SERGIO GILBERTO RODRÍGUEZ SORIA

**ESTUDOS SOBRE PERSONALIZAÇÃO DA FUNÇÃO DE
TRANSFERÊNCIA RELATIVA À CABEÇA EM SISTEMAS
BIAURAIS DE REPRODUÇÃO ACÚSTICA VIRTUAL**

Dissertação apresentada à Escola
Politécnica da Universidade de São
Paulo para obtenção do Título de Mestre
em Engenharia

São Paulo

2005

SERGIO GILBERTO RODRÍGUEZ SORIA

**ESTUDOS SOBRE PERSONALIZAÇÃO DA FUNÇÃO DE
TRANSFERÊNCIA RELATIVA À CABEÇA EM SISTEMAS
BIAURAIS DE REPRODUÇÃO ACÚSTICA VIRTUAL**

Dissertação apresentada à Escola
Politécnica da Universidade de São
Paulo para obtenção do Título de Mestre
em Engenharia

Área de Concentração:
Sistemas Eletrônicos

Orientador:
Prof. Dr. Miguel Arjona Ramírez

São Paulo
2005

FICHA CATALOGRÁFICA

Rodríguez Soria, Sergio Gilberto
Estudos Sobre Personalização da Função de Transferência Relativa à
Cabeça em Sistemas Biaurais de Reprodução Acústica Virtual, São Paulo,
2005. 155p.

Dissertação (Mestrado) – Escola Politécnica da Universidade de São
Paulo, Departamento de Sistemas Eletrônicos.

1. Processamento de Sinais 2. Função de Transferência Relativa à cabeça (HRTF) 3. Acústica virtual 4.
Som espacial, 5. Áudio 3D.

A mi madre

AGRADECIMENTOS

Ao meu orientador, o Dr. Miguel Arjona Ramírez, pela orientação, estima e consideração recebida, tanto no acadêmico quanto no pessoal, ao longo destes anos.

À minha família, meu pai Gilberto, minha mãe Edith e meu irmão Dante pelo carinho, incentivo e apoio infinito.

Ao meu tio Edgardo Alfaro (*Eddie*), pela confiança e apoio incondicional, sempre.

Aos meus colegas do grupo de áudio do LSI, Régis e Celso, pelas valiosas sugestões na etapa final de correção do texto.

Aos meus amigos peruanos no Brasil, o Gino, o John, o José, e a todos com quem compartilhei estes anos de aprendizagem em todos os sentidos, longe de casa.

À Jacqueline por ter sido minha inesgotável fonte de inspiração.

Ao CNPq pelo apoio financeiro para a realização deste trabalho.

RESUMO

Este trabalho apresenta diversas propostas associadas ao uso ótimo de funções de transferência relativas à cabeça (*HRTFs*) em sistemas de reprodução acústica virtual por fones de ouvido. Estas propostas permitem personalizar a *HRTF* a indivíduos particulares, tomando como base uma combinação da modelagem estrutural e morfológica de *HRTFs*. Dentro do contexto da modelagem estrutural, o presente trabalho se concentrou no estudo da contribuição do *pinna* à *HRTF*. O *pinna* é a estrutura anatômica responsável pela percepção de elevação. Assim, o primeiro passo foi extrair um conjunto de funções de transferência relativas ao *pinna* (*PRTFs*) das *HRTFs* de uma base de dados. Para tanto, foram usadas diversas técnicas como análise preditiva linear para rastrear as ressonâncias, janelamento para eliminar a influência do torso, funções de autocorrelação e de atraso de grupo para salientar as antirressonâncias, e outros algoritmos para combinar ressonâncias e antirressonâncias em apenas uma magnitude espectral. Usando essa nova base de dados de *PRTFs* e parâmetros antropométricos propostos mais outros registrados na base de dados, um espaço vetorial correspondente à antropometria do *pinna* foi mapeado linearmente em um espaço vetorial correspondente às características espectrais da *PRTF*, calculando-se assim várias transformações lineares para estimação de novas *PRTFs* fora da base de dados. A estimação atingiu 66% de reconstrução no grupo de treino. O trabalho está orientado à exploração das características espectrais importantes na percepção de elevação, portanto, está limitado ao plano médio do hemisfério frontal, onde não existem diferenças interaurais significativas nem efeitos difrativos da cabeça. Finalmente é proposto um sistema de testes de localização de fonte sonora para validar o modelo.

ABSTRACT

This work presents several proposals associated with the optimal use of head-related transfer functions (*HRTF*) in virtual auditory spaces presented via headphones. These proposals lead to personalization of the *HRTF* to particular individuals, using a combination of the structural and morphological modeling techniques. In the context of structural modeling, this work focuses on modeling the contribution of the *pinna* to the *HRTF*. The *pinna* is the anatomical structure responsible for vertical sound localization. Thus, the first step was to extract a set of *pinna*-related transfer functions (*PRTFs*) from *HRTFs* published in a database. This was accomplished using several techniques like linear prediction analysis for tracking the resonances, windowing for eliminating the torso influence, autocorrelation and group delay functions for emphasizing the notches and other algorithms for combining resonances and notches in only one magnitude response. Using this novel database of *PRTFs* and a set of proposed anthropometric parameters plus some others registered in the database, a vector space corresponding to *pinna* anthropometry is linearly mapped into a vector space corresponding to spectral features of the *PRTF*, being calculated, in this way, several linear transformations for estimation of new *PRTFs*, outside the database. The estimation attains 66% of reconstruction in the training group. The work focuses on the exploration of spectral characteristics important for elevation perception, therefore, it is limited to the median plane where there are no meaningful interaural differences nor head diffraction effects. Finally, a system for sound localization tests is proposed in order to validate the model.

SUMÁRIO

RESUMO.....	1
ABSTRACT.....	3
SUMÁRIO	4
LISTA DE FIGURAS	8
LISTA DE TABELAS.....	10
LISTA DE ABREVIATURAS E SIGLAS	11

CAPÍTULO 1

INTRODUÇÃO	13
1.1 APRESENTAÇÃO DO PROBLEMA	14
1.2 REVISÃO DA BIBLIOGRAFIA	16
1.3 METAS E OBJETIVOS DO TRABALHO	18
1.4 SUMÁRIO DO TRABALHO	19
1.5 NOTAÇÕES.....	20

CAPÍTULO 2

FUNDAMENTOS DA AUDIÇÃO ESPACIAL	22
2.1 INTRODUÇÃO	22
2.2 SISTEMA DE COORDENADAS	24
2.3 LOCALIZAÇÃO DE FONTE SONORA.....	25
2.3.1 <i>Localização Horizontal</i>	27
2.3.2 <i>Localização Vertical e resolução de ambigüidades frente/trás</i>	29
2.3.3 <i>Movimentos da Cabeça</i>	32
2.3.4 <i>Informação Visual</i>	33
2.3.5 <i>Influência do recinto</i>	34
2.4 FUNÇÃO DE TRANSFERÊNCIA RELATIVA À CABEÇA <i>HRTF</i>	34

2.5	MEDIÇÃO DA <i>HRTF</i> E ASPECTOS RELATIVOS AO PROCESSAMENTO DE SINAIS	35
2.5.1	<i>Pré-processamento de HRTFs</i>	36
2.5.2	<i>Base de dados pública</i>	37
2.5.3	<i>Aproximações de Fase Mínima</i>	38
2.6	VISUALIZAÇÃO DOS DADOS: REPRESENTAÇÕES UNIDIMENSIONAIS E BIDIMENSIONAIS DE <i>HRTFs</i>	38
2.7	OTIMIZAÇÃO DE SISTEMAS BIAURAIS BASEADOS EM <i>HRTFs</i>	40
2.7.1	<i>Problemas dos sistemas baseados em HRTFs</i>	41
2.7.1.1	Problemas computacionais.....	41
2.7.1.2	Problemas de Qualidade de som	42
2.7.2	<i>Modelagem da HRTF</i>	42
2.7.2.1	Modelagem por Pólos e Zeros.....	43
2.7.2.2	Modelagem numérica	43
2.7.3	<i>Interpolação espacial de HRTFs</i>	45
2.8	APLICAÇÕES DO SOM 3D	45
2.8.1	<i>Estudo científico de localização sonora</i>	46
2.8.2	<i>Avaliação Clínica</i>	47
2.8.3	<i>Transferência de informação espacial para comando/controle</i>	47
2.8.4	<i>Entretenimento</i>	48
2.8.5	<i>Sistemas de Telecomunicações e Realidade Virtual</i>	49
2.9	SUMÁRIO	49

CAPÍTULO 3

SOLUÇÕES PROPOSTAS AO PROBLEMA DA PERSONALIZAÇÃO DA <i>HRTF</i>		51
3.1	INTRODUÇÃO	51
3.2	SOLUÇÃO POR MÉTODOS NUMÉRICOS.....	52
3.3	SOLUÇÃO POR ESCALONAMENTO EM FREQUÊNCIA.....	53
3.4	SOLUÇÃO POR COTEJAMENTO COM BASE DE DADOS.....	56
3.5	SOLUÇÃO POR ABORDAGEM MORFOLÓGICA	58
3.6	SOLUÇÃO POR MODELAGEM ESTRUTURAL.....	60
3.6.1	<i>Análise estrutural</i>	61

3.6.2	<i>Contribuição da cabeça e torso à HRTF</i>	62
3.6.3	<i>Contribuição do pinna à HRTF</i>	65
3.7	SUMÁRIO	70

CAPÍTULO 4

TÉCNICAS E ALGORITMOS DESENVOLVIDOS E APLICADOS NA ESTIMAÇÃO DA PRTF.....72

4.1	INTRODUÇÃO	72
4.2	DETALHES OPERACIONAIS	73
4.3	CLASSIFICAÇÃO EM PADRÕES DE REPRESENTAÇÕES BIDIMENSIONAIS DA <i>HRTF</i>	75
4.3.1	<i>Aproximação da PRTF através de processamento bidimensional</i>	79
4.3.2	<i>Método de reforço de contraste para observação de contornos de antirressonâncias em representações bidimensionais de HRTFs</i>	84
4.3.3	<i>Padrões encontrados no plano médio</i>	86
4.4	HIPÓTESES E SIMPLIFICAÇÕES ADOTADAS	88
4.5	SOLUÇÃO PROPOSTA A PARTIR DE BASE DE DADOS <i>CIPIC</i>	91
4.6	ESTÁGIO DE EXTRAÇÃO DA <i>PRTF</i>	93
4.6.1	<i>Rastreamento de ressonâncias da HRTF</i>	94
4.6.2	<i>Extração das frequências centrais de antirressonâncias causadas pelo pinna na HRTF</i> 95	
4.6.3	<i>Aproximação da magnitude espectral da PRTF</i>	99
4.6.4	<i>Resultados finais dos processos de classificação e extração</i>	102
4.7	ESTÁGIO DE MODELAGEM DA <i>PRTF</i>	105
4.8	ESTÁGIO DE MAPEAMENTO LINEAR	109
4.8.1	<i>Antropometria do pinna</i>	109
4.8.2	<i>Correlação entre dados antropométricos e espectrais do pinna</i>	111
4.8.2.1	Correlação Individual	111
4.8.2.2	Derivação de novos parâmetros	113
4.8.2.3	Regressão linear multivariável	115
4.8.3	<i>Algoritmo de movimentação de contornos de antirressonâncias</i>	118
4.9	ESTÁGIO DE ESTIMAÇÃO DE NOVAS <i>PRTFs</i>	120

4.10	AVALIAÇÃO DE DESEMPENHO	120
4.11	SUMÁRIO	121

CAPÍTULO 5

PROPOSTA DE TESTES AUDITIVOS DE LOCALIZAÇÃO DE FONTE SONORA123

5.1	INTRODUÇÃO	123
5.2	SISTEMAS DE REPRODUÇÃO ACÚSTICA VIRTUAL (VADs)	124
5.3	MODELO DE TRANSMISSÃO DE SOM NA REPRODUÇÃO BIAURAL: FONES DE OUVIDO	125
5.4	IMPLEMENTAÇÃO DO SISTEMA DE TESTES	128
5.4.1	<i>Software de espacialização.....</i>	<i>129</i>
5.4.2	<i>Reprodução de fontes virtuais: Fones de ouvido.....</i>	<i>131</i>
5.4.3	<i>Reprodução de fontes reais: Alto-falantes</i>	<i>134</i>
5.4.4	<i>Recinto de reprodução</i>	<i>135</i>
5.4.5	<i>Procedimento dos testes auditivos</i>	<i>136</i>
5.5	SUMÁRIO	138

CAPÍTULO 6

CONCLUSÕES E TRABALHOS FUTUROS139

6.1	SUMÁRIO DO TRABALHO	139
6.2	APRESENTAÇÃO DA CONTRIBUIÇÃO DE MESTRADO.....	141
6.3	CONCLUSÕES	143
6.4	PERSPECTIVAS PARA CONTINUIDADE DA PESQUISA	145

BIBLIOGRAFIA.....148

LISTA DE FIGURAS

Figura 1.1. Diferentes sistemas de reprodução de som 3D.....	14
Figura 2.1. Sistema interaural de coordenadas.	25
Figura 2.2. Esquema simples da origem dos indicadores interaurais.....	28
Figura 2.3. Esquema bidimensional do cone de confusão.....	30
Figura 2.4. Caminho direto (linha tracejada) e caminho refletido (linha contínua) do som até atingir o canal auditivo.....	31
Figura 2.5. Localização e <i>Blur</i> de Localização	33
Figura 2.6. Sistema de medição de <i>HRTFs</i> de alta qualidade.....	37
Figura 2.7. Métodos de visualização de <i>HRTFs</i>	40
Figura 3.1. Visualização espectral bidimensional de <i>HRTFs</i>	52
Figura 3.2. <i>DTFs</i> do plano médio	55
Figura 3.3. Medidas antropométricas do <i>pinna</i> registradas na base de dados <i>CIPIC</i> para 45 indivíduos.	57
Figura 3.4. Percentagem de explicação da variância com respeito ao número de componentes principais	59
Figura 3.5. <i>HRIR</i> direita correspondente ao azimute de 0°, elevação de 45°, indivíduo 010...62	
Figura 3.6. Composição das <i>HRTFs</i> do manequim <i>KEMAR</i>	63
Figura 3.7. Regiões definidas pelo cone de sombra do torso no modelo “ <i>snowman</i> ”.	64
Figura 3.8. Os seis modos normais identificados por Shaw.	67
Figura 3.9. Modos ressonantes do <i>pinna</i> em visualizações bidimensionais de <i>HRTFs</i> e seu correspondente modelo autorregressivo de 12 coeficientes.....	69
Figura 3.10. Antirressonâncias causadas pelo <i>pinna</i> mostradas em visualizações bidimensionais da <i>HRTF</i> e do espectro do sinal residual LP.....	70
Figura 4.1. Diagrama de fluxo do algoritmo de alinhamento de <i>HRIRs</i>	79
Figura 4.2. Aplicação do algoritmo de alinhamento nas <i>HRIRs</i>	79
Figura 4.3. Estágios de processamento da abordagem bidimensional de aproximação da <i>PRTF</i> a partir de análise preditiva linear, e “ênfase” temporal.....	81
Figura 4.4. Algoritmo para aproximação bidimensional da <i>PRTF</i>	82
Figura 4.5. Algumas <i>PRTFs</i> aproximadas a partir da abordagem bidimensional.....	83
Figura 4.6. Janelamento de <i>HRIR</i> com meia janela Kaiser para eliminar reflexão do torso	85

Figura 4.7. Alinhamento e janelamento de <i>HRIRs</i> em visualizações temporal e espectral.....	86
Figura 4.8. Padrões encontrados em representações bidimensionais de <i>HRTFs</i> do plano médio.....	87
Figura 4.9. Amostra dos 3 padrões encontrados no plano médio.	88
Figura 4.10. Esquema de posições espaciais modeladas no plano médio, e de eventuais posições ipsilaterais	91
Figura 4.11. Esquema geral da solução proposta ao problema de personalização da <i>HRTF</i> . ..	93
Figura 4.12. Algoritmo de extração da <i>PRTF</i> aproximada.....	98
Figura 4.13. Processo de extração de frequências centrais de antirressonâncias causadas pelo <i>pinna</i> na <i>HRTF</i>	99
Figura 4.14. Influência da classificação em padrões no processo de extração da <i>PRTF</i>	100
Figura 4.15. Ajuste de frequências centrais de antirressonâncias causadas pelo <i>pinna</i>	101
Figura 4.16. Aproximação da magnitude espectral da <i>PRTF</i> esquerda do indivíduo 044	101
Figura 4.17. <i>PRTFs</i> extraídas de 64 <i>HRTFs</i> correspondentes ao <i>pinna</i> esquerdo e direito de 32 indivíduos da Base de dados <i>CIPIC</i>	102
Figura 4.18. <i>PCA</i> dos 64 vetores <i>PRTF</i> extraídos no caso em estudo	108
Figura 4.19. Mapeamento antropométrico - espectral do <i>pinna</i>	110
Figura 4.20. Exemplo de novo parâmetro antropométrico do <i>pinna</i>	114
Figura 4.21. Algoritmo de movimentação de contornos de antirressonâncias	119
Figura 4.22. <i>PRTFs</i> originais e <i>PRTFs</i> estimadas e ajustadas	119
Figura 4.23. Avaliação de desempenho do método para o grupo de vetores de treino.	121
Figura 5.1. Transmissão anatômica do som desde o fone de ouvido, passando através do <i>pinna</i> , até o tímpano.....	126
Figura 5.2. Circuito acústico equivalente de Thévenin da transmissão anatômica do som desde fones de ouvido, através do <i>pinna</i> , até o tímpano.	126
Figura 5.3. Média e desvio padrão calculados em grupo de 40 <i>PTFs</i> medido em 14 fones de ouvido comerciais	129
Figura 5.4. Esquema geral do sistema de testes de localização de fonte sonora.....	130
Figura 5.5. Patch em PD para convolução binaural.	133
Figura 5.6. Estrutura fixa de alto-falantes e estrutura giratória de alto-falantes.	135

LISTA DE TABELAS

Tabela 2.1. Definições importantes em audição espacial.	24
Tabela 2.2. Regiões espaciais importantes no sistema de coordenadas interaural.....	25
Tabela 2.3. Principais trabalhos em projeto de filtros digitais para <i>HRTFs</i>	44
Tabela 4.1. Percentagem de ocorrência dos 3 padrões do plano médio	87
Tabela 4.2. Dados relevantes da classificação em padrões da base de dados <i>CIPIC</i>	103
Tabela 4.3. Dimensões do <i>pinna</i> registradas na base de dados <i>CIPIC</i>	110
Tabela 4.4. PCPs e seus melhores preditores antropométricos.....	112
Tabela 4.5 Índice de importância das dimensões antropométricas	113
Tabela 4.6. Novos parâmetros antropométricos.....	114
Tabela 4.7. Índice de importância dos novos parâmetros antropométricos.....	115
Tabela 4.8. PCPs e seus melhores preditores antropométricos considerando os novos parâmetros.	115
Tabela 4.9. FCA e seus melhores preditores antropométricos.....	115
Tabela 4.10. Coeficientes de correlação entre PCPs originais e PCPs estimados através dos mínimos quadrados.	117
Tabela 4.11. Coeficientes de correlação entre FCAs originais e FCAs estimados através dos mínimos quadrados.	117

LISTA DE ABREVIATURAS E SIGLAS

AUTOREGRESSIVE	AR
AUTOREGRESSIVE MOVING AVERAGE	ARMA
CENTER FOR IMAGE PROCESSING AND INTEGRATED COMPUTING	CIPIC
COMPONENTE PRINCIPAL	CP
DIRECTIONAL TRANSFER FUNCTION	DTF
ENTRADA SIMPLES E SAÍDA DUPLA	ESSD
FAST FOURIER TRANSFORM	FFT
FINITE IMPULSE RESPONSE	FIR
FREE-AIR EQUIVALENT COUPLING	FEC
FREQUENCIA CENTRAL DE ANTIRRESSONÂNCIA	FCA
HEAD AND TORSO	HAT
HEAD-RELATED IMPULSE RESPONSE	HRIR
HEAD-RELATED TRANSFER FUNCTION	HRTF
HEAD TRACKER	HT
HEADPHONE TRANSFER FUNCTION	PTF
HIGH FIDELITY	HI FI
INFINITE IMPULSE RESPONSE	IIR
INTERAURAL LEVEL DIFFERENCE	ILD
INTERAURAL TIME DIFFERENCE	ITD
INTER-SUBJECT SPECTRAL DIFFERENCE	ISSD
KARHUNEN LOÈVE EXPANSION	KLE
KNOWLES ELECTRONIC MANEQUIN FOR ACOUSTIC RESEARCH	KEMAR
LINEAR AND TIME INVARIANT	LTI
MAXIMUM LENGTH SEQUENCES	MLS
OPTIMAL SCALE FACTOR	OSF
PESO DO COMPONENTE PRINCIPAL	PCP
PRINCIPAL COMPONENT ANALYSIS	PCA
PRESSURE DIVISION RADIO	PDR
PINNA-RELATED IMPULSE RESPONSE	PRIR

<i>PINNA</i> -RELATED TRANSFER FUNCTION	PRTF
PURE DATA	PD
SINGLE DELAY AND ADD	SDA
SINGULAR VALUE DECOMPOSITION	SVD
VIRTUAL AUDITORY DISPLAY	VAD
VIRTUAL AUDITORY SPACE	VAS

Capítulo 1

INTRODUÇÃO

Bem como a visão a audição tem três dimensões. Em condições normais de audição o ser humano pode estimar a direção de procedência do som, isto é, a posição das fontes sonoras que se encontram ao seu redor. Essa habilidade e a análise dela são aproveitadas pela tecnologia para diversos fins. Um deles é produzir máquinas que possam escutar. Um outro é produzir sistemas de comunicação onde o aumento da imersão e o acréscimo de dimensões perceptivas comuniquem mais e melhor informação. Nesse cenário, a audição espacial vem-se convertendo num campo de pesquisa importante e muito frutífero nas últimas décadas. Porém, muitas questões ainda permanecem sem resposta.

Quando da invenção do som estéreo, o objetivo já era criar uma sensação de imersão sonora, tornando o ambiente aural mais envolvente. Bem como o estéreo, nos anos seguintes, apareceu uma grande diversidade de tecnologias relativas à imersão sonora. Pela sua importância comercial em aparelhos domésticos e cinema, atualmente as técnicas multicanal são as mais conhecidas; entre elas os sistemas *5.1*, *7.1* e *10.2* [1]. Contudo, existem também outras técnicas multicanal de melhor desempenho onde o objetivo não é apenas produzir aquela sensação envolvente no ouvinte, senão também posicionar fontes sonoras virtuais em pontos específicos daquele espaço virtual que se tenta produzir. Além disso, o “ponto doce”¹ é bem maior quando comparado aos sistemas comerciais, sendo que seu tamanho e a qualidade da reprodução dependem do número de canais usados. Exemplos dessas técnicas, são: *Ambisonics*, *Ambiophonics* e *Wave Field Synthesis* [1]. A idéia dessas tecnologias é reproduzir o som através de arranjos de alto-falantes, modificando temporal e espectralmente

¹ Do inglês *sweet spot*.

o sinal de cada canal, de tal modo que somados reconstruam as frentes de onda produzidas por uma ou mais fontes no espaço virtual. Sistemas de realidade virtual, cinema, e de reprodução de som doméstico as usam para reproduzir som tridimensional (3D).

Por outro lado, existe uma outra tecnologia onde o som 3D pode ser reproduzido de maneira mais personalizada e sobretudo mais econômica, através de fones de ouvido: a tecnologia biaural². Ela é baseada na função de transferência relativa à cabeça ou *HRTF*, e busca reproduzir nas entradas dos ouvidos os indicadores³ naturais que o sistema auditivo usa na localização de fontes sonoras. Neste caso, também é necessário modificar os sinais no tempo e na frequência a fim de reproduzirem esses sinais ingressantes aos ouvidos. A tecnologia biaural é usada em pesquisa básica sobre audição espacial, novos sistemas de comunicação multimídia, análise clínica, realidade virtual, monitoramento, entre outras aplicações.

Finalmente, quando sinais biaurais são reproduzidos através de dois alto-falantes, a tecnologia é chamada de transaural. Nesse caso, filtros para cancelamento de eco⁴ são necessários a fim de minimizar a influência do canal esquerdo no canal direito e vice-versa. Na Figura 1.1 é apresentada a diferença esquemática entre os três sistemas de reprodução de som 3D referidos.

Figura 1.1. Diferentes sistemas de reprodução de som 3D.

1.1 Apresentação do problema

² Do inglês *binaural*

³ Do inglês *cue*

⁴ Do inglês *crosstalk cancellation*

Dada a sua maior versatilidade e economia, e a grande quantidade de aplicações que abrange, o presente trabalho se concentra na tecnologia binaural como meio de reprodução de som 3D. A pesquisa nesse campo traz consigo um maior entendimento do mecanismo governante da audição espacial, o qual é importante no entendimento de nossas capacidades perceptivas, ainda bem desconhecidas.

A função de transferência relativa à cabeça, ou *HRTF* (*Head-Related Transfer Function*), descreve a transformação do sinal sonoro desde a fonte sonora até os tímpanos do ouvinte; é, portanto, um sistema de entrada simples e saída dupla (ESSD), e pode ser dividido em *HRTF* esquerda e *HRTF* direita. O sistema depende da frequência, da posição da fonte em relação ao ouvinte e da geometria corporal do ouvinte. Essa função fornece os indicadores de localização para o sistema auditivo identificar a posição aparente de uma ou mais fontes sonoras no espaço. Assim, o fundamento da tecnologia binaural é a filtragem de um sinal anecóico monoaural de som através da *HRTF* correspondente à direção espacial onde se pretende posicionar o som. No caso da aplicação precisar de mais realismo, podem-se aplicar módulos de reverberação ou de simulação acústica de salas nesse sinal espacializado, convertendo o sistema num sistema auralizador [2].

Neste cenário, é claro que todas as aplicações procuram por métodos rápidos, econômicos e adaptativos, sendo de vital importância a análise de *HRTFs* e o desenvolvimento de modelos e algoritmos que otimizem sistemas binaurais. Em geral, um sistema binaural de espacialização de som visa atingir dois paradigmas: eficiência na implementação e desempenho de localização aceitável. Inicialmente, a maior preocupação foi atingir eficiência, reduzindo o número de dados e o número de medições no processo de aquisição direta de *HRTFs*. Aquelas tarefas já apresentaram diversos resultados, mas nunca totalmente satisfatórios. Além disso, a potência computacional disponível atualmente permite pôr em segundo plano a eficiência. Por outro lado, o desempenho de localização do sistema mede o grau de coincidência entre a posição espacial que o sistema pretende simular e a posição espacial que o usuário percebe. O desempenho de localização de qualquer modelo ou algoritmo binaural será sempre avaliado subjetivamente por cada usuário do sistema. Dada a natureza intrinsecamente subjetiva dessa métrica e os importantes problemas de qualidade de som produzidos pelo uso de *HRTFs* genéricas, a busca de métodos de cálculo ou medição de *HRTF* individuais, ou métodos para personalizar a *HRTF* de maneira automática, é o ponto para o qual convergem as pesquisas atuais em áudio 3D. No entanto, o problema se apresenta com um grande grau de complexidade devido à complicada interação das ondas acústicas com a forma irregular do

corpo, especialmente do pavilhão auditivo ou *pinna*. Essa dificuldade é maior considerando a grande diferença inter-individual do *pinna*, e o fato de que uma pequena diferença dimensional pode ocasionar uma grande diferença na magnitude e na fase da *HRTF*. Neste cenário, é importante identificar na *HRTF* as características que variam entre indivíduos e que são realmente relevantes no processo de localização sonora. O problema a resolver então é implementar modelos que capturem essas características e que se adaptem às particularidades de cada usuário a partir de dados simples.

Normalmente, a obtenção das *HRTFs* individuais é através da medição direta. O processo de medição requer tempo, equipamento custoso e considerável experiência em medições acústicas [3]. Portanto, quando a aplicação requer tempo real ou tecnologia para consumo em massa, essa opção é inviável. Diversos métodos computacionais têm sido desenvolvidos a fim de personalizar a *HRTF* adaptando uma *HRTF* genérica ou um grupo de *HRTFs* a certas características individuais. A idéia é que essa adaptação seja realizada nessas características perceptualmente relevantes. No entanto, até hoje, nenhum método tem gerado resultados satisfatórios.

1.2 Revisão da bibliografia

O problema da personalização da *HRTF* foi tratado diretamente só recentemente. Apesar de as pesquisas em audição espacial terem começado nos primeiros anos do século XX [4], somente nos últimos 15 anos essas pesquisas vêm sendo aproveitadas no desenvolvimento de aplicações multimídia de espacialização de som [5]. No entanto, os princípios biofísicos da capacidade perceptiva humana não são ainda plenamente conhecidos. Embora persista essa incerteza, hoje já existem sistemas de reprodução acústica virtual (*VADs – Virtual Auditory Displays*) que podem sintetizar um ambiente imersivo aural de alta fidelidade (*HI FI – High Fidelity*) [6]. Um *VAD HI FI* deve cumprir três requerimentos básicos: espacialização ou externalização do som (realismo), sensação envolvente (imersão) e localização exata de fontes sonoras (informação útil). Dependendo da aplicação, alguns desses requisitos podem ter mais importância que os outros. No caso de sistemas binaurais, um *VAD HI FI* deve levar em conta o uso de *HRTFs* individuais [7], e a equalização individual dos transdutores usados na reprodução [8]. O problema associado à equalização individual dos transdutores pode ser

minimizado usando fones intra-auriculares⁵ de alta qualidade na hora da reprodução [6], enquanto que o problema do uso de HRTFs individuais encontra-se ainda aberto. Assim, dada a alta complexidade do processo de medição acústica de *HRTFs*, diversos métodos computacionais têm sido apresentados com diferentes graus de sucesso. Todos esses métodos tentam estimar a *HRTF* individual a partir de dados já existentes, sejam *HRTFs* genéricas, um conjunto de *HRTFs* de vários indivíduos (base de dados de *HRTFs*), parâmetros antropométricos do indivíduo, intervenção ativa do usuário ou até uma combinação de todas as anteriores. Esses métodos são chamados de métodos de personalização ou customização da *HRTF*.

Uma abordagem heurística para resolver o problema é a modelagem estrutural. A idéia é modelar cada estrutura corporal envolvida na conformação temporal e espectral da *HRTF*, usando como parâmetros de modelagem as dimensões corporais específicas de cada estrutura. Em 1998, Brown e Duda apresentam um modelo estrutural da *HRTF* baseado na observação de dados medidos experimentalmente e de certos padrões encontrados na resposta impulsiva relativa à cabeça (*HRIR – Head-related Impulse Response*) [9]. Esse modelo apresenta um mínimo grau de personalização e está conformado por módulos diferenciados para a cabeça, o torso e o *pinna*. Seu ponto fraco é a falta de abrangência dada a reduzida população de indivíduos a partir da qual os padrões foram observados: 3 indivíduos apenas.

Outra das tentativas importantes para personalizar a *HRTF* e relacioná-la com antropometria é aquela apresentada pelo Middlebrooks em 1999 [10] [11]. Este autor trabalhou com a hipótese de que a forma geral da magnitude espectral da *HRTF* é similar em todos os indivíduos, variando apenas a escala de frequência de acordo com o tamanho do indivíduo. Assim, seria possível escalonar em frequência uma *HRTF* para adaptá-la a qualquer indivíduo. Embora os testes psicoacústicos mostrassem que, na média, o método melhora o desempenho de *HRTFs* genéricas, o fator de escalonamento ótimo não é calculado a partir de dados antropométricos. Contudo, esse fator apresenta boa correlação com a antropometria (coeficientes de correlação próximos a 0,75 para dois parâmetros antropométricos específicos: raio da cabeça e altura total do *pinna*).

Em 2003 pesquisadores da Universidade de Maryland apresentam mais uma tentativa para personalizar a *HRTF* a partir de dados antropométricos [12], a partir da base de dados *CIPIC* (*Center for Image Processing and Integrated Computing*) com 1250 *HRTFs* em total e 27

⁵ Do inglês *in-ear phones*

dimensões antropométricas de 45 indivíduos [13]. Primeiro, são medidas as dimensões antropométricas de um novo indivíduo. Na sequência, o sistema busca na base de dados o indivíduo mais parecido ao novo indivíduo em relação às dimensões dos *pinnae*⁶. Finalmente, as *HRTFs* desse indivíduo são atribuídas ao novo indivíduo. No entanto, os testes psicoacústicos não revelaram um bom desempenho do método.

O método numérico é o método mais promissor. Ele de fato calcula a *HRTF* individual resolvendo a equação acústica de onda para a superfície do corpo. Apesar da excelente exatidão do método, essa abordagem apresenta sérios problemas: o elemento de superfície relativo à taxa de amostragem espacial do corpo deve ser muito pequeno a fim de salientar os detalhes, tornando o tempo de processamento muito longo, bem distante do tempo real. Além disso, outros detalhes do processo tornam o procedimento ainda mais demorado do que a medição acústica direta de *HRTFs* [14].

Outra das principais tentativas é a abordagem morfológica, apresentada pelos pesquisadores da Universidade de Sydney [15]. O método consiste em fazer um mapeamento linear entre dados que representam as *HRTFs* e dados que representam antropometria numa base de dados de 36 indivíduos. A partir das cifras apresentadas (coeficientes de correlação próximos a 0.9) o método parece ser muito bom, porém, ele não tem sido validado psicoacusticamente com novos indivíduos, pelo menos com resultados publicados.

Finalmente, existem mais duas tentativas. Uma delas é a aproximação geométrica do *pinna* através de sete polígonos refletivos, também carente de validação psicoacústica [16]. A outra é uma abordagem não automática, e consiste em o usuário definir psicoacusticamente certas mudanças espectrais nas *HRTF* usadas para espacialização do som [17] [18], ou em o usuário definir psicoacusticamente o fator ótimo de escalonamento em frequência [19].

1.3 Metas e objetivos do trabalho

O principal objetivo do trabalho é aprofundar estudos sobre a personalização automática da *HRTF* a partir de dados antropométricos, visando propor um método original para a resolução do problema. Neste cenário, nos concentraremos na exploração do problema mais complexo e dependente do indivíduo: a localização vertical. Dada a coerência na sua

⁶ Plural de *pinna*

concepção, nossa proposta se baseará na combinação das abordagens estrutural e morfológica. Contudo, podemos visualizar diferentes metas, as quais são listadas a seguir:

- Seguindo a abordagem estrutural e considerando que atualmente já existem modelos de relativo sucesso sobre a contribuição da cabeça e o torso à *HRTF*, visamos modelar a contribuição do *pinna*. Assim, a primeira tarefa é extrair das *HRTFs* da base de dados *CIPIC*, as funções de transferência relativas ao *pinna* ou *PRTFs* (*Pinna-related Transfer Functions*).
- Estudar a variação das *PRTFs* extraídas com as características antropométricas disponíveis dos indivíduos através de abordagens numéricas como análise em componentes principais e o cálculo de coeficientes de correlação.
- A partir dos resultados anteriores, propor um método capaz de sintetizar *PRTFs* de novos indivíduos a partir de seus próprios dados antropométricos.
- Introduzir e definir os métodos para avaliação de desempenho de localização espacial do modelo, isto é: os testes auditivos de localização de fonte sonora.

1.4 Sumário do trabalho

O trabalho está dividido em seis capítulos, sendo o presente capítulo – Introdução - o primeiro deles. O capítulo 2 está destinado a apresentar conceitos básicos da audição espacial e sua problemática. Serão desenvolvidos vários aspectos relativos à *HRTF*, tanto psicológico-experimentais quanto relativos ao processamento de sinais. Este capítulo ajudará a compreender a importância da personalização da *HRTF* e o porquê da sua dificuldade. Além disso, faremos uma revisão das possíveis aplicações envolvidas.

Dado que a personalização da *HRTF* é um problema que já foi abordado por diferentes autores, como foi brevemente tratado na seção 1.2, o capítulo 3 está destinado a apresentar um resumo das técnicas desenvolvidas na área, as quais se destacam dentro da literatura e servem como referência para o presente trabalho. Além disso, apresentaremos uma análise estrutural da *HRTF* e mostraremos o porquê da abordagem escolhida.

O capítulo 4 representa a descrição de todos os algoritmos aplicados e desenvolvidos na concepção do método de personalização automática da *HRTF*. Dado que o problema é complicado tivemos que adotar certas simplificações e hipóteses que tornassem viável o trabalho. Elas também serão detalhadas neste capítulo. .

Uma proposta de testes psicoacústicos de localização de fonte sonora, cujo objetivo é estimar estatisticamente o desempenho de localização de um sistema de espacialização de som é apresentada no capítulo 5. Abordaremos também uma descrição detalhada da reprodução binaural através de fones de ouvido.

Finalmente, o capítulo 6 é destinado às conclusões finais da pesquisa, detalhando a contribuição do Mestrado e o que é considerado mais importante neste trabalho: as perspectivas para a continuidade de pesquisas futuras.

1.5 Notações

Na presente dissertação, as seguintes notações foram usadas:

- Matrizes são denotadas com letras maiúsculas. Exemplo: D , W .
- Vetores são denotados com letras minúsculas e itálicas. Exemplo: d , w
- Escalares podem ser denotados com letras minúsculas, maiúsculas, itálicas ou não itálicas. Exemplo: d_1 , d_2 , P_s , f .
- Funções de transferência e respostas em frequência são denotadas com letras maiúsculas e itálicas. Exemplo: $H(e^{j\omega})$.
- Sinais no domínio do tempo são denotados com letras minúsculas e itálicas. Exemplo $h_1(n)$.
- Índices temporais ou de frequência entre parênteses são denotados com letras itálicas. Exemplo: $h_2(n)$

- Siglas ou abreviações são denotadas com letras maiúsculas. Exemplo: *HRTF*.
- Em geral, as siglas provêm da abreviação dos nomes em inglês. Por ser essa a notação encontrada na literatura e a usada mundialmente, desconsideramos traduzir as siglas para o português. Exemplo: “Função de transferência relativa ao *pinna*” em português seria abreviada como FTRP, porém, a sigla usada é *PRTF* (*Pinna-related transfer function*). Certas siglas que não são usadas mundialmente, e que foram propostas no curso do presente trabalho, foram abreviadas do português. Exemplo: “Frequência central de antirressonância” é FCA; ou “Peso do componente principal” é PCP. As siglas provenientes do inglês são escritas com letras itálicas, as provenientes do português não.
- Palavras ou frases em inglês carentes de tradução conhecida ou apenas por fins explicativos são escritas em inglês com letras itálicas. Exemplo: *Wave Field Synthesis*
- O sub-índice i é sempre usado para indivíduos e/ou *pinnae*.
- O sub-índice j é sempre usado para ângulos de azimute.
- O sub-índice n é usado para ângulos de elevação. Quando n for usado como índice é indicativo de amostra temporal.
- O sub-índice k é sempre usado para parâmetros antropométricos.
- O sub-índice r é sempre usado para componentes principais.

Capítulo 2

FUNDAMENTOS DA AUDIÇÃO ESPACIAL

2.1 Introdução

Audição espacial é a habilidade para perceber os sons de maneira tridimensional. Considera o conjunto de características associadas aos sinais acústicos na recepção pelo sistema auditivo humano, tendo por finalidade o de adquirir informação referente à posição, número, trajetória e distância das fontes sonoras que se encontram ao seu redor [4]. Nos últimos cem anos, têm surgido muitas teorias e trabalhos experimentais sobre o mecanismo governante [4] [20]. Em geral, as geometrias da cabeça, torso e *pinnae* são fundamentais no processo de audição espacial. Em particular, as diferenças de pressão sonora (*ILD* – *Interaural Level Difference*) e de tempo de chegada do som (*ITD* - *Interaural Time Difference*) entre ambos os ouvidos determinam o posicionamento horizontal ou de azimute. A distância está determinada também pelo *ILD* e muitas vezes pela reverberação produzida no local de reprodução [21]. O posicionamento vertical ou de elevação é determinado principalmente pelas mudanças espectrais produzidas no som, pelos *pinnae* e pela cabeça e torso (*HAT* – *Head and Torso*). A resolução de ambigüidades frente/trás para diferenças interaurais iguais no cone de confusão⁷ é resolvida também pelas mudanças espectrais causadas pelo *pinna*. Assim, uma vez que cada pessoa possui uma função de transferência própria dependente da geometria e tamanho do seu corpo, e aprendida provavelmente ao

⁷ Superfície de cone espacial onde a diferença em comprimento entre os caminhos até os dois ouvidos é constante. Ver Figura 2.3

longo do crescimento pelo sistema nervoso central, a função de localização espacial é individual [4].

Além da mecânica do fenômeno baseada nas transformações dos sinais através da periferia auditiva, existem estudos biofísicos que exploram o processo de localização auditiva no estágio de transdução mecano-elétrica através das células ciliadas⁸ [22]. A partir dessa perspectiva, é possível supor que a localização sonora não é um processo aprendido apenas uma vez. Essa hipótese seria confirmada por experimentos onde é provado que o processo pode ser aprendido várias vezes [23]. Nessa mesma corrente, a simples pergunta de como o sistema auditivo assume o processo de crescimento corporal e, portanto, a mudança dos indicadores espectrais produzidos pela anatomia em amadurecimento, já foi explorada em roedores [24]. Para exemplificar, os indicadores acústicos espaciais que resultariam se a cabeça de um ouvinte fosse maior que a normal, provavelmente seriam exagerados em relação aos indicadores naturais usados por esse ouvinte. Codificar a posição espacial das fontes sonoras com esses indicadores pode melhorar a resolução espacial auditiva percebida. Contudo, se deve tomar cuidado com esses esquemas de codificação não naturais, pois eles precisam de treinamento nos indivíduos e não podem ser muito diferentes dos esquemas naturais porque poderiam desvirtuar totalmente a percepção das fontes virtuais. Finalmente, experimentos sugerem que a resolução auditiva espacial é limitada em muitos casos por fatores centrais cognitivos, tais como memória e atenção. Eles vêm a ser mais importantes que a sensibilidade periférica aos indicadores acústicos espaciais [25].

No contexto da tecnologia de informação, a audição espacial apresenta vantagens a serem aproveitadas no processamento de sinais acústicos. Maior inteligibilidade da fala, o efeito *cocktail party* [26], informação sobre o ambiente ou sala [27], ótima recriação da cena sonora e imersão são algumas dessas vantagens [5]. Assim, um sistema de som 3D tenta reproduzir ou analisar as alterações sofridas pelo som desde sua emissão pela fonte até chegar a cada ouvido, seja num ambiente específico (auralização) ou em campo livre (espacialização).

A primeira parte do presente capítulo apresenta informação sobre o processo de localização espacial de fonte sonora e os indicadores periféricos que o sistema auditivo usa. A seguir, é discutida a *HRTF* e seus aspectos relevantes, bem como os principais processos de otimização de sistemas de som 3D baseados nessas funções. Por fim, algumas aplicações do estudo da audição espacial são descritas.

⁸ Do inglês *auditory hair cells*

Na Tabela 2.1 são apresentadas algumas definições importantes encontradas nas seções seguintes [4].

Tabela 2.1. Definições importantes em audição espacial.

Termo	Definição
Monoaural	Aqueles fatos ou atributos referentes a um ouvido
Biaural	Aqueles fatos ou atributos referentes aos dois ouvidos.
Evento sonoro	Emissão sonora real desde uma fonte sonora no espaço físico real, i.e. espaço sonoro.
Evento auditivo	Percepção de um som no espaço por parte do sistema auditivo. Ele é produzido no espaço “imaginário” do ouvinte, i.e. espaço auditivo.
Localização	Correspondência espacial entre o evento sonoro e o evento auditivo
<i>Blur</i> de localização	Diferença espacial entre o evento sonoro e o evento auditivo. O conceito também está relacionado inversamente à resolução espacial do sistema auditivo.
Indicadores de localização	Atributos acústicos ou não acústicos que o sistema auditivo utiliza para localizar a fonte sonora.
Cone de confusão	Superfície de cone espacial onde a diferença em comprimento entre os caminhos até as duas orelhas é constante. Ver Figura 2.3.
Contralateral	Relativo ao <i>pinna</i> mais afastado da fonte sonora
Ipsilateral	Relativo ao <i>pinna</i> mais próximo da fonte sonora

2.2 Sistema de coordenadas

Em geral, a audição espacial é definida através do sistema de coordenadas esférico com respeito à cabeça humana. Nesse sistema, chamado comumente de sistema de coordenadas interaural, o centro do segmento que une os dois canais auditivos é a origem. Em um indivíduo comum, essa origem fica normalmente localizada embaixo e atrás do centro de massa da cabeça do indivíduo. Assim, uma posição qualquer no espaço é definida por três escalares: o ângulo lateral ou de azimute θ , o ângulo vertical ou de elevação ϕ , e o raio ou distância r . A Figura 2.1 apresenta este conceito [13]. O ângulo de azimute varia desde -90° à esquerda da cabeça até 90° à direita da cabeça, e o ângulo de elevação varia desde -90° embaixo da cabeça até completar a volta nos 270° , também embaixo da cabeça. Na Tabela 2.2 definem-se as regiões espaciais importantes com respeito a esse sistema de coordenadas.

Figura 2.1. Sistema interaural de coordenadas com ângulo de elevação ϕ , ângulo de azimute θ e raio ou distância r .

Tabela 2.2. Regiões espaciais importantes no sistema de coordenadas interaural.

Região	Definição (com respeito à Figura 2.1)
Plano médio	Plano formado pelos eixos X_2 e X_3 .
Plano horizontal	Plano formado pelos eixos X_1 e X_2 .
Plano vertical lateral	Plano formado pelos eixos X_1 e X_3 .
Semi-hemisfério frontal inferior	Semi-hemisfério compreendido entre os ângulos de elevação de -90° e 0°
Semi-hemisfério frontal superior	Semi-hemisfério compreendido entre os ângulos de elevação de 0° e 90°
Semi-hemisfério posterior superior	Semi-hemisfério compreendido entre os ângulos de elevação de 90° e 180°
Semi-hemisfério posterior inferior	Semi-hemisfério compreendido entre os ângulos de elevação de 180° e 270°

2.3 Localização de Fonte Sonora

Como explicado na introdução, no processo de localização de fonte sonora estão envolvidos tanto fatores físicos (aqueles relacionados à mecânica dos sinais sonoros e à periferia auditiva até atingir o tímpano), quanto psicológicos, i.e. cognitivos. Segundo pesquisas recentes, os melhores indicadores de localização dependem de fatores externos [28]

(ambientais ou dependentes da fonte sonora). Porém, a decodificação desses indicadores é um processo individual, e depende em grande medida da qualidade de nossas antenas acústicas: os *pinnae*.

Recentemente [28], tem sido introduzida uma nova divisão dos indicadores de localização: indicadores dependentes da fonte sonora, indicadores dependentes do ambiente e indicadores dependentes do indivíduo. Os indicadores dependentes da fonte sonora são as características temporais, espectrais ou espaciais dos eventos sonoros. É diferente localizar um tom isolado do que um som complexo de maior banda passante; ou um evento sonoro na frente da pessoa do que um acima da cabeça ou no cone de confusão. Os indicadores dependentes do ambiente são as características do evento sonoro associadas ao local ou ambiente onde é produzido. Podem ser: som direto ou som refletido. É diferente localizar um som em campo livre, em uma sala de aula ou em uma sala de concertos. Finalmente, os indicadores dependentes do indivíduo são os processamentos que o indivíduo aplica ao evento sonoro para convertê-lo em evento auditivo. Estão subdivididos em indicadores binaurais - o *ILD* e o *ITD*; a *HRTF*, que descreve as mudanças espectrais do som desde a fonte até os dois ouvidos; os indicadores dinâmicos, que são os movimentos do corpo ou da cabeça que o ouvinte faz para confirmar a localização; e os indicadores não acústicos que incluem os fatores cognitivos, como o conhecimento ou familiaridade que o ouvinte tem com o evento sonoro, ou a informação visual que ele tem sobre o evento. No processo de localização todos esses indicadores são usados de forma combinada [28].

No cenário da modelagem, os modelos de fontes sonoras e de ambientes existem com diferente grau de complexidade e funcionalidade. No entanto, modelos que reproduzam e expliquem os indicadores dependentes do indivíduo são desenvolvidos com certas restrições [29]. Embora os processos de localização horizontal e vertical sejam complementares, para fins de pesquisa e de modelagem muitas vezes é necessário separá-los. A localização horizontal é bem modelada, como pode ser comprovado pelo fato de que todos os sistemas binaurais conseguem lateralizar o som com considerável grau de exatidão (espacialização 2D). No entanto, a localização vertical, e em menor grau a resolução das ambigüidades frente/trás, ainda ficam como problemas a serem resolvidos.

Neste ponto é importante salientar que todos os sistemas de espacialização binaural de som buscam modelar ou reproduzir o processo de localização de fonte sonora no aspecto mecânico, isto é, considerando a cadeia auditiva apenas até a chegada das ondas acústicas no tímpano. Sendo esse o paradigma, o objetivo é reproduzir fielmente os sinais no tímpano de

tal forma que o resto da cadeia auditiva se cumpra naturalmente. Na sequência consideraremos uma descrição sumária dos fenômenos de localização horizontal e localização vertical em separado, e os indicadores de localização relevantes que os governam até a chegada ao tímpano, i.e. desconsiderando qualquer fator cognitivo. Além disso, apresentamos uma breve resenha dos outros indicadores de localização que consideramos mais relevantes no presente trabalho: movimentos da cabeça, a visão e a influência do recinto.

2.3.1 Localização Horizontal

A teoria Duplex proposta por John Strutt (*Lord Rayleigh*) [20] há 100 anos, explica quase totalmente o fenômeno de localização horizontal. Os dois indicadores interaurais, o *ILD* e o *ITD*, se complementam para definir a suposta posição horizontal da fonte sonora. Na Figura 2.2 é mostrado um esquema simples para derivação do *ITD* [4]. Considerando uma frente de onda plana de uma fonte afastada que chega a uma cabeça esférica de raio r , a partir de uma direção especificada pelo ângulo horizontal θ em graus, existe uma diferença de comprimento nos caminhos até cada ouvido, definida por $(a + b)$ e dada pela Equação (2.1):

$$(a+b) = \left(\frac{\theta}{360} \right) \cdot 2\pi \cdot r + r \cdot \sin(\theta) \quad (2.1)$$

Essa diferença convertida em retardo de tempo é o *ITD* aproximado, dado pela Equação (2.2):

$$ITD = \frac{r}{c} \left(\frac{2\pi\theta}{360} + \sin(\theta) \right) \quad (2.2)$$

onde c é a velocidade do som no ar, 340m/s, e $-90^\circ \leq \theta \leq 90^\circ$. Quando θ tende a zero, o *ITD* é zero, quando θ tende a 90° , o *ITD* é maximizado. Supondo o raio da esfera r de 8,75 cm e θ igual a 90° , $(a + b)$ é 22,5 cm e o *ITD* associado é 0,656 ms. Este é um valor típico para o *ITD* quando a fonte está do lado direito da cabeça ($\theta = 90^\circ$).

Figura 2.2. Esquema simples da origem dos indicadores interaurais.

Lord Rayleigh também resolveu a equação de onda na superfície de uma esfera rígida para mostrar o efeito difrativo da cabeça. Isto leva ao cálculo do *ILD*. Em baixas frequências, quando o comprimento de onda é grande em relação ao diâmetro da cabeça, não existem diferenças significativas de pressão sonora. Em altas frequências, a diferença passa a ser até de 20 dB, sendo chamado este efeito de sombreamento da cabeça⁹. Por outro lado, o sistema auditivo é sensível ao *ITD* até frequências abaixo de 1,5 kHz, onde a fase do som incidente é determinada univocamente; em altas frequências esta fase é ambígua [29]. Assim se apresenta a complementaridade do *ILD* e do *ITD* na localização horizontal.

Existem diversas técnicas de estimação do *ITD* a partir de *HRTFs* medidas, tais como o cálculo das diferenças interaurais de fase, o método de correlação cruzada ou o cálculo do retardo do máximo impulso [29].

Segundo pesquisas psicoacústicas, a resolução espacial da capacidade localizadora humana no sentido horizontal depende da frequência e da direção, chegando até 1° como máxima resolução para sinais provindos da frente da cabeça (em posições próximas à posição determinada pelo ângulo de azimuth 0° e o ângulo de elevação 0°).

⁹ Do inglês *head shadow*

Os indicadores interaurais não são importantes apenas na localização horizontal. Duda [30] propõe que estes indicadores são fundamentais em nossa percepção da distância e na localização vertical. Inclusive no plano médio onde não existem indicadores interaurais significativos, é bem provável que essa diferença nula também seja usada pelo sistema auditivo para reconhecer sons provenientes desse setor [4]. Porém, como veremos na seção seguinte, tem sido amplamente demonstrado que os indicadores monoaurais, i.e. indicadores espectrais contidos na *HRTF*, são os principais agentes desse processo.

2.3.2 Localização Vertical e resolução de ambigüidades frente/trás

Os indicadores primários para localização vertical e para resolução de ambigüidades frente/trás são monoaurais. Estes dois processos são equivalentes quando neles não existem indicadores interaurais significativos, isto é, no plano médio e no cone de confusão. Já nesses casos é possível separar contundentemente os processos de localização horizontal e vertical. Portanto, com fins explicativos serão considerados principalmente esses casos. Na Figura 2.3 é apresentado o conceito de ambigüidades frente/trás no cone de confusão. Intuitivamente, é possível dizer que a assimetria do *pinna* em relação ao plano vertical lateral é fundamental nesse processo. Do mesmo jeito, na localização no plano médio, a forma irregular do *pinna* cumpre uma tarefa fundamental.

O *pinna* atua como um filtro acústico, modificando espectralmente os sinais dependendo da frequência e da direção de procedência. As cavidades ressonantes do *pinna* amplificam a magnitude espectral em algumas frequências, e por sua vez, a complexa geometria dele produz efeitos de interferência atenuando a magnitude em outras frequências. Para exemplificar: nos casos da Figura 2.4, uma onda sonora senoidal (um tom puro) percorre dois caminhos desde a fonte sonora até o canal auditivo, um caminho direto e um caminho refletido mais longo. Intermediária

Como é observado, o comprimento desse caminho depende do ângulo de elevação. Em moderadas baixas frequências ($3kHz \leq f \leq 6kHz$)¹⁰, os sinais dos dois caminhos atingem o conduto auditivo com a mesma fase; em frequências mais altas, o sinal refletido pode aparecer fora de fase com o sinal direto, ocorrendo interferência destrutiva. A maior interferência acontece quando a diferença de comprimento entre os dois caminhos d é meio comprimento

¹⁰ Em audição espacial as frequências até 5 kHz são consideradas frequências baixas, e as frequências até 10 kHz são consideradas frequências intermediárias [4].

de onda, tal que $f = c / 2d$. Nesse caso uma antirressonância se produz nessa frequência. Para valores típicos de d , a frequência da antirressonância está entre 6 kHz e 16 kHz, dependendo da elevação. Esse modelo foi chamado de modelo do retardo simples e soma (*SDA – Single Delay and Add*), e foi proposto por Hebrank em 1974 [31] a partir da teoria de reflexões de Batteau [32].

Figura 2.3. Esquema bidimensional do cone de confusão: superfície espacial em forma de cone onde todas as posições da fonte sonora apresentam um mesmo *ITD*.

Tem sido provado através de estudos psicoacústicos, comportamentais e neurofisiológicos que tanto ressonâncias quanto antirressonâncias proeminentes na magnitude espectral da *HRTF* são importantes na localização vertical e na resolução de ambigüidades frente/trás. Esses estudos foram realizados em seres humanos e animais. O primeiro passo foi provar que o *pinna* é fundamental nesses processos. Assim, experimentos de oclusão das cavidades do *pinna* foram apresentados em [33], provando que a oclusão progressiva das cavidades minimiza as habilidades de localização, até atingir certo treino com os “novos” *pinnae* [23]. Como complemento, já foi provado que o conteúdo espectral de altas frequências é necessário na localização ótima de fontes sonoras [31]. Hebrank inclusive encontrou três indicadores específicos: uma antirressonância entre 4 kHz e 16 kHz para reconhecer o semi-hemisfério

frontal superior, uma ressonância entre 7 kHz e 9 kHz para reconhecer sons procedentes de acima da cabeça, e uma ressonância entre 10 kHz e 12 kHz para reconhecer o semi-hemisfério posterior superior. O incremento da elevação no semi-hemisfério frontal superior é caracterizado pelo incremento da frequência de corte da antirressonância. Esse comportamento seria consistente com a teoria das reflexões de Batteau. Um outro experimento de Hebrank provou que indivíduos monoaurais podem ser facilmente treinados para localizar sons na vertical como se tivessem os dois ouvidos operativos. Isso prova que a localização vertical pode ser estritamente monoaural.

Figura 2.4. Caminho direto (linha tracejada) e caminho refletido (linha contínua) do som até atingir o canal auditivo. O comprimento dos caminhos varia com a elevação [32].

Experimentos comportamentais feitos em chinchilas [34] e morcegos [35] confirmam a importância do *pinna* na localização vertical e na resolução frente/trás dos mamíferos.

Outras pesquisas tentaram encontrar quais indicadores são mais importantes, antirressonâncias ou ressonâncias. Moore sugere que as ressonâncias são mais importantes, porém mudanças na frequência central das antirressonâncias são detectáveis ainda quando os

contornos das antirressonâncias são estreitos [36]. Por outro lado, Macpherson diz que um modelo baseado apenas na primeira antirressonância não é adequado [37]. Estudos feitos em gatos demonstram que as fibras do nervo auditivo simples são sensíveis à presença de antirressonâncias e, sobretudo à sua movimentação [38]. Isto sugere que a habilidade de localização é afinada quando há movimentação relativa entre a fonte sonora e a cabeça.

Na seção 3.6.3 será feita uma análise quantitativa de todas as ressonâncias e antirressonâncias produzidas pelo *pinna*, a partir dos resultados do presente estudo e dos trabalhos de Shaw [39] e de Lopez-Poveda [40]. Essa análise será de grande importância no desenvolvimento do método de personalização de *HRTFs* apresentado no Capítulo 4.

Por outro lado, estudos recentes mostram também que a reflexão do sinal sonoro no torso, pode ser usada como indicador de localização vertical para sinais de baixas frequências ($f \leq 3\text{kHz}$) [41]. A cabeça e o torso são as únicas estruturas que podem interagir com comprimentos de onda maiores, correspondentes a frequências abaixo dos 3 kHz, onde o *pinna* não exerce qualquer influência. A seção 3.6.3 apresenta mais informações.

Finalmente, segundo pesquisas psicoacústicas, a resolução espacial da capacidade localizadora humana no sentido vertical é bem menor do que no horizontal, atingindo no máximo 4° , quando o sinal é ruído branco. Porém, essa resolução é também dependente da frequência e de outros fatores cognitivos [4]. As diferenças quanto à resolução na localização horizontal e na localização vertical podem ser apreciadas na Figura 2.5. Aí é apresentado o conceito de *Blur* de localização, definido na Tabela 2.1.

2.3.3 Movimentos da Cabeça

Há quase 40 anos as teorias de movimento vêm sendo estudadas em audição espacial [42]. É bastante comum, como pode ser comprovado cotidianamente, que movimentos involuntários da cabeça sejam executados a fim de melhorar a localização de som [28]. Inclusive movimentos quase imperceptíveis de apenas alguns graus são suficientes para aumentar a exatidão da localização, especialmente em casos ambíguos. O *Blur* de localização é minimizado quando a cabeça é movimentada, já que o evento auditivo fica restrito intuitivamente a uma região onde a audição espacial é mais “fina”. Além dos movimentos involuntários ou reflexivos, movimentos totalmente conscientes para refinar a localização auditiva também se apresentam comumente.

Em [42] foi apresentada uma grande diversidade de experimentos sobre movimentos da cabeça. Esses movimentos foram divididos em rotativos, pivotantes e de inclinação. Movimentos rotativos, isto é ao redor do eixo X_3 da Figura 2.1, permitem resolver ambigüidades frente/trás. Movimentos pivotantes, isto é ao redor do eixo X_2 , revelam se o evento auditivo está no hemisfério superior ou inferior. Finalmente, movimentos de inclinação, ao redor do eixo X_1 , revelam informação sobre a elevação das fontes sonoras. Os resultados dos experimentos mostraram que movimentos rotativos, ou movimentos rotativos mais de inclinação, ou movimentos pivotantes são os mais freqüentes. Além disso, a maioria dos indivíduos move sua cabeça para frente e para trás a fim de determinar melhor a direção da fonte. Segundo estes estudos, os benefícios da movimentação da cabeça são muito menores na melhoria da localização vertical do que na resolução de ambigüidades frente/trás.

Figura 2.5. Localização e *Blur* de Localização. (a) No plano horizontal para 600 – 900 indivíduos de teste, pulsos de ruído branco de 100ms de duração. (b) No plano médio para sete indivíduos aplicando fala contínua de uma pessoa familiar. Figura modificada de [32].

2.3.4 Informação Visual

É evidente que a visão exerce influência poderosa no processo de localização de fonte sonora. Em muitos casos a informação visual predomina sobre os indicadores auditivos na localização. Por exemplo, ao assistir a um filme o som parece emanar da tela, ainda quando

ele provém realmente dos alto-falantes nas laterais. Esse efeito é chamado de “efeito ventríloquo” e consiste no estímulo auditivo parecer provir do mesmo ponto do estímulo visual.

Por outro lado, a informação visual também fornece uma grande melhoria da percepção de distância. No entanto, em alguns casos a informação visual pode produzir confusão, por exemplo, num auditório de concertos a congruência entre o espaço visual e o espaço auditivo é geralmente inconsistente devido à influência do recinto, i.e. reverberação local. Assim, enquanto a pessoa fecha os olhos, a música parece soar melhor, porque a visão não induz nenhum tipo de distorção. Da mesma maneira, é possível se concentrar melhor na música de olhos fechados. Finalmente, o “efeito reflexivo audiovisual” é aquele movimento involuntário da mirada na direção de onde o som provém. Esse reflexo é desencadeado pela percepção auditiva da primeira frente de onda.

2.3.5 Influência do recinto

Eventos auditivos podem ser produzidos em campo livre ou, na maioria dos casos, num local específico. Assim, a geometria e consistência desse local também exercem uma importante influência nas capacidades localizadoras. As reflexões iniciais, abaixo dos 100ms, são percebidas como reverberação. Essa reverberação traz consigo a sensação de “profundidade” do som, portanto, o ouvinte pode perceber melhor a distância das fontes sonoras ao seu redor [27]. Porém, em alguns casos como salas pequenas, as reflexões iniciais enfatizam erros de localização no plano médio ou no cone de confusão.

2.4 Função de Transferência relativa à cabeça *HRTF*

A *HRTF* caracteriza os indicadores de localização criados pelo corpo do ouvinte a partir do contato e dispersão das ondas acústicas nele. Ela descreve a transformação do sinal acústico desde um ponto arbitrário no espaço até os tímpanos do ouvinte. Normalmente, é obtida através de um trabalhoso processo de medição onde é requerido equipamento especializado.

O procedimento realizado pelos sistemas de espacialização binaurais é o seguinte: primeiro, deve-se obter um conjunto de *HRTFs*. O conjunto pode ser guardado como respostas

impulsivas (*HRIRs*), ou comprimido por meio de técnicas de modelagem ou de redução de dados. Geralmente a distribuição espacial do conjunto de *HRTFs* é dada pelo sistema de coordenadas esféricas, descrito na seção 2.2, onde a origem é o centro do segmento interaural. Portanto, a *HRTF* dependerá do ângulo de azimuth, do ângulo de elevação e da distância desde a fonte até a origem. Quando o sistema necessita sintetizar uma fonte sonora de tal modo que ela pareça provir de dada uma direção, o sinal sonoro é convoluído com as respostas impulsivas correspondentes àquela direção (uma para cada ouvido). O número e qualidade das respostas impulsivas do conjunto determinarão a resolução espacial e o desempenho de localização do sistema respectivamente.

Matematicamente, a *HRTF* relaciona o sinal de pressão sonora invariante no tempo, $x(n)$, da fonte sonora em campo livre, com o sinal de pressão sonora $x_E(n)$ e $x_D(n)$ nos tímpanos esquerdo e direito do ouvinte. A convolução do sinal da fonte sonora com as respostas impulsivas relativas à cabeça $h_E(n)$ e $h_D(n)$ correspondentes ao ouvido esquerdo e direito, se converte no domínio da frequência no produto das transformadas de Fourier, conforme Equação (2.3):

$$\begin{aligned} X_E(e^{j\omega}) &= H_E(e^{j\omega}) \cdot X(e^{j\omega}) \\ X_D(e^{j\omega}) &= H_D(e^{j\omega}) \cdot X(e^{j\omega}) \end{aligned} \tag{2.3}$$

Assim, as *HRTFs* esquerda e direita podem ser expressas respectivamente como $H_E(e^{j\omega}, \phi, \theta, r)$ e $H_D(e^{j\omega}, \phi, \theta, r)$ onde θ é o ângulo de azimuth, ϕ é o ângulo de elevação e r é a distância.

2.5 Medição da *HRTF* e aspectos relativos ao processamento de sinais

Empiricamente, a *HRTF* é medida usando microfones de teste pequenos, posicionados nos canais auditivos de um indivíduo de teste ou de uma cabeça/torso artificial ¹¹. Tem sido mostrado que a transmissão dentro do canal auditivo é independente da direção, ao menos

¹¹ Do inglês *dummy head*

desde a entrada do canal. Portanto, o ponto de medição dentro deste é secundário [4]. Blauert sugere o tímpano como ponto de medição, dado que ele contém toda a informação espacial. Porém, o ponto na entrada do canal auditivo fechado ou aberto é a posição mais atrativa dada sua maior facilidade de acesso. Geralmente é construído um sistema de suporte para alto-falantes. Estes devem ser posicionados numa distribuição esférica, de tal modo que se possa variar ou girar sua posição de emissão com respeito à elevação e ao azimuth. Assim, o indivíduo de teste ou cabeça/torso artificial deve ser cuidadosamente posicionado para seu centro interaural coincidir com o centro de rotação dos alto-falantes, i.e. a origem do sistema. Os alto-falantes emitem um sinal de grande banda passante, como por exemplo, códigos Golay, seqüências de comprimento máximo (*MLS – Maximum Length Sequences*), entre outros. O ótimo é realizar a medição numa sala anecóica, visando minimizar a energia refletida. A resposta impulsiva binaural é então obtida mediante a resposta de pressão dos microfones e gravada digitalmente em estéreo. Ambos os ouvidos devem ser medidos simultaneamente. A mesma operação de medição é realizada em cada posição espacial desejada. A resposta de cada ouvido é formatada e carregada num filtro *FIR (Finite Impulse Response)* para a modificação espectral do respectivo ouvido. A fonte sonora a ser espacializada é então filtrada através desses filtros *FIR* durante a reprodução, realizada através de fones de ouvido ou alto-falantes. Na Figura 2.6 é apresentado o sistema de medição de *HRTFs* descrito em [3].

2.5.1 Pré-processamento de *HRTFs*

A transmissão do som numa medição de *HRTFs* inclui as características de muitos subsistemas, os quais devem ser compensados para atingir a resposta desejada. As funções de transferência dos alto-falantes emissores do som, do microfone, e do canal auditivo (se a posição de medição estiver no canal auditivo aberto) têm que ser equalizadas apropriadamente. Se for desejado um conjunto de *HRTFs* ainda mais geral, devem-se aplicar outras técnicas de equalização, tais como equalização com respeito a uma direção determinada, ou equalização em campo difuso. Em geral, as técnicas de pré-processamento de *HRTFs* podem ser divididas em pré-processamento temporal (reconstruções de fase mínima, implementação do *ITD*, etc.) e pré-processamento espectral (suavização, ponderação, deformação, etc.) [29].

Figura 2.6. Sistema de medição de *HRTFs* de alta qualidade apresentado em [3].

2.5.2 Base de dados pública

Existem diversas bases de dados públicas de *HRTFs*. Podemos citar a conhecida base de dados de *HRTFs* medida na cabeça/torso artificial *KEMAR* (*Knowles Electronic Mannequin for Acoustic Research*) do MIT [43]; ela já foi usada em diversas pesquisas e sistemas de espacialização de som. Uma outra base de dados pública conhecida é aquela publicada pelo IRCAM da França [44]. Por outro lado, existem outras bases de dados não públicas, como aquela da Universidade de Helsinki na Finlândia [3].

Neste trabalho usaremos a Base de dados pública do Laboratório de Interface do *CIPIC* da Universidade de Califórnia – Davis [13]. Ela apresenta um conjunto de *HRTFs* de grande resolução espacial, com medições de 45 indivíduos em 25 ângulos de azimutes e 50 ângulos de elevação (1250 direções) com incrementos angulares de 5° aproximadamente. A distância r foi fixada em 1m. Além disso, a base de dados contém vinte e sete dimensões antropométricas de cada indivíduo, dezessete da cabeça e torso e dez do *pinna*. O azimute foi amostrado desde -80° até 80° , e a elevação desde -45° até $230,625^\circ$. Foram usados códigos Golay como sinal de excitação. A escolha

da base de dados se deve aos seguintes fatores: boa resolução espacial, existência de medidas antropométricas e documentação.

2.5.3 Aproximações de Fase Mínima

Geralmente, as *HRTFs* são aproximadas como sistemas de fase mínima já que as *HRIRs* contêm a maior parte de energia nos primeiros instantes, i.e. no impulso correspondente ao som direto [45]. Analiticamente quer dizer que o sistema contém todos os pólos e zeros no interior do círculo unitário. Como é sabido, os sistemas de fase mínima relacionam a fase espectral e o *log* da magnitude espectral do sistema através da transformada de *Hilbert*, assim, é possível trabalhar apenas com um destes elementos e depois derivar o outro.

Tem sido provado experimentalmente que a aproximação de fase mínima da *HRTF* não produz diferenças perceptuais relevantes no desempenho de localização em relação à *HRTF* medida [45]. Assim, obtemos uma importante simplificação: trabalhar apenas com a magnitude espectral da *HRTF*, aproximando o retardo *ITD* como um filtro passa - tudo (um retardo de fase) para atuar em cascata com a *HRTF* de fase mínima, na hora da reconstrução.

Sendo a resposta em frequência da *HRTF* o produto da sua magnitude e fase espectrais, segundo Equação (2.4):

$$H(e^{j\omega}) = |H(e^{j\omega})| \cdot e^{j \cdot \arg\{H(e^{j\omega})\}} \quad (2.4)$$

ela pode ser expressa por meio da Equação (2.5),

$$H(e^{j\omega}) = |H(e^{j\omega})| \cdot e^{j \left(\arg\{H_{pt}(e^{j\omega})\} + \arg\{H_{fm}(e^{j\omega})\} \right)} \quad (2.5)$$

onde $H_{pt}(e^{j\omega})$ é o filtro passa - tudo correspondente ao *ITD*, e $H_{fm}(e^{j\omega})$ é o filtro de fase mínima.

2.6 Visualização dos dados: Representações unidimensionais e bidimensionais de *HRTFs*

A visualização dos dados é importante no descobrimento das características macroscópicas que são perceptualmente relevantes na *HRTF*, como por exemplo, ressonâncias ou antirressonâncias na magnitude espectral. Além disso, uma boa visualização serve para comparar técnicas de modelagem, respostas de indivíduos diferentes, etc.

Geralmente os métodos de visualização expõem a *HRTF* em duas ou três dimensões dado seu caráter multidimensional (azimute, elevação, frequência, magnitude, distância, etc.). Neste trabalho usaremos principalmente dois métodos de visualização, que serão chamados de forma simplificada: visualização espectral e visualização temporal. Esses métodos em particular caracterizam a variação da magnitude espectral das *HRTFs* de um mesmo indivíduo com a elevação. Além desses métodos, usaremos também o método unidimensional clássico para visualização de sinais (amplitude vs. tempo ou amplitude vs. frequência).

A visualização espectral consiste no seguinte: para cada *pinna* de cada indivíduo, uma imagem artificial bidimensional correspondente a um ângulo de azimute, será apresentada. Nela, cada coluna será a magnitude espectral em dB da *HRTF* correspondente a uma elevação dada, e o valor dessa magnitude será expresso em níveis de cinza. Portanto, se tivermos uma resposta em frequência de 100 pontos (até a metade da frequência de amostragem), a imagem seria de 100 x 50 pontos, considerando 50 ângulos de elevação. A visualização temporal segue a mesma filosofia trocando o eixo de frequência pelo eixo temporal, e os valores em dB por valores normalizados da amplitude do sinal [13].

A escolha dos métodos de visualização está baseada nas seguintes razões: através desses métodos é possível identificar características estruturais; nosso interesse é modelar o efeito de elevação; trabalharemos com a magnitude espectral utilizando reconstruções de fase mínima quando necessário; e faremos algumas explorações com técnicas de processamento de imagens aplicadas nesse tipo de imagens. A Figura 2.7 apresenta um exemplo dos dois métodos de visualização de *HRTFs*. As *HRTFs* apresentadas correspondem ao plano médio, indivíduo 10 da base de dados *CIPIC*. As características estruturais estão marcadas, elas serão tratadas detalhadamente nas seguintes seções.

Figura 2.7. Métodos de visualização de *HRTFs*. (a) Visualização espectral bidimensional, azimute 0° , indivíduo 10. (b) Visualização temporal bidimensional correspondente ao azimute de 0° , indivíduo 10. As características estruturais estão marcadas.

2.7 Otimização de sistemas biaurais baseados em *HRTFs*

A maior limitação da tecnologia biaural é que todos os ouvintes são diferentes. Sinais biaurais gravados no indivíduo A podem, e de fato não funcionam bem no indivíduo B. Porém, existem poucos sistemas biaurais personalizados para um ouvinte qualquer (com *HRTFs* personalizadas). Ao invés disso, é usada uma “cabeça de referência” a qual possa representar um ouvinte típico na média. Isto é chamado de sistema não-individualizado¹² (com *HRTFs* genéricas) [46]. Assim, o uso de sistemas não individualizados apresenta sérios problemas de qualidade de som; o grande desafio em termos de tecnologia biaural é desenvolver métodos práticos pelos quais sinais biaurais possam ser personalizados para ouvintes específicos ou *HRTFs* individuais calculadas de maneira eficiente. Além disso, em relação ao primeiro paradigma de um sistema de espacialização de som (eficiência na implementação), existem mais duas tarefas para atingir otimização: modelagem e interpolação. Nesta seção faremos uma breve descrição dessas duas tarefas, antecedida por um breve sumário dos problemas já expostos.

¹² Do inglês *non-individualized*

2.7.1 Problemas dos sistemas baseados em *HRTFs*

Embora a teoria do uso da *HRTF* para espacializar som seja simples, existem muitos problemas na prática. Estes podem ser divididos em problemas computacionais e problemas de qualidade de som. Os primeiros podem ser resolvidos através de técnicas eficientes de modelagem e interpolação espacial, quanto aos segundos, eles precisam da personalização da *HRTF* ou da obtenção da *HRTF* individual.

2.7.1.1 Problemas computacionais

A seguir enumeramos os principais problemas computacionais apresentados em sistemas binaurais baseados em *HRTFs*:

- O sistema precisa do armazenamento de muitos dados. Considerando na média um sistema com 1250 posições espaciais, consistindo cada posição numa resposta impulsiva de 512 pontos, e sendo cada ponto gravado em formato fixo de 16 bits, o sistema precisaria $16 \times 512 \times 1250 = 10\,240.000,00$ bits. Isto representa mais de 10 MB de memória, o qual é impensável, por exemplo, num terminal celular que atinge no máximo uma memória de 1 MB.
- O sistema precisa de grande potência computacional para o processamento (convolução) de muitos dados em tempo real. Portanto, geralmente é requerido hardware dedicado. Quando o sistema sintetiza sons em movimento, o processamento é maior, e ainda mais quando o sistema leva em conta o “seguimento da cabeça” através de rastreadores de movimento.
- O processo de medição direta de *HRTFs* é muito custoso, tanto temporalmente quanto economicamente. Além disso, considerável experiência em medições acústicas é requerida. Por exemplo, um sistema básico precisaria no mínimo: um computador potente (atualmente pelo menos um Pentium III); uma placa de som externa multicanal (pelo menos 4 canais); caixas acústicas com alto-falantes (de

boa qualidade e pelo menos duas); microfones miniatura de alta qualidade; uma estrutura de suporte para as caixas acústicas e alto-falantes; provavelmente um sistema de rotação das caixas que pode ser controlado por um motor de passos; e, finalmente, tratamento acústico do local quando não for possível fazer a medição numa câmara anecóica.

2.7.1.2 Problemas de Qualidade de som

A seguir enumeramos os principais problemas de qualidade de som apresentados em sistemas biaurais baseados em *HRTFs*, que já foram mencionados nas seções anteriores:

- O sistema não espacializa o som, isto é, apresenta localização “dentro da cabeça”. Em 1996, Hartmann comparou estímulos reais reproduzidos através de alto-falantes numa câmara anecóica e estímulos virtuais reproduzidos através de fones de ouvido abertos, isto é, fones separados do *pinna*. Quando ele conseguiu que os estímulos virtuais não fossem diferenciados dos estímulos reais, o erro de localização ficou dentro dos limites da percepção. Porém, para atingir esse objetivo em termos de externalização foi imprescindível usar *HRTFs* individuais [47].
- Ambigüidade frente/trás. Esta ambigüidade se apresenta fortemente no cone de confusão, onde as diferenças interaurais são iguais para as fontes na superfície do cone. Neste caso, o uso de *HRTFs* individuais é necessário [6].
- Fraca espacialização vertical. O sistema funciona bem apenas para ângulos de elevação nulos, ou seja, no plano horizontal. Esse comportamento é comum em sistemas que usam *HRTFs* genéricas.

2.7.2 Modelagem da *HRTF*

Em pesquisas de audição espacial, tem-se tentado criar representações funcionais da *HRTF*. As tentativas de modelagem da *HRTF* pretendem descrevê-la em termos de

determinadas características: detalhes perceptualmente relevantes, a posição espacial, parâmetros antropométricos, padrões numéricos ou parâmetros espectrais. Cada uma dessas tentativas apresenta diversas vantagens e dificuldades de implementação. Dado que as tentativas de representação da *HRTF* em termos dos parâmetros antropométricos estão vinculadas aos processos de personalização, nesta seção serão descritas apenas as tentativas de modelagem em termos de parâmetros espectrais (por pólos e zeros) e de padrões numéricos.

2.7.2.1 Modelagem por Pólos e Zeros

As *HRTFs* são as saídas de sistemas lineares invariantes no tempo (*LTI – Linear and Time Invariant*). Assim, em teoria, as *HRIRs*, quando processadas adequadamente, podem serem representadas diretamente por filtros não recursivos *FIR*. Devido aos problemas computacionais relatados anteriormente, as aproximações da *HRTF* são muito comuns. Isto pode ser feito através de técnicas convencionais de projeto de filtros digitais. Porém, é necessário tomar cuidado em relação a certos aspectos. O principal problema a ser resolvido é: quais aspectos da *HRTF* devem ser conservados? Como explicado na seção 2.3.1 o *ITD* é importante embaixo dos 1,5 kHz, e o *ILD* é fundamental em frequências maiores. Além disso, indicadores espectrais de altas frequências são importantes na percepção da elevação, na externalização e na resolução de ambigüidades frente/trás. Assim, não fica claro como combinar todos esses elementos e qual é o grau de exatidão que cada um deles necessita ao se modelar a *HRTF* a fim de preservar os indicadores de localização importantes. A modelagem por pólos e zeros guarda estreita relação com essa implementação de filtros digitais para simulação binaural, porém ela não leva em conta os aspectos importantes na localização espacial. Em geral, o objetivo principal é ajustar o modelo à envoltória espectral e reduzir o quanto for possível o erro de aproximação (medido através de uma norma matemática) e a ordem do modelo. Na Tabela 2.3, são apresentados os principais trabalhos de modelagem por pólos e zeros. A maioria não tem validação psicoacústica.

2.7.2.2 Modelagem numérica

A modelagem numérica resolve o problema de estimação da *HRTF* através de métodos numéricos. A abordagem numérica mais promissora é a resolução exata da equação acústica de onda para a superfície do corpo. A fim de atingir esse objetivo, são usados métodos como elementos de contorno, elementos finitos ou re-expansão de múltiplos pólos. Estes métodos vêm sendo usados também em acústica de salas para a resolução da equação acústica e cálculo da resposta impulsiva do recinto. Já que essa é uma das principais abordagens para calcular a *HRTF* individual, ela será tratada com mais detalhe no seguinte capítulo. No entanto, existem outras abordagens numéricas. Um método usado ao longo da pesquisa tem sido a análise em componentes principais (*PCA – Principal Component Analysis*). Em geral, a *PCA* tem sido aplicada ao *log* da magnitude espectral da *HRTF*. Martens [6] foi quem a usou pela primeira vez para modelar a *HRTF*, concentrando-se na variação da *HRTF* com o azimuth. Kistler e Wightman [45] confirmaram os resultados de Martens e aplicaram *PCA* em *HRTFs* de distintas elevações. A expansão de Karhunen-Loève (*KLE – Karhunen Loève Expansion*), cuja formulação matemática é a mesma que aquela da *PCA*, mas usando todos os componentes, foi usada por Chen [48], quem apresentou resultados diferentes aplicando a *KLE* no valor complexo da *HRTF*. A decomposição em valores singulares (*SVD – Singular Value Decomposition*) também foi usada por Abel para modelar a *HRTF* [49].

Tabela 2.3. Principais trabalhos em projeto de filtros digitais para *HRTFs*.

Autores	Filtro digital	Ordem
Begault, 1991	Biaural/ <i>FIR</i>	82-512
Sandvad e Hammershoi, 1994-1	Biaural/ <i>FIR</i>	72
Kulkarni e Colburn, 1995, 1997	Biaural/ <i>FIR</i>	64
Hartung e Raab, 1996	Biaural/ <i>FIR</i>	48
Asano, 1990	Biaural/ <i>FIR</i>	>40
Sandvad e Hammershoi, 1994-2	Biaural/ <i>IIR</i>	48
Blommer e Wakefield, 1994	Biaural/ <i>IIR</i>	14
Jot, 1995	Biaural/ <i>IIR</i>	10-20
Ryan e Furlong, 1995	Biaural/ <i>IIR</i>	24
Kulkarni e Colburn, 1995, 1997	Biaural/ <i>IIR</i>	6
Kulkarni e Colburn, 1995, 1997	Biaural/ <i>IIR</i>	25 (autorregressivo)
Hartung e Raab, 1996	Biaural/ <i>IIR</i>	34/10
Mackenzie, 1996	Biaural/ <i>IIR</i>	10
Blommer e Wakefield, 1997	Biaural/ <i>IIR</i>	40

Haneda, 1999	Biaural/IIR	60
Liu, 2001	Biaural/IIR	24

2.7.3 Interpolação espacial de *HRTFs*

Uma outra otimização importante é a interpolação espacial de *HRTFs*, isto é, a síntese de *HRTFs* correspondentes a direções espaciais não medidas, através do aporte ponderado de *HRTFs* de direções próximas. Existem muitos métodos apresentados, mas a maioria deles não apresenta uma preocupação rigorosa pela conservação das características perceptualmente relevantes, bem como os métodos de modelagem. A interpolação espacial é especialmente importante no contexto de reprodução de ambientes sonoros dinâmicos, i.e. fontes sonoras em movimento e de simuladores biaurais em tempo real. Diferentes métodos têm sido apresentados com diferentes graus de sucesso. Entre eles: Huopaniemi em 1999 [29], Biscainho em 2002 [50], entre outros.

Processos de interpolação podem ser definidos no domínio do tempo, da frequência, dos componentes principais, ou de qualquer outra característica funcional da *HRTF*. Além disso, outro tipo de interpolação é possível: no domínio das dimensões antropométricas. No entanto, a validade de qualquer método não é garantida ao menos que testes psicoacústicos formais sejam realizados.

2.8 Aplicações do som 3D

Atualmente já é possível sintetizar campos acústicos virtuais (*VASs* – *Virtual Auditory Spaces*) com grande grau de precisão, i.e. *VAS HI FI*. É claro que o custo e o tempo de implementação desses sistemas às vezes não é justificável, principalmente pela necessidade de uso de *HRTFs* individuais. A maioria dos *VADs* usa fones de ouvido na hora da reprodução. Outro tipo de *VADs* usam alto-falantes, porém o uso de fones permite o controle preciso do estímulo, atingindo boas simulações, especialmente se o sinal acústico mudar de acordo com o movimento do ouvinte. Normalmente, o sistema baseado em alto-falantes estéreo é mais eficaz quando a posição da cabeça é fixada numa posição. Em algumas aplicações, o mais importante é que o estímulo sonoro seja realista, isto é, que o som pareça provir de locais externos ao ouvinte. Nesse caso, obviamente a solução mais confiável é o uso de um sistema de alto-falantes, seja estéreo ou multicanal. Em outras aplicações, o mais

importante é que a informação espacial seja precisa e convenientemente transportada ao ouvinte. Neste caso, o uso de fones de ouvido e de *HRTFs* individuais apresenta vantagens importantes. Na sequência, três parâmetros a serem usados na caracterização de sistemas de som 3D dependendo da aplicação, são definidos:

- Realismo: Implica na externalização do som e ambientação.
- Imersão: Implica na sensação envolvente seja ou não seja controlada.
- Controle preciso dos objetos sonoros: Implica em um posicionamento preciso a fim de transmitir informação precisa.

A partir desses parâmetros, são definidos três tipos de *HRTFs* que devem ser usados em concordância com a aplicação a ser implementada:

- *HRTFs* genéricas: Em geral, *HRTFs* medidas numa cabeça-torso artificial que capture as características antropométricas de uma população abrangente.
- *HRTFs* personalizadas: Estimações da *HRTF* individual ou adaptações de *HRTF* genéricas a indivíduos específicos.
- *HRTFs* individuais: *HRTF* medida ou calculada em um indivíduo específico.

Além do tipo de *HRTF*, o uso de módulos para simulação acústica de salas, ou de simples efeitos de reverberação, também deve ser considerado como fator de projeto numa aplicação de som 3D. Na sequência apresentam-se algumas áreas genéricas de aplicações atuais onde são usados *VADs*. Citaremos algumas aplicações específicas.

2.8.1 Estudo científico de localização sonora

No início das pesquisas sobre localização sonora, a maioria de estudos foi realizada usando indicadores acústicos simplificados apresentados através de fones de ouvido ou

usando fontes sonoras reais posicionadas ao redor do ouvinte. O primeiro desses métodos caracterizava completamente apenas o *ITD* e o *ILD*, causando nenhum efeito em relação à externalização do som [47]. Por outro lado, estudos em locais têm permitido caracterizar o desempenho de localização em condições reais. Nesses estudos é quase impossível determinar a contribuição de cada indicador acústico natural à percepção. Nos últimos vinte anos, o desenvolvimento de técnicas de implementação de *VADs HI FI* tem permitido o estudo do mecanismo governante da localização auditiva natural. Contudo, ainda hoje existem muitas lacunas que continuam sendo pesquisadas. Nesse cenário, os sistemas baseados em *HRTFs* têm permitido pesquisar a importância dos diferentes indicadores de localização. Da mesma maneira, estudos psicoacústicos usando *VADs* baseados em *HRTFs* tentam determinar como os diferentes indicadores afetam a resposta neural em diferentes níveis da cadeia auditiva [51]. Neste caso, o custo e o tempo de implementação dos sistemas de espacialização não é um fator importante na escolha dos métodos de implementação.

2.8.2 Avaliação Clínica

Clinicamente, a percepção auditiva espacial pode ser um fator importante na predição do funcionamento de um paciente em condições normais de ambientes ruidosos [52]. Testes de localização espacial têm sido sugeridos para avaliar diferentes fatores relativos à audição, desde as funções de fones de ouvido especiais até as deficiências no processamento cerebral auditivo. Um dos principais benefícios da audição espacial em pessoas de audição normal é a capacidade de atender seletivamente fontes em direções particulares, i.e. o efeito *cocktail party* [26]. Para propósitos clínicos, a informação mais importante provida pelos testes de localização auditiva é se os pacientes podem ou não usar os indicadores de localização. Nesse caso, não é necessário simular as fontes sonoras com precisão para todas as possíveis direções, mas é importante que as fontes sejam facilmente localizáveis e, portanto, que a tarefa de treinamento dos pacientes para assistir aos testes seja simples e rápida. Assim, o sistema deve ser econômico e fácil de usar.

2.8.3 Transferência de informação espacial para comando/controle

Uma grande quantidade de tarefas demanda um operador analisar e responder a informação espacial. Controladores de tráfego aéreo devem rastrear vários aviões sob seu controle. Pilotos militares devem navegar sua aeronave enquanto rastreiam naves inimigas, objetos hostis ou quaisquer outros objetos. Operadores de veículos remotos ou de tele - robôs devem levar em conta os dados transmitidos desde essas locações distantes a fim de construir um modelo mental do sítio remoto. Informação espacial tem sido apresentada, na grande maioria dos casos, usando displays visuais. Assim, em geral, o canal visual está sempre sobrecarregado com operadores monitorando miríades, calibradores e displays gráficos. Conseqüentemente, atualmente existe um grande interesse no uso de displays auditivos para estas tarefas. Neste cenário de aplicações de comando/controle, o objetivo principal é prover o operador de informação precisa e que possa eliminar ambigüidades. Portanto, realismo não é fundamental, exceto no caso que facilite a tarefa do operador. E, por outro lado, a resolução espacial é crítica. Nestas aplicações, esquemas de processamento digital de sinais que possam realçar a quantidade de informação transferida ao operador, são extremadamente úteis, ainda quando o estímulo não seja tão natural assim. Neste caso o uso de *HRTFs* individuais ou, no mínimo, personalizadas, é fundamental.

2.8.4 Entretenimento

Muitas aplicações de entretenimento usam múltiplos alto-falantes para simular indicadores acústicos espaciais. Em sistemas domésticos de reprodução e em cinemas a tecnologia padrão *Dolby Surround* usa cinco alto-falantes mais um canal de reforço de baixas frequências. Essa tecnologia está se popularizando também no mercado do computador pessoal. Uma grande variedade de fabricantes vende produtos que usam algoritmos proprietários para gerar efeitos espaciais sobre alto-falantes, como por exemplo, *Spatializer Audio Laboratories* ou *Creative Labs*. Por outro lado, o áudio espacial baseado em fones de ouvido e *HRTFs* é vendido por fabricantes como *Sennheiser Electronics*, ou *AMD Inc*. As considerações mais importantes no caso de aplicações para entretenimento é o custo dos sistemas e o impacto perceptual da simulação. Nestas aplicações, as fontes sonoras devem parecer reais, isto é, devem ser externalizadas. Portanto, a precisão na simulação não é tão importante. Ao invés disso, a ênfase está no desenvolvimento de sistemas baratos que possam gerar a experiência mais realista ao usuário.

2.8.5 Sistemas de Telecomunicações e Realidade Virtual

As aplicações em sistemas de comunicações se centram, evidentemente, na reprodução da informação, i.e. a reprodução de voz. Sistemas de reprodução espacial são usados em videoconferência para reprodução de múltiplos canais de voz [53] e em realidade aumentada usada em sistemas de comunicação móveis para separação da conversação e das fontes reais presentes no redor do usuário [54]. Isto é útil dada a maior inteligibilidade da fala quando o objetivo é a localização da fonte em diferentes posições espaciais na presença de ruído, que em muitos casos podem ser outras fontes de voz não importantes para o usuário nesse momento. Esse fenômeno tem relação, mais uma vez, com o efeito *cocktail party* [26]. Dado que neste caso é suficiente separar espacialmente os canais, com pouca importância na precisão do controle espacial, sistemas genéricos podem ser usados. A economia do sistema é mais importante.

No caso da realidade virtual a transmissão de informação espacial pode requerer de diferentes graus de exatidão e, portanto, não existe um critério unificado como nos casos anteriores. Porém, o paradigma da realidade virtual é reproduzir tão fielmente quanto seja possível uma cena visual-aural qualquer. Assim, o realismo pode ser mais importante que o controle espacial preciso. Às vezes, a reverberação e os ecos artificiais não permitem um controle espacial ótimo. Isto deve ser aperfeiçoado de acordo com a aplicação.

2.9 Sumário

Neste capítulo foram apresentados os fundamentos de audição espacial que consideramos mais importantes no presente trabalho. A base do trabalho é a perspectiva mecânica do processo de audição espacial, isto é a transformação do sinal acústico através do ouvido externo, desde seu ingresso ao sistema auditivo através do *pinna* até atingir o tímpano. Considerando que a tecnologia binaural tem como paradigma a reprodução dos sinais no tímpano como se eles fossem emitidos por fontes reais e, portanto, a captura dos indicadores relevantes no processo de localização de fonte sonora nesse caminho, a cadeia auditiva percorrida no ouvido médio e no ouvido interno não é relevante. A primeira seção apresentou uma breve introdução à problemática da audição espacial, as diversas teorias, e algumas evidências neurofisiológicas, psicoacústicas e comportamentais das capacidades humanas de

localização sonora. A seção dois descreveu o sistema de coordenadas clássico usado em trabalhos e sistemas de áudio 3D. A seção três apresentou as principais teorias sobre localização de fontes sonoras e os indicadores relevantes nesse processo, fazendo uma separação necessária entre localização biaural, i.e. horizontal, e localização monoaural, i.e. vertical. Da seção quatro a sete foram desenvolvidos todos os conceitos relativos à *HRTF* relevantes neste trabalho. Finalmente a seção oito apresentou um sumário das principais aplicações do áudio 3D e os fatores a serem levados em conta dependendo dos requerimentos: *HRTFs* genéricas, *HRTFs* personalizadas ou *HRTFs* individuais.

Capítulo 3

SOLUÇÕES PROPOSTAS AO PROBLEMA DA PERSONALIZAÇÃO DA *HRTF*

3.1 Introdução

A *HRTF* é uma função individual que pode ser comparada à impressão digital de uma pessoa. Por tanto, não existem duas *HRTFs* iguais. Na Figura 3.1 mostramos a visualização espectral das *HRTFs* de três indivíduos da base de dados *CIPIC*, correspondentes aos ângulos de azimute de 0°, 45° e 80°. As diferenças são claras.

A pergunta é como personalizar a *HRTF* ou obter de maneira prática a *HRTF* individual? O caminho evidente é a medição direta descrita na seção 2.5. Contudo, pesquisas atuais em qualidade de medições de *HRTFs* demonstram que é bem difícil medir a *HRTF* certa de um indivíduo. Uma infinidade de fatores influencia no processo de medição, desde o cabelo e a roupa até a posição exata do microfone e pequenos movimentos involuntários da cabeça [55]. Mesmo quando não se tenha mudado nada quanto ao indivíduo, sua roupa, o equipamento, os sinais, etc. medições em dias distintos podem produzir resultados diferentes num intervalo de desvio de até 0,5dB [56]. Em qualquer caso, esse desvio é insignificante em termos perceptuais, e o comportamento geral da *HRTF* de um mesmo indivíduo não varia grandemente se o sistema de medição for o mesmo. Assim, por enquanto, a melhor maneira possível de sintetizar um *VAS HI FI*, é através da *HRTF* individual medida no indivíduo – ouvinte, usuário do sistema.

Figura 3.1. Visualização espectral bidimensional de *HRTFs* dos indivíduos 003, 010 e 020 da base de dados *CIPIC*, correspondentes aos ângulos de azimutes de 0°, 45° e 80°.

O outro caminho evidente é a resolução da equação de onda na superfície do corpo. Embora essa solução resulte ainda mais complexa do que a medição direta de *HRTFs* é bem provável no futuro esse método atingir resultados importantes. Além dessas abordagens, têm surgido vários métodos que tentam estimar uma *HRTF* individual por meios computacionais, isto é, personalizar a *HRTF* ao usuário atual do sistema. Essa estimação, obviamente, tem que ser feita a partir de dados concretos. Mas, quais são esses dados? Existem vários elementos possíveis: uma *HRTF* genérica ou um conjunto de *HRTFs* de diferentes indivíduos (base de dados); parâmetros antropométricos que descrevam adequadamente a superfície do corpo, ou até a “decisão psicoacústica” dos indivíduos-ouvintes. Contudo, o uso de parâmetros antropométricos é fundamental a fim de tornar a personalização da *HRTF* automática sem recorrer à intervenção do usuário. Agora, o problema a resolver é: quais parâmetros antropométricos são os melhores? e, sobretudo, como sintetizar ou estimar a *HRTF* de um novo indivíduo a partir desses parâmetros e/ou dessa *HRTF* genérica? Qual é essa relação matemática entre os parâmetros antropométricos e as características relevantes na percepção presentes inequivocamente na *HRTF*? Na sequência vamos apresentar as principais abordagens para personalizar a *HRTFs* encontradas na literatura. Nossa proposta nasce da combinação de duas delas, as abordagens estrutural e morfológica.

3.2 Solução por métodos numéricos

A modelagem numérica é a única abordagem exata para calcular a *HRTF* individual através de processos computacionais sem recorrer a medições acústicas. A técnica, desenvolvida e trabalhada principalmente por Kahana desde 1997 [57], usa um laser tridimensional para produzir uma representação geométrica exata da cabeça e do *pinna*. Na resolução da equação de onda são usadas as técnicas de modelagem por elementos finitos [14] e por elementos de contorno [57]. Com esse método, *HRTFs* podem ser determinadas computacionalmente com a mesma exatidão que medidas acusticamente, inclusive em altas frequências. Usando um modelo de 15 000 elementos da cabeça e *pinna*, Kahana demonstrou grande exatidão nas *HRTFs* calculadas até 15 kHz.

No entanto, a técnica apresenta muitos problemas na implementação prática. A varredura da cabeça é complicada pela presença de cabelo, e pelo obscurecimento das regiões atrás dos *pinnae* e interiores a eles. Por isso, replicar as características interiores ao *pinna* requer a confecção de moldes e a varredura separada deles. Após a união das diferentes varreduras, o número de elementos do modelo deve ser restrito a quantidades computacionalmente viáveis, sem comprometer a resolução espacial. Finalmente, a solução das equações acústicas requer uma quantidade significativa de computação. Assim, por todas essas razões, esta abordagem atualmente requer mais tempo e recursos que a medição acústica.

3.3 Solução por escalonamento em frequência

Esta abordagem busca estimar a *HRTF* de um indivíduo qualquer a partir da adaptação de uma outra *HRTF*. Essa adaptação é baseada num fator de escalonamento ótimo (*OSF* – *Optimal Scale Factor*) que, em teoria, pode ser estimado a partir da antropometria do novo indivíduo. Além da proposta numérica, esta proposta é baseada em hipóteses relativamente consistentes. O trabalho foi desenvolvido desde os primeiros anos da década de 1990 pelo grupo de pesquisa da Universidade de Michigan, liderado por John Middlebrooks.

O método trabalha com funções de transferência direcionais (*DTFs* – *Directional Transfer Functions*). Elas são versões da *HRTF* onde foi eliminada a componente comum a todas as direções. As hipóteses adotadas pelo método estão baseadas no primeiro trabalho do Middlebrooks [58]. Ele concluiu que o ganho de amplitude correspondente a uma direção específica num indivíduo particular pode ser parecido ao ganho de amplitude correspondente à mesma direção em outro indivíduo em frequências diferentes. Essas frequências tendem a ser inversamente proporcionais ao tamanho médio dos indivíduos. Assim: “sons de

freqüências menores, com comprimentos de onda maiores, interagem com cabeças e *pinnae* maiores numa maneira parecida em que sons de freqüências maiores, com comprimentos de onda menores interagem com cabeças e *pinnae* menores” [10]. Segundo Middlebrooks é possível supor que *DTFs* de diferentes indivíduos difiram, basicamente, nas freqüências onde características espectrais como ressonâncias e antirressonâncias se apresentam [10]. Portanto, seria possível alinhar aproximadamente duas *HRTFs* através da movimentação de uma delas (ou das duas) no eixo de freqüência. Esse procedimento foi chamado de escalonamento em freqüência e é governado por um fator de escalonamento, que é o fator que representa quantitativamente a movimentação da *HRTF* na freqüência. Na Figura 3.2 é apresentado o conceito de escalonamento em freqüência para dois indivíduos da base de dados medida pelo Middlebrooks. A figura foi obtida de [10].

O trabalho usa *DTFs* de 45 indivíduos diferentes correspondentes a 393 direções diferentes no sistema de coordenadas esférico. O intervalo de freqüência onde são avaliadas as diferenças espectrais entre indivíduos (*ISSDs* – *Inter-Subject Spectral Differences*) é 3,7 kHz – 12,9 kHz, que é o intervalo de freqüência mais importante na localização vertical de fonte sonora [4]. Essa *ISSD* foi calculada a partir da variância de 64 componentes freqüenciais, correspondentes ao intervalo de freqüência mencionado acima, do espectro obtido a partir da diferença aritmética entre as magnitudes espectrais das respostas de cada par de indivíduos. Assim, a *ISSD* entre dois indivíduos é a média dessa variância para as 393 direções testadas.

Segundo o método, é possível reduzir as *ISSDs* através de escalonamento em freqüência das *HRTFs*. O escalonamento em freqüência foi feito na escala de freqüência de oitavas obtendo como resultado a redução da *ISSD* numa média de 15,5% ao longo de todos os pares de indivíduos, e com uma redução de mais de 50% em 9,5% dos pares de indivíduos. O *OSF* de dois indivíduos está bem correlacionado com a razão entre os *ITDs* máximos desses indivíduos (0,71), a razão da altura total do *pinna* dos indivíduos (0,72) e a razão da largura da cabeça dos indivíduos (0,67). Desde esses resultados, é possível prever o *OSF* a partir do tamanho relativo dos indivíduos em relação a essas dimensões antropométricas, a partir da Equação (3.1):

$$\log_2(\text{OSF}) = 0,34 * \log_2(p_X/p_Y) + 0,527 * \log_2(c_X/c_Y) \quad (3.1)$$

onde p_X e p_Y correspondem à altura total do *pinna* do indivíduo X e do indivíduo Y respectivamente e c_X e c_Y correspondem à largura da cabeça do indivíduo X e do indivíduo Y respectivamente.

Figura 3.2. *DTFs* do plano médio. No painel esquerdo são mostradas as *DTFs* dos indivíduos 35 (linha delgada) e 7 (linha grossa) sem escalonamento. No painel direito são mostradas as *DTFs* dos mesmos indivíduos alinhadas a partir do fator de escalonamento ótimo *OSF*. Esse fator reduz as diferenças espectrais entre indivíduos *ISSD*. Figura obtida de [10].

A correlação entre o *OSF* estimado por essa equação e o *OSF* calculado a partir das *DTFs* medidas acusticamente é 0.82. Dado que o *OSF* reduz a *ISSD*, em princípio, foi possível melhorar o desempenho de localização de um sistema com *HRTFs* genéricas através do escalonamento em frequência [11]. O resultado dos testes psicoacústicos revelou que quando comparado com o desempenho de localização de um sistema de *HRTFs* genéricas, o sistema de *HRTFs* escalonadas reduz os erros de quadrante (resolução da ambigüidade frente/trás) desde 12,9% até 1,4%, e o erro de localização vertical médio desde 38,4° até 29,0° RMS. Portanto, todas as hipóteses foram comprovadas.

Contudo, podemos vislumbrar alguns problemas no método. Primeiramente, em geral, a melhoria produzida com o *OSF* em relação à *ISSD* não é tão consistente assim (15,5%). Porém, em termos perceptuais, aparentemente ela produz uma melhoria aceitável na resolução de ambigüidades frente/trás, mas não tanto na localização vertical. O fato de predizer o *OSF* a partir de apenas dois parâmetros antropométricos (largura da cabeça e altura do *pinna*) é muito simplista dada a complexa forma do *pinna*. Além disso, é bem provável que a consistência do resultado tenha sua maior concentração na predição de indicadores interaurais, importantes na localização horizontal. É conhecido que a largura da cabeça pode predizer com grande grau de exatidão as diferenças interaurais [9]. No entanto, provas contundentes de melhoria de desempenho exclusivamente no plano médio não foram apresentadas. Também não foram apresentados testes psicoacústicos com *OSFs* calculados a partir da Equação (3.1).

3.4 Solução por cotejamento com base de dados

A proposta de personalização da *HRTF* por cotejamento com base de dados foi feita pelo grupo de pesquisa em áudio espacial da Universidade de Maryland [12]. A proposta se baseia no cotejamento de certos parâmetros antropométricos do *pinna* de um indivíduo usuário qualquer com os mesmos parâmetros registrados em 45 indivíduos diferentes na base de dados *CIPIC*. As *HRTFs* daquele indivíduo da base de dados que apresentar “maior parecido antropométrico” com o indivíduo usuário, são usadas para a espacialização respectiva do som. Além disso, um modelo geométrico da *HAT* é adicionado para corrigir o comportamento em baixas frequências das *HRTFs* medidas experimentalmente da base de dados [59]. A Figura 3.3 apresenta as dimensões antropométricas do *pinna* registradas na base de dados *CIPIC*; o presente método usa apenas as 7 primeiras dimensões, desconsiderando as dimensões de profundidade, d8, d10 e a dimensão de rotação do *pinna*, d9. O primeiro passo do método é medir através de visão computacional as dimensões no indivíduo usuário. Isto é feito através de 3 fotografias do indivíduo, uma do torso e cabeça e uma de cada *pinna* em primeiro plano. As dimensões de 9 indivíduos de teste foram medidas e depois comparadas com medições manuais, validando o processo. Supondo que $\hat{d}_k, k = 1 \dots 7$, é o valor da k -ésima dimensão na imagem do *pinna* do indivíduo usuário, e d_k^i é o valor da mesma dimensão medida no i -ésimo indivíduo da base de dados, então é calculado o fator E^i através da Equação (3.2):

$$E^i = \sum_{k=1}^7 \frac{(\hat{d}_k - d_k^i)^2}{\sigma_k^2} \quad (3.2)$$

Figura 3.3. Medidas antropométricas do *pinna* registradas na base de dados *CIPIC* para 45 indivíduos.

onde o termo σ_k^2 é a variância da k -ésima dimensão ao longo de todos os indivíduos. Assim, o indivíduo $i, i = \arg \min_i E^i$, é eleito como o “mais parecido”. Neste caso, foram realizados testes psicoacústicos com 9 indivíduos, visando validar o método. Foram comparadas a *HRTF* genérica do *KEMAR* [43], a *HRTF* personalizada através do método, a *HRTF* personalizada através do método mais o modelo da *HAT* para corrigir as baixas frequências e a *HRTF* genérica do *KEMAR* mais o modelo da *HAT*. O principal objetivo foi melhorar a percepção de elevação com respeito às *HRTF* genéricas do *KEMAR*. Em geral, o modelo da *HAT* melhorou a resposta, principalmente em posições afastadas do plano médio confirmando os resultados apresentados em [59]. Porém, o método de personalização não teve bons resultados em todos os indivíduos. A média do erro em ângulo de elevação RMS entre todos os indivíduos foi

12,24° para a *HRTF* genérica, 12,46° para a *HRTF* personalizada e 10,36° para a *HRTF* personalizada mais o modelo da *HAT*.

Problemas evidentes do método são que, embora se preocupe pela personalização a partir de antropometria, não se preocupa por associar essa antropometria às características perceptualmente relevantes nos processos de localização sonora, i.e. ressonâncias e antirressonâncias presentes na *HRTF* em altas frequências. Além disso, o método não usa dimensões de profundidade do *pinna*, i.e. profundidade da concha, como *d8* e *d10*, provavelmente pela dificuldade de medição. Porém, essas dimensões são muito importantes na predição das ressonâncias e antirressonâncias como é demonstrado em [40]. Por fim, os testes psicoacústicos descartam o método para futuras aplicações. No entanto, o modelo geométrico da *HAT* confirma sua importância na melhoria do desempenho de localização em sistemas de espacialização baseados em *HRTFs*.

3.5 Solução por abordagem morfológica

Neste trabalho um modelo estatístico gerador de *DTFs* para uma população de 36 indivíduos provê uma base para a variação sistemática de cotejamento entre *DTFs* de teste (geradas pelo modelo) e as *DTF* individuais (que geraram o modelo) [15]. Usando esse modelo foram realizados experimentos psicoacústicos a fim de examinar a sensibilidade do desempenho de localização humano às diferenças inter-individuais nas *DTFs*. Adicionalmente, o mapeamento entre parâmetros antropométricos da periferia auditiva (cabeça, pescoço e *pinnae*) foi explorado.

A base de dados de *DTFs* consiste em magnitudes espectrais de 400 pontos de frequência, num total de 393 posições espaciais igualmente distribuídas sobre a superfície esférica, para 36 indivíduos de teste. As *DTFs* esquerda e direita foram concatenadas para cada indivíduo e para cada posição espacial, formando vetores de 800 pontos. Na sequência, análise em componentes principais (*PCA*) foi aplicada para os 14148 vetores (36 x 393), reduzindo os 800 pontos de cada vetor a 40 pontos (pesos dos componentes principais). As 393 *DTFs* comprimidas de cada indivíduo foram concatenadas novamente formando 36 vetores de 15720 pontos (393 x 40). Assim, cada um dos 36 vetores resultantes provê uma representação única da *DTF* de cada indivíduo dentro do espaço esférico todo. A seguir, uma segunda *PCA* foi realizada a fim de analisar a variação dos dados ao longo dos 36 indivíduos. A Figura 3.4a mostra a percentagem acumulada de explicação da variância dos 36 vetores

com respeito ao número de componentes principais usados na reconstrução. A diferença de outros trabalhos com *PCA* e *HRTFs*, onde os 5 primeiros componentes explicavam mais do 90% da variância [45] e em particular o primeiro vetor explicava mais do 80%, os resultados da abordagem morfológica mostram que são necessários mais de vinte vetores para chegar ao 90% de explicação da variância, e em particular o primeiro vetor explica apenas o 22% da variância. Resultados consistentes com o presente método de abordagem morfológica foram obtidos com nosso método, como será apresentado no seguinte capítulo.

Figura 3.4. Percentagem de explicação da variância com respeito ao número de componentes principais entre os 36 indivíduos usados no método de personalização por abordagem morfológica. (a) Variação dos vetores *DTFs*. (b) Variação dos vetores de morfologia antropométrica. Figura modificada de [15].

Parâmetros antropométricos foram medidos para 24 dos 36 indivíduos de teste. Esses parâmetros consistiram em vinte pontos definidos num espaço cartesiano (x, y, z) com origem na entrada do canal auditivo, e corresponderam a medidas da cabeça, *pinna* e pescoço. Em geral, esses parâmetros são consistentes com as dimensões registradas na base dados *CIPIC* (ver Figura 3.3). Vetores de 60 pontos (3×20) correspondentes às características morfológicas de cada indivíduo foram formados. Por fim, *PCA* foi realizada para comprimir esses vetores em 24 pontos. A Figura 3.4b mostra a percentagem acumulada de explicação da variância dos 24 vetores com respeito ao número de componentes principais usados na reconstrução.

Testes psicoacústicos de localização de fonte sonora foram realizados em cinco indivíduos a fim de determinar quantos componentes principais são necessários para reconstruir os vetores de *DTFs* obtendo boa localização. Os resultados mostraram que o uso de sete componentes principais provê boa localização. Segundo a Figura 3.4a, sete componentes possibilitam um grau de reconstrução da variância de 60%. Na média dos casos testados, o uso de mais componentes não apresentou uma grande melhoria na localização e, às vezes, introduz certa degradação na métrica usada para testar o desempenho de localização. Por outro lado, o mapeamento linear entre os vetores de morfologia e os sete primeiros componentes principais resultou alentador, atingindo coeficientes de correlação acima de 0,9. Porém, no artigo não ficou claro entre que fatores se deram esses coeficientes. Além disso, não foram apresentados resultados de testes psicoacústicos com novos indivíduos para validar o método.

3.6 Solução por modelagem estrutural

Heuristicamente é possível conceber a *HRTF* como uma soma de aportes de diferentes estruturas corporais como o *pinna*, a cabeça e o torso. Esses aportes podem ser observados no domínio do tempo ou da frequência (ver Figura 3.6) e assumem diferentes formas, devido a fenômenos físicos específicos, em especial a fenômenos refletivos e difrativos. A idéia básica é modelar cada estrutura usando como parâmetros dimensões antropométricas específicas a fim de personalizar o modelo.

Batteau apresentou a primeira aproximação estrutural; ele modelou o *pinna* como um sistema refletor de três canais de dois retardos cada, e um acoplador acústico [32]. Genuit desenvolveu um modelo formado por um banco de filtros, correspondentes ao torso, ombros,

pinna, concha, cabeça e canal auditivo [60]. Brown e Duda apresentaram um modelo estrutural em termos de quatro funções matemáticas: sombreamento e retardo causado pela cabeça, reflexões do ombro e reflexões do *pinna* [9]. Embora existisse uma forte interação entre as ondas acústicas que chegam a cada estrutura, em geral, a modelagem isolada destas provê uma boa aproximação. A seguir, apresentamos uma análise detalhada da composição estrutural da *HRTF* e como essa análise serve para a solução proposta no capítulo seguinte.

3.6.1 Análise estrutural

De maneira geral no plano médio onde não existem indicadores interaurais, a composição da *HRTF* pode ser observada na Figura 2.7, em termos da difração da cabeça, reflexão do torso, efeitos do *pinna* e reflexão do joelho. Todas essas características já foram confirmadas experimentalmente com a cabeça/torso artificial *KEMAR* [61] (ver Figura 3.6), medindo as respostas com diferentes partes removidas (*pinnae*, e/ou cabeça-torso), e analiticamente através de modelos geométricos [20] [59]. Na Figura 3.5 é possível observar a atividade depois do som direto devida à difração em torno da cabeça e às reflexões nas paredes do *pinna*, efeito que será discutido posteriormente. A difração da cabeça é explicada no domínio da frequência pelo modelo esférico de Lord Rayleigh [20] e é possível observá-la melhor em *HRTFs* contra-laterais afastadas do plano médio, como as observadas na Figura 3.1 (imagem das *HRTFs* esquerdas correspondentes aos ângulos de azimute de 45° e 80°). A segunda borda clara é a reflexão do torso entre 1ms e 2ms, observando-se que ela varia consistentemente com a elevação. O atraso entre o som direto e o som refletido pelo torso é máximo acima da cabeça (elevação de 90°) e decresce nos lados. No domínio da frequência o efeito desse atraso se apresenta como a estrutura de antirressonâncias com forma de arco ao longo do espectro todo. Esta onda refletida é fraca para elevações negativas devido à sombra causada pelo torso. A atividade observada depois de 2ms se deve à reflexão do joelho, dado que as medições foram feitas com pessoas sentadas. Em geral, esta reflexão é considerada sem importância no processo de localização [41]. Na Figura 3.5 essas reflexões podem ser observadas numa *HRIR* típica. Nessa figura, a região marcada como som direto está conformada pelo impulso inicial (o som direto) e as primeiras reflexões (tanto picos quanto vales) causadas pelas paredes do *pinna* e pela difração da cabeça. Na sequência, a região marcada como reflexão do torso abrange a mesma estrutura anterior, mas notoriamente atenuada. Finalmente a reflexão do joelho apresenta uma estrutura parecida, mas quase

imperceptível. Os outros indicadores principais são observados no domínio da frequência, e são as antirressonâncias presentes acima de 5 kHz. No domínio temporal, elas correspondem àquelas características constituintes da região som direto da Figura 3.5, além do impulso inicial. Três antirressonâncias proeminentes são observadas na Figura 2.7a para elevações entre -45° e 90° , elas variam diagonalmente na imagem; portanto a sua variação com a elevação é também evidente. Experimentos com a cabeça/torso artificial *KEMAR* confirmam que elas são causadas pelo *pinna* [61] (ver Figura 3.6).

Figura 3.5. *HRIR* direita correspondente ao azimuth de 0° , elevação de 45° , indivíduo 010.

Diversos modelos têm sido propostos para explicar a variação dessas antirressonâncias com respeito à elevação. Geralmente, esses modelos têm sido baseados em fenômenos refletivos e difrativos do *pinna*, em especial da concha ou cavidade principal. Entre eles temos os modelos de Batteau [32], Hebrank [31] e Lopez-Poveda [40]. Finalmente, o *pinna* também causa as principais ressonâncias da *HRTF* como demonstrado por Shaw [39]. Na sequência apresentaremos uma análise individual a partir da literatura e de algumas observações originais feitas com respeito ao aporte separado do *pinna* e da *HAT* à *HRTF*.

3.6.2 Contribuição da cabeça e torso à *HRTF*

O efeito do torso é apreciado na magnitude espectral da *HRTF* como aquele padrão de arcos que aparece quase sempre na imagem toda (ver Figura 2.7a). Por outro lado, o efeito da cabeça é reconhecido como as antirressonâncias verticais que aparecem nas imagens de *HRTFs* contralaterais. É lógico que em *HRTF* ipsilaterais esse padrão não apareça dado que o som atinge diretamente o *pinna* e não percorre um caminho difratado pela cabeça. O papel da

cabeça na *HRTF* e, em geral, no processo de localização sonora, é muito importante na conformação do *ILD* através do fenômeno de “sombreamento da cabeça”, e está bem entendido faz muitos anos. Contudo, o papel do torso na localização é menos entendido. O fato de o torso distorcer o campo acústico incidente em baixas frequências tem sido reconhecido faz tempo [62]. Porém, os efeitos do torso são relativamente fracos e os experimentos para comprovar sua influência perceptual como indicadores importantes em baixas frequências têm apresentado resultados não conclusivos.

Figura 3.6. Composição das *HRTFs* do manequim *KEMAR*. Figura modificada de [61].

Em 1987, Kuhn usou o *KEMAR* com e sem *pinna* e torso para estudar o comportamento da *HRTF* em todas as elevações no plano médio [62]. Ele mostrou que o torso fornece uma resposta variante com a elevação abaixo de 2 kHz e que o efeito do *pinna* aparece acima dos 4 kHz. Porém, a importância perceptual do torso não foi provada. Fora do plano médio, Genuit mostrou que o torso introduz efeitos variantes com a direção e a distância e estão limitados abaixo dos 3 kHz. Na sequência ele incluiu sub-modelos separados do torso e dos ombros no

seu modelo estrutural da *HRTF* [60]. Brown também incluiu um modelo refletivo do torso no seu modelo estrutural [9]. Finalmente, em [41] Algazi apresenta resultados a partir de provas psicoacústicas com *HRTFs* individuais e com modelos geométricos simplificados de *HRTFs* que mostram a existência de indicadores perceptuais significativos em baixas frequências para fontes afastadas do plano médio. Assim, a reflexão provida pelo torso é importante no lado ipsilateral e a difração da cabeça é mais forte no lado contralateral, onde o som direto é dramaticamente atenuado. Em [41] é apresentado um modelo de cabeça esférica e torso elipsoidal e em [59] é apresentado um modelo de cabeça e torsos esféricos, chamado de Homem de Neve¹³. Esse modelo foi usado no método de personalização descrito na seção 3.4. Nele, a cabeça esférica e o torso esférico são separados por certa distância (pescoço). O modelo usa dois algoritmos diferentes dependendo se a fonte estiver dentro ou fora da sombra do torso. Um diagrama com a separação das regiões do modelo é mostrado na Figura 3.7.

Figura 3.7. Regiões definidas pelo cone de sombra do torso no modelo “snowman”.

¹³ Do inglês *snowman*

A fonte fora da sombra do torso tem um caminho direto entre a fonte e o *pinna* e um caminho indireto refletido pelo torso. A fonte dentro da sombra do torso deve difratar o som ao redor do corpo até atingir o *pinna*. Todos os caminhos de propagação até o *pinna* contralateral são eventualmente sombreados pela cabeça.

Para o acoplamento entre um modelo do *pinna* para altas frequências ou *HRTFs* medidas experimentalmente, e o Homem de Neve, deve-se usar a fase do Homem de Neve ao longo de todo o eixo frequencial e gradualmente trocar a magnitude dele pela magnitude do outro modelo. Assim, a magnitude do Homem de Neve deve ser usada até 500 Hz. Desde 500 até 3000 Hz a magnitude do modelo do *pinna* é progressivamente (linearmente) incorporada e acima dos 3 kHz é usada apenas a magnitude do modelo do *pinna*. Esse modelo já provou sua eficácia, porém ele não provê indicadores relevantes no plano médio [12].

3.6.3 Contribuição do *pinna* à *HRTF*

A análise do aporte do *pinna* deve ser feito no domínio da frequência. Na década de 1970, Shaw [39] fez a mais ampla pesquisa no domínio da frequência usando um modelo de látex do *pinna* bem como *pinnae* naturais. Ele mediu a pressão sonora sob várias condições de som incidente, determinando com sucesso quais eram as causas das principais ressonâncias e antirressonâncias. Porém, em [39] ele apenas trabalhou com sons incidentes desde o plano vertical lateral. O resultado mais importante desse trabalho foi identificar as frequências onde são produzidas as principais ressonâncias na magnitude espectral da resposta do *pinna* isolado. A primeira ressonância aparece em 3 kHz aproximadamente, ela corresponde à ressonância de um quarto de comprimento de onda ($\lambda/4$) de um tubo de 30 mm fechado numa terminação. Esse comprimento resultou ser 33% maior do que o comprimento do tubo que foi usado para simular o canal auditivo no experimento. Assim, o *pinna* funcionaria como a prolongação do canal, ocorrendo um efeito de abertura. A segunda ressonância aparece em aproximadamente 5 kHz. Nesse caso, um máximo de pressão se apresenta na cavidade da concha toda. A distribuição da pressão é bem similar àquela produzida quando o canal auditivo é fechado. Shaw notou que oscilações da mesma fase ocorriam em toda a superfície de incidência do campo sonoro, assim, essa primeira ressonância da concha é amortecida por uma alta resistência de radiação, o que conduz a um contorno da ressonância relativamente largo. Ondas longitudinais estacionárias são as características essenciais das três seguintes ressonâncias de alta frequência, as quais aparecem aproximadamente em 9 kHz, 11 kHz e 13

kHz. Nós de pressão sonora aparecem e dividem a concha em várias regiões. Essas ressonâncias de alta frequência têm fator de qualidade bem maior que as duas primeiras. Esse fato é explicado pelo fraco acoplamento ao campo sonoro e, conseqüentemente, pelas perdas de radiação menores. Além do estudo feito com o modelo, Shaw também mediu o campo sonoro de seis *pinnae* naturais e conseguiu confirmar as duas primeiras ressonâncias. É bem provável que nesses *pinnae* naturais as três ressonâncias seguintes tenham ficado combinadas em apenas uma ressonância bem mais larga. Algumas dessas ressonâncias ocorrem em diferentes frequências nos *pinnae* naturais. Logicamente isto se deve à diferente geometria dos *pinnae* e à diferente terminação acústica do tímpano entre os indivíduos. Embora Shaw conseguisse medir a dependência da função de transferência do ouvido externo com a posição da fonte sonora, ele explicou a dependência com a elevação apenas no caso da antirressonância que aparece ao redor dos 8 kHz. Esse resultado estaria em concordância com a teoria refletiva do Bateau, pois Shaw atribuiu essa variação a um fenômeno de interferência. Além do Shaw, o Blauert [4] mediu também a função de transferência do ouvido externo com um aparelho experimental parecido ao usado pelo Shaw. Esse experimento examinou a variação da resposta em relação à variação da fonte sonora no plano horizontal. Ele observou que a segunda ressonância correspondente à concha mantém amplitude constante entre os ângulos de azimute de 0 e 90°, caindo entre 15 e 20 dB até o azimute de 110° e permanecendo nesse nível baixo até 180°. Assim, som procedente da frente excita esse modo mais fortemente que sons provenientes de trás. Além disso, também foi observado que o padrão de ondas estacionárias na região do *pinna* depende da direção do som, mas esse não é o caso do canal auditivo como foi confirmado posteriormente em [56]. A transmissão dentro do canal, por ser de natureza unidimensional, é independente da direção. Os estudos posteriores de Shaw [63] confirmaram os resultados anteriores, apresentando medições em dez indivíduos de teste, e estabeleceram os seis modos normais ressonantes do *pinna*, os quais aparecem na grande maioria dos casos e são excitados em frequências e elevações particulares. A Figura 3.8 mostra as regiões do *pinna* que produzem esses modos.

Nesse cenário, um modelo autorregressivo (AR) de 12 coeficientes seria capaz de rastrear as ressonâncias produzidas pelo *pinna* na *HRTF*. Assim, na Figura 3.9 apresentamos a variação da magnitude espectral da *HRTF* com a elevação e o modelo AR de 12 coeficientes originado pela respectiva *HRTF*. O objetivo é ressaltar os modos ressonantes. Nesse caso, apresentamos as visualizações espectrais correspondentes aos ângulos de azimute de 0° (plano médio) e de 80° (muito perto ao plano vertical lateral). Na figura se observa como os modos

ressonantes se movem com respeito à imagem, sendo que no caso do azimute de 80° as ressonâncias se apresentam em frequências relativamente constantes ao longo do intervalo de elevação todo, revelando um comportamento não dependente da elevação. Por outro lado, no plano médio, os modos ressonantes são maximamente excitados em intervalos específicos de elevação, revelando um comportamento dependente da elevação. Além disso, a variação da disposição dos modos ressonantes com o azimute revela importância na conformação do *ILD* e, portanto na localização horizontal.

Figura 3.8. Os seis modos normais identificados por Shaw. Para cada modo são mostrados: o número de modo, a frequência de ressonância, o ângulo de elevação onde é maximizada a resposta e o nível em dB da ressonância. Figura obtida de [63].

O modelo *AR* rastreia razoavelmente bem as ressonâncias da *HRTF*, portanto, será usado no método de personalização de *HRTF* apresentado no seguinte capítulo.

Como tratado na seção 2.3.2 existe evidência psicoacústica, neurofisiológica e comportamental que bem como ressonâncias, as antirressonâncias são muito importantes na percepção da elevação. Além dos trabalhos do Hebrank e Batteau discutidos brevemente naquela seção, um trabalho muito importante foi o realizado por Lopez-Poveda em 1996 [40].

Nele, um modelo físico aproximado da função de transferência da concha é desenvolvido. A formulação inclui difração, reflexão e interferência na cavidade. O desempenho do modelo é comparado com o modelo *SDA* do Batteau, de acordo com o grau de predição em comparação com as medições experimentais num modelo metálico com forma de espiral. Os resultados mostraram que o modelo de difração-reflexão se desempenha consideravelmente melhor em relação à predição da frequência central de antirressonâncias e do espaçamento relativo entre elas. O mesmo modelo é aplicado a uma forma realista da concha e suas predições são comparadas com *HRTFs* medidas experimentalmente com fontes sonoras variantes em azimuth e elevação. Nesse caso, o modelo prediz as características espectrais dependentes da elevação relativas às dimensões transversais da concha. Dados experimentais medidos na cabeça-torso *KEMAR* mostraram a presença de uma antirressonância proeminente cuja posição em frequência varia entre 6,5 kHz e 10 kHz para ângulos de elevação entre -40° e 40° . Esse resultado é consistente com o apresentado por Shaw [39] no plano vertical lateral e Hebrank [31] no plano médio. Adicionalmente, no plano médio, existe uma segunda antirressonância centrada em 10,8 kHz e presente apenas até a elevação de -10° . Finalmente, uma terceira antirressonância aparece entre 14 kHz e 17 kHz e sua frequência central varia suavemente com a elevação. Para elevações maiores que 40° a forma espectral é mais parecida a um declive gradual que decresce notavelmente em altas frequências. O modelo prediz com grande grau de exatidão as frequências centrais da primeira e terceira antirressonância, mas não da segunda. Isto se pode dever ao fato de que o *crus-helias* não foi incluído na formulação do modelo. O *crus-helias* é uma estrutura que divide a concha em duas partes. Além disso, a amplitude tampouco é predita com exatidão, o que poderia se dever às simplificações feitas na formulação acústica.

Bem como os modos ressonantes do *pinna*, numerados de 1 a 6 e apresentados na Figura 3.8 e 3.9, no presente trabalho denominaremos os contornos das quatro primeiras antirressonâncias observadas na *HRTF* como A, A2, B e C. Além disso, detalharemos diversas observações feitas em *HRTFs* medidas no plano médio com respeito a essas antirressonâncias, e confirmaremos as observações feitas por Lopez-Poveda nas *HRTFs* do *KEMAR* com respeito às três primeiras antirressonâncias (A, A2 e B), e as observações de Hebrank [31], Macpherson [37] e Shaw [63] com respeito à primeira antirressonância (A). A Figura 3.10 mostra quatro conjuntos de *HRTFs* ipsilaterais apresentadas através da visualização espectral, correspondente cada uma delas a um ângulo de azimuth diferente, 0° , 30° , 50° e 80° , respectivamente. Cada visualização é apresentada junto à visualização

correspondente ao espectro do sinal residual da análise preditiva linear processado através de janelamento e de filtragem de auto-reforço. Esse espectro e o respectivo processamento serão apresentados no seguinte capítulo; a particularidade dele é que conserva e isola as antirressonâncias causadas pelo *pinna* na imagem. Na figura é possível observar que no plano médio são mostrados os quatro principais contornos de antirressonâncias correspondentes a A, A2, B e C; nos ângulos de azimute de 30 e 50° o contorno correspondente a A2 desaparece completamente e os de B e C ficam bem fracos; no ângulo de azimute de 80° somente fica o de A.

Figura 3.9. Modos ressonantes do *pinna* em visualizações bidimensionais de *HRTFs* e seu correspondente modelo autorregressivo de 12 coeficientes. *HRTFs* direitas correspondentes ao indivíduo 010.

Tudo o que foi mencionado acima nos leva a formular nossas principais hipóteses e simplificações no seguinte capítulo, visando a propor e implementar o método de personalização automática da *HRTF*. Em geral, tudo nasce a partir das referências encontradas na literatura com respeito aos modos ressonantes e aos contornos de antirressonâncias e, sobretudo, a partir das confirmações dessas referências e demais observações feitas no presente trabalho, em relação às *HRTFs* da base de dados *CIPIC*.

Figura 3.10. Antirressonâncias causadas pelo *pinna* mostradas em visualizações bidimensionais da *HRTF* e do espectro do sinal residual LP. *HRTFs* direitas correspondentes ao indivíduo 010.

3.7 Sumário

No presente capítulo, os principais métodos de personalização de *HRTFs*, encontrados na literatura, foram descritos. Mesmo que existam algumas abordagens adicionais, consideramos que as mais importantes para nosso trabalho são as tratadas aqui. Além disso, um detalhamento da abordagem estrutural e sua importância no processo de personalização foram tratados na seção 3.6. A partir de todas as considerações anteriores, é possível concluir que as características antropométricas são definitivamente os dados a partir dos quais uma nova *HRTF* deve ser sintetizada ou aproximada. Quais características são as mais importantes?

Aparentemente na produção das ressonâncias da *HRTF* todas as dimensões do *pinna* são importantes; por outro lado, na produção das antirressonâncias são fundamentais as dimensões da concha. Além dessas características, é claro que uma base de dados de *HRTFs* correspondente a uma população de indivíduos que seja relativamente representativa, é necessária. Por fim, um modelo personalizado da *HRTF* que vise a superar os problemas tratados na seção 2.7.1.2 deve levar em conta: a reflexão do torso, a qual é importante para localização vertical abaixo dos 3 kHz quando a fonte sonora está afastada do plano médio, e as ressonâncias e antirressonâncias causadas pelo *pinna*, as quais definem a localização vertical e a resolução de ambigüidades frente/trás acima dos 3 kHz. Mesmo que alguns modelos geométricos da cabeça e torso tenham sido apresentados nos últimos 4 anos com diferentes graus de sucesso, não existe modelo nenhum do *pinna* que complete a cadeia.

Capítulo 4

TÉCNICAS E ALGORITMOS DESENVOLVIDOS E APLICADOS NA ESTIMAÇÃO DA PRTF

4.1 Introdução

Do exposto nos capítulos anteriores, foi possível entender a importância da personalização automática da *HRTF*. No contexto da abordagem estrutural, dados os modelos geométricos da *HAT*, é necessário desenvolver um modelo personalizado da *PRTF* que capture os indicadores perceptualmente relevantes no processo de localização sonora, especialmente na localização vertical e resolução de ambigüidades frente/trás. Segundo estudos psicoacústicos, comportamentais e neurofisiológicos, essas características são as ressonâncias e antirressonâncias da *HRTF* causadas pela interação da onda acústica com o *pinna*. Neste cenário, que significa capturar as ressonâncias e antirressonâncias da *HRTF* causadas pelo *pinna*? Significa implementar um método que possa sintetizar uma resposta de magnitude onde as amplitudes relativas e as posições em frequência dessas características espectrais dependam do indivíduo. E quais parâmetros podem representar esse indivíduo? Parâmetros antropométricos que, em conjunto, definam univocamente cada *pinna*.

O presente trabalho adota uma abordagem estatística baseada na base de dados *CIPIC*. A idéia é mapear o espaço vetorial correspondente às características antropométricas escolhidas no espaço vetorial correspondente a uma representação espectral da *HRTF* (ou neste caso, da

PRTF). Dessa maneira, seria possível mapear novos vetores antropométricos através das relações achadas. A solução será mais bem detalhada nas seguintes seções.

O presente capítulo apresenta inicialmente algumas observações sobre três padrões encontrados em representações bidimensionais da *HRTF* do plano médio. Esses padrões estão relacionados com os contornos de antirressonâncias causadas pelo *pinna* e sua evolução com a elevação. Essa evolução é muito mais evidente do que aquela apresentada por ressonâncias, como mostrado nas Figuras 3.9 e 3.10. Os contornos de antirressonâncias variam diagonalmente com respeito à elevação, levando-nos à hipótese que a frequência central dessas antirressonâncias denota uma elevação determinada para o sistema auditivo. Essa hipótese já foi comprovada por diversos estudos em [36] [38] como mencionado na seção 2.3.2. Além disso, na literatura já foram registradas algumas dessas antirressonâncias, mas até hoje não foram compendiadas da maneira feita no presente trabalho. Os três padrões encontrados no plano médio estão definidos a partir do número e distribuição desses contornos de antirressonâncias. A classificação pode ser estendida a posições fora do plano médio observando principalmente *HRTFs* ipsilaterais, porém nessas posições vários dos contornos observados no plano médio desaparecem ou se fundem apenas em um contorno como visto no final do capítulo anterior. A importância do plano médio em nossa análise reside no fato de que nele o processo de localização vertical é isolado contundentemente devido a que não existem indicadores interaurais significativos. Além disso, a cabeça não exerce difração importante e a reflexão do torso tampouco fornece indicadores de localização nesse setor [41]. Como veremos nas próximas seções, a tarefa de classificação será importante nos seguintes estágios do trabalho.

Após a classificação em padrões, serão apresentadas quantitativamente as hipóteses e simplificações adotadas pelo método, o qual será descrito na sequência, diferenciando os vários estágios de processamento e de modelagem. A seguir, esses estágios serão desenvolvidos em separado, finalizando com uma avaliação preliminar do método através de algumas métricas introduzidas. Mas primeiro, são apresentados os detalhes operacionais que estabeleceram o marco de trabalho.

4.2 Detalhes operacionais

Na sequência, são listados os detalhes operacionais que determinaram o cenário do presente trabalho:

- As respostas impulsivas relativas à cabeça *HRIRs* usadas (base de dados *CIPIC*) foram gravadas com frequência de amostragem de 44,1 kHz, resolução de 16 bits. Cada resposta impulsiva consta de 200 amostras, isto é, 4,53 ms. Além disso, todas as *HRIRs* foram normalizados à escala de amplitude $<-1,1>$ para cada indivíduo.
- No cálculo das respostas em frequência, *HRTFs*, foi usada a função *FFT* (*Fast Fourier Transform*) fornecida pelo pacote de software MATLAB¹⁴. As *FFTs* foram calculadas com 1024 pontos de frequência (512 até a metade da frequência de amostragem), usando *zero-padding* para completar a sequência de 200 amostras temporais das *HRIRs*.
- Todos os processamentos foram realizados no ambiente de desenvolvimento MATLAB.
- As representações bidimensionais das *HRTFs* são matrizes de 512 x 50 pontos para cada *pinna*. Isto se deve aos 512 pontos calculados até a metade da frequência de amostragem, 22,05 kHz, e aos 50 ângulos de elevação fornecidos na base de dados.
- Todos os processamentos na magnitude espectral foram realizados com a amplitude em escala logarítmica (dB), exceto a Análise em Componentes Principais (*PCA*) que foi realizada com escala de amplitude linear.
- A análise das *HRTFs* foi feita apenas no plano médio como será justificado na seção 4.4. Essa simplificação pressupõe que dado que não existem indicadores interaurais nesse setor, as *HRIRs* esquerda e direita de um mesmo indivíduo devem estar alinhadas, i.e. o pico correspondente ao som direto deve estar alinhado. No entanto, a fim de assegurar que esta condição seja cumprida, foi aplicado um segundo pré-processamento em todas as *HRIRs*: um algoritmo de alinhamento para todas as respostas impulsivas. O algoritmo é descrito no diagrama de fluxo da Figura 4.1, e um exemplo com uma *HRIR* real é apresentado na Figura 4.2. No exemplo, o algoritmo não entra no segundo laço e é alinhado a partir do primeiro pico. Esse comportamento

¹⁴ Informação sobre o pacote MATLAB pode ser encontrada no sítio web <http://www.mathworks.com>

é necessário a fim de evitar o alinhamento a partir de um falso pico inicial que aparece em muitas *HRIRs*. A amplitude desse pico inicial normalmente não alcança o 10% do pico principal produzido pelo som direto. Aparentemente, a causa do aparecimento desse falso pico inicial seria a natureza do microfone de prova ER-7C da Etymotic (e do tubo de prova) com que foram gravadas as respostas impulsivas nos canais auditivos dos indivíduos [64].

4.3 Classificação em padrões de representações bidimensionais da *HRTF*

A partir da observação cuidadosa da base de dados de *HRTFs* em formato bidimensional, e depois do processamento detalhado na seção 4.3.2, foi possível identificar três contornos de antirressonâncias principais (A, B e C em ordem de aparição) e dois secundários (A2 entre A e B, e C2 depois de B) nas *HRTFs* do plano médio, hemisfério frontal. Os principais contornos foram assim catalogados pelo fato de aparecerem na maioria dos indivíduos. Os secundários nem sempre aparecem ou às vezes aparecem quase imperceptíveis para a vista. Como explicado anteriormente, os contornos principais já foram registrados na literatura, especialmente o primeiro contorno, A [35 – 39]. A presença e orientação desses contornos na imagem são os fatores que definem os três padrões encontrados no plano médio e que serão descritos na seção 4.3.3. Os fatos observados que auxiliaram essa definição são listados a seguir:

- Na elevação de -45° , o menor ângulo de elevação registrado, o contorno A sempre aparece entre 4,5 kHz e 7 kHz.
- Na elevação de -45° , o contorno B sempre aparece entre 8 kHz e 12 kHz, mas sua maior ocorrência fica entre 10 kHz e 11 kHz.
- O contorno C é o mais imprevisível. Na elevação de -45° , ele sempre aparece entre 13 kHz e 17 kHz. Porém, às vezes dois contornos aparecem nesse intervalo de frequência, denominando C2 ao outro contorno. No caso de C e C2 aparecerem, indistintamente um deles aparece sempre antes de 15 kHz e o outro numa frequência maior. O

contorno principal C será sempre o mais profundo ou forte e o contorno secundário C2 o mais fraco.

- Os contornos A e B estão presentes em todos os indivíduos
- Os contornos A e B evoluem até aproximadamente 30° elevação. Depois, eles desaparecem.
- O contorno A é o contorno mais diagonal e o mais longo com respeito à imagem. Ele é relativamente horizontal até aproximadamente 0°, caindo vertiginosamente depois dessa elevação.
- Em geral, os contornos B e C são sempre mais horizontais que A.
- Em geral, o contorno C é o contorno mais curto desaparecendo logo após 10° de elevação. Porém, em alguns casos ele atinge 30° elevação diagonalmente, bem como o contorno A.
- Normalmente, todos os contornos principais caem com a elevação, exceto o contorno B que pode subir. Porém, esse caso não é muito comum.
- Na maioria dos indivíduos, um contorno secundário denominado A2 aparece entre A e B. Ele geralmente fica bem próximo a um dos contornos (A ou B), muitas vezes fundindo-se com ele. Sua evolução acaba em 0°, e raramente em 10° de elevação.
- No hemisfério posterior não foi possível estabelecer padrões comuns nas *HRTFs*. Além disso, o processo de localização nesse setor não tem sido estudado com profundidade na literatura.

Na sequência serão apresentados os pré-processamentos realizados nas imagens para a melhor observação dos contornos das antirressonâncias e, portanto, para auxílio à classificação em padrões, e finalmente a descrição de tais padrões. Mas primeiro,

apresentamos a primeira tentativa de aproximação da $PRTF$ através de uma abordagem bidimensional.

Figura 4.1. Diagrama de fluxo do algoritmo de alinhamento de *HRIRs*.

Figura 4.2. Aplicação do algoritmo de alinhamento nas *HRIRs*. Na figura são mostradas apenas as amostras iniciais. *HRIR* direita correspondente ao azimuth de 0°, elevação de 45°, indivíduo 010.

4.3.1 Aproximação da *PRTF* através de processamento bidimensional

A primeira tentativa da fase inicial da pesquisa foi isolar os contornos das antirressonâncias através de processamento bidimensional, isto é, a partir de técnicas de processamento de imagens aplicadas às imagens artificiais correspondentes à visualização espectral da *HRTF*. Os resultados do trabalho foram publicados em [65]. Essa tentativa, embora considerada exótica, representa a primeira vez que a *HRTF* é analisada e processada como uma função de mais de uma dimensão (neste caso, frequência e elevação). Dada a natureza multidimensional da *HRTF* (frequência ou tempo, elevação e azimuth), consideramos que foi uma tentativa válida e, no futuro, se pretende continuar explorando esses métodos.

Os contornos das antirressonâncias causadas pelo *pinna* foram aproximadamente isolados, transformando as imagens em níveis de cinza em imagens binárias onde as regiões

correspondentes a antirressonâncias ficaram pretas (nível 0) enquanto o resto da imagem virava branco (nível 1). O objetivo final daquele trabalho já era aproximar a resposta do *pinna* através da conformação de uma resposta formada por ressonâncias e antirressonâncias. Conforme discutido na seção 3.6.3, as ressonâncias são rastreadas por um modelo autorregressivo de 12 coeficientes, dados os 6 modos ressonantes do *pinna*. Quanto às antirressonâncias, elas são rastreadas pelo espectro do sinal residual da análise preditiva linear como demonstrado em [64]. A fim de isolar a resposta do *pinna*, se tentou minimizar a influência da reflexão do torso no espectro do sinal residual. Isto é feito através de processamento temporal. Se uma função $f(n)$, dependente do tempo, multiplicasse o sinal residual da *HRIR* poder-se-ia enfatizar a primeira parte do sinal e minimizar a parte final, onde a reflexão do torso se apresenta. Esse processo é parecido ao feito por um simples janelamento, mas sem eliminar completamente a parte final do sinal, dado que a influência do *pinna* também pode atingir esse intervalo final. Baseados em [64], no presente trabalho, chamamos aquele processo de processo de ênfase temporal governado pela função $f(n)$. Essa função é definida através da Equação (4.1):

$$f(n) = r^n \quad (4.1)$$

Onde r é o fator de ênfase e n é o índice temporal. Com r variando no intervalo $\langle 0,9 - 0,97 \rangle$ é possível enfatizar a resposta do *pinna* sobre a resposta do torso. Contudo, o fator r depende do indivíduo e às vezes é necessário somar várias imagens binárias correspondentes a diferentes fatores de ênfase para atingir uma imagem final onde as antirressonâncias, observadas por inspeção na imagem origem, sejam adequadamente isoladas. Na Figura 4.3 é apresentada uma imagem para cada estágio de processamento. Logo após a ênfase, a imagem é binarizada através de um simples limiar. Usando o histograma da imagem em níveis de cinza, o limiar foi escolhido como o menor nível de cinza (mais perto ao preto) abrangendo 11% da imagem. Essa porcentagem foi selecionada dado que, na média, apenas 11% da imagem do sinal residual enfatizado corresponde às antirressonâncias do *pinna*. Finalmente, as regiões da imagem das *HRTFs* originais correspondentes às antirressonâncias, marcadas pelos setores pretos nas imagens binárias, foram somadas à imagem correspondente ao modelo autorregressivo, conseguindo uma imagem aproximada da *PRTF*. O algoritmo é apresentado

no diagrama de fluxo da Figura 4.4. O grande problema da abordagem bidimensional está na hora de diferenciar o tipo de contorno (A, A2, B, C, ou C2), dado que às vezes são introduzidas certas antirressonâncias espúrias pelo processo de ênfase temporal e, outras vezes, acontece a não detecção de alguns contornos de antirressonâncias importantes.

Figura 4.3. Estágios de processamento da abordagem bidimensional de aproximação da *PRTF* a partir de análise preditiva linear, e “ênfase” temporal. *HRTFs* esquerdas correspondentes ao indivíduo 010.

Figura 4.4. Algoritmo para aproximação bidimensional da *PRTF* apresentado em [65].

As frequências centrais das antirressonâncias causadas pelo *pinna* não são identificadas para futuros estágios de comparação ou atribuição antropométrica, devido a que os contornos de antirressonâncias são grosseiramente aproximados pela binarização. Uma vantagem do método é que as *PRTFs* de todas as posições no plano médio são aproximadas simultaneamente, porém, esse fato pode introduzir inexatidão nas *PRTFs*. No entanto, uma das hipóteses adotadas, como será visto na seção 4.4, é que todas as posições espaciais devem ser modeladas em separado a fim de analisar isoladamente a origem de cada padrão de *HRTFs*. Na Figura 4.5 são apresentados resultados do método bidimensional.

Figura 4.5. Algumas *PRTFs* aproximadas a partir da abordagem bidimensional.

Devido a que a presente abordagem bidimensional não foi implementada no método final, ela não será mais detalhada. Maiores detalhes podem ser encontrados em [65]. Contudo, a pesquisa dos métodos bidimensionais deixou dois importantes fatos a serem usados no método final:

- A comprovação de que o espectro do sinal residual LP preserva as antirressonâncias presentes na *HRTF*, como apresentado em [64].
- A derivação de um método de reforço de contraste para a melhor observação das antirressonâncias e, portanto, para auxílio da classificação manual em padrões. Esse método será apresentado na sequência.

4.3.2 Método de reforço de contraste para observação de contornos de antirressonâncias em representações bidimensionais de *HRTFs*

O método consiste basicamente em janelar unidimensionalmente a *HRIR* para eliminar a influência do torso e aplicar filtragem bidimensional de auto-reforço a fim de aumentar o contraste na imagem. Dessa maneira, as regiões escuras ficam mais escuras e as regiões claras, mais claras.

No método desenvolvido foi usada a janela de Kaiser dada sua maior versatilidade e uso em aplicações modernas. Ela nos permite variar gradualmente a forma temporal da janela sem variar o comprimento desta. Com essa variação é possível buscar uma compensação entre a largura do lóbulo central e o tamanho dos lóbulos secundários na resposta de magnitude da janela. A janela de Kaiser é descrita pela Equação (4.2):

$$w_k(n) = \frac{I_0 \left[\alpha \sqrt{1 - \left(\frac{|2n - N + 1|}{N - 1} \right)^2} \right]}{I_0(\alpha)}, 0 \leq n \leq N - 1 \quad (4.2)$$

$$I_0(x) = \sum_{k=0}^{\infty} \left(\frac{x^k}{2^k \cdot k!} \right)^2$$

onde α é o parâmetro que controla a forma temporal da janela e I_0 a função de Bessel do primeiro tipo de ordem zero. O parâmetro α foi fixado em 5 para todos os indivíduos, dado que esse valor funciona na maioria dos casos. No caso de se pretender aperfeiçoar o método, α deve ser variado de acordo com o indivíduo. Um outro fato importante ao respeito do janelamento é o uso de apenas meia janela, isto é, a janela centrada em tempo zero. Em forma parecida ao processo de ênfase temporal da seção anterior, aqui se pretende não distorcer os instantes iniciais do sinal a ser janelado (a *HRIR*), e sim minimizar e/ou eliminar os instantes finais do sinal, onde se apresenta a reflexão do torso.

A Figura 4.6 mostra uma janela de Kaiser típica de 1,1 ms e $\alpha = 5$; a *HRIR* direita do indivíduo 10, elevação 45° e azimuth 0°; a mesma *HRIR* alinhada (ver seção 4.5.1) junto com meia janela Kaiser e finalmente a *HRIR* alinhada e janelada. Na Figura 4.7 são apresentados

os três estágios de pré-processamento temporal até aqui detalhados, em formato bidimensional, bem como a resposta de magnitude espectral respectiva. Note-se a eliminação da borda correspondente à reflexão do torso na subfigura d da figura 4.6, e do padrão em forma de arco no terceiro espectro da subfigura f da figura 4.7.

Logo após o janelamento é aplicada uma filtragem bidimensional de auto-reforço. Matematicamente a filtragem de auto-reforço define-se como aquela onde o resultado do processo de filtragem (convolução) é adicionado à imagem original.

Figura 4.6. Janelamento de *HRIR* com meia janela Kaiser para eliminar reflexão do torso.
HRIR direita correspondente ao azimuth de 0°, elevação de 45°, indivíduo 10.

Neste caso, é utilizado um filtro passa-altas a fim de reforçar as componentes de altas frequências que dizem respeito à região das bordas, ou mudanças bruscas de nível. Esse processamento é descrito pela Equação (4.3):

$$G(x, y) = f(x, y) + \beta \cdot [h(x, y) * f(x, y)] \quad (4.3)$$

onde $G(x, y)$ é a função bidimensional resultante da aplicação do filtro de auto-reforço, $f(x, y)$ é a função bidimensional original, $h(x, y)$ é a resposta ao impulso que define o filtro passa-altas

e β é um escalar. Essa filtragem bidimensional é aplicada à imagem correspondente à resposta de magnitude espectral resultante do alinhamento e janelamento das HRIRs, a fim de diferenciar as regiões de antirressonâncias. O parâmetro β foi fixado em 0.2, porém, maior aperfeiçoamento do método requereria adaptação desse parâmetro a cada indivíduo. O filtro passa-altas implementado foi do tipo *FIR*. Maiores especificações sobre o método e o filtro usado podem ser encontrados em [66]. Finalmente é obtida uma imagem onde é possível identificar com facilidade os contornos de antirressonâncias definidos anteriormente.

Figura 4.7. Alinhamento e janelamento de *HRIRs* em visualizações temporal e espectral.
HRIRs e *HRTFs* direitas correspondentes ao azimute de 0°, indivíduo 010.

4.3.3 Padrões encontrados no plano médio

A classificação foi realizada por simples inspeção. Em futuros estágios do trabalho se pretende implementar algoritmos de classificação que automatizem o processo. O Padrão 1 apresenta os três contornos principais A, B e C, e o contorno secundário A2 atingindo

indistintamente A ou B antes de 30° elevação. O Padrão 2 é igual ao Padrão 1, mas sem o contorno A2. O Padrão 3 é também igual ao Padrão 1 se diferenciando no fato que o contorno secundário A2 evolui independentemente de A e de B, sem atingi-los. O contorno secundário C2 pode aparecer em qualquer padrão, em qualquer indivíduo. A Figura 4.8 apresenta um diagrama esquemático dos três padrões e de suas diferenças. Uma amostra de cada padrão, com as imagens processadas segundo a seção anterior, pode ser apreciada na Figura 4.9. Os contornos de antirressonâncias são identificados e marcados com linhas tracejadas.

Figura 4.8. Padrões encontrados em representações bidimensionais de *HRTFs* do plano médio.

A Tabela 4.1 apresenta a percentagem de ocorrência dos três padrões na base de dados *CIPIC*, *HRTFs* do plano médio. Foram incluídas no cálculo as 90 *HRTFs* correspondentes aos *pinnae* esquerdos e direitos dos 45 indivíduos. Embora as *HRTFs* esquerda e direita de um mesmo indivíduo sejam bem parecidas, em 5% dos casos elas não pertencem ao mesmo padrão.

Tabela 4.1. Percentagem de ocorrência dos 3 padrões do plano médio

Padrão 1	Padrão 2	Padrão 3
69,66%	21,34%	8,98 %

Figura 4.9. Amostra dos 3 padrões encontrados no plano médio.

4.4 Hipóteses e simplificações adotadas

Na presente seção, as hipóteses e simplificações adotadas no método de personalização são enumeradas. A maioria delas já foi mencionada direta ou indiretamente nas seções anteriores, portanto, a seguinte será principalmente um resumo do que foi feito até aqui.

- a) Pretende-se estudar apenas a localização vertical. Esse processo é o mais dependente do indivíduo dado que se baseia em detalhes anatômicos finos, i.e. o *pinna* e suas cavidades. Assim, desconsidera-se qualquer análise ou estudo de indicadores interaurais e da localização horizontal. O estudo será feito, portanto, no plano médio.
- b) Dado que o processo de localização vertical no hemisfério posterior não tem modelos desenvolvidos nem apresenta nenhum padrão de características coerente (número e distribuição de ressonâncias e antirressonâncias), o estudo será feito apenas no hemisfério frontal, desconsiderando a análise de ambigüidades frente/trás.
- c) A solução proposta é uma combinação da abordagem estrutural e da abordagem morfológica. Da primeira abordagem se usará a separação de modelos: *HAT* e *pinna*. A *HAT* já foi modelada com sucesso através de aproximações geométricas [59], no entanto, o *pinna* não tem sido modelado com bons resultados. Da abordagem morfológica se adotará a proposta de estabelecer mapeamentos lineares entre espaços vetoriais correspondentes a uma população de *HRTFs* (em nosso caso: *PRTFs*) e aos parâmetros antropométricos dos indivíduos onde as *HRTFs* foram medidas.
- d) Dado que já foi provado que a *HAT* não fornece indicadores de localização vertical em posições afastadas do plano médio [41], ela não será considerada no estudo nem tampouco nos eventuais testes psicoacústicos. O estudo se concentrará apenas em modelar e testar a *PRTF*.
- e) O modelo considerará tanto os modos ressonantes quanto os contornos de antirressonâncias na modelagem da *PRTF*. Porém, dado que a variação das antirressonâncias com a elevação é bem mais notável, a preocupação principal será o correto posicionamento delas no eixo de frequências. A variância inter-individual da frequência onde as ressonâncias aparecem é bem menor do que aquela das antirressonâncias. Por outro lado, em geral, as ressonâncias têm largura de faixa bem maior do que as antirressonâncias de alta frequência.

- f) Segundo as observações feitas na seção anterior, apenas os contornos de antirressonâncias A e B aparecem em todos os indivíduos. Assim, se tomará maior cuidado na estimação de suas frequências centrais.
- g) A maioria dos trabalhos tem modelado a *HRTF* em todas as posições espaciais simultaneamente [10] [15]. Embora na média dos resultados essas abordagens conseguissem melhorias em relação às *HRTFs* genéricas, a modelagem separada de cada posição pode trazer melhorias quantitativas ainda maiores. Além disso, cada posição espacial pode apresentar maior dependência de certas características antropométricas e não de todas as características disponíveis. Assim, apresentaremos o método de personalização tomando como estudo de caso apenas uma posição espacial no plano médio: -23° . A escolha se deve a que, como visto na Figuras 4.6 e 4.7, nessa posição espacial todas as ressonâncias e antirressonâncias aparecem claramente. O método pode ser usado para modelar posições fora do plano médio, mas deve-se tomar cuidado com os padrões de contornos de antirressonâncias variáveis entre indivíduos. Nesse caso, necessita-se modelar apenas a *HRTF* ipsilateral e derivar a *HRTF* contralateral. A *HRTF* ipsilateral possui uma imagem mais “limpa” dado que não apresenta difração da cabeça e o som atinge diretamente o *pinna*. Estudos psicoacústicos revelam claramente que a *HRTF* ipsilateral é muito mais importante em qualquer processo de localização sonora, sendo possível inclusive localizar monoauralmente apenas com esse ouvido habilitado [35]. A Figura 4.10 apresenta um esquema das posições modeladas.
- h) Trabalharemos apenas na modelagem da magnitude espectral da *PRTF*. Reconstruções de fase mínima são usadas para derivar as respostas impulsivas correspondentes, i.e. respostas impulsivas relativas ao *pinna* – *PRIRs* (*Pinna-Related Impulse Response*). Dado que trabalharemos no plano médio, esta simplificação é ainda mais viável já que não teremos que modelar o *ITD*.

Figura 4.10. Esquema de posições espaciais modeladas no plano médio, e de eventuais posições ipsilaterais. Figura modificada de [13].

4.5 Solução proposta a partir de Base de Dados CIPIC

O esquema geral da proposta é apresentado na Figura 4.11. A idéia básica por trás da personalização da *HRTF* a partir de dados antropométricos é que é bem mais fácil, eficiente e econômico adquirir imagens e calcular parâmetros através de técnicas de visão computacional, do que medir acusticamente *HRTFs* ou modelá-las numericamente. Assim, nossa proposta está dividida em sete fases. As fases 1 e 7 têm a ver com a implementação pratica do método, e serão detalhadas mais profundamente nos seguintes capítulos. Por outro lado, as fases 2 a 5 são técnicas de processamento desenvolvidas no presente trabalho e serão detalhadas a partir da seguinte seção. A fase 6 já foi tratada brevemente na seção 3.6.2..

Na seqüência, cada fase será brevemente descrita.

- Fase 1: A fim de implementar um sistema completo, é necessário, inicialmente, desenvolver um sistema de aquisição e processamento de imagens. O sistema deve adquirir imagens dos *pinnae*, cabeça e torso de um indivíduo de teste, e calcular os

parâmetros necessários usados pelas técnicas de processamento (vetores de parâmetros antropométricos).

- Fase 2: Nesta fase é iniciado o trabalho com a base de dados. Em nosso caso, usamos a base de dados *CIPIC* [13]. Dado que vamos modelar apenas a *PRTF*, é necessário conformar uma base de dados de *PRTFs* baseada na base de dados de *HRTFs*. O processo de extração de *PRTFs* é realizado através de técnicas de processamento de sinais [67] [2].
- Fase 3: As *PRTFs* devem ser modeladas a fim de obter uma representação reduzida que possa conformar o “espaço vetorial *PRTF*”. Essa representação deve capturar a variância inter-individual das *PRTFs* e desconsiderar a parte comum aos indivíduos todos. Análise em componentes principais (*PCA*) foi a escolha de modelagem.
- Fase 4: Será realizado um mapeamento linear entre o “espaço vetorial *PRTF*” e o “espaço vetorial antropométrico”. O último é conformado por vetores que representem detalhes antropométricos relevantes do *pinna* de cada indivíduo. O problema dos mínimos quadrados será resolvido para calcular as transformações lineares entre espaços vetoriais. Em próximas fases da pesquisa se planeja aplicar outros tipos de mapeamento.
- Fase 5: O vetor antropométrico do novo indivíduo obtido na fase 1 é mapeado ao “espaço vetorial *PRTF*” através das transformações lineares calculadas na fase anterior. A nova *PRTF* é conformada e a *PRIR* reconstruída com fase mínima. No caso de modelar a *PRTF* fora do plano médio, o *ITD* também deve ser implementado.
- Fase 6: Em posições fora do plano médio, é necessário implementar um modelo da *HAT*. Isto é feito através dos modelos geométricos descritos anteriormente [59], e dos parâmetros antropométricos da *HAT* obtidos na fase 1. O modelo da *HAT* pode ser agregado ao modelo do *pinna* através do procedimento descrito em [59]. Dado que no plano médio a *HAT* não faz diferença nenhuma, não a descreveremos com detalhe.

- Fase 7: Finalmente, a fim de comprovar a eficácia do método em relação às *HRTFs* genéricas é necessário realizar testes de localização com o novo indivíduo. Esses testes serão descritos no seguinte capítulo.

A seguir, serão descritas com detalhe as fases 2 a 5.

Figura 4.11. Esquema geral da solução proposta ao problema de personalização da *HRTF*.

4.6 Estágio de Extração da *PRTF*

O processo de extração da *PRTF* está baseado num algoritmo publicado recentemente para extrair as frequências centrais das antirressonâncias causadas pelo *pinna* na *HRTF* [67], junto com outros algoritmos desenvolvidos no presente trabalho. A validação do presente método só pode ser realizada através de medições acústicas com modelos do *pinna* separados do resto do corpo como as realizadas em [61]. No entanto, são assumidas como verdadeiras as conclusões dos diversos trabalhos que dizem respeito à resposta do *pinna* e sua importância na localização vertical de fontes sonoras. Esses estudos já foram discutidos amplamente nas seções anteriores. O algoritmo a ser detalhado é de natureza unidimensional, isto é, o processamento é aplicado em cada *HRTF*. Primeiro, será descrito o processamento realizado nas *HRTFs* para capturar as ressonâncias. A seguir, o algoritmo de extração das frequências centrais de antirressonâncias causadas pelo *pinna* será descrito. Finalmente será apresentado o método de aproximação da resposta de magnitude espectral da *PRTF* a partir das ressonâncias e antirressonâncias extraídas previamente.

4.6.1 Rastreamento de ressonâncias da *HRTF*

Até aqui é claro que as ressonâncias da *HRTF* são causadas pela interação da onda acústica com o *pinna*. Como antecipado na seção 3.6.3, um modelo autorregressivo de 12 coeficientes é bom candidato na conformação de uma resposta que rastreie apenas as 6 ressonâncias da *HRTF*. Assim, é aplicada análise preditiva linear nas *HRTFs*, com o modelo autorregressivo descrito pela Equação (4.4):

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + \sum_{k=1}^p a_k \cdot z^{-k}} \quad (4.4)$$

onde $A(z)$ é o filtro de análise e a_k os coeficientes de predição linear. A ordem do modelo, p , deve ser 12 dado que se procuram 6 ressonâncias no máximo. Se p for maior, ressonâncias espúrias podem aparecer. No caso de estudo (-23° , plano médio) marcado com linhas tracejadas na Figura 4.8, o modelo autorregressivo rastreia bem as ressonâncias em 87% dos casos considerando 90 *HRTFs* correspondentes aos *pinnae* esquerdos e direitos de 45

indivíduos. Isto foi verificado por inspeção. O 13% restante foi desconsiderado nos estágios seguintes (ver Tabela 4.2).

4.6.2 Extração das frequências centrais de antirressonâncias causadas pelo *pinna* na *HRTF*

O presente sub-estágio é baseado no algoritmo de [67]. Ele calcula primeiro o sinal residual da predição linear (LP) da *HRTF*. Esse sinal é descrito através da Equação (4.5):

$$e(n) = x(n) - \hat{x}(n) = x(n) + \sum_{k=1}^p a_k \cdot x(n-k) \quad (4.5)$$

onde $x(n)$ denota o sinal original, i.e. a resposta impulsiva relativa à cabeça *HRIR*, e $\hat{x}(n)$ a versão predita da *HRIR* através do modelo autorregressivo. O sinal residual representa a parte imprevisível da *HRIR*. Assim, a resposta de magnitude espectral desse sinal preserva apenas as antirressonâncias da *HRTF*, como mostrado na Figura 4.3. Nesse espectro, as ressonâncias não exercem influência na formação de vales, facilitando o rastreamento de frequências das verdadeiras antirressonâncias. Na sequência, é necessário eliminar as antirressonâncias causadas por outras estruturas corporais que não são o *pinna*, i.e. a *HAT*. Então, é aplicada meia janela de Kaiser como feito e explicado na seção 4.2.1.1 e mostrado na Figura 4.6. O comprimento da janela e o parâmetro α são variáveis entre indivíduos e, por exemplo, janelas de comprimento maior do que o necessário podem causar a preservação de antirressonâncias espúrias. Por facilidade, fixamos o parâmetro α em 5; no entanto, o comprimento da meia janela variou desde 0,7 ms. até 1,1 ms, sendo a ocorrência maior em 1 ms. Por outro lado, a operação de janelamento diminui a resolução espectral, isto é, suaviza a magnitude espectral alargando consideravelmente os contornos das ressonâncias e antirressonâncias. Em geral, janelas de comprimento menor dramatizam esse efeito. Isto pode mascarar ou até alterar as frequências a serem achadas pelo algoritmo, por isso, Raykar em [67] propõe a aplicação da função de autocorrelação ao sinal residual janelado, $e(n)$, descrita pela Equação (4.6):

$$R(-m) = \sum_{n=m}^{N-1} e(n) \cdot e(n-m), \text{ e} \quad (4.6)$$

$$R(m) = R(-m)$$

onde $e(n)$ é o sinal ao qual é aplicada a autocorrelação, neste caso, o sinal residual janelado. Toma-se apenas a parte positiva do sinal autocorrelação, isto é, se usa o mesmo número de amostras, N , que o sinal original. Assim, o efeito de alargamento das características espectrais devido ao janelamento é significativamente atenuado quando a janela é aplicada no sinal de autocorrelação [67]. Isto se deve ao fato que, em geral, o sinal de autocorrelação possui amplitudes que decrescem em forma assintótica quando existem componentes refletidos no sinal original, auxiliando no cômputo da função de atraso de grupo e preservando a maioria de detalhes da envoltória espectral. Assim, é aplicada meia janela Kaiser com os mesmos parâmetros que a primeira vez.

A partir do estágio prévio de classificação em padrões é conhecido o tipo ao qual pertence cada *HRTF* (1, 2 ou 3). Assim, é possível saber quantos e quais contornos de antirressonâncias devem ser extraídos, faltando apenas encontrar as frequências centrais onde eles aparecem. Neste cenário, é possível aplicar uma primeira janela de comprimento determinado e dependendo do número de frequências extraídas ser maior que o devido, aplicar progressivamente janelas de menor comprimento até encontrar o número certo de frequências centrais.

A seguir, a função de atraso de grupo é calculada para o sinal de autocorrelação janelado do sinal residual janelado. Essa função é descrita através da Equação (4.7):

$$\tau(w) = -\frac{d\theta(w)}{dw} \quad (4.7)$$

onde $\theta(w)$ é a função de fase espectral do sinal ao qual é aplicada a função de atraso de grupo, neste caso, o sinal de autocorrelação janelado do sinal residual janelado. A característica interessante da função de atraso de grupo é que enfatiza as ressonâncias e antirressonâncias, tornando menos esforçada a extração de mínimos locais [67]. Na sequência, os mínimos locais são achados através de um algoritmo que busca alterações de inclinação ao longo do intervalo de frequência onde aparecem as antirressonâncias do *pinna*: 5 kHz a 17 kHz. Às vezes são encontrados mais ou menos mínimos do que os devidos (conhecidos pelo processo de classificação), sendo necessária a aplicação de um limiar. A

escolha do limiar é auxiliada interativamente pelo tipo ao qual pertence a *HRTF* analisada, dado que, bem como no caso do janelamento, o número de antirressonâncias a serem achadas já é conhecido previamente. Normalmente o limiar varia entre -2 dB e 2 dB, sendo a maior ocorrência em 1 dB. Na prática, somente foi implementado o processo iterativo para detecção do limiar do atraso de grupo e não aquele para estabelecimento do comprimento das janelas. O algoritmo completo é ilustrado na Figura 4.12.

Finalmente, devido aos diversos processamentos realizados na *HRTF*, é comum que as frequências centrais das antirressonâncias mudem ligeiramente. Assim, como estágio final é aplicado um ajuste dessas frequências extraídas. Esse ajuste é realizado procurando o mínimo local da *HRTF* original mais próximo à frequência extraída. Essa busca também se baseia na alternância de inclinação bem como a busca de mínimos locais na função de atraso de grupo. Finalmente, as frequências centrais de antirressonâncias ficam como aquelas frequências dos mínimos das *HRTFs* originais mais próximos às frequências extraídas pelo algoritmo.

A Figura 4.13 ilustra o algoritmo completo com exemplos das magnitudes espectrais processadas.

Já a Figura 4.14 mostra a influência da classificação em padrões no processo de detecção por limiar usado na função de atraso de grupo. Enquanto a *HRTF* direita do indivíduo 28 pertence ao padrão 2 (contornos A, B e C), a *HRTF* direita do indivíduo 27 pertence ao padrão 1 (contornos A, A2, B e C), como pode ser visto nas sub-figuras a e d. Assim, usando o limiar de 0dB na função de atraso de grupo do indivíduo 28, é introduzida uma antirressonância espúria entre B e C. Então, para achar o número correto de antirressonâncias é preciso mudar o limiar para -1 dB. Com esse valor, A, B e C são achadas corretamente. Por outro lado, usando -1 dB como limiar no indivíduo 27, o contorno B não é rastreado. Portanto, é necessário mudar o limiar para 0 dB a fim de achar corretamente A, A2, B e C. A classificação em padrões diz quando o limiar deve ser mudado a fim de achar o número correto de contornos de antirressonâncias.

Como pode ser visto na Figura 4.14 também, no caso do indivíduo 27 é preciso um considerável ajuste das frequências centrais das antirressonâncias, principalmente no caso de A2. A Figura 4.15 mostra mais de perto como é necessário esse pequeno ajuste na maioria das frequências centrais extraídas.

Figura 4.12. Algoritmo de extração da *PRTF* aproximada. Para simplificação, alguns detalhes repetidos da Figura 4.4 foram omitidos.

Figura 4.13. Processo de extração de frequências centrais de antirressonâncias causadas pelo *pinna* na *HRTF*. *HRTF* correspondente ao azimuth de 0° , elevação de -23° , indivíduo 010.

4.6.3 Aproximação da magnitude espectral da *PRTF*

A partir da combinação do modelo *AR* da *HRTF* e das frequências centrais de antirressonâncias causadas pelo *pinna*, é possível conformar uma magnitude aproximada da *PRTF*. O algoritmo leva em conta três elementos: a magnitude espectral do modelo *AR*, as frequências centrais de antirressonâncias achadas no processo anterior e a magnitude espectral original da *HRTF*.

Inicialmente, a *PRTF* é assumida como a magnitude do modelo *AR*, porém, nas frequências centrais de antirressonâncias, o valor da *PRTF* adota o valor da *HRTF* original. Na sequência, a partir dessas frequências centrais, o algoritmo busca, progressiva e regressivamente, a coincidência de valor em dBs entre a magnitude da *HRTF* original e a magnitude do modelo *AR*. Às vezes acontece que essas duas magnitudes nunca coincidem ou que a coincidência está muito longe da frequência central. Então, são determinados dois valores de tolerância, um para a diferença aritmética entre as duas magnitudes e outro para o intervalo de frequência onde o algoritmo busca a coincidência. A tolerância de magnitude é

0.5dB, quando o algoritmo atinge essa diferença entre as duas magnitudes ele pára. A tolerância de frequência é 500 Hz para cada lado da frequência central. Se o algoritmo não atinge o valor de 0.5dB para a diferença entre as magnitudes, ele pára em 500 Hz à esquerda ou 500 Hz à direita da frequência central. Assim, no máximo, o intervalo de busca da coincidência entre as magnitudes é de 1 kHz por frequência central. Se não for atingida nenhuma das duas tolerâncias, o algoritmo encontra duas frequências (correspondentes às coincidências das magnitudes citadas anteriormente), uma à direita e outra à esquerda da frequência central. Essas frequências determinarão o intervalo de frequências procurado pelo algoritmo. Dentro desse intervalo, a *PRTF* adota o valor da *HRTF* original. Como complemento, é aplicado um filtro gaussiano com $Q = 20$ para suavizar a magnitude da *PRTF*. A Figura 4.16 mostra o processo completo. O resultado do processo de extração das 64 *PRTFs* válidas é mostrado na Figura 4.17. A trajetória formada pelas frequências centrais de antirressonâncias A e B é tracejada ao longo das 64 *PRTFs*. As “zonas de influência” dos contornos A2, C e C2 são também marcados: ficam evidentes as diferenças entre indivíduos. As *PRTF* esquerda e direita de cada indivíduo ficaram juntas na figura.

Figura 4.14. Influência da classificação em padrões no processo de extração da *PRTF*.

Figura 4.15. Ajuste de frequências centrais de antirressonâncias causadas pelo *pinna*. Cruzes indicam mínimos locais da *HRTF* original e pontos pretos indicam as frequências centrais extraídas pelo algoritmo. *HRTF* correspondente ao indivíduo 050.

Figura 4.16. Aproximação da magnitude espectral da *PRTF* esquerda do indivíduo 044.

Cruzes indicam frequências centrais de antirressonâncias e pontos pretos indicam as “coincidências” entre a magnitude da *HRTF* e o modelo *AR*. Neste caso, já que de fato os espectros coincidem, nem a tolerância de frequência nem a de magnitude foi requerida.

Figura 4.17. *PRTFs* extraídas de 64 *HRTFs* correspondentes ao *pinna* esquerdo e direito de 32 indivíduos da Base de dados *CIPIC*. O resto de *HRTFs* foi excluído do processo. Ver Tabela 4.2.

4.6.4 Resultados finais dos processos de classificação e extração

Finalmente, na Tabela 4.2 são apresentados os dados mais relevantes da classificação em padrões da base de dados *CIPIC*, do rastreamento de ressonâncias realizado pelo modelo autorregressivo descrito na seção 4.5.1 e da extração de frequências centrais de antirressonâncias descrita na presente seção. Notem-se alguns detalhes importantes:

- Como mencionado anteriormente, o modelo autorregressivo não sempre rastreia corretamente as ressonâncias da *HRTF*.
- O padrão não está sempre em concordância com o tipo de ressonâncias extraídas. Por exemplo, no caso No. 9 o padrão reconhecido por inspeção é o 1, ou seja, a *HRTF* possui os três contornos de antirressonâncias principais A, B e C e o contorno secundário A2. Porém, o algoritmo extrai apenas A, A2 e C, sem detectar B. Nestes

casos, o ajuste de parâmetros variantes entre indivíduos (processo iterativo) que foi descrito brevemente na seção 4.6.2, tampouco funciona e é necessária uma supervisão do resultado para determinar o tipo de contorno extraído. Assim, às vezes o algoritmo não conseguiu rastrear o número de antirressonâncias esperado.

No final desse estágio já se conta com uma representação reduzida e particular de cada *PRTF* (e, portanto, de cada *HRTF*): as frequências centrais das principais antirressonâncias causadas pelo *pinna*. No presente trabalho, tais frequências são definidas como FCAs (Frequências Centrais de Antirressonâncias), e são listadas também na Tabela 4.2.

Tabela 4.2. Dados relevantes da classificação em padrões da base de dados *CIPIC*

No.	Indivíduo	Rastreamento de ressonâncias	Número de antirressonâncias extraídas	Tipo de antirressonâncias extraídas	Padrão
1	'003'L	Bem: 4 de 5	3	A-B-C	2
2	'003'R	Bem: 4 de 5	3	A-B-C	2
3	'010'L	Bem: 4 de 5	3	A-B-C	2
4	'010'R	Bem: 4 de 5	3	A-B-C	1
5	'018'L	Bem: 4 de 5	4	A-A2-B-C	3
6	'018'R	Bem: 4 de 5	4	A-A2-B-C	3
7	'020'L	Bem: 4 de 5	3	A-B-C	1
8	'020'R	Bem: 4 de 5	3	A-B-C	1
9	'021'L	Bem: 5 de 6	3	A-A2-C	1
10	'021'R	Bem: 5 de 6	3	A-A2-C	1
11	'027'L	Bem: 5 de 6	4	A-A2-B-C	3
12	'027'R	Bem: 5 de 6	4	A-A2-B-C	3
13	'028'L	Bem: 5 de 6	3	A-B-C	2
14	'028'R	Bem: 5 de 6	3	A-B-C	2
15	'033'L	Bem: 5 de 6	4	A-A2-B-C	1
16	'033'R	Bem: 5 de 6	4	A-A2-B-C	1
17	'040'L	Bem: 5 de 6	3	A2-B-C	1
18	'040'R	Bem: 5 de 6	3	A2-B-C	1
19	'044'L	Bem: 5 de 6	3	A-B-C	1
20	'044'R	Bem: 5 de 6	3	A-B-C	2
21	'048'L	Bem: 5 de 6	4	A-A2-B-C	3
22	'048'R	Bem: 5 de 6	4	A-A2-B-C	1

23	'050'L	Bem: 6 de 6	3	A-B-C	1
24	'050'R	Bem: 6 de 6	3	A-B-C	1
25	'051'L	Mal		A-C	2
26	'051'R	Mal		A-C	2
27	'058'L	Bem: 4 de 5	3	A-B-C	2
28	'058'R	Bem: 4 de 5	3	A-B-C	1
29	'059'L	Mal		A-A2-B	1
30	'059'R	Mal		A-A2-B-C	1
31	'060'L	Mal		A-A2-B-C	1
32	'060'R	Mal		A-A2-B-C	1
33	'061'L	Bem: 5 de 5	3	A-A2-B	2
34	'061'R	Bem: 4 de 6	4	A-A2-B	2
35	'065'L	Bem: 6 de 6	3	A-B-C	1
36	'065'R	Bem: 6 de 6	3	A2-B-C	1
37	'119'L	Bem: 6 de 6	3	A-B-C	2
38	'119'R	Bem: 6 de 6	3	A-B-C	2
39	'124'L	Bem: 6 de 6	3	A2-B-C	1
40	'124'R	Bem: 6 de 6	4	A-A2-B-C	3
41	'126'L	Bem: 4 de 5	3	A-A2-B	1
42	'126'R	Bem: 6 de 6	4	A-A2-B-C	1
43	'127'L	Bem: 4 de 5	4	A-A2-B-C	1
44	'127'R	Mal	4	A-A2-B-C	1
45	'131'L	Bem: 4 de 5	3	A-B-C	2
46	'131'R	Bem: 4 de 5	3	A-B-C	2
47	'133'L	Bem: 6 de 6	3	A-B-C	2
48	'133'R	Bem: 4 de 5	3	A-B-C	2
49	'134'L	Bem: 6 de 6	4	A-A2-B-C	1
50	'134'R	Bem: 4 de 5	4	A-A2-B-C	1
51	'135'L	Bem: 6 de 6	3	A-B-C	1
52	'135'R	Bem: 6 de 6	3	A-B-C	1
53	'137'L	Bem: 4 de 5	4	A-A2-B-C	3
54	'137'R	Bem: 4 de 5	4	A-A2-B-C	3
55	'147'L	Bem: 4 de 5	3	A-A2-C	2
56	'147'R	Bem: 4 de 5	3	A-B-C	1
57	'148'L	Bem: 3 de 5	3	A-B-C	1
58	'148'R	Bem: 4 de 5	3	A-A2-C	1
59	'152'L	Bem: 4 de 5	4	A-A2-B-C	1
60	'152'R	Bem: 4 de 5	4	A-A2-B-C	1

61	'153'L	Bem: 4 de 5	3	A-B-C	1
62	'153'R	Bem: 4 de 5	4	A-A2-B-C	1
63	'154'L	Bem: 4 de 5	4	A-A2-B-C	1
64	'154'R	Bem: 4 de 5	4	A-A2-B-C	1
65	'155'L	Bem: 3 de 4	4	A-A2-B-C	1
66	'155'R	Mal	4	A-A2-B-C	1
67	'156'L	Bem: 3 de 4	4	A-A2-B-C	1
68	'156'R	Bem: 4 de 5	3	A-B-C	1
69	'162'L	Bem: 4 de 5	3	A2-B-C	1
70	'162'R	Bem: 4 de 5	3	A2-B-C	1
71	'163'L	Bem: 4 de 5	4	A-A2-B-C	1
72	'163'R	Bem: 4 de 5	3	A-B-C	1
73	'165'L	Bem: 4 de 5	4	A-A2-B-C	1
74	'165'R	Bem: 4 de 5	4	A-A2-B-C	1

Indivíduos sem dados antropométricos

No.	Indivíduo	Rastreamento de ressonâncias	Número de antirressonâncias extraídas	Tipo de antirressonâncias extraídas	Padrão
	008'L	Bem: 4 de 5	3	A-B-C	2
	008'R	Bem: 4 de 5	3	A-A2-B-C	2
	009'L	Mal	4	A-A2-B-C	1
	009'R	Mal	4	A-A2-B-C	1
	011'L	Bem: 4 de 5	4	A-A2-B-C	1
	011'R	Bem: 4 de 5	4	A-A2-B-C	1
	012'L	Bem: 4 de 5	4	A-A2-B-C	1
	012'R	Bem: 4 de 5	4	A-A2-B-C	1
	015'L	Bem: 4 de 5	4	A-A2-B-C	1
	015'R	Bem: 4 de 5	4	A-A2-B-C	1
	017'L	Bem: 4 de 5	4	A-A2-B-C	1
	017'R	Bem: 4 de 5	4	A-A2-B-C	1
	019'L	Mal	4	A-A2-B-C	1
	019'R	Mal	4	A-A2-B-C	1
	158'L	Bem: 4 de 5	4	A-A2-B-C	1
	158'R	Bem: 4 de 5	4	A-A2-B-C	1

4.7 Estágio de Modelagem da *PRTF*

O seguinte passo é modelar as *PRTF*s extraídas a fim de reduzir a quantidade de dados e conseguir uma segunda representação reduzida e particular de cada *PRTF* (além das FCAs do estágio anterior). Foi escolhida a técnica de Análise em Componentes Principais (*PCA*) por duas razões:

- O método separa a parte variante (pesos dos componentes principais) da parte comum (componentes principais) em um grupo de vetores N-dimensional [68].
- O método já foi usado em vários trabalhos com *HRTF*s [45] [69].

Assim, definimos um espaço vetorial formado por vetores de 512 dimensões correspondentes à magnitude espectral em dB das 64 *PRTF*s extraídas. No presente trabalho esses vetores são chamados de “vetores *PRTF*”. No caso em estudo, 64 vetores 512-dimensionais (um para cada *pinna*).

Até aqui, todos os trabalhos com *PCA* aplicada em conjuntos de *HRTF*s agruparam todas as posições espaciais em um aglomerado de dados. No presente trabalho uma abordagem diferente foi explorada: aplicar *PCA* em cada posição espacial em separado. Dessa forma, são ressaltadas quantitativamente as diferenças do processamento que cada *pinna* faz em uma mesma posição espacial, visando a relacionar essa medida com dimensões antropométricas dos *pinnae*. Essa abordagem também pode auxiliar no caso de que, dependendo da posição espacial, certas dimensões antropométricas sejam mais importantes do que outras na conformação espectral da *HRTF*.

Seja P_j^i a matriz formada pelos vetores *PRTF* (vetores coluna) correspondentes ao *pinna* i e ao ângulo de azimuth j , conforme Equação (4.8):

$$P_j^i = \begin{bmatrix} p_{j1}^i & p_{j2}^i & \dots & p_{jN}^i \end{bmatrix} \quad (4.8)$$

onde p_{jn}^i é o vetor *PRTF* correspondente ao *pinna* i , ângulo de azimuth j , e ângulo de elevação n , com $n = 1, 2, \dots, N$. No caso máximo: $i = 1, 2, \dots, 90$ (90 *pinnae*, 45 indivíduos); $N = 50$ (50 ângulos de elevação) e $j = 1, 2, \dots, 25$ (25 ângulos de azimuth). A matriz de dados de entrada para cada processo *PCA* é descrita pela Equação (4.9):

$$E_{jn} = \begin{bmatrix} p_{jn}^1 & p_{jn}^2 & \dots & p_{jn}^L \end{bmatrix} \quad (4.9)$$

onde L é o número de vetores *PRTF* considerados no cálculo e previamente corrigidos pela média, isto é, previamente é calculada e subtraída de cada vetor, a média simples do grupo de vetores. No caso em estudo: $L = 64$. Finalmente, teremos $N \times J$ processos *PCA*, onde J é o número de ângulos de azimuth modelados, isto é $j=1,2,\dots,J$. Em princípio $N = 9$, abrangendo os ângulos de elevação desde -45° até 90° (hemisfério anterior) marcados na Figura 4.8. Por outro lado, no presente trabalho J será igual a 1, considerando apenas o plano médio pelas razões expostas na seção 4.4. Segundo trabalhos anteriores com *PCA* em *HRTFs*, é comum o primeiro componente principal explicar, na média, 90% da variabilidade das *HRTFs*, e os sete primeiros componentes principais explicar 95% dessa variação [45] [69]. No entanto, o caso em estudo apresentou resultados diferentes. Os três primeiros componentes principais explicaram 22,9%, 17,96% e 13,6% da variância total respectivamente. Os primeiros dez componentes principais explicaram 92,5% e os vinte primeiros 99,28% da variância total. A fim de atingir reconstrução quase perfeita, foram considerados no processo os vinte primeiros componentes principais. A percentagem de explicação da variância vs. Número de componentes principais é mostrado na Figura 4.18.

Então, é obtida a matriz de componentes principais considerados (20) e a matriz de pesos dos componentes principais considerados, dadas respectivamente, pelas Equações (4.10) e (4.11):

$$C_{jn} = \begin{bmatrix} c1_{jn} & c2_{jn} & \dots & c20_{jn} \end{bmatrix} \quad (4.10)$$

$$W_{jn} = \begin{bmatrix} w_{jn}^1 & w_{jn}^2 & \dots & w_{jn}^{64} \end{bmatrix} \quad (4.11)$$

onde $c1_{jn}, c2_{jn}, \dots, c20_{jn}$ são os vetores coluna base ou componentes principais (CP) do processo *PCA* definido pelos ângulos de azimuth j e de elevação n ; e w_{jn}^i é o vetor de pesos dos componentes principais correspondente ao *pinna* i , com $i=1,2,\dots,64$ no caso de estudo. Neste caso o vetor w_{jn}^i tem vinte dimensões, por ser considerados apenas vinte CPs. Portanto,

todos os CPs terão dimensão 512, e cada vetor w terá dimensão 20. Assim, para recuperar os vetores *PRTFs*, com uma distorção menor que 1% e corrigidos pela média, basta aplicar a Equação (4.12):

$$\hat{E}_{jn} = C_{jn} \cdot W_{jn} \quad (4.12)$$

Onde $\hat{E}_{jn}$ é a matriz recuperada (estimada) com perda de informação causada pela redução do dimensionamento dos dados, correspondente à matriz E_{jn} .

Figura 4.18. *PCA* dos 64 vetores *PRTF* extraídos no caso em estudo

A *PCA* foi aplicada com a magnitude dos vetores em escala linear (reconversão da escala em dB) e com a frequência em escala linear. Foram testados processos *PCA* no mesmo grupo de vetores reformulados com escala logarítmica e envolta (*wrapped*) para frequência e escala em dB para magnitude. Além disso, foram aplicadas diversas técnicas de modelagem para suavizar a resposta em magnitude: modelos *AR*, *AR* envolta (*wrapped AR*) e *ARMA* (*Auto-Regressive Moving Average*), e a resposta processada por filtros gaussianos com fator de

qualidade, Q , em 8, 10 e 12. A idéia por trás da aplicação desses processamentos foi conseguir uma maior explicação da variância com um número menor de CPs. Em geral, essas tentativas não representaram uma diferença importante com o caso original (escalas lineares em magnitude e frequência sem processamento adicional).

No final desse estágio, foi encontrada a segunda representação reduzida e particular de cada *PRTF* (e, portanto, de cada *HRTF*): os pesos dos componentes principais a partir da *PCA* aplicada no grupo de vetores *PRTF* correspondentes à mesma posição espacial em *pinnae* diferentes. No presente trabalho, tais pesos dos componentes principais são definidos como PCP. Assim, a partir das duas representações reduzidas e particulares das *PRTFs* extraídas são definidos dois grupos de vetores adicionais: os “vetores FCA” e os “vetores PCP”. Cada vetor FCA, bem como cada vetor PCP, corresponde a um *pinna* diferente e constituem a representação espectral dos *pinnae*.

4.8 Estágio de Mapeamento linear

O presente estágio visa a mapear uma representação reduzida da antropometria de cada *pinna* no espaço vetorial FCA e PCP em separado, isto é, na representação espectral dos *pinnae*. A Figura 4.19 mostra a idéia de maneira simplificada. Mesmo que os dados contidos no vetor FCA tenham dado origem aos vetores PCP (os vetores *PRTF*, ingressantes no processo *PCA* que deu origem aos vetores PCP, foram aproximados com a contribuição dos contornos de antirressonâncias ao redor das FCAs achadas), em geral, essas duas representações consideram-se independentes. Na sequência se faz uma revisão dos dados antropométricos registrados na base de dados *CIPIC* e das tentativas de mapeamento feitas até agora.

4.8.1 Antropometria do *pinna*

A Figura 3.3 apresenta as dez dimensões antropométricas do *pinna* registradas na base de dados *CIPIC*. Naquela seção, o método de cotejamento com base de dados para personalização da *HRTF* foi descrito. As dimensões registradas são novamente enumeradas na Tabela 4.3.

Figura 4.19. Mapeamento antropométrico - espectral do *pinna*.

Tabela 4.3. Dimensões do *pinna* registradas na base de dados *CIPIC*

Notação	Dimensão do <i>pinna</i>
d1	Altura da cavidade da concha
d2	Altura da cymba da concha
d3	Largura da cavidade da concha
d4	Altura da fossa
d5	Altura do <i>pinna</i>
d6	Largura do <i>pinna</i>
d7	Largura inter-tragal
d8	Profundidade da concha
d9	Ângulo de rotação do <i>pinna</i>
d10	Ângulo de vôo do <i>pinna</i>

A matriz D é a matriz de dimensões antropométricas, conforme Equação (4.13):

$$D = \begin{bmatrix} d^1 & d^2 & \dots & d^L \end{bmatrix} \quad (4.13)$$

onde d^i é o vetor coluna de dez dimensões correspondente às dimensões antropométricas do *pinna* i , segundo Equação (4.14):

$$d^T = [d1 \quad d2 \quad \dots \quad d10] \quad (4.14)$$

No caso em estudo $L = 64$. Portanto, a matriz D é de 10×64 dimensões. Por outro lado, para cada posição espacial temos a matriz W_{jn} de $R \times L$ dimensões da Equação (4.7), onde R é o número de componentes principais considerados (20). A idéia do presente método é relacionar a informação dessas duas matrizes, mas primeiro, a matriz D deve ser reordenada e depurada.

4.8.2 Correlação entre dados antropométricos e espectrais do *pinna*.

4.8.2.1 Correlação Individual

A partir de trabalhos anteriores [40] se sabe que a estrutura fundamental na conformação das antirressonâncias da *HRTF* é a cavidade principal ou concha. Segundo esse trabalho, duas das principais antirressonâncias têm origem nos efeitos de interferência originados na concha. Por tanto, a frequência onde essas interferências aparecem depende das dimensões da cavidade.

A primeira hipótese usada nesta parte do trabalho é que as dimensões da concha deveriam apresentar maior correlação com as características espectrais da *PRTF*, no mínimo, com as FCAs. Neste caso, como se pode ver na Figura 3.3, essas dimensões são $d1$, $d2$, $d3$ e $d7$. Na Tabela 4.4 é apresentado o coeficiente de correlação entre cada PCP e seu melhor preditor dentre as dimensões antropométricas.

Como observado na tabela, $d4$ aparece quatro vezes, $d10$ e $d6$ três vezes e $d3$, $d1$ e $d7$ apenas duas vezes cada. No entanto, esse não é um fator confiável para julgar a importância das dimensões. É necessária uma métrica global que considere a significância de cada PCP, isto é, o autovalor relativo a esse CP. Assim, é introduzido o índice de importância que leva em consideração o autovalor associado a cada PCP. Esse índice é descrito pela Equação (4.15):

$$I_k = \frac{\sum_{r=1}^{20} \lambda_r |\rho_{rk}|}{\sum_{r=1}^{512} \lambda_r}, 1 \leq k \leq 10 \quad (4.15)$$

Onde λ_r é o autovalor associado ao r -ésimo CP, e ρ_{rk} é o coeficiente de correlação entre o r -ésimo PCP e a dimensão do *pinna* denotada por k , com $1 < k < 10$, dadas as dez dimensões registradas. Esta vez, a importância de cada dimensão é computada ao longo dos vinte PCPs considerados para modelar as *PRTFs*. A Tabela 4.5 apresenta os resultados para tal cálculo.

Agora, as dimensões mais importantes são d1, d5 e d9. Apenas d1 é relacionada com a concha. Além disso, da Tabela 4.5, mesmo que os coeficientes de correlação não sejam tão baixos para os primeiros quatro PCPs (os quais explicam o 65% da variância), eles estão bem longe de ser confiáveis para realizar estimações de novas PCPs através de regressão linear. A solução aparente seria procurar por novos parâmetros antropométricos do *pinna* que o representem melhor.

Tabela 4.4. PCPs e seus melhores preditores antropométricos

PCP	Melhor Preditor	Coefficiente de Correlação ρ
w1	d5	-0.35
w2	d4	-0.43
w3	d3	-0.31
w4	d10	-0.35
w5	d4	0.15
w6	d1	-0.22
w7	d7	0.31
w8	d3	0.36
w9	d2	0.34
w10	d6	-0.25
w11	d4	0.32
w12	d7	0.19
w13	d1	-0.31
w14	d10	-0.17
w15	d4	0.28
w16	d10	-0.21
w17	d6	0.32
w18	d6	0.17
w19	d9	0.18
w20	d8	-0.29

Tabela 4.5 Índice de importância das dimensões antropométricas

Notação	Índice de Importância
d1	0.18
d2	0.12
d3	0.15
d4	0.14
d5	0.19
d6	0.11
d7	0.16
d8	0.13
d9	0.19
d10	0.11

4.8.2.2 Derivação de novos parâmetros

É bem provável que dimensões simples como as registradas na base de dados *CIPIC* não representem com exatidão as complexas estruturas anatômicas do *pinna*. Um outro problema importante é a falta de um ponto de referência para as dimensões, de maneira que uma coordenada cartesiana ou esférica possa representá-las. Isto foi feito em [15] com sucesso. No entanto, a base de dados de [15] não é publica.

Baseados na hipótese de que parâmetros bidimensionais ou até tridimensionais, porém sem ponto de referência, possam representar melhor cada *pinna*, e cada concha, foram derivados doze parâmetros adicionais, baseados nas dimensões existentes. Eles são enumerados na Tabela 4.6. Um exemplo deles, d13, é apresentado na Figura 4.20.

Mesmo que alguns parâmetros não tenham muito sentido físico como d15 ou d17, em muitos casos, eles caracterizam melhor o *pinna* que representam do que as dimensões registradas na base de dados. O índice de importância definido na Equação (4.11) e calculado para as dez dimensões originais é calculado também para os novos parâmetros. A Tabela 4.7 apresenta o resultado desse cálculo. A Tabela 4.8 apresenta os coeficientes de correlação dos dez primeiros PCPs com seus respectivos melhores preditores.

Em termos numéricos, só foi conseguida uma pequena melhoria. Porém, agora d13 e d12 (melhores preditores dos dois PCPs mais importantes) estão mais bem correlacionados com a concha. Além disso, também calculamos os coeficientes de correlação entre os parâmetros antropométricos e as FCAs. Em princípio, dado que as FCAs estão mais diretamente relacionadas às características espectrais da *HRTF*, elas deveriam estar mais bem

correlacionadas com os parâmetros antropométricos. Os resultados são apresentados na Tabela 4.9.

Mesmo que os coeficientes de correlação das FCAs sejam muito melhores do que os dos PCPs, ainda são baixos para realizar regressão linear no cálculo de FCAs de novos indivíduos. A seguinte tentativa será somar o aporte de todas as dimensões através de análise multivariável.

Tabela 4.6. Novos parâmetros antropométricos.

Notação	Novos parâmetros antropométricos
d11	$d1 + d2$
d12	$d1 + d2 + d4$
d13	$(d1 + d2).d3$
d14	$(d1+d2).d3.d8$
d15	$(d1+d2).d3.d10$
d16	$(d1+d2).d7.d8$
d17	$(d1+d2).d7. d10$
d18	$d1.d3$
d19	$d5.d6$
d20	$d5.d6.d8$
d21	$d5.d6. d10$
d22	$d4.d6$

Figura 4.20. Exemplo de novo parâmetro antropométrico do *pinna*

Tabela 4.7. Índice de importância dos novos parâmetros antropométricos

Notação	Índice de Importância
d11	0.2
d12	0.17
d13	0.2
d14	0.2
d15	0.18
d16	0.14
d17	0.18
d18	0.2
d19	0.16
d20	0.16
d21	0.16
d22	0.13

Tabela 4.8. PCPs e seus melhores preditores antropométricos considerando os novos parâmetros.

PCP	Melhor Preditor	Coefficiente de Correlação ρ
w1	d13	-0.37
w2	d12	-0.45
w3	d18	-0.31
w4	d10	-0.35
w5	d4	0.15
w6	d1	-0.22
w7	d16	0.34
W8	d13	0.4
w9	d2	0.35
w10	d6	-0.25

Tabela 4.9. FCA e seus melhores preditores antropométricos.

FCA	Melhor Preditor	Coefficiente de Correlação ρ
A	d15	-0.46
A2	d8	-0.76
B	d14	-0.57
C	d11	-0.44

4.8.2.3 Regressão linear multivariável

É necessário redefinir a matriz D, formulada na Equação (4.9). Ela deve ser estendida para incluir os doze novos parâmetros de tal forma a se converter numa matriz de 22 x 64 dimensões. No entanto, realizados os cálculos para relacioná-la com a matriz W, foi

descoberto que seria necessário eliminar as dimensões d2 e d4 da matriz. A razão é que elas não podem coexistir como linhas na mesma matriz com d1, d11 e d12 já que produziriam dependência linear, tornando a matriz T (resultado do relacionamento entre W e D), a ser definida na Equação (4.12), singular. Além disso, essas dimensões são as menos importantes daquele grupo (d1, d2, d4, d11 e d12) como visto nas Tabelas 4.5 e 4.8. A partir dessas tabelas, se sabe que os parâmetros mais importantes agora são d18, d13 e d14, os quais estão bem relacionados com a concha. Finalmente, a matriz D, no caso em estudo, fica com dimensão 21 x 64: uma fila de uns (1) é adicionada à matriz para introduzir um termo constante no modelo. A regressão linear multivariável, então, é realizada resolvendo a Equação (4.16).

$$D^T \cdot T + \mathcal{E} = W^T \quad (4.16)$$

Sendo que a matriz D^T é de 64 x 21 dimensões e a matriz W^T é de 64 x 20 dimensões, a matriz T a ser calculada deve ser de 21 x 20 dimensões, com a matriz residual $\mathcal{E}$ de 64 x 21 dimensões. Cada coluna de T é calculada resolvendo o problema dos mínimos quadrados para a matriz D^T e cada coluna de W^T , conforme Equação (4.17):

$$D^T \cdot t_r + \mathcal{E}_r = w_r \quad (4.17)$$

Onde t_r e $\mathcal{E}_r$ são as r-ésimas colunas das matrizes T e $\mathcal{E}$ respectivamente, e w_r a r-ésima coluna da matriz W^T que corresponde ao vetor formado pelo r-ésimo PCP de todas as *PRTFs*. Cada problema dos mínimos quadrados é resolvido por decomposição QR. Assim, finalmente, achada a matriz T, é possível estimar a matriz W e, portanto os PCPs, através da Equação (4.18):

$$D^T \cdot T = \hat{W}^T \quad (4.18)$$

Onde $\hat{W}^T$ é a matriz estimada dos PCPs. A Tabela 4.10 apresenta os coeficientes de correlação entre os PCPs originais e os PCPs estimados.

Tabela 4.10. Coeficientes de correlação entre PCPs originais e PCPs estimados através dos mínimos quadrados.

PCP	PCP estimado	Coeficiente de Correlação ρ
w1	$\hat{w}1$	0.7237
w2	$\hat{w}2$	0.7118
w3	$\hat{w}3$	0.6064
w4	$\hat{w}4$	0.6975
w5	$\hat{w}5$	0.6005
w6	$\hat{w}6$	0.5969
w7	$\hat{w}7$	0.5738
w8	$\hat{w}8$	0.7247
w9	$\hat{w}9$	0.6612
w10	$\hat{w}10$	0.6150
w11	$\hat{w}11$	0.6358
w12	$\hat{w}12$	0.5445
w13	$\hat{w}13$	0.7565
w14	$\hat{w}14$	0.4795
w15	$\hat{w}15$	0.6230
w16	$\hat{w}16$	0.6915
w17	$\hat{w}17$	0.5982
w18	$\hat{w}18$	0.5482
w19	$\hat{w}19$	0.5220
w20	$\hat{w}20$	0.5851

Um processo semelhante é realizado para as FCAs. A partir das observações feitas nos padrões e dos estudos associados às antirressonâncias do *pinna*, foi decidido estimar apenas as FCAs A e B (elas aparecem no 99% das *HRTFs* observadas). Neste caso, dada a natureza escalar das FCA, é estimado apenas um vetor, correspondente à matriz T no caso dos PCPs. Esses vetores, de dimensão 21 no caso em estudo, são chamados u_A e u_B para A e B respectivamente. A Tabela 4.11 apresenta os coeficientes de correlação entre as FCA originais e as FCAs estimadas. Como observado nas Tabelas 4.10 e 4.11, a melhoria atingida é considerável.

Tabela 4.11. Coeficientes de correlação entre FCAs originais e FCAs estimados através dos mínimos quadrados.

FCA	FCA estimada	Coeficiente de Correlação ρ
A	$\hat{A}$	0.74
B	$\hat{B}$	0.81

4.8.3 Algoritmo de movimentação de contornos de antirressonâncias

Tendo calculada a matriz de estimação para os PCPs é possível estimar a *PRTF* de um novo indivíduo através do processo de reconstrução da *PCA*. Contudo, sendo que a estimação das FCAs apresenta melhores coeficientes de correlação, e em teoria a variação das FCAs deveria ter maior impacto perceptual do que a variação dos PCPs (isso poderia ser comprovado experimentalmente como será visto no capítulo final), foi ideado um método para misturar essas duas estimações. A idéia é estimar inicialmente a *PRTF* através dos PCPs e ajustar as FCAs dessa *PRTF* e os contornos respectivos através das FCAs estimadas. O algoritmo pode-se resumir nos seguintes passos (ver Figura 4.21):

1. Encontrar o mínimo local na magnitude da *PRTF* estimada (vetor *PRTF*) mais próximo a cada FCA estimada: A e B. É usado o algoritmo de busca de alterações de inclinação usado na busca dos mínimos locais na função de atraso de grupo da seção 4.6.2.
2. Calcular a distância entre a FCA estimada e o mínimo local achado: x
3. Se o mínimo local estiver à direita da FCA estimada, calcular x pontos à esquerda da FCA estimada e x pontos à direita do mínimo local, e marcar os limites do intervalo de ação do algoritmo: 1 e 2 respectivamente na Figura 4.18.
4. Converter os $2x$ pontos entre 1 e o mínimo local em x pontos através de uma simples operação de dizimação.
5. Converter os x pontos entre o mínimo local e 2 em $2x$ pontos através de uma simples operação de interpolação linear.
6. Se o mínimo local estiver à esquerda da FCA estimada, realizar o processo contrário.

Figura 4.21. Algoritmo de movimentação de contornos de antirressonâncias

O algoritmo de movimentação é aplicado nas FCAs A e B. Na Figura 4.22 são mostradas algumas *PRTFs* estimadas e ajustadas através dos processos descritos anteriormente (linhas tracejadas) junto com *PRTFs* originais (linhas contínuas). Em termos gerais, as FCAs A e B são corretamente estimadas. Embora a amplitude geral das *PRTFs* estimadas e ajustadas não apresente exatidão, sobretudo em altas frequências, a envoltória em médias e baixas frequências é consideravelmente aceitável.

Figura 4.22. *PRTFs* originais (linhas contínuas) e *PRTFs* estimadas e ajustadas (linhas tracejadas).

4.9 Estágio de Estimação de novas *PRTFs*

Para estimar a *PRTF* de um novo indivíduo, será necessário medir as mesmas dez dimensões antropométricas registradas na base de dados *CIPIC*. Isto pode ser feito tomando 2 fotografias do cada *pinna*, em concordância com a Figura 3.3. e medir manualmente as distâncias envolvidas. Na sequência, o resto de parâmetros antropométricos será calculado segundo as fórmulas da Tabela 4.6. A seguir, a matriz *D* será formada segundo a seção 4.2.3 e os PCPs estimados através da Equação (4.13) (matriz *T*). Da mesma forma, as FCAs serão estimadas para os contornos A e B (vetores u_A e u_B). Finalmente o algoritmo de movimentação de contornos de antirressonâncias é aplicado para ajustar as posições das FCAs.

4.10 Avaliação de Desempenho

O cenário ideal para testar o desempenho do processo de estimação de *PRTFs* é o seguinte:

1. Medir a *HRTF* de um novo indivíduo e extrair a *PRTF* através dos métodos apresentados na seção 4.6.
2. Estimar a *PRTF* através do processo apresentado na seção 4.9
3. Comparar as duas *PRTFs* em base à diferença em dB ao longo do intervalo de frequência de interesse.
4. Comparar as duas *PRTFs* através de testes subjetivos de localização de fonte sonora aplicados no indivíduo de teste.

A proposta de testes auditivos de localização de fonte sonora será descrita no seguinte capítulo. No entanto, uma medida simples da estimação nos “dados de treino” é apresentada na sequência: o índice de estimação descrito pela Equação (4.19):

$$E = \frac{\sum_{r=1}^{20} \lambda_r |\rho_r|}{\sum_{r=1}^{512} \lambda_r}, \quad (4.19)$$

Onde ρ_r é o coeficiente de correlação entre o r-ésimo PCP e sua versão estimada, e λ_r é o autovalor associado a esse PCP. A Figura 4.23 apresenta o índice de estimação, E , da Equação (4.2) vs. o número de CPs usado na modelagem das *PRTFs*, no caso em estudo. Para vinte CPs atingimos 66% de estimação, o máximo possível. Esse índice calcula a percentagem de estimação atingida pelo método, levando em conta a importância do PCP (autovalor) e o grau de semelhança entre os PCP e seu valor estimado (coeficiente de correlação).

Figura 4.23. Avaliação de desempenho do método para o grupo de vetores de treino.

4.11 Sumário

O presente capítulo apresentou todos os métodos, algoritmos e observações originais feitas no contexto de estimação de novas *PRTFs*. No início foi sumarizado o trabalho feito em processamento bidimensional de imagens da *HRTF*. Esse trabalho está bem resumido em [65]. O objetivo atingido foi um método para observação de antirressonâncias da *HRTF* e auxílio na classificação em padrões. Esses padrões são uma das principais contribuições do mestrado, pelo menos no plano médio. No futuro, pretende-se estender essa classificação ao espaço todo. Além disso, também foi elaborado um método bidimensional para extração da *PRTF*, porém, ele não foi usado nos seguintes estágios devido às inexatidões observadas. Na sequência foram apresentadas as hipóteses do trabalho e o método de maneira geral. A seguir cada estágio do método foi descrito. O método está baseado no algoritmo apresentado em [67] junto com diversas contribuições originais, entre elas:

- O processo iterativo para encontrar o limiar do atraso de grupo auxiliado pela classificação em padrões.
- O algoritmo para construir uma *PRTF* estimada a partir do modelo *AR*, as *FCAs* e a *HRTF* original.
- O algoritmo de movimentação de contornos de antirressonâncias.
- A derivação de novos parâmetros antropométricos a partir dos já existentes.
- O relacionamento linear (através da resolução do problema dos mínimos quadrados) entre antropometria e espectro.

Todos esses métodos e algoritmos já foram apresentados com sucesso em conferências internacionais [70] [71] [72].

Finalmente é descrito brevemente o processo de estimação de uma nova *PRTF* e a métrica introduzida para avaliar o desempenho do método nos “dados de treino”: o índice de estimação.

Capítulo 5

PROPOSTA DE TESTES AUDITIVOS DE LOCALIZAÇÃO DE FONTE SONORA

5.1 Introdução

Num sentido geral, testes auditivos de localização de fonte sonora são procedimentos estruturados e concebidos para avaliar a capacidade de localização espacial auditiva de uma pessoa, a partir do cálculo do erro médio de localização ou de outras métricas usadas. No entanto, testes auditivos de localização também são usados para avaliar o desempenho de localização de um sistema de reprodução acústica virtual (VAD). Neste caso, esses sistemas podem basear-se num modelo computacional (cognitivo, estatístico, matemático, etc.) a ser avaliado, ou num novo método de reprodução a ser testado. É importante salientar a diferenciação entre testes auditivos de localização de fontes sonoras reais e testes auditivos de modelos computacionais que simulam a emissão sonora real. No segundo caso, o teste é aplicado sobre a percepção que o ouvinte tem sobre o que ele pensa estar escutando, isto é sobre a imagem auditiva produzida na construção mental do espaço [73]. Neste contexto, na realização de um teste auditivo que seja confiável, é necessário fornecer ao ouvinte uma referência física, isto é, a oportunidade de comparar o estímulo sonoro real presente no espaço sonoro e o estímulo simulado presente só no espaço auditivo. Contudo, apesar dos avanços enormes em tecnologia binaural nos últimos vinte anos, ainda não é possível conseguir uma perfeita simulação, sendo possível apenas conseguir uma simulação que forneça ao ouvinte os indicadores suficientes para reconhecer de onde provém o estímulo simulado. Assim a reconstrução perfeita é ainda inalcançável [73]. O objetivo do presente capítulo é apresentar

sucintamente os principais componentes e considerações técnicas de testes de localização de fonte sonora para *VADs*, visando avaliar o modelo de personalização da *HRTF* proposto no Capítulo 4. Na segunda seção do presente capítulo são apresentados os principais conceitos relativos a *VADs*. Na sequência, detalhes importantes associados ao uso de fones de ouvido em reprodução binaural são detalhados. A seguir, são descritas as considerações gerais da implementação de um teste de localização de fonte sonora para *VADs*, incluídos seus principais componentes. Finalmente, é apresentado o sumário do capítulo.

5.2 Sistemas de reprodução acústica virtual (*VADs*)

Um sistema de reprodução acústica virtual é aquele capaz de produzir um ambiente aural imersivo entregue através de fones de ouvido ou arranjos especiais de alto-falantes. Nele, o campo acústico percebido pelo usuário é comparável (mas nunca igual) àquele produzido em campo livre (ou num recinto específico, se esse for o objetivo) por uma ou mais fontes sonoras localizadas em pontos arbitrários do espaço. Como referido no primeiro capítulo, o uso de arranjos de alto-falantes é comum em aplicações de realidade virtual e cinema onde o “ponto doce” requerido é maior [74]. Fones de ouvido é a tecnologia mais acessível e portátil para aplicações de pesquisa em neurociência, telecomunicações e entretenimento [6]. Essa tecnologia é chamada de binaural e visa reproduzir os sinais nos tímpanos do ouvinte produzidos pela fonte sonora em campo livre [75]. As vantagens dos fones de ouvido são que mantêm a separação dos dois canais e eliminam a influência do recinto onde é feita a reprodução. A desvantagem é o efeito de localização “dentro da cabeça”, inerente ao uso dos fones. Além disso, a reprodução e o “ponto doce” são estritamente individuais.

Pelo exposto nos capítulos anteriores, o requisito de alta fidelidade para reprodução binaural deve cumprir duas regras [6]:

1. A *HRTF* usada deve ser individual, ou pelo menos personalizada através de um método eficaz, isto é, cada usuário deve assistir à reprodução binaural com a própria *HRTF*, seja medida ou estimada aproximadamente através de processos computacionais.
2. Deve-se compensar a função de transferência dos fones de ouvido (*PTF* – *Headphone Transfer Function*), a qual descreve a transmissão entre a voltagem nos

terminais dos fones e o sinal medido na entrada do canal auditivo aberto. Segundo estudos da *PTF*, ela também depende do indivíduo e a sua caracterização é, às vezes, ainda mais complexa do que a da *HRTF*. A diferença é que a *PTF* é independente da direção [76], mas dependente da posição dos fones de ouvido no *pinna* [77].

É importante ressaltar que existem três vertentes em termos de alta fidelidade de *VADs*, e é claro que os requisitos do sistema (externalização, imersão e controle do posicionamento) dependem da aplicação, como analisado no Capítulo 2. Alguns *VADs* buscam a maior adequação e eficiência em termos de implementação, outros buscam o maior realismo possível, enquanto outros precisam da entrega da maior quantidade de informação espacial com finalidade de monitoramento. Assim, é importante diferenciar os três esquemas já mencionados ao longo da dissertação: o uso de *HRTFs* genéricas para uma grande população de ouvintes; o uso de *HRTF* personalizadas para ouvintes específicos através de dados antropométricos, acústicos ou psicoacústicos; e o uso de *HRTFs* individuais.

5.3 Modelo de transmissão de som na reprodução binaural: Fones de ouvido

A Figura 5.1 mostra o esquema típico da reprodução por fones de ouvido. A Figura 5.2 mostra o circuito equivalente de Thévenin desde a excitação acústica produzida pelos fones até o tímpano. A impedância da fonte sonora é representada por Z_{fone} e a pressão de Thévenin por P_1 . Normalmente essa pressão não existe fisicamente, mas se o canal auditivo estiver fechado, a pressão de Thévenin poder-se-ia medir na parte externa da fechadura. A fim de achar a pressão sonora na entrada do canal auditivo aberto, denotada por P_2 , a pressão de Thevenin é dividida pela impedância da fonte mais a impedância de entrada ao canal auditivo, denotada por $Z_{\text{canal auditivo}}$, como apresentado na Equação (5.1):

$$P_2 = P_1 \cdot \frac{Z_{\text{canal auditivo}}}{Z_{\text{canal auditivo}} + Z_{\text{fone}}} \quad (5.1)$$

Figura 5.1. Transmissão anatômica do som desde o fone de ouvido, passando através do *pinna*, até o tímpano.

Figura 5.2. Circuito acústico equivalente de Thévenin da transmissão anatômica do som desde fones de ouvido, através do *pinna*, até o tímpano.

O canal auditivo atua como um terminal acústico de duas portas, terminando na impedância do tímpano $Z_{\text{tímpano}}$, de pressão acústica P_3 . Segundo Hammershøi [78], gravações biaurais necessárias para obter empiricamente *HRTFs* e *PTFs*, podem ser feitas em qualquer ponto do canal auditivo, dado que a transmissão dentro dele é independente da direção. Nesse contexto, a pressão de Thevenin medida na entrada do canal auditivo fechado é a mais indicada para determinar *HRTFs* e *PTFs*, devido a:

- Ela contém toda a informação direcional.
- É a pressão com menor influência das variações entre indivíduos.
- O ponto de medição é mais facilmente acessível, além da maior segurança em relação ao posicionamento do microfone.

Assim, quando a gravação for feita na entrada do canal auditivo fechado, é possível reproduzir o sinal correto no tímpano através da compensação, ou filtragem inversa, definida pela Equação (5.2):

$$G = \frac{1}{M \cdot P_1 / E_{\text{fone}}} \cdot \frac{Z_{\text{canal auditivo}} + Z_{\text{fone}}}{Z_{\text{canal auditivo}} + Z_{\text{radiação}}} \quad (5.2)$$

onde M é a função de transferência voltagem/pressão do microfone de gravação e $Z_{\text{radiação}}$ é a impedância de radiação em campo livre vista pelo canal auditivo. O termo P_1 / E_{fone} é a função de transferência eletroacústica do fone de ouvido, ou *PTF*, medida na entrada do canal auditivo fechado. O termo que envolve as quatro impedâncias é a razão entre as divisões da pressão de Thévenin em dois casos:

- Com o *pinna* em campo livre.
- Com o fone de ouvido sobreposto no *pinna*.

Essa razão é chamada de Rázão de Divisão de Pressão (*PDR – Pressure Division Ratio*). Assim, a equação anterior pode ser escrita conforme a Equação (5.3):

$$G = \frac{1}{M.PTF}.PDR \quad (5.3)$$

Na situação ideal, *PDR* é 1. Esses fones são chamados de *FEC (Free-Air Equivalent Coupling)* [75]. Por outro lado, *M* pode ser aproximada à unidade enquanto o microfone tiver resposta em frequência plana na banda de interesse, sendo este o caso geral. Assim, a função a ser compensada se reduz à *PTF*. No entanto, a *PTF* deve ser medida individualmente, pois ela apresenta uma grande variabilidade entre indivíduos como comprovado na Figura 5.3. Ela mostra os resultados das medições de 40 *PTFs* apresentados por Møller em [75] para um grupo de 14 fones de ouvido comerciais. A linha sólida representa a *PTF* média dentre as 40 *PTFs*, e as região cinza representa o desvio padrão do grupo. É claro que cada fone produz um processamento diferente em cada *pinna*. Em frequências acima de 7 kHz a variância entre indivíduos é ainda maior o que distorceria com toda certeza as FCAs estimadas pelo modelo proposto no Capítulo 4, presentes de 5 kHz a 17 kHz.

5.4 Implementação do sistema de testes

O esquema geral do sistema de testes proposto é apresentado na Figura 5.4. Em geral, o teste inteiro está governado por dois elementos fundamentais que devem funcionar de maneira complementar: o software de espacialização e o procedimento de teste. Um depende do outro. Quem gerenciará os testes, ingressando os parâmetros no software de espacialização e, portanto, aplicando o procedimento de teste, deve ser uma pessoa com profundo conhecimento da tarefa a ser realizada.

Objetivo do trabalho é propor uma plataforma de testes auditivos de baixo custo. Como conhecido, esses sistemas se caracterizam por seu grande custo econômico por causa do hardware de áudio profissional (fones, placas de som, alto-falantes e caixas), do recinto onde devem ser realizados os testes e do rastreador de cabeça.

Na sequência detalharemos os pontos críticos do sistema de testes.

Figura 5.3. Média (linha preta) e desvio padrão calculados em grupo de 40 *PTFs* medido em 14 fones de ouvido comerciais. Figura obtida de [75].

5.4.1 Software de espacialização

A implementação do software de espacialização é bem flexível quanto à linguagem de programação, os aspectos ergonômicos e as interfaces de controle. As funcionalidades do software estão diretamente associadas aos parâmetros gerais que devem ser inseridos pelo operador dos testes:

- O tipo de sinal de teste (neste caso: ruído branco).

Figura 5.4. Esquema geral do sistema de testes de localização de fonte sonora.

- A duração de cada sinal (uma configuração adequada quando o sinal de teste é ruído branco é: entre três e cinco rajadas de 100ms com intervalos de silêncio de 100ms; ou três rajadas de 1s com intervalos de silêncio de 1s [12])
- A ordem em que serão tocados os sinais de teste, se eles forem mais de um. (há experimentos realizados com sinais de fala [4]).

- A posição espacial a ser testada e a ordem das posições espaciais a serem testadas caso haja mais de uma posição, podendo ser em ordem aleatória.
- A origem das *HRTFs* a serem testadas (*HRTF* genérica: *KEMAR*; *HRTF* personalizada: Método 1; *HRTF* personalizada: Método 2; ... *HRTF* personalizada: Método N; *HRTF* Individual, etc.).
- Dependendo da estrutura física do arranjo de alto-falantes, também será necessário configurar o movimento/rotação dos alto-falantes (o passo, a velocidade, etc.).
- Escolha dos fones de ouvido e, portanto, da *PTF* a ser compensada.
- Outros parâmetros de pré-processamento e pós-processamento opcionais: equalização, filtragem passa-altas, filtragem passa-baixas, mixagem de vários sinais de teste, etc.

O protótipo do software de espacialização foi implementado na plataforma MATLAB e depois migrado para a plataforma *Pure Data* (PD) ¹⁵ a fim de aperfeiçoar seu desempenho em tempo real. A janela inicial é mostrada na Figura 5.5.

5.4.2 Reprodução de fontes virtuais: Fones de ouvido

O fator mais crítico é achar empiricamente a *PTF* individual de fones de ouvido específicos da maneira mais econômica possível, a fim de implementar um *VAD HI FI*. Como mencionado na seção 5.3, esse requerimento é imprescindível visando avaliar objetivamente o desempenho do modelo proposto da *HRTF*. Se a *PTF* não for equalizada individualmente, ou pelo menos aproximada individualmente, ela de fato introduzirá novas antirressonâncias em alta frequência no espectro do sinal entregue ao usuário, como relatado por Møller em [75]. Dado que o modelo proposto se baseia principalmente na estimação da posição em frequência das antirressonâncias como fator fundamental na localização vertical,

¹⁵ ¹⁵ Informação sobre o pacote *Pure Data* pode ser encontrada no sítio web <http://www.pd.org>

uma *PTF* genérica desvirtuaria totalmente o desempenho do modelo. A seguir, são propostas quatro soluções ao problema:

1. Usar fones de ouvido intra-auriculares, tais como os Etymotic ER-2 usados em [6]. Eles são posicionados na entrada do canal auditivo, o que elimina a necessidade de compensação individual, pois o *pinna* exerceria pouca influência sobre o sinal. Além disso, segundo os fabricantes, os ER-2 têm resposta em frequência plana, eliminando a necessidade de qualquer equalização. O grande problema é o custo dos fones ER-2 (825 US\$), fabricados especialmente para propósitos de pesquisa. Soluções alternativas são os fones Etymotic ER-1 (540 US\$) ou Shure E5c (500 US\$). É claro que a diferença de preço envolve diferença em qualidade de reprodução. No entanto, seja qual for o fone usado, a condição de intra-auricular reduz o problema a uma equalização genérica só, isto é, não individual.
2. Usar fones de ouvido Circum-Aurais regulares tais como o Sennheiser HD 250, avaliados por Møller [75]. Obviamente, essa solução requereria equalização individual. Para tanto é necessário também um microfone miniatura tal como o Sennheiser KE 4-211-2, usado por Møller em [75] e Carlile em [6]. Além disso, um processo de calibração estruturado e minucioso é imprescindível, a fim de compensar corretamente o sinal entregue pelo fone para cada indivíduo. Em [75] são avaliados 14 fones de ouvido, determinando quais são os de resposta mais constante entre indivíduos, quais são os mais facilmente compensados (devido a sua resposta mais plana) e quais são os que têm maiores propriedades *FEC*. A partir desse estudo é possível escolher o fone mais indicado. Um bom candidato é o fone de ouvido Circum-Aural Stax SR Lambda profissional, devido a sua resposta em frequência razoavelmente plana e as suas propriedades razoavelmente *FEC*. Porém, ele apresenta grande variação entre indivíduos. Por outro lado, o fone AKG Acoustics K500 apresenta menor variação entre indivíduos, mas o problema dele é que não possui propriedades *FEC* nem é facilmente compensável devido às ressonâncias e antirressonâncias de alto fator de qualidade *Q*. No entanto, o custo dos aparelhos é ainda mais alto do que na solução 1, com o microfone miniatura custando aproximadamente 1500 US\$, e o fone de ouvido por volta dos 200 US\$.

Figura 5.5. Patch em PD para convolução biaural.

3. A terceira solução usa uma compensação aproximada através de uma *PTF* média. Técnicas para esse processamento são apresentadas em [76] e [75]. Obviamente essa solução não é a ideal, mas é a mais econômica. Neste caso, seria usado um fone de ouvido com baixa variação entre indivíduos, como por exemplo, o AKG Acoustics K500. O fato de esse fone não ter propriedades *FEC* nem facilidade de equalização seria compensado pelo fato que a distorção dependente da frequência introduzida pelo *PDR* diferente de 0 e pela compensação inexata, ocorreria em todos os indivíduos por igual. O custo monetário dessa solução seria o custo dos fones, isto é aproximadamente 200 US\$. Porém, os sinais (*PTF*) deveriam ser solicitados ao Laboratório de Acústica de Universidade de Aalborg [75].

4. Uma outra solução alternativa é não usar compensação nenhuma para fones de ouvido, como feito em [12] (Sennheiser HD 470). Neste caso, seria avaliado só um comportamento geral do modelo, sem resultados confiáveis.

No entanto, segundo pesquisas sobre variabilidade da *PTF* para um mesmo indivíduo com fones de ouvido circum-aurais e supra-aurais [77], essa função também depende da posição em que o fone de ouvido é colocado no *pinna*. Nessa pesquisa, os autores concluem que se o requerimento for um *VAD HI FI*, não é confiável usar uma única filtragem inversa para compensação dos fones de ouvido, pois ela depende em grande medida da posição dos fones. Ainda usando uma média da *PTF* para varias posições, a filtragem inversa não é confiável. Mas eles não propõem solução nenhuma.

Como conclusão, a solução mais atraente seria o uso de fones intra-auriculares (solução 1), sendo o problema a superar o custo dos aparelhos.

5.4.3 Reprodução de fontes reais: Alto-falantes

Uma vez resolvido o problema da reprodução biaural, deve ser direcionado o problema do *set-up* de alto-falantes. A melhor forma de testar o desempenho de localização é comparar diretamente os pareceres subjetivos do indivíduo ante uma fonte sonora real e ante a fonte sonora virtual reproduzida através dos fones de ouvido [6]. Assim, temos duas alternativas para reproduzir o som da fonte sonora real em posições espaciais determinadas:

- Usar vários alto-falantes com uma estrutura fixa, como apresentado na Figura 5.6a.
- Usar 1 alto-falante e uma estrutura de rotação vertical para ele, como ilustrado na Figura 5.6b.

A desvantagem da primeira solução é a necessidade de hardware extra a fim de direcionar vários alto-falantes: uma placa de som multicanal. A desvantagem da segunda solução é a dificuldade na construção da estrutura giratória e o controle dela.

O objetivo dessas estruturas é reproduzir o som da fonte real no ponto espacial que depois será reproduzido pelo modelo na síntese da fonte virtual. A única especificação técnica

iniludível das estruturas é que a distância entre os alto-falantes e o centro do eixo interaural (ponto de origem do sistema de coordenadas da base de dados *CIPIC*) deve ser 1m.

O modelo de *HRTF* desenvolvido funcionará, por agora, no plano médio, nos ângulos de elevação de -45° , -33.125° , -22.5° , -11.125° , 0° , 11.125° , 22.5° , 33.125° e 45° . Portanto, deveram ser consideradas pelo menos essas posições espaciais ao momento de desenhar e construir a estrutura de alto-falantes. Se for possível, e dependendo da dificuldade da construção, a estrutura contemplará mais posições intermediárias, a fim de testar métodos de interpolação de *HRTFs*.

Figura 5.6. (a) Estrutura fixa de alto-falantes; (b) estrutura giratória de alto-falantes.

5.4.4 Recinto de reprodução

Um outro problema a ser resolvido é a resposta do recinto onde serão realizados os testes com os alto-falantes. O ideal seria realizar o teste numa câmara anecóica onde não existisse reverberação. A solução imediata seria agregar um módulo de reverberação à cadeia de reprodução da fonte virtual. O módulo deveria reproduzir as características reverberantes do recinto onde é realizado o teste, isto é, onde são emitidos os sons reais. Para tanto, seria

necessária informação sobre as dimensões do recinto e sobre as características exatas de absorção e espalhamento dos materiais. Além disso, um software de simulação acústica seria necessário para conseguir a resposta impulsiva requerida a partir desses dados levando em conta as posições das fontes e do indivíduo de teste (centro da estrutura de alto-falantes). Uma opção boa, substituindo a modelagem acústica da sala, é medir a resposta impulsiva do recinto nas posições adequadas e adicionar o módulo à cadeia de reprodução do sistema virtual. No entanto, a consideração do recinto adiciona outro ponto de incerteza na cadeia de reprodução.

5.4.5 Procedimento dos testes auditivos

A idéia dos testes é a seguinte: o indivíduo deve ouvir o sinal de teste (ruído branco) emitido pelo alto-falante sem informação visual da posição da fonte, dar um parecer subjetivo da localização da fonte, e depois fazer o mesmo, mas com o sinal entregue através dos fones de ouvido, correspondente à mesma posição espacial, e sintetizado de duas formas:

- Através da *HRTF* personalizada através do modelo desenvolvido.
- Através de uma *HRTF* genérica.

O indivíduo apontará com a cabeça na direção de onde ele acha que o som provém, isto é, a posição da fonte sonora, determinando assim o que chamaremos “posição subjetiva”. A posição subjetiva será registrada através de um rastreador de cabeça (*HT – Head Tracker*) [6] [13]. Assim, serão registradas três posições subjetivas para cada posição espacial, uma da emissão real através do alto falante, uma da emissão virtual através dos fones usando o modelo desenvolvido, e outra da emissão virtual através dos fones usando a *HRTF* genérica.

Dado que os testes serão realizados apenas no plano médio, a medida importante registrada pelo *HT* será o ângulo de elevação, porém, o ângulo de azimuth também será guardado, podendo este ser subjetivamente diferente de zero. Consideremos ϕ_{real}^n o ângulo de elevação registrado para a posição subjetiva correspondente à emissão real através do alto-falante na posição real determinada pelo ângulo de elevação de índice n . Do mesmo jeito, ϕ_{per}^n o ângulo de elevação registrado correspondente à emissão virtual através dos fones usando a *HRTF* personalizada, e ϕ_{gen}^n o ângulo de elevação registrado correspondente à

emissão virtual através dos fones usando a *HRTF* genérica. Calcularemos um índice geral (RMS) de desempenho de localização de fonte sonora para ambos as *HRTFs*, onde a Equação (5.4):

$$IDL_{per} = \frac{\sqrt{\sum_{n=1}^N (\phi_{real}^n - \phi_{per}^n)^2}}{N} \quad (5.4)$$

corresponderia ao índice para o modelo desenvolvido e a Equação (5.5):

$$IDL_{gen} = \frac{\sqrt{\sum_{n=1}^N (\phi_{real}^n - \phi_{gen}^n)^2}}{N} \quad (5.5)$$

corresponderia ao índice para a *HRTF* genérica. Neste equacionamento, N é o número total das posições testadas, sendo pelo menos 2 e como máximo 9. A comparação desses índices determinará o resultado final da pesquisa.

Resumindo, o procedimento do teste será o seguinte:

1. O indivíduo, de olhos fechados, senta na cadeira no centro da estrutura de alto-falantes.
2. O alto-falante emite o ruído branco para uma posição aleatória dentre as posições possíveis (entre 2 e 9).
3. O indivíduo aponta a cabeça na direção de onde ele acha que o som provém. O *HT* registra essa posição.
4. Os pontos 2. e 3. são repetidos para todas as posições possíveis da estrutura de alto-falantes (entre 2 e 9).

5. São colocados os fones de ouvido no indivíduo.
6. É emitido o mesmo sinal pelos fones de ouvido, numa ordem aleatória dentre as posições possíveis. Essas posições serão virtualmente as mesmas do que as posições reais dos alto-falantes.
7. O indivíduo aponta a cabeça na direção de onde ele acha o som provém. O *HT* registra essa posição.
8. Os pontos 6. e 7. são repetidos para o resto das posições.
9. São calculados e comparados os índices de desempenho de localização.

5.5 Sumário

A forma mais confiável de fazer reprodução biaural, visando avaliar um *VAD HI FI*, é usar fones de ouvido intra-auriculares para minimizar a variação entre indivíduos da *PTF*. Além disso, é desejável um fone intra-auricular de alta qualidade com resposta em frequência plana, evitando a necessidade de qualquer equalização. Além dos fones intra-auriculares, um teste de localização aceitável precisa um rastreador de cabeça (*HT*) para registrar o parecer subjetivo do ouvinte. Em caso de dificuldade para construir a estrutura giratória de alto-falantes, ela pode ser trocada por um *set-up* simples de dois alto-falantes, com a distância entre o ponto de origem do sistema de referência e o centro de cada alto-falante fixa em 1m, e os ângulos de elevação fixados em dois ângulos considerados pelo modelo, por exemplo -23° e 23° . Os alto-falantes não precisam ser caros nem especiais. Um amplificador simples também seria adequado para direcionar os alto-falantes. O software de espacialização deve incluir: geração dos sinais de teste, interface com o usuário, convolução com as *HRTF*, direcionamento biaural e dos alto-falantes, captação dos dados do *HT* e cálculo dos índices de desempenho de localização. Finalmente, um local adequado e tratado acusticamente é necessário para a realização dos testes. O procedimento do testes é simples e foi descrito na seção 5.4.4.

Capítulo 6

CONCLUSÕES E TRABALHOS FUTUROS

6.1 Sumário do trabalho

O presente trabalho apresentou técnicas para estimar a função de transferência relativa ao *pinna* (*PRTF*) de um indivíduo qualquer, a partir de uma base de dados de *HRTFs* e de certos dados antropométricos do indivíduo. As *PRTFs* estimadas pertencem a um intervalo espacial limitado, no plano médio do hemisfério frontal.

No capítulo 2 foi feita uma ampla revisão dos princípios fundamentais de audição espacial, tanto daqueles que governam a localização de fontes sonoras no sentido horizontal quanto daqueles que a governam no sentido vertical. O capítulo 3 apresentou uma revisão dos princípios de personalização da *HRTF* a partir de antropometria e das diversas abordagens apresentadas na literatura. Dentre elas, salientamos a abordagem estrutural e a morfológica. A primeira abrange uma análise efeito-causa das características temporais e espectrais da *HRTF* associadas a certas estruturas anatômicas, i.e. cabeça, torso e *pinnae*. Quanto à segunda abordagem, ela propõe mapeamentos lineares entre o espaço vetorial correspondente a dados antropométricos, e o espaço vetorial correspondente à magnitude espectral de *HRTFs*. No capítulo 4, a solução proposta é apresentada como uma combinação das abordagens estrutural e morfológica. Dada a existência de modelos geométricos da contribuição da *HAT* à *HRTF*, o presente trabalho concentrou-se na elaboração de um modelo da contribuição do *pinna*, que inclui, portanto, a estimação da *PRTF*. Assim, calcularam-se uma série de transformações lineares através do mapeamento linear de espaços vetoriais antropométricos para espaços vetoriais correspondentes à magnitude espectral da resposta acústica causada pelo *pinna*,

PRTF. Essa resposta acústica é estimada pragmaticamente a partir das *HRTF* correspondentes, baseando-se nos estudos psicológicos, neurofisiológicos e comportamentais e nas medições experimentais que a descrevem há mais de 30 anos. A extração da *PRTF*, apresentada na seção 4.6, pode ser considerada como um processo de simplificação da *HRTF* para preservar apenas os detalhes relevantes na percepção de elevação. Uma outra vantagem do uso de *PRTFs* é que as funções de transferência de todos os indivíduos são consideravelmente mais suaves que as *HRTFs* de origem, o que uniformiza o conjunto. A extração da *PRTF* foi atingida através de diversas técnicas como análise preditiva linear, funções de autocorrelação e de atraso de grupo, janelamento, limiarização, interpolação, dizimação, e outras. Na sequência os dados foram reduzidos através de *PCA*, como visto na seção 4.7. Assim, se conseguiu uma representação compacta das *PRTFs* estimadas, isto é, um conjunto de pesos de componentes principais (PCPs) e as frequências centrais das antirressonâncias causadas pelo *pinna* (FCAs). As FCAs foram achadas através de um algoritmo desenvolvido a partir de um outro recentemente publicado. Por outro lado, a modelagem (mapeamento linear) foi feita para *PRTFs* de diferentes indivíduos correspondente à mesma posição espacial. Achou-se importante ensaiar uma solução para cada posição espacial em separado, prevendo-se para trabalhos futuros a análise das implicações de transmissão, reflexão e difração acústica associadas à posição espacial, onde certas dimensões ou estruturas antropométricas podem ser mais importantes que outras. Tentou-se evitar uma aproximação estatística de todas as posições espaciais como um aglomerado de dados, esquema que vem se aplicando em todos os trabalhos anteriores. A seguir, na seção 4.8, foram calculados coeficientes de correlação entre as características espectrais extraídas (PCPs e FCAs) e dimensões antropométricas do *pinna* registradas na base de dados *CIPIC*. As dimensões que apresentaram melhor correlação junto com outros parâmetros antropométricos derivados (áreas e volumes), foram usadas na regressão linear multivariável para calcular as transformações lineares que permitirão estimar PCPs e FCAs de novos indivíduos a partir da coleção dos dados antropométricos correspondentes.

Novamente foram calculados coeficientes de correlação, sendo desta vez entre as PCPs e FCAs originais e suas versões estimadas. Esses coeficientes representam a percentagem de estimação no “grupo de dados de treino”, isto é, a base de dados *CIPIC*. No caso em estudo, as FCAs resultaram em melhores coeficientes de correlação (0,74 e 0,81 para A e B respectivamente). Isto se deve a que as FCAs estão diretamente associadas às dimensões antropométricas que geram os fenômenos de interferência na concha, segundo estudos

passados. Por outro lado, os PCPs não tiveram coeficientes de correlação muito promissores (apenas 0,72 no caso máximo: w_1), já que eles dependem mais do grupo de dados usado na PCA e provavelmente não existe uma relação poderosa entre um PCP e uma dimensão antropométrica em particular. Além disso, o grupo de dados foi pequeno já que se realizou a PCA para cada posição espacial isoladamente. Isto resultou em menos de 70 vetores para cada PCA quando normalmente, em trabalhos com *HRTFs* e PCA, são usados mais de 300 vetores sendo inclusas todas as posições espaciais. A fim de calcular a percentagem unificada de estimação usando todos os PCPs, se introduz uma nova métrica chamada índice de estimação calculada a partir dos autovalores de cada PCP e dos coeficientes de correlação respectivos. O máximo índice de estimação atingiu 66% nos dados de treino, percentagem não tão promissora, porém interessante e muito melhor do que aquela calculada só com as dimensões antropométricas registradas na base de dados: 45%. Desta maneira, os parâmetros antropométricos derivados a partir das dimensões originais (áreas e volumes como o mostrado na Figura 4.20) comprovaram sua utilidade.

Finalmente, as três transformações lineares calculadas, uma para os PCPs e mais duas para as FCAs consideradas, podem ser usadas para calcular novas *PRTFs*. Os resultados dessas estimações foram combinados através do algoritmo de movimentação de contornos de antirressonâncias apresentado na seção 4.8.3. Cabe destacar que as dimensões e parâmetros antropométricos escolhidos, em geral, tiveram considerável relação com a concha, cavidade que aparece como a mais importante no fenômeno refletivo e difrativo da onda acústica no *pinna*, sendo imprescindível para a percepção de elevação. Agora, é possível estimar a *PRTF* de qualquer *pinna* fora da base de dados a partir de duas fotografias, extraindo manualmente as dimensões usadas pelo método. A realização de testes auditivos de localização de fonte sonora é passo a ser realizado, a fim de validar ou de descartar o método. Diversas considerações sobre esses testes, bem como o estado atual dos desenvolvimentos foram apresentados no capítulo 5.

6.2 Apresentação da contribuição de mestrado

O presente trabalho apresentou diversas contribuições na busca de métodos para personalizar a *HRTF* a partir de antropometria, entre elas:

- Foi proposta uma solução genérica para personalização da *HRTF*. Dada a natureza modular da solução, os módulos podem ser modificados isoladamente. Por exemplo: o relacionamento entre antropometria e espectro que pode ser atingido usando outras técnicas como técnicas não lineares.
- Foi introduzido o conceito de processamento bidimensional em grupos de *HRTFs* de um mesmo indivíduo para diferentes elevações. Este processamento pode ser usado na extração de características importantes da *HRTF* como frequências onde aparecem ressonâncias ou antirressonâncias, ou para “limpar” a imagem a fim de realizar um melhor reconhecimento dessas características. Além disso, sugere-se um futuro trabalho de reconhecimento bidimensional de padrões nas imagens de *HRTFs*.
- Foram classificadas e identificadas as antirressonâncias do plano médio presentes em todas as *HRTFs*. A partir do número e distribuição dessas antirressonâncias, 3 padrões diferentes foram reconhecidos. Essa classificação também pode ser estendida a *HRTF* ipsilaterais fora do plano médio.
- Foi proposto um método para extração de uma *PRTF* aproximada de uma *HRTF* original. Esse método está baseado em diversas técnicas de processamento de sinais como análise preditiva linear, funções de autocorrelação e de atraso de grupo, janelamento, dizimação, interpolação e limiarização.
- Foi proposta a derivação de novos parâmetros antropométricos a partir dos existentes na base de dados *CIPIC* a fim de obter mais parâmetros relativos à concha. O objetivo final foi atingir melhores coeficientes de correlação e, portanto, comprovar que a concha é a principal “fornecedora” de antirressonâncias na *HRTF*.
- Foram relacionados antropometria e espectro através de álgebra linear, solução proposta anteriormente, mas nunca realizada com características específicas e diretamente associadas aos fenômenos perceptualmente relevantes como as FCAs.

- Foram introduzidas algumas métricas originais como o índice de importância, o índice de estimação, e o índice de desempenho de localização.
- Foi iniciado o projeto e implementação de um sistema de testes de localização de fonte sonora de baixo custo. Geralmente esses sistemas são proibitivos para qualquer laboratório no mundo.

6.3 Conclusões

A presente seção é considerada a mais importante do trabalho, tendo em vista que o projeto de personalização da *HRTF* é um processo longo e até agora, em vinte anos de pesquisa, não se conseguiram resultados realmente relevantes. A abordagem da proposta apresentada é apenas uma primeira aproximação à resolução do problema. Considerando o caráter inovador da busca de métodos de personalização da *HRTF* e as contribuições originais feitas no campo da psicoacústica e percepção, cada conclusão, apresentada na seqüência, vem acompanhada de uma perspectiva de continuidade de pesquisa apresentada na seção final 6.4.

- No trabalho apresentado em [15] o mapeamento linear entre antropometria e espectro atingiu uma correlação geral de 0,9 considerando apenas os primeiros 7 CPs. Nesse trabalho, os 7 primeiros CPs permitem uma reconstrução da variância de 60% na *PCA*, percentagem que segundo esse estudo fornece boa localização em 5 indivíduos dentro do grupo de treino. Considerando 0,6 de reconstrução e 0,9 de correlação o índice de estimação aproximado ficaria em 0,54. O índice de estimação encontrado no presente trabalho usando vinte CPs resultou em 0,66 no caso em estudo, com uma percentagem de reconstrução da variância de 99%. No entanto, outros trabalhos com *PCA* no passado, mostraram que boa localização é só atingida com pelo menos 90% de reconstrução da variância nas *HRTFs*. Isto nos levaria à hipótese de que qualquer índice de estimação calculado deve levar em conta pelo menos 90% de reconstrução de variância. E por outro lado, o mapeamento linear entre antropometria e espectro resulta em coeficientes fracos, sendo válido só quando testes psicoacústicos sérios fossem realizados, especialmente com indivíduos fora do “grupo de treino”. Assim, é

possível intuir que mapeamentos não lineares podem ser mais adequados para estabelecer uma relação poderosa entre antropometria e espectro.

- O uso de parâmetros antropométricos fora dos registrados na base de dados *CIPIC* resultou num melhor índice geral de estimação (0.66 vs. 0.45). Esses parâmetros foram calculados e “idealizados” a fim de representar melhor a concha, cavidade principal do *pinna* que, segundo estudos passados, aparece como a maior “fornecedora” de antirressonâncias na *HRTF*. Além de se comprovar esse fato, duas conclusões foram tiradas da modelagem antropométrica:
 - Considerando os dados fornecidos na base de dados *CIPIC*, dimensões simples são úteis, mas áreas ou volumes são mais adequados para realizar os mapeamentos.
 - De preferência, um ponto de referência deve ser fornecido ao sistema de medição antropométrica como feito em [15].
- Segundo o observado ao longo do trabalho e segundo as teorias de origem das antirressonâncias na *HRTF* (modelos *SDA* e refletivo-difrativo), é bem possível que a importância dos parâmetros antropométricos dependa da posição espacial, isto é: os melhores preditores de características espectrais variam com respeito à posição de onde provém o som. Isto é lógico considerando que a principal estrutura refletiva “fornecedora” de antirressonâncias na *HRTF* é a parede posterior da concha quando o som provém do hemisfério frontal. Dado que a concha não tem uma estrutura simétrica em plano nenhum, a distância percorrida pela onda refletida e que causa a interferência com a onda direta é diferente para cada direção (ver Figura 2.4). Assim, é provável que uma dimensão simples possa prever pelo menos a primeira FCA (A): a distância até a parede posterior da concha que varia com a elevação. Isto já vem sendo explorado em pesquisas atuais e segundo esses resultados iniciais, a abordagem pode apresentar bons resultados no futuro [67]. Neste caso em particular, não adianta realizar a estimação a partir de áreas e volumes. As áreas e volumes podem prever

melhor a forma geral da magnitude espectral calculada em nosso caso através de reconstrução por componentes principais.

- Mesmo que a abordagem estrutural não tenha mostrado contundentemente sua eficácia neste trabalho nem em trabalhos passados, o uso de modelos antropométricos separados para a *HAT* e o *pinna* aparece como uma opção atrativa e simplificadora em relação à modelagem da *HRTF* inteira, dado que são usados apenas os detalhes perceptualmente relevantes: ressonâncias e antirressonâncias.
- A classificação em padrões das imagens Elevação Vs. Frequência das *HRTFs* resultou num processo viável e simples. A partir da observação foi possível identificar as antirressonâncias do plano médio (A, A2, B, C e C2) e caracterizá-las dentro de um comportamento quase-comum em todos os indivíduos da base de dados *CIPIC*.
- Observando os resultados atingidos com os parâmetros espectrais usados (PCPs e FCAs) é claro que não se deve somente procurar mais e melhores parâmetros antropométricos. A busca de melhores parâmetros espectrais é outra linha de pesquisa importante. Outro tipo de modelagem da magnitude espectral, além da *PCA*, pode resultar em melhores coeficientes de correlação. Assim, ao invés de PCPs, podem ser obtidos dados espectrais mais diretamente relacionados com ressonâncias e antirressonâncias da *HRTF* e que, à sua vez, permitam a reconstrução da função de transferência completa. Modelagem autorregressivo de média móvel (*ARMA*) pode ser uma boa opção para as *PRTFs*.
- Não todas as populações têm características antropométricas iguais, portanto deve-se levar isso em consideração na hora de usar bases de dados específicas para espacializar som a populações específicas.

6.4 Perspectivas para continuidade da pesquisa

Na sequência são apresentadas as perspectivas de trabalhos futuros. Cada uma delas está associada a uma conclusão da seção anterior.

- Sugerem-se realizar tentativas de relacionamento entre antropometria e espectro com redes neurais, algoritmos genéticos e outras técnicas não lineares. Um bom ponto de referência é estudar a solução acústica da equação de onda aplicada no cálculo de *HRTFs* (cálculo numérico de *HRTFs* através de elementos finitos ou elementos de contorno). Isto permitiria estabelecer com maior critério a escolha da técnica a ser usada.
- Outros tipos de parametrização antropométrica são possíveis como evidenciado em [79] onde foi usada a transformada tridimensional elíptica de Fourier para calcular um conjunto de parâmetros unívocos que representem a estrutura tridimensional do *pinna*. O cálculo de novos parâmetros vem acompanhado da proposta de implementação de um sistema de aquisição automática de parâmetros antropométricos, que inclui o desenvolvimento de sistema de aquisição e análise de imagens do *pinna* e da cabeça e torso. Esse sistema deve ser integrável com o software de espacialização de som descrito no capítulo 5. A idéia é capturar imagens (seja a partir de fotografias ou a partir de imagens filmadas em vídeo) para calcular parâmetros antropométricos específicos, por exemplo, a área da cavidade principal ou concha, ou o perímetro completo do *pinna* [70]. Uma possível tentativa será a aplicação de algoritmos desenvolvidos para segmentação de imagens médicas. Eles abrangem fases de pré-processamento das imagens usando técnicas como *wavelets* e operadores derivativos para redução do ruído e elevação de contraste respectivamente. Sequencialmente, será procurada a segmentação do *pinna* mediante algoritmos como *watershed* ou operadores de gradiente morfológico. Uma outra possibilidade ainda mais atrativa, porém mais custosa em termos de tempo e pesquisa, é o desenvolvimento de um sistema de formatação (*shaping*) 3D do *pinna* a partir de imagens de vídeo ou de determinados planos do *pinna* capturados através de fotografias. A idéia neste caso é evitar o uso de escaners laser para a varredura 3D do *pinna*, e concentrar o processamento no algoritmo computacional. Com a forma obtida seria possível determinar uma diversidade de parâmetros físicos e, inclusive, calcular diferentes parâmetros dependendo da posição espacial a ser modelada. Isto nos leva à seguinte conclusão.

- Sugere-se que o tipo de parâmetros antropométricos usado para estimar características espectrais dependa da posição espacial.
- Sugere-se continuar procurando um processo de extração automática da *PRTF* usando *HRTFs* medidas. Isto deve ser complementado otimamente pela informação obtida a partir da classificação bidimensional em padrões.
- Devem-se realizar novos trabalhos em reconhecimento de padrões baseados nas características espectrais (ressonâncias e antirressonâncias), usando técnicas modernas de reconhecimento de padrões. Além disso, um trabalho futuro é estender essa classificação em padrões a posições fora do plano médio, levando em consideração *HRTFs* (ou *PRTFs*) ipsilaterais.
- Quanto à busca de novos parâmetros espectrais é possível aplicar diversas técnicas de pré-processamento. Por exemplo, o uso de bancos de filtros para trabalhar em bandas de frequência (bandas que poderiam ser determinadas, por exemplo, pelas antirressonâncias), o cálculo de fatores de escala em diferentes bandas a fim de normalizar as *PRTFs*, (esses fatores também poderiam ser parâmetros espectrais), entre outros. Um outro ponto importante é determinar a importância psicoacústica da variação de cada um desses parâmetros espectrais. Isto pode ser atingido através de um esquema de processamento de sinais que permita a variação dinâmica de parâmetros espectrais em testes de localização de fonte sonora em uma população considerável de indivíduos de teste. Em princípio, os parâmetros espectrais a serem variados podem ser alguns evidentes: frequências centrais de ressonâncias e antirressonâncias, distância entre antirressonâncias e outros. O algoritmo de movimentação de contornos pode ser aperfeiçoado para esses fins.
- Um estudo estatístico da antropometria do *pinna* associada a características raciais, de idade e de gênero é necessário. No longo prazo, o ideal é medir uma base de dados de *HRTFs* tão abrangente que inclua como parâmetros de seleção e associação todas essas características adicionais.

BIBLIOGRAFIA

- [1] R.E. Miller III, "Contrasting ITU 5.1 and Panor-ambiophonic 4.1 Surround Sound Recording Using OCT and Sphere Microphones", Preprints of the 112th Audio Engineering Society Convention, Munich, Preprint. 5577, 2002.
- [2] M. Kleiner, B.-I. Dalenbäck, P. Svensson, "Auralization - An Overview", Journal of the Audio Engineering Society, Vol.41, pp 861-875, 1993.
- [3] K.J.A. Riederer, "Head-related transfer functions measurements", MSc. Dissertation, Helsinki University of Technology, 1998.
- [4] J. P. Blauert, "Spatial Hearing [Revised edition]", MIT Press, Cambridge, 1997.
- [5] D. R. Begault, "3-D Sound for Virtual Reality and Multimedia", Academic Press, Boston, 1994.
- [6] S. Carlile, C. Jin, V. Harvey, "The generation and validation of high fidelity virtual auditory space", Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Hong Kong, Vol. 3, pp. 1090-1095, 1998.
- [7] H. Møller, M. F. Sørensen, C. B. Jensen, D. Hammershøi, "Binaural technique: Do we need individual recordings?", Journal of the Audio Engineering Society, Vol. 44, No. 6, pp. 451-469, 1996.
- [8] A. Kulkarni, H.S. Colburn, "Variability in the characterization of the headphone transfer function", Journal of the Acoustical Society of America, Vol. 107, No. 2, pp 1071-1074, 2000.
- [9] C.P. Brown, R.O. Duda, "A Structural Model for Binaural Sound Synthesis", IEEE Transactions on Speech and Audio Processing, Vol. 6, No. 5, pp.476-488, 1998.

- [10] J.C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency", *Journal of Acoustical Society of America*, Vol. 106, No. 3, pp.1480-1492, 1999.
- [11] J.C. Middlesbrooks, "Virtual Localization improved by scaling non-individualized external ear transfer function in frequency", *Journal of Acoustical Society of America*, Vol. 106, No. 3, pp. 1493-1510, 1999.
- [12] D. Zotkin, J. Hwang, R. Duraiswami, L.S. Davis, "HRTF personalization using anthropometric measurements", *Proceedings of the 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, pp. 157-160, 2003.
- [13] V.R. Algazi, R.O. Duda, D.M. Thompson, C. Avendano, "The *CIPIC* HRTF Database", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, pp.99-102, 2001.
- [14] Y. Kahana, P. A. Nelson, M. Petyt, S. Choi, "Numerical Modeling of the Transfer Functions of a Dummy-head and of the External Ear", *Proceedings of the Audio Engineering Society 16th International Conference on Spatial Sound Reproduction*, Rovaniemi, pp. 330–334, 1999.
- [15] C. Jin, P. Leong, J. Leung, A. Corderoy, S. Carlile, "Enabling individualized virtual auditory space using morphological measurements", *Proceedings of the 1st IEEE Pacific-Rim Conference on Multimedia*, pp. 235-238, 2000.
- [16] Ch.T. Chang, O.T.Chen, "An efficient 3D sound using head and *pinna* models", *Proceedings of the 1998 IEEE Asia-Pacific Conference on Circuits and Systems*, Chiangmai, pp. 427-430, 1998.
- [17] Ch. J. Tan, W. S. Gan, "User-defined spectral manipulation of HRTF for improved localization in 3D sound systems", *Electronic Letters*, Vol. 34, No. 25, pp. 2387-2389, 1998.

- [18] N. Gupta, A. Barreto, C. Ordonez, “Improving Sound Spatialization by Modifying Head-Related Transfer Functions to Emulate Protruding *Pinnae*”, Proceeding of the IEEE Southeast Conference, Columbia, pp. 446-450, 2002.
- [19] J.C. Middlesbrooks, E.A. Macpherson, Z.A. Onsan, “Psychophysical customization of directional transfer function for virtual sound localization”, Journal of Acoustical Society of America, Vol. 108, No. 6, pp. 3088-3091, 2000.
- [20] J.W. Strutt, “On the acoustic shadow of a sphere”, Philosophy Transactions of the Royal Society of London, No. 203A, pp. 87-97, 1904.
- [21] B.G. Shinn-Cunningham, “Distance cues for virtual auditory space”, IEEE Pacific-Rim Conference on Multimedia, Sydney, pp. 227-230, 2000.
- [22] F. Rattay, C. Gebershuber, “The mammalian auditory hair cell: A simple electric circuit model”, Journal of the Acoustical Society of America, Vol. 103, No. 3, pp. 1158-1165, 1998.
- [23] F.L. Wightman, D.J. Kistler, “Of vulcan ears, human ears and earprints”, Nature Neuroscience, Vol. 1, No. 5, pp. 337-339, 1998.
- [24] J.W.H. Schnupp, J. Booth, A.J. King, “Modeling individual differences in ferret external ear transfer functions”, Journal of the Acoustical Society of America, Vol. 114, No. 4, pp. 2021-2030, 2003.
- [25] B. G. Shinn-Cunningham, N. I. Durlach, R. M. Held, “Adapting to supernormal auditory localization cues I: Bias and resolution” Journal of the Acoustical Society of America, Vol. 103, No. 6, pp. 3656-3666, 1998.
- [26] M.L. Hawleyb, R.Y. Litovskyc, J.F. Culling, “The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer”, Journal of the Acoustical Society of America, Vol. 115, No. 2, pp. 833-843, 2004.
- [27] H. Kuttruf, “Room Acoustics”, Elsevier Science Publishers Ltd, New York, 1991.

- [28] P. Mackensen, "Auditive Localization. Head movements, an additional cue in Localization", PhD. Thesis, Technische Universität Berlin, 2004.
- [29] J. Huopaniemi, "Virtual Acoustics and 3D sound in multimedia signal processing", PhD. Thesis, Helsinki University of Technology, 1999.
- [30] R.O Duda, W.L. Martens, "Range dependence of the response of a spherical head model", Journal of the Acoustical Society of America, Vol. 104, No. 5, pp. 3048-3058, 1998.
- [31] J. Hebrank, D. Wright, "Spectral cues used in the location of sound sources on the median plane" Journal of the Acoustical Society of America, Vol. 56, No. 6, pp.1829–1834, 1974.
- [32] D.W. Batteau, "The role of the *pinna* in human localization", Proceedings of the Royal Society of London, No. 168B, pp.158-180, 1967.
- [33] M.B. Gardner, R.S. Gardner, "Problem of localization in the median plane: Effect of *pinna* cavity occlusion", Journal of the Acoustical Society of America, Vol. 53, No. 2, pp. 400-408, 1973.
- [34] R.S. Heffner, H.E. Heffner, G. Koay, "Sound Localization in chinchillas. III. Effect of *pinna* removal", Hearing Research, Vol. 99, pp. 13-21, 1996.
- [35] M.F. Zoltan, "Monoaural and binaural spectral cues created by the external ears of the pallid bat", Hearing Research, vol. 95, pp. 1-17, 1996.
- [36] B.C.J. Moore, S.R. Oldfield, G. Dooley, "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz", Journal of the Acoustical Society of America, Vol. 85, pp. 820-836, 1989.

- [37] E.A. Macpherson, “On the role of head-related transfer function spectral notches in the judgment of sound source elevation”, Proceedings of the 2nd International Conference on Auditory Display, Santa Fe, pp.187-194, 1994.
- [38] P.W.F. Poon, J.F. Brugge, “Sensitivity of auditory nerve fibers to spectral notches”, Journal of Neurophysiology, Vol. 70, pp. 655-666, 1993.
- [39] E. A. G. Shaw, R. Teranishi, “Sound Pressure generated in an external-ear replica and real human ears by a nearby sound source”, Journal of the Acoustical Society of America, Vol. 44, pp.240–249, 1968.
- [40] E.A. Lopez-Poveda, R. Meddis, “A physical model of sound diffraction and reflections in the human concha”, Journal of the Acoustical Society of America, Vol. 100, No. 5, pp.3248-3259, 1996.
- [41] V.R. Algazi, C. Avendano, R.O. Duda, “Elevation localization and head-related transfer function analysis at low frequencies”, Journal of the Acoustical Society of America, Vol. 109, No. 3, pp.1110–1122, 2001.
- [42] W.R. Thurlow, J.W. Mangels, P.S. Runge, “Head movements during sound localization”, Journal of the Acoustical Society of America, Vol. 42, No. 2, pp. 489-493, 1967.
- [43] W.G. Gardner, K. D. Martin, “HRTF Measurements of a KEMAR”, Journal of the Acoustical Society of America, Vol. 97, No.6, pp.3907-3908, 1995.
- [44] <http://recherche.ircam.fr/equipes/salles/listen/index.html> (revisado em junho de 2003).
- [45] D.J. Kistler, F.L. Wightman, “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction”, Journal of the Acoustical Society of America, Vol. 91, No. 3, pp.1637-1647, 1992.

- [46] E. Wenzel, M. Arruda, M. Kistler, D.J. Wightman, "Localization using non-individualized head-related transfer functions", *Journal of the Acoustical Society of America*, Vol. 94, No. 1, pp.111-123, 1993.
- [47] W.M. Hartmann, A. Wittenberg, "On the externalization of sound images", *Journal of the Acoustical Society of America*, Vol. 99, No. 6, pp. 3678–3688, 1996.
- [48] J. Chen, B. Veen, K.E. Hecox, "A spatial feature extraction and regularization model of the head-related transfer function", *Journal of the Acoustical Society of America*, Vol. 97, No. 1, pp.439-452, 1995.
- [49] S.M. Abel, V.H. Hay, "Sound localization: The interaction of aging, hearing loss, and hearing protection" *Scandinavian Audiology*, Vol. 25, No. 1, pp.3-12, 1996.
- [50] L.W. Biscainho, F.P. Freeland, P. S. Diniz, "Using inter-positional transfer functions in 3-D sound", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2, Salt Lake City, pp.1961-1964, 2002.
- [51] P.W.F. Poon, J.F. Brugge, "Virtual-space receptive fields of single auditory nerve fibers", *Journal of Neurophysiology*, Vol. 70, pp. 667-676, 1993.
- [52] R. Hausler, H. S. Colburn, E. Marr, "Sound localization in subjects with impaired hearing" *Acta Oto-Laryngologica*, Vol. 400, pp. 1-62, 1983.
- [53] J.E. West, J. Blauert, D.J. MacLean, "Teleconferencing system using head-related signals", *Applied Acoustics*, Vol. 36, pp. 327-333, 1992.
- [54] A. Harma, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, G. Lorho, "Augmented Reality Audio for Mobile and Wearable Appliances", *Journal of the Audio Engineering Society*, Vol. 52, No. 6, pp. 618-639, 2004.
- [55] K.J.A. Riederer, "Repeatability analysis of HRTF measurements", *Preprints of the 105th Audio Engineering Society Convention*, San Francisco, Preprint. 4846, 1998.

[56] H. Møller, "Fundamentals of binaural technology", *Applied Acoustics*, Vol. 36, No. 3/4, pp. 171-218, 1992.

[57] Y. Kahana, P.A. Nelson, M. Petyt, "Boundary element simulation of HRTFs and sound fields produced by virtual acoustic imaging", *Preprints of the 105th Audio Engineering Society Convention*, San Francisco, Preprint 4817, 1998.

[58] J.C. Middlebrooks, J.C. Makous, D.M. Green, "Directional sensitivity of sound-pressure levels in the human ear canal", *Journal of the Acoustical Society of America*, Vol. 86, No. 1, pp. 89-108, 1989.

[59] V. R. Algazi, R. O. Duda, D. M. Thompson, "The use of head-and-torso models for improved spatial sound synthesis", *Preprints of the 113th Audio Engineering Society Convention*, Los Angeles, preprint No. 5712, 2002.

[60] K. Genuit, "Ein Modell zur Beschreibung von Aussenohrübertragungseigenschaften", *PhD Thesis*, Rheinisch-Westfälischen Technischen Hochschule Aachen, 1984.

[61] V.R. Algazi, R.O. Duda, R.P. Morrison, D.M. Thompson, "Structural composition and decomposition of HRTFs", *Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, pp.103–106, 2001.

[62] G.F. Kuhn, R.M. Guernsey, "Sound pressure distribution about the human head and torso", *Journal of the Acoustical Society of America*, Vol. 73, No. 1, pp. 95–105, 1983.

[63] E.A.G. Shaw, "Acoustical features of the human external ear", *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey and T. R. Anderson, Erlbaum, pp. 25–47, 1997.

[64] V.C. Raykar, R. Duraiswami, L. Davis, B. Yegnanarayana, "Extracting significant features from the HRTF", *Proceedings of the 9th International Conference on Auditory Display*, Boston, pp. 115 – 118, 2003.

- [65] S.R. Soria, M.A. Ramírez, “Feature identification techniques for HRTF individualization”, Proceedings of the 2004 International Workshop on Telecommunications, Santa Rita de Sapucaí, pp.188-193, 2004.
- [66] R. Gonzales, R. Wood, “Digital Image Processing”, Prentice-Hall, Inc., 2a ed., New Jersey, 2002.
- [67] V.C. Raykar, R. Duraiswami, B. Yegnanarayana, “Extracting frequencies of the *pinna* spectral notches in measured head related impulse responses”, Technical report CS-TR-4609, Perceptual Interfaces and Reality Laboratory, University of Maryland, 2004.
- [68] I. T. Jolliffe, “Principal component analysis”, Springer-Verlag, New York, 1986.
- [69] W. Martens, “Principal components analysis and resynthesis of spectral cues to perceived direction”, Proceedings of the International Computer Music Conference, Champaign/Urbana, pp.274-281, 1997.
- [70] S. R. Soria, M.A. Ramírez, “HRTF Individualization by Solving the Least Squares Problem”, Preprints of the 118th Audio Engineering Society Convention, Barcelona, Preprint 6738, 2005.
- [71] S. R. Soria, M.A. Ramírez, “Extracting and modeling approximated *pinna*-related transfer functions from HRTF data”, Proceeding of the 11th Meeting of the International Conference in Auditory Displays, Limerick, pp. 269-273, 2005.
- [72] S. R. Soria, M.A. Ramírez, “Linear Relationships Between Spectral Characteristics and Anthropometry of the External Ear”, Proceeding of the 11th Meeting of the International Conference in Auditory Displays, Limerick, pp. 336-339, 2005
- [73] W. Martens, “Perceptual Evaluation of Filters Controlling Source Direction: Customized and Generalized HRTFs for Binaural Synthesis”, Acoustical Science and Technology, Vol. 24, No. 5, pp. 220-232, 2003.

[74] A.J. Berkhout, D. de Vries, P. Vogel, "Acoustic control by wave field synthesis," Journal of the Acoustical Society of America, Vol. 7, No. 2, pp.188-196, 1999.

[75] H. Møller, "Fundamentals of binaural technology", Applied Acoustics, Vol. 36, No. 3/4, pp. 171-218, 1992.

[76] H. Møller, D. Hammershøi, C. B. Jensen, M. F. Sørensen, "Transfer characteristics of headphones measured on human ears", Journal of the Audio Engineering Society, Vol. 43, No. 4, pp. 203-217, 1995.

[77] A. Kulkarni, H.S. Colburn, "Variability in the characterization of the headphone transfer function", Journal of the Acoustical Society of America, Vol. 107, No. 2, pp 1071-1074, 2000.

[78] D. Hammershøi, H. Møller, "Sound transmission to and within the human ear canal", Journal of the Acoustical Society of America, Vol. 100, No. 1, pp. 408-427, 1996.

[79] C. Hetherington, A.I. Tew, T. Yufei, "Three-dimensional elliptic Fourier methods for the parameterization of human *pinna* shape", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 5, Hong Kong, pp. 612-615, 2003.