

RESTAURAÇÃO DIGITAL DE SINAIS DE ÁUDIO PROVENIENTES DE
GRAVAÇÕES MUSICAIS DEGRADADAS

Luiz Wagner Pereira Biscainho

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 DOUTOR EM CIÊNCIAS
EM ENGENHARIA ELÉTRICA.

Aprovada por:

Prof. Paulo Sérgio Ramirez Diniz, Ph.D.

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

Prof. Jacques Szczupak, Ph.D.

Prof. Márcio Nogueira de Souza, D.Sc.

Prof. Weiler Alves Finamore, Ph.D.

RIO DE JANEIRO, RJ - BRASIL
DEZEMBRO DE 2000

BISCAINHO, LUIZ WAGNER PEREIRA

Restauração Digital de Sinais de Áudio Provenientes de Gravações Musicais Degradadas [Rio de Janeiro] 2000

XI, 210 p., 29,7 cm (COPPE/UFRJ, D.Sc., Engenharia Elétrica, 2000)

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 Doutor em Ciências (D.Sc.)

RESTAURAÇÃO DIGITAL DE SINAIS DE ÁUDIO PROVENIENTES DE GRAVAÇÕES MUSICAIS DEGRADADAS

Luiz Wagner Pereira Biscainho

Dezembro/2000

Orientador: Paulo Sérgio Ramirez Diniz

Programa de Engenharia Elétrica

Este trabalho apresenta um conjunto de técnicas de processamento digital de sinais de áudio orientadas para a restauração de gravações musicais deterioradas. Mesclando aspectos teóricos e práticos, discute-se e afere-se o desempenho das técnicas correntes e são propostas contribuições no sentido de melhor interpretá-las e analisá-las, reduzir seu esforço computacional ou melhorar seu desempenho. São abordados tratamentos de ruído de fundo, ruído impulsivo e pulsos longos, envolvendo principalmente modelagem estatística de sinais, representação de sinais na frequência, sistemas adaptativos, sistemas multitaxa e *wavelets*. Como coprodutos, traçam-se discussões sobre modelos, cujas aplicações transcendem a restauração de áudio. Há um esforço em aplicar, quando possível, métodos objetivos adequados à avaliação de resultados práticos.

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

DIGITAL RESTORATION OF AUDIO SIGNALS FROM DEGRADED
MUSICAL RECORDINGS

Luiz Wagner Pereira Biscainho

December/2000

Advisor: Paulo Sérgio Ramirez Diniz

Department: Electrical Engineering

This work investigates a myriad of digital signal processing techniques applied to restoration of degraded musical recordings. Mixing theoretical and practical aspects, current techniques have their performance discussed and analyzed, leading to new contributions aiming new interpretations and analysis methods, reduced computational effort or better performance. Elimination of background noise, impulsive noise and long duration pulses are achieved by employing several techniques, mainly through statistical signal modelling, signal frequency description, adaptive systems, multirate systems and wavelets. As by-products, results related to ARMA models in sub-bands are advanced, of which potential applications transcend audio restoration. An effort is made in employing, whenever possible, appropriate objective methods to evaluate the quality of the restored audio signals.

Dedico este trabalho à Música e aos músicos, que, artistas que são, fazem de sua loucura uma mensagem à Humanidade.

Agradeço a todas as pessoas que possibilitaram a elaboração deste trabalho: minha família, minha analista Ana Maria Rozzante de Castro e meus amigos extra-trabalho, pelo suporte pessoal; meus alunos-amigos que se aventuraram comigo nesta pesquisa, Paulo Antônio Andrade Esquef, Arthur Haerdy Jr., Fábio Pacheco Freeland, Mauro Flores de Carvalho e Cristiano Nogueira dos Santos, sem os quais não teria realizado a tese; meus outros alunos, que fazem parte do lado mais gratificante da minha profissão; meu amigo duas vezes orientador Paulo Sérgio Ramirez Diniz, pela inspiração e confiança; meu amigo Eduardo Antônio Barros da Silva, pelo estímulo e exemplo; demais colegas-professores do Departamento de Eletrônica, pelo apoio irrestrito que me proporcionaram; funcionários e demais alunos que também me honraram com sua colaboração. Gostaria ainda de destacar que minha participação no LPS tem sido inesgotável fonte de energia para minha evolução profissional e pessoal.

Sumário

Folha de Rosto	i
Ficha Catalográfica	ii
Resumo	iii
Abstract	iv
Dedicatória	v
Agradecimentos	vi
Sumário	vii
1 Apresentação	1
1.1 Uma Breve História do Áudio Gravado	2
1.2 O Problema-Objetivo	4
1.3 Elementos da Descrição de Sinais Musicais de Áudio	6
1.3.1 Domínio Contínuo	6
1.3.1.1 Modelo Estatístico	6
1.3.1.2 Conteúdo Espectral	7
1.3.1.3 Faixa Dinâmica	8
1.3.1.4 Percepção	8
1.3.2 Domínio Discreto	9
1.3.2.1 Modelo Estatístico	9
1.3.2.2 Taxa de Amostragem	9
1.3.2.3 Número de Bits de Representação	9
1.3.2.4 Percepção	9

I Sistemas e Modelos 11

2 Efeito de Zeros, Pólos e Pares Zero-Pólo sobre uma Função de Transferência Discreta 13

2.1 Efeito de Zero ou Pólo Individual 14

2.2 Efeito de Par Zero-Pólo 20

2.3 Exemplo 28

3 Efeito da Quantização sobre a Obtenção do Modelo AR para um Sinal de Áudio 29

3.1 Apresentação do Problema 30

3.1.1 Indício: Interpolação de Amostras Faltantes Baseada em Modelo 30

3.1.2 Origem: Cálculo do Modelo AR 31

3.1.3 Testes Adicionais 31

3.2 Premissas Teóricas 32

3.2.1 Descrição do Erro de Quantização 32

3.2.2 Estimação LS de Parâmetros de Modelo AR 33

3.2.3 Desvio na Estimativa de Parâmetros devido a Ruído Aditivo . 34

3.2.4 Desvio dos Pólos do Filtro Gerador devido ao Desvio nos Parâmetros do Modelo AR 36

3.3 Explicações e Solução 39

4 Modelo para um Processo AR Analisado em Oitavas por Banco de Filtros 41

4.1 Motivação: Restauração de Áudio em Sub-Bandas 41

4.2 Caso Geral: Modelo para um Processo ARMA Dividido em Sub-Bandas 42

4.3 Modelo para um Processo AR Dividido em Oitavas 45

4.4 Correlação entre Sub-Bandas 55

4.5 Uma Representação Hierárquica 55

4.6 Discussão da Validade do Modelo 59

4.7 Mapeamento de Zeros e Pólos em cada Sub-Banda 62

4.7.1	Banco de Filtros Ideais	62
4.7.2	Banco de Filtros Realizáveis	68
4.7.3	Exemplos Práticos e Constatações	71
5	<i>Aliasing</i> e Reconstrução Perfeita no Processamento Multitaxa	91
II	Restauração	96
6	Psicoacústica e Avaliação Objetiva de Qualidade	98
6.1	Avaliação da Qualidade de Sinais de Áudio	99
6.2	Psicoacústica	100
6.2.1	Audibilidade	101
6.2.2	Mascaramento	102
6.2.3	Bandas Críticas	103
6.2.4	Espalhamento	103
6.2.5	Compressão	103
6.3	PAQM	107
6.4	Proposta: PAQM Adaptada à Aferição da Qualidade de Sinais Res- taurados	109
7	Interpolação de Amostras Faltantes Baseada em Modelo	111
7.1	Formulação do Problema	111
7.2	Técnica de Base Determinística: LS	113
7.3	Estimativa de ML	114
7.4	Estimação por EM	114
7.5	A Necessidade do Vetor-Parâmetro	115
7.6	Estimativa pelo Amostrador de Gibbs	115
7.7	Da Implementação	116
7.8	Investigação Realizada	116
7.8.1	Testes e Resultados	116
7.8.2	Interpolação de Sinal Quantizado	117
8	Tratamento de Ruído Impulsivo	121
8.1	Caracterização	121

8.2	Restauração sem Modelo: Filtragem por Mediana	122
8.3	Restauração com Modelagem do Sinal	123
8.3.1	Notação Adotada	123
8.3.2	Detecção por Filtragem Inversa do Modelo AR + Interpolação	124
8.3.3	Restauração Bayesiana	128
8.4	Aperfeiçoamento da Detecção por Inversão do Modelo AR	129
8.4.1	Determinação da Extensão da Falha	130
8.4.2	Reiteração e Parada do Algoritmo por Bloco	132
8.4.3	Desempenho do Algoritmo Modificado	134
8.4.3.1	Não-Detecção e Falso Alarme	135
8.4.3.2	Avaliação Subjetiva	135
8.4.4	Avaliação Objetiva por Critérios Perceptivos	136
8.5	Tratamento em Sub-bandas	137
8.5.1	Descrição e Discussão	137
8.5.2	Implementação e Avaliação	140
9	Tratamento de Pulsos Longos	142
9.1	Restauração por Similaridade	144
9.1.1	Testes Realizados	146
9.2	Supressão por Separação de Processos AR	148
9.3	Detecção Rápida	150
10	Restauração de Sinais Contaminados por Ruído de Fundo	155
10.1	Métodos Baseados no Espectro	156
10.1.1	Fundamentos	156
10.1.2	Subtração Espectral	156
10.1.3	Filtragem de Wiener	157
10.1.4	Investigação Realizada	159
10.1.4.1	Avaliação Inicial de Desempenho	159
10.1.4.2	Proposições	161
10.1.4.3	Avaliação das Proposições	163
10.2	Método Baseado em <i>Wavelet: Wavelet Shrinkage</i>	165
10.2.1	Princípios da DWT	165

10.2.2	Compressão de Informação Usando <i>Wavelets</i>	169
10.2.3	A Redução de Ruído de Fundo	170
10.2.4	Métodos de Cálculo dos Limiares	170
10.2.4.1	Método Minimax e Variante Prática	171
10.2.4.2	Método SURE (<i>Stein's Unbiased Risk Estimate</i>)	172
10.2.4.3	Método Híbrido	173
10.2.5	Investigação: Avaliação da <i>Wavelet Shrinkage</i>	174
10.2.5.1	Sinais e Estratégia	174
10.2.5.2	Medidas	175
10.2.5.3	Conclusões	179
10.3	Comentário Final	181
11	Supressor Adaptativo de Ruído	182
11.1	Filtragem Adaptativa	182
11.2	Supressor Adaptativo de Ruído	184
11.3	Investigação Realizada	185
11.3.1	Testes Práticos sobre Sinais Reais	185
11.3.2	Expectativas de Desempenho	185
12	Avanços nos Métodos em Sub-bandas	189
12.1	Tratamento de Ruído Impulsivo Baseado em Modelo Realizado em Sub-bandas Revisitado	189
12.2	Tratamento de Ruído de Fundo por Filtragem de Wiener Baseada na DFT numa Estrutura em Oitavas	193
12.2.1	Filtragem de Wiener Baseada na DFT	193
12.2.2	Filtragem de Wiener diretamente nas Sub-bandas	194
12.2.3	Composição Proposta	195
12.2.3.1	Descrição	195
12.2.3.2	Avaliação	196
	Referências Bibliográficas	202

Capítulo 1

Apresentação

A idéia para o tema desta tese resultou, inicialmente, de um antigo interesse pessoal em música—quanto às obras, às interpretações e, como extensão natural, à sua gravação. A formação em Engenharia Eletrônica e sua subsequente especialização em Processamento de Sinais forneceram o arsenal técnico que permitirá contribuir um pouco para a preservação da integridade de realizações artísticas únicas no tempo.

Cronologicamente, o desenrolar do trabalho partiu de uma ampla pesquisa dos métodos mais modernos e em aplicação na restauração digital de gravações musicais degradadas, na forma em que estavam documentados na literatura científica. O passo seguinte, que envolveu maciças doses de bom-senso e criatividade, foi implementar os métodos de maior interesse e testá-los em condições reais. Detectados seus principais problemas e limitações, buscaram-se meios de explicá-los, reduzi-los e/ou solucioná-los. Alguns dos meios teóricos empregados acabaram por merecer qualificação como tópicos independentes, por sua generalidade. A avaliação quantitativa dos resultados práticos também mereceu especial cuidado.

A organização final do texto assumiu uma ordem bem diferente daquela em que o trabalho foi realizado. Preferiu-se apresentar na Parte I os desenvolvimentos e resultados mais genéricos referentes aos modelos associados ao processamento, deixando para a Parte II o que diz respeito à restauração propriamente dita. Embora isso não pareça óbvio preliminarmente, todos os Capítulos se interligam. Por outro lado, excetuando-se o Capítulo final, houve um esforço em torná-los autocontidos, ao menos quanto ao seu assunto principal, ocasionando um mínimo de redundância.

As publicações do autor associadas a este trabalho são [1] (Capítulo 2); [2] (Capítulo 3); [3] e [4] (Capítulo 4); [5] e [6] (Capítulo 8); [7] (Capítulo 9); e [8] (Capítulo 10).

O autor tentou fazer com que o texto seguisse estritamente o vocabulário recomendado em [9]. A edição foi feita em LaTeX [10, 11].

1.1 Uma Breve História do Áudio Gravado

Como forma de ambientar historicamente o alvo deste trabalho, é apresentada a seguir uma cronologia [12, 13] bastante simplificada do som gravado.

1877 Edison faz a primeira gravação num **fonógrafo** para cilindros de latão.

1885 Bell & Tainter patenteiam o **grafofone** para cilindros de cera, com sulcos cortados verticalmente.

1888 Berliner patenteia o **gramofone** para discos de vulcanite de 7" de diâmetro, com sulcos cortados lateralmente e 1 só lado gravado a 70 rpm, para produção em massa.

1897 Os discos passam a ser feitos de goma-laca, ainda reproduzidos a partir de agulhas de aço de pressão equivalente a 270 g.

1898 Poulser patenteia o **telegrafone**, gravador magnético para fio de aço.

1901 A Victor Co. lança discos de 10" a 78 rpm.

1904 A Odeon produz em massa discos gravados em ambos os lados.

1906 É lançada a **vitrola**, com corneta embutida.

1912 Edison obtém qualidade acústica superior com o uso de agulhas de diamante.

1913 Edison fabrica discos.

1925 São produzidos os primeiros discos gravados eletricamente, pelo sistema Western Electric, desenvolvido nos Bell Labs da AT&T.

Vitaphone & Warner Brothers demonstram discos de alumínio coberto de acetato a 33 1/3 rpm para uso em sincronismo com filmes.

1926 Brush vende agulhas piezoelétricas superleves.

1927 A AMI Co. faz as primeiras *jukeboxes*.

1928 Pleufmer patenteia a aplicação de pó magnético à fita de papel.

1929 Edison fabrica os últimos cilindros e discos.

A RCA se funde à Victor.

1931 Blumlein patenteia um método de gravação binaural (estéreo).

A RCA apresenta discos de Victrolac com sulco largo a 33 1/3 rpm, que não se popularizam.

1932 A BASF e a AEG-Telefunken desenvolvem a gravação em fita magnética a partir da patente de Pleufmer, vindo a lançar o **magnetofone** em 1935.

1934 É cunhada a expressão **alta-fidelidade** (Hi-Fi).

1939 Camras inventa independentemente o gravador de fio, que se torna muito popular até 1955 entre amadores.

1948 A Columbia apresenta o disco *long-play* (LP) de 12" com microsulco a 33 1/3 rpm em vinilite.

É lançada a fita coberta de óxido férrico 3M Scotch 111.

A Magnecord faz o primeiro gravador de fita de rolo estéreo.

1949 A RCA Victor apresenta o disco *extended-play* (EP) de 7" com microsulco a 45 rpm em vinilite, mais tarde em poliestireno.

1951 AMPEX e RCA começam a trabalhar num *video-tape recorder* (VTR), fabricando a primeira unidade móvel em 1959.

1954 A RCA Victor vende as primeiras fitas de rolo estéreo pré-gravadas.

1958 O padrão para discos em estéreo é estabelecido.

1963 A Philips demonstra o primeiro audiocassete compacto com fita de 1/8" de largura a 1 7/8 ips.

1969 O sistema Dolby NR (*Noise Reduction*) é introduzido em fitas pré-gravadas.

A Sony apresenta o primeiro videocassete.

1972 A Philips apresenta o primeiro reprodutor de *laser-disc* (LD).

1975 A Sony apresenta o *video-cassete recorder* (VCR) Betamax doméstico com fita de 1 hora de duração.

1976 A JVC apresenta o VCR VHS doméstico com fita de 1 hora de duração.

1977 RCA anuncia fitas de vídeo de 4 horas de duração.

1978 A Pioneer desenvolve o LD.

1979 A Sony apresenta o *walkman*.

1980 São comercializados *compact-discs* (CD's) de 5".

1983 É lançado o Beta HiFi VCR com som FM de alta qualidade.

1987 É apresentada a *digital audio tape* (DAT).

1997 O *digital video disc*, depois *digital versatile disc* (DVD), se torna disponível comercialmente como meio de armazenamento de áudio e vídeo e de dados de computador.

1.2 O Problema-Objetivo

Com a tecnologia digital [14, 15], o armazenamento de áudio sofreu sua maior revolução desde a invenção do fonógrafo. Pela primeira vez, foi possível gerar cópias rigorosamente iguais, garantir a perenidade de uma matriz, utilizar recursos de processamento antes impensáveis e ter o controle sobre o binômio quantidade x qualidade da informação. Por razões históricas e artísticas, imediatamente se quis estender às gravações analógicas as conquistas da nova era, quanto à possibilidade de preservá-las e, num passo adiante, de restaurar ou mesmo aumentar sua qualidade. Curiosamente, a mais ousada destas metas—aumentar a qualidade—motivou a primeira utilização de processamento digital em música, na primeira metade dos anos

70 [16, 13], quando se aplicou a modelagem física dos sistemas de gravação do início do século para, por filtragem inversa, reconstituir o sinal original.

Considerando que o objetivo principal deste trabalho é o tratamento de sinais de áudio pré-gravados com vistas à recuperação de qualidade, faz-se necessário definir, em termos gerais, o conjunto-universo onde ele se insere.

O primeiro aspecto a delimitar é o tipo das matrizes de áudio de que se ocupará o trabalho: cilindros, discos e fitas magnéticas [17]. Sendo essas as origens dos sinais, podemos classificar grosseiramente as falhas encontráveis em dois grandes grupos: distribuídas e localizadas.

As falhas distribuídas podem ser de diversos tipos: ruído na forma de um chiado de fundo, de espectro largo em frequência, aproximadamente plano (ruído branco) ou não (ruído colorido), associado, em geral, ao uso de fitas magnéticas como meio de armazenamento; ruído de fundo de baixa frequência, associado, em geral, a vibração de motor no equipamento de gravação ou reprodução; flutuações de velocidade periódicas, originadas por excentricidade de disco, e não-periódicas, origináveis na edição e mixagem de diversas tomadas de som; e distorções globais, originadas em não-linearidades inerentes aos processos de gravação e reprodução.

As falhas localizadas também podem ser de diversos tipos: ruído na forma de estalidos breves que, se muito próximos, podem dar a impressão de continuidade, caracterizado por pequenos pulsos dispersos aleatoriamente ao longo do sinal, e estalos não-periódicos, caracterizados por pulsos de duração relativamente longa, associados, em geral, a resíduos agregados ou falhas no material do disco ou cilindro; estalos periódicos, também caracterizados por pulsos de duração relativamente longa, originados por arranhão ou quebra do disco ou cilindro; ruídos periódicos de duração muito longa, por vezes assemelhando-se a uma raspagem, originados, em geral, por resíduos agregados no material ou desgaste não-homogêneo do disco ou cilindro; distorção por saturação, caracterizada por limitação rígida na amplitude do sinal, associada, em geral, a desgaste do disco ou cilindro ou a ajuste inadequado do equipamento de gravação.

Vale destacar que nem todos os ruídos presentes numa gravação são objeto de processamento. Pode ser desejável, por exemplo, preservar os ruídos da platéia em uma execução ao vivo, em nome da fidelidade.

1.3 Elementos da Descrição de Sinais Musicais de Áudio

O objetivo dessa Seção é prover alguma familiaridade com os sinais que se pretenderá, em geral, processar e com alguns princípios e considerações que podem ser empregados em sua modelagem.

1.3.1 Domínio Contínuo

1.3.1.1 Modelo Estatístico

Descrever matematicamente um sinal de áudio musical gravado típico de forma completa é uma tarefa impossível, se se consideram as infinitas possibilidades envolvendo seu espectro em frequência (relacionado diretamente às notas musicais e aos timbres dos instrumentos) e sua potência (relacionada diretamente à dinâmica requerida)—termos, aliás, imprecisos para essa categoria de sinais com características tão fortemente variantes no tempo.

Uma forma razoavelmente cautelosa de modelar um tal sinal é considerá-lo uma função-amostra de um processo aleatório não-estacionário [18, 19]. Nesse ponto, pode ser útil uma pequena discussão:

A maneira mais direta de associar esse modelo ao problema de restauração, dadas diversas cópias de uma matriz, é interpretar cada uma como uma função-amostra de um processo composto de um sinal determinístico (o sinal, propriamente dito) e outro, aleatório (as imperfeições a eliminar). Contudo, na maioria dos casos, trata-se de restaurar uma única cópia disponível (ou de poucas cópias disponíveis) da gravação, o que reduz a utilidade dessa abordagem.

Alternativamente, pode-se atribuir um caráter aleatório ao próprio sinal de áudio (o que é mais usual). Porém, sem considerações adicionais, a interpretação “física” do processo fica bem mais difícil: o que formaria o *ensemble*? Avançando mais um passo nas hipóteses, se existe um intervalo de tempo máximo para o qual o processo ainda pode ser considerado aproximadamente estacionário (o qual, para áudio musical típico, costuma ser tomado entre 25 e 50 ms, conforme sua variabilidade maior ou menor no tempo) e se supõe a ergodicidade nesse intervalo (permitindo obter valores esperados a partir de médias temporais), está reabilitada a

interpretação: uma única cópia do sinal (função-amostra) pode responder pelo processo não-estacionário se janelas finitas permitirem caracterizar aproximadamente subprocessos estacionários ergódicos. Isso dispensa, na verdade, a interpretação do *ensemble*, pois restringe a análise ao tempo.

Feitas tais considerações, tornam a fazer sentido prático (para cada janela) potência e espectro em frequência, permitindo que se prossiga com a caracterização do sinal.

1.3.1.2 Conteúdo Espectral

Costuma-se adotar os limites da audição humana como 20 Hz e 20 kHz [20]. Assim sendo, parece razoável, para efeito de gravação de áudio musical, buscar fidelidade ao menos dentro desses limites. Contudo, este é um tópico polêmico, gerando até hoje discussões quanto ao limite superior útil do espectro em frequência da informação a gravar. Defendem alguns que frequências não decodificadas convencionalmente a partir do aparelho auditivo podem ser “sentidas” de outras formas pelo corpo. Paralelamente, estudos têm registrado que o espectro de instrumentos musicais (especialmente os de percussão) pode atingir os 200 kHz com conteúdo ainda considerável!

Outro dado importante, ligado ao refinamento necessário à descrição dos sinais musicais em frequência, se refere à organização da escala musical. Atualmente, a música ocidental tradicional utiliza a escala temperada, que, adotando uma nota-padrão, individualiza 12 semitons dentro de cada oitava espaçados geometricamente. Por exemplo, escolhendo-se como padrão o **lá** em 440 Hz, o próximo **lá** estará em 880 Hz (a uma oitava de distância). E a razão entre as frequências de quaisquer 2 semitons contíguos, por exemplo, **si bemol** e **lá**, é sempre $2^{\frac{1}{12}} \approx 1,06$, ou seja, um semitom de distância corresponde a uma diferença de 6% em frequência.

Em função do caráter predominantemente tonal da música, a inspeção do espectro em frequência de um sinal de áudio musical indica, usualmente, a presença de picos ressonantes e vales suaves. Representando-se esse sinal como a saída de um filtro excitado por ruído branco (entrada com espectro plano), o filtro apresentará preferencialmente pólos a zeros. Outra representação possível para o sinal, de inspiração próxima, é como uma combinação de senóides individualmente moduladas em

amplitude e frequência [21, 22]. Alternativamente, a expansão do sinal como combinação linear de funções-base apropriadas, como *wavelets* [23, 24], pode fornecer uma descrição que equilibra boa definição no tempo e na frequência.

1.3.1.3 Faixa Dinâmica

A audição humana responde a diferenças de pressão sonora, entre o limiar da audição e o limiar da dor, de aproximadamente 120 dB. Uma orquestra sinfônica numa sala de concertos pode ultrapassar 100 dB de faixa dinâmica. Tais valores propõem um limite ideal para o armazenamento do áudio.

1.3.1.4 Percepção

Embora esse tópico vá ser abordado com algum detalhe no Capítulo 6.2, vale mencionar dois aspectos [20], ligados às Subseções 1.3.1.2 e 1.3.1.3.

- A frequência percebida auditivamente (*pitch*) não é, em geral, a mesma emitida pela fonte sonora, podendo a diferença percentual entre elas variar com a própria frequência emitida, com o conjunto de frequências eventualmente emitidas em conjunto com ela, com a intensidade do som e até mesmo com o estado emocional do ouvinte!
- O aparelho auditivo tem uma espécie de controle automático de ganho, de tal forma que em ambientes silenciosos prepara-se para “ouvir mais alto” (aumentando a capacidade de percepção de detalhes) e em ambientes mais ruidosos prepara-se para “ouvir mais baixo” (protegendo o sistema auditivo contra continuada pressão excessiva). Esse mecanismo tem uma inércia considerável; por isso, permanecemos um pouco surdos por um tempo após sair de um *show* em volume muito elevado.

Isso dá uma dimensão de quanto podem ser equívocas as medidas quantitativas usuais na caracterização do som ouvido.

1.3.2 Domínio Discreto

1.3.2.1 Modelo Estatístico

Considerando o sinal de áudio discretizado, o modelo na forma de um filtro gerador excitado por ruído branco mencionado na Subseção 1.3.1.2 recai na forma autorregressiva (AR), que utiliza um filtro IIR só-pólos; em geral, conforme o conteúdo espectral menos ou mais rico, uma ordem entre 20 e 50 é suficiente para descrever o sinal. Outros modelos mais sofisticados podem ser adotados [16], como o autorregressivo com média móvel (ARMA), que acrescenta zeros ao filtro gerador [18, 25], e variações do AR e do ARMA incorporando não-linearidades.

1.3.2.2 Taxa de Amostragem

Segundo o Critério de Nyquist [26], a menor taxa de amostragem de um sinal contínuo de espectro limitado superiormente na frequência f_{sup} que permite sua perfeita recuperação como sinal contínuo é $2f_{sup}$. Tomando o limite superior da audição como 20 kHz, o padrão mais popular ainda adotado para o áudio comercial de alta qualidade (o CD [14]) realiza a amostragem a 44,1 kHz, após limitar convenientemente em frequência o sinal de áudio original.

1.3.2.3 Número de Bits de Representação

O fato de o ouvido comprimir os sinais em potência na percepção sugere que a quantização do áudio seja realizada de forma não-linear, dando maior detalhe às diferenças entre pequenas amplitudes e menor detalhe às diferenças entre grandes amplitudes; pelo mesmo motivo, parece desejável representar digitalmente suas amostras em ponto flutuante. Contudo, o padrão do CD [14] realiza quantização linear e representação em ponto fixo com 16 bits, o que resulta numa faixa dinâmica teórica de cerca de 96 dB.

1.3.2.4 Percepção

Testes realizados com ouvintes especializados indicaram que a taxa de amostragem mínima e o número de bits mínimo para que fosse imperceptível a diferença entre uma gravação analógica e uma gravação digital eram 96 kHz e 24 bits, res-

pectivamente. Deve-se notar que isso só determina o limite da polêmica entre as gravações digitais e analógicas, mas não entre as gravações digitais e o som original. De qualquer modo, a tecnologia já permite alcançar aqueles valores no padrão do DVD.

Importância crescente tem sido dada aos aspectos perceptivos na codificação de áudio, permitindo atingir elevadas taxas de compressão com alta qualidade.

1.3.2.5 Padrão Adotado neste Trabalho

Nos testes práticos efetuados ao longo deste trabalho, a menos de menção contrária, os arquivos de áudio a processar eram monaurais, amostrados a 44,1 kHz e representados em ponto fixo com 16 bits. O processamento foi feito na precisão da máquina e os arquivos de saída foram levados de volta ao formato original.

Parte I

Sistemas e Modelos

Nesta Parte, figuram os tópicos teóricos que se originaram das investigações em restauração de áudio da Parte II e tinham um grau de generalidade tal que justificou sua apresentação à parte. A ordem adotada foi aproximadamente a de crescente especificidade.

O estudo sistemático do efeito de zeros e pólos em funções de transferência discretas (assunto do Capítulo 2) surgiu do interesse em julgar a possibilidade de simplificar a ordem do modelo derivado para o processo AR dividido em sub-bandas de frequência (assunto do Capítulo 4). Este modelo resultou da idéia de realizar o tratamento de ruído impulsivo em sub-bandas, apresentada na Seção 8.5. O estudo dos efeitos da quantização do sinal na obtenção de seu modelo AR (assunto do Capítulo 3) buscou explicar o insucesso da interpolação de sinais de áudio sob certas condições, constatado na Subseção 7.8.2. Por fim, a discussão sobre a reconstrução perfeita veio à tona após o uso bem-sucedido de bancos de filtros sem essa propriedade no processamento de ruído de fundo, na Subseção 10.1.4.2.

Capítulo 2

Efeito de Zeros, Pólos e Pares

Zero-Pólo sobre uma Função de Transferência Discreta

Em inúmeras aplicações práticas, associam-se modelos (justificados fisicamente ou apenas matematicamente) a sinais e sistemas reais, de forma a permitir realizar algum tipo de processamento. A complexidade dos modelos determina, em última análise, a velocidade com que se pode realizar o processamento, com conseqüências sobre o alcance de sua eficácia. Assim, a utilização de um modelo pressupõe quase sempre sua máxima simplificação sem comprometer o refinamento que dele se espera na aplicação em vista.

Tendo em mente o interesse em reduzir a ordem do modelo adotado para um sistema discreto no tempo (aplicando-se, assim, diretamente, aos resultados do Capítulo 4), este Capítulo tenta quantificar o efeito que o abandono de um zero, um pólo ou um par zero-pólo da função de transferência correspondente causaria sobre a resposta em frequência do modelo. Como parâmetros de avaliação, escolheram-se os máximos desvios de módulo e de fase que os candidatos a descarte impõem, individualmente, à função de transferência.

A estratégia adotada será, partindo de um sistema de fase mínima, sem perda de generalidade, isolar a parcela que se deseja eliminar de sua função de transferência global $H(z)$ e compará-la com a função de transferência unitária. Os máximos desvios em relação a esta representarão o erro cometido com a simplificação.

Os diversos casos a seguir serão apresentados, grosso modo, dos mais particulares para os mais gerais, de forma a facilitar sua interpretação, embora pudessem ser todos reunidos no último, mais abrangente. Em [27, 28], a busca de fórmulas para cálculo de $|H(e^{j\omega})|_{\text{máx}}$ e $|H(e^{j\omega})|_{\text{mín}}$ para escalamento de filtros com funções de transferência $H(z)$ acaba por chegar a expressões equivalentes às de desvio de módulo deste último caso—equações (2.46), (2.47) e (2.48)—, embora descritas em função dos coeficientes do filtro. Na seqüência, contudo, além de se formular também o desvio de fase, os resultados serão descritos em função dos módulos e fases dos zeros e pólos do filtro. Outro trabalho em linha próxima foi recentemente publicado em [29].

2.1 Efeito de Zero ou Pólo Individual

Inicialmente, será avaliada a eliminação de zeros ou pólos. Pergunta a responder: qual a contribuição do termo “abandonado” à função de transferência global? Serão examinados os casos de um zero real, dois zeros complexos conjugados, um pólo real e dois pólos complexos conjugados. Partindo do módulo e da fase de sua resposta em freqüência isolada, serão obtidos em relação a ω o módulo mínimo m e o módulo máximo M , levando ao desvio máximo em relação ao módulo unitário (ideal)

$$\Delta M = \text{máx}(1 - m, M - 1), \quad (2.1)$$

e o desvio máximo $\Delta\theta$ em relação à fase nula (ideal). Tais parâmetros dependem da localização dos zeros ou pólos, em função da qual serão descritos analítica e graficamente.

Com referência ao zero real a , considere-se a transferência parcial

$$H_{zr}(z) = \frac{z - a}{z}, \quad |a| < 1. \quad (2.2)$$

Sua resposta em freqüência pode ser descrita por

$$|H_{zr}(e^{j\omega})|^2 = 1 - 2a \cos \omega + a^2 \quad (2.3)$$

e

$$\angle H_{zr}(e^{j\omega}) = \text{arctg} \frac{a \text{ sen } \omega}{1 - a \cos \omega}. \quad (2.4)$$

“Abandonar” o zero a corresponderia a fazer

$$H_{zr}(z) \approx 1. \quad (2.5)$$

Definindo-se m como o mínimo e M como o máximo valor de $|H_{zr}(e^{j\omega})|$, chega-se a

$$m = 1 - |a|, \text{ em } \omega = \begin{cases} \pi, & \text{se } a < 0 \\ 0, & \text{se } a > 0 \end{cases} \quad (2.6)$$

e

$$M = 1 + |a|, \text{ em } \omega = \begin{cases} 0, & \text{se } a < 0 \\ \pi, & \text{se } a > 0. \end{cases} \quad (2.7)$$

O desvio máximo de módulo, obtido pela equação (2.1), é

$$\Delta M = |a|. \quad (2.8)$$

Definindo-se $\Delta\theta$ como o máximo valor do desvio $|\angle H_{zr}(e^{j\omega})|$, chega-se a

$$\Delta\theta = \arcsen |a|, \quad (2.9)$$

na frequência $\omega = \arccos a$.

Com isso, obtiveram-se as expressões que descrevem o efeito do zero real a sobre a resposta em frequência global.

Com referência ao par de zeros complexos conjugados $Ae^{\pm j\alpha}$, considere-se a transferência parcial

$$H_{2zc}(z) = \frac{z^2 - 2A \cos \alpha z + A^2}{z^2}, \quad A < 1. \quad (2.10)$$

Sua resposta em frequência pode ser descrita por

$$|H_{2zc}(e^{j\omega})|^2 = 1 - 4A \cos \alpha \cos \omega + 2A^2[2 \cos^2 \alpha + \cos(2\omega)] - 4A^3 \cos \alpha \cos \omega + A^4 \quad (2.11)$$

e

$$\angle H_{2zc}(e^{j\omega}) = \text{arctg} \frac{2A \cos \alpha \sin \omega - A^2 \sin(2\omega)}{1 - 2A \cos \alpha \cos \omega + A^2 \cos(2\omega)}. \quad (2.12)$$

“Abandonar” os zeros $Ae^{\pm j\alpha}$ corresponderia a fazer

$$H_{2zc}(z) \approx 1. \quad (2.13)$$

Definindo-se m^2 como o mínimo e M^2 como o máximo valor¹ de $|H_{2zc}(e^{j\omega})|$, chega-se a

$$m = \begin{cases} \sqrt{1 - 2A|\cos \alpha| + A^2}, & \text{se } |\cos \alpha| > \frac{2A}{1+A^2}, \text{ em } \omega = \begin{cases} 0, & \text{se } 0 < \alpha < \frac{\pi}{2} \\ \pi, & \text{se } \frac{\pi}{2} < \alpha < \pi \end{cases} \\ \sqrt{(1 - A^2)|\sin \alpha|}, & \text{se } |\cos \alpha| < \frac{2A}{1+A^2}, \text{ em } \omega = \arccos\left(\frac{1+A^2}{2A} \cos \alpha\right), \end{cases} \quad (2.14)$$

e

$$M = \sqrt{1 + 2A|\cos \alpha| + A^2}, \text{ em } \omega = \begin{cases} \pi, & \text{se } 0 < \alpha < \frac{\pi}{2} \\ 0, & \text{se } \frac{\pi}{2} < \alpha < \pi. \end{cases} \quad (2.15)$$

O desvio máximo de módulo ΔM se obtém pela equação (2.1).

Definindo-se $2\Delta\theta$ como o máximo valor² do desvio $|\angle H_{2zc}(e^{j\omega})|$, chega-se a

$$\Delta\theta = \frac{1}{2} \text{máx}(|\angle H_{2zc}(e^{j\omega_0})|), \quad (2.16)$$

sendo

$$\omega_0 = \arccos \frac{(1 + 3A^2) \cos \alpha \pm \sqrt{(1 - 10A^2 + 9A^4) \cos^2 \alpha + 8A^2(1 - A^2)}}{4A} \quad (2.17)$$

com $0 \leq \omega_0 \leq \pi$.

Com isso, obtiveram-se as expressões que descrevem o efeito dos zeros $Ae^{\pm j\alpha}$ sobre a resposta em frequência global.

A Figura 2.1 ilustra, respectivamente, os desvios de módulo e fase resultantes do “abandono” de zeros, como funções de A , para $\alpha = 0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}$ e π .

Alguns exemplos:

- Admitindo-se um desvio $\Delta M = 0,1$ no módulo, o módulo máximo de cada zero descartável vai de $A = \sqrt{0,19} \approx 0,4359$, para $\alpha = \frac{\pi}{2}$, até $A = 0,1$, para $\alpha = 0, \pi$.
- Admitindo-se um desvio $\Delta\theta = \frac{\pi}{20}$ na fase, o módulo máximo de cada zero descartável vai de $A = \sqrt{\sin \frac{\pi}{10}} \approx 0,5559$, para $\alpha = \frac{\pi}{2}$, até $A = \sin \frac{\pi}{20} \approx 0,1564$, para $\alpha = 0, \pi$.

¹O quadrado foi empregado por se tratar de um par de zeros, a fim de preservar a compatibilidade com os parâmetros definidos para o zero real simples.

²O fator 2 foi empregado por se tratar de um par de zeros, a fim de preservar a compatibilidade com o parâmetro definido para o zero real simples.

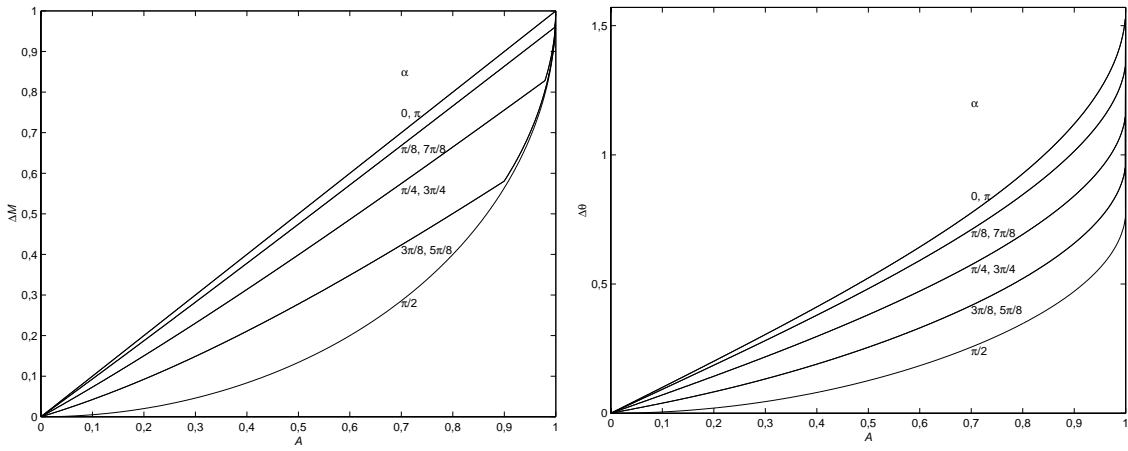


Figura 2.1: Efeito de zero sobre a transferência.

Com referência ao pólo real b , considere-se a transferência parcial

$$H_{\text{pr}}(z) = \frac{z}{z - b}, \quad |b| < 1. \quad (2.18)$$

Sua resposta em frequência pode ser descrita por

$$|H_{\text{pr}}(e^{j\omega})|^2 = \frac{1}{1 - 2b \cos \omega + b^2} \quad (2.19)$$

e

$$\angle H_{\text{pr}}(e^{j\omega}) = \text{arctg} \frac{-b \text{sen } \omega}{1 - b \cos \omega}. \quad (2.20)$$

“Abandonar” o pólo b corresponderia a fazer

$$H_{\text{pr}}(z) \approx 1. \quad (2.21)$$

Definindo-se m como o mínimo e M como o máximo valor de $|H_{\text{pr}}(e^{j\omega})|$, chega-se a

$$m = \frac{1}{1 + |b|}, \quad \text{em } \omega = \begin{cases} 0, & \text{se } b < 0 \\ \pi, & \text{se } b > 0 \end{cases} \quad (2.22)$$

e

$$M = \frac{1}{1 - |b|}, \quad \text{em } \omega = \begin{cases} \pi, & \text{se } b < 0 \\ 0, & \text{se } b > 0. \end{cases} \quad (2.23)$$

O desvio máximo de módulo, obtido pela equação (2.1), é

$$\Delta M = \frac{|b|}{1 - |b|} \quad (2.24)$$

Definindo-se $\Delta\theta$ como o máximo valor do desvio $|H_{\text{pr}}(e^{j\omega})|$, chega-se a

$$\Delta\theta = \arcsen |b|, \quad (2.25)$$

na frequência $\omega = \arccos b$.

Com isso, obtiveram-se as expressões que descrevem o efeito do pólo real b sobre a resposta em frequência global.

Com referência ao par de pólos complexos conjugados $Be^{\pm j\beta}$, considere-se a transferência parcial

$$H_{2\text{pc}}(z) = \frac{z^2}{z^2 - 2B \cos \beta z + B^2}, \quad B < 1. \quad (2.26)$$

Sua resposta em frequência pode ser descrita por

$$|H_{2\text{pc}}(e^{j\omega})|^2 = \frac{1}{1 - 4B \cos \beta \cos \omega + 2B^2[2 \cos^2 \beta + \cos(2\omega)] - 4B^3 \cos \beta \cos \omega + B^4} \quad (2.27)$$

e

$$\angle H_{2\text{pc}}(e^{j\omega}) = \arctg \frac{-[2B \cos \beta \sen \omega - B^2 \sen(2\omega)]}{1 - 2B \cos \beta \cos \omega + B^2 \cos(2\omega)}. \quad (2.28)$$

“Abandonar” os pólos $Be^{\pm j\beta}$ corresponderia a fazer

$$H_{2\text{pc}}(z) \approx 1. \quad (2.29)$$

Definindo-se m^2 como o mínimo e M^2 como o máximo valor³ de $|H_{2\text{pc}}(e^{j\omega})|$, chega-se a

$$m = \frac{1}{\sqrt{1 + 2B|\cos \beta| + B^2}}, \quad \text{em } \omega = \begin{cases} \pi, & \text{se } 0 < \beta < \frac{\pi}{2} \\ 0, & \text{se } \frac{\pi}{2} < \beta < \pi. \end{cases} \quad (2.30)$$

e

$$M = \begin{cases} \frac{1}{\sqrt{1 - 2B|\cos \beta| + B^2}}, & \text{se } |\cos \beta| > \frac{2B}{1+B^2}, \quad \text{em } \omega = \begin{cases} 0, & \text{se } 0 < \beta < \frac{\pi}{2} \\ \pi, & \text{se } \frac{\pi}{2} < \beta < \pi \end{cases} \\ \frac{1}{\sqrt{(1-B^2)|\sen \beta|}}, & \text{se } |\cos \beta| < \frac{2B}{1+B^2}, \quad \text{em } \omega = \arccos \left(\frac{1+B^2}{2B} \cos \beta \right). \end{cases} \quad (2.31)$$

O desvio máximo de módulo ΔM se obtém pela equação (2.1).

³O quadrado foi empregado por se tratar de um par de pólos, a fim de preservar a compatibilidade com os parâmetros definidos para o pólo real simples.

Definindo-se $2\Delta\theta$ como o máximo valor⁴ do desvio $|\angle H_{2pc}(e^{j\omega})|$, chega-se a

$$\Delta\theta = \frac{1}{2} \max(|\angle H_{2pc}(e^{j\omega_0})|), \quad (2.32)$$

sendo

$$\omega_0 = \arccos \frac{(1 + 3B^2) \cos \beta \pm \sqrt{(1 - 10B^2 + 9B^4) \cos^2 \beta + 8B^2(1 - B^2)}}{4B} \quad (2.33)$$

com $0 \leq \omega_0 \leq \pi$.

Com isso, obtiveram-se as expressões que descrevem o efeito dos pólos $Be^{\pm j\beta}$ sobre a resposta em frequência global.

A Figura 2.2 ilustra, respectivamente, os desvios de módulo e fase resultantes do “abandono” de pólos, como funções de B , para $\beta = 0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}$ e π .

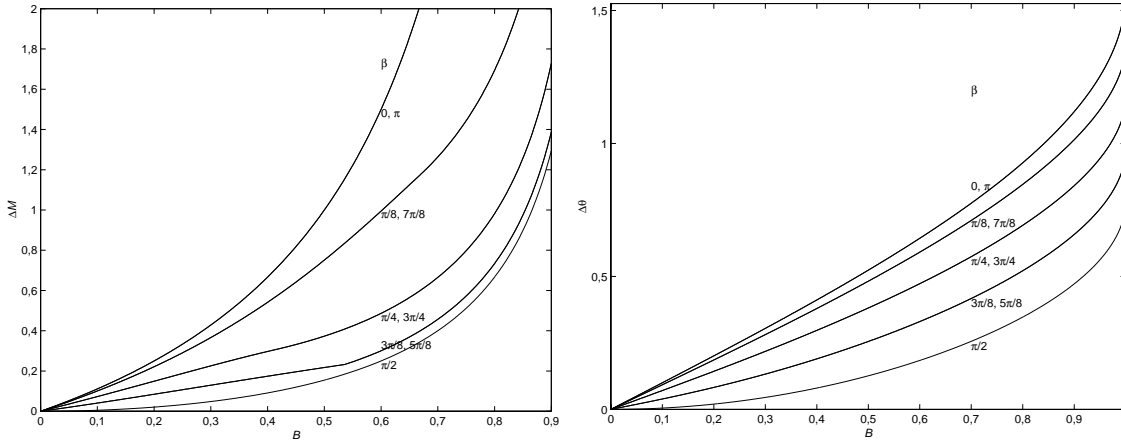


Figura 2.2: Efeito de pólo sobre a transferência.

Alguns exemplos:

- Admitindo-se um desvio $\Delta M = 0,1$ no módulo, o módulo máximo de cada pólo descartável vai de $B = \frac{\sqrt{21}}{11} \approx 0,4166$, para $\beta = \frac{\pi}{2}$, até $B = \frac{1}{11} \approx 0,0909$, para $\beta = 0, \pi$.
- Admitindo-se um desvio $\Delta\theta = \frac{\pi}{20}$ na fase, o módulo máximo de cada pólo descartável vai de $B = \sqrt{\sin \frac{\pi}{10}} \approx 0,5559$, para $\beta = \frac{\pi}{2}$, até $B = \sin \frac{\pi}{20} \approx 0,1564$, para $\beta = 0, \pi$.

⁴O fator 2 foi empregado por se tratar de um par de pólos, a fim de preservar a compatibilidade com o parâmetro definido para o pólo real simples.

Nota-se que zeros e pólos de fases próximas de zero ou próximas de π têm efeito mais significativo que zeros e pólos de fases próximas de $\frac{\pi}{2}$. Assim, é possível formular um teste simplificado que adote um único módulo como limiar de “abandono”, qualquer que seja a fase do zero ou pólo; nesse caso, deve ser usado o de pior caso, correspondente à fase nula.

2.2 Efeito de Par Zero-Pólo

Na seqüência, será avaliado o cancelamento entre zeros e pólos próximos. Pergunta a responder: qual a contribuição do quociente “abandonado” à função de transferência global? Serão examinados os casos de um zero real com um pólo real e dois zeros complexos conjugados com dois pólos complexos conjugados. Partindo do módulo e da fase da resposta em freqüência isolada da razão entre zero e pólo, serão obtidos em relação a ω o módulo mínimo m e o módulo máximo M , levando ao desvio máximo ΔM em relação ao módulo unitário (ideal), e o desvio máximo $\Delta\theta$ em relação à fase nula (ideal). Tais parâmetros dependem da localização dos zeros e pólos, em função da qual serão descritos analítica e graficamente.

Com relação ao zero real a e ao pólo real b , considere-se a transferência parcial

$$H_{zpr}(z) = \frac{z - a}{z - b}, |a|, |b| < 1. \quad (2.34)$$

Sua resposta em freqüência pode ser descrita por

$$|H_{zpr}(e^{j\omega})|^2 = \frac{1 - 2a \cos \omega + a^2}{1 - 2b \cos \omega + b^2} \quad (2.35)$$

e

$$\angle H_{zpr}(e^{j\omega}) = \text{arctg} \frac{(a - b) \text{sen } \omega}{1 - (a + b) \cos \omega + ab}. \quad (2.36)$$

“Cancelar” o pólo b com o zero a corresponderia a fazer

$$H_{zpr}(z) \approx 1. \quad (2.37)$$

Definindo-se m como o mínimo e M como o máximo valor de $|H_{zpr}(e^{j\omega})|$, chega-se a

$$m = \begin{cases} \frac{1+a}{1+b}, & \text{se } a < b, \text{ em } \omega = \pi \\ \frac{1-a}{1-b}, & \text{se } a > b, \text{ em } \omega = 0 \end{cases} \quad (2.38)$$

e

$$M = \begin{cases} \frac{1-a}{1-b}, & \text{se } a < b, \text{ em } \omega = 0 \\ \frac{1+a}{1+b}, & \text{se } a > b, \text{ em } \omega = \pi. \end{cases} \quad (2.39)$$

O desvio máximo de módulo, obtido pela equação (2.1), é

$$\Delta M = \frac{|a-b|}{1-|b|} \quad (2.40)$$

Definindo-se $\Delta\theta$ como o máximo valor do desvio $|\angle H_{zpr}(e^{j\omega})|$, chega-se a

$$\Delta\theta = \arcsen \frac{|a-b|}{1-ab}, \quad (2.41)$$

na frequência $\omega = \arccos \frac{a+b}{1+ab}$.

Com isso, obtiveram-se as expressões que descrevem o efeito de razão entre o zero real a e o pólo real b sobre a resposta em frequência global.

Com referência ao par de zeros complexos conjugados $Ae^{\pm j\alpha}$ e ao par de pólos complexos conjugados $Be^{\pm j\beta}$, considere-se a transferência parcial

$$H_{2zpc}(z) = \frac{z^2 - 2A \cos \alpha z + A^2}{z^2 - 2B \cos \beta z + B^2}, \quad A, B < 1. \quad (2.42)$$

Sua resposta em frequência pode ser descrita por

$$|H_{2zpc}(e^{j\omega})|^2 = \frac{1 - 4A \cos \alpha \cos \omega + 2A^2[2 \cos^2 \alpha + \cos(2\omega)] - 4A^3 \cos \alpha \cos \omega + A^4}{1 - 4B \cos \beta \cos \omega + 2B^2[2 \cos^2 \beta + \cos(2\omega)] - 4B^3 \cos \beta \cos \omega + B^4} \quad (2.43)$$

e

$$\angle H_{2zpc}(e^{j\omega}) = \arctg \frac{\sen 2\omega - 2A \cos \alpha \sen \omega}{\cos 2\omega - 2A \cos \alpha \cos \omega + A^2} - \arctg \frac{\sen 2\omega - 2B \cos \beta \sen \omega}{\cos 2\omega - 2B \cos \beta \cos \omega + B^2}. \quad (2.44)$$

“Cancelar” os zeros $Ae^{\pm j\alpha}$ com os pólos $Be^{\pm j\beta}$ corresponderia a fazer

$$H_{2zpc}(z) \approx 1. \quad (2.45)$$

Definindo-se m^2 como o mínimo e M^2 como o máximo valor⁵ de $|H_{2zpc}(e^{j\omega})|$, chega-se a

$$m = \sqrt{\min(|H_{2zpc}(1)|, |H_{2zpc}(-1)|, |H_{2zpc}(e^{j\omega_{11}})|, \dots, |H_{2zpc}(e^{j\omega_{1I}})|)} \quad (2.46)$$

⁵O quadrado foi empregado por se tratar de um par de zeros e um par de pólos, a fim de preservar a compatibilidade com os parâmetros definidos para o par zero-pólo real simples.

e

$$M = \sqrt{\text{máx}(|H_{2\text{zpc}}(1)|, |H_{2\text{zpc}}(-1)|, |H_{2\text{zpc}}(e^{j\omega_{11}})|, \dots, |H_{2\text{zpc}}(e^{j\omega_{1I}})|)}, \quad (2.47)$$

sendo ω_{1i} , $i = 1, \dots, I$, as I possíveis soluções de

$$\begin{aligned} & 4AB[A(1+B^2)\cos\beta - B(1+A^2)\cos\alpha]\cos^2\omega + \\ & + 2[4A^2B^2(\cos^2\alpha - \cos^2\beta) - (A^2 - B^2)(1 - A^2B^2)]\cos\omega + \\ & + A(1+A^2)[(1-B^2)^2 + 4B^2\cos^2\beta]\cos\alpha - \\ & - B(1+B^2)[(1-A^2)^2 + 4A^2\cos^2\alpha]\cos\beta = 0 \end{aligned} \quad (2.48)$$

com $0 \leq \omega_1 \leq \pi$. O desvio máximo de módulo ΔM se obtém pela equação (2.1).

Definindo-se $2\Delta\theta$ como o máximo valor⁶ do desvio $|\angle H_{2\text{zpc}}(e^{j\omega})|$, chega-se a

$$\Delta\theta = \frac{1}{2} \text{máx}(|\angle H_{2\text{zpc}}(e^{j\omega_{01}})|, \dots, |\angle H_{2\text{zpc}}(e^{j\omega_{0J}})|), \quad (2.49)$$

sendo ω_{0j} , $j = 1, \dots, J$, as J possíveis soluções de

$$\begin{aligned} & 4AB[A(1-B^2)\cos\beta - B(1-A^2)\cos\alpha]\cos^3\omega - \\ & - 2(A^2 - B^2)(1 + A^2B^2 + 4AB\cos\alpha\cos\beta)\cos^2\omega + \\ & + [1 + 3(A^2 + B^2) + A^2B^2 + 4AB\cos\alpha\cos\beta][A(1-B^2)\cos\alpha - B(1-A^2)\cos\beta]\cos\omega - \\ & - (A^2 - B^2)(1 - A^2 - B^2) + 2[B^2(1 - A^4)\cos^2\beta - A^2(1 - B^4)\cos^2\alpha] = 0 \end{aligned} \quad (2.50)$$

com $0 \leq \omega_0 \leq \pi$.

Com isso, obtiveram-se as expressões que definem o efeito da razão entre os zeros $Ae^{\pm j\alpha}$ e os pólos $Be^{\pm j\beta}$ sobre a resposta em frequência global.

Para ilustrar os desvios de módulo e fase resultantes do cancelamento aproximado de zeros por pólos, dois casos particulares serão examinados. As Figuras 2.3 e 2.4 mostram, respectivamente, os desvios de módulo e fase quando a fase dos zeros é igual à fase dos pólos como funções de A , e as Figuras 2.5 e 2.6 mostram, respectivamente, os desvios de módulo e fase quando o módulo dos zeros é igual ao módulo dos pólos como funções de α , para $B = 0, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}$ e $0,9$ e $\beta = 0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}$ e π .

Alguns exemplos:

⁶O fator 2 foi empregado por se tratar de um par de zeros e um par de pólos, a fim de preservar a compatibilidade com o parâmetro definido para o par zero-pólo real simples.

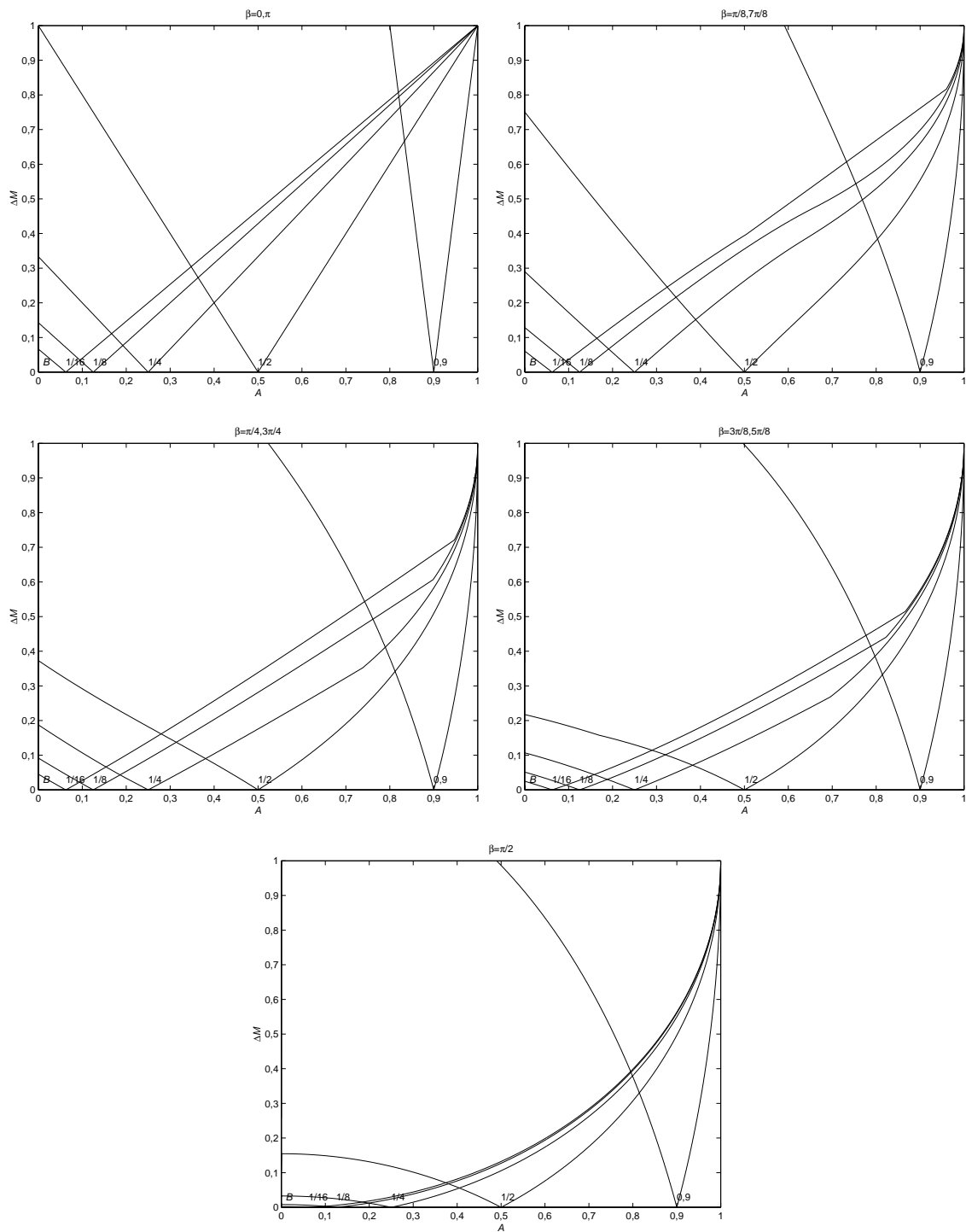


Figura 2.3: Efeito de zero e pólo de fases iguais no módulo da função de transferência.

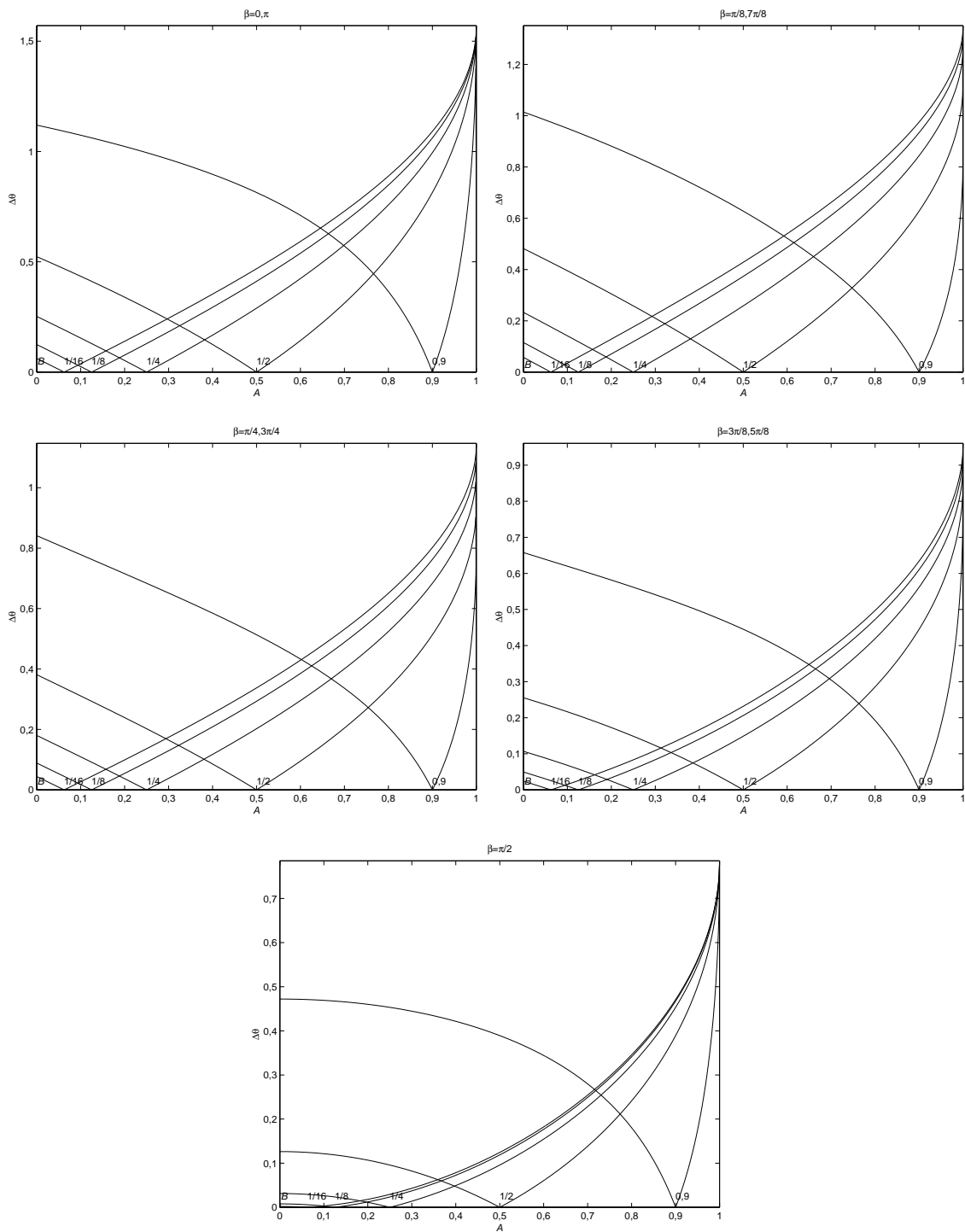


Figura 2.4: Efeito de zero e pólo de fases iguais na fase da função de transferência.

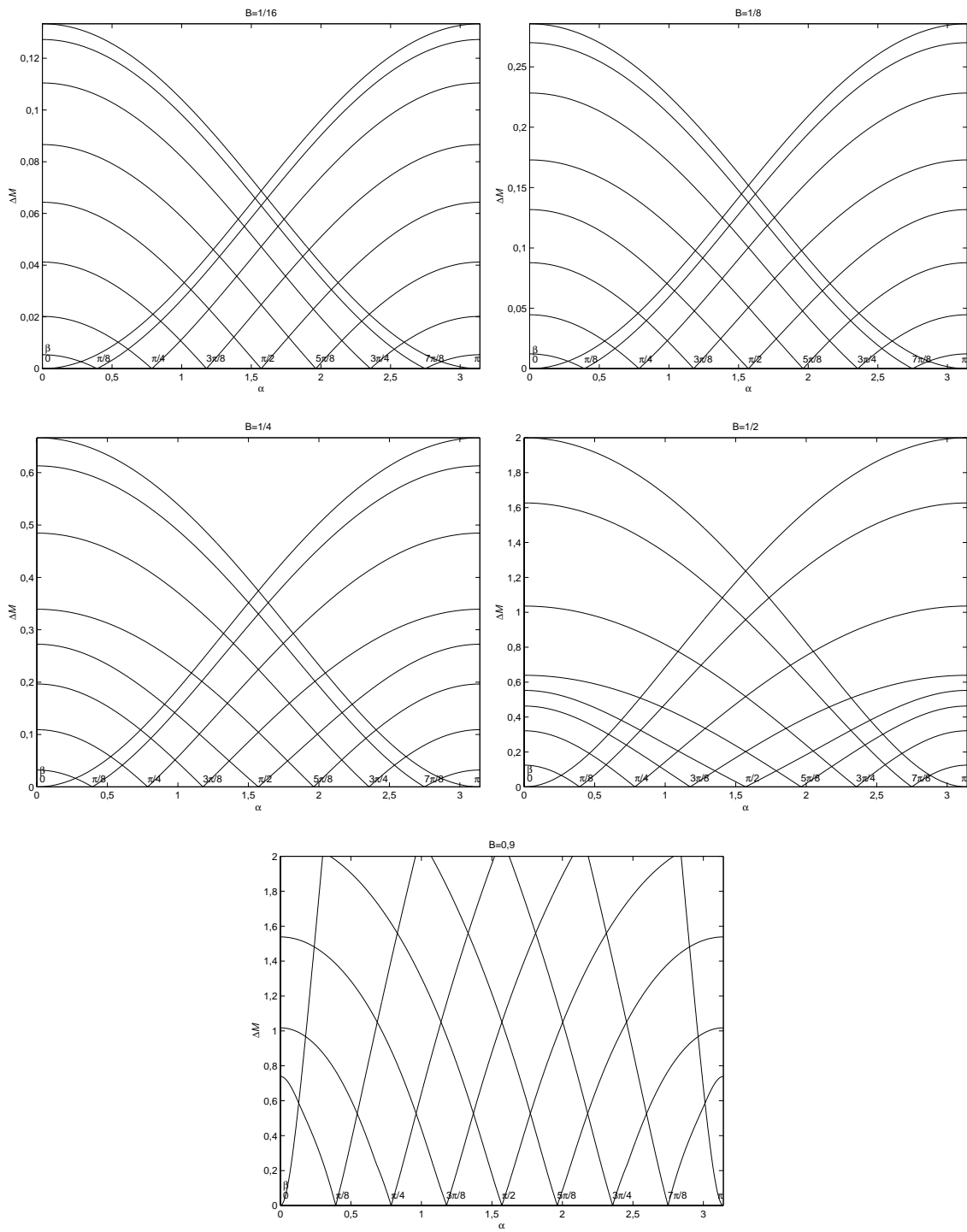


Figura 2.5: Efeito de zero e pólo de módulos iguais no módulo da função de transferência.

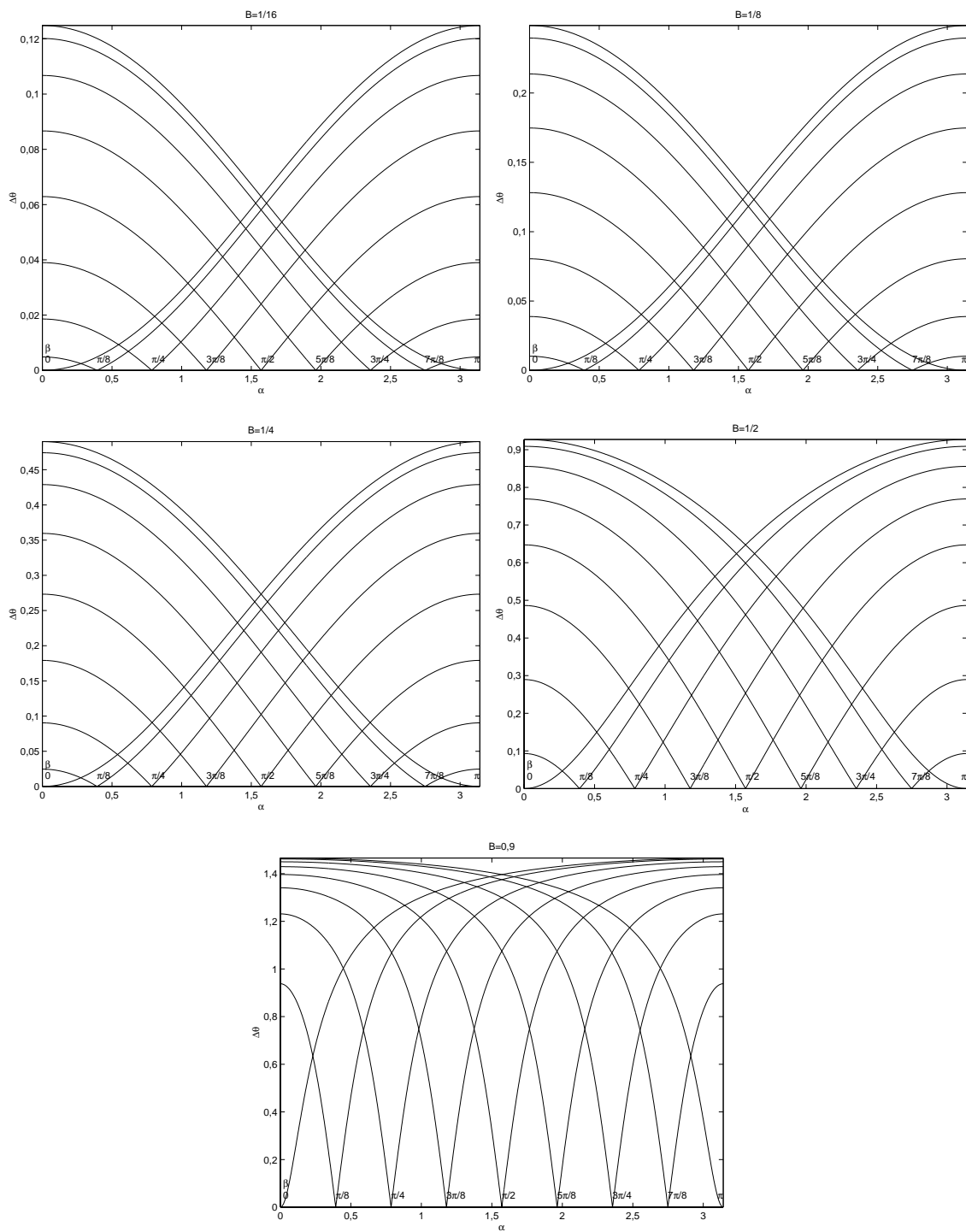


Figura 2.6: Efeito de zero e pólo de módulos iguais na fase da função de transferência.

- Admitindo-se um desvio $\Delta M = 0,1$ no módulo, o módulo de um zero que cancele aproximadamente um pólo de módulo $B = 0,9$, tendo ambos fase nula, vai de $A = 0,89$ até $A = 0,91$.
- Admitindo-se um desvio $\Delta\theta = \frac{\pi}{20}$ na fase, o módulo de um zero que cancele aproximadamente um pólo de módulo $B = 0,9$, tendo ambos fase nula, vai de $A = \frac{0,9 - \text{sen}(\frac{\pi}{20})}{1 - 0,9\text{sen}(\frac{\pi}{20})} \approx 0,8654$ até $A = \frac{0,9 + \text{sen}(\frac{\pi}{20})}{1 + 0,9\text{sen}(\frac{\pi}{20})} \approx 0,9261$.
- Admitindo-se um desvio $\Delta M = 0,1$ no módulo, a fase de um zero que cancele aproximadamente um pólo de fase $\beta = \frac{\pi}{2}$, tendo ambos módulo igual a $0,9$, vai de $\alpha \approx \frac{\pi}{2} - 0,04045$ até $\alpha \approx \frac{\pi}{2} + 0,04045$.
- Admitindo-se um desvio $\Delta\theta = \frac{\pi}{20}$ na fase, a fase de um zero que cancele aproximadamente um pólo de fase $\beta = \frac{\pi}{2}$, tendo ambos módulo igual a $0,9$, vai de $\alpha \approx \frac{\pi}{2} - 0,03345$ até $\alpha \approx \frac{\pi}{2} + 0,03345$.

No caso de fases iguais, nota-se que: para uma dada diferença relativa entre seus módulos, a razão entre zeros e pólos de fases próximas de 0 ou próximas de π tem efeito mais significativo que a razão entre zeros e pólos de fases próximas de $\frac{\pi}{2}$; e o efeito da diferença percentual entre os módulos do zero e do pólo se acentua com o aumento do módulo do pólo. Assim, é possível formular um teste simplificado que adote para o zero um único módulo como limiar de cancelamento de um pólo, qualquer que seja sua fase; então, deve ser usado o de pior caso, correspondente à fase nula.

No caso de módulos iguais, nota-se que: o efeito da diferença entre as fases do zero e do pólo se acentua com o aumento do módulo do pólo; para uma dada diferença entre suas fases, o efeito da razão entre zeros e pólos tende a independe da fase do pólo, à medida que seu módulo aumenta. Assim, é possível formular um teste simplificado que adote para o zero uma diferença de fase única como limiar de cancelamento de um pólo, qualquer que seja sua fase; então, por simplicidade, pode ser usado, ainda que não sendo o pior caso, o correspondente à fase igual a $\frac{\pi}{2}$.

No caso mais geral (e realista) em que módulos e fases são diferentes, é possível adotar como teste simplificado a combinação dos dois testes simplificados anteriores.

2.3 Exemplo

Para ilustrar o uso dos critérios sugeridos para redução da ordem de modelos, seja um modelo cuja função de transferência inclui um pólo complexo $0,9e^{j0,25\pi}$ e um zero complexo $0,89e^{j0,27\pi}$. Deseja-se saber se este par pode ser desprezado, caso se aceitem desvios $\Delta M < 0,2$ e $\Delta\theta < 0,2\pi$ na resposta em frequência global. O cálculo direto fornece $\Delta M = 0,1948$ e $\Delta\theta = 0,0889\pi$, ou seja, o par pode ser desprezado.

Adotando-se o critério combinado baseado nos módulos e fases do par, proposto no final da Seção 2.2: supondo $\alpha = \beta = \frac{\pi}{4}$, $B = 0,9$ resultaria em $\Delta M = 0,0462$ e $\Delta\theta = 0,0088\pi$; supondo $A = B = 0,9$, $\beta = \frac{\pi}{4}$ resultaria em $\Delta M = 0,1771$ e $\Delta\theta = 0,0918\pi$. Tendo satisfeito os 2 testes simultaneamente, o par seria considerado desprezível.

Alternativamente, adotando-se o critério combinado simplificado: supondo fases iguais e adotando as restrições referentes a $\beta = 0$, $B = 0,9$ resultaria em uma faixa $|A - 0,9| < 0,02$ para que se admitisse o cancelamento do par; supondo módulos iguais e adotando as restrições referentes a $\beta = \frac{\pi}{2}$, $B = 0,9$ resultaria em uma faixa $|\alpha - 0,25\pi| < 0,025\pi$ para que se admitisse o cancelamento do par. Ambas as condições tendo sido atendidas, o teste simplificado também autorizaria a redução da ordem do modelo.

Capítulo 3

Efeito da Quantização sobre a Obtenção do Modelo AR para um Sinal de Áudio

Uma situação recorrente no processamento digital de sinais de voz e áudio é a necessidade de se estimar um modelo estatístico para descrever esses sinais. Um modelo freqüentemente adotado é o processo autorregressivo (AR), obtido pela passagem de ruído branco por um filtro digital linear só-pólos. E, tipicamente, as versões discretas desses sinais são representadas digitalmente em ponto fixo. Daí o interesse em estudar o efeito do erro de quantização dos sinais sobre seu modelo AR.

Este Capítulo parte de um problema prático encontrado num contexto de restauração de sinais de áudio, mais especificamente na interpolação de amostras faltantes baseada em modelo (v. Seção 7.8.2), o qual mostra-se ter origem na estimativa polarizada do modelo AR para sinais que sofreram quantização. Revisam-se aspectos relacionados com o desvio dos coeficientes do modelo AR devido à quantização do sinal modelado, permitindo descrever o desvio nos coeficientes em função do número de bits da representação do sinal e, potencialmente, corrigi-lo. Investiga-se, ainda, a sensibilidade da posição dos pólos do filtro gerador associado em relação aos coeficientes, explicando a ocorrência preferencial do desvio de pólos em dadas regiões do espectro.

3.1 Apresentação do Problema

3.1.1 Indício: Interpolação de Amostras Faltantes Baseada em Modelo

Em determinado estágio deste trabalho, no qual se estudava a interpolação de amostras faltantes num sinal baseada em modelo (Subseção 7.8.2), verificou-se que a quantização do sinal pode prejudicar a interpolação.

O exemplo que denunciou o problema consistia em interpolar 400 amostras num sinal formado pela soma de 2 senóides de freqüências distintas. O resultado, reproduzido na Figura 3.1, indica que a senóide de freqüência inferior não foi interpolada entre as amostras 375 e 775.

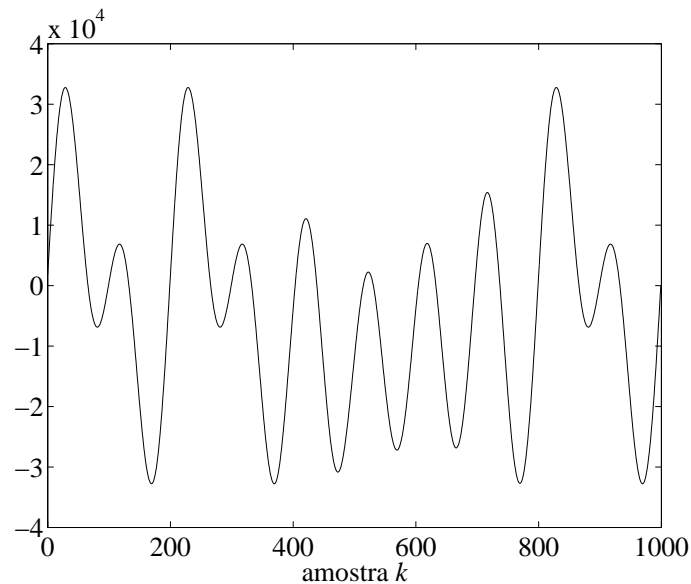


Figura 3.1: Sinal quantizado interpolado.

Testes preliminares revelaram que a anomalia era gradativamente eliminada com o aumento do número de bits com que se quantizava o sinal a interpolar (originando-se, portanto, nesta quantização) e ocorria preferencialmente com freqüências baixas. Por outro lado, não era afetada pelo número de amostras a interpolar nem pelo tamanho do bloco, sugerindo que poderia não ser propriamente a interpolação a falhar, mas apenas o cálculo do modelo nela envolvido.

3.1.2 Origem: Cálculo do Modelo AR

O sinal do exemplo citado na Subseção 3.1.1 é um processo AR de 4a. ordem degenerado na soma de 2 senóides, cujo filtro gerador teórico associado teria pólos $e^{\pm j\frac{\pi}{100}}$ e $e^{\pm j\frac{\pi}{50}}$. Calcularam-se os pólos referentes à estimativa por Mínimos Quadrados (LS) dos coeficientes do modelo AR para o bloco completo do sinal quantizado em 16 bits, obtendo-se aproximadamente 1, $-0,093$ e $e^{\pm j\frac{\pi}{53}}$. Isso confirmou a observação de que só a senóide na frequência superior era interpolada—de fato, só ela era modelada, enquanto a outra dava origem a 2 pólos reais.

Com isso, o objetivo passou a ser explicar a relação entre o erro na modelagem e a quantização do sinal a modelar, além do motivo da ocorrência preferencial do problema em baixas frequências.

3.1.3 Testes Adicionais

Abandonou-se o caso especial da senóide e efetuaram-se testes sobre processos AR genéricos.

Processos de primeira ordem não trouxeram informações de interesse.

Independendo de quantização, processos de segunda ordem com 2 pólos complexos forneceram indicações de cunho geral:

- O modelo falha para fase θ exageradamente baixa, abaixo de $\frac{\pi}{1000}$; a título de ilustração, no caso de áudio amostrado a 44,1 kHz, uma senóide de frequência $\frac{\pi}{1000}$ corresponde a 22,05 Hz, ou seja, no limite inferior da audição humana.
- A estimativa da localização dos pólos é tão menos precisa quanto mais o módulo m se afasta da unidade; isso não é necessariamente um problema, uma vez que pólos de módulo reduzido contribuem pouco na resposta em frequência (v. Seção 2.1).

Passando ao caso de quarta ordem quantizado, em que se enquadra o exemplo já discutido das 2 senóides, fizeram-se outras constatações.

- Embora houvesse a preferência genérica pelas fases reduzidas, a anomalia ocorria em função dos valores das duas fases tomados em conjunto; por exemplo, $\theta_1 = \frac{\pi}{100}$ não era modelada se combinada com $\frac{\pi}{100} < \theta_2 < \frac{\pi}{20}$, enquanto que $\theta_1 = \frac{\pi}{200}$ não era modelada se combinada com $\frac{\pi}{200} < \theta_2 < \frac{\pi}{10}$.

- O uso de modelo de ordem superior à teórica pode solucionar o problema; no caso real, isso apenas significa genericamente superdimensionar o modelo.

O comportamento do estimador para ordens mais altas segue as linhas esboçadas acima.

3.2 Premissas Teóricas

3.2.1 Descrição do Erro de Quantização

Considere-se inicialmente um sinal aleatório de média zero $x(k)$ ¹, com amplitude contínua limitada por

$$|x(k)| < X_{\text{máx}}, \quad (3.1)$$

para o qual se obtém uma representação uniformemente quantizada em L níveis com passo δ , tal que

$$x_q(k) = \left\lceil \frac{x(k)}{\delta} + \frac{1}{2} \right\rceil \delta. \quad (3.2)$$

O erro de quantização

$$e_q(x) \triangleq x_q(k) - x(k) \quad (3.3)$$

terá média zero e amplitude contínua limitada por

$$|e_q(k)| \leq \delta. \quad (3.4)$$

Atribuindo-se ao erro distribuição uniforme, sua função densidade de probabilidade é

$$f(e_q(k)) = \begin{cases} \frac{1}{\delta}, & -\frac{\delta}{2} < e_q(k) \leq \frac{\delta}{2} \\ 0, & (e_q(k) \leq -\frac{\delta}{2}) \wedge (e_q(k) > \frac{\delta}{2}), \end{cases} \quad (3.5)$$

¹A rigor, a notação utilizada para uma variável ou um processo aleatório [18, 19] deve ser em grifo (por exemplo, em letra maiúscula), contrastando com um valor da variável ou uma realização do processo (por exemplo, em letra minúscula). Por simplicidade, não será feita distinção entre os dois casos ao longo do trabalho.

permitindo obter [30] sua variância

$$\sigma_{e_q}^2 = \frac{\delta^2}{12}. \quad (3.6)$$

No caso de se adotar representação binária em ponto fixo para $x_q(k)$, pode-se supor $x(k)$ previamente escalado de forma que

$$X_{\text{máx}} = 1. \quad (3.7)$$

Sendo B o número total de bits da representação, então

$$\delta = 2^{-(B-1)} \quad (3.8)$$

e, portanto,

$$\sigma_{e_q}^2 = \frac{2^{-2B}}{3}. \quad (3.9)$$

3.2.2 Estimação LS de Parâmetros de Modelo AR

Considere-se o sinal aleatório $x(k)$ descritível por um processo AR de ordem I

$$x(k) = \sum_{i=1}^I a(i)x(k-i) + e(k), \quad (3.10)$$

onde a excitação $e(k)$ é ruído branco com média zero e variância σ_e^2 . Redescrivendo-se o problema como a predição linear de $x(k)$ a partir de suas I amostras anteriores, $e(k)$ seria o erro de predição correspondente.

A estimativa LS \mathbf{a}^{LS} do vetor de coeficientes

$$\mathbf{a} \triangleq [a(1) \quad a(2) \quad \cdots \quad a(I)]^T, \quad (3.11)$$

conhecendo-se $x(k)$, $k = 0, 1, \dots, K-1$, é obtida pela minimização da energia $\sum_{k=I}^{K-1} e^2(k)$ do erro de predição [16], e vale

$$\mathbf{a}^{\text{LS}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{x}_1, \quad (3.12)$$

sendo

$$\mathbf{x}_1 \triangleq [x(I) \quad x(I+1) \quad \cdots \quad x(K-1)]^T \quad (3.13)$$

e

$$\mathbf{X} \triangleq \begin{bmatrix} x(I-1) & x(I-2) & \cdots & x(0) \\ x(I) & x(I-1) & \cdots & x(1) \\ \vdots & \vdots & \ddots & \vdots \\ x(K-2) & x(K-3) & \cdots & x(K-I-1) \end{bmatrix}. \quad (3.14)$$

3.2.3 Desvio na Estimativa de Parâmetros devido a Ruído Aditivo

Considere-se novamente o sinal aleatório $x(k)$ definido na Subseção 3.2.2.

Conhecendo-se a estatística de $x(k)$, seria possível calcular a solução MMSE (de mínimo erro quadrático médio) \mathbf{a}^{MMSE} para o vetor de coeficientes pela minimização da potência do erro de predição $E[e^2(k)]$ [31], resultando na conhecida solução das equações de Yule-Walker

$$\mathbf{a}^{\text{MMSE}} = \mathbf{R}_x^{-1} \mathbf{r}_x, \quad (3.15)$$

onde

$$\mathbf{R}_x \triangleq \begin{bmatrix} r_x(0) & r_x(1) & \cdots & r_x(I-1) \\ r_x(1) & r_x(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_x(1) \\ r_x(I-1) & \cdots & r_x(1) & r_x(0) \end{bmatrix} \quad (3.16)$$

e

$$\mathbf{r}_x \triangleq [r_x(1) \ r_x(2) \ \cdots \ r_x(I)]^T, \quad (3.17)$$

com

$$r_x(\Delta) \triangleq E[x(k)x(k-\Delta)]. \quad (3.18)$$

Se, entretanto, se conhecesse apenas a estatística de uma versão de $x(k)$ contaminada por ruído branco aditivo $n(k)$ de variância σ_n^2 ,

$$y(k) \triangleq x(k) + n(k), \quad (3.19)$$

a solução MMSE $\hat{\mathbf{a}}^{\text{MMSE}}$ obtida por $y(k)$

$$\hat{\mathbf{a}}^{\text{MMSE}} = \mathbf{R}_y^{-1} \mathbf{r}_y, \quad (3.20)$$

onde

$$\mathbf{R}_y \triangleq \begin{bmatrix} r_y(0) & r_y(1) & \cdots & r_y(I-1) \\ r_y(1) & r_y(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_y(1) \\ r_y(I-1) & \cdots & r_y(1) & r_y(0) \end{bmatrix} \quad (3.21)$$

e

$$\mathbf{r}_y \triangleq [r_y(1) \ r_y(2) \ \cdots \ r_y(I)]^T, \quad (3.22)$$

com

$$r_y(\Delta) \triangleq E[y(k)y(k-\Delta)], \quad (3.23)$$

seria uma versão polarizada daquela obtida por $x(k)$. Sendo $n(k)$ não-correlacionado com $x(k)$ e $e(k)$, e sendo $e(k)$ não-correlacionado com amostras passadas de $x(k)$, pode-se mostrar [32] que

$$\mathbf{a}^{\text{MMSE}} = (\mathbf{I} - \sigma_n^2 \mathbf{R}_y^{-1})^{-1} \hat{\mathbf{a}}^{\text{MMSE}}. \quad (3.24)$$

Num caso real, $r_y(\Delta)$ não seriam conhecidas, e as soluções MMSE poderiam dar lugar às estimativas LS, que delas se aproximam assintoticamente com $K \rightarrow \infty$. Para K elevado,

$$\mathbf{a}^{\text{LS}} \approx (\mathbf{I} - \sigma_n^2 K (\mathbf{Y}^T \mathbf{Y})^{-1})^{-1} \hat{\mathbf{a}}^{\text{LS}}, \quad (3.25)$$

sendo

$$\hat{\mathbf{a}}^{\text{LS}} = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{x}_1 \quad (3.26)$$

a estimativa LS dos coeficientes AR obtida a partir de $y(k)$,

$$\mathbf{y}_1 \triangleq [y(I) \ y(I+1) \ \cdots \ y(K-1)]^T \quad (3.27)$$

e

$$\mathbf{Y} \triangleq \begin{bmatrix} y(I-1) & y(I-2) & \cdots & y(0) \\ y(I) & y(I-1) & \cdots & y(1) \\ \vdots & \vdots & \ddots & \vdots \\ y(K-2) & y(K-3) & \cdots & y(K-I-1) \end{bmatrix}. \quad (3.28)$$

Essa expressão permite, potencialmente, corrigir a polarização da estimativa de coeficientes devido à contaminação do sinal observado por ruído branco aditivo, desde que se conheça a potência deste.

3.2.4 Desvio dos Pólos do Filtro Gerador devido ao Desvio nos Parâmetros do Modelo AR

Na tentativa de associar quantitativamente o desvio dos pólos do filtro gerador de um processo AR ao desvio dos coeficientes do processo, são apresentadas a seguir as expressões para a sensibilidade dos módulos e fases dos pólos em relação aos coeficientes, para as ordens mais baixas, definindo-se genericamente a sensibilidade de c em relação a d como

$$S_d^c \triangleq \frac{\frac{\partial c}{c}}{\frac{\partial d}{d}} = \frac{\partial c}{\partial d} \cdot \frac{d}{c}. \quad (3.29)$$

As expressões serão explicitadas em função dos próprios módulos e fases, a cujos valores se quer associar sua sensibilidade.

Para o filtro de primeira ordem e pólo p com

$$H(z) = \frac{z}{z - p} = \frac{z}{z - a_1}, \quad (3.30)$$

$$S_{a_1}^p = 1. \quad (3.31)$$

Para o filtro de segunda ordem e pólos $me^{\pm j\theta}$ com

$$H(z) = \frac{z^2}{z^2 - 2m \cos \theta z + m^2} = \frac{z^2}{z^2 - a_1 z - a_2}, \quad (3.32)$$

as sensibilidades são

$$S_{a_1}^m = 1, \quad (3.33)$$

$$S_{a_2}^m = \frac{1}{2}, \quad (3.34)$$

$$S_{a_1}^\theta = -\frac{\cos \theta}{\theta \text{sen} \theta} \quad (3.35)$$

e

$$S_{a_2}^\theta = 0. \quad (3.36)$$

Para o filtro de terceira ordem e pólos p e $me^{\pm j\theta}$ com

$$H(z) = \frac{z^3}{(z - p)(z^2 - 2m \cos \theta z + m^2)} = \frac{z^3}{z^3 - a_1 z^2 - a_2 z - a_3}, \quad (3.37)$$

as sensibilidades são

$$S_{a_1}^p = \frac{2m \cos \theta + p}{p}, \quad (3.38)$$

$$S_{a_2}^p = \frac{m + 2p \cos \theta}{2p \cos \theta}, \quad (3.39)$$

$$S_{a_3}^p = 1, \quad (3.40)$$

$$S_{a_1}^m = \frac{2m \cos \theta + p}{2m \cos \theta}, \quad (3.41)$$

$$S_{a_2}^m = \frac{m + 2p \cos \theta}{2m + 2p \cos \theta}, \quad (3.42)$$

$$S_{a_3}^m = \frac{1}{2}, \quad (3.43)$$

$$S_{a_1}^\theta = -\frac{2 \cos \theta + p}{2m\theta \operatorname{sen} \theta}, \quad (3.44)$$

$$S_{a_2}^\theta = -\frac{m + 2p \cos \theta}{2p\theta \operatorname{sen} \theta} \quad (3.45)$$

e

$$S_{a_3}^\theta = 0. \quad (3.46)$$

Para o filtro de quarta ordem e pólos $m_1 e^{\pm j\theta_1}$ e $m_2 e^{\pm j\theta_2}$ com

$$H(z) = \frac{z^4}{(z^2 - 2m_1 \cos \theta_1 z + m_1^2)(z^2 - 2m_2 \cos \theta_2 z + m_2^2)} = \frac{z^4}{z^4 - a_1 z^3 - a_2 z^2 - a_3 z - a_4}, \quad (3.47)$$

as sensibilidades são

$$S_{a_1}^{m_1} = \frac{m_1 \cos \theta_1 + m_2 \cos \theta_2}{m_1 \cos \theta_1}, \quad (3.48)$$

$$S_{a_2}^{m_1} = \frac{m_1^2 + m_2^2 + 4m_1 m_2 \cos \theta_1 \cos \theta_2}{2m_1^2 + 4m_1 m_2 \cos \theta_1 \cos \theta_2}, \quad (3.49)$$

$$S_{a_3}^{m_1} = \frac{m_1 \cos \theta_2 + m_2 \cos \theta_1}{2m_1 \cos \theta_2 + m_2 \cos \theta_1}, \quad (3.50)$$

$$S_{a_4}^{m_1} = \frac{1}{2}, \quad (3.51)$$

$$S_{a_1}^{\theta_1} = -\frac{m_1 \cos \theta_1 + m_2 \cos \theta_2}{m_1 \theta_1 \text{sen} \theta_1}, \quad (3.52)$$

$$S_{a_2}^{\theta_1} = -\frac{m_1^2 + m_2^2 + 4m_1 m_2 \cos \theta_1 \cos \theta_2}{4m_1 m_2 \theta_1 \text{sen} \theta_1 \cos \theta_2}, \quad (3.53)$$

$$S_{a_3}^{\theta_1} = -\frac{m_1 \cos \theta_2 + m_2 \cos \theta_1}{m_2 \theta_1 \text{sen} \theta_1}, \quad (3.54)$$

$$S_{a_4}^{\theta_1} = 0, \quad (3.55)$$

$$S_{a_1}^{m_2} = \frac{m_1 \cos \theta_1 + m_2 \cos \theta_2}{m_2 \cos \theta_2}, \quad (3.56)$$

$$S_{a_2}^{m_2} = \frac{m_1^2 + m_2^2 + 4m_1 m_2 \cos \theta_1 \cos \theta_2}{2m_2^2 + 4m_1 m_2 \cos \theta_1 \cos \theta_2}, \quad (3.57)$$

$$S_{a_3}^{m_2} = \frac{m_1 \cos \theta_2 + m_2 \cos \theta_1}{m_1 \cos \theta_2 + 2m_2 \cos \theta_1}, \quad (3.58)$$

$$S_{a_4}^{m_2} = \frac{1}{2}, \quad (3.59)$$

$$S_{a_1}^{\theta_2} = -\frac{m_1 \cos \theta_1 + m_2 \cos \theta_2}{m_2 \theta_2 \text{sen} \theta_2}, \quad (3.60)$$

$$S_{a_2}^{\theta_2} = -\frac{m_1^2 + m_2^2 + 4m_1 m_2 \cos \theta_1 \cos \theta_2}{4m_1 m_2 \theta_2 \cos \theta_1 \text{sen} \theta_2}, \quad (3.61)$$

$$S_{a_3}^{\theta_2} = -\frac{m_1 \cos \theta_2 + m_2 \cos \theta_1}{m_1 \theta_2 \text{sen} \theta_2}, \quad (3.62)$$

e

$$S_{a_4}^{\theta_2} = 0. \quad (3.63)$$

Esses casos simples são indicativos do comportamento para ordens mais elevadas.

3.3 Explicações e Solução

Retornando à questão da relação entre a quantização e o desvio nos coeficientes do modelo AR, é possível ligar o exposto nas Subseções 3.2.4 e 3.2.2 se o ruído aditivo for o próprio erro de quantização, ou seja,

$$n(k) = e_q(k), \quad (3.64)$$

com

$$y(k) = x_q(k). \quad (3.65)$$

Nesse caso, é fácil obter a variância do ruído

$$\sigma_n^2 = \sigma_{e_q}^2 = \frac{2^{-2B}}{3}, \quad (3.66)$$

que permite potencialmente realizar a correção na estimativa polarizada de \mathbf{a} através de

$$\mathbf{a}^{\text{LS}} \approx \left(\mathbf{I} - \frac{2^{-2B}}{3} K(\mathbf{X}_q^T \mathbf{X}_q)^{-1} \right)^{-1} \hat{\mathbf{a}}^{\text{LS}}, \quad (3.67)$$

com

$$\hat{\mathbf{a}}^{\text{LS}} = (\mathbf{X}_q^T \mathbf{X}_q)^{-1} \mathbf{X}_q^T \mathbf{x}_{q1}, \quad (3.68)$$

sendo

$$\mathbf{x}_{q1} \triangleq [x_q(I) \quad x_q(I+1) \quad \cdots \quad x_q(K-1)]^T \quad (3.69)$$

e

$$\mathbf{X}_q \triangleq \begin{bmatrix} x_q(I-1) & x_q(I-2) & \cdots & x_q(0) \\ x_q(I) & x_q(I-1) & \cdots & x_q(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_q(K-2) & x_q(K-3) & \cdots & x_q(K-I-1) \end{bmatrix}. \quad (3.70)$$

Isso pode ser confirmado praticamente. A fim de que seja atendida a hipótese de que o erro de quantização é ruído branco, será examinado o exemplo de um sinal aleatório que se aproxime do sinal com 2 senóides que originou toda a discussão. A Tabela 3.1 mostra os pólos originais do filtro gerador de um processo AR, juntamente com os que resultam das estimativas LS dos coeficientes do processo, sem

e com correção. Os elementos duplos se referem a 2 pólos reais onde se esperavam 2 pólos complexos conjugados. Vê-se que os pólos são reposicionados com sucesso pelo recálculo do modelo de acordo com a equação (3.67). Deve ser notado que a estimação corrigida pressupõe bom condicionamento das matrizes e blocos suficientemente longos de dados. Conforme o módulo dos pólos e o intervalo em que se possa atribuir estacionaridade ao sinal, tais condições podem não ser atendidas.

Tabela 3.1: Pólos do filtro gerador de um processo AR.

pólos	m_1	θ_1	m_2	θ_2
teóricos	0,99	$\pm \frac{\pi}{100}$	0,99	$\pm \frac{\pi}{50}$
estimados	1 e 0,33	0 e π	1	$\pm \frac{\pi}{63}$
corrigidos	0,99	$\pm \frac{\pi}{105}$	0,99	$\pm \frac{\pi}{51}$

Por fim, quanto à questão da ocorrência preferencial do problema em baixas frequências, basta que se recorra à Subseção 3.2.4. Das expressões para as sensibilidades, o que se pode inferir é que, no caso geral, as variações das fases de muito baixo valor com os coeficientes (exceto o último) são extremamente elevadas, pela ocorrência sistemática de termos da forma $\theta \text{sen} \theta$ nos denominadores de S_a^θ . Já a interdependência entre os valores das fases capazes de propiciar o problema decorre da forma com que elas compõem conjuntamente o desvio dos pólos; os denominadores das expressões para $S_{a_2}^\theta$ são exemplos indicativos disso.

Estes resultados têm aplicação direta em problemas de restauração de áudio que envolvam a modelagem AR de sinais quantizados, incluindo a interpolação de amostras faltantes.

Capítulo 4

Modelo para um Processo AR Analisado em Oitavas por Banco de Filtros

A modelagem proposta neste Capítulo tem como alvo a realização em sub-bandas do processamento digital de sinais baseado em modelo.

Inicialmente, tendo em vista particularmente esta aplicação, examinam-se as possíveis vantagens da realização da restauração de áudio em sub-bandas. Deriva-se um modelo geral simples na taxa reduzida para as sub-bandas resultantes da análise de um processo autorregressivo com média móvel (ARMA) em N bandas por banco de filtros. Em seguida, detalha-se o caso particular da análise de um processo autorregressivo (AR) em oitavas por banco de filtros, recorrente em processamento de áudio. Por fim, esse modelo tem sua validade aferida e enseja importantes interpretações, com reflexos sobre a redução de sua complexidade.

4.1 Motivação: Restauração de Áudio em Sub-Bandas

Da mesma forma que a codificação e a compressão de áudio aplicam princípios de Psicoacústica para determinar “quanto o áudio pode ser piorado sem que se perceba” [33], a restauração de áudio pode utilizá-los para determinar “quanto o áudio precisa ser melhorado que ainda se perceba”, a fim de evitar esforço computacional

desnecessário e até mesmo perda de qualidade por excessivo processamento. Um princípio fundamental é o Mascaramento [34], propriedade de um som inibir a percepção de outro próximo no tempo ou na frequência. Sabe-se que só uma estreita faixa de frequências, chamada Banda Crítica [35], em torno de um tom contribui para o mascaramento deste. Um conjunto fixo de bandas críticas é convencionalmente adotado para modelar a percepção sonora, levando a uma configuração correspondente em sub-bandas.

Como já visto no Capítulo 1, a restauração de áudio proveniente de gravações deterioradas [16] lida com problemas que podem ser genericamente classificados em localizados ou globais, dos quais os exemplos mais comuns são, respectivamente, *clicks* (ruído impulsivo) e *hiss* (ruído de fundo de banda larga). Métodos de supressão de ruído impulsivo geralmente usam modelos autorregressivos [36, 16] e métodos de redução de ruído de fundo se baseiam frequentemente em sistemas em sub-bandas, de forma direta [36] ou indireta [37]. Um sistema multitaxa adequado pode favorecer a realização da supressão de *clicks* baseada em modelo em conjunto com a redução de *hiss* em sub-bandas.

Os distúrbios podem, às vezes, se encontrar preferencialmente em determinadas faixas de frequência; então, o processamento em sub-bandas pode possibilitar redução de processamento ao dar mais “atenção” às sub-bandas mais afetadas.

Pode-se conseguir ainda maior redução de esforço computacional se as operações mais complexas dos sistemas em banda completa tiverem equivalentes em sub-bandas que requeiram no total menor processamento.

4.2 Caso Geral: Modelo para um Processo ARMA Dividido em Sub-Bandas

Nesta Seção, é derivado um modelo na taxa reduzida para os sinais resultantes da análise de um processo ARMA em sub-bandas.

Inicialmente, seja o sistema multitaxa [38, 39] em N bandas mostrado na Figura 4.1, empregado na análise do processo ARMA

$$s(k) = \sum_{i=1}^I a(i)s(k-i) + \sum_{j=0}^J b(j)e(k-j). \quad (4.1)$$

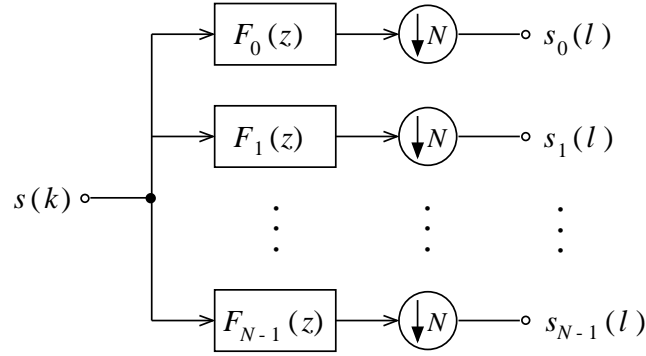


Figura 4.1: Processo ARMA analisado em N bandas por banco de filtros com decimação.

Para obtenção do modelo, seguem-se os passos mostrados na Figura 4.2.

- (a) Considerar o subsistema correspondente à sub-banda n .
- (b) Representar $s(k)$ explicitamente pelo filtro IIR adequado tendo o ruído branco $e(k)$ como entrada, e agrupá-lo com o filtro de análise $F_n(z)$.
- (c) Obter a decomposição polifásica [39] para o filtro resultante $\bar{F}_n(z)$.
- (d) Mover o decimador para antes do somador.
- (e) Para cada polifase p , mover o decimador para antes do filtro $\bar{F}_{np}(z)$ [38], representando explicitamente como $e_p(l)$ a respectiva componente de $e(k)$.
- (f) Sendo σ_e^2 a variância de $e(k)$, é fácil concluir que $e_p(l)$, $p = 0, 1, \dots, N - 1$, são processos de ruído branco mutuamente decorrelacionados, com variâncias

$$E[e_p^2(l)] = E[e^2(Nl + p)] = E[e^2(k)] = \sigma_e^2. \quad (4.2)$$

Uma vez que os sinais $s_{np}(l)$, $p = 0, 1, \dots, N - 1$, são versões coloridas de $e_p(l)$, respectivamente, também são mutuamente decorrelacionados. Então, a autocorrelação de $s_n(l)$ pode ser escrita sem os termos cruzados como

$$R_{s_n s_n}(\Delta) = \sum_{p=0}^{N-1} R_{s_{np} s_{np}}(\Delta), \quad (4.3)$$

levando à densidade espectral de potência [18, 19]

$$\mathcal{S}_{s_n s_n}(z) = \sum_{p=0}^{N-1} \mathcal{S}_{s_{np} s_{np}}(z) = \sum_{p=0}^{N-1} \bar{F}_{np}(z) \bar{F}_{np}(z^{-1}) \sigma_e^2 = \mathcal{S}_{s_n s_n}(-z). \quad (4.4)$$

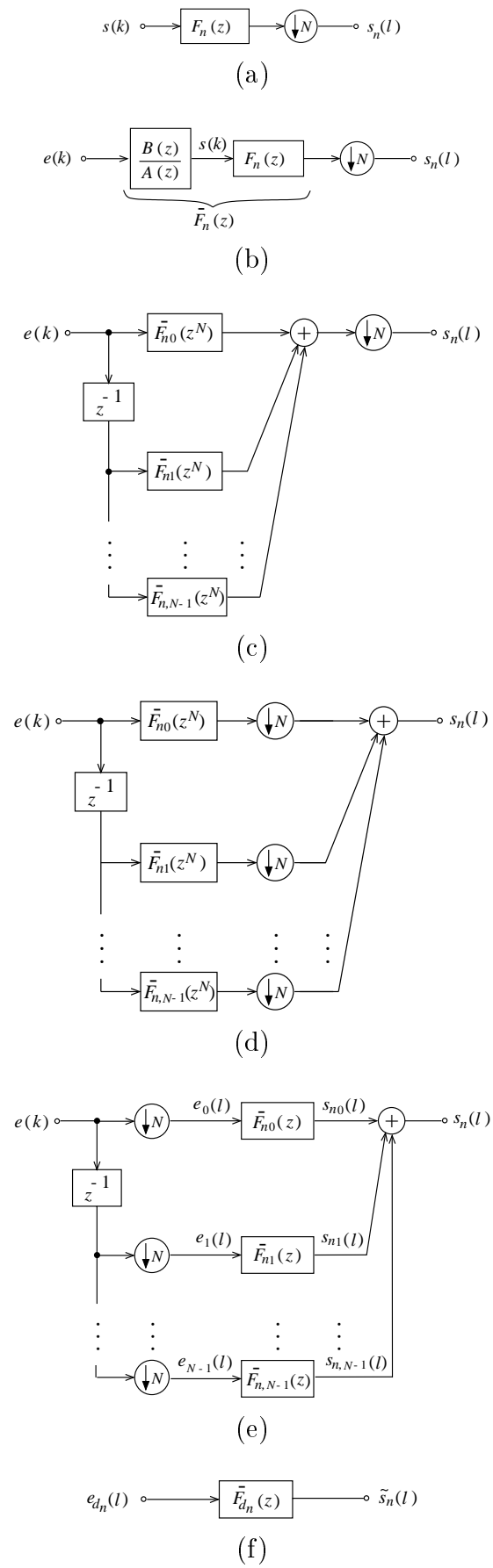


Figura 4.2: Geração de $s_n(l)$: da representação multitaxa ao modelo na taxa reduzida.

Fatorando-se a expressão 4.4 na forma

$$\mathcal{S}_{s_n s_n}(z) \triangleq \bar{F}_{d_n}(z) \bar{F}_{d_n}(z^{-1}) \sigma_e^2, \quad (4.5)$$

obtém-se a representação de $\tilde{s}_n(l)$ —equivalente a $s_n(l)$ quanto à densidade espectral de potência—como um processo ARMA definido pelo filtro de fase mínima $\bar{F}_{d_n}(z)$ ¹ tendo como entrada o ruído branco $e_{d_n}(l)$ ² de variância σ_e^2 .

Então, um modelo equivalente completo consistiria de N subsistemas similares ao mostrado em paralelo, um para cada sub-banda. Entretanto, esse modelo não descreve a correlação entre diferentes sub-bandas. A seguir, obtém-se tal descrição.

Sejam duas sub-bandas distintas n_1 e n_2 , cada uma descrita na forma da Figura 4.2(e). A correlação cruzada entre seus sinais é dada por

$$R_{s_{n_1} s_{n_2}}(\Delta) = \sum_{p=0}^{N-1} E[s_{n_1 p}(l) s_{n_2 p}(l + \Delta)] = \sum_{p=0}^{N-1} E[(\bar{f}_{n_1 p} * e_p)(l) (\bar{f}_{n_2 p} * e_p)(l + \Delta)], \quad (4.6)$$

resultando na densidade espectral de potência cruzada

$$\mathcal{S}_{s_{n_1} s_{n_2}}(z) = \left[\sum_{p=0}^{N-1} \bar{F}_{n_1 p}(z) \bar{F}_{n_2 p}(z^{-1}) \right] \sigma_e^2. \quad (4.7)$$

Isso completa o modelo.

4.3 Modelo para um Processo AR Dividido em Oitavas

Aqui, examina-se com detalhe este caso particular do modelo derivado na Seção 4.2, tanto por seu particular interesse no processamento de áudio quanto pela facilidade interpretativa ensejada por sua modularidade.

Seja o processo autorregressivo $s(k)$ definido por

$$s(k) = \sum_{i=1}^I a(i) s(k - i) + e(k), \quad (4.8)$$

onde $e(k)$ é ruído branco, do qual se efetua a divisão em oitavas por um banco de filtros de análise em árvore binária [39], conforme indicado na Figura 4.3.

¹O índice d se refere à decimação por N , evitando confusão com $\bar{F}_n(z)$ anteriormente definido.

²O índice d foi incorporado para evitar confusão com $e_n(l)$ anteriormente definido.

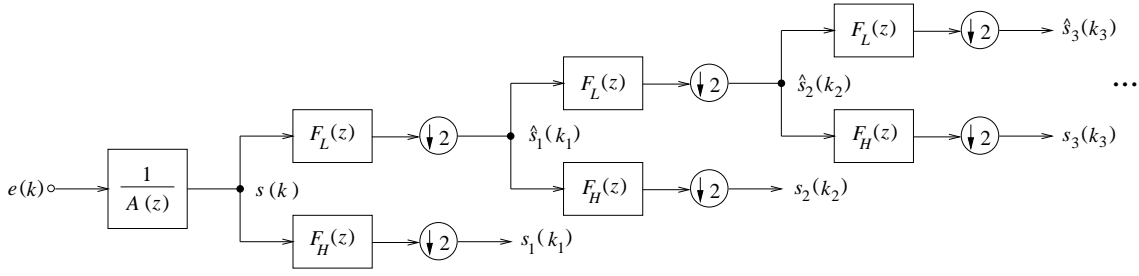


Figura 4.3: Processo autorregressivo analisado em oitavas por banco de filtros.

A cada estágio, os filtros de análise são o passa-altas $F_H(z)$ e o passa-baixas $F_L(z)$, supostos FIR de ordem Q . Após M estágios, há $M + 1$ sub-bandas, de $s_1(k_1)$ até $s_M(k_M)$ e $\hat{s}_M(k_M)$.

A seguir, será obtida a representação polifásica [39] do sistema que conduz ao sinal em cada sub-banda.

O primeiro estágio está representado na Figura 4.4, que pode ser transformada na Figura 4.5. Nesta, tem-se

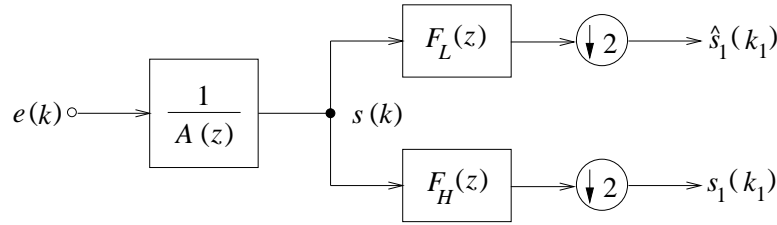


Figura 4.4: Primeiro estágio da divisão do processo AR em oitavas.

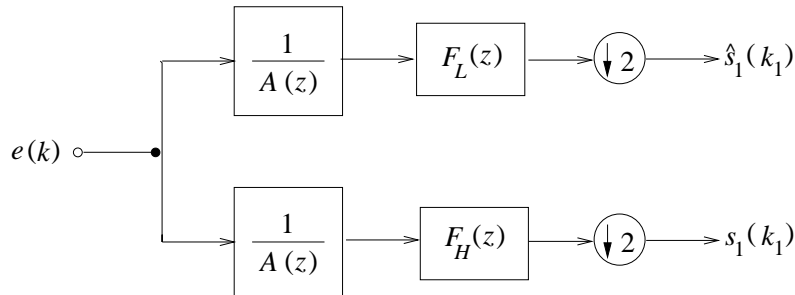


Figura 4.5: Duas sub-bandas resultantes do primeiro estágio da divisão do processo AR em oitavas.

$$\frac{F_H(z)}{A(z)} = \frac{F_H(z)A(-z)}{A(z)A(-z)} \triangleq \frac{B_H(z)}{A_1(z^2)} \triangleq \frac{B_{10}(z^2) + z^{-1}B_{11}(z^2)}{A_1(z^2)} \quad (4.9)$$

e

$$\frac{F_L(z)}{A(z)} = \frac{F_L(z) A(-z)}{A(z) A(-z)} \triangleq \frac{B_L(z)}{A_1(z^2)} \triangleq \frac{\hat{B}_{10}(z^2) + z^{-1} \hat{B}_{11}(z^2)}{A_1(z^2)}. \quad (4.10)$$

Com isso, chegou-se à representação polifásica do modelo que conduz a $s_1(k_1)$ e $\hat{s}_1(k_1)$, mostrada na Figura 4.6, que pode ser transformada [39] na forma equivalente mostrada na Figura 4.7.

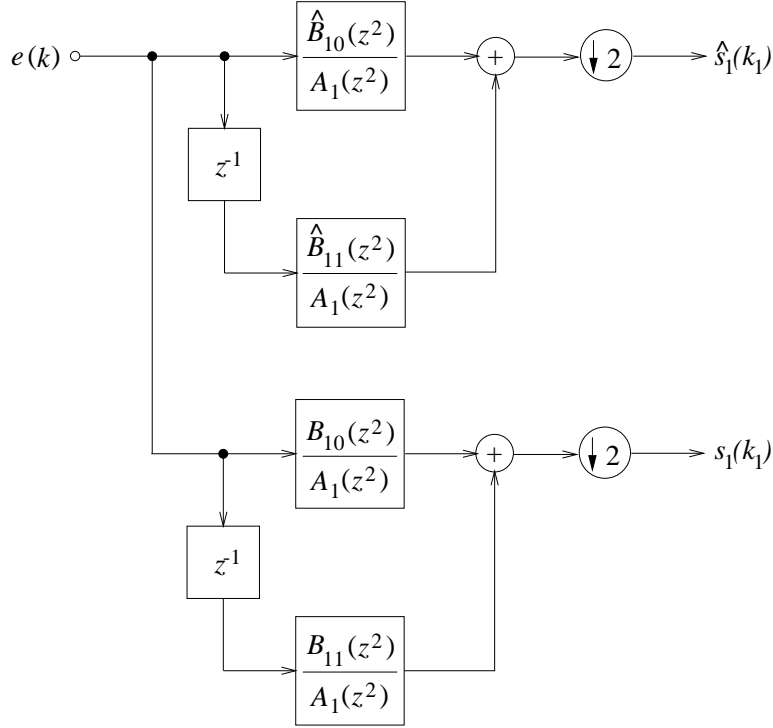


Figura 4.6: Representação polifásica do modelo para as duas sub-bandas resultantes do primeiro estágio da divisão do processo AR em oitavas.

O segundo estágio, limitado à metade superior da Figura 4.7, está representado na Figura 4.8, que pode ser transformada na Figura 4.9. Nesta, tem-se

$$F_H(z) \frac{\hat{B}_{10}(z)}{A_1(z)} = F_H(z) \frac{\hat{B}_{10}(z) A_1(-z)}{A_1(z) A_1(-z)} \triangleq \frac{B_{10H}(z)}{A_2(z^2)} \triangleq \frac{B_{20}(z^2) + z^{-1} B_{22}(z^2)}{A_2(z^2)}, \quad (4.11)$$

$$F_H(z) \frac{\hat{B}_{11}(z)}{A_1(z)} = F_H(z) \frac{\hat{B}_{11}(z) A_1(-z)}{A_1(z) A_1(-z)} \triangleq \frac{B_{11H}(z)}{A_2(z^2)} \triangleq \frac{B_{21}(z^2) + z^{-1} B_{23}(z^2)}{A_2(z^2)}, \quad (4.12)$$

$$F_L(z) \frac{\hat{B}_{10}(z)}{A_1(z)} = F_L(z) \frac{\hat{B}_{10}(z) A_1(-z)}{A_1(z) A_1(-z)} \triangleq \frac{B_{10L}(z)}{A_2(z^2)} \triangleq \frac{\hat{B}_{20}(z^2) + z^{-1} \hat{B}_{22}(z^2)}{A_2(z^2)} \quad (4.13)$$

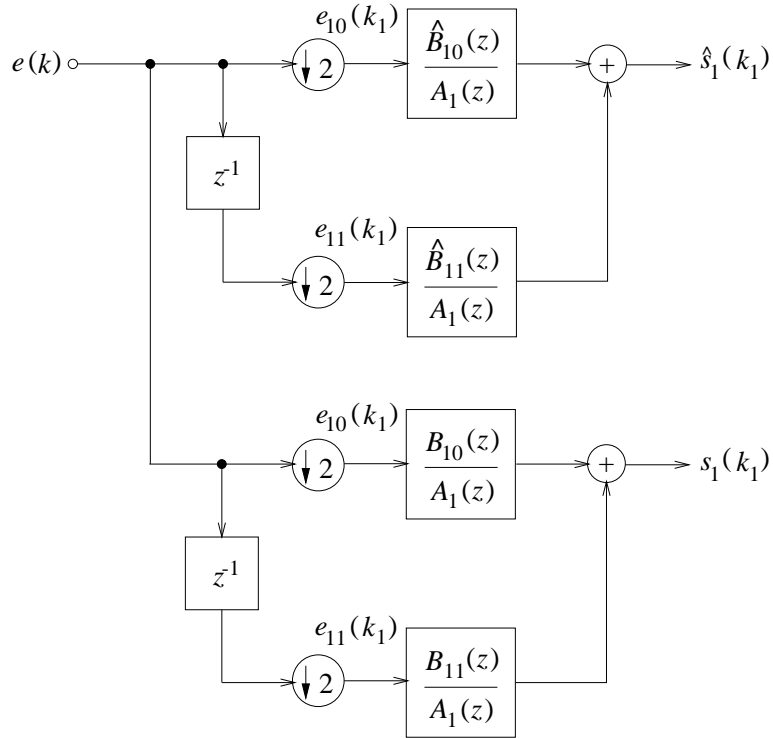


Figura 4.7: Representação polifásica modificada do modelo para as duas sub-bandas resultantes do primeiro estágio da divisão do processo AR em oitavas.

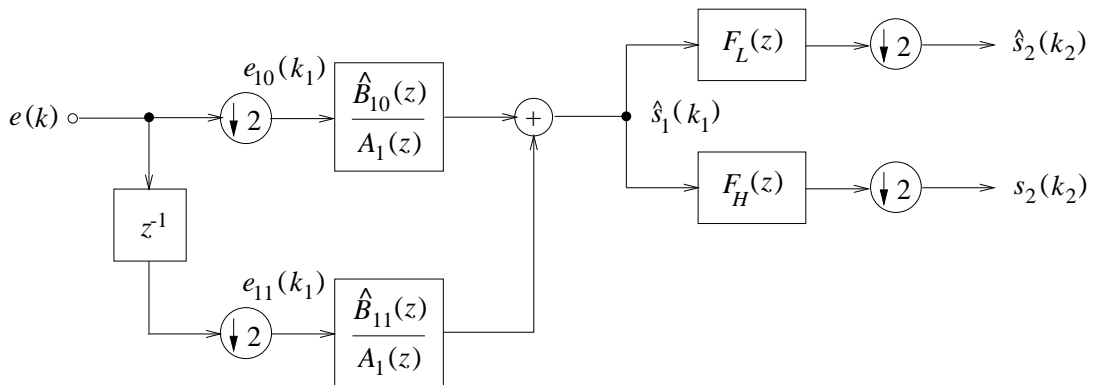


Figura 4.8: Segundo estágio da divisão do processo AR em oitavas.

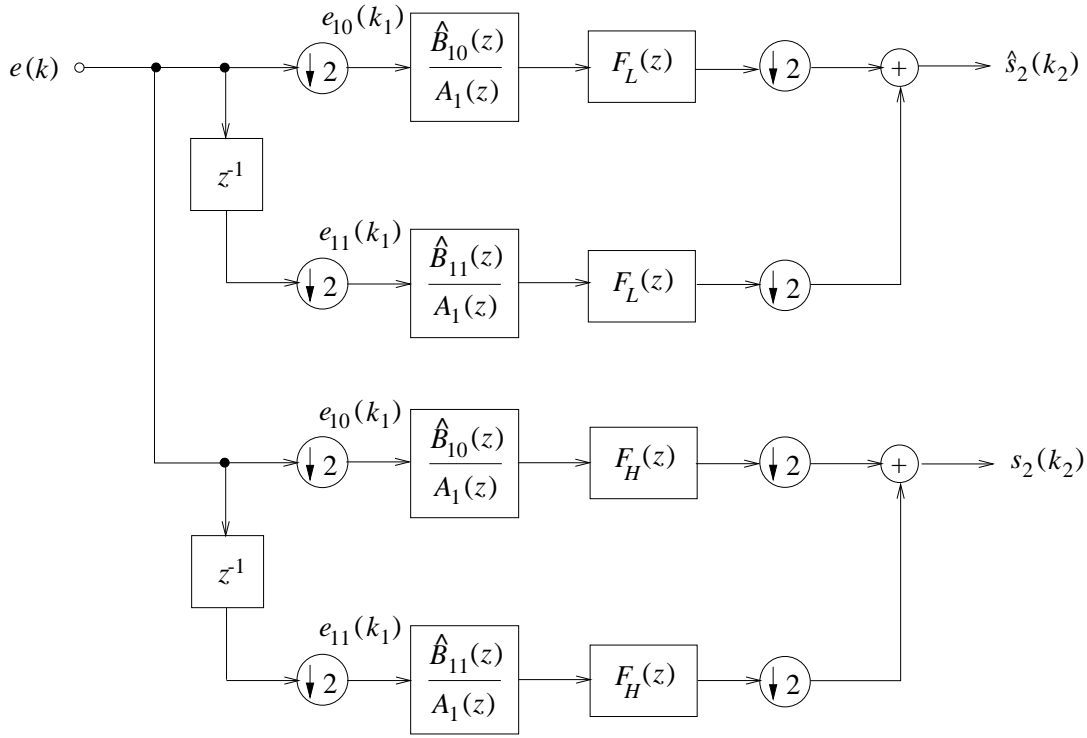


Figura 4.9: Duas sub-bandas resultantes do segundo estágio da divisão do processo AR em oitavas.

e

$$F_L(z) \frac{\hat{B}_{11}(z)}{A_1(z)} = F_L(z) \frac{\hat{B}_{11}(z)}{A_1(z)} \frac{A_1(-z)}{A_1(-z)} \triangleq \frac{B_{11L}(z)}{A_2(z^2)} \triangleq \frac{\hat{B}_{21}(z^2) + z^{-1} \hat{B}_{23}(z^2)}{A_2(z^2)}. \quad (4.14)$$

Com isso, chegou-se à representação polifásica do modelo que conduz a $s_2(k_2)$ e $\hat{s}_2(k_2)$, mostrada na Figura 4.10, que pode ser transformada na forma equivalente mostrada na Figura 4.11.

Prosseguindo sucessivamente, chega-se à representação polifásica do modelo que conduz a $s_M(k_M)$ e $\hat{s}_M(k_M)$, mostrada na Figura 4.12, já em sua forma modificada.

Então, após um m -ésimo estágio genérico, o sinal em uma sub-banda será representado pela soma de 2^m processos ARMA; os respectivos filtros geradores

$$\mathcal{M}_{mp}(z) \triangleq \frac{B_{mp}(z)}{A_m(z)}, \quad p = 0 \text{ a } 2^m - 1, \quad (4.15)$$

são IIR, com numeradores de ordem

$$\begin{cases} O_{\text{num}}(m) = \lceil \frac{O_{\text{num}}(m-1) + I}{2} \rceil, & m \geq 2 \\ O_{\text{num}}(1) = \lceil \frac{Q+I}{2} \rceil \end{cases} \quad (4.16)$$

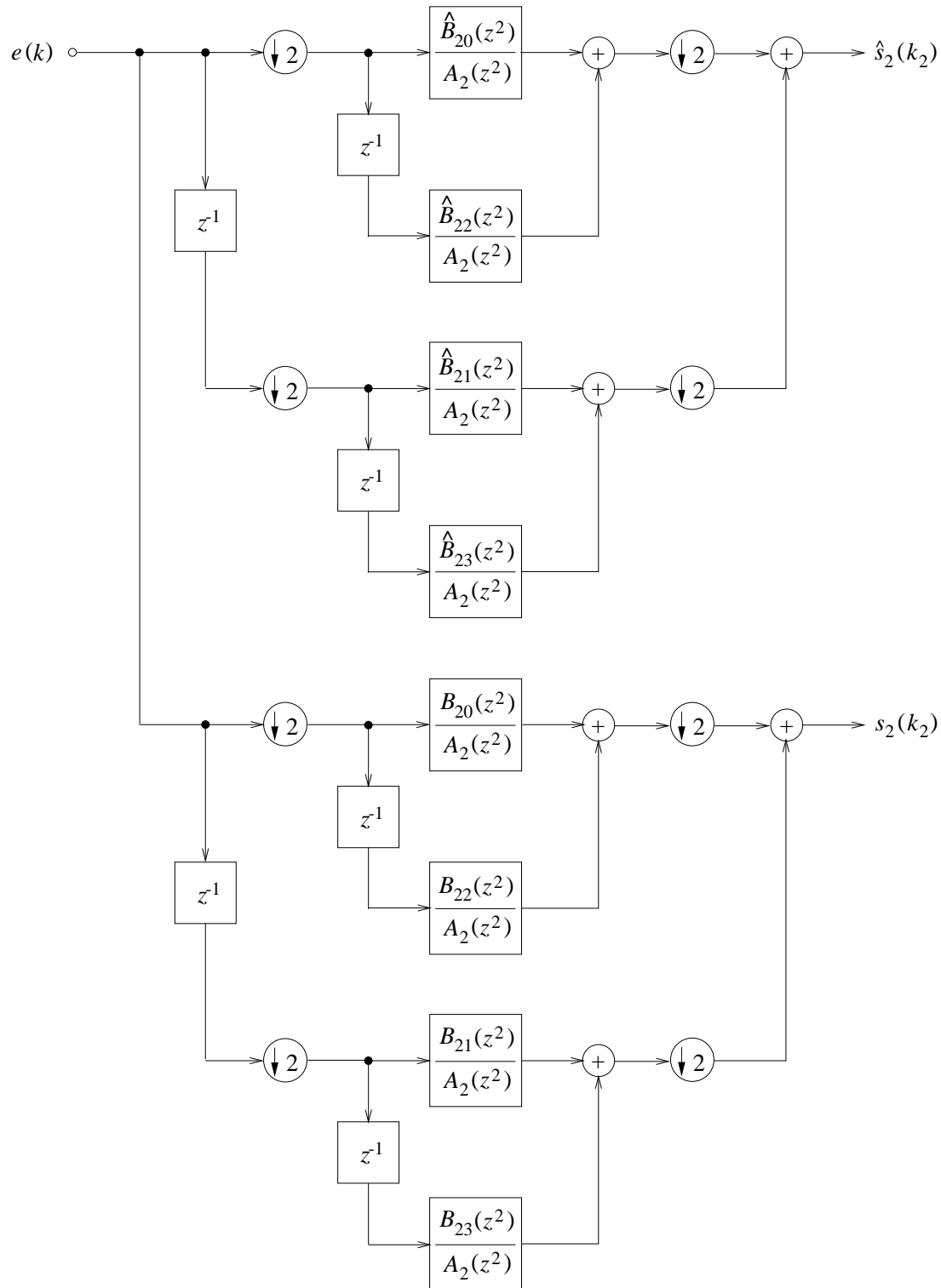


Figura 4.10: Representação polifásica do modelo para as duas sub-bandas resultantes do segundo estágio da divisão do processo AR em oitavas.

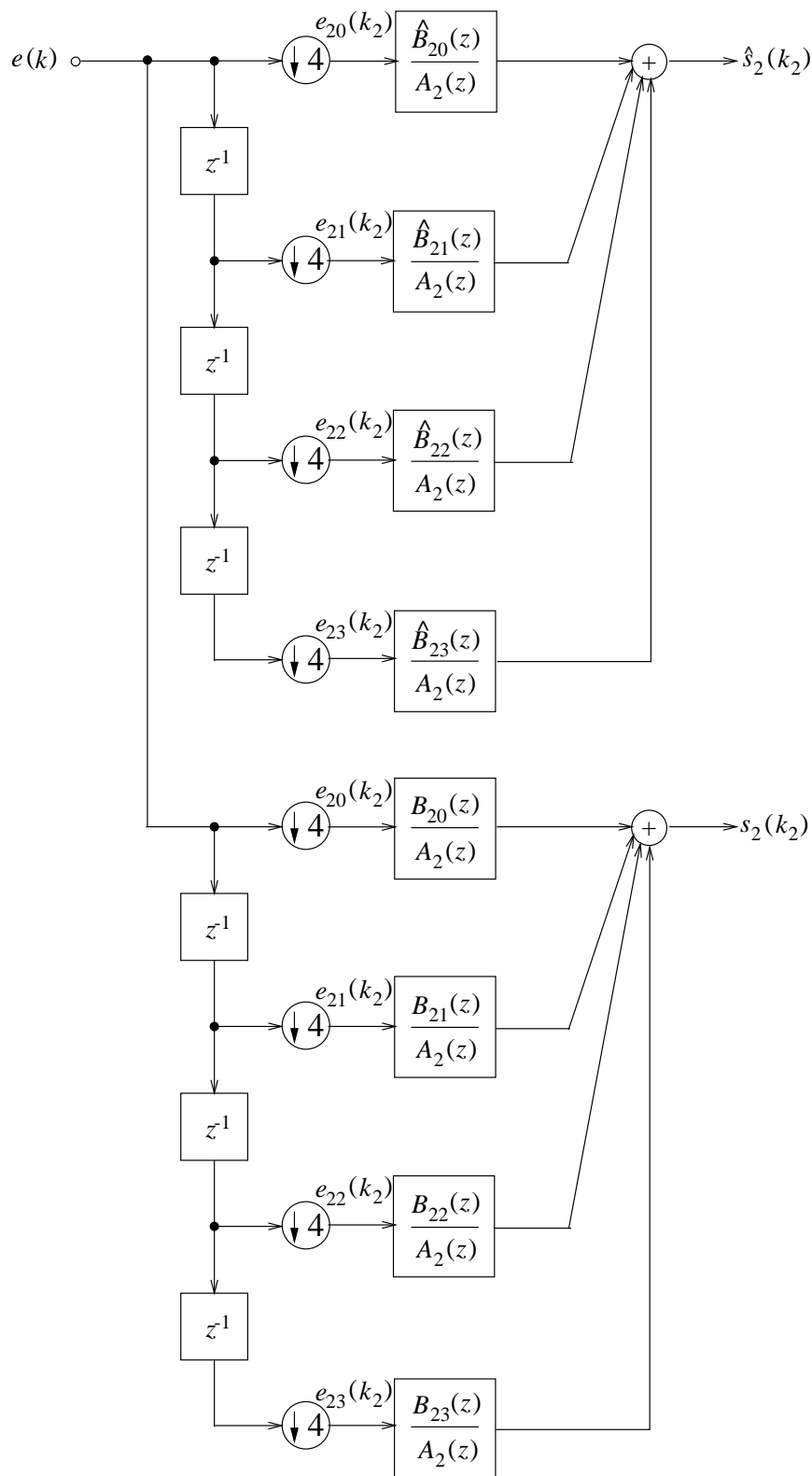


Figura 4.11: Representação polifásica modificada do modelo para as duas sub-bandas resultantes do segundo estágio da divisão do processo AR em oitavas.

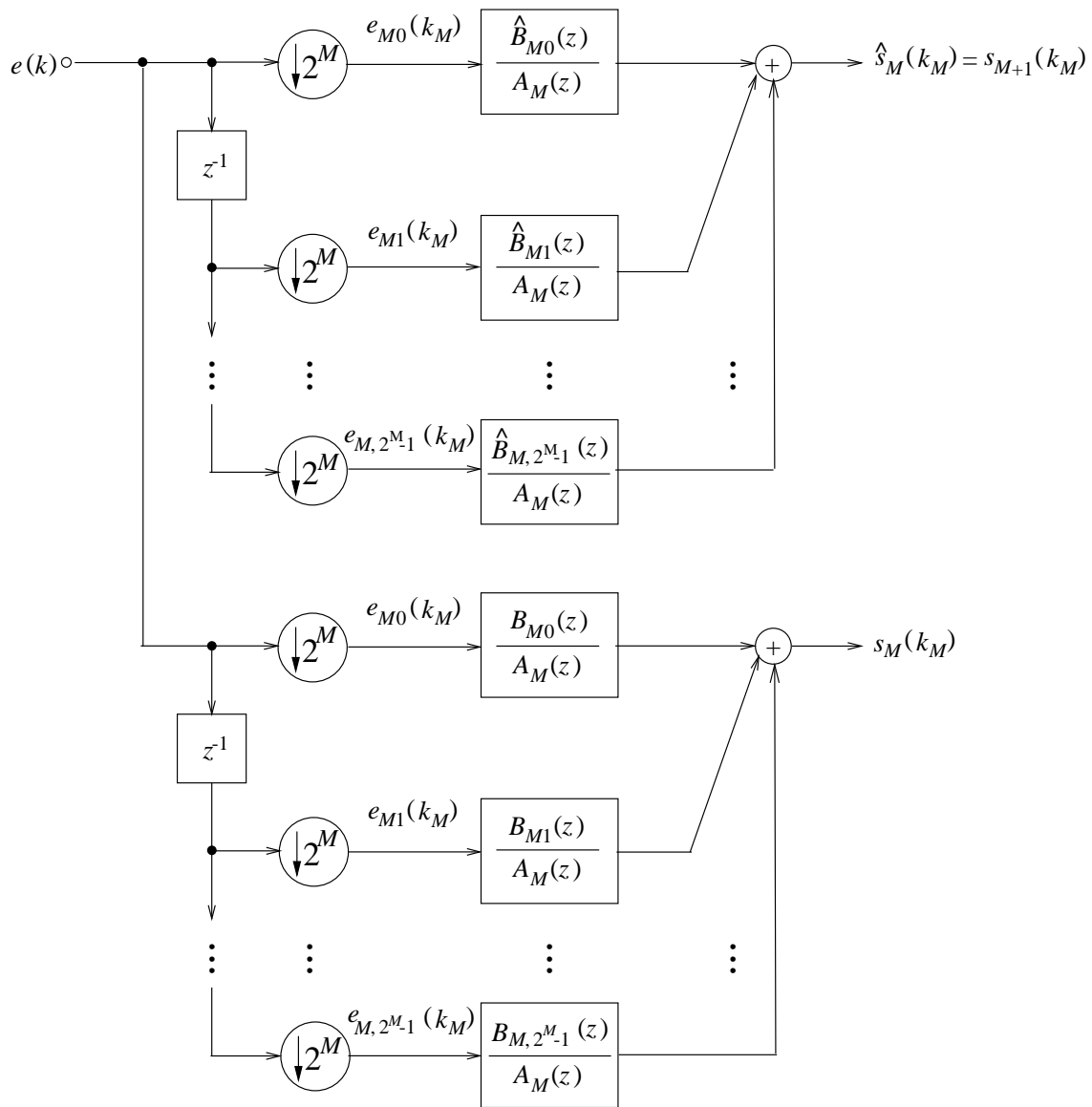


Figura 4.12: Representação polifásica modificada do modelo para as duas sub-bandas resultantes do M -ésimo estágio da divisão do processo AR em oitavas.

e denominadores iguais de ordem

$$O_{\text{den}}(m) = I, \quad (4.17)$$

a menos de cancelamentos. Convencione-se, aqui, que $\mathcal{M}(z) = \mathcal{Z}[m(k)]$.

Utilizando a densidade espectral de potência [18, 19], será derivado para cada sub-banda um modelo integrado a partir da representação em componentes polifásicas.

Seja a sub-banda genérica m , mostrada na Figura 4.13.

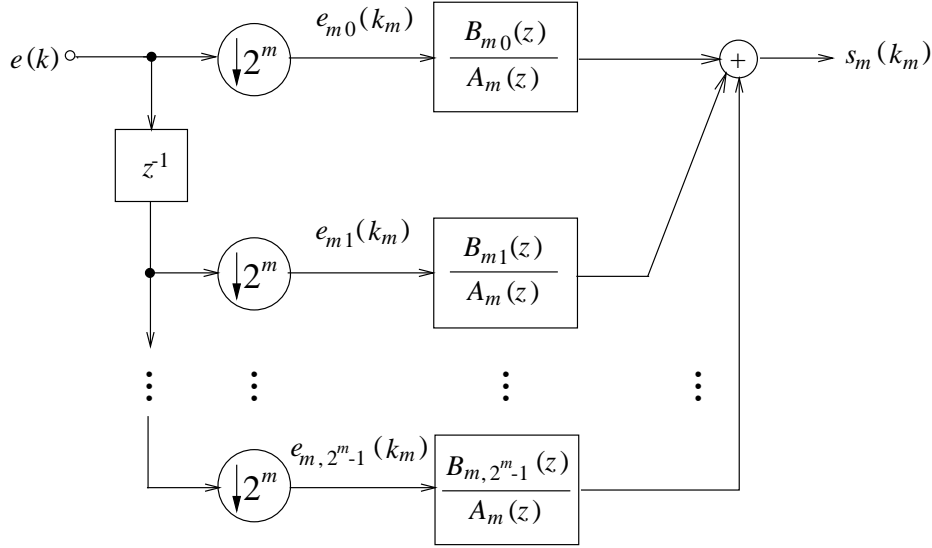


Figura 4.13: Representação polifásica do modelo para a m -ésima sub-banda resultante da divisão do processo AR em oitavas.

O processo $s_m(k_m)$, estacionário no sentido amplo, pode ser descrito como

$$s_m(k_m) = \sum_{p=0}^{2^m-1} (m_{mp} * e_{mp})(k_m). \quad (4.18)$$

Sua autocorrelação, portanto, é

$$R_{s_m s_m}(\Delta_m) = E \left[\sum_{p_1=0}^{2^m-1} (m_{mp_1} * e_{mp_1})(k_m) \sum_{p_2=0}^{2^m-1} (m_{mp_2} * e_{mp_2})(k_m - \Delta_m) \right]. \quad (4.19)$$

A independência entre as excitações $e_{mp}(k_m)$ permite a eliminação dos termos cruzados, levando a

$$R_{s_m s_m}(\Delta_m) = \sum_{p=0}^{2^m-1} E [(m_{mp} * e_{mp})(k_m)(m_{mp} * e_{mp})(k_m - \Delta_m)]. \quad (4.20)$$

A densidade espectral de potência associada a $s_m(k_m)$ é, portanto,

$$\mathcal{S}_{s_m s_m}(z) = \sum_{p=0}^{2^m-1} \mathcal{M}_{mp}(z) \mathcal{M}_{mp}(z^{-1}) \mathcal{S}_{e_{mp}}(z). \quad (4.21)$$

Mas $e_{mp}(k_m)$ são versões decimadas de $e(k)$, ruído branco de variância σ_e^2 ; então,

$$\mathcal{S}_{e_{mp}}(z) = \sigma_e^2. \quad (4.22)$$

Com isso,

$$\begin{aligned} \mathcal{S}_{s_m s_m}(z) &= \left(\sum_{p=0}^{2^m-1} \mathcal{M}_{mp}(z) \mathcal{M}_{mp}(z^{-1}) \right) \sigma_e^2 = \\ &= \frac{\sum_{p=0}^{2^m-1} B_{mp}(z) B_{mp}(z^{-1})}{A_m(z) A_m(z^{-1})} \sigma_e^2. \end{aligned} \quad (4.23)$$

Agora, por inspeção, tem-se que o numerador da equação (4.23)

$$P(z) \triangleq \sum_{p=0}^{2^m-1} B_{mp}(z) B_{mp}(z^{-1}) = P(z^{-1}). \quad (4.24)$$

Isso mostra que suas raízes são recíprocas duas a duas, permitindo fatorá-lo, então, como

$$P(z) = B_m(z) B_m(z^{-1}), \quad (4.25)$$

que, levado à equação (4.23), resulta em

$$\mathcal{S}_{s_m s_m}(z) = \frac{B_m(z) B_m(z^{-1})}{A_m(z) A_m(z^{-1})} \sigma_e^2 = \frac{B_m(z)}{A_m(z)} \frac{B_m(z^{-1})}{A_m(z^{-1})} \sigma_e^2. \quad (4.26)$$

Isso significa que se pode substituir o modelo composto da soma de 2^m processos ARMA independentes para o sinal $s_m(k_m)$ por um modelo ARMA único na taxa reduzida, equivalente em termos da densidade espectral de potência, com filtro gerador de fase mínima

$$\mathcal{M}_m(z) \triangleq \frac{B_m(z)}{A_m(z)}, \quad (4.27)$$

como se vê na Figura 4.14, obtido pela fatoração de $\mathcal{S}_{s_m s_m}(z)$ e de ordem igual às de $\mathcal{M}_{mp}(z)$, a menos de cancelamentos.

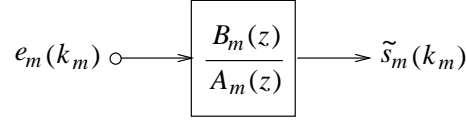


Figura 4.14: Modelo integrado para a m -ésima sub-banda resultante da divisão do processo AR em oitavas.

4.4 Correlação entre Sub-Bandas

O modelo obtido na Subseção 4.3 ainda não prevê a correlação entre sub-bandas. A seguir, efetua-se sua formulação.

Considerem-se os sinais resultantes do m -ésimo estágio de subdivisão, s_m e \hat{s}_m . A sua correlação cruzada é

$$R_{\hat{s}_m s_m}(\Delta_m) = E \left[\sum_{p_1=0}^{2^m-1} (\hat{m}_{mp_1} * e_{mp_1})(k_m) \sum_{p_2=0}^{2^m-1} (m_{mp_2} * e_{mp_2})(k_m - \Delta_m) \right]. \quad (4.28)$$

Novamente, a independência entre as excitações $e_{mp}(k_m)$ permite a eliminação dos termos cruzados, levando a

$$R_{\hat{s}_m s_m}(\Delta_m) = \sum_{p=0}^{2^m-1} E [(\hat{m}_{mp} * e_{mp})(k_m)(m_{mp} * e_{mp})(k_m - \Delta_m)]. \quad (4.29)$$

A densidade espectral de potência cruzada associada é, portanto,

$$\mathcal{S}_{\hat{s}_m s_m}(z) = \sum_{p=0}^{2^m-1} \hat{\mathcal{M}}_{mp}(z) \mathcal{M}_{mp}(z^{-1}) \mathcal{S}_{e_{mp}}(z) = \sum_{p=0}^{2^m-1} \hat{\mathcal{M}}_{mp}(z) \mathcal{M}_{mp}(z^{-1}) \sigma_e^2, \quad (4.30)$$

já de acordo com a equação (4.22).

Isso completa o modelo.

4.5 Uma Representação Hierárquica

É possível representar o modelo completo em taxa reduzida que leva aos dois sinais gerados no m -ésimo estágio de divisão do espectro. Esse é o objeto desta Subseção.

Após o $(m - 1)$ -ésimo estágio, já se pode dispor do modelo para o sinal $\hat{s}_{m-1}(k_{m-1})$ nos moldes da Subseção 4.3, conforme a Figura 4.15. Pode-se, então,

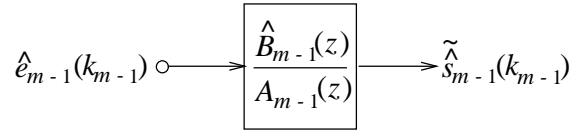


Figura 4.15: Modelo para o sinal $\hat{s}_{m-1}(k_{m-1})$, gerado na $(m - 1)$ -ésima divisão do processo AR em oitavas.

partir de $\tilde{s}_{m-1}(k_{m-1})$ na elaboração dos modelos para os sinais decorrentes do m -ésimo estágio de divisão, representado na Figura 4.16, que pode ser transformada na Figura 4.17. Nesta, tem-se

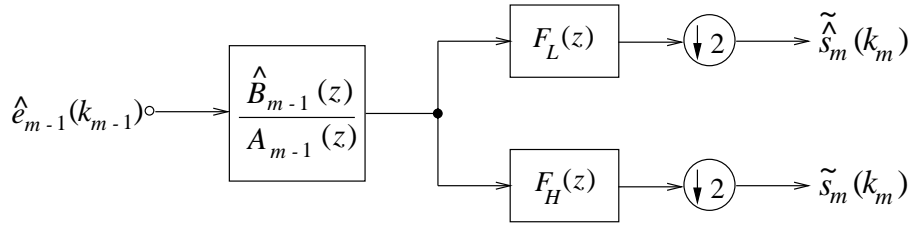


Figura 4.16: M -ésimo estágio da divisão do processo AR em oitavas.

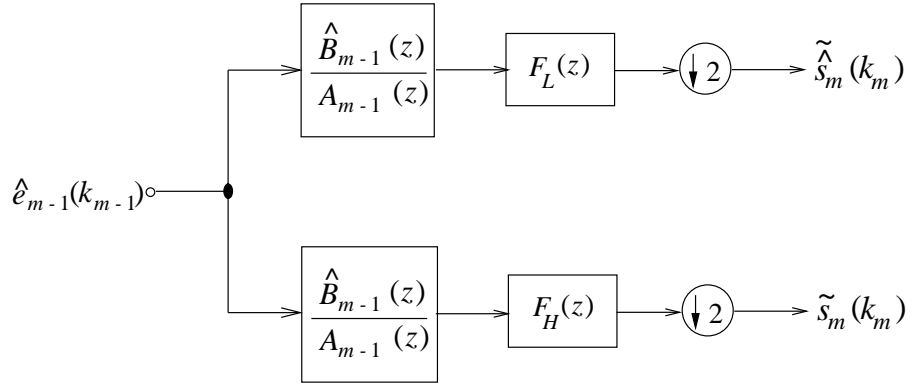


Figura 4.17: Duas sub-bandas resultantes do m -ésimo estágio da divisão do processo AR em oitavas.

$$\frac{F_H(z)\hat{B}_{m-1}(z)}{A_{m-1}(z)} = \frac{F_H(z)\hat{B}_{m-1}(z)A_{m-1}(-z)}{A_{m-1}(z)A_{m-1}(-z)} \triangleq \frac{B_{m_H}(z)}{A_m(z^2)} \triangleq \frac{C_{m0}(z^2) + z^{-1}C_{m1}(z^2)}{A_m(z^2)} \quad (4.31)$$

e

$$\frac{F_L(z)\hat{B}_{m-1}(z)}{A_{m-1}(z)} = \frac{F_L(z)\hat{B}_{m-1}(z)A_{m-1}(-z)}{A_{m-1}(z)A_{m-1}(-z)} \triangleq \frac{B_{mL}(z)}{A_m(z^2)} \triangleq \frac{\hat{C}_{m0}(z^2) + z^{-1}\hat{C}_{m1}(z^2)}{A_m(z^2)}. \quad (4.32)$$

Com isso, chegou-se à representação polifásica do modelo que conduz a $\tilde{s}_m(k_m)$ e $\tilde{\tilde{s}}_m(k_m)$, mostrada na Figura 4.18, que pode ser transformada na forma equivalente mostrada na Figura 4.19.

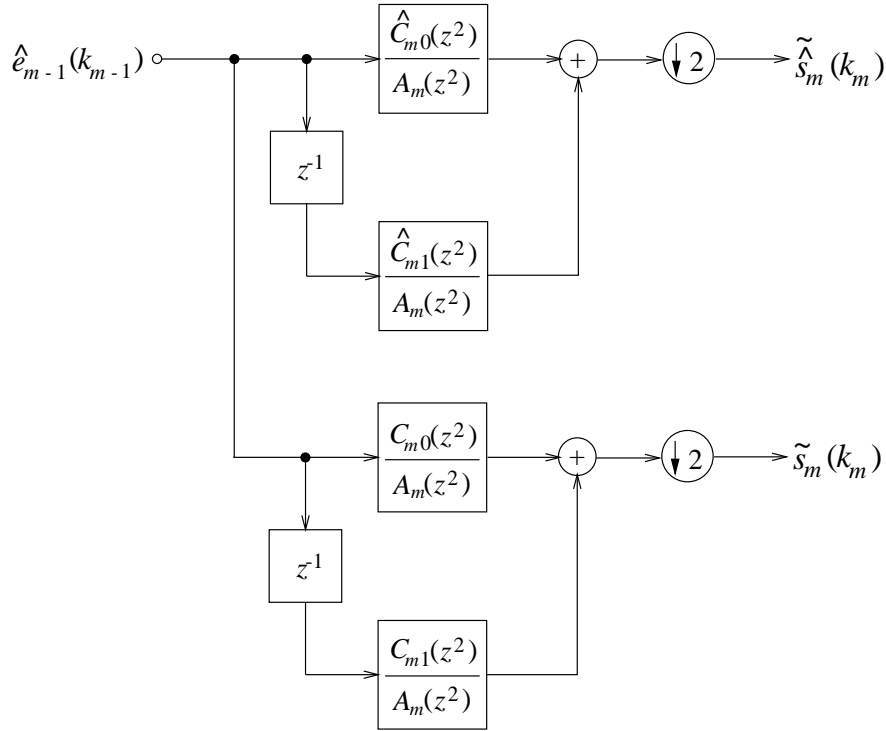


Figura 4.18: Representação polifásica do modelo para as duas sub-bandas resultantes do m -ésimo estágio da divisão do processo AR em oitavas.

A rede redesenhada na Figura 4.20 representa completamente os modelos em taxa reduzida $\tilde{s}_m(k_m)$ e $\tilde{\tilde{s}}_m(k_m)$, equivalentes respectivamente a $s_m(k_m)$ e $\hat{s}_m(k_m)$ quanto à densidade espectral de potência referida na Subseção 4.3 e à densidade espectral de potência cruzada referida na Subseção 4.4.

Em outras palavras, no que diz respeito unicamente aos sinais em mesma taxa obtidos a cada etapa da análise em oitavas de um processo AR, uma rede composta por 4 filtros alimentada por 2 processos de ruído branco de mesma variância é um modelo suficiente na taxa reduzida.

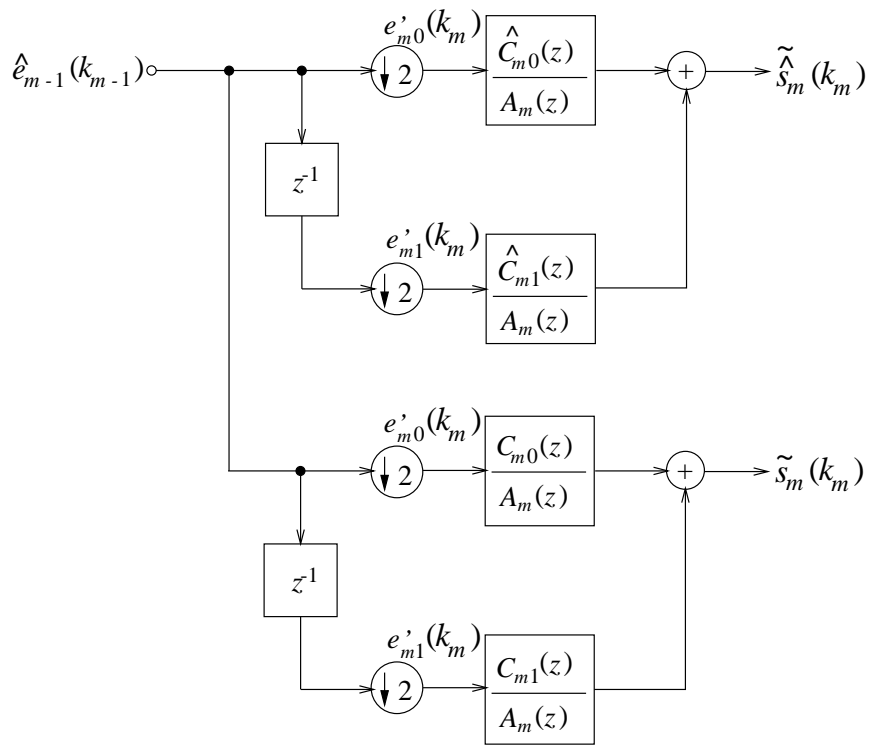


Figura 4.19: Representação polifásica modificada do modelo para as duas sub-bandas resultantes do m -ésimo estágio da divisão do processo AR em oitavas.

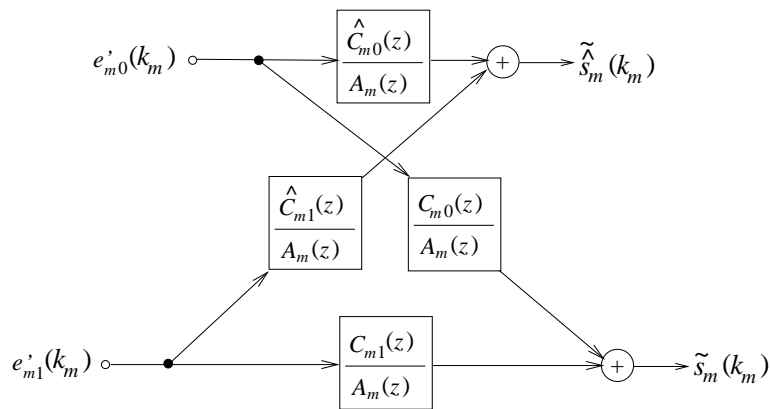


Figura 4.20: Modelo completo para os sinais decorrentes do m -ésimo estágio da divisão do processo AR em oitavas.

4.6 Discussão da Validade do Modelo

Supondo, inicialmente, que modelos ARMA são capazes de representar adequadamente os sinais nas sub-bandas, a resposta em frequência dos filtros geradores descritos pela equação (4.27) corresponde à dos filtros que seriam obtidos por estimação direta dos coeficientes do modelo a partir dos dados?

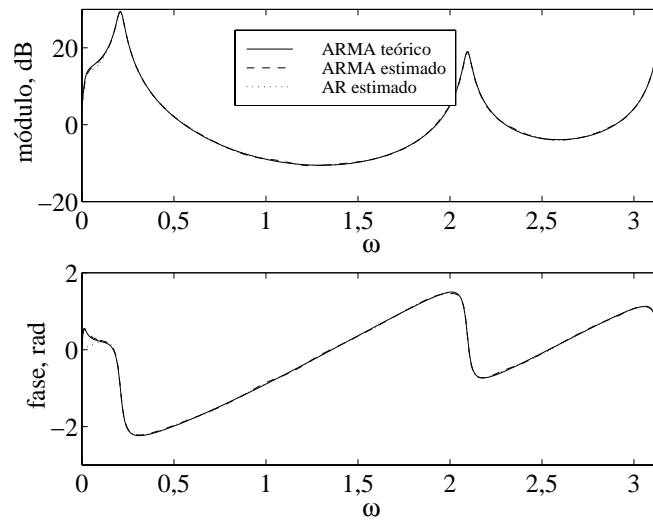
Em caso positivo, os modelos ARMA poderiam ser substituídos por modelos AR sem acréscimo de complexidade e ainda manter, com boa aproximação, a resposta em frequência?

Por fim, os filtros geradores correspondentes aos modelos ARMA nas sub-bandas, invertidos, branqueiam os sinais nas sub-bandas?

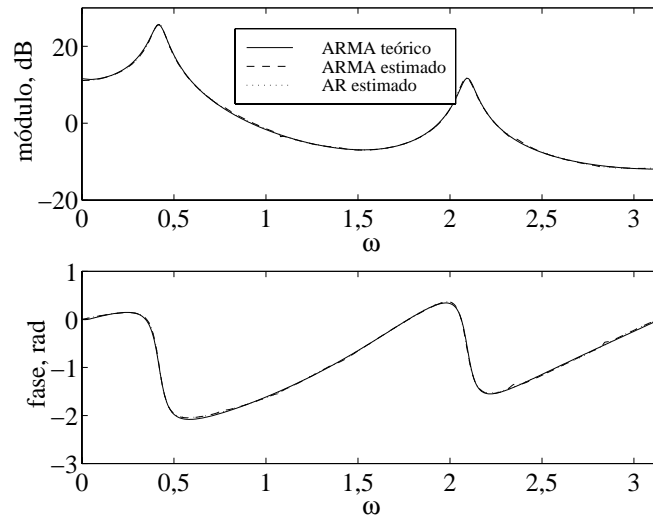
Um exemplo prático pode ilustrar as três respostas. Um processo AR de 22ª ordem foi artificialmente formado com os pólos obtidos pela combinação dos módulos 0,5 e 0,99 com cada uma das fases 0 , $\frac{\pi}{30}$, $\frac{\pi}{3}$, $\frac{\pi}{2}$, $2\frac{\pi}{3}$ e $29\frac{\pi}{30}$ e analisado em sucessivos estágios de bancos de filtros Haar [39] num esquema como o da Figura 4.3.

Para o sinal em cada sub-banda, obteve-se o filtro gerador correspondente de três modos distintos: pelo modelo ARMA teoricamente descrito pela equação (4.27); por estimação direta dos coeficientes de um modelo ARMA de mesma ordem, em simulação; e por estimação direta dos coeficientes de um modelo AR com regressão de mesma ordem, em simulação. A Figura 4.21 compara as 3 respostas de módulo e fase correspondentes, para os sinais resultantes após 2 estágios de divisão. As respostas em frequência dos modelos ARMA foram praticamente indistinguíveis, validando o modelo teórico. A possibilidade de utilização de modelos AR de complexidade menor também se evidenciou, embora a ordem escolhida não tenha sido capaz de conformar perfeitamente a resposta em muito baixas frequências de \mathcal{M}_1 .

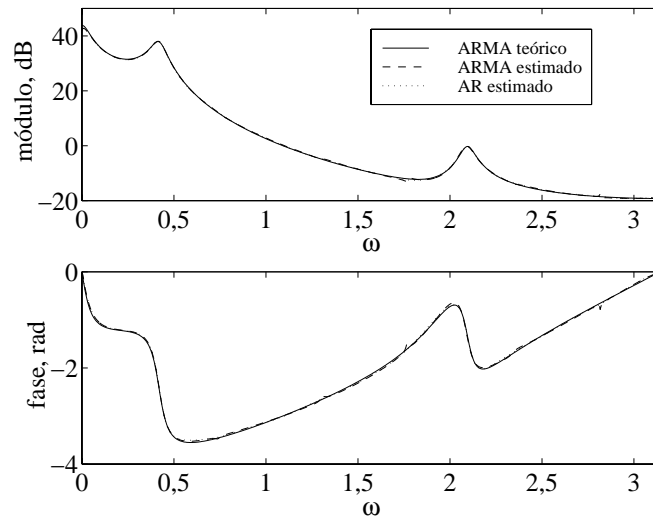
No mesmo exemplo, realizou-se a filtragem inversa do sinal em cada sub-banda, em simulação, pelo filtro gerador previsto pela equação (4.27) e calculou-se a autocorrelação temporal do sinal de excitação resultante. A Figura 4.22 mostra as autocorrelações normalizadas para as excitações correspondentes a 2 estágios de divisão. Os sinais podem ser considerados, para todos os efeitos, ruídos brancos, confirmando a adequação do modelo.



(a) $\mathcal{M}_1(z)$

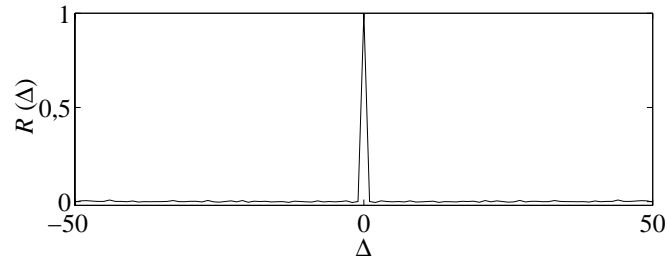


(b) $\mathcal{M}_2(z)$

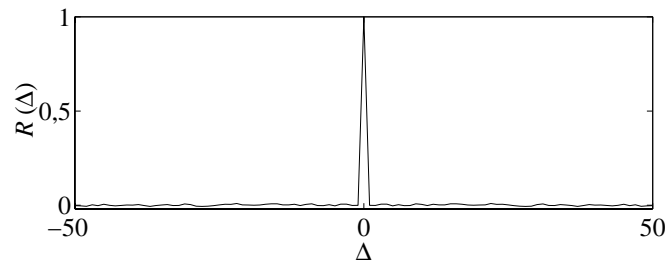


(c) $\hat{\mathcal{M}}_2(z)$

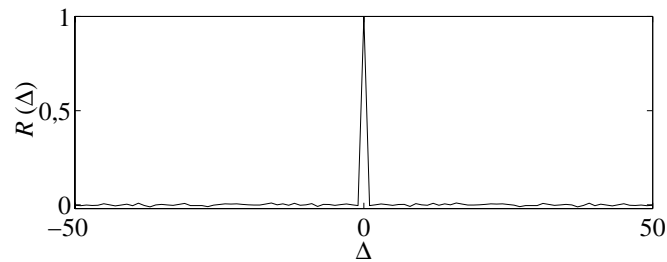
Figura 4.21: Respostas de módulo e fase: modelo ARMA teórico e modelos ARMA e AR estimados.



(a) $e_{d_1}(k_1)$



(b) $e_{d_2}(k_2)$



(c) $\hat{e}_{d_2}(k_2)$

Figura 4.22: Autocorrelações dos sinais obtidos por filtragem inversa dos sinais nas sub-bandas.

4.7 Mapeamento de Zeros e Pólos em cada Sub-Banda

Nesta Subseção, em busca de subsídios para avaliar a possibilidade de realizar simplificações, é examinada a localização dos pólos e zeros do modelo integrado para cada sub-banda. Será utilizada a mesma notação encontrada a partir da Figura 4.3.

4.7.1 Banco de Filtros Ideais

Inicialmente, considere-se que os filtros $F_H(z)$ e $F_L(z)$ do banco de análise foram escolhidos, respectivamente, passa-altas e passa-baixas de meia-banda ideais, como nas Figuras 4.23 e 4.24.

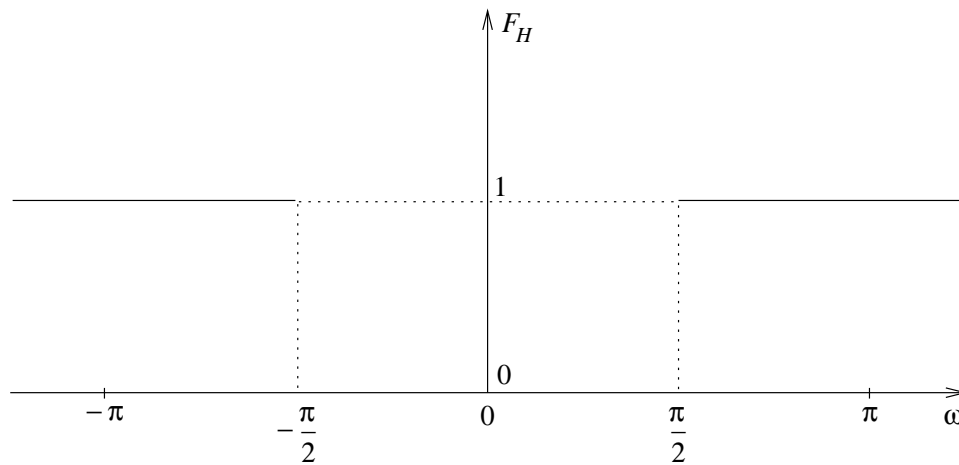


Figura 4.23: Resposta em frequência do passa-altas ideal.

Seja o caso trivial em que o sinal analisado é uma cossenóide, ou seja, o limite de um processo AR de segunda ordem quando os pólos do filtro gerador são complexos conjugados tendendo à circunferência unitária. Tem-se, então,

$$s(k) = C \cos(\theta k), \quad (4.33)$$

cujo filtro gerador se caracteriza pelo diagrama de pólos e zeros da Figura 4.25 e pela resposta em frequência da Figura 4.26.

É fácil concluir que, após a pré-filtragem por $F_H(z)$ e $F_L(z)$ e a decimação por 2 do sinal mostrado no exemplo, obtêm-se, respectivamente,

$$s_1(k_1) = 0 \quad (4.34)$$

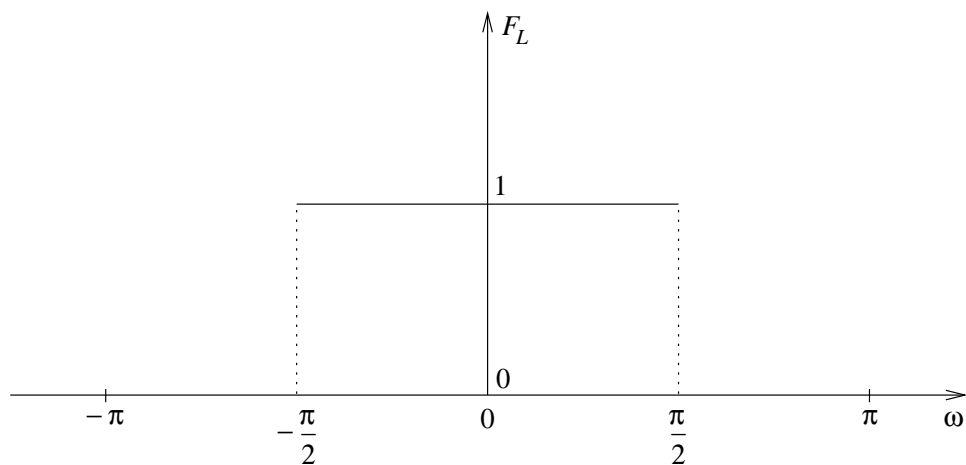


Figura 4.24: Resposta em freqüência do passa-baixas ideal.

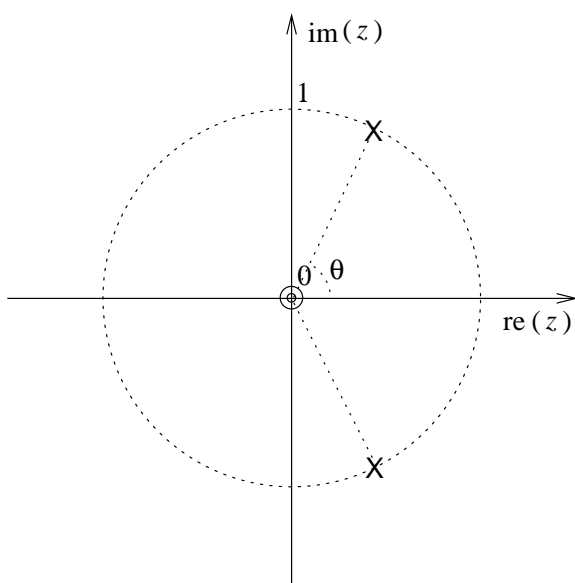


Figura 4.25: Diagrama de pólos e zeros do filtro gerador de $s(k)$.

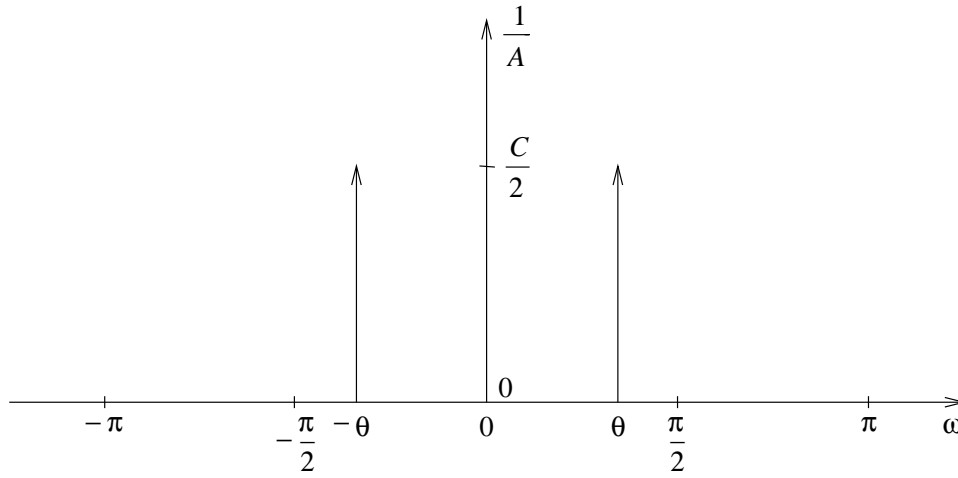


Figura 4.26: Resposta em frequência do filtro gerador de $s(k)$.

e

$$\hat{s}_1(k_1) = C \cos(2\theta k_1), \quad (4.35)$$

este com filtro gerador caracterizado pelo diagrama de pólos e zeros da Figura 4.27 e pela resposta em frequência da Figura 4.28.

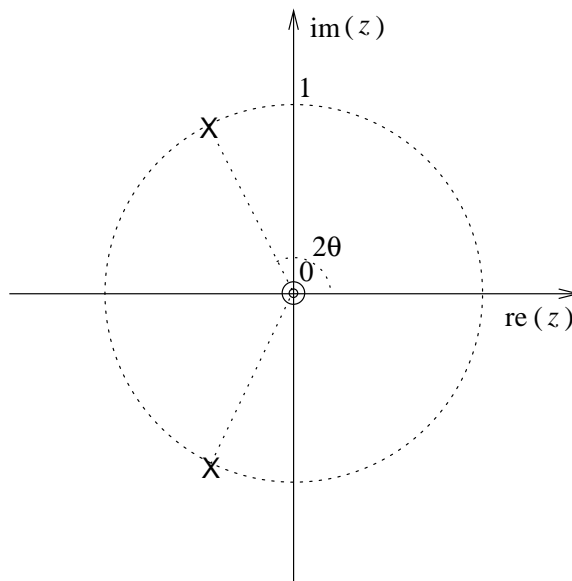


Figura 4.27: Diagrama de pólos e zeros do filtro gerador de $\hat{s}_1(k_1)$.

No próximo estágio, após a pré-filtragem por $F_H(z)$ e $F_L(z)$ e a decimação por 2 de $\hat{s}_1(k_1)$ no exemplo, obtêm-se, respectivamente,

$$s_2(k_2) = C \cos(4\theta k_2) \quad (4.36)$$

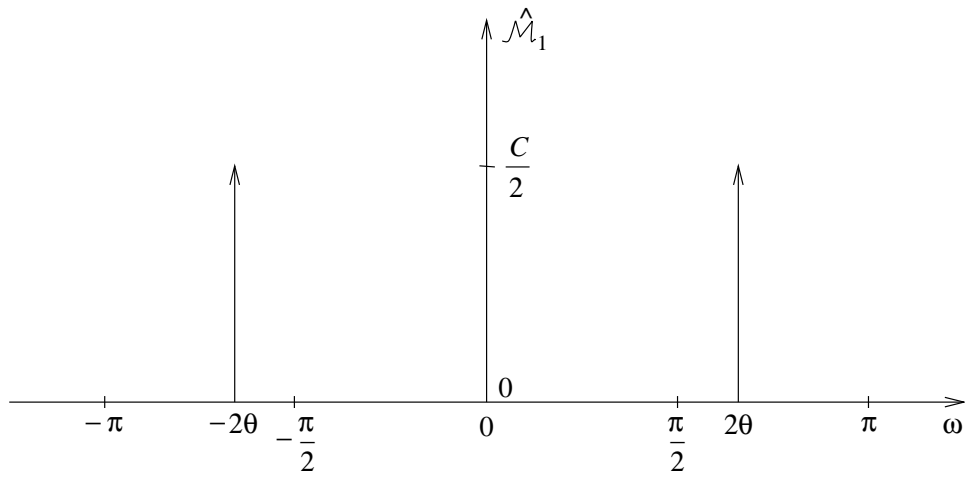


Figura 4.28: Resposta em freqüência do filtro gerador de $\hat{s}_1(k_1)$.

e

$$\hat{s}_2(k_2) = 0, \quad (4.37)$$

aquele com filtro gerador caracterizado pelo diagrama de pólos e zeros da Figura 4.29 e pela resposta em freqüência da Figura 4.30.

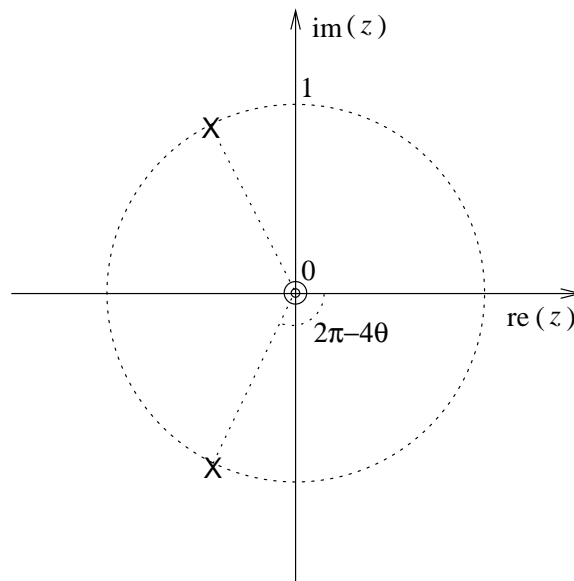


Figura 4.29: Diagrama de pólos e zeros do filtro gerador de $s_2(k_2)$.

A generalização do procedimento acima permite afirmar que uma cossenóide de freqüência θ analisada sucessivamente em oitavas por bancos de filtros ideais aparecerá em $\hat{s}_m(k_m)$ com sua freqüência dobrada, até ser “separada” como uma

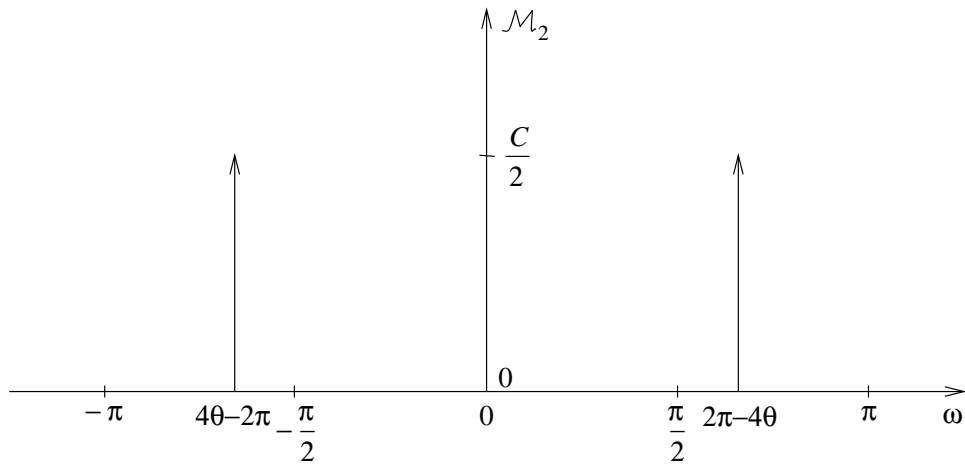


Figura 4.30: Resposta em freqüência do filtro gerador de $s_2(k_2)$.

componente aparecendo exclusivamente no sinal $s_{m_\theta}(k_{m_\theta})$ da sub-banda m_θ , com freqüência $2\pi - 2^{m_\theta}\theta$, sendo

$$m_\theta = \left\lceil \log_2 \frac{\pi}{2\theta} \right\rceil + 1. \quad (4.38)$$

Seja, agora, um caso geral em que $s(k)$ é perfeitamente modelado por um processo AR genérico cujo filtro gerador tem pólos z_{p_i} , $i = 0, \dots, I - 1$, e apresenta a resposta em freqüência de módulo esquematicamente mostrada na Figura 4.31.

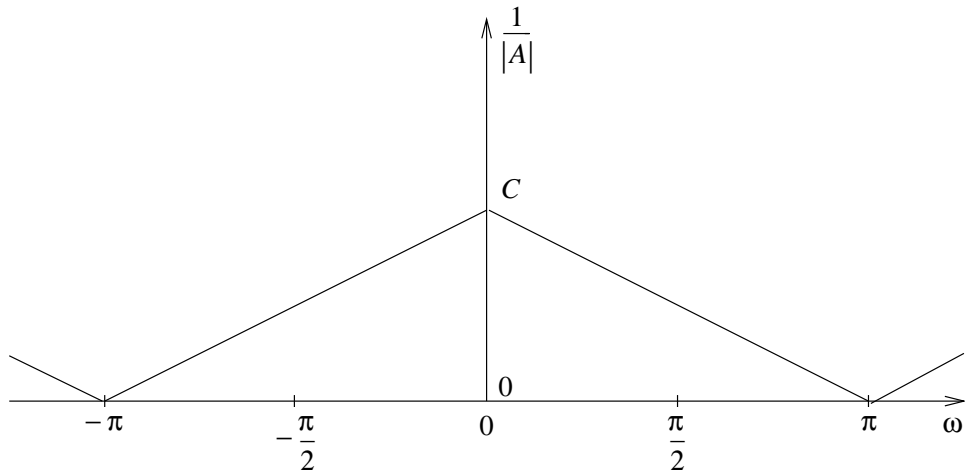


Figura 4.31: Resposta em freqüência esquemática do módulo do filtro gerador de um processo AR genérico.

Após a pré-filtragem por $F_H(z)$ e $F_L(z)$ e a decimação por 2 do sinal mostrado no exemplo, obtêm-se, respectivamente, $s_1(k_1)$, com filtro gerador caracterizado

pela resposta em frequência de módulo da Figura 4.32, e $\hat{s}_1(k_1)$, com filtro gerador caracterizado pela resposta em frequência de módulo da Figura 4.33.

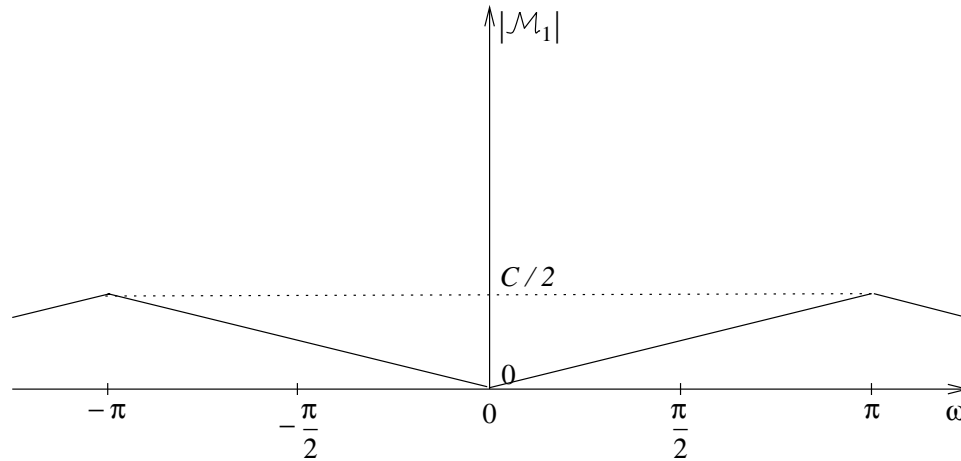


Figura 4.32: Resposta em frequência do módulo do filtro gerador de $s_1(k_1)$.

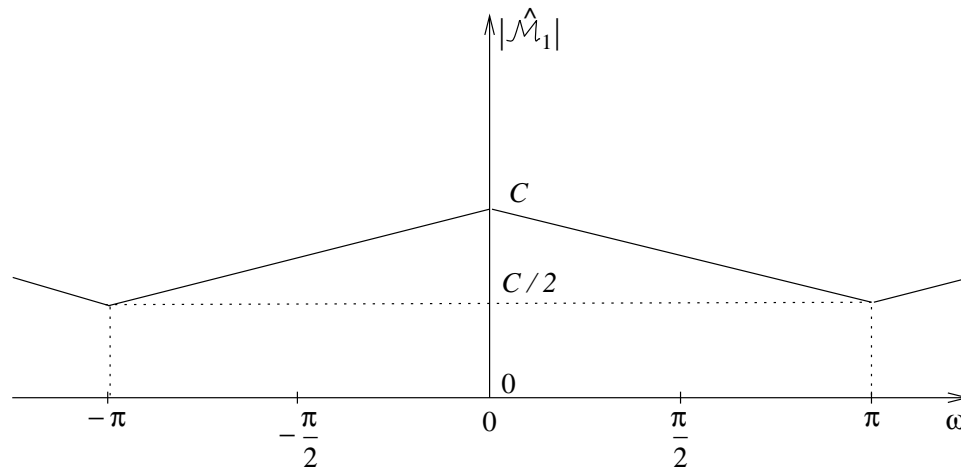


Figura 4.33: Resposta em frequência do módulo do filtro gerador de $\hat{s}_1(k_1)$.

Para que se possa abstrair da reflexão sofrida pela metade superior do espectro original na sua versão transformada em $s_1(k_1)$, será discutida apenas a transformação da metade inferior do espectro original em $\hat{s}_1(k_1)$, sem perda de generalidade. Pode-se dizer, então, que a equação

$$\hat{\mathcal{M}}_1(z) = \frac{1}{A\left(z^{\frac{1}{2}}\right)} \quad (4.39)$$

é válida, porém *apenas dentro dos limites permitidos pela pré-filtragem*. Observando-se a resposta em frequência, é possível inferir que o diagrama de pólos e zeros correspondente mostraria:

- os pólos $z_{p_i}^2$, ou seja, os de $\frac{1}{A(z)}$ elevados ao quadrado;
- os zeros suficientes para terem eliminado a metade superior do espectro antes da decimação, também elevados ao quadrado.

Após os estágios seguintes, o mesmo efeito pode ser observado sucessivamente.

Conclui-se que os efeitos dos pólos originais não podem ser perfeitamente isolados pelas sub-bandas; os pólos só podem ser de fato “distribuídos” entre elas no caso-limite de módulo unitário (o cosseno).

4.7.2 Banco de Filtros Realizáveis

Considere-se, agora, que os filtros $F_H(z)$ e $F_L(z)$ do banco de análise são, respectivamente, passa-altas e passa-baixas FIR não-ideais.

Seja o caso simples em que o sinal analisado é perfeitamente modelado por um processo AR de primeira ordem com filtro gerador descrito por

$$\frac{1}{A(z)} = \frac{1}{1 - az^{-1}}, \quad (4.40)$$

submetido às operações indicadas na Seção 4.3.

É fácil mapear os pólos a cada estágio.

Sendo

$$A(z) = 1 - az^{-1}, \quad (4.41)$$

após o primeiro estágio de análise,

$$A_1(z^2) = A(z)A(-z) = (1 - az^{-1})(1 + az^{-1}) = 1 - a^2z^{-2} \Rightarrow \quad (4.42)$$

$$\Rightarrow A_1(z) = 1 - a^2z^{-1}. \quad (4.43)$$

Isto quer dizer que o pólo em $z = a$ é mapeado em $z = a^2$.

No caso de um processo AR de segunda ordem com filtro gerador descrito por

$$\frac{1}{A(z)} = \frac{1}{(1 - re^{j\theta}z^{-1})(1 - re^{-j\theta}z^{-1})} = \quad (4.44)$$

$$= \frac{1}{1 - 2r\cos\theta z^{-1} + r^2z^{-2}}, \quad (4.45)$$

pode-se verificar que

$$A(z) = 1 - 2r\cos\theta z^{-1} + r^2 z^{-2}, \quad (4.46)$$

após o primeiro estágio de análise, resulta em

$$A_1(z) = 1 - 2r^2\cos(2\theta)z^{-1} + (r^2)^2 z^{-2}. \quad (4.47)$$

Isto quer dizer que o modo natural complexo do filtro dobra de frequência, com a duplicação do valor da fase dos pólos, e tem a seletividade a ele associada reduzida, com a redução da sua distância à origem de $r < 1$ para $r^2 < r$.

Esse efeito pode ser generalizado para qualquer número de pólos, e justifica as observações feitas no início da Seção 4.7. Naturalmente, ao longo dos estágios seguintes o mesmo efeito pode ser observado sucessivamente.

Uma questão mais difícil a responder é qual o posicionamento dos zeros introduzidos nos filtros geradores dos modelos, a cada estágio, e o efeito de sua interação com os pólos.

Foi visto que o mapeamento dos zeros dos filtros geradores dos sinais obtidos a cada estágio de análise resulta de sucessivas composições entre versões modificadas do denominador do filtro gerador do processo AR original e os filtros de análise. Levando-se este fato em conta e reavaliando-se de forma mais frouxa o que se viu no caso de bancos de filtros ideais, é possível traçar algumas expectativas quanto a esse mapeamento e seus efeitos.

Ignorando por ora a decimação, a largura de banda do sinal original $s(k)$ vai sendo progressivamente reduzida ao longo da filtragem de análise. Num estágio qualquer, o efeito individual de um pólo do filtro gerador de origem $\frac{1}{A(z)}$ sobre a resposta em frequência se distribui, por assim dizer, entre a banda passante e a banda de rejeição do filtro de análise equivalente. Se, por exemplo, tal efeito se encontra praticamente todo contido na banda de rejeição, a filtragem extremamente eficaz observada pode ser vista como resultado do cancelamento daquele pólo por um zero na mesma posição. Retomando, agora, a decimação, é preciso lembrar de que após o m -ésimo estágio cada pólo é representado por sua versão m vezes quadrada, mas a idéia geral é a mesma.

Portanto, à medida que os diversos estágios de divisão se sucedem, cada vez mais os filtros geradores equivalentes devem tender a apresentar pólos cancelados

por zeros. Naturalmente, isso conduz à simplificação gradual dos modelos dos sinais nas sub-bandas. E da mesma forma que se observa na ordem do numerador das funções de transferência dos filtros geradores na equação (4.16), que pára de crescer com o número de estágios de divisão e tende a um valor constante I , o efeito dos sucessivos filtros de análise deve se cristalizar numa configuração estável, à medida que se tenda à situação em que não há mais pólos a cancelar.

É claro que o emprego de bancos de análise de filtros mais seletivos é mais eficaz em separar os efeitos individuais dos pólos; mas, não sendo estes efeitos, eles próprios, necessariamente muito localizados na freqüência, é prematura a expectativa de que, em regra, os filtros mais seletivos serão mais eficazes na simplificação dos modelos dos sinais nas sub-bandas.

Por fim, pensando-se exclusivamente no seu efeito sobre uma dada sub-banda, seria defensável desprezar zeros ou pólos que fossem muito interiores à circunferência unitária, independentemente de cancelamentos. Mas nesse caso, é preciso ter em mente a conseqüente perda de acurácia do modelo global. Como exemplo, considere-se que foi adotado para o sistema em oitavas um critério em que se abandonam quaisquer zeros e pólos com módulo menor que um dado valor, igual para todas as sub-bandas. A cada estágio de subdivisão em freqüência, os pólos vão se afastando da circunferência unitária, ficando cada vez mais sujeitos a serem desprezados. Então, um pólo que tem originalmente fase próxima de 0 tem mais chances de ser descartado do modelo ao longo das diversas subdivisões do que um pólo que tem originalmente o mesmo módulo e fase próxima de π , pois este já pode ser “separado” pelo passa-altas do primeiro estágio.

Então, pelo já exposto, uma vez encontrado o modelo para uma particular sub-banda, é justificável efetuar simplificações de dois tipos nas funções de transferência do respectivo filtro gerador: eliminação de zeros ou pólos “com baixo fator de qualidade Q ” (entendendo-se dessa forma aqueles muito afastados da circunferência unitária) e eliminação de zeros e pólos por cancelamento, dentro de certa aproximação. No Capítulo 2, discute-se uma forma de avaliação do efeito de tais simplificações sobre a função de transferência global com base nos seus conseqüentes desvios máximos de módulo e fase.

4.7.3 Exemplos Práticos e Constatações

Será ilustrativo, nesse ponto, examinar alguns exemplos práticos.

Para três sinais de exemplo, serão mostrados os efeitos da análise em 4 estágios por 2 conjuntos de bancos de filtros (I e II) cujos filtros passa-baixas e passa-altas são denotados $F_L(z)$ e $F_H(z)$, respectivamente.

O Banco I, descrito na Tabela 4.1 e na Figura 4.34, será Haar, exemplificando separação mais relaxada do espectro em duas metades.

Tabela 4.1: Respostas impulsivas dos filtros do Banco I.

i	f_L	f_H
0	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
1	$\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$

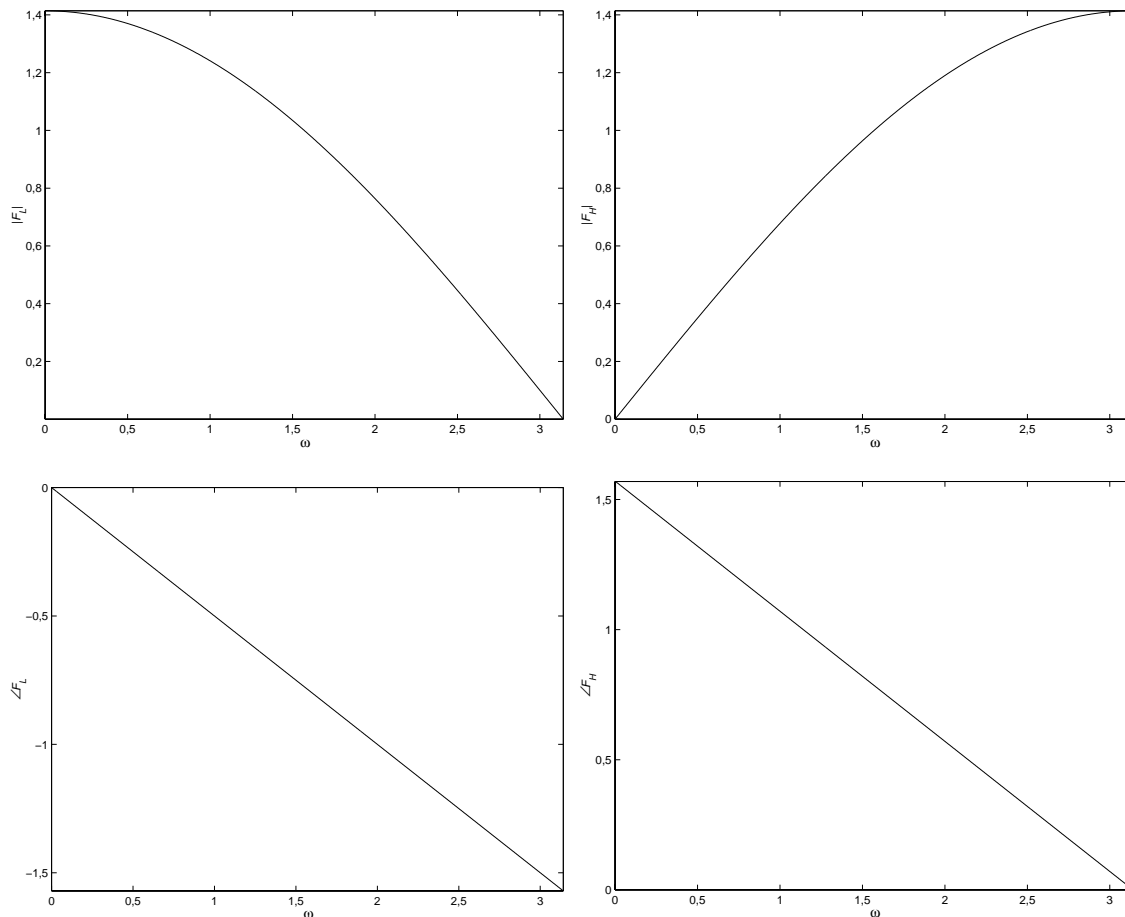


Figura 4.34: Respostas de módulo e fase dos filtros do Banco I.

O Banco II, descrito na Tabela 4.2 e na Figura 4.35, será de filtros de meia

banda, FIR de ordem 15, *equiripple*, com fase linear e com banda de transição de $\frac{\pi}{10}$, exemplificando separação mais abrupta do espectro em duas metades.

Tabela 4.2: Respostas impulsivas aproximadas dos filtros do Banco II.

i	f_L	f_H
0	-0,03069489015956	-0,03069489015956
1	-0,06235666533939	0,06235666533939
2	0,05003272754162	0,05003272754162
3	0,04036732258759	-0,04036732258759
4	-0,05905007705128	-0,05905007705128
5	-0,08851617614778	0,08851617614778
6	0,14913924986610	0,14913924986610
7	0,44949098790812	-0,44949098790812
8	0,44949098790812	0,44949098790812
9	0,14913924986610	-0,14913924986610
10	-0,08851617614778	-0,08851617614778
11	-0,05905007705128	0,05905007705128
12	0,04036732258759	0,04036732258759
13	0,05003272754162	-0,05003272754162
14	-0,06235666533939	-0,06235666533939
15	-0,03069489015956	0,03069489015956

Com base nos exemplos, tentar-se-á fazer a interpretação das transformações sofridas pelos sinais, bem como verificar as especulações feitas na Subseção 4.7.2.

Inicialmente, será examinado o caso de processos AR com predomínio de pólos de fase alta (entre $\frac{\pi}{2}$ e π), de Q alto e baixo, o Caso A. Seja, então, o sinal $s(k)$ perfeitamente modelado pelo processo AR cujo filtro gerador tem os pólos indicados na Tabela 4.3 e na Figura 4.36.

As Figuras 4.37 e 4.38 mostram os diagramas de pólos e zeros dos filtros geradores resultantes de cada estágio da análise do sinal do Caso A pelos Bancos I.

Após o primeiro estágio da análise, os pólos com fases originalmente próximas de π , devido à “duplicação” de suas fases, aparecem com fases próximas de 0. Em s_1 ,

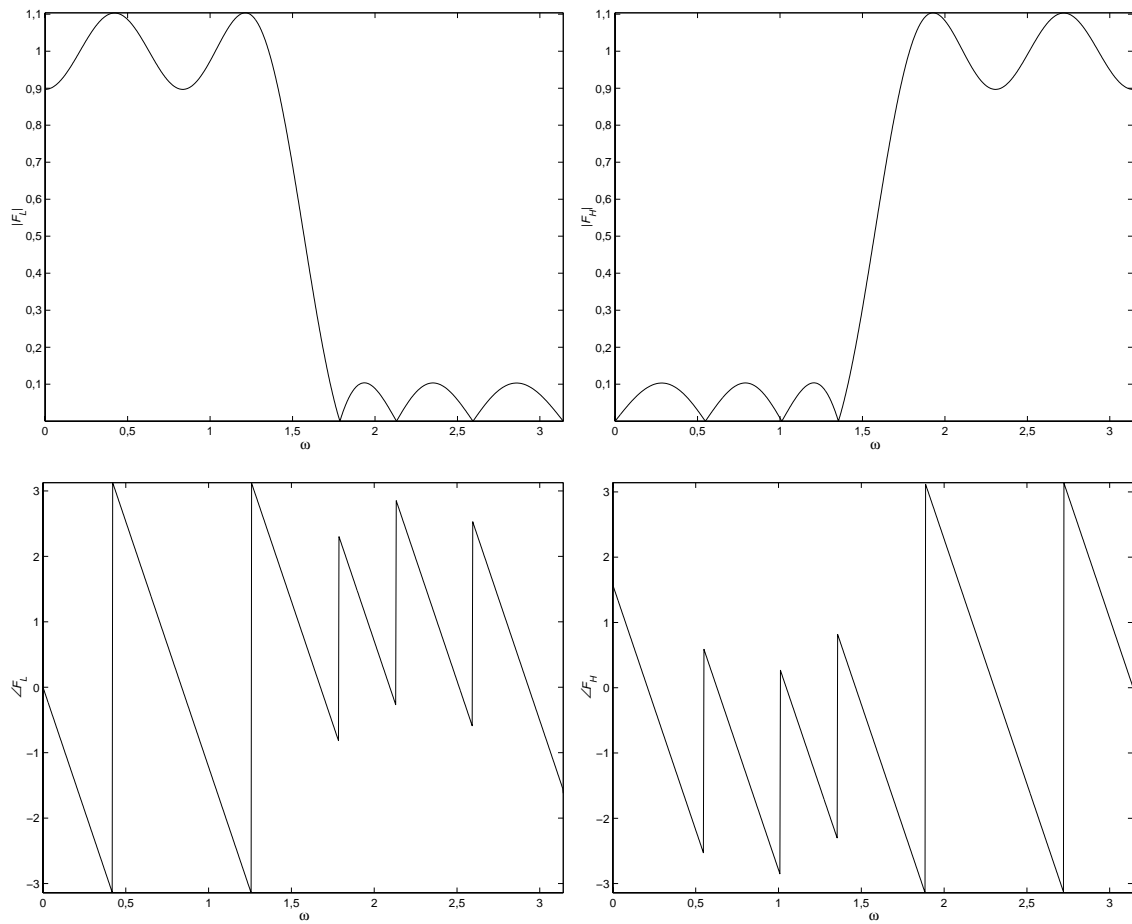


Figura 4.35: Respostas de módulo e fase dos filtros do Banco II.

Tabela 4.3: Pólos do filtro gerador do sinal de exemplo do Caso A.

r	θ
0,5	$\pm \frac{\pi}{2}$
0,95	$\pm \frac{\pi}{2}$
0,5	$\pm \frac{2\pi}{3}$
0,95	$\pm \frac{2\pi}{3}$
0,5	$\pm \left(\pi - \frac{\pi}{180}\right)$
0,95	$\pm \left(\pi - \frac{\pi}{180}\right)$

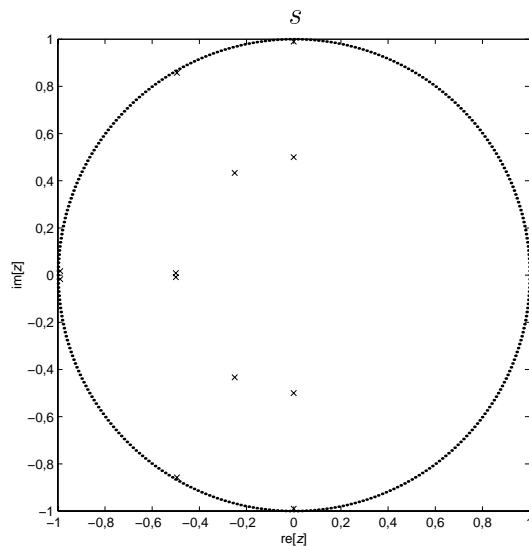


Figura 4.36: Pólos do filtro gerador do sinal de exemplo do Caso A.

que contém a metade superior do espectro, não aparecem acompanhados de zeros próximos; em \hat{s}_1 , que contém a metade inferior do espectro, são uma parcela de *aliasing* e aparecem acompanhados de um zero em $z = 1$, que “tenta eliminá-los”.

É fácil constatar-se o efeito de deslocamento de pólos (tanto pelo dobramento de suas fases quanto pelo quadramento de seus módulos) acompanhando-se, por exemplo, os de fase próxima de 0 ao longo dos estágios 1, 2 e 3.

A existência de um par de pólos originais com fase $\frac{2\pi}{3}$ também propicia o acompanhamento particular de seu módulo, porquanto suas fases dobradas resultam nos mesmos ângulos. É possível, então, verificar, nas sucessivas subdivisões inferiores do espectro \hat{s}_2 , \hat{s}_3 e \hat{s}_4 , um par de zeros “perseguido-os, tentando cancelá-los”.

As Figuras 4.39 e 4.40 mostram os diagramas de pólos e zeros dos filtros geradores resultantes de cada estágio da análise do sinal do Caso A pelos Bancos II.

O uso de bancos de filtros mais seletivos pode favorecer os cancelamentos, como se pode ver acompanhando os pólos com fase original $\frac{2\pi}{3}$ ao longo dos sinais \hat{s}_2 , \hat{s}_3 e \hat{s}_4 nas Figuras 4.37, 4.38 (Bancos I), 4.39 e 4.40 (Bancos II).

A Tabela 4.4 compara as complexidades dos modelos referentes ao Caso A analisado pelos Bancos I e II³. São mostrados o modelo dito completo—sem qualquer simplificação ou cancelamento—e o modelo dito reduzido—que teve cancelados zeros, pólos e pares zero-pólo que implicassem individualmente $\Delta M < 0,1$ e $\Delta\theta < \frac{\pi}{20}$.

³Implicamente, só estão sendo considerados zeros e pólos não-nulos.

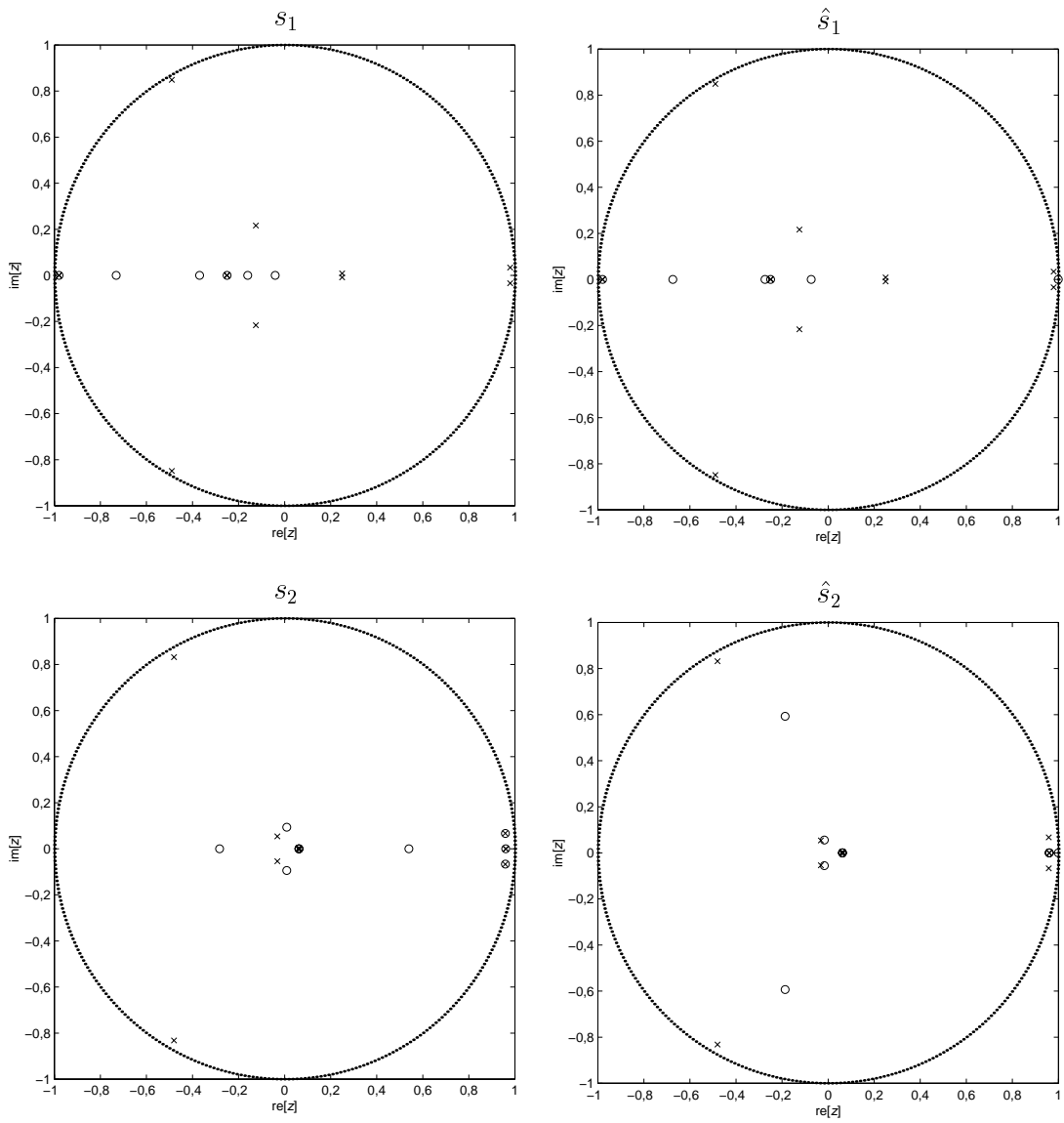


Figura 4.37: Análise do sinal de exemplo do Caso A pelos Bancos I—estágios 1 e 2: zeros e pólos.

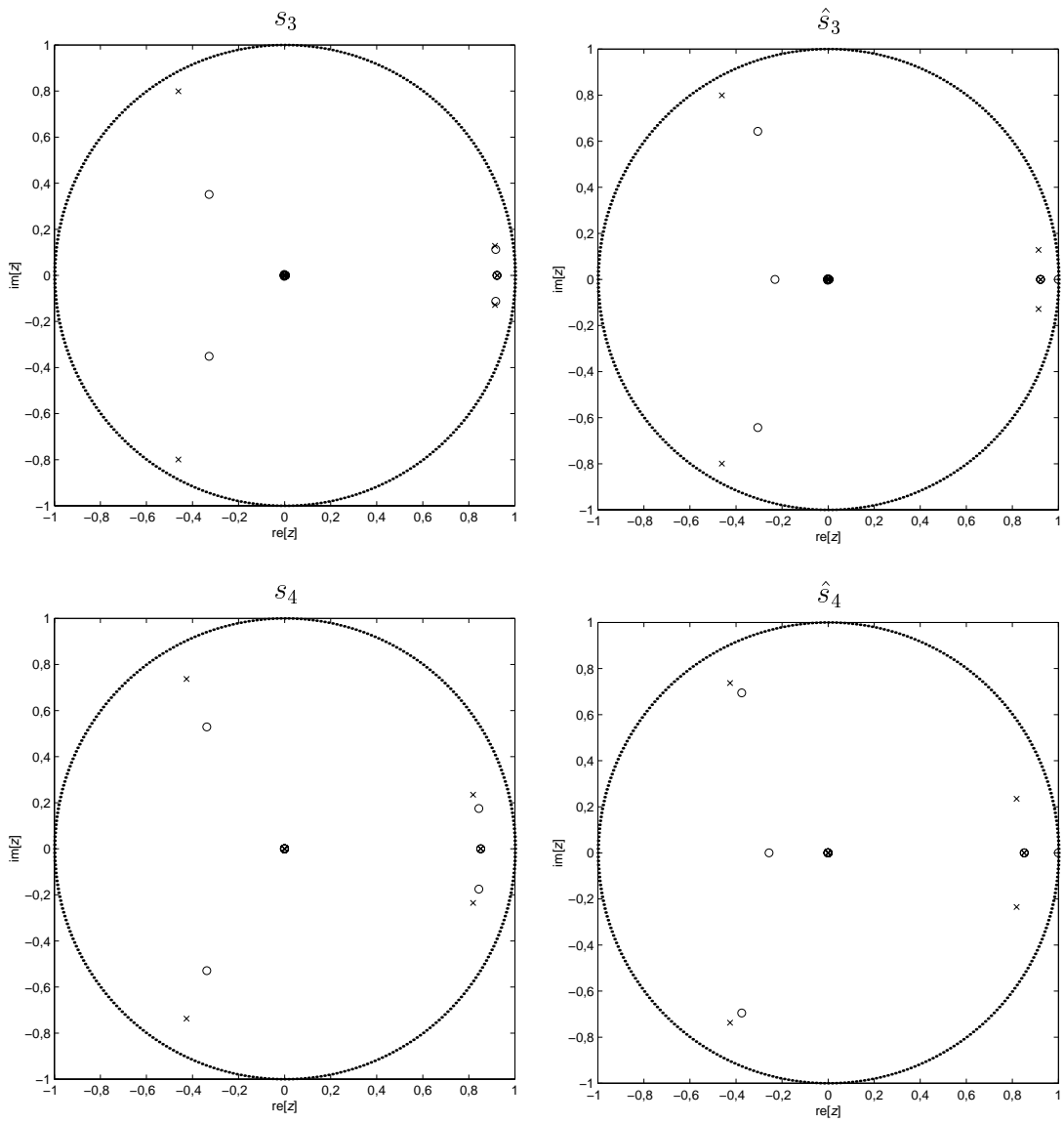


Figura 4.38: Análise do sinal de exemplo do Caso A pelos Bancos I—estágios 3 e 4: zeros e pólos.

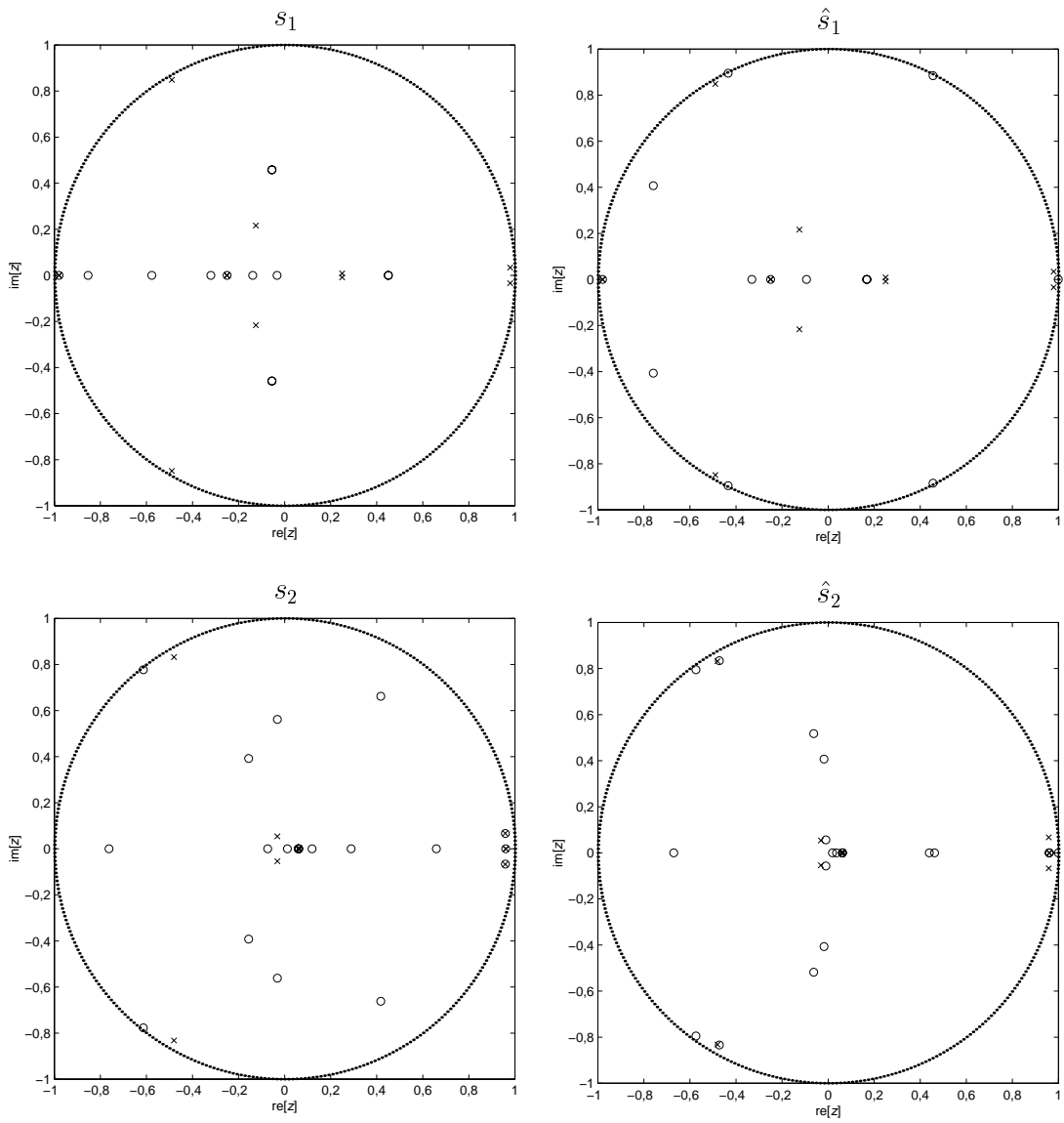


Figura 4.39: Análise do sinal de exemplo do Caso A pelos Bancos II—estágios 1 e 2: zeros e pólos.

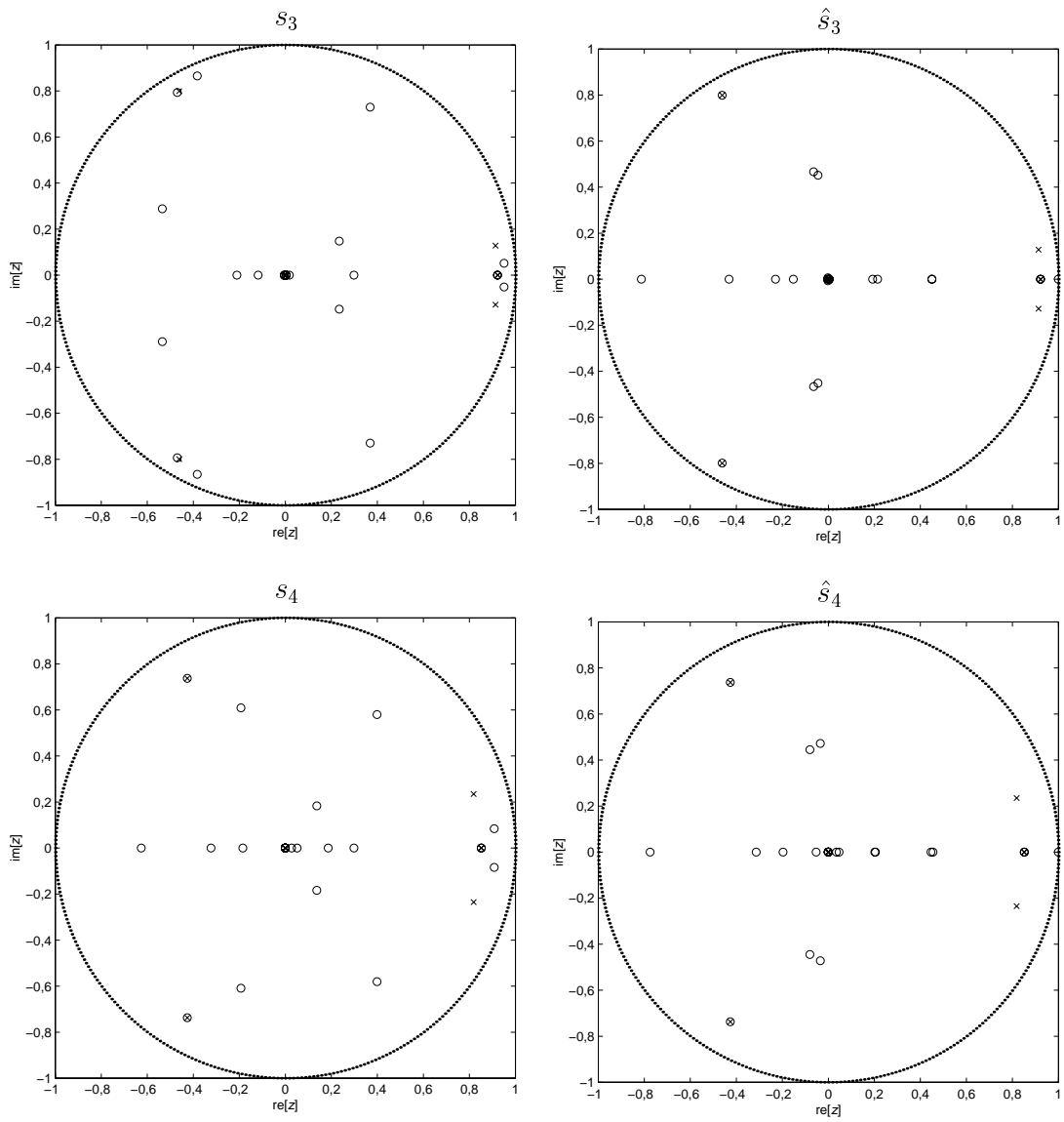


Figura 4.40: Análise do sinal de exemplo do Caso A pelos Bancos II—estágios 3 e 4: zeros e pólos.

Tabela 4.4: Número de zeros e número pólos dos filtros geradores - Caso A.

filtro gerador	Banco I				Banco II			
	completo		reduzido		completo		reduzido	
	zeros	pólos	zeros	pólos	zeros	pólos	zeros	pólos
\mathcal{M}_1	6	12	3	10	13	12	10	10
$\hat{\mathcal{M}}_1$	6	12	3	10	13	12	10	10
\mathcal{M}_2	9	12	2	3	20	12	12	3
$\hat{\mathcal{M}}_2$	9	12	3	4	20	12	8	2
\mathcal{M}_3	11	12	2	2	23	12	13	2
$\hat{\mathcal{M}}_3$	11	12	4	4	23	12	13	2
\mathcal{M}_4	12	12	4	4	25	12	13	2
$\hat{\mathcal{M}}_4$	12	12	4	4	25	12	12	2

Constata-se a gradual redução do número de pólos ao longo dos estágios de divisão. Para esse exemplo, pode-se dizer que a maior ordem dos filtros dos Bancos II não correspondeu a uma simplificação tão mais drástica dos modelos nas sub-bandas, em relação à obtida com os Bancos I; em contrapartida, criou-se um grande número de zeros não ligados ao cancelamento de pólos.

Em seguida, será examinado o caso de processos AR com predomínio de pólos de fase baixa, de Q alto e baixo (Caso B). Seja, então, o sinal $s(k)$ perfeitamente modelado pelo processo AR cujo filtro gerador tem os pólos indicados na Tabela 4.5 e na Figura 4.41.

Tabela 4.5: Pólos do filtro gerador do processo AR do sinal de exemplo do Caso B.

r	θ
0,5	$\pm \frac{\pi}{180}$
0,95	$\pm \frac{\pi}{180}$
0,5	$\pm \frac{\pi}{3}$
0,95	$\pm \frac{\pi}{3}$
0,5	$\pm \frac{\pi}{2}$
0,95	$\pm \frac{\pi}{2}$

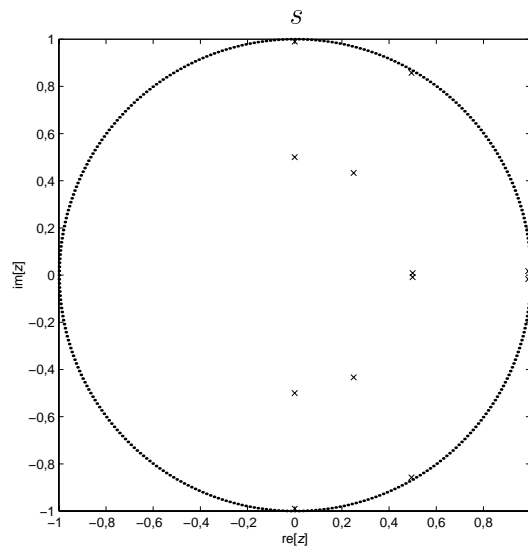


Figura 4.41: Pólos do filtro gerador do processo AR do sinal de exemplo do Caso B.

As Figuras 4.42 e 4.43 mostram os diagramas de pólos e zeros dos filtros geradores resultantes de cada estágio da análise do sinal do Caso B pelos Bancos I.

Após o primeiro estágio da análise, os pólos com fases originalmente próximas de 0, com sua “duplicação”, ainda aparecem com fases próximas de 0. Em s_1 , que contém a metade superior do espectro, aparecem acompanhados de um zero em $z = 1$, que “tenta eliminá-los”; em \hat{s}_1 , que contém a metade inferior do espectro, não aparecem acompanhados de zeros próximos.

Resumidamente, no Caso B (Figuras 4.42 e 4.43) o que se constata é aproximadamente o dual do comportamento do Caso A (Figuras 4.37 e 4.38).

As Figuras 4.44 e 4.45 mostram os diagramas de pólos e zeros dos filtros geradores resultantes de cada estágio da análise do sinal do Caso B pelos Bancos II.

Novamente, no Caso B (Figuras 4.44 e 4.45) o que se constata é aproximadamente o dual do comportamento do Caso A (Figuras 4.39 e 4.40).

Percebe-se, ainda, a formação gradual de uma configuração de zeros fixos com a evolução dos estágios de divisão (\hat{s}_3 e \hat{s}_4 , por exemplo), não ligados ao cancelamento de quaisquer pólos.

A Tabela 4.6 compara as complexidades dos modelos referentes ao Caso B analisado pelos Bancos I e II⁴. São mostrados o modelo dito completo—sem qualquer simplificação ou cancelamento—e o modelo dito reduzido—que teve cancelados

⁴Impliciteamente, só estão sendo considerados zeros e pólos não-nulos.

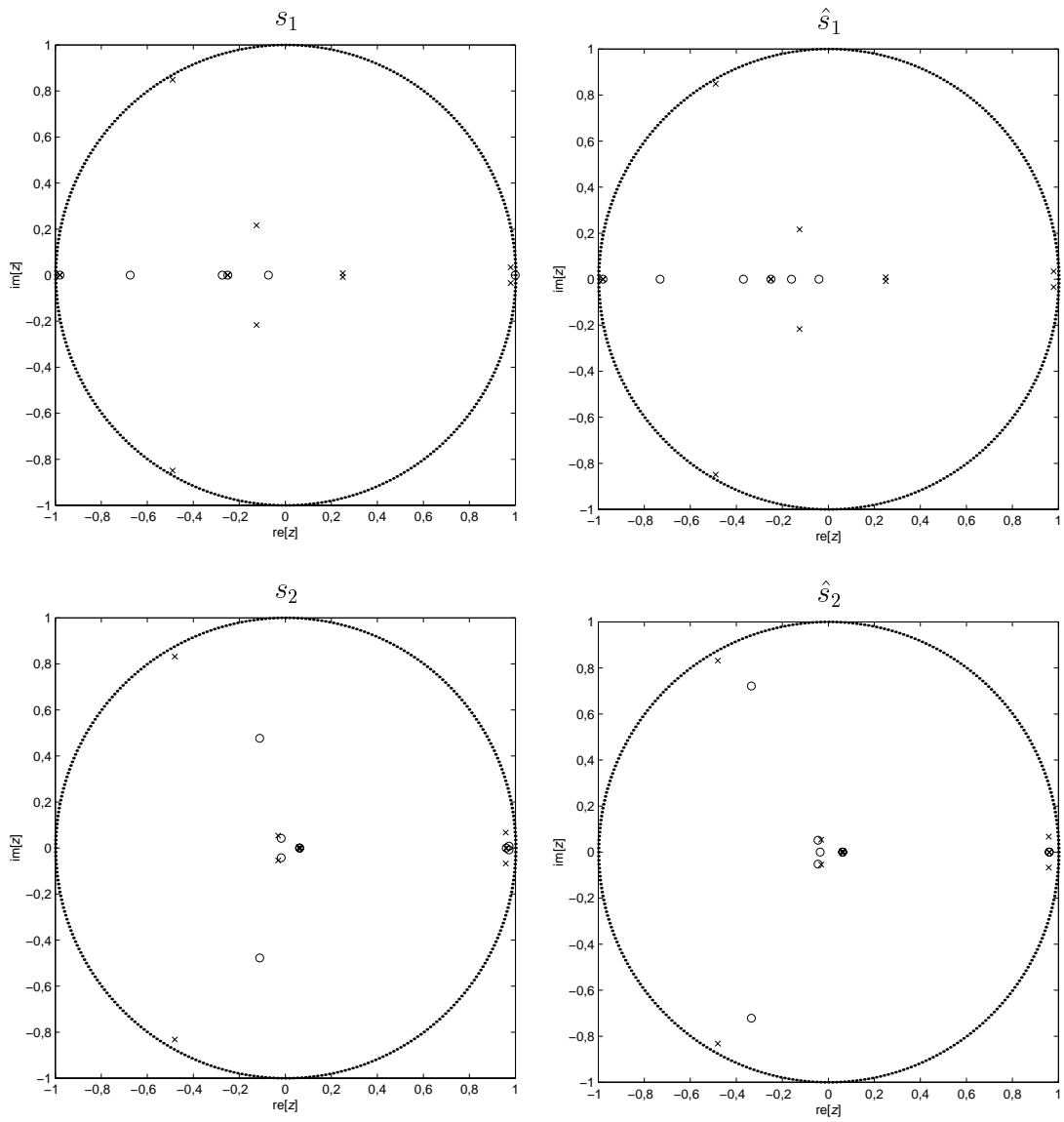


Figura 4.42: Análise do sinal de exemplo do Caso B pelos Bancos I—estágios 1 e 2: zeros e pólos.

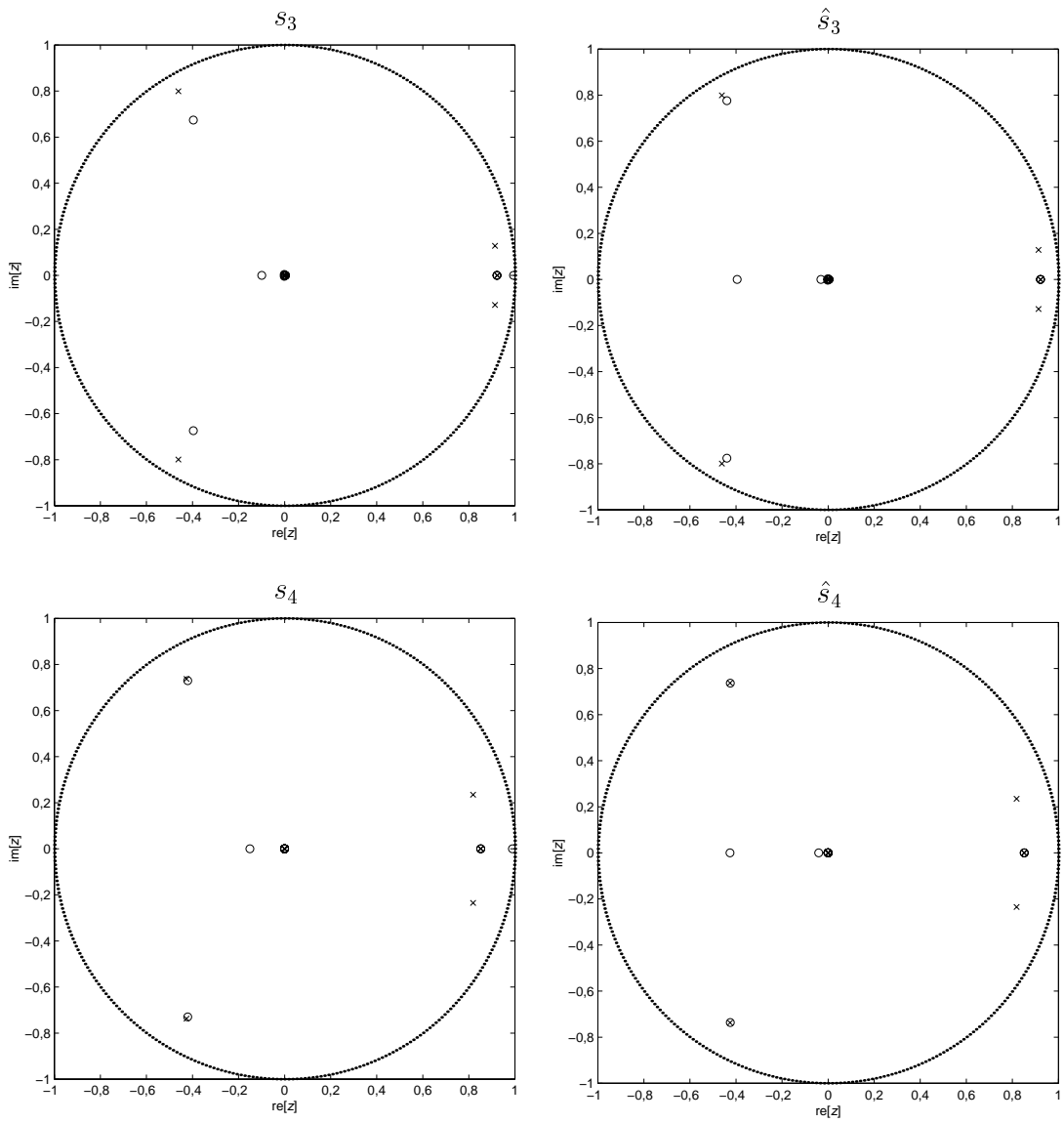


Figura 4.43: Análise do sinal de exemplo do Caso B pelos Bancos I—estágios 3 e 4: zeros e pólos.

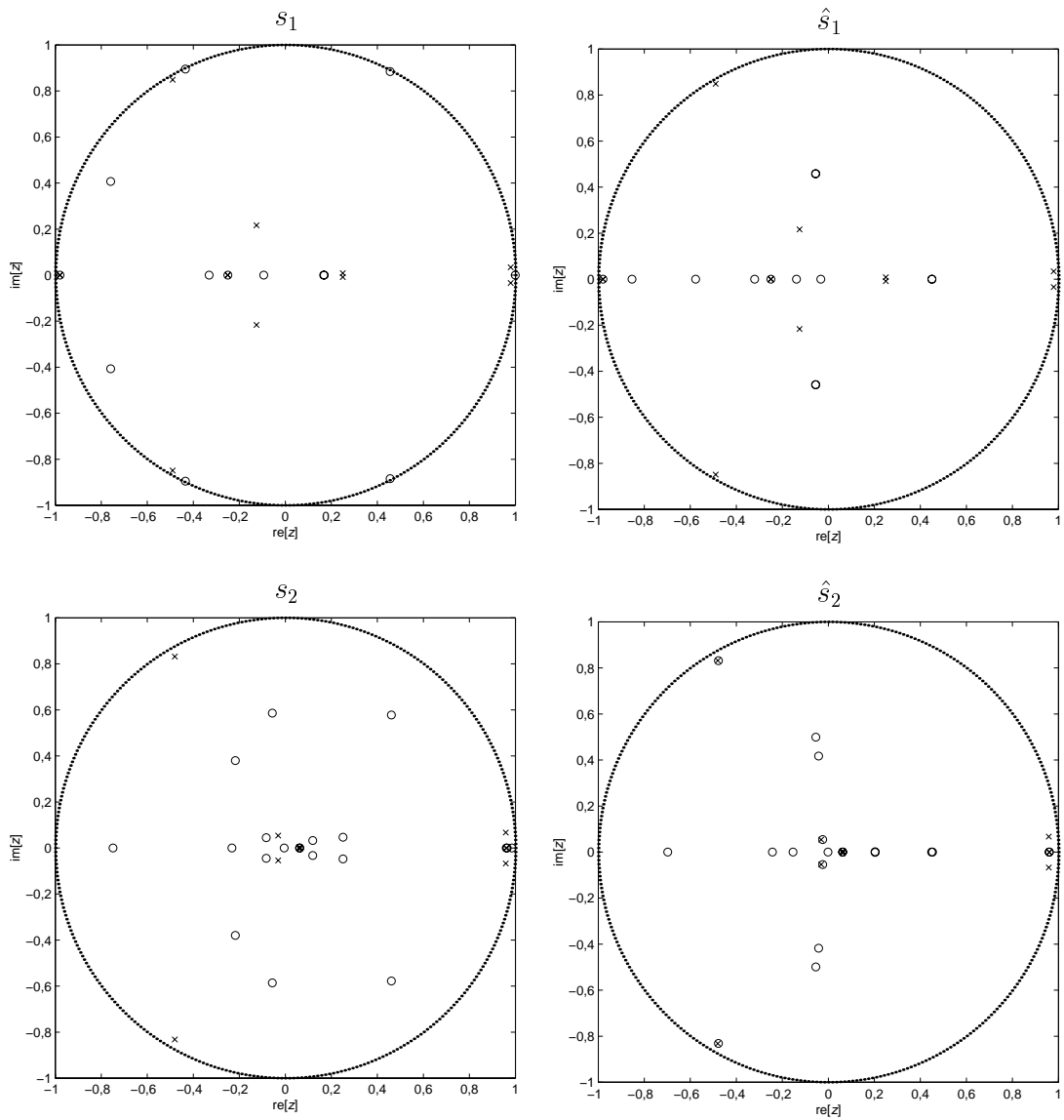


Figura 4.44: Análise do sinal de exemplo do Caso B pelos Bancos II—estágios 1 e 2: zeros e pólos.

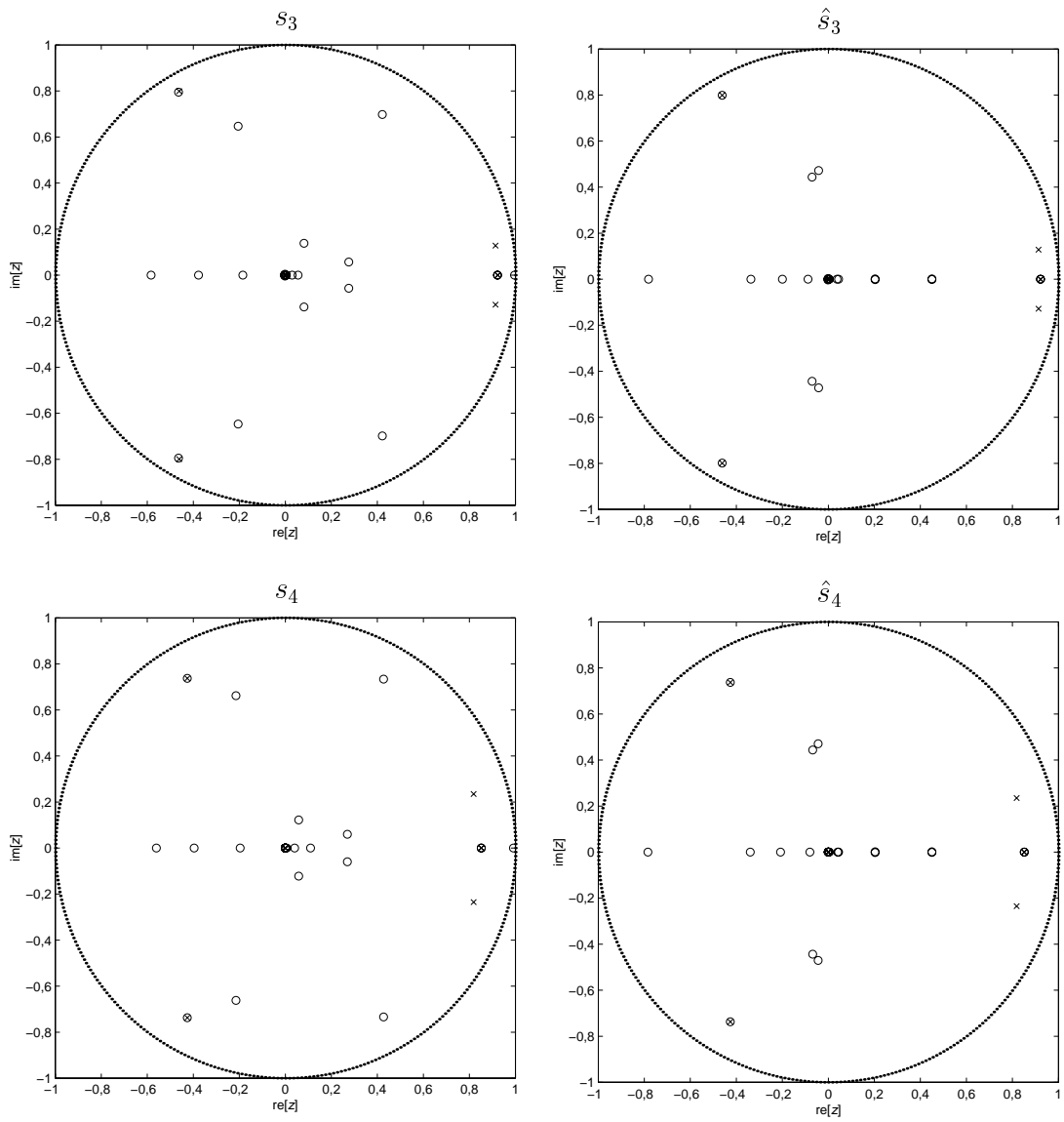


Figura 4.45: Análise do sinal de exemplo do Caso B pelos Bancos II—estágios 3 e 4: zeros e pólos.

zeros, pólos e pares zero-pólo que implicassem individualmente $\Delta M < 0,1$ e $\Delta\theta < \frac{\pi}{20}$.

Tabela 4.6: Número de zeros e número pólos dos filtros geradores - Caso B.

filtro gerador	Banco I				Banco II			
	completo		reduzido		completo		reduzido	
	zeros	pólos	zeros	pólos	zeros	pólos	zeros	pólos
\mathcal{M}_1	6	12	3	10	13	12	10	10
$\hat{\mathcal{M}}_1$	6	12	3	10	13	12	10	10
\mathcal{M}_2	9	12	4	5	20	12	11	4
$\hat{\mathcal{M}}_2$	9	12	2	4	20	12	11	2
\mathcal{M}_3	11	12	3	4	23	12	10	2
$\hat{\mathcal{M}}_3$	11	12	3	4	23	12	11	2
\mathcal{M}_4	12	12	2	2	25	12	11	2
$\hat{\mathcal{M}}_4$	12	12	1	2	25	12	11	2

Valem as mesmas observações que para o Caso A.

Por fim, será examinado o caso de um pequeno trecho de um sinal de áudio musical típico, com conteúdo espectral rico, proveniente de uma gravação de som orquestral (Caso C). Para este $s(k)$, foi obtido um modelo AR de ordem 40 cujo filtro gerador estimado tem os pólos indicados na Figura 4.46.

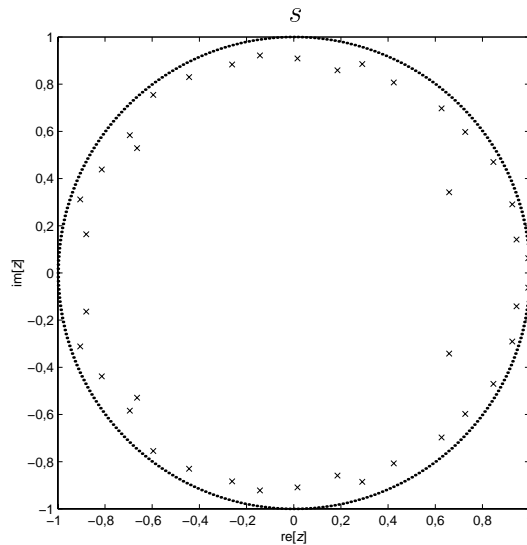


Figura 4.46: Pólos do filtro gerador do processo AR do sinal de exemplo do Caso C.

Deve ser observado que a configuração de pólos obtida resulta, em parte,

das próprias limitações do cálculo do modelo. Os pólos mais significativos são os mais próximos da circunferência unitária. No mais, a distribuição aproximadamente homogênea de pólos de Q médio tenta modelar aspectos pouco seletivos do espectro. Pólos de Q muito baixo são pouco significativos (v. Capítulo 2), e o fato de já não aparecerem naturalmente não compromete a modelagem.

As Figuras 4.47 e 4.48 mostram os diagramas de pólos e zeros dos filtros geradores resultantes de cada estágio da análise do sinal do Caso C pelos Bancos I.

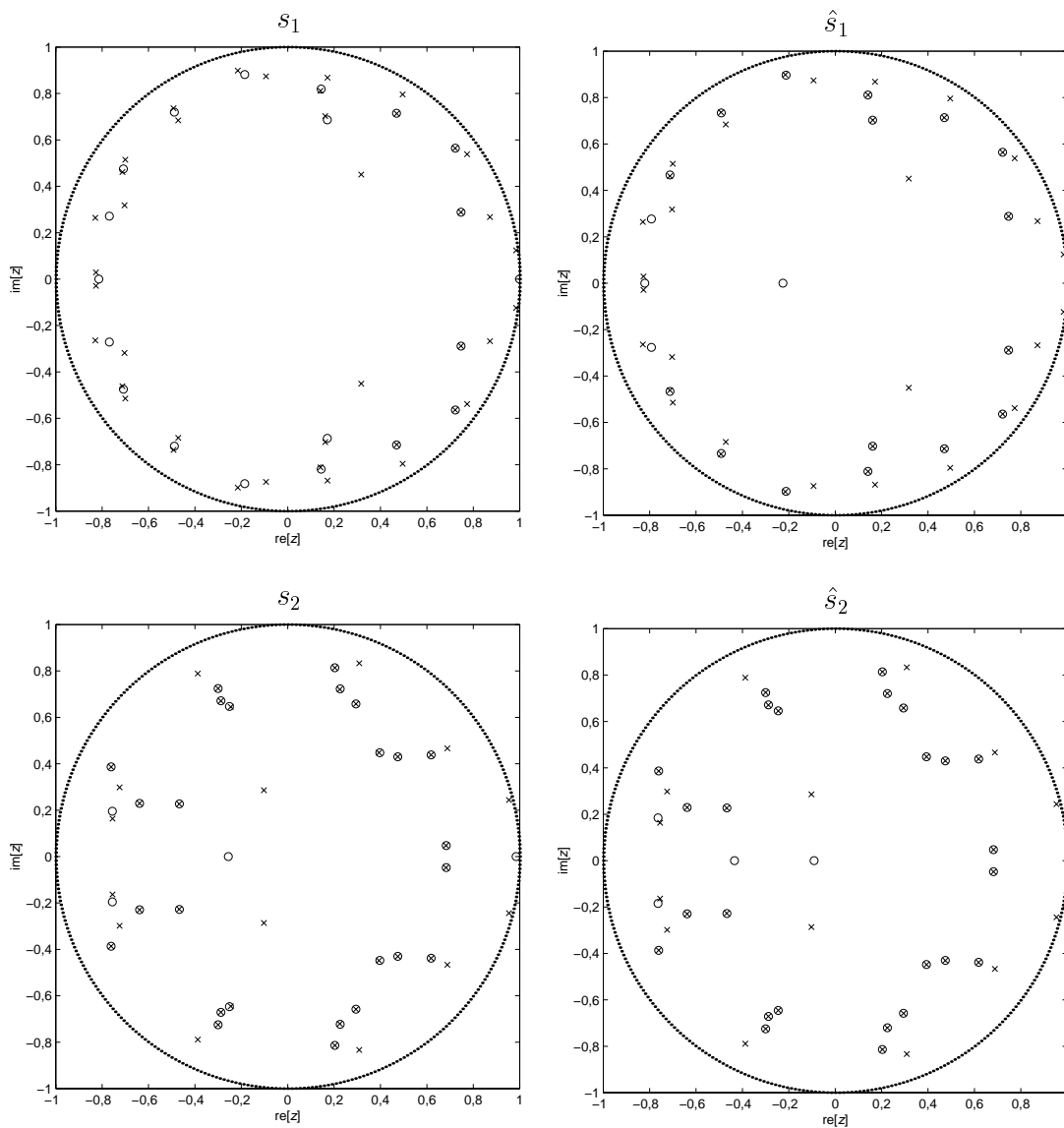


Figura 4.47: Análise do sinal de exemplo do Caso C pelos Bancos I—estágios 1 e 2: zeros e pólos.

As Figuras 4.49 e 4.50 mostram os diagramas de pólos e zeros dos filtros geradores resultantes de cada estágio da análise do sinal do Caso C pelos Bancos II.

A Tabela 4.7 compara as complexidades dos modelos referentes ao Caso C

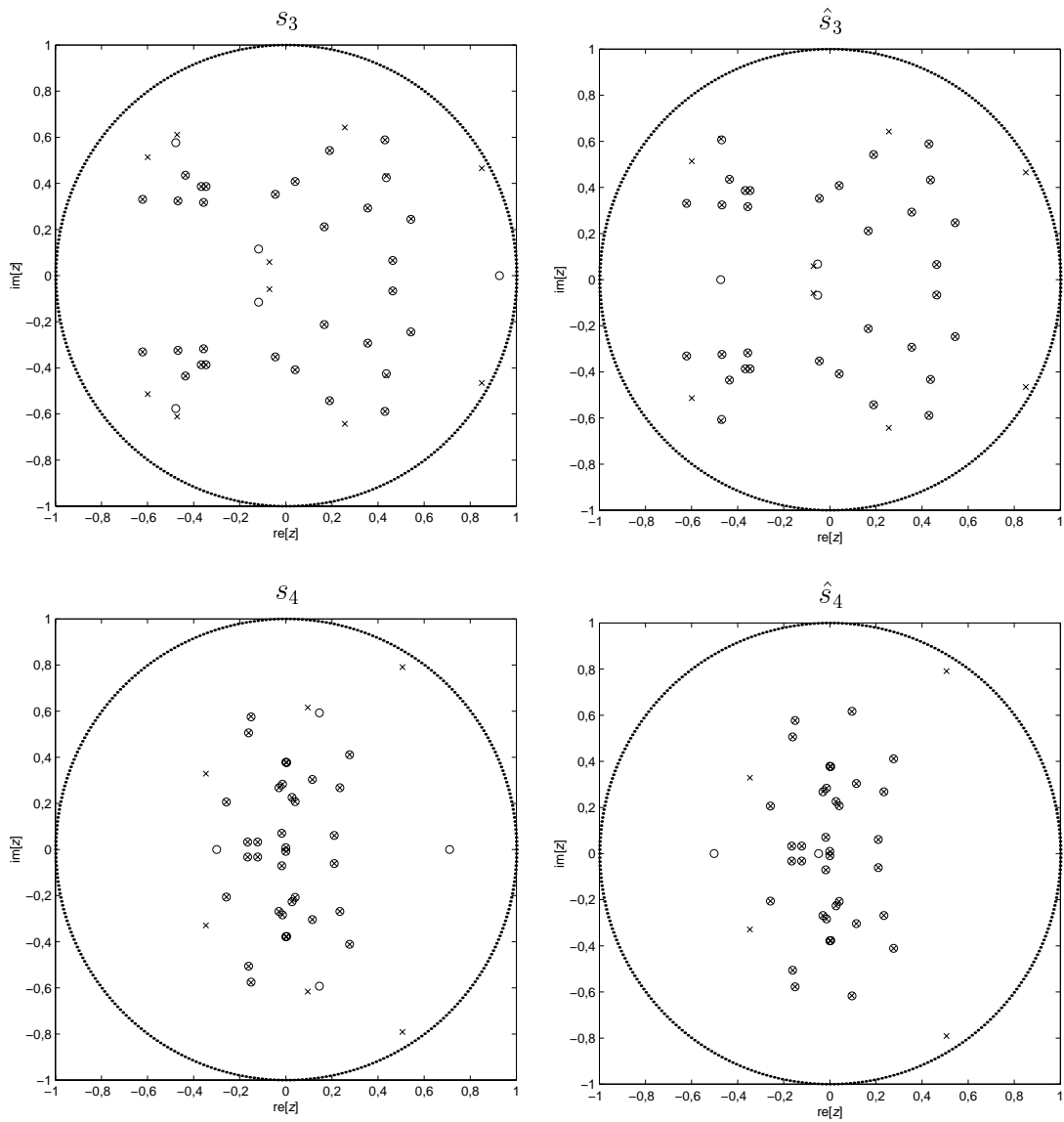


Figura 4.48: Análise do sinal de exemplo do Caso C pelos Bancos I—estágios 3 e 4: zeros e pólos.

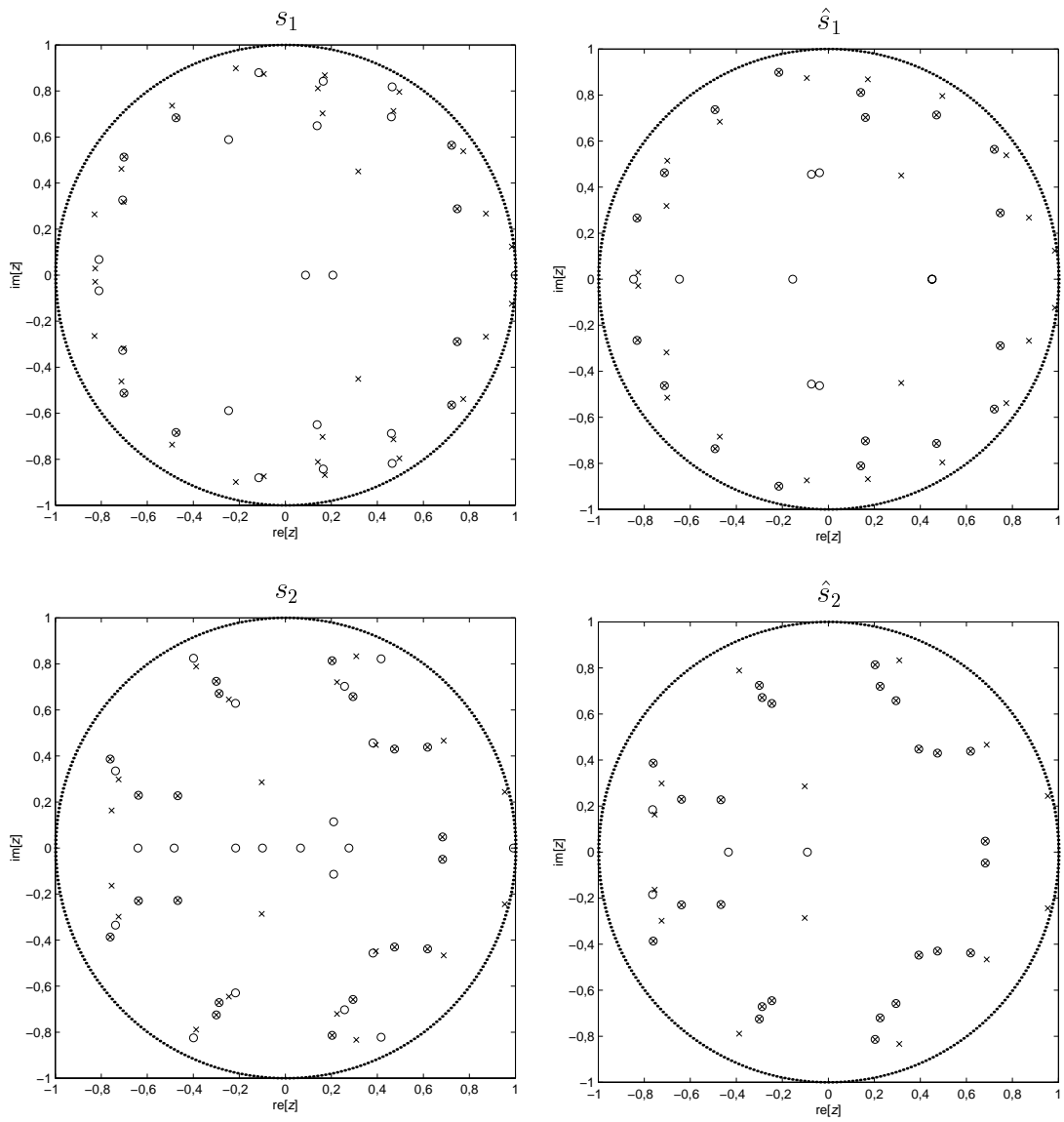


Figura 4.49: Análise do sinal de exemplo do Caso C pelos Bancos II—estágios 1 e 2: zeros e pólos.

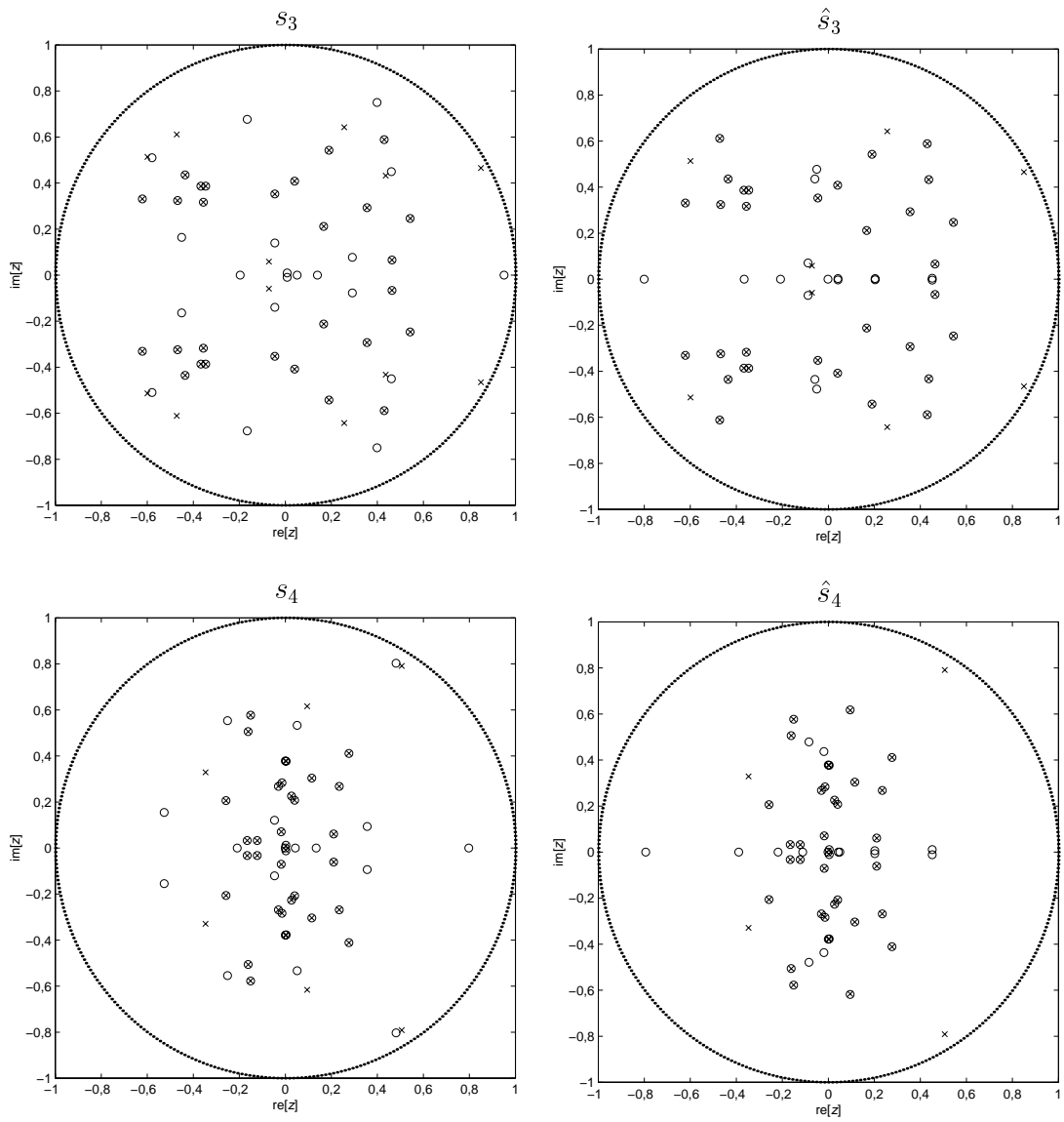


Figura 4.50: Análise do sinal de exemplo do Caso C pelos Bancos II—estágios 3 e 4: zeros e pólos.

analisado pelos Bancos I e II⁵. São mostrados o modelo dito completo—sem qualquer simplificação ou cancelamento—e o modelo dito reduzido—que teve cancelados zeros, pólos e pares zero-pólo que implicassem individualmente $\Delta M < 0,1$ e $\Delta\theta < \frac{\pi}{20}$.

Tabela 4.7: Número de zeros e número pólos dos filtros geradores - Caso C.

filtro gerador	Banco I				Banco II			
	completo		reduzido		completo		reduzido	
	zeros	pólos	zeros	pólos	zeros	pólos	zeros	pólos
\mathcal{M}_1	20	40	6	26	27	40	8	22
$\hat{\mathcal{M}}_1$	20	40	3	27	27	40	9	22
\mathcal{M}_2	30	40	2	12	41	40	11	12
$\hat{\mathcal{M}}_2$	30	40	1	12	41	40	12	14
\mathcal{M}_3	35	40	3	6	48	40	13	8
$\hat{\mathcal{M}}_3$	35	40	1	12	48	40	11	6
\mathcal{M}_4	38	40	2	4	51	40	11	4
$\hat{\mathcal{M}}_4$	38	40	1	4	51	40	10	2

Não há nenhum aspecto especial a destacar, exceto que, dadas as características especializadas do modelo original, desta vez o uso de filtros mais seletivos (Banco II) levou, a partir do estágio 3, a uma simplificação bem maior dos modelos das sub-bandas.

Para sinais reais típicos, é possível adotar modelos ARMA de ordem decrescente ao longo das subdivisões do espectro. Após as possíveis simplificações, os zeros remanescentes dos filtros geradores do modelo permitem moldar com facilidade os “vales” dos espectro dos sinais nas sub-bandas; contudo, é possível empregar filtros geradores só-pólos com um número superior de pólos (isto é, substituir os modelos ARMA por modelos AR de ordem superior), com boa aproximação.

⁵Implicitamente, só estão sendo considerados zeros e pólos não-nulos.

Capítulo 5

Aliasing e Reconstrução Perfeita no Processamento Multitaxa

Neste Capítulo, são discutidas a necessidade de reconstrução perfeita e a forma de eliminação de *aliasing* nos sistemas em multitaxa quando se realiza processamento independente dos sinais nas sub-bandas.

Em diversas situações, um sinal a ser processado tem características tão variadas ao longo do espectro de frequência que justificam sua prévia separação em faixas diferenciadas para posterior processamento.

Um sistema assim constituído, suposto linear e invariante no tempo, pode assumir a forma geral da Figura 5.1, onde N é o número de sub-bandas, $F_n(z)$ compõem o banco de filtros de análise e $P_n(z)$ são os processamentos diferenciados por sub-banda. Sua operação pode ser descrita como

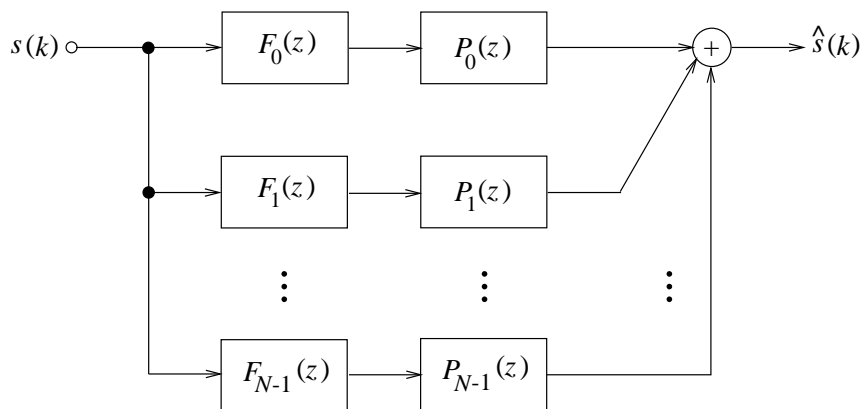


Figura 5.1: Sistema de processamento em N bandas.

$$\hat{S}(z) = \left(\sum_{n=0}^{N-1} P_n(z) F_n(z) \right) S(z). \quad (5.1)$$

Para que o sistema permita reconstrução perfeita (PR) do sinal de entrada na ausência de processamento nas sub-bandas, o que é desejável, basta que

$$\sum_{n=0}^{N-1} F_n(z) = z^{-\Delta}, \quad (5.2)$$

onde Δ é o atraso gerado pelo banco de análise.

Então, caso se deseje processar $s(k)$ de modo a obter $\hat{s}(k)$ com o uso da estrutura da Fig. 5.1 satisfazendo a reconstrução perfeita, é possível projetar as funções $P_n(z)$ que realizem o processamento desejado em cada sub-banda, a partir do conhecimento do banco dos filtros do banco de análise. Isso pode ser verificado, por exemplo, para o caso simples de 2 sub-bandas, em que a equação (5.1) se torna

$$\hat{S}(z) = [P_1(z)z^{-\Delta} + (P_0(z) - P_1(z)) F_0(z)] S(z). \quad (5.3)$$

Isso já indica um tipo de relacionamento entre as funções $P_n(z)$, mas não propriamente restritivo.

Com o intuito de reduzir a complexidade do sistema e sob a inspiração do Teorema de Nyquist [26], é possível elaborar um sistema multitaxa [38, 39] em que o processamento seja efetuado em sinais de taxas menores que a original. Agora, além de permitir a realização de processamentos de complexidades adequadas a cada sub-banda, com a redução de taxa o sistema em sub-bandas pode se mostrar mais eficiente do que uma solução equivalente em banda completa.

Um sistema típico, em que as sub-bandas sejam obtidas após iguais decimações, maximamente decimado, pode assumir a forma da Figura 5.2, onde $G_n(z)$ compõem o banco de filtros de síntese. A descrição de sua operação é bem mais elaborada, mas bastará ao propósito desta Seção examinar o caso simples em 2 sub-bandas ($N = 2$). Com referência às variáveis definidas na Figura 5.2, tem-se

$$\begin{aligned} X_0(z) &= F_0(z)S(z) \\ X_1(z) &= F_1(z)S(z) \end{aligned} \Rightarrow \quad (5.4)$$

$$\begin{aligned} \Rightarrow X_{0_d}(z) &= \frac{F_0(z^{1/2})S(z^{1/2}) + F_0(-z^{1/2})S(-z^{1/2})}{2} \\ \Rightarrow X_{1_d}(z) &= \frac{F_1(z^{1/2})S(z^{1/2}) + F_1(-z^{1/2})S(-z^{1/2})}{2} \end{aligned} \Rightarrow \quad (5.5)$$

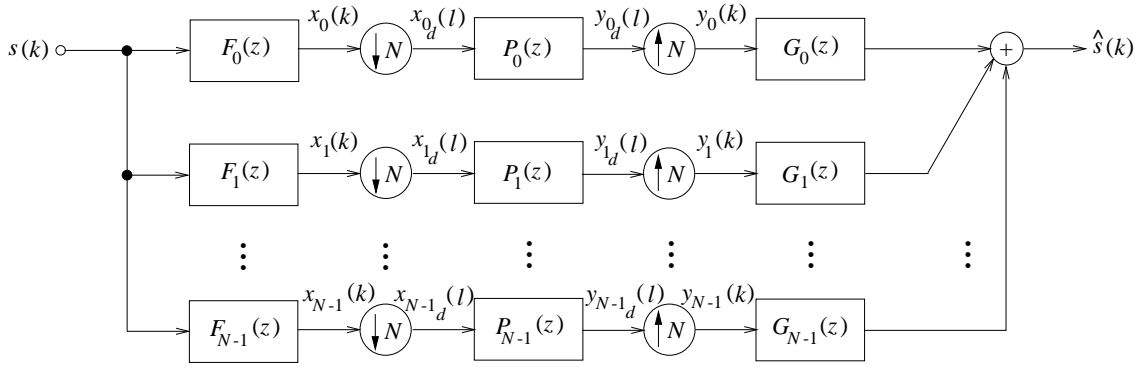


Figura 5.2: Sistema de processamento multitaxa em N sub-bandas.

$$\Rightarrow \begin{aligned} Y_{0_d}(z) &= P_0(z) \frac{F_0(z^{1/2})S(z^{1/2}) + F_0(-z^{1/2})S(-z^{1/2})}{2} \\ Y_{1_d}(z) &= P_1(z) \frac{F_1(z^{1/2})S(z^{1/2}) + F_1(-z^{1/2})S(-z^{1/2})}{2} \end{aligned} \Rightarrow \quad (5.6)$$

$$\Rightarrow \begin{aligned} Y_0(z) &= P_0(z^2) \frac{F_0(z)S(z) + F_0(-z)S(-z)}{2} \\ Y_1(z) &= P_1(z^2) \frac{F_1(z)S(z) + F_1(-z)S(-z)}{2} \end{aligned} \Rightarrow \quad (5.7)$$

$$\begin{aligned} \Rightarrow \hat{S}(z) &= G_0(z)P_0(z^2) \frac{F_0(z)S(z) + F_0(-z)S(-z)}{2} + G_1(z)P_1(z^2) \frac{F_1(z)S(z) + F_1(-z)S(-z)}{2} = \\ &= \frac{G_0(z)F_0(z)P_0(z^2) + G_1(z)F_1(z)P_1(z^2)}{2} S(z) + \frac{G_0(z)F_0(-z)P_0(z^2) + G_1(z)F_1(-z)P_1(z^2)}{2} S(-z). \end{aligned} \quad (5.8)$$

Na equação (5.8), o fator que multiplica $S(z)$ representa o processamento resultante sofrido pelo sinal e a segunda parcela representa o *aliasing* [40] causado pelo sistema.

Para que o sistema permita reconstrução perfeita do sinal de entrada na ausência de processamento nas sub-bandas, basta que

$$G_0(z)F_0(-z) + G_1(z)F_1(-z) = 0, \quad (5.9)$$

para que não ocorra *aliasing*, e que

$$G_0(z)F_0(z) + G_1(z)F_1(z) = z^{-\Delta}, \quad (5.10)$$

para que não ocorra distorção.

Assim, retornando à equação (5.8),

$$\hat{S}(z) = \frac{P_1(z^2)z^{-\Delta} + G_0(z)(P_0(z^2) - P_1(z^2))F_0(z)}{2} S(z) + \frac{G_0(-z)(P_0(z^2) - P_1(z^2))F_0(z)}{2} S(-z). \quad (5.11)$$

Isso significa que, ocorrendo processamento nas sub-bandas do sistema com bancos de filtros com PR genéricos, o *aliasing* só é eliminado se

$$P_1(z^2) = P_0(z^2). \quad (5.12)$$

Se um dos argumentos para a realização do processamento em sub-bandas é a possibilidade de particularizar o processamento conforme a faixa de frequência, a condição encontrada de igualdade dos processamentos nas sub-bandas suscita o questionamento da “verdadeira” reconstrução perfeita.

Relaxando-se a condição de PR, a simples eliminação do *aliasing* requer

$$G_0(z)F_0(-z)P_0(z^2) + G_1(z)F_1(-z)P_1(z^2) = 0. \quad (5.13)$$

Isso significa que, ocorrendo processamento nas sub-bandas do sistema com bancos de filtros genéricos, o *aliasing* só é eliminado se

$$P_1(z^2) = -\frac{G_0(z)H_0(-z)}{G_1(z)H_1(-z)}P_0(z^2). \quad (5.14)$$

Essa dependência indesejável entre os processamentos nas sub-bandas questiona, portanto, até mesmo a “verdadeira” eliminação de *aliasing*.

Por outro lado, preservando a generalidade e a independência de $P_n(z)$, o *aliasing* só é eliminado se

$$G_0(z)F_0(-z) = 0 \quad (5.15)$$

e

$$G_1(z)F_1(-z) = 0, \quad (5.16)$$

o que só poderia ser atendido por filtros ideais.

Vale a pena ressaltar, nesse ponto, que a solução para sistemas com $N > 2$ chegaria a restrições similares, apenas em maior número.

Então:

- O atendimento da PR pelos bancos de filtros não garante a eliminação do *aliasing* para processamentos genéricos realizados nas sub-bandas.
- A eliminação total do *aliasing* para processamentos genéricos realizados nas sub-bandas não é possível para bancos de filtros realizáveis.

Com isso, mostramos que nos sistemas multitaxa em que se realiza o processamento independente dos sinais nas sub-bandas a condição de reconstrução perfeita dos bancos de filtros pode não ser tão importante quanto freqüentemente se a considera e que mesmo apenas as condições para eliminação do *aliasing* no processamento resultante só podem ser atendidas aproximadamente.

Esta conclusão é exemplificada em restauração de áudio na Subseção 10.1.4.2.

Referindo-se à codificação de sinais, em [41] mostra-se que os bancos de filtros ótimos na ocorrência de quantização não apresentam reconstrução perfeita, propondo-se como alternativa filtros perceptivamente superiores.

Parte II

Restauração

Esta Parte trata da restauração de áudio propriamente dita, e inicia abordando Psicoacústica e avaliação objetiva de qualidade (Capítulo 6), importante instrumento auxiliar de validação. A seguir, trata-se da supressão de falhas localizadas, a saber, ruído impulsivo (Capítulo 8) e pulsos longos (Capítulo 9), passando pela interpolação de amostras faltantes (Capítulo 7), parte integrante de algumas técnicas. Em seguida, aborda-se o ruído de fundo (Capítulo 10) e uma técnica adaptativa aplicável quando se dispõe de mais de uma matriz da gravação a tratar (Capítulo 11). Por fim, avança-se na discussão da realização dos tratamentos de ruído impulsivo e de ruído de fundo em sub-bandas (Capítulo 12).

Capítulo 6

Psicoacústica e Avaliação Objetiva de Qualidade

O uso de técnicas de restauração de som gravado é freqüentemente questionado nos seguintes termos: qual o limite para a interferência na matriz gravada tal que não altere suas qualidades intrínsecas, isto é, que não comprometa a sua autenticidade como obra de arte? Essa discussão é comum à restauração de arte, em geral, e passa pela definição do seu objetivo.

Qual o alvo da restauração ideal de uma gravação musical?

1. Retornar ao estado original da matriz gravada. Este pode ser o objetivo do tratamento no caso em que o uso e/ou envelhecimento deteriorou o estado da matriz disponível.
2. Retornar ao som emitido por ocasião da gravação. Este pode ser o objetivo do tratamento quando imperfeições inerentes às próprias técnicas de registro podem ser eliminadas *a posteriori*.

À primeira vista o primeiro objetivo parece inquestionável, enquanto que o segundo soa como se envolvesse uma componente de adivinhação. Mas, em qualquer caso, a única forma de responder pela **fidelidade** buscada é comparar o sinal tratado a um padrão.

Na maior parte das vezes, o “padrão” disponível é a associação da audição com a memória, gerando uma “expectativa de qualidade”. A variabilidade do julgamento, considerado nesses termos, é muito grande, influenciada por fatores que vão

dos biológicos aos culturais: as pessoas ouvem e reagem ao que ouvem diferentemente. Essa componente imponderável, além de provocar a suspeita do ouvinte-alvo, dificulta um julgamento “científico” das técnicas de restauração.

Neste Capítulo faz-se uma breve exposição sobre as formas usualmente aceitas de aferição da qualidade do áudio em diversas aplicações. São revistos princípios de Psicoacústica, uma ciência que tenta caracterizar a percepção do som pelo ser humano “médio” e tem sido cada vez mais útil no processamento de áudio. Em seguida, descreve-se o moderno método para comparação perceptiva entre sinais de áudio proposto em [42]. Por fim, mostra-se como adaptar e aplicar esse método à avaliação objetiva e comparação do desempenho de algoritmos de restauração de áudio.

6.1 Avaliação da Qualidade de Sinais de Áudio

A maior parte do tempo, ao longo deste trabalho, referimo-nos a áudio tendo em mente áudio musical gravado. Em termos amplos, o áudio compreende todo fenômeno perceptível pelo ouvido humano, o que torna a fala um subconjunto do áudio. Mas a importância particular da fala como forma de comunicação justifica que se a trate em separado. E de fato, o que se espera como qualidade [43] da fala ou do “áudio” podem ser coisas distintas: processar um sinal de voz pode ter por objetivo torná-lo inteligível sem necessariamente ser agradável, enquanto o que se busca no áudio musical gravado é a fidelidade de reprodução. Contudo, sua modelagem e seu processamento têm muito em comum, e muitos conceitos e fundamentos ligados às formas de avaliação de sua qualidade são compartilhados.

Duas grandes famílias de processamento de áudio requerem avaliação de qualidade: a **codificação/síntese**, após a qual se precisa saber quanto foi alterada para pior a qualidade do sinal, e a **valorização** (*enhancement*), após a qual é preciso avaliar sua melhoria. Em ambos os casos, ela pode ser feita através de medidas **objetivas** ou **subjetivas**, dependendo do interesse do avaliador e das características do sinal. As medidas subjetivas se baseiam no ouvinte e avaliam o sinal através de uma nota dada, dentro de uma escala previamente combinada, por um grupo de pessoas normalmente especializadas no assunto. As objetivas utilizam alguma comparação

matemática, que, por sua vez, se baseia em dois métodos, principalmente: distância numérica e modelos de percepção da audição.

É importante frisar que a credibilidade de qualquer medida adotada deve exibir **consistência**, ou manutenção das mesmas condições de realização, e **representatividade**, ou abrangência dos testes.

A preferência pelos métodos objetivos tem como motivação superar as principais limitações práticas das medidas subjetivas: a necessidade de grande quantidade de tempo e de pessoas e a difícil reprodutibilidade. Mas a qualidade de um avaliador objetivo está na sua capacidade de reproduzir, o mais proximamente possível, o que seriam os resultados da avaliação humana.

Métodos de distância numérica tendem a comparar sinais segundo critérios inteiramente matemáticos, distanciando-se da avaliação subjetiva. Suponha-se, por hipótese, que sejam geradas 2 versões de um sinal contaminadas, respectivamente, com 2 ruídos de mesma potência, um acima de 20kHz e outro dentro do espectro audível. Perceptivamente, o primeiro soa intacto e o segundo deteriorado, embora ambos tenham a mesma razão sinal-ruído. Na Seção 10.2.5.2, outros exemplos reais ilustram esse ponto.

Métodos baseados na percepção podem empregar a modelagem biofísica extremamente complexa da audição (o que ainda não se tem mostrado prático) ou princípios psicoacústicos (de utilização crescente).

6.2 Psicoacústica

A Psicoacústica é uma matéria multidisciplinar, originada no início deste século, que tenta obter modelos para a percepção do som pela representação dos sinais de áudio na forma com que supostamente o faz o sistema auditivo humano, isto é, passando de um domínio físico para um psicofísico. Sua aplicação à compressão de sinais de áudio (só representando o que é perceptível), por exemplo, tem propiciado taxas de compressão antes inimagináveis com elevada qualidade [33].

Os princípios básicos de Psicoacústica mais importantes na avaliação de qualidade são descritos a seguir.

6.2.1 Audibilidade

A percepção da intensidade sonora varia segundo muitos fatores, sendo os mais importantes e de mais fácil quantificação a intensidade do som (pressão sonora) e sua frequência.

O Diagrama de Fletcher [20, 44], mostrado na Figura 6.1, descreve quantitativamente a sensação auditiva humana média, tomando o limiar da audição como 0 dB e prosseguindo em gradação relativa até o limiar da dor, em 120 dB. Cada curva representa qual o nível sonoro de um tom puro em dB *versus* frequência em Hz (dentro dos limites admitidos para a audição humana, 20 Hz e 20 kHz) que resulta num dado valor de audibilidade em dB. A calibração é feita de tal forma que 0 dB seja a intensidade correspondente ao limiar da audição em 1 kHz.

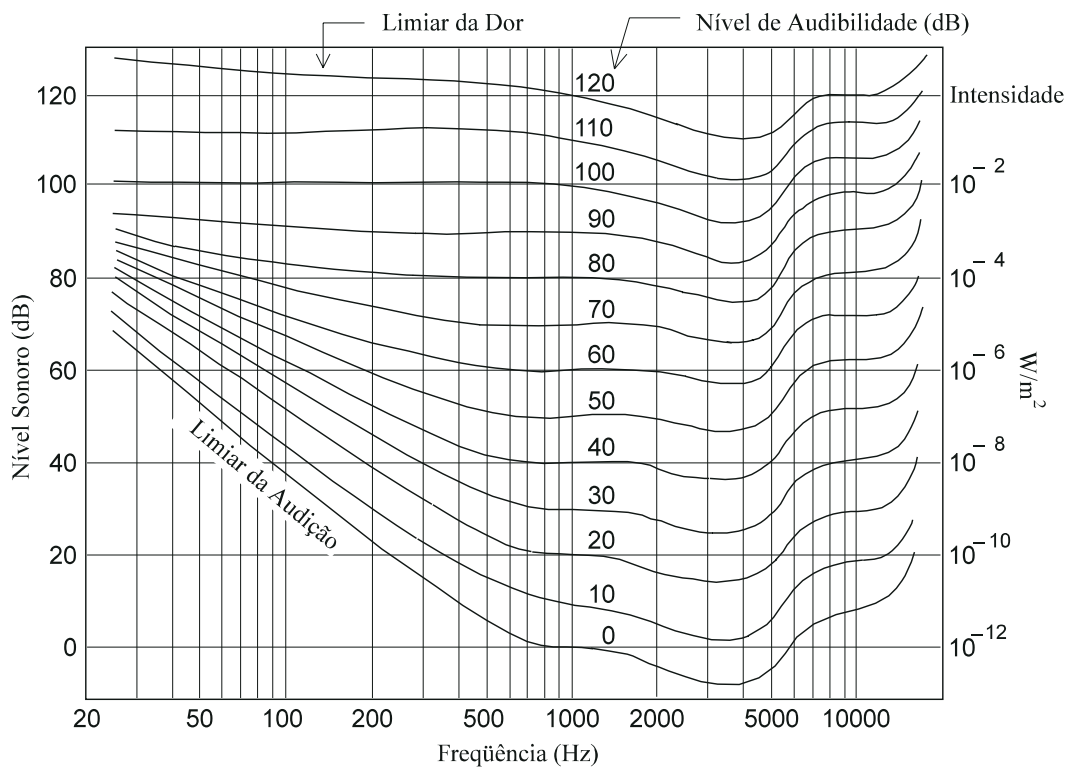


Figura 6.1: Diagrama de Fletcher para audibilidade.

Pode-se observar que a maior sensibilidade ocorre para frequências médias (com pico entre 3 e 4 kHz) e a variação da sensibilidade com a frequência se reduz

com o aumento da intensidade sonora.

6.2.2 Mascaramento

O **mascaramento** [34] é a propriedade de um som de maior intensidade (mascarador) inibir a percepção auditiva de outro (mascarado), que lhe seja “próximo”. Essa proximidade se refere tanto ao tempo quanto à frequência.

Um sinal pode mascarar outro que ocorra antes ou depois dele, embora no primeiro caso, anti-causal, o efeito seja bem menos significativo. O fenômeno se mede, em geral pela emissão de breves pulsos sonoros próximos no tempo (v. Figura 6.2).

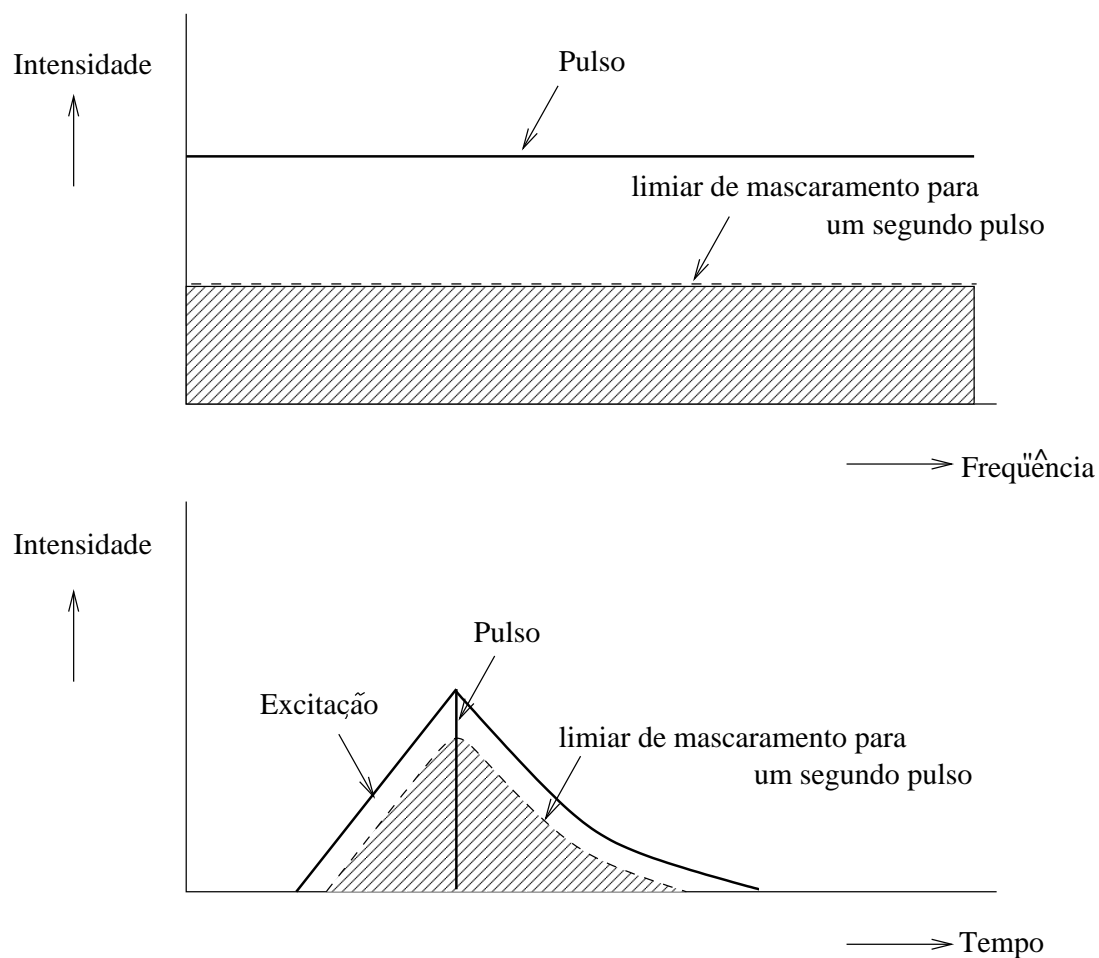


Figura 6.2: Mascaramento temporal.

Os primeiros estudos sobre mascaramento em frequência tratavam de tom mascarando tom, e indicaram que uma frequência pode mascarar frequências infe-

riores ou superiores a ela, neste último caso numa faixa pouco mais larga. A fim de evitar batimentos, o fenômeno se mede, em geral, com o tom mascarador substituído por ruído de banda estreita (v. Figura 6.3). Também há resultados relevantes sobre tom mascarando ruído de banda larga e vice-versa—como na obtenção da razão crítica (v. Subseção 6.2.3).

6.2.3 Bandas Críticas

Na definição formal, **banda crítica** [35] é a largura de banda capaz de contribuir para o mascaramento de um tom. Para suas primeiras medidas, gerava-se ruído branco de potência controlada, variando-a até ocorrer o mascaramento do tom de teste; a largura da banda de ruído que igualasse a potência do tom era a banda crítica em torno dele. Esta grandeza passou a ser chamada **razão crítica**, com a constatação de que a banda crítica não é, de fato, plana—e, conseqüentemente, é mais larga.

Bandas críticas podem ser obtidas para qualquer freqüência. Por comodidade, adotou-se um conjunto de bandas contíguas fixas, definindo-se com base nelas a unidade espectral Bark (v. Tabela 6.1).

6.2.4 Espalhamento

No caso real em que os sinais não podem ser, simplesmente, restritos a pulsos estreitos e tons senoidais, uma forma prática de representar o mascaramento é através do **espalhamento**, que descreve genericamente como a energia em um ponto do plano (tempo, freqüência) resulta na excitação em outro ponto. Os espalhamentos no tempo e na freqüência podem ser calculados separadamente. A Figura 6.4 mostra a representação composta do efeito devido a um sinal genérico.

6.2.5 Compressão

A **compressão** [45] é um mecanismo central ao processo auditivo humano e também responde pelo efeito final do mascaramento. Levou à definição por Zwicker da unidade Sone de densidade comprimida de excitação pela aplicação de uma operação não-linear adequada à densidade de excitação. A Figura 6.5 ilustra o mas-

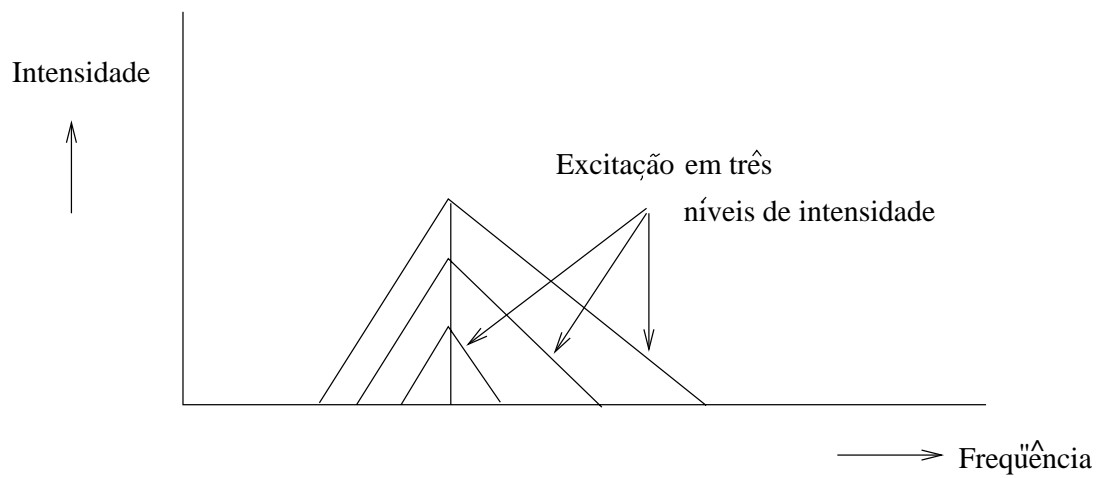
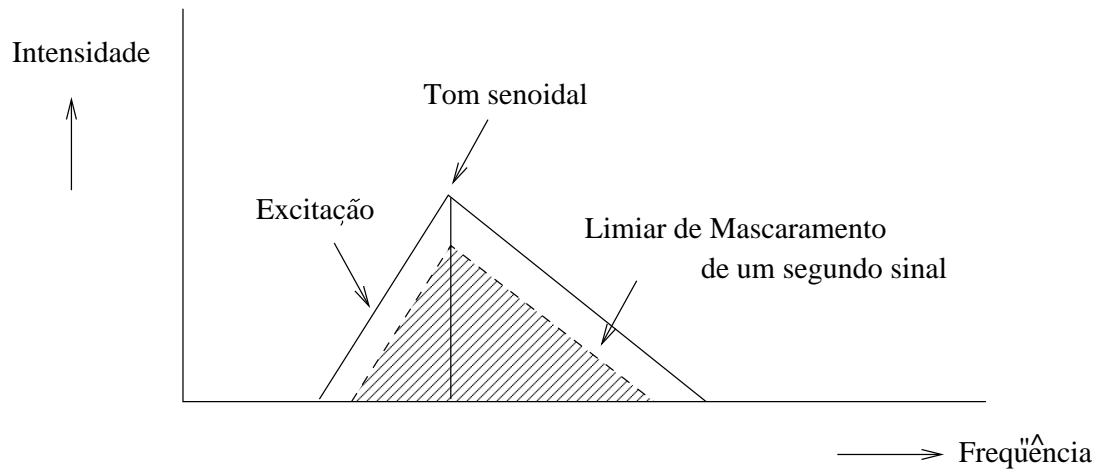
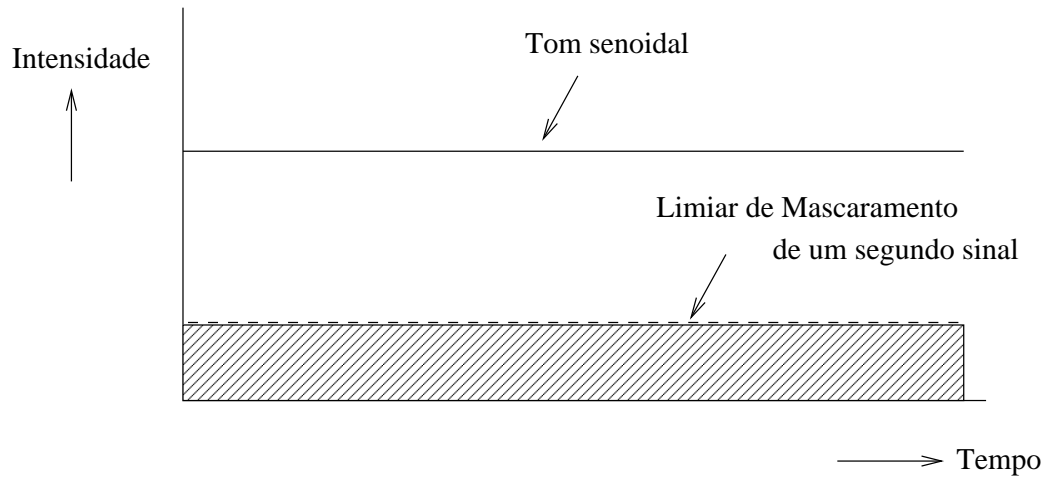


Figura 6.3: Mascaramento frequencial.

Tabela 6.1: Bandas críticas e índice Bark.

índice Bark	limite inferior (Hz)	freqüência central (Hz)	limite superior (Hz)
0	0	50	100
1	100	150	200
2	200	250	300
3	300	350	400
4	400	450	510
5	510	570	630
6	630	700	770
7	770	840	920
8	920	1000	1080
9	1080	1170	1270
10	1270	1370	1480
11	1480	1600	1720
12	1720	1850	2000
13	2000	2150	2320
14	2320	2500	2700
15	2700	2900	3150
16	3150	3400	3700
17	3700	4000	4400
18	4400	4800	5300
19	5300	5800	6400
20	6400	7000	7700
21	7700	8500	9500
22	9500	10500	12000
23	12000	13500	15500
24	15500	19500	24000

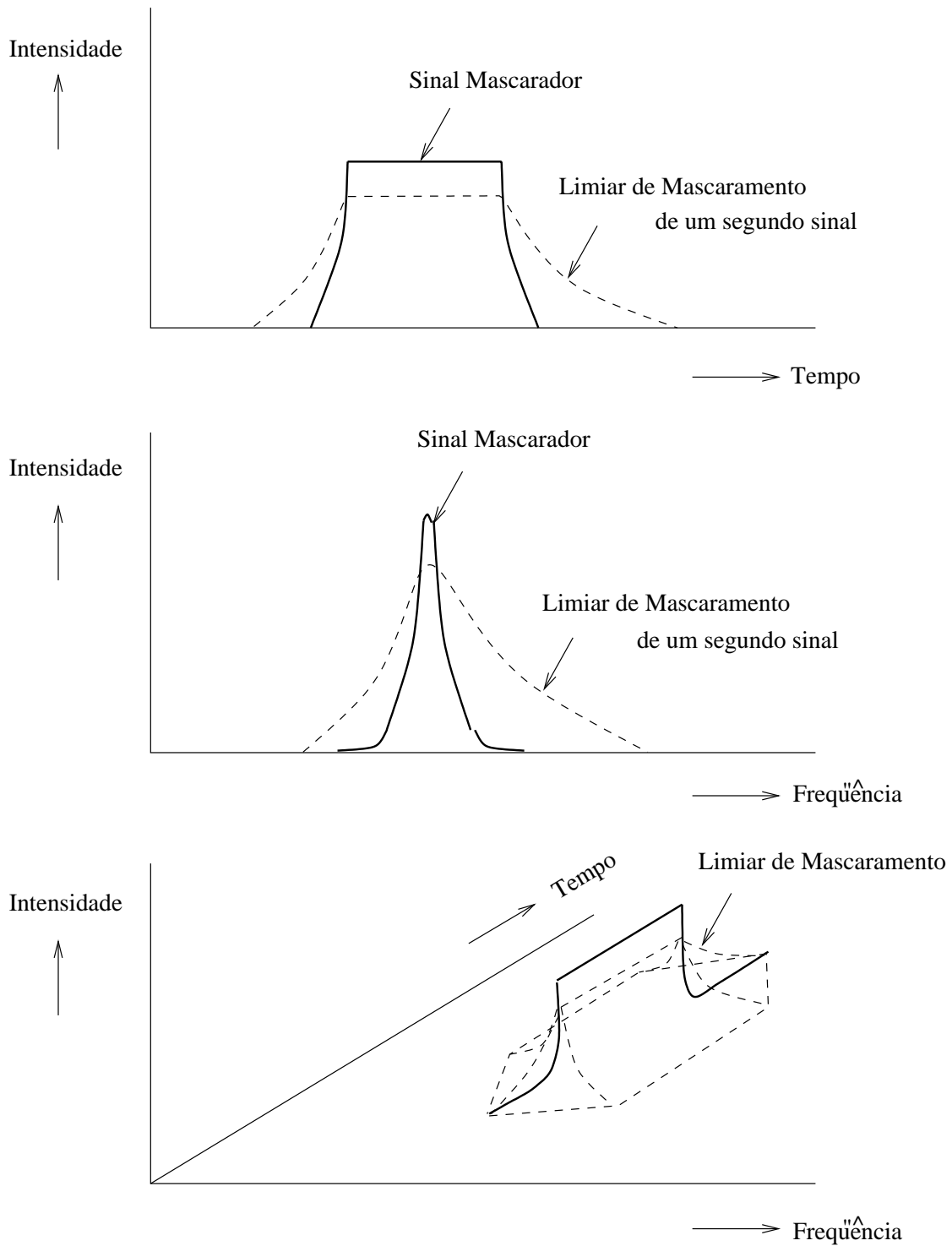


Figura 6.4: Mascaramento devido ao espalhamento da excitação.

caramento (freqüencial, a título de exemplo) como a composição do espalhamento e da compressão da excitação. Na figura, X representa o sinal senoidal mascarador e N, o sinal senoidal mascarado; P se refere ao espectro de potência, E à excitação e ζ à excitação comprimida. Os valores foram atribuídos apenas para dar uma noção quantitativa das operações. Na primeira linha, representam-se os espectros de potência do sinal composto e de cada um dos dois sinais individuais; na segunda linha, mostram-se as excitações correspondentes aos três casos; na terceira linha, vêm-se as excitações comprimidas. Por fim, o efeito do mascaramento é constatado comparando-se a diferença entre a excitação comprimida do sinal composto e a excitação comprimida do sinal mascarador com a excitação do sinal mascarado, esta muito maior que aquela. Interpretando: O que se percebe do sinal mascarado no sinal composto é muito menos do que dele se perceberia em separado.

6.3 PAQM

Com base na Psicoacústica, Beerends elaborou [42, 46] uma medida de qualidade perceptível de áudio (PAQM), sintetizada na Figura 6.6, que nada mais é do que uma medida da diferença perceptível entre dois sinais. Sendo estes um sinal de referência e uma versão sua modificada, a PAQM permite aferir a perda de qualidade sofrida por um sinal que foi processado (comprimido, por exemplo). A medida consiste simplesmente em obter uma representação quantitativa de “quanto se ouve” de cada um dos sinais ao longo da freqüência e do tempo e calcular a diferença média na forma de um número.

A seguir, faz-se uma breve (suficiente para os objetivos deste texto) descrição do cálculo da PAQM, que remete aos conceitos da Seção 6.2. A seqüência de processamento de cada um dos dois sinais discretizados é:

- Divisão em quadros (*frames*) com 50% de interseção; e em cada quadro:
- aplicação de janela suavizadora;
- cálculo do espectro de potência;
- conversão do espectro à escala bark;
- filtragem pela transferência do ouvido externo para o ouvido interno;

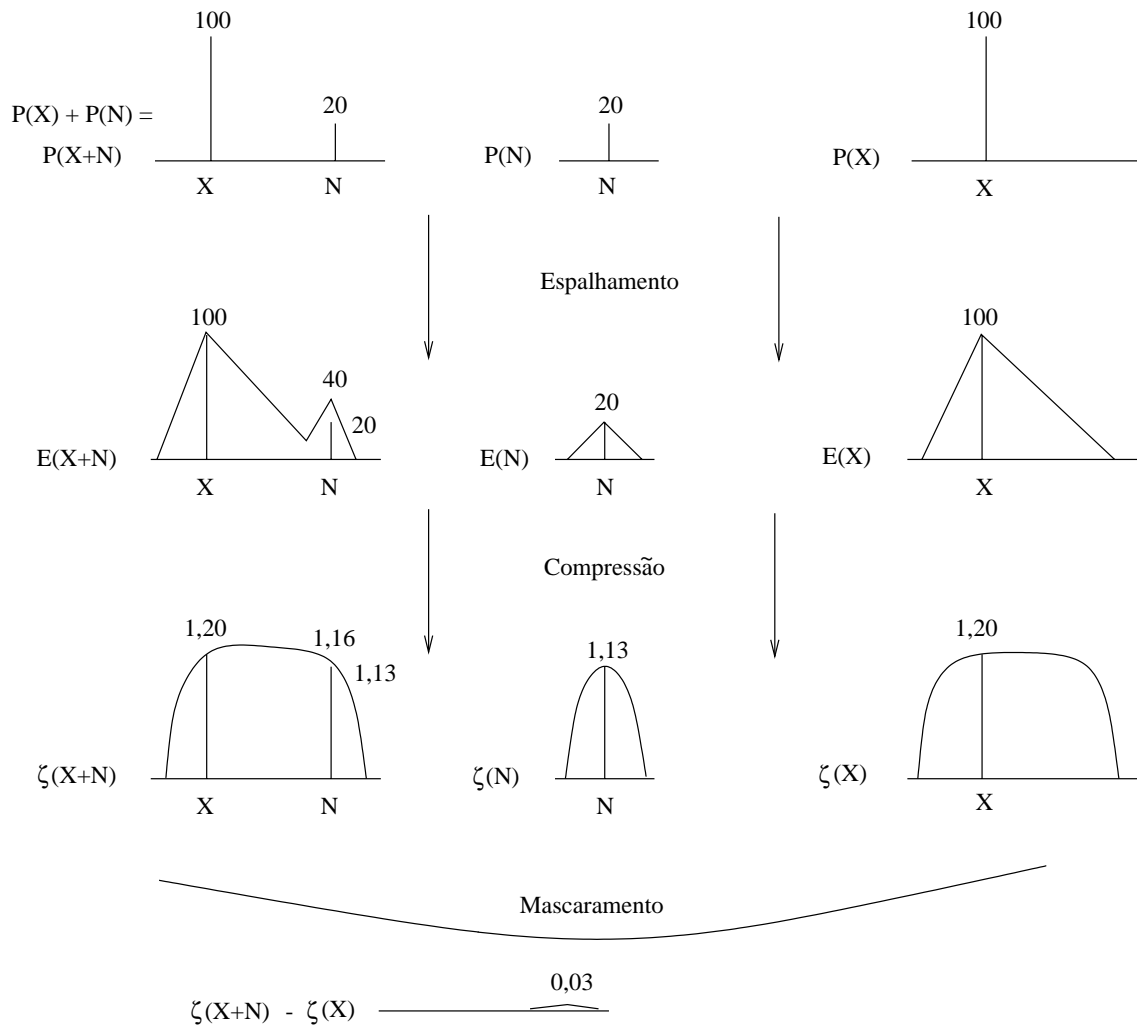


Figura 6.5: Esquema detalhando o mascaramento como resultado de espalhamento e compressão.

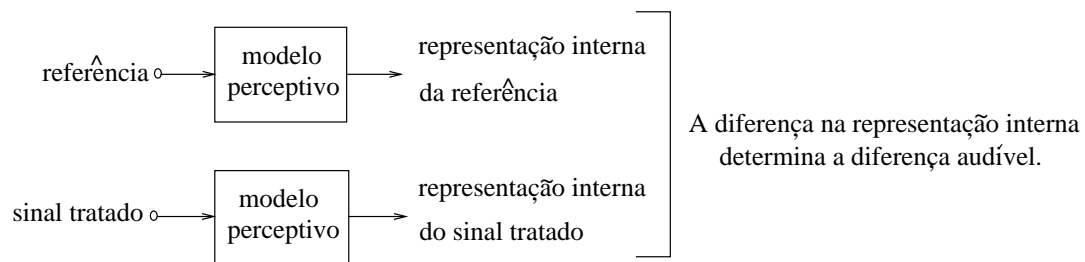


Figura 6.6: Esquema do PAQM.

- obtenção da curva de excitação pelo espalhamento do espectro Bark no tempo e na frequência;
- conversão à audibilidade comprimida em Sones.

Em seguida, a representação interna do sinal processado é subdividida e ponderada em três faixas Bark diferentes, de acordo com sua importância na representação. O módulo da diferença da representação interna da referência e da representação interna ponderada do sinal tratado é obtido para cada par (quadro, frequência Bark). As diferenças são acumuladas no valor $\eta \geq 0$ e a PAQM é definida como $10 \cdot \log(\eta)$. Dessa forma, sinais iguais resultam no valor $-\infty$, e maiores diferenças, em números progressivamente maiores.

6.4 Proposta: PAQM Adaptada à Aferição da Qualidade de Sinais Restaurados

A fim de comparar através de medida objetiva e confiável os resultados da restauração de sinais de áudio por algoritmos ou *set-ups* diferentes de um mesmo algoritmo, recorreu-se à PAQM.

Inicialmente, a utilização da PAQM neste trabalho imporá como requisito a existência de um padrão de qualidade. Nesse caso, os resultados do tratamento de um sinal pelos algoritmos de restauração serão avaliados contra a versão não-corrompida do sinal, que deverá estar disponível. Desde que se utilize um conjunto suficientemente representativo de sinais-padrão (com suas versões contaminadas) para permitir aferir o processamento, isso não é uma limitação grave.

O esquema adotado será, então, o mostrado na Figura 6.7. Como um ponto de partida, pode-se calcular a dessemelhança η_c entre o sinal padrão s_p e sua versão contaminada s_c anterior ao processamento. Cada uma das versões tratadas s_{t_i} de s_c é similarmente comparada com s_p , resultando em η_{t_i} . Qualquer tratamento eficaz deverá resultar na redução da dessemelhança; e quanto maior a redução, melhor o desempenho do tratamento realizado.

A fim de preservar o significado dimensional das medidas, de forma a poderem ser comparadas por diferença direta, eliminamos da medida a aplicação do logaritmo.

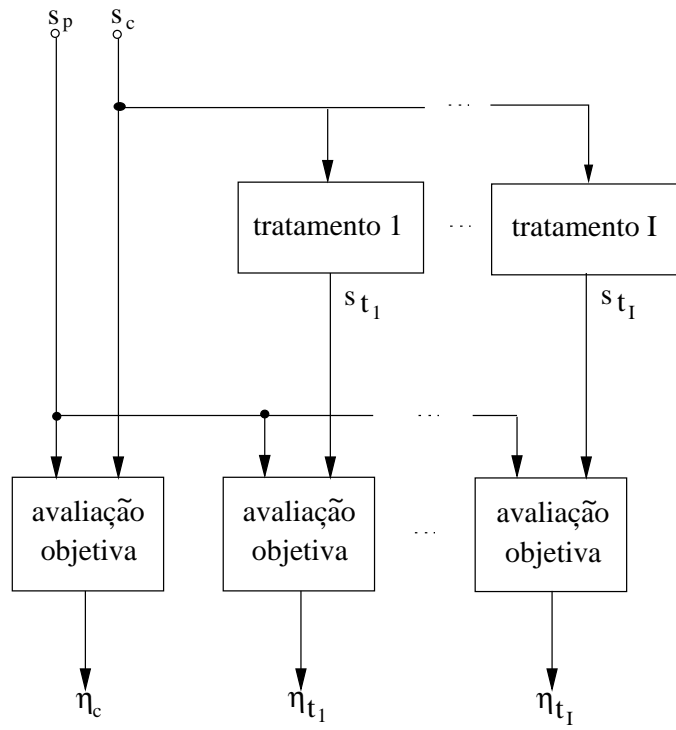


Figura 6.7: PAQM adaptada à avaliação da restauração de sinais.

Com isso, a restauração ideal resultaria no valor 0.

Este avaliador objetivo é utilizado nos Capítulos 8 e 10.

Capítulo 7

Interpolação de Amostras

Faltantes Baseada em Modelo

As técnicas para recomposição de amostras de um sinal que tenham sido, por algum motivo, perdidas encontram aplicações diversas. Em particular, na restauração de sinais corrompidos por falhas localizadas (como *clicks* de gravações em disco), quando as amostras corrompidas são dadas como inaproveitáveis, a sua substituição por amostras coerentes com o entorno é uma etapa inerente ao processo.

Neste Capítulo, após uma formulação geral do problema, são revistos alguns métodos mais usuais de efetuar a recomposição de amostras faltantes em restauração de áudio: LS, de cunho determinístico; e ML, EM e GS, de cunho estocástico. Em seguida, descreve-se a investigação do assunto realizada neste trabalho, que envolveu a comparação do desempenho dos métodos e levou à constatação de um problema na estimação do modelo de sinais de áudio quantizados (discutido em detalhe e solucionado no Capítulo 3).

7.1 Formulação do Problema

Nesta Seção, apresenta-se um modelamento matemático para o tratamento do problema [36, 47, 16].

Inicialmente, o sinal de áudio discretizado $x(k)$ será considerado como um processo aleatório discreto com média zero e estacionário [18, 19] por até K amostras.

Seja

$$\mathbf{x} = \left[x(0) \ x(1) \ \cdots \ x(I) \ \cdots \ x(k_f) \ \cdots \ x(k_f + F) \ \cdots \ x(K - 1) \right]^T \quad (7.1)$$

o vetor correspondente a um trecho de sinal com K amostras e seja

$$\mathbf{x}_1 = \left[x(I) \ x(I + 1) \ \cdots \ x(K - 1) \right]^T. \quad (7.2)$$

Adotando-se para o sinal um modelo autorregressivo (AR) de ordem I , pode-se fazer

$$x(k) = \sum_{i=1}^I a(i)x(k - i) + e(k), \quad (7.3)$$

com $k \geq I$, sendo $a(i)$ os coeficientes do modelo e $e(k)$ o sinal de excitação, ruído branco com média zero e variância σ_e^2 . Na forma matricial,

$$\mathbf{e} = \mathbf{x}_1 - \mathbf{X}\mathbf{a}, \quad (7.4)$$

onde

$$\mathbf{e} = \left[e(I) \ e(I + 1) \ \cdots \ e(K - 1) \right]^T, \quad (7.5)$$

$$\mathbf{X} = \begin{bmatrix} x(I - 1) & x(I - 2) & \cdots & x(0) \\ x(I) & x(I - 1) & \cdots & x(1) \\ \vdots & \vdots & \ddots & \vdots \\ x(K - 2) & x(K - 3) & \cdots & x(K - I - 1) \end{bmatrix} \quad (7.6)$$

e

$$\mathbf{a} = \left[a(1) \ a(2) \ \cdots \ a(I) \right]^T. \quad (7.7)$$

Com isso, pode-se definir a energia de excitação

$$\varepsilon \triangleq \mathbf{e}^T \mathbf{e} = \mathbf{x}_1^T \mathbf{x}_1 - 2\mathbf{x}_1^T \mathbf{X}\mathbf{a} + \mathbf{a}^T \mathbf{X}^T \mathbf{X}\mathbf{a} \quad (7.8)$$

em função de \mathbf{a} .

Chamando, agora,

$$\mathbf{x}_d = \left[x(k_f) \ x(k_f + 1) \ \cdots \ x(k_f + F - 1) \right]^T \quad (7.9)$$

à falha de comprimento F após as primeiras k_f amostras (o subscrito d corresponde a “desconhecido”) e

$$\mathbf{x}_c = \left[x(0) \ x(1) \ \cdots \ x(k_f - 1) \ x(k_f + F) \ \cdots \ x(K - 1) \right]^T \quad (7.10)$$

ao que resta de \mathbf{x} retirado \mathbf{x}_d (o subscrito c corresponde a “conhecido”) e reescrevendo a equação (7.4) como

$$\mathbf{e} = \mathbf{A}\mathbf{x}, \quad (7.11)$$

onde a matriz $(K - I) \times K$

$$\mathbf{A} = \begin{bmatrix} -a(I) & \cdots & -a(1) & 1 & 0 & \cdots & 0 \\ 0 & -a(I) & \cdots & -a(1) & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \cdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -a(I) & \cdots & -a(1) & 1 \end{bmatrix}, \quad (7.12)$$

pode-se redefinir a energia de excitação

$$\varepsilon \triangleq \mathbf{e}^T \mathbf{e} = \mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{x}_c^T \mathbf{A}_c^T \mathbf{A}_c \mathbf{x}_c + 2\mathbf{x}_c^T \mathbf{A}_c^T \mathbf{A}_d \mathbf{x}_d + \mathbf{x}_d^T \mathbf{A}_d^T \mathbf{A}_d \mathbf{x}_d \quad (7.13)$$

em função de \mathbf{x}_d , sendo \mathbf{A}_d a submatriz formada pelas colunas de $k_f + 1$ até $k_F + 1$ de \mathbf{A} e sendo \mathbf{A}_c o que resta de \mathbf{A} retirada \mathbf{A}_d , tais que

$$\mathbf{A}\mathbf{x} = \mathbf{A}_d \mathbf{x}_d + \mathbf{A}_c \mathbf{x}_c. \quad (7.14)$$

7.2 Técnica de Base Determinística: LS

É possível reinterpretar o modelo para o sinal, descrito pela equação (7.3), como um preditor de ordem I com erro de predição $e(k)$.

Sob essa visão, as amostras utilizadas na interpolação podem ser obtidas pela minimização da energia do erro

$$\varepsilon = \sum_{k=I}^K e^2(k), \quad (7.15)$$

descrita na forma da equação (7.13). A solução deste problema de mínimos quadrados (LS) é [16]

$$\mathbf{x}_d^{\text{LS}} = (\mathbf{A}_d^T \mathbf{A}_d)^{-1} \mathbf{A}_d^T \mathbf{A}_c \mathbf{x}_c. \quad (7.16)$$

7.3 Estimativa de ML

A verossimilhança de \mathbf{x}_1 com modelo $\{\mathbf{a}, \sigma_e\}$ é definida por¹

$$L(\mathbf{x}_1; \mathbf{a}, \sigma_e) \triangleq f_{\mathbf{x}_1|\mathbf{a}, \sigma_e}(\mathbf{x}_1 | \mathbf{a}, \sigma_e). \quad (7.17)$$

Um modo de interpolar no sinal uma estimativa para suas amostras faltantes é utilizar o sinal mais provável, conhecido seu modelo. Este resulta da maximização de $L(\mathbf{x}_1; \mathbf{a}, \sigma_e)$. Se $e(k)$ for Gaussiano, tal solução de máxima verossimilhança (ML) será dada por [48]

$$\mathbf{x}_d^{\text{ML}} = (\mathbf{A}_d^T \mathbf{A}_d)^{-1} \mathbf{A}_d^T \mathbf{A}_c \mathbf{x}_c. \quad (7.18)$$

Nessas condições, então, $\mathbf{x}_d^{\text{ML}} = \mathbf{x}_d^{\text{LS}}$ (v. Seção 7.2), o que implica estar sendo minimizada a energia de excitação.

7.4 Estimação por EM

Outro modo de interpolar no sinal uma estimativa para suas amostras faltantes pode ser utilizar as amostras desconhecidas mais prováveis do sinal, conhecidas as outras. Esta pode ser obtida pela maximização da densidade preditiva de \mathbf{x}_d , definida por $f_{\mathbf{x}_d|\mathbf{x}_c}(\mathbf{x}_d | \mathbf{x}_c)$. Por ser formulada como a alternância de dois estágios, um de Expectância e outro de Maximização, a solução para este problema é dita por EM [48]. Se $e(k)$ for Gaussiano e $f_{\mathbf{a}}(\mathbf{a})$ for suposta uniforme por ser desconhecida na ausência de dados, tal solução será

$$\mathbf{x}_d^{\text{EM}} = (\sigma_e^2 \mathbf{T} + \mathbf{A}_d^T \mathbf{A}_d)^{-1} (\sigma_e^2 \mathbf{q} + \mathbf{A}_d^T \mathbf{A}_c \mathbf{x}_c), \quad (7.19)$$

sendo a matriz \mathbf{T} e o vetor \mathbf{q} formados por manipulações dos elementos de \mathbf{x} , parte dos quais está sendo estimada—o que confere caráter iterativo ao algoritmo.

Nessas condições, é possível prever [48] que a solução por EM tenderá à solução de ML, com a conseqüente minimização da energia de excitação.

¹Ao longo deste trabalho, nota-se por $F_v(v)$ a distribuição cumulativa e por $f_v(v)$ a densidade de probabilidade da variável aleatória v [18, 19].

7.5 A Necessidade do Vetor-Parâmetro

Os três estimadores para \mathbf{x}_d vistos (LS, ML e EM) pressupõem o conhecimento de \mathbf{a} . Uma solução prática é obter sua estimativa LS a partir da minimização de ε descrita na forma da equação (7.8), que resulta em

$$\mathbf{a}^{\text{LS}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{x}_1. \quad (7.20)$$

Novamente, tal solução se confunde com a de ML no caso Gaussiano.

Um algoritmo completo pode alternar os cálculos das estimativas de \mathbf{a} e \mathbf{x}_d até que se atinja uma condição de convergência.

7.6 Estimativa pelo Amostrador de Gibbs

Outro método, mais elaborado, de interpolar no sinal uma estimativa para suas amostras faltantes é utilizar amostras geradas de uma estimativa da distribuição conjunta $f_{\mathbf{x}_d, \mathbf{a}, \sigma_e | \mathbf{x}_c}(\mathbf{x}_d, \mathbf{a}, \sigma_e | \mathbf{x}_c)$, a qual descreve integralmente o relacionamento entre as variáveis do problema. Se $e(k)$ for Gaussiano e $f_{\mathbf{a}, \sigma_e}(\mathbf{a}, \sigma_e)$ for suposta uniforme por ser desconhecida na ausência de dados, as distribuições condicionais de \mathbf{x}_d , \mathbf{a} e σ_e serão analiticamente tratáveis.

Então, o algoritmo do Amostrador de Gibbs (GS) [49, 50, 48] consistirá em, partindo de $(\mathbf{x}_d^0, \mathbf{a}^0, \sigma_e^0)$, gerar alternadamente amostras das distribuições condicionais abaixo, por um determinado número de vezes (cuja determinação não é trivial):

- \mathbf{x}_d^{GS} de $f_{\mathbf{x}_d | \mathbf{a}, \sigma_e, \mathbf{x}_c}(\mathbf{x}_d | \mathbf{a}^{\text{GS}}, \sigma_e^{\text{GS}}, \mathbf{x}_c)$;
- \mathbf{a}^{GS} de $f_{\mathbf{a} | \sigma_e, \mathbf{x}_d, \mathbf{x}_c}(\mathbf{a} | \sigma_e^{\text{GS}}, \mathbf{x}_d^{\text{GS}}, \mathbf{x}_c)$;
- σ_e^{GS} de $f_{\sigma_e | \mathbf{x}_d, \mathbf{a}, \mathbf{x}_c}(\sigma_e | \mathbf{x}_d^{\text{GS}}, \mathbf{a}^{\text{GS}}, \mathbf{x}_c)$.

A distribuição dessas amostras tende assintoticamente a $f_{\mathbf{x}_d, \mathbf{a}, \sigma_e | \mathbf{x}_c}(\mathbf{x}_d, \mathbf{a}, \sigma_e | \mathbf{x}_c)$.

O que distingue esse método dos anteriores é o fato de considerar a energia de excitação como aleatória, realizando sua estimação em lugar de sua minimização explícita (como no LS) ou implícita (como no ML e no EM).

7.7 Da Implementação

A implementação das técnicas acima tem alto custo de memória para armazenamento dos dados envolvidos no processamento e computacional para a realização das operações matemáticas, particularmente a inversão das matrizes $\mathbf{X}^T\mathbf{X}$ ($I \times I$) e $\mathbf{A}_d^T\mathbf{A}_d$ ($F \times F$).

Entretanto, os dados exibem particularidades estruturais que permitem o emprego de técnicas especializadas [51] capazes de elevar a eficiência de armazenamento e cálculo:

- A matriz \mathbf{A} é esparsa.
- A matriz $\mathbf{X}^T\mathbf{X}$ é simétrica.
- A matriz $\mathbf{A}_d^T\mathbf{A}_d$ é simétrica, podendo ser tornada Toeplitz por reformulação do problema.

Vale notar, ainda, que a formulação pressupõe que se disponha de p amostras válidas antes da falha.

7.8 Investigação Realizada

7.8.1 Testes e Resultados

Implementaram-se testes [52] sobre sinais artificiais e sinais naturais, sempre consistindo de um bloco de K amostras do qual se descartavam F amostras. Efetuaram-se, entre outros, os testes de [48, 50]. A interpolação foi realizada com os algoritmos ML (ou LS) e EM incluindo a estimação LS do vetor \mathbf{a} e com o algoritmo GS, que também a inclui na inicialização.

Os sinais artificiais, de alta energia frente ao maior valor representável em 16 bits com ponto fixo, foram:

- Processo AR de ordem baixa com pólos afastados da circunferência unitária;
- processo AR de ordem baixa com pólos próximos à circunferência unitária;
- uma senóide;

- uma senóide com ruído branco aditivo;
- soma de duas senóides de frequências diferentes.

Os sinais naturais, também representados em 16 bits com ponto fixo, foram:

- Sinal musical com variada instrumentação e alta energia;
- sinal musical com conteúdo predominantemente vocal e média energia;
- sinal musical com instrumentação simples e baixa energia com falha extensa (da ordem da estacionaridade assumida para o sinal).

O ML e o EM apresentaram resultados virtualmente idênticos, não compensando a maior sofisticação do EM.

A maior autocorrelação dos sinais, associada ao filtro gerador do processo autorregressivo $x(k)$ a partir da excitação $e(k)$ com pólos próximos à circunferência unitária, torna mais fácil a restauração.

Excetuando-se os casos da senóide (sinal determinístico, e portanto maximamente correlacionado) e do sinal vocal (similarmente constituído), o ML e o EM apresentaram resultados inaceitáveis, tendendo a reconstituir o trecho da falha somente com informação correlacionada; por exemplo, a falha na senóide com ruído era preenchida com uma senóide pura. Isso se deveu a ambos tenderem a minimizar a energia de excitação, que representa justamente a parcela não-preditível do sinal. Em contrapartida, o GS tende a interpolar um sinal com características comuns às do entorno presumidamente não corrompido. A Figura 7.1 compara os três algoritmos para um dos sinais de teste.

O GS mostrou-se, assim, sempre superior, só não tendo sido eficaz (como também o ML e o EM) em dois casos: o da falha extensa, que viola as hipóteses feitas na Seção 7.1 (como a de estacionaridade curta); e o das duas senóides, que será examinado separadamente a seguir.

7.8.2 Interpolação de Sinal Quantizado

O exemplo em que todos os algoritmos falharam nas simulações realizadas consistia em interpolar as amostras de índices $375 \leq k < 775$, retiradas de um bloco

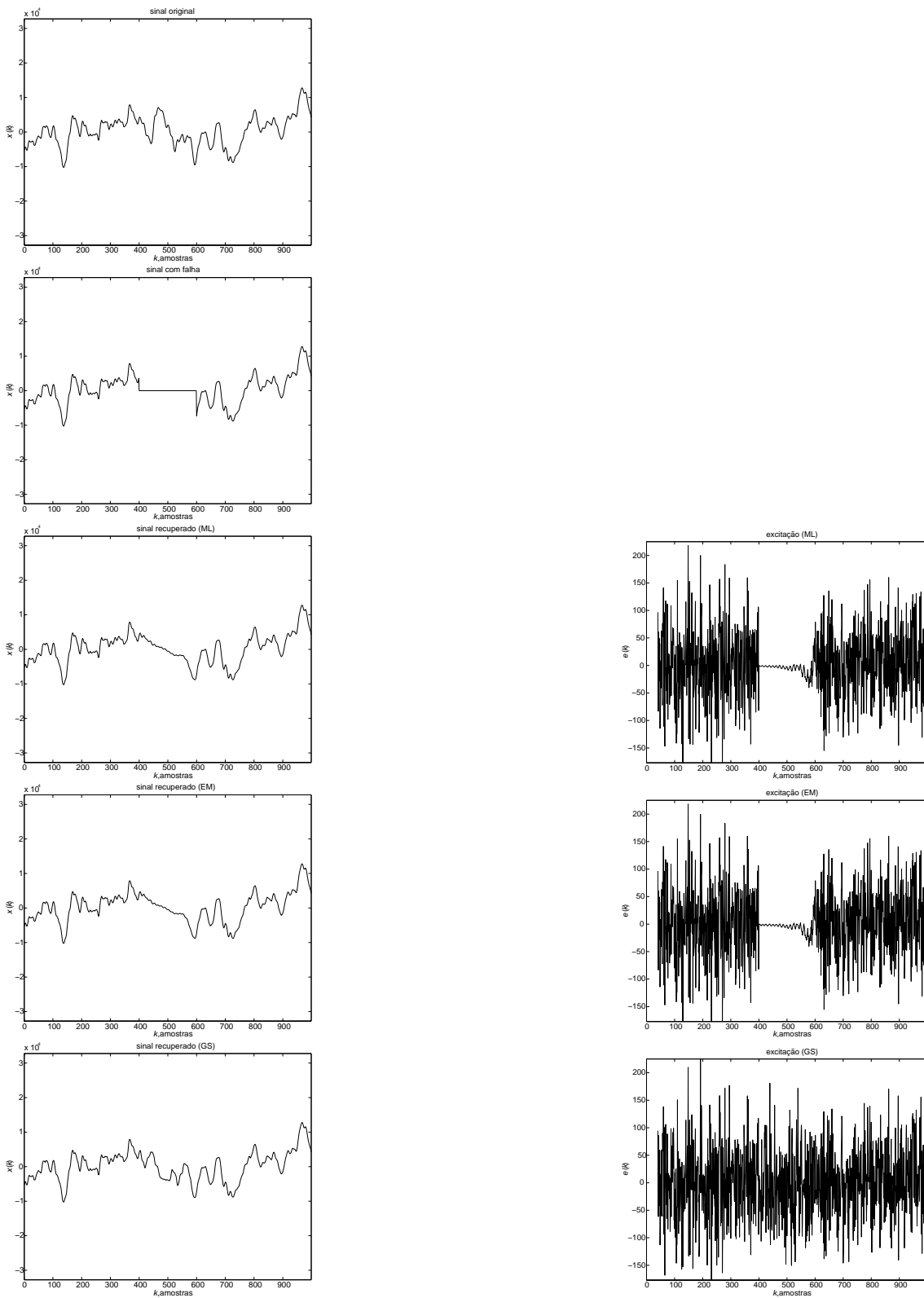


Figura 7.1: Sinal musical com variada instrumentação e alta energia.

formado pelas amostras de índices $0 \leq k < 1000$ do sinal definido por

$$x(k) = A \left[\text{sen} \left(\frac{\pi k}{100} \right) + \text{sen} \left(\frac{\pi k}{50} \right) \right].$$

A Figura 7.2 mostra os resultados encontrados em nossas simulações, em que consistentemente a senóide de frequência inferior parece não ter sido interpolada.

É importante destacar que tais resultados contradiziam aqueles apresentados em [48] para simulações idênticas, que indicam perfeito desempenho dos três algoritmos!

Inúmeros testes foram realizados na tentativa de definir os contornos do problema. A variação das frequências mostrou que a anomalia ocorria sempre que ambas eram muito baixas relativamente à taxa de amostragem e independia (salvo exageros) do número de amostras a interpolar e do tamanho do bloco.

Ocorre que aqui, assim como os sinais reais, os sinais artificiais estavam sendo quantizados em 16 bits, embora os cálculos fossem efetuados com precisão elevada em ponto flutuante. Verificou-se que aumentando o número de bits da representação do sinal a modelar, a anomalia era gradualmente eliminada. Então, o problema se originava na interpolação e, provavelmente, em [48] não se efetuara a quantização dos sinais, todos artificiais.

Todos os 3 algoritmos de interpolação testados têm em comum uma etapa de estimação de um modelo AR a partir das amostras válidas do sinal. Uma vez fora constatado que o tamanho do bloco e do segmento a interpolar, fatores cruciais para a interpolação propriamente dita, não afetavam o problema, deduziu-se—corretamente—que só o cálculo do modelo já respondia pelo insucesso dos algoritmos. A análise do efeito da quantização sobre a estimativa do modelo AR e a proposição de uma solução para o problema são o tema do Capítulo 3. A correção do modelo lá proposta permite reconstruir o sinal de forma que o torna visualmente indistinguível do sinal original da Figura 7.2, razão pela qual não foi incluída aqui a ilustração correspondente.

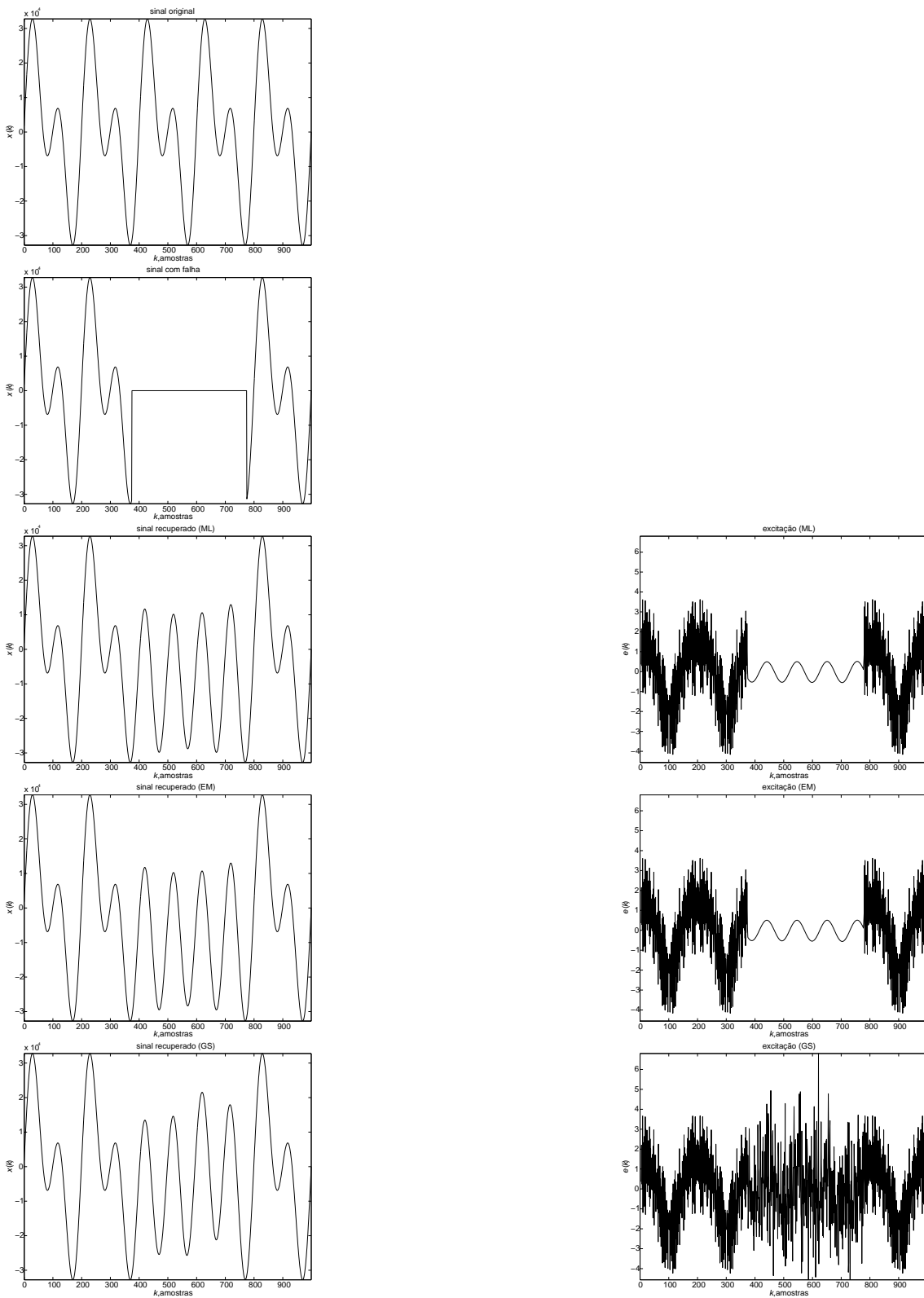


Figura 7.2: Soma de duas senóides com freqüências diferentes.

Capítulo 8

Tratamento de Ruído Impulsivo

8.1 Caracterização

Um problema muito comum em gravações analógicas é a degradação do sinal de áudio por pulsos de curta duração (menor que 1 ms) com subida e descida relativamente rápidas, dispersos aleatoriamente ao longo do tempo¹. Chamados imprecisamente de ruído impulsivo, esses distúrbios apresentam grandes variações na duração e na amplitude.

O ruído impulsivo pode ser resultante de pequenas imperfeições, rugosidades ou deposições de resíduos na superfície de cilindros ou discos. A sensação auditiva por ele causada é a de uma seqüência de breves estalidos que, se muito freqüentes, podem dar uma impressão contínua.

O percentual de contaminação por ruído impulsivo normalmente não ultrapassa 10% das amostras do sinal discretizado, valor que já corresponde a uma degradação muito severa se todos os distúrbios forem perceptíveis. Sendo um tipo de falha localizada, este tipo de contaminação permite classificar razoavelmente as amostras do sinal em corrompidas e não-corrompidas. Em conseqüência, as técnicas de processamento digital para supressão de ruído impulsivo se compõem, normalmente, de uma etapa de Detecção dos trechos corrompidos seguida de outra onde se efetua a Reconstrução do sinal.

Naturalmente, um processamento eficaz deve atender ao compromisso entre

¹Não se incluem aqui os pulsos de maior duração com conteúdo importante de baixa freqüência originados, por exemplo, por arranhões ou quebra em disco ou cilindro (tema do Capítulo 9).

corrigir o máximo de amostras corrompidas (ao custo de “corrigir” amostras não-corrompidas) e evitar ao máximo “corrigir” amostras não-corrompidas (ao custo de deixar de corrigir amostras corrompidas). Por isso, a localização precisa dos pulsos de ruído é fundamental à restauração.

O tratamento de ruído impulsivo inspirou soluções variadas (de diversos graus de eficiência e generalidade) que vão desde a detecção de distúrbios após a filtragem passa-altas do sinal contaminado seguida da reconstrução por interpolação dos coeficientes da DWT que representam os trechos detectados [53] até o uso de redes neurais [54].

A seguir serão, inicialmente, revistas três técnicas para eliminação de ruído impulsivo em áudio gravado: a filtragem por mediana, extremamente simples e de desempenho limitado na mesma medida; a detecção por inversão do modelo autorregressivo + interpolação, que atinge um bom compromisso entre desempenho e custo computacional e em cujo princípio se baseiam produtos comerciais; e a técnica Bayesiana, mais robusta mas de complexidade bastante elevada. Reconhecendo a detecção por inversão do modelo AR como o ponto mais merecedor de atenção e sempre com base em situações reais, investigam-se detalhes de sua implementação. Caracterizados seus principais problemas, apresenta-se uma série de alternativas ao algoritmo original (pesquisa em colaboração [55]), que em sua forma final resultaram em desempenho indiscutivelmente superior sem aumento significativo de complexidade. Finalmente, aborda-se o processamento em sub-bandas.

8.2 Restauração sem Modelo: Filtragem por Mediana

Dada uma variável aleatória x com função cumulativa de probabilidade $F_x(x)$, a Mediana [18] de x é uma estatística não-linear dada pelo menor valor de x tal que $F_x(x) \geq 0,5$. Feitas as devidas considerações, pode-se tomar como a mediana de uma seqüência de números o valor da amostra central de sua versão ordenada crescentemente, uma medida essencialmente insensível a distúrbios impulsivos.

Assim, a Filtragem por Mediana [36] é uma abordagem clássica para remoção de distúrbios impulsivos de seqüências temporais de variação relativamente lenta, e

consiste em substituir o valor da amostra central de uma janela (preferencialmente de tamanho ímpar) que desliza ponto a ponto pela seqüência pelo valor de sua mediana. Sua principal vantagem é não requerer prévio conhecimento sobre os mecanismos de geração do sinal, não lhe atribuindo qualquer modelo estatístico. Seu desempenho pode ser melhorado pela utilização de um limiar adaptativo proporcional ao módulo da mediana da janela, de forma que a substituição da amostra central pela mediana só ocorra quando a diferença entre elas exceder o limiar. A escolha da janela é um fator crítico: ela precisa ser curta o suficiente para não suavizar indevidamente o sinal e muito mais longa que os ditos “impulsos”. Na prática, o compromisso é inconciliável.

8.3 Restauração com Modelagem do Sinal

8.3.1 Notação Adotada

Como em [16], o sinal ruidoso será considerado corrompido aditivamente, como em

$$y(k) = x(k) + c(k)n(k), \quad (8.1)$$

onde $x(k)$ é o sinal não-corrompido, $c(k)$ é um processo binário tal que $c(k) = 0$ quando a amostra $y(k)$ está intacta e $c(k) = 1$ quando a amostra $y(k)$ está corrompida. Então, adotando-se os índices d para “desconhecido” e c para “conhecido”, pode-se convencionar

$$y_d(k) = x_d(k) + n_d(k) \quad (8.2)$$

e

$$y_c(k) = x_c(k). \quad (8.3)$$

Para blocos de K amostras, o sinal $x(k)$ será considerado estacionário e modelável como um processo AR de ordem I tal que

$$x(k) = \sum_{i=1}^I a(i)x(k-i) + e_x(k), \quad (8.4)$$

com $k \geq I$, sendo $a(i)$ os coeficientes do modelo e $e_x(k)$ o sinal de excitação, ruído branco com média zero e variância $\sigma_{e_x}^2$.

Definam-se, ainda,

$$\mathbf{x} = \begin{bmatrix} x(0) & x(1) & \cdots & x(K-1) \end{bmatrix}^T, \quad (8.5)$$

$$\mathbf{x}_1 = \begin{bmatrix} x(I) & x(I+1) & \cdots & x(K-1) \end{bmatrix}^T \quad (8.6)$$

e

$$\mathbf{A} = \begin{bmatrix} -a(I) & \cdots & -a(1) & 1 & 0 & \cdots & 0 \\ 0 & -a(I) & \cdots & -a(1) & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \cdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -a(I) & \cdots & -a(1) & 1 \end{bmatrix}, \quad (8.7)$$

convenientemente particionada para que

$$\mathbf{A}\mathbf{x} = \mathbf{A}_d\mathbf{x}_d + \mathbf{A}_c\mathbf{x}_c. \quad (8.8)$$

8.3.2 Detecção por Filtragem Inversa do Modelo AR + Interpolação

Como já foi dito, uma das formas de tratar o ruído impulsivo é operar em duas etapas: a Detecção das amostras corrompidas e a subsequente Reconstrução do sinal segundo algum procedimento. Aqui, a reconstrução consiste na substituição das amostras dos distúrbios por amostras “corrigidas” através da interpolação do sinal, que é o tema do Capítulo 7. Será discutida, então, apenas a detecção das falhas [36, 56, 16].

A primeira idéia que vem à mente quando se pensa na detecção de pulsos imersos no sinal de áudio é a comparação direta da amplitude do sinal ruidoso com um limiar, na expectativa de que os pulsos ultrapassem o limiar, mas o sinal, não. Principalmente no caso dos pequenos pulsos que compõem o ruído impulsivo, não é isso que ocorre: na maior parte do tempo, os pulsos não se destacam no sinal (que, ademais, pode sofrer grandes variações de potência ao longo do tempo). A detecção pela inversão do modelo AR tem como fim reduzir a razão sinal-ruído para facilitar a identificação dos pulsos de ruído em meio ao sinal.

O processo AR $x(k)$ pode ser revertido ao sinal de excitação $e(k)$ que o originou pelo filtro inverso FIR

$$A(z) = 1 - \sum_{i=1}^I a(i)z^{-i}. \quad (8.9)$$

Se o sinal ruidoso $y(k)$ fosse filtrado por $A(z)$, resultaria em

$$e(k) = e_x(k) + c(k)n(k) - \sum_{i=1}^I a(i)c(k-i)n(k-i). \quad (8.10)$$

Nota-se que cada distúrbio tende a ser espalhado pela resposta do filtro inverso. Agora, considerando a razão sinal-ruído SNR como a razão entre a potência originada no sinal e a potência originada no ruído, é possível calcular [52] a SNR no domínio do sinal

$$\text{SNR}_{\text{sin}} = \frac{E[x^2(k)]}{E[c^2(k)n^2(k)]} = \frac{\sigma_{e_x}^2 + E\left[\left(\sum_{i=1}^I a(i)x(k-i)\right)^2\right]}{E[c^2(k)n^2(k)]} \quad (8.11)$$

e a SNR no domínio da excitação

$$\text{SNR}_{\text{exc}} = \frac{\sigma_{e_x}^2}{E\left[\left(c(k)n(k) - \sum_{i=1}^I a(i)c(k-i)n(k-i)\right)^2\right]}. \quad (8.12)$$

Com a hipótese (certamente discutível, no caso real) de não-correlação entre as amostras de $c(k)n(k)$,

$$\text{SNR}_{\text{exc}} = \frac{\sigma_{e_x}^2}{\left(1 + \sum_{i=1}^I a^2(i)\right) E[c^2(k)n^2(k)]}. \quad (8.13)$$

Com isso, a razão entre a SNR no domínio da excitação e a SNR no domínio do sinal resulta em

$$\frac{\text{SNR}_{\text{exc}}}{\text{SNR}_{\text{sin}}} = \frac{1}{\left(1 + \frac{E\left[\left(\sum_{i=1}^I a(i)x(k-i)\right)^2\right]}{\sigma_{e_x}^2}\right) \left(1 + \sum_{i=1}^I a^2(i)\right)} < 1, \quad (8.14)$$

indicando a redução da razão sinal-ruído com a passagem do sinal para a excitação. Isso se explica pela eliminação da parcela preditível do áudio (predominante na sua potência), com a filtragem inversa. Para sinais reais, uma redução típica é da ordem de 100. Na situação real, não se dispõe exatamente de $a(i)$, que podem ser estimados por um método como o LS (v. Seção 3.2.2).

Em [36], propõe-se aumentar ainda mais a detectabilidade dos distúrbios por um procedimento baseado em filtragem casada. Entretanto, esta resulta num espalhamento excessivo dos pulsos, prejudicando sua precisa localização, o que a torna inadequada. A Fig. 8.1 compara os sinais envolvidos num exemplo real.

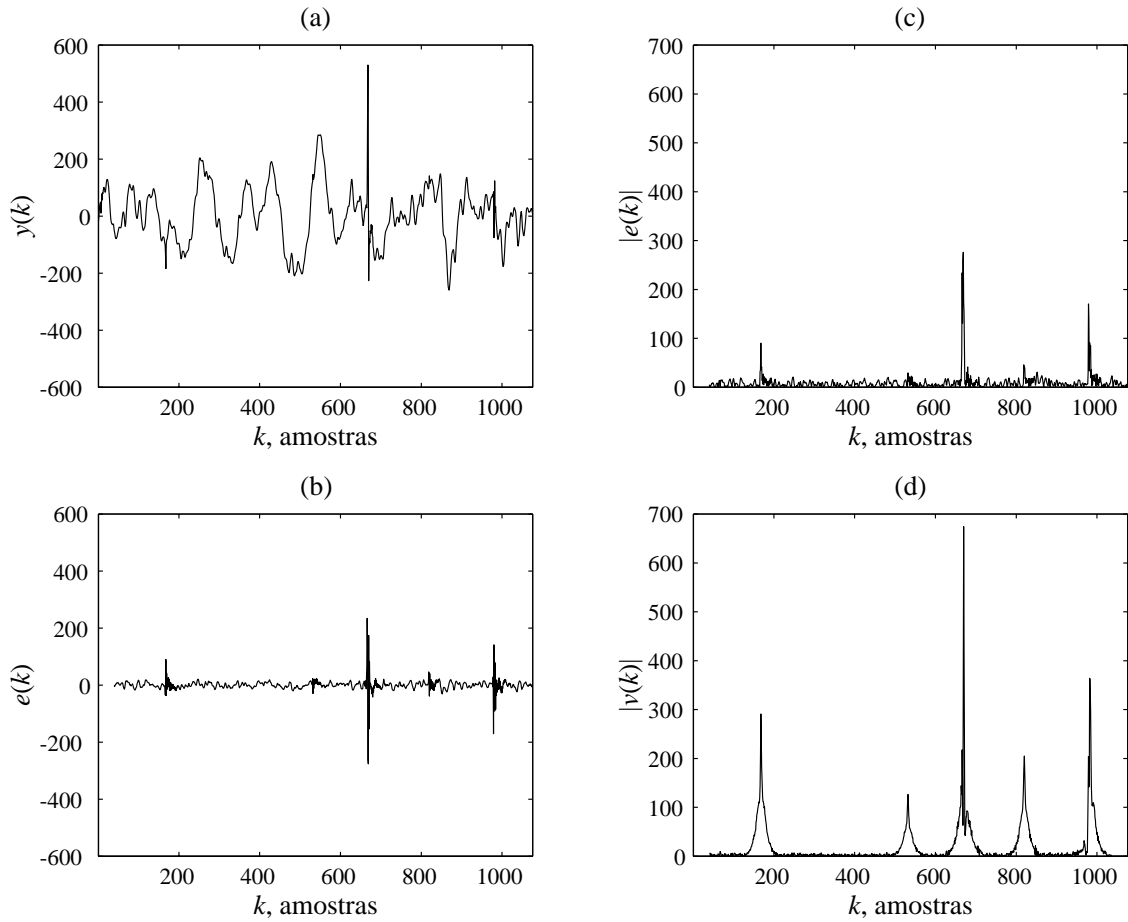


Figura 8.1: (a) Sinal de áudio corrompido por ruído impulsivo. (b) Sinal de excitação obtido por filtragem inversa. (c) Módulo do sinal de excitação. (d) Sinal obtido após a filtragem da excitação pelo filtro casado.

Evidenciados os pulsos, resta, então, adotar um limiar coerente para o teste da amplitude de $|e(k)|$. De acordo com as hipóteses feitas, $e_x(k)$ é Gaussiano. Assim, um limiar proporcional ao seu desvio-padrão

$$\lambda = C\sigma_{e_x} \quad (8.15)$$

é de interpretação imediata. Amostras acima de λ em $e(k)$ (isto é, distantes C desvios-padrões da média zero) sinalizariam possíveis distúrbios em $y(k)$. Novamen-

te, na situação real não se dispõe de σ_{e_x} , que precisa ser estimada. O método mais ingênuo de fazê-lo é calcular a média temporal de $e^2(k)$ no bloco; mas a presença dos distúrbios impulsivos tende a polarizar a estimativa, elevando impropriamente o limiar. Uma alternativa é arbitrar um percentual máximo ($P = 10\%$, por exemplo) para as amostras “fora da curva” e calcular a média temporal apenas sobre os $(1 - P)\%$ menores valores de $e^2(k)$ no bloco. Ainda menos custoso é estimar σ_{e_x} por $1,4826 \cdot \text{mediana}[|e(k)|]$, com base em [37].

Conhecido o índice k_0 da amostra acima de λ com maior amplitude em $e(k)$, pode-se tomar como pertencentes ao distúrbio um número fixo F de amostras em torno de $y(k_0)$ [36], procedimento que tende a superdimensionar a duração dos pulsos, já que F tem de ser no mínimo igual ao maior pulso admissível. É mais preciso eleger como corrompidas todas as amostras contíguas a $y(k_0)$ cuja excitação ainda exceda λ [16]. A correta determinação da extensão das falhas é retomada na Subseção 8.4.1. É possível, ainda, detectar as falhas conjuntamente, em vez de uma a uma, bastando para isso que se localizem os máximos locais na amplitude de $e(k)$ acima de λ , e não apenas o máximo absoluto. A possibilidade de ocorrência de distúrbios nas extremidades dos blocos é prevista permitindo-se pequena interseção entre estes.

Da mesma forma, a Interpolação pode ser realizada em conjunto, tenham as falhas sido detectadas individualmente ou não. A formulação do Capítulo 7 é geral, bastando que se admita mais de uma falha no interior do vetor \mathbf{x} , com as conseqüências sobre $\mathbf{x}_d(k)$ e $\mathbf{x}_c(k)$.

Cada vez que se realiza a interpolação de um ou mais trechos reconhecidos como corrompidos, torna-se possível estimar melhor o modelo $a(i)$, além de haver uma tendência à redução do limiar em conseqüência da redução na energia de $e(k)$. Isso pode ser aproveitado, se desejado, ao longo do processamento de um bloco. No caso da interpolação conjunta, enseja a realização de um novo ciclo, mais refinado. Em qualquer caso, espera-se idealmente que, não havendo mais falhas a tratar, a amplitude da excitação fique toda abaixo do limiar e se passe ao processamento do bloco subsequente. Na verdade, não há garantia de que isso ocorra, e este ponto será reexaminado na Subseção 8.4.2.

8.3.3 Restauração Bayesiana

A técnica Bayesiana [57, 58, 59, 47, 16] tem como fundamento a realização da Detecção das falhas empregando modelagem estatística para o ruído, além de para o sinal. Como conseqüências, determina com maior acerto a duração das falhas e, ao invés de descartar completamente as amostras atingidas para posterior interpolação, resgata a informação útil nelas contida, como se “separasse” sinal de ruído.

A Detecção Bayesiana consiste em escolher o processo de chaveamento $c(k)$ que maximiza sua probabilidade condicionada ao sinal ruidoso $y(k)$. Na forma vetorial, considerando-se um bloco de K amostras, busca-se a solução \mathbf{c}^{MAP} que maximiza a probabilidade *a posteriori*

$$P(\mathbf{c}|\mathbf{y}) = \frac{f_{\mathbf{y}|\mathbf{c}}(\mathbf{y}|\mathbf{c})P(\mathbf{c})}{f_{\mathbf{y}}(\mathbf{y})}, \quad (8.16)$$

dada pelo Teorema de Bayes. O procedimento, então, consiste em: para cada um dos 2^K possíveis vetores \mathbf{c} , calcular $f_{\mathbf{y}|\mathbf{c}}(\mathbf{y}|\mathbf{c})$, atribuir um valor para $P(\mathbf{c})$ e substituir ambos em 8.16; sendo $f_{\mathbf{y}}(\mathbf{y})$ constante para cada bloco, escolher \mathbf{c}^{MAP} .

A probabilidade *a priori* $P(\mathbf{c})$ deve refletir qualquer informação que se conheça do processo de chaveamento, como a forma das salvas em que ocorrem os pulsos e o percentual de contaminação do sinal. Não sendo possível estimá-la melhor, pode-se adotar $P(\mathbf{c}) = \frac{1}{2^K}$, para qualquer \mathbf{c} (uma hipótese nada razoável).

Supondo que $x(k)$ é Gaussiano, que se conhecem as p primeiras amostras de \mathbf{x} e que a distribuição $f_{\mathbf{n}_d}(\mathbf{n}_d)$ é Gaussiana com média zero e matriz-correlação \mathbf{R}_{n_d} , a densidade de probabilidade *a posteriori* é

$$f_{\mathbf{y}|\mathbf{c}}(\mathbf{y}|\mathbf{c}) = \frac{e^{-\frac{1}{2\sigma_e^2}[\mathbf{y}_c^T \mathbf{A}_c^T \mathbf{A}_c \mathbf{y}_c + \sigma_e^2 \mathbf{y}_d^T \mathbf{R}_{n_d}^{-1} \mathbf{y}_d + (\mathbf{A}_d^T \mathbf{A}_c \mathbf{y}_c - \sigma_e^2 \mathbf{R}_{n_d}^{-1} \mathbf{y}_d)^T x_d^{\text{MAP}}]}}{\sqrt{(2\pi\sigma_e^2)^{N-p-c^T \mathbf{c}} (\mathbf{R}_{n_d} \mathbf{A}_d^T \mathbf{A}_d + \sigma_e^2)}}, \quad (8.17)$$

sendo

$$\mathbf{x}_d^{\text{MAP}} = (\mathbf{A}_d^T \mathbf{A}_d + \sigma_e^2 \mathbf{R}_{n_d}^{-1})^{-1} (-\mathbf{A}_d^T \mathbf{A}_c \mathbf{y}_c + \sigma_e^2 \mathbf{R}_{n_d}^{-1} \mathbf{y}_d). \quad (8.18)$$

Uma simplificação adicional pode ser obtida com a hipótese (novamente pouco razoável, dada a óbvia correlação entre amostras da mesma falha, por exemplo) de que $n_d(k)$ são independentes com variância σ_n^2 , tornando

$$\mathbf{R}_{n_d} = \frac{1}{\sigma_n^2} \mathbf{I}. \quad (8.19)$$

Deve ser observado que o problema da Detecção posto dessa forma resolve conjuntamente o problema de Reconstrução—que, afinal, é o objetivo. No caso simplificado, em que se pode considerar $\frac{\sigma_e}{\sigma_n} \rightarrow 0$, então $\mathbf{x}_d^{\text{MAP}} \rightarrow \mathbf{x}_d^{\text{LS}}$ (v. Seção 7.2).

A maior desvantagem da abordagem Bayesiana em blocos é o esforço computacional excessivo requerido para calcular, para cada bloco de sinal de K amostras, 2^K probabilidades *a posteriori* $P(\mathbf{c}|\mathbf{y})$, o que torna praticamente inviável o processamento de sinais de áudio de longa duração. Para contornar esse problema, é desenvolvida em [57] uma abordagem seqüencial sub-ótima, na qual $f_{\mathbf{y}|\mathbf{c}}(\mathbf{y}|\mathbf{c})$ é calculada, recursivamente, a cada vez que uma nova amostra de \mathbf{y} é disponibilizada. Lá, o algoritmo realiza uma busca na árvore binária que representa todas as possibilidades para o vetor de detecção \mathbf{c} , eliminando, entretanto, as rotas em que a probabilidade *a posteriori* é baixa, promovendo, dessa forma, uma redução da carga computacional, em relação à busca exaustiva.

Outro aspecto a destacar é que, ainda que feitas as hipóteses e simplificações mencionadas, a implementação do algoritmo requer que se estimem parâmetros como \mathbf{a} e σ_e^2 , que não apresentam maiores problemas, e σ_n^2 , nada trivial. Também é possível preestimar aproximadamente o vetor \mathbf{c} por um método como o da Subseção 8.3.2.

Mas, afinal, todas as simplificações propostas para acelerar o processamento Bayesiano comprometem justamente a generalidade que seria sua principal vantagem.

8.4 Aperfeiçoamento da Detecção por Inversão do Modelo AR

Esta Seção se liga ao trabalho em [55].

Todos os aspectos a que se referiu a Subseção 8.3.2 foram exaustivamente verificados em simulações com sinais reais [52]. Os dois pontos que se mostraram mais frágeis no algoritmo foram a determinação correta da duração dos distúrbios pelo limiar e os critérios de reiteração e parada do algoritmo num dado segmento de sinal, ambos tendo merecido ser reelaborados [55, 60].

8.4.1 Determinação da Extensão da Falha

Os testes realizados com o método de localização baseado num único limiar fixo

$$\lambda = C\hat{\sigma}_{e_x} \quad (8.20)$$

mostraram a dificuldade de se encontrar uma constante de proporcionalidade C que permitisse extrair da comparação do sinal de excitação com λ o tamanho real das falhas, com suas variadas características. O limiar não pode ser muito baixo, na tentativa de reconhecer todas as amostras corrompidas do sinal, sob pena de detectar como inválidas amostras não-corrompidas (falsos alarmes); por outro lado, um limiar elevado tende a subdimensionar a largura das falhas, deixando intocadas amostras corrompidas (não-deteccões) que vão, em última análise, influir na interpolação das amostras detectadas.

Para atender esse compromisso, imaginou-se a utilização de dois limiares: o limiar **de deteção** λ_d , calculado, como antes, pela equação (8.20), e um limiar reduzido **de localização**

$$\lambda_l = \beta\lambda_d, \text{ com } 0 \leq \beta \leq 1. \quad (8.21)$$

Inicialmente, a idéia foi aplicada à Deteção individual de cada distúrbio: sua presença era indicada na excitação pela amostra de maior amplitude acima de λ_d ; sua extensão, pelas amostras contíguas com amplitude acima de λ_l . Na Deteção conjunta, mostrou-se mais eficiente aplicar primeiro λ_l e depois validar os conjuntos de amostras contíguas detectadas como distúrbios pela comparação com λ_d .

A Fig. 8.2 ilustra a Deteção de um distúrbio impulsivo com dois limiares.

Além de aumentar muito a precisão na determinação da extensão dos pulsos de ruído, o duplo limiar tornou a escolha de C mais robusta—o mesmo valor podendo ser utilizado com sucesso para sinais de características bem distintas. Tipicamente, os valores $C = 5$ e $\beta = 0,5$ fornecem bons resultados perceptivos.

Viu-se na Subseção 8.3.2 que a filtragem inversa leva a algum espalhamento dos pulsos de ruído na excitação. Esse espalhamento pode ocasionar interferência construtiva ou destrutiva entre amostras de um mesmo distúrbio ou de distúrbios próximos [16]. Não é raro, então, o sinal de excitação correspondente a um distúrbio

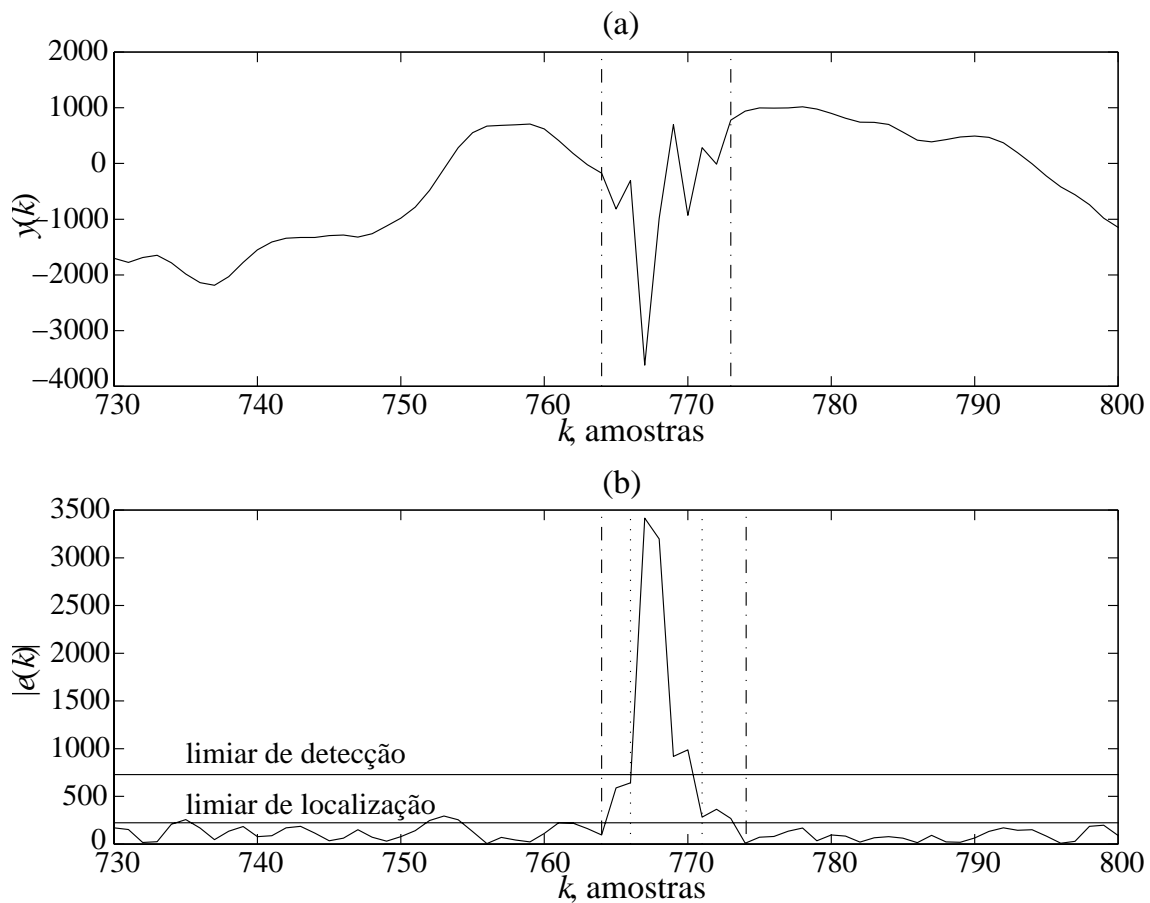


Figura 8.2: Exemplo de Detecção com dois limiares. (a) Trecho de sinal de áudio corrompido por distúrbio impulsivo—delimitado pelas linhas mistas verticais. (b) Módulo da excitação comparado aos dois limiares—as linhas mistas verticais delimitam o distúrbio localizado via λ_l , enquanto as linhas pontilhadas delimitam o distúrbio como seria localizado via λ_d .

relativamente duradouro exibir amostras intermediárias com módulos menores que o valor de λ_l , devido à interferência destrutiva. Nesses casos, mesmo com o uso de dois limiares, pode-se obter uma determinação imprecisa da duração do distúrbio ou, o que é pior, sua detecção na forma de dois ou mais distúrbios, separados por amostras consideradas não-corrompidas, mas que, de fato, são corrompidas.

A estratégia idealizada para contornar este problema foi promover uma união forçada de quaisquer distúrbios adjacentes separados por menos de k_{sep} amostras. Para isso, basta incorporar a cada distúrbio detectado todas as amostras adjacentes à de maior módulo até que haja, em ambos os lados, pelo menos k_{sep} amostras consecutivas da excitação com módulos menores que o limiar de localização.

A Fig. 8.3 mostra um exemplo real, em que o comportamento da amplitude da excitação na região do distúrbio (em torno da amostra $k = 300$) indicaria 4 distúrbios distintos, mas a união de distúrbios adjacentes permitiu realizar corretamente a Detecção.

Em geral, unir distúrbios separados por 2 ou 3 amostras, a uma taxa de amostragem de 44100 Hz, é suficiente; entretanto, para distúrbios de maior duração (mais de 30 amostras), esse número pode ser pequeno. Uma opção é fazer k_{sep} dependente da duração F que se determinaria para o distúrbio sem o critério de união, por exemplo, na forma

$$k_{\text{sep}} = k_{\text{sep}_{\text{mín}}} + \left\lfloor \frac{F}{g} \right\rfloor, \quad (8.22)$$

onde $k_{\text{sep}_{\text{mín}}}$ é o menor valor inteiro admitido para k_{sep} , g é um fator de controle da extensão de k_{sep} em relação à extensão de F e $\lfloor \cdot \rfloor$ é o operador “maior inteiro menor que ou igual a”.

8.4.2 Reiteração e Parada do Algoritmo por Bloco

Conforme já se disse na Subseção 8.3.2, é vantajoso iterar o processamento em cada bloco do sinal, a fim de refinar o tratamento até incluir os distúrbios impulsivos de menor amplitude.

Como a etapa de Reconstrução tende a reduzir a energia do sinal de excitação nos trechos onde houve distúrbios impulsivos suprimidos, espera-se que trechos já recuperados do sinal não sejam mais detectados numa nova etapa de Detecção que

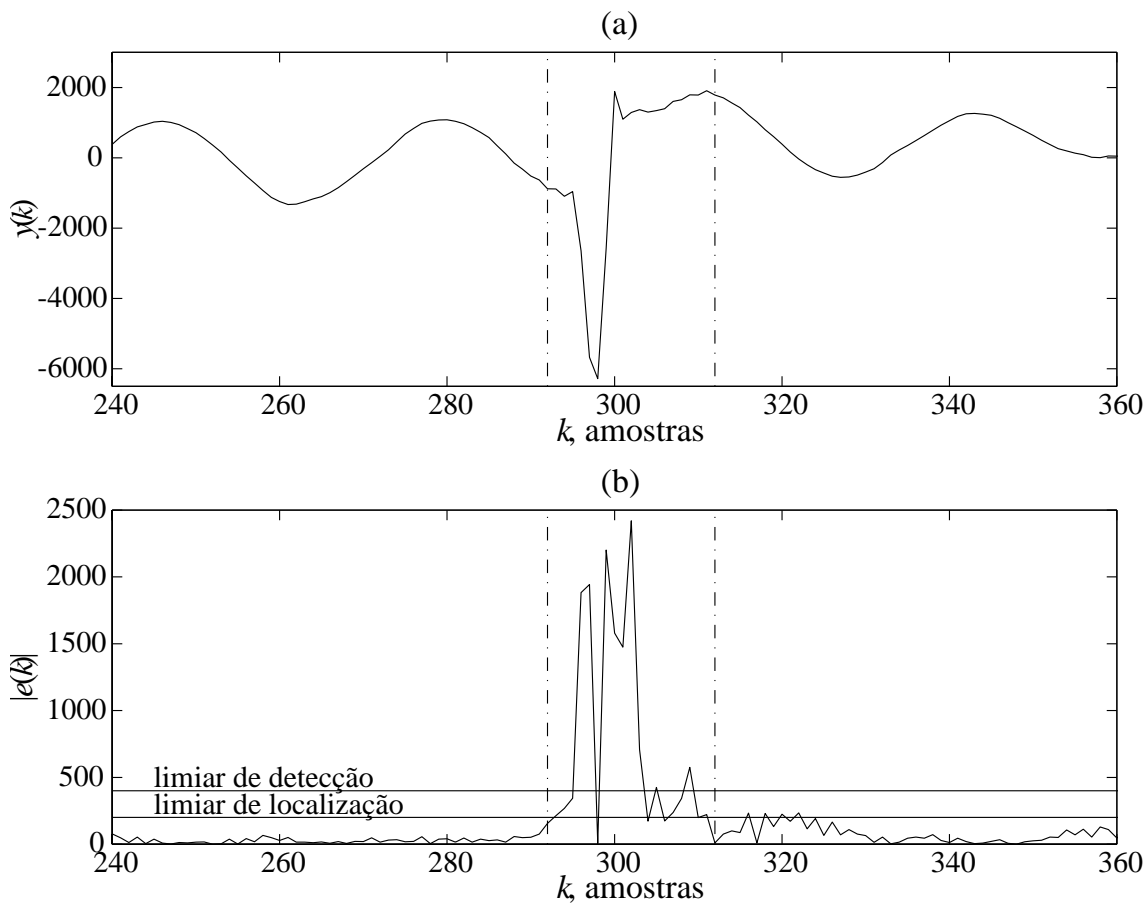


Figura 8.3: Interferência destrutiva e Detecção correta pela união de distúrbios adjacentes. (a) Sinal corrompido—por inspeção visual, o distúrbio impulsivo é delimitado pelas linhas mistas verticais. (b) Módulo da excitação e limiares—o distúrbio localizado usando união de distúrbios adjacentes (com $k_{\text{sep}} = 2$) é delimitado pelas linhas mistas verticais.

utilize idênticos parâmetros de processamento. Então, o critério mais óbvio para interrupção das iterações num mesmo bloco é a ocorrência de $|e(k)| < \lambda_D, \forall k$. Entretanto, não há garantia de que tal critério seja satisfeito por todos os blocos processados.

De fato, alguns distúrbios de pequena amplitude podem ser detectados com duração subestimada e, uma vez reconstruído o sinal no trecho correspondente, não ocorrer a esperada redução do sinal de excitação, tornando-os indefinidamente detectados. A forma mais simples de contornar esse problema é interromper as iterações quando não houver mais distúrbios aparentes (critério do parágrafo anterior) ou quando o número de iterações atingir uma contagem arbitrária $i_{\text{máx}}$.

É possível aumentar as chances de tratar convenientemente os pequenos pulsos de ruído mencionados acima antes de uma parada forçada no processamento permitindo-se a redução do limiar de localização. Isso pode ser feito diminuindo-se a razão β entre λ_d e λ_l conforme o número de iterações i é incrementado, por exemplo, na forma

$$\beta(i) = r^{\lfloor \frac{i}{f} \rfloor} \cdot \beta(1), \quad i = 2 \text{ a } i_{\text{máx}}, \quad (8.23)$$

onde r é o fator de redução e f , um parâmetro que controla a freqüência com que β é alterado. A escolha dos parâmetros $i_{\text{máx}}$, f e r está relacionada a observações experimentais. Além de permitir a supressão de distúrbios de menor amplitude, a utilização de β dependente da iteração também evita repetições inócuas de processamento, e contribui, dessa forma, para menores tempos de processamento.

8.4.3 Desempenho do Algoritmo Modificado

Os algoritmos de Detecção de Ruído Impulsivo por Filtragem Inversa do Modelo AR Convencional (DC) e Modificado (DM) foram implementados e aplicados a diferentes trechos de sinal com características distintas. A menos do fato de não empregar dois limiares e união de distúrbios adjacentes, a DC operou com os mesmos parâmetros (tamanho do bloco, ordem do modelo AR, fator C e limite das iterações por bloco $i_{\text{máx}}$) que a DM.

Três tipos de aferição foram feitas. Inicialmente, buscou-se avaliar a eficácia dos algoritmos na Detecção através dos percentuais de não-detecção e de falso alarme

encontrados nos resultados do tratamento de sinais artificialmente corrompidos. Em seguida, comparou-se subjetivamente (pela audição) a qualidade da supressão de ruído impulsivo alcançada com auxílio da DC e da DM. Por último, a comparação foi feita objetivamente utilizando-se a PAQM (medida perceptiva de qualidade de áudio), conforme descrita no Capítulo 6.

8.4.3.1 Não-Detecção e Falso Alarme

Para os testes, partiu-se de 6 sinais musicais (I a VI) de alta qualidade, abrangendo diversas características de espectro e dinâmica. Todos foram contaminados artificialmente por ruído impulsivo extraído de sinais naturalmente corrompidos, com as amplitudes ajustadas convenientemente para o efeito perceptivo desejado. Os sinais de I a IV foram somados ao ruído RA (percentual de corrupção menor que 1%) e os sinais V e VI foram somados ao ruído RB (percentual de corrupção em torno de 5%).

Aplicadas as rotinas DM e DC a cada sinal (com os mesmos parâmetros em todos os sinais), calcularam-se as porcentagens de não-detecção

$$ND = \frac{\text{n.º de amostras corrompidas não detectadas}}{\text{n.º de amostras corrompidas}} \quad (8.24)$$

e de falso alarme

$$FA = \frac{\text{n.º de amostras não-corrompidas indevidamente detectadas}}{\text{n.º de amostras não-corrompidas}} \quad (8.25)$$

para os fatores $C = 5$ e $C = 4,75$. A Tabela 8.1 mostra os resultados.

Os percentuais de não-detecção associados à DM foram entre 4 e 16 vezes menores que os apresentados pela DC. Já os percentuais de falso alarme associadas à DM foram entre 2 e 3 vezes maiores que os apresentados pela DC. Foi alcançada, portanto, a desejada redução no índice de não-detecção, ao custo de algum aumento do índice de falsa detecção era esperado—previsível, já que as modificações propostas tendem a superestimar, ligeiramente, a duração de alguns distúrbios. A comparação entre os resultados para os dois valores de C ilustra o balanço entre FA e ND .

8.4.3.2 Avaliação Subjetiva

A avaliação subjetiva foi feita comparando-se auditivamente os resultados da restauração de diversos sinais com a etapa de Detecção realizada pelos métodos

Tabela 8.1: Comparação entre os métodos de detecção convencional e modificado.

Sinais de Teste	$C = 5$				$C = 4,75$			
	DM		DC		DM		DC	
	$ND(\%)$	$FA(\%)$	$ND(\%)$	$FA(\%)$	$ND(\%)$	$FA(\%)$	$ND(\%)$	$FA(\%)$
I	2,34	3,13	14,18	1,06	1,83	4,03	13,56	1,25
II	3,68	1,66	18,20	0,63	2,97	2,49	17,21	0,76
III	0,84	2,02	14,01	0,72	0,76	2,73	13,37	0,86
IV	5,03	2,80	20,53	0,85	4,05	3,88	19,79	1,02
V	1,56	6,86	20,03	2,97	1,33	8,09	18,49	3,23
VI	2,43	7,20	17,60	3,20	2,18	8,43	16,66	3,46

Convencional e Modificado e a Reconstrução por interpolação LS (v. Seção 7.2).

Inicialmente, foram tratados sinais reais naturalmente contaminados de características espectrais e dinâmicas variadas; os resultados obtidos com a DM foram superiores em todos os casos, sendo mais acentuada a diferença nos sinais mais severamente contaminados. Em seguida, realizou-se a Reconstrução nos exemplos contaminados artificialmente da Subseção 8.4.3.1, com a mesma conclusão. Ficou claro que a redução no índice de não-deteção é, de fato, responsável pelos resultados perceptivamente superiores obtidos com a DM. Em contrapartida, o pequeno aumento do índice de falsa deteção, idealmente indesejável, embora eleve um pouco o esforço computacional, não afeta de modo significativo a qualidade do sinal de áudio; afinal, se o modelo AR é fiel, o processamento tenderá a substituir as amostras falsamente corrompidas por amostras praticamente iguais a elas.

8.4.4 Avaliação Objetiva por Critérios Perceptivos

No Capítulo 6, revisou-se uma medida objetiva da qualidade de sinais de áudio baseada em aspectos psicoacústicos (a PAQM) [42], e se propôs adaptá-la à aferição do desempenho de algoritmos de restauração de áudio. Esta medida será empregada a seguir no intuito de confirmar os resultados encontrados por avaliação subjetiva na Subseção 8.4.3.2.

Sendo necessária a disponibilidade do sinal de referência não-corrompido, lançou-se mão dos mesmos sinais utilizados na Subseção 8.4.3.1. Na Tabela 8.2,

lêem-se os valores da PAQM dos sinais corrompidos, tratados com DM e tratados com DC. São dados, ainda, como referência, os melhores valores alcançáveis pela PAQM com o método de Reconstrução empregado, correspondentes à Detecção ideal.

Tabela 8.2: PAQM dos sinais corrompidos por ruído impulsivo antes e depois do tratamento.

Sinais de Teste	Corrompidos	Tratados				
		$C = 4,75$		$C = 5$		Detecção
		DM	DC	DM	DC	Ideal
I	0,3891	0,0715	0,0750	0,0607	0,0756	0,0045
II	0,2183	0,0361	0,0462	0,0302	0,0428	0,0037
III	0,5774	0,0319	0,0947	0,0270	0,0982	0,0028
IV	0,3037	0,0461	0,0628	0,0377	0,0647	0,0037
V	1,1544	0,0644	0,4743	0,0584	0,5000	0,0133
VI	1,3969	0,0708	0,6664	0,0642	0,6777	0,0139

A PAQM emulou corretamente o julgamento subjetivo.

Novamente, DM levou a melhores resultados que DC. Para os sinais com alto percentual de contaminação por ruído impulsivo, o método convencional foi extremamente ineficaz, enquanto que o modificado manteve o padrão de qualidade dos outros sinais, confirmando o aumento da robustez na Detecção. Por fim, a DC levou a resultados melhores com $C = 5$ que com $C = 4,75$, indicando que neste caso o balanço entre a diminuição do percentual de não-detecções e o aumento do percentual de falsos alarmes (v. Tabela 8.1) foi desfavorável.

8.5 Tratamento em Sub-bandas

8.5.1 Descrição e Discussão

É bem conhecida a concentração preferencial da potência dos sinais de áudio musical nas regiões inferiores do espectro de frequências. Em contrapartida, os estreitos pulsos que formam o ruído dito impulsivo têm conteúdo espectral espalhado

com predomínio de freqüências elevadas. Estes fatos, aliados ao Mascaramento (v. Seção 6.2.2), sugerem que se possa atribuir importância variável com a freqüência à supressão dos distúrbios impulsivos. Especificamente, espera-se evoluir gradualmente das regiões de mais baixa freqüência, onde o sinal pode mascarar a maior parte do ruído, reduzindo a necessidade de tratamento, até as de mais alta freqüência, onde os distúrbios devem predominar, facilitando sua detecção ao mesmo tempo em que seu tratamento é indispensável. Este raciocínio levou à idéia de realizar o tratamento de ruído impulsivo em sub-bandas². Continuando a linha adotada anteriormente, decidiu-se partir da técnica de detecção por filtragem inversa do modelo + interpolação e suas variantes (Subseção 8.3.2 e Seção 8.4).

A forma mais simples de realizar o processamento de ruído impulsivo em sub-bandas é transportar diretamente o algoritmo original em banda completa para cada uma delas [61, 55]. A organização da música ocidental se fez em escalas baseadas na oitava, em decorrência de a seletividade da audição humana ser aproximadamente geométrica em freqüência [20]. Então, preferiu-se realizar a divisão do espectro pela separação de oitavas sucessivas através de bancos de filtros em árvore, como no esquema da Fig. 8.4. A cada estágio: na análise, os filtros passa-baixas $H_0(z)$ e passa-altas $H_1(z)$ dividem o espectro ao meio, e os sinais resultantes são (maximamente) decimados por 2; após o processamento, os sinais são interpolados por 2, filtrados pelo banco de síntese formado por $G_0(z)$ e $G_1(z)$ e reunidos.

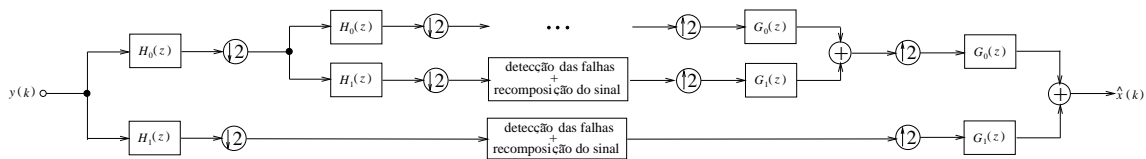


Figura 8.4: Esquema geral para tratamento do ruído impulsivo numa estrutura multitaxa em oitavas.

Algumas considerações precisam ser feitas nessa passagem do esquema em banda completa para o esquema em sub-bandas.

²O conteúdo desta Seção antecede cronologicamente o desenvolvimento do Capítulo 4, que foi apresentado anteriormente no texto por conveniência de organização. O assunto será revisitado à luz daquele desenvolvimento no Capítulo 12.

Um primeiro aspecto é a duração dos distúrbios a detectar. Se um pulso de ruído fosse apenas um impulso de alta amplitude, sua passagem por um par de filtros de análise resultaria nas respectivas respostas impulsivas; pode-se concluir, então, que a passagem do sinal pelos bancos de análise pode alongar os distúrbios. Esse efeito é “dividido à metade”, por assim dizer, pela decimação por 2. Por exemplo, supondo que os filtros são FIR de ordem Q , uma falha de duração F pode assumir, no pior caso, a duração de $\frac{F+Q}{2}$ amostras após um estágio de divisão do espectro. Mas, de qualquer modo, é importante ter em vista que se devem evitar filtros com resposta impulsiva significativa por muitas amostras.

Outro ponto muito importante se refere ao intervalo de tempo durante o qual se pode considerar estacionário o sinal na etapa de determinação do modelo para realização do processamento, agora em cada sub-banda. Tipicamente, adota-se para os sinais de áudio um tempo de estacionaridade máximo admissível entre 25 e 50 ms (v. Subseção 1.3.1.1). No sistema de processamento, esse tempo determina o número de amostras K de cada bloco, como definido na Subseção 8.3.1. Ora, considerando-se estritamente esse dado, a cada estágio de análise o número máximo de amostras permissíveis por bloco em que se poderiam dividir os sinais nas sub-bandas (uma vez filtrados e decimados por 2) é reduzido à metade. Após sucessivos estágios, os blocos podem resultar tão curtos a ponto de inviabilizar a estimação de um modelo para os sinais nas sub-bandas. É desejável, portanto, manter um número suficientemente elevado de amostras por bloco em todas as sub-bandas—igual, talvez. Mas faz sentido atribuir tempos de estacionaridade diferentes (em segundos) às sub-bandas, na verdade crescentes em direção às baixas frequências? Pode-se construir o seguinte argumento: a própria divisão do sinal em faixas de frequência tende a confinar, ou representar preferencialmente, variações rápidas no modelo nas sub-bandas superiores (de taxa mais alta), enquanto que as sub-bandas inferiores (de taxa mais baixa) descrevem principalmente as variações de longo prazo. Assim, a atribuição de tempos gradativamente maiores para preservação dos modelos em direção às sub-bandas inferiores não deve incorrer em prejuízo da descrição das variações de curto prazo do modelo. Pelo contrário, isso permite descrever melhor as mais lentas, já que há amostras suficientes para que se realizem as estimativas. Optou-se, então, por manter constante o número de amostras de cada bloco em

todas as sub-bandas, o que equivale dobrar o tempo de estacionaridade aceitável a cada divisão sucessiva do espectro.

A questão de mais difícil enfrentamento, contudo, é a alteração das características do modelo estatístico atribuído ao sinal em banda completa. O sinal que podia ser representado por um processo AR se converte em que nas sub-bandas? O uso de bancos de análise de filtros FIR, por exemplo, já confere características ARMA aos sinais; mas a decimação, que é uma operação variante no tempo, ainda permite que se associe um modelo ARMA aos sinais nas sub-bandas? E de que forma eles se interrelacionam? Essa discussão teórica (que originaria, posteriormente, o Capítulo 4) foi, inicialmente, deixada de lado. Uma série de medidas realizadas sobre sinais reais autorizaram uma postura relativamente ingênua: a utilização de modelos AR de ordem suficiente nas sub-bandas. A notar, apenas, o sentimento de que a complexidade do modelo poderia ser reduzida ao longo das sucessivas divisões do espectro.

8.5.2 Implementação e Avaliação

Foi realizada uma quantidade muito grande de testes do sistema proposto sobre sinais reais. De forma ampla, pode-se dizer que o conjunto de escolhas que representou o melhor compromisso entre desempenho e complexidade foi o uso de 3 estágios de divisão por bancos QMF FIR de ordem 15. Os resultados observados [55] foram bastante animadores.

De forma bem resumida, podem ser destacados alguns aspectos:

- A escolha dos diversos parâmetros de processamento (v. Seção 8.4), que, em geral, tiveram de ser diferenciados para cada sub-banda, não foi tão simples quanto no sistema em banda completa.
- Em algumas situações, observaram-se artefatos de baixa freqüência após o processamento nas regiões tratadas, isto é, onde havia distúrbios, a princípio atribuídos ao espalhamento (ou talvez “diluição”) dos distúrbios nas sub-bandas de mais baixas freqüências.
- Foi possível, especialmente com sinais com informação originalmente pobre em altas freqüências, simplificar o procedimento de detecção de distúrbios na

sub-banda mais aguda (até mesmo realizando a detecção diretamente sobre o sinal).

- A redução de esforço computacional se mostrou em geral significativa, em especial quando foi possível aplicar um esquema simplificado de detecção.

Retorna-se a essa idéia de realizar o tratamento de ruído impulsivo em sub-bandas no Capítulo 12, já com o apoio do Capítulo 4.

Capítulo 9

Tratamento de Pulsos Longos

Danos sérios nos sulcos de uma matriz em disco ou cilindro, como os originados por fendas ou arranhões profundos na superfície ou mesmo por quebra, excitam o sistema de reprodução da matriz com estímulos curtos e de altas amplitudes em relação às dos sulcos gravados. Como resposta, são gerados pulsos de longa duração e elevada amplitude, chamados aqui **pulsos longos**, que ocasionam fortes estalos na audição—um dos tipos de distúrbios mais desagradáveis encontrados em gravações analógicas de áudio.

O tratamento digital de pulsos longos deve ser o primeiro a ser realizado, nos casos em que há também a tratar ruído impulsivo (assunto do Capítulo 8) e ruído de fundo (assunto do Capítulo 10), pois sua alta energia tende a polarizar as estimativas necessárias às outras técnicas de restauração.

Como o ruído impulsivo, os pulsos longos são distúrbios localizados e que, portanto, admitem ser processados em 2 etapas, detecção de falha e recomposição das amostras danificadas. Entretanto, ocorrem com frequência muito reduzida (por exemplo, um arranhão radial num disco de 78 rpm produz estalos espaçados de cerca de 765 ms), e têm grande duração (em geral, de mais de 5 ms). Tipicamente, consistem de uma descontinuidade inicial impulsiva de amplitude muito alta seguida de um transitório assemelhado a oscilações amortecidas de frequência baixa e decrescente, como se vê na Figura 9.1. No espectrograma do sinal corrompido, suas ocorrências geram faixas espectrais largas localizadas no tempo, como indica a Figura 9.2.

A recomposição de um número elevado de amostras contíguas não é um

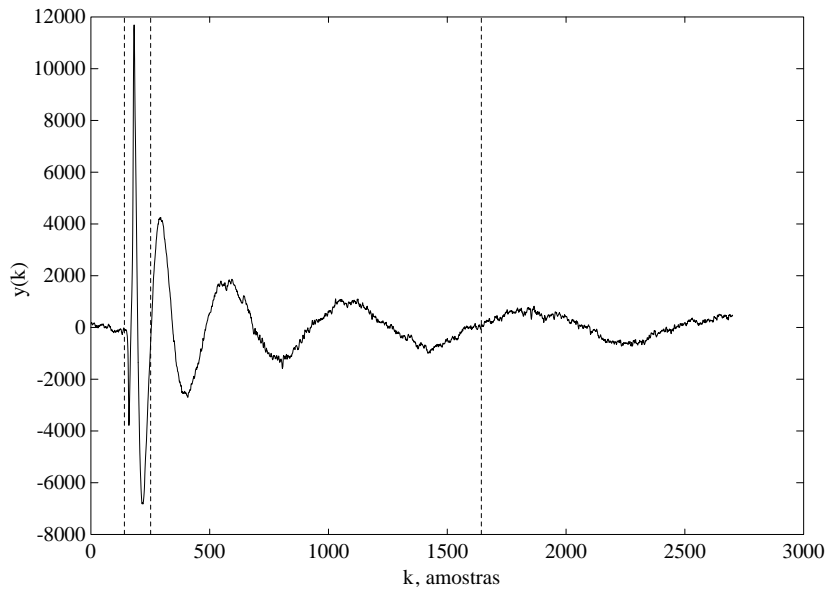


Figura 9.1: Exemplo de pulso longo: as linhas verticais delimitam aproximadamente as regiões de descontinuidade inicial e de transitório de baixa frequência.

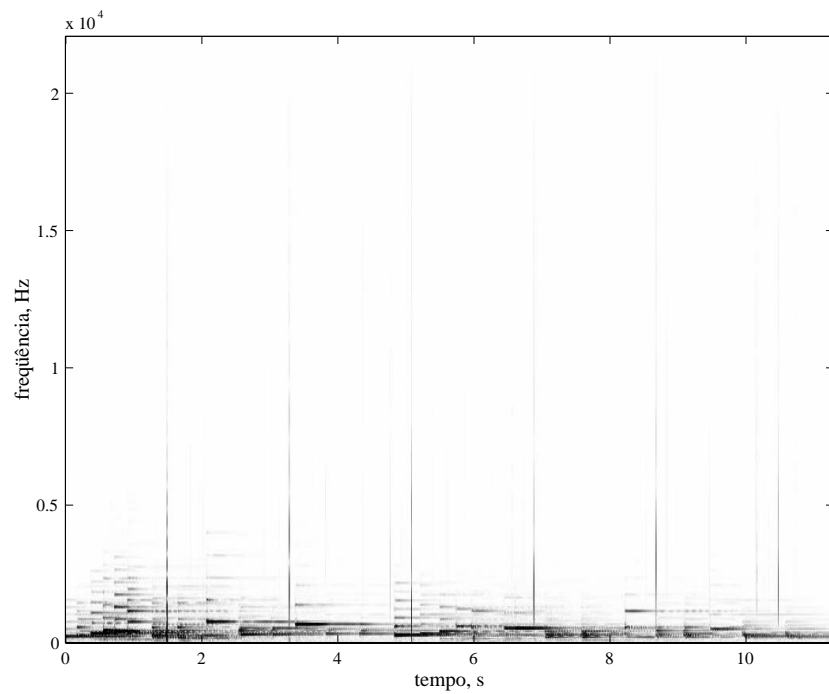


Figura 9.2: Espectrograma de um sinal de áudio originalmente contaminado por pulsos longos: blocos de 1024 amostras, FFT de 1024 pontos e superposição de 50%.

problema trivial, mas as ocorrências raras de altas energias sugerem que um processo genérico de localização possa ser simples e rápido.

Este Capítulo sobre restauração de sinais contaminados por pulsos longos inicialmente revisa uma técnica de restauração por pulso-padrão, que foi avaliada em simulações, e uma técnica para recomposição das amostras corrompidas. Em seguida, discute propostas para localização eficiente das falhas, em particular operando no domínio da frequência.

9.1 Restauração por Similaridade

A duração dos pulsos longos, em geral, é da ordem de grandeza do tempo de estacionaridade máximo que se pode atribuir ao sinal de áudio (da ordem de dezenas de milissegundos), dificultando a modelagem estatística do sinal. Uma alternativa pode ser a realização do processamento em duas etapas: uma pré-correção dos pulsos, em que se suprime o longo transitório de baixa frequência, seguida do tratamento do resíduo inicial de característica impulsiva pelas técnicas apropriadas (v. Capítulo 8). Esta Seção trata de um método para realizar a pré-correção.

O método de detecção e supressão de pulsos longos por similaridade ou por pulso-padrão apresentado em [36] se baseia na hipótese de que os pulsos longos que corrompem aditivamente o sinal são muito semelhantes, pois ditados pela resposta impulsiva do sistema usado na reprodução da matriz. Consiste em subtrair de cada pulso a suprimir uma versão sincronizada e de amplitude adequada de um pulso-padrão similar.

Os pulsos-padrões podem provir de uma biblioteca disponível ou ser obtidos de uma gravação particular da seguinte forma: localiza-se por algum procedimento um primeiro pulso no sinal, que é adotado como padrão; buscam-se ao longo do sinal trechos cujos coeficientes de correlação com o padrão sejam localmente máximos e próximos de 1; cada ocorrência é acumulada no padrão, que se vai tornando progressivamente mais “limpo”.

A busca e sincronização dos pulsos a suprimir com um dado pulso-padrão é igual à descrita no parágrafo anterior. Sendo a duração dos pulsos, já sincronizados,

igual a K amostras, o sinal no trecho pode ser descrito como

$$y(k) = x(k) + n(k), \quad (9.1)$$

onde $x(k)$ é o sinal original e $n(k)$ é o pulso de ruído a suprimir, supostos não-correlacionados. Sendo o pulso-padrão $\bar{n}(k)$ representativo da família de $n(k)$, eles diferem essencialmente em amplitude, isto é,

$$n(k) \approx \mu \bar{n}(k); \quad (9.2)$$

pode-se, portanto, obter uma estimativa para $x(k)$

$$\hat{x}(k) = y(k) - \mu \bar{n}(k), \quad (9.3)$$

desde que se consiga uma estimativa de μ . A correlação temporal do trecho ruidoso com o pulso-padrão fornece

$$\sum_{k=0}^{N-1} y(k) \bar{n}(k) = \sum_{k=0}^{N-1} x(k) \bar{n}(k) + \mu \sum_{k=0}^{N-1} [\bar{n}(k)]^2. \quad (9.4)$$

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

$$\mu \approx \frac{\sum_{k=0}^{N-1} y(k) \bar{n}(k)}{\sum_{k=0}^{N-1} [\bar{n}(k)]^2}, \quad (9.5)$$

que, levado à equação (9.3), soluciona o problema.

A limitação na eficácia deste método está em requerer uma vasta biblioteca de pulsos-padrões representativos e/ou depender da invariância dos pulsos a suprimir. Em geral, os pulsos longos variam tanto em duração quanto no formato, e na ocorrência eventual de dois arranhões bem próximos, a superposição dos pulsos pode inviabilizar seu processamento [57].

O fato de se precisar percorrer o sinal amostra-a-amostra na busca dos distúrbios acaba por requerer uma carga de processamento incompatível com a raridade de ocorrência dos distúrbios. No caso de arranhões aproximadamente radiais num disco, os pulsos longos ocorrem separados de aproximadamente um período de rotação da matriz, o que permite limitar sua busca ao entorno das regiões estimadas dessa forma; mas esse não é um procedimento genérico.

9.1.1 Testes Realizados

Realizaram-se experimentos com sinais de áudio provenientes de discos de 33 1/3 rpm com arranhões, a fim de valiar o desempenho do método descrito [61]. Pulsos-padrões foram determinados para cada gravação.

O que se observou, apesar de a reprodução das matrizes ter sido feita pelo mesmo sistema, foi que o grau de disparidade dos pulsos dificulta tanto a formação do pulso-padrão quanto a supressão dos distúrbios, apesar de sua fácil detecção. Como consequência, após o resíduo impulsivo previsto, ainda permanece uma contaminação audível de baixa amplitude na região do transitório eliminado, extremamente difícil de ser eliminada.

Referindo-se a um dos exemplos práticos simulados, a Figura 9.3 mostra o pulso-padrão obtido a partir das ocorrências de pulsos longos no trecho musical gravado; a Figura 9.4 mostra 6 dessas ocorrências, superpondo o trecho de sinal corrompido, o pulso-padrão invertido com amplitude corretamente ajustada e o trecho de sinal pretratado resultante da soma de ambos.

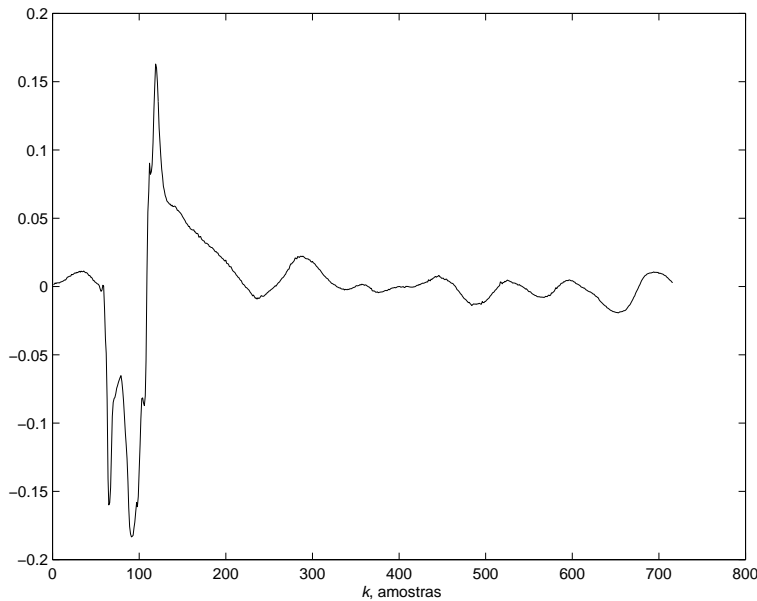


Figura 9.3: Pulso-padrão calculado para um sinal corrompido.

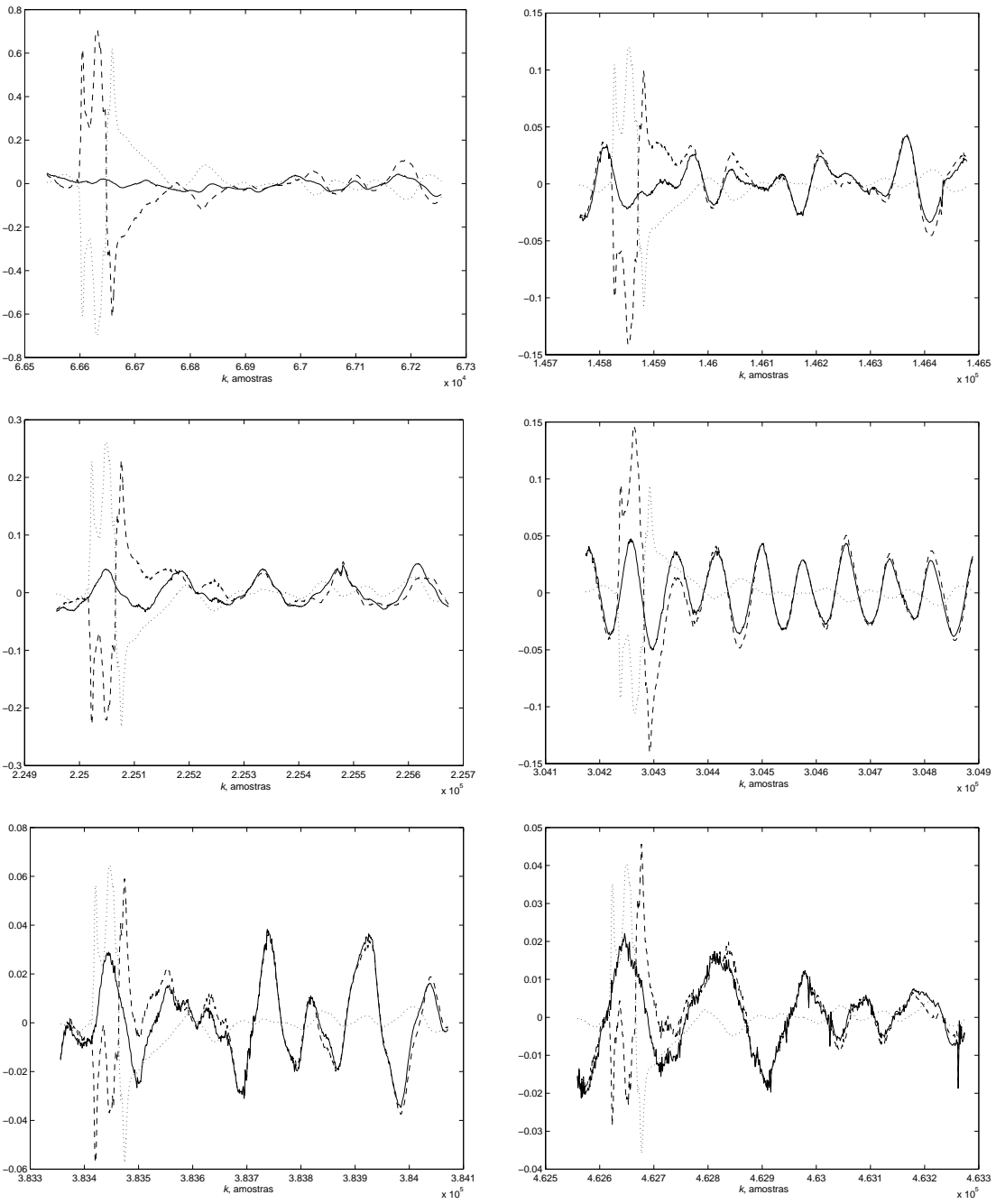


Figura 9.4: Seis ocorrências de pulso longo e seu pré-processamento. Tracejado: sinal corrompido; pontilhado: padrão invertido a somar; contínuo: sinal resultante da soma.

9.2 Supressão por Separação de Processos AR

A supressão de pulsos longos através de separação de processos autorregressivos (AR) proposta em [57, 62] é mais geral que o método de detecção/supressão por pulso-padrão [36]; entretanto, as localizações dos distúrbios 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, modelado como um processo AR de ordem entre 40 e 80, através de um algoritmo Bayesiano.

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

$$\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2, \quad (9.6)$$

sendo \mathbf{y}_1 e \mathbf{y}_2 originados por processos AR Gaussianos independentes análogos ao da equação (7.3), com ordens I_1 e I_2 , vetores de coeficientes \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 I_1 primeiras amostras de \mathbf{y}_1 e as I_2 primeiras amostras de \mathbf{y}_2 , é possível expressar suas densidades de probabilidade como

$$f_{\mathbf{y}_1}(\mathbf{y}_1) = \frac{1}{\sqrt{(2\pi\sigma_{e_1}^2)^{K-I_1}}} e^{-\frac{1}{2\sigma_{e_1}^2} \mathbf{y}_1^T \mathbf{A}_1^T \mathbf{A}_1 \mathbf{y}_1} \quad (9.7)$$

e

$$f_{\mathbf{y}_2}(\mathbf{y}_2) = \frac{1}{\sqrt{(2\pi\sigma_{e_2}^2)^{K-I_2}}} e^{-\frac{1}{2\sigma_{e_2}^2} \mathbf{y}_2^T \mathbf{A}_2^T \mathbf{A}_2 \mathbf{y}_2}, \quad (9.8)$$

sendo \mathbf{A}_1 e \mathbf{A}_2 definidas de forma análoga à equação (7.12). A obtenção da estimativa mais provável *a posteriori* de \mathbf{y} , $\mathbf{y}_1^{\text{MAP}}$, decorre da maximização de $f_{\mathbf{y}_1|\mathbf{y}}(\mathbf{y}_1|\mathbf{y})$ e resulta em

$$\mathbf{y}_1^{\text{MAP}} = \left(\frac{\mathbf{A}_1^T \mathbf{A}_1}{\sigma_{e_1}^2} + \frac{\mathbf{A}_2^T \mathbf{A}_2}{\sigma_{e_2}^2} \right)^{-1} \frac{\mathbf{A}_2^T \mathbf{A}_2}{\sigma_{e_2}^2} \mathbf{y}. \quad (9.9)$$

Passando ao problema de restauração do sinal já descrito na equação (9.1), agora na forma vetorial

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \quad (9.10)$$

pode-se associar $x(k)$ (e \mathbf{x}) com coeficientes \mathbf{a}_x (e matriz \mathbf{A}_x) a \mathbf{y}_1 , e $n(k)$ (e \mathbf{n}) com coeficientes \mathbf{a}_n (e matriz \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 K componentes binárias que indiquem por 1 as amostras de alta amplitude da descontinuidade inicial do pulso e por 0 as demais amostras, pode-se obter a matriz de autocorrelação $\mathbf{R}_{e_n e_n}$ de $e_n(k)$, que é diagonal com elementos

$$\lambda_{kk} = \sigma_{e_{n_0}}^2 + i(k)(\sigma_{e_{n_1}}^2 - \sigma_{e_{n_0}}^2). \quad (9.11)$$

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

$$f_{\mathbf{n}}(\mathbf{n}) = \frac{1}{\sqrt{(2\pi)^{K-I_2} |\mathbf{R}_{e_n e_n}|}} e^{-\frac{1}{2} \mathbf{n}^T \mathbf{A}_n^T \mathbf{R}_{e_n e_n}^{-1} \mathbf{A}_n \mathbf{n}}, \quad (9.12)$$

levando à nova solução

$$\mathbf{x}^{\text{MAP}} = \left(\frac{\mathbf{A}_x^T \mathbf{A}_x}{\sigma_{e_x}^2} + \mathbf{A}_n^T \mathbf{R}_{e_n e_n}^{-1} \mathbf{A}_n \right)^{-1} \mathbf{A}_n^T \mathbf{R}_{e_n e_n}^{-1} \mathbf{A}_n \mathbf{y}. \quad (9.13)$$

Os coeficientes do modelo AR associado ao sinal não-ruído podem ser obtidos por estimativa LS (v. Subseção 3.2.2) usando-se 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-se detecção por filtragem inversa, tal como na detecção de ruído impulsivo (v. Subseção 8.3.2), associando-se valor 1 aos elementos de \mathbf{i} com índices iguais aos das amostras da excitação cujos módulos excedem um limiar de valor alto.

A supressão de pulsos longos por separação de processos AR objetiva ser menos dependente da hipótese de invariabilidade dos pulsos longos, o que permite suprimir tanto pulsos com transitórios diferentes como pulsos superpostos, e ainda realiza a supressão dos pulsos em uma única etapa. Deve ser notado que a recomposição das amostras corrompidas não as descarta para posterior recriação, o que permite lidar com distúrbios de longa duração sem a necessidade de longas interpolações.

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 [57], o autor adota o caso limite $\sigma_{e_{n_1}}^2 \rightarrow \infty$, que leva a zero os elementos de $\mathbf{R}_{e_n e_n}^{-1}$ 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. (Mas isso retoma, em parte, os problemas associados ao método da Seção 9.1!)

9.3 Detecção Rápida

A busca amostra-a-amostra inerente à restauração por similaridade de sinais contaminados por pulsos longos e a necessidade de associar uma localização eficiente dos pulsos ao método Bayesiano de supressão sugeriram a investigação de formas rápidas de detecção.

A busca em blocos pareceu uma forma econômica de identificar a ocorrência dos pulsos, bastando para isso extrair deles informações que permitam distinguir com contraste suficiente os blocos corrompidos dos não-corrompidos. A seguir, descreve-se a evolução de um método eficiente de detecção em blocos no domínio da frequência, original de [55] com coorientação do autor da presente tese.

Dada a elevada amplitude dos pulsos a localizar, uma primeira idéia simples é aplicar ao módulo $|y(k)|$ do sinal contaminado um limiar λ que, ultrapassado, sinalize a ocorrência dos distúrbios (método I). Mas as variações de amplitude do sinal, a ponto de os pulsos longos “submergirem”, embora permanecendo audíveis, inviabiliza este procedimento.

Um primeiro método em blocos consiste em aplicar o limiar de detecção à energia

$$\varepsilon_y(b) = \sum_{k=(1-\alpha)Kb=0}^{K-1} y^2(k), \quad 0 < \alpha < 1 \quad (9.14)$$

de cada quadro b de K amostras do sinal, com a fração de superposição entre quadros dada por α (método II). Entretanto, novamente a ocorrência de trechos de potência elevada impede a discriminação dos pulsos.

Uma alternativa que tenta destacar o ruído do sinal, baseada num método para detecção de ruído impulsivo (v. Subseção 8.3.2), consiste em aplicar o limiar de detecção à energia

$$\varepsilon_e(b) = \sum_{k-(1-\alpha)Kb=0}^{K-1} e^2(k), \quad 0 < \alpha < 1 \quad (9.15)$$

do sinal de excitação $e(k)$ associado ao modelo autorregressivo obtido para cada quadro do sinal (método III). Na excitação, o sinal subjacente dá lugar a ruído branco e os pulsos pouco se alteram. Como se espera, agora os pulsos se destacam melhor, mesmo nos trechos de potência elevada. Porém, a aplicação de um limiar fixo ainda não parece possível.

Pelo teorema de Parseval, a energia do sinal também poderia ser calculada pela soma dos módulos quadrados das componentes $Y_b(w)$ da FFT do quadro. Como vimos na Figura 9.2, a ocorrência de pulsos longos gera raias ao longo de quase todo o espectro nos respectivos blocos; por outro lado, a potência dos sinais de áudio se concentra predominantemente em baixas frequências. Então, restringir à faixa de frequências $\omega_1 \leq \omega \leq \omega_2$ em que tipicamente é pior a razão sinal-ruído o cálculo da energia da cada quadro do sinal pela FFT

$$\varepsilon_{yf}(b) = \sum_{\omega=\omega_1}^{\omega_2} |Y_b(w)|^2 \quad (9.16)$$

antes da aplicação do limiar (método IV) deve favorecer o contraste entre blocos corrompidos e não-corrompidos. (Na verdade, isso equivale a pré-filtrar o sinal antes de calcular sua potência, porém sem reverter ao domínio do tempo.) Agora, a eficácia se compara à do método anterior, mas com menor processamento.

Considerando apenas as componentes selecionadas da FFT, como anteriormente, o que se deseja é conseguir distinguir facilmente um vetor não-contaminado, com poucas componentes elevadas (as componentes tonais do sinal) de sua versão ruidosa, acrescida de múltiplas componentes reduzidas (correspondentes ao pulso espalhado em frequência). Sejam, por analogia, o vetor \mathbf{v}_1 com 2 componentes de amplitude 10 e 18 componentes nulas e o vetor \mathbf{v}_2 com 2 componentes de amplitude 10 e 18 componentes de amplitude 1. O cálculo de suas normas $\|\mathbf{v}_1\|_1 = 20$, $\|\mathbf{v}_1\|_2 = 200$, $\|\mathbf{v}_2\|_1 = 38$ e $\|\mathbf{v}_2\|_2 = 218$ mostra que a norma 1 permite contraste

maior entre os 2 vetores que a norma 2. Isso sugere o uso da soma dos módulos

$$\mu_{y_f}(b) = \sum_{\omega=\omega_1}^{\omega_2} |Y_b(\omega)| \quad (9.17)$$

em substituição à soma dos quadrados dos módulos das componentes de interesse da FFT antes da aplicação do limiar (método V). De fato, mesmo os pulsos mais difíceis de serem detectados se tornam bem destacados do entorno. Mas ainda é inviável a aplicação de um limiar fixo.

A filtragem por mediana retira as ocorrências localizadas de valores atípicos de uma seqüência. Daí veio a idéia de aplicar o limiar à seqüência de parâmetros de comparação subtraída de sua versão filtrada por mediana com uma janela adequada,

$$\Delta\mu_{y_f}(b) \triangleq \mu_{y_f}(b) - \text{mediana}[\mu_{y_f}(b)], \quad (9.18)$$

a qual se espera que contenha praticamente só as ocorrências atípicas (método V'). Dessa forma, um limiar fixo é capaz de detectar os pulsos longos. Vale observar que a mediana também pode ser utilizada como auxiliar nos métodos III e IV, com seus novos parâmetros definidos analogamente à equação 9.18. Nessas condições, o método III' se mostra superior ao IV', mas o V' ainda supera o III'.

Para exemplificar a aplicação de todos os métodos e os comentários acima, utilizamos um trecho de sinal com faixa dinâmica ampla contaminado com pulsos longos, o que determina ocorrências tanto de fácil quanto de difícil identificação. As Figuras 9.5 e 9.6 mostram ao longo do tempo os parâmetros de comparação associados a cada método. Os limiares, quando apresentados, servem apenas como orientação.

Para um caso típico, com blocos de 2048 amostras e modelo autorregressivo de ordem 40, o método V é mais de 12 vezes mais rápido que o método III [55].

Com isso, mostrou-se que a associação de uma medida espectral com a filtragem por mediana pode resultar num método extremamente eficaz e rápido para detecção de blocos de sinal contaminados por pulsos longos.

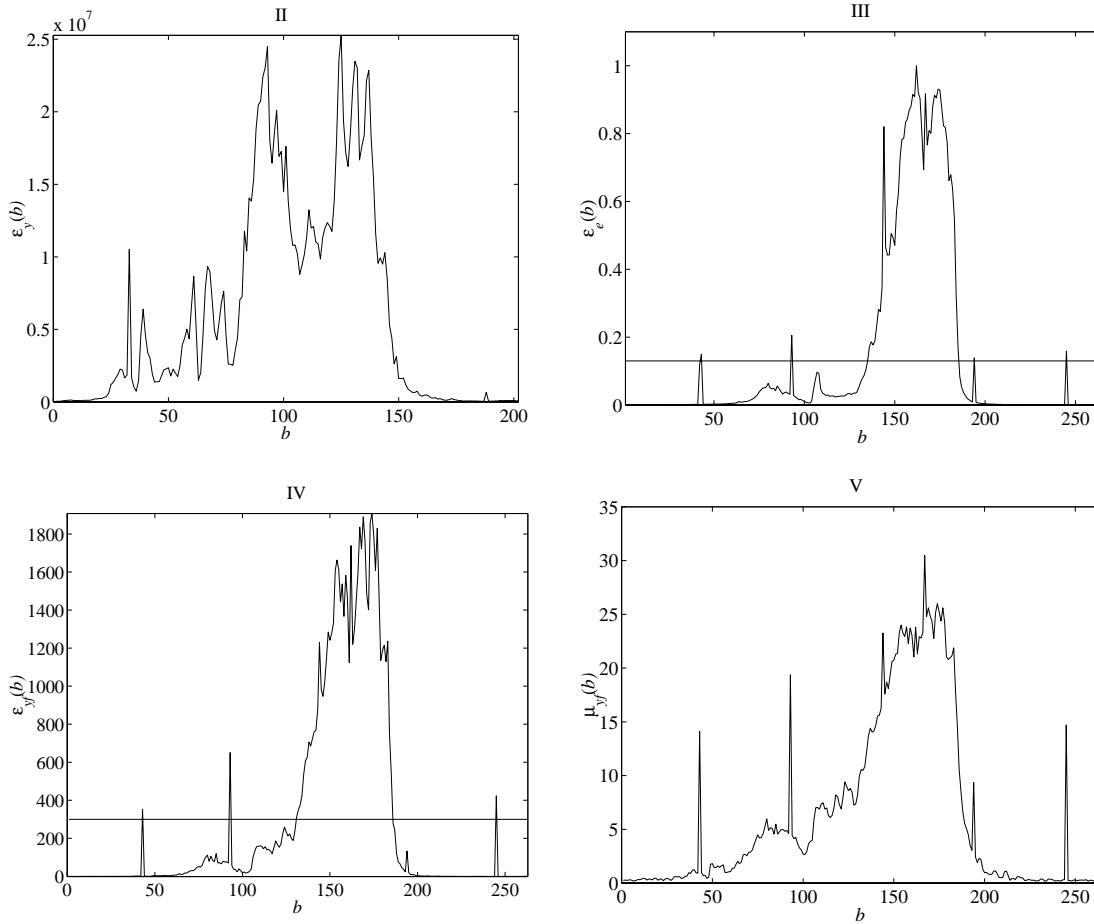


Figura 9.5: Parâmetros de comparação utilizados nos métodos de II a V ao longo do tempo.

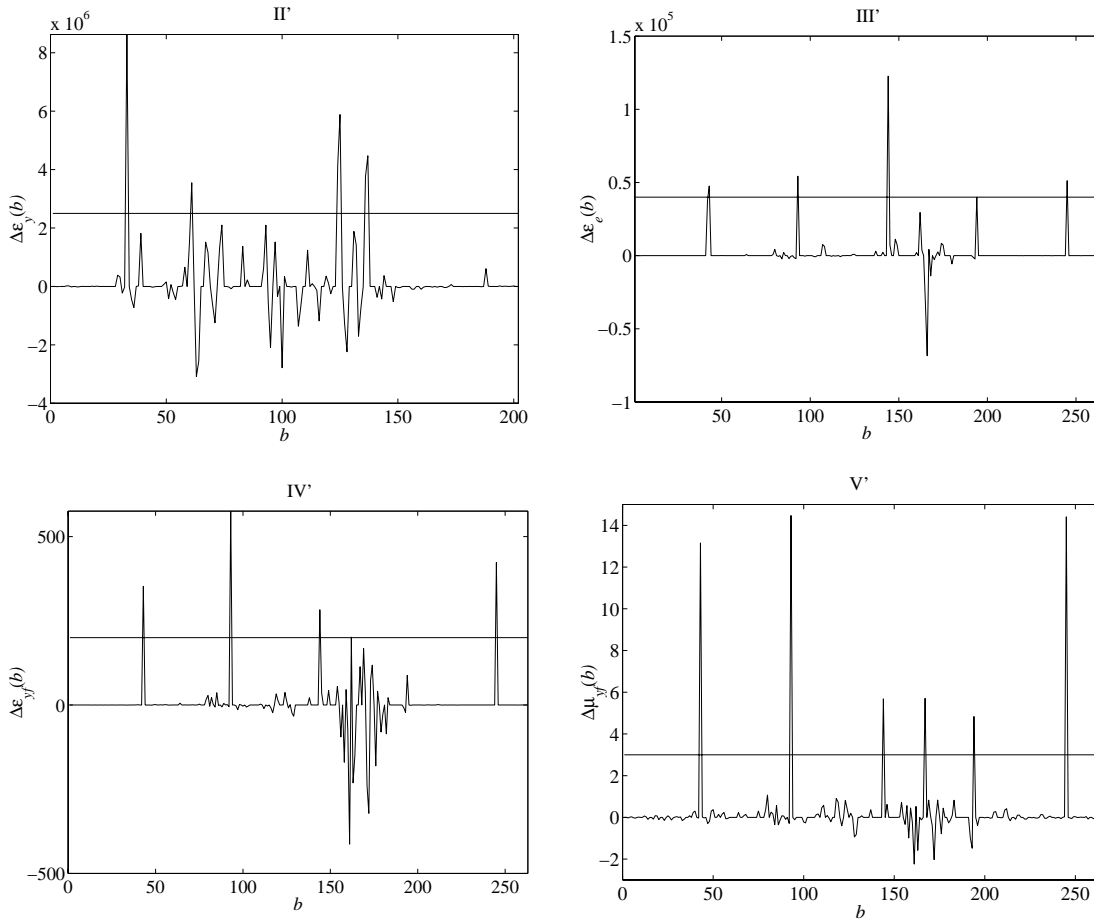


Figura 9.6: Parâmetros de comparação utilizados nos métodos de II' a V' ao longo do tempo.

Capítulo 10

Restauração de Sinais

Contaminados por Ruído de

Fundo

Um tipo de degradação bastante usual em gravações de áudio consiste num chiado de tonalidade indistinguível presente ao longo do sinal, que pode ter sido gerado pelos próprios equipamentos envolvidos na gravação e reprodução ou ser inerente às fitas magnéticas analógicas, se utilizadas nesses processos.

Esse ruído apresenta, em geral, características estacionárias durante longos intervalos de tempo e espectro de densidade de potência bastante largo e de contorno suave, isto é, sem picos pronunciados. Se for possível considerar esse espectro aproximadamente constante por toda a faixa de frequências de interesse, poder-se-á tratá-lo como ruído branco. Ruído de banda larga é um problema intratável por filtragem linear fixa, uma vez que não pode ser separado do sinal, em frequência.

Outro aspecto complicador do tratamento de ruído de fundo é sua ocorrência ao longo de todo o sinal, ao contrário de distúrbios localizados como ruído impulsivo (tema do Capítulo 8). Por isso, sua remoção implica alterar o sinal em toda a sua duração, aumentando as chances de comprometer a qualidade original de uma gravação em consequência do próprio tratamento.

Dentro do conjunto de métodos digitais para restauração de sinais contaminados por ruído de fundo, muitos deles desenvolvidos para realizar processamento de fala e de imagem, este trabalho aborda duas famílias. Na primeira parte,

revisam-se duas técnicas clássicas de fundamentação espectral, faz-se a avaliação de seu desempenho e propõem-se-lhes alternativas vantajosas quanto à complexidade. Na segunda, discute-se um método mais recente, baseado na DWT; investiga-se seu comportamento com sinais musicais reais e contra a variação de diversos parâmetros inerentes ao processamento, com o auxílio do algoritmo de avaliação apresentado no Capítulo 6.

10.1 Métodos Baseados no Espectro

10.1.1 Fundamentos

Sejam $y(k)$, $x(k)$ e $n(k)$ as seqüências correspondentes, respectivamente, ao sinal ruidoso, ao sinal original e ao ruído branco aditivo que definem nosso problema, tais que

$$y(k) = x(k) + n(k). \quad (10.1)$$

Assumindo que $x(k)$ e $n(k)$ são amostras retiradas de processos aleatórios conjuntamente estacionários e não-correlacionados com média zero, o espectro de densidade de potência de $y(k)$ é dado por

$$S_{yy}(\omega) = S_{xx}(\omega) + S_{nn}(\omega). \quad (10.2)$$

As técnicas apresentadas a seguir [36, 16] partem dessas premissas e empregam estimativas para $S_{yy}(\omega)$ e $S_{nn}(\omega)$. A primeira, $\hat{S}_{yy}(\omega)$, é obtida do conhecimento de $y(k)$. A segunda, $\hat{S}_{nn}(\omega)$, pode ser obtida de duas formas: em períodos de silêncio do sinal, quando só há ruído—caso que pressupõe ruído estacionário, embora não necessariamente branco; ou do final do espectro, se valer a hipótese de que nas freqüências altas há pouco conteúdo do sinal—caso que pressupõe ruído aproximadamente branco, embora de variância possivelmente variável no tempo.

10.1.2 Subtração Espectral

Da Equação 10.2, pode-se escrever

$$S_{xx}(\omega) = S_{yy}(\omega) - S_{nn}(\omega), \quad (10.3)$$

que define o algoritmo.

A partir das estimativas $\hat{S}_{nn}(\omega)$ e $\hat{S}_{yy}(\omega)$, é possível, então, obter uma estimativa do espectro de densidade de potência do sinal original $\hat{S}_{xx}(\omega)$, pela operação

$$\hat{S}_{xx}(\omega) = \hat{S}_{yy}(\omega) - \hat{S}_{nn}(\omega). \quad (10.4)$$

Admitindo-se como válidas as hipóteses de estacionaridade e ergodicidade do sinal genérico $p(k)$ durante quadros de K amostras, pode-se estimar (a menos de uma constante) sua densidade espectral de potência $S_{pp}(\omega)$ pelo periodograma $|P(\omega)|^2$ de cada quadro; a versão amostrada desse periodograma pode ser obtida pelo quadrado do módulo da DFT [63, 64], que por uma liberdade de notação, ainda será chamado de $|P(\omega)|^2$.

Assim, a implementação da Subtração Espectral parte da divisão do sinal a tratar em blocos de K amostras com superposição, aplicando-lhes janelas de suavização, para reduzir os efeitos da blocagem abrupta sobre o espectro. Para cada bloco, calcula-se a DFT $Y(\omega)$ do sinal e, com a DFT estimada $\hat{N}(\omega)$ do ruído, obtém-se $|X(\omega)|$ por

$$|\hat{X}(\omega)|^2 = |Y(\omega)|^2 - |\hat{N}(\omega)|^2. \quad (10.5)$$

Estimando a fase de $\hat{X}(\omega)$ pela fase de $Y(\omega)$, calcula-se $\hat{x}(k)$ pela IDFT de $\hat{X}(\omega)$. O sinal é reconstruído respeitando a superposição.

Deve ser notado que, sendo o processamento realizado sobre o espectro de densidade de potência, desconsidera a parcela de informação referente à fase presente no sinal; além disso, o reaproveitamento da fase do sinal corrompido se justifica tão menos quanto maior sua contaminação.

10.1.3 Filtragem de Wiener

Se se deseja obter uma estimativa $\hat{x}(k)$ de $x(k)$ com erro quadrático médio mínimo aplicando $y(k)$ à entrada de um filtro com resposta ao impulso $a(k)$, a solução é o conhecido Filtro de Wiener [18, 19, 25], dado por

$$A(\omega) = \frac{S_{xx}(\omega)}{S_{yy}(\omega)}. \quad (10.6)$$

O Filtro de Wiener não opera sobre a fase do sinal, pois sua função de transferência é real. Reescrevendo-se a equação (10.6) em função da razão sinal-ruído¹ $\text{SNR}(\omega)$, obtém-se

$$A(\omega) = \frac{1}{1 + \frac{1}{\text{SNR}(\omega)}}, \quad (10.7)$$

que pode ser interpretado como um filtro que atua mais fortemente sobre as frequências de baixa Razão Sinal-Ruído. Isso é conveniente psicoacusticamente, aproveitando a propriedade de mascaramento do ruído (ver Subseção 6.2.2).

Novamente, a partir das estimativas $\hat{S}_{nn}(\omega)$ e $\hat{S}_{yy}(\omega)$, é possível, então, obter uma aproximação para $A(\omega)$, dada por

$$\hat{A}(\omega) = \frac{\hat{S}_{yy}(\omega) - \hat{S}_{nn}(\omega)}{\hat{S}_{yy}(\omega)}. \quad (10.8)$$

Fazendo as mesmas considerações da Subseção 10.1.2, a implementação da Filtragem de Wiener envolve agora o cálculo de

$$\hat{A}(\omega) = \frac{|Y(\omega)|^2 - |\hat{N}(\omega)|^2}{|Y(\omega)|^2}, \quad (10.9)$$

que filtra cada quadro de $y(k)$ no domínio da DFT, levando a $\hat{x}(k)$.

É possível realizar uma implementação alternativa do Filtro de Wiener, no tempo. Nesta, apresentada na Figura 10.1, um banco de filtros divide o sinal a tratar em B faixas de frequência suficientemente estreitas para que se possa considerar plano o espectro de densidade de potência em cada uma. Nesse caso, a razão de densidades do fator de Wiener em cada faixa pode ser substituída pela razão das potências correspondentes.

A estimativa da potência \mathcal{P}_p^b para cada sinal $p^b(k)$ na banda genérica b pode ser obtida recursivamente por

$$\mathcal{P}_p^b(k) = \alpha \mathcal{P}_p^b(k-1) + (1-\alpha)|p^b(k)|^2, \quad (10.10)$$

sendo $0 < \alpha < 1$ um fator de esquecimento.

Assim, o Filtro de Wiener para cada banda se reduz a um fator independente da frequência

$$\hat{A}^b(k) = \frac{\mathcal{P}_y^b(k) - \mathcal{P}_{\hat{n}}^b(k)}{\mathcal{P}_y^b(k)}, \quad (10.11)$$

¹Aqui, entende-se a razão sinal-ruído como “razão entre potência originada pelo sinal e potência originada pelo ruído”.

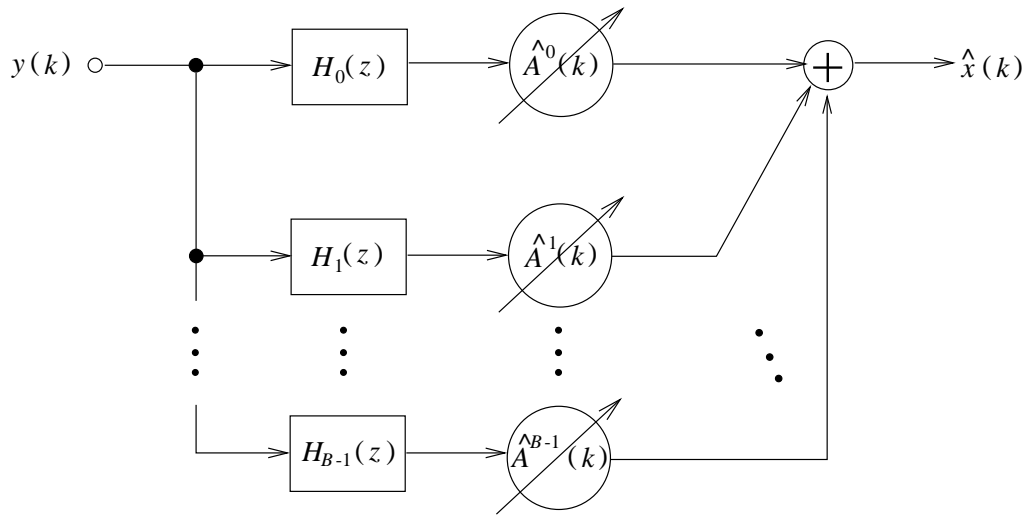


Figura 10.1: Filtro de Wiener implementado no tempo.

requerendo apenas as estimativas das potências médias recentes do sinal ruidoso e do ruído.

O sinal é reconstituído pela soma das amostras processadas nas diversas bandas.

10.1.4 Investigação Realizada

10.1.4.1 Avaliação Inicial de Desempenho

Numa primeira etapa do trabalho referente aos métodos discutidos nas Subseções de 10.1.1 a 10.1.3, implementaram-se testes [52] que aferissem seus desempenhos e dessem noções sobre os parâmetros a empregar num sistema prático.

As simulações foram realizadas sobre sinais discretizados a uma taxa de 44,1 kHz, de três categorias: sinais contaminados artificiais, sinais naturais contaminados artificialmente e sinais contaminados naturais. A qualidade final dos resultados decresceu, nesta ordem. Comparando-se os dois primeiros casos, isso se explica pela maior complexidade dos sinais naturais, dos quais parte do conteúdo espectral (que responde pelos timbres dos instrumentos) pode ser indistinguível do ruído. Comparando-se os dois últimos casos e com base em algumas medições auxiliares, pode-se afirmar que muitas vezes o espectro do ruído de fundo presente nos sinais naturais não pode ser considerado plano; caso não haja como estimá-

lo adequadamente, isso pode restringir a aplicabilidade das técnicas espectrais de restauração.

Nas implementações da Subtração Espectral e da Filtragem de Wiener na frequência, o tamanho escolhido para as janelas (considerando o compromisso entre a hipótese de estacionaridade e a definição em frequência) foi de 2048 amostras com 50% de superposição. Na implementação da Filtragem de Wiener no tempo, o número de sub-bandas adotado (considerando a separação entre elas e a complexidade dos filtros) foi 8, portanto com largura de 2756,25 Hz (correspondendo ao intervalo de um tom da escala musical no extremo superior do espectro).

Entre Subtração Espectral e Filtragem de Wiener na frequência, a segunda (cuja justificativa teórica é mais consistente) tem melhor desempenho. Contudo, sendo ambas baseadas, em última análise, em potência, sofrem do mesmo inconveniente: ignorar completamente o processamento da fase—o que equivale a “creditar boa qualidade” à fase do sinal corrompido. Essa aproximação é tão pior quanto menor a razão sinal-ruído; e, de fato, a redução da razão sinal-ruído leva a resultados progressivamente piores.

Deve ser mencionado que os métodos implementados no domínio da DFT deixam, em regra, um tipo peculiar de resíduo, após o processamento: o chamado ruído musical. Este consiste aproximadamente numa série de componentes senoidais de baixa amplitude e frequências variantes com o tempo, extremamente perturbadoras por introduzirem uma espécie de movimentação tonal aleatória. O principal motivo para isso é o fato de se eliminarem diversas componentes do sinal, tomadas como ruído, entre as quais podem restar outras, isoladas—estas sendo responsáveis pelo efeito auditivo característico. Em [65], por exemplo, tenta-se reduzir o ruído musical através da elevação da parcela a retirar do espectro na Subtração Espectral, de acordo com o limiar de mascaramento auditivo.

Entre as duas formas da Filtragem de Wiener, a implementação em frequência inicialmente resultou comparável à implementada no tempo. O acréscimo de alguns cuidados propostos em [16] (v. Subseção 12.2.1) melhorou seu desempenho.

De qualquer modo, não se pode dizer que nenhum dos métodos tenha sido sempre inteiramente satisfatório perceptivamente, devido às distorções eventualmente introduzidas pelo processamento.

10.1.4.2 Proposições

A partir da avaliação do desempenho dos métodos espectrais de supressão de ruído de fundo, foram implementadas [52] diversas alternativas a eles, as quais são discutidas a seguir, didaticamente separadas.

- **Fator de atenuação da estimativa de ruído**

Com o objetivo de minimizar o resíduo artificial de ruído de fundo característico das técnicas revistas, acrescentou-se um fator de atenuação $\gamma \leq 1$ à estimativa de ruído em ambas as formulações. A idéia é trocar supressão “em excesso” do ruído (com os artefatos decorrentes), causada pela má estimativa das potências, por supressão parcial, em busca de um resíduo mais homogêneo e natural.

- **Separação em oitavas**

A tonalidade na música ocidental é organizada em subdivisões (geometricamente iguais, no sistema temperado) do intervalo de oitava, (compreendido entre as frequências f e $2f$). Desse modo, sem aprofundar a questão psicoacústica (abordada no Capítulo 6), pode-se pensar em “atribuir mesma importância” a intervalos entre frequências que tenham a mesma razão. De acordo com esse raciocínio, faz mais sentido um sistema de processamento de áudio em bandas que compreendam um mesmo intervalo geométrico (oitavas, por exemplo) que em bandas de mesma largura em Hz. Propôs-se, então, visando a melhor distribuição do esforço de processamento, que a divisão do espectro, quando fosse o caso, fosse feita pela separação sucessiva de suas oitavas superiores até restar a banda inferior desejada.

- **Multitaxa**

Considerando a Filtragem de Wiener implementada no tempo por meio de um banco de filtros de análise, o Teorema de Nyquist assegura que se possa trabalhar com taxa reduzida em cada sub-banda. Assim, como forma de relaxar os requisitos de computação do algoritmo original, propôs-se sua alternativa multitaxa—que, além do banco de filtros de análise antes da decimação, exige um banco de filtros de síntese após a interpolação. Como exemplo, a versão multitaxa da estrutura da Fig. 10.1 é mostrada na Fig. 10.2.

- **Tipo de banco de filtros**

Reunindo as idéias da divisão do espectro em oitavas e de um sistema multitaxa, implementou-se a Filtragem de Wiener no tempo por uma estrutura multitaxa

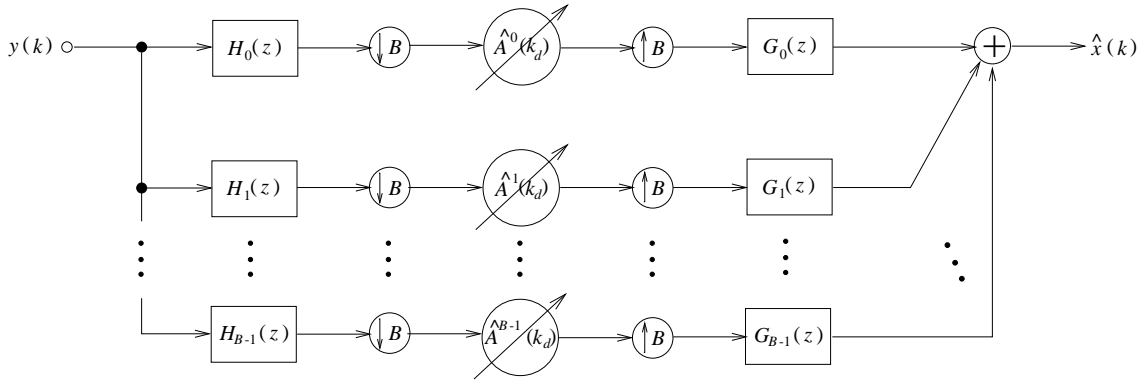


Figura 10.2: Filtro de Wiener implementado no tempo em multitaxa.

de bancos de filtros FIR em árvore binária, como na Fig. 10.3. Compararam-se os

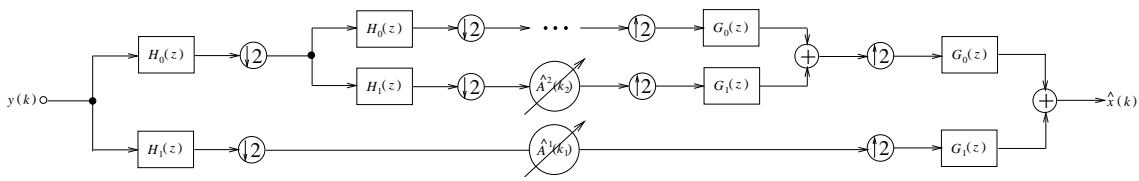


Figura 10.3: Filtro de Wiener implementado no tempo em multitaxa em oitavas sucessivas.

resultados obtidos com dois tipos de bancos: os *Quadrature Mirror Filters* (QMF) e os *Conjugate Quadrature Filters* (CQF) [39, 40].

Os QMF têm fase linear e eliminam *aliasing*, mas não permitem reconstrução perfeita. São obtidos partindo-se de um filtro FIR $H_0(z)$ simétrico de comprimento L par escolhido de forma a aproximar

$$H_0^2(z) - H_0^2(-z) \approx 2z^{-k_0} \quad (10.12)$$

e forçando-se

$$H_1(z) = H_0(-z), \quad (10.13)$$

$$G_0(z) = H_0(z) \quad (10.14)$$

e

$$G_1(z) = -H_0(-z). \quad (10.15)$$

Os CQF se caracterizam por permitir reconstrução perfeita, mas sem possuir fase linear. São obtidos partindo-se de um filtro FIR $H_0(z)$ de comprimento L par tal que

$$H_0(z)H_0(z^{-1}) + H_0(-z)H_0(-z^{-1}) = 2 \quad (10.16)$$

e forçando-se

$$H_1(z) = z^{(L-1)}H_0(-z^{-1}), \quad (10.17)$$

$$G_0(z) = -z^{(L-1)}H_0(z^{-1}) \quad (10.18)$$

e

$$G_1(z) = H_0(-z). \quad (10.19)$$

10.1.4.3 Avaliação das Proposições

Com ambos os algoritmos, o uso do fator de atenuação γ se mostrou, realmente, capaz de melhorar o resíduo percebido do ruído de fundo. Os melhores resultados, dependendo do sinal, foram obtidos com $0,5 \leq \gamma \leq 0,8$. A seguir, analisam-se especificamente as simulações para teste das alternativas propostas para o Filtro de Wiener no tempo.

Para testar a substituição das bandas de mesma largura pelas oitavas sucessivas, o sistema original de teste em 8 faixas de mesma largura descrito na Subseção 10.1.4.1, em que a faixa mais grave ia de 0 a 2756,25 Hz (mais de 7 oitavas a partir de 20 Hz) e a mais aguda vai de 20293,75 a 22050 Hz (menos de um tom na escala musical), teve o banco de filtros reorganizado em 4 faixas, mantendo-se a mais grave e separando-se nas outras as 3 oitavas superiores do espectro (de 2756,25 a 5512,5 Hz, de 5512,5 a 11025 Hz e de 11025 Hz a 22050 Hz). O bom desempenho do algoritmo depende de as faixas de frequência serem suficientemente estreitas para que se possam considerar planos seus espectros. Como as faixas superiores se alargaram, houve, como esperado, algum prejuízo perceptível à qualidade final.

Testes auditivos indicaram que os sistemas em multitaxa preservaram virtualmente a qualidade final dos sinais restaurados, apesar da substancial redução do número de amostras processadas em cada sub-banda pelo seu fator de decimação.

Associando-se o sistema multitaxa e a divisão em oitavas sucessivas, sempre garantindo uma boa separação entre sub-bandas adjacentes e empregando filtros FIR de ordens em torno das dezenas, os resultados das simulações apontaram para o uso preferencial de bancos de QMF, sugerindo que a fase linear pode importar mais ao processamento que a reconstrução perfeita. Como se viu no Capítulo 5, a realização do processamento dos sinais nas sub-bandas já descaracteriza, de certa forma, a idéia de reconstrução perfeita. Porém, mais que isso, os requisitos impostos aos filtros dos bancos de análise e síntese para garantir a reconstrução perfeita podem tornar o sistema muito mais sensível aos desvios causados pelo processamento do que se os filtros tiverem fase linear, ainda que apenas aproximem a reconstrução ao perfeita.

Ainda nesse caso, considerando que a sub-banda inferior era excessivamente larga, invalidando inteiramente a hipótese de espectro plano, pareceu razoável gastar algum processamento adicional para subdividi-la; adotou-se, então, uma rede com 10 sub-bandas, de forma a ter a sub-banda mais grave entre 0 e 43 Hz (contendo, portanto, aproximadamente a oitava inferior do espectro audível). O ganho de qualidade conseqüente foi inferior à perda anteriormente provocada pelo alargamento das sub-bandas superiores, quando se adotou a divisão do espectro em oitavas. Uma possível explicação para isso é que nas sub-bandas inferiores a SNR é maior, pela maior concentração de potência do sinal de áudio; em conseqüência, a supressão de ruído, mesmo que seja mal dimensionada, é muito pequena, não causando tanto prejuízo. Por outro lado, nas sub-bandas superiores, o ganho de Wiener atinge valores mais baixos, ou seja, a supressão é mais atuante; desse modo, sempre que este começa a se elevar, com a estrutura em oitavas, uma banda de ruído equivalente a uma oitava é deixada em torno do sinal cuja presença elevou o fator (e que pode ter banda bastante estreita). O que se ouve, um “véu” de ruído cuja potência acompanha a do sinal, parece corroborar esta explicação.

Como já se disse, os métodos para redução de ruído de fundo aqui avaliados não levam sempre a resultados perceptivos de excelente qualidade, especialmente para sinais fortemente corrompidos. E as alterações aqui propostas buscaram apenas reduzir dramaticamente a carga de processamento sem prejuízo notável da qualidade.

Na busca de um desempenho superior, retorna-se à Filtragem de Wiener no

Capítulo 12, na forma de uma composição de sua implementação na frequência com a estrutura em sub-bandas.

10.2 Método Baseado em *Wavelet*: *Wavelet Shrinkage*

10.2.1 Princípios da DWT

A Transformada Discreta de *Wavelets* (DWT) [66, 24, 23, 67, 68] realiza a representação de uma função contínua $f(t)$ por coeficientes discretos $d_{j,k}$ utilizando como base da transformação versões expandidas e transladadas de uma *wavelet*-mãe. As *wavelets* (pequenas ondas) são funções contínuas com energia concentrada num certo intervalo de tempo, sendo, por isso, apropriadas para análise de fenômenos transitórios, não-estacionários ou variantes no tempo.

No caso geral (dito biortogonal por envolver duas bases duais, mutuamente ortogonais), temos

$$d_{j,k} = \int f(t) \tilde{\psi}_{j,k}(t) dt, \quad (10.20)$$

sendo $\tilde{\psi}_{j,k}(t)$ as *wavelets* (que formam a base) de análise, e

$$f(t) = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} d_{j,k} \psi_{j,k}(t), \quad (10.21)$$

sendo $\psi_{j,k}(t)$ as *wavelets* (que formam a base) de síntese.

A forma mais usual de derivação das funções de base é a diádica, em que

$$\mathcal{F}_{j,k}(t) = 2^{\frac{j}{2}} \mathcal{F}(2^j t - k), \quad (10.22)$$

sendo $\mathcal{F}(t)$ a função-mãe, aqui $\tilde{\psi}(t)$ ou $\psi(t)$. Os inteiros j e k são responsáveis pela localização “na frequência” (correspondendo às escalas) e no tempo (correspondendo aos deslocamentos), respectivamente.

A propriedade da multirresolução é inerente aos sistemas de *wavelets*, e está associada à existência de outra família de funções, chamadas funções de escalamento, tais que

$$c_{j,k} = \int f(t) \tilde{\varphi}_{j,k}(t) dt, \quad (10.23)$$

sendo $\tilde{\varphi}_{j,k}(t)$ as funções de escalamento de análise, e

$$\sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{j_0-1} d_{j,k} \psi_{j,k}(t) = \sum_{k=-\infty}^{\infty} c_{j_0,k} \varphi_{j_0,k}(t), \quad (10.24)$$

sendo $\varphi_{j,k}(t)$ as funções de escalamento de síntese, todas novamente derivadas na forma da equação (10.22). Então, se uma família de sinais é representável por uma soma ponderada de $\varphi_{j_0,k}(t)$, $-\infty < k < \infty$, uma família mais ampla (que inclui a original) pode ser representada por uma soma ponderada de $\varphi_{j_0+1,k}(t)$, $-\infty < k < \infty$. A complementação de um conjunto ao outro se dá pelo acréscimo das *wavelets* da escala j_0 à expansão. Isso permite a representação alternativa

$$f(t) = \sum_{k=-\infty}^{\infty} c_{0,k} \varphi_{0,k}(t) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} \psi_{j,k}(t). \quad (10.25)$$

A Figura 10.4, em que Φ_j e Ψ_j são os subespaços gerados respectivamente por $\varphi_{j,k}(t)$ e $\psi_{j,k}(t)$, ilustra as relações vistas.

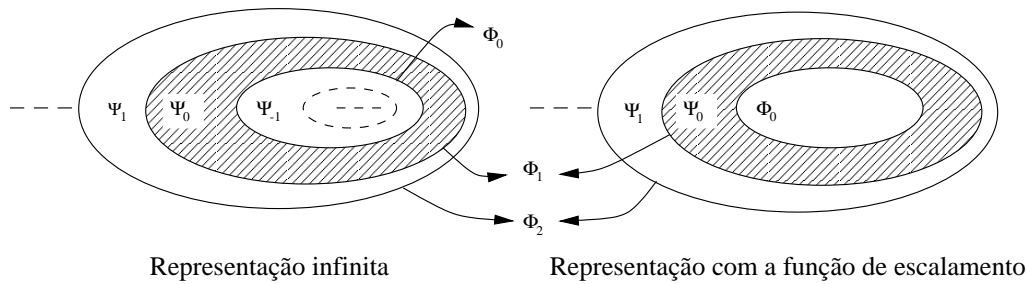


Figura 10.4: Esquema da representação sem e com a função de escalamento.

Caso se possa garantir que o sinal $f(t)$ pode ser representado pela série utilizando apenas um número finito de escalas, digamos, de 0 a J (isto é, que não há mais detalhes do que elas podem representar), pode-se limitar a expansão a

$$f(t) = \sum_{k=-\infty}^{\infty} c_{0,k} \varphi_{0,k}(t) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^J d_{j,k} \psi_{j,k}(t). \quad (10.26)$$

Se o sinal $f(t)$ necessitar de mais escalas do que sua expansão possui, esta apenas o aproximará.

A Figura 10.5 mostra como exemplo o conjunto de funções-mãe de um sistema biortogonal² de *wavelets*.

²A nomenclatura adotada neste Capítulo segue a do *toolbox* correspondente do MatlabTM.

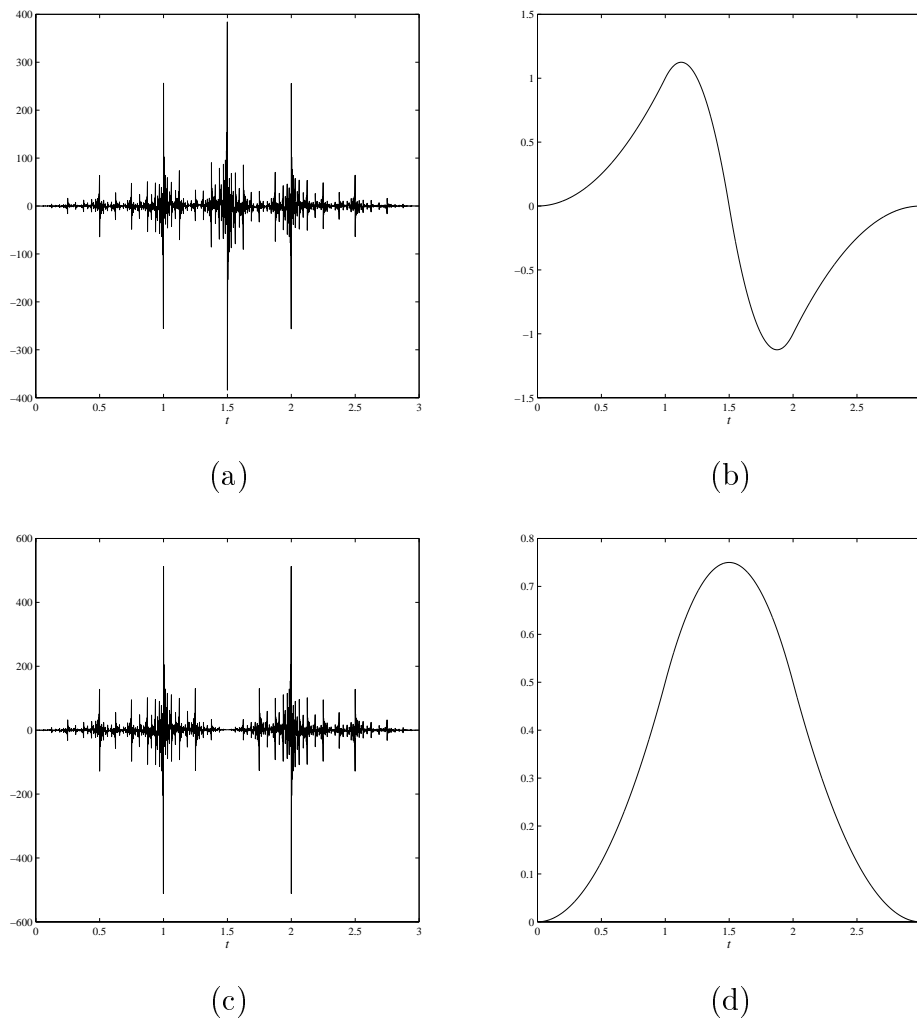


Figura 10.5: Funções-mãe para DWT biortogonal: (a) *wavelet* de análise 3.1 $\tilde{\psi}(t)$; (b) *wavelet* de síntese 3.1 $\psi(t)$; (c) função de escalamento de análise 3.1 $\tilde{\varphi}(t)$; (d) função de escalamento de síntese 3.1 $\varphi(t)$.

As famílias de *wavelets* ortogonais têm as funções de análise e síntese iguais. Foram as primeiras a serem obtidas e se tornaram muito populares. A Figura 10.6 mostra como exemplos os pares de funções-mãe de dois sistemas ortogonais de *wavelets*.

A DWT se tornou bastante utilizada em aplicações como análise, codificação e compressão de sinal e redução de ruído por propiciar um bom compromisso entre localização no tempo e em frequência. A liberdade de escolha de funções de base adequadas à representação compacta de sinais específicos lhe confere generalidade. É desejável, contudo, que seu cálculo seja eficiente (como é o da FFT, por exemplo). Um fato importante nesse aspecto é a possibilidade de calcular recursivamente seus

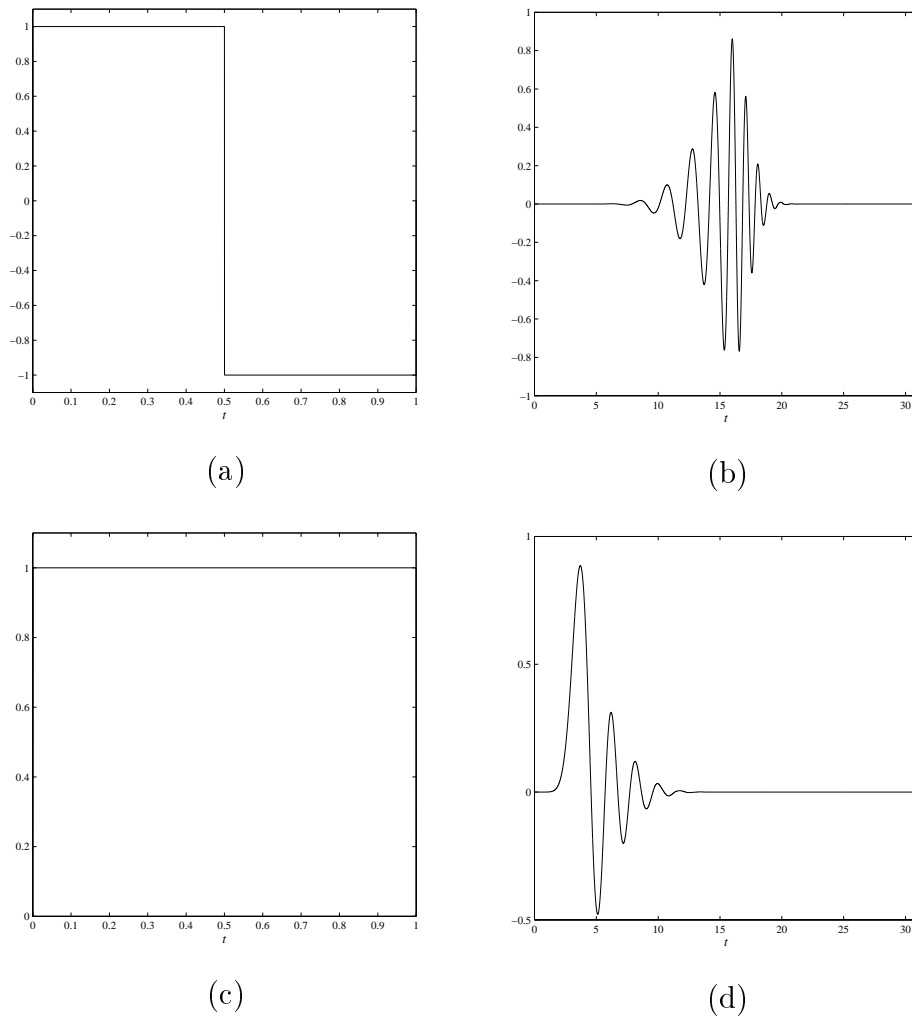


Figura 10.6: Funções-mãe para DWT ortogonal: (a) *wavelet* Haar $\psi(t)$; (b) *wavelet* Daubechies 32 $\psi(t)$; (c) função de escalamento Haar $\varphi(t)$; (d) função de escalamento Daubechies 32 $\varphi(t)$.

coeficientes por uma estrutura binária de bancos de filtros digitais, mostrada na Figura 10.7. Os filtros de análise $h_0(n)$ e $h_1(n)$ são a solução das equações

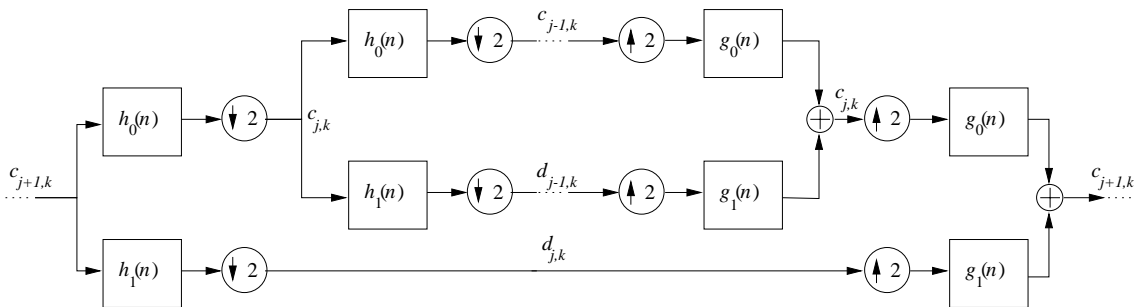


Figura 10.7: Estrutura de banco de filtros em árvore.

$$\varphi(t) = \sum_n h_0(-n)\sqrt{2}\varphi(2t - n), \quad n \in \mathbb{Z} \quad (10.27)$$

e

$$\psi(t) = \sum_n h_1(-n)\sqrt{2}\varphi(2t - n), \quad n \in \mathbb{Z}, \quad (10.28)$$

e os de síntese $g_0(n)$ e $g_1(n)$ são convenientemente calculados para que cada par de bancos de filtros tenha reconstrução perfeita.

A DWT, como foi definida, é uma representação para funções contínuas no tempo. Uma representação de sinais discretos que pudesse se basear na representação em *wavelets* seria muito útil. A passagem do domínio contínuo ao discreto pode ser feita partindo-se do sistema em banco de filtros da Figura 10.7 tendo na entrada os coeficientes $c_{J+1,k}$ calculados para o sinal $f(t)$.

Deve-se notar que, dispondo-se da versão já discretizada $f(kT)$ de $f(t)$, ao se fazer diretamente a igualdade $c_{J+1,k} = f(kT)$, não se obtém a DWT de $f(t)$, mas sim a DWT de um sinal contínuo $\hat{f}(t)$ com representação limitada às escalas de 0 a J , que tem como coeficientes $c_{J+1,k}$ a seqüência $f(kT)$ [69, 70]. É muito comum, contudo, ver-se na literatura o nome DWT dado à transformação direta do sinal discreto nos coeficientes $d_{j,k}$ —e na seqüência deste Capítulo se tomará esta liberdade.

10.2.2 Compressão de Informação Usando *Wavelets*

A utilização de *wavelets* para realizar a compressão de sinais pressupõe que a *wavelet* escolhida consiga, com o menor número possível de coeficientes, representar bem o sinal, de forma que os coeficientes de valores reduzidos possam ser eliminados sem prejuízo significativo dos detalhes representados.

A forma de efetuar essa redução do número de coeficientes utilizados para representar o sinal envolve um procedimento não-linear baseado em limiar, chamado *wavelet shrinkage* [24, 68]. Um limiar é calculado segundo um método coerente e, em seguida, comparado com os coeficientes. Se a amplitude de um coeficiente for pequeno, menor que o do limiar calculado, esse coeficiente é substituído por zero—isto é, descartado, portanto reduzindo a quantidade de informação.

A aplicação do limiar pode ser de duas formas: os coeficientes com amplitude

acima do limiar são mantidos os mesmos no *Hard-Thresholding*, descrito por

$$\theta_h(d_{j,k}) = \begin{cases} 0, & \text{se } |d_{j,k}| \leq \lambda_h \\ d_{j,k}, & \text{se } |d_{j,k}| > \lambda_h \end{cases}, \quad (10.29)$$

ou decrementados do valor do limiar no *Soft-Thresholding*, descrito por

$$\theta_s(d_{j,k}) = \begin{cases} 0, & \text{se } |d_{j,k}| \leq \lambda_s \\ \text{sign}(d_{j,k})(|d_{j,k}| - \lambda), & \text{se } |d_{j,k}| > \lambda_s \end{cases}, \quad (10.30)$$

sendo λ o valor do limiar calculado.

10.2.3 A Redução de Ruído de Fundo

Quando se tenta representar pela DWT sinal contaminado por ruído branco aditivo (para o qual, naturalmente, a *wavelet* escolhida não dá uma representação compacta), este dá uma pequena contribuição dispersa por um número elevado de coeficientes. Assim sendo, uma operação adequada de “compressão” terá o efeito de reduzir o ruído de fundo [71, 72, 73, 74], constituindo uma das técnicas atualmente em voga para diversas aplicações.

Aqui, os limiares são calculados de acordo com uma estimativa do ruído, e podem levar em conta o tamanho do segmento do sinal utilizado. As formas de cálculo dos limiares distinguem as técnicas desse tipo.

10.2.4 Métodos de Cálculo dos Limiares

Considere-se o sinal determinístico $x(k)$ contaminado por um ruído aditivo branco Gaussiano de variância σ^2 , representado por

$$y(k) = x(k) + \sigma u(k), \quad (10.31)$$

onde $u(k)$ é ruído branco Gaussiano de média zero e variância unitária e $y(k)$ é a versão corrompida do sinal.

A aplicação da DWT fornece

$$Y_{j,k} = X_{j,k} + \sigma U_{j,k}, \quad (10.32)$$

sendo $(\cdot)_{j,k}$ a transformada de $(\cdot)(k)$.

Sendo a DWT ortogonal, pode-se provar que $U_{j,k}$ é também branco e Gaussiano de média zero e variância unitária; e pelo teorema de Parseval, pode-se concluir que, se for calculada uma estimativa $\hat{X}_{j,k}$ para $X_{j,k}$ a partir de $Y_{j,k}$, a estimativa $\hat{x}(k)$ para $x(k)$ obtida pela transformação inversa de $\hat{X}_{j,k}$ terá o mesmo erro quadrático que a estimativa no domínio das *wavelets*, ou seja,

$$\|\hat{\mathbf{X}} - \mathbf{X}\|^2 = \|\hat{\mathbf{x}} - \mathbf{x}\|^2. \quad (10.33)$$

Os três métodos a seguir empregam esta formulação [68].

10.2.4.1 Método Minimax e Variante Prática

Pode-se obter um estimador $\hat{x}(k)$ para o sinal $x(k)$ aplicando-se um atenuador $\alpha(j, k)$ (que, de alguma forma, depende da energia do ruído) aos coeficientes $Y_{j,k}$, como em

$$\hat{x}(k) = \text{IDWT}[\alpha(j, k)Y_{j,k}], \quad (10.34)$$

de forma a eliminar o ruído [68].

Nesse caso, a minimização do erro médio quadrático

$$\text{E}[\|\hat{\mathbf{x}} - \mathbf{x}\|^2] = \text{E}[\|\hat{\mathbf{X}} - \mathbf{X}\|^2] \quad (10.35)$$

pode ser realizada pela de cada termo individual $\text{E}[(\hat{X}_{j,k} - X_{j,k})^2]$. Com o auxílio das equações (10.32) e (10.34), pode-se mostrar que

$$\text{E}[(\hat{X}_{j,k} - X_{j,k})^2] = \text{E}[(\alpha(j, k)Y_{j,k} - X_{j,k})^2] = X_{j,k}^2(1 - \alpha(j, k))^2 + \sigma^2\alpha^2(j, k). \quad (10.36)$$

Um estimador muito simples pode ser obtido fazendo-se $\alpha(j, k) \in \{0, 1\}$, isto é, escolhendo-se alguns coeficientes e eliminando outros, segundo regra coerente. A função que minimiza a equação (10.36) é, então,

$$\alpha(j, k) = \begin{cases} 0, & \text{se } X_{j,k}^2 \leq \sigma^2 \\ 1, & \text{se } X_{j,k}^2 > \sigma^2 \end{cases}. \quad (10.37)$$

Como não se tem, a princípio, informação sobre o sinal não-corrompido $x(k)$, essa forma de cálculo não pode ser implementada diretamente, já que requer o conhecimento de $X_{j,k}$.

Um estimador alternativo sub-ótimo prático para $x(k)$ pode ser descrito por

$$\hat{x}(k) = \text{IDWT}[\theta_h(Y_{j,k})], \quad (10.38)$$

empregando o *Hard-Thresholding*, definido na equação (10.29), com $\lambda_h = \hat{\sigma}$ (uma estimativa para σ). Mas seu erro se afasta imprevisivelmente do mínimo.

Em [71] Donoho e Johnstone mostram que, para um sinal com K amostras, o erro do estimador sub-ótimo não ultrapassa o valor do erro do caso ideal multiplicado por um fator $2 \ln K$, se

$$\lambda = \sigma \sqrt{2 \ln K}. \quad (10.39)$$

10.2.4.2 Método SURE (*Stein's Unbiased Risk Estimate*)

Conforme Stein, é possível obter uma estimativa não-polarizada para o sinal, mesmo a partir de uma estimativa polarizada do ruído. Isso permite, então, formular uma estimativa para os coeficientes $X_{j,k}$ como a soma dos coeficientes contaminados $Y_{j,k}$ com uma estimativa $g(Y_{j,k})$ obtida a partir destes para o simétrico do erro, para funções $g(\cdot)$ que atendem a condições bastante amplas:

$$\hat{X}_{j,k} = Y_{j,k} + g(Y_{j,k}). \quad (10.40)$$

Novamente invocando o teorema de Parseval e normalizando a potência do sinal, tem-se para o erro médio quadrático [72]

$$\text{E}[||\hat{\mathbf{x}} - \mathbf{x}||^2] = \text{E}[||\hat{\mathbf{X}} - \mathbf{X}||^2] = K + \text{E}[||\mathbf{g}(\mathbf{Y})||^2 + 2(\nabla \cdot \mathbf{g})(\mathbf{Y})], \quad (10.41)$$

onde K é o número de amostras do sinal ruidoso e $(\nabla \cdot \mathbf{g})(\cdot)$ é o divergente de g em relação a (\cdot) . Aproximando-o pelo valor de uma função-amostra,

$$\text{E}[||\hat{\mathbf{X}} - \mathbf{X}||^2] \approx ||\hat{\mathbf{X}} - \mathbf{X}||^2 = K + ||\mathbf{g}(\mathbf{Y})||^2 + 2(\nabla \cdot \mathbf{g})(\mathbf{Y}). \quad (10.42)$$

Convencionando que há $K_{<\lambda}$ coeficientes abaixo do limiar λ , formando o vetor $\mathbf{Y}_{<\lambda}$:

- se for adotado o *Hard-Thresholding*, definido na equação (10.29),

$$\hat{X}_{j,k} = \theta_h(Y_{j,k}), \quad (10.43)$$

ou seja,

$$g(Y_{j,k}) = \theta_h(Y_{j,k}) - Y_{j,k}, \quad (10.44)$$

e a função a minimizar será

$$\|\hat{\mathbf{X}} - \mathbf{X}\|^2 = K + \|\mathbf{Y}_{<\lambda_h}\|^2 - 2K_{<\lambda_h}. \quad (10.45)$$

- se for adotado o *Soft-Thresholding*, definido na equação (10.30),

$$\hat{X}_{j,k} = \theta_s(Y_{j,k}), \quad (10.46)$$

ou seja,

$$g(Y_{j,k}) = \theta_s(Y_{j,k}) - Y_{j,k}, \quad (10.47)$$

e a função a minimizar será

$$\|\hat{\mathbf{X}} - \mathbf{X}\|^2 = K + \|\mathbf{Y}_{<\lambda_s}\|^2 + (K - K_{<\lambda_s})\lambda_s^2 - 2K_{<\lambda_s}; \quad (10.48)$$

A estimativa SURE [37] consiste em obter o valor de $K_{<\lambda}$ (e, conseqüentemente, λ) que realiza a minimização.

10.2.4.3 Método Híbrido

O método SURE apresentado na Subseção 10.2.4.2 tem seu resultado comprometido em casos que a Razão Sinal-Ruído é muito baixa. Isso ocorre porque o ruído acaba por contribuir muito na forma da função-objetivo a minimizar, com efeito sobre o cálculo do limiar. Como resultado, o erro obtenível se afasta exageradamente do mínimo.

Uma forma encontrada para solucionar tal problema foi a de se alternar a forma de cálculo entre o método SURE quando a razão sinal-ruído for alta e o método Minimax quando a razão sinal-ruído for baixa.

Como critério de decisão para se definir quando é baixa ou alta a razão sinal-ruído, em [37] adota-se uma razão sinal-ruído crítica heurística de valor

$$\text{SNR}_{\text{crit}} = \frac{(\log_2 K)^{3/2}}{\sqrt{K}}. \quad (10.49)$$

A estimativa da razão sinal-ruído necessita da estimativa da varância do ruído, já inerente ao algoritmo, pois necessária à determinação do limiar pelo método Minimax.

10.2.5 Investigação: Avaliação da *Wavelet Shrinkage*

Na literatura, a avaliação dos métodos de redução de ruído baseados em *wavelet* é feita, comumente, utilizando sinais-padrão representativos de características específicas contra as quais se deseja aferir o desempenho dos métodos [37] em conjunto com medidas objetivas de qualidade, como razão sinal-ruído (SNR). Contudo, é imprescindível testar os algoritmos para processamento de áudio sobre os sinais reais, bem mais complexos, a que se destinam. Além disso, apesar de seu claro significado matemático, medidas objetivas usualmente não refletem a qualidade subjetiva dos sinais, enquanto que a avaliação humana requer a realização de uma quantidade inviável de testes sob condições idênticas.

Esta Subseção traça um quadro mais realista do alcance da *Wavelet Shrinkage*. Para tanto, parte de um conjunto representativo de sinais de áudio de alta qualidade (banda larga e dinâmica ampla) corrompidos artificialmente por ruído branco; efetua sua restauração por diversas instâncias do algoritmo com diferentes conjuntos de parâmetros; e emprega a PAQM (medida objetiva de qualidade baseada em Psicoacústica descrita no Capítulo 6) para avaliar os resultados. Além de se reafirmar a aplicabilidade da própria PAQM, obtém-se um bom conjunto de dados sobre o desempenho da *Wavelet Shrinkage* quando aplicada a sinais de alta fidelidade corrompidos e sobre o peso de cada parâmetro testado na qualidade final alcançável.

10.2.5.1 Sinais e Estratégia

Empregou-se nos testes um conjunto de sinais musicais de alta qualidade, cinco deles puramente instrumentais (I1 a I5) e o sexto de voz *a capella*, isto é, sem acompanhamento (V). Sua escolha teve como objetivo cobrir uma série de aspectos relevantes em sinais de áudio reais.

- I1 tem conteúdo espectral rico e variações de potência reduzidas;
- I2 tem espectro concentrado em baixas e altas frequências e variações de potência reduzidas;
- I3 pode ser considerado estacionário por longos intervalos de tempo;

- I4 só pode ser considerado estacionário em intervalos muito curtos de tempo;
- I5 tem conteúdo espectral rico e apresenta variações extremas de potência;
- V visa a explorar os efeitos do processamento em características particulares da voz, como inteligibilidade.

Seqüências de ruído pseudo-branco Gaussiano foram utilizadas para obter versões aditivamente corrompidas (indexadas $\{\cdot\}_c$) de cada sinal, com diferentes valores globais de SNR, sobre as quais foram efetuados os testes. Como referência inicial, seus valores originais de PAQM são dados na Tabela 10.1. Perceptivamente, podem-se considerar os sinais com SNR = 30 dB como fortemente contaminados.

Tabela 10.1: PAQM dos sinais corrompidos.

Sinal	SNR = 30 dB	SNR = 40 dB
I1 _c	0,5236	0,1462
I2 _c	0,2944	0,0610
I3 _c	0,2541	0,0335
I4 _c	0,2249	0,0330
I5 _c	0,2317	0,1247
V _c	0,1845	0,0471

Para cada sinal corrompido, efetuou-se a restauração para diversas versões do algoritmo, combinando diferentes números de escalas, métodos de cálculo e aplicação do limiar e famílias e suportes no tempo das *wavelets*. Cada sinal restaurado foi comparado com sua versão não-corrompida pela PAQM.

10.2.5.2 Medidas

Do extenso conjunto de simulações realizado, apenas um grupo selecionado dos valores da PAQM relativos aos sinais tratados (indexados $\{\cdot\}_d$) é apresentado a seguir, para mais fácil comparação. Didaticamente, optou-se por mostrar em primeiro lugar na Tabela 10.2 os resultados da restauração de cada sinal referido na Tabela 10.1, empregando os parâmetros básicos $J = 2$ (3 escalas), *soft thresholding*, cálculo do limiar pelo método SURE e *wavelet* Daubechies 32. Em seguida, apenas

para os sinais com SNR = 30 dB, cada parâmetro aparece individualmente variado, enquanto os outros são mantidos fixos.

Tabela 10.2: PAQM das versões tratadas de sinais com diferentes valores de SNR ($J = 2$, *soft*, SURE, Daubechies 32).

Sinal	SNR = 30 dB	SNR = 40 dB
I1 _d	0,2118	0,0766
I2 _d	0,1205	0,0373
I3 _d	0,0955	0,0247
I4 _d	0,1076	0,0197
I5 _d	0,2014	0,1170
V _d	0,1439	0,0401

- **Número de escalas**

De acordo com a notação adotada, há $J + 1$ escalas no sistema de processamento, correspondendo a $J + 2$ sub-bandas no diagrama da Fig. 10.7.

A Tabela 10.3 mostra os resultados para $J = 2$ e $J = 7$.

Tabela 10.3: PAQM dos sinais tratados obtidos com diferentes números de escalas (SNR = 30 dB, *soft*, SURE, Daubechies 32).

Sinal	$J = 2$	$J = 7$
I1 _d	0,2118	0,2123
I2 _d	0,1205	0,1205
I3 _d	0,0955	0,0955
I4 _d	0,1076	0,1076
I5 _d	0,2014	0,2018
V _d	0,1439	0,1432

- **Aplicação do limiar**

A Tabela 10.4 compara *soft* com *hard thresholding*.

- **Cálculo do limiar**

Tabela 10.4: PAQM dos sinais tratados obtidos com *soft* e *hard thresholding* (SNR = 30 dB, $J = 2$, SURE, Daubechies 32).

Sinal	<i>Soft</i>	<i>Hard</i>
I1 _d	0,2118	0,4395
I2 _d	0,1205	0,2322
I3 _d	0,0955	0,1897
I4 _d	0,1076	0,1811
I5 _d	0,2014	0,2075
V _d	0,1439	0,1633

A Tabela 10.5 compara os métodos Minimax, SURE e Híbrido de cálculo do limiar.

Tabela 10.5: PAQM dos sinais tratados obtidos por diferentes métodos de cálculo do limiar (SNR = 30 dB, $J = 2$, *soft*, Daubechies 32).

Sinal	Minimax	SURE	Híbrido
I1 _d	0,6546	0,2118	0,2118
I2 _d	0,6592	0,1205	0,1205
I3 _d	0,4699	0,0955	0,0955
I4 _d	0,3741	0,1076	0,1076
I5 _d	0,4256	0,2014	0,2014
V _d	0,5309	0,1439	0,1439

A fim de aferir quanto se afastavam dos limiares ótimos os calculados, fez-se uma busca do multiplicador α que deveria ser aplicado ao limiar calculado para atingir a PAQM mínima. A Tabela 10.6 compara as PAQMs obtidas com os limiares calculados pelo método SURE e com os limiares ótimos e apresenta os valores correspondentes de α .

• **Família e suporte no tempo da *wavelet***

As famílias de *wavelets* costumam ser classificadas segundo dois critérios: o número de momentos desvanescentes (*vanishing moments*) e o tamanho do su-

Tabela 10.6: PAQM dos sinais tratados com limiar SURE e limiar ótimo com α respectivo (SNR = 30 dB, $J = 2$, *soft*, Daubechies 32).

Sinal	SURE	ótimo (α)
I1 _d	0,2118	0,1807 (1,43)
I2 _d	0,1205	0,1202 (1,04)
I3 _d	0,0955	0,0953 (1,06)
I4 _d	0,1076	0,0912 (1,47)
I5 _d	0,2014	0,2009 (1,14)
V _d	0,1439	0,1413 (1,35)

porte temporal da *wavelet*. Quanto maior o número de momentos desvanescentes, mais rapidamente decaem os valores dos coeficientes da transformada em direção às escalas superiores—isto é, uma *wavelet* com um número grande de momentos desvanescentes pode permitir representar com menos escalas um sinal. Quanto menor o tamanho do suporte (tamanho) da *wavelet* no tempo, maior a precisão com que se podem localizar as singularidades presentes no sinal. No entanto, o aumento dos número de momentos desvanescentes aumenta o tamanho do suporte da *wavelet*, o que estabelece um compromisso.

A família de *wavelets* que tem a melhor relação de compromisso entre o número de momentos desvanescentes e o tamanho do suporte é a família Daubechies, já que, dado um número de momentos desvanescentes, a *wavelet* de Daubechies é a que tem o menor tamanho possível de suporte. A Haar, que equivale à Daubechies com 1 momento desvanescente, foi a base dos testes realizados com *wavelets* ortogonais, juntamente com as Daubechies 16, 32 e 64 (com 8, 16 e 32 momentos desvanescentes, respectivamente). Para as *wavelets* biortogonais, escolheram-se dois tipos: a Biortogonal 3.7, que tem funções-mães parecidas com a da Daubechies 32 (v. Fig. 10.8), e a Biortogonal 3.1, que tem características bem diferenciadas entre as funções de análise e síntese (já vistas na Fig. 10.5).

A Tabela 10.7 apresenta os resultados para as diversas *wavelets* [66] citadas.

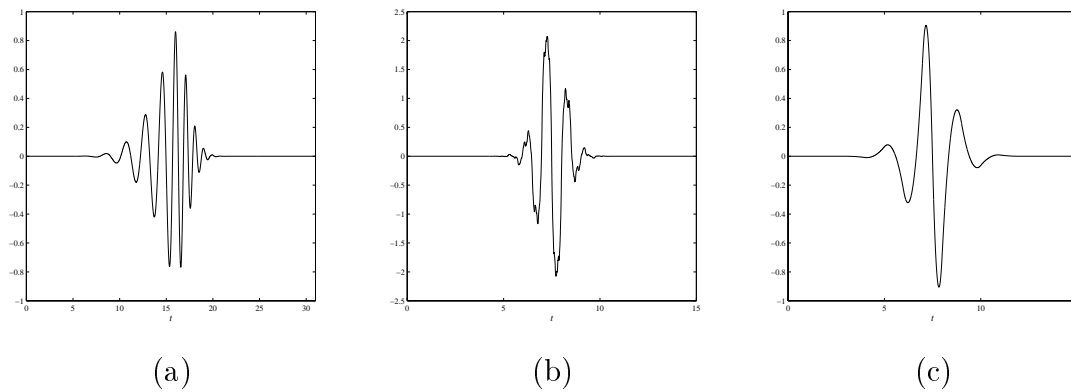


Figura 10.8: Comparação entre as *wavelets* (a) Daubechies 32, (b) Biortogonal 3.7 (análise) e (c) Biortogonal 3.7 (síntese).

Tabela 10.7: PAQM dos sinais tratados obtidos para diferentes famílias e suportes temporais de *wavelet* (SNR = 30 dB, $J = 2$, *soft*, SURE).

Sinal	Haar	Daubechies			Biortogonal	
		16	32	64	3.1	3.7
I1 _d	0,4920	0,2371	0,2118	0,1972	0,3715	0,3133
I2 _d	0,2742	0,1242	0,1205	0,1279	0,1854	0,1514
I3 _d	0,2133	0,1009	0,0955	0,1021	0,1862	0,1303
I4 _d	0,1932	0,1138	0,1076	0,1064	0,1618	0,1309
I5 _d	0,2173	0,2023	0,2014	0,2004	0,2564	0,2223
V _d	0,1556	0,1449	0,1439	0,1432	0,1986	0,1690

10.2.5.3 Conclusões

Já ao primeiro exame, a Tabela 10.1 aponta a incoerência entre a SNR e a avaliação perceptiva, já que sinais distintos com a mesma SNR são pontuados bem diferentemente pela PAQM.

A avaliação subjetiva dos resultados mostrou que a *wavelet shrinkage* nem sempre tem desempenho aceitável sobre sinais de alta qualidade contaminados por ruído de fundo, especialmente aqueles com baixa SNR. Este fato é confirmado pela comparação entre as medidas das 2 colunas da Tabela 10.2: a PAQM final é pior (mais positiva) para os sinais originalmente mais contaminados (SNR = 30 dB em contraste com SNR = 40 dB).

Um efeito perceptível bastante comum após o processamento é que o ruído original, relativamente homogêneo e de alta amplitude, dá lugar a um resíduo de baixa amplitude, mas extremamente perturbador por sua característica variante no tempo.

Vozes e sinais com grandes variações de potência são de mais difícil tratamento, enquanto que o contorno espectral e a estacionaridade não se mostraram fatores críticos. Coerentemente, a comparação entre as Tabelas 10.2 e 10.1 indica que os sinais de I1 a I4 tiveram uma melhora de qualidade bem superior à de I5 e I6.

O algoritmo PAQM adaptado emulou bastante bem a opinião subjetiva, embora se deva ter em mente que ele mede a **diferença** perceptível entre o sinal sob teste e sua versão original, e não diretamente a **falta de qualidade** do primeiro.

O aumento do número de escalas teve pouco efeito no desempenho, após $J = 2$ (v. Tabela 10.3).

Como esperado, *soft thresholding* levou a melhores resultados que *hard thresholding* (v. Tabela 10.4).

Minimax é tido como um método melhor para baixa SNR. O fato de que o cálculo SURE se mostrou melhor em todas as simulações e levou aos mesmos resultados que o método Híbrido (v. Tabela 10.5) quer dizer que os 30 dB ainda estão sendo considerados uma SNR alta, diante de SNR_{crit} . Mesmo em testes adicionais sobre uma versão de $I1_c$ com $SNR = 20$ dB (para áudio, uma SNR extremamente baixa em termos práticos), os métodos Minimax e Híbrido não atingiram a eficácia do SURE. De acordo com a Tabela 10.6, os limiares ótimos encontrados foram sempre maiores (entre 5 e 50 %) que os calculados, embora levando em quase todos os casos a melhora muito pequena medida pela PAQM.

Dentre as famílias de *wavelets*, a Daubechies 32, de forma suave e próxima à de um trecho típico de áudio (permitindo representação mais compacta), atingiu o melhor desempenho, sem requerer suporte mais longo no tempo; e a Biorthogonal similar em forma à Daubechies 32 apresentou resultados próximos (v. Tabela 10.7). Ficou evidente que a forma da *wavelet* (de síntese, no caso biortogonal) desempenha o papel fundamental na qualidade do processamento.

10.3 Comentário Final

É claro que se podem trazer para uma formulação comum os algoritmos baseados no espectro e em *wavelet*, ambos passíveis de implementação em sub-bandas e empregando a redução ou eliminação de coeficientes na medida de sua contaminação pelo ruído. Mas é importante distinguir as diferentes filosofias de base. A Subtração Espectral e a Filtragem de Wiener dependem da divisão do espectro em faixas de frequência o mais possível estreitas e fortemente separadas. A *wavelet shrinkage* depende muito mais da “similaridade” entre as funções-mães e o sinal a representar, resultando em compactação e conseqüente eficácia na eliminação do ruído representado nos coeficientes abaixo do limiar.

Outros métodos recentes orbitam em torno dos descritos neste Capítulo, como [75] (que itera o processamento por *wavelets*), [76] (que faz uma translação nas *wavelets*), [77] (baseado em *wavelet packets*) e [78].

Capítulo 11

Supressor Adaptativo de Ruído

Quando, para uma gravação que se deseja restaurar, se dispõe de múltiplas cópias que possam ser utilizadas como matrizes, contaminadas por ruídos não-correlacionados, é possível aplicar um **supressor adaptativo de ruído**. Uma situação prática comum é a disponibilidade de uma única matriz mono; nesse caso, ainda, sua leitura por um sistema estéreo pode fornecer 2 matrizes para as quais a hipótese de não-correlação ainda é amplamente satisfeita.

Neste Capítulo, após uma breve revisão do supressor, apresenta-se a investigação do assunto realizada neste trabalho, em particular o esboço de algumas expectativas teóricas sobre seu desempenho.

11.1 Filtragem Adaptativa

Filtros digitais de coeficientes fixos [63] podem ser ineficientes no processamento de sinais com propriedades estatísticas desconhecidas e/ou não estacionárias (como os sinais de áudio, em geral). Os filtros adaptativos [31, 79] são uma alternativa extremamente versátil com aplicações em múltiplas áreas, como controle e telecomunicações.

A Figura 11.1 mostra o esquema geral de um filtro adaptativo. O sinal de entrada $x(k)$ é processado de acordo com os coeficientes do filtro, que são recalculados ao longo do tempo de forma a minimizar uma função do erro $e(k)$ entre o sinal de saída $y(k)$ e um sinal de referência (desejado) $d(k)$. Diferentes funções-objetivo levam a diferentes algoritmos de adaptação.

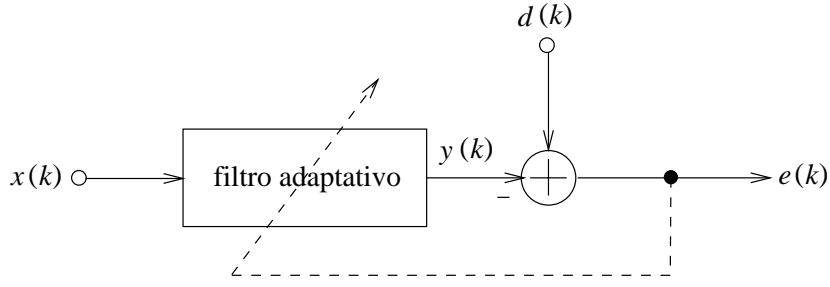


Figura 11.1: Filtro adaptativo genérico.

Os filtros adaptativos FIR, cujo processamento se dá segundo

$$y(k) = \mathbf{w}^T(k)\mathbf{x}(k), \quad (11.1)$$

sendo

$$\mathbf{x}(k) \triangleq [x(k) \quad x(k-1) \quad \dots \quad x(k-M)]^T, \quad (11.2)$$

$$\mathbf{w}(k) \triangleq [w_0(k) \quad w_1(k) \quad \dots \quad w_M(k)]^T \quad (11.3)$$

e M sua ordem, são mais largamente utilizados que os IIR por serem incondicionalmente estáveis, embora requeiram ordens mais elevadas para atender a especificações mais exigentes. Os algoritmos mais populares para filtros FIR são o LMS (*Least-Mean-Square*), que minimiza o erro quadrático instantâneo $e^2(k)$, e o RLS (*Recursive Least-Squares*), que minimiza uma média dos erros ponderada no tempo, $\sum_{i=0}^k \lambda^{k-i} e^2(k)$. Para sinais estacionários, ambos tendem, na média, ao filtro ótimo de Wiener

$$\mathbf{w}_o = \mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{p}, \quad (11.4)$$

onde

$$\mathbf{R}_{\mathbf{xx}} \triangleq E[\mathbf{x}(k)\mathbf{x}^T(k)] \quad (11.5)$$

e

$$\mathbf{p} \triangleq E[d(k)\mathbf{x}(k)], \quad (11.6)$$

obtido pela minimização do erro médio quadrático $E[e^2(k)]$.

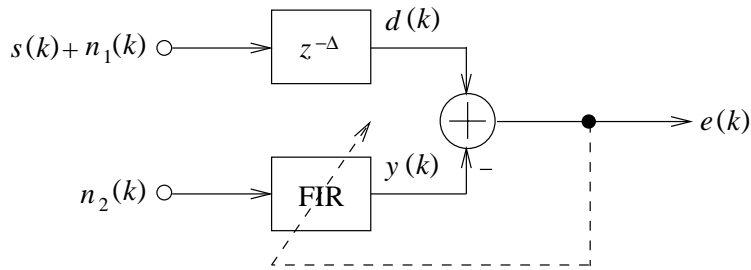


Figura 11.2: Cancelador adaptativo de ruído.

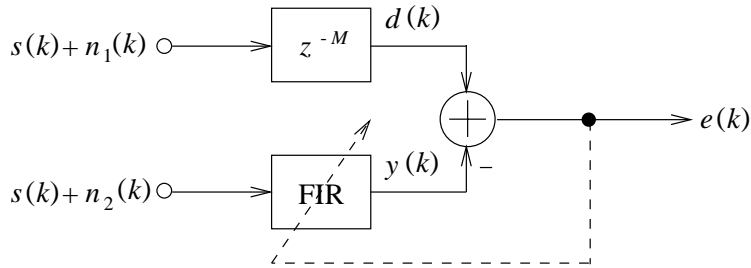


Figura 11.3: Supressor adaptativo de ruído.

11.2 Supressor Adaptativo de Ruído

Uma estrutura comumente usada para recuperar um sinal $s(k)$ contaminado por ruído aditivo $n_1(k)$ quando se dispõe do ruído $n_2(k)$ correlacionado com o primeiro (como para extrair interferência da rede elétrica sobre eletrocardiogramas) é o cancelador adaptativo de ruído (ANC) [36], mostrado na Figura 11.2. Com a convergência, espera-se que $y(k) \rightarrow n_1(k - M)$ e, portanto, $e(k) \rightarrow s(k - M)$.

Nosso problema de restauração, quando se dispõe de 2 matrizes sincronizadas do sinal $s(k)$ contaminadas respectivamente por ruídos aditivos $n_1(k)$ e $n_2(k)$ não-correlacionados entre si e com $s(k)$, leva à estrutura modificada do supressor adaptativo de ruído (ANS), mostrado na Figura 11.3. A idéia é que, na tentativa de minimizar o erro, o filtro tenda a subtrair a parte do sinal de referência correlacionada com o sinal de entrada e, na convergência, se tenha $y(k) \rightarrow s(k - M)$.

11.3 Investigação Realizada

11.3.1 Testes Práticos sobre Sinais Reais

Avaliou-se o desempenho do ANS [61] sobre gravações musicais reais de 2 categorias que garantiam a sincronização: duas versões do mesmo sinal contaminadas respectivamente por dois ruídos aditivos artificiais não-correlacionados; e dois sinais provenientes da leitura estéreo de uma matriz mono naturalmente contaminada.

De um modo geral, sob avaliação subjetiva, os resultados se mostraram apenas regulares. Em particular, a eventual ocorrência de distúrbios correlacionados nos sinais naturalmente contaminados compromete, nessas regiões, a eficácia do tratamento.

Os filtros FIR capazes de realizar o processamento são em geral muito longos, de ordem superior a 1000.

A comparação entre o LMS e o RLS acusou a superioridade do segundo à custa de sua maior complexidade. Grande cuidado requer a sintonia da velocidade de adaptação dos algoritmos por seus respectivos parâmetros (como o passo de adaptação μ do LMS e o fator de esquecimento λ do RLS), que precisa ser feita caso a caso, comprometendo o automatismo. A adaptação deve ser rápida o suficiente para acompanhar os detalhes do sinal e lenta o suficiente para suprimir os distúrbios. Um grau de eficácia aceitável na restauração leva, em geral, a alguma suavização do sinal. As características não-estacionárias dos sinais poderiam ser melhor acompanhadas por algoritmos com passo variável [80, 81, 82].

11.3.2 Expectativas de Desempenho

Esta Seção [61] tenta obter analiticamente algum *insight* sobre o desempenho do ANS pela comparação de suas relações sinal-ruído (SNR) de entrada e saída, já que a solução de Wiener está associada à minimização da sinal-ruído.

Nesse caso, estamos definindo a sinal-ruído como a razão entre a potência originada do sinal e a potência originada do ruído. Assim, em se tratando da saída de um filtro, deve-se ter em mente que a medida nada diz da distorção eventualmente sofrida pelo sinal de entrada.

Além disso, a SNR não é um bom avaliador da qualidade do áudio em termos

perceptivos (v. Seção 6.1).

Inicialmente, é possível provar que, dados dois sinais

$$x_1(k) = s(k) + n_1(k) \quad (11.7)$$

e

$$x_2(k) = s(k) + n_2(k), \quad (11.8)$$

sendo $n_1(k)$ e $n_2(k)$ processos aleatórios de média zero não-correlacionados entre si e com $s(k)$, a simples obtenção da média de $x_1(k)$ e $x_2(k)$ resulta num sinal com SNR maior que a de cada um dos sinais originais, desde que a razão entre estas esteja entre $\frac{1}{3}$ e 3. Isso pode ser tomado como uma expectativa mínima em relação a um processamento sofisticado como o realizado pelo ANS.

Agora, relativamente ao ANS, façamos a hipótese de que a potência $\sigma_{n_1}^2$ de $n_1(k)$ é maior que a potência $\sigma_{n_2}^2$ de $n_2(k)$ e, portanto, a SNR de $x(k) = s(k) + n_2(k)$, SNR_x , é maior que a SNR de $d(k)$, SNR_d . Com isso, a avaliação de desempenho pode ser obtida pela simples comparação da SNR de $x(k)$ com a SNR de $y(k)$, SNR_y , que (como se verá) independe de $\sigma_{n_1}^2$. Definam-se

$$\mathbf{s}(k) \triangleq [s(k) \quad s(k-1) \quad \dots \quad s(k-M)]^T, \quad (11.9)$$

$$\mathbf{R}_{ss} \triangleq E[\mathbf{s}(k)\mathbf{s}^T(k)], \quad (11.10)$$

$$\mathbf{n}_2(k) \triangleq [n_2(k) \quad n_2(k-1) \quad \dots \quad n_2(k-M)]^T \quad (11.11)$$

e

$$\mathbf{R}_{n_2n_2} \triangleq E[\mathbf{n}_2(k)\mathbf{n}_2^T(k)]. \quad (11.12)$$

A potência de saída para \mathbf{w} fixo é

$$E[y^2(k)] = E \left[\left(\sum_{m=0}^M w_m (s(k-m) + n_2(k-m)) \right)^2 \right] = \mathbf{w}^T \mathbf{R}_{ss} \mathbf{w} + \mathbf{w}^T \mathbf{R}_{n_2n_2} \mathbf{w}. \quad (11.13)$$

Com isso, a SNR na saída vale

$$\text{SNR}_y = \frac{\mathbf{w}^T \mathbf{R}_{ss} \mathbf{w}}{\mathbf{w}^T \mathbf{R}_{n_2n_2} \mathbf{w}}. \quad (11.14)$$

O vetor de coeficientes ótimo, de acordo com a equação (11.4), é

$$\mathbf{w}_o = \mathbf{R}_{xx}^{-1} E[(s(k-M) + n_1(k-M))(s(k) + \mathbf{n}_2(k))] = \mathbf{R}_{xx}^{-1} \mathbf{r}_{ss}(:, M), \quad (11.15)$$

sendo $\mathbf{r}_{ss}(:, M)$ a M -ésima coluna de \mathbf{R}_{ss} . Substituindo (11.15) em (11.14), obtém-se

$$\text{SNR}_y = \frac{\mathbf{r}_{ss}^T(:, M) (\mathbf{R}_{xx}^{-1})^T \mathbf{R}_{ss} \mathbf{R}_{xx}^{-1} \mathbf{r}_{ss}(:, M)}{\mathbf{r}_{ss}^T(:, M) (\mathbf{R}_{xx}^{-1})^T \mathbf{R}_{n_2 n_2} \mathbf{R}_{xx}^{-1} \mathbf{r}_{ss}(:, M)}. \quad (11.16)$$

A expressão não é de fácil comparação com

$$\text{SNR}_x = \frac{\sigma_s^2}{\sigma_{n_2}^2}. \quad (11.17)$$

Assim sendo, lançaremos mão de alguns casos particulares representativos, que permitam interpretar melhor esse resultado.

Chamando o processo AR tendo filtro gerador associado com pólo em 0,9 de sinal tipo A e o ruído branco de sinal tipo B, a Tabela 11.1 ilustra os resultados obtidos para algumas combinações de $s(k)$ e $n_2(k)$.

Tabela 11.1: ANS: Comparação de SNR_x e SNR_y para alguns casos simples.

$s(k)$	$n_2(k)$	SNR_x	SNR_y
A	A	1	1
A	A	10	10
B	B	1	1
B	B	10	10
A	B	1	1,886
A	B	10	15,095

Pode-se inferir dos resultados que:

- a eficácia do ANS depende dos tipos de sinal envolvidos;
- os resultados são tão melhores quanto mais distintos forem $s(k)$ e $n_2(k)$, do ponto de vista estatístico;
- o aumento percentual na SNR é tão maior quanto menor a SNR original.

Isso permite afirmar que, ainda que sinal e ruído sejam não-correlacionados, é mais difícil suprimir ruído de sinal com modelo próximo (uma dependência um tanto

surpreendente para sistemas adaptativos), bem como prever que não é possível realizar processamento muito refinado. Em todo caso, utilizou-se aqui como parâmetro de avaliação a sinal-ruído, que não reflete necessariamente qualidade subjetiva (v. Capítulo 6).

Cumprir observar que esse é um resultado apenas preliminar, estando longe de esgotar o assunto. O uso de configurações adaptativas mais elaboradas por si só já pode conduzir a resultados superiores.

Capítulo 12

Avanços nos Métodos em Sub-bandas

Neste Capítulo, o único não-autocontido deste trabalho, retorna-se ao tratamento de ruído impulsivo e ao tratamento de ruído de fundo para apresentar novas propostas para sua realização em sub-bandas. No caso do ruído impulsivo, busca-se avançar no caminho aberto na Seção 8.5 quanto ao processamento baseado em modelo implementado em sub-bandas, já à luz da discussão de modelos do Capítulo 4. Para o ruído de fundo, propõe-se a realização da filtragem de Wiener no domínio da frequência (v. Subseção 10.1.3) em cada oitava do espectro do sinal.

12.1 Tratamento de Ruído Impulsivo Baseado em Modelo Realizado em Sub-bandas Revisitado

No Capítulo 8, reviu-se o tratamento de ruído impulsivo com a detecção realizada por inversão do modelo AR e a recomposição das amostras afetadas através de interpolação e apresentou-se sua versão aperfeiçoada, de desempenho muito superior. Lá também se discutiu uma implementação alternativa em sub-bandas do algoritmo modificado. Na forma descrita, o método consiste em: subdividir o sinal a processar em sub-bandas separando cada oitava superior do espectro seqüencialmente através de bancos de filtros multitaxa, como mostrado na Figura 12.1; aplicar em

cada sub-banda o algoritmo de restauração de banda completa, transposto diretamente; reconstituir o sinal restaurado pelos respectivos bancos de filtros de síntese.

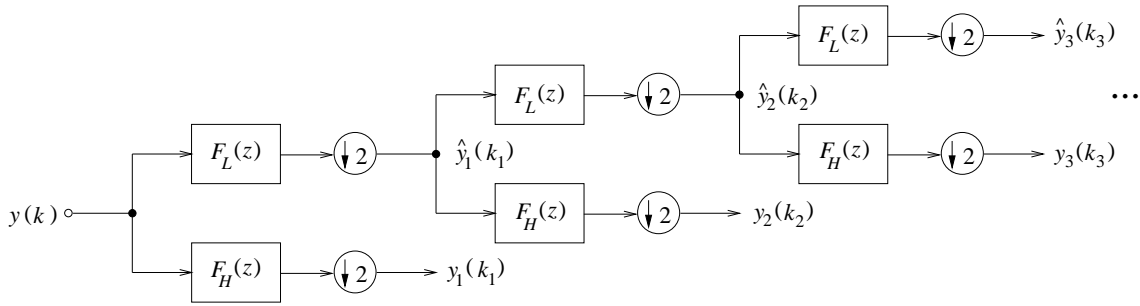


Figura 12.1: Processo autorregressivo analisado em oitavas por banco de filtros.

As hipóteses feitas quanto à estacionaridade dos sinais na estrutura proposta já foram comentadas. Entretanto, a viabilidade de se adotarem modelos AR independentes para as sub-bandas de forma a realizar o processamento em cada uma com o mesmo algoritmo empregado em banda completa foi endossada apenas pelo exame de uma série de medidas realizadas sobre sinais reais [55], tendo sido deixada de lado, naquele ponto, sua verificação analítica. O trabalho realizado no Capítulo 4 (apresentado na Parte I por conveniência de organização) se originou da necessidade de realizá-la. É da reunião da modelagem do Capítulo 4 com a implementação da Seção 8.5 que trata esta Seção.

De acordo com a modelagem derivada nas Seções de 4.3 a 4.5, os sinais nas sub-bandas da estrutura proposta admitem modelos convencionais exatos na taxa reduzida. Na forma completa, cada par de sub-bandas de mesma taxa pode ser gerado a partir de um par de sinais de ruído branco mutuamente não-correlacionados (sinais de excitação) apresentados como entradas de um banco de 4 filtros IIR, 2 diretos e 2 cruzados; o sinal em cada sub-banda é a soma das saídas dos respectivos filtros direto e cruzado.

O estudo apresentado na Seção 4.7 mostrou de que forma se comportam os zeros e pólos dos filtros geradores associados ao modelo, que vão sendo progressivamente cancelados e interiorizados ao longo das subdivisões do espectro, inteiramente de acordo com o que se poderia esperar da filtragem encadeada. Isso pode possibilitar a gradual redução das ordens dos modelos nas sub-bandas. Mais que isso,

é possível, dentro de certos limites, substituir os filtros geradores IIR por aproximações só-pólos, o que simplifica a inversão do modelo e se aproxima da estrutura anteriormente proposta.

Mas o que dizer da interrelação entre as sub-bandas? É razoável desprezá-la e processar independentemente os subsinais? Pode-se dizer que o problema causado por desprezar a correlação entre sub-bandas será tão maior quanto mais impreciso for o esquema de detecção e interpolação adotado em cada uma. Obviamente, se a restauração de cada subsinal fosse realizada idealmente, a correlação entre os subsinais originais seria, por assim dizer, implicitamente preservada. É justamente a imperfeição prática dos processamentos individuais que pode dar origem a incoerências entre os resultados, com o conseqüente surgimento de artefatos audíveis nas regiões tratadas do sinal. É freqüentemente citado e explorado o fato de que a DWT pode resultar em coeficientes não-correlacionados, dependendo da escolha das funções-base [83, 84]¹. Como se podem obter os coeficientes da DWT através de uma estrutura em bancos de filtros, pode-se deduzir que a escolha adequada dos bancos de filtros pode ter o efeito de descorrelacionar os sinais nas sub-bandas, podendo favorecer seu processamento independente.

Em síntese, o esquema da Seção 8.5 se justifica, certamente, dentro das limitações acima. Fica, contudo, a impressão de que alguma sofisticação adicional pode favorecer o desempenho do método em sub-bandas.

A primeira possível medida é empregar filtros IIR com zeros, ainda que de ordem reduzida.

A segunda é não abrir mão dos filtros cruzados. Mas como eles poderiam ser levados em conta? Pode-se pensar num esquema recursivo que execute o processamento a cada estágio de divisão do espectro. O modelo para os sinais após o m -ésimo estágio seria como mostra a Figura 12.2. Uma estratégia seria empregada para modelar conjuntamente os 4 filtros. Com base no modelo integrado, seria feita a localização dos distúrbios e a interpolação de amostras em substituição àquelas dadas como corrompidas em ambos os sinais, simultaneamente. O procedimento seria repetido a cada redivisão do sinal até ser atingida a sub-banda inferior.

Esse método garantiria a coerência dos sinais parciais após a restauração.

¹Essa afirmativa só é aceitável sob condições aproximadamente estacionárias.

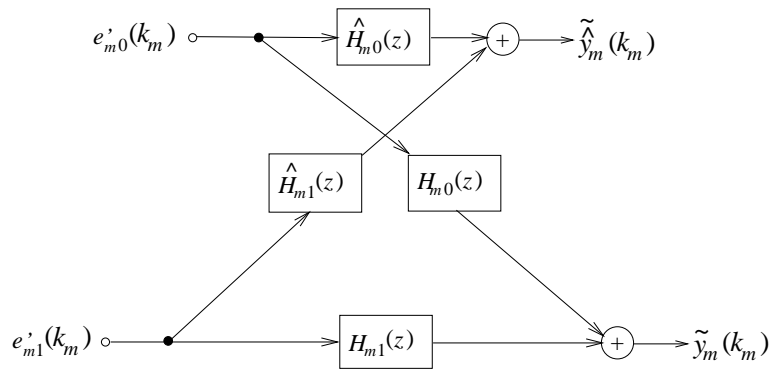


Figura 12.2: Modelo completo para os sinais decorrentes do m -ésimo estágio da divisão do sinal $s(k)$ em oitavas.

A implementação e os testes deste algoritmo serão efetuados na continuação deste trabalho.

É preciso ter em mente algumas ressalvas quanto ao processamento de ruído impulsivo em sub-bandas. As seguidas filtragens do sinal reduzem o fator de qualidade Q associado aos pólos dos filtros geradores associados aos modelos, o que dificulta a obtenção de modelos acurados, especialmente na forma autorregressiva. Informações provenientes de muito baixas frequências, então, podem vir a ser pior modeladas, por serem “separadas” somente após um número razoável de divisões do espectro, podendo afetar os trechos interpolados nas bandas inferiores. Da mesma forma, os impulsos a detectar vão sendo alongados (apesar das decimações) e suavizados no tempo por essas mesmas filtragens, o que pode acabar dificultando sua evidenciação no sinal de excitação, nas bandas inferiores. Entretanto, uma dos aspectos característicos desse esquema aplicado ao ruído impulsivo é justamente a expectativa de concentrar a maior parte da restauração nas bandas superiores, menos afetadas pelos fatores mencionados, onde há predomínio dos distúrbios sobre o sinal. De qualquer forma, pode ser necessário adotar uma estratégia adicional para lidar com a dispersão dos pulsos.

12.2 Tratamento de Ruído de Fundo por Filtragem de Wiener Baseada na DFT numa Estrutura em Oitavas

12.2.1 Filtragem de Wiener Baseada na DFT

Apresentado na Subseção 10.1.3, o método consiste em dividir o sinal em blocos com superposição e, a cada bloco, aplicar uma janela de suavização, calcular a DFT, aplicar o filtro de Wiener estimado e calcular a IDFT.

Em geral, a janela de Hanning é utilizada, juntamente com superposição de 50%. Deixando de lado a questão da estimativa do espectro do ruído, o grande fator limitante da implementação da filtragem de Wiener baseada na DFT é o tamanho do bloco.

Sendo o sinal variante no tempo, para que o espectro calculado seja representativo, a janela não pode ser longa demais; dito de outra forma, é preciso manter uma razoável localização no tempo. O limite superior adotado costuma ser de cerca de 50 ms a intervalos de 25 ms [85, 16]. Na freqüência, isso corresponderia a dividir o espectro em componentes igualmente espaçadas de 20 Hz. A partir do final do espectro audível (20 kHz) para baixo, isso corresponde a 1,7% de semitom, uma diferença absolutamente imperceptível; do início do espectro audível (20 Hz) para cima, entretanto, isso corresponde a toda a primeira oitava. Em outras palavras, a escolha refina excessivamente as freqüências altas e é grosseira para as freqüências baixas. Isso certamente pode contribuir para as distorções pouco naturais causadas no sinal pelo método.

Considerando janelas de N amostras com 50% de superposição, a complexidade do algoritmo fica em torno de $9 + 3 \log_2 N$ multiplicações por amostra.

Valores práticos: taxa de amostragem de 44,1 kHz, espectro útil de 0 a 22,05 kHz; janelas de 44,6 ms, $N = 2048$; componentes espaçadas de 21,5 Hz, ou seja, 1,6% de semitom no final do espectro e 1 oitava no início; 42 multiplicações por amostra.

Algumas modificações [16] podem melhorar o desempenho da técnica, com impacto desprezível sobre a complexidade computacional: usar a janela de suavi-

zação também antes da IDFT, para tornar ainda mais naturais as transições entre quadros; adotar um limite inferior positivo para o fator de atenuação de Wiener—v. equação (10.9)—aplicado às componentes da DFT, deixando um “chão” de ruído mais homogêneo que o resultante da supressão total; e usar como estimativa do espectro do sinal para cálculo do fator de Wiener a mediana dos espectros de quadros adjacentes, suavizando a estimativa.

12.2.2 Filtragem de Wiener diretamente nas Sub-bandas

Apresentado na Subseção 10.1.3 e já incorporando algumas de nossas proposições (v. Subseção 10.1.4.2), o método consiste em: separar o sinal em oitavas sucessivas ao longo de M estágios através de um banco de filtros de análise organizado em árvore binária maximamente decimado; multiplicar cada amostra de cada sub-banda pelo fator de Wiener—v. equação (10.11)—conveniente; recompor o sinal em banda completa pelo respectivo banco de filtros de síntese.

Em geral, utilizam-se bancos de filtros FIR de ordem Q . Deixando novamente de lado a questão da estimativa do espectro do ruído, o grande fator limitante da implementação da filtragem de Wiener diretamente nas sub-bandas é a largura das sub-bandas.

Como se viu, a divisão do espectro em oitavas (digamos, nas 10 faixas de 0 a 40 Hz, 40 a 80Hz, 80 a 160 Hz,... e 10 a 20 kHz) teve a intenção de detalhar mais homogêneo o espectro do que permite a divisão em bandas de mesma largura [36], que favorece as altas freqüências em detrimento das baixas, mantendo o número de sub-bandas dentro de limites razoáveis. Entretanto, a hipótese de espectro plano ao longo de uma oitava é extremamente permissiva, o que certamente contribui para o “véu” de ruído que permanece após o processamento, e cuja intensidade acompanha aproximadamente a potência do sinal.

Considerando os parâmetros enunciados e a realização da pré- e da pós-filtragem na FFT, a complexidade do algoritmo fica em torno de $5 + \sum_{m=1}^M \frac{16}{2^m} \log_2 Q$ multiplicações por amostra.

Valores práticos: taxa de amostragem de 44,1 kHz, espectro útil de 0 a 22,05 kHz; 9 estágios sucessivos de divisão do espectro ao meio, ou seja, primeira sub-banda de 0 a 39,0625 Hz e última de 11,025 a 22,05 kHz; filtros de ordem 63;

12.2.3 Composição Proposta

12.2.3.1 Descrição

Objetivando alcançar um compromisso entre a baixa complexidade da FFT e a mais bem distribuída divisão do espectro em oitavas, imaginou-se um algoritmo híbrido em que a versão baseada na DFT da filtragem de Wiener fosse aplicada a cada uma das oitavas do espectro do sinal a processar, numa estrutura empregando bancos de filtros em árvore binária, como mostrado no diagrama da Fig. 12.3.

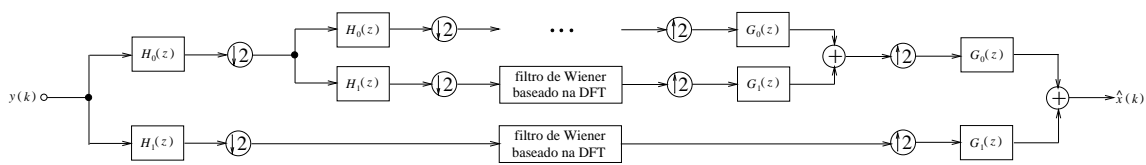


Figura 12.3: Filtragem de Wiener baseada na DFT numa estrutura multitaxa em oitavas sucessivas.

Considerando a oitava compreendida no intervalo $f_0 \leq f < 2f_0$, a aplicação de uma DFT de N pontos resultaria em componentes linearmente espaçadas de $\frac{2f_0}{N}$. Adotou-se como critério para escolha de N a garantia de que o espaçamento entre componentes fosse menor que uma fração de semitom γ para toda a oitava (o pior caso é seu limite inferior), isto é,

$$\frac{2f_0}{N} < f_0 \left(2^{\frac{\gamma}{12}} - 1 \right). \quad (12.1)$$

Em termos práticos, adotou-se $\gamma = \frac{1}{4}$, correspondendo a um espaçamento mínimo de um oitavo de tom no interior de cada oitava, um intervalo bastante estreito. Isso requereria $N = 138$; a fim de favorecer a FFT, adotou-se $N = 128$, com mínimo prejuízo do intervalo inicialmente escolhido. A estrutura proposta conserva as 10 faixas de frequência descritas na Subseção 12.2.2, com taxa de amostragem de 44,1 kHz.

Nesses termos, a representação global do sinal está sendo feita por 1280 componentes complexas (640 reais). A divisão linear do espectro nesse mesmo número

de componentes resultaria num espaçamento de 34,5 Hz. Da forma adotada, o espaçamento mínimo é de aproximadamente 0,673 Hz (de 0 a 86,1 Hz), que requereria 65536 componentes complexas linearmente espaçadas; o espaçamento máximo é de 172 Hz (de 11,025 a 22,05 kHz). Se fosse possível espaçar geometricamente as componentes, seriam necessárias 960 para chegar ao intervalo de um oitavo de tom; portanto, a solução híbrida proposta não é tão distante da ideal, sob esse aspecto.

Entretanto, deixou-se de lado até aqui a discussão das durações das janelas resultantes dos valores sugeridos acima. Compostas de 128 amostras, elas evoluem de 5,80 ms na sub-banda mais aguda (a de taxa mais alta) até 1,48 s nas 2 sub-bandas mais graves (as de taxa mais baixa). Como justificar isso face ao clássico valor-limite de 50 ms? O argumento é simples². A não-estacionaridade dos sinais pode ser implicitamente descrita pela sua própria representação em frequência. Em outras palavras, o que varia rapidamente pode ser convenientemente descrito pelas componentes de alta frequência nas sub-bandas superiores, assim como as variações lentas pelas de baixa frequência nas sub-bandas inferiores. Ora, a janela mais curta adotada é 10 vezes menor que o valor adotado em banda completa; portanto, a descrição dos fenômenos rápidos está sendo bastante favorecida. A grande diferença agora está na propriedade de descrever mais acuradamente os fenômenos lentos, para o que não havia amostras suficientes na janela de banda completa, que, afinal, tinha de atingir um compromisso entre as exigências de todo o espectro.

A complexidade do algoritmo, para M estágios, filtros de ordem Q e DFTs de comprimento N , fica em torno de $9 + 3 \log_2 N + \sum_{m=1}^M \frac{16}{2^m} \log_2 Q$ multiplicações por amostra. Com $M = 9$, $Q = 63$ e $N = 128$, isso corresponde a 126 multiplicações por amostra.

12.2.3.2 Avaliação

Foi realizada uma vasta quantidade de testes comparativos entre as 3 técnicas descritas anteriormente sobre sinais reais.

Escolheram-se 2 casos especialmente representativos em torno dos quais se traçará a discussão de desempenho nesta Subseção:

- Sinal A: trecho orquestral com altíssima variação de potência, que sai de um

²Uma discussão semelhante a esta foi delineada na Subseção 8.5.1.

murmúrio das cordas em extremo pianíssimo até atingir um *fortissimo* com grande impacto da percussão grave e volta a se desvanecer. O trecho inicial requer muito detalhe para ser percebido com clareza. Não há muita informação em frequências elevadas na maior parte do tempo. O trecho é extremamente lento, quase estático, o que o faz facilmente perturbável perceptivamente.

- Sinal B: solo de flauta acompanhado por orquestra barroca, com pouca variação de potência mas ocupação importante de todo o espectro. Destacam-se a flauta e o cravo nas frequências mais altas. O trecho é muito movimentado, tanto por ter andamento elevado quanto por ser bastante rítmico, não havendo praticamente repouso.

Ambos foram contaminados com ruído pseudobranco aditivo de forma a atingir razão sinal-ruído média de 30dB, o que corresponde a um nível bastante elevado de contaminação, em termos perceptivos. Suas versões corrompidas foram, então, restauradas de diversas formas e os resultados correspondentes comparados, subjetivamente (por audição informal) e objetivamente (com auxílio do PAQM, descrita nas Seções 6.3 e 6.4).

As três implementações da Filtragem de Wiener **FW-DFT** (baseada na DFT), **FW-SB** (realizada diretamente nas sub-bandas) e **FW-DFTSB** (aqui proposta) foram confrontadas.

Para a FW-DFT, a configuração que se mostrou preferencial emprega blocos de 2048 pontos com 50 % de superposição, janela de Hanning e estimação do espectro do sinal para cálculo do fator de Wiener pela mediana de 7 espectros de quadros adjacentes. Utilizou-se, ainda, um limite inferior ζ para o fator de Wiener (que, a princípio, poderia variar de 0 a 1).

No caso da FW-SB, adotaram-se 10 sub-bandas, bancos de QMF e fator de esquecimento $\alpha = 0,99$. O fator de supressão γ não foi utilizado, tendo seu papel substituído de forma aproximativa pelo parâmetro ζ definido anteriormente.

Por fim, a FW-DFTSB operou com 10 sub-bandas e bancos de QMF. Em cada sub-banda, empregou blocos de 128 pontos com 50 % de superposição, janela de Hanning e estimação do espectro do sinal para cálculo do fator de Wiener pela mediana de 7 espectros de quadros adjacentes, além de ζ .

A Tabela 12.1 mostra as medidas objetivas referentes ao Sinal A, cuja versão

corrompida resultara em $PAQM = 1,2646$. A segunda coluna de resultados apresenta os melhores valores da PAQM com os respectivos valores de ζ entre parênteses (ζ_o).

Tabela 12.1: PAQM dos sinais A tratados por diferentes implementações e parâmetros.

Tratamento	$\zeta = 0$	$\zeta = \zeta_o$	$\zeta = 0,2$
FW-SB, QMF64	0,4405	0,2790 (0,145)	0,3260
FW-SB, QMF16	0,2635	0,2212 (0,100)	0,3240
FW-DFT	0,4355	0,3734 (0,125)	0,4112
FW-DFTSB, QMF64	0,4753	0,3840 (0,125)	0,4141
FW-DFTSB, QMF16	0,3972	0,3638 (0,105)	0,4229

A Tabela 12.2 mostra as medidas objetivas referentes ao Sinal B, cuja versão corrompida resultara em $PAQM = 0,2923$. A segunda coluna de resultados apresenta os melhores valores da PAQM com os respectivos valores de ζ entre parênteses (ζ_o).

Tabela 12.2: PAQM dos sinais B tratados por diferentes implementações e parâmetros.

Tratamento	$\zeta = 0$	$\zeta = \zeta_o$
FW-SB, QMF64	0,1593	0,1184 (0,450)
FW-SB, QMF16	0,0993	0,0933 (0,300)
FW-DFT	0,7249	0,1694 (0,625)
FW-DFTSB, QMF64	0,3024	0,1466 (0,590)
FW-DFTSB, QMF16	0,2728	0,1535 (0,560)

Nas tabelas, QMF64 e QMF16 se referem a filtros FIR de ordens iguais a 63 e 15, respectivamente.

Os resultados, analisados a seguir, foram muito interessantes e novamente permitiram a avaliação da própria PAQM na presente aplicação.

Inicialmente, pode-se observar que os sinais A e B corrompidos, embora apresentando a mesma SNR, apresentaram valores muito diferentes de PAQM. Isso foi inteiramente coerente com o que se percebe auditivamente: o Sinal A, quase estático

e com um trecho de baixíssima intensidade, deixa o chiado muito mais a descoberto que o Sinal B, rápido, farto em agudos e de potência pouco variante.

Os testes começaram com o Sinal A, mais “difícil”.

Sendo a FW-DFT o método preferido na literatura, foi tomado como primeira referência. Seu melhor resultado subjetivo, de acordo com audição informal, foi obtido com $\zeta = 0,2$: fica algum ruído de fundo homogêneo, mas a qualidade geral da gravação é aceitavelmente preservada; a clareza do início de baixa potência fica um pouco prejudicada pelo chiado. Valores mais baixos de ζ , mesmo quando apresentaram melhor PAQM, levaram à substituição do chiado por ruído musical, um entrecruzamento de sinais tonais quase puros de frequências variáveis, levíssimos (v. Subseção 10.1.4.1). Com $\zeta = 0,125$, algo da clareza é resgatada com a redução do chiado, mas o ruído musical é muito mais perturbador, porque acrescenta movimentação “aleatória” o sinal. Isso parece apontar para uma “fragilidade” da PAQM: corretamente, ela dá o sinal tratado com $\zeta = 0,125$ como mais parecido que o original que tratado com $\zeta = 0,2$, o que efetivamente pode ser percebido; mas é claro que os efeitos psicológicos ligados aos diferentes tipos de resíduo não são avaliáveis pela medida.

A FW-DFTSB levou a resultados muito próximos aos da FW-DFT. Auditivamente, para valores baixos de ζ , o ruído musical se mostrou pouco maior após a FW-DFTSB que após a FW-DFT. Subjetiva e objetivamente, o uso de filtros menos seletivos foi preferível (para o bem de complexidade computacional). O melhor resultado perceptivo de todos com $\zeta = 0,2$ foi obtido com a FW-DFTSB/QMF16.

Os testes subjetivos realizados com a FW-SB resultaram bastante ruins: os sinais tratados soaram roucos, devido ao “véu” de ruído que acompanha a potência do sinal; o uso de $\zeta = 0$ acrescenta uma distorção intermitente ao início de baixa intensidade. Mesmo quando a escolha de *zeta* levou a baixos valores da PAQM, o resíduo de ruído resultou bastante perturbador. Isso indica que maior similaridade perceptiva não corresponde sempre a um resultado perceptivamente mais agradável.

Prosseguiu-se com o sinal B, e o panorama mudou inteiramente.

Por analogia, decidiu-se começar os testes pela FW-DFT com $\zeta = 0,2$. O resultado foi um fracasso: perderam-se drasticamente os agudos da gravação original— a PAQM foi pior que a do sinal corrompido! Liberou-se gradualmente o fator ζ até

a faixa em torno de 0,6: retornaram os agudos, juntamente com algum resíduo, homogêneo.

O comportamento foi muito similar com a FW-DFTSB, mas mesmo com baixos valores de ζ a perda de agudos não foi tão radical (veja-se a PAQM na coluna de $\zeta = 0$, que confirma isso). Subjetivamente, dentre os resultados de mínima PAQM, FW-DFTSB/QMF64 foi o melhor, com ligeira vantagem sobre QMF16.

É importante dizer que não se observou o ruído musical em nenhum resultado. Nos casos de ζ mais elevado, isso era previsível. Mas nos demais, explica-se pelas características do Sinal B, que acabam por mascará-lo.

A grande “surpresa”, contudo, ficou por conta da FW-SB. Já com fator de Wiener pleno ($\zeta = 0$) a impressão auditiva era de que o sinal estava limpo e não sofrera perda de altas freqüências, e a PAQM já surpreendentemente baixa de início corroborou isso; um pouco mais de atenção revelaria pequena rouquidão no cravo (o “véu” residual). A busca de ζ maior pela PAQM levou a resultados subjetivos e objetivos surpreendentes, ainda superiores com filtros menos rigorosos (QMF16). O que há de tão especial no Sinal B que possa ter favorecido tanto a FW-SB? Tanto a flauta quanto o cravo, os dois principais personagens em altas freqüências na gravação, apresentam um “chiado” como parte de seu som próprio—a flauta pelo sopro e o cravo pela sua reverberação particular. (Esse resíduo não-correlacionado é extremamente importante na síntese de instrumentos musicais naturais [86].) Pois bem, o “véu” de ruído residual que acompanha a potência do sinal em cada sub-banda na FW-SB acaba por emular esse chiado, só na presença dos instrumentos, ou seja, justamente “quando necessário”; com isso, o ouvido e a PAQM se deixam enganar muito bem...

Como conclusão geral, claro que com base em muitos outros exemplos aqui não descritos, o método FW-DFTSB, ao custo de uma complexidade um pouco maior que a dos outros, mostrou-se mais robusto a variações no tipo de sinal e levou, em geral, a resultados perceptivamente superiores (resíduo mais natural). A PAQM lhe foi favorável e, em maioria, o uso de filtros menos rigorosos para separar as sub-bandas foi preferível, com vantagem quanto os requisitos computacionais. Deve ser destacado que há uma linha de trabalhos que defende a supressão do ruído musical *a posteriori*, pelo pós-processamento de sinais que o adquiriram em razão do

uso de técnicas para eliminação de ruído de fundo; em [87], por exemplo, propõe-se a associação de *wavelet packets* e Filtragem de Wiener com esse fim.

Referências Bibliográficas

- [1] BISCAINHO, L. W. P., DINIZ, P. S. R., “On the Effects of Zero-Pole Pairs and Individual Zeros and Poles on Discrete-Time Transfer Functions”. In: *Proceedings of the International Symposium on Circuits and Systems (ISCAS-2000)*, v. I, pp. 363–366, IEEE, Geneve, Switzerland, May 2000.
- [2] BISCAINHO, L. W. P., DINIZ, P. S. R., CARVALHO, M. F. D., “Efeitos da Quantização de Sinais na Obtenção de seu Modelo Autorregressivo”. In: *Anais do XIII Congresso Brasileiro de Automática*, pp. 907–912, SBA, Florianópolis, SC, Brasil, Set. 2000.
- [3] BISCAINHO, L. W. P., DINIZ, P. S. R., ESQUEF, P. A. A., “A Model for an ARMA Process Split in Subbands”. In: *Proceedings of the International Symposium on Circuits and Systems (ISCAS-2000)*, v. III, pp. 97–100, IEEE, Geneve, Switzerland, May 2000.
- [4] BISCAINHO, L. W. P., DINIZ, P. S. R., ESQUEF, P. A. A., “Investigação do Modelo para um Processo ARMA Dividido em Sub-bandas”. In: *Anais do XIII Congresso Brasileiro de Automática*, pp. 580–585, SBA, Florianópolis, SC, Brasil, Set. 2000.
- [5] BISCAINHO, L. W. P., DINIZ, P. S. R., ESQUEF, P. A. A., “Aspectos da Localização de Distúrbios Impulsivos em Sinais de Áudio Usando Técnicas Baseadas em Limiar”. In: *Anais do XVII Simpósio Brasileiro de Telecomunicações*, pp. 185–190, SBrT, Vila Velha, ES, Brasil, Set. 1999.
- [6] ESQUEF, P. A. A., BISCAINHO, L. W. P., DINIZ, P. S. R. *et al.*, “A Double-Threshold-Based Approach to Impulsive Noise Detection in Audio Signals”. In:

- Proceedings of the 2000 EUSIPCO – European Signal Processing Conference*, v. IV, pp. 2041–2044, EURASIP, Tampere, Finland, Sept. 2000.
- [7] ESQUEF, P. A. A., BISCAINHO, L. W. P., DINIZ, P. S. R., “Detecção de Pulsos Longos em Sinais de Áudio”. In: *Anais do XVII Simpósio Brasileiro de Telecomunicações*, pp. 191–196, SBrT, Vila Velha, ES, Brasil, Set. 1999.
- [8] BISCAINHO, L. W. P., FREELAND, F. P., ESQUEF, P. A. A. *et al.*, “Wavelet Shrinkage Denoising Applied to Real Audio Signals under Perceptual Evaluation”. In: *Proceedings of the 2000 EUSIPCO – European Signal Processing Conference*, v. IV, pp. 2062–2064, EURASIP, Tampere, Finland, Sept. 2000.
- [9] ACADEMIA BRASILEIRA DE LETRAS, *Vocabulário Ortográfico da Língua Portuguesa*. 2 ed., Bloch, 1998.
- [10] LAMPORT, L., *LaTeX: A Document Preparation System*. 2 ed., Addison-Wesley, 1994.
- [11] GOOSSENS, M., MITTELBAACH, F., SAMARIN, A., *LaTeX Companion*. Addison-Wesley, 1994.
- [12] SCHÖNHERR, S., “Recording Technology History”, <http://history.acusd.edu/gen/reading/notes.html>, Aug. 2000, webpage.
- [13] BLOOM, P. J., “High-Quality Digital Audio in the Entertainment Industry: An Overview of Achievements and Challenges”, *IEEE Audio, Speech and Signal Processing Magazine*, v. 2, n. 4, pp. 2–25, Oct. 1985.
- [14] BERKHOUT, P. J., EGGERMONT, L. D. J., “Digital Audio Systems”, *IEEE Audio, Speech and Signal Processing Magazine*, v. 2, n. 4, pp. 45–67, Oct. 1985.
- [15] IEEE Audio and Electroacoustics Technical Committee, “The Past, Present, and Future in Audio Signal Processing”, *IEEE Signal Processing Magazine*, v. 14, n. 5, pp. 30–57, Sept. 1997. M. Kahrs, ed.
- [16] GODSILL, S. J., RAYNER, P. J. W., *Digital Audio Restoration*. Springer, 1998.
- [17] EARGLE, J., *Sound Recording*. New York, NY, USA, Van Nostrand Reinhold, 1980.

- [18] PAPOULIS, A., *Probability, Random Variables and Stochastic Processes*. 3 ed., New York, NY, USA, McGraw-Hill, 1991.
- [19] PEEBLES, Jr., P. Z., *Probability, Random Variables and Random Signal Principles*. 3 ed., New York, NY, USA, McGraw-Hill, 1993.
- [20] LEIPP, E., *Acoustique et Musique*. 2 ed., Paris, France, Masson, 1984.
- [21] SERRA, X., “Musical Sound Modeling with Sinusoids plus Noise”. In: Roads, C., Pope, S. T., Piccialli, A. *et al.* (eds.), *Musical Signal Processing*, chapter 3, Swets & Zeitlinger, p. 24 pages, 1997.
- [22] QUATIERI, T. F., McAulay, R. J., “Audio Signal Processing Based on Sinusoidal Analysis/Synthesis”. In: Brandenburg, K., Mahrs, M. (eds.), *Applications of Digital Signal Processing to Audio and Acoustics*, chapter 9, Norwell, MA, USA, Kluwer, pp. 344–416, 1998.
- [23] VAIDYANATHAN, P. P., DJOKOVIC, I., “Wavelet Transforms”. In: Chen, W.-K. (ed.), *The Circuits and Filters Handbook*, chapter 6, Boca Raton, FL, USA, CRC, pp. 134–219, 1995.
- [24] STRANG, G., NGUYEN, T., *Wavelets and Filter Banks*. Wellesley, MA, USA, Wellesley-Cambridge, 1996.
- [25] SHANMUGAN, K. S., BREIPOHL, A. H., *Random Signals: Detection, Estimation and Data Analysis*. New York, NY, USA, John Wiley & Sons, 1988.
- [26] OPPENHEIM, A., WILLSKY, A. I., *Signals and Systems*. 2 ed., Upper Saddle River, NJ, USA, Prentice-Hall, 1996.
- [27] ANANDA MOHAN, P. V., RAMACHANDRAN, V., SWAMY, M. N. S., “Formulas for Dynamic Range Evaluation of Second-Order Discrete-Time Filters”, *IEEE Transactions on Circuits and Systems*, v. CAS-30, n. 5, May 1983.
- [28] LAAKSO, T., “Comments on ‘Comments on ‘Calculations of L-Norms for Scaling Second-Order State-Space Digital Filter Sections’ ””, *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, v. 39, n. 4, Apr. 1992.

- [29] LA BARRA, B. A. L. D., PRIETO, R., “On the Frequency Response of Scalar Discrete-Time Systems”, *Automatica*, v. 35, pp. 1843–1853, 1999.
- [30] HAYKIN, S., *Communication Systems*, 3 ed., section 6.7, New York, NY, USA, John Wiley, pp. 374–378, 1994.
- [31] DINIZ, P. S. R., *Adaptive Filtering—Algorithms and Practical Implementation*. Norwell, MA, USA, Kluwer, 1997.
- [32] ZHENG, W. X., “A Least-Squares Based Method for Autoregressive Signals in Presence of Noise”, *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, v. 46, n. 1, pp. 81–85, Jan. 1999.
- [33] PAINTER, T., SPANIAS, A., “Perceptual Coding of Digital Audio”, *Proceedings of the IEEE*, v. 88, n. 4, pp. 451–513, Apr. 2000.
- [34] 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.
- [35] 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.
- [36] VASEGHI, S. V., *Algorithms for Restoration of Archived Gramophone Recordings*. Ph.D. dissertation, University of Cambridge, Cambridge, UK, Feb. 1988.
- [37] DONOHO, D. L., *Wavelet Shrinkage and W. V. D.: A Ten-Minute Tour*, Report 416, Department of Statistics, Stanford University, Stanford, CA, USA, June 1993.
- [38] CROCHIERE, R. E., RABINER, L. R., *Multirate Digital Signal Processing*. Englewood Cliffs, NJ, USA, Prentice-Hall, 1983.
- [39] VAIDYANATHAN, P. P., *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ, USA, Prentice-Hall, 1993.

- [40] NGUYEN, T. Q., “Aliasing-free Reconstruction Filter Bank”. In: Chen, W.-K. (ed.), *The Circuits and Filters Handbook*, chapter 85, Boca Raton, FL, USA, CRC, pp. 2682–2717, 1995.
- [41] GOSSE, K., SAINT-MARTIN, F. M. D., DUROT, X. *et al.*, “Subband Audio Coding with Synthesis Filters Minimizing a Perceptual Distortion”. In: *Proceedings ICASSP 97*, v. I, pp. 347–350, IEEE, Munich, Germany, 1997.
- [42] BEERENDS, J. G., STEMERDINK, J. A., “A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation”, *J. Audio Eng. Soc.*, v. 40, n. 12, pp. 963–978, Dec. 1992.
- [43] DELLER, JR., J. R., PROAKIS, J. G., HANSEN, J. H. L., “Speech Quality Assessment”. In: *Discrete-Time Processing of Speech Signals*, chapter 9, Nova Iorque, NY, EUA, Macmillan, pp. 568–595, 1993.
- [44] VARGAS, L. F. C., *Codificação de Áudio de Alta Fidelidade Utilizando Transformadas com Superposição*. Tese de mestrado, Universidade de Brasília - Faculdade de Tecnologia, Out. 1993.
- [45] ZWICKER, E., ZWICKER, U. T., “Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System”, *J. Audio Eng. Soc.*, v. 39, n. 3, pp. 115–126, Março 1991.
- [46] BEERENDS, J. G., “Audio Quality Determination Based on Perceptual Measurement Techniques”. In: Brandenburg, K., Mahrs, M. (eds.), *Applications of Digital Signal Processing to Audio and Acoustics*, chapter 1, Norwell, MA, USA, Kluwer, pp. 1–38, 1998.
- [47] 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, pp. 133–194, 1998.
- [48] Ó RUANAIDH, J. J. K., *Numerical Bayesian Methods Applied to Signal Processing*. Ph.D. dissertation, University of Cambridge, Cambridge, UK, Oct. 1994.

- [49] Ó RUANAIDH, J. J. K., FITZGERALD, W. J., “Interpolation of Missing Samples for Audio Restoration”, *Electronic Letters*, v. 30, n. 8, pp. 622–623, Apr. 1994.
- [50] 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.
- [51] PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. *et al.*, *Numerical Recipes in C—The Art of Scientific Computing*. 2 ed., Cambridge, 1992.
- [52] ESQUEF, P. A. A., *Implementação e Investigação Comparativa de Técnicas para Restauração Digital de Áudio*. Projeto final, UFRJ – EE, Fev. 1996. Sob orientação do autor desta tese.
- [53] MONTRÉSOR, S., VALIÈRE, J. C., BAUDRY, M., “Détection et Suppression de Bruits Impulsionnels Apliqués à la Restauration d’Enregistrements Anciens”, *Colloque de Physique C2*, v. 51, n. 2 (suppl.), pp. 757–760, Fév. 1990.
- [54] CZYZEWSKI, A., “Learning Algorithms for Audio Signal Enhancement: Part 1 Neural Network Implementation for the Removal of Impulse Distortions”, *Journal of the Audio Engineering Society*, v. 45, n. 10, pp. 815–831, Oct. 1997.
- [55] ESQUEF, P. A. A., *Restauração de Sinais de Áudio Degradados por Ruído Impulsivo*. Tese de mestrado, UFRJ – COPPE, Mar. 1999. Sob orientação de P. S. R. Diniz e coorientação informal do autor desta tese.
- [56] 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.
- [57] GODSILL, S. J., *The Restoration of Degraded Audio Signals*. Ph.D. dissertation, University of Cambridge, Cambridge, UK, Dec. 1993.
- [58] GODSILL, S. J., RAYNER, P. J. W., *Statistical Reconstruction and Analysis of Autoregressive Signals in Impulsive Noise*, Report CUED/F-INFENG/TR 233, University of Cambridge, Cambridge, UK, 1995.

- [59] GODSILL, S. J., RAYNER, P. J. W., “Robust Noise Reduction for Speech and Audio Signals”. In: *Proceedings of the 1996 IEEE International Conference on Acoustics, Speech, & Signal Processing*, pp. 625–628, IEEE, Atlanta, GA, USA, May 1996.
- [60] FREELAND, F. P., *Técnicas para Restauração Digital de Áudio Musical*. Projeto final, UFRJ – EE, Set. 1999. Sob orientação do autor desta tese.
- [61] HAERDY JR., A., *Técnicas para Restauração Digital de Áudio*. Projeto final, UFRJ – EE, Set. 1998. Sob orientação do autor desta tese.
- [62] GODSILL, S. J., TAN, C. H., “Removal of Low Frequency Transient Noise from Old Recordings Using Model-Based Signal Separation Techniques”. In: *Proc. IEEE Workshop on Audio and Acoustics*, IEEE, Mohonk, NY, USA, Oct. 1997.
- [63] OPPENHEIM, A., SHAFER, R., *Discrete-Time Signal Processing*. Englewood Cliffs, NJ, USA, Prentice-Hall, 1989.
- [64] JENKINS, W. R., “Fourier Series, Fourier transforms and the DFT”. In: Chen, W.-K. (ed.), *The Circuits and Filters Handbook*, chapter 4, Boca Raton, FL, USA, CRC, pp. 597–797, 1995.
- [65] LORBER, M., HOELDRICH, R., “A Combined Approach for Broadband Noise Reduction”. In: *Proc. IEEE Workshop on Audio and Acoustics*, IEEE, Mohonk, NY, USA, Oct. 1997.
- [66] BURRUS, C. S., GUO, H., GOPINATH, R. A., *Introduction to Wavelet and Wavelet Transform—A Primer*. Prentice-Hall, 1997.
- [67] VETTERLI, M., KOVAČEVIĆ, J., *Wavelet and Subband Coding*. New Jersey, Prentice Hall, 1995.
- [68] MALLAT, S., *A Wavelet Tour of Signal Processing*. San Diego, CA, USA, Academic Press, 1997.
- [69] ABRY, P., FLANDRIN, P., “On the Initialization of the Discrete Wavelet Transform Algorithm”, *IEEE Signal Processing Letters*, v. 1, n. 2, pp. 32–34, Feb. 1994.

- [70] DONOHO, D. L., *Interpolating Wavelet Transforms*, Report, Department of Statistics, Stanford University, Stanford, CA, USA, Oct. 1992.
- [71] DONOHO, D., JOHNSTONE, I., “Ideal Spatial Adaptation via Wavelet Shrinkage”, *Biometrika*, v. 81, pp. 425–455, Dec. 1994.
- [72] DONOHO, D. L., JOHNSTONE, I., “Adapting to Unknown Smoothness via Wavelet Shrinkage”, *Journal of the American Statistical Association*, v. 90, pp. 1200–1224, Dec. 1995.
- [73] DONOHO, D. L., “De-noising by Soft Thresholding”, *IEEE Transactions on Information Theory*, v. 41, n. 3, pp. 613–627, 1995.
- [74] DONOHO, D. L., JOHNSTONE, I., KERKYACHARIAN, G. *et al.*, “Wavelet Shrinkage: Asymptopia?”, *J. Roy. Stat. Soc. B*, v. 57, n. 2, pp. 301–337, 1995.
- [75] RAMARAPU, P. K., MAHER, R. C., “Methods for Reducing Audible Artifacts in a Wavelet-Based Broad-Band Denoising System”, *Journal of the Audio Engineering Society*, v. 46, n. 3, pp. 178–190, Mar. 1998.
- [76] COIFMAN, R. R., DONOHO, D. L., “Translation-Invariant De-Noising”. In: Antoniadis, A. (ed.), *Wavelets and Statistics*, Springer-Verlag Lecture Notes, 1995.
- [77] BERGER, J., COIFMAN, R. R., GOLDBERG, M. J., “Removing Noise from Music Using Local Trigonometric Bases and Wavelet Packets”, *J. Audio Eng. Soc.*, v. 42, n. 10, pp. 808–818, Oct. 1994.
- [78] GHAEL, S. P., SAYEED, A. M., BARANIUK, R. G., “Improved Wavelet Denoising via Empirical Wiener Filtering”. In: *Proc. of SPIE*, San Diego, July 1997.
- [79] HAYKIN, S., *Adaptive Filter Theory*. 3 ed., Upper Saddle River, NJ, USA, Prentice-Hall, 1996.
- [80] BISCAINHO, L. W. P., *Processamento Adaptativo em Multibandas Incluindo Algoritmo Rápido para Filtros Adaptativos FIR*. Tese de mestrado, UFRJ – COPPE, Mai. 1990. Sob orientação de P. S. R. Diniz.

- [81] DINIZ, P. S. R., BISCAINHO, L. W. P., “Optimal Convergence Factor for the LMS/Newton Algorithm”. In: *Proceedings ICASSP 91*, v. 3, pp. 2221–2224, IEEE, Toronto, Canada, May 1991.
- [82] DINIZ, P. S. R., BISCAINHO, L. W. P., “Optimal Variable Step-Size for the LMS/Newton with Application to Subband Adaptive Filtering”, *IEEE Transactions on Signal Processing*, v. 40, n. 11, pp. 2925–2829, Nov. 1992.
- [83] WORNELL, G. W., *Signal Processing with Fractals—A Wavelet-based Approach*. Upper Saddle River, NJ, USA, Prentice-Hall, 1996.
- [84] ZHANG, J., WALTER, G., “A Wavelet-Based KL-like Expansion for Wide-Sense Stationary Random Processes”, *IEEE Transactions on Signal Processing*, v. 42, n. 7, pp. 1737–1745, Jul. 1994.
- [85] CAPPÉ, O., *Techniques de Réduction de Bruit pour la Restauration d’Enregistrements Musicaux*. Thèse de doctorat, École Nationale Supérieure des Télécommunication, Sep. 1993.
- [86] GOODWIN, M., “Residual Modeling in Music Analysis and Synthesis”. In: *Proc. of the ICASSP-96*, v. II, pp. 1005–1008, IEEE, Atlanta, GA, USA, May 1996.
- [87] JEBARA, S. B., BENAZZA-BENYAHIA, A., KHELIFA, A. B., “Reduction of Musical Noise Generated by Spectral Subtraction by Combining Wavelet Packet Transform and Wiener Filtering”. In: *Proceedings of the 2000 EUSIPCO – European Signal Processing Conference*, v. II, pp. 749–752, EURASIP, Tampere, Finland, Sept. 2000.