

REGIS ROSSI ALVES FARIA

**APLICAÇÃO DE *WAVELETS* NA ANÁLISE DE GESTOS MUSICAIS
EM TIMBRES DE INSTRUMENTOS ACÚSTICOS TRADICIONAIS**

Dissertação apresentada à Escola
Politécnica da Universidade de São Paulo
para obtenção do título de Mestre em
Engenharia

Área de Concentração:
Engenharia de Sistemas Eletrônicos

Orientador: Prof. Dr. João Antônio Zuffo

São Paulo
1997

Àqueles que têm a consciência da evolução.

AGRADECIMENTOS

Ao Prof. Dr. João Antônio Zuffo, pelo interesse e apoio à esta linha de pesquisa.

À violinista Soraya Mancini Landim, pela interpretação artística das partes de violino.

A meu pai, pela influência científica. A minha mãe, pela influência artística. A ambos, pela presença constante e inesgotável estímulo, e a minha família, pela "cobrança" frequente.

À Luciana Albieri, pela inenarrável companhia, e apoio na revisão bibliográfica.

A Gustavo Adolfo M. Alvarez, pela sua inestimável amizade e presença constante.

A Adrian Boller, pela produtiva convivência e amizade.

Ao engenheiro Rubem Dutra R. Fagundes, pela amizade e pelas valiosas revisões.

Ao compositor e colega Ruggero Ruschioni, pelos livros e informações transmitidas.

Ao músico e violinista Murilo Muniz, pela avaliação sonora das análises musicais.

À Mylene Melly, pela revisão de texto.

Aos amigos e funcionários do LSI - Laboratório de Sistemas Integráveis da Escola Politécnica da USP, pelo suporte e valorização a este trabalho.

À CAPES e ao LSI, pelo suporte financeiro.

A todos aqueles que contribuíram para a finalização deste trabalho.

SUMÁRIO

LISTA DE SÍMBOLOS

LISTA DE FIGURAS

RESUMO

ABSTRACT

1 INTRODUÇÃO	1
1.1 Motivação.....	5
1.2 Objetivos	7
1.3 Organização da dissertação	12
2 TEORIA WAVELET E ANÁLISE EM MULTIRESOLUÇÃO	14
2.1 Histórico da teoria <i>wavelet</i>	14
2.2 <i>Wavelets</i> e a Transformada <i>Wavelet</i>	18
<i>Requisitos básicos</i>	19
<i>Wavelets contínuas</i>	20
<i>Transformada Wavelet Contínua (TWC)</i>	21
<i>Transformada Enjanelada de Fourier X Transformada Wavelet</i>	21
<i>Análise com largura de banda relativa constante</i>	23
<i>Implementação da transformada wavelet</i>	27
<i>Wavelets discretas</i>	27
<i>Transformada Wavelet Discreta (TWD)</i>	29
<i>Tipos de wavelets</i>	29
2.3 <i>Wavelets</i> em processamento de sinais.....	34
<i>Wavelets na Análise e Síntese de sons e timbres musicais</i>	36
2.4 Análise em multiresolução	38
<i>Teoria da multiresolução</i>	39
<i>O espaço de detalhes ou espaço wavelet: W_j</i>	40
2.5 Análise em multiresolução com <i>wavelets</i>	42
<i>O algoritmo de decomposição e reconstrução wavelet</i>	42
2.5.1 Implementação via banco de filtros.....	44
2.5.2 Abordagem matricial	47
3 TIMBRES DE INSTRUMENTOS ACÚSTICOS E A MÚSICA INSTRUMENTAL	49
3.1 Instrumentos acústicos tradicionais.....	52
<i>Seleção dos timbres instrumentais</i>	52
3.1.1 A flauta	52
3.1.2 O violino	55
3.2 Dos timbres à música instrumental	59

3.2.1	Composição de timbres: modelos elementares.....	59
3.2.2	A técnica de interpretação como fontes de expressividade.....	62
3.3	A análise e síntese de timbres musicais	64
4	ANÁLISE EM MULTIRESOLUÇÃO DE GESTOS MUSICAIS COM WAVELETS	69
4.1	Gestos musicais, padrões sônicos e níveis <i>wavelets</i>	69
	<i>A percepção de objetos sônicos</i>	70
	<i>A influência da interpretação sobre os gestos musicais</i>	72
4.2	Análise de gestos musicais em partes de flauta e violino	73
4.2.1	Seleção do material sonoro para análise.....	73
	<i>Seleção dos eventos musicais</i>	73
	<i>Ataques (notas individuais)</i>	74
	<i>Fraseados</i>	76
	<i>Sustentação</i>	80
4.2.2	Metodologia e implementação da análise das amostras	81
4.2.2.1	Descrição do sistema de análise <i>wavelet</i>	85
	<i>Implementação algorítmica</i>	87
	<i>WaveLab</i>	87
	<i>Descrição da interface gráfica, funções e códigos desenvolvidos</i>	88
	<i>Seleção das bases wavelets</i>	90
4.2.2.2	Interpretação gráfica e sônica das análises	94
	<i>Interpretação gráfica</i>	94
	<i>Interpretação sônica</i>	95
5	RESULTADOS, DISCUSSÕES E CONCLUSÕES	98
5.1	Resultados das análises	98
	<i>Análises selecionadas:</i>	
	<i>Flauta: Fraseado, Não ligado, Trêmulo dedilhado</i>	104
	<i>Flauta: Sustentação com Frulato</i>	111
	<i>Violino: Ataque Expressivo com Acento (Sforzato) com D16</i>	117
	<i>Violino: Ataque Expressivo com Acento (Sforzato) com D4</i>	122
5.2	Apreciação e discussão dos resultados.....	125
	<i>A projeção dos eventos musicais nos níveis wavelets</i>	125
	<i>Os ruídos</i>	138
	<i>O imageamento ou geração de bandas espelhadas</i>	144
	<i>O efeito harmonizador</i>	147
	<i>Análise tempo-escala</i>	150
	<i>Reconstruções seletivas e compressão de dados</i>	151
	<i>Diferenças relevantes entre a análise com D4 versus D16</i>	154
	<i>Considerações sobre a qualidade e a intenção das gravações</i>	155
	<i>Enumeração alternativa para os níveis</i>	157
	<i>Considerações sobre o WaveLab</i>	158
5.3	Conclusões.....	159
	<i>Tabela de distribuição de gestos/eventos musicais por níveis wavelets</i>	164

5.4 Trabalhos futuros.....	167
5.5 Epílogo	171
REFERÊNCIAS BIBLIOGRÁFICAS	174

GLOSSÁRIO

APÊNDICES

I DIFERENÇAS ENTRE TOM, PITCH E NOTA MUSICAL

II CENTROS DE PESQUISA EM COMPUTAÇÃO MUSICAL

III LINGUAGENS COMPUTACIONAIS PARA MÚSICA

IV MÉTODOS PARA ANÁLISE, SÍNTESE E TRANSFORMAÇÃO MUSICAL

V WAVESYNT - PACOTE PARA ANÁLISE WAVELET EM MULTIRESOLUÇÃO DE SINAIS MUSICAIS

LISTA DE SÍMBOLOS

$\ \Psi\ $	norma de uma função
$\phi_{j,k}(t)$	função escaladora discreta
$\psi(t)$	<i>wavelet</i> mãe, descrita no domínio do tempo
$\phi(t)$	função escaladora (ou <i>wavelet</i> pai), descrita no domínio do tempo
$\Psi(w)$	transformada de Fourier da <i>wavelet</i> mãe
ω_a	frequência angular de amostragem normalizada. $\omega_a=2\pi$
$\psi_{a,b}(t)$	família de <i>wavelets</i> contínuas
$\Delta f_j(t)$	projeção de $f(t)$ no subespaço W_j (detalhe)
$\psi_{j,k}(t)$	família de <i>wavelets</i> discretas
a	parâmetro escala (para <i>wavelets</i> contínuas)
b	parâmetro deslocamento (para <i>wavelets</i> contínuas)
$c_{j,k}$	coeficientes de escalamento
$d_{j,k}$	coeficientes <i>wavelet</i> - coeficientes da transformada <i>wavelet</i>
f	variável frequência (Hertz)
$F(a,b)$	transformada <i>wavelet</i> contínua (TWC), onde b é o deslocamento da <i>wavelet</i> no tempo ou espaço (posição). F é o sinal transformado.
$f(t)$	sinal descrito no domínio do tempo
$F(w,b)$	transformada enjanelada (TEF) de $f(t)$, onde b é o deslocamento da janela. F é o sinal transformado.
f_a	frequência de amostragem (Hertz). Nas análises adotou-se $f_a=44100$ Hz.
$f_j(t)$	projeção de $f(t)$ no subespaço V_j
$G, g(k)$	filtro passa-alta associado à <i>wavelet</i>
$H, h(k)$	filtro passa-baixa associado à função escaladora
j	índice de escala ou nível (para <i>wavelets</i> discretas)
k	índice de deslocamento (para <i>wavelets</i> discretas)
n, k	variável tempo discreta
t, t	variável tempo contínua
T_a	período de amostragem (segundos). $T_a=1/f_a$
w, ω	variável frequência angular normalizada (radianos). $\omega=2\pi f \cdot T_a$ ^[52]
W^T	matriz transposta

LISTA DE FIGURAS

Figura 1 - Exemplo de uma <i>wavelet</i> (não normalizada).....	19
Figura 2 - Suporte de um átomo de tempo-frequência no plano tempo x frequência	22
Figura 3 - Janelas de análise no plano tempo-frequência para a TEF e para as <i>wavelets</i>	24
Figura 4 - Cobertura do espectro de frequências pela TEF (acima) e pela TW (abaixo).....	24
Figura 5 - Fisiologia do aparelho auditivo humano	26
Figura 6 - Discretização do plano tempo-escala (<i>b versus a</i> em escala logarítmica).....	28
Figura 7 - <i>Wavelet</i> de Haar	30
Figura 8 - <i>Wavelet</i> de Daubechies (D4)	30
Figura 9 - <i>Wavelets</i> de Daubechies D4 e D16, e respectivas transformadas de Fourier (FFT's)	32
Figura 10 - Associação dos espaços V_j e W_j numa análise em multiresolução.....	40
Figura 11 - Árvore de conexão entre os espaços V_j e W_j	41
Figura 12 - Esquema de decomposição e reconstrução <i>wavelet</i>	43
Figura 13 - Análise em multiresolução com banco de filtros.....	44
Figura 14 - Flauta básica: dedilhado e respectivos tons produzidos	53
Figura 15 - Flauta transversal metálica, afinada em dó, de 3 partes montáveis e chaves vazadas	54
Figura 16 - Anatomia de um violino: partes constituintes principais.....	56
Figura 17 - Corte seccional do violino sob o cavalete.	57
Figura 18 - Esquema de um sintetizador analógico modular básico	60
Figura 19 - Envelope de amplitude x tempo, ou ADSR.....	61
Figura 20 - Fluxograma de um simples algoritmo descritor de instrumento em síntese digital.....	62
Figura 21 - Exemplos de Fraseados Ligados Normais.....	77
Figura 22 - Exemplo de notação em partitura de um portamento e/ou glissando	77
Figura 23 - Ilustração das notações usuais para o trêmulo.....	78
Figura 24 - Ilustração de uma melodia contendo frases não ligadas com <i>staccatos</i> e acentos.....	79
Figura 25 - Gravação e preparação das amostras contendo os eventos musicais para análise.....	84
Figura 26 - Diagrama de blocos geral do sistema para análise de eventos musicais	85
Figura 27 - Decomposição do sinal musical $f(n)$ em coeficientes <i>wavelet</i> $d_{j,k}$ na árvore de filtros	86
Figura 28 - Geração do sinal no nível <i>wavelet</i> p : a projeção de $f(n)$ no nível <i>wavelet</i> W_p	87
Figura 29 - Interface gráfica desenvolvida sobre MATLAB: janela de comandos.....	89
Figura 30 - Cobertura em frequência das <i>wavelets</i> D16: $\Psi_{11,0}$ a $\Psi_{16,0}$ (em escala logarítmica)	93
Figura 31 - Diagrama de blocos do sistema para ressíntese seletiva de níveis específicos.....	97
Figura 32 - Eventos musicais amostrados e analisados para o violino.....	98
Figura 33 - Eventos musicais amostrados e analisados para a flauta	99
Figura 34 - Exemplo de página inicial da análise AWMR de um evento musical com extensão 2^{17}	101
Figura 35 - Parte do sinal original (topo) e do sinal ao nível 16 (abaixo).....	106
Figura 36 - Parte do sinal ao nível 15: intervalo entre 0,94 s e 1,63 s (da 6 ^a à 9 ^a notas)	107
Figura 37 - Parte do sinal ao nível 13: da 6 ^a à 9 ^a notas (si-sol-si-sol).....	108
Figura 38 - FFT do nível 7 para a análise de fraseado não ligado, trêmulo dedilhado	110
Figura 39 - FFT do nível 15 (sustentação com frulato).....	113
Figura 40 - Primeira banda relevante da FFT do sinal original (esquerda) e do nível 11 (direita)	115
Figura 41 - Transição da 5 ^a para a 6 ^a notas nos 6 ^o (à esquerda) e 2 ^o níveis mais finos (à direita).....	133
Figura 42 - FFT do sinal original (em verde) e do 8 ^o nível mais fino (em vermelho) entre 0 e 300 Hz..	143
Figura 43 - Decomposição seguida de reconstrução (canal passa-alta do algoritmo <i>wavelet</i>).....	146
Figura 44 - Harmônicos da série de fá2 (F2) como notas/tons na escala musical.....	147
Figura 45 - Melodia tocada - Análise de fraseado ligado normal (flauta).....	148
Figura 46 - Melodia percebida no 5 ^o nível mais fino (análise de fraseado ligado normal, flauta).....	148
Figura 47 - Melodia percebida no 6 ^o nível mais fino (análise de fraseado ligado normal, flauta).....	148
Figura 48 - Melodia percebida no 7 ^o nível mais fino (análise de fraseado ligado normal, flauta).....	148
Figura 49 - Intervalo das terceira e quarta oitavas da escala temperada (registro em clave de sol).....	159

RESUMO

A expressividade é um elemento chave para o transporte de emoções em música, e seu modelamento, vital para a concepção de sistemas de síntese mais realistas. Gestos musicais executados durante a interpretação usualmente portam a informação responsável pela expressividade percebida, e podem ser rastreados por meio de padrões sônicos a eles associados em diversas escalas de resolução.

Um conjunto relevante de gestos musicais expressivos foi estudado através de uma análise em multiresolução utilizando-se a transformada *wavelet*. A escolha deve-se principalmente à capacidade natural desta ferramenta em realizar análises de tempo-escala/frequência, e suas semelhanças com o processamento dos estágios primários do sistema auditivo.

Vinte e sete eventos musicais foram capturados em interpretações de violino e flauta, e analisados com o objetivo de avaliar a aplicabilidade desta ferramenta na identificação e segregação de padrões sônicos associados a gestos musicais expressivos.

Os algoritmos *wavelet* foram implementados na plataforma MATLAB utilizando-se bancos de filtros organizados em esquema piramidal. Rotinas para análises gráfica e sônica e uma interface ao usuário foram também implementadas.

Verificou-se que as *wavelets* permitem a identificação de padrões sônicos associados a gestos expressivos exibindo diferentes propriedades em níveis diferentes da análise. A técnica mostrou-se útil para isolar ruídos oriundos de fontes diversas, extrair transientes associados a gestos súbitos e/ou intensos, e para segregar a estrutura harmônica de tons musicais, entre outras potencialidades não menos importantes.

Particularidades da técnica e efeitos secundários observados são discutidos, e os padrões sônicos observados nos níveis *wavelets* são correlacionados com os gestos musicais que lhes deram origem.

São propostos trabalhos futuros objetivando a investigação de certos eventos musicais e fenômenos verificados, bem como o estudo de implementações alternativas.

ABSTRACT

Expressiveness is a key element for emotion transportation in music, and its modeling necessary to conceive more realistic synthesis systems. Musical gestures executed during a performance carry the information answering for expressiveness, and may be tracked by means of sonic patterns associated to them within several resolution scales.

A relevant set of musical gestures was studied through a multiresolution analysis using the wavelet transform. The choice for this tool is mainly due to its natural ability to perform time-scale/frequency analysis, and for its similarities with early auditory processing stages.

Twenty seven musical events were captured from violin and flute performances, and analyzed in order to evaluate the applicability of this tool for identification and segregation of sonic patterns associated with expressive musical gestures.

The wavelet algorithms were implemented on the MATLAB platform, employing filter banks organized in a pyramidal scheme. Graphical and sonic analysis routines and a user interface were carried out over the same platform.

It was verified that wavelets enable the identification of sonic patterns associated to musical gestures revealing different properties on different levels of the analysis. The technique showed up useful to isolate noise from different sources, extract transients associated to sudden and/or intense gestures, and segregate the tonal harmonic structure, among other important features.

Particularities of the technique and secondary effects observed are discussed, and sonic patterns on wavelet levels are correlated with the musical gestures which produced them.

Future works are proposed addressing further investigation of certain musical events and phenomena observed, as well as the study of alternative implementations.

1 INTRODUÇÃO

“A questão central para os compositores não é se o comportamento humano pode ser duplicado, mas qual novo efeito musical pode ser alcançado com a interação computacional que não podia ser alcançado pelos meios existentes anteriores.”

David A. Jaffe

No trecho citado acima, Jaffe retoma uma questão polêmica sobre o papel dos instrumentos artificiais eletrônicos, dotados de um certo nível de inteligência, e que fazem frente à atividade ou função dos músicos humanos.

Durante uma certa fase, desde a introdução comercial de instrumentos musicais eletrônicos, questionou-se muito sobre a ameaça que o “músico artificial” representava aos instrumentistas humanos de carne e ossos, tomando-lhes o espaço e mesmo o emprego, e competindo com vantagens em patamares onde certas qualidades humanas são limitadas.

Hoje, no entanto, a visão acerca desta ameaça parece estar ultrapassada quando a sociedade passa, numa velocidade vertiginosa, a absorver os preceitos tecnológicos e a enxergar seus frutos como ferramentas capazes de estimularem a criatividade e o desenvolvimento do ser humano em frentes antes inimaginadas, porque até então tínhamos de ocupar nossas mentes e corpos em funções de mais baixo nível. Os instrumentos e sistemas musicais computadorizados de hoje se tornam ferramentas de sonoplastas, compositores, artistas e instrumentistas que não mais se sentem ameaçados pelo “poder sintético” e pela precisão e incansabilidade dos “eletrônicos”.

A produção musical no século XX foi e tem sido profundamente marcada e influenciada pelo uso dos meios eletrônicos, inicialmente analógicos e, posteriormente, digitais. A *mídia* eletrônica abriu perspectivas inéditas no campo da criação e produção sonora, e introduziu novos conceitos de gravação, armazenamento e reprodução de som.

Compositores, instrumentistas e artistas multimídia aproximam-se do novo meio interessados em explorar seu potencial e sua diversidade, em empregá-lo na criação de ambientações sonoras inéditas, na síntese de novos estilos, na incorporação de linguagens e representações musicais inovadoras, na aplicação de timbres e efeitos musicais exóticos, artificialmente moldados.

A tecnologia eletrônica motivou o desenvolvimento da engenharia de áudio e possibilitou o soerguimento de uma indústria voltada para a concepção e o desenvolvimento de *hardware* e *software* para produção musical, dando origem a sofisticados equipamentos para gravação, síntese, processamento, edição, masterização, armazenamento e reprodução de áudio.

Estes equipamentos eletrônicos hoje dividem espaço nos grandes estúdios e nos espetáculos com compositores, músicos, engenheiros de áudio, técnicos e sonoplastas, bem como com instrumentos musicais (muitos dos quais essencialmente eletrônicos). A eletrônica ganhou espaço decisivo nas diversas fases da produção musical - *da composição à reprodução do áudio* - participando intensamente em todas, conquistando definitivamente a aceitação da comunidade musical e sendo naturalmente absorvida no cotidiano.

A eletrônica pode estar presente no processo de composição musical, através do emprego de técnicas de composição algorítmica, ou como auxiliar no processo de escrita e depuração da partitura. Pode também estar presente na síntese e no processamento de sons naturais, na simulação de instrumentos acústicos tradicionais, ou no projeto de instrumentos virtuais, cujo paradigma de síntese não é acústico, mas matemático. Pode estar presente na interface, no instrumento ou controlador que o músico utiliza para gerar sua música. Está presente na aquisição de vozes e de sons acústicos naturais. Está presente na criação de ambientes reverberantes, nos simuladores de posicionamento (3D-áudio), na aplicação de efeitos especiais, na replicação de vozes, na transposição de escalas, na perversão de ritmos, nas harmonias exóticas e nas ilusões sônicas (segundo Risset, p.148 ^[44], e p.67 ^[64]).

A eletrônica está maciçamente presente nos ambientes de estúdio: mesas de gravação, mixadores, amplificadores, monitores, equalizadores, processadores de efeitos, compressores, sequenciadores, gravadores digitais e diversas outros equipamentos digitais e analógicos destinados ao polimento e acabamento das produções musicais, até que cheguem aos ouvintes, seja por meio de audições ao vivo (quase sempre acompanhadas por instrumentos eletrônicos ou instrumentistas virtuais criados por computador), seja por meio de um veículo armazenador adequado, como os antigos e analógicos discos de vinil (LP's - *Long Play*), os atuais CD's digitais (*Compact Disc*), os DCC's (*Digital Compact Cassete*), DAT's (*Digital Audio Tape*), os mais recentes

DVD's (*Digital Video Disc*), ou qualquer outro meio passível de introdução comercial futura (como, por exemplo, os *silicon music chips*, dispositivos armazenadores baseados em estado sólido, cujo acesso dispensa movimentação mecânica de peças)¹.

Diversos algoritmos para processamento e manipulação de sinais discretos surgiram em decorrência da pesquisa estimulada pelo advento da computação digital neste século. A clássica transformada de Fourier, conhecida desde a primeira metade do século XIX, tornou-se um instrumento popular e indispensável em análise espectral em decorrência do invento do algoritmo da transformada rápida de Fourier, a FFT (*Fast Fourier Transform*), publicado primeiro por Cooley e Tukey em 1965, hoje contando com inúmeras variações, adições e versões melhoradas acessíveis na literatura específica.

A evolução tecnológica é acompanhada por uma crescente demanda por sistemas computacionais sofisticados e algoritmos rápidos e eficientes para se processar vastos volumes de dados, que afloram em todos os ramos da atividade humana. “*O processamento de sinal tornou-se uma parte essencial na atividade tecnológica e científica contemporânea*”.^[47]

Sob a égide da tecnologia eletrônica o mundo musical nunca mais foi o mesmo, e um fluxo interminável de contribuições teóricas e práticas passou a estimular o desenvolvimento de modelos mais sofisticados para representar os fenômenos acústicos musicais, construindo uma estrutura de referência em constante evolução, sobre a qual novos sistemas computacionais, *hardware* e *software*, foram e são criados para atender a necessidades gerais e específicas no âmbito da análise, criação e interpretação musical, assistindo o compositor e o músico em tarefas como a simulação de ambientes musicais complexos e a obtenção de texturas tímbricas não naturais, exóticas e densas; a manipulação de sons naturais, a criação de estilos musicais não convencionais, ritmos e harmonias especiais; a criação de arranjos e a composição de obras inéditas, permitindo

¹ O DCC consiste num meio de gravação digital que utiliza fitas cassete (K7) padrão, mas que não se popularizou no mercado de áudio devido a diversos fatores. O DAT se tornou o meio de gravação digital mais utilizado em estúdios profissionais, devido à versatilidade intrínseca às fitas magnéticas, à portabilidade do sistema e à qualidade do áudio. O DVD é um padrão recente que utiliza como meio armazenador discos de leitura óptica (CD's) com maior densidade de trilhas, duas faces e duas camadas de leitura. Protótipos atuais exibem capacidade de armazenamento de 4,7 a 17 GB de dados, sons e imagens. O estabelecimento e especificação do padrão DVD comercial ainda está sob desenvolvimento por um consórcio de empresas, embora alguns equipamentos comerciais já estejam sendo comercializados. Os *silicon music* ou *silicon audio chips* são uma tecnologia ainda por emergir que se utilizarão das novas tecnologias de memória de estado sólido (ROM) de alta capacidade. O estágio atual é de intensa pesquisa e desenvolvimento. A NEC introduziu um primeiro protótipo no início da década, mas prevê a introdução comercial do padrão para o final deste milênio.

verificar imediatamente a relevância e adequação, bem como avaliar o impacto, do material sonoro recém editado.

As contribuições não se limitam ao desenvolvimento de sintetizadores e *softwares* para síntese e composição. Vão muito mais além, com o desenvolvimento de:

- **instrumentos musicais eletrônicos**, como o pioneiro (analogico) *Thérémin* inventado pelo russo Leon Thérémin ^[27], o *Sequential Drum* e o violino eletrônico de Max Mathews ^[64], os teclados de Robert Moog ^[16], etc.
- **instrumentos e interfaces baseados em MIDI**: teclados polifônicos, *drum machines*, instrumentos de sopro, guitarras, placas de som para computadores, etc.
- **sistemas interativos** e instrumentos virtuais baseados no mapeamento de movimentos em música: as **interfaces gestuais** - *luvas sensoriais, sensores gestuais ópticos, ultra-sônicos, de pressão e deformação* - e também na transdução de sinais biológicos (*Biomuse*, 1990, *BodySynth*, etc.) ^[48]
- **dispositivos para regência** ^[48], sistemas para **análise, transcrição e notação** musical ^[64], programas para **gravação de interpretação** (em instrumentos MIDI), **edição e reprodução** musicais (*Cakewalk*, *Cubase*, etc.), para **reconhecimento automático de partituras** (como o *MIDISCAN*) ^[71], e para **edição e impressão de partituras** (*SCORE*, *Finale*, *Encore*, etc.) ^[64]
- **processadores digitais de efeitos** ^[22; 23], **harmonizadores vocais** ^[22], as **estações de trabalho para áudio digital** (*Digital Audio Workstations*) - pacotes integrados (*software* e *hardware*) para produção de áudio digital, como o Music Kit do CCRMA (^[89], e p.63 ^[70]), o Pro-Tools da DigiDesign (<http://www.digidesign.com>), e os mais recentes Sound Forge (da Sonic Foundry), *Samplitude-Pro* (da Samplitude), o *Mediasound* (da Timeline), o *Musicator Audio*, a *Audio Prisma Workstation* (da Spectral, Inc., <http://www.spectralinc.com>), o *Sonic Solutions* (da Sonic Solutions, <http://www.sonic.com>), o *Alchemy* ^[23], e muitos outros editores digitais de áudio, além de uma infinidade de novos dispositivos baseados em MIDI, como o *MIDIvox* (conversor voz-MIDI), processadores de efeitos MIDI, etc.

O número de equipamentos e arquiteturas proprietárias disponíveis comercialmente é enorme, e não é objetivo compreender uma abordagem aprofundada neste tópico. O leitor interessado poderá buscar informações atualizadas em revistas e periódicos

especializados em áudio e tecnologia musical (*Computer Music Journal*; *Future Music*; *Keyboard*; *Áudio, Música e Tecnologia*, etc.) e em inúmeros *sites* na Internet.

1.1 Motivação

Um desafio que merece intensa investigação está em modelar-se corretamente o transporte de emoções por meio da música, compreender como sentimentos e padrões associados ao comportamento humano são transmitidos dos músicos aos ouvintes, induzindo uma espécie de "ressonância" nestes últimos. Até então os sistemas de síntese baseados em computador têm oferecido poucos graus de liberdade ao músico instrumentista, refletindo na produção de uma música de qualidade "mecânica", marcada por uma monotonicidade, um objetivismo e uma perfeição artificial, frequentemente pobre em expressividade, e desprovida do realismo humano tão presente nas audições acústicas.

A possibilidade de se conduzir a síntese de timbres não naturais impregnando-os com padrões de execução semelhantes aos que ocorrem naturalmente na execução ao vivo de instrumentos acústicos alargaria substancialmente os recursos sônicos, oferecendo a possibilidade de se imprimir uma dinâmica específica, intensificar a percepção de realismo, simulando uma "presença viva" no material sonoro, e conseguir efeitos de notável expressividade, como, por exemplo, numa execução de um concerto para timbres complexos, construídos pela associação de canto vocal e metais, mas executando fraseados numa interpretação marcada com a qualidade de cordas tocadas com arco, como nos violinos. Esta seria uma tarefa somente alcançável com o emprego da computação musical.

Há uma corrida estabelecida à busca de metodologias eficazes, velozes, conceitualmente modernas, culturalmente contemporâneas. Busca-se a redução do tempo empregado em todas as fases: no processo de composição, na escolha do ritmo e harmonização; na definição do corpo orquestral, nos gestos expressivos, empregando timbres e arranjos complexos e interessantes; na equalização, mixagem e na produção de efeitos sonoros especiais; no processamento sobre vozes humanas e timbres instrumentais, a fim de realçar propriedades estéticas, eliminar ruídos e componentes

indesejáveis, ou criar uma atmosfera psicológica específica, focalizando um ou mais temas desejados.

A pesquisa no campo da computação musical busca satisfazer esta corrida tecnológica, abrindo novas fronteiras de aplicações que preencham as demandas atuais e mostrem novos rumos de desenvolvimento. É nesse contexto que se inserem as teorias e ferramentas matemáticas como as transformadas de Fourier, *Wavelet* e outras que encontram vasto campo de aplicações no modelamento e representação de música.

Uma característica intrínseca à música é seu caráter multidimensional, composta de múltiplos objetos e estruturas sonoras de diferentes durações, intensidades e composições espectrais. Como num mapa geográfico, onde as cidades maiores são facilmente identificáveis numa escala maior, que revela o traçado do todo, e os menores acidentes geográficos e vilarejos só são visíveis numa escala menor, onde os detalhes estão presentes. Assim também se estrutura a música. ^[20]

A análise em multiresolução com *wavelets* é uma técnica que permite visualizar o sinal musical em níveis de resolução (escala) e conteúdo espectral diferentes. Uma implementação adequada pode mostrar-se útil para isolar vozes, timbres e estruturas relevantes na percepção de gestos expressivos, acentos, ataques, modulações e outros eventos discerníveis projetados em níveis *wavelets* distintos, levando-se à concepção de um modelo para mapeamento de eventos sônicos passíveis de representação em linguagens de alto nível (como na notação musical tradicional) em objetos rastreáveis a nível de sinal acústico, isolados em níveis de resoluções distintas. Tal sistema consistiria numa ferramenta de análise musical poderosa, e permitiria uma compreensão mais profunda dos fenômenos associados à interpretação sinfônica, além de abrir caminhos para novas e sofisticadas técnicas de síntese digital, capazes de modelar a execução humana com mais precisão, realismo, e em tempo real - objetivo perseguido por inúmeros pesquisadores em música computacional e desenvolvedores de sintetizadores.

O emprego de *wavelets* em música computacional também oferece vantagens óbvias em comparação às outras técnicas, dado a baixa complexidade algorítmica da transformada discreta, implicando numa alta velocidade de processamento, à facilidade de implementação digital, e por ser uma ferramenta matemática (uma transformada) capaz

de localizar características e propriedades musicais no domínio do tempo (das durações), e no domínio frequencial (da coloração espectral).

Some-se a isso o fascínio pela exploração básica num campo de aplicações novas: as regras pelas quais o sistema auditivo humano percebe eventos acústicos e realiza o reconhecimento de padrões seletivamente dentro de um conjunto de eventos simultâneos ainda não estão claramente definidas, e portanto merecem uma pesquisa mais aprofundada a fim de se revelar os segredos da cognição auditiva humana. A psicoacústica é o campo que se ocupa de investigações sistemáticas em torno destes tópicos, havendo hoje uma atividade febril na realização de diversos experimentos.²

1.2 Objetivos

Este trabalho pretende estudar e avaliar as potencialidades das *wavelets* em tarefas de análise musical, verificar sua eficácia como ferramenta para modelar aspectos expressivos e dinâmicos na música instrumental, rastrear gestos musicais, e avaliar seu potencial como técnica para processamento e síntese de música. Visando a este propósito, organizou-se uma metodologia que endereça satisfazer os seguintes objetivos específicos:

- Realizar a análise em multiresolução de trechos musicais executados por instrumentos acústicos tradicionais utilizando-se a transformada *wavelet* como técnica de decomposição do sinal sônico (análise) em "multiníveis *wavelets*" (níveis de resolução e escalas distintas).
- Identificar eventos musicais, padrões sônicos e constituintes elementares característicos do timbre do instrumento tocado segundo técnicas de interpretação específicas, e verificar a concentração, o isolamento e a distribuição destes objetos sônicos pelos multiníveis *wavelets* gerados na análise.
- Estudar e introduzir a transformada *wavelet* na análise de sinais musicais, avaliar sua aplicabilidade no reconhecimento e isolamento de padrões sônicos oriundos de gestos

² Parte da euforia atual se deve à admissão de que há mecanismos sensoriais capazes de perceber frequências além do clássico limite superior da audição de 20 KHz, fato que está atraindo a indústria do áudio digital no sentido de se empregar uma taxa de amostragem de 96 KHz com resolução de 24 bits por amostra, figuras bem superiores ao padrão do atual *Compact Disc* (CD) de 44,1 KHz de frequência de amostragem e largura de palavra digital de 16 bits.

musicais presentes em ataques, fraseados, efeitos e modulações em timbres instrumentais, e estabelecer as bases para a concepção de um futuro sistema de síntese sonora baseada em *wavelets*.

JUSTIFICATIVAS

A música é um elemento vital no processo de compreensão do mundo que nos cerca. Constitui-se num veículo essencial de aquisição e troca de informações complexas entre o indivíduo e o meio físico em que atua, e frequentemente influi na direção evolutiva de uma sociedade.

A análise e síntese de música sempre estiveram associadas à vanguarda tecnológica, utilizando-se das ferramentas e técnicas mais sofisticadas disponíveis em cada época. A representação da música, e seus intrínsecos componentes psicológicos e perceptuais associados, através de sinais mensuráveis e processáveis eletronicamente (analógica ou digitalmente) abriu perspectivas novas e inéditas na forma de se reorganizar a informação musical, e no processo de criação e composição.

Sinais musicais podem conter uma infinidade de informações complexas que podem ser visualizadas e interpretadas sob diferenciados ângulos, e representáveis distintamente em diferentes linguagens.

Por exemplo, durante a execução de um movimento sinfônico ficam registradas uma variedade de impressões psicológicas: sentimentos de tensão, momentos de expectativa, euforia e triunfo, relaxamento e tranquilidade. Imagens e situações são sugeridas a todo momento, objetos que pertencem ao universo perceptual do ouvinte, um universo lingüístico conceitual, crítico, de representação sutil e subjetiva, encarnado por uma simbologia emocional, e portanto dependente da cultura musical do ouvinte.

Sob um outro ponto de vista registram-se contornos melódicos, temas, fórmulas rítmicas, modulações, fraseados expressivos, progressões, padrões de harmonia e orquestração que evoluem no tempo: objetos interpretáveis à luz da musicologia e da notação musical tradicional.

Ainda, sob outra perspectiva, registra-se uma complexidade espectral variante no tempo e a ocorrência mista de padrões harmônicos e periódicos, em diversas escalas, associados a padrões arrítmicos, transientes e não periódicos, objetos visíveis à luz da análise física dos sinais acústicos, expressáveis numa linguagem intrinsecamente mais

técnica e matemática. Mesmo uma simples melodia, solada por um único instrumento, consiste num sinal complexo, cuja evolução espectral no tempo mistura elementos periódicos e não periódicos.

Vemos, portanto, que a análise musical está associada à especificação de um nível conceitual ou contexto que se deseja abordar. Uma ferramenta que auxiliasse a mapear elementos identificáveis num contexto lingüístico (ou perceptual) em elementos representáveis em outro contexto (ou nível), estabelecendo uma ponte entre estas linguagens, seria de extrema valia para o aperfeiçoamento dos sistemas de análise e síntese de música baseados em computador, contribuindo para encurtar a distância entre uma definição abstrata ou subjetiva e a geração de um som acústico real que se aproxime daquela definição.

Transformadas, como a *Wavelet*, que permitem um desmembramento do sinal acústico em sinais descritos em diferentes níveis de resolução (escalas) com conteúdo espectral distintos, segregando assim diferentes padrões e objetos sônicos em diferentes níveis, parecem adequadas para este desafio. Em especial, as *wavelets* atraem as atenções pela sua flexibilidade na escolha das bases, possibilitando a obtenção de bases adaptadas e otimizadas para tratarem famílias específicas de sinais.

Entre as principais justificativas para se aplicar técnicas de análise *wavelet* em sinais musicais estão:

- **Suporte local** da transformada matemática *wavelet*, desejável para a caracterização e análise de transientes, rajadas (*bursts*) e outras discontinuidades em sinais não periódicos. As técnicas de Fourier revelam-se adequadas para se analisar sinais periódicos e estacionários. Singularidades e transições abruptas se revelam espalhadas por todo o espectro de frequência. Este “suporte global” é a principal desvantagem da transformada de Fourier para representar uma função $f(t)$ localizada no tempo: são necessários muitos coeficientes de Fourier para representar o sinal, o que implica num alto custo computacional para se obter uma precisão razoável.
- **Implementação natural de uma análise em multiresolução** do sinal (tempo-escala) através do algoritmo de decomposição da transformada *wavelet*. Numa análise multiresolução com *wavelets* é possível visualizar o sinal em diversos níveis *wavelets*, isto é, em diversas *escalas*, o que permite uma analogia com mapas

geográficos³. Cada nível está associado a uma banda de frequências distinta, e a resolução temporal e frequencial num escalograma *wavelet*⁴ é variável ao longo dos eixos temporal e frequencial (ao contrário dos espectrogramas baseados em Fourier, que exibem resoluções temporal e frequencial constantes ao longo do tempo e da frequência).

- **Baixa complexidade algorítmica** e, conseqüentemente, **baixo custo computacional**, fatores preponderantes em diversas aplicações de processamento de sinais discretos, especialmente para implementações em tempo real, que requerem algoritmos velozes.
- **A propriedade de se isolar padrões distintos contidos no sinal** em níveis (da análise multiresolução) também distintos. As características e especificações das bandas de frequência, como largura de banda, forma, sintonia, taxa de subida e decaimento, suavidade, regularidade e outros parâmetros matemáticos estão diretamente relacionadas com as propriedades da *wavelet* ora empregada. Em decorrência disto, cada nível está associado a uma banda de frequências distintas, e exibe padrões ondulatórios associados àquela banda, uma característica bastante atraente e desejável para a caracterização de sinais musicais e à extração de seus componentes elementares formadores, assim como de ruídos, transientes e outros processos oscilatórios presentes no sinal.
- A possibilidade de **implementação através de bancos de filtros digitais** ortogonais e/ou biortogonais, com propriedades de reconstrução perfeita.
- A propriedade de **análise em frequência com largura de banda relativa ($\Delta f/f$) constante**, similar ao processo de "leitura" do estímulo acústico realizado pela membrana basilar da cóclea, órgão auditivo localizado no ouvido interno, o que determina um grau de similaridade com o processo de reconhecimento e de percepção sonora realizado pelo sistema auditivo humano.

³ Em mapas geográficos o conceito de escala é natural. Para se ter uma visão panorâmica num mapa usualmente adota-se um fator de escala maior. Se no entanto é desejado visualizar detalhes de uma região específica, uma escala menor é preferível. O conceito se aplica à análise em multiresolução com *wavelets*, onde cada nível corresponde a uma escala.

⁴ Escalogramas *wavelet* usualmente mapeiam **escala X tempo**, com a **intensidade** (amplitude) usualmente codificada em gradações de cinza (*gray scale*) ou em cores. A escala é proporcional ao inverso da frequência, ou seja, para se observar frequências altas requer-se uma escala menor, mais fina, enquanto flutuações de baixa frequência são facilmente vistas numa escala maior; o eixo vertical portanto pode mapear escala ou frequência (inverso do parâmetro escala).

Há ainda um aspecto comercial a considerar que justifique a pesquisa de aplicações de *wavelets* em sistemas eletrônicos musicais, no sentido de desenvolver e disponibilizar esta tecnologia para a comunidade musical.

Neste final de século verifica-se uma explosão na oferta de sistemas eletrônicos integrados digitais de alta fidelidade para produção sonora, frutos da absorção de sofisticadas tecnologias disponibilizadas pela pesquisa e desenvolvimento em laboratórios, universidades e corporações privadas.

O mercado musical, tanto do ponto de vista dos compositores e grupos musicais, quanto dos produtores, engenheiros de som e consumidores finais, facilmente adapta-se aos níveis de qualidade e sofisticação providos por uma alta tecnologia eletrônica, e passa, incessante e naturalmente, a exigir mais qualidade, recursos e níveis de sofisticação cada vez maiores. Este comportamento, mesmo não sendo exclusividade do mundo da música, torna-se potencializado neste domínio: a sofisticação e a busca da perfeição é um exemplo recorrente na história da música.

Tal oferta, acompanhando as tendências do mercado de computadores e periféricos, têm disponibilizado, a um custo cada vez menor, uma gama imensa de equipamentos de alta qualidade, precisão e poder de processamento, abrindo chances para novos produtores, músicos e profissionais do mundo da música menos afortunados, que não contam com o alto poder aquisitivo das grandes gravadoras e produtoras musicais.

1.3 Organização da dissertação

No capítulo 1 apresentou-se uma breve introdução ao universo da música sob a intensa influência da tecnologia eletrônica neste século, de forma a situar melhor o leitor no contexto desta área de pesquisa e para clarear a importância e o alcance do trabalho. A motivação em se trabalhar no campo da análise e síntese de música, os objetivos gerais e específicos da pesquisa, e as justificativas em se empregar aqui as técnicas de *wavelets* são apresentados neste capítulo.

No capítulo 2 apresenta-se a ferramenta de engenharia deste trabalho: as *wavelets*, sua transformada e a análise em multiresolução. Um traçado histórico das *wavelets* até os dias de hoje é compilado. Na sequência apresenta-se as *wavelets*, a transformada matemática *wavelet*, suas especificidades e aplicações em várias áreas de processamento de sinais. A teoria da análise em multiresolução é abordada a seguir, confluindo com a utilização de *wavelets* para implementá-la.

O capítulo 3 trata do alvo básico deste trabalho: os timbres de instrumentos acústicos reais e a música instrumental que deles emana. A flauta e o violino foram os instrumentos escolhidos como exemplos para este estudo. Neste capítulo apresentam-se alguns modelos frequentemente usados para gerar timbres nos sintetizadores eletrônicos, discute-se a influência da técnica interpretativa sobre a impressão de expressividade e apresenta-se o clássico método de análise e síntese de timbres musicais.

No capítulo 4 os gestos musicais são então introduzidos como alvos específicos para análise neste trabalho. São apresentados os eventos musicais representativos de um conjunto relevante de gestos musicais expressivos, que serão analisados através de uma decomposição em multiresolução com *wavelets*. Objetiva-se com isso a identificação dos padrões sônicos presentes nos trechos musicais analisados que estejam correlacionados a tais eventos. O material sonoro que foi amostrado a partir de interpretações de flauta e violino e o sistema de análise *wavelet* são apresentados, juntamente com a metodologia adotada para realizar as análises e interpretá-las. A descrição dos eventos musicais é feita sob o ponto de vista musical, associando a técnica interpretativa aos gestos musicais.

No capítulo 5 passa-se à apresentação, apreciação e discussão dos resultados, e às conclusões finais. Procede-se à discussão de alguns temas adequados à avaliação da análise *wavelet* em multiresolução aplicada a sinais musicais, endereçando a sua capacidade em identificar e isolar estruturas sônicas em níveis diferentes da análise, a fenomenologia associada ao processo da análise, propriedades e particularidades. Finalmente, tendo como pano de fundo o futuro, discute-se alguns projetos de pesquisa aplicáveis, abordagens alternativas e perspectivas.

A bibliografia consultada é apresentada na sessão seguinte.

Os apêndices vêm em seguida, acrescentando informações relacionadas à área de computação musical e abordando temas paralelos à essa área de pesquisa. O pacote de análise wavelet em multiresolução (AWMR) para sinais musicais desenvolvido para este trabalho - o *Wavesynt* - é também introduzido nos apêndices.

Anexo a este texto também se encontra um CD-ROM (PC-compatível) e disquete (1,44MB, PC-compatível) onde se encontra a íntegra das análises realizadas de flauta e violino (texto e figuras), gravadas em formato Microsoft Word-compatível e comprimidas em formato ZIP.

Ainda, no CD-ROM, incluiu-se cópias dos pacotes de análise WaveLab e **Wavesynt** para MATLAB, uma cópia da dissertação, cópias dos arquivos sonoros originais de flauta e violino (trechos musicais analisados, em formato AIFF), bem como arquivos sonoros (em formato MATLAB) e figuras das análises (em formato GIF), de forma a possibilitar ao leitor/usuário a verificação e reprodução das análises e resultados obtidos.

2 TEORIA *WAVELET* E ANÁLISE EM MULTIRESOLUÇÃO

“A transformada wavelet é uma ferramenta que fatia dados ou funções ou operadores em componentes frequenciais diferentes, e então estuda cada componente com uma resolução casada com sua escala”

I. Daubechies ^[13]

A teoria *wavelet* se baseia em representar funções genéricas em termos de “*blocos construtivos básicos*”, ou “*átomos*”, fixos em diferentes escalas e posições.

O tipo de sinal sob atenção determinará as ferramentas coadjuvantes necessárias para uma melhor identificação, reconhecimento e tratamento de suas características. Em imagens, por exemplo, as bordas, contornos, texturas, cores e padrões geométricos são elementos úteis no mapeamento e na classificação de funções visuais. Para sons, a composição harmônica, padrões de dinâmica, envelopes, gestos, progressões, texturas e outros objetos percebidos pela audição são elementos importantes para se compreender e classificar sinais musicais.

Esta abordagem é útil em diversas áreas diferentes, por exemplo em técnicas de filtragem QMF (*Quadrature Mirror Filtering*), nos esquemas piramidais e de codificação sub-banda, na teoria dos estados coerentes, entre outros ^[35]. A teoria *Wavelet* representa hoje um corpo sintético destas teorias, unificando-as sob uma mesma estrutura de referência.

2.1 Histórico da teoria *Wavelet*

A aplicação de *wavelets* em processamento de sinais é relativamente nova, datando da década de 1980. Entretanto, uma investigação retroativa na história da matemática irá revelar pelo menos sete origens diferentes para a análise *wavelet* ^[47]. A idéia de se focalizar um sinal sob várias escalas diferentes e analisá-lo com várias resoluções surgiu independentemente em várias áreas de matemática, física e engenharia.

Em 1910, enquanto trabalhava na construção de bases para representar funções integráveis quadraticamente, Haar construiu a primeira família *wavelet* ortonormal de suporte compacto de que se tem notícia, e que leva o seu nome.

Calderón (1960) abordou em seus trabalhos a *decomposição atômica*, introduzindo a *identidade de Calderón*, uma decomposição de um operador de identidade ^[47]. Seu trabalho foi redescoberto 20 anos mais tarde por Grossman e Morlet, como veremos a seguir.

No início dos anos 80, Strömberg (1982) construiu as primeiras *wavelets* ortogonais, trabalhando num contexto para melhor compreender os espaços de Hardy ^[35]. Suas *wavelets* exibem decaimento exponencial e estão em C^k (k arbitrário mas finito)⁵. Infelizmente sua construção foi pouco notada naquele momento^[13].

Sob esta perspectiva histórica, foi possível relacionar entre si as várias teorias - *da decomposição de Littlewood-Paley (1930) à identidade de Calderón (1960)* - e uma primeira abordagem unificadora surge com os trabalhos de Grossman e Morlet (1984), levando à primeira síntese que encampava o termo *wavelet*.⁶

Em 1982, o geofísico francês Jean Morlet empregou *wavelets* em seus trabalhos em geoexploração. Em seguida, Alex Grossman, Morlet e Yves Meyer estudaram a transformada *Wavelet* e perceberam que as técnicas da teoria de Calderón-Zygmund, em particular as representações de Littlewood-Paley, poderiam levar a uma concepção unificada de muitos resultados da análise harmônica, e mais: que poderiam substituir as séries de Fourier em aplicações numéricas^[35]. Daí construíram as bases matemáticas da teoria *Wavelet*, com ênfase nas representações de sinais por “blocos construtivos”, a que Grossman e Morlet chamaram de *wavelets*.

Em 1987, Tchamitchian construiu o primeiro exemplo do que se pode chamar de bases *wavelets* biortogonais.

A atenção da comunidade de Processamento de Sinais foi atraída para o novo “filão” quando, no outono de 1986, S. Mallat e Y. Meyer desenvolveram a teoria da análise em multiresolução, que proporcionou uma explanação satisfatória para todas essas construções, e disponibilizou uma ferramenta para a construção de outras bases^[13]. Seu trabalho pode ser visto como uma segunda síntese, mais completa, englobando as técnicas numéricas de processamento de sinais e imagens^[47].

⁵ Funções em C^k ou de classe k são funções contínuas com k derivadas contínuas.

⁶ O termo original provém do Francês "*ondelette*", portado para o inglês como "*wavelet*".

Nos últimos 15 anos, várias bases *wavelets* ortonormais para $L^2(\mathbf{R})$ ⁷ foram construídas, as quais verificam as melhores propriedades das bases de Haar e de Littlewood-Paley: estas novas bases possuem excelente localização tanto no tempo quanto na frequência. A descoberta de bases ortonormais da forma $2^{j/2}\psi(2^j t-k)$, $j,k \in \mathbf{Z}$, $\psi(t)$ de classe C^1 , e de suporte compacto (não nulas em um intervalo finito), é devida a Ingrid Daubechies.

Daubechies (1987), inspirada nos trabalhos de Mallat, estendeu o trabalho de Haar, sintetizando esta família de *wavelets* ortonormais, possibilitando uma análise e síntese mais eficiente do que a obtida com outros sistemas (como o de Haar). Desde então o número de contribuições teóricas e práticas no campo das *wavelets* cresceu exponencialmente ano após ano, assim como a difusão de seu uso em inúmeras áreas.

A partir de 1994 verifica-se uma tendência de vários segmentos da comunidade científica em explorar o potencial das *wavelets* principalmente em aplicações de codificação, representação e compressão de sinais. Observa-se a partir de então um interesse em torno de aplicações gráficas, especialmente compressão de imagens e processamento gráfico em multiresolução^[68; 69].

Paralelamente aos trabalhos em aplicações específicas, diversas contribuições endereçando novos algoritmos e estratégias de implementação alternativas foram introduzidas para a transformada *wavelet* contínua e discreta. Implementações computacionais propuseram versões para a transformada rápida discreta (utilizando-se bancos de filtros), algoritmos melhorados para resolver efeitos espúrios de bordas⁸ e esquemas para obter maior particionamento do espectro, de forma a aumentar a resolução nas faixas de frequências mais altas, utilizando-se técnicas como os "pacotes de *wavelets*" (*wavelet-packets*), e *multiwavelets*.

A pesquisa por novas bases *wavelets* e a formalização de técnicas para construção de bases adquirem também importância na medida em que várias aplicações apresentam melhores desempenhos. Bases *wavelets* são desenvolvidas tais que satisfaçam diversos

⁷ $L^2(\mathbf{R})$ é o espaço de todas as funções quadraticamente integráveis em \mathbf{R} (números reais), como na equação (3) adiante. Nesta classe de funções encontram-se a maior parte dos sinais reais (de energia finita) de importância em aplicações de processamento de sinais.

⁸ Várias técnicas existem para estender sinais finitos de forma a prepará-los para o processamento e atenuar os indesejados efeitos de bordas, as mais empregadas sendo a extensão por acolchoamento com zeros (*zero padding*), por extensões periódicas (replicação periódica, *wraparound* ou convolução circular), por extensões simétricas (*mirror-image replication*) e extrapolação linear^[45; 80; 78, p.263]. Neste trabalho o algoritmo de decomposição utiliza convolução circular, e adicionalmente algumas amostras foram estendidas por *zero-padding* para atingir o comprimento adequado exigido pelo algoritmo.

critérios desejáveis, como rapidez de processamento, boa localização espacial/temporal, regularidade e suavidade, boa localização em frequência, independência, e melhor descrição de classes específicas de sinais.

Entre 1989 e 1994 o *método da melhor base* é introduzido e vários algoritmos baseados nele são desenvolvidos. Wickerhauser (1994) aborda o algoritmo da melhor base (*best basis algorithm*) em livro de sua autoria ^[88]. Várias bases *wavelets* são concebidas e publicadas na literatura - bases biortogonais, bases adaptativas, as bases de Malvar, Vaidyanathan, Beylkin, entre outras.

Surgem por volta de 1995 novas técnicas de implementação e síntese de *wavelets* que não recorrem aos métodos de Fourier para sua construção, introduzindo novos paradigmas de implementação e geração de bases *wavelets*, que alguns autores consideram uma segunda onda das *wavelets*: as *wavelets* de segunda geração. Sweldens (1995) introduz assim o esquema ou algoritmo de *lifting*, uma nova filosofia para construção de *wavelets* biortogonais independente da transformada de Fourier, até então importante instrumental na construção de *wavelets*. ^[79]

Implementações em *hardware* começam a aparecer no mercado pela metade da década de 1990. A Analog Devices lança o dispositivo ADV601, um *chip* para aplicações em compressão com *wavelets*. A empresa norte-americana Aware patenteia algoritmos de melhor base *wavelet*, desenvolve sistemas de transmissão para telecomunicações baseado numa tecnologia proprietária denominada *Discrete Wavelet Multitone Technology* (DWMT), e também lança *chips* para processamento *wavelet* no mercado. Diversas propostas de implementação de *wavelets* em dispositivos ASIC/VLSI aparecem na literatura.

A atividade em torno das *wavelets* é febril. Numerosos congressos, seminários, *workshops*, publicações especiais e conferências internacionais endereçando aspectos teóricos e aplicativos das *wavelets* são realizados e propostos em vários países. Diversos *sites* na Internet são dedicados à teoria *wavelet*, à disseminação de informação, artigos, trabalhos científicos, pesquisas em aplicações específicas, à divulgação de eventos, grupos de pesquisa, e suporte a “usuários” das *wavelets*. O *site* www.wavelet.org - o *Wavelet Digest* - é um dos pioneiros, contando 9418 assinantes (em 16 de maio de 1997), na sua maioria pesquisadores e usuários de *wavelets*.

O leitor interessado numa visão histórica mais completa até 1993 encontrará no livro de Meyer (1993) ^[47] uma excelente retrospectiva histórico-matemática sobre o assunto. A Internet, por sua vez, constitui hoje o melhor e mais rápido veículo para se obter informações sobre quaisquer áreas de desenvolvimento endereçando as *wavelets* e suas aplicações.

No item 2.2, que se segue, procede-se a uma introdução à teoria *wavelet*, ao conceito de *wavelet*, às *wavelets* contínuas e discretas e às transformadas contínua e discreta.

No item 2.3 descreve-se brevemente a aplicação de *wavelets* em processamento de sinais, ilustrando com referências a outros trabalhos que serviram de inspiração ao presente.

2.2 *Wavelets* e a Transformada *Wavelet*

Wavelets são "ondas pequenas", funções obtidas a partir de uma função protótipo - a *wavelet* mãe - $\psi(t) \in L^2(\mathbf{R})$, por meio de dilatações (contrações ou escalamento) e translações (deslocamentos). O objetivo primordial em se realizar estas operações é o de se obter uma família de funções base para descrever outras funções $\in L^2(\mathbf{R})$.

O termo "*wavelet*" emergiu da literatura francesa na área de geofísica ^[78], mais especificamente em trabalhos de geoexploração desenvolvidos por Grossman e Morlet, que primeiro cunharam a palavra "*ondelette*" referindo-se às "ondas pequenas" ou blocos construtivos básicos que estavam utilizando ^[35]. O terminologia inglesa *wavelet* deriva portanto do vocábulo francês⁹.

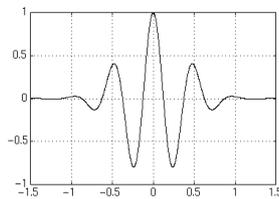
Na Figura 1 (veja abaixo) mostra-se um exemplo de uma *wavelet*-mãe (a), uma versão escalada (b) e uma versão escalada e deslocada (c). A *wavelet* da Figura 1 é uma *wavelet* de Morlet, cuja primitiva é a função $\psi(t) = e^{-\alpha t^2} \cdot e^{j\omega t}$. Introduzindo dilatações por meio de um fator de escalamento a , e translações por meio de uma variável b deduz-se a forma geral de uma família de *wavelets*

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad - (1)$$

⁹ Alguns países de língua espanhola traduzem o termo como "ondaletas".

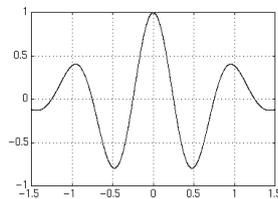
onde $\psi(t)$ é a *wavelet* primitiva ou mãe.

Figura 1 - Exemplo de uma *wavelet* (não normalizada)



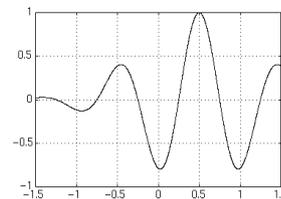
a) *wavelet* mãe

$$\psi(t) = e^{-\alpha t^2} \cdot e^{j\omega t}$$



b) *wavelet* escalada

$$\psi(t/a), \quad a = 2.0.$$



c) *wavelet* escalada e deslocada

$$\psi\left(\frac{t-b}{a}\right), \quad a=2, b=0.5$$

A utilidade das *wavelets* está na sua possibilidade de atuarem como funções base na decomposição de outras funções $\in L^2(\mathbf{R})$. As bases *wavelets* podem ser redundantes ou ortonormais, cobrindo e descrevendo completamente o espaço $L^2(\mathbf{R})$, de uma forma mais sofisticada que as bases senoidais dos métodos de Fourier. Na sua forma contínua, como veremos a seguir, as *wavelets* são basicamente redundantes. Em forma discreta podem dar origem às chamadas “molduras” (*frames*: bases discretas redundantes), a bases ortonormais, biortogonais, e outras bases *wavelets*.^[13]

REQUISITOS BÁSICOS

Além desses aspectos, outras propriedades são desejadas e exigidas das famílias *wavelets* para que sejam úteis em aplicações práticas em processamento de sinais e na análise de espaços funcionais. Neste contexto, *wavelets* devem satisfazer requisitos como^[47]:

- possuírem energia finita, de forma que a análise preserve a energia, e que a síntese leve a uma reconstrução perfeita
- possuírem certo grau de regularidade (suavidade)
- serem nulas no infinito
- possuírem um certo número de momentos nulos¹⁰
- que sejam funções de classe C^k ($0 < k < \infty$)
- que tenham suporte compacto, no tempo e na frequência

Existem inúmeras possibilidades de escolha de *wavelets* para análise, entretanto a escolha não é arbitrária, e obedece a critérios matemáticos bem definidos, como os supracitados.

Analisa-se a seguir o caso contínuo, que leva à **transformada *wavelet* contínua ou integral**.

WAVELETS CONTÍNUAS

Seja $\psi(t) \in L^2(\mathbf{R})$ uma *wavelet* mãe. Para que possa dar origem a uma família de *wavelets* exige-se que^[38]:

a *wavelet* seja absolutamente integrável

$$\int_{-\infty}^{+\infty} \psi(t) dt < \infty \quad - (2)$$

que possua energia finita

$$\int_{-\infty}^{+\infty} |\psi(t)|^2 dt < \infty \quad - (3)$$

e que satisfaça a uma condição de admissibilidade

$$C = 2\pi \int_{-\infty}^{+\infty} |\Psi(w)|^2 \frac{dw}{|w|} < \infty \quad - (4)$$

Esta última condição implica na prática que a *wavelet* oscila, integra-se a zero e não possui componente contínuo ($DC = 0$), ou $\Psi(w = 0) = 0$, logo $\int_{-\infty}^{+\infty} \psi(t) dt = 0$ ^[13].

Uma família de *wavelets* duplamente indexadas surge através do escalamento e translações da *wavelet* mãe:

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right) \quad a, b \in \mathbf{R}, a \neq 0 \quad - (5)$$

A normalização é escolhida de forma que $\|\psi_{a,b}\| = \|\psi\|$, para todo $a, b \in \mathbf{R}, a \neq 0$, e para fins práticos fazemos $\|\psi\| = 1$. ^[13]

¹⁰ $f(t)$ tem \mathbf{m} momentos nulos quando $\int_{-\infty}^{+\infty} f(t) \cdot t^k dt = 0$ para $k=0,1,\dots,\mathbf{m}$

Para a análise de voz e sons musicais muitas vezes é conveniente extrair informações sobre a distribuição de energia e comportamento de fase, o que pode ser feito utilizando-se uma *wavelet* complexa, tal que $\psi(t)$ possua somente componentes de frequência positiva, $\Psi(w = 0) = 0$, de forma que a parte imaginária seja a transformada de Hilbert da parte real. [38]

No escopo do presente trabalho, entretanto, é explorado o uso de **bases *wavelets* reais**, ortonormais, em implementação discreta utilizando-se bancos de filtros.

TRANSFORMADA WAVELET CONTÍNUA (TWC)

A transformada *wavelet* contínua (ou integral) com respeito à família de *wavelets* da equação 5 acima é

$$F(a, b) = |a|^{-1/2} \int_{-\infty}^{+\infty} f(t) \cdot \bar{\psi}\left(\frac{t-b}{a}\right) dt = \langle f(t), \bar{\psi}_{a,b}(t) \rangle \quad - (6) \text{ TWC}$$

onde $\bar{\psi}$ é o complexo conjugado de ψ . A função $f(t)$ pode ser recuperada de $F(a,b)$ através da *resolução de identidade de Calderón* [13] - uma **transformada *wavelet* (contínua) inversa (TWCI)**, como abaixo:

$$f(t) = \frac{1}{C} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(a, b) \cdot \psi_{a,b}(t) \frac{da \cdot db}{a^2} \quad - (7) \text{ TWCI}$$

A recuperação de $f(t)$ de $F(a,b)$ só é possível se a constante C acima representada (constante na *identidade de Calderón*) for finita, donde deriva a **condição de admissibilidade** expressa na equação (4) acima [78]. Na transformada enjanelada de Fourier (TEF)¹¹, C é obtido a partir da norma da função janela g , isto é, $C = \|g\|^2$. Para *wavelets*, C é definido como na equação (4).

TRANSFORMADA ENJANELADA DE FOURIER (TEF) X TRANSFORMADA WAVELET

Comparando-se a transformada *wavelet* (TW) com a transformada enjanelada de Fourier (TEF), podemos verificar semelhanças e diferenças importantes que clareiam certos aspectos vantajosos no uso das *wavelets*. Uma comparação direta de suas fórmulas

¹¹ Também encontrada sob o nome de *Short Time Fourier Transform* (STFT), a transformada de Fourier em intervalos curtos.

mostra uma correspondência e semelhança entre as janelas moduladas da TEF e as *wavelets* da TW:

Transformada Enjanelada de Fourier

Transformada *Wavelet* Contínua

$$F(w, b) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(t) \cdot g(t-b) \cdot e^{-j\omega t} dt \quad F(a, b) = |a|^{-1/2} \int_{-\infty}^{+\infty} f(t) \cdot \bar{\psi}\left(\frac{t-b}{a}\right) dt = \langle f(t), \bar{\psi}_{a,b}(t) \rangle$$

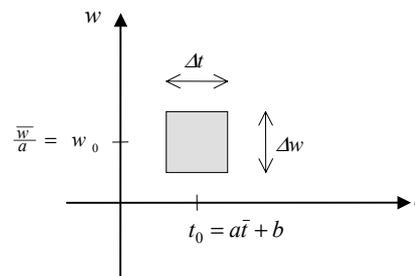
Ambas as transformações levam a implementações de análise em tempo-frequência, ou, simplesmente, análise tempo-frequência. Numa análise tempo-frequência o sinal analisado pode ser expresso por meio de uma combinação linear de átomos de tempo-frequência, caracterizados por uma duração finita no tempo (Δt) e uma duração finita na frequência (Δw).

Janelas moduladas, como $g(t-b) \cdot e^{j\omega t}$, e *wavelets* $\psi_{a,b}(t)$ são *átomos de tempo-frequência*. O exemplo mais conhecido talvez seja a "*wavelet* de Gabor":

$$G_{w_0, t_0}(t) = e^{jw_0 t} g_h(t - t_0), \quad \text{onde} \quad g_h(t) = \pi^{-1/4} e^{-\frac{t^2}{2h^2}}$$

Um átomo de tempo-frequência ocupa uma região finita no plano tempo-frequência que possui suporte no intervalo $[t_0 - \Delta t/2, t_0 + \Delta t/2]$ e $[w_0 - \Delta w/2, w_0 + \Delta w/2]$, como abaixo:

Figura 2 - Suporte de um átomo de tempo-frequência no plano tempo x frequência



Funções não podem ao mesmo tempo serem limitadas no tempo e na frequência, bem como serem simultaneamente determinadas com precisão no tempo e na frequência [13; 47], mas certas situações práticas equivalem a esse suporte compacto no tempo-frequência, e podem ser expressas por condições menos severas como:

$$\int_{-\infty}^{+\infty} (t - t_0)^2 \cdot |\Psi(t)|^2 dt = c_1^2 \Delta t^2 \leq A^2 \Delta T^2$$

$$\int_{-\infty}^{+\infty} (w - w_0)^2 \cdot |\Psi(w)|^2 dt = c_2^2 \Delta w^2 \leq 2\pi \cdot A^2 \Delta T^{-2}$$

onde A , C_1 , C_2 e ΔT são constantes finitas.

Isso quer dizer que os “átomos” extraem informações sobre o sinal analisado que estejam contidas nos intervalos de tempo e frequência delimitados acima, localizando desta forma porções específicas do sinal no espaço tempo-frequência.

A principal diferença entre os átomos da TEF e as *wavelets* é que nestas, ao invés de uma variável frequência (w), temos uma variável escala (a). As *wavelets* ao invés de moduladas são escaladas, e na verdade os átomos são de tempo-escala.

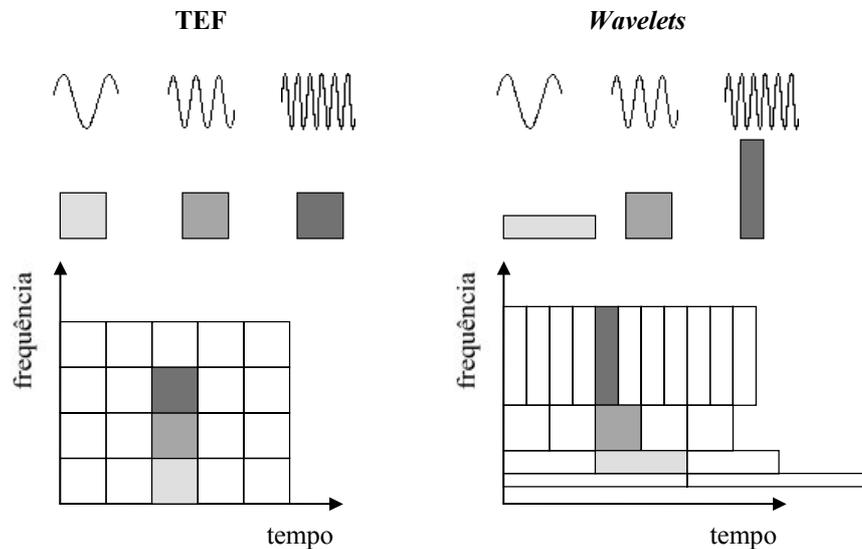
A “*wavelet* de Gabor” (na prática uma janela modulada em w para $h=1$) analisa sinais na região de alta frequência com uma janela contendo um número de ciclos muito maior do que contém a janela para análise em baixa frequência, o que leva à instabilidade numérica na representação de transientes de alta frequência [38]. Além disso, as “*wavelets* de Gabor” não são bem separadas umas das outras, falhando em serem ortogonais entre si, logo não levando à construção de uma base ortonormal para $L^2(\mathbf{R})$ [47], aspectos que limitam seu uso em análise de sinais em tempo-frequência, em especial quando o algoritmo é discreto e deseja-se um significativo poder de decorrelacionar ou reduzir a entropia na representação *wavelet* e habilitar a reconstrução perfeita.

Estas desvantagens não ocorrem com as *wavelets*, que reescalam sua base analisadora para cada (banda de) frequência, sem alterar o número de ciclos da onda. Na prática, essa diferença implica num desempenho superior da análise tempo-frequencial com *wavelets*.

ANÁLISE COM LARGURA DE BANDA RELATIVA CONSTANTE

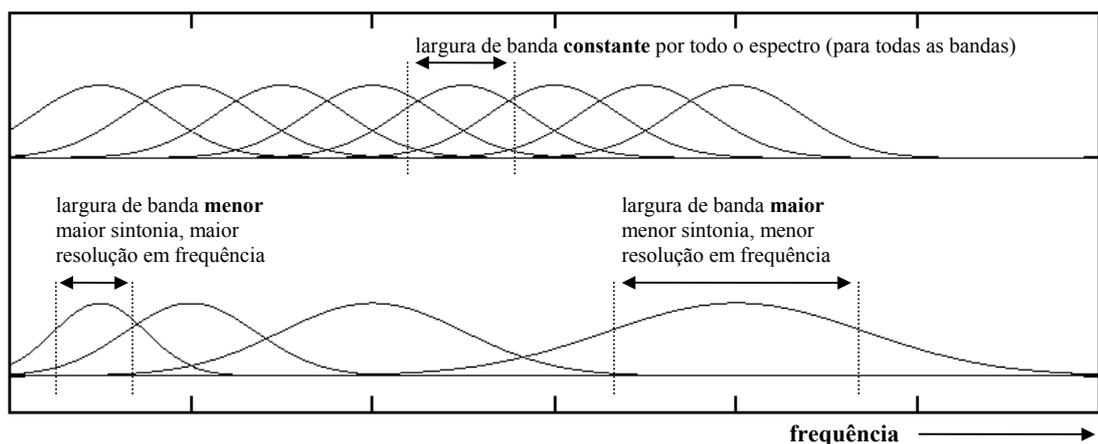
Outra diferença chave está na resolução da cobertura do plano tempo-frequência. A TEF divide o plano tempo-frequência em porções de dimensões regulares e constantes, analisando em frequências altas com as mesmas resoluções que analisa frequências baixas. A transformada *wavelet*, por outro lado, permite uma cobertura do plano tempo-frequência com resolução temporal (e frequencial) nas faixas de baixas frequências diferente da resolução em altas frequências. Alguns autores consideram esta a principal diferença entre a TEF e a transformada *wavelet*. [74] A Figura 3 abaixo ilustra as janelas de análise no plano tempo-frequência comparativamente para a TEF e para as *wavelets*. A resolução no plano tempo-frequência para as *wavelets* é mais fina no tempo para altas frequências, e mais sintonizada ou seletiva em frequência para baixas frequências. [84]

Figura 3 - Janelas de análise no plano tempo-frequência para a TEF e para as *wavelets*



Quando se analisa frequências mais altas a resolução temporal é mais fina do que quando se analisa em frequências baixas. Já a resolução em frequência fica melhor nas frequências mais baixas, onde a segmentação do espectro é mais fina e resolvida do que em altas frequências. A Figura 4, logo abaixo, é auxiliar na compreensão desta segmentação do espectro.

Figura 4 - Cobertura do espectro de frequências pela TEF (acima) e pela TW (abaixo)



Esse efeito pode ser explicado de outra maneira: na análise em frequências mais altas a transformada emprega *wavelets* mais finas, comprimidas e de curta duração, portanto permitindo a visualização do sinal analisado numa escala mais detalhada, localizada, com melhor resolução temporal. Em frequências mais baixas, *wavelets* mais largas e dilatadas enquadram o sinal numa escala maior, menos detalhada, permitindo visualizar características globais.^[60] Nesta faixa, a análise apresenta uma resolução de frequência

mais fina do que a temporal, tornando mais fácil a localização em frequência. Esse aspecto funcional das *wavelets* justifica o título de *microscópio matemático* que muitos pesquisadores lhe conferem.

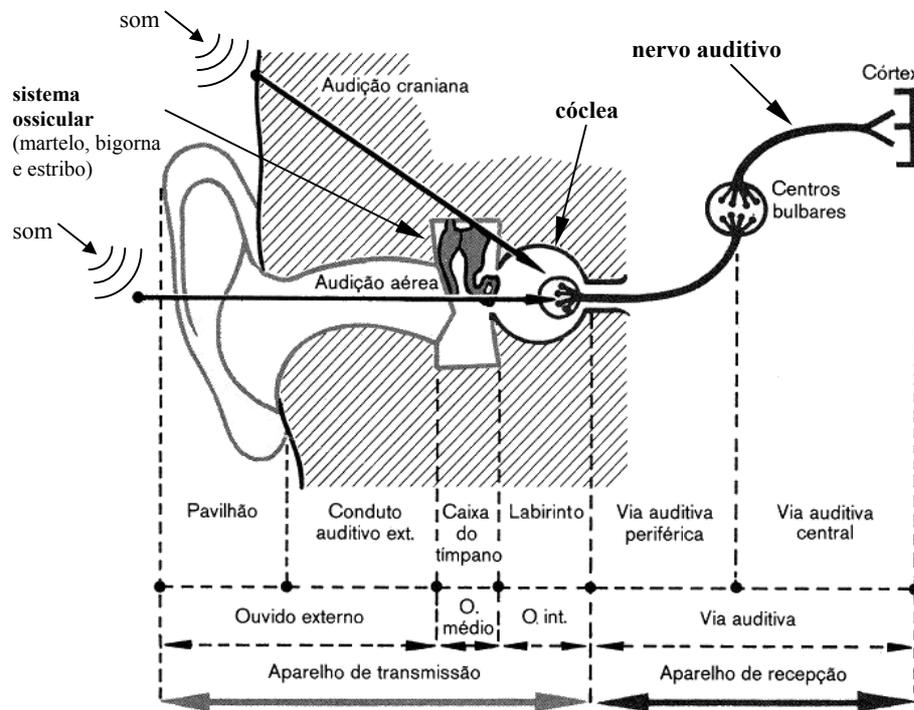
O mesmo processo de segmentação do espectro, com resoluções temporal e frequencial variáveis ao longo do plano tempo-frequência, é realizado pela membrana basilar da cóclea - *órgão perceptivo auditivo humano* - perfazendo uma análise espectral dos sinais acústicos com fator de qualidade *Q-constante* ^[60], isto é, com uma relação $\Delta f/f$ constante ao longo do eixo da frequência, como ilustra a Figura 4, acima.

Graças a esse esquema de processamento e análise do estímulo acústico (semelhante a uma análise *wavelet*) pelo ouvido é que podemos discernir melhor 2 *pitchs* (ou frequências individuais) muito próximos entre si nas faixas de baixa frequência, do que *pitchs* separados pelo mesmo intervalo nas faixas de alta frequência. Também por isso é mais fácil interpretar e acompanhar transições e eventos muito rápidos em *pitchs* agudos (alta frequência) do que entre *pitchs* graves (baixa frequência) onde a resolução em frequência é pior.

O espectro fica segmentado numa escala logarítmica, onde a resolução e seletividade de cada banda é mais fina (sintonizada) na direção das baixas frequências. Para timbres muito agudos possuímos melhor sensibilidade temporal e menor frequencial. É mais fácil distinguir eventos curtos e seqüenciais e mais difícil detectar desafinações ou deslocamentos de frequências. Nos timbres graves é mais fácil distinguir microtons e avaliar intervalos de frequência que separam *pitchs* (ou frequências individuais), isto é, a seletividade em frequência é melhor nas faixas mais graves, como sugerem as bandas de passagem dos filtros da TW na Figura 4 acima.

A Figura 5 mostra um esquema do sistema auditivo humano mostrando suas divisões anatômicas e funcionais. A modalidade de audição por condução aérea é a mais importante. Por esta via, o estímulo acústico (ondas sonoras *aéreas*) adentra pelo conduto auditivo no ouvido externo, é amplificado (em cerca de 22 vezes ^[28]) pelo sistema ossicular do ouvido médio e transmitido, como estímulo mecânico (ondas sonoras *líquidas*), ao líquido perilinfático no interior do labirinto coclear, localizado no ouvido interno. ^[29]

Figura 5 - Fisiologia do aparelho auditivo humano



O sistema auditivo consta basicamente de 3 etapas de processamento por onde o estímulo acústico é captado, amplificado, transmitido, analisado e interpretado.

A cóclea, dentro do ouvido interno, aloja a membrana basilar e o órgão de Corti, tecidos importantes no processo de transdução da energia acústica em energia eletroquímica. É nesta etapa que ocorre a análise e filtragem do sinal acústico (mecânico) similar à análise *wavelet*¹². Uma vez transduzida em impulsos nervosos¹³, a informação acústica segue seu trajeto primeiro pela via auditiva periférica, onde centros nervosos periféricos atuam no pré-processamento, em algumas funções interpretativas e na codificação de impulsos mais complexos para o sistema auditivo central, localizado no córtex.

Esta similaridade entre a análise da membrana basilar da cóclea e a análise *wavelet* é um fator que realça a importância do emprego de *wavelets* no estudo de processos acústicos e no modelamento e compreensão do sistema auditivo humano.^[47]

¹² A cóclea pode ser vista do ponto de vista funcional como um banco paralelo de filtros passabanda, cujas funções de transferência são determinadas em função do padrão de deslocamento verificado ao longo da membrana basilar (ressonância vibratória) para cada faixa de frequência. Acima de 800 Hz para humanos as respostas ao impulso destes filtros estão relacionadas entre si por uma dilatação, como na análise *wavelet*^[91]. Abaixo de 500 Hz o escalamento de frequência torna-se progressivamente mais linear^[91], e o mecanismo de interpretação do estímulo acústico é mais complexo^[28].

¹³ processo realizado pelas *células ciliadas* do órgão de Corti: “as células nobres do ouvido interno”^[29], altamente especializadas na transdução do som em impulsos eletroquímicos.

IMPLEMENTAÇÃO DA TRANSFORMADA WAVELET

A transformada *wavelet* integral (contínua), em se tratando de aplicações com sinais ópticos (analógicos), pode ser implementada através de filtros ópticos, à velocidade da luz, com todas as vantagens (e desvantagens) inerentes a esse tipo de processamento analógico. A implementação da transformada contínua em sistemas eletrônicos ficará entretanto sujeita aos fatores inerentes ao processamento elétrico, como limitações na resolução e precisão, margem de ruído estocástico, efeitos não lineares, etc.

Em princípio, a implementação da transformada *wavelet* contínua em sistemas digitais implica num processamento no qual as funções envolvidas, respectivos cálculos e transformações são realizados de forma a simular discretamente o processamento analógico. Simulações digitais de sistemas contínuos normalmente são acompanhadas de uma alta carga computacional a fim de se reproduzir o efeito analógico tão próximo quanto possível, e não se esquivam das limitações impostas pelo interfaceamento análogo-digital quanto à resolução da amostragem, erros de quantização, precisão da conversão, etc.

A alternativa natural é procurar abordar a transformada *wavelet* em sua forma matemática discreta, cuja implementação computacional é direta, dispensando os estágios sensíveis a erros e de difícil implementação em tempo real.

A transformada *wavelet* pode ser tratada sob a forma discreta sem prejuízo de suas qualidades e propriedades, e de maneira numericamente estável.

WAVELETS DISCRETAS

Tradicionalmente discretizam-se os parâmetros a (escala) e b (deslocamento), variáveis do sinal transformado $F(a,b)$, os **coeficientes *wavelet***. Para a toma-se valores inteiros (positivos e negativos), potências de um parâmetro fixo a_0

$$a = a_0^j, \quad a_0 > 1 \text{ e } j \in \mathbf{Z}$$

A discretização de b deve depender de j tal que *wavelets* estreitas (de alta frequência) seja deslocadas por passos pequenos, e *wavelets* largas (de baixa frequência) sejam deslocadas por passos maiores. Assim, uma escolha natural é

$$b = k b_0 a_0^{-j} \quad b_0 > 0, \text{ fixo, } \quad j, k \in \mathbf{Z}$$

A *wavelet* discreta fica então

$$\Psi_{j,k}(t) = a_0^{j/2} \Psi(a_0^j t - kb_0) \quad - (8) \text{ wavelet discreta}$$

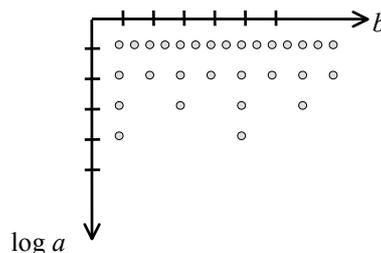
No caso discreto, a *resolução de identidade* - na equação (7) - não mais existe, e surge a questão básica sobre a possibilidade de se representar $f(t)$ em termos de $F(a,b)$, e de se recuperá-la de tais coeficientes. A resposta é sim, desde que satisfeitas algumas condições sobre o suporte e a regularidade da *wavelet*, e atendidos alguns requisitos matemáticos. Essencialmente a condição de admissibilidade permanece válida¹⁴. Existem formas diferentes de se trabalhar com *wavelets* discretas, e de se implementar a transformada discreta. Uma delas é através da utilização de bancos de filtros organizados num esquema piramidal, que levará também a uma representação em multiresolução do sinal. É importante ressaltar que neste caso tais requisitos matemáticos (acima mencionados) são traduzidos em restrições impostas sobre os coeficientes dos filtros utilizados na decomposição e reconstrução *wavelet*.

O fator a_0 não pode ser arbitrário. Diferentes valores de a_0 levam a *wavelets* diferentes, e bases ortonormais de *wavelets* só são conhecidas para valores racionais de a_0 [13]. A escolha mais natural é $a_0=2$. O fato das escalas musicais tradicionais se basearem em potências de 2 (oitavas) e a base numérica digital ser 2 é mais que uma coincidência em relação a esta escolha. Fazendo-se $b_0=1$, a *wavelet* da equação (8) fica:

$$\Psi_{j,k}(t) = 2^{j/2} \Psi(2^j t - k) \quad - (9) \text{ wavelet diádica}$$

A *wavelet* da equação acima, escala em oitavas, é conhecida por *wavelet diádica*. O plano tempo-escala (frequência) neste caso fica amostrado por uma **grade diádica**, como abaixo:

Figura 6 - Discretização do plano tempo-escala (b versus a em escala logarítmica)



¹⁴ Estas questões específicas extrapolam o escopo deste texto, e o leitor poderá encontrar explicações completas sobre elas em [13].

Wavelets diádicas constituem bases ortonormais, e permitem a caracterização de um sinal $f(t)$ sem redundância^[13]. A prova deste fato pode ser desenvolvida utilizando-se a análise em multiresolução como ferramenta, mostrando que qualquer função $f(t)$ pode ser aproximada numa precisão arbitrária por combinações lineares de *wavelets* ortonormais.^[13]

TRANSFORMADA WAVELET DISCRETA (TWD)

A transformada *wavelet* discreta (TWD) e sua inversa (TWDI) podem ser expressas da seguinte forma:

$$d_{j,k} = \langle \Psi_{j,k}(t), f(t) \rangle = \int f(t) \Psi_{j,k}(t) dt \quad - (10) \text{ (TWD)}$$

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \Psi_{j,k}(t) \cdot d_{j,k} \quad - (11) \text{ (TWDI)}$$

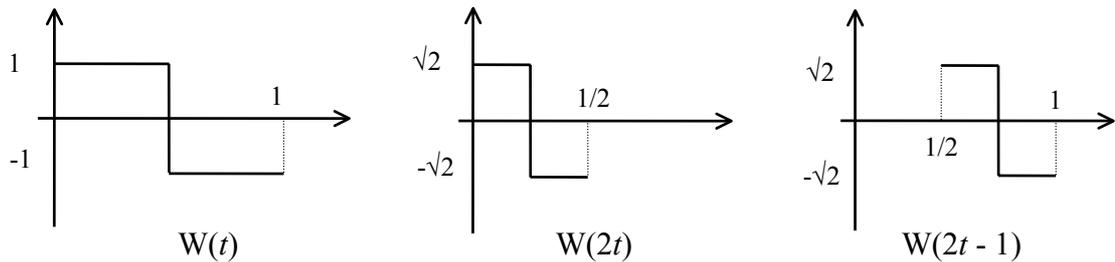
onde $d_{j,k}$ são os coeficientes *wavelets*, correspondentes a $F(a,b)$ da transformada integral.

TIPOS DE WAVELETS

Existem vários tipos de *wavelets* citados na literatura. O uso de uma ou outra está associado à aplicação. Regras de construção de *wavelets* estão sendo propostas por vários pesquisadores, segundo as restrições e necessidades que cada aplicação específica impõe. Isto nos leva a concluir que podemos gerar uma infinidade de *wavelets* diferentes, e particularmente construir um conjunto de *wavelets* adequado ao processamento de um tipo de sinal ou aplicação específica, levando à obtenção de resultados melhores.

Na Figura 7 mostra-se a *wavelet* de Haar, a mais simples das *wavelets*, introduzida por Haar por volta de 1910. Mostra-se graficamente as operações essenciais de contração (dilatação) e deslocamento, gerando $W(2t)$ e $W(2t-1)$.

Figura 7 - Wavelet de Haar



O próximo nível (não mostrado) contém $W(4t)$, $W(4t-1)$, $W(4t-2)$ e $W(4t-3)$, de onde exprimimos uma forma geral para esta família de *wavelets*

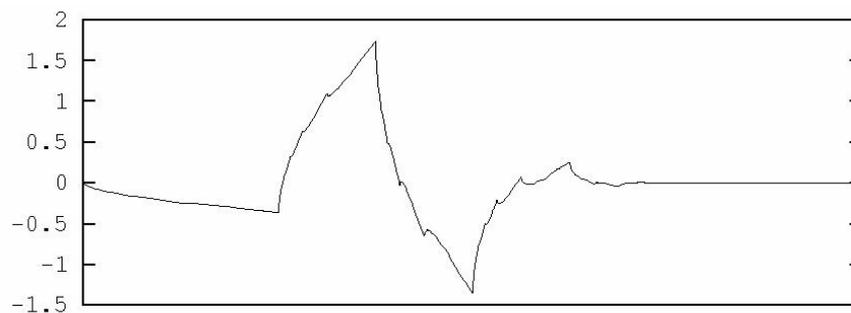
$$W_{j,k}(t) = 2^{j/2} W(2^j t - k) \quad j, k \in \mathbf{Z} \quad - (12)$$

Estas *wavelets* são contínuas por partes e constituem uma base para $L^2(\mathbf{R})$. Isto significa que as translações e deslocamentos de W são mutuamente ortogonais para todos os j e k

$$\int_{-\infty}^{+\infty} W(t) \cdot W(2t - k) dt = 0 \quad - (13)$$

Na figura 8 mostra-se uma *wavelet* ortonormal de suporte compacto de Daubechies (uma D4).

Figura 8 - Wavelet de Daubechies (D4)



As *wavelets* de Daubechies apresentam uma capacidade de análise e síntese muito mais efetiva do que as de Haar por possuírem maior regularidade (suavidade) e aproximarem melhor funções (suaves) em $L^2(\mathbf{R})$ [47; 58]. No caso de Haar, funções regulares¹⁵ são aproximadas por uma função com severas discontinuidades, o que introduz efeitos e artefatos indesejados na representação do sinal [47].

¹⁵ Sinais musicais, como tons estáveis (estacionários) de instrumentos de cordas e sopro, podem ser vistos como funções com certo nível de suavidade.

Como veremos adiante, a transformada *wavelet* pode ser produzida por um esquema de filtragem utilizando-se bancos de filtros. As origens das *wavelets* de Daubechies estão ligadas a famílias de filtros com propriedades especiais [78]. Dois canais de filtros existem nas implementações por banco de filtros, um estando associado às *wavelets* e outro associado às funções escaladoras ¹⁶. Os filtros de Daubechies são ortogonais, e exibem máxima planura (*maximum flatness*) em $\omega=0$ e $\omega=\pi$ [78], isto é, maximizam a suavidade nas funções escaladoras associadas maximizando a taxa de decaimento de suas transformadas de Fourier. [88]

Se desejarmos que as *wavelets* sejam úteis para análise de sinais regulares e suaves, é necessário impor condições sobre os filtros associados a elas além das exigências da capacidade de reconstrução perfeita e da preservação de energia. Condições impostas sobre o grau de regularidade da *wavelet*, sua taxa de decaimento no infinito e seu número de momentos nulos irão habilitá-las a melhor aproximarem e analisarem um maior número de classes de sinais e funções, bem como produzirem melhores resultados e desempenho. [47]

O grau de regularidade da *wavelet* e a sua taxa de decaimento é governada pelo número de momentos nulos que apresenta. Esta propriedade é importante para deduzir as propriedades de aproximação exibidas pela *wavelet* nos espaços de multiresolução. Momentos nulos também impõem uma condição necessária para que as *wavelets* sejam N vezes diferenciáveis, isto é, de classe C^N . [58]

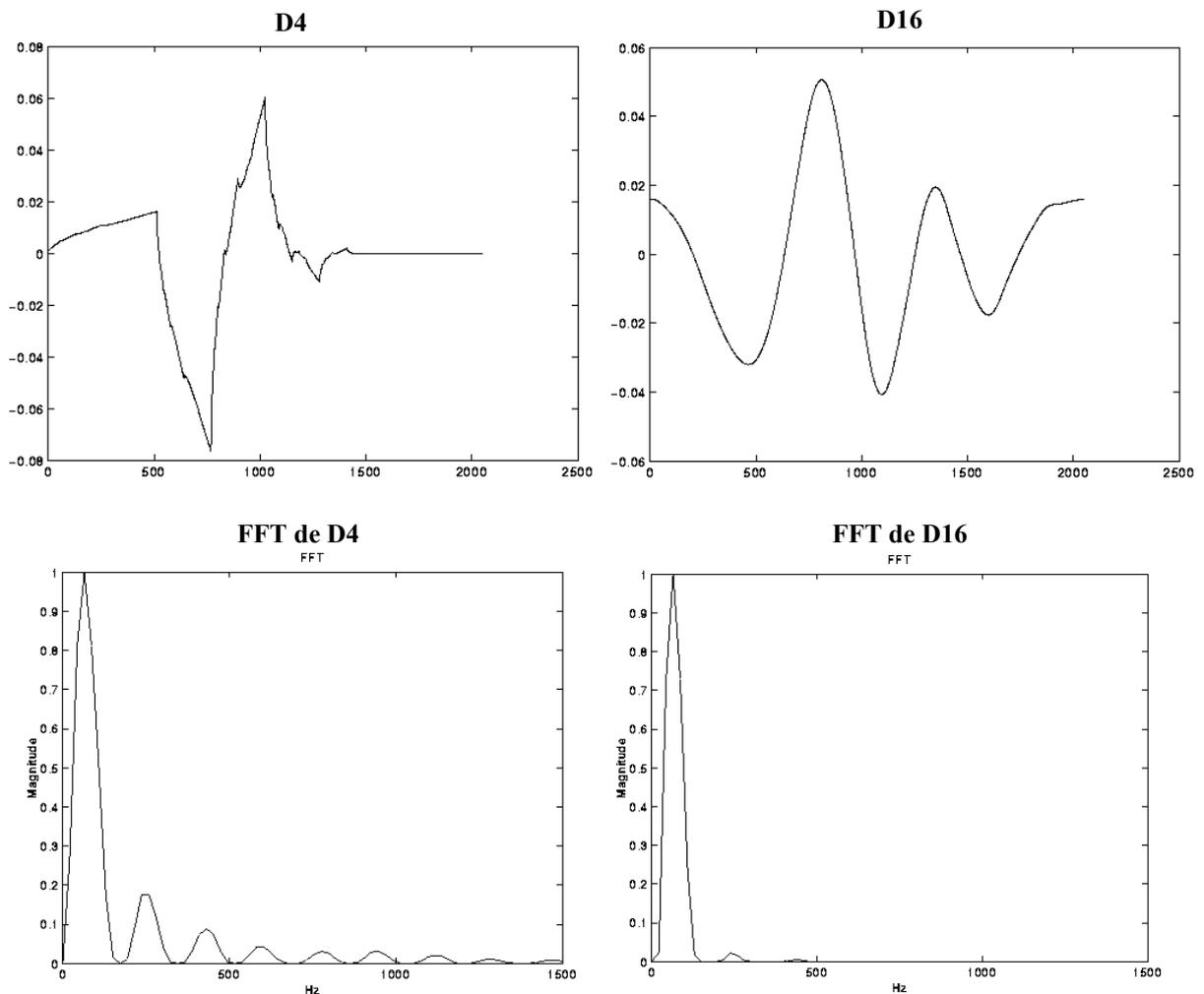
As *wavelets* de Daubechies são numeradas em função do número de momentos nulos que possuem. O índice em si corresponde ao número de coeficientes que os filtros associados possuem, que também é o número de derivadas nulas do filtro associado à função escaladora em $\omega=\pi$ (o filtro tem um zero em $\omega=\pi$ de ordem $N-1$). Assim, a *wavelet* D4 possui 2 momentos nulos, e os filtros associados possuem 4 zeros em $\omega=\pi$, e 4 *taps* (4 coeficientes). [58; 78]

As *wavelets* D16 possuem 8 momentos nulos, e seus filtros associados têm 16 coeficientes. A taxa de decaimento e grau de suavidade para esta *wavelet* são maiores do que para a D4, e elas se mostram mais eficientes para representar sinais musicais que

¹⁶ Estas serão introduzidas adiante na sessão dedicada à análise em multiresolução. Por ora, o objetivo é clarear a origem das *wavelets* de Daubechies para que se possa melhor descrevê-las em função de suas propriedades.

wavelets de menor regularidade. A Figura 9 abaixo mostra *wavelets* D4 e D16, juntamente com suas transformadas de Fourier¹⁷. A *wavelet* D16 exibe visivelmente uma forma de onda mais suave e uma melhor sintonia no espectro coberto que a D4.

Figura 9 - Wavelets de Daubechies D4 e D16, e respectivas transformadas de Fourier (FFT's)



As *wavelets* de Daubechies não possuem formas analíticas fechadas que as descrevam. Isso ocorre com muitas *wavelets* práticas inventadas e frequentemente encontradas na literatura.

Entre as *wavelets* mais celebradas e algumas referências bibliográficas que remetam a elas (sem a menor presunção de compilar uma lista completa) podemos citar:

- **Haar**, que pode ser considerada como sendo uma D2 (Daubechies 2): a primeira *wavelet*, com um único momento nulo; [6; 9; 12; 13; 14; 35; 42; 47; 60; 68; 69; 78; 88]

¹⁷ Transformadas rápidas de Fourier (*FFT*), calculadas com frequência de amostragem (f_a) em 44100 Hz, ou frequência de Nyquist em 22050 Hz, figuras utilizadas no padrão do *Compact Disc* (CD).

- **Daubechies**, de suporte compacto e suavidade "regulável"; [6; 9; 12; 13; 14; 35; 42; 47; 60; 68; 69; 78]
- **Coiflets**, projetada para satisfazer certo número de momentos nulos; [6]
- **Coifman**, cujos "filtros foram projetados tais que tanto a *wavelet* quanto a função escaladora tenham momentos nulos"; [13; 47; 88]
- **Beylkin**, cujos filtros apresentam raízes na proximidade da frequência de Nyquist; [6; 88]
- **Vaidyanathan**, cujo filtro exibe reconstrução exata apesar de não satisfazer nenhuma condição de momentos; o filtro foi otimizado para aplicações de voz; [6; 88]
- **Meyer**, que derivam das chamadas *wavelets* de Shannon, ou *Sinc wavelets*, que são suavemente enjaneladas na frequência tal que o decaimento no tempo (t) possa ser mais rápido que qualquer potência de t ; [47; 78]
- **Malvar**, *wavelets* ortonormais cuja descoberta se insere na estrutura geral de referência da análise enjanelada de Fourier, vindo a constituir um algoritmo de análise em tempo-frequência, em oposição à análise tempo-escala. Henrique Malvar (1987), brasileiro, professor na Universidade de Brasília, atualmente vice presidente de uma empresa norte-americana, desenvolveu as *wavelets* que levam seu nome em trabalho de doutorado no MIT. A semelhança entre a sua construção e a estrutura dos *grãos sonoros* ou átomos, possuindo um ataque, um período estacionário e um decaimento, estabelecem conexões com a teoria da "decomposição atômica", onde os átomos são de tempo-frequência. Os algoritmos para implementação de *wavelets* a partir de esquemas de filtragem todavia são mais simples, e trabalham com planos de tempo-escala. O *modus operandi* de uma análise com *wavelets* de Malvar extrapolam os objetivos deste trabalho, que também são os de utilizar a abordagem das *wavelets* via bancos de filtros. Meyer, entretanto, chama a atenção para o fato de que as *wavelets* de Malvar poderiam ser úteis para aplicações de voz e som, mas introduz uma abordagem às *wavelets* de Malvar em favor mais da síntese e transmissão e menos da análise; [47]
- ***wavelets* biortogonais**, para as quais a restrição da ortogonalidade é relaxada; [47; 78]
- ***wavelets* simétricas** ou **Simlets** (*symmlets*), ou *wavelets* "menos assimétricas", de suporte compacto e número de momentos nulos variando de 4 a 10; [6; 88]

- **bases *wavelets* ótimas**, que fornecem a decomposição mais compacta (comprimida) de um sinal, ^[47; 88]

e mais uma diversidade de variantes, que usualmente recebem os nomes de seus desenvolvedores ou das técnicas que lhes dão origem, como é o caso com as novas "*wavelets* de segunda geração" (usando a terminologia de Sweldens (1995)), obtidas pelo esquema de *lifting*, o qual não leva necessariamente a *wavelets* que sejam dilatações e translações de uma matriz. ^[79]

As *wavelets* D16 apresentaram os melhores resultados nas análises de eventos musicais realizadas neste trabalho, superiores aos alcançados utilizando-se D4 ou *wavelets* de baixa regularidade, e semelhantes aos resultados obtidos com outras *wavelets* com idêntico número de momentos nulos (regularidade). Isso se deve a sua melhor sintonia (suporte compacto na frequência), suavidade e desempenho computacional. Por essas razões foram utilizadas maciçamente neste trabalho. O fato de ser talvez a *wavelet* mais celebrada e freqüente na literatura, cujas propriedades são conhecidas e bem documentadas, também contribuiu para sua escolha como principal *wavelet* de análise neste trabalho.

Em momento oportuno a questão da escolha da *wavelet* será melhor abordado adiante.

2.3 *Wavelets* em processamento de sinais

Wavelets tem sido amplamente exploradas em um grande número de aplicações em processamento de sinais. Várias delas já foram, e serão ainda comentadas neste texto. A seguir, um esboço das principais aplicações de *wavelets* em processamento de sinais. Devido a multiplicidade de referências a uma mesma aplicação comum, optou-se por citar somente as referências mais importantes disponíveis e/ou que abordem o tema com mais cuidado.

Talvez as *wavelets* mais populares têm sido as de Daubechies, pela sua ortogonalidade e suporte compacto. As *wavelets* de Gabor também se mostram populares em aplicações em segmentação de imagens¹⁸. As ferramentas de implementação variam de aplicação

¹⁸ Veja o servidor Internet do grupo de visão computacional (*Computer Vision Group, Computer Science III*) da Universidade de Bonn, na Alemanha: <http://www-dbv.informatik.uni-bonn.de/image/overview.html> para detalhes sobre a utilização destas *wavelets* em aplicações de segmentação de imagens.

para aplicação, entretanto nota-se uma grande utilização de bancos de filtros e de algoritmos piramidais de filtragem para implementação da transformada *wavelet*. Há aplicações que requerem muitas vezes a transformada contínua (cf. [87]).

Em processamento de imagens há versões naturalmente extrapoladas para 2 dimensões dos algoritmos *wavelets*, empregando, em especial, os esquemas piramidais baseados em implementações em multiresolução. Em computação gráfica *wavelets* têm sido usadas para edição de curvas, manipulação de superfícies [75; 76], análise de textura¹⁹, compressão de imagens [13; 26; 35; 77; 78], segmentação, segmentação de imagens texturizadas²⁰, quantização [88], enfoque, magnificação, interpolação, pintura em multiresolução [55], representação de curvas, análise de superfícies, representação de fluxo de luz, radiosidade, modelamento geométrico, modelos fractais [69], visão computacional [31; 42] e óptica difrativa, entre outras.

Mallat (1989) descreveu o uso de *wavelets* em discriminação de textura e análise fractal em imagens [42]. Stollnitz e outros (1995) também mostraram as capacidades seletivas das *wavelets* na edição de curvas em imagens, alterando aspectos globais de contorno sem afetar detalhes na imagem [75; 76].

Wavelets têm sido aplicadas para remoção de ruídos, reconhecimento e extração de padrões, e realce de características em sinais [26; 31; 78; 85; 88]. Graps (1995) e Vidakovic (1991) abordaram o uso de *wavelets* em aplicações de extração de ruídos de sinais de ressonância magnética nuclear [26] e de sinais sísmicos [85].

Têm sido empregadas em interpretação de imagens e sinais biomédicos, detecção de sinais [31], em interpretações sísmicas e em geofísica [31; 85], em análise de voz, acústica, análise e síntese, música, representações auditivas, identificação de *pitch*, psicoacústica e identificação de padrões sônicos [18; 19; 26; 31; 33; 36; 37; 38; 50; 72; 78; 91; 92], como abordado mais detalhadamente a seguir.

¹⁹ Veja o servidor Internet em <http://www.ruca.ua.ac.be/~VisionLab/WTA.html>.

²⁰ Com aplicações em robótica, visão guiada automática, inspeção de qualidade de produtos, diagnósticos médicos, análise de imagens remotas, análise de movimento, visão estereoscópica, reconhecimento de objetos e interpretação de cena *-scene interpretation*. Todas áreas que dependem de uma segmentação de imagens de alta qualidade.

WAVELETS NA ANÁLISE E SÍNTESE DE SONS E TIMBRES MUSICAIS

J.-C. Risset (1989) destaca o potencial de aplicações desta ferramenta no campo da computação musical ^[61]. Kronland-Martinet (1987/88) iniciou a demarcação de um território para aplicações musicais de *wavelets* ^[38]. Scott Levine (1994) reconhece na transformada *wavelet* uma ferramenta adequada para analisar sinais de áudio pela sua similaridade e adequação aos modelos atuais do ouvido humano ^[89]. Mas o interesse por *wavelets* na área de áudio resulta da verificação de seu extraordinário desempenho em aplicações para reconhecimento de voz, identificação de padrões, e análise de sinais acústicos. ^[78]

Em 1985, S. Mallat (1985) estabeleceu uma conexão da teoria *wavelet* com outras teorias amplamente empregadas na área de processamento de sinais, abrindo o caminho para Ingrid Daubechies (1988) e suas celebradas *wavelets* de suporte compacto ^[12]. Desde então o número de contribuições teóricas e práticas cresceu substancialmente, existindo um grande número de trabalhos documentados abordando a aplicação de *wavelets* em processamento de áudio e música. A seguir, uma compilação de contribuições por área de aplicação:

Acústica e música

Newland (1994) propõe as *wavelets musicais* ^[50], com discriminação de frequência superior à das *wavelets harmônicas* ^[51] (também de sua autoria), cujos intervalos de frequência (destas últimas) são oitavas. Kronland-Martinet (1988) aborda a transformada *wavelet* para análise, síntese e processamento de voz e sons musicais ^[38]. Karlsen (1995) examina a aplicação de *wavelets* na identificação de veículos terrestres pela análise dos sinais acústicos emitidos (no que se refere como identificação de assinaturas acústicas). ^[37]

Detecção de pitch

Yip (1995) propõe um algoritmo para detecção de *pitch* em sinais de voz em ambientes ruidosos utilizando o espectro de fase da análise tempo-frequência com *wavelets*, verificando um desempenho notável, superior a outros métodos ^[92]. Shelby (1995) também explora o uso de *wavelets* na detecção de *pitch* de voz, obtendo resultados comparáveis a outras técnicas, e avaliando o impacto da escolha da base *wavelet* no desempenho final do sistema ^[67]. Kadambe (1992) descreve um detetor de *pitch* de voz

baseado na transformada *wavelet*, constatando um desempenho superior em comparação a outros métodos clássicos, e com vantagens de maior imunidade a ruídos e baixa complexidade computacional. [36]

Representações auditivas

Drake (1993) considera o uso de uma representação *wavelet* de voz em um algoritmo para compensar uma deficiência auditiva que altera a faixa dinâmica percebida, e menciona vantagens da abordagem com *wavelets* neste tipo de aplicação [18]. Yang e outros (1992) apresenta uma estrutura de referência geral, analiticamente tratável, para descrever as transformações que os sinais acústicos sofrem nos primeiros estágios do sistema auditivo, e desenvolveu modelos algorítmicos para representar os sinais em cada estágio do processamento, utilizando processamento multiescalar, o método de projeções convexas, e representações *wavelet* para interpretar etapas do processamento coclear. [91]

Irino e Kawahara (1993) também utilizam a transformada *wavelet* para simular o sistema auditivo periférico humano (modelando a resposta em frequência da cóclea) para experimentos psicofísicos. Eles propõem um método para modificar sinais acústicos manipulando sua representação auditiva, e, partindo do pressuposto de que "a transformada *wavelet* simula as características da periferia²¹ do sistema auditivo quando uma *wavelet* de análise é apropriadamente selecionada a partir da resposta impulsiva de um modelo auditivo", definem uma "**transformada *wavelet* auditiva**" (*AWT - auditory wavelet transform*) introduzindo as características do modelo coclear na transformada *wavelet* [33]. Uma das vantagens do uso da transformada *wavelet* em representações do sistema auditivo está exatamente na semelhança existente entre a sua resposta em frequência e a da cóclea, ambas perfazendo uma análise espectral com fator de qualidade Q constante ao longo da frequência. [33, 60, 91]

Compressão de áudio

Sinha (1993) descreve o uso de procedimento para compressão de áudio baseado na seleção adaptativa da base *wavelet* em conjunto com um abordagem de dicionário dinâmico, obtendo para áudio em qualidade de CD (amostrado a 44,1KHz e resolução de 16 bits) taxas de transmissão de 48-66 Kbits/s [72]. Scholl (1994) empregou um algoritmo variante da transformada *wavelet* para compressão de sinais de áudio (que

denomina *circular wavelet-packets*) reportando a eliminação de grande parte dos efeitos (indesejáveis) de reconstrução de bordas.^[66]

Extração de características (feature extraction)

Evangelista (1993) introduz uma nova representação *wavelet* baseada numa representação vetorial *pitch*-síncrona, permitindo-se identificar e separar os segmentos pseudo-periódicos e aperiódicos do sinal de voz ou música, e servindo à extração de padrões, à separação de consonantes vocais e à compressão de dados^[19]. Seu trabalho neste campo serviu de estímulo à consecução da presente pesquisa, e é comentado em maiores detalhes no capítulo 4 (Análise em multiresolução de gestos musicais com *wavelets*) mais adiante.

Além destas áreas, *wavelets* têm encontrado aplicações não menos importantes em óptica não linear, análise de séries temporais, estatística e mecânica estatística (sistemas dinâmicos Hamiltonianos, sistemas Hamiltonianos não lineares²²), matemática aplicada e computacional²³, física matemática, análise de espaço de estados, análise de processos estacionários, solução de equações diferenciais e integrais, análise matricial combinatória e aplicações em ciências biológicas, operadores, teoria dos estados coerentes, teoria dos grafos, teoria de grupos, teoria de jogos, álgebra booleana, química, biologia, ecologia, oceanografia, aerodinâmica, meteorologia, engenharia petrolífera, de águas profundas e contaminação, eletromagnetismo, ondas aquáticas, e solução de problemas com valores de contorno^[8; 25; 31]. A alternativa mais veloz atualmente para buscas de novidades nestes campos está nos milhares de servidores Internet espalhados pelo globo.

2.4 Análise em Multiresolução

Até aqui mostraram-se representações de uma função contínua $f(t)$ decomposta sobre uma base de *wavelets* contínuas $\psi_{j,k}(t)$, e expressa então por coeficientes discretos $d_{j,k}$. O interesse, no entanto, é obter um algoritmo que nos permita representar sinais discretos $f(n)$ em termos de uma combinação linear de *wavelets* discretas $\psi_{j,k}(n)$.

²¹ primeiros estágios.

²² Meiss, J.D., professor de matemática aplicada, pesquisas da Universidade da Califórnia, Berkeley.

A análise em multiresolução leva naturalmente a um esquema rápido e hierárquico para a computação dos coeficientes *wavelets*, como veremos a seguir. A estrutura de implementação é identificada com os algoritmos de codificação sub-banda, utilizados em compressão de voz ^[60], e com os algoritmos piramidais, empregados em processamento de imagens ^[47] e visão computacional.

Numa análise em multiresolução um sinal $f(t)$, $t \in \mathbf{R}$, é decomposto em aproximações sucessivas de resolução cada vez menor, numa sequência de estágios de processamento consecutivos.

Nos itens que se seguem apresenta-se a teoria da multiresolução para sinais contínuos, igualmente válida para o caso discreto. Com a utilização de bancos de filtros pode-se implementar uma análise em multiresolução com *wavelets* na forma discreta.

TEORIA DA MULTIRESOLUÇÃO

Uma análise em multiresolução consiste numa sequência de espaços (fechados) de aproximações sucessivas V_j . Cada subespaço V_j está contido no próximo subespaço V_{j+1} . Uma função em um subespaço está em todos os subespaços mais finos:

$$\dots V_{-1} \subset V_0 \subset V_1 \subset \dots \subset V_j \subset V_{j+1} \subset \dots$$

Uma função $f(t)$ decomposta nestes espaços tem um pedaço em cada subespaço. Este pedaço - a projeção de $f(t)$ em V_j - é $f_j(t)$. A união de todos os subespaços é $L^2(\mathbf{R})$, e as interseções entre eles é o espaço vazio ($\bigcap_{j \in \mathbf{Z}} V_j = \{0\}$).

Há ainda requisitos adicionais para haver multiresolução ^[9; 13; 35; 78]:

- Completeza: $f_j(t) \rightarrow f(t)$ para $j \rightarrow \infty$, e nulidade: $\|f_j(t)\| \rightarrow 0$ para $j \rightarrow -\infty$
- V_{j+1} compõe-se de todas as funções reescaladas em V_j : $f(t) \in V_j \Rightarrow f(2t) \in V_{j+1}$
- Invariância ao deslocamento: $f(t) \in V_j \Rightarrow f(t - 2^j \cdot k) \in V_j$
- Há uma base ortonormal $\{\phi_{j,k}, j, k \in \mathbf{Z}\}$ para cada subespaço V_j , isto é, as funções $f_j(t)$ - projeções de $f(t)$ no nível j - podem ser descritas como combinações lineares de $\phi_{j,k}$.

²³ Veja o servidor Internet http://amath.colorado.edu/appm/department/fac_summ.html para informações suplementares nestes campos e outros.

Chamamos ϕ a “**função escaladora**” da análise em multiresolução, que dá origem à família ortogonal $\phi_{j,k}$

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k) \quad - (14)$$

que cobre todo nível j e consiste numa versão escalada e deslocada da função escaladora $\phi(t)$, também referenciada como a *wavelet* pai, em contrapartida à terminologia adotada para $\psi(t)$, a *wavelet* mãe.

O ESPAÇO DOS DETALHES OU ESPAÇO WAVELET: W_j

A função $f_{j+1}(t) \in V_{j+1}$ possui uma resolução melhor que $f_j(t) \in V_j$. A parte que falta para aproximar $f_{j+1}(t)$ de $f_j(t)$ é o **detalhe** $\Delta f_j(t)$ que se encontra num novo espaço complementar a V_j : o **espaço** W_j . Portanto,

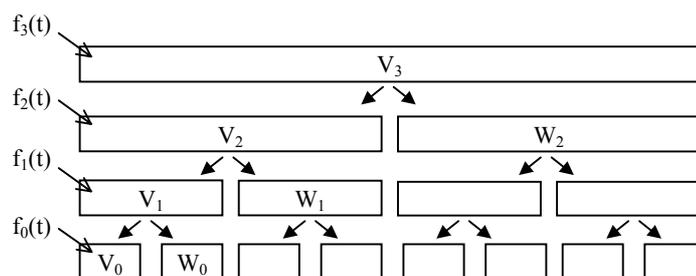
$$\Delta f_j(t) = f_{j+1}(t) - f_j(t), \quad \text{onde } \Delta f_j(t) \in W_j.$$

Do ponto de vista dos subespaços,

$$V_j \oplus W_j = V_{j+1}.$$

O subespaço W_j consiste no complemento ortogonal de V_{j+1} em V_j . A Figura 10 abaixo ilustra esta hierarquia de relações entre os espaços V_j e W_j . Os espaços mais finos (no alto) se decompõem em uma versão menos resolvida - um "molde estrutural" - e num conjunto de detalhes. O molde estrutural ainda pode ser novamente decomposto em duas novas versões - um novo molde e um novo conjunto de detalhes. O processo, como sugerido, é iterado sucessivamente, culminando com a completa decomposição do espaço mais fino. Na direção oposta, isto é na síntese, o molde é enriquecido de detalhes, ampliando assim sua resolução e formando os níveis imediatamente mais finos.

Figura 10 - Associação dos espaços V_j e W_j numa análise em multiresolução



Segue que

$$V_{j+1} = W_j \oplus W_{j-1} \oplus W_{j-2} \oplus \dots \quad , \text{o que implica em}$$

$$f_{j+1}(t) = \Delta f_j + \Delta f_{j-1} + \dots + \Delta f_1 + \Delta f_0 + \Delta f_{-1} + \dots = \sum_{-\infty}^j \Delta f_j(t) \quad - (15)$$

ou seja, a versão $f_{j+1}(t)$ pode ser descrita em termos da somatória das contribuições de todos os detalhes nos níveis inferiores.

Naturalmente sucede que a união de todos os W_j também produz o espaço $L^2(\mathbf{R})$, e os requisitos acima impostos à V_j também se aplicarão à família de subespaços (fechados) W_j .

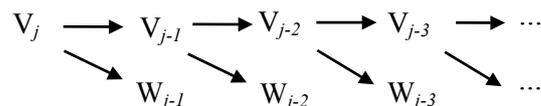
A família de funções $\{\psi_{j,k}(t), k \in \mathbf{Z}\}$ constitui uma base ortonormal para o subespaço W_j . Uma extensão deste fato nos permite afirmar que a coleção inteira $\{\psi_{j,k}(t), j, k \in \mathbf{Z}\}$ constitui uma base ortonormal para $L^2(\mathbf{R})$, que é chamada de **base wavelet** de $L^2(\mathbf{R})$, com

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \quad - (16)$$

mantendo a coerência com $\phi_{j,k}(t)$.

A estrutura que conecta os subespaços V_j e W_j pode ser ilustrada também num formato de árvore, como abaixo:

Figura 11 - Árvore de conexão entre os espaços V_j e W_j



As *wavelets* são uma base para o espaço $L^2(\mathbf{R})$ inteiro, mas a função escaladora ϕ em $j=0$ e as *wavelets* com $j \geq 0$ são uma base mais prática. Tomando o nível V_0 como o nível de menor resolução da análise, eliminamos os cálculos para $j < 0$, e a estrutura piramidal pára no nível mais baixo V_0 . Assim,

$$V_{j+1} = W_j \oplus W_{j-1} \oplus W_{j-2} \oplus \dots \oplus V_0$$

e podemos recuperar $f(t)$, decomposta num conjunto de subespaços V_j e W_j , através de um número (infinito) de operações:

$$f(t) = \sum_{-\infty}^{+\infty} \Delta f_j(t) = \sum_{-\infty}^{+\infty} \sum_{-\infty}^{+\infty} d_{j,k} \psi_{j,k}(t) = f_0(t) + \sum_{j=0}^{+\infty} \Delta f_j(t) = f_0(t) + \sum_{j=0}^{+\infty} \sum_{-\infty}^{+\infty} d_{j,k} \psi_{j,k}(t) \quad - (17)$$

para $j \geq 0$, onde $d_{j,k}$ são os **coeficientes wavelet de $f(t)$** .

De $f_0(t)$ (nível V_0) e todos os Δf_j ($j > 0$) extraem-se as outras $f_j(t)$ via operações matemáticas através da pirâmide (ou árvore, como na Figura 11 acima). Podemos parar em alguma escala 2^{-j} (nível J , onde está $f_j(t)$) com suficientes componentes de alta frequência (resolução fina) para reproduzir o sinal tão exatamente quanto possível, isto é, na prática podemos tomar $f(t) = f_j(t)$ no nível mais fino (J), e operarmos diretamente sobre $f(t)$. Na direção oposta (descendo a pirâmide), decompõem-se f_j em sucessivas aproximações mais grosseiras e obtém-se descrições menos finas de $f(t)$ (numa escala maior, exatamente como em mapas geofísicos), até o limite de pior resolução em f_0 .^[78]

2.5 Análise em multiresolução com wavelets

O ALGORITMO DE DECOMPOSIÇÃO E RECONSTRUÇÃO WAVELET

Precisamos obter uma descrição de $f(t)$ em diferentes escalas. É desejada a habilidade de se trafegar de uma aproximação de baixa resolução de $f(t)$ em direção às mais finas, onde mais detalhes estejam disponíveis (melhor resolução) e vice-versa, bem como ter acesso aos sinais em qualquer escala. A estrutura de referência da multiresolução, agora confinada entre o nível de mais alta resolução $j=J$ e o de mais baixa $j=0$, oferece o meio necessário para se implementar esse processamento. Neste esquema, as projeções de $f(t)$ nos subespaços V_j e W_j para $j \in [J,0]$ (intervalo finito) são relacionadas por:

$$f_j = f_{j-1} + \Delta f_{j-1}, \quad \text{o que por iteração fornece} \quad f_j = \Delta f_{j-1} + \Delta f_{j-2} + \dots + \Delta f_0 + f_0.$$

Há uma relação íntima entre $\phi(t) \in V_0$ e $\phi(2t-k) \in V_1$ conhecida por *relação entre duas escalas*, a qual decorre diretamente do fato de $\phi \in V_0 \subset V_1$. Esta relação é expressa por meio da **equação de dilatação**, também chamada de equação de refinamento, porque mostra $\phi(t)$ num espaço mais fino V_1

$$\phi = \sum_n h_n \phi_{1,n} = \sum_n 2^{1/2} h(n) \phi(2t - n) \quad \text{-(18) eq. de dilatação}$$

Uma relação entre duas escalas semelhante também ocorre entre $\psi(t) \in W_0$ e $\psi(2t-k) \in V_1$, a qual dá origem à **equação wavelet**, decorrente de $\psi \in W_0 \subset V_1$

$$\psi = \sum_n g_n \phi_{1,n} = \sum_n 2^{1/2} g(n) \phi(2t - n) \quad - (19) \text{ eq. wavelet}$$

Destas relações, derivam duas novas sequências: $h(n)$ e $g(n)$, que vêm a ser os coeficientes dos filtros associados respectivamente à função escaladora $\phi(t)$ e à *wavelet* $\psi(t)$.²⁴

Imediatamente, segue-se que uma função $f_1(t) \in V_1$, pode ser descrita como combinação de duas versões um nível de resolução abaixo, porque $V_1 = V_0 \oplus W_0$. Isto, juntamente com as relações de duas escalas acima, leva às fórmulas de **decomposição** [13; 78]:

$$c_{j-1,k} = \langle f, \phi_{j-1,k} \rangle = \sum_n h_{n-2k} c_{j,n} \quad - (20)$$

$$d_{j-1,k} = \langle f, \psi_{j-1,k} \rangle = \sum_n g_{n-2k} c_{j,n} \quad - (21)$$

onde $d_{j,k}$ são os **coeficientes wavelet** de $f(t)$. Definimos agora $f_j(t)$ e $\Delta f_j(t)$ como

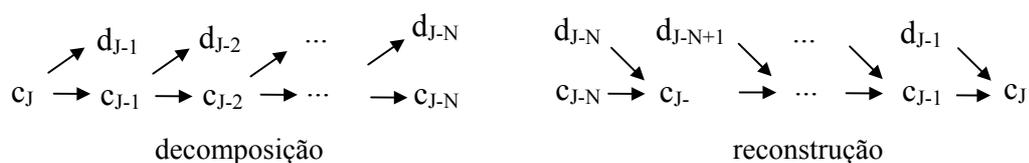
$$f_j(t) = \sum_k c_{j,k} \phi_{j,k}(t) \quad \text{e} \quad \Delta f_j(t) = \sum_k d_{j,k} \psi_{j,k}(t)$$

onde $c_{j,k} \in V_j$ e $d_{j,k} \in W_j$. Sendo $f_j = f_{j-1} + \Delta f_{j-1}$, o algoritmo de **reconstrução** fica [13; 78]:

$$c_{j+1,n} = \sum_k [\overline{h_{n-2k}} c_{j,k} + \overline{g_{n-2k}} d_{j,k}] \quad - (22)$$

Estas expressões são o cerne da decomposição (transformada direta) e reconstrução (transformada inversa) *wavelet*, e podem ser esquematizadas de forma análoga aos algoritmos piramidais e de codificação sub-banda, como na figura abaixo:

Figura 12 - Esquema de decomposição e reconstrução wavelet



²⁴ Como veremos adiante, estas sequências serão a chave para uma implementação prática de uma análise em multiresolução com bancos de filtros.

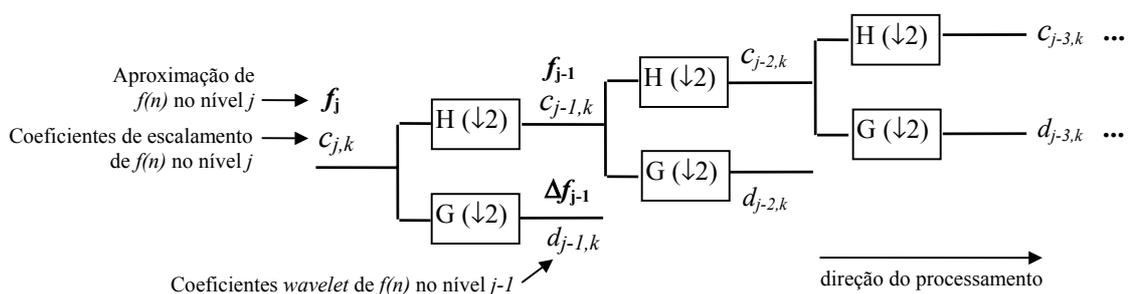
A figura sugere um esquema hierárquico para se obter todos os coeficientes *wavelet* de um sinal eliminando a necessidade de se calcular o produto interno $\langle f, \psi_{j,k}(t) \rangle$ para cada $d_{j,k}$. É a natureza recursiva deste algoritmo *wavelet* que o torna computacionalmente veloz e eficiente, atraindo a atenção da comunidade de processamento de sinais. A introdução da teoria da multiresolução abriu o caminho para descobrir-se as conexões entre os algoritmos piramidais, as estruturas de codificação sub-banda, os filtros de espelhamento de quadratura (QMF) e a teoria *wavelet*.^[42; 47]

Strang (1993) analisa em^[77] as implementações matriciais da transformada *wavelet*, mostrando que a matriz de transformação pode ser fatorada em matrizes esparsas, num processo que é equivalente ao algoritmo piramidal acima, dando origem à *transformada wavelet rápida*, que permite uma redução na complexidade algorítmica de $O(n \cdot \log n)$ para $O(n)$, onde n é a extensão do sinal (discreto). O algoritmo acima pode ser implementado utilizando-se bancos de filtros num arranjo piramidal, como veremos a seguir.

2.5.1 Implementação via banco de filtros

Uma análise em multiresolução pode ser vista como um sistema de filtros, esquematizados num arranjo piramidal (em formato de árvore) como abaixo^[78]:

Figura 13 - Análise em multiresolução com banco de filtros



Um **banco de filtros** é um conjunto de filtros conectados por operadores amostradores ou decimadores ($\downarrow 2$) e, em alguns casos, por módulos de atraso^[78]. Os filtros H e G neste arranjo formam um banco de filtros, e podem ser escolhidos tal que realizem os algoritmos de decomposição das equações 20 e 21 acima.

O banco opera sobre um sinal de entrada f_j filtrando-o por dois canais distintos, geralmente um passa-alta (G) e outro passa-baixa (H), e iterando-se o processo no canal

passa-baixa, como mostra a figura acima. Em cada etapa geram-se 2 aproximações de f_j de menor resolução: $f_{j-1} \in V_{j-1}$ e $\Delta f_{j-1} \in W_{j-1}$.

Em implementações práticas, trabalhamos diretamente com os coeficientes $c_{j,k}$ e $d_{j,k}$ - representativos do sinal original- na árvore de filtros. Começamos com um nível $j=J$ suficientemente fino²⁵ ($f_j=f_J$, descrito na escala 2^{-J} , ao nível J) e descemos a árvore até o nível $j=0$. Se o vetor de entrada $c_{J,k}$ tem extensão $N=2^J$, atingimos o nível $j=1$ com 2 elementos, quase no nível de resolução mais baixa da árvore. [78]

H é o canal passa-baixa, que computa suavizações (médias), gera os subespaços V_j , e reduz a resolução da sequência de entrada pela metade em cada etapa da decomposição. G é o canal passa-alta, que extrai os detalhes (diferenças) em cada etapa e gera os subespaços W_j , os subespaços *wavelet*. Os passos de subamostragem ou decimação ($\downarrow 2$) selecionam os componentes pares da sequência de entrada, e dobram a escala em cada etapa:

$$(\downarrow 2)x(k) = x(2k)$$

O processo de decimação, ao contrário da filtragem que é linear e invariante no tempo, é variante no tempo e não invertível. Seu uso em banco de filtros, entretanto, não prejudica as propriedades desse arranjo especial, entre as quais a sua invertibilidade ou capacidade de reconstrução perfeita.

O banco de filtros implementa uma ***análise wavelet em multiresolução*** quando os coeficientes dos filtros H e G correspondem àqueles dos filtros associados às sequências $h(k)$ e $g(k)$ previamente apresentadas: os coeficientes dos filtros associados à função escaladora e *wavelet*. ***Análise em multiresolução*** porque o repetido reescalamto em cada etapa produz detalhes em todas as escalas/resoluções. ***Análise wavelet*** porque no limite do processo de filtragem do banco, a iteração do canal passa-baixo produzirá a função escaladora e a *wavelet*. Em outras palavras, a conexão entre o banco de filtros discretos e as *wavelets* contínuas está no limite da árvore de filtragem.

A vantagem em se utilizar bancos de filtros para realizar uma análise *wavelet* em multiresolução está na facilidade da sua implementação na forma discreta, empregando-se filtros digitais. Com a escolha apropriada dos filtros, o processo de filtragem passa a realizar na verdade uma transformação *wavelet*, decompondo o sinal de entrada em

coeficientes $d_{j,k}$, indexados pelos parâmetros escala (j) e deslocamento (k). Os requisitos impostos às *wavelets* no item 2.2 para que haja a transformada e a inversa ficam traduzidos no contexto dos bancos de filtros em restrições sobre os coeficientes dos filtros a serem satisfeitas.

O primeiro passo é aproximar a função $f(t)$ por uma função $f(n)$, discreta, com resolução arbitrariamente fina, suportada numa extensão finita $[0, N-1]$ e constante por partes em intervalos $[\ell N, (\ell+1)N]$. Esse procedimento consiste numa amostragem do sinal original $f(t)$ gerando uma aproximação discreta $f(n)$, que pode ser tão fina (ou resolvida) quanto se desejar. Em certas aplicações podemos tomar os valores de $f(n)$ diretamente como vetor de entrada, i.é, fazendo os coeficientes $c_{j,k} = f(n)$. Isto corresponde a aproximar $f(n)$ por seus coeficientes $c_{j,k}$ num nível J arbitrariamente fino.

A análise proverá no próximo nível uma versão menos resolvida de $f(n)$ numa escala duas vezes maior - $c_{j-1,k}$ - e também os detalhes para que esta versão menos fina possa reconstruir o sinal original, no nível superior mais fino - $d_{j-1,k}$ (os coeficientes *wavelet*). O processo iterado no canal passa-baixa levará à geração de todos os coeficientes *wavelet* $d_{j,k}$ de $f(n)$, que contém por toda a árvore os detalhes de $f(n)$ separados em bandas de frequências.

Na síntese inverte-se a direção do processo, exatamente como na síntese em multiresolução contínua. Substitui-se a subamostragem ($\downarrow 2$) por uma superamostragem ($\uparrow 2$)²⁶ e os filtros de análise H e G pelos de síntese \bar{H} e \bar{G} .

\bar{H} e \bar{G} podem satisfazer diversos critérios de formação e relações com o banco de análise H e G . A operação de síntese é inversa à de análise. O banco de síntese dever ser portanto o inverso do banco de análise se a reconstrução perfeita é desejada. Parte do projeto dos filtros consiste em garantir que $(\downarrow 2)Hf(n) = u(k)$ e $(\downarrow 2)Gf(n) = v(k)$ irão produzir toda a informação necessária para recuperar $f(n)$ no processo inverso, com $\bar{H}(\uparrow 2)u(k)$ e $\bar{G}(\uparrow 2)v(k)$. No caso de transformações ortogonais os bancos de filtros também serão ortogonais, e a relação de invertibilidade entre síntese e análise ganha

²⁵ tão fino quanto possível para consistir numa excelente aproximação (f) para o sinal original sob análise ($f(n)$ ou $f(t)$)

²⁶ Superamostragem^[78]: $y = (\uparrow 2)x(k) = \begin{cases} y(2k) = x(k) \\ y(2k+1) = 0 \end{cases}$

uma expressão matemática direta e simples, melhor visualizada sob uma abordagem matricial dos bancos.

2.5.2 Abordagem matricial

As operações num banco de filtros como acima podem ser representadas matricialmente, através de multiplicação de matrizes. Para se transformar uma sequência de extensão N nos seus N coeficientes (gerados pela decomposição) é necessário uma matriz $N \times N$. A transformada inversa envolve o uso da matriz inversa:

$$\text{Síntese: } \mathbf{x} = \mathbf{W} \cdot \mathbf{b} \qquad \text{Análise: } \mathbf{b} = \mathbf{W}^{-1} \cdot \mathbf{x}$$

onde \mathbf{b} é o vetor transformado (com N coeficientes *wavelet*), \mathbf{W} é a matriz de transformação *wavelet* montada com os coeficientes dos filtros do banco ($h(n)$ e $g(n)$) e \mathbf{x} é o vetor de entrada (de extensão N).

A escolha do filtro passa-alta e passa-baixa (equivalente à escolha da base *wavelet*) exercerá uma forte influência nas propriedades verificadas pelas matrizes. Por exemplo, se H e G são ortogonais, o banco de filtros é ortogonal e a matriz correspondente será também ortogonal. Por meio de uma normalização a ortogonalidade passa a ortonormalidade, e temos que

$$\mathbf{W}^T \cdot \mathbf{W} = \mathbf{I} \Rightarrow \mathbf{W}^{-1} = \mathbf{W}^T \quad - (23)$$

isto é, o banco de síntese é o transposto (conjugado) do banco de análise. Isto permite que as matrizes possam ser fatoradas em matrizes menores e mais esparsas, gerando uma transformada rápida, e diminuindo o número de operações necessárias para a transformação *wavelet* ^[77]. Pode ser demonstrado que o número de multiplicações na transformada rápida fica limitado a no máximo $2 \cdot T \cdot N$, onde T é o número de coeficientes dos filtros ^[78]. Strang (1996) realiza uma conexão cuidadosa e aprofundada entre a teoria dos bancos de filtros e a teoria *wavelet* em ^[78].

Também examina com detalhes a questão da reconstrução e os métodos de geração dos filtros H , G , \bar{H} e \bar{G} , tais que constituam bancos (inversos) satisfazendo esta propriedade. Um caminho possível de projeto resulta na geração de bancos de filtros

QMF - ou filtros de espelhamento de quadratura. Outro caminho, utilizando filtros inversos alternados (*alternating flip*) conduz a bancos de filtros ortogonais.^[78]

Os filtros de Daubechies, por exemplo, seguem este padrão, que leva G a ser definido a partir de H como abaixo:

$$h(k) \text{ e } g(k) = (-1)^k h(N-1-k)$$

Nas análises experimentais de gestos musicais contidos em interpretações de flauta e violino - *mais adiante neste trabalho*- os filtros de Daubechies foram maciçamente empregados. O esquema de análise empregado pelo pacote computacional adotado baseia-se na implementação de uma análise *wavelet* em multiresolução via banco de filtros, como exposto acima. Os filtros de Daubechies D16, que conduzem a uma análise *wavelet* com a família Daubechies 16, foram os mais empregados nesta fase, exibindo os melhores resultados dentre as famílias avaliadas.

No capítulo seguinte aborda-se os timbres instrumentais: o conceito de timbre, os instrumentos selecionados para as análises deste trabalho, o paradigma de representação de timbres instrumentais em sistemas eletrônicos de síntese, e a importância da interpretação sobre a qualidade e naturalidade dos sons instrumentais.

3 TIMBRES DE INSTRUMENTOS ACÚSTICOS E MÚSICA INSTRUMENTAL

Dissertar sobre timbres musicais implica na premissa de prover antes uma definição para o termo “timbre”. Esta tem sido uma tarefa difícil já há um bom tempo.

A *American Standards Association*²⁷ em 1960 definiu timbre como “aquele atributo da sensação auditiva em termos do qual um ouvinte pode julgar que dois sons similarmente apresentados e tendo a mesma intensidade e *pitch* são dissimilares.” Esta definição entretanto deixa muito há desejar. Bregman (1990), comentando sobre a definição do termo, classifica a definição acima de inócua, e mesmo sugere uma “nova versão” para ela: “Nós não sabemos definir timbre, mas ele não é intensidade e não é *pitch*.”^[5]

Pitch e intensidade são parâmetros acústicos importantes em música, mas não transportam toda a informação necessária para permitir a classificação de timbres, ou servir de guia na tarefa de reconhecê-los. Numerosos fatores informam sobre a percepção do timbre: a amplitude do envelope do som, a forma e conteúdo do ataque, as ondulações periódicas devido a um vibrato ou um trêmulo, a estrutura dos formantes, a amplitude percebida, duração, e o espectro variante no tempo.^[63]

Duração e amplitude exercem uma importante influência sobre o timbre percebido de um instrumento. Por exemplo, diferenças na intensidade podem transformar um tom de flauta, que a 60 dB equivalia a uma nota, num “barulho intenso” quando em 120 dB. Um tom que dure 30 ms pode possuir a mesma forma de onda periódica de um tom que dure 30 segundos, mas para um ouvinte pode ser muito difícil determinar se tais tons provém da mesma fonte sonora.^[63]

Timbre é portanto uma entidade de complicada definição. Certamente ele depende até certo ponto do espectro do som, mas vai além disso. Por exemplo, é possível reconhecer diferentes instrumentos tocados através de um radinho de pilha monoaural barato e de baixa fidelidade, o qual distorce seriamente e corta o espectro original dos sons. Também o fato de aumentarmos o volume do som neste radinho não alterar nossa percepção da dinâmica do som original (se é um *pianíssimo* ou um *fortíssimo*) constitui outra confirmação sobre a complexidade da percepção do timbre.^[56]

²⁷ Atualmente ANSI - *American National Standards Institute*.

A classificação de timbres é historicamente mais antiga que a tentativa de se atribuir uma definição clara para o termo (na esperança de que isso sistematizaria a categorização dos sons musicais). Os chineses há séculos faziam classificações dos timbres baseado numa taxonomia de suas fontes sonoras. Desenvolveram sofisticadas descrições dos timbres e levaram em conta a influência dos diferentes tipos de “toques” (formas de ataques, puxadas e vibratos) na conformação do timbre. ^[63]

O espectro e o timbre são conceitos relacionados, mas não equivalentes ^[63]. O espectro refere-se às propriedades físicas (a distribuição da energia em função da frequência) dos sons musicais. As ferramentas computacionais de análise baseadas no espectro tradicional encontram limites em tarefas de alto nível em musicologia, como análise de estilos, percepção e identificação de características em interpretações musicais.

A análise destes elementos, transportados nos timbres instrumentais, encontra alicerce em grande parte nos paradigmas da psicoacústica. Psicólogos usam o termo *timbre* para se referir às qualidades percebidas e aos mecanismos perceptuais que permitem classificar os sons em famílias. A psicoacústica dá mais ênfase à descrição dos timbres em termos de suas qualidades perceptuais, e nos remete a considerar fenômenos como a *integração seqüencial*, *fusão* e *mascaramento*. Neste contexto, Roads (1996) considera certamente mais fácil discutir-se *timbre* dentro do domínio dos tons instrumentais tradicionais, onde quase toda a pesquisa passada estava focada ^[63]. Bregman (1990) também sugere que a classificação de timbres, referenciando-os pelo rótulo do instrumento que os produzem, é perfeitamente satisfatória ^[5]. Existem várias técnicas criadas no intuito de se permitir classificar o universo de timbres por categorias ou famílias. ^[56; 63]

Na música ocidental tradicional o paradigma de classificação baseado em famílias instrumentais é amplamente empregado. Os instrumentos acústicos tradicionais são classificados como pertencentes a quatro famílias básicas, tomando-se os materiais de construção dos instrumentos como parâmetro categorizador: os **metais** (trompetes, tubas, trombones, trompas, etc.), as **madeiras** (flautas, oboés, fagotes, clarinetas, saxofones, etc.), as **cordas** (violinos, violas, violoncelos, violões, guitarras, baixos, etc.), e as **percussões** (tímpanos, tambores, taróis, tablas, pratos, tamborins, tantãs, gamelões, etc.).

Esta classificação, no entanto, não reflete muitos dos avanços tecnológicos em música neste século, bem como não atendem a todos os timbres existentes (ou sintetizáveis eletronicamente) permanecendo hoje válida como herança da tradição musical de séculos passados. Instrumentos como a harpa, o acordeão ou o complexo piano, por exemplo, não encontram lugar nesta classificação, bem como as modernas flautas transversais, na sua maioria feitas de ligas metálicas, não se enquadram bem sob o termo “madeiras”.²⁸

No domínio da música eletrônica, entretanto, os timbres precisam de um sistema de descrição mais exato, com parâmetros definidos de tal forma que possam servir para controlar um aparato sintetizador e dirigir a síntese de um sinal acústico. Esta necessidade impulsionou, no decorrer da segunda metade deste século, o desenvolvimento de um número de modelos representativos para timbres musicais enquadrados numa estrutura de referência matemática e de engenharia, isto é, modelos expressos em linguagens matemáticas e de engenharia, e que pudessem ser implementados através de algoritmos ou fluxogramas de controle.

Os primeiros sintetizadores analógicos, e as primeiras versões digitais que lhes sucederam, baseavam-se em modelos descritivos de timbres simples e determinísticos. A inexistência de dispositivos de alto poder computacional no início, somado à imaturidade dos sistemas e teorias explanativas sobre os fenômenos musicais, limitaram bastante os primeiros passos em direção à síntese eletrônica de música: os resultados sônicos eram pobres e inexpressivos, reflexos claros do emprego de modelos reducionistas e super simplificados para representar complexos fenômenos acústicos. Não fosse a sintomática vontade humana da persistência não teríamos chegado aos atuais sistemas musicais computadorizados de hoje.

Neste capítulo apresentam-se os instrumentos acústicos cujos timbres foram escolhidos como alvo de análise: a flauta e o violino. Faz-se então uma breve introdução aos modelos elementares de timbres instrumentais nos sistemas de síntese eletrônicos. Em seguida adiciona-se uma discussão sobre a influência da interpretação na música instrumental: o papel preponderante dos gestos musicais realizados pelo intérprete, que constituem o diferencial não alcançado pelos sintetizadores eletrônicos no sentido de se

²⁸ O mesmo ocorre com o saxofone, instrumento inventado por volta de 1840, também classificado em Orquestração na família das madeiras, apesar de ser feito inteiramente de metal (com exceção do bucal onde fica a palheta).

produzir uma música rica em detalhes, natural na evolução e expressiva no contexto psicológico. Para completar o capítulo, introduz-se a técnica da Análise e Síntese, ferramenta já tradicional em computação musical, com a qual procura-se compreender melhor a estrutura de formação dos sons musicais.

3.1 Instrumentos acústicos tradicionais

SELEÇÃO DOS TIMBRES INSTRUMENTAIS

Pretende-se abordar neste trabalho os timbres de dois instrumentos acústicos sinfônicos para gerar as amostras para análise: um de sopro, a flauta transversa, e um de cordas, o violino.

As formas de excitação, de controle da vibração e emissão sonora são distintas em cada instrumento, embora haja uma semelhança nos princípios físicos de geração sonora, fundamentados pelas teorias da física clássica de ondas e propagação de ondas mecânicas em meios deformáveis ou elásticos. Por pertencerem a famílias de instrumentos diferentes, seus timbres são notadamente diferentes, possuindo registros e “colorações tonais” distintos.

3.1.1 A flauta

A flauta em verdade pertence à família dos instrumentos de sopros. Vários autores, seguindo a nomenclatura tradicional, se referem à flauta como pertencente à família das madeiras, numa alusão ao material de que eram construídas no passado, mantendo-se a nomenclatura mesmo nos dias atuais quando há flautas feitas de diversos materiais, as mais nobres construídas de metais, o que lhes proporciona um registro sônico mais amplo e um timbre mais nítido e brilhante²⁹. A flauta caracteriza-se por ser um instrumento de timbre essencialmente harmônico, de qualidade clara, capaz de uma

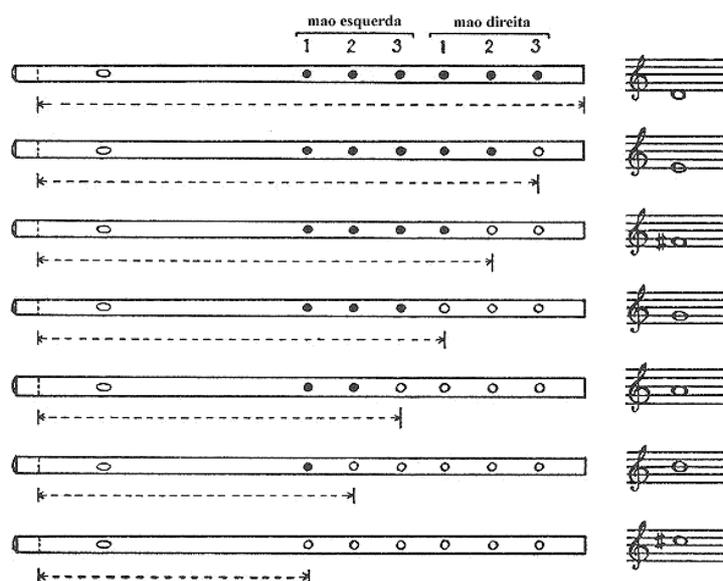
²⁹ Existem flautas construídas da mais variada gama de materiais, incluindo prata, ouro, platina, madeiras, e até vidros e materiais cerâmicos. A extensão da influência do material sobre a qualidade do timbre é um campo ainda inconclusivo, embora na prática, por exemplo, hajam suficientes indicativos de que bucais de prata ou platina produzem tons mais claros e límpidos. Por outro lado, o artífice, o projeto, a perícia na manufatura e o acabamento, estes indubitavelmente são preponderantes sobre a qualidade final do timbre do instrumento.

extrema flexibilidade de variações tonais e velocidade nas articulações. O mecanismo físico de geração da escala (tonal) baseia-se na variação do comprimento da coluna de ar vibrante em intervalos discretos, dando origem a uma sequência de modos de vibração: os tons e semitons, que exibem uma série harmônica e um *pitch*. Em síntese, a flauta (transversa) comporta-se como um tubo vibrante, de uma extremidade fechada e a outra aberta, onde o comprimento vibrante é variável pelo arranjo do dedilhado.

Através do sopro induz-se vibrações simpáticas por ressonância sintonizada na frequência (*pitch*) da coluna de ar. A altura da frequência (e o *pitch*) é função do comprimento da coluna de ar vibrante. Furos devidamente dispostos no tubo da flauta em intervalos regulares servem para provocar o efeito da variação do comprimento da coluna de ar em intervalos também regulares. A produção da escala é obtida pelo posicionamento dos dedos sobre os furos, fechando-os ou abrindo-os em arranjos determinados (posições ou dedilhados).

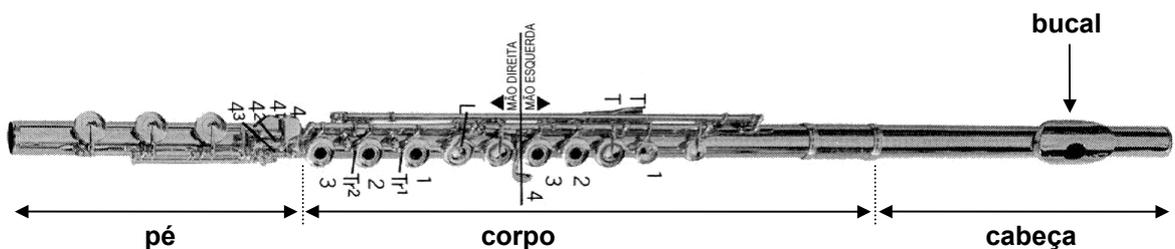
A Figura 14 mostra uma flauta (transversal) básica (didática) e alguns dos tons produzidos para algumas posições de dedilhado. O dedo indicador é o número 1 para ambas as mãos, o dedo médio, o 2, e assim por diante ^[57]. A indicação das notas correspondentes aos dedilhados é feita nas partituras à direita, em clave de sol (registro típico para flautas). Os segmentos pontilhados logo abaixo de cada um dos 7 arranjos mostrados mostram o comprimento (efetivo) da coluna de ar vibrante para cada nota/dedilhado.

Figura 14 - Flauta básica: dedilhado e respectivos tons produzidos



Nas flautas modernas foram adicionadas chaves especiais para facilitarem o acesso a disposições de dedilhados mais difíceis, o que contribuiu muito para melhorar a interface do instrumento para o manuseio do instrumentista, conseqüentemente oferecendo-lhes mais liberdade e velocidade na interpretação [57]. A Figura 15, abaixo, mostra uma flauta transversal moderna, um modelo com as chaves vazadas e com a nota mais baixa acessível em si bemol (Bb2 na afinação oficial de referência). O instrumento cobre quatro oitavas, e existem posições de dedilhados alternativas que produzem a mesma nota³⁰ (dedilhados *isotonais*) como recurso para incrementar a agilidade nos fraseados que normalmente exigiriam mudanças rápidas de posições envolvendo muitos dedos.

Figura 15 - Flauta transversal metálica, afinada em dó, de 3 partes montáveis e chaves vazadas



A excitação na flauta transversal é feita através do sopro transversal ao bucal do instrumento. O controle da entonação na flauta é manejado na embocadura pelos lábios e pela regularidade do fluxo de sopro. Existem amplas possibilidades de se impor efeitos especiais e modulações simplesmente alterando-se a pressão do sopro, a disposição dos lábios sobre o bucal, o formato da abertura labial e o ângulo do sopro em direção à fenda do bucal.

A embocadura refere-se à formação dos lábios e músculos vizinhos, de forma a produzir um tom em instrumentos de sopro. A disposição dos lábios, maxilar, queixo e garganta é extremamente importante na formação de uma embocadura correta para guiar o sopro para a flauta transversal.

A língua é responsável pela textura e articulação do ataque. Existem vários ataques possíveis para uma nota, dependendo da forma como o instrumentista inicia o fluxo de

³⁰ As diferenças em *pitch* estando numa escala microtonal, perfeitamente ajustável pelo intérprete modulando-se a embocadura e a pressão de sopro. Vale lembrar que a entonação é também um processo pessoal de afinação, que admite variações e estilos diferentes entre instrumentistas diferentes.

ar projetado no bucal. Os mais comuns são obtidos liberando-se o fluxo de ar como se pronunciando a sílaba “*tu*”, ou “*ku*”.

A continuidade ou ritmo do fluxo de ar depende do controle sobre a respiração, em grande parte manejada pelo diafragma, um músculo vigoroso da região abdominal. A intensidade (dinâmica) é função do volume de ar soprado, e sua regularidade, especialmente em notas longas, também vai depender do controle que o instrumentista tem sobre sua respiração.

As posições na escala tonal são determinadas pelo dedilhado e pela embocadura, que também influencia na determinação da altura da nota (*pitch*). Por exemplo, é possível tocar-se um lá⁴ (A₄) ou um mi⁵ (E₅) com a mesma posição do dedilhado, simplesmente alterando-se a embocadura e o fluxo de ar no bucal.

A flauta transversal utilizada para amostrar os eventos musicais abordados neste trabalho foi uma **Armstrong** modelo 104, de fabricação americana, de chaves fechadas e afinada em dó.

3.1.2 O violino

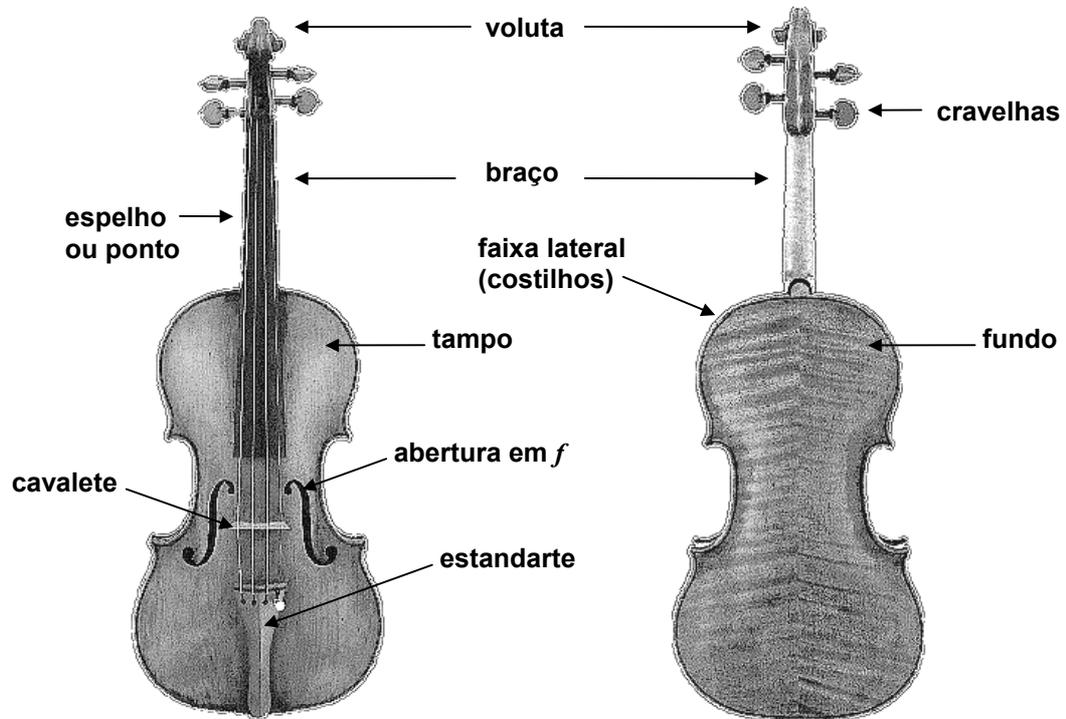
O violino pertence à família das cordas, e caracteriza-se por ser um instrumento de timbre denso, de grande alcance dinâmico, e imensa capacidade de expressividade e possibilidade de variação contínua na tonalidade.

Não é muito claro quem inventou o violino. Pode ter sido Andrea Amati, que fundou uma importante escola de *luthiers* de violinos, em Cremona, Itália. Amati morreu em 1580, e cerca de 150 anos depois seus “pupilos” desenvolveram a arte da fabricação de violinos a um extraordinário alto nível, particularmente Antonio Stradivari e Giuseppe Guarneri. Naquela época, o pouco conhecimento da física do som não influenciou sobremaneira o desenvolvimento do instrumento. Hoje, entretanto, a desenvolvida ciência da acústica é aplicada abertamente para compreender o violino e seu processo de fabricação, bem como contribuir para o seu aperfeiçoamento.

Em essência, o violino é um conjunto de quatro cordas montadas sobre uma caixa de madeira compreendendo um espaço de ar quase fechado. Alguma energia das vibrações induzidas nas cordas é transmitida à caixa e ao espaço de ar incluso no qual são estabelecidas vibrações correspondentes.

A Figura 16 abaixo mostra a anatomia de um violino, indicando suas partes componentes.

Figura 16 - Anatomia de um violino: partes constituintes principais



A 1ª corda (primeira à direita na figura acima) é a corda afinada em mi: a mais aguda. A 2ª é a corda em lá, a 3ª, em ré, e a 4ª (última à esquerda, também chamada bordão), em sol, a mais grave ^[73]. As cordas tradicionalmente são feitas de uma alma (que pode ser de tripa de porco ou *perlon*) envolvida por finos fios trançados de prata ou alumínio.

Cordas vibrantes têm sido estudadas desde a época de Pitágoras. No século XIX, O físico alemão Hermann von Helmholtz muito contribuiu ao tema elucidando os tipos de vibrações que distinguem a corda puxada (*pizzicato*) da tocada pelo arco. Seu trabalho foi mais tarde complementado e enriquecido por muitos outros pesquisadores.

A corda vibrante por si só não possui área de superfície suficiente para imprimir uma pressão sonora apreciável, e sozinha, sem amplificação, soaria desprezivelmente. Cerca de 10% da energia suprida pelo violinista é comunicada aos corpos ressonantes de madeira por meio de um mecanismo de movimentação do cavalete. Excitada pelo arco, as vibrações na corda encampam dezenas de harmônicos energéticos, e embora seu movimento possa parecer simples³¹ nenhuma solução geral foi firmada para ele, por

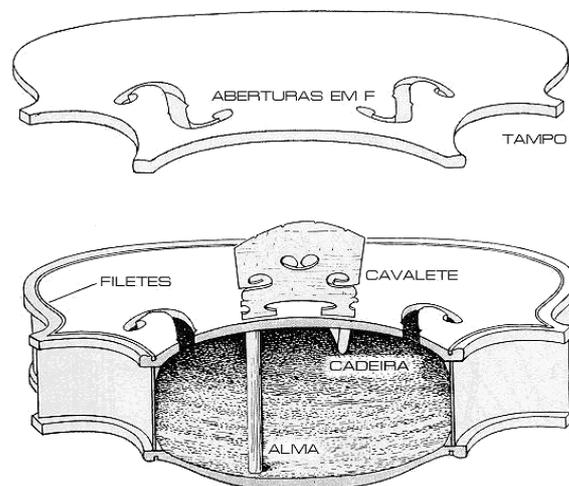
³¹ muitas vezes considerado simplesmente um deslocamento em forma de onda triangular ou dente de serra.

causa da extrema não linearidade existente na fricção, e pelo complexo relacionamento desta com a velocidade de arraste do arco.

A caixa sonora vibrante é formada por um tampo - *construído de abeto laminado no sentido do comprimento da tora, arqueado para fora e vazado por 2 furos em forma de f* - por um fundo - *também arqueado para fora, normalmente esculpido de um bloco de ácer curado por vários anos, cuja espessura varia de 6 mm no centro a 2 mm nas bordas* - e por finas faixas laterais de suporte, de espessura próxima de 1mm, construídas de ácer crespo, envergadas e coladas a blocos de abetos ou madeira de salgueiros. O braço, de ácer crespo, é a base para o espelho, este normalmente de ébano, onde o violinista dedilha. O cavalete, de ácer maciço, é um elemento fundamental para a formação do som no violino.

Na parte interior da caixa existem ainda dois elementos de importantes funções acústicas, exercendo relevante influência sobre a qualidade do tom: a **alma** (*sound post* na língua inglesa; *âme*, em francês), um cilindro de madeira encaixado (aproximadamente) sob um dos pés do cavalete, e sem o qual o violino passa a soar mais parecido a um violão, e a **cadeira** ^[73] ou **barra harmônica** (*bass bar*), uma faixa de madeira sob o outro pé do cavalete que se estende sob boa parte do comprimento do tampo. "Ambas as estruturas, além de darem mais solidez ao tampo, melhoram o som: a barra harmônica, o das notas graves; a alma, o das agudas". ^[73] A Figura 17, abaixo, ilustra um corte seccional do violino mostrando a localização destes componentes internos à caixa.

Figura 17 - Corte seccional do violino sob o cavalete.



O arco, ferramenta (imprescindível) à parte, possui o lenho de madeira e o "cabelo" tradicionalmente formado de crina de cavalo. Outros ingredientes, como colas, vernizes e enchimentos, também exercem sua parcela contribuinte sobre a qualidade do timbre.

Itokawa e Kumagai (1952) lembram que uma fonte sonora usualmente possui três elementos necessários: um *sistema vibratório*, uma *parte comunicativa* (ou transmissiva) que geralmente inclui sistemas filtrantes e ressonantes, e uma terceira parte: o *radiador*. No caso do violino as cordas fazem o papel da primeira parte, o cavalete age transmitindo e filtrando as vibrações, e a terceira parte cabe à caixa, ao tampo, fundo e o volume de ar contido no corpo do violino, que consiste num ressonador de Helmholtz juntamente com as aberturas em f .^[30, p.55]

As aberturas em f , de cada lado do instrumento, exercem duas funções acústicas fundamentais: **(1)** reduzir a rigidez da base do tampo, onde o cavalete se ergue, e **(2)** formar uma ressonador de Helmholtz^[30]. Elas não consistem simplesmente em saídas para o som irradiar-se, mas juntamente com o as paredes do corpo do violino formam um vibrador harmônico ou cavidade ressonante, que ressona em faixas graves de frequência.

A extensa faixa de frequências comunicada pelas cordas à caixa forçam-na a vibrar em sintonia. A estrutura do corpo entretanto possui seu próprio conjunto de frequências ressonantes, e transferência expressiva de energia ocorre na coincidência dos harmônicos ressonantes. O *luthier* de violinos está principalmente interessado na principal frequência de ressonância do volume de ar (PFA-Principal Frequência do Ar), que é a frequência do ressonador de Helmholtz. Quanto maior o volume de ar fechado, menor será a frequência de ressonância, e quanto maior for a área das aberturas em f , maior será esta frequência.

A excitação do instrumento normalmente ocorre pelo arraste de um arco sobre as cordas presas no corpo do instrumento. A afinação das tonalidades fundamentais das cordas é determinada pela tensão a que estão submetidas, embora a tonalidade possa ser continuamente variada pelo instrumentista alterando o comprimento vibrante das cordas com os dedos.

Existe uma supremacia dos instrumentos de cordas sobre os outros em orquestras sinfônicas, assim como uma "discreta preferência" pelos compositores (eruditos) por eles. A "coloração tonal" do grupo das cordas é equitativamente homogênea de cima a

baixo, enquanto os instrumentos de sopro apresentam características sonoras individuais muito mais pronunciadas. ^[57]

As possibilidades de expressão dinâmica destes instrumentos, bem como os detalhes sobre suas respectivas técnicas de interpretação são assuntos cuja abordagem extrapola o objetivo deste texto. O leitor interessado pode encontrar maiores detalhes em ^[57]. Também dúvidas mais acirradas e um interesse maior pela física do instrumento podem ser satisfeitas, ainda que inicialmente, pela compilação dos mais relevantes trabalhos abordando a acústica do violino, a acústica da corda arqueada, os materiais de fabricação e propriedades, e o papel de importantes elementos do instrumento (como o cavalete e a alma) em ^[30].

O instrumento utilizado para capturar os eventos musicais abordados neste trabalho foi um violino Hopf, de 1730, com a corda sol revestida de prata, o interior de perlon, as demais cordas revestidas de alumínio.

3.2 Dos timbres à música instrumental

O som de uma sinfonia para orquestra é composto pela somatória dos tons individuais de cada instrumento, que são emitidos segundo uma sequência temporal organizada, obedecendo a um padrão rítmico, harmônico e melódico. O tom musical de um instrumento é formado por padrões oscilatórios periódicos e aperiódicos atuando em diferentes faixas de frequência e intervalos de tempo durante o período de sua duração. Timbres de instrumentos acústicos variam na qualidade percebida do início ao fim do tom, isto é, o timbre evolui no curso de sua duração. Tons assim produzidos são ditos possuir espectro variante no tempo. ^[54, p.292]

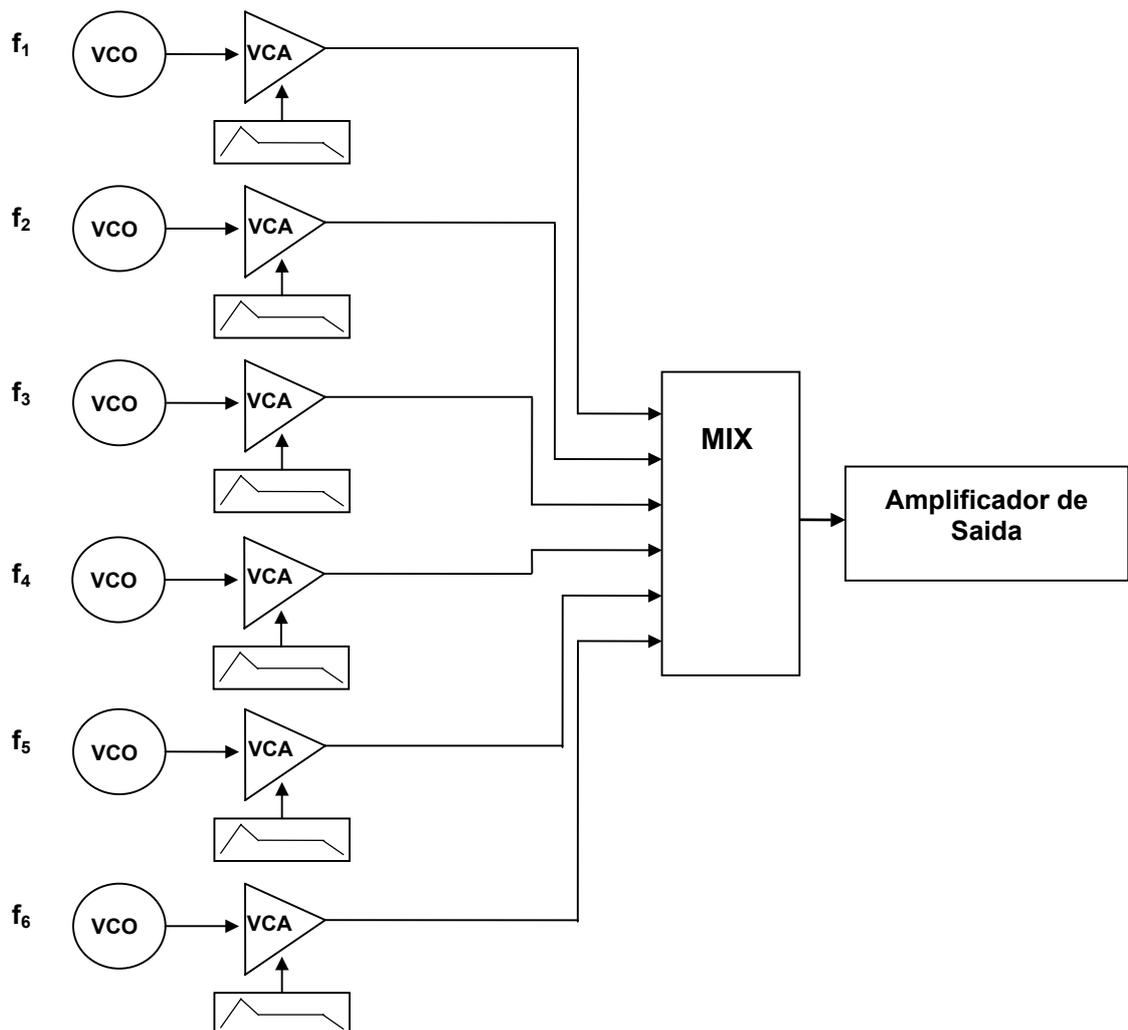
3.2.1 Composição de timbres: modelos elementares

A compreensão dos elementos constituintes do tom é essencial para construir um modelo realista de um instrumento musical e seu timbre para objetivo de síntese. Em particular, será necessário aqui introduzir e definir as regiões comumente presentes em tons musicais que serão objetos de análise e discussão nos capítulos seguintes.

Os antigos sintetizadores analógicos baseavam-se num conceito simples de geração sonora, procurando imitar o modo natural de evolução das frequências parciais componentes do tom. Neles, geradores de frequências (VCOs) geravam a fundamental e parciais do tom. A saída de cada um era então modulada em amplitude (através de VCAs) de forma que cada frequência tivesse um envelope único, e finalmente misturadas para compor o tom final, consistindo-se portanto numa forma de síntese aditiva.

A Figura 18 abaixo mostra o esquema de roteamento dos sinais entre estes módulos, ilustrando um sintetizador analógico elementar, realizando uma síntese aditiva de um tom com 6 frequências parciais (f_1 a f_6 na saída dos VCO's) devidamente "envelopadas" por VCA's ^[54, p.241]. Evidentemente, com VCO's produzindo formas de onda mais complexas as possibilidades de síntese se alargam.

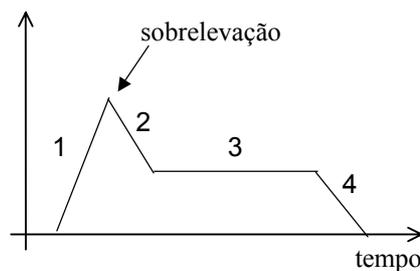
Figura 18 - Esquema de um sintetizador analógico modular básico



O módulo do gerador de envelope tipicamente esculpia um contorno de 4 estágios para o sinal: uma etapa de ataque (*attack*), um queda inicial (*decay*), um período de sustentação (*sustain*) e a queda final (*release*), e é frequentemente conhecido pela sigla (inglesa) ADSR ^[54]. Este modelo descritivo de tons naturais, que busca imitar o envelope de tons naturais, sofreu variações no decorrer das gerações de sintetizadores, e foi naturalmente portado para os sistemas digitais. Os envelopes temporais são ainda largamente utilizados nos sintetizadores digitais comerciais atuais, diferenças e sofisticacões na arquitetura de implementação variando de fabricante para fabricante.

A Figura 19 abaixo ilustra o envelope temporal e seus estágios ou regiões básicas descritas sob um ponto de vista físico.

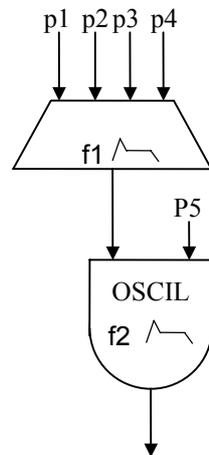
Figura 19 - Envelope de amplitude x tempo, ou ADSR



- **ataque (1)** - a **primeira região de subida** do tom até um pico determinado (sobrelevação); refere-se ao período transitório de excitação quando vibrações em várias faixas de frequência são estabelecidas e um determinado padrão harmônico, solicitado.
- **decaimento (2)** - o período que segue do pico do ataque (sobrelevação) à etapa de regime permanente em que as vibrações convergem para modos estacionários.
- **sustentação (3)** - o período referente ao estado de **regime permanente**, onde as vibrações são sustentadas pela duração (e sob efeito das modulações) que se desejar impor ao tom/timbre.
- **liberação (release) (4)** - o período de **colapso**, que se inicia ao fim da excitação e corresponde ao desvanecimento do tom, associado ao desmonte dos padrões vibratórios, até o seu completo término.

A Figura 20 mostra um fluxograma de um algoritmo simples descritivo de um instrumento gerado em sistema digital ^[54]. Uma associação modular semelhante à do diagrama do sintetizador analógico da Figura 18 é igualmente aplicável neste caso.

Figura 20 - Fluxograma de um simples algoritmo descritor de instrumento em síntese digital



Parâmetros de Controle:

- p1 - Tempo de ataque
- p2 - Tempo de decaimento
- p3 - Duração da nota
- p4 - Amplitude

- p5 - Frequência para o oscilador

- f1 - Envelope para o oscilador
- f2 - Nota/tom final

Em verdade, nas interpretações acústicas reais, o número de etapas com identidades próprias e diferenciadas pode ser bem maior. Apesar da abordagem generalista e simplificadora deste modelo, verifica-se que a qualidade do som gerado sob este paradigma depende diretamente do esquema de implementação dos estágios do envelope, motivo pelo qual encontra-se no mercado sintetizadores com qualidades drasticamente distintas, numa classificação que vai do decepcionante e monótono ao satisfatório e estimulante. Isto porque cada fabricante implementa uma arquitetura diferente no arranjo de unidades geradoras, filtros digitais e/ou analógicos, bancos de formas de onda (amostras de timbres naturais em memória), interfaces e fluxo de controle e dados, otimizando alguns aspectos.

Nos sintetizadores com mais recursos as etapas do envelope podem ser configuradas individualmente (editadas), criando texturas e promovendo efeitos sutis ou drásticos sobre o som produzido. O custo é um fator influente na complexidade final do sistema, e costuma crescer em proporcionalidade direta com a qualidade sonora do sintetizador.

3.2.2 A técnica de interpretação como fonte de expressividade

Existem várias técnicas de síntese empregadas para gerar tons tão próximos quanto possível de tons instrumentais naturais, conforme já abordado neste texto. Nos modelos de síntese mais simplistas, mesmo reducionistas, verifica-se a geração de tons e timbres claramente artificiais, que não raro excluem informações necessárias para que sejam reconhecidos, associados ou referenciados a um(s) instrumento(s) em particular, devido a inexpressividade incomparável aos tons naturais extraíveis de um instrumento real.

Nos métodos mais sofisticados, como nos de modelamento físico e espectral, adotam-se algoritmos de síntese mais elaborados, que levam em conta mais parâmetros de controle associados à interpretação do instrumento e também os fenômenos físicos acústicos ocorridos no instrumento durante a emissão dos tons. Conseqüentemente, consegue-se um melhor mapeamento entre a reprodução sintética e os processos associados à interpretação.

Todavia, mesmo os métodos de síntese mais complexos, disponíveis em alguns sintetizadores comerciais, não oferecem graus de liberdade e flexibilidade suficientes para o compositor ou músico utilizar todo o seu potencial criativo ou técnica interpretativa. O registro disso está no baixo grau de realismo e inovação, e na pouca expressividade e variabilidade dinâmica que conseguem impor aos tons que produzem.

Tais qualidades podem ser encontradas nas interpretações musicais ao vivo, e estão ocultas codificadas em escalas diferentes no material sonoro. As inúmeras possibilidades de controle do instrumento numa execução real, a infinidade de parâmetros físicos envolvidos na excitação e vibração dos instrumentos acústicos, e mais a impressão de elementos psicológicos e de expressividade do(s) intérprete(s) transformam a experiência musical num fenômeno complexo cujo modelamento passa pelo mapeamento dos eventos musicais associados à dinâmica, expressividade, realismo e técnica de interpretação em eventos de baixo nível, registrados como padrões sônicos em escalas diferentes.^[20]

É evidente que o exame ou interpretação de fatores subjetivos como “expressividade” e “realismo” são feitos à luz da descrição e da representação que a musicologia e a orquestração em música fazem destes elementos. Mesmo sendo de interpretação flexível, de uma correspondência não bijetora num sentido estritamente matemático, estes elementos se manifestam através de eventos musicais modeláveis e rastreáveis com o uso de ferramentas de análise adequadas para representar processos estacionários e transitórios, localizados no tempo e na frequência, como são as *wavelets*.

3.3 A análise e síntese de timbres musicais

Em 1946/47, o físico Dennis Gabor, trabalhando numa teoria da audição, sugeriu que os sons são formados por partículas elementares, os *quanta* acústicos, numa forma de representação quântica do som.

Em 1971, Iannis Xenakis, em seu livro *Formalized Music*, introduziu uma teoria de composição a partir de *grãos* de sons, aproximando-se do modelo de Gabor, e sugerindo o uso de computadores para gerar estes grãos ^[90]. Estava nascida a *síntese granular*, uma técnica de síntese aditiva de sons complexos a partir de grãos elementares (formas de ondas simples), que foi pesquisada e empregada por Roads (1978; 1985), Jones & Park (1988) e Truax (1988), entre outros ^[62]. Uma desvantagem reconhecida está no seu difícil controle, dado a imensa quantidade de dados requerida para efetivar a síntese (Truax, 1988).

As técnicas de análise e síntese estão entre as mais festejadas e antigas técnicas de síntese empregadas em computação musical. Em 1969 Risset e Mathews realizaram estudos para desvendar aspectos intrigantes dos timbres de trompetes e violinos que os tornavam inimitáveis (Risset e Mathews, 1969). Estudaram a sua estrutura espectral, determinando sua composição frequencial e sua evolução no tempo. O que mais os intrigava era a “relutância” que esses timbres em particular tinham em serem imitados por processos de síntese artificial. Segundo Risset, “*os sons instrumentais foram analisados, e a relevância destas análises foi checada pela síntese: se elas retêm os parâmetros 'auditivamente' significativos, elas devem permitir uma imitação razoavelmente boa do som analisado*”.^[61] Desta forma podiam verificar a relevância dos dados da análise bem como validar o modelo de síntese empregado. Em 1982 Risset e Wessel foram mais a fundo, explorando timbres pela análise e síntese (Risset e Wessel, 1978).

A análise e síntese, em poucas palavras, consiste num método para se extrair informações parametrizáveis do sinal musical existentes no domínio do tempo (análise), e utilizá-las como parâmetros controladores em um processo inverso (síntese), através do qual se reproduz o sinal original. Quando se usa as técnicas de Fourier na análise, os parâmetros extraídos revelam o conteúdo espectral do sinal, e portanto estão descritos no domínio da frequência.

Risset foi um dos pioneiros na utilização de métodos de análise pela síntese em computação musical. Sua tese em Física abordou a análise do timbre do trompete pela técnica da análise e síntese, que Mathews considera a técnica mais poderosa para analisar sons musicais naturais. ^[64, p.8]

As técnicas de análise e síntese frequentemente levam a representações reveladoras sobre a estrutura dos sons. Utilizando-a no estudo do trompete, Risset pôde verificar, por exemplo, que o aumento na intensidade (*loudness*) do som implicava num alargamento do espectro do timbre ^[61], enriquecendo-o. Também verificou que os parciais harmônicos mais altos apareciam mais tarde no ataque, depois dos parciais baixos, e também terminavam antes que estes. Embora variações randômicas da amplitude dos parciais provassem de nenhuma importância para o ouvido, variações randômicas das frequências destes parciais foram importantes na síntese de sons realistas de trompete. Risset também demonstrou que variações muito curtas nas amplitudes de vários parciais, bem como as rajadas curtas de ruídos (freqüentes no início dos ataques de trompetes) não representavam muito para o ouvido em termos de reconhecimento do timbre. ^[56]

A análise normalmente vale-se de um algoritmo matemático que permita mapear os aspectos estruturais do som no domínio do tempo em componentes descritas no domínio da frequência. Em linguagem matemática: obter uma representação espectral de uma função descrita no domínio do tempo.

A Transformada Enjanelada de Fourier (TEF) e a FFT são técnicas comumente empregadas em análises de séries e funções temporais, através da expansão da função em termos de blocos construtores básicos (funções base), no caso senóides. Rioul e Vetterli (1991) revêem a transformada enjanelada de Fourier - que chamam de "transformada de Fourier de tempo curto" (*STFT - Short Time Fourier Transform*) - e suas propriedades de análise sob resolução fixa ^[60]. A função original, num processo inverso à análise, pode ser aproximada por uma soma de contribuições de funções bases. Nisto consiste a síntese, conduzida pelos dados extraídos na análise (os coeficientes da transformada direta).

O principal problema nesta abordagem é que a análise pode gerar uma quantidade enorme de dados, uma "explosão de informação", tornando a síntese praticamente impossível em tempo real sem o emprego de arquiteturas computacionais de alto

desempenho. Evidentemente que existem formas de se agrupar alguns componentes, eliminar componentes não significativos perceptualmente (baseado em paradigmas da psicoacústica) e correlacionar vários componentes em modelos simplificadores, reduzindo assim a quantidade de informação necessária para uma síntese satisfatória, mas não há para isso uma receita padronizada de uso geral, aplicável em todos os casos. Evangelista (1993), por exemplo, utiliza um esquema para identificar e agrupar padrões periódicos em sinais acústicos de forma a reduzir a quantidade de informação necessária para representá-los ^[19]. Uma solução adequada, à parte de suposições sobre o modelo estrutural do sinal, está no emprego de técnicas de compressão de dados, ou na escolha de algoritmos naturalmente redutores da quantidade de informação, como a codificação sub-banda, os algoritmos piramidais, e os algoritmos *wavelets*.

A análise *wavelet* pode gerar um grande número de coeficientes nulos ou aproximadamente nulos, implicando numa redução considerável no número de dados gerados ^[77]. No caso de sinais discretos, a análise *wavelet* pode ser implementada através de esquemas de codificação sub-banda, com bancos de filtros, expandindo-se o sinal original num número de coeficientes igual ou menor ao número de amostras do sinal original. Muitos deles podem ser nulos, ou aproximadamente nulos, e por meio de técnicas de “limiarização” (*thresholding*) podem ser desconsiderados sem que isso represente perda de informação essencial.

Essa característica do processamento com *wavelets* atraiu o interesse de especialistas interessados em desenvolver técnicas para altas taxas de compressão de dados, admitindo-se uma pequena perda na informação (compressão com perdas) ^[26; 77; 88]. Entretanto, nas aplicações de caracterização e análise estrutural dos sinais a maior preocupação não é com a redução no conteúdo de informação, mas, ao contrário, com a garantia em se preservar qualquer informação que contribua para uma descrição mais precisa da estrutura interna do sinal, e que possa levar a uma melhor compreensão e ao modelamento dos fenômenos associados com a geração do sinal.

Numa análise *wavelet* os parâmetros extraídos também revelam o conteúdo espectral do sinal, porém sob uma roupagem mais elaborada, não podendo os coeficientes *wavelets* serem interpretados direta e simplesmente como parâmetros de frequência. Dado o trânsito da teoria *wavelet* por várias áreas da matemática, física e engenharia, é natural que possamos abordá-la sobre diferentes linguagens ou interpretações, escolhendo a

mais conveniente para enfatizar uma propriedade específica, ou explicar seu modo operante.

Sob um ponto de vista matemático, a transformada *wavelet* realiza um produto interno com o sinal, projetando sobre bases *wavelets* o sinal analisado. O produto interno dessa forma pode ser visto como uma medida de similaridade entre o sinal e as *wavelets*, expressa pelos coeficientes *wavelets* gerados.^[60]

Sob a perspectiva da codificação sub-banda, a transformada *wavelet* consiste num banco de filtros passabanda aplicados ao sinal, e cada canal dá origem a uma versão filtrada do sinal original, numa escala e nível de resolução específicos, representada pelos coeficientes *wavelets*. Ambas as interpretações são corretas, e clareiam aspectos diferentes da mesma técnica.

O processo de análise e síntese provê um *insight* na compreensão das identidades sonoras e frequentemente leva a representações reveladoras. Além disso, pode levar à redução da quantidade de informação (compressão de dados) e a um aspecto mais interessante: a capacidade de se expandir o vocabulário sônico, alterando aspectos do som enquanto preservando sua riqueza e densidade - como na técnica de **síntese cruzada** (Grey e Moorer, 1977) e na produção de ilusões e paradoxos sônicos (Risset 1978, 1985) -, bem como a capacidade de se produzir variantes e transformações íntimas no som original. (como por exemplo preservar o timbre e alterar o *pitch*, a duração e a articulação^[61], ou alterar a duração preservando-se o *pitch*³²). Vários processos de análise e síntese, baseados em técnicas de Fourier, foram desenvolvidos com estes propósitos, como o *Phase-Vocoder* (Portnoff, 1976; Moorer, 1978), abordado em^[17], e o *processo de predição linear* (Moorer, 1977; Lansky, 1981), mais conhecido pela sigla LPC (*Linear Prediction Coding*).³³

Os métodos de Fourier tornaram possível analisar digitalmente uma variedade de timbres de instrumentos musicais tradicionais, levando a uma melhor compreensão da natureza física e perceptual dos sons musicais, assim como à melhoria nas técnicas de

³² O termo "*pitch*" encerra uma conotação em música mais intrincada do que simplesmente a frequência de um tom. Uma definição mais adequada para este termo é abordada nos apêndices.

³³ O *Phase-Vocoder* perfaz uma análise espectral do sinal sônico, gerando funções de amplitude e frequência para cada harmônico do tom analisado. Em aplicações de voz e canto, permite a extração das frequências formantes^[64, p.460]. A análise com LPC permite extrair o *pitch* (que modela a excitação), e os coeficientes de um filtro recursivo (que modela as cavidades nasais e a garganta), gerando um modelo

síntese de som digital. Entretanto, como notado por Gabor em 1946, é paradoxal analisar-se um som limitado no tempo por meio de sons (as senóides) que duram eternamente, como ocorre na análise espectral aplicada à música: um número infinito de senóides infinitas (definidas de $-\infty$ a $+\infty$) são necessárias para cancelarem-se mutuamente nas regiões onde o sinal é zero.

Métodos que incorporem um senso de localização, tanto no domínio do tempo quanto no da frequência, com funções analisadoras compactamente representáveis num plano tempo-frequência, seriam mais adequados para estas tarefas. As *wavelets* vêm preencher, com vantagens, esta vaga, candidatando-se como funções bases de suporte compacto, no tempo e na frequência. A transformada *wavelet*, como será verificado neste trabalho, oferece meios eficazes para caracterizar o conteúdo espectral de sinais musicais, levando a representações compactas e reveladoras do sinal, e permitindo o acesso a níveis de detalhes e resoluções diferentes dentro do sinal, como uma sonda seletiva.

O próximo capítulo trata da análise em multiresolução de gestos musicais. Inicialmente compila-se um pequeno histórico de aplicações de *wavelets* em sinais acústicos para situar o leitor em relação a prévios trabalhos envolvendo *wavelets* e aplicações em áudio; em seguida introduz-se propriamente os **gestos musicais** - *eventos musicais com teor expressivo, alvos de análise do presente trabalho*. Aborda-se então os objetos sônicos a eles associados, a sua associação com níveis *wavelets* específicos (onde sua visualização torna-se mais evidente) e questões relevantes sobre a percepção de eventos musicais pelo sistema auditivo humano. A percepção auditiva torna-se significativa neste contexto, uma vez que a análise *wavelet* é similar à análise realizada pela membrana basilar da cóclea dos estímulos acústicos. As *wavelets* são então referendadas como ferramenta apropriada para análise de eventos musicais expressivos. Na sequência, introduz-se os eventos musicais expressivos que foram amostrados de interpretações de flauta e violino e que constituem o material básico de análise. A metodologia e forma de implementação das análises é então detalhadamente apresentada, finalizando o capítulo.

satisfatório do trato vocal, útil também principalmente em aplicações para voz ^[61; 64, p.16 e 436]. Ambos, entretanto, sofrem do mal da geração explosiva de dados na análise.

4 ANÁLISE EM MULTIRESOLUÇÃO DE GESTOS MUSICAIS COM *WAVELETS*

A idéia de se separar sinais, eventos sônicos de identidade própria, reconhecer e isolar componentes elementares formadores e outras classes de eventos constituintes de um material sonoro utilizando-se *wavelets* não é nova.

Evangelista (1993) introduziu uma nova representação *wavelet* baseada numa representação vetorial *pitch*-síncrona, permitindo identificar e separar os segmentos pseudo-periódicos e aperiódicos do sinal de voz ou música, e servindo à extração de padrões, à separação da voz do ruído em consonantes vocais, e à compressão de dados. Ele ilustra esta característica do processamento com *wavelets* decompondo a palavra “Susan” em níveis de resoluções diferentes, e mostra como é possível separar o ruído fricativo dos componentes vocais associados a consonante vocálica /z/. No mesmo exemplo, observa a concentração de informação a nível silábico num nível específico, e a concentração de componentes fricativos nas escalas menores. Num outro exemplo ele separa o ruído do arco das vibrações harmônicas do som de um violino, e relata outras experiências separando parciais representativas do ataque de um instrumento, e componentes inarmônicos.^[19]

A análise *wavelet* em multiresolução (AWMR) apresenta-se como uma ferramenta adequada para se analisar trechos musicais e isolar componentes constituintes periódicos e aperiódicos, como vozes, ataques instrumentais, efeitos expressivos, eventos dinâmicos, fraseados e modulações, separando-os pelos níveis da análise em multiresolução. Padrões ondulatórios distintos - *em cada nível da análise* - podem estar associados a eventos musicais específicos e que dispõem de notação em partitura, o que permite uma parametrização (ou mapeamento) dos objetos musicais notados em pauta em “objetos ondulatórios” ou padrões sônicos presentes nos níveis da análise.

4.1 Gestos Musicais, padrões sônicos e níveis *wavelets*

Pretende-se caracterizar neste trabalho padrões sônicos associados a gestos musicais expressivos executados em instrumentos acústicos tradicionais. Um bom músico ao executar uma peça musical faz uso de técnicas de interpretação que o permitem dominar o instrumento e produzir, a partir de atos gestuais executados no instrumento, gestos

musicais correspondentes, compreendendo todas as variações, efeitos e modulações exigidos para uma correta interpretação da sequência de notas da peça e necessários para que se possa transmitir os efeitos dramáticos e psicológicos associados, bem como estabelecer um senso de progressão, sentido de direção ou finalidade.

Quando os **eventos musicais** produzidos possuem uma qualidade expressiva relevante em música, os **gestos musicais** correspondentes são ditos **expressivos**. Na interpretação de um violino a manipulação das cordas pelos dedos e o arraste do arco sobre as cordas - ato controlado pelo punho e braço do violinista - contribuem com os principais gestos físicos que levarão à produção dos mais diversos gestos musicais possíveis para este instrumento.

Na flauta, o sopro e o dedilhado do instrumentista são os gestos físicos ativos que vão dirigir a produção dos gestos musicais. O teor de expressividade do gesto musical é um componente de difícil julgamento, mas está sem dúvida alguma diretamente correlacionado à técnica com que o instrumentista "manipula" o instrumento e explora sua capacidade em produzir eventos musicais.

Eventos musicais são construídos de estruturas sonoras menores, objetos sônicos ou sons individuais, partes constituintes que se agrupam numa unidade sonora clara e definida. Os objetos sônicos elementares podem estar descritos ou contribuir em faixas de frequência restritas e limitadas, e portanto serão melhor visualizados e identificados em certos níveis *wavelets* que cubram estas bandas de frequência. Os eventos musicais que se formam a partir deles, portanto, se distribuem ou concentram-se em certos níveis *wavelets* de uma análise em multiresolução, e estarão melhor caracterizados nestas escalas de resolução.

A PERCEÇÃO DE OBJETOS SÔNICOS

A identificação de uma identidade sonora nova, com propriedades claras e definidas, faz parte do trabalho de construir uma representação auditiva de um evento acústico ou objeto musical real. A construção de uma representação mental sobre eventos acústicos do mundo exterior implica na identificação e segregação dos estímulos acústicos que compartilham propriedades ou qualidades semelhantes, e que nos dizem sobre um mesmo evento musical. ^[5]

A forma como tais estímulos são agrupados pelo sistema nervoso guiará a elaboração de um novo padrão percebido. Surgem em decorrência deste agrupamento **propriedades emergentes** antes não atribuídas a nenhum dos estímulos ou objetos sônicos individualmente, e que agora passam a caracterizar o novo padrão que surge. Normalmente, em percepção, propriedades emergentes constituem-se em descrições precisas das propriedades dos objetos (acústicos) em nosso ambiente ^[5]. A correta associação, fusão ou agrupamento de tais objetos pelo sistema auditivo levará à correta representação (ou não) do evento ou objeto acústico real em curso no ambiente externo. O sistema auditivo, assim como o visual, possui um mecanismo sofisticado para elaborar sobre as relações entre objetos sonoros e a percepção de grupos, progressões, cursos ou sucessões de objetos sonoros (*streams*) com propriedades características. A medida deste agrupamento, do grau de relacionamento entre objetos sonoros e da percepção de grupos ou sucessões (*streams*) é objeto de estudo das linhas de pesquisa de Análise de Cena Auditiva. Bregman (1990) introduz e elabora sistematicamente sobre esta linha de pesquisa em ^[5].³⁴

A percepção portanto de uma unidade sonora clara e definida depende de como propriedades ou qualidades semelhantes dos sons ou objetos acústicos individuais vão se fundir ou agrupar numa única representação perceptual, num padrão percebido. Em música, percebemos uma evolução de padrões associados em diferentes escalas, construindo uma representação mais complexa, rica em conteúdo expressivo e dotada de um senso de direção, um objetivo. Contornos melódicos, fórmulas rítmicas, sequências de timbres, combinações harmônicas, efeitos, e uma infinidade de gestos musicais nascidos pelas técnicas de interpretação do(s) músico(s) são exemplos de elementos da Orquestração em Música analisáveis sob a óptica de padrões percebidos.

Belkin (1988), tendo visualizado esta correlação, sugere, pois, a construção de uma teoria mista de orquestração com percepção^[3]. Um grande número de padrões orquestrais organizados segundo uma sequência temporal sugerem a percepção de progressões e evoluções, ou proporcionam uma sensação de destaque, um contraste ou continuidade, e tecem assim uma trama expressiva, envolvente, criando uma

³⁴ Seu livro é um verdadeiro tratado sobre o tema, empregando argumentações de base da psicologia Gestalt e da psicoacústica, e indo mais além elaborando sobre questões como a integração sequencial de sons, integração simultânea de sons, integração baseada em esquema aprendido (*scheme-based*

representação (mental) emotiva. Por exemplo, sons curtos e percussivos podem ser usados para atrair a atenção ou refrescá-la após um movimento muito longo. Sons que evoluem, como se perfazendo uma contínua avaliação e trazendo um afluxo de informação, podem dar origem a um padrão de evolução. Sons que crescem ou diminuem em intensidade sugerem uma movimentação numa dada direção, uma progressão dinâmica. ^[3]

A INFLUÊNCIA DA INTERPRETAÇÃO SOBRE OS GESTOS MUSICAIS

O sucesso de uma peça musical dependerá completamente da habilidade do músico em usar plenamente suas técnicas de interpretação na execução de fraseados, efeitos, modulações e outros eventos musicais. Através delas, o músico pode fazer uso de ataques repentinos e intensos (*staccatos*, acentos), pode alterar a intensidade e entonação do som que produz (*crescendo*, *diminuindo*, efeitos de entonação e dinâmica), pode tocar sequências de tons continuamente (*legatos*, portamento, *glissando*), em pulsos ou saltos (*pizzicato*, acentos, *staccatos*), modulados em frequência (vibratos) ou alternar tons distintos rapidamente (trinados, trêmulos).

Esta é uma amostra das imensas possibilidades que o intérprete tem sobre a variabilidade na produção de tons e variações timbrísticas. A maior parte destas possibilidades sonoras acima podem ser representados em notação musical convencional, e constituem para efeitos práticos objetos ou eventos musicais, tratáveis à luz da musicologia e linguagens musicais associadas, bem como da "linguagem tradicional" em engenharia para representação de sinais (não determinísticos).

Dada a intrínseca composição hierárquica destes eventos musicais, construídos de objetos sônicos menores com propriedades frequenciais e delimitados em intervalos temporais finitos, a análise *wavelet* em multiresolução provê o arcabouço ideal para segregá-los e investigar as regras que regem sua construção. Faria (1996) aborda a análise de eventos musicais associados a técnicas de interpretação de instrumentos acústicos reais, e a emergência da qualidade de expressividade e realismo em padrões sônicos visíveis em diferentes níveis de resolução numa análise *wavelet*. ^[21]

integration: a influência da cultura e do conhecimento adquirido sobre a percepção auditiva), a organização auditiva na música e na percepção de voz.

4.2 Análise de gestos musicais em partes de flauta e violino

Inspirado pelos trabalhos em acústica já citados anteriormente, realizou-se no presente a implementação de uma análise *wavelet* em multiresolução (AWMR) para investigar a capacidade de segregação e a distribuição de eventos musicais - *em particular, gestos musicais expressivos associados a técnicas de interpretação tradicionais para flauta e violino* - em níveis/escalas de resolução diferentes gerados por este tipo de análise. A análise proposta para este trabalho não pretendeu o levantamento e a caracterização formal completa de um mapeamento deste tipo, objetivo este que solicita uma investigação mais complexa e abrangente, e possivelmente o emprego de técnicas complementares, como os pacotes de *wavelets* (*wavelet-packets*) e teorias de percepção cognitiva, sendo portanto alvo cabível de um trabalho futuro mais extenso.

4.2.1 Seleção do material sonoro para análise

SELEÇÃO DE EVENTOS MUSICAIS

Dadas as peculiaridades de cada instrumento, alguns eventos musicais se aplicam a um e não a outro, em função das técnicas de interpretação completamente diferentes que empregam. Todavia, uma grande classe de eventos discerníveis e executáveis em um podem ser executadas no outro, de forma diferente mas produzindo um efeito musical semelhante.

Existem praticamente uma infinidade de eventos, fraseados, modulações e efeitos produzíveis por estes instrumentos, entre os quais foram selecionados um grupo de grande importância expressiva e intenso emprego, tanto na música sinfônica quanto nos estilos contemporâneos e populares. Abordou-se os seguintes eventos musicais, discriminados na classificação abaixo:

Tabela 1 - Eventos musicais abordados

Eventos Musicais	Classificação	
Ataques (notas individuais)	Normais	
	Expressivos	com acentos
		com <i>staccatos</i>
		com <i>spiccatos</i>
Fraseados	Ligados (<i>legato</i>)	normais
		com acentos
		com trêmulos
	Não ligados (<i>non legato</i>)	normais
		com acentos
		com <i>staccatos</i>
		com trêmulos
Sustentação	com dinâmica	
	com vibrato	
	com frulato	
	com trinado	

Previamente às análises destes eventos, é necessário introduzi-los de forma a familiarizar o leitor com a terminologia, com sua notação convencional em música e, principalmente, com a descrição dos gestos musicais, sua fisiologia de produção pelo intérprete e o sentido musical que transmitem. É o que se faz a seguir.

Ataques (notas individuais)

Análise dos ataques de notas individuais emitidas fora de um contexto melódico, geralmente não associadas a uma sequência ou agrupamento de notas/tons definidos como ocorreria em uma progressão ou num tema musical. Duas classes de ataques foram abordadas: os **normais** e os **expressivos**.

Normais

Ataques normais são tocados em condições normais de interpretação, sem emprego de dinâmicas especiais, efeitos ou modulações.

Expressivos

Acentos São ataques realizados com gestos intensos. Leva-se em consideração a dinâmica da nota (intensidade). Acentos fazem mais sentido num

aspecto musical, dentro de uma hierarquia de valores (dinâmica) do que no aspecto técnico (da técnica de interpretação) por si só.

O fato de os ataques acentuados serem obtidos por meio de gestos usualmente súbitos e energéticos influenciam na conformação de outros aspectos e propriedades dos tons produzidos, como na forma da onda, na taxa de subida (taxa de variação de intensidade no tempo), na "coloração tonal" (expressa pela riqueza e variabilidade do espectro na região do ataque) e na qualidade do timbre por toda a duração da nota, uma vez que um padrão acentuado exercerá influência não só sobre o início da nota, mas também nos segmentos seguintes do envelope, como a região de sustentação e colapso.

Para o violino foram analisados acentos do tipo *sforzato*, uma técnica para execução de acentos em violino. Piston (1955) aborda com mais detalhes esta técnica em ^[57]. Para a flauta, usualmente o acento implica num ataque súbito, energético e constante, marcado por um adicional na pressão do sopro e por uma focalização de embocadura que permita que modos de vibração mais intensos sejam excitados no tubo vibrante.

Staccatos São ataques destacados e pontuais. Fazem sentido como elemento de articulação no contexto da técnica de interpretação, influenciando no destaque e na pronúncia das notas.

Existem dois tipos de *staccatos* para o violino: o tocado com o **arco na corda** (aproximado ou “**apoiado**”) ou com o **arco saltante** (também conhecido pela palavra italiana *spiccato*). Existe muito desacordo quanto à exata aplicação destes termos, mas felizmente para o compositor ou orquestrador basta a indicação de pontos sobre as notas para referenciar o *staccato* ^[57], incluindo sinais de dinâmica e outros caso deseje especificar melhor a qualidade da música a ser interpretada.

No tipo saltante (*spiccato*), em velocidade moderada e nível de dinâmica do *piano* ao *mezzo forte*, baixa-se o arco sobre a corda com um movimento semicircular pela metade do arco. O arco então repica na corda embora seja também parcialmente levantando pelo pulso. Em tempo acelerado não ocorre o levantamento do arco, mas sim rápidos movimentos para baixo e para cima dirigidos pelo pulso.

Nos *staccatos* apoiados não se permite que o arco abandone a corda, e o ataque é muito veloz, com início e fim abruptos. O som destes poderia ser imaginado na forma de um bloco, com extremidades quadradas^[57]. Alguns autores, como Piston^[57], consideram a técnica de *martelé* (*martellato*) como sendo um *staccato apoiado*. Alguns instrumentistas consideram ainda o *martelé* um tipo de acento, dado o destaque e a dinâmica envolvidos³⁵. Há ainda outros tipos de *staccatos*, como o *staccato ligado* que deve ser realizado com o arco saltante, com paradas bem pronunciadas e abruptas pelo pulso, e uma infinidade de combinações possíveis.

Para a flauta, os *staccatos* são realizados por movimentos rápidos de retração e avanço da língua sobre a parte posterior dos incisivos superiores, o que produz chaveamentos rápidos, súbitos e destacados no fluxo de sopro. Quanto melhor a técnica de articulação do intérprete, tanto melhor a qualidade do *staccato*.

Fraseados

Fraseados são eventos musicais combinados (em sequência ou simultâneos) que produzem algum efeito psicológico, perceptual ou atingem um objetivo musical em conjunto, e não separadamente. Os eventos musicais são ditos formarem uma *frase*, um arranjo que alcança - assim como na formação de frases e sentenças a partir de vocábulos individuais - um nível de "complexidade informacional" mais alto, transmitindo uma identidade informativa nova, com propriedades emergentes associadas. Existem frases ligadas (*legato*, em italiano) e não ligadas (*non legato*). Nos fraseados ligados o movimento do arco é contínuo e encampa todas as notas ligadas. Nos não ligados geralmente há uma mudança no sentido do arco para cada nota do fraseado.

O que se deseja estudar nas análises de fraseados é como se dá a conexão entre os vocábulos simples, isto é, como são as transições entre as notas individuais que formam a frase. Estas ligações serão diferentes nos fraseados ligados e não ligados.

³⁵ Comunicações pessoais ouvidas de alguns violinistas.

Ligado (*legato*)

normal passagens onde as notas numa composição melódica são interpretadas sem empregar acentos, modulações ou efeitos especiais. Para o violino existem dois tipos especiais de fraseados ligados: o portamento e o glissando. A fronteira entre ambos é estreita e suscetível de ambiguidade em alguns casos.

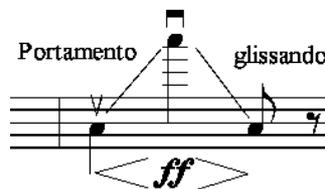
A Figura 21 abaixo ilustra frases ligadas normais. A primeira (à esquerda) consiste num trecho de 4 notas (de mesma duração) tocadas em seqüência. A segunda (à direita) mostra um trecho com duas mínimas ligadas - um lá³ (A3) seguido por um dó⁴ (C4) - havendo indicações de uma variação de dinâmica (a segunda nota sob interpretação fortíssimo) e para que a frase seja *sempre legato*.³⁶

Figura 21 - Exemplos de Fraseados Ligados Normais



O glissando consiste numa mudança contínua no tom (*pitch*) de uma nota a outra, deslizando-se o dedo pela corda da altura de uma nota à altura da outra sob arco contínuo, e com isso soando todos os *pitchs* intermediários entre elas.³⁷ O *pitch* varia aproximadamente linearmente no tempo, a curva exata de variação estando fortemente correlacionada à curva de velocidade do deslize do dedo sobre a corda.

Figura 22 - Exemplo de notação em partitura de um portamento e/ou glissando



³⁶ Nesta seção algumas partituras serão introduzidas para ilustrar a notação em Música para os eventos sob descrição. As partituras mais simples na música ocidental consistem de uma pauta com 5 linhas onde são escritas as notas musicais. A clave, símbolo no início da pauta, indica o registro (altura em frequência) para o qual está se escrevendo. Para flauta e violinos o registro típico é o da clave de sol (♩). Adicionalmente notações periféricas à pauta exprimem aspectos dinâmicos, suprem informações quanto a gestos específicos, articulações e outros atos para guiar a interpretação.

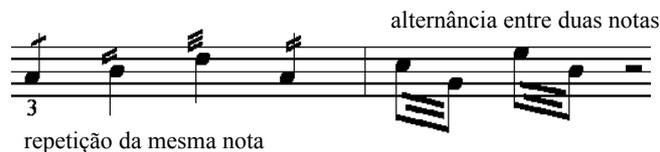
³⁷ O glissando é uma técnica mais natural nos instrumentos de corda, embora na 1ª metade do século XIX fosse popular na Inglaterra e Alemanha para a flauta. Na Inglaterra Vitoriana havia também uma técnica de glissando para flauta, conhecida como "rush": uma combinação de *pitch bending* com o vagaroso alçar dos dedos, em desuso atualmente. (informações do FAQ -*Frequently Asked Questions*- da **FLUTE list**, do servidor de listas listserv.syr.edu).

O portamento também consiste numa transição contínua entre duas notas, mas com a intenção de prover uma maneira elaborada de conectar os tons melódicos ligados, sem no entanto considerar a cobertura de todo o intervalo. Assim, o portamento pode associar um glissando veloz e mais objetivo no início - sem a intenção de conferir expressividade durante a transição, mas simplesmente levando o dedo de uma posição a outra sem interromper o som ligado- e um pequeno salto já próximo da segunda nota. É executado principalmente quando o fraseado exige uma mudança de posição (empunhadura e dedilhado) ou quando a composição assim o determina especificamente para fins expressivos. O glissando, pelo maior impacto expressivo que imprime, foi escolhido para efeito de análise neste trabalho (veja partitura ilustrativa acima, à Figura 22).

acentuado passagens onde as notas são interpretadas empregando-se algum tipo de acentuação, como nos acentos já abordados em Ataques (notas individuais) anteriormente, mas atentando-se ao fato de que agora eles são executados num contexto melódico, sobre uma frase. Oportunamente será discutido a diferença entre notas soltas ou individuais daquelas imersas num contexto fraseado.

trêmulo dedilhado (flauta) O trêmulo consiste na repetição rápida de uma nota dentro do seu tempo, ou a alternância entre 2 notas (de intervalo harmônico normalmente maior que um tom, ou uma segunda maior) dentro do tempo de uma delas, numa velocidade estipulada ou não (*tremulando*).^[57]

Figura 23 - Ilustração das notações usuais para o trêmulo



O **trêmulo dedilhado** em flautas consiste na execução de duas notas no tempo de uma, e pode ser ligado, quando executado num sopro contínuo dando origem a um fraseado do tipo ligado, ou não ligado, quando há interrupção no sopro entre cada nota. O trêmulo ligado, aqui abordado, possui uma qualidade diferente daquele interpretado sem ligaduras, e por envolver a excitação de dois tons diferentes

distantes de um tom ou mais de forma contínua implica numa interpretação mais difícil.

Não Ligado (non *legato*)

normal passagens onde as notas numa composição melódica são interpretadas sem empregar acentos, modulações ou efeitos especiais, em condições normais de interpretação. Para a flauta a transição não ligada baseia-se na interrupção do fluxo de sopro a cada nota. Para o violino um alternância no sentido do arco é esperada a cada nota.

acentuados frases (não ligadas) com acentos.

staccatos *staccato* normal para flauta e *spiccato* para o violino

Figura 24 - Ilustração de uma melodia contendo frases não ligadas com *staccatos* e acentos



trêmulo dedilhado (flauta) alternância rápida entre duas notas excedendo intervalo maior que um tom, sem ligaduras (sopro contínuo)

trêmulo de arco (violino) quando o trêmulo é entre duas notas que estão além do alcance do dedo, faz-se necessário o uso de 2 cordas para executá-lo, e o arco deve portanto atuar sobre as duas cordas alternadamente. Muito utilizado também quando o propósito do trêmulo é a **repetição iterada** da mesma nota dentro do tempo de uma, novamente o arco sendo crucial em realizar os ataques velozes de cada iteração, alternando o sentido em cada movimento de descida agora sobre uma mesma corda. O exemplo abordado para análise consta da repetição iterada da mesma nota.

Sustentação

Com dinâmica com *crescendo* (*cresc.*) e/ou *diminuindo* (*decresc.*), isto é, com uma variação na intensidade (magnitude ou amplitude) da(s) nota(s) tocada(s) durante a execução³⁸. A figura abaixo ilustra a notação usual.



Com vibratos O vibrato no violino é o resultado de uma combinação de impulsos dos músculos da mão, pulso e braço, e seu efeito é o de uma oscilação no *pitch* da nota quase imperceptível. É aplicado em todas as notas, exceto as de curta duração, e não necessita indicação para seu uso. Entretanto, quando deseja-se suprimir o vibrato indica-se com o termo italiano *senza vibrato*^[57]. Fisicamente, o vibrato pode ser definido como uma baixa modulação de frequência por volta de 6 Hz em uma extensão inferior a 1%.^[61] Para a flauta o vibrato consiste mais numa modulação de amplitude aplicada por uma variação na pressão de sopro e/ou na embocadura, esta última podendo provocar também alguma flutuação de frequência.³⁹

Com trinados Alternância rápida entre notas adjacentes, obtida pelo chaveamento veloz no dedilhado.⁴⁰

Com frulatos (flauta) Considerado um trêmulo especial em instrumentos de sopro, consiste numa repetição muito rápida da(s) nota(s) tocada(s) obtida modulando-se o fluxo de ar pela vibração da língua, como na

³⁸ Existe uma gradação ou escala de intensidade cujos níveis são representados por combinações das letras *p* (piano ou fraco) e *f* (forte). Uma gradação usual, do menos intenso ao mais intenso, é: ppp, pp, p, mp, mf, f, ff, fff. O símbolo ">" indica uma queda gradal na dinâmica, e "<", um aumento gradual.

³⁹ O vibrato na flauta é uma técnica que se popularizou no século XX. O século XIX não faz muitas menções ao vibrato, exceto pelo vibrato de dedilhado. (informações do FAQ -*Frequently Asked Questions*- da **FLUTE list**, do servidor de listas listserv.syr.edu) .dois

⁴⁰ Começar o trinado pela nota auxiliar mais alta (em *pitch*) era mais comum do que começá-lo pela nota principal, pelo menos até aproximadamente 1830. O trinado é uma técnica que pode ter seu estilo variável segundo a época. No sistema de trinados para flauta do Século XIX, alguns trinados eram estreitos e entediantes, em contraste com os trinados vivos e largos nas flautas barrocas. (informações do FAQ -*Frequently Asked Questions*- da **FLUTE list**, do servidor de listas listserv.syr.edu)

pronúncia de um “r”. Uma sequência de ataques reincidentes é então impressa às notas sobre frulato. Efeito mais destacado na flauta entre os instrumentos de sopro. ^[57]

4.2.2 Metodologia e implementação da análise das amostras

Este trabalho fundamenta-se em **dois módulos básicos**, um **teórico** - cuja contribuição está na abordagem à teoria *wavelet* e à técnica de análise e síntese em computação musical - e um **experimental**, que consiste na análise em multiresolução de gestos musicais em timbres instrumentais empregando a transformada *wavelet*. O tema central é a transformada *wavelet* como ferramenta de análise de sinais musicais, e a aplicação proposta é a análise de timbres instrumentais objetivando a caracterização de gestos musicais e padrões sônicos associados às técnicas de interpretação instrumental.

O procedimento no **módulo teórico** incluiu:

- uma introdução à área de computação musical, aos timbres musicais de dois instrumentos acústicos tradicionais (flauta e violino), e à técnica de análise e síntese;
- uma abordagem introdutória à teoria *wavelet*, suas aplicações em processamento de sinais e outras áreas;
- uma abordagem à análise em multiresolução com *wavelets* e sua implementação discreta através de bancos de filtros;
- uma abordagem conceitual aos gestos musicais, fraseados e modulações associadas a técnicas de interpretação instrumental;
- uma introdução aos algoritmos para decomposição e reconstrução *wavelets*, e à análise de timbres em multiresolução com *wavelets*.

O procedimento utilizado no **módulo experimental** consistiu de:

- seleção dos instrumentos (timbres), seleção do material sonoro a ser amostrado e analisado - **trechos musicais** contendo ataques, fraseados e efeitos de sustentação, e uma breve descrição destes eventos e gestos musicais;
- gravação e preparação das amostras (trechos musicais);

- descrição da plataforma computacional, apresentação dos algoritmos de decomposição (pacote de *software*) e das bases *wavelets* empregadas;
- implementação da análise das amostras (trechos musicais);
- análise e discussão dos resultados.

A flauta foi escolhida por ser um instrumento extremamente difundido e polivalente, empregado em inúmeros estilos e modalidades musicais, nas orquestras sinfônicas, nas danças e cortejos, no *rock*, na *new age*, na música erudita contemporânea, na música popular dos países europeus, no norte da África, sul da Ásia, pelos povos indígenas das Américas, e também aclamado na música orquestral brasileira - presente no trabalho de expoentes como Carlos Gomes e Villa Lobos entre outros- assim como -não poderia deixar de ser- na música popular brasileira, onde ecoa sonora na voz de inúmeros maxixes, choros, sambas, na bossa nova, e no *rock* brasileiro. É um instrumento de origem muito antiga, presente em várias sociedades desde a pré-história, de timbre essencialmente harmônico, melódico e alvissareiro. A condição de ser eu próprio um instrumentista, buscando caracterizar nuances das técnicas de interpretação em flauta, vem se somar a estes motivos no peso da escolha.

O violino, instrumento de origem na Europa do século XVII, é considerado o mais importante dos instrumentos de cordas. Tradicionalmente empregado em orquestras sinfônicas e de câmaras, restrito a círculos nobres, foi progressivamente incorporado a outros estilos neste século, e com o advento dos captadores elétricos a tecnologia musical ganhou terreno em novos estilos musicais, incluindo aí o *rock* e a música eletroacústica. O apogeu do artífice do instrumento, entretanto, foi atingido nos séculos XVII e XVIII, nas mãos de importantes *luthiers*, e desta época datam os melhores instrumentos fabricados, até hoje não ultrapassados. A escolha por ele recai naturalmente no interesse em verificar a influência dos gestos musicais e técnicas de arco sobre a qualidade do som emitido.

A metodologia adotada para a **gravação e preparação das amostras** de trechos musicais contendo os eventos/gestos a estudar é descrita a seguir:

Foram escolhidos trechos musicais que contivessem os eventos e gestos musicais objetos de análise (sumariados na tabela de eventos musicais, no item 4.2.1, e ilustrados no item 5.1, Figura 32 e Figura 33). A gravação foi realizada no estúdio do

Departamento de Cinema, Rádio e Televisão, da Escola de Comunicação e Artes da USP. Apesar de seu condicionamento acústico orientado para as atividades sonoras deste departamento, e não propriamente para fins musicais, o estúdio forneceu o isolamento acústico mínimo necessário para uma boa gravação.

O equipamento de gravação empregado foi um DAT (*Digital Audio Tape recorder*) marca TEAC, modelo DA-P20. O microfone utilizado foi um Neuman, modelo MKH 416 P 48 U, com resposta de frequência de 40 a 20000 Hz, característica direcional supercardióide. A gravação foi realizada com 48 KHz de taxa de amostragem (f_a), palavra digital com resolução de 16 bits por amostra.

A preparação das amostras para análise em computador a partir da gravação em DAT envolveu as seguintes etapas:

- Transporte da gravação de áudio digital serial em fita magnética para meio de armazenamento em disco rígido, realizada numa estação Indy, da Silicon Graphics;
- conversão do áudio em DAT estéreo 48 KHz para arquivos em formato AIF mono 44,1 KHz de taxa de amostragem, realizada numa estação Indy, da Silicon Graphics e também num PC Pentium 100 MHz;
- segmentação e edição dos arquivos, para isolar os eventos musicais em arquivos AIF individuais menores, realizada numa Indy e em PC;
- normalização de amplitude, realizada em parte numa estação Indy, rodando o aplicativo SoundEditor ou MediaConvert, e em parte num PC rodando o aplicativo CoolEdit;⁴¹
- conversão dos arquivos do formato AIF para formato MATLAB, utilizando-se o aplicativo Goldwave, em PC Pentium 100 MHz;
- normalização da extensão dos arquivos para valores em potência de 2. Duas técnicas foram utilizadas com esse objetivo:
 - a **prévia produção e edição** do arquivo AIF (ou MAT) para uma extensão (número de amostras digitais ou pontos) **exatamente em potência de dois**, empregada principalmente quando o arquivo original excedia levemente em

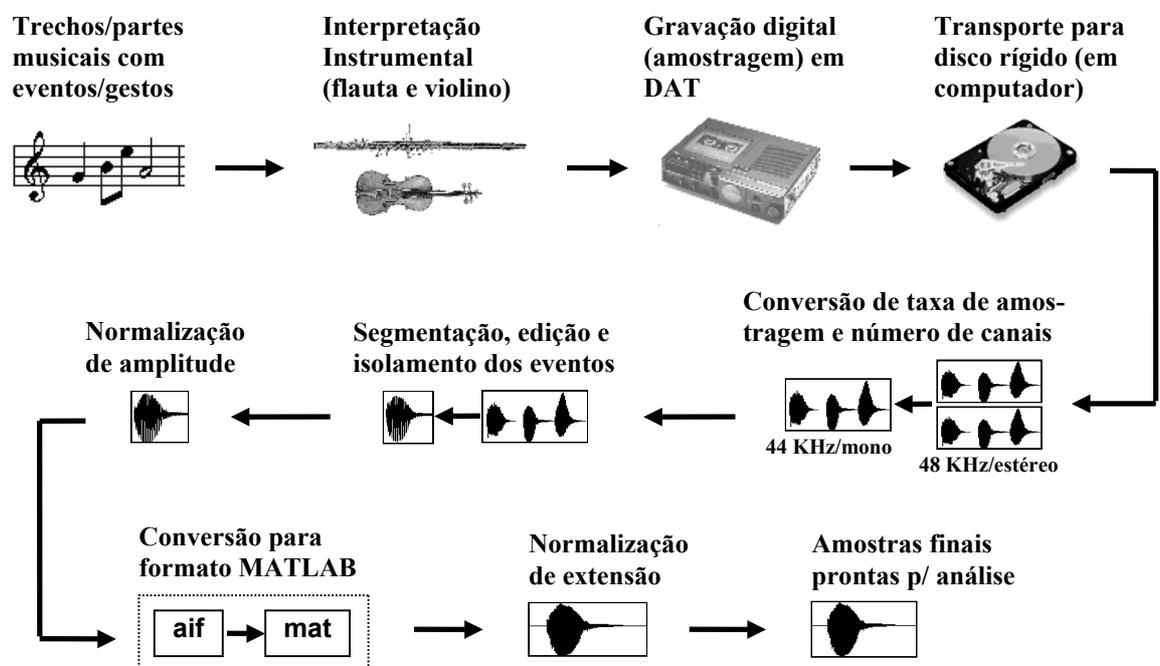
⁴¹ SoundEditor é um aplicativo (*software*) para edição e processamento básico de áudio digital da Silicon Graphics. CoolEdit (<http://www.syntrillium.com>) e GoldWave (<http://web.cs.mun.ca/~chris3/goldwave/>) são aplicativos *shareware* (disponíveis na Internet) para edição e processamento simples de áudio digital em PC's.

extensão uma potência de dois próxima, ou quando o evento musical em questão podia ser isolado dentro de uma exata extensão em potência de dois, com as bordas do sinal (início e fim) nulas ou próximas de zero;

- a técnica de **acolchoamento por zeros** (*zero padding*) até atingir-se a extensão da potência de dois superior mais próxima, adicionando pontos nulos (zeros) ao início e ao fim da amostra (trecho musical editado).

A Figura 25 abaixo ilustra estas etapas de gravação e preparo das amostras musicais para análise:

Figura 25 - Gravação e preparação das amostras contendo os eventos musicais para análise



O procedimento para as análises das amostras (arquivos com eventos musicais) incluiu:

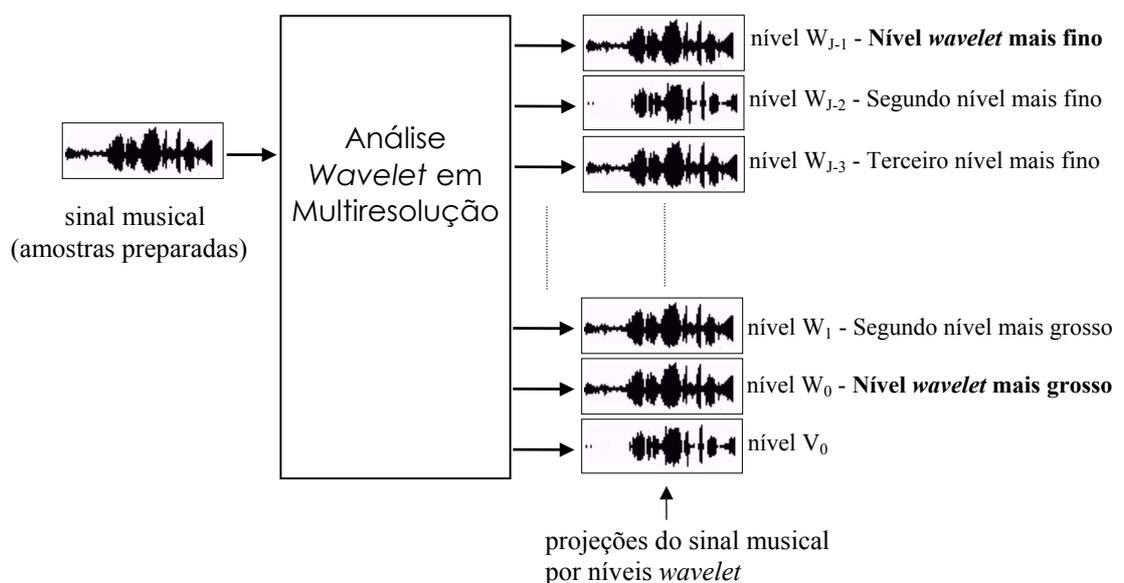
- decomposição *wavelet* de cada amostra em multiníveis, utilizando as rotinas ortogonais do pacote WaveLab para MATLAB, desenvolvido na Universidade de Stanford;
- a análise gráfica e sonora dos sinais em cada nível da decomposição em multiresolução, observando o padrão de presença e a concentração do evento musical em cada nível analisado;
- identificação gráfica e sonora de padrões (objetos) correspondentes aos eventos musicais analisados nos níveis;

- verificação do isolamento e da distribuição dos eventos por níveis *wavelets*, avaliando a energia de cada nível, os padrões de distribuição gráficos e o conteúdo sonoro dos níveis;
- eventuais análises espectrais complementares como ferramenta auxiliar;
- discussão dos resultados e fenômenos observados.

4.2.2.1 Descrição do Sistema de Análise *Wavelet*

O diagrama geral do sistema de análise implementado está esquematizado na Figura 26 abaixo. Os eventos musicais de interesse, já editados, normalizados e armazenados em sinais digitais segundo o procedimento ilustrado na Figura 25, são os sinais de entrada. O algoritmo computacional de decomposição *wavelet* utiliza um banco de filtros e operadores de decimação arranjados sucessivamente num esquema piramidal (árvore de filtros), como introduzido no item 2.5.1. Os sinais de saída são os sinais em cada nível *wavelet*, isto é, as projeções do sinal original em cada nível da análise em multiresolução.

Figura 26 - Diagrama de blocos geral do sistema para análise de eventos musicais



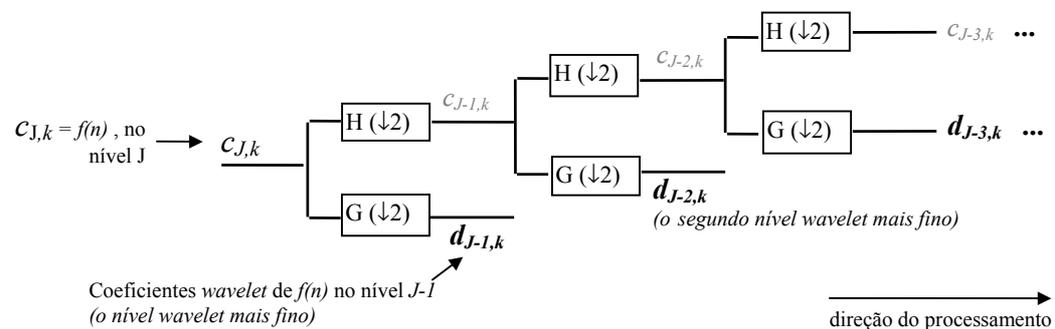
No topo da árvore de decomposição, o algoritmo toma os sinais de entrada como **coeficientes de escalamento** $c_{J,k}$ na escala mais fina possível: **a original, ao nível J**. O

processamento começa a partir daí, gerando todos os coeficientes $c_{j,k}$ até o nível mais baixo de resolução: o nível mais grosso (ou menos fino), em $j=0$.

Se, por exemplo, o vetor de entrada tem extensão $N=2^3$, isto é $c_{3,k}$ possui oito elementos no nível mais fino $J=3$; o próximo nível ($j=2$, o **nível wavelet mais fino**) terá 4 coeficientes ($d_{2,k}$), o nível seguinte $j=1$ terá 2 elementos ($d_{1,k}$), e o nível mais baixo de resolução - $j=0$ - terá 1 coeficiente *wavelet* ($d_{0,0}$ do **nível wavelet menos fino**) e 1 coeficiente de escalamento $c_{0,0}$, resíduo final do iterado processo de suavização, na verdade uma figura representativa da média de todos os coeficientes originais.

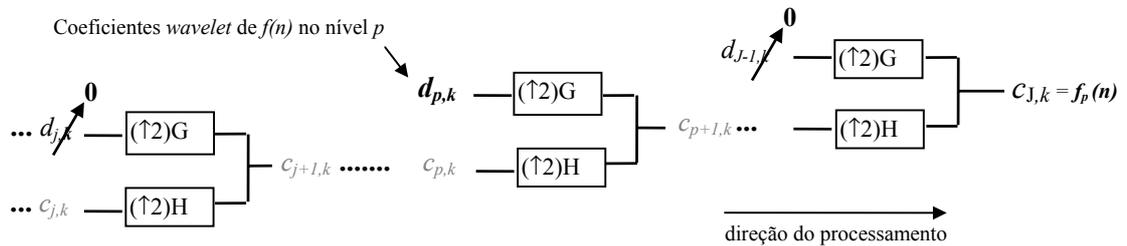
Na decomposição são gerados coeficientes *wavelets* para os níveis *wavelets* (W_j) em escalas sucessivas, numa razão de escalamento 2, isto é, em oitavas. Cada nível de coeficientes possui a metade do número de amostras do nível anterior resultado do processo iterado de decimação. O número de coeficientes gerados ao final da transformação é igual ao número de amostras (elementos) do vetor de entrada.

Figura 27 - Decomposição do sinal musical $f(n)$ em coeficientes wavelet $d_{j,k}$ na árvore de filtros



Os coeficientes *wavelet* gerados nesta decomposição não correspondem propriamente às projeções do sinal original em cada nível de resolução. Estas são geradas realizando-se uma síntese, o processamento inverso, considerando um nível de coeficientes por vez, zerando-se todos os outros. Assim, para gerar a projeção do sinal no nível j , considera-se os coeficientes *wavelet* calculados no nível j ($d_{j,k}$) e perfaz-se uma síntese trafegando a árvore em sentido inverso, até obter-se um sinal no nível original J . Este novo sinal, gerado somente a partir das informações do nível j , é a projeção do sinal original naquele nível de resolução (veja Figura 28 abaixo). Este procedimento é utilizado para gerar todos os níveis da análise em multiresolução, os sinais de saída da Figura 26.

Figura 28 - Geração do sinal no nível *wavelet* p : a projeção de $f(n)$ no nível *wavelet* W_p



O sistema de análise *wavelet* em multiresolução (AWMR) acima foi totalmente implementado em plataforma computacional MATLAB. Os sinais de entrada foram convertidos em vetores no formato MAT. Os sinais de saída, também vetores em formato MAT, foram analisados gráfica e sonicamente utilizando-se o próprio ambiente MATLAB, algumas rotinas desenvolvidas para avaliar o conteúdo energético e espectral de cada nível, e programas editores, processadores e reprodutores de áudio digital em plataforma PC, como o CoolEdit e GoldWave, este último capaz de ler arquivos sonoros em formato MAT.

IMPLEMENTAÇÃO ALGORÍTMICA

Os algoritmos de decomposição, de geração das aproximações do sinal em cada nível W_j , e de reconstrução *wavelet* que foram empregados têm origem nas rotinas do pacote computacional WaveLab, desenvolvidos para a plataforma computacional para processamento de sinais MATLAB, e de domínio público, acessíveis na Internet.

WAVELAB

WaveLab é uma biblioteca de rotinas MATLAB para análise *wavelet*, para análise com pacotes de *wavelets* (*wavelet-packet analysis*), para análise com pacotes de cossenos e busca de casamento (*matching pursuit*) [6]. Sua biblioteca reúne ao todo mais de 800 arquivos de extensão .m (em linguagem MATLAB), documentados e inter-referenciados de várias formas. Várias pesquisadores contribuíram e contribuem no desenvolvimento deste *software*, entre eles Jonathan Buckheit, Shaobing Chen e David Donoho, da Universidade de Stanford. O *software* está disponível através de FTP para Stanford no endereço `fpt://playfair.stanford.edu/pub/wavelab`. [6] Uma cópia da versão 0.701 está inclusa no CD-ROM anexo, gentilmente cedida pela sua equipe de desenvolvimento.

As rotinas estão codificadas em linguagem MATLAB, o que possibilita o trabalho de leitura e compreensão da implementação algorítmica, e sua modificação para fins específicos, adaptando-as às necessidades do trabalho em questão. Desta forma desenvolveu-se um conjunto de rotinas alternativas, melhor adaptadas às tarefas de análise dos sinais musicais, assim como uma interface de comandos gráfica sobre o MATLAB, que incorporasse as funções do WaveLab juntamente com uma série de outros comandos orientados a tarefas específicas de configuração de parâmetros, entrada e saída de dados.

Rotinas originais do WaveLab foram reescritas e suas funções modularizadas visando a focalizar o serviço reduzindo o número de operações a um núcleo mínimo. O impacto mais importante foi na redução do custo computacional do processamento na decomposição, na geração de figuras e nas tarefas de ressíntese seletiva⁴². Rotinas de visualização gráfica foram incrementadas tal que possibilitassem a plotagem de cada nível sob análise numa escala conveniente e numa janela (na tela do computador) específica. Outras, para reprodução e gravação de áudio, foram criadas atendendo às demandas da análise sônica.

As operações de decomposição e reconstrução wavelet baseiam-se nas rotinas ortogonais do WaveLab, localizadas nos seus diretórios `\ortho` e `\ortho.1`, especialmente as rotinas `fw_t_po.m` (*forward wavelet transform, periodized, orthogonal*) e `iwt_po.m` (*inverse wavelet transform, periodized, orthogonal*), que implementam a transformada direta e inversa, respectivamente.

DESCRIÇÃO DA INTERFACE GRÁFICA, FUNÇÕES E CÓDIGOS DESENVOLVIDOS

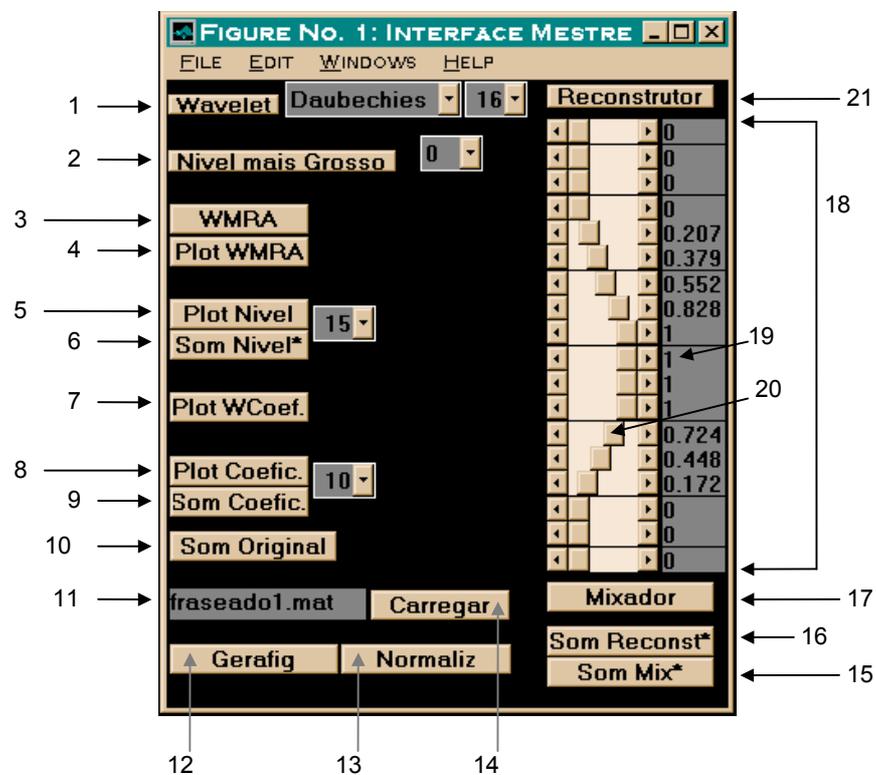
Uma interface gráfica para MATLAB foi desenvolvida para se sistematizar e agilizar o trabalho de análise, provendo um painel de comandos orientado às funções (tarefas) de configuração de parâmetros para a análise constantemente empregadas. A rotina que gera a interface gráfica e as janelas diferenciadas para plotagem gráfica dos níveis, coeficientes, reconstruções seletivas e mixagens é `interface.m`, e seu código fonte está disponível no CD-ROM e no disquete anexos, juntamente com todas as outras rotinas desenvolvidas e/ou incrementadas.

⁴² A reconstrução ou ressíntese seletiva será abordada no item “análise sônica”, logo adiante.

O conjunto de todas elas forma um pacote novo, que se nomeou **Wavesynt** - *uma contribuição em software deste trabalho, para a análise wavelet em multiresolução de sinais musicais* - e que roda sobre o MATLAB utilizando-se do WaveLab previamente instalado.

A Figura 29 abaixo mostra a janela de comandos principal (ou interface mestre) do **Wavesynt** na sua atual versão, indicando os botões, campos numéricos disponíveis e respectivas funções:

Figura 29 - Interface gráfica desenvolvida sobre MATLAB: janela de comandos



1 Seleção da <i>wavelet</i>	2 Seleção de nível mais grosso
3 Decomposição <i>wavelet</i> (geração de coeficientes <i>wavelet</i> para todos os níveis da análise)	4 Gráfico de todos os níveis da análise <i>wavelet</i> em multiresolução (AWMR)
5 Seleção e plotagem de nível específico da AWMR	6 Sonorização do nível selecionado
7 Gráfico dos coeficientes por nível	8 Seleção e plotagem dos coeficientes de nível específico
9 Sonorização dos coeficientes do nível selecionado	10 Sonorização do sinal original
11 Seleção de nome de arquivo para entrada e saída (arquivos com extensão .MAT)	12 Gerador de figura com todos os gráficos de todos os níveis da AWMR
13 Normalização em extensão do vetor de entrada (sinal original)	14 Carregar arquivo selecionado
15 Sonorização da mixagem	16 Sonorização da reconstrução
17 Mixador de níveis	18 Editor de pesos para reconstrução e mixagem
19 Seleção numérica para o peso do nível	20 Seleção via botão deslizante para o peso do nível
21 Reconstrutor de níveis	

Os objetivos imediatos da interface foram otimizar:

- a entrada de dados (os arquivos sonoros contendo as amostras musicais);
- a escolha da *wavelet* de análise;
- a escolha do nível mais grosso para a análise (usualmente o nível 0);
- a saída de dados;
- a audição dos sinais;

bem como facilitar a programação de parâmetros extras, como as escalas para plotagem.

Foram implementadas:

- funções para decomposição e plotagem de todos os níveis da análise
- funções para geração de cada nível e cálculo de sua energia
- função de reconstrução seletiva (ressíntese), com paleta para edição dos pesos de cada nível
- função misturadora, para mixar diretamente os sons dos níveis
- funções para impressão dos níveis da análise em arquivo, em formato gráfico.

Tanto o WaveLab quanto o **Wavesynt** são facilmente instaláveis em qualquer plataforma onde rode o MATLAB. O **Wavesynt** é abordado nos apêndices, e informações técnicas mais detalhadas sobre ele e sobre o WaveLab estão disponíveis no CD-ROM em anexo.

SELEÇÃO DAS BASES WAVELETS

Wavelets diferentes levam a representações diferentes do sinal analisado. Conseqüentemente, a separação de padrões sônicos (confinamento dos eventos musicais em níveis de resoluções diferentes) pode ser melhor verificada e destacada para algumas bases específicas. Selecionou-se um grupo amostral de famílias *wavelets* para avaliar qual dentre elas exibia melhor desempenho na tarefa de segregação de componentes e estudar a textura e características dos níveis *wavelets* produzidos por elas, com a finalidade de se selecionar a(s) base(s) a ser adotada para a consecução das análises propostas.

A seleção das *wavelets* traz também um impacto sobre a eficiência e a estabilidade computacional dos algoritmos de decomposição e reconstrução; sobre a obtenção de representações não redundantes e várias taxas de compressão; sobre a suavidade e continuidade das representações transformadas (as aproximações do sinal original em

cada escala), e sobre a resolução da localização temporal e frequencial^[55]. *Wavelets* são frequentemente classificadas em famílias pelo número de momentos nulos que possuem^[26]. A regularidade e suavidade das *wavelets* estão associadas com este parâmetro, que também influencia na localização em frequência e na convergência das aproximações.^[55]

As seguintes famílias de bases foram avaliadas^[6; 13; 69; 84; 88].

- Daubechies (D4, D8, D12, D16)
- Coiflets (C2, C4)
- Beylkin
- Simlets

Verificaram-se diferenças significativas entre *wavelets* de uma mesma família com número de momentos diferentes. As *wavelets* mais regulares propiciaram uma melhor sintonia da janela de análise, isolando melhor faixas de frequências por níveis. Isto se deve pelas características dos filtros associados (do banco de filtros) que possuem um decaimento mais pronunciado nas zonas de corte, levando a uma melhor localização em frequência das *wavelets* correspondentes.

Isto pode ser verificado pela inspeção dos espectros cobertos por exemplo pelas *wavelets* D4 e D16, como mostrados na Figura 9 (item 2.2, à página 32). A cobertura da D16 é mais localizada ou sintonizada, exibindo uma banda (passante) principal mais intensa, as bandas secundárias bem mais atenuadas que as da D4. Isto conduz a uma menor supercobertura, ou redundância na representação frequencial. O resultado prático desse atributo está na simplificação do conteúdo espectral dos níveis *wavelets* gerados na análise, que cobrem faixas mais localizadas de frequências, acabam por isolar melhor alguns componentes em níveis diferentes, e propiciam uma textura sonora “mais aveludada”, um som “menos áspero”.

Foi realizado um experimento comparativo de desempenho para qualificar atributos diferentes gerados na análise de um evento musical de sustentação com variações de dinâmica em violino com as *wavelets* ortogonais D4, Simlet4, Coiflet2, Beylkin e Vaidyanathan⁴³. A textura sonora revelou-se levemente diferente entre elas, com exceção de D4, que reproduziu sons mais ásperos nos níveis mais finos. Os envelopes e

⁴³ Note que D4 possui somente 2 momentos nulos, ao passo que Coiflet2 e Simlet4 possuem 4 momentos nulos.

formas de onda produzidos pela Simlet4 mostraram-se mais regulares e menos ruidosos que com D4. Coiflet2 produziu níveis com formas de onda semelhantes a Simlet4. Beylkin realçou alguns picos em algumas formas de onda mais que Simlet4. D4 revelou formas de onda mais ruidosas, menos identificadas com as formas de onda original (para o nível 10 da análise).

As rápidas conclusões que extraem-se deste e outros experimentos facilmente implementáveis com a interface gráfica são que:

- entre as *wavelets* ortogonais comparadas de mesmo número de momentos nulos, isto é, mesmo grau de regularidade e suavidade, as formas de onda dos níveis são semelhantes;
- pequenas variações de cobertura em frequência levam a um realce de alguns objetos em alguns níveis (com componentes espectrais sob as bandas passantes);
- a cobertura espectral das *wavelets* ortogonais comparadas de mesmo número de momentos nulos é bastante similar, levando a representações aproximadamente idênticas;
- o parâmetro de maior impacto sobre a qualidade sônica dos níveis (para as *wavelets* ortogonais comparadas) é o número de momentos nulos satisfeitos pela *wavelet*.

Estudos mais amplos, que englobassem uma análise comparativa para um grupo maior de eventos musicais diferentes, provavelmente apontariam à conclusão de que a escolha da família *wavelet* deveria ser feita em função do tipo de evento musical que se deseja caracterizar, e que o grau de regularidade seria um parâmetro importante na obtenção de melhor separação de objetos, suavidade das formas de onda e textura sônica. Isto se deve também ao fato de que *wavelets* com formas de onda assemelhadas às formas de onda dos eventos analisados conduzem a representações mais compactas, com representações mais precisas deste eventos com menor número de coeficientes.^[79]

No estudo proposto a variabilidade do espaço amostral é extenso: os eventos musicais que dispusemos a caracterizar exibem variações e conformações diferentes tanto no tempo quanto na frequência. A escolha de muitas variantes de *wavelets* para casar melhor cada classe de eventos tornaria o estudo muito amplo, alargando substancialmente o espaço de variáveis a analisar e os resultados. Além disso não existem trabalhos na literatura que apontem quais *wavelets* atualmente existentes

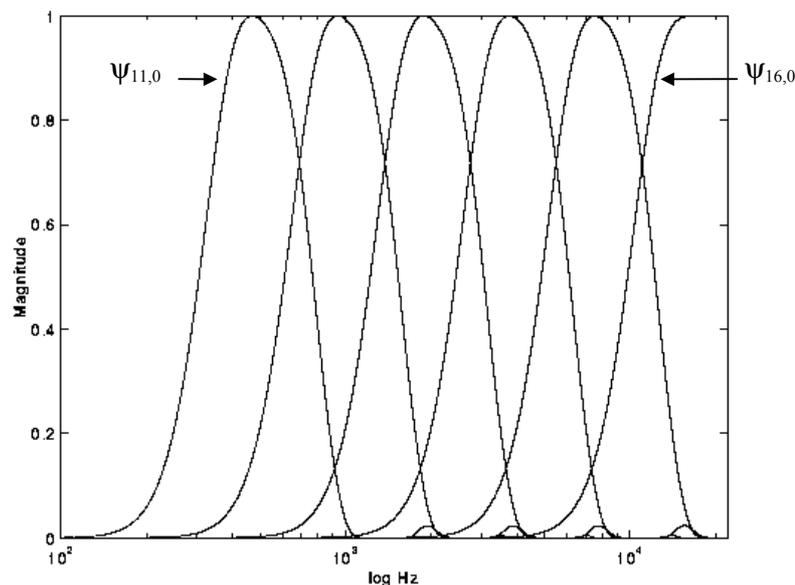
produzem melhor desempenho analisando classes específicas de eventos. O tema é inédito.

A saída natural recai portanto na escolha de uma família única *wavelet* a ser empregada em todas as análises. Os benefícios são óbvios: **(1)** o estabelecimento de uma referência que possibilite uma avaliação geral da aplicabilidade e do desempenho da *wavelet* para diversos tipos de eventos, e **(2)** a simplificação direta na complexidade deste trabalho.

A regularidade é parâmetro decisivo na escolha da *wavelet*. Ortogonalidade é um atributo indispensável para implementar os algoritmos de geração dos níveis da análise em multiresolução e as ressínteses seletivas (a serem ainda abordadas). A facilidade de implementação e a disseminação na literatura da base *wavelet* escolhida são outros fatores desejáveis. A escolha, baseado nestes aspectos, recai sobre o emprego das *wavelets* de Daubechies, em especial a **D16** que possui 8 momentos nulos, um grau de regularidade nem muito baixo e nem extremamente alto, tal que se possa avaliar o desempenho de análises sob graus de regularidade médios. Além disso, com 16 coeficientes no filtro *wavelet* correspondente, estabelece-se um compromisso entre custo computacional e suavidade na representação.

A Figura 30 abaixo mostra a cobertura em frequência de 6 *wavelets* D16: $\Psi_{11,0}$ a $\Psi_{16,0}$. Foram plotadas as FFT's de *wavelets* D16 com extensão de 131072 pontos amostrais de extensão (2^{17}), adequadas para análise de sinais discretos de mesma extensão. Note a semelhança de suas formas de onda quando plotadas em escala logarítmica.

Figura 30 - Cobertura em frequência das *wavelets* D16: $\Psi_{11,0}$ a $\Psi_{16,0}$ (em escala logarítmica)



Vários textos abordam questões sobre ortogonalidade, biortogonalidade, simetria, suporte compacto, suavidade, regularidade e extensão dos filtros *wavelets*, analisando vantagens e desvantagens de cada família de *wavelets* em aplicações específicas. O leitor interessado achará abordagens apropriadas sobre estas questões principalmente em [13; 35; 47; 55; 78; 84; 88], embora quase todos os textos sobre *wavelets* discutam estes aspectos.

4.2.2.2 Interpretação Gráfica e Sônica da Análise

INTERPRETAÇÃO GRÁFICA

A análise gráfica consiste na interpretação e estudo dos **níveis *wavelets* das análises**: as **representações gráficas** bidimensionais de **intensidade versus** pontos amostrais (**tempo**) das projeções do sinal original $f(n)$ em cada nível da análise em multiresolução, isto é, os sinais $f_j(n)$, também referenciados simplesmente como *os níveis* da análise. A partir destes gráficos proceder-se-á à interpretação dos sinais em cada nível da análise, avaliando a distribuição e o grau de isolamento dos eventos pelos níveis, gerando-se resultados e conclusões.

A análise espectral dos níveis *wavelets* também é outra ferramenta suplementar que foi adotada na análise de alguns dos eventos, para permitir identificar padrões harmônicos existentes nos sinais, avaliar as bandas e/ou frequências mais intensas. A modalidade mais comum adotada foi a análise estática via o cálculo da FFT do sinal, a qual não fornece informações sobre a evolução das frequências no tempo, mas sim de um panorama geral da composição espectral no trecho analisado. Espectrogramas, que permitem o mapeamento dinâmico das frequências, foram empregados somente na análise de glissandos, em violino.

Os níveis *wavelets*, em tese, contém informações confinadas a uma banda limitada do espectro, logo é de se esperar que eventos estritamente produzidos por vibrações ou processos ondulatórios regidos por um conjunto finito e limitado de frequências devam se confinar em alguns níveis, isto é, manifestar-se com realce em alguns níveis. Como a partição do espectro no caso da análise *wavelet* simples se dá em oitavas, havendo uma melhor partição do espectro nas baixas frequências do que nas altas, os níveis mais

finos (bandas largas nas altas frequências) têm maior probabilidade de confinar padrões associados a um maior número de eventos e processos vibratórios que participem da mesma (larga) banda de frequências, e se misturam numa única forma de onda. A separação dos eventos então é menos trivial.

Para os níveis mais baixos entretanto, cobrindo bandas mais estreitas, a confirmação do confinamento de um evento é menos difícil. A primeira vista pode parecer então que uma partição do espectro em oitavas, como feita pela análise *wavelet* simples, não seja útil na extração de características e isolamento de padrões, mas é exatamente essa segmentação não linear do espectro que provê o diferencial da análise *wavelet* sobre as técnicas de Fourier convencionais, e que serve à investigação de inúmeras classes de sinais na natureza em dezenas de aplicações, conforme abordado em itens anteriores. Vale também ressaltar que nossos ouvidos, através de um sofisticado sistema aperfeiçoado pela natureza por milhares de anos, perfazem uma análise dos estímulos acústicos através de uma segmentação do espectro semelhante a uma segmentação de uma análise *wavelet*, e dela conseguem extrair toda a informação que necessitam para compreender o complexo mundo sonoro ao nosso redor. Esse, entre outros, consiste na principal motivação para se adotar também a análise sônica, suplementando a análise gráfica.

INTERPRETAÇÃO SÔNICA

Risset, em seus trabalhos de análise de timbres, utilizou uma metodologia para investigar a composição harmônica dos timbres, que foi a análise da evolução (no tempo) das frequências harmônicas, e uma metodologia de validação da sua análise pela síntese de conjuntos de coeficientes, extraíndo suas conclusões sobre a relevância da ressíntese pela percepção do som ressintetizado, conforme sugere seu artigo^[61].

É fato estabelecido entre os profissionais do ramo musical, tanto a nível técnico quanto musicológico, que o instrumento de análise mais apurado e definitivo em quaisquer avaliações de materiais sonoros é o ouvido humano. A existência de uma base tecnológica oferecendo uma variedade de ferramentas matemáticas e de engenharia exatas e confiáveis para atestar a qualidade, pureza e outros aspectos acústicos da música não pressupõe a dispensa da avaliação auditiva. Pelo contrário, o ouvido frequentemente é quem dá a última palavra, e o que se tem presenciado desde a

introdução dos primeiros equipamentos eletrônicos sonoros são acaloradas discussões sobre as diferenças percebidas na qualidade dos sons sintetizados, gravados e processados de forma digital dos sons processados em sistemas analógicos, como também dos sons naturais nas audições acústicas em espaços reais.

Atualmente, com a recente tendência em se estender os padrões adotados para frequência de amostragem em áudio de 44.1 KHz (ou 48 KHz em alguns equipamentos profissionais) para 96 KHz, objetivando uma representação mais completa, novamente o ouvido tem-se revelado o instrumento de referência para nortear as aferições e definir o espectro daquilo que é perceptualmente rico e expressivo.

Essas razões levam ao estabelecimento de uma metodologia de aferição baseada no julgamento das estruturas resintetizadas através da sua audição e comparação com os sons originais, utilizando o ouvido para mensurar o grau de presença ou isolamento de gestos musicais expressivos em determinados níveis da análise em multiresolução, em outras palavras, um *benchmark* para a classificação validada pela audição.

Para efeito de prova ou validação do grau de confinamento de certo padrão sônico ou evento musical num determinado nível j (ou conjunto de níveis: j, i, \dots) empregamos o sistema de edição dos coeficientes *wavelet* embutido na interface gráfica desenvolvida, no qual zera-se ou determina-se um limiar próximo de zero para todos os coeficientes nos níveis diferentes daquele(s) que se deseja pesquisar, e então procede-se à reconstrução (transformada inversa) gerando um sinal na extensão do sinal original que revele as contribuições daquele(s) nível(eis) pesquisados. A confirmação do confinamento se dá via audição da resíntese dos níveis pesquisados.

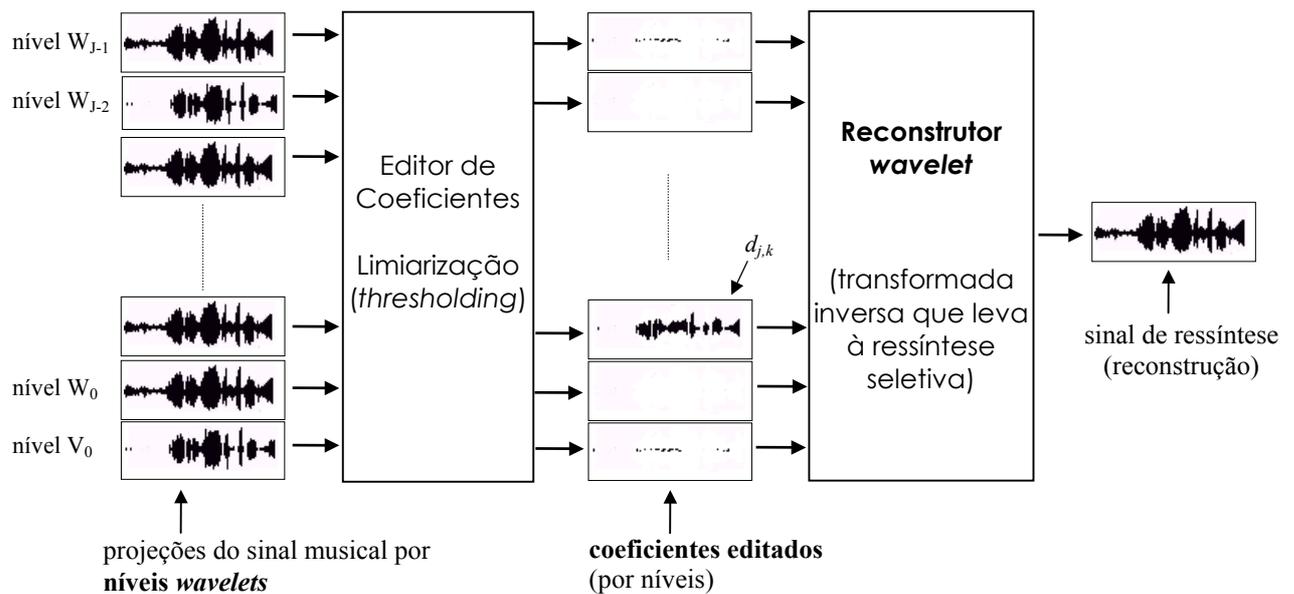
Um procedimento comparativo proposto para este objetivo é descrito abaixo:

- amostra-se o mesmo trecho musical tocado (1) com expressividade ou realce no evento musical que se deseja caracterizar, e (2) sem expressividade no evento musical em questão (execução normal, sem efeito ou modulação)
- analisa-se os trechos com o mesmo procedimento e bases *wavelets*, gerando uma decomposição *wavelet*.
- edita-se os coeficientes dos níveis desejados em ambos os trechos, mantendo a contribuição do(s) nível(is) sob pesquisa e reduzindo os coeficientes dos outros níveis a zero ou a um limiar próximo de zero

- resintetiza-se as amostras a partir dos coeficientes editados ou “limiarizados”
- compara-se os trechos resintetizados (1) e (2) de forma a validar a relevância da contribuição do(s) nível(eis) pesquisado(s) na representação do evento estudado.

O diagrama de blocos deste esquema para edição e ressíntese de trechos musicais está mostrado na Figura 31 abaixo.

Figura 31 - Diagrama de blocos do sistema para ressíntese seletiva de níveis específicos



O editor de coeficientes foi convenientemente implementado na interface gráfica (veja Figura 29) para servir aos propósitos acima. Pode-se editar os coeficientes de cada nível, atribuindo valores de atenuação (pesos) numa faixa contínua de 0 (0%) a 1 (100%), ou mesmo amplificando-os além de 100% de suas intensidades originais. A entrada do valor de atenuação (peso) pode ser feita diretamente pela escrita do numeral no campo numérico, ou deslocando-se o botão deslizante ao lado (na interface gráfica), o qual age como um potenciômetro imediatamente atualizando o valor no campo numérico (entre 0 e 1).

No capítulo seguinte apresenta-se propriamente as análises dos eventos musicais propostos para a flauta e para o violino, seguidas dos resultados, discussões e conclusões finais.

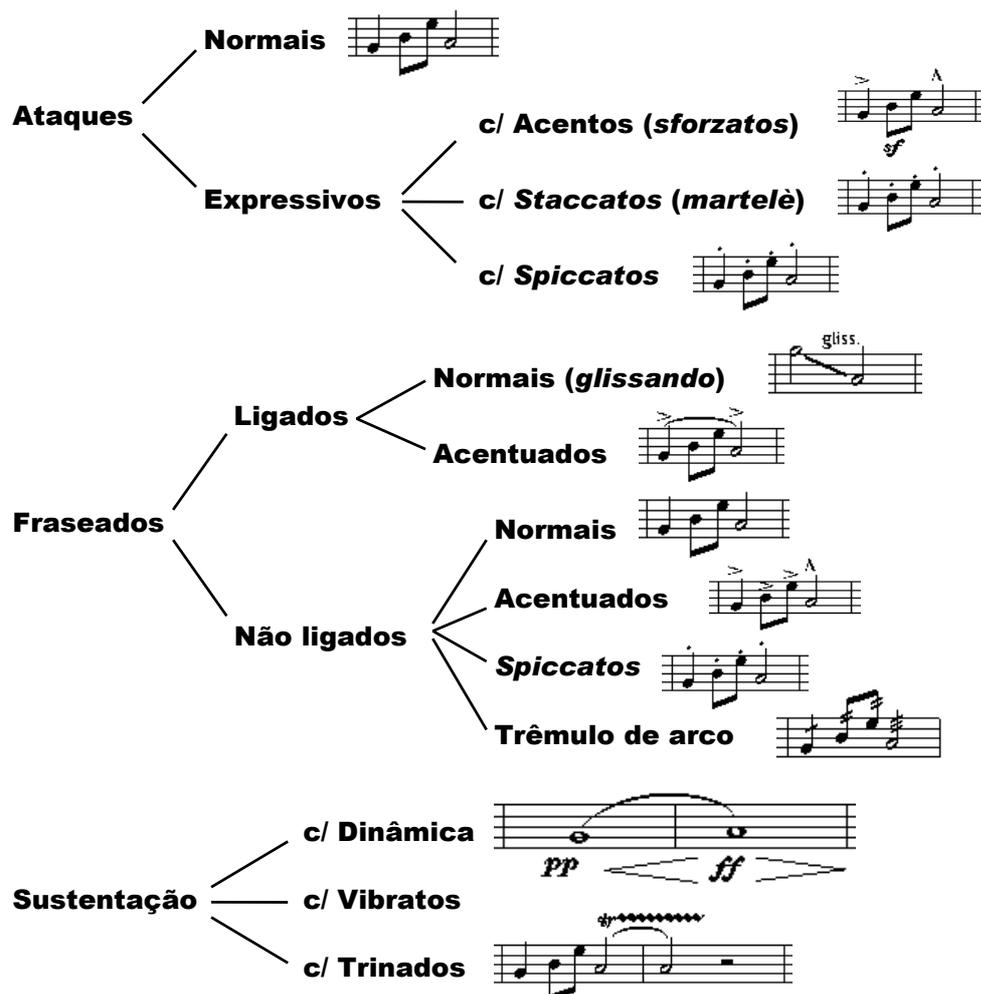
5 RESULTADOS, DISCUSSÕES E CONCLUSÕES

Na sessão de resultados apresenta-se os resultados das análises dos eventos musicais abordados. Na sessão de discussão faz-se uma apreciação dos resultados das análises, fenômenos associados, e compila-se um cruzamento de informações advindas da interpretação das mesmas. Na sessão de conclusões, tece-se opiniões, comentários finais e perspectivas futuras.

5.1 Resultados das análises

A Figura 32 abaixo resume todos os tipos de eventos musicais amostrados e analisados para o violino. São ao todo 13 eventos analisados:

Figura 32 - Eventos musicais amostrados e analisados para o violino

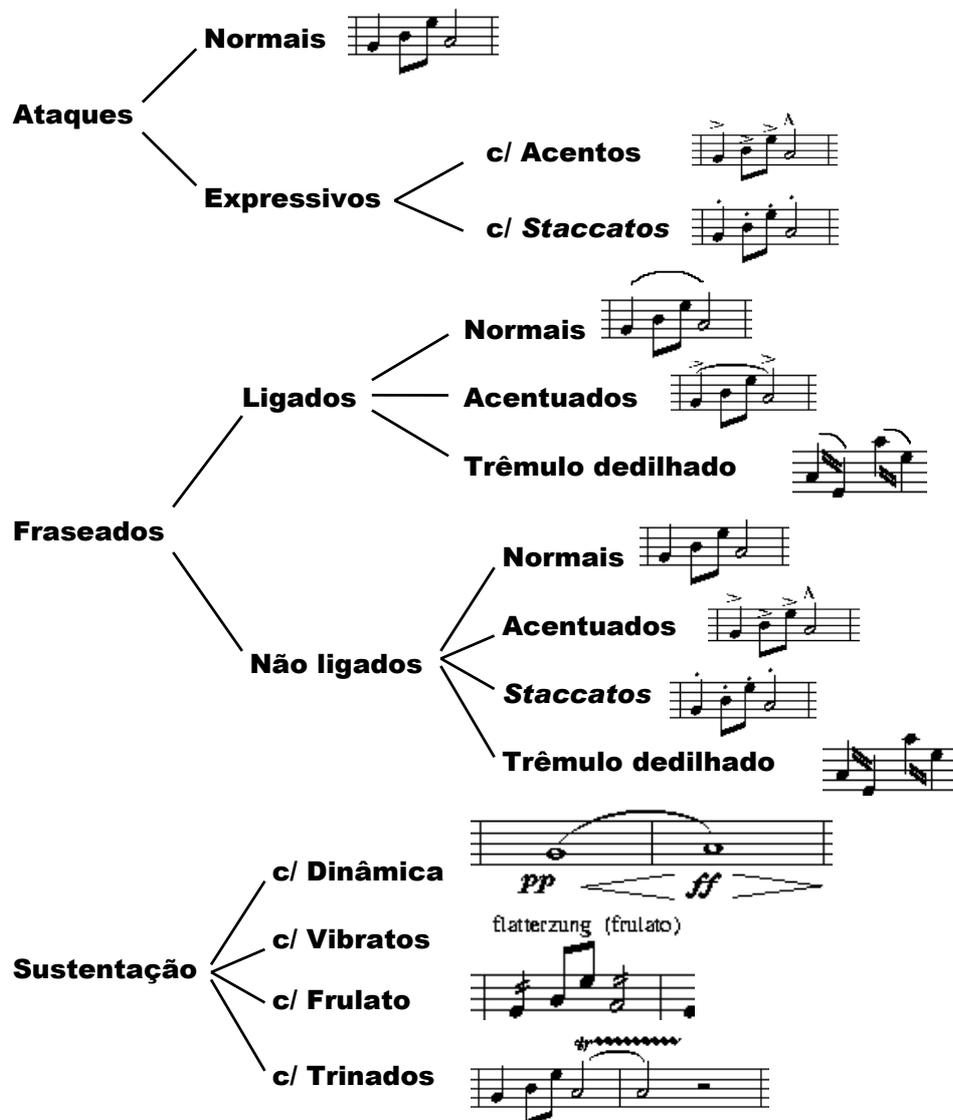


Os ataques abordam a análise dos períodos de ataques de notas individuais. Os fraseados estudam as transições entre notas embutidas num contexto melódico. As sustentações analisam um conjunto de efeitos sobre notas sustentadas.

As partituras ao lado de cada evento meramente ilustram a simbologia adotada em música para representá-los, isto é, a notação escrita para guiar a interpretação, não consistindo propriamente dos trechos musicais tocados.

Mantendo a mesma metodologia, a Figura 33 abaixo sumaria todos os tipos de eventos musicais amostrados e analisados para a flauta, ao todo 14 eventos:

Figura 33 - Eventos musicais amostrados e analisados para a flauta



A análise *Wavelet* em multiresolução (AWMR) de cada evento musical compreende:

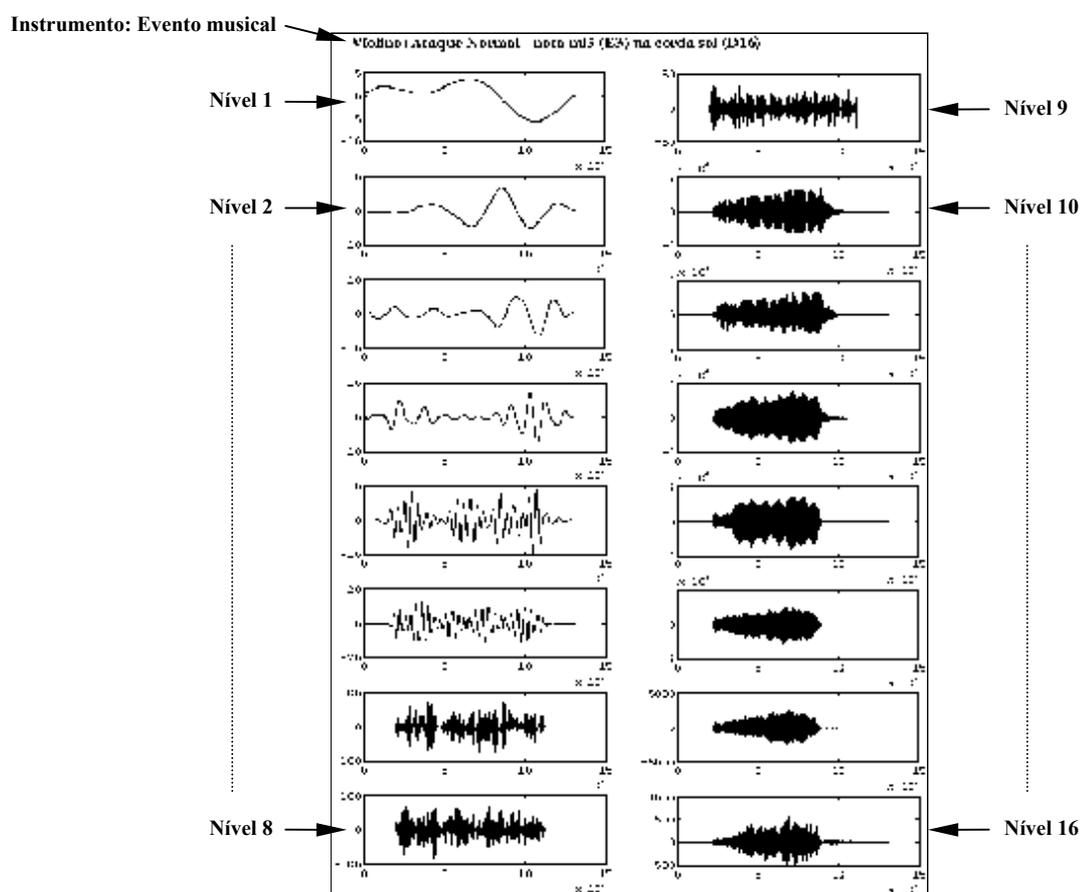
- os gráficos dos níveis *Wavelet* da análise,

- uma descrição do **evento** (uma descrição da interpretação e/ou sobre o trecho tocado),
- uma ilustração da **partitura** representativa do trecho tocado (quando aplicável),
- a representação gráfica do **som original**,
- a **extensão da amostra** (sinal musical),
- a *Wavelet* de análise,
- os **resultados individuais** por níveis (pontos mais relevantes e revelações em cada nível *Wavelet*),
- gráficos mostrando a **distribuição de energia** pelos níveis,
- e eventuais **espectrogramas** realizados.

A página inicial de cada análise consta de um **título**, que especifica o **instrumento** e o **evento musical** abordado, e dos **gráficos dos níveis *Wavelets*** da análise do evento, isto é, a projeção do sinal em cada um dos subespaços da AWMR, do **segundo nível *Wavelet* mais grosso** (nível 1) ao **mais fino** (nível N-1, onde N é o \log_2 da extensão do sinal musical). Por exemplo, para sinais de extensão 2^{17} pontos amostrais, o nível 16 será o mais fino (W_{16}) e o nível 1 (W_1), o segundo mais grosso (ou menos fino). O nível 0 não é mostrado, e corresponde à informação encerrada nos subespaços W_0 e V_0 da AWMR. O sinal nestes níveis são desprezíveis em termos de energia, e consistem basicamente na forma de onda da *Wavelet* de análise, em nada contribuindo à descrição ou formação do sinal.

A Figura 34 que segue ilustra uma página inicial de uma análise de evento musical, indicando a disposição e ordenação em que os níveis *wavelets* são plotados. O primeiro nível acima à esquerda é o nível *wavelet* 1. O último nível (abaixo à direita) é o nível mais fino (que pode ser o 17, 16 ou 15 para sinais de extensão 2^{18} , 2^{17} ou 2^{16} , respectivamente).

Figura 34 - Exemplo de página inicial da análise AWMR de um evento musical com extensão 2¹⁷



O eixo horizontal de cada gráfico mostra a extensão do sinal (número de pontos amostrais ou amostras discretas), e corresponde diretamente à dimensão temporal (44100 pontos = 1 segundo). O eixo vertical mostra a intensidade (amplitude ou magnitude) em unidades (adimensionais) em escala linear.

Da segunda página em diante apresentam-se os outros itens da análise (textos, partituras e gráficos). Usualmente o item mais longo é aquele que discorre sobre os resultados e revelações individuais de cada nível: a **análise dos níveis**. O texto é detalhado, abordando tanto aspectos genéricos - *muitas vezes comuns a outras análises, como descrições do envelope, de padrões periódicos e estruturas produzidas por gestos de interpretação, avaliações sobre a suavidade e regularidade da forma de onda, da textura, taxa de variação de subida ou descida, qualidade percebida do timbre, pitches percebidos via audição, textura sonora, quantidade e características de ruído presente* - quanto aspectos exclusivos - *como análises espectrais suplementares, caracterização de série harmônica associada, análise da evolução da forma de onda acompanhando*

sua regularidade e/ou variabilidade gráfica e sônica, correlações com gestos de interpretação, relação de transientes e objetos sônicos relevantes e comparações com outros níveis.

Ao final de cada análise, dois gráficos mostram a **distribuição de energia** do sinal pelos níveis *wavelets*; o **primeiro**, traçado em escala linear, evidencia os níveis mais energéticos do sinal musical; no **segundo**, o eixo vertical (normalizado) mostra o logaritmo da energia de cada nível sobre o logaritmo da energia total, desta forma comprimindo grandes variações de intensidade no topo do gráfico, permitindo que os níveis menos energéticos apareçam e possa-se inferir sobre a distribuição relativa e a contribuição dos níveis que não aparecem sob o escalamento linear. O escalamento neste último é portanto logarítmico.

A quantidade de informação gerada é relativamente extensa. Sua inclusão na íntegra dentro do corpo do texto quebraria a sequência e ritmo da narrativa, e produziria um efeito indesejado de torná-lo excessivamente extenso e prolixo.

Assim, por motivos de clareza, espaço e concisão, optou-se por reproduzir nesta sessão apenas uma seleção de quatro análises significativas, duas de flauta e duas de violino, as quais apresentam alguns resultados genéricos relevantes, comuns a outras análises, bem como revelações específicas importantes. Estas análises ainda oferecem subsídios para os temas que são abordados nas discussões da sessão seguinte. Ao mesmo tempo, ilustram o formato e o nível de detalhamento adotado na maioria das análises. São as seguintes as análises apresentadas a seguir:

Flauta: Fraseado, Não ligado, Trêmulo dedilhado - notas sol⁴ (G⁴) e si⁴ (B⁴)

Flauta: Sustentação com Frulato - nota fá⁴ (F⁴)

Violino: Ataque Expressivo com Acento (*Sforzato*) - nota mi⁴ (E⁴), corda lá (D¹⁶)

Violino: Ataque Expressivo com Acento (*Sforzato*) - nota mi⁴ (E⁴), corda lá (D⁴)

O material integral -*as análises de todos os eventos abordados nesta pesquisa*- está disponibilizado para consulta, armazenado em meio digital anexo (**CD-ROM** e **disquete**, em formato Microsoft Word for Windows 95 - versão 7.0). As análises de **flauta** estão no arquivo "**an_flauta.doc**", e as de **violino** no arquivo

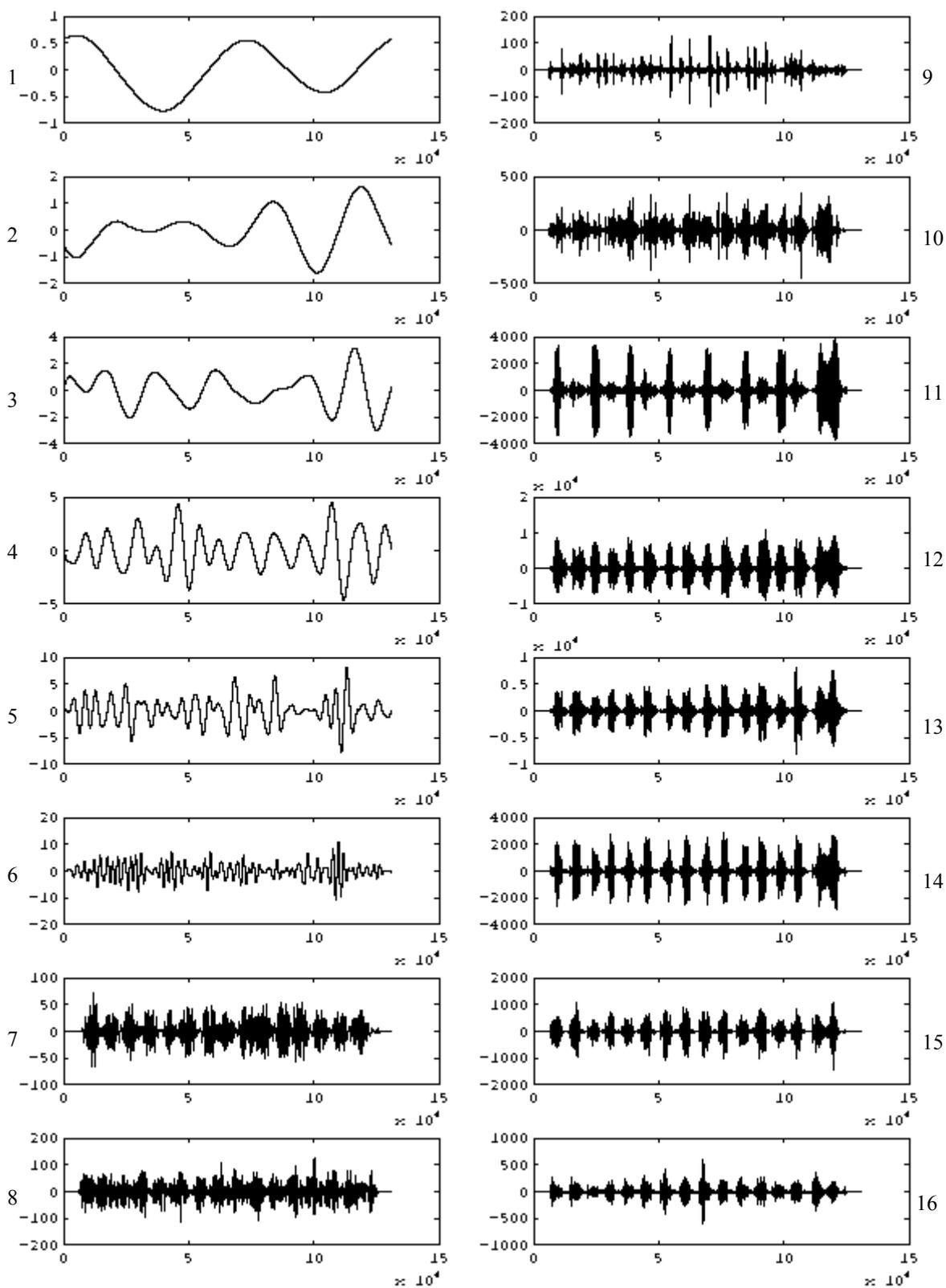
"**an_violin.doc**". Versões em formato texto (ASC-II), sem gráficos ou figuras, estão nos arquivos "**an_flauta.txt**" e "**an_violin.txt**".⁴⁴

Estes arquivos consistem num amplo material de referência para musicólogos, engenheiros de som/áudio, projetistas e desenvolvedores de tecnologia musical interessados em conhecer os detalhes da análise *wavelet* de sinais musicais, avaliar sua aplicabilidade em projetos comerciais e/ou científicos específicos, e elaborar um roteiro de especificações iniciais para nortear explorações mais profundas, focalizadas, e orientadas à concepção de produtos que utilizem a tecnologia.

Adicionalmente, inclui-se uma cópia do pacote WaveLab (versão 7.01 para MATLAB, gentilmente cedida pela sua equipe de desenvolvimento), do módulo **Wavesynt**, desenvolvido para análise AWMR para sinais musicais (1D), e das amostras musicais preparadas (contendo os eventos/gestos musicais de flauta e violino analisados) em formato de som AIF e em formato MAT (MATLAB, já preparadas para análise pelo **Wavesynt**). O leitor/usuário do MATLAB que então desejar reproduzir as análises realizadas, avaliar seus resultados, realizar reconstruções seletivas, sonorizar os níveis *wavelets* e utilizar as capacidades e funções do WaveLab e **Wavesynt** sobre seus próprios arquivos sonoros terá então totalmente à sua disposição o ferramental e material sonoro básico necessários.

⁴⁴ Sendo arquivos extensos, estarão comprimidos no formato **.zip**, do utilitário para DOS pkzip/pkunzip (compactador/descompressor), ou da sua versão para Windows, o Winzip.

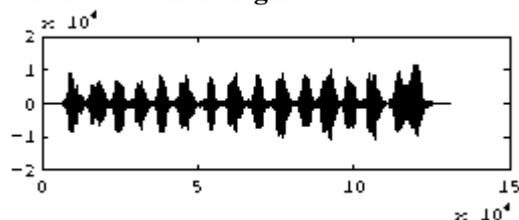
Flauta: Fraseado, Não ligado, Trêmulo dedilhado - notas sol4 (G4) e si4 (B4)



Evento: Fraseado não ligado: trêmulo dedilhado. Exemplo consiste na execução de um trêmulo com as notas sol (G4) e si (B4) da segunda oitava da flauta, executados em tempo rápido para se caracterizar a natureza das transições e da emissão desses tons num fraseado tremulado. O trêmulo é do tipo "trêmulo de dedilhado" consistindo na interpretação de duas notas com intervalo maior que 1 tom.

A última nota do exemplo possui uma identidade diferenciada, com uma leve modulação aplicada alongando-a, e produzindo um final dinamicamente mais relevante.

Gráfico do sinal original:



Extensão da amostra: 2^{17} pontos (131072). **Wavelet de análise:** D16

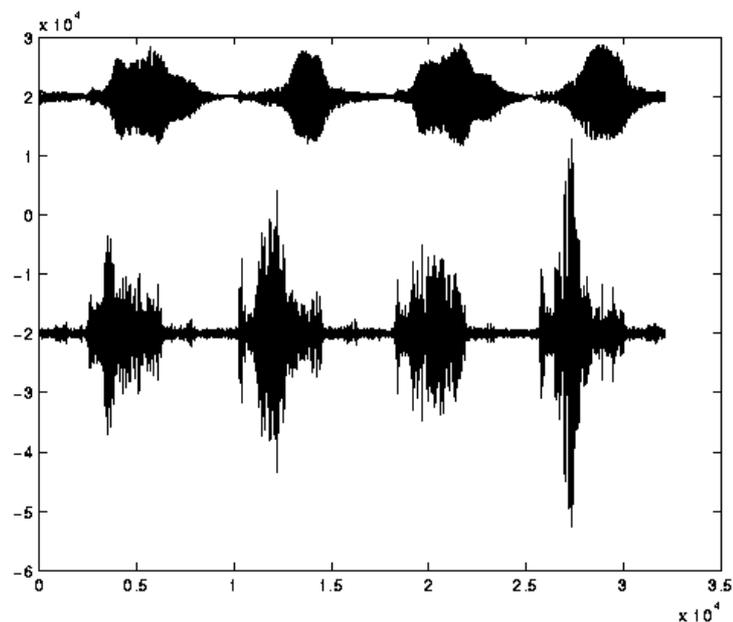
Análise dos níveis:

Nível 16: Percebe-se bastante modulação de sopro, isto é, o sopro é fundamental na formação do envelope do trêmulo não ligado. O ciclo do sopro envolve o ataque e o colapso forçado sucessivos. O nível mais fino portanto é importante para se caracterizar trêmulos não ligados, onde observamos a presença do ciclo completo do sopro, do ataque ao colapso forçado, em tempo rápido.

Uma inspeção visual mais próxima (magnificação) de algumas notas mostrará o envelope característico do ataque em fraseados não ligados, com uma rampa de subida íngreme, uma sobrelevação seguida de um decaimento.

Logo em seguida vem a região de sustentação com altas intensidades, superiores às da sobrelevação, e, em seguida, pela diminuição da pressão de sopro devido ao avanço da língua sobre o céu da boca reduzindo o sopro, uma queda progressiva nas amplitudes, entrando na região de colapso. A interrupção completa do sopro, que é cíclica no trêmulo não ligado, provoca o decaimento súbito ao final da região de colapso, conforme podemos observar na Figura 35, mostrando, para um mesmo intervalo de tempo (4 notas), o sinal original (sinal de cima) e o sinal no nível 16 (sinal de baixo).

Figura 35 - Parte do sinal original (topo) e do sinal ao nível 16 (abaixo)



Os "nódulos inter-notas", isto é, as pequenas elevações situadas entre as notas, correspondem aos gestos bucais que ocorrem quando temos um movimento de língua rápido avançando e retraindo em relação ao céu da boca, no sentido de se modular a passagem de ar e assim o fluxo de sopro. A pressão de sopro é máxima no centro das notas, onde a intensidade é mais alta.

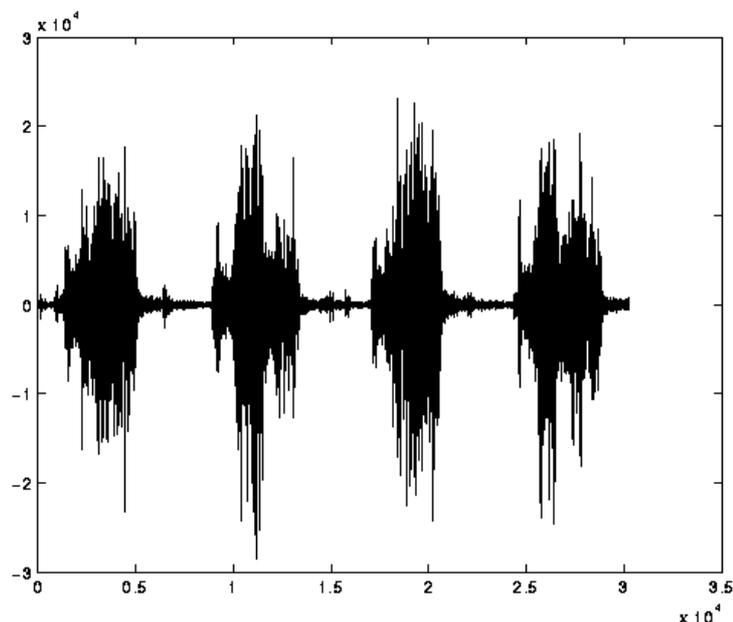
Nível 15: Verificamos que os envelopes das notas sol4 (G4) possui uma depressão de amplitudes ao centro da região de sustentação. Previamente ao colapso, as amplitudes crescem novamente e então caem subitamente em virtude do colapso forçado.

Estruturas intermediárias, isto é, pequenos nódulos ou elevações ainda são vistas entre as notas, embora bem mais sutilmente do que no nível anterior.

Para as notas si, observamos o ataque característico, com sobrelevação e decaimento, e em seguida um progressivo aumento de amplitudes em direção ao colapso, quando então caem abruptamente. Os contornos da forma de onda, como no nível anterior, não são suaves: são irregulares e com vários picos locais (de máxima amplitude).

A Figura 36 mostra uma parte magnificada do nível 15, entre as 6^a e 9^a notas inclusive, uma sequência de si-sol-si-sol (entre 0.944 e 1.630 segundos).

Figura 36 - Parte do sinal ao nível 15: intervalo entre 0,94 s e 1,63 s (da 6ª à 9ª notas)



Textura sônica ruidosa, como o nível anterior, mas apresentando já alguma contribuição das séries harmônicas das notas do trêmulo.

Nível 14: Primeiro nível que permite uma melhor percepção das tonalidades. Os envelopes das notas si4 (B4) diferem em forma dos envelopes das notas sol4 (G4), estes sendo ligeiramente mais suaves, com menos irregularidades e transientes, possuindo ataques discretos, e uma flutuação ou modulação de baixa frequência sobre as amplitudes na região de sustentação, com uma progressiva elevação que culmina nas maiores elevações para cada nota (sol4) do centro em diante, na direção do colapso.

Já as notas si4 apresentam ataques mais discretos, embora bem caracterizados, não apresentam flutuações de baixa frequência na amplitude após o decaimento, e sim as amplitudes crescem rapidamente e mantém uma média intensa por toda a sustentação, decaindo após no colapso.

O colapso para ambas as notas consiste numa queda de elevações menos abrupta do que nos níveis anteriores, e são estendidos por uma pequena reverberação ou eco, como num pulso após o corte da excitação.

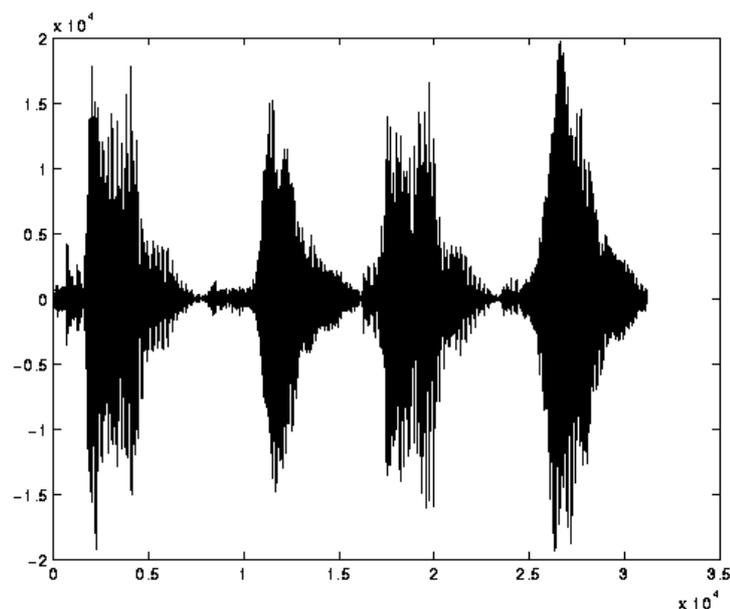
Ouvidas a 11025 Hz as duas notas apresentam diferenças nas regiões transitórias, do ataque, previamente à região estacionária. As notas sol4 apresentam um transitório mais complexo até estabelecerem o tom correto e soá-lo com regularidade. Esta complexidade no ataque implica num maior tempo necessário para atingir a zona de sustentação, e explica-se em parte pela dificuldade inercial de se excitar uma nota mais baixa (maior coluna de ar) no tubo vibrante. O fato de termos um ciclo de interrupção do sopro a todo momento, as vibrações que se processam no tubo estão sofrendo transitórios constantes, e as zonas de vibração estacionárias são mais curtas, especialmente se o tempo da interpretação for rápido.

Nível 13: As notas si apresentam uma modulação em frequência facilmente constatada por inspeção visual do gráfico, fruto do processamento da análise *Wavelet* em multiresolução. Também seus envelopes são mais irregulares, com ataques discretos, zonas de sustentação com envelopes quase quadrados, com subidas e descidas íngremes, e uma região de colapso mais longa, que decai gradualmente. Neste nível o colapso se mistura à região de sustentação, seu

início podendo ser considerado uma segunda parte daquela região onde as amplitudes são menores.

A Figura 37 ilustra o intervalo cobrindo da sexta à nona notas (si-sol-si-sol) mostrando os contornos e aspectos da forma dos envelopes. Os envelopes das notas sol apresentam ataques mais discretos, com pequenas elevações, e uma região de sustentação mais intensa, com as amplitudes gradualmente reduzindo-se em seguida, misturando-se à zona de colapso.

Figura 37 - Parte do sinal ao nível 13: da 6ª à 9ª notas (si-sol-si-sol)



Nível 12: Nível mais energético. Boa sonoridade, nível concentra energia dos parciais harmônicos. As sobrelevações nos ataques das notas si4 apresentam amplitudes destacadas no exemplo, contrastando com o envelope das notas sol4 (aproximadamente com um formato de gaussiana), com elevação e queda variando mais suavemente, de forma gradual.

Nível 11: As notas si4 apresentam intensidades bem menores que as notas sol4, em média 6 vezes menos intensas, assim como ocorreu no exemplo com fraseado ligado com trêmulo dedilhado.

Observando o gráfico do nível da sexta à oitava notas, percebe-se que as regiões centrais das notas sol4 (mais baixas em frequência) são as mais destacadas neste nível, e que as notas si4 apresentam formas de onda de envelopes amorfos, contornos não suaves.

Outro fato a ser observado, inclusive em níveis anteriores, é a não observância de silêncios completos entre as notas, em todos os níveis até o presente. Se o trêmulo fosse mais lento poderíamos visualizar momentos de pausa, uma vez que o tubo ressonante teria tempo o suficiente para colapsar todas as vibrações que nele trafegam.

Nível 10: Alguns dos picos de máximos no nível situam-se sobre as zonas centrais de ambas as notas do trêmulo. No entanto, existem também transientes associados às zonas de transição entre notas e ausência de máximos de amplitude pontiagudos sobre algumas notas. Essa irregularidade sugere que os picos estão rastreando eventos musicais e/ou gestos transitórios que não estão aparentes ou facilmente identificáveis no sinal original.

Os eventos ou gestos em questão seriam então aperiódicos, isto é, produzidos de forma aleatória, sua existência estando condicionada à natural manipulação do instrumento, a eventos

de respiração e movimentos bucais aperiódicos, e não aos processos vibratórios correlacionados aos modos ressonantes do tubo vibrante. Dado à textura já ruidosa e grave do som do nível (que cobre faixas de frequência mais baixas), é de se esperar que tais gestos impliquem em contribuições relevantes de frequências mais baixas.

Estudos mais específicos serão necessários para investigar a natureza dos picos de máximo associando-os a gestos específicos na interpretação do instrumento.

Nível 9: Sinal não suave e de contornos irregulares, mas apresentando certa periodicidade na ocorrência de transientes ou pulsos que se distribuem regularmente por todo o exemplo, com capacidade de localização (isto é, não extrapolam os limites do sinal original real invadindo a região de *zero padding*).

Os pulsos localizam as zonas centrais das transições entre notas e das regiões de sustentação, onde encontram-se pontos críticos associados à intermitência do sopro, quais seja os de máxima intensidade e pressão (quando a língua permite a passagem do maior fluxo de ar) e os de intensidade mínima ou nula, onde a língua interrompe por curtíssimo período de tempo o fluxo de ar.

Os estalidos e outros ruídos bucais associados aos gestos bucais da interpretação são os eventos mais prováveis de estarem visíveis neste nível, uma vez que são periódicos mas exibem frequência de periodicidade inconstante exatamente por serem processos sincronizados pelos ritmos naturais dos movimentos do intérprete. Este nível portanto revela-se mais útil que o anterior para rastrear tais gestos.

Cabe notar entretanto que em audições acústicas reais, onde vários instrumentos são tocados num ambiente adequado, tais gestos não produzem ruídos perceptíveis porque seriam fortemente atenuados pela impedância acústica do ambiente real, e também mascarados pela contribuição mais energética das vibrações harmônicas dos instrumentos. Só microfones com capacidade de captação direcional, como o que foi utilizado, podem revelar tais ruídos característicos da manipulação do instrumento ou ruídos de origem bucal.

Nível 8: O contorno do envelope do nível consiste numa aproximação ao envelope original: depressões ou estrangulamentos no sinal original que ocorrem nas zonas entre notas, e elevações correspondentes às zonas centrais das notas também ocorrem no sinal deste nível. Como o nível é baixo, já representa de certa forma uma média do sinal original projetada num nível de resolução específico onde as flutuações de larga escala do sinal original são mais visíveis. Em outras palavras, as frequências cobertas por este nível integram o sinal original e contribuem no delineamento ou conformação do envelope de larga escala do sinal original (mesmo tratando-se de nível pouco energético).

As notas entretanto neste nível não soam mais diferentes e seus envelopes são preenchidos por um conteúdo de frequências comum a todas as notas, isto é, um mesmo tom é modulado em amplitude pelo envelope do sinal original. Dois *pitchs* ou bandas de frequências preponderantes são encontradas no espectro do nível: 46 Hz e 40 Hz, a diferença entre estas raias de aproximadamente 6 Hz corresponde a grosso modo à periodicidade das notas, que ocorrem numa frequência de aproximadamente 6 por segundo.

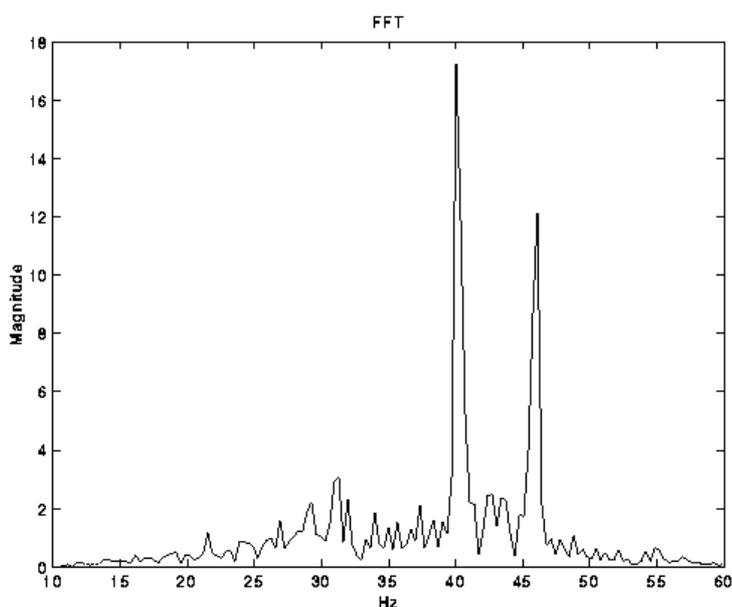
Portanto, quando temos uma interpretação bem periódica, com precisão no respeito à grade de tempos e de duração das notas - *especialmente se a partitura consiste em seqüências de notas de mesma duração* - a frequência de produção de notas aparecerá como simples diferença entre os raios do espectro deste nível. Se a frequência for inconstante, então a sua extração por

intermédio deste simples cálculo (o do batimento entre os *pitchs* preponderantes do nível) não será eficaz, e técnicas mais sofisticadas serão necessárias.

Somente com esta sequência de 15 notas é que primeiro pudemos descobrir a sensibilidade deste nível a periodicidades na interpretação de notas, agindo a grosso modo como um **contador de notas**.

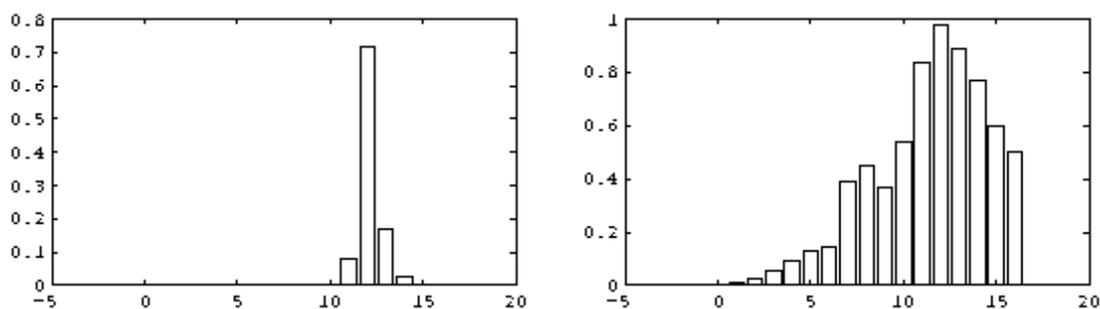
Nível 7: Também revela periodicidade, contando notas, com picos do espectro em 40 Hz (mais intenso) e 46 Hz, exatamente as mesmas frequências do nível anterior. O espectro do nível está na Figura 38 abaixo.

Figura 38 - FFT do nível 7 para a análise de fraseado não ligado, trêmulo dedilhado



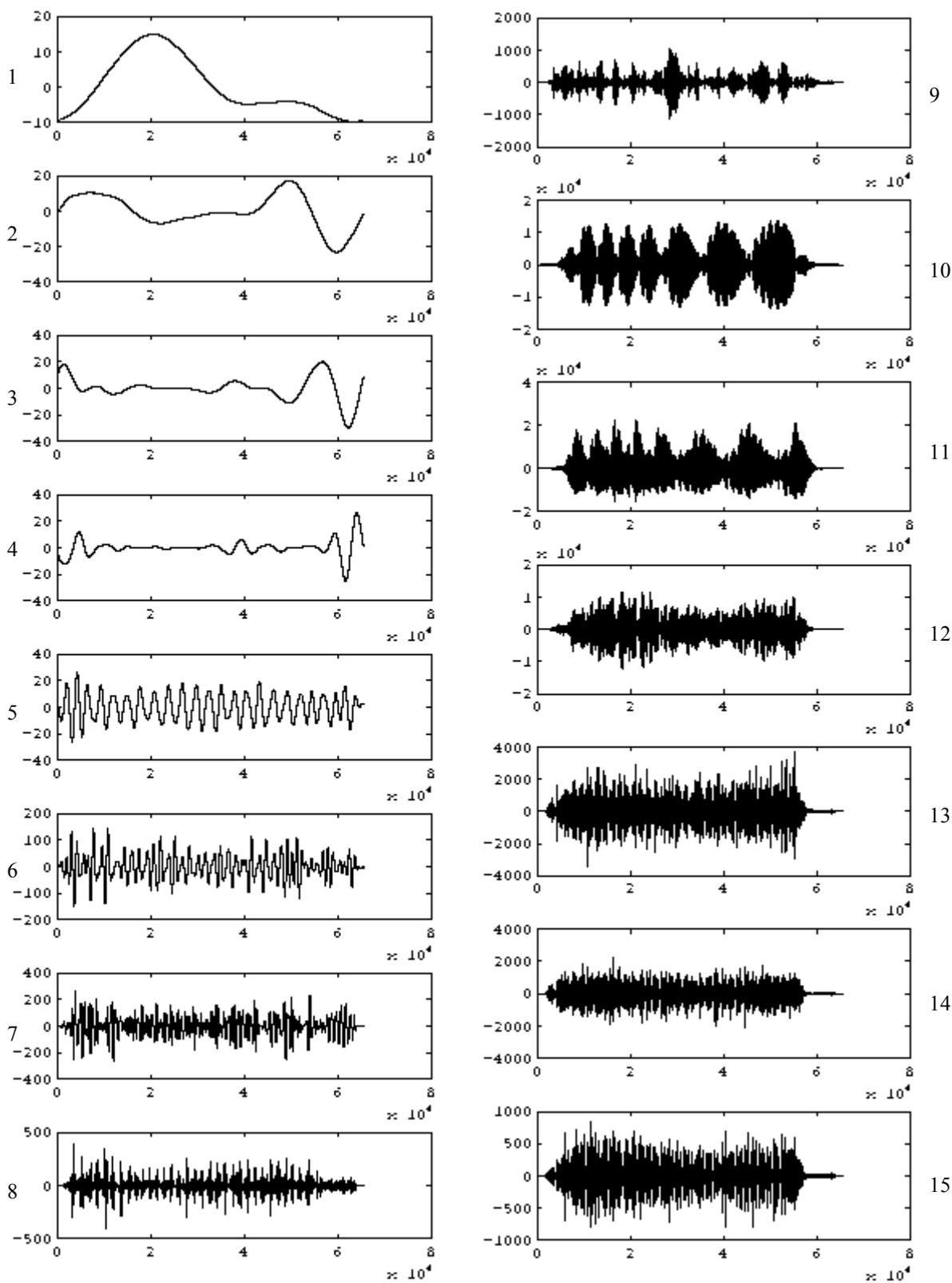
Nível 6: Inaudível (exceto pelos *clicks*), se espalha pela zona de acolchoamento, não permite localizar com precisão eventos em relação ao sinal original

Distribuição de energia:



Mais de 99% de toda a energia do exemplo se concentra nos níveis de 11 a 14, sendo o nível 12 o mais energético.

Flauta: Sustentação com Frulato - nota fá4 (F4)

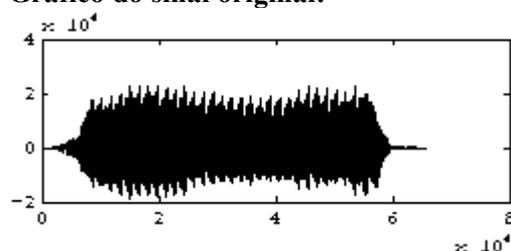


Evento: Sustentação com frulato. Exemplo de um nota executada com efeito de frulato, desde o ataque até o final da nota. Ataque em frulato do tipo "tu" à nota fá4 (F4) e sustentação do efeito.

Ouvido a 11025 é notável o efeito: inicialmente ouve-se o sopro e logo em seguida, quando quebra-se a inércia da coluna de ar, estabelecem-se os padrões estacionários, que são periodicamente modulados por um jato intermitente de sopro. O efeito de intermitência é produzido pela vibração da língua contra o céu da boca, interrompendo o fluxo de ar periodicamente. A cada ciclo as vibrações estacionárias tendem a colapsar, e o retorno do fluxo de ar ocorre antes que morram completamente. No entanto, a cada retorno do fluxo de ar o ruído de sopro é novamente introduzido, e temos um realce das componentes do sopro a cada ciclo.

Uma análise FFT da nota mostra a presença de uma série harmônica predominante, e obscurece o fato de que há um frulato executado na nota. Três raias se destacam: o **segundo harmônico do fá3** (F3) que na verdade é a fundamental desta nota F4 (em 691 Hz, amplitude 5456 unidades), o **quarto harmônico** (em 1387,5 Hz, amplitude 1616 unidades) e o **sexto harmônico** (em 2100,8 Hz, amplitude 472 unidades). Há portanto uma predominância dos parciais pares, no entanto as raias são relativamente mais largas comparadas às notas interpretadas sem o frulato, com picos locais distribuídos ao redor das frequências centrais discriminadoras da raia, acusando a presença de bandas laterais de uma modulação de anel, resultado dos desvios em frequência provocados pelo frulato. A frequência com que o frulato é executado não é entretanto de fácil verificação no espectro da nota - por exemplo, pelo menos duas frequências moduladoras diferentes podem ser verificadas na FFT entre 650 e 750 Hz.

Gráfico do sinal original:



Extensão da amostra: 2^{16} pontos (65536). **Wavelet de análise:** D16

Análise dos níveis:

Nível 15: O ataque consiste numa região de subida gradual até um ponto máximo do envelope. Na ponta inicial a presença do sopro é predominante. No decorrer da nota, no período de sustentação, observa-se uma textura pouco suave no envelope, com muitos picos localizados de amplitude e uma sequência de pulsos periodicamente distribuídos pela nota até a região do colapso, quando cessa a modulação do frulato e dá-se início ao colapso propriamente.

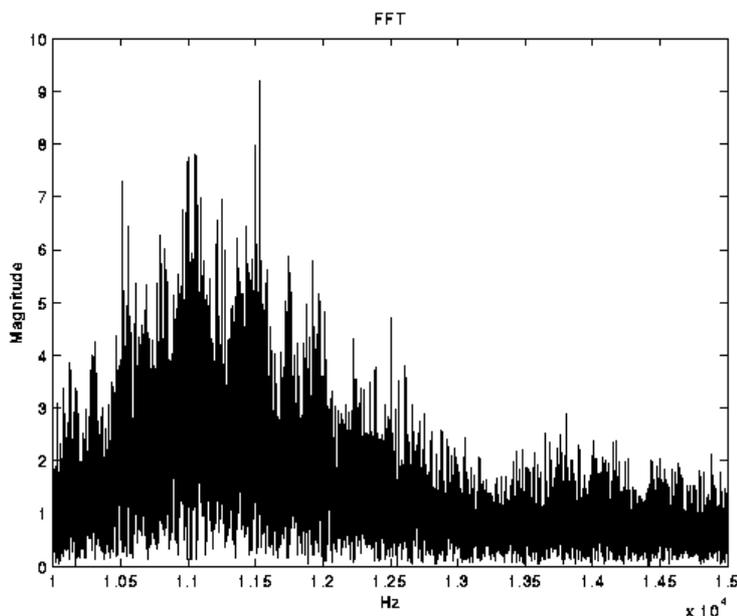
O colapso revela-se neste nível como um "apêndice" com amplitudes baixas, consistindo num ruído de fundo na sua maior parte.

O espectro do nível revela uma densa população de frequências acima de 5000 Hz até 22050 Hz, com picos duplos resultado de uma modulação de anel (modulação de amplitude sem portadora) gerada pelo efeito de frulato. O pico mais amplo está em 11535 Hz e sua banda lateral espelhada se encontra 42 Hz abaixo, em 11493 Hz, evidenciando uma modulação de amplitude de 21 Hz, a frequência estimada para o movimento lingual que oriunda o frulato.

A mesma frequência é extraível da segunda raia mais ampla, com pico principal em 11046 Hz, e sua banda lateral inferior espelhada, também a 42 Hz abaixo, em 11004 Hz, novamente

mostrando a frequência do ciclo intermitente de sopro do frulato. A FFT do nível está ilustrada na Figura 39, mostrando o espectro entre 10000 e 15000 Hz, onde se encontram as raiais principais do som neste nível, e pode-se visualizar a modulação de anel implícita citada acima, com picos gêmeos separados de 42 Hz.

Figura 39 - FFT do nível 15 (sustentação com frulato)



A análise do espectro do som original não torna evidente a presença da frequência do frulato, e o sinal no nível 15, tipicamente um nível que realça as componentes do sopro, é capaz de clarear os aspectos cíclicos envolvidos no sopro intermitente do frulato.

O som do nível revela a periodicidade do sopro do frulato, tornando-se sua principal característica. Nota-se que a periodicidade é estabelecida desde o início da nota, caracterizando o ataque de frulato a partir do *onset* da nota. Somente na região do colapso o frulato é interrompido. O som lembra um chocalho de cascavel quando reproduzido em 11025 Hz.

Um nível mais elevado de energia para este nível mais fino revela que grande parte da energia do frulato fica esboçado no sopro.

Nível 14: O envelope do nível se assemelha ao do nível anterior, mostrando entretanto um nível de ruído de fundo menor (visível pelo apêndice do colapso). O som do nível é semelhante ao de um chocalho de cascavel.

A análise espectral estática do nível mostra que ele ainda representa uma significativa concentração de componentes do sopro, dada a ampla distribuição de frequências por uma faixa larga do espectro acima de 3000 Hz.

Nível 13: A presença do ruído de sopro ainda é forte como pode-se verificar pela FFT (espectro estático) do nível, e a periodicidade do ciclo intermitente do frulato é a mesma calculada nos níveis anteriores. O espectro (não mostrado) revela uma densa população de frequências entre 2000 e 8000 Hz principalmente, mas raiais "mais comportadas" já aparecem claramente exibindo um padrão de série harmônica. O pico máximo do espectro se dá em 3411,7 Hz, e entre 3300e 3600 Hz aparecem várias raiais onde a decifração visual da frequência do frulato associada se torna mais complexa.

Ao redor de $1/16$ da frequência de amostragem (2756,25 Hz) encontramos bandas laterais espelhadas, com ligeiras diferenças em amplitudes, um efeito de duplicação de bandas causado pelo algoritmo de análise *Wavelet* em multiresolução⁴⁵.

A audição do nível sugere que há um *pitch* predominante e a frequência do frulato é aproximadamente a mesma dos níveis anteriores. Cálculos da distância entre picos, numa tentativa de estimar a frequência modulante associada ao frulato, revelam valores entre 18 e 27 Hz de frequência para o frulato. A estimativa da frequência de frulato pela simples análise visual dos picos e raias do espectro do nível é, entretanto, praticamente inviável, requerendo para tal o uso de técnicas mais complexas, como por exemplo o emprego de análise *pitch*-síncrona (*pitch synchronous*) sobre o sinal do nível, como Evangelista (1993) realizou em seus experimentos (veja citação de seu trabalho nas referências bibliográficas da dissertação).

A audição do nível a 11025 Hz revela que os componentes harmônicos do tom (vibrações estacionárias) surgem após 3 ou 4 ciclos do frulato, e que permanecem presentes até o fim da nota, flutuando sob a modulação do frulato, conforme já dito anteriormente.

Nível 12: O envelope deste nível revela duas características novas:

1. O ataque se processa numa ascensão gradual de amplitude
2. Existem duas regiões de máximos e uma de mínimo (amplitudes relativas) na fase de sustentação. O efeito entretanto sobre a amplitude percebida do som é desprezível, e até observa-se que a região de menores intensidades na sustentação é portadora de um som mais puro e regular.

A visualização da forma de onda revela que a frequência modulante do frulato no nível pode ser na verdade uma combinação de frequências modulantes, uma vez que pulsos perfeitamente distanciados não são verificados, e sim verifica-se a existência de uma modulação mais complexa. A audição do nível a 11025 Hz revela um início soprado seguido, após alguns ciclos, pela ascensão dos parciais harmônicos. Na região do colapso também verifica-se uma ressonância após o fim da excitação (sopro) produzindo uma espécie de eco ou reprodução menos intensa do último ciclo do som.

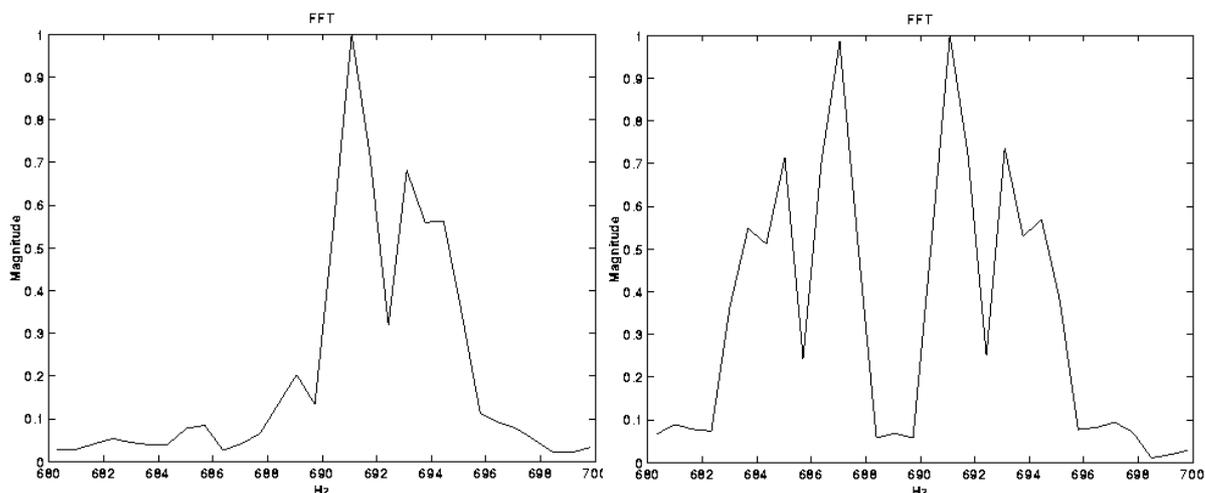
Nível 11: A inspeção gráfica da onda deste nível revela pelo menos duas regiões com diferentes modulações de amplitude, separadas aproximadamente ao meio da nota. De fato, o que se verifica é que a frequência modulante simples do frulato, calculada anteriormente em aproximadamente 21 Hz, não se aplica a todos os níveis, o que sugere a ocorrência de diferentes modulações de amplitude com diferentes frequências modulantes aplicadas a cada faixa do espectro audível. No nível 11 por exemplo nenhuma frequência modulante aparece com destaque, como 21 Hz o fez no nível 15. O espectro do nível basicamente revela as mesmas bandas existentes no espectro do sinal original, a única diferença está no aparecimento das raias espelhadas em torno de frequências que são razão de números inteiros da frequência de amostragem ($f_a=44100$ Hz): $1/64.f_a$ (~689 Hz) e $1/32.f_a$ (~1378 Hz). Essa modulação, que tem origem no algoritmo usado para a análise *wavelet*, é a causa da modulação em amplitude, tanto visualizada no gráfico da onda quanto percebida via audição.

A Figura 40 mostra a primeira banda espectral relevante do sinal original e do sinal ao nível 11, respectivamente (espectros normalizados). Note que o pico em 691 Hz (fundamental da nota)

⁴⁵ Leia discussão sobre este efeito no capítulo 5, item 5.2 da dissertação.

aparece em ambas as figuras, mas ao nível 11 ela aparece espelhada em torno de $1/64 \cdot f_a$ (689,0625 Hz).

Figura 40 - Primeira banda relevante da FFT do sinal original (esquerda) e do nível 11 (direita)



A análise auditiva do sinal no nível 11 mostra um sinal mais puro e desprovido de ruídos, com uma só voz percebida, e uma modulação em amplitude percebida como um batimento com frequência em torno de 9 Hz para uma primeira região (aproximadamente de 0 a 680ms) e 4,5 Hz para uma segunda região (de 680ms ao final).

Nível 10: O nível apresenta características gerais semelhantes ao nível 11, como a fácil audição e visualização de diferentes frequências modulantes em forma de batimento; entretanto apresenta melhor suavidade, "pureza" de tom e o envelope possui contornos mais "contínuos". O batimento verificado no nível anterior fica muito mais explícito neste nível, onde as cristas e vales da onda ficam melhor caracterizados.

Uma análise mais cuidadosa da forma de onda apresentada no nível (e da modulação em amplitude em forma de batimento que molda seu envelope) mostra que existe uma variação contínua da frequência percebida de batimento. No início da nota esta frequência é máxima, e decresce a medida que a nota progride e o efeito do frulato se estabiliza. Isto sugere **(1)** que o regime permanente de um tom emitido sob frulato requer algum tempo para se estabilizar, ou **(2)** que o gesto musical de frulato evolua, alterando seu padrão modular no tempo. O batimento mais rápido ao início mostra que os modos harmônicos (representados neste nível) oscilam em amplitude respondendo ao ciclo de excitação e corte (sopro e ausência de sopro) do frulato. Esta frequência de oscilação tende a diminuir a medida que os modos vão ganhando energia em cada ciclo, e podem portanto manter suas oscilações por mais tempo. Esse comportamento pode ser vital para se produzir um frulato sintetizado com qualidade bem próxima ao natural.

Nível 9: Cerca de 600 vezes menos energético que o nível anterior, o nível 9 apresenta um envelope também diferenciado daqueles dos níveis anteriores, e seu som não guarda semelhança com os sons de níveis anteriores.

O som do nível não revela a preponderância dos modos harmônicos do tom, como nos níveis mais energéticos. O envelope não é suave, apresenta várias "descontinuidades", revela ataques e colapsos semelhantes aos dos níveis anteriores, com início e final graduais. A presença de uma modulação de amplitude em forma de batimento é entretanto uma característica impressa devida ao frulato, e inexistente de forma tão regular em notas interpretadas com outras técnicas.

Nível 8: A presença da frequência característica do frulato se faz novamente relevante agora no nível 8, muito menos energético que os anteriores. O envelope do nível revela inúmeras "descontinuidades", e o colapso, onde o frulato é cessado, não apresenta os picos localizados que se distribuem por toda a nota onde o frulato existe - inclusive no início, uma vez que a nota foi atacada sob frulato.

O som é característico dos níveis baixos menos energéticos, e realça os componentes ruidosos de baixa frequência que existem no som original, sejam eles oriundos de ruídos ambientes, das fases de gravação, conversão e preparo das amostras, seja oriundo de vibrações graves excitadas nos instrumentos durante a emissão dos tons, ou de gestos musicais devido a manipulação de arcos, sopros, respiração e dedilhados. O espectro do nível revela uma densa população de frequências entre 50 e 250 Hz, com picos em 109,0 e 134,6 Hz.

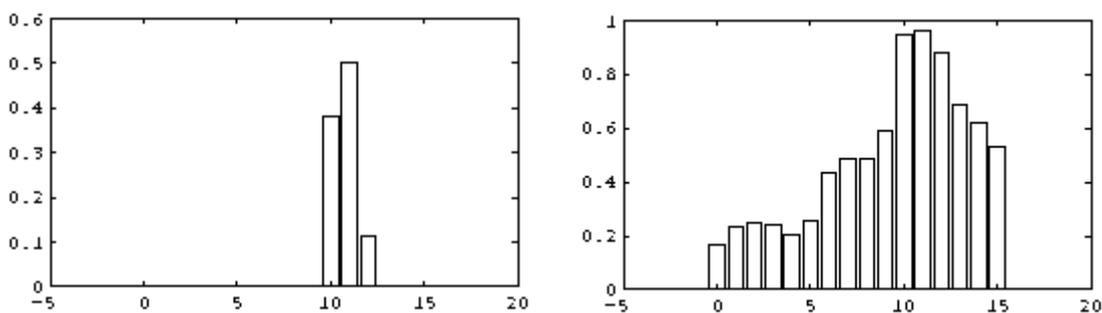
Nível 7: Não revela aspectos importantes do sinal musical, e guarda semelhança com os sinais de outras notas sem frulato.

Nível 6: Espectro revela pico de frequência em 27,59 Hz, seguido por outro em 40,37 Hz, frequências muito baixas. Não há contaminação de 60 Hz.

Nível 5: Inaudível (exceto pelo *click* do início e fim). Mostra um onda contínua, com frequência (*pitch*) central em 15,47 Hz, e um segundo pico local em torno de 20 Hz, um valor muito próximo à frequência de frulato estimada no nível 15. Não pode-se todavia afirmar que a frequência modulante do frulato fica evidente neste nível.

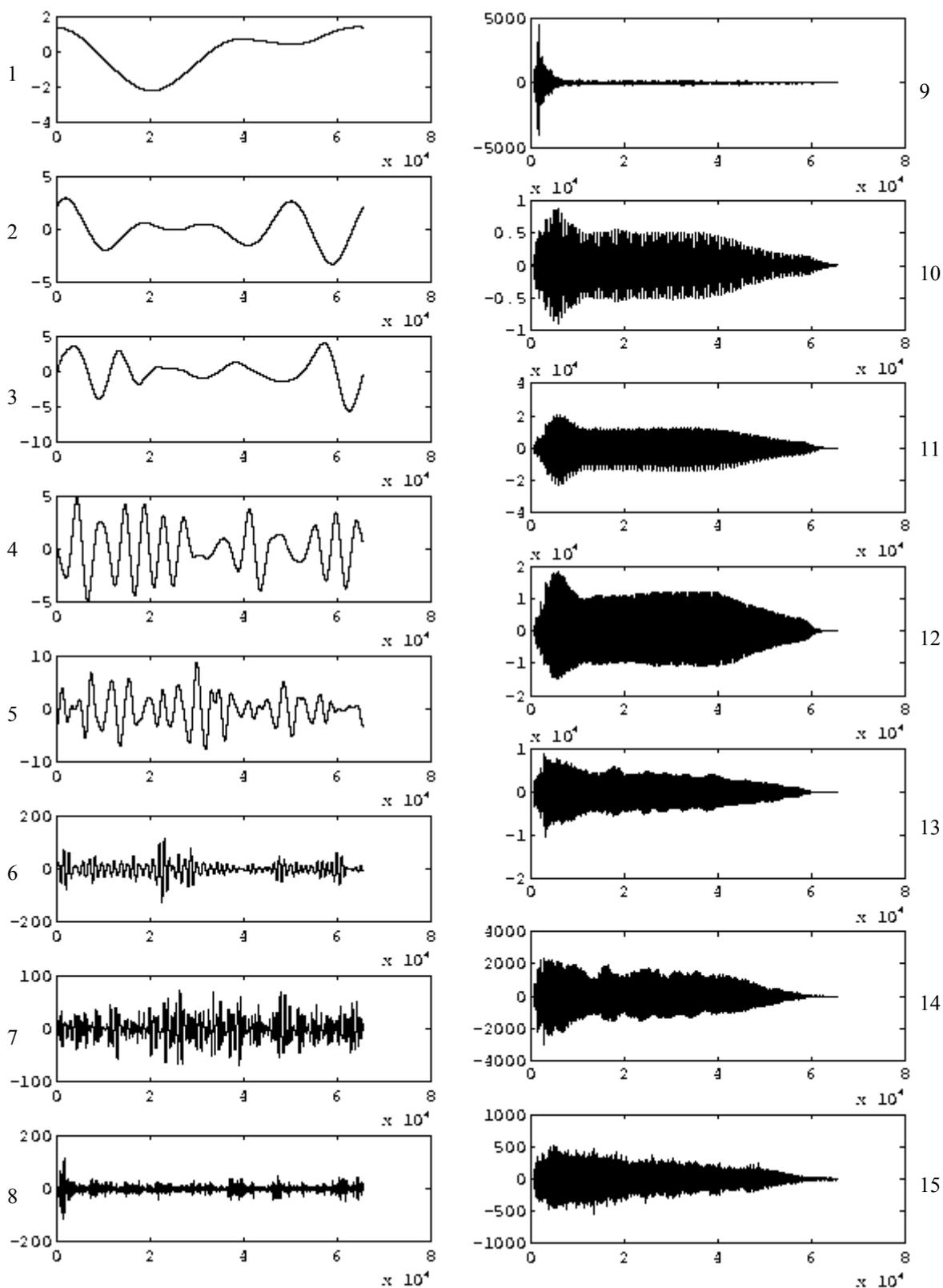
Do nível 4 para baixo as formas de onda não exibem as flutuações periódicas observadas nos níveis superiores, e não mostraram-se importantes para a verificação do frulato.

Distribuição de energia:



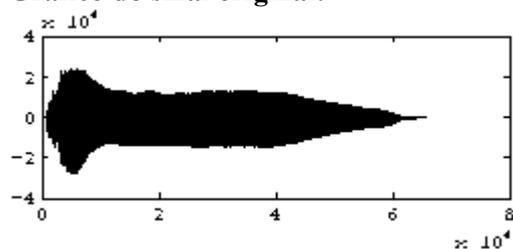
Quase toda a totalidade da energia do exemplo se concentra nos níveis de 10 a 12, sendo o nível 11 o mais energético (com aproximadamente 50% de toda a energia). Os níveis mais baixos -*em especial de 6 a 9*- revelam-se mais energéticos neste exemplo de frulato do que em outros eventos abordados neste trabalho, sugerindo que a modulação provocada por este adorno musical se imprime por um número maior de níveis *wavelets*.

Violino: Ataque Expressivo com Acento (*Sforzato*) - nota mi4 (E4), corda lá



Evento: ataque acentuado (*sforzato*), nota única, ao primeiro mi (E4) na corda lá.

Gráfico do sinal original:



Extensão da amostra: 2^{16} pontos (65536). **Wavelet de análise:** D16

Análise dos níveis:

No ataque do nível mais fino (15) é maior a participação de parciais mais altos, efeito que se dilui da região de sustentação em diante. A maior sobrelevação neste nível ocorre entre 100 e 120 ms, e não no início, logo após a subida inicial. A textura sônica denota a presença do ruído do arco e a composição harmônica do tom, principalmente na região inicial onde soam mais intensamente os parciais mais altos. O formato do ataque, sobrelevação e decaimento no início não estão bem caracterizados, havendo em verdade uma sequência destes padrões na fase inicial da nota. Essa complexidade da forma de onda inicial reflete a complexidade das vibrações iniciais que podem ser estabelecidas pelo engate e arraste do arco na corda do violino.

Os níveis de 13 a 10 são os mais importantes para a formação desta nota acentuada. Pela audição (principalmente a 11025 Hz) e observação gráfica destes níveis, verifica-se que há um **período inicial de engate**, onde o arco “ataca” a corda e não são perceptíveis *pitchs* estáveis (padrões vibratórios com altura definida) mas sim padrões irregulares e ruidosos, e um período consecutivo onde *pitchs* definidos são estabelecidos, havendo a percepção de tonalidades.

No nível 13, ouvido a 11025 Hz, percebe-se a grosso modo 3 fases compondo a região inicial do tom: uma **primeira**, que cobre o momento do engate, uma **segunda** porção onde parciais mais altos são emitidos juntamente com 2 *pitchs* centrais, e uma **terceira** porção, onde parte do brilho sônico anterior se reduz e melhora-se a percepção de 2 *pitchs* centrais. Destes dois *pitchs*, o mais baixo fica levemente mascarado durante a segunda fase pelas frequências mais altas.

O nível 12 exibe um formato de envelope onde pode-se facilmente observar uma região de ataque, uma sobrelevação, um leve decaimento e as regiões de sustentação e colapso. Contudo, estas regiões, normalmente associadas aos processos transitórios que se desenvolvem antes do período de sustentação, ocorrem na verdade antes desta maior sobrelevação visível. Isto porque o nível 12 é um dos mais energéticos, exibindo a evolução dos padrões harmônicos e preterindo os componentes transitórios. Sendo natural que o surgimento e evolução dos padrões harmônicos se dê realmente após os transitórios iniciais da nota, a forma de onda observada ao nível 12 não dá informações sobre a "turbulenta" fase inicial, e não exibe os padrões característicos da região de ataque para o *sforzato*.

Uma visão mais detalhada dos primeiros 250ms da nota ao nível 12 revela uma fase inicial, com forma de onda aproximadamente irregular, não exibindo padrões de ataque, sobrelevações ou decaimentos, mas sim exibindo um sutil padrão periódico que cresce de amplitude em cada período, e ganha relevância e regularidade a partir de 188 ms do início.

O envelope visto no nível 12 é portanto uma versão escalada do envelope ideal. Esta interpretação em particular do *sforzato* contribuiu muito para moldar a forma de onda

observada: a intérprete caracterizou a qualidade de acento com um incremento na pressão e/ou velocidade do arraste de arco após o engate inicial do arco à corda, produzindo assim a sobrelevação máxima também após a região inicial, onde já soam modos estáveis, e que ocorre no momento em que a energia do arco encontra menos resistência à sua transferência à caixa ressonante.

A influência do intérprete sobre a caracterização dos acentos é portanto inegável, e as regiões de máxima intensidade de *sforzatos* podem ocorrer bem após a região de ataque, no início da nota. Isto mostra que a síntese de notas acentuadas nos violinos deve abordar uma fase de engate do arco e independentemente uma fase de incremento no arraste para caracterizar o *sforzato*, esta podendo ocorrer dissociadamente do ataque inicial.

O engate do arco à corda é bem perceptível pela audição do nível 10, onde dois *pitchs* centrais são perceptíveis. Também nos níveis 11 e 12 são ouvidos dois *pitchs* soando.

Os níveis 9 e 8 (respectivamente o 7º e o 8º níveis mais finos) localizam e caracterizam bem o ataque acentuado *sforzato*, especialmente o engate inicial do arco, apesar de serem muito menos energéticos. Parte da qualidade sonora do *sforzato* está portanto centrada em níveis pouco energéticos. O 7º nível mais fino nas análises de violino (9 nesta análise) usualmente revela os processos acústicos que ocorrem no corpo de madeira do instrumento.

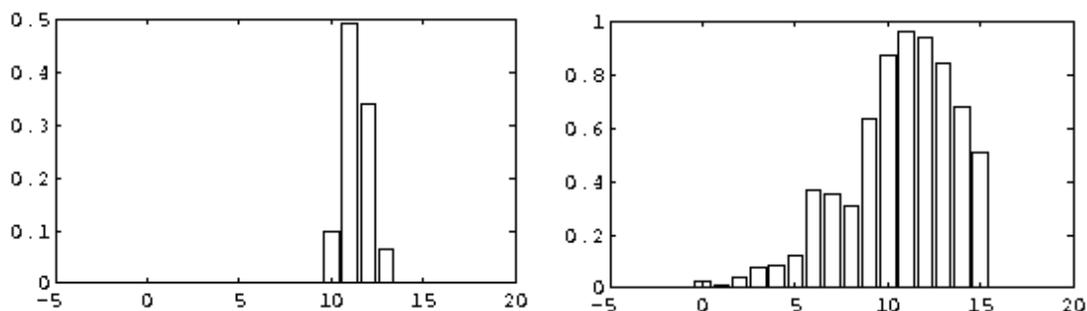
Violinistas são capazes de segregar as diversas sonoridades que emergem de pontos diferentes do instrumento, e identificar a parte do violino que está gerando aquele som, seja ele transiente ou estacionário. Os sons associados à caixa ressonante do violino, seu tampo e fundo, especialmente os transientes e pulsos, são enfatizados neste sétimo nível mais fino da análise *Wavelet*, podendo-se informalmente referenciá-lo pelo "**nível da madeira**", pela sua textura sônica, como se deixando passar somente a contribuição da caixa de madeira. Isto está de acordo com a percepção deste nível escutado por um violinista.

O nível 8 exhibe basicamente uma textura sônica de ruído de fundo.

O nível 7 da análise revela um ruído de fundo disperso por todo o exemplo, descorrelacionado com o envelope da nota tocada.

O nível 4 exhibe suas maiores amplitudes no início, exatamente sobre a região de sobrelevação do *sforzato*, e é inaudível.

Distribuição de energia:



O primeiro gráfico apresenta no eixo vertical a intensidade em escala linear. Mais de 99% de toda a energia da amostra analisada se concentra nos níveis de 10 a 13, sendo o nível 11 o mais energético.

No segundo gráfico o eixo vertical contém o logaritmo da energia de cada nível sobre o logaritmo da energia total. Reconstruções do sinal original desprezando-se os coeficientes dos

níveis 9 e inferiores (de 8 a 0) não implicam em perda de qualidade significativa no som final ressampleado (em comparação ao original).

Outras reconstruções relevantes foram realizadas para esse evento, inclusive algumas tomando-se alguns níveis (na verdade os coeficientes) da análise da mesma nota sob ataque normal, de forma a avaliar a importância sônica de alguns níveis para caracterizar o *sforzato*.

Uma delas foi feita zerando-se os níveis de 0 a 8 e também o nível mais fino (15); O sinal reconstruído foi praticamente idêntico à nota original, com diferença abaixo do limiar de percepção. Outra reconstrução zerando-se os níveis de 0 a 9 (inclusive) e também o nível mais fino (15) mostra que o nível 9 não é tão importante para a efetivação da percepção do *sforzato*.

O experimento mais significativo para se avaliar a concentração da qualidade do *sforzato* em alguns níveis utilizou a rotina **cruzar.m** para realizar reconstruções com níveis tomados de arquivos sonoros diferentes, implementando uma forma de **síntese cruzada** com *Wavelets*.

Dois arquivos sonoros, exemplificando eventos diferentes, foram empregados: (**A**) a nota mi na corda lá da presente análise, e (**B**) a nota mi também tocada na corda lá, sob interpretação normal⁴⁶. Os arquivos sonoros foram editados e devidamente acolchoados com zeros ao início e fim dos vetores, tais que o início de ambas as notas coincidisse no tempo, seus ataques comesçassem simultaneamente, e a duração dos vetores fosse igual em ambos. Em seguida, ambos os sinais foram decompostos com o sistema de AWMR, e a rotina **cruzar.m** foi empregada para se construir **sinais mistos**, contendo contribuições de ambos os eventos (ataque *sforzato*, A, e ataque normal, B).

Analisando-se comparativamente os níveis de **A** e **B**, nota-se que no nível 11 ambos apresentam mesmo *pitch*, mas exibem diferenças na região de ataque de **A** e na elevação ao final de **B**. O nível 10 difere bastante nos dois eventos: no ataque normal ele apresenta um envelope mais largo também no final, e varia de intensidade por toda a duração; no ataque com *sforzato* há um pico bem destacado no início. O nível 12 em ambos apresenta envelopes regulares e apresenta um *pitch* semelhante; o sinal com *sforzato* apresenta, contudo, um segundo *pitch* em destaque, associado à "mordida" do arco à corda.

Os sinais construídos em síntese cruzada foram em seguida ouvidos, interpretados e comparados com ambas as amostras originais de ataque normal e ataque acentuado com *sforzato*. A tabela abaixo mostra as mais importantes sínteses cruzadas realizadas e os arquivos onde se encontram armazenados seus sinais de áudio (disponíveis no CD-ROM em anexo):

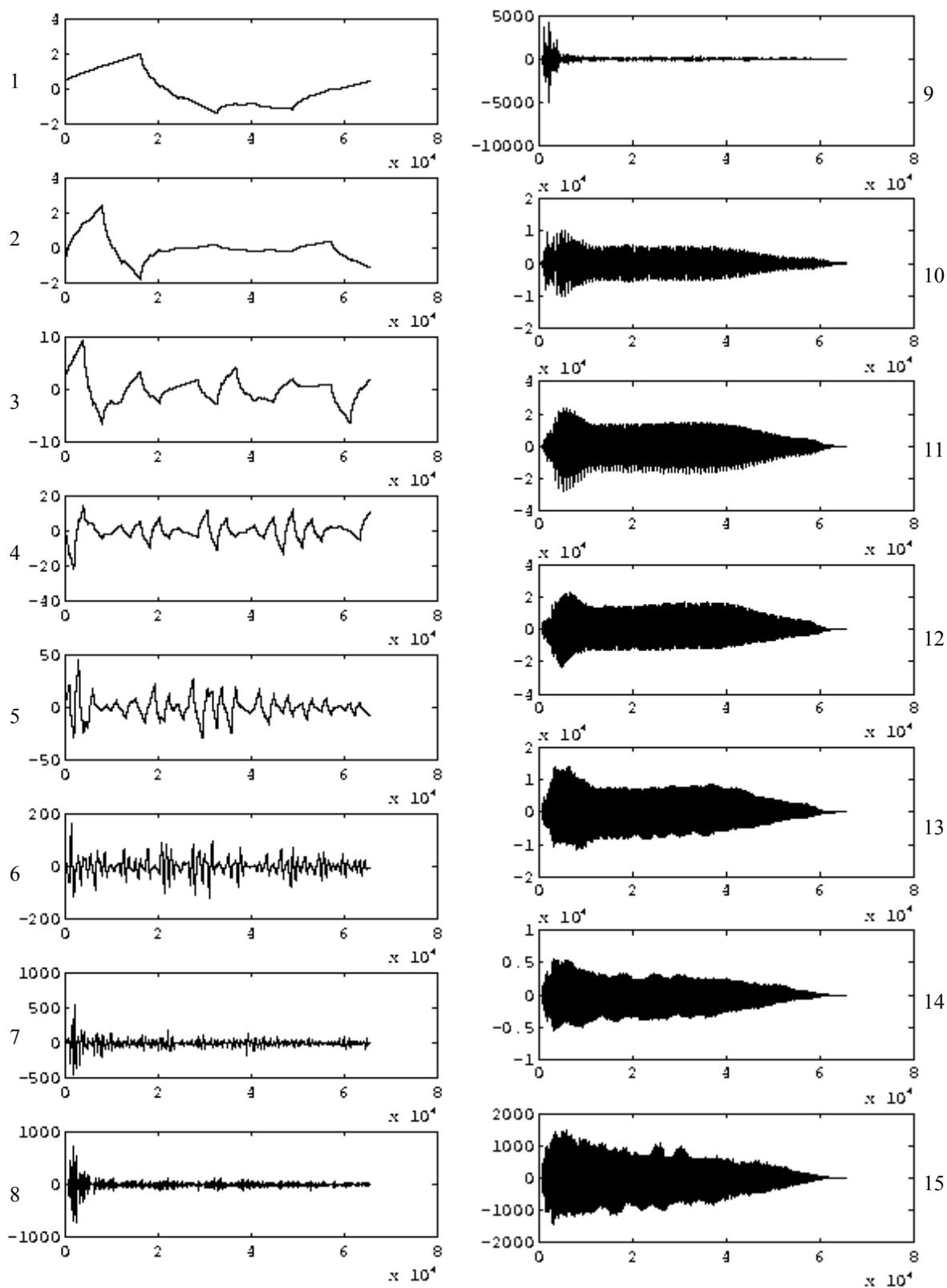
⁴⁶ Amostra não selecionada para análise neste trabalho. No entanto, é aqui utilizada por haver sido tocada também na corda lá, como a amostra de ataque com *sforzato* em análise o foi. **A** e **B** estão disponíveis no CD-ROM em formato AIFF (som) e MAT (MATLAB).

Arquivo	Níveis tomados de A (<i>sforzato</i>)	Níveis tomados de B (<i>normal</i>)
norm-sf1	10, com peso amplificado em 10 vezes (peso 10)	todos, exceto o 10
norm-sf2	10, 11 e 12, com peso 1	todos, exceto 10, 11, 12
norm-sf3	11, 12, 13, com peso 1	todos, exceto 11, 12, 13
norm-sf4	11, 12, com peso 1	todos, exceto 11 e 12
norm-sf5	10, com peso 3	todos, exceto o 10
norm-sf6	11, 12, com peso 1	13, com peso 0,225; 14 com peso 0,543; 15 com peso 0,552; os outros com peso 1
norm-sf7	8 e 9, com peso 3; 10, com peso 2,5	todos, exceto 8, 9 e 10

Somente com os níveis 11 e 12 da amostra A já é possível notar que uma qualidade de *sforzato* é adicionada à nota, mas existe uma degradação no som, evidenciada pela percepção de alguns *pitchs* (tons ou vozes adicionais) que não se integram (fundem num timbre ou *stream* único). A causa principal disso está no balanço deficiente dos pesos na reconstrução, o que é usualmente um motivo para se "dividir" uma nota em uma espécie de acorde onde outras vozes soam simultaneamente.

O *sforzato* ainda é perceptível se somente o nível 10 da amostra A for substituído com um peso 3 (3 vezes maior), como no arquivo **norm-sf5** acima na tabela. O timbre geral ainda degrada-se um pouco, distorcendo levemente a percepção de um violino, mas a qualidade do *sforzato*, em especial a região do ataque, é adicionada, mostrando a importância deste nível na formação da percepção do acento *sforzato*.

Violino: Ataque Expressivo com Acento (*Sforzato*) - nota mi4 (E4), corda lá (D4)



Evento: Ataque *sforzato*, nota única, ao primeiro mi (E4) na corda lá. Mesmo amostra sonora da análise anterior, desta vez analisada com a *Wavelet* D4.

Gráfico do sinal original: (veja análise anterior)

Extensão da amostra: 2^{16} pontos (65536). **Wavelet de análise:** D4

Análise dos níveis:

Os gestos observados e características gerais verificadas em cada nível são semelhantes às da análise anterior, da mesma nota acentuada com D16, especialmente para os níveis mais energéticos.

O nível mais fino encampa frequências mais baixas não cobertas pelo nível 15 da análise com D16, o que é perceptível pela audição comparativa dos mesmos. Seu som apresenta uma textura menos aguda que o som do mesmo nível para D16.

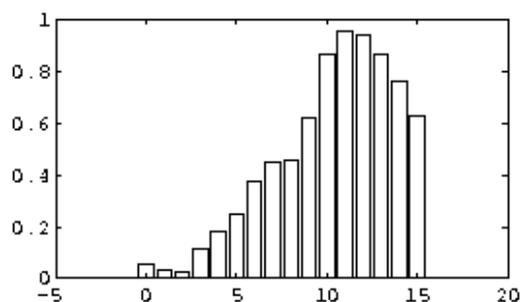
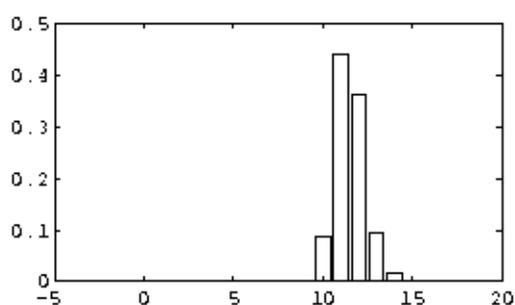
No nível 11 fica mais destacado o ataque de *sforzato*, sua audição apresentando a percepção de um tom ao início e estabilizando em outro: o ataque no sinal ao nível 11 apresenta 2 porções, cada qual com um *pitch* central dominante (mais pronunciado). O segundo tom permanece constante durante o exemplo, só reduzindo em intensidade no final.

O nível 10 caracteriza mais ainda o ataque *sforzato*. Comparando este nível com o nível correspondente da amostra da nota mi tocada na corda mi solta (da primeira análise de ataque normal, atrás) nota-se claramente a diferença provocada pela presença do *sforzato* neste nível. A intensidade no ataque acentuado (o presente) é mais forte, e além disso para aquela amostra foi utilizada a *Wavelet* D16.

O nível 7 com D4 permite uma localização do acento *sforzato*, guardando semelhança como nível imediatamente mais fino (8).

O nível 5 e inferiores também exibem formas de onda que já se assemelham à *Wavelet* D4, menos regular.

Distribuição de energia:



Mais de 99% de toda a energia do exemplo concentra-se nos níveis de 10 a 14, sendo o 11^o o mais energético. Comparativamente à análise com D16, a energia se distribui mais uniformemente pelos níveis, numa gradação mais suave entre a energia de um nível e do nível adjacente (degrau menos alto), como pode-se verificar comparando a distribuição (em escala logarítmica) do segundo gráfico acima com seu correspondente na análise anterior (mesmo sinal analisado com D16).

Como na análise anterior, algumas reconstruções seletivas também foram realizadas com os coeficientes obtidos com D4, as quais relata-se abaixo:

1. reconstrução com níveis 9 e 10 com peso 0,5 e os níveis 11, 12, 13 e 14 com peso 0,586: não elimina a qualidade percebida do acento *sforzato*.
2. com níveis 10 a 15 intactos, o restante zerado: leva a uma reconstrução praticamente idêntica à original, diferenças abaixo do limiar perceptivo.
3. níveis 10 a 14 intactos, o restante em zero: leva a uma reconstrução quase idêntica, havendo certa granulosidade na textura do som.
4. níveis 12, 13 e 14 intactos, o restante em zero.
5. níveis 13 e 14 intactos, o restante em zero: o tom fica muito degradado, falta muito do espectro harmônico, e o *sforzato* fica bastante atenuado, ainda que não eliminado.

Várias reconstruções seletivas foram feitas eliminando-se os níveis 9 e 10 e deixando-se diferentes pesos para os níveis de 11 a 15, mas nenhuma delas eliminou o *sforzato*. Ele aparece em todos estes níveis, com características e composição diferentes.

No item "Apreciação dos resultados e discussões" a frente é feita uma avaliação geral da análise *wavelet* em multiresolução aplicada a sinais musicais. São abordados aspectos importantes sobre a fenomenologia associada ao processo da análise e particularidades sobre as estruturas e objetos musicais identificados em cada nível, procurando sempre identificar a relação de causa e efeito verificada entre os gestos executados na interpretação instrumental e os objetos sônicos identificados nos níveis.

Nas conclusões finais elabora-se uma *tabela de distribuição de gestos/eventos musicais por níveis wavelets* que relaciona os objetos sônicos, estruturas, eventos e gestos musicais identificados aos níveis onde mais se concentram e são visíveis e/ou audíveis. Apresentam-se as estruturas sônicas que aparecem destacadas em cada nível, a percepção gráfica e sônica do nível, e uma lista de sensibilidade de cada nível, sumariando os resultados mais relevantes deste trabalho.

5.2 Apreciação e discussão dos resultados

As análises geraram um apreciável montante de dados sobre as propriedades dos níveis *wavelets* e sobre as características dos gestos e eventos musicais analisados, e naturalmente trouxeram à tona alguns aspectos importantes sobre o comportamento da análise *wavelet* em multiresolução aplicada a sinais musicais, sobre os tipos de objetos sônicos e estruturas musicais que ela permite identificar e isolar, sobre a influência da técnica de interpretação na conformação das estruturas musicais, e finalmente sobre o algoritmo e pacote computacional utilizado. Estes aspectos merecem pois uma apreciação e discussão a fim de que se compreenda a fenomenologia associada ao processo da análise *wavelet* e as particularidades associadas às análises realizadas.

Alguns temas são assim adequados para uma reflexão e discussão dos resultados das análises, os quais tratam sobre:

- A projeção dos eventos musicais nos níveis *wavelets*
- Os ruídos
- O imageamento ou geração de bandas espelhadas
- O efeito harmonizador
- Análise tempo-escala
- Reconstruções seletivas e compressão de dados
- Diferenças relevantes entre a análise com D4 *versus* D16
- Considerações sobre a qualidade e a intenção das gravações
- Enumeração alternativa para os níveis
- Considerações sobre o WaveLab

A projeção dos eventos musicais nos níveis *wavelets*

Para avaliar a capacidade e eficiência da AWMR em identificar e segregar padrões musicais em níveis *wavelets* distintos requer que o leitor esteja a par dos diversos objetos, estruturas e/ou padrões mais relevantes que foram recursivamente verificados, observados e/ou percebidos (gráfica e/ou sonicamente) nos diversos níveis das análises

realizadas. A leitura das íntegras das análises de flauta e violino é portanto necessária para se inteirar dos resultados detalhados de todas as análises.

Pode-se realizar avaliações dos resultados das análises a partir de dois pontos de vistas diferentes: o dos **eventos musicais** e o dos **níveis *wavelets***. No primeiro, aborda-se o "comportamento" dos níveis *wavelets* em função do evento analisado. Procura-se identificar os objetos e estruturas sônicas mais relevantes associados ao evento musical analisado e sua distribuição pelos níveis *wavelets*. A presença de determinado objeto em determinado nível e sua relação com o gesto musical que lhe deu origem são consideradas. Desta forma estaremos projetando os diversos gestos musicais e suas características nos níveis da análise que os destacam, e por conseguinte identificando os níveis onde determinados gestos se concentram.

O segundo ponto de vista - *o dos níveis wavelets* - é endereçado no item conclusivo seguinte. Nele os gestos e eventos musicais identificados em todas as análises são apresentados por níveis onde se destacam numa tabela, indicando as classes de objetos e estruturas sônicas que cada nível permite identificar e/ou concentra.

A seguir, a apresentação dos objetos e estruturas sônicas mais relevantes identificados por eventos analisados para a flauta, para o violino, ou para ambos. Discussões sobre aspectos sônicos percebidos, comparações entre resultados produzidos por *wavelets* diferentes, e resultados de reconstruções seletivas são também considerados.

ATAQUES NORMAIS

No nível mais fino das análises para flauta verifica-se um intervalo de tempo perceptível onde o instrumento está respondendo à excitação do sopro, estabelecendo vibrações transientes a caminho do estabelecimento de ondas estacionárias. A qualidade do som é ruidosa, retendo a maior parte do ruído característico do sopro, em especial nos ataques.

No nível mais energético, a pureza do som, os envelopes regulares (suaves), e a ausência de texturas ruidosas mostram serem os níveis mais energéticos (que concentram mais de 99% de toda a energia) aqueles que possuem a melhor textura sonora percebida. Também nestes níveis verifica-se uma maior insensibilidade a ruídos e transientes.

Os níveis abaixo do 9º nível mais fino (incluso) não contribuem significativamente para a formação do som original. Reconstruções realizadas desprezando-os ou mesmo intensificando seus coeficientes mostra somente que eles contribuem para um ruído de fundo grave (especialmente os 8º e 9º níveis mais finos), não sendo importante para os eventos musicais do exemplo.

A medida que o nível é mais energético ele também concentra mais da qualidade vibratória do timbre, isto é, da energia dos modos harmônicos vibrantes.

Os níveis mais baixos são capazes de revelar transientes oriundos de gestos labiais e da respiração associados à interpretação em flauta. De uma forma geral esta constatação se aplica aos outros eventos analisados.

ATAQUES EXPRESSIVOS COM ACENTOS

O nível mais fino normalmente concentra grande parte do ruído da excitação (arraste de arco para o violino, e sopro, para flauta) e permite monitorar a presença ou não desta, assim como variações na sua intensidade. No caso de acentos, a localização dos momentos de ataque é nítida, com uma súbita elevação na amplitude, que evolui no curso da região de sustentação.

Nas análises de flauta, os transientes associados ao colapso se destacaram mais neste nível, enquanto para o violino, não exibiram tanto destaque, nem no ataque nem no colapso. As maiores amplitudes associadas ao acento, nos dois instrumentos, estão na região de sustentação das notas. Isto mostra que os intérpretes procuram sempre por imprimir a intensa dinâmica associada ao acento não na região do ataque, mas logo após quando o tom já se estabelece. De fato, para a flauta por exemplo, um intenso sopro inicial pode excitar tons indesejados, a turbulência pode provocar a excitação de modos harmônicos que não os desejados, e portanto é natural que o acento progrida do ataque à sustentação.

Nos próximos dois níveis mais finos a região de ataque fica melhor destacada, com subidas íngremes e sobrelevações mais intensas. Os ataques acentuados, para ambos os instrumentos, exibiram taxas de crescimento no ataque maiores do que com ataques normais.

Os acentos exercem forte influência nos 7º, 8º, e 9º níveis mais finos, onde ficam bem evidentes e apresentam transientes amplos localizando o ataque inicial, e também outras

variações súbitas dos gestos de sopro (como nos colapsos). A altura da nota em questão determinará qual nível concentrará mais da qualidade súbita e intensa do acento, e dos outros gestos transientes de sopro.

Várias reconstruções seletivas sobre a amostra de acentos *sforzatos* (para violino) foram realizadas no intuito de se verificar a concentração deste gesto sobre alguns níveis. Tentativas de se eliminar a qualidade deste acento pela direta eliminação de diversos conjuntos de níveis mostraram que o *sforzato* aparece com características e padrões diferentes em cada nível, e não pode ser eliminado completamente da nota simplesmente editando-se os pesos dos níveis.

ATAQUES EXPRESSIVOS COM STACCATOS

O ruído de fundo verificado nos níveis mais finos desta análise para a flauta é indesejado, e de certa forma impossibilitou a observação dos componentes de ruído oriundos do sopro, uma vez que eles se fundem. *Wavelet-packets* podem ser úteis para separar estes componentes, haja visto que uma inspeção visual mais acurada das regiões de pausa (entre notas) e do início das notas mostra evidências de que as formas de onda apresentam espectros diferentes nestas regiões.

Os gestos bucais que caracterizam o pulso de sopro inicial dos *staccatos* são facilmente identificáveis nos dois primeiros níveis mais finos. A presença de "ruídos bucais" em fraseados velozes é frequente, e o 2º nível mais fino é particularmente sensível a tais gestos.

Na execução de *staccatos* em violinos, o padrão de alternância de sentido de arco pode estar evidente nos 8º, 9º e 10º níveis mais finos da análise *wavelet*. Destes, o 10º nível mostrou-se mais sensível à identificação destes gestos musicais. Neste tipo de ataque o intérprete usualmente muda o sentido arco para cada nota atacada. O arco para baixo normalmente consiste num movimento mais enérgico, e conduz a um gesto musical acentuado; o ponto e o ângulo de contato arco-corda, diferente em cada sentido, também influem sobre a qualidade e textura do timbre produzido, e conseqüentemente na sua composição espectral. Todavia, os tons das notas produzidas por sentidos opostos de arco serão tão diferentes quanto o intérprete o desejar, refletindo-se proporcionalmente na capacidade de identificação destes gestos pelos níveis acima.

Para *staccatos spiccatos*, a técnica interpretativa ensina que, para cada nota, o arco deve atacar súbito e sair da corda, deixando-a vibrar. Este gesto musical conduz a um **colapso natural** da nota após o arco abandonar a corda, produzindo "regiões reverberantes" características que podem ser verificadas gráfica e auditivamente principalmente nos 5º e 6º níveis mais finos da análise (para notas inscritas na 4ª oitava).

FRASEADOS LIGADOS NORMAIS - GLISSANDO (VIOLINO)

Nas análises de glissando o que chama a atenção é um efeito sonoro onde as linhas espectrais mais pronunciadas se deslocam em frequência, algumas para cima, outras para baixo, produzindo padrões sonoros transitórios interessantes. Nos tons mais altos em frequência, o número de linhas espectrais mais pronunciadas é menor, e a separação entre elas é maior, em oposição ao que se verifica nos tons mais baixos, com mais raias espectrais intensas presentes e uma menor separação (intervalo de frequência) entre elas.

Em alguns níveis - em particular no 7º nível mais fino ou 5º mais energético - o efeito do glissando é equivalente à imposição de um envelope de amplitude sobre o sinal.

FRASEADOS LIGADOS NORMAIS (FLAUTA)

A segmentação de notas é inexistente no nível mais fino, uma vez que no fraseado ligado o ruído de sopro existe por todo o fraseado. A segmentação, entretanto vai melhorando nos níveis menos finos subsequentes da análise, até atingir sua melhor forma nos níveis mais energéticos, onde a regularidade da forma de onda e suavidade do envelope são mais destacados.

Somente a primeira nota de um fraseado ligado apresentará o formato característico da região de ataque, com uma rampa de subida íngreme, uma região de sobrelevação de amplitude e decaimento. Tal padrão é observado em vários níveis do 2º ao 7º mais fino. Nos 9º e 10º níveis mais finos o momento do ataque da primeira nota é localizado por um transiente intenso.

Em vários temas musicais, quando se deseja impressioná-los com expressividade, faz-se uso de recursos dinâmicos alterando a intensidade de algumas notas chaves e imprimindo um vibrato associado, como ocorre na última nota do fraseado ligado

normal analisado para a flauta. Os níveis onde as notas com dinâmicas mais fortes contribuirão mais em energia são aqueles que cobrem as oitavas onde localizam-se os modos harmônicos do tom (e nos níveis adjacentes), bem como, naturalmente, naqueles mais energéticos.

FRASEADOS LIGADOS ACENTUADOS

Nos fraseados onde há uma transição ligada de uma nota normal para uma expressiva acentuada verifica-se a AWMR fornece meios para localizar o instante de *onset* (ataque) da nota acentuada, as regiões intensas que ocorrem nos níveis mais energéticos após o decaimento, bem como caracterizá-los nos diversos níveis da análise, expresso em termos de variação de intensidade do sinal. Acentos em flauta estão mais associados com a formalização do transitório inicial do ataque do que nos violinos, que além disso exibem amplitudes elevadas nos níveis mais energéticos, após a região de decaimento da nota.

FRASEADOS LIGADOS - TRÊMULO DEDILHADO (FLAUTA)

No nível mais fino os transientes com alta declividade, e que são periódicos no trêmulo, localizam os instantes em que há novo ataque na alternância das notas envolvidas no trêmulo.

Os picos de máximo (e de mínimo) ao 7º sétimo nível mais fino (nível 10 para este exemplo) se situam **sobre as regiões centrais das notas**, e não sobre o ataque. Isso mostra que o nível é sensível aos componentes harmônicos que só são realçados ou alcançam máxima excitação nas regiões de sustentação de cada nota, quando o padrão harmônico estacionário já se estabeleceu. Tanto que, nas regiões de transição observamos um estrangulamento substancial nas amplitudes, marcando os pontos onde as vibrações estacionárias são débeis porque o padrão vibratório está variando.

FRASEADOS NÃO LIGADOS NORMAIS

O nível mais fino mostra uma segmentação razoável das notas e torna possível localizar com precisão os momentos do ataque, colapso e pausas. As rampas de subida são íngremes, isto é, as taxas de crescimento no ataque são elevadas. A textura sônica é ruidosa, e pela audição não se diferencia os tons das notas: o som é um ruído de alta

frequência. O envelope das notas varia em amplitude na medida em que a pressão de sopro é maior, demonstrando uma extrema sensibilidade à presença do ruído de sopro. O nível também permite avaliar o comportamento e evolução do sopro na execução das notas identificando modulações na sua intensidade realizadas pelo flautista. Modulações de amplitude semelhante a vibratos (de amplitude, para a flauta) são observados em alguns níveis.

Na análise do 3º nível mais fino observou-se que a frequência modulante do “vibrato” varia dentro de uma única nota. A frequência de batimento inicial pode ser inicialmente mais lenta, e acelerar em direção ao final. Isto sugere a existência de uma **correlação com o aumento de intensidade no decorrer da nota** (*crescendo*) percebido nos níveis mais finos, em especial no 2º nível mais fino. Nas regiões onde o sopro é menos intenso o "vibrato" é mais lento, e cresce em velocidade (frequência de batimento) nas regiões onde há maior energia de sopro aplicada.

FRASEADO NÃO LIGADO ACENTUADO

O envelope típico de uma nota acentuada em fraseado não ligado é essencialmente semelhante ao ataque de notas individuais acentuadas.

Na análise de **fraseado não ligado acentuado para flauta**, o acento é marcado por um pulso, objeto sonoro com uma elevada taxa de variação de amplitude no ataque, e também por uma sobrelevação (pico de máxima intensidade) logo após.

Os ataques de todas as notas ficam melhor caracterizados no 2º nível mais fino, com uma intensa e repentina elevação nas amplitudes do sinal no início do ataque, atingindo uma sobrelevação inicial (que caracteriza o acento) seguida por um decaimento, que finaliza a região do ataque.

Uma análise com *wavelet-packets* do sinal neste nível, fragmentando melhor o segundo nível mais fino, poderia revelar melhor a composição frequencial em bandas de frequência. No entanto, não é nosso objetivo no instante avaliar aspectos quantitativos das modulações existentes utilizando esta técnica. Na reprodução do nível a 44100 Hz fica mais difícil de se perceber a modulação em amplitude via audição.

As descrições dos envelopes de ataques referenciando formas de onda com intensas variações de amplitude em curtos intervalos de tempo, seguidas de sobrelevações e decaimentos, se aplicam a todos os tipos de ataques não ligados.

Ao 4^o nível mais fino (da mesma análise para flauta) observa-se uma tendência das regiões de sustentação de todas as notas exibirem amplitudes semelhantes.

Ataques acentuados ficam melhor caracterizados nos 3 primeiros níveis mais finos, onde as rampas de subida, sobrelevação e região de decaimento e transientes de gestos de sopro associados tornam-se mais visíveis e as formas (estruturas), mais pronunciadas. Também a informação auditiva evidencia a qualidade de ataque acentuado, especialmente no 3^o nível mais fino, onde é maior a regularidade da forma de onda e menor a quantidade de ruído.

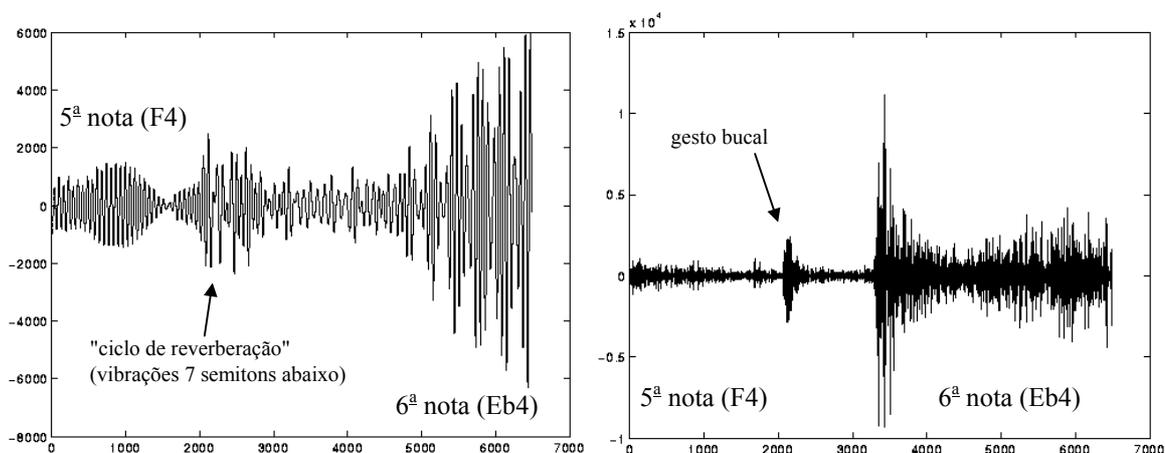
Estruturas caracterizando gestos bucais existentes entre as notas não contribuem relevantemente no 4^o nível mais fino desta análise para flauta.

Observou-se que as vibrações do tom entre um gesto bucal (modulando o sopro) e o final do decaimento do ataque da 6^a nota para esta análise não correspondem aos parciais harmônicos da 6^a nota, que só se estabelecem na região de sustentação desta nota.

Em fraseados não ligados na flauta, usualmente há um gesto bucal ou de sopro articulando as notas, de forma a separá-las, forçando o colapso da anterior e produzindo um ataque para a nota seguinte. Observou-se nos primeiros níveis mais energéticos (em especial no 6^o nível mais fino) que certas vibrações (vários semitons abaixo do tom exibido durante a sustentação) aparecem tão logo a nota entre em colapso, isto é, tão logo o sopro seja cortado, e soam como um “ciclo de eco ou reverberação”. Estas vibrações experimentam um desvanecimento natural, e que pode perdurar o suficiente para “invadir” a região de ataque da nota seguinte, enquanto as tonalidades estacionárias desta não se firmam.

A Figura 41 abaixo mostra um exemplo disto, ilustrando a região de transição entre a 5^a e a 6^a notas do tema da **análise de fraseado não ligado acentuado (flauta)**, respectivamente no 6^o e no 2^o nível mais fino desta análise, e cobrindo o mesmo intervalo de tempo em ambos os níveis:

Figura 41 - Transição da 5ª para a 6ª notas nos 6º (à esquerda) e 2º níveis mais finos (à direita)



Verificou-se que as vibrações 7 semitons abaixo de F4 soam tão logo a excitação das vibrações de F4 terminam, o que coincide com a localização do gesto bucal existente entre as 5ª e 6ª notas no 2º nível mais fino, e vão terminar após iniciado o ataque da 6ª nota (Eb4), mais especificamente após a região de decaimento desta nota.

O 6º nível mais fino (nível 12 da análise em questão) focaliza melhor as componentes harmônicas do som, localizando melhor as regiões onde estas são mais energéticas. O 2º nível mais fino (nível 16 desta análise) permite localizar os gestos bucais e os momentos onde há transições e chaveamentos no sopro. Note que as vibrações harmônicas da 6ª nota (Eb4) no 6º nível mais fino ganham energia a partir do final do decaimento observado no 2º nível mais fino, após o ataque inicial. Esta é uma informação que pode ser bastante útil para orientar a síntese natural de fraseados não ligados com acentos em sistemas computacionais.

Este comportamento, entretanto, não se repetiu nas transições de todas as notas daquele exemplo, o que sugere que o fenômeno seja visível ou não num certo nível em função dos intervalos tonais entre as notas. Algumas transições assim seriam melhor visualizadas em níveis específicos, que dêem cobertura às faixas de frequências relevantes na transição. Uma análise com *wavelet-packets*, que permite uma melhor segmentação do espectro poderia permitir uma melhor visualização deste fenômeno para um número maior de transições tonais, e verificar a ocorrência do mesmo fenômeno, tal que a hipótese possa ser estendida a uma classe maior de transições não ligadas, ou mesmo generalizada. Um mapeamento completo deste comportamento extrapola os limites deste trabalho, ficando como sugestão para futuras pesquisas.

Ao 8^o nível mais fino as notas mais baixas (da 1^a a 5^a) apresentam seus instantes de ataque bem marcados por pulsos transientes. A localização de transientes associados a gestos bucais de ataques é também verificada ao 11^o nível mais fino (7^o nível mais grosso desta análise).

Nos níveis mais grossos, em especial no 6^o nível mais grosso, a segmentação de notas verificada simplesmente não corresponde à segmentação do sinal original, exatamente por extrapolar os limites originais dos eventos musicais do sinal original. Isso entretanto não impede que o nível não localize ou aponte regiões de eventos importantes. De uma maneira geral, pode-se estender essa conclusão para todas as frases ligadas analisadas, que não exibam a capacidade de localizar eventos ou notas.

No violino, na análise de fraseado não ligado acentuado, o 4^o nível mais fino da análise sugere a existência de uma compartimentação de cada nota em regiões com padrões vibratórios e amplitudes relativas diferentes. Nesta análise é feita uma descrição detalhada dos gestos musicais associados à execução de fraseados não ligados acentuados, interpretando seus efeitos sobre a modelagem de cada região das notas.

Acentos (*sforzatos*) são melhor caracterizados já na região de sustentação das notas, e não propriamente na subida ou sobrelevação do ataque. Quando o ataque à corda é acentuado, sua duração é bem curta (assim como a duração da sobrelevação e decaimento iniciais) e a fase de sustentação é estimulada mais cedo. A qualidade acentuada vai se manifestar pouco depois do ataque, contando com as mais altas amplitudes (em especial nos níveis mais energéticos). O transiente do contato inicial à corda é melhor visualizado nos primeiros níveis mais finos e no 8^o nível mais fino.

FRASEADOS NÃO LIGADOS COM STACCATOS

Há uma sutil diferença entre os ataques *staccatos* e acentuados no nível mais fino: os acentuados apresentam uma subida com declividade mais acentuada, de contorno bem aproximado por um envelope quadrilátero, e projetam uma sobrelevação mais intensa seguida de um decaimento; os *staccatos*, apesar de exibirem aproximadamente a mesma descrição, apresentam declividades de subida menos acentuada e sobrelevações mais discretas. Alguns deles neste exemplo neste nível apresentam rampas íngremes de subida, outros, especialmente os associados às notas intermediárias de menor

intensidade (veja figura do som original), apresentam subidas mais discretas, com intensidades em elevação gradual.

Uma característica marcante do nível mais fino está na sua sensibilidade a objetos correlacionados aos gestos de sopro, os quais são bem localizados. Particularmente dois tipos de gestos bucais e de sopro são rastreados neste nível: o movimento da língua, gerando um estalido na interrupção do jato de ar, marcando o final do colapso das notas, e os estalidos que ocorrem nas transições, previamente à liberação de novo jato de sopro.

Também no nível mais fino de análises de *staccatos spiccatos* de violino observa-se que ataques pontuais e energéticos à corda geralmente intensificam a energia dos parciais harmônicos mais altos, aumentando a resposta nas altas frequências.

Ao segundo nível mais fino observa-se claramente estalidos nas transições das notas, os quais são objetos acústicos transientes provocados pela movimentação da língua nos instantes de chaveamento do sopro, em regiões fora das notas.

A região de colapso para a maioria das notas exhibe uma queda gradual, sem decréscimos súbitos e acentuados. Isso se deve à fisiologia do movimento lingual que modula o fluxo de ar para o *staccato*: a língua, ao mover-se em direção à parte posterior dos incisivos superiores para interromper o jato de ar, promove uma redução gradual no fluxo/pressão do sopro no instrumento, que culmina com o fechamento total do sopro, marcado por um estalido rastreável como um transiente logo após a região do colapso. O início do colapso portanto não fica marcado por transientes ou pulsos provocados pela língua ou lábios, e ele se torna portanto uma extensão da região de sustentação onde as intensidades estão decrescendo.

Desta forma, para notas emitidas sob *staccatos* não se observa marcadores (transientes) de início de colapso, mas sim marcadores de fim. Estalidos semelhantes podem ocorrer também no momento de retração da língua previamente à liberação de novo jato de sopro (portanto previamente ao ataque da próxima nota).

Os níveis mais energéticos exibem contornos mais suaves, logo envelopes mais suaves. Exibem também uma maior sensibilidade em destacar os processos sônicos que ocorrem na região central das notas: a região de sustentação, onde as vibrações harmônicas apresentam maior contribuição. Adicionalmente possuem um som “mais puro”, sem a textura granulosa e áspera características dos primeiros níveis mais finos.

Para *staccatos spiccatos* em violinos, as observações anteriores feitas para ataques expressivos com *staccatos spiccatos* valem igualmente no contexto de fraseado. A alternância de sentido de arco e o abandono do arco à corda para cada nota faz até mais sentido quando inserida no contexto de um fraseado não ligado do que propriamente para notas individuais.

FRASEADO NÃO LIGADO - TRÊMULO DEDILHADO (FLAUTA)

Os dois primeiros níveis mais finos são importantes para se caracterizar o ciclo completo de sopro (do ataque ao colapso) para trêmulos não ligados, evidenciando o envelope característico do ataque em fraseados não ligados, com um transiente inicial (uma rampa de subida íngreme) e uma sobrelevação seguida de um decaimento.

A região de sustentação usualmente apresenta altas intensidades, superiores às da sobrelevação do ataque, e, em seguida, pela diminuição da pressão de sopro devido ao avanço da língua contra os incisivos superiores reduzindo o sopro, apresenta uma queda progressiva nas amplitudes, entrando na região de colapso. A interrupção completa do sopro, que é cíclica no trêmulo não ligado, provoca o decaimento súbito ao final do colapso.

Ao terceiro nível mais fino o colapso para ambas as notas do trêmulo consiste numa queda de elevações menos abrupta -*mas ainda acentuada*- do que nos níveis anteriores mais finos, e são estendidos por uma pequena reverberação ou “eco”, como um pulso após o corte da excitação. A queda acentuada da região de sustentação para a região de colapso se deve ao desvanecimento das vibrações harmônicas das notas.

No 8º oitavo nível mais fino, usualmente aquele que concentra um ruído grave de fundo, verificou-se a sensibilidade a gestos bucais da interpretação, marcados por transientes. Estes são os eventos musicais mais prováveis de estarem visíveis neste nível, por serem periódicos mas apresentando frequência de periodicidade inconstante, exatamente por serem processos sincronizados pelos ritmos naturais dos movimentos do intérprete.

Observa-se também que quando um número maior de notas está presente o 9º nível mais fino da análise mostra-se sensível a periodicidades na interpretação de várias notas, agindo a grosso modo como um “contador de notas”. Os níveis mais grossos além deste podem ser úteis na extração de informações musicológicas em trechos mais longos, uma

vez que são versões do sinal analisado vistas sob escalas maiores, e assim podem localizar movimentos e estruturas mais longas (distribuídas no tempo).

SUSTENTAÇÃO COM DINÂMICA

Uma avaliação do 9º nível mais fino da análise de sustentação com dinâmica para a flauta mostra que ele é mais energético que o nível anterior mais fino, não é sensível à variação dinâmica, e que ruídos e vibrações de baixa frequência sem correlação alguma com o material musical sob análise se encontram presentes no som original, e neste nível concentram sua maior contribuição. Pode-se eliminá-los do som desprezando-se os coeficientes deste nível, uma vez que a sua ausência não prejudica a composição harmônica do som musical. Uma reconstrução assim foi realizada, e, no sistema de áudio utilizado, não foi perceptível qualquer diferença com o sinal original. Diferenças, porém, podem ficar evidentes se um sistema mais sofisticado de áudio digital, em estúdio apropriado, for utilizado. Cabe a um trabalho futuro dar cabo a este experimento.

Para o violino, verificou-se que as variações dinâmicas na região de sustentação estão presentes nos 7 primeiros níveis mais finos da análise, mostrando que uma larga faixa de frequências é sensível a alterações da dinâmica e aos gestos musicais que as produzem, principalmente as variações na velocidade e pressão de arraste do arco, além do ponto de contato do arco na corda. O 8º nível mais fino mostrou-se sensível ao transiente provocado pelo abandono do arco à corda, ao final da nota.

SUSTENTAÇÃO COM VIBRATO (FLAUTA)

O vibrato para flauta constitui-se fundamentalmente numa modulação de amplitude intencional, que é verificada por uma larga faixa de escalas ou níveis *wavelet*, do 1º ao 10º níveis mais finos, sendo que o 7º nível mais fino apresenta maior sensibilidade aos gestos transientes do que propriamente à modulação do vibrato. Em alguns níveis pode-se extrair uma estimativa da frequência do vibrato por meio de simples avaliação das distâncias entre linhas espectrais.

SUSTENTAÇÃO COM FRULATO

Ao oitavo nível mais fino, o som é característico dos níveis baixos menos energéticos, e realça os componentes ruidosos de baixa frequência que existem no som original, sejam eles oriundos de ruídos ambientes, ou produzidos nas fases de gravação, conversões e preparo das amostras (mais provável), seja oriundo de vibrações graves excitadas nos instrumentos durante a emissão dos tons, ou de gestos de arco, sopro, respiração e dedilhados.

Testes de análise *wavelet* em multiresolução em sinais senoidais sintetizados (1000-2000 Hz) modulados em amplitude por frequências baixas (20 Hz) mostram que as frequências modulantes não ficam evidenciadas nos níveis baixos, por exemplo no quinto nível mais grosso e níveis vizinhos.

Os ruídos

Os ruídos considerados neste tópico consistem em sinais não suaves com forma de onda e contornos irregulares, que geralmente não seguem um envelope definido, e apresentam um espectro espalhado por uma faixa larga, seja em alta frequência (observado em níveis mais finos) ou nas baixas (em níveis menos finos). O som usualmente não transmite informação musical relevante, é percebido como um estímulo sonoro desagradável, constante e desinteressante, e a textura sônica é granulosa, “áspera”.

Excluem-se desta discussão os transientes, estalidos e outros sons impulsivos produzidos por gestos bucais e de arco que tenham sido referenciados nas análises pela denominação genérica de “ruídos”. Este constituem uma classe de objetos altamente correlacionados aos gestos musicais e não exibem propriedades de processos aleatórios.

Basicamente duas classes de ruídos foram observadas: uma em **alta frequência (agudos)**, manifestando-se nos níveis mais finos, e uma em **baixa frequência (graves)**, em alguns níveis mais baixos (menos finos). O ruído característico do sopro (para a flauta) e do arraste/atrito do arco (para o violino) integram a classe de ruídos verificados nos níveis mais finos, em particular concentrando-se no nível mais fino da AWMR.

Na classe de ruídos graves dois tipos básicos foram identificados: um de **origem instrumental**, que usualmente se manifesta nas análises de violino no 8^o (oitavo) nível mais fino e níveis adjacentes, e outro de **origem extra-instrumental**, cuja causa pode

estar associada a ruídos ambiente de fundo durante a gravação em estúdio, a ruídos de quantização e/ou aos diversos processos realizados na manipulação e preparo das amostras digitais.

RUÍDO DE SOPRO EM FRASEADOS LIGADOS

Gestos de sopro são produzidos pela ação dos lábios (embocadura), língua (moduladora do fluxo de ar) e pulmões (pressão de ar); usualmente promovem variações de intensidade e direcionamento no jato de sopro sobre o bucal, e servem a propósitos musicais diversos.

O flautista, na iminência de executar um pulo intervalar que esteja associado a uma grande mudança de posição (dedilhado), e que geralmente implica numa também longa variação do comprimento da coluna de ar excitada, perfaz um **ajuste de foco da embocadura**, usualmente associado a uma breve interrupção parcial do fluxo de ar ou a uma leve variação (acréscimo ou decréscimo) na pressão de sopro. Esta modulação breve no fluxo de ar constitui um **gesto de sopro**, e pode servir aos seguintes objetivos:

1. Certificar que a próxima nota terá um leve ataque, sem comprometer a condição de fraseado ligado, de forma a garantir sua correta entonação e destaque na frase, e
2. Pontuar ou destacar a próxima nota fornecendo-lhe uma expressividade especial e às vezes realçando o ruído de sopro, um componente natural à flauta e característico de seu timbre.

Geralmente o gesto de sopro ocorre ainda sobre o colapso da nota anterior, alterando o padrão vibratório presente.

Quando a interrupção ou desfocalização do sopro ocorrerem, serão melhor visualizados nos níveis mais finos, porque o ruído de sopro sofrerá modulação evidente. Os ajustes de foco da embocadura e modulações na pressão de sopro que provoquem direta variação na intensidade das vibrações harmônicas serão melhor localizados, porém, nos níveis mais energéticos.

ORIGENS PROVÁVEIS DE RUÍDOS

Ruídos podem ter diversas origens, dentre as quais citam-se as mais principais:

1. **ruído ambiente do estúdio.** O estúdio do departamento de cinema e televisão da ECA é um estúdio adequado para gravações de jornalismo, e a despeito de apresentar boa isolamento externa e baixa reflexão das ondas sonoras⁴⁷ pode ter apresentado baixo isolamento acústico em faixas de alta e baixa frequências
2. **sensibilidade do microfone.** O microfone utilizado foi um Neuman modelo MKH 416 P 48 U, com acentuada característica direcional, sensível a movimentos sutis e capaz de captar uma variedade de eventos acústicos e pequenos ruídos causados pela respiração, movimento dos dedos sobre os instrumentos e a própria manipulação do instrumento, além de ruídos de baixa frequência (aquém do registro dos instrumentos), cochichos, manipulação de papéis, arraste de pés, etc.
3. **qualidade do DAT utilizado na gravação.** O DAT TEAC modelo DA-P20 é semiprofissional, e pode-se notar em várias gravações por ele feitas a presença de ruídos. Fontes possíveis de ruído num equipamento como esse estão diretamente ligadas à qualidade e precisão do conversor analógico-digital, especificamente à grandeza do ruído de quantização, ao circuito de codificação PAM (*Pulse Amplitude Modulation*) e eventuais estágios analógicos de pré-filtragem.
4. **Transferência de meio.** A material sonoro gravado originalmente em estéreo, em fita DAT, com taxa de amostragem de 48 KHz e largura de resolução de 16 bits/amostra, foi lido por outro equipamento (DAT Silicon Graphics) e transferido para disco rígido de computador, armazenado neste meio como arquivos de som estéreo, no formato proprietário da Apple (AIFF). Nesta fase de transferência existe a possibilidade de ocorrer um fenômeno de falta de sincronia na reprodução do *clock* original. Nas palavras digitais codificadas pelo DAT a informação de sincronia vem misturada aos dados do sinal amostrado, e em alguns casos, principalmente quando o equipamento de leitura é diferente do de gravação, existe a possibilidade de os bits de sincronismo serem mal recuperados ou produzirem um *clock* inconstante para o conversor digital-analógico, causando uma distorção não linear conhecida pelo nome de "*jitter*". Nos sistemas mais modernos, entretanto, o nível de *jitter* é baixíssimo, graças à melhor precisão dos circuitos PLL's e à baixa deformação do sinal elétrico (digital) no circuito de transporte. A codificação em formato AIFF não altera o conteúdo original de

⁴⁷ no jargão técnico diria-se que é um "estúdio morto".

informação, somente adicionando um cabeçalho (*header*) específico às palavras digitais tal que sejam reconhecidas pelo sistema operacional como arquivos de som (AIF).

5. **Conversão de taxa de amostragem.** A gravação original foi feita com taxa de amostragem de 48000 amostras/segundo (48 KHz) em dois canais (estéreo). Previamente à realização das análises as amostras sofreram uma mudança de taxa de amostragem, sendo convertidas para 44100 amostras/segundo (44,1 KHz, padrão CD), monaurais (um canal). Foram utilizados dois tipos de equipamentos diferentes nesta fase: uma estação de trabalho Indy, da Silicon Graphics, rodando os *softwares* Soundfiler ou MediaConvert, e um PC Pentium 100 MHz, rodando o programa *shareware* Cool Edit. A redução da taxa de amostragem e conversão para padrão monoaural foi necessária para realizar as análises em microcomputador, com placa de som operável até 44,1 KHz.⁴⁸ Não foi realizada nenhuma conversão de largura de resolução dinâmica, originalmente em 16 bits e mantida inalterada, o que exclui a possibilidade de algum algoritmo de "*dithering*" ou modelamento de ruído (*noise shaping*) ter adicionado ruído.
6. **normalização das amostras.** Os arquivos AIFF convertidos foram editados, e fragmentados em trechos menores, contendo somente os eventos analisáveis desejados. Seguindo a edição, foram normalizados em amplitude, num processo que requer a localização da maior amplitude no intervalo e a amplificação do trecho proporcionalmente à maior amplitude. Apesar de ser uma etapa necessária para se preparar as amostras para a análise, o processo pode realçar ruídos gerados nas fases anteriores, amplificando-os, e às vezes introduzindo distorções não lineares (distorção harmônica diferencial).

Dentre estas, a transferência de meio digital, conversão de taxa de amostragem e normalização das amostras são as etapas com maior potencial para geração de ruído. Também o próprio ruído de fundo acrescido dos sons oriundos de fontes que não os instrumentos musicais mas que exibam intensidades próximas ao nível de amplitude dos sons musicais quando estes são baixos (como respiração, vozes, arraste dos pés,

⁴⁸ para permitir a análise sônica. Os pacotes de análises (WaveLab e Wavesynt) todavia não impõem limite de frequência de taxa de amostragem para o processamento.

manipulação de papéis, etc.) constituem uma classe de ruídos que se pronuncia em diversos das amostras gravadas e analisadas neste trabalho.

SOBRE O RUÍDO DE FUNDO VERIFICADO NO OITAVO NÍVEL MAIS FINO E ADJACÊNCIAS

No 8^o (oitavo) nível mais fino, e também nos níveis adjacentes, verifica-se que um ruído grave se manifesta de forma mais intensa, especialmente nas gravações onde originalmente este também já era perceptível. Todas as origens de ruídos abordadas acima podem estar contribuindo para um aumento no nível de ruído verificado.

Estes níveis - *em especial o 8^o e o 9^o níveis mais finos da análise* - apresentam uma textura sônica ruidosa, concentrando e segregando componentes ruidosos associados aos ruídos de fundo ambiente, à manipulação do instrumento, a ressonâncias e vibrações de baixa frequência, e a gestos e modulações de baixa periodicidade (em frequências geralmente subsônicas).

Por exemplo, verificou-se que se concentra nestes dois níveis a maior parte do ruído grave, de fundo ambiente, originário da gravação - *contribuições não procedentes das vibrações musicais dos instrumentos*.

Em gravações baixas, i.é, onde a dinâmica é fraca e/ou existe uma baixa relação sinal/ruído no som, o processo de normalização de amplitude acentua a contribuição dos sinais menos intensos e ruidosos codificados nos bits menos significativos das amostras digitais, evidenciando também o **ruído de quantização** que é gerado no processo de amostragem e codificação digital linear PCM (*Pulse Coded Modulation*). Esta contribuição ruidosa também concentra-se principalmente nos 8^o e 9^o níveis mais finos da análise.

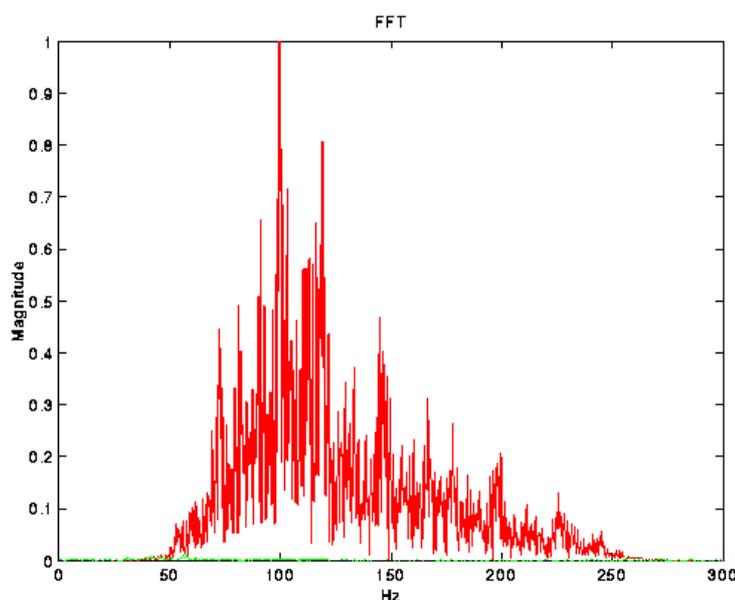
Em várias análises de flauta e violino, também verificou-se a sensibilidade ao ruído de quantização nos 11^o e/ou 12^o níveis mais finos da análise, que exibiram as maiores amplitudes das formas de onda (envelopes) sobre as áreas onde a dinâmica era menor, e as menores amplitudes sobre as regiões de sustentação das notas, exatamente onde os modos harmônicos estacionários apresentam maior energia.

Em reconstruções seletivas onde o 9^o nível mais fino foi amplificado (i.é, seus coeficientes foram amplificados) em até 12 vezes seu valor original, verificou-se que contribui com o ruído grave de fundo existente, não se mostrando importante na qualificação de gestos musicais.

Nas gravações originalmente pouco ruidosas, onde o nível dinâmico é alto o suficiente para preencher os bits mais significativos de cada amostra digital, o nível de ruído apresentado nestes níveis é menor, ainda que existente devido às outras fontes de ruído - *em especial as conversões de taxa de amostragem e resolução* - que tendem a potencializar ruídos presentes originais.

Tome como exemplo o 8º nível mais fino da análise de fraseado ligado, trêmulo dedilhado em flauta (o nível 9 naquela análise). Trata-se de um nível pouco energético, de textura irregular com contornos não suaves, de um som grave e ruidoso, e não demonstra segmentação de notas. A Figura 42 abaixo mostra o espectro (FFT) do sinal original (em verde) e do 8º nível mais fino (em vermelho) para esta análise, traçado entre 0 e 300 Hz.

Figura 42 - FFT do sinal original (em verde) e do 8º nível mais fino (em vermelho) entre 0 e 300 Hz



Há um realce nas frequências no 8º nível mais fino, particularmente entre 50 e 250 Hz, faixa que no sinal original não apresenta amplitudes significativas. Estas bandas não consistem em bandas espelhadas geradas pela análise/reconstrução *wavelet*. Note, contudo, que o gráfico está normalizado, porque as amplitudes absolutas das linhas espectrais ruidosas são muito pouco intensas.

PERCEPÇÃO SONORA RUIDOSA

A sonorização e subsequente análise sônica destes níveis podem, entretanto, conduzir a uma má interpretação de suas características e propriedades intrínsecas, mascaradas que

podem ser pela atribuição de uma qualidade ruidosa dominante percebida no som gerado.

A rotina MATLAB preparada para sonorizar e converter o vetor do sinal em arquivo sonoro implementa uma normalização de amplitude simples, promovendo um reescalamiento dos níveis de amplitude para a faixa (linear) de 1-65536 níveis discretos (codificação para 16 bits/amostra). O MATLAB trabalha normalmente com uma resolução de 8 bytes/amostra (64 bits/amostra) e uma redução para 16 bits/amostra normalmente adiciona erros (de quantização, usualmente aleatórios) ao sinal, traduzindo pequenas flutuações de amplitude no sinal de alta resolução em ruído no sinal de baixa resolução. Felizmente, antes de serem convertidas para o formato MATLAB de alta resolução todas as amostras foram previamente tratadas e pré-normalizadas, de forma a atenuar sobremaneira a possível potencialização de ruídos espúrios nesta fase de sonorização, e garantir a fidelidade das análises.

Um procedimento aplicável para se corrigir o nível de ruído em trabalhos futuros seria a utilização de técnicas de modelamento de ruído (*noise shaping*), muitas vezes referenciadas na literatura de processamento de áudio e imagens genericamente como *dithering*. Os algoritmos para se implementar tal processamento não são, todavia, triviais, sua implementação implica em custos significativos, e os melhores algoritmos/sistemas de *noise shaping* atualmente são proprietários (marcas registradas), desenvolvidos por indústrias do mercado de áudio profissional.

Há evidências ainda de que gestos de sopro podem ser localizados nestes níveis mais baixos ao redor do oitavo e nono mais finos, confirmadas pela localização de objetos sonoros que casam com estruturas localizadas em níveis mais finos, onde o ruído e gestos de sopro são mais visíveis.

O imageamento ou geração de bandas espelhadas

Em praticamente todas as análises e em vários níveis verificou-se a ocorrência de um fenômeno gerador de uma forma de modulação de amplitude sobre os sinais musicais.

A análise espectral (FFT) destes níveis mostra linhas espectrais organizadas em bandas lateralmente distribuídas (à esquerda e à direita) em relação a uma frequência central, não populada (usando-se a terminologia apropriada: com a portadora suprimida, i.é, sem a presença da linha espectral central).

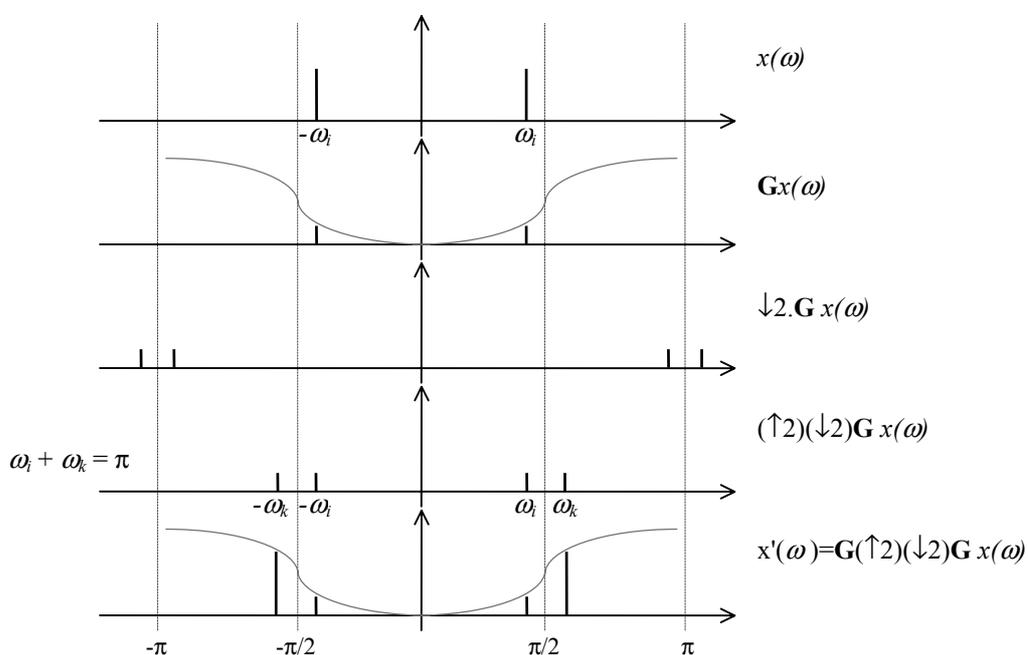
Em alguns exemplos uma flutuação periódica era observada no gráfico do nível e perceptível auditivamente, sugerindo que houvesse uma frequência ou grupo de frequências modulantes bem mais baixas que uma frequência (central) portadora, tal que a distância entre as linhas (bandas) laterais do seu espectro fosse curta, usualmente inferior a 100 Hz, e conseqüentemente a taxa de flutuação (batimento) fosse em baixa frequência, o suficiente para ser perceptível gráfica e auditivamente. Noutros casos a identificação de bandas laterais no espectro e a visualização da flutuação no sinal eram mais complexas, impedindo um claro reconhecimento das supostas frequências modulantes e portadora, ou mesmo sugerindo a existência de múltiplas portadoras.

Em primeira análise, a modulação em amplitude poderia ser oriunda de gestos musicais modulando o sopro (para a flauta) ou o arraste do arco (para o violino) de maneira sutil e imperceptível, mas que fossem reveladas pela análise *wavelet* na forma de flutuações periódicas nos sinais dos níveis *wavelets*. Neste caso, a frequência modulante seria única e de baixa frequência, enquanto as linhas espectrais do sinal musical comporiam um grupo de portadoras, em oposição à modulação de amplitude convencional onde uma única portadora é modulada por um grupo de frequências modulantes de baixa frequência. Todavia, a ausência de portadora e a constatação de modulações numa larga faixa de frequências (particularmente entre 70 e 4000 Hz, além das limitações fisiológicas para a produção de gestos físicos com esta periodicidade) afastaram esta hipótese.

De fato, as características que os sinais modulados apresentavam não sugeriam que uma modulação de amplitude (AM - *Amplitude Modulation*) convencional estivesse ocorrendo, porque o efeito não se manifestava regularmente para todos os tons/notas, nem para todos os níveis, assim como também afastava a possibilidade de uma forma de modulação de amplitude mais simples conhecida por **modulação de anel** (*ring modulation*), onde a portadora geralmente é suprimida. Além disso, não se manifestara o padrão e orientação das bandas laterais esperados de uma AM para quando um sinal de baixa frequência (gestos) modula sinais de frequência bem mais alta (vibrações musicais). A modulação produzida, entretanto, apresentava uma característica comum em todas as suas ocorrências, independente do tom ou nível onde ocorria, fato que norteou sua decifração e determinação de sua causa.

A causa está no algoritmo de decomposição e reconstrução *wavelet* utilizando-se estruturas piramidais, em particular no processo de reconstrução onde as etapas de superamostragem ($\uparrow 2$) provocam um **efeito de imageamento** ou **replicação de bandas**, usualmente espelhadas em relação a uma frequência central que é um divisor da frequência de amostragem, i.é, uma fração racional da frequência de amostragem ($1/64f_a$, $1/128f_a$, etc.). A figura abaixo mostra como cada etapa do processamento no algoritmo piramidal age sobre um sinal $x(\omega)$ no domínio da frequência.

Figura 43 - Decomposição seguida de reconstrução (canal passa-alta do algoritmo *wavelet*)



Na decomposição o sinal é filtrado (Gx) e decimado ($(\downarrow 2)Gx$). A decimação provoca um dobramento de frequências, esticando $x(\omega)$ em $x(\omega/2)$, e também gera uma imagem $x(\omega/2 + \pi)$ que aparece mais próxima de $x(\omega/2)$, isto é, gera *aliases*. Caso $x(\omega)$ não seja adequadamente limitado em frequência ($x(\omega)=0$ p/ $|\omega| > \pi/2$) haverá superposição destas bandas (*aliasing*).

Na reconstrução a superamostragem ($\uparrow 2$) provoca a compressão do espectro e imagens das bandas comprimidas aparecem próximas a elas, o suficiente para, em sinais de áudio, equivaler a um efeito de modulação de amplitude. Veja na figura acima (decomposição seguida de reconstrução sem edição dos coeficientes) como o sinal $x'(\omega)$ reconstruído passa a apresentar bandas laterais espelhadas ao redor de frações racionais divisoras da frequência de amostragem ($\omega_a = 2\pi \Leftrightarrow f_a = 44100$ amostras/s).

Algumas bandas presentes no sinal podem ser realçadas, e a baixa eficiência da filtragem em eliminar as frequências causadoras de *aliasing* é um problema detectado neste esquema de decomposição, particularmente para *wavelets* menos regulares. Strang (1996) aborda o efeito de imageamento associado ao processo de reconstrução.^[78] A análise de **sustentação com frulato para flauta** documenta e caracteriza bem a ocorrência do efeito de imageamento.

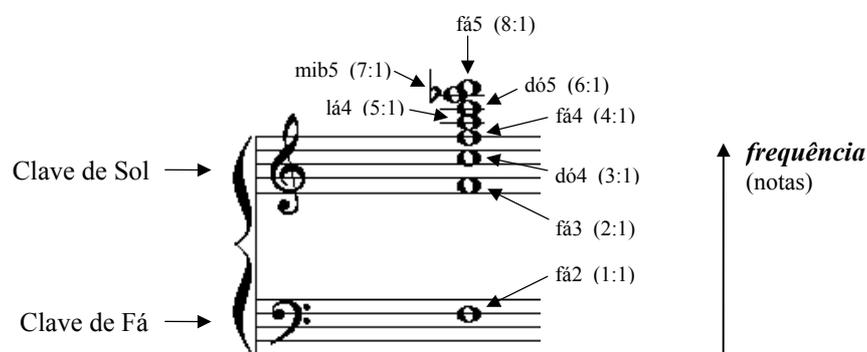
O efeito harmonizador

A audição do tema melódico no 5º nível mais fino da **análise de fraseado ligado normal para flauta** (nível 13 para este exemplo) revela dois aspectos importantes:

1. Surgem novas vozes para algumas notas, arranjadas em acorde (tendendo a compor uma função harmônica), e
2. A melodia no nível é diferente da melodia original, pois que algumas notas estão em tons diferentes, havendo migrado intervalos variando de um semitom a vários tons.

As notas parecem haver migrado porque o que se ouve num nível específico são aqueles harmônicos que passam pela banda de passagem do nível, além é claro das possíveis bandas espelhadas (imagens) que são percebidas como *pitchs* de novos tons, formando novas vozes. Os parciais da série harmônica da nota original podem naturalmente consistir em fundamentais para outras notas que estão a distâncias intervalares fixas da nota original na escala musical. A figura abaixo mostra um exemplo ilustrativo disto:

Figura 44 - Harmônicos da série de fá2 (F2) como notas/tons na escala musical



A nota **fá2** (F2, inscrita na pauta de baixo, em clave de fá) possui uma série harmônica onde seus modos parciais podem ser considerados fundamentais de outras notas na escala. Por exemplo, a nota **fá3** (F3), que está uma oitava acima de fá2, possui como

fundamental exatamente o segundo harmônico de fá2; da mesma forma, **dó4** (C4) tem no terceiro harmônico de fá2 a sua fundamental, e assim por diante, até a nota **fá5** (F5), que tem como fundamental o oitavo harmônico da série de fá2.

Adicionalmente, a geração de bandas laterais espelhadas ao redor de divisores (frações racionais) da frequência de amostragem faz surgir *pitchs* inexistentes no som original, que por sua vez podem consistir em fundamentais de outras notas na escala musical.

A Figura 45 mostra a melodia tocada para esta análise de fraseado ligado normal (flauta), e a Figura 46 seguinte mostra uma boa aproximação para a melodia/harmonia gerada ao 5^o nível mais fino: ⁴⁹

Figura 45 - Melodia tocada - Análise de fraseado ligado normal (flauta)



Figura 46 - Melodia percebida no 5^o nível mais fino (análise de fraseado ligado normal, flauta)



A Figura 47 mostra o que é ouvido no 6^o nível mais fino:

Figura 47 - Melodia percebida no 6^o nível mais fino (análise de fraseado ligado normal, flauta)



A Figura 48 mostra a partitura sugerida para a melodia no 7^o nível mais fino:

Figura 48 - Melodia percebida no 7^o nível mais fino (análise de fraseado ligado normal, flauta)



A função harmônica ou tipo de acorde que se forma vai depender de vários fatores:

⁴⁹ aproximação porque as notas e suas alturas (*pitchs*) podem ser percebidas diferentemente, e diferentes pessoas podem atribuir a elas diferentes tons, microtons, interpretar uma mudança na tonalidade central, ou mesmo admitir a existência de uma escala não temperada.

- (1) a **tonalidade da nota original**, que irá determinar a posição no espectro da série harmônica da nota, e sob quais níveis *wavelet* ela irá contribuir com maior energia
- (2) as **intensidades relativas de cada um de seus harmônicos**; alguns deles, mais realçados, poderão produzir bandas espelhadas também realçadas, e que podem eventualmente constituir-se no *pitch* mais relevante, uma nova voz que se sobressai
- (3) o **timbre do instrumento**, que influi no padrão exibido de intensidades relativas dos vários parciais harmônicos.
- (4) a ***wavelet* utilizada** e o seu grau de regularidade, que irão determinar a quantidade de sintonia nas bandas cobertas pelos níveis, e assim reduzir os efeitos de imageamento de bandas. *Wavelets* mais regulares propiciam bandas de passagem mais sintonizadas e uma atenuação maior nas frequências acima desta banda. *Wavelets* menos regulares apresentam uma banda de passagem definida e individualizada, mas não apresentam um corte efetivo para as frequências acima desta banda, admitindo uma série de “pequenas bandas” em atenuação progressiva à medida que cresce a frequência, como mostra a Figura 9 no item 2.2 (p.32).

O arranjo com que os parciais harmônicos de um timbre instrumental (como flauta ou violino) organizam suas intensidades relativamente entre si propicia à percepção de um timbre único e com uma identidade individualizada. A análise *wavelet* em multiresolução produz níveis onde as intensidades relativas entre os parciais ali presentes (i.é, o arranjo de harmônicos por nível) não levam a uma fusão de forma a permitir a percepção do timbre, segregando a composição harmônica dos tons. Os níveis passam a exibir um timbre dissociado do original, dificultando a percepção do instrumento original. A percepção do timbre será mais fácil pela análise dos transientes, análise de envelopes e da textura sônica dos níveis, que forneçam indícios sobre o tipo de excitação presente, sobre a forma de estabelecimento das vibrações e sobre as regiões de colapsos.

Neste processo de dissociação, vozes adicionais são geradas, produzindo acordes ou funções harmônicas pela sua combinação. Uma possibilidade é a de que alguns harmônicos venham a produzir uma nova voz num nível específico, eventualmente compondo um acorde com outras vozes derivadas de outros harmônicos originais ou das bandas espelhadas destes, as quais na maioria das vezes não constituem parte integrante

da série harmônica original, mas sim da série harmônica de outras notas da escala cromática. A intensidade relativa entre os harmônicos presentes num nível parece ser o fator mais importante que irá determinar quais deles se integrarão numa só voz perceptível.

Assim a dissociação da série harmônica e o imageamento causado pelo processamento *wavelet* constitui-se numa espécie de **ferramenta harmonizadora**, produzindo acordes pela replicação e fusão de (bandas de) frequências. Mais estudos dirigidos às leis de formação de acordes devem ser feitos tal que possa-se utilizar deste efeito harmonizador com um maior grau de controle sobre o processo.

Nas análises realizadas verificou-se que os tons gerados por esse processo estavam inscritos na escala cromática, i.é, eram percebidos como notas afinadas, em harmonia com outros tons (veja por exemplo a análise de ataques normais de flauta, nota si4).

Análise tempo-escala

Uma alternativa útil para se interpretar os efeitos modulantes gerados pelo processamento na reconstrução *wavelet* é abordar cada nível da análise em multiresolução como uma versão do sinal original vista sob uma determinada escala. Oriundos de uma análise de tempo-escala, os níveis corresponderiam a projeções do sinal original vistas sob uma escala específica. Esta é por sinal a essência da análise tempo-escala.

A geração de escalas no algoritmo utilizado é baseado em **grades diádicas**, isto é, as escalas são potências de 2, e são obtidas por um processo simples de filtragem e decimação reduzindo a resolução pela metade enquanto dobra a escala. Na decimação, certas razões das frequências presentes originalmente podem ser realçadas, num fenômeno de *aliasing*. A baixa eficiência da filtragem em eliminar as frequências causadoras de *aliasing* é um problema detectado neste esquema de decomposição. Na síntese, quando se geram também os sinais dos níveis, ocorre o imageamento (ou replicação) de bandas, o qual leva definitivamente às raias espelhadas entre si ao redor de uma frequência que sempre é uma razão racional da frequência de amostragem, e geralmente apresentando amplitudes diferentes.

Assim vibratos (de amplitude, para a flauta) poderiam estar facilmente visíveis em alguns níveis específicos, como se eles modulassem mais certas faixas de frequências

do espectro. Desse ponto de vista, modulações de amplitude podem "residir" nalgumas escalas, enquanto ausentam-se noutras.

O fato de que a transformada é reversível (desde que se não se edite os coeficientes entre a decomposição e a reconstrução) nos assegura que não existe um erro *a priori* no algoritmo. Como dito anteriormente, o banco de filtros de reconstrução realiza o processamento inverso do banco de decomposição, e os coeficientes dos filtros são escolhidos de maneira a cancelar os efeitos de *aliasing* e distorção, permitindo uma reconstrução perfeita. A equipe do pacote WaveLab, da Universidade de Stanford, foi informada sobre o efeito de imageamento gerado pela rotina de análise em multiresolução (`plotmra.m`), produzindo novas frequências e um efeito de modulação de amplitude sobre os sinais dos níveis, mas até o desfecho final deste texto não havia comentado o fenômeno.

Reconstruções seletivas e compressão de dados

Verificou-se em todas as análises que os níveis mais baixos, usualmente do 1º (primeiro) nível mais grosso ao 9º (nono) nível mais grosso⁵⁰, não contribuem significativamente ao conteúdo energético do sinal original, e reconstruções realizadas desprezando-se os coeficientes destes níveis (zerando-os) mostraram que não há perda significativa na qualidade do som, seja na textura do timbre, ou na informação quanto a eventos musicais existentes na amostra. Em algumas análises esta faixa de insensibilidade foi mesmo maior, alcançando também o 10º (décimo) nível mais grosso. Em todas as análises a eliminação dos coeficientes do 1º (primeiro) ao 8º (oitavo) nível mais grosso para efeito de reconstruções seletivas não provocou nenhuma perda de qualidade perceptível do som, igualando-se estas aos sinais originais. A eliminação destes níveis consiste numa forma direta (e trivial) de compressão de dados, o que é possibilitado pela natural capacidade descorrelacionadora da transformada *wavelet*, reduzindo a entropia presente no sinal pela codificação. Para sinais de extensão 262144 pontos amostrais (2^{18}), desprezar-se coeficientes do 1º ao 9º níveis mais grossos implica na eliminação de somente 1024 pontos amostrais, o que representa uma redução de

⁵⁰ contando-se do nível mais grosso em direção aos níveis mais finos.

aproximadamente 0,4% na quantidade de dados. Para sinais de extensão 131072 (2^{17}) a redução seria de 0,78%, e para sinais de extensão 65536 (2^{16}), de 1,56%.⁵¹

Os níveis mais finos *-em particular o nível mais fino-* são muito importantes para efetivar a transmissão das informações musicais contidas nos gestos e eventos musicais, e também para garantir a preservação da qualidade do timbre, sua constituição espectral e textura perceptível. A eliminação destes níveis em reconstruções seletivas trouxe perdas significativas, alterando por completo o som. Todavia, conforme verificado em várias reconstruções seletivas, a eliminação somente do nível mais fino *-especialmente para wavelets de análise mais regulares-* não leva a perdas significativas na informação gestual, e o sinal reconstruído eventualmente não apresentará diferenças relevantes na sua qualidade sonora em comparação ao original⁵².

A eliminação dos níveis mais finos para efeito de reconstrução seletiva usualmente leva a uma perda das informações codificadas nas bandas mais agudas, e influi principalmente na textura sônica percebida, gerando sons "ásperos" ou "granulosos", embora tal efeito seja sutil, e seja mais evidente para a *wavelet* de análise D4 do que para D16. Este fato nos leva às seguintes constatações:

- (1) o nível mais fino contém detalhes que são importantes na textura final do timbre, e
- (2) o nível mais fino apresenta amplitudes médias e energia inferiores aos níveis menos finos subsequentes, e sua ausência numa reconstrução altera sutilmente a qualidade do som final, sem no entanto implicar numa perda significativa de qualidade.

A percepção deste nível como um ruído agudo, com pouca informação audível sobre os eventos musicais tocados, é compreensível, haja visto estarmos submetendo o nosso sistema auditivo ao som do nível mais fino e agudo da análise, o qual não apresenta componentes nas faixas de maior sensibilidade do sistema auditivo (1000 a 3000 Hz).

⁵¹ A eliminação do 9º nível mais grosso para sinais dessa extensão entretanto já produz maiores perdas. Evidentemente que outras técnicas adicionais tratando os coeficientes dos níveis superiores devem ser empregadas a fim de se utilizar *wavelets* para finalidades de compressão de dados, produzindo taxas de compressão muito mais elevadas (acima de 100:1).

⁵² A perda será maior quanto menos regular for a *wavelet* e mais alto em frequência for o registro da nota musical, e usualmente reflete-se numa textura granulosa imposta ao sinal.

Os mecanismos de transmissão da energia acústica/mecânica nos fluidos auditivos e na membrana basilar apresentam uma atenuação de frequências nas faixas mais altas, implicando numa menor sensibilidade nesta faixa devido à menor capacidade de transferência de energia para as células ciliadas, as "células nobres do ouvido interno" que são os transdutores da energia mecânica em pulsos eletroquímicos neurais (p.261, cap.30, ^[29]). Enviar ao ouvido o nível mais fino de uma análise *wavelet* significa enviar um som contendo somente componentes que vão sensibilizar uma região específica e limitada da membrana basilar: aquela próxima à base da membrana, onde a resolução em frequência é menor, bem como o mecanismo de interpretação é mais complexo. Também é maior a impedância mecânica da membrana à transmissão de frequências altas, repercutindo numa menor sensibilidade.

Eventualmente, quando os tons em análise situavam-se num registro mais alto (oitava 4 ou 5) as reconstruções eliminando o nível mais fino (e também o segundo nível mais fino) provocaram o aparecimento de vozes adicionais devido à alteração na organização original dos modos da série harmônica (principalmente no que diz respeito às amplitudes relativas dos modos). O efeito direto disso está na quebra da fusão das linhas/bandas espectrais em um tom e timbre únicos.

Os níveis finos intermediários *-entre o segundo nível mais fino e os 9 primeiros níveis mais grossos-* frequentemente coincidem com os níveis mais energéticos (aqueles que concentram em conjunto mais de 99% da energia do sinal), e sua eliminação em reconstruções seletivas alteram significativamente o sinal reconstruído, seja pela atenuação das informações musicais, seja pela perda do timbre único perceptível do instrumento. A eliminação de um só deles usualmente produz um som onde a sua ausência reflete-se na geração de vozes (tons) adicionais, de timbres diferentes do timbre do instrumento original. Se o nível concentrar consideravelmente informações sobre algum gesto importante, geralmente ocorrerá atenuação na percepção deste. A percepção disso pode entretanto numa primeira audição ficar mascarada, porque o efeito da deterioração do timbre original e a geração de tons adicionais é mais proeminente.

Usualmente, quanto maior for a diferença dos pesos de reconstrução de níveis vizinhos, maior será a desintegração do timbre e mais perceptível será o efeito harmonizador, pela geração de vozes mais proeminentes.

A eliminação de dois níveis adjacentes geralmente produz um arranjo tonal diferente, corrigindo os efeitos causados quando um só nível é desprezado. Isto pode ser verificado por um simples experimento, qual seja o de analisar uma senóide pura na faixa de áudio, analisá-la com o sistema de AWMR, e em seguida reconstruí-la desprezando um nível intermediário. No som reconstruído usualmente aparecerão tons adicionais, diferentes da frequência da senóide original. Se se realizar outra reconstrução desta vez eliminando mais um outro nível adjacente observa-se que os tons da reconstrução anterior são atenuados ou mesmo eliminados, e outros surgem. Eventualmente, dependendo da frequência da senóide e dos níveis que se edita, a produção de tons adicionais pode ser bastante atenuada se dois níveis vizinhos são desprezados, em oposição a um só.

A eliminação de níveis intermediários alternadamente (isto é, zera-se um e mantém-se o outro) usualmente produz reconstruções onde novas vozes (tons) surgem em acorde e o timbre original não é reconhecível. Quanto mais níveis são eliminados menor será a percepção do timbre original e a perda da informação musical.

Isto mostra que a transformada *wavelet* implementada por este algoritmo pode transformar um som completamente pela simples edição de seus coeficientes, alterando sua composição espectral, a fusão dos modos harmônicos, os eventos musicais e o timbre percebidos. O potencial para trabalhos de síntese é grande, mas o controle deste processo não é trivial, e pode necessitar de manipulações espectrais adicionais. A escolha da *wavelet* ou conjunto de *wavelets* apropriadas também é crítica.

Diferenças relevantes entre a análise com D4 versus D16

As análises com D4 mostraram ataques melhor caracterizados nos níveis mais finos, em particular no nível mais fino, onde foi observado uma maior contribuição do conteúdo harmônico do tom em relação ao conteúdo harmônico observado para este nível na análise com D16. O nível mais fino da análise com D4 cobre frequências mais baixas não cobertas pelo nível mais fino da análise com D16, o que é perceptível pela audição comparativa dos mesmos. Seu som apresenta ainda uma textura menos aguda que o som do mesmo nível para D16.

Isto mostra que *wavelets* diferentes encampam diferentes propriedades e detalhes do sinal por apresentarem bandas passantes de formato e alcance também diferentes. D4 é

menos seletiva, e permite a passagem de componentes que são atenuadas por D16, alterando o conteúdo frequencial de cada nível da análise, e conseqüentemente o seu som.

Evidentemente, a exclusão do nível mais fino da análise com D4 em reconstruções seletivas levará a uma maior degradação do timbre, percebida como uma “granulosidade” ou “leve aspereza” na textura sônica, e eventualmente ocorrerá a geração de vozes pela desintegração dos harmônicos mais altos da série harmônica encampados na sua banda de passagem (efeito harmonizador).

A avaliação do padrão exibido de distribuição de energia linear é praticamente a mesma tanto para a *wavelet* de análise D16 quanto para a *wavelet* de análise D4. No entanto, quando observamos o gráfico de distribuição de energia em escala logarítmica, verificamos que a energia se distribui mais uniformemente pelos níveis de uma análise com D4, numa gradação mais suave entre a energia de um nível e do nível adjacente (degrau menos alto), como pode-se verificar comparando os gráficos de distribuição (em escala logarítmica) dos mesmos sinais analisados ora com D4, ora com D16.

Um outro aspecto a ser notado é a melhor resolução propiciada pelas *wavelets* mais regulares nos sinais dos níveis mais grossos. D16, por exemplo, por ser mais regular, é suave o suficiente para permitir ainda uma boa aproximação do sinal original nestes níveis, com formas de onda também mais regulares, e ampliando as possibilidades de extrair informações relevantes sobre os eventos e estruturas musicais nos níveis mais baixos.

O formato das estruturas em níveis mais baixos (em especial do 1º ao 6º níveis mais grossos) ficam muito semelhantes à *wavelet* utilizada, tornando menos óbvia a extração de informações musicais do sinal quando a *wavelet* é menos regular. D4, por exemplo, leva a representações do sinal menos suaves, já insinuando a sua forma de onda mais intensamente.

Considerações sobre a qualidade e a intenção das gravações

Numa produção musical existe sempre a preocupação em se valorizar determinados aspectos e componentes musicais de forma a dar a forma desejada ao som, atendendo a exigências ou especificações do compositor, do produtor e dos instrumentistas. Frequentemente o engenheiro de áudio deve agir no sentido de propiciar o destaque

adequado a determinado instrumento, capturar detalhes especiais, sons acessórios e amplificar estruturas musicais pouco perceptíveis normalmente. Para tanto, faz uso de diferentes técnicas de microfonação, gravação, edição e processamento do som, que produzem substancial impacto sobre a qualidade da produção final.

Cabe notar que cada região de um instrumento musical irradia um som diferente em amplitude e composição espectral. Numa flauta, o som proveniente do bucal é bem diferente daquele que se ouve no pé do instrumento. Num violino, o som das cordas apresenta um alcance dinâmico e qualidade tímbrica muito diferente do som que é irradiado nas aberturas em *f*. Por exemplo, captar-se o som de um fraseado apontando um microfone direcional para o bucal da flauta permite a captura dos mínimos ruídos de origem bucal, lingual e respirações, enquanto a captura do som irradiado pela extremidade oposta (o pé) valoriza algumas faixas mais graves da série harmônica e excluem substancialmente os ruídos e transientes ouvidos no bucal.

As amostras sonoras que foram analisadas neste trabalho foram gravadas com um microfone direcional, extremamente capaz de capturar estruturas sonoras normalmente não percebidas em audições de auditório, como ruídos bucais, ruídos produzidos pela mudança de dedilhados, respirações, e também variações de posição.

Para as gravações de violino, o microfone foi posicionado a uma distância entre 40 e 60 cm do instrumento, em altura do chão aproximadamente igual à altura do instrumento, e com foco centralizado sobre o cavalete. Esta microfonação permitiu a captura de ruídos dos gestos interpretativos, da irradiação das cordas e aberturas em *f*, arraste do arco, e eventualmente respirações da intérprete.

Para as gravações de flauta, o microfone foi posicionado entre 40 a 60 cm do instrumento, em altura ligeiramente superior à altura do instrumento em relação ao piso, apontado para a cabeça em ângulo de aproximadamente 45° com o eixo da flauta, com foco no tubo cerca de 6 a 8 cm abaixo do bucal. Isto permitiu a captura do sopro, da irradiação mais "brilhante" (faixas mais agudas), e também de estalidos, gestos bucais diversos, respiração e ruídos de dedilhado.

Originalmente, gravações foram realizadas em quatro posições: uma com foco sobre o bucal, outra com foco sobre o corpo do instrumento (onde se localizam as teclas), outra com foco na região da abertura ao fim do tubo (pé), e na posição padrão descrita acima. A escolha por adotar esta última conciliou diversos aspectos no que tange ao que se

desejava evitar em excesso (ruídos bucais, respiração, estalidos intensos, movimentação da embocadura, etc.) e aquilo que desejava-se capturar (“brilho” sonoro, sopro, timbre característico da flauta transversa).

Verificou-se que a adoção de outras posições para o microfone exerce substancial influência sobre a captura de gestos musicais, e conseqüentemente sobre a capacidade de analisá-los e correlacioná-los com as técnicas de interpretação. Se a intenção da análise repousa sobre objetivos didáticos e estudos de ordem musicológica, onde a avaliação dos gestos mais discretos reveste-se de importância para melhorar a técnica de interpretação, a microfonação a ser utilizada deve focalizar os pontos dos instrumentos onde ocorre a excitação.

Se, no entanto, a intenção é analisar o timbre geral e o impacto de gestos musicais sobre a expressividade percebida e variações no tempo do espectro irradiado, as melhores opções aconselham evitar o foco nas regiões de excitação e manipulação de dedilhados, dispor o microfone em ângulos menor ou igual a 45° com o eixo do instrumento (aproximando-o de um eixo paralelo), adotar distâncias maiores tal que se capte a irradiação proveniente de vários pontos, e mesmo utilizar microfones com características direcionais alternativas (respostas polares diferentes, como padrões cardióide, supercardióide, etc.).

Enumeração alternativa para os níveis

Uma forma alternativa de se enumerar os níveis de uma AWMR, atendendo simplesmente a uma maior facilidade notacional, consistiria em tomar o **nível mais fino** como o **primeiro nível (nível 1)** e o menos fino ou menos resolvido como o nível X (se a amostra original possuir 2^X pontos). Esta nomenclatura alternativa poderia facilitar a referência a níveis correspondentes de análises de sinais diferentes que possuam extensões também diferentes, agilizando a identificação dos diversos níveis da análise.

Esta forma de enumeração é mais prática no sentido de ser independente do número de níveis ou da extensão da amostra sonora, uma vez que o nível mais fino sempre será o nível 1 e os outros subsequentes são numerados em ordem crescente, até o nível menos fino. A tabela de distribuição de gestos/eventos musicais por níveis *wavelets*, compilada no item conclusivo a seguir, utiliza preferencialmente esta ordenação para enumerar os níveis de uma AWMR.

Considerações sobre o WaveLab

O pacote WaveLab atendeu como ferramenta aos objetivos propostos de análise de gestos musicais. Suas vantagens principais estão na portabilidade dos códigos e rotinas para diversas plataformas computacionais que suportam o MATLAB, e na filosofia adotada de permitir a reprodução da pesquisa e seus resultados eletronicamente, disponibilizando não somente o documento em texto, mas o *software* em si. Adicionalmente, os algoritmos codificados baseiam-se em implementações discretas da transformada com bancos de filtros, um esquema veloz e com baixo custo computacional.

Desvantagens estão nas limitações da plataforma MATLAB, principalmente em relação ao esquema de entrada e saída de dados, de sonificação dos sinais (especialmente em plataformas PC's), do gerenciamento de memória e na tradução dos códigos em MATLAB para códigos executáveis de baixo nível. O MATLAB é indubitavelmente uma plataforma adequada para experimentações, pesquisas e validações, não para implementações de sistemas endereçando a otimização de velocidade e eficiência de entrada e saída. Adicionalmente, o algoritmo piramidal utilizado pelo WaveLab produz efeitos de imageamento na geração dos níveis *wavelets* de um AWMR ou quando há a edição dos coeficientes entre a decomposição e reconstrução.

Existem atualmente diversos outros pacotes e/ou sistemas computacionais para análise *wavelet* disponíveis tanto na Internet (gratuitamente) quanto para aquisição (sistemas comerciais), cada qual com suas próprias características, vantagens, desvantagens, e orientação a aplicações e máquinas específicas. Entre muitos podemos citar o **Wavelet Toolbox** para MATLAB da MathWorks (<http://www.mathworks.com>), o **WaveBox**, etc. As alternativas existentes para se implementar algoritmos *wavelets* são muitas, e a escolha de um sistema ou paradigma de implementação deve atender às necessidades e requisitos do projeto a se desenvolver. As características que motivaram e justificaram o uso do pacote WaveLab neste trabalho atenderam ao propósito desta pesquisa acadêmica, mas não atenderão diversos requisitos práticos para implementação de sistemas de análise mais robustos, portáteis, adaptados ao processamento de sinais específicos, e orientados para operação em tempo real, devido principalmente a limitações em velocidade, dependência de outro pacote computacional (MATLAB),

limitações da linguagem, deficiências na documentação, inadequação da circuitaria eletrônica (projeto do *hardware*) e do algoritmo para aplicações específicas.

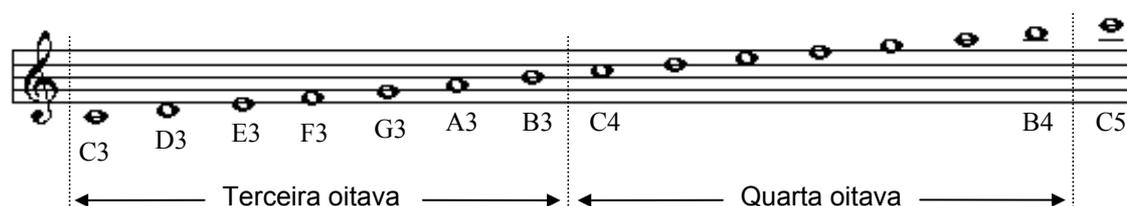
5.3 Conclusões

As análises realizadas não objetivaram uma caracterização completa dos eventos e gestos musicais abordados. A análise qualitativa visual e auditiva dos níveis é o principal objeto de interesse no momento, fornecendo subsídios para fundamentar trabalhos futuros de análise e síntese de timbres musicais.

Sem dúvida alguma a análise com *wavelets* comprovou a complexidade de um tom natural de violino: as estruturas periódicas e transitórias em cada faixa de frequências (nível *wavelet*) possuem uma evolução não trivial, sensível aos gestos musicais cujas propriedades mudam continuamente (pressão de arraste, dedilhado, velocidade, etc.), cultuando a variabilidade da música. Semelhante constatação se aplica à flauta.

A maioria dos eventos amostrados compreenderam notas musicais situadas nas 3^a e 4^a oitavas de frequências segundo o padrão de referência adotado para a escala temperada.⁵³ A figura abaixo mostra, em partitura na clave de sol, a extensão deste intervalo que compreende essas duas oitavas. O dó central do piano (C3) corresponde ao primeiro dó na flauta transversa (o dó mais grave); o lá padrão, em 440 Hz, é o lá3 (A3), também indicado na figura.

Figura 49 - Intervalo das terceira e quarta oitavas da escala temperada (registro em clave de sol)



É importante frisar o intervalo de frequência (ou registro tonal) onde a maioria dos eventos aqui analisados se situam, porque um deslocamento de uma (ou mais) oitava(s) para cima ou para baixo implicará diretamente num deslocamento dos padrões sônicos

⁵³ A escala temperada é apresentada no Glossário.

de um ou mais níveis, para cima ou para baixo, alterando a identificação e a percepção dos componentes e estruturas musicais.

As notas distribuem-se por vários níveis, mas a concentração de suas características e energia em certos níveis é função do seu registro tonal: notas mais graves se concentrarão em níveis mais baixos do que as notas mais agudas. Se duas notas possuem a mesma expressividade e foram geradas com uma mesma dinâmica, mesmo tipo de ataque, modulações e efeitos, então elas possivelmente terão descrições semelhantes numa sequência de níveis consecutivos da análise *wavelet*. Entretanto, se foram tocadas em oitavas diferentes, suas sequências consecutivas não coincidirão em ocupar os mesmo níveis, exatamente por elas ocuparem registros tonais diferentes, estando deslocadas entre si de um certo número de níveis, para cima ou para baixo.

Por exemplo, transientes presentes nos 7^o e 8^o níveis mais finos, localizando um ataque acentuado de uma nota mi₄ (E₄) em violino, vão se transferir para os 8^o e 9^o níveis mais finos caso a nota acentuada seja um mi₃ (E₃), ou para os 6^o e 7^o níveis mais finos, caso seja um mi₅ (E₅).

Na maioria das vezes as notas - tomadas as análises dos períodos de sustentação, excluindo-se os períodos transitórios do ataque e colapso - distribuem-se numa AWMR concentrando quase a totalidade de sua energia em 4 níveis consecutivos, possuindo pouca energia distribuída nos outros níveis da análise.

Estas são características da análise *wavelet* que a aproxima do tipo de representação usada em partituras, onde a frequência (tons/registo) é organizada verticalmente, e o tempo horizontalmente. Os tons se distribuem numa escala vertical, do nível mais grosso (grave) até o mais fino (agudo), e o tempo se distribui na escala horizontal.

Todas as notas executáveis tanto na flauta como no violino apresentam séries harmônicas descritas nos primeiros 7 níveis mais finos de uma AWMR. A nota mais baixa executável - *um sol₂ (G₂) no violino, com pitch percebido em 196 Hz (na prática, a fundamental)* - apresentará uma série harmônica distribuída a partir do 7^o nível mais fino da análise, que cobre esta faixa de frequências. Os níveis mais baixos fundamentalmente não apresentam contribuições das séries harmônicas.

Em acentos, subidas e descidas observadas nas fases de ataque apresentam altas taxas de variação na amplitude, refletindo-se em estruturas como rampas íngremes nas representações gráficas do sinal nos primeiros oito níveis mais finos da análise.

O nível mais fino usualmente contém a maior parte dos ruídos provocados pela excitação dos instrumentos, seja o ruído do arraste do arco no violino, ou o ruído do sopro na flauta.

Um aspecto importante é que este nível quase não fornece informação auditiva para se determinar o tom de notas diferentes situadas até a 3^a oitava da escala musical, frequentemente exibindo-as com mesmo *pitch* e qualidade de timbre idênticas, a de um ruído áspero e agudo. A causa fundamental deste efeito está na contribuição irrelevante da série harmônica do tom neste nível, impedindo a sua identificação. Para notas situadas em registros mais altos (acima da 4^a oitava) verifica-se alguma contribuição da série harmônica no nível mais fino, e a diferenciação das tonalidades torna-se possível.

Transientes associados a gestos de sopro e arco mostraram-se presentes nos três primeiros níveis mais finos, no 7^o e 8^o níveis mais finos, e eventualmente no 9^o nível mais fino em análises de acentos.

O 7^o nível mais fino para as análises de violino mostrou-se capaz de revelar sons e processos sônicos em ação na caixa ressonante do instrumento, merecendo a denominação de nível da madeira em função desta propriedade.

O 8^o nível mais fino para a maioria das análises mostrou concentrar ruídos de baixa frequência, normalmente ruídos de fundo ambiente e de quantização.

Os níveis mais energéticos são aqueles que conjuntamente concentram mais de 99% de toda a energia do sinal analisado. Usualmente apresentam quantidade de energia variando de 1 a 4 ordens de grandeza acima dos níveis menos energéticos, a diferença sendo maior em relação aos níveis mais grossos abaixo, e menor em relação aos níveis mais finos não energéticos acima. Os gráficos de distribuição de energia em escala linear permitem uma rápida identificação destes níveis.

A audição dos níveis mais energéticos mostra serem eles os que concentram a maior contribuição da série harmônica dos sinais musicais, isto é, onde a maior parte da energia das vibrações musicais do instrumento se concentra. Adicionalmente, são também nestes níveis onde ficam menos aparentes transientes, alterações bruscas e “descontínuas” associadas a gestos súbitos e intensos. Tais objetos são melhor visualizados em outros níveis, acima e abaixo dos mais energéticos.

A segmentação das notas num fraseado analisado vai melhorando do nível mais fino em direção aos menos finos abaixo, até atingir sua melhor forma nos níveis mais

energéticos, onde a regularidade da forma de onda e suavidade do envelope são maiores, e piorando novamente nos níveis mais grossos subsequentes.

O nível mais energético revela o envelope mais regular, formas de onda suaves, e os tons mais puros, de percepção clara e límpida, sem *clicks* ou estalos.

Dois fatores concorrem para esta percepção auditiva espetacular dos níveis mais energéticos: o fato de as componentes harmônicas mais intensas e influentes sobre o *pitch* percebido dos tons se situarem nas faixas de frequências cobertas por estes níveis, o que leva a um evidente destaque da composição harmônica (estacionária) presente, e o fato de usualmente os níveis mais energéticos cobrirem a faixa de frequências onde é melhor a audibilidade humana: entre 1000 e 3000 Hz (KOBRAÇ apud ^[29]).

De posse das análises de eventos e gestos musicais expressivos realizados para interpretações de flauta e violino, e após a avaliação de seus resultados mais relevantes, pode-se traçar as seguintes conclusões gerais:

1. A análise *wavelet* em multiresolução (AWMR) é útil para separar ruídos (agudos e graves), transientes, e a porção estacionária e mais energética normalmente associada à série harmônica das vibrações musicais.
2. É útil para identificar e isolar em níveis específicos os gestos musicais (gestos de sopro, bucais, de arco e dedilhados) produzidos em adornamentos, efeitos, ataques e fraseados expressivos
3. É útil para analisar o sinal em bandas de frequências diferentes, acompanhar a evolução dos sinais musicais em cada uma delas, e extrair informações importantes sobre eventos e processos vibratórios distintos que ocorram em faixas de frequência diferentes.
4. É útil para estudar articulações velozes e complexas, com trinados, trêmulos, frulatos e outros efeitos, e identificar os níveis onde seus efeitos são mais sensíveis.
5. É útil para determinar a presença de alguns tipos de eventos musicais expressivos numa passagem musical, como *staccatos*, *spiccatos*, acentos, variações dinâmicas e vibratos.
6. É útil como ferramenta para se aferir a qualidade do timbre de instrumentos complexos e de sofisticada manufatura (como violinos).

7. É útil como ferramenta modificadora de timbres e seus aspectos expressivos, para efeito de síntese
8. É útil como ferramenta compressora de informação e codificadora de som. Apresenta adicionalmente potencial para aplicações em modeladores de ruídos de quantização (*dithering*) em virtude de sua capacidade de segregá-los.
9. É útil para identificar e separar os níveis de resolução ou oitavas mais energéticas de um evento musical. Notas mais altas em frequência (num registro mais agudo) apresentam maior concentração de energia em níveis mais altos (finos), e as mais baixas em níveis inferiores (menos finos). Um deslocamento do padrão de distribuição da nota pelos níveis será tão mais evidente quanto maior for o intervalo tonal em questão. A grosso modo, uma diferença de uma oitava para baixo implicaria no deslocamento do padrão de distribuição de um nível para baixo (em direção aos níveis mais grossos).
10. A análise com *wavelets* exibe a propriedade de isolar transientes, caracterizar acentos e identificar gestos de sopro e de arco; a ocorrência do fenômeno de imageamento de bandas produz ainda um prático efeito harmonizador com algumas aplicações musicais em potencial.

Finalmente, a tabela abaixo relaciona os objetos sônicos associados aos gestos/eventos musicais por níveis *wavelets*, procedendo a um **sumário dos resultados das análises a partir do ponto de vista dos níveis *wavelets***, isto é, tomando-se cada nível como referência e relatando-se os objetos e estruturas que nele se destacaram ou concentraram.

A primeira coluna compila uma lista de sensibilidade (insensibilidade) do nível a determinados objetos/padrões sônicos. A segunda coluna sumaria os aspectos e objetos que o nível destaca ou concentra. A terceira coluna fornece indícios sobre como o nível é percebido, gráfica e/ou sonicamente. Finalmente a última busca uma mensuração qualitativa quanto ao impacto que a ausência do nível terá sobre a reconstrução do som.

Tabela 2 - Distribuição de gestos/eventos musicais por níveis *wavelets*

Nível	Sensível a (↑) Insensível a (↓)	Concentra ou destaca	Percepção	Impacto de sua ausência em reconstrução
1º	↑ foco de embocadura, variação de sopro, dinâmica (intensidade dos tons), localização do instante de ataques p/ trêmulos dedilhados ligados (flauta). ↓ variação tonal de glissandos.	ruído de fundo, ruído de sopro, ruído de arco, ruído de gestos bucais, acentos; série harmônica (para tons altos); intensifica parciais mais altos em ataques pontuais e energéticos (violino); baixo poder de segmentação das notas em fraseados ligados.	ruído agudo; fase inicial transitória da excitação das vibrações dos tons; timbres originais não reconhecíveis; diferenças tonais (para notas diferentes) são imperceptíveis (exceto para notas altas); <i>pitches</i> dos tons imperceptíveis.	som áspero, granuloso, mais evidente se D4 for usada.
2º	↑ variações transitórias na pressão de sopro, foco de embocadura, sobrelevações de acentos <i>sforzato</i> (violino), ataques <i>sforzato</i> muito rápidos (<8ms), gestos espúrios indesejados portamentos, rangidos, movimentos bruscos de arco) (violino).	ataques <i>staccatos</i> , acentos, gestos bucais, chaveamento de sopro; parte alta da série harmônica (principalmente p/ tons na 4 oitava).	padrão característico do ataque (rampa, sobrelevação e decaimento); acentos; <i>pitches</i> (tonalidades) são perceptíveis, notas são diferenciáveis em tonalidades.	perdas na constituição harmônica; geração de vozes (tons em função harmônica).
3º	↑ sobrelevação destacada para acentos (<i>sforzato</i>), ataques <i>sforzatos</i> muito rápidos (<8ms).	acentos; região de sustentação (flauta); séries harmônicas; caracteriza curvas da região de colapso (violino).	padrão característico do ataque (rampa, sobrelevação e decaimento); variações de sopro sobre a sustentação (variações de amplitude, flauta); envelope não é suave, apresenta muitos transientes e aspecto amorfo em fraseado não ligado com acentos p/ flauta).	perdas na constituição harmônica; geração de vozes (tons em função harmônica).
4º	-	série harmônica; objetos e estruturas moldados pela composição harmônica dos tons.		perdas na constituição harmônica; geração de vozes (tons em função harmônica).

5º	se entre os níveis mais energéticos: ↑ vibrações harmônicas; constituição das séries harmônicas.	série harmônica; regiões de sustentação; regiões de colapso reverberantes (quando corda é abandonada); destacam a evolução dos modos estacionários.	se entre os níveis mais energéticos: exibe textura sônica "limpa", envelopes e contornos mais suaves, regulares e uniformes (envelopes padrão ADSR).	perdas na constituição harmônica; geração de vozes (tons em função harmônica).
6º	↑ vibrações harmônicas (maior contribuição das séries harmônicas); sobrelevações de acentos (flauta); deslocamentos de ar (variação no sopro). ↓ transientes.	espectro harmônico; curvas dos colapsos bem caracterizadas (violino e flauta); se entre os níveis mais energéticos: destaque às flutuações e modulações de qualquer origem.	som mais puro, menos ruidoso, tons definidos (permite diferenciar tonalidades); admite diferentes agrupamentos sequenciais dos tons (evolução percebida dos tons).	perdas na constituição harmônica; geração de vozes (tons em função harmônica).
7º	↑ componentes da série harmônica (estados estacionários), acentos (<i>sforzato</i>), transientes de respiração, gestos de arco (ataques).	ataques <i>sforzato</i> ; curvas de colapso (flauta e violino); ataques não ligados; série harmônica.	"Nível da madeira"; sopro é percebido em reprodução a 11025 KHz; regiões de máxima amplitude localizam regiões centrais (sustentação) das notas (em trêmulo dedilhado, flauta).	perdas na constituição harmônica; geração de vozes (tons em função harmônica).
8º	↑ ataques <i>staccatos</i> , acentos (<i>sforzato</i>); abandono do arco à corda (violino); gestos produzindo variações de velocidade e contato de arco-corda; ↓ variações de dinâmica (violino), gestos bucais (estalidos e outros, flauta).	ataques <i>staccatos</i> ; gestos bucais; ruído de fundo (ambiente e/ou quantização); ataques <i>sforzato</i> (subida); transiente de ataques acentuados (flauta, notas na 3ª oitava); transiente de chaveamento de sopro no colapso.	sinal "contínuo", espalhado por toda a extensão do sinal, com propriedades ruidosas; textura ruidosa; ruído de fundo ambiente, ruído de quantização em gravações baixas.	-
9º	↑ transientes de ataques; transientes de sopro; transitórios em geral; alternância de sentido de arco.	ruído de fundo (ambiente e/ou quantização); gestos de sopro; ataques <i>sforzato</i> (subida); transitório de ataques acentuados (flauta).	ruído grave, textura ruidosa; os níveis abaixo deste contribuem pouco para a formação dos tons e eventos musicais; a participação energética é desprezível; e sua ausência não causa perdas qualitativas expressivas; ruído de quantização em gravações baixas.	-

10 ^o	↑ transiente de ataque; alternância de sentido de arco; sentido do arco (<i>staccatos</i> , violino); variação velocidade do arco; abandono do arco à corda.	gestos de arco; transientes de ataques acentuados (flauta).	ruído grave.	-
11 ^o	↑ gestos de arco (aumento de pressão, variação no contato arco-corda); padrões interpretativos (sucessão de gestos); ruídos de fundo (ambiente, processamento, quantização); dinâmica baixa em sinais ruidosos (baixa relação S/R).	contribuição do ruído de quantização (em sinais de baixa dinâmica); gestos bucais de notas muito acentuadas.	"negativo do sinal"; amplitude é menor nas regiões de sustentação quando o sinal apresenta contaminação por ruídos (baixa relação S/R e/ou dinâmica baixa).	desprezível
12 ^o	↑ variações de velocidade, pressão e contato arco-corda (a confirmar em trabalhos futuros); dinâmica baixa em sinais ruidosos (baixa relação S/R).	contribuição do ruído de quantização em baixa frequência (violino).	inaudível; deste nível para baixo o formato das estruturas, objetos, e formas de onda assemelham-se à wavelet de análise, tornando menos óbvia a extração de informação musical nestes níveis; "negativo do sinal" - em sinais com baixa dinâmica e/ou baixa relação S/R: amplitudes menores nas regiões de sustentação dos tons, e maiores nas áreas onde a dinâmica é menor.	desprezível para este nível e todos os demais abaixo (mais grossos).
13 ^o			todos os níveis deste para baixo são inaudíveis.	
14 ^o	-	-	-	-
15 ^o	↑ informações a nível musicológico (violino, a confirmar em trabalhos futuros).	indicativo de padrões musicais longos (repetições, progressões).		

5.4 Trabalhos Futuros

Mesmo populado de "pulsos" e picos locais, aparentemente distribuídos aleatoriamente sem revelar correlação com os eventos musicais mais importantes, o sinal aos 8º e 9º níveis mais finos pode sofrer algum tratamento ou filtragem no tempo (ou frequência) adicional que facilite a localização de gestos musicais. Um tratamento suplementar ao sinal poderia mostrar correlações importantes entre os picos intensos e eventos musicais. A verificação de tais hipóteses é deixada para um trabalho futuro, contando com a aplicação de técnicas adicionais e complementares ao processamento com *wavelets*.

A possibilidade de que os envelopes e pontos de máximos verificados aos 10º e 11º níveis mais finos forneçam alguma informação relevante sobre os gestos musicais deve ser melhor explorada em trabalhos futuros, envolvendo inclusive o uso de ferramentas algorítmicas adicionais no processamento do sinal nestes níveis.

Wavelet-packets podem ser úteis para segmentar melhor as largas bandas cobertas pelos níveis mais finos - *em especial o mais fino* - e permitir a separação e identificação de componentes e estruturas sônicas que estejam fundidas nestas escalas mais finas.

A separação de estruturas sônicas em tons inscritos em registros mais altos - *da 5ª oitava em diante* - deve ser mais eficaz com *wavelet-packets* do que com a presente análise *wavelet* em multiresolução simples, com separação de bandas por oitavas. Os tons mais altos naturalmente se decompõem e concentram sua energia nos níveis mais altos/finos, que cobrem faixas mais largas de frequência. Seus objetos sônicos e componentes harmônicos estarão portanto fundidos dentro de uma banda (escala) maior, inviabilizando a sua identificação e separação. A aplicação de *wavelet-packets* para implementar uma segmentação mais fina das bandas mais altas, poderia viabilizar a análise de processos que ocorram em bandas mais curtas, subescalas dentro de uma banda larga, e assim facilitar a separação de componentes sônicos dos tons mais altos.

Os níveis mais energéticos também devem ser alvos de análises com *wavelet-packets*, haja visto sua importância na formação dos tons.

As possíveis correlações entre a amplitude do sinal musical verificada no 8º nível mais fino das análises de violino e parâmetros de manipulação do arco, como velocidade e ponto de engate do arco na corda, ainda não estão bem descritas, e merecem um estudo

mais criterioso para se avaliar o impacto destes parâmetro sobre os sinais neste nível, e conseqüentemente as possibilidades de síntese expressiva decorrentes.

Wavelet-packets são mais indicadas para aplicações de rastreamento de eventos musicais associados com gestos musicais, porque possuem uma divisão mais fina do espectro, fragmentando o espectro em níveis com largura de banda menores e mais sintonizados.

Também de extrema importância é a escolha da base *wavelet* e do número de coeficientes de seu filtro. Filtros *wavelets* com mais coeficientes tendem a apresentar uma maior regularidade e suavidade na representação dos sinais, além de possuírem melhor localização em frequência. O aspecto de sintonia e cobertura em frequência dos filtros *wavelets* é relevante para implementar um bom sistema de análise e síntese de música com *wavelets*. Bases *wavelets* construídas de acordo com os modelos de representação auditiva das membranas da cóclea⁵⁴ podem se revelar mais adaptadas para a tarefa de interpretação (e síntese) de estruturas musicais.

Irino e Kawahara (1993) construíram bases *wavelets* que simulam as características de resposta em frequência do sistema auditivo periférico, utilizando-se para tal de um modelo coclear adicionado de um filtro de ouvido médio (para casamento de impedância)^[33]. Seu objetivo era o de conceber uma representação para o sistema auditivo periférico que também provesse um esquema para síntese de sinais para experimentos psicofísicos, entretanto a aplicabilidade dos filtros *wavelet* que produziram deve se estender à análise e síntese de música, possivelmente com vantagens sobre outros sistemas pela similaridade com o filtro auditivo real, abrindo assim frente nova de experimentações.

O uso de bases *wavelets* adaptativas e "dedicadas" (i.é, para uso específico orientado a classes específicas de sons) também poderiam ser outra frente de pesquisa interessante para a caracterização de gestos musicais expressivos. Abry e Aldroubi (1994) comentam que as *wavelets* de Daubechies, de Battle-Lemarié ou de *splines* são as mais comuns em implementações da transformada *wavelet* discreta, enquanto que para decomposições contínuas uma maior variedade de *wavelets* é encontrada. Motivados por essa constatação, propõem vários métodos para sintetizar *wavelets* desejadas (semi-ortogonais), mantendo a estrutura computacional piramidal recursiva e a grade diádica

de amostragem.^[1] Outras estruturas algorítmicas existem para ser exploradas no campo de aplicações musicais e da fala, como os bancos de filtros iterados. Blu (1993) afirma que os esquemas de bancos de filtros iterados (diádicos, como empregados neste trabalho) são "*muito grossos para analisar sinais como áudio de alta qualidade e fala*" porque realizam uma decomposição em bandas de oitavas, e sugere o uso de bancos de filtros iterados racionais, uma arquitetura diferente de computação onde o sinal é separado num conjunto de bandas de frequências mais denso (resolução em frequência mais fina).^[4] Sua abordagem, entretanto, também levanta problemas e limitações associados a essa arquitetura, a qual não implementa propriamente uma transformada *wavelet*.^[4]

Os sistemas capazes de reconhecer gestos musicais e investigar sua estrutura de formação encontram aplicação direta nos sistemas de síntese, para propósitos de reproduzir elementos das técnicas de interpretação instrumental de músicos humanos, provendo recursos para se introduzir expressividade e naturalidade, enriquecendo as possibilidades de síntese de música. Aplicações interessantes estariam em habilitar determinados instrumentos (timbres) a executar fraseados e efeitos característicos de outros instrumentos (timbres), fornecendo-lhes graus de liberdade expressiva adicionais. Muitos aspectos da síntese de timbres acústicos para interpretação de peças musicais estão altamente relacionados não somente à fisiologia da interpretação, i.é, aos aspectos funcionais e limitantes físicos, mas também ao estilo da composição e uma variedade de adornamentos, variações rítmicas, portamentos, dedilhados alternativos e pequenas alterações tonais e temporais que ela admite. Por exemplo, observa-se muitas vezes no sistema interpretativo do Século XIX para flauta sequências de notas colcheias tocadas diferentemente num arranjo alternativo com semínima pontuada, semi-colcheia e colcheias. Acelerações no tempo também eram apropriadas para passagens excitantes, enquanto retardos no tempo o eram para passagens subjugadas ou reprimidas. A articulação do fraseado onde os instrumentistas procuram emular cantores ou oradores também imprime uma dinâmica interpretativa própria à composição e modificam por completo sua realização, o produto sonoro e suas impressões psicológicas.

Estes aspectos constituem uma classe mais elaborada ou elevada de parâmetros controladores e moduladores para interpretações sintéticas, e devem ser melhor

⁵⁴ em especial a membrana basilar.

compreendidos e modelados recorrendo-se ao ferramental provido pela área de Inteligência Artificial e outras abordagens cognitivas, a fim de se quantificar e possibilitar sua reprodução, guiando a síntese a níveis de expressividade e beleza sonora mais aperfeiçoados.

Ansiedade, curiosidade, humor, fúria, expectativa, dúvida e outros objetos psicológicos podem ser abstraídos de uma execução acústica, e estão presentes na música através de um mapeamento que ainda não está bem definido e compreendido. Existe uma concordância entre vários pesquisadores de que a análise *wavelet* pode ser útil na investigação deste mapeamento, pela sua semelhança com o processamento de estágios primários no sistema auditivo humano, e por suas propriedades de localizar eventos no espaço tempo-frequência. A facilidade em se proceder a implementações algorítmicas discretas, de complexidade computacional da ordem do tamanho das amostras ($O(N)$), torna as implementações em dispositivos integrados (VLSI ou com DSP's) possibilidades viáveis e atraentes objetivando aplicações de alto desempenho e em tempo real. Diversas propostas de implementações de algoritmos *wavelet* em VLSI podem ser encontradas na literatura escrita, bem como na Internet. PARHI e outros (1993) propõe duas arquiteturas diferentes para implementação de transformadas *wavelet* discretas de uma dimensão (1-D) em VLSI.^[53]

SOBRE A APLICABILIDADE EM ANÁLISE MUSICOLÓGICA DOS NÍVEIS MAIS GROSSOS

Níveis abaixo do 7^o nível mais fino usualmente apresentam envelopes para as notas que não guardam semelhança com os envelopes dos níveis anteriores ou com o sinal original, apresentando formatos irregulares e não correlacionados com o modelo tradicional de envelope ADSR (ataque, *decay*, *sustain*, *release*) empregado em sintetizadores.

Os níveis ainda mais baixos, cobrindo faixas subsônicas, apresentam informações que poderiam justificar o seu emprego para analisar aspectos musicológicos, como por exemplo inferir sobre a intensidade de determinado trecho tocado em relação a outros. Todavia, somente a análise de partes orquestrais mais significativas e longas poderão mostrar sua eficácia neste sentido.

Apesar de serem descartáveis numa reconstrução, praticamente sem repercussões sobre a qualidade do som reconstruído frente ao original, os níveis mais baixos consistem em

médias aproximadas do sinal original e dos níveis anteriores mais finos, observadas em escalas maiores, onde os detalhes foram extraídos e restaram os contornos básicos, os descritores de envelopes, os esboços do sinal original. Por cobrirem faixas de frequências várias oitavas abaixo do registro energético da série harmônica do timbre, não participam da formação das vibrações harmônicas. Também não permitem rastrear os transitórios muito finos e eventos de alta frequência.

Encaixam-se nesta categoria usualmente os níveis a partir do 12^o nível mais fino, que exibem formas de onda por vezes completamente diferentes do envelope do som original. Reconstruções utilizando-se pesos para estes níveis até 15 vezes maior que 1 (1500%) mostraram que eles não repercutem significativamente nos aspectos de larga escala do som, como poderia se esperar. No entanto, verificações mais direcionadas, utilizando-se passagens musicais mais longas e endereçando sua influência sobre a dinâmica geral do trecho, seriam necessárias para se avaliar sua utilidade para efeito de síntese.

5.5 Epílogo

Extensa ênfase dada à apresentação da ferramenta - *a transformada wavelet e a teoria da análise em multiresolução* - bem como a sua história, aplicações, e tópicos correlatos à aplicação em música computacional - *a técnica de análise e síntese, os eventos e definições musicais* - deve-se a pouca literatura existente em língua portuguesa até o momento que os aborde em profundidade, principalmente quando abordados conjuntamente. As áreas de música computacional e engenharia de áudio são relativamente novas no Brasil, e o desenvolvimento de tecnologia de áudio é emergente no país.

Existe no mercado uma grande variedade de timbres musicais disponíveis em sintetizadores comerciais. A maioria destes timbres é gerada por técnicas convencionais como a síntese FM, a associação de unidades geradoras (*hardware e/ou software*), através de algoritmos descritivos (síntese de sinais), ou utilizando amostras de timbres naturais armazenados num banco de memória (*wavetable synthesis*). Muitos sintetizadores comerciais são desfavorecidos pela pobreza e artificialidade de seus timbres, por utilizarem técnicas que não permitem incrementar ou adicionar

expressividade, realismo e “*vida*” aos tons gerados, e por usarem modelos tímbricos simplistas e limitados.

A eventual adequação das *wavelets* como método para permitir a incorporação destes aspectos dinâmicos aos sistemas de síntese musical a tornaria uma ferramenta imprescindível em música computacional, tanto para a academia (musicólogos, maestros e compositores), como instrumento de auxílio à análise de orquestração em estilos contemporâneos, quanto para o sofisticado mercado musical e fonográfico, empregada em estações digitais de áudio, em processadores de som, módulos sintetizadores, teclados e outros instrumentos eletrônicos musicais.

O Brasil, pela sua relevante produção musical, ocupa lugar de destaque no setor musical do planeta. Sua música popular é extremamente rica e expressiva. Seus instrumentistas e compositores estão entre os mais capacitados, criativos e requisitados. No entanto, a produção de tecnologia que possibilite uma participação expressiva do país no setor produtivo da moderna indústria musical e de áudio está muito aquém das suas possibilidades e potencialidades. O país não tem tradição na produção de instrumentos musicais⁵⁵, nem de sofisticados equipamentos de áudio e vídeo para estúdios, teatros ou instalações a céu aberto.

Os grandes canais de televisão do país, entre os melhores do mundo, dispõem da mais moderna tecnologia para produção de áudio e vídeo, para o tratamento, manipulação e armazenamento de imagem e som, mas sua estrutura operacional está praticamente baseada em equipamentos concebidos e produzidos no exterior. Não existe até então no país nenhuma instalação industrial de porte que produza sintetizadores, teclados, instrumentos musicais eletrônicos, estações digitais de áudio e vídeo, processadores digitais de áudio, ou *software houses* que desenvolvam e comercializem produtos voltados para este nicho.

Existe, entretanto, no Brasil uma elite de profissionais capacitados a mudar este clássico perfil de importador de tecnologia musical, tanto na área técnica e de engenharia, quanto na área artística e musical. Existem já grupos organizados no sentido de se integrar a comunidade da música computacional, agregando músicos e engenheiros, artistas e cientistas, para fomentar a geração de tecnologia de áudio e incentivar o emprego de tecnologia para produção musical, mas talvez o direcionamento de suas

atividades e interesses não tenham ainda admitido massa crítica o suficiente para ultrapassar diferenças de ordem estética e cobiçar patamares mais estratégicos e desafiadores.

Estas questões por si só justificam o uso desta dissertação como veículo para difundir nos meios acadêmico e produtivo esta linha de pesquisa, mostrar as novas perspectivas do emprego direto da tecnologia computacional e eletrônica no segmento de áudio e música, e o insaciável mercado para tal tecnologia.

Esta contribuição atende diretamente a anseios do autor em estimular novos pesquisadores, investidores do mercado produtivo, engenheiros, músicos e profissionais da Informática à produzir tecnologia e contribuir para o desenvolvimento da integração arte e ciência.

⁵⁵ com exceção de instrumentos percussivos.

REFERÊNCIAS BIBLIOGRÁFICAS

- 1 ABRY, P.; ALDROUBI, A. Designing multiresolution analysis-type wavelets and their fast algorithms. **Journal of Fourier Analysis and Applications**, v.2, p.135-59, 1995.
- 2 ARFIB, D.; DELPRAT, N. Musical transformations using modification of time-frequency images. **Computer Music Journal**, v.17, n.2, p.66-72, Summer 1993.
- 3 BELKIN, A. Orchestration, perception, and musical time: a composer's view. **Computer Music Journal**, v.12, n.2, p.47-53, Summer 1988.
- 4 BLU, T. Iterated filter banks with rational rate changes: connection with discrete wavelet transforms. **IEEE Transactions on Signal Processing**, v.41, n.12, p.3232-44, Dec. 1993.
- 5 BREGMAN, A.S. **Auditory scene analysis: the perceptual organization of sound**. Cambridge, MIT Press, 1990. 773p.
- 6 BUCKHEIT, J. et al. About WaveLab. Stanford, Stanford University, Nov. 1995. **Internet**, "<ftp://playfair.stanford.edu/pub/wavelab/AboutWavelab.pdf>", Nov. 1995. 35p.
- 7 BUCKHEIT, J.; DONOHO, D. WaveLab architecture: Version 0.700. Stanford, Stanford University, Nov. 1995. **Internet**, "<http://playfair.stanford.edu/reports/wavelab/WaveLabArch.pdf>", 1996. 38p.
- 8 CHIANN, C.; MORETTIN, P.A. A wavelet analysis for stationary processes. /Submetido ao **Journal of Nonparametric Statistics**; sob revisão/
- 9 CHUI, C.K. **An introduction to wavelets**. San Diego, Academic Press, 1992. 226p.
- 10 CODY, M.A. A wavelet analyzer: an alternative to the FFT-based spectrum analysis. **Dr. Dobb's Journal**, p.44-54, 82-91, Apr. 1993.
- 11 CODY, M.A. The fast wavelet transform: beyond Fourier transforms. **Dr. Dobb's Journal**, p.16-28, 100-1, Apr. 1992.
- 12 DAUBECHIES, I. Orthonormal basis of compactly supported wavelets. **Communications on Pure and Applied Mathematics**, v.41, p.909-96, 1988.
- 13 DAUBECHIES, I. **Ten lectures on wavelets**. Philadelphia, SIAM, 1992. (CBMS-NSF Regional Conference Series on Applied Mathematics) 357p.
- 14 DAUBECHIES, I. The wavelet transform, time-frequency localization and signal analysis. **IEEE Transactions on Information Theory**, v.36, n.5, p.961-1005, Sept. 1990.
- 15 DE VORE, R.A.; JAWERTH, B.; LUCIER, B.J. Image compression through wavelet transform coding. **IEEE Transactions on Information Theory**, v.38, n.2, p.719-46, Mar. 1992.
- 16 DOERSCHUK, R.L. Bob Moog. **Keyboard**, v.21, n.2 (issue 226), p.92-100, Feb. 1995.
- 17 DOLSON, M. Fourier-transform-based timbral manipulations. In: MATHEWS, M.V.; PIERCE, J.R., eds. **Current directions in computer music research**. Cambridge, MIT Press, 1989. p.105-12.
- 18 DRAKE, L.A. et al. Wavelet analysis in recruitment of loudness compensation. **IEEE Transactions on Signal Processing**, v.41, n.12, p.3306-12, Dec. 1993.
- 19 EVANGELISTA, G. Pitch-synchronous wavelet representation of speech and music signals. **IEEE Transactions on Signal Processing**, v.41, n.12, p.3313-30, Dec. 1993. Special edition.

- 20 FARIA, R.R.; ZUFFO, J.A. Wavelets as a multiresolution analysis and synthesis technique for sound timbres edition. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO E MÚSICA, 2º / CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 15º, Canela, 1995. **Anais**. Porto Alegre, IIUFRGS, 1995. p.198-204.
- 21 FARIA, R.R.A.; RUSCHIONI, R.A.; ZUFFO, J.A. Wavelets in music analysis and synthesis: timbres analysis and perspectives. **Proceedings of SPIE**, v.2825, pt.2, p.950-61, 1996. /Apresentado ao Wavelet Applications in Signal and Image Processing IV; SPIE 1996, Denver, 1996/
- 22 FUTURE MUSIC. Somerset, n.38, Dec. 1995. /Internet, "<http://www.futurenet.co.uk/>"
- 23 FUTURE MUSIC. Somerset, n.39, Jan. 1996. /Internet, "<http://www.futurenet.co.uk/>"
- 24 GARNET, G.E. **An overview of CNMAT research**. Berkeley, California University, Center for New Music and Audio Technologies (CNMAT), s.d.
- 25 GOMES, S.M.; CORTINA, E. Wavelet transform: a local time-frequency analysis. In: SEMINÁRIO BRASILEIRO DE ANÁLISE, 40º, Rio de Janeiro, 1994. **Separatas**. Rio de Janeiro, UFRJ, 1994. p.95-142.
- 26 GRAPS, A. An introduction to wavelets. **IEEE Computational Science & Engineering**, v.2, n.3, p.50-61, Summer 1995.
- 27 GRIFFITHS, P. **A música moderna: uma história concisa e ilustrada de Debussy a Boulez**. Trad. de Clóvis Marques. Rio de Janeiro, Jorge Zahar, 1987. 206p.
- 28 GUYTON, A.C. **Tratado de fisiologia médica**. 7.ed. Rio de Janeiro, Guanabara Koogan, 1989. cap.61, p.582-90.
- 29 HUNGRIA, H. **Otorrinolaringologia**. 7. ed. Rio de Janeiro, Guanabara Koogan, 1995. cap.29-32, p.243-95.
- 30 HUTCHINS, C.M., comp. **Benchmark papers in acoustics/5: Musical acoustics, part I: violin family components**. Stroudsburg, Dowden, Halsted Press, 1975. 478 p.
- 31 INTERNATIONAL CONFERENCE ON WAVELETS: TIME FREQUENCY METHODS AND PHASE SPACE, Marseille, 1987. **Proceedings**. 2. ed. Berlin, Springer, 1989. 331p.
- 32 IRCAM. **Institut de Recherche et de Coordination Acoustique/Musique**, Paris, Aubin Imprimeur, 1994. /Impresso Informativo/
- 33 IRINO, T.; KAWAHARA, H. Signal reconstruction from modified auditory wavelet transform. **IEEE Transactions on Signal Processing**, v.41, n.12, p.3549-54, Dec. 1993.
- 34 JAFFE, D.A. An overview of criteria for evaluating synthesis and processing techniques. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO E MÚSICA, 2º / CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 15º, Canela, 1995. **Anais**. Porto Alegre, IIUFRGS, 1995. p.53-61.
- 35 JAWERTH, B.; SWELDENS, W. An overview of wavelet based multiresolution analysis. **SIAM Review**, v.36, n.3, p.377-412, 1994. /Disponível via Internet, "ftp://ftp.math.sc.edu/pub/imi_93/imi93_1.ps", Columbia, University of South Carolina, site do Dept. of Mathematics, 1997. 39p./
- 36 KADAMBE, S.; BOUDREAUX-BARTELS, G.F. Application of the wavelet transform for pitch detection of speech signals. **IEEE Transactions on Information Theory**, v.38, n.2, p.917-24, Mar. 1992. Special edition.
- 37 KARLSEN, R.E.; GERHART, G.R. et al. Wavelet analysis of ground vehicle acoustic signatures. **Proceedings of SPIE**, v.2491, pt.1, p.560-70, 1995. /Apresentado ao Wavelet Applications II, Orlando, 1995/
- 38 KRONLAND-MARTINET, R. The wavelet transform for analysis, synthesis, and processing of speech and music sounds. **Computer Music Journal**, v.12, n.4, p.11-20, Winter 1988.

- 39 LANCASTER, D. Hardware hacker: understanding transforms, video compression secrets, video crosshatch generator, power electronics resources and more wavelet breakthroughs. **Radio Electronics**, v.62, p.68-73, July 1991.
- 40 LEHRMAN, P.D. Microcomputer applications in music. In: BIRNES, W.J., ed. **PC hardware and systems**. New York, McGraw-Hill, 1989. p.75-92.
- 41 MALLAT, S.; HWANG, W.L. Singularity detection and processing with wavelets. **IEEE Transactions on Information Theory**, v.38, n.2, p.617-43, Mar. 1992.
- 42 MALLAT, S.G. A theory for multiresolution signal decomposition: the wavelet representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.11, n.7, p.674-93, July 1989.
- 43 MALLAT, S.G. Multiresolution approximation and wavelet orthogonal bases of $L^2(\mathbb{R})$. **Transactions of the American Mathematical Society**, v.315, n.1, p.69-87, Sept. 1989.
- 44 MATHEWS, M.V.; PIERCE, J.R., eds. **Current directions in computer music research**. Cambridge, MIT Press, 1989. 432 p.
- 45 MCGILL, K.C.; TASWELL, C. Length-preserving wavelet transform algorithms for zero-padded and linearly-extended signals. Palo Alto, Rehabilitation Research and Development Center, Veterans Affairs Medical Center, Mar. 1992. **Internet**, "<http://www.wavbox.com/LPWTA92.ps>", Nov. 1996. 20p.
- 46 MEYER, Y. Book reviews: *An introduction to wavelets* (Chui, C.K.) and *Ten lectures on wavelets* (Daubechies, I.). **Bulletin (New Series) of the American Mathematical Society**, v.28, n.2, p.350-60, Apr. 1993.
- 47 MEYER, Y. **Wavelets: algorithms and applications**. Philadelphia, SIAM, 1993. 133p.
- 48 MULDER, A. Virtual musical instruments: accessing the sounds synthesis universe as a performer. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO E MÚSICA, 1ª / CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 14ª, Caxambu, 1994. **Anais**. Belo Horizonte, UFMG, 1994. p.243-50.
- 49 NASON, G.P.; SILVERMAN, B.W. The discrete wavelet transform in S. **Journal of Computational and Graphical Statistics**, v.3, n.2, p.163-91, 1994.
- 50 NEWLAND, D.E. Harmonic and musical wavelets. **Proceedings of the Royal Society London A**, v.444, p.605-20, 1994.
- 51 NEWLAND, D.E. Harmonic wavelet analysis. **Proceedings of the Royal Society London A**, v.443, p.203-25, 1993.
- 52 OPPENHEIM, A.V.; SCHAFER, R.W. **Discrete-time signal processing**. Englewood Cliffs, Prentice-Hall, 1989. 879p.
- 53 PARHI, K.K.; NISHITANI, T. VLSI architectures for discrete wavelet transforms. **IEEE Transactions on VLSI Systems**, v.1, n.2, p.191-202, June 1993.
- 54 PELLMAN, S. **An introduction to the creation of electroacoustic music**. Belmont, Wadsworth, 1994. 441 p.
- 55 PERLIN, K.; VELHO, L. **B-spline wavelet paint**. New York, Computer Science Dept., New York University; Rio de Janeiro, Instituto de Matemática Pura e Aplicada (IMPA), 1994. (Technical Report) 23p.
- 56 PIERCE, J.R. **The science of musical sound**. New York, Scientific American Books, 1983. 242 p.
- 57 PISTON, W. **Orchestration**. Nova York, W.W. Norton, 1955. 477 p.
- 58 REISSEL, L-M. Multiresolution and Wavelets. In: SIGGRAPH 94. INTERNATIONAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 21., Orlando, 1994. **Wavelets and their applications to computer graphics**. New York, ACM, 1994. (Course notes, 11). p.31-61.

- 59 RIOUL, O.; DUHAMEL, P. Fast algorithms for discrete and continuous wavelet transforms. **IEEE Transactions on Information Theory**, v.38, n.2, p.569-86, Mar. 1992. Special issue.
- 60 RIOUL, O.; VETTERLI, M. Wavelets and signal processing. **IEEE Signal Processing Magazine**, v.8, n.4, p.14-35, Oct. 1991.
- 61 RISSET, J.C. The computer, music and sound models. In: INTERNATIONAL CONFERENCE ON WAVELETS: TIME FREQUENCY METHODS AND PHASE SPACE, Marseille, 1987. **Proceedings**. 2. ed. Berlin, Springer, 1989. p.102-23.
- 62 ROADS, C. Introduction to granular synthesis. **Computer Music Journal**, v.12, n.2, p.11-3, Summer 1988.
- 63 ROADS, C. et al. **The computer music tutorial**. Cambridge, MIT Press, 1996. 1234 p.
- 64 ROADS, C., ed. **The music machine: selected readings from *Computer Music Journal***. Cambridge, MIT Press, 1989. 725 p.
- 65 SAHINER, B.; YAGLE, A.E. Image reconstruction from projections under wavelets constraints. **IEEE Transactions on Signal Processing**, v.41, n.12, p.3579-83, Dec. 1993. Special edition.
- 66 SCHOLL, J.F.; ROGOVIN, D. Audio signal compression with Circular Wavelet Packets. **Proceedings of SPIE**, v.2303, p.518-29, 1994. /Apresentado a Wavelet Applications in Signal and Image Processing II, San Diego, 1994/
- 67 SHELBY, G.A.; ADHAMI, R.R. Tone detection using Wavelet transforms. **Proceedings of SPIE**, v.2491, pt.1, p.615-26, 1995. /Apresentado a Wavelet Applications II, Orlando, 1995/
- 68 SIGGRAPH 94. INTERNATIONAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 21., Orlando, 1994. **Wavelets and their applications to computer graphics**. New York, ACM, 1994. (Course notes, 11) 162p.
- 69 SIGGRAPH 95. INTERNATIONAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 22., Los Angeles, 1995. **Wavelets and their applications to computer graphics: CD-ROM**. New York, ACM, 1995. (Course notes, 26)
- 70 SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO E MÚSICA, 1º / CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 14º, Caxambu, 1994. **Anais**. Belo Horizonte, UFMG, 1994. 251 p.
- 71 SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO E MÚSICA, 2º / CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 15º, Canela, 1995. **Anais**. Porto Alegre, IIUFRGS, 1995. 303p.
- 72 SINHA, D.; TEWFIK, A.H. Low bit rate transparent audio compression using adapted wavelets. **IEEE Transactions on Signal Processing**, v.41, n.12, p.3463-79, Dec. 1993. Special issue.
- 73 SINZIG, PEDRO, Frei. **Dicionário musical**. 2.ed. Rio de Janeiro, Kosmos, 1959. p.261-3, p.601-2.
- 74 SOLBACH, L.; WÖHRMANN, R.; KLIEWER, J. The complex-valued continuous wavelet transform as a preprocessor for auditory scene analysis. **Internet**, "ftp://ftp.ti6.tu-harburg.de/pub/paper/ijcai95-casa-rev1.ps.gz", Mar. 1997. 7p. /Apresentado no IJCAI 95/
- 75 STOLLNITZ, E.J.; DeROSE, T.D.; SALESIN, D.H. Wavelets for computer graphics: a primer, Part 1. **IEEE Computer Graphics and Applications**, v.15, n.3, p.76-84, May 1995.
- 76 STOLLNITZ, E.J.; DeROSE, T.D.; SALESIN, D.H. Wavelets for computer graphics: a primer, Part 2. **IEEE Computer Graphics and Applications**, v.15, n.4, p.75-85, July 1995.
- 77 STRANG, G. Wavelet transforms versus Fourier transforms. **Bulletin (New Series) of the American Mathematical Society**, v.28, n.2, p.288-305, Apr. 1993.
- 78 STRANG, G.; NGUYEN, T. **Wavelets and filter banks**. Wellesley, Wellesley-Cambridge Press, 1996. 490p.

- 79 SWELDENS, W. The lifting scheme: a new philosophy in biorthogonal wavelet constructions. **Proceedings of SPIE**, v.2569, p.68-79, 1995. /Apresentado ao Wavelet Applications in Signal and Image Processing III, 1995/ /Obtido via **Internet**, "<http://www.cs.sc.edu/~fernande/liftpack/liftpack.html>", Dec. 1996/
- 80 TASWELL, C.; MCGILL, K.C. Wavelet transform algorithms for finite-duration discrete-time signals. In: INTERNATIONAL CONFERENCE ON WAVELETS AND APPLICATIONS, Toulouse, 1992. **Proceedings**. Gif-Sur-Yvette, Editions Frontieres, 1992. p.221-4. /Obtido via **Internet**, "<ftp://sccm.stanford.edu/>", Oct. 1993. 21p./
- 81 TEOLIS, A. BARAS, J.S. The wavelet processing workstation: an interactive MATLAB based computational tool for wavelet processing. **Proceedings of SPIE**, v.2491, pt.1, p.592-603, 1995. /Apresentado a Wavelet Applications II, Orlando, 1995/
- 82 UNSER, M. On the optimality of ideal filters for pyramid and wavelet signal approximation. **IEEE Transactions on Signal Processing**, v.41, n.12, p.3591-96, Dec. 1993.
- 83 VAIDYANATHAN, P.P. Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial. **Proceedings of the IEEE**, v.78, n.1, p.56-93, Jan. 1990.
- 84 VETTERLI, M.; HERLEY, C. Wavelets and filter banks: theory and design. **IEEE Transactions on Signal Processing**, v.40, n.9, p.2207-32, Sept. 1992.
- 85 VIDAKOVIC, B.; MÜLLER, P. Wavelets for kids: a tutorial introduction. Durham, Duke University, Institute of Statistics and Decision Sciences, 1991. **Internet**, "<ftp://lorax.isds.duke.edu/pub/Users/brani/papers/wave4kidsA.ps.Z>", Nov. 1996. 26p.
- 86 VILLASENOR, J.D. et al. Wavelet filter evaluation for image compression. **IEEE Transactions on Image Processing**, v.4, n.8, p.1053-60, Aug. 1995.
- 87 WEISS, L.G. Wavelets and wideband correlation processing. **IEEE Signal Processing Magazine**, v.11, n.1, p.13-32, Jan. 1994.
- 88 WICKERHAUSER, M.V. **Adapted wavelet analysis from theory to software**. Wellesley, A.K.Peters, 1994. 486p.
- 89 WOOD, P.; DUYNE, S.V., eds. **Center for Computer Research in Music and Acoustics: overview**. Stanford, Music Dept. of Stanford University, CCRMA, 1994. 72p.
- 90 XENAKIS, I. **Formalized Music: thought and mathematics in composition**. Bloomington, Indiana University Press, 1972. 273 p.
- 91 YANG, X.; WANG, K; et al. Auditory representation of acoustic signals. **IEEE Transactions on Information Theory**, v.38, n.2, p.824-839, Mar. 1992. Special edition.
- 92 YIP, WING-KEI; et al. Pitch detection of speech signals in noisy environment by wavelet. **Proceedings of SPIE**, v.2491, pt.1, p.604-14, 1995. /Apresentado a Wavelet Applications II, Orlando, 1995/
- 93 YOST, W.A. **Fundamentals of hearing: an introduction**. San Diego, Academic Press, 1994. 317p.

GLOSSÁRIO

ADSR (*Attack, Decay, Sustain and Release*) - Modelo elementar de envelope para tons instrumentais, consistindo em 4 estágios ou regiões sucessivas: um ataque, um decaimento, uma região de sustentação e um colapso (queda final).

AIFF - (*Audio Interchange Format File*). Formato digital de áudio, proprietário da Apple.

ASIC (*Application Specific Integrated Circuit*) - Circuito integrado de aplicação específica

AWMR - Análise *Wavelet* em Multiresolução, ou, alternativamente, análise em multiresolução com *wavelets*.

DAT - *Digital Audio Tape*. Equipamento para gravação/reprodução de áudio em fita magnética, em formato digital.

Dinâmica - Em Música, refere-se aos diferentes níveis de intensidade sonora durante a interpretação de um trecho musical, com variações que vão do *fortíssimo* ao *pianíssimo*, quer em progressão brusca ou lenta.

Dithering - Técnica empregada em modeladores de ruído em áudio digital que consiste na adição ao sinal de entrada (áudio) de um sinal randômico adequado, não correlacionado ao sinal de entrada, de forma a eliminar a distorção associada ao erro de quantização.

DSP (*Digital Signal Processing/Processor*) - Processador/Processamento de sinais digitais

Escala musical oficial - A escala (temperada) padrão atual foi estabelecida em uma conferência internacional em Londres, em 1953, quando então definiu-se a altura padrão para a nota Lá3 (o Lá central no piano) em 440 Hz.

Na composição da escala temperada, a relação entre, por exemplo, as frequências das notas Dó e Sol (uma quinta) é igual a relação entre as frequências da quinta formada por Ré # e Lá #, assim como a relação entre as frequências de Dó e Fá (uma quarta) é igual à relação entre as frequências de Fá # e Si. Conseqüentemente, o intervalo entre duas notas consecutivas quaisquer da escala de doze - *um semitom* - guarda sempre uma mesma relação matemática: $N_2=N_1.r$, $N_3=N_1.r^2$, ... , $N_{12}=N_1.r^{11}$. Como a nota N12 é a oitava de N1, i.é, $N_{12}=2 \times N_1$, extrai-se o valor de $r=1,05946$.

A escala temperada trouxe diversos benefícios, como a possibilidade de transposição perfeita para qualquer tom. Todavia, suas notas não permitem a criação dos intervalos acusticamente perfeitos, como as quinta perfeitas regidas pela relação "harmoniosa e divina" de 3:2, tão cultuada nas sociedades asiáticas mais antigas. Assim, um intervalo temperado de quinta (dó-sol, por exemplo) já não mais apresenta a relação 3:2; esta relação na escala temperada será aproximadamente verificada entre, por exemplo, dó e fá#, que não é uma quinta musical. A tabela abaixo mostra o padrão de referência estabelecido na convenção em Londres:

Tabela 3 - Notas da Escala Temperada e frequências de referência *

Notas	Frequência de referência (Hz)
Sol2 (G2)	196,0000
Lá2 (A2)	220,0000
Si2 (B2)	246,9000
Dó3 (C3) (Dó central do piano)	261,6256
Dó#3 (C#3)	277,1826
Ré3 (D3)	293,6648
Ré#3 (D#3)	311,1270
Mi3 (E3)	329,6276
Fá3 (F3)	349,2282
Fá#3 (F#3)	369,9944
Sol3 (G3)	391,9954
Sol#3 (G#3)	415,3047
Lá3 (A3)	440,0000
Lá#3 (A#3)	466,1638
Si3 (B3)	493,8833
Dó4 (C4)	523,2511
Ré4 (D4)	587,3000
Mi4 (E4)	660,0000
Fá4 (F4)	698,4000
Sol4 (G4)	783,9000
Lá4 (A4)	880,0000
Si4 (B4)	987,8000
Dó5 (C5)	1046,5000
Ré5 (D5)	1174,7000
Mi5 (E5)	1318,5000
Fá5 (F5)	1396,9000
Sol5 (G5)	1567,9000
Lá5 (A5)	1760,0000
Si5 (B5)	1975,5000
Dó6 (C6)	2093,0000
Ré6 (D6)	2349,3000
Mi6 (E6)	2637,0000
Fá6 (F6)	2793,0000
Sol6 (G6)	3136,0000

* Mostradas 4 oitavas no registro alcançável do violino (de G2 a G6)
De C3 a C4 mostra-se a escala cromática completa

Fontes: Internet, "<http://www.rionet.com.br/~music-center/escalas.htm>", Abril 1997; e
ITOKAWA, H. e CHIHIRO, K. *On the study of violin and its making* (1952). p.61^[30]

Em língua latina as notas recebem os nomes de dó, ré, mi, fá, sol, lá, e si. Na notação anglo-saxônica empregam-se letras para denominá-las: A (lá), B (si), C (dó), D (ré), E (mi), F (fá), G (sol). Uma nota lá3 (A3) é um lá na 3ª oitava da escala musical oficial.

Embora em termos musicais a oitava do Lá padrão seja a 3ª é comum a referência a ela como pertencente à 4ª oitava. Muitos equipamentos musicais eletrônicos seguem esta referência.

f_a - Frequência de amostragem. Neste trabalho duas foram usadas: 48000 Hz (amostras/s), no equipamento de gravação em estúdio (DAT), e 44100 Hz (amostras/s) nas plataformas computacionais.

FFT (*Fast Fourier Transform*) - Transformada rápida de Fourier

Jitter - Distorção não linear devida a problemas de sincronismo ou problemas no mecanismo ou meio físico (circuito elétrico) de transporte serial de áudio digital.

$L^2(\mathbf{R})$ - É o espaço de todas as funções quadraticamente integráveis em \mathbf{R} (conjunto dos números reais). Na prática, é o espaço das funções com energia finita.

Luthier - Fabricante de instrumentos de corda (vocabulo francês).

Microtom - sonoridade que apresenta uma altura (*pitch*) intermediária entre semitons convencionais da escala musical, e portanto dista (de um semitom ou de outro microtom) de um intervalo (de frequência) menor que o intervalo de um semitom convencional.

MIDI (*Musical Instrument Digital Interface*) - Protocolo para comunicação digital entre instrumentos musicais eletrônicos e equipamentos computacionais.

Música Computacional - Campo de pesquisa e desenvolvimento de tecnologia musical. Referência a um material musical cuja produção utiliza-se de recursos computacionais (*hardware e/ou software*) em uma ou mais de suas diversas fases (composição, transcrição, notação, harmonização, orquestração, gravação, processamento, síntese e reprodução). Alternativamente, o termo **computação musical** pode ser utilizado, mas este orienta a um contexto mais técnico e de computação, fazendo mais referência ao ferramental (técnicas, algoritmos, dispositivos e recursos computacionais) do que ao produto musical.

Música eletrônica - Material sonoro que tenha sido concebido e sintetizado através dispositivos e/ou sistemas eletrônicos operando com frequências na faixa audível. Com o passar dos tempos a produção de música por vias eletrônicas deu origem a sons e formas de organização especiais que levaram ao nascimento de novos estilos musicais, onde não só os aspectos musicais importam, mas também as técnicas de composição, síntese e as formas musicais modeláveis com esta tecnologia. A terminologia **Música Eletroacústica** faz referência a estilos musicais contemporâneos que empregam técnicas e equipamentos eletrônicos (e computacionais) para conceber e sintetizar música. Compositores e músicos adeptos da tecnologia musical, usualmente de formação musical erudita, constituem grupo relevante na produção da música eletroacústica, em suas diversas formas e objetivos.

Pitch - "*Pitch* é aquele atributo das sensações auditivas em termos do qual sons podem ser ordenados numa escala estendendo-se de baixo a alto. Do ponto de vista da psicoacústica, a unidade do *pitch* é o *mel*".^[93] (Leia mais sobre o conceito de *pitch* nos apêndices).

PLL - *Phase-Locked Loop* - Dispositivo eletrônico para capturar a fase de um sinal de referência (*clock*) e possibilitar o sincronismo. Estima a variação de frequência de um sinal de entrada de forma a manter o sincronismo.

QMF (*Quadrature Mirror Filter/Filtering*) - Filtros/Filtragem de espelhamento de quadratura.

R - Conjunto dos números reais.

Relação S/R - Relação sinal/ruído.

TEF - Transformada enjanelada (ou emoldurada) de Fourier. Em síntese, uma transformação de Fourier sobre uma fatia do sinal, obtida pela sobreposição de uma moldura (uma segunda função) sobre uma região de interesse do sinal.

Teoria Wavelet - Campo de pesquisa originário da Matemática - e estreitamente ligado a conceitos da análise funcional - que lida com a representação de funções genéricas em termos de "*blocos construtivos básicos*", ou "*átomos*", fixos em diferentes escalas e posições. O termo

atualmente faz referência a um contexto maior, abrangendo também áreas correlatas e aplicações das *wavelets* nas engenharias, na física, e em biologia. A teoria *Wavelet* representa hoje um corpo sintético de diversas metodologias que endereçam o tratamento e análise de sinais, unificando-as sob uma mesma estrutura de referência. A **Transformada *Wavelet*** é uma ferramenta de implementação matemática baseada na e nascida desta teoria.

TW - Transformada *Wavelet* (termo genérico)

TWC - Transformada *Wavelet* Contínua.

TWCI - Transformada *Wavelet* Contínua Inversa

TWD - Transformada *Wavelet* Discreta.

TWDI - Transformada *Wavelet* Discreta Inversa

VCA - *Voltage Controlled Amplifier*: amplificador controlado por tensão. Módulo modelador de envelope em sistemas de síntese eletrônicos.

VCO - *Voltage Controlled Oscillator*: oscilador controlado por tensão. Módulo gerador de ondas em sintetizadores eletrônicos cuja frequência é controlada por tensão.

VLSI (*Very Large Scale Integration*) - dispositivos com escala de integração muito alta.

Z - Conjunto dos números inteiros.

APÊNDICES

I. DIFERENÇAS ENTRE TOM, PITCH E NOTA MUSICAL

Nota musical se refere a uma posição dentro da escala cromática (ou qualquer outra escala). A nota refere-se mais à posição numa escala do que propriamente ao "conteúdo sônico" emitido quando se a toca. Também refere-se a uma tecla, posição ou arranjo de dedilhado num instrumento musical. Por exemplo, quando dizemos que tocamos um dó C4 num piano, estamos tocando a nota C4, dó central localizado no meio do teclado do piano, que emite um conteúdo sônico que é identificado graficamente na escala por uma posição específica na partitura (representação gráfica).

Tom refere-se ao conteúdo sônico e características tempo-frequenciais da nota musical emitida. Por exemplo pode-se tocar um dó C4 num piano e a mesma nota numa flauta, no entanto os timbres serão diferentes: o tom produzido pela mesma nota em instrumentos diferentes são também tons diferentes, mesmo que sejam percebidos como compartilhando uma mesma altura tonal. Tons podem ser confundidos por notas musicais em determinados contextos. Basicamente o tom de uma nota musical possui uma estrutura harmônica (composição espectral) específica, apresenta uma evolução no tempo (intensidade e espectro variantes), e um *pitch*.

Tom também é definido como um som em uma única frequência, apesar desta definição estar mais para a de *pitch* (a seguir) do que para o conceito de tom. Um tom geralmente possui várias frequências misturadas ou fundidas num único som percebido (timbre).

Pitch é uma propriedade do som que permite identificar ou determinar sua altura em frequência, seja ele um tom ou um "barulho". Esta é uma definição muito direta, que se encaixa melhor dentro de uma abordagem puramente física, quando deseja-se extrair um número exato que represente a frequência do som em questão. Na música e na psicoacústica, no entanto, esta definição é ampliada: o *pitch* ainda consiste numa identificação reducionista, onde se busca classificar o todo por um componente básico essencial que o distinga dos demais, mas ao invés de se referir a uma frequência única, refere-se agora a um padrão frequencial que é percebido como o *pitch* de um tom, ou em outras palavras, sua altura tonal.

Do lado da psicoacústica, o *pitch* é interpretado como propriedade percebida do tom, e mensurável por meio de uma unidade denominada "mel". Do lado físico, o *pitch* sempre leva a uma medida de altura, a frequência do som, mesmo que identificada por meio do padrão de harmônicos presentes e não por uma única componente frequencial.

Existe uma certa confusão no emprego dos termos "*pitch*" e "tom", particularmente quando "tom" é utilizado referindo-se à altura tonal, na verdade uma propriedade sua muito importante para proceder à afinação de instrumentos e à identificação de sua frequência.

Se o som for um tom musical, seja de um instrumento "comportado" como flautas e violinos, ou de instrumentos inarmônicos (como sinos), o *pitch* está atado à periodicidade das ondas sonoras, usualmente indicando a frequência da sua forma de onda básica. Se o som entretanto não for um tom musical (como por exemplo ruídos rosas, alguns sons percussivos, sons de impacto, etc.) a identificação de seu *pitch* é mais complexa, e dependerá do grau de treinamento do ouvido que o analisa.

A determinação do *pitch* envolve o reconhecimento de padrões de vibrações ao longo da membrana basilar no ouvido interno. Para as notas baixas num piano, verificou-se que o *pitch* não pode ser determinado a partir da fundamental. Para notas mais altas (em frequência) a fundamental ou o primeiro harmônico (oitava da fundamental) predominam no julgamento do *pitch* do tom emitido. Na faixa média (*mid-range*) a fundamental é importante quando está presente. E para *pitchs* mais baixos, a fundamental é de pouca importância no seu julgamento, e identifica-se o *pitch* a partir do padrão de harmônicos. ^[56]; Pierce, J.R. *Pitch and Repetition Rate Perception*, in ^[89], p.46).

II. CENTROS DE PESQUISA EM COMPUTAÇÃO E MÚSICA

O *Center for Computer Research in Music e Acoustics (CCRMA)*, localizado na Universidade de Stanford (EUA), reúne uma comunidade de pesquisadores bastante heterogênea, com efetiva representação dos departamentos de Música, Engenharia Elétrica, Mecânica, Ciência da Computação e Psicologia^[89]. O CCRMA atua no magistério, como parte integrante do Depto. de Música, possui atividades nas área de composição, e pesquisa em desenvolvimento de *software* para computação musical, em processamento de sinais digitais e modelamento físico de instrumentos, em desenvolvimento de controladores e instrumentos musicais, em psicoacústica e psicologia da cognição, além de uma extensa lista de áreas de interesse, detalhadas em^[89]. As atividades de pesquisa desenvolvidas no CCRMA desde a sua fundação respondem por contribuições expressivas no campo da música computacional, com uma grande produção científica. Além disso, convênios e projetos conjuntos com corporações privadas permitem a integração direta da pesquisa com a indústria. O centro desenvolveu junto à Yamaha o teclado *Yamaha VLI (Virtual Acoustic Synthesizer)*, um sintetizador em tempo real de instrumentos de sopro metálicos, baseado em técnicas de modelamento físico, e junto à Korg desenvolveu o *Korg Wave Drum*, um instrumento de percussão eletrônico com síntese baseada parte em tabelas de amostras, parte em modelamento físico.

O *Center for New Music e Audio Technologies (CNMAT)*, da Universidade da Califórnia, Berkeley, tem como meta congregar recursos, ferramentas técnicas e científicas e conduzi-las a problemas e questões musicais, com particular ênfase no desenvolvimento de tecnologia para execução musical (interpretação), tópicos relacionados à execução, e problemas como controle e síntese em tempo real. As atividades de pesquisa objetivam a concepção de sistemas controladores e interfaces gestuais mais flexíveis que os controladores MIDI. Outras atividades incluem a difusão de novas tecnologias de som, novos controladores, e apresentação de concertos e trabalhos de composição eletroacústica^[24].

O *Institut de Recherche et Coordination Acoustique/Musique (IRCAM)*, centro de pesquisa em tecnologia musical ligado ao *Centre National d'Art et de Culture Georges-Pompidou*, localizado em Paris, França, foi fundado em 1970 pelo governo francês, e conduzido desde o início por Pierre Boulez.

Em 1974 Boulez declara publicamente os objetivos do instituto, quais sejam o de promover a pesquisa, o desenvolvimento, a experimentação, a criação e o treinamento nas áreas de música e tecnologia. As atividades do instituto centram-se na pesquisa de fenômenos acústicos (acústica instrumental, espacial, psicoacústica), na pesquisa em análise e síntese de sons e estruturas sonoras, em representação musical, no desenvolvimento de ferramentas computacionais e instrumentos para pesquisa e criação musical, bem como em atividades de produção musical e educação.

Dentro da estrutura produtiva do instituto há um enfoque especial orientado à figura do compositor, buscando criar condições para o seu aprimoramento e assistindo-os no curso de suas atividades de criação, coordenando-se de forma a otimizar o processo de produção musical.

O instituto é único no mundo em termos de instalações, contando com uma câmara anecóica para medições acústicas, vários estúdios e laboratórios equipados com computadores UNIX (DEC) e uma rede interligando estações Sun, DEC e NeXT, um espaço de projeções (para concertos, testes e gravações) cujas dimensões volumétricas e características acústicas podem ser modificadas através de um sistema computadorizado, laboratórios de eletrônica e mecânica adequados para produzir modelos e protótipos especiais, entre outras facilidades. Desde o início de suas atividades, vem contribuindo no desenvolvimento de técnicas para análise e a síntese de sons musicais, de programas e métodos para modelar e sintetizar o canto, de sistemas para composição, produção e processamento de som digital, bem como contribuído nas áreas de modelamento e projeto de espaços acústicos^[32].

Além destes, importantes centros de pesquisa, laboratórios e grupos de trabalhos estabelecidos em várias universidades e instituições governamentais e privadas realizam pesquisas no campo da engenharia de áudio, computação musical, psicoacústica, processamento de sinais de áudio (voz e música), bem como na musicologia, na composição e na criação de obras musicais antes inconcebíveis sem as facilidades tecnológicas. Entre muitos, destacamos uma pequena amostragem, porém de participação significativa e relevância na produção científica, pesquisa e desenvolvimento em computação musical, música eletrônica e/ou eletroacústica e tecnologia musical:

- IEEE Computer Society Technical Committee on Computer Generated Music (http://www.computer.org/tab/cgm/tc_cgm.htm)
- Audio Engineering Society - AES (<http://www.aes.org/>)
- International Computer Music Association - ICMA (<http://music.dartmouth.edu/~icma/>)
- NAMM - National Association of Music Merchants (<http://www.namm.com>)
- Hochschule für Musik und Darstellende Kunst (Instituto de Música Eletroacústica e Experimental), Academia de Música de Viena, Áustria
- Studio for Advanced Music Technology (SAMT), Hagenberg/Linz, Áustria
- Music Department, The City University, Londres, Inglaterra
- Korg Research Center, Inglaterra
- AI/Music Group, Fac. of Music and Dept. of Artificial Intelligence, Univ. of Edinburgh, Escócia
- Signal Processing Research Group, Dept. of Electronics, Univ. of York, York, Reino Unido
- Dept. of Music, Keele University, Staffordshire, Inglaterra
- School of Cognitive and Computing Sciences, University of Sussex, Inglaterra
- Laboratorio de Musica & Sonologia, Dipartimento di Matematica Pura ed Applicata, Univ. di L'Aquila, Itália
- AIMI - Associazione di Informatica Musicale Italiana.
- Laboratorio di Informatica Musicale, Dipartimento di Scienze dell'Informazione, Università degli Studi de Milano (LIM-DSI)
- Centro di Sonologia Computazionale, Istituto di Elettrotecnica ed Elettronica, Univ. di Padova, Itália
- Laboratorio de Investigación y Producción Musical (LIPM), Argentina
- Instituto Rosario de Investigaciones en Ciencias de la Educación (IRICE), Argentina
- Laboratoire Bordelais de Recherche en Informatique, Univ. Bordeaux I, França
- LAFORIA, França
- Laboratoire de Mécanique et d'Acoustique, C.N.R.S., França
- Association pour la création et la recherche sur les outils d'expression (ACROE) e Laboratoire d'Informatique Fondamentale et d'Intelligence Artificielle (LIFIA), França
- Institute for Psychoacoustics and Electronic Music, Univ. of Ghent, Bélgica
- Dept. of Music, University of Copenhagen, Dinamarca
- NoTAM - Norwegian network for Technology, Acoustics and Music, Noruega (<http://www.notam.uio.no>)
- Sweelinck Conservatory Amsterdam, Holanda
- Institute of Sonology, Utrecht, Holanda
- Studio for Electronic Music, Musik-Akademie, Basel, Suíça
- Dept. of Speech Communication and Music Acoustics, Royal Institute of Technology (KTH), Estocolmo, Suécia
- Institute for Computer Music and Electronic Media (ICEM), Folkwang Hochschule-Essen, Alemanha
- Lehrstuhl fuer Nachrichtentechnik, Univ. Erlangen-Nuernberg, Alemanha
- Banff Centre for the Arts, Alberta, Canadá
- School of Kinesiology, Simon Fraser University, Canadá
- Centre d'Applications Musicales de l'Informatique (CAMI), Faculté de Musique, Univ. de Montréal, Québec, Canadá
- Structured Sound Synthesis Project (SSSP), Computer Systems Research Group, Univ. of Toronto, Canadá
- Dept. of Computing and Information Science, Queen's University, Ontario, Canadá
- Dept. of Music, Queen's University, Ontario, Canadá

- Department of Communication and Centre for the Arts, Simon Fraser University, Canadá
- Departamento de Ciências da Computação, Universidade de Hong Kong, China
- Media Laboratory, Music Cognition Group, Cambridge, EUA
- Center for Research in Computing and the Arts (CRCA), Univ. of California, San Diego, EUA
- Dept. of Music, State Univ. of New York at Buffalo (NY), EUA
- Dept. of Music, Colgate University, Hamilton (NY), EUA
- Dept. of Music, University of Washington, Seattle, EUA
- Music Department, Princeton University, New Jersey, EUA
- Electrical Engineering and Computer Science Department, Princeton University, New Jersey, EUA
- Computer Audio Research Laboratory, Center for Music Experiment, Univ. of California, San Diego, EUA
- Experimental Music Studio, M.I.T., Cambridge, EUA
- M.I.T. Media Lab, Cambridge, EUA
- Artificial Intelligence Laboratory, M.I.T., Cambridge, EUA
- Dept. Electrical Engineering, Cornell University, Ithaca (NY), EUA
- Lucasfilm Ltd., San Rafael, Califórnia, EUA
- Computer Science Department, Univ. of Rochester, Rochester (NY), EUA
- Dept. of Music, La Trobe University, Victoria, Austrália
- Ingeniería de Sistemas, Univ. Javeriana de Cali, Colômbia
- Department of Microelectronics, Univ. Autónoma de Puebla, México
- Laboratorio de Informática Musical (LIM), Escuela de Música de la Univ. de Guanajuato, México
- Department of Electronics and Telecommunication, Applied Physics Division, Centro de Investigación Científica y de Education Superior de Ensenada, México
- Departamento de Artes, Univ. Federal da Paraíba
- Departamento de Música/CCHLA, Univ. Federal da Paraíba
- Departamento de Música, Univ. Federal de Pernambuco
- Núcleo Interdisciplinar de Computação Sônica (NICS), Depto. de Matemática Aplicada, UNICAMP
- Depto. de Semicondutores, Instrumentos e Fotônica (DSIF), Fac. Eng. Elétrica, UNICAMP
- Instituto de Informática, Univ. Federal do Rio Grande do Sul
- Laboratório de Processamento Espectral, Depto. Ciência da Computação, UNB, Brasília
- Departamento de Música, UNB, Brasília
- Oficina de Análise e Síntese da Imagem e do Som, Depto. Ciência da Computação, UFMG, Belo Horizonte
- Escola de Música da Univ. Federal de Minas Gerais, Belo Horizonte
- Departamento de Formação Artística, Univ. Federal do Espírito Santo
- COPPE/Sistemas, Univ. Federal do Rio de Janeiro
- Escola de Música, Centro de Letras e Artes, Univ. Federal do Rio de Janeiro
- Depto. Ciência da Computação, Inst. de Matemática e Estatística, USP, São Paulo
- Studio PANaroma de Música Eletroacústica da UNESP/FASM, São Paulo
- Laboratório de Linguagens Sonoras / Comunicação e Semiótica, PUC, São Paulo
- Grupo de Computação Musical, Laboratório de Sistemas Integráveis (LSI), USP, São Paulo

A associação interdisciplinar entre engenheiros, matemáticos, físicos, psicólogos, e músicos só se tornou possível graças à transposição das barreiras preconceituosas e segregacionistas, que mantinham esses blocos convenientemente limitados em suas respectivas área de atuação. Esta abordagem holística, que influi nas relações entre as ciências no mundo contemporâneo, floresce em decorrência do inter cruzamento de necessidades, fruto de uma tendência integradora que coloca o artista como co-autor da ferramenta, e o cientista como co-autor da criação artística.

III. LINGUAGENS COMPUTACIONAIS PARA MÚSICA

A descrição formal de sons através de uma linguagem compatível com o computador abriu o caminho para a síntese via *software*, como nas palavras de Risset ^[61]: “sistemas inteligentes que reagem a ordens expressas de forma declarativa, mais fáceis de se formular do que ordens que devam prescrever cada passo que o computador deva fazer”, numa referência à facilidade de se descrever os passos algorítmicos da síntese numa linguagem em nível mais alto, que parametrize aspectos musicais.

Em 1957 Max Mathews desenvolveu o primeiro programa para síntese de música: o MUSIC I. Ele inaugurava a era da síntese digital direta. O MUSIC I foi o primeiro de uma série. Mathews desenvolveu os primeiros *compiladores modulares* para música, incorporados nas versões MUSIC III (1959), MUSIC IV e V (1967). Nestes, a síntese era determinada pela especificação da estrutura física do som desejado (instrumentos) e pela especificação das notas e respectivos argumentos (duração, intensidade, etc.). Isto permitiu de fato a definição de uma linguagem operacional formal para descrição de sons.

Os programas MUSIC introduziram conceitos fundamentais para o desenvolvimento dos sistemas eletrônicos de síntese, como o de *unidade geradora* (no MUSIC III) e a concepção de interfaces gráficas (no MUSIC V), paradigmas amplamente utilizados até hoje no projeto de sintetizadores e programas para síntese e sequenciamento de eventos musicais, como os populares sequenciadores MIDI e os teclados eletrônicos comerciais, produzidos por diversas indústrias de produtos musicais eletrônicos.

O processo da criação musical auxiliada por computador adicionou ao vocabulário do músico contemporâneo termos como *edição, abrir, salvar, cortar, copiar, colar, inserir, formatar*, e outros intensamente disseminados entre os usuários de computadores, bem como um novo conjunto de termos operacionais apropriados à manipulação de objetos musicais, como *gravar, tocar, quantizar, transpor, retroceder, deslocar, inverter, amplificar, filtrar, distorcer, normalizar, reverberar, gerar onda, comprimir, converter formato, analisar espectro, setar base de tempo, volume, pista, canal, envelope, eco, taxa de amostragem...*

Um grande número de linguagens musicais, compiladores, programas e interfaces foram desenvolvidos em vários centros de pesquisa em computação musical e tecnologia de áudio estabelecidos pelo mundo. A tabela a seguir seleciona uma amostra (pequena mas significativa) de importantes contribuições à tecnologia musical, que serviram e servem de base para o desenvolvimento de uma infinidade de produtos musicais eletrônicos, baseados em computadores, e de programas computacionais:

Music10	MUSIC V, de Max Mathews, portado para o PDP10
Music360	(1968)
Music11	(1973)
SCORE	para edição de partitura (L.Smith, 1972)
FORMES	linguagem orientada a objetos, para composição e escalonamento de processos musicais (IRCAM,1984)
CHANT	para síntese de canto (IRCAM, 1984)
MIDI, protocolo	para interfacear sintetizadores e dispositivos digitais (G.Loy, 1985)
CSOUND	para síntese genérica (Vercoe, MIT, 1986)
Common Music	para composição (R.Taube, 1989/90)
CMUSIC	para síntese genérica (R.Moore, 1990)
STELLA	para edição de partitura (Taube, 1993)
Finale	para edição de partitura (produto comercial da CODA)

Acrescente também à lista os diversos ambientes de programação (sistemas de *software*) escritos para o desenvolvimento de aplicações musicais em plataformas Macintosh, NeXT ou UNIX: MAX, LeLisp, Smalltalk, DMIX, Objective C, C, etc.

Referências adicionais são abordadas em ^[64] e ^[89]. Para uma visão do panorama atual no mercado de programas e pacotes computacionais para edição e geração de áudio/música consulte ^[22; 23].

IV. MÉTODOS PARA ANÁLISE, SÍNTESE E TRANSFORMAÇÃO MUSICAL

Os principais métodos utilizados para a análise, síntese e transformação de sons musicais podem ser agrupados nos tipos básicos abaixo, classificados segundo o paradigma generativo de som adotado, e em função dos métodos matemáticos que empregam:

Modelamento Físico

Contempla o modelamento acústico de instrumentos musicais reais. Os modelos físicos orientam-se no sentido de investigar os processos físicos que ocorrem na geração sonora e parametrizar o som na sua fonte física vibratória, por exemplo nas cordas/caixa ressonante de um violino, no trato vocal ou nos elementos vibrantes de qualquer outro instrumento acústico^[89]. Este paradigma tem sido bastante utilizado em muitos modelos de teclados musicais comerciais recentes (por exemplo os Trinity e Prophecy, da Korg^[22]) e em outros instrumentos eletrônicos¹.

Modelamento Espectral

Os modelos espectrais se orientam no sentido de compreender a percepção e a cognição do som no seu destino final -o sistema auditivo, parametrizando-o junto à membrana basilar na cóclea, e levando em consideração também aspectos da psicoacústica e a interpretação da informação acústica pelo cérebro. De uma maneira geral todas as técnicas de análise e síntese fundamentadas na descrição de conteúdo espectral, que operam modelando os sons no domínio da frequência, podem ser admitidas nesta categoria:

- análise e síntese (Risset, 1969)
- métodos das transformadas (Fourier, Karhunen-Loève, Walsh-Radamar, *Wavelet*, etc.)
- síntese aditiva e síntese subtrativa
- síntese granular (Roads, 1978, 1985; Jones & Park, 1988; Truax, 1988; e outros)
- *Phase-Vocoder* (Portnoff, 1976; Moorer, 1978; Dolson, 1986)
- LPC -*Linear Prediction Coding* (Moorer, 1977; Lansky, 1981)

Métodos não-lineares (ou globais)

Os métodos não lineares utilizam modelos abstratos, formulações matemáticas e técnicas generativas não-lineares para construir os sinais acústicos. Estão entre eles:

- Waveshaping (Arfib, 1979)
- Síntese FM (Chowning, 1973), amplamente utilizada em sintetizadores de placas multimídia para PC's, placas MIDI e teclados comerciais (como os da série DX-*n* da Yamaha)
- Síntese Multiplicativa^[64, p.439]
- Síntese por sinais aleatórios^[64, p.434]

Síntese baseada em tabelas de formas de onda, amostras de timbres naturais (MIDI/wavetable based synthesis)

Método baseado na geração do som a partir da reprodução de trechos amostrados (*samples*) de timbres instrumentais. A qualidade do timbre final está condicionada à qualidade da gravação do timbre original amostrado, à resolução da amostragem, e aos níveis de ruído no conversor digital-analógico (DAC) embutido no sintetizador. A qualidade alcançada com esse método é

¹ Como o Korg Wave Drum, com síntese de percussão baseada parte em modelamento físico, parte em tabela de formas de onda (*wavetable*), desenvolvido com a participação do CCRMA de Stanford. Outro exemplo de sintetizador baseado em modelamento físico, também com a participação do CCRMA, é o Yamaha VL1 (*Virtual Acoustic Synthesizer*), capaz de sintetizar com realismo 2 notas em tempo real de metais (instrumentos de sopro metálicos), conforme tive a oportunidade de conferir pessoalmente em visita ao MIDI Studio do CCRMA em agosto de 1995.

razoavelmente satisfatória, estando o julgamento de sua qualidade um passo mais próximo do realismo alcançado na síntese acústica em instrumentos reais. Entretanto, devido à simplificação e tamanhos reduzidos dos bancos amostrais, limitados a interpretações medianas, pouco expressivas, muitas vezes com notas interpretadas alheias a um contexto musical, fica a qualidade final restringida em graus variáveis de inexpressividade e artificialismo, deixando a incômoda impressão de ausência de naturalidade e continuidade nos temas musicais sintetizados.

Métodos mistos (utilizando vários tipos de síntese associadamente)

Observa-se uma rápida absorção do estado da arte na concepção de sistemas de geração sonora, encurtando substancialmente o tempo entre a concepção e o produto. As tendências atuais são a de se propor sistemas mistos, a exploração de técnicas de modelamento físico e de transformadas especiais. A transformada *wavelet*, graças às suas propriedades matemáticas na relação tempo-frequência, e peculiaridades conceituais e estruturais, exhibe vantagens sobre as técnicas clássicas de análise espectral e revela um potencial considerável para aplicações musicais.

As técnicas de modelamento físico, especialmente quando associadas com outros métodos de geração e processamento sônico, implicam numa carga computacional elevada. Em grande parte isso se deve aos processos de simulação acústica em que se baseia o modelamento físico. Sistemas de alto desempenho, capazes de processamento em tempo real, passam a ser uma necessidade real para a obtenção de sistemas de síntese de alta qualidade, de manipulação flexível, com vários graus de liberdade para o músico.

Nesta linha de desenvolvimento destaca-se a necessidade de sistemas que enderecem um melhor mapeamento da linguagem simbólica, da temática e contexto musical em eventos a nível de sinal digital, tanto para efeito de melhoria, desenvolvimento e enriquecimento dos sistemas para análise musical, análise de orquestração, identificação e isolamento de eventos/fraseados musicais elementares e complexos, quanto para efeito da concepção de sistemas de síntese mais eficientes computacionalmente.

A migração do padrão de amostragem para áudio de 44100 amostras/s e 16 bits de resolução para 96000 amostras/s e 24 bits de resolução trará ainda substancial ganho na qualidade sonora e acréscimo de realismo, entre outras importantes melhorias, e impelirá a indústria do áudio na direção de sistemas mais sofisticados.

V. WAVESYNT - PACOTE PARA ANÁLISE WAVELET EM MULTIRESOLUÇÃO DE SINAIS MUSICAIS

O **Wavesynt** consiste na contribuição em *software* deste trabalho. Ele reúne um conjunto mínimo de rotinas em linguagem MATLAB orientada à Análise *Wavelet* em Multiresolução (AWMR) de sinais unidimensionais (1D), especificamente sinais em faixas audíveis (0 a 20 KHz). O pacote foi desenvolvido especialmente para suportar as tarefas de análise de sinais musicais do presente projeto de pesquisa, e roda suportado pelo conjunto de funções e códigos fontes do pacote WaveLab, de Stanford. Vários de seus módulos entretanto são independentes deste último, e podem ser usados sem a prévia instalação do WaveLab.

A chamada a vários deles se dá diretamente a partir da manipulação dos botões e controles da janela principal, a **interface gráfica**, ou console de onde o usuário carrega os arquivos sonoros, define a *wavelet* de análise e dispara tarefas de decomposição *wavelet*, plotagem, gravação e audição de níveis e coeficientes *wavelet*, reconstrução e mixagem. Outros módulos são usualmente chamados diretamente via entrada de comando texto na janela de comandos do MATLAB, seguindo uma sintaxe conforme especificado nos comentários de ajuda (*help*) logo no início da listagem de cada código fonte.

Para utilizar plenamente o **Wavesynt** deve-se **(1)** ter instalado o MATLAB numa plataforma PC ou UNIX, **(2)** instalar o pacote WaveLab, **(3)** copiar todos os códigos fontes (módulos) do **Wavesynt** para um subdiretório *wavesynt* criado dentro do diretório do WaveLab (por exemplo `...matlab\toolbox\Wavelab\Wavesynt`), **(4)** editar o arquivo `...wavelab\wavepath.m` para que ele inclua o novo diretório *wavesynt* no caminho (*path*) do MATLAB², **(5)** disparar o MATLAB e invocar a **interface gráfica** do **Wavesynt** digitando-se "inrface" no *prompt* de comandos.

Os principais códigos fontes do Wavesynt estão relacionados na tabela a seguir, com uma breve descrição de suas respectivas funções. Os códigos fontes compõem ao todo 28 módulos, que conjuntamente integram o pacote **Wavesynt**. Todos os módulos encontram-se disponíveis armazenados nos meios digitais em anexo: uma cópia do **Wavesynt** em CD-ROM (PC-compatível) e em disquete (1,44Mbyte, PC-compatível), arquivos em formato texto/ASCII, com extensão **.m** (executáveis MATLAB).

A versão atual é a 3.0 (junho 1997). Versões futuras deverão incorporar maior automatismo no processo de instalação e padronizar definitivamente a sessão de ajuda presente no início de cada módulo (*help* do MATLAB) tal que exibam informações em uma só língua. Também prevê-se uma versão inglesa e outra em língua portuguesa, além de um pequeno tutorial introdutório para guiar o usuário no uso do pacote.

² Para se incluir o caminho do diretório *wavesynt*, sugere-se por exemplo a inclusão das seguintes linhas ao `wavepath.m` na sessão dentro deste arquivo onde se encontram linhas semelhantes:

```
pref = [MATLABPATHSEPARATOR WAVELABPATH];
pref = [pref 'wavesynt' post];
p = [p pref];
```

Tabela - Principais módulos do Wavesynt e respectivas funções

Módulos	Funções
AMP(var,a,b)	Calcula a amplitude média (absoluta) do vetor var entre os pontos a e b
CRUZAR(wx,wxx,n1,n2,n3,L,qmf,peso)	Reconstrutor para síntese cruzada
ENERGRAF('tipo','nome')	Produz um gráfico de barras de energia de todos os níveis da análise
ENERGY(var,a,b)	Calcula energia do vetor var entre os pontos a e b
GERAFIG3(primeiro,ultimo,'nome','com_energia')	Gera figura (formato GIF) com gráficos dos níveis
INTRFACE.M	Gera a interface gráfica principal (interface mestre)
MIXAR(wx,L,peso,qmf)	Mistura os níveis da análise segundo os pesos atuais na tabela de pesos
NORMALIZ(variável)	Faz um <i>zero-padding</i> sobre a variável a fim de normalizar sua extensão
PWFFT(x,fi,ff,'n')	Calcula e plota a FFT do vetor x
RECONST(wcoef, L, qmf, peso)	Reconstrui o sinal com os coeficientes dos níveis ponderados
SETWLET	Seta a família de <i>wavelets</i> sob uso
SOM(Y,fs,'pr')	Converte vetor Y em som, sonoriza-o, e grava em arquivo
SONORIZE(Y,fs)	Converte vetor Y em arquivo de som AIF e toca-o numa máquina UNIX remota (Silicon Graphics) (<i>para máquinas UNIX em rede</i>)
WNIVEL(wc,n,qmf)	Gera e plota um nível <i>wavelet</i> específico da AWMR