

SOM TRIDIMENSIONAL: MÉTODOS DE GERAÇÃO E MODOS DE  
REPRODUÇÃO

Arnaldo Satoru Gunzi

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

Aprovada por:

---

Prof. Luiz Wagner Pereira Biscainho, D.Sc.

---

Dr. Fábio Pacheco Freeland, D.Sc.

---

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

---

Prof. Márcio Nogueira de Souza, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2008

GUNZI, ARNALDO SATORU

Som Tridimensional: Métodos de Geração e Modos de Reprodução [Rio de Janeiro] 2008

XIII, 146 p., 29,7 cm (COPPE/UFRJ, M.Sc., Engenharia Elétrica, 2008)

Dissertação - Universidade Federal do Rio de Janeiro, COPPE

1.Som Tridimensional 2.Conjuntos de HRTFs 3.Aproximação de Funções de Transferência 4.Panoramização por Amplitude

I.COPPE/UFRJ II.Título (série)

# Agradecimentos

Agradeço à minha família, eterna fonte de renovação de forças e apoio incondicional.

Aos meus orientadores Luiz Wagner Pereira Biscainho e Fábio Freeland, por todo o esforço, zelo e paciência dedicados a este trabalho.

Ao Maj. Av. Tiago Scheel por compreender a importância do mestrado acadêmico e me liberar do serviço sempre que necessário.

À Suzana Yae Ichigo e à Maria Fernanda Caravina, pelo auxílio na preparação dos desenhos.

Ao Alan Tygel, pela ajuda na realização dos testes subjetivos e à todos que participaram deste.

Ao Flávio Martins pela contribuição em geometria plana.

Ao Michel Tcheou por ter me apresentado o Laboratório de Processamento de Sinais (LPS) e por ter contribuído com diversas sugestões.

Ao Tadeu Ferreira pelo auxílio no LaTeX.

A todos os professores, funcionários e colegas do LPS, o meu profundo agradecimento.

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

## SOM TRIDIMENSIONAL: MÉTODOS DE GERAÇÃO E MODOS DE REPRODUÇÃO

Arnaldo Satoru Gunzi

Março/2008

Orientadores: Luiz Wagner Pereira Biscainho

Fábio Pacheco Freeland

Programa: Engenharia Elétrica

O objetivo idealístico do projetista de áudio 3D é ter o controle total do posicionamento de uma fonte virtual, sejam quais forem a posição relativa do ouvinte e o modo de reprodução da fonte sonora. Entretanto, os métodos de geração de áudio 3D dependem fortemente das condições de contorno impostas por esses fatores.

Este trabalho descreve e aperfeiçoa alguns métodos de geração de som tridimensional de acordo com o modo de reprodução adotado. Primeiramente apresentam-se conceitos e definições relacionados ao áudio 3D. Com relação à reprodução por fones de ouvido, faz-se uma comparação e análise entre conjuntos de HRTFs (*Head-Related Transfer Functions*) distintos. Tem-se depois uma comparação entre modelos teóricos e medidas experimentais de ITD (*Interaural Time Difference*). A seguir é feita a análise e reformulação da otimização da estrutura Farrow, que consiste num método de interpolação de HRTFs por parametrização de variáveis espaciais. Por fim, faz-se um estudo e generalização de métodos de panoramização por amplitude para reprodução em caixas acústicas.

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

TRIDIMENSIONAL SOUND: GENERATION METHODS AND  
REPRODUCTION SYSTEMS

Arnaldo Satoru Gunzi

March/2008

Advisors: Luiz Wagner Pereira Biscainho

Fábio Pacheco Freeland

Department: Electrical Engineering

The idealistic objective of a 3D audio designer is to have total control over the positioning of a virtual sound source, whatever are the relative position of the listener and the sound reproduction system. However, methods of generation of 3D audio depend strongly on conditions imposed by these factors.

This thesis describes and improves some methods of generation of three-dimensional sound in agreement with the reproduction system adopted. Firstly, concepts and definitions related to 3D audio are presented. Regarding the reproduction by earphones, comparison and analysis among sets of HRTFs (Head-Related Transfer Functions) are made. Then comparison between theoretical models and experimental measures of ITD (Interaural Time Difference) are shown. An analysis and reformulation in the optimization of the Farrow structure (a method of HRTF interpolation through direct angular parameterization) is carried out. Finally, a study and generalization of amplitude panning methods for loudspeakers is presented.

# Sumário

<b>Agradecimentos</b>	<b>iii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Apresentação . . . . .	1
1.2 Conceitos e Definições . . . . .	2
1.3 Áreas de Pesquisa do Áudio 3D . . . . .	8
1.3.1 Modos de Reprodução . . . . .	8
1.3.2 Métodos de Geração do Áudio 3D . . . . .	10
1.4 Conclusões e Desafios do Áudio 3D . . . . .	11
1.5 Conteúdo desta Dissertação . . . . .	13
<b>2 Análise e Comparação de Conjuntos de HRTFs</b>	<b>15</b>
2.1 Introdução . . . . .	15
2.2 Apresentação dos Conjuntos . . . . .	16
2.2.1 Conjunto de Gardner e Martin . . . . .	16
2.2.2 Conjunto do Projeto Listen . . . . .	18
2.2.3 Conjunto da CIPIC . . . . .	19
2.2.4 Resumo das Características dos Conjuntos . . . . .	21
2.3 Estudo em Relação à Distribuição de Energia . . . . .	23
2.3.1 Comportamento dos Conjuntos . . . . .	24
2.3.2 <i>Bright Spot</i> . . . . .	27
2.4 Estudo em Relação às Componentes Principais . . . . .	29
2.4.1 Representação usando a KLT . . . . .	29
2.4.2 Resultados da Formulação KLT . . . . .	30
2.5 Testes Subjetivos . . . . .	32

2.5.1	Descrição dos Testes . . . . .	33
2.5.2	Resultados e Distribuição Espacial das Respostas . . . . .	34
2.5.3	Análise do Erro Médio Quadrático . . . . .	45
2.5.4	Número de Reversões . . . . .	54
2.6	Conclusões . . . . .	55
<b>3</b>	<b>Modelos Geométricos de ITD</b>	<b>57</b>
3.1	Introdução . . . . .	57
3.2	ITD dos Conjuntos de Medida . . . . .	57
3.2.1	Método de Obtenção de HRTFs de Fase Mínima . . . . .	58
3.2.2	Extração do excesso de fase . . . . .	61
3.3	Apresentação dos Modelos Geométricos . . . . .	65
3.3.1	Trajetória Retilínea entre a Fonte Sonora e as Orelhas . . . . .	65
3.3.2	Trajetória com Efeito da Difração da Cabeça . . . . .	67
3.3.3	Utilização de Geometria Analítica . . . . .	69
3.4	Comparação dos Resultados . . . . .	69
3.5	Aplicação em interpolação de HRIRs . . . . .	77
3.6	Conclusões . . . . .	77
<b>4</b>	<b>Estrutura Farrow Generalizada</b>	<b>79</b>
4.1	Introdução . . . . .	79
4.2	Formulação . . . . .	80
4.2.1	Formulação Original . . . . .	80
4.2.2	Formulação Bidimensional . . . . .	82
4.2.3	Otimização Iterativa no Domínio da Freqüência . . . . .	83
4.2.4	Reformulação da Otimização . . . . .	84
4.2.5	Otimização Direta no Domínio do Tempo . . . . .	86
4.3	Validação da Reformulação da Otimização . . . . .	87
4.3.1	Tempo de Processamento . . . . .	88
4.3.2	Comparações entre as Respostas na Freqüência das Soluções . . . . .	89
4.3.3	Confronto entre Otimização Local e na Esfera Toda . . . . .	93
4.3.4	Investigação dos Efeitos da Variação de Ordens do Modelo . . . . .	97
4.4	Formulação da Estrutura Farrow para 3 Variáveis Espaciais . . . . .	100

4.5	Conclusões . . . . .	101
<b>5</b>	<b>Panoramização por Amplitude para Reprodução por Alto-Falantes</b>	<b>102</b>
5.1	Introdução . . . . .	102
5.2	Métodos Tradicionais de Panoramização por Amplitude . . . . .	103
5.2.1	Lei dos Senos . . . . .	103
5.2.2	Lei das Tangentes . . . . .	107
5.2.3	Panoramização por Amplitude numa Base Vetorial — VBAP .	110
5.2.4	Lei de Chowning . . . . .	111
5.3	Generalização da Lei dos Senos . . . . .	112
5.3.1	Formulação em 3 dimensões . . . . .	112
5.3.2	Generalização para $N$ Caixas Acústicas . . . . .	114
5.3.3	Derivação da Lei dos Senos 2D a partir da Lei dos Senos 3D .	114
5.3.4	Relação entre Lei dos Senos 3D e método VBAP . . . . .	115
5.3.5	Derivação da Lei de Chowning a partir da Lei dos Senos 3D .	116
5.3.6	Derivação da Interpolação Triangular a partir da Lei dos Senos 3D . . . . .	116
5.3.7	Derivação da Interpolação Bilinear a partir da Lei dos Senos 3D	118
5.4	Generalização da Lei das Tangentes . . . . .	120
5.4.1	Formulação em 3 dimensões . . . . .	120
5.4.2	Generalização da Lei das Tangentes 3D para $N$ Caixas Acústicas	120
5.4.3	Derivação da Lei das Tangentes 2D a partir da Lei das Tan- gentes 3D . . . . .	120
5.5	Não-unicidade dos pesos de panoramização . . . . .	121
5.6	Método Alternativo de Panoramização por Amplitude . . . . .	122
5.7	Comparação entre métodos . . . . .	122
5.8	Conclusões . . . . .	127
<b>6</b>	<b>Conclusões</b>	<b>128</b>
<b>A</b>	<b>ITD - Trajetória Reta em 3D</b>	<b>136</b>
<b>B</b>	<b>Desenvolvimento da Lei das Tangentes</b>	<b>139</b>



C	Expansão e Simplificação por Série de Taylor	141
D	Equivalência entre a Lei dos Senos 3D e a Interpolação Triangular	143
E	Equivalência entre a Lei dos Senos 3D e a Interpolação Bilinear	145

# Lista de Figuras

1.1	Planos de referência. . . . .	4
1.2	Coordenadas polares interaurais. . . . .	4
1.3	Efeito do comprimento de onda na percepção de fase. . . . .	6
1.4	Ilustração do cone de confusão. . . . .	7
1.5	Áreas do áudio 3D abordadas neste trabalho. . . . .	8
1.6	Modos de reprodução do áudio 3D. . . . .	9
1.7	Métodos de geração do áudio 3D. . . . .	10
1.8	Conteúdo do trabalho. . . . .	13
2.1	Conteúdo do capítulo sobre HRTFs. . . . .	16
2.2	Posições de medida do conjunto de Gardner . . . . .	17
2.3	HRIR do conjunto de Gardner no ponto $(0,0)$ . . . . .	17
2.4	HRTF do conjunto de Gardner no ponto $(0,0)$ . . . . .	18
2.5	Posições de medida do conjunto do Projeto Listen . . . . .	19
2.6	HRIR do conjunto do Projeto Listen no ponto $(0,0)$ . . . . .	20
2.7	HRTF do conjunto do Projeto Listen no ponto $(0,0)$ . . . . .	20
2.8	Posições de medida do conjunto da CIPIC . . . . .	21
2.9	HRIR do conjunto da CIPIC no ponto $(0,0)$ . . . . .	22
2.10	HRTF do conjunto da CIPIC no ponto $(0,0)$ . . . . .	22
2.11	Comportamento comparativo dos conjuntos no ponto $(0,0)$ . . . . .	23
2.12	Visão isométrica da variação da energia . . . . .	25
2.13	Curvas de nível de energia . . . . .	26
2.14	Energia por azimutes $-180^\circ$ a $180^\circ$ . . . . .	28
2.15	Distribuição de energia por número de autovetores para cada conjunto. . . . .	31

2.16	Posições Espaciais e Convenção Adotada. Coordenadas $(\theta, \phi)$ em graus. Para o azimute, adotaram-se E = esquerda, C = centro, D = direita; para a elevação, adotaram-se A = alto, H = horizontal, B = baixo. . . . .	34
2.17	Distribuição espacial das respostas, para a posição de teste FHC (0, 0). . . . .	41
2.18	Distribuição espacial das respostas, posição de teste MAE . . . . .	42
2.19	Distribuição espacial das respostas, posição de teste MHE . . . . .	43
2.20	Distribuição espacial das respostas, posição de teste TBE . . . . .	44
2.21	MSE por conjunto e posição. . . . .	52
3.1	Conteúdo do capítulo sobre a ITD. . . . .	58
3.2	Fluxograma de obtenção da seqüência de fase mínima. . . . .	60
3.3	Fase do conjunto original e de fase mínima . . . . .	62
3.4	Excesso de fase do conjunto original e fase do atraso puro. . . . .	64
3.5	Modelo de ITD com trajetória retilínea para o caso 2D. . . . .	65
3.6	ITD trajetória retilínea para o caso 3D. . . . .	67
3.7	Detalhe da trajetória retilínea para o caso 3D. . . . .	67
3.8	ITD com difração da cabeça para o caso 2D. . . . .	68
3.9	ITD medida e calculada - Gardner . . . . .	70
3.10	Erros dos modelos de ITD para o conjunto de Gardner. . . . .	71
3.11	ITD medida e calculada - Projeto Listen . . . . .	72
3.12	Erros dos modelos de ITD para o conjunto do Projeto Listen. . . . .	73
3.13	ITD medida e calculada - CIPIC . . . . .	74
3.14	Erros dos modelos de ITD para o conjunto da CIPIC. . . . .	76
4.1	Conteúdo do capítulo sobre a estrutura Farrow. . . . .	80
4.2	Diagrama em blocos da estrutura Farrow original. . . . .	81
4.3	Diagrama em blocos da estrutura Farrow generalizada . . . . .	83
4.4	Caminhos de teste. . . . .	87
4.5	Comparação entre formulações para o caminho horizontal. . . . .	89
4.6	Comparação entre formulações para o caminho diagonal. . . . .	90
4.7	Comparação entre formulações para o caminho vertical. . . . .	91
4.8	Respostas das soluções no tempo e na freqüência . . . . .	92

4.9	Otimização numa região e na esfera toda para o caminho horizontal. . .	93
4.10	Otimização numa região e na esfera toda para o caminho diagonal. . .	94
4.11	Otimização numa região e na esfera toda para o caminho vertical. . .	95
4.12	Respostas na frequência numa região limitada e na esfera toda . . . .	96
4.13	Comparação entre ordens diferentes - caminho horizontal. . . . .	97
4.14	Comparação entre ordens diferentes - caminho diagonal. . . . .	98
4.15	Comparação entre ordens diferentes - caminho vertical. . . . .	99
5.1	Conteúdo do capítulo sobre panoramização por amplitude. . . . .	103
5.2	Diagrama para estudo do mecanismo de estereofonia direcional. . . .	104
5.3	Chegada dos sinais às orelhas. . . . .	105
5.4	Figura ilustrativa para a obtenção da ITD. . . . .	107
5.5	Posições relativas com rotação. . . . .	108
5.6	Chegada dos sinais nas orelhas, com rotação. . . . .	109
5.7	Esquema genérico para Lei dos Senos 3D. . . . .	112
5.8	Relação entre a Lei dos Senos 3D e outros métodos. . . . .	114
5.9	Posições relativas das caixas acústicas e da fonte virtual. . . . .	117
5.10	Grade de posicionamento das caixas acústicas na interpolação bilinear.	119
5.11	Posições dos triângulos utilizados para testes. . . . .	123
5.12	Interpolação para o Triângulo 1, para demonstrar o efeito das altas frequências. . . . .	124
5.13	Resultados obtidos para os triângulos descritos. . . . .	126
6.1	Estrutura detalhada da Dissertação. . . . .	129
A.1	Modelo de ITD com trajetória reta para o caso 3D. . . . .	136
A.2	Representação esquemática da Figura A.1. . . . .	137
A.3	Ângulos do triângulo ADL. . . . .	138

# Lista de Tabelas

1.1	Definições do tipo de sinal de áudio dependendo do estímulo. . . . .	2
2.1	Características dos conjuntos de HRIRs. . . . .	23
2.2	Número de autovetores para representar 99,9% da energia total. . . .	31
2.3	Número de autovetores por porcentagem de energia. . . . .	32
2.4	Coordenadas relativas a cada posição. Valores $(\theta, \phi)$ em graus. . . . .	35
2.5	Tabela de confusão do teste para o conjunto de Gardner - Valores percentuais. . . . .	38
2.6	Tabela de confusão do teste para o conjunto do Projeto Listen - Valores percentuais. . . . .	39
2.7	Tabela de confusão do teste para o conjunto da CIPIC - Valores percentuais. . . . .	40
2.8	Estatísticas dos conjuntos quanto ao azimute . . . . .	49
2.9	Estatísticas dos conjuntos quanto a elevação . . . . .	50
2.10	Estatísticas dos conjuntos, com retirada de reversões. . . . .	51
2.11	Porcentagem de reversões por posição de teste. . . . .	54
3.1	Notação dos tipos de transformações cepstrais. . . . .	59
3.2	Erros dos modelos de ITD para o conjunto de Gardner. . . . .	71
3.3	Erros dos modelos de ITD para o conjunto do Projeto Listen. . . . .	73
3.4	Erros dos modelos de ITD para o conjunto da CIPIC. . . . .	75
4.1	Tempo de processamento da otimização. . . . .	88
5.1	Resumo dos métodos de Panoramização por Amplitude . . . . .	127

# Capítulo 1

## Introdução

### 1.1 Apresentação

O ser humano possui a capacidade de localizar fontes sonoras no ambiente através do sentido da audição. Dado um determinado estímulo sonoro, o ouvinte consegue dizer aproximadamente a direção e a distância de sua fonte geradora, e ainda o tipo de ambiente em que está inserido.

O áudio 3D pode ser utilizado num contexto em que se queira aumentar a sensação de realidade do usuário. Possíveis aplicações são encontradas em jogos eletrônicos, cinemas, aparelhos de *home-theater* e simuladores.

Do ponto de vista comercial, a implantação de um sistema de áudio 3D pode representar uma característica diferencial de um produto em relação à concorrência. No mercado de *home-theater*, por exemplo, há diversos fabricantes e padrões concorrendo para se destacar, e o uso do áudio 3D pode contribuir para tal.

O processamento de áudio 3D trata da modificação de um sinal monofônico para gerar a ilusão de um ambiente virtual e de localização das fontes sonoras, explorando a capacidade humana de ouvir em 3 dimensões.

O objetivo idealístico do projetista de áudio 3D envolve a completa e total manipulação da percepção espacial de uma pessoa [1]. Designações equivalentes de “Som 3D” incluem **acústica virtual**, **áudio binaural** e **som espacializado** [1].

Existem diversos métodos para se produzir o efeito de tridimensionalidade. Cada tipo de aplicação do som 3D tem uma exigência particular. Assim, pode ser necessário processar sons em tempo real, ou ainda realizar uma implementação

variante no tempo, permitindo a simulação de fontes sonoras em movimento [2]. A fim de se aumentar a sensação de realidade, algumas aplicações podem requerer a síntese de reflexões e reverberações de salas [2].

Para se fazer a escolha de um desses métodos, há basicamente três fatores que devem ser considerados: a complexidade computacional envolvida, a qualidade do áudio obtenível e a quantidade de memória necessária. De modo geral, quanto maior a qualidade requerida, mais se exige em termos de processamento e quantidade de dados armazenados; diminuindo-se a exigência de qualidade, pode-se obter métodos menos complexos e que exijam menor banco de dados. Cabe ao projetista escolher o método adequado, dependendo da aplicação a que ele se destina e das condições de contorno como limitações de *hardware*, necessidade de processamento *on-line*, modo de reprodução e outros.

## 1.2 Conceitos e Definições

Esta seção apresenta os principais conceitos e definições adotados nesta dissertação.

### Sinal de Áudio Monoaural e Binaural

Segundo Huopaniemi [3], o termo monoaural refere-se ao estímulo apresentado em apenas um ouvido. Dicótico refere-se ao estímulo que é diferente nos dois ouvidos, diótico é o estímulo que é igual nos ouvidos, e binaural é qualquer estímulo dicótico ou diótico [3]. A Tabela 1.1 sumariza essas definições.

A fim de evitar ambigüidade de terminologia, nesta dissertação “áudio binaural” será equivalente a “som 3D”, e “sinal binaural” será “qualquer estímulo dicótico ou diótico”, conforme a definição da Tabela 1.1.

Tabela 1.1: Definições do tipo de sinal de áudio dependendo do estímulo.

Sinal	Estímulo
Monoaural	em uma orelha
Dicótico	diferente nas orelhas
Diótico	igual nas orelhas
Binaural	dicótico ou diótico

## Panoramização de áudio

A panoramização (ou *panning*) refere-se à técnica de processar um sinal monoaural de forma a convertê-lo em binaural, de modo a projetar o áudio numa determinada posição virtual, criando a ilusão de tridimensionalidade.

### Sistema de Coordenadas Polares Interaural

A fim de se fazer a panoramização, deve-se definir a localização espacial das posições em que se quer colocar o som virtual. Para tal, o Sistema de Coordenadas Polares Interaural é comumente utilizado no contexto do áudio 3D. Tal sistema descreve a localização de um ponto qualquer em relação a um determinado ouvinte.

Os Planos de Referência com relação ao ouvinte são os planos coordenados aos quais cada ponto no espaço é referenciado. Os principais planos de referência são os planos horizontal, mediano e frontal, ilustrados na Figura 1.1.

O centro do sistema de coordenadas é coincidente com o centro do segmento de reta que liga as duas orelhas. O **plano horizontal** é o que passa pelo centro da cabeça, pelo nariz e pelas orelhas do indivíduo. O **plano mediano** é o que passa pelo centro da cabeça e pelo nariz do indivíduo e é perpendicular ao plano horizontal. O **plano frontal** passa pelo centro da cabeça e pelas orelhas do indivíduo e é perpendicular aos outros dois planos, conforme a Figura 1.1.

Um ponto qualquer numa esfera pode ser descrito por dois ângulos: azimute e elevação. O azimute refere-se ao ângulo no plano horizontal, com azimute zero na direção do nariz do indivíduo. A elevação refere-se ao ângulo no plano que passa pelo centro da cabeça e pelo ponto onde se localiza a fonte e é perpendicular ao plano horizontal, arbitrando-se elevação zero à posição no plano horizontal. O azimute é positivo no lado direito da cabeça, e a elevação é positiva acima do plano horizontal, conforme a Figura 1.2. Nesta dissertação, o azimute será denominado  $\theta$  (teta) e a elevação,  $\phi$  (fi).

No ponto diretamente acima do ouvinte, a elevação é de  $90^\circ$ , e no ponto diretamente abaixo é de  $-90^\circ$ . Quanto ao azimute, adota-se  $90^\circ$  como a posição à direita do indivíduo, e  $-90^\circ$  como a posição à esquerda do mesmo.

Os valores de azimute e elevação são capazes de indicar a localização de um ponto sobre uma esfera de raio fixo. Para localizar um ponto em qualquer lugar do



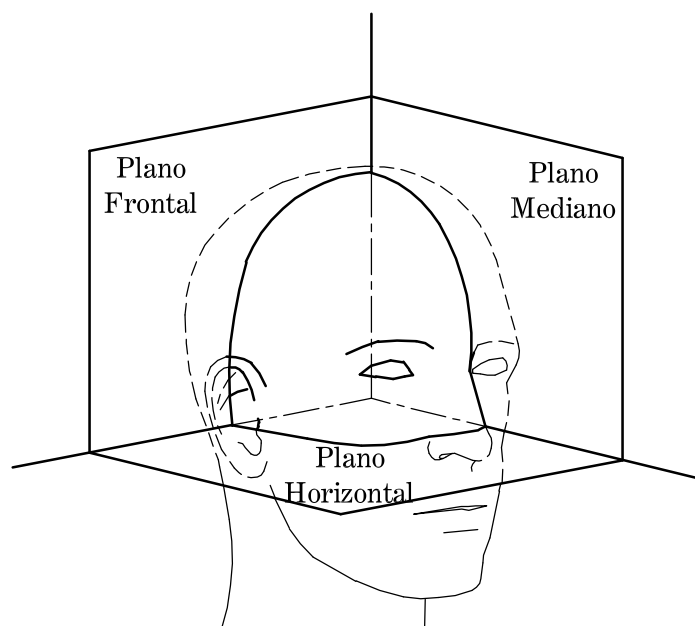


Figura 1.1: Planos de referência.

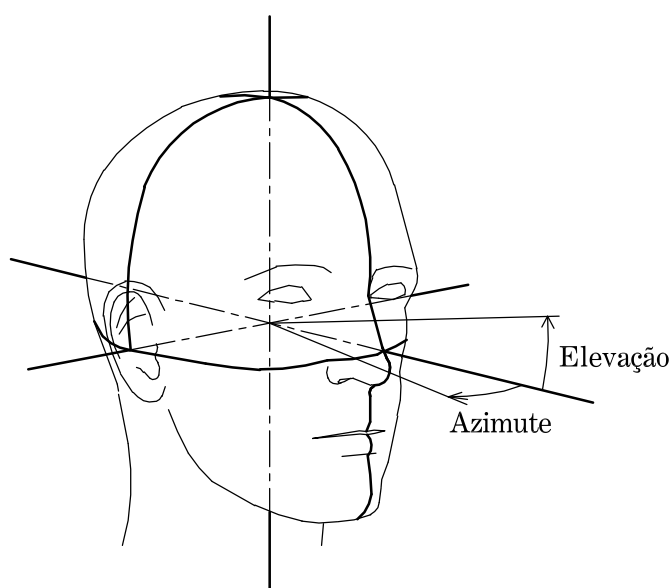


Figura 1.2: Coordenadas polares interaurais.

espaço 3D, há a necessidade de mais uma variável, o raio da esfera, denominado  $r$ .

Nota-se que o sistema de coordenadas descrito muda em função do movimento da cabeça [4].

## Pantofonia e Perifonia

Sendo a panoramização do áudio um estudo complexo, as técnicas de processamento de áudio 3D incluem diversas tentativas de se limitar a região de projeção, a fim de simplificar a formulação do problema.

A pantofonia é a panoramização apenas no plano horizontal [5]. Perifonia refere-se à panoramização numa esfera [5], considerando o raio da esfera fixo.

## Teoria Dúplex

A teoria dúplex foi introduzida por Lord Rayleigh para elucidar o mecanismo de audição direcional no plano horizontal [1, 6]. Segundo a definição anterior, ela é uma forma de pantofonia.

Ela afirma que há duas dicas auditivas principais na localização espacial: a diferença de tempo interaural (ITD) e a diferença de nível interaural (ILD).

A ITD refere-se à diferença do tempo de chegada de uma frente de onda sonora entre a orelha esquerda e a direita, e é freqüentemente citada como a mais importante dica auditiva [7]. Analogamente, a ILD é definida como a diferença de amplitude entre as ondas que chegam nas orelhas esquerda e direita. Na literatura, também encontram-se os termos IID, referindo-se à diferença de intensidade interaural, e ILR, referindo-se à razão de nível interaural.

A rigor, a ILD é uma função dependente da freqüência, já que a presença do torso, da cabeça e da pina produzem efeitos de difração, refração e reflexão diferentes para cada freqüência. Entretanto, é comum na literatura considerar a ILD como sendo um ganho médio, independente da freqüência.

A percepção do posicionamento pela ITD baseia-se na hipótese de a fonte sonora estar mais próxima da orelha em que o som chegar primeiro. Ela é válida principalmente nas baixas freqüências [6], porque, considerando uma senóide pura, o efeito de lateralização é proporcional à fase percebida da senóide. Para altas freqüências a ITD torna-se ambígua, já que a lateralização não mais será proporcional à fase percebida, conforme ilustra a Figura 1.3. Em [6] considera-se a freqüência de 1.500 Hz como a divisora entre altas e baixas freqüências, já que tem um comprimento de onda comparável às dimensões da cabeça humana.

Já a percepção do posicionamento pela ILD baseia-se na hipótese de que

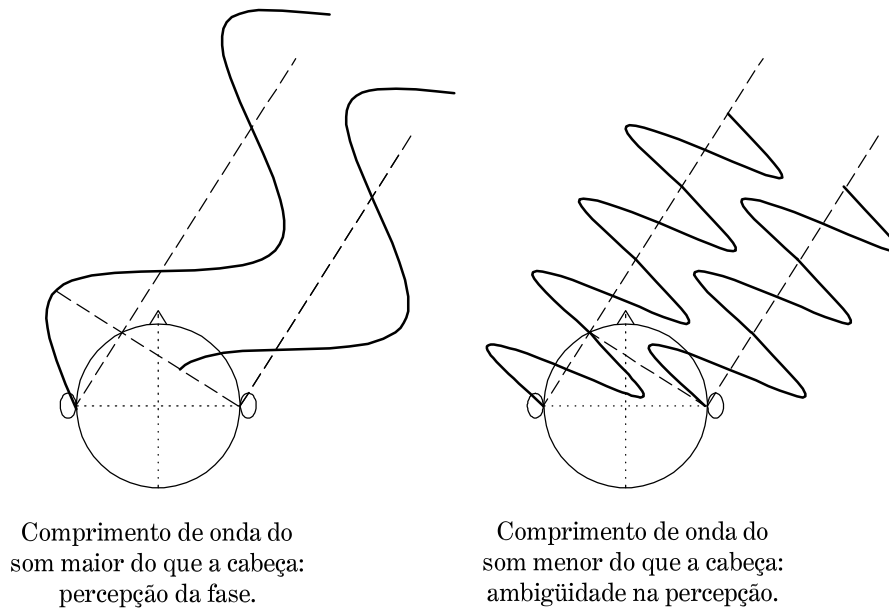


Figura 1.3: Efeito do comprimento de onda na percepção de fase.

o som chegará com mais energia na orelha mais próxima da fonte sonora. Ela é válida principalmente nas altas frequências (acima de 1.500 Hz) [6], uma vez que, nas baixas frequências, ocorre o efeito da difração. A difração é a capacidade de a frente de onda contornar obstáculos, quando o comprimento da onda é maior do que as dimensões do obstáculo, o que faz com que o som chegue com intensidade semelhante em ambas as orelhas.

### Ambigüidade das dicas de ITD e ILD: Cone de Confusão

Há mais do que uma posição espacial possível que gera os mesmos valores de ITD e ILD, considerando a ILD independente da frequência. Imaginando uma simplificação da cabeça humana como uma esfera perfeita, o lugar geométrico dos pontos que geram os mesmos valores de ITD e ILD forma um hiperbolóide de revolução, com origem no centro geométrico da cabeça e focos nas duas orelhas. As retas assíntotas do hiperbolóide formam o “cone de confusão”, com centro na origem do sistema de coordenadas, conforme a Figura 1.4.

Apenas o conhecimento dos valores de ITD e ILD independentemente da frequência não é capaz de resolver a ambigüidade da posição sobre o cone. Para tal, é necessário lidar com outras dicas espaciais da localização [1]. As pinas geram a

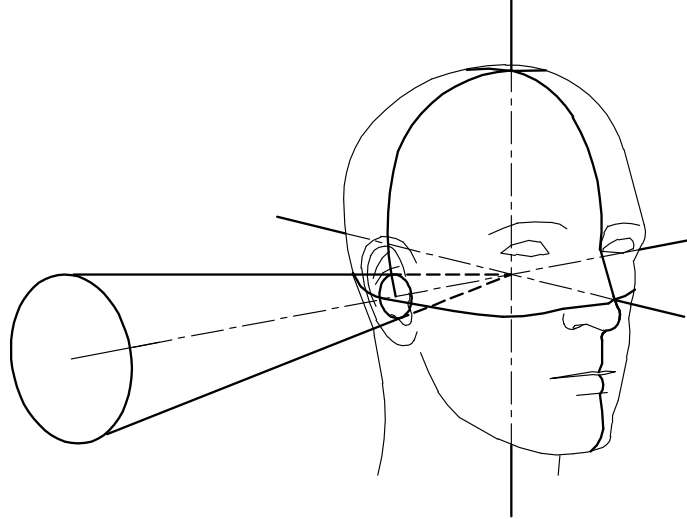


Figura 1.4: Ilustração do cone de confusão.

maior parte das transformações espectrais dependentes da frequência responsáveis pelas dicas auditivas [1].

### ***Head-Related Transfer Function***

O conjunto de modificações espectrais sofridas pelo som no percurso entre uma fonte sonora e o canal auditivo, incluindo as distorções causadas pela presença da cabeça, do torso e da pina, é comumente denominado *Head-Related Transfer Function* (HRTF). A resposta ao impulso correspondente à HRTF é chamada *Head-Related Impulse Response* (HRIR).

As HRTFs variam significativamente não apenas com as variáveis de posição, mas de pessoa para pessoa [8].

Um banco de dados de HRTFs pode ser obtido através de medições com auxílio de um manequim, dentro de uma câmara anecóica, com a fonte sonora em alguns pontos numa esfera ao redor do ouvinte [2].

O uso das HRTFs obtidas através de medidas diretas ou modelagem é componente-chave de um sistema de som 3D para fones de ouvido [1].

A utilização das HRIRs como filtros para se fazer a panoramização é uma técnica de perifonia. Ao se incorporar o comportamento de módulo e fase dependentes da frequência, consegue-se resolver a ambigüidade do cone de confusão, descrito

anteriormente.

Outra forma de utilização das HRTFs pode ser em conjunto com modelos geométricos de propagação do som em salas, incorporando as reflexões da parede e a absorção no ar [2], afim de aumentar a sensação de realidade do ouvinte.

## 1.3 Áreas de Pesquisa do Áudio 3D

Embora haja uma gama bastante ampla de pesquisas relacionadas ao áudio 3D, esta seção terá como escopo a apresentação do tema quanto a duas áreas:

- Modos de reprodução; e
- Métodos de geração da panoramização.

A Figura 1.5 ilustra as áreas de pesquisa mencionadas, que serão melhor discutidas a seguir.

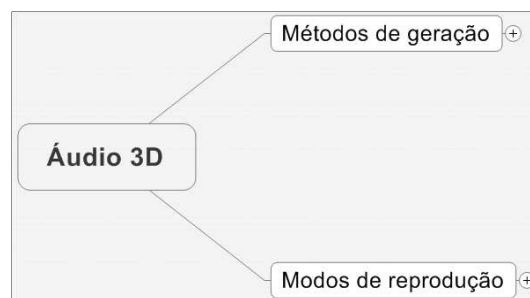


Figura 1.5: Áreas do áudio 3D abordadas neste trabalho.

### 1.3.1 Modos de Reprodução

Há basicamente dois modos de reprodução do som: fones de ouvido e alto-falantes. O conhecimento do modo de reprodução influi diretamente na escolha da técnica para realização da panoramização do som.

Quanto à reprodução por fones de ouvido, a grande vantagem é que o som é entregue ao ouvinte exatamente na forma que o projetista deseja, sem depender do tipo de ambiente em que o ouvinte está inserido.

Os fones de ouvido são a escolha para controle máximo da panoramização do som [9]. Entretanto, não é sempre que se pode fazer com que o ouvinte utilize

fonos de ouvido. Além disso, o som pode parecer internalizado, ou seja, o ouvinte pode ter a impressão de que a fonte virtual está dentro de sua cabeça. Outro fator a se considerar é o movimento do ouvinte. Quando este se move, todo o ambiente virtual parecerá estar se movendo junto, o que nem sempre corresponde à situação desejada. Para compensar o movimento do ouvinte, e reposicionar a fonte virtual, pode-se utilizar algum mecanismo de monitoração da cabeça (*head-tracking*).

O método de reprodução com alto-falantes apresenta a vantagem de ser externo ao ouvinte, não causando incômodo, como poderia ser no caso anterior. Pela mesma razão, a utilização de alto-falantes facilita a externalização do som. Os maiores problemas desse tipo de reprodução são a interferência do ambiente e o efeito da diafonia (*crosstalk*), que devem ser considerados e compensados pelo projetista, para que este tenha total controle sobre a panoramização do som virtual.

Dentre os métodos de reprodução com alto-falantes temos o sistema estéreo, com duas caixas acústicas, o multicanal, com mais de duas caixas acústicas, e o sistema de *array* de alto-falantes. A Figura 1.6 ilustra os modos de reprodução do áudio 3D.



Figura 1.6: Modos de reprodução do áudio 3D.

A maior quantidade de caixas acústicas facilita a panoramização do som virtual; entretanto, afeta o custo de forma diretamente proporcional. Atualmente, o sistema com duas caixas acústicas é o mais comum, sendo portanto o de maior importância. Os sistemas multicanais são utilizados nos sistemas de *home-theater* e em outras situações particulares, como em cinemas, por exemplo. O sistema com *array* de alto-falantes consiste de um número muito grande de alto-falantes dispostos lado a lado, de modo que se tenta produzir uma frente de onda sonora, numa técnica chamada *Wave Field Synthesis* [10, 11] .

### 1.3.2 Métodos de Geração do Áudio 3D

Com relação à geração do efeito de panoramização do som no espaço tridimensional, há basicamente dois fatores a considerar: a localização da fonte sonora e a ambientação. Grande parte dos estudos relativos à localização da fonte sonora considera apenas sua localização angular numa esfera de raio fixo, ou seja, a perifonia.

Uma das formas mais comuns de perifonia consiste na utilização de HRTFs como filtros. O conjunto de HRTFs pode ser obtido através de medidas individuais [12, 13] ou não-individuais [14], por modelagem numérica [15] ou através da solução analítica de modelos físicos simplificados [8, 16], conforme ilustra a Figura 1.7. As medidas são feitas num ambiente anecóico, de forma a não se considerar a influência do ambiente.



Figura 1.7: Métodos de geração do áudio 3D.

As HRTFs individualizadas são obtidas através de medidas no próprio ouvinte. Apesar de o conjunto de medidas descrever bem as noções de posição para o indivíduo em questão, esse conjunto não necessariamente resultará no mesmo desempenho quando aplicado a outro ouvinte. Um grande problema é o fato de as características espaciais de HRTFs diferirem entre indivíduos, e portanto os erros de localização aumentarem com a utilização de HRTFs não-individualizadas [9].

Quanto a um conjunto de HRTFs não-individualizado, o seu objetivo é o de simular um indivíduo médio, o que pode ser feito através de um manequim. O

processo de obtenção das HRTFs é menos custoso, por não ser necessário obter um conjunto novo de medidas para cada indivíduo.

Os modelos numéricos de obtenção das HRTFs consistem em simulações numéricas do efeito de uma fonte virtual sobre um modelo virtual de um indivíduo.

Os métodos simplificados consideram modelos físicos, como por exemplo uma cabeça esférica sem pina, de modo que se possa formular as equações físicas relacionadas e resolvê-las analiticamente.

A utilização de HRTFs é apenas um dos muitos métodos existentes para gerar o som tridimensional. Nesta dissertação também serão estudados métodos de panoramização por amplitude.

Além dos estudos de perifonia, um sistema de geração de áudio 3D pode levar em consideração a noção de distância.

Os métodos de geração descritos até então, com a utilização de HRTFs medidas, são feitos considerando uma fonte virtual num ambiente anecóico. Pode-se inserir efeitos de ambientação aos métodos descritos, o que consiste em fazer com que a fonte sonora esteja inserida num ambiente reverberante, como uma sala de espetáculos, por exemplo. Uma das formas de se considerar o ambiente é incluir um modelo para gerar o efeito reverberante da sala, utilizando um método como o das imagens [17, 18]. Outra forma é obter a resposta ao impulso da sala através de medidas e acrescentá-la posteriormente ao modelo de panoramização.

## 1.4 Conclusões e Desafios do Áudio 3D

Este capítulo mostrou os principais conceitos e definições utilizados no contexto de áudio 3D. A seguir, apresentou as principais áreas de pesquisa abordadas neste trabalho.

No contexto apresentado nas seções anteriores, os maiores erros perceptuais relacionados à reprodução do som 3D estão listados a seguir. O “erro” se refere ao desacordo entre a intenção do projetista e a percepção do ouvinte [9].

- Localização do som virtual interna à cabeça com a reprodução por fones de ouvido [2, 9];



- Reversões frente-trás [2, 9]; e
- Dispersão da localização do som (*blur*) [9].

O primeiro item se refere à impressão de que os sons estão “internalizados” na cabeça do indivíduo. Ocorre principalmente na reprodução por fones de ouvido, já que na reprodução por alto-falantes estes estão distantes da (em outras palavras, externos à) cabeça do indivíduo. Uma das formas de se minimizar essa impressão é inserir no modelo a reverberação da sala.

As reversões frente-trás referem-se à impressão de que um som projetado numa posição à frente ou atrás do indivíduo esteja atrás ou à frente do mesmo, respectivamente. Uma das formas de se minimizar tal problema é através de um sistema 3D com monitoração da cabeça. Isso porque, quando numa situação em que não se tem certeza da posição, o ser humano costuma mover a sua cabeça, de forma a ter mais informações sobre a fonte sonora e resolver tal ambigüidade. O sistema de monitoramento da cabeça coleta dados do movimento do indivíduo, o que permite que o sistema de geração de áudio 3D repositicione a fonte virtual para corresponder à posição relativa desejada, reproduzindo o mesmo efeito que o ser humano perceberia na realidade. Pelo mesmo motivo, o monitoramento da posição da cabeça pode fazer com que as simulações sejam menos dependentes de outros fatores, como o uso de HRTFs medidas não individualizadas e a equalização dos fones de ouvido [2]. Outra proposta para se diminuir o problema da reversão frente-trás é a de treinar as pessoas para se adaptarem às HRTFs não individualizadas.

O erro de dispersão da localização do som refere-se ao espalhamento das posições apontadas pelos ouvintes ao tentar indicar a posição exata do som virtual. O projetista pode aproveitar a característica da percepção do ser humano ser mais imprecisa em umas regiões que em outras, tornando o sistema mais simples sem afetar sua qualidade.

Técnicas como um sistema de monitoramento da cabeça e a simulação de reverberações podem tornar o sistema de geração do áudio 3D melhor e facilitar a percepção de localização das pessoas.

Além das dificuldades citadas, tem-se o desafio de criar um método de panoramização unificado que funcione igualmente bem em todos os modos de reprodução descritos. Tal sistema deve levar em consideração cada particularidade dos modos

de reprodução do som vistos na Subseção 1.3.1, de modo a compensar efeitos indesejados.

## 1.5 Conteúdo desta Dissertação

O diagrama da Figura 1.8 ilustra o conteúdo a ser abordado neste trabalho.



Figura 1.8: Conteúdo do trabalho.

O Capítulo 1 introduziu os conceitos que serão utilizados nesta dissertação, apresentando as definições e o embasamento teórico relacionados ao processamento de áudio 3D, além de descrever algumas das maiores dificuldades e desafios dessa área.

O Capítulo 2 versa sobre a descrição, análise e comparação de diferentes conjuntos de HRTFs disponíveis na literatura. É feita a padronização da notação dos conjuntos, comparação em termos de distribuição de energia e estudo em relação às componentes principais. Por fim, fazem-se testes subjetivos para comparação da percepção da posição estática do som virtual entre os conjuntos.

O Capítulo 3 analisa modelos para estimativa de ITD e testa-os com relação a medidas dos conjuntos de HRTFs apresentados anteriormente. Apresenta-se uma técnica de estimativa de ITD a partir dos conjuntos de HRTFs. Por outro lado, são apresentados modelos obtidos através de considerações geométricas. Faz-se a comparação entre as medidas e os modelos, para testar a validade destes.

O Capítulo 4 faz a análise e a reformulação da otimização de uma técnica de interpolação de áudio 3D para fones de ouvido, a estrutura Farrow. A reformulação torna possível diminuir o tempo para a obtenção dos coeficientes da estrutura. A seguir, é feita a validação da nova formulação contra a antiga. É feita a análise da otimização numa região limitada e da otimização na esfera toda. Finalmente, faz-se uma investigação dos efeitos da variação da ordem do modelo.

O Capítulo 5 trata dos principais conceitos de geração de áudio 3D para o modo de reprodução por alto-falantes. São apresentados diversos métodos de panoramização por amplitude. É mostrada uma generalização da Lei dos Senos e Lei das Tangentes para 3D e estuda-se a sua relação com os métodos apresentados anteriormente.

Finalmente, o Capítulo 6 apresenta as conclusões da dissertação com as propostas de trabalhos futuros.

# Capítulo 2

## Análise e Comparação de Conjuntos de HRTFs

### 2.1 Introdução

Conforme visto no Capítulo 1, um conjunto de HRTFs é fundamental para a síntese do áudio 3D, no contexto de panoramização sobre uma esfera, ou perifonia. Daí surge a dependência entre a qualidade do conjunto de HRTFs e dos sistemas de áudio 3D.

Há diversos conjuntos de HRTFs disponíveis na literatura, cada um com suas características particulares. O objetivo do capítulo é analisar as diferenças e tentar extrair as características comuns aos conjuntos citados.

Primeiramente é feita a apresentação de três conjuntos diferentes de medidas de HRTFs, da literatura. A seguir, é feita uma breve comparação entre eles em termos da distribuição de energia contida nas HRTFs em função da posição espacial. É realizado um estudo em relação à capacidade de representação do conjunto inteiro de HRTFs por um conjunto menor, através de uma decomposição linear, e então, são realizados testes subjetivos de percepção da posição estática do som virtual para comparação dos conjuntos. Por fim, são apresentadas as conclusões. A Figura 2.1 ilustra o estudo a ser apresentado.

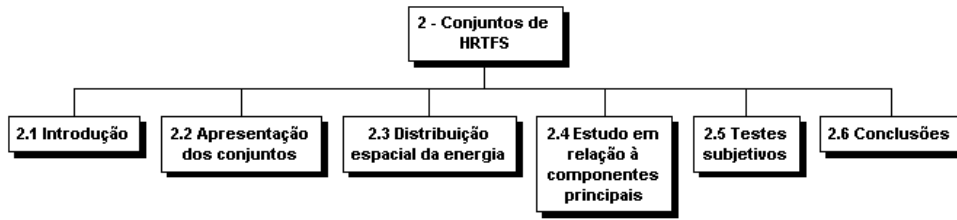


Figura 2.1: Conteúdo do capítulo sobre HRTFs.

## Características comuns aos conjuntos

A fim de padronizar a representação dos conjuntos de medidas, eles foram adaptados de forma a seguir o sistema de coordenadas polares interaural apresentado na Seção 1.2. A faixa de variação de azimutes adotada é de  $-180^\circ$  a  $180^\circ$ , e a faixa de elevações adotada é de  $-90^\circ$  a  $90^\circ$ .

Todas as medições descritas foram realizadas em câmaras anecóicas, com uma frequência de amostragem de 44.100 Hz. Os conjuntos estão compensados dos efeitos dos microfones e outros equipamentos do sistema de medição.

## 2.2 Apresentação dos Conjuntos

### 2.2.1 Conjunto de Gardner e Martin

As medições do conjunto foram feitas no MIT *Media Lab Perceptual Computing*, num manequim da KEMAR (*Knowles Electronic Manikin for Acoustic Research*) [14].

O azimute foi de  $-180$  a  $180^\circ$ , e a elevação cobriu a faixa de  $-40^\circ$  a  $90^\circ$ , com um número variável de medições de azimute para cada elevação, levando a 710 posições diferentes de medição [14]. As posições medidas do conjunto de HRIRs de Gardner tiveram a configuração mostrada na Figura 2.2.

O raio entre os alto-falantes e o centro da cabeça do manequim de referência foi de 1,40 m.

Cada HRIR do conjunto de Gardner, após a compensação dos efeitos do alto-falante e do sistema de medição, apresentou 128 coeficientes.

A Figura 2.3 ilustra o comportamento de uma HRIR do conjunto de Gardner.

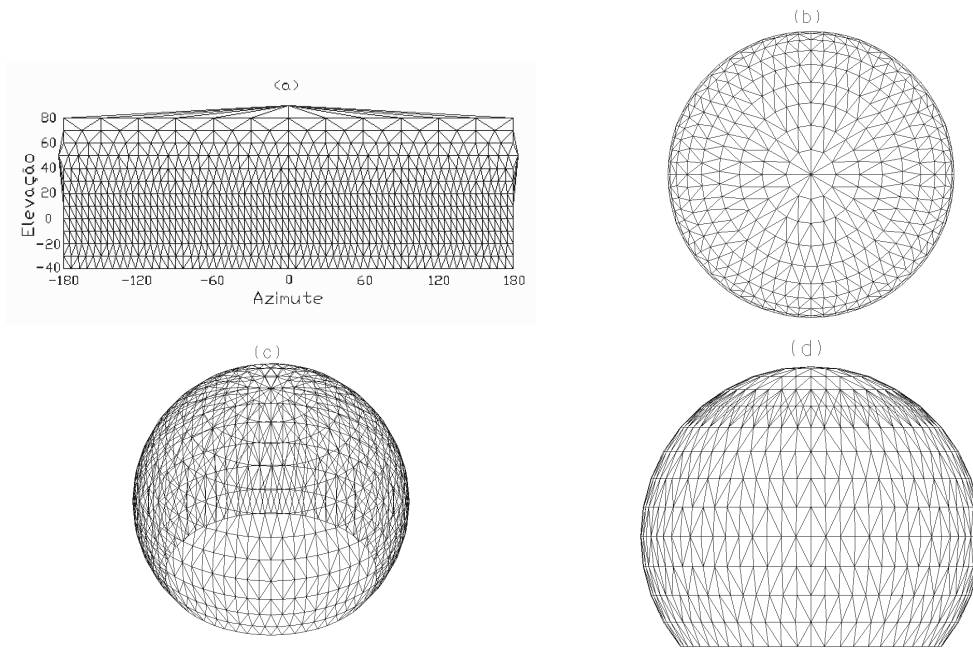


Figura 2.2: Conjunto de Gardner. (a) Visão planificada (b) Visão esférica - topo, (c) Visão esférica - isométrica, (d) Visão esférica - frontal.

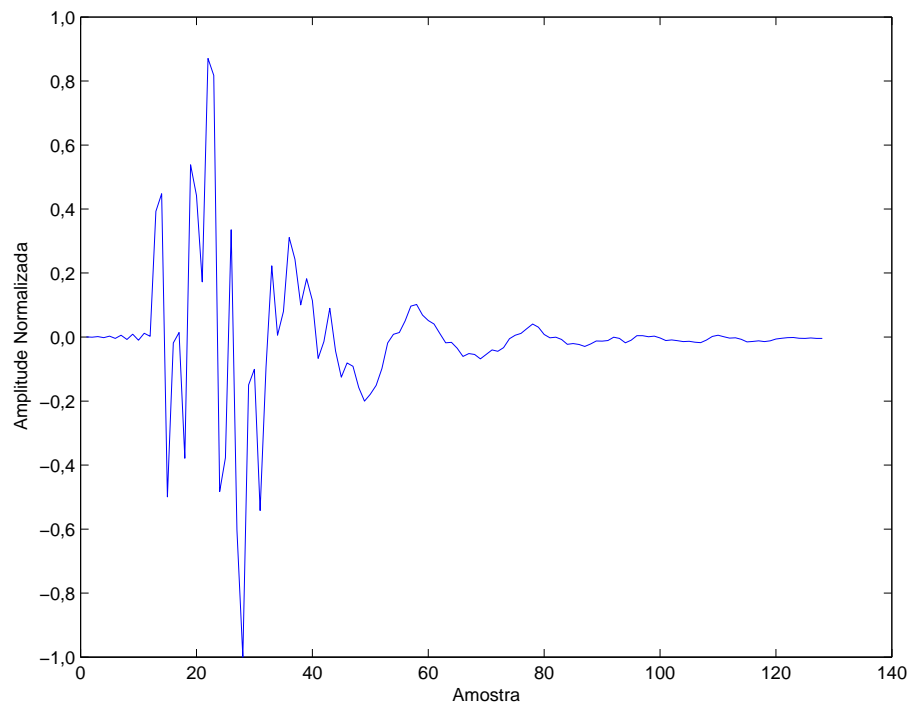


Figura 2.3: HRIR do conjunto de Gardner no ponto  $(0, 0)$ .

A Figura 2.4 representa o comportamento da HRIR anterior no domínio da frequência.

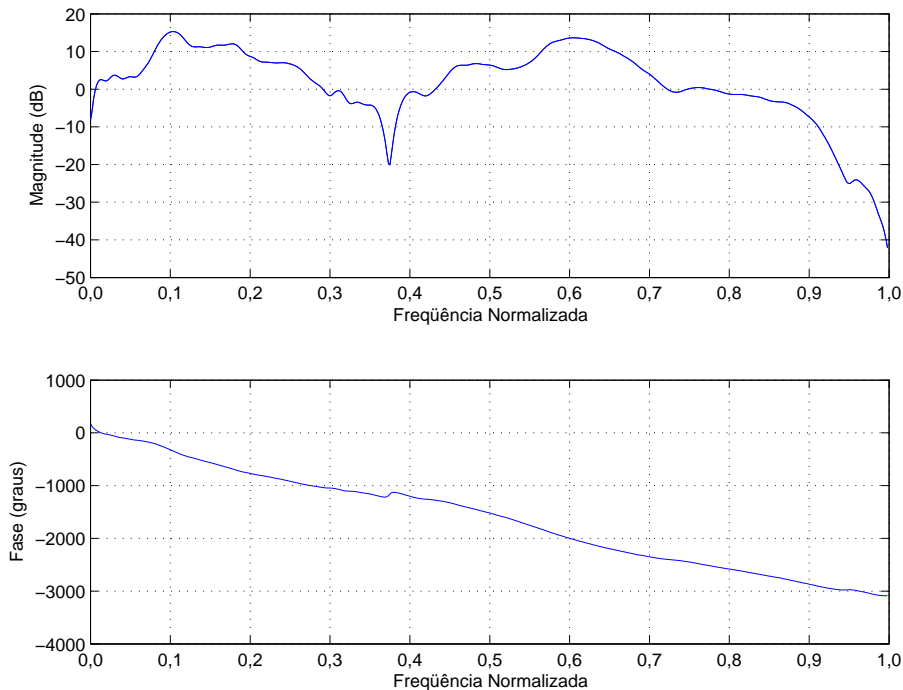


Figura 2.4: HRTF do conjunto de Gardner no ponto  $(0, 0)$ .

## 2.2.2 Conjunto do Projeto Listen

O conjunto de HRIRs apresentado nesta seção foi obtido no contexto do Projeto Listen, que faz parte do *European Commission's Fifth Framework Program* [13]. Foram realizadas medições em 51 pessoas. Para as análises a seguir, escolheu-se aleatoriamente um conjunto de medidas de uma das pessoas submetidas ao ensaio.

A configuração das medições do conjunto do Projeto Listen está representada na Figura 2.5. Nota-se que a grade de medidas é menos densa do que o do conjunto de Gardner.

As medições foram realizadas numa faixa de azimutes de  $-180^\circ$  a  $180^\circ$ , e faixa de elevações de  $-45^\circ$  a  $90^\circ$  graus, com um número variável de medições de azimute para cada elevação, levando a 187 posições diferentes de medição.

O raio correspondente à distância entre o alto-falante e o ouvinte foi de 1,95 m, e cada HRIR, após as compensações do sistema de medidas, ficou com 512 coeficientes.

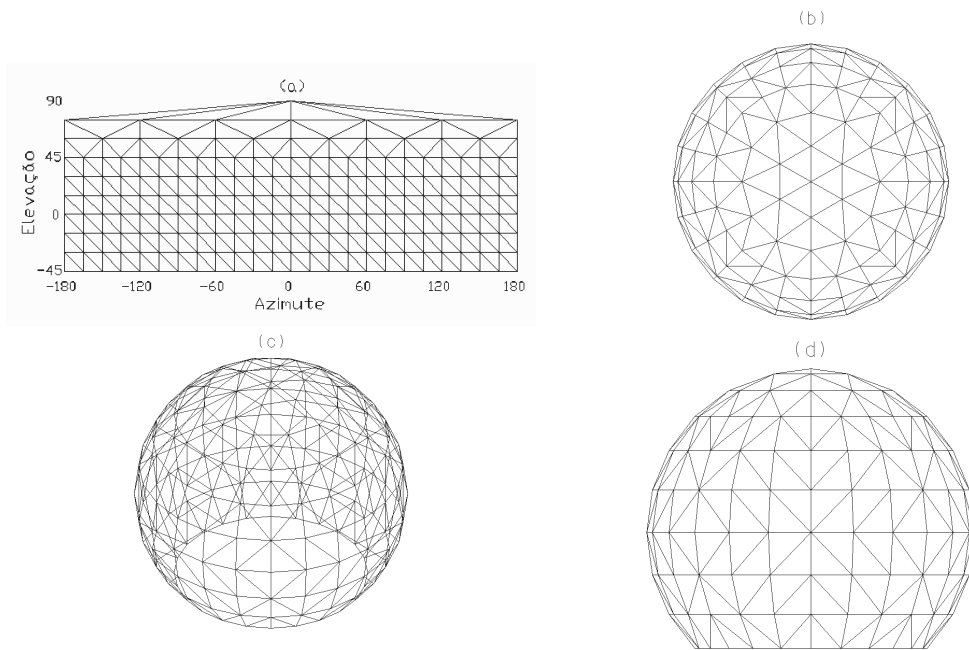


Figura 2.5: Conjunto do Projeto Listen. (a) Visão planificada, (b) Visão esférica - topo, (c) Visão esférica - isométrica, (d) Visão esférica - frontal.

A Figura 2.6 ilustra o comportamento de uma HRIR do conjunto do Projeto Listen. Nota-se que, mesmo a muitas amostras do início, o sinal ainda apresenta amplitude não-nula.

A Figura 2.7 representa o comportamento da HRIR da Figura 2.6 no domínio da frequência. Outra característica da HRTF da Figura 2.7 é apresentar maiores ondulações do que a HRTF da Figura 2.4. A aplicação de um método de suavização espectral como o apresentado em [19] poderia ser uma alternativa para atenuar tais ondulações.

### 2.2.3 Conjunto da CIPIC

O conjunto de HRIRs da CIPIC [12], do Laboratório de Tecnologia da Califórnia, consiste em medições em 45 indivíduos, incluindo o manequim da KEMAR. Para as análises a seguir, escolheu-se aleatoriamente um conjunto de medidas de uma das pessoas submetidas ao ensaio.

As medidas foram numa malha retangular de 25 azimutes e 50 elevações, resultando em 1.250 posições de medida. Elas foram feitas num arco de  $-80^\circ$  a



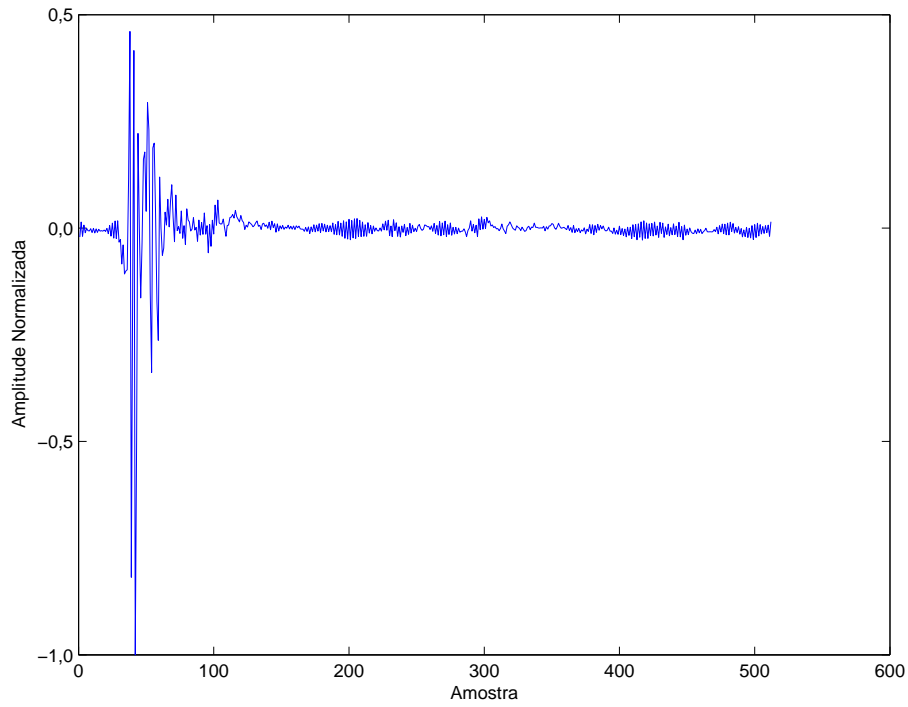


Figura 2.6: HRIR do conjunto do Projeto Listen no ponto (0, 0).

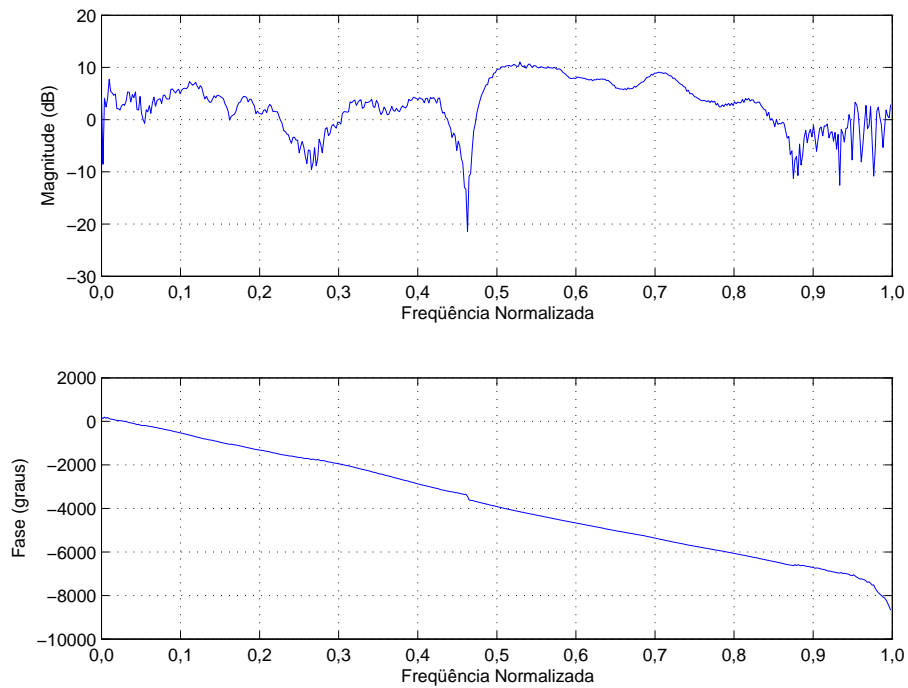


Figura 2.7: HRTF do conjunto do Projeto Listen no ponto (0, 0).

80° de azimute, e de  $-45^\circ$  a  $+230,625^\circ$  de elevação [12]. Devido à simetria, e a fim de normalizar a notação, pode-se considerar que as medições foram efetuadas

variando-se o azimute de  $-180^\circ$  a  $180^\circ$ , e a elevação variando de  $-45^\circ$  a  $90^\circ$  na parte da frente e de  $-50,625^\circ$  a  $90^\circ$  na parte de trás, conforme a Figura 2.8.

O raio da medição foi de 1,00 m.

Cada HRIR apresentou 200 coeficientes, após as compensações do sistema de medidas.

Nota-se que o conjunto da CIPIC é o que tem grade de medidas mais densa, em relação aos conjuntos citados anteriormente.

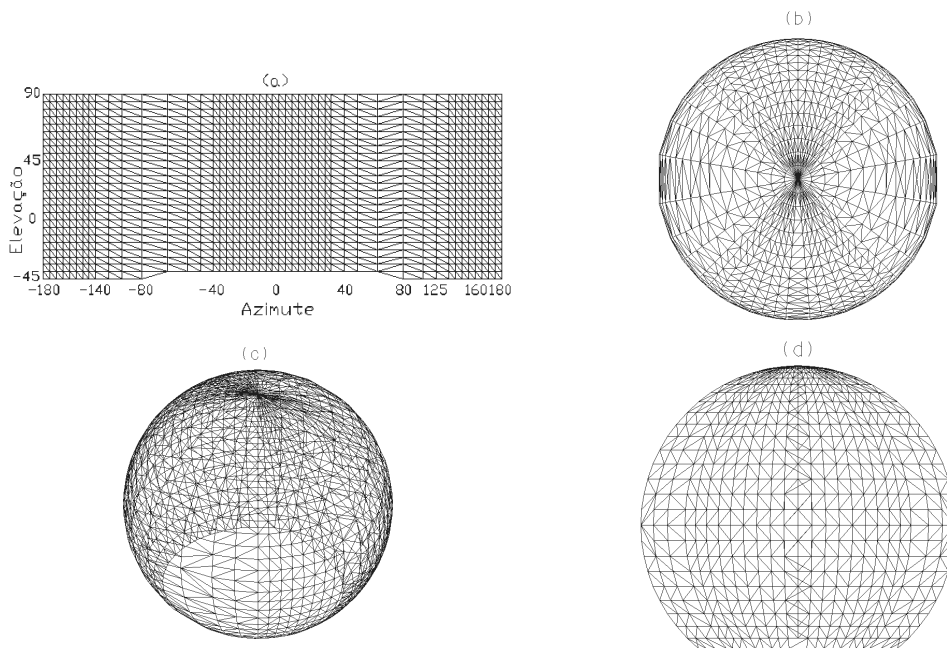


Figura 2.8: Conjunto da CIPIC. (a) Visão planificada (b) Visão esférica - topo, (c) Visão esférica - isométrica, (d) Visão esférica - frontal.

A Figura 2.9 ilustra o comportamento de uma HRIR do conjunto da CIPIC, enquanto que a Figura 2.10 representa a HRTF correspondente. Nota-se que esta HRTF também apresenta respostas sem grandes ondulações, como o conjunto de Gardner.

## 2.2.4 Resumo das Características dos Conjuntos

A Tabela 2.1 resume as principais características de cada um dos conjuntos de HRIRs apresentados.

A Figura 2.11 sobrepõe o comportamento dos conjuntos na frequência, para as

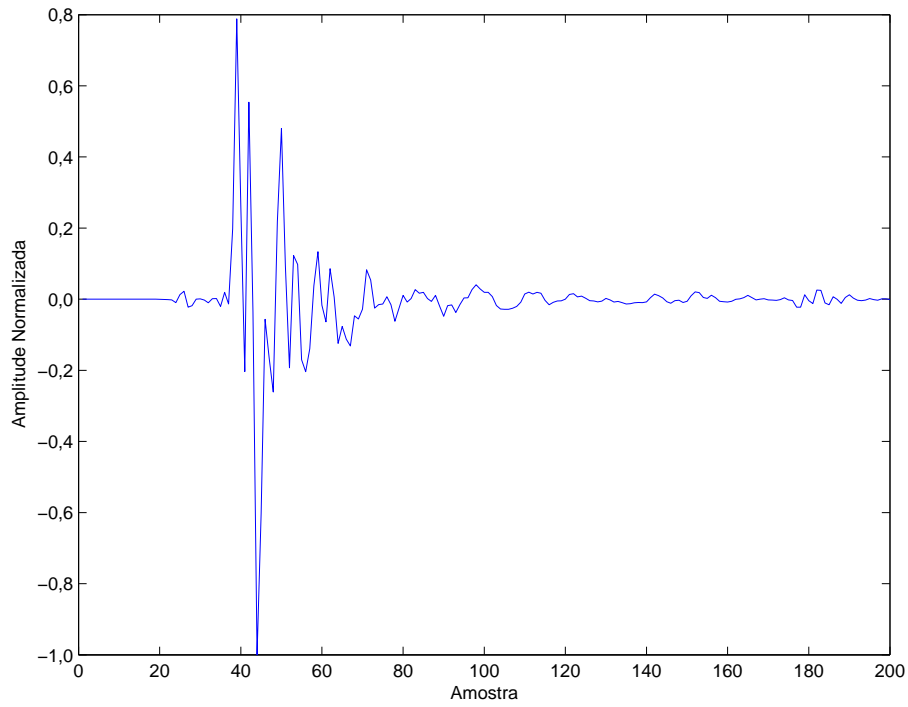


Figura 2.9: HRIR do conjunto da CIPIC no ponto (0,0).

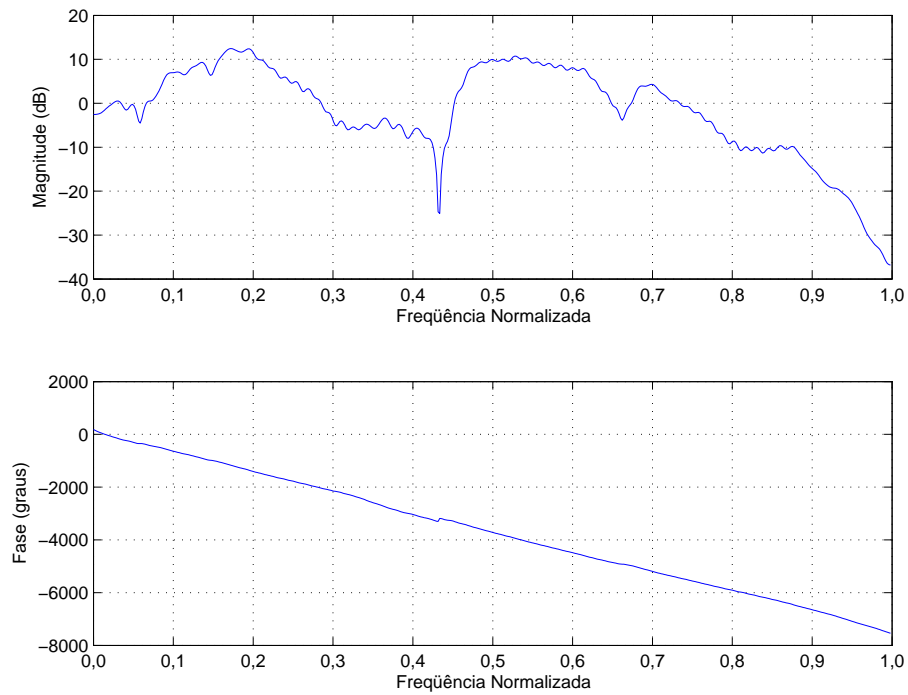


Figura 2.10: HRTF do conjunto da CIPIC no ponto (0,0).

coordenadas (0,0). Nota-se uma certa semelhança nas respostas dos três conjuntos, no que diz respeito à presença do vale em torno de 0,4 e dos picos em torno de 0,1

Tabela 2.1: Características dos conjuntos de HRIRs.

Conjunto de HRIRs	Número de coeficientes	Número de posições	Faixa de azimutes	Faixa de elevações	Raio da esfera
Gardner	128	710	-180° a 180°	-40° a 90°	1,40 m
Projeto Listen	512	187	-180° a 180°	-45° a 90°	1,95 m
CIPIC	200	1.250	-180° a 180°	-45° a 90° (Frente) -50,625° a 90° (Trás)	1,00 m

e em torno de 0,6, valores em termos de frequência normalizada.

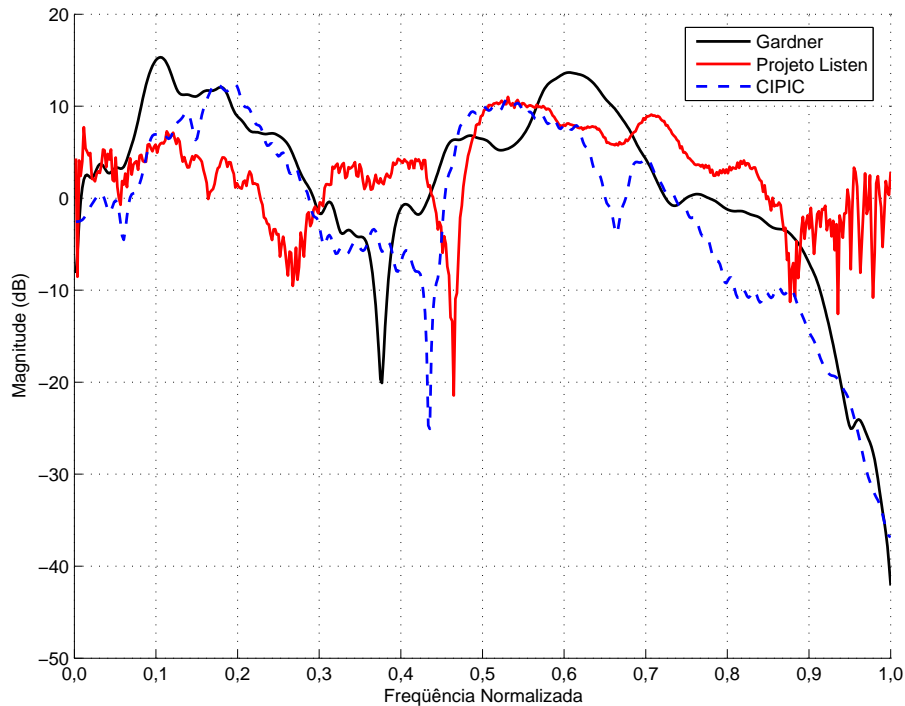


Figura 2.11: Comportamento comparativo dos conjuntos no ponto (0, 0).

## 2.3 Estudo em Relação à Distribuição de Energia

Nesta seção é feita uma análise da distribuição de energia em função da posição da fonte sonora, para os conjuntos de HRTFs apresentados na Subseção 2.2.

Intuitivamente, espera-se que a energia seja maior quanto mais próximo o alto-falante esteja da orelha. Supondo que tomemos a orelha esquerda como referência, o máximo de energia é esperado próximo à posição  $(-90^\circ, 0)$ , enquanto que o mínimo de energia é esperado próximo à posição diametralmente oposta,  $(90^\circ, 0)$ .

Outro fator a se considerar é o efeito de difração em função da frequência. É

sabido que as baixas frequências sonoras conseguem contornar obstáculos com maior facilidade, devido a tal efeito. O mesmo não ocorre com as altas frequências.

### 2.3.1 Comportamento dos Conjuntos

O procedimento adotado foi calcular a energia contida na HRIR, para cada posição da fonte sonora. A energia foi calculada como sendo o somatório das amostras do sinal ao quadrado. A seguir, fez-se a normalização pelo valor máximo da energia de cada conjunto. Os resultados estão descritos em dB. A Figura 2.12 mostra o comportamento geral da distribuição de energia em função da posição, com os azimutes e elevações na representação planificada, para cada conjunto de medidas. Todos os cálculos foram feitos tomando como referência a orelha esquerda do indivíduo de teste. Não foi feita interpolação das posições de medida, de modo que cada conjunto tem uma grade diferente, que segue o que foi exposto na Tabela 2.1.

A Figura 2.13 mostra a Figura 2.12 na forma de curvas de nível. Tanto as figuras na forma de visão isométrica quanto na forma de curvas de nível demonstram comportamento semelhante para os três conjuntos, no sentido de que a energia é maior para posição  $(-90^\circ, 0)$  e decresce ao se afastar desta, tendendo a ser o menor no ponto diametralmente oposto, a menos do efeito do *bright spot*, descrito a seguir.

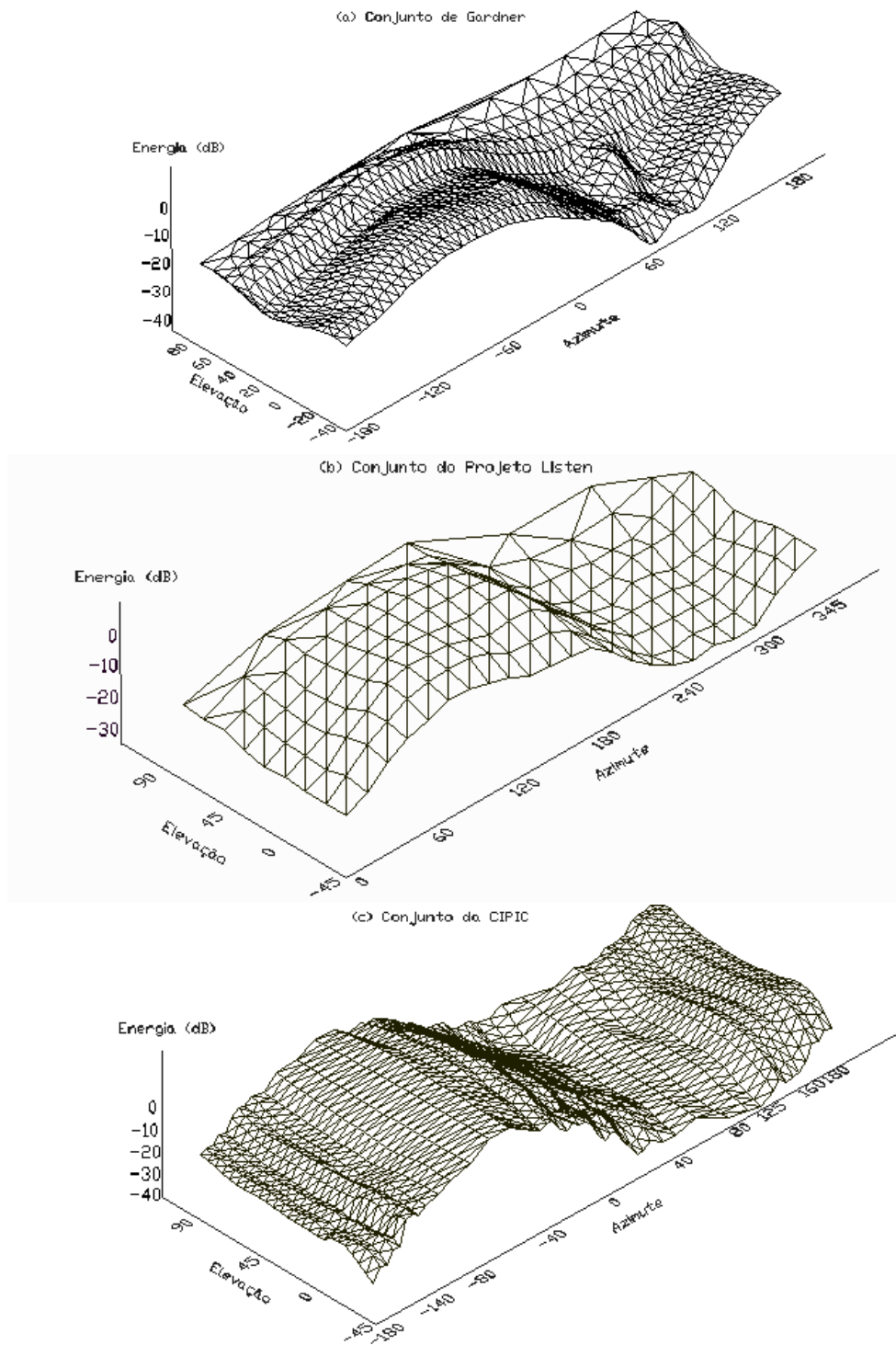


Figura 2.12: Visão 3D isométrica da variação da energia para os conjuntos de medida.

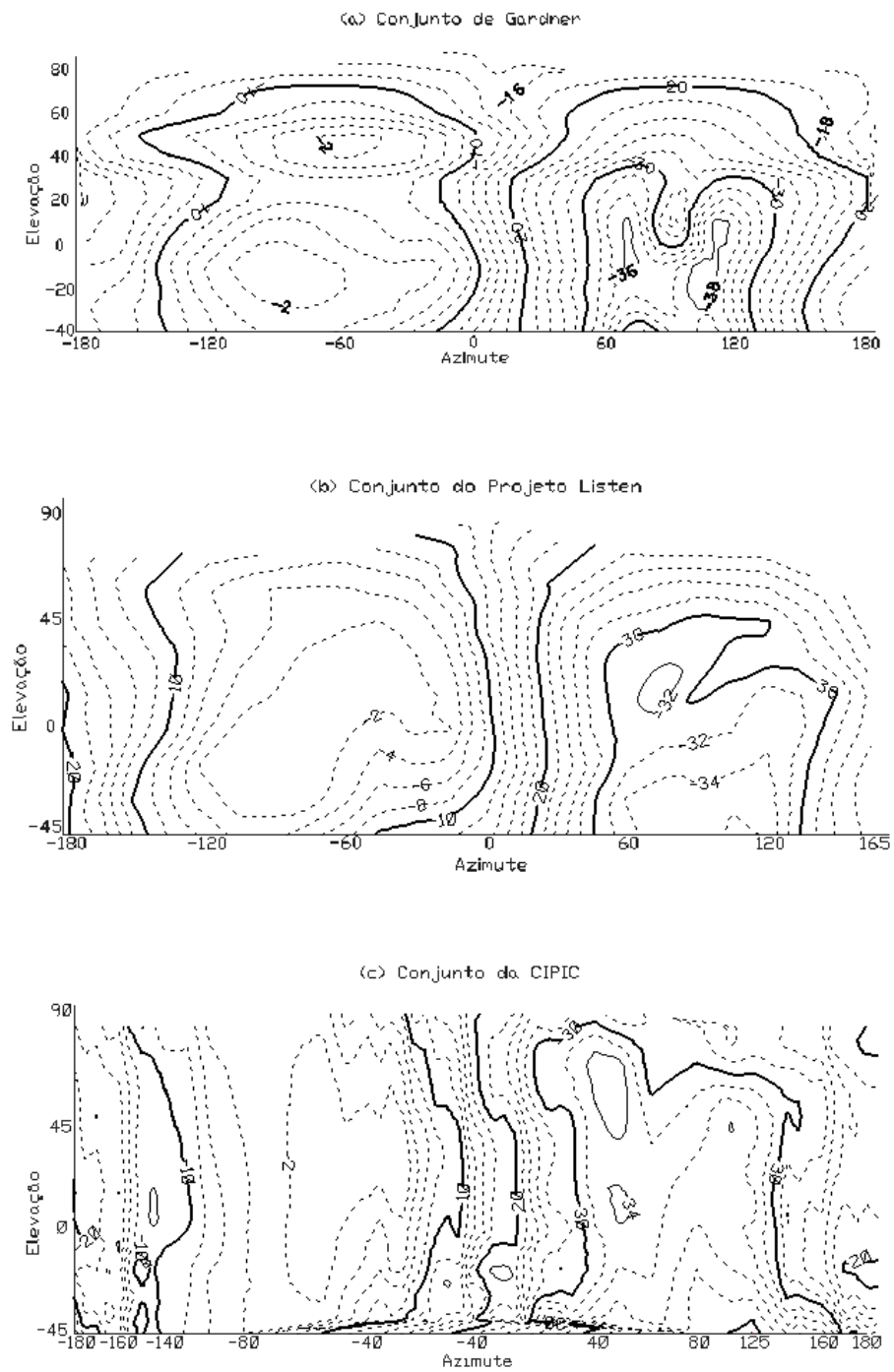


Figura 2.13: Representação em curvas de nível da energia dos conjuntos de medida.

### 2.3.2 *Bright Spot*

De acordo com o comportamento da variação de energia apresentado na Figura 2.12, nota-se que o máximo de energia ocorreu no entorno das coordenadas  $(-90^\circ, 0)$ , conforme esperado, por ser a posição mais próxima entre o alto-falante e a orelha esquerda. Conforme afasta-se a fonte sonora de  $(-90^\circ, 0)$ , o nível de energia diminui.

Entretanto, o mínimo de energia não ocorreu no ponto diametralmente oposto  $(90^\circ, 0)$ . Ao contrário, há um aumento de energia nos arredores de  $(90^\circ, 0)$ , sendo tal ponto um máximo local. Tal máximo local também pode ser notado nas curvas de nível do conjunto de Gardner, na Figura 2.13, e de forma menos evidente para os demais conjuntos. Um possível motivo para tal comportamento é o efeito da difração nas baixas frequências. Para uma fonte sonora posicionada nas coordenadas  $(90^\circ, 0)$ , a orelha esquerda, no ponto  $(-90^\circ, 0)$ , tende a captar as ondas sonoras difratadas igualmente em todos os sentidos: por cima da cabeça, por baixo, pelos lados, etc. Já para a fonte numa posição diferente de  $(90^\circ, 0)$ , há um caminho em que as ondas difratadas chegarão primeiro, e para todos os outros caminhos o percurso a ser percorrido será maior. Portanto, a maior energia devido ao efeito da difração ocorre em  $(90^\circ, 0)$ . O fenômeno descrito é denominado *bright spot* [6, 20].

A Figura 2.14 ilustra o comportamento da energia ao longo de um caminho de elevação fixa e azimute variando de  $-180^\circ$  a  $180^\circ$ .

O efeito de *bright spot* pode ser claramente notado na curva de elevação zero do conjunto de Gardner, ocorrendo de forma mais tímida na curva de elevação  $15^\circ$  do conjunto do Projeto Listen e na curva de elevação  $22,5^\circ$  do conjunto da CIPIC.

Também da Figura 2.14, nota-se que a faixa de variação de energias é maior próximo à elevação zero, e de modo geral, conforme a elevação aumenta, a faixa de variação diminui.

As curvas de energia ao longo de caminhos para o conjunto da CIPIC apresentam um comportamento não suave, com “bicos”, conforme a Figura 2.14. Os demais conjuntos apresentam curvas suaves, conforme era de se esperar.



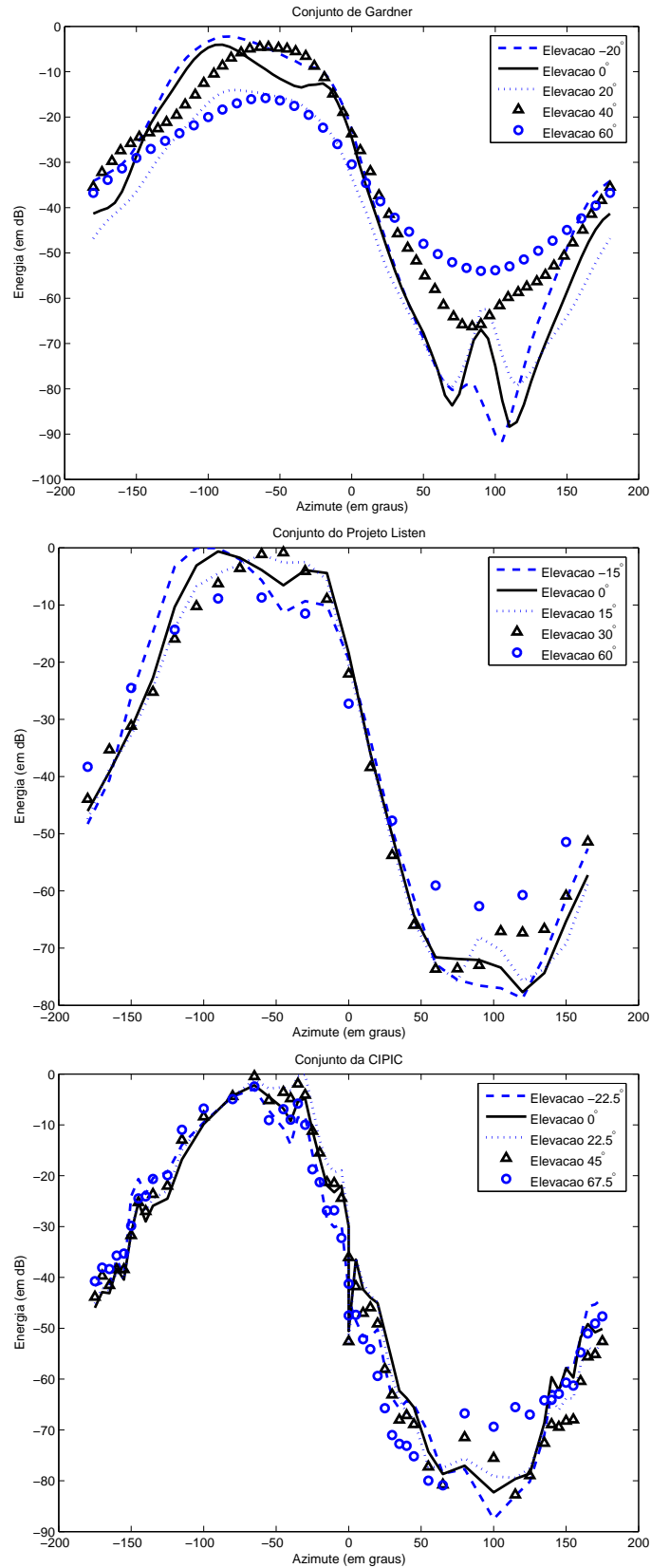


Figura 2.14: Representação da energia percorrendo azimutes  $-180^\circ$  a  $180^\circ$  para os conjuntos analisados.

## 2.4 Estudo em Relação às Componentes Principais

Uma das formas de se fazer a panoramização para fones de ouvido utilizando um conjunto de HRTFs é através da decomposição linear do conjunto num número finito de funções espaciais e filtros básicos [21]. O principal objetivo de tal decomposição é representar o conjunto de HRTFs com menor quantidade de dados do que o conjunto original, perdendo uma quantidade controlada de suas características.

A decomposição e redução de ordem de um conjunto de HRTFs resulta em vantagens como [21]:

- Custo computacional reduzido no contexto da reprodução simultânea de várias fontes,
- Interpolação espacial em tempo real, e
- Representação compacta dos dados de HRTFs: um número menor de funções espaciais, em vez de centenas de HRTFs.

Nesta seção, os conjuntos de HRTFs serão analisados quanto à capacidade de serem representados por um conjunto de dados menor, por meio da transformada de Karhunen-Loève (KLT).

### 2.4.1 Representação usando a KLT

A notação aqui adotada deriva do apresentado em [19].

A decomposição baseia-se em descrever cada HRTF como um somatório de fatores da forma:

$$H(\theta, \phi, f) = \sum_{i=1}^N w_i(\theta, \phi) \Gamma_i(f), \quad (2.1)$$

onde  $w_i(\theta, \phi)$  é a  $i$ -ésima Função Característica Espacial (SCF - *Spatial Characteristic Function*), dependente da posição espacial;  $\Gamma_i(f)$  é a  $i$ -ésima Função de Base da Transformada, dependente da frequência; e  $N$  é o número de funções  $\Gamma_i(f)$ .

Para o caso de um conjunto de HRIRs, cada HRIR é considerada uma amostra do processo aleatório  $\mathbf{H}$ . Assim, a matriz de covariância  $\mathbf{C}$  pode ser decomposta por

$$\mathbf{C} = (\mathbf{H} - \overline{\mathbf{H}})^T (\mathbf{H} - \overline{\mathbf{H}}) = \mathbf{\Psi} \mathbf{\Lambda} \mathbf{\Psi}^T, \quad (2.2)$$

com

$$\mathbf{\Psi}^T \mathbf{\Psi} = \mathbf{I}, \quad (2.3)$$

onde  $\mathbf{I}$  é a matriz identidade;  $\mathbf{\Psi} = [\psi_j(0) \dots \psi_j(N-1)]^T$ , com  $j = 1, \dots, N$ , é a matriz unitária de autovetores de  $\mathbf{C}$ , a matriz da KLT;  $\mathbf{\Lambda}$  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

$$\bar{h}(k) = \frac{1}{P} \sum_{\theta} \sum_{\phi} h(\theta, \phi, k), \quad (2.4)$$

onde  $P$  é o número de HRIRs.

A partir da formulação descrita, pode-se recompor uma HRIR numa posição  $(\theta, \phi)$  segundo a equação

$$\hat{h}(\theta, \phi, k) = \bar{h}(k) + \sum_{j=1}^N w_j(\theta, \phi) \psi_j(k), \quad (2.5)$$

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

A redução na complexidade computacional do processo consiste em utilizar apenas  $M$  autovetores ( $M < N$ ) para representar todo o conjunto, conforme

$$\hat{h}(\theta, \phi, k) = \bar{h}(k) + \sum_{j=1}^M w_j(\theta, \phi) \psi_j(k). \quad (2.6)$$

Escolhem-se apenas os autovetores relacionados aos maiores autovalores, que são a escolha ótima no sentido de mínimos quadrados [22].

## 2.4.2 Resultados da Formulação KLT

A partir da formulação descrita na Subseção 2.4.1, foi feito um estudo com relação ao número de autovetores necessários para representar cada conjunto de HRIRs descrito na Seção 2.2.

A Figura 2.15 apresenta o comportamento da porcentagem da energia total do sistema pelo número de autovetores utilizados.

A Tabela 2.2 resume os resultados obtidos.

O conjunto que necessita de menos autovetores é o de Gardner, seguido pelo conjunto da CIPIC e do Projeto Listen. Os conjuntos de Gardner e da CIPIC

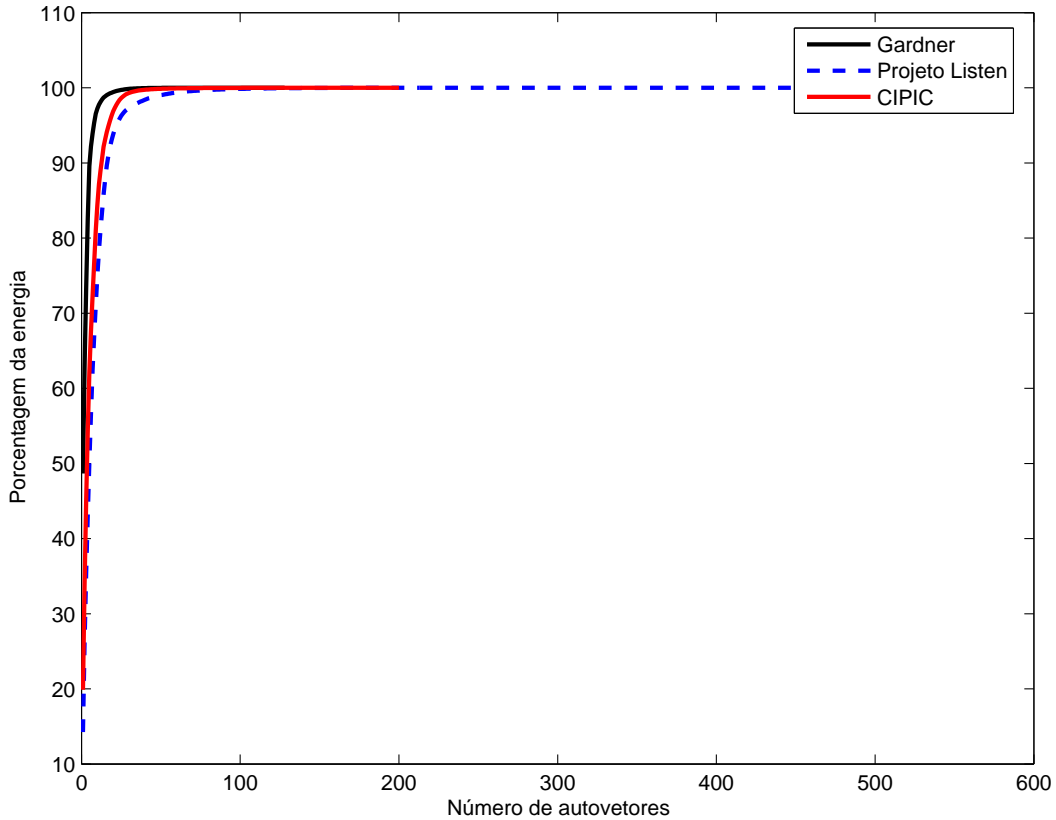


Figura 2.15: Distribuição de energia por número de autovetores para cada conjunto.

Tabela 2.2: Número de autovetores para representar 99,9% da energia total.

Conjunto	Número de autovetores	Número de HRIRs	Porcentagem
Gardner	33	726	4,55%
Projeto Listen	111	187	59,33%
CIPIC	56	1.250	4,48%

tiveram uma redução percentual do número de autovetores equivalente em relação ao número original de HRIRs, conforme indica a coluna “Porcentagem” da Tabela 2.2.

O conjunto de HRIRs restante, do Projeto Listen, apresentou um comportamento destoante dos demais conjuntos analisados. Tal diferença pode ser atribuída ao número de posições de medida (187 posições), relativamente pequeno para representar o grande número de coeficientes das HRIRs (512 coeficientes). Apenas para se ter um parâmetro de comparação, o conjunto de Gardner contempla 710 posições e HRIR com 128 coeficientes em cada posição, como se vê na Tabela 2.1 da Subseção 2.2.4. Portanto, tal conjunto tem mais posições de medida para representar menos informação, o que resulta em maior capacidade de “compressão”.

A Tabela 2.3 indica o número de autovetores necessários para representar uma porcentagem crescente de energia, para cada conjunto analisado.

Tabela 2.3: Número de autovetores por porcentagem de energia.

<b>Conjunto</b>	<b>99,9%</b>	<b>99,99%</b>	<b>99,999%</b>	<b>100%</b>
Gardner	33	53	70	726
Projeto Listen	111	164	183	187
CIPIC	56	105	146	1250

Dentre os conjuntos de medidas, os que apresentaram melhor capacidade de compressão foram o de Gardner e o da CIPIC. Os resultados demonstram que, controlando as perdas dentro do limite estabelecido, pode-se reduzir o conjunto de dados para representar as HRIRs originais. A partir desse conjunto reduzido de dados, pode-se imaginar uma forma de interpolação de HRIRs, à semelhança do estudo realizado em [19].

## 2.5 Testes Subjetivos

Nas seções anteriores, três conjuntos de HRTFs foram descritos e analisados quanto a diversos fatores. A fim de se ter um estudo mais aprofundado do comportamento dos conjuntos e de se confirmar efetivamente a validade dos mesmos, foram realizados testes subjetivos de percepção da posição estática do som virtual.

Os conjuntos de HRTFs utilizados nos testes subjetivos foram os mesmos apresentados anteriormente. Na medida que o Projeto Listen consiste em coleções de conjuntos de medidas de diversos indivíduos, foi escolhido aleatoriamente um conjunto de um dos indivíduos submetidos ao ensaio. O mesmo se aplica ao conjunto da CIPIC, conforme descrito na Subseção 2.2.

Os testes foram realizados no Laboratório de Processamento de Sinais da Universidade Federal do Rio de Janeiro, em 30 pessoas de idades variadas e sem problemas de audição diagnosticados. Os testes foram realizados em grupos de 1 a 6 pessoas. Era possível ao usuário solicitar que a seqüência de teste fosse repetida após a sua reprodução e que, apenas no início do teste, o volume do som fosse alterado.

## 2.5.1 Descrição dos Testes

Foi utilizado ruído rosa como sinal de teste, pela sua característica de manter a energia igualmente distribuída por banda perceptiva. Utilizaram-se fones de ouvido circum-aurais fechados (*closed*) modelo HD265 da Sennheiser<sup>®</sup>, e um amplificador de 8 canais para fones de ouvido modelo Powerplay Pro-8 HA8000, da Behringer<sup>®</sup>. Não foi realizada medição do ruído de fundo no interior da sala, mas com a utilização dos fones de ouvido do tipo fechado, o efeito do ruído ambiente pôde ser desconsiderado. O sinal de teste tinha 4 segundos de duração.

O objetivo dos testes era verificar o comportamento da percepção da posição estática de uma fonte sonora virtual, utilizando os três conjuntos de HRTFs. Para tal, escolheram-se 25 posições de medida no espaço 3D, distribuídas conforme ilustra a Figura 2.16. Na seqüência do texto será utilizada uma identificação mnemônica das posições de teste, como se vê na figura: a primeira letra indica o conjunto Frente, Meio ou Trás; a segunda letra refere-se à elevação, e indica Alto, Horizontal ou Baixo; a terceira letra refere-se ao azimute, e indica Esquerda, Centro ou Direita.

Efetuuou-se o posicionamento do sinal de teste original (ruído rosa) nas 25 posições virtuais da Figura 2.16, considerando os 3 conjuntos de medida, resultando em 75 sinais ao todo. Não foi feita interpolação das HRIRs, utilizando-se as informações originais dos conjuntos. Sendo assim, as posições de medida escolhidas foram as mais próximas possíveis das descritas na Figura 2.16, conforme a Tabela 2.4.

Devido à simetria da cabeça humana, não se apresentou, num mesmo teste, o sinal numa posição do lado direito e a sua posição espelhada no lado esquerdo. Considerou-se possível espelhar o resultado do lado direito para o esquerdo e somar com o resultado do lado esquerdo, na análise dos testes. Portanto, das 25 posições da Figura 2.16 testaram-se apenas 16 em cada teste, para cada um dos 3 conjuntos, resultando em 48 posições testadas de cada vez.

Foi apresentado a cada um dos voluntários um formulário composto de 48 imagens semelhantes à da Figura 2.16. Foi solicitado que o ouvinte assinalasse, em cada imagem, o cubo em que se encontrava a fonte virtual.

Os sinais de teste foram apresentados a cada grupo de pessoas numa ordem aleatória de posição de medida e de conjunto de HRTFs, respeitando o critério de simetria apresentado acima. A seqüência de apresentação dos sinais foi gravada a

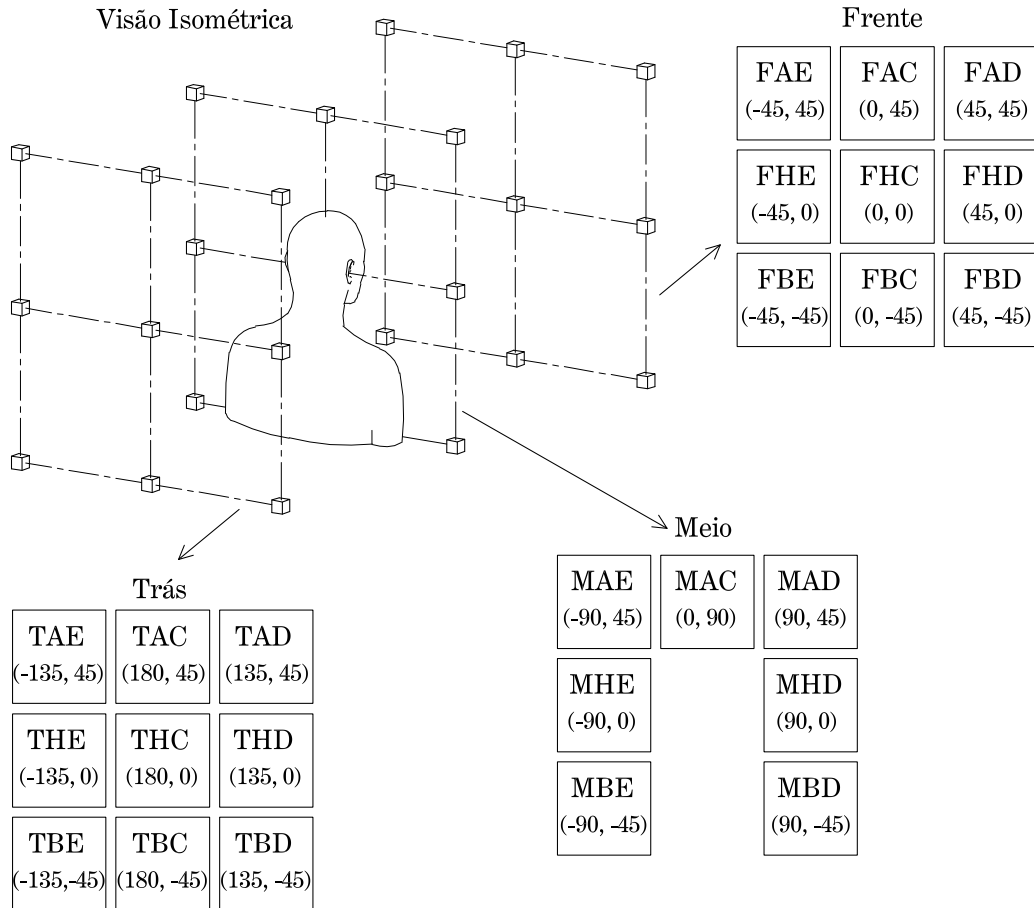


Figura 2.16: Posições Espaciais e Convenção Adotada. Coordenadas  $(\theta, \phi)$  em graus. Para o azimute, adotaram-se E = esquerda, C = centro, D = direita; para a elevação, adotaram-se A = alto, H = horizontal, B = baixo.

cada execução e posteriormente comparada com as respostas dos ouvintes.

## 2.5.2 Resultados e Distribuição Espacial das Respostas

As Tabelas 2.5, 2.6 e 2.7 apresentam os resultados detalhados dos testes, para os conjuntos de medida de Gardner, Projeto Listen e CIPIC, respectivamente. A linha “Nulo” das tabelas se refere a formulários que não foram marcados ou tiveram mais de uma alternativa assinalada.

Nas tabelas apresentadas, as células vazias correspondem a valores que não ocorreram nas respostas dos testes.

As posições MHC (centro da cabeça) e MBC (abaixo da cabeça) foram mantidas na coluna Posição de Teste para que a diagonal principal das tabelas corres-

Tabela 2.4: Coordenadas relativas a cada posição. Valores  $(\theta, \phi)$  em graus.

Posição	Referência	Gardner	Projeto Listen	CIPIC
<b>FAE</b>	(-45, 45)	(-48, 50)	(-45, 45)	(-45, 45)
<b>FAC</b>	(0, 45)	(0, 50)	(0, 45)	(0, 45)
<b>FAD</b>	(45, 45)	(48, 50)	(45, 45)	(45, 45)
<b>FHE</b>	(-45, 0)	(-50, 0)	(-45, 0)	(-45, 0)
<b>FHC</b>	(0, 0)	(0, 0)	(0, 0)	(0, 0)
<b>FHD</b>	(45, 0)	(50, 0)	(45, 0)	(45, 0)
<b>FBE</b>	(-45, -45)	(-45, -40)	(-45, -45)	(-45, -45)
<b>FBC</b>	(0, -45)	(0, -40)	(0, -45)	(0, -45)
<b>FBD</b>	(45, -45)	(45, -40)	(45, -45)	(45, -45)
<b>MAE</b>	(-90, 45)	(-88, 50)	(-90, 45)	(-80, 45)
<b>MAC</b>	(0, 90)	(0, 90)	(0, 90)	(0, 90)
<b>MAD</b>	(90, 45)	(88, 50)	(90, 45)	(80, 45)
<b>MHE</b>	(-90, 0)	(-90, 0)	(-90, 0)	(-80, 0)
<b>MHC</b>	(90, 0)	(90, 0)	(90, 0)	(80, 0)
<b>MHD</b>	(-90, -45)	(-90, -40)	(-90, -45)	(-80, -45)
<b>MBE</b>	(90, -45)	(90, -40)	(90, -45)	(80, -45)
<b>MBC</b>	(-135, 45)	(-136, 50)	(-135, 45)	(-135, 45)
<b>MBD</b>	(180, 45)	(176, 50)	(180, 45)	(175, 45)
<b>TAE</b>	(135, 45)	(136, 50)	(135, 45)	(135, 45)
<b>TAC</b>	(-135, 0)	(-135, 0)	(-135, 0)	(-135, 0)
<b>TAD</b>	(180, 0)	(180, 0)	(180, 0)	(175, 0)
<b>THE</b>	(135, 0)	(135, 0)	(135, 0)	(135, 0)
<b>THC</b>	(-135, -45)	(-135, -40)	(-135, -45)	(-135, -45)
<b>THD</b>	(180, -45)	(180, -40)	(180, -45)	(175, -45)
<b>TBE</b>	(135, -45)	(135, -40)	(135, -45)	(135, -45)

pondesse à mesma posição testada e percebida, e também porque alguns ouvintes assinalaram a posição MHC. A diagonal foi representada em negrito, de modo a facilitar a leitura. As colunas MHC e MBC apresentam 0% porque não fizeram parte das posições testadas.

As posições FHD e TAD da Tabela 2.6 não foram testadas, pois obedecendo o critério de apresentar somente o lado esquerdo ou o direito, escolhidos de modo aleatório, ocorreu de somente as respectivas posições do lado esquerdo FHE e TAE terem sido escolhidas.

### Análise da tabela do conjunto de Gardner

Uma análise breve dos resultados da Tabela 2.5 mostra que houve poucos acertos nas posições testadas da Frente, uma quantidade maior de acertos nas posições do Meio, e grande quantidade de acertos nas posições de Trás. Mais ainda,



ocorreu grande quantidade de reversão de frente para trás, ao passo que o oposto ocorreu com menor intensidade.

Houve uma quantidade grande de acertos nas posições MHE ( $-90^\circ, 0$ ) e MHD ( $90^\circ, 0$ ), em relação às demais posições. Destaca-se também que as posições THE, THC e THD, que se localizam na parte de Trás, no plano horizontal, também tiveram grande quantidade de acertos.

Houve baixa quantidade de acertos nas posições abaixo do plano horizontal, tanto na parte da Frente quanto no Meio e Trás. Para tais posições de teste, os ouvintes tenderam a assinalar posições no plano horizontal. Exemplificando, para as posições de teste MBE e MBD, as posições no plano horizontal MHE e MHD foram bastante assinaladas. Para FBE, FBC e FBD, as posições com maior ocorrência por parte dos ouvintes foram THE, THC e THD, indicando reversão frente-trás e tendência de assinalar posições no plano horizontal. Para TBE e TBD, as posições no plano horizontal MHE e MHD tiveram muitas ocorrências.

### **Análise da tabela do conjunto do Projeto Listen**

Quanto aos resultados da Tabela 2.6, nota-se que também houve uma grande tendência de reversão de frente para trás, e que esta tendência foi maior do que no caso do conjunto do Gardner.

Houve uma quantidade grande de acertos na posição MHE ( $-90^\circ, 0$ ). As posições superiores da parte do Meio apresentaram uma quantidade razoável de acertos. Na parte de Trás houve maior acerto nas posições THE e TBE, que se localizam à esquerda do plano mediano.

Também ocorreu baixo índice de acerto nas posições de teste abaixo do plano horizontal. Parece existir a tendência de os ouvintes assinalarem os testes no plano horizontal. Para a posição de teste FBE, a posição mais assinalada foi a THE. Para MBD, a posição mais assinalada foi THC. Para TBC, a posição mais assinalada foi THC.

### **Análise da tabela do conjunto da CIPIC**

Os resultados da Tabela 2.7 também demonstram uma tendência de reversão de frente para trás. Novamente as posições MHE ( $-90^\circ, 0$ ) e MHD ( $90^\circ, 0$ ) estão

entre as que tiveram maior índice de acerto.

Também com relação a Tabela 2.7, houve baixa quantidade de acertos nas posições abaixo do plano horizontal, tanto na parte da Frente quanto no Meio e Trás. Por exemplo, para as posições TBE, TBC e TBD, as posições com maiores acertos foram, respectivamente, THE, THC e THD, indicando que os ouvintes tiveram a impressão de que as posições se encontravam no plano horizontal.

### **Análise das três tabelas**

As tabelas parecem indicar uma tendência maior de reversão de frente para trás do que o contrário. As reversões fazem parte dos principais erros perceptuais relatados na reprodução do som 3D [2, 9], conforme mencionado na Seção 1.4.

As posições MHE ( $-90^\circ, 0$ ) e MHD ( $90^\circ, 0$ ) tiveram índice de acertos alto, em relação as demais posições, nas três tabelas, indicando serem as mais fáceis de serem localizadas.

As tabelas parecem indicar a tendência de a localização ser mais fácil para posições no plano horizontal. A seguir, para posições acima do plano horizontal. As posições abaixo do plano parecem ser as mais difíceis de se localizar, sendo confundidas com posições no plano horizontal.

Tabela 2.5: Tabela de confusão do teste de percepção da posição para o conjunto de Gardner. Valores percentuais.

Posição percebida	Posição de Teste																												
	FAE	FAC	FAD	FHE	FHC	FHD	FBE	FBC	FBD	MAE	MAC	MAD	MHE	MHC	MHD	MBE	MBC	MBD	TAE	TAC	TAD	THE	THC	THD	TBE	TBC	TBD		
FAE					3,3					20,0																			
FAC		<b>3,3</b>	4,2		6,7			3,3			3,3	5,0									10,0						6,7		
FAD			<b>25,0</b>						5,3		25,0				20,8						3,3	14,3			7,7				
FHE		3,3						3,3			3,3													3,3			3,3		
FHC		3,3			<b>23,3</b>			10,0			40,0										6,7			3,3			20,0		
FHD		6,7	16,7			<b>4,8</b>		3,3	10,5		10,0				4,2			7,1			9,5			7,7				22,7	
FBE								<b>3,3</b>	5,3											3,3									
FBC									<b>5,3</b>						4,2					3,3				3,3				9,1	
FBD													16,7							33,3									
MAE	16,7									<b>20,0</b>						6,3													
MAC		33,3			10,0			16,7	5,3		<b>16,7</b>										33,3	4,8					33,3		
MAD		6,7	16,7			23,8			5,3		30,0				20,8			14,3			6,7	23,8		6,7			3,3	9,1	
MHE	33,3	3,3			3,3		36,4	3,3			3,3		<b>50,0</b>			31,3			11,1				29,4	3,3			50,0		
MHC		3,3			3,3																								
MHD			4,2		3,3	33,3			15,8		6,7	5,0			<b>29,2</b>			28,6			6,7	9,5			15,4			27,3	
MBE							9,1																						
MBC																													
MBD			4,2			9,5					5,0							<b>7,1</b>			3,3			7,7				13,6	
TAE	33,3			22,2	3,3					20,0						6,3			<b>33,3</b>	3,3		23,5			12,5				
TAC		10,0			3,3			13,3			13,3										<b>3,3</b>		6,7			13,3			
TAD			8,3			4,8		3,3	21,1		20,0				4,2			14,3			10,0	<b>28,6</b>		23,1			4,5		
THE	16,7			33,3	13,3		45,5			30,0	3,3		33,3			25,0			22,2			<b>35,3</b>	10,0						
THC		23,3			20,0			30,0			10,0										10,0		<b>26,7</b>				13,3		
THD			16,7			23,8		3,3	15,8					16,7				21,4				9,5		<b>88,5</b>				9,1	
TBE							9,1	3,3		10,0						31,3							11,8			<b>12,5</b>			
TBC		3,3						3,3															10,0				<b>3,3</b>		
TBD			4,2						10,5									7,1										4,5	
Nulo																													
Total	100	100	100	100	100	100	100	100	100	100	100	100	100	0,0	100	100	0,0	100	100	100	100	100	100	100	100	100	100	100	100

Tabela 2.6: Tabela de confusão do teste de percepção para o conjunto do Projeto Listen. Valores percentuais.

Posição percebida	Posição de Teste																										
	FAE	FAC	FAD	FHE	FHC	FHD	FBE	FBC	FBD	MAE	MAC	MAD	MHE	MHC	MHD	MBE	MBC	MBD	TAE	TAC	TAD	THE	THC	THD	TBE	TBC	TBD
FAE	5,0				6,7		8,3			6,3			4,2			8,0			10,0			3,8					
FAC		6,7						3,3											3,3	3,3							
FAD											14,3																
FHE																											
FHC				6,7	3,3		8,3			6,3	3,3		12,5			4,0											3,3
FHD		10,0			3,3			13,3			20,0										6,7		3,8	20,0			10,0
FBE												21,4			33,3								3,3	25,0			3,3
FBC																											
FBD																											
MAE	30,0			16,7	3,3					18,8	6,7		8,3			4,0			16,7	3,3		11,5					9,1
MAC		23,3		3,3	6,7						26,7									20,0							3,3
MAD		3,3	30,0		6,7							28,6															
MHE	15,0			30,0	3,3					18,8	3,3		41,7			44,0			10,0	3,3		15,4					5,3
MHC		3,3																									
MHD		3,3	30,0						50,0			21,4			16,7	4,0		20,0									3,3
MBE				6,7			8,3			12,5					4,0				3,3			7,7					3,3
MBC																											
MBD																											
TAE	25,0			23,3	3,3		20,8			25,0		8,3				12,0			23,3			11,5					3,3
TAC		10,0			10,0						13,3									80,0							10,0
TAD		3,3	30,0									7,1			16,7					3,3							10,0
THE	25,0	6,7		13,3			29,2	3,3		12,5	3,3		16,7			16,0			26,7			34,6	3,3			6,7	18,2
THC							4,2	33,3			10,0									20,0			10,0				30,0
THD	16,7				36,7							7,1			33,3			80,0									18,2
TBE	3,3						20,8				6,7		4,2			4,0			6,7	3,3		11,5				3,3	
TBC	6,7				13,3			13,3			6,7								6,7								13,3
TBD			10,0																								6,7
Nulo													4,2														18,2
Total	100	100	100	100	100	100	100	100	100	100	100	100	100	0,0	100	100	0,0	100	100	100	0,0	100	100	100	100	100	100

Tabela 2.7: Tabela de confusão do teste de percepção para o conjunto da CIPIC. Valores percentuais.

Posição percebida	Posição de Teste																											
	FAE	FAC	FAD	FHE	FHC	FHD	FBE	FBC	FBD	MAE	MAC	MAD	MHE	MHC	MHD	MBE	MBC	MBD	TAE	TAC	TAD	THE	THC	THD	TBE	TBC	TBD	
FAE	29,4	6,7		16,7	3,3		16,7			11,1	3,3																	
FAC		10,0			6,7						3,3	4,8							3,3									
FAD			23,1			12,5		16,7				14,3			15,0					8,7								
FHE	5,9			16,7	3,3		8,3	3,3		22,2	3,3																	
FHC					16,7			13,3			6,7								10,0									
FHD		6,7				8,3					4,8				25,0			17,2			13,0		16,7		4,2		4,2	
FBE											3,3																	
FBC					3,3																		3,3					
FBD						4,2					4,8			10,0			3,4										4,2	
MAE	29,4	16,7		16,7	3,3		16,7				20,0								28,6	10,0		16,7						
MAC		13,3			13,3			16,7			23,3									16,7								
MAD			15,4		3,3	25,0									10,0			20,7			4,3							
MHE	5,9	6,7			3,3		4,2			33,3	3,3		30,0						14,3	3,3		33,3	3,3	12,5		3,3	4,2	
MHC																												
MHD			7,7			25,0					23,8				30,0			31,0			26,1				29,2		3,3	37,5
MBE	5,9	3,3					8,3			11,1			10,0															
MBC																												
MBD					4,2						9,5				5,0						4,3						3,3	4,2
TAE	17,6	16,7		50,0			25,0			11,1	3,3								28,6	6,7		16,7						
TAC		6,7			10,0			10,0			16,7									13,3				10,0				6,7
TAD			30,8			4,2					9,5				5,0			6,9			30,4				8,3			
THE	5,9	13,3					16,7	16,7		11,1	3,3		10,0						14,3	3,3		16,7	6,7			33,3	6,7	
THC					26,7						6,7									23,3				30,0			40,0	
THD			23,1			12,5					23,8						13,8				13,0		6,7	20,8			3,3	41,7
TBE					3,3		4,2	3,3			3,3	4,8	30,0		100,0				14,3			16,7						
TBC					3,3			3,3												3,3			13,3				3,3	
TBD						4,2																			25,0			4,2
Nulo																												
Total	100	100	100	100	100	100	100	100	100	100	100	100	100	0,0	100	100	0,0	100	100	100	100	100	100	100	100	100	100	100

As Figuras 2.17 a 2.20 apresentam a localização das respostas no espaço 3D para algumas posições de medida, a fim de melhor se visualizar a distribuição das mesmas em relação à posição testada.

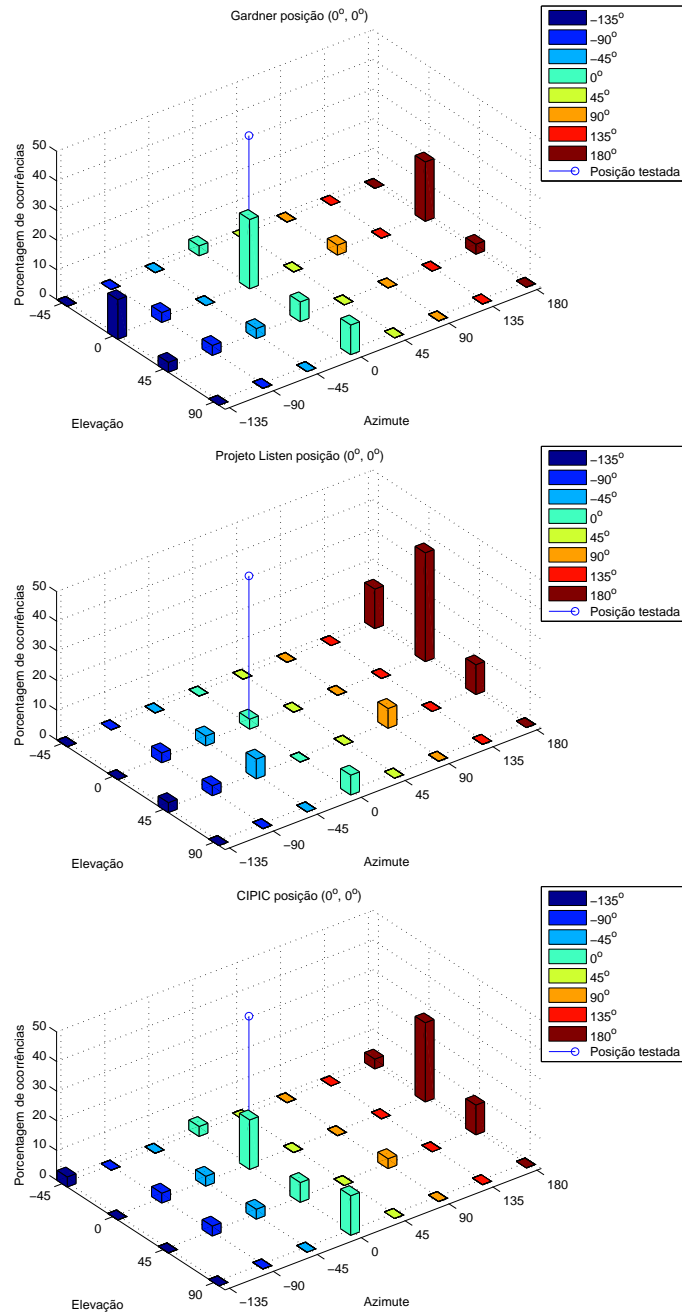


Figura 2.17: Distribuição espacial das respostas, para a posição de teste FHC (0, 0).

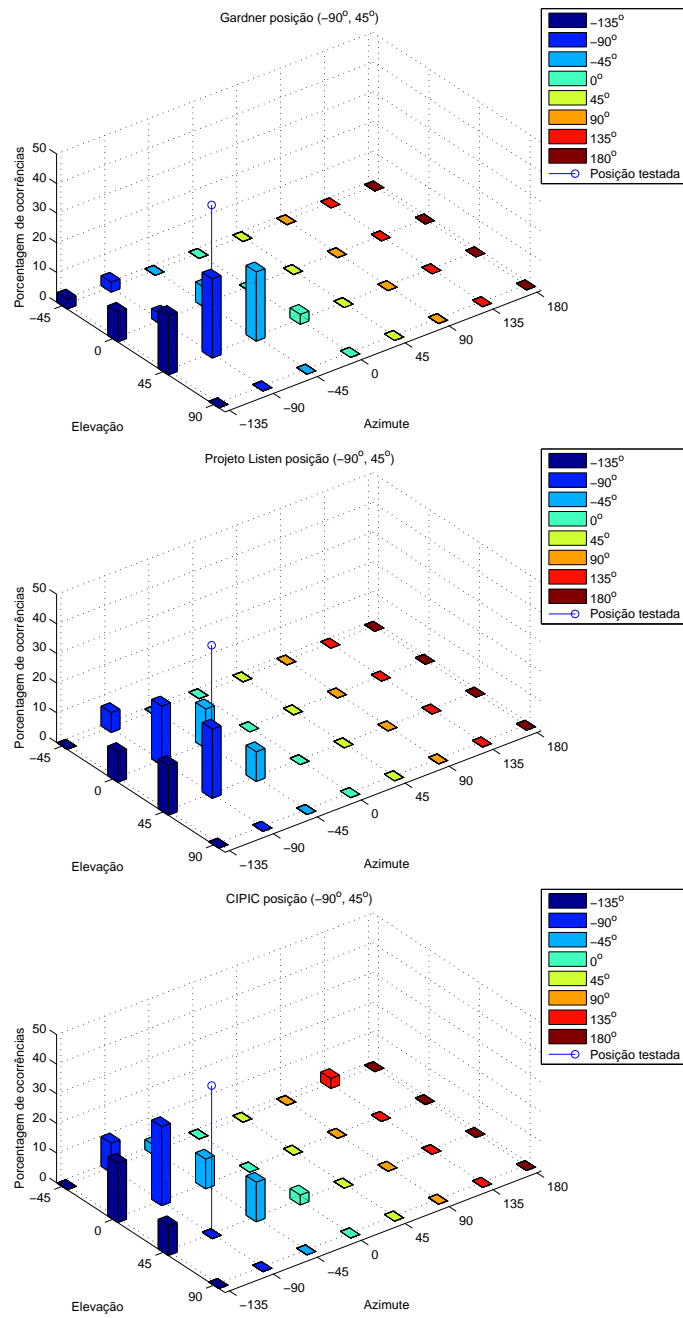


Figura 2.18: Distribuição espacial das respostas, para a posição de teste MAE ( $-90^\circ$ ,  $45^\circ$ ).

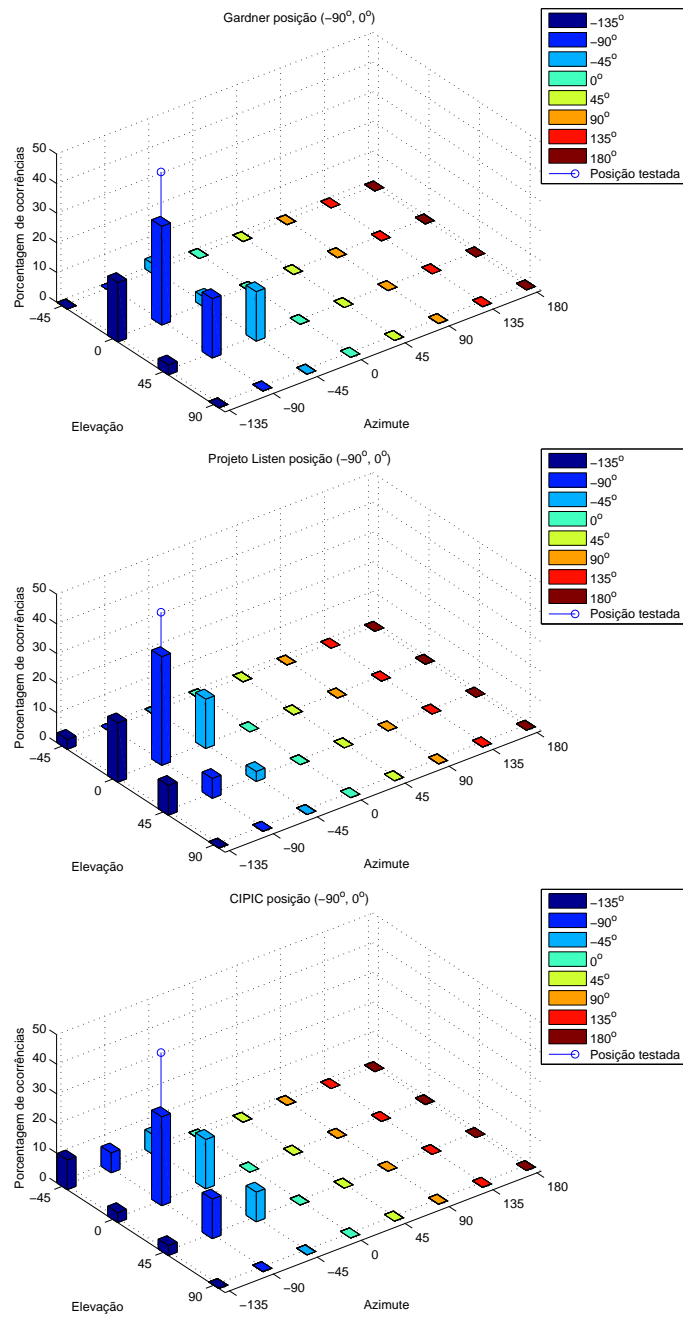


Figura 2.19: Distribuição espacial das respostas, para a posição de teste MHE ( $-90^\circ$ ,  $0^\circ$ ).



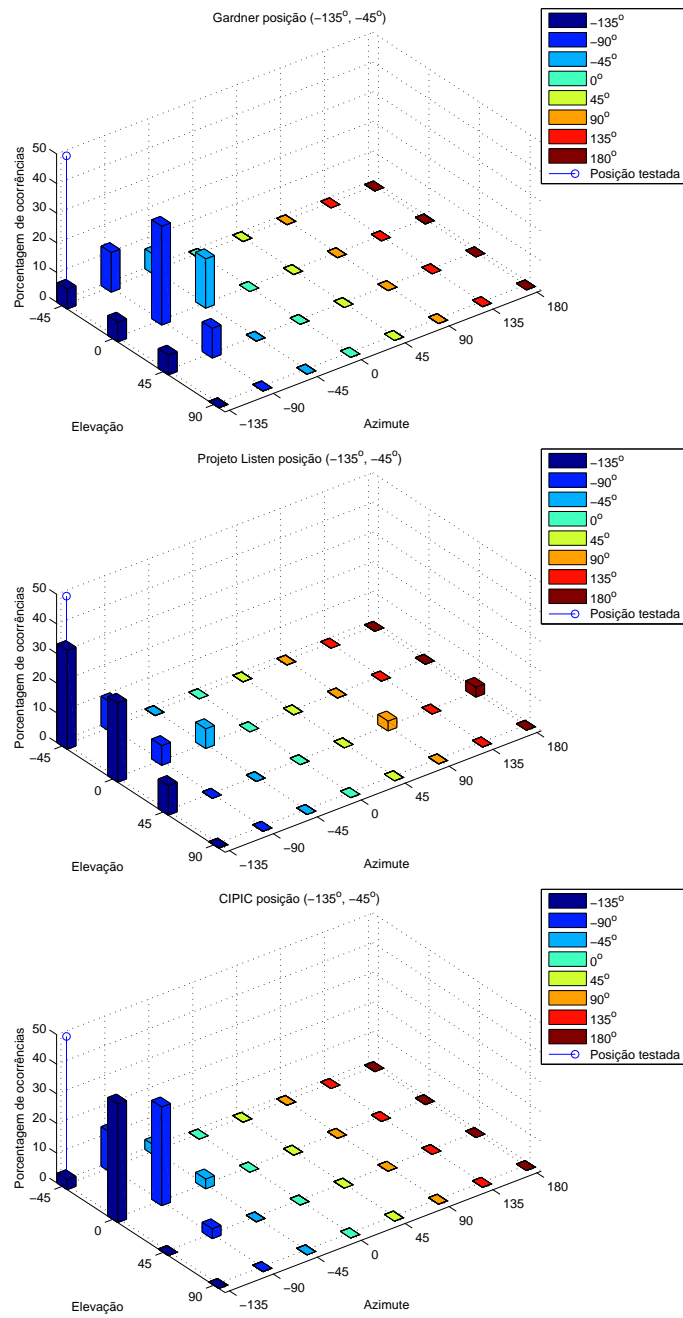


Figura 2.20: Distribuição espacial das respostas, para a posição de teste TBE  $(-135^\circ, -45^\circ)$

Nota-se na Figura 2.17 a ocorrência de reversões frente-trás. A posição de teste FHC (0, 0) apresentou grande quantidade de respostas na posição THC (180°, 0), sendo esta mais assinalada do que a primeira, nos conjuntos do Projeto Listen e da CIPIC. Nota-se também uma dispersão da ocorrência das demais posições, tanto em torno de (0, 0) quanto em torno de (180°, 0).

A Figura 2.18 mostra a dispersão das respostas em torno da posição de teste MAE (−90°, 45°). Nota-se, principalmente para os conjuntos do Projeto Listen e da CIPIC, que há grande quantidade de ocorrências para a elevação zero.

A Figura 2.19 mostra que, mesmo com índice de acerto mais elevado em relação aos demais, os resultados para MHE (−90°, 0) ainda apresentam dispersão em torno da posição de teste.

Na Figura 2.20 novamente tem-se a dispersão ao redor da posição de teste, e mostra também o alto índice de respostas assinaladas em posições no plano horizontal.

Nas Figuras 2.17 a 2.20, nota-se a dispersão da localização do som dos resultados de medida (*blur*) [9], que também é um dos principais erros listados na Seção 1.4. Um fator que influencia a dispersão das respostas é a grande quantidade de alternativas apresentadas, aliada à falta de acurácia de percepção dos ouvintes, uma vez que não foi realizado treinamento prévio dos mesmos.

### 2.5.3 Análise do Erro Médio Quadrático

A fim de auxiliar a análise da diferença de comportamento dos conjuntos quanto à percepção da posição estática, será calculado o *MSE* (*Mean Square Error* - Erro Médio Quadrático) de cada conjunto de medidas. O *MSE* é definido [23] como

$$MSE = E[(\hat{\beta} - \beta)^2], \quad (2.7)$$

onde  $E[\cdot]$  é o valor esperado, e  $\hat{\beta}$  é o estimador do parâmetro  $\beta$ , aqui considerado uma constante.

Pode ser mostrado [23] que

$$E[(\hat{\beta} - \beta)^2] = b^2 + \sigma^2, \quad (2.8)$$

onde  $b$  é a tendenciosidade (*bias*) e  $\sigma^2$  é a variância do estimador.

A tendenciosidade é uma medida de quanto a estimativa do valor esperado de um estimador se afasta do valor que deveria ser encontrado, dá a dimensão do erro sistemático ou tendencioso [23]. A tendenciosidade é dada por

$$b = E[\widehat{\beta}] - \beta. \quad (2.9)$$

A variância é uma medida da dispersão ou variabilidade da distribuição do conjunto [23] — quanto maior a variância, maior a dispersão do mesmo. A variância é dada por

$$\sigma^2 = E[(\widehat{\beta} - E[\widehat{\beta}])^2]. \quad (2.10)$$

No caso em estudo, seja  $P_k$  a variável aleatória “resposta dos ouvintes para a  $k$ -ésima posição testada”, com  $k = 1, \dots, K$ , onde  $K$  é o número de posições testadas.

Seja  $D_k$  a posição nominal da  $k$ -ésima HRTF utilizada no teste, com  $k = 1, \dots, K$ , onde  $K$  é o número de posições testadas.

Como a função densidade de probabilidade não é conhecida, estimam-se os valores de seus parâmetros através dos experimentos, considerando independentes as respostas dos ouvintes dos testes. Assim, seja  $P_{ki}$  a resposta dada pelo  $i$ -ésimo ouvinte, onde  $i = 1, \dots, N$ , e  $N$  é o número de ouvintes.

Devido à quantidade finita de testes realizados, apenas pode-se estimar o valor esperado das respostas dos ouvintes. Dessa forma,  $E[P_k] = \overline{P}_k$  é estimado por

$$\widehat{\overline{P}}_k = \frac{\sum_{i=1}^N P_{ki}}{N}. \quad (2.11)$$

Considerando o estimador da Equação (2.11) e a tendenciosidade dada pela Equação (2.9), define-se o estimador da tendenciosidade como

$$\widehat{b}_k = \widehat{\overline{P}}_k - D_k. \quad (2.12)$$

Seguindo o mesmo raciocínio, o estimador da variância é dado por

$$\widehat{\sigma}_k^2 = \frac{\sum_{i=1}^N (P_{ki} - \widehat{\overline{P}}_k)^2}{N - 1}. \quad (2.13)$$

O estimador do  $MSE$  para a variável aleatória  $P_k$  é dado por

$$\widehat{MSE}_k = \widehat{b}_k^2 + \widehat{\sigma}_k^2. \quad (2.14)$$

Nota-se que  $P_k$  pode se referir somente ao azimute  $\theta$ , somente à elevação  $\phi$  ou a ambos simultaneamente, na forma de um vetor  $(\theta, \phi)$ . Neste último caso, as definições de tendenciosidade e variância devem ser modificadas por

$$\begin{aligned}\widehat{b}_k &= |\widehat{P}_k - D_k| \quad \text{e} \\ \widehat{\sigma}_k^2 &= \frac{\sum_{i=1}^N |P_{ki} - \widehat{P}_k|^2}{N-1},\end{aligned}\tag{2.15}$$

onde  $|\cdot|$  é a norma euclidiana.

Outro fator a ser considerado é a existência de diferentes conjuntos de medidas testados para uma mesma posição  $P_k$ . Seja, portanto,  $P_k^u$  a variável aleatória “respostas dos ouvintes para a  $k$ -ésima posição testada para o  $u$ -ésimo conjunto de HRIRs”, onde  $u = 1, \dots, M$  e  $M$  é o número de conjuntos.

Uma resposta de um ouvinte para uma posição e para um conjunto será denominada “realização”.

Quer-se comparar  $P_k^1, P_k^2, \dots, P_k^M$ ,  $k = 1, \dots, K$ , sob os critérios de tendenciosidade, variância e  $MSE$ .

Utilizando a formulação mostrada, as Tabelas 2.8, 2.9 e 2.10 indicam, para cada posição de teste e cada conjunto de medida, os valores de tendenciosidade, variância e MSE dos testes realizados. A Tabela 2.8 apresenta os resultados dos testes somente quanto ao azimute das posições indicadas pelos ouvintes. A Tabela 2.9 apresenta os resultados somente quanto a elevação, e a Tabela 2.10 leva em conta ambos.

As tabelas foram calculadas espelhando-se os resultados do lado direito e somando com os resultados correspondentes do lado esquerdo, de forma a aproveitar a simetria da cabeça humana.

O cálculo dos dados da Tabela 2.8 desconsiderou as realizações que apresentaram reversão frente-trás, porque o efeito destas na média e na variância resultaria em valores incoerentes. Por exemplo, se o azimute da posição testada fosse 0 graus e se houvesse reversão frente-trás em metade das realizações, o valor médio do azimute seria de 90°, valor que não reflete os resultados do experimento.

Já no cálculo da Tabela 2.9 todas as realizações foram consideradas, inclusive as que apresentaram reversão frente-trás. Isso porque tal reversão influencia somente o azimute e a tabela se refere apenas a valores de elevação.

O cálculo da Tabela 2.10 também desconsidera as realizações que apresentaram reversão frente-trás. Uma vez que a tabela apresenta o efeito conjunto de azimute e elevação, ela também tem a restrição de ser influenciada pela reversão frente-trás. Devido a essa consideração, há pequena diferença entre as médias das elevações apresentadas na Tabela 2.9 e na Tabela 2.10.

Uma posição singular dos testes é a MAC, que se localiza acima da cabeça. Embora seja possível afirmar que sua elevação é de  $90^\circ$ , o azimute pode assumir qualquer valor. Portanto não é possível arbitrar um valor de referência sobre o qual os erros serão calculados. Nas Tabelas 2.8 a 2.10 somente foram computados os valores de elevação da posição MAC. Nas figuras apresentadas anteriormente o valor de MAC foi considerado como  $(0, 90^\circ)$ , somente com o fim de representação gráfica.

Outro cuidado que se tomou diz respeito ao azimute das posições traseiras. As Tabelas 2.8 a 2.10 apresentam os valores de referência entre  $-180^\circ$  e  $180^\circ$ . Entretanto, devem-se desdobrar os ângulos entre  $0$  e  $360^\circ$  para realizar o cálculo da média e da variância em torno de  $180^\circ$ . Analogamente, desdobraram-se os ângulos entre  $0$  e  $-360^\circ$  para se tomar a média e a variância em torno de  $-135^\circ$ .

Fatores como ruído do sistema de reprodução e tipo de fones de ouvido podem influenciar na tendenciosidade de um teste isolado. Entretanto, como estão sendo comparados diferentes conjuntos, as condições de medida os influenciam igualmente.

A Figura 2.21 ilustra os valores de MSE encontrados nas Tabelas 2.8 a 2.10.

Tabela 2.8: Estatísticas dos conjuntos quanto ao azimute, com retirada de reversões.

Posição de Teste	Referência (graus)	Média (graus)			Tendenciosidade (graus)			Variância (graus) <sup>2</sup>			MSE (graus) <sup>2</sup>		
		Gardner	Listen	CIPIC	Gardner	Listen	CIPIC	Gardner	Listen	CIPIC	Gardner	Listen	CIPIC
<b>FAE</b>	-45	-63	-87	-69	-15,0	-42,19	-23,68	725	127	533	950	1.906	1.094
<b>FAC</b>	0	54	64	-19	54,31	64,36	-19,0	3.685	1.709	5.223	6.634	5.852	5.584
<b>FHE</b>	-45	-88	-75	-74	-37,63	-30,11	-28,64	107	1.867	491	1.523	2.774	1.311
<b>FHC</b>	0	8	4	6	7,94	4,10	5,60	3.213	5.977	3.651	3.276	5.994	3.682
<b>FBE</b>	-45	-55	-74	-61	-9,79	-28,64	-15,88	3.127	515	2.010	3.223	1.335	2.262
<b>FBC</b>	0	29	-9	44	29,17	-8,55	43,89	3.697	3.898	2.926	4.547	3.971	4.852
<b>MAE</b>	-90	-89	-92	-81	-0,50	-1,50	-1,0	1.604	1.045	3.128	1.604	1.047	3.129
<b>MAC</b>	0	0	0	0	0,0	0,0	0,0	0	0	0	0	0	0
<b>MHE</b>	-90	-90	-96	-83	0,0	-6,21	-2,50	978	1.117	989	978	1.156	995
<b>MBE</b>	-90	-113	-98	-89	-22,50	-7,50	-8,50	663	2.106	1.045	1.170	2.163	1.117
<b>TAE</b>	-135	-115	-119	-115	20,57	15,58	19,80	520	477	520	943	719	912
<b>TAC</b>	180	152	186	199	-24,13	5,63	23,53	2.278	776	2.546	2.860	808	3.100
<b>THE</b>	-135	-122	-117	-115	12,86	18,33	20,17	429	508	519	594	844	926
<b>THC</b>	180	193	161	185	12,86	-18,53	9,50	757	1.281	831	922	1.624	922
<b>TBE</b>	-135	-102	-133	-111	33,26	1,61	24,11	408	1.122	522	1.514	1.125	1.103
<b>TBC</b>	180	186	173	186	5,63	-7,50	10,87	253	1.174	2.081	285	1.230	2.199

Tabela 2.9: Estatísticas dos conjuntos quanto a elevação, sem retirada de reversões.

Posição de Teste	Referência (graus)	Média (graus)			Tendenciosidade (graus)			Variância (graus) <sup>2</sup>			MSE (graus) <sup>2</sup>		
		Gardner	Listen	CIPIC	Gardner	Listen	CIPIC	Gardner	Listen	CIPIC	Gardner	Listen	CIPIC
<b>FAE</b>	45	21	26	32	-29,0	-19,50	-13,50	801	654	580	1.642	1.034	762
<b>FAC</b>	45	39	31	36	-11,21	-14,46	-9,0	1.798	1.890	975	1.924	2.100	1.056
<b>FHE</b>	0	12	18	18	12,0	18,0	18,0	689	943	922	833	1.267	1.246
<b>FHC</b>	0	17	14	20	17,7	13,97	19,50	945	1.188	1.324	1.237	1.383	1.705
<b>FBE</b>	-45	3	3	23	43,0	48,0	67,50	1.147	1.108	943	2.996	3.412	5.499
<b>FBC</b>	-45	20	8	17	59,50	52,50	61,50	1.870	1.564	1.499	5.410	4.320	5.281
<b>MAE</b>	45	30	20	5	-20,0	-25,50	-40,50	745	794	887	1.145	1.444	2.527
<b>MAC</b>	90	23	27	39	-67,50	-63,0	-51,0	1.371	2.056	1.596	5.927	6.025	4.197
<b>MHE</b>	0	17	8	2	16,50	7,76	1,50	626	444	1.045	898	504	1.047
<b>MBE</b>	-45	-2	6	12	38,50	51,0	57,0	905	521	689	2.388	3.122	3.938
<b>TAE</b>	45	33	20	18	-17,0	-25,50	-27,0	530	933	782	819	1.584	1.511
<b>TAC</b>	45	44	32	30	-6,50	-13,50	-15,0	1.608	1.740	1.331	1.650	1.922	1.556
<b>THE</b>	0	8	3	0	7,50	3,0	0,0	710	829	978	766	838	978
<b>THC</b>	0	24	14	3	24,0	13,97	3,10	2.461	1.559	877	3.037	1.754	887
<b>TBE</b>	-45	-5	-12	-8	35,50	33,0	37,50	887	1.108	431	2.147	2.197	1.837
<b>TBC</b>	-45	41	-3	3	80,50	42,0	48,10	1.940	1.147	424	8.420	2.911	2.738

Tabela 2.10: Estatísticas dos conjuntos, com retirada de reversões.

Posição de Teste	Referência (graus)	Média (graus)			Tendenciosidade (graus)			Variância (graus) <sup>2</sup>			MSE (graus) <sup>2</sup>		
		Gardner	Listen	CIPIC	Gardner	Listen	CIPIC	Gardner	Listen	CIPIC	Gardner	Listen	CIPIC
<b>FAE</b>	(-45, 45)	(-63, 25)	(-87, 28)	(-69, 33)	29,37	45,44	26,48	994	339	754	1.857	2.404	1.455
<b>FAC</b>	(0, 45)	(54, 57)	(64, 57)	(-19, 40)	54,81	65,42	19,58	2.559	1.392	3.837	5.563	5.672	4.220
<b>FHE</b>	(-45, 0)	(-88, 12)	(-75, 12)	(-74, 18)	39,45	32,36	34,4	624	1.834	1.042	2.181	2.881	2.201
<b>FHC</b>	(0, 0)	(8, 27)	(4, 45)	(6, 32)	28,52	45,1	32,25	2.745	2.349	2.804	3.558	4.374	3.844
<b>FBE</b>	(-45, -45)	(-55, 3)	(-74, 8)	(-61, 21)	44,10	60,40	68,6	2.098	592	1.744	4.043	4.240	6.376
<b>FBC</b>	(0, -45)	(29, 35)	(-9, 23)	(44, 45)	80,11	68,4	100,13	2.487	2.188	1.559	8.905	6.817	11.585
<b>MAE</b>	(-90, 45)	(-89, 30)	(-92, 20)	(-81, 5)	20,1	25,54	40,51	2.432	1.904	4.158	2.833	2.557	5.800
<b>MAC</b>	(0, 90)	(0, 23)	(0, 27)	(0, 39)	67,50	63,0	51,0	1.266	2.256	1.553	5.822	6.225	4.154
<b>MHE</b>	(-90, 0)	(-90, 17)	(-96, 8)	(-83, 2)	16,50	9,94	2,92	1.661	1.561	2.107	1.933	1.660	2.115
<b>MBE</b>	(-90, -45)	(-113, -2)	(-98, 6)	(-89, 12)	44,59	51,55	57,63	1.625	2.722	1.796	3.613	5.379	5.117
<b>TAE</b>	(-135, 45)	(-115, 32)	(-119, 16)	(-115, 18)	27,41	33,29	33,48	767	1.285	1.169	1.519	2.393	2.290
<b>TAC</b>	(180, 45)	(152, 21)	(186, 17)	(199, 19)	37,90	28,42	34,85	948	1.208	1.972	2.384	2.016	3.186
<b>THE</b>	(-135, 0)	(-122, 6)	(-117, 2)	(-115, 0)	14,37	18,41	20,17	1.095	1.264	1.531	1.302	1.603	1.938
<b>THC</b>	(180, 0)	(193, 2)	(161, 9)	(185, 0)	13,7	20,81	9,50	854	1.683	998	1.025	2.116	1.088
<b>TBE</b>	(-135, -45)	(-102, -2)	(-133, -13)	(-111, -6)	50,53	32,18	45,49	1.113	2.219	896	3.667	3.254	2.965
<b>TBC</b>	(180, -45)	(186, 20)	(173, -8)	(186, 2)	60,72	38,24	48,20	405	1.784	1.994	4.092	3.246	4.317



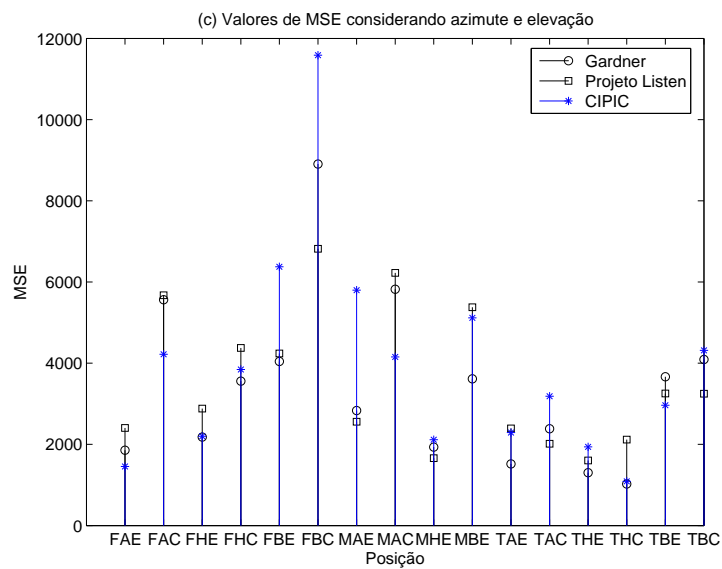
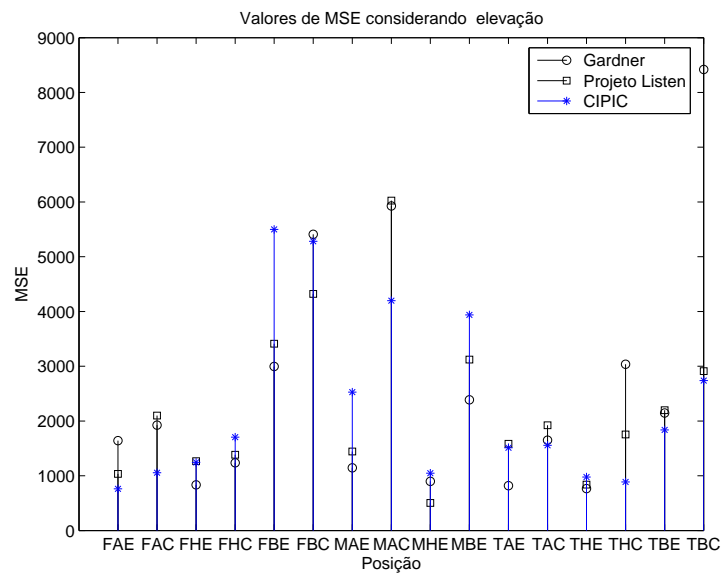
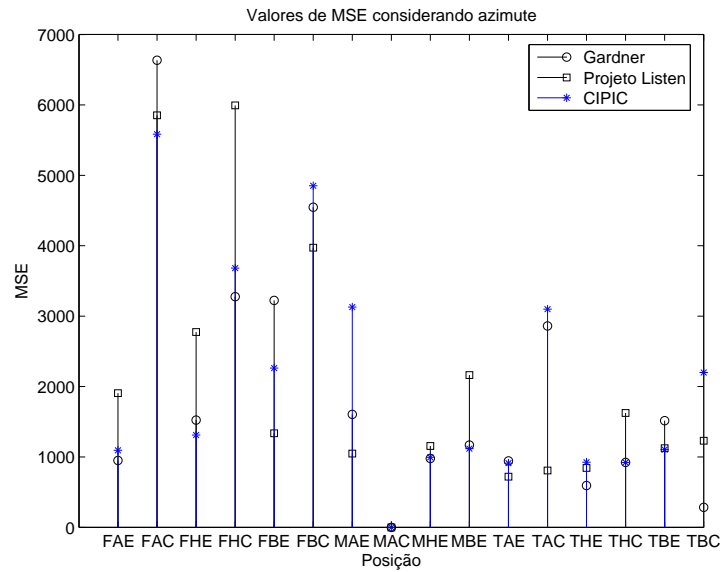


Figura 2.21: MSE por conjunto e posição.

### **Análise quanto ao azimute**

Com relação somente ao azimute, nota-se que o MSE foi maior para a parte da frente do que para a parte do meio e de trás, para todos os conjuntos de medida. Na parte da frente o MSE foi menor para posições com azimute  $-45^\circ$  (posições FAE, FHE e FBE) do que para posições com azimute zero (FAC, FHC e FBC).

Na parte do meio, a posição MHE ( $-90^\circ, 0$ ) obteve MSE baixo em relação a outras posições, indicando ser particularmente mais fácil de ser localizada. A posição MAC não foi considerada, por se localizar acima da cabeça.

Na parte de trás, as posições com azimute  $-135^\circ$  (TAE, THE e TBE) apresentaram MSE mais baixo do que posições com azimute  $180^\circ$  (TAC, THC e TBC). Tanto na parte da frente quanto na de trás, houve a tendência de os ouvintes terem maior dificuldade em localizar o azimute quando a fonte virtual estava posicionada exatamente no plano mediano.

Não parece haver indícios fortes de que um conjunto de medidas seja melhor do que os outros para os testes realizados, quanto à localização do azimute.

### **Análise quanto à elevação**

Observando a Tabela 2.8, de forma geral as posições com elevação de referência na elevação zero (FHE, FHC, MHE, THE e THC) obtiveram o menor valor de MSE, seguido pelas posições acima do plano horizontal (FAE, FAC, MAE, TAE e TAC). Já as posições abaixo do plano horizontal (FBE, FBC, MBE, TBE e TBC) apresentaram valor de MSE elevado, em relação às demais posições. Para as posições de referência com elevação  $-45^\circ$  as elevações médias dos ouvintes foram valores próximos à elevação zero.

A posição MAC, acima da cabeça, mostrou-se especialmente difícil de se localizar. Para esta posição os valores de tendenciosidade e dispersão são elevados, para todos os conjuntos de medida.

Não parece haver indícios fortes de que um conjunto de medidas seja melhor do que os outros para os testes realizados, quanto à localização da elevação.

## Análise quanto ao azimute e elevação

Considerando o efeito conjunto do azimute e da elevação, o maior valor de MSE ocorreu para a posição na parte da frente, no plano mediano e abaixo do plano horizontal FBC, reunindo as piores condições de localização descritas anteriormente. A seguir, veio a posição na parte da frente e abaixo do plano horizontal FBE.

A posição MAC, acima da cabeça, também apresentou alto valor de MSE.

De modo geral, as posições localizadas no plano mediano (FAC, FHC, TAC, THC e TBC) apresentaram maior MSE do que as respectivas posições à esquerda do plano mediano (FAE, FHE, TAE, THE e TBE).

Novamente, não parece haver indícios fortes de que um conjunto de medidas seja melhor do que os outros para os testes realizados, quanto à localização de azimute e elevação.

### 2.5.4 Número de Reversões

A Tabela 2.11 apresenta a quantidade de reversões frente-trás, esquerda-direita e cima-baixo ocorridos por posição. A última linha da Tabela 2.11 apresenta a porcentagem de ocorrências de cada tipo de reversão com relação ao total de medidas de cada conjunto. A análise deste fator demonstra que os conjuntos apresentam a mesma ordem de grandeza de ocorrência de reversões, indicando um desempenho parecido dos 3 conjuntos de HRTFs.

Tabela 2.11: Porcentagem de reversões por posição de teste.

Posição de teste	Frente-Trás			Esquerda-Direita			Cima-Baixo		
	Gardner	Proj. Listen	CIPIC	Gardner	Proj. Listen	CIPIC	Gardner	Proj. Listen	CIPIC
<b>FAE</b>	33,33	46,67	36,67	0,0	0,0	0,0	6,67	3,33	3,33
<b>FAC</b>	36,67	46,67	36,67	0,0	0,0	0,0	3,33	6,67	3,33
<b>FHE</b>	36,67	36,67	26,67	0,0	0,0	0,0	0,0	0,0	0,0
<b>FHC</b>	40,0	63,33	43,33	0,0	0,0	0,0	0,0	0,0	0,0
<b>FBE</b>	50,0	63,33	43,33	3,33	0,0	3,33	23,33	30,0	60,0
<b>FBC</b>	56,67	60,0	66,67	0,0	0,0	0,0	36,67	23,33	26,67
<b>MAE</b>	0,0	0,0	0,0	0,0	0,0	3,33	6,67	6,67	16,67
<b>MAC</b>	0,0	0,0	0,0	0,0	0,0	0,0	0,0	13,33	6,67
<b>MHE</b>	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
<b>MBE</b>	0,0	0,0	0,0	0,0	3,33	0,0	20,0	20,0	33,33
<b>TAE</b>	16,67	13,33	16,67	0,0	0,0	0,0	0,0	10,0	6,67
<b>TAC</b>	23,33	10,0	13,33	0,0	0,0	0,0	6,67	10,0	3,33
<b>THE</b>	6,67	10,0	3,33	0,0	0,0	0,0	0,0	0,0	0,0
<b>THC</b>	10,0	26,67	20,0	0,0	0,0	0,0	0,0	0,0	0,0
<b>TBE</b>	23,33	6,67	6,67	0,0	3,33	0,0	16,67	16,67	3,33
<b>TBC</b>	30,0	16,67	16,67	0,0	0,0	0,0	60,0	16,67	13,33
<b>% Total</b>	<b>22,71</b>	<b>25,0</b>	<b>20,63</b>	<b>0,21</b>	<b>0,42</b>	<b>0,42</b>	<b>11,25</b>	<b>9,79</b>	<b>11,4</b>

Nota-se na Tabela 2.11 que o número de reversões das posições da frente para trás é maior do que o inverso, para todos os conjuntos de medida.

O número de reversões esquerda-direita é desprezível, para todos os conjuntos.

As reversões cima-baixo ocorreram em maior porcentagem para as posições abaixo do plano horizontal (FBE, FBC, MBE, TBE e TBC), indicando a dificuldade da localização destes, para o teste realizado.

O número de reversões frente-trás e as reversões cima-baixo também está relacionado com o alto MSE apresentado pelas posições da frente e pelas posições abaixo do plano horizontal.

Não parece haver haver indícios fortes de que um conjunto de medidas tenha comportamento melhor do que os outros quanto a reversões.

## 2.6 Conclusões

Neste capítulo três conjuntos de medidas de HRTFs disponíveis na literatura foram apresentados e analisados. Além disso, os conjuntos tiveram a notação padronizada, a fim de facilitar a sua utilização em trabalhos futuros.

As diferenças entre o número de pontos de medidas, número de coeficientes em cada HRIR e as faixas de variação de azimutes e elevações foram descritas na Tabela 2.1.

A seguir foi analisada a distribuição de energia das HRIRs com a posição. O comportamento dos conjuntos foi aproximadamente semelhante ao comportamento esperado. Além disso, notou-se o efeito conhecido como *bright spot* em todos os conjuntos.

Outra análise realizada foi relativa à representação das HRIRs por um conjunto de dados menor, através da utilização da KLT. De modo geral, mostrou-se que é possível descrever os conjuntos de HRIRs com menos funções-base do que o conjunto original e com perdas tão pequenas quanto se queira. Alguns dos conjuntos de HRIRs apresentaram maior capacidade de compressão do que outros, devido ao número de posições de medidas relativas à quantidade de coeficientes em cada HRIR.

Foram feitos testes subjetivos de percepção de posição estática, utilizando

os conjuntos de HRTFs. Foram calculados a tendenciosidade, a variância e o erro médio quadrático (MSE) dos mesmos. Além disso, foram calculados os números de reversões apresentados. Para o teste realizado as posições da frente foram mais difíceis de localizar que as posições do meio e de trás; as posições no plano mediano foram mais difíceis de localizar que as posições à esquerda ou à direita do mesmo; e que as posições de baixo foram mais difíceis de localizar que as posições no plano horizontal e acima do mesmo. Não parece haver diferenças significativas entre os conjuntos de medidas de HRTFs pesquisados.

Os resultados aqui descritos servirão de base para estudos realizados em capítulos posteriores.

### **Contribuições deste capítulo:**

- Apresentação de 3 conjuntos de medidas de HRTFs e análise comparativa destes quanto à grade de medidas, ao número de coeficientes, à distribuição espacial da energia e ao comportamento na decomposição KLT.
- Realização de testes subjetivos de percepção de posição estática para os 3 conjuntos apresentados anteriormente. Análise comparativa dos resultados quanto ao erro médio quadrático (MSE) e reversões.

# Capítulo 3

## Modelos Geométricos de ITD

### 3.1 Introdução

A Diferença de Tempo Interaural (ITD) é um dos fatores mais importantes no posicionamento da fonte virtual, conforme a teoria dúplex citada no Capítulo 1.

Por um lado, dos conjuntos de HRTFs apresentados no Capítulo 2 podem ser extraídos valores de ITD, através da obtenção de HRTFs de fase mínima e comparação com o conjunto original.

Por outro lado, existem modelos geométricos para estimar a ITD. Tais modelos utilizam uma simplificação da cabeça humana para calcular o tempo que o som demora para percorrer a trajetória da fonte sonora até as orelhas esquerda e direita, obtendo então a ITD. Serão vistos os modelos que consideram um caminho retilíneo entre a fonte sonora e as orelhas, e os que consideram a difração ocorrida pela presença da cabeça.

Os resultados serão confrontados, de modo a testar a validade dos modelos geométricos em relação aos conjuntos de medida de HRTFs.

Por fim, serão descritas algumas aplicações da teoria exposta. A Figura 3.1 ilustra os assuntos a serem abordados.

### 3.2 ITD dos Conjuntos de Medida

Para se estimar a ITD dos conjuntos de HRTFs, primeiramente obtém-se do conjunto original um conjunto de fase mínima. A diferença de fase entre a versão

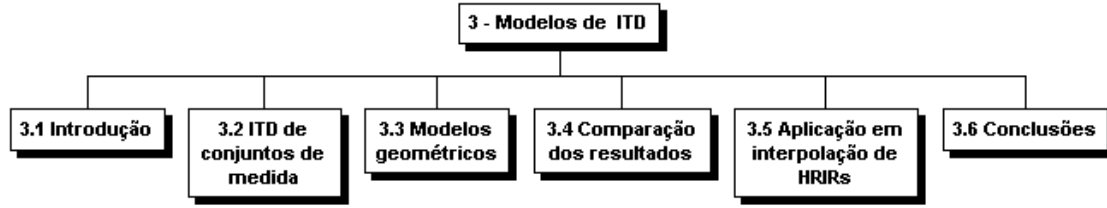


Figura 3.1: Conteúdo do capítulo sobre a ITD.

original e a de fase mínima é analisada, quanto à viabilidade de se representá-la através de um atraso puro. Em caso positivo, cada HRTF pode ser dada através da combinação de um atraso puro e de uma HRTF de fase mínima. A ITD é dada pela diferença dos atrasos dos lados esquerdo e direito do ouvinte.

### 3.2.1 Método de Obtenção de HRTFs de Fase Mínima

Uma seqüência de fase mista  $x[k]$  pode ser descrita por uma convolução entre uma seqüência de fase mínima e uma seqüência passa-tudo [24], conforme

$$x[k] = (x_{\min} * x_{\text{ap}})[k], \quad (3.1)$$

onde  $x_{\min}[k]$  é de fase mínima com o mesmo módulo espectral da seqüência original, e  $x_{\text{ap}}[k]$  é uma seqüência passa-tudo de módulo espectral unitário.

Sejam  $x[k]$  uma HRIR e  $H(e^{j\omega})$  a sua transformada de Fourier (FT).

Um sistema homomórfico é uma generalização de um sistema linear. Engloba classes de sistemas não lineares que obedecem um princípio de superposição generalizado [25]. Uma das funções utilizadas em tais sistemas é o logaritmo. Aplicando-se a função logaritmo neperiano ( $\ln$ ) a  $H(e^{j\omega})$ , tem-se que

$$\ln(H(e^{j\omega})) = \ln(|H(e^{j\omega})|) + j\angle H(e^{j\omega}), \quad (3.2)$$

onde  $|\cdot|$  representa o operador módulo e  $\angle$ , a fase.

Denomina-se por  $\widehat{X}(e^{j\omega})$  a transformada de Fourier do cepstrum complexo de  $x[k]$ . Define-se

$$\widehat{X}(e^{j\omega}) = \ln(|H(e^{j\omega})|) + j\angle H(e^{j\omega}), \quad (3.3)$$

onde  $\angle H(e^{j\omega})$  é uma função contínua periódica em relação a  $\omega$ .

Tabela 3.1: Notação dos tipos de transformações cepstrais.

Transformação	Notação
Cepstrum complexo	$\hat{x}[k]$
FT do cepstrum complexo	$\hat{X}(e^{j\omega})$
Cepstrum	$c[k]$
FT do cepstrum	$C(e^{j\omega})$

O valor de  $\angle H(e^{j\omega})$  da Equação (3.3) é ambíguo, já que para cada  $\omega$  pode-se adicionar um múltiplo de  $2\pi$ . Isso obriga que os valores de fase sejam “desdobrados” pela adição ou subtração de valores inteiros de  $2\pi$  à fase de  $\angle H(e^{j\omega})$ , de modo a resultar numa curva contínua.

O cepstrum complexo  $\hat{x}[k]$  é dado pela transformada de Fourier inversa (iFT) da Equação (3.3), como em

$$\hat{x}[k] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{X}(e^{j\omega}) e^{j\omega k} d\omega. \quad (3.4)$$

Definir-se-á, a seguir, o cepstrum, também denominado cepstrum real.

A transformada de Fourier do cepstrum, denominada por  $C(e^{j\omega})$ , é definida como a parte real da transformada de Fourier do cepstrum complexo da Equação (3.3), conforme

$$C(e^{j\omega}) = \ln(|H(e^{j\omega})|). \quad (3.5)$$

Por sua vez, o cepstrum  $c[k]$  é dado pela transformada de Fourier inversa da Equação (3.5), conforme

$$c[k] = \frac{1}{2\pi} \int_{-\pi}^{\pi} C(e^{j\omega}) e^{j\omega k} d\omega. \quad (3.6)$$

A Tabela 3.1 sumariza a notação apresentada.

Pelas definições apresentadas (Equações (3.3), (3.5) e (3.6)) o cepstrum  $c[k]$  é a transformada de Fourier inversa da parte real de  $\hat{X}(e^{j\omega})$ . Utilizando a propriedade de que a parte real de uma transformada de Fourier é dada pela parte par da seqüência original, tem-se a relação entre o cepstrum e o cepstrum complexo como

$$c[k] = \frac{\hat{x}[k] + \hat{x}[-k]}{2}. \quad (3.7)$$

Se o cepstrum complexo  $\hat{x}[k]$  for causal, pode-se obter seu valor em função do cepstrum  $c[k]$ , conforme a Equação (3.8). Para facilitar a notação, o cepstrum complexo causal será denominado  $\hat{x}_{\min}[k]$ .

$$\hat{x}_{\min}[k] = c[k] l_{\min}[k], \quad (3.8)$$



onde

$$l_{\min}[k] = \begin{cases} 0, & \text{para } k < 0; \\ 1, & \text{para } k = 0; \\ 2, & \text{para } k > 0. \end{cases} \quad (3.9)$$

O cepstrum complexo  $\hat{x}[k]$  tem a propriedade de ser causal se, e somente se, a seqüência  $x[k]$  é de fase mínima [24].

Com base na teoria exposta, sugere-se em [24] uma forma de se obter uma seqüência de fase mínima a partir de uma de fase não-mínima, conforme a Figura 3.2.

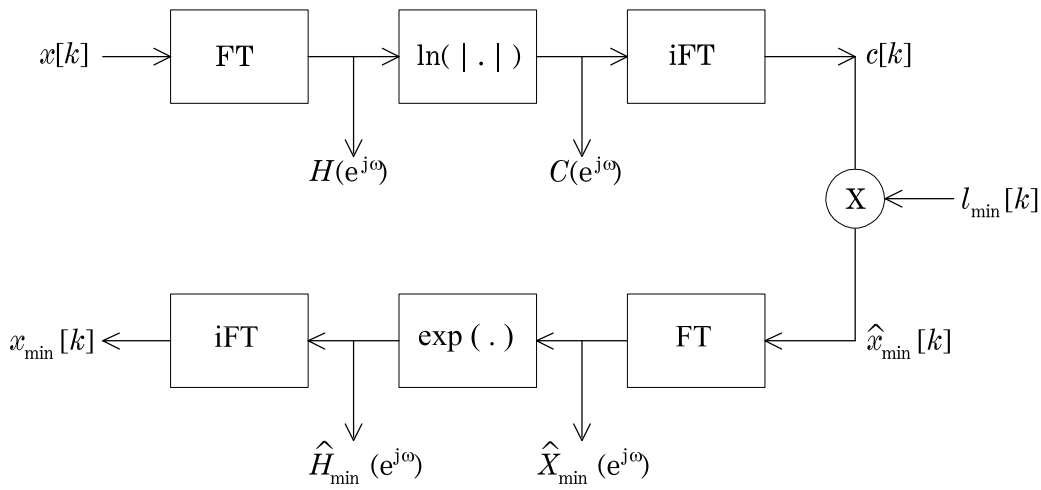


Figura 3.2: Fluxograma de obtenção da seqüência de fase mínima.

Primeiramente faz-se a transformada de Fourier da seqüência original, obtendo-se  $H(e^{j\omega})$ . Faz-se a transformação cepstral do módulo de  $H(e^{j\omega})$ , chegando à transformada de Fourier do cepstrum  $C(e^{j\omega})$ . Faz-se a transformada de Fourier inversa, obtendo-se o cepstrum  $c[k]$ . A seguir, obtém-se do cepstrum o cepstrum complexo de fase mínima  $\hat{x}_{\min}[k]$ , através da aplicação da janela  $l_{\min}[k]$ , conforme a Equação (3.8). O cepstrum complexo de fase mínima conserva a mesma magnitude de  $H(e^{j\omega})$  e tem fase mínima. De  $\hat{x}_{\min}[k]$  faz-se a transformação cepstral inversa, o que resulta na seqüência de fase mínima no domínio do tempo  $x_{\min}[k]$ .

Antes de se aplicar os conceitos descritos, deve-se tomar o cuidado de se fazer um *zero-padding* nas HRIRs, de modo a evitar sobreposição (*aliasing*) no tempo.

Aplicou-se a técnica citada para obter o conjunto de HRTFs de fase mínima a

partir dos conjuntos originais de HRTFs expostos no Capítulo 2. A Figura 3.3 mostra o comportamento obtido, para algumas coordenadas. Nota-se que a diferença entre a fase do conjunto original e o conjunto de fase mínima é semelhante a uma reta, para os três conjuntos de medidas estudados.

### 3.2.2 Extração do excesso de fase

A partir da transformada de Fourier da Equação (3.1), tem-se

$$H_{\text{ap}}(e^{j\omega}) = \frac{H(e^{j\omega})}{H_{\text{min}}(e^{j\omega})}, \quad (3.10)$$

onde  $H_{\text{ap}}(e^{j\omega})$  é a resposta na frequência da seqüência passa-tudo e  $H_{\text{min}}(e^{j\omega})$ , a da seqüência de fase mínima.

A magnitude de  $H_{\text{ap}}(e^{j\omega})$  é, portanto

$$|H_{\text{ap}}(e^{j\omega})| = \frac{|H(e^{j\omega})|}{|H_{\text{min}}(e^{j\omega})|} = 1, \quad (3.11)$$

e a fase

$$\angle H_{\text{ap}}(e^{j\omega}) = \angle H(e^{j\omega}) - \angle H_{\text{min}}(e^{j\omega}), \quad (3.12)$$

que pode ser vista como o excesso de fase de  $H(e^{j\omega})$ .

O excesso de fase entre o sinal original e o de fase mínima foi obtido com o procedimento descrito acima, para os conjuntos de medidas de HRTFs. Alguns dos resultados estão representados na Figura 3.4. Nota-se que  $\angle H_{\text{ap}}(e^{j\omega})$  é um valor dependente da frequência.

Quer-se substituir o excesso de fase  $\angle H_{\text{ap}}(e^{j\omega})$  por um atraso puro, que independa da frequência. Se a transformada de Fourier de  $x[k]$  é  $H(e^{j\omega})$ , pela propriedade de deslocamento no tempo da transformada de Fourier [24], um atraso  $\tau$  no tempo provoca deslocamento de fase  $e^{-j\omega\tau}$ , conforme

$$FT(x(k - \tau)) = e^{-j\omega\tau} H(e^{j\omega}), \quad (3.13)$$

onde  $FT(\cdot)$  é o operador transformada de Fourier.

Quer-se estimar o atraso  $\tau$ , de modo que a reta dada pela fase de  $e^{-j\omega\tau}$  aproxime o excesso de fase  $\angle H_{\text{ap}}(e^{j\omega})$  obtido anteriormente. O atraso puro  $\tau$  foi estimado pela mediana da aproximação da derivada do excesso de fase na DFT,

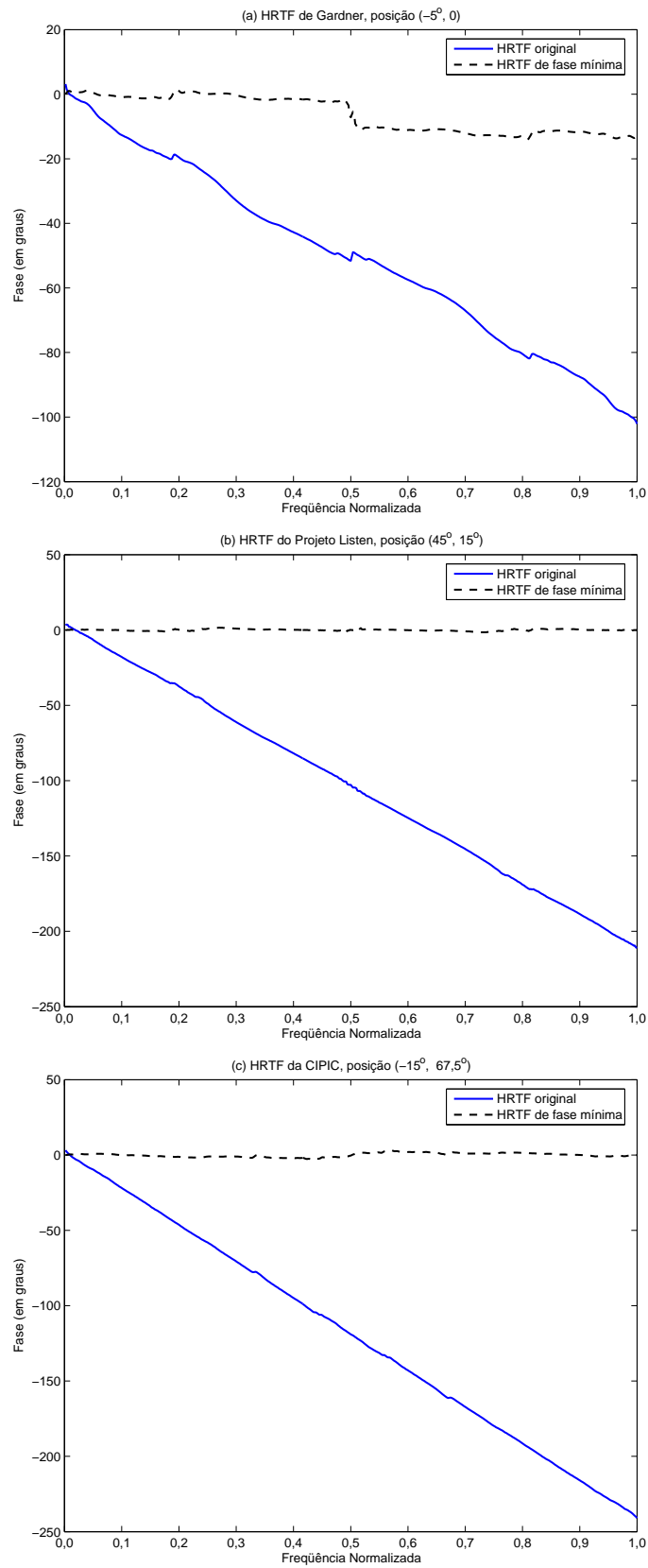


Figura 3.3: Comparações entre a fase do conjunto original e de fase mínima.

conforme

$$\tau = \frac{d}{d\omega} \left[ \angle \left( \frac{H(e^{j\omega})}{H_{\min}(e^{j\omega})} \right) \right] \approx \text{mediana} \left( \frac{H(e^{j\omega_i})}{H_{\min}(e^{j\omega_i})} - \frac{H(e^{j\omega_{i-1}})}{H_{\min}(e^{j\omega_{i-1}})} \right), \quad (3.14)$$

onde  $i$  é o índice da raia da DFT.

Assumiu-se que a reta modelada deveria representar o comportamento da maioria das inclinações da diferença de fase desconsiderando a minoria das inclinações que tiveram comportamento errático. Assim, sob este critério, a mediana é uma heurística melhor do que um método como o de mínimos quadrados, por exemplo, já que no último tais inclinações erráticas influenciariam fortemente na inclinação da reta.

A fase correspondente ao atraso puro obtido através da Equação (3.14) e o excesso de fase entre o conjunto original e o conjunto de fase mínima estão representados na Figura 3.4. O comportamento do excesso de fase dos três conjuntos analisados, para todas as posições de medida, foi semelhante aos gráficos apresentados na Figura 3.4, de modo que parece ser possível representar a diferença de fase por um atraso. Em [2], diz-se que, no caso de HRTFs, o filtro passa-tudo é aproximadamente equivalente a um atraso puro, pelo menos para frequências abaixo de  $8 \sim 10$  kHz.

Portanto, o procedimento para estimativa da ITD dos conjuntos de medida começa estimando-se o excesso de fase  $\angle H_{\text{ap}}(e^{j\omega})$  para as orelhas esquerda e direita do indivíduo. O excesso de fase é substituído por um atraso puro, e obtém-se a ITD do conjunto de medidas para cada posição, subtraindo os atrasos relativos às orelhas esquerda e direita.

Utilizando a técnica citada, obtiveram-se as ITDs para todas as posições de medida de cada um dos conjunto de HRTFs utilizados. Tais valores de ITD serão comparados com resultados de modelos geométricos apresentados a seguir.

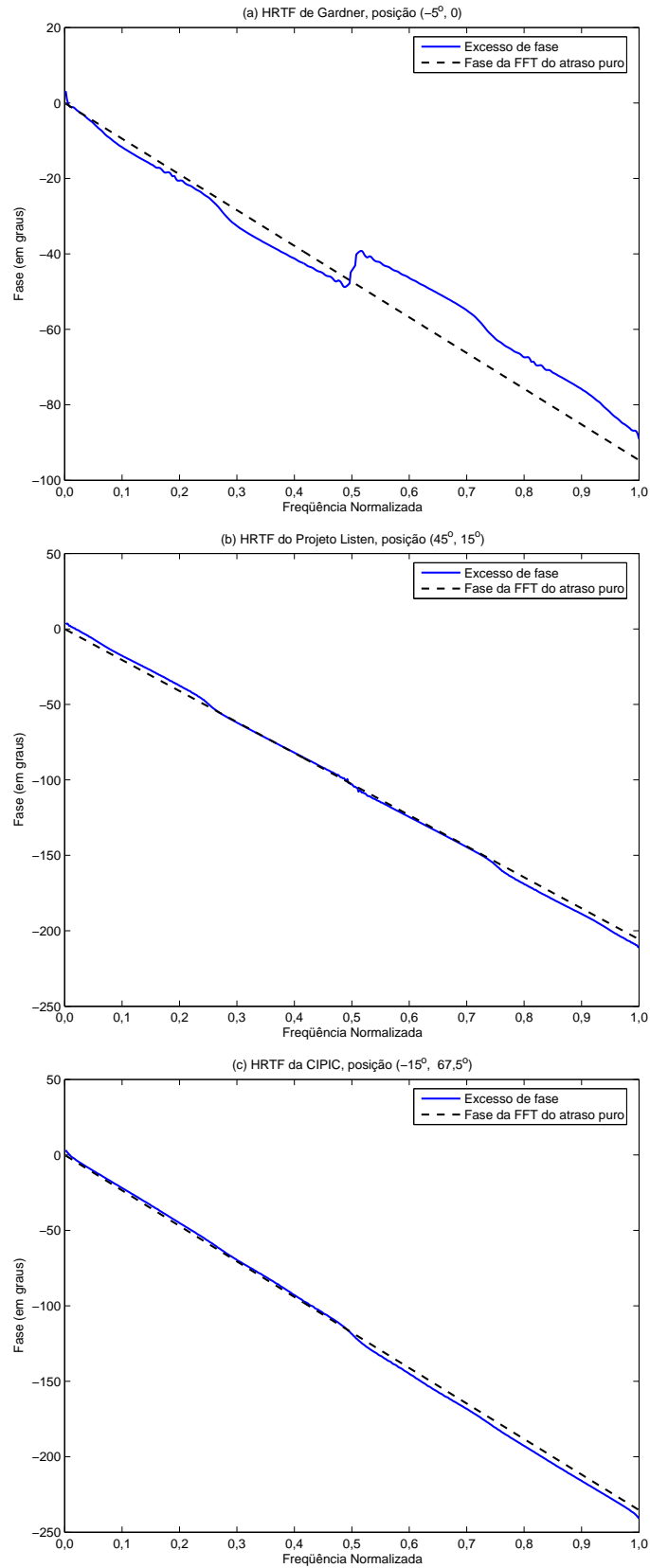


Figura 3.4: Comparações entre o excesso de fase do conjunto original e a fase do atraso puro.

### 3.3 Apresentação dos Modelos Geométricos

Nesta seção serão apresentados modelos geométricos para estimativa da ITD. São dois modelos: o primeiro sugere uma aproximação da trajetória entre a fonte sonora e cada orelha como sendo uma reta, o segundo modelo propõe uma trajetória que leva em conta o efeito da difração.

#### 3.3.1 Trajetória Retilínea entre a Fonte Sonora e as Orelhas

O modelo mais simples de ITD considera que o som percorre uma trajetória retilínea entre a fonte e as orelhas, conforme a Figura 3.5.

O cálculo da ITD é dado pela diferença de tempo que o som leva para sair da fonte virtual e chegar à orelha esquerda e à orelha direita.

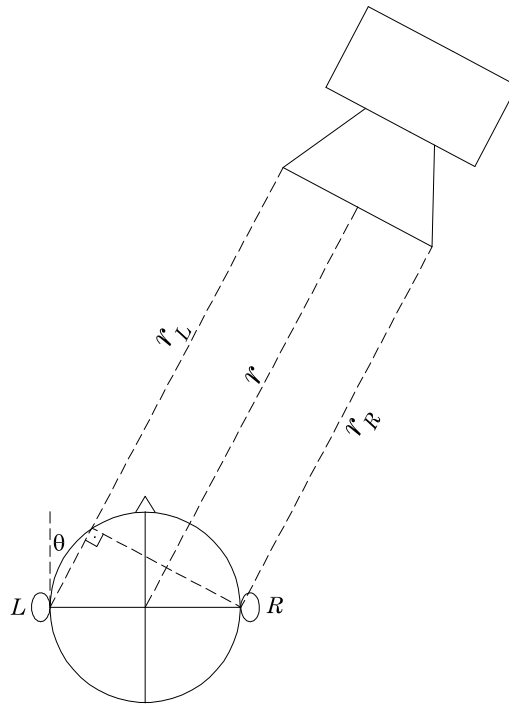


Figura 3.5: Modelo de ITD com trajetória retilínea para o caso 2D.

Para a formulação a seguir, toma-se como hipótese que a distância entre a fonte sonora e o centro geométrico da cabeça é bem maior que a distância entre as orelhas, de modo que as retas  $r_L$ ,  $r$  e  $r_R$  da Figura 3.5 são praticamente paralelas. Outra hipótese é a de que a ITD é independente da frequência.

Primeiramente far-se-á a análise para o caso 2D, estendendo-a a seguir para o 3D. Considera-se que a fonte virtual está localizada nas coordenadas  $(\theta, 0)$ , ou seja, no plano horizontal.

Através da formulação e das hipóteses descritas, a ITD pode ser estimada por

$$\text{ITD} = T_{AL} - T_{AR} = \frac{r_L}{c} - \frac{r_R}{c}, \quad (3.15)$$

onde  $c$  é a velocidade do som no meio,  $T_{AL}$  é o tempo que o som leva para sair da fonte sonora  $A$  até a orelha esquerda  $L$ ,  $T_{AR}$  é o tempo da fonte até a orelha direita  $R$ ,  $r_L$  é a distância entre a fonte e a orelha esquerda,  $r_R$  é a distância entre a fonte e a orelha direita.

O valor da diferença entre as trajetórias  $r_L - r_R$  pode ser obtido por geometria plana, através da Figura 3.5. Desse modo, o valor da ITD é dado pela Equação (3.16).

$$\frac{r_L}{c} - \frac{r_R}{c} = \frac{d \sin \theta}{c}, \quad (3.16)$$

onde  $\theta$  é o azimute da fonte sonora e  $d$  é o diâmetro da esfera de referência.

A Equação (3.16) é uma estimativa de ITD para o caso em que a fonte virtual encontra-se no plano horizontal. Para o caso 3D, a fonte virtual encontra-se numa posição qualquer sobre uma esfera de raio  $r$ , numa posição dada por  $(\theta, \phi)$ , conforme a Figura 3.6. Analogamente ao caso 2D, deve-se formular a diferença de tempo entre a fonte sonora e as orelhas esquerda e direita.

Por hipótese, a distância entre as orelhas  $d$  é bem menor que a distância  $r$  entre o centro geométrico da cabeça e a fonte sonora. Através de considerações geométricas, desenvolvidas no Apêndice A, obtém-se

$$T_{AL} - T_{AR} = \frac{d \sin \theta \cos \phi}{c}, \quad (3.17)$$

conforme a Figura 3.7.

A hipótese de uma trajetória retilínea entre a fonte e as orelhas não é natural, uma vez que tal caminho estaria atravessando a cabeça do indivíduo, conforme ilustrado nas Figuras 3.5 a 3.6.

Apesar de a hipótese de caminho retilíneo não ser a mais adequada, o resultado final de estimativa de ITD é de fácil manipulação algébrica e pode ser uma aproximação suficiente, dependendo da aplicação.

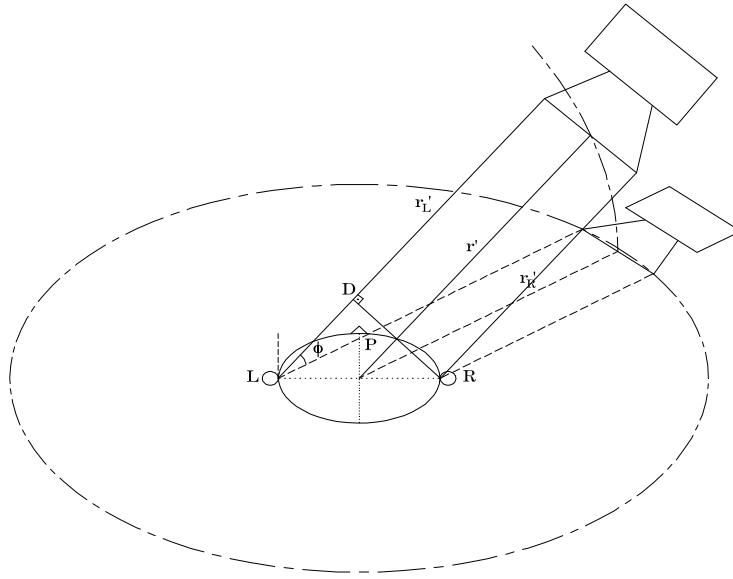


Figura 3.6: ITD trajetória retilínea para o caso 3D.

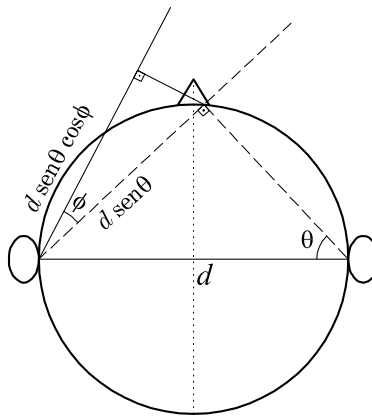


Figura 3.7: Detalhe da trajetória retilínea para o caso 3D.

### 3.3.2 Trajetória com Efeito da Difração da Cabeça

Este modelo de ITD estabelece que o som sofre difração na trajetória da fonte sonora às orelhas, devido à presença da cabeça. Considera-se que o som percorre uma linha reta entre a fonte e o ponto de tangência da cabeça, de onde segue uma trajetória que acompanha o arco de circunferência da cabeça até chegar às orelhas, conforme ilustra a Figura 3.8, para o caso em que a fonte sonora encontra-se no plano horizontal.

Toma-se como hipótese que a cabeça humana pode ser aproximada por uma esfera de diâmetro  $d$ .



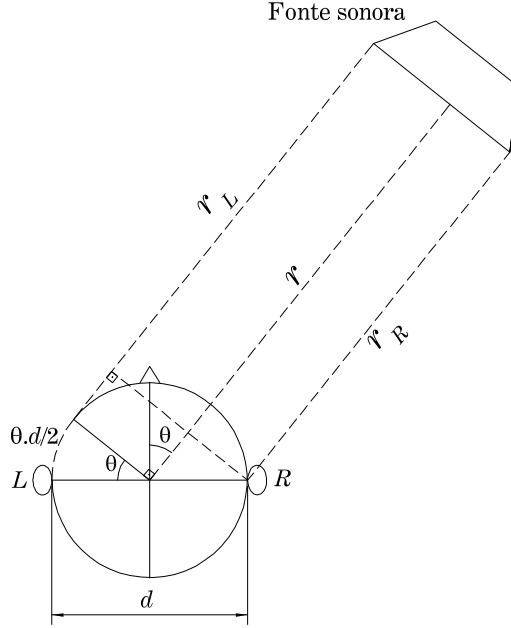


Figura 3.8: ITD com difração da cabeça para o caso 2D.

A partir da Figura 3.8, chega-se a

$$\text{ITD} = \frac{r_L - r_R + \theta d/2}{c}. \quad (3.18)$$

Considera-se que a distância entre a cabeça e a fonte sonora é bem maior do que a distância entre orelhas, de forma que  $r_L$  é paralelo a  $r_R$ .

O valor de  $r_L - r_R$  pode ser calculado por meios geométricos, conforme

$$r_L - r_R = \frac{d}{2} \text{sen } \theta. \quad (3.19)$$

Fazendo a substituição da Equação (3.19) na Equação (3.18), temos a equação que fornece o valor da ITD para o caso da fonte sonora no plano horizontal, considerando a difração provocada pela presença da cabeça:

$$\text{ITD} = \frac{d}{2c} (\theta + \text{sen } \theta). \quad (3.20)$$

A partir da fórmula bidimensional dada pela Equação (3.20), pode-se generalizar para três dimensões por meios geométricos, num modelo esférico de cabeça. O resultado, apresentado em [26], é dado por

$$\text{ITD} = \frac{d}{2} \frac{\arcsen(\text{sen } \theta \cos \phi) + \text{sen } \theta \cos \phi}{c}. \quad (3.21)$$

### 3.3.3 Utilização de Geometria Analítica

É possível tornar os modelos geométricos mais precisos, bastando para tal utilizar geometria analítica. Sabendo-se a posição do ouvinte e das caixas acústicas, pode-se calcular exatamente a distância percorrida pelo som na trajetória até as orelhas da pessoa. Dessa forma, não se faz a hipótese de fonte sonora distante do ouvinte.

Por outro lado, quanto mais preciso o modelo utilizado, mais informações são necessárias para implementá-lo. Seria necessário saber exatamente a posição da cabeça do ouvinte, o que é possível através de um sistema de rastreamento da cabeça, o que pode ser difícil de se ter na prática.

## 3.4 Comparação dos Resultados

A fim de comparar os resultados obtidos entre os modelos teóricos e as medidas empíricas, plotou-se, para cada conjunto considerado, os valores de ITDs medidos e calculados. Para cada gráfico, preservou-se a elevação e variou-se o azimute.

Para se estimar a ITD a partir de modelos geométricos, adotou-se para o diâmetro da cabeça  $d = 18$  cm e para a velocidade do som  $c = 334$  m/s.

Fez-se a medida do erro máximo e do erro médio quadrático (MSE), como parâmetros para auxiliar a avaliação dos modelos geométricos.

O erro máximo foi calculado como o valor absoluto da diferença entre a ITD dada pelo modelo e a ITD obtida do conjunto de medidas, para cada coordenada da grade de medidas dos conjuntos de HRTFs.

Calculou-se o quadrado da diferença entre a ITD do modelo e a ITD obtida do conjunto de medidas, para cada coordenada da grade de medidas dos conjuntos de HRTFs. A seguir, estimou-se o MSE como a média aritmética dos valores quadráticos, variando o azimute para uma elevação fixa.

Alguns dos resultados do conjunto de Gardner são comparados os modelos geométricos na Figura 3.9.

Os valores de erro máximo e MSE para cada elevação do conjunto de Gardner estão descritos na Tabela 3.2.

A Figura 3.10 apresenta de forma gráfica os dados da Tabela 3.2. Para o

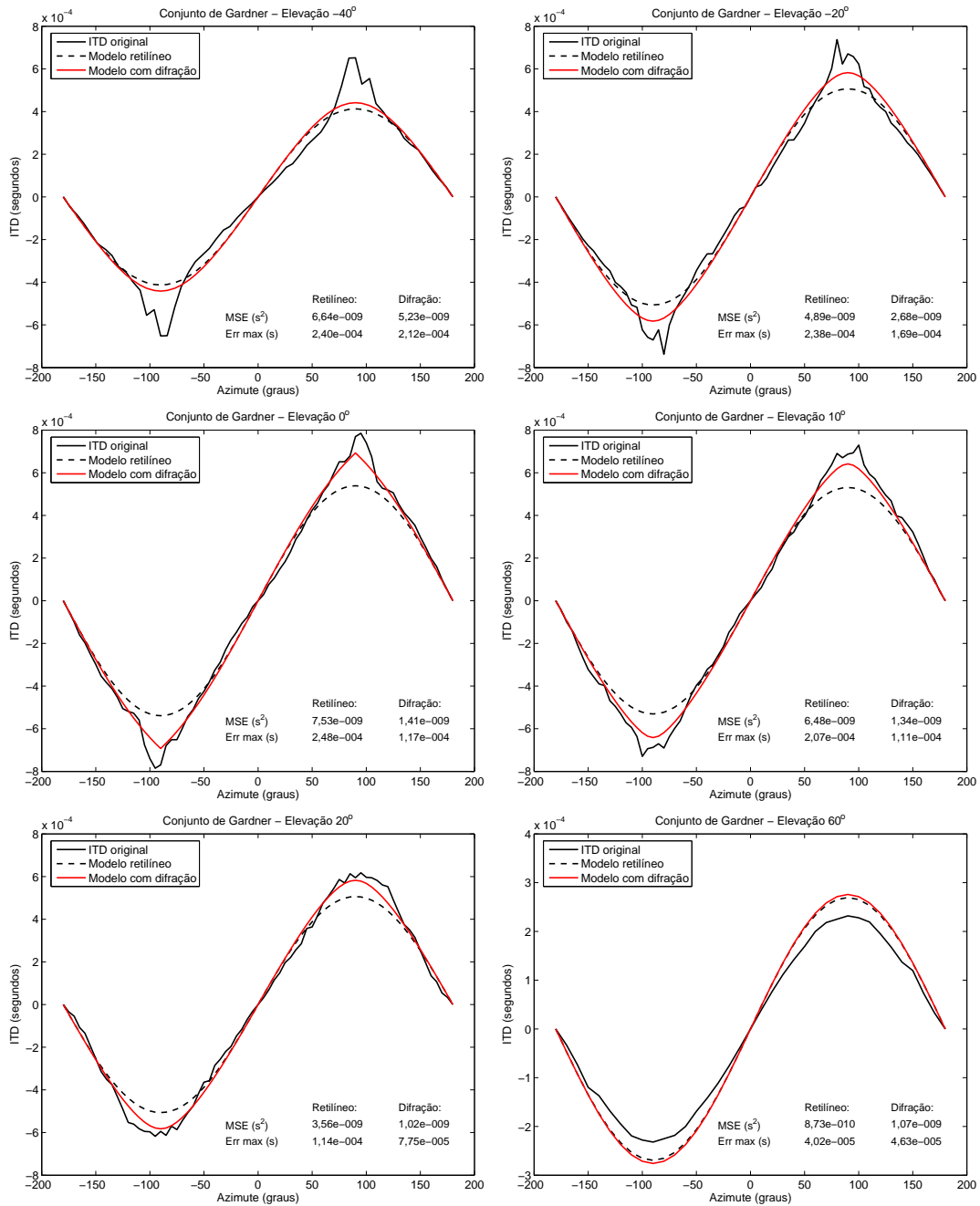


Figura 3.9: Comparações entre as ITDs medida e calculada para o conjunto de Gardner.

conjunto de Gardner, o modelo com difração apresentou erro máximo e MSE menores do que o modelo retilíneo, na maioria das situações. Mesmo quando o modelo retilíneo apresentou erros menores, a diferença foi pequena, conforme a Figura 3.10.

O erro máximo ocorre principalmente nos azimutes  $-90^\circ$  e  $90^\circ$ , onde o comportamento da ITD do conjunto de medida é mais irregular do que nos demais

Tabela 3.2: Erros dos modelos de ITD para o conjunto de Gardner.

Elevação (graus)	Modelo retilíneo		Modelo com difração	
	Erro máximo (s)	MSE ( $s^2$ )	Erro máximo (s)	MSE ( $s^2$ )
-40	2,40e-004	6,64e-009	2,12e-004	5,23e-009
-30	2,09e-004	3,67e-009	1,60e-004	2,55e-009
-20	2,38e-004	4,89e-009	1,69e-004	2,68e-009
-10	2,05e-004	5,11e-009	1,09e-004	1,77e-009
0	2,48e-004	7,53e-009	1,17e-004	1,41e-009
10	2,07e-004	6,48e-009	1,11e-004	1,34e-009
20	1,14e-004	3,56e-009	7,75e-005	1,02e-009
30	5,62e-005	7,32e-010	4,59e-005	3,37e-010
40	4,11e-005	2,08e-010	3,54e-005	2,74e-010
50	2,58e-005	1,59e-010	2,38e-005	1,72e-010
60	4,02e-005	8,73e-010	4,63e-005	1,07e-009
70	1,34e-005	7,68e-011	1,42e-005	9,43e-011
80	2,17e-005	1,68e-010	2,19e-005	1,71e-010

azimutes, sendo, portanto, difícil de se modelar.

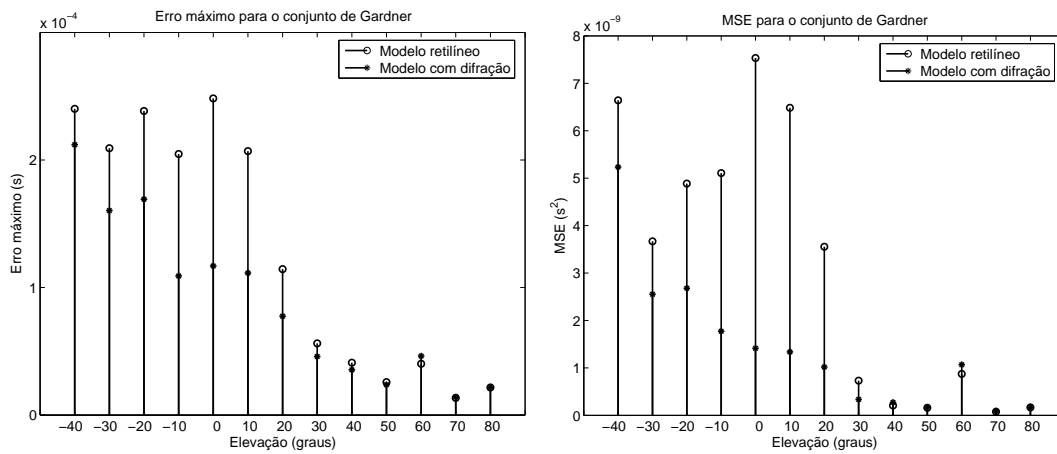


Figura 3.10: Erros dos modelos de ITD para o conjunto de Gardner.

Os modelos geométricos são comparados ao conjunto do Projeto Listen na Figura 3.11.

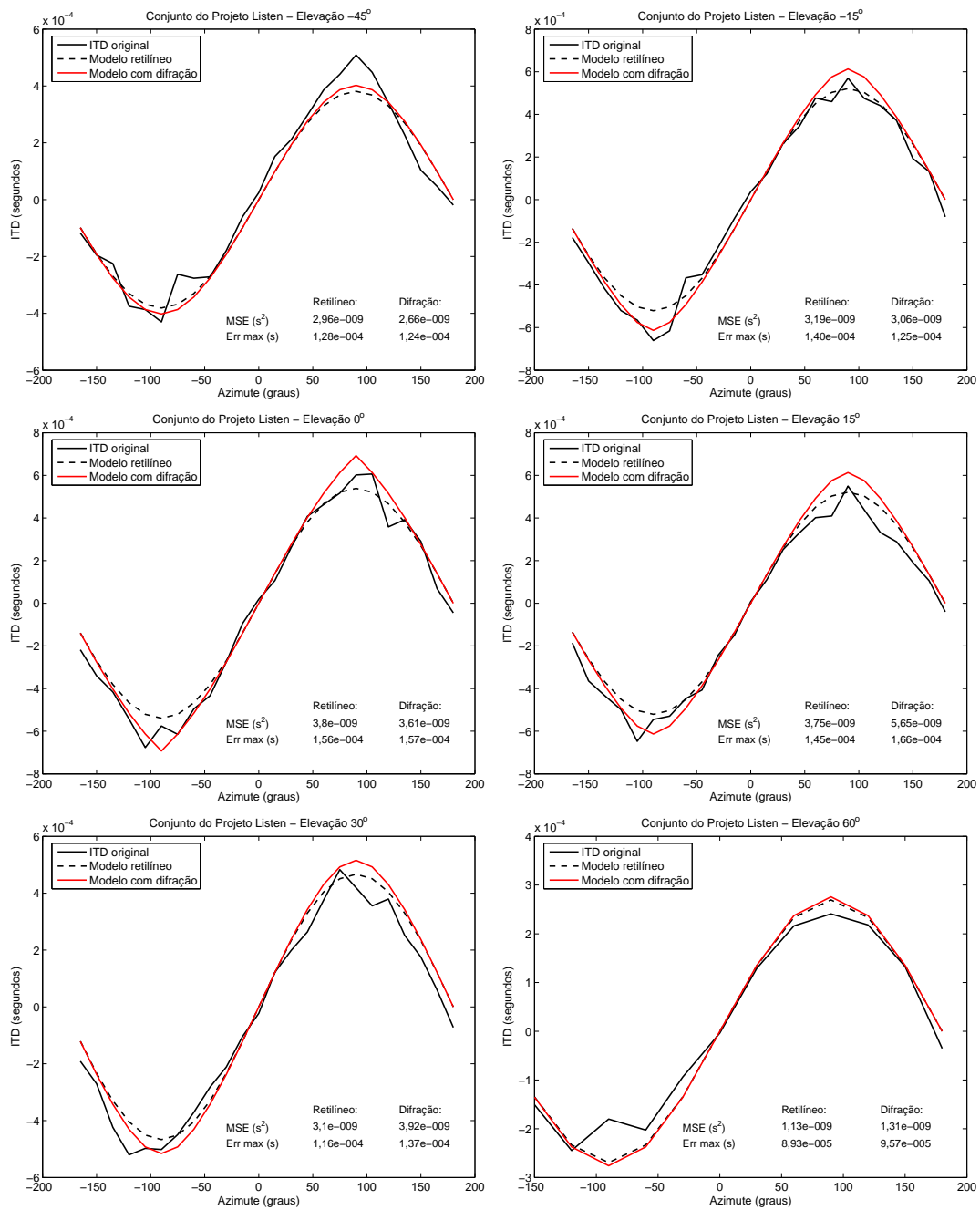


Figura 3.11: Comparações entre as ITDs medida e calculada para o conjunto do Projeto Listen.

Os valores de erro máximo e MSE para cada elevação do conjunto do Projeto Listen estão descritos na Tabela 3.3 e na Figura 3.12.

Tabela 3.3: Erros dos modelos de ITD para o conjunto do Projeto Listen.

Elevação (graus)	Modelo retilíneo		Modelo com difração	
	Erro máximo (s)	MSE (s <sup>2</sup> )	Erro máximo (s)	MSE (s <sup>2</sup> )
-45	1,28e-004	2,96e-009	1,24e-004	2,66e-009
-30	1,85e-004	6,14e-009	1,56e-004	4,81e-009
-15	1,40e-004	3,19e-009	1,25e-004	3,06e-009
0	1,56e-004	3,80e-009	1,57e-004	3,61e-009
15	1,45e-004	3,75e-009	1,66e-004	5,65e-009
30	1,16e-004	3,10e-009	1,37e-004	3,92e-009
45	1,08e-004	2,11e-009	1,21e-004	2,72e-009
60	8,93e-005	1,13e-009	9,57e-005	1,31e-009
75	5,05e-005	5,05e-010	5,10e-005	5,11e-010

No caso do Projeto Listen, não há claramente um modelo geométrico que tenha menores erros máximos e MSE. Para algumas elevações o modelo retilíneo apresentou menores erros, e para outros, o modelo com difração, conforme a Figura 3.12. Novamente, os erros máximos ocorreram em torno dos azimutes 90° e -90°.

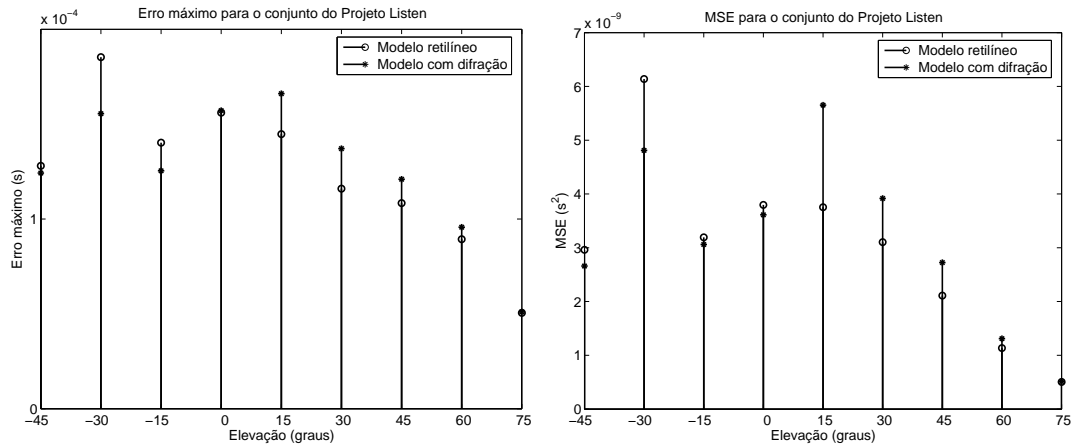


Figura 3.12: Erros dos modelos de ITD para o conjunto do Projeto Listen.

Os resultados do conjunto da CIPIC são comparados aos modelos geométricos na Figura 3.13.

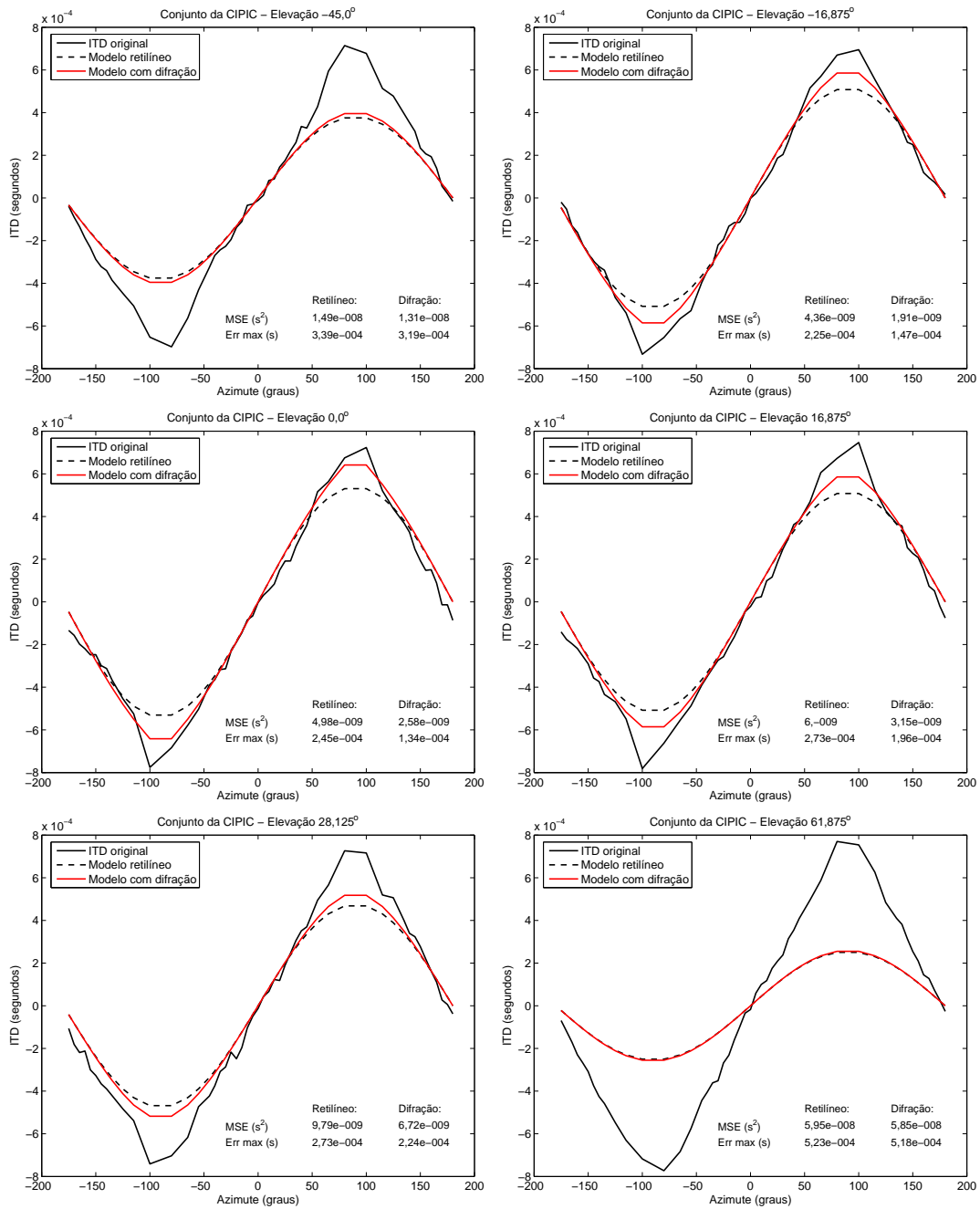


Figura 3.13: Comparações entre as ITDs medida e calculada para o conjunto da CIPIC.

Os valores de erro máximo e MSE para cada elevação do conjunto da CIPIC estão descritos na Tabela 3.4 e na Figura 3.14. Para o conjunto da CIPIC, o modelo

Tabela 3.4: Erros dos modelos de ITD para o conjunto da CIPIC.

Elevação (graus)	Modelo retilíneo		Modelo com difração	
	Erro máximo (s)	MSE (s <sup>2</sup> )	Erro máximo (s)	MSE (s <sup>2</sup> )
-45	3,39e-004	1,49e-008	3,19e-004	1,31e-008
-39	2,91e-004	9,92e-009	2,63e-004	7,92e-009
-34	2,63e-004	7,66e-009	2,25e-004	5,43e-009
-28	2,62e-004	5,73e-009	2,13e-004	3,49e-009
-23	2,33e-004	4,76e-009	1,71e-004	2,52e-009
-17	2,25e-004	4,36e-009	1,47e-004	1,91e-009
-11	2,14e-004	3,45e-009	1,22e-004	1,12e-009
-6	2,37e-004	3,59e-009	1,32e-004	1,40e-009
0	2,45e-004	4,98e-009	1,34e-004	2,58e-009
6	2,14e-004	5,25e-009	1,28e-004	3,04e-009
11	2,29e-004	5,58e-009	1,37e-004	3,20e-009
17	2,73e-004	6,00e-009	1,96e-004	3,15e-009
23	2,74e-004	7,63e-009	2,11e-004	4,57e-009
28	2,73e-004	9,79e-009	2,24e-004	6,72e-009
34	3,43e-004	1,38e-008	3,05e-004	1,07e-008
39	3,60e-004	2,08e-008	3,32e-004	1,78e-008
45	4,15e-004	2,77e-008	3,95e-004	2,52e-008
51	4,57e-004	3,75e-008	4,43e-004	3,55e-008
56	4,97e-004	4,80e-008	4,89e-004	4,66e-008
62	5,23e-004	5,95e-008	5,18e-004	5,85e-008
68	5,72e-004	7,42e-008	5,69e-004	7,36e-008
73	6,01e-004	8,99e-008	6,00e-004	8,96e-008
79	6,67e-004	1,09e-007	6,67e-004	1,09e-007
84	7,05e-004	1,29e-007	7,05e-004	1,29e-007

com difração apresentou erros máximos e MSE menores ou iguais aos do modelo retilíneo. Entretanto, analisando a Figura 3.13 e a Tabela 3.4, nota-se que os erros máximos da estimativa de ITD para este conjunto são consideráveis, principalmente para elevações mais distantes do plano horizontal.

Um comportamento geral a ser descrito é que os modelos de ITD retilíneo e com difração apresentam resultados quase idênticos para elevações acima de 40°, porém comportamentos diferentes para elevações próximas ao plano horizontal, principalmente nas proximidades dos pontos de azimute próximos a 90° e -90°.

Todas as curvas apresentaram uma forma semelhante a uma senóide. Isto ocorre no azimute 90°, tem-se o máximo de ITD com relação à diferença de tempo entre a orelha esquerda e a direita, enquanto o inverso ocorre no azimute -90°



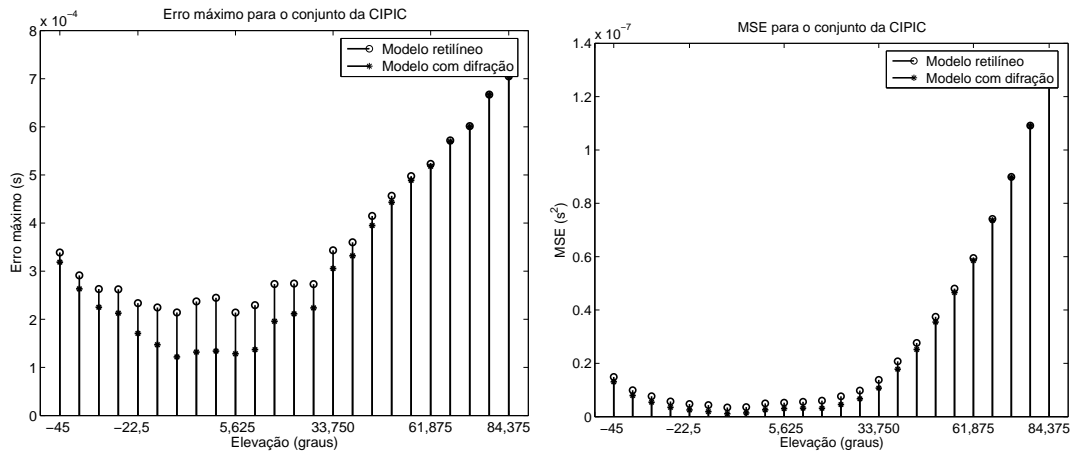


Figura 3.14: Erros dos modelos de ITD para o conjunto da CIPIC.

quando se varia gradativamente a ITD entre esses valores.

A partir do obtido e exposto nas figuras e tabelas analisadas, nota-se que o modelo de ITD com a difração da cabeça apresenta um resultado melhor do que o modelo retilíneo, para o conjunto de Gardner e da CIPIC. Para o conjunto do Projeto Listen, nenhum dos dois modelos descritos apresenta melhor comportamento.

Os valores de MSE descritos nas Tabelas 3.2, 3.3 e 3.4 não podem ser comparados entre si diretamente, porque a grade de medidas dos conjuntos é diferente e os erros foram calculados em cada ponto da mesma.

Um parâmetro que pode ser modificado para melhor adequar um modelo geométrico a um conjunto de medidas é o diâmetro da cabeça considerada nos cálculos.

Os resultados obtidos demonstram que os modelos geométricos apresentados podem ajudar a descrever o comportamento da ITD, principalmente com relação aos conjuntos de Gardner e do Projeto Listen, sendo necessário um cuidado maior com o conjunto da CIPIC.

Nota-se que este capítulo não discute a superioridade de qualquer dos três conjuntos de HRTFS, mas apenas a viabilidade de sua utilização com os modelos geométricos apresentados.

## 3.5 Aplicação em interpolação de HRIRs

As HRIRs obtidas através de medições contemplam apenas alguns pontos discretos no espaço, devido à impossibilidade prática de se medir a resposta para toda e qualquer posição. Para descrever as HRIRs numa posição qualquer, ou para simular uma condição de movimento tomando como base as HRIRs existentes, há diversos métodos de interpolação. Para se fazer a interpolação de HRIRs, deve-se obter a versão de fase mínima das mesmas, para que a interpolação não seja feita sobre versões com atrasos diferentes. A interpolação em versões de fase não-mínima pode criar zeros indesejados e erro na resposta da HRTF interpolada [2, 19]. Depois da interpolação da versão de fase mínima, deve-se inserir a ITD.

O atraso decorrente da ITD, por outro lado, também é medido em alguns pontos discretos do espaço, e também necessita de interpolação ou modelagem para as posições desconhecidas. Uma das formas de se saber qual o atraso a ser inserido é fazer uma interpolação para o atraso separadamente da interpolação das HRTFs.

Modelos geométricos, principalmente o que considera difração da cabeça, podem ser utilizados para simular o atraso a ser inserido.

A função obtida através do método geométrico é contínua, o que garante que a transição de um ponto para outro seja suave em termos de variação de ITD, no caso de simulação de movimento da fonte virtual.

## 3.6 Conclusões

Viu-se neste capítulo uma forma de se obter conjuntos de HRTFs da fase mínima a partir dos conjuntos originais de HRTFs. Empregou-se a técnica nos conjuntos de medidas experimentais do Capítulo 2. A seguir, modelou-se uma reta que melhor representava a diferença de fase entre os conjuntos originais e os de fase mínima. Tal reta no domínio da frequência equivale a um atraso puro no domínio do tempo. O atraso pode ser obtido dos lados esquerdo e direito do indivíduo. Daí, obtém-se a diferença de tempo interaural, ITD, para cada conjunto e para cada posição de medida.

A seguir, mostrou-se a dedução de modelos de ITD obtidos através de aproximações geométricas da cabeça humana, tendo como hipótese que a ITD independe

da frequência. Viu-se também que, quanto mais preciso é o modelo, maior é a necessidade de informação sobre a posição do ouvinte e, portanto, mais elaborado deve ser o sistema de geração de áudio 3D.

Fez-se a comparação dos resultados obtidos através da extração da ITD dos conjuntos originais e dos modelos geométricos, de onde se concluiu que estes podem ajudar a descrever o comportamento da ITD, principalmente com relação aos conjuntos de Gardner e do Projeto Listen. De uma forma geral, os experimentos apresentados demonstram a conformidade dos modelos geométricos com as medidas de ITD extraídas de conjuntos de medidas.

Por fim, mostrou-se uma possível aplicação da teoria e das ferramentas expostas.

#### **Contribuições deste capítulo:**

- Apresentação de comparação entre estimativas de ITD a partir de modelos geométricos teóricos e medidas experimentais.
- Proposta da função mediana como estimador do atraso puro que substitui o excesso de fase (Equação (3.14)).
- Proposta de aplicação dos modelos geométricos em interpolação de HRIRs.

# Capítulo 4

## Estrutura Farrow Generalizada

### 4.1 Introdução

Estudou-se no Capítulo 2 a comparação, análise e tratamento de diversos conjuntos de HRTFs. Um conjunto de HRTFs é utilizado principalmente dentro do contexto da reprodução do áudio 3D através de fones de ouvido.

As medidas de HRTFs são feitas em posições discretas no espaço. A fim de se simular movimento da fonte virtual, ou se faz uma transição brusca entre as medidas ou se faz uma interpolação das HRTFs para as posições faltantes.

Neste capítulo é discutido um método de interpolação de HRTFs baseado em parametrização direta por variáveis espaciais. Ele foi previamente apresentado em [27].

O método apresentado faz uma aproximação polinomial no domínio da frequência de um conjunto de medidas de HRTFs. As suas vantagens incluem o fato de que o número de coeficientes de polinômios a serem armazenados é menor do que o conjunto original, resultando em menor exigência de memória para armazenamento, além de que a função polinomial é contínua, o que assegura a transição suave do movimento de uma posição de medida para outra.

Primeiramente é apresentada a formulação do modelo. A seguir, é apresentada uma reformulação eficiente da otimização do método descrito. Depois são analisados os resultados da aplicação da estrutura apresentada, quanto a três aspectos: comparação da reformulação eficiente contra a formulação anterior, comparação da otimização numa região da esfera contra a otimização na esfera toda, e validação

do modelo para um dos conjuntos de medidas de HRTFs apresentados no Capítulo 2.

A seguir, generaliza-se a formulação para suportar uma terceira variável, o raio da esfera  $r$ , para utilização em conjuntos de HRTFs em que a medição tenha sido feita com um raio variável.

A Figura 4.1 ilustra os temas a serem abordados.



Figura 4.1: Conteúdo do capítulo sobre a estrutura Farrow.

## Observação sobre a notação utilizada

Neste capítulo a notação será modificada, a fim de torná-la mais compacta.

A variável de tempo será representada em subscrito e as demais variáveis serão representadas entre parênteses. Assim, o  $n$ -ésimo coeficiente da HRIR de uma posição  $(\theta, \phi)$  será representado por  $h_n(\theta, \phi)$ .

Os índices dos coeficientes do polinômio serão representados em sobrescrito. Assim,  $C(\alpha) = \sum_{l=0}^L c^l \alpha^l$  representa um polinômio  $C$  na variável  $\alpha$  e com coeficientes  $c^l$ .

## 4.2 Formulação

### 4.2.1 Formulação Original

Em [28], foi proposto um método de parametrização direta das HRTFs pelas variáveis angulares de posição, através da utilização de polinômios. Este método, doravante denominado estrutura Farrow generalizada para interpolação de HRTFs, generaliza para duas variáveis a estrutura polinomial de atraso digital continuamente variável (CVDD) desenvolvida por Farrow [29].

A estrutura Farrow generalizada tem como parâmetros de entrada as variáveis de posição, que modificam a resposta em frequência da estrutura. O sinal monoaural

é filtrado por essa resposta, criando a ilusão de posicionamento.

A função polinomial apresenta a característica de ser uma função analítica, que tem a propriedade de ser contínua e ter todas as suas derivadas contínuas em todos os pontos. Tal característica garante que o movimento da fonte virtual ocorra de maneira suave. Quanto maior for a ordem do polinômio, maior é a semelhança da função interpolada com a função original, porém maior a quantidade de memória requerida para armazenamento dos coeficientes do polinômio.

A estrutura CVDD simula um atraso fracionário  $\tau$  através de um polinômio em  $\alpha$ , sendo  $\alpha = \tau T$ , onde  $-\frac{1}{2} < \alpha < \frac{1}{2}$  e  $T$  é o período de amostragem.

A Figura 4.2 mostra o diagrama em blocos da estrutura original, que considera a dependência apenas da variável  $\alpha$ . O símbolo  $\odot$  representa o operador produto interno.

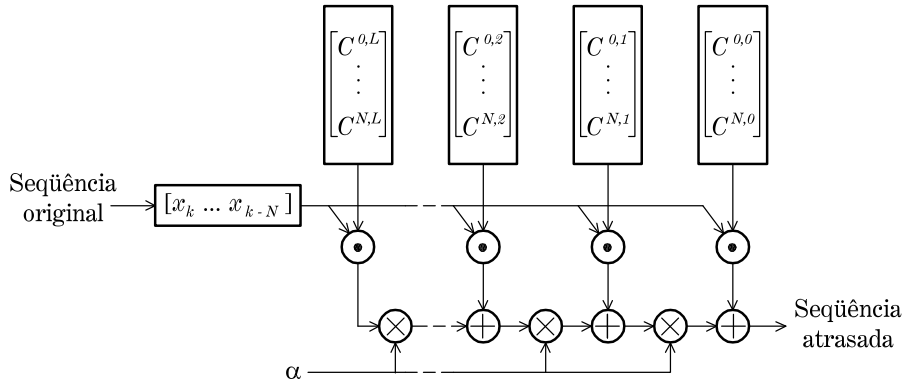


Figura 4.2: Diagrama em blocos da estrutura Farrow original.

Seja a função de transferência do atraso  $\tau$  dada por

$$G(\omega, \tau) = e^{-j\omega\tau}. \quad (4.1)$$

Para estimar  $G(\omega, \tau)$ , utiliza-se um filtro FIR  $F(\omega, \alpha)$  com  $N$  coeficientes  $C_n$  dependentes de  $\alpha$

$$F(\omega, \alpha) = \sum_{n=0}^N C_n(\alpha) e^{-jn\omega T}. \quad (4.2)$$

Estimando cada coeficiente por um polinômio de ordem  $L$  em  $\alpha$ , tem-se

$$C_n(\alpha) = \sum_{l=0}^L c_n^l \alpha^l, \quad (4.3)$$

onde  $c_n^l$  são os coeficientes do polinômio que estima o coeficiente  $C_n$ .

Das Equações (4.2) e (4.3) tem-se

$$F(\omega, \alpha) = \sum_{l=0}^L \alpha^l \sum_{n=0}^N c_n^l e^{-jn\omega T}. \quad (4.4)$$

Os coeficientes  $c_n^l$  podem ser obtidos de forma a minimizar a função-custo

$$J = \int_{-\pi}^{\pi} \int_{-\frac{1}{2}}^{\frac{1}{2}} |F(\omega, \alpha) - G(\omega, \alpha)|^2 d\alpha d\omega. \quad (4.5)$$

## 4.2.2 Formulação Bidimensional

A partir da CVDD, é feita a generalização da estrutura para duas variáveis, para aplicação em HRTFs. A formulação da estrutura Farrow consiste em, a partir das ordens do polinômio em cada variável de interesse, encontrar os coeficientes da estrutura de tal modo que a função polinomial seja a mais próxima possível da resposta de um determinado conjunto de HRTFs. A medida da diferença entre a resposta do conjunto de HRTFs original e a resposta interpolada é dada por uma função-custo. Portanto, os coeficientes do polinômio da estrutura Farrow devem ser escolhidos de forma a minimizar a função-custo. A formulação a seguir deriva de [27, 28].

É importante notar que, para se obter os coeficientes da estrutura, devem-se utilizar versões de fase mínima das HRTFs, sendo os excessos de fase tratados em separado [27, 28], conforme discutido no Capítulo 3.

Considera-se, inicialmente, que a resposta de uma HRIR de ordem  $N$  é definida por

$$G(e^{j\Omega}, \theta, \phi) = \sum_{n=0}^N h_n(\theta, \phi) e^{-j\Omega n}, \quad (4.6)$$

onde  $h_n(\theta, \phi)$  é o  $n$ -ésimo peso da HRIR da posição  $(\theta, \phi)$  e  $\Omega$  é a frequência discreta  $\Omega = \omega T$ .

A estrutura Farrow generalizada aproxima cada peso  $h_n(\theta, \phi)$  por um polinômio de ordem  $L$  em  $\theta$  e ordem  $M$  em  $\phi$ , como em

$$\hat{h}_n(\theta, \phi) \approx \sum_{l=0}^L \sum_{m=0}^M \theta^l \phi^m c_n^{l,m}, \quad (4.7)$$

onde  $c_n^{l,m}$  são os coeficientes a encontrar. A resposta na frequência associada é dada por

$$F(e^{j\Omega}, \theta, \phi) = \sum_{n=0}^N \sum_{l=0}^L \sum_{m=0}^M \theta^l \phi^m c_n^{l,m} e^{-j\Omega n}. \quad (4.8)$$

A Figura 4.3 mostra um diagrama em blocos da estrutura resultante apenas para o canal esquerdo. A saída para o canal direito pode ser obtida simplesmente aplicando  $-\theta$  aos resultados dos produtos internos, aproveitando a simetria da cabeça com relação a  $\theta$ , conforme o sistema de coordenadas descrito na Seção 1.2.

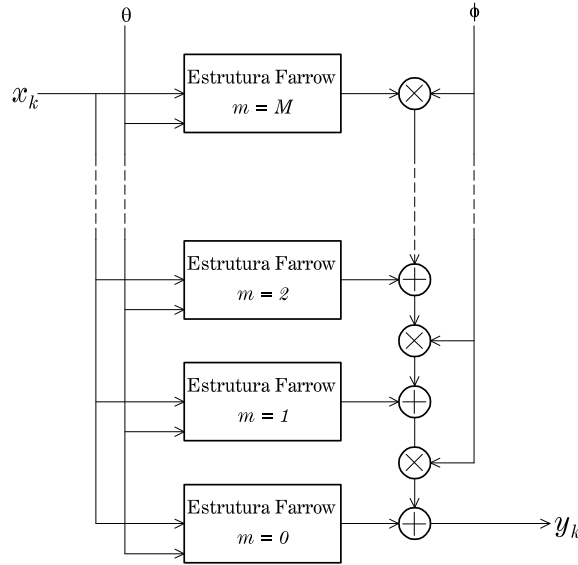


Figura 4.3: Diagrama em blocos da estrutura Farrow generalizada para 2 variáveis espaciais.

Escolhem-se os coeficientes  $c_n^{l,m}$  de tal modo que o erro quadrático da aproximação seja minimizado segundo a função-custo

$$J = \sum_{\Omega} \sum_{\theta} \sum_{\phi} \left| F(e^{j\Omega}, \theta, \phi) - G(e^{j\Omega}, \theta, \phi) \right|^2, \quad (4.9)$$

sendo  $F(e^{j\Omega}, \theta, \phi)$  e  $G(e^{j\Omega}, \theta, \phi)$  dadas pelas Equações (4.6) e (4.8), respectivamente.

### 4.2.3 Otimização Iterativa no Domínio da Frequência

A primeira idéia para obtenção dos coeficientes  $c_n^{l,m}$  consistia em realizar a otimização iterativa por um método *quasi*-Newton, conforme abordado em [19].



Considerando que haja HRIRs de ordem  $N$  medidas para  $A$  posições  $(\theta, \phi)$  e que se adotem ordens  $L$  e  $M$  para cada polinômio (ver Equação (4.7)), a otimização dos  $B = (N + 1)(L + 1)(M + 1)$  coeficientes envolve o cálculo de um vetor-gradiente com  $B$  elementos e uma matriz hessiana com  $B^2$  elementos. Para o conjunto de HRTFs de Gardner [14],  $A = 726$  e  $N = 127$ . Se  $L = M = 6$ , por exemplo, o problema consiste em calcular conjuntamente 6.252 coeficientes; a hessiana tem cerca de  $40 \times 10^6$  elementos e um cálculo de  $J$  envolve mais que  $600 \times 10^6$  multiplicações. Isso dá uma idéia dos custos computacional e de memória extremamente elevados inerentes a essa formulação, tornando-a muito suscetível a problemas numéricos, além de resultar em tempos de processamento exagerados.

Através da otimização iterativa no domínio da frequência, não foi possível utilizar estruturas com ordens maiores que  $L = 6$  e  $M = 6$ ; e ainda assim, foi preciso modelar a esfera em sub-regiões convenientemente ligadas. Isso levou à busca de alternativas, descritas na próxima subseção.

#### 4.2.4 Reformulação da Otimização

É possível reformular a função a otimizar, transformando o problema original de ordem elevada em diversos problemas de ordens menores.

Forçando que a ordem das HRIRs estimadas seja igual à ordem  $N$  das HRIRs medidas, pode-se reescrever a Equação (4.9) combinada com as Equações (4.6) e (4.8) na forma

$$J = \sum_{\Omega} \sum_{\theta} \sum_{\phi} \left| \sum_n \left( \sum_l \sum_m \theta^l \phi^m c_n^{l,m} - h_n(\theta, \phi) \right) e^{-j\Omega n} \right|^2. \quad (4.10)$$

Nota-se que a função-custo (4.9) calcula a energia do erro no domínio da frequência. Pelo Teorema de Parseval, a energia no domínio da frequência tem a sua equivalente no domínio do tempo.

Definindo a DFT como

$$X(\Omega) = \sum_n x_n e^{-j\Omega n}. \quad (4.11)$$

O Teorema de Parseval é dado por

$$\sum_n |x_n|^2 = \frac{1}{N+1} \sum_{\Omega} |X(\Omega)|^2. \quad (4.12)$$

Aplicando a substituição de variáveis

$$x_n = \sum_l \sum_m \theta^l \phi^m c_n^{l,m} - h_n(\theta, \phi) \quad (4.13)$$

na Equação (4.10), chega-se a

$$J = \sum_{\Omega} \sum_{\theta} \sum_{\phi} \left| \sum_n x_n e^{-j\Omega n} \right|^2. \quad (4.14)$$

Nota-se que para  $\theta$  e  $\phi$  fixos, a Equação (4.13) depende apenas de  $n$ .

Aplicando a definição da DFT (Equação (4.11)) na Equação (4.14), chega-se a

$$J = \sum_{\theta} \sum_{\phi} \sum_{\Omega} \left| X(\Omega) \right|^2. \quad (4.15)$$

Aplicando a definição do Teorema de Parseval, Equação (4.12), chega-se a

$$J = (N + 1) \sum_{\theta} \sum_{\phi} \sum_n \left| x_n \right|^2. \quad (4.16)$$

Retirando a substituição de variáveis dada pela Equação (4.13), e notando que, pelo fato de os coeficientes das HRIRs serem número reais (o módulo ao quadrado equivale ao quadrado), a Equação (4.16) torna-se

$$J = (N + 1) \sum_n \sum_{\theta} \sum_{\phi} \left( \sum_l \sum_m \theta^l \phi^m c_n^{l,m} - h_n(\theta, \phi) \right)^2. \quad (4.17)$$

Assim, para cada  $n$ , tem-se uma parcela quadrática da função  $J$ , conforme

$$J = (N + 1)(J_0 + J_1 + \dots + J_N), \quad (4.18)$$

onde

$$J_n = \sum_{\theta} \sum_{\phi} \left( \sum_l \sum_m \theta^l \phi^m c_n^{l,m} - h_n(\theta, \phi) \right)^2. \quad (4.19)$$

A minimização de cada parcela quadrática leva à minimização da função-custo global.

Cada um dos  $(N + 1)$  sub-problemas envolve, agora, o cálculo de  $B' = (L + 1)(M + 1)$  coeficientes. Para  $A = 726$ ,  $N = 127$  e  $L = M = 6$ , por exemplo, tem-se 49 coeficientes a otimizar e 2.041 elementos na hessiana. Dessa forma, os requisitos de memória são bastante reduzidos e espera-se que a redução de ordem melhore bastante o condicionamento numérico do problema.

## 4.2.5 Otimização Direta no Domínio do Tempo

A redução da complexidade do problema inicial da Equação (4.9) ao ser reescrito nos sub-problemas de minimização da Equação (4.19) já permite que se busque sua solução direta por inversão de matriz.

Fixando-se um dado  $n$ , cada elemento do gradiente da função-custo da Equação (4.19) em relação aos coeficientes é dado por

$$\frac{\partial J_n}{\partial c_n^{l_0, m_0}} = 2 \sum_l \sum_m \left( \sum_\theta \sum_\phi \theta^{l_0} \phi^{m_0} \theta^l \phi^m c_n^{l, m} \right) - 2 \sum_\theta \sum_\phi \theta^{l_0} \phi^{m_0} h_n(\theta, \phi). \quad (4.20)$$

Igualando-se o gradiente a zero, tem-se

$$\sum_l \sum_m \left( \sum_\theta \sum_\phi \theta^{l_0} \phi^{m_0} \theta^l \phi^m c_n^{l, m} \right) = \sum_\theta \sum_\phi \theta^{l_0} \phi^{m_0} h_n(\theta, \phi). \quad (4.21)$$

Fazendo-se  $0 \leq l_0 \leq L$  e  $0 \leq m_0 \leq M$ , chega-se a  $(L+1)(M+1)$  equações em  $(L+1)(M+1)$  incógnitas  $c_n^{l, m}$ . Na forma matricial, pode-se reescrever o problema como  $\mathbf{Q}\mathbf{c} = \mathbf{p}$ , onde os elementos da matriz  $\mathbf{Q}$  são

$$q_{i, j} = \sum_\theta \sum_\phi \theta^{v(i)} \phi^{\mu(i)} \theta^{v(j)} \phi^{\mu(j)}, \quad (4.22)$$

e os dos vetores  $\mathbf{c}$  e  $\mathbf{p}$  são, respectivamente,

$$c_{i, n} = c_n^{v(i), \mu(i)}, \quad (4.23)$$

$$p_{i, n} = \sum_\theta \sum_\phi \theta^{v(i)} \phi^{\mu(i)} h_n(\theta, \phi), \quad (4.24)$$

com  $0 \leq i, j \leq (L+1)(M+1)$  e tendo-se definido  $v(i, j)$  como o quociente e  $\mu(i, j)$  como o resto da divisão inteira  $\frac{i, j}{L+1}$ .

Dessa forma, para encontrar os coeficientes resta resolver

$$\mathbf{c}_n = \mathbf{Q}^{-1} \mathbf{p}_n. \quad (4.25)$$

A matriz  $\mathbf{Q}$  de dimensão  $(L+1)(M+1)$  é simétrica, e sua inversa pode ser obtida por métodos robustos adequados.

A otimização de todos os coeficientes  $c_n^{l, m}$  da estrutura requer, então, que se resolva  $N+1$  vezes o problema de baixa ordem acima descrito, variando-se  $0 \leq n \leq N$ . Mas a matriz  $\mathbf{Q}$ , peça-chave da solução, traz em si outro fator importante para

a simplicidade dessa reformulação: ela independe de  $n$  e, portanto, só precisa ser invertida uma vez; as soluções para cada  $n$  são obtidas trocando-se apenas o vetor  $\mathbf{p}$ .

O procedimento aqui apresentado permitiu a obtenção dos coeficientes da estrutura Farrow generalizada para a esfera toda, e para ordens maiores do que  $L = 6$  e  $M = 6$ .

### 4.3 Validação da Reformulação da Otimização

Nesta seção será feita uma comparação entre os resultados obtidos através do método descrito na Subseção 4.2.3 e a reformulação da Subseção 4.2.5.

Dois aspectos serão analisados: o tempo de processamento de cada solução e uma avaliação objetiva da resposta na frequência.

Foram definidos três caminhos, numa esfera. A Figura 4.4 indica as características de cada caminho, batizados de Caminho Horizontal, Caminho Vertical e Caminho Diagonal.

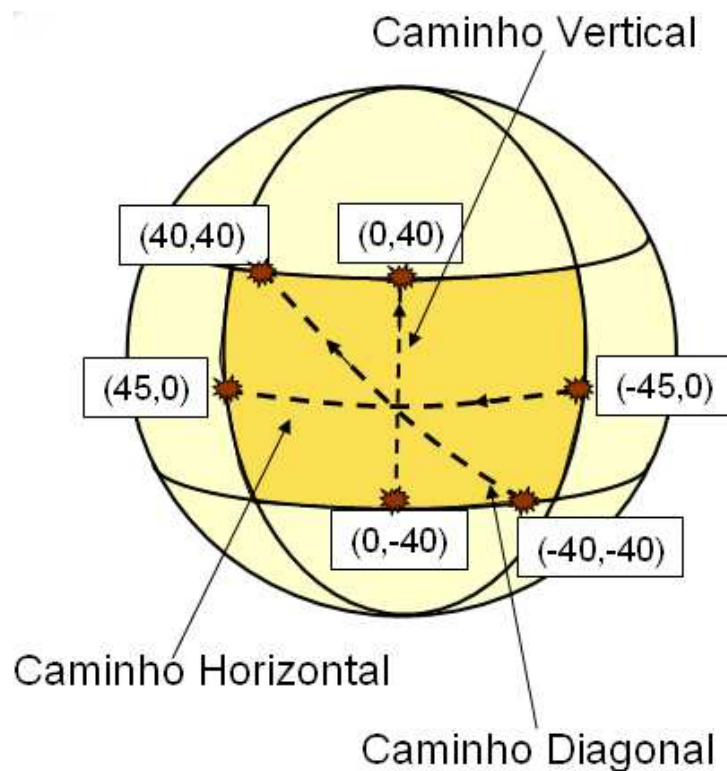


Figura 4.4: Caminhos de teste.

Para cada caminho descrito na Figura 4.4, foram calculadas as HRTFs aproximadas com passos de 1 grau no azimute e na elevação. No caso das HRTFs

de referência, efetuou-se interpolação bilinear [30] em regiões triangulares. Faz-se uma equivalência entre o valor absoluto da HRTF em dB e uma escala de cinza, originando um mapa para cada caminho. Os mapas permitem que se faça uma comparação visual das diversas soluções a serem apresentadas. Vale observar que a estrutura Farrow recalcula as funções nas posições das HRTFs medidas.

Optou-se por trajetórias que pudessem ser parametrizadas de forma simples. Num mesmo gráfico pode-se visualizar a amplitude das HRTFs em dB (níveis de cinza) em todo o espectro nas diversas posições do percurso.

Todos os experimentos desta seção foram realizados sobre o conjunto de HRTFs de Gardner [14].

### 4.3.1 Tempo de Processamento

Para ilustrar a questão da complexidade computacional, a Tabela 4.1 compara os tempos consumidos pela otimização medidos em cada caso, em relação ao de ordem  $L = 5$ ,  $M = 5$  com otimização no tempo na esfera toda, para o conjunto de HRTFs de Gardner.

Tabela 4.1: Tempo de processamento da otimização.

<b>Método de Otimização</b>	<b>Área de Otimização</b>	<b>Ordem Polinomial</b>	<b>Tempo de Processamento</b>
Frequência	Região	$L = 5, M = 5$	$\gg T_0$
Tempo	Região	$L = 5, M = 5$	$\approx T_0/2$
Tempo	Esfera	$L = 5, M = 5$	$T_0$
Tempo	Região	$L = 8, M = 8$	$\approx 2T_0$

Reduzindo-se a região de otimização, que tem aproximadamente 16% das HRIRs do conjunto, houve redução do tempo de processamento à metade. Aumentando-se a ordem polinomial, aumentou-se o tempo de processamento. Ambos os casos consideram o método de otimização no tempo.

Por fim, a otimização na frequência foi ordens de grandeza mais demorada do que qualquer otimização no tempo.

### 4.3.2 Comparações entre as Respostas na Frequência das Soluções

As Figuras 4.5, 4.6 e 4.7 mostram as respostas na frequência para ordens de Estrutura Farrow  $L=5$  e  $M=5$ , para as soluções obtidas no tempo e na frequência, para a região da esfera definida por  $-45^\circ \leq \theta \leq 45^\circ$  e  $-40^\circ \leq \phi \leq 40^\circ$

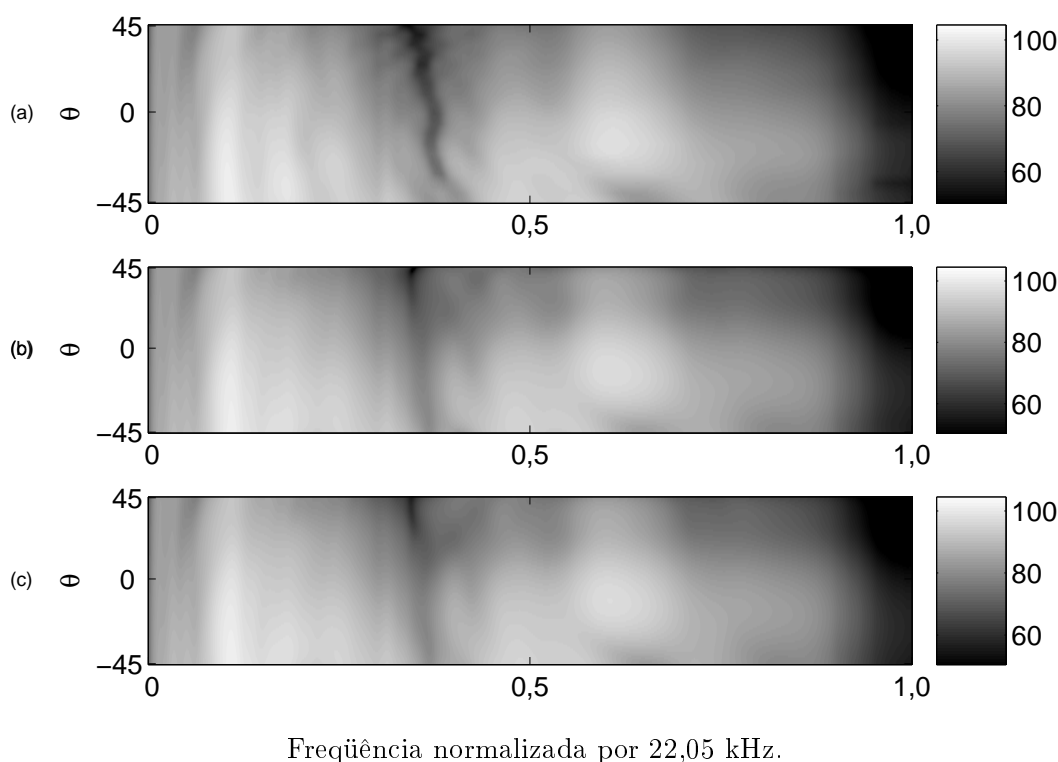


Figura 4.5: Comparação entre formulações para o caminho horizontal. (a) Interpolação bilinear das HRTFs originais. Estrutura Farrow: (b) solução no tempo; (c) solução na frequência.

Primeiramente, pode-se destacar que a continuidade do comportamento das HRTFs é preservada em qualquer solução.

A comparação entre os gráficos (b) e (c) das Figuras 4.5, 4.6 e 4.7 mostra que os resultados obtidos são bastante semelhantes.

A fim de dar uma visão mais detalhada da aproximação realizada sobre as HRTFs, a Figura 4.8 mostra a resposta na frequência das soluções obtidas para algumas coordenadas do espaço.

A comparação das curvas da Figura 4.8 também demonstra que praticamente não há diferença entre os resultados obtidos através da formulação no tempo e na

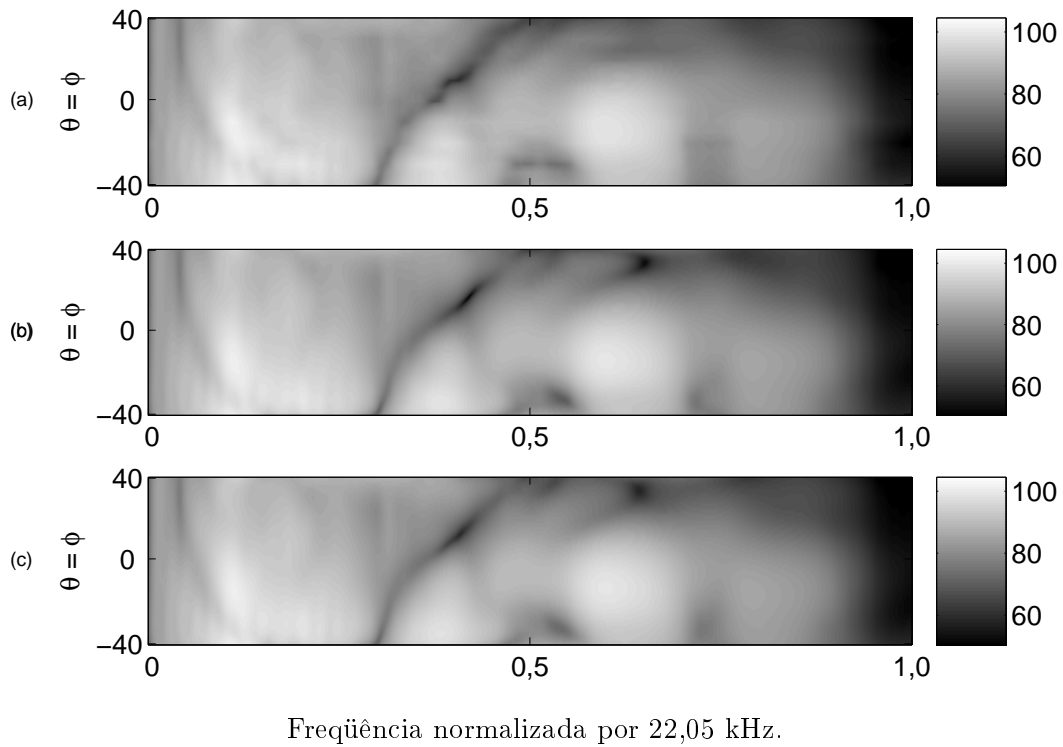
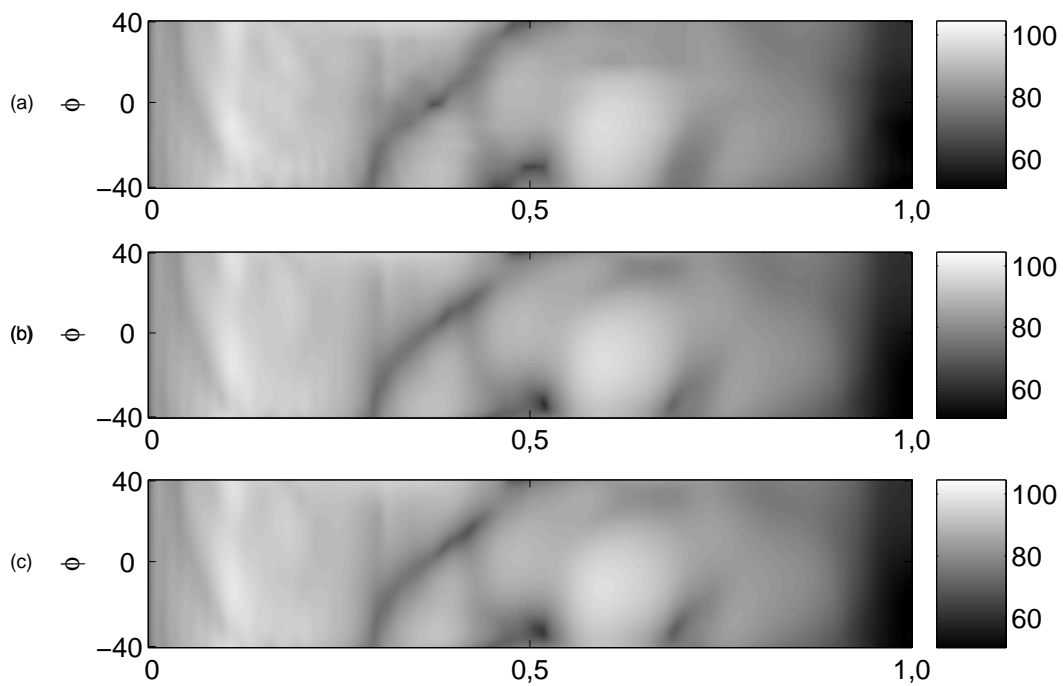


Figura 4.6: Comparação entre formulações para o caminho diagonal. (a) Interpolação bilinear das HRTFs originais. Estrutura Farrow: (b) solução no tempo; (c) solução na frequência.

frequência.

De acordo com as observações anteriores, as soluções de ambos os métodos de cálculo foram muito próximas, dessa forma confirmando a equivalência das otimizações na frequência e no tempo, sendo que a otimização no tempo apresenta a vantagem de ter a menor complexidade de cálculo.



Freqüência normalizada por 22,05 kHz.

Figura 4.7: Comparação entre formulações para o caminho vertical. (a) Interpolação bilinear das HRTFs originais. Estrutura Farrow: (b) solução no tempo; (c) solução na freqüência.



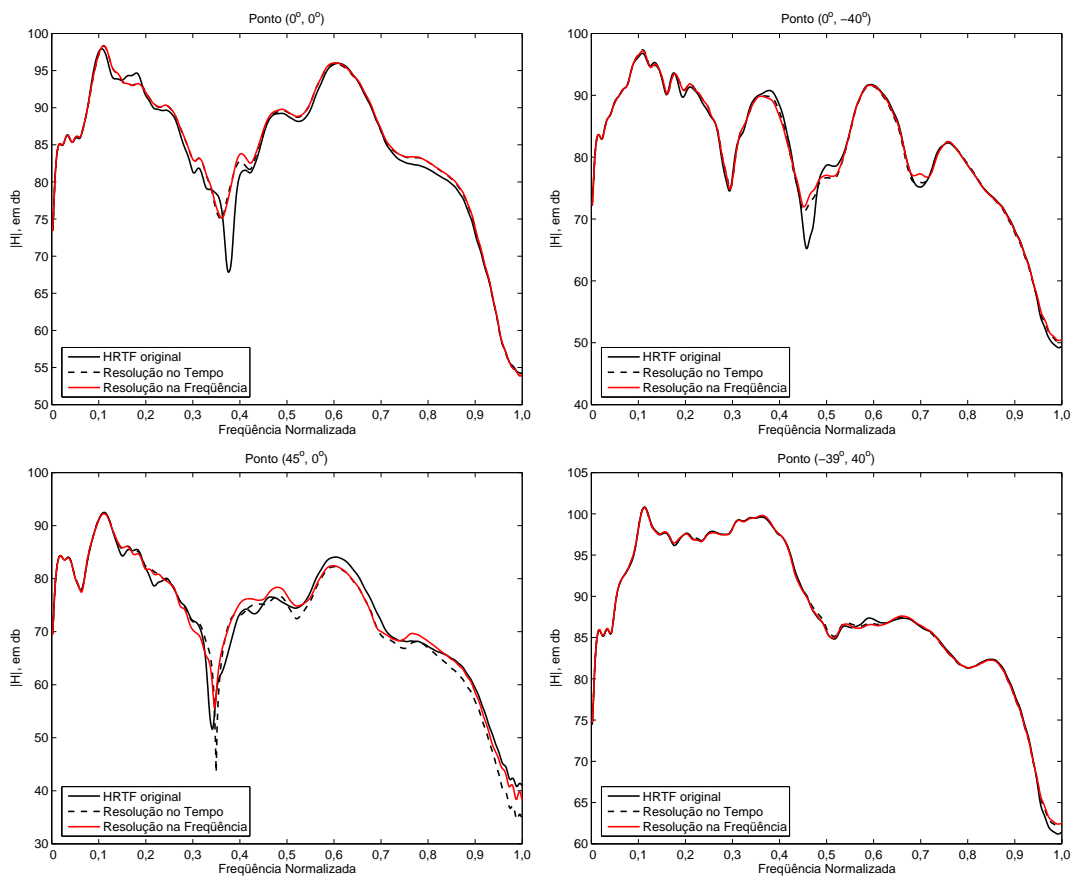


Figura 4.8: Comparação entre as respostas das soluções no tempo e na frequência.

### 4.3.3 Confronto entre Otimização Local e na Esfera Toda

Nesta seção será feita a comparação entre a formulação da estrutura Farrow numa região da esfera, em contraste com a formulação na esfera toda. A região da esfera modelada foi a seção da esfera formada pela faixa de azimutes de  $-45^\circ$  a  $45^\circ$  e elevações de  $-40^\circ$  a  $40^\circ$ . A ordem polinomial foi mantida a mesma para ambos os casos, e novamente foi feita a interpolação bilinear das HRTFs de Gardner [14] submetidas à interpolação bilinear como parâmetro de comparação.

As Figuras 4.9, 4.10 e 4.11 mostram o comportamento ao longo dos caminhos citados na Figura 4.4.

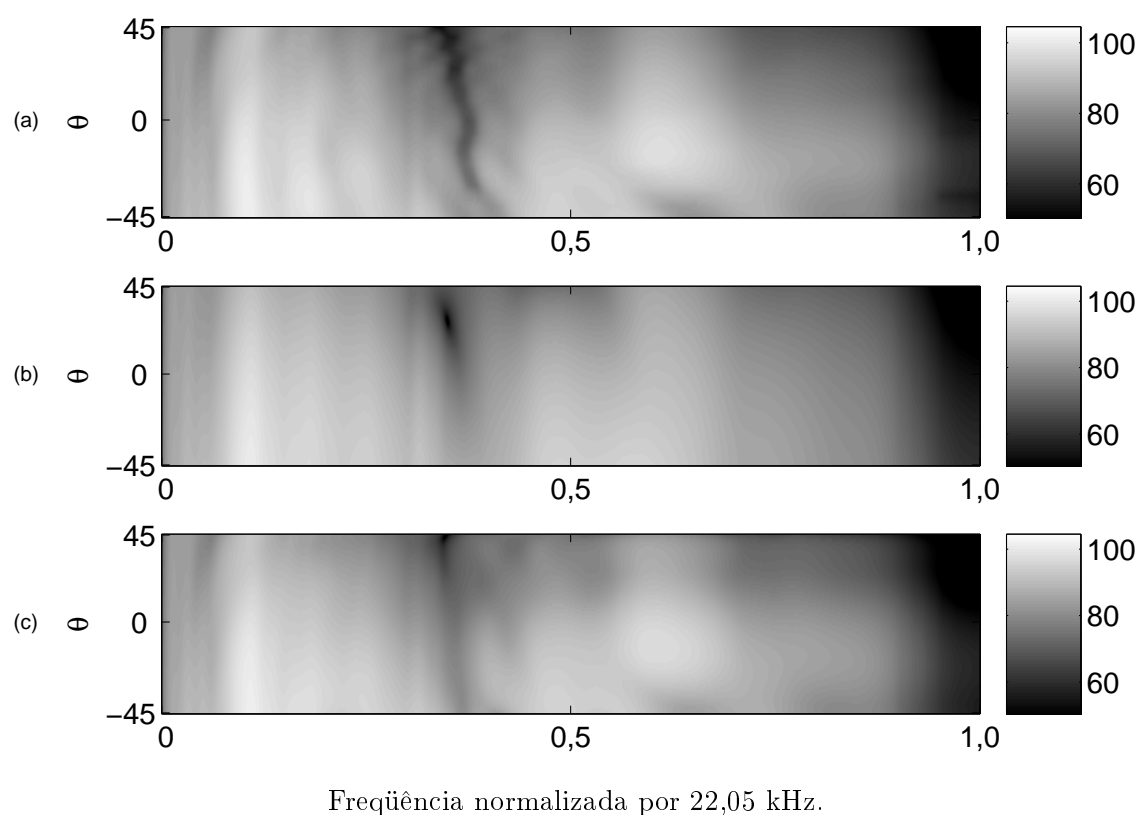


Figura 4.9: Comparação entre a otimização numa região limitada e na esfera toda para o caminho horizontal. (a) Interpolação bilinear das HRTFs originais. Estrutura Farrow; (b) solução na esfera toda; (c) solução numa fração da esfera.

O confronto entre os gráficos (b) e (c) aponta para o fato de que, para uma mesma ordem, comparada à referência (a), a interpolação na esfera toda tende a ser mais suavizada que quando realizada numa região limitada. Nenhuma surpresa aqui, uma vez que a descrição global da esfera exige mais do mesmo número de

coeficientes polinomiais. Assim, é preferível realizar a otimização por regiões, apenas com o cuidado adicional de preservar a continuidade nas transições entre regiões adjacentes [28], por exemplo impondo restrições às fronteiras.

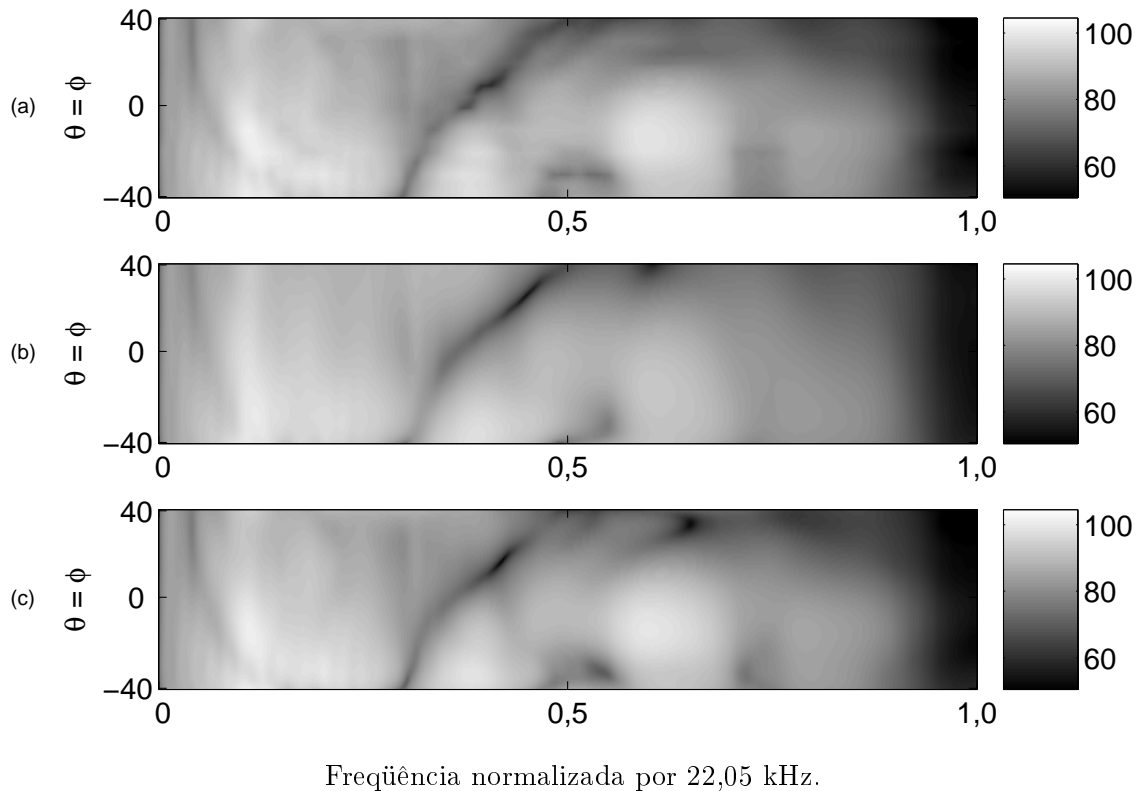
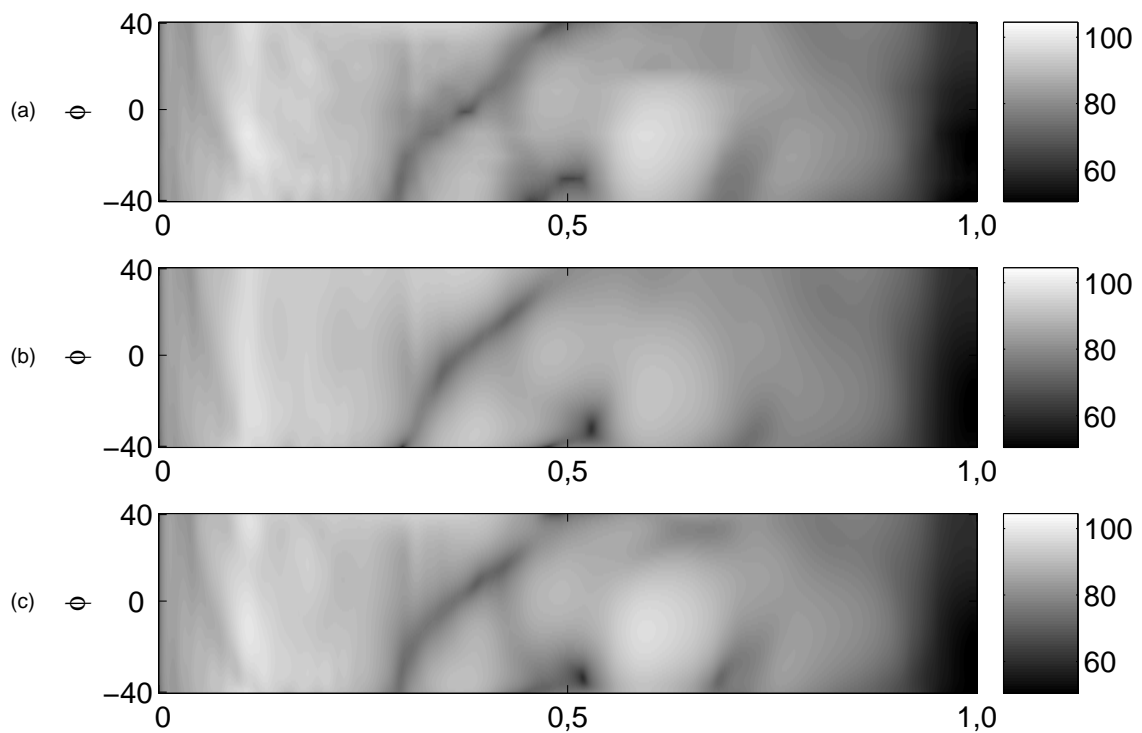


Figura 4.10: Comparação entre a otimização numa região limitada e na esfera toda para o caminho diagonal. (a) Interpolação bilinear das HRTFs originais. Estrutura Farrow; (b) solução na esfera toda; (c) solução numa fração da esfera.

A Figura 4.12 ilustra a resposta na frequência da solução numa região limitada e da solução na esfera toda, para alguns pontos sobre uma esfera.

De acordo com as observações das Figuras 4.9 a 4.12, a solução obtida em apenas uma parte da esfera mostrou resultado superior à modelagem da esfera toda.



Frequência normalizada por 22,05 kHz.

Figura 4.11: Comparação entre a otimização numa região limitada e na esfera toda para o caminho vertical. (a) Interpolação bilinear das HRTFs originais. Estrutura Farrow; (b) solução na esfera toda; (c) solução numa fração da esfera.

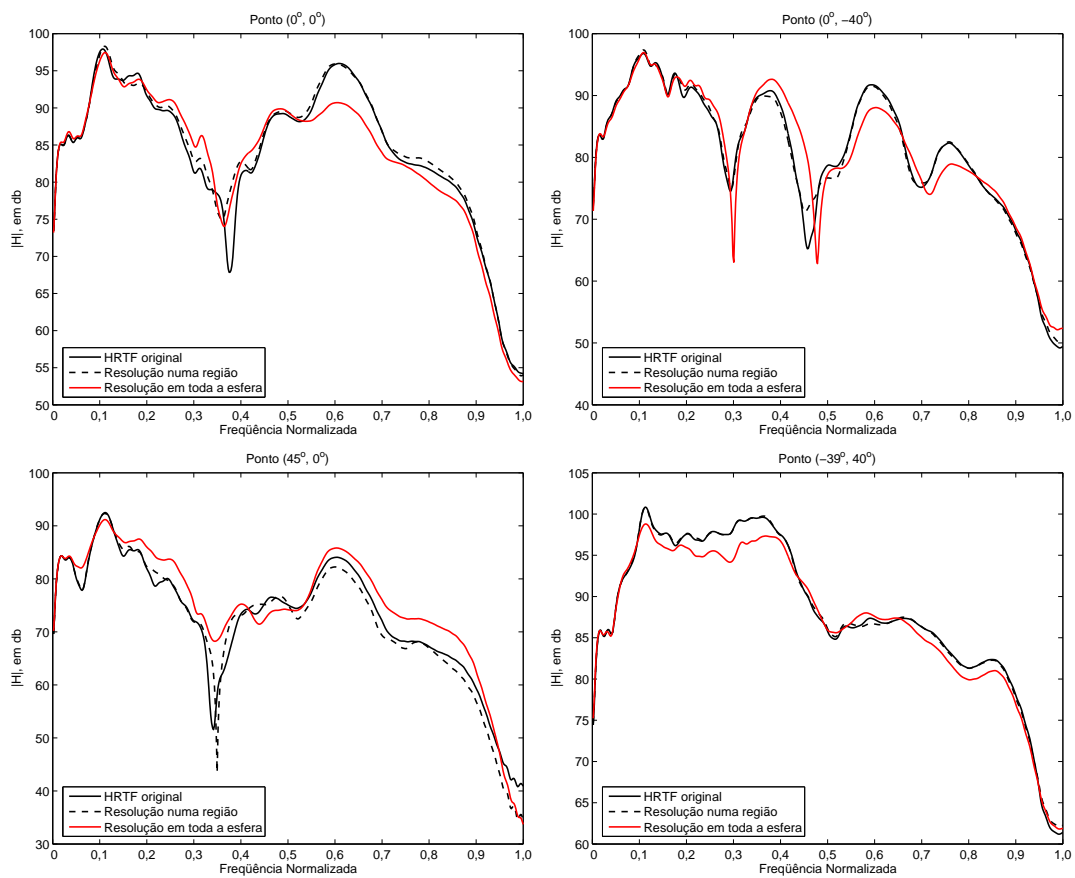


Figura 4.12: Comparação entre as respostas na frequência numa região limitada e na esfera toda.

### 4.3.4 Investigação dos Efeitos da Variação de Ordens do Modelo

Nesta seção serão modeladas uma estrutura de ordem  $L = 5$  e  $M = 5$  e outra de ordem  $L = 8$  e  $M = 8$ , para verificar o comportamento das estruturas com o aumento da ordem polinomial.

As Figuras 4.13 a 4.15 mostram o comportamento da solução obtida sobre o conjunto de Gardner, para os três caminhos da Figura 4.4.

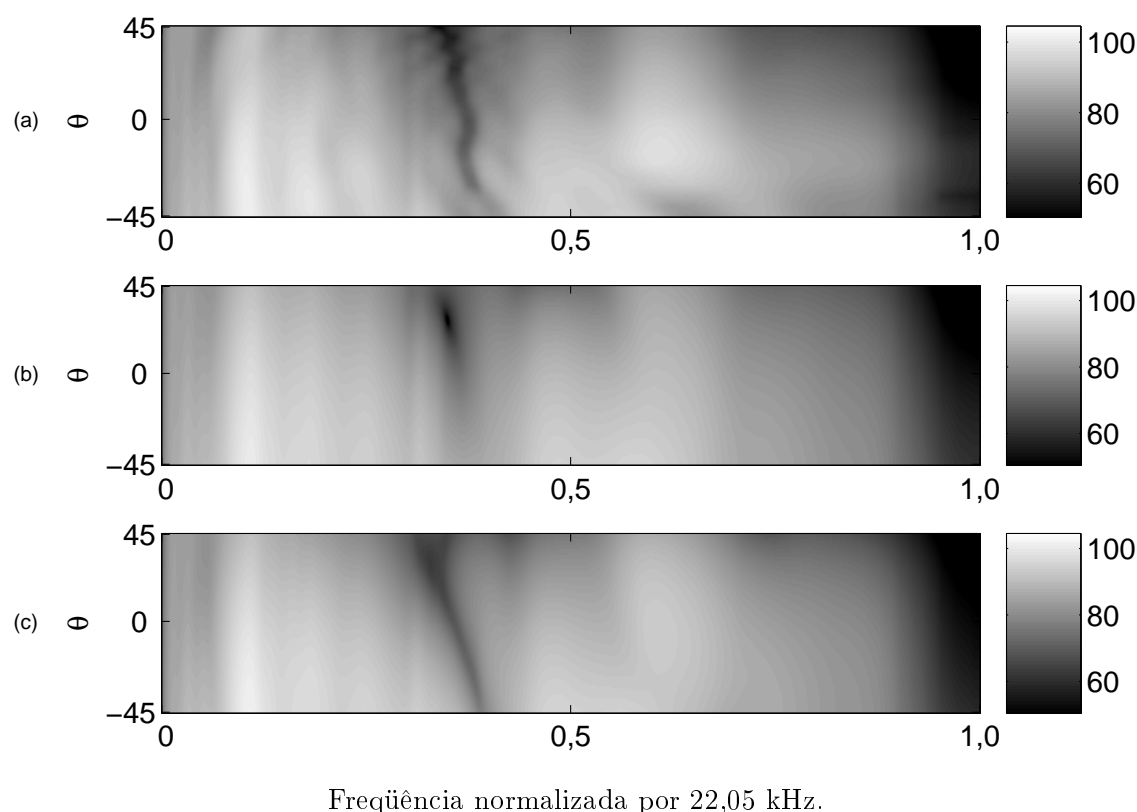
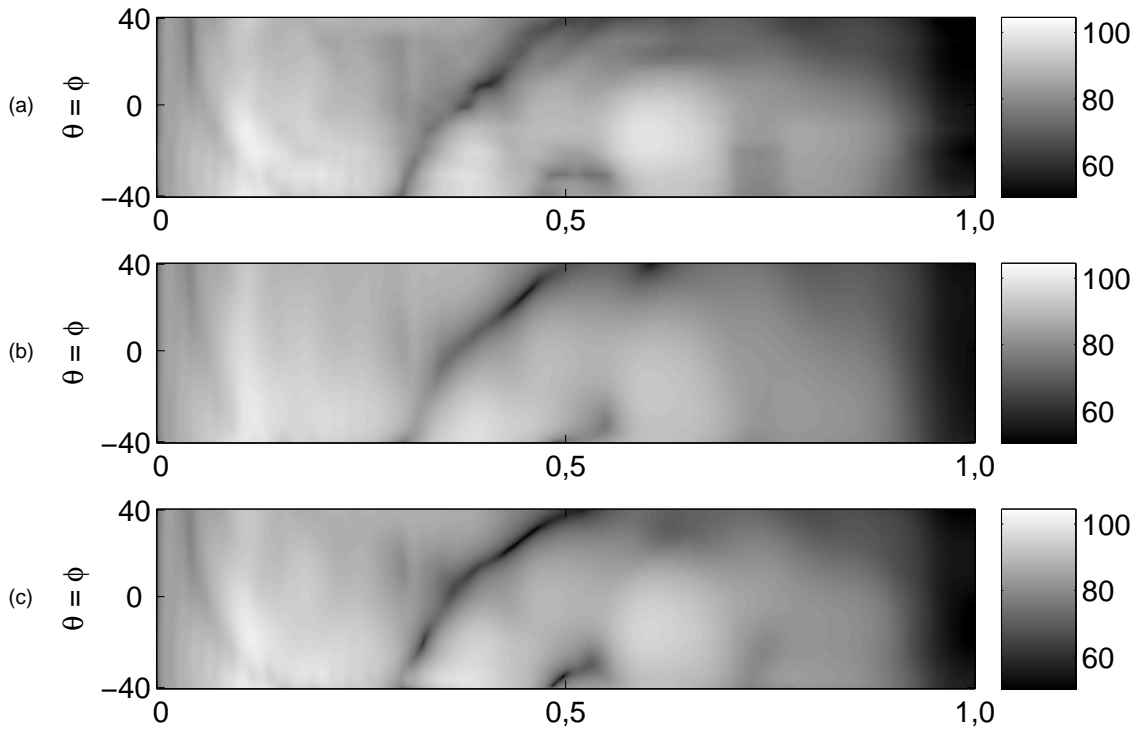


Figura 4.13: Comparação entre ordens diferentes para o caminho horizontal. (a) Interpolação bilinear das HRTFs originais. Estrutura Farrow: (b)  $L=5$ ,  $M=5$ ; (c)  $L=8$ ,  $M=8$ .

Nota-se principalmente nas Figuras 4.14 e 4.15 que a resposta referente ao polinômio de ordem maior apresentou vales mais profundos, embora a característica geral se mantenha a mesma.

Nota-se também que a continuidade e a suavidade das respostas foram mantidas, para os três caminhos de teste.

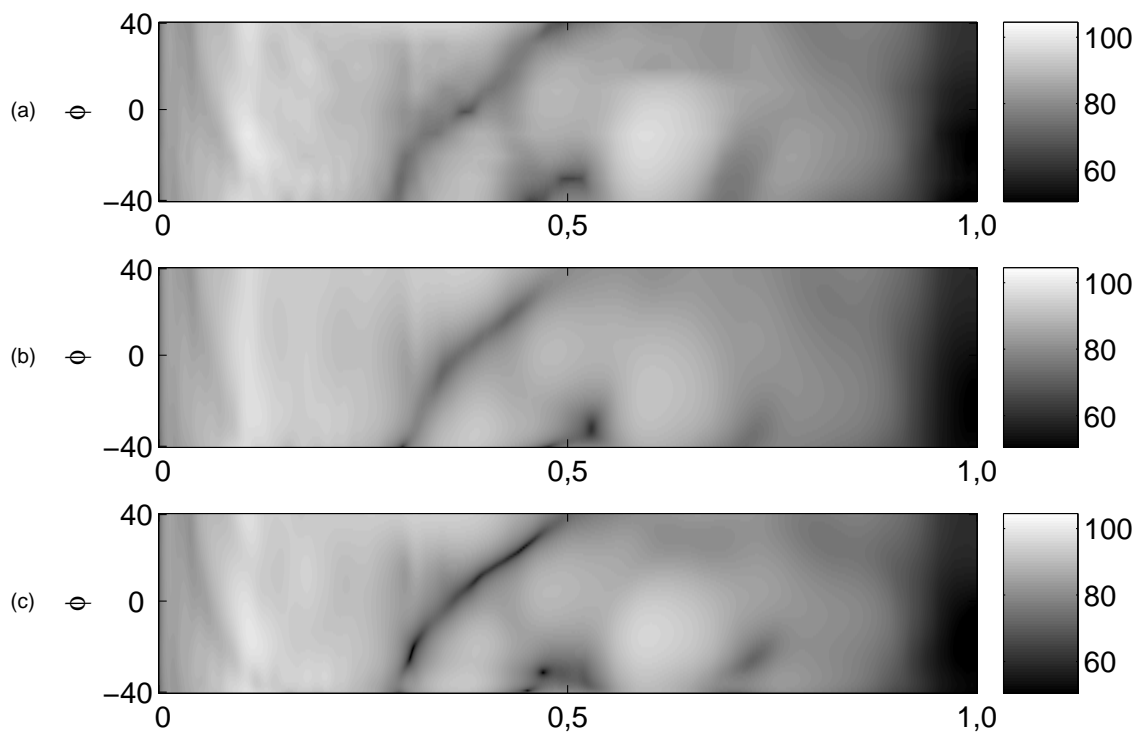
Examinando-se os gráficos (b) e (c) das Figuras 4.13 a 4.15, fica claro que, com



Frequência normalizada por 22,05 kHz.

Figura 4.14: Comparação entre ordens diferentes para o caminho diagonal. (a) Interpolação bilinear das HRTFs originais. Estrutura Farrow: (b)  $L=5$ ,  $M=5$ ; (c)  $L=8$ ,  $M=8$ .

o aumento da ordem, a estrutura Farrow torna-se capaz de delinear muito melhor os acidentes (em particular, os vales) das HRTFs em (a). Nesse caso, contudo, é preciso ter em mente que o aumento da ordem eleva quadraticamente a complexidade do cálculo, podendo torná-lo sensível a erros numéricos a ponto de inutilizar a solução.



Frequência normalizada por 22,05 kHz.

Figura 4.15: Comparação entre ordens diferentes para o caminho vertical. (a) Interpolação bilinear das HRTFs originais. Estrutura Farrow: (b)  $L=5$ ,  $M=5$ ; (c)  $L=8$ ,  $M=8$ .



## 4.4 Formulação da Estrutura Farrow para 3 Variáveis Espaciais

A formulação anterior da estrutura Farrow foi feita especificamente para aplicação de perifonia, uma vez que os conjuntos de HRTFs disponíveis consideram medições com alto-falantes localizados numa esfera. Na formulação apresentada a seguir, considera-se uma terceira variável: a distância  $r$  entre a fonte e o centro do sistema de coordenadas polares interaural, apresentada na Seção 1.2.

A resposta na frequência da HRIR para uma determinada posição seria dada por

$$G(e^{j\omega}, \theta, \phi, r) = \sum_n h_n(\theta, \phi, r) e^{-j\omega}. \quad (4.26)$$

A estrutura Farrow para 3 variáveis teria coeficientes  $c_n^{l,m,p}$  de um polinômio de ordem  $L$  em  $\theta$  e ordem  $M$  em  $\phi$  e ordem  $P$  em  $r$ , cuja resposta na frequência seria dada por

$$F(e^{j\omega}, \theta, \phi, r) = \sum_n \sum_l \sum_m \sum_p \theta^l \phi^m c_n^{l,m,p} e^{-j\omega}. \quad (4.27)$$

A função-custo seria dada pela Equação (4.28)

$$J = \sum_{\omega} \sum_{\theta} \sum_{\phi} \sum_r \left| F(e^{j\omega}, \theta, \phi, r) - G(e^{j\omega}, \theta, \phi, r) \right|^2. \quad (4.28)$$

Substituindo (4.26) e (4.27) em (4.28), teríamos a seguinte equação:

$$J = \sum_{\omega} \sum_{\theta} \sum_{\phi} \sum_r \left| \sum_n \sum_l \sum_m \sum_p \theta^l \phi^m c_n^{l,m,p} e^{-j\omega} - \sum_n h_n(\theta, \phi, r) e^{-j\omega} \right|^2. \quad (4.29)$$

Procedendo de forma análoga ao caso em que se considerava apenas uma esfera, chegar-se-ia à solução dada pelas Equações (4.30) e (4.31).

$$J = (N + 1)(J_{n_0} + J_{n_1} + \dots + J_{n_N}), \quad (4.30)$$

onde

$$J_{n_i} = \sum_{\theta} \sum_{\phi} \sum_r \left( \sum_l \sum_m \sum_p \theta^l \phi^m c_{n_i}^{l,m,p} - h_{n_i}(\theta, \phi, r) \right)^2. \quad (4.31)$$

Portanto, a solução no tempo da Equação (4.28) leva às Equações (4.30) e (4.31), para a otimização dos coeficientes da estrutura Farrow generalizada para 3

variáveis espaciais. Tal estrutura depende da existência de um conjunto de medidas de HRTFs que tenha sido feita com diversos raios. Devido ao fato de os conjuntos de HRTFs do Capítulo 2 terem sido obtidos com raios fixos, nesta dissertação não serão apresentados exemplos da formulação de 3 variáveis mostrada acima.

## 4.5 Conclusões

A estrutura Farrow generalizada [28] aproxima HRTFs ao longo da esfera de referência com pesos polinomiais em função das variáveis angulares de posição. Nesse trabalho modificou-se o método de otimização desses polinômios, reformulando-o no tempo pelo Teorema de Parseval. Dessa forma, tornou-se possível a aproximação conjunta das HRTFs ao redor de toda a esfera de referência, bem como a utilização de ordens polinomiais mais elevadas.

Simulações mostraram que as duas formulações são equivalentes, e que um pequeno aumento na ordem pode melhorar bastante a aproximação; contudo, ainda é preferível realizar a aproximação em sub-regiões da esfera. Naturalmente, é necessária a realização de testes subjetivos sistemáticos do desempenho do sistema proposto, para fontes virtuais em movimento.

A seguir, foi feita a generalização da estrutura Farrow para três variáveis espaciais, através de formulação análoga à mostrada anteriormente.

### Contribuições deste capítulo

- Desenvolvimento e validação de formulação eficiente para a estrutura Farrow.
- Comparação entre estrutura otimizada numa região e na esfera toda.
- Investigação do efeito da variação de ordem da estrutura sobre a interpolação.
- Generalização da estrutura Farrow para 3 variáveis espaciais.

# Capítulo 5

## Panoramização por Amplitude para Reprodução por Alto-Falantes

### 5.1 Introdução

Este capítulo trata de métodos de geração de áudio 3D para utilização com caixas acústicas. Tais técnicas devem contornar diversas dificuldades não existentes no caso de a reprodução ser por fones de ouvido, conforme descrito no Capítulo 1. A principal é a diafonia, o fato de o sinal de um dos alto-falantes ser percebido por ambas as orelhas. Isso faz com que a panoramização deva ser efetuada considerando-se e compensando-se este efeito.

A panoramização por amplitude é uma técnica de se criar o efeito de tridimensionalidade em caixas acústicas variando apenas as amplitudes dos sinais em cada caixa.

Serão apresentados os métodos mais conhecidos de panoramização por amplitude: as Leis dos Senos e das Tangentes [31]. Essas leis se baseiam numa aproximação de primeira ordem da composição dos sinais emitidos pelas caixas acústicas e nas hipóteses simplificadoras de que as caixas estão no plano horizontal, simétricas e equidistantes em relação ao ouvinte.

A Lei dos Senos é mais apropriada para o ouvinte fixo na posição, enquanto a Lei das Tangentes considera que o ouvinte segue a fonte virtual com os olhos, movendo a cabeça [31].

Outro método de panoramização por amplitude é o *Vector Base Amplitude*

*Panning* (VBAP) [5, 32], que considera a posição do som numa esfera como sendo um vetor, e a sua projecção no eixo das caixas acústicas como o ganho a ser aplicado.

Este capítulo mostra a relação existente entre as Leis dos Senos e das Tangentes com o método de panoramização por amplitude VBAP e com outros métodos de panoramização. Estes resultados são obtidos pela generalização das equações originais para 3 dimensões e a aplicação das hipóteses simplificadoras de cada modelo.

Será discutida a unicidade das soluções para as equações de panoramização por amplitude. Em seguida será mostrado um método que deriva de hipóteses diferentes das adotadas anteriormente, e, por fim, será feita uma análise comparativa de alguns dos métodos descritos.

A estrutura do capítulo está mostrada na Figura 5.1.

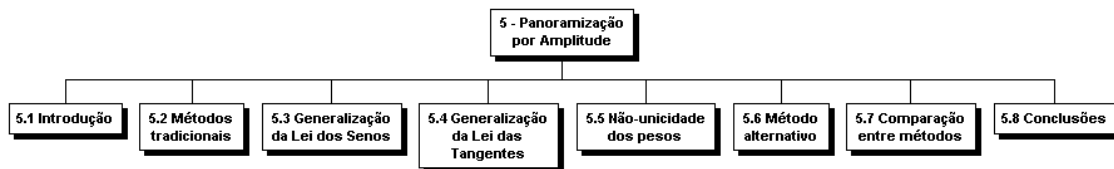


Figura 5.1: Conteúdo do capítulo sobre panoramização por amplitude.

## 5.2 Métodos Tradicionais de Panoramização por Amplitude

### 5.2.1 Lei dos Senos

Nesta seção será mostrada a dedução da Lei dos Senos, conforme o original em [31]. O esquema mostrado na Figura 5.2 foi utilizado para a formulação.

As hipóteses adotadas foram:

- Duas caixas acústicas simétricas e dispostas no plano horizontal;
- Distância entre a cabeça e as caixas bem maior do que distância entre as orelhas;
- Freqüências pequenas comparadas com a maior freqüência do sinal ou função de envoltória; e

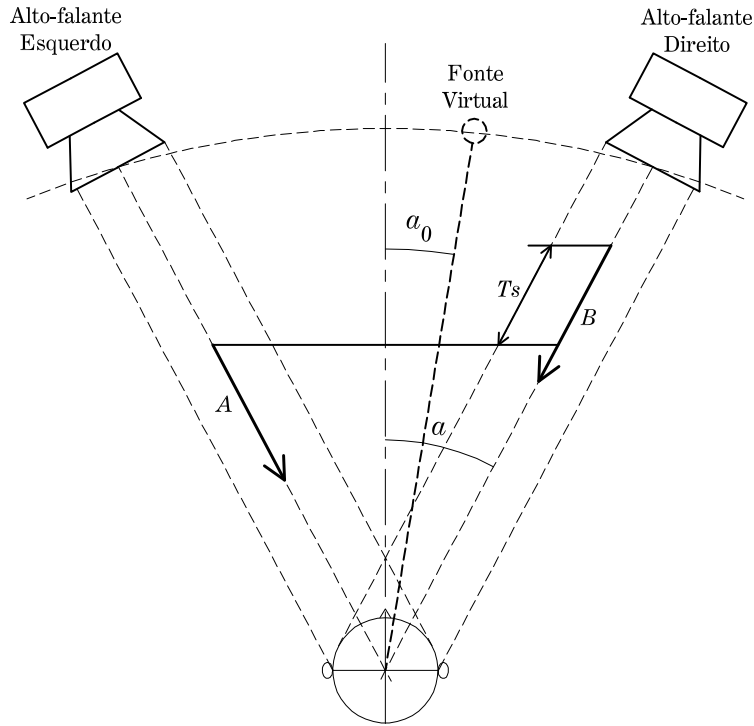


Figura 5.2: Diagrama para estudo do mecanismo de estereofonia direcional.

- d) Som que chega numa das orelhas idêntico ao que chega na outra orelha, a menos de uma diferença de ganho e um atraso.

A hipótese a) pode ser estendida para mais de duas caixas acústicas e fora do plano horizontal, conforme se verá nesta dissertação.

A hipótese b) significa que pode-se considerar a reta que une a orelha direita e a fonte virtual paralela à reta que une a orelha esquerda e a fonte virtual.

A hipótese c) estabelece que a lei é válida para baixas frequências, de modo a não haver ambigüidade na percepção da fase da ITD, conforme descrito na Subseção 1.2.

A hipótese d) assume que as distorções dependentes da frequência introduzidas pela presença do torso, da pina e da cabeça são desconsideradas pelo modelo.

Conforme ilustrado na Figura 5.2, definem-se os ganhos dos alto-falantes esquerdo e direito como  $A$  e  $B$ , respectivamente. O sinal monoaural  $f(t)$  é transmitido a cada alto-falante. Quer-se variar o ganho e o atraso inicial de cada alto-falantes de modo a posicionar a fonte virtual na posição desejada. Sem perda de generalidade, pode-se considerar que o sinal do alto-falante direito tem um atraso inicial  $T_s$  em

relação ao sinal do outro lado.

A relação entre as amplitudes dos sinais emitidos por cada alto-falante pode ser descrita por

$$A \neq B \quad \text{ou} \quad A = kB, \quad (5.1)$$

onde  $k$  é uma constante real.

Do sinal emitido pelo alto-falante esquerdo, a amplitude que chega à orelha direita é  $A_R$ , enquanto a amplitude que chega à orelha esquerda é  $A_L$ . Analogamente, denominam-se  $B_R$  e  $B_L$  às amplitudes dos sinais a partir do alto-falante direito que alcançam as orelhas direita e esquerda, respectivamente.

A partir da Figura 5.2 e das considerações acima, as relações entre os tempos de chegada do som às duas orelhas podem ser esquematizadas conforme a Figura 5.3. Considerou-se que o alto-falante esquerdo emitiu o som primeiro, e a emissão do som pelo alto-falante direito o fez com um atraso  $T_s$ . O valor de  $T_a$  refere-se à ITD do sinal emitido por um alto-falante num azimute de ângulo  $a$ . Devido à hipótese de caixas acústicas dispostas simetricamente em relação ao plano mediano, a ITD relativa ao alto-falante esquerdo é a mesma que para o alto-falante direito.

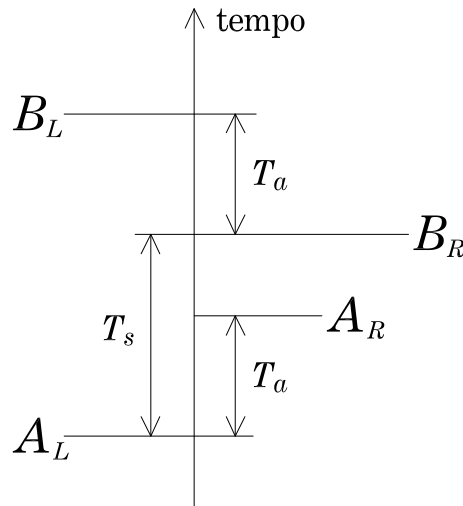


Figura 5.3: Chegada dos sinais às orelhas.

O esquema da Figura 5.3 mostra o efeito da chegada do som das duas caixas acústicas às orelhas, ou seja, considera o efeito da diafonia. Portanto os sinais que chegam às orelhas esquerda e direita são dados por

$$L = A_L f(t) + B_L f(t - (T_a + T_s)), \quad (5.2)$$

$$R = A_R f(t - T_a) + B_R f(t - T_s). \quad (5.3)$$

O resultado final, após a expansão em série de Taylor e o truncamento de primeira ordem, conforme descrito no original [31], é dado por:

$$L \approx (A_L + B_L) f\left(t - \frac{B_L T_a + B_L T_s}{A_L + B_L}\right), \quad (5.4)$$

$$R \approx (A_R + B_R) f\left(t - \frac{A_R T_a + B_R T_s}{A_R + B_R}\right). \quad (5.5)$$

A partir das equações acima, pode-se determinar os valores da diferença de tempo interaural  $\text{ITD}_{\text{caixas}}$  e da diferença de nível interaural  $\text{ILD}_{\text{caixas}}$  causados pelo efeito conjunto de ambas caixas acústicas como

$$\text{ILD}_{\text{caixas}} = \frac{A_L + B_L}{A_R + B_R}, \text{ e} \quad (5.6)$$

$$\text{ITD}_{\text{caixas}} = \frac{A_R T_a + B_R T_s}{A_R + B_R} - \frac{B_L T_a + B_L T_s}{A_L + B_L}. \quad (5.7)$$

Observando as Equações (5.6) e (5.7), nota-se que há duas características passíveis de serem manipuladas para a geração do som virtual: a razão entre as amplitudes  $A$  e  $B$  dos alto-falantes e a diferença de tempo entre a emissão dos sinais  $T_s$ . Os métodos que lidam apenas com a diferença de amplitude serão descritos a seguir. Métodos que lidam com o atraso inicial são de difícil implementação com o uso de alto-falantes, por conta da diafonia e do efeito de “precedência” do som, em que ao invés de se perceber um som virtual, são percebidos dois sons separados [1].

Portanto, para a panoramização por amplitude, o atraso inicial entre os sinais é desconsiderado, ou seja,  $T_s = 0$ .

Para frequências abaixo de 600 Hz o efeito de difração da cabeça é negligível [31]. Isso significa que  $A_L = A_R = A$  e  $B_L = B_R = B$  para tais frequências.

Dessa forma, inserindo as hipóteses de variação apenas em amplitude e para baixas frequências nas Equações (5.6) e (5.7), chega-se a

$$\text{ILD}_{\text{caixas}} = 1, \quad (5.8)$$

$$\text{ITD}_{\text{caixas}} = \frac{A - B}{A + B} T_a. \quad (5.9)$$

Com relação à Equação (5.9), as variáveis  $A$  e  $B$  são as incógnitas a serem determinadas. O  $T_a$  é o atraso da propagação do sinal entre uma orelha e outra.

Se o objetivo é fazer a panoramização da fonte virtual, deve-se igualar o valor de  $ITD_{\text{caixas}}$  do conjunto de caixas acústicas com a  $ITD_0$  da fonte virtual, conforme

$$ITD_{\text{caixas}} = ITD_0. \quad (5.10)$$

Uma ITD pode ser estimada geometricamente considerando-se uma cabeça esférica e um caminho reto entre a fonte sonora e as orelhas, de acordo com o Capítulo 3. A Figura 5.4 ilustra a estimativa de ITD de uma fonte virtual.

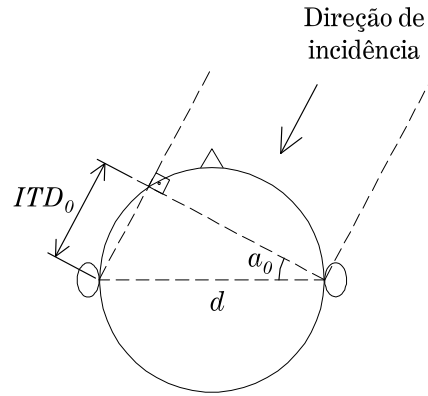


Figura 5.4: Figura ilustrativa para a obtenção da ITD.

Utilizando o esquema da Figura 5.4, a ITD relativa a uma fonte virtual no azimute  $a_0$  é dada por:

$$ITD_0 = \frac{d}{c} \text{sen } a_0, \quad (5.11)$$

onde  $c$  é a velocidade do som no meio.

Analogamente, o atraso  $T_a$  de uma única caixa acústica localizada num azimute  $a$  é dada por

$$T_a = \frac{d}{c} \text{sen } a. \quad (5.12)$$

Substituindo os valores de  $ITD_0$  e  $T_a$  nas Equações (5.9) e (5.10), chega-se à Lei dos Senos:

$$\frac{\text{sen } a_0}{\text{sen } a} = \frac{A - B}{A + B}. \quad (5.13)$$

## 5.2.2 Lei das Tangentes

O método de obtenção da Lei das Tangentes é similar ao da Lei dos Senos, exceto que se considera que o ouvinte acompanha a fonte virtual.



A nova representação esquemática das posições relativas da cabeça e das orelhas com o efeito da rotação da cabeça está demonstrada na Figura 5.5, onde  $L$  e  $R$  referem-se à posição original das orelhas esquerda e direita,  $L'$  e  $R'$  referem-se à posição rotacionada das mesmas.

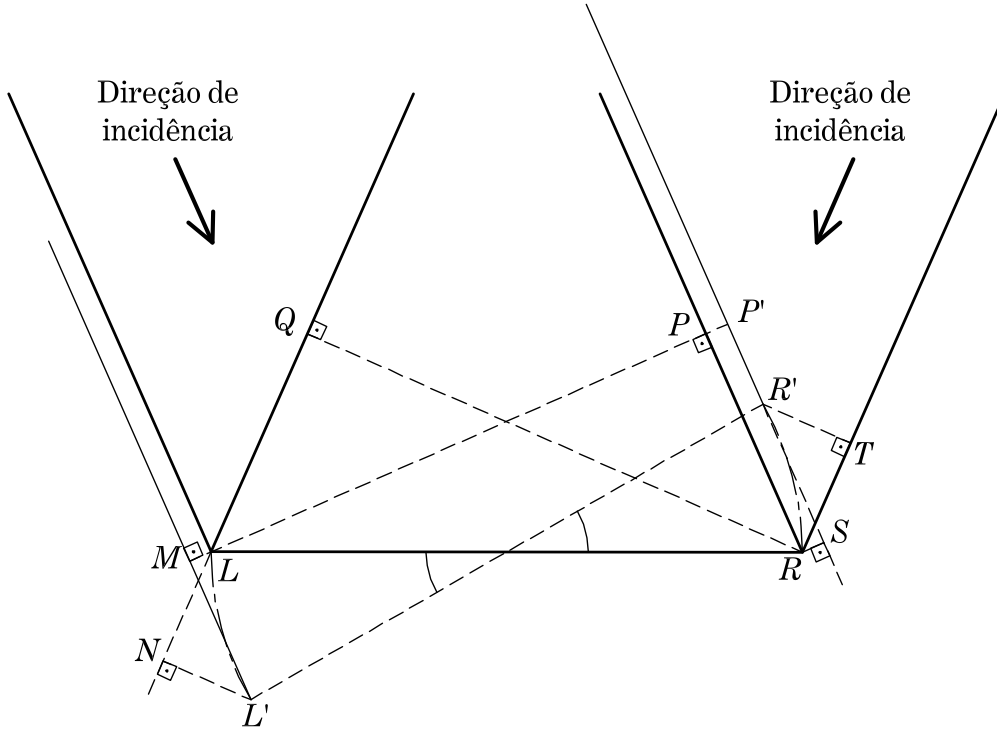


Figura 5.5: Posições relativas com rotação.

A partir da Figura 5.5, pode-se representar o esquema de chegada do som nas orelhas, considerando como referência a posição inicial do ouvinte, conforme a Figura 5.6.

Utilizando novamente a hipótese de que para frequências abaixo de 600 Hz o efeito de difração da cabeça é negligível, tem-se  $A_L = A_R = A$  e  $B_L = B_R = B$ .

A partir do esquema da Figura 5.6, os sinais podem ser descritos por

$$L = Af(t - T_1) + Bf(t - T_2), \quad (5.14)$$

$$R = Af(t + T_3) + Bf(t - T_4). \quad (5.15)$$

Utilizando a expansão em série de Taylor e desprezando os termos de ordem

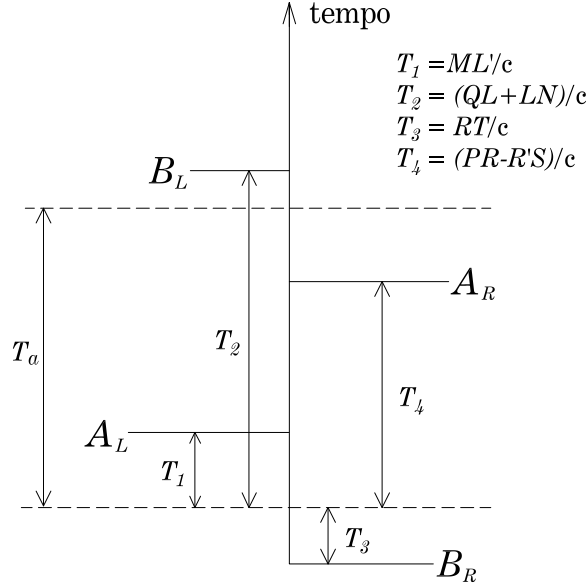


Figura 5.6: Chegada dos sinais nas orelhas, com rotação.

maior do que um, tem-se

$$L \approx (A + B)f\left(t - \frac{AT_1 + BT_2}{A + B}\right), \quad (5.16)$$

$$R \approx (A + B)f\left(t - \frac{AT_4 - BT_3}{A + B}\right). \quad (5.17)$$

Com a hipótese de se tratar de baixas frequências, a ITD e a ILD tornam-se

$$\text{ILD}_{\text{caixas}} = 1, \quad (5.18)$$

$$\text{ITD}_{\text{caixas}} = \frac{AT_4 - BT_3}{A + B} - \frac{AT_1 + BT_2}{A + B}. \quad (5.19)$$

Pelo fato de o ouvinte estar voltado para a fonte virtual, a ITD da fonte virtual é zero.

$$\text{ITD}_0 = 0. \quad (5.20)$$

Igualando-se  $\text{ITD}_{\text{caixas}}$  a  $\text{ITD}_0$  e obtendo-se os valores de  $T_1$  a  $T_4$  por geometria plana, obtém-se a Lei das Tangentes, conforme

$$\frac{\text{tg } a}{\text{tg } a_0} = \frac{A - B}{A + B}. \quad (5.21)$$

O desenvolvimento completo da formulação está apresentado no Apêndice B.

### 5.2.3 Panoramização por Amplitude numa Base Vetorial — VBAP

O método VBAP tem a característica de ser computacionalmente eficiente para posicionamento do som virtual com caixas acústicas [5]. Em muitos métodos de panoramização o número de alto-falantes e suas posições são fixos, de modo que se possa ter fórmulas fixas para fazer o posicionamento da fonte virtual. No caso do VBAP, pode-se estender sua formulação para um número maior de caixas acústicas e em posições livres [5].

A direção do som virtual é definida por um vetor unitário  $\mathbf{p} = [p_1 \ p_2 \ p_3]^T$ .

A posição de cada caixa acústica também é dada por um vetor, conforme

$$\begin{aligned}\mathbf{l}_1 &= [l_{11} \ l_{12} \ l_{13}]^T \\ \mathbf{l}_2 &= [l_{21} \ l_{22} \ l_{23}]^T \\ \mathbf{l}_3 &= [l_{31} \ l_{32} \ l_{33}]^T.\end{aligned}\tag{5.22}$$

Seja  $\mathbf{g} = [g_1 \ g_2 \ g_3]^T$ , onde  $g_1$ ,  $g_2$  e  $g_3$  são os fatores de ganho de cada alto-falante. A seguir, expressa-se a posição da fonte virtual como uma combinação linear dos vetores-posição  $\mathbf{l}_1$ ,  $\mathbf{l}_2$  e  $\mathbf{l}_3$  dos três alto-falantes, conforme

$$\mathbf{p} = g_1\mathbf{l}_1 + g_2\mathbf{l}_2 + g_3\mathbf{l}_3.\tag{5.23}$$

Seja  $\mathbf{L}_{123} = [\mathbf{l}_1 \ \mathbf{l}_2 \ \mathbf{l}_3]$ . Em termos vectoriais, a Equação (5.23) pode ser reescrita como

$$\mathbf{p} = \mathbf{L}_{123}\mathbf{g}.\tag{5.24}$$

O vetor  $\mathbf{g}$  pode ser resolvido a partir da Equação (5.24) como em

$$\mathbf{g} = \mathbf{L}_{123}^{-1}\mathbf{p} = \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ l_{12} & l_{22} & l_{32} \\ l_{13} & l_{23} & l_{33} \end{bmatrix}^{-1} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}.\tag{5.25}$$

A matriz  $\mathbf{L}_{123}^{-1}$  existe sempre que os vetores  $\mathbf{l}_1$ ,  $\mathbf{l}_2$  e  $\mathbf{l}_3$  formarem uma base que gera o espaço tridimensional.

A seguir, o vetor de ganho  $\mathbf{g}$  deve ser normalizado. Há duas formas básicas de normalização: através da amplitude

$$\mathbf{g}^{\text{normalizado}} = \frac{\mathbf{g}}{g_1 + g_2 + g_3}\tag{5.26}$$

e da energia

$$\mathbf{g}^{\text{normalizado}} = \frac{\mathbf{g}}{\sqrt{g_1^2 + g_2^2 + g_3^2}}. \quad (5.27)$$

A normalização por amplitude reproduz o efeito do som numa câmara anecóica enquanto a normalização por energia reproduz o efeito do som numa sala reverberante [33, 34].

De forma geral, os pesos encontrados devem satisfazer

$$\sqrt[p]{\sum_{i=1}^I g_i^p} = 1, \quad (5.28)$$

onde  $I$  é o número de caixas e o valor de  $p$  depende do grau de reverberação do ambiente que se deseja simular [33, 34].

Pode-se aplicar o método VBAP utilizando mais de 3 alto-falantes dispostos em posições arbitrárias. Para tal imagina-se que os alto-falantes particionam o espaço em triângulos, com os alto-falantes nos vértices. A seguir, aplica-se o método VBAP ao triângulo em que a fonte virtual está inserida.

Como a fonte virtual pode pertencer a diversos triângulos, a base de triângulos deve ser escolhida de modo a não haver intersecções e a fornecer a máxima acurácia na localização em cada direção [5].

## 5.2.4 Lei de Chowning

O último método de panoramização por amplitude a ser apresentado foi proposto por Chowning [34], que considera apenas a diferença de azimutes, conforme

$$\begin{aligned} A &= \frac{\theta_B - \theta_0}{\theta_B - \theta_A}, \\ B &= \frac{\theta_A - \theta_0}{\theta_A - \theta_B}, \end{aligned} \quad (5.29)$$

onde  $\theta_A$  e  $\theta_B$  são os azimutes das caixas acústicas de mesmo nome e  $\theta_0$  é o azimute da fonte virtual.

## 5.3 Generalização da Lei dos Senos

### 5.3.1 Formulação em 3 dimensões

Doravante, nesta dissertação, a Lei dos Senos tradicional será chamada Lei dos Senos 2D, enquanto a formulação para 3 dimensões será chamada Lei dos Senos 3D. O esquema adotado será o mesmo da Figura 5.7, em que são consideradas três caixas acústicas não colineares, dispostas numa esfera de raio  $r$ .

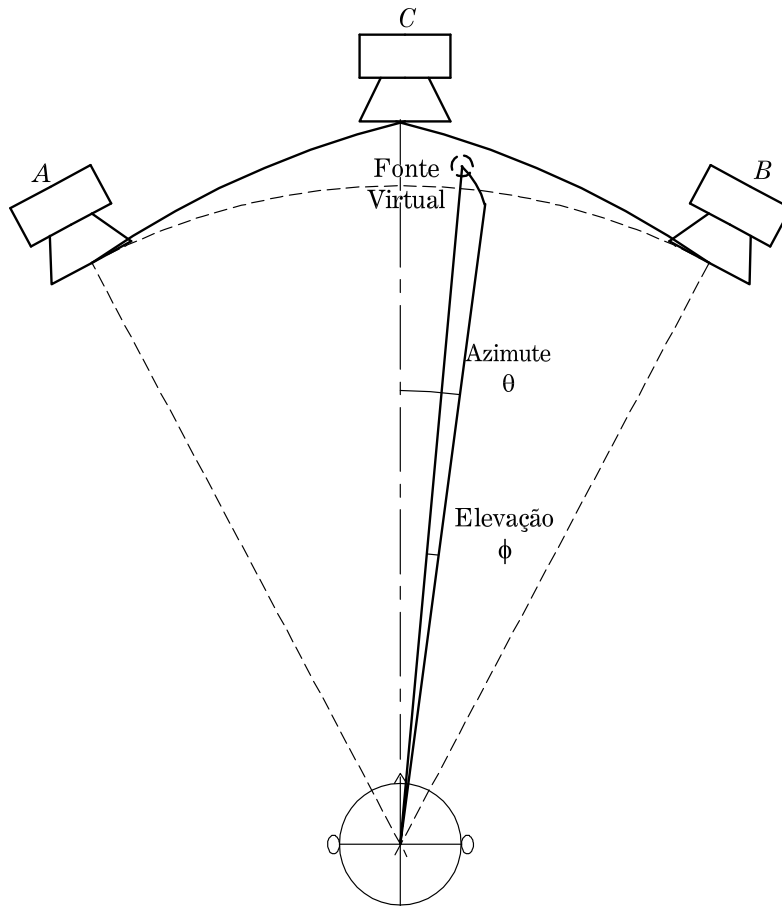


Figura 5.7: Esquema genérico para Lei dos Senos 3D.

Notação:

- $T_{A0}$ ,  $T_{B0}$  e  $T_{C0}$  — atraso inserido para o sinal monofônico respectivamente nas caixas  $A$ ,  $B$  e  $C$ ;
- $A$ ,  $B$  e  $C$  — ganho do sinal monofônico nas caixas de mesmo nome;
- $T_{iL}$  — atraso entre o alto-falante  $i$  e a orelha esquerda, onde  $i = A, B$  ou  $C$ ;

- $T_{iR}$  — atraso entre o alto-falante  $i$  e a orelha direita, onde  $i = A, B$  ou  $C$ ; e
- $c$  — velocidade do som.

Seja  $f(t)$  um sinal monofônico transmitido pelas caixas acústicas com ganhos  $A, B, C$  e atrasos iniciais  $T_{A0}, T_{B0}$  e  $T_{C0}$ , respectivamente. Seja  $L$  a composição dos sinais que chegam à orelha esquerda, e  $R$  a composição dos sinais que chegam à orelha direita. Tais valores são dados por

$$L = A_L f(t - T_{A0} - T_{AL}) + B_L f(t - T_{B0} - T_{BL}) + C_L f(t - T_{C0} - T_{CL}), \quad (5.30)$$

$$R = A_R f(t - T_{A0} - T_{AR}) + B_R f(t - T_{B0} - T_{BR}) + C_R f(t - T_{C0} - T_{CR}). \quad (5.31)$$

A primeira hipótese a ser feita é que se utilizará apenas a diferença em magnitude para a panoramização da fonte virtual, ou seja,  $T_{A0} = T_{B0} = T_{C0} = 0$ .

A segunda hipótese é considerar somente baixas frequências, o que permite afirmar que  $A_L = A_R = A$ ,  $B_L = B_R = B$  e  $C_L = C_R = C$ , pelo efeito de difração nas baixas frequências.

Após expansão em série de Taylor e truncamento dos termos de ordem maior do que 1, chega-se à Equação (5.32), que indica o valor composto da ITD com a presença das caixas acústicas, conforme detalhado no Apêndice C.

$$\text{ITD} = \frac{A(T_{AL} - T_{AR}) + B(T_{BL} - T_{BR}) + C(T_{CL} - T_{CR})}{A + B + C}. \quad (5.32)$$

A partir da formulação demonstrada no Capítulo 3, a diferença de tempo entre as chegadas do som que sai da caixa acústica  $A$  às orelhas direita e esquerda é dada por

$$T_{AL} - T_{AR} = \frac{d \operatorname{sen} \theta \cos \phi}{c}, \quad (5.33)$$

onde  $d$  é o diâmetro da cabeça (considerada esférica) e para uma trajetória retilínea entre a fonte sonora e a orelha. Para as caixas  $B$  e  $C$ , a formulação é análoga.

Suponha uma fonte virtual nas coordenadas  $(\theta_0, \phi_0)$ . Substituindo a expressão da ITD das caixas acústicas na Equação (5.32) e igualando com a ITD da fonte virtual, tem-se

$$\frac{dA \operatorname{sen} \theta_A \cos \phi_A + dB \operatorname{sen} \theta_B \cos \phi_B + dC \operatorname{sen} \theta_C \cos \phi_C}{c(A + B + C)} = \frac{d \operatorname{sen} \theta_0 \cos \phi_0}{c}. \quad (5.34)$$

Eliminando os termos que se cancelam, chega-se à “Lei dos Senos 3D”:

$$\frac{A \operatorname{sen} \theta_A \cos \phi_A + B \operatorname{sen} \theta_B \cos \phi_B + C \operatorname{sen} \theta_C \cos \phi_C}{A + B + C} = \operatorname{sen} \theta_0 \cos \phi_0. \quad (5.35)$$

Considerando que as variáveis  $A$ ,  $B$  e  $C$  estão normalizadas por amplitude, ou seja,  $A + B + C = 1$ , chega-se a

$$A \operatorname{sen} \theta_A \cos \phi_A + B \operatorname{sen} \theta_B \cos \phi_B + C \operatorname{sen} \theta_C \cos \phi_C = \operatorname{sen} \theta_0 \cos \phi_0. \quad (5.36)$$

A seguir será feita a generalização da Lei dos Senos 3D para  $N$  caixas acústicas. Depois será discutida a relação entre a Lei dos Senos 3D e diversos outros métodos de panoramização, conforme ilustrado na Figura 5.8.

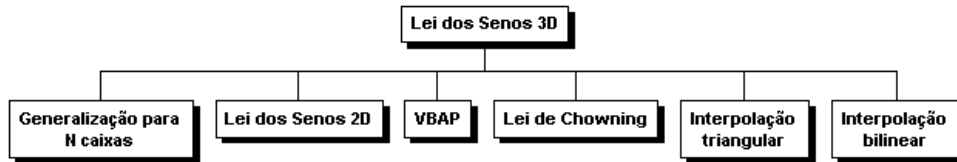


Figura 5.8: Relação entre a Lei dos Senos 3D e outros métodos.

### 5.3.2 Generalização para $N$ Caixas Acústicas

Pode-se generalizar a Equação (5.35) para se fazer o posicionamento da fonte virtual na região convexa formada por  $N$  caixas acústicas, conforme

$$\frac{A_1 \operatorname{sen} \theta_{A_1} \cos \phi_{A_1} + A_2 \operatorname{sen} \theta_{A_2} \cos \phi_{A_2} + \dots + A_N \operatorname{sen} \theta_{A_N} \cos \phi_{A_N}}{A_1 + A_2 + \dots + A_N} = \operatorname{sen} \theta_0 \cos \phi_0. \quad (5.37)$$

### 5.3.3 Derivação da Lei dos Senos 2D a partir da Lei dos Senos 3D

A Lei dos Senos 2D apresenta as hipóteses:

- Dois canais  $A$  e  $B$ ;
- Caixas acústicas simétricas; e
- Caixas acústicas no plano horizontal.

Essas hipóteses se traduzem, respectivamente, nas seguintes considerações:

- O terceiro alto-falante tem peso igual a zero:  $C = 0$ ;
- Os ângulos dos alto-falantes são iguais:  $\theta_B = -\theta_A$ ; e
- A elevação é igual a zero:  $\phi_A = \phi_B = 0$ .

Aplicando as hipóteses acima à Lei dos Senos 3D, dada pela Equação (5.34), obtém-se a Lei dos Senos 2D de modo bastante direto:

$$\begin{aligned} \frac{A \operatorname{sen} \theta_A + B \operatorname{sen}(-\theta_A)}{A + B} &= \operatorname{sen} \theta_0 \\ \implies \frac{\operatorname{sen} \theta_0}{\operatorname{sen} \theta_A} &= \frac{A - B}{A + B}. \end{aligned} \quad (5.38)$$

### 5.3.4 Relação entre Lei dos Senos 3D e método VBAP

Pelo método VBAP, os ganhos aplicados às caixas acústicas devem respeitar a Equação (5.24) reproduzida a seguir, onde os vetores posição estão em coordenadas cartesianas.

$$\mathbf{p} = \mathbf{L}_{123}\mathbf{g}. \quad (5.39)$$

De modo a converter o sistema cartesiano para o sistema de coordenadas polares interaurais apresentado na Seção 1.2, usam-se as seguintes relações:

$$\begin{aligned} x &= r \cos \phi \operatorname{sen} \theta \\ y &= r \cos \phi \cos \theta \\ z &= r \operatorname{sen} \phi. \end{aligned} \quad (5.40)$$

Convertendo a Equação (5.39) em coordenadas polares interaurais, tem-se

$$\begin{bmatrix} r \cos \phi_1 \operatorname{sen} \theta_1 & r \cos \phi_2 \operatorname{sen} \theta_2 & r \cos \phi_3 \operatorname{sen} \theta_3 \\ r \cos \phi_1 \cos \theta_1 & r \cos \phi_2 \cos \theta_2 & r \cos \phi_3 \cos \theta_3 \\ r \operatorname{sen} \phi_1 & r \operatorname{sen} \phi_2 & r \operatorname{sen} \phi_3 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} = \begin{bmatrix} r \cos \phi_0 \operatorname{sen} \theta_0 \\ r \cos \phi_0 \cos \theta_0 \\ r \operatorname{sen} \phi_0 \end{bmatrix}. \quad (5.41)$$

A primeira linha do sistema acima é dada por

$$g_1 \cos \phi_1 \operatorname{sen} \theta_1 + g_2 \cos \phi_2 \operatorname{sen} \theta_2 + g_3 \cos \phi_3 \operatorname{sen} \theta_3 = \cos \phi_0 \operatorname{sen} \theta_0, \quad (5.42)$$



que corresponde exatamente à Lei dos Senos 3D.

Portanto, das Equações (5.39) e (5.42), tem-se que a solução que satisfaz o método VBAP também satisfaz a Lei dos Senos 3D.

### 5.3.5 Derivação da Lei de Chowning a partir da Lei dos Senos 3D

A Lei de Chowning apresenta as seguintes hipóteses:

1. Dois canais  $A$  e  $B$ ;
2. Caixas acústicas no plano horizontal.

Primeiramente toma-se a Lei dos Senos 3D com normalização por amplitude, dada pela Equação (5.35), e aplicam-se as hipóteses  $C = 0$  e  $\phi = 0$ . O resultado para os pesos  $A$  e  $B$  é dado pelas equações

$$A = \frac{\text{sen } \theta_B - \text{sen } \theta_0}{\text{sen } \theta_B - \text{sen } \theta_A}, \quad (5.43)$$

$$B = \frac{\text{sen } \theta_A - \text{sen } \theta_0}{\text{sen } \theta_A - \text{sen } \theta_B}. \quad (5.44)$$

A partir das equações acima, considerando-se uma aproximação  $\text{sen } \alpha \approx \alpha$ , para  $\alpha \approx 0$ , obtém-se a Lei de Chowning:

$$\begin{aligned} A &= \frac{\theta_B - \theta_0}{\theta_B - \theta_A}, \\ B &= \frac{\theta_A - \theta_0}{\theta_A - \theta_B}. \end{aligned} \quad (5.45)$$

A simplificação acima sugere que os alto-falantes devem estar localizados próximos ao azimute zero para a aplicação da Lei de Chowning.

### 5.3.6 Derivação da Interpolação Triangular a partir da Lei dos Senos 3D

A interpolação triangular de HRIRs, descrita em [19], consiste num método de ponderação das HRIRs numa grade de azimutes e elevações não regular, através da divisão da grade numa base de triângulos. A sua aplicação surge no contexto de reprodução com fones de ouvido, mas aqui se estuda sua relação com a Lei dos Senos 3D.

A formulação para a interpolação triangular segue o esquema da Figura 5.9.

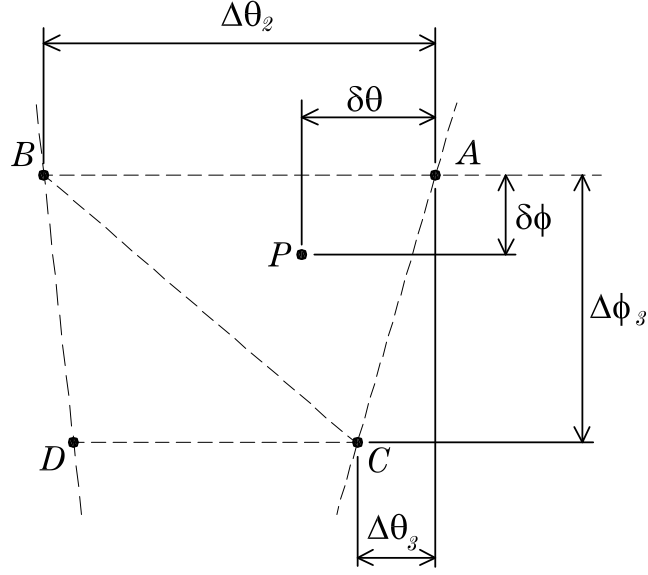


Figura 5.9: Posições relativas das caixas acústicas e da fonte virtual.

A fonte virtual  $P$ , de coordenadas  $(\theta_p, \phi_p)$ , está no triângulo formado pelos alto-falantes  $A$ ,  $B$  e  $C$ , de coordenadas  $(\theta_1, \phi_1)$ ,  $(\theta_2, \phi_2)$  e  $(\theta_3, \phi_3)$ , respectivamente.

Os valores de  $\Delta\theta_2$  e  $\Delta\phi_2$  referem-se às diferenças de azimuth e elevação do ponto  $B$  ao ponto  $A$ . O mesmo ocorre com  $\Delta\theta_3$  e  $\Delta\phi_3$ , entre o ponto  $C$  e o  $A$ . Já  $\delta\theta$  e  $\delta\phi$  referem-se às diferenças de azimuth e elevação da fonte virtual  $P$  ao ponto  $A$ .

Portanto, as coordenadas do sistema em função dos valores relativos ao alto-falante  $A$ , conforme a Figura 5.9, são dadas por:

$$\begin{aligned}
 A &: (\theta_1, \phi_1) \\
 B &: (\theta_2, \phi_2) = (\theta_1 + \Delta\theta_2, \phi_1 + \Delta\phi_2) \\
 C &: (\theta_3, \phi_3) = (\theta_1 + \Delta\theta_3, \phi_1 + \Delta\phi_3) \\
 P &: (\theta_p, \phi_p) = (\theta_1 + \delta\theta, \phi_1 + \delta\phi). \tag{5.46}
 \end{aligned}$$

Substituindo os valores acima na Lei dos Senos 3D, dada pela Equação (5.35), obtém-se a Equação (5.47).

$$\begin{aligned}
 &A \sin \theta_1 \cos \phi_1 + \\
 &B \sin(\theta_1 + \Delta\theta_2) \cos(\phi_1 + \Delta\phi_2) + \\
 &C \sin(\theta_1 + \Delta\theta_3) \cos(\phi_1 + \Delta\phi_3) = \sin(\theta_1 + \delta\theta) \cos(\phi_1 + \delta\phi). \tag{5.47}
 \end{aligned}$$

Considerando um conjunto denso de medidas, pode-se fazer as aproximações:  $\sin \alpha \approx \alpha$ ,  $\cos \alpha \approx 1$ , para  $\alpha \approx 0$ .

Deseja-se obter os valores de  $A$ ,  $B$  e  $C$  a partir da posição dos alto-falantes e da posição da fonte virtual. Considera-se que a fonte virtual encontra-se no interior de um triângulo formado pelos três alto-falantes. Fazendo a expansão das funções trigonométricas e as aproximações por conta da grade densa, chega-se exatamente aos pesos da interpolação triangular descrita em [19]. No Apêndice D mostra-se o desenvolvimento das equações:

$$\begin{aligned} A &= 1 - B - C \\ B &= \frac{\delta\theta}{\Delta\theta_2} - C \frac{\Delta\theta_3}{\Delta\theta_2} \\ C &= \frac{\delta\phi}{\Delta\phi_3}. \end{aligned} \tag{5.48}$$

Embora a interpolação triangular seja originalmente utilizada no contexto de reprodução com fones de ouvido, é curioso notar a relação desta com a Lei dos Senos 3D, sob as hipóteses adequadas.

### 5.3.7 Derivação da Interpolação Bilinear a partir da Lei dos Senos 3D

A interpolação bilinear consiste num método de ponderação das HRIRs numa grade de azimutes e elevações regular, formando um quadrado, conforme ilustrado na Figura 5.10. A interpolação bilinear tem aplicação num contexto de reprodução com fones de ouvido, mas aqui se estuda sua relação com a Lei dos Senos 3D.

A Lei dos Senos 3D para 4 alto-falantes pode ser desenvolvida de acordo com o demonstrado na Subseção 5.3.2, levando à equação

$$A \sin \theta_A \cos \phi_A + B \sin \theta_B \cos \phi_B + C \sin \theta_C \cos \phi_C + D \sin \theta_D \cos \phi_D = \sin \theta_0 \cos \phi_0. \tag{5.49}$$

Utilizando a notação da Figura 5.10 e de modo muito parecido com o desen-

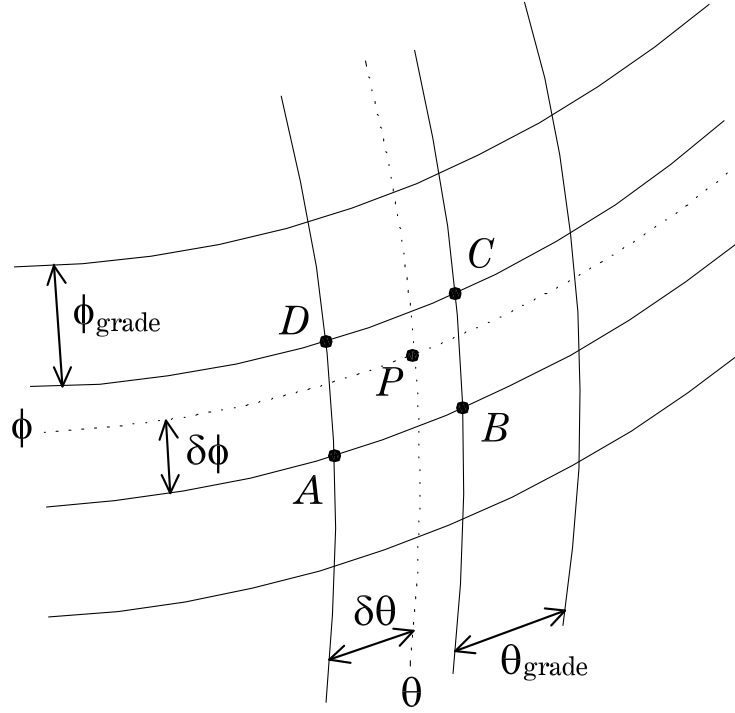


Figura 5.10: Grade de posicionamento das caixas acústicas na interpolação bilinear.

volvimento da Subseção 5.3.6, chega-se a:

$$\begin{aligned}
 & A \operatorname{sen} \theta_A \cos \phi_A + \\
 & B \operatorname{sen}(\theta_A + \theta_{\text{grade}}) \cos \phi_A + \\
 & C \operatorname{sen}(\theta_A + \theta_{\text{grade}}) \cos(\phi_A + \phi_{\text{grade}}) + \\
 & D \operatorname{sen} \theta_A \cos(\phi_A + \phi_{\text{grade}}) = \operatorname{sen}(\theta_A + \delta\theta) \cos(\phi_A + \delta\phi). \quad (5.50)
 \end{aligned}$$

Os passos são análogos ao caso da interpolação triangular, e estão descritos no Apêndice E. O resultado obtido são os pesos da interpolação bilinear:

$$\begin{aligned}
 A &= (1 - d\theta)(1 - d\phi) \\
 B &= d\theta(1 - d\phi) \\
 C &= d\theta d\phi \\
 D &= (1 - d\theta)d\phi, \quad (5.51)
 \end{aligned}$$

onde  $d\theta = \frac{\delta\theta}{\theta_{\text{grade}}}$  e  $d\phi = \frac{\delta\phi}{\phi_{\text{grade}}}$ .

Portanto, embora a interpolação bilinear seja originalmente utilizada no contexto de reprodução com fones de ouvido, mostrou-se a relação desta com a Lei dos

Senos 3D, sob as hipóteses adequadas.

## 5.4 Generalização da Lei das Tangentes

### 5.4.1 Formulação em 3 dimensões

Analogamente à seção anterior, a Lei das Tangentes tradicional será chamada Lei das Tangentes 2D, enquanto a formulação para 3 caixas acústicas não colineares numa esfera será chamada Lei das Tangentes 3D.

A Lei das Tangentes difere da Lei dos Senos por considerar que o ouvinte move a cabeça em direção à fonte virtual. Esta é uma reação natural das pessoas, para aumentar o senso de localização do som.

Pelo fato de o ouvinte estar direcionado para a fonte virtual, a ITD da fonte é nula, enquanto os ângulos de azimute e elevação dos alto-falantes se referem à nova posição da cabeça. Inserindo tais modificações na Lei dos Senos 3D, dada pela Equação (5.35), chega-se à Lei das Tangentes 3D, conforme

$$A \cos(\phi_A - \phi_0) \operatorname{sen}(\theta_A - \theta_0) + B \cos(\phi_B - \phi_0) \operatorname{sen}(\theta_B - \theta_0) + \\ C \cos(\phi_C - \phi_0) \operatorname{sen}(\theta_C - \theta_0) = 0. \quad (5.52)$$

### 5.4.2 Generalização da Lei das Tangentes 3D para $N$ Caixas Acústicas

A generalização da Lei das Tangentes 3D para representar uma fonte virtual na região convexa entre  $N$  caixas acústicas deriva da Equação (5.52), e é dada por

$$A \cos(\phi_A - \phi_0) \operatorname{sen}(\theta_A - \theta_0) + B \cos(\phi_B - \phi_0) \operatorname{sen}(\theta_B - \theta_0) + \dots \\ + N \cos(\phi_N - \phi_0) \operatorname{sen}(\theta_N - \theta_0) = 0. \quad (5.53)$$

### 5.4.3 Derivação da Lei das Tangentes 2D a partir da Lei das Tangentes 3D

Analogamente ao explicado anteriormente, a Lei das Tangentes 2D derivada por Bernfeld [31] apresenta as seguintes hipóteses:

1. O terceiro alto-falante tem peso igual a zero:  $C = 0$ ;
2. Os alto-falantes estão dispostos simetricamente em relação ao azimute zero, implicando  $\theta_A = -\theta_B$ ; e
3. Elevação igual a zero:  $\phi_A = \phi_B = 0$ .

Inserindo essas considerações na Lei das Tangentes 3D, dada pela Equação (5.52), chega-se a

$$A \operatorname{sen}(\theta_A - \theta_0) + B \operatorname{sen}(-\theta_A - \theta_0) = 0. \quad (5.54)$$

Utilizando a identidade trigonométrica da soma de senos, chega-se a

$$(A - B) \operatorname{sen} \theta_A \cos \theta_0 - (A + B) \operatorname{sen} \theta_0 \cos \theta_A = 0. \quad (5.55)$$

A Lei das Tangentes 2D é obtida da Equação (5.55), e é dada por

$$\frac{A - B}{A + B} = \frac{\operatorname{tg} \theta_0}{\operatorname{tg} \theta_A}. \quad (5.56)$$

## 5.5 Não-unicidade dos pesos de panoramização

As Leis dos Senos e das Tangentes, sem a hipótese de grade densa, não tem solução única. Isso se evidencia pelo número de equações e incógnitas envolvidas no processo: 2 equações (a equação da respectiva lei e a restrição de normalização) para 3 incógnitas, no caso de uma disposição triangular de alto-falantes. A não-unicidade da solução ocorre porque essas leis baseiam-se apenas na ITD para realizar o efeito de panoramização da fonte virtual.

O lugar geométrico dos pontos de mesma ITD e ILD é o que se chama de cone de confusão. Portanto, as Leis dos Senos e da Tangentes, ao considerar apenas a dica auditiva da ITD, permitem que qualquer fonte virtual posicionada no cone de confusão produza o mesmo efeito de ITD.

Para o caso de existirem mais do que três caixas acústicas, é a mesma situação. Haverá diversas formas de se fazer a panoramização dentro da região convexa formada pela distribuição inicial dos alto-falantes.

## 5.6 Método Alternativo de Panoramização por Amplitude

Apesar de a formulação da Lei dos Senos 3D justificar a formulação VBAP e a triangular, estas não são as únicas formulações para a Panoramização por Amplitude. Ambas as formulações utilizaram a hipótese de grade densa.

A partir do conhecimento gerado pela generalização das Leis dos Senos e das Tangentes, é possível desenvolver outras formas de se fazer a panoramização, cada qual mais adequada às condições de contorno do problema em questão.

Um método alternativo de panoramização por amplitude consiste numa formulação análoga à Lei dos Senos 3D, mas sem considerar a hipótese de grade de medidas densa. A formulação a seguir deriva da expansão trigonométrica da Equação (5.46), sem considerar essa hipótese:

$$\begin{bmatrix} 1 & \cos \Delta\theta_2 \cos \Delta\phi_2 & \cos \Delta\theta_3 \cos \Delta\phi_3 \\ 0 & \cos \Delta\theta_2 \sin \Delta\phi_2 & \cos \Delta\theta_3 \sin \Delta\phi_3 \\ 0 & \sin \Delta\theta_2 \cos \Delta\phi_2 & \sin \Delta\theta_3 \cos \Delta\phi_3 \\ 0 & \sin \Delta\theta_2 \sin \Delta\phi_2 & \sin \Delta\theta_3 \sin \Delta\phi_3 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} \cos \delta\theta \cos \delta\phi \\ \cos \delta\theta \sin \delta\phi \\ \sin \delta\theta \cos \delta\phi \\ \sin \delta\theta \sin \delta\phi \end{bmatrix}. \quad (5.57)$$

A Equação (5.57) na forma matricial, fica como

$$\mathbf{T}\mathbf{q} = \mathbf{u}. \quad (5.58)$$

Adotando a solução de menor erro no sentido de mínimos quadrados, tem-se a seguinte solução para o vetor de pesos  $\mathbf{q}$ :

$$\mathbf{q} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{u}. \quad (5.59)$$

Nota-se que não foi adotado nenhum procedimento de normalização dos pesos.

## 5.7 Comparação entre métodos

Será feita a seguir uma comparação entre alguns dos métodos de panoramização descritos anteriormente.

O procedimento consiste em tomar como referência uma HRTF de um ponto com medida conhecida. A seguir escolhem-se três pontos ao redor da referência,

formando um triângulo. Tais pontos também têm medidas de HRTF conhecidas. Estima-se a resposta do ponto de referência a partir da interpolação das HRTFs dos vértices do triângulo e compara-se o resultado à resposta da HRTF original.

Tomaram-se três triângulos progressivamente maiores, cujas coordenadas estão descritas na Figura 5.11. Os triângulos têm tamanhos diferentes para auxiliar na análise do efeito da grade de medidas não densa, sendo Triângulo 1 o menor e Triângulo 3 o maior da Figura 5.11.

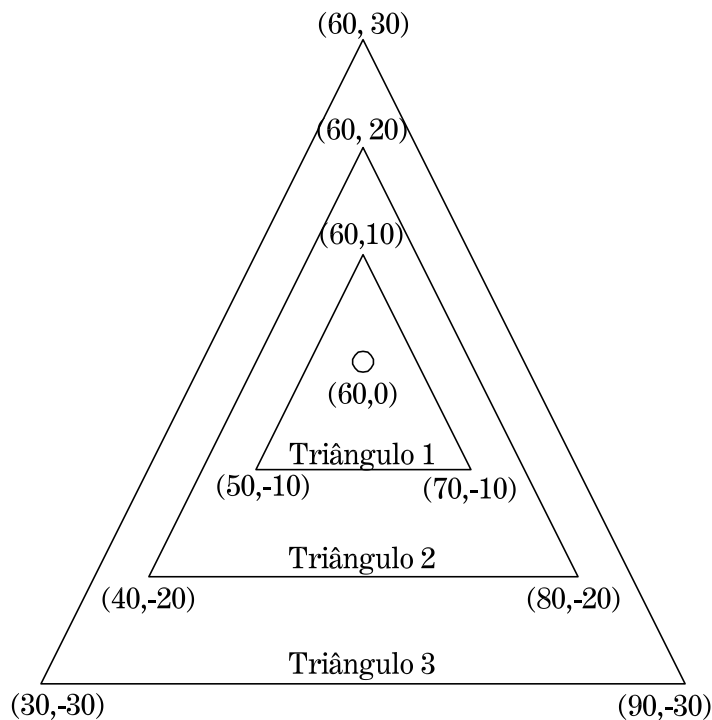


Figura 5.11: Posições dos triângulos utilizados para testes. Coordenadas  $(\theta, \phi)$  em graus.

Para a realização dos testes, foram utilizadas as HRTFs originais do conjunto de Gardner.

Os métodos utilizados para a interpolação foram o triangular, o VBAP e o método alternativo da Seção 5.6. Deve-se observar que, para as posições escolhidas, não seria indicado utilizar a Lei das Tangentes, já que esta supõe que o ouvinte esteja olhando para a fonte virtual, o que somente é possível se o centro dos triângulos estiver na posição  $(0, 0)$ .

A Figura 5.12 demonstra o resultado para a interpolação no Triângulo 1.



Nota-se que para baixas frequências o comportamento das respostas apresenta similitude com a HRTF de referência. Entretanto, para altas frequências as respostas interpoladas tornam-se bastante diferentes da referência, para qualquer método.

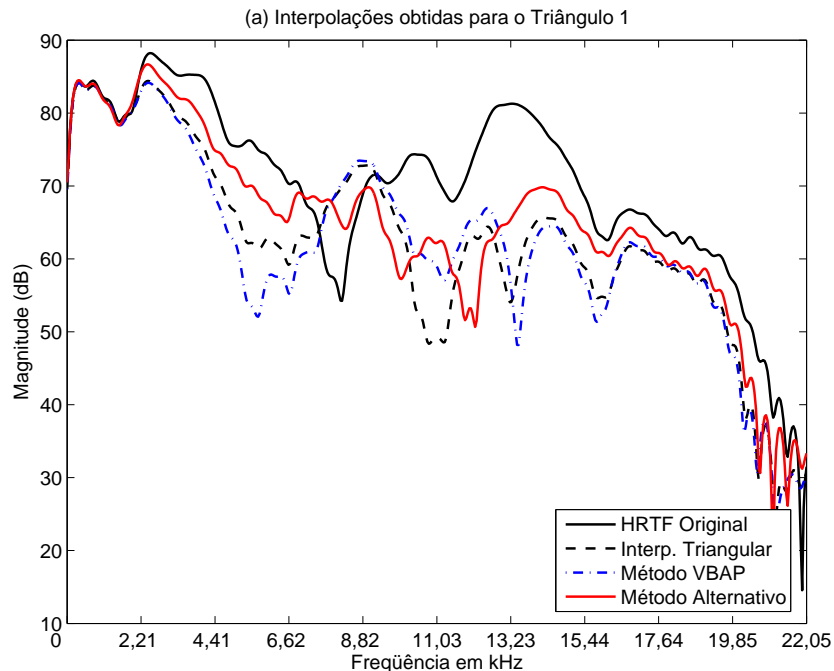


Figura 5.12: Interpolação para o Triângulo 1, para demonstrar o efeito das altas frequências.

Na medida em que, pela própria hipótese de formulação, os métodos valem apenas para baixas frequências, deve-se tomar para análise somente as respostas abaixo de 600 Hz, como demonstram as Figuras 5.13 (a), (b) e (c).

Na Figura 5.13 (a), referente ao Triângulo 1, o método alternativo foi o que se apresentou mais próximo à referência. Os métodos VBAP e interpolação triangular apresentaram resultados bastante próximos entre si, e também próximos à referência. De modo geral, todos os métodos tiveram um comportamento semelhante à referência.

Na Figura 5.13 (b) o método alternativo novamente apresentou comportamento mais próximo à referência do que os outros métodos. Nota-se que os métodos VBAP e interpolação triangular tiveram respostas próximas à referência para frequências abaixo de 200 Hz, mas acima desse valor houve grande diferença.

Na Figura 5.13 (c), referente ao Triângulo 3, o método VBAP apresentou

resposta mais próxima à referência. A interpolação triangular novamente apresentou resposta próxima à referência até a frequência de 200 Hz, sendo que a diferença aumentou para frequências acima desse valor. O método alternativo apresentou diferença bem maior do que os outros métodos.

Nota-se que a panoramização por amplitude não consegue descrever completamente os muitos aspectos dependentes da frequência de um conjunto de HRTFs. De forma geral, a análise realizada indica que as respostas das técnicas de panoramização por amplitude estudadas têm resultados mais próximos da HRTF de referência quanto mais densa a grade de medidas utilizada e na região de baixas frequências do sinal.

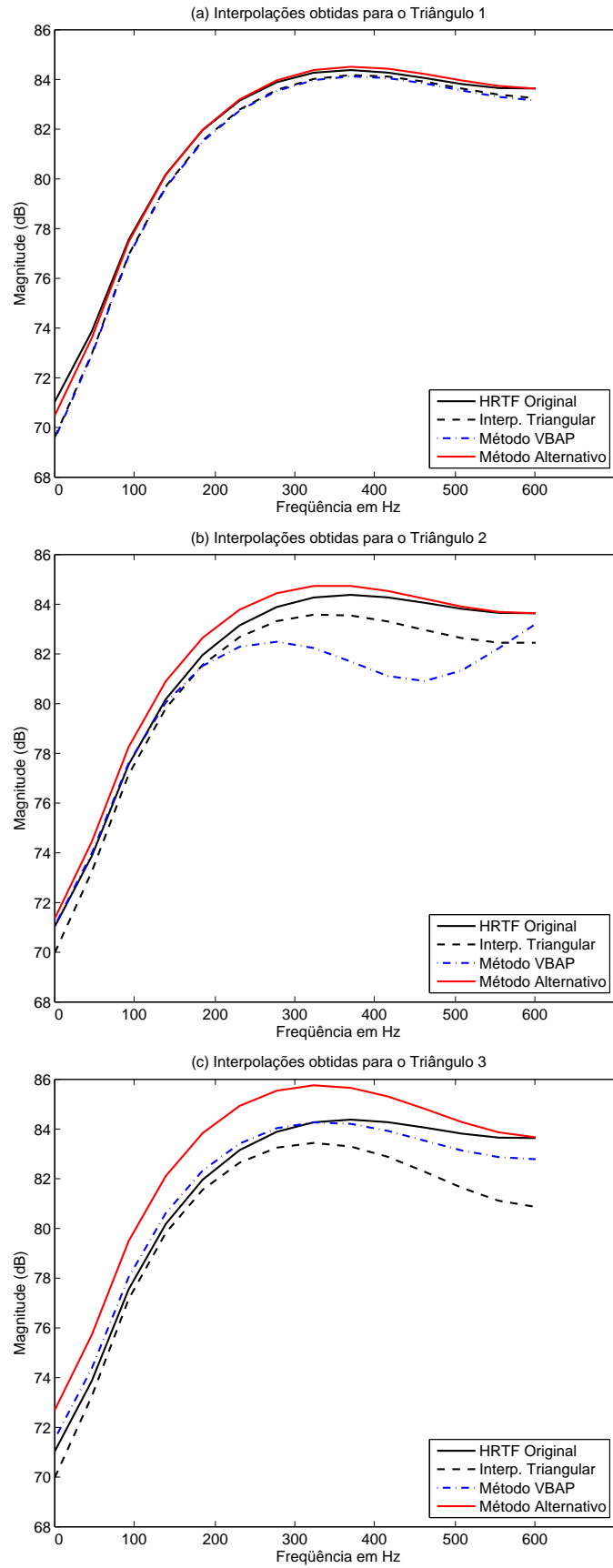


Figura 5.13: Resultados obtidos para os triângulos descritos.

Tabela 5.1: Resumo dos métodos de Panoramização por Amplitude

Hipótese	Método
2 caixas simétricas, no plano horizontal	Lei dos Senos 2D
2 caixas simétricas, plano horizontal e ouvinte acompanha fonte	Lei das Tangentes 2D
2 caixas , plano horizontal e próximo ao azimute zero	Lei de Chowning
3 caixas numa esfera	Interpolação triangular
3 caixas numa esfera	VBAP
3 caixas numa esfera	Lei dos Senos 3D
3 caixas, ouvinte acompanha fonte numa esfera	Lei das Tangentes 3D
4 caixas, grade regular, numa esfera	Interpolação Bilinear

## 5.8 Conclusões

Neste capítulo foram apresentados métodos de geração do áudio 3D baseados em panoramização por amplitude para reprodução em caixas acústicas. Foi feita a generalização das Leis do Seno e da Tangente, e foram pesquisadas as relações destes com os métodos tradicionais.

As maiores restrições dos métodos de panoramização por amplitude são a de serem melhores para baixas frequências e para conjuntos de HRTFs com grade de medidas densa.

O conhecimento das hipóteses de cada método de geração de áudio 3D é importante para que o projetista saiba aproveitar as características inerentes a cada método.

A Tabela 5.1 apresenta as principais hipóteses restritivas relacionados a cada método de panoramização por amplitude.

### Contribuições deste capítulo

- Extensão explícita da formulação das Leis dos Senos e das Tangentes para 3D.
- Derivação de diversas outras regras de panoramização a partir da Lei dos Senos 3D.
- Proposta de método alternativo de panoramização por amplitude.

# Capítulo 6

## Conclusões

Esta dissertação tratou de descrever, analisar e aperfeiçoar métodos de geração de som tridimensional de acordo com o modo de reprodução adotado. Primeiramente, foram apresentados os principais conceitos, definições e desafios relacionados ao áudio 3D.

No Capítulo 2, alguns conjuntos de HRTFs foram apresentados e tiveram a notação normalizada para permitir comparação posterior. Foi feita uma análise da distribuição de energia com a posição. A seguir, foi feita uma análise da representação dos conjuntos com a utilização da KLT. Por fim, foram realizados testes subjetivos de percepção da posição estática, donde se concluiu que não havia diferença significativa nas respostas dos ouvintes para os diferentes conjuntos analisados.

No Capítulo 3 apresentou-se um método de estimar o valor da ITD de um conjunto de medidas, e também modelos de ITD baseados em simplificações geométricas da cabeça humana. Fez-se a comparação dos resultados obtidos através de ambos os métodos, para os conjuntos de medida do capítulo anterior.

Foi apresentada no Capítulo 4 uma reformulação da Estrutura Farrow para interpolação de HRTFs, de modo a possibilitar uma solução com complexidade computacional e exigência de memória várias ordens de grandeza menores do que métodos anteriores. Fez-se a validação do novo modelo em relação ao antigo, e analisou-se a otimização realizada numa região do espaço e na esfera toda. Investigou-se o efeito da variação de ordem da estrutura sobre a interpolação. Por fim, desenvolveu-se a formulação para a Estrutura Farrow considerando a variável “distância”.

Para a aplicação do áudio 3D em alto-falantes, foram apresentados no Capí-

tulo 5 os principais métodos de panoramização por amplitude. A Lei dos Senos foi generalizada explicitamente para 3 dimensões. Foram demonstradas as relações do métodos apresentados com a Lei dos Senos 3D, variando apenas as hipóteses restritivas. Foi desenvolvido um método alternativo, através da mudança de algumas das hipóteses feitas anteriormente.

## Estrutura da Dissertação

A estrutura da dissertação, descrita em seus capítulos e seções, está mostrada na Figura 6.1.

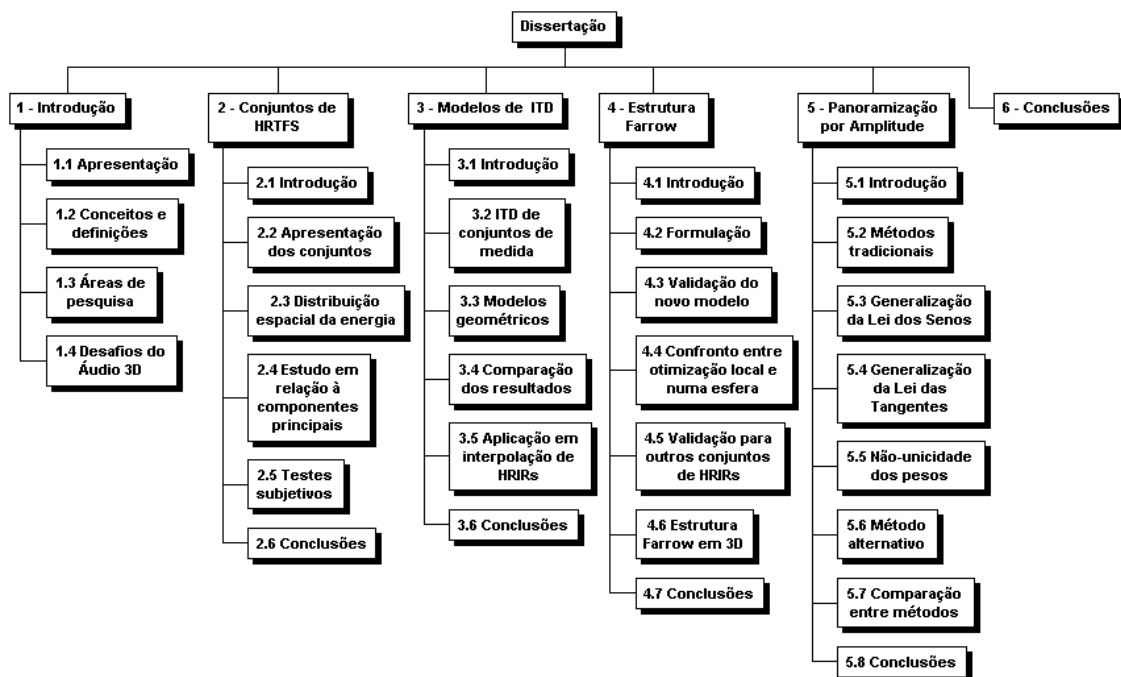


Figura 6.1: Estrutura detalhada da Dissertação.

## Principais contribuições desta dissertação

- Com relação ao estudo de conjuntos de HRTFs:
  - Apresentação de 3 conjuntos de medidas de HRTFs e análise comparativa destes quanto à grade de medidas, ao número de coeficientes, à distribuição espacial da energia e ao comportamento na decomposição KLT.

- Realização de testes subjetivos de percepção de posição estática para os 3 conjuntos apresentados anteriormente. Análise comparativa dos resultados quanto ao erro médio quadrático (MSE) e às reversões.
- Com relação ao estudo da ITD:
  - Apresentação de comparação entre estimativa de ITD a partir de modelos geométricos teóricos e medidas experimentais.
  - Proposta da função mediana como estimador do atraso puro que substitui o excesso de fase (Equação (3.14)).
  - Proposta de aplicação dos modelos geométricos em interpolação de HRIRs.
- Com relação à estrutura Farrow:
  - Desenvolvimento e validação de formulação eficiente para a estrutura Farrow.
  - Comparação entre estrutura otimizada numa região e na esfera toda.
  - Investigação do efeito da variação de ordem da estrutura sobre a interpolação.
  - Generalização da estrutura Farrow para 3 variáveis espaciais.
- Com relação a métodos de panoramização por amplitude:
  - Extensão explícita da formulação das Leis dos Senos e das Tangentes para 3D.
  - Derivação de diversas outras regras de panoramização a partir da Lei dos Senos 3D.
  - Proposta de método alternativo de panoramização por amplitude.

### **Continuação do trabalho**

A continuação do trabalho inclui, entre outros:

- Desenvolvimento de um equalizador espacial, que descreva de forma simples as principais características de variação de energia em função da posição para os conjuntos de HRTFs descritos;

- Buscar elevar a ordem polinomial da estrutura Farrow pelo uso de métodos mais robustos de cálculo;
- Elaborar técnicas para garantir a transição suave entre regiões modeladas separadamente;
- Testar o comportamento da estrutura Farrow para outros conjuntos de HRTFs;
- Ainda na estrutura Farrow, buscar novas formulações para a função de custo;
- Pesquisar o desenvolvimento de outras formas de panoramização por amplitude e a sua relação com outros métodos de geração de áudio 3D; e
- Desenvolver um método de geração que consiga produzir a ilusão de tridimensionalidade independentemente do modo de reprodução, para aplicações onde não se sabe qual será o modo de reprodução utilizado.



# Referências Bibliográficas

- [1] BEGAULT, D. R., *3-D Sound for Virtual Reality and Multimedia*, Technical memorandum, National Aeronautics and Space Administration, April 2000.
- [2] JOT, J. M., LARCHER, E., WARUSFEL, O., “Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony”. In: *98<sup>th</sup> Convention of the AES*, Preprint 3980, Paris, France, February 1995.
- [3] HUOPANIEMI, J., *Virtual Acoustics and 3D Sound in Multimedia Signal Processing*. Ph.D. dissertation, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, 1999.
- [4] MERIMAA, J., *Analysis, Synthesis And Perception of Spatial Sound - Binaural Localization Modeling and Multichannel Loudspeaker Reproduction*. Ph.D. dissertation, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, August 2006. Report 77.
- [5] PULKKI, V., “Virtual Sound Source Positioning Using Vector Base Amplitude Panning”, *J. of the Audio Engineering Society*, v. 45, n. 6, pp. 456–466, July 1997.
- [6] CHENG, C. I., WAKEFIELD, G. H., “Introduction to Head-Related Transfer Functions (HRTFs): Representation in Time, Frequency and Space”, *J. of the Audio Engineering Society*, v. 49, n. 4, pp. 231–249, April 2001.
- [7] MARTIN, K. D., “Estimating Azimuth and Elevation from Interaural Differences”. In: *Proceedings of IEEE Workshop on Applications of Signal Processing to Acoustics and Audio*, pp. 96–99, New York, USA, October 1995.

- [8] BROWN, C. P., DUDA, R. O., “An Efficient HRTF Model for 3-D Sound”. In: *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 298–301, New York, USA, October 1997.
- [9] BEGAULT, D. R., “Auditory and Non-Auditory Factors That Potentially Influence Virtual Acoustic Imagery”. In: *The Proceedings of the AES 16<sup>th</sup> International Conference on Spatial Sound Reproduction*, pp. 13–26, Rovaniemi, Finland, April 1999.
- [10] 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.
- [11] VRIES, D., BOONE, M. M., “Wave Field Synthesis and Analysis using Array Technology”. In: *1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 15–18, New York, USA, October 1999.
- [12] CIPIC, “The CIPIC HRTF Database”, disponível em <http://interface.cipic.ucdavis.edu>, último acesso em Março de 2007.
- [13] LISTEN, “Listen HRTF Database”, disponível em <http://recherche.ircam.fr/equipes/salles/listen/index.html>, último acesso em Março de 2007.
- [14] GARDNER, B., MARTIN, K., *HRTF Measurement of a KEMAR Dummy-Head Microphone*, Report 280, MIT Media Lab Perceptual Computing, May 1994.
- [15] KAHANA, Y., “BEM simulation of ellipsoid HRTFs”, disponível em <http://www.isvr.soton.ac.uk/fdag/vap/ellipsoid/info.htm>, May 1999, último acesso em Setembro de 2007.
- [16] DUDA, R. O., MARTENS, W. L., “Range Dependence of the Response of a Spherical Head Model”, *J. Acoustical Society of America*, v. 104, n. 5, pp. 3048–3058, November 1998.

- [17] ALLEN, J. B., BERKLEY, D. A., “Image Method for Efficiently Simulating Small-Room Acoustics”, *J. Acoustical Society of America*, v. 65, n. 4, pp. 943–950, April 1979.
- [18] BORISH, J., “Extension of the Image Method to Arbitrary Polyhedra”, *J. Acoustical Society of America*, v. 75, n. 6, pp. 1827–1836, June 1984.
- [19] FREELAND, F. P., “Geração Eficiente de Som Tridimensional”, Tese de D.Sc., Universidade Federal do Rio de Janeiro, PEE, COPPE, Dezembro 2005.
- [20] ALGAZI, V. R., DUDA, RICHARD O. AN DURAISWAMI, R., GUMEROV, N. A., *et al.*, “Approximating the Head-Related Transfer Function using Simple Geometric Models of the Head and Torso”, *J. Acoustical Society of America*, v. 112, n. 5, pp. 2053–2064, November 2002.
- [21] LARCHER, V., JOT, J. M., WARUSFEL, O., “Study and Comparison of Efficient Methods for 3D Audio Spatialization Based on Linear Decomposition of HRTF Data”. In: *Audio Engineering Society 108<sup>th</sup> Convention*, Paris, France, January 2000. Paper 5097.
- [22] CHEN, J., VAN VEEN, B. D., 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.
- [23] DEVORE, J. L., *Probability and Statistics for Engineering and the Sciences*. Brooks/Cole, 1982.
- [24] OPPENHEIM, A., SCHAFER, R., BUCK, J., *Discrete-Time Signal Processing*. Prentice Hall, 1989.
- [25] OPPENHEIM, A., SCHAFER, R., *Digital Signal Processing*. 1 ed. Prentice Hall, 1975.
- [26] JOT, J.-M., WARDLE, S., “Approaches to binaural synthesis”. In: *105<sup>th</sup> Convention of the Audio Engineering Society*, San Francisco, USA, September 1998. Paper 4861.

- [27] GUNZI, A. S., FREELAND, F. P., BISCAINHO, L. W., *et al.*, “Modelagem Eficiente de HRTFs por Parametrização Espacial”. In: *Anais do V Congresso de Engenharia de Áudio da AES-Brasil*, pp. 82–88, Maio 2007.
- [28] FREELAND, F. P., BISCAINHO, L. W. P., DINIZ, P. S. R., “HRTF Interpolation Through Direct Angular Parameterization”. In: *ISCAS’2007–International Symposium on Circuits and Systems*, pp. 1823 – 1826, Los Angeles, USA, May 2007.
- [29] FARROW, C. W., “A Continuously Variable Digital Delay Element”. In: *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 2641–2643, Espoo, Finland, June 1988.
- [30] 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, September 2004.
- [31] BERNFELD, B., “Attempts for Better Understanding of the Directional Stereophonic Listening Mechanism”. In: *44<sup>th</sup> Convention of the Audio Engineering Society*, v. 21, Rotterdam, The Netherlands, May 1973. Paper C-4.
- [32] PULKKI, V., “Uniform Spreading of Amplitude Panned Virtual Sources”. In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 187 – 190, New York, USA, October 1999.
- [33] PULKKI, V., “Compensating Displacement of Amplitude-Panned Virtual Sources”. In: *AES 22<sup>th</sup> International Conference on Virtual, Synthetic and Entertainment Audio*, 2000.
- [34] PULKKI, V., *Spatial Sound Generation and Perception by Amplitude Panning Techniques*. Ph.D. dissertation, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, 2001.

# Apêndice A

## ITD - Trajetória Reta em 3D

Para o caso 3D, toma-se a caixa localizada num ângulo  $\theta$  no plano horizontal e eleva-se de um ângulo  $\phi$ , no plano mediano. Por se tratar de localização numa esfera, a reta entre a caixa e o centro geométrico da cabeça tem o mesmo comprimento tanto no caso 2D quanto no 3D.

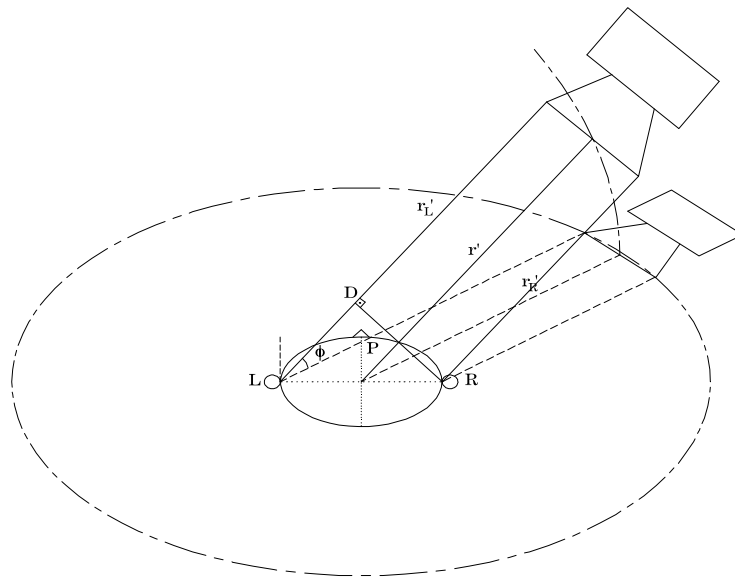


Figura A.1: Modelo de ITD com trajetória reta para o caso 3D.

Analogamente ao caso 2D, dado pela Figura 3.5, considera-se que as retas são paralelas. Quer-se tomar a diferença percorrida pelo som da caixa até as orelhas esquerda e direita. Para tal, traça-se uma reta perpendicular a  $r'_R$ , que corta a reta  $r'_L$  no ponto  $D$ . O ponto  $P$  se localiza no plano horizontal, na intersecção entre  $r_L$  e a perpendicular a  $r_R$ . O esquema simplificado da Figura A.1 está mostrado na

Figura A.2.

Nota-se que  $\overline{LR}$  é um diâmetro e o ângulo  $\widehat{LDR}$  tem  $90^\circ$ , por construção. Portanto, o ponto  $D$  também fica localizado exatamente sobre a circunferência de raio  $r$ .

Localiza-se, então, o ponto médio do segmento  $\overline{LP}$ , denominando-o  $M$ , conforme a Figura A.2. Sendo o triângulo  $LOP$  isósceles, o segmento  $\overline{OM}$  é perpendicular ao segmento  $\overline{LP}$ .

O ângulo  $\widehat{LRP}$  é igual a  $\theta$ , da mesma forma que na Figura 3.7. Portanto, o ângulo  $\widehat{LOP}$  é de  $2\theta$  e o ângulo  $\widehat{LOM}$  é igual a  $\theta$ , pela construção do triângulo  $LOM$ .

O comprimento do segmento  $\overline{OM}$  é obtido por geometria plana, e é dado por

$$\overline{OM} = r \cos \theta. \quad (\text{A.1})$$

O segmento  $\overline{OM}$  é perpendicular ao plano que contém o triângulo  $PLD$ . Portanto, qualquer reta deste plano é perpendicular ao segmento  $\overline{OM}$ . Em particular, o segmento  $\overline{DM}$  é perpendicular a  $\overline{OM}$ . Como  $\overline{OM} = r \cos \theta$  e  $\overline{OD} = r$ , obtém-se o comprimento de  $\overline{DM}$  como

$$\overline{DM} = r \sin \theta. \quad (\text{A.2})$$

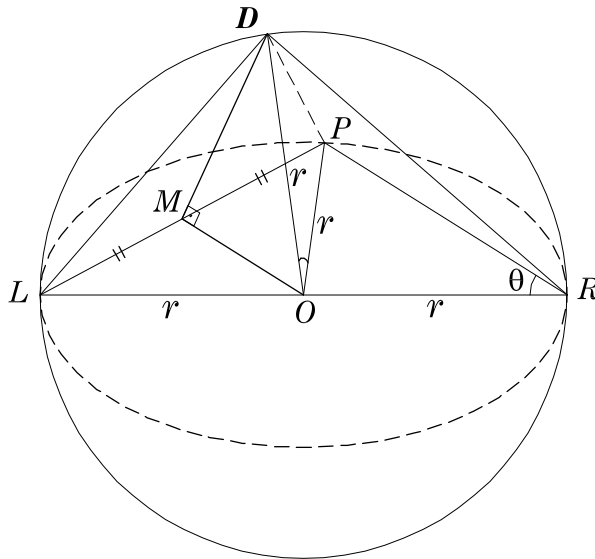


Figura A.2: Representação esquemática da Figura A.1.

O comprimento de  $\overline{PL}$  também pode ser calculado por geometria plana, notando que o triângulo  $PLO$  é isósceles, com ângulo  $\widehat{PLO}$  igual a  $2\theta$  e  $\overline{OP} =$

$$\overline{OL} = r.$$

$$\overline{PL} = 2r \operatorname{sen} \theta. \quad (\text{A.3})$$

Portanto, os segmentos  $\overline{ML}$ ,  $\overline{MP}$  e  $\overline{MD}$  têm o mesmo comprimento, sendo os ângulos do triângulo  $PDL$  conforme a disposição da Figura A.3.

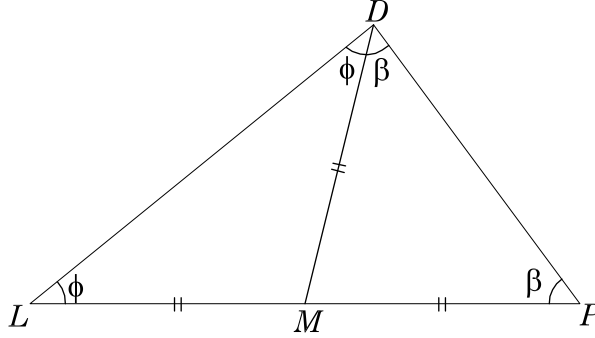


Figura A.3: Ângulos do triângulo ADL.

A soma dos ângulos internos de um triângulo em geometria plana é igual a  $180^\circ$ , portanto

$$2\phi + 2\beta = 180^\circ, \quad (\text{A.4})$$

resultando em

$$\phi + \beta = 90^\circ. \quad (\text{A.5})$$

Portanto, o ângulo  $LDP$  é reto. Sendo  $d$  o diâmetro do círculo, e  $2r = d$ , o segmento  $\overline{LD}$  é dado por

$$\overline{LD} = \overline{LP} \cos \phi = d \operatorname{sen} \theta \cos \phi. \quad (\text{A.6})$$

# Apêndice B

## Desenvolvimento da Lei das Tangentes

A ITD pode ser obtida a partir de

$$\text{ITD}_{\text{caixas}} = \text{ITD}_0, \quad (\text{B.1})$$

onde

$$\text{ITD}_{\text{caixas}} = T_{AL} - T_{AR} = \frac{AT_4 - BT_3}{A + B} - \frac{AT_1 + BT_2}{A + B}, \quad (\text{B.2})$$

e  $\text{ITD}_{\text{caixas}} = 0$ , já que o ouvinte acompanha a fonte virtual.

As equações acima implicam que

$$T_{AL} = T_{AR}. \quad (\text{B.3})$$

Os componentes de  $T_1$  a  $T_4$  podem ser obtidos por meios geométricos através da Figura 5.6, resultando no sistema a seguir.

$$\begin{aligned} T_1 &= \frac{\overline{ML'}}{c} = \frac{d}{c} \operatorname{sen} \frac{a}{2} \cos\left(a_0 - \frac{a}{2}\right) \\ T_2 &= \frac{(\overline{QL} + \overline{LN})}{c} = \frac{d}{c} \operatorname{sen} a_0 + \frac{d}{c} \operatorname{sen} \frac{a}{2} \cos\left(a_0 + \frac{a}{2}\right) \\ T_3 &= \frac{\overline{RT}}{c} = \frac{d}{c} \operatorname{sen} \frac{a}{2} \cos\left(a_0 + \frac{a}{2}\right) \\ T_4 &= \frac{(\overline{PR} - \overline{R'S})}{c} = \frac{d}{c} \operatorname{sen} a_0 - \frac{d}{c} \operatorname{sen} \frac{a}{2} \cos\left(a_0 - \frac{a}{2}\right). \end{aligned} \quad (\text{B.4})$$

Substituindo os valores da Equação (B.4) na Equação (B.2), tem-se

$$\begin{aligned} T_{AL}(A + B) \frac{c}{d} &= A \operatorname{sen} \frac{a}{2} \cos\left(a_0 - \frac{a}{2}\right) + B \operatorname{sen} a_0 + B \operatorname{sen} \frac{a}{2} \cos\left(a_0 + \frac{a}{2}\right) \\ T_{AR}(A + B) \frac{c}{d} &= A \operatorname{sen} a_0 - A \operatorname{sen} \frac{a}{2} \cos\left(a_0 - \frac{a}{2}\right) - B \operatorname{sen} \frac{a}{2} \cos\left(a_0 + \frac{a}{2}\right). \end{aligned} \quad (\text{B.5})$$



Considerando  $T_{AL} = T_{AR}$ ,

$$\frac{A}{B} = \frac{\text{sen } a_0 + 2 \text{sen } \frac{a}{2} \cos(a_0 + \frac{a}{2})}{\text{sen } a_0 - 2 \text{sen } \frac{a}{2} \cos(a_0 - \frac{a}{2})}. \quad (\text{B.6})$$

Desenvolvendo a expressão, chega-se a

$$\frac{A}{B} = \frac{\text{sen}(a_0 + a)}{\text{sen}(a_0 - a)}. \quad (\text{B.7})$$

$$\frac{A - B}{A + B} = \frac{\text{sen}(a_0 + a) - \text{sen}(a_0 - a)}{\text{sen}(a_0 + a) + \text{sen}(a_0 - a)}. \quad (\text{B.8})$$

$$\frac{A - B}{A + B} = \frac{\text{tg } a}{\text{tg } a_0}. \quad (\text{B.9})$$

# Apêndice C

## Expansão e Simplificação por Série de Taylor

Os sinais que chegam aos ouvidos esquerdo  $L$  e direito  $R$ , considerando o efeito composto dos três alto-falantes, são dados respectivamente por

$$L = A_L f(t - T_{A0} - T_{AL}) + B_L f(t - T_{B0} - T_{BL}) + C_L f(t - T_{C0} - T_{CL}) \quad (\text{C.1})$$

$$R = A_R f(t - T_{A0} - T_{AR}) + B_R f(t - T_{B0} - T_{BR}) + C_R f(t - T_{C0} - T_{CR}). \quad (\text{C.2})$$

Após a expansão em série de Taylor e o truncamento dos termos de grau maior do que 1, chega-se a

$$\begin{aligned} L &= A_L f(t) - A_L(T_{A0} + T_{AL})f'(t) + \\ &+ B_L f(t) - B_L(T_{B0} + T_{BL})f'(t) + \\ &+ C_L f(t) - C_L(T_{C0} + T_{CL})f'(t) \end{aligned} \quad (\text{C.3})$$

$$\begin{aligned} R &= A_R f(t) - A_R(T_{A0} + T_{AR})f'(t) + \\ &+ B_R f(t) - B_R(T_{B0} + T_{BR})f'(t) + \\ &+ C_R f(t) - C_R(T_{C0} + T_{CR})f'(t). \end{aligned} \quad (\text{C.4})$$

Reagrupando os termos, tem-se que

$$\begin{aligned} L &= (A_L + B_L + C_L) \left( f(t) - \right. \\ &\left. - f'(t) \frac{A_L(T_{A0} + T_{AL}) + B_L(T_{B0} + T_{BL}) + C_L(T_{C0} + T_{CL})}{A_L + B_L + C_L} \right). \end{aligned} \quad (\text{C.5})$$

e

$$R = (A_R + B_R + C_R) \left( f(t) - f'(t) \frac{A_R(T_{AO} + T_{AR}) + B_R(T_{BO} + T_{BR}) + C_R(T_{CO} + T_{CR})}{A_R + B_R + C_R} \right). \quad (C.6)$$

As Equações (C.5) e (C.6) equivalem a uma expansão em primeira ordem de série de Taylor das funções:

$$L = (A_L + B_L + C_L) f(t - \frac{A_L(T_{AO} + T_{AL}) + B_L(T_{BO} + T_{BL}) + C_L(T_{CO} + T_{CL})}{A_L + B_L + C_L}). \quad (C.7)$$

e

$$R = (A_R + B_R + C_R) f(t - \frac{A_R(T_{AO} + T_{AR}) + B_R(T_{BO} + T_{BR}) + C_R(T_{CO} + T_{CR})}{A_R + B_R + C_R}). \quad (C.8)$$

A partir das Equações (C.7) e (C.8) temos o seguinte valor de ILD:

$$\text{ILD} = \frac{A_L + B_L + C_L}{A_R + B_R + C_R} = 1. \quad (C.9)$$

A ITD é dada pela Equação (C.10).

$$\text{ITD} = \frac{A_L(T_{AO} + T_{AL}) + B_L(T_{BO} + T_{BL}) + C_L(T_{CO} + T_{CL})}{A_L + B_L + C_L} - \frac{A_R(T_{AO} + T_{AR}) + B_R(T_{BO} + T_{BR}) + C_R(T_{CO} + T_{CR})}{A_R + B_R + C_R}. \quad (C.10)$$

A primeira hipótese a ser adicionada é que se utilizará apenas a diferença em magnitude para a panoramização da fonte virtual. O atraso inicial inserido nos alto-falantes será zero, ou seja,  $T_{A0} = T_{B0} = T_{C0} = 0$ .

A segunda hipótese é a de que será considerada apenas a região de baixas frequências, ou seja,  $A_L = A_R$ ,  $B_L = B_R$  e  $C_L = C_R$ .

Considerando essas duas hipóteses, a Equação (C.7) permanece inalterada, mas a Equação (C.8) pode ser reformulada, obtendo-se, dessa forma, a Equação (C.11).

$$\text{ITD} = \frac{A(T_{AL} - T_{AR}) + B(T_{BL} - T_{BR}) + C(T_{CL} - T_{CR})}{A + B + C}. \quad (C.11)$$

# Apêndice D

## Equivalência entre a Lei dos Senos 3D e a Interpolação Triangular

A Equação (D.1) é a Lei dos Senos 3D, com as posições das caixas acústicas e da fonte virtual em função da caixa 1 e das distâncias relativas, conforme o esquema da Figura 5.9.

$$\begin{aligned} & A \operatorname{sen} \theta_1 \cos \phi_1 + \\ & B \operatorname{sen}(\theta_1 + \Delta\theta_2) \cos(\phi_1 + \Delta\phi_2) + \\ & C \operatorname{sen}(\theta_1 + \Delta\theta_3) \cos(\phi_1 + \Delta\phi_3) = \operatorname{sen}(\theta_1 + \delta\theta) \cos(\phi_1 + \delta\phi). \end{aligned} \quad (\text{D.1})$$

Expandindo as funções trigonométricas e aproximando  $\operatorname{sen} \alpha \approx \alpha$  e  $\cos \alpha \approx 1$ , chega-se a

$$\begin{aligned} & A \operatorname{sen} \theta_1 \cos \phi_1 + \\ & B(\operatorname{sen} \theta_1 + \Delta\phi_2 \cos \theta_1)(\cos \phi_1 - \Delta\phi_2 \operatorname{sen} \phi_1) + \\ & C(\operatorname{sen} \theta_1 + \Delta\phi_3 \cos \theta_1)(\cos \phi_1 - \Delta\phi_3 \operatorname{sen} \phi_1) = \\ & = (\operatorname{sen} \theta_1 + \delta\theta \cos \theta_1)(\cos \phi_1 - \delta\phi \operatorname{sen} \phi_1). \end{aligned} \quad (\text{D.2})$$

Nota-se que os valores de seno e cosseno em  $\theta_1$  e  $\phi_1$  formam uma base, de modo que, para haver igualdade, cada um dos termos multiplicativos da base deve

ser igual a zero, conforme o sistema de equações

$$\begin{aligned}
(A + B + C - 1) \operatorname{sen}(\theta_1) \cos(\phi_1) &= 0 \\
(B\Delta\phi_2 + C\Delta\phi_3 - \delta\phi) \operatorname{sen}(\theta_1) \cos(\phi_1) &= 0 \\
(B\Delta\theta_2 + C\Delta\theta_3 - \delta\theta) \cos(\theta_1) \cos(\phi_1) &= 0 \\
(B\Delta\theta_2\Delta\phi_2 + C\Delta\theta_3\Delta\phi_3 - \delta\theta\delta\phi) \cos(\theta_1) \operatorname{sen}(\phi_1) &= 0.
\end{aligned} \tag{D.3}$$

Como os valores de  $\Delta\phi$  e  $\Delta\theta$  são pequenos, pode-se desprezar o valor do produto entre eles. Portanto, chega-se a três equações e três incógnitas:

$$\begin{aligned}
A + B + C &= 1 \\
B\Delta\phi_2 + C\Delta\phi_3 &= \delta\phi \\
B\Delta\theta_2 + C\Delta\theta_3 &= \delta\theta.
\end{aligned} \tag{D.4}$$

Resolvendo o sistema de equações, chega-se a

$$\begin{aligned}
C &= \frac{\delta\phi\Delta\theta_2 - \delta\theta\Delta\phi_2}{\Delta\theta_2\Delta\phi_3 - \Delta\theta_3\Delta\phi_2} \\
B &= \frac{\delta\theta\Delta\phi_3 - \delta\phi\Delta\theta_3}{\Delta\theta_2\Delta\phi_3 - \Delta\theta_3\Delta\phi_2} \\
A &= 1 - B - C.
\end{aligned} \tag{D.5}$$

As Equações (D.5) mostram a solução para o caso geral. Entretanto, nos conjuntos de medidas de HRTFs costuma-se adotar normalmente o passo de elevação fixo, o que significa adotar  $\Delta\phi_2 = 0$ . Disso, obtém-se a formulação final:

$$\begin{aligned}
C &= \frac{\delta\phi}{\Delta\phi_3} \\
B &= \frac{\delta\theta}{\Delta\theta_2} - C\frac{\Delta\theta_3}{\Delta\theta_2} \\
A &= 1 - B - C.
\end{aligned} \tag{D.6}$$

# Apêndice E

## Equivalência entre a Lei dos Senos 3D e a Interpolação Bilinear

A partir da Figura 5.10, formula-se a Lei dos Senos 3D para a grade regular.

$$\begin{aligned} & A \operatorname{sen} \theta_A \cos \phi_A + \\ & B \operatorname{sen}(\theta_A + \theta_{\text{grade}}) \cos \phi_A + \\ & C \operatorname{sen}(\theta_A + \theta_{\text{grade}}) \cos(\phi_A + \phi_{\text{grade}}) + \\ & D \operatorname{sen} \theta_A \cos(\phi_A + \phi_{\text{grade}}) = \operatorname{sen}(\theta_A + \delta\theta) \cos(\phi_A + \delta\phi). \end{aligned} \quad (\text{E.1})$$

Expandindo as funções trigonométricas e aproximando  $\operatorname{sen} \alpha \approx \alpha$  e  $\cos \alpha \approx 1$ , chega-se a

$$\begin{aligned} & A \operatorname{sen} \theta_A \cos \phi_A + \\ & B \operatorname{sen} \theta_A \cos \phi_A + B \theta_{\text{grade}} \cos \theta_A \cos \phi_A + \\ & C \operatorname{sen} \theta_A \cos \phi_A - C \phi_{\text{grade}} \operatorname{sen} \theta_A \operatorname{sen} \phi_A + C \theta_{\text{grade}} \cos \theta_A \cos \phi_A \\ & \quad - C \theta_{\text{grade}} \phi_{\text{grade}} \cos \theta_A \operatorname{sen} \phi_A + \\ & D \operatorname{sen} \theta_A \cos \phi_A - D \phi_{\text{grade}} \operatorname{sen} \theta_A \operatorname{sen} \phi_A = \\ & = \operatorname{sen} \theta_A \cos \phi_A - \delta\phi \operatorname{sen} \theta_A \operatorname{sen} \phi_A + \delta\theta \cos \theta_A \cos \phi_A + \\ & \quad - \delta\theta\delta\phi \cos \theta_A \operatorname{sen} \phi_A. \end{aligned} \quad (\text{E.2})$$

Os valores de seno e cosseno em  $\theta_A$  e  $\phi_A$  formam uma base, de modo que, para haver igualdade, cada um dos termos multiplicativos da base deve ser igual a

zero, resultando em

$$\begin{aligned}A + B + C + D &= 1 \\B\theta_{\text{grade}} + C\theta_{\text{grade}} &= \delta\theta \\-C\theta_{\text{grade}} - D\theta_{\text{grade}} &= -\delta\phi \\C\theta_{\text{grade}}\phi_{\text{grade}} &= \delta\theta\delta\phi.\end{aligned}\tag{E.3}$$

Denominando  $d\theta = \frac{\delta\theta}{\theta_{\text{grade}}}$  e  $d\phi = \frac{\delta\phi}{\phi_{\text{grade}}}$ , tem-se que a solução do sistema de equações é dada por

$$\begin{aligned}D &= (1 - d\theta)d\phi \\C &= d\theta d\phi \\B &= d\theta(1 - d\phi) \\A &= (1 - d\theta)(1 - d\phi).\end{aligned}\tag{E.4}$$