



COPPE/UFRJ

MODELAGEM SENOIDAL DE SINAIS MUSICAIS: TÉCNICAS DE ANÁLISE
E AVALIAÇÃO

Leonardo de Oliveira Nunes

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientadores: Luiz Wagner Pereira
Biscainho
Paulo Antonio Andrade
Esquef

Rio de Janeiro

Agosto de 2009

MODELAGEM SENOIDAL DE SINAIS MUSICAIS: TÉCNICAS DE ANÁLISE
E AVALIAÇÃO

Leonardo de Oliveira Nunes

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Aprovada por:

Prof. Luiz Wagner Pereira Biscainho, D.Sc.

Prof. Paulo Antonio Andrade Esquef, D.Sc.

Prof. Eduardo Antonio de Barros, Ph.D.

Prof. Jacques Szczupak, Ph.D.

RIO DE JANEIRO, RJ – BRASIL

AGOSTO DE 2009

Nunes, Leonardo de Oliveira

Modelagem Senoidal de Sinais Musicais: Técnicas de Análise e Avaliação/Leonardo de Oliveira Nunes. – Rio de Janeiro: UFRJ/COPPE, 2009.

XVII, 143 p.: il.; 29,7cm.

Orientadores: Luiz Wagner Pereira Biscainho

Paulo Antonio Andrade Esquef

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2009.

Referências Bibliográficas: p. 127 – 141.

1. processamento de sinais. 2. processamento de sinais acústicos. 3. análise senoidal. 4. filtros adaptativos. I. Biscainho, Luiz Wagner Pereira *et al.*. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Para:
Domingos José S. de Oliveira

Agradecimentos

Inicialmente, devo agradecer a minha família (e nisto incluo todos os meus amigos) pelo apoio dado durante todos esses anos de educação e, acima de tudo, pelo amor. Meus pais, minha irmã e meus avós criaram o ambiente no qual este trabalho se tornou realidade. Tenho que agradecer a minha namorada, Maria Lucia, pelo carinho e apoio fornecido durante estes dois anos de trabalho.

Ao Prof. Luiz Wagner devo agradecer por ter acreditado em mim e ter me aceitado como aluno de iniciação científica, pela infinita paciência e, acima de tudo, pela amizade. E ao Paulo Esquef, por trazer o tema deste trabalho para o GPA, pela orientação e pelas inúmeras sugestões e idéias.

Não posso deixar de lado todos os amigos que fiz no Grupo de Processamento de Áudio e cujos trabalhos estão de alguma forma associados a este. Agradeço a Fábio Freeland, Filipe Diniz, Alan Tygel, Rafael de Jesus, Alexandre Leizor, Flávio Ávila e a todos os outros membros GPA. Agradeço, também, aos amigos do LPS, especialmente Tadeu Ferreira e André Targino. E a Wallace Martins, pela amizade e inúmeras discussões técnicas.

Por fim, agradeço os Profs. Eduardo Barros e Jacques Szczupak por terem aceitado o convite para participar da banca desta dissertação.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Fundação de Amparo à Pesquisa do Rio de Janeiro (FAPERJ) pelo apoio financeiro.

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

MODELAGEM SENOIDAL DE SINAIS MUSICAIS: TÉCNICAS DE ANÁLISE E AVALIAÇÃO

Leonardo de Oliveira Nunes

Agosto/2009

Orientadores: Luiz Wagner Pereira Biscainho

Paulo Antonio Andrade Esquef

Programa: Engenharia Elétrica

Técnicas de análise de áudio estão presentes em um sem-número de aplicações, como mixagem e edição de gravações, transcrição musical automática, identificação de instrumentos musicais, etc. Em particular, técnicas baseadas em modelos são interessantes por fornecerem uma representação intermediária do sinal musical: mais interpretável que a variação de uma grandeza elétrica ao longo do tempo, e menos simbólica que uma partitura musical. Considerando a natureza “tonal” (espectro de magnitude com predominância de picos) dos sinais musicais, um dos modelos mais favoráveis à sua representação é o senoidal, que descreve o sinal como um somatório de senóides com amplitude e fase (frequência) moduladas ao longo do tempo.

Essa dissertação tem como objetivo apresentar métodos úteis na modelagem senoidal de sinais musicais, bem como para a avaliação objetiva de desempenho dos algoritmos envolvidos. Para a obtenção do modelo senoidal, são apresentados critérios para a seleção de picos senoidais e algoritmos baseados em preditores adaptativos para o rastreamento de trilhas senoidais. Já para a avaliação de desempenho de algoritmos de análise senoidal, são descritos avaliadores específicos para os detectores de picos e para os rastreadores senoidais. Também foi criado um banco de sinais de áudio acompanhados por suas trilhas senoidais que tem como objetivo fornecer dados de referência para os avaliadores.

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

SINUSOIDAL MODELING OF MUSICAL SIGNALS: TECHNIQUES FOR ANALYSIS AND EVALUATION

Leonardo de Oliveira Nunes

August/2009

Advisors: Luiz Wagner Pereira Biscainho

Paulo Antonio Andrade Esquef

Department: Electrical Engineering

Techniques for analysis of audio signals can be found in many applications, such as mixing and editing of recordings, automatic musical transcription, musical instrument identification, etc. In particular, model based techniques are useful for providing an intermediate representation for a musical signal: more interpretable than the time-domain signal, and less symbolic than a musical score. Considering the “tonal” (peaky magnitude spectrum) nature of musical signals, the sinusoidal model, which represents the signal as a sum of amplitude and phase (frequency) modulated sinusoids, is an adequate choice for such a representation.

The target of this dissertation is to present suitable methods for the sinusoidal modeling of music signals as well as for the objective assessment of the performance of sinusoidal analysis algorithms. For the obtention of sinusoidal models, criteria for the selection of sinusoidal peaks and adaptive prediction based sinusoidal trackers are presented. For the performance assessment of sinusoidal analysis algorithms, specific performance assessment methods are described for both peak detection and sinusoidal tracking procedures. A database composed of audio signals coupled with their respective sinusoidal tracks was created in order to provide ancillary reference data for performance assessment methods.

Sumário

Lista de Figuras	xii
Lista de Tabelas	xiv
Lista de Abreviaturas	xvi
1 Introdução	1
1.1 Introdução à Análise Senoidal	2
1.2 Revisão Bibliográfica: Modificações e Aplicações do Modelo Senoidal	3
1.2.1 Aprimoramentos	4
1.2.2 Aplicações	4
1.3 Objetivo e Contribuições desta Dissertação	5
1.4 Estrutura da Dissertação	6
2 Algoritmos para Detecção de Senóides	7
2.1 Introdução	7
2.2 Mapeamento Tempo-Freqüência	8
2.3 Algoritmos para Detecção de Picos Espectrais	9
2.3.1 Estratégia de Detecção de Picos	10
2.3.2 Pré-Processamento Espectral	11
2.3.3 Correção do Espectro	15
2.3.4 Critérios de Seleção de Picos	15
2.4 Estimação de Parâmetros dos Picos	19
2.4.1 Técnicas de Estimação de Freqüência Instantânea	19
2.4.2 Técnicas de Estimação de Amplitude e de Fase	21
2.5 Comparação dos Algoritmos de Estimação da Tendência Espectral . .	22

2.5.1	Geração do Sinal de Teste	23
2.5.2	Esquema de Detecção de Picos	24
2.5.3	Avaliação de Desempenho	24
2.5.4	Resultados	27
2.6	Avaliação Generalizada de Detectores e Estimadores de Picos Espectrais	29
2.6.1	Metodologia de Avaliação	31
2.6.2	Avaliador	33
3	Algoritmos de Rastreamento de Trajetórias Senoidais	38
3.1	Introdução	38
3.1.1	Revisão Bibliográfica	40
3.2	O Algoritmo MQ	43
3.3	Rastreamento de Trilhas Utilizando Predição Linear	45
3.4	Preditores	46
3.4.1	RLS	48
3.4.2	RLS <i>Lattice</i>	50
3.4.3	Inicialização dos Preditores para a Frequência	53
3.4.4	Escolha dos Parâmetros dos Preditores	53
3.5	Algoritmos de Decisão	55
3.5.1	Funções de Decisão	57
3.5.2	Heurísticas Adicionais	58
3.6	Avaliação de algoritmos de rastreamento de trilhas	62
3.6.1	Metodologia	63
3.6.2	Algoritmo de Avaliação	65
3.6.3	Algoritmo de Avaliação Modificado	68
4	Desenvolvimento de um Banco de Trilhas Senoidais	73
4.1	Introdução	73
4.2	Especificação e Categorização dos Sinais-Fonte	74
4.3	Sinais-Fonte Escolhidos	76
4.3.1	Nível 0	76
4.3.2	Nível 1	77
4.3.3	Nível 2	78

4.3.4	Nível 3	80
4.4	Especificação do Banco de Trilhas Senoidais	81
4.5	Especificação dos Métodos de Análise	82
4.5.1	Diretrizes de Análise	82
4.5.2	Especificação dos Erros de Estimacão	83
4.6	Métodos de Análise	87
4.6.1	Diviso em Sub-Bandas	87
4.6.2	Mapeamento Tempo-Freqencia	88
4.6.3	Deteco de Picos	90
4.6.4	Refinamento da Freqencia	91
4.6.5	Refinamento da Amplitude e da Fase	92
4.6.6	Rastreamento	92
4.6.7	Pós-processamento	93
4.7	Trilhas Obtidas	96
4.8	Validaco do BTS	99
4.8.1	Obteno das Trilhas de Referncia	99
4.8.2	Estimaco de Freqencia	100
4.8.3	Estimaco de Amplitude	100
4.8.4	Deteco de <i>Onset</i> e <i>Offset</i>	101
5	Avaliaco de Desempenho de Algoritmos de Análise Senoidal	104
5.1	Introduo	104
5.2	Teste 1: Avaliaco de Métodos de Refinamento da Freqencia	105
5.2.1	Sinais de Teste	105
5.2.2	Descrio do Teste	106
5.2.3	Resultados	108
5.3	Teste 2: Avaliaco de Algoritmos de Rastreamento de Senóides	109
5.3.1	Sinais Escolhidos	109
5.3.2	Descrio do Teste	109
5.3.3	Resultados	111
5.4	Teste 3: Avaliaco de Sistemas de Análise Senoidal	116
5.4.1	Descrio do Teste	116
5.4.2	Resultados	118

6	Conclusões e Trabalhos Futuros	121
6.1	Conclusões	121
6.1.1	Detecção de Senóides	121
6.1.2	Rastreamento de Trilhas	122
6.1.3	Banco de Trilhas	123
6.2	Trabalhos Futuros	123
6.2.1	Detecção de Senóides	123
6.2.2	Rastreamento de Trilhas	124
6.2.3	Banco de Trilhas Senoidais	124
6.2.4	Avaliação de Desempenho	124
	Referências do Autor no Tema da Dissertação	126
	Referências Bibliográficas	127
A	Parâmetros Utilizados no Teste 2 do Capítulo 5	142

Lista de Figuras

1.1	Etapas para obtenção de um modelo senoidal.	3
2.1	Limiar constante aplicado ao espectro de um sinal de áudio.	10
2.2	Estratégia de detecção de picos espectrais	11
2.3	Estimativas da tendência espectral.	16
2.4	Comparação entre o espectro original e sua versão corrigida.	17
2.5	Esquema de avaliação de algoritmos de detecção de picos	23
2.6	Metodologia de avaliação: detecção/estimação de senóides	32
3.1	Frequência dos picos detectados para um sinal sintético de trompete.	39
3.2	Picos detectados do oitavo harmônico de um sinal de trompete.	41
3.3	Visão geral do algoritmo proposto para o rastreamento de trilhas.	46
3.4	Estrutura típica de filtragem adaptativa.	47
3.5	Estrutura <i>lattice a priori</i> para $J = 3$	52
3.6	Exemplo do efeito da inicialização dos preditores.	54
3.7	Erro de previsão (em %) de um filtro adaptativo RLS.	61
3.8	Metodologia de avaliação: algoritmos de rastreamento de senóides.	64
3.9	Possibilidades de superposição de duas trilhas ao longo do tempo.	67
4.1	Curva de erro percentual máximo aceitável na estimativa de f	85
4.2	Incremento necessário para se perceber uma mudança na intensidade.	86
4.3	Pirâmide de Laplace	88
4.4	Resposta em magnitude do filtro $h[n]$ na frequência.	89
4.5	Divisão em sub-bandas utilizada no BTS.	90
4.6	Exemplo de interpolação.	95
4.7	Distribuição de $\Gamma(f_{i,m}, \hat{f}_{i,m})$	101

4.8	Distribuição do erro relativo de estimação de amplitude.	102
4.9	Distribuição do erro de estimação de <i>onset</i> e <i>offset</i>	103
5.1	Valores de ED obtidos por um detector de picos para diferentes <i>d</i> . . .	107
5.2	Trajetórias de frequência para o sinal do Nível 1.	114
5.3	Trajetórias de frequência para o sinal do Nível 3.	115
5.4	Resultado obtido no Teste 3 para os sinais do Nível 1.	118
5.5	Resultado obtido no Teste 3 para os sinais do Nível 2.	119

Lista de Tabelas

2.1	Parâmetros utilizados no caso de estudo.	15
2.2	Indicadores de desempenho obtidos no Teste 1	28
2.3	Indicadores de desempenho obtidos no Teste 2	30
2.4	Indicadores de desempenho obtidos no Teste 3	31
4.1	Informações sobre os sinais escolhidos para o Nível 1 do BTS.	78
4.2	Características dos sinais escolhidos para o Nível 1 do BTS.	78
4.3	Identificação e Localização dos sinais do Nível 2.	79
4.4	Informação das composições executadas nos sinais do Nível 2 do BTS.	79
4.5	Características dos sinais no Nível 2 do BTS.	80
4.6	Identificação e Localização do sinal do Nível 3.	80
4.7	Inicialização dos parâmetros de análise de cada etapa do processamento.	96
4.8	Informações sobre a análise dos sinais de cada nível do BTS.	98
5.1	Valores dos parâmetros da STFT utilizados no Teste 1.	106
5.2	Resultado do Teste 1	108
5.3	Parâmetros do método MQ utilizados na comparação dos rastreadores.	111
5.4	Valores obtidos para E	112
5.5	Valores obtidos para E^{mod}	112
5.6	Valores obtidos para E^{s}	112
5.7	Valores dos parâmetros da STFT utilizados no Teste 3.	117
5.8	Teste 3: Parâmetros utilizados na primeira configuração.	117
5.9	Teste 3: Parâmetros utilizados na terceira configuração.	117
5.10	Resultado obtido no Teste 3 para o sinal do Nível 3.	118
A.1	Parâmetros do preditor Burg utilizados no Teste 2.	142

A.2	Parâmetros do preditor RLS utilizados no Teste 2.	143
A.3	Parâmetros do preditor <i>lattice</i> utilizados no Teste 2.	143

Lista de Abreviaturas

AR	Autorregressivo, p. 11
BTS	Banco de Trilhas Senoidais, p. 74
DFT	Transformada de Fourier Discreta, p. 8
EA	Erro de Associação, p. 33
ED	Erro de Detecção, p. 37
FIR	<i>Finite Impulse Response</i> , p. 13
FN	Falso Negativo, p. 33
FP	Falso Positivo, p. 33
HMM	Modelos Ocultos de Markov, p. 41
JND	<i>Just Noticeable Difference</i> , p. 83
MQ	Algoritmo de McAulay & Quatieri, p. 41
NLRF	Filtro Recursivo Não-Linear, p. 14
RLS	<i>Recursive Least Squares</i> , p. 48
RWC	<i>Real World Computing Database</i> , p. 76
SL	<i>Sensation Level</i> , p. 83
SNR	Razão Sinal-Ruído, p. 18
SSE	Estimador do Espectro Estocástico, p. 13
STFT	Transformada de Fourier de Tempo Curto, p. 8

Capítulo 1

Introdução

A disponibilidade de computadores cada vez mais velozes propiciou o desenvolvimento de técnicas avançadas de processamento digital de sinais de áudio. Em particular, foram beneficiadas a análise e modificação de sinais de áudio, permitindo o surgimento de novas aplicações. Podem-se destacar as seguintes aplicações que têm recebido considerável atenção por pesquisadores tanto do meio acadêmico quanto industrial: transcrição musical automática [1], codificação [2], restauração de sinais [3] e separação de fontes sonoras [4].

Em geral, essas aplicações podem tirar proveito de alguma representação do sinal que concentre a informação mais relevante em um número reduzido de componentes. Além disso, é desejável que essas componentes sejam de fácil manipulação e estejam associadas a aspectos perceptivos do sinal.

Uma representação que se mostrou bastante adequada para sinais de áudio é a modelagem senoidal [5], que expressa as partes determinísticas do sinal (componentes ressonantes), como um somatório de senóides moduladas em amplitude e frequência. A escolha de senóides como elementos básicos leva em consideração a característica em geral quase-periódica (“oscilatória”) desses sinais, que se reflete num espectro cuja magnitude tem predominância de picos. Por sua vez, essa conformação espectral está associada às ressonâncias fisicamente responsáveis pela produção do som. Publicações recentes [6, 7, 8, 9, 10, 11, 12, 13, 14, 15] demonstram que a modelagem senoidal ainda é um tópico de grande atividade na área de processamento de sinais de áudio.

Esta dissertação descreve soluções para a obtenção do modelo senoidal para sinais

de áudio. Também é abordada a questão da avaliação da qualidade do modelo obtido e possíveis métricas que indiquem o desempenho de sistemas de análise senoidal.

O restante deste capítulo será organizado da seguinte forma: Na próxima seção será feita uma breve introdução à análise senoidal. Em seguida, são apresentadas na Seção 1.2 algumas modificações sobre o modelo original juntamente com referências para aplicações que o utilizam. Na Seção 1.3, são descritos o objetivo desta dissertação e as suas principais contribuições. Ao final, na Seção 1.4, é detalhada a estrutura desta dissertação.

1.1 Introdução à Análise Senoidal

Nesta seção será feita uma breve introdução à análise senoidal, sendo que parte de seu conteúdo foi retirada de [P2].

A análise senoidal descreve [5] um sinal de áudio $s(t)$ como a soma de L senóides, moduladas em amplitude e frequência [16]:

$$s(t) = \sum_{l=1}^L A_l(t) \text{sen}(\Psi_l(t)), \quad (1.1)$$

com

$$\Psi_l(t) = \Psi_l(0) + \int_0^t \omega_l(u) du, \quad (1.2)$$

onde $A_l(t)$ e $\Psi_l(t)$ são, respectivamente, a modulação em amplitude e fase do l -ésimo termo do somatório, também chamado de trilha senoidal¹. A modulação em frequência $\omega_l(t)$ é relacionada com a modulação em fase através da equação (1.2).

A natureza contínua da amplitude $A_l(t)$ e da frequência angular $\omega_l(t)$ dificultam o tratamento computacional do problema. Para simplificar a análise, a equação (1.1) é usualmente substituída por um modelo discreto

$$s[n] = \sum_{l=1}^L A_l[n] \text{sen} \Psi_l[n], \quad (1.3)$$

que também é considerado estacionário para curtos períodos de tempo. Para uma dada trilha, e assumindo que $A_l[n]$ e $\Psi_l[n]$ possuem uma banda muito mais estreita que a do sinal $s[n]$, são feitas para um intervalo de N amostras (também chamado de quadro) as aproximações $A_l[n] \approx A_l$ e $\Psi_l[n] \approx \Omega_l n + \Psi_l[0]$, onde A_l e Ω_l são

¹Nesta dissertação, o termo trilha senoidal significará um único termo da equação (1.1).

constantes. Essa hipótese permite uma análise em blocos do sinal de áudio com o objetivo de se encontrar a evolução de A_l e Ω_l ao longo dos quadros, o que determina as trajetórias da amplitude e da frequência, respectivamente.

A Figura 1.1 ilustra as três etapas mais usuais [5] para a obtenção do modelo senoidal para um sinal discreto $s[n]$. A etapa de mapeamento tempo-frequência estima o espectro do sinal para cada quadro de análise. Já a etapa de detecção de picos espectrais busca no espectro obtido em cada quadro os picos que corresponderiam às componentes senoidais a serem modeladas. A etapa de refinamento serve para melhorar as estimativas dos valores de frequência e de amplitude para os picos detectados. Em alguns trabalhos, as etapas de ‘Mapeamento Tempo-Frequência’ e ‘Detecção de Picos’ podem ser substituídas por uma única etapa chamada de ‘Detecção de Senóides’. Ao final, a etapa de ‘Rastreamento das Trajetórias’ encontra, a partir dos picos detectados, as trajetórias correspondentes a cada trilha senoidal. Estas etapas serão descritas mais detalhadamente ao longo desta dissertação.



Figura 1.1: Etapas usualmente empregadas para a obtenção do modelo senoidal de um sinal $s[n]$. O bloco ‘Ref.’ denota a etapa de refinamentos de parâmetros dos picos que usualmente está associada à etapa de detecção dos picos.

1.2 Revisão Bibliográfica: Modificações e Aplicações do Modelo Senoidal

Apesar de ainda não assumir explicitamente o modelo senoidal, atualmente considera-se o *Phase Vocoder* [17] como o primeiro sistema de análise senoidal. Com a introdução dos métodos de análise para sinais discretos em [18] e em [16], o modelo se popularizou. Nesses trabalhos, também são apresentadas as etapas de análise brevemente descritas na seção anterior.

1.2.1 Aprimoramentos

Algumas soluções para se obter o modelo senoidal procuram evitar a etapa de detecção de picos. Dentre os métodos alternativos de análise, os principais são a análise por síntese [19], que busca extrair as trilhas senoidais seqüencialmente; a utilização de *matching pursuit* [20, 21], que utiliza um dicionário de senóides moduladas em amplitude e frequência na decomposição do sinal de entrada; e a utilização de processamento bayesiano [3, 22].

Uma modificação sobre o modelo clássico é o modelo senoidal com multi-resolução [23] que foi proposto para a codificação de sinais de áudio [24]. Nesta modificação, o sinal é dividido em oitavas, tornando o sistema de análise mais similar à percepção humana. Já uma modificação do modelo que se mostrou útil para a síntese é descrita em [25], onde cada senóide do modelo tem a sua “largura de banda” aumentada através da soma de uma componente estocástica. Também é proposto em [25] um esquema de análise que modifica os instantes de início e fim de cada senóide, com a finalidade de melhorar o modelo durante transitórios.

Para tornar o modelo mais abrangente, foram propostas na literatura diversas extensões que procuram modelar não apenas a parcela tonal de um sinal de áudio mas também as parcelas similares a ruído e os transitórios. Em [16], a parcela não-ressonante do sinal é representada por uma realização de ruído cuja densidade espectral de potência é modificada de modo a se assemelhar ao espectro do sinal de áudio após a retirada de sua parcela ressonante. Esse modelo, usualmente chamado de senoidal+ruído, foi expandido através da inclusão de um modelo para os transitórios [26] tornando-se o chamado modelo senoidal+transitórios+ruído [27].

1.2.2 Aplicações

A análise senoidal já foi utilizada em diferentes aplicações com sucesso. No caso de análise e síntese de voz podem ser encontrados na literatura algoritmos para modificação de *pitch* [28, 19] e de entonação [29, 30, 31].

Em codificação, foi padronizado pelo grupo MPEG um codificador baseado no modelo senoidal+ruído [32, 33] para codificação de áudio em taxas muito baixas (até 4 kbits/s). Podem ser encontradas na literatura diversas outras propostas de codificadores de áudio que utilizam o modelo senoidal [34, 35, 36].

Foi abordada em [37, 38] a utilização de trilhas senoidais para a separação de fontes sonoras. Já em [39], é apresentada uma solução de extração de melodia utilizando o modelo senoidal. Soluções para transcrição musical automática utilizando análise senoidal podem ser encontradas em [40] e [41].

Outras aplicações em que o modelo senoidal foi utilizado incluem *watermarking* [6], restauração de áudio [3, 42], identificação de músicas [43] e avaliação de qualidade de sinais de voz [44].

1.3 Objetivo e Contribuições desta Dissertação

Esta dissertação aborda algoritmos para a obtenção de um modelo senoidal para sinais musicais e avaliação da qualidade do modelo obtido. São apresentados algoritmos para a detecção de picos espectrais e para o rastreamento de trajetórias espectrais. Além disso, também são apresentados avaliadores projetados para permitir a comparação quantitativa de diferentes algoritmos para análise senoidal. Esta dissertação não utiliza os métodos desenvolvidos em nenhuma aplicação final; porém, espera-se que as contribuições fornecidas, tanto no formato de novos algoritmos quanto nos novos métodos de avaliação, possam vir a melhorar o desempenho de aplicações que utilizam o modelo senoidal.

A contribuição na etapa de detecção de picos senoidais consiste num critério de seleção de picos espectrais adequado para sinais de áudio e dois avaliadores relacionados: um que avalia apenas falhas na detecção das senóides e outro, que é uma modificação do avaliador proposto em [45], que procura avaliar conjuntamente a detecção das senóides e a estimação de seus parâmetros.

Na etapa de rastreamento de trilhas, é proposta uma solução que utiliza filtros adaptativos. Também são descritas heurísticas apropriadas ao rastreamento adaptativo de trajetórias. Por fim, é revisado um avaliador de desempenho [45] para a etapa de rastreamento de trajetórias sobre o qual é proposta uma modificação.

Para facilitar a avaliação quantitativa de desempenho de sistemas de análise senoidais, foi desenvolvido um banco de trilhas senoidais. O objetivo deste banco é fornecer dados referenciais na forma de sinais de áudio acompanhados de suas trilhas senoidais de forma a permitir a análise de desempenho quantitativa de diferentes

métodos de análise senoidal. São fornecidos também exemplos de como utilizar este banco de trilhas em conjunto com os avaliadores específicos de cada etapa da obtenção de um modelo senoidal.

A pesquisa do autor no tema desta dissertação resultou em três publicações. A primeira [P2] apresenta um critério de seleção para a etapa de detecção e uma avaliação do desempenho do critério. Já as outras duas [P3,P4] descrevem os dois preditores adaptativos e as heurísticas associadas.

1.4 Estrutura da Dissertação

No Capítulo 2 são descritas as soluções para as etapas de mapeamento tempo-frequência e detecção dos picos espectrais, juntamente com métodos de avaliação de desempenho.

No Capítulo 3, são descritos algoritmos de rastreamento de trajetórias e um esquema de rastreamento utilizando predição linear. Também são apresentados dois avaliadores para a etapa de rastreamento de trajetórias.

É detalhado no Capítulo 4 um banco de trilhas senoidais para avaliação de sistemas de análise senoidal. Em particular, após a especificação do banco, são descritos os seus sinais fonte e o método utilizado para a análise dos sinais.

São feitas no Capítulo 5 demonstrações de como o banco de trilhas, em conjunto com os avaliadores apresentados nos Capítulos 2 e 3, pode ser utilizado numa avaliação comparativa de diferentes métodos de análise senoidal.

Por fim, no Capítulo 6, são tiradas as conclusões deste trabalho e são propostas possíveis continuações para a pesquisa.

Capítulo 2

Algoritmos para Detecção de Senóides

2.1 Introdução

Neste capítulo, são apresentados métodos para a detecção de senóides e estimação de seus parâmetros. Conforme discutido no capítulo anterior, esta é a primeira etapa para obtenção do modelo senoidal para sinais de áudio.

Os algoritmos de detecção de senóides geralmente podem operar no domínio do tempo ou no da frequência. Algoritmos que operam no domínio do tempo incluem detectores de cruzamento por zero [46], modelagem autorregressiva [47] e outros modelos paramétricos, como os propostos em [48, 49, 50]. Já os algoritmos de detecção de senóides no domínio da frequência procuram localizar picos espectrais; isso é justificado, já que o espectro de sinais de áudio tende a exibir concentrações de energia em torno das frequências de suas componentes ressonantes.

Na Seção 2.2, é apresentado o algoritmo utilizado neste trabalho para capturar possíveis variações espectrais ao longo do tempo. Um algoritmo de detecção de picos é apresentado na Seção 2.3. Na Seção 2.4, são descritos algoritmos que procuram estimar a amplitude, a frequência e a fase dos picos detectados. O problema de avaliação de algoritmos de detecção de senóides é abordado nas Seções 2.5 e 2.6: na primeira, avalia-se apenas a capacidade de detecção da presença de componentes senoidais; na segunda, uma metodologia mais completa de avaliação também leva em conta a estimação de seus parâmetros.

2.2 Mapeamento Tempo-Frequência

O mapeamento tempo-frequência mais freqüente na literatura é a Transformada de Fourier de Tempo Curto (STFT) [51], que consiste em segmentar um sinal de áudio em quadros de curta duração, possivelmente sobrepostos no tempo, e depois calcular a Transformada de Fourier Discreta (DFT) [52] de cada quadro. Uma janela de suavização pode ser utilizada para reduzir os efeitos da segmentação abrupta no tempo. O comprimento do quadro influencia a resolução tempo-frequência da decomposição: quanto menor for o quadro, maior a resolução temporal, ao custo de uma pior resolução freqüencial, e vice-versa. Além do comprimento do quadro, o tipo de janela de suavização e a sobreposição temporal de quadros adjacentes também desempenham um papel importante na decomposição. Por exibir uma reduzida complexidade computacional, a STFT é vastamente utilizada, não só no processamento de áudio [53], mas no processamento de sinais em geral [51].

Para se definir a STFT matematicamente, primeiramente é construído um segmento do sinal $s[n]$ como

$$s_m[n] = s[n + mH], \quad n = 0, 1, \dots, N, \quad (2.1)$$

onde m é um índice que identifica o segmento, N é a duração do segmento em amostras e H é o salto, em amostras, entre segmentos consecutivos. Usualmente $H \leq N$, de modo que uma determinada amostra de $s[n]$ esteja contida em pelo menos um segmento. A STFT, na sua versão discreta, pode então ser escrita como:

$$S[m, k] = \text{STFT}\{s[n]\} = \frac{1}{N} \sum_{n=0}^{N-1} w[n] s_m[n] e^{-jk \frac{2\pi}{N} n}, \quad (2.2)$$

onde N é o comprimento da DFT utilizada para calcular o espectro do bloco de sinal de índice m , e $w[n]$ é uma janela de suavização [54] tal que $w[n] = 0$ para $|n| > N$.

O resultado da etapa de mapeamento tempo-frequência é uma matriz complexa com elementos, na linha m e na coluna k , $S[m, k]$ contendo o espectro estimado do sinal para cada quadro. Na matriz, m define o índice do quadro e k o índice da frequência.

Outros algoritmos propostos na literatura para este estágio incluem decomposições em multirresolução [23] e transformações bilineares [55]. Outra possível decomposição [56] utiliza o operador de reatribuição, que é apresentado na Seção 2.4.1.

2.3 Algoritmos para Detecção de Picos Espectrais

A etapa de detecção dos picos espectrais consiste em encontrar os picos no espectro que interessam à modelagem do sinal. Uma vez encontrados todos os picos espectrais, eles devem ser caracterizados como picos genuínos ou espúrios. Os picos genuínos são aqueles associados a uma componente senoidal de fato presente no sinal original. Os picos espúrios, ao contrário, decorrem de ruído presente no sinal ou de distorções provocadas pelo mapeamento tempo-freqüência, como, por exemplo, lobos laterais. Um bom detector, então, deve ser capaz de detectar corretamente todos os picos genuínos presentes no espectro (com mínimo percentual de falsos-positivos) e, ao mesmo tempo, rejeitar todos os picos espúrios (com mínimo percentual de falsos-negativos).

Diversas estratégias para discriminar entre picos genuínos e picos espúrios podem ser encontradas na literatura. Uma das mais triviais é selecionar apenas os P picos com maior magnitude no espectro de um quadro. A vantagem desta estratégia é limitar o número de picos a considerar, reduzindo a complexidade geral do sistema de análise. A desvantagem é a seleção indiferente de picos genuínos e espúrios.

Outra abordagem possível é considerar apenas os picos cujas magnitudes estão acima de um determinado limiar. Usualmente, o limiar é escolhido T dB abaixo do pico com maior magnitude [16]. Note que, por esse critério, picos espúrios podem ser detectados como genuínos em blocos que só contêm ruído. Ademais, o número total de senóides detectadas, neste caso, é variável e influenciado pelo parâmetro T .

O desempenho de um algoritmo de detecção de picos espectrais pode melhorar caso seja associado a um modelo psicoacústico, de maneira que apenas os picos espectrais audíveis sejam selecionados [57].

Os métodos descritos anteriormente utilizam apenas a informação da magnitude espectral para realizar a detecção. Em [14], é proposto um algoritmo que utiliza descritores complementares sobre um pico, como por exemplo largura de banda, coerência da freqüência e duração, para melhor discriminar entre os picos genuínos e os espúrios. Outra estratégia de detecção que pode ser mencionada é o chamado teste F [58], que projeta o sinal numa base ortonormal e utiliza critérios estatísticos para classificar os picos espectrais. Outro método, largamente utilizado pela comunidade de processamento de fala, é o teste de correlação cruzada [38].

A seguir, é apresentado um método de detecção de picos espectrais que utiliza apenas a informação da magnitude do espectro. Inicialmente, é descrito um pré-processamento que ajusta o espectro do sinal para a aplicação de um limiar; em seguida, adota-se um critério para a escolha deste limiar. Esse método foi apresentado inicialmente em [P2,P1]. O texto desta seção se baseia fortemente no de [P2].

2.3.1 Estratégia de Detecção de Picos

Como já foi visto, em esquemas de detecção de picos por limiar, o procedimento utilizado consiste em selecionar os picos espectrais que ultrapassem um determinado limiar pré-definido de energia. Embora simples, tal critério não é adequado quando utilizado para sinais de áudio, que, comumente, possuem um espectro em que a energia decresce com a frequência [59]. Devido a isso, muitos picos genuínos são descartados nas frequências altas. Este efeito é ilustrado na Figura 2.1, onde um limiar que se mostra adequado para a detecção de picos em baixas frequências não é adequado para os picos em altas frequências.

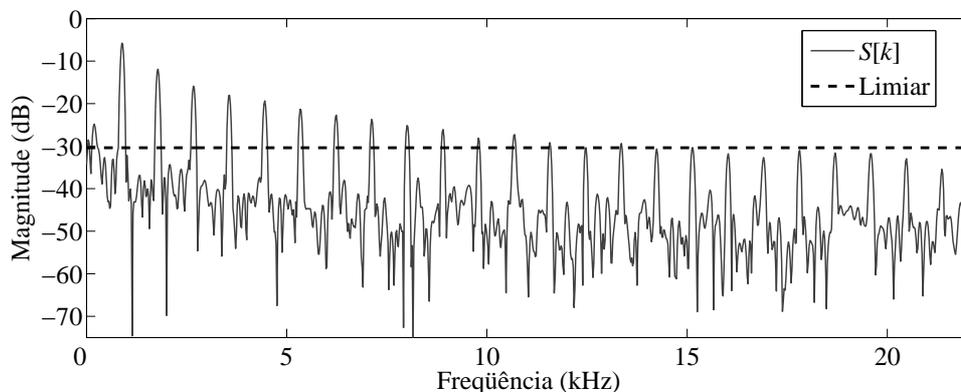


Figura 2.1: Limiar constante aplicado ao espectro de um sinal de áudio.

O método aqui proposto se baseia na seleção de picos espectrais através de um limiar variável de energia. A estratégia adotada se divide em duas etapas: um pré-processamento espectral e a seleção propriamente dita dos picos, como pode ser visto na Figura 2.2. Ambas as partes serão descritas nas seções seguintes.

O bloco de pré-processamento tenta compensar a tendência espectral do sinal a partir de uma estimativa da conformação espectral do chão de ruído do sinal.

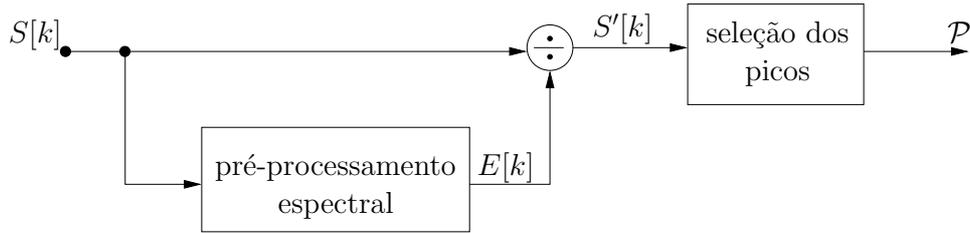


Figura 2.2: Diagrama de blocos ilustrando a estratégia de detecção de picos espectrais. Na figura, S denota o espectro original do sinal, E é uma aproximação da tendência espectral de S , S' é o espectro sem tendência espectral e \mathcal{P} é um conjunto contendo os picos espectrais selecionados.

Alternativamente, pode-se tentar estimar a tendência espectral do sinal a partir de uma estimativa da envoltória espectral da parte determinística do sinal [P2].

Uma vez que o método utiliza apenas a informação presente num quadro, para aliviar a notação utilizada o sub-índice m relativo ao quadro será omitido na discussão a seguir. Além disso, nesta seção $S[k]$ será utilizado para denotar a magnitude do espectro.

2.3.2 Pré-Processamento Espectral

As seções a seguir descrevem quatro métodos propostos na literatura para estimar o colorimento do ruído de fundo ou a envoltória espectral. Para os dois casos, a estimativa em questão será denominada $E[k]$.

Modelagem AR de Baixa Ordem

Modelos autorregressivos (AR) são amplamente utilizados em processamento de sinais de áudio, como, por exemplo, na modelagem do trato vocal em processamento de fala [60, 46].

O procedimento consiste em ajustar um modelo AR de baixa ordem ao sinal no domínio do tempo, $s[n]$. A envoltória desejada $E^{\text{ar}}[k]$ é a magnitude do espectro desse modelo AR de baixa ordem.

Em termos matemáticos, podemos assumir que o sinal $s[n]$ seja representado pelo seguinte modelo AR

$$s[n] = \sum_{u=1}^{q^*} a[u]s[n-u] + r[n], \quad (2.3)$$

onde q^* é a ordem do modelo suficiente para branquear o erro de modelagem $r[n]$. Para se obter a envoltória desejada, é necessário escolher uma ordem $q < q^*$. Usualmente, bons resultados são obtidos com $4 \leq q \leq 12$ para sinais de áudio.

Os parâmetros do modelo podem ser estimados através de qualquer estimador AR padrão [61], como os métodos de Burg [62] e Yule-Walker [63]. Uma vez obtido o modelo¹ $A(z) = [1 - \sum_{u=1}^q a_u z^{-u}]^{-1}$, a envoltória espectral pode ser obtida através de $E^{\text{ar}}[k] = |A(e^{j\omega_k})|$, onde $\omega_k = \frac{2\pi k}{N}$.

Two-Pass Split Window

A filtragem TPSW (do inglês, *Two-Pass Split Window*) foi originalmente proposta para estimação espectral de ruído em sistemas de sonar [64]. O procedimento pode ser decomposto em três etapas. Na primeira etapa, o espectro $S[k]$ é filtrado por uma *split window* descrita por [64]

$$h^{\text{sw}}[k] = \begin{cases} 0, & |k| < M^{\text{sw}} \\ 1, & M^{\text{sw}} \leq |k| < N^{\text{sw}}, \end{cases} \quad (2.4)$$

com $0 \leq M^{\text{sw}} < N^{\text{sw}}$, sendo N^{sw} e M^{sw} inteiros positivos que controlam, respectivamente, o comprimento da *split window* e a quantidade de zeros em seu interior.

A saída da primeira etapa $\tilde{S}[k]$ é, então, modificada de acordo com o critério

$$\hat{S}[k] = \begin{cases} S[k], & \text{if } S[k] \leq \alpha \tilde{S}[k] \\ \tilde{S}[k], & \text{if } S[k] > \alpha \tilde{S}[k], \end{cases} \quad (2.5)$$

onde $\alpha \geq 1$ é um parâmetro relacionado com a rejeição de picos no espectro observado.

No último estágio, o espectro modificado $\hat{S}[k]$, que supostamente está livre dos picos mais proeminentes do espectro $S[k]$, é filtrado por um filtro de média móvel convencional, com o mesmo comprimento da *split window*. A saída desta terceira etapa corresponde à estimativa desejada, $E^{\text{tpsw}}[k]$. Os filtros FIRs utilizados no primeiro e no terceiro estágio devem ser normalizados de modo a possuir ganho unitário em DC.

¹Ao contrário do que é indicado na Figura 2.2, $E^{\text{ar}}[k]$ é obtido diretamente de $s[n]$, sem requerer a computação de $S[k]$.

Na filtragem TPSW, considera-se que $S[k]$ está contido entre 0 e π . O espectro deve ser estendido em aproximadamente 20%, de modo a evitar efeitos de bordas durante as filtrações. Para isto, parte do espectro em ambas as extremidades é espelhada. Os atrasos causados pela filtragem são compensados de modo a garantir o sincronismo entre a entrada e a saída. Isto pode ser feito tomando-se apenas a parte central da correlação dos resultados da filtragem. Ao final, as extensões são descartadas de modo a retornar o espectro ao seu tamanho original.

Em relação aos parâmetros para a *split window*, a suavidade da estimativa $E^{\text{tpsw}}[k]$ estará associada diretamente ao comprimento da janela N^{sw} . O parâmetro M^{sw} deve ser escolhido de modo que o número de zeros na janela corresponda à largura (em número de *bins*) dos picos mais proeminentes. O valor de α deve ser pequeno o suficiente para garantir que a componente $\alpha\tilde{S}[k]$ esteja abaixo da amplitude média dos picos e grande o suficiente para que $\alpha\tilde{S}[k]$ fique acima do nível do ruído. Usualmente, são obtidos resultados satisfatórios para $2 \leq \alpha \leq 8$.

Estimação do Espectro Estocástico

O estimador do espectro estocástico (SSE) é outro estimador não-linear para o colorimento do espectro do ruído. O método foi apresentado em [65, 11].

Resumidamente, o método SSE consiste em:

1. Passar $S[k]$ por um filtro de média móvel de três coeficientes, de modo a obter $S^1[k]$ possivelmente livre de valores nulos;
2. Computar $R[k] = \frac{1}{S^1[k]}$;
3. Obter $R^1[k]$, uma versão suavizada de $R[k]$, através da convolução cíclica entre $R[k]$ e um filtro de média móvel FIR com N^{sse} coeficientes;
4. Calcular a estimativa desejada: $E^{\text{sse}}[k] = \frac{1}{R^1[k]}$.

Assim como o método TPSW, a suavidade de $E^{\text{sse}}[k]$ aumenta conforme o valor de N^{sse} . Além disto, todos os filtros FIR devem ser normalizados para ganho DC unitário. Diferentemente dos esquemas anteriores, no método SSE $S[k]$ deve ser considerado dentro do intervalo de $-\pi$ a π , para que se possa utilizar a convolução cíclica. O intervalo de 0 a π pode ser adotado, mas nesse caso a convolução

cíclica precisa ser substituída por uma filtragem convencional, adotando-se o mesmo esquema de extensão nas bordas utilizado pelo método TPSW.

Filtro Recursivo Não-Linear

Em [66] um filtro recursivo não-linear (NLRf) é proposto para estimar o espectro de ruído colorido de fundo na presença de picos espectrais. O filtro assume que a densidade espectral de potência do ruído varia lentamente em função da frequência. Com isto, o filtro limita em módulo a primeira derivada (*slew rate*) das amostras do espectro em função da frequência. O filtro recursivo não-linear que implementa esta solução é obtido através de

$$E^{\text{nrlf}}[k] = E^{\text{nrlf}}[k-1] \beta^{\text{sinal}(S[k]-E^{\text{nrlf}}[k-1])}, \quad (2.6)$$

onde $E^{\text{nrlf}}[k]$ é a estimativa desejada, β é uma constante pouco maior que a unidade e:

$$\text{sinal}(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ -1, & \text{se } x < 0 \end{cases}. \quad (2.7)$$

O parâmetro β pode ser escrito como $\beta = 1 + \lambda$. Em teoria, λ deve ser escolhido de modo a que se exceda o *slew rate* máximo associado à densidade de potência do espectro do ruído. Na prática, como pode ser visto na Equação (2.6), o valor de λ controla o fator de esquecimento do filtro. Com isso, quanto maior for o valor de λ maior será a variância de $E^{\text{nrlf}}[k]$. Uma estimativa suave o suficiente para a aplicação em questão pode ser obtida escolhendo-se $\lambda = 0,05$ [67].

Devido ao esquema de recursão, a inicialização do filtro deve ser feita com cautela. A inicialização inadequada do filtro pode levar a uma polarização da estimativa nas primeiras amostras de $E^{\text{nrlf}}[k]$ e, com isso, degradar o desempenho do algoritmo. Uma possível solução para a inicialização é estender $S[k]$ nas bordas, como descrito para o TPSW (neste caso a extensão só é necessária no início do espectro), e inicializar a recursão com $E^{\text{nrlf}}[-1] = S[0]$. A extensão do espectro deve ser longa o suficiente para que qualquer influência de uma má inicialização seja mitigada. Após a filtragem, os valores estimados para $E^{\text{nrlf}}[k]$ correspondentes à extensão do sinal são descartados. Alternativamente, $E^{\text{nrlf}}[k-1]$ pode ser escolhido como a mediana das primeiras C amostras de $S[k]$. Em ambos os casos, o algoritmo considera que o espectro $S[k]$ está limitado entre 0 e π radianos.

2.3.3 Correção do Espectro

Uma vez de posse da curva $E[k]$, a tendência espectral de um sinal pode ser corrigida através de

$$S'[k] = \frac{S[k]}{E[k]}, \quad (2.8)$$

onde $S'[k]$ é o espectro com a tendência espectral compensada. Deve-se notar que diferenças no nível médio de $E[k]$ não influenciam a etapa da detecção dos picos, uma vez que apenas seu formato é importante para a correção espectral.

Como um caso de estudo, os quatro métodos descritos anteriormente serão utilizados para se obter estimativas da tendência espectral de um sinal gerado artificialmente. O sinal em questão é composto por uma senóide com frequência fundamental de 1 kHz e harmônicos preenchendo o espectro até a frequência máxima de 22,05 kHz; com amplitudes decrescendo exponencialmente com a frequência. Ruído aditivo rosa foi utilizado para formar a parte estocástica do espectro.

Tabela 2.1: Parâmetros utilizados no caso de estudo para os métodos de obtenção da tendência espectral.

Método	Parâmetros
TPSW	$N^{\text{sw}} = 51$, $M^{\text{sw}} = 8$, e $\alpha = 4$
AR	$q = 10$
SSE	$N^{\text{sse}} = 101$
NLRF	$\beta = 1,01$

O espectro do sinal foi obtido pela DFT de 2048 pontos de um bloco de sinal de mesmo tamanho, multiplicado pela janela de Hann. Na Tabela 2.1, podem ser observados os parâmetros utilizados em cada método. Na Figura 2.3, vêem-se as curvas obtidas e curvas produzidas pelos estimadores de tendência espectral. Como pode ser visto, todos os métodos fornecem estimativas adequadas.

2.3.4 Critérios de Seleção de Picos

Como demonstrado na Figura 2.2, a seleção de picos é realizada sobre $S'[k]$ que, idealmente, é uma versão compensada do espectro com distribuição de energia (da

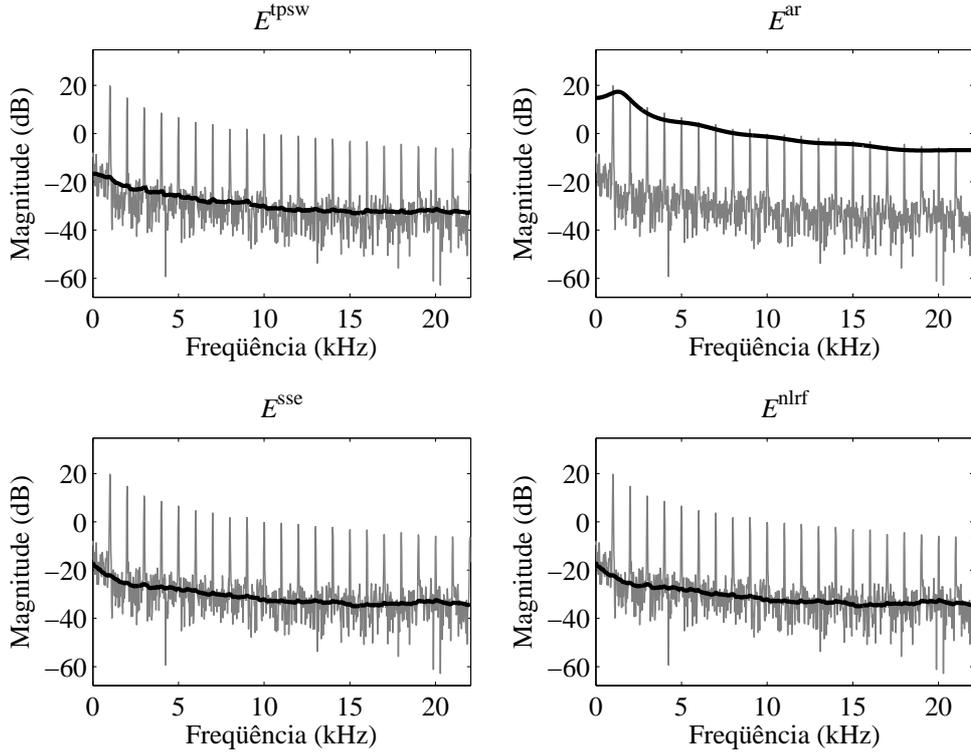


Figura 2.3: Estimativas da tendência espectral de um sinal sintético utilizando os métodos da Tabela 2.1.

parte estocástica) mais uniforme em função da frequência. A Figura 2.4 compara o espectro original $S[k]$ e sua versão compensada $S'[k]$, em um exemplo onde $E^{\text{sse}}[k]$ foi utilizada.

O critério para a seleção dos picos pode ser descrito da seguinte maneira, assumindo inicialmente a seqüência $k \in \{2, 3, \dots, (N/2 - 1)\}$ de índices de $S'[k]$: Para todos os elementos de k , agrupe em um subconjunto \mathcal{P}_m os índices k_{picos} que satisfazem simultaneamente as seguintes condições:

$$\begin{aligned}
 S'[k] &> S'[k - 1] \\
 S'[k] &> S'[k + 1] \\
 S'[k] &> d\mu.
 \end{aligned} \tag{2.9}$$

As duas primeiras condições garantem que os índices selecionados correspondem a picos espectrais. A terceira condição seleciona entre os picos espectrais aqueles que satisfazem um determinado critério de energia: ser superior a um limiar $d\mu$, sendo d um multiplicador escolhido empiricamente e μ uma estimativa do nível médio do

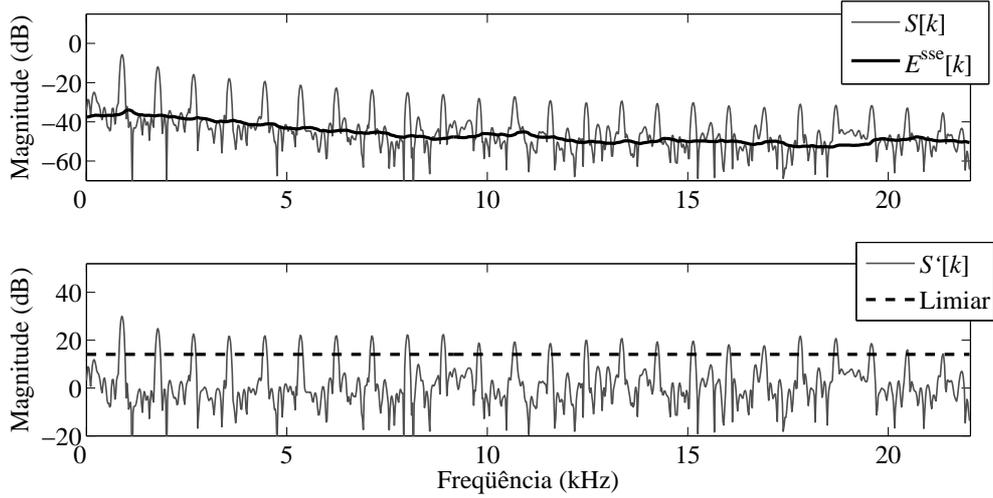


Figura 2.4: Comparação entre o espectro original (em cima) e sua versão corrigida (embaixo). Pode-se observar que a compensação do espectro favorece a detecção de picos com o uso de um limiar constante.

chão de ruído do espectro observado em $S'[k]$. Como o espectro utilizado é corrigido, o nível médio do chão de ruído pode ser considerado aproximadamente constante ao longo da frequência. O fator multiplicador d coloca o limiar d vezes acima do nível médio do chão de ruído em $S'[k]$ e deve ser escolhido de modo que o limiar $d\mu$ fique acima dos picos provocados pelo ruído e abaixo dos picos genuínos. Ao final, k_{picos} são os índices associados aos picos detectados em $S'[k]$.

O valor de μ pode ser obtido por qualquer estimador robusto o suficiente para prover estimativas para o nível médio de $S'[k]$, apesar da presença de muitos *outliers* (neste caso, os picos espectrais). Um possível estimador é a mediana de $S'[k]$. Porém, experimentalmente observou-se que, quando tratando um sinal densamente povoado por picos genuínos, a mediana fornece uma estimativa polarizada (superestimada) do nível médio. Um estimador que é menos suscetível a essa polarização surge a partir dos próprios métodos utilizados para o pré-processamento espectral (com exceção do modelo AR de baixa ordem). Dentre os métodos utilizados, o SSE mostrou ser o menos afetado pela presença de picos genuínos ao estimar o ruído de fundo. A estratégia para se estimar μ utilizando o estimador SSE é

$$\mu = \text{m\u00e9dia}(\bar{E}^{sse}[k]), \quad (2.10)$$

onde $\bar{E}^{sse}[k]$ é a curva fornecida pela aplicação do m\u00e9todo SSE ao espectro compen-

sado $S'[k]$. Este será o estimador utilizado nesta dissertação.

Posteriormente, observou-se que μ pode ser obtido *a priori* de uma forma mais simples a partir da própria curva $E[k]$ (com exceção da curva de $E^{\text{ar}}[k]$), que já é uma estimativa do nível médio *local* da parte estocástica de $S[k]$. Para isto, basta modificar a equação (2.8) para $S'[k] = \mu \frac{S[k]}{E[k]}$, sendo μ escolhido como um valor qualquer de $E[k]$, por exemplo, $\mu = \max(E[k])$.

Para o valor de d , pode-se assumir que o pré-processamento espectral foi bem-sucedido em “branquear” o espectro do ruído, e ainda que este pode ser considerado gaussiano. Assumindo que μ seja um estimador confiável do desvio-padrão deste ruído, ao se escolher $2 \leq d \leq 5$ estabelece-se um intervalo de confiança maior que 95% para que os picos espúrios estejam abaixo do limiar [68].

Na realidade, para uma dada razão sinal-ruído (SNR), quanto maior for o número de picos genuínos presentes no sinal, mais difícil será distinguir os picos genuínos do ruído de fundo, devido ao compartilhamento da energia do sinal entre os picos genuínos. Isto favorece a ocorrência de erros de detecção e requer uma escolha mais cuidadosa de d . De maneira inversa, a tarefa de escolher o limiar é simplificada quando o número de picos genuínos é menor, até mesmo para SNRs baixas.

As condições descritas acima motivam a seguinte estratégia para selecionar um valor adequado de d :

1. Calibrar d de modo a garantir uma detecção satisfatória considerando um sinal composto apenas por ruído;
2. Atribuir a $d_{\text{mín}}$ o valor de d obtido no item anterior;
3. Calcular $\rho = \max \left\{ 1, \left(\frac{\max(S'[k]) - \mu}{B} \right)^{0,5} \right\}$, com $S'[k]$, μ e B em dB;
4. Fazer $d = \rho d_{\text{mín}}$.

Apesar de adequado para casos com muitos picos espectrais, o valor de $d_{\text{mín}}$ tende a ser muito baixo para $S'[k]$ contendo poucos picos genuínos. Como consequência, muitas falsas detecções podem ocorrer.

No cálculo de ρ , $\max(S')$ e μ são expressos em dB. Assim, a quantidade $(\max(S'[k]) - \mu)$ pode ser interpretada como o intervalo (em magnitude) disponível

entre o máximo espectral e o nível médio do ruído. O parâmetro B pode ser interpretado como a tolerância máxima aceita para que a magnitude de um pico espúrio fique acima do nível médio estimado o chão de ruído. Tipicamente, o valor de B fica em torno de 10 dB. Logo, quando $(\max(S') - \mu) > B$, o multiplicador $\rho > 1$ contribui para elevar o limiar. Esta situação ocorre quando existem poucos picos genuínos em $S'[k]$ com magnitude bem acima do chão de ruído, evitando a ocorrência de falsos alarmes. Caso o intervalo seja menor que B dB, $\rho = 0$ e $d = d_{\min}$. Na Seção 2.5, será avaliado o desempenho do detector apresentado, para diferentes métodos de obtenção de tendência espectral.

2.4 Estimação de Parâmetros dos Picos

Algoritmos de refinamento de parâmetros são utilizados para melhorar a localização de componentes senoidais no plano tempo-freqüência, possivelmente aumentando a correção das trilhas senoidais obtidas. Os parâmetros que podem ser refinados para cada componente senoidal são sua amplitude, freqüência e fase. Algoritmos que realizam estes refinamentos são abordados nas Seções 2.4.1 e 2.4.2.

2.4.1 Técnicas de Estimação de Freqüência Instantânea

Como a detecção de picos é realizada no domínio da STFT, a medida da freqüência de qualquer pico detectado fica inicialmente restrita a valores pertencentes à grade uniformemente espaçada com valores $\omega_k = \frac{2\pi}{N}$, onde N é o tamanho da DFT usada na análise. Este espaçamento fixo impõe uma indesejável quantização dos valores assumidos pela freqüência instantânea $\Omega_l[n]$ da l -ésima componente senoidal. Em geral os algoritmos de estimação de freqüência instantânea procuram remover essa quantização indesejável, encontrando um valor de freqüência que melhor localize o máximo local (pico) do espectro na região de ocorrência de uma ressonância de interesse.

Pelo menos duas famílias de estimadores de freqüência são descritas na literatura: métodos que utilizam informação de fase do sinal e métodos baseados em interpolação espectral. Estes últimos [69, 70] procuram ajustar uma função analítica

(usualmente um polinômio) a amostras do espectro de potência no entorno de uma ressonância de interesse. Estimativas mais precisas para a frequência e a amplitude de um dado pico espectral são obtidas, então, através das coordenadas do valor máximo assumido pela função analítica.

Já a família de estimadores baseados na fase do sinal [71, 72, 73, 74, 12] procura estimar implicitamente a frequência através da derivada da fase instantânea do sinal. Matematicamente, definindo-se a STFT de tempo contínuo como

$$S(t, \omega) = \int_{-\infty}^{\infty} s(\tau) w(\tau - t) e^{-j\omega(\tau - t)} d\tau = A(t, \omega) e^{j\Phi(t, \omega)}, \quad (2.11)$$

onde $w(t)$ é a versão em tempo contínuo da janela de suavização e $s(t)$ é o sinal a ser modelado, a frequência instantânea [75, 9] de um sinal pode, então, ser obtida através de

$$\hat{\omega}(t, \omega) = \frac{\partial \Phi(t, \omega)}{\partial t}. \quad (2.12)$$

Diversas aproximações [71, 72, 73, 74, 12] para a derivada da equação (2.12) podem ser encontradas na literatura. A equivalência teórica de várias dessas aproximações é apresentada em [9] juntamente com uma comparação do desempenho dos diferentes estimadores obtidos a partir dessas aproximações. Outras comparações entre esses diferentes estimadores podem ser encontradas em [76] e em [77].

A seguir será apresentado o operador de reatribuição da frequência, um dos estimadores de frequência instantânea baseados na fase. A idéia é reatribuir os pontos da frequência do espectro amostrado para os centros de gravidade do espectro contínuo do sinal. Sua primeira formulação foi realizada para bancos de filtros em [78], e sua generalização para diversas distribuições tempo-frequência foi feita em [79, 80].

A derivação do estimador é

$$\hat{\omega}_r = \frac{\partial \Phi(t, \omega)}{\partial t} = \frac{\partial \Im\{\log(S(t, \omega))\}}{\partial t} = \Im\left\{\frac{1}{S(t, \omega)} \frac{\partial S(t, \omega)}{\partial t}\right\}, \quad (2.13)$$

onde $\Im(\cdot)$ denota a parte imaginária. Substituindo a expressão analítica para a STFT contínua (2.11) na expressão acima, obtém-se [9]

$$\hat{\omega}_r(t, \omega) = \omega - \Im\left\{\frac{S_{w'}(t, \omega)}{S(t, \omega)}\right\}, \quad (2.14)$$

onde $S_{w'}(t, \omega)$ é a STFT do sinal utilizando a janela $w'(t)$ (a derivada no tempo da janela $w(t)$). A versão discreta dessa estimativa é aproximada através de

$$\hat{\Omega}_r[m, k] = \frac{2\pi k}{N} - \Im \left\{ \frac{S_{w'}[m, k]}{S[m, k]} \right\}. \quad (2.15)$$

Neste caso, $S_{w'}[m, k]$ é o espectro discreto obtido utilizando-se a derivada da janela. A derivada da janela pode ser aproximada de maneira computacionalmente eficiente a partir da multiplicação do espectro da janela por uma rampa na frequência [73]. Com isso, a janela $w'[n]$ pode ser obtida através de

$$w'[n] = \text{IDFT} \left\{ \frac{jk}{F_s} \text{DFT}\{w[n]\} \right\}. \quad (2.16)$$

Outros métodos podem ser utilizados para se obter a derivada da janela, sendo este, no entanto, o mais difundido na literatura.

2.4.2 Técnicas de Estimação de Amplitude e de Fase

Conhecendo-se a frequência da senoide, sua amplitude pode ser obtida através de algoritmos de otimização [81, 15]. Em particular, considerando o sinal de entrada como um somatório de exponenciais complexas com fase e amplitude constantes em cada quadro m , diversos estimadores, com diferentes graus de complexidade, podem ser utilizados para obter a amplitude e a fase [81].

Inicialmente, o sinal num quadro m é modelado como

$$s_m[n] = \sum_{l=0}^{L-1} A_l e^{j\Omega_l n} + \nu[n], \quad n = 0, \dots, N-1, \quad (2.17)$$

onde A_l é a amplitude complexa da l -ésima senoide com frequência Ω_l conhecida, N é o comprimento do quadro e $\nu[n]$ é um ruído de observação, considerado estacionário e possivelmente colorido. Em forma matricial pode-se representar o problema através de

$$\begin{bmatrix} s_m[0] \\ s_m[1] \\ \vdots \\ s_m[N-1] \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ e^{j\Omega_1} & \cdots & e^{j\Omega_L} \\ \vdots & \vdots & \vdots \\ e^{j\Omega_1(N-1)} & \cdots & e^{j\Omega_L(N-1)} \end{bmatrix} \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_{L-1} \end{bmatrix} + \begin{bmatrix} \nu[0] \\ \nu[1] \\ \vdots \\ \nu[N-1] \end{bmatrix} \quad (2.18)$$

ou, mais compactamente,

$$\mathbf{s}_m = \mathbf{W}\mathbf{a} + \mathbf{v}. \quad (2.19)$$

O vetor \mathbf{a} que minimiza o erro quadrático da expressão acima é dado por [81]

$$\hat{\mathbf{a}} = (\mathbf{W}^H\mathbf{W})^{-1} \mathbf{W}^H\mathbf{s}_m, \quad (2.20)$$

onde $(.)^H$ denota o conjugado transposto de $(.)$. Devido à necessidade de inversão de matriz, esta solução é computacionalmente custosa. Caso as frequências das senóides sejam suficientemente espaçadas, é possível simplificar o resultado acima considerando a estimação da amplitude das senóides uma de cada vez. Simplificando a expressão (2.20) para $L = 1$, obtém-se

$$\hat{A}_l = \frac{1}{N} \sum_{n=0}^{N-1} s_m[n] e^{-j\Omega_l n}, \quad (2.21)$$

que nada mais é que a transformada de Fourier do sinal $s_m[n]$ avaliada na frequência Ω_l . Outros possíveis estimadores, que procuram modelar também o ruído de observação, são derivados em [81]. Em [82, 15], podem ser encontrados estimadores que utilizam modelos diferentes para o sinal num quadro. Neste trabalho, será utilizada a solução fornecida pela equação (2.21).

2.5 Comparação dos Algoritmos de Estimação da Tendência Espectral

Esta seção descreve um esquema experimental utilizado para avaliar o desempenho do sistema de detecção de picos. O objetivo é avaliar os diferentes estimadores de tendência espectral. Essa seção é fortemente baseada em [P2].

Como pode ser visto na Figura 2.5, o esquema de teste consiste de um gerador de sinais sintéticos, cuja saída $x[n]$ é fornecida a um analisador baseado na STFT que, por sua vez, fornece o espectro de magnitude $S[m,k]$ para o algoritmo de detecção de picos (ver Figura 2.2). De posse das frequências referenciais f_i de cada componente senoidal presente no sinal, um avaliador de desempenho analisa o resultado da detecção de picos em termos quantitativos, contando quantos picos foram corretamente identificados. O sistema emite três métricas de desempenho, κ , γ e ζ , que serão descritas em 2.5.3.

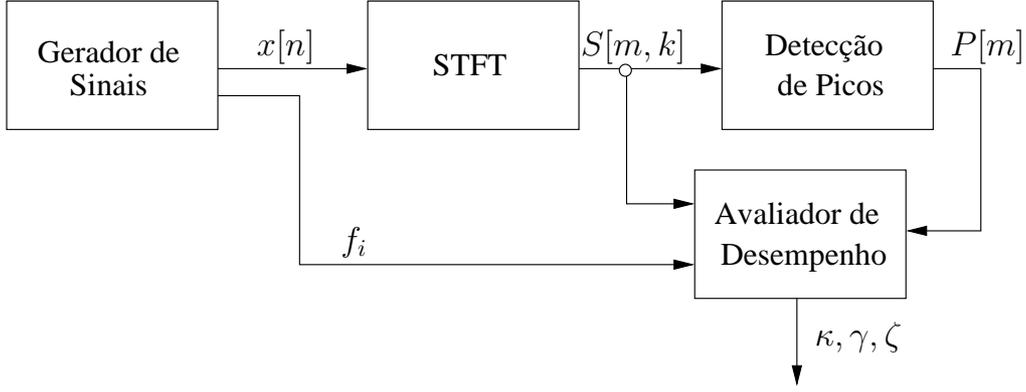


Figura 2.5: Esquema desenvolvido para avaliar o desempenho dos algoritmos de detecção de picos.

2.5.1 Geração do Sinal de Teste

O sinal de teste, amostrado a 44,1 kHz, é composto por duas componentes: uma determinística e outra estocástica. A primeira componente consiste na soma de sinais harmônicos com frequência fundamental $f_{0,i}$. Para cada $f_{0,i}$, que é escolhido aleatoriamente entre 200 Hz e 1 kHz, com probabilidade uniforme, harmônicos são gerados até a frequência de Nyquist. A amplitude dos harmônicos pode ser constante ou decrescente em função da frequência.

A componente estocástica do sinal de teste pode ser ruído aditivo branco ou rosa [83]. Em termos matemáticos, o sinal de teste é definido como

$$x[n] = \sum_{i=1}^I \sum_{j=1}^{J_i} a_{ij} \cos\left(\frac{2\pi j f_{0,i} n}{F_s}\right) + \sigma \eta[n]. \quad (2.22)$$

Os dois somatórios da equação (2.22) representam a componente determinística do sinal, onde F_s é a taxa de amostragem, I é o número de frequências fundamentais presentes na mistura e $J_i = \left\lfloor \frac{F_s}{2f_{0,i}} \right\rfloor$, com $\lfloor \cdot \rfloor$ denotando ‘o maior inteiro menor que ou igual a’, define o número de harmônicos associados a uma dada $f_{0,i}$. A amplitude dos harmônicos é escolhida como a unidade (no caso de amplitude constante) ou decrescente de acordo com $a_{ij} = \frac{1}{j f_{0,i}}$ (no caso de amplitude decrescente em função da frequência).

A segunda parte da equação (2.22) representa a componente de ruído, sendo $\eta[n]$ com $0 \leq n \leq (N - 1)$, uma realização de um processo estocástico [84] cuja densidade espectral de potência pode ser constante ou decrescente de acordo com

$1/f$. A variável σ controla a energia da componente estocástica de modo a forçar uma SNR desejada.

2.5.2 Esquema de Detecção de Picos

A computação de $S[m,k]$ e $S'[m,k]$ é idêntica ao procedimento descrito na Seção 2.3. Os parâmetros utilizados na STFT foram os mesmos que os utilizados no estudo de caso descrito naquela seção. Os parâmetros dos métodos da etapa de pré-processamento espectral exibidos na Tabela 2.1 também foram mantidos.

Quanto ao critério de seleção de picos (ver Seção 2.3.4), o nível médio do chão de ruído no espectro, μ , foi estimado conforme (2.10) com $N^{\text{sse}} = 150$. Ademais, verificou-se experimentalmente que $d_{\text{mín}} = 2$ era uma escolha adequada.

2.5.3 Avaliação de Desempenho

O desempenho da detecção de picos é obtido através da contagem dos picos corretamente detectados e dos falsos alarmes. Aqui, o principal problema a ser considerado é a *detecção* de um determinado pico, sem levar em consideração se a frequência e a amplitude foram precisamente estimadas. De qualquer maneira, ainda é necessário definir condições sob as quais um pico é considerado corretamente detectado.

Para o detector de picos, o domínio de observação é o espectro $S[m,k]$. Logo, a frequência de qualquer pico observado deve estar presente numa das raias de frequência disponíveis, por exemplo, kF_s/N , para $0 \leq k < N/2$. Já no lado da referência, a frequência dos picos pode ser escolhida com a resolução desejada. Para uma avaliação significativa de detectores de picos, os domínios frequenciais de observação (detecção de picos) e de referência (geração de picos) devem ser compatibilizados. A próxima seção apresenta a solução adotada para este problema de forma detalhada.

Alinhamento do Domínio da Referência

Considerando todas as Ω_i associadas à parte determinística do sinal de teste, o primeiro passo consiste em prever quais raias do espectro seriam mais fortemente ativadas pelas frequências Ω_i . Logo, cada Ω_i é quantizado para a frequência da raia

mais próxima, isto é, $\overline{\Omega}_i = k_i F_s / N$, para $k_i = \text{round}(\Omega_i N / F_s)$, onde N é o número de raias e F_s é a frequência de amostragem.

Armazenando todas as ocorrências não-repetidas de k_i no conjunto \mathcal{K} , pode-se definir um vetor de referência preliminar, cujos elementos são definidos como

$$r[k] = \begin{cases} 1, & \text{se } k \in \mathcal{K} \\ 0, & \text{em caso contrário} \end{cases}, \text{ para } 0 \leq k < \frac{N}{2}. \quad (2.23)$$

Deve-se notar, no entanto, que $r[k] = 1$ não necessariamente implica a observação de um pico em $S'[m, k]$, segundo os critérios (duas primeiras condições) de (2.9). Isto se deve ao fato de a posição de um máximo local em $S'[m, k]$, provocado por um determinado Ω_i , ser influenciada não apenas pelo ruído, mas também por outros Ω_i . Com isso, fica evidente que apenas a quantização na frequência não é suficiente para alinhar ambos os domínios. O principal problema, então, é definir uma máscara de referência que corresponda aos picos que são *verdadeiramente* observáveis em $S'[m, k]$.

Um alinhamento de domínio mais adequado pode ser criado a partir de um vetor binário auxiliar, cujos elementos são definidos como

$$c[k] = \begin{cases} 1, & \text{se } k \in \mathcal{O}_m \\ 0, & \text{em caso contrário} \end{cases}, \text{ para } 0 \leq k < \frac{N}{2}, \quad (2.24)$$

onde \mathcal{O}_m é um conjunto com cardinalidade $|\mathcal{O}_m|$ contendo todos os índices das raias associadas aos picos *observáveis* em $S'[m, k]$, sejam eles genuínos ou espúrios. Desta forma, o vetor de alinhamento é obtido através de

$$r_a[k] = r_0[k] \oplus r_{-1}[k] \oplus r_{+1}[k], \quad (2.25)$$

com $r_0[k] = r[k] \wedge c[k]$, $r_{-1}[k] = r[k-1] \wedge c[k]$, e $r_{+1}[k] = r[k+1] \wedge c[k]$, para $1 \leq k < (N/2 - 1)$, onde os símbolos \wedge e \oplus significam as operações booleanas ‘E’ e ‘OU EXCLUSIVO’, respectivamente.

As partes de $r_a[k]$ podem ser interpretadas da seguinte maneira:

- $r_0[k]$ anula $r[k]$ se $c[k] = 0$, isto é, quando raias ativas na referência não são picos observados;
- $r_{-1}[k]$ move uma raia ativa em $r[k]$ para a raia adjacente à esquerda se $c[k] = 0$ mas $c[k-1] = 1$;

- $r_{+1}[k]$ move uma raia ativa em $r[k]$ para a raia adjacente à direita se $c[k] = 0$ mas $c[k + 1] = 1$.

Métricas de Avaliação

Na prática, os índices k para os quais $r_a[k] = 1$ indicam as referências de localização com as quais confrontar os picos detectados em $S'[m,k]$. Uma vez que $r_a[k]$ é obtido, a porcentagem de picos corretamente detectados pode ser calculada através de

$$\gamma = \frac{G}{Q}, \quad (2.26)$$

onde $Q = \sum_{k=0}^{\frac{N}{2}} r_a[k]$ é a contagem dos picos no vetor de referência e $G = \sum_{k=0}^{\frac{N}{2}} g[k]$ é a contagem de todos os picos corretamente detectados, com

$$g[k] = \begin{cases} r_a[k], & \text{se } k \in \mathcal{P}_m \\ 0, & \text{em caso contrário} \end{cases}, \quad (2.27)$$

onde \mathcal{P}_m é um conjunto de cardinalidade $|\mathcal{P}_m|$ contendo todos os índices de todos os picos detectados em $S'[k,m]$.

A porcentagem de falsos alarmes pode ser calculada através de

$$\zeta = \frac{|\mathcal{P}_m| - G}{|\mathcal{O}_m| - Q}, \quad (2.28)$$

onde $|\mathcal{P}_m| - G$ é a contagem de todos os picos incorretamente detectados e $|\mathcal{O}_m| - Q$ é a contagem de todos os picos observáveis em $S'[k,m]$ que não deveriam ser detectados.

Ainda é possível definir a seguinte métrica

$$\kappa = \gamma - \zeta, \quad (2.29)$$

que agrega as duas métricas previamente definidas numa única figura de mérito. Idealmente, $-1 \leq \kappa \leq 1$, e um detector perfeito é alcançado quando $\kappa = 1$. Valores de κ próximos de 1 indicam a ocorrência de mais picos corretamente detectados do que falsos alarmes. Ao contrário, valores de κ próximos de -1 indicam uma contagem excessiva de falsos alarmes em comparação com o número de picos corretamente detectados.

2.5.4 Resultados

Todos os quatro métodos de compensação de tendência espectral apresentados na Seção 2.3.3 foram testados com um mesmo conjunto de sinais de teste, de modo a investigar seus efeitos sobre o desempenho de detecção de picos. Nos experimentos aqui realizados foram utilizados os parâmetros de processamento indicados na Tabela 2.1 (pré-processamento) e na Seção 2.3.3 (critérios de detecção de picos). Além disso, os resultados obtidos sem o estágio de pré-processamento também foram examinados, sendo indicados como ‘Nenhum’.

Cada sinal de teste foi criado de modo a acomodar 10 quadros de análise. Para cada teste, 500 realizações foram geradas. Os resultados obtidos correspondem à média dos indicadores de desempenho medidos para cada quadro do sinal. Por conveniência, os valores de γ e ζ são expressos em %. Três casos de teste serão apresentados.

O **Teste 1** procura avaliar o desempenho dos detectores de picos para diferentes valores de SNR. O sinal de teste escolhido consiste de uma única série harmônica ($I = 1$) com fundamental $f_{0,i}$, tal que $a_{ij} = \frac{1}{j f_{0,i}}$ (ver Seção 2.5.1), adicionada de ruído rosa. A SNR vai de 0 até 20 dB em passos de 5 dB. Os resultados obtidos podem ser vistos nas Tabelas 2.2 (a) a 2.2 (c).

Os valores de γ demonstram que a maioria dos picos são corretamente classificados mesmo em condições de SNR baixa. Como esperado, quanto menor o valor da SNR, maior o valor obtido para ζ , demonstrando uma tendência à detecção de picos espúrios. Todos os 4 métodos obtiveram resultados similares de κ , demonstrando que todos operam similarmente bem sob condições idênticas de teste. É importante ressaltar os valores mais baixos para κ quando nenhum esquema de compensação de tendência espectral é utilizado.

O **Teste 2** mede o desempenho dos algoritmos de detecção de picos com respeito à variação do número de sinais harmônicos na mistura. Usaram-se sinais de teste contendo $3 \leq I \leq 7$ frequências fundamentais, com $a_{ij} = \frac{1}{j f_{0,i}}$ (ver 2.5.1), adicionados de ruído rosa para uma SNR = 10 dB.

As Tabelas 2.3 (a) a 2.3 (c) resumem os resultados encontrados. Pode ser observado que γ tende a decrescer com o aumento de I . Quanto a ζ , com exceção do método AR, a métrica tende a crescer conforme I aumenta. Dentre os métodos

Tabela 2.2: Indicadores de desempenho obtidos no **Teste 1**.

(a)

Pré-proc. \ SNR(dB)	γ (%)				
	0	5	10	15	20
TPSW	85,6	98,9	99,3	99,5	99,6
AR	88,9	99,7	99,8	99,8	99,8
NLRF	83,3	97,1	99,2	99,3	99,2
SSE	85,6	97,9	99,2	99,3	99,3
Nenhum	38,3	63,0	84,9	93,6	98,5

(b)

Pré-proc. \ SNR(dB)	ζ (%)				
	0	5	10	15	20
TPSW	17,5	4,6	0,5	0,3	0,2
AR	22,5	13,4	1,4	0,2	0,1
NLRF	15,0	1,7	0,7	0,4	0,29
SSE	18,1	2,5	0,7	0,5	0,3
Nenhum	6,3	2,3	0,9	0,4	0,2

(c)

Pré-proc. \ SNR(dB)	κ				
	0	5	10	15	20
TPSW	0,68	0,94	0,99	0,99	0,99
AR	0,66	0,86	0,98	1,00	1,00
NLRF	0,68	0,95	0,99	0,99	0,99
SSE	0,68	0,95	0,99	0,99	0,99
Nenhum	0,32	0,61	0,84	0,93	0,98

testados, o método NLRFF obteve o pior desempenho, enquanto o método AR obteve o melhor, demonstrando alguma robustez deste método à densidade de componentes senoidais no sinal.

O **Teste 3** procura estressar os métodos com a substituição do ruído rosa anteriormente utilizado por ruído branco gaussiano. Neste teste, $I = 3$, $a_{ij} = \frac{1}{j f_{0,i}}$ e $\text{SNR} = 10$ dB. Os resultados são exibidos na Tabela 2.4. Como pode ser observado, o desempenho de todos os métodos piora sob estas condições de teste. O método AR exibe altos valores para γ , uma vez que $E^{\text{ar}}[k]$ tende a decair com a frequência, favorecendo a detecção de picos genuínos imersos no ruído. O efeito colateral é um alto valor para ζ que acaba por reduzir o valor obtido para κ . De maneira geral, de acordo com κ , o método TPSW obteve o melhor desempenho neste teste.

De forma geral, os testes executados demonstraram que a utilização de algum método de pré-processamento leva a uma melhora no desempenho da detecção dos picos. No entanto, os resultados dos testes não permitem afirmar que o desempenho de um determinado método de estimação de tendência espectral é superior ao desempenho dos demais.

2.6 Avaliação Generalizada de Detectores e Estimadores de Picos Espectrais

Na seção anterior, foi utilizado um esquema de avaliação de desempenho que considerou apenas a detecção das senóides propriamente dita sem levar em conta os erros na estimação dos parâmetros senoidais. Nesta seção, é revisada a proposta apresentada em [45] de um algoritmo de avaliação da etapa de detecção de senóides que também contempla a estimação de seus parâmetros.

Usualmente, algoritmos de detecção de senóides são avaliados utilizando metodologias similares à apresentada na Seção 2.5. As figuras de mérito escolhidas são tipicamente os percentuais de picos corretamente detectados e os de falso positivo [85, 76]. Já para a etapa de estimação de parâmetros é possível [45] obter limites teóricos para os estimadores, desde que o caso de teste seja extremamente simplificado. Os critérios mais comuns para a avaliação dos estimadores são a média e a variância do erro [76, 75, 9].

Tabela 2.3: Indicadores de desempenho obtidos no **Teste 2**.

(a)

Pré-proc. \ I	γ (%)				
	3	4	5	6	7
TPSW	97,2	91,6	85,7	78,0	71,0
AR	98,3	92,5	86,4	78,4	71,4
NLRF	91,4	84,6	79,4	72,5	66,4
SSE	94,3	88,5	82,9	75,5	69,0
Nenhum	48,4	38,2	33,1	25,6	24,0

(b)

Pré-proc. \ I	ζ (%)				
	3	4	5	6	7
TPSW	3,0	3,5	3,6	3,8	4,0
AR	4,5	4,3	3,9	3,7	3,5
NLRF	2,1	3,6	4,5	5,3	5,8
SSE	2,0	2,9	3,3	3,8	4,1
Nenhum	2,2	2,8	3,3	3,9	4,5

(c)

Pré-proc. \ I	κ				
	3	4	5	6	7
TPSW	0,94	0,88	0,82	0,74	0,67
AR	0,94	0,88	0,83	0,75	0,68
NLRF	0,89	0,81	0,75	0,67	0,61
SSE	0,92	0,86	0,80	0,72	0,65
Nenhum	0,46	0,35	0,30	0,25	0,21

Tabela 2.4: Indicadores de desempenho obtidos no **Teste 3**.

	Método				
	TPSW	AR	NLRF	SSE	Nenhum
γ (%)	64,2	76,9	44,1	41,6	38,9
ζ (%)	11,0	31,5	2,1	2,1	0,2
κ	0,53	0,46	0,42	0,39	0,39

O objetivo do avaliador que será apresentado é permitir uma análise conjunta de desempenho das etapas de detecção e estimação de parâmetros de senóides. Na Seção 2.6.1, é feita uma introdução da metodologia de avaliação de algoritmos de detecção de senóides originalmente apresentada em [45]. Já na Seção 2.6.2, é proposta uma modificação para o avaliador de [45], de modo a corrigir um defeito/ponto fraco nele observado.

2.6.1 Metodologia de Avaliação

A metodologia apresentada em [45] tem como objetivo permitir a avaliação comparativa de algoritmos de detecção/estimação de senóides. Uma visão geral da metodologia pode ser observada na Figura 2.6, onde o bloco ‘Detecção de Senóides’ contempla ambas as etapas de detecção e de estimação dos parâmetros. A metodologia pode ser resumida nos passos descritos abaixo.

1. Definição de um cenário de teste.
2. Determinação do conjunto de picos senoidais de referência $\overline{\mathcal{P}}$ e do sinal $s[n]$ a partir do cenário de teste.
3. Obtenção do conjunto de senóides detectadas \mathcal{P} .
4. Comparação do conjunto de senóides obtidas com o conjunto de referência.
5. Cálculo das figuras de mérito.

O avaliador utilizado e as figuras de mérito são descritos na próxima seção. Já os conjuntos $\overline{\mathcal{P}}$ e \mathcal{P} são definidos a seguir.

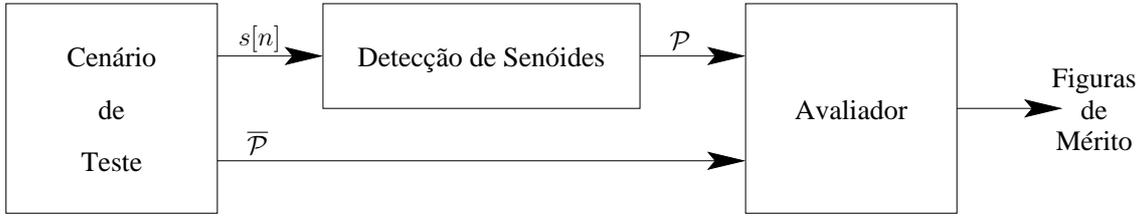


Figura 2.6: Metodologia de avaliação de algoritmos de detecção/estimação de senóides [45].

O conjunto de referência, $\overline{\mathcal{P}}$, pode ser definido arbitrariamente ou obtido de sinais naturais previamente analisados, dependendo do cenário de teste escolhido. Analogamente, o sinal a ser analisado $s[n]$ pode ser gerado artificialmente ou ser um sinal natural. No Capítulo 4 é apresentado um banco de trilhas senoidais cujo objetivo é disponibilizar dados de teste variados para sistematizar a avaliação de desempenho de sistemas de análise senoidal. O cenário de teste determina tipos e níveis de dificuldades aos quais o algoritmo de detecção/estimação de senóides sob teste será submetido. Diferentes objetivos poderiam ser contemplados pelos cenários de teste, por exemplo:

- avaliação da robustez do algoritmo a picos provocados pela parcela não ressonante do sinal;
- avaliação da capacidade do algoritmo de detectar duas (ou mais) senóides cujas frequências são próximas;
- avaliação da precisão da estimação da frequência.

No Capítulo 5, discutem-se mais detalhadamente os possíveis cenários de teste.

Para que os diferentes algoritmos de detecção/estimação de senóides possam ser comparados, é necessário definir o que o avaliador considerará como uma senóide. Desta maneira, para fins de avaliação, uma senóide detectada é definida através dos seguintes parâmetros

$$p_{i,m} = (f_{i,m}, a_{i,m}, \phi_{i,m}), \quad (2.30)$$

onde m é o índice do quadro em que esta senóide foi detectada, i é um índice que identifica esta senóide no quadro e $f_{i,m}$, $a_{i,m}$ e $\phi_{i,m}$ são os valores estimados para a frequência, a magnitude e a fase da senóide, respectivamente. Aqui, assume-se

um sinal sob análise com excursão (de amplitude) entre -1 e 1. Logo, o parâmetro a situa-se no intervalo $(0,1]$ em escala linear. A frequência também é considerada normalizada entre 0 (DC) e 1 (metade da frequência de amostragem).

Pode-se, então, definir o conjunto

$$\mathcal{P}_m = \bigcup_i \{p_{i,m}\} \quad (2.31)$$

que contém todas as senóides detectadas no quadro m . O conjunto \mathcal{P} , definido como

$$\mathcal{P} = \bigcup_m \mathcal{P}_m, \quad (2.32)$$

contém os picos detectados em todo o sinal. O conjunto de referência é definido analogamente, porém é denotado como $\overline{\mathcal{P}}$. Define-se ainda $\overline{\mathcal{P}}_m$ correspondentemente, de forma que os elementos de \mathcal{P}_m e $\overline{\mathcal{P}}_m$ se referem a um mesmo bloco m .

2.6.2 Avaliador

O avaliador de desempenho consiste num algoritmo que associa os elementos do conjunto de referência $\overline{\mathcal{P}}$ aos elementos do conjunto sob teste \mathcal{P} . É atribuído um custo a cada uma dessas associações e são obtidas figuras de mérito para quantificar a diferença entre \mathcal{P}_m e $\overline{\mathcal{P}}_m$. O avaliador apresentado é uma modificação do proposto em [45]. As modificações realizadas sobre o avaliador serão explicitadas ao longo de sua descrição.

As diferenças entre \mathcal{P}_m e $\overline{\mathcal{P}}_m$ podem ser tipificadas em três figuras de mérito:

- Erro de Falso Negativo (FN) – quando um pico está presente em $\overline{\mathcal{P}}_m$ mas não em \mathcal{P}_m .
- Erro de Falso Positivo (FP) – quando um pico está presente em \mathcal{P}_m mas não em $\overline{\mathcal{P}}_m$.
- Erro de Associação (EA) – erro cometido ao se associar um pico p_i (em \mathcal{P}_m) com frequência f_i no lugar de um pico p_j (em $\overline{\mathcal{P}}_m$) com frequência f_j .

Inicialmente, é necessário definir quando um elemento de \mathcal{P}_m poderá ser associado a um elemento de $\overline{\mathcal{P}}_m$. Em [45], esta associação é feita por um critério que considera exclusivamente a maior proximidade frequencial entre dois picos. Entretanto,

não se define um limite máximo aceitável para esse afastamento. Um problema decorrente disto é a possibilidade de associação (válida, mas indesejável) entre dois picos muito distantes na frequência. Uma ocorrência desse tipo pode ter natureza propagatória. Para evitar esta propagação de erro, é utilizado aqui um critério de proximidade em frequência em que uma associação entre picos só é aceita se seu afastamento frequencial for menor que um limite máximo pré-estabelecido. Este critério é inspirado em algoritmos de rastreamento de senóides, que serão apresentados no Capítulo 3. Considerando que a distância relativa entre os picos $p_{j,m} \in \overline{\mathcal{P}}_m$ e $p_{i,m} \in \mathcal{P}_m$ é

$$d(p_{i,m}, p_{j,m}) = \frac{|f_{i,m} - f_{j,m}|}{f_{j,m}}, \quad (2.33)$$

um pico $p_{j,m}$ só pode ser associado a um pico $p_{i,m}$ se

$$d(p_{i,m}, p_{j,m}) \leq \Lambda_f, \quad (2.34)$$

onde Λ_f é um limiar a ser determinado. Caso o critério seja satisfeito, o custo da associação é o próprio $d(p_{i,m}, p_{j,m})$. É necessário ressaltar que Λ_f deve ser escolhido com cuidado devido ao seu grande impacto sobre as figuras de mérito fornecidas pelo avaliador. No Capítulo 5, é apresentado um esquema experimental para se determinar um valor adequado para Λ_f . O uso deste limiar é a principal modificação realizada sobre o avaliador proposto em [45].

A estratégia de associação entre os elementos de \mathcal{P}_m e $\overline{\mathcal{P}}_m$ envolverá um procedimento iterativo, guiado pela proximidade frequencial entre os picos. Sua implementação segue o Algoritmo 2.1, que utiliza cinco variáveis auxiliares:

- $\overline{\mathcal{P}}_r$ – contém os picos do conjunto de referência que ainda não foram associados;
- \mathcal{P}_r – contém os picos do conjunto sob teste que não ainda foram associados;
- e_m^s – acumula o erro de cada associação;
- e_m^f – é incrementado de 1 toda vez que um elemento de $\overline{\mathcal{P}}_r$ não consegue ser associado a um elemento de \mathcal{P}_r ;
- T_a – armazena o número de associações realizadas pelo algoritmo.

Os passos do algoritmo dentro do *loop* ‘enquanto’ podem ser resumidos da seguinte forma:

1. Escolher-se o pico p_j do conjunto auxiliar $\overline{\mathcal{P}}_r$ com maior amplitude.
2. Escolher-se o pico p_i do conjunto \mathcal{P}_r que minimiza a função (2.33) e obedece à desigualdade (2.34).
3. Caso um pico p_i seja encontrado, a associação entre p_j e p_i é realizada e o custo da associação é acumulado em e_m^s . A variável T_a é incrementada, indicando que mais uma associação ocorreu. Em seguida, p_i é removido de \mathcal{P}_r de modo a evitar que ele seja associado a outros picos.
4. Caso um pico p_i não seja encontrado, e_m^f é incrementado de uma unidade, indicando que um elemento de $\overline{\mathcal{P}}_r$ não conseguiu se associar a nenhum elemento de \mathcal{P}_r .
5. Ao final, p_j é removido do conjunto $\overline{\mathcal{P}}_r$.

O *loop* termina quando o conjunto $\overline{\mathcal{P}}_r$ fica vazio.

Após a execução do Algoritmo 2.1, o conjunto \mathcal{P}_r pode não estar vazio. Isso indica que alguns picos deste conjunto não foram associados a nenhum elemento do conjunto de referência. Os picos remanescentes em \mathcal{P}_r , então, podem ser considerados como falsos positivos. Assim, pode-se assumir que a quantidade de falsos negativos no quadro m é igual à cardinalidade do conjunto \mathcal{P}_r ao final do algoritmo.

Utilizando-se os três erros obtidos pelo algoritmo, três figuras de mérito podem ser obtidas:

$$EA_m = \frac{e_m^s}{T_a} \quad (2.35)$$

$$FP_m = \frac{|\mathcal{P}_r|}{|\mathcal{P}_m|} \quad (2.36)$$

$$FN_m = \frac{e_m^f}{|\overline{\mathcal{P}}_m|}, \quad (2.37)$$

onde $|\mathcal{A}|$ indica a cardinalidade do conjunto \mathcal{A} . Supondo um total de M quadros analisados, pode-se obter três medidas consolidadas do desempenho de um algoritmo

Algoritmo 2.1 – Associação dos elementos de \mathcal{P}_m aos elementos de $\overline{\mathcal{P}}_m$

$$e_m^s \leftarrow 0$$

$$e_m^f \leftarrow 0$$

$$T_a \leftarrow 0$$

$$\overline{\mathcal{P}}_r \leftarrow \overline{\mathcal{P}}_m$$

$$\mathcal{P}_r \leftarrow \mathcal{P}_m$$

enquanto $\overline{\mathcal{P}}_r \neq \emptyset$ **faça**

 escolha $p_j \in \overline{\mathcal{P}}_r$ tal que $a_j = \max_{p_k \in \overline{\mathcal{P}}_r} a_{k,m}$

 encontre $p_i \in \mathcal{P}_r$ tal que $d(p_{i,m}, p_{j,m}) = \min_{p_k \in \mathcal{P}_r} d(p_{k,m}, p_{j,m})$

se $d(p_{i,m}, p_{j,m}) \leq \Lambda_f$ **então**

$$e_m^s \leftarrow e_m^s + d(p_{i,m}, p_{j,m})$$

$$T_a \leftarrow T_a + 1$$

$$\mathcal{P}_r \leftarrow \mathcal{P}_r - \{p_i\}$$

senão

$$e_m^f \leftarrow e_m^f + 1$$

fim se

$$\overline{\mathcal{P}}_r \leftarrow \overline{\mathcal{P}}_r - \{p_j\}$$

fim enquanto

de detecção de senóides, tais como

$$EA = \frac{1}{M} \sum_{m=1}^M EA_m \quad (2.38)$$

$$FP = \frac{1}{M} \sum_{m=1}^M FP_m \quad (2.39)$$

$$FN = \frac{1}{M} \sum_{m=1}^M FN_m. \quad (2.40)$$

Abaixo, podem ser vistos alguns exemplos de como essas figuras de mérito podem ser interpretadas.

- $EA = 0,1$ indica que, em média, um erro relativo de 10 % foi cometido ao se estimar a frequência dos picos.
- $FP = 0,2$ indica que, em média, 20 % dos picos detectados no conjunto sob teste eram falsos positivos.
- $FN = 0,2$ indica que, em média, 20 % dos picos do conjunto de referência não foram detectados.

Os erros FP e FN ainda podem ser unificados em apenas uma figura de mérito, relacionada à quantidade de picos que não foram corretamente detectados. O erro de detecção (ED) é definido como

$$ED = \frac{FP + FN}{2}. \quad (2.41)$$

Desta forma, o avaliador fornece duas figuras de mérito: uma relacionada ao número de picos que foram erroneamente detectados e outra relacionada com o desempenho da estimação da frequência. Esta também é uma modificação sobre o avaliador proposto em [45], que fornece apenas uma única figura de mérito.

Deve-se notar que a figura de mérito ED calculada pelo avaliador apresentado nesta seção é similar às métricas fornecidas pelo método de avaliação descrito na Seção 2.5. Em última análise, é possível que os resultados fornecidos pelo método da Seção 2.5 possam ser replicados pelo avaliador aqui proposto, dada uma escolha criteriosa de Λ_f . O avaliador descrito nesta seção, no entanto, é mais geral, pois além de considerar ED ele também mede EA.

O avaliador descrito nesta seção será utilizado nos experimentos descritos no Capítulo 5.

Capítulo 3

Algoritmos de Rastreamento de Trajetórias Senoidais

3.1 Introdução

O objetivo deste capítulo é apresentar algoritmos para o rastreamento de trajetórias senoidais. Cada uma dessas trajetórias, também chamadas de trilhas senoidais, representa uma senóide modulada em amplitude e em frequência utilizada no modelo senoidal que foi apresentado no Capítulo 1.

Até esse ponto, na análise senoidal, apenas a informação da localização das componentes senoidais detectadas na forma de pontos no plano tempo-frequência é conhecida. O problema de ligação de picos espectrais pode, então, ser descrito como uma busca que associa os picos (os pontos no plano tempo-frequência) encontrados num quadro m aos picos encontrados no quadro $m + 1$, tipicamente utilizando algum critério de proximidade de amplitude e de frequência, como será visto mais tarde. Dessa maneira, ao se ligar os pontos no plano tempo-frequência, obtém-se o comportamento da amplitude e da frequência ao longo do tempo para cada componente senoidal.

Para ilustrar o problema de rastreamento de parciais, a Figura 3.1 exibe para cada quadro no tempo a frequência dos picos encontrados para um sinal de trompete tocando a nota Mi bemol (segunda oitava, 155 Hz) com *vibrato*. O sinal foi gerado artificialmente utilizando-se um sintetizador de sinais de sopro [86]. Algumas das dificuldades enfrentadas pelos algoritmos de rastreamento de trilhas podem ser

observadas neste exemplo:

1. a variação dos parâmetros ao longo do tempo (neste caso, a frequência) que dificulta a ligação dos picos através de um critério de proximidade dos parâmetros;
2. a falta de picos em alguns quadros e em determinadas regiões da frequência;
3. a presença de picos espúrios.

Além das dificuldades citadas, no caso de sinais com a ocorrência simultânea de mais de um nota musical, pode ocorrer o cruzamento na frequência de parciais.

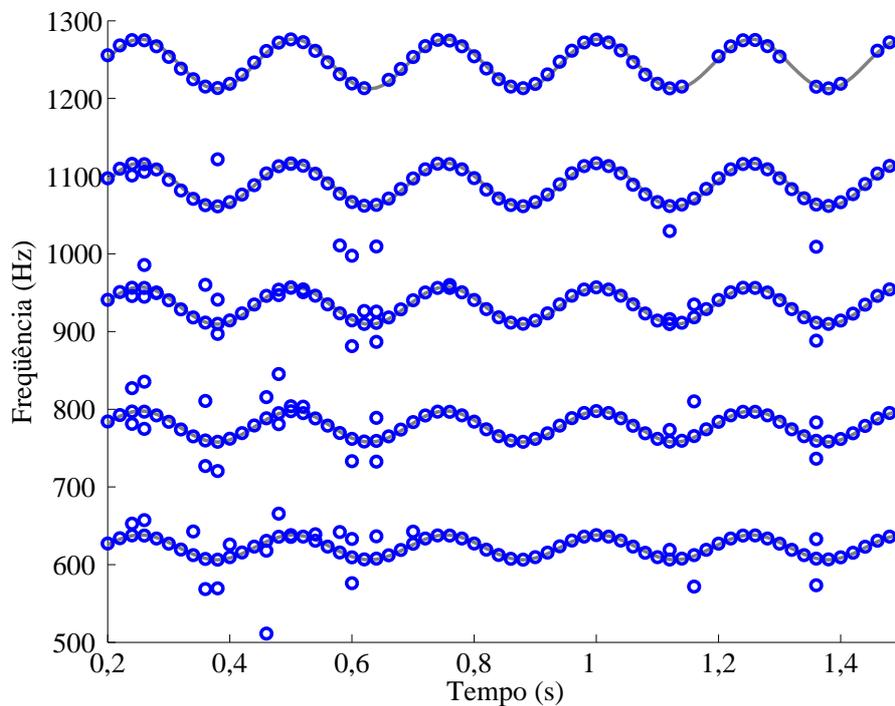


Figura 3.1: Frequência dos picos ('o') detectados em cada quadro de um sinal sintético de trompete exibindo *vibrato*. As linhas correspondem à trajetória na frequência das trilhas utilizadas para gerar o sinal.

O problema de encontrar as ligações que minimizam algum critério de proximidade entre os picos considerando-se todos os picos encontrados em todos os quadros do sinal tem complexidade computacional exponencialmente crescente com o número de picos [39]. Pode-se pensar em reduzir esta complexidade computacional procurando-se as continuações dos picos para um número reduzido de quadros ou, de maneira mais restrita, apenas procurar as associações de picos em quadros adjacentes. A solução usualmente mais adotada é restringir a busca aos quadros adjacentes,

o que, além de reduzir a complexidade do problema, permite a execução *online* do algoritmo de rastreamento. Observando-se a Figura 3.1, um critério aparentemente adequado para restringir ainda mais a busca pela continuação de uma trilha seria a proximidade na frequência. Ainda assim, uma determinada região da frequência pode possuir mais de um pico como possível candidato para a continuação de uma trilha. Neste caso, o pico escolhido como continuação para uma trilha costuma ser o pico que minimiza alguma função da distância entre a frequência, amplitude ou algum outro parâmetro do pico medido e os parâmetros correspondentes da trilha. Para ilustrar isso, a Figura 3.2 mostra uma faixa de frequências que isolaria os picos que formam, aproximadamente, o 8º. harmônico do exemplo anterior, i.e. aqueles cujas frequências se situam no intervalo [900; 960] Hz. Para este exemplo, pode-se observar que um critério que utilizasse a proximidade da amplitude dos picos seria capaz de rejeitar os picos espúrios. Na Figura 3.2, os pontos marcados ‘A’ e ‘B’ ilustram situações em que isso ocorre.

As soluções descritas neste capítulo para o problema de rastreamento de trilhas utilizam a proximidade de frequência e de amplitude para encontrar a melhor continuação de uma determinada trilha. Além disso, um esquema utilizando predição linear é apresentado. Idealmente, este esquema aumenta a robustez do algoritmo de rastreamento a picos espúrios e melhora o desempenho para sinais exibindo trilhas com trajetórias próximas na frequência.

Este capítulo se organiza da seguinte forma: após uma revisão bibliográfica de métodos de rastreamento de trilhas, é apresentado na Seção 3.2 o algoritmo utilizado como ponto de partida para o trabalho realizado nesta dissertação. Em seguida, na Seção 3.3 a solução proposta é apresentada. Os algoritmos de predição linear utilizados neste trabalho são detalhados na Seção 3.4. Já a Seção 3.5 apresenta algumas heurísticas para a decisão da escolha do pico que sirva de melhor continuação para uma determinada trilha. Por fim, o capítulo termina com a descrição de um avaliador de desempenho para algoritmos de rastreamento de trilhas.

3.1.1 Revisão Bibliográfica

Nesta seção, os principais métodos de rastreamento de trilhas serão brevemente descritos.

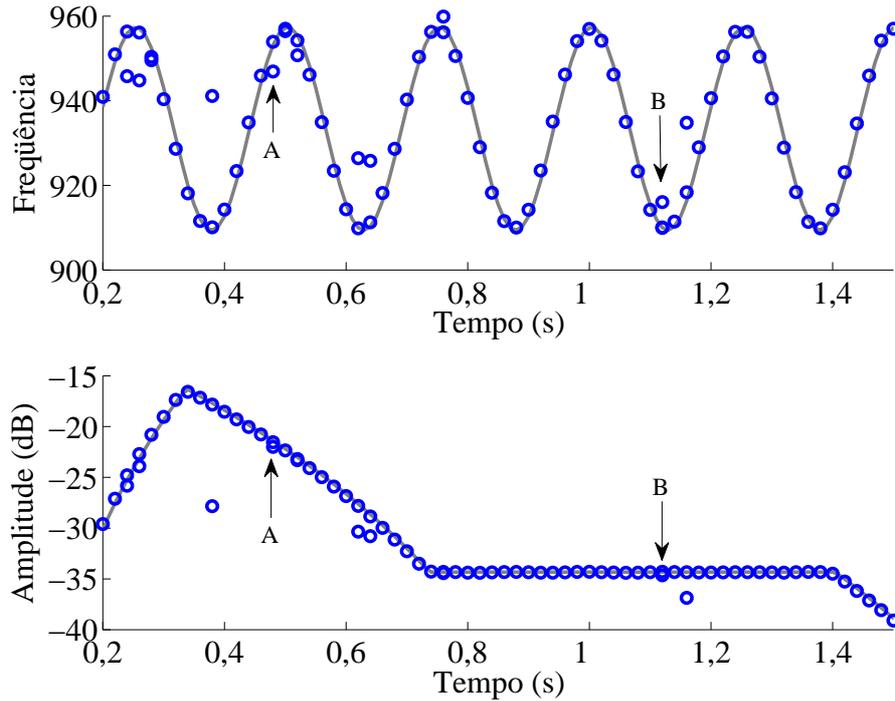


Figura 3.2: Freqüência (em cima) e Amplitude (embaixo) dos picos ('o') detectados para o mesmo sinal da Figura 3.1 com freqüência dentro do intervalo $[900; 960]$ Hz. As trajetórias da freqüência e amplitude utilizadas para sintetizar o sinal podem ser vistas na figura através da linha cinza.

O algoritmo mais tradicional de rastreamento de trilhas é o algoritmo de McAulay & Quatieri (algoritmo MQ), proposto para sinais de fala em [28] e generalizado para sinais não-harmônicos em [16]. Uma descrição detalhada do algoritmo descrito em [16] é feita na Seção 3.2 devido a sua importância e por servir como base para os outros algoritmos apresentados nesta dissertação. Resumidamente, o algoritmo MQ utiliza a proximidade da freqüência em quadros adjacentes para encontrar a continuação de uma trilha.

Em [87] o algoritmo MQ é estendido de modo a utilizar mais parâmetros durante a busca pela continuação de uma dada trilha. Além da freqüência, são utilizados a amplitude e um parâmetro chamado *sweep rate* que mede quanto a freqüência de uma componente senoidal varia dentro de um quadro de análise.

Uma abordagem diferente para o problema utiliza modelos ocultos de Markov [88] (HMM) para obter a probabilidade de um pico ser a continuação de uma trilha num determinado quadro de análise. Uma vez calculada esta probabilidade para um

determinado número de quadros, o algoritmo de Viterbi pode ser utilizado para selecionar a trajetória mais provável para uma trilha. O rastreamento de parciais via HMM foi recentemente melhorado [8] através de novas funções-custo que deixam o algoritmo mais robusto a sinais não-estacionários e ruidosos.

Devido à alta complexidade computacional do método proposto em [88], uma tentativa utilizando modelos mais simples foi proposta em [89]. Neste caso, um modelo gaussiano é utilizado para se atribuir uma probabilidade de transição de uma trilha para um pico. Esta probabilidade de transição é utilizada para se determinar a melhor continuação para a trilha numa determinada região de frequência.

Um solução por filtragem de Kalman foi proposta em [90]. Nesta solução, o conhecimento de um modelo que controla a evolução das parciais para um dado instrumento (ou família de instrumentos) é usada para prever a amplitude, frequência e fase de uma trilha num quadro. A principal desvantagem deste método é a necessidade da obtenção de um modelo para cada tipo de instrumento musical, o que limita sua aplicação. Diversas melhorias para este algoritmo são propostas em [91].

Uma solução possível para rastreamento de trilhas consiste na “clustering” [92, 39] dos picos espectrais. Assim, os picos presentes num determinado quadro são conjuntamente classificados (de acordo com a fonte sonora) e agrupados em trilhas senoidais. Isto traz vantagens, uma vez que a informação sobre a fonte auxilia na criação das trilhas e vice-versa. O agrupamento é realizado utilizando informação de picos espectrais em um número determinado de quadros através de uma métrica que leva em consideração a distância na frequência e na amplitude dos picos e a relação harmônica entre as frequências dos picos.

Outra solução [38] encontrada na literatura propõe a utilização de estimativas das derivadas da amplitude e da frequência para a escolha do melhor candidato para a continuação da trilha. Em [13] é proposta uma solução que utiliza a localização espacial das fontes em gravações binaurais para auxiliar o rastreamento das trilhas.

Por fim, uma solução que se mostrou particularmente eficaz para o problema de rastreamento de parciais foi a utilização de predição linear [93, 94] para os valores de amplitude e de frequência das trilhas. Nestes trabalhos, a predição dos parâmetros das trilhas (amplitude e frequência) é realizada utilizando um modelo autorregressivo obtido através do algoritmo de Burg [61]. Para uma dada trilha, o

pico de continuação é escolhido utilizando-se um critério baseado na distância [39] do valor predito para o valor da frequência e da amplitude dos picos candidatos. As vantagens deste método são a dispensa de um modelo de fonte sonora a ser treinado, a estabilidade do algoritmo de predição e a utilização de informação de mais de um quadro para a escolha da melhor continuação.

3.2 O Algoritmo MQ

Nesta seção, o algoritmo MQ é descrito com o objetivo de familiarizar o leitor com o problema de rastreamento de trilhas. Parte da notação adotada para formular o problema de rastreamento de trilhas também será apresentada nesta seção.

Denotam-se as frequências dos picos encontrados num determinado quadro m por $\mathbf{f}[m]$, sendo o p -ésimo elemento deste vetor denotado por $f_p[m]$. Já o vetor contendo as frequências atribuídas à i -ésima trilha é representado por $\bar{\mathbf{f}}_i[m]$, sendo que o elemento atribuído a este vetor no quadro m é denotado por $\bar{f}_i[m]$. Deve-se enfatizar que a trilha não precisa necessariamente existir para todos os quadros do sinal, sendo definida apenas a partir de um quadro inicial $m_{0,i}$.

O algoritmo MQ considera que as trilhas podem estar em três estados: ‘surgindo’, ‘ativa’ e ‘desaparecendo’. Todo pico que não é associado a nenhuma trilha dá origem a uma trilha no estado ‘surgindo’. Caso uma trilha no estado ‘surgindo’ seja associada a um número pré-determinado de picos ela passa para o estado ‘ativa’. Caso uma trilha no estado ‘ativa’ não encontre uma continuação, ela passa para o estado ‘desaparecendo’. Se uma trilha no estado ‘desaparecendo’ não encontra uma continuação em um alguns quadros, ela é considerada extinta e o algoritmo pára de procurar por uma continuação para esta trilha. Os passos abaixo resumem o algoritmo MQ operando num quadro m :

1. Para cada trilha i , o pico p cuja frequência está mais próxima da frequência da trilha, respeitando $|f_p[m] - \bar{f}_i[m - 1]| \leq \Delta f$, é escolhido como pertencente à rota de sua continuação. Quando duas trilhas disputam um mesmo pico, aquela mais próxima dele em frequência ganha a disputa. A outra trilha procura outro pico mais próximo. Em caso de empate, a trilha com maior

amplitude ganha a disputa.

2. Uma trilha no estado ‘surgindo’ é criada para acomodar qualquer pico que não tenha sido associado a uma trilha. Se uma trilha fica no estado ‘surgindo’ por mais de E quadros, seu estado muda para ‘ativa’. Se uma trilha ‘surgindo’ não encontra uma continuação após E quadros, ela é descartada.
3. Se num quadro m uma trilha não encontra nenhum pico, ela entra no estado ‘desaparecendo’. Seus parâmetros no quadro $m - 1$ são replicados para o quadro m . Se a trilha encontra uma continuação natural em até S quadros, ela sai do estado ‘desaparecendo’; caso contrário, ela é considerada extinta. Quando uma trilha é considerada extinta, o algoritmo pára de buscar-lhe continuações nos próximos quadros.

O desempenho do algoritmo é altamente dependente da escolha dos parâmetros $\{\Delta f, E, S\}$. O papel de cada parâmetro é descrito abaixo:

- O parâmetro Δf controla a máxima variação da frequência de uma trilha entre dois quadros adjacentes. Normalmente este parâmetro é escolhido em função da própria frequência da trilha; por exemplo, uma escolha usual é $\Delta f = 0,03\bar{f}_i[m - 1]$, que corresponde a uma variação de aproximadamente um quarto de tom em torno de $\bar{f}_i[m - 1]$.
- O parâmetro S evita a segmentação de uma trilha supostamente longa em diversas outras de menor duração, devido a descontinuidades indesejáveis na trajetória frequencial da trilha, decorrentes de falhas do estágio de detecção de picos espectrais. Deve-se levar em consideração que um valor alto para S (da ordem de dezenas de quadros) pode levar à ligação errônea tanto entre picos genuínos, como com aqueles provocados por ruído. Assim, aumenta-se a possibilidade de criação de trilhas inexistentes no sinal original. Um valor muito baixo para este parâmetro, por outro lado, pode provocar trilhas excessivamente curtas, que não capturam corretamente as trajetórias espectrais presentes no sinal.
- O parâmetro E é responsável por remover trilhas curtas, possivelmente formadas por picos erroneamente detectados (falsos alarmes). A escolha deste

parâmetro pode levar em consideração critérios psicoacústicos, já que um sinal senoidal com duração inferior a 100 ms pode ser compreendido como um impulso sonoro, em vez de um sinal ressonante.

O procedimento descrito acima pode falhar em identificar a continuação de uma trilha em sinais que apresentem *vibrato* ou em sinais polifônicos. A ocorrência de *vibrato* pode levar a grandes variações de frequência entre quadros adjacentes. Em situações como esta, aumentar o valor de Δ_f facilita o rastreamento das trilhas. Já para sinais polifônicos, a ocorrência de componentes espectrais pouco espaçadas na frequência (até mesmo com cruzamento das trajetórias frequenciais de suas trilhas) pode requerer a adoção de valores pequenos para a magnitude de Δf . Além disso, por apenas considerar a evolução da frequência das trilhas, o algoritmo MQ ignora a continuidade da amplitude, o que pode levar a distorções no sinal ressintetizado.

3.3 Rastreamento de Trilhas Utilizando Predição Linear

A solução proposta neste trabalho, que segue a estratégia desenvolvida originalmente em [10], procura utilizar a informação já acumulada na trilha para auxiliar o processo de continuação de sua trajetória. Uma visão geral dessa estratégia é apresentada nesta seção.

Os algoritmos propostos nas próximas seções podem ser divididos em duas etapas: predição e decisão. Na etapa de predição, os valores passados de amplitude e de frequência são utilizados para realizar uma estimativa dos valores futuros dessas mesmas quantidades. As soluções estudadas para esta tarefa são descritas na Seção 3.4. Na etapa de decisão, os picos são associados às trilhas de acordo com heurísticas cuidadosamente escolhidos. Estas heurísticas são detalhadas na Seção 3.5.

A Figura 3.3 ilustra o papel dos estágios de predição e de decisão no esquema de rastreamento de trilhas proposto. Para uma determinada trilha i , utilizando-se a trajetória da frequência ($\bar{f}_i[m-1]$) e da amplitude ($\bar{A}_i[m-1]$) obtidas até o quadro $m-1$, é feita uma predição para o valor que a trilha assumiria no quadro m para ambos parâmetros ($\hat{f}_i[m]$) e $\hat{A}_i[m]$). O decisor, então, compara os valores preditos para a frequência e amplitude com os respectivos valores dos picos detectados no

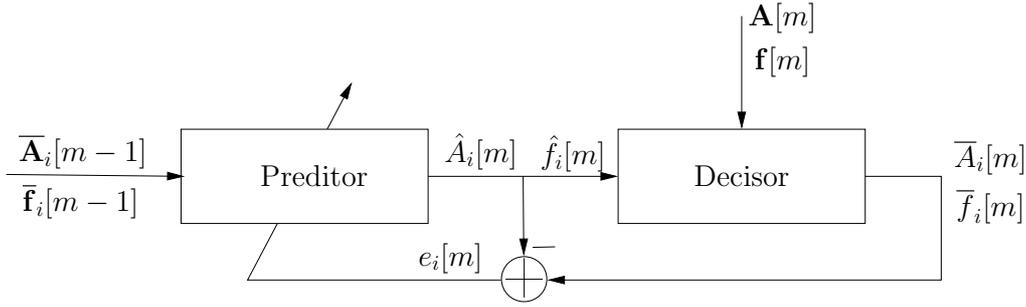


Figura 3.3: Visão geral do algoritmo proposto para o rastreamento de trilhas.

quadro m , armazenados nos vetores $\mathbf{A}[m]$ e $\mathbf{f}[m]$, respectivamente. Uma vez tomada a decisão, as trajetórias da trilha são atualizadas com os valores escolhidos: $\bar{\mathbf{A}}_i[m]$ e $\bar{f}_i[m]$. Esses valores, então são utilizados para atualizar o preditor através do sinal auxiliar $e_i[m]$.

O próprio algoritmo MQ, resumido na seção anterior, pode ser descrito em uma etapa de predição e uma etapa de decisão. Neste caso, a etapa de predição consistiria em replicar no quadro $m + 1$ o valor de frequência atribuído à trilha no quadro m . A etapa de decisão, por sua vez, consistiria nos passos descritos na Seção 3.2.

O uso de preditores mais elaborados do que o empregado no esquema de MQ, i.e. que consideram um histórico mais duradouro dos parâmetros de uma trilha para melhor estimar a tendência de evolução de sua trajetória, favorecem o esquema de rastreamento. No caso de sinais com modulações de frequência (*vibrato*) e de amplitude (*tremolo*), espera-se que os preditores capturem estas modulações. Assim, a região de busca pela continuação de uma trilha pode ser reduzida, possivelmente aumentando a robustez do algoritmo à polifonia. Além disso, o algoritmo proposto utiliza informações de frequência e de amplitude ao realizar esta busca, procurando manter a suavidade na evolução temporal dos parâmetros em questão.

3.4 Preditores

Nesta seção, são descritos os preditores utilizados. Para isto, uma breve introdução sobre predição linear e filtragem adaptativa é realizada. Em seguida, os diferentes preditores utilizados são discutidos em detalhe.

Um preditor linear procura obter uma estimativa para uma amostra futura de

um sinal através de uma combinação linear de amostras passadas do sinal. Assim, para um sinal $x[n]$, um preditor linear de ordem M estima a amostra do sinal no tempo $n + s$, $s \geq 1$, utilizando as M amostras mais recentes, através de

$$\hat{x}[n + s] = \sum_{j=0}^{M-1} w_j x[n - j], \quad (3.1)$$

onde w_j são os coeficientes de predição. A diferença entre os diferentes preditores a serem apresentados está no algoritmo de otimização utilizado para encontrar estes coeficientes. Em geral, os algoritmos podem ser classificados em duas grandes famílias: os que minimizam o erro médio quadrático, como por exemplo o algoritmo de Burg [62], e os que minimizam o erro quadrático [95], como os algoritmos apresentados nesta seção. Além dessa classificação, pode-se agrupar os algoritmos de duas outras formas: de solução por blocos e de solução sequencial [63]. Na solução por blocos, os coeficientes ótimos são encontrados para um conjunto de amostras do sinal; já nas soluções sequenciais, os coeficientes são atualizados com a chegada de cada nova amostra.

Uma possível maneira de se encontrar os coeficientes de predição é através de um filtro adaptativo [96]. Um filtro adaptativo é geralmente implementado através de um filtro digital cujos coeficientes são ajustados ao longo do tempo de acordo com algum critério de otimização. Esta técnica é especialmente vantajosa para situações em que a filtragem deve se adaptar a algum ambiente variante no tempo. O diagrama de um filtro adaptativo típico pode ser visto na Figura 3.4.

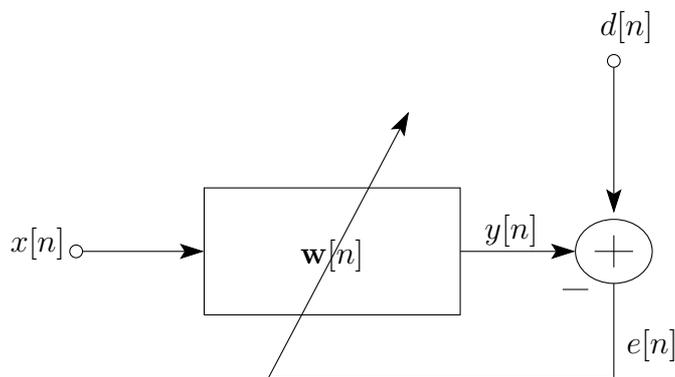


Figura 3.4: Estrutura típica de filtragem adaptativa.

O sinal de erro, $e[n]$, é a diferença entre o sinal de saída do filtro, $y[n] = \sum_{m=0}^M w_m[n]x[n - m]$ (para um filtro FIR de ordem M), e um dado sinal desejado

$d[n]$. Conforme o sinal de entrada $x[n]$ evolui, os coeficientes do filtro, $\mathbf{w}[n]$, são estimados seqüencialmente de modo a minimizar alguma função-custo que tipicamente envolve a potência do sinal de erro. Em particular, ao se escolher $d[n] = x[n + 1]$ o sistema se torna num preditor linear que estima a próxima amostra do sinal através de $\hat{x}[n + 1] = y[n]$. Deve-se notar que neste caso, o vetor contendo os coeficientes do filtro (que fazem o papel dos coeficientes de predição) é estimado para cada nova entrada do sinal; por isso, a necessidade de se escrever os coeficientes como um vetor, sendo o instante de tempo em que foram calculados denotado entre colchetes. Um elemento deste vetor é denotado através do sub-índice m . A seguir, são apresentados os filtros adaptativos utilizados neste trabalho.

3.4.1 RLS

Nesta seção, é descrita uma solução utilizando um filtro adaptativo RLS (do inglês, *Recursive Least Squares*) com regularização [P3]. Duas formas de predição são realizadas com este filtro, uma predição conjunta da freqüência e da amplitude das trilhas e uma predição independente destas duas quantidades. Primeiramente, é descrito o esquema de predição conjunta. O sub-índice i denota que a predição está sendo realizada para a i -ésima trilha.

Definindo o vetor de saída $\mathbf{y}_i[m] = \begin{bmatrix} \hat{A}_i[m] & \hat{f}_i[m] \end{bmatrix}$ e o vetor de entrada $\mathbf{x}_i[m] = \begin{bmatrix} \bar{\mathbf{A}}_i^T[m-1] & \bar{\mathbf{f}}_i^T[m-1] \end{bmatrix}$, com

$$\bar{\mathbf{A}}_i[m-1] = \begin{bmatrix} \bar{A}_i[m-1] & \bar{A}_i[m-2] & \cdots & \bar{A}_i[m-J] \end{bmatrix}^T \quad (3.2)$$

$$\bar{\mathbf{f}}_i[m-1] = \begin{bmatrix} \bar{f}_i[m-1] & \bar{f}_i[m-2] & \cdots & \bar{f}_i[m-J] \end{bmatrix}^T, \quad (3.3)$$

pode-se escrever

$$\mathbf{y}_i[m] = \mathbf{x}_i[m] \mathbf{W}_i[m-1], \quad (3.4)$$

onde $\mathbf{W}_i[\cdot]$ é a matriz $2J \times 2$ de coeficientes.

Dada uma escolha cuidadosa de um fator de regularização $\alpha > 0$ e um fator de esquecimento $0 \ll \lambda \leq 1$, o problema de mínimos quadrados ponderados com regularização [95] procura a matriz de coeficientes $\mathbf{W}_i[m]$ que minimiza a função-custo

$$\lambda^{k+1} \mathbf{W}_i^T[m] \mathbf{\Pi}_J^{-1} \mathbf{W}_i[m] + \sum_{l=0}^k \lambda^{k-l} \|\mathbf{d}_i[l] - \mathbf{x}_i[l] \mathbf{W}_i[m]\|^2, \quad (3.5)$$

onde $\mathbf{d}_i[k] = [\bar{A}_i[k+1] \ \bar{f}_i[k+1]]$ é um vetor contendo o sinal desejado, e $\mathbf{\Pi}_J^{-1} = \alpha^{-1}\mathbf{I}_J$, com \mathbf{I}_J sendo uma matriz identidade de dimensão J .

A solução no quadro m pode ser calculada através dos seguintes passos:

$$\gamma_i[m] = (1 + \lambda^{-1}\mathbf{x}_i[m]\mathbf{P}_i[m-1]\mathbf{x}_i^T[m])^{-1} \quad (3.6)$$

$$\mathbf{g}_i[m] = \lambda^{-1}\mathbf{P}_i[m-1]\mathbf{x}_i^T[m]\gamma_i[m] \quad (3.7)$$

$$\mathbf{e}'_i[m] = \mathbf{d}_i[m] - \mathbf{x}_i[m]\mathbf{W}_i[m-1] \quad (3.8)$$

$$\mathbf{W}_i[m] = \mathbf{W}_i[m-1] + \mathbf{g}_i[m]\mathbf{e}'_i[m] \quad (3.9)$$

$$\mathbf{P}_i[m] = \lambda^{-1}\mathbf{P}_i[m-1] - \mathbf{g}_i[m]\mathbf{g}_i^T[m]\gamma_i^{-1}[m]. \quad (3.10)$$

A matriz de coeficientes pode ser definida explicitamente como

$$\mathbf{W}_i[m] = \begin{bmatrix} \mathbf{w}_{i,AA} & \mathbf{w}_{i,fA} \\ \mathbf{w}_{i,Af} & \mathbf{w}_{i,ff} \end{bmatrix}, \quad (3.11)$$

onde cada vetor $\mathbf{w}_{i,bc}$ representa o efeito de c sobre a predição de b , podendo ser cada um tanto a amplitude quanto a frequência. Desta maneira, este esquema utiliza a informação da frequência para melhorar a predição da amplitude, e vice-versa. Isto, em última análise, pode melhorar o resultado da predição, uma vez que estes parâmetros podem ser controlados simultaneamente pelo músico [97].

Dependendo do tipo de fonte sonora ou do nível de contaminação do sinal por ruído, o sinal de amplitude pode se comportar mais erráticamente do que o parâmetro de frequência correspondente, atrapalhando sua estimação. Para estes casos, uma possível alternativa é desacoplar os estimadores. Isto pode ser feito através de uma simplificação na estrutura da matriz de coeficientes, que pode ser reescrita como

$$\mathbf{W}_i[m] = \begin{bmatrix} \mathbf{w}_{i,AA} & \mathbf{0} \\ \mathbf{0} & \mathbf{w}_{i,ff} \end{bmatrix}. \quad (3.12)$$

Assim, nenhuma informação cruzada da evolução da amplitude e da frequência é compartilhada. Os coeficientes podem, então, ser estimados através de dois filtros adaptativos independentes, um para a frequência e outro para a amplitude. Conseqüentemente, a ordem de predição de cada um destes filtros e os parâmetros do filtro adaptativo podem ser escolhidos de modo a melhor se adequarem à característica da evolução temporal da amplitude e da frequência. Nesta configuração, a ordem do preditor de frequência seria denotada por J^f e a do preditor de amplitude

por J^A . Os parâmetros do filtro também receberiam estes índices, de acordo com o preditor em questão.

3.4.2 RLS *Lattice*

O filtro adaptativo descrito na seção anterior pode ser considerado como um algoritmo com ordem fixa, no sentido de que apenas atualizações no tempo são realizadas nos coeficientes do filtro. Desta maneira, apenas quantidades relacionadas à solução de uma predição de ordem fixa são propagadas pelo filtro. A solução *lattice* [98, 99], no entanto, utiliza atualizações de tempo e de ordem (até uma ordem pré-determinada) para obter os valores preditos através da solução de problemas de predição linear de ordem crescente. A solução assim obtida exibe diversas vantagens sobre a solução apresentada anteriormente, incluindo melhores propriedades numéricas, estabilidade e, principalmente, complexidade computacional reduzida [95].

A notação utilizada até aqui será modificada devido à natureza recursiva em ordem do filtro *lattice*. Um sub-índice j , denotando uma solução de ordem j para uma dada quantidade, será utilizado nesta seção.

Um filtro *lattice a priori*¹ foi empregado para realizar a predição independente da amplitude e da frequência. Este esquema seria equivalente ao preditor desacoplado da seção anterior. Para um dado $\alpha > 0$ e um fator de esquecimento $0 \ll \lambda \leq 1$, o sistema adaptativo usando filtro *lattice* obtém os vetores de pesos $\mathbf{w}_{i,J}[m]$ que minimizam a seguinte função-custo de mínimos quadrados de ordem J [95]:

$$\lambda^{m+1} \mathbf{w}_{i,J}[m]^T \mathbf{\Pi}_J^{-1} \mathbf{w}_{i,J}[m] + \sum_{l=0}^m \lambda^{m-l} \|d_i[l] - \mathbf{x}_i^T[l] \mathbf{w}_{i,J}[m]\|^2, \quad (3.13)$$

onde $\mathbf{x}_i[m]$ pode ser tanto $\bar{\mathbf{A}}_i[m-1]$, para o preditor da amplitude, quanto $\bar{\mathbf{f}}_i[m-1]$, para o preditor de frequência. O sinal desejado no tempo m é denotado por $d_i[m]$, e pode ser tanto $\bar{A}_i[m]$ quanto $\bar{f}_i[m]$, dependendo da quantidade sendo predita. Já

$$\mathbf{\Pi}_J^{-1} = \alpha^{-1} \begin{bmatrix} \lambda^{-2} & & & & \\ & \lambda^{-3} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \lambda^{-(J+1)} \end{bmatrix} \quad (3.14)$$

¹Que minimiza o erro *a priori*.

é a matriz de regularização utilizada.

A solução de ordem j deste problema pode ser obtida para a i -ésima trilha através das seguintes equações [95]:

$$\zeta_{i,j}^f[m] = \lambda \zeta_{i,j}^f[m-1] + \alpha_{i,j}^2[m] \gamma_{i,j}[m-1] \quad (3.15)$$

$$\zeta_{i,j}^b[m] = \lambda \zeta_{i,j}^f[m-1] + \beta_{i,j}^2[m] \gamma_{i,j}[m] \quad (3.16)$$

$$\delta_{i,j}[m] = \lambda \delta_{i,j}[m-1] + \alpha_{i,j}[m] \beta_{i,j}[m-1] \gamma_{i,j}[m-1] \quad (3.17)$$

$$\rho_{i,j}[m] = \lambda \rho_{i,j}[m-1] + e_{i,j} \beta_{i,j}[m] \gamma_{i,j}[m-1] \quad (3.18)$$

$$\beta_{i,j+1}[m] = \beta_{i,j}[m-1] - \kappa_{i,j}^b[m-1] \alpha_{i,j}[m] \quad (3.19)$$

$$\alpha_{i,j+1}[m] = \alpha_{i,j}[m] - \kappa_{i,j}^f[m-1] \beta_{i,j}[m-1] \quad (3.20)$$

$$e_{i,j+1}[m] = e_{i,j}[m] - \kappa_{i,j}[m-1] \beta_{i,j}[m] \quad (3.21)$$

$$\gamma_{i,j+1}[m] = \gamma_{i,j}[m] - (\gamma_{i,j}[m] \beta_{i,j}[m])^2 / \zeta_{i,j}^b[m] \quad (3.22)$$

$$\kappa_{i,j}^b[m] = \delta_{i,j}[m] / \zeta_{i,j}^f[m] \quad (3.23)$$

$$\kappa_{i,j}^f[m] = \delta_{i,j}[m] / \zeta_{i,j}^b[m-1] \quad (3.24)$$

$$\kappa_{i,j}[m] = \rho_{i,j}[m] / \zeta_{i,j}^b[m], \quad (3.25)$$

com $\gamma_{i,0}[m] = 1$, $\beta_{i,0}[m] = \alpha_{i,0}[m] = x_i[m]$ e $e_{i,0}[m] = d_i[m]$. A solução de ordem J no quadro m pode ser calculada iterando-se as equações acima com j variando de 0 até $J-1$. Quando uma nova trilha é criada, os preditores associados a esta trilha devem ser inicializados. A seguinte inicialização dos parâmetros do filtro *lattice* é necessária:

$$\gamma_{i,j}[-1] = 1 \quad (3.26)$$

$$\delta_{i,j}[-1] = 0 \quad (3.27)$$

$$\rho_{i,j}[-1] = \beta_{i,j}[-1] = \bar{\beta}_{i,j}[-1] = \bar{\alpha}_{i,j}[-1] = 0 \quad (3.28)$$

$$\kappa_{i,j}^f[-1] = \kappa_{i,j}^b[-1] = \kappa_{i,j}[-1] = 0 \quad (3.29)$$

$$\zeta_{i,j}^f[-1] = \alpha^{-1} \lambda^{-2} \quad (3.30)$$

$$\zeta_{i,j}^b[-1] = \alpha^{-1} \lambda^{-j-2}, \quad (3.31)$$

com j variando de 0 até $J-1$. Um exemplo da estrutura *lattice* para $J = 3$ utilizando as equações exibidas acima pode ser visto na Figura 3.5.

Como pode ser observado, o filtro *lattice* não encontra explicitamente o vetor de coeficientes ótimo. O vetor de coeficientes e o valor predito poderiam ser encontrados

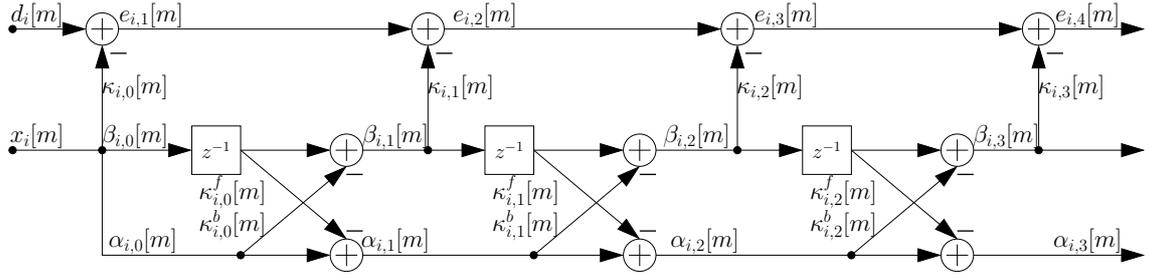


Figura 3.5: Estrutura *lattice a priori* para $J = 3$.

através de uma recursão adicional [95]. No entanto, neste trabalho, é adotada uma solução mais simples para obter o valor predito após uma atualização de tempo do filtro adaptativo. Para isto, a quantidade chave é o erro *a priori*, definido como

$$e_{i,J}[m] = d_i[m] - \mathbf{x}_i^T[m] \mathbf{w}_i[m-1]. \quad (3.32)$$

O valor predito para o quadro m pode ser escrito como

$$y_i[m] = \mathbf{x}_i^T[m] \mathbf{w}_i[m-1], \quad (3.33)$$

onde $y_i[m]$ pode ser tanto $\hat{A}_i[m]$ ou $\hat{f}_i[m]$, o que leva a

$$y_i[m] = -e_{i,J}[m] \Big|_{d_i[m]=0}. \quad (3.34)$$

Com isso, para se obter o valor predito do filtro no quadro m , é calculada uma nova quantidade numericamente equivalente ao erro *a priori* para um sinal desejado nulo. Desta maneira, o valor predito pode ser obtido através das seguintes recursões de ordem:

$$\bar{\alpha}_{i,j+1}[m] = \bar{\alpha}_{i,j}[m] - \kappa_{i,j}^f[m-1] \bar{\beta}_{i,j}[m-1] \quad (3.35)$$

$$\bar{\beta}_{i,j+1}[m] = \bar{\beta}_{i,j}[m-1] - \kappa_{i,j}^b[m-1] \bar{\alpha}_{i,j}[m] \quad (3.36)$$

$$y_{i,j+1}[m] = y_{i,j}[m] + \kappa_{i,j}[m-1] \bar{\beta}_{i,j}[m] \quad (3.37)$$

com $y_{i,0}[m+1] = 0$, sendo $\bar{\beta}_{i,0}[m] = \bar{\alpha}_{i,0}[m]$ inicializados como $\bar{A}[m-1]$ ou $\bar{f}[m-1]$, dependendo da variável que está sendo predita. Deve-se notar que todas as quantidades envolvidas neste cálculo estão disponíveis no tempo $m-1$, após a atualização de tempo correspondente da estrutura *lattice*.

3.4.3 Inicialização dos Preditores para a Frequência

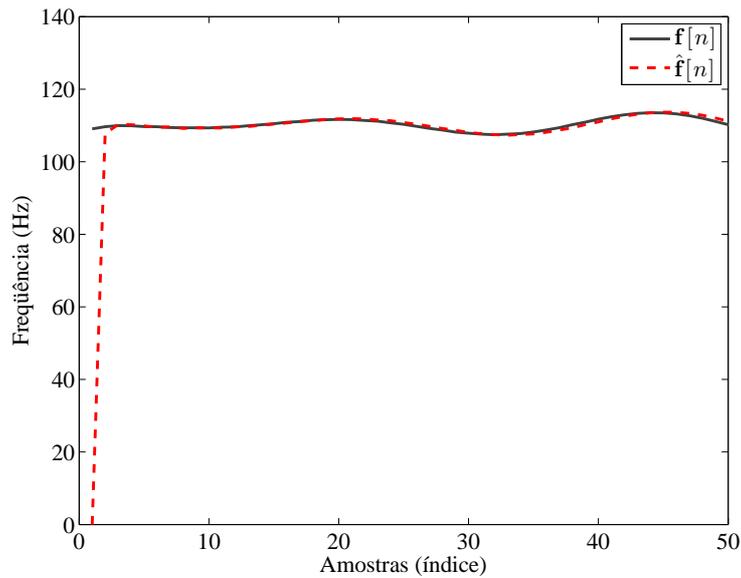
A modulação em frequência de uma trilha normalmente consiste de uma variação em torno de uma frequência média. Desta forma, caso um filtro adaptativo seja inicializado com coeficientes nulos, decorrerá certo tempo até a saída do filtro alcançar o nível médio da modulação em frequência [96]. Para evitar esse atraso, duas soluções são possíveis: inicializar os coeficientes do filtro com o primeiro valor de frequência atribuído à trilha ou subtrair este primeiro valor das amostras na entrada do filtro. A primeira solução pode ser inconveniente devido à utilização da regularização, principalmente no caso do filtro *lattice* [100]. Já a segunda solução se mostrou mais simples, e pode ser facilmente implementada tanto no caso *lattice* quanto para o filtro RLS. Com isso, subtrai-se o valor da frequência do primeiro pico atribuído à trilha de todos os valores de frequência submetidos ao filtro. As amostras preditas, por sua vez, são somadas com este mesmo valor.

Para ilustrar o efeito descrito acima, um filtro adaptativo foi utilizado para prever um sinal sintético simulando um *vibrato* com frequência média de 110 Hz. A Figura 3.6 mostra o sinal de teste e o sinal predito para dois casos: com e sem a inicialização informada. Para o caso sem inicialização, podemos observar o período que o filtro adaptativo leva para sair do estado inicial relaxado até o nível médio do sinal. Já no caso que utiliza a inicialização descrita no parágrafo anterior (subtração do valor inicial), o filtro não precisa sair de uma condição inicial nula, o que evita o atraso inicial.

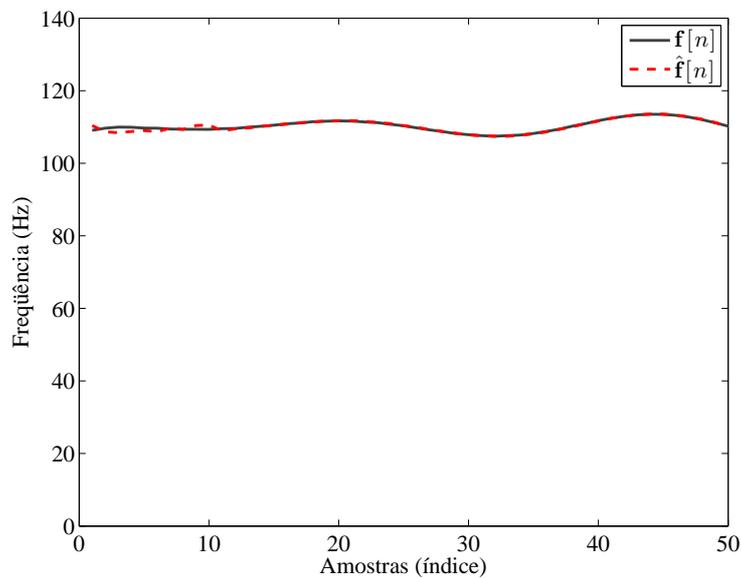
3.4.4 Escolha dos Parâmetros dos Preditores

Nesta seção, é fornecida uma breve explicação dos parâmetros necessários para uma predição adequada da amplitude e da frequência. Os parâmetros dos filtros adaptativos utilizados são a ordem do filtro, J , o fator de esquecimento, λ , e o fator de regularização α .

O valor de J é usualmente escolhido entre 2 e 10 [10]. Um alto valor para J pode reduzir o erro de predição, mas pode implicar um período muito longo de treinamento para o filtro. A escolha de $2 \leq J \leq 6$ tanto para os preditores da amplitude quanto da frequência se mostrou um bom compromisso entre as questões mencionadas.



(a) Sem inicialização



(b) Com inicialização

Figura 3.6: Exemplo do efeito da inicialização empregada no predictor da frequência.

O fator de esquecimento controla a influência das amostras passadas do sinal na predição. Quanto mais próximo de 1 é o fator de esquecimento, maior a influência das amostras passadas. Um valor $\lambda = 0,98$ se mostrou adequado para este problema de predição.

A regularização empregada nos filtros adaptativos permite uma diminuição do tempo de convergência do filtro em troca de uma possível polarização dos valores preditos pelo filtro [96]. O parâmetro α permite o controle entre o aumento do tempo

de convergência e a polarização da solução: quanto mais elevado for α , menor a influência da regularização na solução. Um α entre 500 e 1000 se mostrou adequado para a predição das trilhas.

3.5 Algoritmos de Decisão

Nesta seção, são descritos mecanismos de escolha do pico mais adequado a continuar uma determinada trilha. Estes algoritmos podem ser vistos como uma extensão do algoritmo MQ que permite utilizar informação fornecida por preditores e considerar na decisão um maior número de parâmetros, como, por exemplo, a informação da amplitude. São propostas novas heurísticas de modo a melhor utilizar estas informações adicionais.

De maneira geral, os seguintes passos ilustram o processo de associação do pico p à trilha i , num quadro m :

1. Escolher os picos candidatos à continuação da trilha de maneira que todos obedeçam a um critério de proximidade de frequência.
2. Selecionar o pico p , dentre os candidatos, que forneça o menor valor para uma função de distância entre parâmetros preditos a partir da trilha e do pico. Caso haja conflitos, solucioná-los utilizando um critério pré-determinado.
3. Incorporar os parâmetros do pico selecionado à seqüência de parâmetros da trilha. Caso nenhum pico tenha sido escolhido, incorporar os próprios valores preditos.
4. Predizer os valores que os parâmetros da trilha assumirão no quadro $m + 1$.
 1. Caso nenhuma continuação adequada tenha sido encontrada, utilizar uma heurística pré-determinada para permitir realizar a predição.

Cada um dos passos descritos acima será detalhado a seguir. Nas próximas seções, serão apresentados algoritmos para executar as tarefas mencionadas.

Assim como o MQ, o esquema de decisão proposto procura a melhor continuação para uma determinada trilha primeiramente de acordo com um critério de proximidade de frequência. Apenas os picos para os quais a diferença entre a frequência

estimada e o valor de frequência predito para a trilha está abaixo de um determinado limiar são considerados possíveis candidatos à continuação da trilha. O limiar, de maneira similar ao parâmetro Δf do algoritmo MQ, é escolhido em função da frequência predita. O p -ésimo pico é considerado um candidato à continuação da i -ésima trilha senoidal apenas se

$$|f_p[m] - \hat{f}_i[m]| \leq \Delta f, \quad (3.38)$$

onde Δf é usualmente escolhido como uma fração de $\hat{f}_i[m]$. Desta maneira, o problema é simplificado, pois apenas os picos cujas frequências são próximas da frequência predita para a trilha são considerados na busca. Além disso, como os critérios a serem adotados para a escolha do melhor candidato a continuação de uma trilha consideram a proximidade de frequência, não há nenhum prejuízo ao desempenho do rastreador de trilhas.

Em geral, a adequação de um pico como continuação de uma trilha é medida através de uma função de decisão. Por exemplo, no caso do algoritmo MQ, esta função seria a distância da frequência estimada do pico para a frequência atual da trilha. De modo mais geral, pode-se escrever as possíveis funções que desempenham este papel na seguinte forma: $c_m(\Theta_p, \hat{\Theta}_i)$, onde Θ_p denota um conjunto de parâmetros do p -ésimo pico e $\hat{\Theta}_i$ denota a estimativa destes mesmos parâmetros para a trilha i . O sub-índice m na função denota que a função está sendo calculada para as quantidades referentes ao m -ésimo quadro do sinal. Comumente, estas funções são selecionadas de maneira que quanto menor o seu valor mais adequada seja a escolha do p -ésimo pico ser a continuação da i -ésima trilha. Algumas funções que podem realizar esta tarefa serão especificadas na Seção 3.5.1.

A decisão de qual pico deve ser selecionado como continuação de uma trilha deve levar em consideração a escolha das outras trilhas num mesmo quadro. Isto requer estratégias para resolver a ocorrência de conflitos, isto é, quando duas trilhas possuem como melhor continuação um mesmo pico. No algoritmo MQ, a solução destes conflitos é realizada da seguinte maneira: a trilha com o menor custo c_m em relação àquele pico ganha o conflito. A trilha perdedora do conflito procura o pico com o segundo menor custo. Este procedimento pode levar a novos conflitos, já que a trilha perdedora do conflito pode selecionar um pico que já havia sido previamente alocado para outra trilha. Ao final, esta função de decisão precisa ser recursiva, já

que um conflito pode reiniciar o processo de decisão de outra trilha que pode gerar novamente um conflito, e assim por diante.

Em [10], é proposta uma solução alternativa que evita uma recursão para a resolução destes conflitos. Para isto, os picos são ordenados de acordo com o valor da amplitude do último pico atribuído à trilha. À trilha com a maior amplitude, então, é atribuído o pico com menor custo em relação a ela. Em seguida, a trilha com a segunda maior amplitude repete este procedimento, e assim por diante. A idéia deste método é privilegiar as trilhas com maior amplitude na escolha da melhor continuação, evitando o surgimento de conflitos.

A informação dos preditores é utilizada no momento da decisão, ao se compararem os parâmetros preditos para a trilha com os parâmetros dos picos detectados num quadro. Os parâmetros do pico escolhido como continuação para uma determinada trilha, por sua vez, são incorporados ao histórico de parâmetros da trilha. Neste esquema, dois problemas relativos à predição devem ser mencionados. O primeiro diz respeito à atualização do valor predito quando uma continuação para uma determinada trilha não é encontrada. O outro é quão confiáveis são os valores fornecidos pelo preditor, ou seja, como determinar se a convergência da resposta do filtro adaptativo utilizado na predição foi alcançada. Soluções para estes problemas são apresentadas na Seção 3.5.2.

3.5.1 Funções de Decisão

A escolha de uma função que indique o quão adequado é um pico para continuar uma determinada trilha deve levar em conta a suavidade da evolução temporal da amplitude e da frequência da trilha. Nesta seção, são descritas algumas funções estudadas e outras desenvolvidas, juntamente com suas vantagens. Todas as funções utilizam informação sobre a frequência e/ou a amplitude para encontrar a melhor continuação, logo, os parâmetros utilizados pela função são $\Theta_p = \{A_p[m], f_p[m]\}$ e $\hat{\Theta}_i = \{\hat{A}_i[m], \hat{f}_i[m]\}$.

No caso do algoritmo MQ, o pico escolhido como continuação é aquele para o qual é minimizada a diferença entre a sua frequência e a frequência predita para a trilha no quadro m (desde que não haja nenhum conflito). Desta maneira tem-se

$$c_m^{\text{MQ}}(f_p[m], \hat{f}_i[m]) = |\hat{f}_i[m] - f_p[m]|. \quad (3.39)$$

Em [94], a diferença entre a amplitude predita para a trilha e a amplitude estimada para o pico é utilizada para escolher a melhor continuação para a trilha. Assim, a função de decisão seria

$$c_m^{\text{Amp}}(A_p[m], \hat{A}_i[m]) = |\hat{A}_i[m] - A_p[m]|. \quad (3.40)$$

A estratégia proposta nesta dissertação [P3] considera os parâmetros de amplitude e de frequência para escolher a melhor continuação. A função escolhida é uma combinação convexa das distâncias relativas dos parâmetros, ou seja,

$$c_m^{\text{Comb}}(A_p[m], f_p[m], \hat{A}_i[m], \hat{f}_i[m]) = (1-q) \frac{|\hat{f}_i[m] - f_p[m]|}{\hat{f}_i[m]} + q \frac{|\hat{A}_i[m] - A_p[m]|}{\hat{A}_i[m]}. \quad (3.41)$$

O parâmetro q controla a ênfase que deve ser dada às continuidades da amplitude e da frequência na escolha da continuação. Como os picos submetidos a esta função já passaram por um critério de proximidade de frequência, é usual escolher-se este parâmetro com valor maior que 0,5. Deve-se observar que para $q = 0$ esta função recai na função utilizada no caso MQ, e para $q = 1$ a função proposta em [94] é obtida.

Uma função que difere das anteriores é definida em [89], onde é proposto o produto das diferenças entre as amplitudes e a diferença entre as frequências estimadas para um pico e preditas para a trilha. Desta forma, tem-se

$$c_m^{\text{Prod}}(A_p[m], f_p[m], \hat{A}_i[m], \hat{f}_i[m]) = \sqrt{\left(\hat{f}_i[m] - f_p[m]\right)^2 \left(\hat{A}_i[m] - A_p[m]\right)^2}. \quad (3.42)$$

Esta função é interessante pois uma das diferenças acaba por ponderar a outra. Além disso, esta função será utilizada no método de avaliação de algoritmos de rastreamento de trilhas discutido na Seção 3.6.

3.5.2 Heurísticas Adicionais

Nesta seção, são discutidas algumas heurísticas adicionais necessárias para a utilização dos preditores no rastreamento de trilhas. O primeiro problema ocorre devido ao período de treinamento dos filtros adaptativos, após o qual os valores preditos pelo filtro passam a ser considerados confiáveis. O segundo problema é relativo a como atualizar o preditor quando nenhuma continuação para uma trilha é encontrada num determinado quadro. São apresentadas soluções para ambos os problemas.

Os filtros adaptativos utilizados neste trabalho necessitam de um período de treinamento, também chamado de período de convergência do filtro. Durante este período, os valores preditos pelo filtro são pouco confiáveis, o que pode ser observado através dos valores elevados do erro de predição que ocorrem nesta fase. Este problema é altamente indesejável para o rastreamento de trilhas, já que é exatamente neste período que as trilhas senoidais estão sendo criadas. A escolha da função-custo dos filtros adaptativos utilizados neste trabalho levou em conta este problema, uma vez que eles apresentam um período de treinamento reduzido [95]. Usualmente, o período de treinamento destes filtros é da ordem de grandeza do comprimento do filtro [96].

Uma primeira solução [P3] utiliza este fato sobre o filtro para descartar os valores preditos durante o treinamento. Assim, os primeiros ξJ , com $\xi \in \mathbb{N}^*$, valores preditos do filtro são descartados, onde J é a ordem do filtro. No lugar destes valores, o valor de frequência mais recente atribuído à trilha é utilizado, como no algoritmo MQ. Desta maneira, se uma trilha i foi inicialmente detectada no quadro m_0 , tem-se para a frequência

$$\hat{f}_i[m] = \begin{cases} \bar{f}_i[m-1], & \text{se } \xi_f J_f > m - m_0 \\ \hat{f}_i[m], & \text{se } \xi_f J_f \leq m - m_0 \end{cases} \quad (3.43)$$

e para a amplitude

$$\hat{A}_i[m] = \begin{cases} \bar{A}_i[m-1], & \text{se } \xi_A J_A > m - m_0 \\ \hat{A}_i[m], & \text{se } \xi_A J_A \leq m - m_0 \end{cases}. \quad (3.44)$$

Este critério, no entanto, não assegura que a convergência dos filtros foi atingida, nem tampouco que os valores preditos estão próximos dos valores da trilha após a passagem dos ξJ quadros. Levando isto em conta, outra solução proposta [P4] verifica a diferença entre o valor predito e o último valor atribuído à trilha. Caso esta diferença seja maior que um limiar pré-estabelecido, o valor predito é descartado e o último valor atribuído é utilizado. Neste caso, tem-se para a frequência

$$\hat{f}_i[m] = \begin{cases} \bar{f}_i[m-1], & \text{se } |\bar{f}_i[m-1] - \hat{f}_i[m]| > \nu_f \\ \hat{f}_i[m], & \text{em caso contrário,} \end{cases} \quad (3.45)$$

e para a amplitude

$$\hat{A}_i[m] = \begin{cases} \bar{A}_i[m-1], & \text{se } |\bar{A}_i[m-1] - \hat{A}_i[m]| > \nu_A \\ \hat{A}_i[m], & \text{em caso contrário,} \end{cases} \quad (3.46)$$

onde ν_f e ν_A são limiares previamente escolhidos para a frequência e para a amplitude, respectivamente. Desta maneira, os valores utilizados para a busca do melhor candidato estão garantidamente na vizinhança dos últimos valores atribuídos à trilha. Deve-se observar que a escolha para o parâmetro da amplitude deve ser menos restritiva do que a escolha para o parâmetro da frequência, já que é esperada uma variação menor ao longo do tempo para a modulação de frequência do que para a modulação de amplitude.

A Figura 3.7 ilustra o comportamento dos dois critérios descritos acima. Neste exemplo, um filtro adaptativo RLS de ordem 5 foi utilizado para prever uma trajetória de amplitude de um sinal sintético, o erro de predição em % é exibido. A linha tracejada vertical mostra o índice a partir do qual as amostras preditas pelo filtro seriam utilizadas caso fosse utilizado o critério da equação (3.43) com $\xi_A = 2$. Já a linha tracejada/pontilhada denota o erro abaixo do qual os valores preditos pelo filtro seriam utilizados caso fosse utilizado o critério definido pela equação (3.45) com $\nu_A = 0,1$. Pode-se perceber no exemplo que para o sinal utilizado, ambos os critérios seriam suficientes para descartar os valores preditos que possuem erro elevado. Uma vantagem do segundo critério é o fato dele limitar superiormente o maior erro cometido pelo filtro, como pode ser observado na figura.

Durante o rastreamento de uma trilha, é possível que não se encontre nenhuma continuação adequada para a trilha. Isto pode ocorrer devido a três motivos:

1. a trilha deixou de existir, logo não há nenhuma continuação a ser encontrada;
2. um pico que deveria pertencer à trajetória da trilha foi atribuído, erroneamente, a outra trilha;
3. um pico não foi detectado, devido a, por exemplo, algum erro do algoritmo de detecção de picos.

A falta de um pico deve ser contornada para que se possa buscar possíveis continuações da trilha em quadros futuros (nos casos 2 e 3, acima) ou parar de procurar

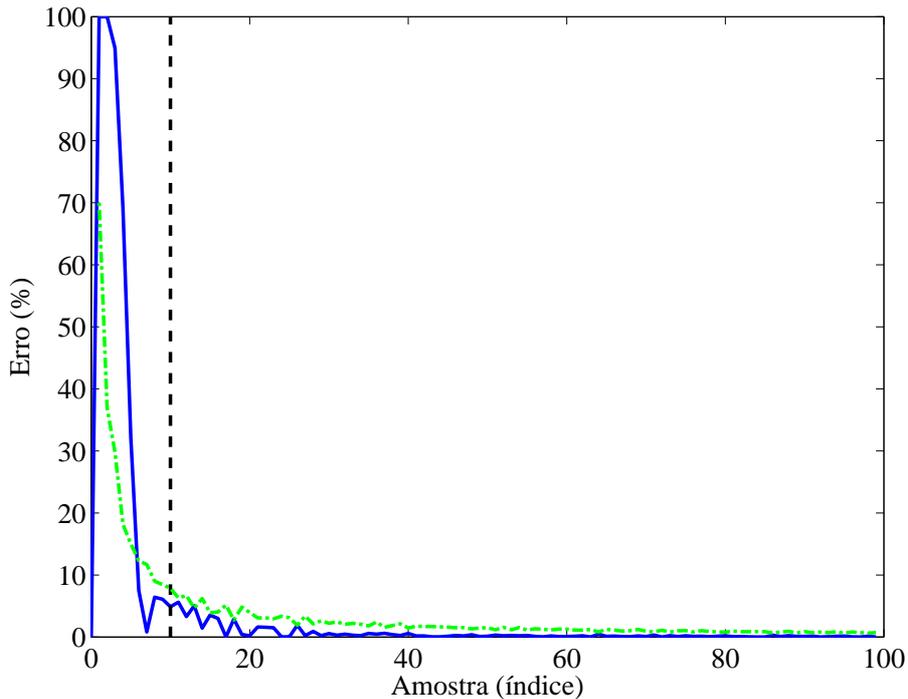


Figura 3.7: Erro de predição (em %) de um filtro adaptativo RLS de ordem 5 para as primeiras 100 amostras do sinal. A linha vertical tracejada (---) ilustra o critério de convergência determinado pela ordem do filtro. Já a linha tracejada/pontilhada (- · -) ilustra o critério determinado pelo maior erro permitido.

uma continuação para a trilha em questão (no caso 1, acima). O parâmetro S do algoritmo MQ (ver Seção 3.2) fornece uma primeira solução para este problema, permitindo que uma trilha continue buscando uma continuação mesmo que ela não tenha sido associada a nenhum pico no quadro anterior. Ainda que seja utilizada uma heurística como esta, é necessário definir como atualizar o preditor no caso de nenhum pico ser associado à trilha num determinado quadro.

Uma possível solução [P4] consiste na utilização de preditores de mais de um passo, ou seja, preditores que buscam encontrar o valor predito para o quadro $m + s$ com informações adquiridas até o quadro m . Assim, caso um pico não seja detectado num quadro m , um preditor com $s = 2$ é treinado de modo a realizar a predição para o quadro $m + 1$ utilizando apenas a informação adquirida até o quadro $m - 1$. Caso a trilha permaneça sem encontrar uma continuação no quadro $m + 1$, o mesmo esquema é utilizado com $s = 3$, e assim por diante até a trilha ser considerada extinta ou encontrar uma continuação.

Formalmente, se uma trilha está no estado ‘desaparecendo’ durante s quadros consecutivos e permanece neste estado no quadro m , o seguinte preditor é utilizado para a fornecer a amostra predita no quadro $m + 1$,

$$y_i[m] = \mathbf{x}_i^T[m - (s - 1)]\tilde{\mathbf{w}}_i[m - s], \quad (3.47)$$

onde os coeficientes de predição são encontrados minimizando-se a função-custo do filtro adaptativo escolhido, com $d_i[m] = x_i[m - (s - 1)]$, onde $x_i[.]$ pode ser a frequência $\bar{f}_i[.]$ ou a amplitude $\bar{A}_i[.]$ da i -ésima trilha. Desta maneira, os valores preditos ótimos são sempre usados na busca pela continuação da trilha, dada a informação disponível.

Assim como no algoritmo MQ, se uma continuação para a trilha não é encontrada após S quadros, ela é considerada extinta e removida do algoritmo de rastreamento. Durante os quadros em que a trilha está no estado ‘desaparecendo’, os valores preditos são atribuídos à trilha. Note que esses valores não são utilizados para treinar o preditor, logo não ocorre a realimentação do erro de predição não-nulo no filtro adaptativo. Os mesmos esquemas de verificação da convergência do filtro descritos anteriormente são aplicáveis aos preditores de mais de um passo.

3.6 Avaliação de algoritmos de rastreamento de trilhas

Nesta seção, é apresentada uma metodologia para a avaliação comparativa de algoritmos de rastreamento de trilhas senoidais. Para isto, procura-se isolar o estágio de rastreamento dos demais estágios de um sistema de análise senoidal durante a avaliação.

Usualmente, os algoritmos de rastreamento de trilhas senoidais são avaliados inspecionando-se visualmente o resultado da análise [18, 93, 8], exibindo-se a variação de amplitude ou de frequência estimada para cada trilha senoidal obtida e comparando-a com o espectrograma do sinal de interesse. Com isso, a avaliação é feita de maneira subjetiva. Para comparar o desempenho de diferentes algoritmos de rastreamento de trilhas, no entanto, é altamente desejável uma medida objetiva de qualidade.

Em [74, 94, 10] um esquema de avaliação objetivo é utilizado na comparação de diferentes sistemas de análise senoidal. Este esquema utiliza a diferença entre o sinal sintetizado a partir das trilhas senoidais obtidas e o sinal sob análise (um sinal sintético) para obter uma figura de mérito. A desvantagem deste esquema é que o sistema de análise/síntese senoidal como um todo acaba sendo avaliado e não apenas a etapa de rastreamento de trilhas.

Para que a avaliação ocorra de maneira isolada, o algoritmo de avaliação a ser apresentado considera como entrada um conjunto de trilhas senoidais de referência que é comparado com outro conjunto de trilhas senoidais obtido por um algoritmo de rastreamento de senóides. A metodologia utilizada para a comparação destes conjuntos, originalmente apresentada em [45], é descrita na Seção 3.6.1. Na Seção 3.6.2, é descrito o algoritmo de avaliação proposto em [45]. Já na Seção 3.6.3, é sugerida uma modificação do avaliador.

3.6.1 Metodologia

A metodologia adotada [45] para avaliação de algoritmos de rastreamento de trilhas é similar à adotada para algoritmos de detecção de senóides, anteriormente apresentada na Seção 2.6. Deve-se ressaltar que a avaliação de algoritmos de rastreamento de trilhas é mais difícil de ser realizada. Primeiramente, as trilhas são compostas por uma coleção de picos senoidais ao longo do tempo. A correspondência entre uma trilha detectada com uma trilha presente no sinal de referência, por exemplo, deve considerar não apenas os parâmetros de cada pico atribuído à trilha mas também o suporte temporal compreendido entre os quadros em que a trilha começou e deixou de existir.

A Figura 3.8 apresenta a metodologia de avaliação utilizada neste trabalho. A metodologia pode ser resumida nos passos descritos abaixo.

1. Definição de um cenário de teste.
2. Geração de um conjunto de picos senoidais \mathcal{P} e de trilhas de referência $\bar{\mathcal{L}}$ a partir do cenário de teste.
3. Obtenção do conjunto de trilhas senoidais \mathcal{L} .
4. Comparação do conjunto de trilhas obtidos com o conjunto de referência.

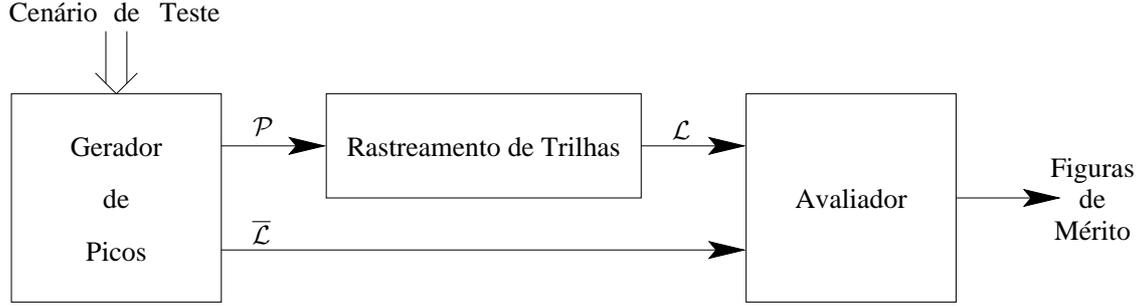


Figura 3.8: Metodologia utilizada para a avaliação de algoritmos de rastreamento de senóides.

5. Cálculo das figuras de mérito.

Os conjuntos de referência podem ser gerados sinteticamente ou obtidos de sinais naturais previamente analisados. No Capítulo 4, é apresentado um banco de trilhas senoidais cujo objetivo é disponibilizar sinais de teste variados para sistematizar a avaliação de desempenho de sistemas de análise senoidal. A escolha do cenário de teste determina tipos e níveis de dificuldades aos quais os rastreadores de trilhas serão submetidos. Poderiam ser criados para os cenários objetivos como:

- avaliação da robustez do algoritmo a picos provocados pela parcela não ressonante do sinal;
- avaliação da sensibilidade do algoritmo a *vibrato* e a *tremolo*;
- avaliação da capacidade do algoritmo de resolver o problema de parciais com cruzamento de trajetórias de frequência.

No Capítulo 5, é realizado um estudo de caso utilizando o algoritmo de avaliação apresentado.

Seguindo em ordem os passos da metodologia, é necessário definir o conjunto de trilhas senoidais utilizado na avaliação. Para começar, a i -ésima trilha senoidal pertencente a um conjunto de trilhas pode ser definida através de

$$l_i[m_i] = \{\mathbf{f}_i[m_i], \mathbf{A}_i[m_i], \Phi_i[m_i]\}, \quad (3.48)$$

com

$$m_i = \{b_i, b_i + 1, \dots, e_i - 1, e_i\}, \quad (3.49)$$

onde $\mathbf{f}_i[m]$, $\mathbf{a}_i[m]$ e $\Phi_i[m]$ são vetores que contêm os valores de frequência, amplitude e fase atribuídos a trilha i ; e os valores b_i e e_i definem os quadros onde ocorrem o início e o fim da trilha, respectivamente. Assim como no avaliador de algoritmos de detecção de senóides, é esperado que os valores de amplitude e de frequência das trilhas estejam normalizados no intervalo $[0,1]$.

O conjunto \mathcal{L} pode, então, ser definido como a união de todas as trilhas detectadas no sinal, ou seja,

$$\mathcal{L} = \bigcup_i \{l_i[m_i]\}. \quad (3.50)$$

O conjunto de referência é definido de forma análoga, porém será denotado como $\bar{\mathcal{L}}$.

3.6.2 Algoritmo de Avaliação

Nesta seção, é apresentado o avaliador originalmente proposto em [45]. Inicialmente será definido um erro cometido ao se associar uma trilha de um conjunto de trilhas com uma trilha do conjunto de referência. Em seguida, é descrito um esquema para associar iterativamente as trilhas do conjunto de referência às trilhas do conjunto sob teste.

Seguindo os passos utilizados em [45], o erro cometido ao se associar a trilha $l_i \in \mathcal{L}$ à trilha $l_j \in \bar{\mathcal{L}}$ é definido como

$$d(l_i, l_j) = \sum_{m_1=n_2}^{n_3} |F_i[m_1] - F_j[m_1]| \cdot |A_i[m_1] - A_j[m_1]| + \sum_{m_2=n_1}^{n_2} \max(A_i[m_2], A_j[m_2]) + \sum_{m_3=n_3}^{n_4} \max(A_i[m_3], A_j[m_3]), \quad (3.51)$$

onde

$$n_1 = \min(b_i, b_j) \quad (3.52)$$

$$n_2 = \min(\max(b_i, b_j), \min(e_i, e_j)) \quad (3.53)$$

$$n_3 = \max(\max(b_i, b_j), \min(e_i, e_j)) \quad (3.54)$$

$$n_4 = \max(e_i, e_j). \quad (3.55)$$

A Figura 3.9 exibe três casos de sobreposição das trilhas l_i e l_j ao longo do tempo e as escolhas dos índices n_1 , n_2 , n_3 e n_4 para cada um dos casos. O erro $d(l_i, l_j)$ pode ser dividido em três parcelas:

1. entre n_1 e n_2 , quando uma das trilhas já começou porém a outra trilha não, o erro é considerado como a amplitude da trilha que já começou (a amplitude da trilha que ainda não começou é considerada nula);
2. entre n_2 e n_3 , quando ambas as trilhas existem, o erro é definido como o produto entre a diferença ponto-a-ponto da amplitude e da frequência, sendo elevada quando ambos os parâmetros das trilhas diferirem por quantidades elevadas;
3. entre n_2 e n_3 , quando uma das trilhas já se extinguiu porém a outra não, o erro é considerado como a amplitude da trilha que ainda não se extinguiu.

Deve-se ressaltar que, para o Caso 2 ilustrado na Figura 3.9, apenas os somatórios com limites de n_1 até n_2 e n_3 até n_4 são utilizados na obtenção do custo $d(l_i, l_j)$.

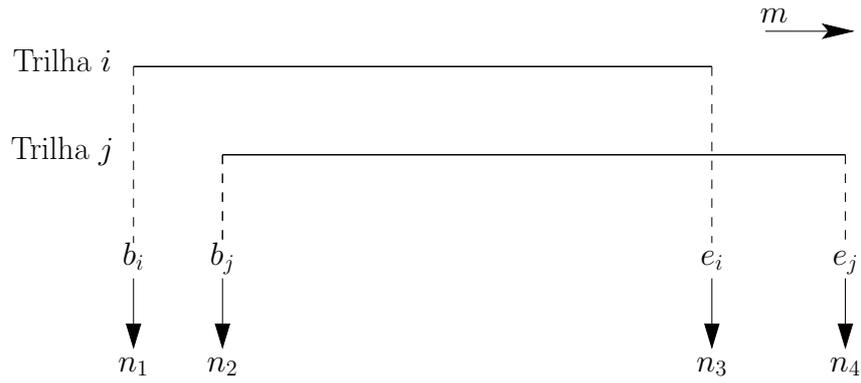
Seguindo o desenvolvimento realizado para o caso de detecção de senóides, é necessário formular um método que permita comparar dois conjuntos de trilhas senoidais e chegar a um custo “total” de seu casamento. Considerando o conjunto de referência $\overline{\mathcal{L}}$ e o conjunto sob avaliação \mathcal{L} , o Algoritmo 3.1 procura associar os elementos destes dois conjuntos de modo a minimizar o custo desta associação. A ordem da busca é realizada de acordo com a amplitude total das trilhas do conjunto de referência, definida para uma trilha $l_j \in \overline{\mathcal{L}}$ como [45]

$$A_j^T = \sum_{m=b_j}^{e_j} A_j[m]. \quad (3.56)$$

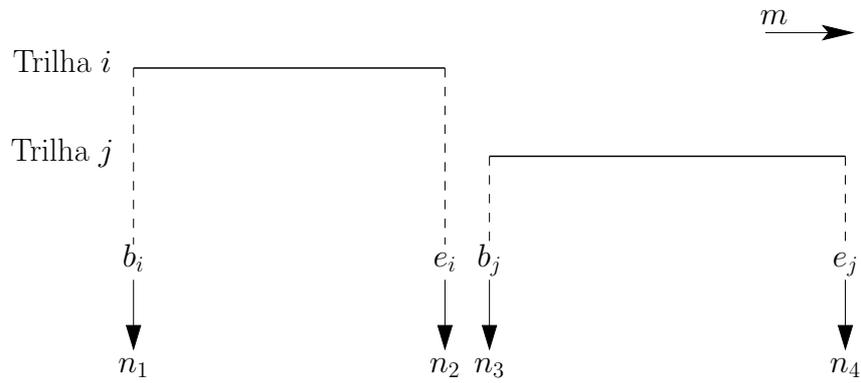
No Algoritmo 3.1, são utilizados dois conjuntos auxiliares \mathcal{L}_r e $\overline{\mathcal{L}}_r$ que armazenam as trilhas que ainda não foram associadas. Os passos seguidos pelo Algoritmo 3.1 dentro do *loop* ‘enquanto’ podem ser resumidos da seguinte maneira:

1. Escolher a trilha l_j do conjunto auxiliar $\overline{\mathcal{L}}_r$ com maior valor de (3.56);
2. Escolher a trilha l_i do conjunto \mathcal{L}_r que minimiza a função (3.51);
3. Acumular o erro de associação da trilha l_i à trilha l_j na variável e ;
4. Remover as trilhas l_i e l_j de \mathcal{L}_r e $\overline{\mathcal{L}}_r$, respectivamente;

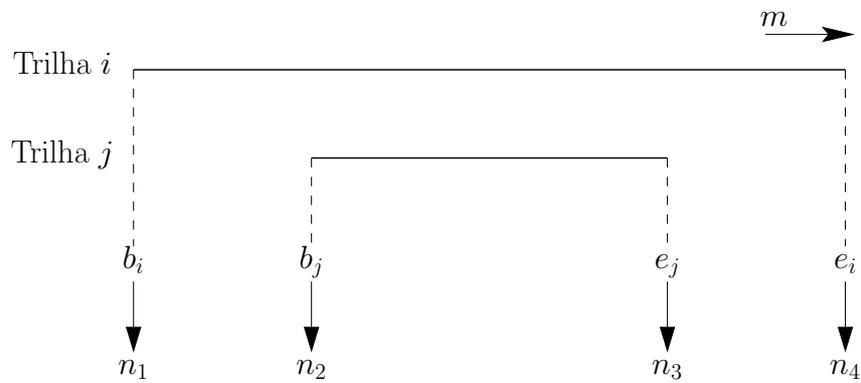
O *loop* termina quando \mathcal{L}_r ou $\overline{\mathcal{L}}_r$ fica vazio, indicando que todas as trilhas de um destes conjuntos foram associadas.



(a) Caso 1



(b) Caso 2



(c) Caso 3

Figura 3.9: Três diferentes possibilidades de superposição de duas trilhas ao longo do tempo e seus respectivos índices obtidos através das expressões de (3.52) até (3.55).

Após a execução do algoritmo, $\bar{\mathcal{L}}_r$ ou \mathcal{L}_r podem não estar vazios. Caso $\bar{\mathcal{L}}_r$ não esteja vazio, algumas trilhas do conjunto de referência não foram associadas a nenhuma trilha do conjunto sob teste. Por outro lado, caso \mathcal{L}_r não esteja vazio, trilhas do conjunto sob teste não foram associadas a nenhuma trilha do conjunto de

Algoritmo 3.1 – Associação dos elementos de \mathcal{L} aos elementos de $\overline{\mathcal{L}}$.

$e \leftarrow 0$

$\overline{\mathcal{L}}_r \leftarrow \overline{\mathcal{L}}$

$\mathcal{L}_r \leftarrow \mathcal{L}$

enquanto $\mathcal{L} \neq \emptyset$ e $\overline{\mathcal{L}}_r \neq \emptyset$ **faça**

 escolha $l_j \in \overline{\mathcal{L}}_r$ tal que $A_j^T = \max_{l_k \in \overline{\mathcal{L}}_r} A_k^T$

 encontre $l_i \in \mathcal{L}_r$ tal que $d(l_i, l_j) = \min_{l_k \in \mathcal{L}_r} d(l_k, l_j)$

$e \leftarrow e + d(l_i, l_j)$

$\overline{\mathcal{L}}_r \leftarrow \overline{\mathcal{L}}_r - \{l_j\}$

$\mathcal{L}_r \leftarrow \mathcal{L}_r - \{l_i\}$

fim enquanto

referência. Dois novos custos são definidos para quantificar estes efeitos, i.e.,

$$e^o = \sum_{l_k \in \mathcal{L}_r} A_k^T, \quad (3.57)$$

$$e^u = \sum_{l_k \in \overline{\mathcal{L}}_r} A_k^T. \quad (3.58)$$

Combinando-se os três erros, pode-se chegar a uma única figura de mérito para avaliar algoritmos de rastreamento de trilhas senoidais. Definindo-se

$$E = \frac{e + e^o + e^u}{N}, \quad (3.59)$$

onde N é o número de elementos do conjunto $\overline{\mathcal{L}}$ de referência, obtém-se uma figura de mérito E que representa o erro médio por trilha no casamento entre os elementos de \mathcal{L} e $\overline{\mathcal{L}}$.

3.6.3 Algoritmo de Avaliação Modificado

Nesta seção, é proposta uma modificação sobre o avaliador descrito anteriormente. É identificado um tipo de erro que não era tratado corretamente pelo avaliador, e em seguida, são propostas modificações para melhorar o desempenho do avaliador.

Dois erros comuns ocorrem em rastreadores de trilhas senoidais:

- Segmentação de uma trilha única de longa duração em várias trilhas curtas;
- União de diversas trilhas de curta duração em uma única trilha longa.

Ao avaliar rastreadores que cometeram erros deste tipo, o Algoritmo 3.1 é muito severo, já que uma trilha de longa duração só pode ser associada a uma única trilha de curta duração. As demais trilhas de curta duração vão contribuir para os erros de e^o ou e^u ou vão ser erroneamente associadas a outras trilhas.

Para contornar este problema, uma modificação no avaliador de [45] é proposta no Algoritmo 3.3. A principal modificação no algoritmo ocorre na maneira como as trilhas são associadas. Enquanto anteriormente todo o suporte temporal de uma trilha era associado a todo o suporte temporal de outra trilha, agora a associação é feita apenas durante a sobreposição temporal das trilhas. O erro de associação modificado de uma trilha l_i com uma trilha l_j , então, é definido como

$$d^{\text{mod}}(l_i, l_j) = \sum_{m_1=n_2}^{n_3} |F_i[m_1] - F_j[m_1]| \cdot |A_i[m_1] - A_j[m_1]|. \quad (3.60)$$

Caso as trilhas não possuam qualquer sobreposição ao longo do tempo (Caso 2 da Figura 3.9) o erro é definido como infinito².

Para que uma trilha de longa duração possa ser associada a diversas trilhas de curta duração, após a associação as trilhas são segmentadas através do Algoritmo 3.2. As trilhas segmentadas, então, são mantidas nos conjuntos de busca, criando a oportunidade de serem associadas a outras trilhas. Caso o erro calculado para uma trilha do conjunto de referência $\bar{\mathcal{L}}$ seja infinito (indicando que nenhuma trilha do conjunto \mathcal{L} possui sobreposição temporal com esta trilha) esta trilha é removida do conjunto de busca $\bar{\mathcal{L}}_r$. Isto garante a condição de parada do avaliador. Os seguintes passos resumem o Algoritmo 3.3:

1. Escolher a trilha l_j do conjunto auxiliar $\bar{\mathcal{L}}_r$ com maior valor de (3.56).
2. Escolher a trilha l_i do conjunto \mathcal{L}_r que minimiza a função (3.51).
3. Se o erro de associação for infinito, indicando que não há sobreposição temporal da trilha l_j com nenhuma trilha do conjunto \mathcal{L}_r , a trilha l_j é removida do conjunto auxiliar $\bar{\mathcal{L}}_r$. Neste caso, o erro e^u é atualizado com o valor A_j^T , de forma similar ao que foi feito no avaliador apresentado anteriormente.

²Isto pode ser implementado na prática definindo-se infinito como um valor negativo qualquer, já que a função d^{mod} só assume valores positivos.

4. Se o erro de associação não for infinito, as trilhas l_i e l_j são segmentadas e as trilhas correspondentes às regiões temporais dessas trilhas que não foram associadas (regiões que não possuem sobreposição temporal) são adicionadas aos conjuntos \mathcal{L}_r e $\overline{\mathcal{L}}_r$, enquanto que as trilhas originais l_i e l_j são removidas. O erro de associação entre as trilhas l_i e l_j é acumulado na variável e .

O algoritmo termina quando o conjunto auxiliar $\overline{\mathcal{L}}_r$ estiver vazio, ou seja, todas as trilhas do conjunto de referência são removidas do conjunto $\overline{\mathcal{L}}_r$ seja devido a uma associação, seja por não possuírem sobreposição temporal com nenhuma trilha do conjunto sob teste. Deve-se ressaltar que caso ocorra um casamento temporal perfeito entre duas trilhas, o algoritmo de segmentação retornará dois conjuntos vazios e, conseqüentemente, nenhuma nova trilha será adicionada aos conjuntos $\overline{\mathcal{L}}_r$ e \mathcal{L}_r . Neste caso particular, o avaliador proposto se comporta como o avaliador descrito em [45].

Ao final da execução do Algoritmo 3.3, ainda resta computar o erro associado às trilhas remanescentes do conjunto \mathcal{L}_r , que é dado por

$$e^o = \sum_{l_k \in \mathcal{L}_r} A_k^T. \quad (3.61)$$

O erro médio por trilha no casamento dos elementos de \mathcal{L} e $\overline{\mathcal{L}}$ passa a ser definido como

$$E^{\text{mod}} = \frac{e + e^o + e^u}{N}, \quad (3.62)$$

onde, novamente, N corresponde ao número de trilhas no conjunto de referência $\overline{\mathcal{L}}$.

É esperado que com as mudanças o avaliador se comporte de maneira mais adequada para os dois tipos de erros citados no início desta seção. Por exemplo, no avaliador proposto uma trilha de longa duração do conjunto de referência pode ser associada a diversas trilhas do conjunto sob teste. O erro de amplitude total só é utilizado para as regiões temporais da trilha de longa duração que não conseguiram se associar a nenhuma trilha do conjunto sob teste.

Uma quantificação apenas dos erros citados no início da seção pode ser útil para a comparação de diferentes rastreadores. Para isto, duas variáveis são utilizadas: s_i^a e s_j^c . A variável s_i^a é incrementada de uma unidade toda vez que uma trilha $l_i \in \mathcal{L}$ é associada a um elemento de $\overline{\mathcal{L}}$, indicando se uma trilha no conjunto de referência é associada a mais de uma trilha no conjunto sob teste. Em contrapartida, s_j^c é

Algoritmo 3.2 – Segmentação de uma trilha l_i em duas outras trilhas definidas pelos quadros n_1, n_2, n_3 e n_4 .

Entradas: l_i, n_1, n_2, n_3 e n_4

se $b_i \leq n_1$ e $n_1 \neq n_2$ **então**

$l_a[m] \leftarrow l_i[m]$ para $n_1 \leq m < n_2$

senão

$l_a \leftarrow \emptyset$

fim se

se $e_i \geq n_4$ e $n_3 \neq n_4$ **então**

$l_b[m] \leftarrow l_i[m], n_3 < m \leq n_4$

senão

$l_b \leftarrow \emptyset$

fim se

$\mathcal{L}_{\text{aux}} = \{l_a, l_b\}$

Saída: \mathcal{L}_{aux}

incrementada toda vez que uma trilha $l_j \in \overline{\mathcal{L}}$ é associada a um elemento de \mathcal{L} . Idealmente, após a execução do algoritmo, essas duas variáveis deveriam possuir o valor 1 para todas as trilhas em ambos conjuntos. De modo a consolidar s^a e s^c em uma única medida, propõe-se fazer

$$E^s = \frac{1}{M} \sum_{i=1}^M (s_i^a - 1) + \frac{1}{N} \sum_{j=1}^N (s_j^c - 1), \quad (3.63)$$

onde M é o número de trilhas no conjunto \mathcal{L} .

Idealmente, as figuras de mérito E^{mod} e E^s seriam medidas mais confiáveis do que apenas E para se analisar o desempenho de algoritmos de rastreamento de trilhas. Isto acontece porque no avaliador proposto é considerado um número maior de erros que podem ser cometidos por um rastreador. Isto permite que seja dado um tratamento mais adequado aos diferentes tipos de erros. No entanto, a definição de quão “parecidas” são duas trilhas ainda é um problema em aberto, e não se pode afirmar que o avaliador proposto sempre dará resultados mais consistentes do que o avaliador descrito em [45].

Algoritmo 3.3 – Associação dos elementos de \mathcal{L} aos elementos de $\overline{\mathcal{L}}$ considerando as modificações propostas.

$e \leftarrow 0$

$e^u \leftarrow 0$

$\overline{\mathcal{L}}_r \leftarrow \overline{\mathcal{L}}$

$\mathcal{L}_r \leftarrow \mathcal{L}$

enquanto $\overline{\mathcal{L}}_r \neq \emptyset$ **faça**

 escolha $l_j \in \overline{\mathcal{L}}_r$ tal que $A_j^T = \max_{l_k \in \overline{\mathcal{L}}_r} A_k^T$

 encontre $l_i \in \mathcal{L}_r$ tal que $d^{\text{mod}}(l_i, l_j) = \min_{l_k \in \mathcal{L}_r} d(l_k, l_j)$

se $d^{\text{mod}}(l_i, l_j) = \inf$ **então**

$e^u \leftarrow e^u + A_j^T$

$\overline{\mathcal{L}}_r \leftarrow \overline{\mathcal{L}}_r - \{l_j\}$

senão

$e \leftarrow e + d^{\text{mod}}(l_i, l_j)$

 Aplicar o Algoritmo 3.2 em l_j obtendo \mathcal{L}_j

$\overline{\mathcal{L}}_r \leftarrow \overline{\mathcal{L}}_r - \{l_j\}$

$\overline{\mathcal{L}}_r \leftarrow \overline{\mathcal{L}}_r \cup \mathcal{L}_j$

 Aplicar o Algoritmo 3.2 em l_i obtendo \mathcal{L}_i

$\mathcal{L}_r \leftarrow \mathcal{L}_r - \{l_i\}$

$\mathcal{L}_r \leftarrow \mathcal{L}_r \cup \mathcal{L}_i$

fim se

fim enquanto

Capítulo 4

Desenvolvimento de um Banco de Trilhas Senoidais

4.1 Introdução

A comparação de algoritmos de análise senoidal utilizando sinais naturais é difícil, principalmente por não se saber, antes da análise, quais trilhas senoidais deveriam ser extraídas do sinal. Por exemplo, em [101] é realizada uma avaliação comparativa de diversos sistemas de análise utilizando-se sinais de teste naturais. Para se comparar os diversos sistemas de análise, as trilhas obtidas pelos diferentes sistemas foram reconstruídas por um sintetizador padronizado. A avaliação foi feita de acordo com critérios pré-definidos utilizados em testes subjetivos de audição sobre os sinais sintetizados. A avaliação feita em [101], no entanto, não consegue fornecer valores quantitativos que possam auxiliar na escolha de um determinado sistema de análise senoidal em detrimento de outro. Além disso, a informação fornecida pelo teste subjetivo pode ser de pouca valia, dependendo da aplicação final a que se destinam as trilhas senoidais.

Para a comparação objetiva de sistemas de análise senoidal, uma possível solução consiste na utilização de sinais em que se conhece, *a priori*, um conjunto de trilhas que o represente. Desta forma, podem ser utilizadas medidas objetivas tais como as apresentadas nas Seções 2.6 e 3.6. Sinais sintéticos se enquadram no requisito acima. Para um sinal natural, no entanto, as trilhas senoidais precisam ser extraídas por algum sistema ‘padrão’ de análise senoidal para, então, servirem como referência

para outros sistemas.

Neste capítulo, é apresentado o desenvolvimento de um banco de referência contendo sinais-fonte com representações por trilhas senoidais associadas. Tal corpus, chamado BTS (Banco de Trilhas Senoidais) forneceria os dados necessários para uma possível avaliação objetiva de desempenho de sistemas de análise senoidal. O BTS será gerado a partir de sinais sintéticos e naturais. É necessário definir cuidadosamente quais métodos de análise senoidal serão utilizados, podendo-se utilizar intervenção manual para assegurar que as trilhas extraídas são, de fato, representativas.

Na Seção 4.2 são definidas as características desejadas dos sinais a serem analisados. Os sinais escolhidos são descritos na Seção 4.3. Na Seção 4.4 são especificadas as trilhas do BTS. Já na Seção 4.5, é definida a especificação para os métodos de construção do BTS, incluindo as diretrizes que deverão guiar a análise e, principalmente, os erros tolerados para a estimação dos parâmetros das trilhas. Já na Seção 4.6, são descritas as etapas que levam à geração do BTS. Na Seção 4.7, são apresentadas as trilhas obtidas, bem como o procedimento utilizado para sua extração. Finalmente, na Seção 4.8 são realizados testes para se verificar se o BTS obedece à especificação realizada.

No Capítulo 5, serão apresentadas avaliações de sistemas de análise senoidal com o objetivo de demonstrar a utilização do BTS em conjunto com avaliadores de qualidade.

4.2 Especificação e Categorização dos Sinais-Fonte

Nesta seção são definidas as características desejadas dos sinais de áudio a serem incluídos no BTS. Estes sinais-fonte são especificados de acordo com níveis crescentes de complexidade de conteúdo, o que permite avaliar o desempenho de um sistema de análise senoidal sob diferentes graus de dificuldade. Assim, diversos desafios (como polifonia, *vibrato*, etc) para sistemas de análise senoidal serão introduzidos na passagem de um nível de complexidade para o outro.

Seguindo a divisão proposta em [102], quatro níveis de crescente complexidade

são especificados abaixo.

Nível 0 – Contém sinais criados a partir de um sintetizador de sons de instrumentos¹ musicais.

Nível 1 – Contém gravações de notas musicais isoladas emitidas por instrumentos musicais, um por vez.

Nível 2 – Contém gravações de frases musicais isoladas executadas em instrumentos musicais, um por vez.

Nível 3 – Contém gravações de trechos de música executados por mais de um instrumento musical por vez.

O Nível 0 é incluído para fornecer dados referenciais que permitam verificar se os métodos de análise são capazes de estimar trilhas dentro das especificações pré-definidas para o BTS. Para isto, os parâmetros de síntese devem ser fornecidos num formato que permita a comparação com as trilhas obtidas no BTS. Idealmente, cada sinal do Nível 0 deve estressar uma determinada especificação do BTS. O Nível 1 contempla o caso mais simples envolvendo sinais naturais. O Nível 2 já permite a avaliação de desempenho de algoritmos de análise senoidal para casos com *vibrato* e *tremolo*, além do comportamento dos algoritmos durante mudanças de notas musicais. Por fim, o Nível 3 permite a avaliação do desempenho de algoritmos de análise senoidal para sinais exibindo diferentes graus de polifonia e, possivelmente, o cruzamento de trilhas senoidais.

Os sinais dos Níveis de 1 a 3 deverão ser gravações contendo emissões de instrumentos acústicos (não-sintéticos). A duração dos sinais não deve ser inferior a 2 segundos nem superior a 15 segundos, de modo a permitir a realização de testes subjetivos [103].

Todas as gravações devem possuir apenas 1 canal de áudio, com taxa de amostragem igual a 44,1 kHz e precisão de 16 bits. Os sinais gravados com uma taxa de amostragem e/ou uma precisão maior que as especificadas devem ser convertidos para a taxa e precisão desejadas. Os sinais de áudio devem ser normalizados, de modo que o módulo de seu valor máximo seja igual a 1, e ser armazenados no formato WAVE [104].

¹Por simplicidade, será omitido o termo “som” quando o sentido estiver claro.

4.3 Sinais-Fonte Escolhidos

Nesta seção, são descritos os sinais-fonte escolhidos para compor uma primeira versão do BTS. A seleção foi feita de modo a obedecer às especificações feitas na Seção 4.2. Outros sinais poderão ser adicionados ao banco posteriormente, e deverão ser escolhidos de forma a complementar os sinais aqui descritos.

Os sinais dos Níveis de 1 a 3 são excertos de gravações retirados do banco de sinais *Real World Computing Database* (RWC) [105, 106, 107, 108]. Essa escolha foi motivada pelas seguintes propriedades deste banco:

1. O formato dos arquivos obedece às especificações do BTS;
2. As gravações foram realizadas em estúdios profissionais;
3. Há grande variedade de sinais presentes na RWC, o que permite escolher sinais para os Níveis de 1 a 3 do BTS;
4. Está disponível na literatura [107] documentação detalhada sobre os sinais;
5. Dispõe-se de anotações de *onset* e *offset* e transcrição em partitura [108];

A seguir, serão descritos em separado os sinais escolhidos para cada nível.

4.3.1 Nível 0

Os sinais do Nível 0 foram gerados especificamente para o BTS. Para isto, foi escolhido um sintetizador descrito em [86], baseado numa modificação de síntese aditiva, que se mostrou adequado para emular instrumentos de sopro. O sintetizador utiliza um banco de osciladores senoidais, cada um representando uma parcial do sinal que está sendo gerado. Dois sinais de controle são fornecidos a cada oscilador: os valores da amplitude e da frequência em cada instante de tempo. As principais vantagens deste esquema de síntese são a sua simplicidade e o fato de os parâmetros utilizados na síntese serem análogos aos parâmetros obtidos na análise senoidal.

Escolheu-se sintetizar os sinais com um modelo de trompete, devido ao elevado número de parciais proeminentes exibido por este instrumento, resultando num número de trilhas senoidais igualmente elevado. Em [109], são descritos detalhadamente o método de síntese utilizado e os seus parâmetros.

Foram criados três sinais sintéticos para o BTS:

1. um sinal exibindo *vibrato* para o teste de conformidade do BTS com a especificação do erro de estimação de frequência em condição não-estacionária;
2. um sinal exibindo *tremolo* para o teste de conformidade do BTS com a especificação do erro de estimação de amplitude em condição não-estacionária;
3. um sinal contendo um trecho de silêncio contaminado com ruído para a verificação da conformidade do detector de início e fim das trilhas.

O sinal 1 (*vibrato*) dura 2 segundos, possui frequência fundamental igual a 156 Hz e um total de 26 parciais. O *vibrato* foi emulado utilizando-se uma taxa de 4 Hz e uma profundidade dependente do harmônico que estava sendo sintetizado. Matematicamente, o *vibrato* forçado sobre o n -ésimo harmônico pode ser descrito por uma variação de frequência igual a $(n + 1) \text{sen}(2\pi 4t)$. Estes valores escolhidos para o *vibrato* são similares aos utilizados por sintetizadores relatados na literatura [110].

O sinal 2 (*tremolo*) é similar ao sinal 1 em termos de frequência fundamental e duração. No lugar do *vibrato*, um *tremolo* com taxa de 8 Hz foi adicionado ao sinal. A extensão do *tremolo* é a mesma para todos os harmônicos, sendo igual a 0,5.

O sinal contendo trechos de silêncio foi sintetizado com três notas de *pitch* fixo executadas sequencialmente. A primeira nota possui frequência fundamental igual a 156 Hz, 26 parciais e duração de 0,5 s. A segunda nota, que começa imediatamente após a primeira, possui uma frequência fundamental de 262 Hz, 21 parciais e duração de 1 s. Foi inserido entre a segunda e a terceira nota um trecho de 0,2 s de silêncio. A última nota possui frequência fundamental igual a 622 Hz, 10 parciais e duração de 0,2 s. A duração total do sinal sintetizado é de 2 s. Foi adicionado ruído rosa ao sinal sintetizado de modo a se obter uma SNR de 60 dB.

Uma vez sintetizados, os sinais foram armazenados em arquivos WAVE com uma frequência de amostragem igual a 44,1 kHz e precisão de 16 bits.

4.3.2 Nível 1

Na Tabela 4.1, podem ser vistas informações sobre os sinais escolhidos para o Nível 1 do BTS. Os sinais escolhidos foram retirados do banco de sons de instrumentos

musicais do RWC [106]. Conforme especificado, cada sinal contém a emissão de apenas uma nota musical. Na tabela, a coluna ‘Ref. RWC’ indica o DVD de onde o sinal foi retirado, enquanto a coluna ‘Arquivo’ indica o nome do arquivo selecionado. É exibida na coluna ‘f0’, a frequência fundamental da nota sendo executada.

Tabela 4.1: Informações sobre os sinais escolhidos para o Nível 1 do BTS.

Nome	Ref. RWC	Arquivo	f0	Duração
Violino	RWC-MDB-I-2001-W05	151VNNVM	247 Hz	3,1 s
Piano	RWC-MDB-I-2001-W01	011PFNOF	207 Hz	2,2 s
Saxofone	RWC-MDB-I-2001-W08	283BANOF	148 Hz	4,2 s
Acordeão	RWC-MDB-I-2001-W02	071ACR3F	312 Hz	2,5 s

Em geral, os sinais selecionados são tonais e, aparentemente, deveriam ser bem modelados por um sistema de análise senoidal. Procurou-se escolher sinais contendo emissões de instrumentos musicais com diferentes características e de diferentes famílias. Além disso, os registros exibem um elevado número de parciais devido à baixa frequência fundamental, permitindo que um número elevado de trilhas sejam potencialmente extraídas de cada sinal. A Tabela 4.2, resume as características dos sinais do Nível 1.

Tabela 4.2: Características dos sinais escolhidos para o Nível 1 do BTS.

Nome	Característica
Violino	Presença de <i>vibrato</i> e uma dinâmica <i>mezzo forte</i> .
Piano	Execução convencional, sem pedal, e com dinâmica <i>forte</i> .
Saxofone barítono	Execução convencional com dinâmica <i>forte</i> .
Acordeão	Execução convencional com dinâmica <i>forte</i> .

4.3.3 Nível 2

Os sinais selecionados para o Nível 2 foram retirados dos bancos de gravações de música clássica e jazz, ambos pertencentes ao RWC [107]. A Tabela 4.3 informa

sobre a identificação e localização desses sinais no RWC. Na tabela, a coluna ‘Ref. RWC’ indica o CD de onde o sinal foi retirado, enquanto a coluna ‘Número’ indica o número da faixa escolhida. São exibidos na coluna ‘Segmento’ os instantes de início e fim dos trechos extraídos de cada gravação.

Os sinais possuem gravações de frases musicais executadas por apenas um instrumento musical. São exibidos na Tabela 4.4 o compositor e a composição executada, para cada um dos sinais. Para que a duração dos sinais respeitasse a especificação do BTS, foi utilizado apenas um trecho de cada composição.

Tabela 4.3: Identificação e localização no RWC dos sinais escolhidos para o Nível 2 do BTS.

Nome	Ref. RWC	Número	Segmento (min:s)	Duração
Órgão	RWC-MDB-C-2001-M02	11	0:06 até 0:17	10,9 s
Violino	RWC-MDB-C-2001-M06	36	0:36 até 0:40	3,4 s
Piano	RWC-MDB-J-2001-M01	1	0:00 até 0:10	10,2 s
Violoncelo	RWC-MDB-C-2001-M06	41	0:00 até 0:10	10,2 s

Tabela 4.4: Informação das composições executadas nos sinais do Nível 2 do BTS.

Nome	Compositor	Composição
Órgão	Johann S. Bach	Passacaglia e Fuga em Dó menor
Violino	Johann S. Bach	Partitura nº 6 para Violino Solo
Piano	Não Identificado	Identificada como <i>Jive</i>
Violoncelo	Johann S. Bach	Suite nº 1 para Violoncelo Solo em Sol maior

Cada sinal deste nível foi escolhido por possuir uma característica importante para a avaliação de um sistema de análise senoidal. O sinal de órgão possui notas graves, incluindo a execução de uma nota com frequência fundamental de 62 Hz. O sinal de violoncelo possui transições rápidas entre notas, o que dificulta a detecção de início e fim das trilhas. Já o sinal de violino exibe *vibrato* e *tremolo*, o que pode dificultar o rastreamento de trilhas. Por fim, no sinal de piano acordes são executados, sendo este o primeiro sinal do BTS que exibe polifonia. De modo geral,

os sinais selecionados para o Nível 2 apresentam uma dificuldade de análise superior aos sinais do Nível 1. Deve-se ressaltar, no entanto, que os sinais contêm emissões de instrumentos musicais que podem ser bem modeladas por trilhas senoidais. Na Tabela 4.5, é feito um resumo das características dos sinais do Nível 2.

Tabela 4.5: Características dos sinais no Nível 2 do BTS.

Nome	Característica
Órgão	Notas graves, porém sem acordes. Andamento lento.
Violino	Notas exibindo <i>vibrato</i> . Andamento moderado.
Piano	Presença de acordes. Andamento rápido.
Violoncelo	Notas de curta duração e <i>vibrato</i> . Andamento moderado.

4.3.4 Nível 3

Na Tabela 4.6, podem ser vistas informações sobre o sinal selecionado para o Nível 3 do BTS, que foi retirado do banco de gravações de jazz pertencente ao RWC [107]. O sinal possui um trecho de uma gravação da música *Crescent Serenade* de H. Kobayashi. Para que a duração dos sinais respeitasse a especificação do BTS, apenas um trecho da composição foi utilizado. Na tabela, a coluna ‘Ref. RWC’ indica o CD de onde o sinal foi retirado, enquanto a coluna ‘Número’ indica o número da faixa escolhida. São exibidos na coluna ‘Segmento’ os instantes de início e fim do trecho extraído da gravação.

Tabela 4.6: Identificação e localização no RWC do sinal selecionado para o Nível 3 do BTS.

Nome	Ref. RWC	Número	Segmento (min:s)	Duração
Quarteto	RWC-MDB-J-2001-M02	24	0:38 até 0:41	3,1 s

O sinal escolhido envolve a execução dos seguintes instrumentos musicais: piano, baixo acústico, bateria e saxofone tenor. A principal dificuldade de modelagem imposta por este sinal, além da polifonia, é a presença de um instrumento percussivo, que, idealmente, não deve ser modelado.

Futuramente, é prevista a adição de mais sinais neste nível. Em particular, pretende-se incluir um sinal com cruzamento de parciais. Outros sinais que serão adicionados neste nível incluem, por exemplo, gravações de orquestra sinfônica e de duetos de violinos.

4.4 Especificação do Banco de Trilhas Senoidais

Uma trilha senoidal será composta por três vetores, contendo amostras da trajetória temporal dos seguintes parâmetros: amplitude (em escala linear), frequência (em Hz) e fase (em radianos). Além desses vetores, uma trilha deve armazenar o instante de tempo (em segundos) em que foi detectado o primeiro elemento de cada vetor. O banco de trilhas não considerará o tempo de ocorrência de cada elemento destes vetores como um possível quarto parâmetro, uma vez que o avanço temporal da janela de observação (ver Seção 2.2) é mantido constante. Assim, algoritmos como o descrito em [111] não serão empregados como ferramenta de análise para a obtenção do BTS.

A taxa de amostragem dos parâmetros das trajetórias das trilhas do BTS também deve ser especificada. No caso de um sistema de análise que utilize uma janela para observação do sinal, a taxa de amostragem dos parâmetros das trilhas é definida pelo salto entre quadros consecutivos. Usualmente, espera-se que o conteúdo espectral das trajetórias seja limitado em 20 Hz [112], o que levaria a uma taxa de amostragem mínima de 40 Hz. No entanto, são utilizadas taxas mais elevadas em sistemas de análise senoidal, pois facilitam a obtenção das trilhas. Além disso, é provável que os parâmetros das trilhas do BTS tenham que ser subamostrados de modo a permitir a comparação com trilhas obtidas por diferentes algoritmos de análise. Assim, para evitar a introdução de erros indesejáveis nos parâmetros no caso de uma possível necessidade de interpolação de suas trajetórias, é razoável que as trilhas fornecidas pelo banco possuam uma taxa de amostragem superior às de trilhas obtidas por sistemas de análise relatados na literatura. Considerando que taxas de até 200 Hz são reportadas em [10], a taxa de amostragem de 500 Hz foi considerada adequada para os parâmetros das trilhas do BTS.

4.5 Especificação dos Métodos de Análise

Nesta seção, são realizadas especificações com o objetivo de guiar a escolha do sistema de análise senoidal a ser utilizado. Através desta especificação, espera-se garantir a qualidade do BTS e sua aplicabilidade à avaliação de sistemas de análise senoidal. Cabe ressaltar aqui que só é possível verificar se as especificações realizadas nesta seção são atendidas no caso dos sinais sintéticos. É esperado que se os métodos de análises empregados se adequem às especificações para os sinais sintéticos, eles também o conseguirão para os demais sinais.

Na Seção 4.5.1, são definidas as diretrizes que deverão guiar o processo de análise, enquanto na Seção 4.5.2 são especificados valores aceitáveis para os erros de estimação dos parâmetros (amplitude, frequência e fase) de uma trilha.

4.5.1 Diretrizes de Análise

Devido às diferenças entre os sinais presentes no BTS e às diversas características dos sistemas de análise, definir um conjunto de especificações para a extração das trilhas poderia favorecer um determinado algoritmo de análise senoidal. Isto é indesejável, pois poderia polarizar comparações entre sistemas de análise senoidal utilizando o BTS. Para isso, no lugar de especificações, são fornecidas diretrizes que devem auxiliar na extração das trilhas.

No BTS, uma trilha de um sinal-fonte deverá iniciar e terminar juntamente com a percepção do evento musical que a provocou. No caso de a(s) nota(s) sendo executada(s) exibir(em) *glissando* ou *legato*, as trilhas devem capturar estes efeitos, se eles forem percebidos como um mesmo evento sonoro (i.e. sem a percepção de ataque de novas notas).

Para se determinar quais picos senoidais devem ser detectados num determinado instante de tempo, pode ser utilizado um algoritmo de estimação do chão de ruído [64]. Devem ser utilizados para formar as trilhas apenas os picos espectrais cujas magnitudes se situam acima do chão de ruído estimado.

No caso de a inobservância de picos espectrais por certo intervalo de tempo resultar em interrupção obviamente imprópria de uma trilha, pode-se lançar mão de um algoritmo de interpolação para estimar os parâmetros faltantes. Já que os

parâmetros das trilhas são sobreamostrados, o erro cometido por tal interpolação será baixo se ela for utilizada um número reduzido de vezes. Poderá ser interpolada de uma dada falha informação de até 8 ms (o que equivale a 4 amostras consecutivas dos parâmetros). Esta duração foi adotada em função do limite abaixo do qual, de acordo com [113, 114], seres humanos não percebem emudecimentos em sinais senoidais [113, 114], de forma que apenas uma falha não perceptível fosse ligada. Apesar de este valor depender da frequência e da intensidade sonora da senóide utilizada no teste, o limiar superior de 8 ms foi reportado para frequências acima de 200 Hz e intensidades sonoras acima de 50 dB SL², condições consideradas típicas para o BTS, daí ter sido este o valor adotado. Caso um intervalo superior a 8 ms tenha que ser interpolado, a trilha em questão deve ser dividida em duas, mesmo que isso não obedeça ao critério sobre o início e fim das trilhas definido abaixo.

Para se determinar o instante de início (*onset*) e de fim (*offset*) de uma trilha, poderão ser utilizados algoritmos de detecção de *onset* [115]. Usualmente, esses algoritmos possuem uma tolerância de 50 ms [115, 116, 117]. Este erro, no entanto, é elevado se comparado com a resolução temporal humana, que pode chegar a 8 ms, como discutido no parágrafo anterior. Como um valor de compromisso, adota-se um erro máximo de 20 ms para a estimação dos instantes inicial e final de uma trilha.

4.5.2 Especificação dos Erros de Estimação

Nesta seção, são especificados os erros máximos tolerados para os valores estimados dos parâmetros das trilhas. O objetivo é fornecer um intervalo de confiança sobre os parâmetros a serem estimados. Assim, avaliações de desempenho que utilizam o BTS podem levar em conta a incerteza associada aos parâmetros de referência.

Serão utilizados critérios psicoacústicos para a especificação dos erros máximos permitidos. A idéia é usar a chamada JND (*Just Noticeable Difference*), que fornece, para uma determinada quantidade física, a menor diferença de intensidade perceptível por um ser humano. Serão aplicados resultados relatados na literatura para a percepção de frequência e intensidade sonora (amplitude). Em princípio, caso

²Intensidade em dB acima do limiar de audibilidade [114] (do inglês, *Sensation Level*) para uma dada frequência. Duas senóides de frequências distintas com a mesma intensidade sonora em dB SL, são, idealmente, percebidas com a mesma intensidade por um ouvinte.

o erro nas estimativas de parâmetros seja menor que as respectivas JNDs, então o erro pode ser considerado imperceptível.

Frequência

A JND de frequência de tons puros [114] fornece o incremento necessário (δf) sobre uma determinada frequência (f) para que ela seja percebida como uma frequência diferente por um ser humano. Este incremento é obtido através de testes subjetivos e é dependente [114] do estímulo escolhido no teste. Neste trabalho, os resultados obtidos para estímulos senoidais foram escolhidos pois fornecem os valores mais restritivos para a JND.

Para senóides de curta duração, foi observado experimentalmente que a JND é função da frequência e da intensidade sonora do estímulo [118], sendo que, para uma dada frequência, quanto maior a intensidade sonora menor a JND. Em [119], foi reportada uma dependência linear do logaritmo do incremento com a raiz quadrada da frequência:

$$\log_{10} \delta f = a\sqrt{f} + b, \quad (4.1)$$

onde parâmetros a e b dependem da intensidade do estímulo e são escolhidos de modo a aproximar resultados experimentais.

O erro máximo aceitável para as estimativas de f é definido como

$$e^{\text{freq}}(f) = \frac{\delta f}{f}, \quad (4.2)$$

onde δf é obtida através da expressão (4.1).

A intensidade sonora da reprodução de uma trilha (após a síntese) é desconhecida, por isso foi escolhida a maior intensidade testada em [119], i.e. 80 dB SL, e que leva aos incrementos mais restritivos. Para esta intensidade, os valores de a e b obtidos são 0,028 e $-0,696$, respectivamente.

A Figura 4.1 mostra o erro relativo máximo especificado para frequências variando de 0,02 até 10 kHz. O menor valor ocorre em $f = 962$ Hz, onde o erro especificado é de aproximadamente 0,15 %. Será utilizada para fornecer a tolerância para o erro de estimação de frequência a equação (4.2), sendo, assim, a tolerância variável com a frequência.

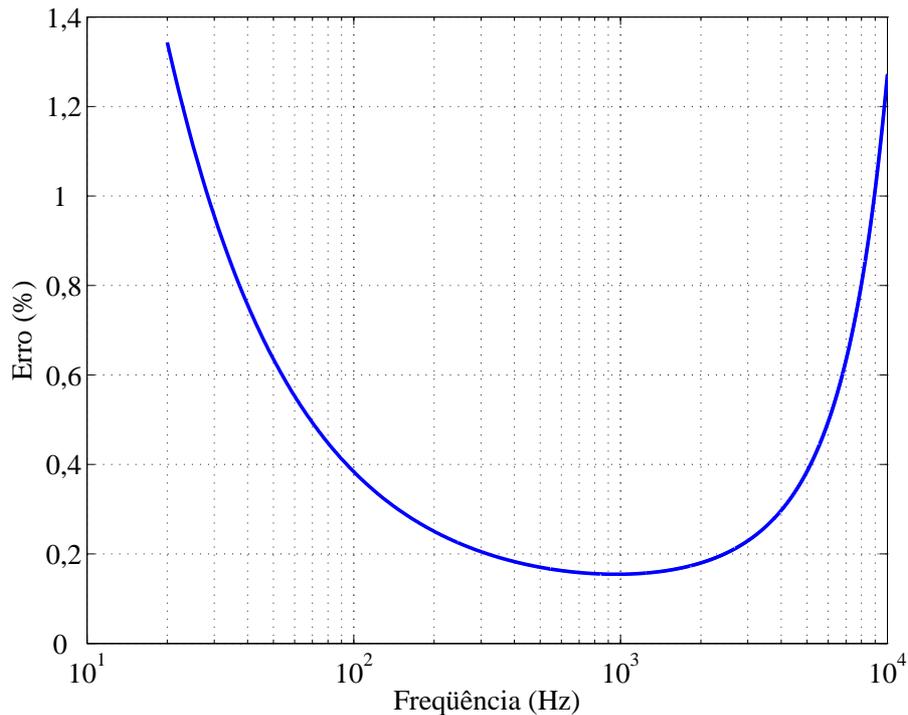


Figura 4.1: Curva de erro percentual máximo (Eq. (4.2) para 80 dB SL) aceitável na estimativa de f .

Amplitude

O procedimento para determinar o maior erro tolerado de amplitude seguirá os mesmos passos adotados para a definição das tolerâncias nos erros de estimação da frequência. Inicialmente, será descrito um estudo psicoacústico sobre a percepção da intensidade sonora. A partir deste estudo, será definido o erro máximo permitido para a estimação da amplitude das trilhas ao longo do tempo. Assim como para a frequência, a JND da intensidade sonora percebida é dependente do tipo de estímulo [114]. Novamente, foram escolhidos resultados obtidos com senóides de curta duração.

Em [120], foram feitos testes subjetivos para estimar o menor incremento de intensidade δI sobre uma senóide de intensidade I que ainda é percebido por um ouvinte. O estudo procurou estimar δI para diversos valores de intensidade sonora I e para senóides com diferentes frequências. Uma influência desprezível da frequência da senóide sobre δI é relatada, o que está de acordo com outros resultados encontrados na literatura [121, 114].

A dependência de δI em relação a I encontrada em [120] pode ser descrita através da seguinte expressão

$$\frac{\delta I}{I} = 0,463I^{-0,072}. \quad (4.3)$$

Os valores 0,463 e 0,072 foram encontrados de modo a aproximar os resultados de testes subjetivos. O comportamento da razão $\frac{\delta I}{I}$ em função de I (em dB SL) pode ser visto na Figura 4.2. Como pode ser observado, a razão $\frac{\delta I}{I}$ é tão menor quanto maior for a intensidade sonora.

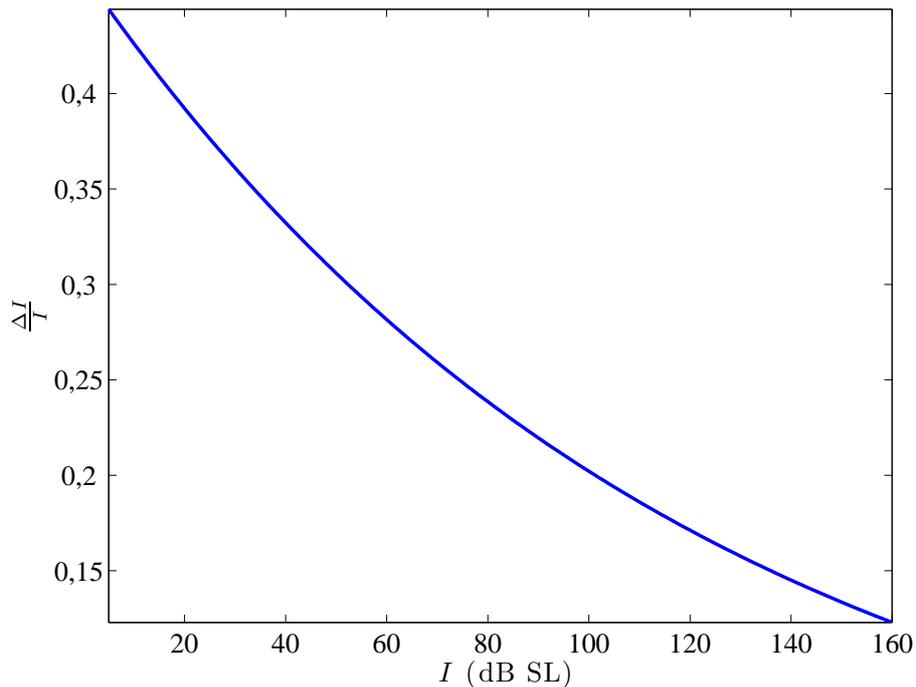


Figura 4.2: Incremento relativo necessário para que uma mudança em intensidade sonora seja percebida em um estímulo senoidal.

O erro relativo máximo da estimação da amplitude foi escolhido como 20 % e, portanto, situa-se sempre abaixo da curva da Figura 4.2 para intensidades sonoras tão elevadas quanto 100 dB SL. Apesar de conservadora, essa intensidade foi escolhida para contornar a incerteza associada à intensidade sonora com que seria reproduzido o sinal resintetizado a partir de seu modelo senoidal.

Fase

As diferenças em fase absoluta e fase relativa de sons tonais são pouco percebidas por seres humanos [114, 121]. Contudo, a estimação de fase pode ser importante

para algoritmos de síntese e de modificação das trilhas senoidais [28, 18]. Por esse motivo, o banco de trilhas senoidais inclui estimativas da fase das trilhas para cada instante de tempo analisado. A especificação do erro máximo, no entanto, se faz pouco necessária, já que os valores de fase variam muito pouco ao longo da frequência na proximidade de uma componente ressonante [112, 71].

4.6 Métodos de Análise

Nesta seção, são descritos os métodos de análise senoidal adotados para se construir o BTS, que foram escolhidos seguindo os seguintes critérios:

1. Ser capaz de extrair trilhas que respeitem as especificações definidas nas Seções 4.4 e 4.5;
2. Permitir o ajuste manual dos parâmetros para cada sinal a ser analisado.

A seguir, serão apresentadas as etapas de processamento do sinal na ordem em que são aplicadas ao sinal, sendo definidos os parâmetros de análise dos métodos que não são ajustados manualmente. Já na Seção 4.7, são comentados os parâmetros restantes e o critério utilizado para ajustá-los.

Cada método utilizado foi implementado como uma função do *software* Matlab® [122].

4.6.1 Divisão em Sub-Bandas

O primeiro passo da obtenção do BTS consiste na divisão do sinal em R sub-bandas de frequência. Esta estratégia foi adotada para permitir um controle maior dos parâmetros de análise e o tratamento por faixas de frequência dos erros decorrentes da estimação dos parâmetros. Para a divisão espectral do sinal, foi utilizado o método descrito em [123, 23], que emprega uma estrutura chamada Pirâmide de Laplace, que é composta por níveis como o mostrado na Figura 4.3. O filtro *anti-aliasing* $h[n]$ é um filtro passa-baixas FIR com fase linear de ordem 256 e atraso d igual a 128, projetado utilizando a função `firpm` do Matlab. A sua faixa de passagem vai até $0,4\pi$ rad e sua faixa de rejeição começa em $0,45\pi$ rad com uma atenuação de 100 dB. Na Figura 4.4, pode ser observada a resposta em magnitude do filtro na

freqüência. A divisão em sub-bandas é obtida através da aplicação consecutiva da estrutura exibida na Figura 4.3 ao sinal $s_l[n]$.

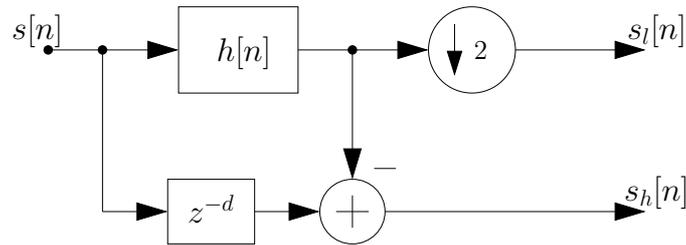


Figura 4.3: Elemento básico da estrutura utilizada para se dividir o sinal em sub-bandas por oitavas. O sinal $s[n]$ é dividido em duas sub-bandas: uma contendo o conteúdo de alta-freqüência $s_h[n]$ e outra com o de baixa freqüência expandido $s_l[n]$. Novas sub-bandas podem ser obtidas a partir da reaplicação da estrutura ao sinal $s_l[n]$.

Na Figura 4.5, é exibido um exemplo para a divisão em freqüência resultante para o caso de quatro sub-bandas. Pode-se perceber pela figura que existe uma sobreposição entre as sub-bandas, característica que será utilizada favoravelmente durante o rastreamento das trilhas. Por conveniência, as sub-bandas são numeradas em ordem decrescente de taxa de amostragem, sendo denotada por $r = 1$ a sub-banda com maior taxa.

Cada uma das etapas descritas da Seção 4.6.2 até a Seção 4.6.6 será executada individualmente para cada sub-banda.

Deve-se ressaltar que no atual conjunto de sinais-fonte não foi necessário rastrear uma trilha com picos detectados em mais de uma sub-banda. Com isso, é feito um simples agrupamento, descrito na Seção 4.6.7, das trilhas obtidas em diferentes sub-bandas. O problema do rastreamento de uma trilha presente em mais de uma sub-banda será tratado futuramente, caso se faça necessário. Em [124], é descrita uma solução para um problema similar, porém aplicada à síntese sonora.

4.6.2 Mapeamento Tempo-Freqüência

O espectro de cada sub-banda é obtido utilizando-se uma STFT (ver Seção 2.2). Para isto, o sinal é segmentado em blocos de comprimento N de acordo com (2.1), utilizando-se uma janela de suavização de Hann [54]. Procurou-se escolher um salto

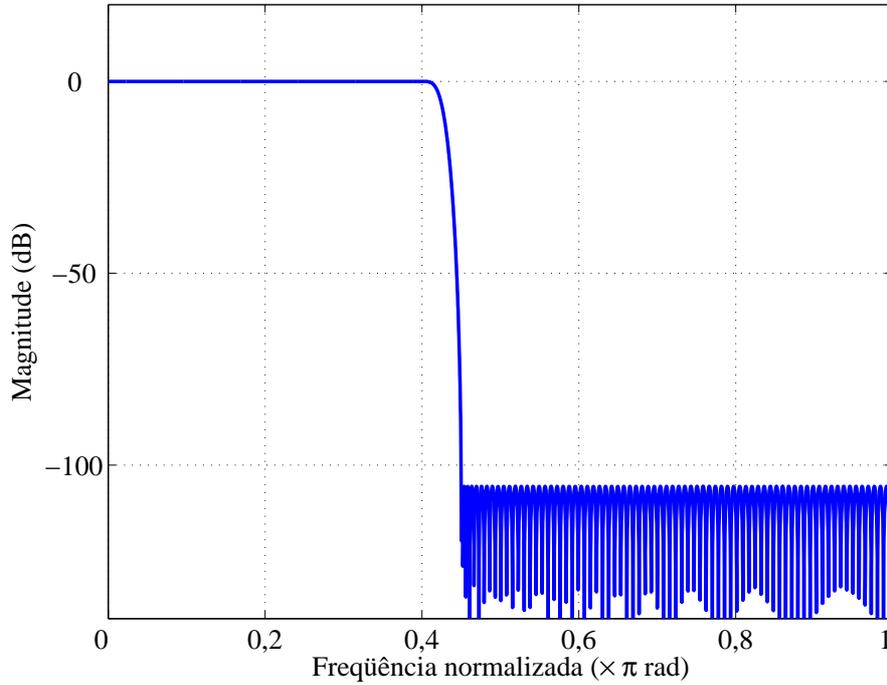


Figura 4.4: Resposta em magnitude do filtro $h[n]$ na frequência.

em amostras que fosse o mais próximo possível da taxa de amostragem de 500 Hz especificada para os parâmetros das trilhas (Seção 4.4). Foi utilizado para a r -ésima sub-banda um salto em amostras de

$$H^r = \left\lfloor \frac{1}{500} \times 2^{-(r-1)} F_s \right\rfloor, \quad (4.4)$$

onde F_s é a taxa de amostragem do sinal original, $2^{-(r-1)} F_s$ é a taxa de amostragem da sub-banda r e $\lfloor b \rfloor$ denota o maior inteiro menor que ou igual a b . Desta forma, o salto utilizado em cada sub-banda é sempre menor que os 2 ms especificados. Uma desvantagem deste esquema é a geração de trilhas senoidais com parâmetros com taxas ligeiramente diferentes para cada sub-banda. Na Seção 4.6.7, é apresentada uma solução para este problema.

Uma vez definido o salto, o espectro é obtido através da DFT de comprimento $Z_p N$ (N em amostras) de cada bloco, onde Z_p é o fator de *zero padding*. Na Seção 4.6.4, são definidos os valores escolhidos para Z_p . O comprimento da janela N é escolhido diferentemente para cada sub-banda e para cada sinal. Em geral, a duração da janela em segundos é maior para as sub-bandas com menor taxa de amostragem.

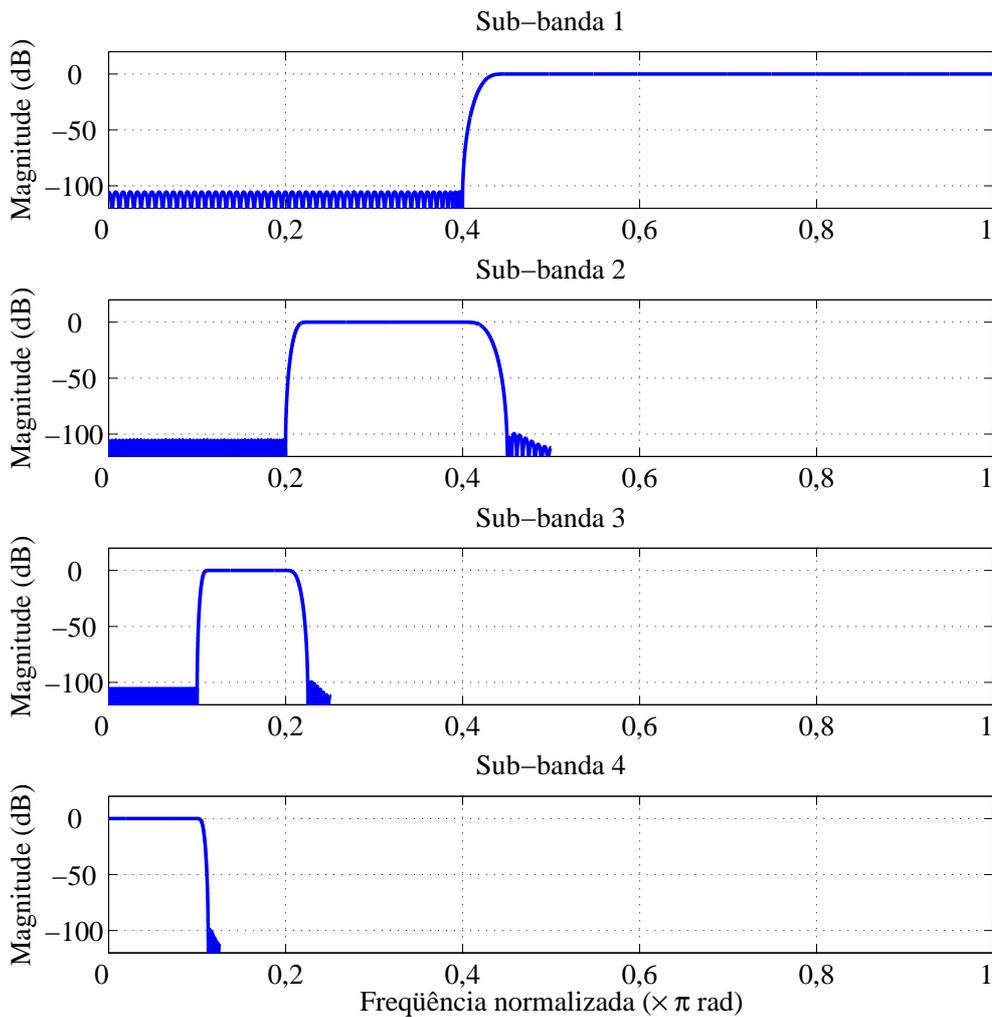


Figura 4.5: Resposta em magnitude na frequência para 4 sub-bandas obtidas a partir da estrutura da Figura 4.3. As sub-bandas estão sendo exibidas na taxa de amostragem do sinal de entrada.

4.6.3 Detecção de Picos

Empregou-se para se construir o BTS um esquema de detecção de picos que utiliza dois limiares: um absoluto e outro local.

O limiar absoluto garante que apenas picos acima de um chão de ruído sejam detectados (conforme diretriz dada na Seção 4.5). Para estimá-lo, seleciona-se inicialmente um trecho de “silêncio” do sinal. Em seguida, o método SSE (ver Seção 2.3.2) é aplicado ao espectro em magnitude médio desse trecho, resultando no limiar ab-

soluto $T_a[k]$.

O limiar local $T_1[m,k]$ é obtido através da aplicação do método SSE ao espectro em magnitude de cada janela do sinal.

Os picos senoidais, então, são obtidos através da seleção dos pontos do espectro de magnitude $S[m,k]$ do sinal que satisfazem

$$\begin{aligned}
 S[m,k] &> S[m,k-1] \\
 S[m,k] &> S[m,k+1] \\
 S[m,k] &> 3T_a[k] \\
 S[m,k] &> d[m,k]T_1[m,k].
 \end{aligned} \tag{4.5}$$

As duas primeiras condições garantem que os pontos selecionados são picos espectrais. A terceira condição restringe a seleção a picos que estejam quase sempre acima do chão de ruído do sinal (estimado através do limiar absoluto), enquanto que a quarta condição obriga que a magnitude do ponto selecionado esteja acima de $d[m,k]T_1[m,k]$. O multiplicador $d[m,k]$ permite o ajuste manual do limiar local para diferentes quadros e diferentes raias de DFT.

Pode-se optar por limitar o número de picos que satisfazem os critérios de (4.5). Desta forma, são selecionados no quadro m apenas os $\Psi[m] \in \mathbb{N}$ picos com maior amplitude. Por convenção, caso $\Psi[m] = 0$ o número de picos selecionados no quadro m não é artificialmente limitado.

4.6.4 Refinamento da Freqüência

Para se gerar o BTS, foi escolhido um estimador de freqüência baseado na interpolação espectral através de um polinômio de segundo grau [125, 126]. Este estimador foi escolhido por permitir o controle do erro de estimação a partir do fator de *zero padding*. Utilizando os resultados relatados em [125], é possível escolher o *zero padding* de maneira que o erro de estimação de freqüência sempre fique abaixo do especificado pela equação (4.2). O código do estimador utilizado foi obtido em [127]; já o código que fornece o menor Z_p necessário para se limitar superiormente o erro de estimação foi obtido em [128]. Este código recebe como argumentos de entrada o tipo da janela (no caso, ‘hanning’), o comprimento da janela N em segundos e os maiores erros relativos tolerados para a freqüência e para a amplitude. O maior

erro tolerado para a amplitude é o mesmo para todas as sub-bandas, escolhido como 0,2; já o valor do maior erro tolerado para a frequência é escolhido diferentemente para cada sub-banda, de acordo com o menor valor assumido pela curva definida por (4.2) nas frequências contidas na sub-banda em questão. Este esquema permite que fatores elevados de Z_p sejam utilizados apenas quando necessário.

4.6.5 Refinamento da Amplitude e da Fase

Uma vez estimada a frequência instantânea, os valores de amplitude e de fase de cada pico são atualizados de acordo com a equação (2.21). O processo de divisão em sub-bandas pode distorcer a amplitude e a fase estimadas da senóide, principalmente se a frequência da senóide for próxima da região de transição do filtro $h[n]$. Para corrigir esta possível distorção causada pelo filtro, a amplitude complexa $\hat{A}_{m,p}$ estimada para o p -ésimo pico no quadro m é modificada de acordo com

$$\hat{A}_{m,p}^{\text{mod}} = \frac{\hat{A}_{m,p}}{\overline{H}(e^{j\hat{\Omega}_{m,p}})}, \quad (4.6)$$

onde $\hat{\Omega}_{m,p}$ é a frequência estimada para o pico e $\overline{H}(\cdot)$ é a resposta em frequência complexa do filtro $h[n]$ quando decimado por 2. Esta correção na amplitude complexa não é realizada para a primeira sub-banda do sinal. Deve-se ressaltar que a equação (4.6) corrige apenas o efeito provocado pelo último filtro utilizado para se obter a sub-banda em questão; ela não corrige a distorção global causada por todos os filtros utilizados para decompor o sinal até aquele nível. Esta aproximação foi considerada satisfatória, uma vez que o filtro $h[n]$ possui uma resposta razoavelmente plana na banda de passagem.

4.6.6 Rastreamento

Foi utilizado como rastreador de trilhas uma versão do algoritmo MQ (ver Seção 3.2). A seleção do melhor pico para continuação da trilha é feita pela minimização da função-custo da equação (3.41). Foram ajustados manualmente para cada sinal-fonte e sub-banda os seguintes parâmetros:

- q – que controla o peso dado entre a amplitude e a frequência ao se escolher a melhor continuação para uma trilha (ver Seção 3.5.1);

- Δ_f – que define a região de busca na frequência por candidatos a continuação de uma determinada trilha (ver Seção 3.2);
- E – que define a duração mínima que uma trilha deve possuir (ver Seção 3.2).

Escolheu-se como 4 o parâmetro S (ver Seção 3.2), que indica por quantos quadros uma trilha deve buscar uma continuação antes de ser considerada extinta. Esta escolha foi feita de modo a atender as especificações definidas na Seção 4.5. Quando uma continuação não é encontrada para a trilha, uma amplitude igual a ‘-1’ lhe é atribuída. Após o rastreamento de todas as trilhas, o algoritmo de interpolação descrito na próxima seção é utilizado para definir os pontos faltantes.

4.6.7 Pós-processamento

Após serem obtidas, as trilhas de cada sub-banda podem ser submetidas às modificações descritas abaixo.

1. Interpolação dos parâmetros faltantes de cada trilha.
2. Seleção manual das trilhas por inspeção visual (através da comparação das trilhas obtidas com o espectrograma do sinal sob análise) e audição do sinal sintetizado. Assim, são removidas as trilhas que não corresponderem às diretrizes da Seção 4.5.1.
3. Correção manual de trilhas que foram erroneamente concatenadas através de uma inspeção visual do espectrograma e das trilhas obtidas. Caso um erro seja observado, as trilhas são separadas e novas concatenações (corretas) são feitas manualmente.
4. Remoção de trilhas duplicadas em diferentes sub-bandas, sendo sempre removida a trilha duplicada extraída na sub-banda com maior frequência de amostragem.
5. Redução na taxa de amostragem dos parâmetros das trilhas para 500 Hz.
6. Junção das trilhas nas diferentes sub-bandas numa única estrutura de dados.

Deve-se ressaltar que as etapas de 2 até 4 são altamente dependentes da intervenção humana, já que utilizam inspeções visuais e auditivas. Abaixo são descritos os algoritmos utilizados nas etapas 1 e 5.

Interpolação de Dados Desconhecidos

A interpolação de parâmetros faltantes é necessária para “completar” as trajetórias temporais de amplitude, frequência e fase de uma determinada trilha quando há falta de dados. Isto ocorre toda vez que o algoritmo de rastreamento de trilhas não associa um pico a uma trilha num quadro. A interpolação [3] utiliza o modelo AR da trajetória do parâmetro em questão. A formulação do método será apresentada para a trajetória da frequência, mas pode ser diretamente replicada para as trajetórias de amplitude e de fase. Para o vetor $\bar{\mathbf{f}}_i$ contendo M amostras da trajetória frequencial da i -ésima trilha, o seu modelo AR (ver equação (2.3)) de ordem q pode ser escrito em forma vetorial como

$$\mathbf{e} = \mathbf{P}\bar{\mathbf{f}}_i, \quad (4.7)$$

onde \mathbf{e} é um vetor de erro de comprimento $(M - q)$ e \mathbf{P} é uma matriz $(M - q) \times M$ contendo os coeficientes do modelo e definida como

$$\mathbf{P} = \begin{bmatrix} -a_q & \cdots & -a_1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & -a_q & \cdots & -a_1 & 1 \end{bmatrix}. \quad (4.8)$$

Considerando que existem $l \ll (M - l)$ amostras faltantes no vetor $\bar{\mathbf{f}}_i$ a partir do instante m_0 , pode-se dividir $\bar{\mathbf{f}}_i$ em dois vetores: $\bar{\mathbf{f}}_i^k$ contendo as $M - l$ amostras conhecidas e $\bar{\mathbf{f}}_i^u$ contendo as l amostras desconhecidas. Desta forma, através de uma partição adequada da matriz \mathbf{P} , pode-se escrever

$$\mathbf{e} = \mathbf{P}\bar{\mathbf{f}}_i = \mathbf{P}^k\bar{\mathbf{f}}_i^k + \mathbf{P}^u\bar{\mathbf{f}}_i^u, \quad (4.9)$$

onde \mathbf{P}^k e \mathbf{P}^u são as partições de \mathbf{P} correspondentes a $\bar{\mathbf{f}}_i^k$ e $\bar{\mathbf{f}}_i^u$, respectivamente. O vetor $\bar{\mathbf{f}}_i^u$ é obtido [3] através da minimização do erro quadrático $\mathbf{e}^T\mathbf{e}$ em relação ao próprio $\bar{\mathbf{f}}_i^u$. Deve-se ressaltar que o método que foi brevemente descrito utiliza toda a informação acumulada na trajetória para estimar os dados desconhecidos.

Para a interpolação da trajetória dos parâmetros de uma trilha, verificou-se que uma ordem $q = 4$ é suficiente para que se obtenham resultados satisfatórios para falhas de até 4 amostras, como demonstrado na Figura 4.6.

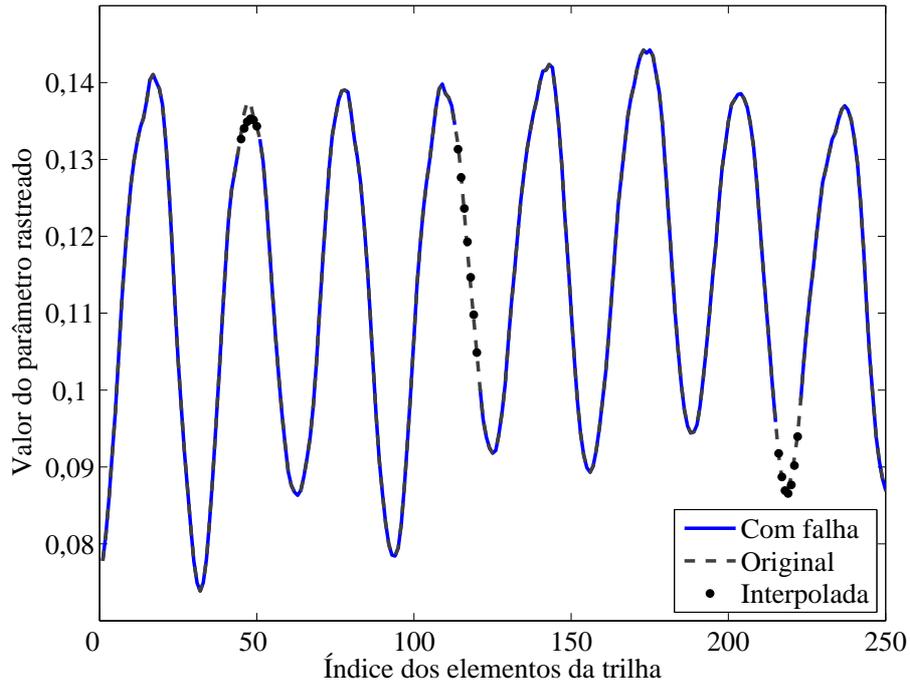


Figura 4.6: Exemplo de interpolação dos parâmetros de uma trilha.

Reamostragem das Trilhas

Considerando que as trajetórias dos parâmetros das trilhas senoidais possuem taxas de amostragem levemente diferentes, dependendo da sub-banda em que foram rastreadas (ver Seção 4.6.2), foi implementado um método simples de reamostragem.

A redução da taxa de amostragem das trilhas é feita considerando-se como linear por partes o sinal contínuo subjacente das trajetórias dos parâmetros. Feita esta aproximação, as trajetórias são reamostradas de modo a se obter amostras espaçadas temporalmente de exatamente 2 ms (frequência de amostragem igual a 500 Hz). Para garantir o sincronismo entre as trilhas, o instante de tempo inicial de cada trilha é arredondado para o múltiplo inteiro mais próximo de 2 ms. Este método simples de mudança de taxa de amostragem foi escolhido porque os parâmetros das trilhas já se encontram sobre-amostrados (ver Seção 4.4). Além disso, nenhuma diferença audível foi percebida entre os sinais sintetizados a partir das trilhas na taxa original e a partir das trilhas na taxa modificada.

4.7 Trilhas Obtidas

Nesta seção, são apresentadas informações sobre como o BTS foi construída utilizando-se os métodos apresentados na seção anterior. Também é descrita a forma como os parâmetros foram ajustados.

Para cada sinal, os parâmetros de análise foram inicializados com um valor padrão que é ajustado posteriormente, durante a análise. Na Tabela 4.7, estão listados os parâmetros iniciais. Deve-se ressaltar que com exceção do número de sub-bandas R , todos os outros parâmetros são inicializados da mesma forma em todas as sub-bandas.

Tabela 4.7: Inicialização dos parâmetros de análise de cada etapa do processamento.

Etapa	Parâmetros
Divisão em Sub-bandas	$R = 4$
STFT ³	$N = 40$ ms
Detecção de Picos	$d[m,k] = 3$ e $T_p[m] = 0$
Refinamentos	–
Rastreamento	$\Delta_f = 0,03$, $E = 30$ ms e $q = 0,5$

O primeiro parâmetro de análise escolhido para cada sinal é R , conforme o procedimento abaixo:

1. O sinal é dividido em quatro sub-bandas.
2. É obtido o espectrograma do sinal na sub-banda com menor taxa de amostragem.
3. Se as trilhas estiverem muito aglomeradas, pode-se tomar a decisão de elevar R .

Uma vez escolhido R , os demais parâmetros de análise são ajustados para cada sub-banda. A sub-banda com menor taxa de amostragem é analisada primeiro, e

³Deve-se ressaltar que, uma vez escolhido N , o parâmetro Z_p é obtido através do método descrito em [125].

assim por diante. A duração da janela de análise N é definida observando-se o espectro do sinal. Para ajustá-la, busca-se um compromisso entre as resoluções temporal e freqüencial do espectrograma obtido. Uma vez definida a janela, é necessário ajustar o limiar local de detecção de picos através do parâmetro $d[m,k]$. Usualmente, este parâmetro é modificado de acordo com mudanças abruptas no sinal, tal como o *onset* de uma nota musical. Nestas mudanças, o limiar pode ser elevado para evitar a detecção de picos provocados pela parcela não-tonal do sinal. Se o espectro do sinal exibe picos claramente provocados pelo lóbulos laterais da janela de análise, após o ajuste de $T_l[m]$, $\Psi[m]$ (ver equação (4.5)) pode ser ajustado para que esses picos não sejam detectados. A escolha de ajustar $\Psi[m]$ é de cunho puramente prático, para permitir um ajuste mais fácil.

Para a etapa de rastreamento, os parâmetros são escolhidos de modo que as trilhas obtidas se aproximem ao máximo do que é esperado pela observação do espectrograma. Na prática, os parâmetros foram escolhidos de maneira permissiva, de modo a garantir que todas as trilhas da parcela tonal do sinal fossem corretamente rastreadas, mesmo que isso provocasse a criação de trilhas associadas à parcela não tonal do sinal, já que estas trilhas indesejáveis podem facilmente ser removidas na etapa de pós-processamento.

Na Tabela 4.8 de (a) até (d), podem ser vistos para cada sinal do BTS o parâmetro R , as sub-bandas das quais foram extraídas trilhas e o total de trilhas extraídas. Pode-se observar que apenas em dois sinais foi utilizado um R superior a 4. Além disso, pode-se observar que para alguns sinais, nenhuma trilha foi obtida para a sub-banda com maior taxa de amostragem. Isto acontece devido à característica “passa-baixas” dos sinais de áudio, o que faz com que a energia da parcela tonal dos sinais em altas-freqüências seja muito baixa, deixando os picos provocados por ela abaixo do chão de ruído e, por conseguinte, abaixo do limiar absoluto $T_a[k]$.

Para cada sinal, foram documentados em uma rotina do Matlab[®] todos os ajustes de parâmetros feitos no procedimento de análise. Desta forma, todo o processo de obtenção do BTS pode ser facilmente refeito e a metodologia utilizada aferida.

Tabela 4.8: Informações sobre a análise dos sinais de cada nível do BTS.

(a) Nível 0

Sinal	R	Sub-bandas com trilhas	Número de Trilhas
<i>vibrato</i>	4	3 e 4	26
<i>tremolo</i>	4	3 e 4	26
frase musical	4	2, 3 e 4	58

(b) Nível 1

Sinal	R	Sub-bandas com trilhas	Número de Trilhas
Violino	4	2, 3 e 4	60
Piano	4	1, 2 3 e 4	82
Saxofone	4	1, 2 3 e 4	212
Acordeão	4	1, 2 3 e 4	96

(c) Nível 2

Sinal	R	Sub-bandas com trilhas	Número de Trilhas
Órgão	5	1, 2, 3, 4 e 5	576
Violino	4	1, 2, 3 e 4	654
Piano	4	2, 3 e 4	1040
Cello	4	2, 3 e 4	2180

(d) Nível 3

Sinal	R	Sub-bandas com trilhas	Número de Trilhas
Quarteto	5	3, 4 e 5	330

4.8 Validação do BTS

Nesta seção, os sinais sintéticos são utilizados para aferir se o método de análise descrito na Seção 4.6 respeita as especificações feitas na Seção 4.5. Em particular, são testadas três especificações: a do erro máximo para estimação de frequência, a do erro máximo para a estimação da amplitude e a do erro máximo para a detecção do início e do fim de uma trilha senoidal. O objetivo desta seção não é fazer uma validação extensiva dos métodos de análise utilizados, mas sim verificar se para o caso simples dos sinais sintéticos as especificações feitas conseguem ser cumpridas.

Inicialmente, é apresentado um método para se obter trilhas senoidais a partir dos sinais de controle utilizados no sintetizador. Estas trilhas, então, são consideradas como referência nos testes realizados em seguida.

4.8.1 Obtenção das Trilhas de Referência

As trilhas de referência utilizadas na validação do BTS foram obtidas do sintetizador empregado para se gerar os sinais do Nível 0 (ver Seção 4.3.1). Estes sinais controlam a amplitude e a frequência dos osciladores associados a cada parcial gerada pelo sintetizador e estão na taxa do sinal sintetizado (44,1 kHz).

A conversão dos sinais de controle em trilhas senoidais é direta: apenas uma conversão da taxa de amostragem se faz necessária. Foi utilizado na conversão de taxa o algoritmo descrito na Seção 4.6.7 para se ajustar a taxa de amostragem das trilhas.

Após a conversão da taxa, o início e fim de cada trilha é determinado. O ponto inicial (final) de cada trilha é determinado pelo instante em que a sua trajetória de amplitude fica acima (abaixo) de um limiar pré-estabelecido. No caso dos sinais de *vibrato* e de *tremolo*, o limiar é determinado pelo nível do ruído de quantização infligido pela representação digital com um número finito de bits. Como ambos os sinais são quantizados em 16 bits, o ruído de quantização possui um nível máximo de aproximadamente -96 dB (*full scale*). Já para o sinal composto por diversas notas musicais, o limiar depende da frequência da trilha em questão, e é determinado pela densidade espectral de potência do ruído, que foi estimada através de um

periodograma⁴ [58].

4.8.2 Estimação de Frequência

Para aferir se o erro de estimação dos valores de frequência é inferior ao especificado na Seção 4.5.2, é realizado um teste utilizando o sinal sintético de *vibrato*. Compara-se ponto a ponto a trajetória da frequência obtida no BTS com a trajetória da frequência utilizada para sintetizar o sinal (ver seção anterior). Sejam $\hat{f}_{i,m}$ a frequência estimada para a i -ésima trilha no quadro m e $f_{i,m}$ a frequência correspondente utilizada para sintetizar o sinal; o erro relativo da estimação é obtido como

$$\frac{|\hat{f}_{i,m} - f_{i,m}|}{f_{i,m}}. \quad (4.10)$$

É calculado, então, em quanto o erro relativo supera o limite máximo estabelecido, $e^{freq}(f_{i,m})$, para a frequência $\hat{f}_{i,m}$ (equação (4.2)). Assim, adotou-se a figura de mérito

$$\Gamma(f_{i,m}, \hat{f}_{i,m}) = \frac{|\hat{f}_{i,m} - f_{i,m}|}{f_{i,m}} - e^{freq}(f_{i,m}), \quad (4.11)$$

que se faz necessária porque o erro relativo máximo especificado é uma função da frequência.

Na Figura 4.7, pode ser vista a distribuição da figura de mérito utilizada para todos os pontos das trajetórias em frequência das trilhas obtidas para o sinal sintético exibindo *vibrato*. Como pode-se observar, apenas valores negativos foram obtidos, indicando que para o sinal sob teste o erro cometido na estimação da frequência ficou abaixo do erro máximo especificado. Com isso, as especificações são atendidas para este sinal, indicando que o método de estimação escolhido parece ser adequado.

4.8.3 Estimação de Amplitude

A aferição dos valores estimados para amplitude pode ser feita de maneira mais direta. Foram comparados os valores estimados para a trajetória da amplitude do sinal sintético contendo *tremolo* com os valores utilizados para sintetizar o sinal. Sendo $\hat{A}_{i,m}$ o valor estimado de amplitude para a i -ésima trilha no quadro m e $A_{i,m}$

⁴Qualquer estimador para densidade espectral de potência poderia ser utilizado para esta tarefa.

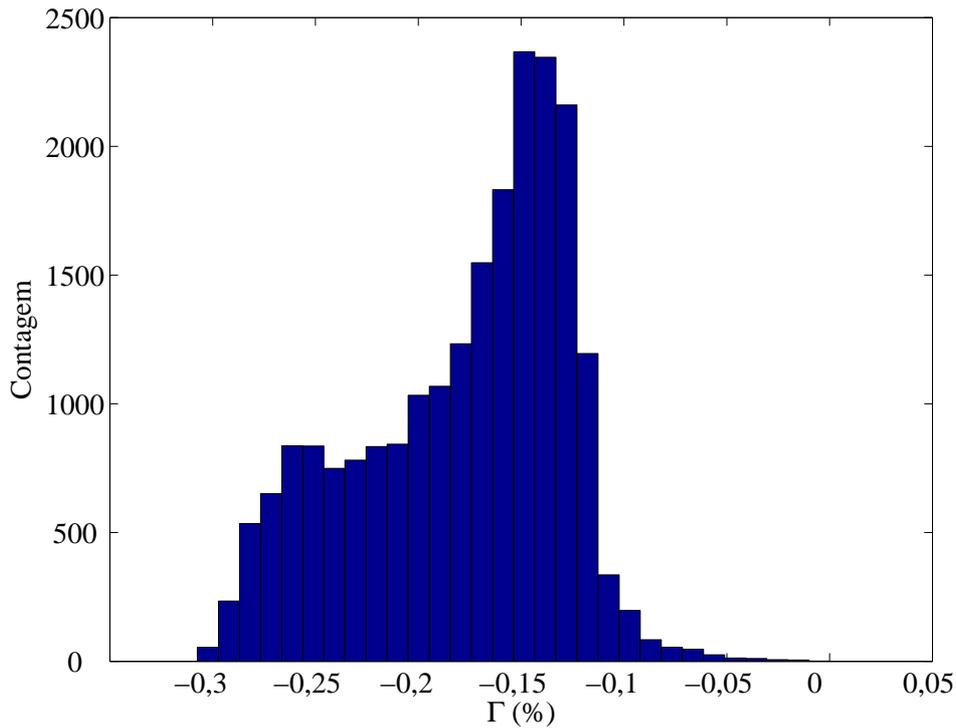


Figura 4.7: Distribuição da figura de mérito $\Gamma(f_{i,m}, \hat{f}_{i,m})$ em porcentos para as trilhas do sinal sintético exibindo *vibrato*. Como só há valores negativos, pode-se afirmar que os erros cometidos ficaram abaixo do erro máximo especificado.

o valor correspondente utilizado para sintetizar o sinal, o erro relativo de amplitude é obtido através de

$$\frac{|\hat{A}_{i,m} - A_{i,m}|}{A_{i,m}}. \quad (4.12)$$

Na Figura 4.8, é exibido o histograma do erro relativo de amplitude referente ao conjunto de todas as medidas de amplitude sobre as trilhas identificadas para o sinal com *tremolo*. Lembrando que o erro relativo máximo especificado é de 20 %, pode-se observar que o erro dos valores estimados se situa abaixo dele. O erro relativo médio observado foi de 2,8 %, portanto, bem menor que o limite máximo. Logo, o estimador de amplitude adotado atende às necessidades do problema.

4.8.4 Detecção de *Onset* e *Offset*

O erro de estimação dos valores de início e de fim das trilhas senoidais foi aferido utilizando-se o sinal sintético de uma frase musical. Para isso, foi calculada a diferença absoluta entre o instante inicial das trilhas e o instante inicial das trilhas de

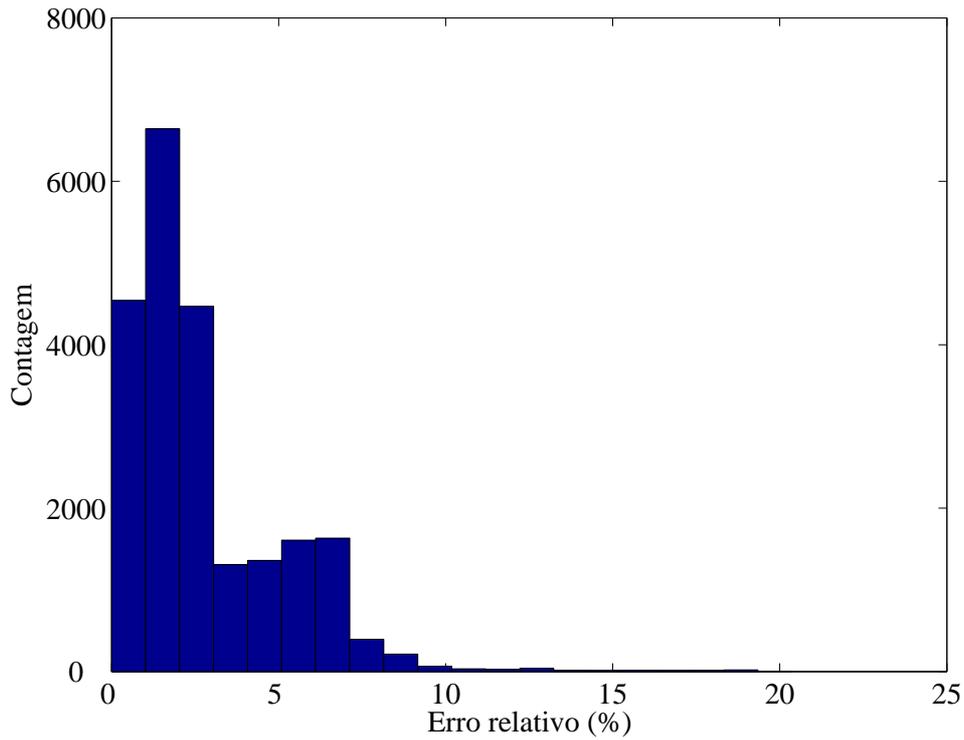


Figura 4.8: Distribuição do erro relativo da estimação da amplitude para o sinal sintético exibindo *tremolo*. Deve-se observar que nenhum valor foi estimado com erro acima de 20 %, que é o erro máximo especificado.

referência. O mesmo foi feito para os instantes finais da trilha.

Na Figura 4.9, vê-se o histograma das medidas de erro absoluto, referentes a instantes de *onset* e *offset*, computadas para todas as trilhas identificadas no sinal. O maior erro encontrado foi de 8 ms, que fica abaixo da tolerância de 20 ms especificada na Seção 4.5.1. Foi calculado um erro médio de 1,2 ms, o que permite afirmar que os tempos de início e fim das trilhas obtidas para o sinal sob teste estão abaixo da tolerância especificada.

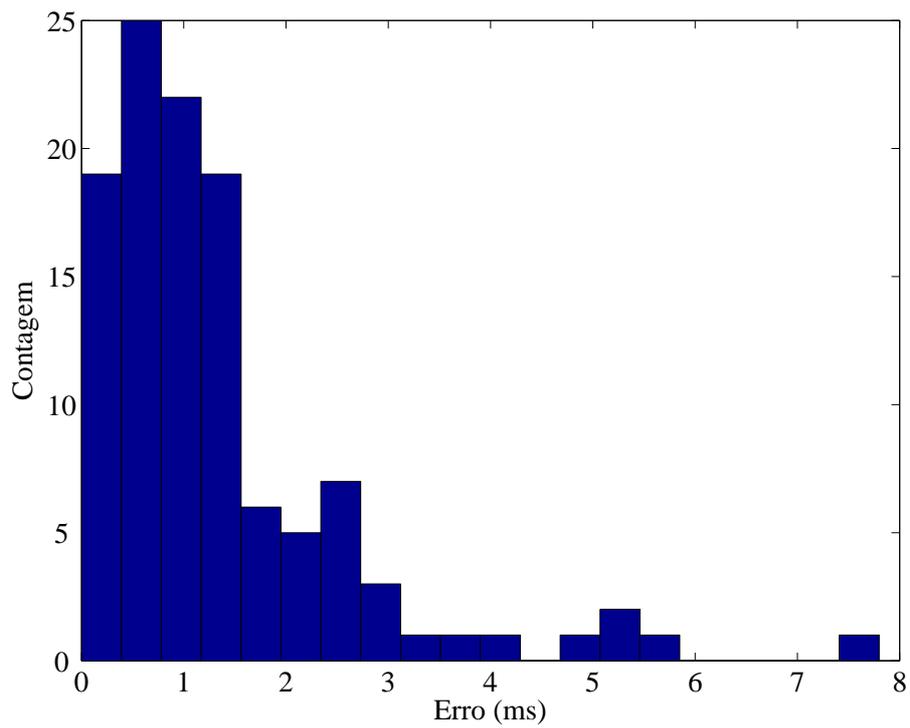


Figura 4.9: Distribuição do erro absoluto de estimação de início e fim das trilhas senoidais para o sinal sintético de uma frase musical. Deve-se observar que o maior erro de estimação cometido ficou abaixo da especificação do maior erro admissível (20 ms).

Capítulo 5

Avaliação de Desempenho de Algoritmos de Análise Senoidal

5.1 Introdução

Neste capítulo, são apresentados três casos de estudo que ilustram de que modo o BTS, apresentado no capítulo anterior, e os avaliadores, apresentados nas Seções 2.6 e 3.6, podem ser utilizados para medir o desempenho de diferentes etapas de um sistema de análise senoidal. Os testes comparam alguns dos métodos para análise senoidal descritos nos capítulos anteriores. Entretanto, uma avaliação extensiva destes métodos não é o foco deste capítulo. Os testes foram realizados utilizando o sistema de análise senoidal FlexSM [P1].

Na Seção 5.2, é descrito um teste que utiliza o avaliador generalizado de detectores e estimadores de picos espectrais, apresentado na Seção 2.6, para medir o desempenho de diferentes métodos de refinamento de frequência. Na Seção 5.3, é apresentado um teste em que rastreadores adaptativos de trilhas são avaliados utilizando-se o método descrito na Seção 5.3. Por fim, na Seção 5.4 é feita a comparação de diferentes sistemas de análise senoidal. Em todos os testes, os sinais e as trilhas de referência são fornecidos pelo BTS descrito no capítulo anterior.

A descrição de cada teste seguirá a estrutura abaixo:

1. Escolha dos sinais de entrada;
2. Configuração dos sub-sistemas sob teste;

3. Configuração do avaliador;
4. Análise de resultados.

5.2 Teste 1: Avaliação de Métodos de Refinamento da Frequência

Neste teste é demonstrado como o avaliador generalizado de detectores e estimadores de picos espectrais apresentado na Seção 2.6 pode ser utilizado para testar o desempenho de diferentes métodos de refinamento de frequência. São comparados os desempenhos de três estimadores de frequência descritos em [9]. O BTS é utilizado para obter um conjunto de picos espectrais que servem de referência para o avaliador e fornecer os sinais de áudio utilizados no teste.

São utilizadas no teste as duas figuras de mérito fornecidas pelo avaliador descrito no Capítulo 2: EA e ED. A primeira é a de principal interesse para o teste a ser realizado, já que mede diretamente o erro de estimação da frequência dos picos. Já a segunda, associada ao número de falsos alarmes e falsos positivos, é utilizada de forma auxiliar, conforme descrito na próxima seção.

Inicialmente, na Seção 5.2.1, são apresentados os sinais escolhidos para os testes. Na Seção 5.2.2, é feita uma descrição do teste realizado. Os resultados são discutidos na Seção 5.2.3.

5.2.1 Sinais de Teste

Para se realizar o teste 1, foi escolhido apenas um sinal de cada nível do BTS (ver Seção 4.3). Os sinais selecionados são:

- Nível 0 – sinal 3 (ver Seção 4.3.1).
- Nível 1 – acordeão (ver Tabela 4.1).
- Nível 2 – órgão (ver Tabela 4.3).
- Nível 3 – quarteto (ver Tabela 4.6).

Foram escolhidos sinais com um número elevado de trilhas de modo a permitir que um número elevado de picos pudessem ser potencialmente detectados.

5.2.2 Descrição do Teste

Em termos gerais, o teste é composto por duas etapas:

1. Submete-se um dado sinal de teste a um método de detecção/estimação de picos;
2. Compara-se o conjunto de picos estimados com o de referência.

A seguir, serão detalhadas as escolhas realizadas em cada uma destas etapas.

Método de Detecção/Estimação de Picos

Neste teste, será utilizado o algoritmo de detecção de picos proposto em [P2] e descrito na Seção 2.3. Podem ser vistos na Tabela 5.1 os valores dos parâmetros da STFT utilizados.

Tabela 5.1: Valores dos parâmetros da STFT utilizados no Teste 1.

Parâmetro	Valor
Janela	Hann com duração de 40 ms
Salto entre janelas adjacentes	10 ms
Comprimento da DFT	2048 pontos

Para estimar a tendência espectral do sinal sob teste, foi utilizado o método SSE (ver Seção 2.3.2) com um filtro de média móvel de ordem 50.

O estimador SSE (agora com um filtro média móvel de ordem 150) também foi utilizado para obter o nível médio do chão de ruído em cada quadro de análise do sinal. O fator multiplicativo d , que determina o posicionamento do limiar de detecção, foi selecionado pelo procedimento descrito a seguir.

Para cada sinal de teste, o parâmetro d foi escolhido de modo a minimizar o erro do tipo ED (ver Seção 2.6). Esse esquema procura garantir um número elevado de picos corretamente detectados e reduzir o número de picos de ruído, que podem influenciar negativamente o resultado do teste. Assim, é possível se obter de maneira mais precisa a figura de mérito EA. Variou-se d de 1 até 10 em passos de 0,5, escolhendo-se o valor de d que fornecesse o menor valor de ED. Como exemplo, a

Figura 5.1 exibe ED para diferentes valores de d sobre o sinal de *acordeão* do Nível 1 do BTS: o melhor desempenho foi obtido para $d = 3$, tendo-se alcançado ED=0,13.

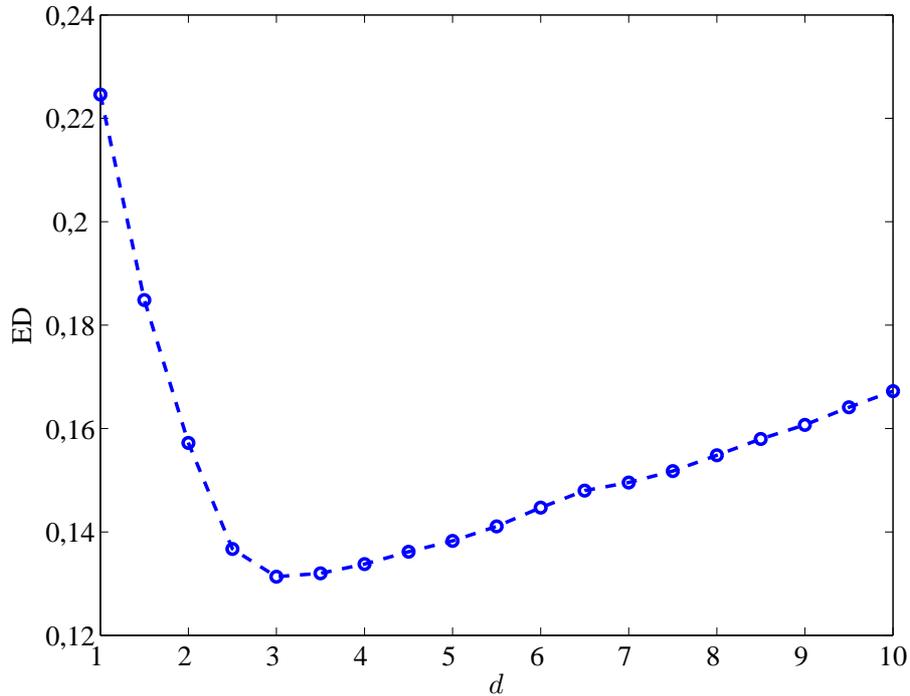


Figura 5.1: Valores de ED obtidos por um algoritmo de detecção de picos para diferentes valores de d .

Uma vez escolhido um valor adequado para d , os picos espectrais do sinal sob teste foram detectados e a frequência instantânea de cada pico foi refinada utilizando-se os seguintes algoritmos [9]: reatribuição de frequência, DFT1 e diferença de fases.

Ajuste do Avaliador

Para realizar a avaliação dos métodos de refinamento é necessário um conjunto de picos de referência para cada sinal. No entanto, o BTS não fornece um conjunto de picos de referência, mas trilhas. Logo, é necessário converter as trilhas fornecidas pelo BTS num conjunto de picos de referência. Para um dado instante t referente ao centro de uma janela de análise, os picos de referência são obtidos por amostragem dos parâmetros das trilhas no instante de tempo em questão.

Por fim, é necessário ajustar o parâmetro Λ_f do avaliador (Seção 2.6). Para este teste, foi escolhido, arbitrariamente, um $\Lambda_f = 6 \%$, o que equivale a um intervalo de um semitom e foi selecionado de maneira similar à escolha do parâmetro Δ_f dos rastreadores.

5.2.3 Resultados

A Tabela 5.2 resume os resultados de EA para os quatro sinais sob teste. Na tabela, a coluna cujo nome é NE exibe os resultados obtidos quando nenhum método de refinamento de estimativa freqüencial é utilizado.

Observando-se a tabela, é possível notar que EA cresce para os sinais dos níveis mais elevados do BTS, que apresentam maior dificuldade para a análise. De forma geral, os três métodos de refinamento de freqüência testados foram capazes de reduzir EA. Para os sinais do Nível 0 e 1, o método DFT1 obteve um resultado pior que os obtidos com os outros dois métodos de refinamento. Já para os outros dois níveis, não é possível afirmar que um determinado método de refinamento obteve um resultado notavelmente superior aos dos demais, resultado este que está de acordo com o relatado em [9]. Deve-se ressaltar a grande redução no valor de EA para o sinal do Nível 2 do BTS quando qualquer um dos métodos de refinamento é utilizado. Este sinal exibe uma grande quantidade de picos com elevada energia em baixa freqüência que são de difícil detecção. Além disso, foi obtida uma pequena redução de EA para o sinal do Nível 3, indicando que os métodos de refinamento utilizados podem não ser robustos à polifonia.

Tabela 5.2: Valores obtidos para EA (em %) para os estimadores de freqüência instantânea utilizados. NE = ‘Nenhum Estimador’, RF = ‘Reatribuição de Freqüência’, DFT1 = ‘DFT da Derivada do Sinal’ e DF = ‘Diferença de Fases’.

Nível \ Método	EA (%)			
	NE	RF	DFT1	DF
0	$2,22 \times 10^{-1}$	$3,67 \times 10^{-2}$	$1,15 \times 10^{-1}$	$3,67 \times 10^{-2}$
1	$2,33 \times 10^{-1}$	$9,94 \times 10^{-2}$	$1,07 \times 10^{-1}$	$9,95 \times 10^{-2}$
2	$1,10 \times 10^0$	$2,71 \times 10^{-1}$	$2,74 \times 10^{-1}$	$2,71 \times 10^{-1}$
3	$1,21 \times 10^0$	$9,21 \times 10^{-1}$	$9,19 \times 10^{-1}$	$9,21 \times 10^{-1}$

5.3 Teste 2: Avaliação de Algoritmos de Rastreamento de Senóides

Neste teste, demonstra-se de que forma os dois avaliadores propostos na Seção 3.6 podem ser utilizados, em conjunto com o BTS apresentado no Capítulo 4, para medir a eficácia de rastreadores de trilhas senoidais. Serão comparados aqueles que fazem uso dos preditores adaptativos apresentados na Seção 3.4, do preditor apresentado em [94] e o (de ordem-zero) do algoritmo MQ (ver Seção 3.2). Um segundo objetivo do teste é comparar as figuras de mérito fornecidas pelos avaliadores descritos na Seção 3.6.

Na Seção 5.3.1 são descritos os sinais escolhidos para o teste. Na Seção 5.3.2, descreve-se o teste realizado. Os resultados são apresentados na Seção 5.3.3.

5.3.1 Sinais Escolhidos

Para se realizar este teste, foi escolhido apenas um sinal de cada do nível do BTS (ver Seção 4.3):

- Nível 0 – sinal 1 (ver Seção 4.3.1).
- Nível 1 – violino (ver Tabela 4.1).
- Nível 2 – violino (ver Tabela 4.3).
- Nível 3 – quarteto (ver Tabela 4.6).

Procurou-se escolher sinais que exibissem variações amplas ao longo do tempo nos parâmetros das trilhas.

5.3.2 Descrição do Teste

Tratamento dos Dados de Entrada

Para se realizar o teste, as trajetórias das trilhas dos sinais sob teste são “amostradas” no instante de tempo mais próximo ao do centro da janela de análise. Em seguida, é criado um conjunto de picos senoidais a partir das amostras das trajetórias de frequência e amplitude de cada trilha da seguinte forma:

1. Seleciona-se a i -ésima trilha de referência. Supondo que a trilha começa num quadro m_b e termina num quadro m_e , convertem-se as suas trajetórias de amplitude e de frequência em $m_e - m_b$ picos senoidais. Para tanto, cada pico $p_{i,m}$ é definido como a i -ésima trilha no m -ésimo quadro ou, equivalentemente, $p_{i,m} = \{f_i[m], A_i[m]\}$ para $m_b \leq m \leq m_e$, onde f_i e A_i são, respectivamente, as trajetórias de frequência e amplitude da trilha em questão.
2. Para cada quadro m_k , percorrem-se todas as trilhas e agrupam-se os picos correspondentes, formando o conjunto \mathcal{P}_{m_k} . Matematicamente, $\mathcal{P}_{m_k} = \bigcup_{\forall i} p_{i,m_k}$.

Os conjuntos \mathcal{P}_m , para cada quadro m , são fornecidos como entrada aos diferentes rastreadores sob teste.

Configuração dos Rastreadores sob Teste

Com exceção do algoritmo MQ, os rastreadores empregados diferem apenas no preditor empregado. Todos os rastreadores utilizam a função de decisão da equação (3.41) e a heurística definida pela equação (3.45). Foram utilizados os seguintes preditores:

- Burg – descrito em [94].
- RLS conjunto – descrito na Seção 3.4.1.
- *Lattice* – descrito na Seção 3.4.2.

Os parâmetros relacionados ao algoritmo MQ e à função de decisão são comparilhados pelos rastreadores sob teste e foram ajustados de maneira a fornecer um desempenho satisfatório (avaliado através de inspeção visual das trilhas obtidas) para o algoritmo MQ. Na Tabela 5.3, podem ser vistos os valores destes parâmetros utilizados para cada sinal. Os parâmetros específicos de cada preditor foram ajustados empiricamente para cada sinal de teste de maneira similar ao esquema utilizado para se ajustar os parâmetros do algoritmo MQ. No Apêndice A, podem ser vistos os valores escolhidos para cada parâmetro dos preditores utilizados neste teste.

Avaliação de Desempenho

A avaliação é feita comparando-se, através dos avaliadores descritos na Seção 3.6, as trilhas obtidas pelos diferentes rastreadores com as trilhas de referência sub-

Tabela 5.3: Parâmetros do método MQ utilizados na comparação dos rastreadores.

Nível \ Parâmetro	S (ms)	E (ms)	Δ_f (%)	q
0	20	50	3,5	0,5
1	30	50	3	0,5
2	40	50	4,5	0,75
3	40	50	3	0,75

amostrados fornecidas pelo BTS.

As figuras de mérito fornecidas pelos avaliadores utilizados são: E , E^{mod} e E^{s} . Como se viu na Seção 3.6, a figura de mérito E fornece uma nota geral para o desempenho do rastreador sob teste. Já as figuras de mérito E^{mod} e E^{s} do avaliador modificado são mutuamente complementares: a primeira calcula erros de estimação cometidos ao longo das trajetórias de frequência e amplitude das trilhas e ainda erros de detecção; já a segunda calcula erros de segmentação/união de trilhas. Considerando o teste que está sendo executado, é esperada uma preponderância de erros do tipo E^{s} , uma vez que não há erros de estimação envolvidos (os picos de entrada são obtidos a partir das próprias trilhas de referência).

Deve-se ressaltar que a figura de mérito E pode ser interpretada isoladamente, enquanto E^{s} e E^{mod} sempre devem ser interpretadas em conjunto.

5.3.3 Resultados

Na Tabela 5.4, podem ser vistos os valores de E fornecidos pelo avaliador proposto em [45], para cada par {sinal, rastreador} considerado. Já nas Tabelas 5.5 e 5.6, são exibidas, respectivamente, as figuras de mérito E^{mod} e E^{s} do avaliador modificado.

Para auxiliar a interpretação dos resultados fornecidos pelos avaliadores, são exibidas as trajetórias de frequência obtidas para alguns sinais utilizados no teste. Nas figuras, as trajetórias de referência são exibidas na cor cinza e em linhas espessas, enquanto as estimadas estão escuras e em linhas finas. Idealmente, deseja-se que as trajetórias estimadas se sobreponham às de referência, logo sendo possível visualizar as trajetórias em azul “sombreadas” pelas trajetórias de referência. As setas apontam para alguns erros de rastreamento que serão comentados ao longo do texto.

Tabela 5.4: Valores obtidos para E .

Sinal \ Método	E			
	MQ	Burg	RLS	Lat.
Nível 0	$4,58 \times 10^{-6}$	$4,28 \times 10^{-6}$	$4,28 \times 10^{-6}$	$4,28 \times 10^{-6}$
Nível 1	$1,94 \times 10^{-3}$	$2,46 \times 10^{-3}$	$2,46 \times 10^{-3}$	$1,36 \times 10^{-3}$
Nível 2	$7,17 \times 10^{-2}$	$5,66 \times 10^{-2}$	$5,25 \times 10^{-2}$	$4,95 \times 10^{-2}$
Nível 3	$1,05 \times 10^{-1}$	$8,94 \times 10^{-2}$	$8,74 \times 10^{-2}$	$8,60 \times 10^{-2}$

Tabela 5.5: Valores obtidos para E^{mod} .

Sinal \ Método	E^{mod}			
	MQ	Burg	RLS	Lat.
Nível 0	$4,58 \times 10^{-6}$	$4,28 \times 10^{-6}$	$4,28 \times 10^{-6}$	$4,28 \times 10^{-6}$
Nível 1	$8,36 \times 10^{-5}$	$1,77 \times 10^{-3}$	$1,77 \times 10^{-3}$	$3,07 \times 10^{-5}$
Nível 2	$2,59 \times 10^{-3}$	$1,16 \times 10^{-3}$	$7,16 \times 10^{-4}$	$1,18 \times 10^{-3}$
Nível 3	$3,73 \times 10^{-3}$	$3,60 \times 10^{-3}$	$2,30 \times 10^{-3}$	$3,38 \times 10^{-3}$

Tabela 5.6: Valores obtidos para E^{s} .

Sinal \ Método	E^{s}			
	MQ	Burg	RLS	Lat.
Nível 0	$0,00 \times 10^0$	$0,00 \times 10^0$	$0,00 \times 10^0$	$0,00 \times 10^0$
Nível 1	$1,02 \times 10^0$	$3,98 \times 10^{-1}$	$3,98 \times 10^{-1}$	$3,90 \times 10^{-1}$
Nível 2	$3,93 \times 10^0$	$1,06 \times 10^0$	$9,97 \times 10^{-1}$	$9,42 \times 10^{-1}$
Nível 3	$4,18 \times 10^0$	$2,46 \times 10^0$	$2,30 \times 10^0$	$2,40 \times 10^0$

Na Figura 5.2, são exibidas as trajetórias de referência juntamente com as trajetórias estimadas para o sinal do Nível 0, quando é utilizado o preditor *lattice*. Pode-se perceber que não há quase nenhuma diferença visível entre as trajetórias de referência e as encontradas, e este fato é refletido nos baixos valores encontrados para as figuras de mérito. Os poucos erros que podem ser observados ocorrem no final das trajetórias, como o apontado pela seta.

Nas Figuras 5.3 (a) e (b) podem ser vistas as trajetórias de frequência para o sinal do Nível 3 para os métodos MQ e *lattice*, respectivamente. Para facilitar a visualização, são exibidas apenas as trajetórias com frequências entre 0 e 1 kHz. De forma geral, as trilhas foram corretamente rastreadas, sendo possível perceber uma melhora para as trilhas obtidas pelo rastreador *lattice*. A figura de mérito E que mede esse desempenho “global” do rastreador é condizente com o que é observado: um decréscimo de aproximadamente 20 % no valor de E atribuído ao rastreador com preditor *lattice*, em comparação com o resultado para o método MQ. Dois tipos de problemas localizados também podem ser observados nas figuras e serão ressaltados. Um problema que ocorreu nas trilhas obtidas pelos dois métodos é a perda de trilhas de curta duração (caso exemplificado pelas trajetórias apontadas pelas setas “B”). Nas trilhas obtidas pelo método MQ, pode-se perceber que algumas trajetórias foram “confundidas”, principalmente as situadas acima de 600 Hz (caso exemplificado pelas trajetórias apontadas pelas setas “A”). As figuras de mérito E^s e E^{mod} são coerentes com esses defeitos. Idealmente, a figura de mérito E^{mod} apontaria o erro causado pelas trilhas perdidas, enquanto que E^s quantificaria o efeito das trajetórias “confundidas”. De fato, E^s apresenta uma redução em aproximadamente 42 % para o método *lattice* quando comparado com o MQ, ao passo que E^{mod} apresenta uma redução percentual muito inferior.

Comparando-se as figuras de mérito para os diferentes métodos, pode-se perceber que as maiores reduções ocorreram para os sinais dos Níveis 2 e 3, que apresentariam maior dificuldade em ter suas trilhas corretamente rastreadas. No Nível 1, observou-se um aumento em E^{mod} quando os métodos Burg e RLS são utilizados, fato este que é compensado pela diminuição de E^s . De forma geral, pode-se verificar que o desempenho dos três preditores testados foi similar, sendo que o método de Burg apresentou um desempenho levemente inferior aos dos preditores adaptativos,

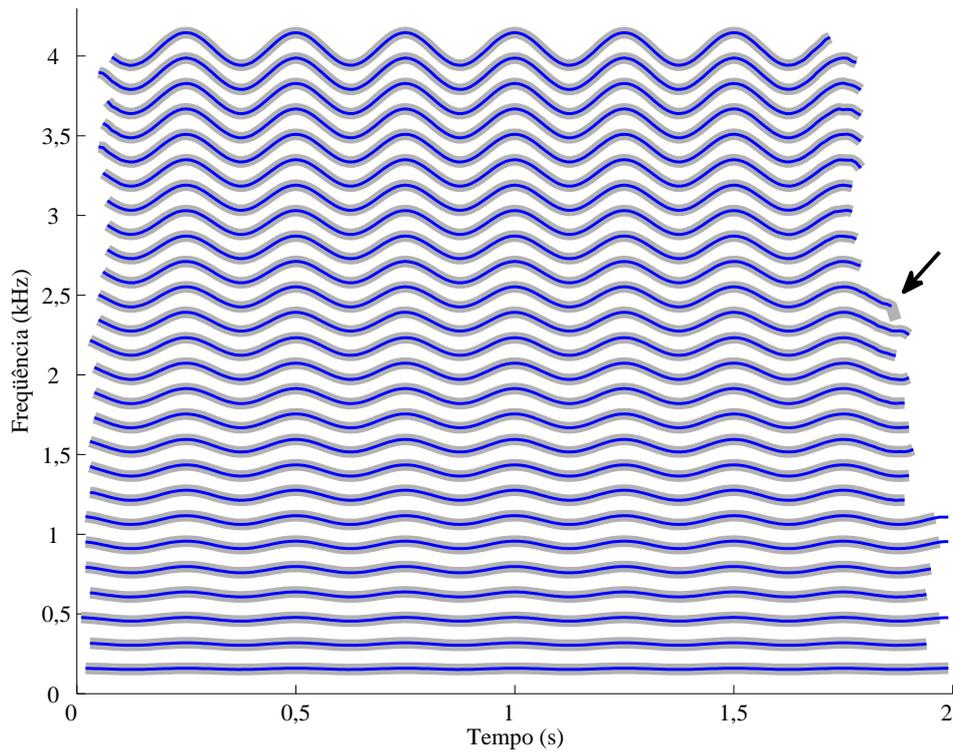
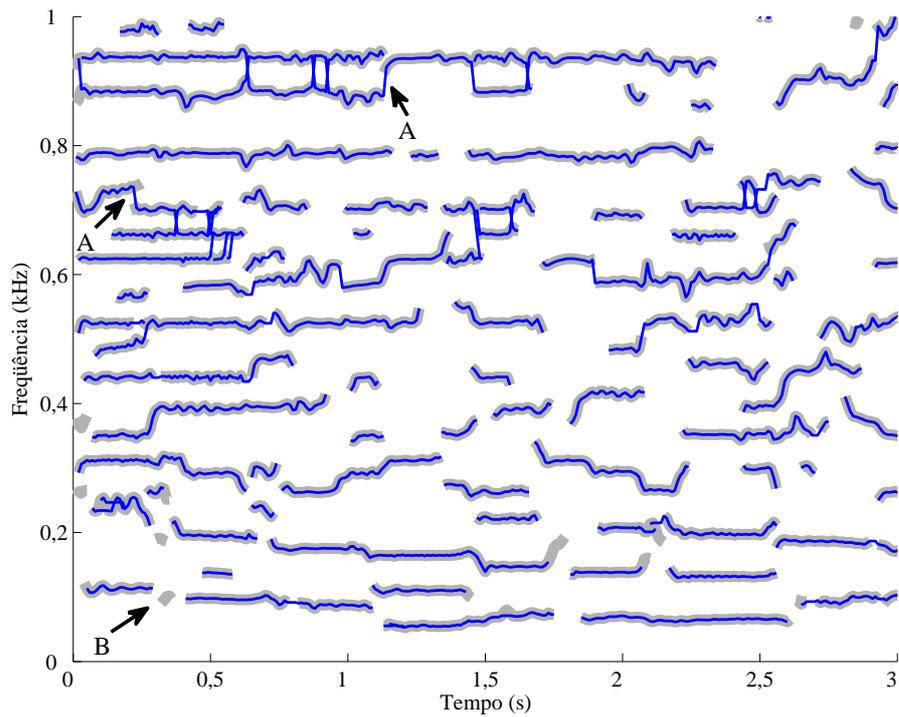
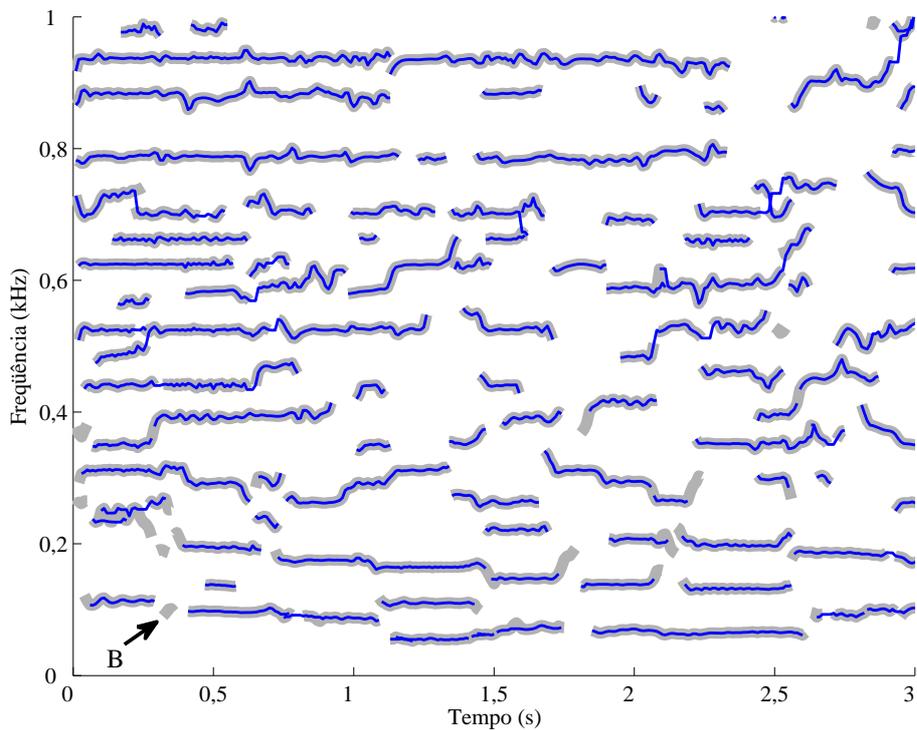


Figura 5.2: Trajetórias de frequência para as trilhas de referência (cinza) e as obtidas pelo método *lattice* para o sinal do Nível 1. A seta aponta para um erro em que o final de uma trajetória não foi corretamente rastreado.

notadamente para os sinais dos Níveis 2 e 3. No entanto, este teste não foi capaz de indicar vantagem significativa de um determinado preditor em relação aos demais.



(a) Método MQ



(b) Método *lattice*

Figura 5.3: Trajetórias de frequência para as trilhas de referência (cinza) e as obtidas pelo método MQ (a) e *lattice* (b) para o sinal do Nível 3. As setas “A” apontam para trajetórias que foram “confundidas” pelo rastreador MQ. Já as setas “B” apontam para uma trilha que não foi detectada tanto para o método MQ quanto para o *lattice*.

5.4 Teste 3: Avaliação de Sistemas de Análise Senoidal

Nas duas seções anteriores foram descritos testes por meio dos quais pôde-se avaliar o desempenho de uma etapa específica de um sistema de análise senoidal; nesta seção, é descrito um teste que visa a medir de modo integral o desempenho de um sistema de análise senoidal. Serão comparadas três configurações distintas de um sistema de análise utilizando o mesmo avaliador que o teste anterior. São escolhidos valores usuais para os parâmetros dos três sistemas, de modo a permitir a verificação da sensibilidade dos parâmetros aos diferentes sinais sob teste.

Serão utilizados no teste todos os sinais dos Níveis 1, 2 e 3 do BTS. Os sinais do Nível 0 não serão avaliados, devido a sua simplicidade.

Na Seção 5.4.1, é feita uma descrição do teste realizado. Os resultados são apresentados na Seção 5.4.2.

5.4.1 Descrição do Teste

Tratamento dos Dados de Entrada

São utilizados como dados de entrada do teste os sinais fonte do BTS sem nenhuma modificação.

Configurações do Sistema sob Teste

O teste utiliza três configurações de um sistema de análise senoidal: a partir de uma configuração básica, são introduzidas sequencialmente duas melhorias no sistema.

As três configurações utilizam uma STFT, com parâmetros definidos na Tabela 5.7, para obter uma representação no domínio tempo-freqüência do sinal sob teste.

A **primeira configuração** utiliza o detector de picos descrito na Seção 2.3.4 sem nenhum método de pré-processamento espectral. A estimativa do nível médio do chão de ruído é fornecida pelo método SSE com um filtro média móvel de ordem 150 e foi utilizado um $d = 3$. Foi empregada a reatribuição de freqüência para refinar a estimativa da freqüência de cada pico detectado, cuja amplitude também

Tabela 5.7: Valores dos parâmetros da STFT utilizados no Teste 3.

Parâmetro	Valor
Janela	Hann com duração de 40 ms
Salto entre janelas adjacentes	10 ms
Comprimento da DFT	2048 pontos

foi refinada através da equação (2.21). Para a etapa de rastreamento, foi escolhido o algoritmo MQ. Os parâmetros do sistema podem ser vistos na Tabela 5.8.

Tabela 5.8: Parâmetros utilizados na primeira configuração do sistema de análise.

Parâmetros				
d	S (ms)	E (ms)	Δ_f (%)	q
3	40	60	3	0,5

A **segunda configuração** acrescenta à primeira um método de pré-processamento espectral: foi utilizada a estimativa de tendência espectral fornecida pelo método SSE com um filtro média móvel de ordem 50, tendo-se escolhido $d = 3$.

Na **terceira configuração**, foi adicionado o preditor RLS conjunto com os parâmetros escolhidos exibidos na Tabela 5.9.

Tabela 5.9: Parâmetros utilizados na terceira configuração do sistema de análise.

Parâmetros			
Δ_f (%)	J	λ	α
1,5	4	0,98	500

Avaliação de Desempenho

A avaliação foi feita comparando-se as trilhas fornecidas por cada uma das configurações com as trilhas de referência fornecidas pelo BTS. Para tal, as trilhas de referência tiveram que ser sub-amostradas de modo a possuírem um período de amostragem igual a 10 ms.

5.4.2 Resultados

Nas Figuras 5.4 e 5.5, podem ser observados os resultados obtidos com as três configurações para os sinais dos Níveis 1 e 2, respectivamente. Na Tabela 5.10, podem ser vistos os resultados para o sinal do Nível 3.

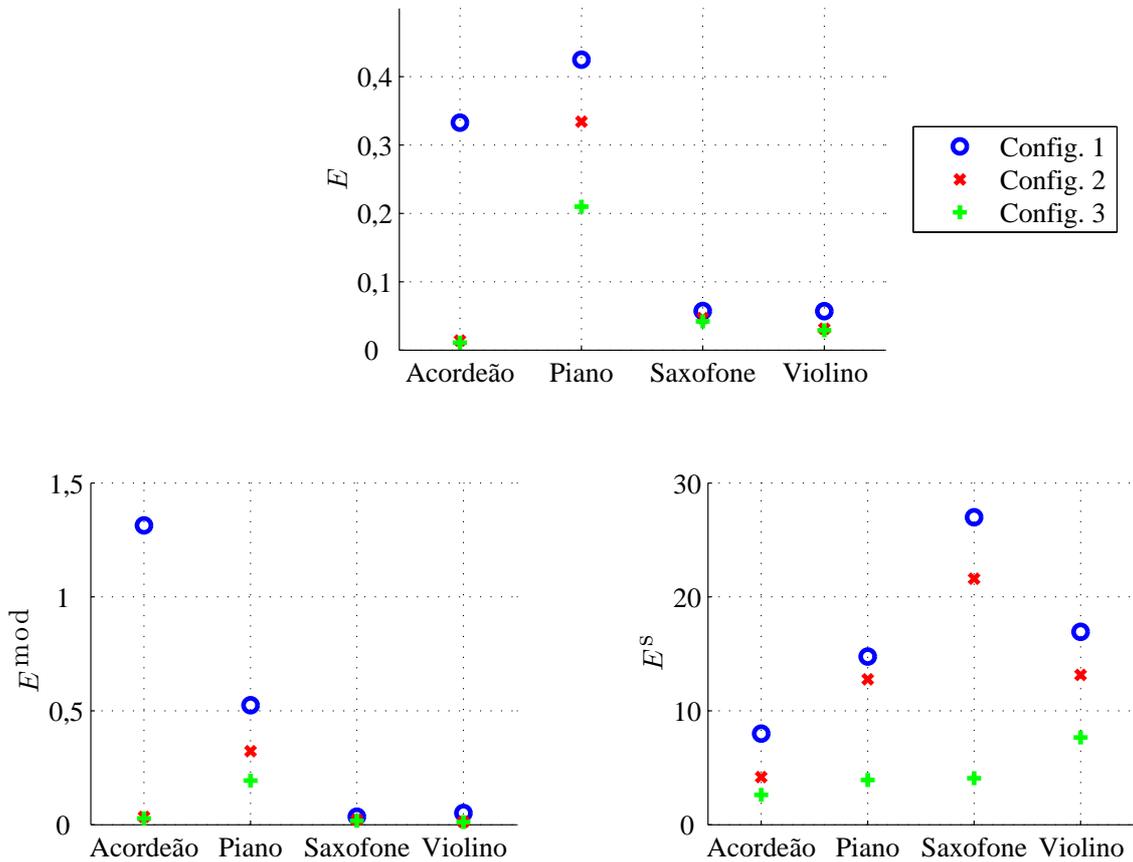


Figura 5.4: Resultado obtido no Teste 3 para os sinais do Nível 1.

Tabela 5.10: Resultado obtido no Teste 3 para o sinal do Nível 3.

Configuração	Figuras de Mérito		
	E	E^{mod}	E^s
1	$1,57 \times 10^{-1}$	$3,10 \times 10^{-2}$	$1,66 \times 10^1$
2	$1,42 \times 10^{-1}$	$2,32 \times 10^{-2}$	$1,56 \times 10^1$
3	$1,29 \times 10^{-1}$	$2,01 \times 10^{-2}$	$8,38 \times 10^1$

Pelas Figuras 5.4 e 5.5, pode-se perceber que para todos os sinais os resultados melhoraram progressivamente conforme a “sofisticação” da configuração. Além

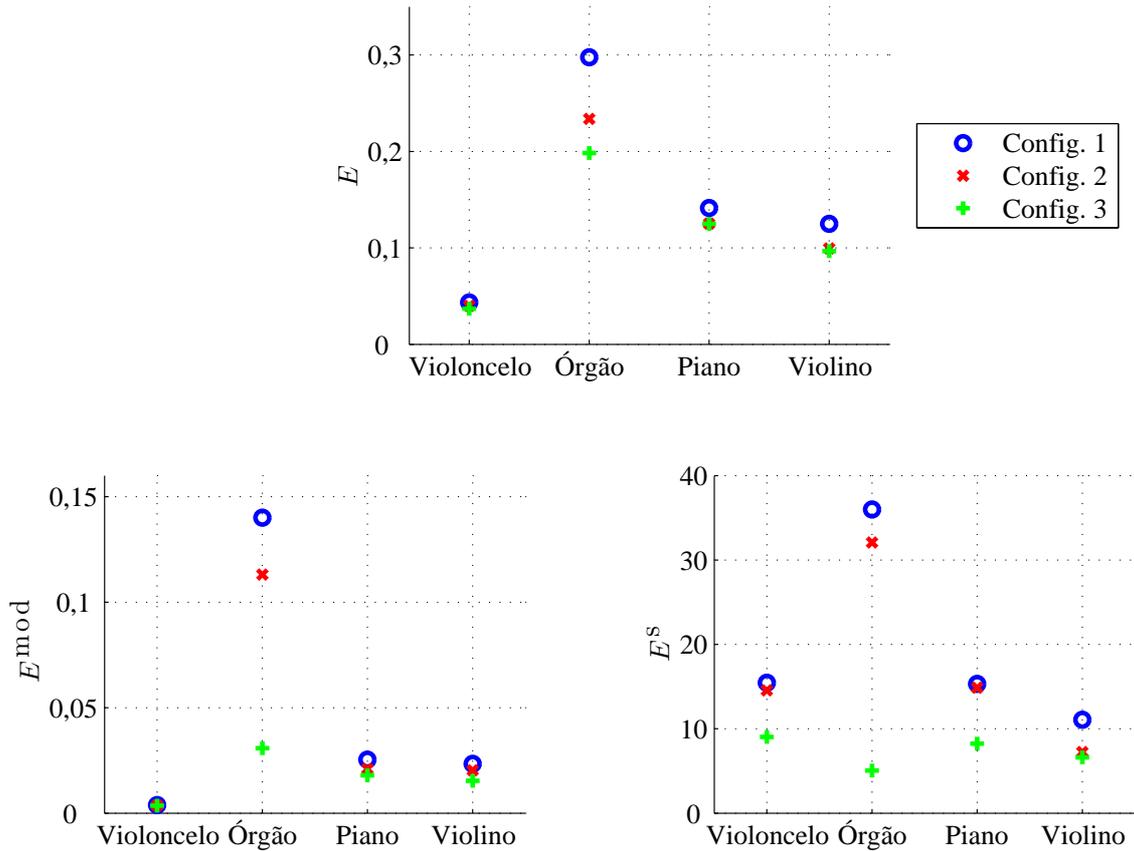


Figura 5.5: Resultado obtido no Teste 3 para os sinais do Nível 2.

disso, os resultados obtidos para a terceira configuração parecem ser menos dependentes do sinal sob teste. Já para a primeira configuração, a utilização de um único conjunto de parâmetros para a análise dos sinais se mostrou inadequada, indicando que é necessário fazer um ajuste fino dos parâmetros conforme o sinal sob análise. Dentre as figuras de mérito, E^{s} apresentou a maior diferença absoluta para as três configurações, principalmente da segunda para a terceira. Isto indica que o rastreador adaptativo consegue compensar algumas falhas na detecção de picos, levando a uma redução em erros de segmentação/união de trilhas.

Deve-se ressaltar os resultados obtidos para alguns sinais. No Nível 1, o sinal de acordeão apresentou um valor elevado para E^{mod} quando analisado com a primeira configuração, mas obteve o menor valor dentre os sinais do mesmo nível para E^{s} . Esse comportamento pode ter sido causado pelos parâmetros utilizados para análise, indicando que foram cometidos muitos erros de estimação. No Nível 2, o sinal de órgão apresentou uma diferença elevada entre as três configurações para E^{mod} e E^{s} ,

o que já não foi observado em E . Isso sugere que se estude com mais detalhe de que forma se inter-relacionam as figuras de mérito.

Capítulo 6

Conclusões e Trabalhos Futuros

6.1 Conclusões

Esta dissertação apresentou soluções para a obtenção e avaliação da qualidade de um modelo senoidal de um sinal de áudio. As principais contribuições são:

1. descrição de um esquema de seleção de picos senoidais;
2. proposição de um método de rastreamento de trilhas que utiliza filtros adaptativos [P3];
3. proposição de melhorias em algoritmos de avaliação de desempenho;
4. criação de um banco de trilhas senoidais para a avaliação de desempenho.

Estas contribuições serão detalhadas a seguir.

6.1.1 Detecção de Senóides

No Capítulo 2, foi apresentado um algoritmo para detectar picos senoidais em sinais de áudio. O algoritmo opera no espectro do sinal, obtido através de uma STFT, e detecta os picos de acordo com um limiar pré-estabelecido na magnitude. É proposto um esquema de pré-processamento do espectro que remove a tendência espectral de sinais de áudio. São descritas algumas estratégias para o posicionamento do limiar em magnitude após o pré-processamento que incluem diferentes estimativas do nível médio do chão de ruído. Também são brevemente mostradas soluções para a estimação da frequência e da amplitude dos picos detectados.

Avaliou-se o desempenho do algoritmo de detecção de picos quando utilizado em conjunto com os quatro métodos de estimação da tendência espectral apresentados. Foram criados um gerador de sinais e métricas específicas para esta avaliação. Os quatro estimadores de tendência espectral testados favoreceram a detecção de picos, sem contudo a contribuição de qualquer deles se diferenciar significativamente. Foram descritos em [P2] o algoritmo de detecção de picos e a metodologia experimental empregada.

O Capítulo 2 termina com a descrição de uma metodologia unificada de avaliação que contempla a etapa de detecção e de estimação dos parâmetros de uma senóide. Em particular, é proposta uma modificação sobre o avaliador descrito em [45] que fornece duas figuras de mérito associadas ao desempenho de algoritmos de detecção de senóides: uma relacionada à detecção (falsos positivos e negativos) e outra à estimação de frequência. Mais adiante, já no Capítulo 5, este avaliador é satisfatoriamente utilizado na comparação de três algoritmos de estimação de frequência, sendo os resultados obtidos compatíveis com os reportados na literatura.

6.1.2 Rastreamento de Trilhas

No Capítulo 3, foi descrito um algoritmo de rastreamento de trilhas senoidais que utiliza a informação passada da amplitude e da frequência de uma trilha para auxiliar a busca pela sua melhor continuação. O esquema de rastreamento apresentado se divide em uma etapa de predição, quando a informação passada da trilha é utilizada para prever um valor futuro de sua trajetória, e uma etapa de decisão, quando o valor predito para a trajetória é comparado com os picos detectados e é escolhida uma continuação da trilha. Para a etapa de predição, foram propostas duas soluções utilizando preditores adaptativos RLS. A primeira (descrita em [P3]) permite a predição conjunta dos valores de frequência e de amplitude da trilha, o que pode levar a resultados melhores por aproveitar uma correlação entre estes dois parâmetros. Embora este filtro permita a predição isolada das trajetórias dos parâmetros de uma trilha, a segunda solução (descrita em [P4]), que emprega uma estrutura *lattice*, possui menor complexidade computacional, sendo mais indicada para este caso. Para melhorar o desempenho do esquema de rastreamento, foram propostas heurísticas tanto para verificar a convergência dos preditores quanto para atualizá-los no caso

de haver dados faltantes. Para a etapa de decisão, são descritas diversas funções que auxiliam a escolher o pico que melhor continua uma determinada trilha.

No final do capítulo é descrita uma metodologia de avaliação de desempenho de métodos de rastreamento de trilhas que permite avaliar apenas o desempenho de rastreadores, sem a influência das demais etapas necessárias à obtenção do modelo senoidal. É apresentado um avaliador já descrito na literatura [45] e é proposta uma modificação sobre este avaliador. Esses dois avaliadores são utilizados nos testes descritos no Capítulo 5, no qual são comparados os dois preditores, RLS e RLS-*lattice*, com o descrito em [129].

6.1.3 Banco de Trilhas

No Capítulo 4, foi detalhado o desenvolvimento de um banco de trilhas (BTS) para a avaliação de sistemas de análise senoidal e foram especificados tanto os sinais-fonte quanto os métodos de análise. Um grupo inicial de sinais-fonte foi selecionado para gerar o primeiro conjunto de trilhas de referência, obtidas por um esquema de análise senoidal cujos parâmetros eram manualmente selecionados para cada sinal. Foi possível verificar que os erros cometidos na estimação dos parâmetros de um conjunto restrito de sinais sintéticos estavam abaixo das tolerâncias especificadas.

No Capítulo 5, o BTS foi utilizado em três avaliações de desempenho que demonstraram como as trilhas de referência podem ser utilizadas diferentes cenários de teste. Os testes indicam que os avaliadores, quando utilizados em conjunto com o BTS, são uma valiosa ferramenta para a medição de desempenho de algoritmos de análise senoidal.

6.2 Trabalhos Futuros

Possíveis continuações para a pesquisa realizada serão detalhadas para cada contribuição deste trabalho.

6.2.1 Detecção de Senóides

Na etapa de detecção de senóides, pode-se investigar novos algoritmos para a estimação do espectro de áudio, que substituiriam a STFT. Em particular, o método

descrito em [130] parece ser adequado para a tarefa em questão. Novos métodos paramétricos como os descritos em [50] também têm aparecido com frequência na literatura e merecem uma investigação mais aprofundada no contexto de análise senoidal.

Para a detecção de picos, deve-se estudar as maneiras mais adequadas para o cálculo do nível médio de ruído, especialmente para quadros onde há presença exclusiva de ruído ou de ressonâncias. Também é necessária a investigação da utilização de outras informações que não apenas a magnitude dos picos, conforme proposto em [14].

6.2.2 Rastreamento de Trilhas

Na etapa de rastreamento de trilhas, pode-se melhorar o desempenho do rastreador se a busca por uma continuação for realizada em mais de um quadro por vez [10]. Pode-se também investigar uma solução *lattice* para o problema de predição conjunta das trajetórias de amplitude e frequência de uma trilha, apesar do aparente incremento em complexidade computacional que isso gerará [131]. O esquema descrito em [132] pode ser utilizado para melhorar o desempenho do preditor *lattice* com um incremento computacional negligenciável e parece ser uma continuação natural para o seu desenvolvimento. Também pode-se pesquisar a utilização do esquema de predição proposto em conjunto com outras informações sobre a trilha [87, 13], além da amplitude e da frequência.

6.2.3 Banco de Trilhas Senoidais

O banco de trilhas senoidais pode ter o seu número de sinais-fonte expandido, aumentando o número de sinais de maior dificuldade de análise (Nível 3). Além disso, podem-se incluir sinais projetados para avaliar problemas específicos, como trajetórias de frequência se cruzando.

6.2.4 Avaliação de Desempenho

As metodologias de avaliação de desempenho e o banco de trilhas podem ser utilizados para uma comparação detalhada de diferentes métodos de análise senoidal.

Em particular, propõe-se comparar quantitativamente o desempenho de algoritmos de rastreamento de trilhas.

Referências do Autor no Tema da Dissertação

- [P1] NUNES, L. O., “Analisador Senoidal de Sinais Musicais”, Projeto Final, September 2007, Departamento de Engenharia Eletrônica e de Computação, Escola Politécnica, UFRJ. Rio de Janeiro, Brasil.
- [P2] NUNES, L. O., ESQUEF, P. A. A., BISCAINHO, L. W. P., “Evaluation of Threshold-Based Algorithms for Detection of Spectral Peaks in Audio”. In: *Anais do 5o Congresso de Engenharia de Áudio*, pp. 66–73, São Paulo, Brazil, May 2007.
- [P3] NUNES, L. O., MERCHED, R., BISCAINHO, L. W. P., “Recursive Least-Squares Estimation of the Evolution of Partial in Sinusoidal Analysis”. In: *Proceedings of the 2007 IEEE Conference on Acoustics, Speech, and Signal Processing*, v. I, pp. 253–256, Honolulu, USA, April 2007.
- [P4] NUNES, L. O., ESQUEF, P. A. A., BISCAINHO, L. W. P., MERCHED, R., “Partial Tracking in Sinusoidal Modeling: An Adaptive Prediction-Based RLS Lattice Solution”. In: *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pp. 84–91, Porto, Portugal, July 2008.

Referências Bibliográficas

- [1] KLAPURI, A., DAVY, M., (eds), *Signal Processing Methods for Music Transcription*. Springer, 2006.
- [2] BOSI, M., GOLDBERG, R. E., *Introduction to Audio Coding and Standards*. Kluwer, 2003.
- [3] GODSILL, S. J., RAYNER, P. J. W., *Digital Audio Restoration*. Springer, 1998.
- [4] VIRTANEN, T., *Sound Source Separation in Monaural Music Signals*, Ph.D. Thesis, Tampere University of Technology, Tampere, Finland, November 2006.
- [5] ESQUEF, P. A. A., BISCAINHO, L. W. P., “Spectral-Based Analysis and Synthesis of Audio Signals”. In: *Advances in Audio and Speech Signal Processing: Technologies and Applications*, pp. 56 – 92, Hershey: Idea Group, 2007.
- [6] LIU, Y.-W., SMITH III, J. O., “Audio Watermarking through Deterministic plus Stochastic Signal Decomposition”, *European Journal of Information Systems*, , n. 1, January 2007.
- [7] BONADA, J., “Wide-band Harmonic Sinusoidal Modeling”. In: *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, October 2008.
- [8] KERELIUK, C., DEPALLE, P., “Improved Hidden Markov Model Partial Tracking Through Time-Frequency Analysis”. In: *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, September 2008.

- [9] LAGRANGE, M., MARCHAND, S., “Estimating the Instantaneous Frequency of Sinusoidal Components Using Phase-Based Methods”, *Journal of the Audio Engineering Society*, v. 55, n. 5, pp. 385–399, May 2007.
- [10] LAGRANGE, M., MARCHAND, S., RAULT, J.-B., “Enhancing the Tracking of Partial for the Sinusoidal Modeling of Polyphonic Sounds”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 15, n. 5, pp. 1625–1634, July 2007.
- [11] LAURENTI, N., DE POLI, G., “A Nonlinear Method for Stochastic Spectrum Estimation in the Modeling of Musical Sounds”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 2, n. 15, pp. 531 – 541, February 2007.
- [12] MARCHAND, S., DEPALLE, P., “Generalization of the Derivative Analysis Method to Non-stationary Sinusoidal Modeling”. In: *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, September 2008.
- [13] RASPAUD, M., EVANGELISTA, G., “Binaural Partial Tracking”. In: *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, 2008 2008.
- [14] ZIVANOVIC, M., RÖBEL, A., RODET, X., “Adaptive Threshold Determination for Spectral Peak Classification”. In: *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07)*, Bordeaux, France, September 2007.
- [15] DAVID, B., BADEAU, R., “Fast Sequential LS Estimation for Sinusoidal Modeling and Decomposition of Audio Signals”. In: *Proceedings of the 2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 211–214, New Paltz, USA, October 2007.
- [16] SMITH III, J. O., SERRA, X., “PARSHL: An Analysis/Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation”. In: *Proceedings of the International Computer Music Conference*, v. 76 (6), pp. 1738–1742, Champaign-Urbana, USA, 1987.

- [17] FLANAGAN, J. L., MEINHART, D. I. S., GOLDEN, R. M., et al., “Phase Vocoder”, *The Journal of the Acoustical Society of America*, v. 38, n. 5, pp. 939–940, November 1965.
- [18] MCAULAY, R., QUATIERI, T., “Speech Analysis/Synthesis Based on a Sinusoidal Representation”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 34, n. 4, pp. 744–754, August 1986.
- [19] GEORGE, E. B., SMITH, M. J. T., “Speech Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model”, *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. 5, n. 5, pp. 389–206, September 1997.
- [20] ANDERSON, D. V., CLEMENTS, M. A., “Efficient Multi-resolution Sinusoidal Modeling”. In: *Proceedings of the Joint Meeting of the 4th World Multiconference on Systemics, Cybernetics, and Informatics (SCI2000) and the 6th International Conference on Informatin Systems Analysis and Synthesis (ISAS2000)*, Orlando, USA, 2000.
- [21] GOODWIN, M. M., “Multiscale Overlap-and-Add Sinusoidal Modeling Using Matching Pursuit and Refinements”. In: *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (WASPAA01)*, pp. 207–210, New Paltz, USA, October 2001.
- [22] HAYES, M. H., *Statistical Digital Signal Processing and Modeling*. Wiley, 2001.
- [23] LEVINE, S. N., VERMA, T. S., SMITH III, J. O., “Multiresolution Sinusoidal Modeling for Wideband Audio with Modifications”. In: *Proceedings of the 1998 IEEE Conference on Acoustics, Speech, and Signal Processing*, v. 6, pp. 3585–3588, Washington, USA, May 1998.
- [24] LEVINE, S. N., *Audio Representations for Data Compression and Compressed Domain Processing*, Ph.D. Thesis, Stanford University, Palo Alto, USA, November 1998.
- [25] FITZ, K., *The Reassigned Bandwidth-Enhanced Method of Additive Synthesis*, Ph.D. Thesis, Department of Electrical and Computer Engineering,

University of Illinois at Urbana-Champaign., Urbana-Champaign, USA, 1999.

- [26] VERMA, T. S., MENG, T. H. Y., “An Analysis/Synthesis Tool for Transient Signals that Allows a Flexible Sines+Transients+Noise Model for Audio”. In: *Proceedings of the 1998 IEEE Conference on Acoustics, Speech and Signal Processing*, v. 6, pp. 3573–3576, Seattle, USA, May 1998.
- [27] VERMA, T. S., MENG, T. H. Y., “Extending Spectral Modeling Synthesis with Transient Modeling Synthesis”, *Computer Music Journal*, v. 24, n. 2, pp. 47–59, Fall 2000.
- [28] QUATIERI, T. F., MCAULAY, R. J., “Speech Transformations Based on a Sinusoidal Representation”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 34, n. 6, pp. 1449–1464, December 1986.
- [29] CANO, P., LOSCOS, A., BONADA, J., et al., “Voice Morphing System for Impersonating in Karaoke Applications”. In: *Proceedings of the International Computer Music Conference*, Berlin, Germany, August 2000.
- [30] BONADA, J., “F-2 Voice Solo to Unison Choir Transformation”. In: *Presented at the 118th Convention of the Audio Engineering Society*, Barcelona, Spain, May 2005.
- [31] DRIOLI, C., TISATO, G., COSI, P., et al., “Emotions and Voice Quality: Experiments with Sinusoidal Modeling”. In: *Proceedings of Voice Quality: Functions, Analysis, and Synthesis Workshop*, pp. 127–132, Geneva, Switzerland, August 2003.
- [32] PURNHAGEN, H., MEINE, N., “HILN - The MPEG-4 Parametric Audio Coding Tools”. In: *Proceedings of the 2000 IEEE International Symposium on Circuits and Systems*, v. 3, pp. 201–204, Geneva, Switzerland, May 2002.
- [33] ISO/IEC, *Coding of Audio-Visual Objects - Part 3: Audio*, Int. Std. 144963:2001, MPEG 4 (Audio Edition 2001), 2001.

- [34] VERMA, T. S., *A Perceptually Based Audio Signal Model with Application to Scalable Audio Compression*, Ph.D. Thesis, Stanford University, 2000.
- [35] HEUSDENS, R., VAJN, R., KLEIJRZ, W. B., “Sinusoidal Modeling of Audio and Speech Using Psychoacoustic-adaptive Matching Pursuits”. In: *Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, v. 5, pp. 3281–3285, Salt Lake City, USA, May 2001.
- [36] HEUSDENS, R., JENSES, J., KLEIJN, W. B., et al., “Bit-Rate Scalable Intra-frame Sinusoidal Audio Coding Based on Rate-Distortion Optimization”, *Journal of the Audio Engineering Society*, v. 54, n. 3, pp. 167–188, March 2006.
- [37] VIRTANEN, T., KLAPURI, A., “Separation of Harmonic Sound Sources Using Sinusoidal Modeling”. In: *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing*, v. 2, pp. 765–768, Istanbul, Turkey, May 2000.
- [38] VIRTANEN, T., *Audio Signal Modeling with Sinusoids Plus Noise*, Master’s Thesis, Tampere University of Technology, Tampere, Finland, March 2001.
- [39] LAGRANGE, M., MARTINS, L. G., MURDOCH, J., et al., “Normalized Cuts for Predominant Melodic Source Separation”, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 16, n. 2, pp. 278–290, February 2008.
- [40] HAINSWORTH, S. W., *Techniques for the Automated Analysis of Musical Audio*, Ph.D. Thesis, University of Cambridge, Cambridge, UK, September 2004.
- [41] WEN, X., SANDLER, M., “A Partial Searching Algorithm and its Applications for Polyphonic Music Transcription”. In: *Proceedings of the 6th International Conference on Music Information Retrieval*, v. 690–695, London, UK, September 2005.

- [42] CZYZEWSKI, A., KOSTEK, B., MAZIEWSKI, P., et al., “Accidental Wow Defect Evaluation Using Sinusoidal Analysis Enhanced by Artificial Neural Networks”. In: *Proceedings of the First International Conference on Rough Sets and Knowledge Technology*, Chengdu, China, July 2006.
- [43] BETSER, M., COLLEN, P., RAULT, J.-B., “Audio Identification Using Sinusoidal Modeling and Application to Jingle Detection”. In: *Proceedings of the 8th International Conference on Music Information Retrieval*, Vienna, Austria, September 2007.
- [44] SCHOLZ, K., LEUTELT, L., HEUTE, U., “Speech-Codec Detection by Spectral Harmonic-Plus-Noise Decomposition”. In: *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems, and Computers*, v. 2, pp. 2295–2299, Pacific Grove, USA, November 2004.
- [45] LAGRANGE, M., MARCHAND, S., “Assessing the Quality of the Extraction and Tracking of Sinusoidal Components: Towards an Evaluation Methodology”. In: *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, Montreal, Canada, September 2006.
- [46] RABINER, L. R., SCHAFER, R. W., “Digital Processing of Speech Signals”, *Journal of the Acoustical Society of America*, v. 67, n. 4, pp. 1406 – 1407, April 1980.
- [47] HAYES, M. H., *Statistical Digital Signal Processing and Modeling*. Wiley, 1996.
- [48] EMIYA, V., BADEAU, R., DAVID, B., “Multipitch Estimation of Quasi-Harmonic Sounds in Colored Noise”. In: *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07)*, Bordeaux, France, September 2007.
- [49] DAVID, B., BADEAU, R., RICHARD, G., “HRHATRAC Algorithm for Spectral Line Tracking of Musical Signals”. In: *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing*, v. 3, pp. 45–48, Toulouse, France, May 2006.

- [50] BADEAU, R., DAVID, B., RICHARD, G., “High-Resolution Spectral Analysis of Mixtures of Complex Exponentials Modulated by Polynomials”, *IEEE Transactions on Signal Processing*, v. 54, n. 4, pp. 1341–1350, April 2006.
- [51] COHEN, L., *Time Frequency Analysis: Theory and Applications*. Prentice Hall, 1995.
- [52] DINIZ, P. S. R., DA SILVA, E. A. B., NETTO, S. L., *Digital Signal Processing: System Analysis and Design*. Cambridge, 2002.
- [53] MASRI, P., BATEMAN, A., CANAGARAJAH, N., “The importance of the time–frequency representation for sound/music analysis–resynthesis”, *Organized Sound*, v. 2, n. 3, pp. 207–214, 1997.
- [54] HARRIS, F. J., “On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform”, *Proceedings of the IEEE*, v. 66, n. 1, pp. 51–83, January 1978.
- [55] PIELEMEIER, W., WAKEFIELD, G., “A high-resolution time-frequency representation for musical instrument signals”, *Journal of the Acoustical Society of America*, v. 99, n. 4, pp. 2382–2396, April 1996.
- [56] FULOP, S. A., FITZ, K., “Algorithms for Computing the Time-Corrected Instantaneous Frequency (Reassigned) Spectrogram, with Applications”, *Journal of the Acoustical Society of America*, v. 119, n. 1, pp. 360–371, January 2006.
- [57] PAINTER, T., SPANIAS, A., “Sinusoidal Analysis-Synthesis of Audio Using Perceptual Criteria”, *EURASIP Journal on Applied Signal Processing*, v. 1, pp. 15–20, 2003.
- [58] THOMSON, D. J., “Spectrum estimation and harmonic analysis”, *Proceedings of the IEEE*, v. 70, n. 9, pp. 1055 – 1096, September 1982.
- [59] GREY, J. M., GORDON, J. W., “Perceptual Effects of the Spectral Modifications of Musical Timbres”, *Journal of the Acoustical Society of America*, v. 61, n. 5, pp. 1493 – 1500, May 1978.

- [60] MAKHOUL, J., “Spectral Analysis of Speech by Linear Prediction”, *IEEE Transactions on Audio and Electroacoustics*, v. 21, n. 3, pp. 140–148, June 1973.
- [61] MAKHOUL, J., “Linear prediction: A tutorial review”, *Proceedings of the IEEE*, v. 63, n. 4, pp. 561 – 580, April 1975.
- [62] HAYKIN, S., *Adaptive Filter Theory*. 4th ed. Prentice Hall, 2001.
- [63] VAIDYANATHAN, P. P., *The Theory of Linear Prediction*. Morgan & Claypool, 2008.
- [64] STRUZINSKI, W. A., LOWE, E. D., “A Performance Comparison of Four Noise Background Normalization Schemes Proposed for Signal Detection Systems”, *Journal of the Acoustical Society of America*, v. 76, n. 6, pp. 1738–1742, December 1984.
- [65] LAURENTI, N., POLI, G. D., “A Method for Spectrum Separation and Envelope Estimation of the Residual in Spectrum Modeling of Musical Sound”. In: *Proceedings of the 3rd COST-G6 Conference on Audio Digital Effects (DAFx)*, Verona, Italy, December 2000.
- [66] MACLEOD, M. D., “Nonlinear Recursive Smoothing Filters and Their Use for Noise Floor Estimation”, *IEEE Electronics Letters*, v. 28, n. 21, pp. 1952 – 1953, October 1992.
- [67] MACLEOD, M. D., “Performance Analysis of Simple Non-Linear Recursive Smoothing Filters”. In: *Proceedings of the IEEE Colloquium on Non-Linear Filters*, v. 5, pp. 1 – 6, London, UK, May 1994.
- [68] DONOHO, D. L., *Wavelet shrinkage and W.V.D.: a 10-minute tour*, Tech. Rep. 416, Department of Statistics, Stanford University, June 1993.
- [69] SERRA, X., “Musical Sound Modeling with Sinusoids plus Noise”. In: *Musical Signal Processing*, Swets & Zeitlinger Publishers, 1997.
- [70] KEILER, F., ZÖLZER, U., “Extracting Sinusoids from Harmonic Signals”, *Journal of New Music Research*, v. 30, n. 3, pp. 243–258, September 2001.

- [71] BROWN, J. C., PUCKETTE, M. S., “A High Resolution Fundamental Frequency Determination Based on Phase Changes of the Fourier Transform”, *Journal of the Acoustical Society of America*, v. 94, n. 2, pp. 662 – 667, August 1993.
- [72] DAVID, P. A. M.-S., SZCZUPAK, J., “Refining the Digital Spectrum”. In: *Proceedings of the IEEE 39th Midwest Symposium on Circuits and Systems*, v. 2, pp. 767 – 770, Ames, USA, August 1996.
- [73] HAINSWORTH, S. W., WOLFE, P. J. ., “Time-Frequency Reassignment for Music Analysis”. In: *Proceedings of the International Computer Music Conference*, pp. 14–17, Havana, Cuba, September 2001.
- [74] LAGRANGE, M., MARCHAND, S., RAULT, J.-B., “Improving Sinusoidal Frequency Estimation Using a Trigonometric Approach”. In: *Proceedings of the 8th International Conference on Digital Audio Effects (DAFx’05)*, pp. 110–115, Madrid, Spain, September 2005.
- [75] MARCHAND, S., LAGRANGE, M., “On the Equivalence of Phase-Based Methods for the Estimation of Instantaneous Frequency”. In: *Proceedings of the 14th European Conference on Signal Processing (EUSIPCO 2006)*, Florence, Italy, September 2006.
- [76] HAINSWORTH, S. W., MACLEOD, M. D., “On Sinusoidal Parameter Estimation”. In: *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx)*, pp. 151–156, London, UK, September 2003.
- [77] BETSER, M., COLLEN, P., RICHARD, G., et al., “Review and Discussion on Classical STFT-Based Frequency Estimators”. In: *Presented at the 120th Convention of the Audio Engineering Society*, 2006, preprint no. 6765.
- [78] KODERA, K., GENDRIN, R., DE VILLEDARY, C., “Analysis of Time-Varying Signals with Small BT Values”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 26, n. 1, pp. 64 – 78, February 1978.
- [79] AUGER, F., FLANDRIN, P., “Generalization of the Reassignment Method to All Bilinear Time-Frequency and Time-Scale Representations”. In: *Pro-*

- ceedings of the 1994 IEEE International Conference on Acoustics, Speech, and Signal Processing*, v. 4, pp. 317–320, Adelaide, Australia, April 1994.
- [80] AUGER, F., FLANDRIN, P., “Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method”, *IEEE Transactions on Signal Processing*, v. 43, n. 5, pp. 1068–1089, May 1995.
- [81] STOICA, P., LI, H., LI, J., “Amplitude Estimation of Sinusoidal Signals: Survey, New Results, and an Application”, *IEEE Transactions on Signal Processing*, v. 48, n. 2, pp. 338–352, February 2000.
- [82] QUATIERI, T. F., MCAULAY, R. J., “Audio Signal Processing Based on Sinusoidal Analysis and Synthesis”. In: *Applications of Digital Signal Processing to Audio and Acoustics*, pp. 312–416, Kluwer, 1998.
- [83] COHEN, L., “The History of Noise”, *IEEE Signal Processing Magazine*, v. 22, n. 6, pp. 20–45, November 2005.
- [84] PEEBLES JR., P., *Probability, Random Variables, and Random Signal Principles*. 4th ed. McGraw-Hill, 2000.
- [85] KEILER, F., MARCHAND, S., “Survey on Extraction of Sinusoids in Stationary Sounds”. In: *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx’02)*, v. 1, pp. 51–58, Hamburg, Germany, September 2002.
- [86] HORNER, A., AYERS, L., “Modeling Acoustic Wind Instruments with Contiguous Group Synthesis”, *Journal of the Audio Engineering Society*, v. 46, n. 10, pp. 868–879, October 1998.
- [87] PURNHAGEN, H., “Parameter Estimation and Tracking for Time-Varying Sinusoids”. In: *Proceedings of the 1st IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002)*, pp. 5–8, Leuven, Belgium, November 2002.
- [88] DEPALLE, P., GARCIA, G., RODET, X., “Tracking of Partial for Additive Sound Synthesis Using Hidden Markov Models”. In: *Proceedings of the*

1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, v. 1, pp. 225–228, Minneapolis, USA, April 1993.

- [89] LAGRANGE, M., MARCHAND, S., RAULT, J.-B., “Partial Tracking Based on Future Trajectories Exploration”. In: *Presented at the 116th Convention of the Audio Engineering Society*, Berlin, Germany, May 2004, preprint no. 6046.
- [90] STERIAN, A., WAKEFIELD, G. H., “A model-based approach to partial tracking for musical transcription”. In: *Proceedings of the 1998 SPIE Annual Meeting*, San Diego, USA, July 1998.
- [91] SATAR-BOROJENI, H., *Partial Tracking of Music Signals using Robust Kalman Filtering*, Ph.D. Thesis, Northeastern University, 2006.
- [92] LAGRANGE, M., TZANETAKIS, G., “Sound Source Tracking and Formation Using Normalized Cuts”. In: *Proceedings of the 2007 IEEE Conference on Acoustics, Speech, and Signal Processing*, v. I, pp. 61 – 64, Honolulu, USA, April 2007.
- [93] LAGRANGE, M., MARCHAND, S., RASPAUD, M., et al., “Enhanced Partial Tracking Using Linear Prediction”. In: *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx’03)*, London, UK, September 2003.
- [94] LAGRANGE, M., MARCHAND, S., RAULT, J.-B., “Using Linear Prediction to Enhance The Tracking of Partials”. In: *Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, v. 4, pp. 241 – 244, Montreal, Canada, April 2004.
- [95] SAYED, A., *Fundamentals of Adaptive Filtering*. Wiley-IEEE, 2003.
- [96] DINIZ, P. S. R., *Adaptive Filtering*. 3rd ed. Springer, 2008.
- [97] VERFAILLE, V., GUASTAVINO, C., DEPALLE, P., “Perceptual Evaluation of Vibrato Models”. In: *Proceedings of the Conference on Interdisciplinary Musicology*, pp. 149–151, Montreal, Canada, March 2005.

- [98] FRIEDLANDER, B., “Lattice Methods for Spectral Estimation”, *Proceedings of the IEEE*, v. 70, n. 9, pp. 990–1017, September 1982.
- [99] FRIEDLANDER, B., “Lattice Filters for Adaptive Processing”, *Proceedings of the IEEE*, v. 70, n. 8, pp. 829–867, August 1982.
- [100] MERCHED, R., “Extended RLS Lattice Adaptive Filters”, *IEEE Transactions on Signal Processing*, v. 51, n. 9, pp. 2294–2309, September 2003.
- [101] WRIGHT, M., BEAUCHAMP, J., FITZ, K., et al., “Analysis/synthesis comparison”, *Organized Sound*, v. 5, n. 3, pp. 173–189, December 2000.
- [102] HERRERA-BOYER, P., “Setting Up an Audio Database for Music Information Retrieval Benchmarking”, In: *The MIR/MDL Evaluation Project White Paper Collection*, 3rd ed., pp. 53–55, Available online, 2005, http://www.music-ir.org/evaluation/wp3/wp3_entire.pdf.
- [103] BECH, S., ZACHAROV, N., *Perceptual Audio Evaluation – Theory, Method and Application*. Wiley, 2006.
- [104] MICROSOFT, “Multiple Channel Audio Data and WAVE Files”, webpage, <http://www.microsoft.com/whdc/device/audio/multichaud.msp>.
- [105] GOTO, M., HASHIGUCHI, H., NISHIMURA, T., et al., “RWC Music Database: Popular, Classical, and Jazz Music Databases”. In: *Proceedings of the 3rd International Conference on Music Information Retrieval*, pp. 287–288, Paris, France, October 2002.
- [106] GOTO, M., HASHIGUCHI, H., NISHIMURA, T., et al., “RWC Music Database: Music Genre Database and Musical Instrument Sound Database”. In: *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, pp. 229–230, 2003.
- [107] GOTO, M., “Development of the RWC Music Database”. In: *Proceedings of the 18th International Congress on Acoustics*, v. I, pp. 553–556, Kyoto, Japan, April 2004.

- [108] GOTO, M., “AIST Annotation for the RWC Music Database”. In: *Proceedings of the 7th International Conference on Music Information Retrieval*, Victoria, Canada, October 2006.
- [109] NETO, A. M. M., “Sintetizador de Instrumentos de Sopro”, Projeto Final, 2004, Departamento de Engenharia Eletrônica e de Computação, Escola Politécnica, UFRJ. Rio de Janeiro, Brazil.
- [110] JÄRVELÄINEN, H., “Perception-based control of vibrato parameters in string instrument synthesis”. In: *Proceedings of the International Computer Music Conference*, pp. 287–294, Gothenburg, Sweden, September 2002.
- [111] FITZ, K., HAKEN, L., “On the Use of Time-Frequency Reassignment in Additive Sound Modeling”, *Journal of the Audio Engineering Society*, v. 50, n. 11, pp. 879–893, November 2002.
- [112] ELLIS, D. P. W., *A Perceptual Representation of Audio*, M.Sc. Thesis, Massachusetts Institute of Technology, February 1992.
- [113] MOORE, B. C. J., PETERS, R. W., GLASBERG, B. R., “Detection of temporal gaps in sinusoids: Effects of frequency and level”, *Journal of the Acoustical Society of America*, v. 93, n. 3, pp. 1563–1570, March 1993.
- [114] MOORE, B. C. J., *An Introduction to the Psychology of Hearing*. 5th ed. Elsevier, 2004.
- [115] BELLO, J. P., DAUDET, L., ABDALLAH, S., et al., “A Tutorial on Onset Detection in Music Signals”, *IEEE Transactions on Speech and Audio Processing*, v. 13, n. 5, pp. 1035–1047, September 2005.
- [116] DIXON, S., “Onset Detection Revisited”. In: *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, pp. 133–137, Montreal, Canada, September 2006.
- [117] COLLINS, N., “A Comparison of Sound Onset Detection Algorithms with Emphasis on Psychoacoustically Motivated Detection Functions”. In: *Presented at the 118th Convention*, Barcelona, Spain, May 2005, Convention Paper 6363.

- [118] SEK, A., MOORE, B. C. J., “Frequency discrimination as a function of frequency, measured in several ways”, *Journal of the Acoustical Society of America*, v. 97, n. 4, pp. 2479–2486, April 1995.
- [119] WIER, C. C., JESTEADT, W., GREEN, D. M., “Frequency discrimination as a function of frequency and sensation level”, *Journal of the Acoustical Society of America*, v. 61, n. 1, pp. 178–184, January 1977.
- [120] JESTEADT, W., WIER, C. C., GREEN, D. M., “Intensity discrimination as a function of frequency and sensation level”, *Journal of the Acoustical Society of America*, v. 61, n. 1, pp. 169–177, January 1977.
- [121] WARREN, R. M., *Auditory Perception: An Analysis and Synthesis*. Third edition ed. Cambridge, 2008.
- [122] MATHWORKS, “Matlab Help”, webpage, 2008, available at: <http://www.mathworks.com/access/helpdesk/help/techdoc/matlab.shtml>.
- [123] LEVINE, S. N., VERMA, T. S., SMITH III, J. O., “Alias-free, Multiresolution Sinusoidal Modeling for Polyphonic, Wideband Audio”. In: *Proceedings of the 1997 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA01)*, New Paltz, USA, October 1997.
- [124] TRAUTMANN, L., VALIMÄKI, V., “A Multirate Approach to Physical Modeling Synthesis Using the Functional Transformation Method”. In: *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 221–224, New Paltz, USA, October 2003.
- [125] ABE, M., SMITH III, J. O., “Design Criteria for Simple Sinusoidal Parameter Estimation based on Quadratic Interpolation of FFT Magnitude Peaks”. In: *Presented at the 117th Convention of the Audio Engineering Society*, v. 58, pp. 104 – 117, San Francisco, USA, October 2004.
- [126] ABE, M., SMITH III, J. O., *CQIFFT: Correcting Bias in a Sinusoidal Parameter Estimator based on Quadratic Interpolation of FFT Magnitude Peaks*, Technical Report STAN-M-117, Stanford University, October 2004.

- [127] SMITH III, J. O., “Matlab for Parabolic Peak Interpolation”, webpage, May 2009, http://ccrma.stanford.edu/~jos/sasp/Matlab_Parabolic_Peak_Interpolation.html.
- [128] SMITH III, J. O., “Matlab listing: zpfmin.m”, webpage, May 2009, http://ccrma.stanford.edu/~jos/sasp/Matlab_listing_zpfmin_m.html.
- [129] LAGRANGE, M., MARCHAND, S., RAULT, J.-B., “Tracking Partial for the Sinusoidal Modeling of Polyphonic Sounds”. In: *Proceedings of the 2005 IEEE International Conference on Acoustics, Speech and Signal Processing*, v. 3, pp. 229–232, Philadelphia, USA, March 2005.
- [130] XIAO, J., FLANDRIN, P., “Multitaper Time-Frequency Reassignment for Nonstationary Spectrum Estimation and Chirp Enhancement”, *IEEE Transactions on Signal Processing*, v. 55, n. 6, pp. 2851–2860, June 2007.
- [131] PARKER, S. R., KAYRAN, A. H., “Lattice Parameter Autoregressive Modeling of Two-Dimensional Fields-Part I: The Quarter-Plane Case”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 32, n. 4, pp. 872–885, August 1984.
- [132] SINGER, A. C., FEDER, M., “Universal Linear Prediction by Model Order Weighting”, *IEEE Transactions on Signal Processing*, v. 47, n. 10, pp. 2685–2699, October 1999.

Apêndice A

Parâmetros Utilizados no Teste 2 do Capítulo 5

Neste apêndice podem ser encontrados os parâmetros escolhidos para os preditores utilizados no Teste 2 (ver Seção 5.3) do Capítulo 5. Nas Tabelas A.1, A.2 e A.3, são exibidos os valores escolhidos para os parâmetros dos preditores Burg [94], RLS (ver Seção 3.4.1) e *lattice* (ver Seção 3.4.2), respectivamente.

Tabela A.1: Parâmetros do preditor Burg utilizados na comparação dos rastreadores.

Parâmetro \ Nível	0	1	2	3
Ordem do Preditor (frequência)	6	6	8	8
Tamanho da Janela de Observação (frequência)	20	20	30	30
Ordem do Preditor (amplitude)	4	2	4	4
Tamanho da Janela de Observação (amplitude)	20	15	30	30

Tabela A.2: Parâmetros do preditor RLS utilizados na comparação dos rastreadores.

Parâmetro \ Nível	0	1	2	3
Ordem do Preditor (J)	6	4	6	6
Fator de Esquecimento (λ)	0,98	0,98	0,98	0,98
Fator de Regularização (α)	4000	3000	2000	2000

Tabela A.3: Parâmetros do preditor *lattice* utilizados na comparação dos rastreadores.

Parâmetro \ Nível	0	1	2	3
Ordem do Preditor (frequência, J^f)	8	6	6	6
Fator de Esquecimento (frequência, λ^f)	0,99	0,99	0,99	0,99
Fator de Regularização (frequência, α^f)	2000	2000	2000	1000
Ordem do Preditor (amplitude, J^A)	4	4	2	2
Fator de Esquecimento (amplitude, λ^A)	0,99	0,98	0,98	0,98
Fator de Regularização (amplitude, α^A)	2000	2000	2000	2000