# Auralização de fontes sonoras móveis usando HRTFs

Gustavo Henrique Montesião de Sousa

Dissertação apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do título
de
Mestre em Ciências

Programa: Ciência da Computação
Orientador: Prof. Dr. Marcelo Gomes de Queiroz

Durante o desenvolvimento deste trabalho, o autor recebeu auxílio financeiro de CNPq e FAPESP

São Paulo, maio de 2010

# Auralização de fontes sonoras móveis usando HRTFs

Esta versão definitiva da dissertação
contém as correções e alterações sugeridas pela
Comissão Julgadora durante a defesa realizada
por Gustavo Henrique Montesião de Sousa em 29/4/2010.

Comissão Julgadora:

- Prof. Dr. Marcelo Gomes de Queiroz (orientador) - IME-USP
- Prof. Dr. Miguel Arjona Ramirez - EP-USP
- Prof. Dr. Francisco José Fraga da Silva - UFABC

# Agradecimentos

Agradeço primeiramente a minha família, pelo amor incondicional que vivemos. Aos meus amigos, particularmente os que fiz nesses anos todos em que estudei em São Paulo, por tudo o que passamos e por tudo o que me ensinaram. Ao professor Marcelo Queiroz, pela paciência e pelas boas idéias partilhadas. Agradeço, por fim, a cada brasileiro, por pagar meus estudos com tanta luta e sofrimento.

# Resumo

Este trabalho tem por objetivo desenvolver ferramentas que permitam gerar em fones-de-ouvido o efeito psicoacústico de fontes sonoras locomovendo-se no espaço, por meio da auralização do sinal monofônico original. Embora a auralização binaural possa ser feita empregando variações de atraso (chamadas ITD – *interaural time difference*, ou diferença de tempo interaural) e de intensidade (chamadas ILD – *interaural level difference*, ou diferença de nível interaural) entre os canais, melhores resultados psicoacústicos podem ser obtidos ao se utilizar filtros digitais conhecidos como HRTFs (*head related transfer functions*, ou funções de transferência relativas à cabeça). Uma HRTF insere no sinal monofônico informações que possibilitam ao sistema auditivo identificá-lo como proveniente de uma direção específica, direção esta que é única para cada HRTF. Para posicionar uma fonte estática em uma direção específica, bastaria, então, filtrar o sinal original pela HRTF da direção desejada. Se, no entanto, for desejável que a fonte se locomova em uma trajetória contínua, um número infinitamente grande de filtros seria necessário. Como eles são, normalmente, obtidos empiricamente, um número arbitrariamente alto deles não está disponível. Disso surge a necessidade de técnicas de interpolação de HRTFs, que possibilitem gerar os filtros intermediários não disponíveis.

Este trabalho apresenta três novas técnicas de interpolação de HRTFs, para assim alcançar o objetivo de auralizar fontes sonoras móveis: a *interpolação triangular*, que é uma técnica de interpolação linear baseada na técnica de panorama sonoro VBAP (*vector-based amplitude panning*, ou panorama sonoro baseado em vetores); o *método das movimentações discretas*, que busca explorar o limiar de percepção do nosso sistema auditivo para, com isso, gerar uma técnica extremamente barata computacionalmente; e a *interpolação espectral*, que altera continuamente as estruturas das HRTFs para gerar filtros interpolados. São apresentadas também as implementações feitas dessas novas técnicas desenvolvidas, bem como os testes numéricos realizados para medir sua eficácia.

**Palavras-chave:** auralização, HRTF, fontes sonoras móveis.

# Abstract

The goal of this work is the development of tools that allow simulating through headphones the psychoacoustic effect of sound sources moving in space, by the auralisation of the original monophonic signals. Although binaural auralisation can be implemented using variations in delays (called ITD – interaural time difference) and in intensities (called ILD – interaural level difference) among channels, better psychoacoustic results can be achieved using digital filters known as HRTFs (head related transfer functions). A HRTF inserts in the monophonic signal information that allow the auditory system to perceive this signal to be as if coming from a specific direction, which is unique for each single HRTF. Thus, to position a static sound source at a specific direction, filtering the original signal with the HRTF from the desired direction would be enough. Nevertheless, if it is desired that the sound source moves in a continuous trajectory, an infinitely large amount of filters would be necessary. Since they are usually obtained by measurements, such an arbitrarily large amount of them is not available. In this case, HRTF interpolation techniques that generate intermediary filters must be used.

This work presents three new HRTF interpolation techniques in order to auralise moving sound sources: the *triangular interpolation*, a linear interpolation technique based on the VBAP amplitude panning technique; the *discrete movements method*, an extremely efficient technique that exploits the auditory system's limitations in perceiving very small changes in direction; and the *spectral interpolation*, that alters continuously the structures of the HRTFs to generate interpolated filters. Implementations of these techniques are discussed and numerical tests are also presented.

**Keywords:** auralisation, HRTF, moving sound sources.

# Sumário

**Lista de Figuras** **ix**

**Lista de Tabelas** **xi**

**1 Introdução** **1**

1.1 Auralização e o emprego de HRTFs . . . . . 3

1.2 O banco de dados CIPIC . . . . . 4

1.3 Trabalhos relacionados . . . . . 6

**2 Interpolações temporais** **9**

2.1 *Amplitude panning* . . . . . 9

2.2 VBAP . . . . . 10

2.3 Interpolação triangular . . . . . 11

2.4 Método das movimentações discretas . . . . . 15

**3 Interpolação espectral** **19**

3.1 Duas maneiras de interpolar pólos e zeros . . . . . 23

3.2 Associação de pólos e de zeros . . . . . 27

3.3 Construindo um banco de HRTFs IIR . . . . . 28

3.4 Quanto à configuração dos filtros . . . . . 31

**4 Implementação e testes** **35**

4.1 Implementação e testes da interpolação triangular . . . . . 36

4.2 Implementação e testes do método das movimentações discretas . . . . . 43

4.3 Implementação e testes da interpolação espectral . . . . . 44

**5 Conclusões** **57**

**A Gráficos da interpolação triangular** **61**

**B Gráficos da interpolação espectral** **67**

**Referências Bibliográficas** **79**

# Lista de Figuras


1.1 Desenho esquemático dos pontos de medição das HRTFs do banco de dados CIPIC. . . 5

2.1 Exemplo de como uma interpolação feita sem correção dos atrasos relativos pode distorcer o resultado final. . . . . 14

2.2 A distância das fontes entre si e a percepção de direção. . . . . 16

3.1 Interpolação linear de passa-banda . . . . . 20

3.2 Interpolação espectral de passa-banda . . . . . 22

3.3 Exemplo de interpolação espectral com dois filtros simples. . . . . 24

3.4 O sistema tratado por Kalman. . . . . 29

3.5 Exemplo de interpolação espectral. . . . . 33

4.1 Disposições dos triângulos usadas durante os testes . . . . . 37

4.2 Comparação da interpolação para o ouvido esquerdo e direito. . . . . 38

4.3 Resultado da interpolação triangular para o ouvido esquerdo . . . . . 39

4.4 Continuação dos dados da figura 4.3. . . . . 40

4.4 Continuação dos dados da figura 4.3. . . . . 41

4.5 Diferenças das superfícies espaciais de resposta em freqüência para o ouvido direito . . 42

4.6 Evolução dos erros relativos das interpolações em função da freqüência. . . . . 43

4.7 Movimentações discretas. Ouvido esquerdo. Valores em decibéis. . . . . 45

4.8 Comparação dos erros das técnicas de interpolação triangular e das movimentações discretas . . . . . 46

4.9 Distribuição dos filtros segundo pólos e zeros reais . . . . . 47

4.10 Filtros interpolados com 14 pólos e 14 zeros. Direção (-5°, 90°) . . . . . 51

4.11 Filtros interpolados com 14 pólos e 14 zeros. Direção (10°, 73,125°) . . . . . 52

4.12 Filtros interpolados com 12 pólos e 12 zeros. Direção (-30°, -39,375°). . . . . 53

4.13 Diagramas de pólos e zeros para a direção (-30°, -39,375°) . . . . . 54

4.14 Filtros interpolados com 14 pólos e 14 zeros. Direção (40°, -5,625°). . . . . 55

# Lista de Tabelas

3.1 Distribuição dos filtros com 2 pólos e 2 zeros reais . . . . . 32

4.1 Erros da interpolação espectral. Filtros com 10 pólos e 10 zeros. . . . . 48

4.2 Erros da interpolação espectral. Filtros com 12 pólos e 12 zeros. . . . . 49

4.3 Erros da interpolação espectral. Filtros com 14 pólos e 14 zeros. . . . . 50

4.4 Erros da interpolação espectral para algumas configurações de triângulo para a direção (40°, -5,625°) e filtros de 14 pólos e 14 zeros. . . . . 54

## Capítulo 1

# Introdução

A pesquisa sobre espacialização sonora vem adquirindo importância crescente desde a última metade do século passado, tanto do ponto de vista tecnológico quanto artístico [Roa96]. Seu emprego ganha notoriedade em áreas que vão desde sistemas de imersão e realidade virtual, passando por jogos de computador, até a composição de peças musicais, área em que o emprego da espacialização sonora insere-se como uma dimensão a mais de exploração artística [Sto61, Xen92].

Uma maneira simples e direta de se gerar o efeito de espacialização sonora é fazendo uso de um conjunto de alto-falantes adequadamente posicionados ao redor do ouvinte: o sinal sonoro seria, então, emitido pelo alto-falante que estivesse posicionado na direção desejada. Se nos detivéssemos apenas nesses elementos de utilização desta idéia, veríamos rapidamente que seu leque de aplicações ficaria bastante reduzido, porquanto seria necessário possuir um alto-falante para cada direção de onde se deseja que provenha o sinal sonoro. Para evitar o emprego de um número excessivamente grande de alto-falantes, tornando seu uso bastante oneroso ou mesmo inviável, algumas técnicas foram desenvolvidas visando a gerar o efeito de direcionalidade da fonte sonora a partir de um número reduzido de alto-falantes. Uma delas, muito simples e amplamente utilizada, é a chamada *amplitude panning*, que busca a criação de um panorama sonoro por meio do controle do ganho do sinal em cada alto-falante [Moo90]. Uma variante dessa técnica é a VBAP, a qual permite uma generalização da *amplitude panning* para diversas disposições de alto-falantes, inclusive tridimensionais [Pul01].

O principal senão do emprego dessas técnicas está em que elas pressupõem que o ouvinte está localizado no centro da configuração de alto-falantes. Para qualquer posição diferente, o efeito gerado já não será mais o desejado. Isso dificulta seu emprego para uma audiência com várias pessoas: o campo sonoro não seria uniforme dentro do espaço de escuta, gerando variações nem sempre desejáveis

entre o que é captado por cada ouvinte.

Para aplicações nas quais é desejável um controle maior sobre o campo sonoro no entorno de cada ouvinte, pode ser mais interessante a utilização de fones-de-ouvido. Com eles, ganha-se controle sobre a informação sonora que chegará a cada um dos indivíduos que compõem a audiência. Outra vantagem é não ter que se preocupar com a influência do espaço real de escuta no sinal sonoro, adquirindo-se, desta maneira, um domínio mais preciso sobre o resultado final do campo sonoro.

Embora alternativas simples, como *amplitude panning* e a inserção de ITD tenham sido aplicadas a fones-de-ouvido com o intuito de gerar o efeito de espacialização, resultados muito mais ricos psicoacusticamente podem ser conseguidos por meio da transformação do sinal emitido pelos fones-de-ouvido através de filtros digitais chamados HRTFs, que normalmente são obtidos por meio de medições. Esses filtros carregam consigo as informações acerca das mudanças sofridas pelo sinal sonoro desde sua emissão até sua captação pelo ouvido e, assim, podem ser utilizados para incutir essas mesmas mudanças no sinal emitido por cada um dos alto-falantes de um fone-de-ouvido. A seção 1.1 deste trabalho faz uma apresentação de tais filtros.

A HRTF imprime no sinal sonoro as transformações que ele sofreria caso se propagasse em um espaço de escuta real através de um padrão de propagação determinado. Tendo em conta que sons oriundos de direções diferentes sofrem transformações distintas, cada HRTF está associada a uma única direção, não podendo uma mesma HRTF ser usada para auralizar fontes sonoras virtuais posicionadas em direções distintas.

Uma vez que o objetivo deste trabalho é simular fontes sonoras locomovendo-se no espaço em trajetórias contínuas, um sem-número de HRTFs seria necessário para produzir esses deslocamentos. Todavia, bancos de dados de HRTFs como o CIPIC [ADTA01], embora possuam grande quantidade de HRTFs, não as possuem em número ilimitado, o que é natural, dado que as HRTFs de tais bancos são obtidas empiricamente. Decorre disto a necessidade de técnicas de interpolação de HRTFs, técnicas que permitiriam a mudança gradual de uma HRTF para outra, preenchendo as lacunas deixadas pelos bancos de dados, de modo a tornar possível o emprego desses filtros na produção de curvas suaves para os deslocamentos das fontes sonoras.

Este problema da interpolação de HRTFs vem sendo estudado há algum tempo, e propostas diferentes já foram apresentadas (cf., como exemplo, [SHLV99, FBD02, FBD04]). O presente trabalho exibirá três novas técnicas de interpolação de HRTFs, que trazem vantagens sobre as técnicas já disponíveis.

Na próxima seção, será mostrado como as HRTFs podem ser utilizadas na espacialização sonora em fones-de-ouvido. Depois, será apresentado o banco de HRTFs chamado CIPIC, que é o banco utilizado nas implementações das técnicas desenvolvidas aqui. Por fim, na última seção é feita uma breve descrição de outras técnicas de interpolação de HRTFs, disponíveis na literatura.

## 1.1 Auralização e o emprego de HRTFs

Um sinal sonoro sofre alterações quando percorre o caminho desde sua emissão por uma fonte até sua captação pelo ouvinte. Entre essas transformações, incluem-se absorções, reflexões e refrações pelo ambiente e pelo corpo do ouvinte, destacando-se, aqui, o papel desempenhado pelas orelhas e sua geometria peculiar [Bla97, Sor05]. Além dessas alterações na própria estrutura da onda sonora, o ouvinte capta um outro tipo de informação decorrente do deslocamento do sinal sonoro através do espaço de propagação: as diferenças de fase (ITD) e de intensidade (ILD) dos sinais que chegam a seu ouvido direito e esquerdo. A inserção artificial desses efeitos em um sinal sonoro, com o intuito de reconstruir ou modelar uma cena sonora real, é chamada auralização [Eve00, Vor07].

O conjunto das informações mencionadas acima é percebido pelo sistema sensorial do ouvinte e por ele utilizado para determinar a direção da fonte sonora, e é tão importante para tal determinação, que, sendo possível imprimir no sinal sonoro original essas mesmas transformações por meio de alguma técnica de auralização, o sistema auditivo gerará a ilusão de direcionalidade desejada. Tais transformações artificiais podem ser conseguidas por meio de funções de transferência chamadas HRTFs, comumente obtidas por meio de medições empíricas [Bla97].

Pensando em termos de processamento de sinais, uma HRTF é, idealmente, um filtro que modifica um sinal de entrada $x(n)$ incutindo-lhe as mesmas transformações supracitadas, de modo a tornar a saída $y(n)$ igual ao sinal sonoro que chega ao ouvido. Sendo $h(n)$ a resposta impulsiva desse filtro, a saída desejada será, então, calculada através da convolução do sinal de entrada com $h(n)$, definida por:

$$y(n) = (x * h)(n) = \sum_{k=0}^{N-1} h(k)x(n-k), \tag{1.1}$$

sendo $N$ a duração da resposta impulsiva $h(n)$ medida empiricamente.

Normalmente, as HRTFs são obtidas medindo-se a resposta impulsiva (captada em cada ouvido de um indivíduo) a um estímulo curto, chamada HRIR (*Head Related Impulse Response*, ou resposta impulsiva relativa à cabeça) [ADTA01]. Por serem armazenadas como uma resposta impulsiva finita,

as HRTFs são filtros FIR, pois cada amostra da resposta impulsiva coincide com um coeficiente do filtro. A relação entre uma HRTF e sua HRIR pode ser estabelecida por meio da transformada de Fourier, cuja versão discreta é definida nos seguintes termos:

$$\mathscr{F}\{x\}(k) = \begin{cases} \sum_{n=0}^{N-1} x(n)e^{-\frac{2\pi i}{N}nk}, & 0 \leq k \leq N-1 \\ 0, & c.c. \end{cases} \tag{1.2}$$

Uma HRTF é, assim, o resultado da aplicação da transformada de Fourier à HRIR correspondente:

$$H(k) = \mathscr{F}\{h\}(k). \tag{1.3}$$

Por sua vez, uma HRIR pode ser obtida a partir de sua HRTF através da transformada inversa de Fourier, cuja versão discreta é como segue:

$$\mathscr{F}^{-1}\{f\}(n) = \begin{cases} \frac{1}{N}\sum_{k=0}^{N-1} f(k)e^{\frac{2\pi i}{N}kn}, & 0 \leq n \leq N-1 \\ 0, & c.c. \end{cases} \tag{1.4}$$

A HRIR é calculada, então, a partir da HRTF da seguinte maneira:

$$h(n) = \mathscr{F}^{-1}\{H\}(n). \tag{1.5}$$

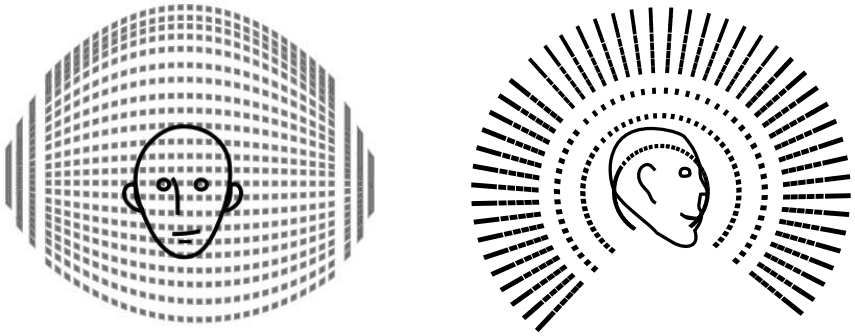
Uma HRTF não é somente indissociável de uma direção do espaço, ela também está relacionada de maneira única ao indivíduo no qual ela foi medida, do que decorre que os resultados psicoacústicos do uso desses filtros só serão ótimos se cada HRTF for usada em auralizações para aquele indivíduo no qual ela foi medida, e somente para ele. Esse fato levou ao desenvolvimento de técnicas que permitissem a personalização de HRTFs genéricas baseada em medidas antropométricas [Sor05]. Não obstante isso, é comum o emprego de HRTFs de uso geral em pesquisas de auralização, dada a facilidade de acesso a bancos de dados públicos de HRTFs (nos quais se podem obter HRTFs medidas para um certo grupo de indivíduos) e também dada a dificuldade de obtenção de HRTFs personalizadas na maioria dos casos.

## 1.2 O banco de dados CIPIC

Diversos bancos de dados de HRTFs estão disponíveis publicamente. Tais bancos são geralmente resultado de medições feitas em indivíduos específicos, o que torna o resultado de sua aplicação

ótimo apenas para esses indivíduos. Como as técnicas desenvolvidas aqui são pensadas para serem utilizadas em aplicações de uso geral, a extração de um banco de HRTFs para cada potencial usuário dessas aplicações não é normalmente viável. Entretanto, como dito na seção anterior, é comum o emprego de HRTFs não individualizadas com resultados satisfatórios, ainda que não ótimos. Essas considerações levaram à decisão de se utilizar, neste trabalho, um desses bancos públicos de HRTFs, chamado CIPIC [ADTA01]. Ele contém respostas impulsivas de 1250 direções feitas em 45 indivíduos distintos (27 homens, 16 mulheres e mais dois conjuntos de medições realizadas com um boneco KEMAR [GM95]).

As HRTFs foram medidas tendo os indivíduos se posicionado no centro de uma esfera de 1 m de raio com diâmetro alinhado com o eixo interauricular do indivíduo (cf. figura 1.1). Alto-falantes de 5,8 cm de raio foram colocados em diversas posições ao longo da esfera (aproximadamente nos pontos exibidos na figura 1.1). Os canais auditivos dos indivíduos foram bloqueados e microfones instalados em seus ouvidos foram utilizados para captar o sinal emitido pelos alto-falantes. Os sinais captados foram digitalizados a uma freqüência de amostragem de 44,1 kHz e a 16 bits de resolução. Cada HRIR tem 200 amostras, correspondendo a aproximadamente 4,5 ms.



(a) Visão frontal (b) Visão lateral

Figura 1.1: Desenho esquemático dos pontos de medição das HRTFs do banco de dados CIPIC.

A localização dos alto-falantes foi especificada em coordenadas polares interaurais [MA84]. Esse sistema de coordenadas é levemente diferente do sistema de coordenadas polares vertical tradicional, como pode ser observado na figura 1.1 (que contém um esquema da disposição dos pontos de medição), e em especial na figura 1.1b. A elevação foi distribuída uniformemente em passos de 5,625°, variando

de -45° a 230,625°. Já com relação ao azimute, foram realizadas medições em -80°, -65°, -55°, de -45° a 45° em passos de 5° e, por fim, em 55°, 65° e 80°.

Adicionalmente às medições das HRIRs, o CIPIC disponibiliza também um conjunto de medições antropométricas para cada indivíduo, de modo a tornar possível a investigação da influência do corpo de cada pessoa no resultado da HRTF. Foram tomadas medições em 17 pontos distintos da cabeça e do torso, e em 10 pontos distintos das orelhas. Esse conjunto de medições permite considerar ainda a possibilidade de apresentar ao usuário de um sistema que implemente as técnicas desenvolvidas aqui os 45 conjuntos de HRTFs, juntamente com suas medidas antropométricas, e deixar que o usuário escolha o conjunto cujas medidas mais se aproximem das medidas do corpo do usuário, diminuindo, desta maneira, os efeitos de se utilizarem HRTFs não individualizadas.

## 1.3 Trabalhos relacionados

A interpolação de HRTFs vem suscitando diversos estudos, e diferentes soluções já foram propostas. A seguir, serão apresentadas algumas dessas soluções, que são representativas do que se está desenvolvendo nessa área, o que servirá para compreender melhor as inovações das propostas apresentadas aqui.

A interpolação bilinear [Huo99, SHLV99] gera uma HRIR interpolada através de uma interpolação linear das quatro HRIRs mais próximas. Ou seja, em um banco de HRTFs com passo de $\theta_{grid}$ para o azimute e $\phi_{grid}$ para a elevação, para gerar uma HRIR interpolada para uma certa direção ($\theta$,$\phi$), este método usa as quatro HRIRs mais próximas dessa direção ($\theta$,$\phi$) (digamos, $h_A$, $h_B$, $h_C$ e $h_D$), gerando assim a nova HRIR por meio da equação

$$\hat{h}(n) = (1 - c_\theta)(1 - c_\phi)h_A(n) + c_\theta(1 - c_\phi)h_B(n) + c_\theta c_\phi h_C(n) + (1 - c_\theta)c_\phi h_D(n),$$

sendo $c_\theta = (\theta \bmod \theta_{grid})/\theta_{grid}$ e $c_\phi = (\phi \bmod \phi_{grid})/\phi_{grid}$, em que $a \bmod b$ é o resto da divisão de $a$ por $b$, definido como $x - y\lfloor x/y \rfloor$.

Runkle e colaboradores [RBW95] estudaram uma técnica de interpolação baseada na movimentação de pólos e zeros no plano complexo. Se duas HRTFs $H_a$ e $H_b$ são filtros IIR (*infinite impulse response*, ou de resposta impulsiva infinita), um modelo IIR para a HRTF interpolada $\hat{H}$ é obtida por meio do deslocamento dos pólos e zeros de $H_a$ até os pólos e zeros de $H_b$ através de uma trajetória determinada por uma busca por gradiente: minimizando a discrepância entre o modelo IIR e a HRTF interpolada $\hat{H}$ (que se supõe conhecida).

Exemplificando, considere que $H_a$ e $H_b$ são filtros de apenas um par de pólos complexos-conjugados ($p_a$ e $p_a^*$ para $H_a$ e $p_b$ e $p_b^*$ para $H_b$). Partindo de $p_a$ para $p_b$, uma busca pelo método do gradiente definiria um caminho entre $p_a$ e $p_b$. O filtro interpolado seria aquele definido por um pólo posicionado nesse caminho entre $p_a$ e $p_b$ e pelo seu complexo-conjugado. Determinado o caminho, há, contudo, diferentes pontos onde se poderia posicionar o pólo interpolado. O melhor ponto é escolhido por meio de um critério psicoacústico.

Esse método possui semelhanças com a interpolação espectral desenvolvida no capítulo 3. Entretanto, ele só lida com interpolações em segmentos de reta ligando duas HRTFs, não possuindo extensão para superfícies, como é o caso da interpolação espectral. Contudo, a grande desvantagem desse método é a necessidade de se conhecer, *a priori*, a HRTF que se está querendo interpolar.

Em *Moving Sound Source Synthesis for Binaural Electroacoustic Music Using Interpolated Head-Related Transfer Functions (HRTFs)* [CW01], é apresentada uma técnica de interpolação baseada nas Superfícies Espaciais de Resposta em Freqüência (*Spatial Frequency Response Surfaces*, ou SFRS) [CW99]. Neste tipo de visualização, escolhem-se freqüências para as quais serão feitos gráficos e, então, as magnitudes das respostas em freqüência das HRTFs são exibidas como função do azimute e da elevação para cada uma das freqüências escolhidas. As superfícies para cada freqüência são geradas interpolando linearmente as magnitudes em função do azimute e da elevação. Nesse trabalho, a interpolação usada para gerar as superfícies usa sempre 3 pontos, executando triangulações no azimute e elevação.

Uma HRTF interpolada para uma certa direção ($\theta$,$\phi$) seria, então, obtida extraindo-se os pontos ($\theta$,$\phi$) de cada uma das SFRS, cada ponto definindo o valor de uma amostra da resposta em freqüência da HRTF interpolada. Como o processo de suavização das superfícies é feito usando uma interpolação linear com triângulos, esta técnica tem certas semelhanças com a interpolação triangular apresentada na seção 2.3.

*Martin e McAnally* [MM07] desenvolveram uma técnica de interpolação de HRTFs que utiliza os espectros dos filtros. Diferentemente da interpolação espectral, apresentada no capítulo 3 deste trabalho, sua técnica executa uma interpolação linear das magnitudes e fases das HRTFs, o que é equivalente a uma interpolação de HRIRs (no domínio do tempo), devido à linearidade da Transformada de Fourier (cf. discussão apresentada no início do capítulo 3). E, ao contrário das técnicas desenvolvidas aqui, eles apenas lidam com interpolação em segmentos de reta ligando duas HRTFs.

Feita esta introdução, nos próximos capítulos serão apresentadas as técnicas desenvolvidas neste

trabalho e suas implementações, segundo a seguinte estrutura: no capítulo 2, serão exibidas duas técnicas de interpolação de HRTFs no domínio do tempo, chamadas de interpolação triangular e de método das movimentações discretas. No capítulo 3, será vista a técnica de interpolação de HRTFs baseada em modelos IIR. E, no capítulo 4, serão discutidos aspectos das implementações e serão apresentados testes objetivos comparando as diversas técnicas.

## Capítulo 2

# Interpolações temporais

Os dois próximos capítulos (este e o capítulo 3) visam a apresentar as técnicas desenvolvidas e estudadas neste trabalho. A divisão da exposição em dois diferentes capítulos busca evidenciar uma diferença essencial dessas técnicas na abordagem ao problema da interpolação de HRTFs. Tal diferença consiste no modo como as técnicas tratam os filtros em interpolação: em um caso, os *coeficientes* dos filtros são usados para gerar o filtro interpolado; em outro, são usados seus *pólos e zeros*. A primeira abordagem será apresentada no presente capítulo (*Interpolações temporais*), enquanto que a segunda será apresentada no capítulo seguinte (*Interpolação espectral*).

As técnicas de interpolação temporal que serão apresentadas aqui são duas: uma chamada *interpolação triangular* e a outra chamada *método das movimentações discretas*. A interpolação triangular é baseada em uma técnica tradicional de espacialização multicanal chamada VBAP. Depois dela, será apresentado o método das movimentações discretas, que se vale das limitações do sistema auditivo em perceber pequenas diferenças de variação da direção da fonte sonora.

Para apresentar a interpolação triangular, vamos começar introduzindo na seção 2.1 a técnica de *amplitude panning*, da qual o VBAP é um caso particular. A seção 2.3 apresentará a técnica de interpolação triangular e a seção 2.4 exibirá o método das movimentações discretas.

### 2.1 *Amplitude panning*

A técnica de *amplitude panning* (*panorama sonoro por amplitude*, em uma tradução livre para o português) é uma técnica tradicional de posicionamento virtual de fontes sonoras utilizada quando se tem um conjunto de alto-falantes devidamente instalados ao redor da cabeça do ouvinte. Para criar a ilusão de que a fonte sonora esteja posicionada em algum ponto no plano que contém os alto-falantes,

a técnica propõe um balanceamento das amplitudes do som oriundo desses alto-falantes. Vale dizer, sendo $x(t)$ o sinal sonoro original, o sinal $x_n(t)$ oriundo do alto-falante $n$, será obtido por [Pul01]:

$$x_n(t) = g_n x(t),\ 1 \leq n \leq N, \tag{2.1}$$

sendo $g_n$ o ganho no alto-falante $n$, e $N$ o número de alto-falantes utilizados. A percepção desejada de direção da fonte sonora é conseguida, então, através da escolha correta dos ganhos $g_n$.

## 2.2 VBAP

A técnica de VBAP é um tipo particular de *amplitude panning* que oferece um método para se calcular os ganhos da equação 2.1. É aplicável tanto para configurações bidimensionais quanto tridimensionais de alto-falantes. No caso tridimensional, ela pode ser formulada do seguinte modo [Pul01]: em um sistema de coordenadas cuja origem é a posição do ouvinte, sejam os alto-falantes $n$, $m$ e $k$ dispostos em uma superfície esférica com centro na origem e que não se encontrem, os três, no equador da esfera. nas direções $\boldsymbol{l^n} = (l_1^n, l_2^n, l_3^n)^t$, $\boldsymbol{l^m} = (l_1^m, l_2^m, l_3^m)^t$ e $\boldsymbol{l^k} = (l_1^k, l_2^k, l_3^k)^t$, Seja $\boldsymbol{p} = (p_1, p_2, p_3)^t$ a direção desejada para a fonte virtual. Está-se considerando que $\boldsymbol{p}$ está no triângulo formado pelos pontos $l^n$, $l^m$ e $l^k$, e que o efeito da distância da fonte ao ouvinte será tratado posteriormente à determinação dos ganhos dos alto-falantes. Satisfeitos os pressupostos acima, $\boldsymbol{l^n}$, $\boldsymbol{l^m}$ e $\boldsymbol{l^k}$ formam uma base de $\mathbb{R}^3$. Portanto, o vetor $\boldsymbol{p}$ pode ser escrito como uma combinação linear deles:

$$\boldsymbol{p} = g_n\ \boldsymbol{l^n} + g_m\ \boldsymbol{l^m} + g_k\ \boldsymbol{l^k}, \tag{2.2}$$

ou, em notação matricial,

$$\boldsymbol{p} = \boldsymbol{L}_{nmk}\ \boldsymbol{g}, \tag{2.3}$$

sendo $\boldsymbol{g} = (g_n,\ g_m,\ g_k)^t$ e $\boldsymbol{L}_{nmk} = [\boldsymbol{l^n}\ \boldsymbol{l^m}\ \boldsymbol{l^k}]$.

Conhecendo-se $\boldsymbol{p}$, $\boldsymbol{l^n}$, $\boldsymbol{l^m}$ e $\boldsymbol{l^k}$, os valores de $g_n$, $g_m$ e $g_k$ podem, então, ser obtidos por

$$\boldsymbol{g} = \boldsymbol{L}_{nmk}^{-1}\ \boldsymbol{p} = \begin{bmatrix} l_1^n & l_1^m & l_1^k \\ l_2^n & l_2^m & l_2^k \\ l_3^n & l_3^m & l_3^k \end{bmatrix}^{-1} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}. \tag{2.4}$$

A técnica de VBAP então afirma que o vetor $\boldsymbol{g}$ pode ser identificado com os ganhos dos alto-

falantes da equação 2.1.

Em outras palavras, tendo uma certa configuração tridimensional de alto-falantes, a técnica de VBAP permite criar a sensação de que uma fonte sonora está posicionada em qualquer direção contida na região definida pelos alto-falantes procedendo da seguinte forma:

- estabelecendo qual trio de alto-falantes define o menor triângulo que contém a direção na qual se deseja posicionar a fonte;
- calculando o valor de $\boldsymbol{g}$ para esse trio de alto-falantes dada a direção da fonte desejada;
- usando os valores das componentes de $\boldsymbol{g}$ como os ganhos de cada alto-falante.

## 2.3 Interpolação triangular

A interpolação triangular proposta neste trabalho foi elaborada como uma transposição da técnica de VBAP de sua aplicação original (balanceamento de ganhos de alto-falantes) para uma aplicação binaural, na qual são utilizadas HRTFs em lugar dos alto-falantes [dSQ09]. Na técnica de VBAP variam-se as intensidades dos alto-falantes para se criar a ilusão de posicionamento da fonte sonora; na técnica de interpolação triangular, um sinal filtrado por uma HRTF correspondente a uma certa direção $\boldsymbol{i}$ e reproduzido em um fone-de-ouvido é tratado da mesma forma que um sinal oriundo de um alto-falante posicionado nessa mesma direção $\boldsymbol{i}$. Pode-se, disso, extrair as seguintes considerações:

- sendo a HRIR $\boldsymbol{h_i}(n)$ aquela correspondente à direção $\boldsymbol{i}$ e $\boldsymbol{s}(n)$ o sinal monofônico original, o sinal $\boldsymbol{s_i}(n) = (\boldsymbol{h_i} * \boldsymbol{s})(n)$ é o correspondente, no contexto binaural da interpolação triangular, ao sinal oriundo do alto-falante $i$ da técnica de VBAP;
- se na técnica de VBAP usam-se, concomitantemente, três alto-falantes dispostos em direções linearmente independentes para gerar o som de uma fonte virtual em um espaço de escuta tridimensional, aqui também se utilizarão três HRIRs correspondentes a direções linearmente independentes para gerar o som de uma fonte virtual que será reproduzido nos fones-de-ouvido.

Dado isso, a técnica se deriva imediatamente: sendo $\boldsymbol{i}$, $\boldsymbol{j}$ e $\boldsymbol{k}$ as direções não-colineares, o sinal $\hat{\boldsymbol{s}}$, resultado da interpolação, será obtido como

$$\hat{\boldsymbol{s}}(n) = g_i \boldsymbol{s_i}(n) + g_j \boldsymbol{s_j}(n) + g_k \boldsymbol{s_k}(n), \tag{2.5}$$

sendo $\boldsymbol{g} = (g_i,\, g_j,\, g_k)$ obtido por meio da equação 2.4 e $\boldsymbol{s_m}(n) = (\boldsymbol{h_m} * \boldsymbol{s})(n)$ para $m \in \{i,j,k\}$.

Vale aqui fazer uma simplificação, para economia de recursos computacionais. Utilizando a propriedade de linearidade da operação de convolução,

$$\begin{gathered}\hat{\boldsymbol{s}}(n) = g_i\boldsymbol{s_i}(n) + g_j\boldsymbol{s_j}(n) + g_k\boldsymbol{s_k}(n) \Leftrightarrow \\ \hat{\boldsymbol{s}}(n) = g_i(\boldsymbol{h_i} * \boldsymbol{s})(n) + g_j(\boldsymbol{h_j} * \boldsymbol{s})(n) + g_k(\boldsymbol{h_k} * \boldsymbol{s})(n) \Leftrightarrow \\ \hat{\boldsymbol{s}}(n) = [(g_i\boldsymbol{h_i} + g_j\boldsymbol{h_j} + g_k\boldsymbol{h_k}) * \boldsymbol{s}](n).\end{gathered} \tag{2.6}$$

Definindo

$$\hat{\boldsymbol{h}}(n) = g_i\boldsymbol{h_i}(n) + g_j\boldsymbol{h_j}(n) + g_k\boldsymbol{h_k}(n), \tag{2.7}$$

chega-se a

$$\hat{\boldsymbol{s}}(n) = (\hat{\boldsymbol{h}} * \boldsymbol{s})(n). \tag{2.8}$$

Assim obtém-se, por fim, a técnica:

1. dividir a superfície abarcada pelo banco de dados CIPIC (cf. seção 1.2) em triângulos cujos vértices coincidam com as direções cobertas pelas HRTFs do banco;
2. dada uma direção $\boldsymbol{p}$ na qual se quer posicionar a fonte virtual, determinar qual triângulo (dentre os estabelecidos no item 1) a contém;
3. calcular o vetor $\boldsymbol{g}$ conforme equação 2.4, usando as posições dos vértices do triângulo determinado no item 2;
4. calcular $\hat{\boldsymbol{h}}$ segundo a equação 2.7, usando o vetor $\boldsymbol{g}$ calculado no item 3 e as HRTFs associadas aos vértices do triângulo estabelecido no item 2;
5. proceder à convolução de $\hat{\boldsymbol{h}}$ com o sinal monofônico original (conforme equação 2.8).

O cálculo de $\hat{\boldsymbol{h}}$ no passo 4 e a sua utilização para a filtragem do sinal no passo 5 podem ser interpretados como se o passo 4 levasse à produção de uma nova HRIR, a qual, por ser combinação linear de três HRIRs próximas, herda características das três. A proposição aqui, novamente em analogia com a técnica de VBAP, é a de que o sinal produzido no passo 5 é capaz de recriar a sensação subjetiva de direcionalidade associada à direção $\boldsymbol{p}$.

As HRTFs do banco de dados CIPIC são obtidas através da medição de resposta a estímulos sonoros provenientes de direções específicas. Esses estímulos podem demorar frações de tempo diferentes para chegarem a cada uma das orelhas do ouvinte, dependendo da direção de onde provêm, o que gera uma diferença no atraso relativo entre as respostas impulsivas do banco. Calcular a interpolação triangular usando essas HRIRs com diferentes atrasos relativos pode distorcer o resultado final (ver figura 2.1).

Esse problema pode ser contornado através da utilização de HRIRs de fase mínima correspondentes às HRIRs do banco de dados CIPIC. Um sistema de fase mínima pode ser definido como aquele cujos pólos e zeros, todos eles, têm valor absoluto menor que 1, ou seja, cujos pólos e zeros residam dentro do círculo unitário [OSB89]. Sistemas de fase mínima possuem propriedades muito úteis, duas das quais são de interesse aqui.

A primeira das propriedades relacionadas a um sistema de fase mínima é a de que todo sistema racional pode ser decomposto em um sistema de fase mínima e um sistema passa-tudo:

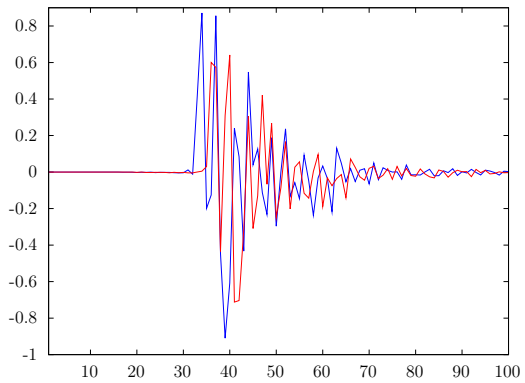
$$H(z) = H_{min}(z)H_{pt}(z). \tag{2.9}$$

Uma vez que a magnitude da resposta em freqüência de um sistema passa-tudo é constante, a equação acima nos diz que a magnitude da resposta em freqüência de um sistema racional é igual à magnitude da resposta em freqüência do sistema de fase mínima correspondente. Assim, as informações da magnitude da resposta em freqüência de um sistema são preservadas no sistema de fase mínima correspondente.

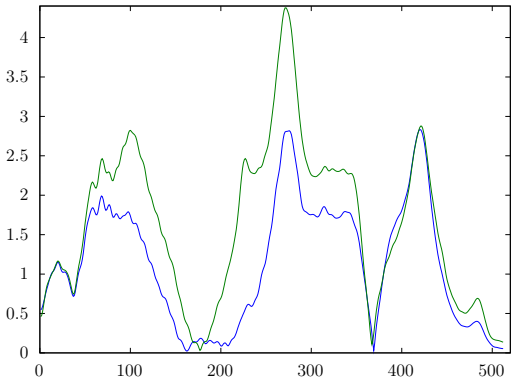
A segunda das propriedades relacionadas a um sistema de fase mínima é que sua energia tende a se concentrar perto do instante inicial. Pra ser mais preciso, sendo $h_{min}(n)$ a resposta impulsiva de um sistema de fase mínima e $h(n)$ a resposta impulsiva de qualquer outro sistema racional que possua a mesma magnitude da resposta em freqüência do sistema de fase mínima, tem-se que

$$\sum_{n=0}^{k} |h_{min}(n)|^2 \geq \sum_{n=0}^{k} |h(n)|^2, \forall k \in \mathbb{N}. \tag{2.10}$$

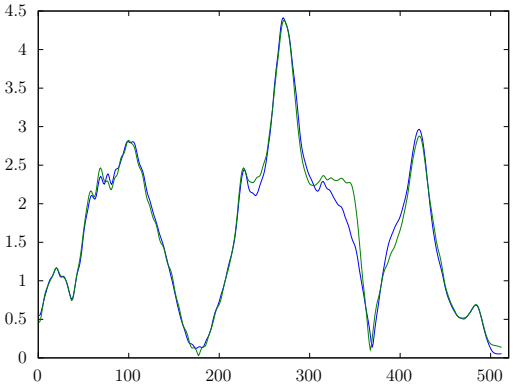
Ou seja, dada uma certa magnitude de resposta em freqüência, o sistema de fase mínima correspondente, dentre todos os sistemas que possuem tal magnitude, é aquele que mais tende a concentrar a informação de sua resposta impulsiva nas vizinhanças da origem.




(a) As duas HRIRs com diferença de atrasos.




(b) Em azul, a magnitude da resposta em freqüência da média das duas HRIRs acima, feita sem corrigir os atrasos relativos; em verde, a magnitude da resposta em freqüência correta.




(c) Em azul, a magnitude da resposta em freqüência da média dos filtros de fase mínima correspondentes às duas HRIRs acima; em verde, a magnitude da resposta em freqüência correta.

Figura 2.1: Exemplo de como uma interpolação feita sem correção dos atrasos relativos pode distorcer o resultado final.

Portanto, dentre todos os sistemas que preservam a informação da magnitude da resposta em freqüência de uma HRTF, o sistema de fase mínima é aquele que terá o menor atraso. Por isso é útil utilizar tais filtros para contornar o problema dos atrasos relativos na interpolação triangular.

Filtros de fase mínima podem ser obtidos de filtros FIR extraindo os zeros do polinômio da equação do filtro e substituindo todos os zeros $q$ tais que $|q| > 1$ por $1/q^*$ [OSB89], sendo $q^*$ o complexo-conjugado de $q$.

## 2.4 Método das movimentações discretas

O método das movimentações discretas difere substancialmente da interpolação triangular apresentada na seção 2.3. Ela não é baseada na técnica de *amplitude panning* como a interpolação triangular, nem tampouco é uma técnica de interpolação de diferentes HRTFs. Sua motivação é a idéia, muito simples, de explorar o limiar de percepção de diferentes direções de fontes sonoras pelo sistema auditivo [Gra86, GHE03, HBBS88, LM94, PM89, PT88, PS90].

O exemplo a seguir pretende tornar mais clara essa idéia. Considere dois alto-falantes dispostos nas direções $\boldsymbol{d_1}$ e $\boldsymbol{d_2}$ em um sistema de coordenadas cuja origem seja a cabeça do ouvinte. Tal disposição é mostrada na figura 2.2. Em um primeiro momento, as duas direções são suficientemente distantes uma da outra (configuração (a) da figura), o que faz o ouvinte distinguir claramente o som proveniente de uma e outra direções. Agora considere que os alto-falantes vão se aproximando até chegarem às direções $\boldsymbol{d_3}$ e $\boldsymbol{d_4}$. Nessa nova configuração, as fontes estão demasiado próximas. Tão próximas, que é possível que o sistema sensorial do ouvinte já não consiga diferenciar a direção $\boldsymbol{d_3}$ da direção $\boldsymbol{d_4}$, fazendo-o crer que já não há duas direções distintas, mas tão-somente uma única direção de onde provém o som. Em uma tal situação, não sendo possível diferenciar a direção $\boldsymbol{d_3}$ de $\boldsymbol{d_4}$, tampouco será possível, ao ouvinte, perceber um deslocamento da fonte de $\boldsymbol{d_3}$ para $\boldsymbol{d_4}$. Se isso ocorre, qualquer técnica de interpolação entre as HRTFs das direções $\boldsymbol{d_3}$ e $\boldsymbol{d_4}$ torna-se um processo desnecessário do ponto de vista psicoacústico.

Essa observação é o princípio que norteia o método das movimentações discretas. Tendo em mãos um banco de HRTFs suficientemente denso, a interpolação de HRTFs vizinhas passaria a ser um processo irrelevante.

O método, considerando essas premissas, fica assim:

- dada uma direção ($\theta_1$,$\phi_1$), utilize como filtro a HRTF mais próxima a ela (chamada aqui $\boldsymbol{H_1}$);

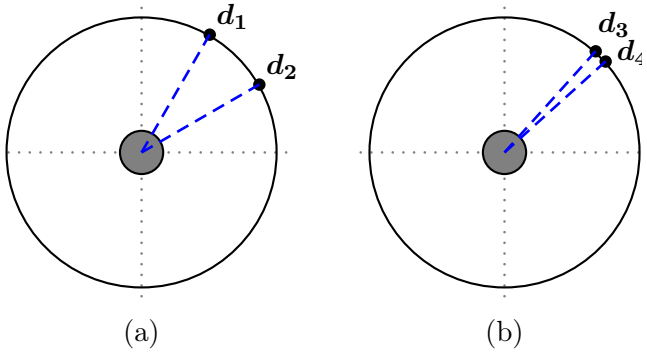



Figura 2.2: A distância das fontes entre si e a percepção de direção: em (a), as fontes estão distantes o bastante uma da outra para que o ouvinte perceba claramente que os dois sons provêm de direções distintas; em (b), as fontes estão muito mais próximas, o suficiente para que o ouvinte não seja capaz de distinguir o som oriundo do alto-falante em $\boldsymbol{d_3}$ do som oriundo do alto-falante em $\boldsymbol{d_4}$, tudo se passando, para ele, como se o som viesse de uma única direção, e não mais de duas.

- se a fonte se deslocar da direção ($\theta_1$,$\phi_1$) para uma direção ($\theta_2$,$\phi_2$) e, durante essa trajetória, a HRTF mais próxima a ela passar a ser outra — por exemplo, $\boldsymbol{H_2}$ —, troque o filtro utilizado, de $\boldsymbol{H_1}$ para $\boldsymbol{H_2}$;
- para evitar cliques audíveis, durante a troca efetue um *crossfade* de uns poucos instantes entre os sinais filtrados pela $\boldsymbol{H_1}$ e pela $\boldsymbol{H_2}$.

Essa é uma técnica simples e extremamente barata computacionalmente. Sua principal desvantagem com relação à interpolação triangular é seu universo de aplicação mais restrito, dada a exigência de uma malha densa de HRTFs.

Sob condições ideais (o sinal posicionado exatamente em frente ao ouvinte, e composto por uma banda larga de freqüências) o chamado *mínimo ângulo audível* (ou, na sigla em inglês, MAA, de *minimum audible angle*) é de cerca de 1° [HBBS88, Gra86, LM94], valor muito inferior ao passo de amostragem do azimute e da elevação nos bancos de HRTFs disponíveis. Em particular, o CIPIC não possui amostragem inferior a 5°, e isso apenas em algumas regiões do azimute. Entretanto, esse resultado vale apenas para fontes sonoras estáticas. Uma vez que este trabalho tem por objetivo a auralização de fontes móveis, deve ser considerado o limiar de percepção para fontes em movimento.

Neste caso, o mínimo ângulo audível (que, para fontes em movimento, é chamado de MAMA, do inglês *minimum audible movement angle*) fica entre 2° e 5° sob condições ideais (velocidade da fonte inferior a 20°/s; fonte com uma banda larga de freqüências) [Gra86, PT88, PM89]. Já no plano vertical, com fontes exatamente em frente ao ouvinte, o MAA sob condições ótimas é de cerca de 4° [PS90, Bla97] e o MAMA é de cerca de 6° [SP90, GHE03].

O resultado do MAMA é bastante dependente da velocidade da fonte sonora. À velocidade de 128°/s ele é algo entre 50% e 100% maior que a velocidades entre 8°/s e 16°/s [PT88], nunca ficando abaixo de 6° a 128°/s.

Vê-se que, sob certas condições, é possível aplicar o método das movimentações discretas a bancos de HRTFs já existentes, como o CIPIC. Seria possível, por exemplo, utilizar o CIPIC com o método caso a fonte se desloque rapidamente, ou caso se restrinja a movimentação da fonte sonora a um plano vertical, no qual o MAMA não é inferior a 6°, já que o passo do ângulo de elevação no CIPIC é de 5,625°. Se, contudo, a trajetória da fonte deixar o plano vertical, ainda que passando para um plano levemente menos inclinado, os resultados podem já não ser bons, pois é sabido que em trajetórias diagonais, mesmo a inclinações quase verticais, como 80°, o MAMA já diminui significativamente [PS90, GHE03]. Portanto, caso seja desejável uma maior liberdade para as movimentações da fonte sonora, um banco com medições mais finas faz-se necessário.

## Capítulo 3

# Interpolação espectral

No capítulo precedente, foram apresentadas algumas propostas de interpolação de HRTFs. Uma delas, a interpolação triangular, foi elaborada baseando-se em uma técnica de espacialização multicanal chamada VBAP. A analogia com tal técnica permitiu que se chegasse a uma interpolação que é, simplificadamente, uma soma ponderada de três HRIRs vizinhas, ou seja, de três HRTFs em sua representação temporal. Assim, para fazer uma fonte sonora variar de um ponto a outro dentro de um triângulo definido por três HRIRs, na interpolação triangular variam-se continuamente os pesos da soma ponderada dessas HRIRs.

A interpolação triangular pertence à classe das interpolações lineares, por aplicar apenas transformações lineares às HRTFs usadas no processo. Há algumas peculiaridades no comportamento espectral dos filtros interpolados por esses métodos que ainda não foram examinadas neste trabalho, mas que devem ser consideradas durante a auralização de fontes móveis, conforme se verá a seguir.

As figuras 3.1a e 3.1b exibem as respostas em freqüência de dois filtros passa-banda bastante simples, com 2 pólos e 2 zeros cada. Suponha que se deseje efetuar uma transição suave do primeiro para o segundo filtro. Ao se fazer isso empregando uma interpolação triangular, os filtros intermediários gerados não apresentam comportamento passa-banda (ver figura 3.1c).

O exemplo acima nos possibilita adquirir uma informação importante sobre os processos de interpolação linear, nos quais está inserida a interpolação triangular: eles podem não conservar algumas características dos filtros originais; os filtros gerados pela interpolação podem possuir um comportamento espectral distinto do comportamento dos filtros que estão sendo usados para a interpolação. Em decorrência disso, uma variação contínua de um filtro para outro usando a interpolação triangular pode não estar associada a uma variação contínua das características espectrais dos filtros, tais como

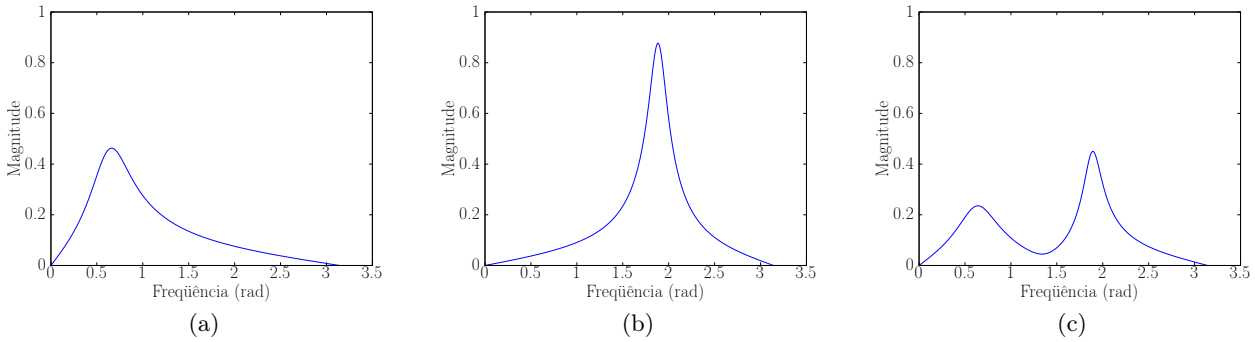



Figura 3.1: Exemplo de interpolação linear de filtros passa-banda. Nos gráficos (a) e (b) vêem-se os filtros usados no processo de interpolação. O gráfico (c) exibe o resultado da interpolação linear desses dois filtros.

bandas de ampliação ou rejeição de freqüências.

É sabido que algumas das informações utilizadas pelo sistema auditivo para determinar a localização de fontes sonoras estão relacionadas com a movimentação de picos e vales no espectro das HRTFs [Beg94]. Ou seja, não são apenas as características estáticas das HRTFs que devem ser levadas em consideração, mas também o modo como elas evoluem em função do deslocamento da fonte sonora. Uma vez que o objetivo primeiro deste trabalho é a auralização de fontes locomovendo-se no espaço, essas informações que o sistema auditivo extrai do movimento da fonte precisam ser levadas em consideração. Ora, são justamente essas informações que as técnicas lineares de interpolação podem não preservar.

Para entender melhor o porquê desse comportamento, é útil notar o modo como a interpolação triangular trata as HRTFs que estão sendo interpoladas. Para ela, nenhuma informação acerca das características dos filtros é usada. A única informação necessária são as coordenadas da fonte virtual, para poder calcular os pesos usados durante a interpolação. Nenhum conhecimento sobre as HRTFs, sobre seu espectro e seu comportamento, precisa ser usado durante a técnica, porque ela não modifica os filtros que são interpolados, apenas os pesos. Vimos acima que, procedendo assim, certas características dos filtros podem não ser preservadas.

É útil, portanto, poder usar uma abordagem diferente. Em vez de a alteração limitar-se aos

*pesos* da soma ponderada, mantendo intactas as HRTFs originais, é bom poder variar os próprios *espectros* das HRTFs, alterar continuamente a estrutura mesma da resposta em freqüência de um filtro, moldando-a, até que ela assuma a conformação de um seu vizinho. Este capítulo é dedicado a propor uma nova técnica de interpolação de HRTFs denominada *interpolação espectral*, que procura fazer exatamente isso [dSQ09].

Para alterar continuamente a estrutura da resposta em freqüência das HRTFs não é suficiente substituir as HRIRs por HRTFs na equação do filtro interpolado da interpolação triangular (equação 2.7), produzindo a seguinte equação:

$$HRTF_{int}(n) = g_1 HRTF_1(n) + g_2 HRTF_2(n) + g_3 HRTF_3(n). \tag{3.1}$$

Ainda que assim se consiga uma interpolação de espectros, a interpolação executada dessa maneira é idêntica àquela processada pela interpolação triangular, dada a linearidade da transformada de Fourier:

$$\begin{aligned} &g_1 HRTF_1 + g_2 HRTF_2 + g_3 HRTF_3 = g_1 \mathscr{F}\{HRIR_1\} + g_2 \mathscr{F}\{HRIR_2\} + g_3 \mathscr{F}\{HRIR_3\} = \\ &\mathscr{F}\{g_1 HRIR_1\} + \mathscr{F}\{g_2 HRIR_2\} + \mathscr{F}\{g_3 HRIR_3\} = \mathscr{F}\{g_1 HRIR_1 + g_2 HRIR_2 + g_3 HRIR_3\}, \end{aligned} \tag{3.2}$$

sendo, aqui, $\mathscr{F}\{HRIR_n\}$ a transformada de Fourier da enésima HRIR.

Interpolar espectros à maneira de uma interpolação triangular é, deste modo, equivalente a calcular o espectro da interpolação triangular correspondente. Essa interpolação (que nada mais é que a própria interpolação triangular), portanto, manterá o mesmo comportamento exibido no começo do capítulo, de não preservar certas características dos filtros.

Para uma interpolação diferente das vistas até aqui, faz-se necessária uma representação alternativa para as HRTFs, de modo a se obterem novos elementos para serem usados na interpolação, os quais permitam manipular as HRTFs em sua própria estrutura, e não mais apenas seus pesos. Se, no exemplo da figura 3.1, o filtro intermediário fosse gerado não por meio de uma interpolação linear, mas definindo que seus pólos e zeros estejam em uma posição intermediária entre os pólos e zeros dos filtros originais, sua resposta em freqüência se comportaria como desejado, como um filtro passa-banda. Isso é o que mostra a figura 3.2, em que os mesmos filtros da figura 3.1 (exibidos novamente nas figuras 3.2a e 3.2b) são usados agora para gerar o filtro da figura 3.2c. Esse filtro interpolado é gerado, desta vez, posicionando seus pólos e zeros em uma localização intermediária (cf. figuras 3.2d a 3.2f). Note que o filtro interpolado, gerado dessa forma, preserva as características dos filtros

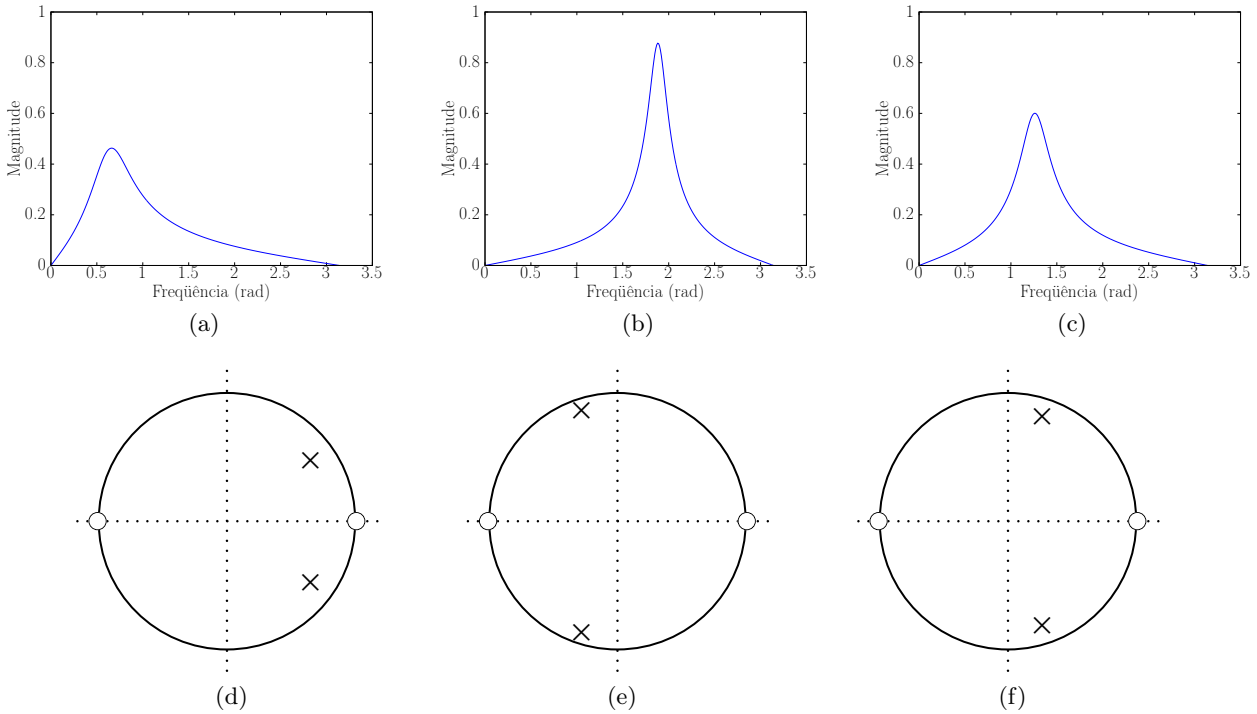



Figura 3.2: Neste exemplo, os mesmos filtros da figura 3.1 são novamente usados para gerar um filtro intermediário (figura (c)). Desta vez, contudo, ele é gerado posicionando seus pólos e zeros em uma localização intermediária com relação aos pólos e zeros dos filtros originais (figuras (d) a (f)). O filtro da figura (a) possui zeros em $-1$ e 1 e pólos em 0,8 $e^{i\pi/5}$ e 0,8 $e^{-i\pi/5}$. Já o filtro da figura (d) possui zeros em $-1$ e 1 e pólos em 0,9 $e^{3i\pi/5}$ e 0,9 $e^{-3i\pi/5}$. O filtro interpolado possui, assim, zeros em $-1$ e 1 e pólos em 0,85 $e^{2i\pi/5}$ e 0,85 $e^{-2i\pi/5}$.

originais, possuindo o comportamento que, intuitivamente, se esperaria de um filtro intermediário.

A interpolação espectral parte dessas observações para gerar uma técnica nova, bastante diferente das interpolações lineares. Para tanto, ela pressupõe que as HRTFs utilizadas são filtros IIR (*Infinite Impulse Response*, ou de resposta impulsiva infinita, em português) e, assim, do mesmo modo como foi feito no exemplo da figura 3.2, faz variarem continuamente as posições dos pólos e zeros constituintes dos filtros, deslocando-os das posições em que estão em um filtro para as posições em que se apresentam no outro, moldando a estrutura do primeiro filtro até que esta passe a ter a configuração do segundo.

Essa variação dos pólos e zeros é conseguida através da interpolação dos pólos e zeros dos filtros originais usados pela técnica. Duas variantes dessa interpolação de pólos e zeros são apresentadas a seguir, uma que usa a representação *polar* dos zeros e pólos, e outra que usa sua representação *cartesiana*.

## 3.1 Duas maneiras de interpolar pólos e zeros

A interpolação dos pólos e zeros dos filtros através de sua representação polar se dá da seguinte maneira: considere duas HRTFs — a $H_a(n)$ e a $H_b(n)$ — associadas a duas direções próximas, $\boldsymbol{a}$ e $\boldsymbol{b}$. Suponha, para efeitos didáticos, que cada uma dessas HRTFs seja um filtro IIR de apenas um par de pólos complexos-conjugados e nenhum zero não-trivial. Sejam $p_a$ e $p_a^*$ os pólos da $H_a(n)$, e $p_b$ e $p_b^*$ os pólos da $H_b(n)$. Dado que tais pólos são números complexos, eles podem ser escritos da seguinte forma:

$$\begin{aligned} p_a &= r_a e^{i\theta_a} \\ p_a^* &= r_a e^{-i\theta_a} \\ p_b &= r_b e^{i\theta_b} \\ p_b^* &= r_b e^{-i\theta_b}. \end{aligned} \tag{3.3}$$

Para este caso, querendo deslocar uma fonte sonora virtual da direção $\boldsymbol{a}$ para a direção $\boldsymbol{b}$, na variante polar da interpolação espectral altera-se linearmente o módulo $r_a$ até ele atingir o valor de $r_b$, fazendo o mesmo com o argumento $\theta_a$, variando-o em direção a $\theta_b$, percorrendo, deste modo, uma trajetória elíptica no plano complexo. Procedendo assim, a configuração dos pólos de $H_a(n)$ é alterada continuamente até ficar igual à configuração dos pólos de $H_b(n)$. Dito de maneira explícita, o filtro interpolado é aquele definido pelos dois pólos calculados segundo a seguinte equação:

$$\begin{aligned} \hat{p} &= (g_1 r_a + g_2 r_b) e^{i(g_1\theta_a + g_2\theta_b)} \\ \hat{p}^* &= (g_1 r_a + g_2 r_b) e^{-i(g_1\theta_a + g_2\theta_b)}, \end{aligned} \tag{3.4}$$

sendo que $g_1$ e $g_2$ são pesos variando entre 0 e 1 e $g_1 + g_2 = 1$. Sua função de transferência, obtida a partir desses pólos, é a seguinte:

$$\hat{H}(z) = \frac{1}{(1 - \hat{p}z^{-1})(1 - \hat{p}^* z^{-1})}. \tag{3.5}$$

Como os pólos produzidos nesse processo são um par complexo-conjugado, é suficiente calcular apenas um deles, sendo o outro derivado deste por conjugação. Portanto, a partir daqui será feita menção

apenas a um dos elementos do par complexo-conjugado, estando implícita a obtenção do outro por meio de conjugação.

A equação 3.5 mostra que o deslocamento dos pólos do filtro gera uma transformação não-linear em seu espectro. Por conseguinte, tal procedimento é diferente daquele executado na interpolação triangular. Se lá se trabalha com soma ponderadas de HRTFs, aqui se alteram estruturas de HRTFs. A figura 3.3 ilustra o presente exemplo para valores específicos de pólos e dois valores diferentes para os pares de pesos da equação 3.4.

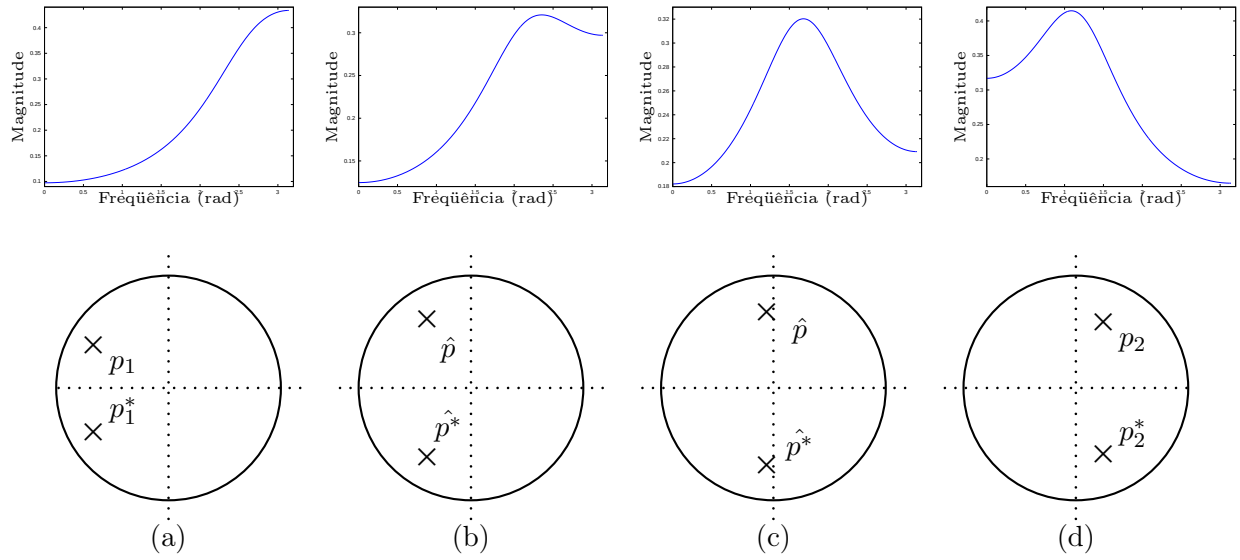



Figura 3.3: Exemplo de interpolação espectral usando dois filtros simples de apenas um par de pólos complexos-conjugados cada um. Seguindo o exemplo do texto acima, e sendo o caso (a) o filtro da direção $\boldsymbol{a}$ e o caso (d) o filtro da direção $\boldsymbol{b}$, o caso (b) é obtido ao atribuir os valores $g_1 = 2/3$ e $g_2 = 1/3$ aos pesos da equação 3.4. De modo similar, o caso (c) é obtido ao atribuir aos pesos os valores $g_1 = 1/3$ e $g_2 = 2/3$.

Em um espaço de escuta tridimensional, a fonte nem sempre se encontrará no segmento de reta que liga duas HRTFs próximas. Assim, em um exemplo mais fiel ao que se dá realmente, considere agora que se quer simular uma fonte sonora deslocando-se no interior de um triângulo definido por três direções ($\boldsymbol{a}$, $\boldsymbol{b}$ e $\boldsymbol{c}$) linearmente independentes. Considere que haja uma HRTF associada a cada uma dessas direções (chamadas $H_a(n)$, $H_b(n)$ e $H_c(n)$) e que cada uma dessas HRTFs é, novamente,

um filtro formado por apenas um par de pólos complexos-conjugados. Sejam $p_a$, $p_b$ e $p_c$ esses pólos (deixando implícito o tratamento dos outros pólos dos pares complexos-conjugados), cuja representação polar é dada por

$$p_j = r_j e^{i\theta_j} \tag{3.6}$$

para $j \in \{a, b, c\}$. Assim, em uma extensão imediata da equação 3.4, o pólo da HRTF interpolada é calculado da seguinte maneira:

$$\hat{p} = (g_a r_a + g_b r_b + g_c r_c) e^{i(g_a\theta_a + g_b\theta_b + g_c\theta_c)}. \tag{3.7}$$

Os pesos $g_i$ utilizados aqui (e, de resto, em toda implementação da interpolação espectral) são os mesmos utilizados na interpolação triangular, obtidos por meio da equação 2.4 do capítulo 2, segundo a técnica de VBAP.

A utilização de filtros com alguns pólos e alguns zeros impõe um grau extra de complexidade. A interpolação espectral é feita, como se pôde observar acima, tomando trios de pólos e de zeros das HRTFs (cada HRTF contribui com 1 elemento do trio). Tendo alguns pólos e alguns zeros, diversas são as possibilidades de se estabelecerem esses trios, sendo que pode ocorrer de nem todas serem adequadas. Uma discussão detalhada sobre essa questão é feita na seção 3.2. Por enquanto, vamos admitir que já temos uma associação adequada e que os pólos e zeros das HRTFs já estejam ordenados segundo essa associação, ou seja, que o pólo de índice $i$ de $H_a(n)$ esteja associado com os pólos de índice $i$ de $H_b(n)$ e de $H_c(n)$.

Seja $p_k^j$ o pólo de índice $k$ da HRTF $j$ (sendo $j \in \{a,b,c\}$) e $q_k^j$ seu zero de índice $k$. Considere que sua representação polar é

$$\begin{aligned} p_k^j &= \alpha_k^j e^{i\theta_k^j} \\ q_k^j &= \beta_k^j e^{i\phi_k^j}. \end{aligned} \tag{3.8}$$

A variante polar da interpolação espectral define o pólo e o zero de índice $k$ do filtro resultante como:

$$\hat{p}_k = (g_a\alpha_k^a + g_b\alpha_k^b + g_c\alpha_k^c) e^{i(g_a\theta_k^a + g_b\theta_k^b + g_c\theta_k^c)} \tag{3.9}$$

$$\hat{q}_k = (g_a\beta_k^a + g_b\beta_k^b + g_c\beta_k^c) e^{i(g_a\phi_k^a + g_b\phi_k^b + g_c\phi_k^c)}. \tag{3.10}$$

Cada filtro IIR contém também um coeficiente de ganho $b_0$. Na interpolação espectral, o coeficiente

de ganho do filtro interpolado é definido como

$$\hat{b}_0 = g_a b_0^a + g_b b_0^b + g_c b_0^c, \tag{3.11}$$

de modo análogo ao que ocorre com os pólos e zeros.

Esse conjunto de pólos e zeros define uma função de transferência, que se obtém a partir da seguinte equação:

$$\hat{H}(z) = \hat{b}_0 \frac{\prod_{k=1}^{Q}(1 - \hat{q}_k z^{-1})}{\prod_{k=1}^{P}(1 - \hat{p}_k z^{-1})}, \tag{3.12}$$

que pode ser reescrita como:

$$\hat{H}(z) = \frac{\sum_{k=0}^{Q} \hat{b}_k z^{-k}}{1 + \sum_{k=1}^{P} \hat{a}_k z^{-k}}, \tag{3.13}$$

a partir do que se extraem os coeficientes para a equação do filtro:

$$\hat{y}(n) = \sum_{k=0}^{Q} \hat{b}_k x(n-k) - \sum_{k=1}^{P} \hat{a}_k \hat{y}(n-k). \tag{3.14}$$

Por sua vez, a variante cartesiana interpola os zeros e pólos dos filtros da seguinte maneira: considere novamente que se quer posicionar a fonte sonora virtual no interior de um triângulo definido por três direções ($\boldsymbol{a}$, $\boldsymbol{b}$ e $\boldsymbol{c}$) distintas e não-colineares, e que haja uma HRTF associada a cada uma dessas direções ($H_a(n)$, $H_b(n)$ e $H_c(n)$). Seja, mais uma vez, $p_k^j$ o pólo de índice $k$ da HRTF $j$ (com $j \in \{a,b,c\}$) e $q_k^j$ seu zero de índice $k$, e admita que eles já estão ordenados segundo uma associação adequada (o pólo de índice $i$ de $H_a(n)$ faz trio com os pólos de índice $i$ de $H_b(n)$ e de $H_c(n)$, e o mesmo valendo para os zeros). Isso posto, na variante cartesiana da interpolação espectral o pólo e o zero de índice $k$ do filtro resultante são definidos através das seguintes equações:

$$\hat{p}_k = g_a p_k^a + g_b p_k^b + g_c p_k^c \tag{3.15}$$

$$\hat{q}_k = g_a q_k^a + g_b q_k^b + g_c q_k^c, \tag{3.16}$$

que possuem paralelo com as equações 3.9 e 3.10 da variante polar. Aqui também, como no caso anterior, os coeficientes $g_i$ são obtidos pela técnica de VBAP através da equação 2.4.

## 3.2 Associação de pólos e de zeros

A interpolação espectral, assim como a interpolação triangular, produz um filtro interpolado através do emprego de três HRTFs concomitantemente. Também como a interpolação triangular, utiliza os pesos calculados pela técnica de VBAP, apresentada na seção 2.2. Entretanto, diferentemente da interpolação triangular, na interpolação espectral se interpolam os zeros e pólos das três HRTFs. Como cada HRTF pode ter diversos pólos e zeros, e a interpolação espectral toma um pólo ou zero de cada uma das três HRTFs para gerar um novo pólo ou zero interpolado, é preciso escolher um pólo ou um zero de cada HRTF para compor cada um dos trios, agrupando esses elementos três a três. A escolha arbitrária dos elementos que comporão cada trio pode gerar artefatos sonoros indesejáveis, pois, escolhendo, por exemplo, zeros muito distantes entre si para formarem um trio, sua interpolação poderia gerar variações muito bruscas no espectro, mesmo com deslocamentos pequenos da fonte sonora. Não é óbvia, contudo, qual é a escolha ótima a ser feita em geral.

Embora outros critérios de associação possam ser explorados futuramente, este trabalho utilizou, como primeira alternativa de solução, um critério de mínima energia, o que significa dizer que, entre todas as associações possíveis, foi escolhida aquela que minimizava a distância total percorrida pelos elementos dos trios entre uma configuração e outra.

Formalizando esse critério de associação, sejam $H_a(n)$, $H_b(n)$ e $H_c(n)$ as três HRTFs usadas pela interpolação. Vai-se aqui estabelecer a associação para os pólos apenas, uma vez que o procedimento é o mesmo para os zeros. Seja $P$ o número de pólos da HRTF $j$ (para $j \in \{a,b,c\}$) e seja $\{p_k^j\}$, com $1 \leq k \leq P$, o conjunto ordenado desses pólos. Considere que $\sigma_a(k)$ e $\sigma_b(k)$ sejam permutações dos índices $k$. A associação de pólos escolhida serão as permutações $\sigma_a(k)$ e $\sigma_b(k)$ que minimizarem

$$\sum_{k=1}^{P}[d(p^a_{\sigma_a(k)},p^b_{\sigma_b(k)}) + d(p^a_{\sigma_a(k)},p^c_k) + d(p^b_{\sigma_b(k)},p^c_k)], \tag{3.17}$$

sendo $d(\boldsymbol{x},\boldsymbol{y})$ a distância euclidiana entre $\boldsymbol{x}$ e $\boldsymbol{y}$. Tal procedimento é executado tanto para os pólos como para os zeros das HRTFs. Logo, teremos dois conjuntos de permutações, $\sigma_a^p(k)$ e $\sigma_b^p(k)$ para os pólos, e $\sigma_a^z(k)$ e $\sigma_b^z(k)$ para os zeros.

A associação pode ser feita em uma etapa de pré-processamento, na qual são armazenadas as permutações para cada trio. Portanto, esse procedimento não interfere nos custos computacionais da interpolação espectral.

## 3.3 Construindo um banco de HRTFs IIR

A interpolação espectral possui ainda alguns problemas de ordem prática a serem resolvidos. Essa técnica pressupõe que as HRTFs utilizadas são filtros IIR. No entanto, não estão disponíveis bancos de dados de HRTFs IIR que possam ser usados na interpolação. Isso leva à necessidade de obter filtros IIR a partir das HRTFs do CIPIC, que são filtros FIR. São diversos os métodos que possibilitam a aproximação de filtros FIR por filtros IIR [BKC92, SA92, LXYS99], alguns dos quais pensados especialmente para aplicações relacionadas a HRTFs [MHVK97, KC04]. A interpolação espectral, como desenvolvida nas seções acima, é independente do método escolhido de geração de filtros IIR. Pode-se aplicar a interpolação usando um banco de HRTFs IIR gerado por qualquer método. É preciso ter em mente, contudo, que qualquer aproximação de um filtro FIR por um filtro IIR (comumente de ordem reduzida) envolve um erro que deve ser levado em conta. Para os testes desenvolvidos neste trabalho, foi utilizado um método proposto originalmente por R. E. Kalman [Kal58, KC04], que será apresentado a seguir.

O problema de aproximar um filtro FIR por um filtro IIR pode ser formulado nos seguintes termos: seja $H(z)$, para $z \in \mathbb{C}$, a representação espectral do filtro FIR que se deseja modelar. Sendo $q$ o número de zeros e $p$ o número de pólos desejados para o filtro IIR, o objetivo é obter uma nova função de transferência $\hat{H}(z)$

$$\hat{H}(z) = \frac{B(z)}{A(z)}, \tag{3.18}$$

na qual

$$B(z) = \sum_{k=0}^{q} b_k z^{-k} \tag{3.19}$$

e

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}, \tag{3.20}$$

que aproxime, sob alguma métrica dada, $H(z)$. Em outras palavras, quer-se encontrar um $\hat{H}(z)$ que possa ser escrito como na equação 3.18 e que melhor aproxime a relação

$$\hat{H}(z) \approx H(z). \tag{3.21}$$

Kalman [Kal58, KC04] propôs uma solução para o sistema representado pelo diagrama da figura 3.4.

Minimizar o erro desse sistema é encontrar um $B(z)$ e um $A(z)$ que melhor aproximem a relação

$$B(z) \approx H(z)A(z). \tag{3.22}$$

Caso se encontre um $B(z)$ e um $A(z)$ tais que $\frac{B(z)}{A(z)}$ seja exatamente igual a $H(z)$, o quociente $\frac{B(z)}{A(z)}$ será também solução exata de $\hat{H}(z) = H(z)$ (cf. equação 3.21), mas esse geralmente não é o caso. Ainda assim, uma solução aproximada para a relação 3.21 também será, com menor ou maior exatidão, uma solução aproximada para a relação 3.22.

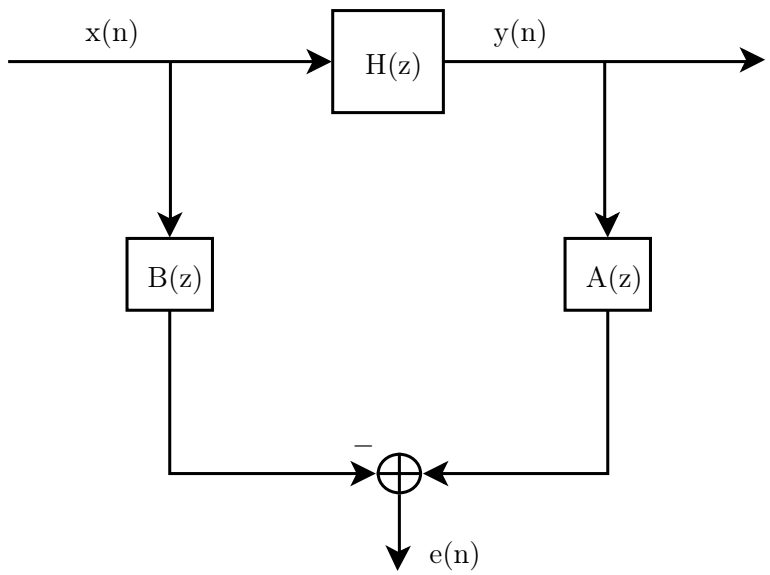



Figura 3.4: O sistema tratado por Kalman.

A solução de Kalman consiste em minimizar o erro quadrático

$$\sum_{k=0}^{N} |e(k)|^2, \tag{3.23}$$

sendo

$$e(k) = \sum_{l=0}^{q} b_l x(k-l) - \sum_{l=1}^{p} a_l y(k-l) - y(k).$$

Aqui, $y(\cdot)$ é a HRIR (conhecida) do filtro FIR que se deseja aproximar. Observe que $e(k) = \hat{y}(k) - y(k)$, sendo que

$$\hat{y}(k) = \sum_{l=0}^{q} b_l x(k-l) - \sum_{l=1}^{p} a_l y(k-l) \tag{3.24}$$

é a saída calculada pelo filtro IIR $\hat{H}$, dada a entrada impulsiva $x(\cdot)$ e a saída $y(\cdot)$ do filtro original.

A solução é como segue: seja

$$\boldsymbol{v}(k) = \begin{pmatrix} x(k) \\ \vdots \\ x(k-q) \\ y(k-1) \\ \vdots \\ y(k-p) \end{pmatrix}. \tag{3.25}$$

Considere a matriz V e o vetor de correlação $\mathbf{c}$ definidos como

$$V = \sum_{k=0}^{n} \boldsymbol{v}(k)\boldsymbol{v}(k)^t \tag{3.26}$$

e

$$\boldsymbol{c} = \sum_{k=0}^{n} y(k)\boldsymbol{v}(k). \tag{3.27}$$

Assim sendo, a solução $\boldsymbol{\delta}$ é obtida por

$$\boldsymbol{\delta} = V^{-1}\boldsymbol{c}, \tag{3.28}$$

sendo

$$\boldsymbol{\delta} = \begin{pmatrix} b_0 \\ \vdots \\ b_q \\ -a_1 \\ \vdots \\ -a_p \end{pmatrix}. \tag{3.29}$$

A partir do vetor $\boldsymbol{\delta}$ chega-se a $B(z)$ e $A(z)$:

$$B(z) = \sum_{k=0}^{q} b_k z^{-k} \tag{3.30}$$

e

$$A(z) = \sum_{k=1}^{p} a_k z^{-k}. \tag{3.31}$$

De posse desses coeficientes, podemos obter as posições dos pólos e zeros, ao reescrever $B(z)$ e $A(z)$ como

$$B(z) = b_0 \prod_{k=1}^{q} (1 - q_k z^{-1}) \tag{3.32}$$

e

$$A(z) = \prod_{k=1}^{p} (1 - p_k z^{-1}). \tag{3.33}$$

Aplicando esta técnica a todas as HRTFs de um banco de dados, como o CIPIC, por exemplo, podemos construir um banco de dados de HRTFs IIR para o qual as técnicas de associação de pólos e zeros da seção 3.2 e de interpolação espectral da seção 3.1 podem ser aplicadas.

## 3.4 Quanto à configuração dos filtros

O método de Kalman não impõe qualquer restrição à distribuição no plano complexo dos pólos e zeros gerados por ele. Ao aplicar o método ao banco de dados CIPIC, observou-se que ele produzia filtros com número variado de pólos e zeros reais (ver tabela 3.1). Ocorre que a interpolação espectral não permite que sejam usados na interpolação filtros com número diferente de pólos ou zeros reais. O seguinte exemplo ilustra uma tal situação: considere que se deseja interpolar duas HRTFs, uma delas com um par de pólos complexos-conjugados e um par de zeros complexos-conjugados, e a outra com um par de pólos complexos-conjugados e um par de zeros reais (conforme figuras 3.5a e 3.5d). Os pólos podem ser tranqüilamente interpolados segundo os critérios da interpolação espectral vistos neste capítulo. Contudo, a interpolação dos zeros, tal como estão dispostos, introduz uma problemática nova: se as disposições dos zeros na primeira e na última HRTFs da figura 3.5 são disposições perfeitamente válidas, o mesmo já não se pode dizer das disposições geradas nos estágios intermediários da interpolação (figuras 3.5b e 3.5c). Isto porque as HRTFs intermediárias produzidas pela interpolação espectral possuem zeros que não formam um par complexo-conjugado (como deixam

evidentes os casos (b) e (c) da figura 3.5). Essa configuração de zeros produz filtros com coeficientes complexos, o que de maneira nenhuma é admissível para as aplicações desejadas aqui.

| Filtros com 6 pólos e 6 zeros | | |
|---|---|---|
| **pólos reais** | **zeros reais** | **total de filtros** |
| 0 | 0 | 106 |
| 0 | 2 | 973 |
| 0 | 4 | 28 |
| 2 | 0 | 2 |
| 2 | 2 | 140 |
| 2 | 4 | 1 |

| Filtros com 8 pólos e 8 zeros | | |
|---|---|---|
| **pólos reais** | **zeros reais** | **total de filtros** |
| 0 | 0 | 154 |
| 0 | 2 | 451 |
| 0 | 4 | 70 |
| 2 | 0 | 63 |
| 2 | 2 | 415 |
| 2 | 4 | 97 |

| Filtros com 10 pólos e 10 zeros | | |
|---|---|---|
| **pólos reais** | **zeros reais** | **total de filtros** |
| 0 | 0 | 142 |
| 0 | 2 | 539 |
| 0 | 4 | 48 |
| 2 | 0 | 98 |
| 2 | 2 | 372 |
| 2 | 4 | 50 |
| 4 | 2 | 1 |

| Filtros com 12 pólos e 12 zeros | | |
|---|---|---|
| **pólos reais** | **zeros reais** | **total de filtros** |
| 0 | 0 | 77 |
| 0 | 2 | 417 |
| 0 | 4 | 41 |
| 2 | 0 | 38 |
| 2 | 2 | 533 |
| 2 | 4 | 138 |
| 4 | 2 | 5 |
| 4 | 4 | 1 |

| Filtros com 14 pólos e 14 zeros | | |
|---|---|---|
| **pólos reais** | **zeros reais** | **total de filtros** |
| 0 | 0 | 17 |
| 0 | 2 | 38 |
| 0 | 4 | 5 |
| 2 | 0 | 29 |
| 2 | 2 | 778 |
| 2 | 4 | 366 |
| 2 | 6 | 1 |
| 4 | 0 | 1 |
| 4 | 2 | 11 |
| 4 | 4 | 4 |

Tabela 3.1: Distribuição dos filtros gerados pelo método de Kalman segundo número de pólos e de zeros reais, para filtros de algumas ordens diferentes.

A inadequação da disposição dos pólos (ou zeros) em filtros interpolados surge sempre que o

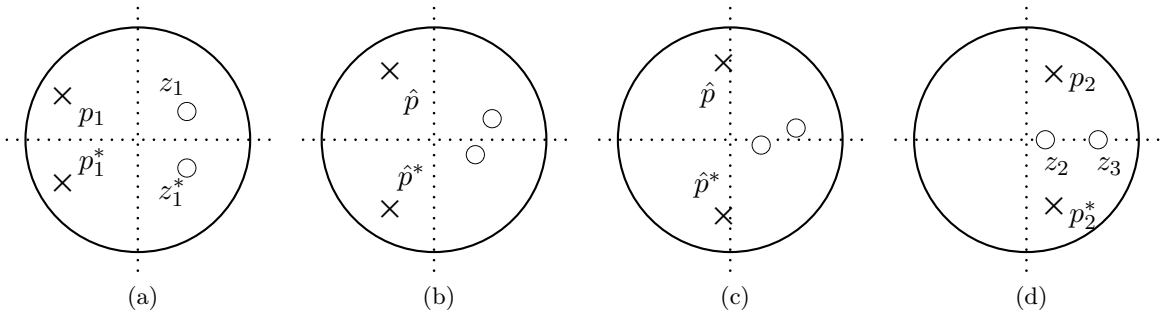



Figura 3.5: Exemplo de interpolação de dois filtros com configurações diferentes. Em (a) há dois zeros complexos-conjugados. Em (d), os dois zeros são reais. A interpolação espectral empregada usando esses dois filtros gera filtros intermediários — figuras (b) e (c) — que possuem zeros que não são reais nem tampouco pares complexos-conjugados.

número de pólos (ou de zeros) reais não for o mesmo nos filtros usados pela interpolação, e é inerente à técnica mesma e a sua proposta. Esse tipo de problema só pode ser evitado impondo-se restrições à disposição dos pólos e zeros, como, por exemplo, elegendo uma configuração padrão e forçando os filtros a se conformarem a ela, ou seja, estabelecer o número de zeros e pólos reais que um filtro deverá ter, se é que os deverá ter. Ocorre que nenhum dos métodos encontrados na literatura para se obterem filtros IIR a partir de filtros FIR impõe restrições ao número de pólos e zeros reais que produz.

A única maneira encontrada até o momento para forçar uma tal homogeneidade de configurações é através da substituição de pares de pólos (ou zeros) reais por pares complexos-conjugados que deformem o mínimo possível os filtros IIR obtidos pelo método de Kalman. Tal condição pode ser alcançada fazendo, para cada par ($p_A$, $p_B$) de pólos (ou zeros) reais que se deseja substituir por um par complexo-conjugado, uma busca exaustiva sobre os valores $(Ae^{\varphi}, Ae^{-\varphi})$, calculando a função de transferência correspondente à substituição e seu erro em relação à função de transferência original, e escolhendo os melhores valores de $A$ e $\varphi$. Este trabalho, embora computacionalmente muito custoso, só precisa ser feito ao se montar o banco de dados de IIRs que serão posteriormente utilizadas na interpolação, e, portanto, este custo não interfere no custo da interpolação propriamente dita.

Nos experimentos computacionais preliminares não foram obtidas boas substituições de pólos e zeros usando esta técnica, sendo que os erros ficavam freqüentemente acima de 30% do valor da

norma das HRTFs. Por esta razão, os resultados numéricos apresentados na próxima seção utilizam apenas o conjunto de filtros IIR para os quais o método original de Kalman produziu configurações adequadas. O problema de produzir filtros IIR a partir de filtros FIR com uma estrutura de pólos e zeros pré-definida será alvo de investigações futuras.

## Capítulo 4

# Implementação e testes

A qualidade das técnicas apresentadas nos dois capítulos anteriores foi avaliada por meio de testes numéricos. Esses testes utilizaram uma metodologia comum à pesquisa em interpolação de HRTFs: gerar HRTFs interpoladas que possam ser comparadas com as HRTFs correspondentes no banco de dados original. Para poder avaliar objetivamente a qualidade de uma HRTF interpolada para uma certa direção, é bastante útil poder compará-la com a HRTF original dessa mesma direção; por isso, os testes executados aqui geraram HRTFs interpoladas para direções para as quais já existiam HRTFs originais no banco de dados, utilizando no processo de interpolação HRTFs vizinhas também presentes no banco. Há variações nas aplicações dos testes para cada uma das técnicas de interpolação. Entretanto, há o seguinte procedimento básico que todas elas seguem:

- as direções usadas para posicionar a fonte virtual foram escolhidas dentre aquelas para as quais o banco de dados possui medições; em outras palavras, para cada direção escolhida para posicionar a fonte virtual, havia já uma HRTF original no banco de dados;
- para poder gerar HRTFs interpoladas para direções já contempladas no banco de dados, a densidade da malha de HRTFs foi reduzida à metade, o que significa que, se a malha do banco de dados possui medições na seqüência de azimutes $\theta_n$, $\theta_{n+1}$, $\theta_{n+2}$, $\theta_{n+3}$ e $\theta_{n+4}$, nos testes foram usados, por exemplo, apenas os azimutes $\theta_n$, $\theta_{n+2}$ e $\theta_{n+4}$; da mesma forma, se o banco de dados possui medições na seqüência de elevações $\phi_n$, $\phi_{n+1}$, $\phi_{n+2}$, $\phi_{n+3}$ e $\phi_{n+4}$, nos testes foram usados, por exemplo, apenas as elevações $\phi_n$, $\phi_{n+2}$ e $\phi_{n+4}$;
- usando essa nova malha, mais esparsa, novas HRTFs foram geradas para as direções escolhidas a partir de HRTFs adjacentes;

- a HRTF interpolada foi, então, comparada com a HRTF original associada à direção escolhida.

Isso dá uma medida objetiva da qualidade numérica da interpolação executada por cada técnica. Se a HRTF original é aquela medida experimentalmente, então é objetivo da interpolação estar o mais próximo possível dela.

Todos os testes foram implementados usando o programa Octave[1].

## 4.1 Implementação e testes da interpolação triangular

Baseado no procedimento descrito anteriormente, os testes para a interpolação triangular consistiram em gerar HRTFs interpoladas e então compará-las com as HRTFs originais correspondentes. Isso foi feito da seguinte forma: dada uma direção $D$ do banco de dados, foram selecionados triângulos $ABC$ que contivessem $D$ e cujos vértices estivessem associados a HRTFs ($H_A(n)$, $H_B(n)$ e $H_C(n)$) adjacentes a $H_D(n)$. Em seguida, $H_A(n)$, $H_B(n)$ e $H_C(n)$ foram usadas para gerar uma HRTF interpolada $\hat{H}_D(n)$ para a mesma direção $D$. Note que as HRTFs $H_A(n)$, $H_B(n)$, $H_C(n)$ e $H_D(n)$ são todas previamente conhecidas. Deste modo, pôde-se comparar a HRTF $\hat{H}_D(n)$, que é resultado da interpolação para a direção $D$, com a HRTF $H_D(n)$ de fato medida para essa mesma direção.

Para a interpolação triangular, são usadas 3 HRTFs concomitantemente. Entretanto, são 8 os seus vizinhos mais próximos (ver figura 4.1). Isso faz que sejam possíveis diversas escolhas para as 3 HRTFs. Como não se pode dizer, *a priori*, qual escolha trará os melhores resultados, foram estabelecidas 8 configurações distintas de triângulos que contivessem a direção $D$, que foram usados na técnica de interpolação triangular para gerar 8 HRTFs $\hat{H}_D(n)$ distintas. Gerou-se, também, uma HRTF média a partir dessas 8 HRTFs. As HRTFs interpoladas foram, então, comparadas com a $H_D(n)$ original. Os triângulos utilizados são exibidos na figura 4.1.

As figuras de número 4.3 a 4.5 exibem os gráficos das superfícies geradas a partir das HRTFs originais do banco de dados e das médias de cada conjunto de 8 HRTFs interpoladas. Elas contém as superfícies espaciais de resposta em freqüência para esse conjunto de HRTFs (ver seção 1.3 para mais detalhes acerca dessas superfícies). Os valores de freqüência utilizados para a geração das superfícies foram extraídos do artigo "Spatial Frequency Response Surfaces (SFRS's): An Alternative Visualization and Interpolation Technique for Head-Related Transfer Functions (HRTF's)" [CW99]. A coluna da esquerda exibe as HRTFs originais, enquanto que a coluna do meio exibe a média das

[1] Programa de código aberto disponível em http://www.gnu.org/software/octave

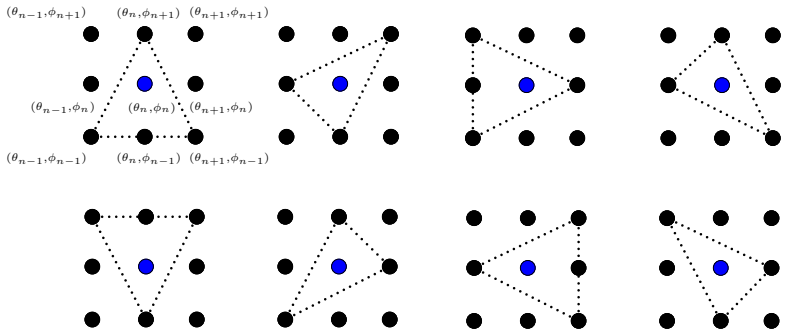



Figura 4.1: Configurações dos 8 triângulos que foram utilizados para os testes da interpolação triangular. Os pontos em azul são aqueles para os quais a técnica gera uma HRTF interpolada — conforme o texto, os pontos em azul são o ponto ($\theta$,$\phi$). Os 8 pontos ao redor do ponto azul são seus vizinhos imediatos no banco de dados.

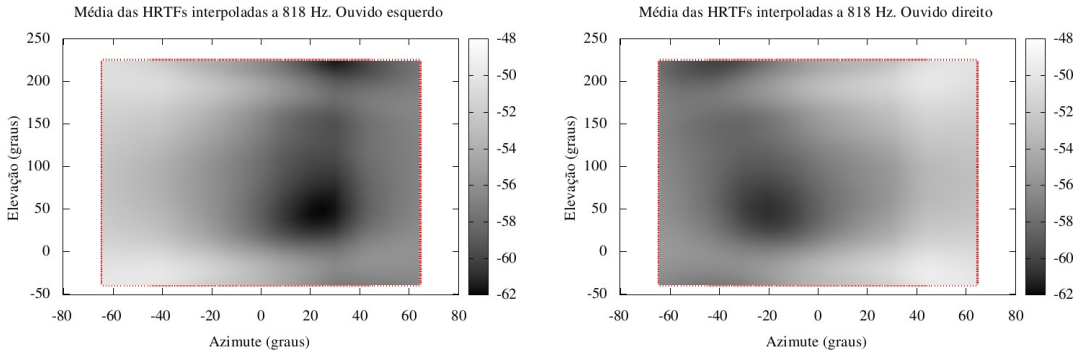
HRTFs interpoladas. A coluna da direita, por sua vez, exibe o valor absoluto da diferença das magnitudes das HRTFs originais e das médias das HRTFs interpoladas, para facilitar a identificação dos erros cometidos pela interpolação. Como, no geral, os gráficos referentes ao ouvido direito são praticamente imagens espelhadas dos gráficos referentes ao ouvido esquerdo (cf. figura 4.2), apenas os gráficos das diferenças são exibidos para o ouvido direito (figura 4.5).

Os gráficos das HRTFs interpoladas para o ouvido direito, bem como a relação completa dos gráficos para diversas freqüências e para os ouvidos esquerdo e direito encontra-se no apêndice A.

Pode-se verificar que, no geral, as médias das HRTFs interpoladas aproximam bem o comportamento dos picos e vales das HRTFs originais. Algumas observações mais pontuais, contudo, podem ser extraídas:

- em geral, os gráficos das médias das HRTFs interpoladas são mais suaves que os gráficos das HRTFs originais. Esse efeito de suavização é esperado, uma vez que cada HRTF interpolada é construída a partir das 8 HRTFs vizinhas a ela ao se usar as médias das interpolações;

- na maioria dos gráficos, os maiores valores dos erros encontram-se nas extremidades dos azimutes. Isso se deve ao fato de a malha do CIPIC não ser homogeneamente espaçada, mas, pelo contrário, ser menos densa justamente nas extremidades dos azimutes. O erro nas demais áreas da superfície (parte central dos gráficos) é, em geral, significativamente menor (ver, por




(a) Ouvido esquerdo (b) Ouvido direito

Figura 4.2: Resultado da interpolação para o ouvido esquerdo e direito. É possível notar a simetria entre os gráficos.

ex., figuras 4.3i, 4.3l, 4.4f e 4.4i). Há uma grande exceção, contudo: o gráfico 4.4r, no qual há uma faixa de erro elevado bem em sua região central; em tal região, há um vale profundo na superfície original que chega a -85 dB (figura 4.4p) que, apesar de ser capturado pela superfície das HRTFs interpoladas (figura 4.4q), não atinge nesta superfície a mesma intensidade da faixa na superfície original (não passa de -75 dB na interpolada) e é mais larga que a original; esse é também resultado do efeito de suavização mencionado no item anterior;

- o gráfico 4.3l exibe uma situação interessante: embora os valores dos erros sejam quase todos relativamente baixos, há uma pequena região com erro bastante elevado (chegando a até 12dB). Observando as superfícies das HRTFs originais (figura 4.3j) e interpoladas (figura 4.3k), vê-se que esse erro tem duas origens: uma delas é devida ao fato de que a área do vale profundo é menor na superfície original do que na superfície interpolada; a outra é a profundidade desse vale, que na superfície original chega a -80 dB, e na interpolada não passa de -66 dB. Ou seja, o erro aqui tem as mesmas características que o erro apontado no item anterior, e certamente se deve ao mesmo efeito.

As observações apontadas acima indicam que picos altos e vales profundos limitados a pequenas regiões das superfícies de resposta das HRTFs originais são mais difíceis de serem interpolados corretamente, dado o fato de que as superfícies interpoladas são construídas através da informação obtida

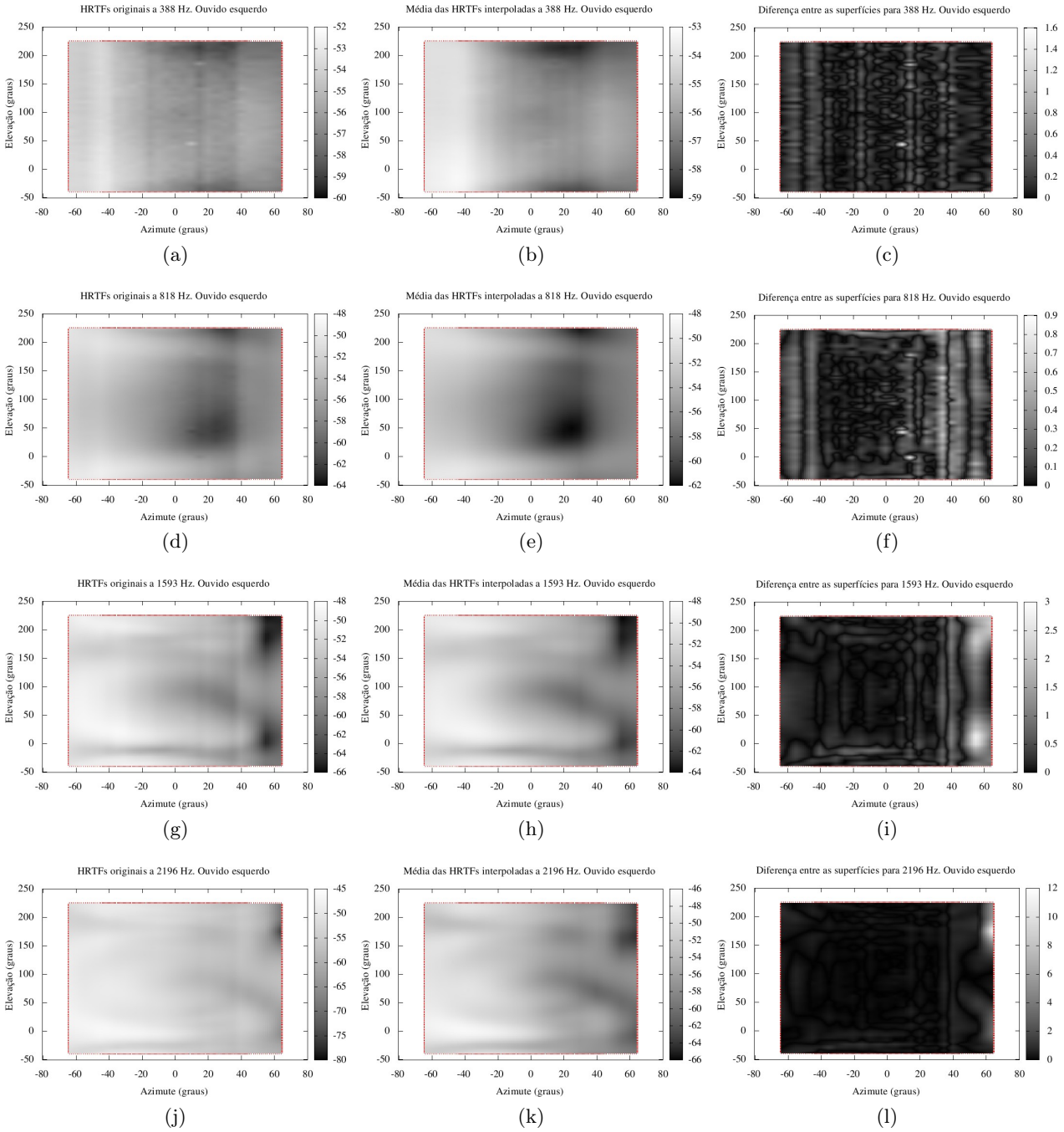



Figura 4.3: Resultado da interpolação triangular para o ouvido esquerdo. A primeira coluna contém as superfícies das HRTFs originais; a segunda coluna contém as superfícies das HRTFs interpoladas; e a terceira coluna contém os valores absolutos das diferenças entre as duas superfícies. Os valores são dados em decibéis.

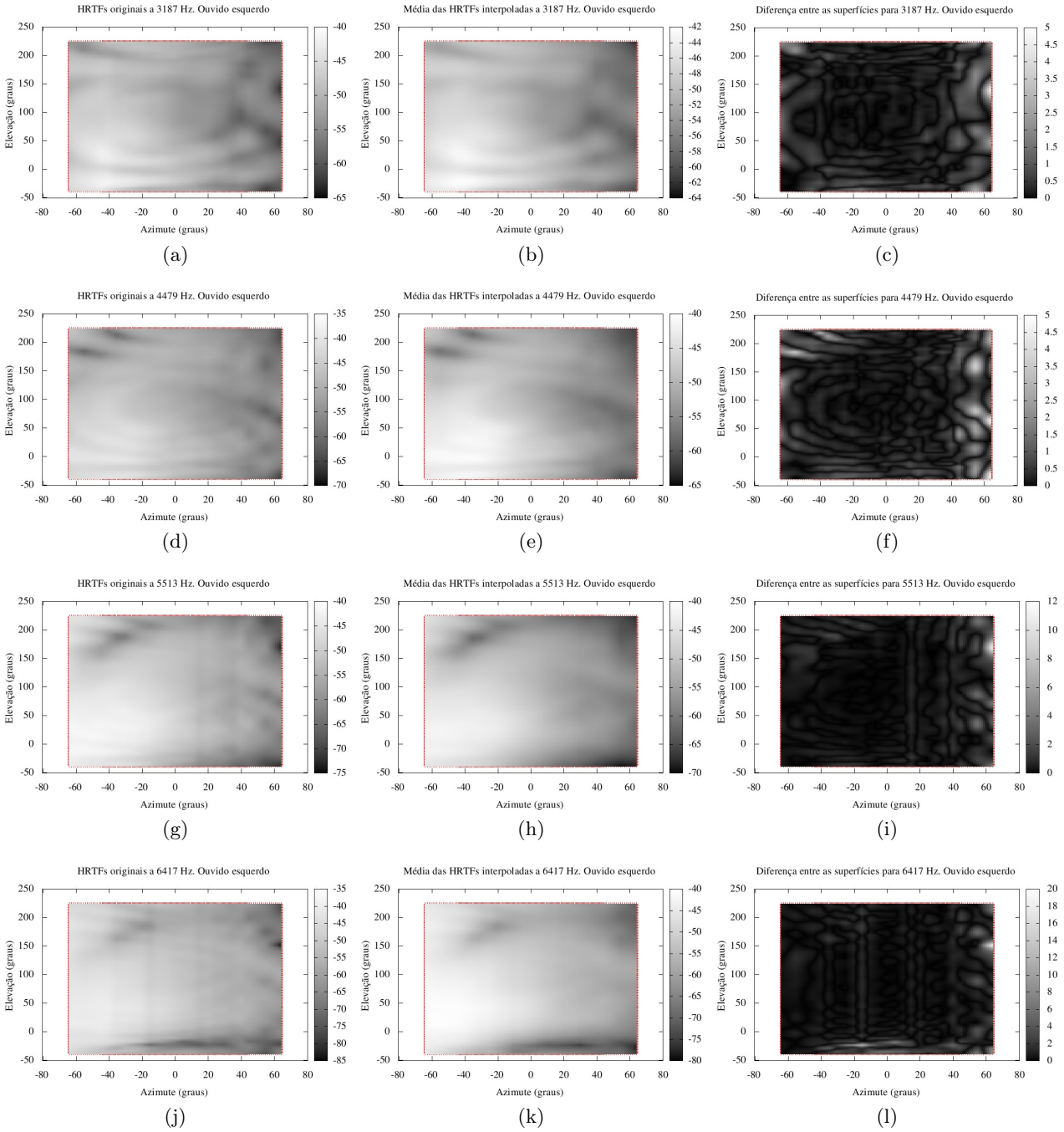



Figura 4.4: Continuação dos dados da figura 4.3.

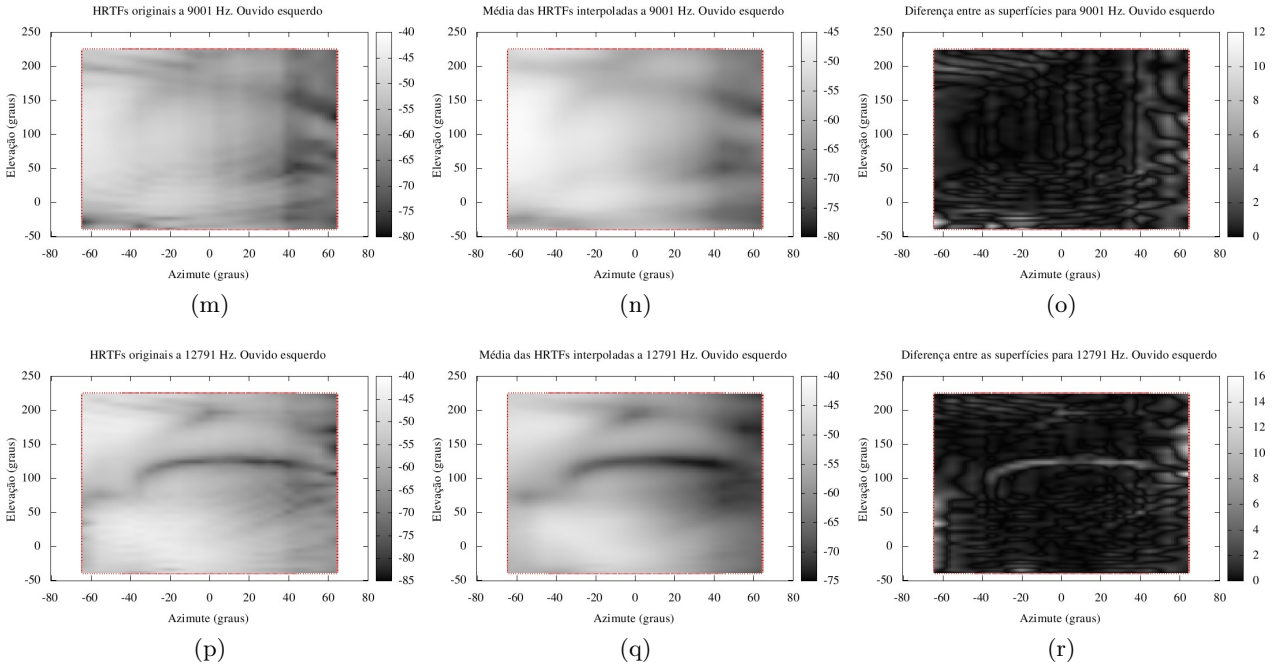



Figura 4.4: Continuação dos dados da figura 4.3.

de pontos vizinhos.

Outra característica interessante, mas que é menos perceptível através das superfícies espaciais de resposta em freqüência, é a relação entre o aumento do erro da interpolação e o aumento da freqüência. Para tornar essa característica mais nítida, a figura 4.6 exibe gráficos com os valores dos erros relativos das superfícies de diferença em função da freqüência. Assim é possível verificar que, de fato, o erro aumenta com a freqüência, ainda que não seja monotonicamente. Isso vale tanto para o ouvido esquerdo como para o direito.

Uma explicação para esse fato é que a distância entre os pontos medidos no CIPIC é normalmente menor que 10 cm. Portanto esse valor é da mesma ordem de grandeza que o comprimento de onda de sinais sonoros a freqüências mais elevadas[2]. Logo, HRTFs vizinhas tendem a ser menos semelhantes entre si nas freqüências mais altas do que nas mais baixas. Isso acaba refletindo no desempenho da

[2]A 6,4 kHz, por exemplo, o comprimento de onda de um sinal sonoro é de cerca de 5 cm.

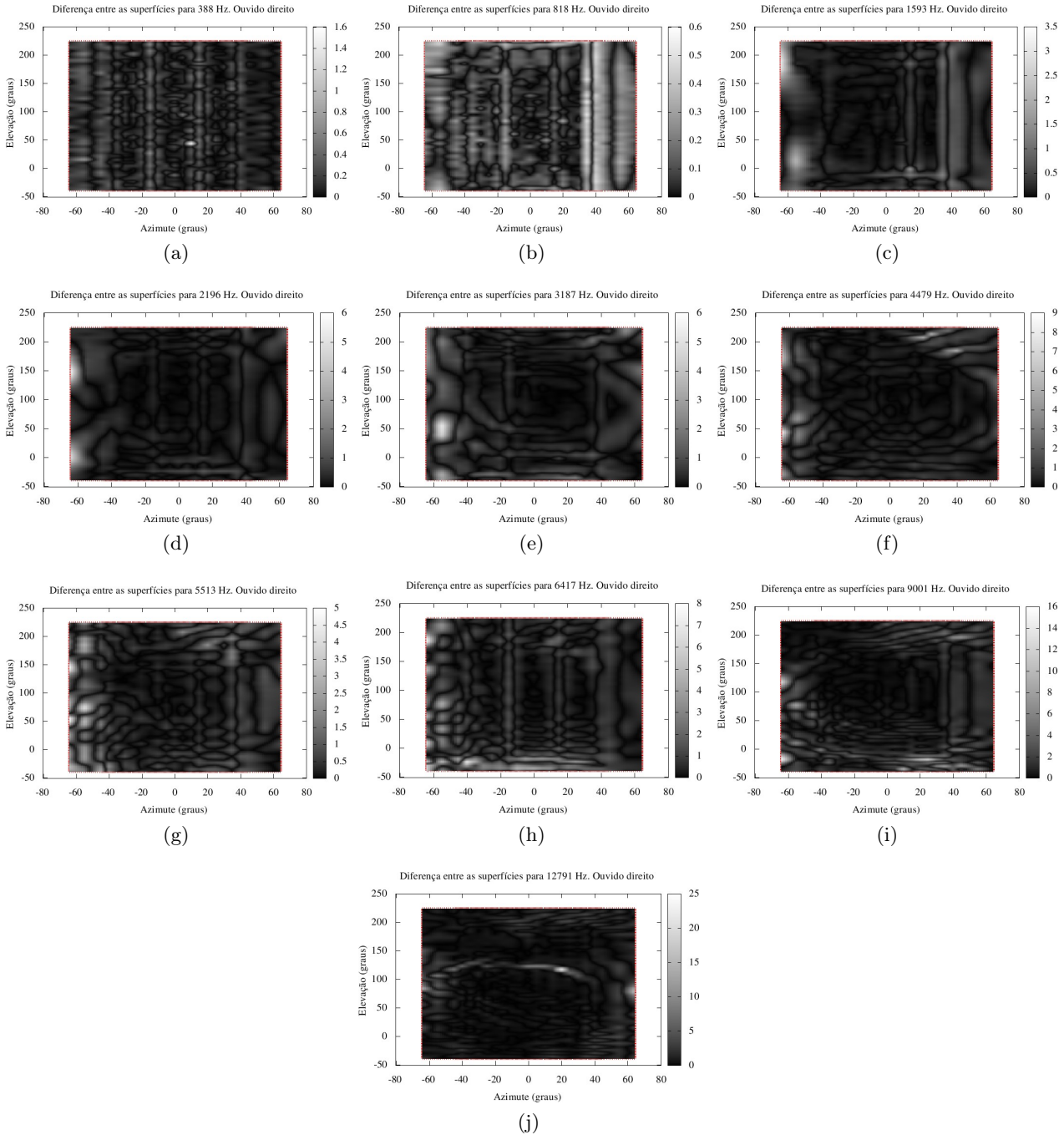
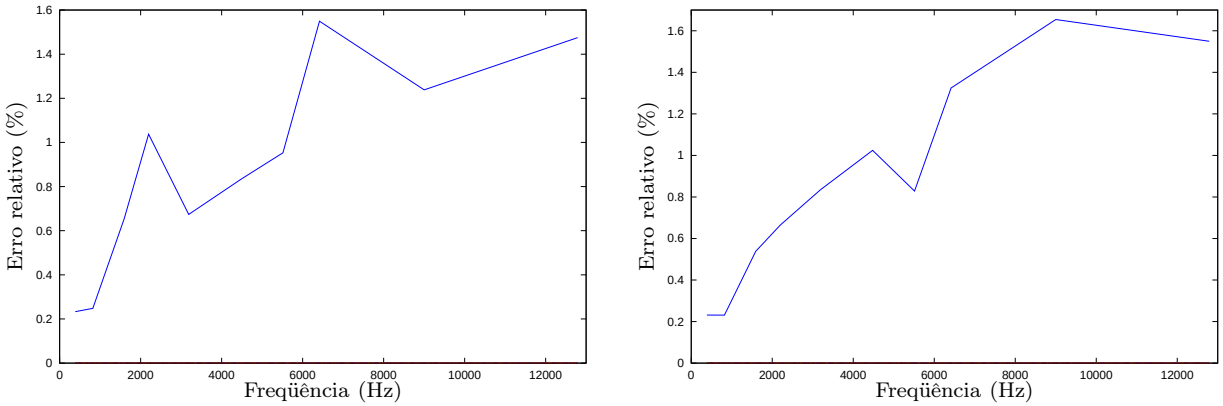



Figura 4.5: Diferenças das superfícies espaciais de resposta em freqüência para o ouvido direito. Valores em decibéis.




(a) Ouvido esquerdo (b) Ouvido direito

Figura 4.6: Evolução dos erros relativos das interpolações em função da freqüência.

técnica de interpolação nessa região do espectro.

Esse fenômeno já foi observado em um trabalho de Ajdler e colaboradores [AFSV05], que estudaram qual a menor distância angular entre duas medições de HRTFs necessária para que elas possam ser interpoladas adequadamente. O estudo conclui haver uma dependência inversa entre essa distância angular e a largura de banda dos sinais que se quer usar. Isso está relacionado justamente ao fato de os erros das interpolações aumentarem para freqüências mais elevadas.

## 4.2 Implementação e testes do método das movimentações discretas

O método das movimentações discretas, diferentemente da interpolação triangular, só necessita de uma HRTF em sua implementação. O procedimento dos testes é, deste modo, bastante simplificado, consistindo em selecionar a HRTF mais próxima dentre suas 4 vizinhas imediatas.

O intervalo de medições do CIPIC, quando restrito a azimutes de -45° a 45°, é regular, consistindo em passos de 5° no azimute e 5,625° na elevação. Sendo assim, o método das movimentações discretas, no caso específico destes testes e restrito ao intervalo azimutal mencionado acima, vai sempre selecionar um vizinho seu de azimute. Ou seja, se $(\theta_n,\phi_n)$ é a direção da fonte virtual, o método escolhe sempre ou a HRTF associada à direção $(\theta_{n-1},\phi)$ ou aquela associada à direção $(\theta_{n+1},\phi)$. Particularmente

na implementação usada nestes testes, o método vai sempre selecionar a HRTF da direção ($\theta_{n-1}$,$\phi$) quando estiver dentro da região de azimute entre -45° e 45°. Desta maneira, gerar gráficos das superfícies de resposta, como feito para a interpolação triangular, não acrescenta muita informação sobre a qualidade da técnica, uma vez que produziria apenas gráficos deslocados para a direita quando comparados com os originais. Por isso, a figura 4.7 apresenta somente os gráficos das diferenças das superfícies.

Uma vez que os resultados do método das movimentações discretas não são construídos a partir de uma média ponderada de HRTFs vizinhas, esta técnica não apresenta os mesmos problemas decorrentes da suavização das superfícies de resposta que foi observado nos resultados da interpolação triangular. Comparando, como exemplo, a superfície de diferenças da figura 4.7b das movimentações discretas com a superfície de diferenças da figura 4.3l da interpolação triangular, vê-se que aquela não possui o mesmo pico de erro de 12 dB que esta possui. Pelo contrário, seus erros nunca passam de 3 dB. Da mesma forma, os resultados do método das movimentações discretas para 3.187 Hz, 5.513 Hz, 6.417 Hz, 9.001 Hz e 12.791 Hz todos possuem valor máximo de erro inferior aos obtidos na interpolação triangular. Isso pode dar a impressão de que o método das movimentações discretas apresenta resultados melhores de interpolação que os da interpolação triangular, o que seria bastante estranho, já que as movimentações discretas não chegam a interpolar HRTFs. Na verdade, embora os valores máximos de erro na interpolação triangular sejam muitas vezes maiores, na média esses erros ficam aquém dos erros produzidos pelo método das movimentações discretas. Para uma melhor comparação, a figura 4.8 exibe a evolução das normas das diferenças das movimentações discretas lado a lado com os gráficos da interpolação triangular.

## 4.3 Implementação e testes da interpolação espectral

A interpolação espectral foi implementada de modo análogo ao que foi feito para a interpolação triangular. Os mesmos 8 triângulos foram utilizados, gerando 8 conjuntos de dados e mais um, correspondendo à média desses 8. Aqui, diferentemente do que ocorreu nas outras técnicas, os resultados da interpolação espectral não foram comparados diretamente com as HRTFs originais do CIPIC. Isso porque estariam sendo sobrepostos dois conjuntos de erros nessa comparação: um referente à conversão do CIPIC, que é um banco de HRTFs FIR, para um banco de HRTFs IIR; e outro que é o erro cometido pela interpolação propriamente. Por isso as HRTFs originais usadas para comparar o resultado da interpolação foram as HRTFs IIR geradas pelo método de Kalman e não as HRTFs FIR do CIPIC.

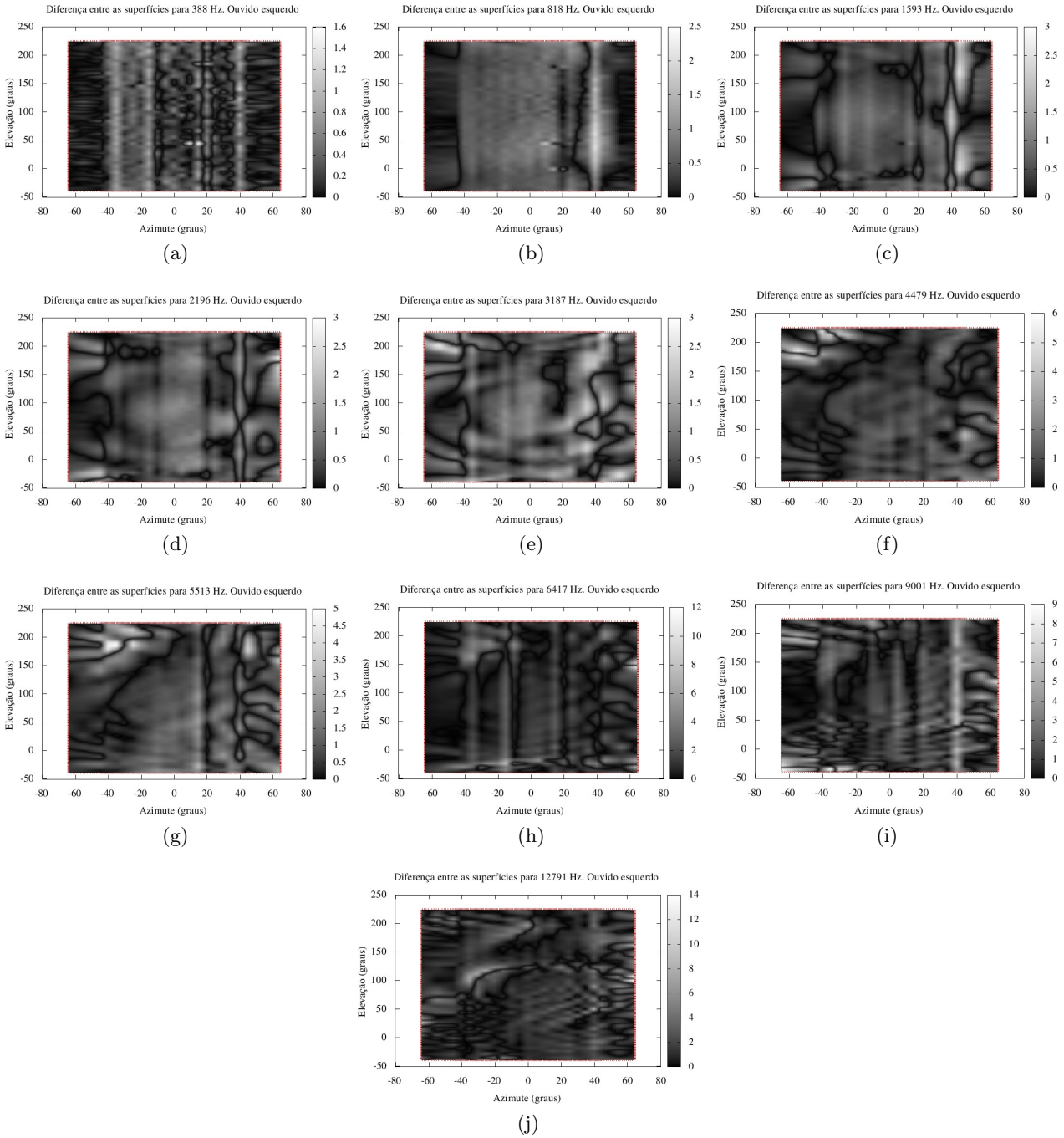
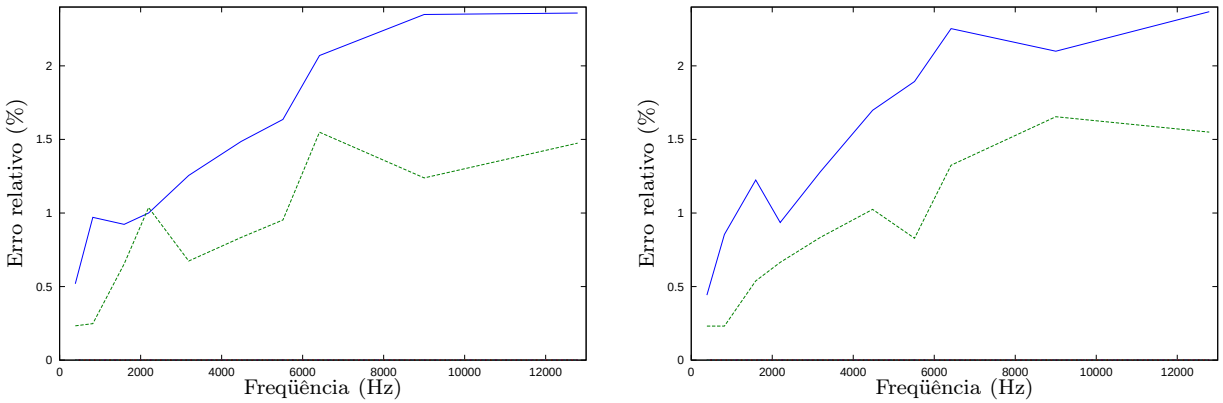



Figura 4.7: Movimentações discretas. Ouvido esquerdo. Valores em decibéis.




(a) Ouvido esquerdo. (b) Ouvido direito

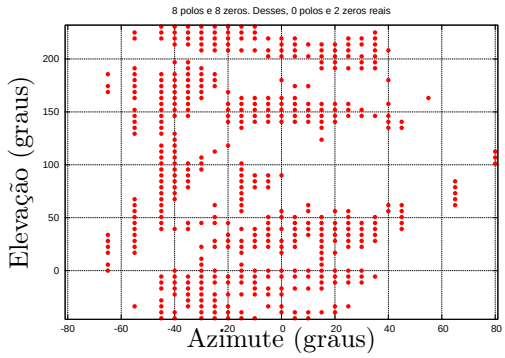
Figura 4.8: Comparação dos erros das técnicas de interpolação triangular e das movimentações discretas. Em azul, os erros do método das movimentações discretas; em verde, os erros da interpolação triangular.

Durante a implementação dessa técnica, verificou-se que o método de Kalman, usado para gerar as HRTFs IIR (cf. seção 3.3) não produz resultados com a homogeneidade necessária para permitir testes em todo o banco de dados (cf. seção 3.4), o que impediu que se gerassem superfícies de resposta nos moldes do que foi feito com a interpolação triangular.

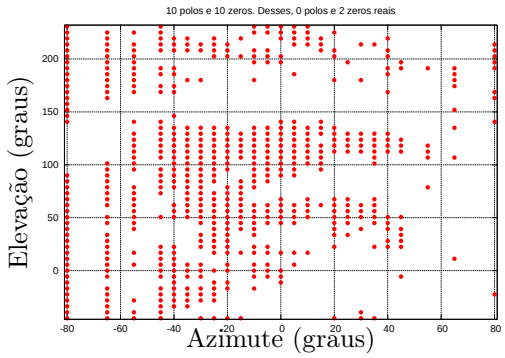
A solução adotada para estes testes foi aplicar a interpolação espectral usando apenas os filtros cuja configuração de pólos e zeros reais é a mais comum entre todos os filtros IIR gerados pelo método de Kalman. Isso restringiu a aplicação dos testes a apenas algumas regiões do banco de HRTFs. A figura 4.9 mostra quais são as regiões utilizadas em função da ordem do filtro.

Note que esses filtros geralmente não estão agrupados em uma única região do banco de HRTFs, mas estão espalhados por ele. Essa dispersão impossibilitou a geração de superfícies espaciais de resposta em freqüência, como foi feito para a interpolação triangular. Por isso, para a geração de gráficos para os testes da interpolação espectral, foram escolhidos pontos somente da região que tivesse o maior número de filtros (a região representada na figura 4.9), e esses pontos foram escolhidos de tal modo que eles cobrissem de maneira mais ou menos homogênea toda essa região.

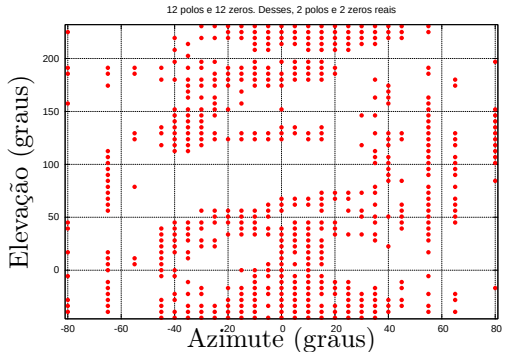
Os testes foram executados para filtros de até 14 pólos e 14 zeros. As tabelas 4.1, 4.2 e 4.3 exibem



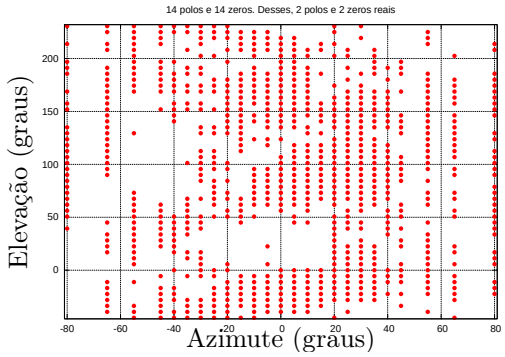

(a) Filtros com 8 pólos e 8 zeros, dos quais 2 zeros são reais.




(b) Filtros com 10 pólos e 10 zeros, dos quais 2 zeros são reais.




(c) Filtros com 12 pólos e 12 zeros, dos quais 2 pólos e 2 zeros são reais.




(d) Filtros com 14 pólos e 14 zeros, dos quais 2 pólos e 2 zeros são reais.

Figura 4.9: Distribuição espacial, no banco de dados, dos filtros que possuem a configuração mais comum de pólos e zeros (cf. tabela 3.1).

os valores dos erros da interpolação nestes testes, para filtros de algumas ordens diferentes.

As seguintes informações podem ser extraídas diretamente das tabelas 4.1, 4.2 e 4.3:

- no geral, os erros cometidos ficam abaixo dos 15%: das 158 medições de erro, apenas 6 ficaram acima desse valor, correspondendo às direções (30°, -22,5°) e (5°, 56,25°) para filtros de 10 pólos e 10 zeros, e (30°, -39,375°) para filtros de 12 pólos e 12 zeros; além disso, das 158 medições, apenas 11 apresentam erro acima de 12%;

| | **Erro relativo** | |
|---|---|---|
| **Direção** | **Variante cartesiana** | **Variante polar** |
| (-40°, -28,125°) | 6,35% | 6,47% |
| (-40°, 84,375°) | 4,85% | 5,22% |
| (-40°, 213,75°) | 1,73% | 0,46% |
| (-35°, 0°) | 11,53% | 10,36% |
| (-35°, 101,25°) | 6,98% | 7,64% |
| (-35°, 129,375°) | 11,22% | 11,18% |
| (-30°, -22,5°) | 23,19% | 24,34% |
| (-30°, 56,25°) | 6,92% | 7,31% |
| (-20°, 50,625°) | 6,71% | 6,89% |
| (-20°, 123,75°) | 13,13% | 13,70% |
| (-10°, 56,25°) | 4,66% | 5,37% |
| (-10°, 112,5°) | 3,89% | 8,38% |
| (-5°, 95,625°) | 9,61% | 12,64% |
| (5°, 56,25°) | 20,48% | 20,04% |
| (5°, 129,375°) | 0,52% | 0,10% |
| (5°, 225°) | 2,76% | 2,79% |
| (10°, 106,875°) | 0,07% | 0,81% |
| (25°, 56,25°) | 4,27% | 4,60% |
| (25°, 118,125°) | 6,83% | 6,61% |
| (35°, 123,75°) | 1,30% | 0,67% |
| (40°, 118,125°) | 4,72% | 3,67% |

Tabela 4.1: Erros da interpolação espectral. Filtros com 10 pólos e 10 zeros. Eles foram calculados a partir da média das interpolações para os 8 triângulos utilizados.

- nenhuma das variantes apresentou erros consistentemente menores que a outra; entretanto, a variante cartesiana foi a que mais vezes apresentou erros menores: das 47 direções analisadas, a variante cartesiana apresentou erros menores 26 vezes, enquanto que a variante polar apresentou resultados melhores em 20 vezes, tendo havido um valor de erro que foi igual nos dois casos.

Vê-se, portanto, que, tipicamente, os resultados dos testes da interpolação espectral reproduzem relativamente bem o comportamento dos filtros originais, com erros que raramente excedem 12%. Vê-se também que, pelos dados coletados, não se pode determinar com clareza que alguma das variantes da técnica de interpolação seja a melhor que a outra.

| | **Erro relativo** | |
|---|---|---|
| **Direção** | **Variante cartesiana** | **Variante polar** |
| (-40°, 33,75°) | 4,98% | 5,89% |
| (-35°, 140,625°) | 10,11% | 10,13% |
| (30°, -39,375°) | 30,35% | 37,89% |
| (-20°, 50,625°) | 7,75% | 7,21% |
| (-15°, -33,75°) | 10,69% | 9,90% |
| (-15°, 225°) | 11,10% | 10,45% |
| (-5°, 180°) | 10,32% | 9,41% |
| (0°, 219,375°) | 5,61% | 11,62% |
| (5°, -16,875°) | 2,29% | 2,21% |
| (10°, 185,625°) | 0,24% | 0,50% |
| (10°, 225°) | 1,01% | 5,70% |
| (15°, 213,75°) | 3,85% | 3,32% |
| (20°, -39,375°) | 10,96% | 10,67% |
| (35°, -33,75°) | 8,68% | 11,62% |
| (55°, 50,625°) | 0,72% | 1,10% |

Tabela 4.2: Erros da interpolação espectral. Filtros com 12 pólos e 12 zeros.

Além do cômputo dos erros, os testes geraram gráficos dos filtros interpolados. O apêndice B contém o conjunto completo dos gráficos dos testes da interpolação espectral para filtros de 14 pólos e 14 zeros. Lá é possível visualizar as respostas em freqüência dos filtros interpolados em comparação com os originais, para ambas as variantes da técnica. É possível verificar que, visualmente, as duas variantes produzem resultados praticamente idênticos, sendo, normalmente, muito difícil notar as diferenças entre elas. É possível ver também que os filtros interpolados, apesar de variações nos resultados, reproduzem relativamente bem os filtros originais, no que se refere às regiões de ressonância e antirressonância do espectro, à fase, e ao posicionamento dos pólos e zeros no plano complexo.

A seguir, uma breve amostra dos resultados recolhidos no apêndice B será apresentada, com o intuito de exibir situações representativas do teste todo para facilitar a visualização de casos típicos.

A primeira observação a se fazer é que, na maior parte das direções testadas, os resultados da interpolação espectral reproduzem relativamente bem o comportamento dos filtros originais. A figura 4.10 exibe gráficos da direção (-5°, 90°), para a qual os testes geraram resultados satisfatórios. São exibidos a magnitude e fase da resposta em freqüência e o diagrama de pólos e zeros do filtro interpolado com

| | **Erro relativo** | |
|---|---|---|
| **Direção** | **Variante cartesiana** | **Variante polar** |
| (-40°, 45°) | 8,62% | 7,31% |
| (-25°, 225°) | 6,58% | 6,22% |
| (-15°, 174,375°) | 5,91% | 6,32% |
| (-10°, -28,125°) | 8,42% | 8,81% |
| (-5°, 90°) | 8,33% | 8,18% |
| (5°, 151,875°) | 0,32% | 0,04% |
| (10°, 73,125°) | 12,51% | 12,21% |
| (25°, 0°) | 9,31% | 9,62% |
| (30°, 101.25°) | 0,07% | 0,08% |
| (30°, 163,125°) | 0,10% | 0,10% |
| (40°, -5,625°) | 5,82% | 6,58% |

Tabela 4.3: Erros da interpolação espectral. Filtros com 14 pólos e 14 zeros.

um dos 8 triângulos utilizados nos testes. São exibidas também a magnitude e a fase da média da respostas em freqüência para os 8 triângulos. Note que a resposta em freqüência do filtro interpolado aproxima-se bem da resposta em freqüência do filtro original. Note também que a disposição dos pólos e zeros do filtro interpolado é próxima à disposição dos pólos e zeros do filtro original. Para alguns pólos e zeros, há mesmo uma correspondência exata, mas mesmo os demais estão posicionados próximo aos pólos e zeros originais. Uma observação cabe aqui: nota-se descontinuidades nos gráficos das fases do filtro original. Isso é devido ao fato de o argumento de uma variável complexa ser uma *função multivalorada*, podendo assumir valores distintos na imagem para um mesmo valor no domínio. Para que $arg(z)$ passe a ser uma função, é necessário restringir sua imagem a uma certa faixa de valores (aqui, o intervalo $[-\pi, \pi)$ ), gerando as descontinuidades.

Como dito acima, um resultado típico dos testes não apresenta diferenças visíveis consideráveis entre a variante cartesiana e a polar. A figura 4.11 exemplifica uma tal situação, que neste caso ocorre para a direção (10°, 73,125°).

O pior caso encontrado nos testes, conforme se pode ver na tabela 4.2, foi o da direção (-30°, -39,375°) para filtros de 12 pólos e 12 zeros. A figura 4.12 exibe esse caso extremo. O erro cometido pela interpolação aqui foi bastante grande, de 30,35% na variante cartesiana e de 37,89% na variante polar.

Esse erro tão discrepante pode ser melhor compreendido ao se observar a figura 4.13, que contém

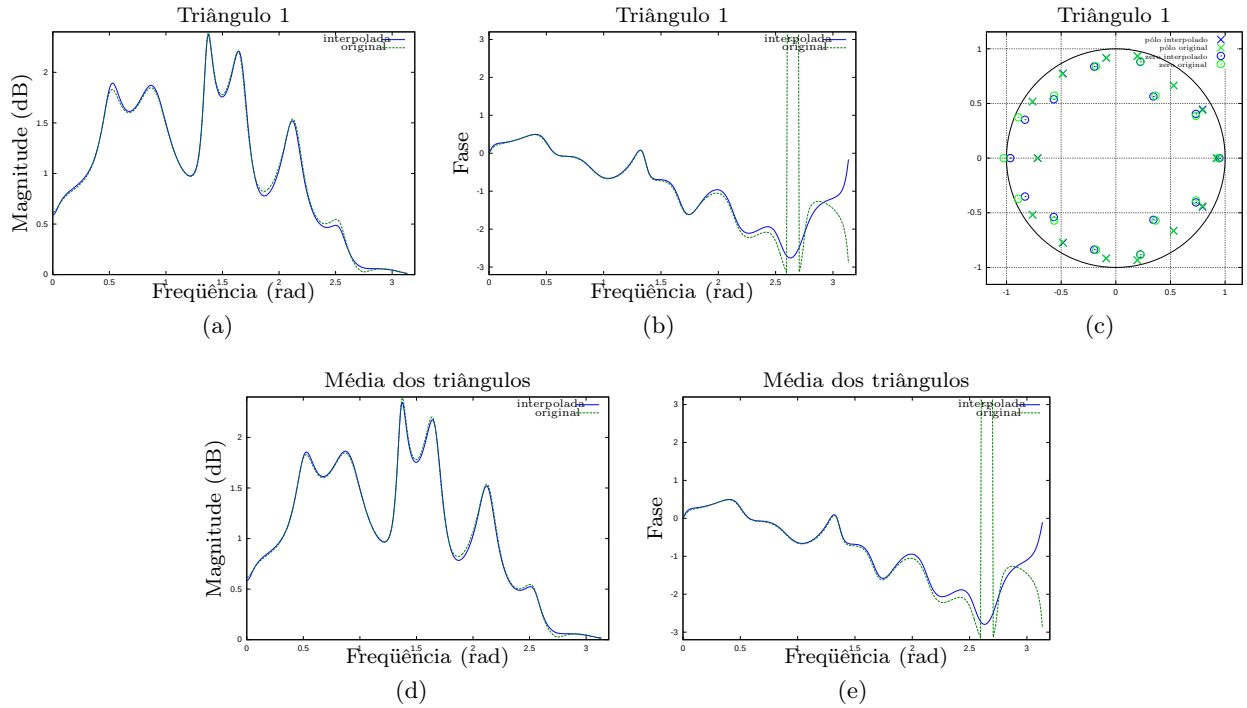
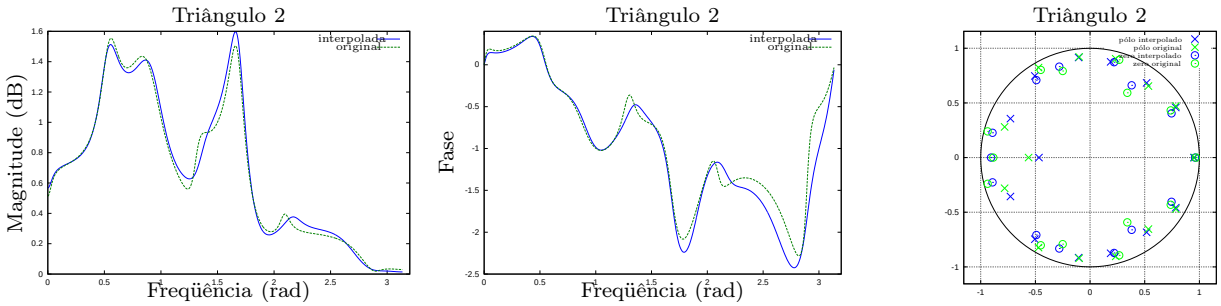
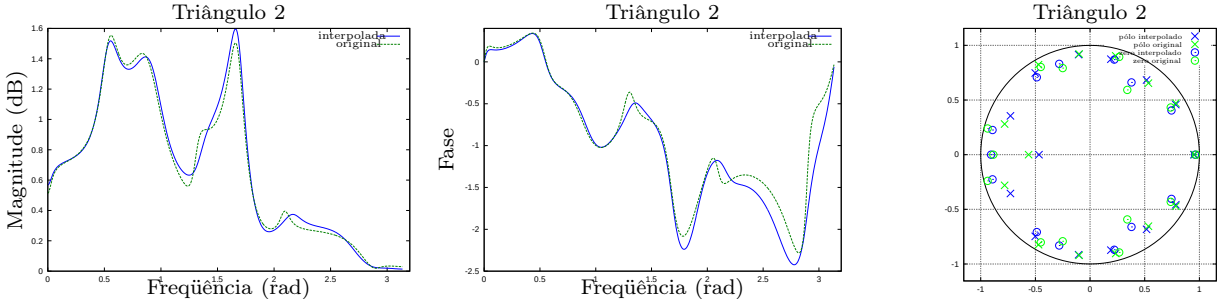



Figura 4.10: Filtros interpolados com 14 pólos e 14 zeros. Direção (-5°, 90°). Aqui se pode ver o gráfico da magnitude dos espectros dos filtros original e interpolado (4.10a), da fase (4.10b), e o diagrama de pólos e zeros dos dois filtros (4.10c), todos eles para a primeira das oito configurações de triângulo. São apresentados também os gráficos da magnitude (4.10d) e da fase (4.10e) da média dos triângulos.

os diagramas de pólos e zeros do filtro original, do filtro interpolado e dos 3 filtros utilizados para gerar o filtro interpolado. No filtro B, os 2 zeros reais estão posicionados próximos à região de freqüência $2\pi$ do espectro, enquanto que nos demais (incluindo aí o filtro original) há sempre um zero real próximo à região de freqüência 0. Disso resulta que o filtro interpolado ficou com um pólo isolado perto da freqüência 0 no círculo unitário, enquanto que no filtro original esse pólo é acompanhado de um zero. Isso fez aumentar muito a energia nas freqüências graves do espectro do filtro interpolado. Em contrapartida, os demais pólos do filtro interpolado estão todos muito próximos de um zero, o que reduz sua influência no espectro. Já no filtro original, os pólos estão posicionados entre os zeros, o que contribui para a formação dos picos no espectro do filtro original e que o filtro interpolado não é




(a) Variante polar.



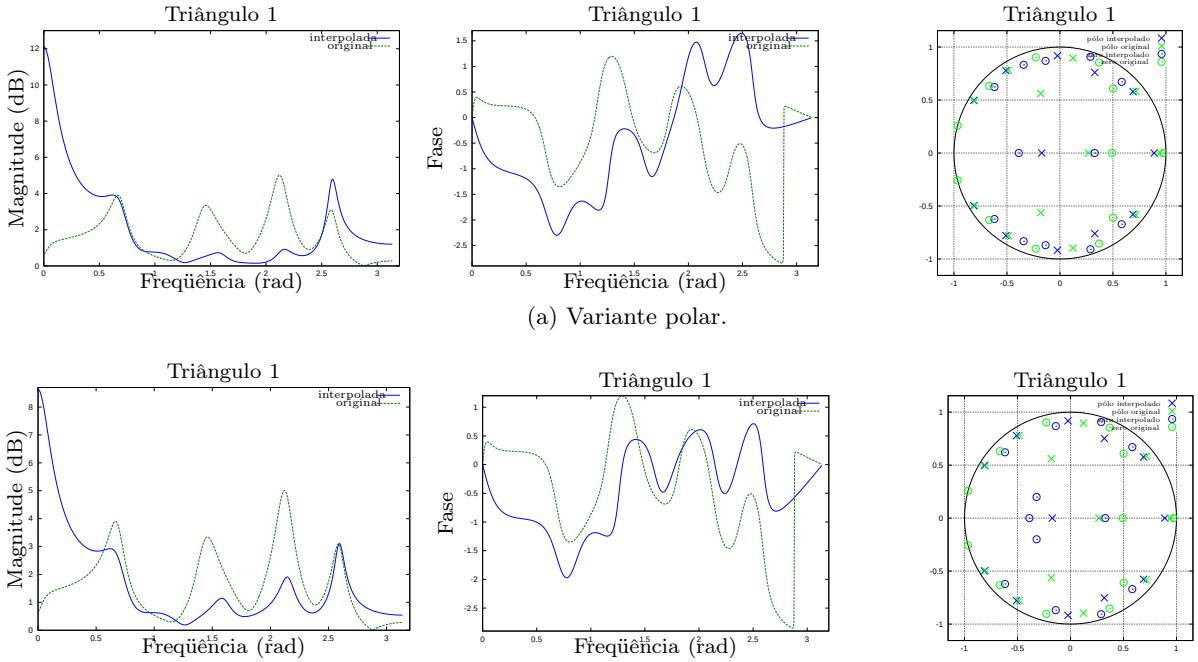

(b) Variante cartesiana.

Figura 4.11: Filtros interpolados com 14 pólos e 14 zeros. Direção (10°, 73,125°). Comparação entre a variante polar da interpolação espectral e a variante cartesiana. Em um caso típico, como este, não diferenças sensíveis entre uma e outra.

capaz de reproduzir. Esta já é uma situação mais complexa, porque envolveu uma combinação muito particular da disposição dos pólos e dos zeros no filtro interpolado, fazendo o efeito dos pólos ser atenuado pelos zeros.

Os resultados da direção (-30°, -39,375°) apresentam ainda um outro comportamento atípico: os resultados da variante cartesiana e da variante polar diferem bastante, e tais diferenças podem ser conferidas mesmo visualmente. A figura 4.12a exibe os gráficos para a variante polar e a figura 4.12b exibe os gráficos para a variante cartesiana. As diferenças são visíveis e claras. O mais notável é o que ocorre nos diagramas de pólos e zeros, em que há uma diferença muito grande no posicionamento de um par complexo-conjugado nas duas variantes (a polar e a cartesiana).

Os resultados apresentados nas tabelas 4.1, 4.2 e 4.3 são, no geral, bastante representativos do




(a) Variante polar.

(b) Variante cartesiana.

Figura 4.12: Filtros interpolados com 12 pólos e 12 zeros. Direção (-30°, -39,375°).

comportamento real dos filtros interpolados. Entretanto, como os erros exibidos lá são calculados a partir das médias dos resultados para as 8 configurações de triângulos, há alguns casos em que a média dos resultados apresenta bons resultados mas acaba mascarando algumas variações na interpolação para cada triângulo. A figura 4.14 exibe um desses casos, correspondente à direção (40°, -5,625°). São mostrados os resultados de 3 configurações de triângulos, que diferem sensivelmente entre si, e é mostrado também o resultado da média das 8 configurações. Todas essas configurações possuem erro ligeiramente maior que o erro da média, conforme pode ser visto na tabela 4.4.

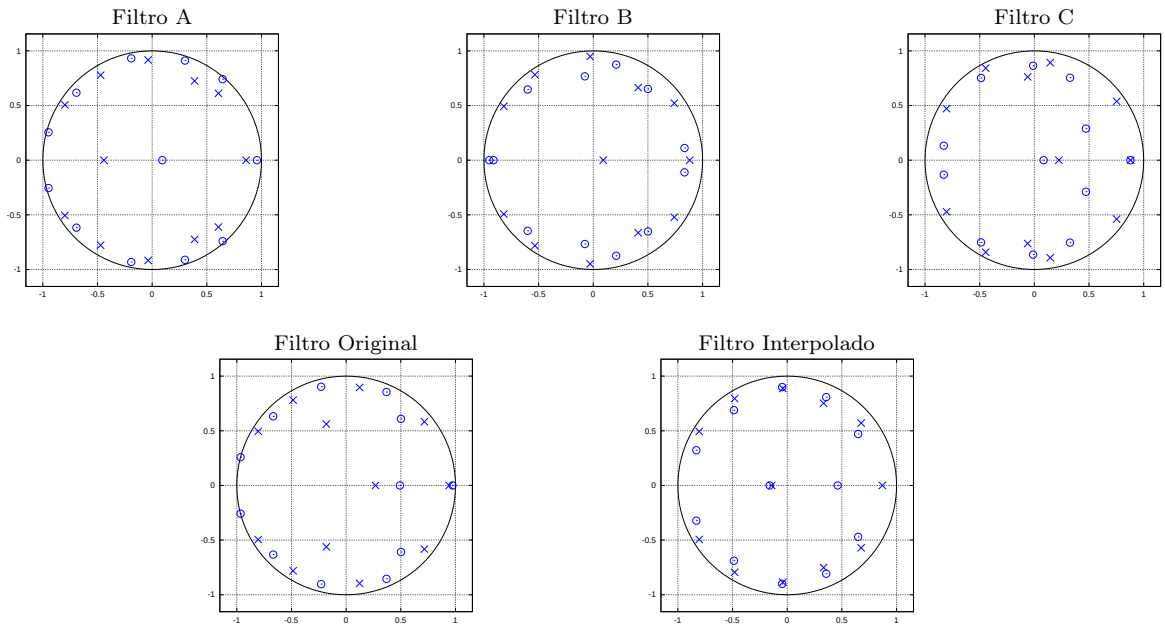



Figura 4.13: Diagramas de pólos e zeros para a direção (-30°, -39,375°) dos filtros original, interpolado, e dos 3 filtros usados na interpolação.

| **Triângulo** | **Erro** |
|---|---|
| 4 | 7,04% |
| 6 | 7,20% |
| 7 | 6,41% |
| Média dos 8 triângulos | 5,82% |

Tabela 4.4: Erros da interpolação espectral para algumas configurações de triângulo para a direção (40°, -5,625°) e filtros de 14 pólos e 14 zeros.

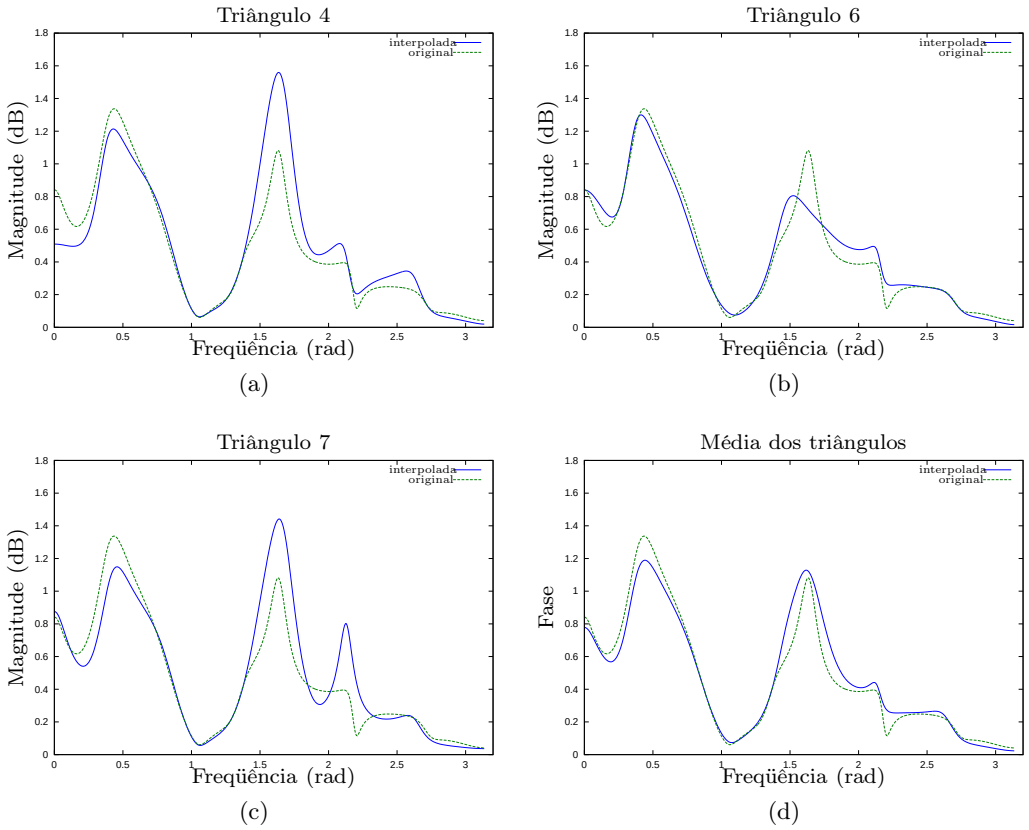



Figura 4.14: Filtros interpolados com 14 pólos e 14 zeros. Direção (40°, -5,625°).

Isto conclui a exposição dos casos representativos. Como dito anteriormente, para quem desejar uma visão mais detalhada dos resultados, o conjunto completo dos gráficos dos testes da interpolação espectral para filtros de 14 pólos e 14 zeros encontra-se no apêndice B.

## Capítulo 5

# Conclusões

Este trabalho apresentou três novas técnicas de interpolação de HRTFs, com o intuito de auralizar fontes sonoras móveis para reprodução binaural. Ao empregar HRTFs para criar o efeito de fontes sonoras se locomovendo, a interpolação faz-se necessária para se gerar trajetórias contínuas para o movimento, dado que cada HRTF está associada de maneira única a uma direção específica. Embora a interpolação de HRTFs não seja um conceito novo, as técnicas desenvolvidas aqui trazem idéias ainda não exploradas em outros trabalhos.

A interpolação triangular apresentou a possibilidade de utilização de configurações triangulares de HRTFs na interpolação, através do emprego da técnica de VBAP, que fora originalmente concebida para espacialização multicanal. Essa mesma idéia pôde, depois, ser empregada com sucesso na interpolação espectral. Comparada com a interpolação bilinear [Huo99, SHLV99], a interpolação triangular apresenta economia de recursos computacionais de cerca de 25%. Os testes executados mostraram que as HRTFs geradas pela interpolação triangular, no geral, conseguem reproduzir bem o comportamento das HRTFs originais.

O método das movimentações discretas apresenta uma idéia bastante simples para o problema de auralizar fontes móveis: a cada deslocamento da fonte, usar como filtro para auralização a HRTF que estiver mais próxima à direção atual da fonte, sem interpolações ou geração de HRTFs estimadas. A economia de recursos computacionais aqui é evidente. Para conseguir alcançar esse nível de simplificação, a técnica pressupõe que os pontos de medição do banco de HRTFs estejam mais próximos entre si do que a menor distância perceptível pelo sistema auditivo humano.

Um problema encontrado ao se usar o método das movimentações discretas é que ele depende de um banco de HRTFs mais refinado do que os disponíveis atualmente. Sem isso, seu universo de

aplicações fica bastante reduzido, sendo necessário impor restrições à movimentação da fonte sonora, como, por exemplo, restringi-la a um plano vertical, onde o limiar de percepção da movimentação da fonte é mais elevado. Contudo, o fato de o limiar de percepção da direção diminuir rapidamente com a diminuição da inclinação do plano de movimentação da fonte (indo de uma inclinação vertical até uma horizontal) indica que a ITD e a ILD podem ter papel primordial na determinação do deslocamento da fonte. Isso aponta para a possibilidade de se estudar uma variação do método na qual se combinem HRTFs de fase mínima com ITD+ILD. Se a troca de HRTFs com a movimentação da fonte fica limitada à densidade do banco de dados, isso já não é verdade para a ITD e a ILD, que podem ser razoavelmente bem estimadas através de modelos geométricos, podendo, essas sim, variarem continuamente conforme a fonte se desloca. Essa combinação de informações talvez forneça elementos suficientes para o sistema sensorial perceber a movimentação da fonte sonora mesmo sem a troca de HRTFs. Experimentos psicoacústicos que serão conduzidos em estudos futuros poderão avaliar a viabilidade dessa idéia.

Uma outra possibilidade de se contornar a falta de um banco de HRTFs suficientemente denso seria gerar as HRTFs faltantes através de interpolação. Neste caso, seria possível, por exemplo, reduzir o menor ângulo entre HRTFs vizinhas para 1°, possibilitando a aplicação do método das movimentações discretas. Uma vez que o processo de interpolação seria conduzido em uma etapa de pré-processamento, o desempenho do método seria preservado, o que permitiria utilizá-lo em aplicações que exijam alto desempenho. Esta é uma idéia que deve ser também estudada em trabalhos futuros.

A interpolação espectral difere substancialmente das técnicas anteriores. Ela procura alterar as estruturas das HRTFs através da interpolação de seus pólos e zeros, ao invés de simplesmente executar uma soma ponderada de HRTFs. Dado que há uma mudança contínua da posição dos pólos e zeros do filtro, há também uma alteração gradual da influência de cada um nas regiões do espectro do filtro interpolado, promovendo uma alteração contínua no seu comportamento espectral de uma maneira tal que pode não ser conseguida através de interpolações lineares.

Um problema inerente à interpolação espectral, e que restringiu muito sua implementação neste trabalho, foi sua incapacidade de interpolar corretamente filtros com número diferente de pólos ou de zeros reais, gerando filtros com coeficientes complexos ao invés de reais. Nos testes implementados aqui, foi necessário escolher um subconjunto do banco de HRTFs IIR no qual todos os filtros possuíam número igual de pólos reais e número igual de zeros reais. Um método para homogeneizar as estruturas dos filtros IIR em termos do número de pólos e zeros reais foi proposto na seção 3.3, mas será necessária

uma investigação mais aprofundada antes de se poder aplicar esta técnica a um banco inteiro de filtros HRTF.

Uma outra limitação está no algoritmo de associação de pólos e zeros. Embora os resultados tenham se mostrado satisfatórios para filtros de até 14 pólos e 14 zeros, o aumento na ordem dos filtros introduz um crescimento muito rápido nas possibilidades de associações, tornando mais complexa a determinação da melhor associação.

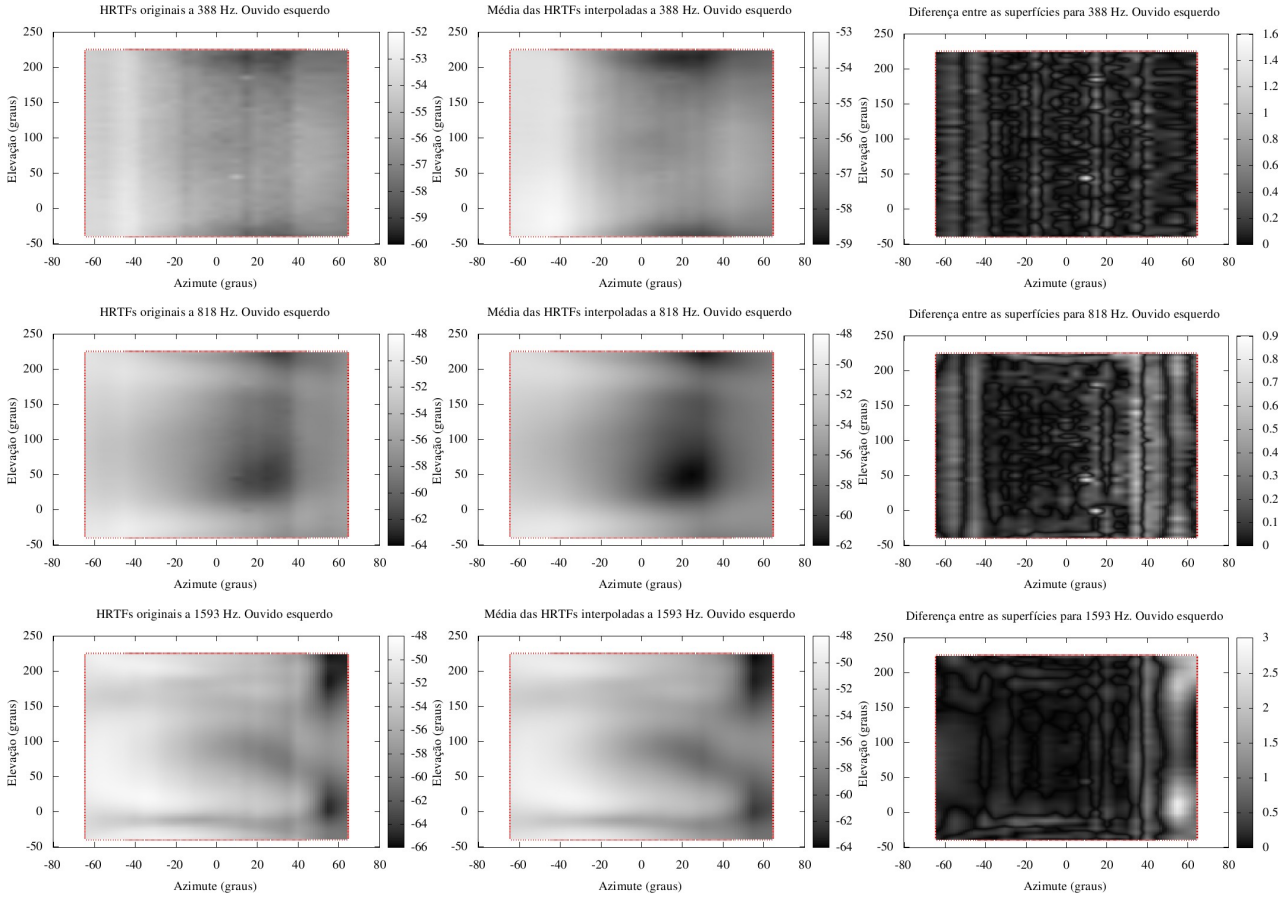
Estudos futuros devem trabalhar no desenvolvimento de uma variante do método de Kalman que permita impor restrições às soluções, possibilitando, por exemplo, estabelecer uma faixa de valores para o argumento de cada pólo ou zero. Assim, se o filtro possui 6 pólos (tendo 3 deles parte imaginária positiva), as restrições à solução poderiam impor que o argumento de um dos pólos tivesse valor entre 0 e $\pi/3$, que o argumento de outro pólo estivesse entre $\pi/3$ e $2\pi/3$ e que o argumento do terceiro pólo estivesse entre $2\pi/3$ e $\pi$. Conseguindo isso, o algoritmo de associação dos pólos e dos zeros seria trivial, bastando agrupar aqueles cujo argumento estivesse na mesma faixa de valores. Isso resolveria ambos os problemas descritos no parágrafo anterior.

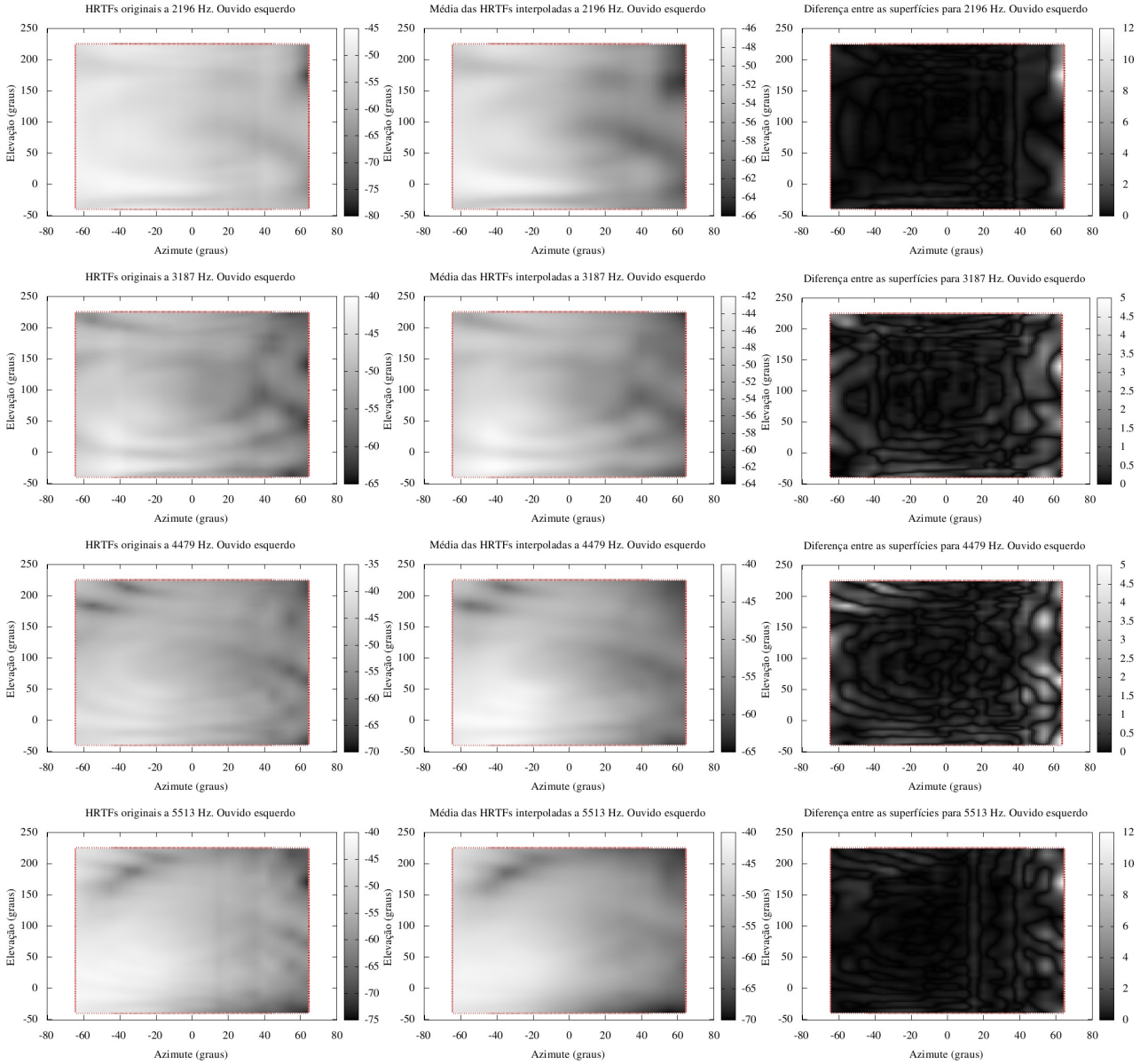
Serão feitas implementações dessas técnicas na forma de *externals* para o `Pure Data`, que permitirão seu emprego em aplicações em tempo real por toda a comunidade de usuários desse sistema. Tais *externals* serão publicados e estarão disponíveis a todos na internet, bem como seu código-fonte.

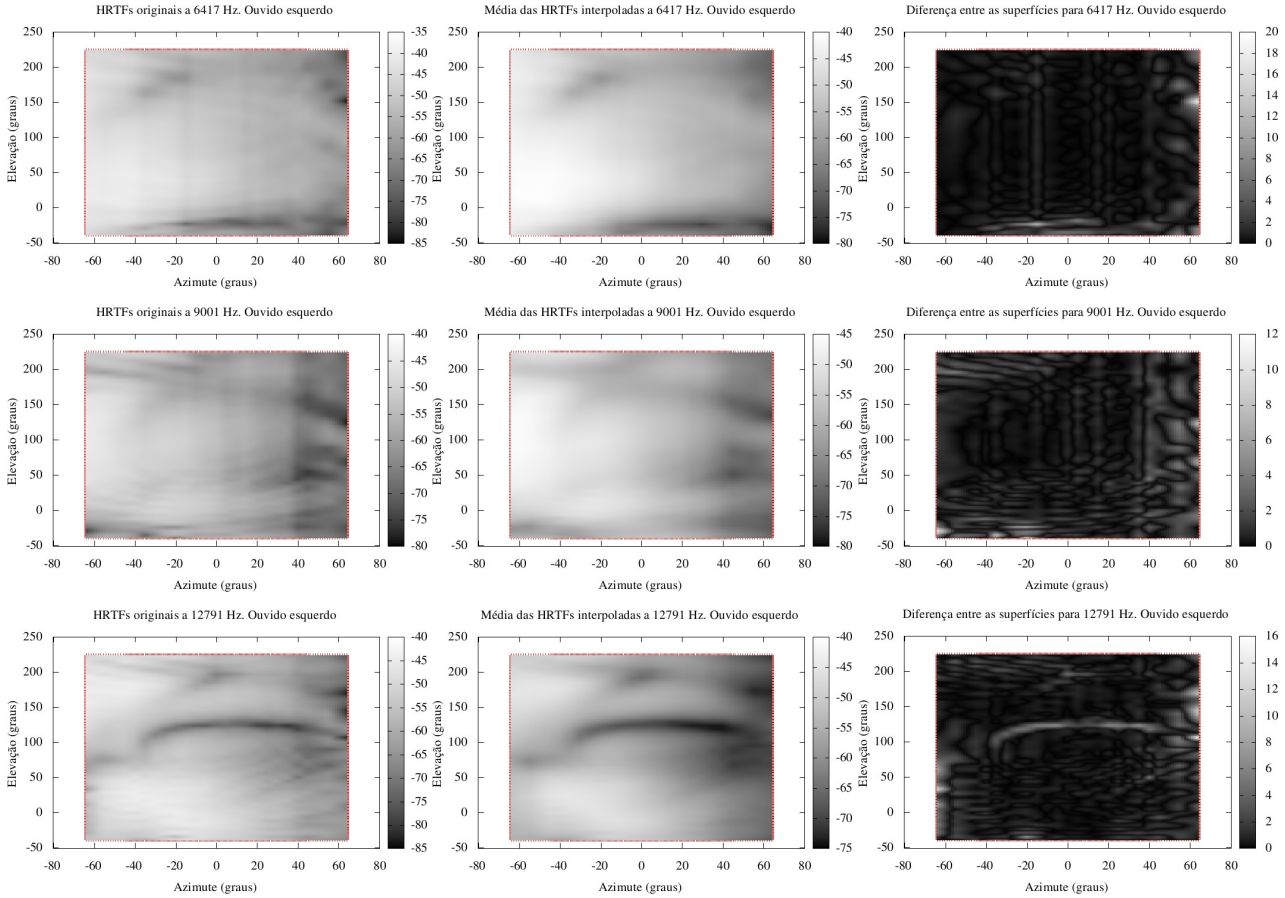
# Apêndice A

# Gráficos da interpolação triangular

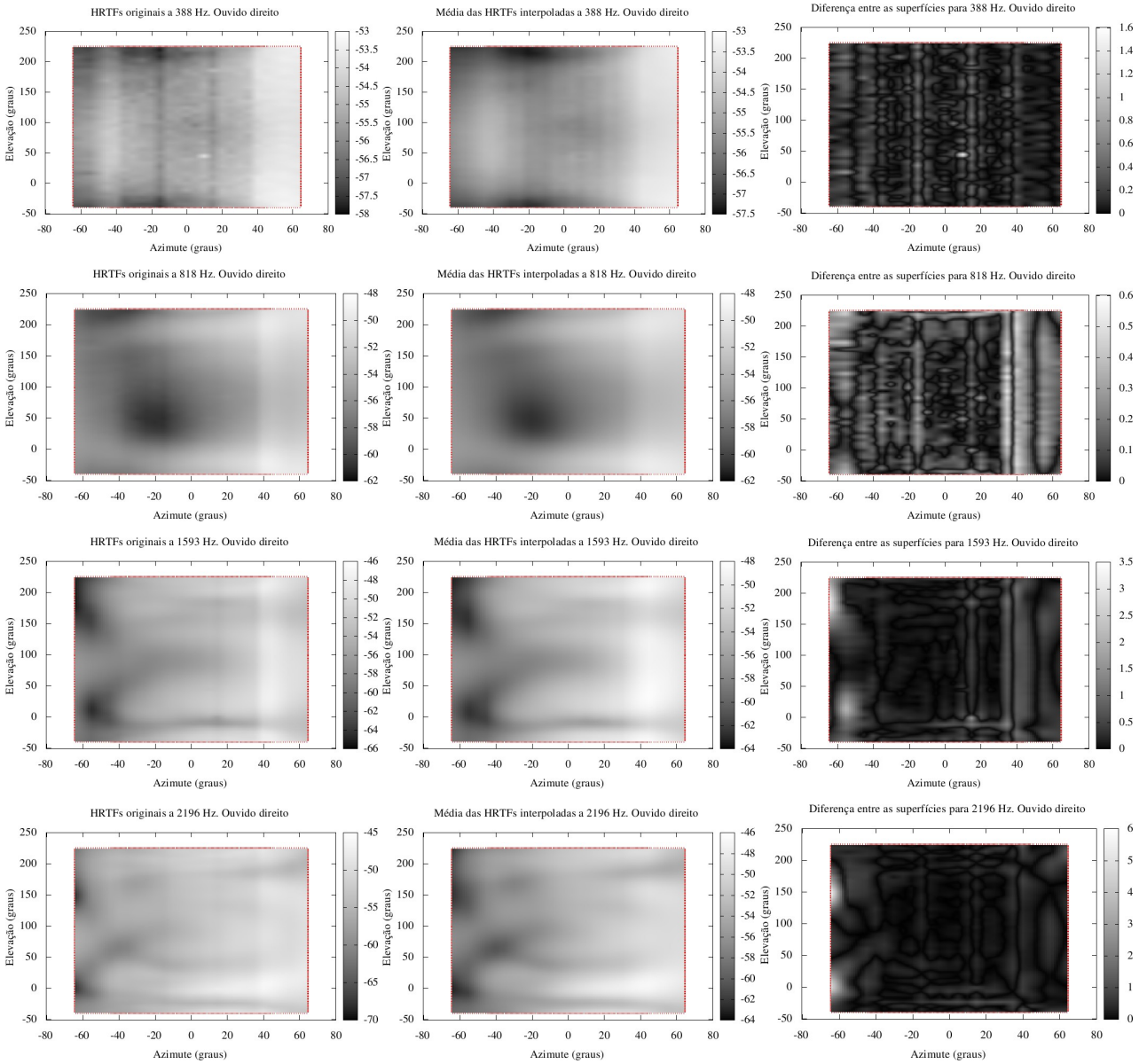
## Ouvido esquerdo

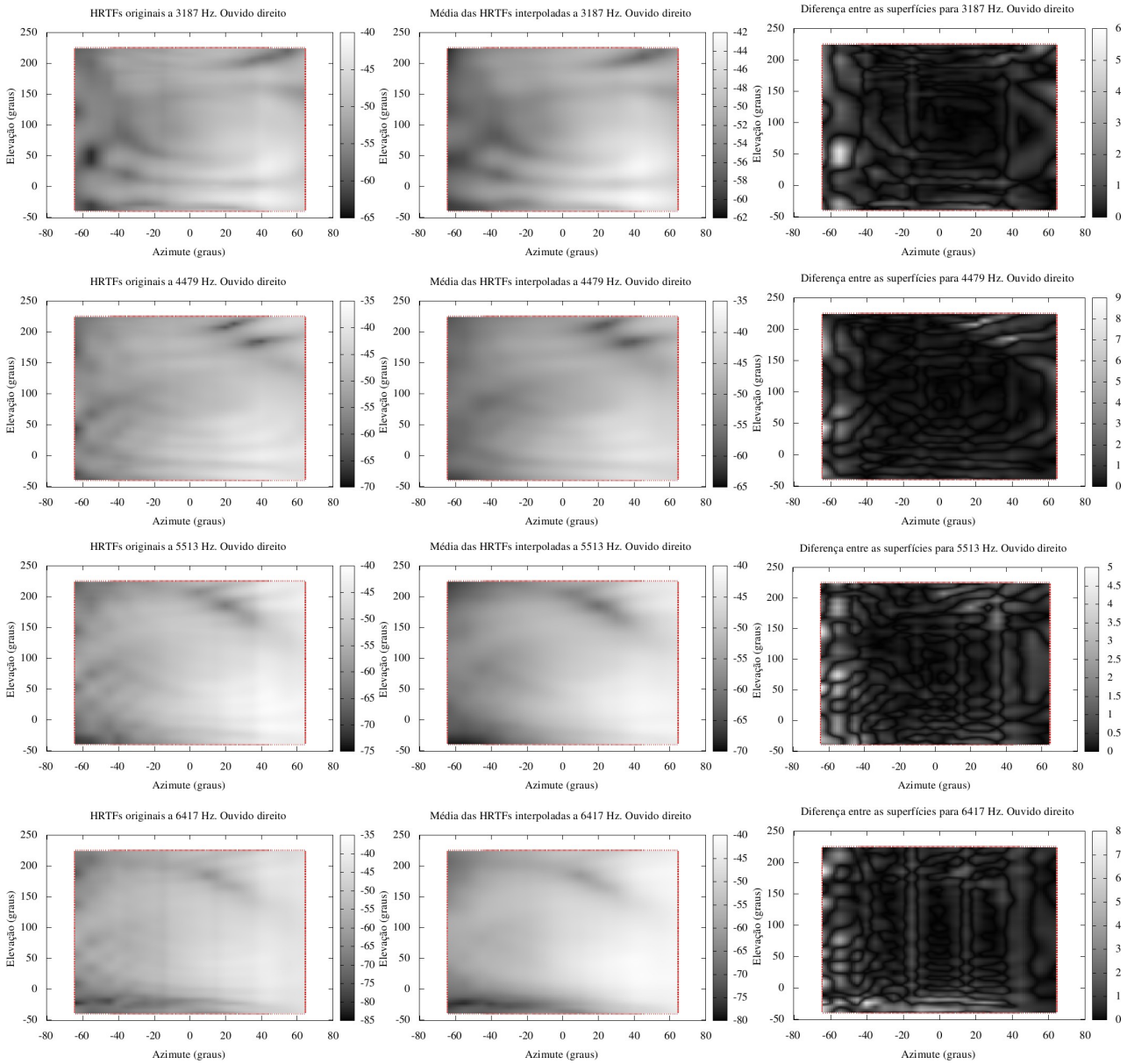


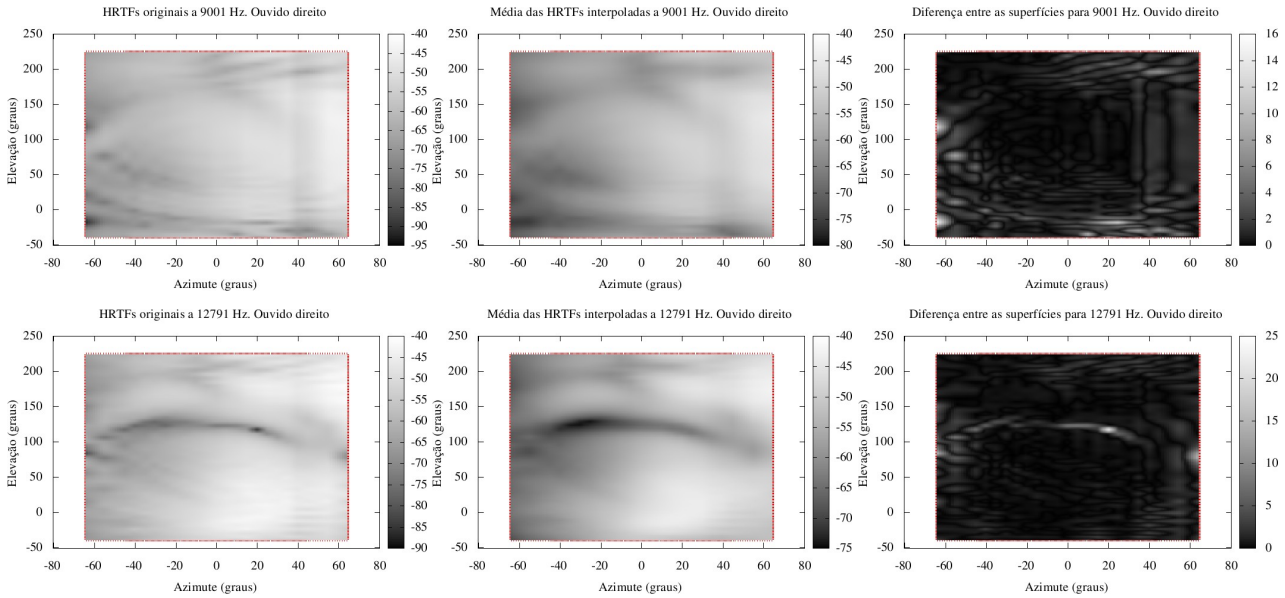

HRTFs originais a 2196 Hz. Ouvido esquerdo
Média das HRTFs interpoladas a 2196 Hz. Ouvido esquerdo
Diferença entre as superfícies para 2196 Hz. Ouvido esquerdo
HRTFs originais a 3187 Hz. Ouvido esquerdo
Média das HRTFs interpoladas a 3187 Hz. Ouvido esquerdo
Diferença entre as superfícies para 3187 Hz. Ouvido esquerdo
HRTFs originais a 4479 Hz. Ouvido esquerdo
Média das HRTFs interpoladas a 4479 Hz. Ouvido esquerdo
Diferença entre as superfícies para 4479 Hz. Ouvido esquerdo
HRTFs originais a 5513 Hz. Ouvido esquerdo
Média das HRTFs interpoladas a 5513 Hz. Ouvido esquerdo
Diferença entre as superfícies para 5513 Hz. Ouvido esquerdo
Elevação (graus)
Azimute (graus)


HRTFs originais a 6417 Hz. Ouvido esquerdo
Média das HRTFs interpoladas a 6417 Hz. Ouvido esquerdo
Diferença entre as superfícies para 6417 Hz. Ouvido esquerdo
HRTFs originais a 9001 Hz. Ouvido esquerdo
Média das HRTFs interpoladas a 9001 Hz. Ouvido esquerdo
Diferença entre as superfícies para 9001 Hz. Ouvido esquerdo
HRTFs originais a 12791 Hz. Ouvido esquerdo
Média das HRTFs interpoladas a 12791 Hz. Ouvido esquerdo
Diferença entre as superfícies para 12791 Hz. Ouvido esquerdo
Elevação (graus)
Azimute (graus)

## Ouvido direito



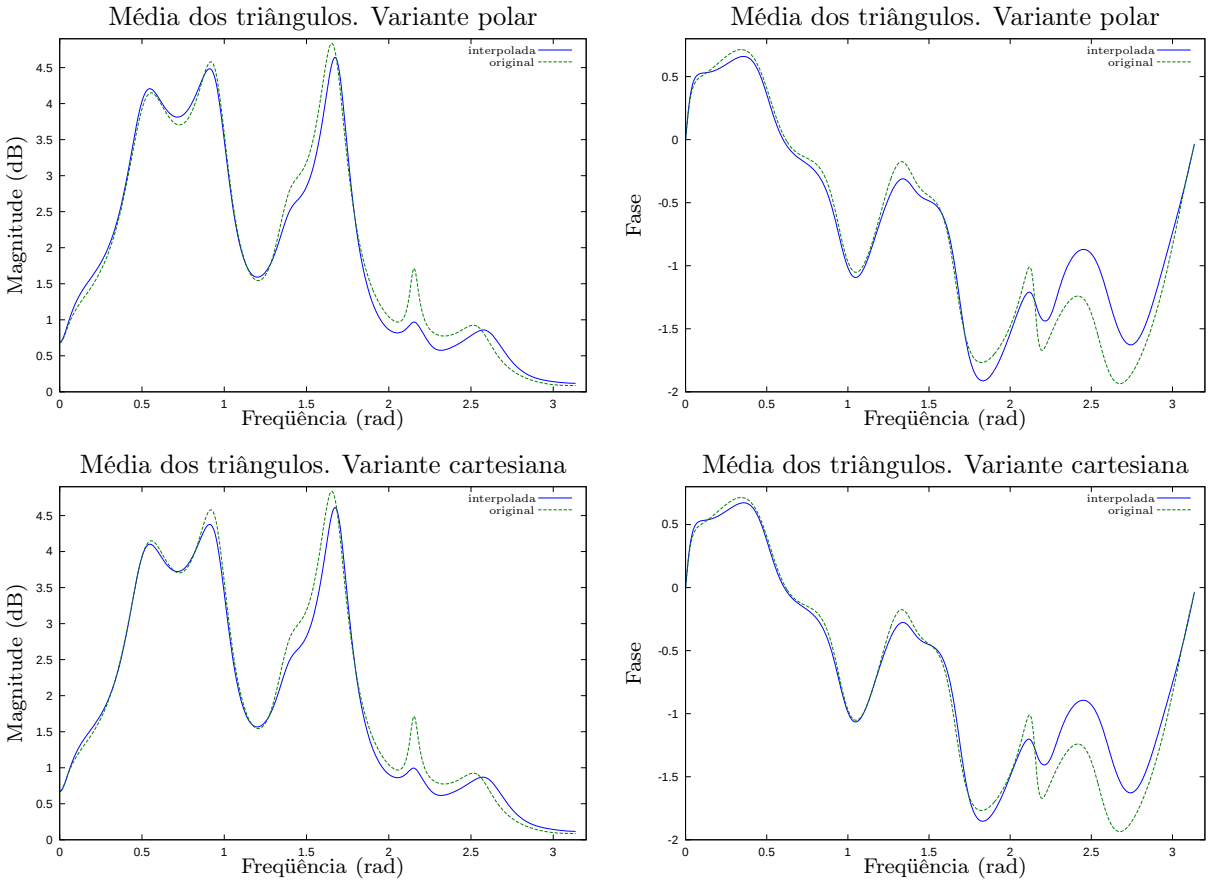

HRTFs originais a 3187 Hz. Ouvido direito
Média das HRTFs interpoladas a 3187 Hz. Ouvido direito
Diferença entre as superfícies para 3187 Hz. Ouvido direito
HRTFs originais a 4479 Hz. Ouvido direito
Média das HRTFs interpoladas a 4479 Hz. Ouvido direito
Diferença entre as superfícies para 4479 Hz. Ouvido direito
HRTFs originais a 5513 Hz. Ouvido direito
Média das HRTFs interpoladas a 5513 Hz. Ouvido direito
Diferença entre as superfícies para 5513 Hz. Ouvido direito
HRTFs originais a 6417 Hz. Ouvido direito
Média das HRTFs interpoladas a 6417 Hz. Ouvido direito
Diferença entre as superfícies para 6417 Hz. Ouvido direito
Elevação (graus)
Azimute (graus)


HRTFs originais a 9001 Hz. Ouvido direito
Média das HRTFs interpoladas a 9001 Hz. Ouvido direito
Diferença entre as superfícies para 9001 Hz. Ouvido direito
HRTFs originais a 12791 Hz. Ouvido direito
Média das HRTFs interpoladas a 12791 Hz. Ouvido direito
Diferença entre as superfícies para 12791 Hz. Ouvido direito
Elevação (graus)
Azimute (graus)

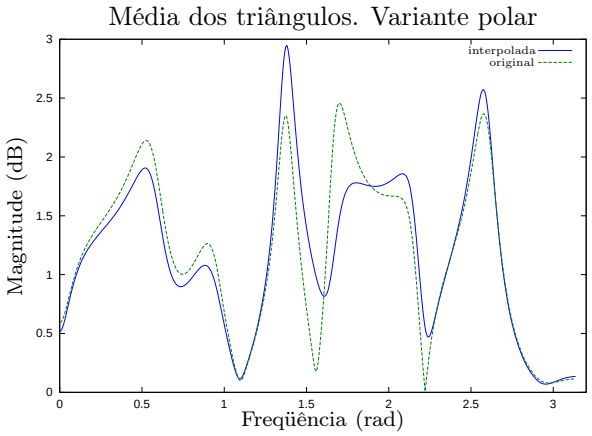
# Apêndice B

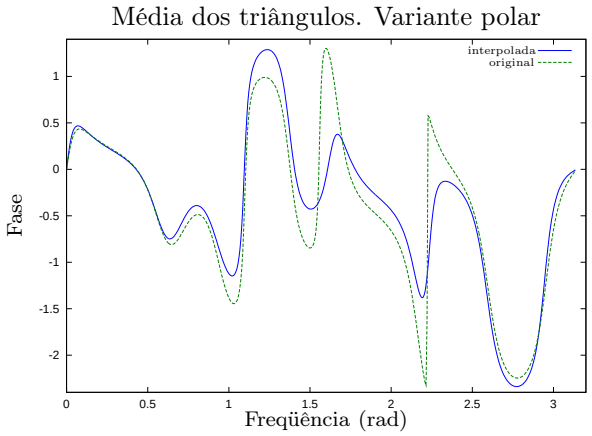
# Gráficos da interpolação espectral

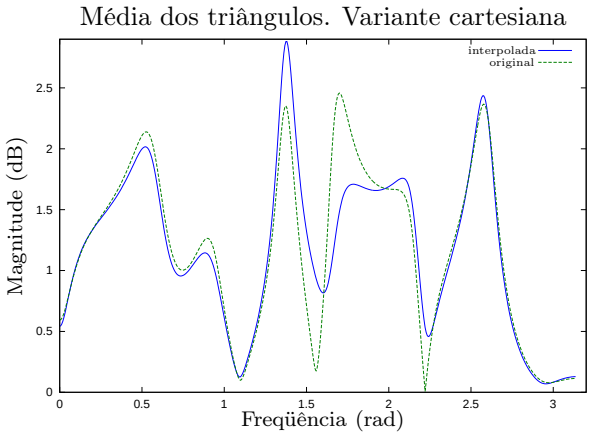
## Direção (-40°, 45°)

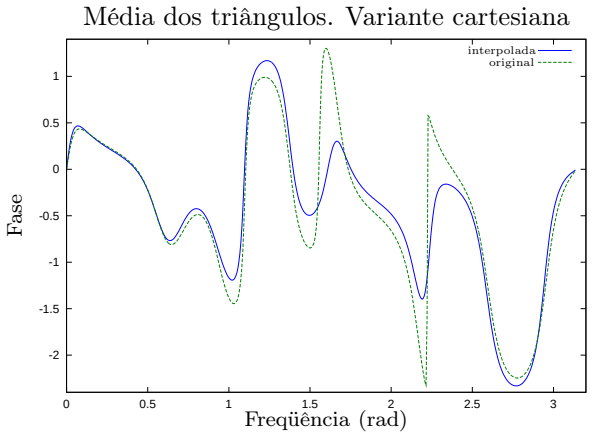


## Direção (-25°, 225°)

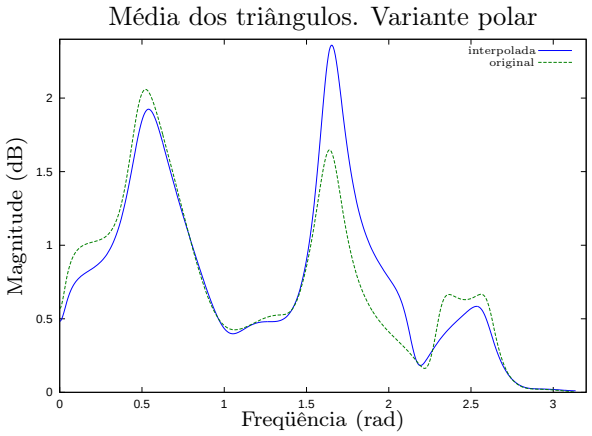


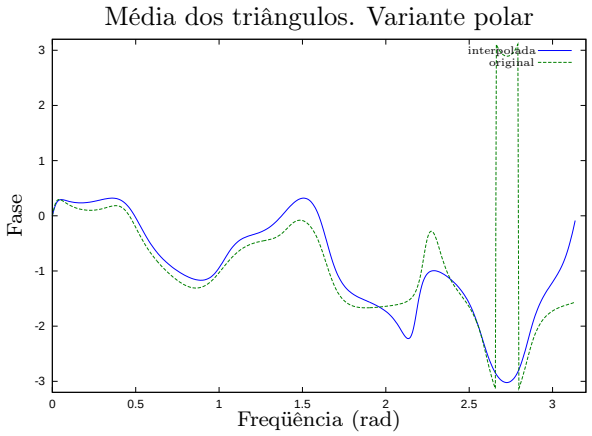



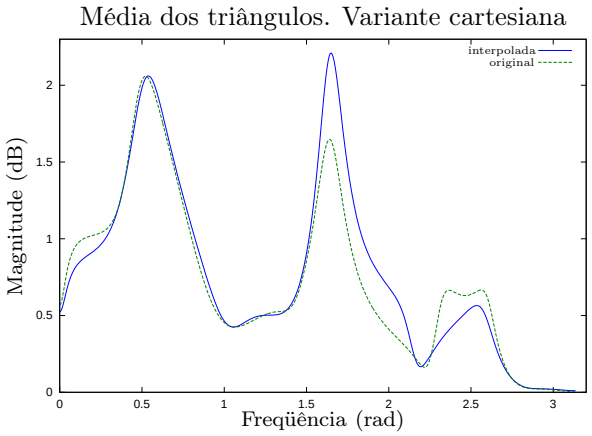



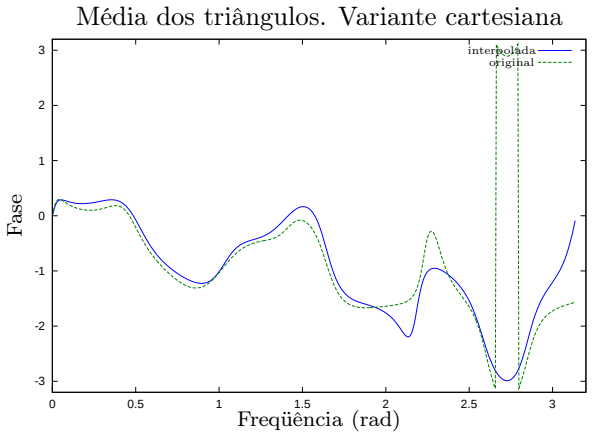



## Direção (-15°, 174,38°)

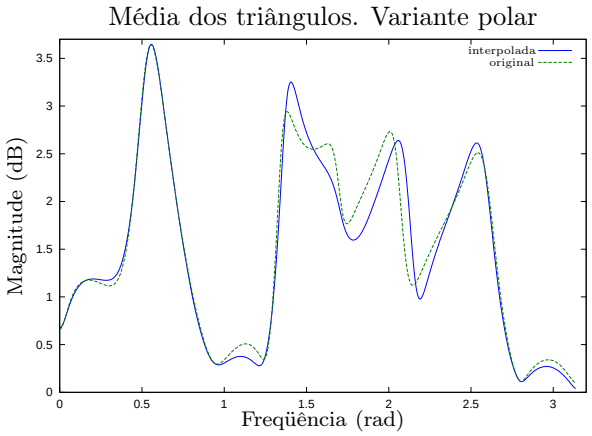


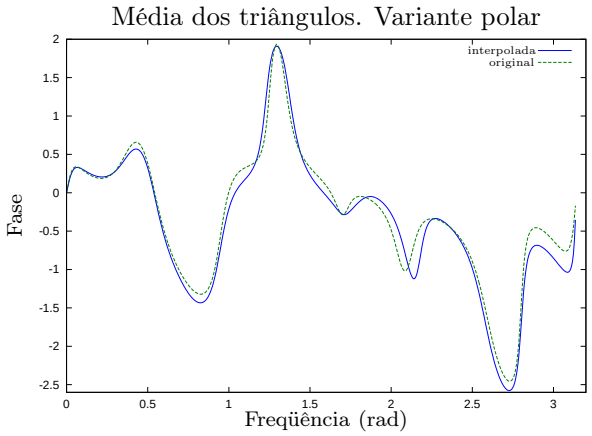



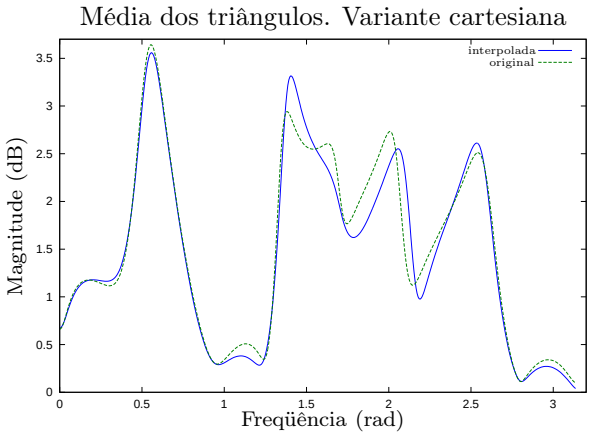



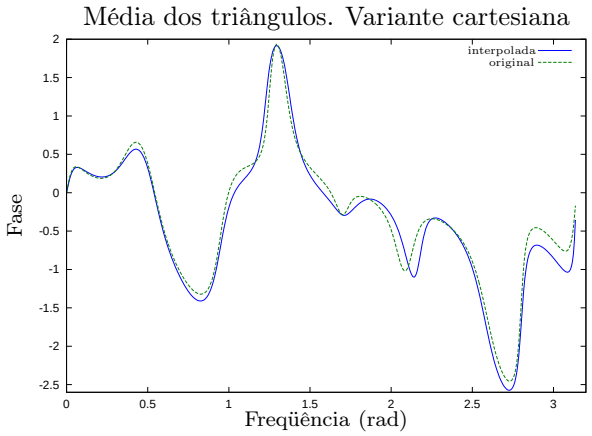



**Direção (-10°, -28,125°)**

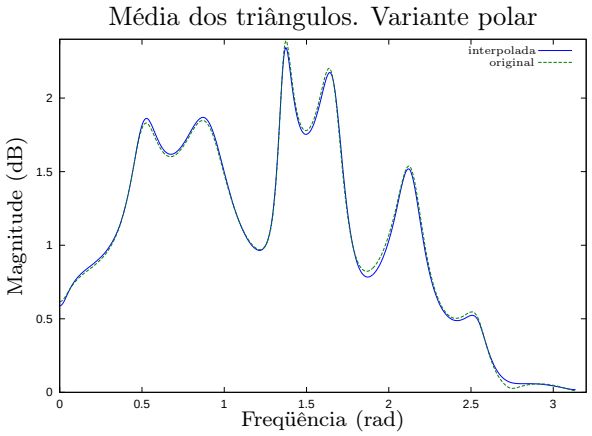


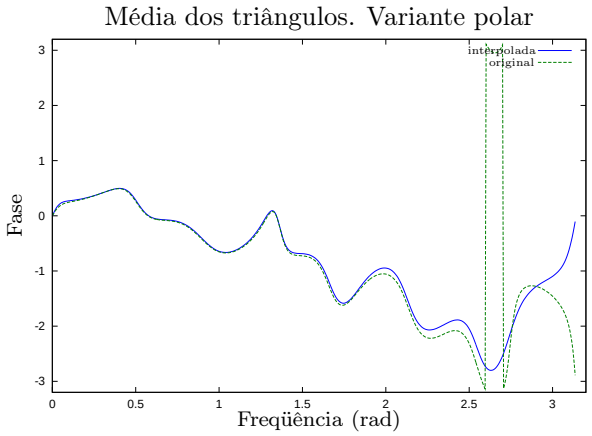



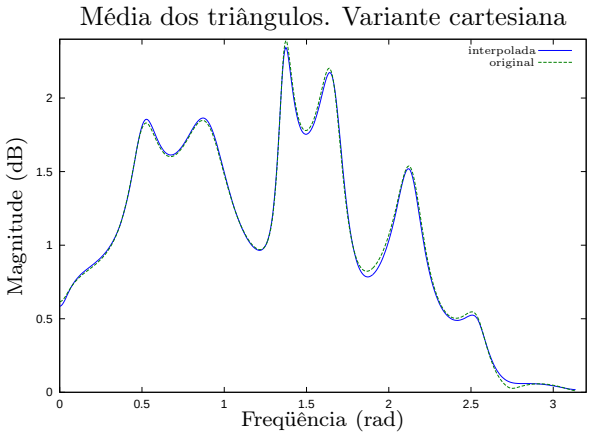



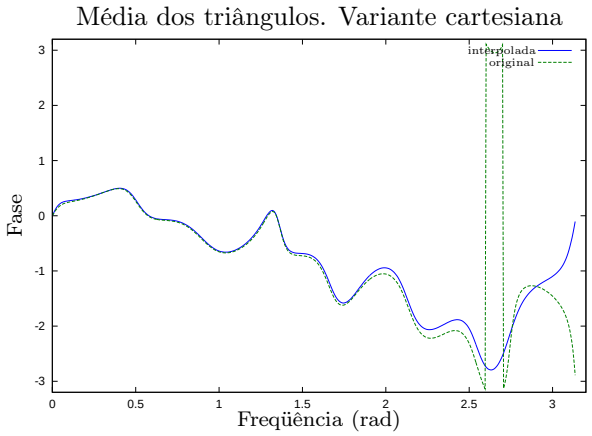



## Direção (-5°, 90°)

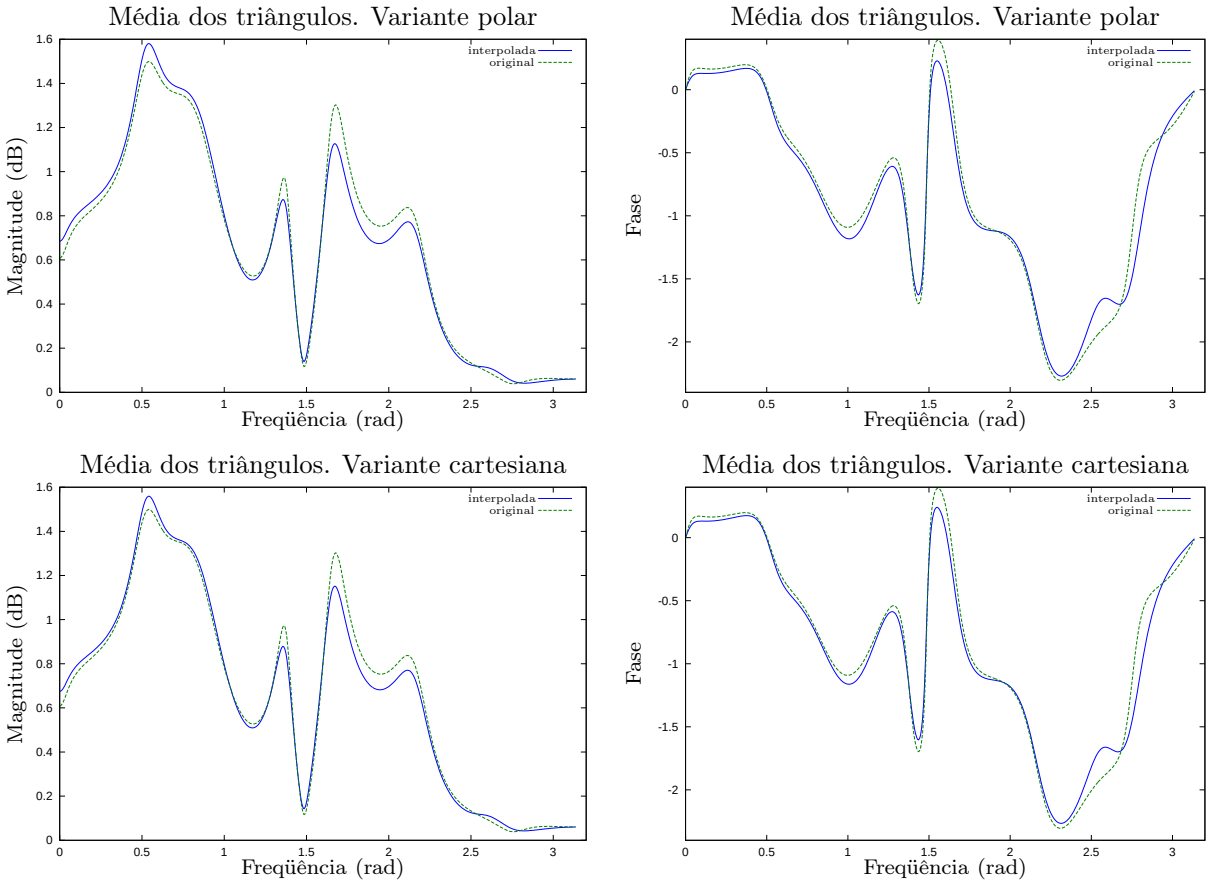

Média dos triângulos. Variante polar
interpolada
original
Magnitude (dB)
Freqüência (rad)


Média dos triângulos. Variante polar
interpolada
original
Fase
Freqüência (rad)

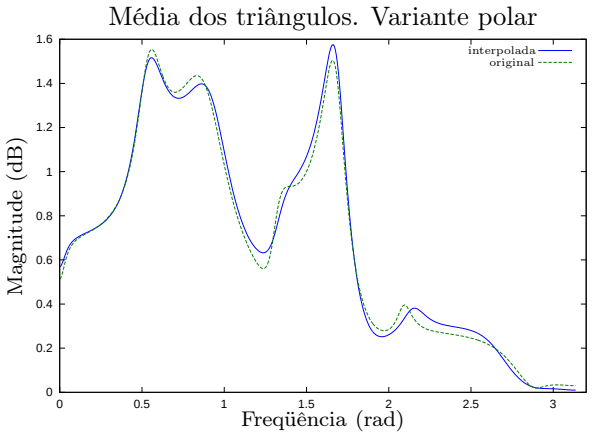

Média dos triângulos. Variante cartesiana
interpolada
original
Magnitude (dB)
Freqüência (rad)

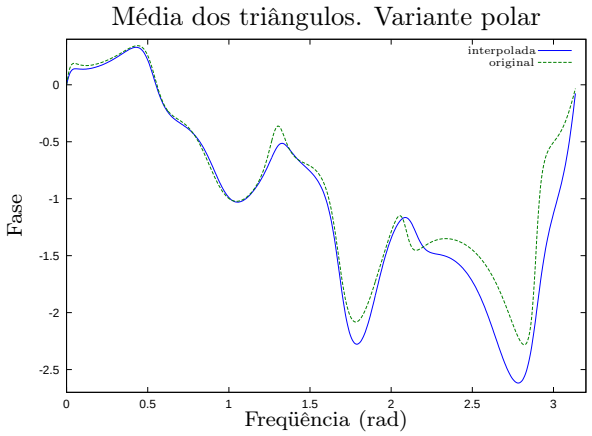

Média dos triângulos. Variante cartesiana
interpolada
original
Fase
Freqüência (rad)

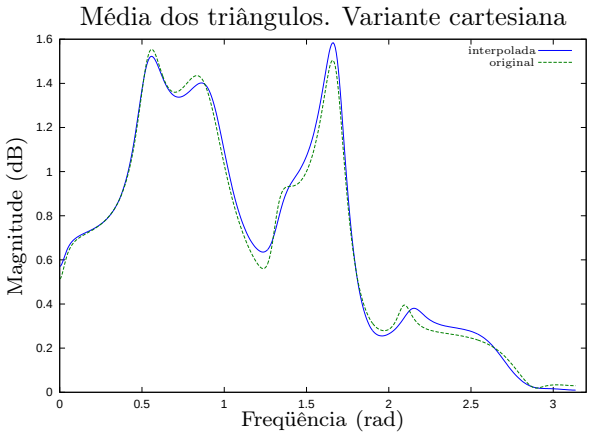
## Direção (5°, 151,88°)

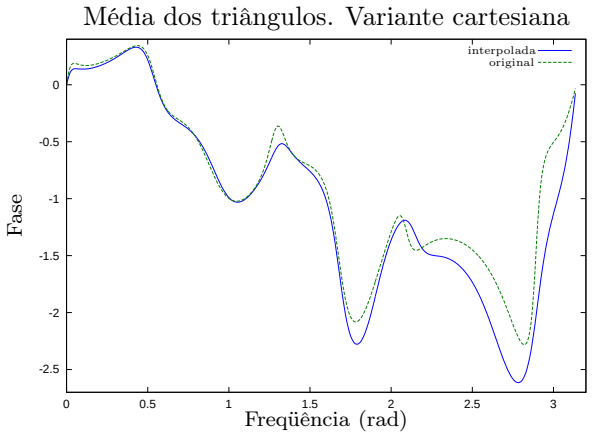


## Direção (10°, 73,125°)

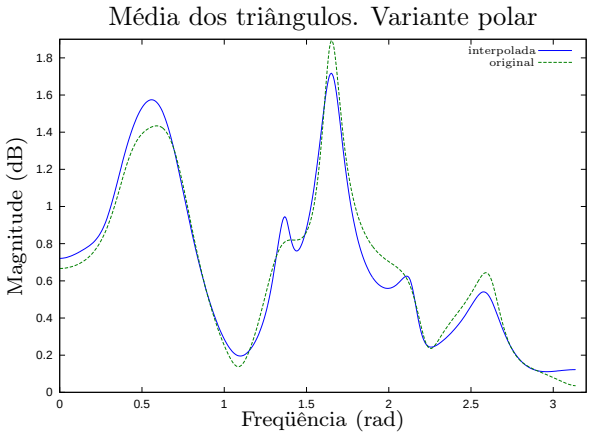


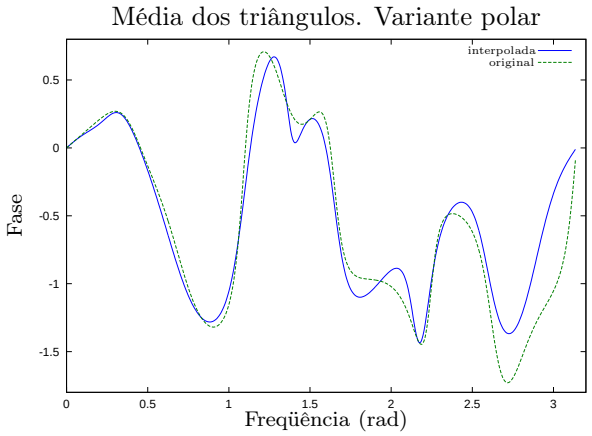



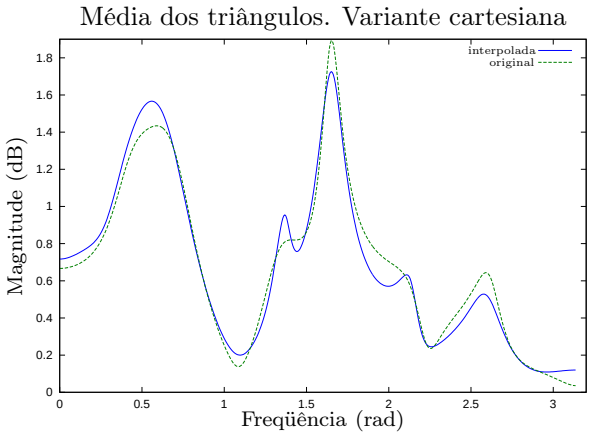



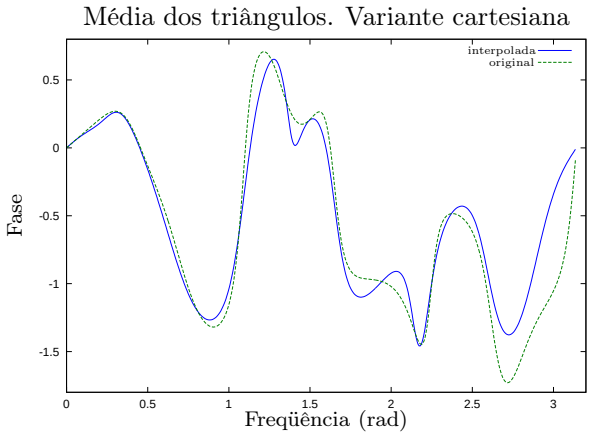



## Direção (25°, 0°)

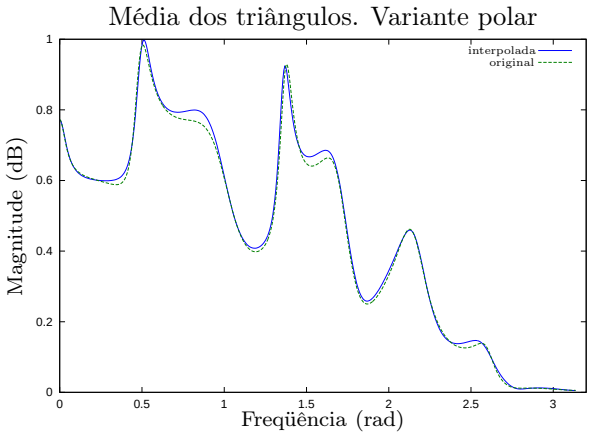

Média dos triângulos. Variante polar
interpolada
original
Magnitude (dB)
Freqüência (rad)

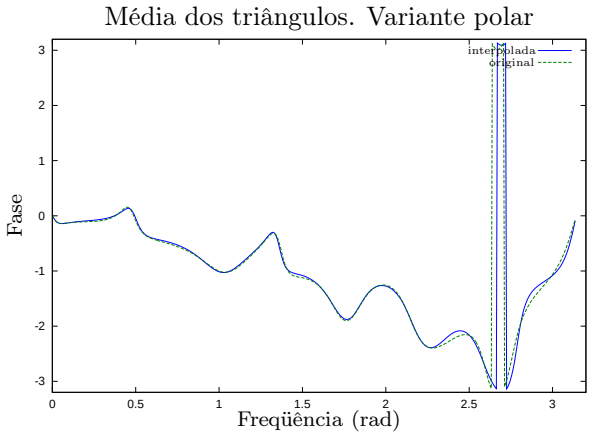

Média dos triângulos. Variante polar
interpolada
original
Fase
Freqüência (rad)

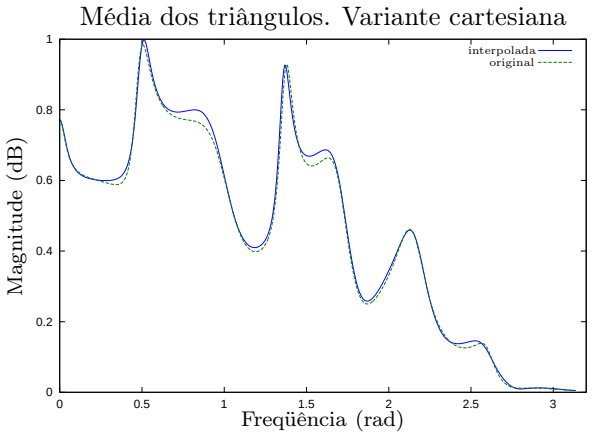

Média dos triângulos. Variante cartesiana
interpolada
original
Magnitude (dB)
Freqüência (rad)

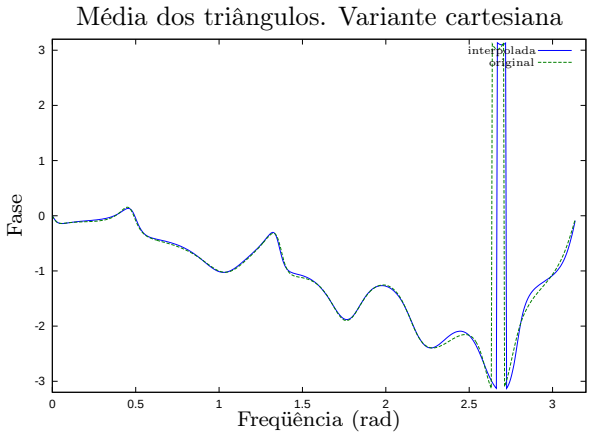

Média dos triângulos. Variante cartesiana
interpolada
original
Fase
Freqüência (rad)

## Direção (30°, 101,25°)

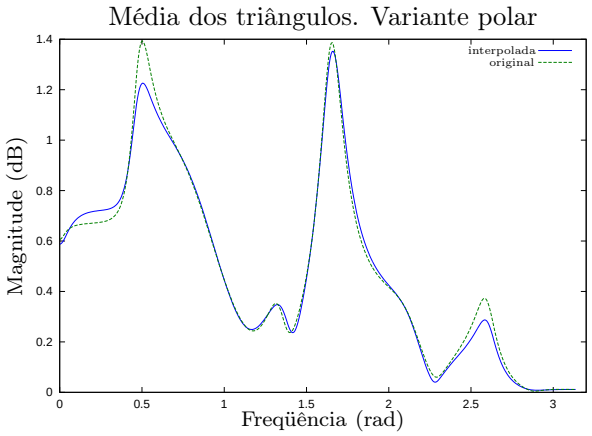


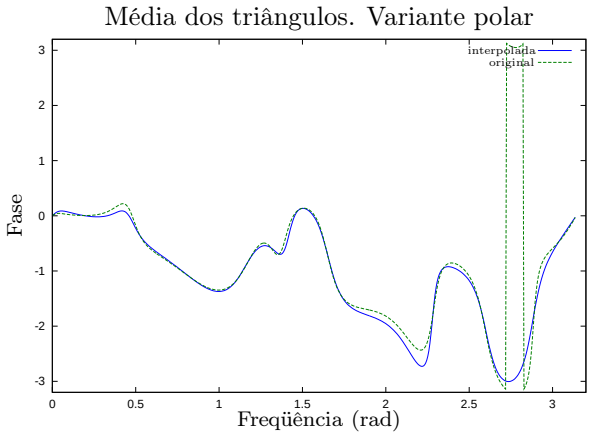



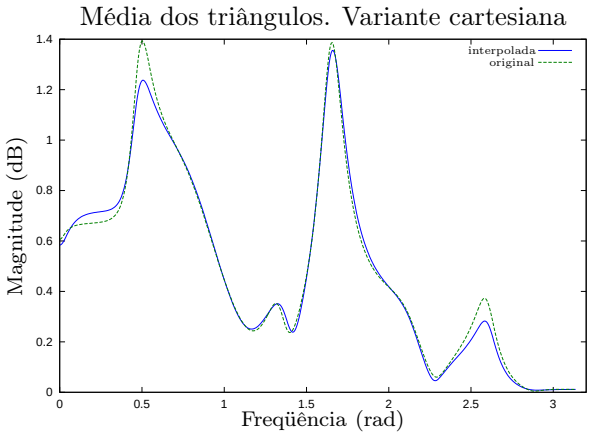



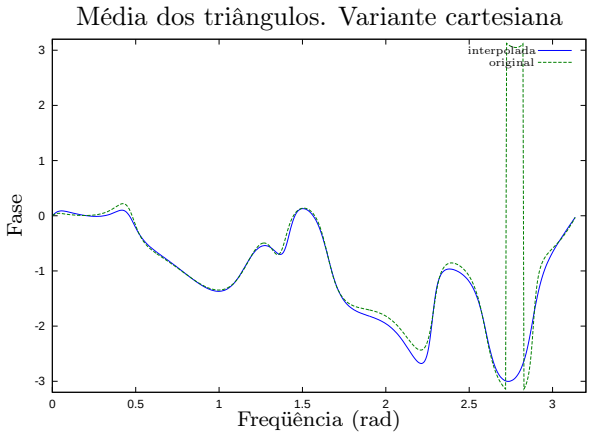



**Direção (30°, 163,12°)**

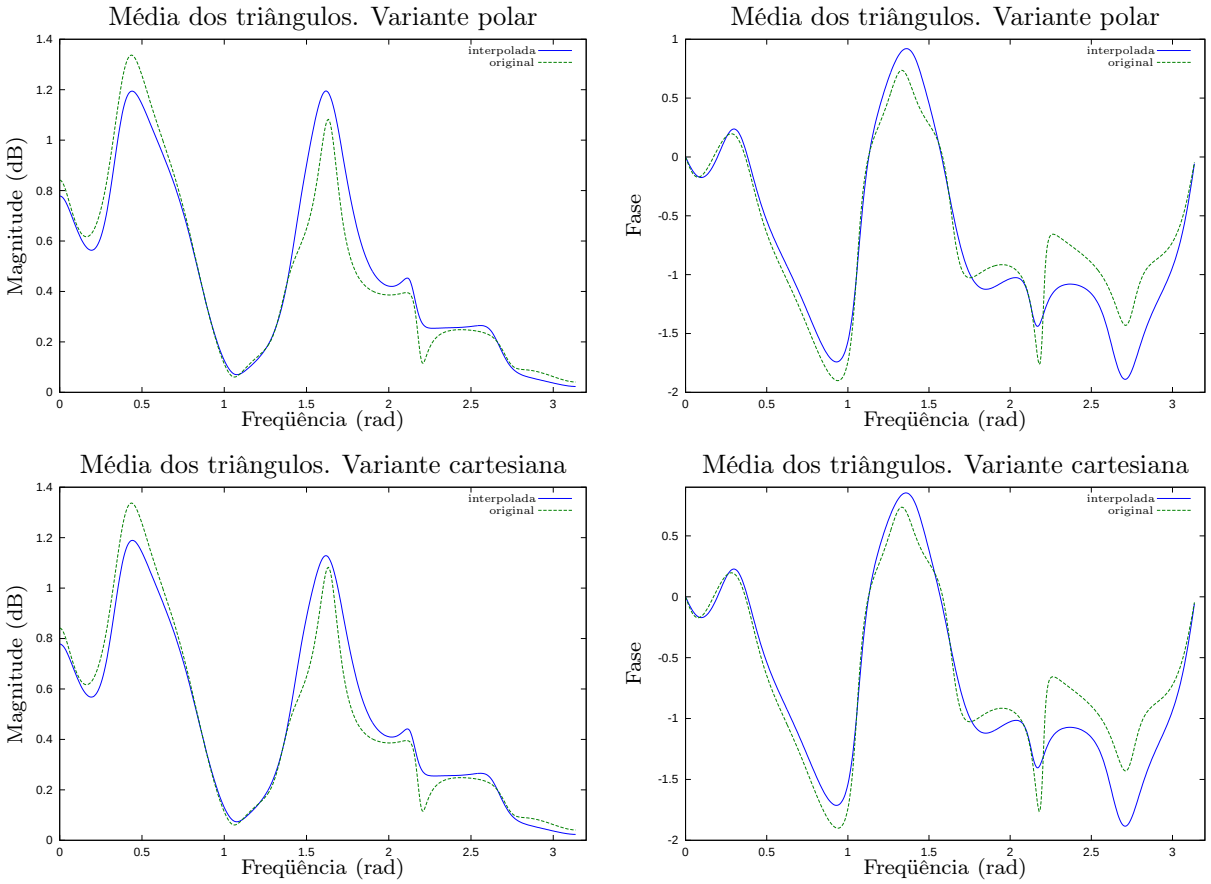


## Direção (40°, -5,625°)



# Referências Bibliográficas


[ADTA01] R. Algazi, R. O. Duda, D. M. Thompson, e C. Avendano. The CIPIC HRTF database. Em *WASSAP '01*. 2001 IEEE ASSP Workshop on Application of Signal Processing to Audio and Acoustics, 2001. 2, 3, 5

[AFSV05] T. Ajdler, C. Faller, L. Sbaiz, e M. Vetterli. Interpolation of Head Related Transfer Functions Considering Acoustics. Em *118th Convention of the Audio Engineering Society, J. Audio Eng. Soc.(Abstracts)*, volume 53, página 662, 2005. 43

[Beg94] D. R. Begault. *3-D sound for virtual reality and multimedia*. Academic Press Professional, Inc., San Diego, CA, USA, 1994. 20

[BKC92] B. Beliczynski, I. Kale, e G. Cain. Approximation of FIR by IIR digital filters: an algorithm based on balanced model reduction. *IEEE Transactions on Signal Processing*, 40:532–542, Março 1992. 28

[Bla97] J. Blauert. *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997. 3, 17

[CW99] C. Cheng e G. Wakefield. Spatial Frequency Response Surfaces (SFRS's): An Alternative Visualization and Interpolation Technique for Head-Related Transfer Functions (HRTF's). Em *Proceedings of the 16th Audio Engineering Society (AES) International Conference on Spatial Sound Reproduction*, páginas 147–159, 1999. 7, 36

[CW01] C. Cheng e G. Wakefield. Moving Sound Source Synthesis for Binaural Electroacoustic Music Using Interpolated Head-Related Transfer Functions (HRTFs). *Computer Music Journal*, 25(4):57–80, 2001. 7

[dSQ09] G. H. M. de Sousa e M. Queiroz. Two approaches for HRTF interpolation. Em *Anais do 12º Simpósio Brasileiro de Computação Musical*, 2009. 11, 21

[Eve00] F. A. Everest. *The master handbook of acoustics*. McGraw-Hill Professional, 4 edição, 2000. 3

[FBD02] F. P. Freeland, L. W. P. Biscainho, e P. S. R. Diniz. Efficient HRTF interpolation in 3D moving sound. Em *WASSAP '01*. AES 22nd International Conference: Virtual, Synthetic, and Entertainment Audio, 2002. 2

[FBD04] F. Freeland, L. Biscainho, e P. Diniz. Interpolation of head-related transfer functions (HRTFs): A multi-source approach. Em *Proc. of the EUSIPCO'2004-European Signal Processing Conference*, páginas 1761–1764, 2004. 2

[GHE03] D. W. Grantham, B. W. Y. Hornsby, e E. A. Erpenbeck. Auditory spatial resolution in horizontal, vertical, and diagonal planes. *The Journal of the Acoustical Society of America*, 114(2):1009–1022, 2003. 15, 17

[GM95] W. Gardner e K. Martin. HRTF measurements of a KEMAR. *Journal of the Acoustical Society of America*, 97(6):3907–3912, 1995. 5

[Gra86] D. W. Grantham. Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *The Journal of the Acoustical Society of America*, 79(6):1939–1949, 1986. 15, 16, 17

[HBBS88] E. R. Hafter, T. N. Buell, D. A. Basiji, e E. E. Shriberg. Discrimination of direction for complex sounds presented in the free-field. Em *Basic Issues in Hearing: Proceedings of the 8th International Symposium on Hearing, edited by H. Duifhuis, JW Horst, and HP Wit (Academic, London)*, páginas 394–401, 1988. 15, 16

[Huo99] J. Huopaniemi. *Virtual acoustics and 3-D sound in multimedia signal processing*. Tese de Doutorado, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, 1999. 6, 57

[Kal58] R. E. Kalman. Design of a self-optimizing control system. Em *Transactions of ASME*, volume 80, páginas 468–478, Fevereiro 1958. 28

[KC04] A. Kulkarni e H. S. Colburn. Infinite-impulse-response models of the head-related transfer function. *The Journal of the Acoustical Society of America*, 115(4):1714–1728, Abril 2004. 28

[LM94] R. Litovsky e N. Macmillan. Sound localization precision under conditions of the precedence effect: effects of azimuth and standard stimuli. *Journal of the Acoustical Society of America*, 96(2):752–758, 1994. 15, 16

[LXYS99] L. Li, L. Xie, W.-Y. Yan, e Y. C. Soh. Design of low-order linear-phase IIR filters via orthogonal projection. *IEEE Transactions on Signal Processing*, 47:448–457, Fevereiro 1999. 28

[MA84] M. Morimoto e H. Aokata. Localization cues of sound sources in the upper hemisphere. *J. Acoust. Soc. Jpn.(E)*, 5(3):165–173, 1984. 5

[MHVK97] J. Mackenzie, J. Huopaniemi, V. Välimäki, e I. Kale. Low-order modeling of head-related transfer functions using balanced model truncation. *IEEE Transactions on Signal Processing*, 4:39–41, Fevereiro 1997. 28

[MM07] R. Martin e K. McAnally. Interpolation of head-related transfer functions. 2007. 7

[Moo90] F. Moore. *Elements of computer music.* Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1990. 1

[OSB89] A. Oppenheim, R. Schafer, e J. Buck. *Discrete-time signal processing.* Prentice hall Englewood Cliffs, NJ, 1989. 13, 15

[PM89] D. R. Perrott e K. Marlborough. Minimum audible movement angle: Marking the end points of the path traveled by a moving sound source. *The Journal of the Acoustical Society of America*, 85(4):1773–1775, 1989. 15, 17

[PS90] D. Perrott e K. Saberi. Minimum audible angle thresholds for sources varying in both elevation and azimuth. *Journal of the Acoustical Society of America*, 87(4):1728–1731, 1990. 15, 17

[PT88] D. R. Perrott e J. Tucker. Minimum audible movement angle as a function of signal frequency and the velocity of the source. *The Journal of the Acoustical Society of America*, 83(4):1522–1527, 1988. 15, 17

[Pul01] V. Pulkki. *Spatial sound generation and perception by amplitude panning techniques.* Tese de Doutorado, Helsinki University of Technology, 2001. Espoo, Finland. 1, 10

[RBW95] P. Runkle, M. Blommer, e G. Wakefield. A comparison of head related transfer function interpolation methods. Em *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, 1995.*, páginas 88–91, 1995. 6

[Roa96] C. Roads. *The computer music tutorial.* MIT press, 1996. 1

[SA92] V. Sreeram e P. Agathoklis. Design of linear-phase IIR filters via impulse-response gramians. *IEEE Transactions on Signal Processing*, 40:389–394, Fevereiro 1992. 28

[SHLV99] L. Savioja, J. Huopaniemi, T. Lokki, e R. Väänänen. Creating interactive virtual acoustic environments. *J. Audio Eng. Soc.*, 47(9):675–705, 1999. 2, 6, 57

[Sor05] S. G. R. Soria. Estudos sobre personalização da função de transferência relativa à cabeça em sistemas biaurais de reprodução acústica virtual. Dissertação de Mestrado, Escola Politécnica da Universidade de São Paulo, 2005. 3, 4

[SP90] K. Saberi e D. R. Perrott. Minimum audible movement angles as a function of sound source trajectory. *The Journal of the Acoustical Society of America*, 88(6):2639–2644, 1990. 17

[Sto61] K. Stockhausen. *Two lectures*. T. Presser, 1961. 1

[Vor07] M. Vorländer. *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer, 2007. 3

[Xen92] I. Xenakis. *Formalized music*. Pendragon Press Stuyvesant, NY, 1992. 1