalhes são os mesmos apresentados acima, mas não achei conveniente me estender e complicar em demasiado a implementação de uma única particularidade do som. Em vez disso, preferi dar cabo de cobrir um maior número de processos indicando os meios para que eles possam ser desenvolvidos.

5.3.2.3 Shimmer

As flutuações ou micro-variações no desenvolvimento das amplitudes das componentes espectrais são conhecidas como *shimmer*, ou seja, é o correspondente do *jitter* para as amplitudes. Sua implementação seguiu os mesmos passos da implementação de *jitter*: determinação das condicionais que endereçam cada componente para um setor específico, geração dos dados apropriados para cada setor e reendereçamento para o

processo seguinte, o que muda são os parâmetros e o propósito. Aqui o espectro se dividiu em duas partes: frequências iguais ou maiores que 4000 Hz e frequências menores que 4000 Hz. Cada uma dessas regiões do espectro deu origem a um setor que aplicou o *shimmer* de uma maneira diferente. Ambos os setores, *Fast* (região aguda) e *Slow* (região grave), se utilizaram do *opcode jspline* com variações dinâmicas de amplitude e frequência.

```

Shimmer:
if (p5>=4000) goto Fast
if (p5<4000) goto Slow

Fast:
;variações exponenciais de ampli. e freq. mínima e máxima
kamp expon 3,p3,1.2
kfrqmin expseg 4,p3*.2,3,p3*.5,3,p3*.3,1.8
kfrqmax expseg 7,p3*.2,5,p3*.5,5,p3*.3,3.5
kshim jspline kamp,kfrqmin,kfrqmax
goto Synth ;endereçamento para o processo seguinte

Slow:
;variações lineares de freq. mínima e máxima e exponencial de ampli.
kamp expon 2.3,p3,1.8
kfrqmin line 2.3,p3,1.8
kfrqmax line 3.8,p3,2.5
kshim jspline kamp,kfrqmin,kfrqmax
goto Synth ;endereçamento para o processo seguinte

```

Figura 5.15 Implementação de todo o processo de *shimmer*.

As variações dinâmicas de frequência dos setores *Fast* e *Slow* diferem em dois pontos: o tipo de variação de frequência e a correspondência ou não com o envelope. Para a região aguda, as variações são exponenciais e sugerem um modelo de envelope. Essas variações são divididas em três partes, duas das quais, a primeira e a última, denotam uma diminuição progressiva de frequência, enquanto o segmento central se mantém inalterado. Essa é uma sugestão de simulação de variação de frequência do *shimmer*

seguindo uma dependência do envelope dinâmico do parcial, mas não pretende ser fiel³⁰, e sim, experimental. Já a região grave apresenta apenas um segmento linear em desaceleração como variação de frequência, além dos mínimos e máximos serem menores que na outra região. No final de cada parte acontece o endereçamento para o processo final de síntese, no qual se agrupam todos os outros parâmetros.

5.3.2.4 Distribuição Espacial

A implementação da distribuição espacial, conforme o conceito apresentado no capítulo 4, se dá através da utilização do *opcode pan*. Esse *opcode* permite o posicionamento ou movimento de um som no espaço cartesiano inscrito em meio à quatro fontes sonoras. Para isso, são requeridos os seguintes parâmetros: o sinal sonoro que se deseja posicionar, o fator de posicionamento lateral, relativo ao eixo esquerda-direita do ouvinte, o fator de profundidade, relativo ao eixo frente-fundo, o número da *function table* na qual é determinado um padrão de variação de amplitude em cada canal e um fator opcional, usado, que determina a normalização do sinal de entrada. Ambos eixos, lateral e de profundidade, são determinados por fatores no âmbito 0 a 1, 0 significando totalmente à esquerda e totalmente atrás e 1 significando totalmente à direita e totalmente à frente. Como estamos trabalhando unicamente com um par estereofônico frontal das fontes sonoras, o fator de profundidade será fixado em 1 para que só haja distribuição lateral. Os dois outros canais restantes dos quatro indicados pelo *opcode pan*, são descartados.

³⁰ A modelização dessas variações, como já dito, seria extremamente complexa e não encontra consenso na literatura.

```

asynth oscil3 kenw,p5,1,p6
ked init p7
ktf init 1

aesq,adir,a3,a4 pan asynth,ked,ktf,4,1

```

Figura 5.16 Aplicação de distribuição espacial através do *opcode pan*. O sinal *asynth* é distribuído no espaço seguindo os fatores *ked* (controlador esquerda-direita) e *ktf* (controlador trás-frente) resultando nos dois canais de áudio *aesq* e *adir*.

5.3.2.5 Senóides + Ruído

Uma parte importante dos sons é aquela que não pode ser analisada corretamente por meio da FFT e seus derivados ou processos de análise correlatos. É o que costuma ser chamado de residual de um som. Esse residual consiste na parte ‘instável’ do som, que pode ser identificado tanto como transientes quanto como ruídos que integram o evento sonoro. Como já explicado antes, há técnicas de análise/síntese que combinam a síntese aditiva e a síntese subtrativa com o intuito de integrar a recomposição determinística do som com a parte estocástica. Esses processos híbridos, que são chamados normalmente de modelo de senóides mais ruído ou determinístico mais estocástico, são elaborados de maneiras diferentes, utilizam algoritmos variados ou com adaptações e são muito complexos de serem programados. A implementação de um processo como esse exigiria uma grande experiência de programação e conhecimento técnico específico de DSP³¹, o que extrapola muito os objetivos deste trabalho. Assim, duas opções vieram à tona: utilizar um outro *software* que disponibilizasse esse recurso ou criar uma versão desse processo a partir dos conhecimentos e dos recursos dos quais já se dispõe. A opção de utilizar um outro *software* era possível, mas apresentava dois problemas: a falta de maleabilidade no emprego desse processo inserido no contexto do *software* (o SMS³²) e a expansão sempre crescente dos recursos utilizados quando do aparecimento de um novo problema. Uma configuração inicial dos recursos utilizados durante os estudos precisa ser

³¹ *Digital signal processing* – processamento de sinal digital.

³² O SMS é um pacote de processos que faz parte de um projeto maior chamado CLAM – *C++ Library for Audio and Music* que é gratuito, público e é desenvolvido por uma série de pessoas ligadas ao projeto.

mantida por questões de coerência, melhor exploração das qualidades que cada um oferece, facilidade decorrente de não precisar se deparar sempre com um novo sistema informático e ter que aprendê-lo, etc. Não se pode lançar mão de uma nova ferramenta toda vez que se muda o enfoque dos estudos ou que surge um novo, isso não teria fim e seria pouco prático. A segunda opção, adotada, não está liberta de desafios e obstáculos, mas pode, à medida do possível, ser moldada para o contexto apresentado, integrada e relacionada a outros processos. Dos obstáculos, os que mais se destacam são a detecção de fases e a falta de um recurso técnico de obtenção do ruído a partir do som original. Uma das maneiras de execução desse processo se encontra no modelo de senóides + ruído de Xavier Serra (SERRA, Xavier, 1997), mas ele traz tantos desafios técnicos que seria impossível implementar sem ser em linguagem baixa de programação. Outro ponto problemático seria como proceder à subtração, no domínio temporal, da síntese aditiva da parte determinística a partir do som completo sem a reconstrução exata das formas de onda. A alternativa encontrada foi criar um processo que mantivesse a idéia da mistura das partes determinística e estocástica que se completassem em um 'todo'. A síntese sonora que gera a parte determinística, já explicada, será toda baseada na síntese aditiva com os parâmetros fornecidos pelo OpenMusic. Já a criação da parte residual ou ruidosa mereceu uma especial atenção e esse vai ser o enfoque do texto que segue.

O ruído pode ser obtido através de modelagem de suas características e posterior síntese subtrativa utilizando ruído branco (processo que envolve muitas etapas) ou a criação de toda a parte determinística e a sua subtração do som completo no domínio temporal ou frequencial³³. O processo que empreguei é uma adaptação do segundo processo descrito. A adaptação envolveu a subtração da parte determinística sintetizada através de filtragens *bandreject* com parâmetros fornecidos por uma análise de *phase vocoder*. O processo todo consiste no seguinte:

1. realizar uma análise de *phase vocoder* gerando um arquivo de análise;
2. ler as frequências de cada *bin* desse arquivo de análise;
3. carregar o som original para a memória do Csound;

³³ Há várias nuances nesses processos que estão sendo voluntariamente omitidas das descrições devido ao alto grau de complexidade envolvida. Há também formas alternativas de obtenção do residual de um som, mas essas apresentadas acima fornecem as principais idéias.

4. criar um filtro *bandreject* que elimine do som original cada frequência lida da análise;
5. enviar o resultado para o processo seguinte ou para a saída de áudio.

A princípio há dois grandes problemas aqui. Um diz respeito à leitura da análise e envio dos dados para a filtragem. Se eu realizo uma análise de *phase vocoder* com número de FFT igual a 4096, significa que eu terei 4096 canais ou *bins* para ler enquanto a estrutura de um *script* de Csound exige que se indique ‘nominalmente’ qual *bin* se quer ler. Isso não seria necessário se os dados de análise estivessem alocados de forma conveniente no ‘*score*’, mas a leitura de tantos dados de análise de *phase vocoder* só pode ser realizada no ‘*orchestra*’, e tem que ser feita dado por dado. Para não termos que criar 4096 leitores (ou qualquer outro número de FFT empregado nas análises) temos que nos valer da idéia de recursividade, o que será fundamental também para filtrarmos o ‘mesmo som’ 4096 vezes. Se não utilizássemos recursividade na filtragem, estaríamos filtrando 4096 sons originais, cada um deles por uma frequência derivada de um *bin*. No final, teríamos a soma de 4096 sons com uma única e diferente frequência filtrada em cada um. O que é preciso é que o mesmo som seja filtrado a cada vez pela frequência de um diferente *bin* até se eliminar todos os sons analisados pela FFT. A recursividade é aplicada através de um pequeno recurso de programação em linguagem de Csound, dentro do próprio *script*, chamado ‘*opcode*’³⁴. O segundo problema de destaque ligado ao processo proposto diz respeito à associação de dois processos que não possuem uma correspondência direta entre eles, a saber, a análise de parciais que gerou os dados para a síntese aditiva e a análise de *phase vocoder* usada para extração do residual do som. A grande diferença entre uma análise de parciais, que é baseada no processo de FFT, e uma análise de *phase vocoder*, baseada também no mesmo processo, é que a análise de parciais integra um estágio chamado *peak continuation* (continuidade de pico). Nesse estágio são identificadas frequências associadas a picos de amplitude que tenham uma continuidade no tempo criando trajetórias de parciais ‘estáveis’. As trajetórias traçadas fornecem então

³⁴ Este é um ponto que pode causar confusão. *Opcode* tem o sentido de uma rotina que exerce uma determinada função em Csound a qual nos referimos através de um nome que lhe é associado, como exposto antes, mas ‘*opcode*’ também é o nome de um recurso de programação empregado dentro do próprio *script* em linguagem de Csound. Para diferenciá-los, quando se fizer referência ao recurso de programação dentro do *script*, empregarei a palavra ‘*opcode*’ entre aspas simples.

os dados que são exportados na análise de parciais do AudioSculpt enquanto na análise de *phase vocoder*, todos os dados detectados, possuindo uma continuidade ou não, são exportados. O problema surge, dessa maneira, no entrelaçamento dos resultados de dois processos distintos: supressão de todas as frequências identificadas no *phase vocoder* para a obtenção do residual em conjunção com a síntese só de parciais estáveis. O resultado não é, portanto, uma recomposição exata das partes determinísticas e estocásticas de um som, mas a composição de cada uma dessas partes sob um ponto de vista diferente. A fim de evitar grande disparidade nesta recomposição de partes, foi executada uma preparação do som antes da análise de *phase vocoder*. A preparação consiste em eliminar as componentes mais instáveis do som para que a análise de *phase vocoder* da parte restante se aproxime da idéia de estabilidade presente nas análises de parciais. Essa preparação é realizada no *software* SoundHack, que possui um recurso chamado *Spectral Extractor* (ou *Spectral Rate Extractor*) que opera a separação descrita acima separando parciais que possuem uma variação média maior que 5 Hz por janela de análise, denominados transientes, dos parciais que possuem uma variação média menor que 3 Hz, denominados sons estáveis³⁵. Nesse processo de ‘extração espectral’ do SoundHack também podem ser estabelecidos diferentes valores para os parâmetros de análise. O *opcode* utilizado para filtragem foi o *pareq*, na verdade um equalizador paramétrico que pode ser configurado para ser um filtro do tipo *bandreject*. Inicialmente foi considerada a utilização do filtro *bandreject* do tipo Butterworth³⁶, mas ele não leva em conta as amplitudes dos parciais a serem filtrados, realizando um corte igualmente acentuado independentemente da intensidade do parcial original. Já o *opcode* *pareq* pede as especificações completas sobre o corte a ser efetuado: frequência de centro, amplitude e fator Q . A frequência de centro é a frequência que se deseja subtrair (no caso de rejeição de banda), a amplitude é o quanto se deseja adicionar ou subtrair e o fator Q é a razão entre a frequência de centro e a largura de banda em Hertz. Muitas vezes se utiliza

³⁵ Os valores de 5 Hz e 3 Hz são apenas valores sugeridos e que podem ser modificados conforme o interesse.

³⁶ Os filtros Butterworth são implementações de algoritmos bastante eficientes e precisos no corte realizado, mas tomam como base apenas frequência de centro e largura de banda. Os filtros da série Butterworth em Csound são de segunda ordem e atenuam 12 dB por oitava (BUTTERWORTH, 2004). Para mais detalhes sobre filtros, ver DODGE, 1985.

o fator Q em vez de largura de banda porque essa é uma forma de assegurar uma mesma relação entre ambas.

$$Q = Fc/BW$$

Onde Fc é a frequência de centro e BW é a largura de banda (*bandwidth*).

Se mantivermos Q constante enquanto mudamos a frequência de centro, estaremos ajustando automaticamente a largura de banda de forma proporcional. Foi dito acima que o parâmetro referente à amplitude no *opcode pareq* determina adição ou subtração do valor estipulado. O parâmetro de amplitude determinará a adição se seu valor for maior que 1, manterá a amplitude constante se o valor for igual a 1 e subtração se esse valor for menor que 1, mas acima de 0. A leitura dos dados derivados da análise de *phase vocoder* é realizada com o *opcode pvread* e fornece frequência e amplitude do *bin* indicado, a amplitude já na escala linear, ou seja, variando de 0 a 32767. As componentes com valor de amplitude acima de 1 precisam ser ajustadas, quando usadas como fatores do *opcode pareq*, para valores menores que 1, se não, ocorrerá adição de amplitude em vez de subtração e transformaremos o filtro *bandreject* em um *booster*, ou seja, em um amplificador das frequências escolhidas. Podemos fazer os cálculos de adaptação desses valores tomando como base a conversão de dB para a escala linear de amplitude. Se queremos somar 20 dB, usamos o valor 10 em amplitude linear e se queremos subtrair 20 dB, usamos o valor 0.1, o que corresponde à inversão na escala linear: 10^{-1} ou $1/10$. Então os cálculos para adaptarmos valores de amplitude maiores que 1 lidos através da análise através do *pvread* se resumem a invertermos os valores encontrados. Se denominarmos *kamp* a variável correspondente à amplitude lida, podemos estabelecer que $kamp = 1/kamp$. Mas para a amplitude total de um som ser igual ou menor que 32767, cada componente do espectro deve ter uma colaboração pequena de amplitude (tanto menor quanto mais rico for o espectro) e que muitas vezes pode se aproximar de zero. Portanto, muito provavelmente haverá valores de amplitude abaixo de 1 que, se invertermos, constituirá uma adição de amplitude. O que foi feito para solucionar esse impasse foi

estabelecer uma dependência através de operadores condicionais e operadores lógicos em uma linguagem própria do Csound:

$$\mathbf{kamp = (kamp > 1 ? 1/kamp : 1)}$$

Essa expressão significa que se a variável *kamp* for maior que 1, ela será substituída pelo seu inverso ($1/kamp$), se não, ela passará a valer 1, que implica em nenhuma mudança. Se as amplitudes lidas da análise forem menores que 1 em escala linear, elas serão tão pequenas que podem ser ignoradas, por isso a utilização do fator 1.

```
opcode FFTFilter, a, akkpp
asig, ktime, kq, idep, icnt  xin
  if (icnt >= idep) goto next
asig      FFTFilter  asig, ktime, kq, idep, icnt+1
kfrq, kamp  pvread  ktime, "d.pvc", icnt
kamp = (kamp > 1 ? 1/kamp : 1)
asig      pareq  asig, kfrq, kamp, kq
next:
  xout  asig
endop

instr 1
ktime  line 0, p3, p3
kq  init 15
asigf      soundin  "didge1b"
afilt  FFTFilter  asigf, ktime, kq, 1022, 3 ; início no bin 3
out  afilt
endin
```

Figura 5.17 Implementação de leitura e filtragem recursivas através do ‘opcode’.

O filtro *pareq* também tem que estar dentro da estrutura do ‘opcode’ para poder filtrar o som original iterativamente, mas esta é só a definição do que vai acontecer. Mais tarde, dentro do instrumento (*instr 1*), esta definição será usada com o fornecimento dos parâmetros necessários, como a variável na qual está contido o som (nomeada aqui *asigf*),

como se dará a leitura do som no tempo, ou seja, leituras na velocidade e na ordem normal (não reversa (*ktime*)), o fator Q , variável ou não no tempo (*kq*), o número de *bins* a serem lidos pelo *pvread* (1022 neste caso) e em qual *bin* começará a leitura (3). As razões para se iniciar a leitura dos *bins* pelo terceiro se encontra nos erros da análise de *phase vocoder* (através da FFT). Os arquivos de análise contêm distorções de frequência nos dois primeiros *bins* indicando uma série de números negativos e saltos, por isso o início a partir do *bin* 3. Com relação à indicação de leitura de apenas 1022 *bins*, em vez dos 1024 contidos na análise, pode-se entender que, se começamos nossa leitura a partir do terceiro, sobram apenas 1022 *bins* para serem lidos, já que os dois primeiros são descartados.



Som 5.12, 5.13, 5.14 e 5.15 (faixa 17) Sequência de quatro exemplos sonoros baseados no som produzido por um didgeridoo.

O primeiro som do exemplo acima (5.12) corresponde ao som inalterado de um didgeridoo. O segundo exemplo (5.13) consiste na extração das componentes mais instáveis do som original, de caráter ruidoso, através do processo *Spectral Extractor* do software SoundHack e da ressíntese através de *phase vocoder* (utilizando o *opcode pvoc* do próprio Csound). Este exemplo mostra a qualidade do som que resulta do descarte de sua parte ruidosa. É claro que, se tomarmos sons bastante harmônicos e estáveis como exemplo, a diferença gerada pelo descarte da parte ruidosa será pequena ou até mesmo desprezível, pois quase nada será subtraído do som original pelo *Spectral Extractor* e portanto praticamente perfeita será a ressíntese através de *phase vocoder*³⁷. Mas o som do didgeridoo possui componentes bastante instáveis e uma análise do som original sem a eliminação prévia do conteúdo ruidoso levaria à geração de muitos erros de ‘má interpretação’ no processo de análise/ressíntese, como exposto no capítulo 2. O terceiro som do exemplo (5.14) constitui-se apenas da parte ruidosa do didgeridoo extraída pelo método aqui sugerido de filtragem *bandreject* iterativa (implementado pelo par ‘*opcode*’ – *pareq*), como exposto na figura 5.17. O quarto som (5.15) é a verificação da

³⁷ Considerando-se que os parâmetros fornecidos para a análise e ressíntese sejam os mais adequados para o tipo de som escolhido e os algoritmos empregados nesses processos sejam eficientes.

funcionalidade do processo de extração de ruídos. Para isso foi executada a soma balanceada dos dois exemplos sonoros anteriores – a ressíntese de *phase vocoder* sem ruídos e os ruídos extraídos do som original – e que pode ser comparada ao som original do didgeridoo encontrado no exemplo 5.12. O resultado sonoro foi julgado bastante bom apesar da extensão e complexidade do processo todo.

5.3.2.6 Hibridização

O processo de hibridização proposto aqui não tem o sentido que normalmente lhe é associado, a saber, a junção de características de dois sons distintos, em forma de parâmetros, para formar um terceiro som ‘híbrido’. Esse sentido mais comum é muitas vezes chamado de *cross-synthesis* (síntese cruzada). A palavra híbrido, na utilização que se faz neste trabalho, tem a acepção ligada à criação de uma ‘natureza híbrida’ do som quanto ao seu caráter de naturalidade ou artificialidade. Pode-se entender essa hibridização como uma mistura de um som gravado e de uma síntese aditiva que normalmente dele é derivada. Esse tipo de mistura oferece uma pletora de possibilidades que não poderiam ser descritas aqui em sua totalidade. Algumas sugestões serão apontadas e exemplificadas oriundas do meu próprio trabalho.

A hibridização, na verdade, poderia ser vista como a interação som instrumental – som eletrônico, sobre a qual este trabalho é baseado, em tempo diferido. Não quer dizer que as interações propostas aqui até agora sejam fruto de processos de produção ou transformação eletrônica em tempo real, o que acontece é que as ‘interações’ se dão em tempo real, porque os sons de ambos domínios só se misturam no ato da performance, no momento em que serão apreendidos pelos ouvintes. A diferença que se faz perceptível entre as propostas realizadas até este momento e a hibridização é a possibilidade de manipulação antecipada do som instrumental em vez de contar com a junção dos sons instrumental e eletrônico apenas no momento da performance. Quando o som instrumental gravado usado na mistura com a síntese aditiva não é pré-processado, o resultado sonoro é praticamente o mesmo da mistura com o som instrumental performado ao vivo. Distinções claras podem ser encontradas na intenção composicional, no processo

de escritura, na estruturação da obra, que conduzirá o compositor na escolha da melhor ou mais apropriada opção.

As possibilidades da hibridização podem ser separadas em três grandes grupos:

- 1) a mistura do som de um instrumento com a síntese aditiva derivada da análise desse mesmo som
- 2) a mistura do som de um instrumento com a síntese aditiva derivada da análise do som de um outro instrumento reconhecível
- 3) a mistura do som de um instrumento com a síntese aditiva derivada da análise do som de um instrumento não-reconhecível, da mistura de instrumentos ou não derivada de análise

No primeiro caso, as sensações alcançadas com as misturas podem incluir o destaque de certas componentes ou regiões do espectro do som gravado, a expansão das possibilidades naturais do som gravado, alteração das características do som gravado, etc. Os meios de se conseguir esses resultados também são bastante variados, mas sempre incluem a soma do som gravado, em sua forma original ou não, e de síntese aditiva. Alguns exemplos concretos seriam:

- a soma do som gravado em sua forma original com uma síntese aditiva dos parciais mais destacados, de forma que a amplitude da síntese aditiva esteja abaixo de um quinto da amplitude do som gravado, para que haja fusão completa dos sons.
- a soma do som gravado expandido no tempo (através do processo de *time-stretching* baseado em *phase vocoder*³⁸) com a síntese aditiva de parciais com suas frequências trocadas de posição através de permutação e com seus *onsets* e durações expandidas para se ajustar exatamente à nova duração da gravação.
- a soma do som gravado em sua forma original com uma síntese aditiva na qual os parciais tenham início sincrônico com o som gravado e suas durações sejam

³⁸ Esse processo pode ser bastante bem realizado no próprio Csound com a utilização de uma análise de *phase vocoder* feita com o *pvanal* e de sua ressíntese com o *opcode pvoc*.

igualadas ao tamanho desse mesmo som. As amplitudes dos parciais podem variar diferentemente entre si criando flutuações dinâmicas sobre o som gravado.

- a soma do som gravado em sua forma original com uma síntese aditiva na qual os parciais do início do som, podendo ser de um ataque, sejam ‘congelados’ no tempo, assim permanecendo até o final do som ou indo além dele.



Som 5.16 (faixa 18) Som referente ao segundo exemplo exposto acima. Uma seqüência de ataques e trinados de violão é acompanhada pela síntese de seus parciais com frequências permutadas e durações expandidas até o final da seqüência (Frederico Grassano ao violão).



Som 5.17 (faixa 19) Exemplo de som no qual os parciais são ‘congelados’ logo depois do ataque do violão³⁹, neste caso, tocado ao vivo (extraído da peça de minha autoria *Motus Animi*, de 1999, com Frederico Grassano ao violão, para o qual ela foi dedicada).

No segundo caso, o que muda é o caráter da mistura, em vez de somarmos qualidades do mesmo som, somamos qualidades diferentes oriundas de diferentes sons. A utilidade desse processo é mais clara quando consideramos uma obra mista com mais de um instrumento. A referência ao ‘outro’ instrumento poderia se tornar clara na utilização de sua análise na síntese aditiva sobre o primeiro instrumento. Os processos envolvidos podem ser os mesmos daqueles já citados acima, mas o grau de fusão alcançado não seria tão grande por se tratar de sons com características em princípio diferentes⁴⁰. No terceiro caso, o que predomina é a liberdade e variedade de opções, já que qualquer material sintetizado, mesmo que derivado de algoritmos abstratos, pode ser incorporado à mistura.

A parte técnica relativa à implementação da hibridização em Csound pode variar, mas no geral é simples e consiste principalmente da execução dos dois processos

³⁹ Há outros sons presentes além daquele citado no exemplo.

⁴⁰ O maior ou menor grau de fusão obviamente não depende somente do conteúdo espectral do material, permitindo assim que haja possibilidade de se manipular a organização dos parâmetros e da sintaxe do material na composição para que a sensação de fusão seja intensificada.

individualmente (carregar o som original, alterando-o ou não, e proceder à síntese aditiva) e somá-los⁴¹ levando em conta as suas especificidades de tempo e amplitude.



Som 5.18 (faixa 20) Exemplo de hibridização de som de piano filtrado com síntese aditiva não derivada de análise, no início, e com síntese aditiva derivada da análise de outro som, no final⁴² (extraído da peça TEXTVM, de 2000, de minha autoria).

5.3.2.7 Criação do Arquivo de Análise de *Phase Vocoder*

A análise de *phase vocoder*, para ser entendida pelo Csound, precisa usar a ferramenta analisadora do próprio Csound porque há maneiras distintas de se criar o arquivo de análise e elas nem sempre são compatíveis⁴³. Como se pode notar nos *scripts* expostos até agora, as análises (*nome.py* ou *nome.pvc*) aparecem prontas nos *scripts*. O que acontece é que as análises precisam ser realizadas em um processo separado usando *utility programs* (programas utilitários) que podem ser encontrados separadamente ou integrados à própria interface da versão de Csound que se está usando. A interface de Csound que emprego é a versão 1.03 do MacCsound, criada por Matt Ingalls para rodar no sistema OS X Panther da Apple Macintosh e integra a versão 4.23f07 de outubro de 2003⁴⁴. Há uma boa explicação dos *utility programs* no manual do Csound que reproduzo aqui.

Os utilitários de Csound são programas de pré-processamento de arquivos de som que fornecem informações em um arquivo de som ou criam alguma versão analisada dele para serem usados por certos geradores de Csound. Embora

⁴¹ Um processo distinto para a soma dos sons pode não ser necessário se eles forem implementados em instrumentos diferentes e não for imprescindível que eles sejam direcionados para um mesmo instrumento. Dessa forma eles seriam automaticamente somados na saída de áudio.

⁴² Como no exemplo sonoro anterior, aqui também há outros sons presentes que não fazem parte do processo de hibridização.

⁴³ Na verdade, até mesmo diferentes interfaces de Csound podem fornecer arquivos de análise incompatíveis entre si.

⁴⁴ Outra versão mais nova da interface, 1.1b, já está disponível (www.csounds.com/matt/), mas preferi continuar empregando a 1.03 porque muitos testes e processamentos foram efetuados nela e não sei como a nova interface se comportaria. Poderia haver alguma diferença de comportamento ou alguma alteração de sintaxe que prejudicaria o trabalho já realizado.

diferentes em objetivos, eles dividem um mecanismo de acesso comum do arquivo de som e podem ser descritos como um grupo. (VERCOE, 2003, p. 1099)

O programa utilitário usado para se criar análises de *phase vocoder* é o *pvanal*. No MacCsound, o uso do programa é facilitado, pois não requer o uso extensivo de *command lines* (linhas de comando), apesar de termos que definir ainda alguns parâmetros nesse formato. Após abrir o som que se deseja analisar e fornecer os parâmetros extras, como o tamanho da janela de análise (codificado por *-n* nas linhas de comando) e o número de sobreposições das janelas (*-w*), executa-se o programa e ele cria um novo arquivo com o nome indicado. Esse arquivo será usado mais tarde, dentro de um *script*, como fonte de informações das trajetórias de frequências e amplitudes. Quando se fornece o parâmetro referente ao tamanho da janela de análise, como por exemplo 2048 (sempre potência de dois), obtém-se um arquivo que conterá apenas *bins* com a metade do número de amostras na janela (o tamanho da janela) (BOULANGER, 2000, p. 544). O número de *bins* equivaleria ao número de faixas de frequência do espectro analisado, mas se a taxa de amostragem é 44100 Hz, apenas metade desse montante será útil na análise, já que não há como representar mais que 22050 Hz utilizando essa mesma taxa. Então, quando pedimos uma análise de *phase vocoder* com tamanho da janela igual a 2048 amostras (*-n 2048*), estamos definindo também que o arquivo de análise conterá 1024 *bins* que podem ser usados nas leituras feitas pelo *pvread*. Os parâmetros usados nas análises empregados durante este trabalho variaram, no que se refere ao tamanho da janela de análise, dependendo da situação e do som em questão, mas teve o fator de sobreposição sempre mantido fixo em 8 (*-w 8*).

5.3.2.8 Integração dos Processos para a Síntese Final

Finalmente podemos integrar vários dos processos descritos até agora em um *script* com a intenção de dar uma idéia geral e mostrar possíveis conexões entre as partes. É claro que seria uma situação muito forçada a tentativa de integrar todos os processos em um único *script*, pois não se partiria de uma necessidade musical e essa integração

seria complexa demais para ser entendida sem explicações mais minuciosas. Os exemplos abaixo mostram a lógica (o *script* da *orchestra*) de uma possível interação dos vários processos abordados e um pequeno excerto do *score* mostrando os dados exportados pelo OpenMusic.

```
sr=44100
kr=11025
kqaps=4
nchnls=2
gal init 0

opcode FFTFilter, a, akkpp
asig, ktime, kq, idep, icnt  xin
  if (icnt >= idep) goto next
asig      FFTFilter asig, ktime, kq, idep, icnt+1
kfrq, kamp pvread ktime, "d.pvc", icnt
kamp = (kamp > 1 ? 1/kamp : 1)
asig      pareq asig, kfrq, kamp, kq
next:
  xout asig
endop

instr 1
ktime line 0, p3, p3
kq init 15
asig      soundin "didgeib"
afilt FFTFilter asigf, ktime, kq, 1022, 3
outs afilt,afilt
endin

instr 2
if (p3 < .35) goto Quiet
if (p3 >= .35) goto Loud

Quiet:
iamp = aspdb(p4)
iamp = iamp*.75
goto Thresh

Loud:
iamp = aspdb(p4)
iamp = iamp*1.1
goto Thresh
```

continua na página seguinte...

```

Thresh:
iampdb = dbamp(iamp)
iamplin = iamp
ifrq = p5
ift = 1
iph = p6

ithres = (3.64*(ifrq/1000)^(-0.8)) - (6.5*2.71828^(-0.6*((ifrq/1000)-3.3)^2)) + (10^(-3))*(ifrq/1000)^4

iampf = (iampdb < ithres ? 0 : iamplin)

if (p2 < .35) goto SoftTable
if (p2 >= .35) goto HardTable

SoftTable:
kenw oscili iampf,1/p3,3

goto Jitter

HardTable:
kenw oscili iampf,1/p3,2

goto Jitter

Jitter:

if (p5>=5000) goto High
if (p5>900 && p5<5000) goto Med
if (p5<=900) goto Low

High:
kampsu init 3
kamp1 init .85
kamp2 init 1.15
kamp3 init 1
kfrq1 linseg 3,p3*.5,5,p3*.5,2.7
kfrq2 linseg 5,p3*.5,2.9,p3*.5,5.3
kfrq3 expseg 4,p3*.5,6.3,p3*.5,3.9

kjit jitter2 kampsu,kamp1,kfrq1,kamp2,kfrq2,kamp3,kfrq3

goto Shimmer

Med:
kamp init 2
kfrqmin init 2.5
kfrqmax init 4.5

kjit jspline kamp,kfrqmin,kfrqmax

goto Shimmer

```

continua na página seguinte...

```

Low:
kamp init 1.5
kfrqmin init 1.7
kfrqmax init 3.2

kjit jspline kamp,kfrqmin,kfrqmax

goto Shimmer

Shimmer:
if (p5)>=4000 goto Fast
if (p5<4000) goto Slow

Fast:
kamp expon 3,p3,1.2
kfrqmin expseg 4,p3*.2,3,p3*.5,3,p3*.3,1.8
kfrqmax expseg 7,p3*.2,5,p3*.5,5,p3*.3,3.5

kshim jspline kamp,kfrqmin,kfrqmax

goto Synth

Slow:
kamp expon 2.3,p3,1.8
kfrqmin line 2.3,p3,1.8
kfrqmax line 3.8,p3,2.5

kshim jspline kamp,kfrqmin,kfrqmax

goto Synth

Synth:
kfrq init ifrq
klat init p7
kprof init 1
imove = 4

asynth oscil3 kenv+kshim,kfrq+kjit,ift,iph

aesq,adir,a3,a4 pan asynth,klat,kprof,imove,i

aesq = gal + aesq
adir = gal + adir

outs aesq,adir

gal = 0
endin

```

Figura 5.18 Exemplo em Csound da integração de vários dos processos abordados.

```

f1 0 16384 10 1
f2 0 8193 -16 0 164 1 1 573 -1.3 0.73 4833 0 0.73 2522 -1.7 0
f3 0 8193 16 0 2720 -1.3 1 5473 -2.3 0
f4 0 513 7 0 512 1

;p1 p2 p3
i1 0 2.9
i2 0.0580 0.5520 58.5330 947.7830 0.8700
i2 0.0700 0.1220 44.8860 3365.1240 0.5000
i2 0.0750 0.7780 68.8850 657.6340 0.9600
i2 0.0750 0.3200 55.6200 1069.7220 0.7400
i2 0.0750 0.2090 52.8450 1221.6860 0.8800
i2 0.0750 0.4530 47.7660 1366.4360 0.5100
.
.
.
```

Figura 5.19 Excerto de *score* de Csound correspondente à *orchestra* da figura 5.18.

6. Conclusão

CONCLUSÃO

“If we knew what we were doing, it wouldn't be called research, would it?”
Albert Einstein

Este trabalho permite muitas conclusões, sejam estas referentes aos acertos e aos problemas encontrados nas investigações, sejam referentes a perspectivas e apontamentos de direções possíveis a serem tomadas em continuidade a esta pesquisa. Parte das conclusões podem ser tomadas ouvindo-se o CD de exemplos sonoros e julgando-se a pertinência e eficácia de cada proposta ali representada. Um trabalho de composição musical encerrando e ‘coroando’ o trabalho final foi previsto inicialmente e seria de fato desejável, mas em vista do tempo gasto com tamanho volume de dados e um grande número de experimentações, inclusive aquelas de ordem prática ligadas à fatura dos sons, das quais provém o CD, não foi possível alcançar esse objetivo, que se pretende realizado na seqüência das investigações a serem postas em prática no doutorado. Dar-se-á seqüência a algumas considerações teóricas sobre o trabalho.

Uma série de propostas que associam estudos perceptivos à configuração do material eletrônico foi realizada ao lado do levantamento e escolha teórica e empiricamente embasada dos conhecimentos técnicos necessários para a criação de modelos tímbricos instrumentais. Também se passou em revista um conjunto das mais recentes informações acerca da psicoacústica, análise e síntese sonoras e foram sugeridas formas de integração dessas voltadas à composição musical no escopo da música eletroacústica mista. Além disso, as implementações técnicas foram demonstradas de tal forma durante o trabalho que um compositor conhecedor das ferramentas utilizadas e interessado em usar ou levar adiante as propostas aqui apresentadas, terá total condição de fazê-lo.

Problemas encontrados na implementação de algumas propostas e sugestões para possíveis soluções serão agora examinadas, discriminadas por tópicos.

- Quanto à redistribuição da energia do espectro: a alteração da amplitude dos parciais pode levar a resultados muito interessantes com bastante variedade de sonoridades resultantes desde que o som seja suficientemente rico (um som com poucas componentes ou que possua amplitudes próximas não permite muita variação) e a análise de parciais realizada no AudioSculpt seja configurada para selecionar parciais com valores de amplitude menores que -55 dB aproximadamente. Entretanto, parciais originalmente com muito baixa energia são bastante curtos e, quando resintetizados, acabam revelando características pontuais, o que muitas vezes leva à perda de unidade de integração do timbre. Isso poderia ser resolvido de duas formas diferentes: em primeiro lugar, através da configuração dos parâmetros de análise fornecidos no processo de *partial tracking* do AudioSculpt aumentando o tamanho mínimo dos parciais a serem considerados para análise, mas dessa forma corre-se o risco de ampla redução do número de parciais com baixa energia presentes na análise final; e com a adaptação do envelope desses parciais impondo um ataque ‘suave’ sem variações bruscas de amplitude, mas não seria totalmente eficiente mantendo a mesma energia conferida a eles no processo de redistribuição. Um ponto importante a ser implementado seria a variação dinâmica das amplitudes de cada parcial de modo que a soma das amplitudes em cada momento do som resultante se mantivessem inalteradas. Isso criaria sons com modificações interpoladas e contínuas que poderiam levar a gestos sonoros de interesse para o contexto da obra.
- Quanto às micro-variações randômicas dos tempos de início de cada parcial: há um problema relativo à imprecisão temporal da própria análise. A relação temporal de início dos parciais e transientes é muito importante para a correta identificação tímbrica, mas a escala em que se encontram esses tempos são muitas vezes suplantadas pela imprecisão temporal da análise. A precisão temporal acurada tem o custo da imprecisão freqüencial diretamente proporcional causada pela relação dialética tempo-freqüência própria da FFT. Uma solução possível seria a fuga do paradigma da FFT e a adoção das análises de *wavelets*, mas os problemas com esta última não seriam menores.

- Quanto às variações na relação temporal de ADSR dos parciais: essas variações são complicadas no caso do timbre, pois os comportamentos dinâmicos são múltiplos e dependentes do tipo de som e da parte do som onde o parcial se localiza. Pode-se propor uma variedade de tipos de envelope que podem ser empregados e direcionados conforme a necessidade, mas mesmo assim seriam modelos pouco verossímeis. Uma solução ideal seria a modelização real das variações dinâmicas de amplitude de cada parcial, mas seria necessário o emprego de outro *software*, ou mesmo a criação de um, que pudesse cumprir essa função.
- Quanto à quantidade relativa de dados não-determinísticos (transientes e ruídos) acrescentados à parte determinística: apesar da funcionalidade do processo, o ruído extraído não é exatamente complementar ao resultado da síntese aditiva, já que esta se baseia na análise de parciais estáveis (*partial tracking*) e a extração do ruído é derivada da análise de FFT. A proposta seria a implementação de filtros de parciais, mas é necessária a criação de novos algoritmos que executem essa tarefa. Os problemas técnicos para isso incluem programação de uma filtragem recursiva de frequências lidas em arquivo-texto externo ao Csound e sua correta indexação através de *opcodes* apropriados.
- Quanto à mudança de fase dos parciais: tomando como base as técnicas propostas, só é possível criar variações sensoriais distanciadas do som instrumental de origem, pois não se possui o modelo de fases originado dos sons instrumentais. O *software* de análise empregado (AudioSculpt) fornece dados de fase apenas integrados à extensa análise de FFT. Como a recriação em síntese dos modelos sonoros é fundamentada na utilização de parciais derivados do processo de *partial tracking*, não é possível usar os dados de fase, pois os resultados das análises são incompatíveis. As análises de FFT e de *partial tracking* são bastante divergentes em conteúdo, seja esse frequencial, seja de amplitude, pois o primeiro consiste na descrição paramétrica de todos os canais ou *bins* determinados pelo número de FFT, enquanto o segundo vai um passo adiante e seleciona, da análise de FFT previamente realizada, somente aquelas componentes consideradas estáveis o

suficiente para se configurarem parciais¹. Uma sugestão para a resolução desse problema seria a utilização de análises de FFT para as sínteses em vez do emprego das análises de parciais, mas um problema colateral se elevaria: os dados de análise de FFT em forma numérica, tal como exportados do AudioSculpt, precisariam ser filtrados devido ao volume proibitivo dos dados, pois essa carga de informações não é passível de manipulação através do OpenMusic. Embora a outra análise de FFT realizada pelo utilitário PVANAL seja compacta, seu formato codificado não é maleável o bastante, restringindo os âmbito de modificações possíveis.

Uma última consideração a ser feita diz respeito às fontes sonoras das quais se extraem os modelos que serão implementados em síntese. Tratou-se neste trabalho da relação dos meios eletrônicos com a fonte instrumental especificamente, que é aquela para a qual a pesquisa foi direcionada, mas alguns dos processos propostos também se mostram adequados para a exploração dos recursos vocais. Apesar de existirem métodos específicos para a simulação vocal, como a síntese formântica, não se pode descartar o emprego da síntese aditiva como possibilidade alternativa enriquecedora da gama de sonoridades que se podem conseguir a partir do material vocal. Algumas das minhas peças eletroacústicas já fazem uso de sons de caráter vocal obtidas através de síntese aditiva resultante de processos descritos neste trabalho, o que vem me demonstrando a cada vez um caminho promissor a ser investigado.

¹ Segundo os parâmetros fornecidos para a realização do *partial tracking*.

7. Apêndices

Apêndice A – Citações na Língua Original

Capítulo 2 - A Psicoacústica na Interação

Página 23

(MCADAMS, 1996, p. 254) – “The expectancies drive and influence the activation of knowledge structures that affect the way we interpret subsequent sensory information.”

Página 24

(MCADAMS, 1996, p. 255) – “Whereas the nature and organization of these stages are probably similar across cultures in terms of the underlying perceptual and cognitive processing mechanisms involved, the higher-level processes beyond computation of perceptual attributes depend quite strongly on experience and accumulated knowledge that is necessarily culture specific.”

(MCADAMS, 1996, p. 255) – “Auditory comprehension and many aspects of auditory perception are based on generalizations that have been learned from specific experience. These generalizations include implicit knowledge that results from an acculturation within a given acoustic and sociocultural environment. They are abstract in the sense that they are not specific to a given musical pattern that one hears.”

Página 27

(VAGGIONE, 1996, p. 105) – “My position with respect to this is that it is interesting to search for very fine bridges in order to made possible an intimate interaction between the sonic result and the composition process itself, working after the same musical goal, that is, assuming the postulate of a possible convergence, or of a common vectorization.”

(VAGGIONE, 1996, p. 105) – “It is argued often that working with electroacoustic means drives us to an excessive richness of sound, which makes problematic any attempt at integration in a process constituting equally a part of instrumental writing. Thus, from

a point of view contrary to “mixed music”, the electroacoustic sound objects are too complex, and hence not neutral enough to be used in large-scale integrated compositional manipulations. It is also argued that these complex sounds destroy, by the fact of they own singularity, the uniformity that is a necessary condition for a good functioning of a combinatorial process.”

Página 34

(BREGMAN, 1990, p. 642) – “... of deciding how to group them so that each group has been derived from the same environmental event.”

Página 39

(JENSEN, 2001, p. 17) – “The shimmer (irregularity on the amplitude evolution of the partials) adds a quality to the sound, which can be both liveliness, and also additive noise. ... No shimmer makes the sound dead, unnatural, and much shimmer increases the noise and irregularity of the sound.

...

The jitter (irregularity on the frequencies of the partials) adds a different quality than the shimmer to the sound. The jitter gives more low-frequency random pitch variations (with low bandwidth) or it adds roughness (with high bandwidth) ...”

Capítulo 3 - A Análise Paramétrica do Som

Página 47

(ROLAND-MIESZKOWSKI, 2003) - “... the discrete time sequence of a sampled continuous function $\{V(t_n = n \cdot T_s)\}$ contains enough information to reproduce the function $V=V(t)$ exactly provided that the sampling rate ($f_s = 1/T_s$) is at least twice that of the highest frequency contained in the original signal $V(t)$ ”

Página 51

(ZICARELLI, 2004, p. 18) – “Since the quantization error for each sample is usually random ..., we generally hear the effect of quantization error as white noise.”

Página 54

(SERRA, Marie-Hélène, 1997, p. 32) – “Sound analysis encompasses all the techniques that give quantitative descriptions of sound characteristics.”

Páginas 55-56

(ROADS, 1996, p. 545) – “... that complicated vibrations could be analysed as a sum of many simultaneous simple signals. In particular, Fourier proved that any periodic function could be represented as an infinite summation of sine and cosine terms.”

Página 62

(SERRA, Marie-Hélène, 1997, p. 43) – “The ratio F/M (sample rate divided by the DFT size) which separates the frequency bins is called the *analysis frequency* or the *spectral definition* of the DFT. It gives the smallest interval for the distinction between two partials.”

Página 64

(MOMENTUM, 2004) – “In physics, **momentum** is a physical quantity related to the velocity and mass of an object.”

Página 65

(UNCERTAINTY Principle, 2004) – “Consider the following analogy: suppose you have a time-varying signal such as a sound wave, and you want to know the exact frequencies in your signal at an exact moment in time. This is impossible: in order to determine the frequencies accurately, you need to sample the signal for some time and you thereby lose time precision. (In other words, a sound cannot have both a precise time, as in a short pulse, and a precise frequency, as in a continuous pure tone.) The time and frequency of a wave in time are analogous to the position and momentum of a wave in space.”

Página 67

(ROADS, 1996, p. 1108) – “Cooley and Tukey (1965) demonstrated a redundancy in the algebraic structure in the computation of the DFT that could be exploited to speed up the process by several orders of magnitude.”

Página 72

(TRUAX, 1986, p. 157) – “Parameter groupings are just one instance of the introduction of hierarchic levels in a system. Musical structure can usually be thought of as multilevelled, and therefore a computer music system that can address different levels of musical structure effectively will be a powerful compositional tool. And powerful tools lead the creative imagination in new directions. However, the introduction of each level of hierarchy implements a specific musical-acoustic model and ‘prejudices’ the system with a specific hypothesis about how music is structured. The higher the level the greater the prejudice. The lowest level, for instance that of sound samples, is not restricted to music at all (since it can describe any sound), but as each higher-level concept is introduced it contributes to a specific musical model and requires more music-specific knowledge to be implemented in the system.

... general-purpose systems, showing less prejudice because they incorporate less knowledge, are notoriously harder and slower to use since the knowledge they lack (about musical structure) must be embodied in the data provided by the user.”

Capítulo 4 - Simulação e Alteração de Parâmetros Sonoros para a Síntese

Página 111

(SUZUKI, 2004, p. 1) – “... describe the frequency characteristic of sensitivity of our auditory system.”

Página 118

(BOULANGER, 2000, p. 330) – “A sequence of numbers is said to be random if the knowledge of the sequence so far gives no clue as to the next number in the sequence.

Unfortunately when we are using computers we can only create a sequence of numbers by an algorithm or program and that is a deterministic process. The next number can be found from this program and so the sequence cannot be truly random.”

Capítulo 5 - A Síntese

Página 131

(DEPALLE, 1993, p. 122) – “Selon l’acception courante, la synthèse des sons est l’activité qui consiste à créer de nouveaux sons à partir de dispositifs électroniques ou informatiques.”

Página 132

(ROADS, 1996, p. 89) – “UGs are signal processing modules like oscillators, filters, and amplifiers, which can be interconnected to form synthesis *instruments* or *patches* that generate sound signals.”

Página 134

(DEPALLE, 1993, p. 123) – “En synthèse des sons, comme dans toutes les disciplines consistant à simuler et à extrapoler des phénomènes réels, la notion de modèle joue un rôle central et essentiel. Un modèle est une abstraction qui a pour but de condenser de manière formelle l’information véhiculée par le son. Cette information est représentée par un ensemble de valeurs évoluant dans le temps, appelée paramètres, qui identifient le son dans le modèle choisi. Mais un modèle n’est pas seulement un moyen de représenter les sons, c’est également un moyen de définir les caractéristiques générales du dispositif de production qui génère en pratique le son à partir de valeurs de paramètres.”

Páginas 143-144

(SERRA, Xavier, 1997, p. 91) – “Spectral models could be seen as characterizing sound parameters at the basilar membrane of the ear, discarding whatever information the ears seems to discard in the spectrum.”

Páginas 169-170

(VERCOE, 2003, p. 1099) – “The Csound Utilities are *soundfile preprocessing* programs that return information on a soundfile or create some analyzed version of it for use by certain Csound generators. Though different in goals, they share a common soundfile access mechanism and are describable as a set.”

Apêndice B – Relação de Exemplos Sonoros Contidos no CD

Som 4.1, 4.2 e 4.3 – Som de piano original, versão sintetizada do piano e a versão sintética ‘inarmonicizada’ _____	faixa 01
Som 4.4 – Som resultante do processo anterior (inarmônico) transformado de volta em som harmônico _____	faixa 02
Som 4.5, 4.6 e 4.7 – Som de prato tocado com arco, a versão sintetizada e a versão com o espectro expandido _____	faixa 03
Som 4.8 – Som de piano com frequências redistribuídas _____	faixa 04
Som 4.9 e 4.10 – Som de piano ressintetizado com variação randômica dos tempos de ataque e o mesmo som com os tempos de ataque originais _____	faixa 05
Som 4.11, 4.12 e 4.13 – Som de harmônico de piano original, o mesmo som ressintetizado e a versão com redistribuição da energia do espectro _____	faixa 06
Som 4.14 – Som com alterações compassadas de fase das seguintes configurações: em fase, randômica uniforme, fases opostas, distribuição gaussiana e distribuição beta _____	faixa 07
Som 4.15 e 4.16 – Som sintetizado monofônico e o mesmo som com a distribuição estereofônica proposta _____	faixa 08
Som 5.1 – Granulação de voz feminina falando a palavra ‘raiva’ _____	faixa 09
Som 5.2 – Simulação de som de flauta por modelagem física segundo o modelo da figura 5.2 _____	faixa 10
Som 5.3 e 5.4¹ – Som de vozes sem modificação e modulação em anel do mesmo som com um som senoidal com frequência variável de 350 a 320 Hz _	faixa 11
Som 5.5 – Som resultante de um encadeamento de processos básicos de FM _	faixa 12
Som 5.6 – Som senoidal distorcido por <i>waveshaping</i> _____	faixa 13
Som 5.7 – Som resultante de síntese aditiva _____	faixa 14
Som 5.8, 5.9 e 5.10 – Ruído branco e o mesmo ruído com <i>bandreject</i> e com <i>bandpass</i> _____	faixa 15

¹ Som de vozes extraído da gravação da obra *L’Enfant et le Diable* de Aurel Stroe. CD do Ensemble Aleph editado pelo Centre de Creation Sonore, Dammartin, 1992.

Som 5.11 – Som obtido a partir de filtragens <i>bandpass</i> de ruído branco _____	faixa 16
Som 5.12, 5.13, 5.14 e 5.15 – Sequência de quatro exemplos sonoros baseados no som produzido por um didgeridoo _____	faixa 17
Som 5.16 – Sequência de ataques e trinados de violão acompanhada pela síntese de seus parciais com frequências permutadas e durações expandidas __	faixa 18
Som 5.17 – Exemplo de som no qual os parciais são ‘congelados’ logo depois do ataque do violão _____	faixa 19
Som 5.18 – Exemplo de hibridização de som de piano filtrado com síntese aditiva derivada de análises _____	faixa 20

8. Referências Bibliográficas

REFERÊNCIAS BIBLIOGRÁFICAS

- AGON, Carlos et al. *OpenMusic Documentation*. Paris: IRCAM, 2004. 808 p.
- ANDERSEN, T. H.; JENSEN, K. Phase Modeling of Instrument Sounds Based on Psychoacoustic Experiments. In: MOSART WORKSHOP ON CURRENT RESEARCH DIRECTIONS IN COMPUTER MUSIC, 2001, Barcelona. *Proceedings*. Disponível em: <www.diku.dk/musinf/papers/mosart01-phase.pdf>. Acesso em: 13 out. 2004.
- ALTHOFF, Rasmus; KEILER, Florian; ZÖLZER, Udo. Extracting Sinusoids from Harmonic Signals. In: CONFERENCE ON DIGITAL AUDIO EFFECTS (DAFX-99), 1999, Trondheim, Noruega. *Proceedings*. Disponível em <<http://www.tele.ntnu.no/akustikk/meetings/DAFx99/althoff.pdf>>. Acesso em 13 abr. 2004.
- ARFIB, Daniel; KRONLAND-MARTINET, Richard. Transformer le Son: Modeler, Modéliser. In: Les Cahiers de l'Ircam: Recherche et Musique, *La Synthèse Sonore*. Paris: Editions IRCAM – Centre Georges-Pompidou, 1993. Vol. 2, p. 67-82.
- ASHIHARA, Kaoru; KIRYU, Shogo. Audibility of Components Above 22 KHz in a Harmonic Complex Tone. *Acta Acustica United with Acustica*. Gent, Belgium, vol. 89, p. 540-546, 2003.
- ASSAYAG, Gérard. CAO: Vers la Partition Potentielle. In: Les Cahiers de l'Ircam: Recherche et Musique, *La Composition Assistée par Ordinateur*. Paris: Editions IRCAM – Centre Georges-Pompidou, 1993. Vol. 3, p. 13-41.
- ASSAYAG, Gérard. *OpenMusic User Manual*. Paris: IRCAM, 1998. 110 p.

BARRIÈRE, Françoise. Reflections on the State of Electroacoustic Music Today: Aesthetic Evolution and Relation with the Public. In: VVAA, *Aesthetics and Electroacoustic Music*. Paris: Actéon – Editions Mnémosyne, 1996, p. 14-19.

BARRIÈRE, Jean-Baptiste. *Le Timbre: Métaphore pour la Composition*. Paris, Christian Bourgois Éditeur, IRCAM, 1991, 563 p.

BATISTA, Pedro A. G. *An Introduction to Sound Synthesis with Wavelet Packets*. Massachusetts: MIT Press, 2000. CD-ROM HTML de The Csound Book.

BATTIER, Marc. De la Phonographie à la Lutherie Numérique. L'Épiphanie du Son Artificiel. In: JOURNÉE D'INFORMATIQUE MUSICALE – 9^e EDITION, 2002, Marseille. *Proceedings*. Disponível em:
<http://perso.wanadoo.fr/gmem/evenements/jim2002/articles/L01_Battier.pdf>. Acesso em: 10 fev. 2004.

BERTINI, Graziano; MAGRINI, Massimo; TARABELLA, Leonello. Spectral Data Management Tools for Additive Synthesis. In: CONFERENCE ON DIGITAL AUDIO EFFECTS (DAFX-00), 2000, Verona. *Proceedings*. Disponível em:
<<http://profs.sci.univr.it/~dafx/DAFx-final-papers.html>>. Acesso em: 21 ago. de 2004.

BERTOLOTI, Fabio P. *Tools for Heterodyne Data Manipulation*. Massachusetts: MIT Press, 2000. CD-ROM HTML de The Csound Book.

BETA Distribution. Engineering Statistics Handbook. Disponível em
<<http://www.itl.nist.gov/div898/handbook/eda/section3/eda366h.htm>>. Acesso em: 9 dez. 2004.

BORIN, Gianpaolo; DE POLI, Giovanni; SARTI, Augusto. Musical Signal Synthesis. In: ROADS, Curtis et al. (Ed.). *Musical Signal Processing*. Lisse: Swets & Zeitlinger Publishers, 1997. P. 5-30.

BOULANGER, Richard (Ed.). *The CSound Book – Perspectives in Software Synthesis, Sound Design, Signal Processing, and Programming*. Cambridge: The MIT Press, 2000, 740 p.

BREGMAN, Albert S. *Auditory Scene Analysis – The Perceptual Organization of Sound*. Cambridge, Massachusetts: MIT Press, 1990, 773 p.

BUTTERWORTH Filter. Wikipedia, Encyclopedia. Disponível em <http://en.wikipedia.org/wiki/Butterworth_filter>. Acesso em: 21 dez. 2004.

CAESAR, Rodolfo. *Material e Forma na Música Eletroacústica*. Revista Pesquisa e Música do Conservatório Brasileiro de Música, 1997, vol. 3, n. 1.

CAESAR, Rodolfo. *O Zig-Zag Conceitual no Estúdio de Composição*. Anais do II Simpósio Brasileiro de Computação e Música, Canela, 1996, Universidade Federal do Rio Grande do Sul.

CAESAR, Rodolfo. *Re: Ajuda de Mestre!*. Mensagem recebida por <ignacioldc@uol.com.br> em 1 nov. 2004.

CHAFE, Chris. A Short History of Digital Sound Synthesis by Composers in the U.S.A. In: 2004 INTERNATIONAL CONFERENCE CLASSICAL MEETS AVANT-GARDE: INTERPRETATION. *Proceedings*.

Disponível em: <<http://www-ccrma.stanford.edu/~cc/lyon/historyFinal.pdf>>. Acesso em 30 out. 2004.

CHEVEIGNÉ, Alain de. Pitch Perception Models. In: PLACK, Chris; OXENHAM, Andrew (Ed.). *Pitch*. New York: Springer Verlag, 2004.

CHOWNING, John M. The Synthesis of Complex Audio Spectra by Means of Frequency Modulation. *Journal of the Audio Engineering Society*. Vol. 21, n. 7, September 1973.

CHOWNING, John M. Digital Sound Synthesis, Acoustics, and Perception: A Rich Intersection. In: CONFERENCE ON DIGITAL AUDIO EFFECTS (DAFX-00), 2000, Verona. *Proceedings*. Disponível em: <<http://profs.sci.univr.it/~dafx/DAFx-final-papers.html>>. Acesso em 21 ago. 2004. 6 p.

CLOZIER, Christian. The Aesthetic Situation and Outlook of Electroacoustic Music – Related Question: a Definition of Electroacoustic Music. In: VVAA, *Aesthetics and Electroacoustic Music*. Paris: Actéon – Editions Mnémosyne, 1996, p. 31-42.

COHEN-LEVINAS, Danielle. Entretien avec Marco Stroppa. In: Les Cahiers de l'Ircam: Recherche et Musique, *La Composition Assistée par Ordinateur*. Paris: Editions IRCAM – Centre Georges-Pompidous, 1993. Vol. 3, p. 99-117.

COHEN-LEVINAS, Danielle. Entretien avec Philippe Manoury. In: Les Cahiers de l'Ircam: Recherche et Musique, *Composition et Environnements Informatiques*. Paris: Editions IRCAM – Centre Georges-Pompidous, 1992. Vol. 1, p. 35-47.

COOK, Perry R. *Real Sound Synthesis for Interactive Applications*. Natick, Massachusetts: A K Peters, 2002, 263 p.

DANIEL, P.; WEBER, R. Psychoacoustical Roughness: Implementation of an Optimized Model. *Acustica – Acta Acustica*. Vol. 83, p. 113-123, 1997.

DEPALLE, Philippe; RODET, Xavier. De la Voix aux Instruments. In: Les Cahiers de l'Ircam: Recherche et Musique, *La Synthèse Sonore*. Paris: Editions IRCAM – Centre Georges-Pompidous, 1993. Vol. 2, p. 121-141.

DIGITAL Signal Processing: *site*. Disponível em: <<http://www.dsptutor.freeuk.com/aliasing/aliasingDemo.html>>. Acesso em: 14 jun. 2004.

DODGE, Charles; JERSE, Thomas A. *Computer Music – Synthesis, Composition, and Performance*. New York: Schirmer Books, 1985, 383 p.

ECKEL, Gerhard. La Maîtrise de la Synthèse Sonore: Outils pour l'Élaboration et l'emploi du Matériau Sonore Synthétique dans la Composition Musicale Contemporaine. In: Les Cahiers de l'Ircam: Recherche et Musique, *La Synthèse Sonore*. Paris: Editions IRCAM – Centre Georges-Pompidou, 1993. Vol. 2, p. 97-106.

EMMERSON, Simon. Acoustic/Electroacoustic: The Relationship with Instruments. *Journal of New Music Research*. Vol. 27, n. 1-2, p. 146-164, 1998.

ESCUADERO, Juan García. Continuous and Discrete Fourier Spectra of Aperiodic Sequences for Sound Modeling. In: CONFERENCE ON DIGITAL AUDIO EFFECTS (DAFX-00), 2000, Verona. *Proceedings*. Disponível em:
<<http://profs.sci.univr.it/~dafx/DAFx-final-papers.html>>. Acesso em: 21 ago. 2004.

FISCHMAN, Rajmil. The Phase Vocoder: Theory and Practice. *Organised Sound*, Cambridge, UK, vol. 2, n. 2, p. 127-145, ago. 1997.

FITZ, Kelly; HAKEN, Lippold; CHRISTENSEN, Paul. *Transient Preservation Under Transformation in an Additive Sound Model*. Disponível em:
<<http://www.cerlsoundgroup.org>>. Acesso em: 5 maio 2004.

GARCIA, Denise. *Modelos Perceptivos na Música Eletroacústica*, 1998. Tese, 249 f. (doutorado em comunicação e semiótica) – Programa de Estudos Pós-Graduados em Comunicação e semiótica, Pontifícia Universidade Católica, São Paulo.

GAUSS Curve. Wikipedia, Encyclopedia. Disponível em
<<http://www.2dcurves.com/exponential/exponentialg.html>>. Acesso em: 10 out. 2004.

GERZSO, Andrew. De Nouveaux Environnements Musicaux. In: Les Cahiers de l'Ircam: Recherche et Musique, *Composition et Environnements Informatiques*. Paris: Editions IRCAM – Centre Georges-Pompidou, 1992. Vol. 1, p. 49-55.

GRIBONVAL, R. et al. *Analysis of Sound Signals with High Resolution Matching Pursuit*.

HANAPPEUNDER, Peter. *AudioSculpt User's Manual*. Paris: IRCAM, 1996. 102 p.

HANNA, Pierre; DESAINTE-CATHERINE SCRIME, Myriam. Influence de la Densité Spectrale sur la Synthèse de Sons Bruités. In: JOURNÉE D'INFORMATIQUE MUSICALE – 9^a EDITION, 2002, Marseille. *Proceedings*. Disponível em: <http://www.labri.fr/Documents/Publications/Publi_Labri_2064.pdf>. Acesso em: 10 fev. 2004.

HARVEY, Jonathan. The Mirror of Ambiguity. In: EMMERSON, Simon (Ed.). *The Language of Electroacoustic Music*. London: The Macmillan Press Ltd., 1986. P. 175-190.

HEISENBERG, Werner. *Física e Filosofia*. Brasília: Editora Universidade de Brasília, 3^a. ed., 1995, 160 p.

HOWARD, David M.; ANGUS, James. *Acoustics and Psychoacoustics*. Oxford: Focal Press, 1998, 365 p.

HOWARD, David M.; TYRRELL, Andy M. Psychoacoustically Informed Spectrography and Timbre. *Organised Sound*. Cambridge, UK, vol. 2, n. 2, p. 65-76, ago. 1997.

IRCAM - Institut de Recherche et Coordination Acoustique/Musique. *AudioSculpt Newsletter Issue 5*. Paris, 1997. 38 p.

JEHAN, Tristan. *Perceptual Synthesis Engine: An Audio-Driven Timbre Generator*, 2001. Dissertação, 76 f. (Master of Science in Media Arts and Sciences) – MIT, Cambridge.

JENSEN, Kristoffer. *Perceptual and Physical Aspects of Musical Sounds*. Journal of the ITC-SRA, vol. 16, 2002.

JENSEN, Kristoffer; MARENTAKIS, Georgios. *Hybrid Perception*. 1st Seminar on Auditory Models. Lyngby, Denmark, 2001.

JOHNSON, Micah Kimo. *The Spectral Modeling Toolbox: A Sound Analysis/Synthesis System*, 2002. Dissertação, 98 f. (Master of Arts in Electro-Acoustic Music) - Dartmouth College, Hanover, New Hampshire.

KENDALL, Roger A. *Cognitive Psychology of Music*. Disponível em: <<http://www.ethnomusic.ucla.edu/courses/ESM172a/Files172A/WEEK4.HTM>>. Acesso em: 2 jan. 2005.

KRONLAND-MARTINET, R.; GUILLEMAIN, Ph., YSTAD, S. Modelling of Natural Sounds by Time-Frequency and Wavelet Representations. *Organized Sound*. Cambridge, UK, vol. 2, n. 3, p. 179-191, dez. 1997.

KÜPPER, Léo. Analysis of the Spatial Parameter – Psychoacoustic Measurements in Sound Cupolas. In: VVAA, *Composition/Diffusion in Electroacoustic Music*. Paris: Actéon - Editions Mnémosyne, 1998, p. 289-314.

LAGRANGE, Mathieu; MARCHAND, Sylvain. Real-Time Additive Synthesis of Sound by Taking Advantage of Psychoacoustics. In: CONFERENCE ON DIGITAL AUDIO EFFECTS (DAFX-01), 2001, Limerick, Ireland. *Proceedings*. Disponível em: <<http://www.csis.ul.ie/dafx01/proceedings/navig/toc.htm>>. Acesso em: 12 ago. 2004. 5 p.

LEIPP, E. *La Machine à Écouter: Essay de Psycho-Acoustique*. Paris: Masson e C^{ie}, 1977, 261 p.

LES CAHIERS DE L'IRCAM: COMPOSITEUR D'AUJOURD'HUI, Philippe Manoury. Paris: Editions IRCAM, vol. 8, 1995.

MARENTAKIS, Georgios; JENSEN, Kristoffer. *Sinusoidal Synthesis Optimization*. In: INTERNATIONAL COMPUTER MUSIC CONFERENCE - ICMC, 2002, Göteborg, Sweden. *Proceedings*. Disponível em: <www.dcs.gla.ac.uk/~georgios/icmc02-sinopt.pdf>. Acesso em 21 set. 2004.

MALT, Mikhail. *Les Mathématiques et la Composition Assistée par Ordinateur – Concepts, Outils et Modèles*, 2000. Tese, 874 f. (doutorado) - Ecole des Hautes Etudes en Sciences Sociales. Recebido do próprio autor em 9 ago. 2003.

MALT, Mikhail. *Composition Assistée par Ordinateur: OpenMusic*. Paris: IRCAM/Centre Georges Pompidou, 1999. 21 p. Apostila.

MANOURY, Philippe. The Arrow of Time. *Contemporary Music Review*. London, vol. 1, p. 131-145, 1984.

MANOURY, Philippe. The Role of the Conscious. *Contemporary Music Review*. London, vol. 1, p. 147-156, 1984.

MANOURY, Philippe. *Cours de Composition*. Paris: IRCAM/Centre Georges Pompidou, 1999. 15 p. Apostila.

MASRI, Paul; BATEMAN, Andrew; CANAGARAJAH, Nishan. A Review of Time-Frequency Representation, with Application to Sound/Music Analysis-Resynthesis. *Organised Sound*, Cambridge, UK, v. 2, n. 3, p. 193-205, dez. 1997a.

MASRI, Paul; BATEMAN, Andrew; CANAGARAJAH, Nishan. The Importance of the Time-Frequency Representation for Sound/Music Analysis-Resynthesis. *Organised Sound*, Cambridge, UK, v. 2, n. 3, p. 207-214, dez. 1997b.

MCADAMS, Stephen. Audition: Cognitive Psychology of Music. In: LLINÁS, Rodolfo; CHURCHLAND, Patricia (Ed.). *The Mind-Brain Continuum – Sensory Processes*. Cambridge, Massachusetts, MIT Press, 1996. P. 250-279.

MCADAMS, Stephen; WINSBERG, Suzanne. Psychophysical Quantification of Individual Differences in Timbre Perception. In: *Contributions to Psychological Acoustics*. SCHICK, August et al. (Ed.). Oldenburg: 2000. P. 165-181.

MENEZES, Flo. Fusão e Contraste entre a Escritura Instrumental e as Estruturas Eletroacústicas. In: MENEZES, Flo (Ed.). *Atualidade Estética da Música Eletroacústica*. São Paulo: Editora da Universidade Estadual Paulista, 1999. P. 13-20.

MIRANDA, Eduardo Reck. *Computer Sound Design – Synthesis Techniques and Programming*. Oxford: Focal Press, Second Edition, 2002, 264 p.

MOMENTUM. Wikipedia, Encyclopedia. Disponível em <<http://en.wikipedia.org/wiki/Momentum>>. Acesso em: 11 jun. 2004.

MOREIRA, Ildeu de Castro. Fractais. In: NUSSENZVEIG, H. Moysés (Org.), *Complexidade e Caos*. Rio de Janeiro: Editora UFRJ, COPEA, 1999. Cap. 3, p. 51-82.

MORPHOLOGY. EARS: Electroacoustic Research Site. Disponível em <<http://www.mti.dmu.ac.uk/EARS/Data/node412.html#4922>>. Acesso em: 27 nov. 2004.

MURAIL, Tristan. Au Fil des Oeuvres. In: SZENDY, Peter (Org.). *Tristan Murail*. Paris: L'Harmattan, IRCAM-Centre Pompidou, 2002. P. 103-152.

NELSON, Peter; MONTAGUE, Stephen. Live-Electronics. In: Osborne, Nigel, *Contemporary Music Review*. Paris: Harwood Academic Publishers, 1991. v. 6, 237 p.

PAINTER, Ted; SPANIAS, Andreas. *Perceptual Coding of Digital Audio*. Proceedings of the IEEE, vol. 88, n. 4, April 2000.

PEETERS, Geoffroy. Analyse et Synthèse des Sons Musicaux par la Méthode PSOLA. In: JOURNÉES D'INFORMATIQUE MUSICALE, JIM98, 1998, Agelonde. Ata. Agelonde.

PEETERS, Geoffroy; RODET, Xavier. *SINOLA: A New Analysis/Synthesis Method using Spectrum Peak Shape Distortion, Phase and Reassigned Spectrum*. Inédito. Recebido do próprio autor em 9 maio 2004.

POTARD, Yves; BAISNÉE, P. F.; BARRIÈRE, Jean-Baptiste. *Méthodologie de Synthèse du Timbre: l'Exemple des Modèles de Résonance*. In: VVAA, *Le Timbre - Métaphore pour la Composition*. Paris: I.R.C.A.M., 1991, p. 135-163.

PRESSNITZER, Daniel; MCADAMS, Stephens. *Acoustics, Psychoacoustics and Spectral Music*. Disponível em:
<<http://www.ircam.fr/pcm/pressnitzer/publications.html>>. Acesso em: 30 abril 2004.

PRESSNITZER, Daniel; MCADAMS, Stephens. *Influence of Phase Effects on Roughness Modeling*. ICMC: International Computer Music Conference, Thessaloniki, Greece, September, 1997.

PRIGOGINE, Ilya. *O Fim das Certezas: Tempo, Caos e as Leis da Natureza*. São Paulo: Editora da Universidade Estadual Paulista, 1996, 202 p.

PROBABILITY Distribution. Wikipedia, Encyclopedia. Disponível em
<http://en.wikipedia.org/wiki/Probability_distribution>. Acesso em: 9 dez. 2004.

PURWINS, Hendrik; BLANKERTZ, Benjamin; OBERMAYER, Klaus. Computing Auditory Perception. *Organised Sound*. Cambridge, UK, vol. 5, n. 3, p. 159-171, 2000.

RIMELL, Andrew; HAWKSFORD, Malcolm O. *Reduction of Loudspeaker Polar Response Aberrations Using Psychoacoustically Motivated Adaptive Sub-Band Algorithms*. 97th Convention of the AES, San Francisco, November, 1994, p. 10-14.

RISSET, Jean-Claude. Synthèse et Matériau Musical. In: Les Cahiers de l'Ircam: Recherche et Musique, *La Synthèse Sonore*. Paris: Editions IRCAM – Centre Georges-Pompidou, 1993. Vol. 2, p. 43-65.

ROADS, Curtis. *Microsound*. Cambridge, Massachusetts: The MIT Press, 2001, 409 p.

ROADS, Curtis. *The Computer Music Tutorial*. Cambridge, Massachusetts: The MIT Press, 1996, 1234 p.

ROADS, Curtis. Sound Transformation by Convolution. ROADS, Curtis et al. *Musical Signal Processing*. Lisse: Swets & Zeitlinger Publishers, 1997, P. 411-438.

ROBINSON, David J. M.; HAWKSFORD, Malcolm O. J. Psychoacoustic Models and Non-Linear Human Hearing. AES 109th CONVENTION, Los Angeles, 2000. Disponível em: <www.mp3-tech.org/programmer/docs/non-linear_human_hearing.pdf>. Acesso em: 8 maio 2004.

RODET, Xavier. Sound Analysis, Processing and Synthesis Tools for Music Research and Production. COLLOQUIUM ON MUSICAL INFORMATICS – XIII CIM, 2000, Aquila. *Proceedings*. Disponível em: <www.ircam.fr/equipes/analyse-synthese/listePublications/articlesRodet/CIM00/RodetCIM2000.ps>. Acesso em: 21 maio 2004.

RODET, Xavier; DEPALLE, Philippe; GARCIA, Guillermo. New Possibilities in Sound Analysis and Synthesis. In: ILLINOIS MATHEMATICS AND SCIENCE ACADEMY - IMSA, 1995. *Proceedings*.

Disponível em: <<http://mediatheque.ircam.fr/articles/textes/Rodet95b/>>. Acesso em: 1 abril 2004.

ROEBEL, Axel. *AudioSculpt 2.04/SuperVP 2.10: Transient Detection and Preservation*. Paris: IRCAM, 2003.

ROEDERER, Juan. *Introdução à Física e Psicofísica da Música*. São Paulo: Edusp, 1998, 311 p.

ROLAND-MIESZKOWSKI, Marek. Consequences of Nyquist Theorem for Acoustic Signals Stored in Digital Format. In: CAA CONFERENCE, 1991, Edmonton, Canada. *Proceedings*. Disponível em: <<http://www.digital-recordings.com/publ/pubneq.html>>. Acesso em: 20 out. 2003.

ROSSIGNOL, S. et al. *Features Extraction and Temporal Segmentation of Acoustic Signals*. Inédito. Recebido do próprio autor em 10 maio 2004.

RUVIARO, Bruno. *The Spell of Speech*, 2004. Dissertação, 129 f. (Master of Arts in Electro-Acoustic Music) - Dartmouth College, Hanover.

SCHAEFFER, Pierre. *Traité des objets Musicaux*. Editions du Seuil, Paris, 1966.

SCHILINGI, Jacopo Baboni; MALT, Mikhail; ASSAYAG, Gérard. *OpenMusic Tutorial*. Paris: IRCAM, 1998. 131 p.

SCHWARTZ, Diemo; RODET, Xavier. *Spectral Envelope Estimation and Representation for Sound Analysis-Synthesis*. Inédito. Recebido do próprio autor em 14 maio 2004.

SELF, Graham. *Wavelets for Sound Analysis and Re-Synthesis*, 2001. Monografia de graduação, 63 f. (Science Dual Honours in Computer Science and Mathematics) – Department of Computer Science, The University of Sheffield. Disponível em: <<http://www.dcs.shef.ac.uk/teaching/eproj/ug2001/pdf/u8gds.pdf>>.

Acesso em: 12 jun. 2004.

SERRA, Marie-Hélène. Introducing the Phase Vocoder. In: ROADS, Curtis et al. (Ed.). *Musical Signal Processing*. Lisse: Swets & Zeitlinger Publishers, 1997. P. 31-90.

SERRA, Xavier. Musical Sound Modeling with Sinusoids plus Noise. In: ROADS, Curtis et al. (Ed.). *Musical Signal Processing*. Lisse: Swets & Zeitlinger Publishers, 1997. P. 91-122.

SERRA, Xavier. Spectral Modeling Synthesis: Past and Present. In: INTERNATIONAL COMPUTER MUSIC CONFERENCE, 1989, Ohio, EUA. Disponível em: <<http://www.iaa.upf.es/~xserra/cursos/CCRMA-workshop/xjs/lectures/11-SMS-related-research.pdf>>. Acesso em 20 maio 2004.

SMALLEY, Denis. Spectro-Morphology and Structuring Processes. In: EMMERSON, Simon (Ed.). *The Language of Electroacoustic Music*. London: The Macmillan Press Ltd., 1986. P. 61-93.

SMALLEY, Denis. Spectromorphology: Explaining Sound-Shapes. *Organised Sound*. Cambridge, UK, vol. 2, n. 2, p. 107-126, ago. 1997.

SMITH III, Julius Orion. Observations sur l'Histoire de la Synthèse Numérique du Son. In: Les Cahiers de l'Ircam: Recherche et Musique, *La Synthèse Sonore*. Paris: Editions IRCAM – Centre Georges-Pompidou, 1993. Vol. 2, p. 83-95.

STEWART, Ian. *Será que Deus Joga Dados?: A Nova Matemática do Caos*. Rio de Janeiro: Jorge Zahar Editor, 1991, 336 p.

STROPPA, Marco. Paradigms for the High-Level Musical Control of Digital Signal Processing. In: CONFERENCE ON DIGITAL AUDIO EFFECTS, 2000 (DAFX-00), Verona. *Proceedings*. Disponível em:

<<http://marco.stroppa.free.fr/download/articles/Paradigms.pdf>>. Acesso em: 1 ago. 2003.

STROPPA, Marco; LEMOUTON, Serge; AGON, Carlos. omChroma : vers une Formalisation Compositionnelle des Processus de Synthèse Sonore. In: JOURNÉE D'INFORMATIQUE MUSICALE – 9^a EDITION, 2002, Marseille. *Proceedings*. Recebido do próprio autor em 15 abril 2004.

SUZUKI, Yôiti et al. Precise and Full-range Determination of Two-dimensional Equal Loudness Contour. Disponível em: <www.nedo.go.jp/itd/grant-e/report/00pdf/is-01e.pdf>. Acesso em 10 out. 2004.

SZENDY, Peter (org.). *Brian Ferneyhough*. Paris: L'Harmattan, Ircam-Centre Pompidou, 1999, 144 p.

SZENDY, Peter (org.). *Tristan Murail*. Paris: L'Harmattan, Ircam-Centre Pompidou, 2002, 161 p.

TAKESHIMA, Hisashi; SUZUKI, Yôiti. Equal-Loudness-Level Contours for Pure Tones. *The Journal of the Acoustical Society of America*. Vol. 116, Issue 2, p. 918-933, August 2004.

TENÓRIO, Robinson M. *Cérebros e Computadores: A Complementaridade Analógico-Digital na Informática e na Educação*. São Paulo: Escrituras Editora, 1998, Coleção Ensaio Transversais, 214 p.

TERHARDT, Ernst. *Auditory Acquisition of Information*. Disponível em:

<<http://www.mmk.ei.tum.de/persons/ter/top/audiinfo.html>>. Acesso em: 20 fev. 2004a.

- TERHARDT, Ernst. *Auditory Roughness*. Disponível em:
<<http://www.mmk.ei.tum.de/persons/ter/top/roughness.html>>. Acesso em: 20 fev. 2004b.
- TERHARDT, Ernst. *Definition of Pitch*. Disponível em:
<<http://www.mmk.ei.tum.de/persons/ter/top/defpitch.html>>. Acesso em: 20 fev. 2004c.
- TERHARDT, Ernst. *Spectral Pitch*. Disponível em:
<<http://www.mmk.ei.tum.de/persons/ter/top/specpitch.html>>. Acesso em: 20 fev. 2004d.
- TERHARDT, Ernst. *Virtual Pitch*. Disponível em:
<<http://www.mmk.ei.tum.de/persons/ter/top/virtualp.html>>. Acesso em: 20 fev. 2004e.
- THORNBURG, Harvey. *Detection and Modeling of Transients*. Stanford, 2003. Stanford University, Center for Computer Research in Music and Acoustics (CCRMA). Disponível em: <<http://www-ccrma.stanford.edu/~jos/tvar/tvar.pdf>>. Acesso em: 15 mar. 2004.
- TRUAX, Barry. Computer Music Language Design and the Composing Process. In: EMMERSON, Simon (Ed.). *The Language of Electroacoustic Music*. London: The Macmillan Press Ltd., 1986. P. 155-173.
- TURICCHIA, L.; DE POLI, G.; MIAN, G. A. Audio Analysis by a Model of the Physiological Auditory System. In: CONFERENCE ON DIGITAL AUDIO EFFECTS, 2000 (DAFX-00), Verona. *Proceedings*. Disponível em:
<<http://profs.sci.univr.it/~dafx/DAFx-final-papers.html>>. Acesso em: 21 ago. 2004.
- TUTSCHKU, Hans. *Analyse et Synthèse Sonore: AudioSculpt*. Paris: IRCAM/Centre Georges Pompidou, 1999. 22 p. Apostila.
- UNCERTAINTY Principle. Wikipedia, Encyclopedia. Disponível em:

<http://en.wikipedia.org/wiki/Heisenberg_uncertainty_principle>. Acesso em: 11 jun. 2004.

VAGGIONE, Horacio. About the Electroacoustic Approach. In: VVAA, *Aesthetics and Electroacoustic Music*. Paris: Actéon – Editions MNÉMOSYNE, 1996, p. 99-106.

VAGGIONE, Horacio. Some Ontological Remarks about Music Composition Processes. *Computer Music Journal*. Cambridge, USA, vol. 25, n. 1, p. 54-61, spring 2001.

VERCOE, Barry et al. *The Alternative Csound Reference Manual: Edição 4.23-2*. Boston: MIT Media Lab, 2003. Disponível em:
<<http://www.kevindumpscore.com/download/manual-pdf-4.23.zip>>. Acesso em: 20 fev. 2004.

VINET, Hugues. The representation Levels of Music Information. COMPUTER MUSIC MODELING AND RETRIEVAL, Springer 2004, Montpellier. *Proceedings*. Montpellier: WIIL, Uffe Kock (Ed.), 2004. P. 193-209.

WICKRAMARACHI, Pierre. *Effects of Windowing on the Spectral Content of a Signal*. Disponível em:
<<http://www.dataphysics.com/support/library/downloads/articles/DP-Windowing.pdf>>. Acesso em: 21 maio 2004.

WINCKEL, Fritz. *Music, Sound and Sensation: A Modern Exposition*. New York: Dover Publications, Inc., 1967, 189 p.

WISHART, Trevor. *Audible Design: A Plain and Easy Introduction to Practical Sound Composition*. York: Orpheus the Pantomime Ltd., 1994, 138 p.

WRIGHT, Matthew et al. Analysis/Synthesis Comparison. *Organized Sound*. Cambridge, UK, vol. 5, n. 3, p. 173-189, 2000.

ZICARELLI, David et al. *MSP Getting Started – Tutorials and Topics – Reference*. Disponível em: <<http://www.cycling74.com/products/dldoc.html>>. Acesso em 10 dez. 2004.

ZWICKER, E.; FASTL, H. *Psychoacoustics – Facts and Models*. Berlin: Springer-Verlag, second updated edition, 1999, 417 p.