

RESTAURAÇÃO DE SINAIS DE ÁUDIO DEGRADADOS POR RUÍDO
IMPULSIVO

Paulo Antonio Andrade Esquef

TESE 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. Paulo Sérgio Ramirez Diniz, Ph.D.

Prof. Eduardo Antônio Barros da Silva, Ph.D.

Prof. Weiler Alves Finamore, Ph.D.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 1999

ESQUEF, PAULO ANTONIO ANDRADE

Restauração de Sinais de Áudio Degradados por Ruído Impulsivo [Rio de Janeiro] 1999

IX, 109 pp., 29,7 cm (COPPE/UFRJ, M.Sc., Engenharia Elétrica, 1999)

Tese - Universidade Federal do Rio de Janeiro, COPPE

1. Restauração de Áudio 2. Processamento Digital de Sinais

I. COPPE/UFRJ II. Título (série)

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

RESTAURAÇÃO DE SINAIS DE ÁUDIO DEGRADADOS POR RUÍDO IMPULSIVO

Paulo Antonio Andrade Esquef

Março/1999

Orientador: Paulo Sérgio Ramirez Diniz

Programa de Engenharia Elétrica

Esta dissertação aborda técnicas de processamento digital de sinais voltadas à restauração de sinais de áudio degradados, com ênfase particular em técnicas para detecção e supressão de distúrbios localizados, como ruído impulsivo e pulsos longos produzidos por arranhões. São apresentadas, originalmente: modificações de caráter empírico na técnica clássica de detecção por limiar que aumentam sua eficiência na detecção e localização de ruído impulsivo; uma adaptação da técnica de detecção por limiar para sistema em multitaxa; e uma técnica para detecção de pulsos longos em sinais com conteúdo espectral rico e dinâmica ampla.

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

RESTORATION OF AUDIO SIGNALS DEGRADED BY IMPULSIVE NOISE

Paulo Antonio Andrade Esquef

March/1999

Advisor: Paulo Sérgio Ramirez Diniz

Department: Electrical Engineering

This work features digital signal processing techniques applied to degraded audio signals restoration. It emphasizes restoration techniques for detection and suppression of localized defects, like impulsive noise and low frequency noise pulses. Original contributions presented are: empirical modifications to the classical threshold-based technique that increase performance in detection and location of impulsive noise; threshold-based technique adaptation to a multirate system; and a technique for detection of long noise pulses immersed in audio signals with rich frequency content and wide dynamic range.

a Abdalla Esquef (*in memoriam*)

Agradecimentos

Ao Dr. Paulo Sérgio Ramirez Diniz, pela orientação e o apoio dedicados a este trabalho.

Ao Prof. Luiz Wagner Pereira Biscainho, pela amizade, o apoio e o estímulo, fundamentais durante esta jornada. Obrigado.

Aos professores que contribuíram para a minha formação acadêmica.

A meus pais e irmãos, pelo apoio incondicional.

Aos colegas Alessandro Salvaterra Dutra, Eduardo Cardoso Jr. (Cabelo), Eloane Gonçalves Ramos, Fábio Pacheco Freeland e Rogério Caetano, pelas conversas instigantes e permeadas de idéias criativas e a companhia sempre agradável.

Sumário

1	Restauração de Sinais de Áudio via Processamento Digital de Sinais	1
1.1	Introdução	1
1.2	Tipos de Degradação	3
1.3	Modelo Estatístico para Sinais Musicais de Áudio	4
1.4	Noções de Psicoacústica	5
1.4.1	Audibilidade	6
1.4.2	Mascaramento e Bandas Críticas	7
1.5	Avaliação dos Resultados	9
1.6	Formatos de Áudio e Plataformas Utilizadas	11
2	Algoritmos para Supressão de Ruído Impulsivo	13
2.1	Introdução	13
2.2	Deteção por Filtragem Inversa	15
2.3	Abordagem Bayesiana	18
3	Modificações na Técnica de Deteção por Filtragem Inversa e Limiar	22
3.1	Introdução	22
3.2	Modificações no Método de Deteção via Filtragem Inversa e Limiar	23
3.2.1	Cálculo do Limiar	23
3.2.2	Uso de 2 Limiares	25
3.2.3	União de Distúrbios Adjacentes	27
3.2.4	Reiteração de Processamento num mesmo Bloco	29
3.3	Comparação dos Resultados	31

4	Supressão de Ruído Impulsivo em Multitaxas	35
4.1	Introdução	35
4.2	Sinais de Áudio em Multitaxas	36
4.2.1	Número de Sub-bandas e sua Largura de Banda	36
4.2.2	Tempo de Estacionaridade dos Sinais em Sub-bandas	41
4.2.3	Modelos Estatísticos para Sinais de Áudio em Sub-bandas	44
4.3	Implementação do Sistema e Resultados	58
4.3.1	Tempos de Processamento e Simplificações	61
5	Deteção de Pulsos Longos em Sinais de Áudio	67
5.1	Introdução	67
5.2	Caracterização dos Pulsos Longos	68
5.3	Supressão por Similaridade e por Separação de Processos AR	69
5.3.1	Supressão por Similaridade	69
5.3.2	Supressão por Separação de Processos AR	72
5.4	Deteção de Pulsos Longos	75
5.4.1	Método A	75
5.4.2	Método B	76
5.4.3	Método C	79
5.4.4	Método D	83
5.4.5	Desempenho dos Métodos de Deteção de Pulsos Longos	85
5.4.6	Pulso Longo na Interseção entre Blocos	89
6	Conclusões	91
	Referências Bibliográficas	94
A	Interpolação LSAR	98
A.1	Descrição do Algoritmo	98
A.2	Limitações do Interpolador LSAR	99
A.3	Interpoladores Alternativos	101
B	Fundamentos de Processamento em Multitaxa	103
B.1	Introdução	103
B.2	Decimação e Expansão	103

B.3	Identities Nobres	105
B.4	Bancos de QMF's e CQF's	108

Capítulo 1

Restauração de Sinais de Áudio via Processamento Digital de Sinais

1.1 Introdução

A utilização de Processamento Digital de Sinais (DSP) em problemas relacionados a áudio tem crescido em importância e abrangência desde os trabalhos iniciais em DSP nos anos 60.

Talvez, a mais notória aplicação de DSP em áudio seja o sistema *Compact Disc* (CD), que, além de poder permitir uma qualidade sonora superior, também proporciona um meio de armazenamento bem menos perecível e mais prático que o dos sistemas anteriores. As vantagens dos sistemas digitais de áudio em combinação com o interesse de preservação de gravações históricas levou aos estudos dos primeiros sistemas de restauração de sinais de áudio provenientes de matrizes degradadas, que englobam desde as primeiras gravações de áudio em cilindros de cera, feitas no final do século dezenove, passando pelas gravações em discos (78 rpm e LP, etc.) e as gravações em fita magnética, até as gravações em meio digital.

O objetivo ideal da restauração de áudio é a remoção de qualquer distúrbio ou modificação, criteriosamente julgada indesejável, que esteja corrompendo o sinal original ou não-ruído.

O processo de remoção de ruído é constituído, geralmente, de uma etapa

de detecção do distúrbio e a outra de reconstrução do sinal original, na qual são utilizadas técnicas de estimação de sinais, que visam a recompor o sinal original. É importante ressaltar que também se objetiva preservar, o mais fielmente possível, o conteúdo original das gravações. Nesses termos, é desejável remover, exclusivamente, os ruídos gerados devido aos processos de gravação/reprodução da matriz e manter inalterados, por exemplo, eventos ou ocorrências que fazem ou fizeram parte de uma execução ao vivo, como o “barulho” da platéia.

Além desta Introdução, este Capítulo aborda, na Seção 1.2, os tipos de degradação que mais comumente afetam os sinais de áudio provenientes de gravações históricas; na Seção 1.3, a modelagem estatística para sinais de áudio; na Seção 1.4, aspectos ligados à percepção auditiva humana; na Seção 1.5, aspectos ligados à avaliação da qualidade dos sinais restaurados; e na Seção 1.6, o formato dos sinais de áudio utilizados neste trabalho.

Os demais Capítulos desta dissertação são resumidos a seguir.

No Capítulo 2, são revisadas duas técnicas relevantes de processamento para detecção/supressão de ruído impulsivo — a primeira, baseada em detecção por limiar, e a segunda, mais sofisticada, baseada em detecção Bayesiana. No Capítulo 3, são propostas modificações de caráter empírico na técnica de detecção por limiar que, efetivamente, melhoram seu desempenho na detecção de distúrbios impulsivos. No Capítulo 4, é avaliada uma adaptação da técnica de detecção por limiar a um sistema em multitaxa. No Capítulo 5, são propostas técnicas de detecção de pulsos longos, produzidos por arranhões, imersos em sinais de áudio com conteúdo espectral rico e dinâmica ampla. No Capítulo 6 são tecidas as conclusões finais acerca do trabalho e os possíveis direcionamentos para trabalhos futuros. O Apêndice A aborda o algoritmo de interpolação *Least Squares Autorregressive* (LSAR) e suas variantes, além de outras alternativas de interpoladores para sinais de áudio. O Apêndice B descreve alguns conceitos básicos de sistemas em multitaxas, enfatizando os utilizados neste trabalho.

1.2 Tipos de Degradação

Há vários tipos distintos de degradação de ocorrência comum em sinais de áudio provenientes de matrizes históricas. Essas degradações podem ser grosseiramente classificadas em dois grupos: o de degradações localizadas e o de degradações globais [1].

As degradações localizadas, como o nome já diz, afetam o conteúdo do sinal em trechos temporalmente restritos. São exemplos de degradações localizadas:

- **Ruído Impulsivo** – São descontinuidades de ocorrência aleatória no sinal de áudio, tanto em tempo como em amplitude. São, em geral, produzidos pela rugosidade natural das paredes do sulco da matriz e por material nelas agregado. Chegam a corromper algo em torno de 10% das amostras do sinal digitalizado (ver Seção 1.6) e soam como pequenos estalos.
- **Pulsos Longos** – São grandes estalos produzidos por arranhões, quebras ou danos similares na superfície da matriz. Os pulsos longos exibem, em geral, uma descontinuidade de alta amplitude no seu início que é seguida de um transitório de baixas frequências, associado à resposta impulsiva do sistema de reprodução. Dependendo da disposição do arranhão na superfície da matriz, radial, por exemplo, os pulsos longos podem ocorrer de modo aproximadamente periódico no sinal de áudio.

As degradações globais afetam todas as amostras do sinal de áudio. Como exemplos de degradações globais podem ser citados:

- **Ruído de Fundo** – Também denominado de ruído branco ou ruído de banda larga, é, principalmente, associado a gravações em meios magnéticos, mas ocorre em maior ou menor proporção nos demais meios de gravação.
- **Flutuações** – São alterações na tonalidade ou afinação do sinal de áudio, decorrentes de excentricidades ou variações de velocidade nos motores do sistema mecânico de reprodução.
- **Distorções** – São modificações, em geral de caráter não-linear, que afetam o sinal de áudio. Como exemplo, podem-se citar os efeitos de saturação em amplitude produzidos por desgastes naturais ocorridos nas paredes do sulco.

Em geral, os diversos tipos de degradações ocorrem simultaneamente no sinal de áudio; entretanto, seus tratamentos são realizados separadamente para cada tipo, mantendo alguma ordem de precedência, dependendo do caso.

1.3 Modelo Estatístico para Sinais Musicais de Áudio

Esta Seção apresenta algumas considerações acerca do modelo estatístico mais comumente utilizado para representar sinais de áudio. Definições e terminologias referentes a processos estocásticos não serão explicitamente definidas aqui, mas podem ser encontradas em [2].

A utilização de processamento estatístico voltado à sinais de áudio requer uma série de considerações e simplificações, muitas vezes sem justificativa formal, mas que são necessárias para viabilizar o processamento, em termos práticos.

A primeira barreira com que se se depara é a construção de um processo estocástico para sinais de áudio. A rigor, só se tem disponível uma função-amostra do processo estocástico e, por praticidade, é necessário supor que o processo é ergódico. A ergodicidade do processo, por sua vez, pressupõe um processo estacionário e, no caso de sinais musicais de áudio, a hipótese de estacionaridade só pode ser aceita para trechos de curta duração (até 50 ms) [3, 4].

Considerando, então, um trecho estacionário de sinal de áudio já digitalizado (ver Seção 1.6), o modelo estatístico mais comumente utilizado para representá-lo é o autorregressivo (AR) de ordem p . Conhecidos os coeficientes do modelo, $a(i)$, cada amostra do sinal nos instantes $k = p, p + 1, \dots, N - 1$ é dada por

$$x(k) = \sum_{i=1}^p a(i)x(k-i) + e(k), \quad (1.1)$$

ou seja, uma parcela preditível por uma combinação linear das p amostras anteriores ao instante k e uma parcela não-preditível $e(k)$, que é chamada de erro de predição ou excitação.

Matricialmente, pode-se escrever (1.1) como

$$\mathbf{e} = \mathbf{x}_1 - \mathbf{X}\mathbf{a}, \quad (1.2)$$

onde $\mathbf{x}_1 = [x(p) \cdots x(N-1)]^T$, $\mathbf{e} = [e(p) \cdots e(N-1)]^T$, $\mathbf{a} = [a(1) \cdots a(p)]^T$ e

$$\mathbf{X} = \begin{bmatrix} x(p-1) & x(p-2) & \cdots & x(0) \\ x(p) & x(p-1) & \cdots & x(1) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-2) & x(N-3) & \cdots & x(N-p-1) \end{bmatrix}.$$

Uma das formas de calcular os coeficientes do modelo AR consiste em arbitrar uma função custo, derivá-la em relação aos coeficientes e igualar a derivada da função a zero. Na abordagem por mínimos quadrados ou do inglês *Least Squares* (LS), utiliza-se como função custo $F = \mathbf{e}^T \mathbf{e}$. Nesse caso, a minimização de F por \mathbf{a} resulta em

$$\hat{\mathbf{a}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{x}_1 \quad (1.3)$$

como estimativa dos coeficientes do modelo AR de ordem p de \mathbf{x}_1 .

A escolha do tamanho N do bloco é diretamente associada às considerações de tempo de estacionaridade para trechos de sinal de áudio com curta duração. Intuitivamente, é plausível que o tempo de estacionaridade possa variar bastante de acordo com as características intrínsecas de cada sinal de áudio; de qualquer modo, na maioria dos casos, é apropriado utilizar blocos com 500 a 2000 amostras a 44100 Hz de taxa de amostragem [5, 3].

Tal como a escolha de N , a escolha da ordem p do modelo AR depende das características do sinal. A escolha de uma ordem $p = 100$ pode ser necessária para que o modelo represente adequadamente sinais de áudio com diversidade instrumental e riqueza timbrística, enquanto $p = 30$ pode ser bastante razoável para sinais mais simples.

Além do modelo AR, existem outros modelos que podem ser admitidos para sinais de áudio. Dentre eles, podem ser citados são o modelo Autorregressivo Média-Móvel ou *Autorregressive Moving-Average* (ARMA) [6] e o modelo senoidal [5], que, entretanto, são menos flexíveis, analiticamente, do que o modelo AR.

1.4 Noções de Psicoacústica

A Psicoacústica é uma área multidisciplinar de estudos que engloba aspectos físicos, biológicos, biofísicos, psicológicos e musicais relacionados à percepção

auditiva e à geração/propagação do som.

Como a restauração de áudio está intimamente associada à questão da qualidade sonora e aos fatores ou características que tornam o som agradável ou não, é importante abordar, nesta Seção, conceitos relevantes da percepção auditiva, como Audibilidade e Mascaramento/Bandas Críticas.

1.4.1 Audibilidade

Dentre os diversos fatores que influenciam a percepção da intensidade sonora, os mais evidentes são a pressão sonora e a frequência do som.

O Diagrama de Fletcher [7, 8] mostrado na Fig. 1.1 resume quantitativamente as relações entre Intensidade ou Pressão Sonora, Frequência e Nível de Audibilidade. Por convenção, o diagrama é “calibrado” tomando-se como referência (0 dB) a mínima intensidade sonora necessária para que se atinja o limiar da audição com um tom puro de frequência 1 kHz, que nesse caso vale 10^{-12} W/m² (Watt por metro quadrado). As curvas traçadas no diagrama são de Nível de Audibilidade constante (desde o Limiar da Audição de 0 dB ao Limiar da Dor de 120 dB) em relação à frequência (aproximadamente de 20 Hz a 20 kHz). Assim, como exemplo de leitura do diagrama, tem-se que, enquanto é necessária uma Intensidade Sonora de 10^{-12} W/m² para se atingir o Limiar de Audição a 1 kHz (referência de 0 dB), a 90 Hz é necessária uma Intensidade Sonora de 10^{-8} W/m², ou seja, com um Nível Sonoro 40 dB acima da referência.

Do diagrama de Fletcher pode-se concluir, diretamente, que a sensibilidade auditiva humana é muito baixa em baixas frequências, cresce em direção às frequências médias, atingindo máxima sensibilidade entre 3 kHz e 4 kHz e diminui novamente em direção às altas frequências. Observa-se também que as variações na sensibilidade são menores para Níveis de Audibilidade maiores: enquanto a variação no Nível Sonoro é da ordem de 80 dB para curva de Nível de Audibilidade 0 dB, para a curva de 120 dB a variação é da ordem de 20 dB.

Uma das aplicações em áudio diretamente decorrentes do diagrama de Fletcher é o sistema *loudness*, que age no espectro do sinal de áudio no sentido de compensar as variações de sensibilidade auditiva em relação à frequência e à intensidade sonora.

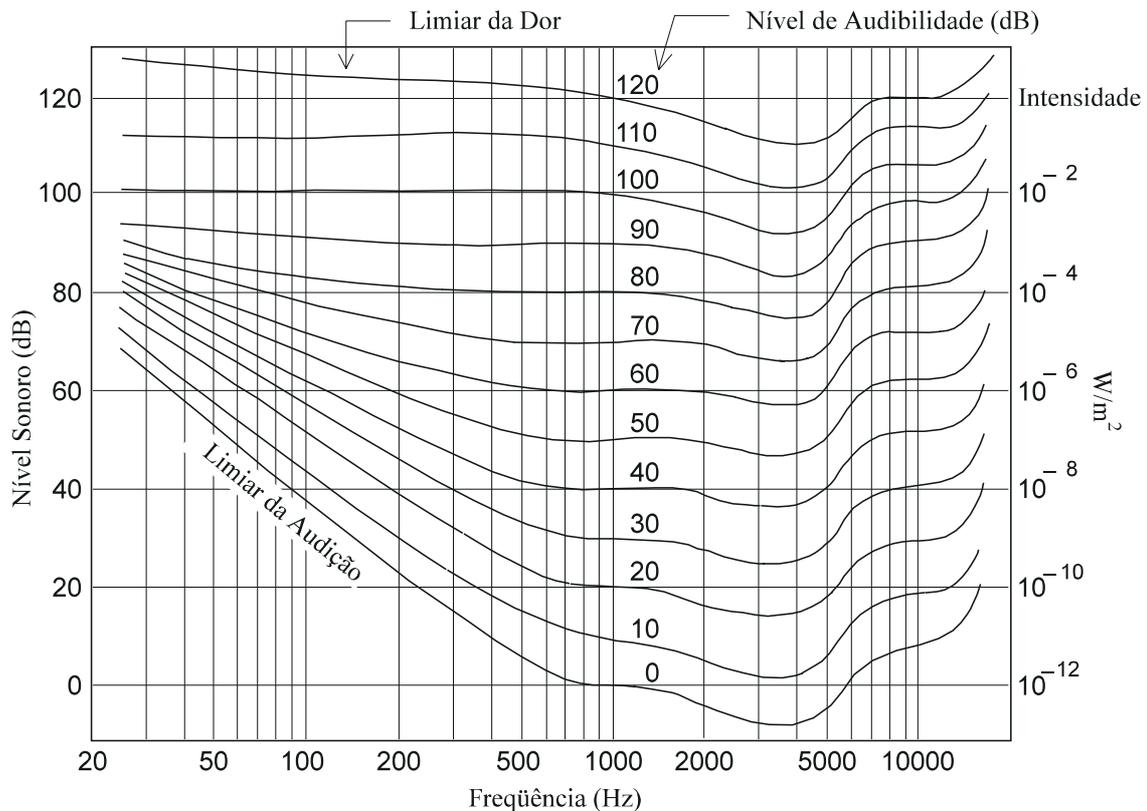


Figura 1.1: Diagrama de Audibilidade de Fletcher

1.4.2 Mascaramento e Bandas Críticas

O fenômeno de Mascaramento [8, 9] consiste no efeito de um sinal de áudio inibir a percepção auditiva de outro sinal de áudio.

As experiências iniciais para a investigação do fenômeno de mascaramento lidavam com tom puro mascarando tom puro. Posteriormente, foram também realizadas experiências, tanto de mascaramento de ruído por tom como de tom por ruído — este com densidade espectral de potência plana na faixa de áudio ou com banda estreita, no intuito de simular o mascaramento entre tons puros sem a ocorrência do fenômeno de batimento entre tons puros com frequências próximas.

Na Fig. 1.2 é apresentada uma adaptação das curvas de mascaramento obtidas por Egan e Hake, em 1950, em que o sinal mascarador é um ruído de banda estreita (90 Hz) centrado em 410 Hz e o sinal mascarado é um tom puro, com frequência variando entre 100 Hz e 5000 Hz, aproximadamente.

As curvas mostram como se altera a curva de Limiar de Audibilidade (0 dB)

para diversas Intensidades Sonoras em dB do ruído mascarador.

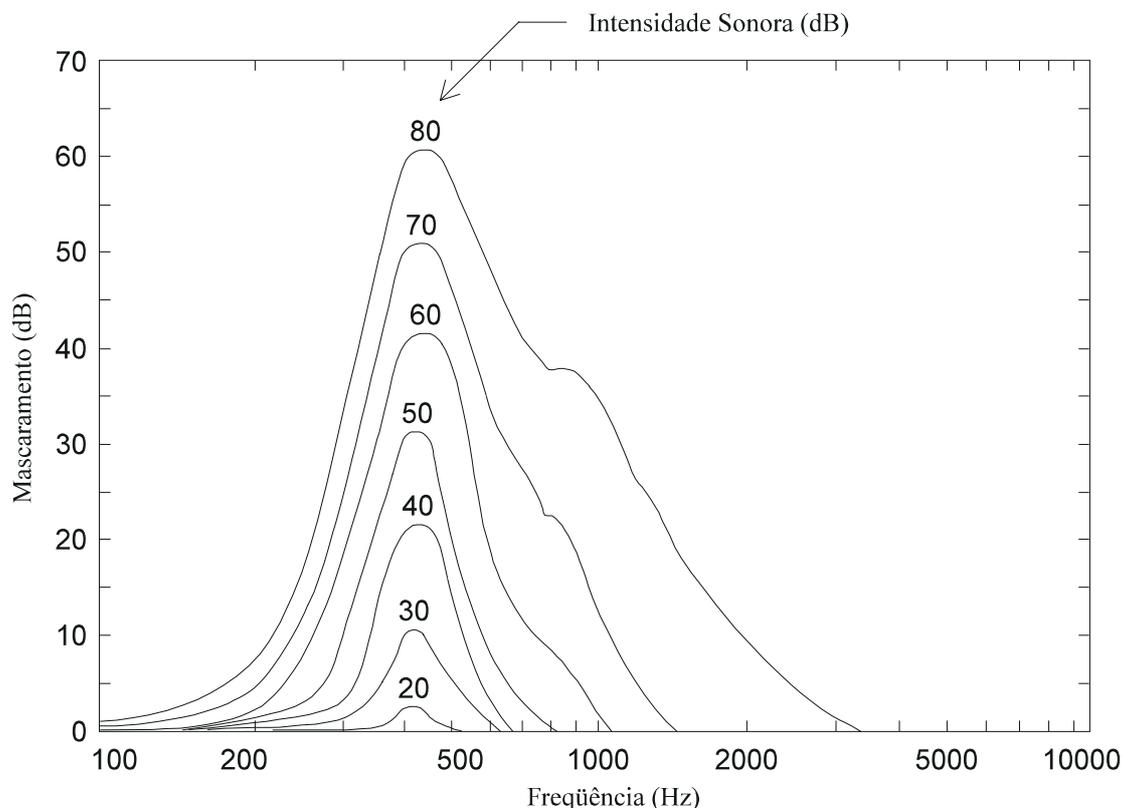


Figura 1.2: Diagrama de Mascaramento

Como pode ser observado, os maiores níveis de mascaramento ocorrem quando a frequência do tom coincide com a frequência central da banda do ruído mascarador. É interessante notar que, para intensidades sonoras elevadas do ruído mascarador (acima de 70 dB), o nível de mascaramento em frequências maiores que a do ruído mascarador é mais elevado do que o nível de mascaramento em frequências menores.

Define-se como Banda Crítica a largura de banda de um ruído mascarador centrado na frequência do tom que efetivamente contribui para o mascaramento do tom.

Os experimentos iniciais realizados por Fletcher em 1940 eram baseados na hipótese da existência de uma banda crítica estreita e de que o início do mascaramento se daria quando a potência do ruído branco na banda crítica se igualasse à potência do tom mascarado. Nesses termos, as medições eram realizadas tomando como sinal a ser mascarado um tom de teste com frequência e potência conhecidas

e como sinal mascarador um ruído branco (banda larga) com variância controlável; a variância do ruído era, então, elevada até se atingir o mascaramento do tom pelo ruído. A Banda Crítica era calculada como a banda do ruído mascarador, centrada na frequência do tom, que correspondesse à mesma potência do tom mascarado.

As Bandas Críticas medidas por Fletcher eram consideradas com resposta em frequência plana. Esse modelo de resposta plana para as bandas é tido, atualmente, como impreciso e leva a bandas mais estreitas do que as obtidas com respostas não-planas. Por esse motivo, as Bandas Críticas de Fletcher são chamadas, hoje, de Razões Críticas.

É importante ressaltar que, a rigor, o número de Bandas Críticas é ilimitado, pois pode-se calcular uma Banda Crítica para qualquer valor contínuo de frequência dentro do espectro de áudio. Na prática, as Bandas Críticas são calculadas para um número limitado de frequências suficientemente próximas, o que permite, por exemplo, usar modelos de análise espectral baseados em Bandas Críticas que podem ser realizados através de Banco de Filtros (ver Apêndice B).

Os conceitos de Audibilidade, Mascaramento e Bandas Críticas são muito utilizados em codificação e compressão de sinais de áudio. Em restauração de áudio, tais conceitos podem ser utilizados para elaboração de técnicas de restauração que gerem resultados psicoacusticamente favoráveis.

1.5 Avaliação dos Resultados

Esta Seção aborda questões relacionadas à avaliação qualitativa/quantitativa dos sinais de áudio restaurados.

Em geral, os processos de avaliação, seja qualitativa ou quantitativa, estão associados a algum tipo comparação entre as versões original e processada de um determinado sinal.

Na avaliação quantitativa, a comparação visa à medição de algum tipo de erro que possa ser traduzido em uma medida de similaridade entre as versões sob comparação; se uma das versões utilizadas for tomada como referência de qualidade absoluta, a medida quantitativa obtida pode ser utilizada como medida qualitativa, como é feito para a medição de perdas em codificação e compressão de sinais.

Em situações reais de restauração de sinais, a versão do sinal que faz o papel de referência absoluta de qualidade não está disponível, tornando inviável medir, quantitativamente, o ganho de qualidade resultante do processamento. Experimentalmente, é usual simular sinais corrompidos adicionando ruído a versões não-corrompidas dos sinais. Nesses casos particulares, dispõe-se das referências absolutas de qualidade que permitem avaliar quantitativamente a eficiência do processamento realizado, pela comparação entre os sinais não-corrompidos e as respectivas versões processadas. Naturalmente, quando se utilizam sinais artificialmente corrompidos, a avaliação quantitativa realizada é polarizada pelas condições artificialmente criadas para o sinal sob avaliação, que não necessariamente correspondem às condições reais de um sinal de áudio naturalmente corrompido. Essas condições artificialmente criadas não levam em consideração aspectos como as limitações intrínsecas do meio de armazenamento e a ocorrência simultânea de diversos tipos de degradação; ademais, os sinais de ruído artificialmente gerados podem ser meras aproximações dos ruídos realmente presentes nos sinais naturalmente corrompidos. Apesar de lidar com aproximações de sinais corrompidos, a avaliação quantitativa de sinais restaurados pode ser um bom indicativo da eficiência do processamento, mesmo quando realizado em sinais naturalmente corrompidos.

Em processamento visando à redução de ruído de fundo, o procedimento usual para simular o sinal degradado é adicionar a um sinal não-corrompido ruído branco Gaussiano com variância conhecida. A avaliação quantitativa de eficiência é realizada pela medida do erro médio quadrático entre as versões processada e não-corrompida do sinal [10, 11].

Em processamento visando à supressão de ruído impulsivo, pode-se corromper artificialmente um sinal não-corrompido com distúrbios impulsivos — gerados artificialmente segundo um modelo estatístico ou resultantes da diferença entre um sinal naturalmente corrompido e uma versão sua processada. Seja como for, se o sinal corrompido for simulado somando-se um sinal de ruído impulsivo conhecido a um sinal não-corrompido, pode-se avaliar quantitativamente o resultado do processamento medindo-se os índices de não-deteção e falsa deteção entre as versões processada e não-corrompida do sinal. Quanto menores forem esses índices maior é a eficiência do algoritmo utilizado em detectar distúrbios impulsivos.

A avaliação qualitativa de um sinal de áudio restaurado é influenciada por critérios subjetivos, propriedades da psicoacústica e uma série de outros fatores.

A subjetividade da avaliação guarda fortes laços com os padrões particulares de qualidade e percepção sonora desenvolvidos por cada indivíduo. Além de poderem variar drasticamente de um indivíduo para outro, devido à forma com que foram desenvolvidos, os padrões de qualidade podem variar fortemente de acordo com as condições biofísicas do sistema auditivo e com o estado emocional do indivíduo.

Um exemplo claro da influência de propriedades da psicoacústica na avaliação qualitativa é a dependência da percepção auditiva em relação às características acústicas do ambiente. Além das características acústicas, fatores como o ruído presente no ambiente também influenciam na percepção auditiva e no julgamento da qualidade sonora.

As considerações anteriores revelam as dificuldades de avaliar, comparativamente, a eficiência de algoritmos de restauração através de um julgamento qualitativo dos sinais de áudio restaurados. Apesar das possíveis divergências de avaliação que possam ocorrer no julgamento qualitativo, sua utilização é imprescindível, mesmo quando se dispõe de medidas quantitativas, pois nem sempre boas medidas quantitativas refletem bons resultados perceptivos, que, em última análise, são o que se deseja.

A evidência de aumento na qualidade do sinal restaurado varia também de acordo com o tipo de ruído ou distorção em questão. Em sinais de áudio degradados por variações ou flutuações na afinação, por exemplo, a própria percepção da degradação pode ser difícil para ouvidos pouco treinados, não tornando evidente o aumento de qualidade produzido pelo processamento de restauração. Por outro lado, a percepção da presença de ruído impulsivo em sinais de áudio é trivial e o aumento de qualidade produzido pela restauração fica bastante evidenciado pelo contraste entre as versões ruidosa e restaurada do sinal de áudio.

1.6 Formatos de Áudio e Plataformas Utilizadas

Os sinais de áudio utilizados neste trabalho foram digitalizados a uma taxa de 44100 Hz, com resolução de 16 bits, utilizando apenas um canal (monoaural) e

armazenados em arquivos no formato WAVE, através de placa de som de computador pessoal. Os sinais provenientes de LP's foram pré-amplificados, para compensação da curva de equalização R.I.A.A. [12], antes de serem digitalizados; já para os sinais provenientes de gravações históricas transferidas para CD's, a obtenção do arquivo no formato WAVE foi feita diretamente por comunicação digital entre o *drive* de CD-ROM e a placa de som.

Os algoritmos de restauração foram implementados no MatLab, usando aritmética de ponto flutuante, na precisão numérica da máquina. Os sinais processados foram quantizados em 16 bits antes de serem reconvertidos para o formato WAVE.

Capítulo 2

Algoritmos para Supressão de Ruído Impulsivo

2.1 Introdução

Um dos maiores problemas associados a gravações musicais históricas é a degradação do sinal de áudio por distúrbios impulsivos decorrentes de imperfeições, rugosidades ou material agregado nos sulcos das matrizes disponíveis.

Os distúrbios impulsivos podem ser caracterizados como descontinuidades localizadas de curta duração (tipicamente menores que 1ms) que corrompem o sinal de áudio em instantes de tempo aleatoriamente distribuídos ao longo do sinal. Apesar da denominação **ruído impulsivo**, os distúrbios impulsivos que corrompem o sinal de áudio apresentam, em geral, durações e amplitudes bastante variadas, conforme a extensão física do defeito na superfície da matriz. Enquanto a rugosidade intrínseca das paredes do sulco pode produzir distúrbios impulsivos de pequenas amplitude e duração, arranhões superficiais e material agregado na superfície da matriz podem produzir distúrbios de maiores amplitude e duração.

A presença dos distúrbios impulsivos não afeta o tempo de duração do sinal de áudio [13] e um modelo genérico para esse tipo de degradação consiste em considerar a corrupção por ruído impulsivo como aditiva, ou seja, o sinal corrompido como

$$y(k) = x(k) + d(k), \quad (2.1)$$

onde $x(k)$ é o sinal não-corrompido e $d(k)$ ruído impulsivo. O ruído impulsivo

também admite um modelo genérico $d(k) = i(k)n(k)$, onde $i(k)$ é um processo de chaveamento que assume valor 1 quando o distúrbio está presente e 0 caso contrário. As amplitudes dos distúrbios são regidas pelo processo $n(k)$.

O percentual de corrupção do sinal de áudio pelo ruído impulsivo pode variar bastante de um sinal degradado para outro, de acordo com o tipo da matriz e com seu estado de conservação. Por exemplo, um sinal degradado já digitalizado (ver Seção 1.6) com 10% das amostras corrompidas por ruído impulsivo é considerado severamente degradado, se todos os distúrbios forem perceptíveis. O efeito perceptivo produzido pelo ruído impulsivo varia de acordo com diversos fatores. Dentre eles podem ser citados o percentual de corrupção, as amplitudes dos distúrbios, o agrupamento das amostras dos distúrbios impulsivos — em geral, um único distúrbio corrompe um certo número de amostras contíguas do sinal e distúrbios de maior amplitude apresentam maior duração —, as características espectrais e de dinâmica do sinal de áudio não-corrompido e as propriedades da Psicoacústica.

É importante ressaltar que distúrbios localizados produzidos por arranhões profundos ou rachaduras na matriz não estão incluídos na classificação **ruído impulsivo** por apresentarem, em geral, longa duração (ver Seção 5.2). Sinais de áudio corrompidos tanto por ruído impulsivo como por ruído produzido por arranhões devem ser inicialmente processados no sentido de primeiro remover os artefatos produzidos pelos arranhões para, posteriormente, promover a supressão do ruído impulsivo.

Técnicas de processamento digital visando à supressão de ruído impulsivo em sinais de áudio discretizado podem, em geral, ser divididas em duas etapas: Detecção e Reconstrução. Na etapa de Detecção, os distúrbios impulsivos são localizados temporalmente ao longo do sinal. A etapa de Reconstrução recebe a informação sobre a localização dos distúrbios e reconstrói o sinal de áudio nessas regiões utilizando, por exemplo, um algoritmo de interpolação apropriado para substituição das amostras corrompidas por outras que melhor aproximem o comportamento do sinal de áudio não-corrompido. A qualidade final do sinal de áudio restaurado depende, obviamente, do bom desempenho de ambas as etapas.

Neste Capítulo, serão revistas duas técnicas para detecção/supressão de ruído impulsivo. A Seção 2.2 descreve a primeira técnica, originalmente apresentada

em [4], que utiliza um procedimento de detecção baseado em limiar e filtragem inversa do sinal de áudio pelo modelo AR. (ver Seção 1.3). A Seção 2.3 descreve resumidamente a técnica apresentada em [13], que adota uma abordagem Bayesiana para a detecção e remoção dos distúrbios impulsivos.

2.2 Detecção por Filtragem Inversa

Neste método, um trecho considerado estacionário do sinal não-corrompido $x(k)$ é modelado com um processo AR de ordem p , tal como em (1.1). Os coeficientes do modelo \mathbf{a} são estimados utilizando o sinal corrompido $y(k)$. Esse, por sua vez, é filtrado pelo Filtro Inverso com função de transferência

$$H(z) = 1 - \sum_{j=1}^p a(j)z^{-j}, \quad (2.2)$$

de modo a se obter o sinal de excitação correspondente

$$e(k) = y(k) - \sum_{j=1}^p y(k-j)a(j). \quad (2.3)$$

Substituindo (2.1) em (2.3), com $d(k) = i(k)n(k)$, tem-se

$$e(k) = e_x(k) + i(k)n(k) - \sum_{j=1}^p i(k-j)n(k-j)a(j), \quad (2.4)$$

onde $e_x(k)$ é a parcela da excitação correspondente apenas ao sinal não-corrompido, que pode ser considerado ruído branco Gaussiano com variância $\sigma_{e_x}^2$, se o sinal $x(k)$ tiver média zero e se o modelo AR adotado for adequado. As outras duas parcelas são, respectivamente, o ruído impulsivo propriamente dito e uma combinação linear das amostras do ruído, que produz um efeito de espalhamento por p amostras no distúrbio impulsivo original.

Para sinais de áudio com amostras fortemente correlacionadas, a excitação $\sigma_{e_x}^2$ tipicamente exibe variância algo em torno de 100 vezes menor que a variância σ_x^2 do sinal não-corrompido [5]. Isso diminui a relação sinal-ruído no domínio da excitação, evidenciando os distúrbios impulsivos e favorecendo a sua detecção, ao custo do espalhamento de p amostras no distúrbio.

A detecção dos distúrbios impulsivos é realizada comparando-se o módulo de $e(k)$ com um limiar dado por

$$\lambda = K \hat{\sigma}_{e_x}, \quad (2.5)$$

onde $\hat{\sigma}_{e_x}^2$ é uma estimativa da variância da parcela do sinal de excitação correspondente ao sinal não corrompido e K é um parâmetro que deve estabelecer um compromisso entre não-deteção (K de valor alto) e a falsa deteção (K de valor baixo). É interessante observar que a dependência do limiar em relação a $\hat{\sigma}_{e_x}$ deve, a princípio, dar conta de adaptar o valor do limiar para cada bloco, conforme as variações de potência ao longo do sinal.

Os distúrbios impulsivos num bloco podem ser detectados, basicamente, de duas formas. Na primeira forma [4] os distúrbios são detectados um a um, começando pelo de maior amplitude na excitação, através dos passos descritos a seguir:

- Detectar o distúrbio de maior amplitude na excitação pelo índice k tal que $|e(k)|$ seja máximo, com $|e(k)| > \lambda$;
- Arbitrar a duração do distúrbio considerando como corrompidas um dado número (fixo para todos os distúrbios) de amostras, adjacentes ao distúrbio detectado;
- Mapear os índices das amostras tidas como corrompidas na excitação para os índices correspondentes no sinal propriamente dito;
- Suprimir o distúrbio (o que vai minimizar a excitação no trecho corrigido);
- Repetir o procedimento para o próximo distúrbio até que $|e(k)| < \lambda, \forall k$.

A segunda forma de deteção de distúrbios [13] consiste em detectar conjuntamente os distúrbios num bloco da excitação, tomando como corrompidas todas as amostras de índice k tais que $|e(k)| > \lambda$. Nessa abordagem, a deteção e a determinação da duração dos distúrbios são realizadas em um único passo. Como na forma anterior, os índices das amostras corrompidas na excitação são mapeados nos índices do bloco de sinal de áudio propriamente dito, onde a reconstrução do sinal é realizada.

Com o intuito de aumentar o desempenho do método de deteção, é proposta em [4] a inserção de um filtro casado após a filtragem inversa. A idéia parte do pressuposto que o distúrbio impulsivo é composto de apenas 1 amostra e que a filtragem inversa transforma o sinal de áudio não-corrompido em ruído branco. Nesses termos, a excitação correspondente ao sinal ruidoso é composta de respostas impulsivas do

filtro inverso imersas em ruído branco. O filtro casado é calculado no intuito de detectar as respostas impulsivas do filtro inverso, que por extensão, correspondem a distúrbios impulsivos do sinal ruidoso. Conhecida uma estimativa dos coeficientes do modelo AR do sinal não ruidoso, $\hat{\mathbf{a}} = [1 \ -\hat{a}(1) \ -\hat{a}(2) \ \cdots \ -\hat{a}(p)]^T$, que corresponde à resposta impulsiva do filtro inverso, os coeficientes do filtro casado são dados pela reversão no tempo dessa resposta, ou seja, $\mathbf{h}_c = [-\hat{a}(p) \ -\hat{a}(p-1) \ \cdots \ 1]^T$.

A filtragem casada é capaz de amplificar a presença de distúrbios impulsivos, principalmente, os de curta duração e pequena amplitude. Em contrapartida, a inserção do filtro casado torna mais crítico o efeito de espalhamento dos distúrbios impulsivos no sinal de excitação filtrado, que passa a ser de $2p$ amostras. Para o critério de seleção de amostras corrompidas adotado em [4], que arbitra uma duração fixa para todas os distúrbios, a utilização da filtragem casada não é tão problemática. Entretanto, quando se desejar detectar e estimar a duração dos distúrbios em um único passo, como descrito anteriormente, mas agora, examinando as amostras do módulo da excitação filtrada que excedem o valor do limiar, o efeito de espalhamento acentuado produzido pela filtragem casada pode inviabilizar a consecução de boas estimativas para as durações dos distúrbios, sendo preferível utilizar, para este propósito, a excitação propriamente dita.

A Fig. 2.1 mostra o aumento na detectabilidade dos distúrbios impulsivos produzido pelas filtrações inversa e casada. Neste exemplo, adotou-se ordem igual a 40 para o modelo AR que caracteriza o sinal de áudio.

Um diagrama esquemático da técnica de detecção/supressão de ruído impulsivo apresentada nesta Seção é mostrada na Fig. 2.2.

Algumas limitações inerentes ao método descrito nesta Seção serão abordadas no Capítulo 3, onde são propostas modificações de caráter empírico no método, com vistas ao aumento do desempenho na detecção e localização de distúrbios impulsivos.

Embora existam diversos algoritmos voltados à interpolação de amostras corrompidas em sinais de áudio [14], interpoladores baseados em modelagem AR para o sinal de áudio têm se mostrado bastante eficientes na reconstrução de sinais de áudio corrompidos por ruído impulsivo. Neste trabalho, utilizou-se como algoritmo de interpolação, unicamente, o *Least Squares Autorregressive-based* (LSAR) [13] (ver Apêndice A).

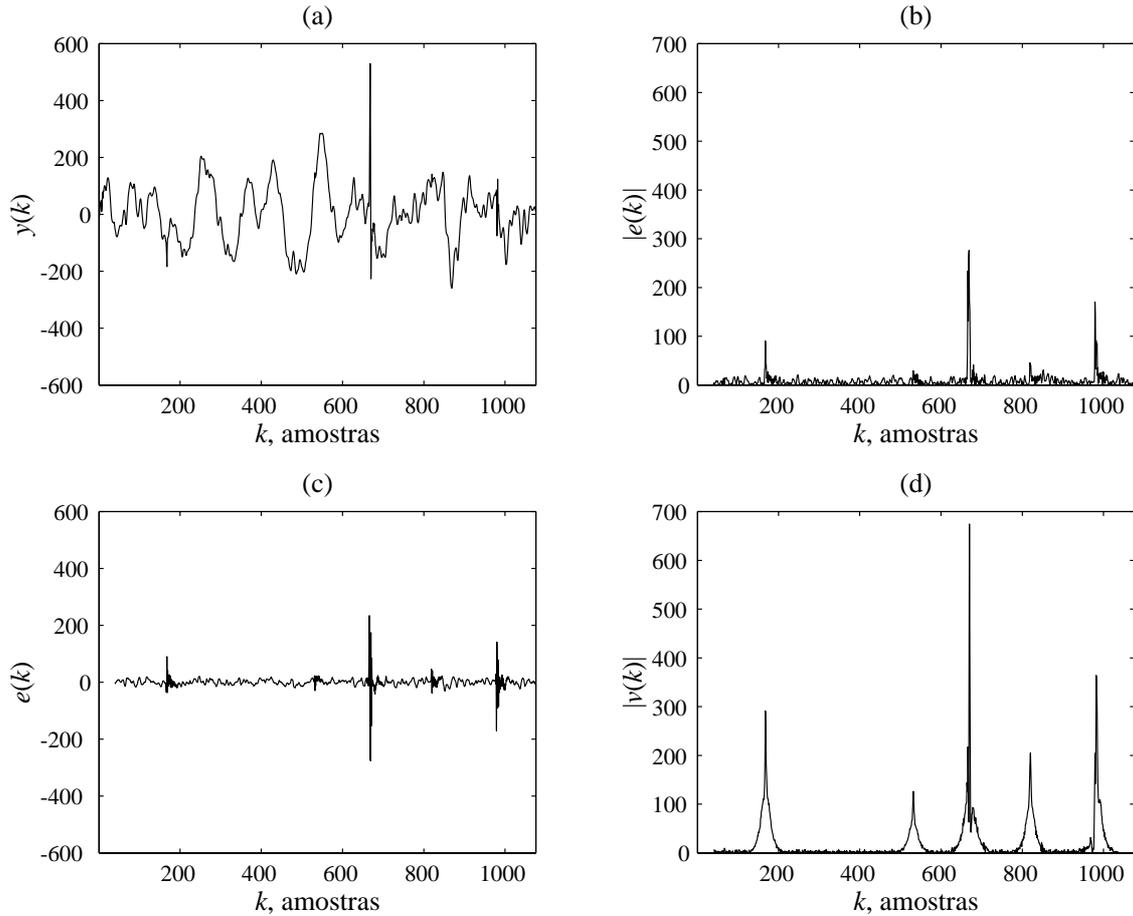


Figura 2.1: (a) Sinal de áudio corrompido por ruído impulsivo. (b) Módulo do sinal de excitação. (c) Sinal de excitação obtido por filtragem inversa. (d) Sinal obtido após a filtragem da excitação pelo filtro casado.

Uma das vantagens do algoritmo LSAR é a capacidade de interpolar conjunta e simultaneamente todos os trechos considerados corrompidos em um dado bloco de sinal. O algoritmo LSAR propriamente dito, suas características, limitações, versões modificadas e algoritmos alternativos são descritos no Apêndice A.

2.3 Abordagem Bayesiana

Na abordagem Bayesiana, a detecção dos distúrbios impulsivos é formulada, explicitamente, como a escolha do processo de chaveamento $i(k)$ (ver Seção 2.1), que maximiza sua probabilidade condicionada ao sinal ruidoso, $y(k)$. A detecção assim obtida é denominada de detecção de Máxima Probabilidade *a Posteriori* (MAP). Considerando a detecção sendo realizada em um bloco sinal com N amostras e es-

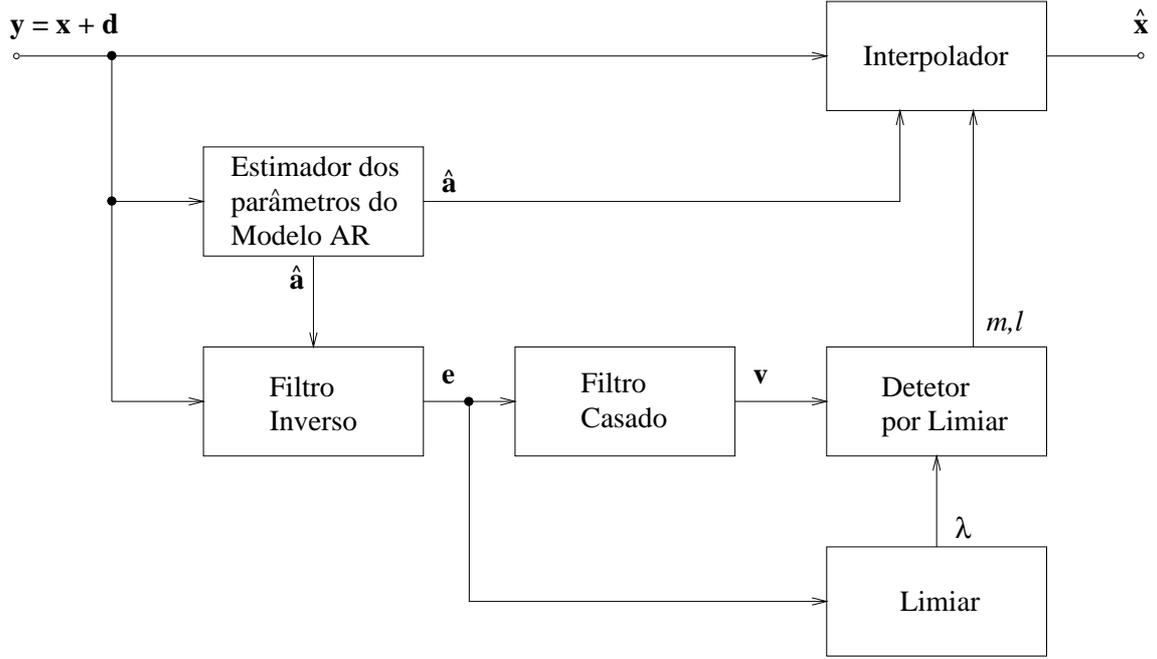


Figura 2.2: Diagrama de Blocos referente a uma iteração do processamento realizado em um bloco de sinal de áudio: \mathbf{y} , \mathbf{d} , \mathbf{x} , $\hat{\mathbf{a}}$, \mathbf{e} , \mathbf{v} e $\hat{\mathbf{x}}$ são, respectivamente, os vetores correspondentes ao sinal ruidoso, sinal de ruído, sinal original, parâmetros estimados do modelo AR, sinal de excitação, sinal de excitação após filtragem casada e sinal restaurado; λ é o limiar calculado, m a posição do início do distúrbio e l sua duração.

crevendo (2.1) vetorialmente, como

$$\mathbf{y} = \mathbf{x} + \mathbf{i}^T \mathbf{n}, \quad (2.6)$$

a detecção MAP vai envolver a escolha do vetor \mathbf{i} que maximiza a probabilidade *a posteriori* $P(\mathbf{i}|\mathbf{y})$. Pelo teorema de Bayes [2], essa probabilidade pode ser expressa por

$$P(\mathbf{i}|\mathbf{y}) = \frac{f_{\mathbf{y}|\mathbf{i}}(\mathbf{y}|\mathbf{i})P(\mathbf{i})}{f_{\mathbf{y}}(\mathbf{y})}. \quad (2.7)$$

Analisando os termos do lado direito de (2.7), tem-se que a probabilidade de detecção *a priori* $P(\mathbf{i})$ deve refletir qualquer informação previamente conhecida a respeito do processo de chaveamento — no caso de ruído impulsivo em sinais de áudio, devem ser incorporadas informações relacionadas à ocorrência dos distúrbios impulsivos em salvas aleatoriamente distribuídas ao longo do tempo, bem como as relacionadas ao percentual de corrupção. A função densidade de probabilidade $f_{\mathbf{y}}(\mathbf{y})$ assume valor constante para qualquer bloco de sinal observado. O principal problema, portanto,

consiste em calcular a função de verossimilhança $\mathbf{i} f_{\mathbf{y}|\mathbf{i}}(\mathbf{y}|\mathbf{i})$. Esse cálculo depende de uma estimativa das amostras do sinal de áudio que correspondem aos valores não-nulos de \mathbf{i} .

Os elementos de \mathbf{y} podem ser reorganizados em dois outros vetores:¹ \mathbf{y}_c , que contém os valores de \mathbf{y} quando $i(k) = 0$ e \mathbf{y}_d , que contém os valores de \mathbf{y} quando $i(k) = 1$. O modelo aditivo para o ruído impulsivo permite escrever $\mathbf{y}_d = \mathbf{x}_d + \mathbf{n}_d$, onde \mathbf{x}_d é o vetor que contém as amostras a serem estimadas do sinal de áudio. Modelando-se o sinal de áudio não-corrompido como um processo AR de ordem p , tal como em (1.1), e assumindo-se para \mathbf{n}_d uma distribuição Gaussiana com média zero e matriz-correlação genérica \mathbf{R}_{n_d} , pode-se escrever a função de verossimilhança como

$$f_{\mathbf{y}|\mathbf{i}}(\mathbf{y}|\mathbf{i}) = \frac{e^{-\frac{1}{2\sigma_e^2}[\mathbf{y}_c^T \mathbf{A}_c^T \mathbf{A}_c \mathbf{y}_c + \sigma_e^2 \mathbf{y}_d^T \mathbf{R}_{n_d}^{-1} \mathbf{y}_d + (\mathbf{A}_d^T \mathbf{A}_c \mathbf{y}_c - \sigma_e^2 \mathbf{R}_{n_d}^{-1} \mathbf{y}_d)^T \mathbf{x}_d^{\text{MAP}}]}}{\sqrt{(2\pi\sigma_e^2)^{N-p-i^T \mathbf{i}} (\mathbf{R}_{n_d} \mathbf{A}_d^T \mathbf{A}_d + \sigma_e^2)}}, \quad (2.8)$$

sendo

$$\mathbf{x}_d^{\text{MAP}} = (\mathbf{A}_d^T \mathbf{A}_d + \sigma_e^2 \mathbf{R}_{n_d}^{-1})^{-1} (-\mathbf{A}_d^T \mathbf{A}_c \mathbf{y}_c + \sigma_e^2 \mathbf{R}_{n_d}^{-1} \mathbf{y}_d), \quad (2.9)$$

\mathbf{A}_d e \mathbf{A}_c partições em colunas da matriz \mathbf{A} que armazena os coeficientes do modelo AR (ver Apêndice A) e σ_e^2 a variância do sinal de excitação.

Uma simplificação possível no cálculo de $f_{\mathbf{y}|\mathbf{i}}(\mathbf{y}|\mathbf{i})$ consiste em supor que as amostras de $n_d(k)$ são independentes com variância σ_n^2 , o que leva a

$$\mathbf{R}_{n_d}^{-1} = \frac{1}{\sigma_n^2} \mathbf{I}. \quad (2.10)$$

Outra simplificação consiste em adotar para \mathbf{i} uma distribuição uniforme, de modo que $P(\mathbf{i}) = \frac{1}{2^N}$, para qualquer \mathbf{i} .

O algoritmo de detecção de ruído impulsivo, em um bloco de N amostras, consiste, portanto, em calcular $P(\mathbf{i}|\mathbf{y})$ para as 2^N possibilidades do vetor \mathbf{i} e escolher o que exibir máxima probabilidade *a posteriori*. É importante observar que para realizar a detecção MAP, é necessário calcular, previamente, $\mathbf{x}_d^{\text{MAP}}$, o que resolve conjuntamente o problema da restauração do sinal de áudio nos trechos corrompidos.

Na abordagem Bayesiana, a incorporação de modelagem estatística para o ruído impulsivo permite extrair informação do sinal de áudio presente nos trechos corrompidos. Nesse sentido, pode-se dizer é realizada uma espécie de separação

¹Os sub-índices d e c se referem, respectivamente, a desconhecido e conhecido.

entre os sinais de áudio e de ruído, ao contrário da técnica descrita na Seção 2.2, em que a reconstrução só leva em consideração o modelo adotado para o sinal de áudio. Considerar apenas o modelo para o sinal de áudio é equivalente a fazer $\sigma_n \rightarrow \infty$ em (2.10), o que leva a $\mathbf{x}_d^{\text{MAP}} \rightarrow \mathbf{x}_d^{\text{LS}}$ (ver (A.7) no Apêndice A).

A maior desvantagem da abordagem Bayesiana em blocos é o esforço computacional excessivo requerido para computar, para cada bloco de sinal de N amostras, 2^N vezes a probabilidade *a posteriori* $P(\mathbf{i}|\mathbf{y})$, o que torna praticamente ineficaz o processamento de sinais de áudio de longa duração. Para contornar esse problema, é desenvolvida em [13] uma abordagem seqüencial sub-ótima, na qual $f_{\mathbf{y}|\mathbf{i}}(\mathbf{y}|\mathbf{i})$ é calculada, recursivamente, a cada vez que uma nova amostra de \mathbf{y} é disponibilizada. O algoritmo seqüencial realiza uma busca numa árvore binária que representa todas as possibilidades para o vetor de detecção \mathbf{i} . Entretanto, o algoritmo elimina as rotas em que a probabilidade *a posteriori* é baixa, promovendo, dessa forma, uma redução da carga computacional, em relação à busca exaustiva.

Capítulo 3

Modificações na Técnica de Detecção por Filtragem Inversa e Limiar

3.1 Introdução

Nesse Capítulo, é retomada a técnica de detecção/supressão de ruído impulsivo baseada em filtragem inversa e limiar, e são propostas algumas modificações de caráter empírico na implementação da etapa de detecção. Essas modificações se mostram eficientes, tanto para aumentar a eficiência na detecção, como para determinar localizações mais apropriadas para os distúrbios impulsivos.

Ao longo deste Capítulo, a utilização do termo **detecção** estará associada à determinação da presença de um distúrbio impulsivo corrompendo o sinal de áudio, enquanto que o termo **localização** será utilizado quando se estiver determinando a duração dos distúrbios detectados, ou seja, os índices das amostras consideradas corrompidas.

É importante ressaltar que as modificações não visam à obtenção de uma detecção ótima, como na abordagem Bayesiana. De fato, elas tendem a produzir localizações ligeiramente superestimadas para os distúrbios impulsivos, com o intuito de reduzir os resultados insatisfatórios produzidos pela etapa de reconstrução, quando os distúrbios são localizados com duração subestimada.

Na Seção 3.2, são descritas as modificações propostas. Na Seção 3.3, a efi-

ciência das modificações propostas é avaliada pelo confronto entre versões restauradas de sinais contaminados por ruído impulsivo, obtidas a partir de processamentos contando e não contando com as modificações propostas para a Etapa de Detecção.

3.2 Modificações no Método de Detecção via Filtragem Inversa e Limiar

O desempenho do método de detecção baseado em filtragem inversa e limiar pode ser aumentado através de modificações de caráter empírico em alguns aspectos de sua implementação, visando, principalmente, a uma maior eficiência na localização dos distúrbios impulsivos. Essas modificações englobam a forma de cálculo do valor do limiar, a utilização de um segundo limiar para auxiliar na localização dos distúrbios, critérios para união de distúrbios adjacentes e critérios de reiteração do processamento num mesmo trecho de sinal.

3.2.1 Cálculo do Limiar

Um dos maiores problemas relacionados à técnica de detecção por filtragem inversa e limiar é, justamente, a obtenção de limiares que realizem um compromisso satisfatório entre os índices de falsa detecção e não-detecção de distúrbios impulsivos, ao longo de todos os blocos do sinal de áudio, dadas a característica altamente não-estacionária dos sinais de áudio e a diversidade de amplitude e duração dos distúrbios impulsivos corruptores, em situações reais.

Em [4], a adaptação do valor do limiar às características locais de um determinado bloco é realizada fazendo o limiar proporcional a uma estimativa do desvio-padrão do sinal de excitação do bloco, como na eq. (2.5). Adicionalmente, em situações reais, a escolha do parâmetro K é feita, experimentalmente, de modo a realizar um compromisso satisfatório entre os índices de falsa detecção e não-detecção, como mencionado anteriormente.

A forma mais direta de se obter uma estimativa da variância da excitação correspondente ao sinal não-ruído é calculá-la utilizando a excitação do sinal ruído. Entretanto, essa forma de estimativa está muito sujeita a polarizações devido à presença da parcela associada ao ruído impulsivo.

Nas investigações experimentais realizadas em [15], observou-se que a presença dos distúrbios no sinal de excitação tende a elevar, impropriamente, o valor do limiar nos trechos de baixa potência, tendendo a elevar o índice de não-deteções nesses trechos. Se, no intuito de compensar os efeitos dessa elevação do limiar, o valor de K for reduzido, o problema é transferido para os trechos de alta potência ou de transitórios bruscos, onde a influência dos distúrbios impulsivos é menor, e a redução do limiar acaba produzindo altos índices de falsa detecção. Nesse mesmo trabalho, é investigado o uso do cálculo do limiar dependente da variância do sinal ruidoso propriamente dito. O limiar calculado dessa forma é mais sensível às variações dinâmicas do sinal ao longo do tempo, o que é psicoacusticamente favorável. Entretanto, é difícil ajustar um valor de K que promova deteções satisfatórias tanto para trechos de baixa potência como de alta potência. Em geral, a escolha dessa forma de cálculo do limiar tende a elevar o índice de não-deteções nos trechos de mais alta potência — o que, de certa forma, não é tão crítico devido ao efeito de mascaramento (ver Seção 1.4). Entretanto, nos trechos de baixa potência, observa-se uma elevação do índice de falsas deteções — o que resulta em distorções audíveis no sinal restaurado.

A consecução de um melhor compromisso entre falsa detecção e não-deteção, nos diversos blocos do sinal a serem processados, passa por uma forma de cálculo do limiar que seja pouco dependente das variações de potência do sinal e praticamente imune aos distúrbios impulsivos na excitação. Nesses termos, uma alternativa que se mostra bastante robusta na estimação da variância da excitação consiste em:

- Armazenar em um vetor as amostras do módulo do sinal de excitação, dispostas em ordem crescente de valor;
- Desprezar um determinado número de amostras de maior valor contidas no final do vetor ordenado, que provavelmente correspondem à parcela do ruído impulsivo presente na excitação;
- Utilizar como estimativa da variância da excitação do sinal não-ruidoso a média aritmética do vetor reduzido.

Na realidade, o número real de amostras corrompidas pode variar muito de bloco para bloco. Na impossibilidade de conhecê-las de antemão, arbitra-se um número

médio de amostras corrompidas por bloco — algo em torno de 10% — e descarta-se esse número de amostras do final do vetor ordenado, como descrito anteriormente.

Uma opção mais enxuta, que dispensa a necessidade de se arbitrar o número médio de amostras corrompidas por bloco, consiste em estimar a variância da excitação pela mediana do vetor de amostras do módulo da excitação do sinal corrompido. Dessa forma, o valor do limiar fica determinado por

$$\lambda = K \text{Mediana} [|e(k)|]. \quad (3.1)$$

De fato, se se considerar o sinal de excitação como um ruído branco Gaussiano estacionário, uma estimativa robusta de seu desvio-padrão é dada por $\sigma_e = \text{Mediana} [|e(k)|] / 0,6745$, como descrito em [16]. Apesar de a consideração anterior ser um tanto grosseira, a utilização de (3.1) se mostrou, efetivamente, menos suscetível, tanto às variações de potência do sinal de áudio, como à presença de ruído impulsivo.

3.2.2 Uso de 2 Limiares

A determinação precisa da duração dos distúrbios é um pré-requisito muito importante no processo de restauração; entretanto, a grande variedade de amplitudes e durações dos distúrbios e o espalhamento produzido pela filtragem inversa nos distúrbios presentes no sinal de excitação, onde serão detectados e localizados, tornam essa tarefa bastante difícil de se realizar.

A atribuição de uma duração fixa para todos os distúrbios presentes em um bloco é uma escolha que não condiz com a variedade de duração dos distúrbios impulsivos observados na realidade. De fato, a escolha de uma duração fixa que contemple distúrbios de maior duração leva, invariavelmente, à correção de um grande número de amostras não corrompidas no entorno dos distúrbios de menor duração.

O critério de detecção e localização em um único passo — dado pela escolha dos k tais que $|e(k)| > \lambda$ [13] —, pode levar a durações subestimadas para alguns distúrbios, principalmente, os de menores amplitudes, apesar do efeito de espalhamento produzido pela filtragem inversa (ver Seção 2.2). Essas determinações subestimadas de duração produzem, em geral, correções insatisfatórias na etapa de Reconstrução do sinal de áudio.

A redução do valor do limiar contribuiria, certamente, para minimizar o problema de determinação de durações subestimadas, entretanto, pode acarretar sérios prejuízos pelo aumento do índice de falsas detecções. A proposta que se faz, aqui, é utilizar a detecção e a localização em dois passos, mas agora estimando a duração dos distúrbios com o auxílio de um segundo limiar λ_L — que será chamado de Limiar de Localização —, menor que o limiar original λ_D — agora chamado de Limiar de Detecção —, por um fator b , ou seja,

$$\lambda_L = b \lambda_D. \quad (3.2)$$

Nesses termos, os distúrbios são detectados um a um, começando pelo de maior amplitude na excitação, através dos passos descritos a seguir:

- Detectar o distúrbio de maior amplitude na excitação pelo índice k para o qual $|e(k)|$ seja máximo, com $|e(k)| > \lambda_D$;
- Determinar a duração do distúrbio pelas amostras contíguas ao distúrbio detectado que exibirem módulo (na excitação) maior que o valor do Limiar de Localização λ_L ;
- Emular a supressão do distúrbio (“zerando” a excitação no trecho considerado corrompido);
- Repetir o procedimento para o próximo distúrbio até que $|e(k)| < \lambda_D, \forall k$.
- Mapear os índices das amostras tidas como corrompidas na excitação para os índices correspondentes no sinal propriamente dito.

A Fig. 3.1 ilustra esse procedimento de detecção e localização para um distúrbio impulsivo. A detecção e a localização do distúrbio são dois passos distintos, realizados individualmente para cada distúrbio. A supressão de cada distúrbio localizado poderia também ser feita individualmente, executando a etapa de Reconstrução a cada novo distúrbio localizado. A supressão do distúrbio se encarregaria de minimizar o trecho do sinal de excitação correspondente e permitiria a detecção dos distúrbios de menores amplitudes. Entretanto, é mais eficiente, computacionalmente, determinar a localização de todos os distúrbios detectáveis em um bloco e promover uma supressão simultânea de todos eles. Uma forma simples de mapear as localizações de

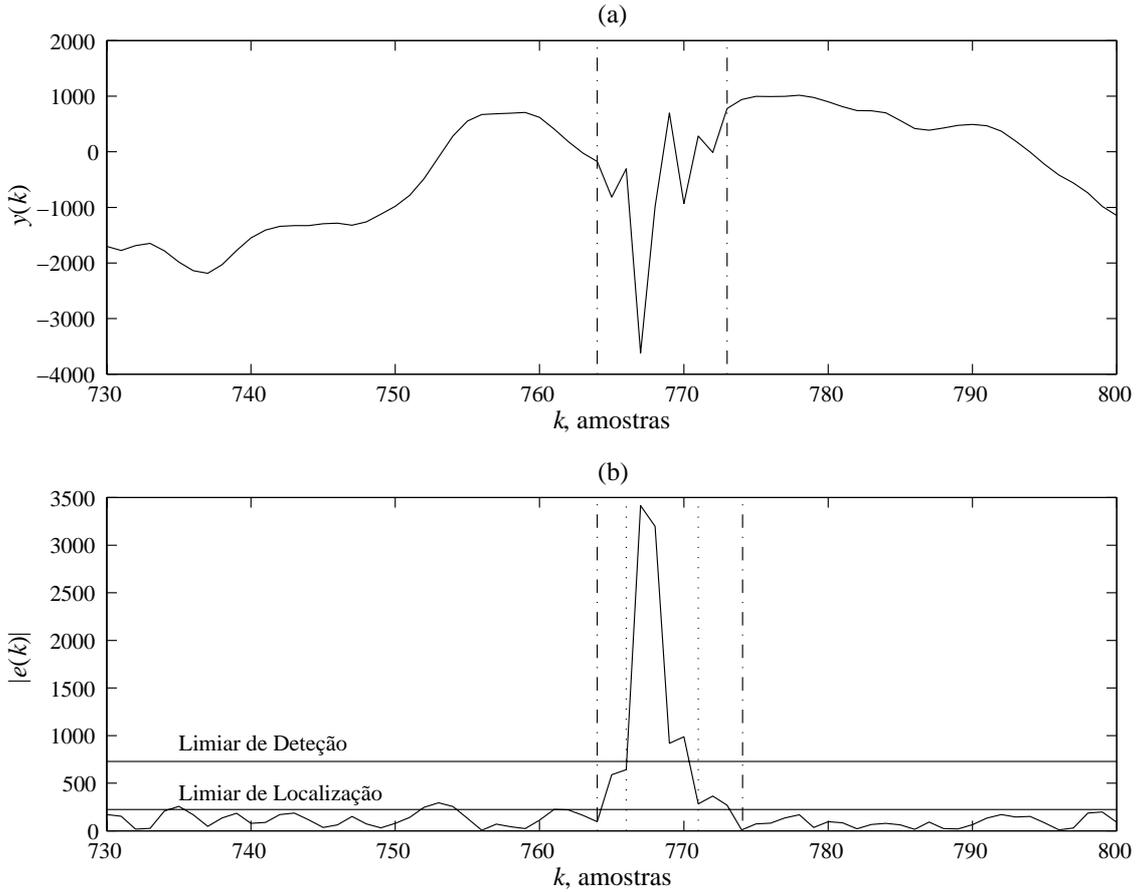


Figura 3.1: (a) Trecho de sinal de áudio corrompido por distúrbio impulsivo – delimitado pelas linhas alternadas verticais. (b) Módulo do sinal de excitação e limiares de Detecção λ_D e Localização λ_L – as linhas alternadas verticais delimitam o distúrbio localizado via λ_L , enquanto as linhas pontilhadas delimitam o distúrbio, caso fosse localizado via λ_D .

todos os distúrbios detectados num dado bloco, sem executar a etapa de Reconstrução para cada um deles, consiste em “zerar” o sinal de excitação correspondente a cada distúrbio localizado, como forma de emular sua minimização pela etapa de Reconstrução. Localizados os distúrbios num dado bloco, pode-se utilizar um procedimento de reconstrução conjunta, como o LSAR (ver Apêndice A), para estimar o sinal não-corrompido nos trechos degradados.

3.2.3 União de Distúrbios Adjacentes

O espalhamento dos distúrbios no sinal de excitação, produzido pela filtragem inversa, pode produzir interferência construtiva ou destrutiva entre amostras de

um mesmo distúrbio ou de distúrbios próximos [13, 5]. No caso de ocorrência de interferência destrutiva, a determinação da duração dos distúrbios através da técnica de limiar fica prejudicada, mesmo se dois limiares forem utilizados. Não é raro o sinal de excitação correspondente a um distúrbio relativamente duradouro exibir amostras intermediárias com módulos menores que o valor do limiar de localização, como mostra a Fig. 3.2 (b) — observar a magnitude do sinal de excitação em torno da posição $k = 300$. Isso acaba por levar a uma determinação incorreta da duração do distúrbio ou, o que é pior, ao invés de apenas 1 distúrbio, à detecção de dois ou mais distúrbios, separados por amostras consideradas não-corrompidas, mas que, de fato, são corrompidas.

Um critério que se mostra bastante eficiente na solução desses casos é promover uma união forçada de distúrbios adjacentes separados por até n amostras. Uma vez arbitrado n , um procedimento possível consiste em incorporar, ao distúrbio detectado, amostras adjacentes à de maior módulo, até que haja, em ambos os sentidos, pelo menos, n amostras consecutivas da excitação com módulos menores que o limiar de localização. Em outras palavras, todo distúrbio localizado dessa forma é ladeado, garantidamente, por, pelo menos, n amostras da excitação com módulos menores que o limiar.

Em geral, unir distúrbios separados por 2 ou 3 amostras¹ parece bastante razoável; entretanto, para distúrbios de maior duração (mais de 30 amostras), esse número pode ser insuficiente. Uma opção é fazê-lo dependente da duração l que se determinaria para o distúrbio, sem qualquer critério de união. O valor variável de n pode ser ditado, por exemplo, por

$$n = n_{min} + \left\lfloor \frac{l}{g} \right\rfloor, \quad (3.3)$$

onde n_{min} é o menor valor inteiro arbitrariamente admitido para n , g é um fator de controle do aumento de n em relação ao aumento de l e $\lfloor \cdot \rfloor$ é o operador “maior inteiro menor que ou igual a”.

A utilização, na etapa de Detecção, do critério de união de distúrbios adjacentes, apresentada nessa Seção, é de fundamental importância para a obtenção de restaurações qualitativamente superiores (ver Seção 3.3).

¹Considerando uma taxa de amostragem de 44100 Hz

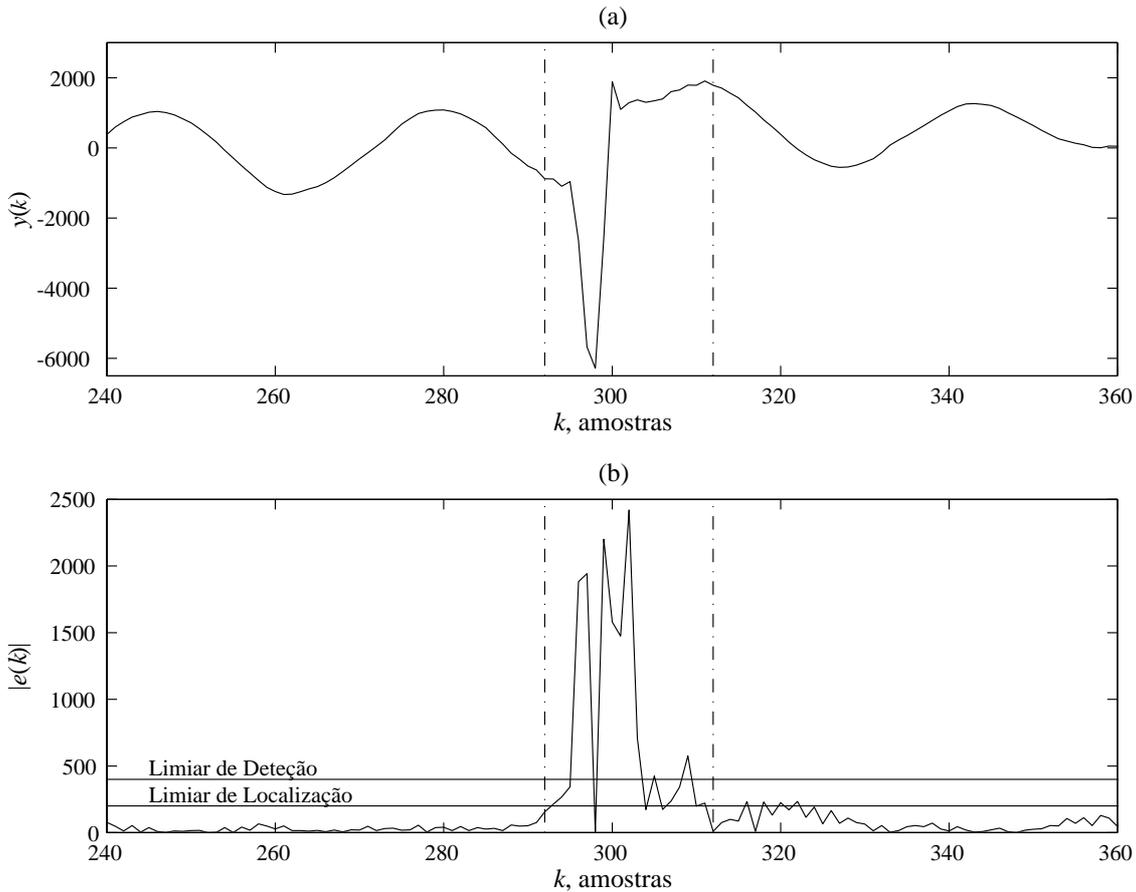


Figura 3.2: (a) Sinal naturalmente corrompido – distúrbio impulsivo delimitado, por inspeção visual, pelas linhas mistas verticais. (b) Módulo do sinal de excitação e limiares de detecção e localização – o distúrbio localizado usando união de distúrbios adjacentes ($n = 2$) é delimitado pelas linhas mistas verticais. Observe que, sem a união de distúrbios adjacentes, seriam localizados 4 distúrbios distintos.

3.2.4 Reiteração de Processamento num mesmo Bloco

Uma vez que as técnicas de detecção descritas nas seções anteriores requerem uma implementação em blocos, é plausível pensar em reiterar, num mesmo bloco, as etapas de Detecção e Reconstrução, visto que uma rodada apenas pode não ser suficiente para detectar todos os distúrbios impulsivos presentes, principalmente os de menor amplitude. Sabendo que a etapa de Reconstrução tende a minimizar os trechos do sinal de excitação correspondentes aos distúrbios impulsivos suprimidos, é de se esperar que trechos do sinal já recuperados não sejam mais detectados numa nova etapa de detecção (utilizando mesmos parâmetros de processamento). É, então, inteiramente justificável adotar a não-ocorrência de amostras do módulo do sinal de

excitação acima do limiar de detecção como critério de interrupção das iterações num mesmo bloco. Entretanto, não há qualquer garantia de que tal critério (Critério 1) será sempre satisfeito para todos os blocos a serem processados.

De fato, quando alguns distúrbios, em geral de pequena amplitude são detectados com duração muito subestimada, a etapa de reconstrução não produz a desejada minimização do sinal de excitação no trecho correspondente, e, por essa razão, tais distúrbios acabam sendo detetados indefinidamente. Uma forma de contornar esse problema é fazer com que as iterações sejam interrompidas quando o Critério 1 for satisfeito **ou** quando o número de iterações atingir uma contagem arbitrária i_{max} . Esse critério misto será chamado de Critério 2.

Observações experimentais mostraram que, nas primeiras iterações realizadas em um bloco, os distúrbios de maiores duração e amplitude presentes são suprimidos satisfatoriamente, restando por serem suprimidos os de menores amplitudes. Esses distúrbios têm maiores chances de serem localizados com durações subestimadas e de produzirem rodadas inócuas do processamento. Isso pode ser evitado promovendo uma localização mais apurada desses distúrbios, que passa, necessariamente, pela redução do valor do limiar de localização. Essa redução pode ser realizada diminuindo-se a razão b entre os limiares de localização e detecção conforme o número de iterações i cresce. Uma forma de implementar essa redução consiste em fazer

$$b[i] = r^{\lfloor \frac{i}{f} \rfloor} \cdot b[i - 1], \quad i = 2 \text{ a } i_{max}, \quad (3.4)$$

onde r é o fator de redução e f um parâmetro que controla a frequência com que b é alterado. A escolha dos parâmetros i_{max} , f e r é arbitrária e está relacionada a observações experimentais.

A utilização de b dependente da iteração realizada, além de proporcionar a supressão de distúrbios de menor amplitude, também evita repetições inócuas de processamento, contribuindo, dessa forma, para a menores tempos de processamento.

3.3 Comparação dos Resultados

A avaliação comparativa dos desempenhos dos métodos de Detecção de distúrbios impulsivos baseados em limiar, com e sem as modificações propostas, foi realizada em um conjunto de sinais de áudio contaminados artificialmente por ruído impulsivo. Duas classes de ruído impulsivo foram obtidas da diferença entre dois sinais de áudio naturalmente corrompidos por ruído impulsivo e suas versões já restauradas. O ruído Classe 1 contém distúrbios impulsivos com amplitudes e durações muito variadas e, quando somado a um sinal de áudio, chega a corromper aproximadamente 0.6% de suas amostras. O ruído Classe 2 contém distúrbios de menores duração e amplitude, entretanto, quando somado a um sinal de áudio, é capaz de corromper aproximadamente 5% de suas amostras. As amplitudes dos sinais de ruído também foram ajustadas de acordo com as dos sinais de áudio, de modo a tornar o ruído claramente audível. O método de Reconstrução utilizado em todos os casos foi o LSAR (ver Apêndice A).

Para avaliar a eficácia da utilização de dois limiares e do critério de união de distúrbios adjacentes, duas versões da etapa de Detecção foram implementadas. A primeira, denominada de ED1, utiliza os parâmetros e critérios descritos a seguir:

- Processamento realizado em blocos de 1024 amostras;
- Ordem do modelo AR igual a 40 e parâmetros calculados por minimização LS;
- Limiar de detecção calculado por (3.1) com $K = 5$;
- União de distúrbios adjacentes separados por n amostras segundo (3.3), com $n_{min} = 2$ e $g = 6$;
- Utilização do Critério 2 para a interrupção das iterações num dado bloco de sinal com $i_{max} = 7$;
- Utilização de redução progressiva do valor do limiar de localização, segundo (3.4), com $b[1] = r = 0.5$ e $f = 3$;

A segunda, denominada de ED2, só difere de ED1 pela utilização de limiar único e pela não-utilização do critério de união de distúrbios adjacentes. O limiar único foi

obtido estabelecendo $b = 1$ em (3.2) e $r = 1$ em (3.4), o que é equivalente a fazer $\lambda_L = \lambda_D$.

A avaliação qualitativa dos sinais de áudio restaurados é, em última análise, uma questão subjetiva associada aos resultados perceptivos produzidos pelo processamento. Neste trabalho, os sinais originalmente não-corrompidos estão disponíveis para permitir o uso de medidas quantitativas que possam corroborar os resultados perceptivos obtidos. Medidas como as taxas de não-deteção e de falsa deteção são simples e efetivas para esse propósito.

Sejam $n_o(k)$ e $n_p(k)$, respectivamente, o ruído que corrompe os sinais de áudio e o ruído residual após o processamento — obtido pela diferença entre as versões processada e não-ruidosa do sinal de áudio. As percentagens de não-deteção são calculadas como a razão entre o número de amostras não-nulas em $n_o(k)$ que assim permanecem em $n_p(k)$ e o número total de amostras não-nulas em $n_o(k)$. As percentagens de falsa deteção são calculadas como a razão entre o número de amostras nulas em $n_o(k)$ que se tornam não-nulas em $n_p(k)$ e o número total de amostras nulas em $n_o(k)$.

Definindo a função t_1 tal que

$$t_1(\alpha) = \begin{cases} 0, & \text{se } \alpha = 0 \\ 1, & \text{em caso contrário} \end{cases}$$

e a função t_2 tal que

$$t_2(\alpha) = \begin{cases} 0, & \text{se } \alpha \leq 0 \\ 1, & \text{em caso contrário} \end{cases},$$

o percentual de não-deteção (PND) é dado por

$$\text{PND} = \frac{\sum_{k=1}^N t_2(t_1(n_o(k)) - t_1(n_p(k) - n_o(k)))}{\sum_{k=1}^N t_1(n_o(k))} \quad (3.5)$$

e o percentual de falsa deteção (PFD) é dado por

$$\text{PFD} = \frac{\sum_{k=1}^N t_2(t_1(n_p(k) - n_o(k)) - t_1(n_o(k)))}{N - \sum_{k=1}^N t_1(n_o(k))}, \quad (3.6)$$

onde N em (3.5) e (3.6) é o número total de amostras do sinal de áudio.

A Tabela 3.1 mostra os percentuais de falsas deteções e não-deteções obtidos utilizando-se as versões ED1 e ED2 da etapa de Deteção. Os Sinais 1 a 4 foram

corrompidos pelo sinal de ruído da Classe 1, enquanto que os Sinais 5 e 6 foram corrompidos pelo sinal de ruído da Classe 2. É importante enfatizar que tanto os parâmetros usados em ED1 como os usados em ED2 foram mantidos fixos para processar os seis sinais.

Tabela 3.1: Medidas Comparativas dos percentuais de falsa detecção e não-deteção relativos às versões restauradas usando ED1 e ED2.

	Não-Deteção (%)		Falsa Deteção (%)	
	ED1	ED2	ED1	ED2
Sinal 1	1.5595	16.4133	3.2641	1.1650
Sinal 2	2.0952	20.3880	1.5609	0.6784
Sinal 3	1.5638	10.7057	1.8313	0.8733
Sinal 4	4.2105	21.5205	2.4005	0.9307
Sinal 5	1.6270	14.8647	6.6537	3.5418
Sinal 6	1.8487	12.1126	8.0029	4.4480

Todos os sinais restaurados usando ED1 apresentaram resultados perceptivamente superiores em relação aos restaurados usando ED2. Os resultados de ED1 ainda apresentaram uma qualidade perceptiva mais homogênea, a despeito da diversidade dos sinais de áudio avaliados, das duas classes distintas de ruídos corruptores e da manutenção dos mesmos parâmetros para processar todos os sinais. Isso indica que as modificações propostas aumentam a robustez da etapa de Deteção quanto a variações nas características dos sinais de áudio e de ruído.

As medidas de PND associadas a ED1 indicam uma redução por um fator entre 5 e 10 no índice de não-deteção em relação às mesmas medidas associadas a ED2. Já as medidas de PFD associadas a ED1 indicam um aumento por um fator entre 2 e 3 no índice de falsa detecção em relação às mesmas medidas associadas a ED2. A desejável redução no índice de não-deteção é, de fato, responsável pelos resultados perceptivamente superiores dos sinais restaurados por ED1. O aumento do índice de falsa detecção, idealmente indesejável, era esperado, visto que as modificações propostas tendem a superestimar, ligeiramente, a duração de alguns distúrbios. Os resultados perceptivos, entretanto, mostram que tais durações superestimadas não

afetam de modo significativo a qualidade do sinal de áudio; ao contrário, são capazes de evitar restaurações insatisfatórias, como a mostrada na Fig. 3.3 (c).

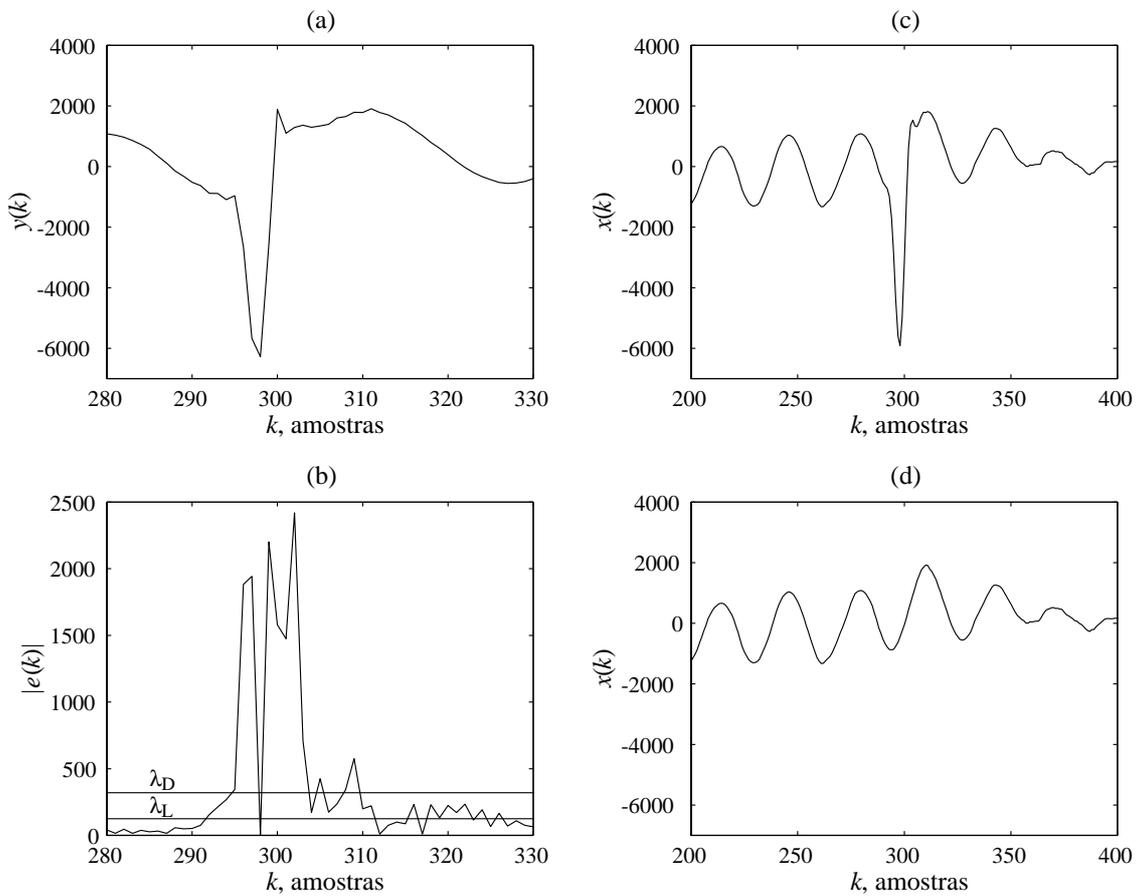


Figura 3.3: (a) Detalhe do distúrbio impulsivo corrompendo o sinal de áudio. (b) Excitação do sinal mostrado em (a) e limiares de detecção, λ_D , e localização, λ_L . (c) Sinal restaurado usando ED2. (d) Sinal restaurado usando ED1.

Apesar de os resultados apresentados nesta seção serem relativos a sinais de áudio artificialmente corrompidos, as modificações propostas para a etapa de Detecção produzem resultados igualmente satisfatórios na restauração de sinais naturalmente corrompidos por ruído impulsivo.

Capítulo 4

Supressão de Ruído Impulsivo em Multitaxas

4.1 Introdução

Sistemas multitaxas têm sido largamente utilizados em processamento de sinais de áudio com os objetivos de compressão e redução de ruído de fundo [17, 18, 10, 19, 11, 20]. A divisão do sinal de áudio em faixas de frequência e seu subsequente processamento pode exibir vantagens, em relação ao processamento realizado em banda completa, devido às propriedades psicoacústicas do sistema auditivo humano (ver Seção 1.4) e às propriedades intrínsecas dos sinais musicais de áudio, principalmente no que tange à concentração de energia em baixas frequências. Essas propriedades justificam, de certa forma, a análise multirresolução ou em sub-bandas do sinal de áudio e até o uso de processamentos diferenciados em cada sub-banda.

A finalidade deste Capítulo é investigar as possibilidades de adaptação da técnica de detecção via limiar e filtragem inversa para um sistema multitaxas. A Seção 4.2 aborda questões relacionadas à adaptação desejada, como a estacionariedade dos sinais em sub-bandas e a possibilidade de utilização de modelo AR para esses sinais. Na Seção 4.3, são descritos aspectos da implementação do sistema, da utilização de diferentes tipos de algoritmos de reconstrução conforme o conteúdo espectral do sinal de áudio e do tempo de processamento. Alguns conceitos básicos de sistemas multitaxas utilizados neste Capítulo são descritos no Apêndice B.

4.2 Sinais de Áudio em Multitaxas

A adaptação da técnica baseada em limiar apresentada no Capítulo 3 nos sinais em sub-bandas é, idealmente, direta, se se dispuser de uma caracterização estatística adequada para esses sinais. Considerando, isoladamente, cada sinal em sub-banda, resultantes da decomposição multiresolução de um sinal de áudio, o processamento de detecção/supressão de ruído impulsivo seria idêntico ao realizado em banda completa, ou seja: obter um modelo estatístico adequado para um bloco do sinal em questão; realizar, se possível, uma filtragem inversa de modo a obter o sinal de excitação correspondente ao bloco; promover a detecção dos distúrbios impulsivos pela comparação do módulo do sinal de excitação com um limiar adequadamente calculado; e reconstruir os trechos do sinal de áudio considerados corrompidos por um algoritmo de interpolação apropriado. A síntese dos sinais em sub-bandas processados resultaria no sinal de áudio restaurado em banda completa.

Na realidade, os sinais em sub-bandas, obtidos pela passagem do sinal de áudio por um banco de análise, sofrem alterações relacionadas ao tempo de estacionaridade e à modelagem estatística, que estão diretamente associadas às características do banco de filtros de análise utilizado [21].

A configuração de um sistema para processamento em sub-bandas envolve a escolha, nem sempre evidente, do número de sub-bandas do banco de filtros, da largura de banda de cada sub-banda, do tipo de banco de filtros (Ortogonal, Bior-togonal, QMF, ...) e das ordens dos filtros que compõem o banco. Essas escolhas devem estabelecer, na realidade, um compromisso entre as premissas estabelecidas pela Psicoacústica, visando melhores resultados perceptivos, e aspectos computacionais, visando maior rapidez de processamento. As subseções seguintes tecem algumas considerações sobre essas questões.

4.2.1 Número de Sub-bandas e sua Largura de Banda

A determinação do número de sub-bandas e de sua largura de banda para um banco de filtros que seja mais apropriado ao sistema auditivo humano deveria levar em consideração propriedades da Psicoacústica, como a Audibilidade e as Bandas Críticas [8, 9], a finalidade do processamento a ser realizado, características dos sinais

envolvidos, dentre outros fatores. Só as propriedades de Audibilidade já impõem limites, por exemplo, à adoção de uma taxa de amostragem para sinais de áudio que garanta fidelidade na região considerada audível de seu espectro. Por outro lado, é importante ressaltar que os conceitos e formulações associadas às Bandas Críticas são restritos a sinais bastante simples, podendo não ser muito adequados a sinais reais de áudio.

O modelamento do sistema auditivo humano não faz parte do escopo deste trabalho, entretanto, é intuitivamente plausível que haja uma forte correlação entre a construção das escalas musicais e as características de percepção do sistema auditivo humano.

Na maioria das escalas musicais, dois tons de freqüências cuja razão é 2 são ditos separados por uma oitava e classificados como a mesma nota musical. Esse tipo de estrutura é bastante indicativa de que a resolução do sistema auditivo humano, no que tange à percepção de variações freqüenciais, não é igual em toda a faixa audível do espectro de freqüências. Isso significa, por exemplo, que uma variação de 110 Hz numa dada freqüência F é percebida de formas diferentes, conforme a localização de F no espectro audível de freqüências. De fato, considerando a escala temperada — em que uma oitava é sub-dividida regularmente¹ em 12 semi-tons — os tons puros com freqüências 110 Hz e 220 Hz são percebidos como separados por uma oitava ou 12 semi-tons; já dois tons puros com freqüências entre 1760 Hz e 1870 Hz por exemplo, são percebidos como separados por apenas um semi-tom, a despeito da mesma variação de 110 Hz entre as freqüências. Pelo o exposto, é intuitivo que a resolução freqüencial do ouvido humano seja gradativamente menor em direção às altas freqüências e, nesses termos, parece interessante promover uma divisão no espectro condizente com esse comportamento.

Uma forma possível de promover uma divisão espectral com resolução freqüencial crescente em direção às baixas freqüências é conseguida dividindo, idealmente, ao meio o espectro de freqüências e iterativamente, dividindo ao meio as metades inferiores resultantes da divisão anterior. Por razões práticas, não se pode iterar indefinidamente; a adoção de um número finito de iterações implica mesma largura para as duas bandas inferiores. A multiresolução tempo-freqüência é conseguida

¹A razão entre as freqüências de dois semi-tons adjacentes é $\sqrt[12]{2}$.

reduzindo a taxa de amostragem proporcionalmente à largura de cada banda (amostragem crítica). Bancos de CQF ou QMF (em oitavas) ou análise freqüencial usando *Wavelets* são apropriados para esse fim. Mesmo não entrando no mérito das características dos filtros que compõem o banco de análise, a divisão em 11 sub-bandas já seria suficiente para contemplar a faixa de freqüências percebida pelo sistema auditivo humano, situada entre 20 Hz e 20000 Hz.

Em se tratando de um sistema visando à supressão de ruído impulsivo, que objetiva, primordialmente, evidenciar a ocorrência de distúrbios impulsivos, talvez não seja necessária a utilização de 11 sub-bandas, como citado anteriormente. Apesar de a característica impulsiva do ruído lhe conferir um espectro largo, as propriedades de Audibilidade e Mascaramento fazem com que sua percepção seja mais evidente em médias e altas freqüências. De fato, a concentração de informação em baixas e médias freqüências associada a sinais naturais de áudio mascara as componentes espectrais do ruído nessas freqüências, o que não ocorre, em geral, nas altas freqüências. Desta forma, parece desnecessário estender a subdivisão em oitavas na direção das baixas freqüências até a obtenção de 11 sub-bandas, bastando, talvez, subdividir o sinal em até 5 ou 6 sub-bandas.

As Figs. 4.1 e 4.2 mostram exemplos de sinais em sub-bandas correspondentes à decomposição, em oitavas, por um banco de análise de QMF's, de um trecho de sinal de áudio contaminado por ruído impulsivo, respectivamente, em 4 e 8 sub-bandas.

É interessante notar que os distúrbios impulsivos ficam bastante evidenciados nas sub-bandas agudas², favorecendo, por essa razão, sua detecção nessas sub-bandas, talvez simplificarmente, sem necessidade de realizar filtragem inversa para detectá-los no domínio da excitação. Já nas sub-bandas graves, os distúrbios impulsivos parecem estar bem menos evidenciados, como também devem ser seus efeitos perceptivos, podendo até ser desnecessário realizar qualquer processamento nessas sub-bandas.

Os coeficientes do filtro protótipo, utilizado para construir os bancos de QMF's, foram obtidos em [22] (pp. 402, tabela 7.2, conj. 16A). A magnitude

²As denominações sub-banda aguda e sub-banda grave numa divisão em oitavas se referem, respectivamente, à sub-banda de menor fator de decimação e à de maior fator de decimação.

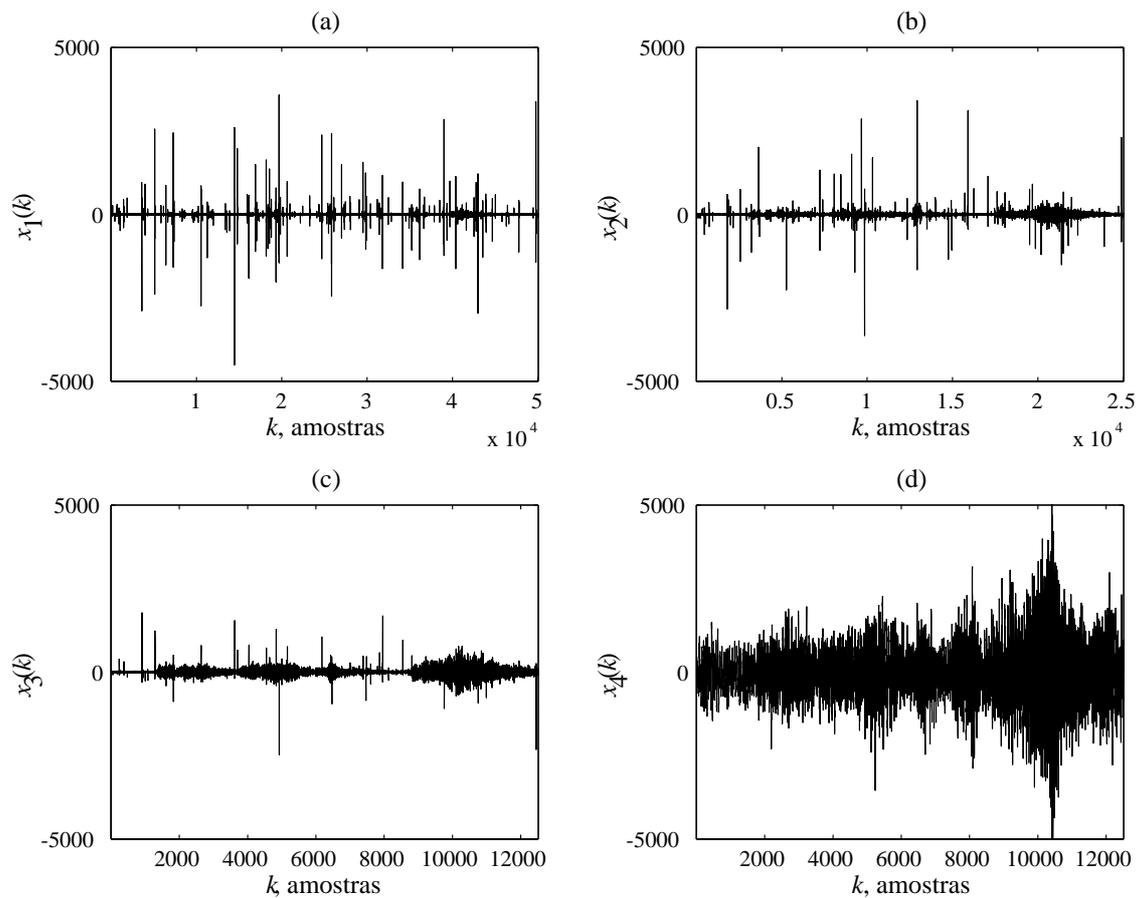


Figura 4.1: Análise em 4 sub-bandas de sinal de áudio contaminado por ruído impulsivo: (a), (b), (c) e (d) mostram, respectivamente, os sinais nas sub-bandas, da mais aguda para a mais grave.

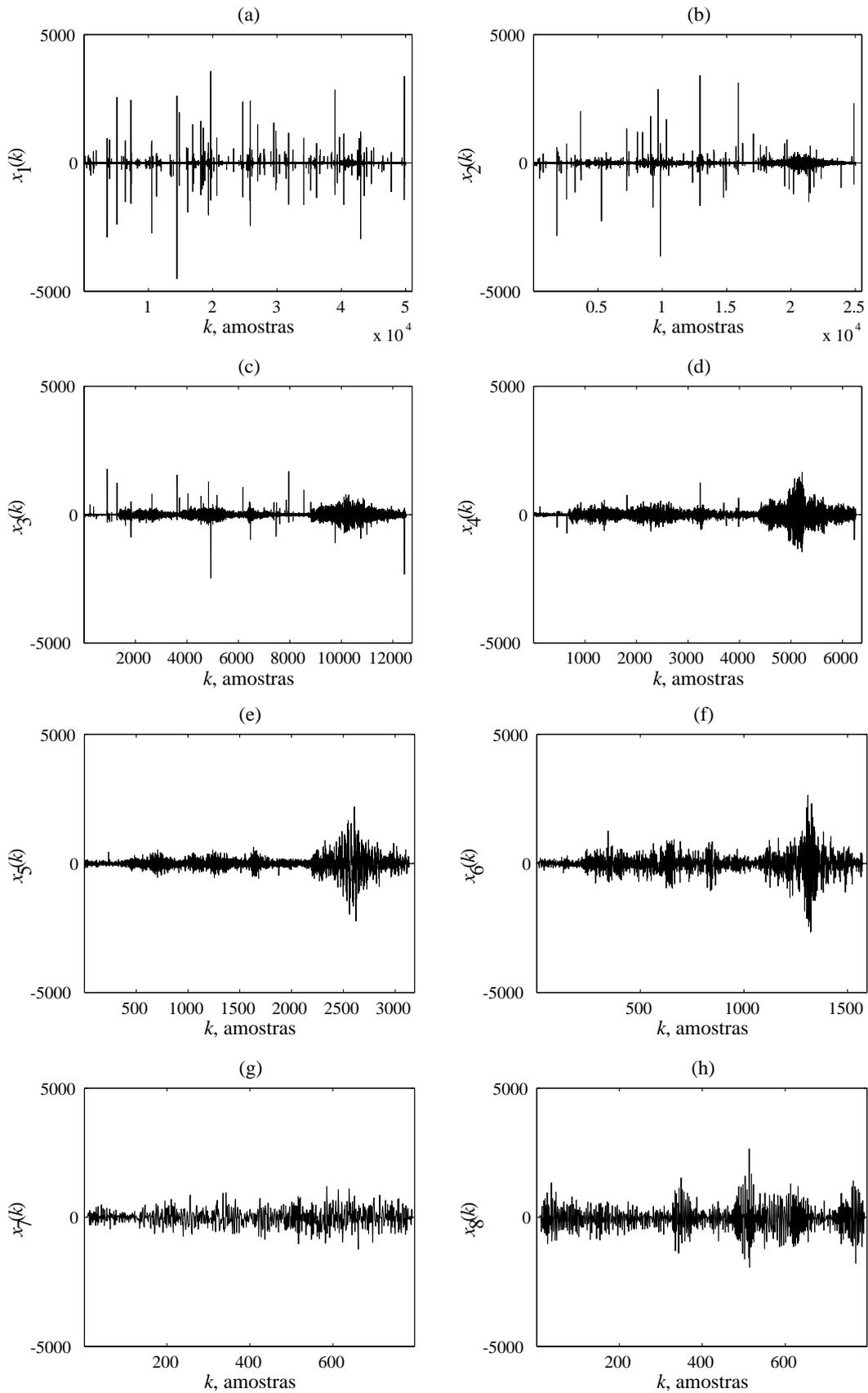


Figura 4.2: Análise em 8 sub-bandas de sinal corrompido por ruído impulsivo: de (a) a (h) são mostrados, respectivamente, os sinais nas sub-bandas, da mais aguda para a mais grave.

da resposta em frequência do protótipo e a primeira metade de seus coeficientes são mostradas na Fig. 4.3.

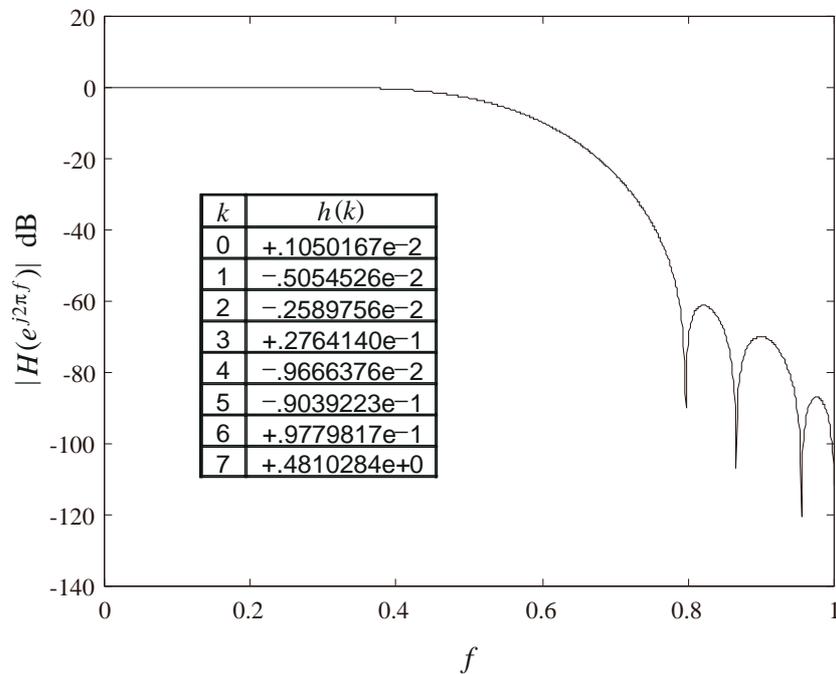


Figura 4.3: Magnitude (em dB) da resposta em frequência e primeira metade dos coeficientes do filtro protótipo do banco de QMF's.

A superposição das magnitudes das respostas em frequência dos filtros resultantes de um banco de análise de 4 sub-bandas é mostrada na Fig. 4.4. Na Fig. 4.5, tem-se o mesmo para 8 sub-bandas.

4.2.2 Tempo de Estacionaridade dos Sinais em Sub-bandas

A Seção 1.3 abordou questões e considerações acerca da estacionaridade dos sinais de áudio em banda completa. Nesta Seção, é considerada a estacionaridade dos sinais em sub-bandas, resultantes de análise em multirresolução de sinal de áudio.

Inicialmente, a primeira hipótese que pode ser assumida como válida é a de que o tempo de estacionaridade independe da banda do sinal e de sua localização dentro do espectro de áudio. Nesses termos, se um sinal estacionário fosse decomposto em sub-bandas, os sinais em sub-bandas manteriam o mesmo tempo de estacionaridade do sinal em banda completa e, desta forma, o número de amostras

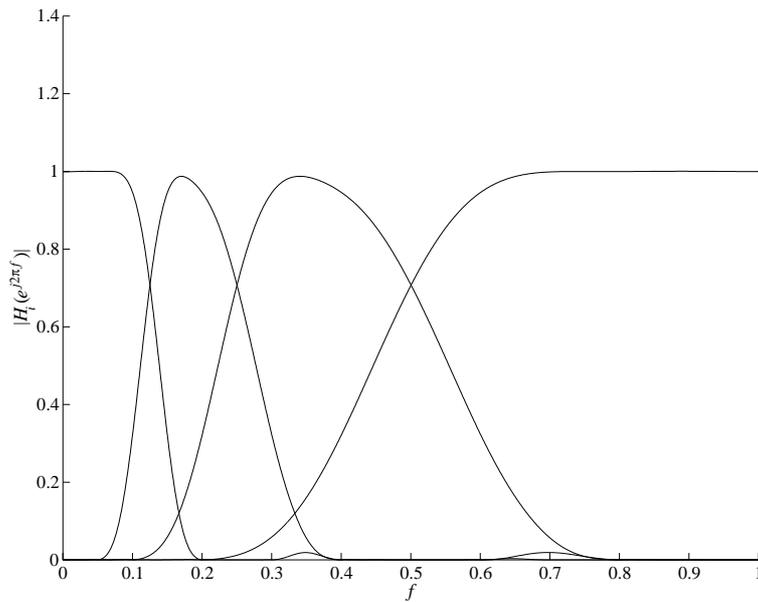


Figura 4.4: Superposição das magnitudes das respostas em frequência dos filtros equivalentes para uma decomposição em 4 sub-bandas em oitavas, utilizando banco de análise QMF com o protótipo da Fig. 4.3.

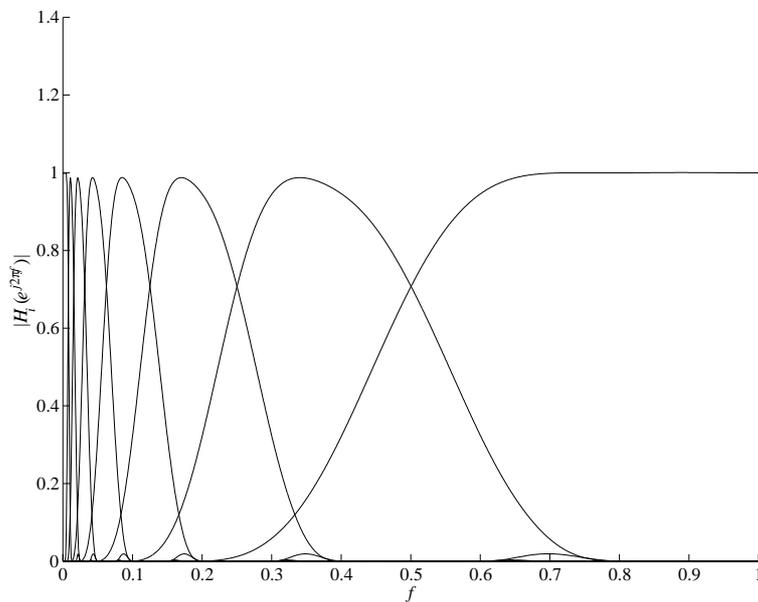


Figura 4.5: Superposição das magnitudes das respostas em frequência dos filtros equivalentes para uma decomposição em 8 sub-bandas em oitavas, utilizando banco de análise QMF com o protótipo da Fig. 4.3.

de um bloco estacionário em sub-bandas sofreria redução diretamente proporcional à redução da taxa de amostragem utilizada.

A aceitação da hipótese anterior tem implicações interessantes. Suponha um bloco de sinal de áudio de 2048 amostras, considerado estacionário em banda completa; uma decomposição em oitavas por um banco de filtros de análise com 11 sub-bandas faria com que o sinal correspondente à sub-banda de frequências mais baixas tivesse apenas uma amostra. O número reduzido de amostras dos sinais, supostamente estacionários, presentes nas sub-bandas de baixas frequências, pode tornar proibitiva, por insuficiência de dados, a realização de algum processamento que necessite de modelagem estatística e, de certa forma, acaba impondo limites à decomposição frequencial do sinal em um número grande de sub-bandas — o que, a princípio, não constituiria problema para o processamento visando à supressão de ruído impulsivo, como visto na seção 4.2.1.

A consideração de um tempo de estacionaridade para sinais de áudio está fortemente associada à possibilidade de representar o sinal por um modelo estatístico, como, por exemplo, o AR, de tal modo que seus coeficientes permaneçam inalterados, por um determinado período de tempo. Seguindo essa linha de raciocínio e adicionando o fato de os registros mais graves, em um sinal de áudio, possuírem, em geral, uma tendência de variação bem menor que a dos registros agudos, é plausível supor que um modelo estatístico que represente registros graves possa ser válido, sem necessidade de alteração, por um período de tempo maior do que o período no qual se possa considerar a invariabilidade de um modelo que represente registros agudos.

A utilização do violino e do contrabaixo numa composição musical é um exemplo que corrobora as diferentes tendências de variação dos registros grave e agudo. O contrabaixo é, em geral, utilizado como base harmônica ou rítmica na estrutura musical, devido a seu registro grave. O violino, por sua vez, é geralmente escalado para enunciar linhas melódicas ou executar ornamentações de andamento ligeiro, funções adequadamente talhadas a seu registro agudo. Os maiores tempos de ataque e de decaimento associados à execução das notas graves no contrabaixo, em relação aos do violino nas notas agudas, são um dos principais motivos da ineficácia de uma inversão de papéis entre os dois instrumentos — a menos que se deseje obter

algum efeito sonoro específico —, fato que, provavelmente, causaria grande perda de clareza das linhas melódicas e das ornamentações.

Das considerações anteriores, é intuitivo supor que o tempo de estacionariedade varie de acordo com a faixa de frequências efetivamente presente nos sinais em sub-bandas, sendo grande para sinais com conteúdo espectral limitado a uma faixa de baixas frequências e pequeno para sinais com conteúdo espectral limitado a altas frequências.

4.2.3 Modelos Estatísticos para Sinais de Áudio em Sub-bandas

A aplicação direta da técnica de supressão de ruído impulsivo descrita no Capítulo 3 nos sinais em sub-bandas requer a adoção de modelos estatísticos que representem de modo satisfatório o comportamento estatístico desses sinais. Para os sinais de áudio em banda completa, a adoção do modelo AR é, em geral, muito adequada, como mostra a Fig. 4.6, sendo a eficiência da modelagem AR medida, em parte, pela capacidade do modelo de descorrelacionar o sinal de excitação correspondente ao sinal de áudio.

A atribuição de um modelo estatístico a sinais em sub-bandas provenientes de decomposição de um sinal áudio por um banco de filtros não é trivial. Uma investigação dos efeitos estatísticos da filtragem FIR seguida da operação de decimação, que consiste em um módulo básico da estrutura do banco de filtros de análise, se faz necessária.

Inicialmente, convém supor filtros passa-baixas e passa-altas ideais no banco de filtros de análise para tecer algumas considerações sobre os efeitos da filtragem FIR seguida de decimação na localização dos pólos e zeros de um possível modelo estatístico para os sinais em sub-bandas.

Os pólos de um processo AR associados a um trecho de sinal de áudio em banda completa se encontram, em geral, bastante próximos da circunferência unitária, como mostra a Fig. 4.7. A densidade espectral de potência desse sinal deve exibir picos correspondentes aos pólos presentes na proximidade da circunferência unitária e uma filtragem passa-baixas ideal teria o efeito de só preservar a faixa da densidade espectral de potência correspondente à banda de passagem do filtro e, portanto,

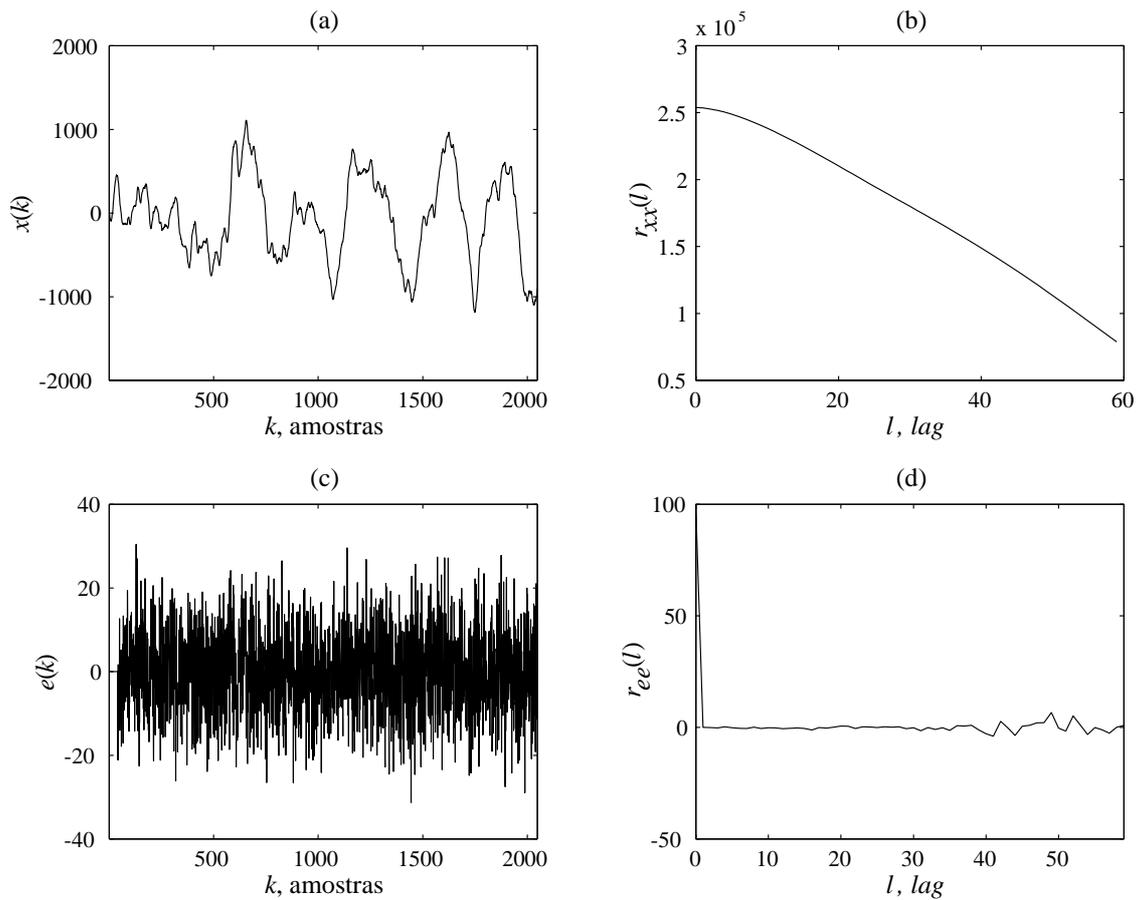


Figura 4.6: (a) Trecho de 2048 amostras de um sinal de áudio. (b) Função de autocorrelação do sinal de áudio. (c) Excitação obtida por filtragem inversa (AR, ordem 40) do sinal de áudio. (d) Função de autocorrelação da excitação.

exibindo um número menor de picos. A densidade espectral de potência da versão passa-baixas do sinal, decimada criticamente, vai corresponder exatamente à faixa da densidade anterior preservada pela banda de passagem do filtro.

Por esta primeira aproximação, pode-se intuir que a filtragem passa-baixas anula os efeitos dos pólos situados fora da banda de passagem do filtro e que o número de pólos do sinal em sub-banda resultante da filtragem e da decimação crítica é menor que o número de pólos do sinal em banda completa. Esse raciocínio é igualmente válido para o caso de filtros passa-banda, sendo a redução no número de pólos do sinal em sub-bandas dependente da banda de passagem do filtro.

Uma possibilidade a ser investigada é a utilização de modelagem AR de baixa ordem para os sinais em sub-bandas. Nesse caso, a ordem do modelo a ser atribuído aos sinais de cada sub-banda estaria diretamente ligada aos pólos do modelo em banda completa que permanecessem na sub-banda após a filtragem. Entretanto, devido ao caráter altamente não-estacionário dos sinais de áudio, é praticamente impossível garantir a existência de um determinado número pólos por sub-banda; por essa razão, se torna necessário considerar, para cada bloco estacionário, uma distribuição de pólos aproximadamente constante ao longo da circunferência unitária.

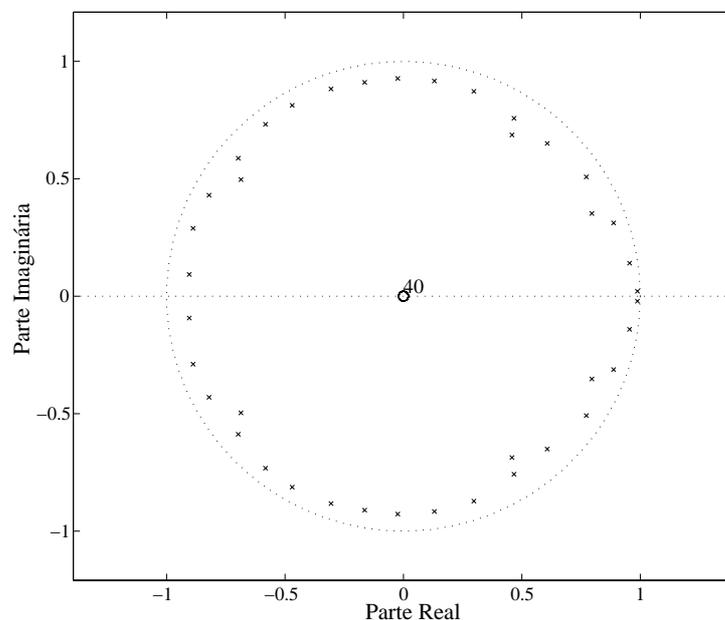


Figura 4.7: Diagrama de pólos e zeros do modelo AR de ordem 40 estimado para 2048 amostras de sinal de áudio em banda completa.

A distribuição de pólos observada na Fig. 4.7 está razoavelmente de acordo

com a consideração anteriormente feita. Neste caso, para uma divisão em oitavas com 6 sub-bandas, considerando o sinal de entrada como AR de ordem 40, uma possível escolha para as ordens dos modelos AR para os sinais em sub-bandas seria 20, 10, 5, 3, 2 e 2, respectivamente, do de maior para o de menor banda. É interessante notar que uma divisão em 11 sub-bandas poderia ser inócua, visto que as sub-bandas mais graves poderiam não conter, na maior parte do tempo, nenhum pólo.

Na realidade, não se pode contar com filtros ideais nos bancos de filtros e torna-se necessária uma análise mais cuidadosa sobre os efeitos estatísticos produzidos nos sinais em sub-bandas pelo procedimento de filtragem seguida de decimação, antes de se concluir pela utilização de modelos AR de baixa ordem para os sinais em sub-bandas. Inicialmente, será analisada a função de transferência entre um sinal de uma sub-banda genérica e a entrada em banda completa, através de suas componentes polifásicas. Em seguida, o uso de modelos AR de baixa ordem para os sinais em sub-bandas será avaliado pelas funções de autocorrelação desses sinais.

A função de transferência entre o sinal em banda completa $x(k)$ e o sinal em qualquer sub-banda i , $x_i(k)$, pode ser modelada, genericamente, pela célula básica da estrutura de um banco de filtros de análise: um filtro FIR $H_i(z)$ seguido por uma operação de decimação por M , como mostra a Fig. 4.8. O fator de decimação pode ser diferente para cada sub-banda e, a rigor, sua notação deveria ser M_i , mas, para não carregar desnecessariamente a notação, utilizar-se-á simplesmente M .



Figura 4.8: Transferência geral entre o sinal em banda completa e o sinal em uma sub-banda, relativa ao banco de análise.

Considerando a entrada $x(k)$ um bloco de sinal modelável por um processo AR de ordem p , como em (1.1), seu espectro é dado por (4.2), onde $E(z)$ é o espectro do sinal de excitação. O espectro do sinal de saída $X_i(z)$, resultante do processo de filtragem e decimação, é dado por (4.3), onde $W_M^k = e^{-j\frac{2\pi k}{M}}$. A interpretação de (4.3) é difícil; entretanto, sua forma especial de somatório suscita uma análise utilizando componentes polifásicas, que é descrita a seguir.

$$A(z) = 1 - \sum_{i=1}^p a(i)z^{-i} \quad (4.1)$$

$$X(z) = \frac{E(z)}{A(z)} \quad (4.2)$$

$$X_i(z) = \frac{1}{M} \sum_{k=0}^{M-1} (X(z^{\frac{1}{M}} W_M^k) H_i(z^{\frac{1}{M}} W_M^k)) \quad (4.3)$$

Como se deseja saber sobre possíveis modelos para os sinais em sub-bandas, uma possibilidade interessante consiste em incorporar o modelo AR de um trecho do sinal de entrada à função de transferência entre sinal em sub-banda e sinal em banda completa, como mostra a Fig. 4.9, para, então, proceder ao cálculo das componentes polifásicas de cada sub-banda.

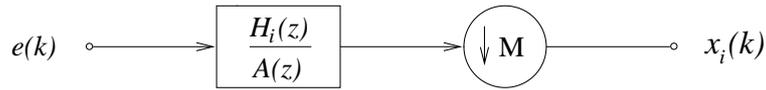


Figura 4.9: Transferência geral entre o sinal em uma sub-banda e o sinal de excitação na entrada do banco de análise.

Qualquer função de transferência FIR, $G(z)$, admite a representação polifásica [23, 24]

$$G(z) = \sum_{j=0}^{M-1} z^{-j} G_j(z^M), \quad (4.4)$$

onde

$$G_j(z) = \sum_{l=-\infty}^{\infty} g(Ml + j)z^{-l}, \quad j = 0 \text{ a } M - 1 \quad (4.5)$$

são as componentes polifásicas de $G(z)$.

Como na equivalência de (4.4) as componentes polifásicas são escritas em termos de potências de z múltiplas de M , pode-se aplicar a identidade nobre de comutação entre filtragem e decimação (ver Apêndice B) e obter a estrutura polifásica mostrada na Fig. 4.10.

Quando $G(z)$ é IIR, como no caso $G(z) = \frac{H_i(z)}{A(z)}$, o procedimento de obtenção das componentes polifásicas é idêntico, bastando, inicialmente, decompor $G(z)$ como

uma razão de polinômios, cujo polinômio em z do denominador exiba coeficientes nulos em potências de z não-múltiplas de M e coeficientes eventualmente não-nulos em potências de z múltiplas de M , ou seja,

$$G(z) = \frac{N(z)}{D(z^M)}. \quad (4.6)$$

As componentes polifásicas de $G(z)$ IIR são, então, obtidas por

$$G_j(z) = \frac{\sum_{l=-\infty}^{\infty} n(Ml + j)z^{-l}}{\sum_{q=-\infty}^{\infty} d(Ml)z^{-q}} \quad (4.7)$$

No caso particular em que se procede a uma decomposição em oitavas, o valor de M para cada sub-banda i assume a forma

$$M_i = 2^{\min(i, L-1)}, \quad i = 1 \text{ a } L, \quad (4.8)$$

onde L é o número de sub-bandas e i o índice da sub-banda ($i = 0$ corresponde à sub-banda mais aguda). Como, nesse caso, M só assume valores pares, a obtenção de $G(z)$ como em (4.6) pode ser conseguida pelo procedimento mostrado a seguir.

Sejam $a(k)$ e $h_i(k)$ seqüências contendo, respectivamente, os coeficientes dos polinômios em z do denominador e do numerador da função de transferência $\frac{H_i(z)}{A(z)}$. Examinando o caso com $M = 2$, por exemplo, $\frac{H_i(z)A(-z)}{A(z)A(-z)}$ tem no denominador um polinômio com coeficientes nulos em potências de z não-múltiplas de 2, ou seja, está na forma $\frac{N(z)}{D(z^2)}$. A transformação consistiu em multiplicar o numerador e o denominador de $\frac{H_i(z)}{A(z)}$ por $A(-z)$. No domínio do tempo, o produto $A(z)A(-z)$ consiste em realizar a convolução linear discreta entre $a(k)$ e $(-1)^k a(k)$.

No caso geral com M par, a obtenção dos coeficientes de $A(z)$ na forma $D(z^M)$ pode ser realizada, no domínio do tempo, computando $n = \log_2(M)$ sucessivas convoluções sobre a seqüência $a(k)$. Essas convoluções estão expressas em (4.9), com $*$ denotando “convolução linear discreta”, e devem ser computadas começando com $l = 1$ e $\alpha_0(k) = a(k)$. Ao final da n -ésima convolução, a seqüência $\alpha_n(k)$ contém os coeficientes na forma desejada.

$$\alpha_l(k) = \alpha_{l-1}(k) * (-1)^{\frac{k}{2^{l-1}}} \alpha_{l-1}(k), \quad l = 1 \text{ a } n \quad (4.9)$$

Para manter a equivalência entre as razões de polinômios $\frac{H_i(z)}{A(z)}$ e $\frac{N(z)}{D(z^M)}$, as convoluções realizadas no denominador têm que ser realizadas igualmente no nu-

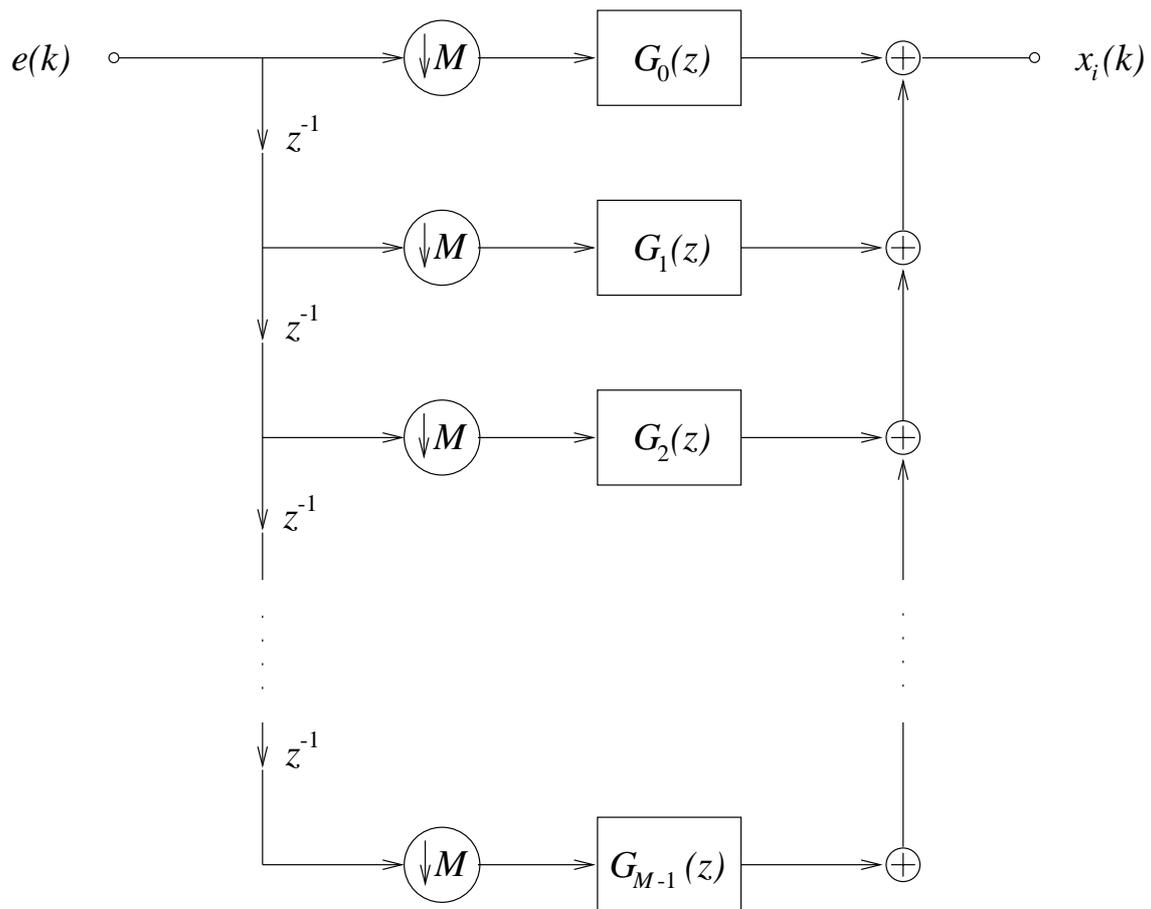


Figura 4.10: Estrutura polifásica equivalente à função de transferência mostrada na Fig. 4.9.

merador. Para obter os coeficientes do numerador $N(z)$ é necessário computar as convoluções regidas por (4.10), desde $l = 1$ a $l = n$, começando com $\eta_0(k) = h_i(k)$ e $\alpha_0(k) = a(k)$. Ao final da n -ésima convolução, $\eta_n(k)$ contém os coeficientes desejados.

$$\eta_l(k) = \eta_{l-1}(k) * (-1)^{\frac{k}{2^{l-1}}} \alpha_{l-1}(k), \quad l = 1 \text{ a } n \quad (4.10)$$

As expressões anteriores permitem calcular as componentes polifásicas de cada sub-banda, com o modelo AR do sinal de entrada incorporado às funções de transferência dos filtros do banco de análise.

Dado os coeficientes do modelo AR da entrada, uma análise dos diagramas de pólos e zeros das componentes polifásicas de uma determinada sub-banda pode ser útil para inferir sobre um possível modelo para o sinal nessa sub-banda.

Dois exemplos de análise em 4 sub-bandas em oitava são apresentados a seguir: um utilizando um banco de análise tipo CQF e com protótipo Haar (Caso Haar) e outro utilizando banco do tipo QMF com o filtro protótipo da Fig. 4.3 (Caso QMF). Em ambos os casos, foram estimados 40 coeficientes AR de um trecho de sinal de áudio em banda completa e, usando esse conjunto de coeficientes, foram calculados os coeficientes de componentes polifásicas de cada sub-banda, através das eqs. (4.7), (4.9) e (4.10).

Os diagramas de pólos e zeros das componentes polifásicas da sub-banda mais grave são mostrados nas Figs. 4.11 e 4.12, respectivamente, para o Caso Haar e o Caso QMF, supracitados. É interessante observar alguns aspectos:

- Os denominadores das componentes polifásicas de uma dada sub-banda são idênticos e não dependem do banco de análise utilizado, mas sim dos coeficientes AR do sinal de entrada e do valor de M ;
- Em uma mesma sub-banda, a localização dos zeros é muito similar de uma componente polifásica para outra;
- Nos dois casos, a maioria dos pólos parece ser “cancelada” por zeros, e os pólos “não-cancelados” são os mesmos em todas as componentes — neste exemplo particular, podem ser observados seis pólos.

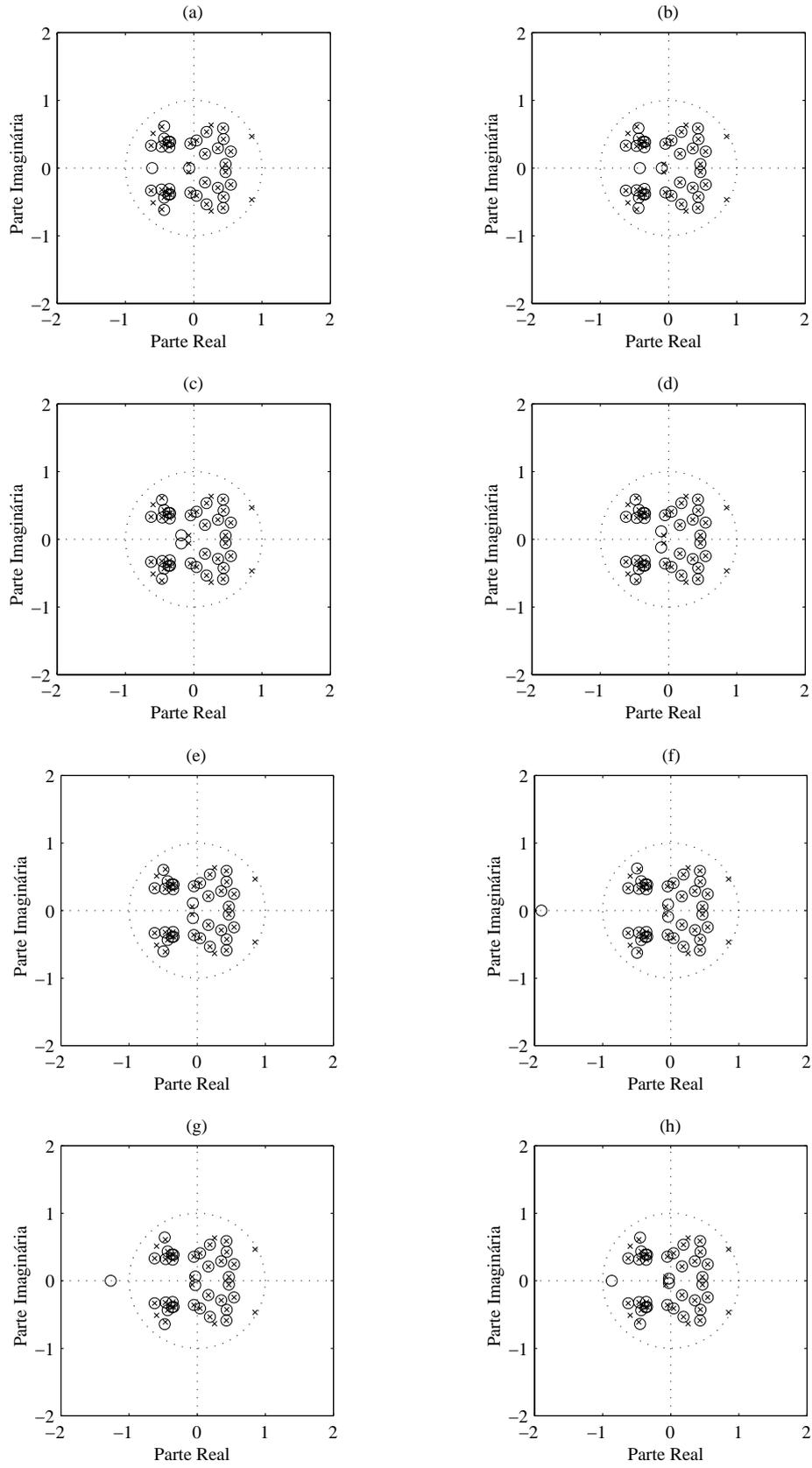


Figura 4.11: Caso Haar: (a) a (h) Detalhe dos diagramas de pólos e zeros das componentes polifásicas da sub-banda mais grave, respectivamente, de $G_0(z)$ a $G_{M-1}(z)$.

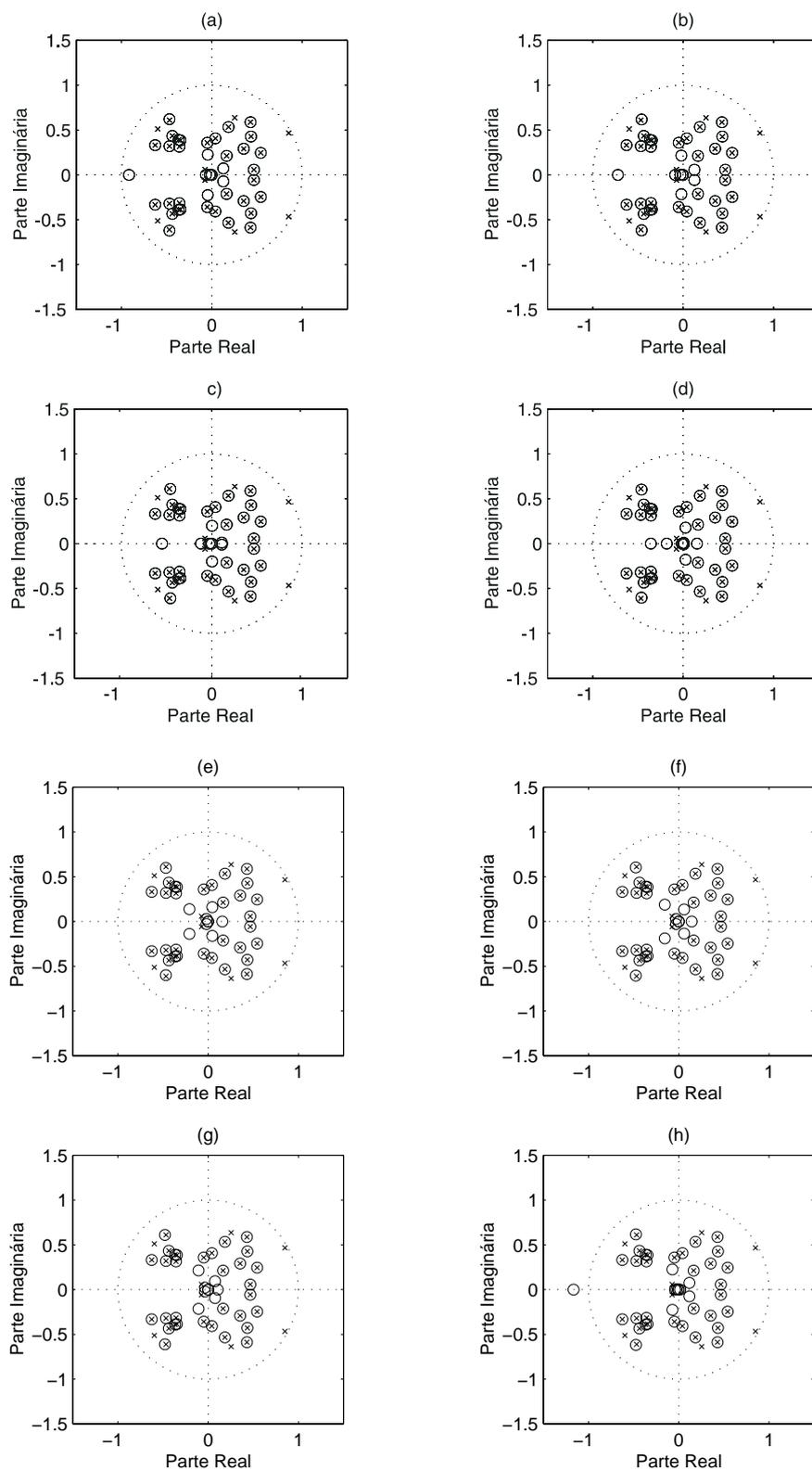


Figura 4.12: Caso QMF: (a) a (h) Detalhe dos diagramas de pólos e zeros das componentes polifásicas da sub-banda mais grave, respectivamente, de $G_0(z)$ a $G_{M-1}(z)$.

Entretanto, é interessante notar que os pólos “não-cancelados” nas componentes polifásicas da sub-banda mais grave têm localizações muito próximas das dos pólos **estimados** de um modelo AR de baixa ordem para o sinal naquela sub-banda, como mostram as Figs. 4.13(d) e 4.14(d).

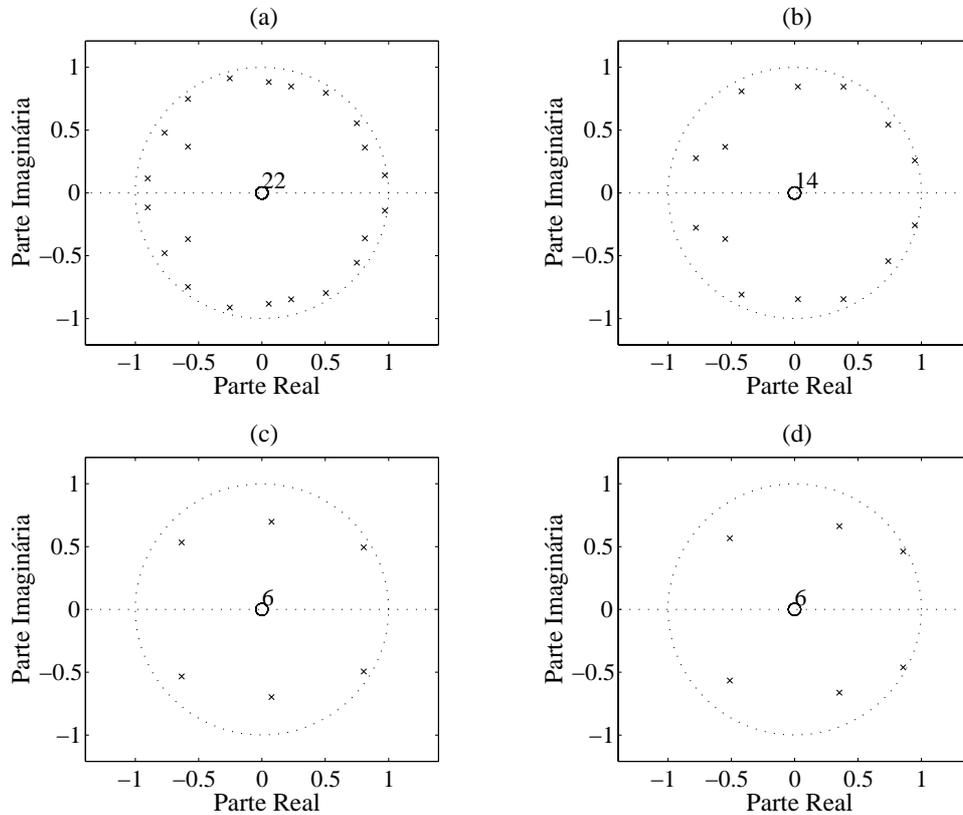


Figura 4.13: Caso Haar: (a) a (d) Diagramas de pólos e zeros correspondentes aos pólos estimados dos modelos AR de ordens 22, 14, 6 e 6 atribuídos aos os sinais em sub-bandas, respectivamente, da sub-banda mais aguda para a mais grave.

Os resultados exibidos pela comparação entre os diagramas de pólos e zeros das Figs. 4.11 e 4.13 e das Figs. 4.12 e 4.14 são muito interessantes, mas não são suficientes para garantir que a escolha de um modelo AR de baixa ordem será adequada para representar o sinal na sub-banda mais grave. Uma boa medida da adequação de um determinado modelo estatístico a um sinal é a capacidade de uma filtragem inversa, baseada nesse modelo, descorrelacionar o sinal de excitação correspondente ao sinal. Essa verificação é realizada a seguir.

As Figs. 4.15 e 4.16 mostram, respectivamente, as funções de autocorrelação relativas aos sinais em sub-bandas e seus respectivos sinais de excitação referentes ao Caso QMF.

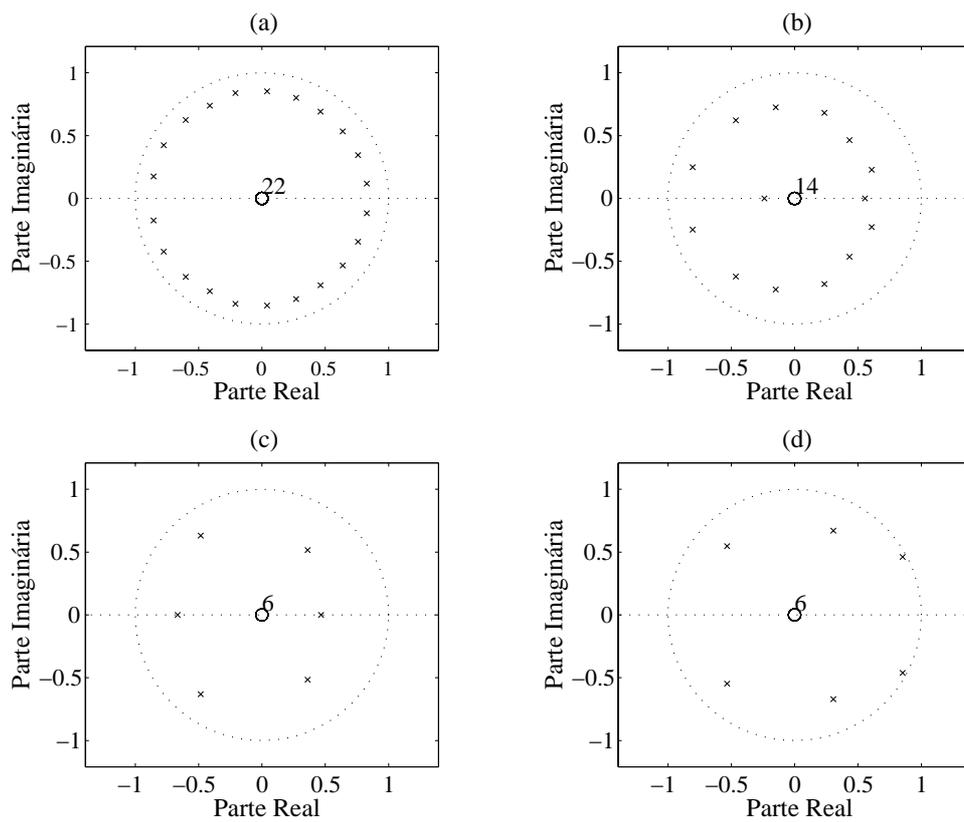


Figura 4.14: Caso QMF: (a) a (d) Diagramas de pólos e zeros correspondentes aos pólos estimados dos modelos AR de ordens 22, 14, 6 e 6 atribuídos aos sinais em sub-bandas, respectivamente, da sub-banda mais aguda para a mais grave.

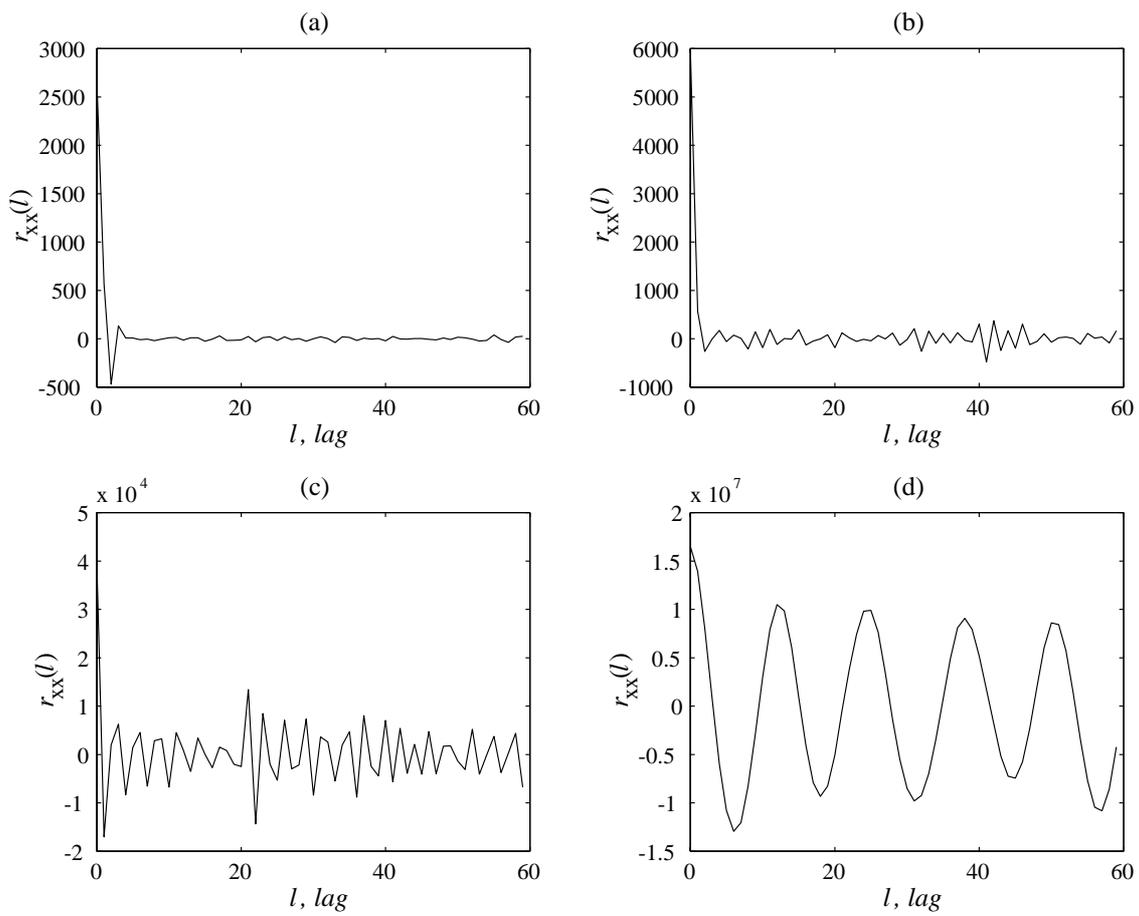


Figura 4.15: Caso QMF: (a) a (d) Funções de autocorrelação dos sinais em sub-bandas, respectivamente da sub-banda mais aguda para a mais grave.

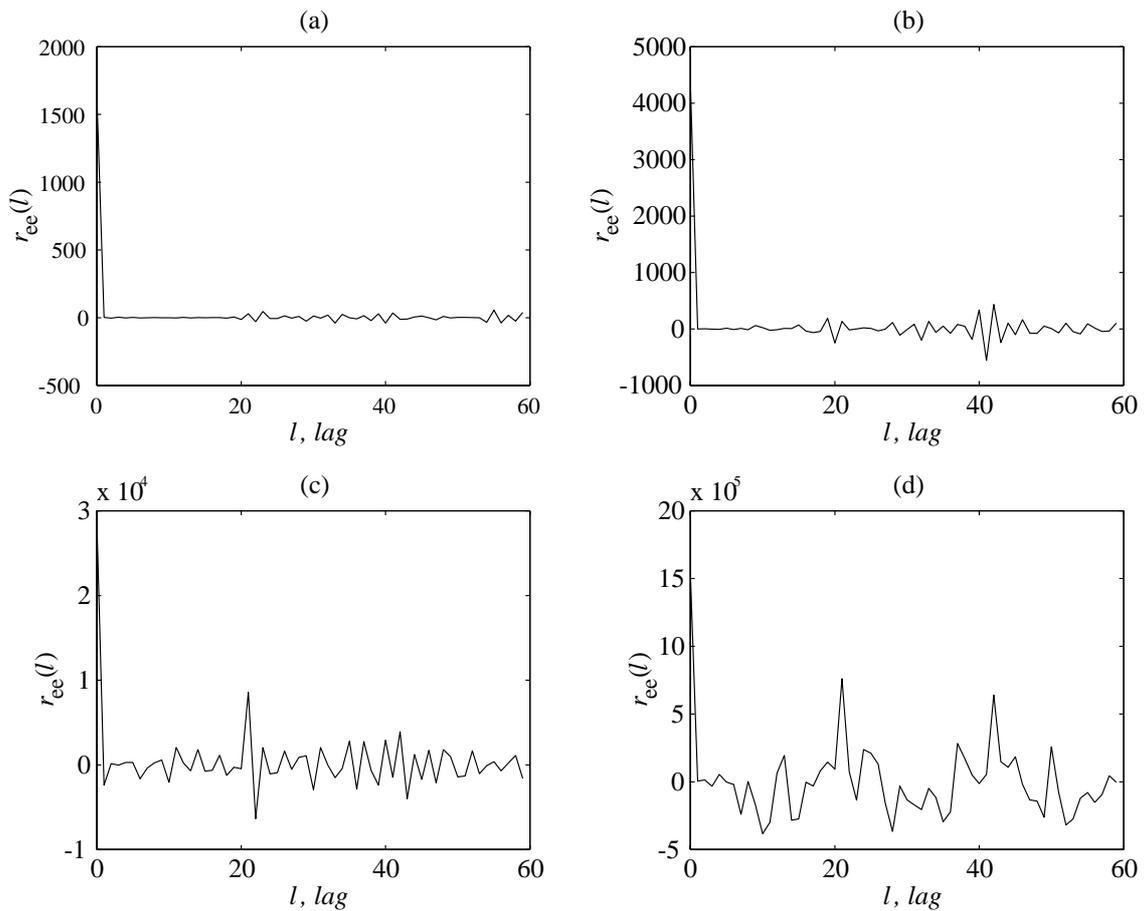


Figura 4.16: Caso QMF: (a) a (d) Funções de autocorrelação dos sinais de excitação correspondentes aos sinais em sub-bandas, respectivamente, da sub-banda mais aguda para a mais grave. A filtragem inversa dos processos AR contou com ordens 20, 10, 5 e 5.

É interessante notar que, examinando as funções de autocorrelação dos sinais das sub-bandas, a correlação entre as amostras dos sinais diminui em direção às sub-bandas mais agudas. A modelagem estatística dos sinais das sub-bandas mais agudas exigiria processos AR de altas ordens. A modelagem do sinal da sub-banda mais aguda como processo AR de ordem 20 é capaz diminuir a correlação das amostras da excitação correspondente, como mostra a Fig. 4.16(a). A correlação entre as amostras do sinal na sub-banda mais grave é mais forte; entretanto, seu modelamento por processo AR de baixa ordem não contribui, significativamente, para decorrelacionar as amostras do sinal de excitação correspondente, como mostra a Fig. 4.16(d). A adoção de ordens maiores para os sinais nas sub-bandas mais graves, sem dúvida, contribuiria para diminuir a correlação entre as amostras das excitações correspondentes; entretanto, o aumento das ordens dos modelos AR é um dos fatores que mais contribuem para o aumento do tempo computacional de processamento, não sendo, portanto, desejável, ainda mais se se levar em conta que os efeitos dos distúrbios impulsivos em baixas frequências é reduzido.

A adoção de modelos estatísticos adequados para os sinais em sub-bandas assume papel importante tanto no processo de detecção dos distúrbios impulsivos como na reconstrução do sinal, principalmente quando os interpoladores levam em consideração a estatística do sinal. A escolha do modelamento AR para os sinais das sub-bandas, apesar de não ser a escolha ótima, traz todas as vantagens provenientes da adaptação direta das etapas do sistema em banda completa para supressão de ruído impulsivo, como a obtenção do sinal de excitação por filtragem FIR e possibilidade de promover interpolação conjunta das amostras corrompidas através do algoritmo LSAR (ver Apêndice A).

4.3 Implementação do Sistema e Resultados

Nesta seção, são apresentados e analisados os resultados dos processamentos obtidos por um sistema em multitaxa para supressão de ruído impulsivo. O sistema inicialmente implementado decompõe o sinal em 4 sub-bandas com largura de banda abrangendo 1 oitava. A escolha de 4 sub-bandas foi arbitrária, mas sustentada pelas considerações feitas na subseção 4.2.1, quando se discutiu a desnecessidade de

decompor o sinal em um número muito grande de sub-bandas, face à não-percepção dos efeitos auditivos do ruído impulsivo nas sub-bandas de baixas frequências. A decomposição foi realizada por banco de análise com o protótipo QMF de ordem 15 mostrado na Fig.4.3.

O sinal de cada sub-banda é processado isoladamente dos sinais das demais sub-bandas; dessa forma, parâmetros de processamento (ver Seção 3.3) são estabelecidos separadamente para cada sub-banda.

Como o processamento em cada sub-banda é independente do processamento realizado nas demais sub-bandas, não há necessidade de segmentar o sinal de entrada em blocos estacionários para realizar a decomposição em sub-bandas; a princípio, decompõe-se todo o sinal de entrada em sub-bandas e realiza-se a segmentação em blocos nos sinais de cada sub-banda.

A escolha dos tamanhos dos blocos de sinal em cada sub-banda levou em consideração a discussão sobre a estacionaridade dos sinais em sub-bandas, apresentada na subseção 4.2.2. Nesta implementação, apostou-se na validade da hipótese de maiores tempos de estacionaridade para sinais mais graves. Dessa forma, os sinais em sub-bandas foram processados em blocos de 1024 amostras. É importante lembrar que 1024 amostras na sub-banda mais grave correspondem a 8192 amostras em banda completa, enquanto que 1024 amostras na sub-banda mais aguda correspondem a apenas 2048 amostras em banda completa.

Os parâmetros e critérios utilizados no processamento em sub-bandas são semelhantes aos utilizados no caso do processamento em banda completa. Em geral, a maioria dos parâmetros e critérios não é alterada conforme a sub-banda sob processamento; fogem a essa regra as ordens dos modelos AR adotados para os sinais em sub-bandas e a escolha do valor da constante de proporcionalidade K , associada ao cálculo do Limiar de Detecção dado por (3.1).

Denominando de $x_1(k)$, $x_2(k)$, $x_3(k)$ e $x_4(k)$ os sinais em sub-bandas, respectivamente, da sub-banda mais aguda para a mais grave, os parâmetros e critérios comuns e que permanecem invariantes a todas elas são descritos a seguir:

- Processamento realizado em blocos de 1024 amostras;
- União de distúrbios adjacentes separados por n amostras segundo (3.3) com $n_{min} = 2$ e $g = 6$;

- Utilização do Critério 2 (ver Seção 3.2.4) para a interrupção das iterações num dado bloco de sinal com $i_{max} = 4$ e redução progressiva do valor do limiar de localização segundo (3.4), com $b[1] = r = 0.5$ e $f = 3$;

As ordens do modelo AR escolhidas para os sinais em sub-bandas foram 20, 10, 5 e 5, respectivamente, para blocos supostamente estacionários de $x_1(k)$, $x_2(k)$, $x_3(k)$ e $x_4(k)$; A escolha do valor de K para cada sub-banda é empírica. Como numa divisão em 4 sub-bandas a sub-banda mais grave concentra parte significativa da informação do sinal de áudio, o valor de K nessa sub-banda é, em geral, feito maior para, justamente, evitar a ocorrência de falsos alarmes.

O processamento foi realizado em três sinais de áudio naturalmente contaminados por ruído impulsivo. Os sinais também estão naturalmente contaminados por ruído de fundo com característica perceptiva similar ao de ruído branco. Os ruídos impulsivos presentes nos sinais variam tanto na duração dos distúrbios quanto no percentual de amostras corrompidas.

A Tabela 4.1 mostra os valores do parâmetro K utilizados em cada sub-banda, no processamento dos três sinais de áudio corrompidos utilizados, aqui denominados de Sinal A, Sinal B e Sinal C.

Tabela 4.1: Comparação entre os valores do parâmetro K utilizados no processamento. Os nomes $x_1(k)$, $x_2(k)$, $x_3(k)$ e $x_4(k)$ se referem aos sinais das sub-bandas, respectivamente, da mais aguda para a mais grave.

	Valor de K			
	$x_1(k)$	$x_2(k)$	$x_3(k)$	$x_4(k)$
Sinal A	4.25	4.25	4.47	5.60
Sinal B	4.25	4.25	5.00	5.00
Sinal C	4.00	4.00	6	10

As Fig. 4.17 mostra uma comparação entre os sinais em sub-bandas correspondentes à decomposição de um trecho do Sinal A, antes e depois do processamento. A Fig. 4.18 mostra a mesma comparação, referente ao do Sinal B. Para se ter noção

das potências dos sinais em cada sub-banda, as escalas dos gráficos são mantidas iguais.

É interessante notar que a maior parte dos distúrbios impulsivos presentes nos sinais em sub-bandas é removido.

As Figs. 4.19 e 4.20 mostram trechos dos Sinais A e B em banda completa, nas versões original e processada, indicando que o sistema é capaz de suprimir os distúrbios impulsivos.

4.3.1 Tempos de Processamento e Simplificações

O tempo necessário ao processamento dos sinais corrompidos por ruído impulsivo varia, em geral, de modo bastante não linear, em relação a diversos fatores, dentre os quais podem-se citar o percentual de amostras corrompidas, as ordens dos modelos AR adotadas e a escolha dos parâmetros de processamento. A comparação entre os tempos de processamento necessários para processar um mesmo sinal usando o processamento em banda completa descrito no Capítulo 3 e o processamento em multitaxa descrito neste Capítulo indica que, para resultados perceptivos equivalentes e utilizando os parâmetros de processamento descritos em cada caso, o processamento em multitaxa pode ser até 2 vezes mais rápido.

A razão entre os tempos de processamento gastos para processar o sinal na sub-banda mais grave e o sinal na sub-banda mais aguda pode chegar a 1/150. Se, por um lado, os distúrbios impulsivos estão bastante evidenciados na sub-banda mais aguda, por outro, os sinais dessas sub-bandas contêm, em geral, pouca informação do sinal de áudio e, entretanto é nela que se gasta mais tempo de processamento. Uma forma de reduzir drasticamente o tempo de processamento nas sub-bandas mais agudas consiste em simplificar a forma de processamento a ser utilizada nos sinais dessas sub-bandas. Sabendo que as amostras dos sinais das sub-bandas mais agudas são pouco correlacionadas, pode-se realizar a detecção diretamente no domínio do sinal, ao invés de no domínio da excitação, e usar amostras de ruído branco Gaussiano com variância proporcional à estimada para o sinal do bloco em questão, para substituir as amostras tidas como corrompidas.

É importante observar que os procedimentos de detecção e interpolação simplificados, citados anteriormente, só devem ser utilizados em sinais com pouco conteúdo

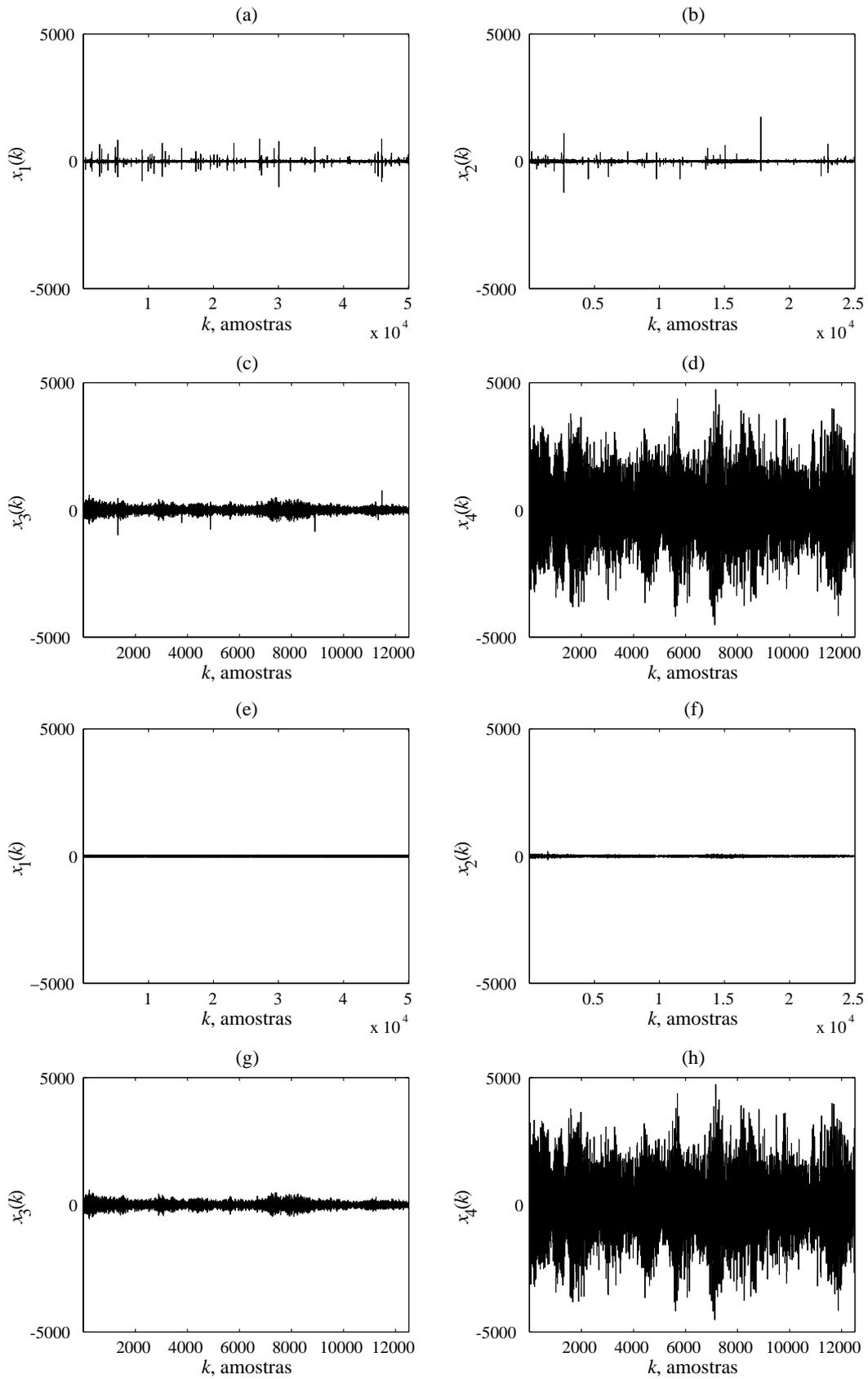


Figura 4.17: (a) a (d) Decomposição de um trecho do Sinal A em 4 sub-bandas, respectivamente, da sub-banda mais aguda para a sub-banda mais grave. (e) a (h) Respetivas versões processadas.

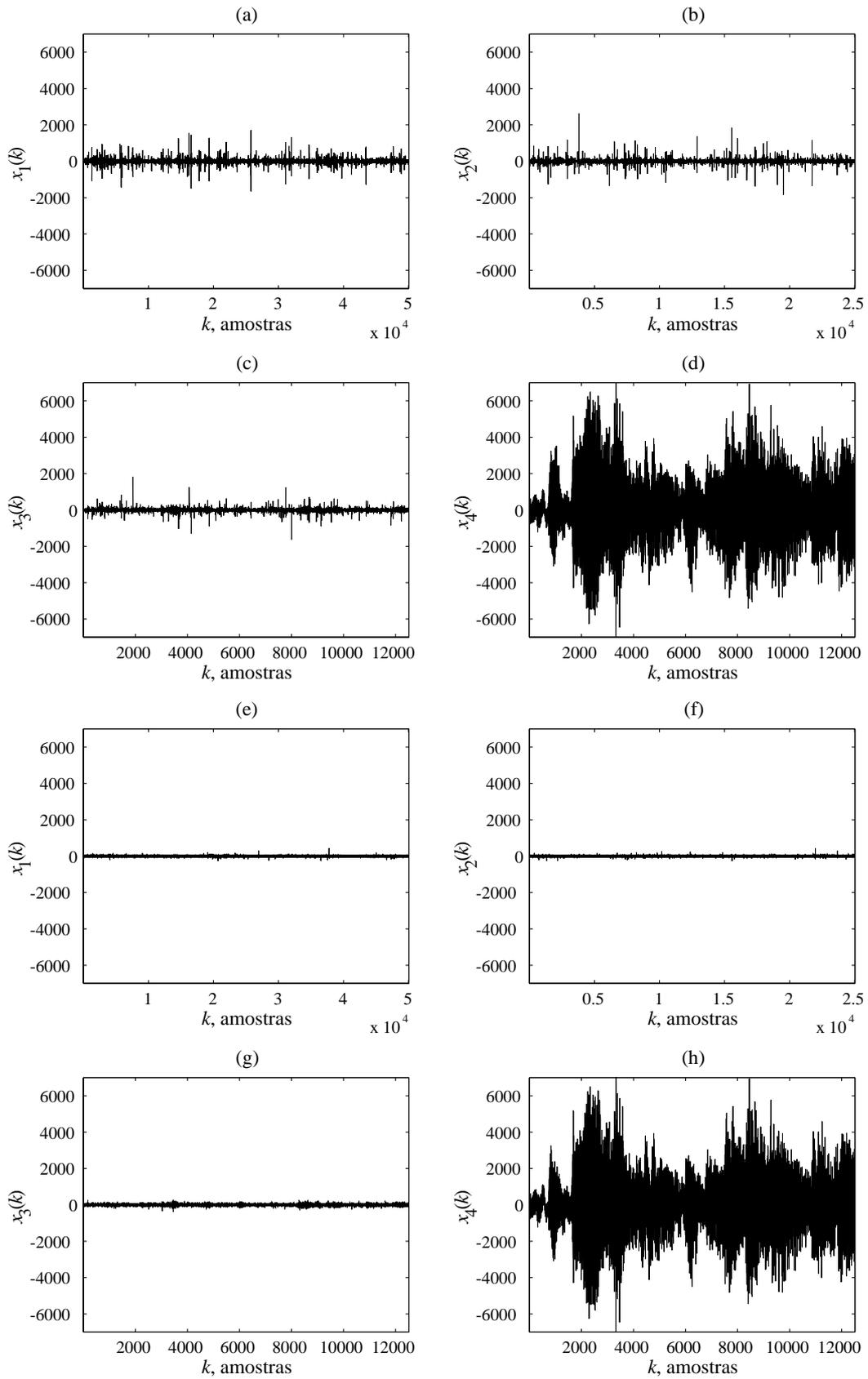


Figura 4.18: (a) a (d) Decomposição de um trecho do Sinal B em 4 sub-bandas, respectivamente, da sub-banda mais aguda para a sub-banda mais grave. (e) a (h) Respective versões processadas.

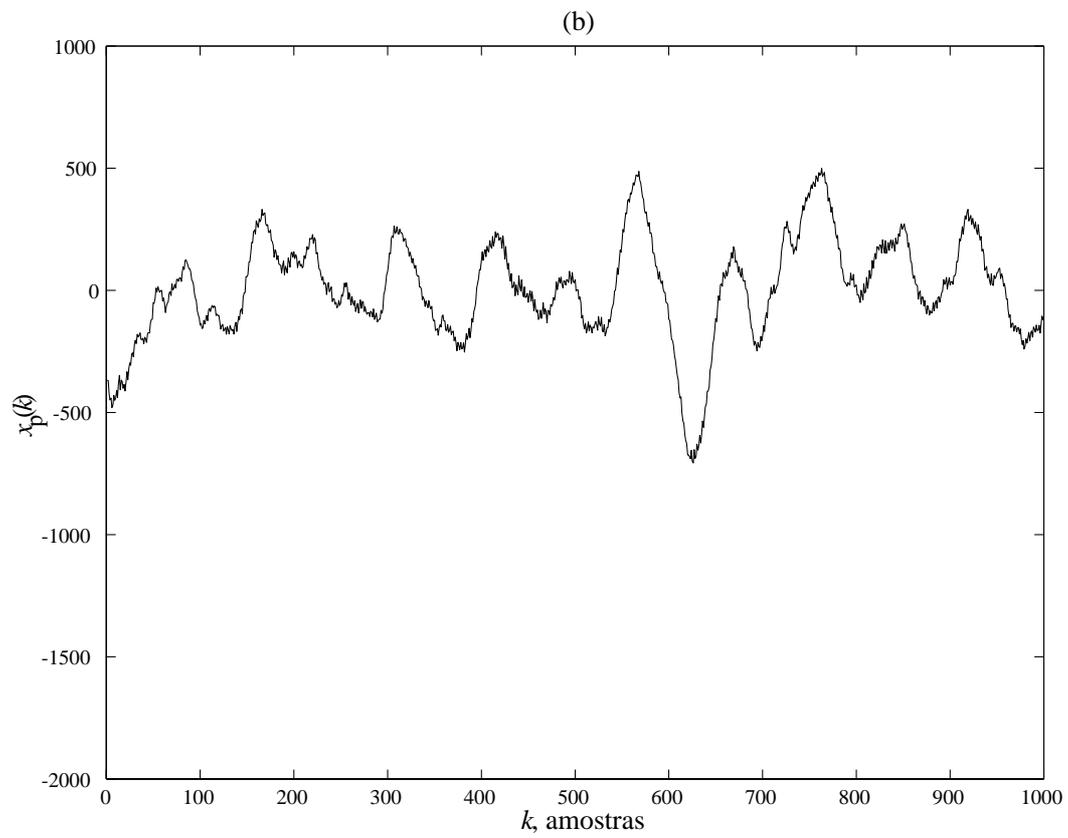
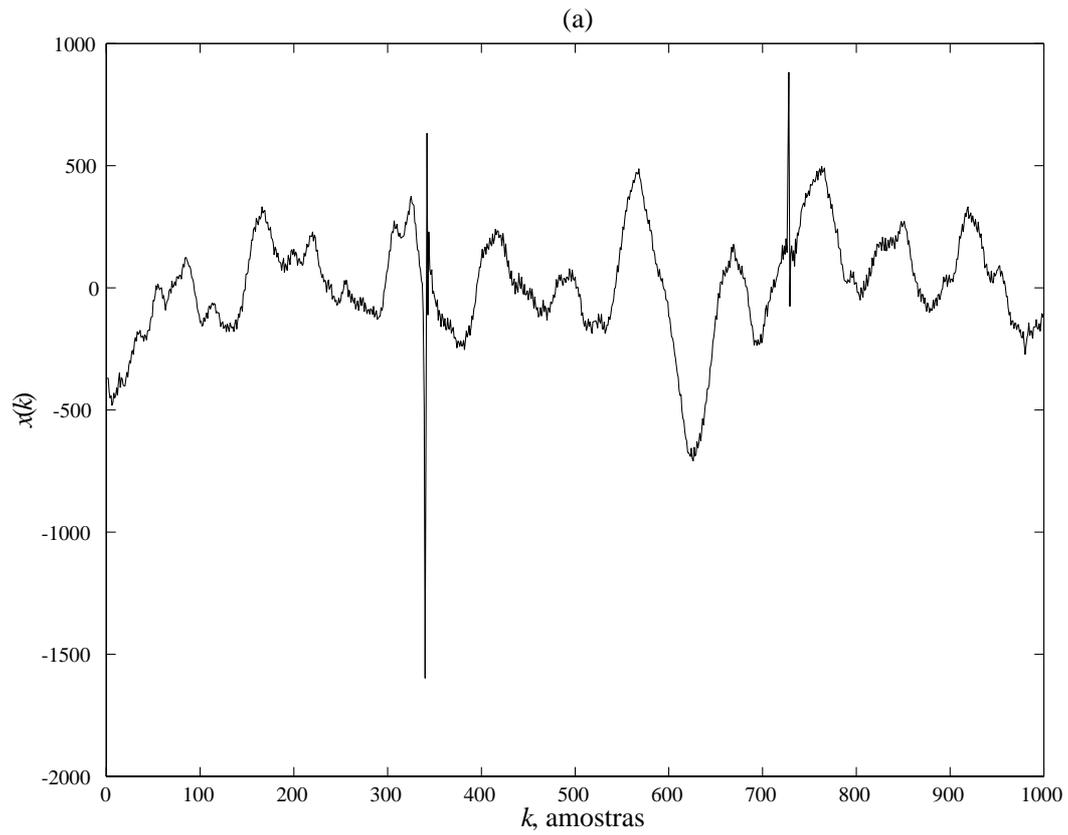


Figura 4.19: (a) Sinal ruidoso e (b) sinal restaurado, correspondentes a um trecho do Sinal A.

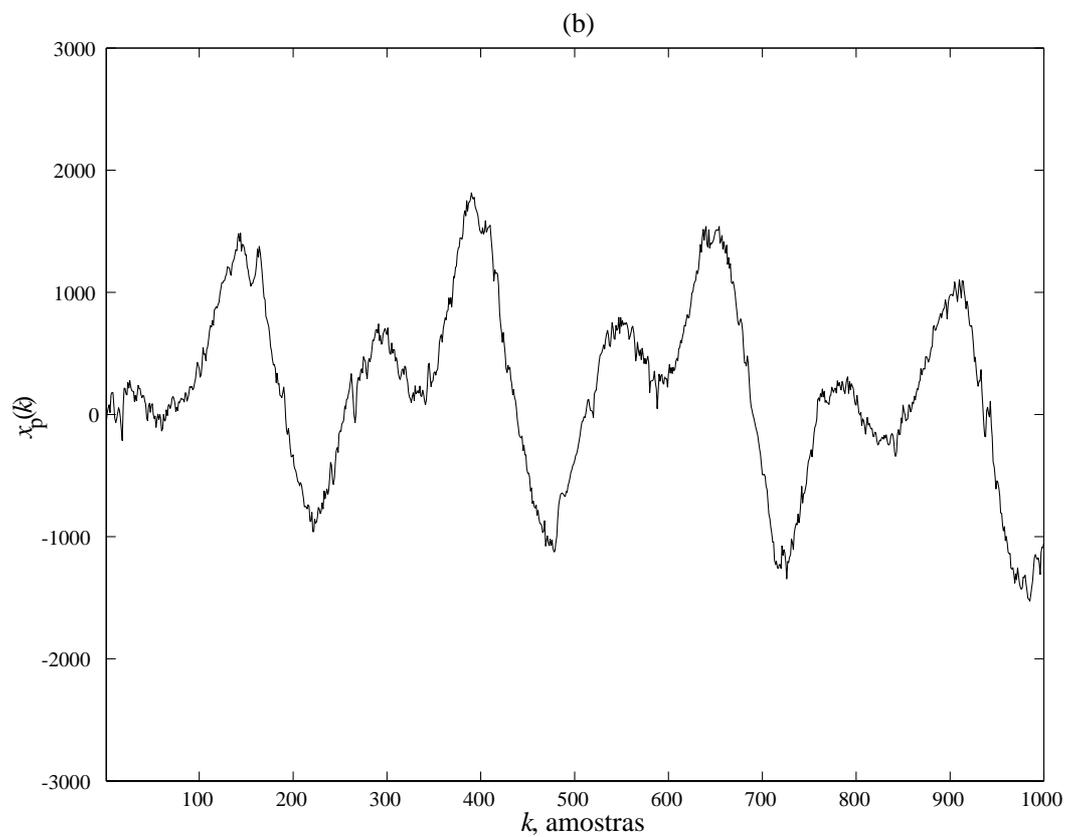
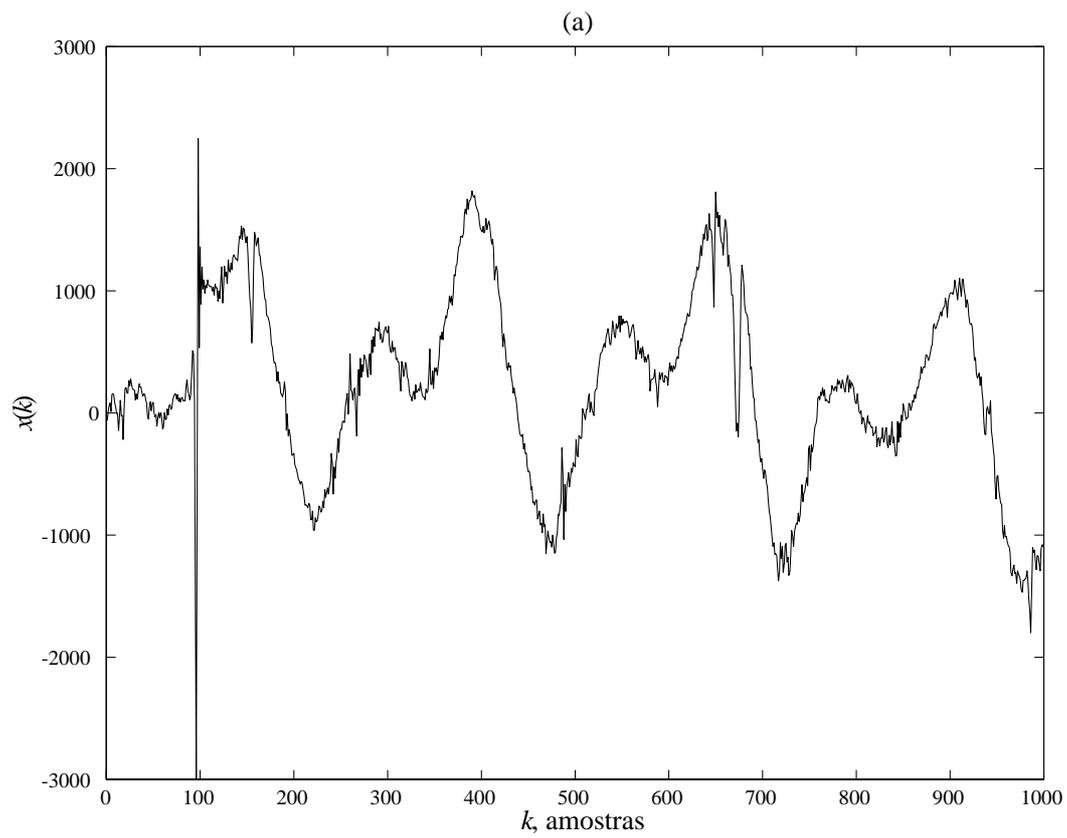


Figura 4.20: (a) Sinal ruidoso e (b) sinal restaurado, correspondentes a um trecho do Sinal B.

espectral em altas frequências. Sua utilização, entretanto, pode tornar o processamento até 20 vezes mais rápido, sem perda significativa da qualidade do sinal reconstruído.

Capítulo 5

Detecção de Pulsos Longos em Sinais de Áudio

5.1 Introdução

Danos sérios no sulco de uma matriz em disco ou cilindro — originados, por exemplo, por fendas ou arranhões profundos acidentalmente criados na superfície ou até mesmo por quebra — excitam o sistema reprodutor de um toca-discos com estímulos de alta amplitude (comparada com as amplitudes originalmente cortadas nas paredes do sulco), semelhantes a impulsos. A resposta do sistema reprodutor a esses estímulos produz distúrbios, em geral de longa duração e alta amplitude, que constituem um dos tipos mais desagradáveis de degradação em gravações analógicas de áudio. Esses distúrbios serão, doravante, denominados de Pulsos Longos.

O processamento visando à supressão de pulsos longos deve ser o primeiro a ser realizado em um sinal também corrompido por ruído branco ou impulsivo. A presença dos pulsos longos pode polarizar erroneamente a estimativa da variância do ruído branco, devido a seu espectro largo (ver Seção 5.2), como também a estimativa dos coeficientes do modelo AR adotado para um dado trecho do sinal (ver Seção 1.3).

Semelhantemente ao processamento para supressão de ruído impulsivo, o processamento visando à supressão de pulsos longos pode ser dividido nas etapas de Detecção e Reconstrução (ver Capítulo 2). Neste Capítulo, entretanto, serão enfatizadas técnicas para a detecção dos pulsos longos.

Técnicas de processamento visando à supressão de pulsos longos já foram

abordadas em [4] — cujo algoritmo de supressão contempla etapas de Detecção e Reconstrução, ambas baseadas em verificação da correlação temporal cruzada entre os pulsos longos e um pulso-padrão — e em [13] — cujo algoritmo se restringe à etapa de Reconstrução, que se baseia em separação de processos AR, sendo as localizações dos pulsos longos no sinal de áudio tomadas como informações previamente conhecidas.

Na Seção 5.2, são abordadas algumas características associadas aos pulsos longos, como suas amplitudes relativas ao sinal de áudio não-corrompido e seu conteúdo espectral. Na Seção 5.3, são revistos os algoritmos abordados em [4] e [13] e, finalmente, na Seção 5.4, são abordadas 4 algoritmos para detecção de pulsos longos em blocos, e proposta uma abordagem baseada em análise de conteúdo espectral, que apresenta desempenho de detecção superior em sinais com conteúdo espectral rico e dinâmica ampla, quando comparado aos outros 3 algoritmos.

5.2 Caracterização dos Pulsos Longos

Em geral, os pulsos longos podem ser divididos, temporalmente, em duas partes. A parte inicial consiste de uma descontinuidade semelhante a um distúrbio impulsivo de alta amplitude. A segunda parte, que sucede a primeira, é formada por oscilações amortecidas de baixas frequências, de longa duração, associadas à resposta impulsiva, em geral não-linear, do sistema de reprodução da matriz.

A Fig. 5.1 mostra um exemplo de pulso longo e ilustra sua divisão nas partes anteriormente mencionadas.

Os pulsos longos parecem contaminar aditivamente [4, 13] o sinal de áudio e podem atingir durações da ordem de grandeza do tempo de estacionaridade máximo suposto para o sinal. Por essa razão, tais distúrbios não podem ser tratados pelas mesmas técnicas empregadas para ruído impulsivo.

A análise do conteúdo espectral dos pulsos longos revela que as descontinuidades iniciais de alta amplitude exibem componentes de altas frequências com considerável energia. A Fig. 5.2 mostra o espectrograma de um sinal de áudio originalmente contaminado por pulsos longos. A ocorrência de faixas espectrais temporalmente localizadas e com conteúdo espectral largo é bastante indicativa da

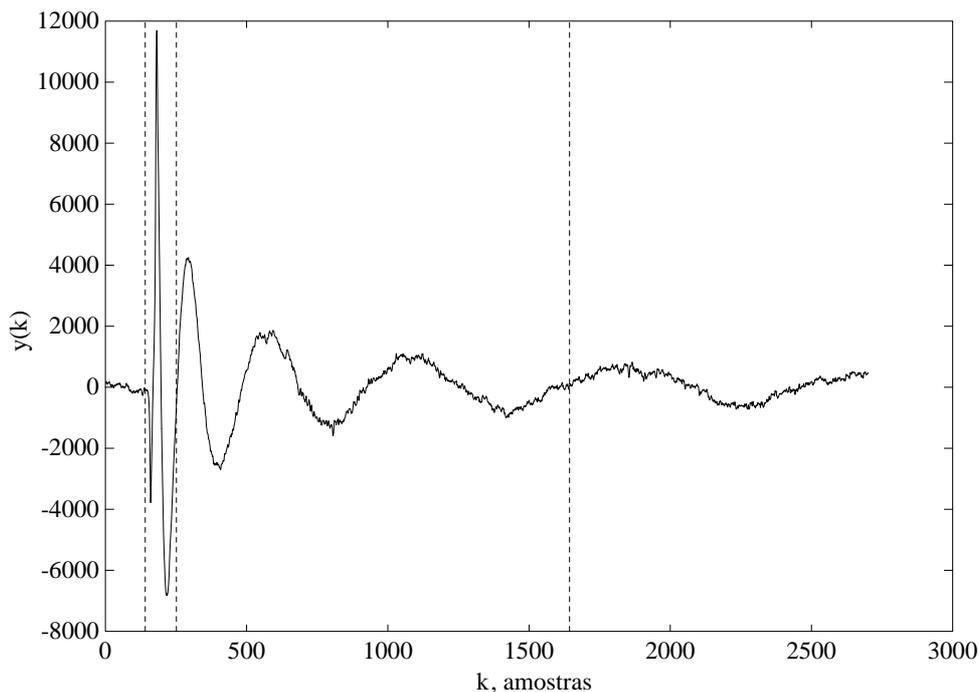


Figura 5.1: Exemplo de pulso longo: as linhas verticais delimitam aproximadamente as regiões de descontinuidade inicial e de transitório de baixa frequência.

presença de pulsos longos e pode ser bastante útil na sua detecção, como será visto na Seção 5.4.

5.3 Supressão por Similaridade e por Separação de Processos AR

5.3.1 Supressão por Similaridade

O método de detecção/supressão de pulsos longos por similaridade ou por pulso-padrão apresentado em [4] se baseia na hipótese de que os pulsos longos só variam em amplitude, mantendo praticamente inalterados seus formatos e durações, ditados pela resposta impulsiva do aparato de reprodução da matriz. Nesses termos, a etapa inicial do método consiste em obter um pulso-padrão, representativo dos demais pulsos presentes no sinal; deslocar, amostra a amostra, esse pulso-padrão ao longo do sinal e calcular o coeficiente de correlação cruzada entre as amostras do pulso padrão e as do sinal correspondentes à posição atual do pulso-padrão em

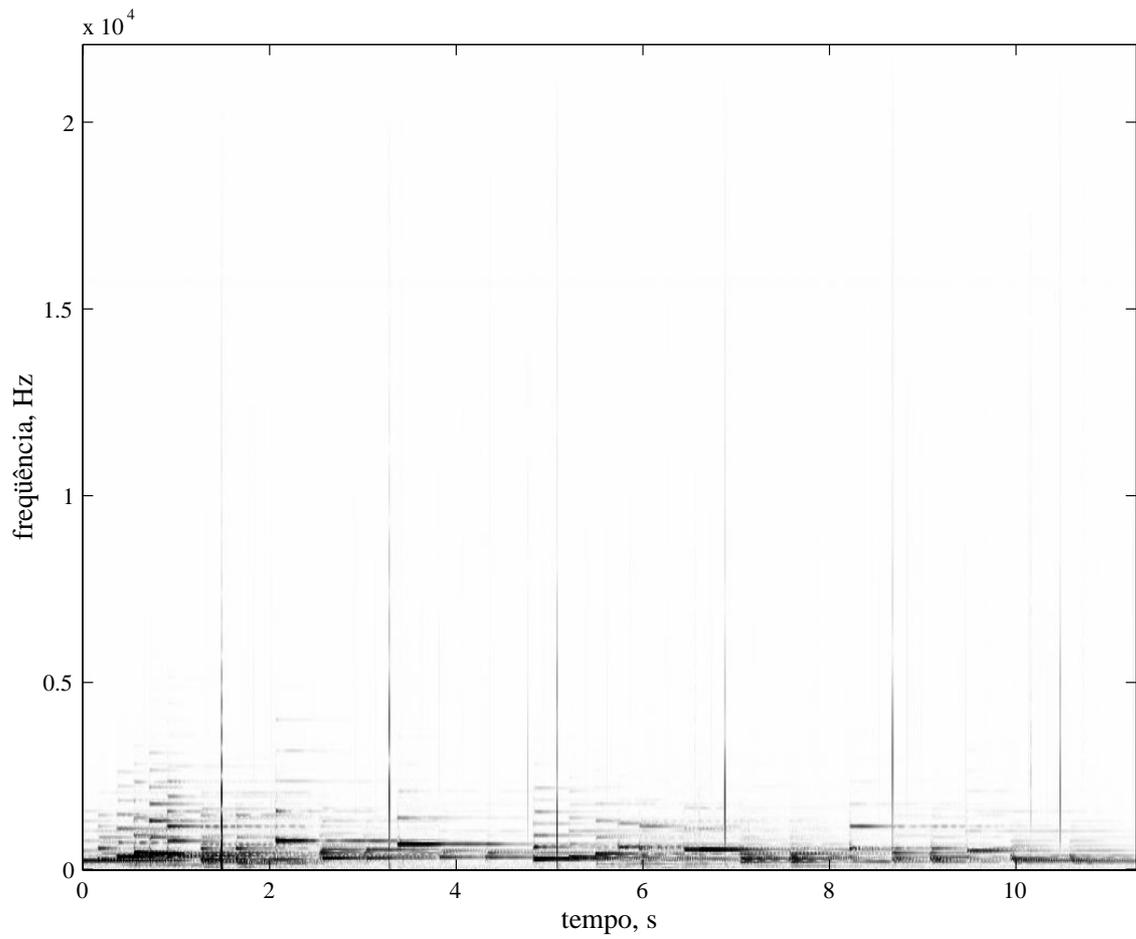


Figura 5.2: Espectrograma de um sinal de áudio originalmente contaminado por pulsos longos: blocos de 1024 amostras, FFT de 1024 pontos e superposição de 50%.

relação ao sinal. A seqüência de valores dos coeficientes de correlação cruzada deve exibir picos com valores próximos a 1, indicando a forte correlação entre o pulso-padrão e o sinal corrompido e, conseqüentemente, a presença de um pulso similar. A detecção dos pulsos longos similares é realizada comparando-se a seqüência de valores dos coeficientes de correlação cruzada com um limiar de valor próximo a 1.

A etapa de supressão é subdividida em duas partes: a primeira consiste em sincronizar no tempo o pulso-padrão e o pulso a ser suprimido e subtrair do sinal corrompido uma versão escalada em amplitude do pulso-padrão; espera-se remover com essa operação o transitório de baixas freqüências do pulso. A segunda parte consiste em reconstruir o trecho correspondente à descontinuidade inicial que ainda restou da operação de subtração do pulso-padrão, e que pode ser tratada como um distúrbio impulsivo.

Como dito anteriormente, os pulsos longos contaminam aditivamente o sinal de áudio e, considerando apenas o trecho corrompido, o sinal ruidoso pode ser modelado como

$$y(k) = x(k) + n(k), \quad (5.1)$$

sendo $x(k)$ o sinal original e $n(k)$ o pulso longo, supostos não-correlacionados.

Por hipótese, o pulso-padrão $\bar{n}(k)$, representativo da família de $n(k)$, difere essencialmente em amplitude dos demais pulsos, isto é,

$$n(k) \approx \mu \bar{n}(k); \quad (5.2)$$

e uma estimativa para o sinal não-ruidoso $x(k)$

$$\hat{x}(k) = y(k) - \mu \bar{n}(k), \quad (5.3)$$

é diretamente obtida, desde que se conheça μ .

O valor de μ pode ser obtido correlacionando no tempo o trecho ruidoso com o pulso-padrão:

$$\sum_{k=0}^{N-1} y(k)\bar{n}(k) = \sum_{k=0}^{N-1} x(k)\bar{n}(k) + \mu \sum_{k=0}^{N-1} [\bar{n}(k)]^2. \quad (5.4)$$

Da não-correlação entre $x(k)$ e $\bar{n}(k)$, resulta

$$\mu \approx \frac{\sum_{k=0}^{N-1} y(k)\bar{n}(k)}{\sum_{k=0}^{N-1} [\bar{n}(k)]^2}, \quad (5.5)$$

que, levado à equação (5.3), soluciona o problema.

A obtenção do pulso-padrão segue um procedimento bastante semelhante ao de detecção de pulsos longos: um pulso representativo dos que corrompem os sinais é escolhido, inicialmente, como padrão e usado para promover a detecção de pulsos similares; assim, começando no início do sinal, cada novo pulso detectado é incorporado ao pulso-padrão — por soma dos dois pulsos sincronizados no tempo —, procede-se, com esse novo pulso-padrão, à detecção do próximo pulso similar, que também é incorporado ao pulso-padrão, e assim por diante. Ao final do processo, espera-se ter um pulso-padrão “limpo”, ou seja, com pouca informação do sinal de áudio não-corrompido.

Quando o arranhão ocorre radialmente na superfície de um disco, os pulsos longos aparecem de modo aproximadamente periódico no sinal de áudio, e suas posições podem ser estimadas levando-se em consideração a taxa de amostragem utilizada na digitalização e a velocidade angular de reprodução da matriz. Desse modo, a busca de pulsos similares, seja no procedimento de obtenção do pulso-padrão ou no de detecção/supressão, não precisa ser por deslocamento do pulso padrão amostra-a-amostra por todo o sinal, mas apenas nas regiões no entorno das posições estimadas de ocorrência dos pulsos longos. Se, por um lado, a utilização desse recurso acelera o processamento, por outro, reduz a generalidade do método em relação à detecção de pulsos em quaisquer posições.

A principal limitação desse método está na hipótese de invariabilidade dos pulsos longos, que, em geral, não é sempre satisfeita. Os pulsos-longos variam tanto em duração quanto no formato e, no caso de dois arranhões bem próximos, os pulsos longos resultantes podem aparecer superpostos, dificultando a detecção/supressão através de pulso-padrão [13].

5.3.2 Supressão por Separação de Processos AR

A supressão de pulsos longos através de separação de processos AR proposta em [13] é mais geral que o método de detecção/supressão por pulso-padrão [4], entretanto, só aborda a supressão dos pulsos longos, ou seja, suas localizações devem ser conhecidas *a priori*. A idéia do método é modelar o pulso longo como um processo AR de baixa ordem (2 ou 3) e promover sua separação do sinal de áudio — mode-

lado como um processo AR de ordem entre 40 e 80 — através de um algoritmo de separação de processos AR.

Descrevendo de modo geral a separação de processos AR, considera-se, inicialmente, o vetor de N amostras

$$\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2, \quad (5.6)$$

sendo \mathbf{y}_1 e \mathbf{y}_2 originados por processos AR Gaussianos independentes análogos ao da equação (1.1), com coeficientes AR \mathbf{a}_1 e \mathbf{a}_2 e variâncias de excitação $\sigma_{e_1}^2$ e $\sigma_{e_2}^2$, respectivamente.

Se, por hipótese, são conhecidas as p_1 primeiras amostras de \mathbf{y}_1 e as p_2 primeiras amostras de \mathbf{y}_2 , é possível expressar suas densidades de probabilidade como

$$f_{\mathbf{y}_1}(\mathbf{y}_1) = \frac{1}{\sqrt{(2\pi\sigma_{e_1}^2)^{N-p_1}}} e^{-\frac{1}{2\sigma_{e_1}^2} \mathbf{y}_1^T \mathbf{A}_1^T \mathbf{A}_1 \mathbf{y}_1} \quad (5.7)$$

e

$$f_{\mathbf{y}_2}(\mathbf{y}_2) = \frac{1}{\sqrt{(2\pi\sigma_{e_2}^2)^{N-p_2}}} e^{-\frac{1}{2\sigma_{e_2}^2} \mathbf{y}_2^T \mathbf{A}_2^T \mathbf{A}_2 \mathbf{y}_2}, \quad (5.8)$$

sendo \mathbf{A}_1 e \mathbf{A}_2 definidas de forma análoga à equação (A.3) (ver Apêndice A).

A obtenção de $\mathbf{y}_1^{\text{MAP}}$ (ver Seção 2.3) decorre da maximização de $f_{\mathbf{y}_1|\mathbf{y}}(\mathbf{y}_1|\mathbf{y})$, que, pelo Teorema de Bayes pode ser escrita como

$$f_{\mathbf{y}_1|\mathbf{y}}(\mathbf{y}_1|\mathbf{y}) = \frac{f_{\mathbf{y}|\mathbf{y}_1}(\mathbf{y}|\mathbf{y}_1)f_{\mathbf{y}_1}(\mathbf{y}_1)}{f_{\mathbf{y}}(\mathbf{y})}. \quad (5.9)$$

A densidade $f_{\mathbf{y}}(\mathbf{y})$ é constante para um dado \mathbf{y} conhecido e funciona como um fator de escalamento; adicionalmente, seria interessante expressar $f_{\mathbf{y}_1|\mathbf{y}}(\mathbf{y}_1|\mathbf{y})$ em função das densidades $f_{\mathbf{y}_1}(\mathbf{y}_1)$ e $f_{\mathbf{y}_2}(\mathbf{y}_2)$. Isso pode ser feito realizando a transformação de variáveis $\mathbf{y}_2 = \mathbf{g}(\mathbf{y}) = \mathbf{y} - \mathbf{y}_1$, que permite escrever

$$f_{\mathbf{y}_2|\mathbf{y}_1}(\mathbf{y}_2|\mathbf{y}_1) = \frac{1}{|\det(\mathbf{J})|} f_{\mathbf{y}|\mathbf{y}_1}(\mathbf{y}|\mathbf{y}_1)|_{\mathbf{y}=\mathbf{g}^{-1}(\mathbf{y}_2)}. \quad (5.10)$$

Como o Jacobiano $\mathbf{J} = \frac{\partial \mathbf{g}(\mathbf{y})}{\partial \mathbf{y}} = \mathbf{I}$ e $\mathbf{g}^{-1}(\mathbf{y}_2) = \mathbf{y}_1 + \mathbf{y}_2 = \mathbf{y}$, tem-se que $f_{\mathbf{y}|\mathbf{y}_1}(\mathbf{y}|\mathbf{y}_1) = f_{\mathbf{y}_2|\mathbf{y}_1}(\mathbf{y}_2|\mathbf{y}_1)$. Da independência entre \mathbf{y}_1 e \mathbf{y}_2 decorre $f_{\mathbf{y}_2|\mathbf{y}_1}(\mathbf{y}_2|\mathbf{y}_1) = f_{\mathbf{y}_2}(\mathbf{y}_2)$ e a maximização de

$$f_{\mathbf{y}_1|\mathbf{y}}(\mathbf{y}_1|\mathbf{y}) = \frac{f_{\mathbf{y}_2}(\mathbf{y}_2) f_{\mathbf{y}_1}(\mathbf{y}_1)}{f_{\mathbf{y}}(\mathbf{y})} \quad (5.11)$$

resulta em

$$\mathbf{y}_1^{\text{MAP}} = \left(\frac{\mathbf{A}_1^T \mathbf{A}_1}{\sigma_{e_1}^2} + \frac{\mathbf{A}_2^T \mathbf{A}_2}{\sigma_{e_2}^2} \right)^{-1} \frac{\mathbf{A}_2^T \mathbf{A}_2}{\sigma_{e_2}^2} \mathbf{y}. \quad (5.12)$$

Passando ao problema de restauração, pode-se associar $x(k)$ com coeficientes AR \mathbf{a}_x a \mathbf{y}_1 e $n(k)$ com coeficientes AR \mathbf{a}_n a \mathbf{y}_2 . Contudo, a característica temporal dos pulsos a processar — início rápido de alta amplitude e curta duração seguido de transitório lento de longa duração — sugere um modelo composto, em que sua excitação $e_n(k)$ alterna entre uma variância elevada $\sigma_{e_{n_1}}^2$ nas regiões iniciais dos pulsos e uma variância reduzida $\sigma_{e_{n_0}}^2$ durante todo o resto do tempo. Recorrendo a um vetor \mathbf{i} de N componentes binárias que indiquem por 1 as amostras de alta amplitude da descontinuidade inicial do pulso e por zero as demais amostras, pode-se obter a matriz-correlação $\mathbf{\Lambda}$ de $e_n(k)$, que é diagonal e com elementos

$$\lambda_{kk} = \sigma_{e_{n_0}}^2 + i(k)(\sigma_{e_{n_1}}^2 - \sigma_{e_{n_0}}^2). \quad (5.13)$$

A densidade de probabilidade de \mathbf{n} se torna

$$f_{\mathbf{n}}(\mathbf{n}) = \frac{1}{\sqrt{(2\pi)^{N-p_2} |\mathbf{\Lambda}|}} e^{-\frac{1}{2} \mathbf{n}^T \mathbf{A}_n^T \mathbf{\Lambda}^{-1} \mathbf{A}_n \mathbf{n}}, \quad (5.14)$$

levando à nova solução

$$\mathbf{x}^{\text{MAP}} = \left(\frac{\mathbf{A}_x^T \mathbf{A}_x}{\sigma_{e_x}^2} + \mathbf{A}_n^T \mathbf{\Lambda}^{-1} \mathbf{A}_n \right)^{-1} \mathbf{A}_n^T \mathbf{\Lambda}^{-1} \mathbf{A}_n \mathbf{y}. \quad (5.15)$$

Os coeficientes do modelo AR associado ao sinal não-ruído podem ser obtidos por estimativa LS usando um trecho não-corrompido imediatamente anterior ao pulso longo a ser suprimido; já para os coeficientes do modelo AR do pulso deve-se utilizar o trecho posterior à ocorrência da descontinuidade inicial de alta amplitude de uma versão “limpa” do pulso. O vetor \mathbf{i} pode ser obtido usando uma abordagem baseada em detecção por filtragem inversa, tal como no processamento de detecção de ruído impulsivo, associando aos índices das amostras da excitação cujos módulos excedem um limiar de valor alto, elementos de \mathbf{i} com valor 1.

As maiores desvantagens do método de supressão de pulsos longos por separação de processos AR são seu custo computacional elevado e a dificuldade de estimar certos parâmetros, como a variância da excitação correspondente ao início do pulso $\sigma_{e_{n_1}}^2$, e de obter uma versão “limpa” do pulso longo de modo a estimar sem polarização seus coeficientes AR \mathbf{a}_n . Em [13], o autor adota o caso limite $\sigma_{e_{n_1}}^2 \rightarrow \infty$, que leva a zero os elementos de $\mathbf{\Lambda}^{-1}$ associados a elementos de \mathbf{i} iguais a 1. Já para obtenção dos coeficientes AR dos pulso longo, o autor diz ser suficiente usar, inicialmente, um pulso de um trecho sem informação do sinal de áudio, como nos intervalos

ou no final do disco; posteriormente, os pulsos suprimidos podem ser utilizados para refinar o modelo.

Em contrapartida, a supressão de pulsos longos por separação de processos AR tem as vantagens de ser menos dependente da hipótese de invariabilidade dos pulsos longos, o que permite suprimir tanto pulsos com transitórios diferentes como pulsos superpostos, e de realizar a supressão dos pulsos em uma única etapa.

5.4 Detecção de Pulsos Longos

A detecção de pulsos longos consiste em localizar temporalmente as regiões do sinal de áudio corrompidas por pulsos longos. Quando se deseja realizar processamento em blocos, a detecção de pulsos longos tem como objetivo identificar em quais blocos do sinal existem ocorrências de pulsos longos. Nesse sentido, a idéia básica de um sistema de detecção em blocos consiste em extrair informações dos blocos do sinal que permitam distinguir, por algum procedimento seletivo, quais blocos estão corrompidos por pulsos longos. Desta forma, o sistema de detecção será tanto melhor quanto maior for o contraste exibido pelas informações extraídas dos blocos corrompidos em relação aos não-corrompidos.

A seguir, são abordados 4 métodos para detecção de pulsos longos, utilizando processamento em blocos. Os Métodos A e B têm apelo quase trivial; os Métodos C e D são apresentados originalmente neste trabalho.

5.4.1 Método A

A idéia do Método A é realizar a detecção dos blocos contaminados por pulsos longos comparando-se o módulo do sinal dos blocos com um limiar de valor arbitrariamente adotado. Os blocos em que pelo menos uma amostra do sinal exibir módulo de valor maior que o valor do limiar são, então, selecionados como corrompidos por pulsos longos.

A justificativa para esse método reside nas amplitudes elevadas apresentadas pelos pulsos longos, logo no início de suas ocorrências, em relação às amplitudes do resto do sinal. Nesses termos, a escolha de um limiar de valor bastante alto é conveniente. Essa forma de detecção pode ser problemática quando o sinal de áudio

tem faixa dinâmica muito ampla, e suas amplitudes podem ser comparáveis ao valor alto adotado para o limiar. Nesse caso, a ocorrência de falsos alarmes seria muito elevada; em contrapartida, a elevação do valor do limiar pode levar a altos índices de não-deteção.

A Fig. 5.3 mostra, respectivamente, 2 sinais reais de áudio contaminados por pulsos longos, denominados daqui para a frente de Sinal A e Sinal B. O Sinal A é naturalmente corrompido por pulsos longos; já o Sinal B é corrompido artificialmente por pulsos longos reais.

O Sinal A tem dinâmica restrita, e nele os pulsos longos estão bastante evidenciados, enquanto que, no Sinal B, com dinâmica ampla, são mascarados pelas amplitudes altas das amostras no início da segunda metade do trecho. O Método A é mais adequado a detectar pulsos longos em sinais com característica de dinâmica similar à do Sinal A; para sinais do tipo do Sinal B, é difícil escolher um limiar único que realize adequadamente um compromisso entre falsa deteção e não-deteção dos blocos corrompidos em todo o sinal.

5.4.2 Método B

O Método B consiste em selecionar os blocos contaminados por pulsos longos comparando a potência média dos sinais nos blocos, definida em (5.16), com um limiar de valor arbitrariamente adotado. É uma abordagem semelhante à utilizada em seleção de trechos de fala vozeada em processamento de voz.

A potência média do sinal de um dado bloco b é definida como:

$$\varepsilon(b) = \frac{1}{N} \sum_{k=0}^{N-1} y_b^2(k), \quad (5.16)$$

onde $y_b(k)$ é o sinal num bloco de índice b , contendo N amostras do sinal de áudio sob processamento.

Os tamanhos adotados para os blocos e para a superposição entre eles influenciam bastante na deteção dos blocos corrompidos por pulsos longos, como mostra a Fig. 5.4. A utilização de blocos pequenos tende a aumentar a resolução temporal do cálculo da potência média, enfatizando os blocos corrompidos por pulsos longos. Naturalmente, o tamanho dos blocos deve ser feito maior que o tamanho típico dos pulsos longos.

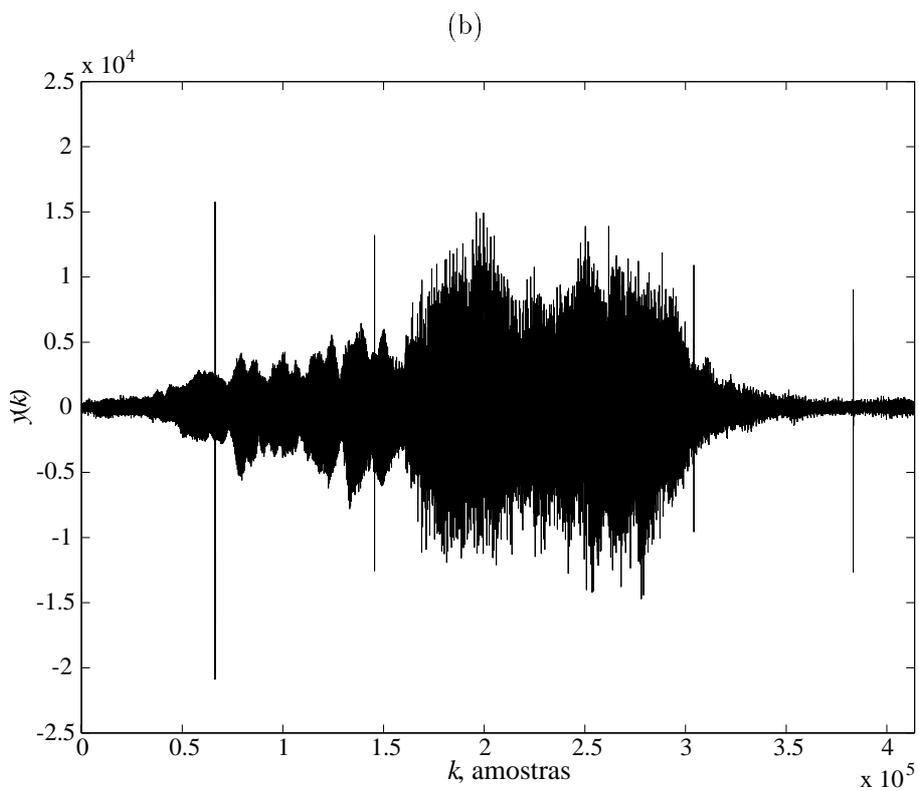
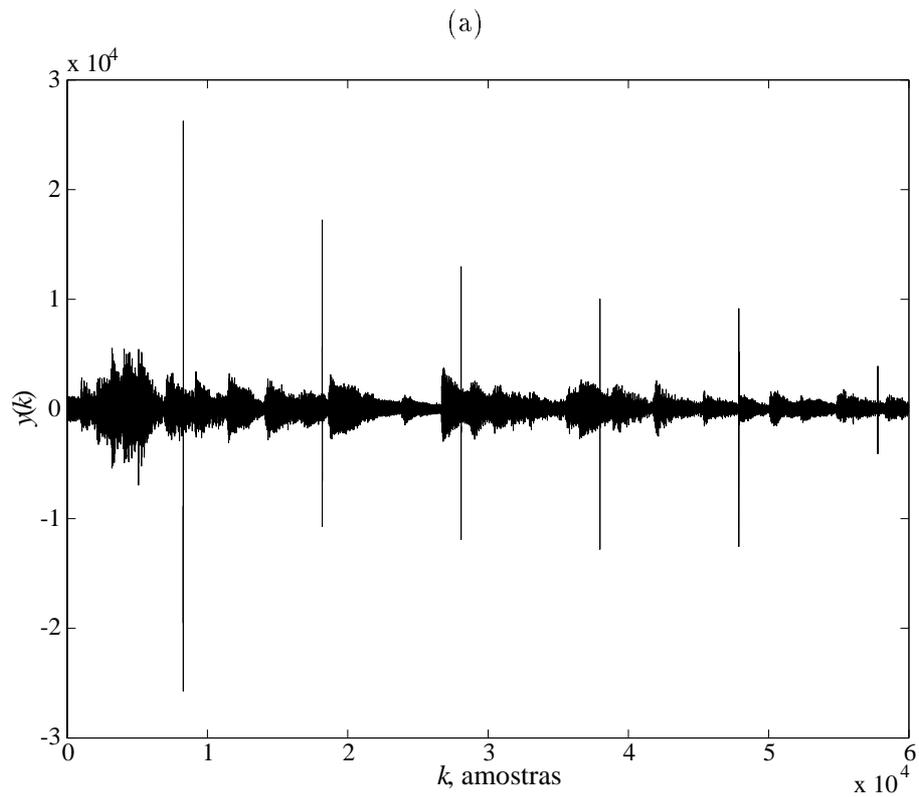


Figura 5.3: (a) Amplitude do Sinal A (dinâmica restrita). (b) Amplitude do Sinal B (dinâmica ampla).

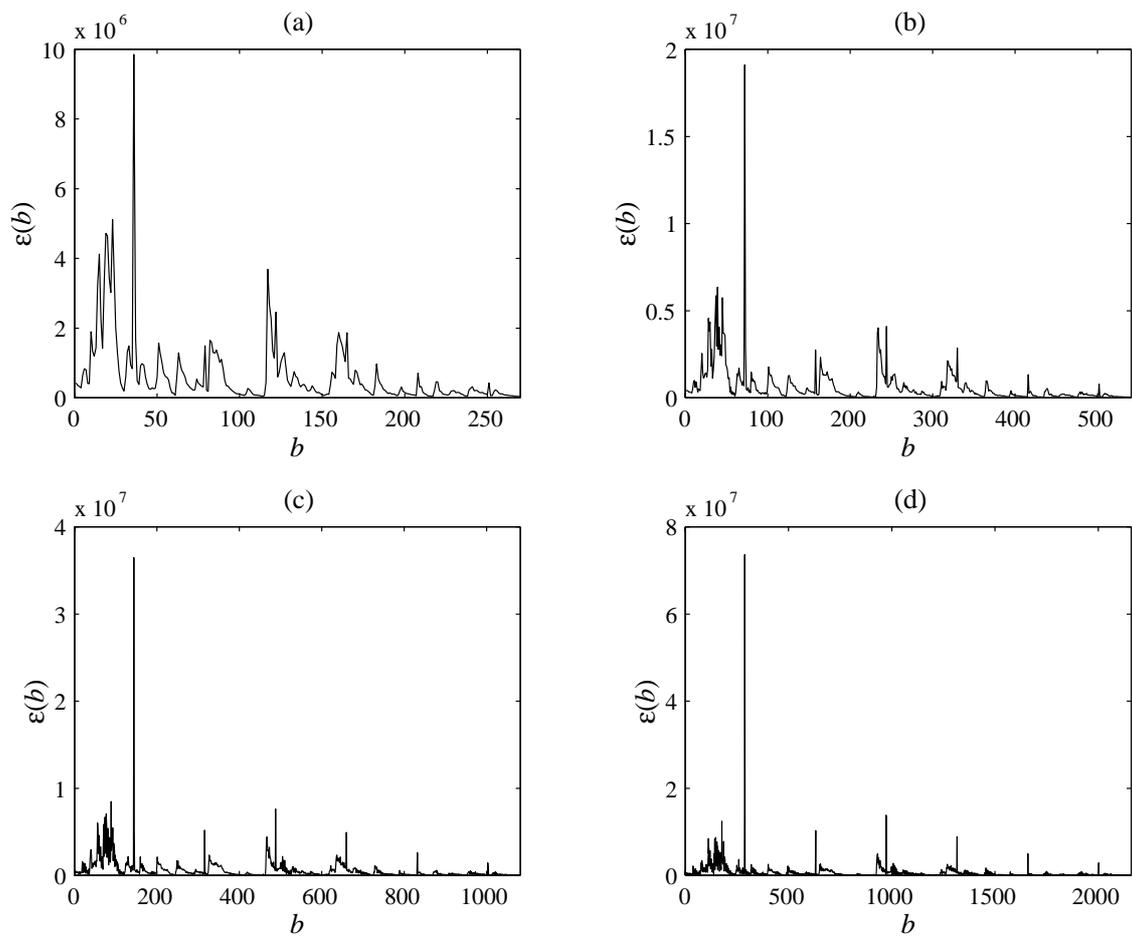


Figura 5.4: Potência média dos blocos do Sinal A: (a) Bloco de 2048 com superposição de 204 amostras. (b) Bloco de 1024 amostras com superposição de 102 amostras. (c) Bloco de 512 amostras com superposição de 51 amostras. (d) Bloco de 256 amostras com superposição de 25 amostras.

A Fig. 5.5 mostra o sinal de potência média associada ao Sinal B, particionado em blocos de 256 amostras, com superposição de 25 amostras.

Como observado no Método A, a dinâmica ampla do Sinal B também dificulta a seleção dos blocos com pulsos longos através da comparação da seqüência de potências médias dos blocos $\varepsilon(b)$ com um limiar. O Método C apresenta desempenho insatisfatório na detecção via limiar de blocos corrompidos por pulsos longos, mesmo em sinais com dinâmica restrita, como pode ser observado na Fig. 5.4.

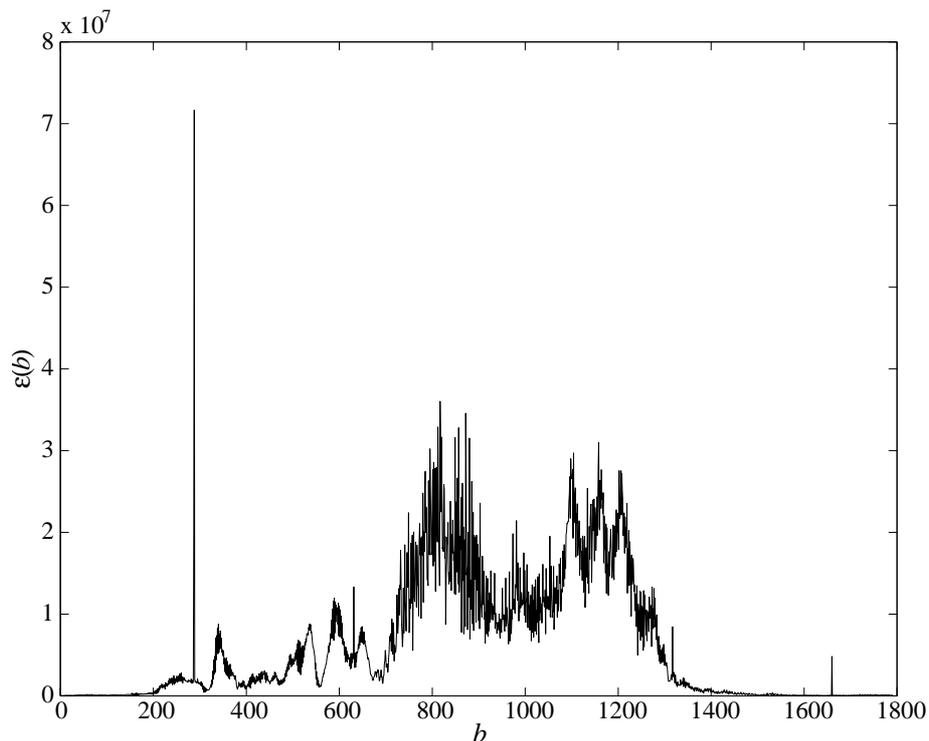


Figura 5.5: Potência média dos blocos do Sinal B: blocos de 256 amostras com superposição de 25 amostras.

5.4.3 Método C

O Método C se baseia nas características espectrais apresentadas pelos pulsos longos, principalmente, o fato de apresentarem energia significativa em uma faixa ampla do espectro, como pode ser observado no espectrograma mostrado na Fig. 5.2.

O método ora proposto consiste em calcular a Transformada Discreta de Fourier (DFT) [25] de cada bloco, selecionar uma faixa de frequências onde a informação dos pulsos longos esteja mais evidente; computar, nessa faixa, a média aritmética

do módulo dos coeficientes; comparar os valores dessas médias, obtidas para cada bloco, com um limiar arbitrariamente adotado; e selecionar, como corrompidos por pulsos longos os blocos cujos valores das médias freqüenciais estiverem acima do valor do limiar.

O Método C requer, além da escolha de um limiar — que pode ser feita diretamente por inspeção visual na seqüência de médias freqüenciais —, a escolha de uma faixa de freqüências na qual será calculada a média dos módulos dos coeficientes da DFT dos blocos.

A média aritmética do módulo dos coeficientes da DFT em uma determinada faixa é então calculada, em um dado bloco b , por

$$\mu(b) = \frac{1}{\beta - \alpha + 1} \sum_{n=\alpha}^{\beta} |Y_b(n)| \quad (5.17)$$

onde $Y_b(n)$ é a DFT do sinal discreto $y_b(k)$ de N amostras, dada por

$$Y_b(n) = \sum_{k=0}^{N-1} y_b(k) e^{-j \frac{2\pi kn}{N}}, \quad n = 0 \text{ a } N - 1 \quad (5.18)$$

e $0 \leq \alpha, \beta \leq \frac{N}{2}$ são os índices (em n) dos coeficientes correspondentes à faixa de freqüências desejada.

Por inspeção visual em espectrogramas de sinais de áudio pode-se constatar que existe uma concentração alta de informação espectral e energia na região de baixas e médias freqüências, até 3 kHz aproximadamente. Essa informação contrasta com a faixa espectral ampla exibida pelos pulsos longos. Deste modo, é conveniente destacar apenas uma faixa de freqüências em que a informação dos pulsos longos esteja mais presente. Testes experimentais mostraram que a adoção de uma faixa com freqüência inferior em torno de 3 kHz é bastante eficaz para acentuar o contraste entre os blocos com e sem contaminação por pulsos longos.

A Fig. 5.6 mostra $\mu(b)$ para o Sinal A em 4 faixas específicas do espectro, comparados com um limiar de valor 0,5. Como pode ser observado, em (c) e (d) os blocos corrompidos por pulsos longos ficam bastante destacados em relação aos demais, facilitando a seleção daqueles por comparação com limiar.

A utilização do Método C para seleção dos blocos corrompidos por pulsos longos no Sinal B é tão problemática quanto a experimentada com a utilização dos Métodos A e B. O problema reside no fato de o trecho forte do Sinal B ser composto

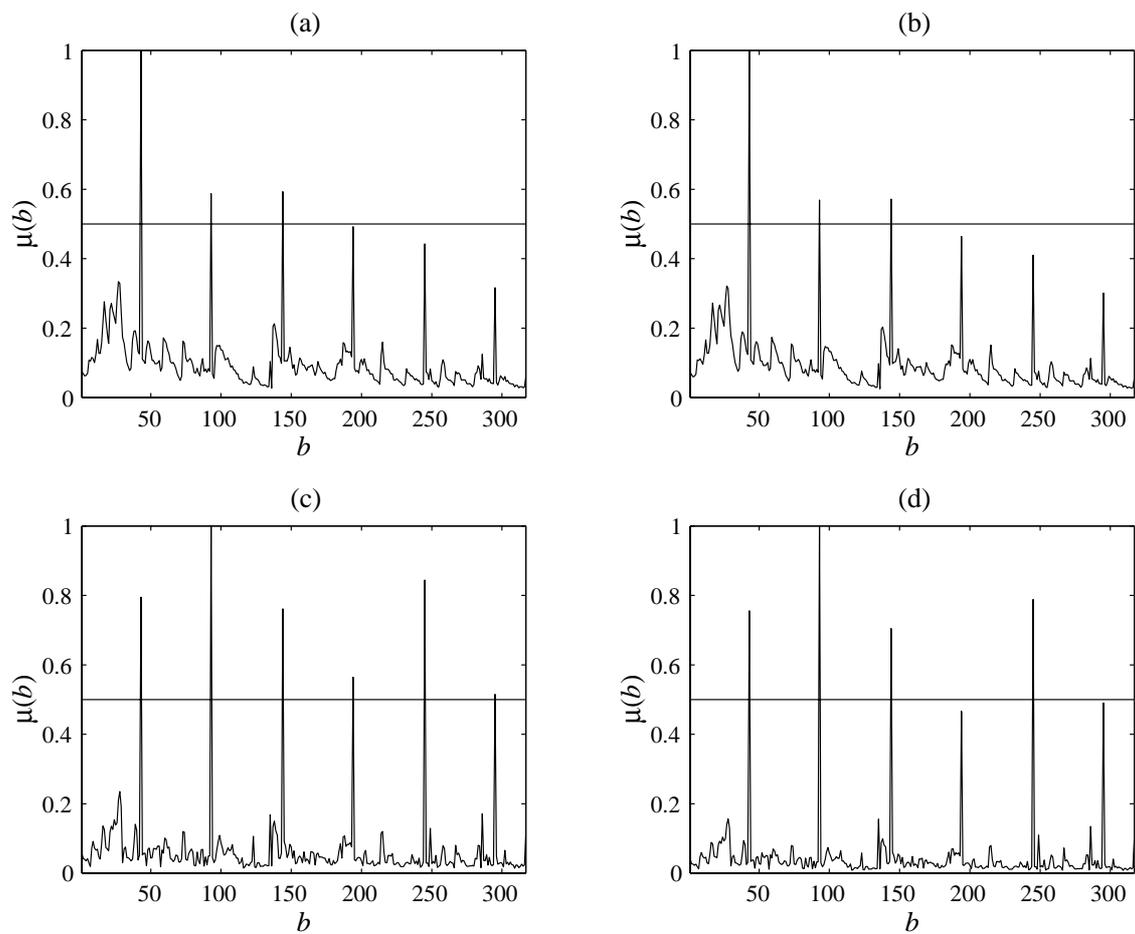


Figura 5.6: Sinal A: Comparação entre um limiar de valor 0,5 e $\mu(b)$ (normalizado) nas seguintes faixas de frequências: (a) 0 Hz a 22,05 kHz; (b) 0 Hz a 10 kHz; (c) 3 kHz a 22,05 kHz e (d) 3 kHz a 10 kHz. Foram utilizados, nesses casos, blocos e DFT's de 2048 amostras e superposição de 480 amostras entre os blocos.

de sons fortes de um naipe de metais que exibem harmônicos em altas frequências com energia significativa. Essa informação em altas frequências produz valores altos de μ nos blocos desses trechos, tornando difícil a seleção dos blocos corrompidos por pulsos longos via comparação com limiar.

A Fig. 5.7 mostra $\mu(b)$ para o Sinal B em 4 faixas específicas do espectro, comparados com um limiar de valor 0,5. Como pode ser observado, em todos os casos, é impossível selecionar os blocos corrompidos por pulsos longos por comparação com limiar, sem incorrer em um alto índice de falsos alarmes.

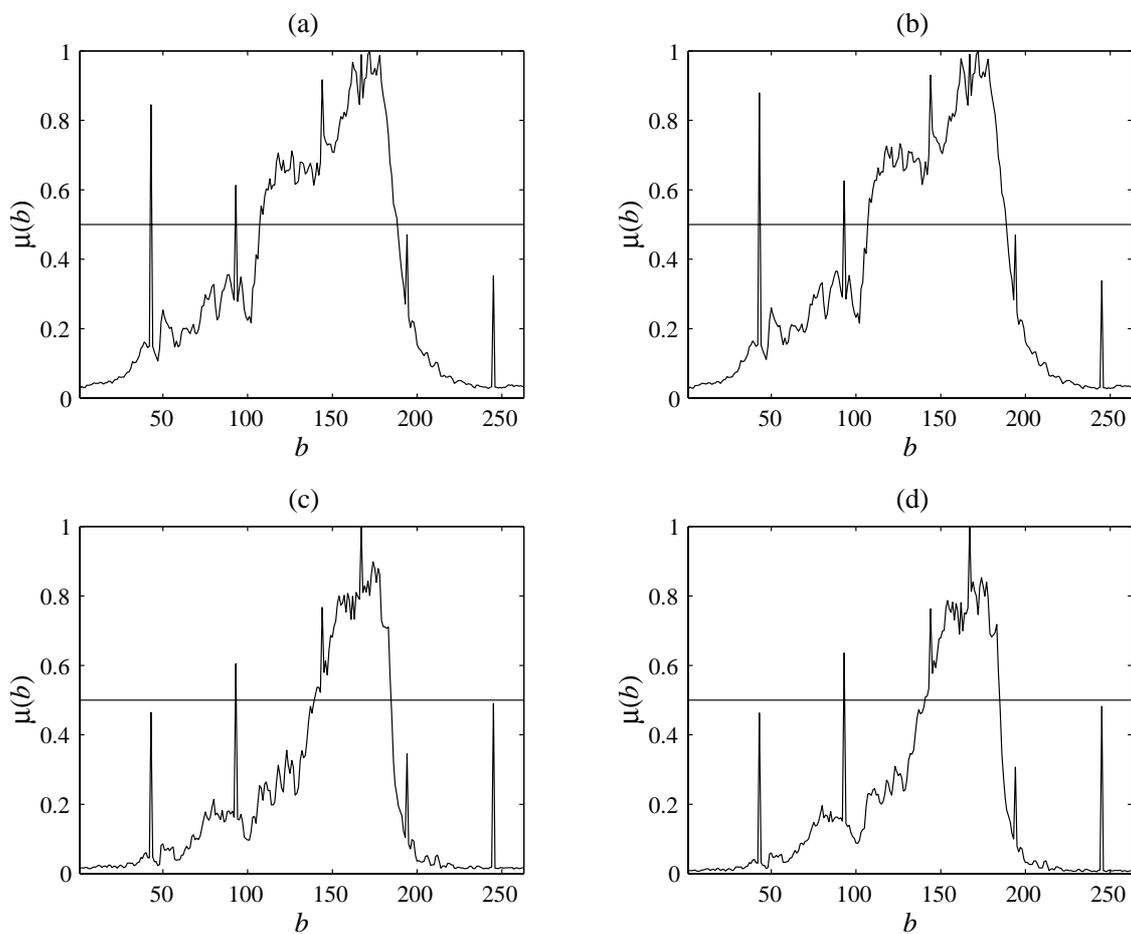


Figura 5.7: Sinal B: Comparação entre um limiar de valor 0,5 e $\mu(b)$ (normalizado) nas seguintes faixas de frequências: (a) 0 Hz a 22,05 kHz; (b) 0 Hz a 10 kHz; (c) 3 kHz a 22,05 kHz e (d) 3 kHz a 10 kHz. Foram utilizados, nesse caso, blocos e DFT's de 2048 amostras e superposição de 480 amostras entre os blocos.

Para contornar o problema de seleção de blocos com pulsos longos em sinais com dinâmica ampla e conteúdo espectral significativo em altas frequências, propõe-

se acrescentar ao Método C o seguinte procedimento:

- Obter $\mu(b)$ como descrito anteriormente;
- Obter um outro sinal $\mu_m(b)$, resultante da filtragem de $\mu(b)$ por um filtro por mediana com janela de tamanho ímpar igual a c amostras;
- Obter $\mu_{dif}(b) \triangleq \mu(b) - \mu_m(b)$;
- Comparar $\mu_{dif}(b)$ com um limiar λ arbitrariamente adotado e selecionar como corrompidos os blocos de índice b tais que $\mu_{dif}(b) \geq \lambda$.

A filtragem por mediana exibe a capacidade de eliminar ocorrências localizadas de valores atípicos em uma seqüência. Essa propriedade é útil, neste caso, pois a elevação dos valores de $\mu(b)$ devido às componentes significativas de alta freqüência do sinal é um fenômeno muito “menos localizado” do que a elevação de $\mu(b)$ devido aos pulsos longos.

Experimentalmente, verificou-se que a filtragem por mediana com $c = 5$ amostras produz resultados bastante satisfatórios, no que diz respeito ao aumento de contraste entre os blocos corrompidos e os blocos não-corrompidos, como pode ser observado na Fig. 5.8.

5.4.4 Método D

A idéia do Método D é realizar a seleção de blocos contaminados por pulsos longos comparando as potências médias dos sinais de excitação correspondentes a cada bloco com um limiar de valor arbitrado. A expectativa é que a ocorrência de valores elevados do sinal de excitação correspondentes aos pulsos longos seja capaz de elevar significativamente a potência do sinal de excitação dos blocos contaminados, destacando-os dos demais.

Para a obtenção do sinal de excitação de cada bloco é necessário estimar os coeficientes do modelo AR associado ao sinal do bloco e realizar uma filtragem inversa (ver Seção 2.2).

Obtido o sinal de excitação $e_b(k)$ de um dado bloco b de N amostras, sua potência pode ser calculada por

$$\sigma^2(b) = \frac{1}{N} \sum_{i=0}^{N-1} e_b^2(i). \quad (5.19)$$

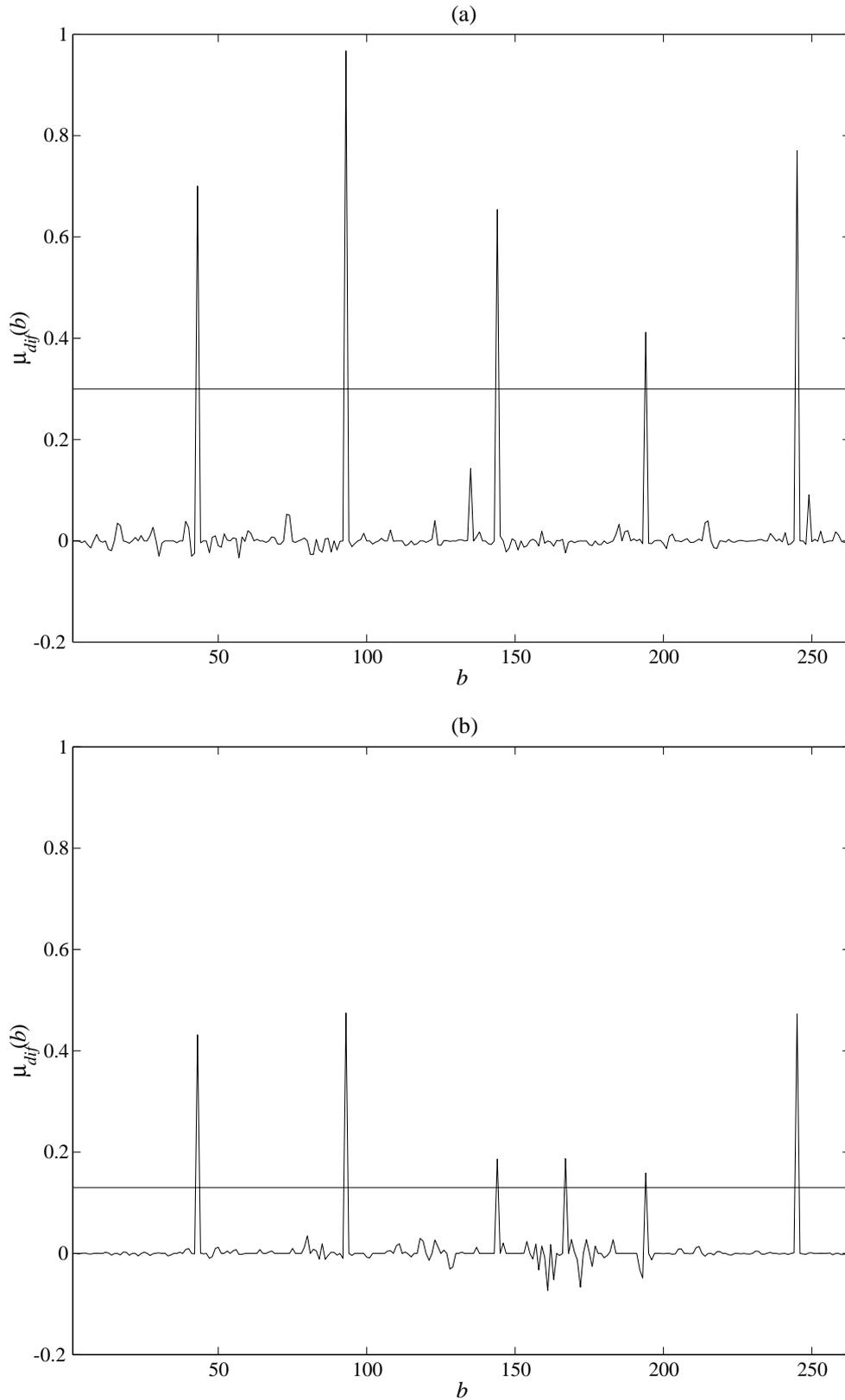


Figura 5.8: Comparação entre $\mu_{dif}(b)$ (normalizado) com limiar λ : (a) $\mu_{dif}(b)$ do Sinal A e $\lambda = 0,3$ e (b) $\mu_{dif}(b)$ do Sinal B e $\lambda = 0,13$. Em ambos os casos a faixa de freqüências utilizada foi de 3 kHz a 10 kHz; o tamanho dos blocos e das DFT's foi de 2048 amostras e o da superposição entre os blocos, de 480 amostras.

Na Fig. 5.9, são confrontadas as seqüências $\sigma^2(b)$ obtidas, respectivamente, para o Sinal A e o Sinal B com limiares arbitrariamente adotados. Como pode ser observado na Fig. 5.9(b), a seleção dos blocos corrompidos por pulsos longos no Sinal B é impossível de ser realizada pela comparação de $\sigma^2(b)$ com um limiar sem incorrer em altos índices de falsos alarmes. A elevação do valor de $\sigma^2(b)$ no trecho forte do Sinal B se deve, em parte, pela ocorrência do fenômeno de subpredição periódica associado a trechos fortemente tonais [13] (ver Apêndice A).

Naturalmente, o procedimento de filtragem por mediana do Método C pode ser aplicado à seqüência $\sigma^2(b)$ para resolver o problema de seleção via limiar de blocos corrompidos por pulsos longos no Sinal B. A Fig. 5.10 mostra a versão de $\sigma^2(b)$, obtida pela diferença entre a versão original e a filtrada por mediana, e a seqüência $\mu_{dif}(b)$, ambas comparadas com um limiar de valor 0,13.

5.4.5 Desempenho dos Métodos de Detecção de Pulsos Longos

A avaliação da eficiência dos 4 métodos, de acordo com o exposto anteriormente, aponta os Métodos C e D como os mais eficientes.

Como visto anteriormente, o desempenho do Método B na localização dos blocos corrompidos por pulsos longos em sinais com pouca dinâmica aumenta com o uso de blocos pequenos (em torno de 250 amostras). Esse tamanho pequeno de blocos acaba sendo um inconveniente, no sentido de que o tamanho dos blocos tipicamente usados nos processamentos de supressão de pulsos longos é da ordem de 2000 amostras e, desta forma, há necessidade de recalculer os índices dos blocos corrompidos de acordo com o tamanho de bloco adotado no processamento posterior. Já nos Métodos C e D, a utilização de blocos da ordem de 2000 amostras não é problemática.

Nos Métodos C e D, o fator determinante do sucesso na localização de blocos corrompidos em sinais com dinâmica ampla e conteúdo espectral rico é a utilização do procedimento auxiliar de filtragem por mediana. Nos testes comparativos entre os dois métodos, o tamanho dos blocos, a superposição entre eles e o tamanho da janela do filtro de mediana utilizados foram os mesmos, isto é, respectivamente, 2048, 480 e 5 amostras.

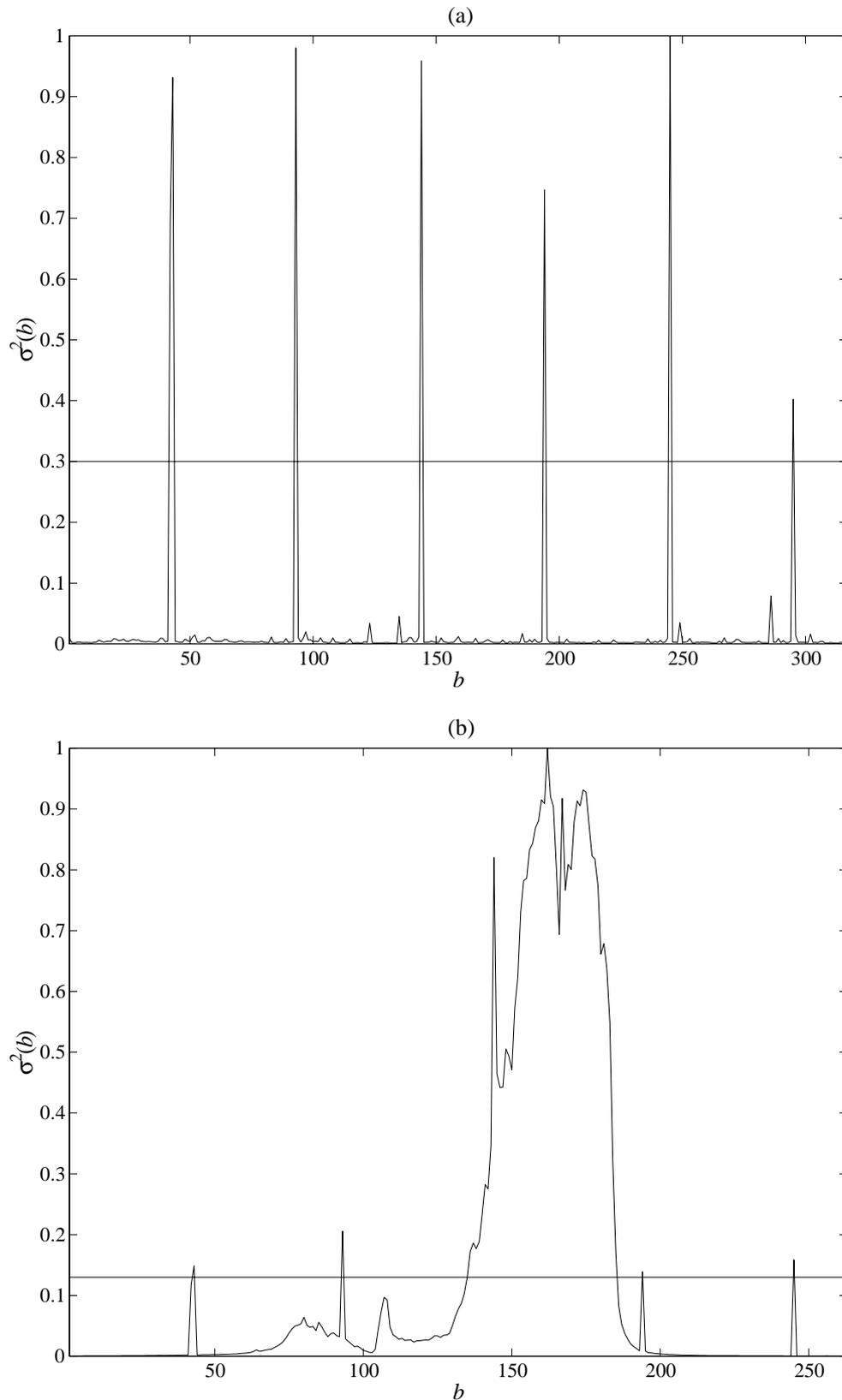


Figura 5.9: Comparação entre $\sigma^2(b)$ (normalizado) com limiar λ . (a) $\sigma^2(b)$ do Sinal A e $\lambda = 0,3$ e (b) $\sigma^2(b)$ do Sinal B e $\lambda = 0,13$. Em ambos os casos, os blocos foram modelados como processos AR de ordem 40; o tamanho dos blocos foi de 2048 amostras e o da superposição entre os blocos, de 480 amostras.

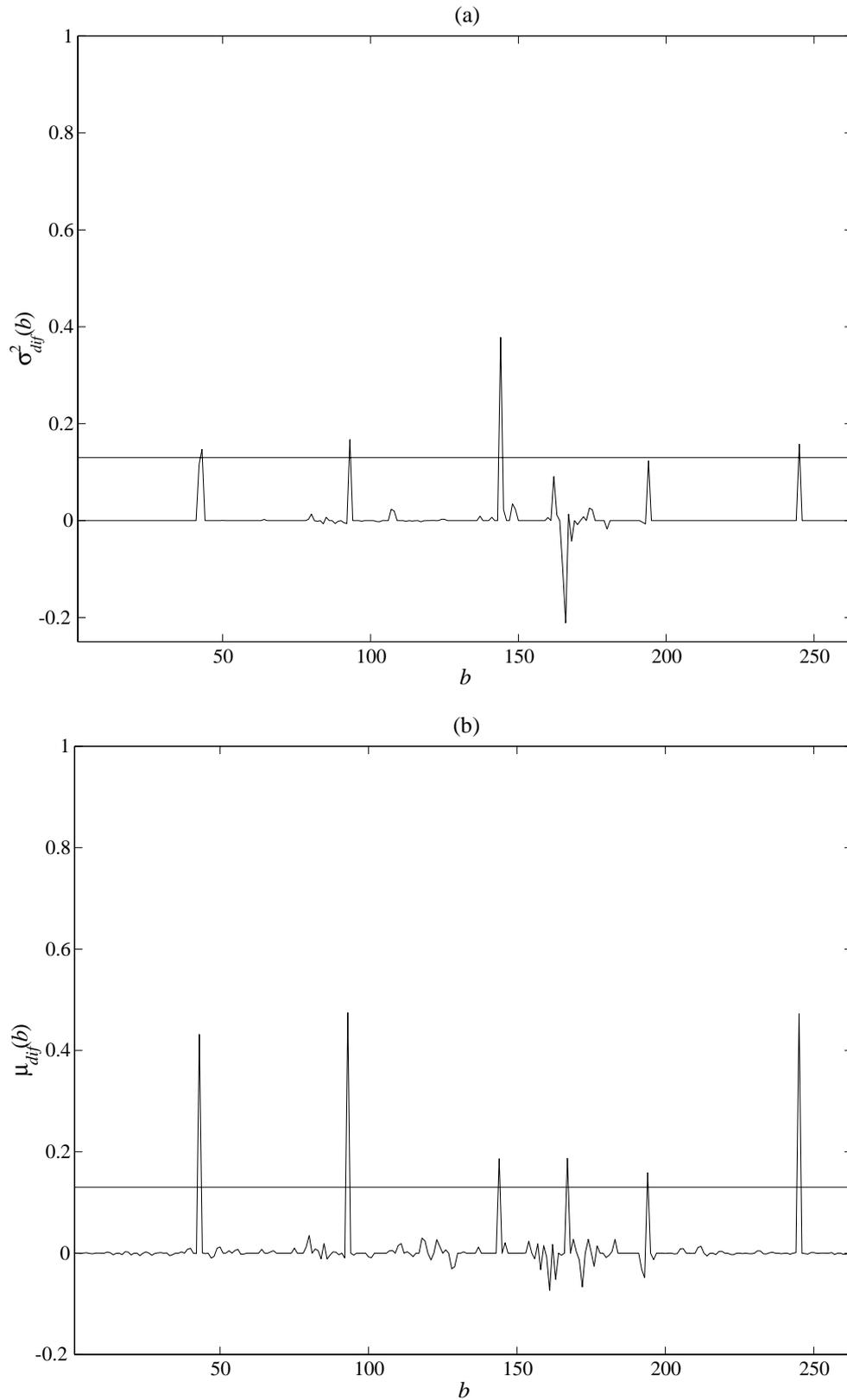


Figura 5.10: Comparação entre: (a) seqüência $\sigma_{dif}^2(b)$ (obtida pelo procedimento de filtragem por mediana do Método C aplicado à seqüência $\sigma^2(b)$) e (b) seqüência $\mu_{dif}(b)$ do Método C, ambas calculadas para o Sinal B. O limiar de referência, nos dois casos, vale 0,13.

No Método C, o espectro do bloco foi calculado através de DFT de 2048 pontos e a faixa do espectro selecionada foi entre 3 kHz e 10 kHz. Já no Método D, o sinal de cada bloco foi modelado como um processo AR de ordem 40 cujos coeficientes foram calculados por estimativa LS (*Least Squares*). Para essa configuração de processamento e lidando com sinais com dinâmica reduzida e conteúdo espectral pobre, os desempenhos dos Métodos C e D podem ser considerados equivalentes. Já para sinais com dinâmica ampla e conteúdo espectral rico em altas frequências, como o Sinal B, o Método C mostrou desempenho superior, como pode ser verificado pela Fig. 5.10.

O Método C ainda tem desempenho muito superior em relação ao tempo de processamento, quando comparado ao Método D. Enquanto no Método D é necessário calcular os coeficientes AR do bloco e realizar a filtragem inversa, no Método C basta computar a DFT do bloco. A computação de $\mu(b)$ e $\sigma^2(b)$ é igualmente custosa nos dois métodos, bem como a filtragem por mediana. Entretanto, a complexidade computacional para estimar os coeficientes AR de um bloco de N amostras é muito maior que a necessária para computar uma DFT. Para resolver (1.3) têm-se que computar, inicialmente, $\mathbf{X}^T \mathbf{X}$ e $\mathbf{X}^T \mathbf{x}_1$:

- $\mathbf{X}^T \mathbf{X}$ é simétrica de ordem p , portanto, só há necessidade de calcular $(p^2 + p)/2$ de seus elementos; só para calcular a sua primeira linha são necessárias $p(N - p)$ multiplicações;
- Para computar $\mathbf{X}^T \mathbf{x}_1$ são necessárias $N - p$ multiplicações;
- Para a solução do sistema propriamente dito são necessárias $p^3/4$ multiplicações [26].

O total subestimado de multiplicações no Método D, até imediatamente antes de se computar a filtragem inversa, é de $p^3/4 - p^2 + p(N - p) + N$, ou seja, para $p = 40$ e $N = 2048$ são necessárias 98328 multiplicações. A computação da DFT no Método C, por sua vez, pode ser realizada por um algoritmo rápido do tipo *Split-Radix*-FFT para sinais reais, que necessita de $N/2(\log_2(N) - 3) + 2$ multiplicações [27], ou seja, 8194 multiplicações, no caso de $N = 2048$. Isso significa que, para esse exemplo, o Método C é, muito subestimadamente, 12 vezes rápido que o Método D.

É interessante observar que, para nenhum dos métodos apresentados, foram feitas quaisquer considerações sobre a periodicidade de ocorrência dos pulsos longos, de modo que a ocorrência não periódica dos pulsos, como no Sinal B, não é limitante ao processamento. A similaridade dos pulsos também não é levada em consideração no processo de detecção dos blocos corrompidos, tornando os métodos aqui apresentados (principalmente os Métodos C e D) mais robustos que o método de detecção apresentado em [4], no que se refere à detecção de pulsos que apresentem diversidade na duração e no formato de transitório.

5.4.6 Pulso Longo na Interseção entre Blocos

A Seção 5.4 apresentou métodos de processamento em blocos para detecção de pulsos longos em sinais de áudio. Essa Seção lida com a ocorrência de pulsos longos nas extremidades dos blocos e descreve como a superposição entre blocos pode ser utilizada para evitar que tais ocorrências dificultem o processamento posterior de supressão dos pulsos.

Supondo, inicialmente, que não exista superposição entre os blocos, se um pulso longo se iniciar em um bloco e terminar no bloco subsequente, os métodos de detecção podem selecionar os dois blocos como corrompidos, o que de fato é verdade. Entretanto, o processamento de pulsos longos nas extremidades dos blocos pode ser bastante problemática. Por exemplo, na supressão de pulsos longos por separação de processos AR, a garantia de continuidade entre o trecho restaurado e o resto do sinal só é conseguida se um determinado número de amostras não-corrompidas antes e depois do distúrbio for levado em consideração nos cálculos [13].

Para evitar que pulsos longos localizados nas extremidades de um bloco contribuam para a seleção do bloco como corrompido, é suficiente adotar uma superposição de $2F$ amostras entre os blocos e “zerar” as F amostras do início e do final de cada bloco. Dessa forma, pulsos longos que se localizarem nas extremidades de um dado bloco serão mascarados pelo procedimento de zerar F amostras do início e do fim do bloco e passarão a ser considerados na região válida (não zerada) de um bloco adjacente. O valor de F está relacionado a uma estimativa do número de amostras do pulso longo de maior duração hipoteticamente presente no sinal e pode ser pensado como um limite máximo para a duração dos pulsos.

Naturalmente, não é descartada a hipótese de um pulso se iniciar na região válida do bloco e terminar na sua região zerada, ou vice-versa. Nesse caso, há possibilidade de dois blocos adjacentes serem selecionados como corrompidos por pulso longo; entretanto, a localização do pulso não vai ser na extremidade de nenhum dos dois blocos, a menos que F não seja uma boa estimativa da duração máxima dos pulsos.

Capítulo 6

Conclusões

Esta dissertação abordou algoritmos de processamento digital de sinais para a detecção e supressão de distúrbios localizados em sinais de áudio degradados. Os algoritmos apresentados realizam processamento em blocos de sinais de áudio, suficientemente curtos de modo que possam ser considerados estacionários. Aspectos da caracterização estatística dos sinais de áudio, como sua modelagem por processo autorregressivo foram abordadas no Capítulo 1. Nesse mesmo Capítulo, foram descritas as principais degradações que afetam os sinais de áudio provenientes de gravações históricas e aspectos associados à Psicoacústica.

Duas técnicas de detecção/supressão de ruído impulsivo apresentadas em [4, 13] foram revisadas no Capítulo 2. A técnica de detecção via filtragem inversa e limiar, tal como apresentada na literatura, apresenta limitações quanto à localização dos distúrbios, fato que pode resultar em restaurações qualitativamente não satisfatórias. Com o intuito de promover uma localização mais adequada dos distúrbios, são propostas modificações de caráter empírico na técnica de detecção por filtragem inversa e limiar. As modificações englobam: a forma de cálculo do limiar de detecção; a introdução de um segundo limiar, destinado exclusivamente à obtenção de uma estimativa mais apropriada da duração dos distúrbios; a adoção de critérios para a união de distúrbios adjacentes, destinados a atenuar os efeitos desastrosos da interferência destrutiva entre as amostras dos distúrbios na excitação; e a utilização de limiar de localização dependente da iteração do processamento de restauração realizada num mesmo bloco de sinal. As modificações propostas se mostram, efetivamente, capazes de aumentar o desempenho da etapa de detecção. Em comparação

com a etapa de detecção implementada de modo convencional, a etapa de detecção implementada com as modificações propostas apresenta uma significativa redução dos índices de não-deteção, o que é desejável, ao custo de uma esperada elevação do índice de falsa detecção, devido à tendência de localização dos distúrbios com durações ligeiramente superestimadas, em razão das modificações propostas. Essa tendência, entretanto, é pouco danosa à reconstrução do sinal, ao contrário do que ocorre quando as durações dos distúrbios são subestimadas. A comparação entre versões restauradas de sinais de áudio corrompidos por ruído impulsivo, utilizando detecção convencional e modificada corrobora as medidas quantitativas realizadas. Os sinais restaurados com a utilização da detecção modificada apresentam, perceptivamente, qualidade bastante superior aos restaurados com detecção convencional, a despeito das características variadas, tanto dos sinais de áudio avaliados, como dos ruídos impulsivos que os corrompiam.

A utilização de processamento em multitaxa visando à restauração de sinais de áudio corrompidos por ruído impulsivo foi abordada no Capítulo 4. No intuito de se adaptar o processamento utilizado em banda completa aos sinais de cada sub-banda, foram tecidas considerações sobre a estacionaridade e a modelagem estatística dos sinais em sub-bandas. A utilização de modelo AR para os sinais em sub-bandas foi avaliada através da análise das componentes polifásicas dos filtros equivalentes do banco de análise, tendo sido a eles incorporado o modelo AR do sinal de entrada, e pela análise das funções de autocorrelação das excitações dos sinais em sub-bandas, obtidas por filtragem inversa dos modelos AR para eles adotados. Apesar de a modelagem AR para os sinais em sub-bandas não poder ser considerada a mais adequada, sua utilização no sistema em multitaxa para restauração de sinais corrompidos por ruído impulsivo não se mostrou insatisfatória. A qualidade dos sinais restaurados em sub-bandas é similar à qualidade apresentada pelos restaurados em banda completa. Para sinais com conteúdo espectral reduzido em altas frequências, como é o caso de gravações históricas do início do século, pode-se, simplificar o processamento a ser realizado na sub-banda mais aguda, prescindindo da filtragem inversa para realizar a detecção dos distúrbios e interpolando de modo bem mais simplificado o sinal nessa sub-banda. Essa simplificações não afetam de modo significativo a qualidade final do sinal restaurado e ainda contribuem para

uma grande redução do tempo de processamento necessário para restaurar os sinais.

Retornando à linha de processamentos em blocos para detecção de distúrbios localizados, foram apresentadas, no Capítulo 5, técnicas visando à detecção e supressão de pulsos longos, que corrompem o sinal de áudio e que são decorrentes da resposta impulsiva do sistema reprodutor a arranhões ou quebras na superfície da matriz. Inicialmente, foram abordadas características de um pulso longo, no tempo e na frequência, e revisadas as técnicas de supressão por similaridade e por separação de processos AR. Em seguida, foram apresentados 4 métodos com processamento em blocos para detecção de pulsos longos: por amplitude (A), por potência média do sinal (B), por conteúdo espectral (C) e por potência média de excitação (D). O desempenho dos 4 métodos com relação à capacidade de detecção de pulsos longos em sinais com dinâmica restrita pode ser considerado equivalente, com exceção do Método B, cujo desempenho não se mostrou satisfatório. Já na detecção pulsos em sinais com dinâmica ampla e conteúdo espectral rico, o Método C, proposto neste trabalho, mostrou ser o mais eficaz, tanto pelo maior desempenho na detecção de pulsos longos quanto pela maior rapidez de processamento. A detecção em blocos de pulsos longos ainda pode ser vantajosa quando utilizada em conjunto com os métodos de supressão ou por similaridade ou por separação de processos AR. No primeiro caso, a etapa de detecção em blocos pode ser utilizada para selecionar previamente os trechos a serem investigados, sem a restrição de ocorrência periódica dos pulsos, contribuindo, assim, para a redução do custo computacional do método. No segundo caso, a associação de um método de detecção em blocos, antecedendo a etapa de separação, vem ao encontro da obtenção de um sistema de tratamento de pulsos longos mais automatizado.

A utilização de processamento de sinais em multirresolução e de técnicas estatísticas no domínio dos sinais ou coeficientes das sub-bandas tem sido bastante investigada. Como exemplo, pode-se citar a utilização de *Hidden Markov Models* no domínio da transformada *Wavelet*, que tem se mostrado útil para uma série de aplicações — incluindo estimação, detecção, classificação, predição e até mesmo síntese de sinais — que são de fundamental importância para a elaboração de novas técnicas visando à restauração de sinais de áudio.

Referências Bibliográficas

- [1] GODSILL, S. J., RAYNER, P. J. W., *Digital Audio Restoration*. London, UK, Springer-Verlag, 1998.
- [2] PAPOULIS, A., *Probability, Random Variables and Stochastic Processes*. 3 ed., New York, NY, USA, McGraw-Hill, Inc., 1991.
- [3] CAPPÉ, O., *Techniques de réduction de bruit pour la restauration d'enregistrements musicaux*. Thèse de doctorat, École Nationale Supérieure des Télécommunications, Paris, Sep. 1993.
- [4] VASEGHI, S. V., *Algorithms for Restoration of Archived Gramophone Recordings*. Ph.D. dissertation, University of Cambridge, Cambridge, UK, Feb. 1988.
- [5] GODSILL, S. J., RAYNER, P. J. W., CAPPÉ, O., “Digital Audio Restoration”. In: Brandenburg, K., Mahrs, M. (eds.), *Applications of Digital Signal Processing to Audio and Acoustics*, chapter 4, Norwell, MA, USA, Kluwer Academic Publishers, pp. 133–194, 1998.
- [6] SHANMUGAN, K. S., BREIPOHL, A. H., *Random Signals: Detection, Estimation and Data Analysis*. New York, NY, USA, John Wiley & Sons, Inc., 1988.
- [7] LEIPP, E., *Acoustique et Musique*. Paris, France, Masson et Cie, 1971.
- [8] JEFFRESS, L. A., “Masking”. In: Tobias, J. V. (ed.), *Foundations of Modern Auditory Theory*, chapter 3, New York, NY, USA, Academy Press, pp. 85–114, 1970.

- [9] SCHARF, B., “Critical Bands”. In: Tobias, J. V. (ed.), *Foundations of Modern Auditory Theory*, chapter 5, New York, NY, USA, Academy Press, pp. 157–202, 1970.
- [10] COIFMAN, R. R., DONOHO, D. L., “Translation-Invariant De-Noising”. In: Antoniadis, A. (ed.), *Wavelets and Statistics*, Springer-Verlag Lecture Notes, 1995.
- [11] GHAEEL, S., SAYEED, A. M., BARANIUK, R. G., “Improved Wavelet Denoising via Empirical Wiener Filtering”. In: *Proceedings of SPIE Wavelet Applications in Signal and Imaging Processing*, San Diego, Jul. 1997.
- [12] EARGLE, J., *Sound Recording*. New York, NY, USA, Van Nostrand Reinhold, 1980.
- [13] GODSILL, S. J., *The Restoration of Degraded Audio Signals*. Ph.D. dissertation, University of Cambridge, Cambridge, UK, Dec. 1993.
- [14] VELDHUIS, R., *Restoration of Lost Samples in Digital Signals*. Prentice-Hall, 1990.
- [15] ESQUEF, P. A. A., *Implementação e Investigação Comparativa de Técnicas para Restauração Digital de Áudio*. Projeto final de graduação, UFRJ – EE/DEL, Rio de Janeiro, RJ, Brasil, Fev. 1997.
- [16] DONOHO, D. L., *Wavelet Shrinkage and W.V.D.: a 10-Minute Tour*, Report 416, Department of Statistics, Stanford University, Jun. 1993.
- [17] DONOHO, D. L., “De-noising by soft-thresholding”, *IEEE Transactions on Information Theory*, v. 41, n. 3, pp. 613–627, 1995.
- [18] DONOHO, D. L., JOHNSTONE, I. M., *Ideal Denoising in Orthonormal Basis Chosen from a Library of Bases*, Report 461, Department of Statistics, Stanford University, Sep. 1994.
- [19] BERGER, J., JOHNNATHAN, COIFMAN, R. R., “Removing Noise from Music Using Local Trigonometric Bases and Wavelets Packets”, *Journal of the Audio Engineering Society*, v. 42, pp. 808–818, Oct. 1994.

- [20] CHENG, C., *Wavelet Signal Processing of Digital Audio with Applications in Electro-Acoustic Music*. M.Sc. dissertation, Dartmouth College, Hanover, New Hampshire, 1996.
- [21] SATHE, V. P., VAIDYANATHAN, P. P., “Effects of Multirate Systems on the Statistical Properties of Random Signals”, *IEEE Transactions on Signal Processing*, v. 41, n. 1, 1993.
- [22] CROCHIERE, R. E., RABINER, L. R., *Multirate Digital Signal Processing*. Englewood Cliffs, Prentice-Hall, Inc., 1993.
- [23] VAIDYANATHAN, P. P., *Multirate Systems and Filter Banks*. Englewood Cliffs, Prentice-Hall, Inc., 1993.
- [24] VAIDYANATHAN, P. P., LIU, V. C., “Classical Sampling Theorems in the Context of Multirate and Polyphase Digital Filter Bank Structures”, *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. 36, n. 9, 1988.
- [25] OPPENHEIM, A., SCHAFER, R., *Discrete-Time Signal Processing*. Englewood Cliffs, Prentice-Hall, Inc., 1989.
- [26] PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. *et al.*, *Numerical Recipes in C : The Art of Scientific Computing*. 2 ed., New York, NY, USA, Cambridge University Press, 1992.
- [27] MALVAR, H. S., *Signal Processing with Lapped Transforms*. Norwood, MA, USA, Artech House, 1992.
- [28] Ó RUANAIDH, J. J. K., FITZGERALD, W. J., *The Restoration of Digital Audio Recordings Using the Gibbs Sampler*, Technical Report CUED/F-INFENG/TR 153, University of Cambridge – Department of Engineering, Cambridge, UK, May 1994.
- [29] Ó RUANAIDH, J. J. K., *Numerical Bayesian Methods Applied to Signal Processing*. Ph.D. dissertation, University of Cambridge, Cambridge, UK, Oct. 1994.

- [30] VASEGHI, S. V., RAYNER, P. J., “Detection and Supression of Impulsive Noise in Speech Communication Systems”, *IEE Proceedings, Part I*, v. 137, n. 1, pp. 38–47, Feb. 1990.

Apêndice A

Interpolação LSAR

A.1 Descrição do Algoritmo

O algoritmo *Least Squares Autorregressive* (LSAR) é voltado à interpolação de sinais discretos que possam ser modelados por um processo autorregressivo (AR).

Seja um sinal de N amostras, $x(k)$, modelável por um processo AR de ordem p . Cada amostra do sinal nos instantes $k = p, p + 1, \dots, N - 1$ é dada, tal como em (1.1), por

$$x(k) = \sum_{i=1}^p a(i)x(k-i) + e(k). \quad (\text{A.1})$$

Vetorialmente, pode-se escrever

$$\mathbf{e} = \mathbf{A}\mathbf{x}, \quad (\text{A.2})$$

onde

$$\mathbf{A} = \begin{bmatrix} -a(p) & \cdots & -a(1) & 1 & 0 & \cdots & 0 \\ 0 & -a(p) & \cdots & -a(1) & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \cdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -a(p) & \cdots & -a(1) & 1 \end{bmatrix}, \quad (\text{A.3})$$

e \mathbf{x} e \mathbf{e} são vetores coluna contendo blocos de N amostras, respectivamente de $x(k)$ e $e(k)$. Suponha, agora, que um distúrbio impulsivo corrompa l amostras de \mathbf{x} e seu início se dê na amostra m , ($p + 1 < m < N - l$). O vetor \mathbf{x} pode, então, ser

subdividido em seções de amostras não-corrompidas ou conhecidas, \mathbf{x}_c , e amostras degradadas ou desconhecidas, \mathbf{x}_d ; chamando de \mathbf{x}_{ca} e \mathbf{x}_{cb} os vetores contendo as amostras de \mathbf{x}_c , respectivamente, antes e depois do distúrbio, \mathbf{x} pode ser reescrito como

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_{ca}^T & \mathbf{x}_d^T & \mathbf{x}_{cb}^T \end{bmatrix}^T, \quad (\text{A.4})$$

e (A.2) se torna

$$\mathbf{e} = \begin{bmatrix} \mathbf{A}_{ca} & \mathbf{A}_d & \mathbf{A}_{cb} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{ca}^T \\ \mathbf{x}_d^T \\ \mathbf{x}_{cb}^T \end{bmatrix}, \quad (\text{A.5})$$

onde \mathbf{A} foi particionada em colunas mantendo correspondência com a partição de \mathbf{x} em (A.4).

Efetuando a multiplicação vetorial de (A.5) e agrupando as seções conhecidas de \mathbf{x} de modo que $\mathbf{x}_c^T = \begin{bmatrix} \mathbf{x}_{ca}^T & \mathbf{x}_{cb}^T \end{bmatrix}$ e $\mathbf{A}_c = \begin{bmatrix} \mathbf{A}_{ca} & \mathbf{A}_{cb} \end{bmatrix}$, tem-se

$$\mathbf{e} = \mathbf{A}_c \mathbf{x}_c + \mathbf{A}_d \mathbf{x}_d \quad (\text{A.6})$$

separado em componentes conhecidas e desconhecidas.

A solução LS é obtida pela minimização da função custo $E = \mathbf{e}^T \mathbf{e}$ com relação ao vetor de amostras desconhecidas \mathbf{x}_d , que leva a

$$\mathbf{x}_d^{\text{LS}} = -(\mathbf{A}_d^T \mathbf{A}_d)^{-1} \mathbf{A}_d^T \mathbf{A}_c \mathbf{x}_c. \quad (\text{A.7})$$

A extensão do procedimento LSAR quando houver mais de um distúrbio impulsivo num bloco é direta e, na maioria dos casos, a interpolação LSAR mostra bom desempenho, pois a minimização da energia de excitação, leva à escolha de amostras desconhecidas com a maior probabilidade segundo o modelo adotado.

A.2 Limitações do Interpolador LSAR

A interpolação LSAR apresenta limitações evidentes em duas situações: na interpolação de falhas de longa duração em sinais que apresentam, naturalmente,

uma parcela com características típicas de ruído branco (amostras pouco correlacionadas), como, por exemplo, trechos fala não-vozeada ou alguns sons de bateria, e na interpolação de sinais com conteúdo fortemente tonal. No primeiro caso, a estimativa LS pode produzir um sinal interpolado muito suave, com amostras fortemente correlacionadas, diferindo da característica original do sinal, principalmente, no meio de uma falha longa, onde, em geral, o erro de predição assume valores pequenos, como mostra a Fig. A.1. Neste exemplo particular, é mostrado um sinal de áudio real, extraído de uma ária de ópera, exatamente em um trecho em que a soprano emite um som não-vozeado, sendo acompanhada por um som suave produzido pelo naipe de cordas.

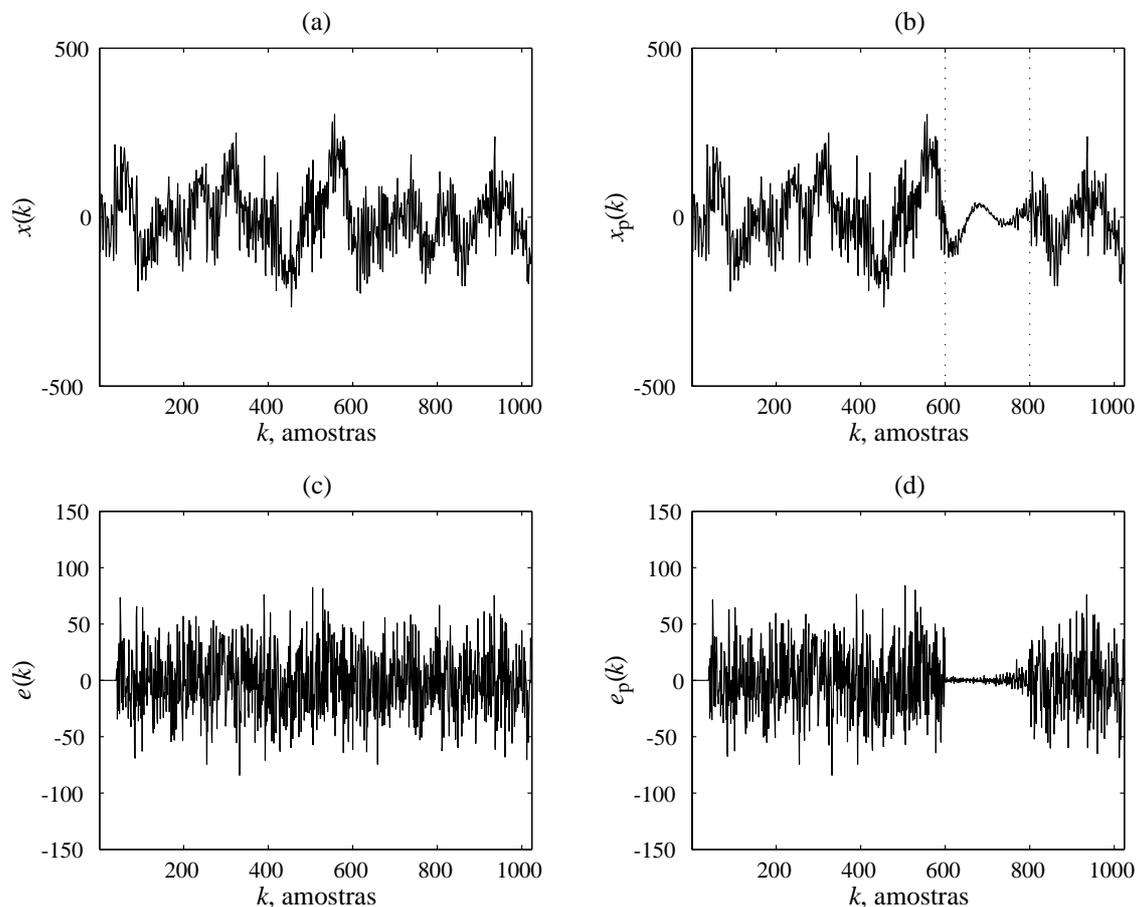


Figura A.1: (a) Trecho de sinal de áudio composto por um som suave de cordas e por um som de fala não-vozeada. (b) Versão reconstruída do sinal por interpolação LSAR de ordem 40 — falha simulada entre as amostras 600 e 800. (c) Excitação correspondente ao sinal original. (d) Excitação correspondente ao sinal interpolado.

No segundo caso, o sinal de excitação típico dos sinais com conteúdo for-

temente tonal, obtido por filtragem inversa do processo AR, exibe subpredições (valores altos de $e(k)$) periódicas, que são associadas à frequência fundamental do sinal ou *pitch*, como mostra a Fig. A.2.

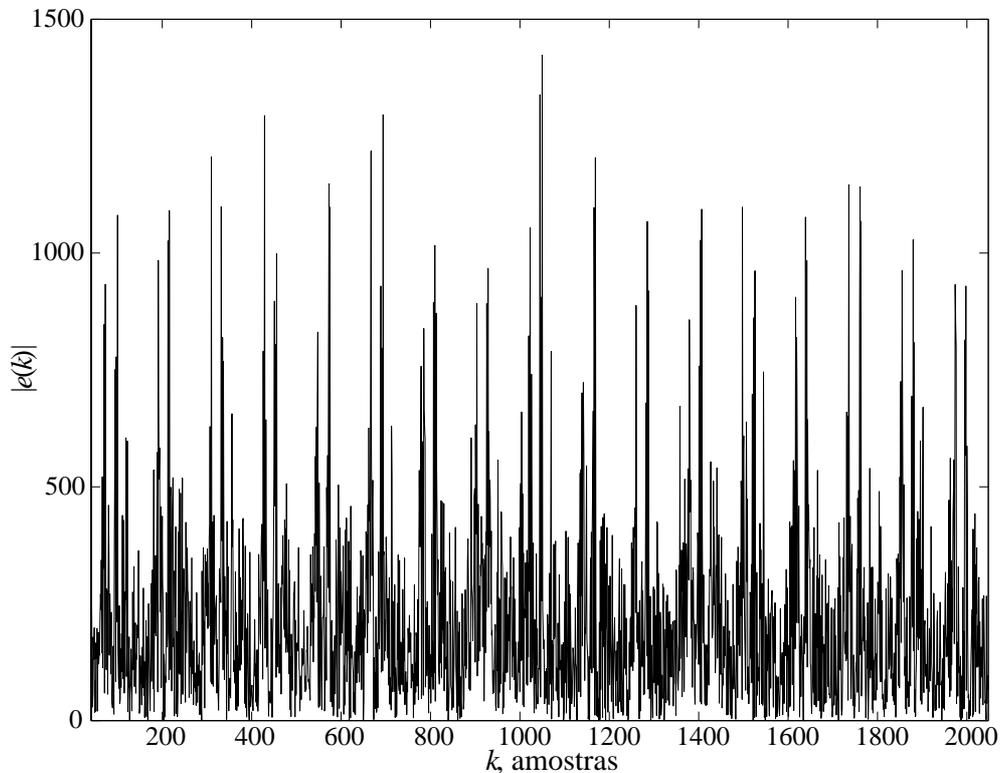


Figura A.2: Fenômeno de subpredição na excitação correspondente a um bloco de sinal com conteúdo fortemente tonal. A ordem do modelo AR utilizada na filtragem inversa foi 40.

No procedimento de detecção por filtragem inversa, essas subpredições tendem a ser detectadas indevidamente como ruído impulsivo. Na reconstrução, a minimização da energia da excitação produzida pelo interpolador LSAR vai de encontro à característica original da excitação, produzindo reconstruções inadequadas.

A.3 Interpoladores Alternativos

Métodos alternativos foram propostos no intuito de resolver os problemas mencionados anteriormente. No problema da minimização da energia da excitação, a solução consiste construir intepoladores que mantenham a energia da excitação coerente com a existente no entorno dos trechos a serem interpolados. Em [5] é apre-

sentado um interpolador que mantém a energia da excitação constante nos trechos interpolados. Outra alternativa é o interpolador baseado em *Gibbs Sampling* (GS) [28, 29], que estima as amostras desconhecidas do sinal tomando amostras de uma aproximação da função densidade de probabilidade *a posteriori* $f_{\mathbf{x}_d, \mathbf{a}, \sigma_e | \mathbf{x}_c}(\mathbf{x}_d, \mathbf{a}, \sigma_e | \mathbf{x}_c)$. O método consiste em estimar tal distribuição por um procedimento iterativo que envolve a amostragem das funções densidades de probabilidade condicionais de \mathbf{x}_d , \mathbf{a} e σ_e^2 — $f_{\mathbf{x}_d | \mathbf{a}, \sigma_e, \mathbf{x}_c}(\mathbf{x}_d | \mathbf{a}, \sigma_e, \mathbf{x}_c)$, $f_{\mathbf{a} | \mathbf{x}_d, \sigma_e, \mathbf{x}_c}(\mathbf{a} | \mathbf{x}_d, \sigma_e, \mathbf{x}_c)$ e $f_{\sigma_e^2 | \mathbf{a}, \mathbf{x}_d, \mathbf{x}_c}(\sigma_e^2 | \mathbf{a}, \mathbf{x}_d, \mathbf{x}_c)$ —, uma por vez, com as outras variáveis fixadas em seus mais recentes valores estimados. Após a convergência, a distribuição dessas amostras tende à distribuição conjunta $f_{\mathbf{x}_d, \mathbf{a}, \sigma_e | \mathbf{x}_c}(\mathbf{x}_d, \mathbf{a}, \sigma_e | \mathbf{x}_c)$, que descreve integralmente o relacionamento entre as variáveis do problema.

Em [30] é proposto um modelo AR estendido para o sinal de áudio, como forma de minimizar o problema de subpredição em trechos fortemente tonais. Esse modelo introduz parâmetros adicionais de predição no entorno do período de *pitch*, T , do sinal em questão. Desse modo, o sinal de áudio não-corrompido, antes descrito por (1.1), fica agora descrito como

$$x(k) = \sum_{i=1}^p x(k-i)a(i) + \sum_{j=-Q}^Q x(k-T-j)b(j) + e(k), \quad (\text{A.8})$$

onde Q é tipicamente menor que p . Naturalmente, o modelo exige uma estimativa de T e a estimação dos parâmetros adicionais $b(j)$, que pode ser obtida estendendo o método descrito na Seção 1.3 para o modelo estendido de (A.8) [1].

Apêndice B

Fundamentos de Processamento em Multitaxa

B.1 Introdução

Nesta seção são abordados conceitos básicos referentes ao processamento em multitaxa, bem como alguns princípios básicos sobre os bancos de *Quadrature Mirror Filters* (QMF's) e *Conjugate Quadrature Filters* (CQF's). As operações de decimação e expansão são descritas na seção B.2 e as identidades nobres, na seção B.3. Na seção B.4 são descritos os fundamentos de projeto dos bancos de QMF's e CQF's.

B.2 Decimação e Expansão

O Critério de Nyquist de Amostragem afirma que a menor taxa com que um sinal analógico com espectro de largura W pode ser amostrado sem perda de informação é

$$f_s = 2W. \quad (\text{B.1})$$

Supondo que um sinal discreto foi dividido em sub-bandas, para cada uma se poderia pensar numa redução proporcional da taxa de amostragem. Os processamentos subseqüentes seriam feitos em taxas mais baixas, com evidente economia de processamento, e ao seu final, se regeneraria a taxa de amostragem original. A redução e o aumento da taxa de amostragem são realizados, respectivamente, pelas operações de decimação e de expansão.

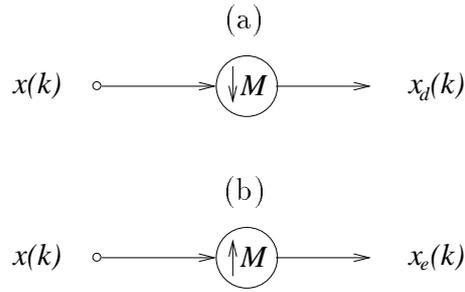


Figura B.1: (a) Decimador por M . (b) Expansor por M

O Decimador por M mostrado na Fig. B.1 (a) opera no tempo conforme

$$x_d(k) = x(Mk), \quad (\text{B.2})$$

o que corresponde, em z , a

$$X_d(z) = \frac{1}{M} \sum_{m=0}^{M-1} X(z^{\frac{1}{M}} e^{-\frac{2\pi m}{M}}). \quad (\text{B.3})$$

O Expansor por M mostrado na Fig. B.1 (b) opera no tempo conforme

$$x_e(k) = \begin{cases} x(\frac{k}{M}) & , k \text{ múltiplo de } M \\ 0 & , \text{ demais } k \end{cases}, \quad (\text{B.4})$$

o que corresponde, em z , a

$$X_e(z) = X(z^M). \quad (\text{B.5})$$

Como se pode observar das expressões em z , se um sinal passa por um Banco de Filtros de Análise e é decimado, sua simples expansão não reconstitui diretamente as componentes aditivas que o regenerariam; é necessário, após a expansão, um Banco de Filtros de Síntese adequado.

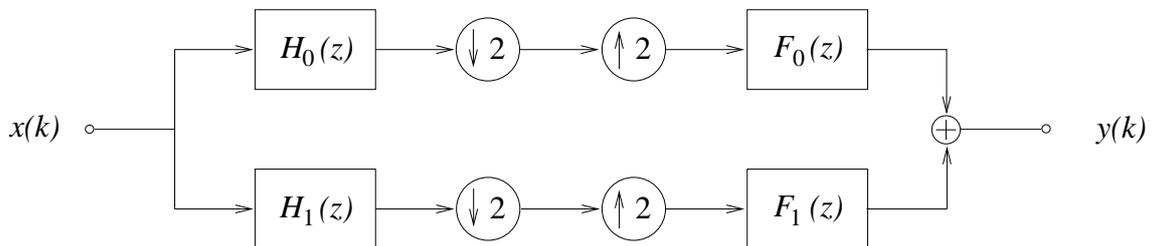


Figura B.2: Estrutura básica de bancos de filtros com 2 sub-bandas.

Examinando a subdivisão em 2 sub-bandas, como mostrado na Fig. B.2, pode-se expressar, em z , a saída do sistema em função da entrada como

$$Y(z) = T(z)X(z) + A(-z)X(-z), \quad (\text{B.6})$$

sendo $T(z)$ a função de Distorção e $A(z)$ a função de *Aliasing* (ligada à repetição do espectro), dadas por

$$T(z) = \frac{1}{2}[F_0(z)H_0(z) + F_1(z)H_1(z)] \quad (\text{B.7})$$

e

$$A(z) = \frac{1}{2}[F_0(z)H_0(-z) + F_1(z)H_1(-z)]. \quad (\text{B.8})$$

Para eliminação do *aliasing*, deve-se ter

$$A(z) = 0, \quad (\text{B.9})$$

e, para eliminação da distorção, deve-se ter

$$T(z) = z^{-k_0}, \quad (\text{B.10})$$

sendo k_0 o número de amostras de atraso. Os bancos de filtros que atendem às 2 condições são ditos com Reconstrução Perfeita.

Os bancos para divisão em 2 sub-bandas, quando utilizados numa implementação modular em árvore binária, na qual as taxas de amostragem vão sendo reduzidas à metade, permitem separar o espectro em 2^D sub-bandas iguais, sendo D o número de níveis de decimação. Dessa forma, uma divisão em 4 sub-bandas iguais, por exemplo, como mostra a Fig. B.3 requer 2 níveis simétricos de decimação/expansão por 2. Já em 4 sub-bandas em oitavas, como mostra a Fig. B.4, a divisão é realizada de modo assimétrico: inicialmente, divide-se o espectro em duas partes e, depois, promovem-se mais duas subdivisões na primeira metade do espectro anteriormente dividido; assim, as 2 sub-bandas de frequências inferiores passam por 3 níveis de decimação/expansão, a seguinte por 2, níveis e a última, por 1 nível.

B.3 Identidades Nobres

As identidades nobres, mostradas graficamente na Fig. B.5, realizam papel importante na manipulação dos principais elementos que compõem uma estrutura em multitaxa (atrasos, filtros, decimadores e expansores). São elas:

Comutação entre Expansor e Decimador: Expansão por M e Decimação por N podem ser comutadas se e somente se M e N forem primos entre si.

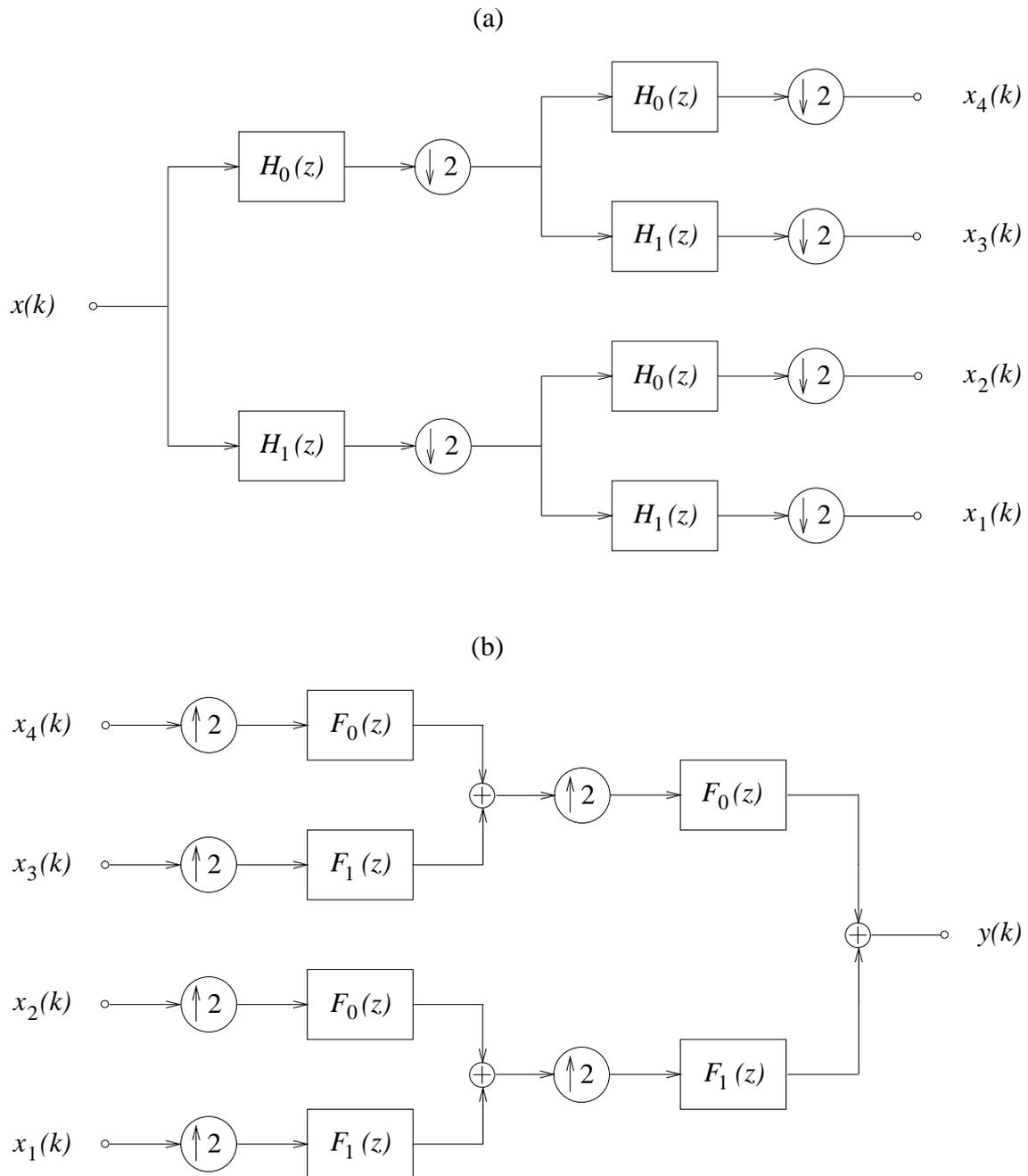


Figura B.3: (a) Banco de análise com estrutura em árvore binária para 4 sub-bandas iguais. (b) Banco de síntese com estrutura em árvore binária para 4 sub-bandas iguais.

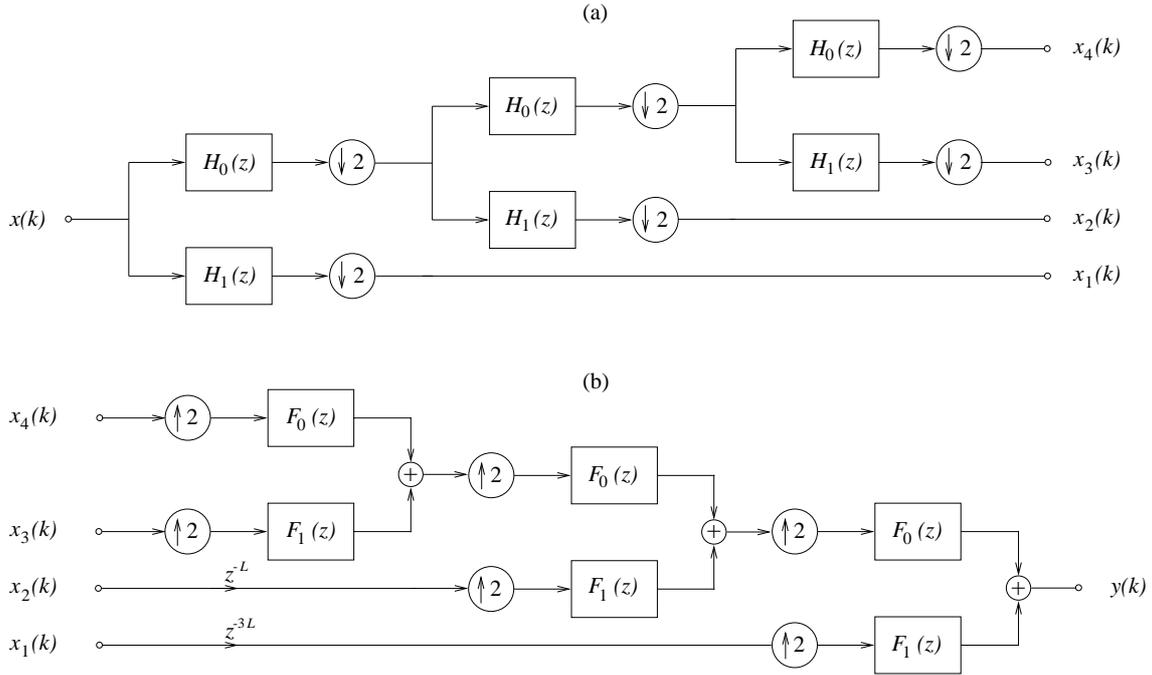


Figura B.4: (a) Banco de análise com estrutura em árvore binária para 4 sub-bandas em oitavas. (b) Banco de síntese com estrutura em árvore binária para 4 sub-bandas em oitavas.

Por (B.3) e (B.5) tem-se que uma expansão por M seguida por uma decimação por N resulta em

$$Y_{e/d}(z) = \sum_{k=0}^{N-1} X(z^{\frac{M}{N}} e^{j\frac{2\pi k}{N}}), \quad (\text{B.11})$$

enquanto que revertendo a ordem entre expansor e decimador obtém-se

$$Y_{d/e}(z) = \sum_{k=0}^{N-1} X(z^{\frac{M}{N}} e^{j\frac{2\pi k M}{N}}). \quad (\text{B.12})$$

A igualdade entre (B.11) e (B.12) ocorre se e somente se M e N forem primos entre si.

Comutação entre Filtro e Decimador: Decimação por M seguida de filtragem por um filtro $H(z)$ é equivalente a realizar uma filtragem por um filtro $H(z^M)$ seguida de decimação por M .

Por (B.3) tem-se que a decimação de um sinal $X(z)$ previamente filtrado por $H(z^M)$ resulta em

$$\sum_{k=0}^{M-1} X(z^{\frac{1}{M}} e^{j\frac{2\pi k}{M}}) H((z^{\frac{1}{M}} e^{j\frac{2\pi k}{M}})^M) = H(z) \sum_{k=0}^{M-1} X(z^{\frac{1}{M}} e^{j\frac{2\pi k}{M}}), \quad (\text{B.13})$$

que é igual a filtrar por $H(z)$ uma versão de $X(z)$ decimada por M .

Comutação entre Filtro e Expansor: Filtragem por $H(z)$ seguida por expansão por M é equivalente a realizar uma expansão por M seguida de uma filtragem por $H(z^M)$.

Por (B.5) conclui-se diretamente que ambos os sistemas com entrada $X(z)$ levam a uma saída $X(z^M)H(z^M)$.

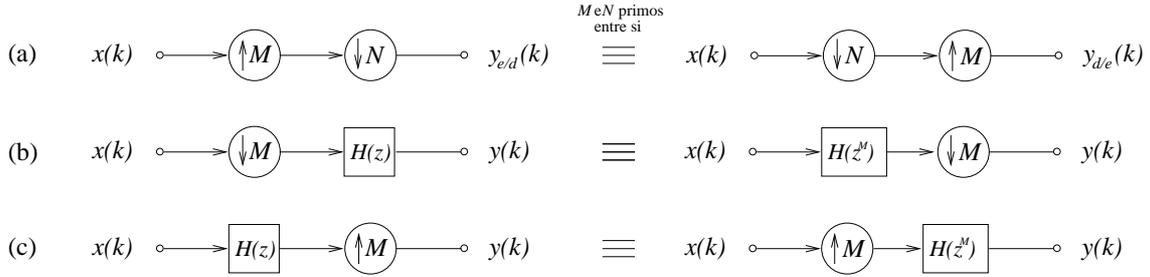


Figura B.5: Identities nobres: (a) Comutação entre Expansor e Decimador. (b) Comutação entre Filtro e Decimador. (c) Comutação entre Filtro e Expansor.

B.4 Bancos de QMF's e CQF's

Dois projetos bastante populares de bancos de filtros para 2 sub-bandas (Ver Fig. B.2) são os *Quadrature Mirror Filters* (QMF) e os *Conjugate Quadrature Filters* (CQF).

Os bancos de QMF's se caracterizam por possuir fase linear (o que pode ser requerido ou vantajoso, conforme o tipo de processamento que se deseja fazer) e eliminar *aliasing*; entretanto, quando realizados em uma estrutura de divisão em duas sub-bandas, não possuem a propriedade de Reconstrução Perfeita. Para esse tipo de estrutura, os QMF's são obtidos partindo-se de um filtro protótipo FIR $H_0(z)$ simétrico de comprimento L par e forçando-se

$$H_1(z) = H_0(-z), \quad (\text{B.14})$$

$$F_0(z) = H_0(z), \quad (\text{B.15})$$

e

$$F_1(z) = -H_0(-z). \quad (\text{B.16})$$

Para aproximar reconstrução perfeita $H_0(z)$ deve ser escolhido de forma que

$$H_0^2(z) - H_0^2(-z) \approx 2z^{-k_0}. \quad (\text{B.17})$$

Os CQF se caracterizam por permitirem Reconstrução Perfeita, mas sem possuir fase linear. São obtidos partindo-se de um filtro protótipo $H_0(z)$ de comprimento L tal que

$$H_0(z)H_0(z^{-1}) + H_0(-z)H_0(-z^{-1}) = 2 \quad (\text{B.18})$$

e forçando-se

$$H_1(z) = z^{-(L-1)} H_0(-z^{-1}), \quad (\text{B.19})$$

$$F_0(z) = -z^{-(L-1)} H_0(z^{-1}), \quad (\text{B.20})$$

e

$$F_1(z) = H_0(-z). \quad (\text{B.21})$$