André Luiz Luvizotto

Modelos de Representação de Sinais Musicais via Transformada *Wavelets*

Faculdade de Engenharia Elétrica e de Computação - Unicamp

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

Orientador: Rafael Santos Mendes
Co-orientador: Jonatas Manzolli

Campinas, SP
2007

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

Luvizotto, André Luiz
F149u Modelos de representação de sinais musicais via transformada Wavelets
André Luiz Luvizotto. – Campinas, SP:
[s.n.], 2007.

Orientadores: Rafael Santos Mendes; Jônatas Manzolli.
Tese (Mestrado) - Universidade Estadual de Campinas,
Faculdade de Engenharia Elétrica e de Computação.
1. 1. Processamento de sinais. 2. Wavelets
(Matemática).
I. Mendes, Rafael Santos. II. Manzolli, Jônatas.
III. Universidade Estadual de Campinas.
Faculdade de Engenharia Elétrica e de Computação.
IV. Título.

Título em Inglês: Representation models of musical signals by means of wavelets transform
Palavras-chave em Inglês: Processing signal, Wavelets (Mathematics)
Área de concentração: Engenharia de Computação
Titulação: Mestre em Engenharia Elétrica
Banca Examinadora: Adolfo Maia Junior, Fernando José Von Zuben e Amauri Lopes
Data da defesa: 17/12/2007
Programa de Pós-Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE MESTRADO

Candidato: André Luiz Luvizotto

Data da Defesa: 17 de dezembro de 2007

Título da Tese: "Modelos de Representação de Sinais Musicais Via Transformada Wavelets"

Prof. Dr. Rafael Santos Mendes (Presidente): ____________________

Prof. Dr. Adolfo Maia Júnior: ____________________

Prof. Dr. Amauri Lopes: ____________________

Prof. Dr. Fernando José Von Zuben: ____________________

Resumo

Neste trabalho, as transformadas *wavelet* são utilizadas para recodificar e explorar sinais musicais. A idéia central do trabalho é mostrar que a reordenação adequada dos coeficientes provenientes das análises possibilita não somente recodificar de modo econômico o sinal como também transitar por sonoridades distintas. Ou seja, relaciona-se o posicionamento dos coeficientes *wavelets*, dentro de cada nível de multiresolução, com a geração de diferentes timbres. Uma vez definido o ordenamento dos coeficientes, duas formas de determinação de seus valores são exploradas. Na primeira forma, essas curvas ascendentes de coeficientes são aproximadas por uma função polinomial de grau p, levando a representações econômicas do sinal musical. Já no segundo modelo, os coeficientes de um sinal distinto (chamado sinal de base) são ordenados de forma ascendente. São então reordenados através das posições equivalentes do sinal original, também chamado de alvo, permitindo a sua reobtenção e também o trânsito entre as duas sonoridades envolvidas. Os resultados deste trabalho destacam a grande importância que o posicionamento dos coeficientes exercem na sonoridade com relação aos seus valores. Com os experimentos realizados, foi possível constatar que valores aproximados de coeficientes, corretamente dispostos no tempo, geram timbres alvos diversos de forma satisfatória. As análises dos resultados são feitas por metodologias diversas e pela audição dos arquivos de áudio gerados, que acompanham o texto em cd. Por fim, uma discussão sobre os resultados obtidos é realizada e uma proposta de continuação da pesquisa é sugerida, baseada em grupos de permutações como forma de síntese.

Palavras-chave: Representação de Sinais Musicais, *Wavelets*, Síntese de Som.

Abstract

Musical signals are encoded and represented using wavelet transforms. Starting upon wavelet analysis, the main research idea is to show how an adequate re-ordination of wavelet coefficients makes possible to decode the signal economically and also to obtain different sonorities. Moreover, we relate a specific set of wavelet coefficients from each multi-resolution level to generate new timbres. Given re-ordination procedures, two forms to determine their values are explored. First one, coefficient curves in ascendent order are approximated by a polynomial function of degree p that leads to an economic representation of musical signals. In the second approach, coefficients from another signal (named "base"signal) are ordinated in ascendent order also. After they are re-ordinate across the positions of the original signal, called "target"signal. The results of this research highlight the great importance of the wavelet coefficient's order to manipulate timbre. Experiments presented here showed that coefficients with approximated values and correctly disposed in time, can be used to generate target timbres in a satisfactory way. The analysis of the results were done by different methodologies and by listening to the sound examples from the CD attached to this dissertation. Finally, a discussion of research aims is presented and a proposal for further work based on Mathematical Group of Permutation is projected as a way to develop a new synthesis method.

Keywords: Musical Signal Representations, Wavelets, Sound Synthesis.

Aos meus pais

Agradecimentos

Aos orientadores Rafael Santos Mendes e Jônatas Manzolli pela dedicação e confiança.

Aos Profs. Amauri Lopes e Fernando Von Zuben por participar da banca.

Ao Prof. Adolfo Maia Jr. pelas aulas de matemática e por participar da banca.

Aos meus pais por todo amor e suporte, sempre e incondicional.

Ao meu irmão George e a Paulinha pelo carinho e suporte quando estive na Alemanha e pela força na finalização do trabalho.

Aos meus "pais"de Campinas tia Tila e tio Gian pelo imenso carinho e amor.

Ao Mauly pelo inglês e doses semanais de insanidade.

Ao Fábio 'coach' Furlanete e sua senhora Carina pelas maravilhosas tardes com música esquisita e degustáveis diversos.

Aos queridos amigos de Campinas Zé, Caio, MilHouse, Xarux, Cesão (Coconot Square Garden), Tu e Mari pelos momentos inesquecíveis de sapiência e inspiração.

À minha namorada Erica pela deliciosa companhia, carinho e dedicação em todas as horas.

Aos amigos do NICS e da FEEC pelo suporte a este trabalho.

À CAPES pelo apoio financeiro.

A todos aqueles que contribuíram para a finalização deste trabalho.

Sumário

Lista de Figuras

Lista de Tabelas

Trabalhos Publicados Pelo Autor

1. LUVIZOTTO, A, L; COSTA, C. R; Context Sensitive Harmonic Processor. In: Proc. of 11th Internacional Symposium on Computer Music, 2007, São Paulo, Brasil. v. 11.

2. LUVIZOTTO, A, L; COSTA, C. R; MANZOLLI, J; MENDES, R: S; VON ZUBEN, F. J; Harmonic Equalizer. In: 11^{a} Convenção Nacional da AES Brasil, 2007, São Paulo, Brasil.

3. LUVIZOTTO, A. L; FURLANETE, F. P; MANZOLLI, J; Microfonia e distorção na guitarra sob a ótica de waveshaping. In: XVI Congresso da Anppom, 2006, Brasilia - Brasil. opus, 2006.

4. LUVIZOTTO, A. L; MANZOLLI, J. ; ICHIZO, R. . Revisitando Waveshaping: implementando um plugin VST para distorcer sons de guitarra. In: Proc. of 10th Internacional Symposium on Computer Music, 2005, Belo Horizonte, Brasil. v. 10. p. 352-354.

Capítulo 1

Introdução

Can One Hear the Shape of a Drum?

Mark Kac

A frase acima foi título de um artigo publicado em 1966 por Mark Kac, no *The American Mathematical Monthly*, onde o autor, através de um estudo sobre as soluções da equação diferencial das ondas, expõe a seguinte questão: é possível, dados os auto-valores e auto-vetores associados a uma solução desta equação, determinarmos a geometria da membrana vibrante? Em outras palavras, é possível ouvir a forma de um tamborim? Podemos, sem muita formalidade, afirmar que a mesma composição espectral pode ser obtida através de diferentes formas, portanto a questão psicoacústica jamais poderá ser resolvida por um método único de análise. A experiência musical não pode ser atrelada a nenhum método de análise. Estamos diante de um conceito muito importante que é a sonoridade de uma forma, de um instrumento, elemento chave para a criação musical, como Kac mesmo escreve em seu trabalho: "soluções desta forma são de especial interesse tanto para os matemáticos quanto para os músicos". Desta forma, esta questão serve de metáfora para o trabalho que se segue, onde abordaremos a utilização de uma ferramenta matemática com o objetivo de melhor entender e representar sinais musicais.

A representação no domínio da freqüência de sons periódicos foi estudada por muitos cientistas, tais como Ohm, Helmholtz e Hermann. De acordo com os estudos de Ohm, as mudanças de fase, mesmo alterando a forma de onda, não participam da nossa percepção sonora. Já Helmholtz desenvolveu um método de análise harmônica com ressonadores acústicos onde a representação de um sinal é feita por somas de senóides. Estes primeiros estudos concluíram que a audição humana é incapaz de perceber mudanças de fase e que o timbre é exclusivamente determinado pelo espectro.

Muito dos estudos desenvolvidos nesta área devem-se particularmente ao ferramental matemático desenvolvido por Joseph Fourier (1768-1730). A transformada de Fourier, embora desvinculada inicialmente do contexto musical, possibilitou a exploração das idéias de Helmholtz e o embasamento matemático da noção de componentes harmônicos.

Em 1946, Dennis Gabor, conhecido pela descoberta do princípio da holografia, escreveu um artigo fundamental onde ele explicitamente define a representação tempo-freqüência de um sinal. Um ponto essencial desta teoria é a existência de uma relação recíproca entre o sinal no tempo e sua representação tempo-freqüência [Arfib, 1991]. Esta nova abordagem foi aplicada à transformada de Fourier, levando a uma representação tempo-freqüencial de um sinal, ou seja, com o trabalho de Gabor tornou-se possível a análise da evolução temporal do espectro de freqüências.

No final da década de 1950 e através de toda a década de 1960, nos laboratórios da Bell, em New Jersey - EUA, um grupo de pesquisadores incluindo-se Max Mathews, John Pierce e Jean-Claude Risset começou a explorar o computador como um novo instrumento musical, o elemento fundamental que iria inserir as teorias matemáticas sobre análise e síntese sonora no contexto musical.

Em 1963, Mathews escreveu: "não há limitações teóricas para a performance de um computador como uma fonte de sons musicais, em contraste à performance de instrumentos conhecidos. Atualmente, as possibilidades da computação musical são limitadas principalmente pelo custo computacional e pelos nossos conhecimentos de psicoacústica"[Mathews, 1963]. Após mais de quarenta anos, esta frase continua totalmente atual. Embora a questão de processamento tenha sido suprida em muitas aplicações, o entendimento psicoacústico permanece como grande limitação nos sons gerados pelo computador, dentro do contexto musical.

Desde então, modelos para a geração sonora vêm sendo criados através de inúmeros caminhos, todos orientados pela obtenção de resultados sonoros que sejam capazes de expressar experiências musicais desejadas. A descrição do que ouvimos ou desejamos ouvir passa inevitavelmente pelo entendimento psicoacústico e de conceitos complexos e cheios de sutilezas, como timbre e formas de representá-lo.

Recentemente, podemos notar a atenção que a comunidade científica vem dando, em diversas áreas, para o novo ferramental matemático proposto pela teoria *wavelet*. As aplicações são inúmeras, da geofísica à medicina, passando com grande potencial pela área musical. A transformada *wavelet* surge como uma nova ferramenta capaz de possibilitar análises com boa resolução em ambos os domínios, tempo e freqüência, sendo uma ferramenta apropriada para inspeções finas de transientes e elementos que são essenciais para um melhor entendimento das características sonoras de um instrumento. Sua flexibilidade e robustez permitem representar sinais musicais de forma extremamente eficaz. Esta será a ferramenta de engenharia utilizada nesta dissertação, pela qual propomos um novo caminho para se obter uma sonoridade desejada: a partir dos coeficientes gerados pela transformada *wavelet* e suas permutações.

1.1 Motivação

Dentre os processos de criação sonora musical, podemos notar que a intenção de imitar características naturais de instrumentos acústicos está sempre presente. A completa reprodução pode nem sempre ser o objetivo principal dos métodos de síntese. Porém, características como inarmonicidade, vibratos, ruídos, distorções provenientes da propagação das ondas no ar, e muitas outras, são elementos que motivam o processo. Ou seja, como qualificar a relevância dessas características ou como manipulá-las frente à criação de um novo som musical são questões profundamente subjetivas cujas iniciativas de resposta acabam recorrendo a análises de sons acústicos.

A questão de como estabelecer um método eficaz e robusto, que seja capaz de suprir as necessidades musicais nas suas diversas vertentes, dá origem a uma busca tecnológica na qual inserem-se as transformadas de Fourier, *wavelets* e outras ferramentas matemáticas que são de vasta aplicação na representação de sinais.

A alta dimensionalidade do espaço sonoro traz consigo um amplo universo de informações, com diferentes estruturas, onde a noção de timbre e sonoridade é fundamental. Porém, tratando-se de um espaço de tão alta dimensionalidade e complexidade é natural imaginar a existência de redundâncias. Alguns efeitos conhecidos, como mascaramentos, curvas de *loudness* e outros fenômenos psico-acústicos, são provas dessas redundâncias. Para analisar tais fenômenos, precisamos de uma ferramenta poderosa, capaz de atuar como um "telescópio"onde a cada nova escala, novos detalhes são revelados.

A análise em multi-resolução por transformada *wavelets* é uma técnica com a qual é possível visualizar um sinal em diferentes escalas, com diferentes níveis de detalhes espectrais, como um telescópio matemático. A sua implementação pode contribuir para a reprodução de atributos sonoros desejáveis, como análise de transientes, descontinuidades, modulações e outras características que não são observáveis através das transformadas de Fourier, comumente empregadas nestas tarefas. As novas possibilidades de síntese geradas por esta ferramenta são inúmeras e as possibilidades de gerar sons musicais vem sendo ampliadas.

1.2 Objetivos

No presente trabalho, propomos um modelo que, através da análise em multi-resolução pela transformada *wavelet*, nos permite trafegar pelas características psicoacústicas de sons distintos, de forma a investigar a relação entre seus timbres, destacando semelhanças e diferenças em suas composições representada pelos coeficientes *wavelet*. Desta forma, buscamos os seguintes objetivos:

- Realizar um estudo da literatura recente dos processos de análise e síntese de sinais musicais, assim como da ferramenta de engenharia envolvida, que são as transformadas *wavelets*.

- Executar a análise em multi-resolução entre pares de sons distintos a fim de detectar propriedades relativas aos seus coeficientes, que possibilitem transitar entre suas sonoridades, identificando semelhanças e diferenças.

- Gerar novos sons a partir de um modelo proposto a partir dos estudos realizados através da análise em multi-resolução.

- Representar um sinal musical através de um modelo simples, fundamentado pela análise em multi-resolução.

1.3 Justificativa

A síntese sonora é um campo onde o processamento digital de sinais tem um papel chave. Nas últimas décadas, testemunhou-se a explosão dos sintetizadores com a definição e a incorporação do protocolo MIDI (*Musical Instrument Digital Interface*), tornando possível a integração entre sistemas digitais dedicados, como os próprios sintetizadores, e mais recentemente entre eles e o computador.

Este campo de pesquisa acabou se espalhando por diversas outras áreas. Hoje podemos encontrar sistemas que são capazes de sintetizar sons nas mais diversas aplicações de multimídia e entretenimento, como efeitos sonoros e musicais de jogos, toques de telefones celulares e alarmes, geração sonora para realidade virtual como forma de terapia, dentre outras. Quase todas as pessoas hoje utilizam no seu dia-a-dia alguma forma de síntese sonora.

As técnicas antigas de síntese ao invés de desaparecerem, estão sendo trazidas de volta com a ajuda de novas técnicas de processamento de sinais digital. A exemplo disso, temos muitos sintetizadores, que foram construídos através de implementações analógicas, sendo implementados em *software* nos últimos anos [Bonada & Serra, 2007].

As aplicações das *wavelets*, por sua vez, estão tomando espaço e mostrando ser eficazes nas mais diversas áreas.

Dentro do processo de representação de sinais e síntese, Xavier Serra estabelece **??** que no seu desenvolvimento devem ser considerados alguns compromissos, tais como:

Qualidade sonora: relacionada à riqueza do som, um som de qualidade seria um som próximo a sons naturais enquanto um som pobre seria um som facilmente percebido como sintético e simples.

Flexibilidade: capacidade do método em obter variações do material sonoro com a variação de seus parâmetros.

Generalidade: a possibilidade de uma técnica gerar qualquer timbre.

Esforço computacional: quantidade de recursos computacionais necessários para a geração do material sonoro.

As *wavelets* permitem a criação de modelos de síntese que preenchem todos estes requisitos de forma eficiente. A representação permite uma boa qualidade sonora e com economia de recursos computacionais, pois os algoritmos são de baixo custo computacional. São altamente flexíveis, pois podem ser implementadas através de bancos de filtros digitais ortogonais e/ou biortogonais com reconstrução perfeita, sem aproximações, podendo explorar a síntese de cada nível de multi-resolução de forma separada e depois reconstruir o sinal de áudio sem perdas. Todos estes fatores somados permitem uma generalidade ampla na geração de timbres diversos, seja no aprimoramento de técnicas consagradas, como síntese aditiva, ou na proposição de novos métodos.

Dessa forma, a aplicação desta ferramenta na proposição de novos horizontes dentro do cenário de síntese de sons está bem focada e na direção das tendências atuais.

1.4 Estruturação da dissertação

Neste primeiro capítulo, apresentamos uma breve introdução sobre o tema e também os aspectos que motivaram este trabalho, com seus objetivos e justificativas.

Após esta introdução, destacaremos no segundo capítulo um panorama geral de como os métodos de síntese sonora vêm se desenvolvendo com o passar dos anos e como as novas ferramentas de análise, mais diretamente a transformada *wavelet*, estão contribuindo na expansão conceitual e no domínio de técnicas mais avançadas de processamento de sinal aplicadas a métodos consagrados, como síntese aditiva, subtrativa e granular.

No terceiro capítulo, um panorama geral das principais propriedades matemáticas necessárias para a compreensão do ferramental que constitui a teoria *wavelet* será apresentado, através de uma revisão breve da literatura deste tópico.

No quarto capítulo, uma descrição detalhada do modelo de representação de sinais musicais proposto é realizada. Nele, as etapas dos dois processos utilizados para a obtenção dos coeficientes *wavelets* são dissecados. Para cada uma das vertentes do modelo, um diagrama de blocos completo é exposto. Os capítulos 2 e 3 servem de base teórica para o modelo. Caso o leitor esteja familiarizado com esta teoria, é possível uma leitura direta do capítulo 4.

Os resultados ficam a cargo do quinto capítulo, onde os experimentos realizados durante a pesquisa são apresentados e avaliados.

Por fim, o último capítulo apresenta as conclusões e reflexões sobre o desenrolar pleno do trabalho e as projeções para futuros desenvolvimentos.

Capítulo 2

Síntese Sonora: Análise e Representação de Sinais Musicais

What Would We Like to See Our
Music Machines Capable of
Doing?

Xavier Rodet

Parece-nos intuitivo imaginar que a idéia de síntese sonora é realizar uma *mímeses* da sonoridade dos instrumentos musicais. Desenvolver um método no qual o resultado sonoro aproxima o som desses instrumentos. Pode-se também imaginar uma outra vertente, a qual se desenvolveu grandemente nos últimos anos, com o objetivo de gerar novas sonoridades. Ou seja, a criação de novos sons.

Os processos de análise e representação aliados aos métodos de processamento digital de sinais se constituem nas bases deste estudo, assim como o cerne do uso de ambientes computacionais para simulação e criação de material novo. Neste contexto o computador configura-se como um novo instrumento musical.

A epígrafe que incia este capítulo foi título de um artigo publicado no *Computer Music Journal* em 1991, onde alguns pesquisadores, dentre eles Xavier Rodet, Miller Pucket, Jean-Claude Rissset, discutem o papel da máquina na assistência às atividades musicais. No artigo, algumas delas são citadas: ouvir, modelar, reconhecer, transcrever, assistência composicional e composição algorítmica. Nos dias de hoje com o atual desempenho computacional podemos processar em tempo real, amostrar e reproduzir sons, também em tempo real, e o mais excitante: produzir novos sons, que podem ser totalmente sintéticos ou gerados através de manipulações de sons acústicos gravados e armazenados na memória do computador. Naquela época discutiam-se novos métodos de implementação em *hardware*, o emprego de redes neurais como nova técnica para a computação musical e o começo da utilização das *wavelets* no cenário musical. Hoje, olhando para o que se produziu nestes anos, podemos ter uma visão retroativa deste campo de pesquisa de forma a situar melhor o nosso trabalho.

Tendo esta proposta, o presente capítulo se divide em 2 seções principais: **Timbre e Sonoridade** onde alguns trabalhos representativos da literatura são revisitados; e **Síntese** onde descrevemos um panorama que aborda, inicialmente, métodos consagrados até nos aproximarmos das recentes inovações que se configuram como estado da arte envolvido na revitalização desta área de conhecimento. Ao final deste capítulo conduzimos o leitor às bases que motivaram a utilização das *wavelets* como ferramenta na expansão do domínio da representação e síntese de sinais sonoros e musicais desta tese.

2.1 Timbre e Sonoridade

A criação de novos sons por meio da tecnologia passa por uma infinidade de conceitos e conhecimentos e se faz necessário o emprego de diversas técnicas. No caso da análise e síntese de sons, com direcionamento musical, necessitamos de entender a estruturação sonora, procurar ferramentas que possibilitem detectar quais características são de maior relevância para os compositores e músicos e então buscar métodos de representação e implementação. Paralelamente, a compreensão dos fenômenos psicoacústicos envolvidos é um objetivo complexo e desejável, motivando muitos trabalhos voltados às questões subjetivas como timbre, sonoridade e suas formas de representação e quantificação.

Ao contrário de outras propriedades sonoras, como altura ou *loudness*, o timbre não pode ser diretamente ligado somente a uma dimensão física. Seu reconhecimento é um fenômeno cognitivo que resulta da interação de muitos atributos e do peso perceptual de cada um deles que, muitas vezes, não é mensurável [De Poli & Prandoni, 1997]. Desta forma, apresentar uma definição consistente para timbre é uma tarefa árdua e mais complexo ainda é desenvolver um modelo computacional satisfatório. A título de exemplo, a *Associação Americana de Padrões* em 1965 formulou a seguinte definição: "timbre é um atributo de sensação através do qual um ouvinte pode julgar dois sons, com mesma intensidade e altura, diferentes". Apesar da aparente simplicidade tal definição ainda perdura em compêndios renomados, o que mostra que há uma defasagem conceitual que deve ser expandida e estudada.

Schaeffer [Schaeffer, 1966] trata os dois termos títulos desta seção, timbre e sonoridade, de forma distinta. Ele define *sonoridade* como um atributo qualitativo e abstrato distintivo de um som, o que lhe garante identidade. Em relação ao termo *timbre*, refere-se às características acústicas que permitem uma associação do som com suas bases físicas, a sua materialidade e concretude. O timbre é o atributo que garante a identificação da fonte via propriedades acústicas. O conceito de sonoridade de Schaeffer inclui e expande a noção do senso comum de timbre, ao tratar o som como uma entidade autônoma. As bases do raciocínio de Schaeffer estão vinculadas a um ponto de vista fenomenológico, ou seja, além da materialidade descrita por parâmetros físicos quantificáveis, o som é um fenômeno por si só no qual fatores psicológicos, culturais, dentre outros, atuam e suportam a experiência auditiva de cada indivíduo.

A notação tradicional tem as suas bases numa representação que prioriza, de forma cartesiana, a distribuição de notas pelo tempo. Muitos compositores usaram altura, duração e intensidade de forma elaborada, mas o timbre parece ter participado, por anos, apenas implicitamente do processo composicional. Há uma complexidade em apriendê-lo de forma eficaz e criativa. Um dos primeiros fatores que levou à expansão do espaço timbrístico foi o desenvolvimento das técnicas de orquestração.

Dentre os diversos tratados de orquestração, os de compositores românticos como Hector Berlioz, ao explorarem os limites da orquestra, abriram as portas para um mundo novo. Outras abordagens mais recentes, especialmente a visão impressionista de Claude Debussy, ao considerar o timbre como parte da expressão musical, passa a articular o discurso musical nas transformações inerentes de blocos sonoros. Como numa pintura, a noção de cor passa a fazer parte da paleta composicional. Arnold Schönberg e a escola de Vienna ampliaram o papel do timbre na música. No pós-guerra, os compositores comumente articularam este atributo como elemento de destaque na estética musical [Hourdin et al., 1997]. Recentemente, no início da década de 1980, nasceu uma escola francesa de composição denominada "espectralismo"ou "música espectral"onde o timbre passa a ser o mais proeminente fator de organização dos processos composicionais [Barriere, 1994]. Para Barriere, o timbre constitui-se na própria metáfora que engendra o projeto criativo.

Em contrapartida, inúmeros pesquisadores de diversas áreas como música, engenharia, matemática, física e psicologia foram atraídos pelo estudo do timbre. Este esforço de pesquisa investiga propriedades capazes de possibilitar a representação sonora, de forma menos abstrata, mais analítica e mensurável. Os primeiros estudos são creditados a Hermann Helmholtz (1863) que obteve resultados pioneiros, mas limitados. Os resultados de seu estudo apontaram para uma modelagem do timbre fortemente ancorada na noção de série harmônica. Na realidade, como mencionaremos a seguir, a distribuição harmônica das componentes espectrais é uma aproximação sendo que a grande maioria dos sons musicais tem, em algum grau, componentes inarmônicas. Depois da primeira síntese digital feita por Max Mathews em 1958, ficou claro que o timbre não dependia somente da sua distribuição espectral, mas também da fase e de outras propriedades temporais desconhecidas.

Nas análises feitas por Jean-Claude Risset de sons de trompete [Risset, 1965], ele descobriu uma importante propriedade. Para sons de instrumentos de sopro, do naipe dos metais, a proporção entre os parciais do espectro sonoro aumenta com a intensidade. Esta descoberta revelou que características espectrais e temporais são inseparáveis.

Outros trabalhos idealizaram um espaço de representação tridimensional para o timbre, atribuindo a cada uma das dimensões espaciais uma característica física audível. Dois trabalho pioneiros se destacam: [Grey, 1977, Grey & Gordon, 1978]. Neles, Grey representou, na figura 2.1, o julgamento psicoacústico subjetivo de similaridade timbristica, a partir de uma população de indivíduos com formação musical.

A partir deste oceano de possibilidades, alguns métodos utilizados em sistemas de extração de conteúdo timbrístico foram empregados na busca de características mais definidas e delimitadas. No trabalho de Sandell [Sandell, 1995] dois sons se fundem gerando um terceiro. Tal propriedade está diretamente relacionada à distribuição espectral dos dois sons envolvidos. Ou seja, essa fusão depende da proximidade entre os centróide espectrais. Quanto menor a diferença, maior a noção de amálgama timbrístico.

Os estudos pioneiros de Grey [Grey & Gordon, 1978] iniciaram a trajetória de uma área do conhecimento denominada de Sonologia. Neste campo, estuda-se a interação entre modelos acústicos/matemáticos e psicoacústica. Aproximadamente vinte anos após essa publicação, De Poli e

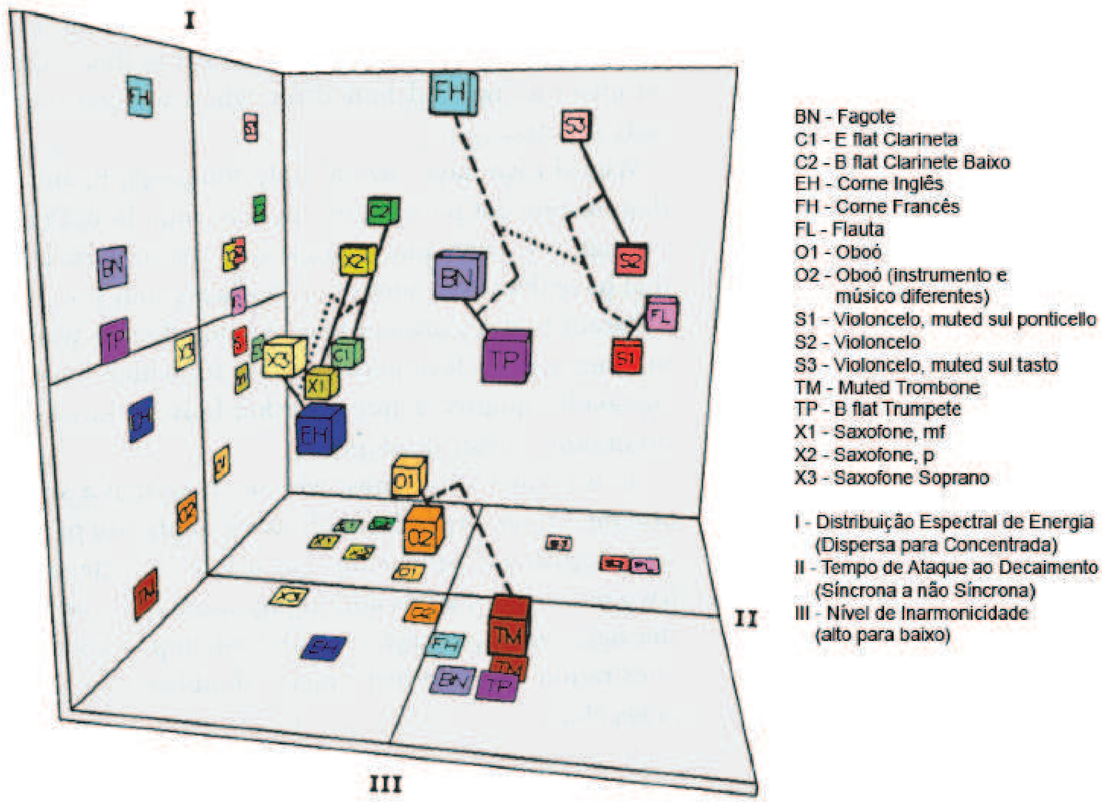

Fig. 2.1: Espaço timbrísico tridimensional proposto por Grey [Grey, 1977]

Prandoni em [De Poli & Prandoni, 1997] apresentaram resultados por eles denominados de modelos sonológicos. A partir de métodos de processamento de sinais de fala e representação de dados como redes neurais e análise de componentes principais (PCA), eles reduziram a complexidade da representação. Num sistema de coordenadasas características principais do timbre estão representadas por vetores que descrevem algumas qualidades do timbre musical de forma analítica sem, necessariamente, relacioná-las a uma avaliação subjetiva do ouvinte.

Nesta mesma direção, outro trabalho sobre representação foi desenvolvido por Loreiro et al em [Loureiro et al., 2004]. Por meio de análises de componentes principais (PCA) e mapas auto-organizáveis, eles representaram trajetórias em um espaço timbrístico as quais foram definidas a partir de coordenadas descritas por três componentes principais (PCA). Desta forma, uma seqüência de notas executadas por uma clarineta, foi utilizada para traçar as trajetórias apresentadas na figura 2.2. Segundo os autores, a criação de subespaços espectrais, utilizando-se de todos os sons possíveis de um instrumento, possibilita uma representação compacta para toda a sua paleta timbrística.

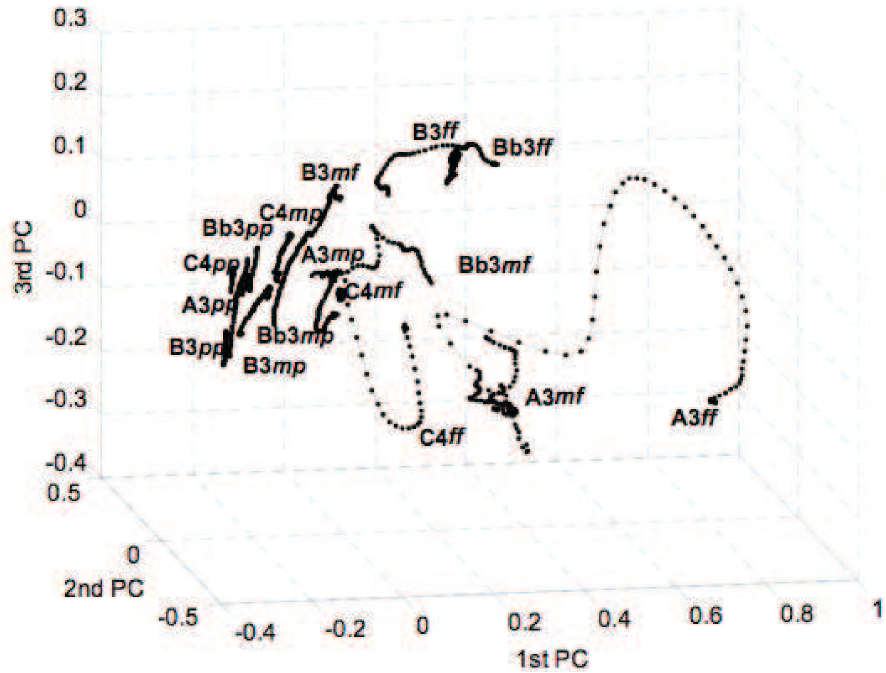

Fig. 2.2: Espaço de representação timbral proposto por [Loureiro et al., 2004]. Nesta figura, cada eixo corresponde a um componente principal.

Outros trabalhos interessantes foram realizados com o auxílio de novas ferramentas, como é o caso dos trabalhos de Su e Jeng [Su & Jeng, 2001], que utilizando-se da transformada *wavelet*, propuseram um novo método para reconhecimento de acordes musicais, baseando-se em um modelo de audição humana por bancos de filtros. Já Kostek et al [Kostek et al., 2002, Kostek et al., 2005] propõem um sistema descritor que se baseia na extração de características que são úteis no reconhecimento de sons musicais ao juntar análises *wavelets* com redes neurais.

A partir de bases psicoacústicas, Spiegelberg [Spiegelberg, 2002] desenvolve um novo método de análise musical vinculado ao estudo das articulações e transientes de fragmentos musicais. Segundo o autor, sua pesquisa serve de suporte para estudos musicológicos focados não só em análises timbrísticas, mas também em psicologia da performance, análise e performance propriamente.

2.2 Síntese Sonora

Todos estes estudos estão alinhados de forma prática a dois objetivos: conhecer e representar melhor elementos sonoros de forma a refinar o processo de expressão musical. O processo, pelo qual um compositor passa, ao imaginar um som, desvenda-lo dentro de uma infinidade de sons existentes, desenvolver um processo simples para obtê-lo, selecionar as tecnologias e softwares para sua criação e então empregá-lo como desejado nutre-se completamente deste binômio análise e representação. Nesta seção serão abordados alguns métodos de síntese tradicional que foram revitalizados nos últimos anos através do emprego de novas tecnologias. Informações complementares sobre os métodos tradicionais de síntese, não abordados aqui, podem ser obtidas pelo leitor em [De Poli, 1983].

O primeiro método de síntese de grande impacto foi a síntese aditiva, que fez uso direto das idéias propostas por Helmholtz e foi comumente implementada baseando-se na transformada de Fourier, onde osciladores senoidais foram empregados na constituição de um espectro de freqüências. Porém, as primeiras implementações resultaram em material sonoro por vezes questionadas nas suas aplicações musicais. Estudos mais aprofundados e que propunham modelos não harmônicos foram desenvolvidos por Risset [Risset, 1965], dentre outros. Revisitar e reformular métodos consagrados como este, têm sido uma prática recorrente durante os últimos anos.

O *Spectral Modeling Synthesis* (SMS), de Serra e Smith [Serra & Smith, 1989], incorpora um modelo de detecção de picos espectrais que sofistica a representação do envelope de cada parcial, assim como um modelo estocástico de representação dos transientes, denominado resíduo, dando origem a um processo de síntese aditiva revitalizada, figura 2.3. Ainda, Bélltran em [Beltran & Beltran, 2003] utilizando um banco de filtro por *wavelets* de Morlet que refina o SMS, incrementando ainda mais o processo de síntese aditiva, figura 2.4.

Os trabalhos realizados por Evangelista e seu grupo de pesquisa em tecnologia sonora, sofisticaram ainda mais os modelos de síntese, incluindo modelos aditivos fractais [Polotti & Evangelista, 2007], modelagem física [I. Testa et al., 2004], com o emprego das transformadas [**?**, Evangelista, 1993, Evangelista, 1996, Evangelista & Cavaliere, 1998a, Evangelista & Cavaliere, 1998b], permitindo a representação de modelos inarmônicos de forma altamente satisfatória[1].

Aliados à síntese aditiva, ou por métodos não-lineares de síntese como *waveshaping* [Lebrun, 1979], modulação em anel [De Poli, 1983] ou modulação em freqüência [Chowning, 1971], bancos de filtros por *wavelets* podem ser empregados para esculpir o espectro sonoro. Neste con-

[1]Os resultados sonoros podem ser conferidos em `http://staffwww.itn.liu.se/~giaev/ST_giaev.html`

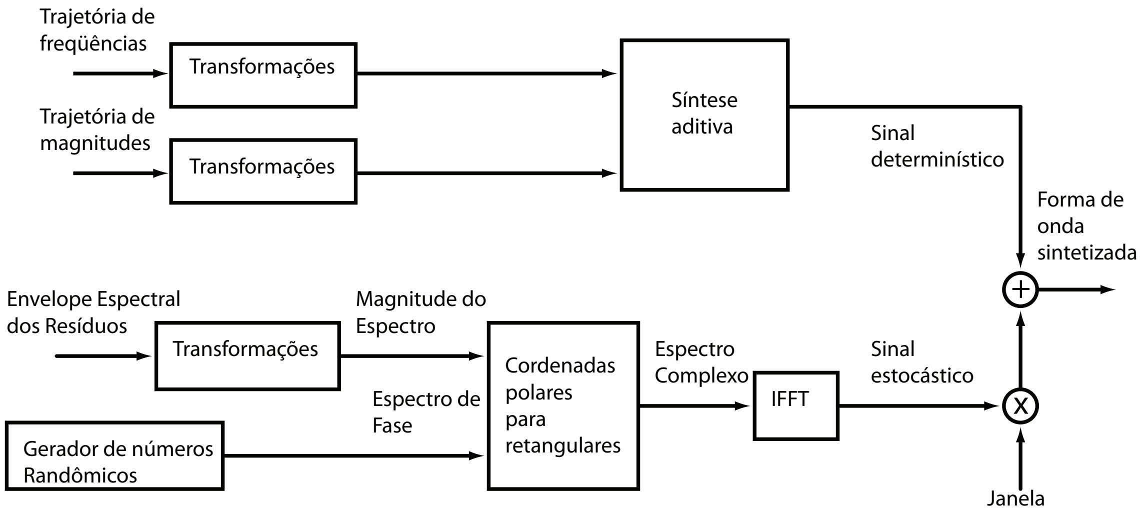

Fig. 2.3: Fluxograma do método SMS, extraído de [Serra & Smith, 1989].

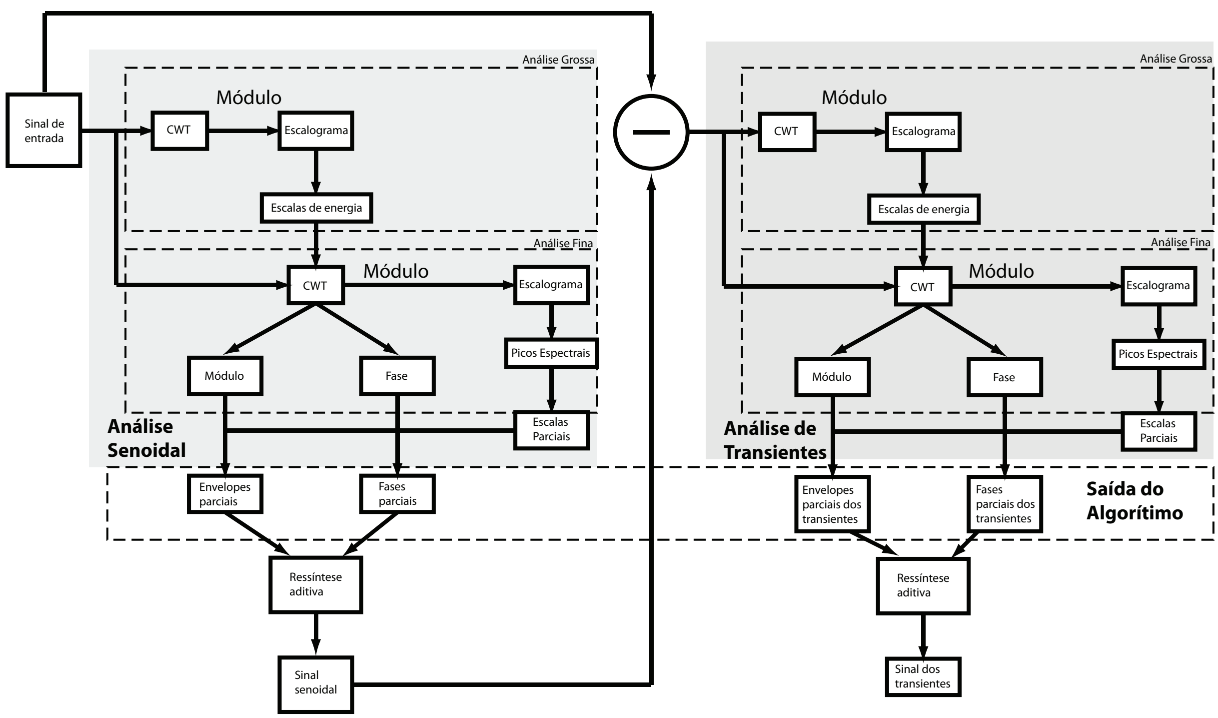

Fig. 2.4: Fluxograma do método de síntese proposto por [Beltran & Beltran, 2003].

texto, um dos subprodutos do nosso trabalho foi um processador de sinais que se utilizou de um banco de filtros desenvolvidos a partir de *wavelets* de Morlet. Como resultado apresentado em [Luvizotto & Costa, 2007, Luvizotto et al., 2007], obtivemos uma ferramenta capaz de explorar tanto a resolução em freqüência como a temporal. Trata-se de um processador formado por um detector de fundamentais aliado a um banco de filtros. A fundamental do sinal de entrada é detectada e o banco de filtros é sintonizado de modo a atenuar ou acentuar cada uma das n primeiras parciais, conforme o usuário desejar. Os parâmetros são ajustados através de controles deslisantes como ilustrado em 2.5. A figura 2.6 mostra a representação do espectro de Fourier em cascata de uma nota D2 de um piano, onde somente a fundamental está presente. Todas as outras parciais foram eliminadas através do banco de filtro .

Podem-se vislumbrar muitas aplicações para este processador. Por exemplo, é possível aplicar a curva da evolução temporal das fundamentais de um sinal de fala em um solo de guitarra, escolhendo-se as parciais que irão ser alteradas a partir da afinação da fala. A sonoridade resultante preserva a variação tonal e rítmica vocal assim como as características originais do solo através da manipulação espectral executada pelos filtros *wavelets*.

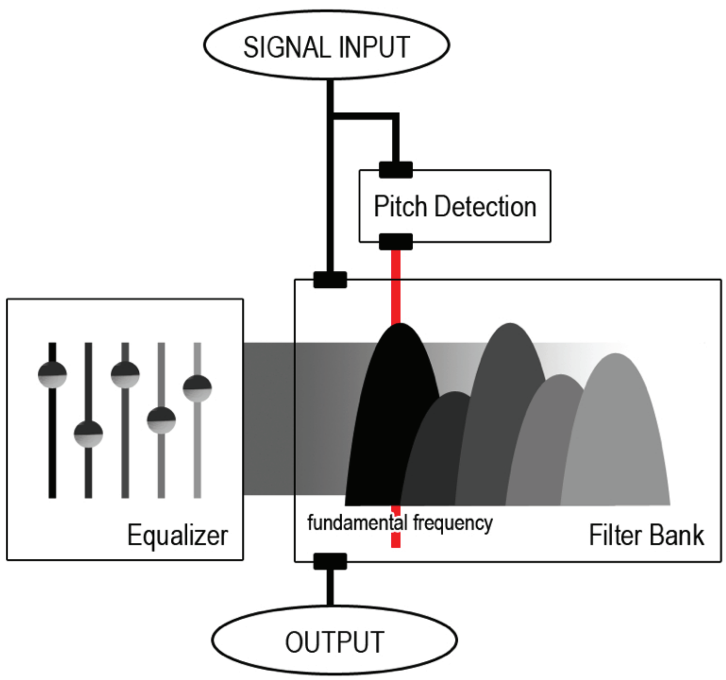

Fig. 2.5: Representação esquemática do processador harmônico proposto em [Luvizotto & Costa, 2007].

Técnicas de síntese com formas de ondas fixas, produzem sons estáticos. Uma característica fundamental de um som musical é a sua evolução timbrística temporal. Uma forma alternativa de representar um sinal sonoro pode ser traçada através de uma seqüência de sons elementares, de durações constantes que o constitui analogamente a um filme, no qual uma imagem em movimento é

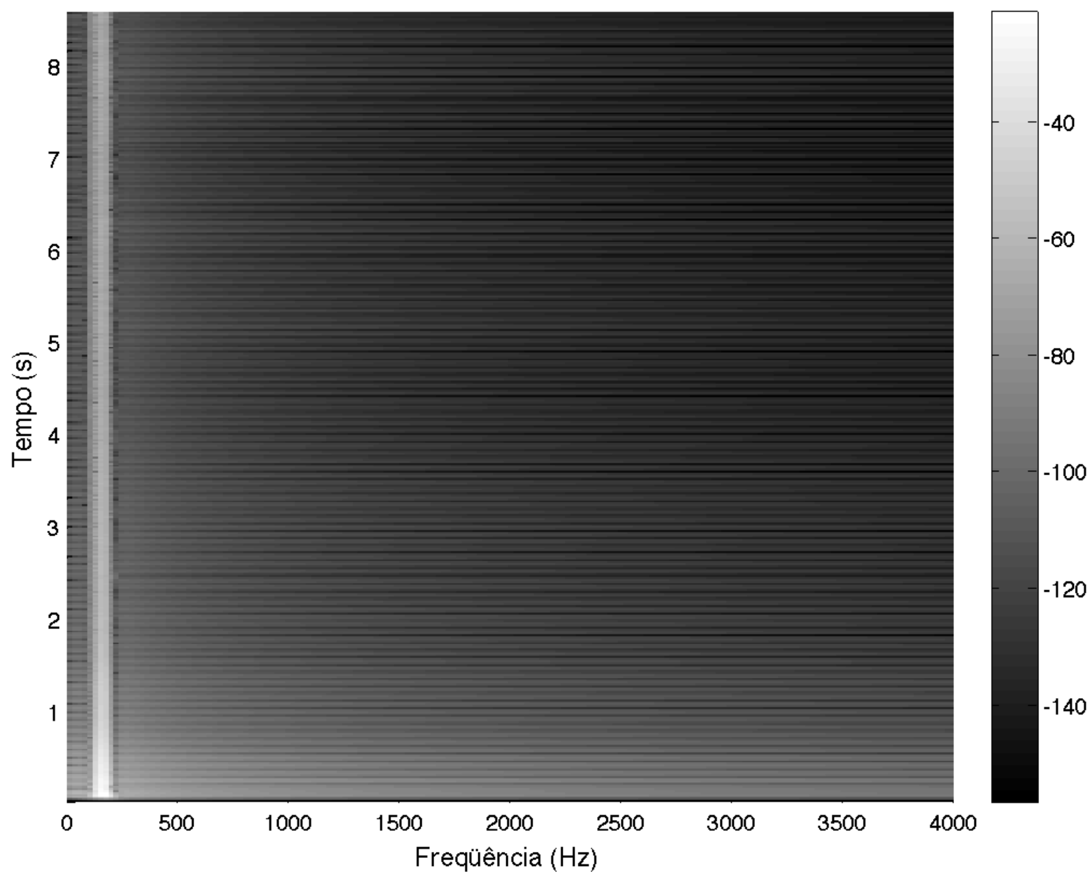

Fig. 2.6: Representação do espectro via sonograma da nota D2 de um piano, onde somente a fundamental está presente, obtido através do banco de filtros desenvolvido por [Luvizotto & Costa, 2007].

produzida por uma seqüência de fotos.

Esta abordagem em música levou às técnicas da síntese granular [Roads, 1978] que obtiveram especial atenção da comunidade científica nas décadas de 1980 e 1990. Esses sons elementares são os chamados *grãos*. Esta abordagem foi inicialmente proposta pela teoria da audição do físico Dennis Gabor [Gabor, 1947b, Gabor, 1947a] que se referiu a um grão como *Quanta Sonoro*, e postulou que qualquer som poderia ser descrito pela teoria granular ou quântica. Esta visão, a qual incorpora um modelo temporal ao som, oferece um modelo mais completo que a abordagem atemporal das análises de Fourier. Ela também sugere uma interpretação alternativa das decomposições e recomposições por *wavelets*. Por esta abordagem, as *wavelets* seriam os grãos e os valores dos coeficientes associados a intensidade com que cada grão participa do processo. A representação final do sinal depende da *wavelet* escolhida como grão, dos coeficientes e também da disposição temporal destes coeficientes. Podemos criar diferentes sons a partir dos mesmos grãos e coeficientes através de diferentes ordenações [Evangelista, 1991].

Os grãos podem ser produzidos por um simples oscilador ou por outros métodos, com durações variáveis na ordem de 5-50 ms. Há duas formas de se implementar um sintetizador granular. A primeira, é organizar os grãos em quadros, como em um filme. A cada quadro, os parâmetros de todos os grãos são atualizados, como na figura 2.7. Esta foi a abordagem adotada por Xenakis, que também foi o primeiro compositor a explicar, em seu livro *Formalized Music* [Xenakis, 1971], uma teoria composicional para a teoria sonora dos grãos. Aplicações em tempo real foram desenvolvidas por Barry Truax em [Truax, 1988] em consonância com esta abordagem desenvolveu-se toda uma vertente estética baseada na noção de paisagem sonora [Truax, 1996].

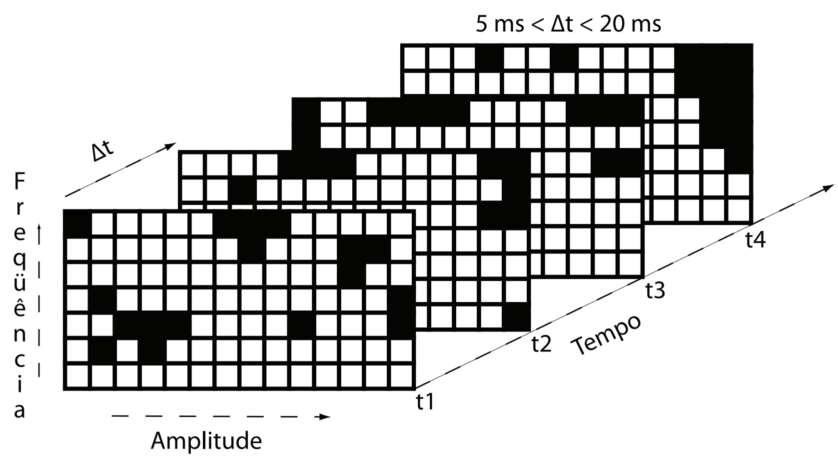

Fig. 2.7: Organização granular por quadros, como adotada por Xenakis em [Xenakis, 1971].

A segunda forma de organização envolve espalhar os grãos com uma máscara, que cerca uma região temporal com freqüência e amplitudes particulares [Roads, 1978]. Assim, os grãos são organizados em eventos, para o quais Roads usou doze parâmetros para caracterizar:

1. Tempo inicial.
2. Duração.
3. Forma de onda inicial.
4. Curva de variação da forma de onda (taxa de transferência de uma senóide para um pulso de banda limitada).

5. Freqüência central inicial.

6. Curva de variação da freqüência.

7. Largura de banda.

8. Curva de variação da largura de freqüência.

9. Densidade inicial dos grãos.

10. Amplitude inicial.

11. Curva de variação da amplitude.

Em um evento, os grãos são espalhados randomicamente através da densidade inicial e de sua curva de variação. Uma vez que os grãos são organizados desta forma, torna-se possível variar os parâmetros *en masse*, ou misturar grãos com diferentes parâmetros para criar nuvens de espectro sonoro [Roads, 1978].

Uma grande variedade de formas de ondas podem ser usadas dentro dos grãos, de simples senóides, passando por sinais modulados em freqüência até formas de ondas gravadas de sons naturais, como no caso das paisagens sonoras [Truax, 1996]. Uma vez que os grãos são obtidos de sons do meio ambiente, eles podem ser recombinados com diferentes espaçamentos, ordens e até misturados utilizandos-se grãos de diversas fontes para compor texturas inéditas ou o que é chamado de paisagens sonoras artificiais.

Como foi apontado na introdução deste capítulo, seu objetivo, ao recapitular os métodos de síntese vigente e fazer um apanhado histórico, foi mostrar que certos problemas que surgiram no desenvolvimento da síntese de som podem ser atacados através do ferramental derivado da teoria de *wavelets*. Então, para avançar no estudo proposto para esta dissertação, fazemos a seguinte divisão: a) teorização das *wavelets* onde se apontam algumas das principais propriedades matemáticas envolvidas a ser apresentada no capítulo a seguir; b) descrição do modelo de representação abordado no capítulo 4.

Capítulo 3

Teoria Wavelet

3.1 Uma breve introdução histórica

Separar um fenômeno complicado em vários outro mais simples para então estudá-lo, é uma abordagem padrão em diversas áreas da ciência. As análises *wavelet* faz uso desta idéia, pois envolve a representação de funções quaisquer em termos de blocos construtivos básicos, de forma fixa, porém em diferentes escalas e posições. Muitos dos trabalhos foram realizados nos anos de 1930. Uma investigação da história da matemática nos revela diferentes origens para a análise *wavelet* [Meyer, 1993].

Em matemática abstrata, é sabido há algum tempo que técnicas baseadas em séries e transformadas de Fourier podem não ser adequadas para a representação de muitos problemas. A primeira menção ao que hoje chamamos de *wavelet* apareceu em 1909, no apêndice da tese doutoramento de A. Haar, que enquanto trabalhava na construção de bases para representar funções integráveis quadraticamente, chegou a primeira *wavelet* ortogonal da qual se tem notícia. Uma outra propriedade desejável da *wavelet* de Haar é o suporte compacto, porém infelizmente ela não é continuamente diferenciável, o que de alguma forma limita sua aplicação.

Na década de 1930, vários grupos independentes trabalharam na representação de funções usando funções bases com escala variável. O físico Paul Levy, ao investigar os movimentos Brownianos constatou que as bases de Haar eram mais eficientes no estudo de pequenos detalhes, de alta complexidade, no movimento Brawniano [Grapes, 1995].

Outros importantes estudos, também datados nos anos de 1930, foram realizados por Littlewood, Paley e Stein envolvendo o cálculo de energia de uma função $f(t)$:

$$Energia = \int_{-0}^{2\pi} \mid f(t) \mid^2 dt \tag{3.1}$$

O cálculo produzia diferentes resultados se a energia estivesse concentrada em alguns pontos ou distribuída sobre um grande intervalo. Este resultado perturbou os cientistas porque indicava que a energia poderia não ser conservada. Os pesquisadores descobriram então uma função que podia variar em escala e conservar a energia funcional. A teoria desenvolvida levou o nome de Littlewood-Paley.

Nas décadas de 1950 e 1960, estes desenvolvimentos forneceram as bases teóricas ao desenvolvimento de poderosas ferramentas para o estudo de outros tópicos, como soluções de equações diferenciais parciais e equações integrais [Jawerth & Sweldens, 1994] . Calderón abordou em seus trabalhos a decomposição atômica, introduzindo a identidade de Calderón, uma decomposição de um operador de identidade.

No começo dos anos de 1980, Strömberg descobriu as primeiras *wavelets* ortogonais, na tentativa de melhor compreender os espaços de Hardy, assim como outros espaços usados para medir o tamanho e a suavidade de funções [Strömberg, 1981]. Apesar disto, o sistema introduzido por Yves Meyer em 1985 recebeu maior reconhecimento, sendo conhecido como base de Meyer.

Yves Meyer e seu grupo de colaboradores chegaram a grandes resultados sobre a teoria de *Calderón-Zygmund*, em particular a representação de *Littlewood-Paley* e perceberam que estas técnicas poderiam levar a uma concepção unificada de muitos resultados da análise harmônica, e mais: que poderiam substituir as séries de Fourier em aplicações numéricas [Jawerth & Sweldens, 1994].

Através dessas pesquisas foi possível relacionar as várias teorias - da decomposição de Littlewood-Paley à identidade de Calderón. Grossman e Morlet em 1984 sugeriram, pela primeira vez, o termo *wavelet* para os blocos construtivos básicos, e o que antes era conhecido por teoria de *Littlewood-Paley*, passava a ser chamado de teoria *wavelet*.

Em 1985, Stephane Mallat deu as *wavelets* um salto adicional através de seus trabalhos em processamento digital de sinais. O autor descobriu algumas relações entre filtros QMF (*Quadrature Mirror Filter*), algoritmos piramidais e bases *wavelets* ortogonais. Em 1986, Mallat e Meyer desenvolveram a teoria da análise em multi-resolução, que proporcionou uma explanação satisfatória para todas essas construções, e disponibilizou uma ferramenta para a construção de outras bases [Daubehies, 1992].

Por volta de 1988, Ingrid Daubechies, utilizando os trabalhos de Mallat, extendeu o trabalho de Haar, construindo sua própria família de wavelets ortogonais, com suporte compacto e contínua, 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* vêm crescendo amplamente ano após ano, assim como a difusão de seu uso em inúmeras áreas, com grande enfôque em processamento digitais de sinais.

3.2 Transformada Wavelet Contínua

As *wavelets*, como o próprio nome sugere, são obtidas através de dilatações e contrações da *wavelet* mãe $\psi(t) \in L^2(R)$ com média nula:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0; \tag{3.2}$$

O objetivo principal é representar funções $\in L^2(R)$ através da decomposição no espaço gerado pela *wavelet* mãe. Uma família de *wavelets* é gerada pela $\psi(t)$ através de dilatações, por meio de um fator de escalamento a, e translações por meio de uma variável b através da expressão

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\Big(\frac{t-b}{a}\Big). \tag{3.3}$$

A transformada contínua *wavelet* (CWT) de $f \in L^2(R)$ no tempo b e escala a é definida por

$$W_f(a,b) = \int_{-\infty}^{+\infty} f(t)\psi^*_{a,b}(t)dt. \tag{3.4}$$

Podendo ser reescrita como um produto de convolução

$$W_f(a,b) = \int_{-\infty}^{+\infty} f(t)\psi^*_{a,b}(t-b)dt = f * \tilde{\psi}_a(b) \tag{3.5}$$

com

$$\tilde{\psi}_a(t) = \frac{1}{\sqrt{a}}\psi^*\Big(\frac{t}{a}\Big). \tag{3.6}$$

onde $*$ denota o complexo conjugado.

A transformada de Fourier de $\psi_a(t)$ é

$$\hat{\psi}_a(\omega) = \sqrt{a}\hat{\psi}(a\omega) \tag{3.7}$$

Como $\hat{\psi}(0) = \int_{-\infty}^{+\infty} \psi(t)dt = 0$, temos que $\hat{\psi}$ é a função de transferência de um filtro passa-faixa. A convolução da equação 3.5 calcula a transformada *wavelet* através de filtros passa-faixas dilatados [Mallat, 1998]. Da mesma forma que a transformada enjanelada de Fourier, a transformada *wavelet* pode medir a evolução temporal dos transientes freqüenciais do sinal. Isto requer o uso de uma *wavelet* analítica complexa, a qual separa os componentes de amplitude e fase, possibilitando medir as freqüências instantâneas do sinal [Kroland-Martinet, 1988]. Em contraste, as *wavelets* reais são freqüentemente utilizadas para detectar abruptas transições no sinal.

Para que uma *wavelet* mãe $\psi(t)$ possa dar origem a uma família de wavelets, *exige-se* que [Kroland-Martinet, 1988]:

- $\psi(t)$ seja absolutamente integrável:

$$\int_{-\infty}^{+\infty} \mid \psi(t) \mid dt < \infty \tag{3.8}$$

- tenha energia finita:

$$\int_{-\infty}^{+\infty} \mid \psi(t) \mid^2 dt < \infty \tag{3.9}$$

- e que satisfaça a condição de admissibilidade:

$$C_\psi = \int_{-\infty}^{+\infty} \frac{\mid \psi(\omega) \mid^2}{\omega} d\omega < \infty. \tag{3.10}$$

Esta última condição foi provada pela primeira vez em 1964 pelo matemático Calderón, por um enfoque distinto. Posteriormente, Grossmann e Morlet provaram o mesmo resultado para a área de processamento de sinais. O teorema de Calderón, Grossmann e Morlet, da onde sai a condição acima, e sua prova podem ser encontrados em [Mallat, 1998]. Para garantir que a integral do sinal seja finita $\hat{\psi}(0)$ tem que ser nulo, o que explica o motivo pelo qual as *wavelets* devem ter média nula. Na prática, esta última condição implica que a wavelet oscila, integra-se a zero e possui valor DC nulo.

Além destas propriedades, outras podem ser desejadas ou exigidas das famílias wavelets para que sejam úteis em aplicações específicas na área de processamento digital de sinais e na análise de espaços funcionais. Abaixo, relacionam-se tais propriedades[Faria, 1997].

- 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

Satisfeita a condição de admissibilidade, é possível recuperar a função $f(t)$ através de $W_(a,b)$ resolvendo-se a identidade de Calderón [Mallat, 1998] [Daubehies, 1992] resultando na transformada *wavelet* (contínua) inversa

$$W_f(a,b) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} W_f(a,b) \psi_{a,b}(t) da \frac{db}{a^2}. \tag{3.11}$$

A constante C_ψ é a constante de Calderón.

3.2.1 A função escala ϕ

Quando $W_f(a,b)$ é conhecido somente para $a < a_0$, a recuperação de f envolve informações complementares sobre $W_f(a,b)$ para $a > a_0$. Isto é obtido através da introdução de uma *função escala* ϕ que é um "acúmulo"de *wavelets* em escalas maiores que 1. O módulo de sua transformada de Fourier é definido por

$$\mid \hat{\phi}(\omega) \mid^2 = \int_1^{+\infty} \mid \hat{\psi}(a\omega) \mid^2 \frac{da}{a} \tag{3.12}$$

A função escala pode então ser interpretada como sendo a resposta ao impulso de um filtro passa-baixas [Mallat, 1998]. Sendo denotada por:

$$\phi_a(t) = \frac{1}{\sqrt{a}}\phi\Big(\frac{t}{a}\Big) \tag{3.13}$$

com $\tilde{\phi}_a(t) = \phi_a^*(-t)$ a aproximação para as baixas freqüências fica:

$$Lf(a,b) = \Big\langle f(t), \frac{1}{\sqrt{a}}\phi\Big(\frac{t}{a}\Big)\Big\rangle = f * \tilde{\phi}_a(t) \tag{3.14}$$

3.2.2 Transformada Wavelet Discreta

Seja $f(t)$ um sinal contínuo no tempo, uniformemente amostrado em intervalos N^{-1} sobre [0,1]. Sua transformada wavelet só pode ser calculada em escalas $N^{-1} < s < 1$. No cálculo discreto, é mais fácil normalizar a distância de amostragem para 1 e considerar o sinal dilatado $f(t) = f(N^1 t)$. Uma mudança de variável na transformada contínua wavelet 3.4 mostra que [Mallat, 1998]:

$$W_f(u,s) = N^{\frac{-1}{2}} W(Nu, Ns) \tag{3.15}$$

Para simplificar a notação, vamos considerar f e denotar por $f[n] = f(n)$ o sinal discreto de tamanho N. Sua transformada *wavelet* discreta é calculada através de escalas $s = a^j$, com $a = 2^{\frac{1}{\nu}}$, que provê ν escalas intermediárias em cada oitava $[2^j, 2^{j+1})$.

Seja $\psi(t)$ uma *wavelet* cujo suporte está incluído em $[-K/2, K/2]$. Para $2 \le a^j \le NK^{-1}$, a *wavelet* discreta escalonada por a^j é definida por

$$\psi_j[n] = \frac{1}{\sqrt{a^j}}\psi\Big(\frac{n}{a^j}\Big). \tag{3.16}$$

Tratando $f[n]$ e $\psi_j[n]$ como sinais periódicos de tamanho N, a transformada discreta pode ser calculada através de uma convolução circular por meio do algoritmo de transformada rápida de Fourier (FFT). Analogamente, a versão discreta da função escala 3.13 fica:

$$\phi_j[n] = \frac{1}{\sqrt{a^j}}\psi\Big(\frac{n}{a^j}\Big) \tag{3.17}$$

para $n \in [-N/2, N/2]$, também calculada através de convolução circular com $f[n]$ por meio de FFT.

3.2.3 Escala Versus Freqüência

A "freqüência instantânea"é freqüentemente considerada como um modo de introduzir dependência freqüencial ao longo do tempo. Dado um sinal com espectro amplo, as freqüências instantâneas contribuem com diferentes componentes espectrais ao decorrer do tempo na sua composição. Para uma maior precisão temporal, é necessário uma representação tempo-freqüência bidimensional $G(\omega, t)$ do sinal $f(t)$ composto pelas características espectrais de cada instante. A transformada de Fourier foi inicialmete adaptada por Gabor [Gabor, 1947b] com esse propósito, definindo $G(\omega, t)$ da seguinte maneira. Considere um sinal $f(t)$ e assuma-o estacionário quando visto através de uma janela $\phi(t)$ de extensão finita, centralizada sobre o tempo b. A transformada de Fourier

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt \tag{3.18}$$

do sinal janelado $f(t)\phi(t-b)$ é dada por

$$G(\omega, b) = \int_{-\infty}^{+\infty} f(t)\phi(t-b) e^{-j\omega t} dt \tag{3.19}$$

chamada de *transformada enjanelada de Fourier*, conhecida por STFT do inglês *Short Time Fourier Transform*.

Gabor utilizou inicialmente a função Gaussiana

$$g_\alpha(t) = \frac{1}{2\sqrt{\pi\alpha}} e^{\frac{-t^2}{4\alpha}} \tag{3.20}$$

com α > 0, como função janela passa-baixas $\phi(t)$. Assim, 3.20 é também conhecido na literatura como transformada de Gabor.

As análises através da STFT dependem criticamente da janela utilizada $\phi(t)$ que pode ser vista, no domínio da freqüência, como uma janela-freqüencial $\hat{\phi}(\omega)$ capaz de percorrer todo o eixo de freqüências de forma que a janela passa-baixa $\phi(t)$ pode ser usada como um filtro passa-faixa. Mais precisamente, enquanto a função temporal modulada

$$e^{j\omega_0 t}\phi(t-b) \tag{3.21}$$

em função de t localiza o sinal $f(t)$ "próximo de $t = b$ ", sua correspondente função janela freqüencial modulada

$$\frac{e^{j\omega_0 b}}{2\pi}e^{-jb\omega}\hat{\phi}(\omega-b) \tag{3.22}$$

em função de ω localiza o espectro (ou transformada de Fourier) $\hat{f}(\omega)$ do sinal próximo a $\omega = \omega_0$ [Chui, 1997], ou seja, a STFT pode ser vista como um banco de filtros modulado [Portnoff, 1980].

Através desta interpretação dual, podemos traçar uma relação entre a resolução no do domínio do tempo e no domínio da freqüência. Assumindo o valor $2\Delta_{\hat{\phi}}$ como sendo a largura RMS da função janela freqüencial $\hat{\phi}(\omega)$ e $2\Delta_{\phi}$ como a duração RMS da janela temporal $\phi(t)$, pelo *princípio da incerteza*, qualquer função janela $\phi(t)$ deve satisfazer

$$2\Delta_{\phi}2\Delta_{\hat{\phi}} \geq \frac{1}{2} \tag{3.23}$$

a igualdade acima garante-se somente quando $\phi(t)$ é dada por 3.20. Em outras palavras, qualquer função janela não pode ter área menor que 2 e qualquer função Gaussiana é uma janela-tempo freqüencial ótima (ou menor) [Chui, 1997] .

A equação 3.23 garante que uma boa resolução temporal leva a uma localização ineficientes no domínio da freqüência e vice-versa. A figura 3.1 mostra a janela tempo-freqüência correspondente a função $\psi(t)$ em várias localizações tempo-freqüênciais b, ω_0. Observe que a área da janela é sempre dada por $4\Delta_{\phi}\Delta_{\hat{\phi}}$.

Já na transformada wavelet a $\phi(t)$ deve necessariamente ser uma função passa-faixa, ao invés de passa-baixa, dado que seu valor DC tem obrigatóriamente que ser nulo. E ainda a largura de sua janela temporal é dada por $2a\Delta_{\psi}$ e aumenta conforme o fator de escala $a > 0$ diminui. No caso das wavelets, a *área* da janela é constante, dada por

$$(2a\Delta_{\psi})\frac{2}{a}\Delta_{\hat{\psi}^+} = 4\Delta_{\psi}\Delta_{\hat{\psi}^+}. \tag{3.24}$$

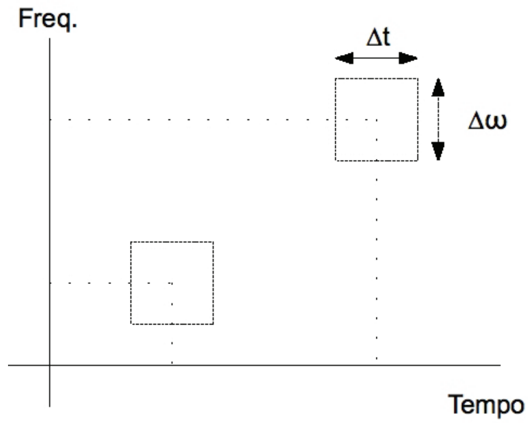

Fig. 3.1: Representação da relação tempo-freqüência na STFT.

Esta janela automaticamente muda seu tamanho para se adaptar às freqüências do sinal analisado [Chui, 1997], como representado nas figuras 3.2 e 3.3.

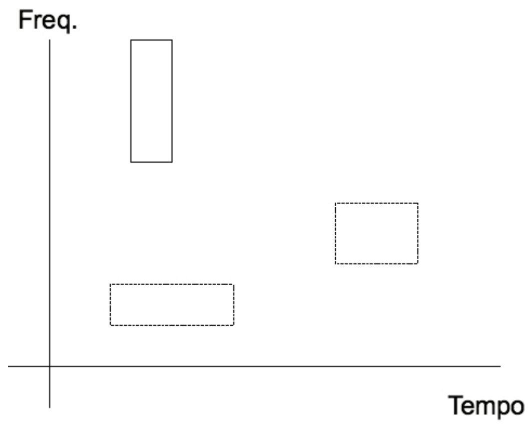

Fig. 3.2: Representação da relação tempo-freqüência na CWT.

Similaridades entre a Transformada de Fourier e a Transformada Wavelet: A transformada discreta de Fourier (DFT) estima a transformada de Fourier de uma função através de um número finito de pontos dela amostrados. As propriedades da DFT são basicamente as mesmas da transformada de Fourier. Ainda, a formula para a transformada inversa pode ser facilmente calculada através da transformada direta pelo fato da estreita semelhança entre as duas expressões. O cálculo da DFT é feito através do algoritmo de FFT [Oppenheim et al., 1999].

Tanto a transformada discreta de Fourier (DFT) quanto a transformada discreta *wavelet* são opera-

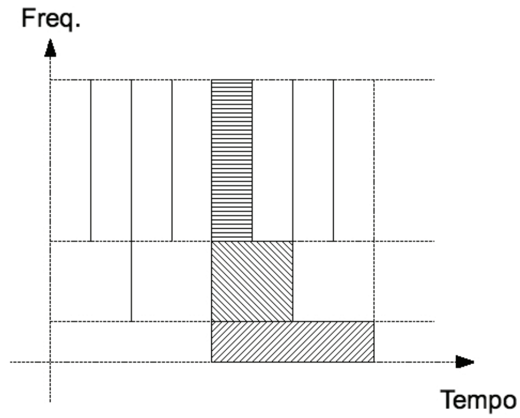

Fig. 3.3: Um outra forma de representação da CWT, com escalonamento diádico, ou seja, com a variando por potência de dois.

dores lineares que geram uma estrutura de dados contendo $log_2 n$ segmentos de vários comprimentos, usualmente de tamanhos 2^n.

As propriedades matemáticas das matrizes envolvidas, também são similares. A matriz da transformada inversa, tanto para a DFT quanto para a DWT é a transposta da original. Como resultado, ambas transformações podem ser vistas como a rotação de um espaço de função para um domínio diferente. Para a DFT este novo domínio contém funções bases que são senos e cossenos. Para a transformada wavelet, este novo domínio contém funções bases *wavelets* [Grapes, 1995]. Ambas transformadas são localizadas em freqüência, o que as tornam ótimas ferramentas para análises espectrais.

Diferenças entre a Transformada de Fourier e a Transformada Wavelet: A diferença mais notável entre esses dois tipos de transformadas é que as wavelets são funções localizadas no tempo, enquanto senos e cossenos não. Ou seja, na STFT uma única janela é utilizada para todas as freqüências, portanto, a resolução da análise é a mesma em todos os locais do plano tempo-freqüência.

Já na transformada *wavelet* as janelas são variáveis. Por exemplo, no caso de isolar uma descontinuidade de um sinal, funções base de suporte bem pequeno são ideais. Por outro lado, para se obter detalhes sobre a distribuição de freqüências mais baixas, funções bases com suporte mais amplo são desejáveis. A transformada *wavelet* oferece este recurso, possibilitando análises finas em ambos os domínios [Grapes, 1995].

3.2.4 Análise em Multi-resolução

Existem pelo menos suas maneiras de abordar a teoria *wavelets*: uma é através das transformadas contínuas, visto na seção 3.2 e outra através de *Análise em Multi-Resolução* . As análises em multi-resolução (MRA) possibilitam decompor um sinal $f(t)$, com $t \in \mathbf{R}$, em aproximações sucessivas de resolução cada vez menor, numa sequência de processos de filtragem consecutivos. Na MRA duas funções são utilizadas: a *wavelet* ψ e a função de escala ϕ, que são ortogonais entre si. A função *wavelet* é utilizada para gerar um filtro passa-altas que dá origem aos coeficientes de detalhe; a função de escala, com oscilações em baixas freqüências, é utilizada para criar um filtro passa-baixas responsável pelos coeficientes de aproximação. Tanto a *wavelet* quanto a função escala são filtros QMF e possibilitam a perfeita reconstrução do sinal [Miner & Caudell, 2002]. Nos próximos itens, apresenta-se a teoria da multiresolução para sinais contínuos, porém igualmente válida para sinais discretos.

3.2.5 A função escala e os subspaços V_n

A análise em multi-resolução de $L^2(\mathbf{R})$ é definida como uma seqüência de subespaços fechados V_n de $L^2(\mathbf{R})$, $n \in \mathbf{Z}$, com as seguintes *propriedades* [Jawerth & Sweldens, 1994, Chui, 1997]:

1. $... \subset V_{-1} \subset V_0 \subset V_1 \subset ...$

2. $f(t) \in V_n \Leftrightarrow f(2t) \in V_{n+1}$

3. $f(t) \in V_n \Leftrightarrow f(t + \frac{1}{2^n}) \in V_n$, para todo n inteiro.

4. o conjunto
   $\{\phi(t-k) : k = 0, \pm 1, ...\}$
   de translações inteiras de $\phi(t)$ é uma *Base de Riesz* de $V_0 = V_{\phi,0} := V_\phi$, e os subespaços fechados $V_n = V_{\phi,n}$ de L^2, satisfazem:

5. $clos_{L^2}(\bigcup_{n=-\infty}^{\infty} V_n) = L^2$

6. $\bigcup_{n=-\infty}^{\infty} V_n = \{0\}$

Vamos fazer algumas simples observações sobre as definições acima.

Seja $\phi(\frac{t}{2})$ expressa como translações inteiras de $\phi(t)$, por

$$\phi\left(\frac{t}{2}\right) = \sum_k p_k \phi(t-k). \tag{3.25}$$

para alguma seqüência $\{p_k\}$ de tal forma que a expressão do lado direito de 3.25 esteja em L^2, segue a definição [Chui, 1997, Mallat, 1989].

Definição 3.2.1 *Se uma função $\phi(t) \in L^2$ satisfaz 3.25 e é estável, então $\phi(t)$ por definição é chamada de função escala e $\{p_k\}$ em 3.25 sua correspondente seqüência bi-escalar.*

Se uma função $\phi(t)$ satisfaz a relação bi-escalar da definição 3.2.1, considerando as escalas a como potências inteiras de 2, é possível escrever a seguinte representação

$$f_n(t) = \sum_k c_{n,k} \phi(2^n t - k) \tag{3.26}$$

e é imediato que a coleção de funções $\{ \ \phi_{n,k} \ | \ k \in \mathbf{Z} \ \}$ forma uma base de Riesz (estável) do espaço V_n [Jawerth & Sweldens, 1994]. Ainda, (com $a = 2^{-n}$, onde n é um inteiro) para $n = n_1$ e $n = n_2$ onde $n_1 < n_2$, em 3.26 um sinal numa escala de menor resolução 2^{-n_1} pode ser representado através de uma escala de resolução mais fina 2^{-n_2}, o que leva a seqüência de espaços aninhados $... \subset V_{-1} \subset V_0 \subset V_1 \subset ...$, descrita acima. A função escala $phi(t)$ gera uma MRA do espaço de energia finita L^2.

Em muitas aplicações não é preciso utilizar a função escala em si, ao invés pode-se somente trabalhar diretamente com a seqüência $\{p_k\}$, descrito pelo teorema seguir [Chui, 1997].

Teorema 3.2.2 *Seja $\phi(t)$ uma função escala com a seqüência bi-escalar $\{p_k\}$ como na definição 3.2.1, e seja $f_n(t) \in V_n$, tal que $... \subset V_{-1} \subset V_0 \subset V_1 \subset ...$. Então $f_n(t) \in V_{n+1}$ e as seqüências de coeficientes*

$$c_n = \{n_{n,k}\} \qquad e \qquad c_{n+1} = \{n_{n+1,k}\} \tag{3.27}$$

do mesmo sinal

$$f_n(t) = \sum_k c_{n,k} \phi(2^n t - k) = \sum_k c_{n+1,k} \phi(2^{n+1} t - k) \tag{3.28}$$

em V_n e V_{n+1} respectivamente satisfazem

$$c_{n+1,k} = \sum_{\ell} p_{k-2\ell} c_{n,\ell} \tag{3.29}$$

Pelo teorema 3.2.2 é possível constatar que o cálculo de c_{n+1} através de c_n é muito eficiente, especialmente quando a seqüência $\{p_k\}$ é finita. Isto requer apenas duas operações, um *upsample* seguido por uma convolução.

3.2.6 A função *Wavelet* e o Espaço dos Detalhes W_n

Seja $\phi(t)$ uma função escala (passa-baixas) que gera uma MRA

$$\{0\} \leftarrow ... \subset V_{-1} \subset V_0 \subset V_1 \subset ... \rightarrow L^2 \tag{3.30}$$

do L^2. Para cada inteiro n, desde que V_{n-1} seja um subespaço próprio de V_n, temos um subespaço não trivial complementar W_{n-1} de V_n relativo à V_{n-1}. Ou seja, $W_{n-1} \subset V_n$ e

$$V_n = V_{n-1} + W_{n-1}, \qquad W_{n-1} \perp V_{n-1} \tag{3.31}$$

Usaremos a notação [Chui, 1997]

$$V_n = V_{n-1} \oplus W_{n-1} \tag{3.32}$$

onde o símbolo $\oplus$ denota soma direta. Em outras palavras, cada elemento de V_{n+1} pode ser escrito, de forma única, como a soma de um elemento de V_n e e um elemento de W_n. Entretanto, os subespaços W_n não são necessariamente únicos, existem várias formas de complementar V_n em V_{n+1} .

Assim, através de 3.30, o espaço L^2 pode ser decomposto como uma soma ortogonal dos subespaços W_n, de forma que

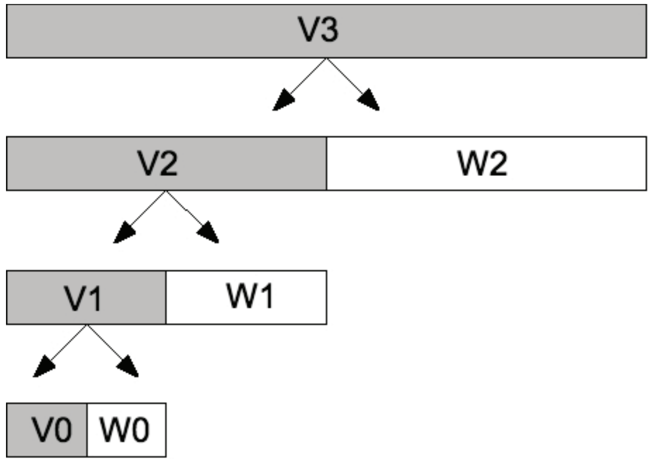

Fig. 3.4: Representação da soma direta dos espaços V_n e seus complementares W_n na decomposição por multiresolução de um sinal.

$$L^2 = \bigoplus_{n=-\infty}^{\infty} W_n. \tag{3.33}$$

Se uma função $\phi(t)$ gera uma base de Riesz, então é possível obter [Chui, 1997], a partir dela, outra função $\phi^{\perp}(t)$, tal que a família

$$\{\phi^{\perp}(t-k) : k = 0, \pm 1, ...\} \tag{3.34}$$

seja ortonormal.

Observe que se $\phi(t)$ gera uma MRA $\{V_n\}$ de L^2, então $\phi^{\perp}(t)$ também a gera. Seja $\{p_k\}$ sua seqüência bi-escalar, ou seja, $\phi^{\perp}(t)$ satisfaz a identidade

$$\phi^{\perp}(t) = \sum_k p_k \phi^{\perp}(2t-k). \tag{3.35}$$

A função de energia finita

$$\psi(t) := \sum_k (-1)^k p_{1-k} \phi^{\perp}(2t-k) \tag{3.36}$$

gera os subespaços complementares ortogonais W_n relativos a MRA descrito em 3.30, como aponta o seguinte teorema:

Teorema 3.2.3 *Seja $\phi(t)$ uma função escala que gera uma MRA $\{V_n\}$ de L^2, e seja $\phi^{\perp}(t)$ sua ortonormalização. Seja também, $\{p_k\}$ a seqêncua bi-escalar de $\phi^{\perp}(t)$ e $\psi(t)$ como definido em 3.36, através de $\phi^{\perp}(t)$ e $\{p_k\}$. Então $\psi(t)$ gera o subespaço ortogonal complementar W_n relativo a MRA $\{V_n\}$. Contudo,*

$$\psi_{j,k}(t) := 2^{\frac{j}{2}}\psi(2^j t - k) \tag{3.37}$$

é uma base ortonormal do espaço de energia finita L^2 [Chui, 1997].

Estes resultados demonstram o potencial da ferramenta no sentido que podemos representar qualquer sinal, de forma exata e sem aproximações, em sucessivas faixas de freqüências, com ótima representação, e com larguras de banda que diminuem por um fator diádico. Através da projeção do sinal nas funções escala ϕ obtemos os coeficientes de aproximação. E pela projeção do sinal na base $\psi_{j,k}$ obtemos os coeficientes referentes aos detalhes. O espectro é analisado conforme ilustrado na figura 3.5.

De acordo com o que foi dito no capítulo 2, exploramos aqui a parte matemática necessária para a compreensão da teoria envolvida no modelo que será exposto adiante. Nos próximo dois capítulos apontaremos como algumas das idéias já apontadas anteriormente foram expandidas durante o nosso estudo.

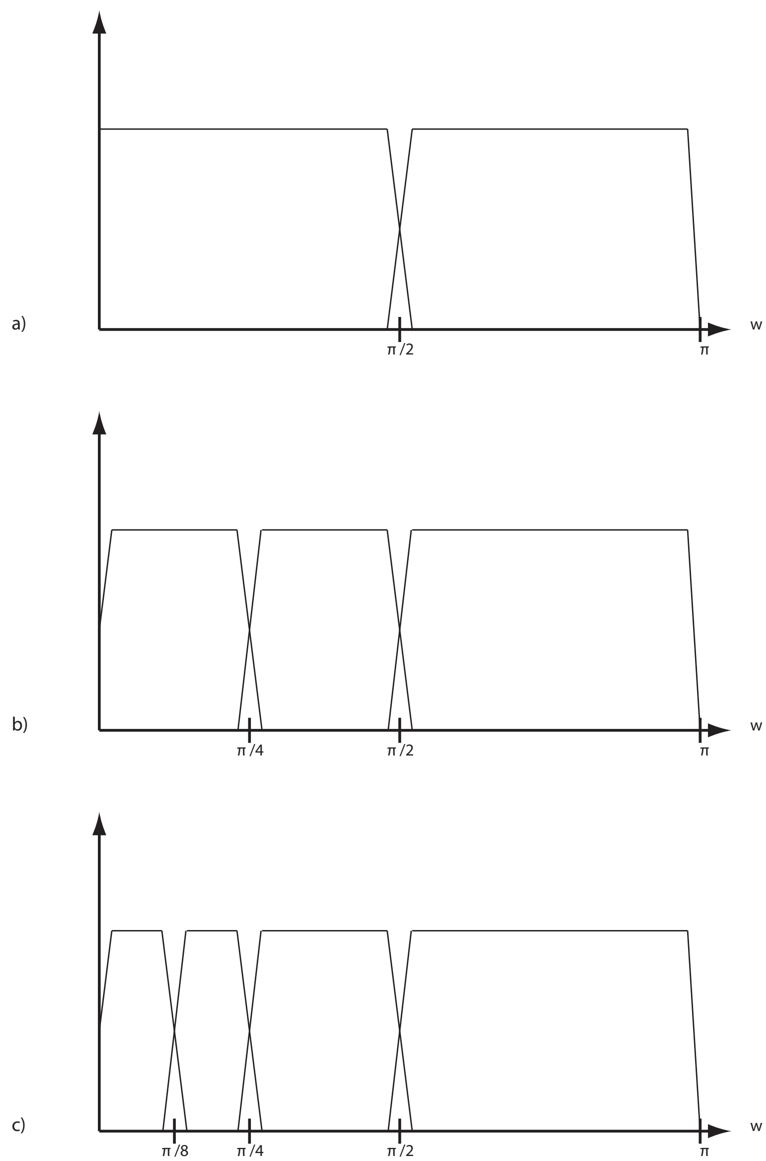

Fig. 3.5: Espectro de freqüências dividido em faixas diádicas, através da análise em multiresolução (MRA).

Capítulo 4

Descrição do Modelo Proposto

4.1 Introdução

Com o objetivo de expandir as aplicações em síntese sonora apresentadas no capítulo 2 e desenvolver um modelo computável a partir da sistemática apresentada no capítulo 3, apresentamos a seguir um modelo de manipulação de bases *wavelets* que possibilita a geração de novos timbres a partir da permutação de um mesmo conjunto de coeficientes.

As análises *wavelets* neste trabalho foram realizadas de forma investigativa dando origem a um novo mecanismo de representação de sinais musicais capaz de transitar por distintas sonoridades, através de um modelo simples. O foco está voltado à destacada importância que o posicionamento dos coeficientes *wavelets* exerce na caracterização timbrística de sons musicais.

Mais especificamente, o modelo se baseia nas análises em multiresolução pela transformada discreta *wavelets*. A representação de sonoridades distintas é dada a partir de reordenações que são feitas em um mesmo grupo de coeficientes, extraídos de um som base ou através de uma aproximação por função polinomial. Ou seja, estamos relacionando o posicionamento dos coeficientes com geração de timbres diversos. Vamos à descrição técnica do modelo.

4.1.1 Análise em Multiresolução das Amostras

Tipicamente instrumentos musicais têm as freqüências fundamentais de suas notas contidas na faixa entre 27.5 Hz e 4186 Hz [1]. Uma representação eficiente de instrumentos com esta tessitura pelas análises em multiresolução requer uma implementação que cubra todo o espectro de freqüências, e ainda que seja capaz de isolar em faixas espectrais distintas o maior número de parciais do sinal. Esta necessidade interfere diretamente no número de níveis no qual o sinal deve ser decomposto. Como podemos notar na figura 4.1, dois sons ilustrativos com fundamental em 50 Hz e outro com fundamental em 80 Hz estão divididos em cinco níveis de multiresolução e, na figura 4.2, em oito níveis. Na primeira ilustração, vemos que, para o som com fundamental em 50 Hz, na faixa de freqüência do nível cinco de aproximação estão contidos seus quinze primeiros harmônicos. E para o som de 80 Hz, os oito primeiros harmônicos. Desta forma, a localização do conteúdo espectral do sinal fica prejudicada, pois várias parciais estão aninhadas numa mesma faixa de freqüência.

Já na figura 4.2, onde são adotados oito níveis, os harmônicos de ambos os sons ficam melhor localizados. No oitavo nível de detalhes temos somente as fundamentais, 50 Hz e 80 Hz, assim como nos níveis seguintes uma distribuição mais segregada das parciais.

Desta forma, escolhemos fazer a decomposição em oito níveis. A faixa de freqüência do oitavo nível de aproximação, o mais grave, termina em 93,75 hertz, suficiente para localizamos com eficiência o espectro de notas de baixa freqüência. Nos experimentos e análises que se seguirão, para cada amostra, serão gerados ao todo nove vetores de coeficientes, sendo oito relativos aos níveis de detalhes e um relativo ao oitavo nível de aproximação.

As *wavelets* escolhidas foram as da família Daubechies, com 16 momentos nulos para $\psi(t)$. Esta família é ortogonal, de suporte compacto, com *wavelets* não simétricas e maior número de momentos nulos para um dado suporte. Os filtros associados são de fase mínima [Daubehies, 1992], figura 4.3.

O tamanho dos vetores depende da taxa de amostragem e da duração do arquivo de áudio utilizados. Para cada segundo de áudio, com taxa de amostragem de 48 kHz, cada vetor do primeiro nível de detalhes possui aproximadamente 24 mil coeficientes, o segundo nível aproximadamente 12 mil coeficientes e assim por diante, decaindo por um fator diádico i.e, se S é o número total de amostras do arquivo de áudio que será analisado, o número $N^{(n)}$ de coeficientes de cada nível n é dado por:

[1] Esta é a faixa das freqüências fundamentais das notas do piano, com afinação igual temperada

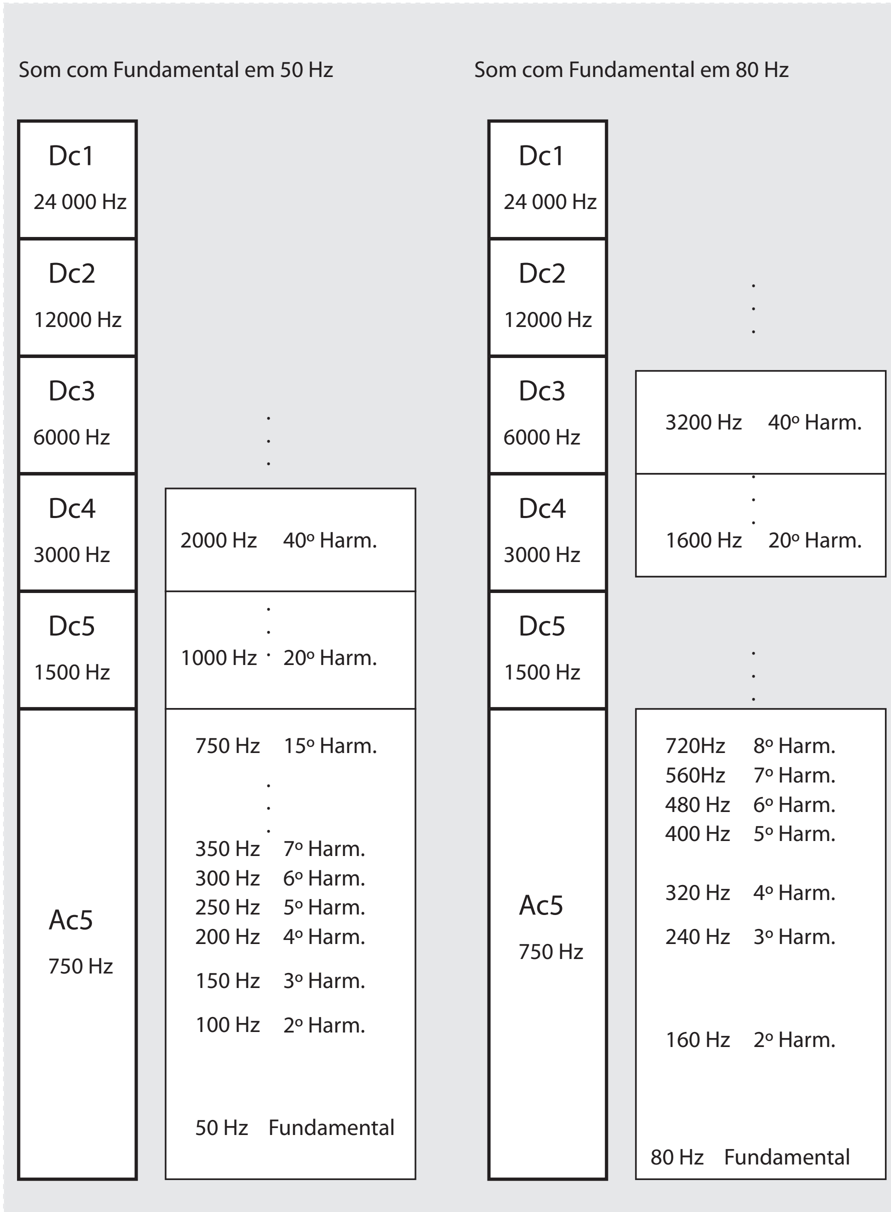

Fig. 4.1: Exemplo de análise em multiresolução com cinco níveis e suas respectivas bandas de freqüências.

$$N^{(n)} = \frac{S}{2^n} \tag{4.1}$$

Seja x um sinal de áudio, representado através de suas amostras temporais. Como explicado anteriormente no capítulo 3, os níveis de detalhes são obtidos através da projeção do sinal na base que gera o subespaço W_n. Igualmente para os níveis de aproximação, no subespaço V_n.

Desta forma, $w_x^{(n)}$ é o vetor dos coeficientes da projeção do sinal x na base *wavelet* $\psi_{j,k}(t)$ que

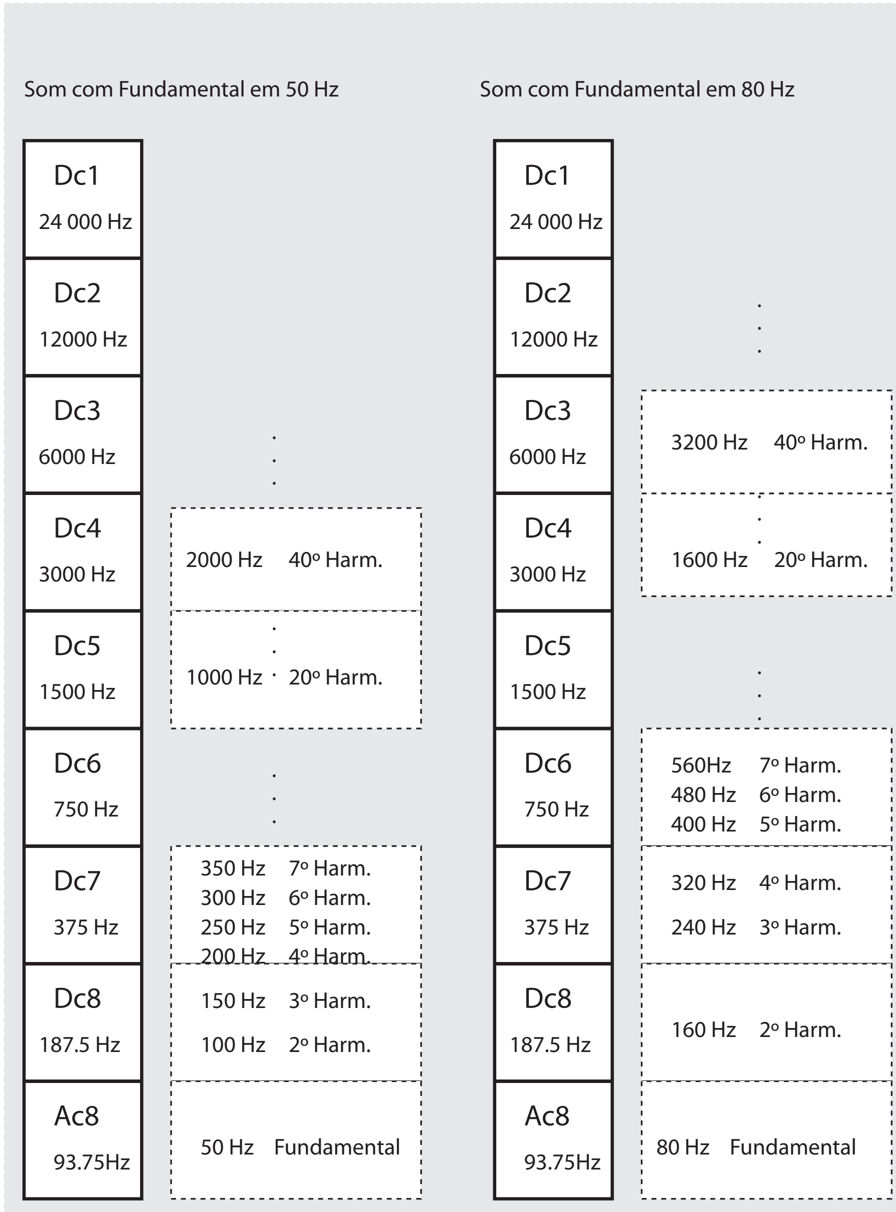

Fig. 4.2: Exemplo de análise em multiresolução com oito níveis e suas respectivas bandas de freqüências .

gera W_n. Analogamente para o caso dos níveis de aproximação onde $v_x^{(n)}$ é o vetor dos coeficientes da projeção do sinal x na base ϕ, que gera a análise em multiresolução em V_n.

De modo geral, neste trabalho o que se busca é modelar o vetor $w_x^{(n)}$ resultando em um outro vetor aproximado $w_s^{(n)}$ obtido de duas formas diferentes, como descrito a seguir.

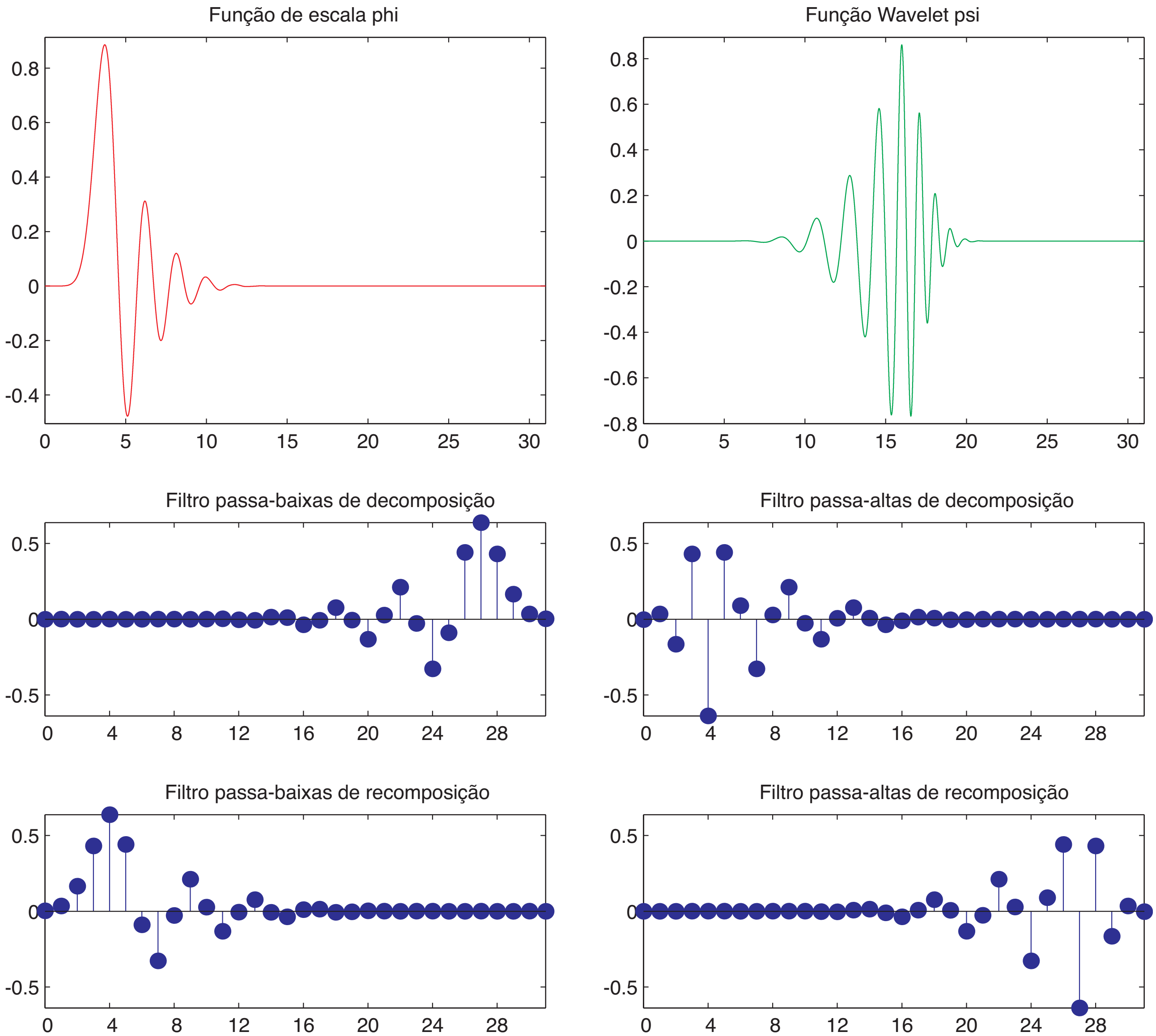

Fig. 4.3: *Wavelet* de Doubechie 16 e seus filtros de decomposição e recomposição.

4.2 Manipulação dos Coeficientes

O modelo proposto está baseado nas permutações dos coeficientes *wavelets* de cada nível n de resolução. Estas operações são realizadas a partir da seguinte definição.

Definição 4.2.1 *Dado* $w^{(n)} = (w_1^{(n)}, w_2^{(n)}, ..., w_N^{(n)})$*, de acordo com o apêndice A existe uma matriz de permutação* $P^{(n)}$ *tal que:*

$$P^{(n)} w^{(n)} = u^{(n)} \tag{4.2}$$

onde $u^{(n)} = (u_1^{(n)}, u_2^{(n)}, ..., u_N^{(n)})$ *é ordenado, isto é,* $u_1^{(n)} \leq u_2^{(n)} \leq ... \leq u_N^{(n)}$.

Então, dado o vetor $w_x^{(n)}$ de coeficientes *wavelets* de detalhes do nível n, do sinal amostrado de áudio x, existe uma matriz que ordena este vetor. Seguindo a definição acima

$$P_x^{(n)} w_x^{(n)} = u_x^{(n)}. \tag{4.3}$$

Nosso objetivo agora é usar esta matriz para manipular um outro vetor $u_y^{(n)}$ de coeficientes ordenados de base, provenientes de uma aproximação polinomial ou de um outro sinal.

Isto significa que estamos transferindo a ordenação do vetor $u_x^{(n)}$, proveniente da $P_x^{(n)}$, para um outro vetor $u_y^{(n)}$ no mesmo nível de resolução n *wavelets*. Denominamos o resultado desta operação

$$(P_x^{(n)})^T u_y^{(n)} = w_s^{(n)} \tag{4.4}$$

onde $w_s^{(n)}$ é o sinal de saída relativo ao nível n de detalhes.

4.2.1 Notação Utilizada

Na tabela 4.1 dispomos, de forma a facilitar a consulta, a notação empregada para as variáveis utilizadas no modelo, com suas descrições.

Tab. 4.1: Varíaveis utilizadas na descrição do modelo.

Notação	Descrição da Variável
x	Sinal de entrada
$w_x^{(n)}$	Vetor com os coeficientes de detalhes do nível n de x
$u_x^{(n)}$	Vetor com os coeficientes de detalhes ordenados do nível n de x
$w_s^{(n)}$	Vetor de saída do modelo.
$P_x^{(n)}$	Matriz de permutação associada ao nível n de detalhes do sinal x

4.3 Visão Geral do Modelo

Consideramos inicialmente um sinal musical x, representado por seus coeficientes *wavelets* em oito níveis de detalhe, conforme discutimos anteriormente.

Seja $w_x^{(n)}$ o vetor de coeficientes do n-ésimo nível de detalhes do sinal musical x. Consideremos a reordenação de seus coeficientes, de modo a produzir um novo vetor $u_x^{(n)}$ ordenado como exposto na definição 4.2.1.

Desse modo, cada nível de representação do sinal x pode ser decomposto em dois aspectos: de um lado, a informação relativa à permutação $P_x^{(n)}$. Do outro lado, a função monotônica $f^{(n)} \sim u_x^{(n)}$.

$$f^{(n)} : N \to R \quad f^{(n)}(k) \sim u_x^{(n)}(k) \tag{4.5}$$

para $k = 1, 2, ...N$.

O modelo proposto neste trabalho é baseado na observação de que o primeiro aspecto descrito acima é fundamental na representação do sinal musical, ao passo que o segundo aspecto pode ser aproximado de diversas maneiras. Em outras palavras, um sinal musical pode ser modelado através das permutações associadas a cada nível de representação cujos coeficientes são aproximados por $f^{(n)}$. Tais aproximações podem ser obtidas por maneiras diversas. Analogamente, para o caso dos coeficientes de aproximação por uma função $g^{(n)}$. Nas próximas seções, duas alternativas de aproximação serão propostas.

4.4 Ordenação dos Coeficientes

A ordenação dos coeficientes é a parte mais importante do processo. Aplicada nível a nível, ela cumpre dois objetivos principais. O primeiro é o de simplificar a representação. Somente assim, podemos aproximar os valores dos coeficientes de base por uma função $f^{(n)}$ simples.

Com os coeficientes na ordenação original, essas curvas são muito complexas, pois são a projeção do sinal na base *wavelet* referente á faixa de freqüência em que se situa aquele nível, de complexidade muito semelhante a própria forma de onda da região de freqüência em questão. O segundo objetivo é obter a matriz de permutação $P_x^{(n)}$.

Na prática, a operação de ordenação é feita através das matrizes de transposição definidas no apêndice A e de forma que as posições originais são guardadas em um vetor de posição. Assim, é possível, posteriormente, efetuar de forma rápida e efetiva a reordenação dos coeficientes, sem que seja necessário trabalhar diretamente com a matriz de permutação P de alta dimensão. Desta forma, contornamos o problema de esforço computacional imposto pela alta dimensionalidade das matrizes de permutação.

4.5 Formas de Aproximação dos Coeficientes de Base

Nesta seção, iremos descrever as duas formas simples utilizadas para obtenção dos coeficientes que serão utilizados como elementos-base do processo de representação musical do sinal alvo. Primeiramente, descreveremos a aproximação por meio de uma função polinomial cujo interesse maior reside em recodificar um sinal musical de forma econômica. Posteriormente, um modelo baseado em sons amostrados será apresentado. Seu destaque está na possibilidade de transitar por diversos sons, conforme os sons alvo e base escolhidos. Este segundo modelo visa prover aos músicos, tanto da área de performance quanto compositores, novas possibilidades de manipulação sonora, não primando pela redução informacional e sim por viabilizar a criação de novos timbres e texturas sonoras. Ainda, ambos os modelos podem ser refinados ou transformados através do dispositivo residual, da seção 4.7, no qual características de um terceiro som podem ser implantadas.

4.5.1 Modelo por Aproximação Polinomial

Neste modelo, os coeficientes do vetor $u_x^{(n)}$ são aproximados por uma função polinomial e utilizados como base. Ou seja, $u_x^{(n)}(k) \sim f^n(k) = u_y^{(n)}(k)$. As aproximações são realizadas por meio de uma função polinomial de grau *p*, refinada pelo método dos mínimos quadrados. Ou seja

$$u_x^{(n)}(k) \sim f^n(k) = \alpha_p k^p + \alpha_{p-1} k^{p-1} + \alpha_{p-2} k^{p-2} + \ldots + \alpha_0 \tag{4.6}$$

para $k = 1, 2, ...N$.

O sistema de equações 4.6 pode ser reescrito, na forma matricial, por:

$$u_x^{(n)} = K\theta^{(n)} \tag{4.7}$$

$$u_x^{(n)} = \begin{pmatrix} u_x^{(n)}(1) \\ u_x^{(n)}(2) \\ u_x^{(n)}(3) \\ \vdots \\ u_x^{(n)}(N) \end{pmatrix} \quad K = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 2^p & 2^{p-1} & 2^{p-2} & \dots & 1 \\ 3^p & 3^{p-1} & 3^{p-2} & \dots & 1 \\ \vdots & \vdots & \vdots & & \\ N^p & N^{p-1} & N^{p-2} & \dots & 1 \end{pmatrix} \quad \theta^{(n)} = \begin{pmatrix} \alpha_p \\ \alpha_{p-1} \\ \alpha_{p-2} \\ \vdots \\ \alpha_0 \end{pmatrix}, \tag{4.8}$$

Os coeficientes $\theta^{(n)}$ podem ser obtidos, através do método dos mínimos quadrados, resultando em:

$$\theta^{(n)} = (K^T K)^{-1} K^T u_x^{(n)}. \tag{4.9}$$

A matriz quadrada $(K^T K)^{-1} K^T$ é chamada de *pseudo inversa* de K. Desta forma, obtemos o vetor coluna $\theta^{(n)}$ que contém os coeficientes do polinômio. Para obter esta aproximação, utilizamos somente o som alvo. A escolha do grau do polinômio depende do sinal alvo. Como veremos no capítulo 5, sons com um comportamento espectral mais regular, como a clarineta, podem ser aproximados de forma satisfatória por um polinômio de grau 5. Já para o piano, cujos transientes são acentuados e seu espectro mais complexo, precisamos aumentar o grau do polinômio para 9.

A escolha do grau do polinômio pode ser feita de várias maneiras. Uma forma eficiente é estabelecer uma medida de distância entre as curvas, que esteja atrelada a um certo padrão de qualidade psicoacústica, para cada tipo de instrumento. Assim, torna-se possível adequar a relação entre a qualidade sonora e a economia de dados, obtida pela representação, conforme o objetivo sonoro final.

Desta forma, a saída do modelo por coeficientes amostrados pode ser escrita como

$$(P_x^{(n)})^T u_y^{(n)} = w_s^{(n)} \tag{4.10}$$

onde $u_y^{(n)}(k) = f^n(k)$, que é a função de aproximação polinomial como discutido acima.

4.6 Modelo a Partir de Coeficientes de Sons Amostrados

Neste modelo, os valores dos coeficientes *wavelets* base $u_y^{(n)}$ são de um segundo sinal de áudio y. A sonoridade obtida é caracterizada pela matriz de permutação P_x^n. Assim, saída do modelo por coeficientes amostrados pode ser escrita como

$$(P_x^{(n)})^T u_y^{(n)} = w_s^{(n)} \tag{4.11}$$

onde $u_y^{(n)}$ são os coeficientes *wavelets* do nível n do sinal amostrado y.

Ajuste de Energia

As permutações propõem reordenações que alteram as características timbrísticas presentes em cada nível. Porém, as energias originais das faixas de freqüência envolvidas são conservadas. Na constituição da sonoridade global, os valores dos ganhos dessas bandas são cruciais. Ou seja, após permutarmos os coeficientes, devemos corrigir as energias de cada nível, de forma a re-equilibrar as contribuições timbrísticas de cada uma delas para o sinal de saída.

O ajuste é feito através de um fator de correção, calculado através da relação entre as energias relativas a cada nível, calculadas através de:

$$E_x^n = \sum_{k=1}^{N} (w_x^{(n)})^2 \tag{4.12}$$

para os níveis do sinal x e:

$$E_y^n = \sum_{k=1}^{N} (w_y^{(n)})^2 \tag{4.13}$$

para os níveis do sinal y.

Assim, o coeficiente de energia G^n é dado por:

$$G^n = \sqrt{\frac{E_x^n}{E_y^n}} \tag{4.14}$$

Os coeficientes G^n são aplicados a cada um dos vetores resultantes da reordenação dos coeficientes base $w_s^{(n)}$ e funcionam como um fator de ganho. Essas correções são aplicadas somente no modelo por sons amostrados. As duas próximas seções apresentadas são aplicadas em ambos os modelos de aproximação dos coeficientes. Elas estão relacionadas à reconstrução do sinal.

4.7 Etapa Residual

Após a construção do sinal através das permutações, podemos manipular diferenças entre o sinal original e o sinal representado através de um modelo residual. Esta etapa consiste em detectar quais coeficientes aproximados são mais distintos, com relação ao sinal alvo. Estes coeficientes, que denominamos de resíduos, são os possíveis responsáveis pelas diferenças psicoacústicas entre o sinal alvo e o sinal gerado pelo modelo. Caso o objetivo seja reproduzir o sinal alvo, podemos substituir esses coeficientes detectados pelos seus correspondentes do som alvo.

Mas, extrapolando esta aplicação, podemos utilizar estas trocas na criação de novos timbres. Através da inserção de coeficientes de outros sons, como veremos no capítulo 5, podemos atribuir novas características timbrísticas à sonoridade alvo, expandindo as possibilidades de geração de material musical.

Existem várias maneiras de detectar semelhanças ou diferenças entre os sinais gerado e original. Uma delas se dá através do cálculo da diferença em módulo entre o vetor C_s, gerado a partir da concatenação dos níveis do sinal de saída e o vetor C_x, gerado a partir da concatenação dos níveis do sinal de entrada.

Efetuando-se ordenação *decrescente* [2] desta diferença através da matriz de permutação Q, obtemos as posições dos R coeficientes mais distintos. Desta forma, podemos substituí-los e então reordená-los através de Q, nas posições originais. Assim, estamos substituindo os R coeficientes mais distintos, entre o som alvo e o permutado, por outros coeficientes que podem ser pegos do som alvo ou de um terceiro sinal.

Esta operação não é mais realizada nível a nível, pois estamos extraindo as informações da concatenação dos níveis. Assim as diferenças são tratadas globalmente, e não mais de forma estratificada como no caso das permutações. Estas trocas nos possibilitam:

[2] A ênfase aqui é dada, pois, nos processos anteriores, utilizamos a ordenação crescente.

- determinar os R possíveis coeficientes responsáveis pela diferenciação entre as duas sonoridades.

- uma exploração sonora entre as principais diferenças timbrísticas dos dois sons envolvidos. Através da recriação dos arquivos de áudio correspondentes, podemos escutar essas diferenças de sonoridade tanto dos coeficientes que foram retirados, como dos coeficientes que foram inseridos.

- utilizar N coeficientes extraídos de um terceiro sinal ao invés do alvo, com o objetivo de inserir novas características ao sinal permutado.

O último item realça a idéia mais atrativa do ponto de vista musical, i.e, interpretar as trocas como um parâmetro de controle de sonoridade. O número R de coeficientes que serão substituídos fica a cargo da criatividade musical do usuário. A sua variação altera as características psicoacústicas, criando novas texturas e sensações sonoras.

Após as trocas, os coeficientes concatenados são finalmente colocados em suas posições originais e submetidos à transformada *wavelet* inversa. Esta é a saída de áudio que será gravada em um arquivo *wave*.

O modelo baseado na aproximação por função polinomial está ilustrado pelo diagrama da figura 4.4. Já o modelo por sons amostrados está representado pelo diagrama da figura 4.5.

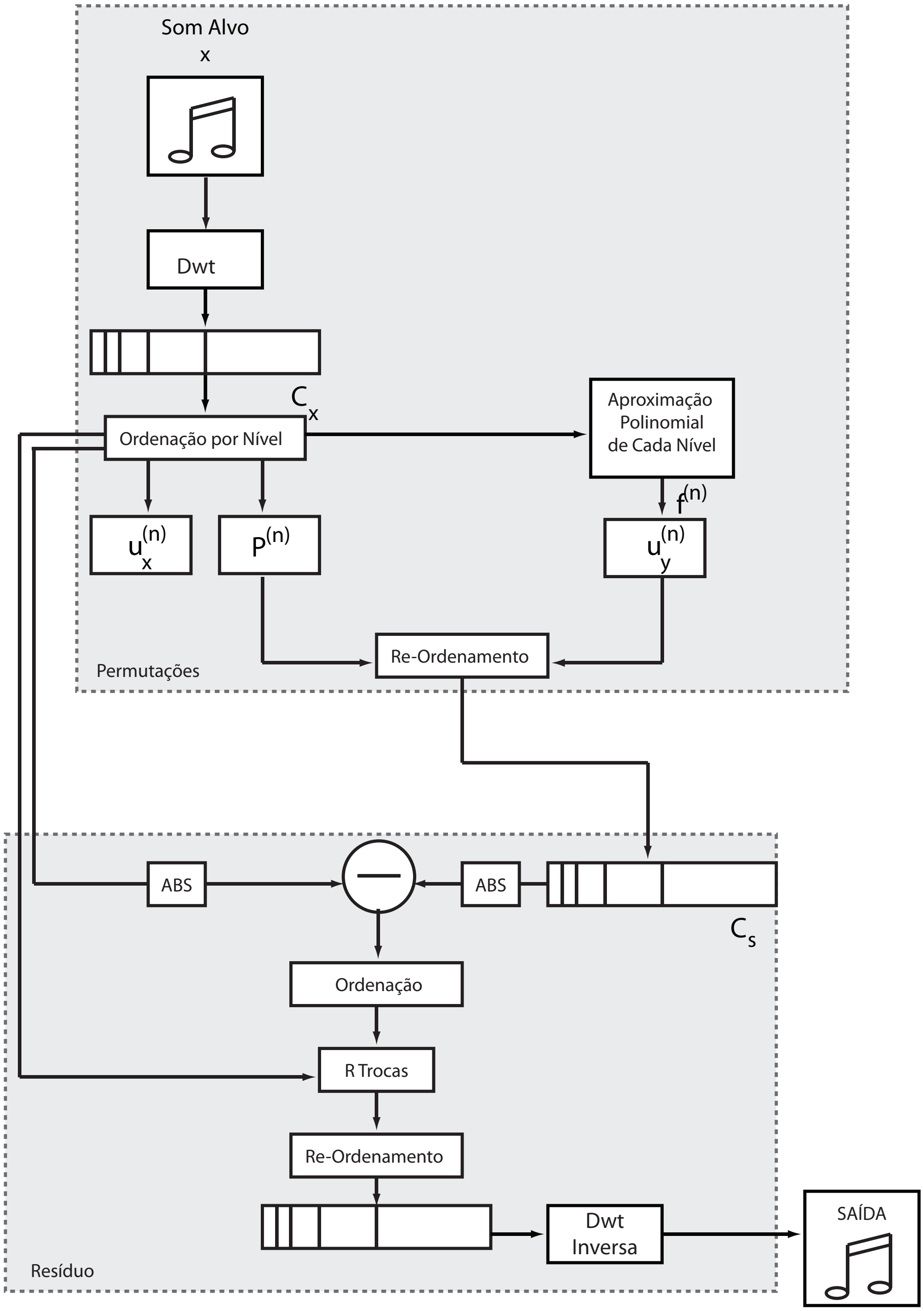

Fig. 4.4: Diagrama completo do modelo polinomial.

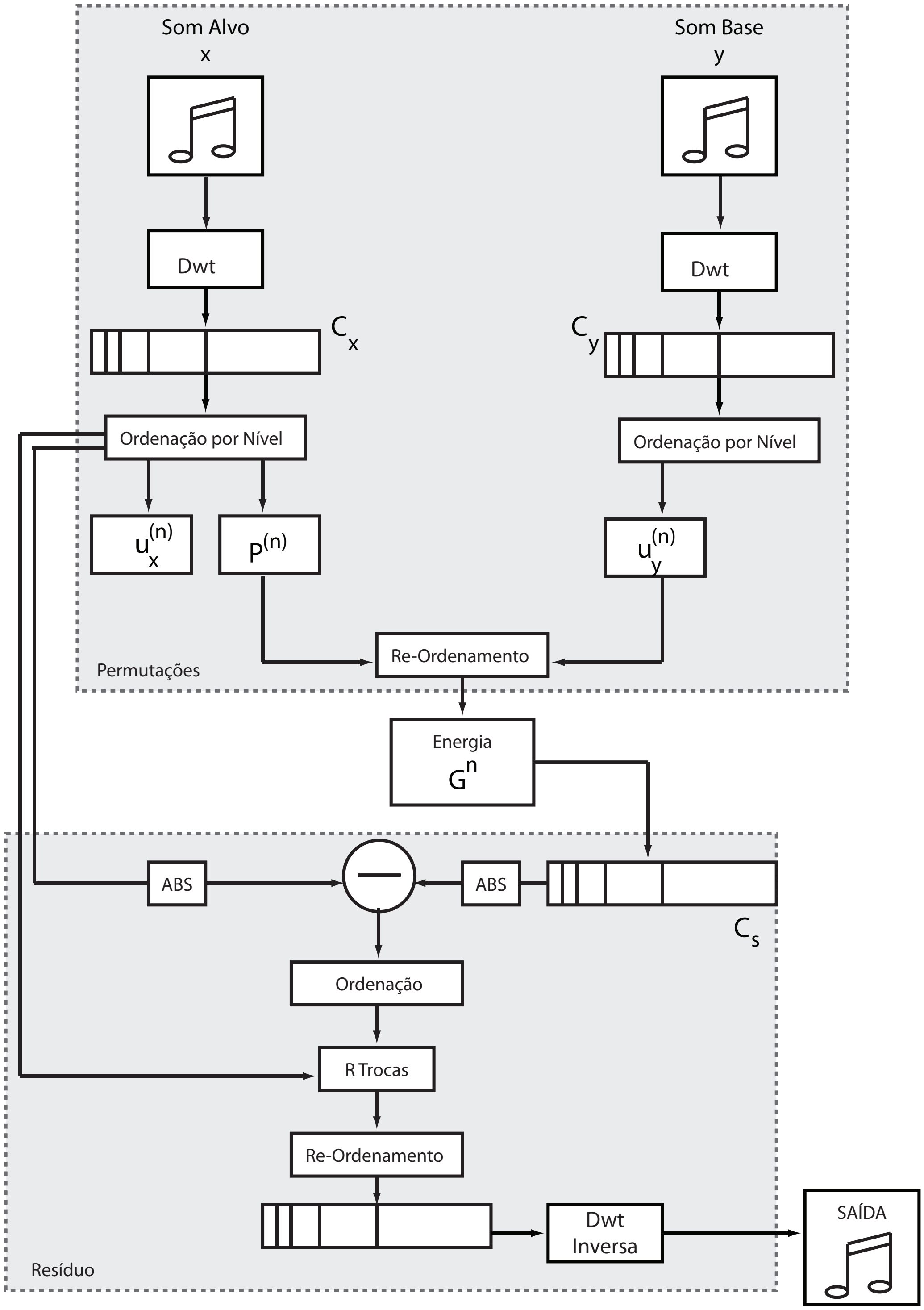

Fig. 4.5: Diagrama completo do modelo por sons amostrados.

Capítulo 5

Experimentos

5.1 Introdução

Neste capítulo, apresentaremos alguns dos resultados obtidos a partir do modelo proposto no capítulo 4. Através dos experimentos que se seguem, avaliamos o potencial do método discutido neste trabalho, destacando algumas de suas diversas formas de aplicação. Seja através da criação e representação sonora pela fusão de materiais diversos, ou pela recodificação econômica de sinais musicais. Neste sentido, os sinais utilizados foram escolhidos visando abranger algumas das principais características presentes e desejáveis no cenário musical.

Dentre os sinais escolhidos estão: (1) a clarineta que possui um comportamento espectral bem determinado, com predominância dos harmônicos ímpares e com transientes suaves; (2) o piano, instrumento predominantemente harmônico, com transientes mais acentuados que os da clarineta, com caráter levemente percussivo; (3) um trecho rítmico de *drum'n bass* executado por uma bateria eletrônica *Roland TR8080*; e (4) um baixo sintético, com mesmo estilo e andamento da bateria. As amostras foram extraídas dos pacotes de sons do aplicativo *Logic Audio* da *Apple*, devido à alta qualidade das gravações e dos instrumentos empregados (piano *Steinway* por exemplo). Foram utilizadas com taxa de amostragem de 48 KHz e 16 bits.

Os experimentos se iniciam pelo modelo polinomial, explorando o potencial ecônomico da representação. Já nas seções seguintes, utilizamos os mesmos sons como material para o modelo por coeficientes amostrados. Nesta parte, os coeficientes da clarienta foram utilizados como base na representação do piano e da bateria eletrônica. Os coeficientes empregados na parte residual foram extraídos da amostra do baixo sintético.

Os resultados do trabalho são apresentados de duas formas. Primeiro, através dos arquivos de áudio gerados, que estão no CD que acompanha a dissertação. Segundo, pelo próprio texto, por meio dos espectros dos arquivos que constam no CD e de tabelas que contêm os valores das distâncias euclidianas entre as curvas produzidas. Todos os experimentos são apresentados através desta mesma estrutura. Assim, o leitor pode navegar pelos resultados de forma prática e rápida.

5.2 Representação Através do Modelo por Aproximação Polinomial

Vamos começar a apresentação dos resultados pelos experimentos realizados através do modelo por aproximação polinomial. A primeira subseção traz a representação do som de clarineta. Posteriormente, exploramos o modelo através de sons de piano. E, por fim, terminamos esta seção com os sons de bateria.

5.2.1 Representação de Som Clarineta

A aproximação dos valores de seus coeficientes foi feita por um polinômio de grau 5. A figuras 5.1 e 5.2 mostram as curvas dos níveis 4 e 7, respectivamente.[1]

Pela inspeção dos gráficos, é possível constatar que a função aproxima de forma satisfatória os valores dos coeficientes. Uma segunda forma de avaliar a qualidade sonora é por meio dos valores das distâncias euclidianas entre.

Para o cálculo das distâncias as curvas foram normalizadas. Isto é, todos os coeficientes foram divididos pelo máximo elemento em módulo dentre eles, o que não altera a relação entre as curvas. Desta forma, tivemos uma variação de valores que ficou entre -1 a 1 e pudemos comparar as distâncias, para todos os níveis.

[1] Por questões de diagramação e organização do texto, colocamos apenas as curvas referentes a alguns níveis para os experimentos. As demais curvas podem ser encontradas no CD que acompanha o texto.

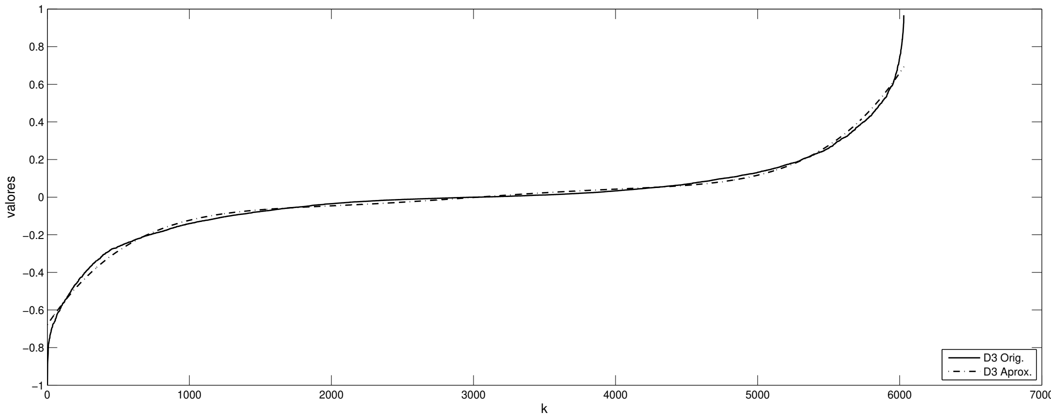

Fig. 5.1: Curva proveniente da ordenação ascendente do nível de detalhes 4 da nota D3 da clarineta e sua aproximação por um polinômio de grau 5.

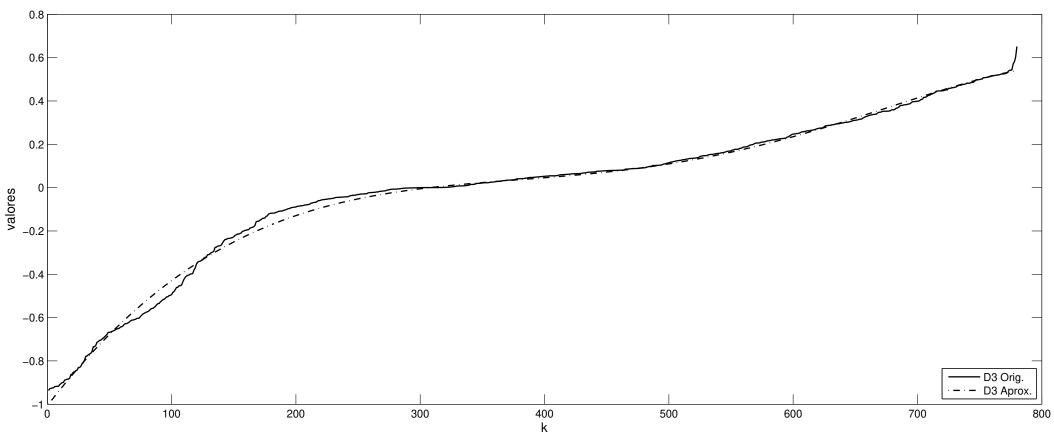

Fig. 5.2: Curva proveniente da ordenação ascendente do nível de detalhes 7 da nota D3 da clarineta e sua aproximação por um polinômio de grau 5.

A métrica foi calculada da seguinte forma. Dados um par de vetores ordenados $wo_n^x = (wo_{n,1}^x, wo_{n,2}^x, ..., wo_{n,L_n}^x)$ e $wo_n^y = wo_{n,1}^y, wo_{n,2}^y, ..., wo_{n,Ln}^y$, a distância é dada por :

$$Dist_n = \sqrt{\frac{\sum_{k=1}^{L_n}(wo_{n,k}^x - wo_{n,k}^y)^2}{L_n}}. \quad (5.1)$$

Na tabela 5.1 estão os valores das distâncias para todos os níveis para o D3 da clarineta.

Para visualizar o espectro, utilizamos a representação em cascata da transformada enjanelada de Fourier. Nas figuras 5.3 e 5.4 estão os espectros do sinal original e final do D3, representado pelo

Vetores	*Comprimento*	Freq.	Dist
$Dc1$	48015	24000	0.0105
$Dc2$	24023	12000	0.0256
$Dc3$	12027	6000	0.019
$Dc4$	6029	3000	0.0205
$Dc5$	3030	1500	0.0252
$Dc6$	1530	750	0.0192
$Dc7$	780	375	0.0215
$Dc8$	405	187.5	0.0333
$Ac8$	405	93.75	0.0763

Tab. 5.1: Distâncias euclidianas entre as curvas ordenadas dos níveis originais e aproximados, para o exemplo da nota alvo D3 da clarineta.

modelo.

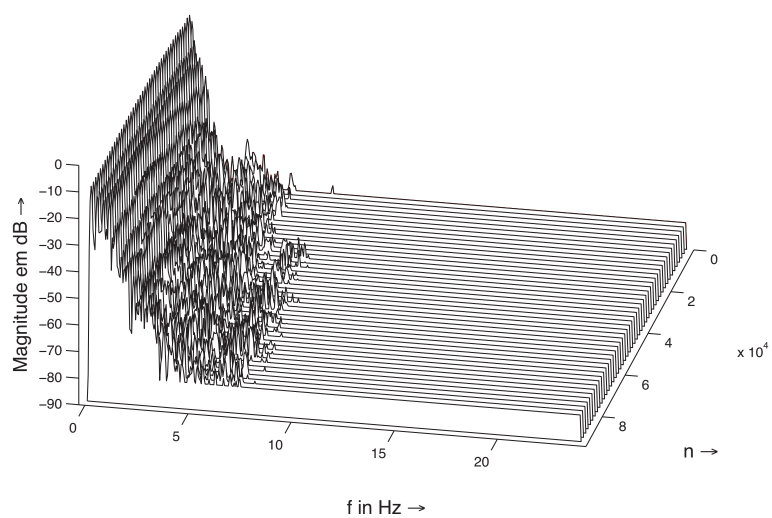

Fig. 5.3: Representação em cascata do espectro da nota D3 original da clarineta.

As diferenças entre os espectros são praticamente imperceptíveis, o que pode ser constatado através da audição das amostras que estão no CD[2]

[2]Uma lista com a disposição das faixas do CD encontra-se no apêndice B, com um texto explicativo, referente às amostras, figuras e tabelas.

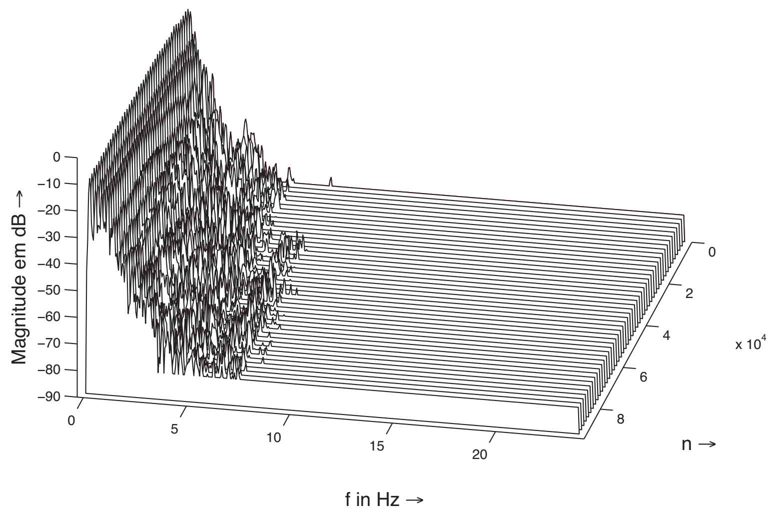

Fig. 5.4: Representação em cascata do espectro da nota D3 da clarineta gerada pelo modelo por aproximação polinomial.

5.2.2 Representação de Som de Piano

A complexidade das envoltórias espectrais deste instrumento afeta diretamente as formas das curvas ordenadas de seus coeficientes *wavelets*. Para aproximá-las, utilizamos uma função polinomial de grau 9. A nota escolhida foi um C4, e as curvas dos níveis 5 e 6 estão nas figuras 5.5 e 5.6.

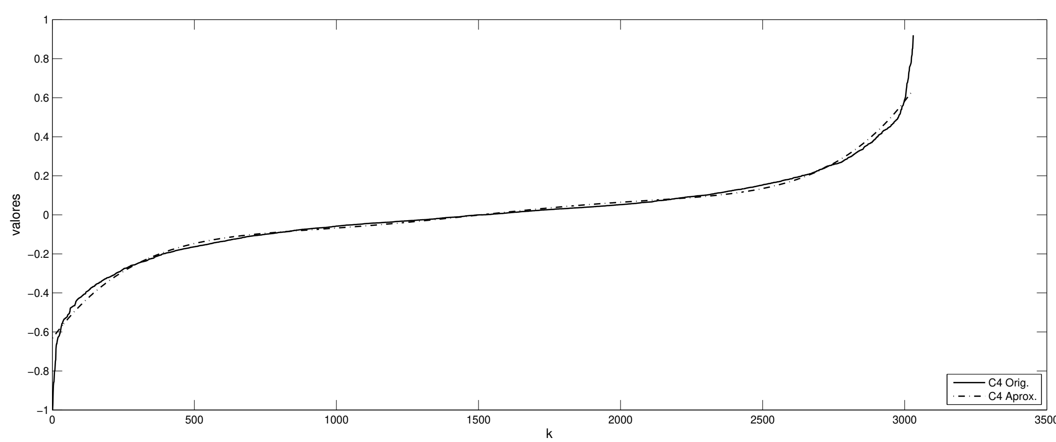

Fig. 5.5: Curva proveniente da ordenação ascendente do nível de detalhes 5 da nota C4 do piano e de sua aproximação por um polinômio de grau 9.

Pela inspeção dos gráficos, é possível constatar que a função aproxima de forma satisfatória os

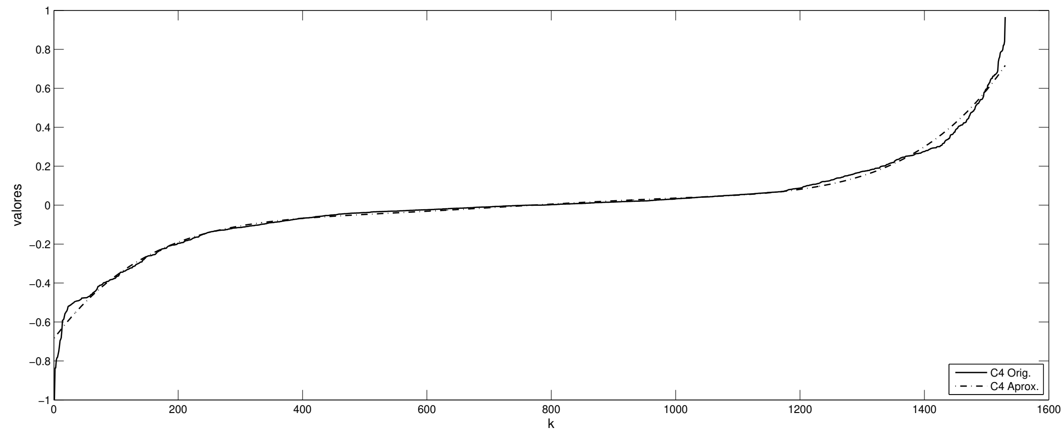

Fig. 5.6: Curva proveniente da ordenação ascendente do nível de detalhes 6 da nota C4 do piano e de sua aproximação por um polinômio de grau 9.

valores dos coeficientes do piano. A tabela 5.2 contém os valores das distâncias entre as curvas ordenadas para todos os níveis do C4.

Vetores	*Comprimento*	Freq.	Dist
$Dc1$	48015	24000	0.0106
$Dc2$	24023	12000	0.0181
$Dc3$	12027	6000	0.016
$Dc4$	6029	3000	0.0155
$Dc5$	3030	1500	0.0148
$Dc6$	1530	750	0.0145
$Dc7$	780	375	0.0122
$Dc8$	405	187.5	0.0259
$Ac8$	405	93.75	0.0297

Tab. 5.2: Distâncias euclidianas entre as curvas ordenadas dos níveis originais e aproximados, para o exemplo da nota alvo C4 do piano.

A representação da sonoridade do piano é mais sensível, com relação aos valores dos coeficientes, do que a da clarineta. Comparando as tabelas 5.1 e 5.2 constatamos que, no geral, as distâncias estão menores para o piano. Porém, podemos ouvir algumas granulações a mais na representação do seu sinal, quando comparado ao da clarineta. Estas granulações são visíveis através da visualização do espectro via sonograma das figuras 5.7 e 5.8. Uma melhor representação pode ser obtida aumentando o grau do polinômio empregado na aproximação, porém com custo computacional maior.

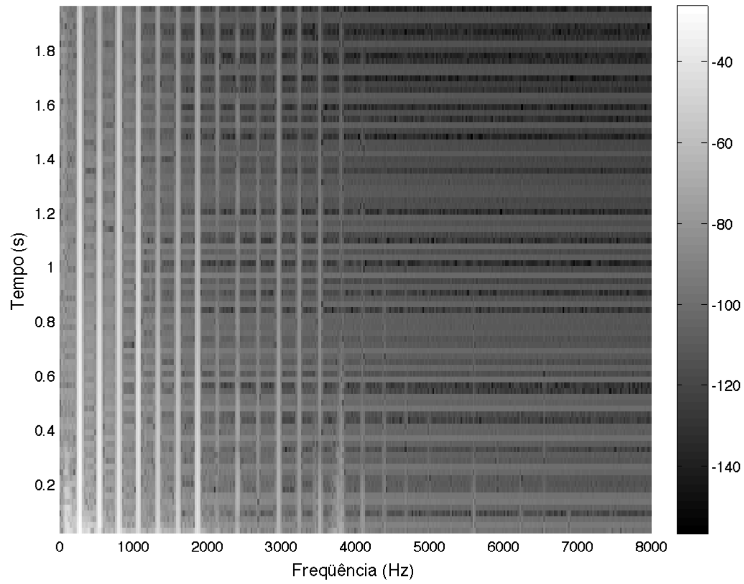

Fig. 5.7: Representação via sonograma da nota C4 original do piano.

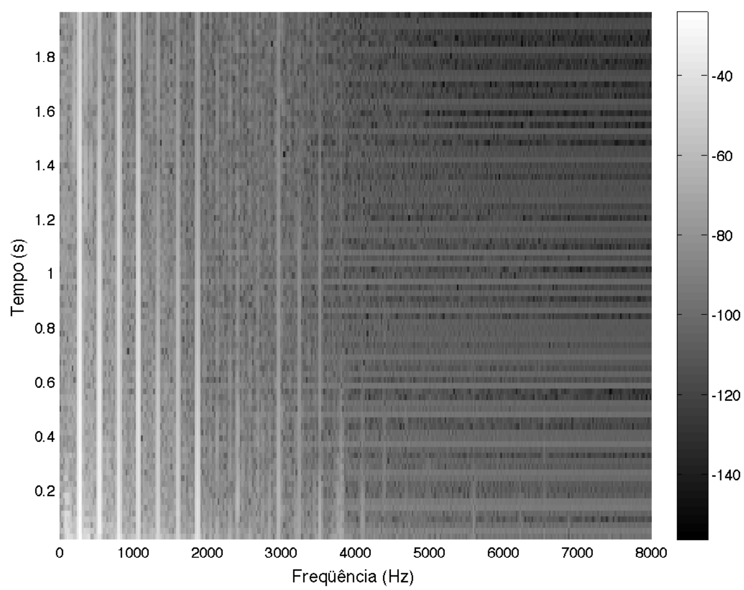

Fig. 5.8: Representação via sonograma da nota C4 do piano gerada pelo modelo por aproximação polinomial.

5.2.3 Representação de um Trecho Rítmico Executado por Bateria

Nesta subseção utilizamos o modelo proposto para reproduzir dois compassos de um trecho rítmico de *drum'n bass*, executado por uma bateria eletrônica *Roland TR808*. As amostras utilizadas são de seis segundos, tempo necessário para os dois compassos.

As possibilidades de representação do modelo foram exploradas ao extremo. Pois, as amostras envolvidas são extremamente complexas, com sons inarmônicos e com envelopes espectrais e temporais de grandes variações. A figura 5.9 traz o gráfico do envelope temporal do som alvo, que ilustra bem a riqueza rítmica que motiva o experimento.

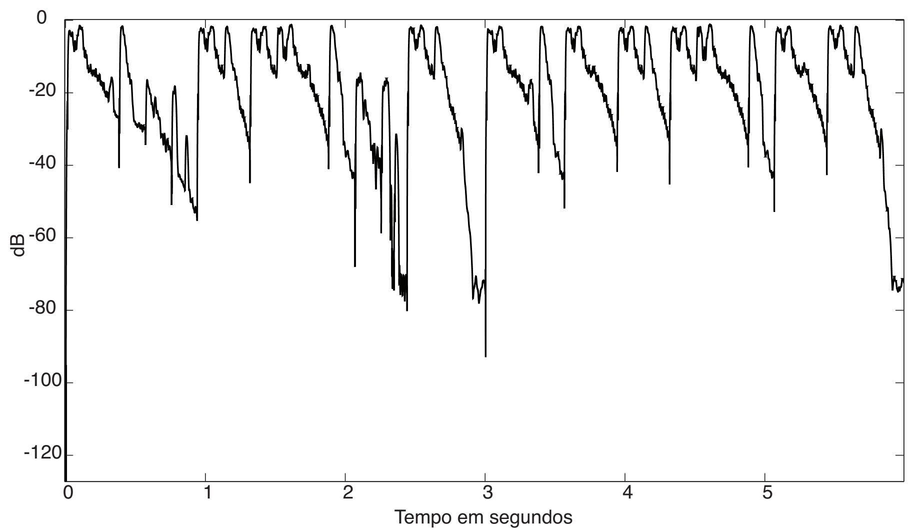

Fig. 5.9: Envelope temporal, em dB, do sinal da bateria.

Nas figuras 5.10 e 5.11 estão representadas as curvas provenientes das ordenações dos coeficientes do sinal da bateria e sua aproximação para os níveis 7 e 8 de detalhes.

Quanto maior a presença de transientes no sinal, maior a concentração de energia na faixa de freqüência responsável pelos ataques. Isto significa que alguns poucos coeficientes, que estão nos níveis relativos àquela faixa de freqüência, oscilam com grandes amplitudes. Como conseqüência, as curvas provenientes das suas ordenações têm extremos acentuados e as regiões centrais, próximas do eixo y, com valores próximos de zero. No som da bateria, os transientes são extremos. Por este motivo, para aproximar seus coeficientes de forma satisfatória, precisamos elevar o grau do polinômio para 11. Caso contrário, as oscilações causadas pelo grau baixo do polinômio implicariam na perda da definição dos ataques. Para exemplificar, figura 5.12 traz a aproximação do mesmo nível 7 de detalhes por uma função polinomial de grau 3.

Os valores das distâncias euclidianas entre as curvas originais e aproximadas estão na tabela 5.3.

E nas figuras 5.13 e 5.14 estão os sonogramas do sinal original e do sinal representado pelo modelo. É possível notar que no sonograma do sinal original há uma maior ocorrência de listras

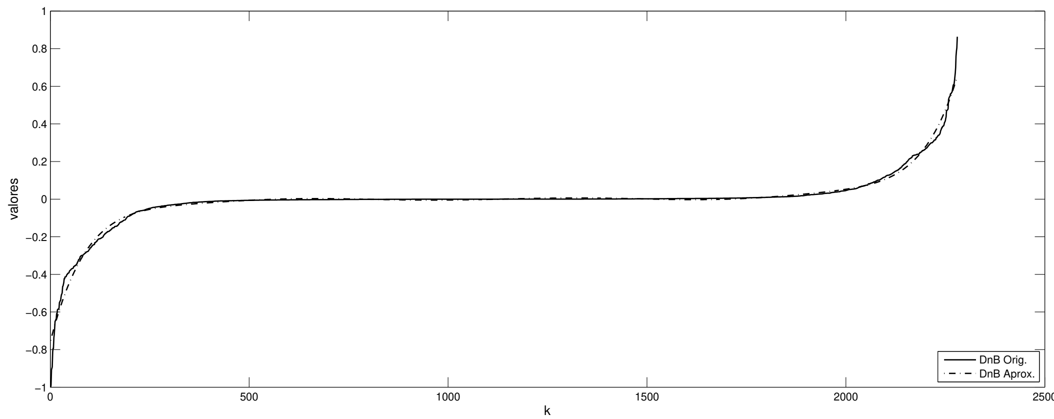

Fig. 5.10: Curva proveniente da ordenação ascendente do nível de detalhes 7 do sinal da bateria e de sua aproximação por um polinômio de grau 11.

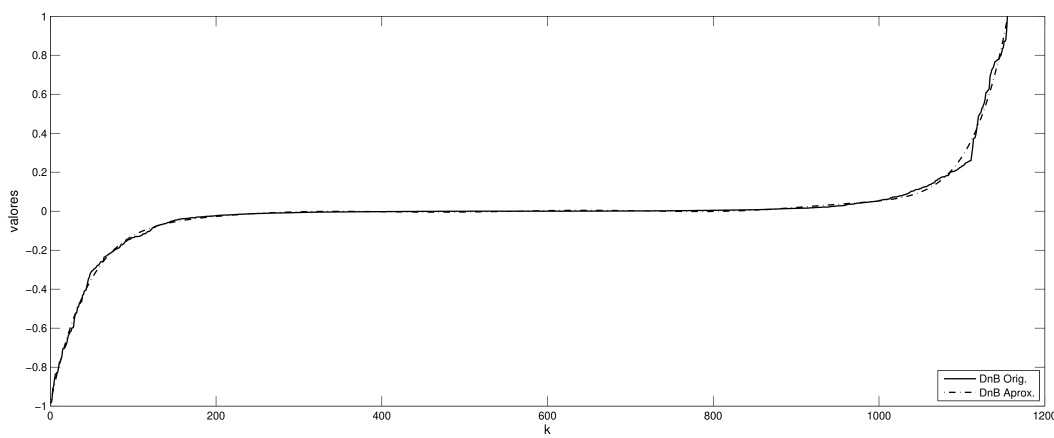

Fig. 5.11: Curva proveniente da ordenação ascendente do nível de detalhes 8 do sinal da bateria e de sua aproximação por um polinômio de grau 11.

horizontais claras (aproximadamente 0 dB), o que indica a presença de pausas no sinal. Já no sinal representado pelo modelo, essas listras não aparecem. Isto está relacionado ao surgimento de um pouco de ruído, gerado pelas aproximações.

Na seção seguinte, apresentaremos os resultados obtidos através do modelo por coeficientes de sons amostrados.

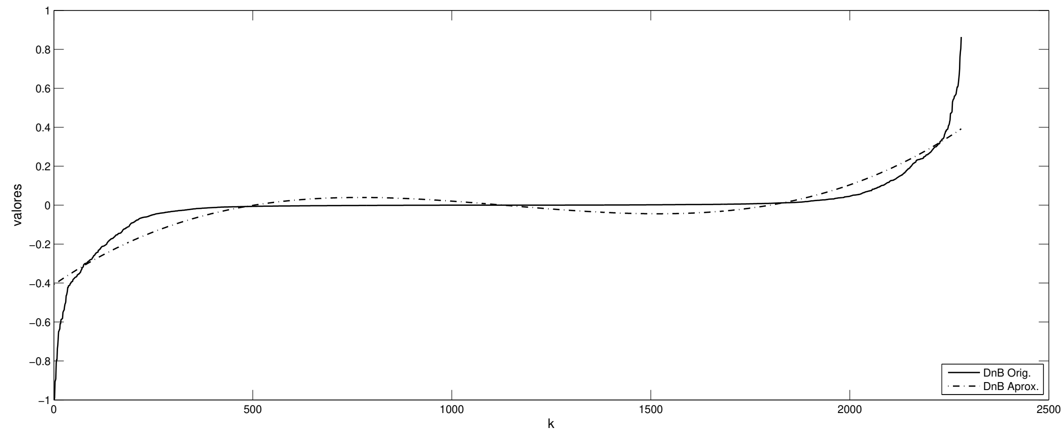

Fig. 5.12: Curva proveniente da ordenação ascendente do nível de detalhes 7 do sinal da bateria e de sua aproximação por um polinômio de grau 3.

Vetores	*Comprimento*	Freq.	Dist
$Dc1$	48015	24000	0.0178
$Dc2$	24023	12000	0.0151
$Dc3$	12027	6000	0.0179
$Dc4$	6029	3000	0.0226
$Dc5$	3030	1500	0.026
$Dc6$	1530	750	0.0199
$Dc7$	780	375	0.018
$Dc8$	405	187.5	0.0139
$Ac8$	405	93.75	0.0109

Tab. 5.3: Distâncias euclidianas entre as curvas ordenadas dos níveis originais e aproximados, para o exemplo da bateria.

5.3 Representação Através do Modelo por Coeficientes de Sons Amostrados

5.3.1 Coeficientes Extraídos da Clarineta

Neste experimento, aproximamos os coeficientes base pela nota G#3 da clarineta. Utilizaremos com alvo duas notas distintas, também da clarineta. Os alvos são: um D3 e um E4, notas extremas da região *chalumeau*.

Nas figuras de 5.15 a 5.16 estão representadas as curvas provenientes das ordenações dos coefici-

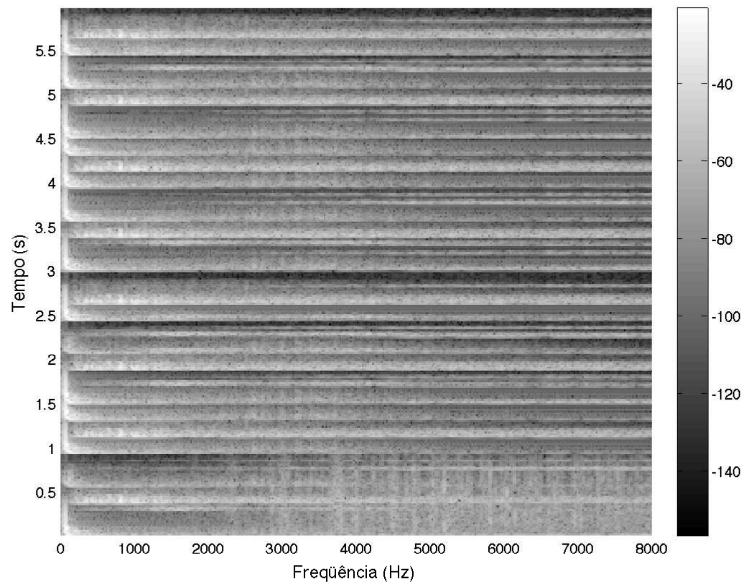

Fig. 5.13: Representação via sonograma do espectro do trecho original executado pela bateria.

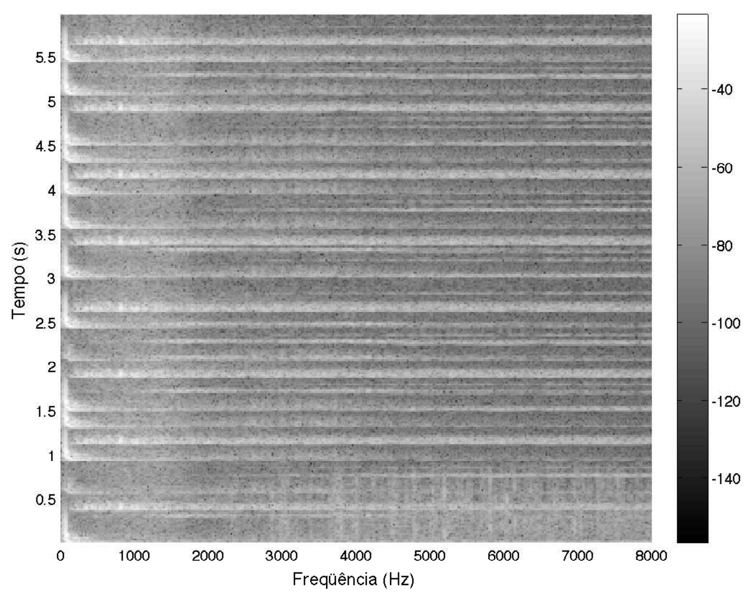

Fig. 5.14: Representação via sonograma do espectro do trecho executado pela bateria gerado pelo modelo por aproximação polinomial.

entes do D3 e do G#3, dos níveis 6 e 8 de detalhes. Os gráficos estão organizados da seguinte forma: nas linhas sólidas estão as curvas ordenadas do D3 e nas linhas pontilhadas as curvas do sinal base G#3; em vermelho estão as curvas geradas pelo ajuste de energia dos coeficientes base. Analogamente, para as figuras 5.17 e 5.18, correspondentes aos níveis de detalhes 5 e 7 da nota E4. Estes níveis foram escolhidos pela relevância espectral dessas notas, nestas faixas de freqüência.

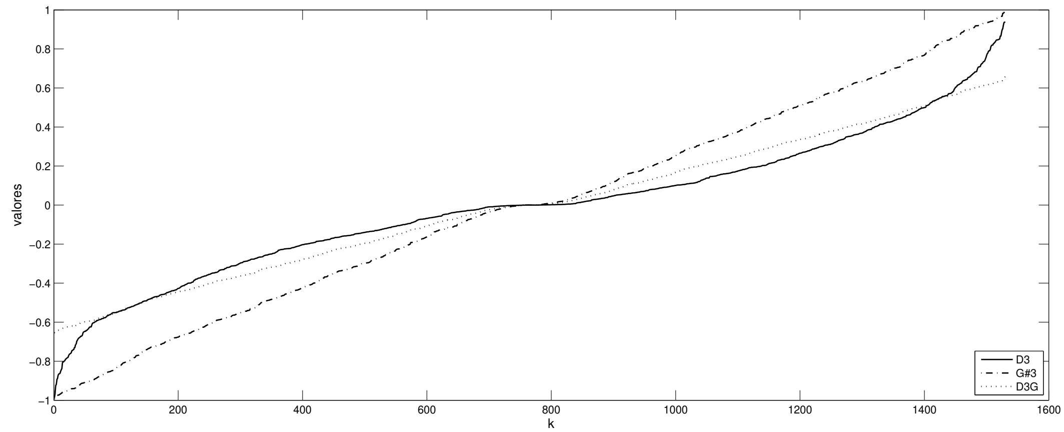

Fig. 5.15: Curva proveniente da ordenação ascendente do nível de detalhes 6 das notas D3 e G#3 da clarineta.

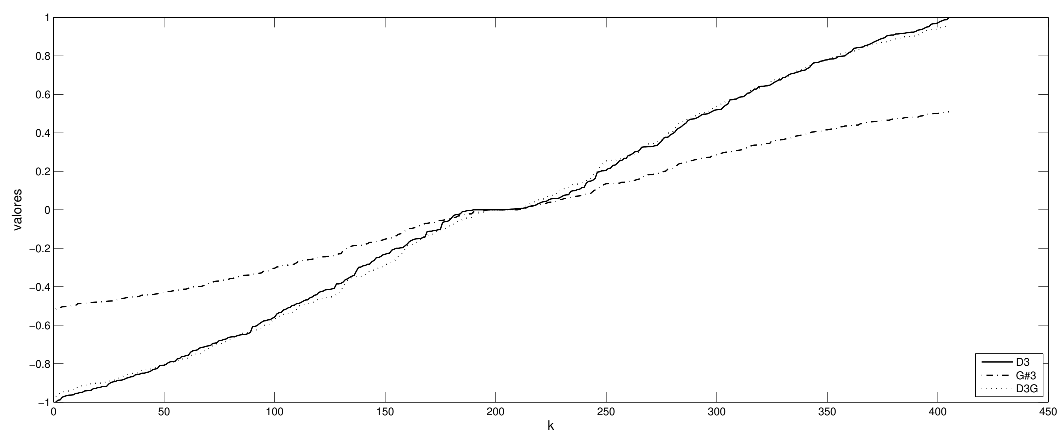

Fig. 5.16: Curva proveniente da ordenação ascendente do nível de detalhes 8 das notas D3 e G#3 da clarineta.

Para os experimentos com coeficientes provenientes de sons amostrados, foram calculados três valores de distância para cada nível, com as seguintes nomenclaturas:

- Dist: entre as curvas ordenadas do alvo e da base;
- DistG: entre as curvas ordenadas do alvo e as curvas obtidas pelo ajuste de energia do som base;
- DistN: entre as curvas ordenados do alvo e a curvas finais, com as $N = 4000$ trocas.

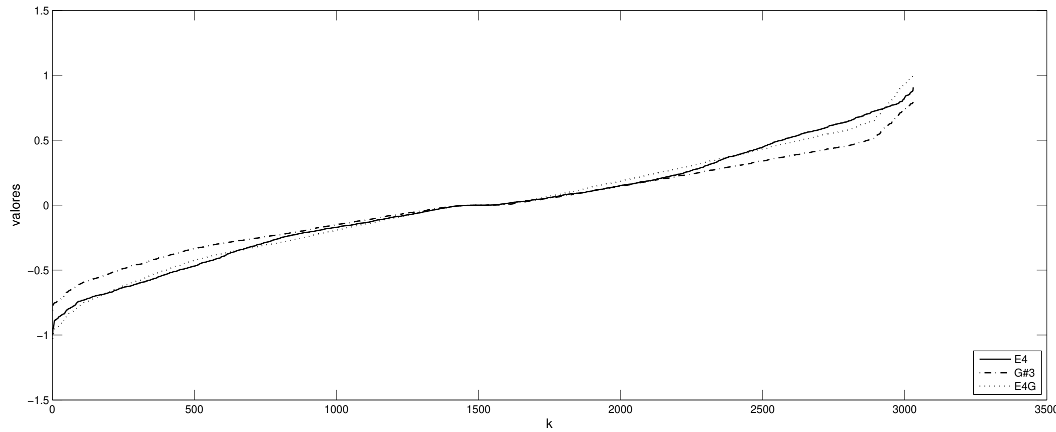

Fig. 5.17: Curva proveniente da ordenação ascendente do nível de detalhes 5 das notas E4 e G#3 da clarineta.

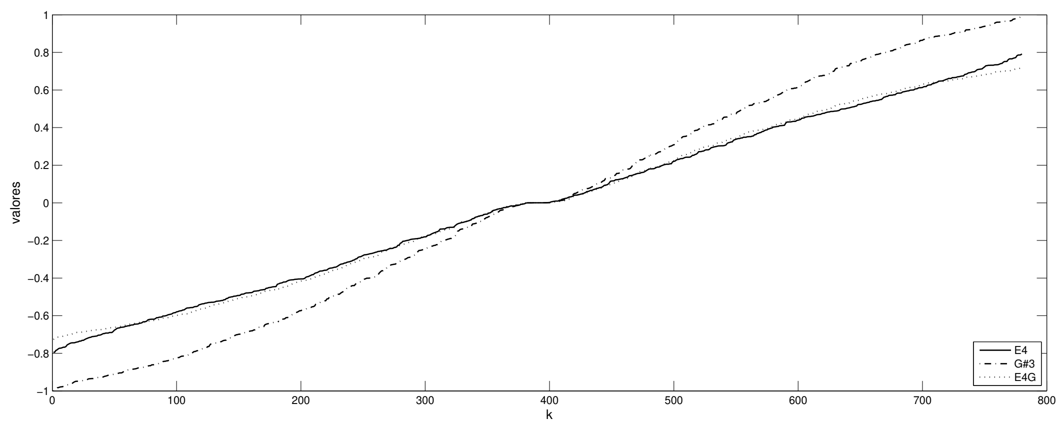

Fig. 5.18: Curva proveniente da ordenação ascendente do nível de detalhes 7 das notas E4 e G#3 da clarineta.

As distâncias foram calculadas através da expressão 5.1. Os valores para exemplo D3 e E4 estão dispostos na tabela 5.4 e 5.4, respectivamente.

Pelos gráficos das figuras 5.15 a 5.18 podemos perceber que há uma forte semelhança entre as curvas originais e as curvas finais, resultante do processo de ajuste de energia, para cada nível n. Isso confere para os demais níveis, tanto para o D3 como para o E4. Ou seja, os coeficientes da nota G#3 constituem uma boa aproximação para as duas notas alvo do experimento. Desta forma, a partir da permutação P_n adequada é possível transitar por essas duas notas, e por conseqüência, pela região *chalumeau* toda. Caso o intuito seja a completa *mimesis* do sinal de entrada, podemos ainda tratar as pequenas diferenças por meio do modelo residual, como apresentado a seguir.

Vetores	*Comprimento*	Freq.	Dist	DistG	DistN
$Dc1$	48015	24000	0.078	0.00579	0.00579
$Dc2$	24023	12000	0.102	0.0106	0.0106
$Dc3$	12027	6000	0.132	0.0112	0.0112
$Dc4$	6029	3000	0.0867	0.0267	0.0225
$Dc5$	3030	1500	0.11	0.0618	0.00524
$Dc6$	1530	750	0.2	0.0654	0.00661
$Dc7$	780	375	0.512	0.0491	0.0074
$Dc8$	405	187.5	0.282	0.0229	0.013
$Ac8$	405	93.75	0.0331	0.0197	0.0153

Tab. 5.4: Distâncias euclidianas entre os coeficientes dos níveis ordenados para o exemplo das notas D3 alvo com G#3 base. Ambas da clarineta

Vetores	*Comprimento*	Freq.	Dist	DistG	DistN
$Dc1$	48015	24000	0.0654	0.0119	0.0119
$Dc2$	24023	12000	0.0582	0.0111	0.0111
$Dc3$	12027	6000	0.167	0.0117	0.0117
$Dc4$	6029	3000	0.214	0.0543	0.0218
$Dc5$	3030	1500	0.0941	0.0345	0.00662
$Dc6$	1530	750	0.334	0.011	0.011
$Dc7$	780	375	0.173	0.0211	0.0102
$Dc8$	405	187.5	0.62	0.00263	0.00263
$Ac8$	405	93.75	0.054	0.00311	0.00311

Tab. 5.5: Distâncias euclidianas entre os coeficientes dos níveis ordenados para o exemplo das notas E4 alvo com G#3 base.

Para o tratamento dos resíduos precisamos antes estabelecer um critério de troca para os coeficientes. Aqui, o objetivo foi a reprodução idêntica do sinal alvo. Então, o critério utilizado foi a comparação psicoacústica objetivando, com um N pequeno, obter a melhor qualidade sonora da saída em relação à entrada. Os testes psicoacústicos foram realizados pelo autor, em um estúdio apropriado, com um sistema de referência de alta fidelidade em sala com reverberação e resposta de freqüência adequadas.

As trocas efetuadas pelo modelo residual aproximam ainda mais as curvas finais das originais, representadas pelo alvo. Podemos verificar estas aproximação através da comparação entre as duas últimas colunas das tabelas 5.4 e 5.5.

As figuras 5.19 e **??** mostram que coeficientes foram trocados para as notas D3 e E4, respecti-

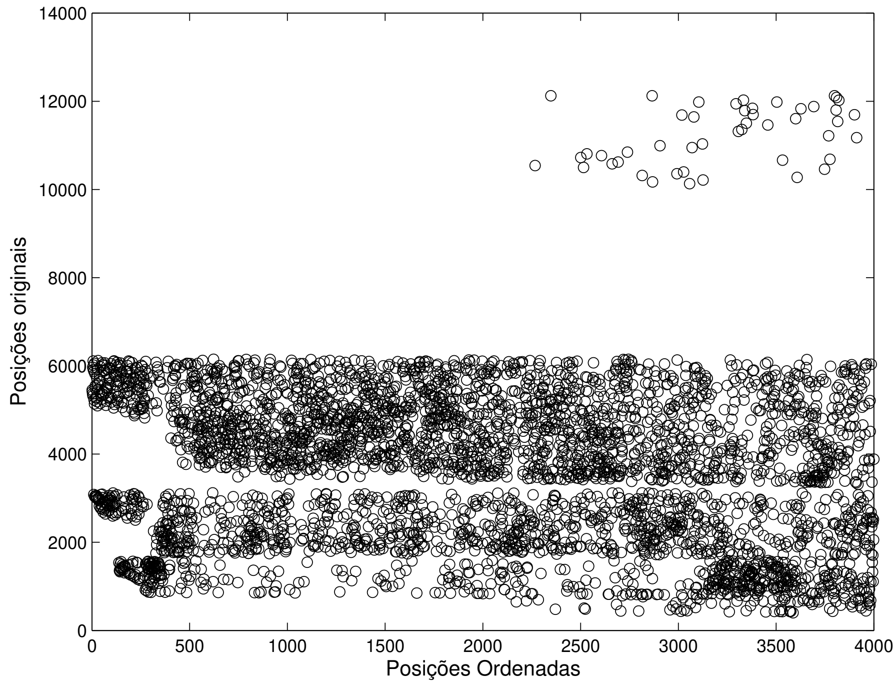

Fig. 5.19: Posição dos $N = 4000$ coeficientes trocados para a nota D3.

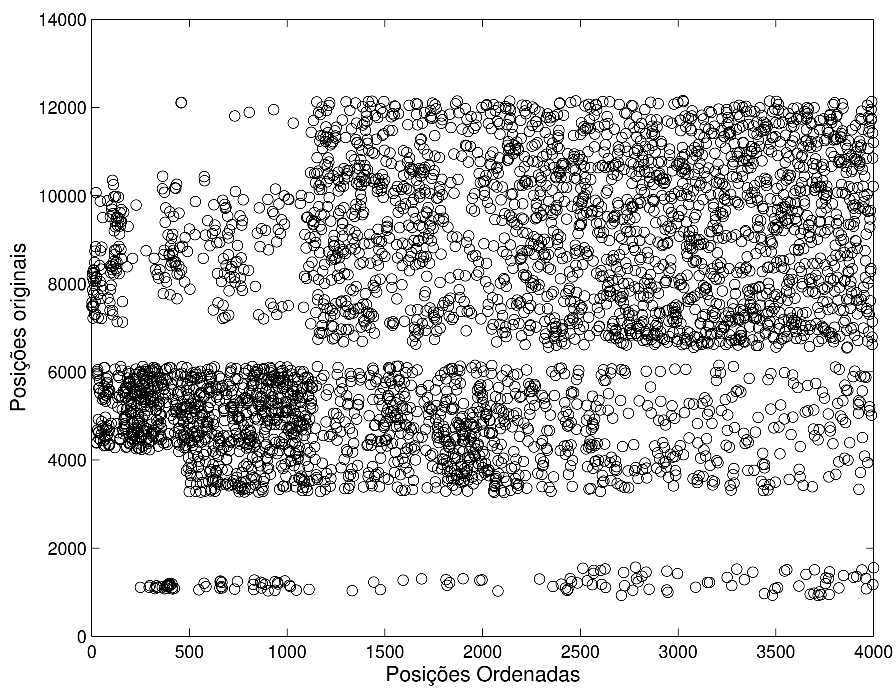

Fig. 5.20: Posição dos $N = 4000$ coeficientes trocados para a nota E4.

vamente. Nestas figuras a numeração das posições do vetor de coeficientes original é feita a partir dos níveis correspondentes às freqüências mais baixas - nível oito de aproximação, em seguida nível oito de detalhes, sete de detalhes, consecutivamente até o nível um de detalhes (ver a figura 4.2 onde os níveis de freqüência mais baixa encontram-se na parte inferior do diagrama). No eixo x estão as posições ordenadas dos coeficientes e no eixo y suas posições originais. A partir destas figuras, é possível saber a posição e o nível de cada um dos R coeficientes mais distintos substituídos pelos originais.

De forma geral, em instrumentos harmônicos as trocas estão concentradas nas primeiras posições do vetor de coeficientes original concatenado. Nos sons predominantemente harmônicos, como a clarineta, a maior parte da energia está concentrada nas primeiras parciais, e portanto o conteúdo de maior relevância do sinal está focado nesta região. Desta forma, é natural que as maiores discrepâncias ocorram nas primeiras posições.

No caso do D3, as mudanças estão concentradas de forma mais acentuada nos níveis de detalhes cinco, seis, sete, oito e oito de aproximação. Já para o caso do E4, pela tabela 5.5 vemos que os níveis sete e oito de detalhes não sofreram alteração no valor de distância após as trocas. Podemos constatar esse fato pelo gráfico da figura 5.19. Na faixa de 0 até aproximadamente 1000, das posições originais, não houve nenhum ponto trocado. Isto acontece porque nesta faixa de freqüência não há nenhum componente espectral desta nota, que tem sua fundamental em em 271.2 Hz e se encontra no sétimo nível de detalhes. As trocas estiveram em níveis mais altos devida a distância entre sua fundamental e a do G#3.

Como forma de visualização do espectro, utilizamos a representação em cascata da transformada enjanelada de Fourier. Nas figuras 5.3, 5.21 e 5.22 temos a nota D3 original, permutada e com as N trocas, respectivamente. Analogamente para para o E4, nas figuras 5.23, 5.21 e 5.22.

As diferenças entre os espectros são extremamente sutis, o que pode ser também constatado através da audição das amostras que estão no CD. Algumas granulações pequenas estão presentes, porém após as trocas estas diferenças ficam praticamente imperceptíveis.

5.3.2 Instrumentos Distintos

Ao utilizarmos como nota alvo um instrumento distinto ao da nota base, verificamos de forma acentuada que as mudanças na sonoridade estão fortemente relacionadas com o posicionamento dos coeficientes. Ao aplicarmos uma permutação adequada a base, constituída pelos coeficientes da clari-

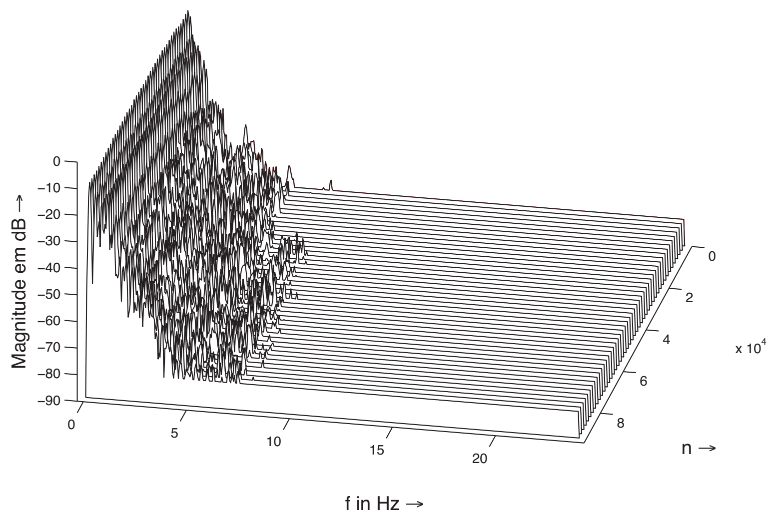

Fig. 5.21: Representação em cascata do espectro da nota D3 reproduzida através das permutações dos coeficientes da nota G#3, ambas da clarineta.

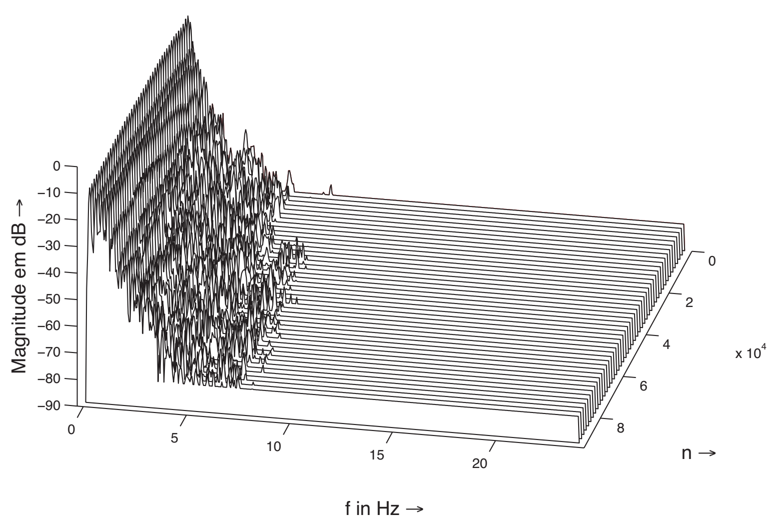

Fig. 5.22: Representação em cascata do espectro da nota D3 da clarineta reproduzida através das permutações dos coeficientes da nota G#3 também da clarineta, com 4000 coeficientes trocados pelos seus respectivos do som alvo.

neta, transitamos pelos timbres de piano e bateria, mudando completamente as características sonoras do material de partida. É isto que veremos nesta subseção.

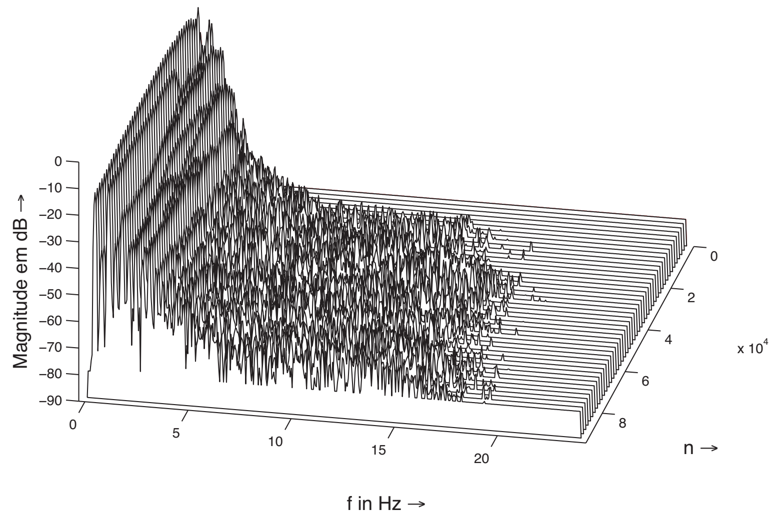

Fig. 5.23: Representação em cascata do espectro da nota E4 original da clarineta.

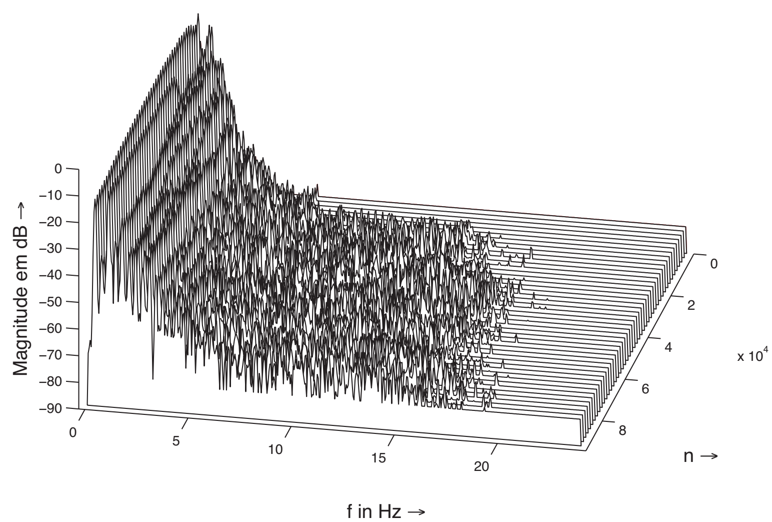

Fig. 5.24: Representação em cascata do espectro da nota E4 reproduzida através das permutações dos coeficientes da nota G#3, ambas da clarineta em Bb.

Clarineta e Piano

Nesta parte dos experimentos, utilizamos como som alvo um som de piano C4, e como som base o mesmo G#3 da clarineta em Bb. O objetivo foi confrontar o modelo na geração de sons harmônicos, porém com transientes mais acentuados que os da clarineta, som base no processo descrito. A figuras 5.26 e 5.27 mostram os níveis ordenados de detalhes 6 e 7 destas amostras.

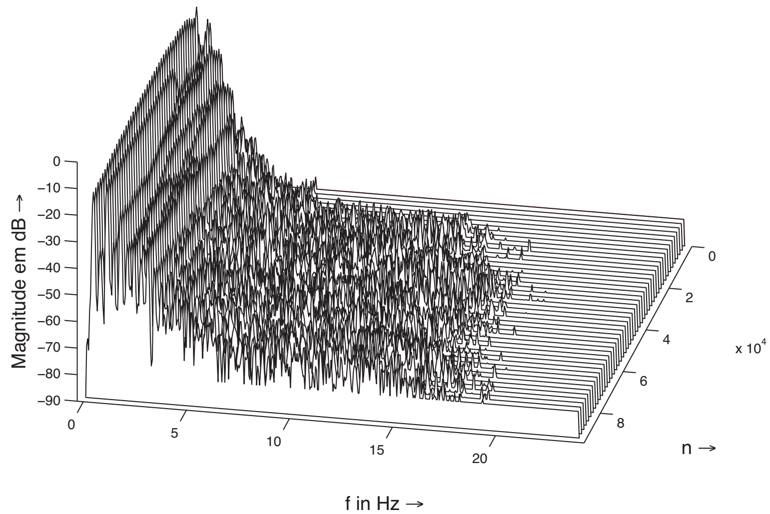

Fig. 5.25: Representação em cascata do espectro da nota E4 da clarineta reproduzida através das permutações dos coeficientes da nota G#3 também da clarineta, com 4000 coeficientes trocados pelos seus respectivos do som alvo.

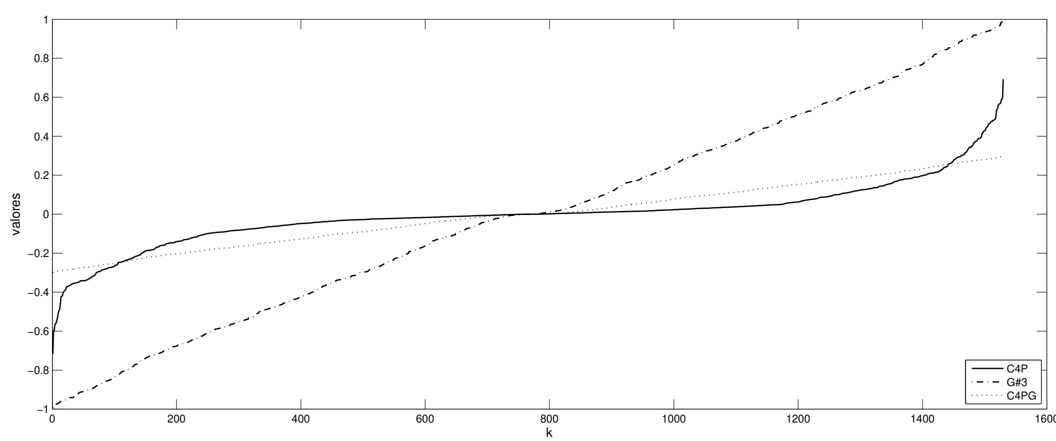

Fig. 5.26: Curva proveniente da ordenação ascendente do nível de detalhes 6 das notas C4 do piano e G#3 da clarineta.

A tabela 5.6 apresenta os valores para as distâncias euclidianas, da mesma forma como exposto anteriormente, com $N = 4000$. O valor de N aqui também foi escolhido de forma experimental, variando a cada tentativa até a obtenção de um valor coerente do ponto de vista psicoacústico, já utilizando-o como parâmetro livre, para criação da sonoridade.

O gráfico da figura 5.28 mostra quais coeficientes foram trocados. No eixo x estão as posições ordenadas de 1 a N e no eixo y as posições *originais* ocupadas no vetor de coeficientes. Através desta

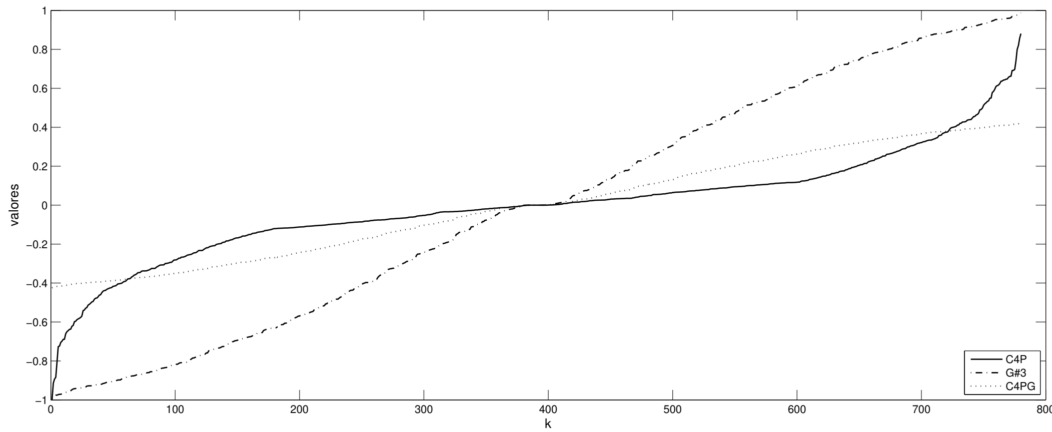

Fig. 5.27: Curva proveniente da ordenação ascendente do nível de detalhes 7 das notas C4 do piano e G#3 da clarineta.

Vetores	*Comprimento*	Freq.	Dist	DistG	DistN
$Dc1$	48015	24000	0.0863	0.00431	0.00431
$Dc2$	24023	12000	0.147	0.00626	0.00626
$Dc3$	12027	6000	0.14	0.042	0.0411
$Dc4$	6029	3000	0.092	0.0485	0.0321
$Dc5$	3030	1500	0.297	0.0315	0.00634
$Dc6$	1530	750	0.396	0.0708	0.00339
$Dc7$	780	375	0.394	0.112	0.00314
$Dc8$	405	187.5	0.603	0.0435	0.0177
$Ac8$	405	93.75	0.0874	0.0872	0.0235

Tab. 5.6: Distâncias euclidianas entre os coeficientes dos níveis ordenados para o C3 do piano como nota alvo e com o G#3 da clarineta como base.

figura e da tabela 5.6 notamos que as mudanças foram realizadas de forma estratificada nos quatro primeiros níveis, posição máxima em 19452 e freqüência máxima de 3000 Hz. Ou seja, conforme o N aumenta no eixo x, as posições dos coeficientes trocados são maiores.

Nas figuras de 5.29 a 5.31 estão os espectros dos sinais originais e gerados pelo modelo, para a nota C4 do piano. Nota-se que nas altas freqüências, do sinal gerado, há uma pequena perda de resolução.

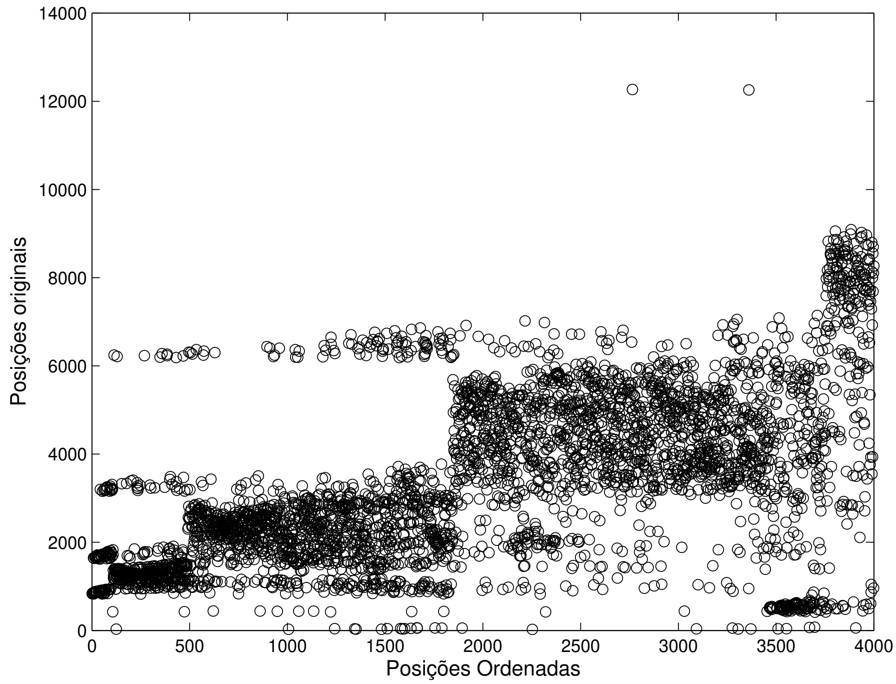

Fig. 5.28: Posição dos $N = 4000$ coeficientes trocados para a nota C4.

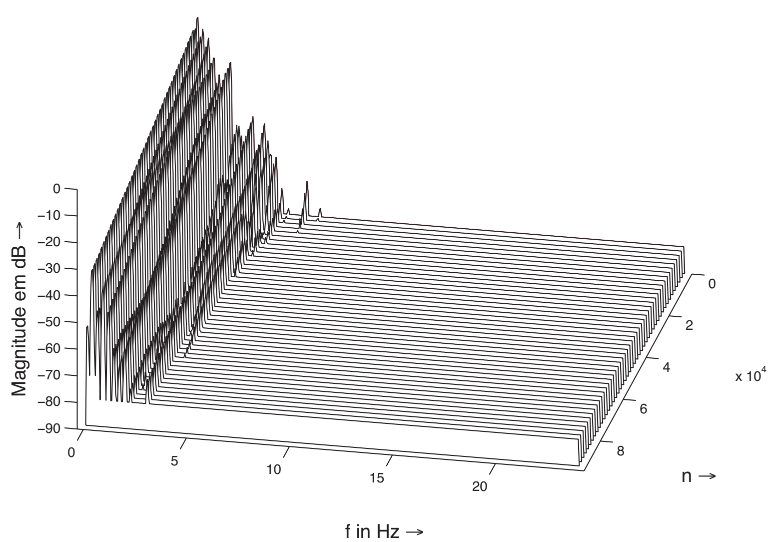

Fig. 5.29: Representação em cascata do espectro da nota C4 original do Piano.

5.3.3 Clarineta e Bateria

Aqui os experimentos foram realizados através da mesma nota G#3 da clarineta, porém a amostra alvo foi extraída dos dois compassos do trecho ritmo de *Drum'n Bass* executado pela bateria eletrô-

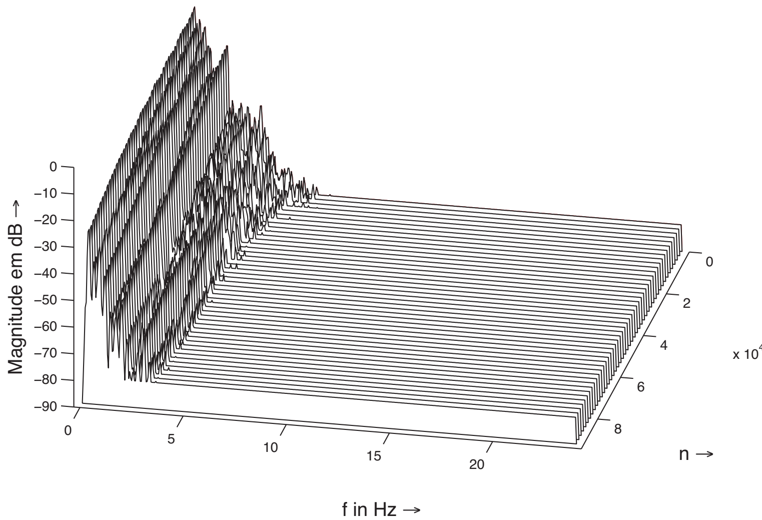

Fig. 5.30: Representação em cascata do espectro da nota C4 do piano reproduzida através das permutações dos coeficientes da nota G#3 amostrada da clarineta.

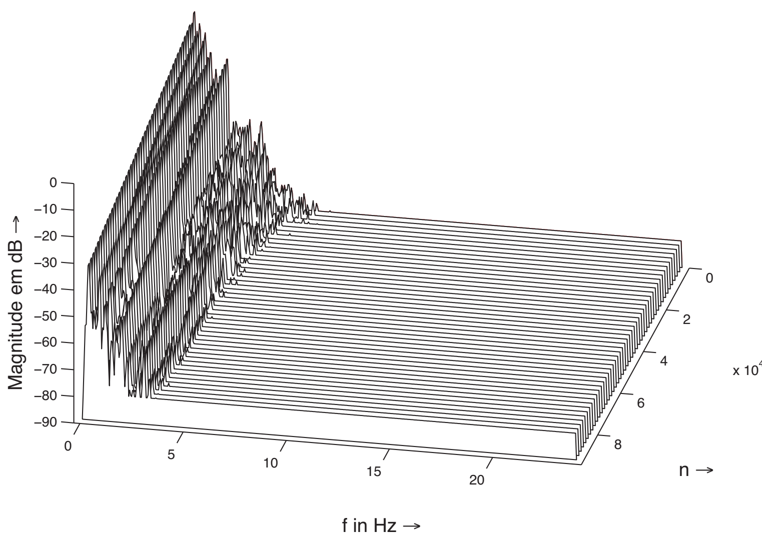

Fig. 5.31: Representação em cascata do espectro da nota C4 do piano reproduzida através das permutações dos coeficientes da nota G#3 da clarineta, com 4000 coeficientes trocados pelos seus respectivos do som alvo.

nica. Os resíduos foram extraídos de um excerto melódico, também do mesmo estilo e andamento, executado por um baixo sintético. Aqui utilizamos amostras com seis segundos, tempo necessário para os dois compassos.

O foco central deste experimento esteve voltado para a exploração do modelo na criação de novas sonoridades. Ao exploramos o modelo residual, pudemos avaliar o potencial sonoro do modelo, gerado pela fusão das características psicoacústicas envolvidas. O resultado foi uma composição sonora extremamente interessante do ponto de vista musical.

As figuras 5.32, 5.33 e 5.34 correspondem às curvas dos ordenamentos dos níveis 6, 7 e 8. No geral, as curvas são bem distintas. Isto confere a sonoridade resultante uma mistura das características espectrais correspondentes as diferenças de cada som envolvido.

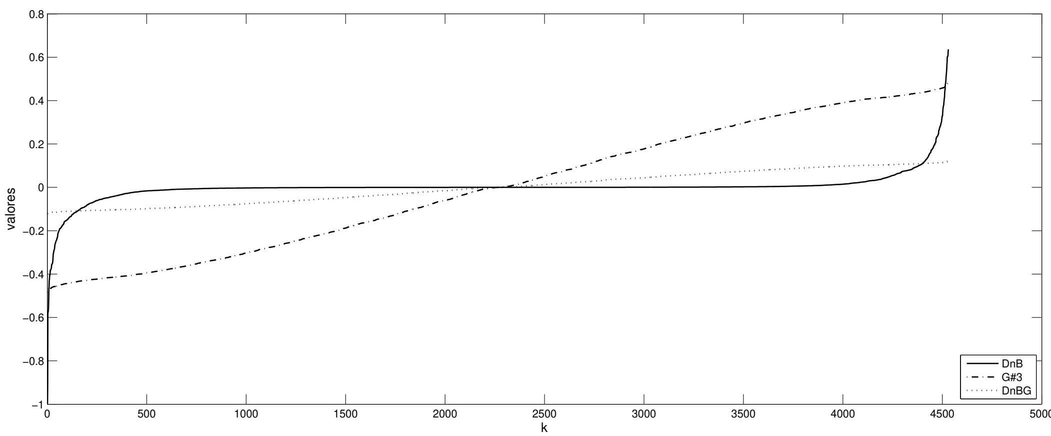

Fig. 5.32: Curva proveniente da ordenação ascendente do nível de detalhes 6 do trecho executado pela bateria e da nota G#3 da clarineta.

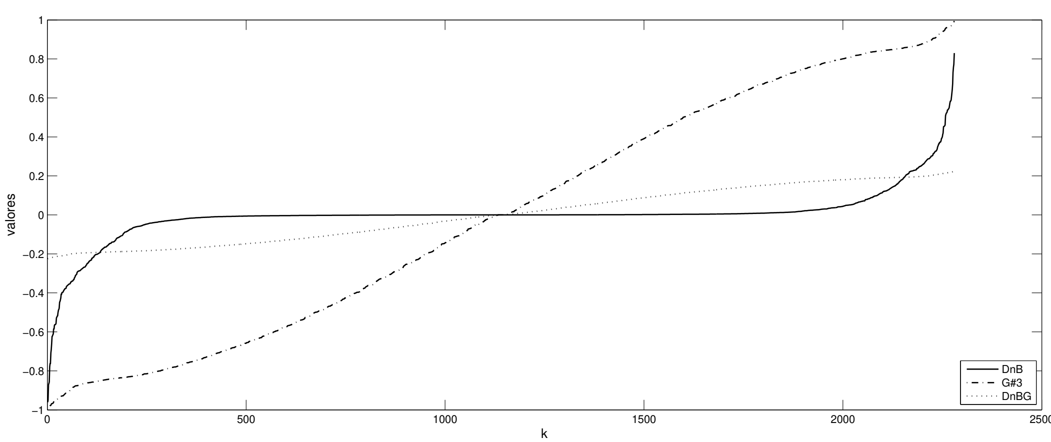

Fig. 5.33: Curva proveniente da ordenação ascendente do nível de detalhes 7 do trecho executado pela bateria e da nota G#3 da clarineta.

A tabela 5.7 apresenta os valores para as distâncias euclidianas, da mesma forma como exposto anteriormente, porém aqui com $N = 12000$. O objetivo das trocas foi produzir uma sonoridade

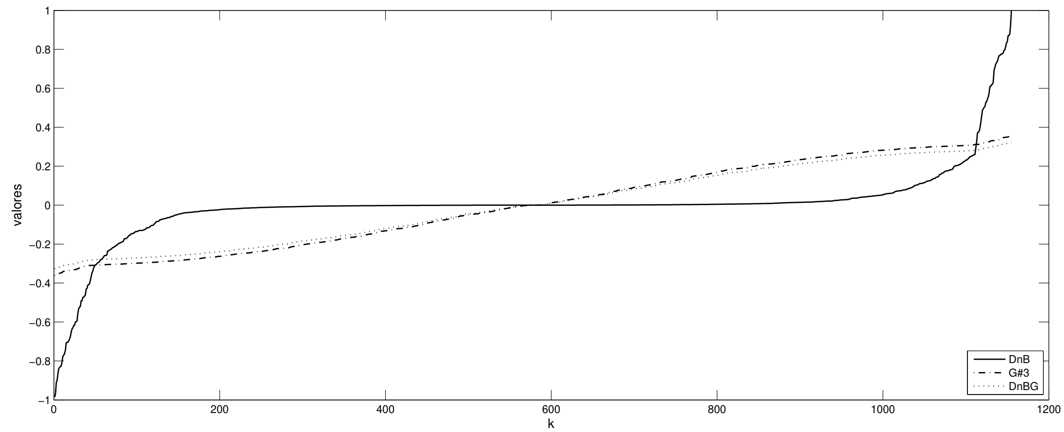

Fig. 5.34: Curva proveniente da ordenação ascendente do nível de detalhes 8 do trecho executado pela bateria e da nota G#3 da clarineta.

diferente, composta por elementos da bateria, clarineta e do baixo utilizado como material de extração dos coeficientes residuais.

Vetores	*Comprimento*	Freq.	Dist	DistG	DistN
$Dc1$	48015	24000	0.0746	0.0462	0.0462
$Dc2$	24023	12000	0.0968	0.0606	0.0645
$Dc3$	12027	6000	0.0653	0.0469	0.0513
$Dc4$	6029	3000	0.0561	0.0545	0.0563
$Dc5$	3030	1500	0.138	0.0831	0.0353
$Dc6$	1530	750	0.262	0.0722	0.0457
$Dc7$	780	375	0.54	0.122	0.0671
$Dc8$	405	187.5	0.184	0.175	0.114
$Ac8$	405	93.75	0.249	0.301	0.0713

Tab. 5.7: Distâncias euclidianas entre os coeficientes dos níveis ordenados para a bateria como alvo, com o G#3 da clarineta como base e com os coeficientes residuais extraídos do som de baixo sintético.

O gráfico da figura 5.35 mostra quais coeficientes foram trocados. As trocas foram realizadas de forma bem esparsa, porém com forte concentração nas primeiras 2000 posições. Este fato se deve a grande concentração de energia existente na região grave da bateria, devido ao bumbo. As correções de energia, realizadas nos níveis do sinal da clarineta, não corrigem totalmente essas diferença, como podemos observar nas figuras 5.32, 5.33 e 5.34.

E por fim, as figuras 5.36, 5.37 e 5.38 trazem os espectros do sinal da bateria original, do baixo e da composição do sinal resultante, produzido pelo modelo.

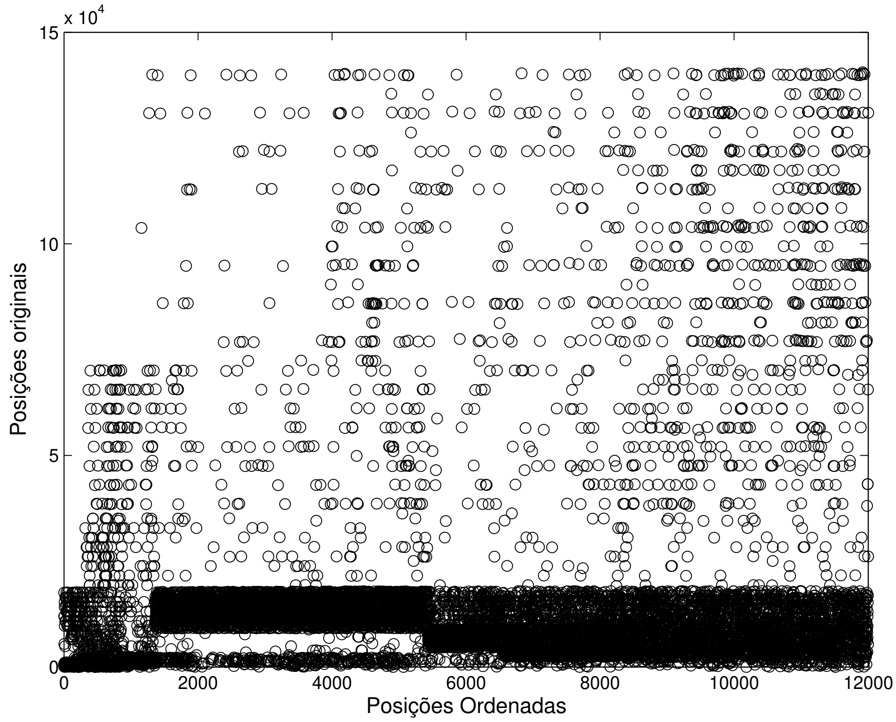

Fig. 5.35: Posição dos $N = 12000$ coeficientes trocados para a composição da sonoridade envolvendo a bateria, clarineta e baixo sintético.

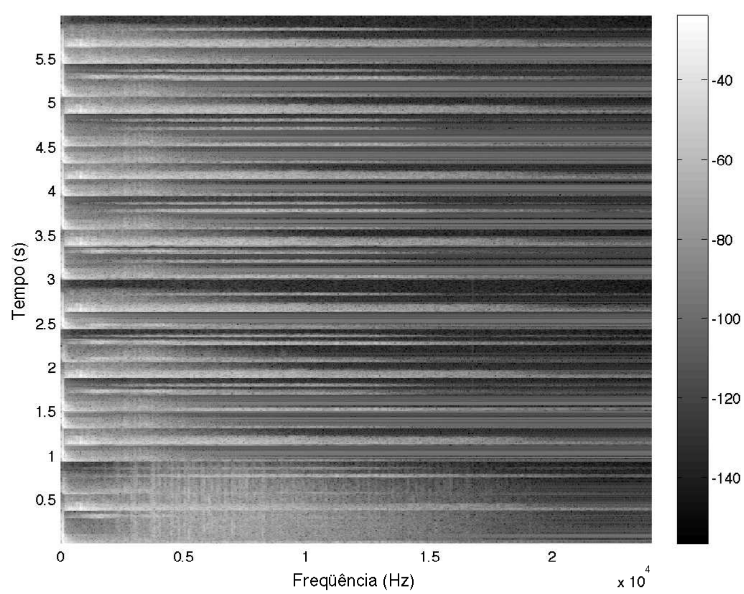

Fig. 5.36: Representação via sonograma do espectro da bateria.

Neste capítulo apresentamos alguns dos resultados obtidos através do modelo proposto no capítulo 4. Pudemos avaliar seu potencial, destacando algumas de suas formas de aplicação. Dentre elas,

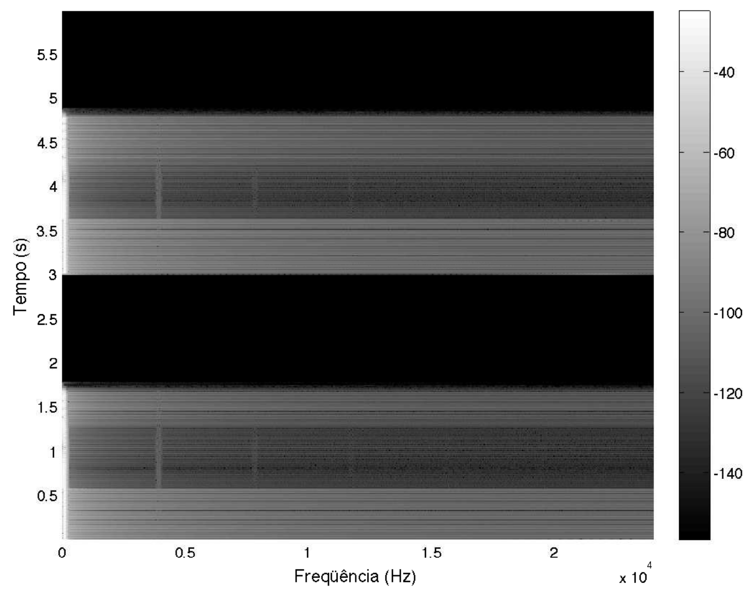

Fig. 5.37: Representação via sonograma do espectro do baixo.

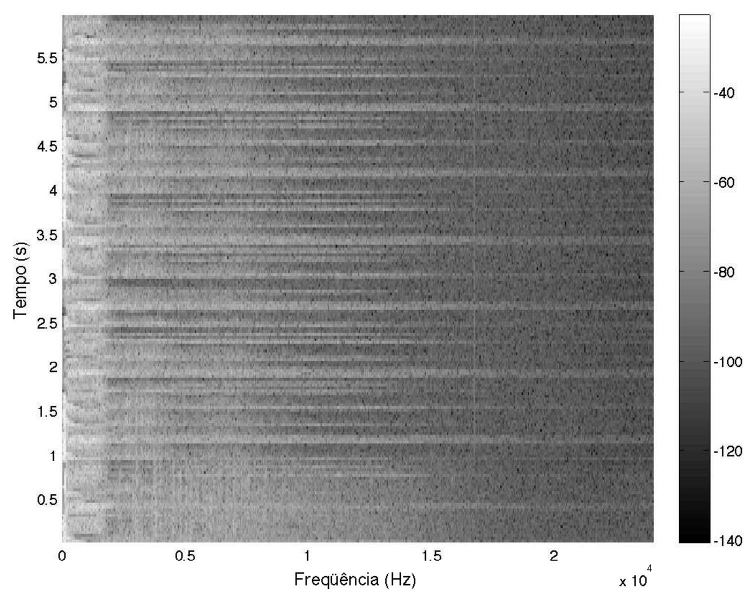

Fig. 5.38: Representação do espectro, via sonograma, da sonoridade produzida através das permutações dos coeficientes da nota G#3 da clarineta, com alvo na bateria eletrônica e com 12000 coeficientes trocados pelos seus respectivos do baixo sintético.

exploramos a criação e representação sonora pela fusão de materiais diversos e também a recodificação econômica de sinais musicais. No capítulo que se segue, faremos as conclusões e as projeções para o futuro deste trabalho.

Capítulo 6

Conclusão

Nesta dissertação, realizamos um estudo sobre a representação de sinais musicais, para propor um método eficaz e robusto, capaz de suprir necessidades técnicas nas diversas vertentes apresentadas no capítulo 2. A partir do que foi exposto, desenvolveu-se um estudo onde elaboramos um método de representação de sinais musicais por transformadas *wavelets*. O uso de *wavelets* foi estratégico porque através desta ferramenta pudemos tratar os transientes sonoros e sua evolução temporal com acuidade. Esta possibilidade de atuar sobre as componentes estáveis do sinal e seus transientes representou um avanço no tratamento de sinais musicais.

A questão de fundo deste trabalho foi qualificar a relevância de características dinâmicas do sinal como inarmonicidade, vibratos, ruídos, distorções provenientes da propagação das ondas no ar dentre outras que motivaram esta pesquisa. Outro ponto foi como manipulá-las, frente à possibilidade de síntese de um som musical. Este entorno gerou questões subjetivas que têm o potencial de encontrar soluções nas análises de sons acústicos. Assim, é necessária uma ferramenta poderosa, capaz de atuar como um "telescópio"onde, a cada nova escala, novos detalhes do fenômeno sonoro são revelados.

O estudo do nosso modelo foi desenvolvido a partir de quatro vetores qualitativos propostos por Serra: qualidade sonora, flexibilidade, generalidade e esforço computacional, como exposto no capítulo 2. Estes vetores representam um objetivo amplo. Porém, a partir deste ponto de vista, buscamos soluções para explorar um subconjunto deste universo. A partir dessas considerações, desenvolvemos os capítulos desta dissertação, descritos abaixo.

Nos três capítulos iniciais, expusemos o estudo da literatura recente dos processos de análise e síntese de sons musicais, assim como da ferramenta de engenharia envolvida, que são as transformadas *wavelets*. Com esta revisão, ampliamos o entendimento teórico sobre o assunto abordado e situamos os pontos teóricos para um possível desenvolvimento deste trabalho, cujo escopo foi a representação de sinais musicais.

No capítulo 4, apresentamos a descrição do modelo estudado, por duas vertentes diferentes. A primeira, foi implementada utilizando-se de aproximação polinomial, refinada com aplicação do método dos resíduos quadráticos. Descrita na seção 4.5.1, esta técnica permite recodificar sinais musicais de forma econômica, através da aproximação polinomial do conjunto de coeficientes do nível de resolução *wavalet* n. A segunda técnica desenvolvida baseou-se na permutação dos coeficientes de sons amostrados para gerar uma nova representação. Este método nos permitiu gerar sonoridades ricas, com alto grau de variabilidade e riqueza espectral, provenientes da fusão de vários elementos sonoros que foram representados pelos níveis de detalhes da transformada *wavelets*.

No decorrer da pesquisa, alguns problemas foram encontrados, como a alta dimensionalidade do espaço representado pelos coeficientes *wavelets*, a qual uma questão delicada que contornamos. Frente a este problema, observamos que trabalhar diretamente com as matrizes de permutação seria inviável computacionalmente. A solução adotada foi utilizar vetores de posição, com N entradas, associados às matrizes $N \times N$. Estes vetores descreveram a posição original ocupada por cada coeficiente do som alvo. Desta forma, reduzimos o número de operações realizadas e o consumo de memória.

Um outro problema encontrado, foi a notação matemática. Trabalhamos com três arquivos de áudio, analisados em oito níveis de resolução *wavelet* de onde resultam vetores de detalhes e de aproximação. Dentro do modelo, esses vetores foram manipulados através de ordenações e reordenações, pela permutação adequada, e tiveram seus coeficientes trocados utilizando-se do método de resíduos. Desta forma, necessitamos de vários descritores para as variáveis e índices. No modelo final, obtivemos uma notação simplificada a qual acreditamos conter o rigor matemático necessário ao escopo do problema.

É necessária a realização de mais testes psicoacústicos para avaliar a extensão dos resultados, com sujeitos músicos e também não treinados. Para sons gerais, que não são gerados por instrumentos musicais, poderemos explorar relações subjetivas. Por exemplo, aplicar a dinâmica de um som natural como o da água em um som de piano, avaliando como as características destes dois sinais são entrelaçadas no resultado sonoro.

Os resultados expostos no capítulo 5 apresentaram a funcionalidade do dispositivo. Dentre os quesitos desejáveis para um método de representação e síntese, os resultados podem ser comparados com os critérios anteriormente mencionados:

Qualidade sonora: as amostras criadas através de ambos os modelos, polinomial ou por sons amostrados, ficaram extremamente condizentes ao som original, preservando suas principais características sonoras. A fusão da bateria eletrônica com o baixo sintético do experimento da seção 5.3.3 demonstrou o potencial criativo do trabalho. Este julgamento psicoacústico poderá ser realizado pelo leitor através da comparação entre os sinais originais e o gerado pelo método. Além dos exemplos incluídos no texto da dissertação, há no cd em anexo à dissertação várias outras amostras, geradas a partir de sons e aproximações distintas.

Flexibilidade: a matriz P de permutação demonstra a capacidade de transferência de sonoridade por um processo de transformação linear. São operações simples, que permitem transitar por sonoridades distintas de forma flexível e imediata.

Generalidade: podemos representar qualquer sinal musical a partir de coeficientes aproximados. Estas aproximações podem ser feitas por um polinômio simples ou por qualquer outro sinal, com boa qualidade sonora. É uma conseqüência da flexibilidade obtida com as matrizes de permutação, que permite apontar para qualquer alvo sonoro, demonstrando a generalidade do processo de representação.

Tempo computacional: os protótipos foram implementados no ambiente *Matlab*. Os algoritmos de análise em multi-resolução são extremamente rápidos, com baixo custo de processamento.

A forma inédia pela qual esta pesquisa abordou o problema da representação de sinais musicais, mostrou-se robusta e capaz de suprir inúmeras necessidades apontadas no capítulo 2. Ressalta-se a possibilidade de recodificar sinais, transitar pelas características psicoacústicas de sons distintos, investigar a relação entre seus timbres e gerar material composicional novo.

As projeções para o desenvolvimento técnico deste trabalho sinalizam para engendrar, de forma mais profunda, as permutações obtidas no processo. Toda permutação está aliada a grupos cíclicos e, do ponto de vista matemático, são estruturas conhecidas e passíveis de um estudo mais aprofundado. A aplicação importante que resultaria desta investigação seria descobrir quais seriam os grupos cíclicos que mais se adaptariam à síntese desejada a partir de um modelo acústico. Além disso, a análise de outros tipos de sinais poderia ser considerada à luz do procedimento aqui apresentado.

Referências Bibliográficas

[Arfib, 1991] Arfib, D. (1991). Analysis, transformation, and resynthesis of musical sounds with the help of a time-frequency representation. In G. De Poli, A. Piccialli, & C. Roads (Eds.), *Representations of Musical Signals* chapter 3, (pp. 87–118). MIT Press.

[Barriere, 1994] Barriere, J. (1994). Le timbre: Metaphore pour la composition. *Computer Music Jornal*, 18(3), 79–82.

[Beltran & Beltran, 2003] Beltran, J. R. & Beltran, F. (2003). Additive synthesis based on the continuous wavelet transform: A sinusoidal plus transient model. In *Proc. of the 6th Int. Conference on Digital Audio Effects*.

[Bonada & Serra, 2007] Bonada, J. & Serra, X. (2007). Synthesis of the singing voice by performance sampling and spectral models. *IEEE SIGNAL PROCESSING MAGAZINE*, 24(2), 67–79.

[Chowning, 1971] Chowning, J. M. (1971). The synthesis of complex audio spectra by means of frequency modulation. *Journal of Audio Engineering Society*, 19(1), 2–6.

[Chui, 1997] Chui, C. K. (1997). *Wavelets: a mathematical tool for signal processing*. Society for Industrial and Applied Mathematics.

[Daubehies, 1992] Daubehies, I. (1992). *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics.

[De Poli, 1983] De Poli, G. (1983). A tutorial on digital sound synthesis techniques. *Computer Music Jornal*, 7(4), 8–25.

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

[Evangelista, 1991] Evangelista, G. (1991). *Wavelet transforms that we can play*. Cambridge, MA, USA: MIT Press.

[Evangelista, 1993] Evangelista, G. (1993). Pitch-synchronous wavelet representation of speech and musical signals. *IEEE Transactions on Signal Processing*, 41(12), 3313–3330.

[Evangelista, 1996] Evangelista, G. (1996). The coding gain of multiplexed wavelet transforms. *IEEE Transactions on Signal Processing*, 44(7), 1681–1692.

[Evangelista & Cavaliere, 1998a] Evangelista, G. & Cavaliere, S. (1998a). Discrete frequency warped wavelets: Theory and applications. *IEEE Transactions on Signal Processing*, 46(4), 874–885.

[Evangelista & Cavaliere, 1998b] Evangelista, G. & Cavaliere, S. (1998b). Frequency-warped filter banks and wavelet transforms: A discrete-time approach via laguerre expansion. *IEEE Transactions on Signal Processing*, 46(10), 2638–2650.

[Faria, 1997] Faria, R. R. A. (1997). A aplicação de *wavelets* na análise de gestos musicais em timbres de instrumentos acústicos tradicionais. Master's thesis, Universidade e São Paulo.

[Gabor, 1947a] Gabor, D. (1947a). Acoustical quanta and the theory of hearing. *Nature*, 159(1044), 591–594.

[Gabor, 1947b] Gabor, D. (1947b). Theory of communication. *Journal of Electrical Engineers Part III*, 93, 429–457.

[Grapes, 1995] Grapes, A. (1995). An introduction do wavelets. *IEEE Computacional Science and Engineering*, 2(2).

[Grey, 1977] Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of Acoustical Society of America*, 61(5), 1270–1277.

[Grey & Gordon, 1978] Grey, J. M. & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of Acoustical Society of America*, 63(5), 1493–1500.

[Hourdin et al., 1997] Hourdin, C., Charbonneau, G., & Moussa, T. (1997). A multidimensional scaling analysis of musical instruments' time-varying spectra. *Computer Music Jornal*.

[I. Testa et al., 2004] I. Testa, G. Evangelista, & S. Cavaliere (2004). Physically Inspired Models for the Synthesis of Stiff Strings with Dispersive Waveguides. *EURASIP Journal on Applied Signal Processing*, 2004(7), 964–977. special issue on Model-Based Sound Synthesis.

[Jawerth & Sweldens, 1994] Jawerth, B. & Sweldens, W. (1994). An overview of wavelet based multiresolution analysis. *SIAM Review*, 36(3), 377 –412.

[Kostek et al., 2005] Kostek, B., Szuczuko, P., & Dalka, P. Z. P. (2005). Processing of musical data employing rough sets and artificial neural networks. *Transactions on Rought Sets*, (pp. 112–133).

[Kostek et al., 2002] Kostek, B., Zwan, P., & Dziubinski, M. (2002). Statistical analysis of musical sound features derived from wavelet representation. In *112Th AES Convention - Munich*.

[Kroland-Martinet, 1988] Kroland-Martinet, R. (1988). The wavelet transform for analysis, synthesis, and processing of speech and music sounds. *Computer Music Jornal*, 12(4), 11–20.

[Lebrun, 1979] Lebrun, M. (1979). Digital waveshaping synthesis. *Journal of Audio Engineering Societyu*, 27(4), 250–266.

[Loureiro et al., 2004] Loureiro, M. A., De Paula, H. B., & Yehia, H. C. (2004). Timbre classification of a single musical instrument. In *5Th International Conference on Music Information Retrieval*.

[Luvizotto & Costa, 2007] Luvizotto, A. L. & Costa, C. R. (2007). Context sensitive harmonic processor. In *In Proc. of 11th Brazilian Symposium on Computer Music*.

[Luvizotto et al., 2007] Luvizotto, A. L., Costa, C. R., Manzolli, J., Mendes, R. S., & Von Zuben, J. F. (2007). Harmonic equalizer. In 11^a *Convenção Nacional da AES Brasil*.

[Mallat, 1998] Mallat, S. (1998). *A Wavelet Tour of Signal Processing*. Academic Press.

[Mallat, 1989] Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7).

[Mathews, 1963] Mathews, M. V. (1963). The digital computer as a musical instrument. *Science*, 142(No.3592), 553–557.

[Meyer, 1993] Meyer, Y. (1993). *Wavelets: algorithms and applications.* Society for Industrial and Applied Mathematics.

[Miner & Caudell, 2002] Miner, N. E. & Caudell, T. P. (2002). A wavelet synthesis technique for creating realistic virtual enviroment sounds. *Presence*, 11(5), 493–507.

[Oppenheim et al., 1999] Oppenheim, A. V., Schafer, R. W., & Buck, J. R. (1999). *Discrete-Time Signal Processing*. Prentice-Hall, 2nd ed edition.

[Polotti & Evangelista, 2007] Polotti, P. & Evangelista, G. (2007). Fractal additive synthesis: A deterministic/stochastic model for sound synthesis by analysis. *IEEE Signal Proc. Mag.*, 24(2), 105–115.

[Portnoff, 1980] Portnoff, M. R. (1980). Time-frequency representation of digital signals and systems based on short-time fourier transform. *IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND AUDIO PROCESSING*.

[Risset, 1965] Risset, J. C. (1965). Computer study of trumpet tones. In *Seventieth Meeting of the Acoustical Society of America Seventieth Meeting of the Acoustical Society of America*.

[Roads, 1978] Roads, C. (1978). Automated granular synthesis of sound. *Computer Music Jornal*, 2(2), 61–62.

[Sandell, 1995] Sandell, G. J. (1995). Roles for spectral centroid and other factors in determining ”blended” instrument pairings in orchestration. *Music Perception*, 13(2), 209–246.

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

[Serra & Smith, 1989] Serra, X. & Smith, J. (1989). Spectral modeling synthesis: A sound analysis / synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Jornal*.

[Spiegelberg, 2002] Spiegelberg, S. C. (2002). *The Psychoacoustics of Musical Articulation*. Phd, Eastman School of Music.

[Strömberg, 1981] Strömberg, J. O. (1981). A modified franklin system and higher order spline systems on R^n as uncondicional bases for hardy spaces. In U. of Chicago (Ed.), *Conference on Harmonic Analysis in Honeor of Antoni Zygmund, B et al.*, volume 2 (pp. 475–494).

[Su & Jeng, 2001] Su, B. & Jeng, S. (2001). Multi-timbre chord classification using wavelet transform and self-organized map neural networks. In I. International (Ed.), *Conference on Acoustics, Speech, and Signal Processing*, volume 5 (pp. 3377 –3380).

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

[Truax, 1996] Truax, B. (1996). Soundscape, acoustic communication and environmental sound composition. *Contemporary Music Review*, 15(1), 44–65.

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

Índice Remissivo de Autores

Apêndice A

Matriz de Permutação

Para uma melhor compreensão da descrição do modelo, faz-se necessário a apresentação de dois teoremas e um corolário que formaliza o conceito de permutação. Vamos à eles.

Definição A.0.1 *Seja $v = (v_1, v_2, ..., v_N)$ de N entradas, então existe uma matriz quadrada $N \times N$ (denominada Matriz de Transposição) $T^{(i,k)}$, tal que:*

$$T^{(i,k)} v = w \tag{A.1}$$

onde

$$\begin{aligned} w_i &= v_j \\ w_j &= v_i \\ e \quad w_k = v_k, &\quad se\ k \neq i, j \end{aligned} \tag{A.2}$$

Teorema A.0.2

$$(T^{(i,k)})^2 = Id \tag{A.3}$$

ou seja,

$$(T^{(i,k)})^{-1} = T^{(i,k)} \tag{A.4}$$

Teorema A.0.3 *Toda matriz de permutação $N \times N$ pode ser fatorada em uma seqüência finita de matrizes de transposição.*

Como conseqüência dos teoremas acima, temos o seguinte corolário.

Em particular, pelos teoremas acima, podemos ordenar um vetor qualquer através de uma permutação. Operacionalmente, pelo teorema A.0.3 podemos obter uma ordenação qualquer aplicando uma seqüência finita de matrizes de transposição. Por exemplo:

$$\begin{pmatrix} 4 & 3 & 2 & 1 \end{pmatrix} \overset{T1}{\Rightarrow}$$
$$\begin{pmatrix} 3 & 4 & 2 & 1 \end{pmatrix} \overset{T2}{\Rightarrow}$$
$$\begin{pmatrix} 3 & 2 & 4 & 1 \end{pmatrix} \overset{T3}{\Rightarrow}$$
$$\begin{pmatrix} 3 & 2 & 1 & 4 \end{pmatrix} \overset{T4}{\Rightarrow}$$
$$\begin{pmatrix} 2 & 3 & 1 & 4 \end{pmatrix} \overset{T5}{\Rightarrow}$$
$$\begin{pmatrix} 2 & 1 & 3 & 4 \end{pmatrix} \overset{T6}{\Rightarrow}$$
$$\begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix}$$

Apêndice B

Organização das Faixas do CD que Acompanha o Texto

B.1 Faixas

As faixas do CD de áudio estão organizadas de acordo com a ordem na qual os exemplos são apresentados no texto da dissertação. Vamos a elas.

B.1.1 Representação Através do Modelo por Aproximação Polinomial

1. D3 Clarineta - Original

2. D3 Clarineta - Permutado

3. C4 Piano - Original

4. C4 Piano - Permutado

5. *Drum'n Bass* - Original

6. *Drum'n Bass* - Permutado

B.1.2 Representação Através do Modelo por Coeficientes de Sons Amostradol

7. D3 Clarineta - Original
8. G#3 Clarineta - Original
9. D3 Clarineta - Permutado
10. D3 Clarineta - Permutado e com 4000 coeficientes trocados pelo som alvo D3.
11. E4 Clarineta - Original
12. G#3 Clarineta - Original
13. E4 Clarineta - Permutado
14. E4 Clarineta - Permutado e com 4000 coeficientes trocados pelo som alvo D3 da Clarineta.
15. C4 Piano - Original
16. G#3 Clarineta - Original
17. C4 Piano - Permutado
18. C4 Piano - Permutado e com 4000 coeficientes trocados pelo som alvo C4 do Piano.
19. *Drum'n Bass* - Original
20. G#3 Clarineta - Original
21. *Drum'n Bass* - Permutado
22. *Drum'n Bass* - Permutado e com 12000 coeficientes trocados pelo som de Baixo Sintético.

Para acessar o conteúdo relativo aos dados, abra a pasta "AmostrasTese". Lá estão todos os exemplos, com os áudios de cada nível de decomposição de todos os exemplos que constam no texto, figuras e tabelas em formato LaTeX.