

GERAÇÃO EFICIENTE DE SOM TRIDIMENSIONAL

Fábio Pacheco Freeland

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

Aprovada por:

Prof. Paulo Sergio Ramirez Diniz, Ph.D.

Prof. Luiz Wagner Pereira Biscainho, D.Sc.

Prof. Sergio Lima Netto, Ph.D.

Prof. Márcio Nogueira de Souza, D.Sc.

Prof. Jacques Szczupak, Ph.D.

Dr. Paulo Antonio Andrade Esquef, D.Sc. (Tech.)

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 2005

FREELAND, FÁBIO PACHECO

Geração Eficiente de Som Tridimensional [Rio de Janeiro] 2005

XVII, 141 pp., 29,7 cm (COPPE/UFRJ, D.Sc., Engenharia Elétrica, 2005)

Tese - Universidade Federal do Rio de Janeiro, COPPE

1.Som Tridimensional 2.Som Multicanal
3.Aproximação de Funções de Transferência

I.COPPE/UFRJ II.Título (série)

Agradecimentos

Gostaria de agradecer, primeiramente, a minha esposa Juliana, a meus pais, Victor e Inês e meus irmãos, Renata, Miriam e Eduardo, que, de uma forma ou de outra, formaram uma base bastante forte que possibilitou o desenvolvimento deste trabalho.

Novamente à Juliana pela paciência, compreensão e companherismo durante todo o decorrer desse trabalho.

Agradeço a meus orientadores, Paulo Sergio Ramirez Diniz e Luiz Wagner Pereira Biscainho, pelo suporte e idéias que fizeram com que esse trabalho fosse viável e trouxesse alguma contribuição para o grupo de Processamento de Sinais, bem como para as pessoas que trabalham na área de áudio. E, ainda, pela tranqüilidade essencial que me passaram. Agradeço a eles também toda a força nos momentos de desânimo.

Agradeço aos amigos e colegas Tadeu N. Ferreira, Leonardo G. Baltar, Filipe C. C. B. Diniz, Michel P. Tcheou, Rodrigo C. Meirelles, Bruno C. Vianna, Cristiano N. dos Santos, Alexandre L. Szczupak, Alessandro J. S. Dutra, Lisandro Lovisolo, Iúri Kothe, Ana Luisa A. Santos, Cassio B. Ribeiro, Miguel B. Furtado Jr., Wallace A. Martins e Bernardo F. P. da Costa, que sempre estiveram dispostos a discutir alguns problemas encontrados no decorrer dessa tese. Além disso, devo agradecer a alguns deles por facilitarem bastante o meu “acesso” à Ilha do Fundão. Ao amigo e colega Paulo Antonio A. Esquef, que, mesmo estando longe, ajudou-me muito com a obtenção de material bibliográfico, além de sempre incentivar o desenvolvimento desse trabalho. Aos amigos e colegas do grupo de áudio formado dentro do LPS, que sempre contribuíram para minha formação com as discussões sobre os variados assuntos em que trabalham. Não posso deixar de mencionar os amigos e colegas Ailton D. Santana Jr., Alexandre G. Ciancio, Aline G. Mello, André C. Vliese, Augusto H. Dantas, Cassio G. G. Duarte, Charles B. do Prado, Décio Ângelo Fonini Jr., Felipe R. Aquino, Lara Christiana R. L. Feio, Maria Heveline V. Duarte, Mauro F. de Carvalho, Ranniery da S. Maia, Rodrigo C. Torres, Rogério Caetano e Solimar S. Silva pela ajuda no ultrapassar dos obstáculos físicos e psicológicos.

Agradeço também a todos os meus amigos, em especial à Ana Luiza (Nana),

ao Carlos Eduardo (Cacá), à Márcia (Marcinha), ao Vinícius (Vini), à Danielle (Dani), ao Saulo, ao Daerson (Tio Dadá), à Daniela, à Fabiana (Faboo), ao Ricardo, ao Weverson (Brubs), ao Wildson (Zé) e ao meu primo Luís Eugênio (Polho), que na maior parte do tempo me ajudaram a manter “um certo nível” de sanidade.

Gostaria de agradecer também aos professores Eduardo Antônio B. da Silva, Marcello Luiz R. Campos e Sergio L. Netto, pois sempre se dispuseram a discutir algumas dúvidas mais específicas que surgiram ao longo desse trabalho.

Agradeço, também, aos outros professores, secretárias e funcionários, que proporcionaram um ambiente profissional bastante agradável.

A todas estas pessoas, muito obrigado. Espero que eu possa retribuir à altura tudo o que fizeram por mim.

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

GERAÇÃO EFICIENTE DE SOM TRIDIMENSIONAL

Fábio Pacheco Freeland

Dezembro/2005

Orientadores: Paulo Sergio Ramirez Diniz

Luiz Wagner Pereira Biscainho

Programa: Engenharia Elétrica

Esta tese tem como objetivo o estudo de técnicas de geração de som tridimensional, principalmente do ponto de vista da localização e movimentação do som. Ela se divide em duas partes: uma que trata da geração binaural de som 3D e outra que relaciona as técnicas de geração binaural e multicanal.

Nesse contexto, propõem-se novas formas de interpolação de HRTFs (*Head-Related Transfer Functions*) que diminuem a complexidade computacional do sistema, ainda que mantendo a sensação de naturalidade para o ouvinte. O primeiro método utiliza uma função auxiliar, a IPTF (*Interpositional Transfer Function*), que pode ser modelada com ordem baixa. O segundo propõe uma interpolação linear incremental sobre os coeficientes da KLT (*Karhunen Loève Transform*) relativos às HRTFs. O terceiro utiliza a estrutura de Farrow, generalizada para duas variáveis de posição. São realizados testes objetivos e subjetivos para avaliar a eficácia dos métodos propostos. Por fim, discute-se a relação entre os sistemas binaural e multicanal, quanto às suas respectivas limitações.

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

EFFICIENT TRIDIMENSIONAL SOUND GENERATION

Fábio Pacheco Freeland

December/2005

Advisors: Paulo Sergio Ramirez Diniz

Luiz Wagner Pereira Biscainho

Department: Electrical Engineering

The main objective of this thesis is to study techniques for generation of tridimensional sound, especially regarding sound localization and motion. It is divided into two parts: one that conveys the generation of binaural 3D sound and another that discusses the relation between the binaural and multichannel techniques.

In this context, new interpolation methods of HRTFs (Head-Related Transfer Functions) that offer low computational complexity without degrading the listener's natural sensation are proposed. The first method uses an auxiliary function, called IPTF (Interpositional Transfer Function), that can be described by a low-order model. The second one proposes a linear incremental interpolation over the KLT (Karhunen-Loève Transform) coefficients related to the HRTFs. The third one uses the Farrow structure, generalized for two position variables. Objective and subjective tests are made to assess the performance of those methods. In the end, binaural and multichannel methods are examined regarding their respective limitations.

Sumário

Lista de Publicações	xvi
Lista de Abreviações	xvii
1 Introdução	1
1.1 Motivação	1
1.2 Geração de Som Tridimensional	2
1.3 Percepção da Posição	4
1.3.1 Principais Características das HRIRs	6
1.3.2 Utilização das HRIRs	9
1.4 Conteúdo desta Tese	10
I Interpolação de HRIRs	14
2 Interpolação de HRIRs	15
2.1 Interpolação Bilinear	16
2.2 Generalização Proposta para Regiões Triangulares	16
2.3 Conclusões	19
3 Conformação das HRIRs	22
3.1 Suavização Espectral	22
3.1.1 Suavização Usando Média Geométrica	24
3.2 HRIRs de Fase Mínima	25
3.2.1 Obtenção de HRIRs de Fase Mínima Suavizadas	26
3.3 Aproximação das HRIRs originais	27
3.4 Conclusões	28

4	Utilização de IPTFs	30
4.1	<i>Inter-positional Transfer Functions</i>	31
4.1.1	Motivação	31
4.1.2	Definição	31
4.2	Interpolação de HRTFs Baseada nas IPTFs	32
4.3	Comparações entre as Interpolações Bilinear e Baseada em IPTFs	35
4.3.1	Configuração dos Testes e Resultados	35
4.3.2	Complexidade Computacional	37
4.4	Conclusão	42
5	Redução de Ordem das IPTFs	43
5.1	Redução do Modelo Balanceado	43
5.2	Algoritmo de Redução de Ordem	46
5.3	Comparações entre as Interpolações Baseada em IPTFs Sem e Com Redução de Ordem	48
5.4	Complexidade Computacional e Desempenho	50
5.5	Conclusões	55
6	Interpolação no Domínio da Transformada	56
6.1	Representação de HRIRs Usando a KLT	57
6.2	Modelo para as SCFs	59
6.2.1	Modelo Linear Incremental para as SCFs	61
6.3	Comparação de Desempenho	65
6.4	Conclusões	66
7	Estrutura Farrow Generalizada para Modelo de HRTF	71
7.1	Estrutura Farrow	72
7.2	Formulação Bidimensional	74
7.3	Otimização da Estrutura Generalizada	77
7.3.1	Otimização para Cada Instante	78
7.4	Comparação de Desempenho	80
7.5	Conclusões	81

8	Testes Subjetivos	87
8.1	Descrição dos Testes	87
8.1.1	Verificação de Mudança de Posição e/ou Timbre	88
8.1.2	Verificação da Percepção do Movimento	89
8.1.3	Verificação da Percepção da Posição Estática	92
8.2	Conclusões	95
II	Integração com Caixas Acústicas	99
9	Interpolação de HRIRs Baseada no VBAP	100
9.1	Posicionamento por Variação de Amplitude, com Três Canais	101
9.2	Interpolação Baseada em Três HRTFs	102
9.3	Método VBAP	103
9.4	Interpolação de HRTFs Baseada no VBAP	104
9.5	Conclusões	108
10	Som Tridimensional com Caixas Acústicas	109
10.1	Leis dos Senos e das Tangentes	110
10.1.1	Lei dos Senos	110
10.1.2	Lei das Tangentes	112
10.2	Limitações e Suas Implicações	115
10.3	Modelo Mais Flexível	116
10.4	Conclusões	120
11	Conclusões	122
	Referências Bibliográficas	124
A	Método de Medição de Funções de Transferência	132
B	Superfícies Espaciais de Resposta na Frequência—SFRSs	134
C	Dedução da Estrutura Farrow Bidimensional	137
C.1	Desenvolvimento da Equação (7.13)	137
C.2	Desenvolvimento da Equação (7.14)	138

Lista de Figuras

1.1	Esquema para obtenção da imagem-fantasma	5
1.2	Esquema simplificado da técnica de síntese de campo acústico.	5
1.3	Esquema simplificado para medição de HRTFs.	7
1.4	Verificação da diferença de tempo interaural (ITD).	8
1.5	ITD e ILD ao se variar a elevação da fonte.	9
1.6	Utilização das HRIRs.	10
2.1	Interpolação bilinear direta.	17
2.2	Detalhe das regiões triangulares sobre a esfera de referência.	18
2.3	Efeito da média ponderada sobre HRIRs de fase não-mínima	20
2.4	Resultado da média ponderada sobre HRIRs de fase mínima.	20
2.5	Estrutura da interpolação triangular direta.	21
3.1	Diagrama em blocos da técnica utilizada no cálculo das HRIRs de fase mínima.	26
3.2	Exemplo do efeito obtido com as HRIRs de fase mínima suavizadas	27
3.3	Exemplo de incorporação do atraso à HRIR de fase mínima	28
4.1	Modelo de HRTF: (a) forma direta e (b) baseado na IATF.	32
4.2	Cálculo da HRTF _f via IPTF: interpretação gráfica.	32
4.3	Detalhes da superfície esférica de referência.	33
4.4	Diagrama em blocos do procedimento de interpolação baseado nas IPTFs.	35
4.5	Comparação entre SFRSs. Bilinear, IPTF e diferença em dB—I.	38
4.6	Comparação entre SFRSs. Bilinear, IPTF e diferença em dB—II.	39
4.7	Comparação entre SFRSs. Bilinear, IPTF e diferença em dB—III.	40
4.8	Comparação entre SFRSs. Bilinear, IPTF e diferença em dB—IV.	41

5.1	Definição de vizinhança entre as coordenadas das HRIRs (vista espacial).	48
5.2	Definição de vizinhança entre as coordenadas das HRIRs (vista planificada).	49
5.3	Comparação entre SFRSs. IPTF sem e com redução de ordem, e diferença em dB—I.	51
5.4	Comparação entre SFRSs. IPTF sem e com redução de ordem, e diferença em dB—II.	52
5.5	Comparação entre SFRSs. IPTF sem e com redução de ordem, e diferença em dB—III.	53
5.6	Comparação entre SFRSs. IPTF sem e com redução de ordem, e diferença em dB—IV.	54
5.7	Histograma para o erro ξ (interpolação com IPTFs de ordem reduzida).	54
6.1	Percentual de energia em função do número de autovetores utilizados no modelo.	59
6.2	Diagrama em blocos da representação pela SFER.	59
6.3	Interpretação gráfica do cálculo de uma das SCF.	61
6.4	Exemplo de aproximação de uma SCF por regiões triangulares.	62
6.5	Erro acumulado ao se mudar de região.	63
6.6	Comparação da complexidade computacional com o aumento do número de fontes.	64
6.7	Comparação entre SFRSs. Triangular direta, Incremental e diferença em dB—I.	67
6.8	Comparação entre SFRSs. Triangular direta, Incremental e diferença em dB—II.	68
6.9	Comparação entre SFRSs. Triangular direta, Incremental e diferença em dB—III.	69
6.10	Comparação entre SFRSs. Triangular direta, Incremental e diferença em dB—IV.	70
6.11	Histograma para o erro ξ (interpolação incremental com a KLT).	70

7.1	Exemplos de atraso fracionário e mudança de taxa de amostragem a partir do sinal original.	73
7.2	Diagrama em blocos da estrutura Farrow.	74
7.3	Diagrama em blocos da estrutura Farrow generalizada.	76
7.4	Norma das HRIRs de todas as posições (θ, ϕ) para cada instante. . .	79
7.5	Erro normalizado mínimo obtido com o método iterativo de otimização para cada instante de tempo	80
7.6	Comparação entre SFRSs. Erro devido à ordem dos polinômios da estrutura Farrow	81
7.7	Comparação entre SFRSs. Triangular direta, com a estrutura Farrow e diferença em dB—I.	82
7.8	Comparação entre SFRSs. Triangular direta, com a estrutura Farrow e diferença em dB—II.	83
7.9	Comparação entre SFRSs. Triangular direta, com a estrutura Farrow e diferença em dB—III.	84
7.10	Comparação entre SFRSs. Triangular direta, com a estrutura Farrow e diferença em dB—IV.	85
7.11	Histograma para o erro ξ (interpolação com a estrutura Farrow generalizada).	86
8.1	Comparação das médias obtidas para a nota de diferença entre o método triangular e todos os outros	90
8.2	Sentidos testados na avaliação de percepção de movimento.	90
8.3	Comparação das taxas de acerto do sentido do movimento obtidas para cada método de interpolação	91
8.4	Posições testadas na avaliação de percepção da posição estática. . . .	94
8.5	Comparação das taxas de acerto da posição estática obtidas para cada método de interpolação	95
9.1	Posicionamento baseado na variação de amplitude com 3 caixas acústicas—ouvido esquerdo.	102
9.2	Interpolação de HRIR a partir de 3 funções conhecidas—ouvido direito.	103
9.3	VBAP com três canais.	104

9.4	Geometria para interpolação triangular de HRIRs.	105
10.1	Esquema de sistema estereofônico convencional.	110
10.2	Mudança nas distâncias devido à rotação da cabeça	114
10.3	Detalhe dos ângulos envolvidos na mudança das distâncias devido à rotação da cabeça	114
10.4	Respostas para a posição $(0^\circ, 0^\circ)$	117
10.5	Respostas para a posição $(60^\circ, 0^\circ)$	118
10.6	Respostas para a posição $(120^\circ, 0^\circ)$	119
B.1	SFRSs a partir das respostas em frequências interpoladas.	135

Lista de Tabelas

4.1	Passos de azimute das medidas (ângulos em graus).	36
8.1	Tabela de confusão do teste de percepção do movimento para o método triangular.	92
8.2	Tabela de confusão do teste de percepção do movimento para o método com IPTFs.	92
8.3	Tabela de confusão do teste de percepção do movimento para o método KLT incremental.	93
8.4	Tabela de confusão do teste de percepção do movimento para o método com a estrutura Farrow generalizada.	93
8.5	Tabela de confusão do teste de percepção da posição para o método triangular.	96
8.6	Tabela de confusão do teste de percepção da posição para o método com IPTFs.	96
8.7	Tabela de confusão do teste de percepção da posição para o método incremental com a KLT.	97
8.8	Tabela de confusão do teste de percepção da posição para o método com a estrutura Farrow generalizada.	97
B.1	Descrição das características das SFRSs por faixa de frequências. . . .	136

Lista de Publicações

- [P1] FREELAND, F. P., *Técnicas de Geração de Som Tridimensional*. Tese de mestrado, Universidade Federal do Rio de Janeiro, Programa de Engenharia Elétrica–COPPE, Julho 2001.
- [P2] FREELAND, F. P., BISCAINHO, L. W. P., DINIZ, P. S. R., “Interpositional Transfer Function for 3D-Sound Generation,” *Journal of the Audio Engineering Society*, v. 52, n. 9, pp. 915–930, Sept. 2004.
- [P3] BISCAINHO, L. W. P., FREELAND, F. P., DINIZ, P. S. R., “Using Interpositional Transfer Functions in 3D-Sound.” In: *IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Orlando, Florida, USA, May 2002.
- [P4] FREELAND, F. P., BISCAINHO, L. W. P., DINIZ, P. S. R., “Efficient HRTF Interpolation in 3D Moving Sound. In: *The Proceedings of The AES 22nd International Conference*, Espoo, Finland, June 2002.
- [P5] FREELAND, F. P., BISCAINHO, L. W. P., DINIZ, P. S. R., “Interpolation of Head-Related Transfer Functions (HRTFs): A Multi-Source Approach”. In: *Proceedings of the XII European Signal Processing Conference*, pp. 1761–1764, EURASIP, Vienna, Austria, Sept. 2004.
- [P6] FREELAND, F. P., BISCAINHO, L. W. P., DINIZ, P. S. R., “Interpolação Bilinear Generalizada de HRTFs para Geração de Som Tridimensional. Anais da VIII Convenção Nacional da Audio Engineering Society Brasil, São Paulo, Brasil, Junho de 2004.

Lista de Abreviações

HRTF	Função de Transferência Relativa à Cabeça (<i>Head-Related Transfer Function</i>)
HRIR	Resposta ao Impulso Relativa à Cabeça (<i>Head-Related Impulse Response</i>)
IPTF	Função de Transferência Interposicional (<i>Interpositional Transfer Function</i>)
IATF	Função de Transferência Interaural (<i>Interaural Transfer Function</i>)
SFRS	Superfície Espacial de Resposta na Frequência (<i>Spatial Frequency Response Surface</i>)
SCF	Função Característica Espacial (<i>Spatial Characteristic Function</i>)
SFER	Extração e Regularização das Características Espectrais (<i>Spectral Feature Extraction and Regularization</i>)
KLT	Transformada de Karhunen-Loève (<i>Karhunen-Loève Transform</i>)
FIR	Resposta ao Impulso Finita (<i>Finite Impulse Response</i>)
IIR	Resposta ao Impulso Infinita (<i>Infinite Impulse Response</i>)
DFT	Transformada de Fourier Discreta (<i>Discrete Fourier Transform</i>)
FFT	Transformada Rápida de Fourier (<i>Fast Fourier Transform</i>)

Capítulo 1

Introdução

1.1 Motivação

Nos últimos anos, a recriação de um ambiente virtual realístico tem sido alvo de vários estudos [1, 2]. Muito aplicadas em sistemas de treinamento, como em simuladores de vôo, algumas das técnicas utilizadas pelos sistemas de realidade virtual começaram a se tornar populares com seu emprego em algumas formas de entretenimento.

O cinema pode ser citado como um dos grandes difusores das técnicas de geração de “realidade”, já que nos últimos anos nota-se um aumento na utilização dessas técnicas, principalmente na finalização dos programas de áudio dos filmes. Além disso, já foram criadas salas de projeção tridimensional nas quais duas imagens são projetadas “simultaneamente” e são direcionadas para cada um dos olhos, por exemplo, através de óculos polarizadores [3]. Apesar de, com esse sistema de projeção, criar-se maior imersão no ambiente do filme, essa forma de projeção ainda é de pouca penetração no mercado cinematográfico. Isso já não acontece quando se trata dos sistemas de reprodução de áudio tridimensional. Este sim, por ser um pouco mais simples de ser implementado, foi o que realmente contribuiu para a evolução da forma de preparação do programa áudio-visual.

Essa evolução, que pôde ser notada especialmente no final do século XX, foi motivada, inicialmente, pela necessidade de uma forma de apresentação sonora nos cinemas mais realística e complexa para fazer frente aos sistemas de reprodução doméstica que vinham evoluindo, ficando mais baratos e com desempenho equiva-

lente. Hoje em dia, com a popularização do DVD (*Digital Versatile Disc*), que pode conter vários programas de áudio para as diversas configurações de reprodução, os mercados cinematográfico e de DVDs já vêm beneficiando-se igualmente das evoluções nessa área.

Os jogos eletrônicos [4], que também têm-se popularizado ultimamente, são outro foco da utilização das técnicas para geração de efeitos sonoros. Nesse caso, o realismo e a imersão são essenciais para a aceitação do jogo pelo público, cada vez mais exigente. Claro que os jogos de ação em primeira pessoa¹ e de simuladores são as aplicações diretamente ligadas ao realismo do ambiente, mas mesmo em outros tipos de jogos os efeitos sonoros podem ser aplicados com o intuito de estimular o usuário com diversas sensações.

Além disso, a simulação do ambiente virtual de gravações ao vivo, que podem ou não ser acompanhadas da gravação do vídeo, tem um apelo bastante interessante pois, dessa forma, tem-se a sensação de estar presente no momento da apresentação.

Portanto, para se tentar contemplar todas estas aplicações, as técnicas para processamento do som de forma a criar a sensação do ambiente desejado vêm evoluindo. Essas técnicas podem ser separadas de maneira geral em dois grandes grupos. O primeiro é o grupo responsável pelo posicionamento da fonte sonora no espaço tridimensional, enquanto o segundo responde pela simulação do efeito do ambiente virtual na propagação do som. Nesta tese, o conjunto de técnicas estudado está ligado à localização da fonte, embora seja chamado genericamente como sistema de geração de som tridimensional.

1.2 Geração de Som Tridimensional

Quando se fala em som *surround* pensa-se automaticamente em áudio tridimensional (3D). Tecnicamente falando, os equipamentos de som *surround* não geram o som tridimensional, eles apenas reproduzem simultaneamente um certo número de canais que podem ou não conter informação sonora gravada de maneira especial [5] para dar a sensação do ambiente desejado. Além disso, a forma de gravação

¹Os jogos considerados em primeira pessoa são aqueles em que o jogador tem a visão que o personagem teria.

é direcionada a uma determinada configuração de reprodução.

Em um ambiente controlado, como uma sala de concerto, por exemplo, esse tipo de gravação é bastante plausível, pois as fontes sonoras de interesse estão em um local bem definido e o ruído ambiente é bastante reduzido. No entanto, conseguir realizar tal gravação em ambientes menos controlados, como em cenas externas de filmes, passa a ser um problema mais complexo, que geralmente é solucionado com a captação individual de cada som de interesse seguida de mixagem. E, ainda, tendo-se alguma informação prévia do posicionamento das caixas acústicas de cada canal² [8, 9, 10], pode-se posicionar fontes virtuais no plano das caixas acústicas [11, 12].

Na maioria dos programas de áudio mixados para esses equipamentos (até mesmo o de filmes), o que se observa é uma tentativa de criar um som mais “envolvente”, o que não implica necessariamente a localização de um determinado evento sonoro. São os profissionais de efeitos sonoros que acrescentam alguns sons bem localizados ou com movimento a determinadas cenas.

Atualmente, é neste ponto que tem entrado a geração de áudio 3D, já que criar todo um programa de áudio com várias componentes distintas (voz, som de fundo do ambiente etc.) fica bastante oneroso. Claro que, em alguns casos, como por exemplo, nos shows em DVD, para os quais o som de cada instrumento (ou pelo menos um certo número deles) pode estar gravado separadamente em um dos canais, pode-se criar um cenário virtual de forma mais simples.

No entanto, em outras aplicações como os simuladores (incluindo alguns jogos nesse conjunto), é desejável que a localização da maioria dos sons seja realizada artificialmente e, ainda, em tempo real. Isso porque existe a interação com o usuário, que comanda a cena virtual a todo instante (o movimento da cabeça do ouvinte [13] é um exemplo de interação existente), motivo pelo qual se deve buscar sistemas de baixo custo computacional.

Mas qual é o princípio de funcionamento desse tipo de sistema e como eles geram o som tridimensional? Primeiramente, tem-se que discutir como nós percebemos a posição dos sons.

²Normalmente, os equipamentos de múltiplos canais indicam em seus manuais [6, 7] a distância angular recomendável entre as caixas.

1.3 Percepção da Posição

O sistema auditivo humano é formado pelo conjunto “orelha externa-orelha média-orelha interna”, responsáveis pela captação, condução e transdução do sinal acústico, respectivamente. O ponto relevante para a geração de som 3D é a distinção entre sons vindos de direções diferentes, que está intimamente ligada a como captamos o som. Sabe-se que isso está também relacionado com a geometria do conjunto cabeça, tronco e pina (orelha externa), em cujas superfícies as frentes de onda sonora sofrem reflexões e difrações que são distintas para cada uma das direções de incidência [14, 15]. São essas reflexões e difrações que alteram o espectro do som agregando a ele a informação de direção de incidência posteriormente interpretada pelo cérebro.

Um sistema de geração de áudio 3D tem como objetivo agregar essa noção de direção ao som original. Isso pode ser feito de várias formas diferentes. De maneira geral, as técnicas empregadas diferem entre si principalmente pela forma de reprodução. Algumas usam várias caixas acústicas em canais independentes, geralmente referidas como sistema multicanal, e outras produzem um som que é aplicado em fones de ouvido, chamadas de sistema binaural.

As técnicas multicanal [11, 12, 16] se valem de um conjunto de caixas acústicas posicionadas ao redor do ouvinte para gerar o som 3D. Duas técnicas muito importantes nessa área são a de Deslocamento através de Variação de Amplitude (AP—*Amplitude Panning*) [11] e a de Geração de Campo Acústico [17, 18, 19].

A primeira utiliza a idéia de imagem-fantasma [20] para gerar o efeito de uma fonte virtual no espaço entre as caixas, como pode ser visto na Figura 1.1. Esse método se utiliza da diferença entre as intensidades e atrasos dos sinais aplicados a cada caixa para posicionar a fonte na posição desejada.

Já a segunda, que, em geral, necessita de um número muito maior de caixas acústicas, tem como intenção simular um determinado campo acústico para algumas posições do ambiente real de forma a criar a sensação de se estar imerso no ambiente virtual desejado. Por exemplo, na Figura 1.2 pode-se ver um esquema simplificado do que ocorre nesse tipo de técnica: o sinal aplicado a cada caixa deve ser tal que, combinado com o das outras caixas, componha o sinal que seria captado pelos ouvidos se o som viesse de uma fonte virtual (que pode estar dentro ou fora do

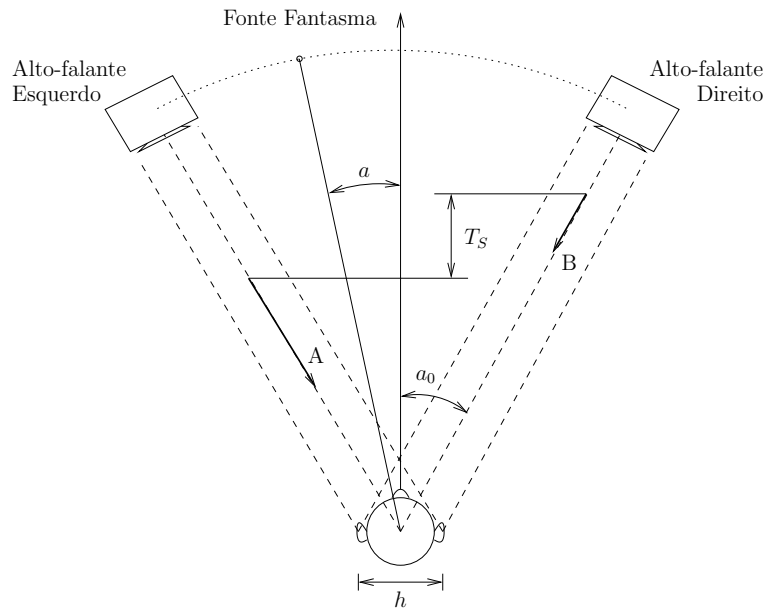


Figura 1.1: Esquema para obtenção da imagem-fantasma. A e B representam os ganhos dos sinais aplicados a cada caixa e T_s é a diferença de atraso.

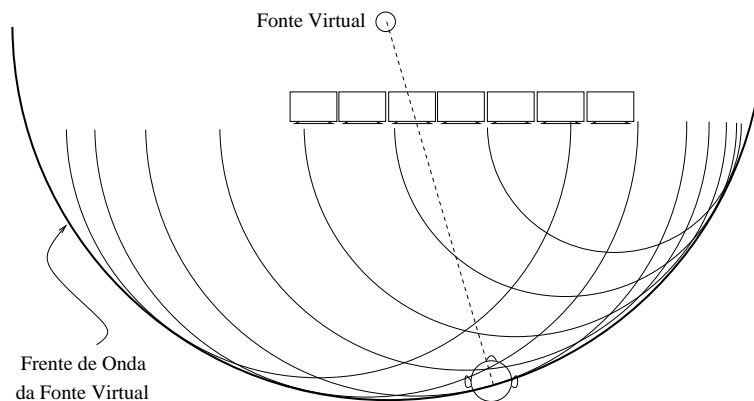


Figura 1.2: Esquema simplificado da técnica de síntese de campo acústico.

ambiente real). Note que, neste caso, é necessário se preocupar com o campo acústico somente nas posições dos ouvidos.

Ambas as técnicas mencionadas são dependentes da posição do ouvinte no ambiente real. Por outro lado, como a forma binaural [14, 21, 15, 22] tem fones de ouvido como transdutores, a posição do ouvinte no ambiente não é crucial como nos casos anteriores³, mas há a necessidade de se dispor de toda a transferência entre a fonte virtual e os ouvidos, incluindo os efeitos de reflexão e difração no conjunto

³Em ambientes imersivos há que se compensar mudança de direção da cabeça do ouvinte. Isso pode requerer um *head-tracker*.

cabeça, tronco e pina. Esse efeito é realizado através das Funções de Transferência Relativas à Cabeça (HRTFs—*Head-Related Transfer Functions*), medidas na forma de suas versões temporais, chamadas HRIRs—*Head-Related Impulse Responses* [23].

De fato, as HRIRs são as responsáveis pela localização da fonte virtual. Normalmente, elas são medidas para várias posições ao redor do ouvinte em ambiente anecóico (livre de reverberação), sempre a uma mesma distância dele, o que significa que as medidas são realizadas para posições sobre uma casca esférica centrada na posição do ouvinte. Elas podem ser medidas utilizando-se pessoas ou um modelo físico médio do conjunto cabeça, tronco e pina (manequim) [24], acoplando-se microfones nos ouvidos. Em ambos os casos, as medidas são realizadas em posições cujas coordenadas são referidas como de azimute e elevação (nessa tese, chamadas de θ e ϕ , respectivamente). O esquema utilizado por Gardner e Martin em [23] para obtenção das HRIRs utilizadas nessa tese é mostrado na Figura 1.3. Naquele trabalho, os autores posicionavam a caixa acústica em cada uma das posições desejadas e emitiam um sinal pseudoaleatório [25, 26] que era gravado através dos microfones acoplados nos ouvidos da cabeça artificial. Após ser gravado, o sinal era processado (veja o Apêndice A), gerando as medidas das HRIRs.

As HRIRs variam de indivíduo para indivíduo. Essas variações são relacionadas com as diferenças na geometria do conjunto cabeça, tronco e pina de cada um. Além disso, a percepção pelos indivíduos dos efeitos gerados com as HRIRs medidas com a cabeça artificial, que representa um indivíduo médio, torna-se também variável. Assim sendo, a adequação individual às sensações transmitidas pelo som gerado artificialmente se faz necessária. No entanto, não demanda muito tempo acostumar-se com o ambiente virtual gerado artificialmente, já que a diferença encontrada acaba por ser compensada após algum tempo de exposição ao ambiente.

1.3.1 Principais Características das HRIRs

Algumas características que indicam ao cérebro a posição da fonte são bem fáceis de serem observadas [P1], verificando-se a diferença existente entre HRIRs de posições diferentes. São elas a diferença de tempo interaural (ITD—*Interaural Time Difference*) e a diferença de nível interaural (ILD—*Interaural Level Difference*). Já o efeito das reflexões no tronco e na pina pode ser observado melhor no domínio da

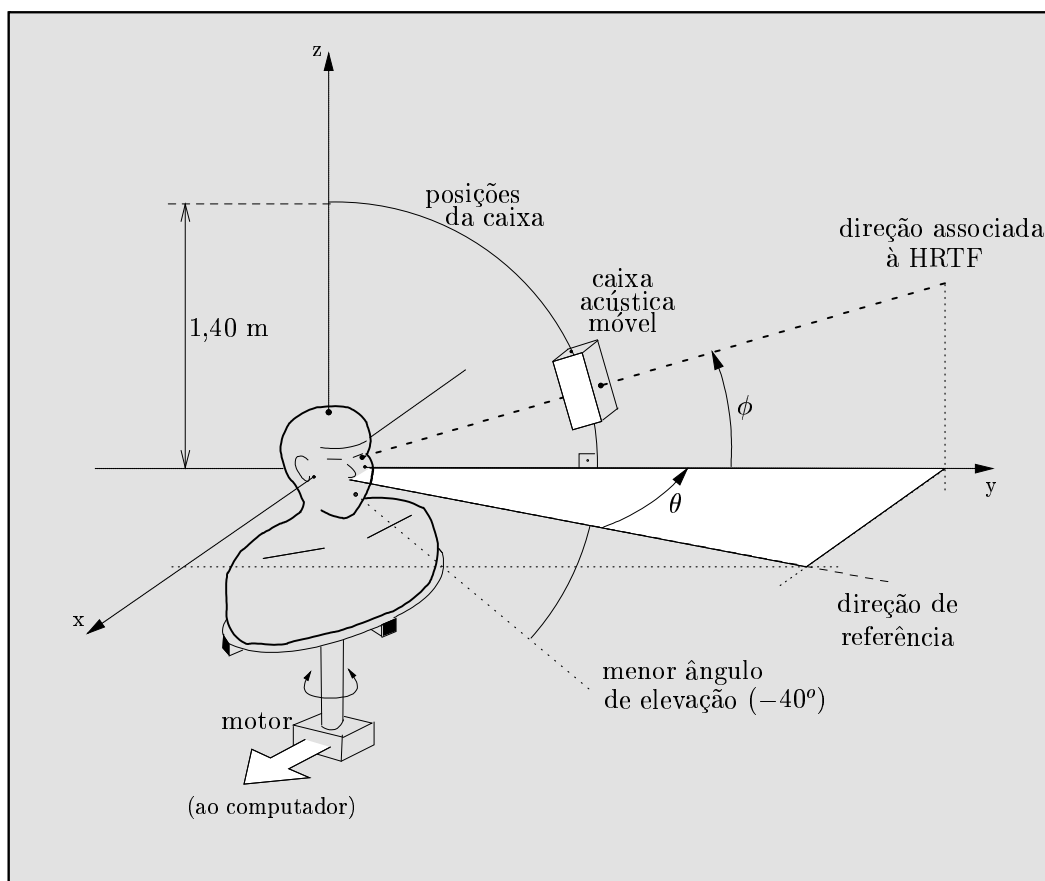


Figura 1.3: Esquema simplificado para medição de HRTFs.

freqüência.

A ITD é o resultado do atraso relativo à diferença de caminho entre a fonte e cada uma das orelhas, que é causada pela distância entre eles e pela difração causada pela cabeça. Como um exemplo, na Figura 1.4, pode-se perceber que a diferença de atrasos existente entre as HRIRs da orelha direita e esquerda para o azimute de 5° (Figura 1.4a) é bem menor que a diferença para o azimute de 45° (Figura 1.4b). Os dois casos mostrados estão na elevação de zero grau. Note que a ITD é bastante relacionada com a variação de azimute, já que para uma mesma elevação e azimutes diferentes, a diferença de atrasos pode mudar muito.

A ILD se caracteriza pela diferença de amplitude do som percebido por cada uma das orelhas. Essa diferença é coerente com a ITD, sendo as duas em conjunto um indicativo do ângulo de azimute da posição da fonte sonora. Na Figura 1.4b pode-se notar a existência da ILD. Comparando-a com a Figura 1.4a, pode-se perceber que, à medida que a fonte vai se deslocando para um dos lados da cabeça, a amplitude

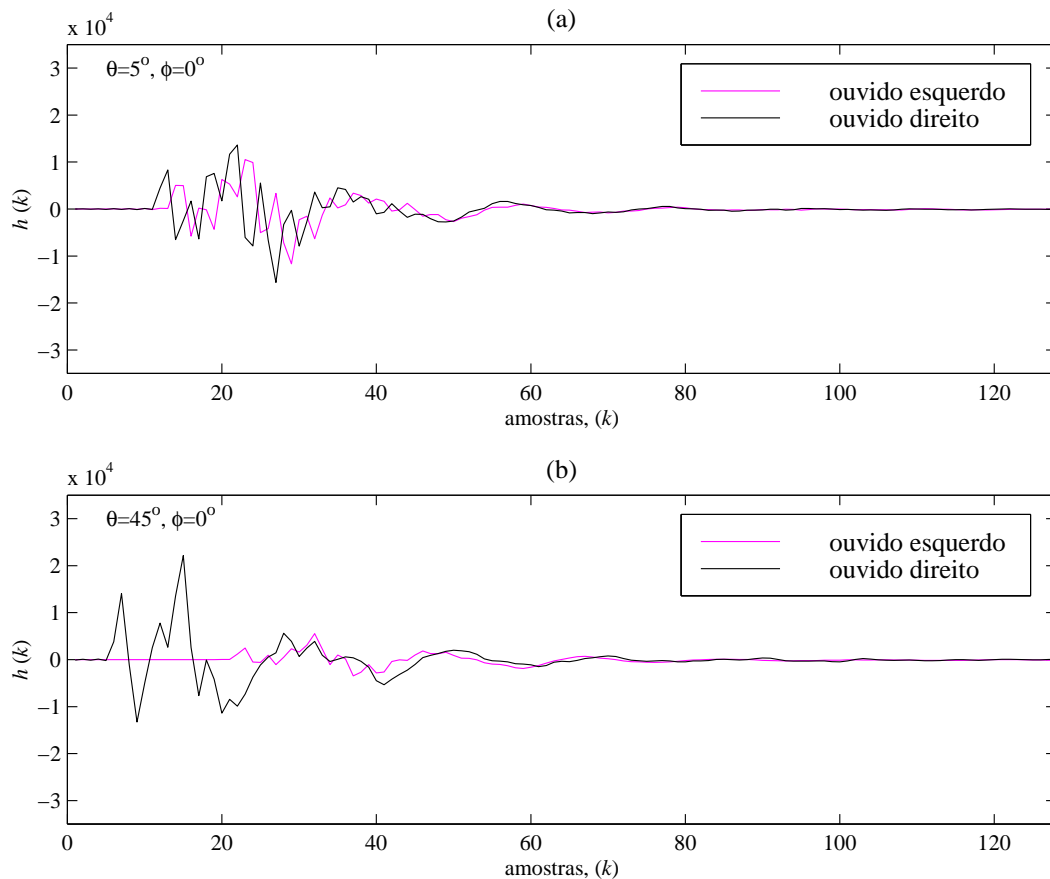


Figura 1.4: Verificação da diferença de tempo interaural (ITD). HRIRs do canal direito e esquerdo na elevação de zero grau e azimutes de 5° (a) e 45° (b).

percebida pelo ouvido do lado oposto diminui. Esse é o efeito chamado de “sombra da cabeça”.

Viu-se que a ITD e a ILD carregam bastante informação sobre o azimute em que se encontra a fonte sonora. No entanto, ao se variar a elevação, essas diferenças não mudam muito. Isso pode ser notado ao se compararem duas HRIRs referentes ao mesmo azimute com elevações diferentes para um mesmo ouvido. Na Figura 1.5a, podem-se observar duas das HRIRs relativas ao azimute zero em elevações diferentes. Como se pode notar na Figura 1.5b, as diferenças são muito mais evidentes no módulo da resposta na frequência. Isso indica que as mudanças das características das HRTFs relacionadas com mudanças de elevação são mais visíveis na resposta à geometria da cabeça artificial, e assim, não só a distância entre os ouvidos (que gera a ITD e parte da ILD) e a sombra da cabeça (que também gera a ILD e modifica um pouco a ITD), mas também as reflexões no tronco e na pina alteram

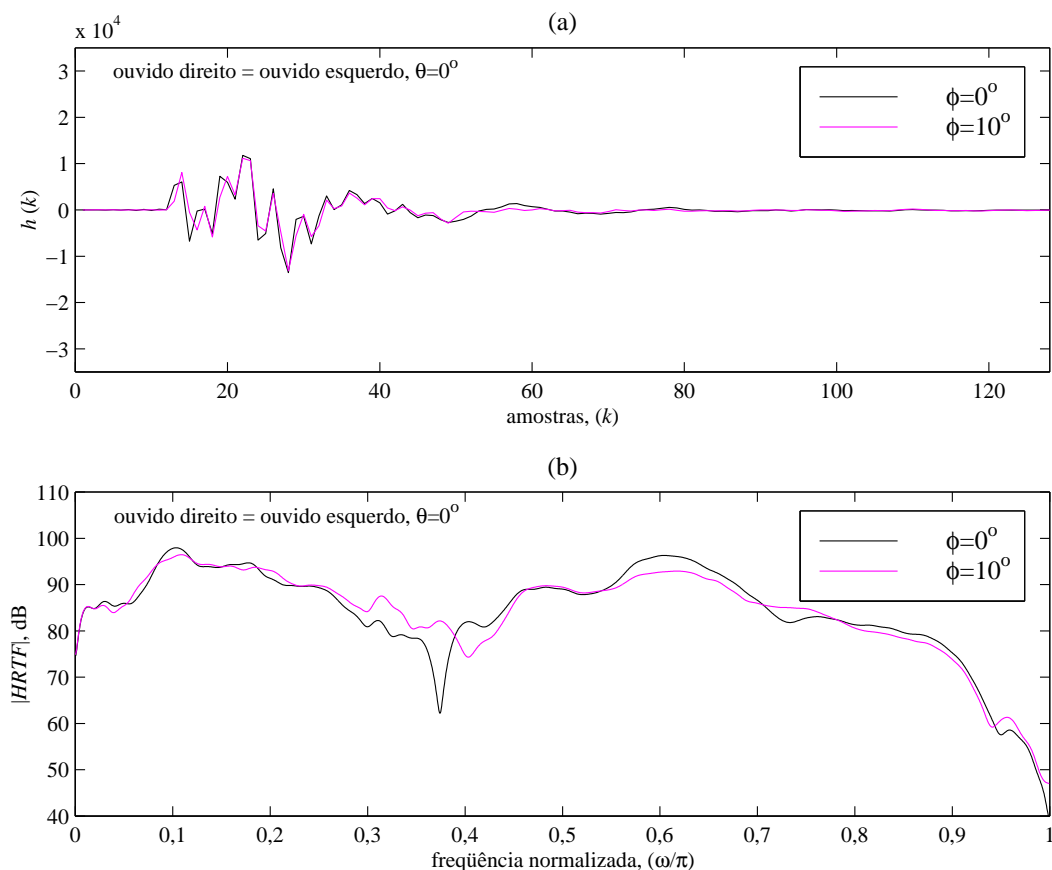


Figura 1.5: ITD e ILD ao se variar a elevação da fonte. Respostas impulsivas relativas ao azimute zero grau (a) e suas respectivas respostas em frequência (b). Essas HRTFs são iguais para o ouvido direito e esquerdo em uma mesma elevação. É possível notar também a importância dos vales e picos para se distinguir em qual elevação se encontra a fonte sonora.

1.3.2 Utilização das HRIRs

No contexto de som tridimensional, as HRIRs são utilizadas como filtros que acrescentam a informação de localização a sinais originalmente monaurais (compostos por um único canal) [14, 21, 15, 27]. Portanto, como pode ser visto na Figura 1.6, cada sinal dá origem a outros dois, alterados pelas HRIRs relativas às orelhas esquerda e direita. Em um modelo simplificado, que considera apenas a localização, o sinal filtrado estéreo é reproduzido com fones de ouvido.

Este procedimento indica a utilização de um único par de HRIRs que tenha sido medido. Assim, para um ponto cujo par de HRIRs seja conhecido, gera-se um

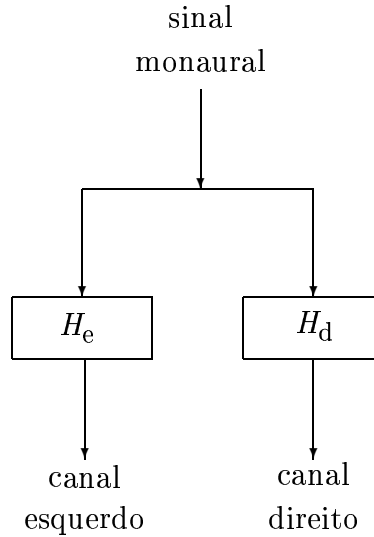


Figura 1.6: Utilização das HRIRs.

som como se a fonte virtual estivesse parada na posição relacionada àquele par.

As formas de posicionamento em qualquer outra localização são tema de estudo da primeira parte desta tese, como é descrito na seção seguinte.

1.4 Conteúdo desta Tese

Esta tese está dividida em duas partes. Na primeira parte, estão os capítulos que tratam de como se pode modelar as transferências desde a fonte virtual até as orelhas, utilizando-se as HRIRs direta ou indiretamente (através de um modelo para o conjunto por elas formado). A segunda parte é referente à utilização de sistemas multicanal para geração de som 3D, sendo lá discutida a formulação empregada nesse tipo de implementação. A seguir, faz-se uma breve descrição de cada capítulo, evidenciando-se as contribuições presentes em cada um deles.

No Capítulo 2, apresenta-se a interpolação de HRIRs diretamente pela ponderação das saídas de um subconjunto delas que define uma região triangular. Nesse capítulo, a maior contribuição fica por conta da formulação dos ponderadores que podem ser utilizados com conjuntos práticos de HRIRs—medidos para posições uniformemente distribuídas em torno do ouvinte, mas com passos fixos para a elevação. Essa formulação é utilizada em vários dos capítulos subseqüentes. Apesar de a formulação da interpolação ter sido mencionada em [22], não foram encontradas em nenhuma outra referência, de forma explícita, as expressões obtidas. Essa formulação

foi descrita em [P2].

No Capítulo 3, é feita uma discussão sobre como pré-processar as HRIRs antes de utilizá-las. São abordadas técnicas para transformar as HRIRs em suas versões de fase mínima e para se fazer a suavização espectral. Dentro desse contexto, uma contribuição desse capítulo está na proposição de utilização da média geométrica ao se fazer a suavização espectral, que pode ser implementada juntamente com o algoritmo que gera as versões de fase mínima das HRIRs. Em [P1], [P3] e [P4] a utilização da média geométrica foi proposta, mas foi em [P2] que ela foi implementada pela primeira vez para todo o conjunto de HRIRs. Além disso, é proposta também uma forma de se aproximar a derivada do excesso de fase entre as HRIRs originais e as funções de fase mínima associadas de forma a possibilitar a recuperação dos atrasos perdidos por elas nesse procedimento.

No Capítulo 4, revisita-se a formulação para interpolação de HRIRs proposta em [P1], que utiliza uma função auxiliar que reduz a complexidade da interpolação. Nesse capítulo, é realizada uma comparação mais cuidadosa entre a interpolação proposta e uma interpolação clássica, de forma a se validar a primeira. É realizado um estudo da complexidade com o intuito de se obter uma ordem máxima para a função auxiliar que ainda permita reduzir a complexidade total. A principal contribuição desse capítulo está na adoção de uma avaliação sistemática [28], que dá uma idéia bem mais abrangente do comportamento da interpolação; com ela foi possível identificar alguns problemas que ocorriam na implementação proposta originalmente em [P1], a qual foi retificada no presente trabalho. Essa comparação também está presente em [P2].

No Capítulo 5, é apresentada a técnica, utilizada também em [P1], para redução de ordem das funções auxiliares mencionadas anteriormente. Além disso, é revisto o procedimento de geração desse conjunto de funções auxiliares com ordem reduzida—levando-se em conta a nova definição de vizinhança adotada entre os pontos para os quais tem-se HRIRs medidas. Outro ponto de interesse nesse capítulo está na comparação feita entre as interpolações realizadas com e sem redução de ordem das funções auxiliares, que resulta na validação da forma de interpolação e da técnica de redução de ordem. Os resultados apresentados nesse capítulo foram também mostrados em [P2].

No Capítulo 6, a transformada de Karhunen-Loève (KLT—Karhunen-Loève *Transform*) é utilizada para representar todo o conjunto de HRIRs. Então, pode-se realizar a interpolação sobre os coeficientes da transformada. Nesse capítulo, propõe-se a utilização de uma interpolação incremental [P5], que reduz a complexidade computacional desse tipo de modelo se comparada com a da interpolação por *splines*. Esse procedimento é a principal contribuição encontrada nesse capítulo.

No Capítulo 7, sugere-se a utilização de um tipo de estrutura proposta por Farrow [29], que pode ter sua transferência controlada através de um parâmetro de entrada. Na realidade, para se tentar modelar o conjunto completo de HRIRs, propõe-se uma generalização da estrutura que a torna dependente das duas variáveis de posição. É realizada uma comparação entre essa forma de interpolação e um método clássico, mostrando que há uma suavização no espectro das HRIRs interpoladas. Esse método ainda deve ser um pouco mais explorado para se tentar obter melhores resultados. A generalização da estrutura Farrow e o estudo da viabilidade da sua utilização como interpolador de HRIRs são as contribuições desse capítulo.

No Capítulo 8, são descritos os testes subjetivos realizados para comparação dos métodos implementados nesta tese, sendo também apresentados os resultados obtidos.

No Capítulo 9, obtém-se uma formulação para a interpolação de HRIRs para uma região triangular qualquer [P6], que se origina no método de Deslocamento por Variação de Amplitude de Base Vetorial (VBAP—*Vector-Based Amplitude Panning*) [11]. Essa analogia se mostra especialmente interessante quando, partindo da expressão mais geral, se obtém a formulação desenvolvida no Capítulo 2. A dedução da formulação geral explicitando a forma de cálculo dos ponderadores, juntamente com a conclusão de que os métodos de cálculo dos ponderadores nos sistemas binaural e multicanal são equivalentes, são as contribuições mais importantes desse capítulo.

Por fim, no Capítulo 10, são rerepresentadas as deduções das Leis dos Senos e das Tangentes, que relacionam os ganhos que devem ser aplicados ao sinal de cada caixa com os ângulos da posição das caixas e o desejado para a fonte virtual. Durante essas deduções, ficam bastante claras as limitações impostas ao se utilizar essas relações em sistemas multicanal. São mostrados alguns exemplos que confirmam

essas limitações. A continuação para o trabalho contido nesse capítulo visa a unificar a formulação relativa às Leis dos Senos e das Tangentes com aquelas utilizadas pelos sistemas VBAP e de interpolação binaural, buscando analisar as diferenças e similaridades entre elas.

Parte I

Interpolação de HRIRs

Capítulo 2

Interpolação de HRIRs

Como foi mencionado anteriormente, para se poder gerar o som tridimensional utilizando fones de ouvido, as mudanças sofridas pelo som ao longo do caminho percorrido desde a fonte até as orelhas podem ser modeladas pelas HRIRs medidas [14, 23].

Em sistemas práticos o conjunto de HRIRs é formado por um número finito de respostas que foram medidas para algumas posições (θ, ϕ) , sendo necessária a interpolação dessas respostas ao impulso para se localizar a fonte virtual em qualquer outra posição. Além disso, para a criação de um ambiente virtual que imite suficientemente bem a realidade é necessário um conjunto grande de HRIRs para não causar descontinuidades ao longo do caminho percorrido.

No caso de uma única fonte virtual¹, um método bastante utilizado para se interpolar as HRIRs é o bilinear [32, 15], que consiste do cálculo da HRIR associada com um dado ponto da esfera de referência como a média ponderada das HRIRs medidas associadas com os quatro pontos mais próximos a ele e que definem uma região que contém a posição desejada.

Na próxima seção, discute-se o método bilinear direto de interpolação como visto na literatura. Na Seção 2.2, faz-se uma generalização da interpolação bilinear para o caso em que as HRIRs são medidas de forma aproximadamente uniforme

¹Alguns métodos, como os descritos em [30, 31], oferecem alternativas ao modelo que se baseia em HRTFs. Esses métodos separam o modelo em duas partes, uma relativa às características espectrais e outra relativa às características espaciais. Essa interpretação pode ser mais eficiente do que usar diretamente as HRIRs no caso de múltiplas fontes, e será abordada no Capítulo 6.

sobre a superfície de uma esfera, gerando espaçamentos angulares variáveis. Algumas considerações são feitas na Seção 2.3.

2.1 Interpolação Bilinear

De posse de um conjunto de HRIRs medidas, uma forma bastante simples de se obter as funções que não foram medidas é a chamada interpolação bilinear direta de HRIRs [15, 32]. Nessa forma de interpolação, a HRIR associada à posição desejada é calculada fazendo-se uma média ponderada das contribuições das HRIRs medidas associadas com as posições mais próximas ao ponto desejado, como ilustra a Figura 2.1.

Assumamos que todas as HRIRs são medidas para pontos de uma superfície esférica de referência centralizada no ponto médio entre as orelhas, com distâncias angulares fixas entre si de θ_{grid} e ϕ_{grid} para o azimute e a elevação, respectivamente. Com relação à Figura 2.1, a aproximação $\hat{h}(k)$ para a HRIR $h(k)$ associada à posição desejada pode ser obtida [33] como

$$\begin{aligned} \hat{h}(k) &= (1 - c_\theta)(1 - c_\phi)h_a(k) + c_\theta(1 - c_\phi)h_b(k) + \\ &+ c_\theta c_\phi h_c(k) + (1 - c_\theta)c_\phi h_d(k), \end{aligned} \quad (2.1)$$

onde $h_a(k)$, $h_b(k)$, $h_c(k)$ e $h_d(k)$ são as quatro HRIRs relativas às posições mais próximas à desejada. Os coeficientes c_θ e c_ϕ podem ser calculados como

$$c_\theta = \frac{C_\theta}{\theta_{\text{grid}}} = \frac{\text{mod}(\theta, \theta_{\text{grid}})}{\theta_{\text{grid}}} \quad (2.2)$$

e

$$c_\phi = \frac{C_\phi}{\phi_{\text{grid}}} = \frac{\text{mod}(\phi, \phi_{\text{grid}})}{\phi_{\text{grid}}}, \quad (2.3)$$

sendo C_θ e C_ϕ as distâncias angulares relativas indicadas na figura e $\text{mod}(a, b)$ é o resto da divisão de a por b .

2.2 Generalização Proposta para Regiões Triangulares

Por razões práticas, um conjunto de HRIRs medidas (como as obtidas por Gardner e Martin [23], utilizadas nessa tese) pode ser obtido para pontos distribuídos

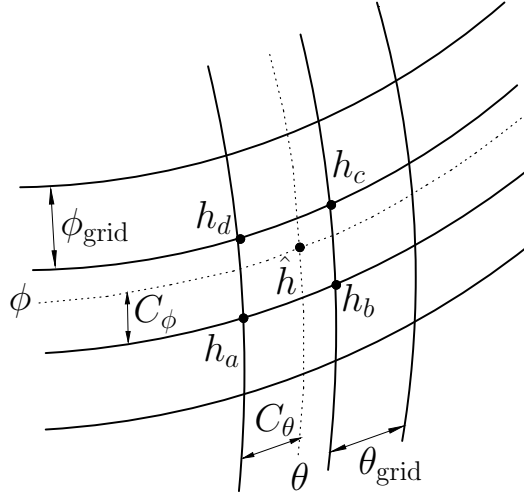


Figura 2.1: Interpolação bilinear direta.

quase homogeneamente sobre a esfera, de forma que a densidade de pontos sobre ela seja aproximadamente constante. Isto implica utilizar-se uma grade com passos angulares variáveis por elevação. Pelo fato de a formulação da Equação (2.1) ser definida para regiões retangulares, para tal conjunto de funções ela não pode ser utilizada.

Para que isso seja possível, propõe-se uma possível generalização para o método bilinear direto [P2, P3]. O procedimento segue a mesma idéia de se ponderar as HRIRs dependendo da sua distância angular ao ponto desejado, mas que utiliza regiões triangulares [22]. Ademais, pode ser utilizado para qualquer conjunto de HRIRs, tendo eles amostragem espacial homogênea ou não. Na Figura 2.2 pode-se ver um setor de uma esfera, sobre a qual foram medidas as HRIRs dos pontos A, B, C e D. Nesse caso, a interpolação da HRIR para o ponto P pode ser realizada fazendo-se

$$\hat{h}_P(k) = w_A h_A(k) + w_B h_B(k) + w_C h_C(k), \quad (2.4)$$

onde

$$w_C = \frac{\Delta\phi}{\Delta\phi_{\text{grid}}}, \quad (2.5)$$

$$w_B = \frac{\Delta\theta}{\Delta\theta_{\text{grid}}}, \quad (2.6)$$

$$w_A + w_B + w_C = 1, \quad (2.7)$$

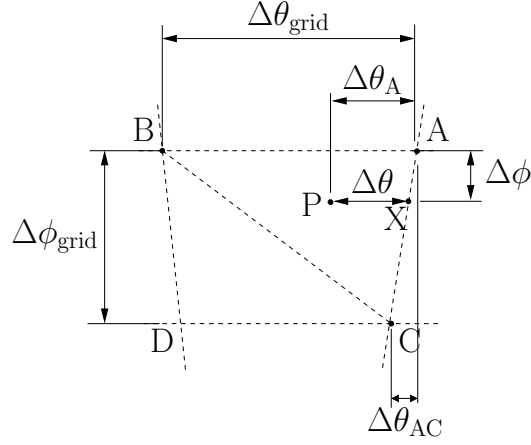


Figura 2.2: Detalhe das regiões triangulares sobre a esfera de referência.

com as distâncias angulares definidas como

$$\Delta\phi = \phi_P - \phi_A, \quad (2.8)$$

$$\Delta\theta = \theta_P - \theta_X, \quad (2.9)$$

$$\Delta\theta_A = \theta_P - \theta_A, \quad (2.10)$$

$$\Delta\theta_{AC} = \theta_C - \theta_A, \quad (2.11)$$

$$\Delta\theta_{\text{grid}} = \theta_B - \theta_A \quad (2.12)$$

e

$$\Delta\phi_{\text{grid}} = \phi_C - \phi_A. \quad (2.13)$$

Como

$$\frac{\Delta\phi}{\Delta\phi_{\text{grid}}} = \frac{\Delta\theta_A - \Delta\theta}{\Delta\theta_{AC}}, \quad (2.14)$$

pode-se obter

$$\Delta\theta = \Delta\theta_A - \frac{\Delta\phi}{\Delta\phi_{\text{grid}}} \Delta\theta_{AC}. \quad (2.15)$$

Note que $\Delta\theta$ é a distância do ponto P até o lado do triângulo que liga as duas elevações a partir do ponto A. É assumido, sem perda de generalidade na prática, que os pontos A e B têm a mesma elevação.

Como já foi mencionado, as HRIRs medidas têm um atraso inicial devido ao tempo de propagação até as orelhas. É importante lembrar que esses atrasos não podem ser desconsiderados, já que a diferença de atraso entre as HRIRs relativas às

duas orelhas (ITD) é crucial para uma boa percepção da direção do som [34]. Apesar disso, a combinação das HRIRs contendo seus atrasos pode gerar uma resposta na frequência com vales que não fazem parte das respostas originais e que, pela similaridade entre HRIRs próximas, não deveriam aparecer na resposta da HRIR desejada. Isso pode ser observado no exemplo da Figura 2.3. Naquela figura, têm-se duas HRIRs adjacentes para a orelha esquerda (Figura 2.3a) e suas respectivas respostas na frequência (Figura 2.3b). Além disso, na Figura 2.3b, é mostrado o módulo da resposta na frequência da média entre as HRIRs da Figura 2.3a, onde se pode perceber a presença dos vales indesejados, além do erro existente na região de frequência mais alta, onde o módulo da média aparece abaixo do valor do módulo das respostas na frequência das HRIRs. Portanto, esse fato constata que os procedimentos de interpolação dados pelas Equações (2.1) e (2.4) devem ser aplicados sobre as versões de fase mínima das HRIRs medidas [35], o que não altera o resultado final se o atraso inicial da HRIR desejada for estimado e utilizado na interpolação como um atraso puro. De fato, o excesso de fase da HRIR original com relação à sua correspondente de fase mínima se aproxima muito de um atraso puro [22]. Então, a estimativa do atraso δ da HRIR desejada pode ser obtido combinando-se apenas as componentes lineares do excesso de cada uma das HRIRs utilizadas na interpolação. Essa estimativa é discutida no Capítulo 3. Na Figura 2.4, pode-se notar que ao se fazer a média sobre as funções de fase mínima o resultado fica mais coerente com o que se espera.

A Figura 2.5 mostra o diagrama em blocos do procedimento de interpolação descrito para um dos canais (esquerdo ou direito) do sistema binaural.

2.3 Conclusões

Foi mostrada nesse capítulo a generalização da formulação da interpolação bilinear para conjuntos de HRIRs medidos para posições uniformemente distribuídas sobre a esfera de referência.

Como essa formulação se utiliza de regiões triangulares, normalmente menores que as retangulares, o resultado da interpolação pode ser considerado mais eficaz, já que no caso retangular uma das funções consideradas está sempre mais distante da

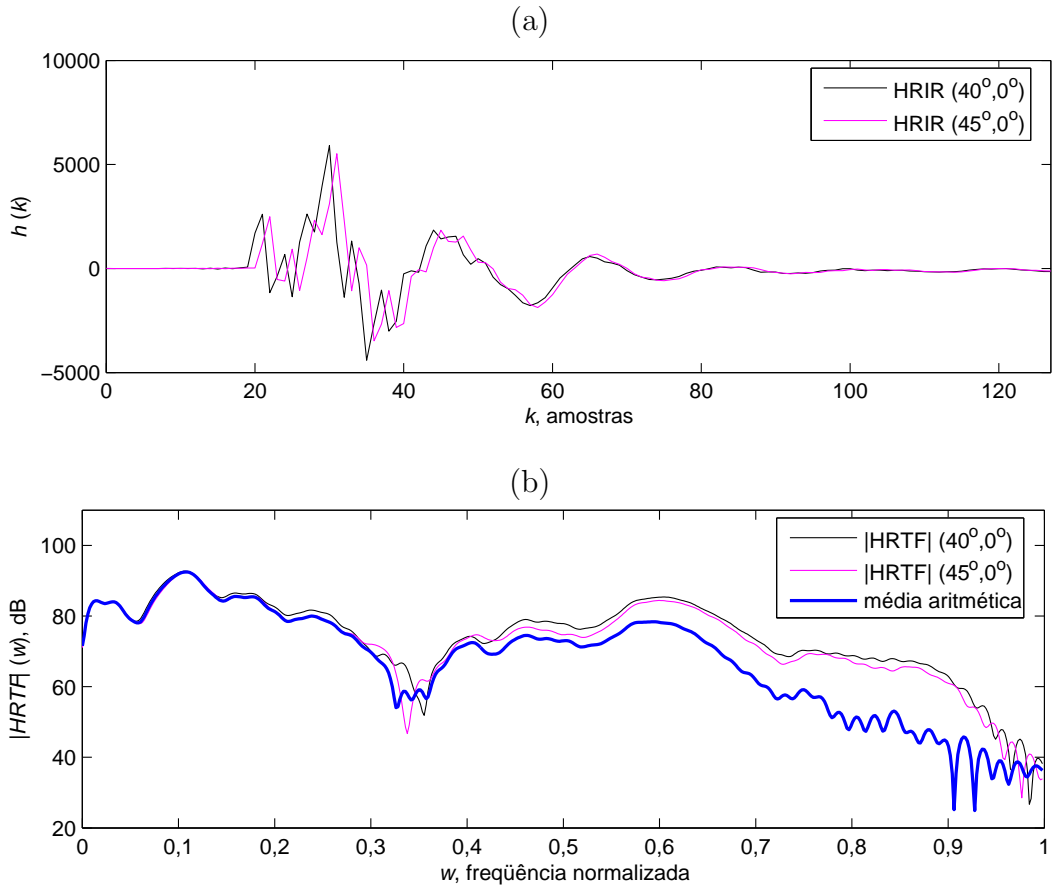


Figura 2.3: Efeito da média ponderada sobre HRIRs de fase não-mínima. (a) HRIRs das posições (θ, ϕ) e (b) respostas na frequência das HRIRs e da média calculada entre elas.

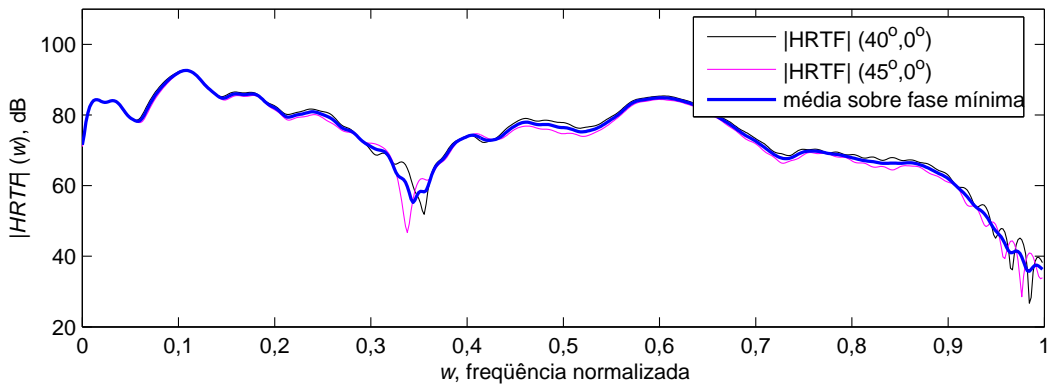


Figura 2.4: Resultado da média ponderada sobre HRIRs de fase mínima.

posição desejada, inserindo no resultado informações menos relevantes a tal posição. Algumas comparações entre esses métodos serão mostradas no Capítulo 4.

Por se utilizarem três ao invés de quatro HRIRs, a complexidade computacional no cálculo das amostras de saída dessa nova forma é menor. Em compensação,

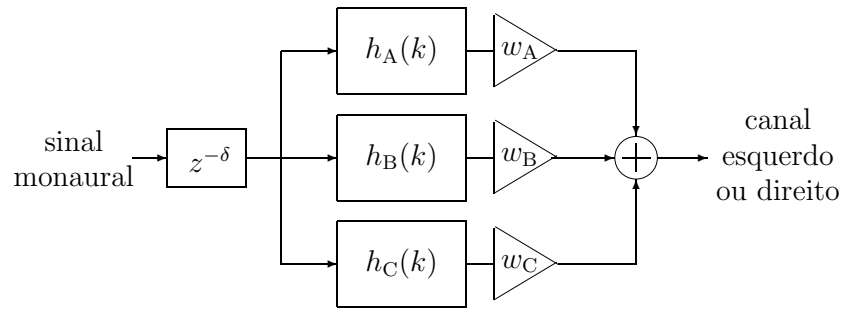


Figura 2.5: Estrutura da interpolação triangular direta.

tende a acontecer um número maior de transições entre regiões diferentes, cujo tempo adicional de cálculo (que inclui principalmente a atualização dos coeficientes dos filtros utilizados) deve ser levado em conta.

No próximo capítulo, fazem-se algumas considerações sobre as HRIRs, além de se mostrarem alguns métodos para o seu pré-processamento, de forma a melhorar o resultado da interpolação.

Capítulo 3

Conformação das HRIRs

Uma vez que as HRIRs são medidas, elas estão sujeitas a imperfeições do sistema de medição, como, por exemplo, as respostas na frequência do amplificador e das caixas acústicas. Esse fato pode fazer aparecerem muitos acidentes indesejáveis na representação espectral dessas respostas ao impulso. Esses acidentes podem atrapalhar o resultado da interpolação, bem como criar algum tipo de erro numérico ao se processarem¹ tais funções.

Na Seção 3.1 desse capítulo, mostram-se algumas formas de se realizar a suavização do espectro das HRIRs. Elas são utilizadas na Seção 3.2 para se obter as versões de fase mínima suavizadas das HRIRs.

3.1 Suavização Espectral

A psicoacústica mostra que a distinção frequencial da audição humana piora com o aumento da frequência. Pode-se supor, por exemplo, que a escala musical logarítmica ocidental é uma consequência desse fenômeno [36, 37]. Além do mais, as medidas das HRTFs tendem a ser menos precisas em altas frequências, onde, graças à atenuação do som, as medidas têm uma razão sinal-ruído mais baixa.

Considerando esses fatos, pode-se pensar em aplicar alguma forma de suavização espectral na resposta das HRIRs de forma que o grau de suavidade aumente

¹No Capítulo 4, é mostrado um tipo de função gerado através das HRIRs. Na geração de tais funções, mostrou-se necessária a suavização espectral das HRIRs para evitar tais problemas numéricos.

com a frequência. Por exemplo, a resposta para cada frequência f pode ser calculada como a média espectral em uma faixa de frequências compreendida entre as frequências

$$f_1 = \frac{f}{\sqrt{K}} \quad \text{e} \quad f_2 = f\sqrt{K}, \quad (3.1)$$

que são equidistantes geometricamente de f pelo fator \sqrt{K} . No domínio da Transformada Discreta de Fourier (DFT—*Discrete Fourier Transform*) com a componente DC indexada como $n = 0$, se f é a frequência associada com a raia espectral n , os índices correspondentes às frequências-limite, f_1 e f_2 , podem ser obtidos por

$$n_1 = \left\lceil \frac{n}{\sqrt{K}} \right\rceil \quad \text{e} \quad n_2 = \lfloor n\sqrt{K} \rfloor, \quad (3.2)$$

respectivamente, denotando $\lceil \cdot \rceil$ como o menor inteiro maior ou igual a (\cdot) e $\lfloor \cdot \rfloor$ como o maior inteiro menor ou igual a (\cdot) . O valor de $K = 2^a$ resulta em uma largura de janela igual a a oitavas; por exemplo, usando $K = 2^{1/3}$ define-se uma largura de um terço de oitava [34, 38].

Dois métodos de suavização são encontrados na literatura: o método da magnitude ao quadrado [34] e o método da resposta na frequência complexa [39]. Em cada caso, a média é calculada sobre as amostras dessas respostas que estão no interior de uma janela deslizante. No método da magnitude ao quadrado, a suavização é obtida calculando-se a média aritmética das amostras da resposta em magnitude elevada ao quadrado no interior da janela e calculando-se a raiz quadrada do resultado, ou seja,

$$|H_s(n)| = \sqrt{\frac{1}{n_2 - n_1 + 1} \sum_{l=n_1}^{n_2} |H_o(l)|^2}, \quad (3.3)$$

onde $H_o(n)$ e $H_s(n)$ são, respectivamente, as respostas na frequência original e suavizada. O método da resposta na frequência complexa leva em conta a informação de fase calculando a média sobre a resposta na frequência complexa, e pode ser expressado como

$$H_s(n) = \frac{1}{n_2 - n_1 + 1} \sum_{l=n_1}^{n_2} H_o(l). \quad (3.4)$$

Em todos os casos, n_1 e n_2 são calculados pela Equação (3.2).

3.1.1 Suavização Usando Média Geométrica

Os dois métodos anteriores baseiam-se em média aritmética, o que resulta em um tratamento diferente para os picos e os vales do módulo das respostas. Uma solução alternativa sugerida em [P1, P4], mas não implementada para todo o conjunto de funções, é trocar a média aritmética pela média geométrica na Equação (3.4). Esses métodos foram comparados em [P3] na suavização das funções interposicionais que serão definidas no Capítulo 4. Pelo fato de essas funções apresentarem uma resposta muito acidentada, os picos apresentavam melhor suavização ao se utilizar a média geométrica. Por esse motivo, ao se pensar na suavização manteve-se a utilização da média geométrica como o padrão. Dessa forma, a suavização com a média geométrica pode ser feita por

$$H_s(n) = \sqrt[n_2 - n_1 + 1]{\prod_{l=n_1}^{n_2} H_o(l)}. \quad (3.5)$$

Aqui, n_1 e n_2 podem ser os mesmos utilizados para as outras formas de suavização.

Pode-se dizer também que a média geométrica aplicada à resposta na frequência complexa pode ser vista como uma média geométrica no módulo, que melhora o tratamento dos picos e vales, e uma média aritmética na fase, mais apropriada.

Por outro lado, um valor de K fixo pode não conseguir atingir a suavidade requerida em altas frequências sem degradar a região de baixas frequências das HRIRs. Para que isso não ocorresse, utilizou-se uma função $K(n)$ que varia linearmente em n , dada por

$$K(n) = 1 - n \frac{1 - \kappa}{L/2 - 1}. \quad (3.6)$$

Nessa equação, L é o tamanho da Transformada Rápida de Fourier (FFT—*Fast Fourier Transform*) utilizada. Nessa tese, foi escolhida uma largura de janela de um sexto de oitava quando $n = L/2 - 1$, fazendo-se $\kappa = 2^{1/6}$.

Essa forma de suavização foi implementada para obtenção do conjunto de HRIRs utilizado nessa tese, com resultados satisfatórios [P2]. Esses resultados serão vistos na Seção 5.3.

3.2 HRIRs de Fase Mínima

Para obter resultados mais consistentes, os métodos de interpolação de HRIRs utilizam as suas versões de fase mínima [14, 38]. Como visto na Figura 2.3, a presença de funções de fase não-mínima pode criar zeros indesejados e erro na resposta da HRIR interpolada. Além disso, pode ocorrer uma certa incoerência entre os atrasos reais provenientes do tempo de percurso do som desde a fonte até as orelhas e os obtidos após a interpolação para a posição desejada, já que a média ponderada é realizada sobre as amostras a cada instante de tempo, e é bastante freqüente o fato de as HRIRs envolvidas na média terem atrasos distintos.

Existem algumas formas de se obter as versões de fase mínima de seqüências finitas. Uma delas [40] gera a seqüência de fase mínima através da transformação inversa do cepstrum complexo $\hat{c}_h(k)$, cuja transformada de Fourier é

$$\hat{C}_h(e^{j\omega}) = \ln |H(e^{j\omega})| + j\angle H(e^{j\omega}). \quad (3.7)$$

Definindo-se o cepstrum real como

$$c_h(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |H(e^{j\omega})| e^{j\omega k} d\omega \quad (3.8)$$

e com base na Equação (3.7), pode-se dizer que $c_h(k)$ é a transformada inversa da parte real de $\hat{C}_h(e^{j\omega})$. Com isso,

$$c_h(k) = \frac{\hat{c}_h(k) + \hat{c}_h^*(-k)}{2}. \quad (3.9)$$

Para seqüências reais tem-se que

$$c_h(k) = \frac{\hat{c}_h(k) + \hat{c}_h(-k)}{2}. \quad (3.10)$$

Uma das propriedades do cepstrum complexo diz que $h(n)$ é de fase mínima se e somente se $\hat{c}_h(k)$ é causal. Como $c_h(k)$ é independente da fase de $H(e^{j\omega})$, considerando $\hat{c}_h(k)$ causal,

$$\hat{c}_h^{\text{mín}}(k) = c_h(k) l_{\text{mín}}(k), \quad (3.11)$$

onde

$$l_{\text{mín}}(k) = \begin{cases} 0, & \text{para } k < 0 \\ 1, & \text{para } k = 0 \\ 2, & \text{para } k > 0. \end{cases} \quad (3.12)$$

Portanto, pode-se obter o cepstrum complexo da versão de fase mínima a partir do cepstrum real da versão de fase não-mínima e, fazendo-se a transformada inversa do cepstrum complexo, obter-se a sequência de fase mínima referente a $h(n)$. Essa técnica é resumida pelo diagrama em blocos da Figura 3.1.

A transformada de Fourier necessária nesse procedimento pode ser aproximada pela DFT. Deve-se apenas tomar o cuidado de se acrescentar um número suficiente de zeros (*zero-padding*) às HRIRs, evitando sobreposição (*aliasing*) no tempo [40].

Na Figura 3.1, $h(n)$ é a HRIR para a qual se deseja obter a versão de fase mínima $h_{\text{mín}}(n)$.

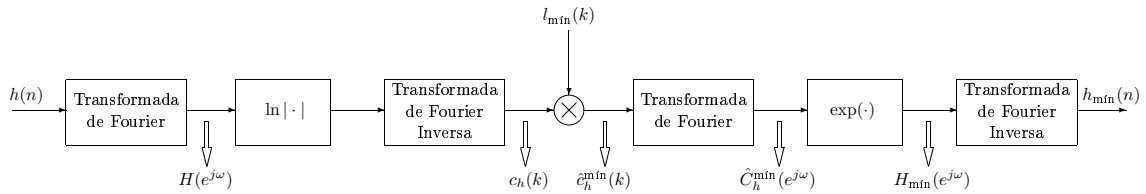


Figura 3.1: Diagrama em blocos da técnica utilizada no cálculo das HRIRs de fase mínima.

3.2.1 Obtenção de HRIRs de Fase Mínima Suavizadas

A operação de suavização indicada pela Equação (3.5) pode ser aplicada ao espectro de cada HRIR, criando a versão suavizada de $H(e^{j\omega})$, antes de se fazer a operação do logaritmo da magnitude (veja Figura 3.1) no processo de transformação delas em suas versões de fase mínima. Desse modo, foram obtidas as versões de fase mínima das HRIRs suavizadas.

Na Figura 3.2, pode-se observar a comparação entre a HRIR original e sua versão de fase mínima suavizada. Na Figura 3.2a, tem-se a comparação das HRIRs e na Figura 3.2b têm-se as respostas em magnitude correspondentes. Pode-se notar que a suavização é bastante coerente. O único fator importante que não é levado em consideração é o atraso inicial. Esse fato será discutido na seção seguinte.

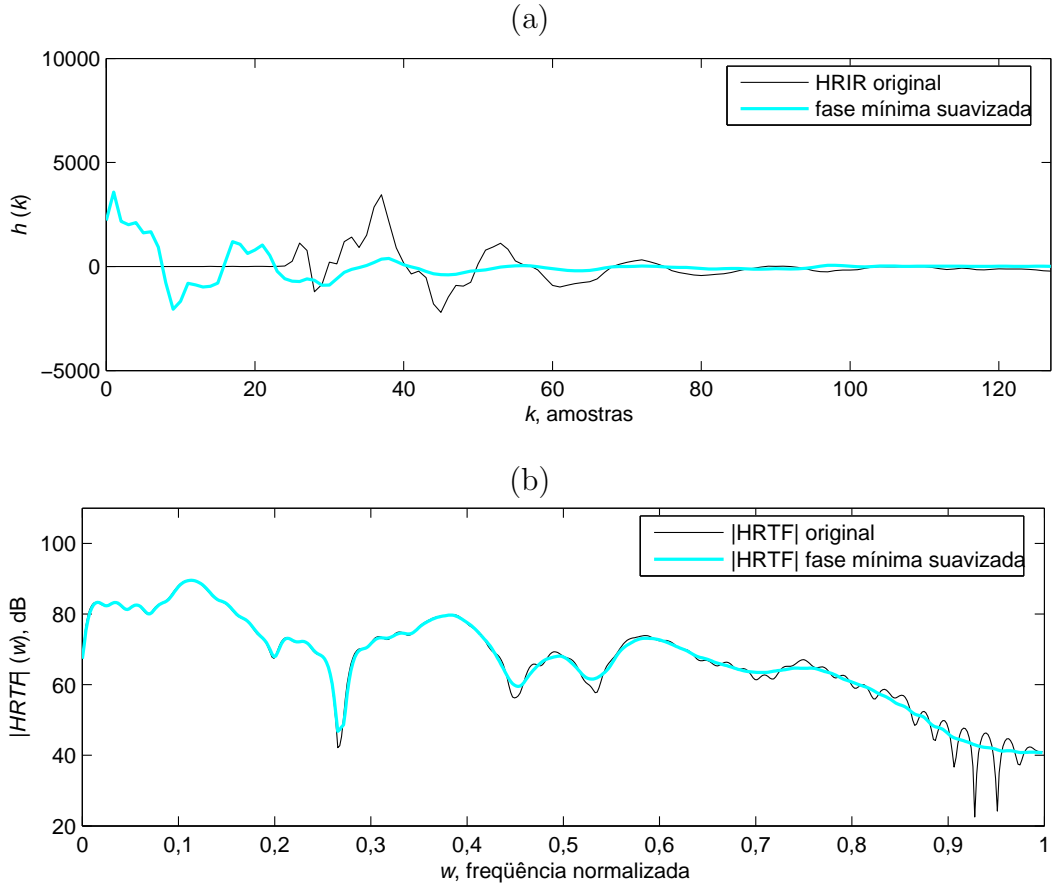


Figura 3.2: Exemplo do efeito obtido com as HRIRs de fase mínima suavizadas. HRIRs para o ouvido esquerdo com $\theta = 70^\circ$ e $\phi = -20^\circ$. (a) HRIRs e (b) suas respostas em magnitude.

3.3 Aproximação das HRIRs originais

Ao se obterem as versões de fase mínima, a informação de fase de cada HRIR original é perdida. Por outro lado, desconsiderando-se o atraso inicial (relacionado com uma componente linear da fase), observa-se que cada HRIR medida pode ser considerada quase de fase mínima [22]. Então, pelo menos esse atraso deve ser resgatado, já que se reflete diretamente na ITD e, portanto, é muito importante no que diz respeito à percepção da posição da fonte virtual [14]. Para recuperar essa informação, estimou-se o atraso de grupo máximo aproximando-se a derivada do excesso de fase por

$$\Delta = \frac{d}{d\omega} \left[\angle \left(\frac{H(e^{j\omega})}{H_{\min}(e^{j\omega})} \right) \right] \approx \min_i \left\{ \angle \left[\frac{H(e^{j\omega_i})}{H_{\min}(e^{j\omega_i})} \right] - \angle \left[\frac{H(e^{j\omega_{i-1}})}{H_{\min}(e^{j\omega_{i-1}})} \right] \right\}, \quad (3.13)$$

onde i é o índice da raia da FFT.

Nessa equação, deve-se notar que as transformadas de Fourier foram aproximadas pela FFT das respostas ao impulso e ω_i e ω_{i-1} são referentes às raiais adjacentes da FFT. Além disso, as fases calculadas no lado direito da equação têm sinal negativo, e por isso o atraso estimado deve ser o mínimo das diferenças.

Na Figura 3.3 mostram-se as mesmas HRIRs da Figura 3.2a, só que agora incorporando-se o atraso calculado pela Equação (3.13). Nota-se que, nesse exemplo, o atraso foi estimado com sucesso. Pode-se notar que existe uma certa diferença de fase próxima a 180° entre as respostas. Isso ocorre entre todas as HRIRs originais e as de fase mínima correspondentes. Todas as HRIRs originais têm um deslocamento de fase 180° desde $\omega = 0$, e essa informação se perde ao serem computadas as versões de fase mínima. Como isso ocorre com todas as HRIRs, nenhum problema é causado se esses 180° forem ignorados na combinação dessas funções.

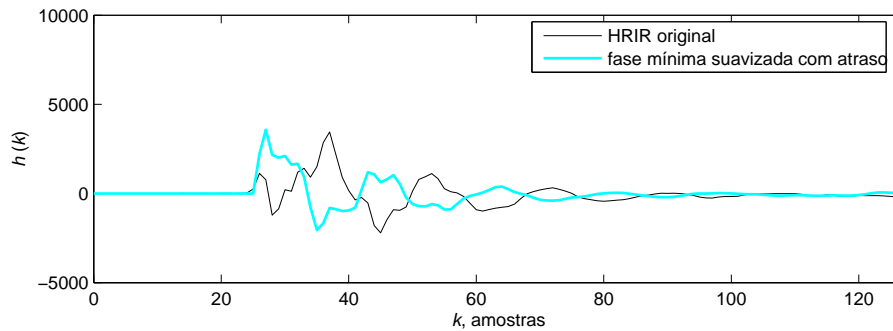


Figura 3.3: Exemplo de incorporação do atraso à HRIR de fase mínima. HRIRs para o ouvido esquerdo com $\theta = 70^\circ$ e $\phi = -20^\circ$.

3.4 Conclusões

Tendo em vista maior eficácia na interpolação de HRIRs, mostraram-se nesse capítulo os procedimentos realizados no pré-processamento do conjunto de HRIRs medidas.

Primeiramente, descreveram-se as técnicas normalmente utilizadas para suavização espectral, seguidas da forma proposta através da média geométrica. Depois, foi mostrada a técnica utilizada para se obter as versões de fase mínima das HRIRs. Propôs-se, então, uma forma de se introduzir suavização nessa técnica.

Além disso, foi indicada uma forma da estimação do atraso inicial da HRIR

original baseada na diferença entre as fases das HRIRs de fase mínima suavizadas e medidas (excesso de fase). Esses atrasos foram utilizados nessa tese para se obter o conjunto de HRIRs de fase mínima suavizadas.

Capítulo 4

Utilização de IPTFs

Os métodos de interpolação de baixa complexidade computacional, que têm sido alvo de muitos trabalhos [41, 34, 42, 43], usualmente focam na redução da ordem das HRTFs. Raramente encontra-se um método que altere também o procedimento de interpolação. Em sua dissertação de Mestrado [P1], o autor desenvolveu uma forma de interpolação alternativa que, com alguns ajustes e desdobramentos, foi posteriormente publicada em [P2, P3]. Essa nova forma, similar ao método de interpolação bilinear direto, altera o esquema de interpolação baseando-se na definição de uma função de transferência auxiliar chamada Função de Transferência Interposicional (IPTF—*Inter-positional Transfer Function*).

Desde que as IPTFs possam ser representadas por modelos de ordem reduzida sem implicar perdas substanciais na acurácia do modelamento, o método proposto atinge uma complexidade computacional bastante reduzida. Neste capítulo, é mostrado como se podem obter e utilizar as IPTFs para a interpolação das HRTFs.

Na Seção 4.1 definem-se as IPTFs em função das HRTFs medidas. Na Seção 4.2, apresenta-se em detalhes o procedimento de interpolação baseado em IPTFs. Comparações entre o desempenho do novo método e o da interpolação bilinear são mostradas na Seção 4.3.

4.1 *Inter-positional Transfer Functions*

4.1.1 Motivação

Para cada posição existe um par de HRIRs, uma para cada uma das orelhas. Na Figura 4.1a, as respostas que representam os caminhos percorridos pelo som da fonte até os ouvidos mais próximo e mais afastado são chamadas de HRTFs ipsi- e contra-laterais, respectivamente. A Figura 4.1b mostra uma representação alternativa para o mesmo par de HRTFs proposta em [43]. Naquele trabalho, os autores mostram que é possível representar um par de HRTFs ipsi- e contra-laterais relativas a uma determinada posição como mostra a Figura 4.1b, onde a razão entre as HRTFs contra- e ipsi-lateral é chamada de Função de Transferência Interaural (IATF¹—*Interaural Transfer Function*). Por ser menos influenciada pela difração e sombra da cabeça, a HRTF ipsi-lateral tem sua energia melhor distribuída ao longo da frequência, podendo assim ser representada por um modelo de ordem menor que a contra-lateral. Por esse motivo, ela é escolhida para ser modelada diretamente. A razão entre as HRTFs ipsi- e contra-lateral modela a diferença entre elas e, por essa diferença não ser grande, o modelo acaba por ter ordem menor que a necessária para se modelar a contra-lateral diretamente. Assim, a configuração baseada na IATF resulta numa realização mais eficiente da HRTF contra-lateral.

4.1.2 Definição

Intuitivamente, pode-se esperar que HRTFs de posições próximas para a mesma orelha sejam mais parecidas entre si do que as HRTFs ipsi- e contra-lateral. Essa observação dá a idéia de se definir uma Função de Transferência Interposicional (IPTF—*Interpositional Transfer Function*) que permita realizar o cálculo da HRTF associada a uma determinada posição como a cascata da HRTF de uma posição adjacente com uma função de erro incremental de ordem reduzida.

A IPTF pode ser definida como

$$\text{IPTF}_{i,f} = \frac{\text{HRTF}_f}{\text{HRTF}_i}, \quad (4.1)$$

¹No trabalho original, a *Interaural Transfer Function* é chamada de ITF. Aqui, a notação IATF é adotada para evitar ambigüidade.

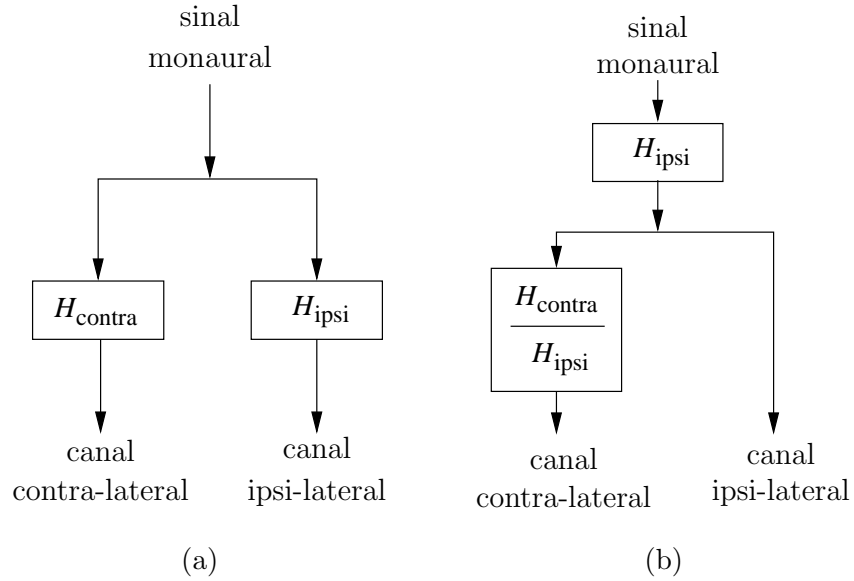


Figura 4.1: Modelo de HRTF: (a) forma direta e (b) baseado na IATF.

onde $HRTF_i$ e $HRTF_f$ são as HRTFs associadas com os pontos inicial e final, respectivamente, como mostra a Figura 4.2. Deve-se notar que todos os vetores estão desenhados com o sentido oposto ao sentido do caminho percorrido pelo som para facilitar a compreensão da geração de $HRTF_f$ partindo da $HRTF_i$.

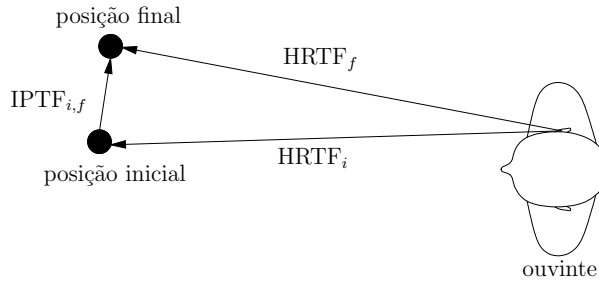


Figura 4.2: Cálculo da $HRTF_f$ via IPTF: interpretação gráfica.

Para garantir a estabilidade das IPTFs, basta que a $HRTF_i$ seja de fase mínima. Isso é satisfeito ao se fatorarem todas as HRTFs de forma aproximada como a cascata de suas versões de fase mínima com um atraso puro, como visto na Seção 2.2.

4.2 Interpolação de HRTFs Baseada nas IPTFs

O mesmo princípio básico utilizado pelo método triangular é aplicado pela interpolação baseada nas IPTFs. No primeiro caso, três HRTFs são utilizadas para se

calcular uma outra. No segundo caso, duas delas podem ser aproximadas usando-se IPTFs de ordem reduzida, o que contribui para a redução da complexidade computacional do procedimento de síntese. Essa é a configuração implementada utilizando regiões triangulares.

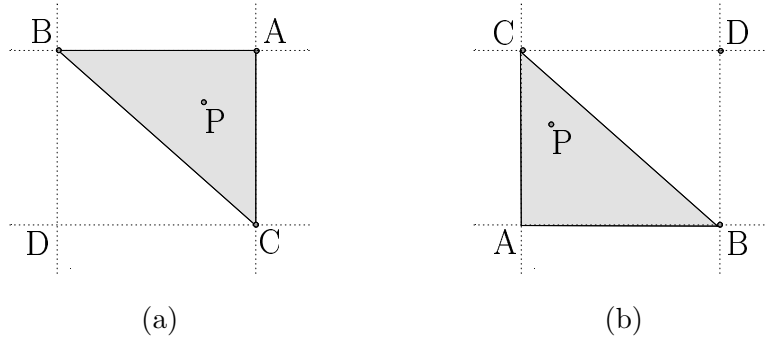


Figura 4.3: Detalhes da superfície esférica de referência: A, B, C, e D têm HRTFs conhecidas, e P é a posição para a qual se quer calcular a HRTF aproximada.

Uma vez que a região triangular foi escolhida, a HRTF, para um dado ponto P dentro da região, pode ser calculada usando-se as HRTFs relativas aos três vértices da região. A Figura 4.3 mostra duas situações diferentes envolvendo a mesma região quadrangular da grade esférica para a qual estão disponíveis as HRTFs medidas para os pontos A, B, C e D, quando se quer estimar as HRTFs relativas ao ponto P. Pode ser visto na Figura 4.3a que, para uma dada posição P ser obtida, os vértices da região triangular a ser utilizada são A, B e C, com A e B tendo a mesma elevação. As coordenadas do ponto mais próximo a P (nesse exemplo, A) serão associadas à versão de fase mínima da HRTF originalmente medida, enquanto que as coordenadas dos outros dois pontos serão associadas com as HRTFs aproximadas por IPTFs. Portanto, as HRTFs relativas ao ponto P podem ser estimadas através da ponderação das HTRFs associadas com A, B e C. É importante notar que para todos os pontos dentro da região sombreada a interpolação se dá utilizando as HRTFs relacionadas com esses três pontos, duas delas aproximadas e uma original. Outro exemplo é mostrado na Figura 4.3b, onde outro ponto P está dentro de uma região diferente. Deve-se notar que os vértices são renomeados para manter a referência aos vértices de mesma elevação da região sombreada como A e B e, assim, preservar a coerência com a Equação (2.4). Nessa nova figura, o ponto mais próximo a P é o vértice C, que será associado à versão de fase mínima da HRTF original.

Seguindo a nomenclatura de vértices explicada anteriormente, as HRTFs relativas ao ponto P, como na Equação (2.4), podem ser descritas por

$$\text{HRTF}_P = w_A \text{HRTF}_A + w_B \text{HRTF}_B + w_C \text{HRTF}_C, \quad (4.2)$$

onde os pesos w_A , w_B e w_C são calculados pelas Equações (2.5)–(2.7).

Chamando de I o ponto mais próximo de P e de F_θ e F_ϕ os outros dois vértices da região que contém P, a interpolação baseada nas IPTFs pode ser escrita como

$$\begin{aligned} \text{HRTF}_P &= \alpha \text{HRTF}_I + \beta \text{HRTF}_{F_\theta} + \gamma \text{HRTF}_{F_\phi} = \\ &= \text{HRTF}_I (\alpha + \beta \text{IPTF}_{I,F_\theta} + \gamma \text{IPTF}_{I,F_\phi}), \end{aligned} \quad (4.3)$$

onde IPTF_{I,F_θ} e IPTF_{I,F_ϕ} são definidas pela Equação (4.1). Os coeficientes α , β e γ são relativos aos vértices I, F_θ e F_ϕ , respectivamente. Por exemplo, na Figura 4.3a, A, B e C são equivalentes a I, F_θ e F_ϕ , respectivamente; dessa forma, a Equação (4.3) fica na forma

$$\begin{aligned} \text{HRTF}_P &= \alpha \text{HRTF}_I + \beta \text{HRTF}_{F_\theta} + \gamma \text{HRTF}_{F_\phi} = \\ &= w_A \text{HRTF}_A + w_B \text{HRTF}_B + w_C \text{HRTF}_C = \\ &= \text{HRTF}_A (w_A + w_B \text{IPTF}_{A,B} + w_C \text{IPTF}_{A,C}). \end{aligned} \quad (4.4)$$

Na Figura 4.4, o diagrama em blocos ilustra o procedimento de interpolação para os dois canais, esquerdo e direito (identificados pelos índices “e” e “d”, respectivamente). Nesse diagrama, os blocos das HRTFs e das IPTFs representam as funções de transferência de fase mínima (ou seja, sem os atrasos inerentes a cada uma delas). Em um procedimento de interpolação simples, os atrasos de $\text{HRTF}_I^{e,d}$, denominados $D_I^{e,d}$, são combinados com os das IPTFs, denominados $d_{I,F_\theta}^{e,d}$ e $d_{I,F_\phi}^{e,d}$, para interpolar separadamente os atrasos das HRTFs desejadas [22]. Deve-se notar que os atrasos de uma IPTF podem ser obtidos através das diferenças dos atrasos das funções HRTF_f e HRTF_i que a geraram. Os atrasos resultantes, dados por

$$\Delta^{e,d} = \alpha D_I^{e,d} + \beta (D_I^{e,d} + d_{I,F_\theta}^{e,d}) + \gamma (D_I^{e,d} + d_{I,F_\phi}^{e,d}) = \quad (4.5)$$

$$= (\alpha + \beta + \gamma) D_I^{e,d} + \beta d_{I,F_\theta}^{e,d} + \gamma d_{I,F_\phi}^{e,d} = \quad (4.6)$$

$$= D_I^{e,d} + \beta d_{I,F_\theta}^{e,d} + \gamma d_{I,F_\phi}^{e,d}, \quad (4.7)$$

podem ser implementados usando-se filtros passa-tudo de atraso fracionário [44] ou então truncados e representados por uma linha de atrasos.

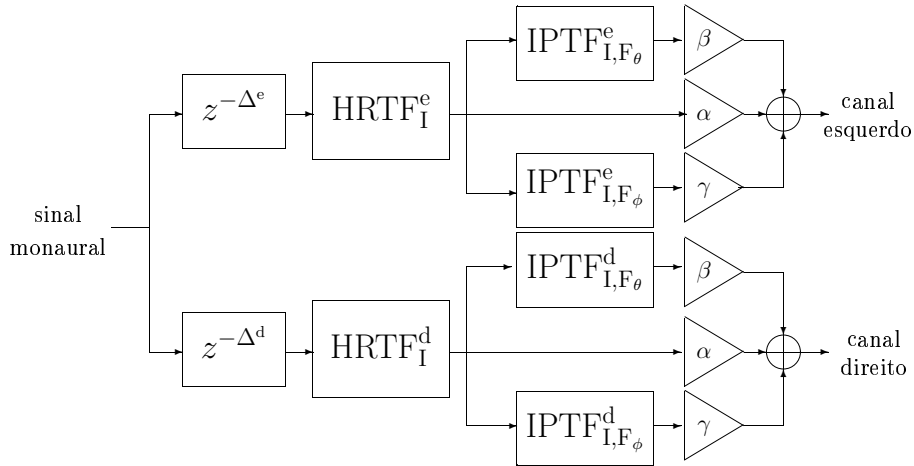


Figura 4.4: Diagrama em blocos do procedimento de interpolação baseado nas IPTFs.

Deve-se perceber também que o uso de IPTFs de ordem baixa é que realmente permite reduzir a complexidade computacional atingida pelo método de interpolação baseado nas IPTFs. Mas o fato de a IPTF ser uma razão entre duas HRTFs muito próximas faz acreditar que é mais simples reduzir a ordem das IPTFs do que das HRTFs diretamente. A redução de complexidade será discutida com detalhes no Capítulo 5.

4.3 Comparações entre as Interpolações Bilinear e Baseada em IPTFs

Na comparação de desempenho, a interpolação baseada nas IPTFs é confrontada com o método bilinear direto através da comparação das suas respostas na frequência para uma faixa de valores de azimuth e elevação.

4.3.1 Configuração dos Testes e Resultados

Inicialmente, somente diferenças na geometria e no cálculo dos pesos são consideradas, isto é, a interpolação baseada nas IPTFs leva em consideração as IPTFs da equação (4.1) como a razão direta entre as HRTFs correspondentes sem considerar simplificação alguma, o que pode ser considerado igual à interpolação

Tabela 4.1: Passos de azimute das medidas (ângulos em graus).

Elevação (ϕ)	Resolução de Azimute (θ_{grid})
-20 a 20	5,00
-30 e 30	6,00
-40 e 40	6,43
50	8,00
60	10,00
70	15,00
80	30,00

triangular aplicada diretamente sobre as HRTFs, na forma

$$\text{HRTF}_P = \alpha \text{HRTF}_I + \beta \text{HRTF}_{F_\theta} + \gamma \text{HRTF}_{F_\phi}. \quad (4.8)$$

A idéia dos testes realizados consiste em comparar as Superfícies Espaciais de Resposta na Frequência² (SFRSs—*Spatial Frequency Response Surfaces*) [28, 45] dos resultados para ambos os métodos de interpolação. Cada SFRS mostra, para uma dada frequência, a dependência da resposta em magnitude com as mudanças na elevação e no azimute.

Já que as equações da interpolação bilinear direta assumem uma grade igualmente espaçada em ϕ e θ , a faixa de valores para a elevação que poderá ser utilizada no decorrer dos testes é restrita a $-20^\circ < \phi < 20^\circ$, de forma a se adequar ao conjunto de HRTFs disponível (veja Tabela 4.1). Esse procedimento foi adotado para que se tivesse uma comparação justa entre os dois métodos.

Todas as simulações nessa tese usam um conjunto de HRTFs de comprimento igual a 128 amostras com frequência de amostragem de 44,1 kHz, obtidas de [23]. Nesse conjunto existem HRTFs para elevações espaçadas de 10° e passos variáveis em azimute segundo a Tabela 4.1.

As Figuras 4.5–4.8 mostram uma comparação visual entre os resultados dos métodos bilinear direto e baseado na IPTF. Na primeira e segunda colunas estão, respectivamente, as SFRSs relativas à interpolação bilinear ($|H_B(\omega, \theta, \phi)|$) e à baseada na IPTF ($|H_{\text{IPTF}}(\omega, \theta, \phi)|$). Escalas distintas de amplitude foram adotadas

²Para um melhor entendimento sobre as SFRSs, veja o Apêndice B.

de acordo com a faixa dinâmica encontrada para cada frequência escolhida. Nelas, somente as respostas para o ouvido esquerdo são consideradas, já que, por simetria, as funções para o outro lado podem ser obtidas invertendo-se os eixos horizontais. As raias frequenciais foram escolhidas segundo o mesmo critério adotado em [28] (Apêndice B).

Para facilitar a comparação, tem-se a terceira coluna contendo a superfície de diferença em dB entre as SFRSs da mesma linha dada por

$$\xi(\omega, \theta, \phi) = 20 \log_{10} \left(\frac{|H_{\text{IPTF}}(\omega, \theta, \phi)|}{|H_{\text{B}}(\omega, \theta, \phi)|} \right). \quad (4.9)$$

Diferentemente das SFRSs, onde as barras laterais mostram a escala em tons de cinza indicando a amplitude em dB, nas figuras de diferenças as regiões mais claras indicam diferenças pequenas (próximas a zero dB) entre as SFRSs. Deve ser notado que para todas as figuras de diferenças foi adotada a mesma escala para melhor comparação entre elas.

Apesar da ocorrência de diferenças um pouco maiores no entorno do ângulo de azimute de 90° (região mais afetada pela difração na cabeça), a equivalência entre os dois métodos pode ainda ser confirmada, já que mesmo essas diferenças observadas não são tão grandes. Pode ser também notado que os piores casos estão associados às regiões de baixa energia das SFRSs.

4.3.2 Complexidade Computacional

Claro que o uso direto das IPTFs obtidas como a razão das HRTFs correspondentes, sem simplificação alguma, não resulta em ganho algum na complexidade computacional. Pelo contrário, até aumentará a complexidade.

Para se ter uma idéia de quanto deve ser a redução de ordem necessária para que se justifique o uso das IPTFs, assumamos que cada HRTF usada diretamente no procedimento de interpolação necessita de N mpa (multiplicações por amostra).

A complexidade computacional da interpolação bilinear direta é, então,

$$C_{\text{bilinear}} = 4N + 12 \text{ mpa}. \quad (4.10)$$

Por outro lado, se cada IPTF aproximada requerer M mpa, o esquema de interpolação baseado nas IPTFs requererá

$$C_{\text{IPTF}} = 2M + N + 6 \text{ mpa}. \quad (4.11)$$

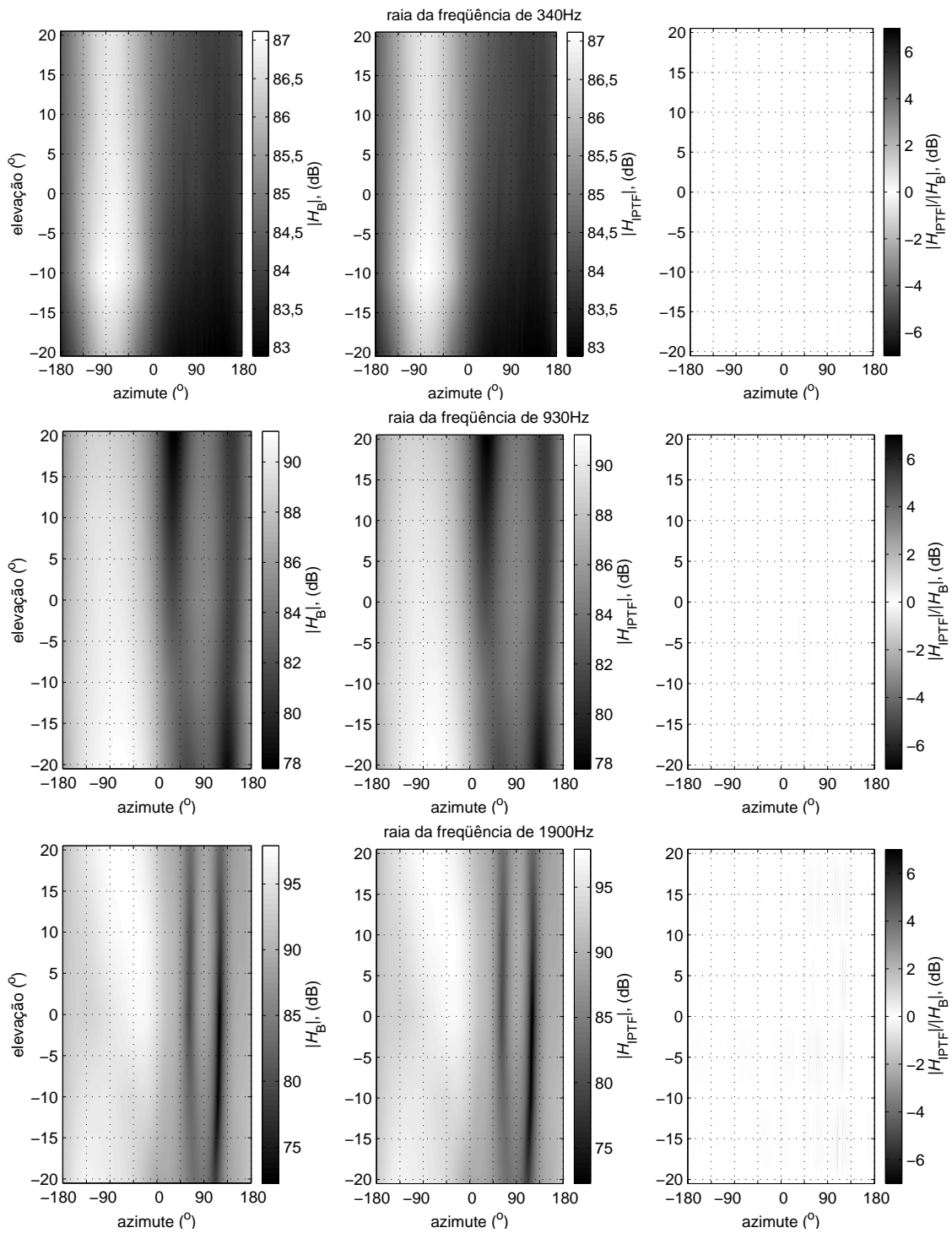


Figura 4.5: Comparação entre SFRs. Bilinear (esquerda), IPTF (centro) e diferença em dB (direita)—I. De cima para baixo: 340 Hz, 930 Hz e 1900 Hz.

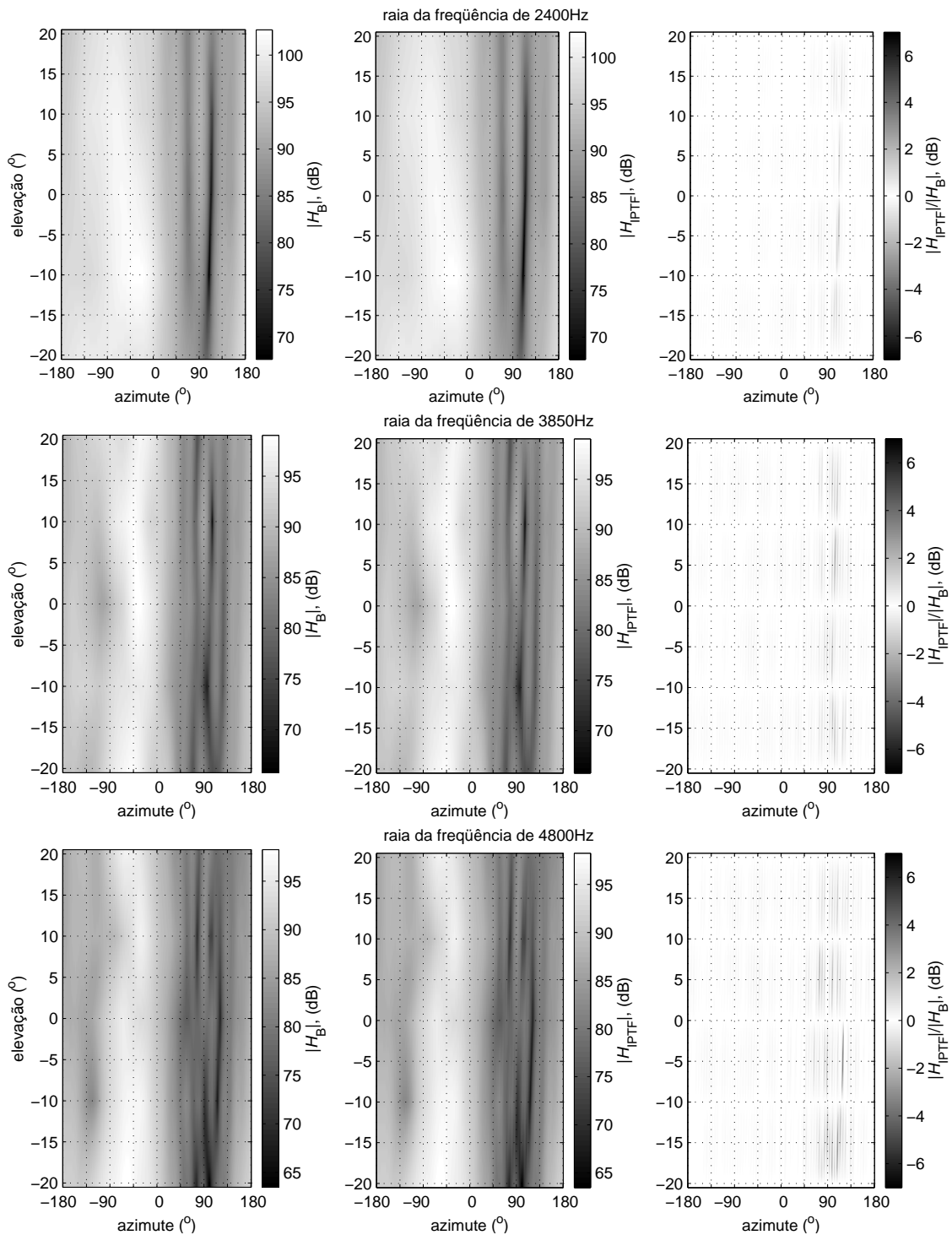


Figura 4.6: Comparação entre SFRSs. Bilinear (esquerda), IPTF (centro) e diferença em dB (direita)—II. De cima para baixo: 2400 Hz, 3850 Hz e 4800 Hz.

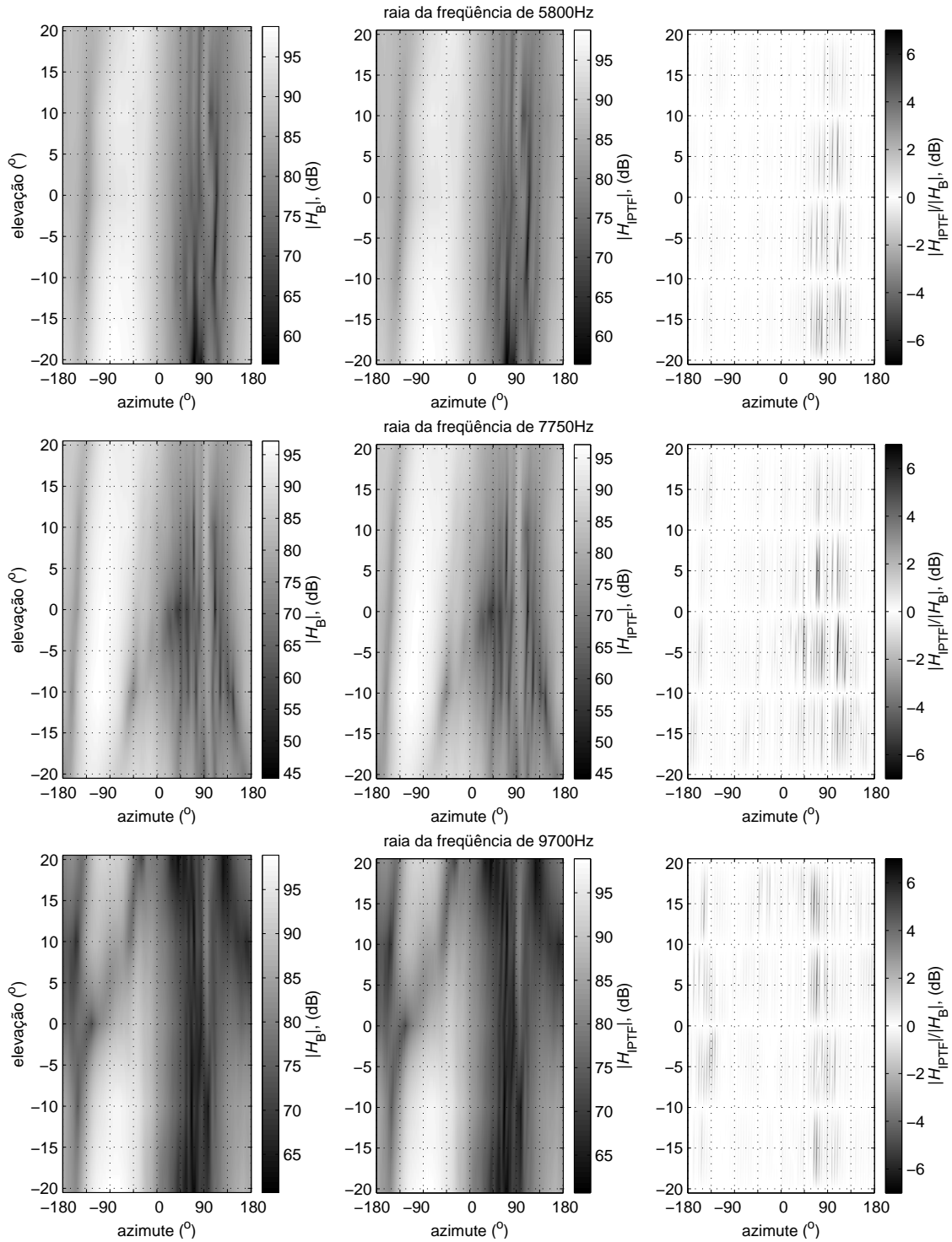


Figura 4.7: Comparação entre SFRSs. Bilinear (esquerda), IPTF (centro) e diferença em dB (direita)—III. De cima para baixo: 5800 Hz, 7750 Hz e 9700 Hz.

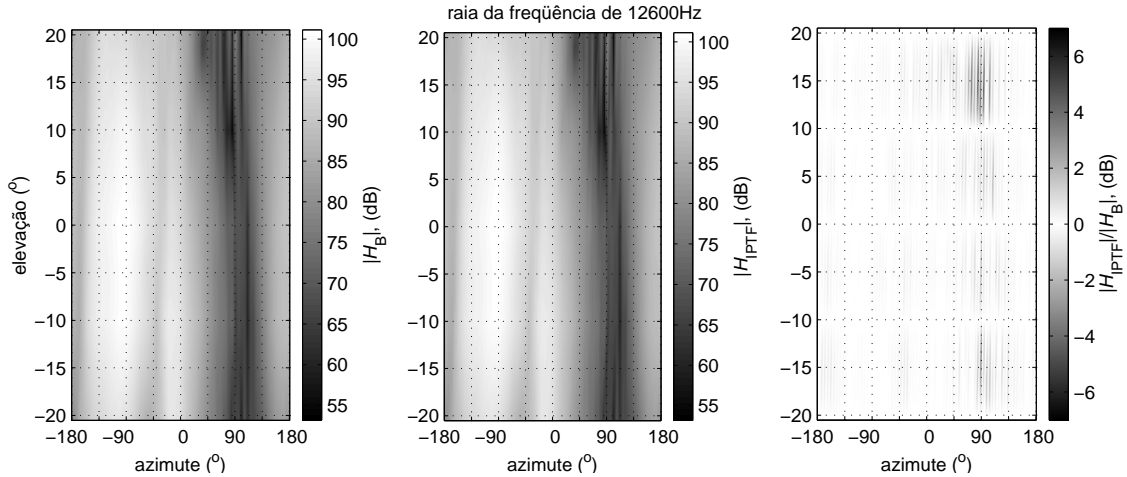


Figura 4.8: Comparação entre SFRSs. Bilinear (esquerda), IPTF (centro) e diferença em dB (direita)—IV. 12600 Hz.

Assim, método com IPTFs terá menor complexidade computacional se

$$M < \frac{3N + 6}{2}. \quad (4.12)$$

Para o caso das HRTFs usadas nas simulações [23] (que têm $N = 128$), isso significa $M \leq 195$. Por outro lado, é possível tirar vantagem de aproximações eficientes de HRTFs [34]. Por exemplo, assumindo $N = 33$, que é o número de multiplicações mais baixo atingido em [34], basta que $M < 53$ para que a interpolação baseada nas IPTFs apresente menor complexidade computacional.

Deve ser notado que como o procedimento de interpolação baseado nas IPTFs usa uma grade triangular, o número de regiões visitadas pela fonte virtual aumenta. Conseqüentemente, qualquer processamento adicional no momento de trocas de regiões (veja a Seção 5.4) aumentaria também a complexidade computacional. Por outro lado, para altas taxas de amostragem, por exemplo 44,1 kHz, e para movimentos não muito rápidos, esse provável aumento na complexidade não pesa muito na média, já que o número de amostras dentro de uma mesma região é muito grande. Por exemplo, com uma velocidade angular de aproximadamente 4π rad/s (duas voltas em 1s), na região mais densa, angularmente falando (próximo a $\phi = 0$), haverá mais de 300 amostras dentro da mesma região. Portanto, o cálculo de uma amostra de saída com a redução de transitório não fará aumentar muito o número de operações na média.

4.4 Conclusão

Nesse capítulo, redefiniu-se a IPTF como pode ser visto em [P1]. Além do que fora realizado naquele trabalho, foi feita uma comparação objetiva utilizando-se as SFRSs calculadas para os métodos de interpolação bilinear direto e aquele que se baseia nas IPTFs, publicado em [P2]. Além disso, foi feita uma comparação mais cuidadosa da complexidade computacional desses métodos, que indicou um limite máximo para o número de multiplicações gastas com as IPTFs de forma que o método que nela se baseia ainda seja competitivo frente ao bilinear direto.

No capítulo a seguir, discute-se uma forma de redução de ordem para as IPTFs testada em [P1] e validada em [P2].

Capítulo 5

Redução de Ordem das IPTFs

Como visto no capítulo anterior, a interpolação com IPTFs só resulta em uma complexidade computacional mais baixa se a ordem dessas funções for reduzida, deixando o número de multiplicações abaixo do limite da Equação (4.12).

Em [P1], utilizou-se a técnica de redução de modelos balanceados para alcançar um conjunto de IPTFs de ordem reduzida que foram utilizadas nessa tese.

Nesse capítulo, o método de redução de ordem é revisto na próxima seção. Na Seção 5.2, é mostrado o procedimento utilizado para criação do conjunto de IPTFs. Alguns resultados comparativos são mostrados na Seção 5.3. Na Seção 5.4, é realizado o estudo sobre a complexidade computacional atingida com a redução de ordem. Na Seção 5.5, apresentam-se as conclusões.

5.1 Redução do Modelo Balanceado

Em [41], os autores aplicam o método de redução do modelo balanceado (BMR—*Balanced Model Reduction*) para reduzir a ordem das HRTFs. Já que as HRTFs são FIR, eles se aproveitam de um algoritmo BMR especializado que aproxima filtros FIR por filtros IIR de ordem baixa, descrito em [46]. Como as IPTFs são formadas pela razão de duas HRTFs, e, portanto, é originalmente um sistema IIR, a formulação mais simples de [46] não pode ser usada. Apesar disso, o método BMR, em sua formulação mais geral, pode ser aplicado a qualquer sistema, sendo utilizado aqui na redução da ordem de IPTFs de um sistema em um espaço de estados.

A representação nos quais os gramianos das suas matrizes de observabilidade

e controlabilidade são ambos diagonais e iguais entre si é chamado de modelo balanceado. A maior vantagem desse tipo de modelo é a possibilidade de se ordenarem os estados com base na sua importância na representação das características do sistema. Esse é um aspecto bastante importante, que permite que se faça a simplificação do sistema diretamente pelo seu truncamento, isto é, descartando-se os estados “menos importantes”.

Considere a representação no espaço de estados de uma IPTF como

$$\begin{cases} \mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{b}u(k) \\ y(k) = \mathbf{c}^T\mathbf{x}(k) + du(k) \end{cases}, \quad (5.1)$$

onde \mathbf{A} , \mathbf{b} , \mathbf{c} e d são a matriz de transição, o vetor de entrada, o vetor de saída, e o coeficiente de transmissão direta, respectivamente. $\mathbf{x}(k)$ é o vetor de estados no instante k , e $u(k)$ e $y(k)$ são os sinais de entrada e de saída no instante k , respectivamente.

Defina

$$\mathbf{W}_o = \left[\mathbf{c} \quad \mathbf{A}^T\mathbf{c} \quad \cdots \quad (\mathbf{A}^T)^i\mathbf{c} \quad \cdots \right]^T \quad (5.2)$$

e

$$\mathbf{W}_c = \left[\mathbf{b} \quad \mathbf{A}\mathbf{b} \quad \cdots \quad \mathbf{A}^i\mathbf{b} \quad \cdots \right], \quad (5.3)$$

para $i = 1, 2, 3, \dots$, como as matrizes de observabilidade e controlabilidade do sistema, respectivamente. Suas respectivas matrizes gramianas são dadas por

$$\mathbf{P} = \mathbf{W}_c\mathbf{W}_c^T \quad (5.4)$$

e

$$\mathbf{Q} = \mathbf{W}_o^T\mathbf{W}_o. \quad (5.5)$$

Se $\text{rank}[\mathbf{A}] = J$, pode-se gerar a matriz de Hankel de posto J

$$\mathbf{H} = \mathbf{W}_o\mathbf{W}_c = \begin{bmatrix} h_1 & h_2 & \cdots \\ h_2 & h_3 & \\ \vdots & & \ddots \end{bmatrix}, \quad (5.6)$$

onde

$$h_i = \mathbf{c}^T\mathbf{A}^{i-1}\mathbf{b}, \quad (5.7)$$

para $i = 1, 2, 3, \dots$, são as amostras da resposta ao impulso do sistema.

Agora, \mathbf{H} pode ser decomposta como

$$\mathbf{H} = \mathbf{V}_1 \mathbf{\Gamma} \mathbf{V}_2, \quad (5.8)$$

onde $\mathbf{\Gamma}$ é a matriz contendo os valores singulares $\sigma_1, \dots, \sigma_J$ de \mathbf{H} nas primeiras J posições da sua diagonal principal, por conveniência organizados em ordem decrescente de magnitude [47]. Seja $\mathbf{\Sigma}$ a sub-matriz $J \times J$ superior esquerda de $\mathbf{\Gamma}$.

Já que o quadrado de cada valor singular de \mathbf{H} , σ_i^2 , é igual ao autovalor correspondente do produto \mathbf{PQ} [48, 46] denotado por λ_i , para $i = 1, 2, 3, \dots$, então

$$\mathbf{PQ} = \mathbf{K} \mathbf{\Sigma}^2 \mathbf{K}^{-1}. \quad (5.9)$$

Assim, pode-se mostrar que os gramianos \mathbf{P}_b e \mathbf{Q}_b do sistema balanceado desejado devem ser tais que

$$\mathbf{P}_b = \mathbf{Q}_b = \mathbf{\Sigma} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_J}) = \text{diag}(\sigma_1, \dots, \sigma_J). \quad (5.10)$$

Então, a transformação que leva o sistema original (\mathbf{A} , \mathbf{b} , \mathbf{c} e d) à sua forma balanceada (\mathbf{A}_b , \mathbf{b}_b , \mathbf{c}_b e d) por

$$\mathbf{A}_b = \mathbf{T}^{-1} \mathbf{A} \mathbf{T}, \quad \mathbf{b}_b = \mathbf{T}^{-1} \mathbf{b}, \quad \mathbf{c}_b = \mathbf{T}^T \mathbf{c} \quad (5.11)$$

é

$$\mathbf{T} = \mathbf{S}^{-1} \mathbf{U} \mathbf{\Sigma}^{1/2}, \quad (5.12)$$

onde \mathbf{S} é obtida da decomposição da matriz \mathbf{Q} na forma

$$\mathbf{Q} = \mathbf{S}^T \mathbf{S}. \quad (5.13)$$

As matrizes \mathbf{U} e $\mathbf{\Sigma}$ podem ser calculadas fazendo-se a decomposição:

$$\mathbf{SPS}^T = \mathbf{U} \mathbf{\Sigma}^2 \mathbf{U}^T, \quad (5.14)$$

onde

$$\mathbf{U} \mathbf{U}^T = \mathbf{I}. \quad (5.15)$$

Deve-se notar que a equação (5.14) é satisfeita, já que $\mathbf{SPS}^T = \mathbf{SPQS}^{-1}$ e, portanto, \mathbf{SPS}^T é simétrica e tem os mesmos autovalores do produto \mathbf{PQ} .

Considere o caso onde, de acordo com algum critério de erro, somente os $j < J$ maiores valores singulares são considerados suficientes para descrever a IPTF. Nesse caso, a matriz Σ pode ser escrita na forma

$$\Sigma = \begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{bmatrix}, \quad (5.16)$$

onde $\mathbf{0}$ são matrizes nulas,

$$\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_j) \quad (5.17)$$

e

$$\Sigma_2 = \text{diag}(\sigma_{j+1}, \dots, \sigma_J). \quad (5.18)$$

Coerentemente, as matrizes do modelo balanceado podem ser decompostas na forma

$$\mathbf{A}_b = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} \end{bmatrix}, \mathbf{b}_b = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}, \mathbf{c}_b = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}, \quad (5.19)$$

sendo $(\mathbf{A}_{1,1}, \mathbf{b}_1, \mathbf{c}_1$ e d) as matrizes de representação no espaço de estados do sub-sistema de ordem j que aproxima acuradamente a IPTF original.

Após a redução de ordem, se desejado, pode-se converter a representação no espaço de estados de volta para a forma de função de transferência.

5.2 Algoritmo de Redução de Ordem

As técnicas de suavização espectral e redução de ordem mencionadas anteriormente são as ferramentas básicas para se obter um conjunto completo de IPTFs de ordem baixa. Por fim, essas IPTFs são utilizadas juntamente com as HRTFs medidas em um sistema eficiente de geração de som tridimensional em movimento.

O procedimento para se obter o conjunto completo de IPTFs consiste de três estágios:

- cada HRTF medida sobre a grade esférica de referência é transformada na sua versão de fase mínima, sendo ela suavizada durante esse processo;
- para cada HRTF suavizada define-se uma IPTF com relação a cada uma das HRTFs da vizinhança (por exemplo, como na Figura 5.1);

- o método BMR é aplicado a cada IPTF definida no passo anterior para se obter um modelo de ordem reduzida que a represente bem.

De fato, essa estratégia contém alguma redundância, já que duas IPTFs podem ser definidas entre duas HRTFs contíguas, dependendo da escolha de qual é a $HRTF_f$ e qual é a $HRTF_i$ na Equação (4.1). Obviamente, somente uma delas precisa ser armazenada por cada par de HRTF¹. No entanto, essas funções redundantes permitem escolher dentre elas qual melhor representa a IPTF em questão. Essa flexibilidade também ajuda a lidar com casos que apresentarem problemas numéricos na redução de ordem.

Para a vizinhança, deve-se adotar um critério em que nem sempre todas as posições das HRIRs próximas são consideradas vizinhas. O que deve ser garantido é que os setores triangulares escolhidos não tenham superposição na casca esférica e cubram toda a superfície da mesma. A escolha de vizinhança adotada nessa tese pode ser melhor visualizada através das Figuras 5.1 e 5.2. Na Figura 5.1, tem-se a seção da esfera aproximada pelas faces triangulares escolhidas². Na Figura 5.2, mostra-se a planificação dessa esfera para ser possível visualizar todas as regiões triangulares escolhidas. Desse modo, as IPTFs computadas são aquelas entre as posições ligadas por segmentos de reta. Deve-se notar que a escolha de quais HRIRs são vizinhas não é única³ e que a forma adotada aqui é apenas uma das possíveis. Portanto, se outra escolha for adotada, as IPTFs calculadas formarão um conjunto distinto do utilizado nessa tese.

É importante ser mencionado que em [P1] não se utilizava nenhuma definição de vizinhança e, ao se passar por algumas posições, era gerada uma pequena descon-

¹Por exemplo, para representar a HRTF de um ponto a utilizando a HRTF do ponto b , utiliza-se a $IPTF_{b,a}$. E para representar a HRTF de um ponto b utilizando a HRTF do ponto a , pode-se usar $IPTF_{a,b} = 1/IPTF_{b,a}$. Isso é o mesmo que trocar o polinômio do numerador com o do denominador.

²Como as HRIRs utilizadas nesse trabalho só foram medidas para elevações a partir de -40° , a base da esfera aparece cortada abaixo de tal limite.

³Como mencionado anteriormente, a escolha da vizinhança só deve garantir que as regiões cubram toda a casca esférica e que não haja superposição entre as regiões. Somente essa restrição não é suficiente para se saber quais são as regiões vizinhas e, com isso, a vizinhança entre HRIRs (segmentos de reta que unem os pontos para os quais se tem as HRIRs medidas).

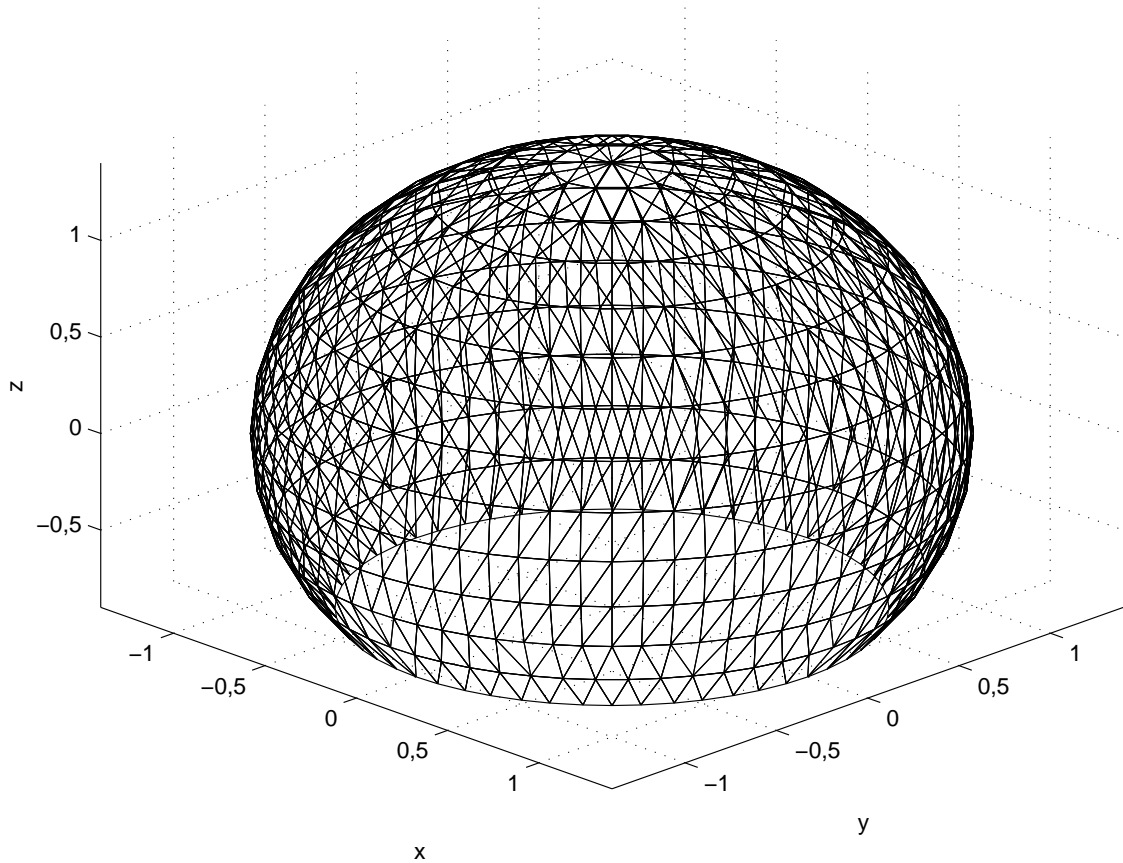


Figura 5.1: Definição de vizinhança entre as coordenadas das HRIRs (vista espacial).

tinuidade no som. Esse problema apontou a necessidade de se adotar uma definição fixa de vizinhança. Esse novo procedimento foi adotado para se obter os resultados que serão apresentados a seguir.

5.3 Comparações entre as Interpolações Baseada em IPTFs Sem e Com Redução de Ordem

Na Seção 4.3, foi feita uma comparação entre os métodos de interpolação bilinear direto e baseado nas IPTFs usando as SFRSs. Na presente seção, os resultados da interpolação com as IPTFs de ordem reduzida são comparados àqueles obtidos com as IPTFs sem redução alguma. Como antes, um conjunto de SFRSs é utilizado para visualização desses resultados. Desta vez, os ângulos de elevação variam de -40° a 90° , assim cobrindo toda a região em que se têm HRTFs medidas. Da mesma forma que anteriormente, as raias freqüenciais foram escolhidas de acordo

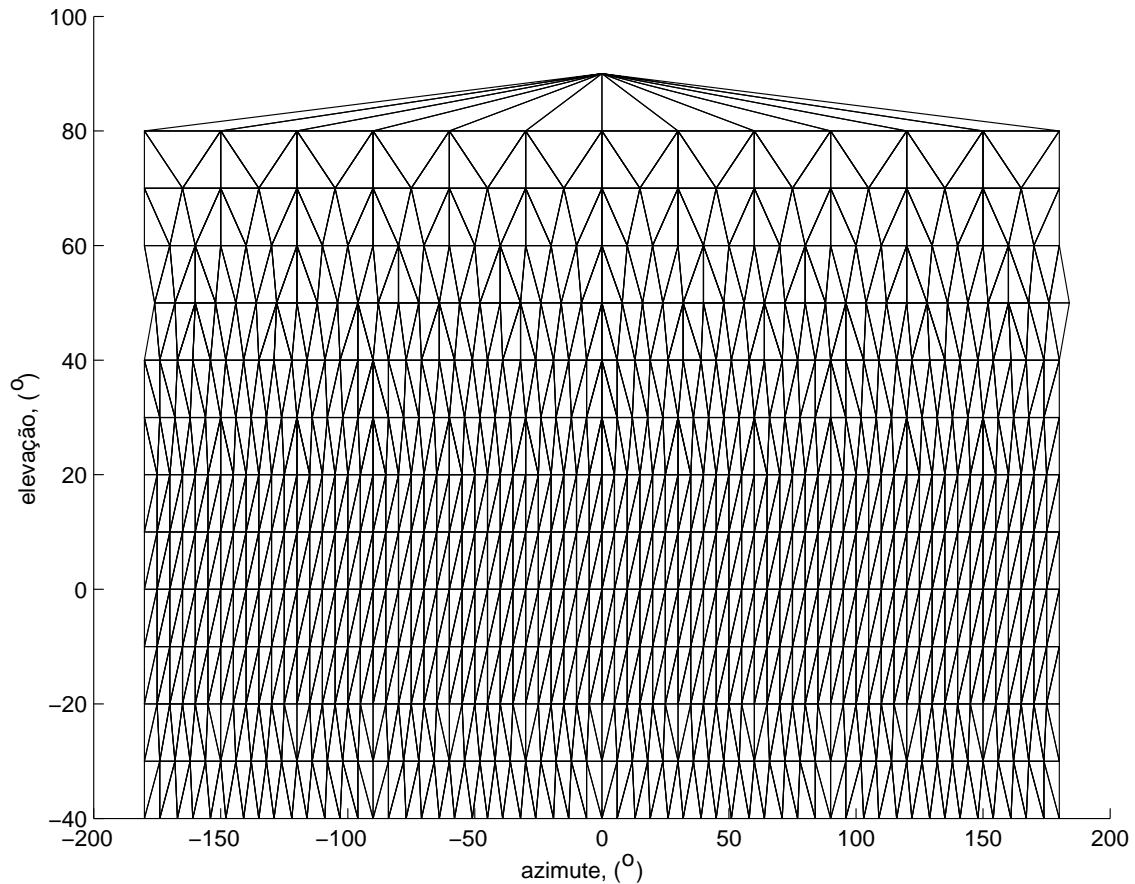


Figura 5.2: Definição de vizinhança entre as coordenadas das HRIRs (vista planificada).

com [28] (Apêndice B).

Um conjunto completo de IPTFs IIR de ordem 10 foi obtido⁴ através do procedimento descrito na Seção 5.2. Cada IPTF desse conjunto corresponde a $M = 21$, satisfazendo os limites de eficiência, considerando $N = 33$ da Seção 4.3 ($M \leq 53$). Assim, o método de interpolação proposto é computacionalmente mais eficiente que o método bilinear direto.

A comparação entre os conjuntos de IPTFs sem redução e com modelos de ordem 10 pode ser vista nas Figuras 5.3–5.6. Como no capítulo anterior, em cada linha mostram-se um par de SFRSs relativas as HRTFs interpoladas para a orelha esquerda e a diferença entre elas obtida por procedimento similar à Equação (4.9),

⁴A ordem 10 para os modelos das IPTFs foi obtida através de testes [P1] onde eram comparados visualmente o módulo das respostas em frequência das IPTFs originais e dos modelos de ordem reduzida.

no presente caso resultando na equação

$$\xi(\omega, \theta, \phi) = 20 \log_{10} \left(\frac{|H_{\text{IPTF}}^{\text{red}}(\omega, \theta, \phi)|}{|H_{\text{IPTF}}(\omega, \theta, \phi)|} \right). \quad (5.20)$$

As SFRSs da primeira coluna correspondem à interpolação com as IPTFs sem redução de ordem, enquanto que as SFRSs da segunda coluna mostram os resultados obtidos pelo mesmo método usando as IPTFs de ordem 10, calculadas de acordo com o procedimento visto no presente capítulo. Com base nas diferenças apresentadas na terceira coluna, pode-se observar que as SFRSs dos dois métodos são bastante parecidas. Apesar da falta de regularidade aparente nas SFRSs, através de alguns testes auditivos o resultado final mostrou-se pouco afetado, dado o pequeno valor do erro em contraste com os altos valores absolutos das SFRSs. Novamente, as regiões mais prejudicadas pelo modelo estão próximas a valores bem baixos nas SFRSs (posições em que já não se tem muita informação sobre as HRTFs).

Por simetria, as SFRSs para a orelha direita podem ser verificadas simplesmente invertendo-se o eixo dos azimutes.

Para que se tenha uma avaliação mais precisa, na Figura 5.7 é mostrado um histograma para o erro relativo da Equação (5.20), calculado em $-180^\circ \leq \theta < 180^\circ$ e $-40^\circ \leq \phi \leq 90^\circ$ com passo de 1° e nas frequências escolhidas para se obter as SFRSs. Esse histograma mostra que mais de 98% dos resultados calculados apresentam um erro menor que 1,2 dB, indicando que o método com IPTFs de ordem reduzida equivale ao método triangular.

5.4 Complexidade Computacional e Desempenho

Um sistema com as características descritas anteriormente foi implementado. Uma avaliação do desempenho através de testes auditivos informais revelou que usar a interpolação baseada em IPTFs com ordem 10 leva a bons resultados, comparáveis aos obtidos com a interpolação triangular de HRIRs. Considerando $N = 33$ [34] e IPTFs de ordem 10 ($M = 21$), pela Equação (4.10) o método bilinear precisa de 144 mpa, enquanto que pela Equação (4.11) o método triangular utilizando IPTFs necessita de 81 mpa, reduzindo em 63 mpa (ou mais de 40%) a complexidade computacional necessária para o método bilinear. Note que, apesar disso, comparando-se a interpolação triangular aplicada diretamente sobre as HRIRs com ela mesma através

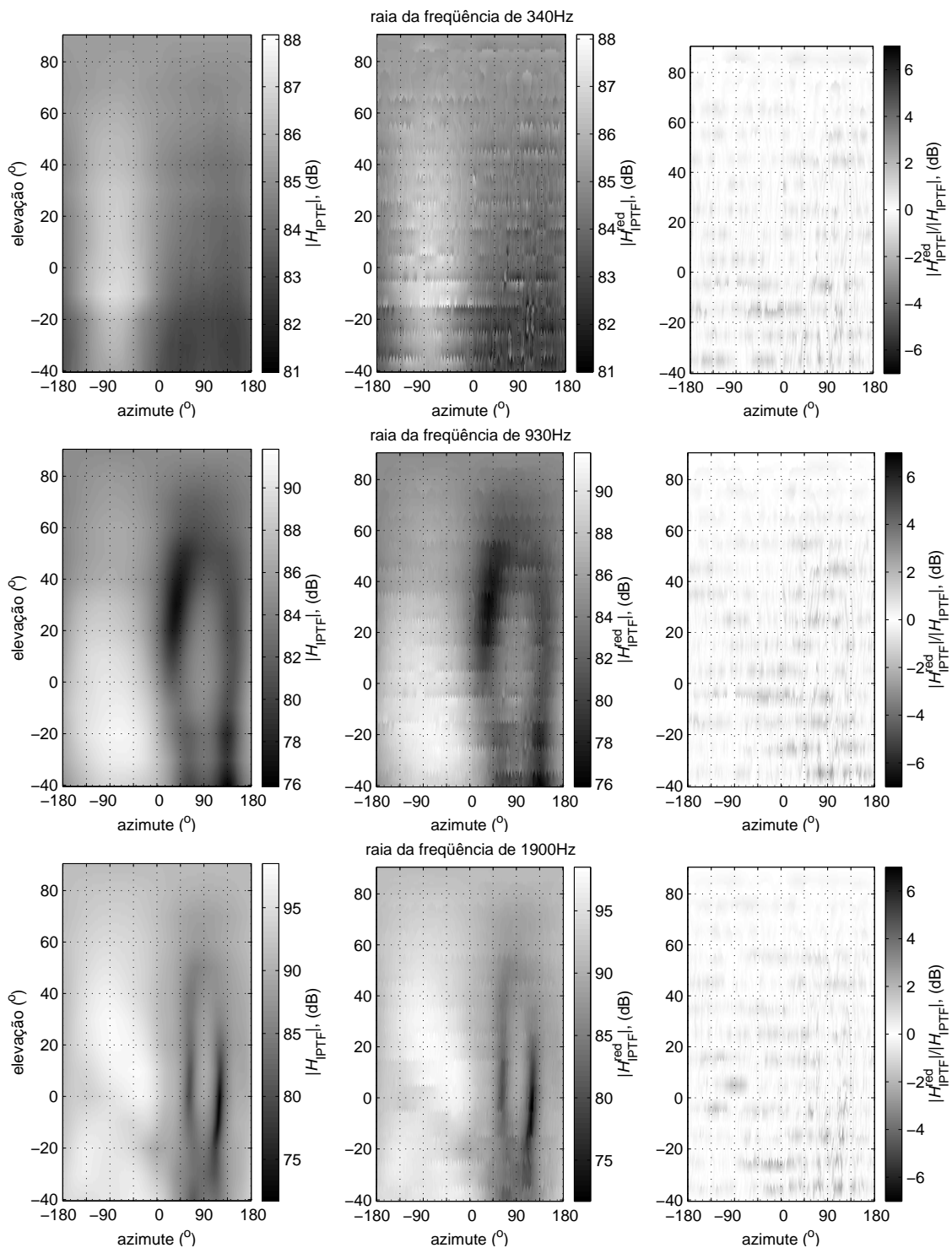


Figura 5.3: Comparação entre SFRSs. IPTF sem (esquerda) e com (centro) redução de ordem, e diferença em dB (direita)—I. De cima para baixo: 340 Hz, 930 Hz e 1900 Hz.

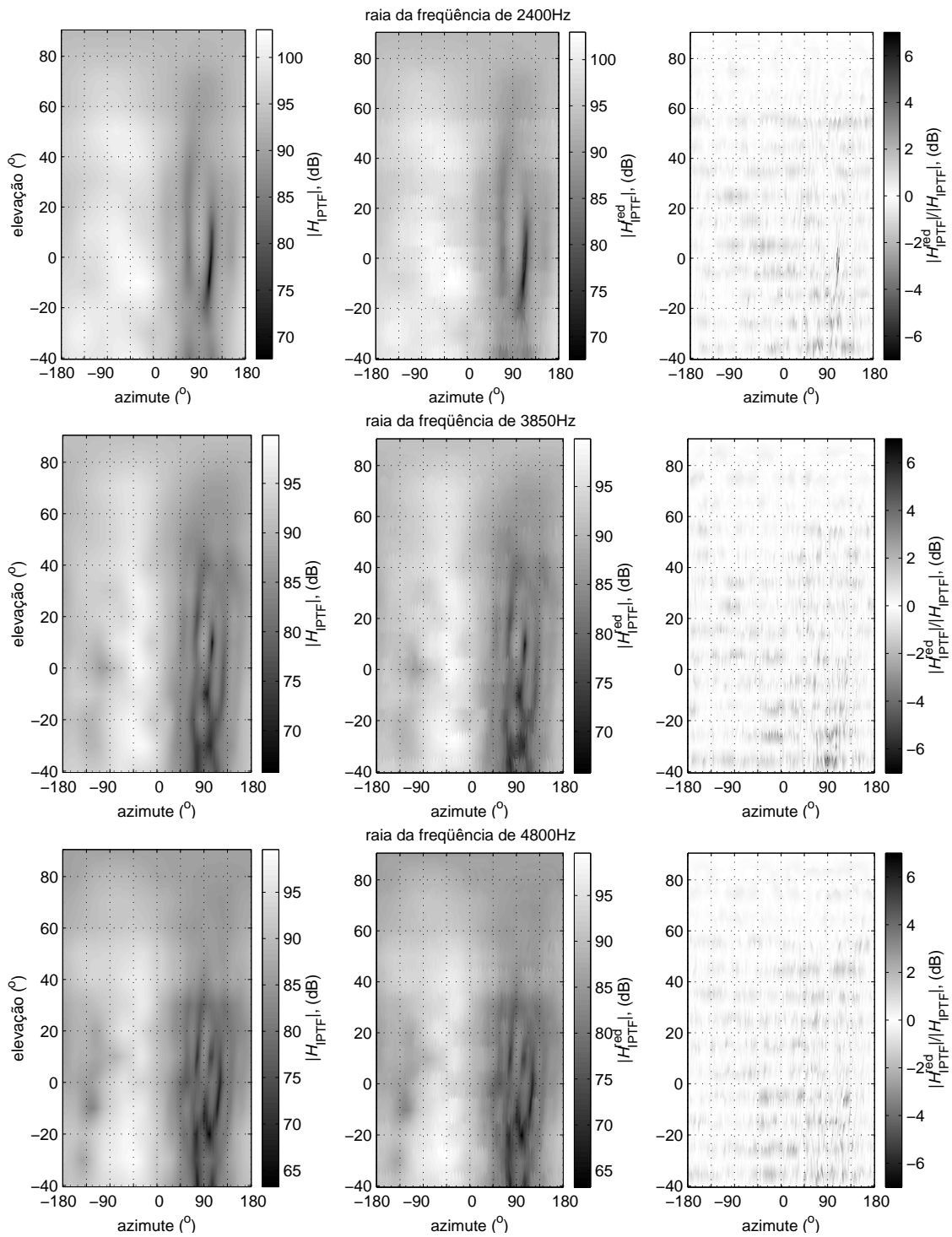


Figura 5.4: Comparação entre SFRSs. IPTF sem (esquerda) e com (centro) redução de ordem, e diferença em dB (direita)—II. De cima para baixo: 2400 Hz, 3850 Hz e 4800 Hz.

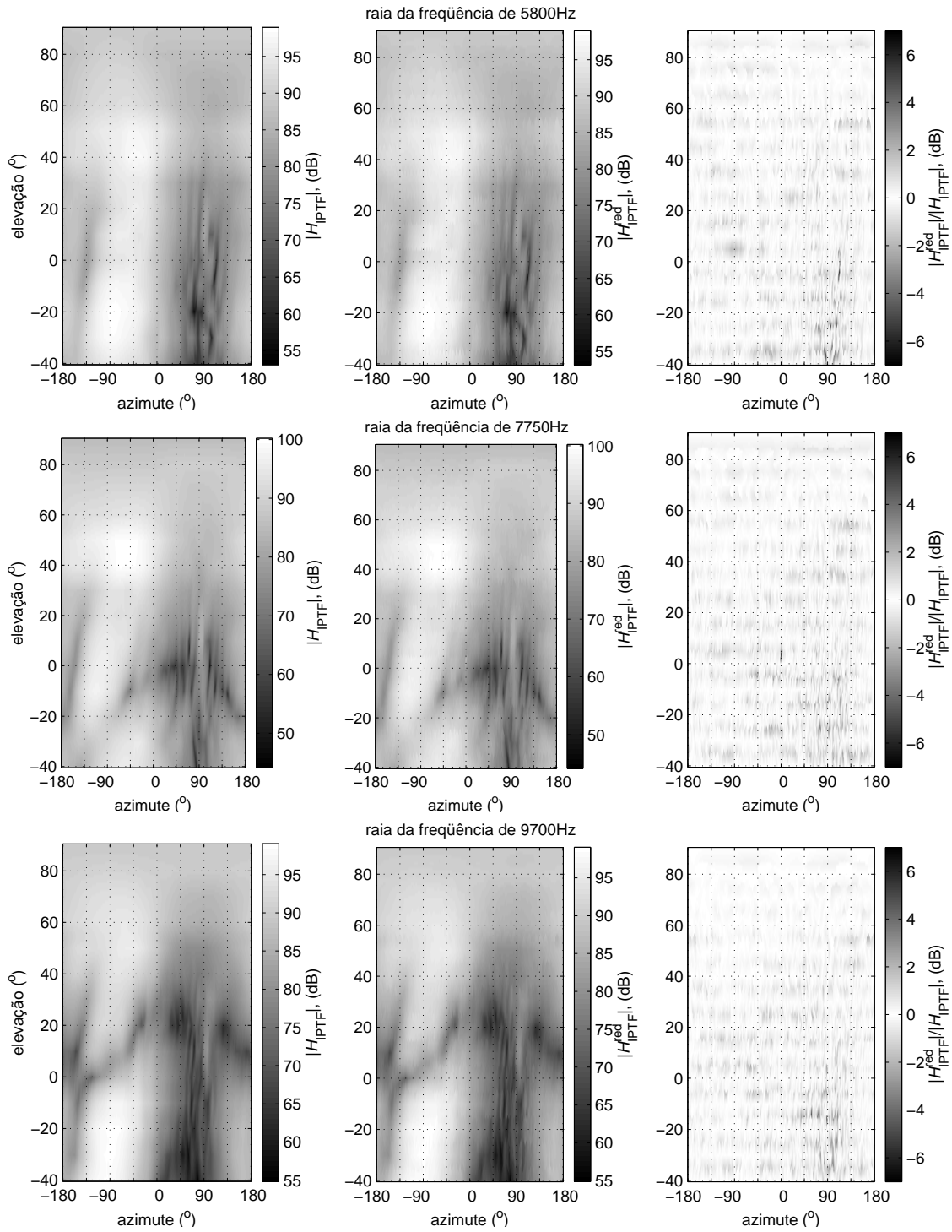


Figura 5.5: Comparação entre SFRSs. IPTF sem (esquerda) e com (centro) redução de ordem, e diferença em dB (direita)—III. De cima para baixo: 5800 Hz, 7750 Hz e 9700 Hz.

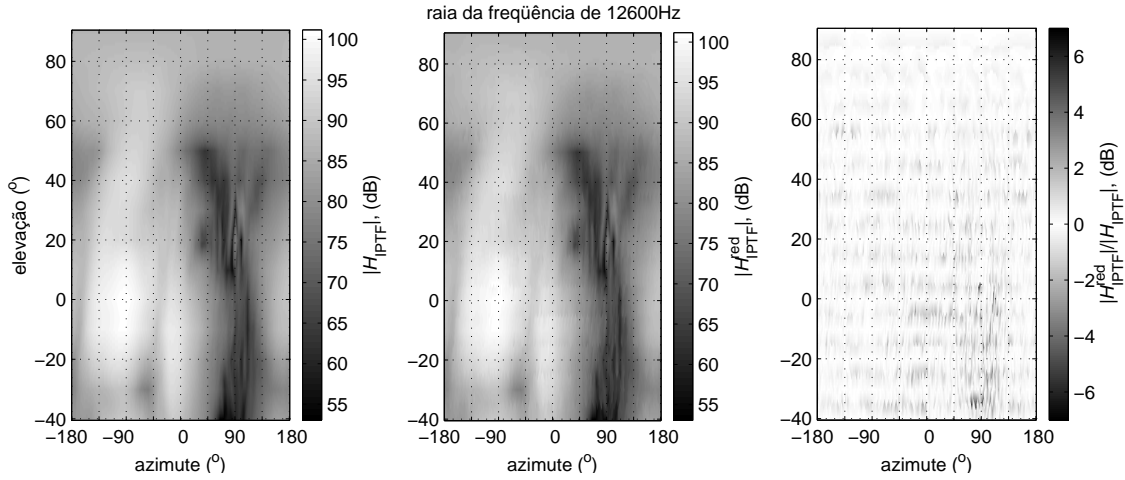


Figura 5.6: Comparação entre SFRSs. IPTF sem (esquerda) e com (centro) redução de ordem, e diferença em dB (direita)—IV. 12600 Hz.

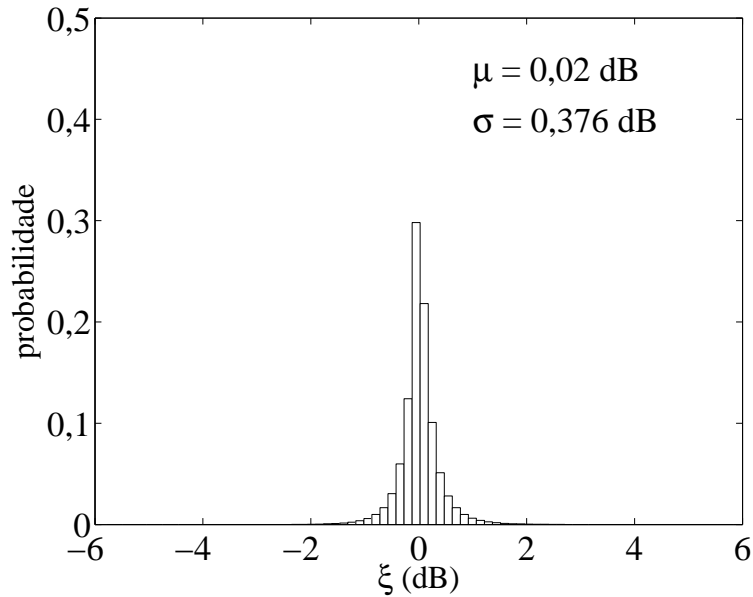


Figura 5.7: Histograma para o erro ξ (interpolação com IPTFs de ordem reduzida).

das IPTFs, o ganho na complexidade computacional é $2 \times (N - M)$, o que significa $2 \times 12 = 24$ mpa para os valores utilizados acima. Portanto, somente por se utilizarem IPTFs tem-se uma queda no número de multiplicações de 24 mpa contra 105 mpa, ou mais de 22%.

Por fim, deve ser lembrado que a troca de sucessivas HRTFs interpoladas ao longo do tempo envolve trocas de filtros IIR, um procedimento que introduz transitórios espúrios que podem causar falhas audíveis no sinal de saída. Esse problema pode ser solucionado como em [49], onde se propõe uma redundância nos estados

do filtro como forma de minimizar tais efeitos. Esse procedimento aumenta momentaneamente o número de multiplicações para se calcular uma nova amostra, ao se passar de uma região para a outra. No entanto, como isso só ocorre apenas nas mudanças de região, o efeito global não afeta a média de multiplicações por amostra de saída. Essa estratégia foi utilizada no sistema, mostrando-se eficaz.

5.5 Conclusões

A partir dos resultados anteriores, é possível realizar-se a interpolação baseada em IPTFs em um sistema prático de geração de som em movimento restrito ao posicionamento angular de uma única fonte monaural. Nesse caso, o sistema recebe ao longo do tempo o sinal monaural e a informação de localização relativa à fonte virtual desejada sobre a esfera de referência. Então, com algumas considerações sobre acurácia de tempo e posição, define-se a posição onde a fonte virtual deve ser situada.

Ao se realizar a interpolação de HRTFs através de IPTFs, o procedimento que deve ser seguido é: para cada ponto P na trajetória da fonte, são escolhidas a HRTF relativa ao ponto mais próximo e as duas IPTFs relativas aos vértices da região triangular, assim formando o triângulo de HRTFs (aproximadas ou não) que contém P; então, a HRTF em P é calculada de acordo com a Equação (4.3) e usada para processar o sinal de áudio.

Como a complexidade para se modelar as IPTFs sobre a esfera de referência pode variar com a localização, é razoável pensar-se em usar IPTFs de diferentes ordens—ordem menor quando possível e ordem maior quando necessário. Com isso, o tempo de processamento variaria com a posição. Nesse caso, ao se implementar o sistema em tempo-real, é necessário levar-se em conta o pior caso.

Alguns exemplos da aplicação do método proposto podem ser encontrados em:

<http://www.lps.ufrj.br/sound/journals/JAES04/>.

Capítulo 6

Interpolação no Domínio da Transformada

Como no domínio do tempo, a interpolação de HRTFs pode ser realizada sobre coeficientes gerados a partir de uma determinada transformação linear. Se for possível reduzir-se o número de funções-base da transformada utilizada, essa forma de interpolação terá uma grande vantagem em comparação com as aplicadas sobre as HRTFs: com um número reduzido de funções poder-se-ia gerar a saída para várias fontes simultaneamente com um mínimo de acréscimo em esforço computacional para cada fonte extra.

É possível considerar as HRTFs como funções de três variáveis: duas coordenadas esféricas de posição (aqui assume-se o raio constante) e a frequência. E, como descrito em [50], a dependência da frequência pode ser separada da dependência das coordenadas de azimute θ e elevação ϕ fazendo-se

$$H(\theta, \phi, f) = \sum_{i=1}^N \omega_i(\theta, \phi) \Gamma_i(f),$$

onde $\omega_i(\theta, \phi)$, $\Gamma_i(f)$ são as componentes dependentes da posição angular e da frequência, respectivamente, e N é o número de funções $\Gamma_i(f)$ (dimensão da base da transformada). As componentes $\omega_i(\theta, \phi)$, também chamadas de Funções Características Espaciais (SCFs—*Spatial Characteristic Functions*) [50], podem ser consideradas os coeficientes de uma transformada para a posição (θ, ϕ) , enquanto que $\Gamma_i(f)$ está relacionada com a função base da transformada.

Em [51, 52], os autores propõem um modelo chamado de Extração e Re-

gularização das Características Espectrais (SFER—*Spectral Feature Extraction and Regularization*), que representa as HRTFs através da Transformada de Karhunen-Loève (KLT—*Karhunen-Loève Transform*) [53], cujas funções-base são utilizadas como as componentes $\Gamma_i(f)$. Naqueles trabalhos, para se obter a interpolação das HRTFs utilizam-se *splines* sobre as SCFs.

Nesse capítulo, desenvolve-se uma maneira de interpolar as HRIRs no domínio da KLT separando-se, como em [51, 52], as componentes dependentes da frequência e do espaço. Ao invés de se utilizar *splines*, propõe-se uma interpolação linear incremental das SCFs utilizando-se suas amostras conhecidas: os coeficientes para cada HRIR medida.

Na próxima seção, mostra-se como se podem representar as HRIRs usando-se a KLT. Na Seção 6.2, apresenta-se o modelo proposto para as SCFs juntamente com a forma utilizada para a interpolação. Em seguida, na Seção 6.3, comparam-se os resultados da interpolação proposta, publicado em [P5], e da interpolação triangular direta, para um determinado conjunto de frequências.

6.1 Representação de HRIRs Usando a KLT

Considerando-se \mathbf{H} a matriz de um processo aleatório que contém em cada linha uma amostra desse processo, a transformação que diagonaliza a sua matriz de covariância \mathbf{C} é a KLT. As funções base da KLT tendem a representar em poucas direções a maior parte da energia do processo. Com isso, tolerando-se um certo erro na representação, pode-se reduzir o número de funções base e, conseqüentemente, o número de coeficientes necessários para tal representação.

Para representar o conjunto de HRIRs com a KLT, considere cada HRIR como uma amostra do processo aleatório representado pela matriz \mathbf{H} , cujas linhas são compostas pelos coeficientes das HRIRs de ordem N . A matriz de covariância \mathbf{C} de \mathbf{H} pode ser decomposta da forma

$$\mathbf{C} = (\mathbf{H} - \bar{\mathbf{H}})^T (\mathbf{H} - \bar{\mathbf{H}}) = \mathbf{\Psi} \mathbf{\Lambda} \mathbf{\Psi}^T, \quad (6.1)$$

com

$$\mathbf{\Psi}^T \mathbf{\Psi} = \mathbf{I}, \quad (6.2)$$

obtendo-se $\Psi = [\psi_j(0) \cdots \psi_j(N-1)]^T$, com $j = 1, \dots, N$, como a matriz unitária de autovetores de \mathbf{C} , chamada também de matriz da KLT. Λ é a matriz diagonal cujos elementos são os autovalores da matriz \mathbf{C} , e $\bar{\mathbf{H}}$ é uma matriz da mesma dimensão de \mathbf{H} que contém em todas as linhas a HRIR média definida por

$$\bar{h}(k) = \frac{1}{I} \sum_{\theta} \sum_{\phi} h(\theta, \phi, k), \quad (6.3)$$

sendo I o número total de HRIRs.

Com as funções base da KLT, torna-se possível aproximar-se a HRIR associada a cada ponto (θ, ϕ) sobre a esfera de referência segundo a equação

$$\hat{h}(\theta, \phi, k) = \bar{h}(k) + \sum_{j=1}^N \omega_j(\theta, \phi) \psi_j(k), \quad (6.4)$$

onde $\omega_j(\theta, \phi)$ são as SCFs.

As SCFs, que seriam contínuas no espaço, não são conhecidas *a priori*; mas para cada uma das HRIRs $h(\tilde{\theta}, \tilde{\phi}, k)$ medidas nas posições $(\tilde{\theta}, \tilde{\phi})$ os coeficientes da KLT $\omega_j(\tilde{\theta}, \tilde{\phi})$ podem ser determinados. Nessas posições, $\hat{h}(\tilde{\theta}, \tilde{\phi}, k) = h(\tilde{\theta}, \tilde{\phi}, k)$. A partir dessas amostras das SCFs, os coeficientes da KLT relativos a qualquer outra HRIR devem ser calculados para se poder localizar a fonte em qualquer outra posição.

O uso da KLT permite uma redução significativa na complexidade computacional do processo de geração do áudio binaural, pois, ao invés de se utilizarem todos os autovetores calculados através da Equação (6.1) para representar as HRIRs, podem-se usar apenas M (com $M < N$), selecionando-se apenas aqueles relacionados com os maiores autovalores [51, 52]. Na Figura 6.1, tem-se o gráfico do percentual de energia em função do número M de autovetores utilizados na representação do conjunto de HRIRs. Nesse capítulo, utilizaram-se os $M = 32$ maiores autovalores, que correspondem a 99,9% da energia total.

A Figura 6.2 mostra o diagrama em blocos que aproxima uma das HRIRs (canal direito ou esquerdo) de um sistema binaural.

Para o caso de múltiplas fontes, espera-se que a aproximação através do modelo SFER reduza a complexidade computacional, se comparada às interpolações bilinear e triangular, já que no modelo SFER somente o número de SCFs (que associam um único multiplicador a cada posição e a cada vetor da base da KLT)

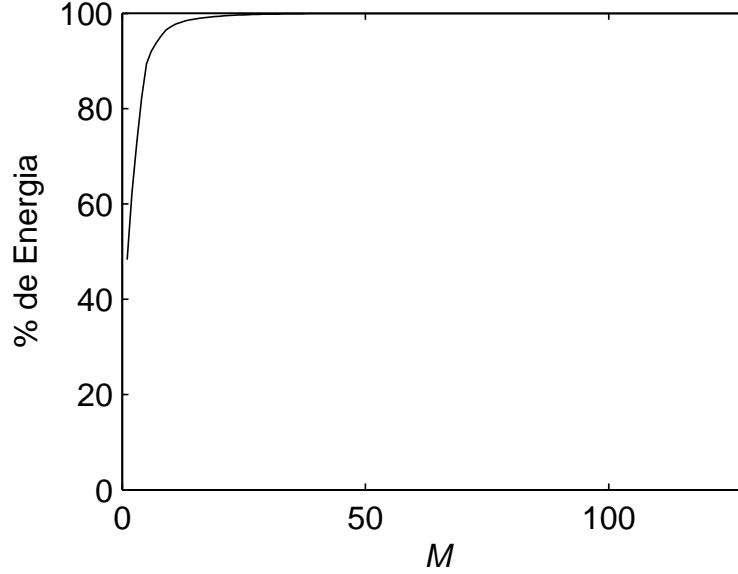


Figura 6.1: Percentual de energia em função do número de autovetores utilizados no modelo.

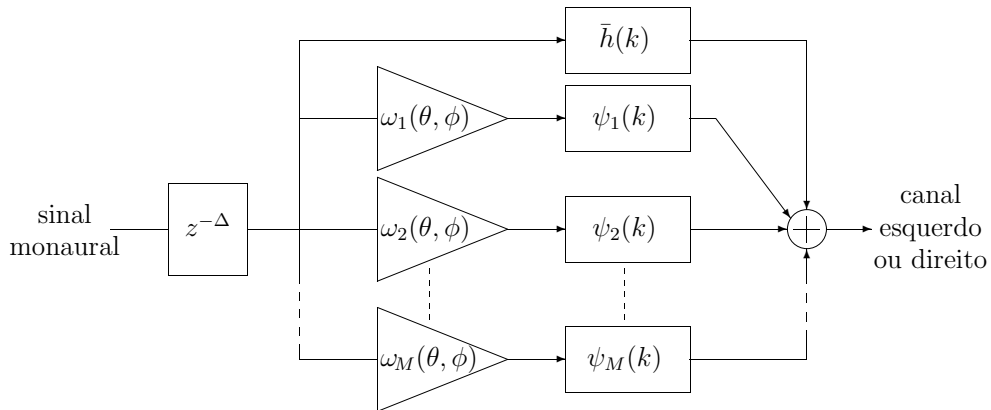


Figura 6.2: Diagrama em blocos da representação pela SFER.

é proporcional ao número de fontes virtuais desejadas (o número de funções $\psi_j(k)$ não varia com o número de fontes), enquanto que nos métodos bilinear e triangular, todos os cálculos são feitos independentemente para cada fonte virtual.

6.2 Modelo para as SCFs

Em [52, 51, 54], as SCFs $\omega_j(\theta, \phi)$ para qualquer posição (θ, ϕ) são calculadas utilizando-se suas amostras conhecidas $\omega_j(\tilde{\theta}, \tilde{\phi})$ através da interpolação por *splines* [55]. Apesar de ter um bom desempenho, o cálculo do valor das SCFs para cada ponto pode requerer um número muito grande de operações, dependendo do tipo

da *spline* utilizada. As *splines* de duas variáveis para as quais o cálculo do valor da função aproximada tem menor custo computacional consideram a grade formada pelas amostras conhecidas como sendo retangular. No caso das SCFs geradas a partir das HRIRs medidas para uma grade não retangular, a *spline* utilizada deve desconsiderar essa hipótese.

A *spline* que se encaixa nesse perfil é a chamada *scattered-translates*, que é definida na forma

$$f(\mathbf{x}) = \sum_{i=1}^{n-k} \Psi(\mathbf{x} - \mathbf{c}_i) a_i + p(\mathbf{x}), \quad (6.5)$$

onde \mathbf{x} é o vetor de coordenadas, a_i são os n coeficientes cujos k de maior índice estão presentes na parte polinomial $p(\mathbf{x})$ da aproximação e \mathbf{c}_i são os pontos centrais que servem como referência para as funções-base radiais $\Psi(\cdot)$.

Uma das formas da *scattered-translates* é a *thin-plate spline* de duas variáveis, para a qual a função base radial é

$$\Psi(\mathbf{x}) = \|\mathbf{x}\|_2^2 \ln \|\mathbf{x}\|_2^2, \quad (6.6)$$

sendo $\|\mathbf{x}\|_2^2$ o quadrado da norma 2 de \mathbf{x} e o polinômio $p(\mathbf{x})$ é

$$p(\mathbf{x}) = x_1 a_{n-2} + x_2 a_{n-1} + a_n, \quad (6.7)$$

o que significa $k = 3$ na Equação (6.5). Com isso o modelo para $\omega_j(\theta, \phi)$ pode ser descrito pela função $f(\mathbf{x})$, onde $\mathbf{x} = [\theta \ \phi]^T$.

Pode-se notar pela Equação (6.5) que, sendo o número $n - k$ de centros \mathbf{c}_i igual ao número de HRIRs utilizadas no cálculo dos coeficientes a_i (no caso, todas as medidas), o número de operações de multiplicação é da ordem de $(4 + \gamma)(n - k) + (k - 1)$, sendo γ o número de multiplicações envolvidas no cálculo do logaritmo. Isso se traduz em algo bem maior que 2800 multiplicações para cada SCF, considerando-se que o conjunto de $n - k$ HRIRs utilizado nesta tese contém um pouco mais de 700 funções e $k = 3$. Isso sem se levar em conta o cálculo do logaritmo¹.

A seguir, é proposta uma alternativa simples [P5] para substituir a interpolação por *splines*.

¹Consideraram-se apenas as operações gastas para se calcular $\|\mathbf{x}\|_2^2$, o produto pelo logaritmo da Equação (6.6) e o produto pelos a_i . Levando-se em conta γ , o número de multiplicações ainda aumentaria de $\gamma(n - k)$.

6.2.1 Modelo Linear Incremental para as SCFs

Se uma fonte virtual está posicionada no interior de uma região triangular ABC (veja a Figura 6.3) sobre a esfera de referência, supondo uma estimativa inicial calculada previamente, a estimativa das SCFs pode ser realizada de uma forma incremental, onde a equação de atualização dos coeficientes é dada por

$$\hat{\omega}_j(\theta_l, \phi_l) = \hat{\omega}_j(\theta_{l-1}, \phi_{l-1}) + \Delta\omega_{j,l-1}, \quad (6.8)$$

sendo l o índice da l -ésima posição angular. O incremento aplicado aos pesos $\hat{\omega}_j$ da posição $l - 1$ para a posição l pode ser calculada como

$$\Delta\omega_{j,l-1} = (\theta_l - \theta_{l-1}) \left. \frac{\partial\omega_j(\theta, \phi)}{\partial\theta} \right|_{\substack{\theta=\theta_{l-1} \\ \phi=\phi_{l-1}}} + (\phi_l - \phi_{l-1}) \left. \frac{\partial\omega_j(\theta, \phi)}{\partial\phi} \right|_{\substack{\theta=\theta_{l-1} \\ \phi=\phi_{l-1}}}. \quad (6.9)$$

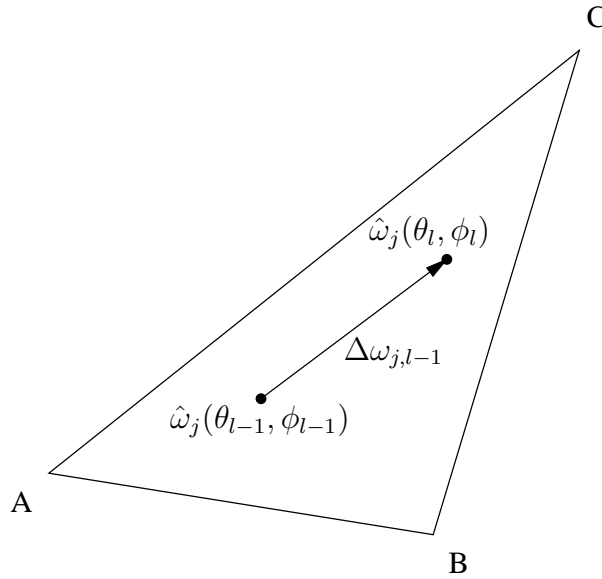


Figura 6.3: Interpretação gráfica do cálculo de uma das SCF.

Como já fora mencionado, as funções $\omega_j(\theta, \phi)$ não são conhecidas e, portanto, as derivadas parciais não podem ser obtidas analiticamente. Apesar disso, cada SCF $\omega_j(\theta, \phi)$ pode ser aproximada por faces triangulares (veja Figura 6.4) cujos vértices são suas amostras conhecidas $\omega_j(\tilde{\theta}, \tilde{\phi})$. Dessa forma, aquelas derivadas podem ser estimadas pelas inclinações da face que contém o ponto $(\theta_{l-1}, \phi_{l-1})$. Para que o cálculo das derivadas não aumente a complexidade computacional em tempo de

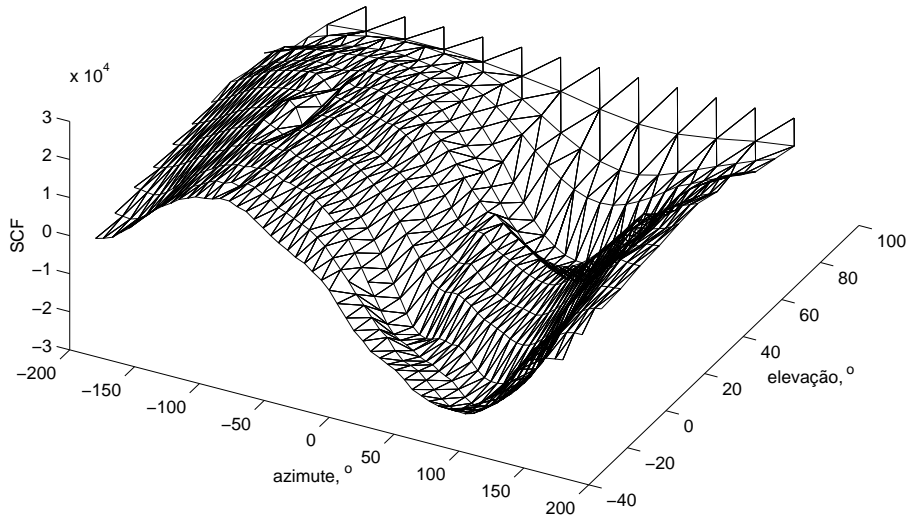


Figura 6.4: Exemplo de aproximação de uma SCF por regiões triangulares.

execução, os valores das inclinações podem ser calculados *a priori* e armazenados em tabelas².

Deve-se tomar cuidado com o erro acumulativo que pode ser gerado ao longo do cálculo do valor atual de $\omega_j(\theta, \phi)$. Como pode ser observado na Figura 6.5, esse erro é gerado principalmente quando a fonte virtual se desloca de uma região triangular para outra vizinha, já que a aproximação das derivadas pelas inclinações das faces triangulares implica descontinuidade nas bordas das regiões. Esse problema pode ser solucionado recalculando-se a estimativa inicial para a nova região diretamente através do método triangular (Equação (2.4)) sobre as SCFs associadas aos vértices A, B e C:

$$\omega_j(\theta, \phi) = w_A \omega_j(\theta_A, \phi_A) + w_B \omega_j(\theta_B, \phi_B) + w_C \omega_j(\theta_C, \phi_C), \quad (6.10)$$

onde w_A , w_B e w_C são calculadas pelas Equações (2.5)–(2.7).

Portanto, o algoritmo para obtenção das SCFs para as posições sobre o ca-

²O modelo em questão leva em conta que a distância entre a fonte virtual e ouvinte é bem maior que as dimensões da cabeça. Tal condição, se garantida, permite o modelamento em separado dos efeitos gerados pela alteração na distância. Nessas condições, basta a referência para a posição angular (0, 0) estar sempre exatamente à frente do ouvinte para que qualquer movimento de rotação do ouvinte ou translação da fonte seja tratado coerentemente.

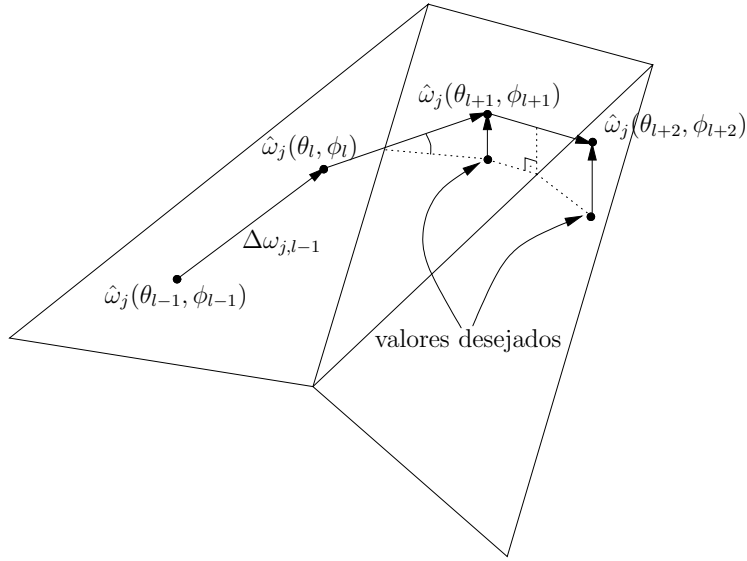


Figura 6.5: Erro acumulado ao se mudar de região.

minho da fonte virtual pode ser resumido como se segue:

1. Para a posição inicial ou se a fonte virtual estiver entrando em uma região triangular diferente da anterior, calculam-se as primeiras SCFs através da Equação (6.10).
2. Se a próxima posição for dentro da mesma região, atualizam-se as SCFs através da Equação (6.8). Se não, retorna-se ao primeiro passo.
3. Retorna-se ao segundo passo.

A vantagem desse método incremental sobre a interpolação por *splines* pode ser facilmente verificada. Primeiramente, desconsideram-se as operações relacionadas com os filtros fixos mostrados na Figura 6.2, que são as mesmas em ambos os casos. Assim, o método incremental precisa somente de $3M$ operações de multiplicação e $5M$ operações de soma por amostra por fonte virtual, enquanto a complexidade da interpolação pela *spline* já mencionada é da ordem de $[(4 + \gamma)(n - k) + (k - 1)]M \approx (4 + \gamma)(n - k)M$ (bem mais que $2800M$, nesse caso) por canal por amostra por fonte virtual. Além disso, se considerarmos o número de multiplicações por amostra para F fontes envolvidas na interpolação linear em regiões triangulares aplicada diretamente sobre as HRIRs da Seção 2.2 igual a

$$C_{\text{triangular}} = (3N + 6)F \quad (6.11)$$

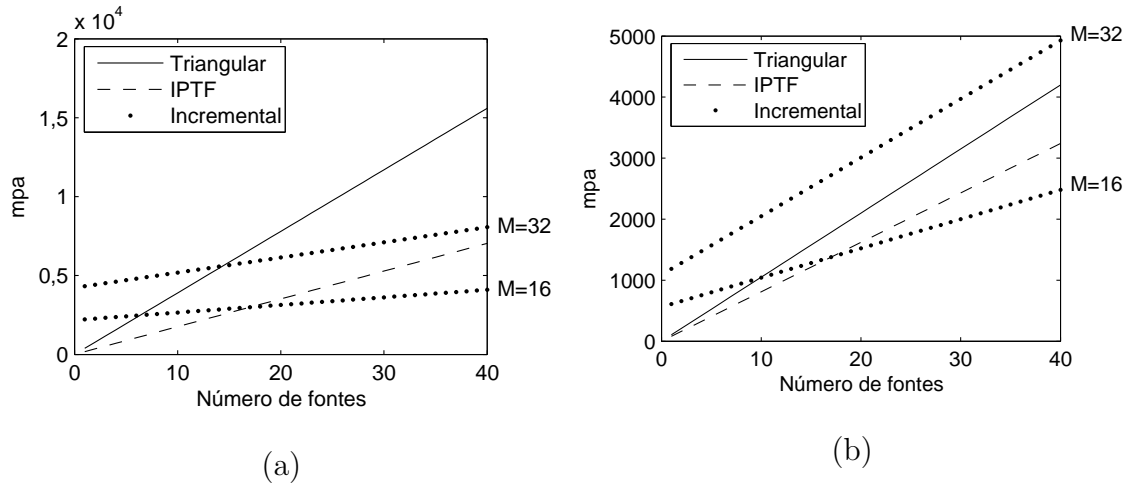


Figura 6.6: Comparação da complexidade computacional com o aumento do número de fontes: (a) com $N = 128$ e (b) com $N = 33$.

e o número de multiplicações por amostra para F fontes envolvidas na interpolação através das SCFs como sendo

$$C_{\text{SCF}} = 3MF + (M + 1)N, \quad (6.12)$$

fazendo $C_{\text{triangular}} > C_{\text{SCF}}$ tem-se que

$$\begin{aligned} (3N + 6)F &> 3MF + (M + 1)N \Rightarrow \\ \Rightarrow (3N - 3M + 6)F &> (M + 1)N \Rightarrow \\ \Rightarrow F &> \frac{(M + 1)N}{3(N - M + 2)}, \end{aligned} \quad (6.13)$$

o que indica que se o número de fontes sonoras virtuais processadas ultrapassar $\frac{(M+1)N}{3(N-M+2)}$, o método incremental proposto alcança uma complexidade computacional menor que a interpolação bilinear em regiões triangulares aplicada diretamente sobre HRIRs. Por exemplo, com $N = 128$ e $M = 32$, esse número limite equivale a 14 (veja a Figura 6.6a), o que, no contexto de modelagem de reflexões próximas através do processo fonte-imagem [33, 15], não é considerado um número elevado.

Na Figura 6.6, mostram-se dois estudos da variação da complexidade computacional, em termos do número de multiplicações envolvido em cada um dos tipos de interpolação, com o aumento do número de fontes virtuais geradas simultaneamente. Em ambos os casos, tem-se a variação do número de multiplicações para a interpolação triangular, a que utiliza IPTFs e dois casos da incremental (com $M = 16$ e $M = 32$, que equivalem a se utilizar 99% e 99,9% da energia dos autovetores, respectivamente).

Nessa figura, pode-se perceber que a maioria das retas para a interpolação incremental tem inclinação menor que as dos outros dois métodos. Isso porque a complexidade quando a ordem dos filtros e o número de fontes aumentam é $O(NF)$ para o caso triangular direto sobre as HRIRs ou com IPTFs, e $O(N + F)$ para o caso incremental, sendo F o número de fontes virtuais.

No caso de se ter um número de fontes da ordem de algumas dezenas e HRIRs com $N = 128$ (Figura 6.6a), nota-se uma grande vantagem ao se utilizar a interpolação incremental, que para $M = 32$ já tem resultado suficientemente acurado em relação à interpolação triangular, como será visto na próxima seção.

Já na Figura 6.6b, tem-se uma comparação entre a complexidade de cada método caso seja possível que o número de multiplicações requeridas para cada filtro $\psi_j(k)$ seja igual ao que se consegue modelando-se as HRTFs na forma de um filtro recursivo ($N = 33$)³. Nota-se que só existirá essa vantagem se o número de fontes for bem maior que no caso anterior ou se $M = 16$, atribuindo-se, com isso, menor acurácia ao modelo. Deve ser notado também que, principalmente nesse último caso citado, a interpolação com IPTFs pode ser ainda considerada bastante competitiva.

6.3 Comparação de Desempenho

As Figuras 6.7–6.10 mostram algumas comparações entre a interpolação triangular direta de HRTFs (coluna esquerda) e a interpolação incremental de SCFs (coluna central), usando novamente as SFRSs [28]. A diferença entre as SFRSs pode ser vista na terceira coluna, como nos capítulos anteriores. Pode-se notar que o erro entre as duas formas de interpolação permanece pequeno. Como no capítulo anterior, a Figura 6.11 mostra um histograma para o erro relativo, agora entre o método incremental e o triangular, dado por

$$\xi(\theta, \phi, f) = 20 \log_{10} \left(\frac{|H_I(\theta, \phi, f)|}{|H_T(\theta, \phi, f)|} \right), \quad (6.14)$$

³No caso de a interpolação ser da forma bilinear ou com IPTFs, isso é possível, como comentado nos capítulos anteriores. Já no caso da interpolação das SCFs, está-se admitindo a possibilidade de os autovetores obtidos para as HRIRs poderem ser modelados da mesma forma, sendo necessárias 33 multiplicações para computar a saída de cada filtro correspondente.

calculado em $-180^\circ \leq \theta < 180^\circ$ e $-40^\circ \leq \phi \leq 90^\circ$ com passo de 1° e nas frequências escolhidas para se obter as SFRSs. $|H_T(\theta, \phi, f)|$ e $|H_I(\theta, \phi, f)|$ são amostras das SFRSs mostradas nas Figuras 6.7–6.10 para as interpolações triangular direta e incremental, respectivamente. Esse histograma mostra que mais de 98% dos resultados calculados apresentam um erro menor que 0,729 dB, confirmando que o método proposto é uma alternativa computacionalmente vantajosa para a implementação de som 3D binaural com múltiplas fontes. Comparando esse histograma com o da Figura 5.7, pode-se notar que o método de interpolação incremental com a KLT gera menos erro que o método que utiliza IPTFs de ordem reduzida.

6.4 Conclusões

Nesse capítulo, mostrou-se um método que realiza a interpolação das funções espaciais características (SCFs) de forma linear incremental, como proposto em [P5], como substituta para as *splines*. Sua validade foi mostrada através de comparações com a interpolação triangular feita diretamente sobre as HRIRs. A interpolação das SCFs, já conhecida pela baixa complexidade ao ser utilizada em um ambiente com múltiplas fontes, é comparada também com relação ao desempenho contra a interpolação triangular direta e com as IPTFs. É indicado o número mínimo de fontes que resultam em ganho na complexidade computacional. Além disso, são mostrados dois gráficos contendo a variação de complexidade com o aumento do número de fontes, que evidenciam esse comportamento e ainda indicam a competitividade da interpolação com IPTFs.

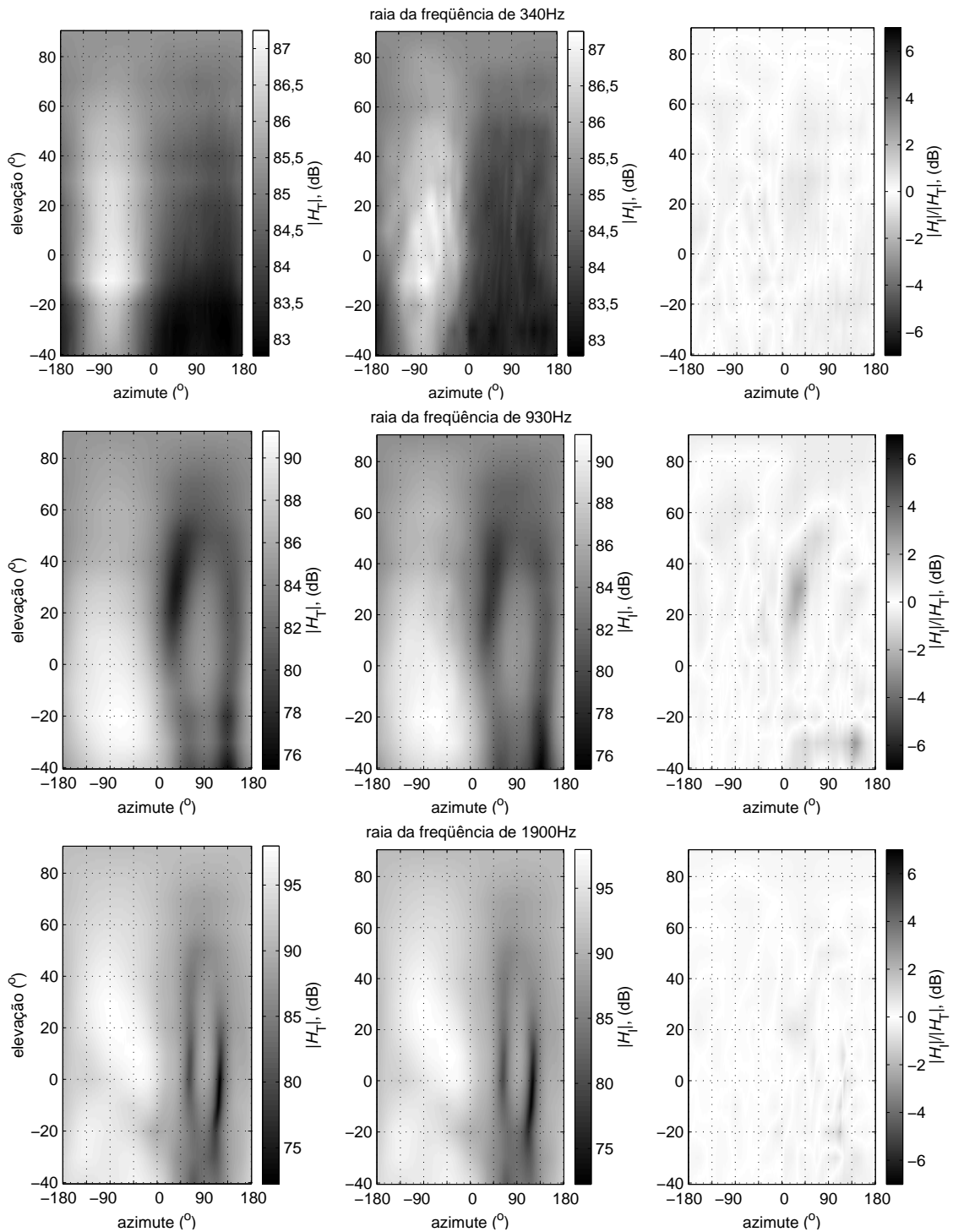


Figura 6.7: Comparação entre SFRSs. Triangular direta (esquerda), Incremental (centro) e diferença em dB (direita)—I. De cima para baixo: 340 Hz, 930 Hz e 1900 Hz.

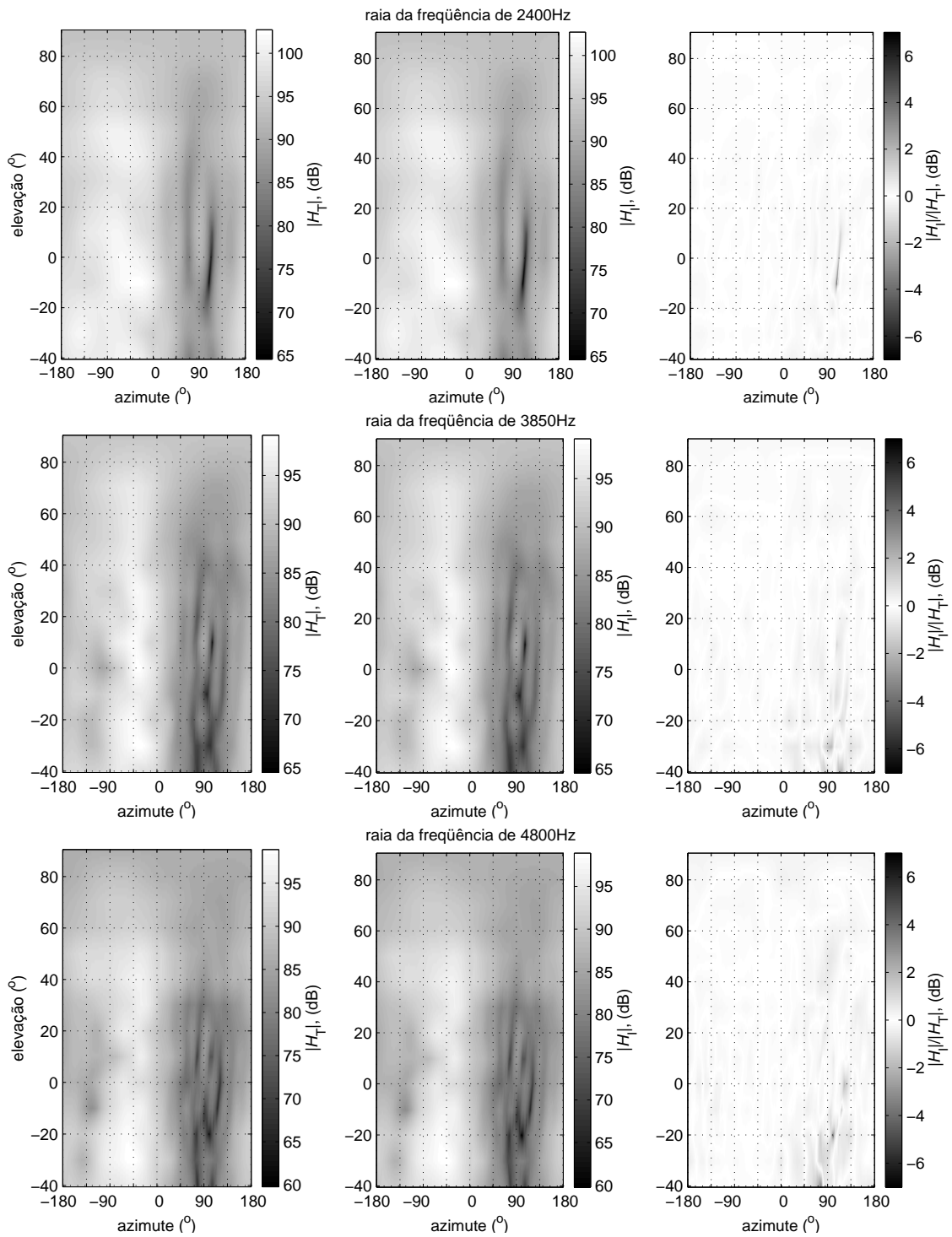


Figura 6.8: Comparação entre SFRSs. Triangular direta (esquerda), Incremental (centro) e diferença em dB (direita)—II. De cima para baixo: 2400 Hz, 3850 Hz e 4800 Hz.

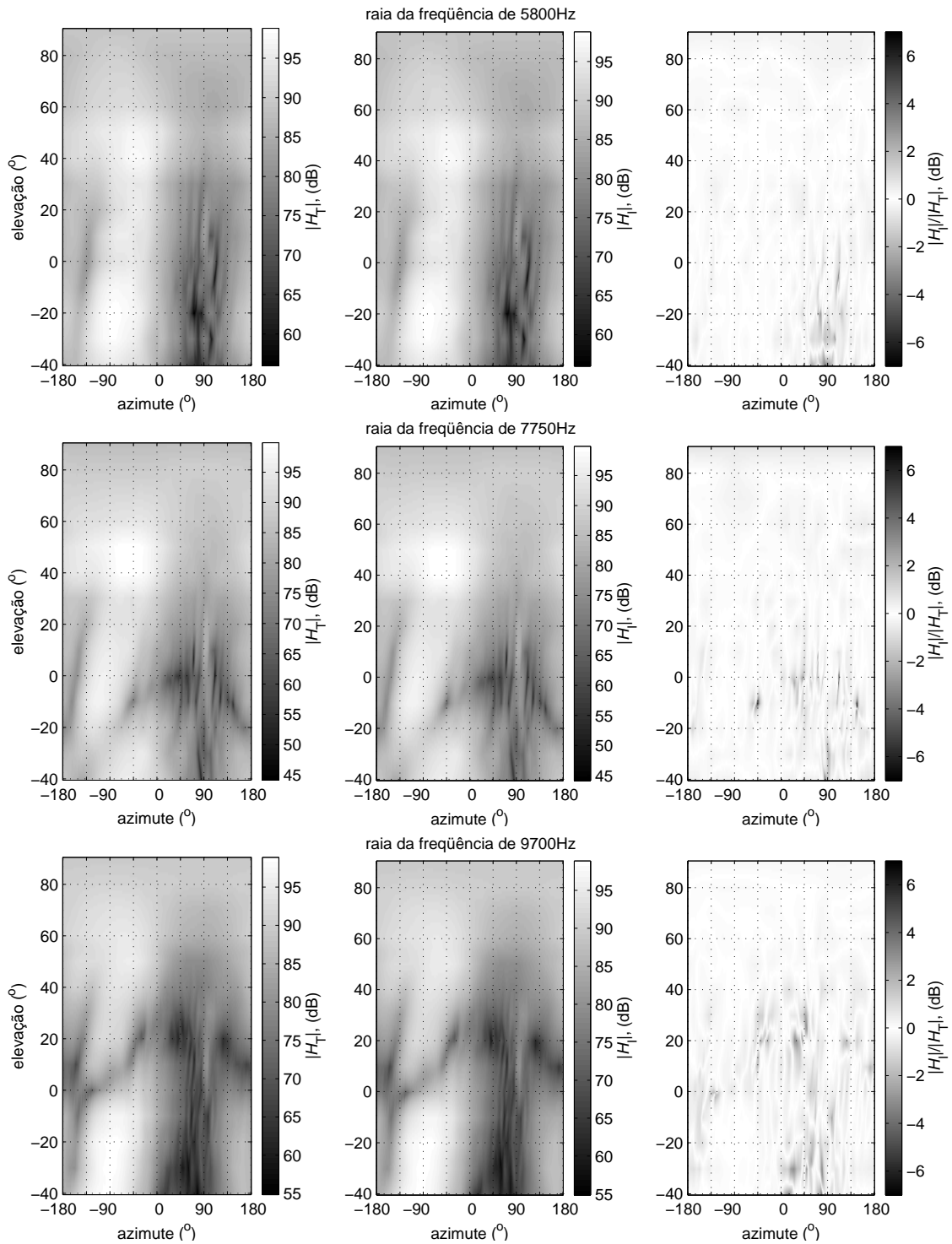


Figura 6.9: Comparação entre SFRSs. Triangular direta (esquerda), Incremental (centro) e diferença em dB (direita)—III. De cima para baixo: 5800 Hz, 7750 Hz e 9700 Hz.

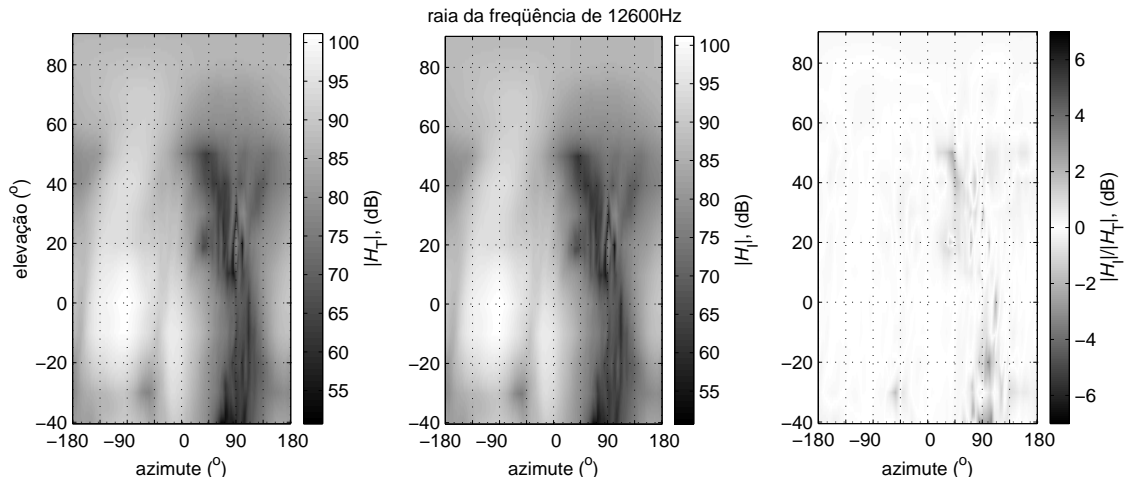


Figura 6.10: Comparação entre SFRSs. Triangular direta (esquerda), Incremental (centro) e diferença em dB (direita)—IV. 12600 Hz.

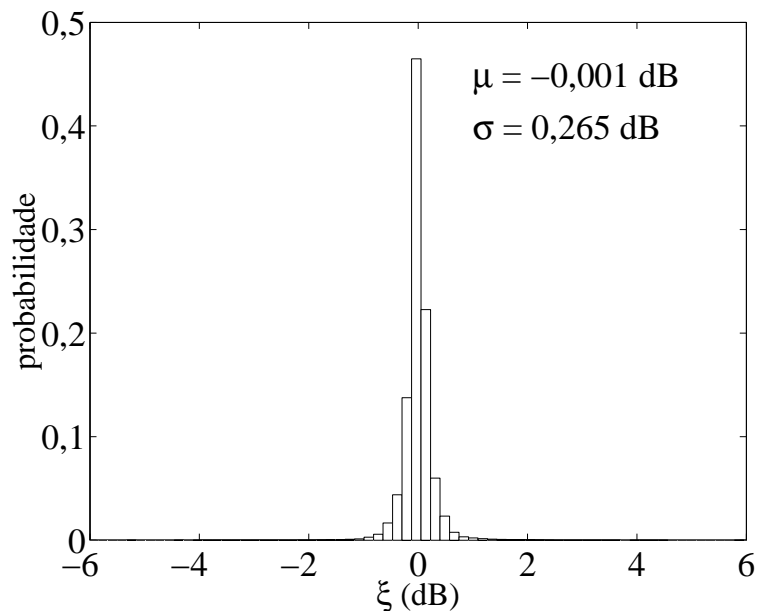


Figura 6.11: Histograma para o erro ξ (interpolação incremental com a KLT).

Capítulo 7

Estrutura Farrow Generalizada para Modelo de HRTF

Como visto anteriormente, a modelagem binaural pode ser realizada sem a utilização direta das HRTFs. No capítulo anterior, as HRTFs foram descritas através dos coeficientes da KLT relativos às HRTFs. No entanto, menos direto ainda, porém mais intuitivo, seria um modelo no qual a resposta em frequência de um certo filtro fosse controlada diretamente pelos parâmetros de posição e cuja resposta correspondesse a uma boa aproximação da resposta de uma HRTF.

Tendo como objetivo principal a implementação de um atraso digital continuamente variável (CVDD—*Continuously Variable Digital Delay*), em [29], Farrow apresenta uma estrutura capaz de calcular a cada instante uma amostra interpolada do sinal de entrada.

Nesse capítulo, após uma breve apresentação da estrutura de Farrow na Seção 7.1, desenvolve-se um estudo com o intuito de se obter um modelo para as HRTFs com base naquela estrutura. Na Seção 7.2, é proposto um modelo para um conjunto de HRIRs que é controlado diretamente pelas variáveis espaciais (θ, ϕ) , partindo de uma formulação da estrutura Farrow generalizada para duas variáveis de controle. Na Seção 7.3, é proposta uma maneira de se obterem os coeficientes da estrutura Farrow generalizada com o objetivo de reduzir a quantidade de memória necessária. Em seguida, mostra-se na Seção 7.4 a comparação entre o método de interpolação com a estrutura Farrow e a interpolação triangular através das SFRSs. Por fim, são apresentadas as conclusões do presente capítulo.

7.1 Estrutura Farrow

Um problema bastante interessante em Processamento Digital de Sinais é a realização do processamento sobre uma versão atrasada de um sinal quando o atraso não é múltiplo do período de amostragem. Um exemplo no qual esse tipo de problema ocorre é o cancelamento de eco [56], onde o atraso essencial para o sucesso do procedimento quase sempre não é múltiplo do período de amostragem. Em outras aplicações em áudio, como a correção de taxa da amostragem, a variação de tempo de execução e a variação de afinação de sinais de áudio [57], o atraso fracionário também pode ser utilizado.

Uma estrutura proposta por Farrow em [29] consegue não só simular um atraso constante qualquer como gerar um sinal com uma taxa de amostragem diferente da taxa do sinal de entrada, bastando, para isso, alterar de forma adequada o parâmetro α de controle instantâneo do atraso a ser aplicado, como é mostrado nas Figuras 7.1a, b e c. Na Figura 7.1a, têm-se o sinal analógico e sua versão amostrada com período T seguidos pelo possível resultado de um atraso fracionário constante (mantendo a mesma taxa de amostragem T), visto na Figura 7.1b, e por um sinal para o qual fora realizada a alteração da taxa de amostragem para T_o na Figura 7.1c.

Esse procedimento, desenvolvido em [29] e discutido também em [58], modela um atraso fracionário τ baseando-se em um polinômio na variável α , onde $\tau = \alpha T$ (T é o período de amostragem do sinal de entrada).

Portanto, para se obter os coeficientes da estrutura Farrow que gera o atraso desejado dependente de α , consideremos a função de transferência de um atraso τ dada por

$$G(\omega, \tau) = e^{-j\omega\tau}. \quad (7.1)$$

Já que um filtro FIR cujo período de amostragem é T tem resposta na frequência igual a

$$F(\omega) = \sum_n C_n e^{-jn\omega T}, \quad (7.2)$$

onde $n = 0, 1, \dots, N$, obtendo-se os coeficientes C_n por polinômios dependentes de α , tem-se

$$F(\omega, \alpha) = \sum_n C_n(\alpha) e^{-jn\omega T}, \quad (7.3)$$

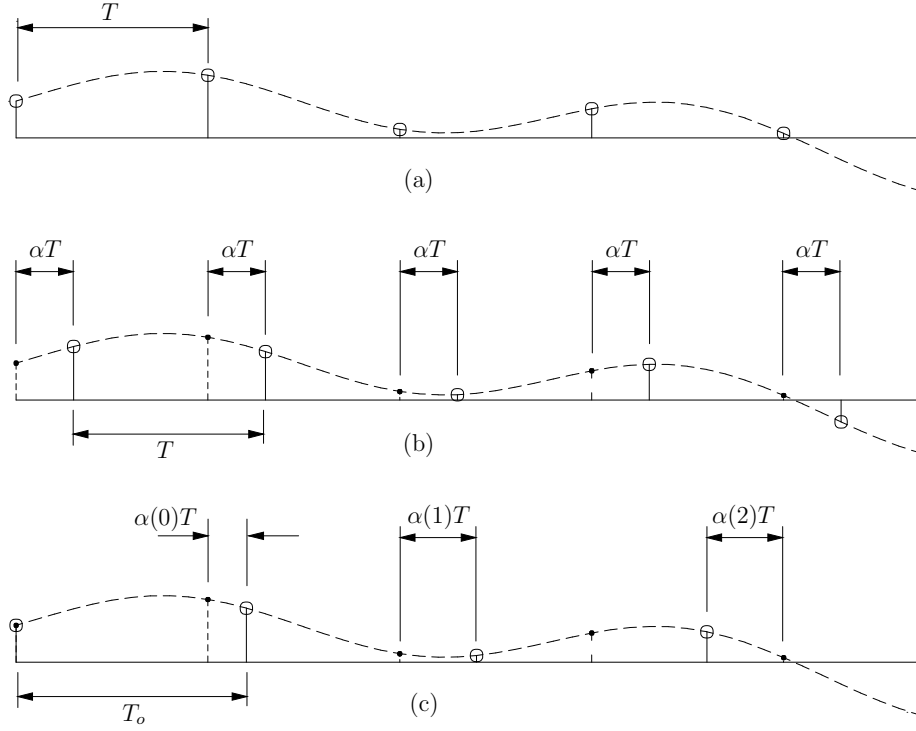


Figura 7.1: Exemplos de atraso fracionário (b) e mudança de taxa de amostragem (c) a partir do sinal original (a).

onde

$$C_n(\alpha) = \sum_l C_{n,l} \alpha^l, \quad (7.4)$$

com $l = 0, 1, \dots, L$. Dessa forma, de (7.3) pode-se obter por manipulação algébrica uma descrição mais direta da estrutura Farrow da Figura 7.2, dada por

$$F(\omega, \alpha) = \sum_{l=0}^L \alpha^l \sum_{n=0}^N C_{n,l} e^{-jn\omega T}. \quad (7.5)$$

Agora, os coeficientes $C_{n,l}$ podem ser otimizados para minimizar a expressão

$$\int_{-\pi}^{\pi} \int_{-\frac{1}{2}}^{\frac{1}{2}} |F(\omega, \alpha) - G(\omega, \tau)|^2 d\alpha d\omega. \quad (7.6)$$

Nesse capítulo, esse método de modelagem é aplicado sobre o conjunto de HRTFs, o que implica uma alteração no polinômio utilizado, já que agora existem duas variáveis de controle: uma para a elevação e outra para o azimute. Como o número de HRTFs é finito, outra alteração é feita na forma de otimização, que passa a ser discreta.

É importante ser mencionado que em [29] o parâmetro α é limitado no intervalo $[-\frac{1}{2}, \frac{1}{2})$ e em [58, 59], no intervalo $[-1, 1)$. Para o caso da interpolação de

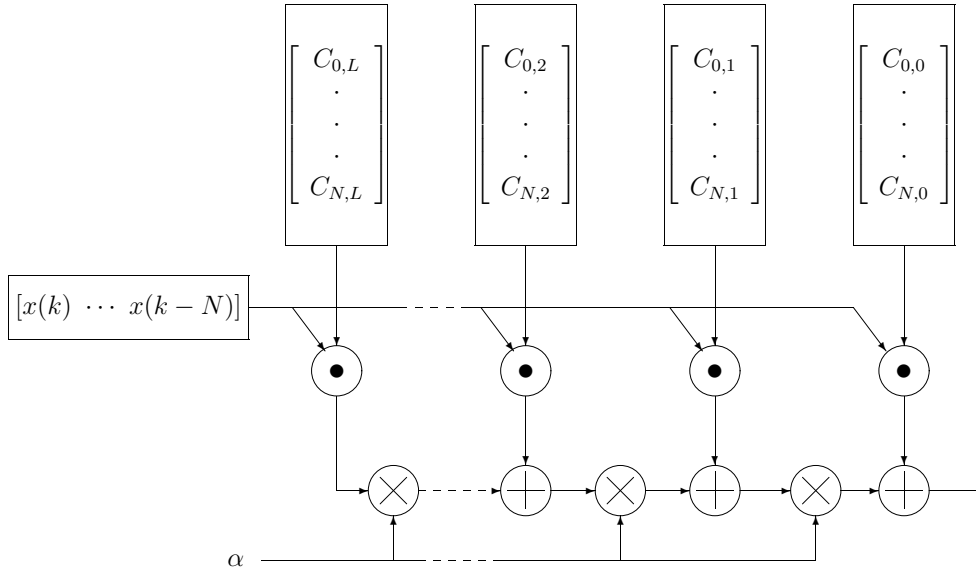


Figura 7.2: Diagrama em blocos da estrutura Farrow. \odot representa a operação de produto interno.

HRIRs, permitiu-se que os parâmetros θ e ϕ assumissem valores do intervalo $[-\pi, \pi]$. No caso da elevação, foi realizado um mapeamento que escala o intervalo $[-40^\circ, 90^\circ]$ para o intervalo $[-\pi, \pi]$. Essa escolha teve como base a observação da faixa dinâmica dos coeficientes ao fim da computação dos coeficientes. Escolheu-se o mapeamento de forma que a faixa dinâmica não aumentasse muito ao se aumentar a ordem do polinômio em θ e ϕ .

7.2 Formulação Bidimensional

A resposta na frequência de um filtro FIR pode ser descrita por

$$G(e^{j\omega}) = \sum_n h_n e^{-j\omega n}. \quad (7.7)$$

Considerando que cada HRTF é representada por um filtro FIR, a resposta em frequência do conjunto de HRTFs fica, então, dependente das coordenadas de posição (θ, ϕ) , e pode ser representada por

$$G(e^{j\omega}, \theta, \phi) = \sum_n h_n(\theta, \phi) e^{-j\omega n}. \quad (7.8)$$

Seguindo o mesmo princípio da estrutura Farrow, pode-se aproximar os coe-

ficientes $h_n(\theta, \phi)$ por um polinômio dado por

$$c_n(\theta, \phi) = \sum_l \sum_m \theta^l \phi^m c_n^{l,m}, \quad (7.9)$$

obtendo-se

$$F(e^{j\omega}, \theta, \phi) = \sum_l \sum_m \theta^l \phi^m \sum_n c_n^{l,m} e^{-j\omega n}. \quad (7.10)$$

como aproximação para $G(e^{j\omega}, \theta, \phi)$.

Uma maneira de se obter os coeficientes $c_n^{l,m}$, é minimizar a função do erro

$$J = \sum_{\omega} \sum_{\theta} \sum_{\phi} |\varepsilon(e^{j\omega}, \theta, \phi)|^2, \quad (7.11)$$

onde

$$\varepsilon(e^{j\omega}, \theta, \phi) = F(e^{j\omega}, \theta, \phi) - G(e^{j\omega}, \theta, \phi). \quad (7.12)$$

Dessa forma, como descrito no Apêndice C, chega-se a uma expressão para J na forma

$$J = \sum_{\omega} \sum_{\theta} \sum_{\phi} \left[\left(\sum_l \sum_m \theta^l \phi^m \sum_n c_n^{l,m} \cos \omega n T - \sum_k h_k(\theta, \phi) \cos \omega k T \right)^2 + \left(- \sum_l \sum_m \theta^l \phi^m \sum_n c_n^{l,m} \sin \omega n T + \sum_k h_k(\theta, \phi) \sin \omega k T \right)^2 \right]. \quad (7.13)$$

A derivada de J com relação a cada coeficiente $c_{n_0}^{l_0, m_0}$ (Apêndice C) fica na forma

$$\frac{\partial J}{\partial c_{n_0}^{l_0, m_0}} = 2 \sum_l \sum_m \sum_n \left\{ \sum_{\omega} \sum_{\theta} \sum_{\phi} \theta^{l_0} \phi^{m_0} \theta^l \phi^m c_n^{l,m} \cos(\omega n_0 T - \omega n T) \right\} - 2 \sum_k \sum_{\omega} \sum_{\theta} \sum_{\phi} \theta^{l_0} \phi^{m_0} h_k(\theta, \phi) \cos(\omega n_0 T - \omega k T). \quad (7.14)$$

Na forma matricial,

$$\vec{\nabla} J = \mathbf{X} \mathbf{c} - \mathbf{p}, \quad (7.15)$$

onde os elementos da matriz \mathbf{X} e do vetor \mathbf{p} são

$$x_{i,j} = \sum_{\omega} \sum_{\theta} \sum_{\phi} \theta^{v(j)} \phi^{\mu(j)} \theta^{v(i)} \phi^{\mu(i)} \cos(\omega \eta(j) T - \omega \eta(i) T) \quad (7.16)$$

e

$$p_i = \sum_k \sum_{\omega} \sum_{\theta} \sum_{\phi} \theta^{v(i)} \phi^{\mu(i)} h_k(\theta, \phi) \cos(\omega \eta(i) T - \omega k T), \quad (7.17)$$

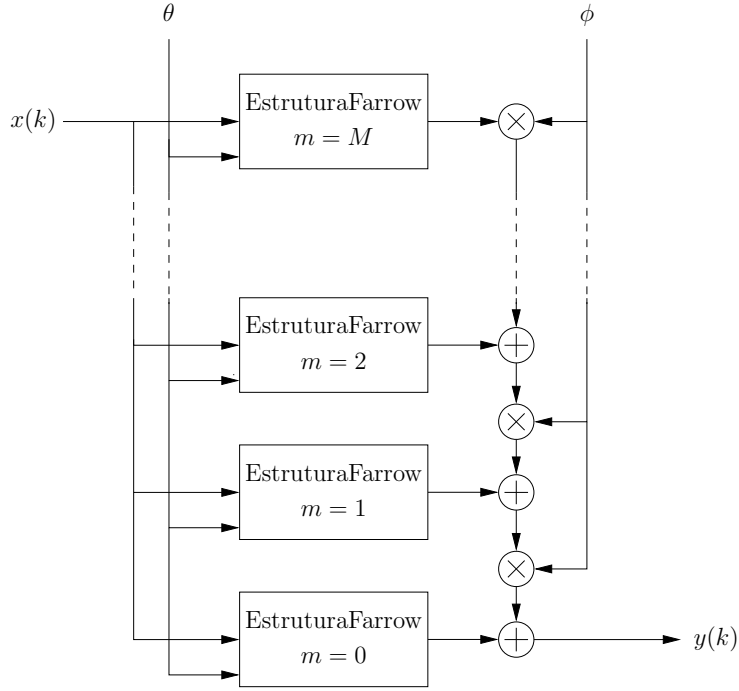


Figura 7.3: Diagrama em blocos da estrutura Farrow generalizada.

respectivamente, sendo $i = j = 0, 1, \dots, (N + 1)(L + 1)(M + 1)$,

$$\eta(a) = \text{mod} [a, (N + 1)], \quad (7.18)$$

$$v(a) = \text{mod} \left[\left\lfloor \frac{a}{N + 1} \right\rfloor, (L + 1) \right] \quad (7.19)$$

e

$$\mu(a) = \left\lfloor \frac{a}{(N + 1)(L + 1)} \right\rfloor, \quad (7.20)$$

sendo $\lfloor \cdot \rfloor$ o maior inteiro menor ou igual a (\cdot) e $\text{mod}(a, b)$ o resto da divisão $\frac{a}{b}$. Por fim, \mathbf{c} é o vetor formado pelos coeficientes $c_n^{l,m}$ de forma que seus elementos sejam

$$c_i = c_{\eta(i)}^{v(i), \mu(i)}, \quad (7.21)$$

com $l = 0, 1, \dots, L$, $m = 0, 1, \dots, M$ e $n = 0, 1, \dots, N$.

Assim, igualando-se o vetor $\overrightarrow{\nabla J}$ a zero, tem-se

$$\mathbf{c} = \mathbf{X}^{-1} \mathbf{p}. \quad (7.22)$$

Na Figura 7.3, pode-se ver a estrutura Farrow generalizada para duas variáveis. Nela, cada bloco de índice m é uma estrutura Farrow de uma variável.

7.3 Otimização da Estrutura Generalizada

Através da Equação (7.22), vê-se a necessidade da inversão da matriz \mathbf{X} para se obterem os coeficientes da estrutura Farrow. Uma das características dessa estrutura é que ela é formada por um conjunto de filtros FIR com um determinado suporte temporal finito. Como o que se está pretendendo interpolar com essa estrutura são filtros FIR, nota-se que, ao se atribuir uma ordem N menor que a desses filtros, acaba-se por introduzir um erro próximo ao de um truncamento da resposta desejada. Para não gerar tal erro, optou-se por manter o suporte temporal das HRIRs, o que significa fazer $N = 127$.

A ordem de \mathbf{X} é igual a $(N + 1)(M + 1)(L + 1)$. Desse modo, vê-se que a inversão da matriz \mathbf{X} não é trivial, exigindo um número de operações muito grande, além de o mau condicionamento da matriz poder tornar o cálculo de sua inversa impreciso.

Durante os testes desse método de obtenção dos coeficientes através da inversão da matriz \mathbf{X} , percebeu-se o tempo demasiado grande para resolução da inversa, que em casos extremos se tornava impraticável. Além disso, nos casos em que se conseguiu chegar a um resultado cuja acurácia não era reduzida pelo mau condicionamento de \mathbf{X} , a ordem dos polinômios não era suficiente para se obter uma interpolação cujo resultado fosse comparável ao obtido com a interpolação triangular. Por exemplo, utilizando-se $M = 3$ e $L = 3$ (ordem de \mathbf{X} igual a 2.048) conseguiu-se obter o resultado da inversa sem que o algoritmo de inversão acusasse mau condicionamento. Mas, com essa ordem para os polinômios, as SFRSs geradas eram muito diferentes das obtidas com a interpolação triangular. Por esse motivo, uma outra maneira de se realizar a otimização dos coeficientes teve de ser proposta.

A primeira tentativa foi utilizar o método iterativo de otimização BFGS (Broyden-Fletcher-Goldfarb-Shanno) [55] para encontrar o mínimo global do erro no domínio da frequência, descrito pela Equação (7.11). Nesse caso, o método esbarrou na quantidade de memória disponível nas máquinas¹ quando se tentava aumentar a ordem dos polinômios além de $M = 6$ e $L = 6$. Apenas as matrizes envolvidas na otimização chegavam a ocupar, no mínimo, algo em torno de 1 GB (um giga *byte*) de

¹A máquina utilizada com maior quantidade de memória tinha 2 GB.

memória, considerando que são utilizadas 3 matrizes da mesma ordem que a matriz \mathbf{X} , com representação de 8 *bytes* para cada elemento $(3 \times (7 \times 7 \times 128))^2 \times 8 \approx 1$ GB). Como não se obteve um resultado que pudesse ser considerado razoável pela comparação feita pelas SFRSs, decidiu-se realizar a otimização de uma outra forma.

7.3.1 Otimização para Cada Instante

Como o suporte temporal dos filtros é mantido igual ao das HRIRs, é possível escrever a resposta ao impulso resultante da estrutura Farrow como

$$\hat{h}_n(\theta, \phi) = \sum_{m=0}^{M-1} \sum_{l=0}^{L-1} c_n^{l,m} \theta^l \phi^m, \quad (7.23)$$

o que significa que a amostra no instante n dessa resposta ao impulso depende somente dos coeficientes correspondentes a esse mesmo instante. Esse fato mostra que se poderia reduzir a ordem das matrizes envolvidas na otimização por $(N + 1)$.

De posse desse dado, conseguiu-se, através do mesmo método iterativo utilizado anteriormente, obter-se um conjunto de coeficientes para polinômios de ordem maior, sem atingir o limite de memória mencionado.

Agora, o erro a ser minimizado deve ser descrito como

$$J'_n = \sum_{\theta} \sum_{\phi} \left| \hat{h}_n(\theta, \phi) - h_n(\theta, \phi) \right|^2. \quad (7.24)$$

O próximo passo foi escolher qual a ordem do polinômio a ser utilizada para a interpolação. Para tanto, elegeram-se alguns valores para L e M e obteve-se para cada instante de tempo o erro normalizado pelo valor da norma das amostras do instante correspondente. Na Figura 7.4, tem-se a norma das amostras das HRIRs para cada instante de tempo.

Escolheram-se $L = 6, 9, 12$ e 15 e $M = 6, 9, 12$ e 15 . Algumas combinações desses valores podem ser vistos na Figura 7.5, onde são mostrados os erros mínimos normalizados obtidos ao final da otimização pelo método BFGS para cada instante da resposta ao impulso.

Pode-se notar que quanto maior é a ordem do polinômio, menor fica o erro, como esperado. Comparando-se os gráficos dessa figura, pode-se notar que para os testes realizados, o melhor valor para L é 12, já que com os valores maiores para M só há aumento significativa no erro quando L passa de 12 para 9. Da mesma forma,

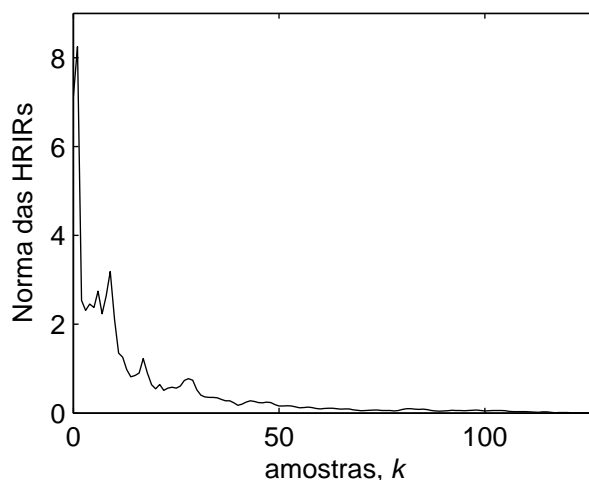


Figura 7.4: Norma das HRIRs de todas as posições (θ, ϕ) para cada instante.

observando-se o gráfico da Figura 7.5a, percebe-se que os erros só têm uma variação significativa quando M passa do valor 9 para 6. Assim, pode-se considerar $M = 9$ um bom valor de compromisso. Então, escolheram-se inicialmente para as ordens do polinômio $L = 12$ e $M = 9$.

No entanto, ao se observar as SFRSs obtidas com esses valores para L e M , notou-se que a escolha baseada no erro normalizado para cada instante de tempo não descreve adequadamente a qualidade da interpolação, já que os erros nas posições para as quais não se têm HRIRs medidas não são avaliados. Um exemplo disso pode ser visto na Figura 7.6, onde é mostrada a comparação entre as SFRSs referentes à interpolação triangular e à estrutura Farrow.

Pode-se notar nessa figura que a suavização esperada nas regiões interpoladas não é bem resolvida, já que aparece uma oscilação bem perceptível na direção das mudanças de elevação. Esse efeito parece indicar uma ordem de polinômio alta demais para interpolar um sinal suave com amostras muito espaçadas. Experimentou-se, então, reduzir os valores de M e L de forma a decidir por uma combinação mais adequada.

Comparando-se as SFRSs geradas para cada conjunto de coeficientes testado, percebeu-se que o comportamento nas regiões interpoladas apresentava um erro menor utilizando-se as ordens $L = 12$ e $M = 6$. Dessa forma, elegeram-se essas ordens para realizar as comparações com os outros métodos de interpolação, que podem ser vistas na seção seguinte.

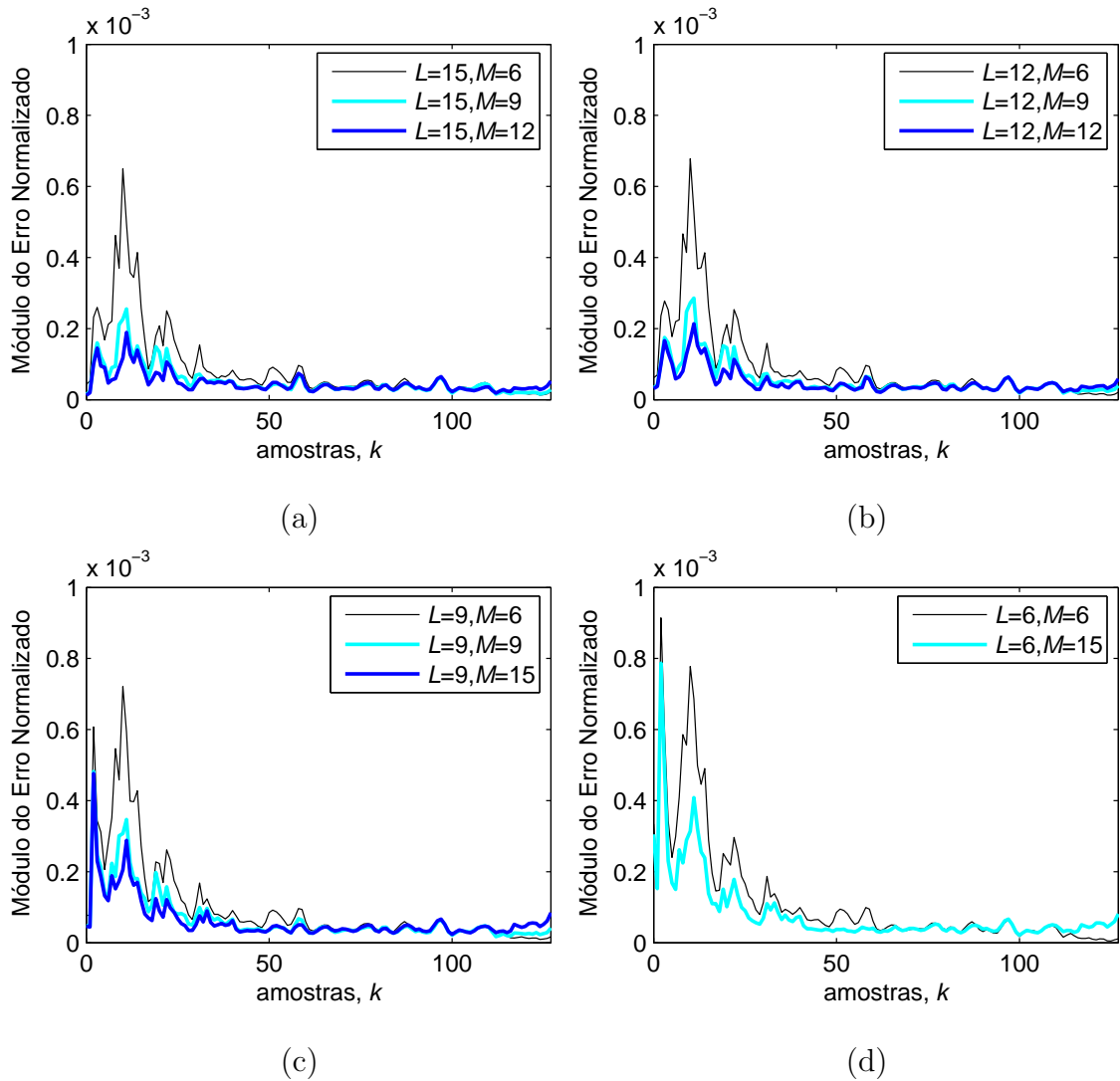


Figura 7.5: Erro normalizado mínimo obtido com o método iterativo de otimização para cada instante de tempo. (a) para $L = 15$, (b) para $L = 12$, (c) para $L = 9$ e (d) para $L = 6$.

7.4 Comparação de Desempenho

As Figuras 7.7–7.10 mostram algumas comparações entre a interpolação triangular direta de HRTFs (coluna esquerda) e a interpolação utilizando a estrutura Farrow (coluna central), usando novamente as SFRSs [28]. A diferença entre as SFRSs pode ser vista na terceira coluna, como nos capítulos anteriores.

Diferentemente das outras formas de interpolação, o resultado com a estrutura Farrow generalizada aparece muito suavizado, já que na escolha da ordem dos polinômios priorizou-se a redução da complexidade do modelo. Isso acaba gerando erros mais elevados nas regiões mais acidentadas, como as de baixa energia. Isso

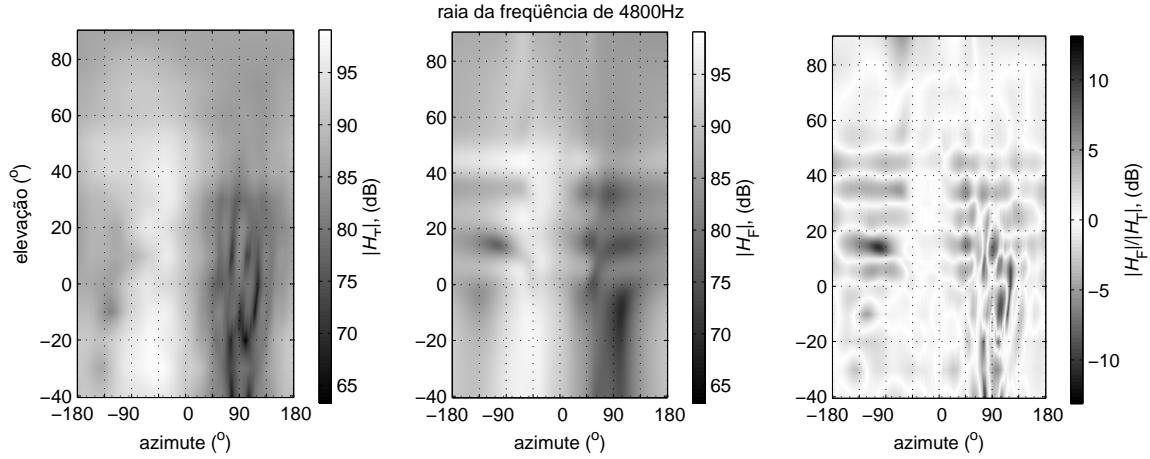


Figura 7.6: Comparação entre SFRSs. Erro devido a ordem dos polinômios da estrutura Farrow. Triangular direta (esquerda), Farrow (centro) e diferença em dB (direita). Para 4800 Hz.

pode ser verificado principalmente nas SFRSs para as frequências mais altas. No entanto, para as frequências baixas/médias, as superfícies mostram erros pequenos em quase todas as regiões. Nas conclusões do presente capítulo e no Capítulo 8, o efeito dos valores escolhidos para M e L é discutido.

Como nos capítulos anteriores, foi feito um histograma do erro relativo, agora entre o método com a estrutura Farrow generalizada e o triangular, dado por

$$\xi(\theta, \phi, f) = 20 \log_{10} \left(\frac{|H_F(\theta, \phi, f)|}{|H_T(\theta, \phi, f)|} \right), \quad (7.25)$$

calculado em $-180^\circ \leq \theta < 180^\circ$ e $-40^\circ \leq \phi \leq 90^\circ$ com passo de 1° e nas frequências escolhidas para se obter as SFRSs. $|H_T(\theta, \phi, f)|$ e $|H_F(\theta, \phi, f)|$ são amostras das SFRSs mostradas nas Figuras 7.7–7.10 para as interpolações triangular direta e com a estrutura Farrow generalizada, respectivamente. Esse histograma mostra que mais de 98% dos resultados calculados apresentam um erro menor que 6,2 dB, bem maior se comparado aos valores obtidos nos capítulos anteriores.

7.5 Conclusões

Nesse capítulo, fez-se uma breve descrição da estrutura Farrow. Além disso, foi mostrada uma generalização de sua formulação para se poder controlar a resposta da estrutura por duas variáveis e, com isso, obter um modelo completo para as HRIRs em função das variáveis de posição (θ, ϕ) .

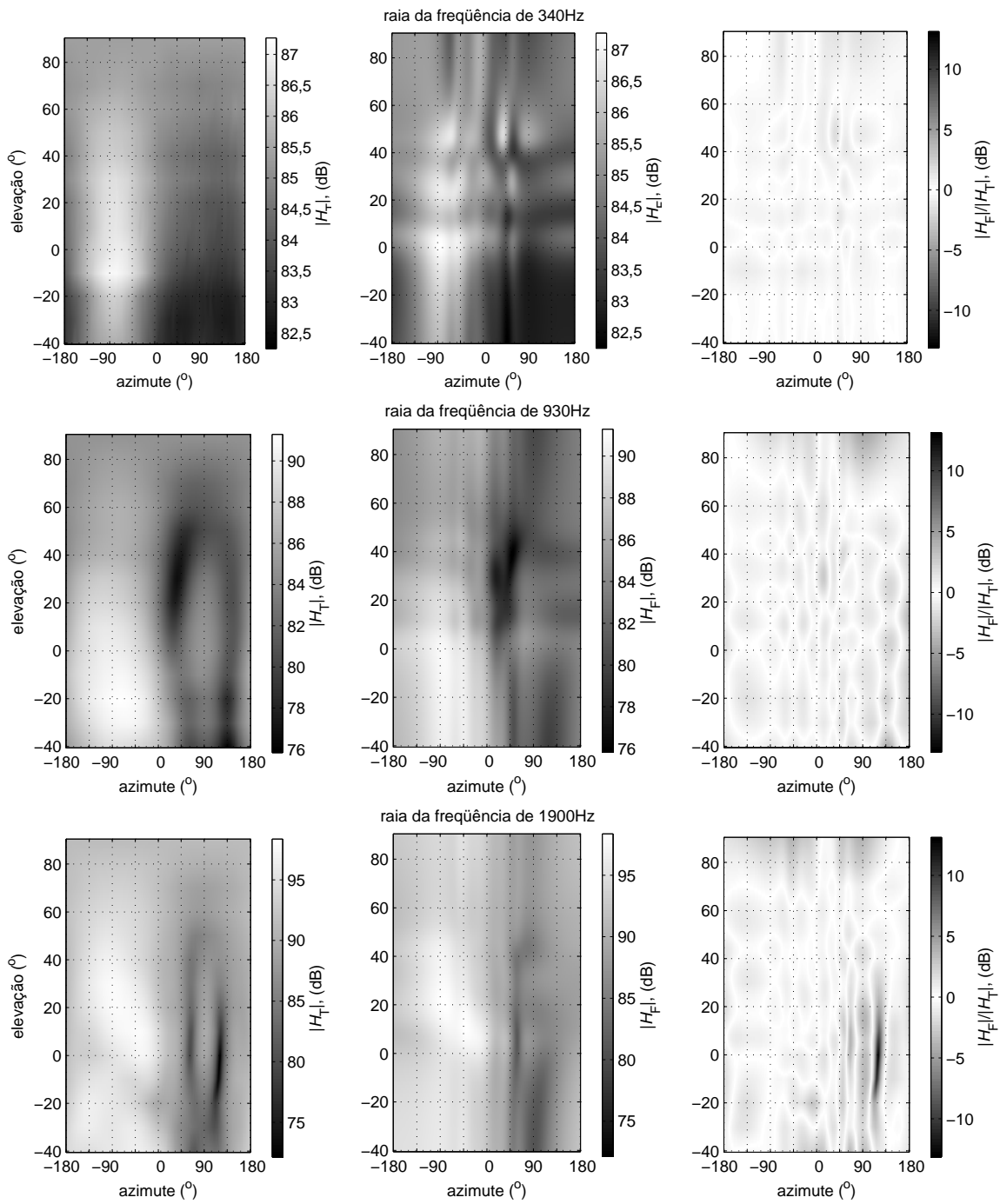


Figura 7.7: Comparação entre SFRSs. Triangular direta (esquerda), com a estrutura Farrow (centro) e diferença em dB (direita)—I. De cima para baixo: 340 Hz, 930 Hz e 1900 Hz.

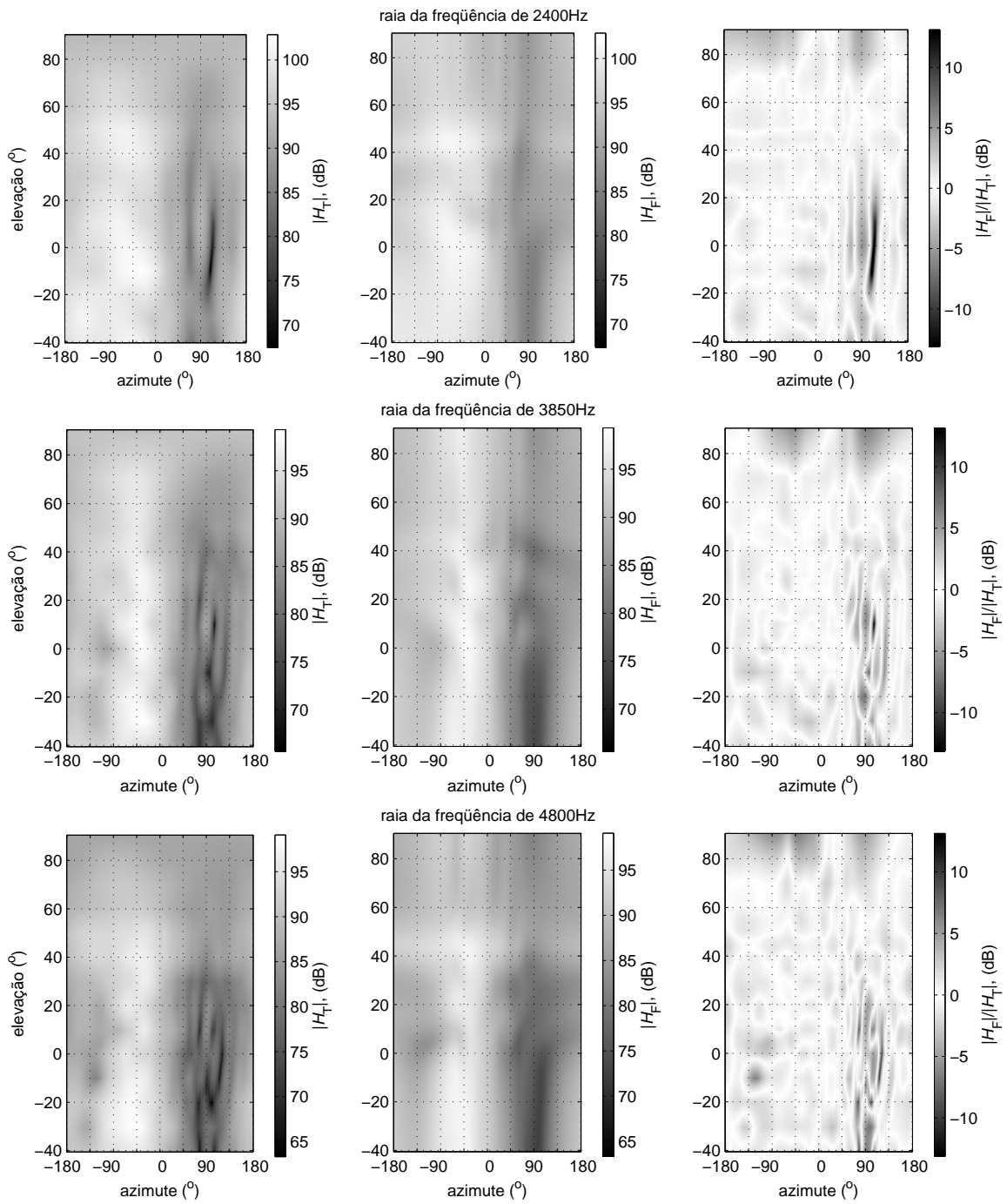


Figura 7.8: Comparação entre SFRSs. Triangular direta (esquerda), com a estrutura Farrow (centro) e diferença em dB (direita)—II. De cima para baixo: 2400 Hz, 3850 Hz e 4800 Hz.

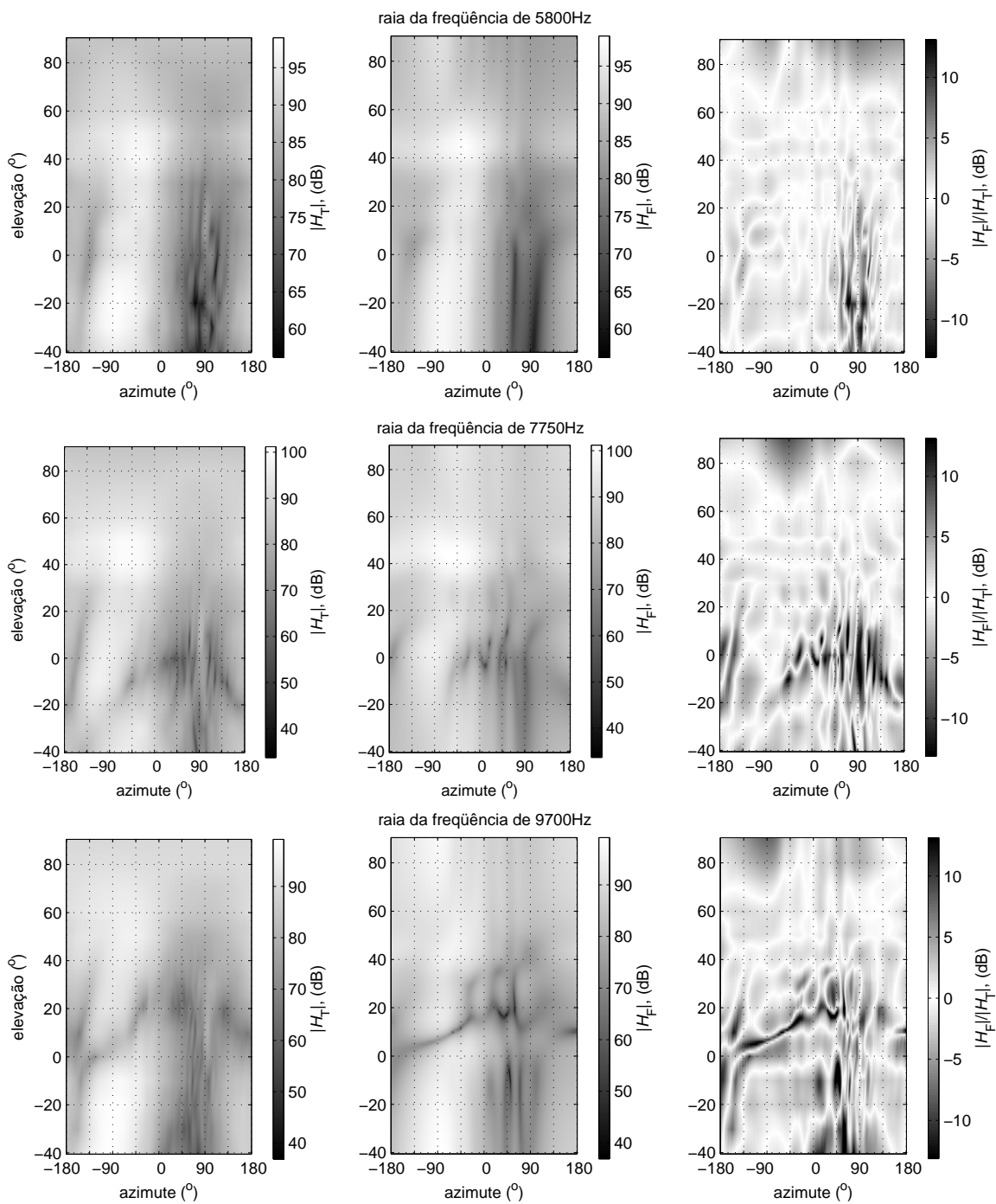


Figura 7.9: Comparação entre SFRSs. Triangular direta (esquerda), com a estrutura Farrow (centro) e diferença em dB (direita)—III. De cima para baixo: 5800 Hz, 7750 Hz e 9700 Hz.

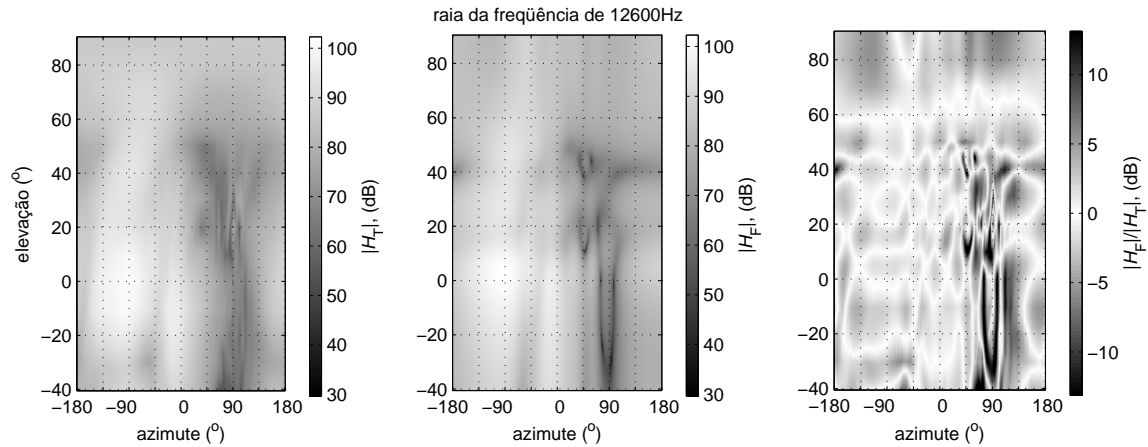


Figura 7.10: Comparação entre SFRSs. Triangular direta (esquerda), com a estrutura Farrow (centro) e diferença em dB (direita)—IV. 12600 Hz.

Parece possível matematicamente obter-se tal modelo. No entanto, problemas de precisão numérica e de necessidade de memória foram encontrados ao se tentar modelar o conjunto completo com polinômios de ordem mais elevada. Buscou-se, então, resolver esse problema através da otimização dos coeficientes de forma independente para cada amostra das HRIRs, obtendo-se um conjunto de coeficientes para o modelo através da estrutura de Farrow generalizada.

De um modo geral, pode-se notar que o erro devido à suavização é mais evidente do que o que ocorre com as outras interpolações. No entanto, para as frequências baixas e médias, a escolha de uma ordem mais baixa pareceu ser acertada, já que os erros encontrados se apresentam menos evidentes. Os maiores erros acontecem nas altas frequências e nas regiões mais acidentadas. Nesses casos, como as mudanças ao longo das posições são mais acentuadas, a ordem utilizada pode ter afetado o desempenho.

Parece existir um compromisso entre os erros nas regiões mais suaves e nas mais acidentadas, o que é coerente com os efeitos nas baixas ou nas altas frequências serem mais ou menos evidentes. A opção pela suavidade em baixas e médias frequências, principalmente entre 1 e 5 kHz, pode ser justificada por ser uma região para a qual a sensibilidade do sistema auditivo é máxima (mínimo das curvas de igual audibilidade).

Por esse motivo, vê-se como um primeiro passo para continuação desse trabalho a busca do melhor compromisso entre as duas regiões. Além da determinação

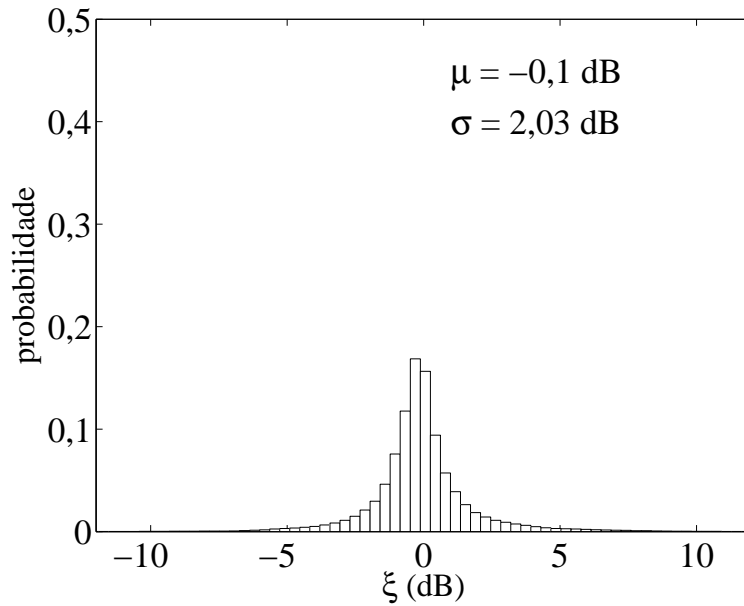


Figura 7.11: Histograma para o erro ξ (interpolação com a estrutura Farrow generalizada).

das melhores ordens para os polinômios, a mudança na forma de se realizar a otimização (escolha de uma outra função objetivo, por exemplo) deve ser explorada. Além disso, através da interpolação triangular, pode-se tentar aumentar a quantidade de funções nas regiões onde não se tem HRIRs medidas e incluí-las no cálculo do erro a ser minimizado.

Uma outra idéia interessante é utilizar a estrutura Farrow como substituta para as IPTFs em uma região próxima a uma HRTF inicial. Novamente, o modelo seria formado por um filtro, cuja resposta é a HRIR da posição mais próxima daquela em que se quer realizar a interpolação, associado a IPTFs, agora modeladas pela estrutura Farrow.

Capítulo 8

Testes Subjetivos

Ao longo da primeira parte dessa tese, desenvolveram-se algumas técnicas de interpolação de HRIRs cujos resultados foram validados objetivamente através das SFRSs. Porém, para uma efetiva validação dessas técnicas se faz necessário algum tipo de avaliação subjetiva.

No presente capítulo, as técnicas de interpolação tratadas nessa tese são comparadas através de três testes subjetivos. Primeiramente, realiza-se a descrição dos testes aplicados, indicando-se o seu objetivo. É realizada, então, a análise dos resultados desses testes, e por fim chega-se a algumas conclusões.

8.1 Descrição dos Testes

De maneira geral, os testes têm como princípio comparar direta ou indiretamente os resultados dos métodos de interpolação. Eles consistem da apresentação do som gerado a pessoas que julgam o que se percebe ao se ouvir o som pré-gravado. A característica a ser julgada deve ser bem esclarecida, e a forma de resposta deve ser a mais simples possível para que a resposta seja quase imediata.

Para que a influência de qualquer diferença seja facilmente percebida, o tipo de sinal a ser apresentado também é importante. O que se faz normalmente é utilizar algum tipo de ruído que excite todos os modos do sistema auditivo. Um tipo de ruído bastante utilizado é o chamado ruído rosa. Esse tipo de ruído tem espectro de potência com decaimento de 3 dB por oitava (10 dB por década) com a frequência. Como a percepção de energia ao longo da frequência é aproximadamente logarítmica,

esse decaimento com a frequência resulta em uma percepção mais uniforme da energia. Com isso, a informação frequencial é percebida sem influência muito maior ou menor de determinada faixa. Em todos os testes realizados utilizou-se ruído rosa obtido de [60].

Trinta e três pessoas com idade entre 20 e 40 anos e com nenhum problema auditivo diagnosticado foram submetidas aos mesmos testes. Nenhuma delas tinha conhecimento específico de som tridimensional, sendo a maioria leiga nesse assunto. Os testes foram realizados em grupos de 3 a 6 pessoas e o controle de apresentação de cada seqüência foi feito pelo autor da presente tese, sendo possível a reapresentação de qualquer seqüência de acordo com a necessidade de algum usuário. A intensidade dos sinais foi regulada previamente, mas aos avaliadores era permitido a alteração do nível de volume. Utilizaram-se fones de ouvido fechado (*closed*), modelo HD265 da Sennheiser, e um amplificador de 8 canais para fones de ouvido da Behringer, modelo Powerplay Pro-8 HA8000. Não foi realizada medição do ruído de fundo no interior da sala, mas com os fones de ouvido do tipo fechado utilizados, o efeito do ruído ambiente pôde ser desconsiderado. Foi ainda sugerido que as pessoas fechassem os olhos a cada seqüência.

8.1.1 Verificação de Mudança de Posição e/ou Timbre

Para avaliar se os métodos de interpolação são equivalentes, o primeiro teste aplicado foi o de simples comparação entre os sinais gerados em uma mesma posição. Nesse teste, cada comparação foi feita entre dois trechos de sinal de 1 segundo de duração, exibidos em seqüência, com um intervalo entre eles também de 1 segundo. As posições foram escolhidas de forma aleatória e independente, segundo uma distribuição uniforme nos intervalos $-180^\circ < \theta < 180^\circ$ e $-40^\circ < \phi < 90^\circ$.

Foram geradas 50 seqüências em posições distintas, das quais 30 contêm uma comparação entre a interpolação triangular realizada diretamente com as HRIRs e uma das outras desenvolvidas na presente tese: a com IPTF de ordem reduzida, a incremental com a KLT e a que utiliza a estrutura Farrow. As outras 20 seqüências são formadas por sinais idênticos gerados com a mesma forma de interpolação, sendo 10 com a interpolação triangular e as outras 10 divididas de maneira aleatória entre os outros métodos interpolação. As seqüências foram apresentadas em uma ordem

aleatória.

O julgamento foi realizado pedindo-se que os avaliadores dessem uma nota de 1 a 4 que indicasse quão perceptível era a diferença entre os sinais da mesma seqüência quanto a mudança na posição e no timbre do ruído. Da maior para a menor, os significados das notas eram “Diferença imperceptível”, “Quase imperceptível”, “Bem evidente” e “Muito acentuada”, respectivamente.

Na Figura 8.1, podem-se ver as notas médias atribuídas a cada um dos métodos e os limites de \pm um desvio-padrão (linhas horizontais acima e abaixo da média). Da esquerda para a direita, vêem-se as médias para os métodos triangular sobre as HRIRs (considerado o padrão), com as IPTFs de ordem reduzida, com a KLT incremental e com a estrutura Farrow. Pode-se notar que, apesar de haver um decrescimento da média, ela ainda está dentro da faixa do desvio da nota para o método triangular. O método de Análise de Variância (ANOVA—*Analysis of Variance*) [61, 62] indicou, com significância maior que 99,99%, que as diferenças encontradas entre as médias eram representativas, não provindo de erro estatístico. Como essas diferenças são pequenas, isso indica equivalência entre os métodos na comparação direta. Pode-se, ainda, confirmar que os resultados são equivalentes pelo fato de a mesma diferença percebida entre o método triangular (padrão) e os outros, ter sido “percebida” entre o método triangular e ele mesmo. Com relação à dispersão das notas em torno da média, nota-se que houve um aumento aproximadamente igual para todos os métodos, comparados ao triangular. Isso indica uma certa diferença entre cada método testado e o triangular, mas insuficiente para alterar tanto a média para os 33 avaliadores.

8.1.2 Verificação da Percepção do Movimento

O segundo teste aplicado procurou avaliar como é percebido o sentido do movimento. Para tanto, geraram-se 32 sinais, 8 para cada tipo de interpolação. Cada um desses 8 partia da posição (0,0) (frente do ouvinte) e seguia por um arco na superfície da esfera em direção a um dos oito pontos cardeais mostrados na Figura 8.2. Cada sinal tinha duração de 5 segundos, sendo que no primeiro e no último segundo a fonte virtual permanecia parada nas posições inicial e final, respectivamente.

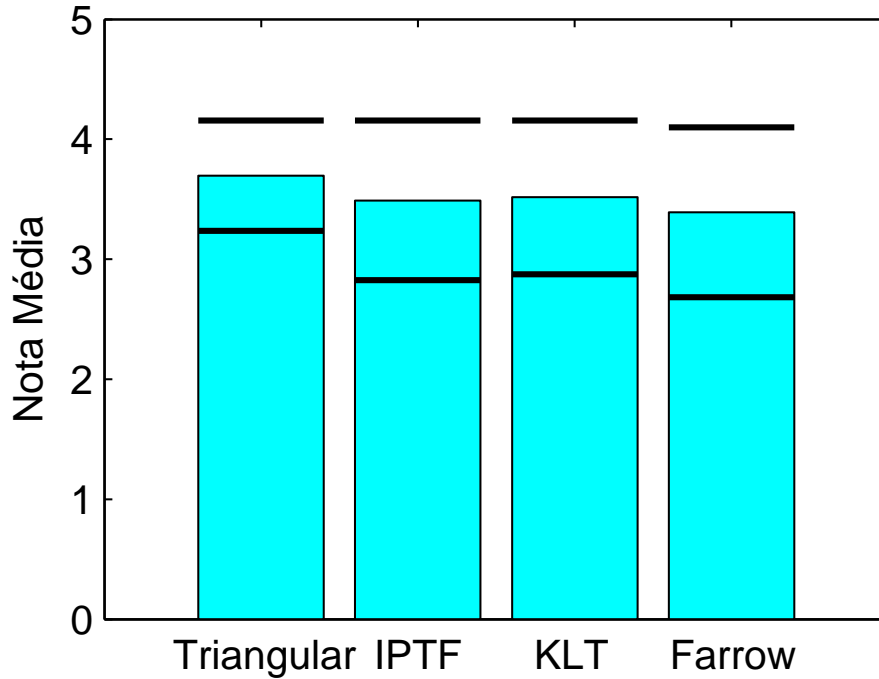


Figura 8.1: Comparação das médias obtidas para a nota de diferença entre o método triangular e todos os outros. Da esquerda para a direita, método: triangular, com IPTF reduzida, KLT incremental e com a estrutura Farrow.

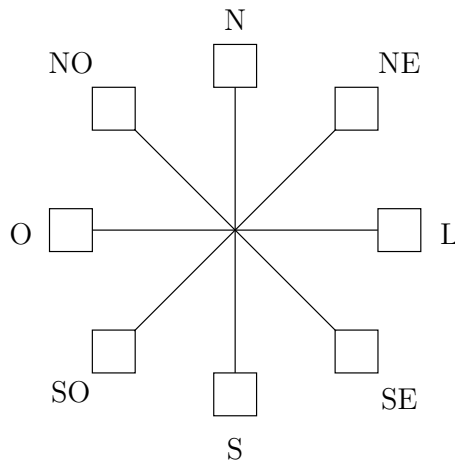


Figura 8.2: Sentidos testados na avaliação de percepção de movimento.

Na Figura 8.3, podem-se ver os resultados das taxas de acerto para cada um dos métodos de interpolação. Nota-se que os métodos de interpolação pela IPTF, com a KLT incremental e com a estrutura Farrow conseguem ser pouco melhores que o triangular. Com esse resultado, pode-se dizer que esses métodos equivalem ao triangular, já que a diferença apresentada não é suficiente para distinguir os três métodos. Também chama a atenção o fato de a estrutura Farrow ter apresentado

resultado equivalente ao dos outros métodos, mesmo tendo ela apresentado SFRSs bastante suavizadas para as frequências mais altas. Isso mostra que a escolha das ordens M e N foi acertada.

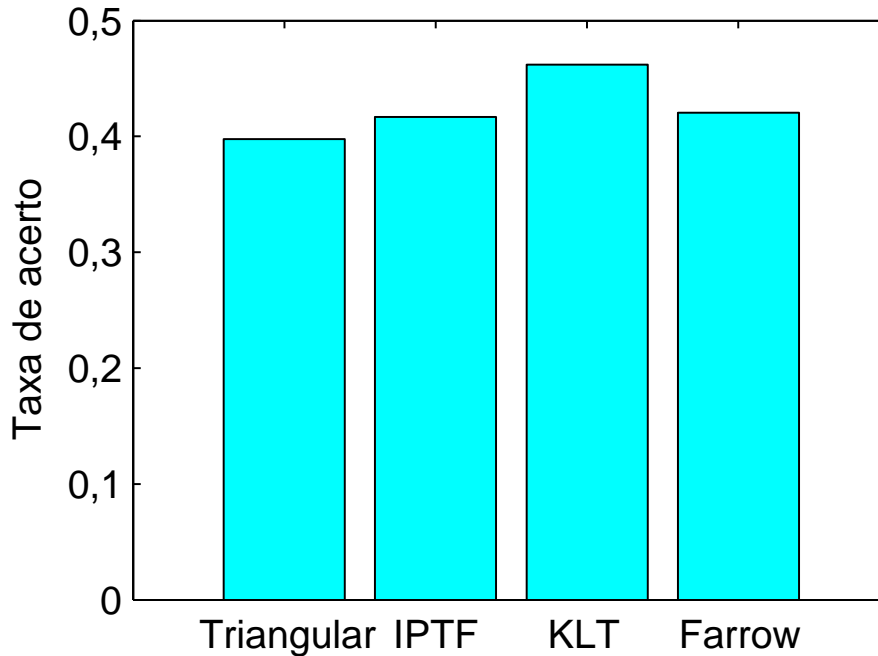


Figura 8.3: Comparação das taxas de acerto do sentido do movimento obtidas para cada método de interpolação. Da esquerda para a Direita: triangular, IPTF reduzida, KLT incremental e com a estrutura Farrow.

Esse teste mostra que o efeito da suavização obtido com o método Farrow generalizado não gerou uma deficiência na percepção global do sentido do movimento e, também, que o método incremental utilizando a KLT pode ser bastante eficaz ao substituir o triangular, já que ele apresentou uma taxa de acerto mais elevada.

Nas Tabelas 8.1–8.4, pode-se observar o percentual das respostas dadas pelos avaliadores para cada sentido gerado. Como mostrado na Figura 8.3, nota-se que há uma pequena melhora nas taxas de acerto (diagonal nas tabelas) para os métodos IPTF, KLT incremental e Farrow, em relação ao método triangular. Isso fica mais evidente para o método KLT incremental. Pode-se perceber, também, que as maiores confusões são entre os sentidos N e S, entre os sentidos NO, SO e O e entre os sentidos NE, SE e L. De fato, a confusão entre cima e baixo é a mais evidente. A distinção lateral é percebida em quase todos os casos, tendo melhor resultado quando se utiliza o método com a estrutura Farrow. No entanto, com a estrutura Farrow, a percepção

dos sentidos N e S fica mais difícil.

Tabela 8.1: Tabela de confusão do teste de percepção do movimento para o método triangular. Valores percentuais.

		Sentido Gerado							
		N	NE	L	SE	S	SO	O	NO
Sentido Percebido	N	51,5	0	0	0	57,6	0	0	6,0
	NE	0	48,5	39,4	42,4	0	0	0	0
	L	0	33,3	42,4	42,4	0	0	0	0
	SE	0	18,2	18,2	15,2	0	0	0	0
	S	45,5	0	0	0	33,3	0	0	0
	SO	3,0	0	0	0	3,0	27,3	30,3	15,2
	O	0	0	0	0	0	36,4	36,4	15,2
	NO	0	0	0	0	6,1	36,3	33,3	63,6

Tabela 8.2: Tabela de confusão do teste de percepção do movimento para o método com IPTFs. Valores Percentuais.

		Sentido Gerado							
		N	NE	L	SE	S	SO	O	NO
Sentido Percebido	N	42,4	0	0	0	36,4	3,0	0	3,0
	NE	6,1	57,6	21,2	33,3	6,0	0	0	0
	L	0	30,3	60,6	39,4	6,1	0	0	0
	SE	12,1	9,1	18,2	27,3	6,0	0	0	0
	S	36,4	3,0	0	0	45,5	0	0	0
	SO	0	0	0	0	0	24,3	30,3	24,3
	O	3,0	0	0	0	0	39,4	36,4	33,3
	NO	0	0	0	0	0	33,3	33,3	39,4

8.1.3 Verificação da Percepção da Posição Estática

O terceiro e último método de avaliação aplicado foi para avaliar a eficácia de cada método com relação à percepção da posição estática. Para esse método,

Tabela 8.3: Tabela de confusão do teste de percepção do movimento para o método KLT incremental. Valores percentuais.

		Sentido Gerado							
		N	NE	L	SE	S	SO	O	NO
Sentido Percebido	N	54,5	0	0	0	51,5	3,0	0	3,0
	NE	6,1	48,5	27,3	27,3	0	0	0	0
	L	3,0	33,3	57,6	27,3	6,0	0	0	0
	SE	6,1	18,2	12,1	45,4	6,1	0	0	0
	S	30,3	0	3,0	0	36,4	0	0	0
	SO	0	0	0	0	0	30,3	15,1	24,3
	O	0	0	0	0	0	39,4	45,5	21,2
	NO	0	0	0	0	0	27,3	39,4	51,5

Tabela 8.4: Tabela de confusão do teste de percepção do movimento para o método com a estrutura Farrow generalizada. Valores Percentuais.

		Sentido Gerado							
		N	NE	L	SE	S	SO	O	NO
Sentido Percebido	N	69,8	0	0	0	24,2	0	0	0
	NE	3,0	45,5	24,2	18,2	24,2	0	0	0
	L	3,0	24,2	51,6	57,6	0	0	0	0
	SE	0	30,3	24,2	24,2	18,2	0	0	0
	S	12,1	0	0	0	27,3	0	0	0
	SO	6,1	0	0	0	0	18,2	30,3	36,4
	O	3,0	0	0	0	0	57,6	57,6	21,2
	NO	3,0	0	0	0	6,1	24,2	12,1	42,4

como no teste anterior, geraram-se 32 sinais, 8 para cada método de interpolação. Para cada método, posicionou-se a fonte virtual em cada um dos cubos mostrados na Figura 8.4. Foi pedido às pessoas que respondessem em qual dos cubos a fonte sonora estava posicionada, considerando que a posição do ouvinte na figura era representada pelo cruzamento dos eixos, que ele estaria olhando na direção da seta e que o plano sombreado passava na altura das orelhas.

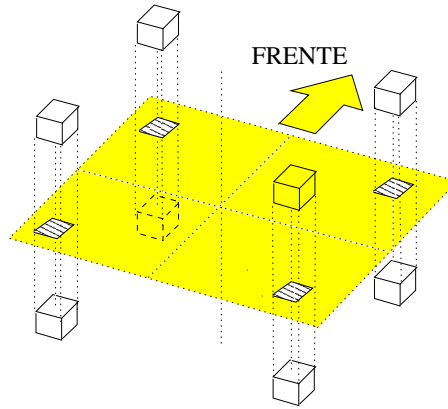


Figura 8.4: Posições testadas na avaliação de percepção da posição estática.

Na Figura 8.5, vêem-se as taxas de acerto para cada um dos métodos de interpolação. Nota-se novamente que os métodos com a IPTF e com a KLT conseguem o mesmo nível de acertos que o triangular, sendo que o que utiliza a KLT ainda é um pouco melhor. O método com a estrutura Farrow generalizada fica com a taxa de acerto um pouco inferior, mas muito próximo à dos outros métodos. Portanto, pode-se deduzir que a escolha das ordens do polinômio da estrutura Farrow generalizada não causou um erro significativo na percepção da posição, apesar dos erros mostrados pelas SFRSs.

Para uma análise mais detalhada, nas Tabelas 8.5–8.8 pode-se ver o percentual das respostas dadas pelos avaliadores para cada uma das posições geradas. Pode-se notar que para nenhum método houve confusão lateral (nenhum sinal pareceu estar vindo do lado oposto àquele em que foi gerado). O que realmente acontece é a confusão frente/trás e cima/baixo. Esse tipo de confusão é considerada normal, já que a fonte foi posicionada em pontos do mesmo cone de confusão. Apesar disso, o método incremental com a KLT obteve um número maior de acertos para a maioria das posições.

Geralmente, para se avaliar se houve erro na identificação da posição, desconsidera-se o efeito da confusão frente/trás [15]. Nesse caso, somando-se os valores percentuais de mesma elevação e azimutes de mesmo sinal (mesmo lado), o método com as IPTFs obtém um número maior de acertos para a maioria das posições, ficando com uma média de acerto igual a 77,7%. O segundo melhor é o método incremental com a KLT (67,5%), praticamente junto com a com o método triangular (66,7%). O método com a estrutura Farrow generalizada ficou com uma taxa de

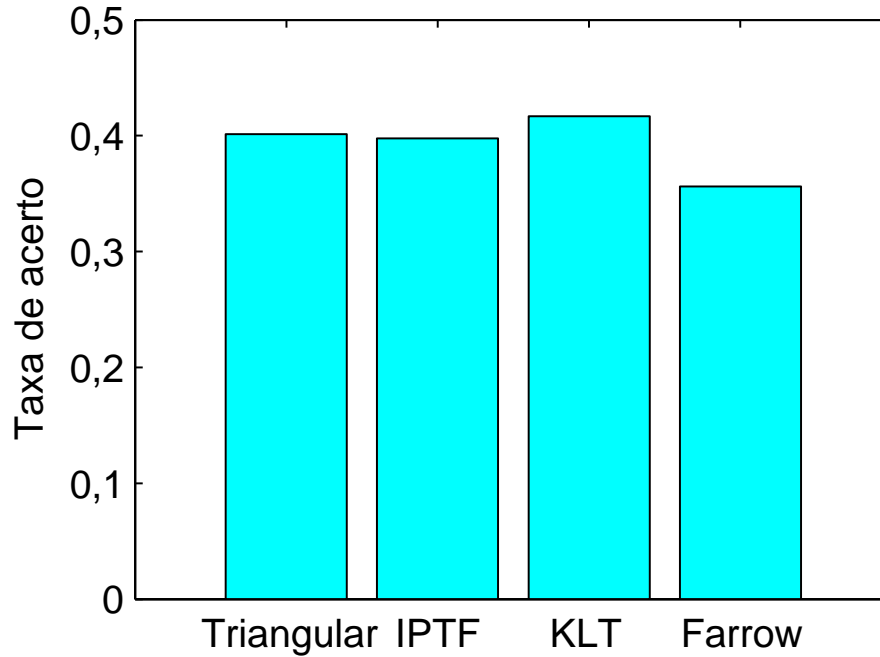


Figura 8.5: Comparação das taxas de acerto da posição estática obtidas para cada método de interpolação. Da esquerda para a Direita: triangular, IPTF reduzida, KLT incremental e com a estrutura Farrow.

acertos de 64,8%, que não é tão diferente das anteriores.

Dessa forma, conclui-se que os métodos podem ser considerados equivalentes, com uma certa vantagem para o incremental com a KLT e, ao se desconsiderar a confusão frente/trás, para o que utiliza as IPTFs. No entanto, para algumas posições, o método com a estrutura Farrow generalizada mostrou-se menos eficaz, apesar de ter obtido resultados bastante próximos aos obtidos com a interpolação triangular. Esse fato indica que diminuindo-se um pouco o erro gerado pela estrutura Farrow, podem-se obter taxas de acerto mais próximas ou maiores que aquelas conseguidas com o método triangular.

8.2 Conclusões

Nesse capítulo, foi apresentada uma descrição dos testes subjetivos, além de uma discussão sobre os resultados obtidos.

Através desses testes, verificou-se que os métodos de interpolação propostos podem ser considerados equivalentes ao método triangular. Os resultados obtidos

Tabela 8.5: Tabela de confusão do teste de percepção da posição para o método triangular. Valores Percentuais.

		Posição Gerada (θ, ϕ)							
		(45,35)	(45,-35)	(135,-35)	(135,35)	(-135,35)	(-45,35)	(-45,-35)	(-135,-35)
Posição Percebida (θ, ϕ)	(45,35)	54,6	27,2	6,1	30,3	0	0	0	0
	(45,-35)	18,2	15,2	21,2	18,2	0	0	0	0
	(135,-35)	3,0	36,4	60,6	12,1	0	0	0	0
	(135,35)	24,2	21,2	12,1	39,4	0	0	0	0
	(-135,35)	0	0	0	0	51,5	36,4	27,2	33,3
	(-45,35)	0	0	0	0	36,3	42,4	48,5	6,1
	(-45,-35)	0	0	0	0	6,1	9,1	9,1	12,1
	(-135,-35)	0	0	0	0	6,1	12,1	15,2	48,5

Tabela 8.6: Tabela de confusão do teste de percepção da posição para o método com IPTFs. Valores Percentuais.

		Posição Gerada (θ, ϕ)							
		(45,35)	(45,-35)	(135,-35)	(135,35)	(-135,35)	(-45,35)	(-45,-35)	(-135,-35)
Posição Percebida (θ, ϕ)	(45,35)	57,6	12,2	6,1	57,6	0	0	0	0
	(45,-35)	0	24,2	21,2	9,1	0	0	0	0
	(135,-35)	0	24,2	60,6	9,1	0	0	0	0
	(135,35)	42,4	39,4	12,1	24,2	0	0	0	0
	(-135,35)	0	0	0	0	48,5	57,6	27,2	27,3
	(-45,35)	0	0	0	0	45,5	27,3	12,1	3,0
	(-45,-35)	0	0	0	0	6,0	9,1	15,2	9,1
	(-135,-35)	0	0	0	0	0	6,0	45,5	60,6

com os métodos incremental com a KLT e que utiliza as IPTFs podem ser considerados um pouco melhores que os outros. Presume-se que seja possível obter melhores resultados com a estrutura Farrow generalizada se for reduzida a suavização gerada por esse método, modificando de forma ainda a ser investigada o algoritmo de obtenção dos coeficientes.

Portanto, chega-se a conclusão que os métodos com IPTFs e KLT incremental

Tabela 8.7: Tabela de confusão do teste de percepção da posição para o método incremental com a KLT. Valores Percentuais.

		Posição Gerada (θ, ϕ)							
		(45,35)	(45,-35)	(135,-35)	(135,35)	(-135,35)	(-45,35)	(-45,-35)	(-135,-35)
Posição Percebida (θ, ϕ)	(45,35)	45,5	21,2	15,2	39,4	0	0	0	0
	(45,-35)	24,1	27,3	15,2	3,0	0	0	0	0
	(135,-35)	15,2	24,2	39,4	9,1	0	0	0	0
	(135,35)	15,2	27,3	30,2	48,5	0	0	0	0
	(-135,35)	0	0	0	0	51,5	30,3	30,3	18,2
	(-45,35)	0	0	0	0	30,3	42,5	12,1	9,1
	(-45,-35)	0	0	0	0	6,1	24,2	21,2	15,1
	(-135,-35)	0	0	0	0	12,1	3,0	36,4	57,6

Tabela 8.8: Tabela de confusão do teste de percepção da posição para o método com a estrutura Farrow generalizada. Valores Percentuais.

		Posição Gerada (θ, ϕ)							
		(45,35)	(45,-35)	(135,-35)	(135,35)	(-135,35)	(-45,35)	(-45,-35)	(-135,-35)
Posição Percebida (θ, ϕ)	(45,35)	63,6	27,3	6,1	51,5	0	0	0	0
	(45,-35)	3,0	9,1	15,1	9,1	0	0	0	0
	(135,-35)	6,1	18,2	48,5	0	0	0	0	0
	(135,35)	27,3	45,4	30,3	39,4	0	0	0	0
	(-135,35)	0	0	0	0	36,4	48,5	57,6	36,4
	(-45,35)	0	0	0	0	33,3	33,3	9,1	3,0
	(-45,-35)	0	0	0	0	18,2	9,1	9,1	15,1
	(-135,-35)	0	0	0	0	12,1	9,1	24,2	45,5

são fortes candidatas a substituir o método triangular. O método incremental com a KLT é especialmente cotado quando se trata do caso com múltiplas fontes, onde sua complexidade o torna bem mais vantajoso.

É importante notar que as taxas de acerto aparentemente baixas (em torno de 40%) devem-se ao fato de não ter sido realizado nenhum treinamento dos ouvintes antes dos testes (os avaliadores foram apresentados aos tipos de som no momento

da avaliação). Além disso, o teste exigia muito da capacidade de abstração de cada um, já que o ambiente virtual não é completo, só tratando da localização da fonte. O modelamento dos outros efeitos como reverberação *early* (as primeiras reflexões) e a compensação do movimento da cabeça [63] podem melhorar muito esses resultados [64].

Parte II

Integração com Caixas Acústicas

Capítulo 9

Interpolação de HRIRs Baseada no VBAP

Como já fora mencionado, um dos métodos multicanal de geração de som 3D [15, 22] é o de Deslocamento por Variação de Amplitude de Base Vetorial (VBAP—*Vector Base Amplitude Panning*) [11]. Esse método posiciona as fontes virtuais utilizando-se de um subconjunto das caixas acústicas dispostas ao redor do ouvinte: pares de caixas, para posicionamento no arco limitado no plano das caixas; ou trios de caixas, para posicionamento no setor esférico limitado pelas caixas.

É possível afirmar que o som que parte de cada caixa acústica sofre a ação de um par de transferências modificadoras até chegar às orelhas e, portanto, a mesma ponderação utilizada pelo método VBAP poderia ser utilizada pelo método de interpolação de HRIRs [14] (binaural). No entanto, deve-se notar que a simples substituição das caixas acústicas pelas HRIRs implica estar-se modelando um ambiente anecóico.

Tendo em vista que a formulação da ponderação das intensidades utilizadas em cada caixa acústica nesses sistemas tem-se mostrado válida, o objetivo desse capítulo é obter uma formulação generalizada para ponderação de HRIRs, dispostas em uma região triangular qualquer, partindo daquela utilizada pelo método VBAP. Além disso, deseja-se apresentar explicitamente o valor dos ponderadores em função das coordenadas angulares, já que tal informação não é mostrada claramente na literatura. Ao final, será mostrado que essa formulação recai na forma triangular utilizada nos capítulos anteriores, mostrando a coerência da interpretação geométrica

utilizada no Capítulo 2.

Assim, na Seção 9.1 discute-se o princípio do posicionamento de uma fonte virtual por variação de amplitude. A seguir, na Seção 9.2, mostra-se que a interpolação de HRTFs pode ser descrita da mesma forma. Na Seção 9.3, revisa-se o método multicanal VBAP. Então, derivam-se as expressões que realizam a interpolação de HRTFs com base no VBAP na Seção 9.4. Por fim, apresentam-se as conclusões.

9.1 Posicionamento por Variação de Amplitude, com Três Canais

Todos os métodos de geração de som tridimensional têm como objetivo básico simular as funções de transferência $H_{PE}(\omega)$ e $H_{PD}(\omega)$ que descrevem os caminhos desde uma posição P até as orelhas esquerda e direita, respectivamente. Para isso, utiliza-se a configuração física do ambiente ou cria-se esse ambiente virtualmente. Assim, tenta-se reproduzir o som de uma fonte virtual localizada em P , atribuindo ao som $S_P(\omega)$, que seria emitido por ela, as mudanças sofridas no caminho até as orelhas, de forma que, no domínio da frequência, se tenham os sinais presentes nas orelhas esquerda e direita dados por

$$\begin{cases} S_E(\omega) = S_P(\omega)H_{PE}(\omega) \\ S_D(\omega) = S_P(\omega)H_{PD}(\omega), \end{cases} \quad (9.1)$$

respectivamente.

No caso de três canais apresentados através de caixas acústicas direcionadas a um ouvinte situado à distância r , como mostra a Figura 9.1, o sinal recebido por cada um dos ouvidos é uma combinação dos sinais emitidos por cada uma das caixas, podendo ser descrito como

$$\begin{cases} \hat{S}_E(\omega) = \sum_{i=1}^3 S_i(\omega)H_{iE}(\omega) \\ \hat{S}_D(\omega) = \sum_{i=1}^3 S_i(\omega)H_{iD}(\omega), \end{cases} \quad (9.2)$$

onde $S_i(\omega)$ é o espectro do sinal emitido pela caixa L_i , e $H_{iE}(\omega)$ e $H_{iD}(\omega)$ são as funções de transferência da caixa L_i aos ouvidos esquerdo e direito, respectivamente.

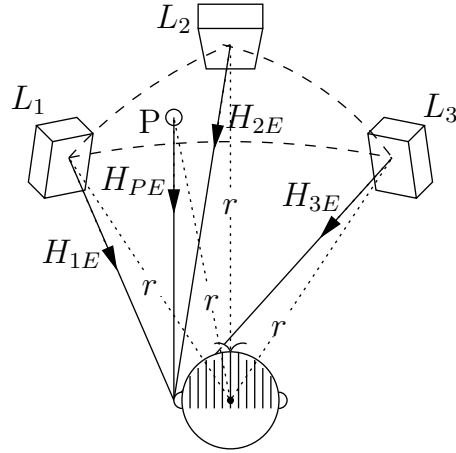


Figura 9.1: Posicionamento baseado na variação de amplitude com 3 caixas acústicas—ouvido esquerdo.

Potencialmente, esta configuração posiciona uma fonte virtual em qualquer ponto P sobre o triângulo esférico determinado pelas três caixas. Um sistema de som 3D eficaz deve garantir que

$$\begin{cases} \hat{S}_E(\omega) \approx S_E(\omega) \\ \hat{S}_D(\omega) \approx S_D(\omega). \end{cases} \quad (9.3)$$

Uma solução intuitiva consiste em aplicar o sinal que se deseja ouvir vindo do ponto P a cada uma das caixas L_i ponderado por uma constante g_i [11], ou seja, $S_i(\omega) = g_i S_P(\omega)$. Nesse caso, a aproximação (9.3) se resume a fazer

$$\begin{cases} H_{PE}(\omega) \approx \sum_{i=1}^3 g_i H_{iE}(\omega) \\ H_{PD}(\omega) \approx \sum_{i=1}^3 g_i H_{iD}(\omega). \end{cases} \quad (9.4)$$

9.2 Interpolação Baseada em Três HRTFs

Como já mencionado nos capítulos anteriores, normalmente, o sistema de geração binaural de som tridimensional utiliza um conjunto de HRIRs medidas para um número finito de posições sobre uma esfera de raio r com o ouvinte no centro [23]. Para se posicionar uma fonte virtual em qualquer ponto P sobre a esfera, atribui-se à HRIR para o ponto P uma combinação linear das HRIRs medidas para os pontos mais próximos de P , isso para cada ouvido.

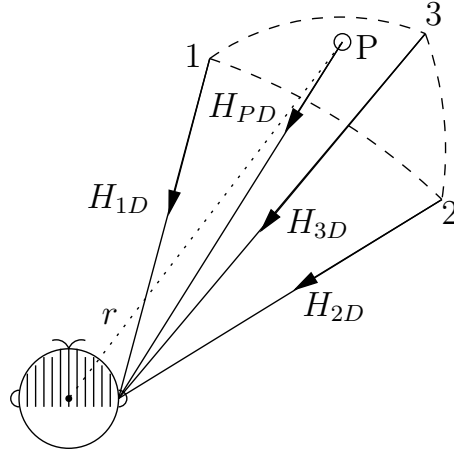


Figura 9.2: Interpolação de HRIR a partir de 3 funções conhecidas—ouvido direito.

A Figura 9.2 mostra um esquema de interpolação baseado em três HRIRs conhecidas.

No domínio da frequência, a aproximação realizada pode ser regida pelas mesmas Equações (9.4), onde as funções $H_{PE}(\omega)$ e $H_{PD}(\omega)$ são estimadas diretamente pela combinação das funções $H_{iE}(\omega)$ e $H_{iD}(\omega)$, com $i = 1, 2, 3$, respectivamente, para os ouvidos esquerdo e direito. Por esse motivo, tendo as funções de transferência a mesma significação física nos casos multicanal e binaural, espera-se que uma solução para os pesos g_i adequada ao caso multicanal também seja apropriada para o caso binaural.

9.3 Método VBAP

A técnica VBAP [11] se baseia na descrição vetorial das posições dos alto-falantes e da fonte virtual em relação ao ouvinte. No caso de três canais, ilustrado na Figura 9.3, os coeficientes g_i que satisfazem as aproximações das Equações (9.4) são calculados através de

$$\mathbf{v}_P = \sum_{i=1}^3 g_i \mathbf{v}_i. \quad (9.5)$$

Essa equação descreve o vetor associado à posição P como uma combinação linear dos vetores \mathbf{v}_i associados às posições das caixas L_i , com $i = 1, 2, 3$, em coordenadas cartesianas. Admite-se, assim, que a mesma combinação é válida para as funções de transferência correspondentes. Explicitamente,

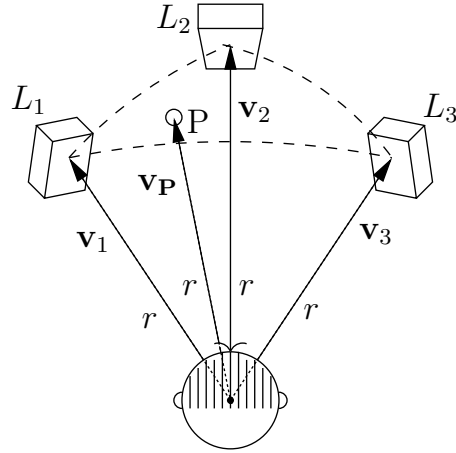


Figura 9.3: VBAP com três canais.

$$\begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}, \quad (9.6)$$

onde (x_i, y_i, z_i) são as coordenadas relacionadas com a posição da caixa L_i .

9.4 Interpolação de HRTFs Baseada no VBAP

Como a interpolação de HRIRs pode utilizar os mesmos ponderadores calculados para um método multicanal com três caixas acústicas, propõe-se usar como base a formulação do VBAP.

Considerando a geometria definida na Figura 9.2, é vantajoso escrever os vetores (x, y, z) em função das coordenadas esféricas:

$$\begin{cases} x = r \cos \phi \cos \theta \\ y = r \cos \phi \sin \theta \\ z = r \sin \phi, \end{cases} \quad (9.7)$$

onde ϕ é o ângulo de elevação relativo ao plano horizontal que passa pelas orelhas, θ é o ângulo de azimute relativo ao plano vertical que corta perpendicularmente a linha que une as orelhas e r é o raio da esfera.

Escolhendo como referência o ponto 1, conforme a Figura 9.4, as coordenadas angulares dos outros pontos ficam iguais a

$$(\theta_2, \phi_2) = (\theta_1 + \Delta\theta_2, \phi_1 + \Delta\phi_2) \quad (9.8)$$

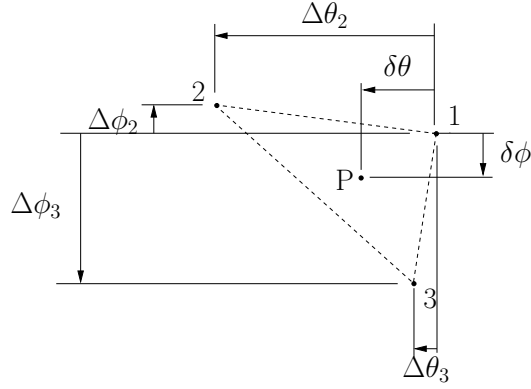


Figura 9.4: Geometria para interpolação triangular de HRIRs.

$$(\theta_3, \phi_3) = (\theta_1 + \Delta\theta_3, \phi_1 + \Delta\phi_3) \quad (9.9)$$

$$(\theta_P, \phi_P) = (\theta_1 + \delta\theta, \phi_1 + \delta\phi). \quad (9.10)$$

Substituindo as Equações (9.8)–(9.10) na Equação (9.7), obtêm-se as coordenadas dos pontos

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} r \cos \phi \cos \theta_1 \\ r \cos \phi \sin \theta_1 \\ r \sin \phi_1 \end{bmatrix} \quad (9.11)$$

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} r \cos(\phi_1 + \Delta\phi_2) \cos(\theta_1 + \Delta\theta_2) \\ r \cos(\phi_1 + \Delta\phi_2) \sin(\theta_1 + \Delta\theta_2) \\ r \sin(\phi_1 + \Delta\phi_2) \end{bmatrix} \quad (9.12)$$

$$\begin{bmatrix} x_3 \\ y_3 \\ z_3 \end{bmatrix} = \begin{bmatrix} r \cos(\phi_1 + \Delta\phi_3) \cos(\theta_1 + \Delta\theta_3) \\ r \cos(\phi_1 + \Delta\phi_3) \sin(\theta_1 + \Delta\theta_3) \\ r \sin(\phi_1 + \Delta\phi_3) \end{bmatrix} \quad (9.13)$$

$$\begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix} = \begin{bmatrix} r \cos(\phi_1 + \delta\phi) \cos(\theta_1 + \delta\theta) \\ r \cos(\phi_1 + \delta\phi) \sin(\theta_1 + \delta\theta) \\ r \sin(\phi_1 + \delta\phi) \end{bmatrix}. \quad (9.14)$$

Agora, considerando que as HRTFs medidas formam um conjunto denso (i.e., foram medidas para pontos bastante próximos¹), pode-se considerar que $\Delta\theta_2, \Delta\theta_3,$

¹Em [23], por exemplo, algo em torno de 700 posições são distribuídas de forma que a distância entre duas posições adjacentes é aproximadamente a mesma em qualquer região da esfera de referência.

$\Delta\phi_2$, $\Delta\phi_3$, $\delta\theta$ e $\delta\phi$ são pequenos. Sendo $\text{sen}(\alpha) \approx \alpha$ e $\text{cos}(\alpha) \approx 1$ para $\alpha \approx 0$, podem-se obter as aproximações para as coordenadas

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} \approx \begin{bmatrix} r(\text{cos } \phi_1 - \Delta\phi_2 \text{sen } \phi_1)(\text{cos } \theta_1 - \Delta\theta_2 \text{sen } \theta_1) \\ r(\text{cos } \phi_1 - \Delta\phi_2 \text{sen } \phi_1)(\text{sen } \theta_1 + \Delta\theta_2 \text{cos } \theta_1) \\ r(\text{sen } \phi_1 + \Delta\phi_2 \text{cos } \phi_1) \end{bmatrix} \quad (9.15)$$

$$\begin{bmatrix} x_3 \\ y_3 \\ z_3 \end{bmatrix} \approx \begin{bmatrix} r(\text{cos } \phi_1 - \Delta\phi_3 \text{sen } \phi_1)(\text{cos } \theta_1 - \Delta\theta_3 \text{sen } \theta_1) \\ r(\text{cos } \phi_1 - \Delta\phi_3 \text{sen } \phi_1)(\text{sen } \theta_1 + \Delta\theta_3 \text{cos } \theta_1) \\ r(\text{sen } \phi_1 + \Delta\phi_3 \text{cos } \phi_1) \end{bmatrix} \quad (9.16)$$

$$\begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix} \approx \begin{bmatrix} r(\text{cos } \phi_1 - \delta\phi \text{sen } \phi_1)(\text{cos } \theta_1 - \delta\theta \text{sen } \theta_1) \\ r(\text{cos } \phi_1 - \delta\phi \text{sen } \phi_1)(\text{sen } \theta_1 + \delta\theta \text{cos } \theta_1) \\ r(\text{sen } \phi_1 + \delta\phi \text{cos } \phi_1) \end{bmatrix}. \quad (9.17)$$

Por fim, substituindo-se as Equações (9.11) e (9.15)–(9.17) na Equação (9.6), podem-se calcular os ponderadores g_i em função das distâncias angulares:

$$\begin{cases} g_3 = \frac{\delta\phi\Delta\theta_2 - \delta\theta\Delta\phi_2}{\Delta\theta_2\Delta\phi_3 - \Delta\theta_3\Delta\phi_2} \\ g_2 = \frac{\delta\theta\Delta\phi_3 - \delta\phi\Delta\theta_3}{\Delta\theta_2\Delta\phi_3 - \Delta\theta_3\Delta\phi_2} \\ g_1 = 1 - g_2 - g_3. \end{cases} \quad (9.18)$$

Nesse caso, o número de operações necessárias para se obter os coeficientes é de 11 somas/subtrações e 8 multiplicações/divisões por nova posição.

Novamente, deve-se notar que essa formulação é válida em ambiente sem reverberação alguma (anecóico), já que somente as transferências entre caixa e ouvidos são levadas em consideração. Em [12], Pulkki mostra que, no caso geral, deve-se manter constante o nível de audibilidade da fonte virtual e que isso depende da reverberação do ambiente. Naquele trabalho, os valores dos pesos calculados devem estar de acordo com a expressão

$$\sqrt[p]{\sum_{i=1}^I g_i^p} = 1, \quad (9.19)$$

onde I é o número de caixas utilizada pelo VBAP e p é o parâmetro que depende do quão reverberante é o ambiente de reprodução. No caso de o ambiente ser anecóico,

$p = 1$ o que revela que a formulação obtida na Equação (9.18) mantém o nível de audibilidade constante, já que a expressão

$$g_1 = 1 - g_2 - g_3 \quad (9.20)$$

é a solução obtida diretamente ao se calcularem os ganhos g_i em função das coordenadas esféricas.

Deve ser notado que no caso geral, as coordenadas angulares dos pontos para os quais as HRIRs foram medidas são quaisquer. Entretanto, na prática, as medidas são realizadas com pelo menos o passo de elevação fixo. Nesse caso particular, pode-se fazer $\Delta\phi_2 = 0$ nas Equações (9.18), o que as torna

$$\begin{cases} g_3 = \frac{\delta\phi}{\Delta\phi_3} \\ g_2 = \frac{\delta\theta}{\Delta\theta_2} - g_3 \frac{\Delta\theta_3}{\Delta\theta_2} \\ g_1 = 1 - g_2 - g_3. \end{cases} \quad (9.21)$$

É importante notar que no Capítulo 2 mostrou-se uma formulação para interpolação de HRIRs que calcula os ponderadores seguindo o mesmo princípio da interpolação bilinear: a relação com as distâncias angulares. Com o resultado obtido pela Equação (9.21), pode-se dizer que as equações obtidas naquele capítulo são equivalentes às do VBAP, evidenciando a estreita relação entre os dois métodos de geração de som tridimensional.

Uma simplificação adicional ainda pode ser feita no caso de as HRIRs serem medidas com ambos os passos, de azimuth e de elevação, constantes, o que permite fazer $\Delta\theta_3 = 0$ nas Equações (9.21), resultando em

$$\begin{cases} g_3 = \frac{\delta\phi}{\Delta\phi_3} \\ g_2 = \frac{\delta\theta}{\Delta\theta_2} \\ g_1 = 1 - g_2 - g_3. \end{cases} \quad (9.22)$$

Assim, quando a geometria da grade de pontos o permite, a complexidade do cálculo dos ponderadores g_i é consideravelmente reduzida.

9.5 Conclusões

Neste capítulo, deduziu-se uma formulação para o cálculo dos ponderadores usados para geração de som tridimensional utilizando-se HRIRs para uma região triangular genérica. Chegou-se às equações dos ponderadores através da analogia com um sistema multicanal que se baseia em três caixas acústicas, chamado de Deslocamento por Variação de Amplitude de Base Vetorial (VBAP).

Os resultados bem sucedidos do VBAP são conhecidos da literatura. Como a interpolação de HRTFs é simplesmente uma combinação linear de funções de transferência medidas para pontos sobre uma esfera, a formulação do VBAP pode ser reutilizada. Já que ambos os métodos atuam sobre o mesmo tipo de função, a expectativa é de que os resultados dos dois métodos sejam equivalentes.

Além da formulação para o caso geral, citaram-se dois casos particulares de menor complexidade, aplicáveis a conjuntos de HRIRs obtidos na prática. Um dos casos particulares, descrito pelas Equações (9.21), recai no mesmo caso do que fora anteriormente aplicado pelos autores em [P2] (formulação rerepresentada no Capítulo 2) a um método de interpolação eficiente de HRIRs, com sucesso. Naquele trabalho, porém, não se fez uma relação direta com a técnica VBAP, o que foi feito em [P6].

Os principais resultados desse capítulo são a confirmação da equivalência entre a formulação utilizada pelo VBAP e pela interpolação de HRIRs descrita no Capítulo 2 e a descrição de forma explícita do cálculo dos pesos utilizados nesses dois tipos de geração de som tridimensional.

Capítulo 10

Som Tridimensional com Caixas Acústicas

No Capítulo 9, formalizaram-se as expressões dos pesos utilizados na interpolação de HRIRs através de uma analogia com o método VBAP [11], que pode ser estendida à maioria dos métodos multicanal. O método VBAP faz uma abordagem independente do número e da posição das caixas de som utilizadas, em que o ganho para cada caixa é calculado a partir da ponderação dos vetores de uma base. Como é evidenciado pelo autor, na configuração de dois canais, esse método recai na chamada Lei das Tangentes, que é normalmente utilizada em sistemas estéreo.

Tanto a Lei das Tangentes quanto a Lei dos Senos são relações entre a intensidade sonora aplicada a cada caixa e as posições angulares das caixas e da fonte virtual. Essas relações, deduzidas em [65], implicam que certas limitações ao bom funcionamento do sistema sejam impostas. Um exemplo de limitação é o fato de essas relações só serem realmente válidas para as baixas frequências ($f_{\text{máx}} < 600$ Hz) [65, 66, 67].

Apesar de existirem tais limitações, a analogia feita no Capítulo 9 continua sendo válida. Essas limitações não implicam erro na interpolação de HRIRs. Pelo contrário, no caso das HRIRs as equações de interpolação são mais coerentes, já que a configuração utilizada na obtenção dos ponderadores não necessita de simplificações, o que não ocorre para os métodos multicanal.

Na Seção 10.1, a dedução para as Leis dos Senos e das Tangentes é revista com o intuito de explicitar as implicações de seu uso nos sistemas de geração

tridimensional com caixas acústicas. Na seção seguinte, o erro causado pelas simplificações utilizadas nas deduções é evidenciado através de alguns exemplos de respostas na frequência entre fontes virtuais e orelhas. Na Seção 10.3, mostra-se como a interpolação de HRIRs é mais completa e não requer o grau de restrição necessário para os métodos com caixas acústicas. Por fim, algumas conclusões são tiradas.

10.1 Leis dos Senos e das Tangentes

10.1.1 Lei dos Senos

O sistema estereofônico pode ser representado por um par de caixas acústicas dispostas de forma simétrica com relação a um ouvinte que se encontra em uma posição equidistante a elas, olhando para o meio do arco que une as caixas, como mostra a Figura 10.1.

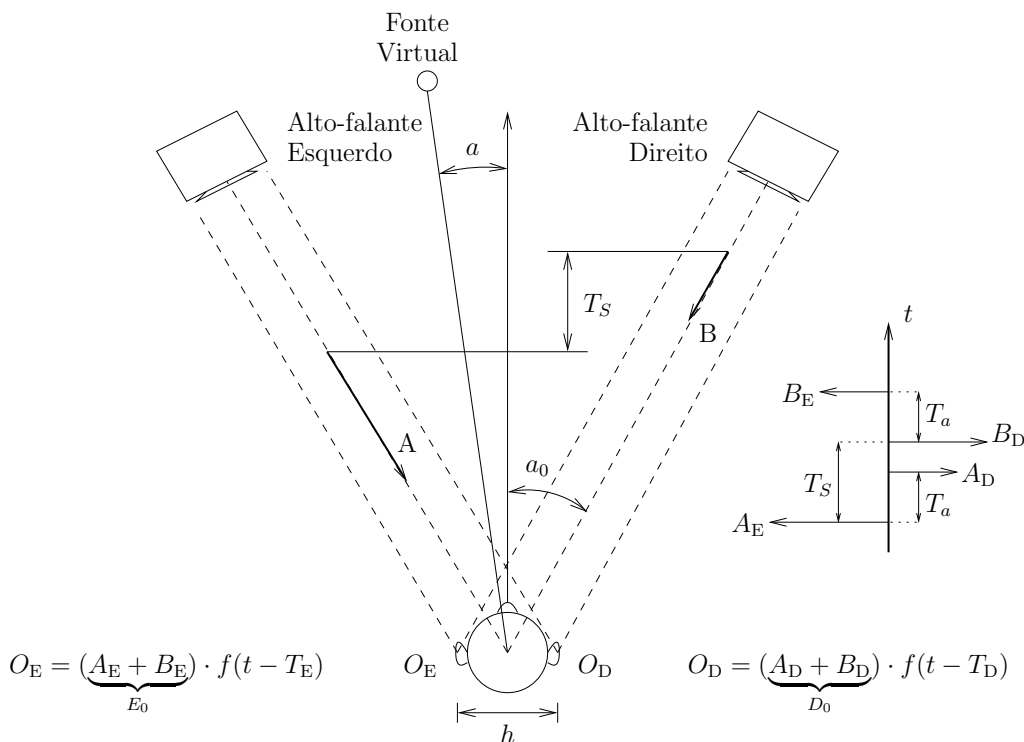


Figura 10.1: Esquema de sistema estereofônico convencional.

Nessa figura, h é a distância interaural, a_0 é o ângulo que a linha que une cada caixa à orelha faz com o eixo central de referência (frente do ouvinte), A e B são as intensidades aplicadas às caixas esquerda e direita, respectivamente, e as

intensidades $A_{(\cdot)}$ e $B_{(\cdot)}$ provêm das caixas esquerda e direita, respectivamente, e chegam à orelha (\cdot) . Além disso, definem-se os tempos T_a como sendo a diferença de tempo interaural (ITD) e T_S como sendo o atraso aplicado entre as caixas acústicas. É importante notar que T_a , pela simetria, é fixo, enquanto que T_S pode ser um dos parâmetros responsáveis pela noção de espacialidade requerida e, portanto, variável.

Note que se os sinais aplicados respectivamente às caixas A e B forem iguais, somente diferenciados pelas amplitudes e pelo atraso T_S , como em

$$C_A = kBf(t); \quad C_B = Bf(t - T_S), \quad (10.1)$$

com $k = A/B$, os sinais que chegam às orelhas esquerda e direita são, respectivamente

$$O_E = A_E f(t) + B_E f[t - (T_a + T_S)] \quad (10.2)$$

$$O_D = A_D f(t - T_a) + B_D f(t - T_S). \quad (10.3)$$

Utilizando-se apenas os dois primeiros termos de sua expansão de Taylor (veja apêndice D), esses sinais podem ser escritos na forma

$$O_E \approx (A_E + B_E) f \left[t - \frac{B_E T_a + B_E T_S}{A_E + B_E} \right] = E_0 f(t - T_E) \quad (10.4)$$

$$O_D \approx (A_D + B_D) f \left[t - \frac{A_D T_a + B_D T_S}{A_D + B_D} \right] = D_0 f(t - T_D). \quad (10.5)$$

Assim, podem-se obter as expressões para a ILD e a ITD

$$\text{ILD} = \frac{E_0}{D_0} = \frac{A_E + B_E}{A_D + B_D} \quad (10.6)$$

e

$$\text{ITD} = T_D - T_E = \frac{A_D T_a + B_D T_S}{A_D + B_D} - \frac{B_E T_a + B_E T_S}{A_E + B_E}. \quad (10.7)$$

Colocando essas equações em função de k (razão interaural) e $m = \frac{A_E}{A_D} = \frac{B_D}{B_E}$ (ILD), tem-se que

$$\text{ITD} = T_D - T_E = \frac{m T_a (k^2 - 1) + k T_S (m^2 - 1)}{m(k^2 + 1) + k(m^2 + 1)} \quad (10.8)$$

e

$$\text{ILD} = \frac{km + 1}{k + m}. \quad (10.9)$$

É importante salientar que, nesse caso, assume-se que a distância interaural é pequena comparada à distância caixa-ouvinete, e que o sistema é simétrico com

relação à posição da cabeça do ouvinte. Além disso, todos os intervalos de tempo são por hipótese bem menores do que o período da maior frequência do sinal reproduzido nas caixas e do que o menor período de transitório. Note que as duas primeiras hipóteses estão relacionadas com parâmetros que se podem controlar na implementação prática do sistema. Por outro lado, as hipóteses relacionadas com os tempos podem não ser garantidas. Por exemplo, consideremos que a ITD relacionada com a posição das caixas (dadas as restrições anteriores relativas às distâncias) é

$$T_a = \frac{h}{c} \operatorname{sen} a_0, \quad (10.10)$$

onde $h \approx 21$ cm estima o diâmetro da cabeça do ouvinte e $c = 340$ m/s é a velocidade do som no ar. Se assumirmos, agora, que a_0 é 30° , teremos um período T_a maior que $300 \mu\text{s}$. Considerando que esse período tem que se enquadrar nas condições acima, isso limitaria a faixa de frequência de funcionamento para algo bem menor que $\frac{1}{T_a} = 3,3$ kHz.

Além dessa limitação, para se chegar à Lei dos Senos, faz-se a restrição de não se aplicarem às caixas sinais atrasados entre si ($T_S = 0$), além de se considerar que $m \approx 1$, o que implica exigir que os sinais aplicados tenham frequências menores que 600 Hz. Assim, chega-se a partir da Equação (10.9) em $ILD \approx 1$ e

$$ITD = \frac{h}{c} \operatorname{sen} a = \frac{k-1}{k+1} T_a = \frac{A-B}{A+B} T_a. \quad (10.11)$$

Assim, sabendo-se que

$$T_a = \frac{h}{c} \operatorname{sen} a_0, \quad (10.12)$$

chega-se à relação chamada de Lei dos Senos:

$$\frac{A-B}{A+B} = \frac{\operatorname{sen} a}{\operatorname{sen} a_0}. \quad (10.13)$$

10.1.2 Lei das Tangentes

Até aqui, admitiu-se que o ouvinte estava com sua cabeça parada, mantendo a simetria do sistema de reprodução. No caso em que o ouvinte tem a liberdade de girar a cabeça, a simetria é perdida. No entanto, restringindo-se essa rotação de forma que o ouvinte esteja sempre olhando para a direção da fonte virtual, podem-se obter as expressões da ITD e ILD.

A Figura 10.2 mostra a mudança nas distâncias relacionadas com a rotação da cabeça: Figura 10.2a com relação à caixa direita e Figura 10.2b com relação à caixa esquerda. Nela pode-se perceber que os atrasos do sinal que chega às orelhas nas posições novas são diferentes daqueles que ocorriam no caso anterior. Essa diferença nos atrasos acarreta os sinais O'_E e O'_D iguais a

$$O'_E = A_E f(t - T_1) + B_E f(t - T_2) \quad (10.14)$$

$$O'_D = A_D f(t - T_4) + B_D f(t + T_3), \quad (10.15)$$

onde

$$T_1 = \frac{\overline{NO'_E}}{c} \quad (10.16)$$

$$T_2 = \frac{\overline{QO_E} + \overline{O_ER}}{c} \quad (10.17)$$

$$T_3 = \frac{\overline{O_DT}}{c} \quad (10.18)$$

$$T_4 = \frac{\overline{LO_D} + \overline{O'_DM}}{c}, \quad (10.19)$$

sendo c a velocidade do som. Colocando esses tempos em função dos ângulos a e a_0 (veja a Figura 10.3), tem-se que

$$T_1 = \frac{h}{c} \operatorname{sen} \left(\frac{a}{2} \right) \cos \left(\frac{a}{2} - a_0 \right) \quad (10.20)$$

$$T_2 = \frac{h}{c} \operatorname{sen} a_0 + \frac{h}{c} \operatorname{sen} \left(\frac{a}{2} \right) \cos \left(\frac{a}{2} + a_0 \right) \quad (10.21)$$

$$T_3 = \frac{h}{c} \operatorname{sen} \left(\frac{a}{2} \right) \cos \left(\frac{a}{2} + a_0 \right) \quad (10.22)$$

$$T_4 = \frac{h}{c} \operatorname{sen} a_0 - \frac{h}{c} \operatorname{sen} \left(\frac{a}{2} \right) \cos \left(\frac{a}{2} - a_0 \right), \quad (10.23)$$

já que os segmentos envolvidos são

$$\overline{NO'_E} = \overline{O_E O'_E} \operatorname{sen} \left(\frac{a}{2} - a_0 + 90^\circ \right) = h \operatorname{sen} \left(\frac{a}{2} \right) \cos \left(\frac{a}{2} - a_0 \right) \quad (10.24)$$

$$\overline{QO_E} = h \operatorname{sen} a_0 \quad (10.25)$$

$$\overline{O_ER} = \overline{O_E O'_E} \cos \left(\frac{a}{2} + a_0 + 90^\circ \right) = h \operatorname{sen} \left(\frac{a}{2} \right) \cos \left(\frac{a}{2} + a_0 \right) \quad (10.26)$$

$$\overline{O_DT} = \overline{O_ER} = h \operatorname{sen} \left(\frac{a}{2} \right) \cos \left(\frac{a}{2} + a_0 \right) \quad (10.27)$$

$$\overline{LO_D} = h \operatorname{sen} a_0 \quad (10.28)$$

$$\overline{O'_DM} = \overline{NO'_E} = h \operatorname{sen} \left(\frac{a}{2} \right) \cos \left(\frac{a}{2} - a_0 \right). \quad (10.29)$$

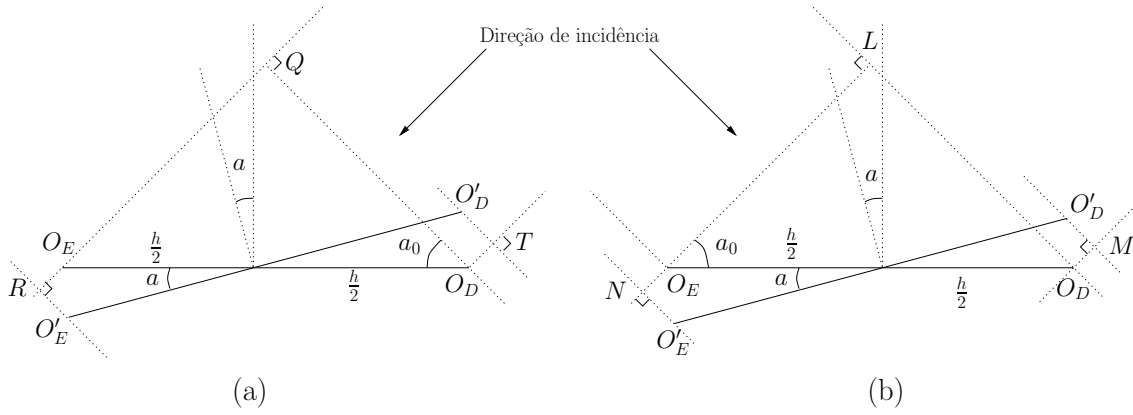


Figura 10.2: Mudança nas distâncias devido à rotação da cabeça: (a) com relação à frente de onda vinda da caixa direita e (b) com relação à frente de onda vinda da caixa esquerda.

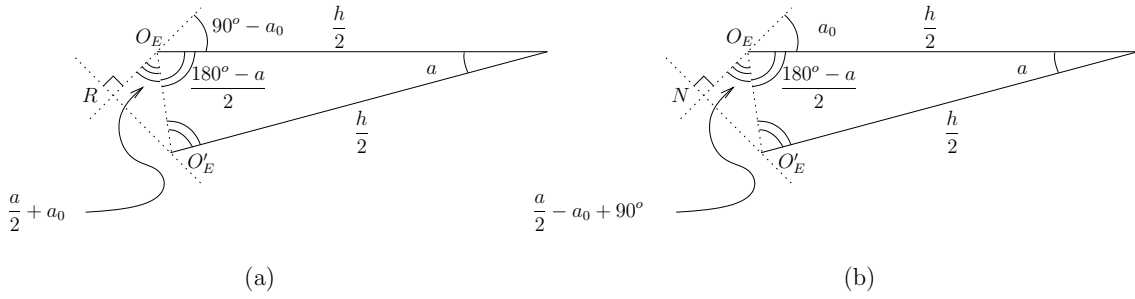


Figura 10.3: Detalhe dos ângulos envolvidos na mudança das distâncias devido à rotação da cabeça: (a) com relação à frente de onda vinda da caixa direita e (b) com relação à frente de onda vinda da caixa esquerda.

Considerando que os sinais são de baixa frequência, $m \approx 1$ e, com isso, $A_E = A_D = A$ e $B_E = B_D = B$. Novamente, considerando os primeiros termos da expansão de Taylor, os atrasos totais nas orelhas esquerda e direita podem ser descritos, respectivamente, como

$$T_E = \frac{AT_1 + BT_2}{A + B} \quad (10.30)$$

$$T_D = \frac{AT_4 - BT_3}{A + B}. \quad (10.31)$$

Substituindo-se os tempos de T_1 a T_4 nessas equações e, notando-se que

$$\text{ITD} = T_D - T_E = 0, \quad (10.32)$$

tem-se

$$\frac{\text{tg } a}{\text{tg } a_0} = \frac{A - B}{A + B}, \quad (10.33)$$

que é a chamada Lei das Tangentes.

10.2 Limitações e Suas Implicações

Ambas as relações obtidas na seção anterior consideram que: os tempos T_a e T_S são pequenos se comparados com o período de tempo gasto na propagação entre as caixas e as orelhas; e que o som proveniente de uma das caixas chega com a mesma intensidade nas duas orelhas ($m \approx 1$). Portanto, deduz-se que tais relações só serão válidas para sinais cujas componentes estejam na região de baixas frequências.

Outro fato importante é que no caso de se utilizar a Lei das Tangentes para gerar múltiplas fontes simultâneas, somente a posição da fonte para a qual o ouvinte estaria olhando (premissa considerada na dedução dessa relação) é que será convenientemente descrita.

Para exemplificar o efeito dessas aproximações sobre as respostas na frequência resultantes para o caminho entre a fonte virtual e as orelhas, fez-se uma comparação entre a resposta na frequência resultante do método VBAP e uma das respostas existentes no conjunto medido de HRTFs. Deve-se notar que para se obter tal resultado pelo método VBAP, utilizaram-se três HRTFs que formavam um setor triangular sobre a esfera (cada uma representando uma das caixas) e que contivesse o ponto de comparação. Para esse exemplo, elegeu-se uma região da esfera onde as Leis dos Senos e das Tangentes eram válidas (isto é, onde é garantido que há simetria e que o ouvinte estaria olhando para a fonte—região em torno da coordenada $(0^\circ, 0^\circ)$) e calculou-se a função de transferência relativa à posição de teste utilizando-se a formulação do Capítulo 9.

Na Figura 10.4, tem-se o exemplo que mostra o efeito das simplificações implícitas nas Leis dos Senos e das Tangentes. Pode-se notar que Figuras 10.4b–10.6b, pode-se perceber que na Figura 10.4b, a partir de frequências próximas a 1 kHz, o erro nas respostas relativas à orelha esquerda e à orelha direita começa a aumentar muito, mostrando que nessa faixa ambas as hipóteses (de que os tempos são desprezíveis e $m \approx 1$) afetam significativamente a resposta em médias e altas frequências.

Outro fato interessante que se pode notar na Figura 10.4a é o erro que aparece

pelo fato de a composição ser realizada sobre as funções reais de fase não-mínima (que são representadas perfeitamente pelas HRIRs): comparando-se os atrasos nas respostas impulsivas, vê-se alguma diferença entre o da resposta medida e o da função resultante do VBAP. O efeito desse erro pode causar o aparecimento de ganhos e atenuações indesejados, como já fora mencionado no Capítulo 3. Note que no caso de sistemas multicanal por variação de amplitude, esse erro não pode ser reduzido, já que as funções entre caixa e orelha são combinadas diretamente na orelha.

Para complementar a análise das limitações do método VBAP, consideraram-se outras duas regiões sobre a esfera em torno do ouvinte. Nas Figuras 10.5 e 10.6, pode-se verificar que a mesma limitação em banda está presente. Note que esses dois últimos exemplos são realizados como se o ouvinte estivesse olhando para a posição $(0^\circ, 0^\circ)$, o que viola o pressuposto pela Lei das Tangentes.

10.3 Modelo Mais Flexível

Pode-se perceber que o modelo adotado para se obter as Leis dos Senos e das Tangentes só se aplica à configuração real se o sinal aplicado às caixas acústicas consistir de apenas uma frequência, já que os ganhos A_E , A_D , B_E e B_D na realidade dependem da frequência. Essa dependência é um dos fatos responsáveis pelas limitações dos métodos multicanal.

O modelo completo para o caso de duas caixas acústicas simétricas em relação ao ouvinte (como o caso da dedução da Lei dos Senos) pode ser descrito por

$$O_E(\omega) = A_E(\omega)F(\omega) + B_E(\omega)F(\omega) \quad (10.34)$$

$$O_D(\omega) = A_D(\omega)F(\omega) + B_D(\omega)F(\omega). \quad (10.35)$$

Considerando que

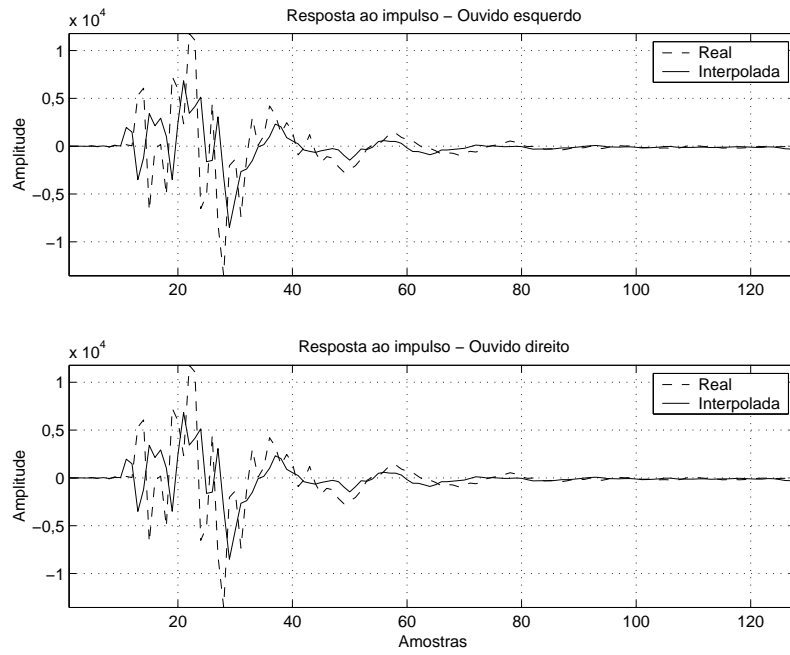
$$A_E(\omega) = A \cdot H_{1E}(\omega), \quad (10.36)$$

$$A_D(\omega) = A \cdot H_{1D}(\omega), \quad (10.37)$$

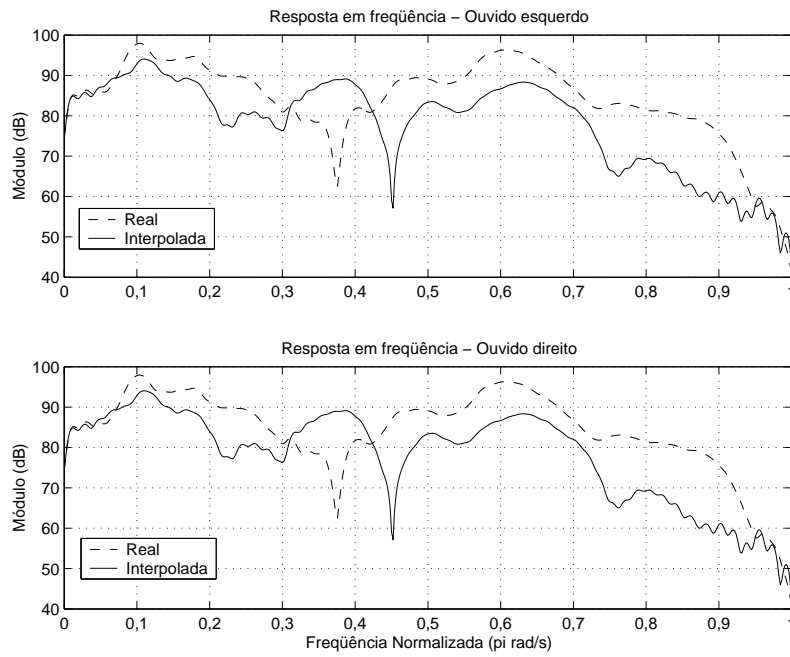
$$B_E(\omega) = B \cdot H_{2E}(\omega) \quad (10.38)$$

e

$$B_D(\omega) = B \cdot H_{2D}(\omega), \quad (10.39)$$

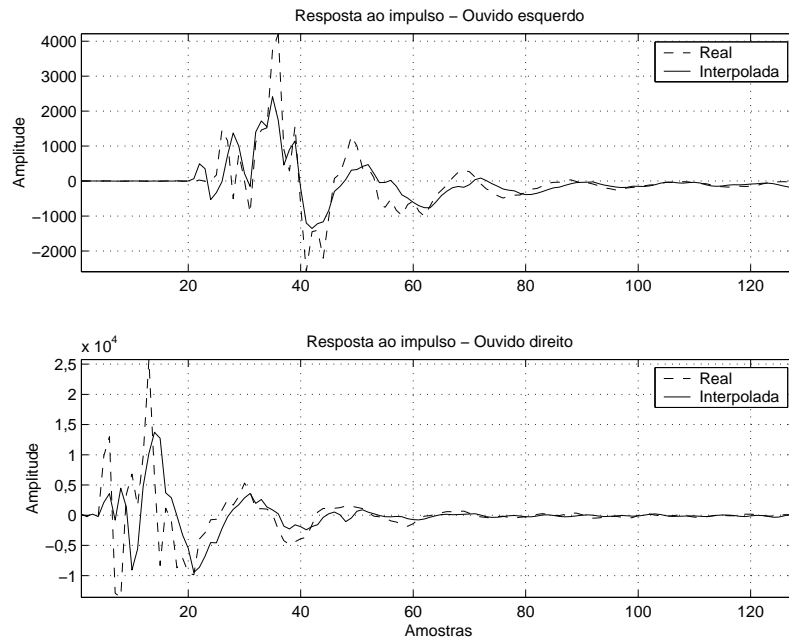


(a)

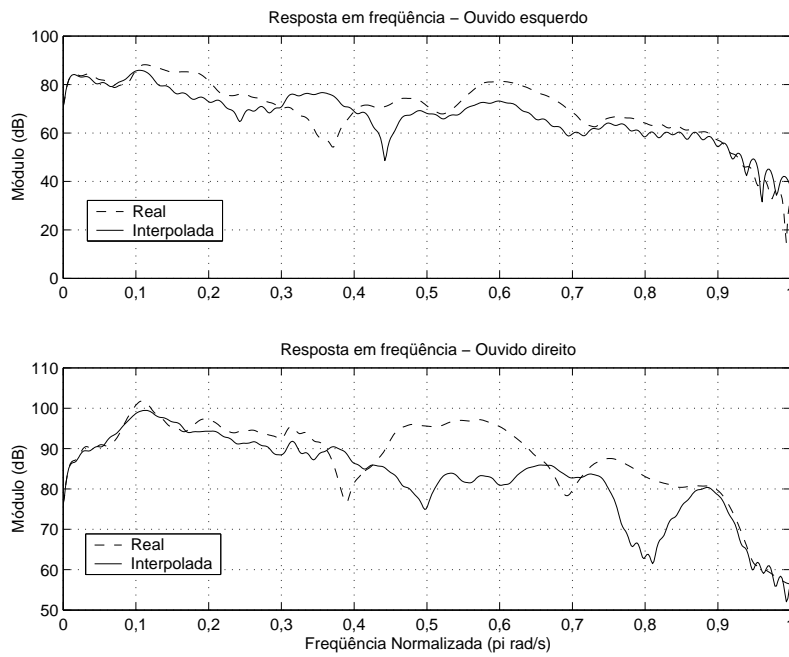


(b)

Figura 10.4: Respostas para a posição $(0^\circ, 0^\circ)$ utilizando o triângulo $\{(0^\circ, 20^\circ), (-15^\circ, -20^\circ), (15^\circ, -20^\circ)\}$: (a) Respostas ao impulso e (b) respostas em frequência.

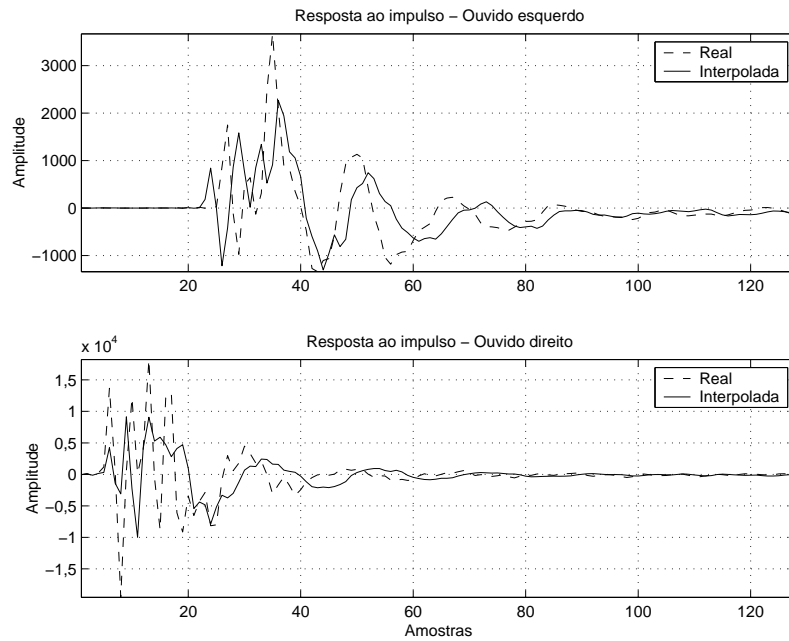


(a)

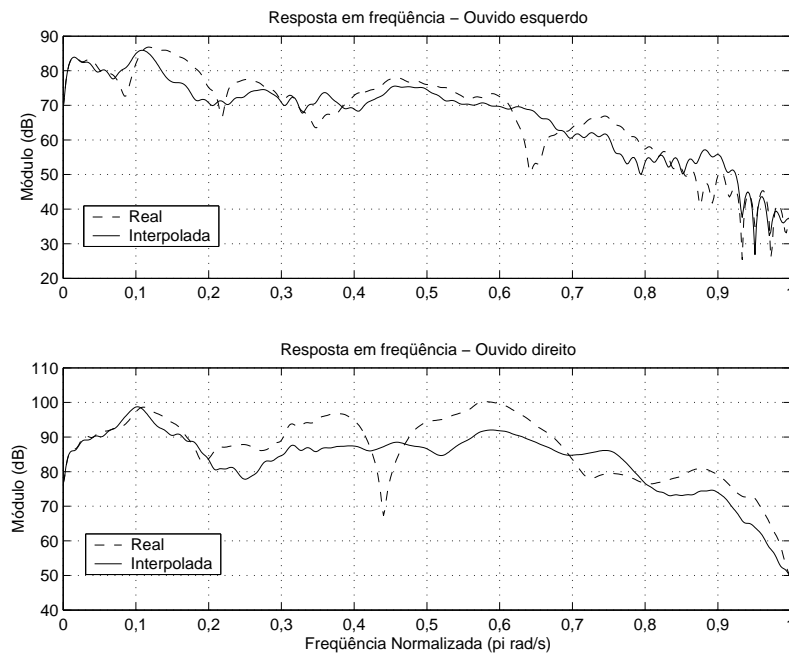


(b)

Figura 10.5: Respostas para a posição $(60^\circ, 0^\circ)$ utilizando o triângulo $\{(60^\circ, 20^\circ), (45^\circ, -20^\circ), (75^\circ, -20^\circ)\}$: (a) Respostas ao impulso e (b) respostas em frequência.



(a)



(b)

Figura 10.6: Respostas para a posição $(120^\circ, 0^\circ)$ utilizando o triângulo $\{(120^\circ, 20^\circ), (105^\circ, -20^\circ), (135^\circ, -20^\circ)\}$: (a) Respostas ao impulso e (b) respostas em frequência.

onde $H_{1E}(\omega)$, $H_{1D}(\omega)$, $H_{2E}(\omega)$ e $H_{2D}(\omega)$ são as transferências entre as caixas 1 e 2 (relativas às intensidades A e B , respectivamente) e as orelhas esquerda (E) e direita (D).

Dessa forma, podem-se escrever as transferências das caixas para as orelhas esquerda e direita, respectivamente, como

$$\frac{O_E(\omega)}{F(\omega)} = A \cdot H_{1E}(\omega) + B \cdot H_{2E}(\omega) \quad (10.40)$$

$$\frac{O_D(\omega)}{F(\omega)} = A \cdot H_{1D}(\omega) + B \cdot H_{2D}(\omega) \quad (10.41)$$

Comparando-se essas expressões com as utilizadas para se obter a formulação das Leis dos Senos e das Tangentes, pode-se deduzir que o fato de se utilizar um ponderador escalar (independente da frequência) leva a uma redução da faixa de frequência para a qual o modelo é válido, como visto anteriormente. Além disso, a combinação linear mostrada nas Equações (10.40) e (10.41) causa efeitos de cancelamento para determinadas frequências pela forma como é realizada a combinação da fase dos sinais que chegam no canal auditivo. Na prática, esses efeitos acabam sendo minimizados em ambientes com pouca ou nenhuma reverberação [68].

Nesse ponto, o modelo binaural mostra-se mais flexível, já que permite o controle dessas transferências, podendo-se fazer as combinações lineares das funções de fase mínima correspondentes e, ainda, realizar a interpolação dos atrasos relativos à ITD separadamente.

10.4 Conclusões

Nesse capítulo, rerepresentou-se a dedução das Leis dos Senos e das Tangentes de forma a explicitar as simplificações adotadas na sua obtenção. Essas simplificações acarretam uma limitação na faixa de frequências para a qual essa relação de ganhos seja válida. Essas limitações são ilustradas através de alguns exemplos práticos que simulam o método VBAP representando a transferência de cada uma das caixas por uma HRIR medida.

Tendo em vista que a formulação do VBAP, que recai na Lei das Tangentes para o caso de dois canais, pode ser utilizada pela interpolação de HRIRs, tem-se a primeira impressão de que os erros decorrentes do não-atendimento das considerações

iniciais ocorreriam na obtenção da HRIR final. De fato, isso não ocorre para o modelo binaural, mais flexível que o multicanal pelo fato de permitir o controle completo das HRIRs.

Uma possível seqüência a este trabalho é mostrar uma formulação unificada que evidencie as similaridades, diferenças e limitações de cada método.

Capítulo 11

Conclusões

Nessa tese, realizou-se um estudo de algumas técnicas para geração de som tridimensional. Primeiramente, como continuação do trabalho desenvolvido em [P1], realizou-se uma avaliação mais completa da interpolação de HRIRs através das IPTFs em regiões triangulares, o que levou a aperfeiçoamentos no procedimento de busca das funções a serem utilizadas na interpolação e resultou na forma final de implementação apresentada no presente trabalho e em [P2].

Novos modelos para o conjunto de HRIRs foram propostos: um através da representação desse conjunto pela KLT [P5] e outro através da estrutura Farrow. Para esse último, fez-se uma generalização da estrutura para o caso de duas variáveis e escolheu-se uma ordem para os polinômios usados na representação, levando em conta uma solução de compromisso que prioriza a redução da complexidade computacional.

Além de uma extensa comparação objetiva entre os métodos propostos e o método clássico, chamado de triangular, através de Superfícies Espaciais de Resposta na Frequência (SFRSs), realizaram-se três tipos de testes subjetivos com 33 pessoas. Esses testes visavam à verificação da diferença entre os sinais produzidos por métodos distintos, da diferença quanto à percepção do sentido do movimento e da diferença quanto à percepção da posição. O método que utiliza a KLT e faz uma interpolação incremental dos seus coeficientes mostrou-se o melhor daqueles propostos, principalmente para o caso de múltiplas fontes simultâneas. Esses testes indicaram também que se pode obter melhores resultados com a estrutura Farrow generalizada se a forma de obtenção dos coeficientes dessa estrutura for revista.

Para o caso da fonte simples, o método com as IPTFs (mais eficiente, nesse caso) se mostrou-se equivalente ao com a KLT.

Além disso, foi realizada também uma discussão sobre a relação entre os métodos binaural e os multicanal, onde se verificou que a formulação utilizada nos sistemas binaurais é equivalente à utilizada nos sistemas multicanal por variação de amplitude. Foi visto também que, ao se aplicar essa formulação em sistemas multicanal, recai-se na formulação clássica das Leis dos Senos e das Tangentes, trazendo consigo as limitações em banda para as quais se pode garantir a acurácia do sistema. Por fim, mostrou-se um dos motivos pelos quais as limitações implícitas nos métodos multicanal não ocorrem nos métodos binaural.

Para continuação do presente trabalho pode-se pensar, principalmente:

- na realização de um estudo mais específico da estrutura Farrow generalizada para que os coeficientes obtidos consigam modelar melhor as HRIRs;
- na busca da redução da complexidade da estrutura Farrow através da representação dos coeficientes obtidos por uma transformada;
- na utilização da estrutura Farrow generalizada como substituta da IPTF, fazendo aquela atuar como o interpolador numa região pequena em torno de uma HRIR conhecida;
- na conclusão da unificação da formulação utilizada nos sistemas multicanal e binaural com as Leis dos Senos e das Tangentes, buscando comparar as limitações envolvidas em cada uma delas;
- no modelamento da reverberação;
- no modelamento da mudança de distância à fonte; e
- na compensação do movimento da cabeça.

Referências Bibliográficas

- [1] AUDIO ENGINEERING SOCIETY, *The Proceedings of the AES 22nd. International Conference on Virtual, Synthetic, and Entertainment Audio*, Espoo, Finland, 15–17th June 2002.
- [2] AUDIO ENGINEERING SOCIETY, *The Proceedings of the AES 16th International Conference on Spatial Sound Reproduction*, Rovaniemi, Finland, 10–12th April 1999.
- [3] IMAX CORPORATION, “The IMAX Experience—How IMAX works?”, <http://www.imax.com>, February 2005.
- [4] NIEMIMUUKKO, U., “Spatial Sound with Personal Computers”, 2003. Internal Communication of Helsinki University of Technology.
- [5] PULKKI, V., HIRVONEN, T., “Localization of Virtual Sources in Multichannel Audio Reproduction”, *IEEE Transactions on Speech and Audio Processing*, v. 13, n. 1, pp. 105–119, January 2005.
- [6] YAMAHA CORPORATION, “Receiver HTR-5130 Owner’s Manual”, <http://www2.yamaha.co.jp/manual/pdf/av/english/re/HTR-5130.pdf>, February 2005.
- [7] YAMAHA CORPORATION, “Receiver RX-V10MKII Owner’s Manual”, <http://www2.yamaha.co.jp/manual/pdf/av/english/re/RX-V10MKII.pdf>, February 2005.
- [8] ITU-R BS.775-1, *Multichannel Stereophonic Sound System with and Without Accompanying Picture*. Geneva, International Telecommunications Union, 1992-94.

- [9] ITU-R BS.1116-1, *Methods for Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*. Geneva, International Telecommunications Union, 1994–97.
- [10] GOLDBERG, A., “Commercial Multichannel Formats and Their Reproduction Using a Single Loudspeaker System”, 2003. Internal Communication of Helsinki University of Technology.
- [11] PULKKI, V., “Virtual Sound Source Positioning Using Vector Base Amplitude Panning”, *Journal of the Audio Eng. Soc.*, v. 45, n. 6, pp. 456–466, June 1997.
- [12] PULKKI, V., *Spatial Sound Generation and Perception by Amplitude Panning Techniques*. Ph.D. thesis, Helsinki University of Technology, Department of Electrical and Communications Engineering, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland, August 2001.
- [13] ALGAZI, V. R., DUDA, R. O., THOMPSON, D. M., “Motion-Tracked Binaural Sound”, *J. Audio Eng. Soc.*, v. 52, n. 11, pp. 1142–1156, November 2004.
- [14] CHENG, C. I., WAKEFIELD, G. H., “Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space”, *J. Audio Eng. Soc.*, v. 49, n. 4, pp. 231–249, April 2001.
- [15] BEGAULT, D. R., *3D Sound for Virtual Reality and Multimedia*. Cambridge, MA, USA, Academic Press, 1994.
- [16] RAUHALA, J., “Overview of Multichannel Reproduction Techniques”, 2003. Internal Communication of Helsinki University of Technology.
- [17] BRANDENBURG, K., BRIX, S., SPORER, T., “Wave Field Synthesis: From Research to Applications”. In: *The Proceedings of the XII European Signal Processing Conference–EUSIPCO’04*, pp. 1369–1376, EURASIP, Vienna, Austria, 2004.
- [18] VRIES, D. D., HULSEBOS, E. M., “Auralization of Room Acoustics by Wave Field Synthesis Based on Array Measurements of Impulse Responses”. In: *The Proceedings of the XII European Signal Processing Conference–EUSIPCO’04*, pp. 1377–1380, EURASIP, Vienna, Austria, 2004.

- [19] BUCHNER, H., SPORS, S., HERBORDT, W., *et al.*, “Wave-Domain Adaptive Filtering for Acoustic Human-Machine Interfaces Based on Wavefield Analysis and Synthesis”. In: *The Proceedings of the XII European Signal Processing Conference–EUSIPCO’04*, pp. 1385–1388, EURASIP, Vienna, Austria, 2004.
- [20] EARGLE, J., *Sound Recording*. Van Nostrand Reinhold Company, 1976.
- [21] HUOPANIEMI, J., *Virtual Acoustics and 3D Sound in Multimedia Signal Processing*. Ph.D. thesis, Helsinki University of Technology, Department of Electrical and Communications Engineering, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland, November 1999.
- [22] JOT, J.-M., LARCHER, V., WARUSFEL, O., “Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony”. In: *Proc. 98th AES Convention*, AES, Paris, France, February 1995. (preprint 3980).
- [23] GARDNER, B., MARTIN, K., *HRTF Measurements of a KEMAR Dummy-Head Microphone*, Technical Report 280, MIT Media Lab., Cambridge, MA, USA, May 1994.
- [24] BURKHARD, M. D. (ed.), *Manikin Measurements*, Industrial Research Products, Inc., February 1978. http://www.knowleselectronics.com/images/products/pdf/KEMAR_Manikin_Measurements.pdf.
- [25] RIFE, D. D., VANDERKOOY, J., “Transfer-Function Measurements using Maximum-Length Sequences”, *J. Audio Eng. Soc.*, v. 37, n. 6, pp. 419–444, June 1989.
- [26] VANDERKOOY, J., “Aspects of MLS Measuring Systems”, *J. Audio Eng. Soc.*, v. 42, n. 4, pp. 219–231, April 1994.
- [27] NIRONEN, H., “Overview of Binaural Reproduction Techniques”, 2003. Internal Communication of Helsinki University of Technology.
- [28] CHENG, C. I., WAKEFIELD, G. H., “Spatial Frequency Response Surfaces (SFRS’S): An Alternative Visualization and Interpolation Technique for Head-Related Transfer Functions (HRTF’s)”. In: *Proc. AES 16th Int. Conf.*, pp. 147–159, Rovaniemi, Finland, April 1999.

- [29] FARROW, C. W., “A Continuously Variable Digital Delay Element”. In: *IEEE Proceedings of the International Symposium in Circuits and Systems*, pp. 2641–2643, 1988.
- [30] CHEN, J., VAN VEEN, B. D., HECOX, K. E., “External Ear Transfer Function Modeling: A Beamforming Approach”, *The J. Acoust. Society of America*, v. 92, n. 4, pp. 1933–1944, October 1992. Part 1.
- [31] LARCHER, V., JOT, J.-M., GUYARD, J., *et al.*, “Study and Comparison of Efficient Methods for 3D Audio Spatialization Based on Linear Decomposition of HRTF Data”. In: *Proc. 108th AES Convention*, AES, Paris, France, February 2000. (preprint 5097).
- [32] SAVIOJA, L., HUOPANIEMI, J., LOKKI, T., *et al.*, “Creating Interactive Virtual Acoustic Environments”, *J. Audio Eng. Soc.*, v. 47, n. 9, pp. 675–705, September 1999.
- [33] SAVIOJA, L., *Modeling Techniques for Virtual Acoustics*. Ph.D. thesis, Helsinki University of Technology, Department of Computer Science and Engineering, Telecommunications Software and Multimedia Laboratory Espoo, Finland, December 1999.
- [34] HUOPANIEMI, J., ZACHAROV, N., KARJALAINEN, M., “Objective and Subjective Evaluation of Head-Related Transfer Function Filter Design”, *Journal of the Audio Eng. Soc.*, v. 47, n. 4, pp. 218–239, April 1999.
- [35] KULKARNI, A., ISABELLE, S. K., COLBURN, H. S., “On the Minimum-Phase Approximation of Head-Related Transfer Functions”. In: *IEEE Workshop on Applications of the Signal Processing to Audio and Acoustics*, IEEE, New Paltz, New York, October 1995.
- [36] ROEDERER, J. G., *The Physics and Psychophysics of Music—An Introduction*. 3 ed. NY, USA, Springer-Verlag, 1994.
- [37] BUTLER, D., *The Musician’s Guide to Perception and Cognition*. NY, USA, Schirmer, 1992.

- [38] HUOPANIEMI, J., SMITH III, J. O., “Spectral and Time-Domain Preprocessing and the Choice of Modeling Error Criteria for Binaural Digital Filters”. In: *The Proceedings of the AES 16th International Conference*, pp. 301–312, AES, Rovaniemi, Finland, April 1999.
- [39] HATZIANTONIOU, P. D., MOURJOPOULOS, J. N., “Generalized Fractional-Octave Smoothing of Audio and Acoustic Response”, *Journal of the Audio Eng. Soc.*, v. 48, n. 4, pp. 259–280, April 2000.
- [40] OPPENHEIM, A. V., SHAFER, R. W., *Discrete-Time Signal Processing*. New Jersey: Prentice Hall, 1989.
- [41] MACKENZIE, J., HUOPANIEMI, J., VÄLIMÄKI, V., *et al.*, “Low-Order Modeling of Head-Related Transfer Functions using Balanced Model Truncation”, *IEEE Signal Processing Letters*, v. 4, n. 2, pp. 39–41, February 1997.
- [42] AVENDANO, C., DUDA, R. O., ALGAZI, V. R., “Modeling the Contralateral HRTF”. In: *Proc. AES 16th Int. Conf.*, pp. 313–318, Rovaniemi, Finland, April 1999.
- [43] LORHO, G., HUOPANIEMI, J., ZACHAROV, N., *et al.*, “Efficient HRTF Synthesis Using an Interaural Transfer Function Model”. In: *Proceedings of X European Signal Processing Conference (EUSIPICO’2000)*, pp. 2213–2216, EUROSIP, Tampere, Finland, September 2000.
- [44] LAAKSO, T. I., VÄLIMÄKI, V., KARJALAINEN, M., *et al.*, “Splitting the Unit Delay”, *IEEE Signal Processing Magazine*, v. 13, n. 1, pp. 30–59, January 1996.
- [45] HARTUNG, K., BRAASCH, J., STERBING, S. J., “Comparison of Different Methods for the Interpolation of Head-Related Transfer Function”. In: *The Proceedings of the AES 16th International Conference*, pp. 319–329, AES, Rovaniemi, Finland, April 1999.
- [46] BELICZYNSKY, B., KALE, I., CAIN, G. D., “Approximation of FIR by IIR Digital Filters: An Algorithm Based on Balanced Model Reduction”, *IEEE Transactions on Signal Processing*, v. 40, n. 3, pp. 532–542, March 1992.

- [47] STRANG, G., *Linear Algebra and Its Applications*. 3 ed. USA, Saunders, 1988.
- [48] GLOVER, K., “All Optimal Hankel-Norm Approximation of Linear Multivariable Systems and Their L^∞ -error Bounds”, *International Journal of Control*, v. 39, n. 6, pp. 1115–1117, 1984.
- [49] VÄLIMÄKI, V., LAAKSO, T. I., “Suppression of Transients in Time-Varying Recursive Filters for Audio Signals”. In: *in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP’98)*, v. 6, pp. 3569–3572, IEEE, Seattle, Washington, May 1998.
- [50] JOT, J.-M., WARDLE, S., LARCHER, V., “Approaches to Binaural Synthesis”. In: *Proceedings of AES 105th Convention*, AES, California, USA, September 1998. (preprint 4861).
- [51] CHEN, J., VEEN, B. D. V., HECOX, K. E., “Synthesis of 3D Virtual Auditory Space Via a Spatial Feature Extraction and Regularization Model”. In: *IEEE Virtual Reality Annual International Symposium*, pp. 188–193, September 1993.
- [52] CHEN, J., VEEN, B. D. V., HECOX, K. E., “A Spatial Feature Extraction and Regularization Model for Virtual Auditory Display”. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, v. 1, pp. 129–132, April 1993.
- [53] JAIN, A. K., *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ, USA, Prentice Hall, 1989.
- [54] NISHINO, T., KAJITA, S., TAKEDA, K., *et al.*, “Interpolating Head-Related Transfer Function In the Median Plane”. In: *Proc. IEEE WASPAA ’1999*, pp. 167–170, New Paltz, NY, USA, October 1999.
- [55] PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T., *et al.*, *Numerical Recipes in C*, 2 ed., Cambridge, 1992.
- [56] DINIZ, P. S. R., *Adaptive Filtering: Algorithms and Practical Implementation*. Kluwer Academic Publishers, 1997.

- [57] LAROCHE, J., “Time and Pitch Scale Modification of Audio Signals”. In: Kahrs, M., Brandenburg, K. (eds.), *Applications of Digital Signal Processing to Audio and Acoustics*, chapter 7, Kluwer Academic Publishers, 1998.
- [58] VESMA, J., *Optimization and Applications of Polynomials-Based Interpolation Filters*. Ph.D. thesis, Tampere University of Technology, Tampere, Finland, 1999.
- [59] VESMA, J., SARAMÄKI, T., “Interpolation Filters with Arbitrary Frequency Response for All-Digital Receivers”. In: *Proc. IEEE Int. Symp. on Circuits and Systems*, pp. 568–571, Atlanta, GA, USA, May 1996.
- [60] MOSHIER, S., Internet, November 2003, <http://www.moshier.net/pink.html>.
- [61] WEISSTEIN, E. W., “ANOVA”, From MathWorld—A Wolfram Web Resource, <http://mathworld.wolfram.com/ANOVA.html>.
- [62] MATLAB, “Statistics Toolbox”, Math Works Inc.
- [63] BEGAULT, D. R., “Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems”, *J. Audio Eng. Soc.*, v. 40, n. 11, pp. 895–904, November 1992.
- [64] TAN, C.-J., GAN, W.-S., “Direct Concha Excitation for the Introduction of Individualized Hearing Cues”, *J. Audio Eng. Society*, v. 48, n. 7/8, pp. 642–653, July/August 2000.
- [65] BERNFELD, B., “Attempts for Better Understanding of the Directional Stereophonic Listening Mechanism”. In: *44th Audio Engineering Society Convention*, AES, Rotterdam, February 1973.
- [66] BERNFELD, B., “Simple Equations for Multichannel Stereophonic Sound Localization”, *Journal of the Audio Eng. Soc.*, v. 23, n. 7, pp. 553–557, September 1975.
- [67] VIANNA, B. C., *Técnicas de Geração de Som Tridimensional*. Projeto final, Universidade Federal do Rio de Janeiro, fevereiro 2005.

- [68] PULKKI, V., KARJALAINEN, M., “Localization of Amplitude-Panned Virtual Sources I: Stereophonic Panning”, *Journal of the Audio Engineering Society*, v. 49, n. 9, pp. 739–752, September 2001.

Apêndice A

Método de Medição de Funções de Transferência

Sendo $x(k)$ uma seqüência pseudo-aleatória de período N e com amostras independentes, pode-se obter a resposta impulsiva $h(k)$ de um determinado sistema, primeiramente, aplicando-se à entrada desse sistema o sinal $x'(k)$ definido por:

$$x'(k) = \begin{cases} x(k+N), & \text{se } k = -N, -N+2, \dots, -1 \\ x(k), & \text{se } k = 0, 1, \dots, N-1 \\ 0, & \text{para outros valores de } k, \end{cases} \quad (\text{A.1})$$

o que equivale a dois períodos da seqüência $x(k)$. Como resultado desse procedimento, obtém-se uma seqüência de $2N$ amostras, das quais as amostras de $k = -N$ a $k = -1$ são descartadas com o intuito de eliminar o transitório. Assim, pode-se dizer que

$$y(k) = h(k) * x'(k), \quad (\text{A.2})$$

com $k = 0, 1, \dots, N-1$, é a seqüência de saída do sistema.

Definindo-se a seqüência $y'(k)$ como sendo a duplicação da seqüência $y(k)$ da forma

$$y'(k) = \begin{cases} y(k+N), & \text{se } k = -N, -N+2, \dots, -1 \\ y(k), & \text{se } k = 0, 1, \dots, N-1 \\ 0, & \text{para outros valores de } k, \end{cases} \quad (\text{A.3})$$

calculando-se a correlação entre $y'(k)$ e $x(k)$ e descartando-se as amostras de $k = -N$ a $k = -1$ dessa correlação, tem-se

$$h'(k) = h(k) * x'(k) * x(-k), \quad (\text{A.4})$$

com $k = 0, 1, \dots, 2N - 1$.

Como a seqüência $x(k)$ aproxima um ruído branco, pode-se dizer que

$$h'(k) = h(k) * \sigma_x^2 \delta(k) = \sigma_x^2 h(k), \quad (\text{A.5})$$

onde $k = 0, 1, \dots, 2N - 1$ e $\delta(k)$ é o impulso unitário definido por

$$\delta(k) = \begin{cases} 1, & \text{se } k = 0 \\ 0, & \text{se } k \neq 0 \end{cases}. \quad (\text{A.6})$$

Dessa forma, pode-se dizer que

$$h'(k) \propto h(k). \quad (\text{A.7})$$

Apêndice B

Superfícies Espaciais de Resposta na Frequência—SFRSs

As Superfícies Espaciais de Resposta na Frequência (SFRSs—*Spatial Frequency Response Surfaces*) foram propostas em [28, 45] como uma alternativa de avaliação qualitativa dos métodos de interpolação de HTRFs.

Cada SFRS representa a magnitude em dB, para uma dada frequência, ao longo das posições no entorno do ouvinte. Portanto, para gerar essas superfícies, deve-se calcular a resposta na frequência do filtro relativo a cada posição e obter o valor da magnitude para cada frequência de interesse. Na Figura B.1, tem-se uma explicação gráfica de como se calcula cada SFRS.

Dessa forma, pode-se comparar para qualquer frequência o comportamento dos métodos de interpolação através dessas superfícies. No entanto, deve-se definir um conjunto de frequências para as quais se consiga julgar os métodos sem que se crie uma infinidade de superfícies.

Em [28], tem-se uma tabela que mostra o resultado global de algumas comparações entre SFRSs geradas a partir de funções medidas para pessoas diferentes. Nessas comparações, percebeu-se que, para uma dada frequência, as SFRSs não se alteram muito de pessoa para pessoa e algumas características se repetem. Com isso, pôde-se reunir nesta tabela as faixas de frequência e as características presentes em cada uma delas. De posse dessas informações, os autores daquele trabalho escolheram um determinado conjunto de dez frequências mais indicadas para a avaliação dos métodos de interpolação, de forma que se tivesse uma maior cobertura

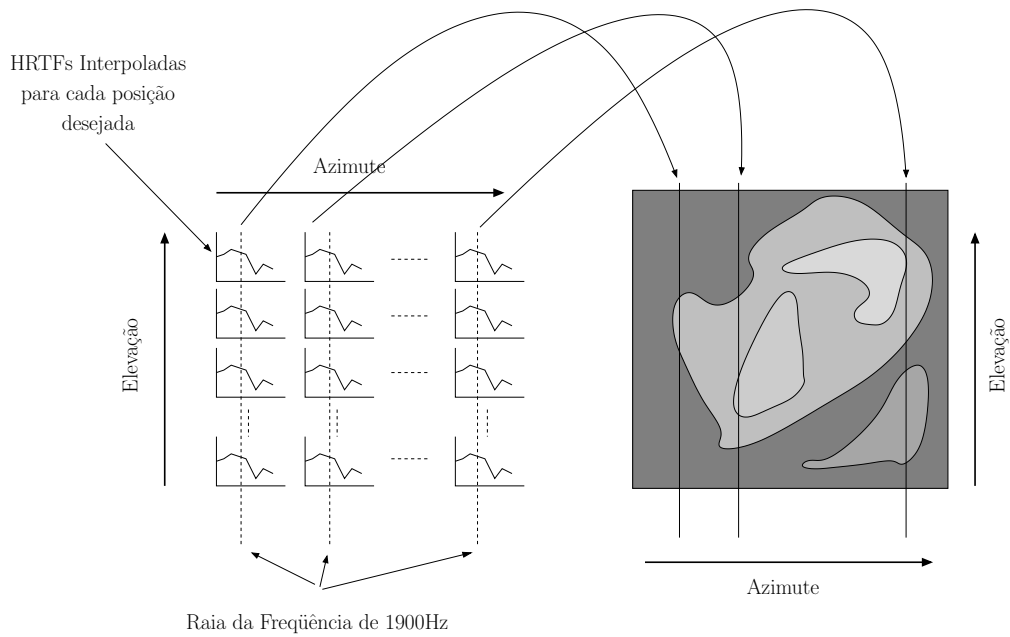


Figura B.1: SFRSs a partir das respostas em frequências interpoladas.

das prováveis diferenças entre eles. Esta lista de faixas de frequências, juntamente com as características mais importantes relativas a cada faixa, está repetida na Tabela B.1 [28].

Na presente tese, reutilizaram-se essas mesmas frequências, coerentemente com as conclusões de [28, 45].

Tabela B.1: Descrição das características das SFRSs por faixa de frequências.

Frequências	Descrição das características
1–600 Hz	Grosso modo, recebe-se de todas as direções uma potência igual.
0,6–1 kHz	Pode-se perceber o efeito da sombra da cabeça, quando a orelha ipsi-lateral recebe mais energia do que a contra-lateral. Efeitos de difração pela cabeça podem ser percebidos no lado contra-lateral da cabeça próximo a ± 100 graus.
1–2 kHz	O efeito da sombra da cabeça começa a aparecer mais fortemente; os efeitos de difração podem ser vistos claramente no lado contra-lateral. Dois ou três picos distintos no lado ipsi-lateral começam a se formar no azimute de 100 graus e elevações de ± 30 graus. Nas elevações baixas, no lado contra-lateral, pode-se notar a atenuação provocada pelo tronco.
2–2,5 kHz	Três picos podem ser vistos no lado ipsi-lateral nos azimutes de 70–80 graus, elevação -30 , -10 e 50 graus. Efeitos de difração podem ainda ser vistos no lado contra lateral próximo ao azimute de 100 graus. O efeito do tronco ainda pode ser percebido.
2,5–4kHz	Os três picos vistos no lado ipsi-lateral estão presentes, agora mais próximos do plano vertical e um pouco para cima (em elevação). Um quarto pico começa a aparecer no azimute de 40 graus e elevação de -50 graus. A difração começa a ficar menos forte à medida que o pico no azimute contra-lateral de 100 começa a desaparecer. Aparecem alguns pontos de baixa energia no lado ipsi-lateral e algum efeito da difração ainda pode ser percebido.
4–5 kHz	Os três picos no lado ipsi-lateral se transformam em um só, mais alto próximo do azimute ipsi-lateral de 50 graus e elevação 0. Os efeitos da difração quase desapareceram, mas o efeito do tronco ainda é proeminente para as elevações mais baixas.
5–6 kHz	O pico supracitado moveu-se para longe do plano central vertical e para cima ao longo da elevação. Esse pico se encontra na posição de azimute 75 graus e elevação 20 graus. O efeito do tronco ainda permanece.
6–8 kHz	O pico se transforma em outros dois menores ainda no azimute ipsi-lateral de 75 graus e elevação de ± 40 graus. Ainda persiste o efeito do tronco.
8–10 kHz	Os dois picos ainda estão presentes, porém o de elevação menor está mais presente do que o outro. Um terceiro pico começa a se formar no entorno do azimute 0 e elevação -30 graus. O efeito do tronco ainda persiste.
10–13 kHz	Agora quatro picos estão aparentes, um sobre o plano vertical no azimute 0 e elevação -20 graus, e outros três no azimute ipsi-lateral 100 nas elevações de -40 , 0 e 40 graus. O efeito do tronco ainda está presente.

Apêndice C

Dedução da Estrutura Farrow Bidimensional

Neste apêndice, descrevem-se os passos intermediários para obtenção das equações do modelo bidimensional para a estrutura Farrow. Esses passos intermediários são mostrados nas duas seções a seguir.

C.1 Desenvolvimento da Equação (7.13)

A partir da Equação (7.11) pode-se obter

$$J = \sum_{\omega} \sum_{\theta} \sum_{\phi} \left| \sum_l \sum_m \theta^l \phi^m \sum_n c_n^{l,m} [\cos \omega n T - j \operatorname{sen} \omega n T] - \sum_k h_k(\theta, \phi) [\cos \omega k T - j \operatorname{sen} \omega k T] \right|^2 = \quad (\text{C.1})$$

$$= \sum_{\omega} \sum_{\theta} \sum_{\phi} \left| \sum_l \sum_m \theta^l \phi^m \sum_n [c_n^{l,m} \cos \omega n T - j c_n^{l,m} \operatorname{sen} \omega n T] - \sum_k [h_k(\theta, \phi) \cos \omega k T - j h_k(\theta, \phi) \operatorname{sen} \omega k T] \right|^2 = \quad (\text{C.2})$$

$$= \sum_{\omega} \sum_{\theta} \sum_{\phi} \left| \sum_l \sum_m \theta^l \phi^m \sum_n c_n^{l,m} \cos \omega n T - j \sum_l \sum_m \theta^l \phi^m \sum_n c_n^{l,m} \operatorname{sen} \omega n T - \sum_k h_k(\theta, \phi) \cos \omega k T + j \sum_k h_k(\theta, \phi) \operatorname{sen} \omega k T \right|^2 = \quad (\text{C.3})$$

$$\begin{aligned}
&= \sum_{\omega} \sum_{\theta} \sum_{\phi} \left[\left(\sum_l \sum_m \theta^l \phi^m \sum_n c_n^{l,m} \cos \omega n T - \sum_k h_k(\theta, \phi) \cos \omega k T \right)^2 + \right. \\
&\quad \left. + \left(- \sum_l \sum_m \theta^l \phi^m \sum_n c_n^{l,m} \sin \omega n T + \sum_k h_k(\theta, \phi) \sin \omega k T \right)^2 \right], \tag{C.4}
\end{aligned}$$

que é a Equação (7.13).

C.2 Desenvolvimento da Equação (7.14)

A partir da Equação (7.13), cujo desenvolvimento pode ser visto na seção anterior, pode-se obter a expressão para a derivada de J com relação aos coeficientes $c_{n_0}^{l_0, m_0}$ fazendo-se

$$\begin{aligned}
\frac{\partial J}{\partial c_{n_0}^{l_0, m_0}} &= \sum_{\omega} \sum_{\theta} \sum_{\phi} \left[2 \left(\sum_l \sum_m \theta^l \phi^m \sum_n c_n^{l,m} \cos \omega n T - \sum_k h_k(\theta, \phi) \cos \omega k T \right) \theta^{l_0} \phi^{m_0} \cos \omega n_0 T - \right. \\
&\quad \left. - 2 \left(- \sum_l \sum_m \theta^l \phi^m \sum_n c_n^{l,m} \sin \omega n T + \sum_k h_k(\theta, \phi) \sin \omega k T \right) \theta^{l_0} \phi^{m_0} \sin \omega n_0 T \right] = \tag{C.5}
\end{aligned}$$

$$\begin{aligned}
&= 2 \sum_{\omega} \sum_{\theta} \sum_{\phi} \theta^{l_0} \phi^{m_0} \left[\cos(\omega n_0 T) \sum_l \sum_m \theta^l \phi^m \sum_n c_n^{l,m} \cos \omega n T - \right. \\
&\quad \left. - \cos(\omega n_0 T) \sum_k h_k(\theta, \phi) \cos \omega k T + \right. \\
&\quad \left. + \sin(\omega n_0 T) \sum_l \sum_m \theta^l \phi^m \sum_n c_n^{l,m} \sin \omega n T - \right. \\
&\quad \left. - \sin(\omega n_0 T) \sum_k h_k(\theta, \phi) \sin \omega k T \right] = \tag{C.6}
\end{aligned}$$

$$\begin{aligned}
&= 2 \sum_l \sum_m \sum_n \left\{ \sum_{\omega} \sum_{\theta} \sum_{\phi} \theta^{l_0} \phi^{m_0} \left[\cos(\omega n_0 T) \theta^l \phi^m c_n^{l,m} \cos \omega n T + \right. \right. \\
&\quad \left. \left. + \sin(\omega n_0 T) \theta^l \phi^m c_n^{l,m} \sin \omega n T \right] \right\} - \tag{C.7} \\
&\quad - 2 \sum_k \sum_{\omega} \sum_{\theta} \sum_{\phi} \theta^{l_0} \phi^{m_0} \left[\cos(\omega n_0 T) h_k(\theta, \phi) \cos \omega k T + \right. \\
&\quad \left. + \sin(\omega n_0 T) h_k(\theta, \phi) \sin \omega k T \right] =
\end{aligned}$$

$$= 2 \sum_l \sum_m \sum_n \left(\sum_\omega \sum_\theta \sum_\phi \theta^{l_0} \phi^{m_0} \{ \theta^l \phi^m c_n^{l,m} [\cos \omega n_0 T \cos \omega n T + \right. \\ \left. + \sin \omega n_0 T \sin \omega n T] \} \right) - \quad (C.8)$$

$$- 2 \sum_k \sum_\omega \sum_\theta \sum_\phi \theta^{l_0} \phi^{m_0} \{ h_k(\theta, \phi) [\cos \omega n_0 T \cos \omega k T + \\ + \sin \omega n_0 T \sin \omega k T] \} =$$

$$= 2 \sum_l \sum_m \sum_n \left\{ \sum_\omega \sum_\theta \sum_\phi \theta^{l_0} \phi^{m_0} \theta^l \phi^m c_n^{l,m} \cos(\omega n_0 T - \omega n T) \right\} - \quad (C.9)$$

$$- 2 \sum_k \sum_\omega \sum_\theta \sum_\phi \theta^{l_0} \phi^{m_0} h_k(\theta, \phi) \cos(\omega n_0 T - \omega k T) =$$

que é a Equação (7.14).

Apêndice D

Desenvolvimento das Equações (10.4) e (10.5) pela Expansão de Taylor

Considere a expansão de Taylor dada por

$$f(x+a) = f(x) + \frac{a}{1!}f'(x) + \frac{a^2}{2!}f''(x) + \dots \quad (\text{D.1})$$

Pode-se, então, obter da Equação (10.4)

$$O_E = A_E f(t) + B_E f[t - (T_a + T_S)] = \quad (\text{D.2})$$

$$= A_E f(t) + B_E f(t) - B_E(T_a + T_S)f'(t) + B_E \frac{(T_a + T_S)^2}{2} f''(t) - \dots \quad (\text{D.3})$$

Desprezando-se os termos de ordem superior, chega-se a

$$O_E = (A_E + B_E) \left[f(t) - \frac{B_E T_a + B_E T_S}{A_E + B_E} f'(t) \right]. \quad (\text{D.4})$$

Notando-se que o termo entre colchetes pode ser considerado igual aos dois primeiros termos de uma expansão de Taylor de $f\left(t - \frac{B_E T_a + B_E T_S}{A_E + B_E}\right)$, tem-se que

$$O_E \approx (A_E + B_E) f\left(t - \frac{B_E T_a + B_E T_S}{A_E + B_E}\right), \quad (\text{D.5})$$

que é a Equação (10.4).

Da mesma forma,

$$O_D = A_D f(t - T_a) + B_D f(t - T_S) \quad (\text{D.6})$$

pode ser escrita pela expansão de Taylor simplificada

$$O_D = A_D f(t) - A_D T_a f'(t) + B_D f(t) - B_D T_S f'(t) = \quad (D.7)$$

$$= (A_D + B_D) f(t) - (A_D T_a + B_D T_S) f'(t) = \quad (D.8)$$

$$= (A_D + B_D) \left[f(t) - \frac{A_D T_a + B_D T_S}{A_D + B_D} f'(t) \right]. \quad (D.9)$$

Como o termo entre colchetes é também formado pelos dois primeiros termos de uma expansão de Taylor, tem-se

$$O_D \approx (A_D + B_D) f\left(t - \frac{A_D T_a + B_D T_S}{A_D + B_D}\right), \quad (D.10)$$

que é a Equação (10.5).