TÉCNICAS DE ANÁLISE ESPECTRAL DE LINHAS MUSICAIS

Iúri Kothe

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Aprovada por:

Prof. Luiz Wagner Pereira Biscainho, D.Sc.

Prof. Sergio Lima Netto, Ph.D.

Prof. Marcio Nogueira de Souza, D.Sc.

Prof. Jacques Szczupak, Ph.D.

RIO DE JANEIRO, RJ - BRASIL
JUNHO DE 2006

KOTHE, IÚRI

Técnicas de Análise Espectral de Linhas Musicais [Rio de Janeiro] 2006

XV, 112 p. 29,7 cm (COPPE/UFRJ, M.Sc., Engenharia Elétrica, 2006)

Dissertação - Universidade Federal do Rio de Janeiro, COPPE

1.Processamento de Sinais

2.Processamento de Audio

3.Análise Espectral

I.COPPE/UFRJ II.Título (série)

Agradecimentos

Agradeço

a toda a minha família, em especial, a minha mãe e minha irmã;

ao meu amigo e orientador LW;

a todos dos grupos LPS e GPA, principalmente Dini, Lonnes, Schmalter e Cristiano;

aos professores Sergio Lima Netto, Calôba, Eduardo e Diniz;

ao pessoal do Portal do Voluntário, Sohaclara, Comunitas, CG Total, Namers, Derrota, CAPES e CNPq;

à minha "família" no Rio: Tigrão, Tocha, Folopo, Inez, Nathalie, Baiano e Serginho;

aos meus amigos de Brasília;

a Villa-Lobos, Bobby McFerrin, Simeon Ten Holt, L Subramaniam, Fela Kuti, Naná Vasconcelos, Glen Velez, Uakti, Bach, Beethoven, Autechre, Murcof, Marlui Miranda e Mambazo;

a todos que acreditam que arte e ciência não são produtos e devem ser livremente compartilhados.

"As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality."

"The significant problems we face cannot be solved at the same level of thinking we were when we created them."

Albert Einstein (1879 - 1955)

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

TÉCNICAS DE ANÁLISE ESPECTRAL DE LINHAS MUSICAIS

Iúri Kothe

Junho/2006

Orientador: Luiz Wagner Pereira Biscainho

Programa: Engenharia Elétrica

A análise espectral de sinais de áudio deve levar em conta as características particulares desse tipo de sinal, que podem envolver desde as propriedades dos instrumentos musicais gravados até a escala musical em que a execução foi realizada. As ferramentas mais populares de análise espectral, como a FFT, nem sempre são as mais indicadas para essa tarefa, sob os pontos de vista da resolução espectral e da distribuição dos canais na freqüência. Este trabalho se dedica a examinar alternativas para a análise espectral de sinais de áudio. São abordadas soluções na forma de transformadas de blocos, como ferramentas de refinamento da DFT; e soluções na forma de bancos de filtros com elevada seletividade e distribuição de canais variável com a freqüência. Os resultados encontram aplicação em equalização digital, transcrição musical automática, reconhecimento de instrumentos e separação de fontes sonoras, entre outras.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

TECHNIQUES FOR AUDIO SPECTRAL ANALYSIS

Iúri Kothe

June/2006

Advisor: Luiz Wagner Pereira Biscainho

Department: Electrical Engineering

Spectral analysis of audio signals must take into account their special characteristics, which may include from the properties of the recorded musical instruments to the musical scale employed in the performance. Popular spectral analysis tools like the FFT are not always the best choice for this job, regardng frequency resolution as well as channel distribution. This work aims at studying alternative methods for the spectral analysis of audio signals. Block-transform methods are approached, in the form of refining tools for the DFT. Additionally, highly selective filter banks with variable channel distribution along the spectrum are examined. The results of the work find application in digital equalization, automatic music transcription, instrument identification, sound source separation etc.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

A análise espectral evoluiu bastante desde sua concepção, utilizando as bases criadas por Joseph Fourier no início do século XIX. A possibilidade de realizar a transformada de Fourier de forma eficiente com o advento dos computadores pela *Fast Fourier Transform* (FFT) tornou-a extremamente popular e atrativa, sendo utilizada para analisar diferentes tipos de sinais.

A análise espectral pode ser entendida como a divisão do sinal em diversas faixas de freqüência e subseqüente descrição de conteúdo. Esse procedimento também pode ser visto como uma filtragem, que é a separação de parte do sinal com determinada característica que se deseja observar.

No caso dos sinais de áudio, para se realizar uma análise espectral adequadamente, deve-se observar as características particulares deste tipo de sinal, que vão desde as propriedades dos intrumentos musicais gravados até a escala musical envolvida na execução. Sob diversos aspectos, que vão da resolução freqüencial à distribuição das faixas de freqüência descritas, a popular transformada de Fourier não é a ferramenta ideal de análise. Este trabalho se dedica a examinar alternativas para análise espectral no contexto de sinais de áudio. São abordadas soluções na forma de transformada de blocos bem como na forma de bancos de filtros.

Como aplicações para os resultados deste trabalho, podem ser citadas: equalização de áudio; transcrição musical automática; reconhecimento de instrumentos musicais; e separação de fontes sonoras;

1.1 Estrutura da Tese

O Capítulo 2 apresenta alguns conceitos associados aos "sons musicais" e à música, como harmônicas, escalas, intervalos e estruturas musicais, além de propriedades do ouvido humano e características do som. Esses conceitos são essenciais para se obter uma ferramenta de análise espectral adequada a sinais de áudio.

Os capítulos seguintes foram agrupados em duas partes, procurando mostrar, na primeira parte, soluções para análise espectral de sinais de áudio por transformada de blocos e na segunda, por bancos de filtros.

Os Capítulos 3, 4 e 5 formam o conjunto sobre Transformada de Blocos. O Capítulo 3 apresenta técnicas de representação do sinal pelo modelo senoidal, que utiliza a transformada de Fourier para descrever o comportamento das linhas musicais ao longo do tempo e que podem ser utilizadas para posterior síntese, que será explicada no Capítulo 5. No Capítulo 4 são apresentadas técnicas de refinamento espectral utilizando a transformada de Fourier baseadas na freqüência instantânea, mostrando-se exemplos comparativos.

A parte de Soluções por Bancos de Filtros é formada pelos Capítulos 6, 7 e 8. O Capítulo 6 apresenta técnicas baseadas em banco de filtros muito seletivos para análise dinâmica de sinais musicais divididos em subgrupos de acordo com a separação espectral linear, geométrica ou linear por oitava. Exemplos comparativos são apresentados no Capítulo 7. Algumas propostas para sintetizar as linhas obtidas pelos métodos de banco de filtros são descritas no Capítulo 8.

O Capítulo 9 conclui analisando a contribuição desta tese e apresenta propostas para sua continuação.

Capítulo 2

Música

Neste capítulo são abordados alguns tópicos básicos de percepção e conceitos de música que são de fundamental importância para se entender como deve ser feita a análise em tempo-freqüência das linhas musicais, que será descrita nos próximos capítulos. Primeiro, na Seção 2.1, compara-se o ouvido humano com um banco de filtros; na Seção 2.2 há uma breve introdução às propriedades do som; a formação das notas pelas harmônicas é descrita na Seção 2.3; em 2.4 são abordadas as escalas e os intervalos musicais; a escala de igual temperamento é descrita na Seção 2.5; e, por fim, a estrutura musical fecha o capítulo na Seção 2.6.

2.1 Ouvido Humano como Banco de Filtros

A membrana basilar localizada dentro da cóclea pode ser associada a um banco de filtros[1] distribuídos ao longo do espectro em escala aproximadamente logarítmica. Assim, à medida que se aumenta a freqüência central do filtro, em progressão geométrica, sua largura de banda (chamada de banda crítica) também aumenta na mesma proporção [1]; Abaixo de 200Hz, porém, a largura de banda dos filtros é constante, em torno de 50Hz. A Figura 2.1 mostra a relação entre a freqüência central de cada ponto da membrana basilar e a largura de banda desse respectivo filtro em escala log-log. Vale ressaltar que a escala de igual temperamento, descrita na Seção 2.5, também segue um padrão logarítmico no espectro.

[1]Bancos de filtros são abordados no capítulo 6.

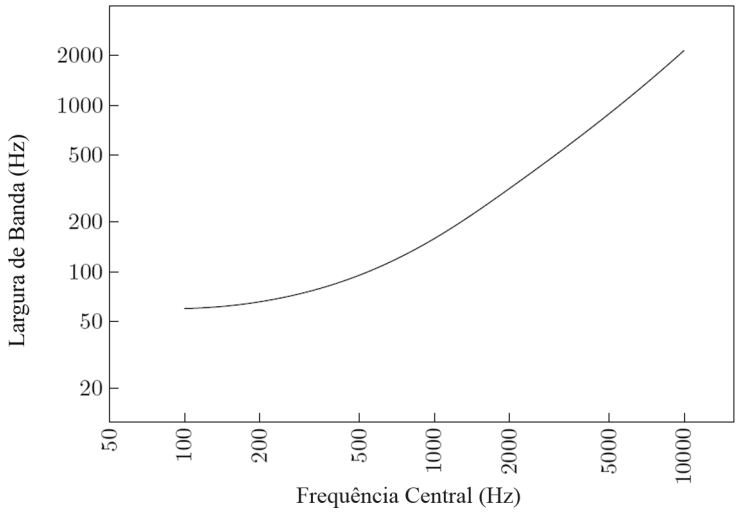

Figura 2.1: Bandas críticas do ouvido humano.

2.2 Propriedades do Som

O som pode ser analisado sob quatro aspectos diferentes, que podem ser observados tanto pelas propriedades físicas como perceptivas [1]. A Tabela 2.1 mostra essas propriedades.

Tabela 2.1: Propriedades do som separadas pelos aspectos físicos e suas relações perceptivas.

Físico	**Percepção**
Amplitude	Intensidade
Freqüência	*Pitch*
Espectro	Timbre
Duração	Tamanho

A amplitude do som diz respeito ao tamanho da vibração sonora propagada, e é percebida como intensidade sonora.

O *pitch* está relacionado com a altura do som (que acaba por se confundir com a nota musical) percebida. Há dois modelos principais que tentam explicar o *pitch*. Segundo o primeiro modelo, chamado *do espaço*, a membrana basilar na cóclea é preferencialmente excitada em regiões associadas biunivocamente com as componentes freqüenciais presentes no sinal; aumentando-se a freqüência geometri-

camente, percorrem-se distâncias aproximadamente iguais na membrana. O segundo modelo, chamado *do tempo*, os disparos elétricos associados à excitação da membrana basilar tendem a se sincronizar com o período do sinal excitante. Nenhum dos dois modelos explica completamente a percepção de *pitch*, principalmente nos casos de *pitch* múltiplo. É muito comum, contudo, associar *pitch* à freqüência fundamental da nota emitida, o que é uma extrema simplificação —já que até mesmo na ausência da fundamental o *pitch* pode ser reconhecido pela distância entre os harmônicos.

O timbre, apelidado de cor do som, é a sensação percebida pela envoltória espectral. Cada fonte sonora possui uma característica específica de formação do espectro, o que possibilita ao ser humano distinguir notas iguais tocadas por instrumentos diferentes. Vale ressaltar que esse espectro característico também depende da intensidade das notas tocadas. Por isso, reconhecer a fonte pela análise espectral não é uma tarefa simples.

Já a duração é a característica temporal do som que é percebida como o início e fim, ou seja, o tamanho de cada nota.

2.3 Harmônicas

Joseph Fourier, no século XIX, descobriu uma propriedade muito interessante da natureza que pode ser aplicada à análise de vibrações periódicas[2]. Ele demonstrou que sinais periódicos, isto é, que se repetem ao longo do tempo, podem ser decompostos em uma soma infinita de senóides com mesma freqüência de vibração (onde freqüência é o inverso do período, i.e., $f_0 = 1/T$) e múltiplas inteiras da mesma. Essas múltiplas da freqüência fundamental f_0 são denominadas harmônicas, podendo ser definidas como $f_n = nf_0$, com $n = 1, 2, 3....$ Isto pode ser observado, por exemplo, pela vibração de uma corda de violão. Como ela é fixa nas duas extremidades, ao tocá-la a amplitude será nula nas extremidades e nas frações inteiras da sua distância que correspondem justamente às harmônicas, como 1/2; 1/3 e 2/3; 1/4, 2/4 e 3/4 etc.

[2]A parte matemática da análise de Fourier é descrita no item 4.3.

2.4 Escalas e Intervalos Musicais

Existem vários tipos de escalas e possíveis afinações para as notas. A mais utilizada, na atualidade, é a escala de igual temperamento, descrita na Seção 2.5. Para escalas de 12 notas, são especificadas três regras básicas de formação:

- Um conjunto de 12 notas é agrupado por, e repetido a cada oitava;
- Existe, dentro desse conjunto, um grupo de sete notas principais, onde duas estão espaçadas por um pequeno incremento (semitom) e cinco por um incremento maior (tons inteiros). Na Tabela 2.2 é possível observar esse subconjunto de 7 notas e onde ocorrem esses dois tipos de incrementos;
- Os intervalos são escolhidos para serem afinados de modo suave e não ocorrerem batimentos[3] . Esse modo suave de afinação é obtido pela busca da maior consonância entre as notas, isto é, obtendo-se um maior número de harmônicas coincidentes.

Tabela 2.2: Intervalos de tons e semitons em uma oitava, para escalas de 12 notas.

Tons inteiros:	I	-	II	-	III	IV	-	V	-	VI	-	VII	(I)
Semitons:	x	x	x	x	x	x	x	x	x	x	x	x	(x)

Na Figura 2.2 é possível observar a relação entre as freqüências harmônicas e seus respectivos valores como notas musicais para a nota de freqüência fundamental C3[4]. Como uma oitava é um intervalo entre duas notas no qual uma tem o dobro da freqüência da outra, as oitavas superiores (15^a, 22^a, 29^a etc) estarão sempre distantes de uma potência de 2. Isso mostra que, coerentemente com o ouvido humano, o espectro do sistema musical ocidental também segue uma distribuição geométrica. Aqui já pode ser percebida a dificuldade que há na análise de áudio causada pela sobreposição de harmônicos, principalmente no caso de oitavas.

[3]O batimento é um fenômeno causado ao se tocar duas freqüências muito próximas Ω_1 e Ω_2: será percebida uma freqüência bastante baixa, que é a semi-diferença entre as duas freqüências, modulando a semi-soma delas, i.e., $\cos(\Omega_1 t) + \cos(\Omega_2 t) = 2\cos(\frac{\Omega_1+\Omega_2}{2}t)\cos(\frac{\Omega_1-\Omega_2}{2}t)$.

[4]Será utilizada a notação musical americana, onde as notas Dó, Ré, Mi, Fá, Sol, Lá, Si são representadas, respectivamente, por C, D, E, F, G, A, B.

A partir da formação das harmônicas é possivel analisar a razão freqüencial entre os intervalos musicais da chamada escala natural. Alguns intervalos são apresentados na Tabela 2.3. A terceira harmônica, que no exemplo da Figura 2.2 é um Sol4, está justamente no meio freqüencial das duas oitavas (harmônicas 2 e 4). Sua prorpoção freqüencial é de 3:2 em relação à fundamental na escala natural, onde a 3^a harmônica da nota mais grave coincide com a 2^a harmônica da mais aguda. Esse intervalo é denominado de quinta justa. Quanto menor for o mínimo múltiplo comum do intervalo, mais consonante ele será, já que possui um maior número de harmônicos coincidentes.

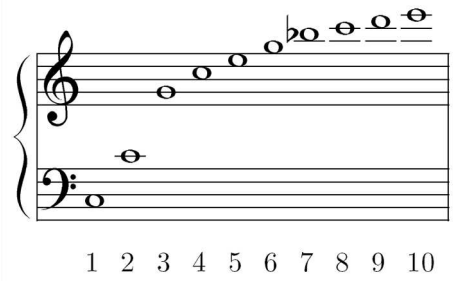

Figura 2.2: C3 e suas freqüências harmônicas.

Descendo uma quinta justa (3:2) a partir de uma oitava de razão 2:1, obtém-se a quarta justa de razão 4:3, como pode ser visto na Figura 2.3. Subtraindo uma quinta justa (3:2) de uma quarta justa (4:3), tem-se um tom inteiro, de razão 9:8. O problema é que não é possível garantir essas afinações indefinidamente para todas as notas em um mesmo sistema, pois ao extrapolar as notas para as oitavas superiores e inferiores, tende a ocorrer uma pequena diferença de afinação, denominada de coma pitagórica, que pode ser vista na Figura 2.3 como a pequena diferença entre $A\flat$ e $G\sharp$. Mais detalhes podem ser encontrados em [2] e [1].

As escalas cujas notas são criadas a partir de intervalos racionais, tendo uma nota de referência, possuem duas características que dificultam a vida do músico:

- Notas enarmônicas com freqüências diferentes, como, por exemplo, $A4\sharp$ e $B4\flat$;
- Semitons de tamanhos diferentes, levando a intervalos diferentes com mesmo nome, conforme as notas envolvidas.

Isso faz com que, dado um instrumento de afinação fixa afinado segundo determinada tonalidade, as músicas não possam ser transpostas livremente. Por

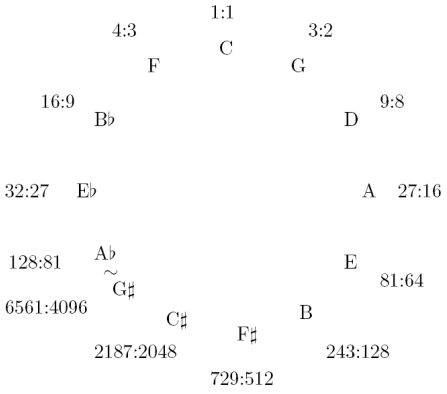

Figura 2.3: Ciclo das quintas, partindo do C.

exemplo, uma peça composta em $F\sharp$ maior, ao ser tocada num instrumento afinado a partir de C maior soaria dissonante e/ou desafinada, justamente porque as coincidências das harmônicas das notas não são preservadas. A solução para isso foi adotar uma razão geométrica fixa entre quaisquer semitons, obtida com a escala de igual temperamento, descrita a seguir.

Tabela 2.3: Comparação, entre as escalas natural e igualmente temperada, das razões de freqüências presentes em alguns intervalos.

Intervalo	Escala natural	Escala igualm. temperada
Segunda maior	$9/8 = 1{,}125$	$2^{2/12} \approx 1{,}122$
Terça maior	$5/4 = 1{,}25$	$2^{4/12} \approx 1{,}260$
Quarta justa	$4/3 \approx 1{,}333$	$2^{5/12} \approx 1{,}335$
Quinta justa	$3/2 = 1{,}5$	$2^{7/12} \approx 1{,}498$
Sexta maior	$5/3 \approx 1{,}667$	$2^{9/12} \approx 1{,}682$
Sétima maior	$15/8 = 1{,}875$	$2^{11/12} \approx 1{,}888$
Oitava justa	2	$2^{12/12} = 2$

2.5 Escala Igualmente Temperada

A escala igualmente temperada, como o próprio nome já diz, tem como princípio distribuir igualmente, em freqüência, todas as 12 notas. Como a oitava é o dobro da freqüência, a escala é dividida em progressão geométrica com um semitom sendo um incremento freqüencial de $2^{1/12}$, isto é, $f_{s+1} = 2^{1/12} f_s$, sendo s o

índice do semitom. Na Figura 2.4 estão apresentadas as notas da escala igualmente temperada, suas freqüências e a extensão de alguns instrumentos musicais e da voz humana.

Essa escala é a mais utilizada atualmente em instrumentos de afinação fixa, como piano e clarinete. Com isso, notas que em outras escalas e instrumentos podem ser diferentes, tais como as enarmônicas $A4\flat$ e $G4\sharp$, na escala igualmente temperada representam a mesma nota. Essa padronização também possibilita a transposição, pois os intervalos preservam sempre a mesma razão freqüencial entre as notas.

O preço a se pagar no igual temperamento é que, excetuando-se a oitava, os intervalos (como, por exemplo, a quinta justa) não possuem mais harmônicos coincidentes, já que eles deixaram de apresentar razões inteiras entre suas notas extremas. Em outras palavras, a escala ficou toda "igualmente dissonante". A relação freqüencial de alguns intervalos musicais é apresentada na Tabela 2.3.

2.6 Estrutura Musical

A estrutura musical possui três características principais: melodia, ritmo e harmonia. A melodia é a sucessão de notas predominante de uma música, ou seja, a voz principal que se destaca em relação a um possível acompanhamento (harmonia). Pela Tabela 2.1 pode-se observar que a melodia está relacionada ao *pitch*. O ritmo se refere ao início, à duração e à intensidade dos sons, o que diz respeito mais à informação temporal. Por fim, a harmonia é um conjunto de notas tocadas em função da melodia, acompanhando-a como elemento enriquecedor da música por meio de acordes, que são combinações de notas simultâneas.

Outro tipo de classificação é quanto à tessitura, que pode ser: mono-, homo- ou polifônica. A monofônica é constituída de apenas uma linha melódica, sendo que o instrumento toca apenas uma nota por vez. A homofônica é uma linha melódica acompanhada de uma harmonia. E a polifônica possui duas ou mais melodias sendo tocadas ao mesmo tempo (com ou sem harmonia).

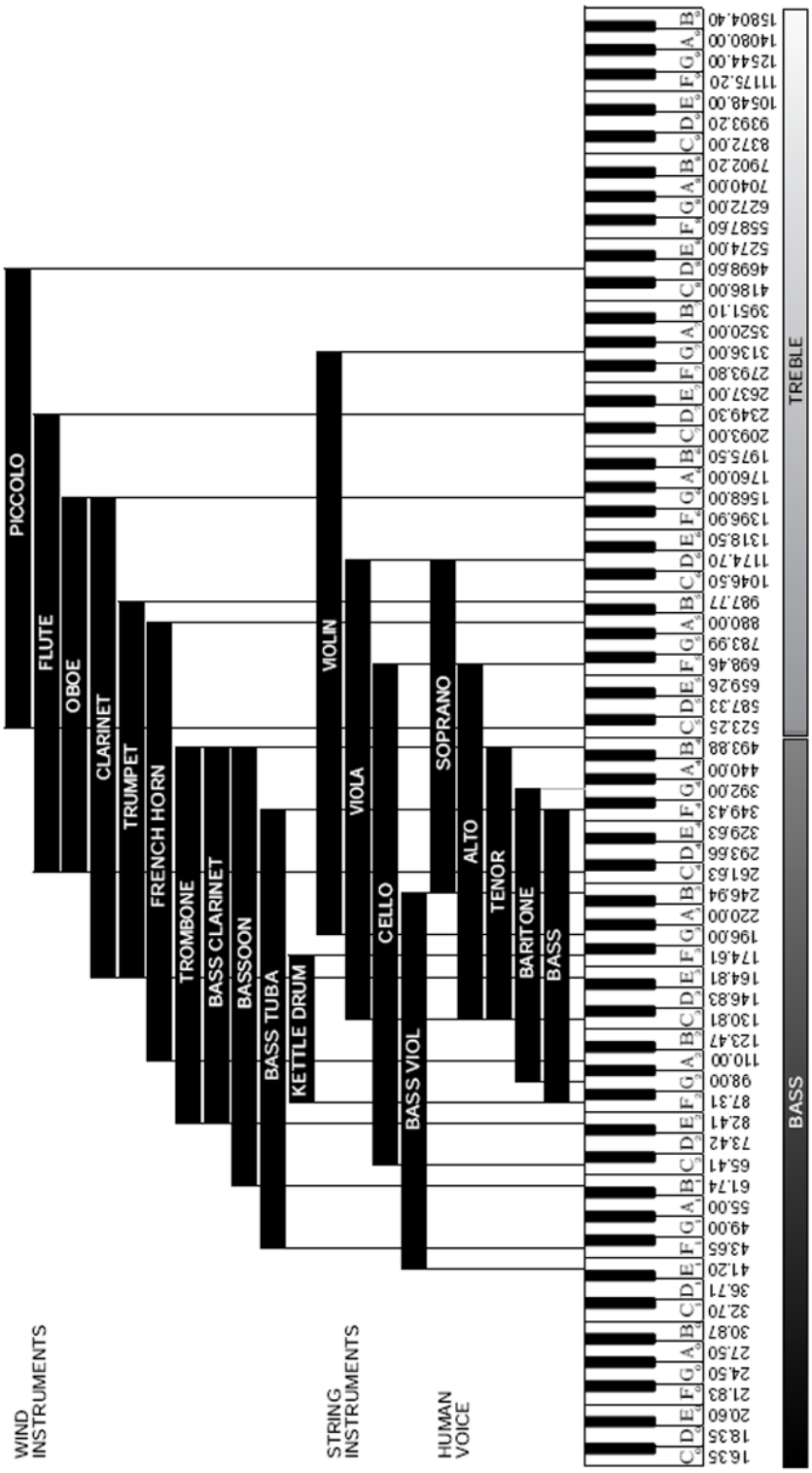

Figura 2.4: Escala igualmente temperada.

Parte I

Soluções por Transformada de Blocos

Capítulo 3

Modelagem Senoidal

A Seção 3.1 apresenta um breve histórico da modelagem senoidal para sinais de áudio e voz, que iniciou com o *Phase Vocoder* (Subseção 3.1.1) e evoluiu para os modelos McAulay-Quatieri (ou MQ, na Subseção 3.1.2) e *Spectral Modeling Synthesis* (ou SMS, na Subseção 3.1.3). A Seção 3.2 descreve a formação das trilhas freqüenciais no tempo e finaliza com um método que utiliza informação da partitura para gerar as trilhas (Subseção 3.2.1).

3.1 Modelagem do Sinal

A modelagem senoidal é o método mais utilizado para análise e síntese de sinais de áudio "tonais" (tais como os de voz e instrumentos musicais), pois estes apresentam picos espectrais bastante evidentes, correspondentes às suas harmônicas ou simplesmente às parciais de sinais não-harmônicos (como, por exemplo, no som de um sino). Outra característica importante dos sinais "tonais" é terem um decaimento lento no tempo, possibilitando descrever o sinal em janelas curtas como sendo uma soma finita de oscilações com freqüências, amplitudes e fases iniciais fixas por janela. Essa forma de modelagem favorece a utilização dos parâmetros obtidos ao se analisar um sinal pela Transformada de Fourier de Curta Duração.

Historicamente essa representação começou com o *Phase Vocoder* [3] em 1966, sendo generalizada por Quatieri e McAulay [4] em 1986 com o modelo MQ, e logo depois, em 1987, por Serra e Smith com o modelo SMS [5]. Esses artigos são as bases do modelagem senoidal e serão brevemente explicados a seguir.

Após os modelos MQ e SMS, vários refinamentos foram propostos, tais como: utilização da FFT inversa (FFT^{-1}) para sintetizar o sinal [6]; uso de Formas de Onda Elementares [7]; método para análise/síntese de sons residuais baseados no modelo auditivo [8]; uso de wavelets para refinar a análise freqüencial do modelo senoidal [9]; uso de um banco de filtros não-uniforme para modelar a parte estocástica [10]; um novo método para o *Phase Vocoder* [11]; uma forma de análise/síntese de sinais ruidosos utilizando senóides de curta duração proposto em [12]; síntese aditiva fractal [13], que visa a descrever também sinais não-harmônicos, como percussão, utilizando a transformada de *wavelets*; síntese aditiva em tempo real usando técnicas lineares e bilineares [14]; método baseado em redes bayesianas dinâmicas [15] utilizado para transcrição de vários intrumentos musicais tocados simultaneamente, porém cada um emitindo uma só nota por vez.

Como a modelagem de sinais de áudio passou a ser descrita por várias técnicas diferentes, Xavier Rodet e Adrian Freed, na década de 90 observaram a necessidade de padronizar os arquivos para análise, processamento e síntese desses sinais. Juntamente com outros colaboradores, criaram, então, o SDIF - *Sound Description Interchange Format*. Esse assunto foge do escopo dessa tese; para mais informações, ver [16] e [17].

3.1.1 Phase Vocoder

O *Phase Vocoder*[1] de Flanagan e Golden [3] foi uma ferramenta criada para sintetizar voz utilizando informações de freqüência, amplitude e fase do sinal em janelas de curta duração. A Transformada de Fourier de Curta Duração, ou STFT ([18] e [19]), de um sinal $x[n]$ janelado pela seqüência $w[n]$ é calculada como

$$\begin{aligned} X[n,\Omega] &= \sum_m w[n-m]x[m]e^{-j\Omega m} \\ &= e^{-j\Omega n}(x * w_{\text{mod}})[n], \end{aligned} \tag{3.1}$$

onde $w_{\text{mod}}[n] = w[n]e^{j\Omega n}$ e $\Omega \in [-\pi, \pi]$ é a freqüência angular.

Pode-se observar a equação (3.1) como sendo uma modulação da janela $w[n]$ para a freqüência Ω (isso é feito com $e^{j\Omega n}$), que filtra o sinal $x[n]$ na forma do filtro

[1]Vocoder é uma abreviação de *Voice Coder*, ou codificador de voz.

passa-banda $w[n]e^{j\Omega n}$, seguindo-se a demodulação do sinal para a banda-base com $e^{-j\Omega n}$. A Figura 3.1 ilustra o diagrama de análise/síntese. A saída desse banco de filtros no domínio do tempo pode ser então vista como senóides discretas moduladas em amplitude e fase pela Transformada de Fourier de Curta Duração (STFT).

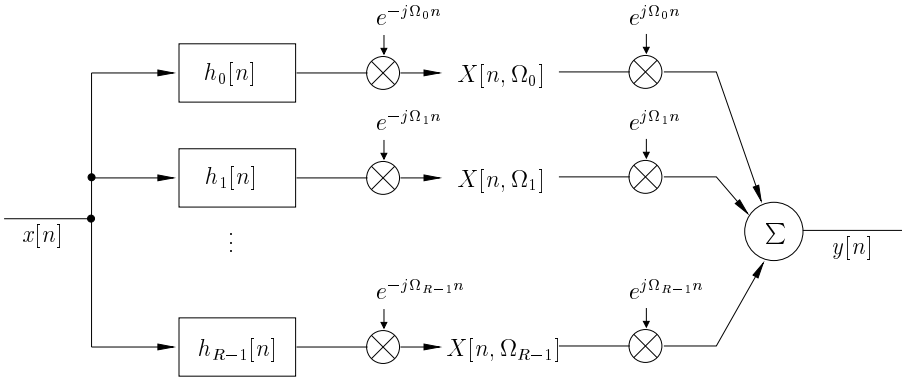

Figura 3.1: Diagrama do *Phase Vocoder*

Particularizando os filtros do banco a serem cópias de um filtro-protótipo $h[n] = w[n]$ moduladas por cossenos harmônicos, pode-se formular cada filtro h_k da seguinte maneira:

$$h_k[n] = w[n]e^{j(2\pi/R)kn}. \tag{3.2}$$

O espaçamento entre os filtros é de $2\pi/R$, onde R é a quantidade de filtros do banco. Assim, tem-se que, para um sinal de entrada $x[n]$, a resposta de um filtro h_k será:

$$\begin{aligned} y_k[n] &= (x * h_k)[n] \\ &= e^{j\Omega_k n} X[n,\Omega_k], \end{aligned} \tag{3.3}$$

que é a equação (3.1) sem a demodulação final para banda-base e com freqüência $\Omega_k = (2\pi/R)k$, que é o centro do filtro h_k.

Como os filtros $h_k[n]$ são complexos, suas respectivas saídas também o serão, com amplitude $a_k[n]$ e fase $\theta_k[n]$ no domínio do tempo (discreto) dadas por:

$$\begin{aligned} a_k[n] &= \mid y_k[n] \mid \\ \theta_k[n] &= \text{arctg}\frac{\text{Im}\,\{y_k[n]\}}{\text{Re}\,\{y_k[n]\}}. \end{aligned} \tag{3.4}$$

Dessa forma, pode-se observar a saída de cada filtro como uma exponencial complexa modulada em amplitude e fase, i.e.

$$y_k[n] = a_k[n]e^{j\theta_k[n]}. \tag{3.5}$$

Por fim, a reconstrução do sinal original $x[n]$ é feita na etapa de síntese, onde são utilizados os parâmetros de amplitude e fase obtidos por cada filtro

$$x[n] = \sum_k a_k[n]e^{j\theta_k[n]}. \tag{3.6}$$

Uma aplicação do *Phase Vocoder* é a expansão/compressão do sinal, originalmente a voz, no tempo sem alterar o *pitch*. Para isto, logo após a etapa de análise, modifica-se o parâmetro temporal, preservando-se o valor das freqüências. As janelas são então sobrepostas e somadas (OLA, do inglês *Overlap-and-Add*) sintetizando-se o sinal modificado. Maiores informações sobre essa aplicação podem ser encontradas em [20], [21] e [22].

Os principais problemas do Phase Vocoder são o de modelar apenas sinais harmônicos da freqüência Ω_1 e não modelar bem sinais com freqüência variável ao longo do tempo, porque a oscilação passa de um filtro para outro. Tendo em vista superar essas dificuldades, foi proposto o método da modelagem senoidal de McAulay e Quatieri, explicado a seguir.

3.1.2 Modelo MQ

Em 1986, McAulay e Quatieri desenvolveram a base da modelagem senoidal (descrita em [4]) fazendo uma generalização do *Phase Vocoder*. Partindo da Série de Fourier, eles propuseram a representação para um sinal $d(t)$ como soma finita de oscilações (senóides complexas) variando lentamente ao longo do tempo em freqüência (ou fase) e amplitude:

$$d(t) = \sum_{p=1}^{P} \mathrm{osc}[f_p(t), a_p(t), \theta_p(0)], \tag{3.7}$$

onde P é o total de oscilações e

$$\mathrm{osc}[f_p(t), a_p(t), \theta_p(0)] = a_p(t)\cos[\theta_p(t)], \tag{3.8}$$

com

$$\frac{d\theta_p}{dt}(t) = 2\pi f_p(t), \tag{3.9}$$

ou seja,

$$\theta_p(t) = \theta_p(0) + 2\pi \int_0^t f_p(u)du. \tag{3.10}$$

Cada oscilação osc_p é denominada de parcial de $d(t)$. As funções f_p, a_p e θ_p são, respectivamente, freqüência, amplitude e fase da p-ésima parcial. A partir de agora esse modelo será denominado de MQ ao longo da tese.

Pelo fato de este modelo não se limitar a harmônicos da freqüência fundamental Ω_1 da STFT (ou DFT) é possível utilizar técnicas de refinamento espectral que podem descrever as linhas freqüenciais com maior precisão. Algumas dessas técnicas serão abordadas no Capítulo 4.

3.1.3 Modelo SMS

A modelagem conhecida como SMS - *Spectral Modeling Synthesis* (em português Síntese por Modelagem Espectral) foi proposta por Serra e Smith [23] estendendo o modelo MQ a análises de sinais com ruído, e consiste em decompor o sinal em uma parte determinística e outra estocástica:

$$x(t) = d(t) + s(t), \tag{3.11}$$

onde:

$$d(t) = \text{ parte determinística;}$$

$$s(t) = \text{ parte estocástica.}$$

A parte determinística $d(t)$ é similar ao modelo MQ descrito na equação (3.7) e modela as parciais. A parte estocástica é a diferença entre o sinal original e a parte determinística, representando porções não-senoidais do sinal, como ataques das notas (transitórios) e ruído em geral, que devem, porém, variar lentamente para serem detectados.

Caso o modelo senoidal tenha sido calculado com a fase de cada parcial, então a parte estocástica é obtida pela subtração do original pelo determinístico no domínio do tempo. Caso contrário, deve-se fazer a subtração no domínio das freqüências, sempre em curtas janelas de tempo. É importante ressaltar que a parte estocástica não deve conter nenhuma parcial. Caso isso aconteça, a representação estocástica não será boa. Assim, antes da estimação deve-se analisar o sinal estocástico e refazer

a modelagem senoidal até que não sobre nenhuma parcial não-modelada na parte determinística [24].

A parte estocástica é, então, representada estimando-se o espectro dessa diferença de sinais. Finalmente, a síntese é feita por um ruído branco que é colorido pela estimação realizada do espectro. Em [23] é mostrado que somente uma aproximação linear já é suficiente para a representação dessa parte estocástica.

O modelo SMS não é capaz de analisar transientes rápidos, como bumbo de bateria ou ataque de xilofone, por exemplo. Para esse tipo de sinal, a parte estocástica deve ser decomposta em ruído e transitório [25]. Em [26] tem-se uma modelagem de transitórios, denominada TMS (do inglês *Transient Modeling Synthesis*) que é utilizada em conjunto com o SMS.

3.2 A formação de trilhas freqüenciais no tempo

A detecção de picos e a estimação de seus parâmetros de amplitude e fase utilizando as técnicas de transformada em blocos serão descritas no Capítulo 4. Elas são realizadas quadro a quadro separadamente, ou seja, não são utilizadas informações dos quadros anteriores. O interesse agora é observar como cada parcial obtida na etapa de análise evolui ao longo do tempo. Para juntar esses quadros é utilizado um método de continuação de picos, formando trilhas no domínio temporal-freqüencial.

Pretende-se obter a história de cada parcial, identificando quando nasceu, como evoluiu no tempo e quando morreu. Existem métodos heurísticos e métodos baseados em regras, assim como soluções estatísticas para esse problema de continuação do pico. Alguns métodos são explicados em [25]. Será apresentado a seguir um método baseado em regras, também descrito em [5], [4] e [24].

A maioria dos métodos baseados em regras tem como critério principal para a formação da trilha a proximidade das freqüências dos picos envolvidos em dois quadros consecutivos. Geralmente, a quantidade de picos obtidos na análise varia a cada quadro, já que alguns picos podem indicar ruído ou ataques rápidos de curta duração. A cada pico de um quadro é, então, atribuído um *status*: trilha emergente, trilha evoluindo ou trilha morrendo. Nos dois primeiros casos as trilhas

são consideradas ativas, e no terceiro, inativa.

No primeiro quadro do sinal, todos os picos são inicializados como trilhas emergentes. Os quadros seguintes são sempre comparados com o seu anterior, ou seja, os picos do quadro m são comparados com as trilhas ativas do quadro $m-1$. Seja, então, o quadro $m-1$ constituído por p picos de freqüências: $f_1, f_2, ..., f_p$. No quadro m há r picos de freqüências: $g_1, g_2, ..., g_r$.

Para todas as i trilhas ativas no quadro $m-1$, procuram-se no quadro m suas possíveis continuações. Isto é, a trilha i busca um pico g_j no quadro m, tal que $|f_i - g_j| < \triangle f_i$. O intervalo $\triangle f_i$ deve ser dependente da freqüência, por exemplo, um semitom em torno de f_i. Têm-se, para esse caso, duas possibilidades:

- Se a trilha i não possuir uma continuação, seu *status* muda de ativa para inativa, ou seja, a trilha está morrendo. Isso é feito da seguinte maneira: no quadro m, é criada uma continuação da trilha i, com mesma frequência f_i, porém com amplitude nula e fase

$$\theta_i[m] = \theta_i[m-1] + 2\pi f_i \frac{\zeta}{N}, \tag{3.12}$$

  onde ζ é a distância, em amostras, entre os quadros, e N é o tamanho do quadro. Assim, cada parcial tem uma morte suave.

- Se a trilha i achar uma continuação, seu *status* permanece como ativa (trilha evoluindo) e o pico g_j, que é o mais próximo de f_i em freqüência, passa a fazer parte da trilha. Porém, há casos em que mais de um pico no quadro m seja candidato à continuação da trilha. Existem, então, duas possibilidades:

  - g_j é um pico livre, ou seja, não foi requisitado por nenhuma outra trilha ativa no quadro $m-1$. Como não há conflito, o pico g_j é imediatamente associado à trilha i.

  - g_j já foi requisitado por outro pico no quadro $m-1$, diferente de f_i. Para resolver esse conflito, calculam-se as distâncias entre o pico g_j e os picos requerentes, decidindo-se de acordo com critérios previamente definidos. Suponha que duas trilhas u e v requisitem o mesmo pico g_j, e v é a trilha que o está atualmente requisitando. Obtêm-se as distâncias $d_u = |f_u - g_j|$ e $d_v = |f_v - g_j|$. Agora, se:

* $d_v > d_u$, a trilha atual, v, perde a disputa e escolhe então o pico mais adequado dentre os que estão disponíveis. Se existe algum pico adequado entre esses picos, a trilha permanece ativa. Se não, troca-se o *status* da trilha para inativa.
* $d_v < d_u$, a trilha atual, v, ganha a disputa e é realizado o procedimento de procura de um pico adequado para a trilha u, que passa a ser a trilha atual. A trilha u irá tentar novamente associar-se ao pico g_j e perderá a disputa. Então, de acordo com o item anterior, associar-se-á ao mais adequado pico dentre os picos disponíveis (se possível) e manter-se-á seu *status* como ativa ou mudar-se-á seu *status* para inativa.

O processo descrito acima é repetido para todas as trilhas ativas no quadro $m-1$ até que o *status* dessas trilhas tenha sido atualizado. Para os picos em m que permanecem não associados, novas trilhas são criadas, com *status* de trilhas emergentes. Similarmente ao que foi feito com as trilhas que estavam morrendo, as novas trilhas que aparecem no quadro m são estendidas ao quadro anterior, $m-1$, onde começam com amplitude zero e as mesmas freqüências às quais foram associadas no quadro m e fase:

$$\theta_i[m-1] = \theta_i[m] - 2\pi f_i \frac{\zeta}{N} \tag{3.13}$$

Um refinamento que pode ser feito no processo descrito consiste em incluir histereses associadas com a decisão de começar uma nova trilha ou terminar uma já existente. Um exemplo onde esse refinamento seria útil é o seguinte: pode acontecer de algumas parciais sofrerem modulação de amplitude. Ocorrendo tal fato, a amplitude da parcial pode permanecer abaixo do limiar de amplitude adotado durante alguns quadros. Sendo assim, o algoritmo de formação de trilhas irá terminar a trilha sempre que o pico a ela associado desaparecer em um dado quadro, começando uma nova trilha com o mesmo pico alguns quadros depois, quando o pico reaparecer. Assim, têm-se como resultado diversas trilhas segmentadas, ao invés de uma só trilha contínua, como deveria ser.

A histerese na mudança de *status* consiste em considerar um certo número de chances para que a trilha então termine, antes de mudar seu *status* para inativa.

Uma maneira prática de implementar essa proposta é a seguinte:

1. Aplicar um contador de quadros à trilha considerada inativa em um dado quadro.

2. Retardar essa mudança de *status* até que o contador atinja determinado valor. Isso implica estender a trilha por quadros sucessivos, inserindo picos com a mesma amplitude e freqüência, e calculando a fase pela equação (3.12).

3. Incrementar o contador a cada quadro processado.

4. Repetir o procedimento até que o contador atinja o valor determinado. Se nesse tempo a trilha encontrar um pico adequado, o contador deve ser zerado e deve-se proceder normalmente. Se não, confirmar a mudança de *status*, retirando os picos que foram criados artificialmente para estender a trilha, que deve, então, ser terminada no quadro estipulado originalmente pelo algoritmo.

Uma estratégia similar é utilizada para trilhas emergentes, com o objetivo de evitar que picos espúrios iniciem trilhas que serão muito curtas. Assim, uma trilha emergente só é confirmada caso permaneça ativa durante um certo número de quadros.

Têm-se, nas Figuras 3.2 e 3.3, representações da formação de trilhas freqüenciais no tempo.

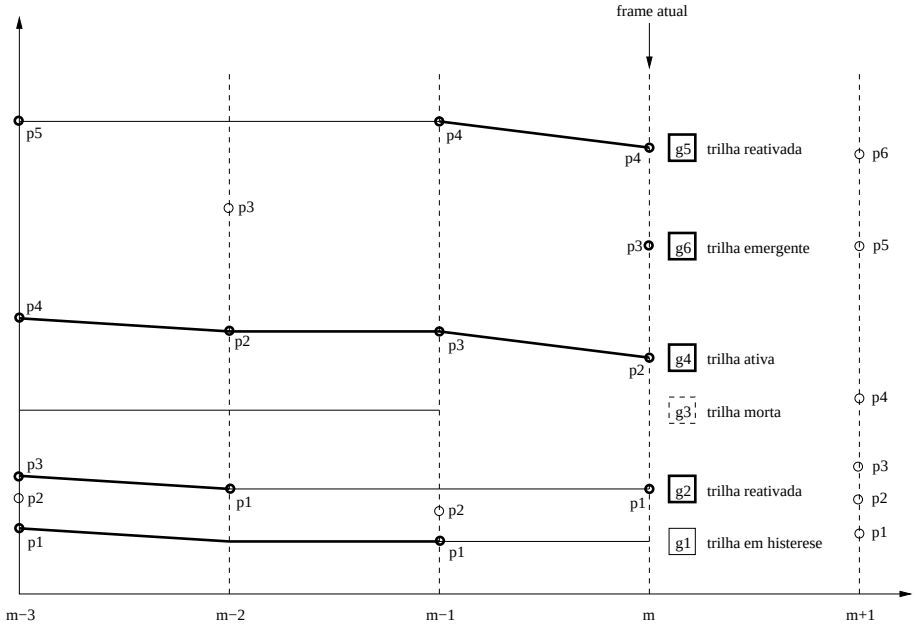

Figura 3.2: Esquema de formação de trilhas, onde g representa as trilhas e p, os picos.

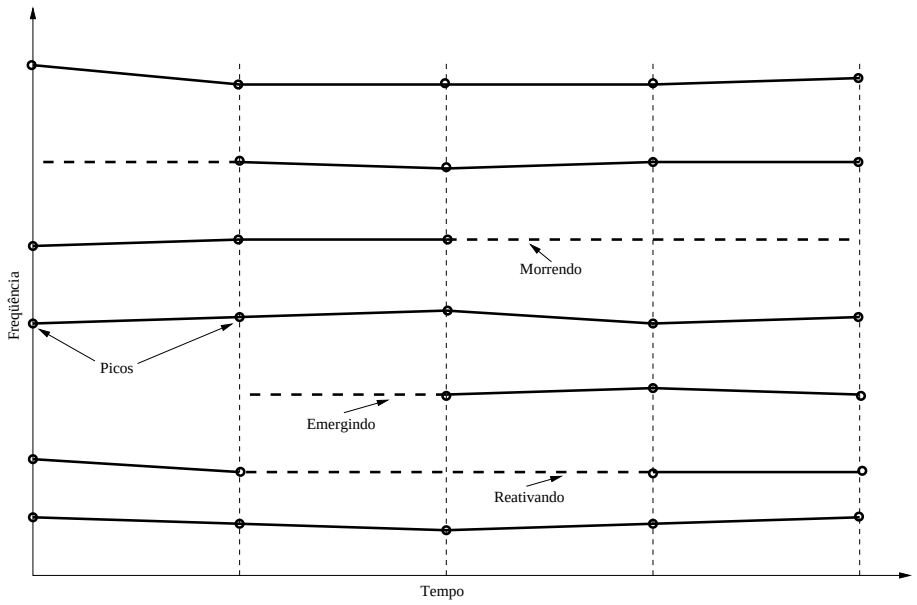

Figura 3.3: Esquema de formação de trilhas, destacando a utilização da estratégia de histerese.

3.2.1 Análise com informação prévia

A tese de mestrado de Eric Scheirer [27] propõe um método sobre análise de expressão musical em gravações de áudio utilizando informação da partitura executada.

A Figura 3.4 demonstra o funcionamento do algoritmo. Um processamento inicial da partitura determina aspectos estruturais da música, tais como notas tocadas em uníssono, quais notas se sobrepõem etc. A partitura também é utilizada para se obter o *pitch* da música. No *loop* principal são realizadas as seguintes iterações:

- Achar decaimentos (*releases*) e amplitudes das notas obtidas previamente;
- Achar início da próxima nota (*onset*) da partitura;
- Reexaminar a partitura, fazendo novas predições sobre o tempo local atual para estimar o início da próxima nota.

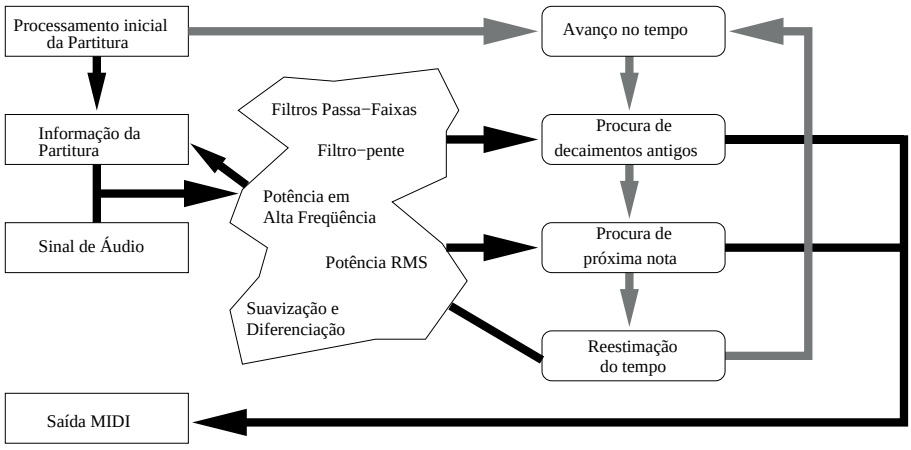

Figura 3.4: Diagrama do sistema de análise com informação prévia.

Assim que todos os inícios de nota estiverem localizados no tempo, procuram-se os decaimentos e obtêm-se as amplitudes de todas as notas. Em seguida os dados extraídos do sinal de áudio são escritos em formato MIDI (do inglês *Musical Instrument Digital Interface*).

O uso prévio de informação das notas é a peça-chave para esse método. Com ele possibilita-se comparar a interpretação dos artistas para uma determinada música. Ele pode ter uma vasta aplicação para extração de expressividade em música

erudita, que já possuem suas partituras vastamente transcritas para MIDI (sem muita expressividade) e músicas populares contemporâneas (como os estilos *rock* e *pop*), que podem ser facilmente transcritas.

O fato de esse método não realizar a síntese do sinal e sim gerar um arquivo MIDI, elimina a necessidade de minimização de ruído e também facilita a separação das informações provenientes de diferentes vozes ou instrumentos musicais, pois as partituras de cada instrumento já são conhecidas e possibilitam extrair apenas a expressividade.

Capítulo 4

Análise Espectral em Blocos a partir da Transformada de Fourier

Inicia-se o capítulo com uma breve apresentação das formas mais usuais de análise freqüencial na Seção 4.1 e de detecção dos picos espectrais na Seção 4.2. A reatribuição de tempo e freqüência para melhorar a resolução em ambas as dimensões é apresentada na Seção 4.3. Em particular, na Seção 4.4 são discutidos métodos de refinamento espectral obtidos a partir da freqüência instantânea. São eles a Reatribuição de Freqüência e a Diferença de Fase com sua versão iterativa. Outro método de refinamento espectral é o uso da Transformada de Fourier utilizando derivadas do sinal, apresentado na Seção 4.5. O capítulo é concluído mostrando-se exemplos que comparam os métodos.

4.1 A Análise Tradicional de Fourier

O sistema descrito no capítulo anterior pressupõe um estágio de detecção de amplitude e freqüência das componentes senoidais do modelo. A forma mais usual de se realizar essa análise é através das representações de Fourier. A Tabela 4.1[1] mostra as quatro representações básicas de Fourier.

Para um sinal de áudio genérico em tempo contínuo $x(t)$, aplica-se a Transformada de Fourier, que o descreve como uma combinação linear contínua de com-

[1]A convenção $< T >$ indica que a integral é realizada em um período T. $< N >$ indica que o somatório é realizado em um período N.

ponentes espectrais de forma $e^{j\omega t}$, isto é, senóides complexas. Caso o sinal $x(t)$ seja periódico, com período fundamental T, é possível recorrer à Série de Fourier, obtendo-se amplitudes complexas (módulo e fase) das componentes $e^{jk\omega_o t}$, com k inteiro e $\omega_0 = \frac{2\pi}{T}$, isto é, o sinal é descrito por exponenciais complexas harmônicas da freqüência fundamental ω_0.

No caso de sinais discretos $x[n]$, pode-se aplicar a Transformada de Fourier de Tempo Discreto, ou DTFT, gerando agora uma análise espectral em função dos componentes $e^{j\Omega n}$. Para sinais discretos e periódicos com período fundamental N, a Série de Fourier de Tempo Discreto, ou DTFS, utiliza componentes $e^{jk\Omega_0 n}$, com k inteiro, onde $\Omega_0 = \frac{2\pi}{N}$.

Tabela 4.1: As Quatro Representações Básicas de Fourier

Tipo	**Equação Espectral**
Transformada de Fourier	$X_{\mathrm{FT}}(j\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt$
Série de Fourier	$X_{\mathrm{FS}}[k] = \frac{1}{T}\int_{<T>} x(t)e^{-jk\omega_0 t}dt$
Transformada de Fourier de Tempo Discreto	$X_{\mathrm{DTFT}}(e^{j\Omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\Omega n}$
Série de Fourier de Tempo Discreto	$X_{\mathrm{DTFS}}[k] = \frac{1}{N}\sum_{n=<N>} x[n]e^{-jk\Omega_0 n}$

Para sinais $x[n]$ de duração finita dada por N, é possível redefinir a DTFS como a Transformada Discreta de Fourier (DFT), sendo um somatório realizado sobre o suporte do sinal, i.e. $\forall n \in [0, N-1]$:

$$X_{\mathrm{DFT}}[k] = \frac{1}{N}\sum_{n=0}^{N-1} x[n]e^{-jk\Omega_0 n}. \tag{4.1}$$

A necessidade de se realizar a análise de um trecho particular de sinal é resolvida de forma geral pelo seu janelamento, multiplicando-o por uma janela (em geral, finita) $w(t)$ ou $w[n]$, levando à definição da Transformada de Fourier de Curta Duração, ou STFT [28]:

$$X_{\mathrm{STFT}}(\tau, j\omega) = \int_{-\infty}^{\infty} x(t)w^*(t-\tau)e^{-j\omega t}dt \tag{4.2}$$

e sua versão discreta no tempo, a DTSTFT:

$$X_{\mathrm{DTSTFT}}[l, e^{j\Omega}] = \sum_{n=-\infty}^{\infty} x[n]w^*[n-l]e^{-j\Omega n}, \tag{4.3}$$

onde τ é a posição no tempo da janela que percorre o sinal $x(t)$, e seu equivalente discreto é a variável l para $x[n]$. A componente espectral resultante é descrita para cada freqüência ω ou Ω ao longo do espectro.

As análises discretas na freqüência citadas até agora dividem o espectro de forma linear. Mas há outras alternativas não-lineares baseadas nas representações de Fourier: a CQT e a BQT. A Transformada de Q Constante (CQT) distribui cada componente espectral de forma logarítmica. A Transformada de Q Limitado (BQT, do inglês *Bounded-Q Transform*) possui uma resolução linear dentro das oitavas, porém dobra essa resolução a cada oitava inferior. Sua discussão será adiada para o Capítulo 6, onde serão comparados métodos lineares e não-lineares de subdivisão do espectro.

4.2 Esquema Usual para Detecção de Picos

A forma geral para busca dos picos que irão gerar as trilhas (ver seção 3.2) pode ser descrita da seguinte forma:

- Calcular a STFT ao longo do tempo. Para cada posição da janela:
  - Buscar os picos mais proeminentes do espectro.
  - Determinar os pares de freqüência e amplitude correspondentes.

Existem várias formas de detectar os picos. A mais simples é escolher todos os picos, ou seja, no espectro positivo considerar todas as componentes que possuírem seus dois *bins* (ou raias) adjacentes menores [24]. Neste caso, tentam-se eliminar os picos ruidosos no estágio seguinte, de formação das trilhas. Porém isso gera inúmeros picos espúrios de baixa amplitude, além de incluir também os lobos laterais dos picos janelados.

Duas outras maneiras de escolher somente os picos mais proeminentes são adicionar um limiar constante ou variável com a freqüência na amplitude das componentes espectrais. O método de limiar constante, no entanto, deve ser calculado a cada quadro, variando entre 60 e 80 dB relativamente ao máximo global [24]. Com isso, tende-se a descartar as componentes harmônicas de alta freqüência com baixa amplitude; mas como os sinais de áudio concentram sua energia em baixa freqüência,

este método já é bastante aceitável. No caso do limiar variável, é possível comparar o espectro com uma estimativa do ruído espectral de fundo ou com uma estimativa do envelope espectral obtida por codificação preditiva linear (LPC), à qual se adiciona um *offset* negativo em dB.

A seguir abordam-se métodos de refinamento da descrição espectral, na tentativa de localizar mais acurada e precisamente os eventos no tempo e na freqüência.

4.3 Reatribuição de Tempo e de Freqüência

Conforme foi visto na seção anterior, a STFT é descrita pela equação (4.2) e possui uma representação contínua do espectro. A resolução de freqüência de cada *bin* da transformada pode ser medida por uma largura de faixa efetiva $\Delta\omega$, e pode ser associada ao espaçamento mínimo para distinguir duas senóides. Já a resolução de tempo $\Delta\tau$ indica uma duração efetiva da janela $w(t)$, supostamente concentrada no tempo. Devido à dualidade tempo-freqüência herdada da Transformada de Fourier, a resolução de freqüência $\Delta\omega$ é inversamente proporcional à resolução de tempo $\Delta\tau$ [28]:

$$\Delta\tau\Delta\omega \geq \frac{1}{2}. \tag{4.4}$$

Essa formulação é proveniente da mecânica quântica, e corresponde ao Princípio de Incerteza de Heisenberg.

Para compreender melhor a equação (4.2) da STFT, pode-se reescrevê-la em coordenadas polares:

$$X_{\mathrm{STFT}}(\tau, j\omega) = a(\tau, j\omega)e^{j\phi(\tau,j\omega)}, \tag{4.5}$$

onde $a(\tau, j\omega)$ é amplitude e $\phi(\tau, j\omega)$ é a fase da componente freqüencial $e^{j\omega t}$ na janela centrada em $t = \tau$.

Calculando-se o módulo da STFT ao quadrado, tem-se a distribuição da energia no espaço tempo-freqüência, denominada de espectrograma:

$$|X_{\mathrm{STFT}}(\tau, j\omega)|^2 = \left| \int_{-\infty}^{\infty} x(t)w^*(t-\tau)e^{-j\omega t}dt \right|^2. \tag{4.6}$$

O espectrograma, porém, devido ao princípio de Heisenberg, não possui uma boa resolução. Os métodos de Reatribuição de Tempo e Freqüência (*Time- and Frequency-Reassignment*) buscam um refinamento a partir do deslocamento das

componentes para o centro de gravidade da distribuição de energia do plano tempo-freqüência [29]. Este corresponde aos pontos estacionários de fase, que, para um dado par (τ, ω), ocorrem na solução do sistema [30]:

$$\frac{\partial}{\partial \omega}[\phi(\tau, j\omega) + \omega \cdot (t - \tau)] = 0 \tag{4.7}$$

$$\frac{\partial}{\partial \tau}[\phi(\tau, j\omega) + \omega \cdot (t - \tau)] = 0. \tag{4.8}$$

Resolvendo-se, os valores realocados de ω e t são, respectivamente:

$$\hat{\omega}(\tau, j\omega) = \frac{\partial}{\partial t}\phi(\tau, j\omega) \tag{4.9}$$

$$\hat{t}(\tau, j\omega) = \tau - \frac{\partial}{\partial \omega}\phi(\tau, j\omega). \tag{4.10}$$

Assim, a correção do par (τ, ω) do centro geométrico implícito da STFT para o centro de gravidade da energia utiliza a freqüência instantânea $\hat{\omega}$ e corrige o tempo pelo atraso de grupo no ponto de interesse.

A seguir, serão apresentados métodos de correção da freqüência com base na freqüência instantânea.

4.4 Análise da Freqüência Instantânea

Analisando-se a fase $\phi(\tau, j\omega)$ da equação (4.2), pode-se obter a freqüência instantânea do sinal [31]:

$$\hat{\omega}(\tau, j\omega) = \frac{\partial}{\partial t}\phi(\tau, j\omega) = \frac{\partial}{\partial t}\mathrm{Im}\left\{\ln X_{\mathrm{STFT}}(\tau, j\omega)\right\}. \tag{4.11}$$

Substituindo (4.2) em (4.11), tem-se então que a freqüência instantânea é

$$\hat{\omega}(\tau, j\omega) = \omega + \mathrm{Im}\left\{\frac{\int_{-\infty}^{\infty} x(t)w'^{*}(t-\tau)e^{-j\omega t}dt}{\int_{-\infty}^{\infty} x(t)w^{*}(t-\tau)e^{-j\omega t}dt}\right\}, \tag{4.12}$$

onde $w'(t)$ é a derivada da janela $w(t)$.

Para obter uma versão discreta do cálculo da freqüência instantânea, pode-se partir da DSTFT, que combina a DFT da equação (4.1) com a DTSTFT da equação (4.3), supondo a janela $w[n]$ com suporte finito N:

$$X_{\mathrm{DSTFT}}[n_0, k] = \frac{1}{N}\sum_{n=0}^{N-1} x[n + n_0]w[n]e^{-jk\Omega_0 n}, \tag{4.13}$$

onde n_0 marca o centro da janela $w[n]$.

Discretizando, então, a equação (4.12), a freqüência instantânea discreta é calculada utilizando duas DSTFTs, uma com a janela $w[n]$, resultando em X^{w}_{DSTFT}, e outra janela com uma ponderação na freqüência $w'[n]$, equivalente a uma derivação no tempo, resultando em $X^{w'}_{\mathrm{DSTFT}}$:

$$\hat{\Omega}[n_0,k] = \Omega[n_0,k] + \frac{-\pi}{N}\mathrm{Im}\left\{\frac{X^{w'}_{\mathrm{DSTFT}}[n_0,k](X^{w}_{\mathrm{DSTFT}}[n_0,k])^*}{|X^{w}_{\mathrm{DSTFT}}[n_0,k]|^2}\right\}. \tag{4.14}$$

Utilizando-se, em particular, uma janela de Hanning, cuja versão discreta pode ser escrita como [32]

$$w[n] = \frac{1}{2}\left[1 - \left(\frac{1}{2}\right)e^{j\Omega_0 n} - \left(\frac{1}{2}\right)e^{-j\Omega_0 n}\right], \tag{4.15}$$

a equação (4.13) pode ser escrita como

$$X^{H}_{\mathrm{DSTFT}}[n_0,k] = \frac{1}{2}\left\{X[k] - \frac{1}{2}X[k+1] - \frac{1}{2}X[k-1]\right\}, \tag{4.16}$$

onde $X[k]$ é a DFT do sinal, calculada conforme a equação (4.1). Para simplificar a notação, a X^{H}_{DSTFT} será escrita como X^H.

4.4.1 Método de Reatribuição de Freqüência

Dispondo-se, então, dos resultados anteriores, pode-se escrever a freqüência instantânea da equação (4.14) utilizando-se uma janela de Hanning [33]:

$$\hat{\Omega}[n_0,k] = \Omega[n_0,k] + \frac{-\pi}{N}\mathrm{Im}\left\{\frac{(X[k-1]-X[k+1])\left(X^H[n_0,k]\right)^*}{2j\,|X^H[n_0,k]|^2}\right\}. \tag{4.17}$$

A equação (4.17) descreve o método de Reatribuição de Freqüência nos domínios do tempo e da freqüência discretos, com janela de Hanning. Ela indica que a freqüência instantânea pode ser obtida com apenas uma DFT pura, pois o espectro de Fourier janelado por Hanning (X^H) também é obtido pela DFT, usando as amplitudes dos três *bins* mais próximos à freqüência escolhida, como mostrado na equação (4.16).

4.4.2 Método da Diferença da Fase

Friedman em 1985 [31] e Brown em 1993 [34] apresentaram um método que utiliza a diferença na fase entre duas janelas adjacentes deslocadas de uma amostra para obter de forma aproximada a freqüência instantânea.

Considerando-se que o sinal em duas janelas deslocadas de m amostras possui um espectro idêntico, a menos de uma diferença de fase, tem-se que $T\{x[n]\} = X[k]$ e então

$$T\{x[n+m]\} \cong e^{jk\Omega_0 m} X[k]. \tag{4.18}$$

Para linhas musicais, essa suposição é bastante aceitável se o deslocamento m for pequeno. Analisando-se o espectro deslocado de uma amostra ($m = 1$), a partir da equação (4.16), essa hipótese leva a

$$\begin{aligned} X^H[n_0+1,k] &= \frac{1}{2}e^{jk\Omega_0}\left\{X[k] - \frac{1}{2}e^{j\Omega_0}X[k+1] - \frac{1}{2}e^{-j\Omega_0}X[k-1]\right\} \\ &\approx e^{jk\Omega_0}X^H[n_0,k], \end{aligned} \tag{4.19}$$

supondo $e^{\pm j\Omega_0} \approx 1$.

Esse resultado valida a propriedade da equação (4.18) para o caso de janelamento por janela de Hanning. Então, pela fase da razão $\frac{X^H[n_0+1,k]}{X^H[n_0,k]}$ é possível estimar a freqüência instantânea[2] como

$$\hat{\Omega}[n_0,k] = \phi[n_0+1,k] - \phi[n_0,k], \tag{4.20}$$

onde

$$\begin{aligned} \phi[n_0,k] &= \operatorname{arctg}\left\{\frac{\operatorname{Im}\left(X^H[n_0,k]\right)}{\operatorname{Re}\left(X^H[n_0,k]\right)}\right\} \\ \phi[n_0+1,k] &= \operatorname{arctg}\left\{\frac{\operatorname{Im}\left(X^H[n_0+1,k]\right)}{\operatorname{Re}\left(X^H[n_0+1,k]\right)}\right\}. \end{aligned} \tag{4.21}$$

Desta vez, duas DFTs puras são necessárias, para o cálculo de $X^H[n_0,k]$ e $X^H[n_0+1,k]$.

4.4.3 Método Iterativo

A substituição, no método da Diferença de Fase, da DFT (freqüência discreta) pela DTFT (contínua) possibilita a realização do refinamento de forma iterativa. É o que se propõe em [35], onde também se incorpora a correção da amplitude. Como demonstração serão apresentados os casos para uma, duas e i senóides complexas.

[2]Naturalmente, não se espera que $\hat{\Omega} = k\Omega_0$, mas sim a freqüência que gerou um pico aparente em $k\Omega_0$.

4.4.3.1 Demonstração

Sejam um sinal $x[n]$ que tem um conteúdo espectral dado por $X(e^{j\Omega})$ e uma janela $w[n] \leftrightarrow W(e^{j\Omega})$. Janelando o sinal e sua versão deslocada de m, tem-se

$$\begin{aligned} x_w[n] = w[n]x[n] &\longleftrightarrow X_w(e^{j\Omega}) = W(e^{j\Omega}) \circledast X(e^{j\Omega}) \\ \tilde{x}_w[n] = w[n]x[n+m] &\longleftrightarrow \tilde{X}_w(e^{j\Omega}) = W(e^{j\Omega}) \circledast \{e^{j\Omega m} X(e^{j\Omega})\} \end{aligned} \quad (4.22)$$

Supondo uma senóide complexa única, $X(e^{j\Omega}) = A_1\delta(\Omega - \Omega_0)$, e

$$\begin{aligned} X_w(e^{j\Omega}) &= W(e^{j\Omega}) \circledast \{A_1\delta(\Omega - \Omega_0)\} \\ &= A_1 W(e^{j(\Omega-\Omega_0)}) \\ \tilde{X}_w(e^{j\Omega}) &= W(e^{j\Omega}) \circledast \{e^{j\Omega_0 n_0} A_1\delta(\Omega - \Omega_0)\} \\ &= A_1 e^{j\Omega_0 m} W(e^{j(\Omega-\Omega_0)}) \\ \frac{\tilde{X}_w(e^{j\Omega})}{X_w(e^{j\Omega})} &= e^{j\Omega_0 m}. \end{aligned} \quad (4.23)$$

A fase da exponencial está, então, diretamente relacionada com a freqüência da senóide, i.e., $\Delta\phi(\Omega_0) = \Omega_0 m$ para uma janela $w[n]$ qualquer.

Para duas senóides complexas, $X(e^{j\Omega}) = A_1\delta(\Omega + \Omega_0) + A_2\delta(\Omega - \Omega_0)$, e

$$\begin{aligned} X_w(e^{j\Omega}) &= A_1 W(e^{j(\Omega+\Omega_0)}) + A_2 W(e^{j(\Omega-\Omega_0)}) \\ \tilde{X}_w(e^{j\Omega}) &= A_1 e^{-j\Omega_0 m} W(e^{j(\Omega+\Omega_0)}) + A_2 e^{j\Omega_0 m} W(e^{j(\Omega-\Omega_0)}) \\ \frac{\tilde{X}_w(e^{j\Omega})}{X_w(e^{j\Omega})} &= \frac{A_1 e^{-j\Omega_0 m} W(e^{j(\Omega+\Omega_0)}) + A_2 e^{j\Omega_0 m} W(e^{j(\Omega-\Omega_0)})}{A_1 W(e^{j(\Omega+\Omega_0)}) + A_2 W(e^{j(\Omega-\Omega_0)})}. \end{aligned} \quad (4.24)$$

Analisando o espectro em $\Omega = \Omega_0$:

$$\frac{\tilde{X}_w(e^{j\Omega})}{X_w(e^{j\Omega})} = \frac{A_1 e^{-j\Omega_0 m} W(e^{j(2\Omega_0)}) + A_2 e^{j\Omega_0 m} W(e^{j0})}{A_1 W(e^{j(2\Omega_0)}) + A_2 W(e^{j0})}, \quad (4.25)$$

e para $\Omega = -\Omega_0$:

$$\frac{\tilde{X}_w(e^{j\Omega})}{X_w(e^{j\Omega})} = \frac{A_1 e^{-j\Omega_0 m} W(e^{j(0)}) + A_2 e^{j\Omega_0 m} W(e^{j(2\Omega_0)})}{A_1 W(e^{j0}) + A_2 W(e^{j(2\Omega_0)})}. \quad (4.26)$$

Se $W(e^{j(2\Omega_0)}) \longrightarrow 0$, então $\Delta\phi(\pm\Omega) = \pm\Omega_0 m$.

Supondo um somatório de senóides complexas $X(e^{j\Omega}) = \sum_i A_i\delta(\Omega - \Omega_i)$, tem-se por analogia que

$$\frac{\tilde{X}_w(e^{j\Omega})}{X_w(e^{j\Omega})} = \frac{\sum_i A_i e^{j\Omega_i m} W(e^{j(\Omega-\Omega_i)})}{\sum_i A_i W(e^{j(\Omega-\Omega_i)})}. \quad (4.27)$$

Para $\Omega = \Omega_0$,

$$\frac{\tilde{X}_w(e^{j\Omega_0})}{X_w(e^{j\Omega_0})} = \frac{\sum_i A_i e^{j\Omega_i m} W(e^{j(\Omega_0-\Omega_i)})}{\sum_i A_i W(e^{j(\Omega_0-\Omega_i)})}. \tag{4.28}$$

Nos casos em que Ω_0 está distante de $\Omega_{i\neq 0}$,

$$\begin{aligned} \frac{\tilde{X}_w(e^{j\Omega_0})}{X_w(e^{j\Omega_0})} &= e^{j\Omega_0 m} \\ \Delta\phi(\Omega_0) &= \Omega_0 m. \end{aligned} \tag{4.29}$$

Pode-se concluir a partir das demonstrações acima que o espectro verdadeiro é obtido a partir da remoção do efeito das janelas. Isso pode ser descrito de forma matricial como sendo a resolução do seguinte sistema linear [18]:

$$\mathbf{x_o} = \mathbf{W}^{-1} \cdot \mathbf{x}, \tag{4.30}$$

onde $\mathbf{x}$ é vetor do espectro aparente (janelado), $\mathbf{x_o}$ é vetor do espectro verdadeiro e $\mathbf{W}$ é a matriz espectral da janela, ou seja,

$$\begin{aligned} \mathbf{x} &\equiv \left[X_w(e^{-j\Omega_1}) \ldots X_w(e^{-j\Omega_n})\ X_w(e^{j\Omega_n}) \ldots X_w(e^{j\Omega_1})\right]^T \\ \mathbf{x_o} &\equiv \left[X(e^{-j\Omega_1}) \ldots X(e^{-j\Omega_n})\ X(e^{j\Omega_n}) \ldots X(e^{j\Omega_1})\right]^T \\ \mathbf{W}(i,j) &= W(e^{j(\Omega_i-\Omega_j)}). \end{aligned} \tag{4.31}$$

Para resolver a equação (4.30) precisa-se conhecer precisamente as freqüências das componentes senoidais do sinal, que podem ser calculadas, então, com o método da diferença de fase, já descrito no item 4.4.2, utilizando duas janelas, uma de referência e outra deslocada de m amostras[3]. A fase de cada componente é calculada na janela de referência, e avança à medida que a janela se desloca ao longo do sinal.

A relação dos espectros verdadeiros avaliados na posição de referência e deslocado de m amostras pode ser escrita de forma matricial para todas as componentes freqüenciais

$$\mathbf{x}_o^{(m)} = \mathbf{X}_o^{(0)} \cdot \theta^{(m)} \rightarrow \theta^{(m)} = \left[\mathbf{X}_o^{(0)}\right]^{-1} \cdot \mathbf{x}_o^{(m)}, \tag{4.32}$$

onde

- $\mathbf{X}_o^{(0)}$: matriz do espectro de referência, i.e., matriz diagonal do espectro verdadeiro com a janela na posição de referência:

$$\mathbf{X}_o^{(0)} \equiv \begin{cases} X^{(0)}(e^{j\Omega})|_{\Omega=\{-\Omega_1 \ldots -\Omega_n \Omega_n \ldots \Omega_1\}}, & i = j \\ 0, & i \neq j. \end{cases} \tag{4.33}$$

[3]No caso descrito no item 4.4.2 o deslocamento era de uma amostra, neste caso só amostra.

- $\mathbf{x}_o^{(m)}$: vetor do espectro deslocado, i.e., vetor do espectro verdadeiro com a janela de análise na posição deslocada de m amostras:

$$\mathbf{x}_o^{(m)} \equiv \left[X^{(m)}(e^{-j\Omega_1}) \dots X^{(m)}(e^{-j\Omega_n}) X^{(m)}(e^{j\Omega_n}) \dots X^{(m)}(e^{j\Omega_1})\right]^T. \quad (4.34)$$

- $\theta^{(m)}$: vetor da rotação de fase, i.e., vetor que contém a rotação de fase com a janela de análise deslocada de m amostras:

$$\theta^{(m)} \equiv \left[e^{j\Omega_1 m} \dots e^{j\Omega_n m}\ e^{-j\Omega_n m} \dots e^{-j\Omega_1 m}\right]^T. \quad (4.35)$$

4.4.3.2 Algoritmo

A equação (4.32) determina a rotação de fase e, portanto, as freqüências instantâneas, com base na relação entre o espectro verdadeiro do sinal nas janelas de referência e deslocada, enquanto que a equação (4.30) calcula o espectro verdadeiro a partir do efeito de vazamento espectral da janela e do espectro aparente com as freqüencias instantâneas.

Combinando-se essas equações, é possível gerar um algoritmo recursivo que, iterativamente encontra o espectro verdadeiro e estima as frequências instantâneas das componentes do sinal.

Enquanto a diferença entre estimativas consecutivas da freqüência instantânea das componentes do sinal for maior que uma tolerância escolhida, faça a seguinte seqüência:

1) Avalie os espectros da janela e dos sinais, para as freqüências estimadas na iteração anterior $(r-1)$:

$$\begin{aligned}
\{\hat{\mathbf{x}}^{(0)}\}_r &\leftarrow X_w^{(0)}(e^{j\hat{\Omega}})|_{\hat{\Omega}=\{-\hat{\Omega}_1 \dots -\hat{\Omega}_n \hat{\Omega}_n \dots \hat{\Omega}_1\}_{r-1}} \\
\{\hat{\mathbf{x}}^{(m)}\}_r &\leftarrow X_w^{(m)}(e^{j\hat{\Omega}})|_{\hat{\Omega}=\{-\hat{\Omega}_1 \dots -\hat{\Omega}_n \hat{\Omega}_n \dots \hat{\Omega}_1\}_{r-1}} \\
\{\hat{\mathbf{W}}\}_r &\leftarrow W^{(0)}(e^{j(\hat{\Omega}_i - \hat{\Omega}_j)})|_{\hat{\Omega}=\{-\hat{\Omega}_1 \dots -\hat{\Omega}_n \hat{\Omega}_n \dots \hat{\Omega}_1\}_{r-1}}.
\end{aligned} \quad (4.36)$$

2) Corrija os espectros dos sinais de referência e atrasado:

$$\begin{aligned}
\{\hat{\mathbf{x}}_o^{(0)}\}_r &= \{\hat{\mathbf{W}}^{-1}\}_r \cdot \{\hat{\mathbf{x}}^{(0)}\}_r \\
\{\hat{\mathbf{x}}_o^{(m)}\}_r &= \{\hat{\mathbf{W}}^{-1}\}_r \cdot \{\hat{\mathbf{x}}^{(m)}\}_r.
\end{aligned} \quad (4.37)$$

3) Avalie a rotação da fase e a freqüência instantânea das componentes do sinal:

$$\begin{aligned} \{\hat{\theta}^{(m)}\}_r &= \frac{\{\hat{\mathbf{x}}_o^{(m)}\}_r}{\{\hat{\mathbf{x}}_o^{(0)}\}_r} \\ \{\hat{\mathbf{\Omega}}\}_r &= -\frac{1}{m}\arg\{\hat{\theta}^{(m)}\}_r. \end{aligned} \tag{4.38}$$

A estimativa inicial das componentes espectrais pode ser obtida por uma FFT, C*Q*T (*Constant-Q Transform*, ver Seção 6.2.1) ou outro método. Vale ressaltar, porém, que as freqüências negativas também devem ser incluídas na análise, pois, como já foi dito anteriormente, seus lobos laterais também influenciam a formação do espectro aparente.

Caso alguma componente não-existente (apenas aparente no espectro) seja selecionada, sua magnitude deve convergir para zero (seu valor verdadeiro) e deverá ser descartada da matriz $\mathbf{X}_o^{(0)}$, evitando-se que seu vazamento distorça todas as componentes espectrais, principalmente as vizinhas.

A rotação de fase requer que todos os ângulos sejam medidos no mesmo ciclo 2π. Assim, para uma janela atrasada por um valor positivo m, deve-se somar 2π à fase das componentes positivas, que giram em sentido anti-horário, e para as negativas (sentido horário) subtraem-se 2π da rotação de fase. Para sinais de banda larga, devido a possíveis variações de curta duração, é utilizada uma janela com apenas uma amostra de atraso, ou seja, $m = 1$.

4.5 Transformada de Fourier Usando Derivadas do Sinal

Como a derivada de uma senóide é uma senóide com um deslocamento na fase, uma outra abordagem para se refinar o espectro é utilizar as derivadas do sinal. Nota-se aqui uma grande similaridade com os métodos de Freqüência Instantânea, que utilizam a derivada da janela.

A partir da descrição do modelamento senoidal apresentado na equação (3.7), derivando-se o sinal no tempo, demonstra-se em [36] que a l-ésima derivada do sinal determinístico $d(t) = \sum_p a_p(t)\cos\theta_p(t)$ é

$$\frac{\partial^l d}{\partial t^l}(t) = \sum_{p=1}^{P} a_p(t) \cdot (2\pi f_p(t))^l \cdot \cos\left(\theta_p(t) + (-l \cdot \frac{\pi}{2})\right). \tag{4.39}$$

Como conseqüência da equação (4.39), cada parcial p possui um máximo de amplitude na l-ésima derivada da Transformada de Fourier FT^l dado pela freqüência f_p

$$f_p = \frac{1}{2\pi} \left| \frac{\mathrm{FT}^{l+1}[f_p]}{\mathrm{FT}^l[f_p]} \right|, \tag{4.40}$$

que para o cálculo contínuo é inútil, pois exige que se saiba de antemão o valor de f_p. Mas para sua versão discreta esse método mostra-se funcional, pois são utilizados os valores dos *bins* mais próximos da parcial p, como será mostrado na equação (4.44).

A formulação da DFT, apresentada na equação (4.1), porém usando a l-ésima derivada do sinal $x[n]$ é definida como [36]

$$X_{\mathrm{DFT}^l}[k] = \frac{1}{N} \sum_{n=0}^{N-1} w[n] \frac{\partial^l x}{\partial t^l}[n] e^{-jk\Omega_0 n}, \tag{4.41}$$

onde a janela $w[n]$ foi acrescentada à definição da DFT, o que formalmente deveria torná-la uma DSTFT, porém foi mantida a convenção da literatura original [36].

Em [36] é demostrado que a diferenciação é uma operação linear que pode ser vista como uma filtragem com ganho

$$|H(e^{j\Omega})| = F_s \sqrt{2(1 - \cos\Omega)}, \tag{4.42}$$

que, tem um erro em relação ao valor ideal $|\Omega|$, crescente com a freqüência. Para corrigir isto, multiplica-se o espectro de magnitude do sinal derivado por um fator F definido como:

$$F(\Omega) = \frac{\Omega}{\sqrt{2(1 - \cos\Omega)}}. \tag{4.43}$$

Utilizando a primeira derivada do sinal[4] obtêm-se a DFT^1. O cálculo corrigido da freqüência f_p é dado, então, pela equação:

$$f_p = \frac{1}{2\pi} \left| \frac{\mathrm{DFT}^1[m_p]}{\mathrm{DFT}^0[m_p]} \right|, \tag{4.44}$$

onde m_p é o índice do *bin* de máxima amplitude na DFT^0 (ou $X_{\mathrm{DFT}}[k]$) correspondente à freqüência f_p. Mais precisamente, m_p é o inteiro mais próximo a $f_p N/F_s$

[4] DFT^1 é uma notação simplificada de $X_{\mathrm{DFT}^1}[k]$

e

$$m_p \frac{F_s}{N} \leq f_p < (m_p + 1)\frac{F_s}{N} \tag{4.45}$$

Caso f_p não satisfaça tais condições, a análise pela DFT[1] falhou para essa freqüência, podendo indicar uma contaminação da raia espectral encontrada, isto é, não há uma única freqüência nesse canal da DFT.

Para corrigir o efeito que a janela impõe sobre a amplitude, é utilizada a seguinte equação:

$$a_p = \frac{a_p^0}{W\left(\left|\frac{f_P - f_p^0}{F_s/N}\right|\right)}, \tag{4.46}$$

onde a_p^0 é o valor obtido através da DFT para a amplitude, f_p^0 é o valor obtido para a freqüência pela DFT e $W(\Omega)$, a DTFT da janela $w[n]$ empregada na análise do sinal amostrado $x[n]$.

O procedimento para a implementação do método baseado na DFT[1] para cada bloco do sinal temporal é:

- Aplicar o janelamento ao sinal original x;
- Obter a DFT[0];
- Computar a derivada de x, denominada de $x^{'}$;
- Aplicar o mesmo janelamento a $x^{'}$;
- Obter a DFT[1];
- Corrigir o espectro de magnitude da DFT[1] pelo fator F, equação (4.43);
- Para cada índice m referente a um máximo na DFT[0]:
    1. Computar a freqüência exata, pela equação (4.44);
    2. Computar a amplitude exata, pela equação (4.46);
    3. Adicionar o par (freqüência, amplitude) à lista de resultados do bloco corrente;

4.6 Exemplos

Nesta seção são apresentados exemplos com senóides sintetizadas e sinais reais para comparar os métodos de refinamento espectral apresentados neste capítulo.

4.6.1 Senóides

O primeiro teste foi realizado com um sinal sintetizado de quatro senóides com freqüências de 263, 526, 789 e 1052Hz, i.e., uma fundamental de 263Hz e suas três primeiras harmônicas (2ª, 3ª e 4ª). O sinal foi gerado a uma taxa de 44100 amostras por segundo e as DFTs de análise para todos os métodos com 1024 pontos. O trecho analisado é estacionário. A Figura 4.1 representa o espectro das quatro senóides.

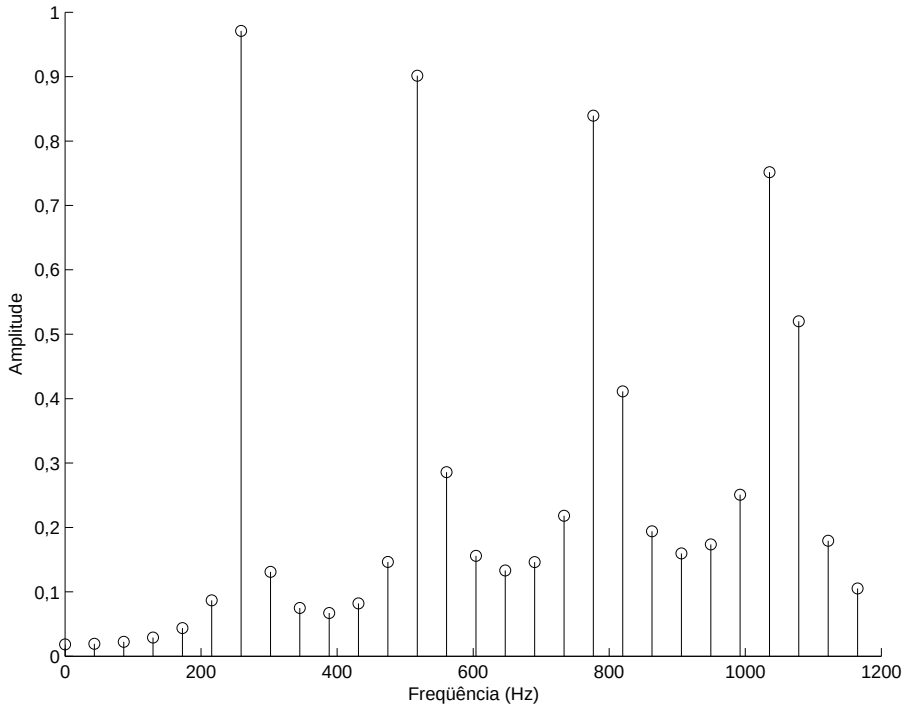

Figura 4.1: Espectro da DFT para as quatro senóides com 1024 pontos.

A Tabela 4.2 mostra as freqüências instantâneas obtidas pelos métodos de Reatribuição de Freqüência, Diferença de Fase e DFT[1] [5] para as quatro senóides mencionadas. Já a Tabela 4.3 mostra o refinamento gerado pelo Método Iterativo da Diferença de Fase.

[5]Nesta seção será utilizada a sigla DFT[1] para indicar o método da Transformada de Fourier com a Primeira Derivada.

Tabela 4.2: Freqüência Instantânea obtida pelos métodos de Reatribuição de Freqüência, Diferença de Fase e DFT[1] para quatro senóides, com uma janela de 1024 pontos. Entre parênteses está o desvio, em porcentagem.

Métodos	Freqüências (Hz)			
	263	526	789	1052
DFT	258,9041 (-1,5574)	517,8082 (-1,5574)	776,7123 (-1,5574)	1035,6164 (-1,5574)
Reatribuição de Freqüência	263,6251 (0,2377)	527,2945 (0,2461)	790,9242 (0,2439)	1054,3002 (0.2187)
Diferença de Fase	263,1167 (**0,0444**)	526,2873 (**0,0546**)	789,4110 (**0,0520**)	1052,2803 (**0,0266**)
DFT[1]	262,6566 (-0,1306)	525,5123 (-0,0927)	788,3775 (-0,0789)	1051,4873 (-0,0487)

Tabela 4.3: Freqüência Instantânea obtida pelo método iterativo da diferença de fase para quatro senóides com 5 iterações

Iterações	Freqüências (Hz)			
	263	526	789	1052
1	263,1167	526,2873	789,4110	1052,2803
2	263,0531	525,9835	788,9905	1051,9352
3	263,054	525,9935	789,0039	1051,9404
4	263,054	525,9932	789,0034	1051,9403
5	263,054 (**0,0205**)	525,9935 (**-0,0012**)	789,0034 (**-4,3093e-4**)	1051,9403 (**-0,0057**)

4.6.2 Senóides com Ruído

O segundo teste foi realizado com uma mistura do mesmo sinal sintetizado no primeiro exemplo (quatro senóides com freqüências de 263, 526, 789 e 1052Hz) com um ruído branco gaussiano, tendo uma razão sinal-ruido de 0 dB. O sinal foi gerado a uma taxa de 44100 amostras por segundo e as DFTs de análise para todos os métodos tinham 1024 pontos.

A Tabela 4.4 mostra as freqüências instantâneas obtidas pelos métodos de Reatribuição de Freqüência, Diferença de Fase e DFT[1] para as quatro senóides mencionadas. Já a Tabela 4.5 mostra o refinamento gerado pelo Método Iterativo da Diferença de Fase.

As Tabelas 4.6 e 4.7 mostram o teste com uma razão sinal-ruído de 10 e 20 dB, respectivamente. Para o método iterativo da diferença de fase, mostra-se apenas o resultado da quinta iteração.

Tabela 4.4: Freqüência Instantânea obtida pelos métodos de Reatribuição de Freqüência, Diferença de Fase e DFT[1] para quatro senóides com ruído branco gaussiano de 0 dB, com uma janela de 1024 pontos. Entre parênteses está o desvio, em porcentagem.

Métodos	Freqüências (Hz)			
	263	526	789	1052
DFT	258,9041 (-1,5574)	517,8082 (-1,5574)	776,7123 (-1,5574)	1035,6164 (-1,5574)
Reatribuição de Freqüência	264,2377 (0,4706)	530,7722 (0,9073)	789,4344 (**0,0551**)	1053,1224 (0,1067)
Diferença de Fase	263,7216 (0,2744)	529,7708 (0,7169)	787,9213 (-0,1367)	1051,097 (**-0,0858**)
DFT[1]	263,2084 (**0,0792**)	524,409 (**-0,3025**)	789,5656 (0,0717)	1057,2461 (0,4987)

Tabela 4.5: Freqüência Instantânea obtida pelo método iterativo da diferença de fase para quatro senóides com ruído branco gaussiano de 0 dB para 5 iterações.

Iterações	Freqüências (Hz)			
	263	526	789	1052
1	263,7216	529,7708	787,9213	1051,097
2	263,9201	529,1467	787,212	1052,693
3	263,9249	525,9935	787,2474	1052,7709
4	263,925	529,1469	787,2456	1052,7745
5	263,925 (**0,3517**)	529,1469 (**0,5983**)	787,2457 (**-0,2223**)	1052,7747 (**0,0736**)

Tabela 4.6: Freqüência Instantânea obtida pelos métodos de Reatribuição de Freqüência, Diferença de Fase, Versão Iterativa e DFT[1] para quatro senóides com ruído branco gaussiano de 10 dB, com uma janela de 1024 pontos. Entre parênteses está o desvio, em porcentagem.

Métodos	Freqüências (Hz)			
	263	526	789	1052
DFT	258,9041 (-1,5574)	517,8082 (-1,5574)	776,7123 (-1,5574)	1035,6164 (-1,5574)
Reatribuição de Freqüência	263,3379 (0,1285)	527,1035 (0,2098)	790,6296 (0,2065)	1055,4835 (0,3311)
Diferença de Fase	262,8289 (-0,0651)	526,0952 (**0,0181**)	789,1155 (0,0146)	1053,4637 (0,1391)
Diferença de Fase Iterativa	262,8968 (**-0,0392**)	525,5512 (-0,0853)	788,9953 (**-5,9569e-004**)	1052,8164 (0,0776)
DFT[1]	262,5021 (-0,1893)	524,6494 (-0,2568)	788,5696 (-0,0546)	1052,2837 (**0,0270**)

Tabela 4.7: Freqüência Instantânea obtida pelos métodos de Reatribuição de Freqüência, Diferença de Fase, Versão Iterativa e DFT[1] para quatro senóides com ruído branco gaussiano de 20 dB, com uma janela de 1024 pontos. Entre parênteses está o desvio, em porcentagem.

Métodos	Freqüências (Hz)			
	263	526	789	1052
DFT	258,9041 (-1,5574)	517,8082 (-1,5574)	776,7123 (-1,5574)	1035,6164 (-1,5574)
Reatribuição de Freqüência	263,7354 (0,2796)	527,1854 (0,2254)	791,3765 (0,3012)	1054,4031 (0,2284)
Diferença de Fase	263,2272 (0,0864)	526,1786 (0,0340)	789,8631 (0,1094)	1052,3822 (0,0363)
Diferença de Fase Iterativa	263,1291 (**0,0491**)	526,1509 (**0,0287**)	789,3357 (**0,0425**)	1051,9430 (**-0,0054**)
DFT[1]	262,6565 (-0,1306)	525,8373 (-0,0309)	788,3081 (-0,0877)	1051,0246 (-0,0927)

4.6.3 Sinal Real

Foi escolhido como exemplo de gravação real, um trecho de violino solo composto por Yann Tiersen, que pode ser observado na Figura 4.2; mais especificamente, analisou-se a emissão de uma longa nota C5 de 523,25Hz. Foram analisadas a freqüência fundamental e suas harmônicas segunda (C6 - 1046,5Hz) e quarta (C7 - 2093Hz), que possuem os maiores picos na primeira janela, como pode ser observado na Figura 4.3.

Foram testados os três métodos (Reatribuição de Freqüência, Diferença de Fase e sua versão iterativa) com a seguinte configuração:

- janela: Hanning com 1024 pontos;
- deslocamento da janela[6]: 80 amostras;
- total de deslocamentos: 500;
- estimativas iniciais: 517,8082 Hz, 1035,6164 Hz e 2114,3836 Hz;

Para o método iterativo foi escolhido um total de cinco iterações. As estimativas iniciais para a primeira janela foram dadas de acordo com o máximo *bin* local encontrado pela DFT.

O método de Reatribuição de Freqüência é ilustrado nas Figuras de número 4.4 a 4.7. As Figuras 4.8 a 4.11 ilustram a análise pela Diferença de Fase. E, por fim, o método iterativo está representado nas Figuras 4.12 a 4.15. Todos os métodos são capazes de descrever a variação de freqüência ao longo do tempo.

[6]Não confundir este deslocamento com o método da Diferença de Fase que usa duas janelas consecutivas. Este é o deslocamento dinâmico, do inglês *hop size*.

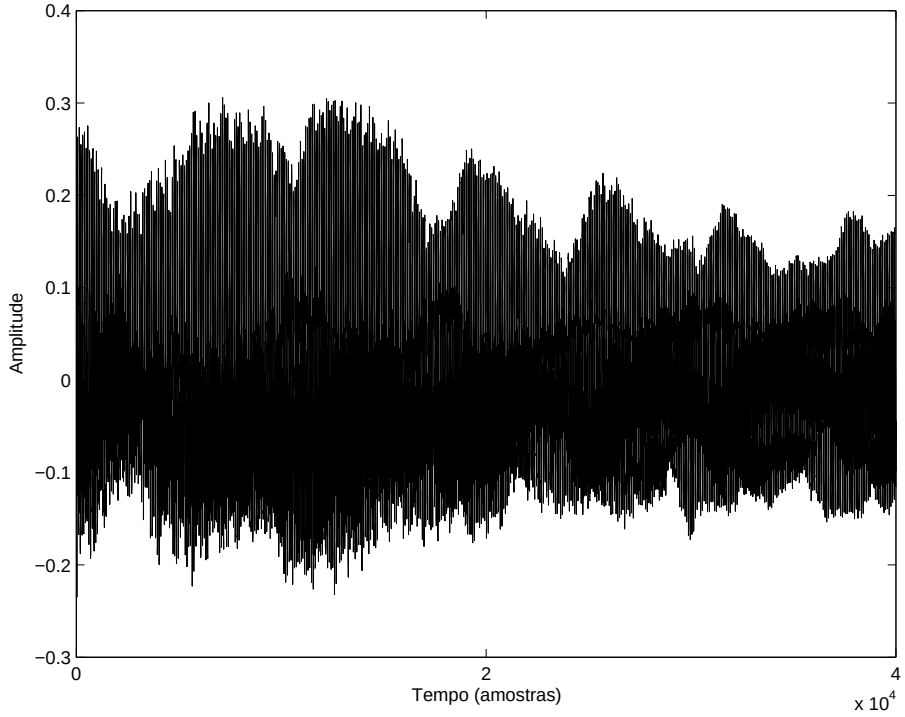

Figura 4.2: Sinal de gravação real de violino no domínio do tempo.

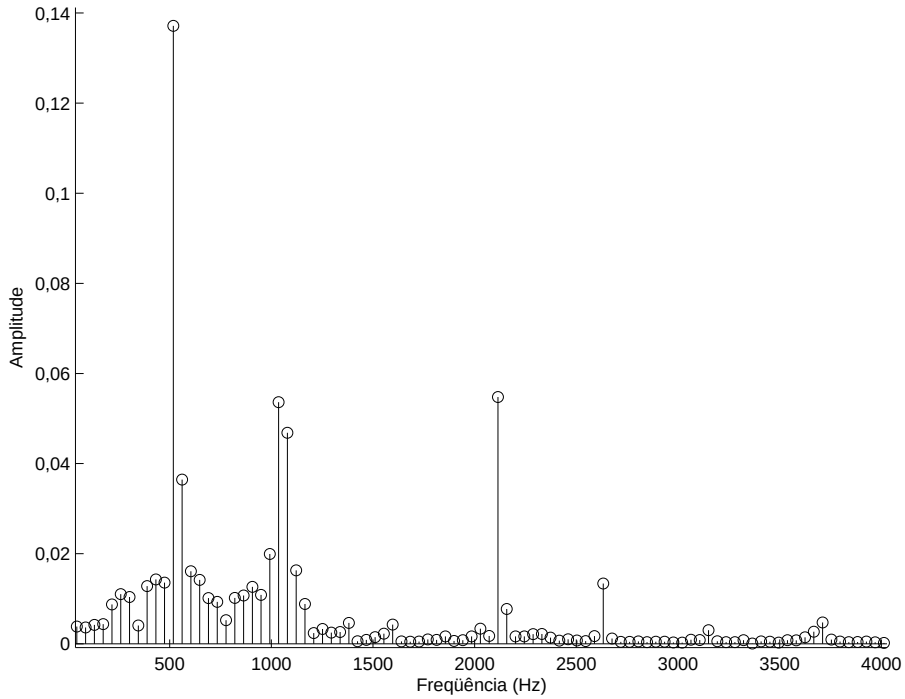

Figura 4.3: Espectro DFT da primeira janela do trecho de um violino tocando um C5 com vibrato.

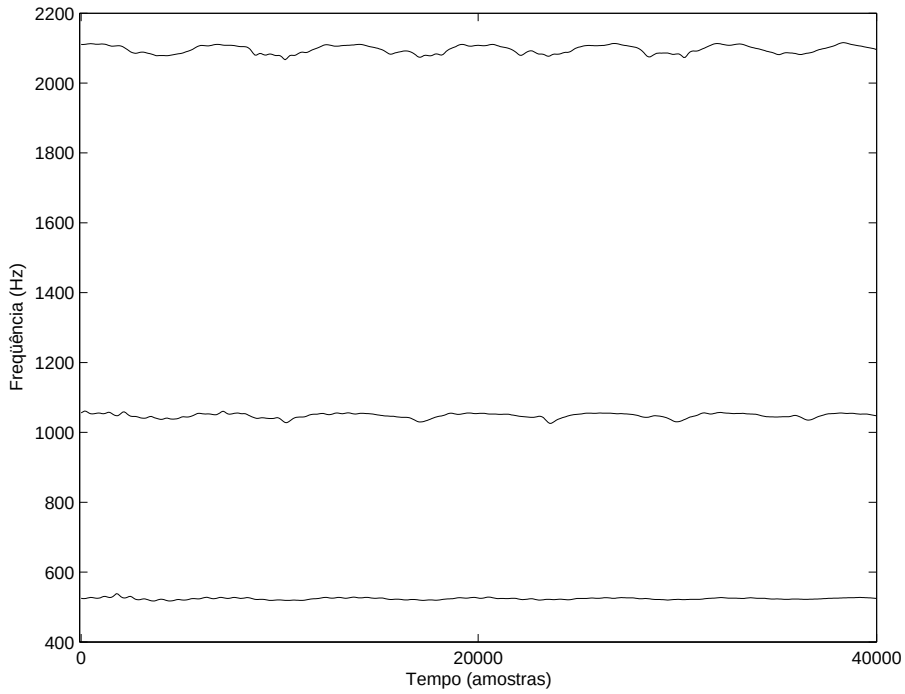

Figura 4.4: Linhas freqüenciais para o C5 do violino pelo método de Reatribuição de Freqüência.

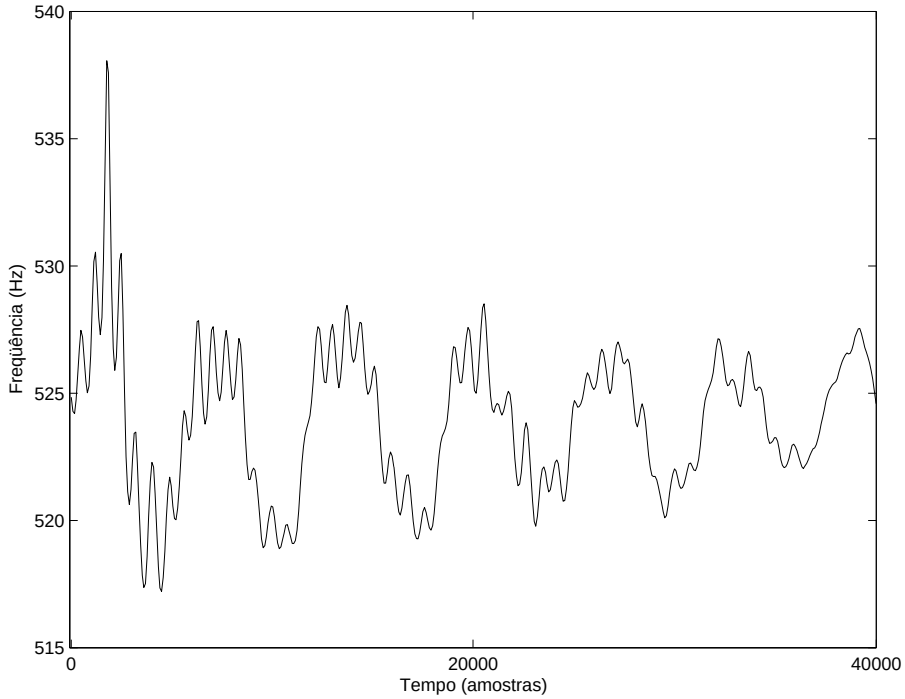

Figura 4.5: Linha freqüencial para o C5 (fundamental) do violino pelo método de Reatribuição de Freqüência.

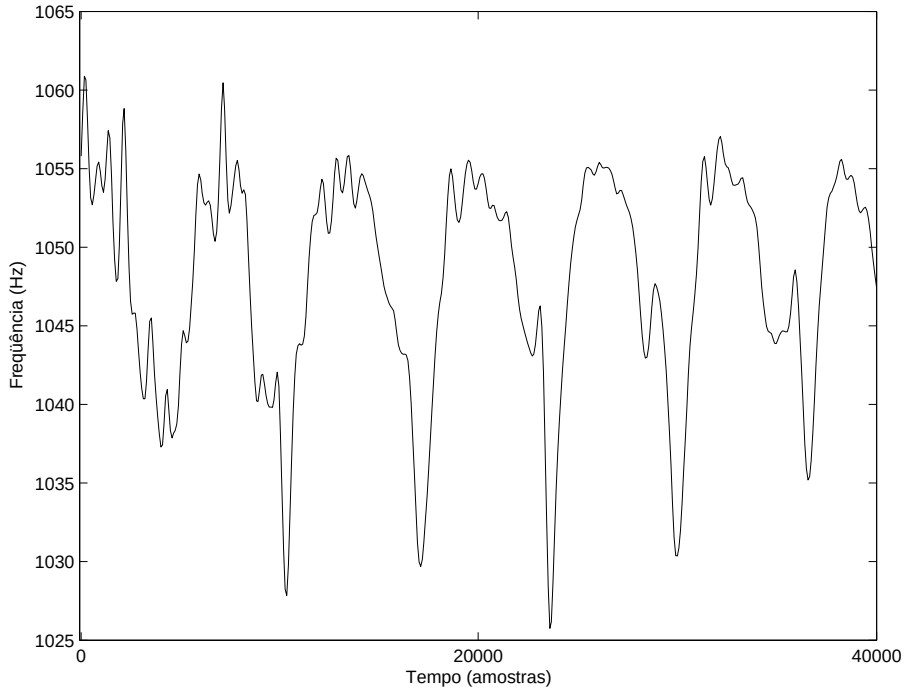

Figura 4.6: Linha freqüencial para o C6 (2ª harmônica) do violino pelo método de Reatribuição de Freqüência.

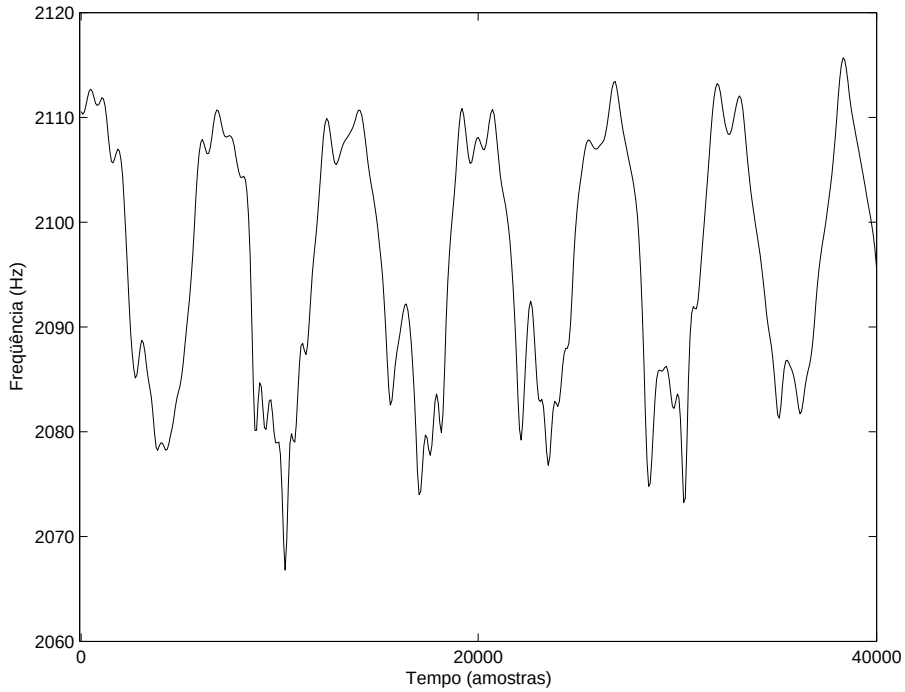

Figura 4.7: Linha freqüencial para o C7 (4ª harmônica) do violino pelo método de Reatribuição de Freqüência.

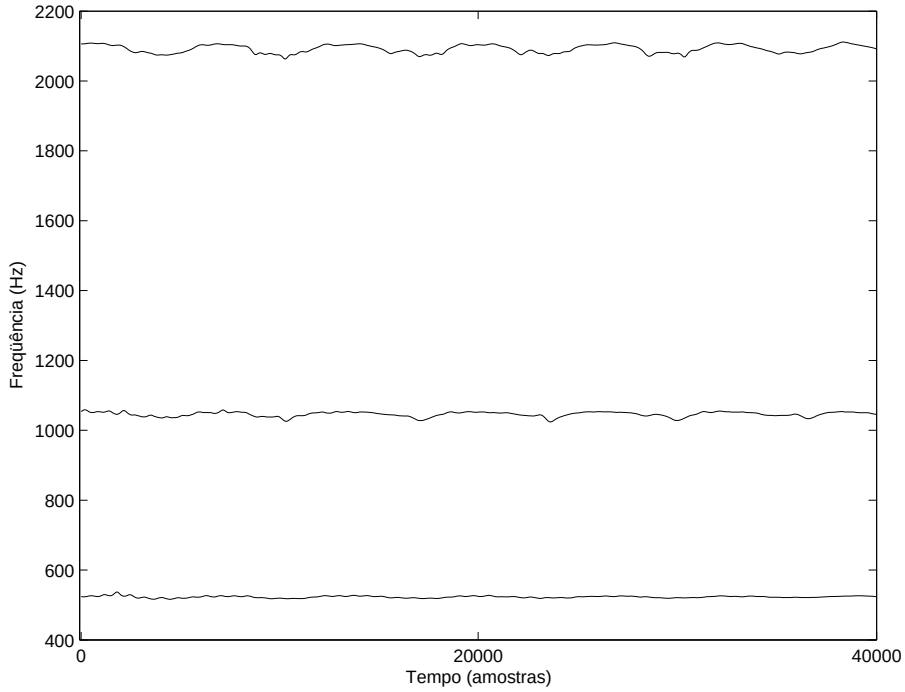

Figura 4.8: Linhas freqüenciais para o C5 do violino pelo método da Diferença de Fase.

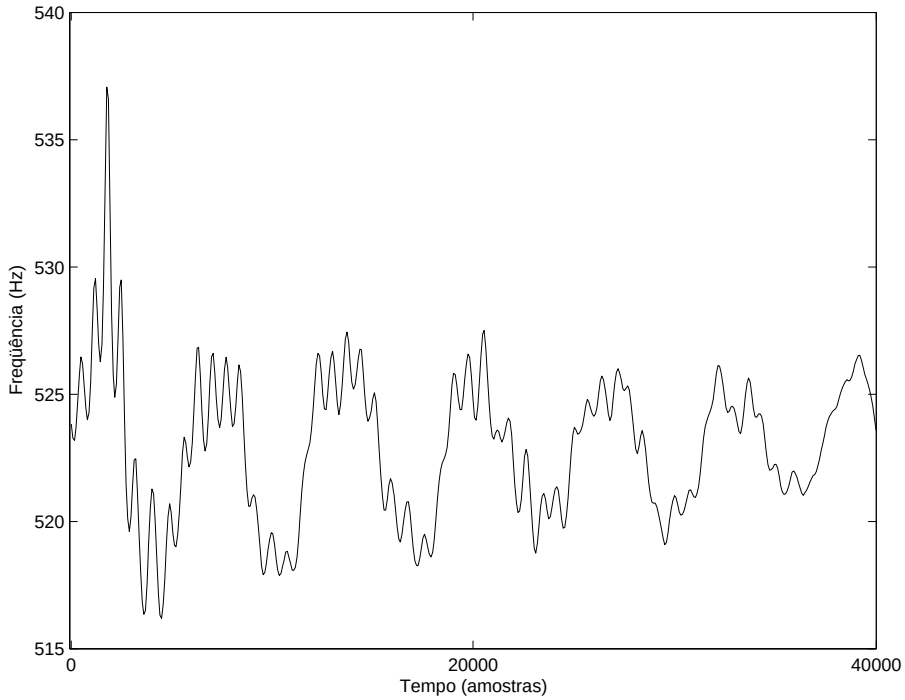

Figura 4.9: Linha freqüencial para o C5 (fundamental) do violino pelo método da Diferença de Fase.

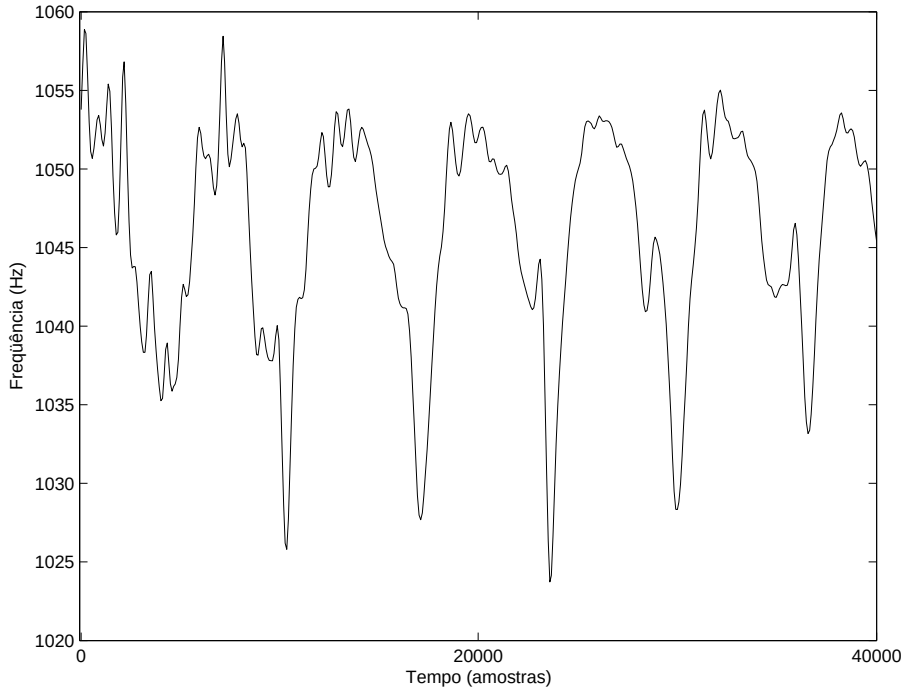

Figura 4.10: Linha freqüencial para o C6 (2ª harmônica) do violino pelo método da Diferença de Fase.

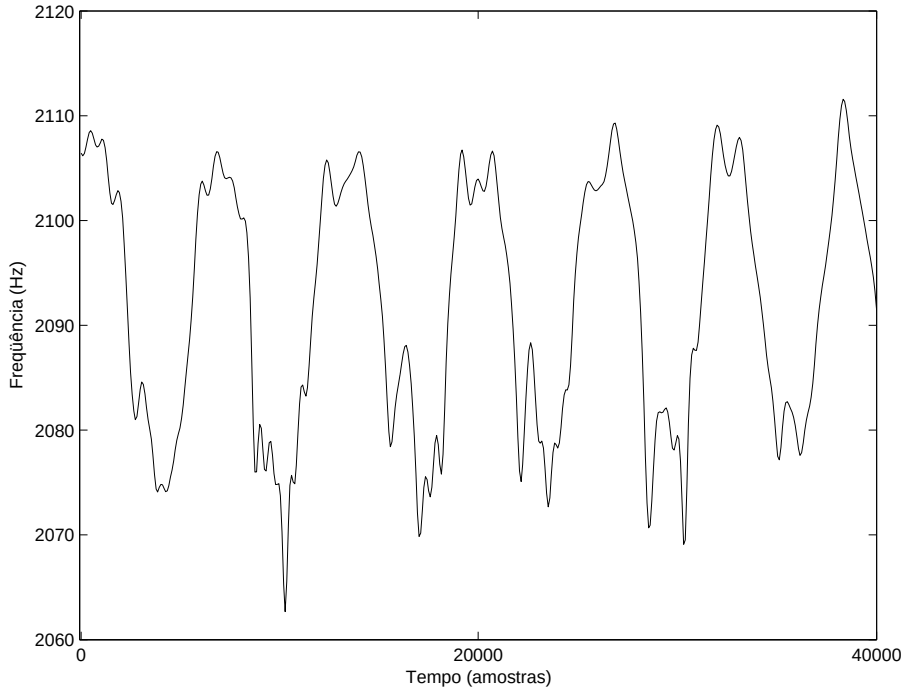

Figura 4.11: Linha freqüencial para o C7 (4ª harmônica) do violino pelo método da Diferença de Fase.

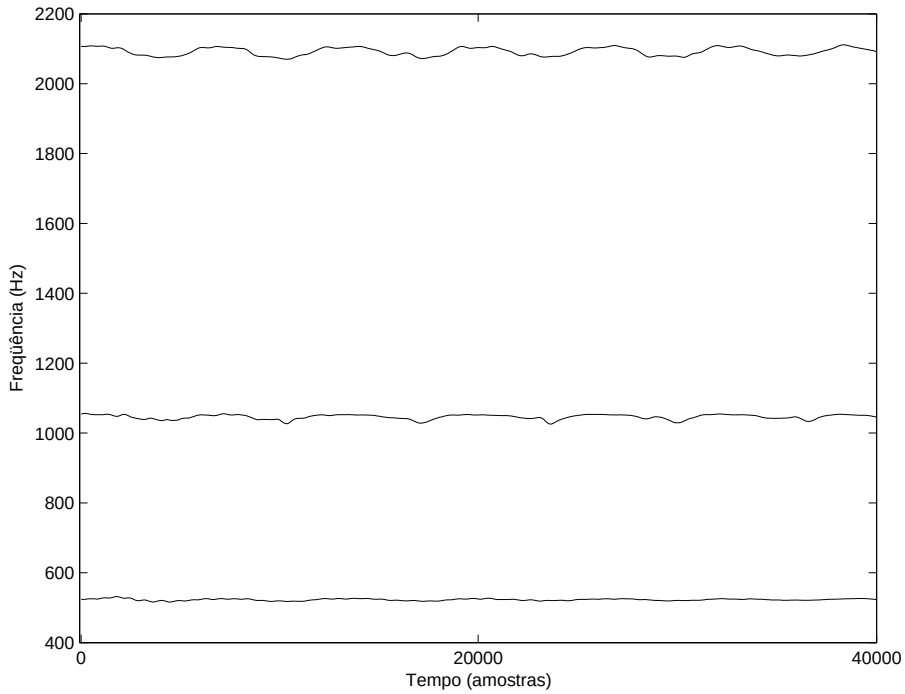

Figura 4.12: Linhas freqüenciais para o C5 do violino pelo método Iterativo da Diferença de Fase.

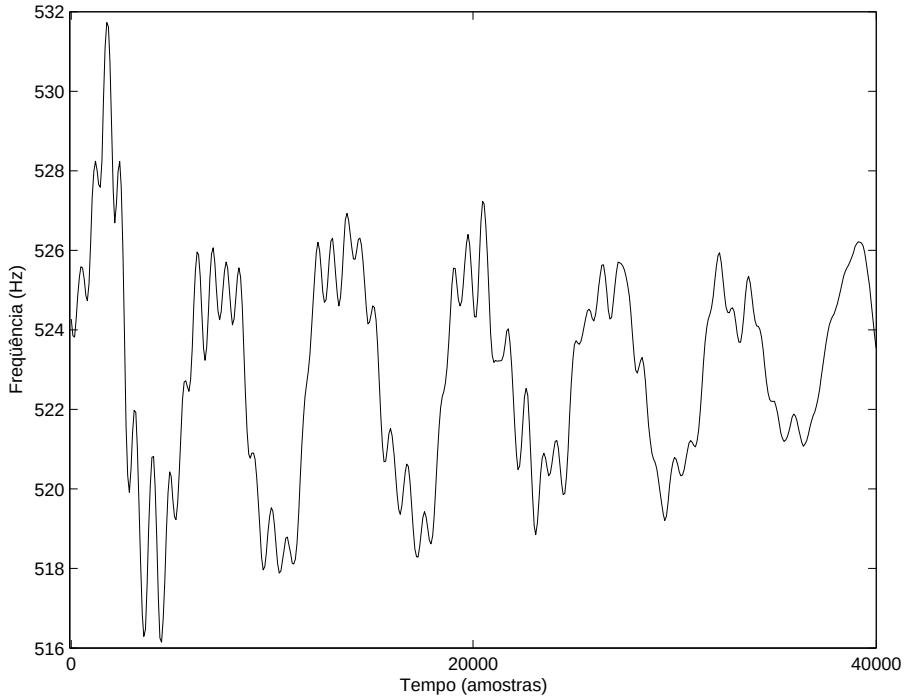

Figura 4.13: Linha freqüencial para o C5 (fundamental) do violino pelo método Iterativo da Diferença de Fase.

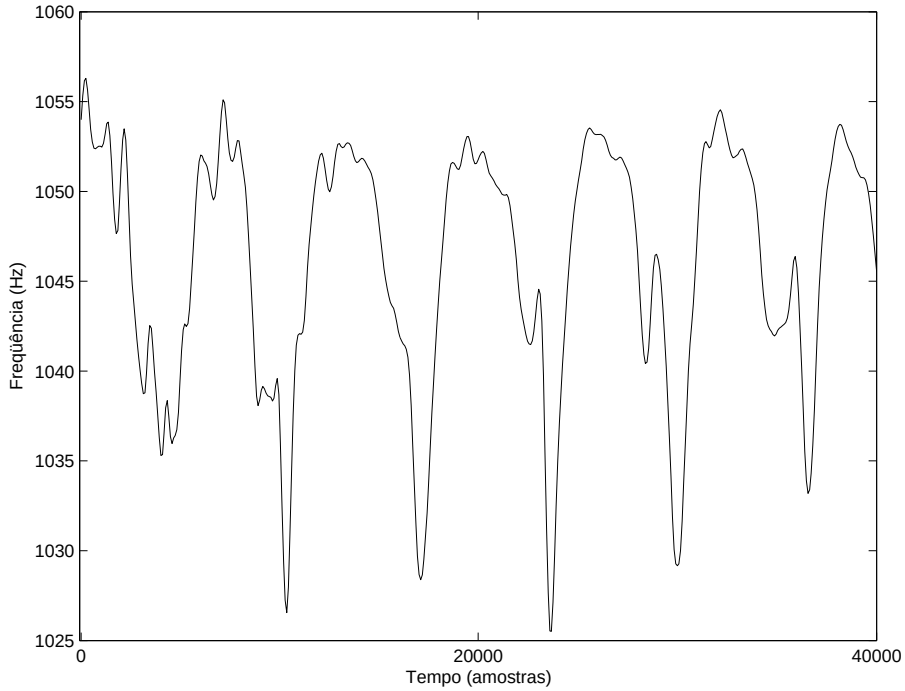

Figura 4.14: Linha freqüencial para o C6 (2ª harmônica) do violino pelo método Iterativo da Diferença de Fase.

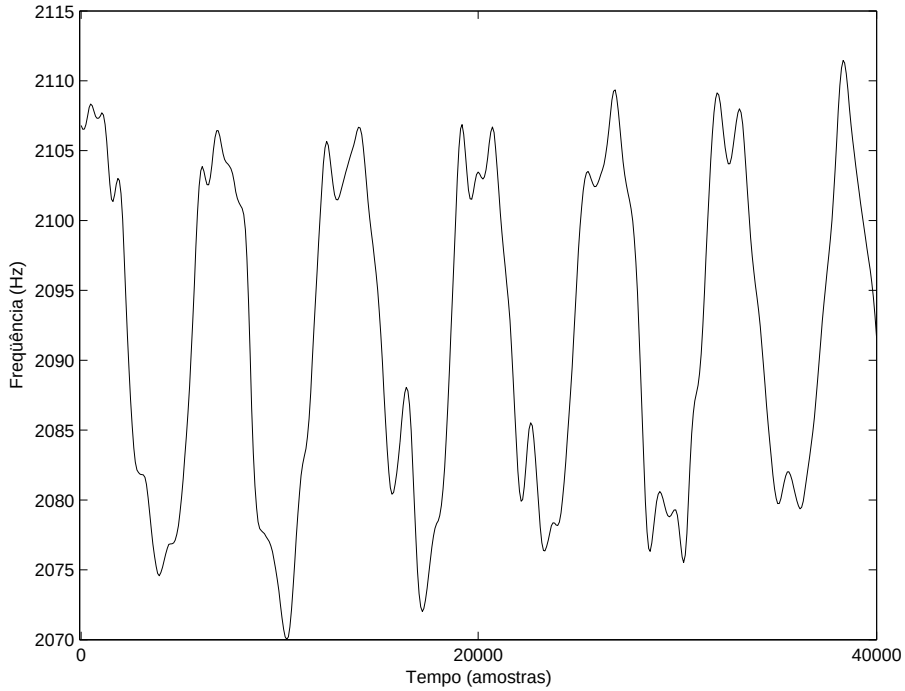

Figura 4.15: Linha freqüencial para o C7 (4ª harmônica) do violino pelo método Iterativo da Diferença de Fase.

4.7 Conclusões

O exemplo das senóides sintéticas sem adição de ruído, visto na Seção 4.6.1, atribui maior precisão ao método da diferença da fase, e principalmente à sua versão iterativa. O método simples (não-iterativo) da diferença de fase utiliza duas DFTs deslocadas de uma amostra, enquanto que a reatribuição de freqüência utiliza apenas uma DFT, mais especificamente, apenas os dados das três raias mais próximas. Como se pode esperar, o melhor resultado decorre de um aumento da complexidade. O método iterativo é mais preciso, porém eleva ainda mais a complexidade.

No caso da DFT[1], também são utilizadas duas DFTs, porém a correção da freqüência da equação (4.43) é feita sem levar em consideração que foi feito janelamento. Isso pode explicar a aproximação pior que no método da diferença de fase.

Nenhum dos métodos se mostrou robusto a ruídos quando a razão sinal-ruído era pior que 20 dB, como pode ser observado no exemplo da Seção 4.6.2, onde cada um dos três métodos foi o melhor em algum caso. Nesses casos, a DFT[1] parece ser um pouco mais robusta à presença de ruído que as demais. Porém, deve-se investigar melhor a sensibilidade dos métodos através de testes estatisticamente sistemáticos. Pela Tabela 4.6 observa-se que para sinais com razão sinal-ruído melhor que 20dB, o método iterativo da diferença de fase é o mais preciso.

Com o exemplo do sinal real visto na Seção 4.6.3, observa-se que tanto o método de Reatribuição de Freqüência como o da Diferença de Fase e sua versão iterativa conseguem descrever a trajetória das linhas de modo semelhante; no entanto, o método iterativo parece possuir maior precisão.

Capítulo 5

Síntese

Todos os métodos de refinamento espectral apresentados no Capítulo 4 podem ser utilizados para gerar as linhas musicais do modelo senoidal descrito no Capítulo 3. Tendo sido obtida uma boa análise do sinal, pode ser de interesse sintetizá-lo, o que é visto brevemente a seguir.

5.1 Síntese pelo Modelo Senoidal

Tendo as informações espectrais dos sinais (completo ou separado por instrumento musical) e os quadros devidamente unidos por trilhas, é possível sintetizá-los a partir dos parâmetros obtidos pela modelagem senoidal (freqüência, amplitude e fase). Basicamente realiza-se uma suave interpolação desses parâmetros, quadro a quadro, evitando assim descontinuidades nesse processo.

5.1.1 Algoritmo de síntese

Conforme explicado no Capítulo 4, os parâmetros calculados pela análise de Fourier e seus refinamentos indicam os valores referentes ao centro de cada quadro (*frame*). Dessa forma, para dois quadros consecutivos k e $k+1$, os parâmetros $(\hat{A}_l^k, \hat{\omega}_l^k, \hat{\theta}_l^k)$ e $(\hat{A}_l^{k+1}, \hat{\omega}_l^{k+1}, \hat{\theta}_l^{k+1})$ indicam o valor central de cada quadro. Considerando que o deslocamento do quadro seja de S amostras[1], o sinal sintetizado $\tilde{s}[n]$ entre os quadros k e $k+1$ é calculado pela seguinte equação [4]:

[1] O deslocamento deve ser, de preferência, de $N/2$ amostras (metade do quadro).

$$\tilde{s}[n] = \sum_{l=1}^{L^k} \hat{A}_l[n] \cos\left(\hat{\theta}_l[n]\right), \tag{5.1}$$

sendo L^k o total de trilhas do quadro k, onde a amostra $n = 0$ representa o centro do quadro k, com seus parâmetros $(\hat{A}_l^k, \hat{\omega}_l^k, \hat{\theta}_l^k)$ referentes à l-*ésima* trilha e a amostra $n = S$ utiliza os parâmetros do quadro $k + 1$. Agora deve-se obter as interpolações das S amostras referentes à amplitude $\tilde{A}_l[n]$ e à fase $\tilde{\theta}_l[n]$.

A interpolação da amplitude é obtida pela seguinte equação:

$$\tilde{A}[n] = \hat{A}^k + \frac{(\hat{A}^{k+1} - \hat{A}^k)}{S} n, \tag{5.2}$$

onde o índice l, referente à trilha, foi omitido por conveniência. Essa interpolação, descrita na equação (5.2), é linear, onde $\tilde{A}[0] = \hat{A}^k$, $\tilde{A}[S-1] \approx \hat{A}^{k+1}$ e $\tilde{A}[S] = \hat{A}^{k+1}$. Como, em sinais de áudio, os valores das amplitudes variam lentamente ao longo do tempo, a abordagem linear é satisfatória para sintetizar as trilhas.

Já a interpolação da freqüência e da fase não é tão trivial, pois a fase medida $\hat{\theta}$ é obtida em módulo 2π. Então deve-se desdobrar a fase para garantir que as trilhas freqüenciais tenham uma transição suave entre os quadros. Para se obter o valor da fase instantânea são necessárias quatro variáveis, sendo elas a freqüência e a fase das duas janelas analisadas ($\hat{\omega}_l^k$, $\hat{\theta}^k$, $\hat{\omega}^{k+1}$ e $\hat{\theta}^{k+1}$). Enquanto a interpolação linear possui apenas um grau de liberdade, nesse caso precisa-se de pelo menos três graus de liberdade. O primeiro passo para resolver esse problema é definir uma função de interpolação de fases que é polinomial cúbica [18]:

$$\tilde{\theta}(t) = \zeta + \gamma t + \alpha t^2 + \beta t^3. \tag{5.3}$$

É conveniente tratar a função de fase em tempo contínuo t, com $t = 0$ correspondendo ao quadro k (novamente, ao centro do quadro) e $t = T$ correspondendo ao quadro $k + 1$. É necessário que a função de fase cúbica da equação (5.3) seja igual às fases das duas janelas no seu limite; o mesmo vale para a sua derivada, sabendo que a derivada da fase é a freqüência, que deve corresponder às freqüências obtidas nos dois quadros.

Analisando a freqüência instantânea como derivada da fase, e usando a definição da equação (5.3), tem-se:

$$\frac{d\tilde{\theta}}{dt}(t) = \dot{\tilde{\theta}}(t) = \gamma + 2\alpha t + 3\beta t^2. \tag{5.4}$$

Então, para o instante $t = 0$,

$$\begin{aligned} \tilde{\theta}(0) &= \zeta = \hat{\theta}^k \\ \dot{\tilde{\theta}}(0) &= \gamma = \hat{\omega}^k; \end{aligned} \tag{5.5}$$

e em $t = T$,

$$\begin{aligned} \tilde{\theta}(T) &= \hat{\theta}^k + \hat{\omega}^k T + \alpha T^2 + \beta T^3 = \hat{\theta}^{k+1} + 2\pi M \\ \dot{\tilde{\theta}}(T) &= \hat{\omega}^k + 2\alpha T + 3\beta T^2 = \hat{\omega}^{k+1}. \end{aligned} \tag{5.6}$$

Como a fase $\hat{\theta}^{k+1}$ é medida em módulo 2π, é necessário aumentá-la pelo termo $2\pi M$ (com M inteiro) para que a função da freqüência seja maximamente suave. A variável M ainda é desconhecida, mas para cada valor de M é possível resolver a função para $\alpha(M)$ e $\beta(M)$ da seguinte forma:

$$\begin{bmatrix} \alpha(M) \\ \beta(M) \end{bmatrix} = \begin{bmatrix} \frac{3}{T^2} & \frac{-1}{T} \\ \frac{-2}{T^3} & \frac{1}{T^2} \end{bmatrix} \begin{bmatrix} \hat{\theta}^{k+1} - \hat{\theta}^k - \hat{\omega}^k T + 2\pi M \\ \hat{\omega}^{k+1} - \hat{\omega}^k \end{bmatrix}. \tag{5.7}$$

Para determinar M, deve-se definir o conceito de transição maximamente suave. A figura 5.1 ilustra um conjunto de funções cúbicas para alguns valores de M. Por intuição, a melhor função escolhida é a de menor variação. Se as freqüências fossem constantes e o sinal estacionário, a função da fase seria linear.

Tem-se, então, que um critério razoável para uma transição maximamente suave é utilizar a segunda derivada da fase $\ddot{\tilde{\theta}}(t)$, ou seja, escolher um M tal que a função

$$f(M) = \int_0^T [\ddot{\tilde{\theta}}(t; M)]^2 dt \tag{5.8}$$

seja mínima.

Apesar de M ser um inteiro, o problema pode ser mais facilmente resolvido minimizando-se $f(x)$ com respeito à variável contínua x e, então, escolhendo-se M como o inteiro mais próximo de x. Após certa álgebra [4], conclui-se que o valor de x que minimiza a função é dado por

$$x^* = \frac{1}{2\pi}[(\hat{\theta}^k + \hat{\omega}^k T - \hat{\theta}^{k+1}) + (\hat{\omega}^{k+1} - \hat{\omega}^k)\frac{T}{2}], \tag{5.9}$$

de onde é então determinado M^*; este é usado na equação (5.7) para determinar $\alpha(M^*)$ e $\beta(M^*)$ e, por conseqüência, a função de interpolação de fase

$$\tilde{\theta}(t) = \hat{\theta}^k + \hat{\omega}^k t + \alpha(M^*)t^2 + \beta(M^*)t^3. \tag{5.10}$$

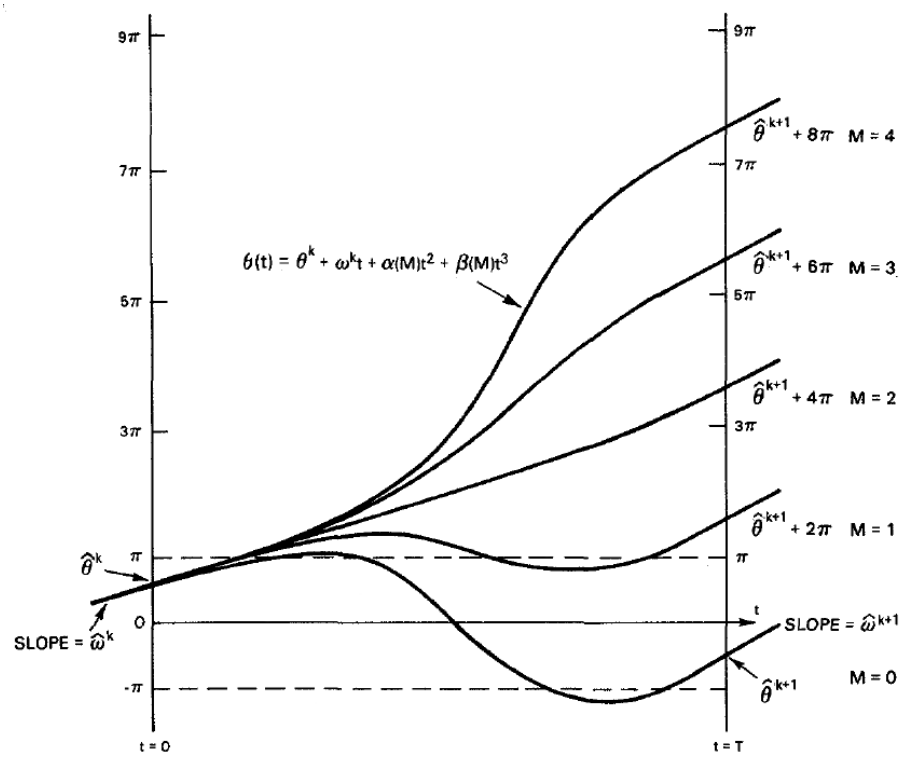

Figura 5.1: Exemplo de funções cúbicas de interpolação de fase para um número de valores de M.

Como a análise começou com a consideração de que a fase desdobrada $\hat{\theta}^k$ corresponde à freqüência $\hat{\omega}^k$ referente ao meio do quadro k, é necessário especificar a inicialização do procedimento de interpolação de fase. Assim, se uma trilha foi detectada no quadro $k+1$ com os parâmetros $(\hat{A}_l^{k+1}, \hat{\omega}_l^{k+1}, \hat{\theta}_l^{k+1})$, são definidos no quadro k, para a trilha correspondente, amplitude igual a zero ($A^k = 0$), e mesma freqüência ($\hat{\omega}^k = \hat{\omega}^{k+1}$); a fase desdobrada no início do quadro k é definida como

$$\hat{\theta}^k = \hat{\theta}^{k+1} - \hat{\omega}^{k+1}\frac{hop}{N}, \tag{5.11}$$

onde *hop* é a distância, em amostras, entre os quadros e N é o tamanho do quadro[2].

Com esse esquema de desdobramento da fase, cada trilha freqüencial terá associada uma fase instantânea de acordo com as rápidas mudanças de fase (freqüência) e com as transições mais lentas.

Obtem-se, então, a versão discreta $\tilde{\theta}[n]$ de $\tilde{\theta}(t)$ na equação (5.10), que será substituída em (5.1).

[2]Isso já foi explicado no item 3.2 pela equação (3.12).

Parte II

Soluções por Bancos de Filtros

Capítulo 6

Técnicas Baseadas em Bancos de Filtros Muito Seletivos para Análise Dinâmica de Sinais Musicais

Uma alternativa à transformada em blocos (apresentada na primeira parte) para o detalhamento do sinal no espectro é a utilização de bancos de filtros de banda estreita. Assim, embora a saída de cada filtro esteja no domínio do tempo, o uso de canais finos faz com que ela equivalha a um *bin* na transformada em blocos. É possível converter as transformadas em blocos de tamanhos constantes em bancos de filtros. Os codificadores perceptivos de áudio do padrão MPEG-1 [37], por exemplo, utilizam a transformada para obter o modelo perceptivo e o banco de filtros para transformar o sinal antes da quantização.

Na Tabela 6.1 tem-se um quadro comparativo entre algumas ferramentas de análise, tanto por transformada como por banco de filtros [38]. A *s*FFT, explicada no item 6.1.1, é a DFT vista como um banco de filtros. Possui baixa complexidade computacional, porém sua resolução é linear e seus filtros são pouco seletivos. A BQT (do inglês *Bounded-Q Transform*) realiza uma DFT para cada oitava, onde as oitavas inferiores são obtidas a partir de decimações do sinal; com isso, dobra-se a resolução a cada oitava inferior. A CQT (do inglês *Constant-Q Transform*) é equivalente a uma DFT com distribuição geométrica dos filtros ao longo do espectro.

Isso é obtido mantendo-se o fator de qualidade $Q = \Delta f / f$ constante. A grande desvantagem dessa ferramenta é sua alta complexidade computacional, como explicado em [38].

A FFB, explicada no item 6.1.2, utiliza os conceitos da sFFT, porém substituindo os filtros de baixa seletividade por outros de maior ordem. A CQFFB é uma CQT que utiliza os filtros mais seletivos da FFB, mas que atinge grande complexidade computacional; será discutida no item 6.2.2. A BQFFB é uma BQT, com FFB, e será explicada detalhadamente no item 6.3.2.

Tabela 6.1: Comparação entre as diferentes técnicas de análise referentes a resolução, seletividade e complexidade. O asterisco indica os métodos baseados na FFB, que possuem uma complexidade maior que os baseados na FFT.

Ferramenta	**Resolução**	**Seletividade**	**Complexidade**
sFFT	linear	baixa	baixa
FFB	linear	alta	baixa (*)
CQT	geométrica	baixa	alta
CQFFB	geométrica	alta	alta (*)
BQT	linear por oitava	baixa	média
BQFFB	linear por oitava	alta	média (*)

6.1 Métodos com separação linear entre os canais

6.1.1 *s*FFT

A transformada em blocos de Fourier, descrita na Seção 4.1, também pode ser vista como um banco de filtros. Esses filtros são igualmente espaçados ao longo do espectro e possuem baixa seletividade. Partindo-se da versão dinâmica da DFT no domínio z [39], tem-se:

$$s\mathrm{DFT}(z) = \frac{1}{N}\sum_{i=0}^{N-1} z^{-i} W_N^{ki} = \frac{1-(z^{-1}W_N^k)^N}{1-z^{-1}W_N^k}, \tag{6.1}$$

onde $W_N^k = e^{-\frac{j2\pi k}{N}}$. Assim, para cada canal k, a sDFT pode ser representada como uma cascata de filtros de mesmo formato

$$s\mathrm{DFT}_k(z) = \frac{1}{N}\prod_{i=0}^{L-1}\left[1 + (z^{-1}W_N^k)^{2i}\right]. \tag{6.2}$$

Para uma DFT com N filtros, sendo N potência de 2, é possível aproveitar algumas propriedades matemáticas e chegar à borboleta da FFT. Dessa forma são utilizadas operações em cascata que visam a minimizar a complexidade computacional [40]. Observando-a então como banco de filtros, denomina-se a FFT com o prefixo s de *sliding*, ou seja, sFFT. Assim, ao invés de N filtros de ordem elevada é possível utilizar vários sub-filtros de ordem reduzida interconectados. O sub-filtro protótipo p é de primeira ordem:

$$G_{s\mathrm{FFT}_p}(z) = 1 + z^{-1}. \tag{6.3}$$

O filtro correspondente a um dado nível da cascata, i, e seu respectivo canal j, é obtido a partir desse protótipo, apenas substituindo-se z por $z^{2^i}W_N^j$ Assim

$$G^{i,j}(z) = 1 + (z^{-2^i}W_N^j), \tag{6.4}$$

onde

$$j = [(2^i k)\mathrm{mod}N]. \tag{6.5}$$

Para um exemplo com $N = 4$ canais, a resposta no domínio z para cada canal é

$$\begin{aligned}
\frac{X_0(z)}{X(z)} &= \frac{1}{N}(1 + W_4^0 z^{-1})(1 + W_4^0 z^{-2}) \\
\frac{X_1(z)}{X(z)} &= \frac{1}{N}(1 + W_4^1 z^{-1})(1 - W_4^0 z^{-2}) \\
\frac{X_2(z)}{X(z)} &= \frac{1}{N}(1 - W_4^0 z^{-1})(1 + W_4^0 z^{-2}) \\
\frac{X_3(z)}{X(z)} &= \frac{1}{N}(1 - W_4^1 z^{-1})(1 - W_4^0 z^{-2}).
\end{aligned} \tag{6.6}$$

A propriedade matemática que demonstra que, no círculo complexo, a parte negativa pode ser alcançada também deslocando-se a metade do círculo no sentido positivo, é formulada da seguinte forma:

$$-W_N^k = W_N^{k+\frac{N}{2}}. \tag{6.7}$$

Aplicando essa propriedade nas equações (6.6), obtêm-se

$$\begin{aligned}
\frac{X_0(z)}{X(z)} &= \frac{1}{N}(1+W_4^0 z^{-1})(1+W_4^0 z^{-2}) \\
\frac{X_1(z)}{X(z)} &= \frac{1}{N}(1+W_4^1 z^{-1})(1+W_4^2 z^{-2}) \\
\frac{X_2(z)}{X(z)} &= \frac{1}{N}(1+W_4^2 z^{-1})(1+W_4^0 z^{-2}) \\
\frac{X_3(z)}{X(z)} &= \frac{1}{N}(1+W_4^3 z^{-1})(1+W_4^2 z^{-2}). \qquad (6.8)
\end{aligned}$$

A estrutura descrita acima pode ser vista na Figura 6.1. Nela, os filtros mais seletivos são realizados a partir de filtros menos seletivos, porém acoplados em cascata, o que, como dito anteriormente, reduz a complexidade computacional.

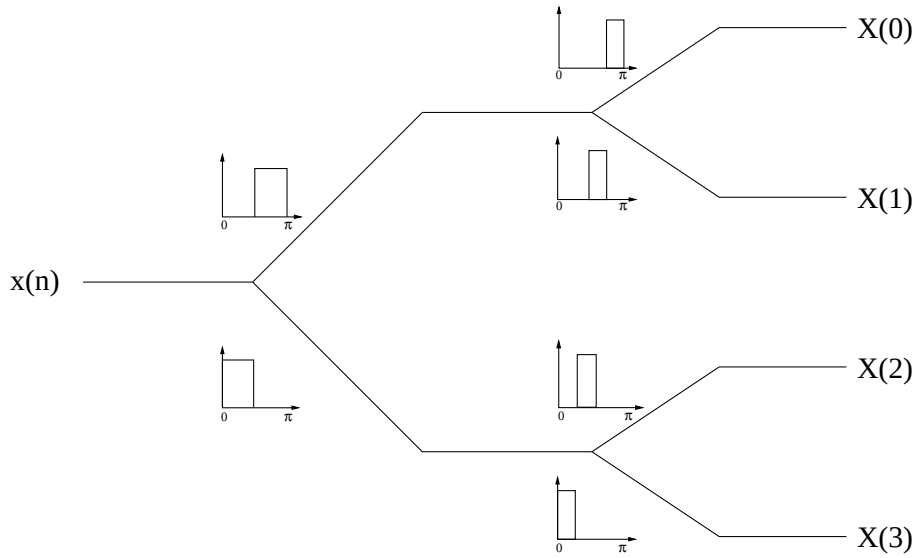

Figura 6.1: Estrutura da *s*FFT como banco de filtros.

6.1.1.1 Seletividade

Na Figura 6.2 (a) observa-se o módulo da resposta em freqüência do canal 7 de uma *s*FFT com 64 canais. Fica evidente como a banda passante modifica o valor real da amplitude devido à sua forma abaulada. O fato de a atenuação mínima da banda de rejeição ser de apenas 13dB é outro fator que afeta a qualidade da análise freqüencial. É usual utilizar-se um janelamento não-retangular, o que traz uma ligeira melhora na atenuação da banda de rejeição, porém o lobo central aumenta, o que prejudica ainda mais a análise seletiva, já que passa a ocorrer maior interseção entre filtros adjacentes.

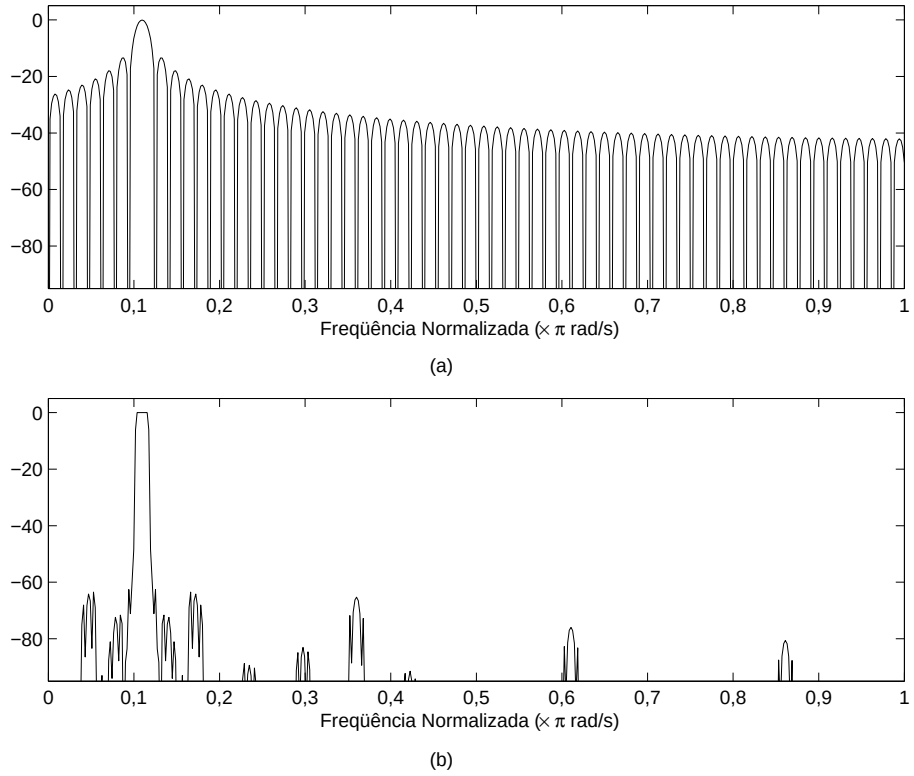

Figura 6.2: Módulo da resposta do 7º canal dos bancos de filtros de 64 canais: (a) sFFT; (b) FFB.

6.1.2 FFB

Procurando melhorar a baixa seletividade da sFFT, Lim e Farhang-Boroujeny desenvolveram o Banco de Filtros Rápido [41], ou FFB, do inglês *Fast Filter Bank*. Aproveitando a mesma estrutura da sFFT, porém utilizando sub-filtros protótipos de ordem maior e diferentes a cada nível i da cascata, consegue-se aumentar a seletividade. A ordem dos sub-filtros protótipos $G^i_{FFB_p}(z)$ diminui à medida que se adentram os níveis i do filtro. Para se obter os sub-filtros da FFB, faz-se $G^{i,j}_{FFB}(z) = G^i_{FFB_p}(zW^k_N)$. Esse aumento na ordem dos filtros tende a elevar a complexidade computacional. Na tentativa de compensar isso, projetam-se filtros $G^i_{FFB_p}$ de meia-banda com resposta ao impulso simétrica e de comprimento ímpar, i.e.

$$
\begin{aligned}
G_{FFB}(z) &= \sum_{n=-\infty}^{\infty} g_{FFB_p}[n] z^{-n} \\
g_{FFB}[n] &= g_{FFB}[-n] \\
g_{FFB}[n] &= 0, \text{ se } n \neq 0 \text{ e } n \text{ par} \\
g_{FFB}[0] &= 1.
\end{aligned} \tag{6.9}
$$

O projeto desses filtros segue o método FRM, do inglês *Frequency Response Masking*. Essa técnica visa a obter filtros com uma faixa de transição pequena e baixa complexidade. Ela parte do princípio de que a resposta de um filtro interpolado $H(z^L)$ é composta por L réplicas periódicas da resposta do filtro $H(z)$ comprimidas por L. Cada réplica possui uma faixa de transição L vezes mais íngreme que a de $H(z)$. É possível projetar um filtro mascarador $H_m(z)$ de complexidade moderada para suprimir as réplicas indesejadas da resposta do filtro intepolado, deixando apenas as bandas passantes de interesse.

A estrutura do banco de filtros da Figura 6.1 se presta perfeitamente a um projeto FRM. A cada nível, um dado filtro interpolado é mascarado pela cascata de filtros subseqüente. O processo recorrente de otimização da estrutura leva naturalmente a filtros protótipos diferentes. Por exemplo, os primeiros sete sub-filtros protótipos $G^i_{FFB_p}$ descritos em [42] e [43] são listados nas equações (6.10):

$$
\begin{aligned}
G^{0,j}_{FFB_p}(z) &= 1 + 0,6275(z + z^{-1}) - 0,1862(z^3 + z^{-3}) + 0,0878(z^5 + z^{-5}) - \\
&- 0,0426(z^7 + z^{-7}) + 0,0186(z^9 + z^{-9}) - 0,0067(z^{11} + z^{-11}) \\
G^{1,j}_{FFB_p}(z) &= 1 + 0,6209(z + z^{-1}) - 0,1688(z^3 + z^{-3}) + 0,0659(z^5 + z^{-5}) - \\
&- 0,0229(z^7 + z^{-7}) + 0,0055(z^9 + z^{-9}) \\
G^{2,j}_{FFB_p}(z) &= 1 + 0,5738(z + z^{-1}) - 0,0753(z^3 + z^{-3}) \\
G^{3,j}_{FFB_p}(z) &= 1 + 0,5654(z + z^{-1}) - 0,0654(z^3 + z^{-3}) \\
G^{4,j}_{FFB_p}(z) &= 1 + 0,5013(z + z^{-1}) \\
G^{5,j}_{FFB_p}(z) &= 1 + 0,5003(z + z^{-1}) \\
G^{6,j}_{FFB_p}(z) &= 1 + 0,5001(z + z^{-1}) \\
G^{7,j}_{FFB_p}(z) &= 1 + 0,5000(z + z^{-1})
\end{aligned} \tag{6.10}
$$

6.1.2.1 Seletividade

As especificações do filtro resultante em [42] estão indicadas na Tabela 6.2.

Tabela 6.2: Especificações do filtro da FFB com dados normalizados para $F_s = 1$

Faixa de passagem	0 a 0,185
Faixa de rejeição	0,315 a 0,5
Ripple na faixa de passagem	0,0139 dB
Atenuação na faixa de rejeição	56 dB

Na Figura 6.2 pode-se comparar a seletividade do filtro do canal 7 da sFFT com o da FFB, ambas com um total de 64 canais. As duas grandes vantagens da FFB em relação à sFFT são a possibilidade de se obter uma banda passante extremamente plana e uma atenuação muito elevada na banda de rejeição.

6.2 Métodos com separação geométrica entre os canais

Definiu-se para esta tese que os cálculos do fator de qualidade

$$Q = \frac{f_k}{\Delta f_k} \tag{6.11}$$

dos filtros terão a freqüência média f_k definida de forma geométrica, isto é,

$$f_k = \sqrt{f_0 f_1} \tag{6.12}$$

e

$$\Delta f_k = f_1 - f_0, \tag{6.13}$$

onde f_0 e f_1 são as freqüências-limite inferior e superior (respectivamente) da banda passante[1].

[1]Convencionou-se que nas extremidades da banda passante ocorre uma queda de 6dB na resposta em freqüência do filtro, de forma a permitir que a soma das respostas adjacentes na intersecção seja 1.

Para se obter a razão freqüencial $r = f_1/f_0$ em função do fator de qualidade Q parte-se de

$$Q = \frac{f_k}{\Delta f_k} = \frac{\sqrt{f_0 f_1}}{f_1 - f_0}. \tag{6.14}$$

Como $f_1 = r f_0$,

$$Q = \frac{\sqrt{r f_0^2}}{r f_0 - f_0} = \frac{f_0\sqrt{r}}{f_0(r-1)} = \frac{\sqrt{r}}{r-1}. \tag{6.15}$$

Elevando-se ao quadrado ambos os lados e depois colocando-se r em evidência

$$Q^2 = \frac{r}{(r-1)^2} = \frac{r}{r^2 - 2r + 1} = \frac{r}{r(r - 2 + 1/r)}, \tag{6.16}$$

chegando-se à equação em r

$$r^2 - \left(2 + \frac{1}{Q^2}\right) r + 1 = 0. \tag{6.17}$$

Resolvendo-se a equação de segundo grau em r

$$r_{1,2} = \frac{\left(2 + \frac{1}{Q^2}\right) \pm \frac{1}{Q}\sqrt{4 + \frac{1}{Q^2}}}{2}; \tag{6.18}$$

como $f_1 > f_0$ na razão $r = \frac{f_1}{f_0}$, a solução r_2 não é válida.

A fórmula da razão geométrica em função do fator de qualidade é, então,

$$r = \frac{2 + \frac{1}{Q^2} + \frac{1}{Q}\sqrt{4 + \frac{1}{Q^2}}}{2}. \tag{6.19}$$

6.2.1 CQT

Em [44], Brown apresenta uma modificação na DFT visando a obter uma distribuição logarítmica no espectro, chamada de Transformada com Q Constante ou CQT (do inglês *constant-Q transform*). Com essa adaptação, o fator de qualidade Q de cada faixa de freqüência k de saída da transformada, dada por

$$Q = \frac{f_k}{\Delta f_k}, \tag{6.20}$$

onde f_k é a freqüência central e Δf_k é a largura da faixa, é mantido fixo ao longo do espectro.

Conforme já foi mostrado na Seção 4.1, a DFT é definida por:

$$X_{\mathrm{DFT}}[k] = \frac{1}{N}\sum_{n=0}^{N-1} x[n] e^{-jk\Omega_0 n}. \tag{6.21}$$

onde $x[n]$ é o sinal no tempo discreto, $X_{\mathrm{DFT}}[k]$ é a resposta (freqüencial) do canal k da DFT, $\Omega_0 = 2\pi/N$ é a freqüência fundamental, e N é o total de canais da DFT.

Na DFT, a largura Δf de cada canal de saída é constante, pois $\Delta f = F_s/N$, onde F_s é a freqüência de amostragem e N é o número de pontos do segmento de sinal sob análise. Já na CQT, como Q é diretamente proporcional a f_k, faz-se com que N seja variável em função de k, isto é,

$$N_k = \frac{F_s}{\Delta f_k} = Q\frac{F_s}{f_k}. \tag{6.22}$$

Desta forma a CQT é gerada substituindo-se f_k/F_s por Q/N_k na equação (6.21), obtendo assim [44]:

$$X_{CQ\mathrm{T}}[k] = \frac{1}{N_k}\sum_{n=0}^{N_k-1} w[k,n]x[n]e^{-j2\pi Qn/N_k}, \tag{6.23}$$

onde $x[n]$ é o sinal no tempo discreto e $w[k,n]$ é a função-janela.

É interessante utilizar um espaçamento entre os canais da CQT que permita distinguir ao menos dois semitons. Para isso, Brown [44] recomenda usar um espaçamento de quarto-de-tom, isto é, com os centros freqüenciais espaçados de $2^{1/24}$. Utiliza-se a raiz quadrada ao invés da metade ($2^{1/12}/2$), por se tratar de freqüências em progressão geométrica. Assim o fator de qualidade é

$$Q = \frac{f_k}{(\Delta f)_{CQ\mathrm{T}}} = \frac{f_k}{(2^{1/48} - 2^{-1/48})f_k} \approx \frac{1}{0,0289} \approx 34,6. \tag{6.24}$$

Em [45], é apresentado um algoritmo eficiente de implementação da CQT que diminui o custo computacional, que será descrito no item 6.4.2. Para implementar a CQT, escolhem-se o fator Q e as freqüências mínima $f_{\mathrm{mín}}$ e máxima $f_{\mathrm{máx}}$ a serem analisadas.

6.2.2 CQFFB

Uma forma de unir a alta seletividade dos filtros FFB com uma distribuição espectral que favorece a localização de notas musicais, i.e. geométrica, utilizando-se os princípios da CQT (explicada no item 6.2.1) é apresentada em [38], [46] e [39].

Vislumbraram-se duas formas de implementar a CQFFB. Em ambas, parte-se de um determinado canal FFB. O canal escolhido depende do fator Q desejado, e o total de canais FFB (N) deve ser tal que $N/2$ seja o menor inteiro maior que

ou igual ao canal Q (também inteiro). As características de simetria e meia-banda dos sub-filtros FFB são mantidas. Os demais canais podem ser obtidos de duas maneiras: reamostrando-se o canal Q (Algorítmo 2) —com isso ele mantém seu fator Q inalterado, porém perde as características de simetria e meia-banda— ou reamostrando-se o sinal de entrada (Algorítmo 1).

6.2.2.1 Primeiro Algoritmo: CQFFB

O algoritmo 1 é descrito a seguir ([38], [42]).

- Defina um fator Q inteiro para alcançar o nível de detalhamento espectral desejado com uma FFB de $N = 2^L$ canais, com L inteiro, tal que $N \geq 2Q$, e utilize o filtro correspondente ao $\lceil Q \rceil$;
- Para cada canal k da CQFFB:
  - Reamostre o sinal de entrada para obter uma taxa de amostragem

$$F_{s_k} = \frac{N}{Q} f_{c_{\text{mín}}} r^{k-1}, \tag{6.25}$$

    onde

$$r = \frac{2 + \frac{1}{Q^2} + \frac{1}{Q}\sqrt{4 + \frac{1}{Q^2}}}{2} \tag{6.26}$$

    é a razão freqüencial entre canais adjacentes (demonstrada nas equações (6.14) a (6.19)) e $f_{c_{\text{mín}}}$ é a freqüência central do canal $k = 1$.
  - Filtre a versão reamostrada do sinal de entrada pelo filtro FFB escolhido na primeira etapa.

Essa reamostragem do sinal para F_{s_k} desloca a faixa espectral desejada para a região de passagem do filtro Q. A principal desvantagem desse método é o seu elevado custo computacional (apresentado no item 6.4.5), além do fato de necessitar de filtros anti-*aliasing*.

6.2.2.2 Segundo Algoritmo: mCQFFB

O Algoritmo 2, apresentado em [39], foi chamado de mCQFFB (do inglês *modified*-CQFFB). Ao invés de reamostrar-se o sinal, reamostra-se o filtro protótipo $H_{\text{mCQFFB}_p}(z)$ com o Q desejado, que já não precisa ser um valor inteiro. Devido à

reamostragem, os filtros perdem as características de baixa complexidade da FFB, porém esta ocorre apenas no projeto, não na execução (como ocorre no algoritmo 1).

A mCQFFB é obtida com a seguinte implementação [42].

- Defina um fator Q inteiro para alcançar o nível de detalhamento espectral desejado com uma FFB de $N = 2^L$ canais, com L inteiro, tal que $N \geq 2Q$, e utilize o filtro correspondente ao $\lceil Q \rceil$;
- Para cada canal k da mCQFFB:
  - Reamostre a resposta ao impulso do filtro escolhido na primeira etapa para obter uma taxa de amostragem de acordo com a equação (6.25).
  - Filtre o sinal de entrada pela versão reamostrada do filtro FFB.

A reamostragem da resposta ao impulso do filtro para F_{s_k} desloca a faixa de passagem do filtro Q para a região espectral do sinal desejada. O cálculo da complexidade é apresentado no item 6.4.6.

6.3 Métodos com separação linear por oitavas

A idéia, aqui, consiste em separar inicialmente o espectro em oitavas geometricamente espaçadas e depois, em cada oitava, realizar uma divisão espectral linear com um algoritmo rápido de FFT ou FFB, dependendo da aplicação. O método é chamado, em inglês, de *bounded-Q* ou BQ, pois limita a variação do fator de qualidade Q. Em cada oitava, o *bin* ou canal com menor Q será o primeiro e o de maior Q, o último. O Q dos filtros do banco fica, assim, compreendendido entre esses dois valores.

Em uma separação de Q constante, dentro de uma oitava a banda do primeiro canal é metade da largura da banda do último canal, e essa variação se dá gradualmente de forma geométrica ao longo dos canais. Para se obter uma resolução de BQ equivalente a uma de R canais por oitava de CQ, parte-se da largura de banda de CQ de

$$BW_{CQ}[n] = f_0 \left[\left(\sqrt[R]{2} \right)^n - \left(\sqrt[R]{2} \right)^{n-1} \right], \tag{6.27}$$

onde f_0 é a freqüência inicial e $n = 1, \ldots, R$ é o índice do canal. No método BQ, para uma resolução linear de $N = 2^l$ canais por oitava, com l inteiro, a largura de banda é dada por

$$BW_{\mathrm{B}Q} = \frac{2f_0 - f_0}{N} = \frac{f_0}{N}. \tag{6.28}$$

Igualando-se $BW_{\mathrm{B}Q} = BW_{\mathrm{C}Q}[1]$ e resolvendo-se em função de N, obtém-se o número mínimo de canais de BQ cuja largura seja menor que ou igual à do canal mais seletivo ($n = 1$) de CQ:

$$N_{\mathrm{mín}} = 2^{\left\lceil \log_2\left(\frac{1}{\sqrt[R]{2}-1}\right)\right\rceil}. \tag{6.29}$$

A Tabela 6.3 mostra os limites freqüenciais para uma separação em dez oitavas (partindo-se da oitava mais alta, com taxa de amostragem de 44100Hz) e a largura de banda de cada canal em uma resolução de quarto-de-tom.

Tabela 6.3: Divisão do espectro de áudio em 10 oitavas, com taxa de amostragem de 44100 Hz.

Índice da oitava (d)	Freqüência inicial (Hz)	Freqüência final (Hz)	Largura de banda para quarto de tom (Hz)
10	11025	22050	459,4
9	5012,5	11025	229,7
8	2756,3	5012,5	114,8
7	1378,1	2756,3	57,4
6	689	1378,1	28,7
5	344,5	689	14,4
4	172,3	344,5	7,2
3	86,1	172,3	3,6
2	43	86,1	1,8
1	21,5	43	0,8

6.3.1 BQT

Uma outra maneira de usar a DFT com uma resolução no espectro mais eficiente na análise de sinais musicais é a BQT (do inglês *bounded-Q transform*)

proposta em [47], que possui um custo computacional (a ser mostrado no item 6.4.3) menor que a CQT. Nesta técnica, restringe-se a separação geométrica às oitavas, linearizando-se a análise no seu interior. Divide-se o espectro em oitavas e, dentro de cada uma, aplica-se uma FFT com a resolução adequada. Dessa forma obtem-se uma FFT por partes, onde a resolução cresce da oitava mais alta para a mais baixa, porém é fixa dentro de cada oitava.

O procedimento para a BQT pode ser explicado da seguinte maneira [48]: primeiro calcula-se uma FFT com resolução $1/T$, onde T é o período. Desse resultado guarda-se apenas a metade superior da faixa de freqüência positiva, que corresponde à oitava mais alta. Decima-se por dois o sinal original, o que significa, na freqüência, remover a oitava superior e dobrar a resolução dos canais restantes para $2/T$. Aplica-se uma FFT a esse sinal decimado e se mantém novamente apenas a metade superior da faixa de freqüências. Realiza-se essa sequência de decimação por dois e cálculo de FFT até se alcançar a menor freqüência desejada.

6.3.2 BQFFB

A idéia da BQFFB é realizar a união entre a BQT, explicada no item 6.3.1, e o uso dos filtros mais seletivos da FFB, proposta por [38] e também descrita em [49]. Seguindo, então, a proposta iniciada por Santos [38], que idealizou a BQFFB como transformada, foi realizada nesta tese a sua implementação na forma de banco de filtros para permitir uma análise espectral ao longo do tempo. Na Figura 6.3 observam-se o diagrama de blocos da BQFFB e a resolução por oitavas.

6.3.2.1 Primeiro Algoritmo: BQFFB

Realiza-se a separação do espectro em oitavas através de uma seqüência de filtragens *anti-aliasing* com freqüência de corte de $\alpha\pi/2$ (com $0 \ll \alpha < 1$), alternadas com decimações por 2. Na subdivisão das oitavas, um banco de filtros FFB de $4N$ canais resulta numa resolução de N canais por oitava. Pode-se descrever a implementação da seguinte maneira:

1. Aplica-se uma FFB de $4N$ canais ao sinal $x[n]$ original.

2. Dos $4N$ filtros que representam o espectro de 0 a 2π, guardam-se os N canais

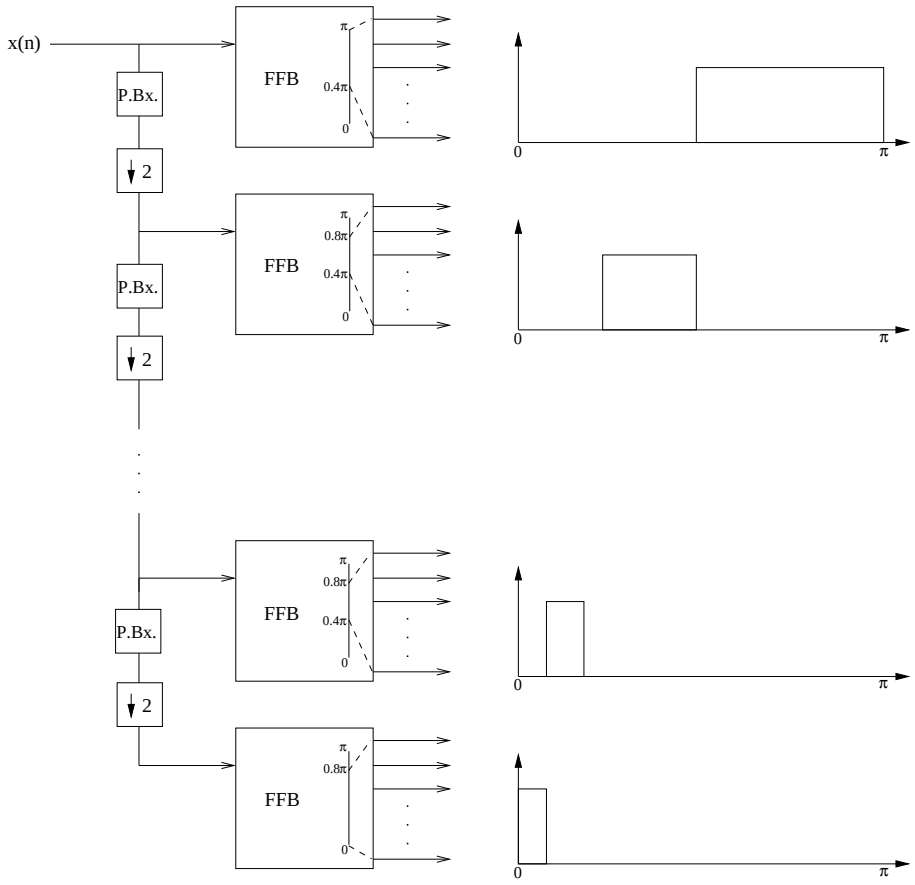

Figura 6.3: Diagrama de blocos da BQFFB.

referentes às altas freqüências da parte positiva ($\alpha\pi/2$ a π);

3. Decima-se o sinal por 2, eliminando-se assim, as altas freqüências, já analisadas no passo anterior;

4. Aplica-se novamente a FFB ao sinal decimado e guardam-se as respostas dos filtros referentes às altas freqüências ($\alpha\pi/2$ a $\alpha\pi$);

5. Calculam-se os passos 2 e 3 quantas vezes for desejado para as oitavas intermediárias; para a última oitava, guardam-se as respostas de grande parte da banda positiva, i.e., os canais referentes ao espectro de 0 a $\alpha\pi$.

Filtro decimador por oitava

Como o filtro decimador (indicado por "P.Bx." na Figura 6.3) não é ideal ($\alpha = 1$) e, portanto, não possui freqüência de corte em exatamente $\pi/2$, o final da banda passante foi ajustado para $0{,}4\pi$ fazendo-se α =0,8; assim, as amostras recolhidas de cada análise FFB decimada não são de $\pi/2$ a π, e sim de $0{,}4\pi$ a $0{,}8\pi$.

Assim evita-se a distorção de amplitude do filtro decimador. O filtro utilizado para a decimação das oitavas é IIR elíptico e sua especificação está na tabela 6.4 e na Figura 6.4 é possível observar a sua resposta em freqüência.

Tabela 6.4: Especificações do filtro decimador

Tipo	IIR Elíptico
Ordem	10
Faixa de passagem	0 a 0,4π
Faixa de rejeição	0,5π a π
Ripple na faixa de passagem	0,05 dB
Atenuação na faixa de rejeição	100 dB

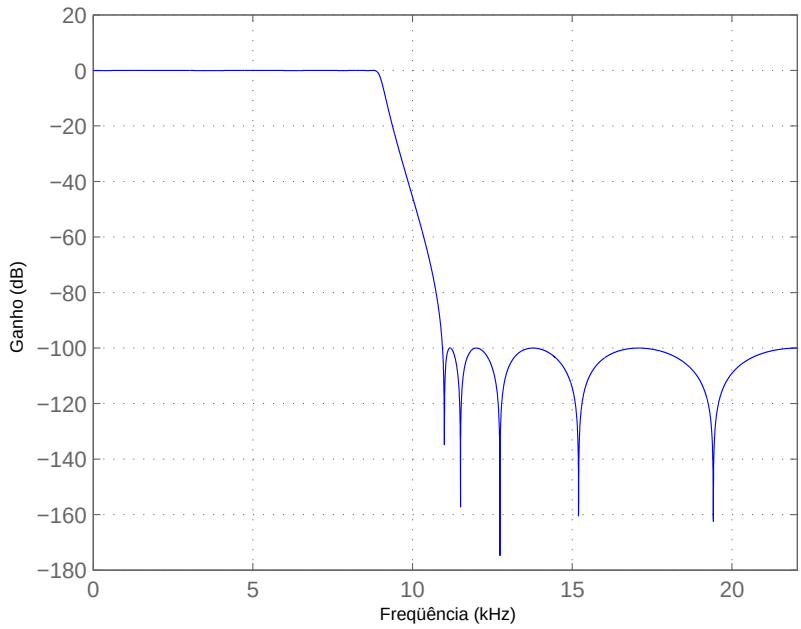

Figura 6.4: Filtro passa-baixas elíptico utilizado na decimação das oitavas no BQFFB. Neste caso é a primeira decimação, com freqüência de amostragem F_s = 44100Hz

Os filtros da BQFFB são os mesmos da FFB. Porém, como o sinal é decimado progressivamente por 2 em direção às oitavas inferiores, a banda do sinal, resultante da decimação, dobra de largura, o que pode ser analisado como uma filtragem duas vezes mais seletiva.

A Tabela 6.5 mostra as bordas freqüenciais e a largura de banda de cada canal para uma BQFFB com 289 canais gerados a partir de 10 oitavas de 32 canais FFB efetivos (de um total de 128 canais). Observa-se claramente o efeito do filtro decimador ao se comparar com proposta original de separação por oitavas de acordo com a freqüência de amostragem apresentada na Tabela 6.3. Outra constatação é que entre as oitavas há uma região (cerca de um semitom) que é descrita conjuntamente pelos filtros adjacentes, porém não mais no platô da faixa de passagem.

Tabela 6.5: Resolução da BQFFB com 10 oitavas e 289 canais, obtidas por 10 FFBs com 32 canais cada.

Índice da oitava (d)	Freqüência inicial (Hz)	Freqüência final (Hz)	Largura de banda para quarto de tom (Hz)
10	9165	22050	344
9	4496	8648	172
8	2248	4324	86
7	1124	2162	43
6	562	1081	21
5	281,0	540,5	10,7
4	140,5	270,0	5,4
3	70,252	135,120	2,691
2	35,126	67,560	1,346
1	0	33,780	0,672

6.3.2.2 Segundo Algoritmo: mBQFFB

Uma outra forma de implementar a BQFFB é proposta em [42]. Com ela elimina-se o problema do filtro *anti-aliasing* de decimação. A idéia é substituir o filtro decimador por uma mCQFFB para separar as oitavas. O canal Q escolhido possui a extensão de uma oitava, e devido à progressão geométrica, a freqüência central do filtro $f_{k1} = \sqrt{2f_0^2}$ onde f_0 é o início da banda de passagem e $2f_0$, o final. O fator de qualidade Q do filtro é, então,

$$Q = \frac{f_{k1}}{(\Delta f)_{CQ}} = \frac{\sqrt{2f_0^2}}{2f_0 - f_0} = \sqrt{2}. \tag{6.30}$$

Pode-se reduzir, porém, o custo computacional utilizando filtros da FFB [42], que dividem o espectro linearmente em duas metades, a oitava superior e o restante. Pode-se observar que numa escala linear, onde o centro do filtro é a média aritmética $f_{k2} = (f_0 + 2f_0)/2$, o mesmo filtro possuiria um fator Q diferente:

$$Q = \frac{f_{k2}}{(\Delta f)_{FFB}} = \frac{3f_0/2}{2f_0 - f_0} = 3/2. \tag{6.31}$$

O algoritmo para escolha desses filtros pode ser descrito da seguinte forma (veja a Figura 6.5):

1. O filtro da oitava superior D é o segundo filtro de uma FFB de 2 canais.

2. Os filtros das oitavas restantes, $d = (D-1), \ldots, 1$, são a cascata do segundo filtro de uma FFB de $2^{(D-d+1)}$ canais com o primeiro filtro de uma FFB de $2^{(D-d)}$ canais.

Os filtros FFB utilizados na separação das oitavas devem ser diferentes dos descritos no item 6.1.2. A Tabela 6.6 indica as variáveis escolhidas para o projeto do filtro. A principal variável alterada foi a faixa de transição, que na especificação da FFB estava muito grande (em 0,13dB —aqui, 0,03dB). Para compensar o possível aumento na ordem dos filtros, relaxou-se o *ripple* da faixa de passagem e também a atenuação da faixa de rejeição. A ordem dos filtros está indicada na Tabela 6.7. Vale ressaltar que em cada oitava é contabilizado apenas mais um filtro, pois os demais podem ser obtidos pela diferença, amostra a amostra, entre o sinal original e as versões já filtradas das oitavas superiores.

Tabela 6.6: Especificações do filtro protótipo para a CQFFB utilizada na mBQFFB com dados normalizados para $F_s = 1$.

Faixa de passagem	0 a 0,235
Faixa de rejeição	0,265 a 0,5
Ripple na faixa de passagem	0,0475 dB
Atenuação na faixa de rejeição	42 dB

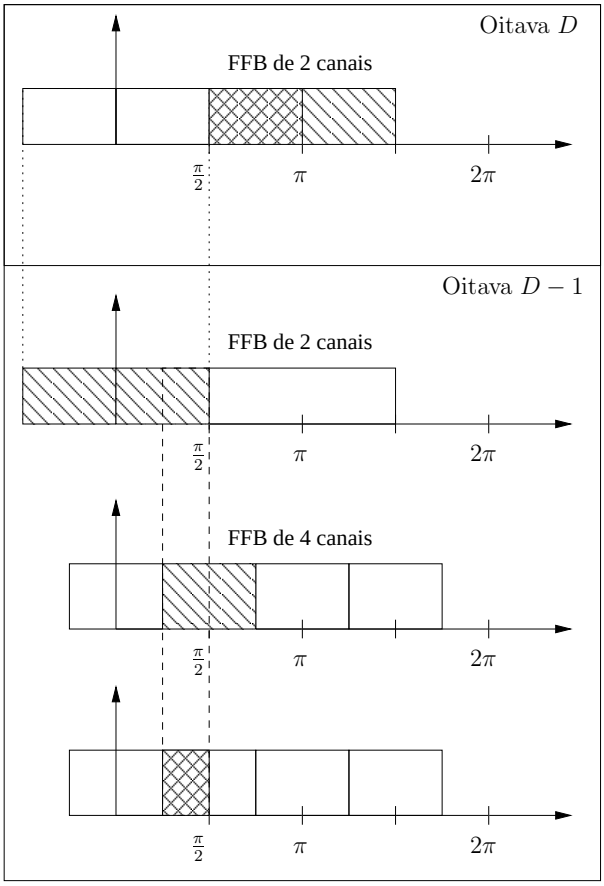

Figura 6.5: Procedimento para criar os filtros CQFFB a partir da FFB para separação das oitavas na mBQFFB.

A oitava superior, de índice D, pode ser obtida por um filtro FFB de 2 canais, que limita o sinal entre $\pi/2$ e $3\pi/2$. Porém, como o sinal de entrada será real, o limite superior fica em π. Para a oitava adjacente de índice $D-1$, como pode ser visto na Figura 6.5, seu filtro é um passa-banda de $\pi/4$ e $\pi/2$, cujo limite superior é obtido com o primeiro filtro de uma FFB de 2 canais, de banda $-\pi/2$ a $\pi/2$. A resposta desse filtro é então limitada inferiormente em $\pi/4$ com o segundo filtro de uma FFB de 4 canais, que possui uma banda passante de $\pi/4$ a $3\pi/4$; porém, a faixa acima de $\pi/2$ já foi cortada na primeira etapa da cascata. Esse procedimento é análogo para as oitavas inferiores.

A Figura 6.6 mostra o espectro de potência da CQFFB para 10 oitavas, cujas bordas freqüenciais são apresentadas na Tabela 6.3. Observa-se que a faixa de transição entre as oitavas possui a largura de apenas um quarto de tom. A Figura 6.7 ilustra a soma dos módulos dos filtros, indicando que não há distorção durante a

Tabela 6.7: Número de coeficientes acumulados para a separação das oitavas por CQFFB utilizada na mBQFFB, onde $d = D$ é a oitava superior.

Número de oitavas (D)	Índice da oitava (d)	Coeficientes na oitava d	Coeficientes acumulados $F(D)$
1	D	22	22
2	$D-1$	3	25
3	$D-2$	2	27
4	$D-3$	2	29
5	$D-4$	2	31
6	$D-5$	2	33
7	$D-6$	2	35
8	$D-7$	2	37
9	$D-8$	2	39
10	$D-9$	2	41

etapa de separação em oitavas.

Após separar o sinal em oitavas de Q constante, cada saída é novamente filtrada por N canais linearmente espaçados de uma FFB. Isto é realizado com o seguinte procedimento:

1. Para todas as oitavas ($d = 1, \ldots, D$), decime o sinal pelo fator $2^{(D-d+1)}$.

2. Filtre cada sinal resultante da etapa 1 por uma FFB de $2N$ canais, obtendo os canais separados para cada oitava d.

Essa decimação das oitavas por fatores diferentes alarga o espectro de todas para a mesma faixa, que é de 0 a 2π. Vale ressaltar que este procedimento não necessita de um filtro decimador, pois os filtros CQFFB já fizeram essa função de evitar o *aliasing*. Como pode ser visto na Figura 6.5, o último filtro de cada CQFFB possui o dobro da largura necessária (por ex., na oitava $D-1$ o filtro é de $\pi/4$ a $3\pi/4$); então deve ser utilizada uma FFB também com o dobro de canais ($2N$) desejados para cada oitava. Porém, isso não é um problema, pois, como já foi dito, a metade superior não possui sinal, já filtrado na primeira etapa da CQFFB. Dos $2N$ canais, que abrangem o espectro de 0 a 2π, apenas os N primeiros são utilizados,

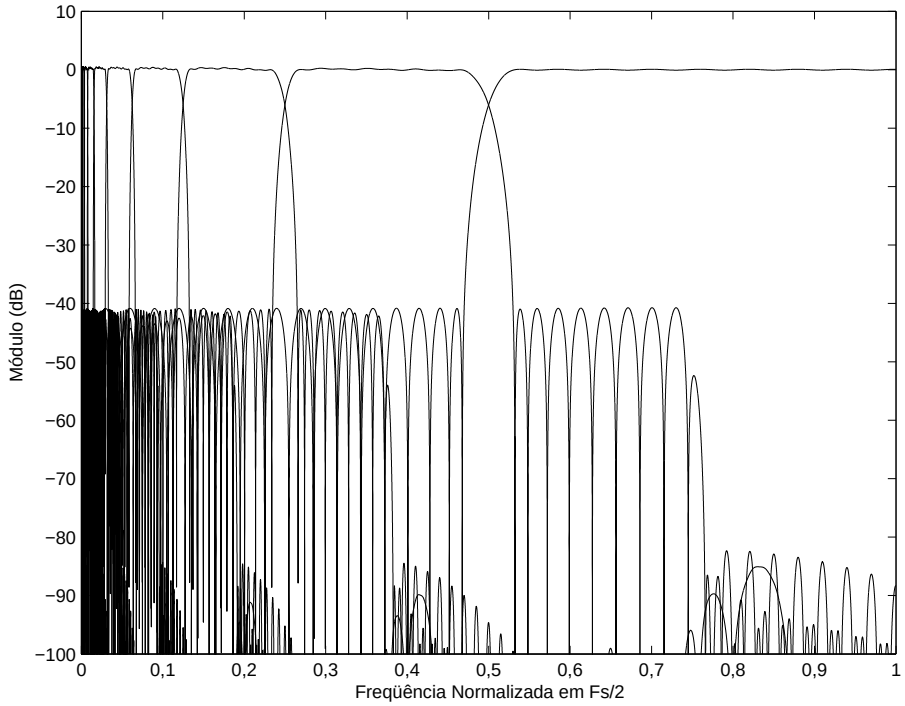

Figura 6.6: Resposta dos filtros CQFFB para separação de 10 oitavas na primeira etapa da mBQFFB.

correspondendo à faixa de 0 a π.

6.4 Complexidade Computacional

A complexidade computacional dos métodos mencionados será detalhada a seguir.

6.4.1 *s*FFT

O produtório da equação (6.2) indica que cada canal k possui um custo computacional de $\log_2 N$. Assim, uma sFFT de N canais ($k = [0, 1, ...N-1]$) possuiria um custo computacional de $N\log_2 N$ multiplicações complexas por amostra. Porém, devido à simetria, reduz-se à metade de canais, obtendo um novo custo de $(N/2)\log_2 N$. Em [40] é alertado o fato de que podem ser reaproveitadas grande parte dos cálculos referentes à amostra anterior. Com isso, é possível chegar a 1 multiplicação complexa por canal por amostra.

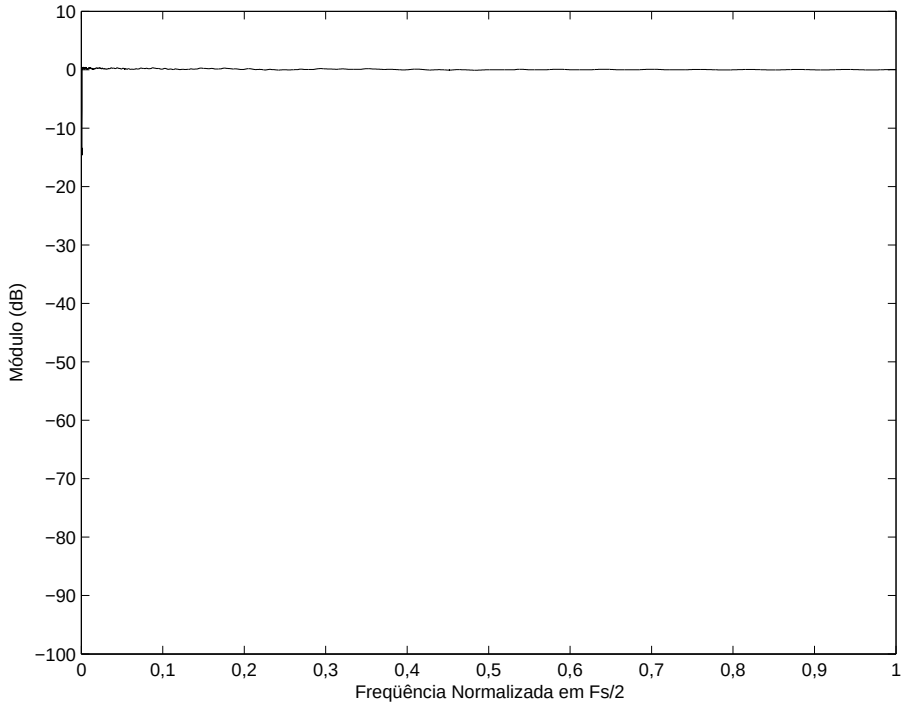

Figura 6.7: Soma dos módulos da resposta dos filtros CQFFB para separação de 10 oitavas na primeira etapa da mBQFFB.

6.4.2 CQT

A CQT possui três parâmetros iniciais que selecionam as freqüências-limite, $f_{\text{mín}}$ e $f_{\text{máx}}$, e o salto entre canais adjacentes, dado por $\alpha = f_{K+1}/f_K$.

A $f_{\text{mín}}$ é a menor freqüencia do canal $k = 0$. Assim, a menor freqüência do canal $k = K$, é $\alpha^k f_{\text{mín}}$, o que faz com que $f_{\text{máx}}$, que deve pertencer ao canal $k = K - 1$, seja

$$\begin{aligned} f_{\text{máx}} &= \alpha^K f_{\text{mín}}, \\ K &= \log_\alpha(\frac{f_{\text{máx}}}{f_{\text{mín}}}). \end{aligned} \tag{6.32}$$

O tamanho da janela nesse método é dependente do canal, i.e. $N_k = F_s/\Delta f_k$, onde $\Delta f_k = f_{\text{mín}}(\alpha^K - \alpha^{K-1})$ o que resulta em

$$N_k = \frac{F_s}{\alpha^K f_{\text{mín}}} \frac{1}{(\alpha - 1)}. \tag{6.33}$$

Para um sinal de M amostras com uma janela N_k, a quantidade total de janelas é M/N_k para um determinado canal k. Cada canal possui um custo computacional de um canal de DFT de N_k amostras, o que resulta em N_k produtos

complexos, como pode ser observado pela equação (6.21) da DFT com uma janela variável N_k.

Para se obter o custo computacional por amostra para uma CQT de K canais (com k=[0,1,...,K-1]), basta dividir o total de produtos pela quantidade total de amostras:

$$\frac{1}{M}\sum_{k=0}^{K-1}\frac{M}{N_k}N_k = K. \tag{6.34}$$

Já para a versão *sliding* CQT, tem-se um custo computacional por amostra de

$$\sum_{k=0}^{K-1} N_k = \frac{F_s}{f_{\text{mín}}}\frac{\alpha^{-1}(1-\alpha^{-K})}{(1-\alpha^{-1})^2}. \tag{6.35}$$

6.4.3 BQT

Para a BQT, a complexidade é calculada como uma sFFT para cada oitava ($d = [0, 1, ...D]$) além das decimações necessárias ($G(d)$). Dado um sinal de M amostras, devido às decimações por dois o sinal apresentado para cada oitava possui um tamanho diferente, i.e, $M(d)$, onde $M(D) = M$, $M(D-1) = M/2$ etc. Assim, a complexidade computacional com uma janela de N amostras (que no caso da BQT não é função de k), resulta em um total de $M(d)/N$ janelas para um canal k de uma oitava d, que é uma DFT de N amostras, resultando em N produtos.

Considerando-se um filtro decimador FIR de ordem L, o custo computacional por canal por amostra é

$$\sum_{d=0}^{D-1}\left(\frac{1}{M(d)}\sum_{k=0}^{K-1}\frac{M(d)}{N}N + G(d)\right), \tag{6.36}$$

onde $G(d) = M(d)(L+1)$.

6.4.4 FFB

Como a FFB leva em conta a estrutura em borboleta da FFT tendo apenas como diferença a ordem dos filtros protótipos de cada nível da cascata l, pode-se comparar facilmente seu custo computacional com o da FFT, apresentado no item 6.4.1, que é de 1 multiplicação complexa por canal por amostra.

A Tabela 6.8 apresenta o número de coeficientes não-nulos por nível da FFB projetados em [41] aproveitando-se o fato de os filtros serem de fase linear [32]. Cada filtro possui um complementar que não é contabilizado, pois é possível obter a sua saída pela diferença (subtração ponto a ponto) entre o sinal de entrada e a resposta do filtro original. Os níveis mais altos possuem filtros menores, e a partir do quinto nível, tem-se apenas dois coeficientes, o que torna vantajoso usar a FFB para um maior número de canais.

Tabela 6.8: Quantidade de coeficientes não-nulos e distintos por nível na estrutura de sub-filtros FFB.

Nível da cascata (l)	Coeficientes por filtro	Filtros	Coeficientes por nível	Total de coeficientes $C_{\mathrm{FFB}}(l)$
1	7	1	7	7
2	6	2	12	19
3	3	4	12	31
4	3	8	24	55
5	2	16	32	87
6	2	32	64	151
7	2	64	128	279
8	2	128	256	535
⋮	⋮	⋮	⋮	⋮
$\log_2 N$	2	$N/2$	N	$2N + 23$

A Tabela 6.9 apresenta o cálculo da complexidade computacional do FFB para um número N genérico de canais, e ainda para alguns valores de N. Para uma FFB com uma quantidade de canais N superior a 16 (i.e. $l \geq 5$), esse número de multiplicações complexas por amostra por canal é de $(2N + 23)/N$, i.e., aproximadamente 2.

6.4.5 C*Q*FFB

Para o cálculo da complexidade da primeira implementação da C*Q*FFB, descrita no item 6.2.2.1, devem-se computar as multiplicações presentes na filtragem pelo canal e as referentes às reamostragens do sinal. Assim, a complexidade para

Tabela 6.9: Complexidade computacional do FFB: número de multiplicações complexas por amostra por canal.

Total de canais FFB (N)	Total de coeficientes $C_{\mathrm{FFB}}(l)$	Multiplicações complexas por canal por amostra
2	7	7/2=3,5
4	19	19/4=4,75
8	31	31/8=3,875
16	$2 \cdot 16 + 23$	55/16=3,4375
32	$2 \cdot 32 + 23$	87/32=2,7188
64	$2 \cdot 64 + 23$	151/64=2,3594
128	$2 \cdot 128 + 23$	279/128=2,1797
256	$2 \cdot 256 + 23$	535/256=2,0899
$\vdots$	$\vdots$	$\vdots$
N	$2N + 23$	$(2N + 23)/N \approx 2,0$

um canal k é dada por [42]

$$C_{\mathrm{CQFFB}}(k) = C_R(k) + (C_Q + 1)\,\gamma(k), \tag{6.37}$$

onde $C_R(k)$ é o custo da reamostragem do sinal de entrada, $\gamma(k)$, o fator de reamostragem, ambos para o canal k, e C_Q é o custo computacional do filtro FFB escolhido na primeira etapa do algoritmo do item 6.2.2.1.

6.4.6 mCQFFB

Com a reamostragem do filtro, perde-se a grande quantidade de coeficientes nulos dos filtros FFB, aumentando o seu custo computacional. Porém os filtros podem ser obtidos apenas uma vez de forma *off-line*, não sendo mais necessário reamostrá-los durante a filtragem. Assim, a complexidade para um determinado canal k é definida como

$$C_{\mathrm{mCQFFB}}(k) = (C_Q + 1)\,\gamma(k), \tag{6.38}$$

e o custo computacional total é

$$C_{\mathrm{mCQFFB,Total}} = \sum_{k=q_1}^{q_2} \left(C_Q r^{-k} + 1\right), \tag{6.39}$$

onde k é o índice do canal, $q_1 = \left\lfloor \log_r \left(2^{-D} \frac{N}{2Q} \right) \right\rfloor$, $q_2 = \left\lceil \log_r \frac{N}{2Q} \right\rceil$, r é a razão freqüencial e D é o número de oitavas. A equação (6.38) mostra que a segunda implementação (mCQFFB) possui um custo computacional menor, pois não necessita de reamostragens durante a filtragem, como a CQFFB.

6.4.7 BQFFB

A complexidade computacional da BQFFB deve levar em conta duas partes: as multiplicações complexas da FFB, explicadas em 6.1.2, e as multiplicações simples relativas às decimações do sinal. Para uma BQFFB de D oitavas com N canais efetivos por oitava, são realizadas D FFBs de $4N$ canais, como pode ser visto no algoritmo da seção 6.3.2.1, já que apenas em torno de $1/4$ dos canais são aproveitados por oitava. Em D oitavas, são realizadas $(D-1)$ decimações, que, multiplicados pelo filtro *anti-aliasing* de ordem L[2], resultam em um total de $(L+1)(D-1)$ multiplicações simples. Unindo essas duas partes, tem-se um custo computacional por canal por amostra de

$$C_{\mathrm{BQFFB}} = (D-1)(L+1) + C_{\mathrm{FFB}}(l+2)D \approx D(L + C_{\mathrm{FFB}}(l+2)), \tag{6.40}$$

onde $C_{\mathrm{FFB}}(l)$ é dado na última coluna da tabela 6.8.

6.4.8 mBQFFB

O cálculo da complexidade computacional da mBQFFB pode ser separado em duas partes, a separação do sinal de entrada em D oitavas e a posterior separação de cada oitava ($d = 1, \ldots, D$) em N canais linearmente espaçados.

Pela Tabela 6.7, tem-se o total de coeficientes não-nulos acumulados $F(D)$ para os filtros CQFFB. Como a filtragem FIR requer, por amostra, um custo computacional do tamanho do filtro acrescido de um, a complexidade por canal por amostra da etapa de separação é de

$$C_{\mathrm{mBQFFB,\ CQFFB}} = (F(D) + D)/D2N, \tag{6.41}$$

[2]O filtro *anti-aliasing* projetado tem ordem $L = 10$, como mostrado na tabela 6.4.

onde cada oitava possui um filtro com $F(d)$ coeficientes gerando um custo de $F(d)+1$ por amostra e, como será visto a seguir, produzindo um total de $D2N$ canais.

Para se obter uma resolução de N canais por oitava, deve ser realizada uma FFB de $2N$ canais, como já foi dito anteriormente. Pela cascata da FFB, e sabendo que $l = \log_2 N$, tem-se um custo computacional por canal por amostra de

$$C_{\mathrm{mBQFFB,\ FFB}} = C_{\mathrm{FFB}}(l+1)D/D2N. \tag{6.42}$$

Entende-se pela equação acima que serão realizadas D FFBs (oitavas) de $2^{l+1} = 2N$ canais, com um total de $D2N$ canais. Unindo essas duas partes que formam a mBQFFB, tem-se que o custo computacional por amostra (sem ser por canal) é de

$$C_{\mathrm{mBQFFB,Total}} = (F(D) + D) + C_{\mathrm{FFB}}(l+1)D. \tag{6.43}$$

6.4.9 Comparações

Na Figura 6.8, comparam-se os custos computacionais da mCQFFB com a mBQFFB (fixa em dez oitavas, variando em múltiplos de dois a quantidade de canais por oitava). Em aplicações de áudio são utilizados entre 100 e 320 canais; com isso a mCQFFB possui um custo computacional até cinco ordens de grandeza maior em relação à mBQFFB.

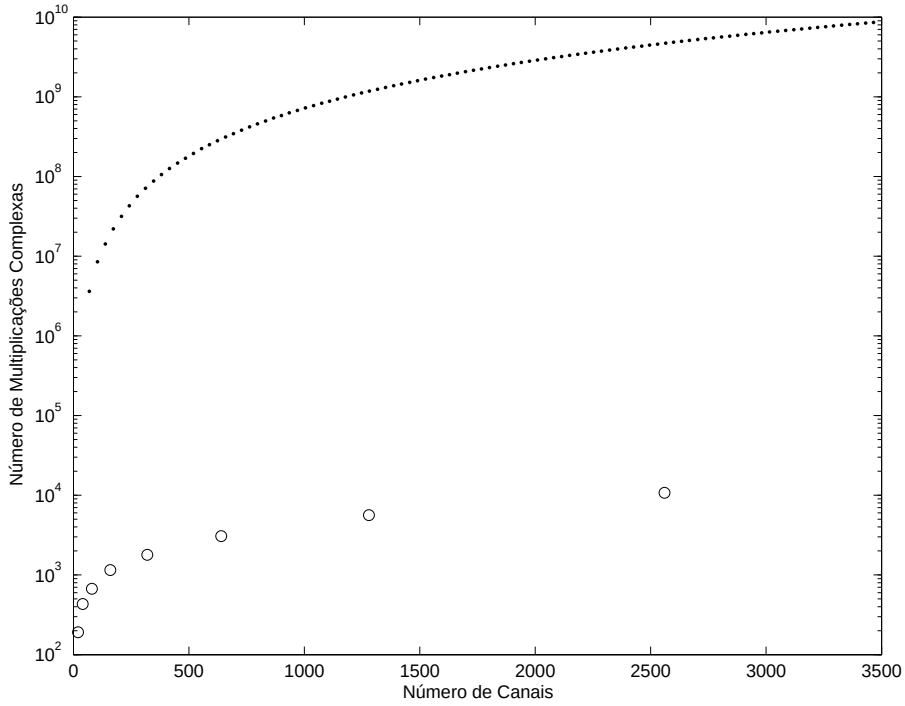

Figura 6.8: Comparação entre os custos computacionais da mCQFFB (linha pontilhada) com a mBQFFB (círculos).

Capítulo 7

Exemplos

Neste capítulo são apresentados exemplos dinâmicos para os métodos de análise discutidos no capítulo anterior. Os exemplos foram divididos em duas partes: Testes Comparativos (Seção 7.1) e Testes Complementares da mBQFFB (Seção 7.2). Para os testes comparativos foram escolhidas a CQT, mCQFFB e mBQFFB.

São apresentados três tipos diferentes de visualização para cada método. Um espectrograma em 2D, para se ter uma visão geral da análise no domínio tempo-freqüência; e duas maneiras de visualização em 3D, nos domínios de tempo-freqüencia-amplitude: uma com as linhas paralelas ao eixo temporal e outra de forma perpendicular. A visualização em paralelo ao tempo indica a saída (em amplitude) de cada canal ou *bin*. Como um canal pode esconder o seu adjacente, mostra-se também o gráfico perpendicular ao tempo, que também facilita a visualização da energia do sinal em determinadas áreas do espectro, principalmente nos métodos com menor resolução freqüencial.

Os sinais escolhidos para os Testes Comparativos da Seção 7.1 foram: senóides sintéticas para observar a resolução em tempo e freqüência dos métodos; e uma gravação para piano solo para analisar ataque e polifonia.

Para os Testes Complementares da mBQFFB (Seção 7.2) foram utilizados dois exemplos mostrando também a partitura: um de flauta solo para observar a expressividade e a dinâmica gradativa; e um trecho rápido de piano solo para analisar o desempenho da mBQFFB para sinais de variação rápida.

7.1 Testes Comparativos

Foram comparados os métodos da CQT, da mBQFFB e da mCQFFB. A seguir são apresentadas as configurações de cada um dos métodos.

As figuras da mBQFFB e mCQFFB são apresentadas retirando-se as amostras referentes a metade do tamanho do filtro, tanto no início como no fim, geradas pela convolução do sinal com o filtro. Além disso, no exemplo do sinal de áudio real, tanto na mBQFFB como na mCQFFB as figuras foram sub-amostradas no tempo para um total de 50 amostras totais, visando a reduzir o tamanho da figura, o que não compromete a visualização da evolução do sinal ao longo do tempo. Preferiu-se apresentar os gráficos da mCQFFB em escala linear.

Para a mCQFFB, devido a limitações no Matlab, realizou-se a análise com 57 canais, em uma resolução de apenas um quarto de tom abrangendo as freqüências de $246,92$Hz (B3) a $1244,51$Hz (D♯6).

A mBQFFB foi configurada para uma resolução de 10 oitavas (como indicadas na Tabela 6.3), com 32 canais efetivos por oitava, o que equivale a uma resolução de um quarto de tom. Foram geradas, para cada exemplo, uma figura em duas dimensões com as oitavas separadas e mais uma figura em três dimensões (Tempo, Freqüência e Amplitude) para cada oitava.

Também na CQT foi utilizada uma faixa de freqüências similar à da mCQFFB, de 246Hz a 1250Hz. O sinal foi sub-amostrado no tempo para ter um total de 50 amostras totais, condizendo, assim, com as outras formas de análise. A separação escolhida foi de quarto-de-tom. Para fazer a análise ao longo do tempo foi utilizada uma janela de Hamming com 4096 amostras e um deslocamento de 600 amostras.

7.1.1 Senóides Sintéticas

Foram escolhidas arbitrariamente duas senóides com freqüências f1=263Hz e f2=296Hz e suas harmônicas segunda, terceira e quarta. O tempo total é de 2 segundos com taxa de amostragem de 44100Hz. A senóide com f1 e suas harmônicas têm duração do início ao fim do sinal. Já a senóide com f2 e suas harmônicas são sintetizadas apenas na segunda metade do sinal, a fim de testar a ativação de um canal.

Pela análise das Figuras 7.1 a 7.3, nota-se claramente a baixa seletividade dos filtros da CQT, pois as senóides acabam ativando pelo menos dois canais adjacentes.

As Figuras 7.4 a 7.6 mostram a análise pela mCQFFB e a mBQFFB é indicada nas Figuras 7.7 a 7.10. Neste exemplo observa-se claramente a alta seletividade da FFB e uma pequena oscilação no início e fim de cada senóide, provavelmente devido ao seu aparecimento abrupto.

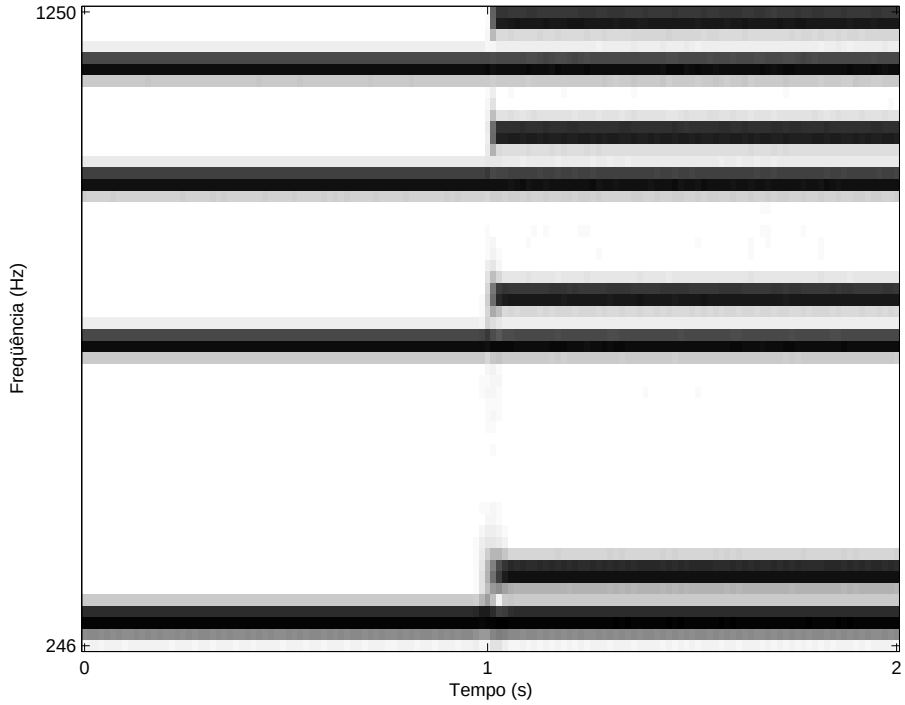

Figura 7.1: Exemplo de senóides analisadas com a CQT.

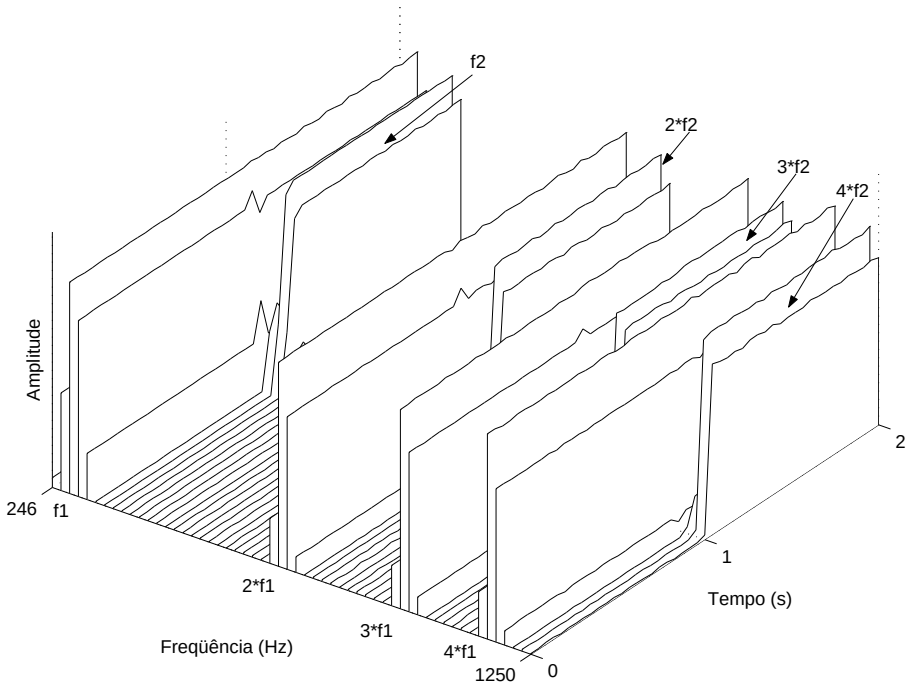

Figura 7.2: Exemplo de senóides analisadas com a CQT, visualização em 3D.

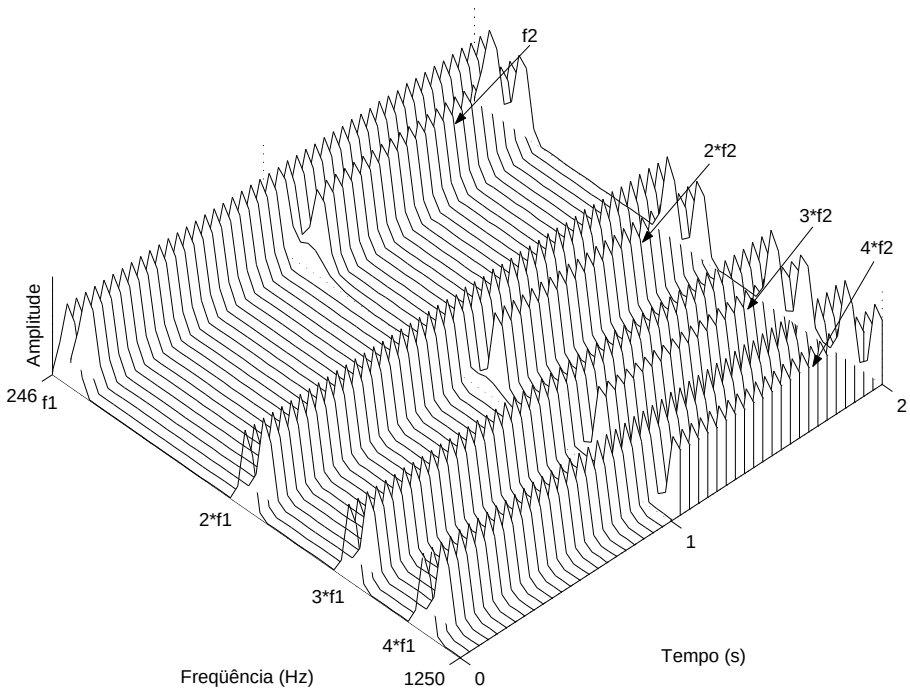

Figura 7.3: Exemplo de senóides analisadas com a CQT, visualização em 3D.

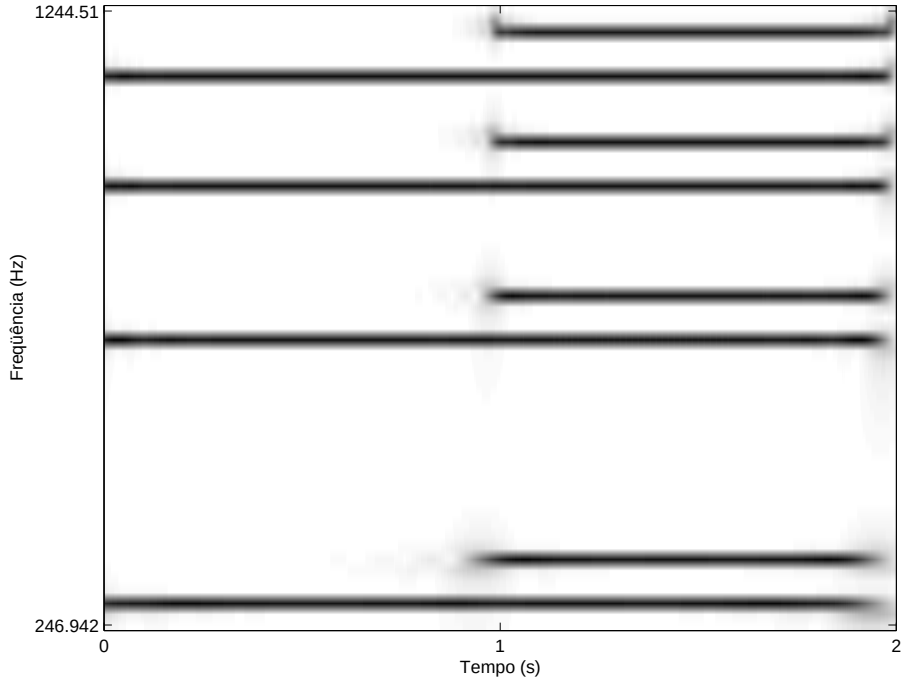

Figura 7.4: Exemplo de senóides analisadas com a mC*Q*FFB.

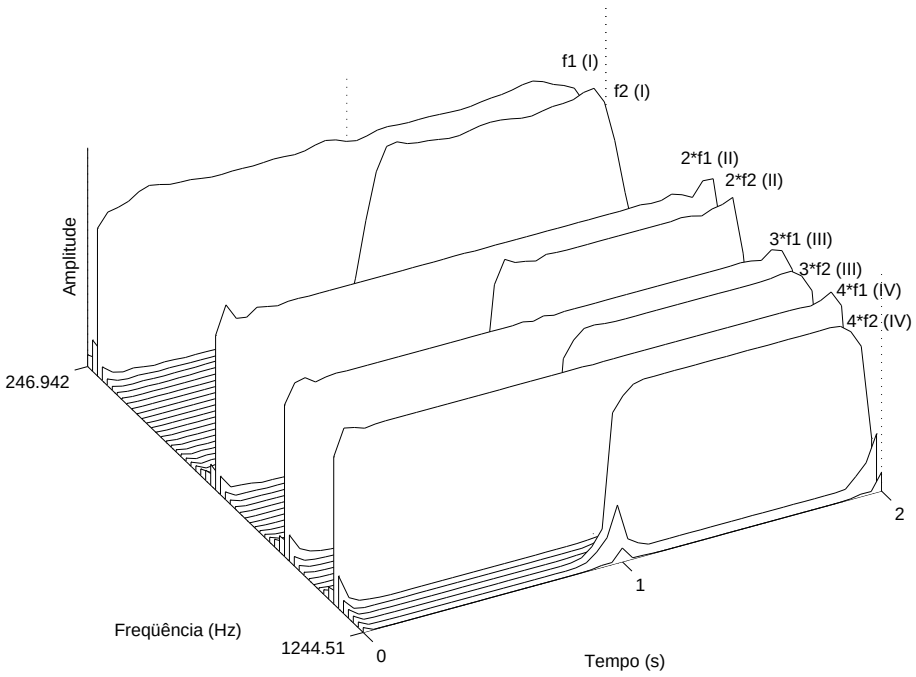

Figura 7.5: Exemplo de senóides analisadas com a mC*Q*FFB, visualização em 3D.

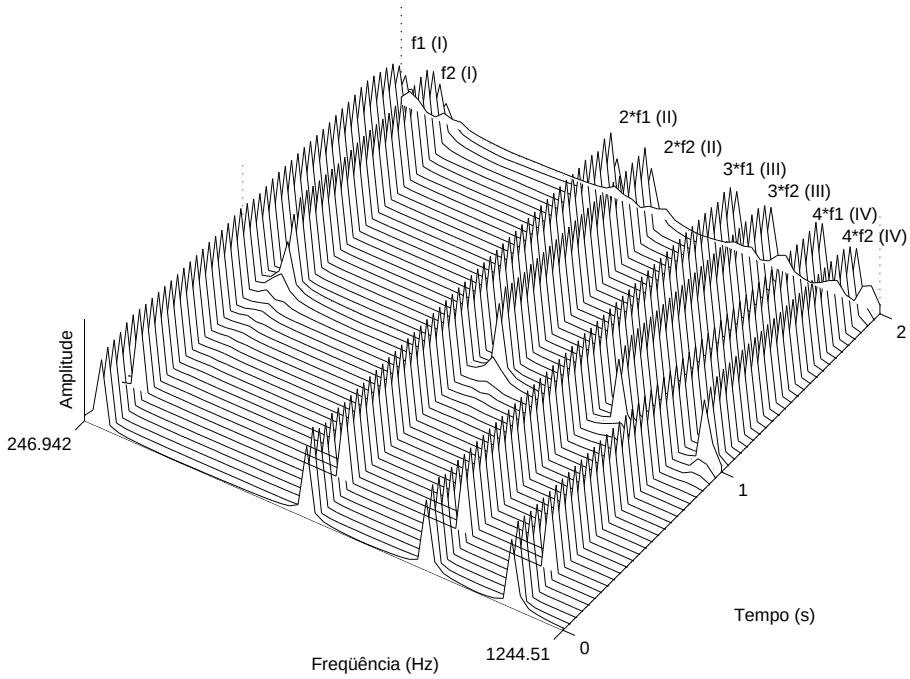

Figura 7.6: Exemplo de senóides analisadas com a mCQFFB, visualização em 3D.

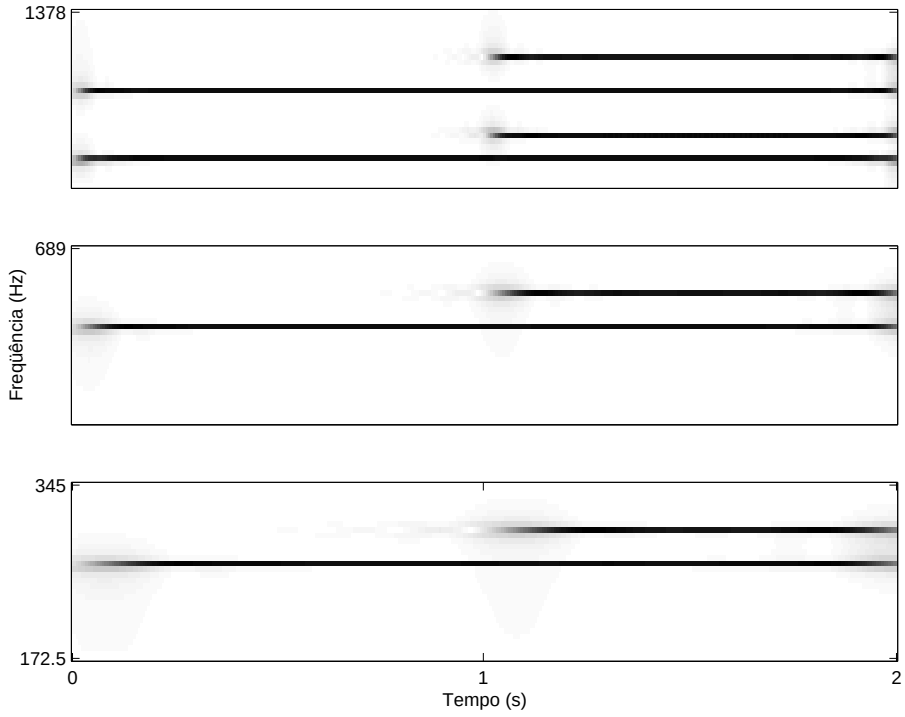

Figura 7.7: Exemplo de senóides analisadas com a mBQFFB.

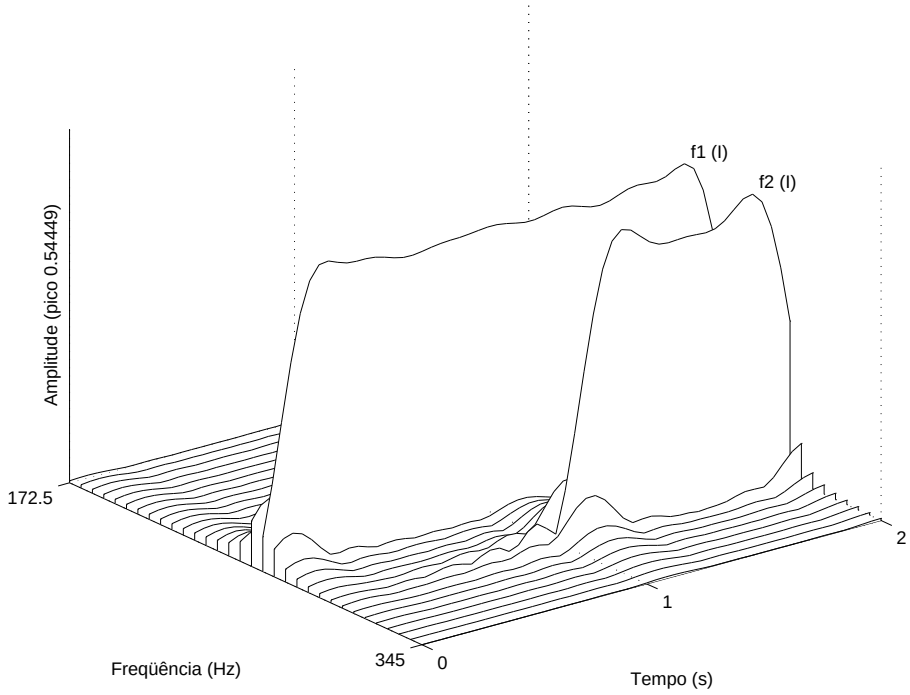

Figura 7.8: Exemplo de senóides analisadas com a mBQFFB, visualização em 3D da oitava $d = 1$ (mais grave).

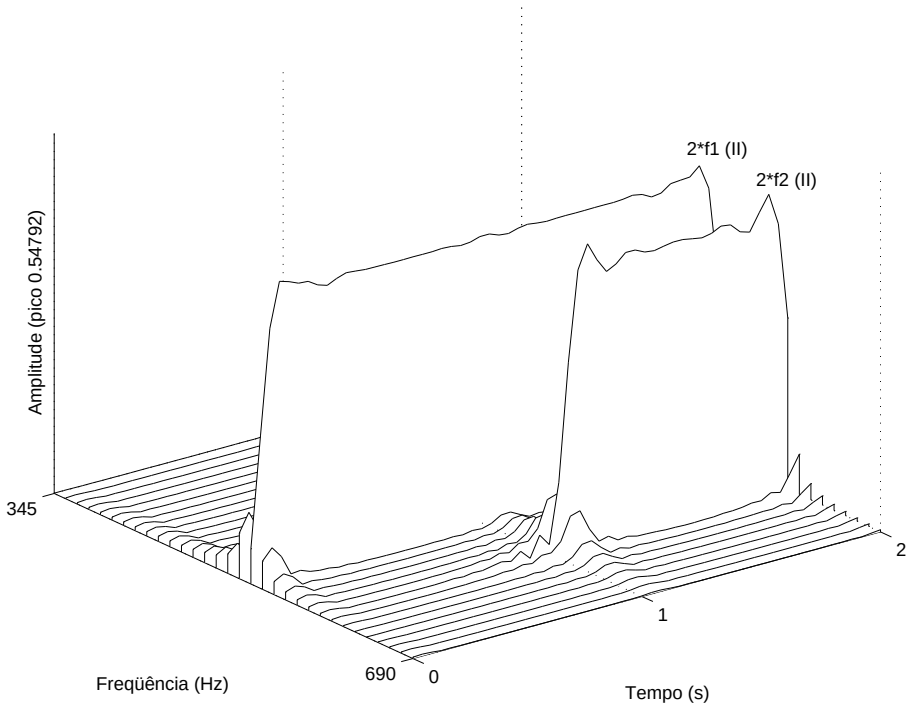

Figura 7.9: Exemplo de senóides analisadas com a mBQFFB, visualização em 3D da oitava $d = 2$ (média).

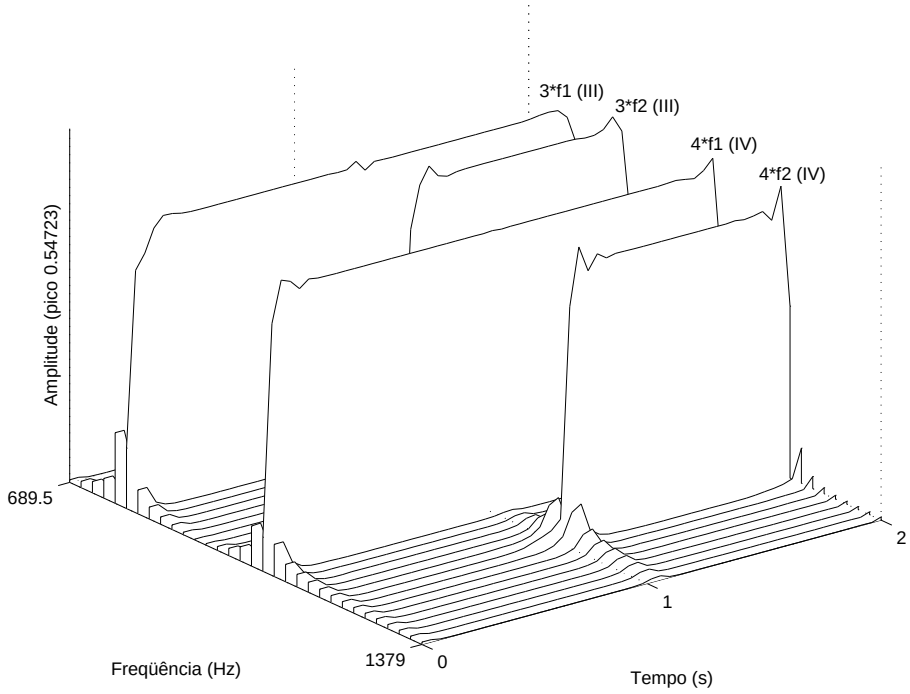

Figura 7.10: Exemplo de senóides analisadas com a mBQFFB, visualização em 3D da oitava $d = 3$ (mais aguda).

7.1.2 Sinal de Áudio Real

Foi escolhido o trecho final de uma gravação do Prelúdio das Bachianas Brasileiras Nº 4 para piano de Villa-Lobos. O sinal inicia com o final do arpejo do penúltimo compasso e logo em seguida com um acorde, e por volta de 3 segundos finaliza com um Ré oitavado ($D3$ e $D4$).

Aqui novamente fica bastante evidente pela análise da Figura 7.12, que os filtros pouco seletivos da CQT com separação de quarto de tom não permitem a identificação das notas. Para cada componente são ativados dois filtros adjacentes. Esse efeito poderia ser minorado com uma separação de semitom, porém não manteria o padrão de comparação com os métodos da FFB, além de tornar a seletividade pior.

A análise com a mCQFFB pode ser vista nas Figuras 7.14 a 7.16 e a mBQFFB é apresentada nas Figuras 7.17 a 7.27.

Comparando-se a Figura 7.11 com a 7.14, nota-se grande melhora na resolução das linha espectrais da mCQFFB em relação à CQT.

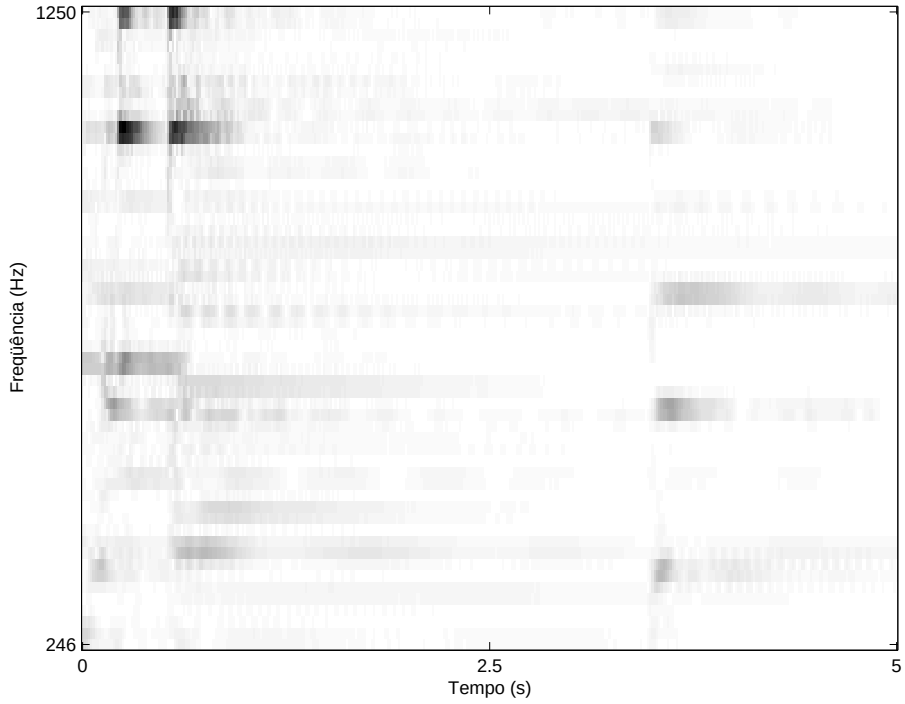

Figura 7.11: Trecho de Villa-Lobos analisado com a CQT.

Conforme explicado anteriormente, somente com a mBQFFB foi possível analisar o extremo inferior do espectro. Como pode ser visto na Figura 7.18, a

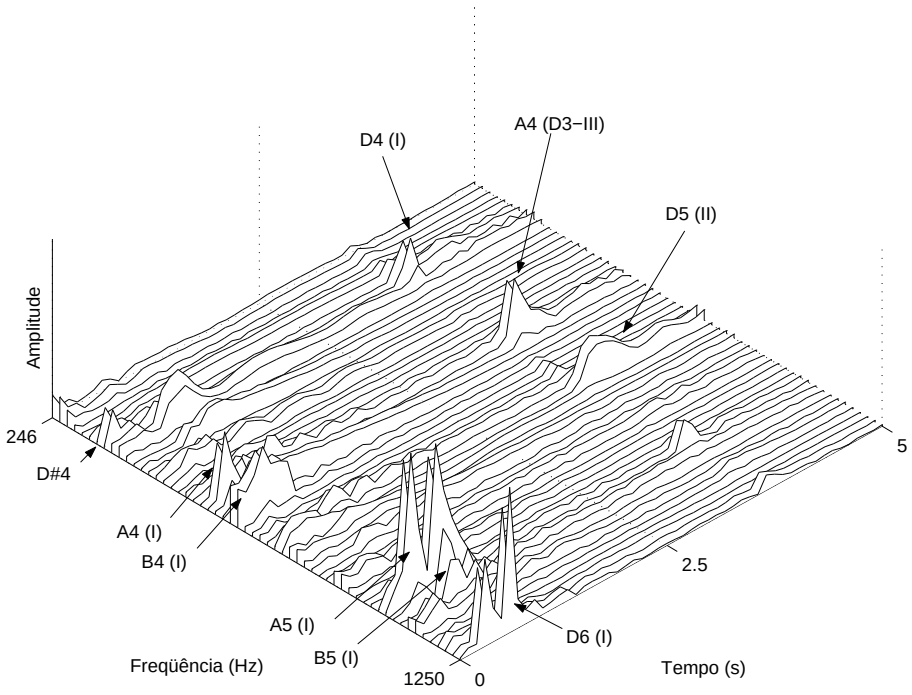

Figura 7.12: Trecho de Villa-Lobos analisado com a CQT, visualização em 3D.

análise com a mBQFFB só não foi suficientemente clara na oitava inferior, onde se encontra a nota mais grave da peça, o $B0$ de $30,87$Hz. Já nas oitavas $d = 2$ e $d = 3$, apresentadas respectivamente nas Figuras 7.19 e 7.20, observa-se claramente a presença das notas.

As duas últimas oitavas (Figuras 7.26 e 7.27) mostram o rápido decaimento das harmônicas de alta freqüência, além de sua baixas amplitudes. Na oitava $d = 10$ observa-se no início de cada canal apenas o transitório dos filtros, de baixa amplitude. O ruído no final da banda é um resíduo de *aliasing*.

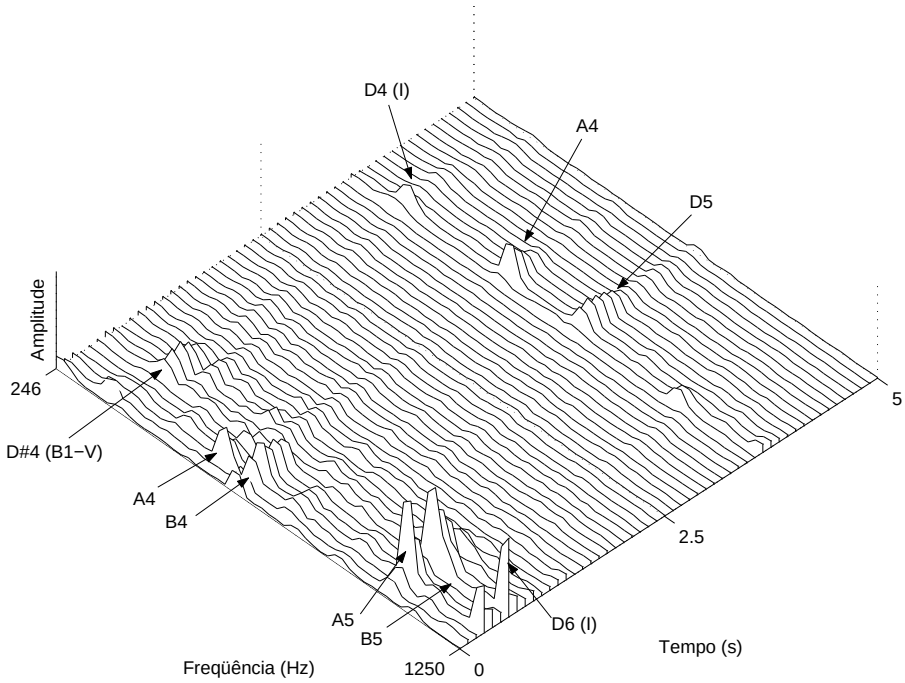

Figura 7.13: Trecho de Villa-Lobos analisado com a CQT, visualização em 3D.

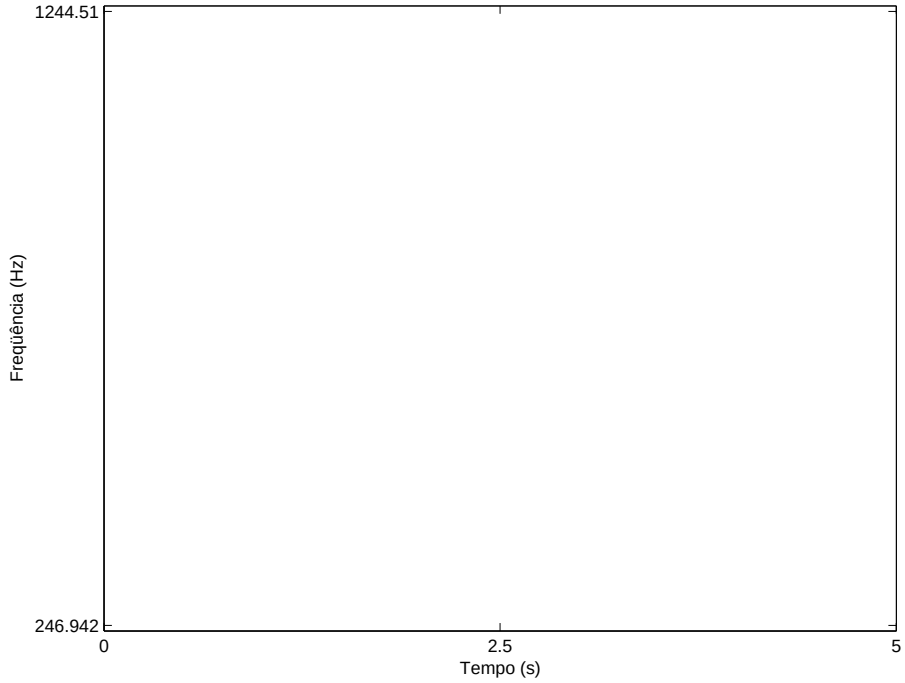

Figura 7.14: Trecho de Villa-Lobos analisado com a mCQFFB.

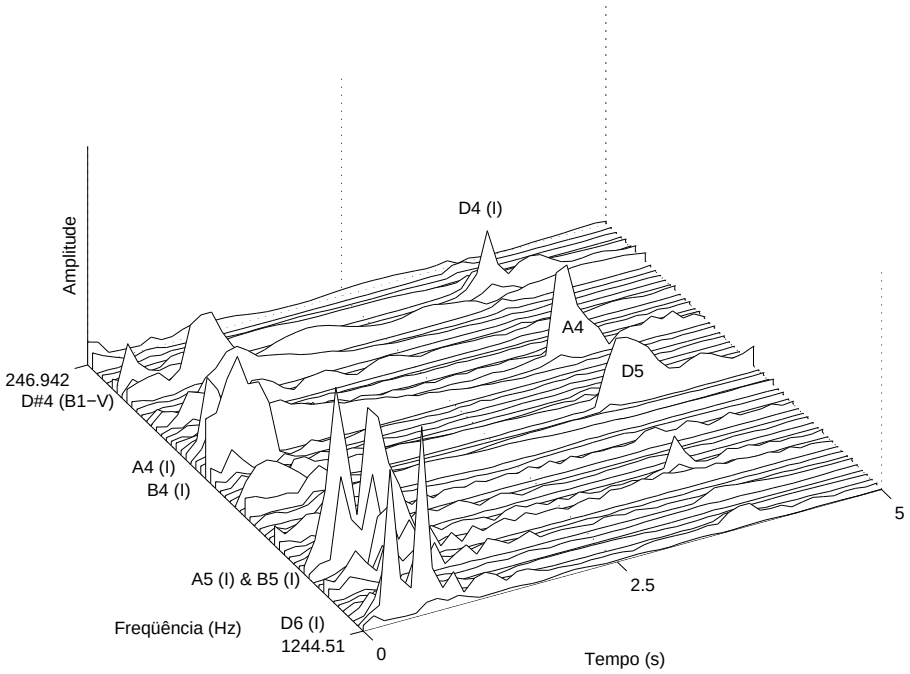

Figura 7.15: Trecho de Villa-Lobos analisado com a mCQFFB, visualização em 3D.

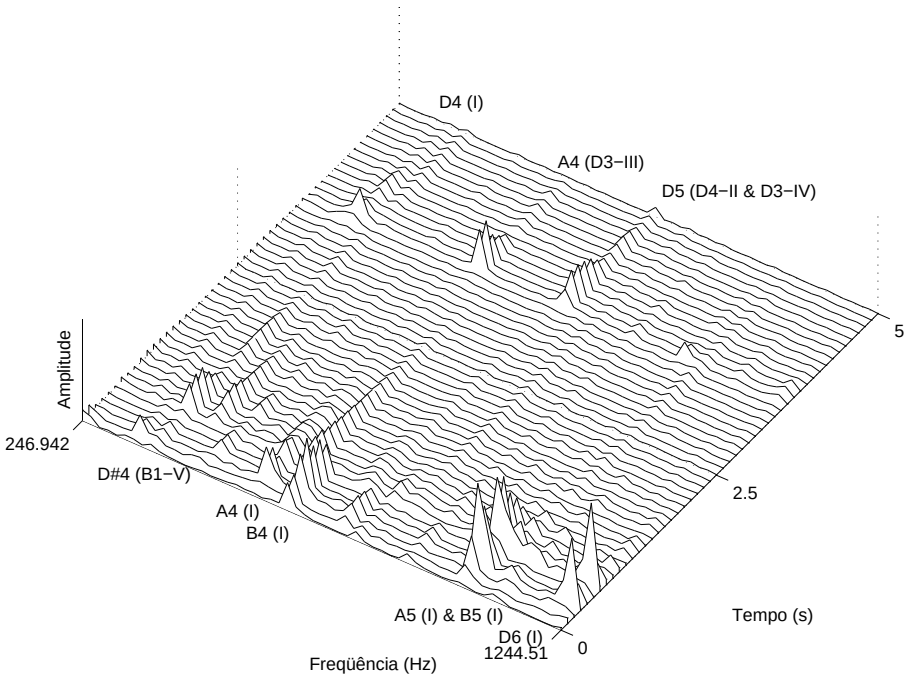

Figura 7.16: Trecho de Villa-Lobos analisado com a mCQFFB, visualização em 3D.

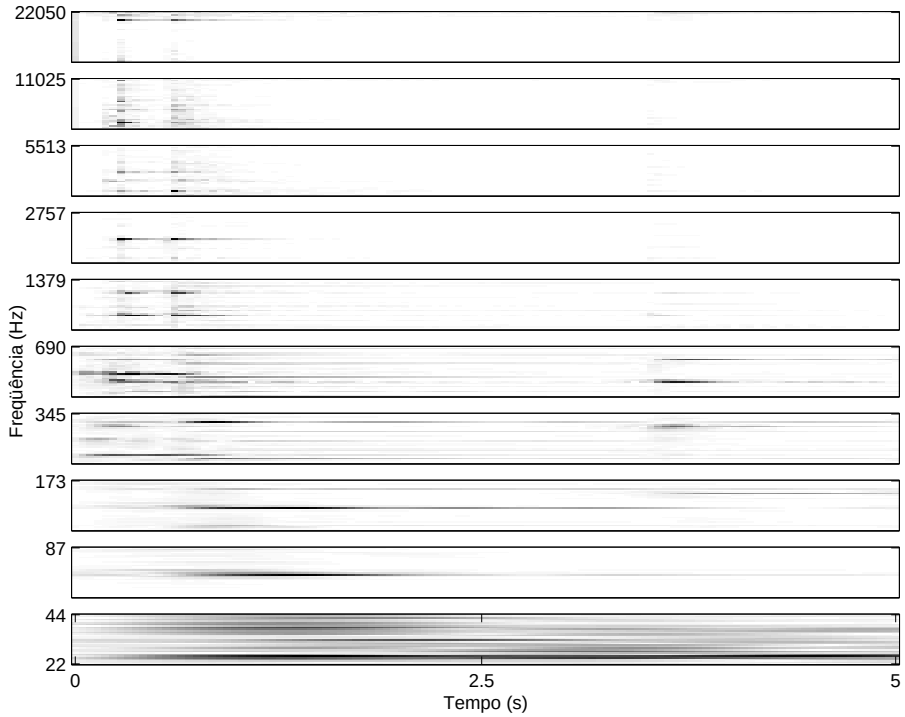

Figura 7.17: Trecho de Villa-Lobos analisado com a mBQFFB, todas as oitavas.

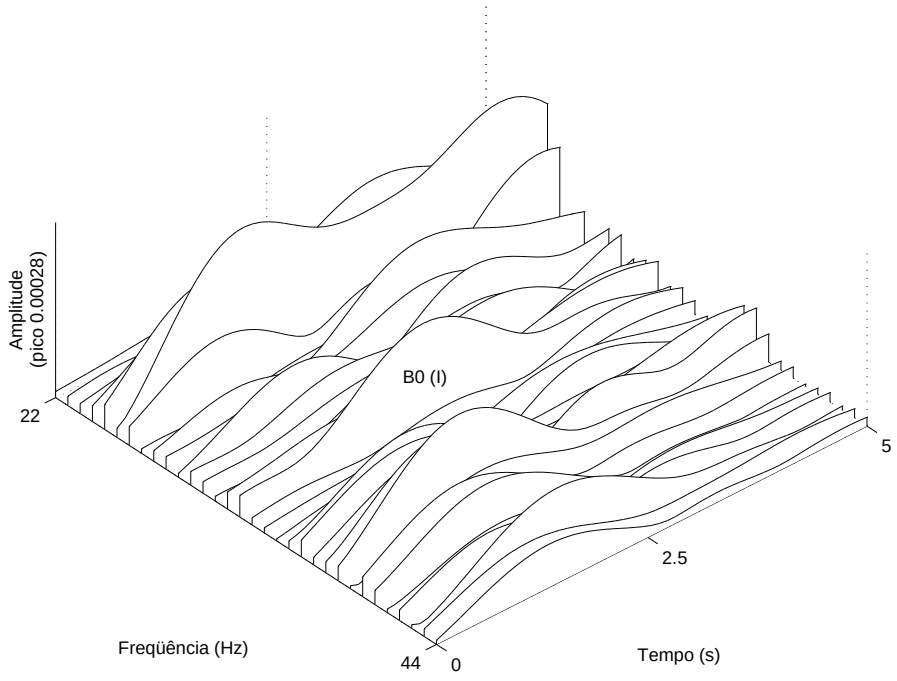

Figura 7.18: Trecho de Villa-Lobos analisado com a mBQFFB, visualização em 3D da oitava $d = 1$ (mais grave).

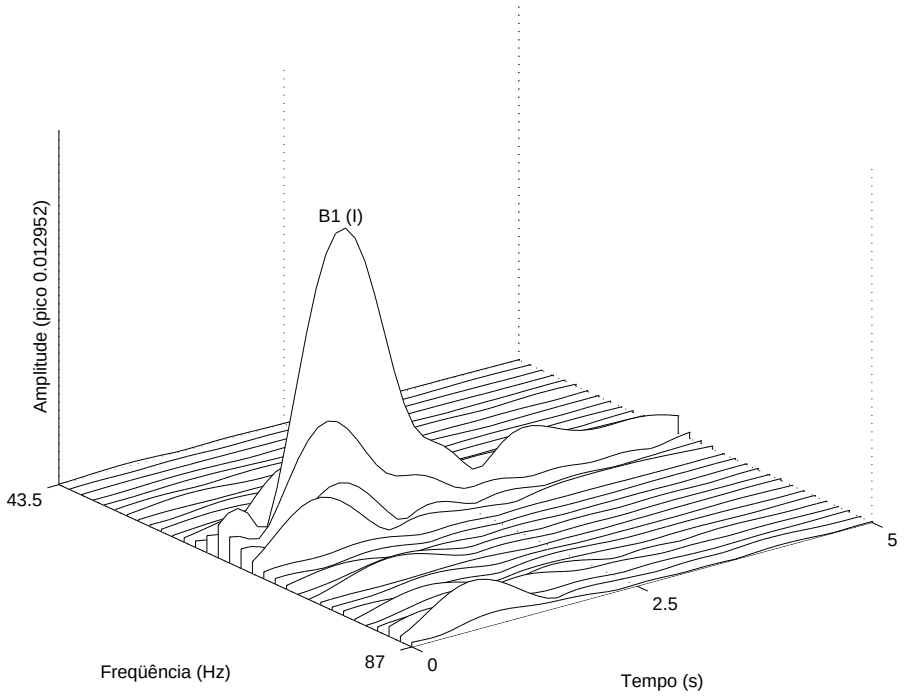

Figura 7.19: Trecho de Villa-Lobos analisado com a mBQFFB, visualização em 3D da oitava $d = 2$.

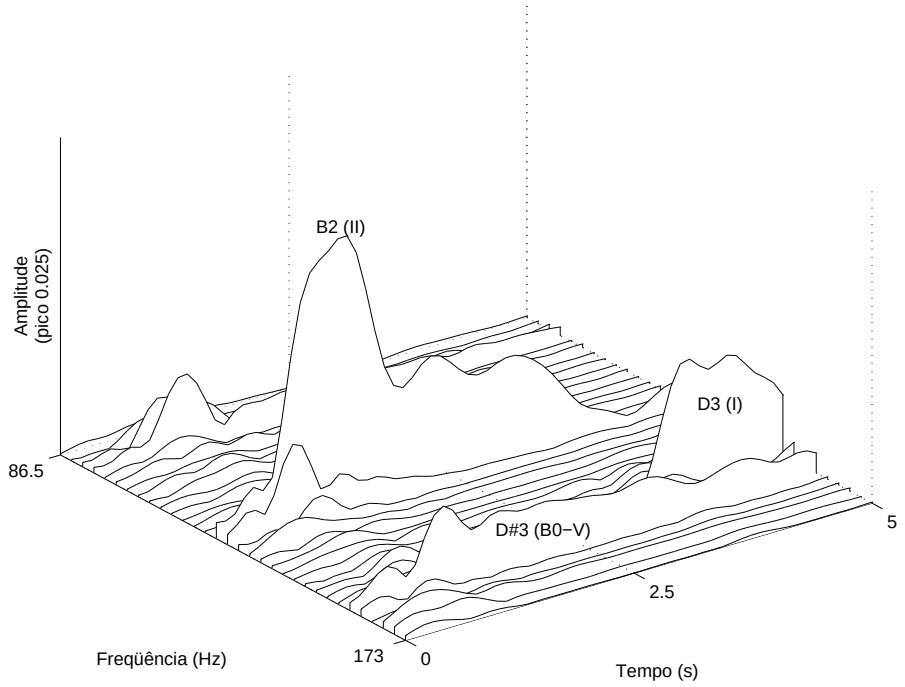

Figura 7.20: Trecho de Villa-Lobos analisado com a mBQFFB, visualização em 3D da oitava $d = 3$.

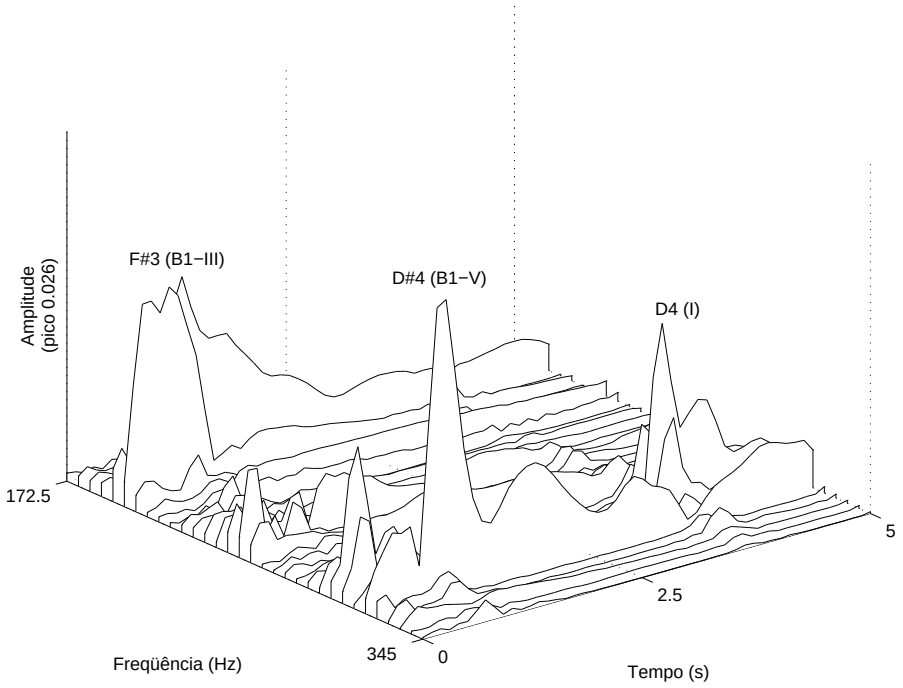

Figura 7.21: Trecho de Villa-Lobos analisado com a mBQFFB, visualização em 3D da oitava $d = 4$.

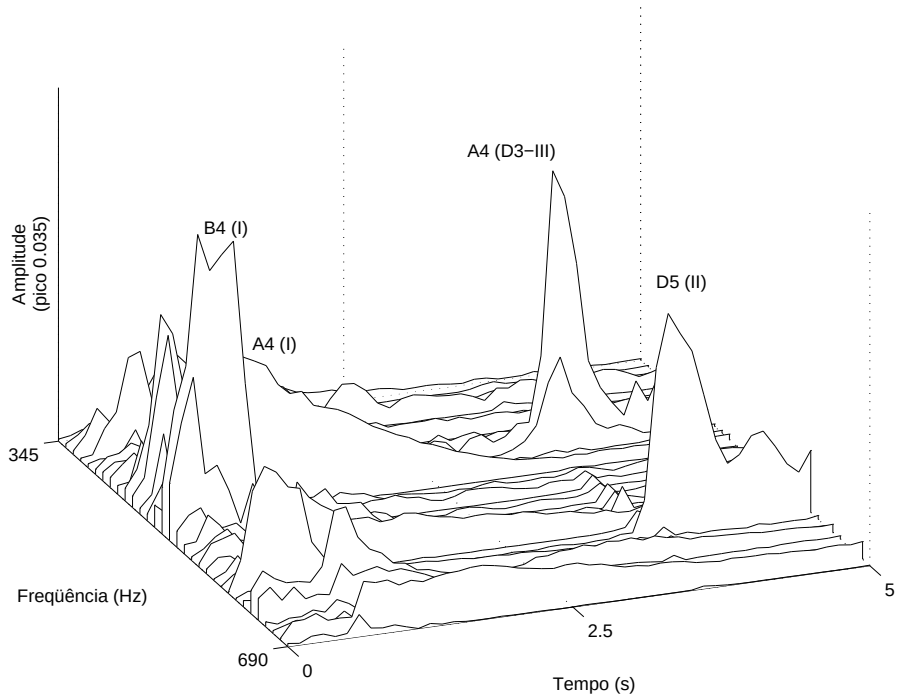

Figura 7.22: Trecho de Villa-Lobos analisado com a mBQFFB, visualização em 3D da oitava $d = 5$.

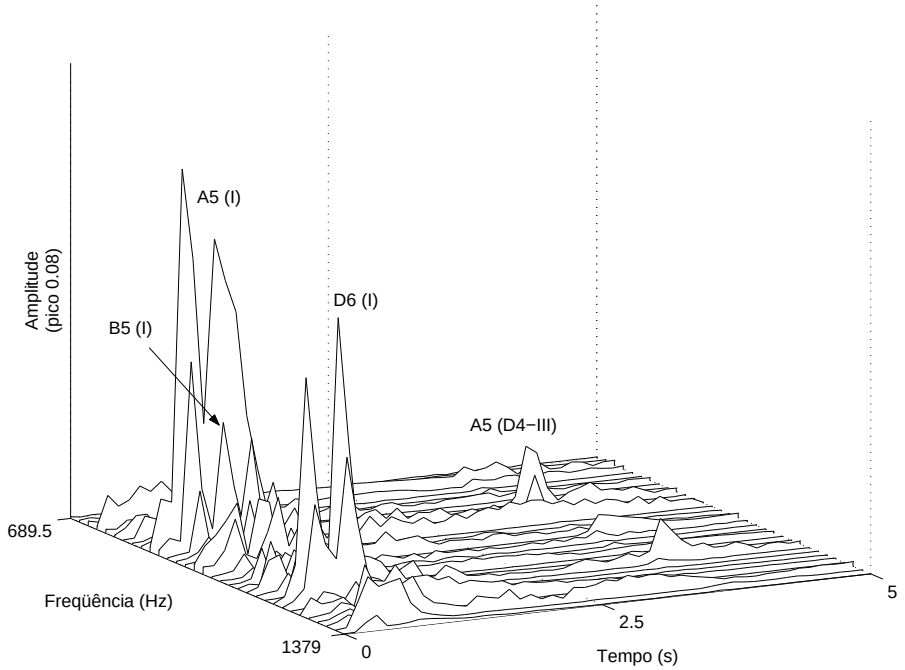

Figura 7.23: Trecho de Villa-Lobos analisado com a mBQFFB, visualização em 3D da oitava $d = 6$.

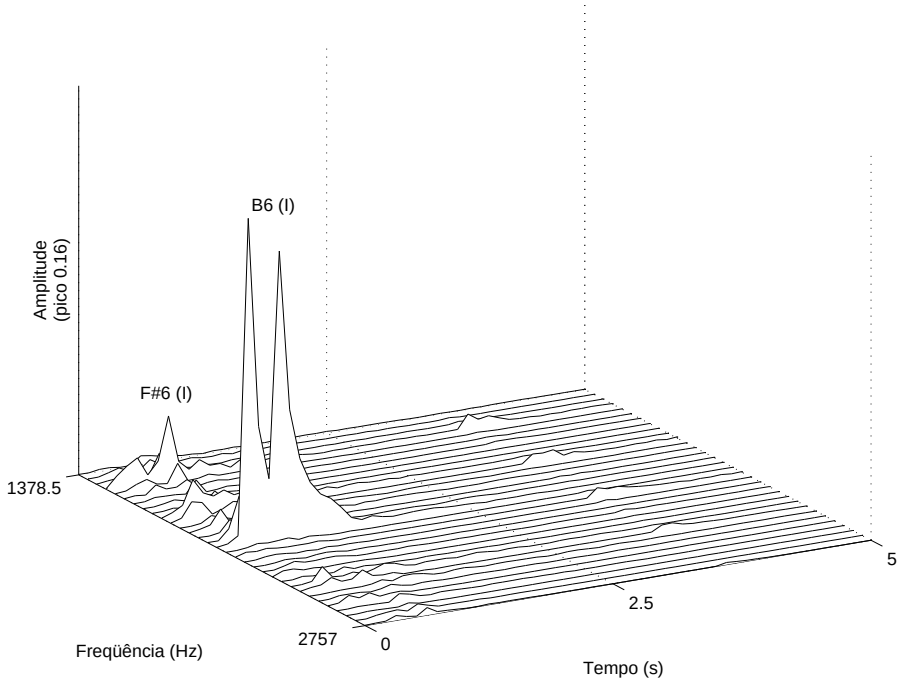

Figura 7.24: Trecho de Villa-Lobos analisado com a mBQFFB, visualização em 3D da oitava $d = 7$.

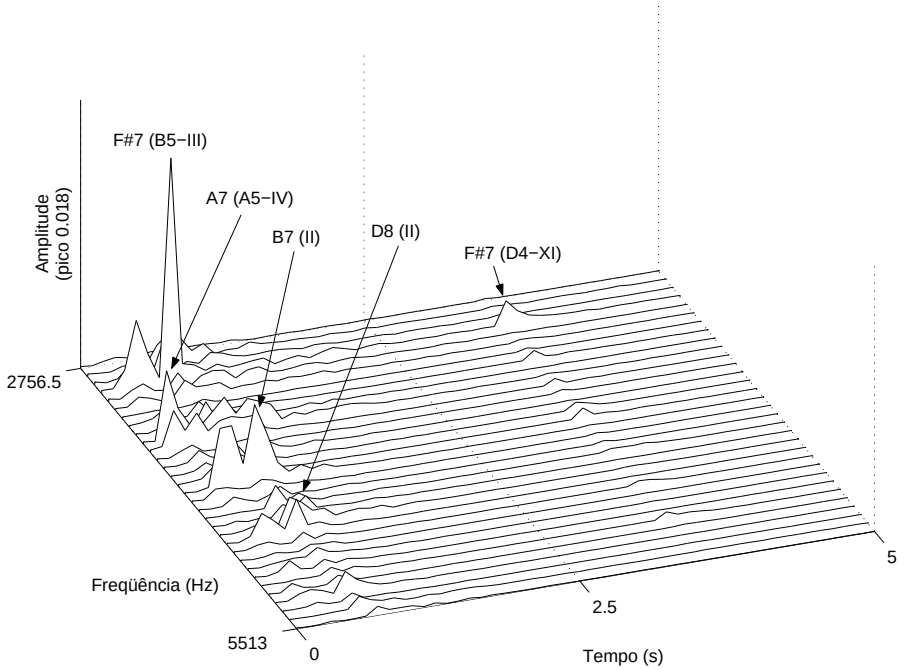

Figura 7.25: Trecho de Villa-Lobos analisado com a mBQFFB, visualização em 3D da oitava $d = 8$.

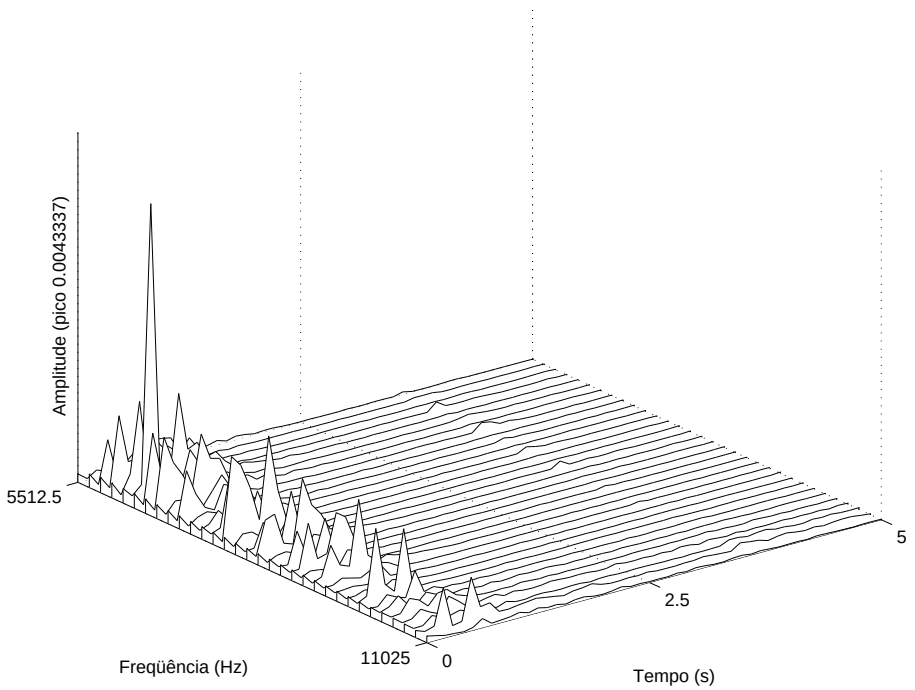

Figura 7.26: Trecho de Villa-Lobos analisado com a mBQFFB, visualização em 3D da oitava $d = 9$.

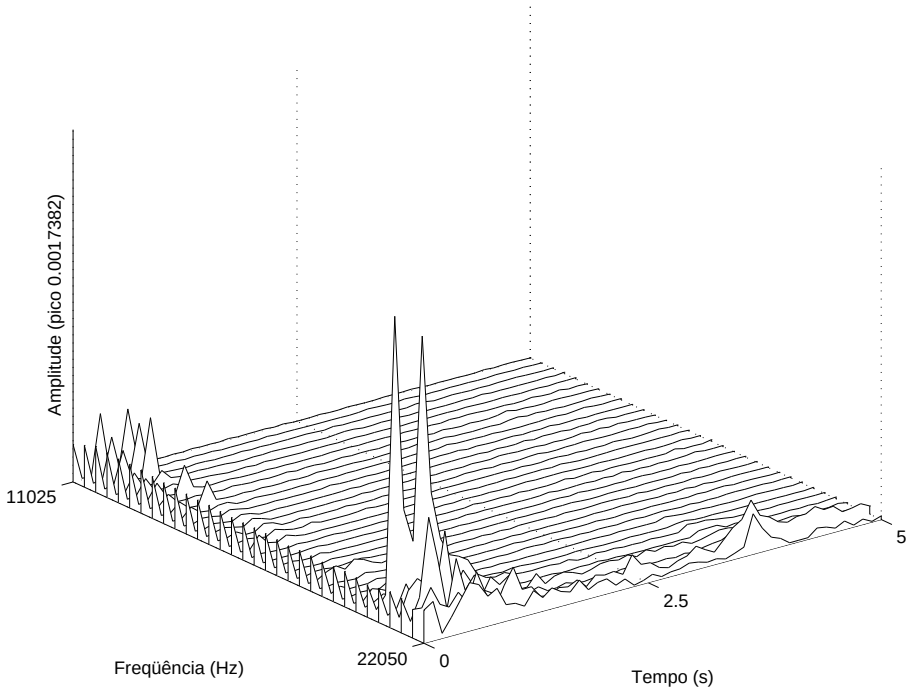

Figura 7.27: Trecho de Villa-Lobos analisado com a mBQFFB, visualização em 3D da oitava $d = 10$ (mais alta).

7.2 Testes Complementares da mBQFFB

Como a mBQFFB apresentou melhor desempenho que as demais técnicas, levando-se em conta o balanço entre número de canais, seletividade e complexidade computacional, resolveu-se observar o seu desempenho em sinais de áudio com características dificultantes. Para melhor entendimento, os exemplos incluem a partitura da peça gravada. Foi utilizada a mesma configuração descrita na Seção 7.1, a menos dos filtros separadores das oitavas, cujas especificações aqui são as mesmas dos filtros internos.

Em cada figura abaixo da partitura é mostrada uma seqüência de gráficos. Cada um representa, em níveis de cinza, a amplitude dos 32 canais (numa escala em Hz) no interior de uma oitava versus tempo (em segundos). Somente as oitavas com conteúdo significativo são mostradas nas figuras. Para facilitar a visualização, como a potência abrange uma faixa dinâmica ampla ao longo das diversas regiões do espectro, o gráfico de cada oitava foi normalizado em amplitude. Isso impede a comparação quantitativa entre as oitavas diferentes.

7.2.1 Flauta Solo

Este é um trecho de uma gravação de uma peça para flauta solo composta por J.S. Bach (1685-1750).

Neste exemplo monofônico, como pode ser visto na Figura 7.28, predominam as freqüências de médias e agudas (entre 345Hz e 1300Hz) executadas em andamento moderado. As notas não estão perfeitamente sincronizadas devido à expressividade do músico. É possível observar claramente as harmônicas de certas notas, principalmente na oitava de 1379Hz a 2757Hz. Na segunda nota, $B5$, indicada com a letra A na figura, observa-se o efeito de *vibrato*, que mostra a nota alternando-se por pelo menos dois canais. A localização das notas é boa devido à acústica seca (pouca reverberação). O trinado do $G\sharp 5$ pode ser observado pouco depois dos 2,5 segundos, indicado pela letra B na figura. Devido à normalização, a oitava de 2757Hz a 5513Hz mostra poucas harmônicas localizadas.

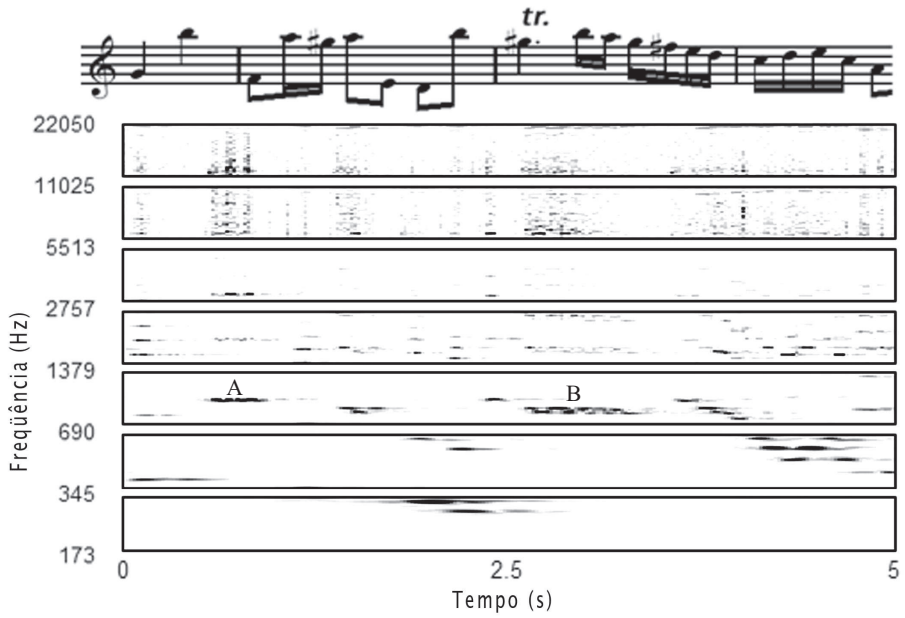

Figura 7.28: Análise mBQFFB: Trecho da Partita em Lá menor para flauta solo, BWV 1013, de J. S. Bach.

7.2.2 Piano Solo

Este exemplo para piano solo mostra a evolução temporal de uma peça de duas vozes que abrange uma ampla faixa espectral. Como pode ser observado na Figura 7.29, a linha melódica da mão direita pode ser facilmente acompanhada no espectro ao longo do tempo (ver letra A na figura); isso graças ao fato de todas as notas terem igual valor, e o pianista ter empregado uma dinâmica restrita e andamento preciso. O trecho é bastante rápido, com a mão direita tocando até 13 notas por segundo; devido à rapidez e ao toque *legato*, percebem-se intersecções de notas. As notas da mão esquerda são observadas nas duas oitavas inferiores (indicada pela letra B na figura), inclusive as suas harmônicas até as oitavas superiores. Isso por causa do *sforzando* (ou *sf*), que indica que a nota deve ser tocada com súbita intensidade. Além disso, mesmo tendo curta duração, apresentam um decaimento mais lento devido à ressonância.

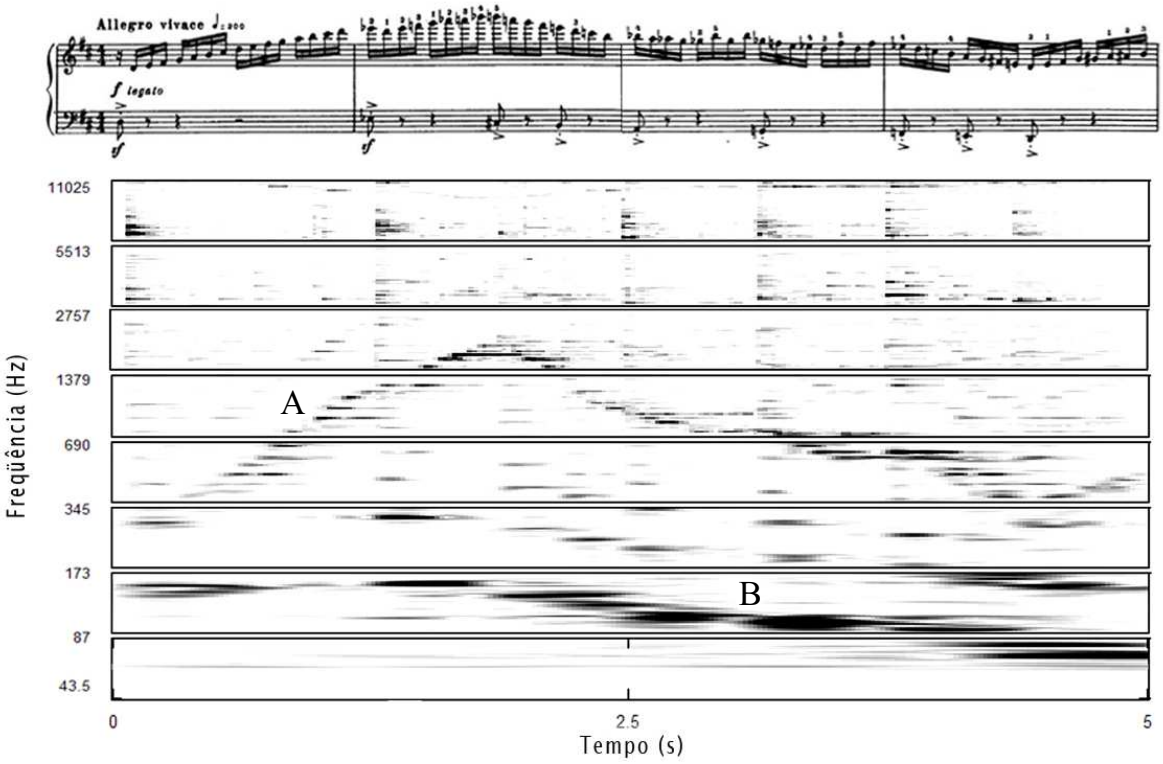

Figura 7.29: Análise por mBQFFB: Trecho do Prelúdio em Ré maior para piano solo, Op. 34/5, de D. Shostakovich.

Capítulo 8

Síntese

A seguir serão sugeridos procedimentos para a síntese associada às ferramentas de análise por Bancos de Filtros abordadas na tese.

Para a FFB, basta somar a saída dos canais, similarmente à *s*FFT. Já os métodos de BQ e CQ não são estruturalmente inversíveis e requerem outras soluções para síntese do sinal.

Na BQFFB deve-se primeiro interpolar adequadamente as oitavas inferiores, de acordo com o grau de decimação realizado em cada subdivisão e, na seqüência, somar as saídas (como na FFB). Porém, é necessário tratar corretamente as regiões-limites das oitavas.

No caso dos métodos de CQFFB, deve ser projetado um banco de filtros de síntese, que, no entanto, poderá apenas aproximar a reconstrução perfeita. Isso acontece, já devido à não-inversibilidade do método original da CQT [44]. Como alternativa, podem-se utilizar os parâmetros de alto nível obtidos pela análise para alimentar um sintetizador genérico.

A síntese por parâmetros de alto nível é um dos tópicos abordados pela área de *Music Information Retrieval*. Denominada de Transcrição Musical Automática (TMA), ela utiliza as informações de tempo e freqüência para gerar uma partitura, i.e., um arquivo MIDI. Não há necessidade de buscar as informações de fase para fazer a continuação das trilhas, porém surgem outras dificuldades como, por exemplo, identificar a nota fundamental a partir do emaranhado de parciais, harmônicas ou não.

Capítulo 9

Conclusão

Este trabalho examinou duas famílias de soluções para o problema da análise espectral de sinais de áudio: baseadas em transformadas de blocos e baseadas em bancos de filtros.

9.1 Nossa Contribuição

Na primeira parte (Soluções por Transformada de Blocos) podem ser citadas como contribuições:

- A análise de diversos métodos de refinamento espectral sob uma notação e interpretação comum;
- A inserção do método da diferença de fase iterativo (originalmente proposto em [35]) nesse contexto, pela primeira vez;
- A análise de desempenho dos métodos quanto à resolução freqüencial, inclusive com ruído.

Isso pavimenta um trabalho de pesquisa mais aprofundado a ser conduzido sobre métodos de análise espectral refinada.

Na segunda parte (Soluções por Bancos de Filtros):

- A partir da CQFFB e BQFFB originariamente definidas como transformada e avaliadas apenas em termos estáticos, concluiu-se a tarefa de criar algoritmos para utilização dos métodos ao longo do tempo, i.e., como bancos de filtros;

- Realizaram-se testes para averiguar o desempenho dinâmico dos métodos, isto é, como representações tempo-freqüência;
- Foi proposto um novo algoritmo, o mBQFFB, que acabou-se tornando o de melhor desempenho, balanceando número reduzido de canais, elevada seletividade e baixa complexidade;
- Detalhou-se a complexidade computacional de todos os métodos.

9.2 Possível Extensão da Pesquisa

Como possível extensão da pesquisa pode-se citar:

Para a primeira parte:

- Realizar testes estatisticamente sistemáticos, que permitam definir acurácia e precisão dos métodos;
- Reformular os métodos diretamente no domínio discreto;
- Procurar uma janela otimizada para associar ao(s) novo(s) método(s);
- Projetar um sistema genérico de análise e síntese que englobe todos os métodos;
- Verificar qual o limite teórico da resolução freqüencial dos métodos.

Para a segunda parte:

- Desenvolver um algoritmo rápido para a CQFFB;
- Projetar um banco de filtros com uma descrição simultaneamente geométrica (para as notas) e linear (para as suas harmônicas);
- Buscar solução adaptativa para a análise espectral, conforme as características locais do sinal e/ou informação a-priori (como a partitura);
- Aplicar os métodos de BQ e CQ para extração de parâmetros de alto nível a serem usados em sistemas de MIR (*Music Information Retrieval*).

Referências Bibliográficas

[1] BENSON, D., *Mathematics and Music*. Livro Eletrônico, 2003. http://www.maths.abdn.ac.uk/∼bensondj/html/maths-music.html.

[2] BUTLER, D., *The Musician's Guide to Perception and Cognition*. Nova Iorque, NY, EUA, Schirmer, 1992.

[3] FLANAGAN, J. L., GOLDEN, R. M., "Phase vocoder", *Bell System Technical Journal*, v. 45, pp. 1493–1509, 1966. http://www.ee.columbia.edu/∼dpwe/e6820/papers/FlanG66.pdf.

[4] QUATIERI, T. F., MCAULAY, R. J., "Speech analysis/synthesis based on a sinusoidal representation", v. 34, n. 4, pp. 744–54, 1986.

[5] J. O. SMITH, I., SERRA, X., "PARSHL: an analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation". In: *Proc. ICMC '87 - International Computer Music Conference*, pp. 290–297, International Computer Music Association, Urbana, IL, EUA, Agosto 1987.

[6] RODET, X., DEPALLE, P., GARCIA, G., "New possibilities in sound analysis and synthesis". In: *Proc. IMSA '95 - International Symposium on Musical Acoustics*, pp. 1–12, Société Française d'Acoustique, Dourdan, France, Julho 1995.

[7] RODET, X., "Musical sound signal analysis/synthesis: sinusoidal+residual and elementary waveform models". In: *Proc. ISTFTSA '97 - International Symposium on Time-Frequency and Time-Scale Analysis*, v. 4, pp. 131–141, IEEE, Coventry, Reino Unido, Junho 1997.

[8] GOODWIN, M., "Residual modeling in music analysis-synthesis". In: *Proc. ICASSP '96 - International Conference on Acoustics, Speech, and Signal Processing*, v. 2, pp. 1005–1008, IEEE, Atlanta, GA, EUA, Maio 1996.

[9] GOODWIN, M., VETTERLI, M., "Time-frequency signal models for music analysis, transformation, and synthesis". In: *Proc. ISTFTSA '96 - International Symposium on Time-Frequency and Time-Scale Analysis*, v. 1, pp. 133–136, IEEE, Paris, França, Junho 1996.

[10] GOODWIN, M., "Nonuniform filterbank design for audio signal modeling". In: *Proc. Asilomar CSSC '96 - Asilomar Conference on Signals, Systems and Computers*, v. 2, pp. 1229–1233, IEEE, Pacific Grove, CA, EUA, Novembro 1996.

[11] LAROCHE, J., DOLSON, M., "New phase vocoder technique for pitch-shifting, harmonizing and other exotic effects". In: *Proc. of WASPAA '99 - Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 91–94, IEEE, New Paltz, NY, EUA, Outubro 1999.

[12] HANNA, P., DESAINTE-CATHERINE, M., "A statistical and spectral model for representing noisy sounds with short-time sinusoids", *EURASIP Journal on Applied Signal Processing*, v. 2005, n. 12, pp. 1794–1806, 2005.

[13] POLOTTI, P., MENZER, F., EVANGELISTA, G., "Inharmonic sound spectral modeling by means of fractal additive synthesis". In: *Proc. DAFX '02 - International Conference on Digital Audio Effects*, pp. 127–132, EU-COST-G6, Hamburgo, Alemanha, Setembro 2002.

[14] WELLS, J. J., MURPHY, D. T., "Real-time partial tracking in an augmented additive synthesis system". In: *Proc. DAFX '02 - International Conference on Digital Audio Effects*, pp. 93–96, EU-COST-G6, Hamburgo, Alemanha, Setembro 2002.

[15] VOGEL, B., JORDAN, M. I., WESSEL, D., "Multi-instrument musical transcription using a dynamic graphical model". In: *Proc. ICASSP '05 - International Conference on Acoustics, Speech, and Signal Processing*, v. 5, pp. 493–496, IEEE, Filadélfia, PA, EUA, Março 2005.

[16] RODET, X., "Sound analysis, processing and synthesis tools for music research and production". In: *Proc. CIM '00 - Colloquium on Musical Informatics*, pp. 1–8, Instituto Gramma, L'Aquila, Itália, Setembro 2000.

[17] WRIGHT, M., CHAUDHARY, A., FREED, A., KHOURY, S., *et al.*, "Audio applications of the sound description interchange format standard". In: *Proc. 107th. AES Convention*, v. 1, pp. 1–30, AES, Nova Iorque, NY, EUA, Setembro 1999.

[18] QUATIERI, T. F., MCAULAY, R. J., "Audio signal processing based on sinusoidal analysis/synthesis". In: Brandenburg, K., Kahrs, M. (eds.), *Applications of Digital Signal Processing to Audio and Acoustics*, chapter 9, Norwell, MA, USA, Kluwer, 1998.

[19] DOLSON, M., "The phase vocoder: a tutorial", *Computer Music Journal*, v. 10, n. 4, pp. 14–27, 1986. http://www.panix.com/∼jens/pvoc-dolson.par.

[20] LAROCHE, J., "Time And pitch scale modification of audio signals". In: Brandenburg, K., Kahrs, M. (eds.), *Applications of Digital Signal Processing to Audio and Acoustics*, chapter 7, Norwell, MA, EUA, Kluwer, 1998.

[21] HAMMER, F., *Time-Scale Modification Using the Phase Vocoder*. Tese de mestrado, Graz University of Music and Dramatic Arts, Institute for Electronic Music and Acoustics, Graz, Áustria, 2001. citeseer.ist.psu.edu/hammer01timescale.html.

[22] TANG, M., WANG, C., SENEFF, S., "Voice transformations: from speech synthesis to mammalian vocalization". In: *Proc. of Eurospeech '01 - European Conference on Speech, Communication and Technology*, pp. 353–356, International Speech Communication Association, Aalborg, Dinamarca, Setembro 2001.

[23] SERRA, X., SMITH, J. O., "Spectral modeling synthesis: a sound analysis/synthesis system based on a deterministic plus stochastic decomposition", *Computer Music Journal*, v. 14, n. 4, pp. 14–24, 1990.

[24] SERRA, X., "Musical sound modeling with sinusoids plus noise". In: Roads, C., Pope, S., Picialli, A., Poli, G. D. (eds.), *Musical Signal Processing*, Lisse, Holanda, Swets and Zeitlinger, pp. 91–122, 1997.

[25] ESQUEF, P. A. A., *Spectral-Based Sound Synthesis—a Review*, Technical report, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Helsinki, Finlândia, 2003.

[26] VERMA, T., LEVINE, S., MENG, T., "Transient modeling synthesis: a flexible analysis/synthesis tool for transient signals". In: *Proc. ICMC '97 - International Computer Music Conference*, pp. 164–167, International Computer Music Association, Grécia, Setembro 1997.

[27] SCHEIRER, E. D., *Extracting Expressive Performance Information from Recorded Music*. Tese de mestrado, MIT Media Laboratory, Cambridge, MA, EUA, 1995. http://web.media.mit.edu/∼eds/thesis.pdf.

[28] HAYKIN, S., VAN VEEN, B., *Sinais e Sistemas*. Porto Alegre, RS, Brasil, Bookman, 2001.

[29] AUGER, F., FLANDRIN, P., "Improving the readability of time-frequency and time-scale representations by the reassignment method", *IEEE Transactions on Signal Processing*, v. 43, n. 5, pp. 1068–1089, 1995.

[30] FITZ, K., HAKEN, L., "On the use of time-frequency reassignment in additive sound modeling", *Journal of the Audio Engineering Society*, v. 50, n. 11, pp. 879–893, 2002.

[31] FRIEDMAN, D., "Instantaneous-frequency distribution vs time: an interpretation of the phase structure of speech". In: *Proc. ICASSP '85 - International Conference on Acoustics, Speech, and Signal Processing*, v. 3, pp. 1121–1124, IEEE, Tampa, FL, EUA, Março 1985.

[32] DINIZ, P. S. R., da SILVA, E. A. B., NETTO, S. L., *Processamento Digital de Sinais: Projeto e Análise de Sistemas*. Porto Alegre, RS, Brasil, Bookman, 2004.

[33] HAINSWORTH, S. W., MACLEOD, M. D., *Time Frequency Reassignment: a Review and Analysis*, Technical Report CUED/F-INFENG/TR.459, Cambridge University Engineering Department, Cambridge, Reino Unido, 2003.

[34] BROWN, J. C., PUCKETTE, M. S., "A high resolution fundamental frequency determination based on phase change of the fourier transform", *Journal of the Acoustical Society of America*, v. 94, n. 2, pp. 662–667, 1993.

[35] DAVID, P. A. M.-S., SZCZUPAK, J., "Refining the digital signal spectrum". In: *Proc. MWSCAS '96 - Midwest Symposium on Circuits and Systems*, v. 2, pp. 767–70, IEEE, Ames, IA, EUA, Agosto 1996.

[36] DESAINTE-CATHERINE, M., MARCHAND, S., "High-precision fourier analysis of sounds using signal derivatives", *Journal of the Audio Engineering Society*, v. 48, n. 7/8, pp. 654–667, 2000.

[37] ISO/IEC 11172-3, "Information Technology: Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1,5 Mbit/s. Part 3: Audio", 1993.

[38] dos SANTOS, C. N., *Representação Espectral de Sinais para Transcrição Musical Automática*. Tese de mestrado, Universidade Federal do Rio de Janeiro, Programa de Engenharia Elétrica do COPPE, Rio de Janeiro, RJ, Brasil, 2004.

[39] dos SANTOS, C. N., NETTO, S. L., BISCAINHO, L. W. P., GRAZIOSI, D. B., "A modified constant-Q transform for audio signals". In: *Proc. ICASSP '04 - International Conference on Acoustics, Speech, and Signal Processing*, v. 2, pp. 469–472, IEEE, Montreal, Canadá, Maio 2004.

[40] FARHANG-BOROUJENY, B., LIM, Y. C., "A comment on computational complexity of sliding FFT", v. 39, n. 12, pp. 875–876, 1992.

[41] LIM, Y. C., FARHANG-BOROUJENY, B., "Fast filter bank (FFB)", v. 39, n. 5, pp. 316–318, 1992.

[42] DINIZ, F. C. C. B., KOTHE, I., BISCAINHO, L. W. P., NETTO, S. L., "High-selectivity filter banks for spectral analysis of music signals", *EURASIP Jour-*

nal on Applied Signal Processing - Special Issue: Music Information Retrieval Based on Signal Processing, , 2006. Submetido para publicação, em revisão.

[43] LEE, E., FOO, S. W., "Transcription of polyphonic signals using fast filter bank". In: *Proc. ISCAS '03 - International Symposium on Circuits and Systems*, v. 3, pp. 241–244, IEEE, Scottsdale, AZ, EUA, Maio 2003.

[44] BROWN, J. C., "Calculation of a constant Q-spectral transform", *Journal of the Acoustical Society of America*, v. 89, n. 1, pp. 425–434, 1991.

[45] BROWN, J. C., "An efficient algorithm for the calculation of a constant Q transform", *Journal of the Acoustical Society of America*, v. 92, n. 5, pp. 2698–2701, 1992.

[46] GRAZIOSI, D. B., dos SANTOS, C. N., NETTO, S. L., BISCAINHO, L. W. P., "A constant-Q spectral transformation with improved frequency response". In: *Proc. ISCAS '04 - International Symposium on Circuits and Systems*, v. 5, pp. 544–547, IEEE, Vancouver, Canada, Maio 2004.

[47] KASHIMA, K., MONT-REYNAUD, B., *The bounded-Q approach to time-varying spectral analysis*, Technical Report STAN-M-23, Stanford University, Department of Music, Stanford, CA, EUA, 1985.

[48] KLAPURI, A., *Automatic Transcription of Music*. Tese de mestrado, Tampere University of Technology, Department of Information Technology, Tampere, Finlândia, 1998.

[49] DINIZ, F. C. C. B., KOTHE, I., BISCAINHO, L. W. P., NETTO, S. L., "A bounded-Q fast filter bank for audio signal analysis". In: *Proc. of ITS '06 - International Telecommunications Symposium*, IEEE, Fortaleza, CE, Brasil, Setembro 2006. Aceito para publicação.