



Universidade do Estado do Rio de Janeiro
Centro de Tecnologia e Ciências
Instituto Politécnico

Filipe Otsuka Taminato

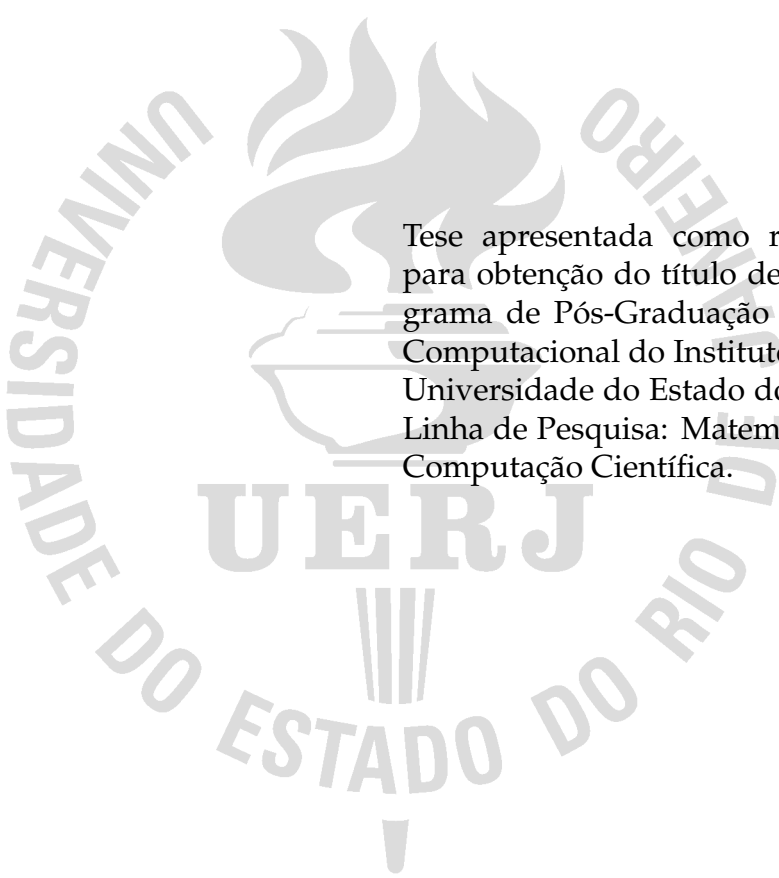
**Redes neurais artificiais aplicadas à modelagem de respostas
impulsivas associadas à cabeça humana para gerar auralização**

Nova Friburgo

2018

Filipe Otsuka Taminato

Redes neurais artificiais aplicadas à modelagem de respostas impulsivas associadas à cabeça humana para gerar auralização



Tese apresentada como requisito parcial para obtenção do título de Doutor ao Programa de Pós-Graduação em Modelagem Computacional do Instituto Politécnico, da Universidade do Estado do Rio de Janeiro. Linha de Pesquisa: Matemática Aplicada e Computação Científica.

Orientador: Prof. Dr. Roberto Aizik Tenenbaum

Nova Friburgo

2018

CATALOGAÇÃO NA FONTE
UERJ / REDE SIRIUS / BIBLIOTECA CTC/E

T158 Taminato, Filipe Otsuka.
Redes neurais artificiais aplicadas à modelagem de respostas impulsivas associadas à cabeça humana para gerar aurilização / Filipe Otsuka Taminato. - 2018.
152 f. : il.

Orientador: Roberto Aizik Tenenbaum.

Tese (Doutorado) – Universidade do Estado do Rio de Janeiro, Instituto Politécnico.

1. Redes neurais (Computação) – Teses. 2. Aurilização - Metodos de simulação – Teses. 3. Percepção auditiva - Metodos de simulação – Teses. 4. Acustica – Teses. 5. Testes de emissão acústica – Teses. 6. Realidade virtual na engenharia – Teses. I. Tenenbaum, Roberto Aizik. II. Universidade do Estado do Rio de Janeiro. Instituto Politécnico. III. Título.

CDU 534.76:004.032.26

Bibliotecária: Sandra Mueller CRB7/3633

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação.

Assinatura

Data

Filipe Otsuka Taminato

**Redes neurais artificiais aplicadas à modelagem de respostas impulsivas associadas
à cabeça humana para gerar auralização**

Tese apresentada como requisito parcial para obtenção do título de Doutor ao Programa de Pós-Graduação em Modelagem Computacional do Instituto Politécnico, da Universidade do Estado do Rio de Janeiro. Linha de Pesquisa: Matemática Aplicada e Computação Científica.

Aprovada em 26 de Fevereiro de 2018.

Banca Examinadora:

Prof. Dr. Roberto Aizik Tenenbaum (Orientador)
Instituto Politécnico – UERJ

Prof. Dr. Antônio José da Silva Neto
Instituto Politecnico – UERJ

Prof. Dr. Diego Campos Knupp
Instituto Politecnico – UERJ

Prof. Dr. Julio Cesar Boscher Torres
Universidade Federal do Rio de Janeiro

Prof. Dra. Mariane Rembold Petraglia
Universidade Federal do Rio de Janeiro

Nova Friburgo

2018

DEDICATÓRIA

Aos meus pais.

AGRADECIMENTOS

Ao meu orientador, Prof. Roberto Aizik Tenenbaum, pela competência, dedicação, paciência, amizade e por me apresentar a esta fascinante área de estudo.

Aos meus pais, Roberto e Márcia, que me proporcionaram a oportunidade de formação pessoal e profissional e me incentivaram durante todo o tempo.

Aos meus irmãos, Rodrigo e Simone, e toda minha família, que sempre me apoiaram nos momentos que precisei e pela compreensão nos momentos de ausência necessários para a concretização desta tese.

Agradeço aos membros da banca examinadora, Prof. Dr. Antônio J. S. Neto, Prof. Dr. Diego C. Knupp, Prof. Dr. Julio C. B. Torres, Prof. Dr. Mariane R. Petraglia, por todas as correções e sugestões, que foram muito importantes para melhorar o texto.

Aos co-autores dos artigos publicados durante a elaboração da tese.

A todos os professores e colegas do departamento, que ajudaram de forma direta e indireta na conclusão deste trabalho.

A todos os colegas do LIDAV, em especial (em ordem alfabética): José F.L. Naranjo, Josué C. Baptista, Pâmela G. Lima, Taciano M.S.M. Barros, Thiago A.O. Roxo e Vera L.D. Ferreira.

A todos os colegas do IPRJ, em especial (em ordem alfabética): Amaury Muñoz Oliva, Gisiane Santos Simão Ferreira, Mauro de Amorim (*in memoriam*) e Stella Oggioni da Fonseca.

A todos os funcionários do IPRJ/UERJ pela colaboração.

Por fim, agradeço à CAPES pelo apoio financeiro prestado durante os dois primeiros anos e à FAPERJ pela Bolsa Nota 10 nos dois últimos anos do curso.

The knower of the mystery of sound
knows the mystery of the whole universe.
Hazrat Inayat Khan

RESUMO

TAMINATO, Filipe Otsuka. *Redes neurais artificiais aplicadas à modelagem de respostas impulsivas associadas à cabeça humana para gerar aurilização*. 2018. 153 f. Tese (Doutorado em Modelagem Computacional) – Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2018.

Um dos objetivos do desenvolvimento e do aperfeiçoamento de técnicas em sistemas de geração de aurilização consiste em reduzir o custo computacional e garantir a qualidade da simulação acústica. Para obter a aurilização é necessário, além de um som anecoico, a determinação das respostas impulsivas biauriculares da sala (BRIRs), para um determinado par fonte-receptor. Nesse sentido, uma maneira de reduzir a complexidade de tais sistemas é diminuindo o custo computacional do cálculo da modificação espectral das respostas impulsivas associadas à cabeça (HRIRs). Para isso, neste trabalho, foram utilizadas as redes neurais do tipo RBF para substituir o método tradicional. Para cobrir todo o espaço auditivo de recepção, um conjunto de redes neurais artificiais foi adotada, sendo que cada rede atua em uma direção específica. Cada rede neural tem como entrada o espectro do raio sonoro em bandas de oitava e como saída a HRIR filtrada da direção correspondente. A estratégia do treinamento dessas redes neurais, a implementação das redes, do modelo e das observações são apresentadas. Através do método proposto obteve-se uma redução do custo computacional de aproximadamente de 70% em relação ao método tradicional, a qual representa uma importante redução na implementação dos sistemas de aurilização. Para verificar a eficiência da abordagem, foram realizadas diversas comparações entre as HRIRs filtradas utilizando as redes RBF e o método tradicional. A similaridade entre os resultados foi avaliada a partir do cômputo do coeficiente de correlação cruzada normalizada. Os resultados simulados demonstram a eficiência da técnica que utiliza as redes neurais, pois esses apresentam valores de correlações próximos de um. Foram observadas discrepâncias em apenas algumas componentes de fase das HRTFs, para algumas direções, no entanto, estes erros não são apreciáveis. São apresentados, também, alguns resultados da intercomparação internacional de simuladores de acústica de salas, denominado *Round Robin on Auralization*. Para o cálculo das BRIRs foram utilizados o simulador de acústica de salas RAIOS (*Room Acoustics Integrated and Optimized Software*) e as redes RBF. Os resultados simulados e medidos são comparados e discutidos.

Palavras-chave: Realidade Virtual Acústica. Respostas Impulsivas Bi-auriculares.

Aurilização. Redes Neurais Artificiais.

ABSTRACT

TAMINATO, Filipe Otsuka. *Artificial neural networks applied to the modeling of head related impulse response to generate auralization*. 2018. 153 f. Tese (Doutorado em Modelagem Computacional) – Instituto Politécnico, Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2018.

One of the objectives of the development and improvement of techniques in systems of generation of auralization consists in the reduction of computational cost and to guarantee the quality of acoustic simulation. In order to obtain the auralization, it is necessary, besides an anechoic sound, the determination of the binaural room impulsive responses (BRIRs), for a determined pair of source-receiver. In this sense, one way to reduce the complexity of such systems is to reduce the computational cost of calculating the spectral modification of the head-related impulse responses (HRIRs). For this, in this work, the RBF neural networks were used to replace the traditional method. To cover the entire auditory reception space, a set of artificial neural networks was adopted, each network acting in a specific direction. Each neural network has as input the acoustic ray's spectrum in octave bands and as output the filtered HRIR of the corresponding direction. The strategy of training these neural networks, the implementation of the networks, the model and the observations are presented. By means of the proposed method, a computational cost reduction of approximately 70% was obtained in relation to the traditional method, which represents a significant reduction in the implementation of the auralization systems. To verify the efficiency of the approach, several comparisons were made between the filtered HRIRs using the RBF neural networks and the traditional method. The similarity between the results was evaluated from the normalized cross-correlation coefficient. The simulated results demonstrate the efficiency of the technique using the neural networks, since these have values of correlations close to one. Only a discrepancy in some phase components of HRTFs has been observed for some directions, however, these errors are unnoticeable. Also presented are some results of the international intercomparison of room acoustics simulators, called Round Robin on Auralization. For the calculation of the BRIRs, the code RAIOS (Room Acoustics Integrated and Optimized Software) and the RBF neural networks were used. The simulated and measured results are compared and discussed.

Keywords: Acoustic Virtual Reality. Binaural Room Impulse Responses. Auralization. Artificial Neural Network.

LISTA DE ILUSTRAÇÕES

Figura 1 - Sensação auditiva 3D	15
Figura 2 - Esquema para gerar aurilização	17
Figura 3 - Medição das HRTFs	18
Figura 4 - Discretização da cabeça utilizada no BEM	19
Figura 5 - Área de audição dos seres humanos	26
Figura 6 - Distribuição temporal da energia sonora para um determinado par fonte-receptor	28
Figura 7 - Classificação dos métodos para simulação da propagação sonora . .	30
Figura 8 - Distribuição da energia do raio incidente pela absorção, reflexão especular e espalhamento	33
Figura 9 - Crescimento e decaimento sonoro em salas	34
Figura 10 - Função impulso unitário	38
Figura 11 - Exemplos de espectros de alguns sinais acústicos	41
Figura 12 - Representação de um sistema que produz uma resposta $y(t)$ baseada em uma entrada $x(t)$	42
Figura 13 - Exemplos de cabeças artificiais de adulto de diversos fabricantes . .	44
Figura 14 - Aparato de medição de respostas impulsivas	45
Figura 15 - Resposta impulsiva medida na mesma posição da cabeça artificial da sala mostrada na Figura 14	45
Figura 16 - Respostas impulsivas biauriculares de acordo com as configurações mostradas na Figura 14	46
Figura 17 - Diagrama de blocos de geração de aurilização.	47
Figura 18 - Janela principal do RAIOS	48
Figura 19 - Aba <i>Sources</i> da janela <i>Data Entry</i> do código RAIOS	49
Figura 20 - Triangularização da superfície	50
Figura 21 - Orientação do vetor normal ao elemento triangular deve estar sem- pre apontada para o interior da sala	51
Figura 22 - Visibilidade de um elemento triangular ao raio acústico	52
Figura 23 - Aba <i>Surfaces</i> da janela <i>Data Entry</i> do código RAIOS	52
Figura 24 - Teste de detecção para verificar se o raio acústico intercepta o recep- tor, cuja posição corresponde ao ponto O	54
Figura 25 - Sistema de coordenadas fixo ao receptor	57
Figura 26 - Sombra acústica gerada pela fenômeno de difração na cabeça	59
Figura 27 - HRIRs e HRTFs para $\phi = 45^\circ$ e $\theta = 0^\circ$	60
Figura 28 - Aba <i>Receivers</i> da janela <i>Data Entry</i> do código RAIOS	61
Figura 29 - Esquema de geração das BRIRs via método tradicional	63

Figura 30 - Principais componentes do neurônio biológico	64
Figura 31 - Representação de um neurônio artificial	65
Figura 32 - Modelo de um neurônio artificial utilizado neste trabalho	66
Figura 33 - Função de ativação linear	67
Figura 34 - Função de ativação logística	68
Figura 35 - Função de ativação tangente hiperbólica	68
Figura 36 - Função de ativação gaussiana	69
Figura 37 - Redes neurais não-recorrentes	70
Figura 38 - Rede neural recorrente	71
Figura 39 - Aprendizado supervisionado	72
Figura 40 - Notações para a dedução do algoritmo <i>backpropagation</i>	75
Figura 41 - Ajuste dos pesos sinápticos da l -ésima camada intermediária	80
Figura 42 - Arquitetura de uma rede neural do tipo <i>Radial Basis Function</i>	86
Figura 43 - Área de operação da rede (gomo)	93
Figura 44 - Resultados gerados pelas abordagens RNA-MLP e MT	94
Figura 45 - Direções do espaço auditivo onde existe uma rede RBF e a escolha desta para uma determinada frente de onda	95
Figura 46 - Comparação das HRIRs filtradas para ($\phi = 2^\circ$ e $\theta = 12^\circ$)	98
Figura 47 - Comparação das HRIRs filtradas para ($\phi = 87^\circ$ e $\theta = 79^\circ$)	98
Figura 48 - Comparação das HRIRs filtradas para ($\phi = 191^\circ$ e $\theta = 67^\circ$)	98
Figura 49 - Comparação das HRIRs filtradas para ($\phi = 274^\circ$ e $\theta = -3^\circ$)	99
Figura 50 - Direção de chegada do raio acústico oriundo das reflexões difusas	100
Figura 51 - Comparação entre a HRIR original com a filtrada (à direita), sendo que uma foi obtida a partir de um raio com maior atenuação nas altas frequências e as outras nas baixas frequências	102
Figura 52 - Redução percentual do número de operações usando a rede RBF em função do número de neurônios na camada intermediária	108
Figura 53 - Performance da rede RBF em função do número de neurônios na camada intermediária para diversas direções	110
Figura 54 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 45^\circ$ e $\theta = 28^\circ$, utilizando os métodos tradicional e da rede RBF	111
Figura 55 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 70^\circ$ e $\theta = -63^\circ$, utilizando os métodos tradicional e da rede RBF	112
Figura 56 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 102^\circ$ e $\theta = 5^\circ$, utilizando os métodos tradicional e da rede RBF	113
Figura 57 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 159^\circ$ e $\theta = -12^\circ$, utilizando os métodos tradicional e da rede RBF	114
Figura 58 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 197^\circ$ e $\theta = 58^\circ$, utilizando os métodos tradicional e da rede RBF	115

Figura 59 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 241^\circ$ e $\theta = -5^\circ$, utilizando os métodos tradicional e da rede RBF	116
Figura 60 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 283^\circ$ e $\theta = 36^\circ$, utilizando os métodos tradicional e da rede RBF	117
Figura 61 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 350^\circ$ e $\theta = -49^\circ$, utilizando os métodos tradicional e da rede RBF	118
Figura 62 - Esquema de geração das BRIRs via redes RBF	120
Figura 63 - Sala do <i>Round Robin 4</i> denominada <i>Seminar room</i> (sala pequena)	122
Figura 64 - Sala do <i>Round Robin 4</i> denominada <i>Chamber music hall</i> (sala média)	123
Figura 65 - Sala do <i>Round Robin 4</i> denominada <i>Auditorium</i> (sala grande)	124
Figura 66 - Comparação das BRIRs medidas e simuladas para o par LS3-MP6 da sala <i>Seminar room</i>	128
Figura 67 - Comparação das BRIRs medidas e simuladas para o par LS7-MP6 da sala <i>Seminar room</i>	129
Figura 68 - Comparação das BRIRs medidas e simuladas para o par LS3-MP6 da sala <i>Chamber music hall</i>	130
Figura 69 - Comparação das BRIRs medidas e simuladas para o par LS7-MP6 da sala <i>Chamber music hall</i>	131
Figura 70 - Comparação das BRIRs medidas e simuladas para o par LS3-MP6 da sala <i>Auditorium</i>	132
Figura 71 - Comparação das BRIRs medidas e simuladas para o par LS7-MP6 da sala <i>Auditorium</i>	133
Figura 72 - Comparação entre as HRTFs de um bebê, uma criança e um adulto no plano horizontal e mediano	138

LISTA DE TABELAS

Tabela	1 - Número de medições e incrementos de azimute para cada elevação.	93
Tabela	2 - Número de amostras da HRTF em cada banda de oitava, para a HRIR com 128, 256 e 512 amostras.	97
Tabela	3 - Número de operações que o método tradicional executa para a HRTF com 512 amostras.	106
Tabela	4 - Número de operações para calcular a saída do j -ésimo neurônio da camada intermediária.	107
Tabela	5 - Número de operações que o método da rede RBF requer.	107
Tabela	6 - Coeficientes de absorção (α) e espalhamento (δ) em bandas de oitava utilizadas nas superfícies da sala <i>Seminar room</i>	123
Tabela	7 - Coeficientes de absorção (α) e espalhamento (δ) em bandas de oitava utilizadas nas superfícies da sala <i>Chamber music hall</i>	124
Tabela	8 - Coeficientes de absorção (α) e espalhamento (δ) em bandas de oitava utilizadas nas superfícies da sala <i>Auditorium</i>	125
Tabela	9 - Posições das fontes e dos receptores nas salas <i>Seminar room</i> , <i>Chamber music hall</i> e <i>Auditorium</i> no sistema de coordenadas cartesianas de cada sala.	125
Tabela	10 - Orientação do eixo principal de radiação das fontes e das cabeças das salas <i>Seminar room</i> , <i>Chamber music hall</i> e <i>Auditorium</i>	125
Tabela	11 - Comparação do parâmetros de qualidade acústica das BRIRs medidas e simuladas, sala <i>Seminar room</i> , par LS3-MP6.	134
Tabela	12 - Comparação do parâmetros de qualidade acústica das BRIRs medidas e simuladas, sala <i>Seminar room</i> , par LS7-MP6.	134
Tabela	13 - Comparação do parâmetros de qualidade acústica das BRIRs medidas e simuladas, sala <i>Chamber music hall</i> , par LS3-MP6.	134
Tabela	14 - Comparação do parâmetros de qualidade acústica das BRIRs medidas e simuladas, sala <i>Chamber music hall</i> , par LS7-MP6.	135
Tabela	15 - Comparação do parâmetros de qualidade acústica das BRIRs medidas e simuladas, sala <i>Auditorium</i> , par LS3-MP6.	135
Tabela	16 - Comparação do parâmetros de qualidade acústica das BRIRs medidas e simuladas, sala <i>Auditorium</i> , par LS7-MP6.	135

LISTA DE ABREVIATURAS E SIGLAS

BEM	<i>Boundary Element Method</i> (Método dos Elementos de Contorno)
BRIR	<i>Binaural Room Impulse Response</i> (Resposta Impulsiva Biauricular da Sala)
EDT	<i>Early Decay Time</i> (Tempo de decaimento inicial)
FDTD	<i>Finite-Difference Time-Domain</i> (Método das Diferenças Finitas no Domínio do Tempo)
FFT	<i>Fast Fourier Transform</i> (Transformada Rápida de Fourier)
FIR	<i>Finite Impulse Response</i> (Resposta ao Impulso Finita)
HRIR	<i>Head Related Impulse Response</i> (Respostas Impulsivas Associadas à Cabeça)
HRTF	<i>Head Related Transfer Function</i> (Funções de Transferência Associadas à Cabeça)
IACC	<i>Inter-Aural Cross-Correlation Coefficient</i> (Correlação cruzada interauricular)
IIR	<i>Infinite Impulse Response</i> (Resposta ao Impulso Infinita)
ILD	<i>Interaural Level Differences</i> (Diferenças de nível interauriculares)
ITD	<i>Interaural Time Differences</i> (Diferenças temporais interauriculares)
ITDG	<i>Initial Time Delay Gap</i> (Intervalo de tempo de atraso inicial)
LF	<i>Lateral Energy Fraction</i> (Fração de energia lateral)
LLE	<i>Locally Linear Embedding</i>
MAA	<i>Minimum Audible Angle</i> (Mínimo Ângulo Audível)
MLP	<i>Multilayer Perceptron</i>
MLS	<i>Maximum Length Sequences</i>
NCCC	<i>Normalized cross-correlation coefficient</i> (Coeficiente de correlação cruzada normalizada)
NMSE	<i>Normalized Mean Square Error</i> (Erro médio quadrático normalizado)
PCA	<i>Principal Components Analysis</i> Análise de Componentes Principais
PQA	Parâmetros de Qualidade Acústica
RAIOS	<i>Room Acoustics Integrated and Optimized Software</i>
RBF	<i>Radial Basis Function</i>
RIR	<i>Room Impulse Response</i> (Resposta Impulsiva da Sala)
RNA	Redes Neurais Artificiais
SLIT	Sistema Linear e Invariante no Tempo
SFER	<i>spatial feature extraction and regularization model</i>
SNR	<i>Signal-to-Noise Ratio</i> (Relação sinal-ruído)
SVD	<i>Singular Value Decomposition</i>

SUMÁRIO

	INTRODUÇÃO	15
1	ACÚSTICA DE SALAS	25
1.1	Introdução	25
1.2	Campo sonoro	27
1.2.1	<u>Modelagem do campo sonoro</u>	30
1.3	Interação onda-superfície	31
1.4	Absorção sonora	32
1.5	Tempo de reverberação	33
1.6	Processamento de sinais	35
1.6.1	<u>Transformada de Fourier</u>	36
1.6.2	<u>Sistemas lineares e invariantes no tempo</u>	42
2	AURILIZAÇÃO DE SALAS	47
2.1	Fontes sonoras	49
2.2	Geometria da sala: superfícies de contorno da sala	50
2.3	Campo sonoro	53
2.3.1	<u>Método do traçado de raios</u>	53
2.3.2	<u>Método das transições de energia</u>	54
2.4	Receptores	56
2.5	Método tradicional de geração das respostas impulsivas biauriculares da sala	61
2.6	Aurilização	63
3	REDES NEURAIS ARTIFICIAIS	64
3.1	As redes neurais biológicas	64
3.2	Neurônio artificial	65
3.2.1	<u>Funções de ativação</u>	67
3.3	Arquitetura	69
3.4	Aprendizado	71
3.5	Redes perceptron de múltiplas camadas	73
3.5.1	<u>Algoritmo <i>backpropagation</i></u>	74
3.5.2	<u>Algoritmo Levenberg-Marquardt</u>	82
3.6	As redes de Funções de Base Radiais	86
3.6.1	<u>Primeira fase do treinamento</u>	89
3.6.2	<u>Segunda fase do treinamento</u>	90
4	MODELAGEM DE HRIRS FILTRADAS COM RNAS	92
4.1	Modelagem utilizando as redes MLP	92
4.2	Modelagem utilizando as redes RBF	95

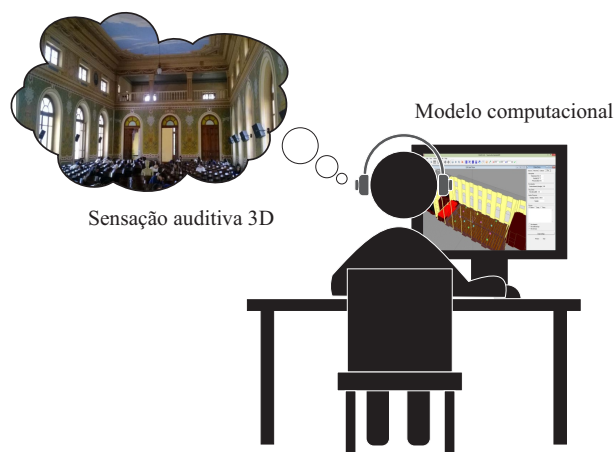
4.2.1	<u>Topologia da rede RBF</u>	96
4.2.2	<u>Ajuste das reflexões difusas para a aplicação do método das redes RBF</u>	99
4.2.3	<u>Padrões de treinamento</u>	100
4.2.4	<u>Algoritmo modificado para o ajuste dos parâmetros livres</u>	103
4.3	Custo computacional	105
4.3.1	<u>Custo computacional do método tradicional de geração das HRIRs filtradas</u>	105
4.3.2	<u>Custo computacional da fase de execução da rede RBF</u>	106
4.4	Resultados dos testes de algumas redes RBF	109
4.5	Construção das BRIRs utilizando as redes RBF	119
5	RESULTADOS DO ROUND ROBIN ON AURALIZATION	121
5.1	<i>Round Robin on Auralization</i>	121
5.2	Dados de entrada para a geração das BRIRs	121
5.3	Comparação dos resultados numéricos com os medidos	126
	CONCLUSÕES E CONTINUIDADE DO TRABALHO	136
	REFERÊNCIAS	139
	APÊNDICE – Cálculo da função exponencial	151

INTRODUÇÃO

Motivação e objetivos

A técnica conhecida como *Realidade Virtual Acústica* (RVA) tem como objetivo recriar ambientes virtuais acústicos a partir de dados medidos ou simulados (VORLÄNDER, 2008). Suas aplicações são inúmeras, varrendo desde o controle de ruído em plantas industriais, a avaliação subjetiva de isolamento acústico entre recintos, o auxílio ao projeto acústico de salas para concertos, até a melhoria da inteligibilidade da palavra em salas de aula. O resultado mais importante em RVA consiste na *aurilização*¹ que nada mais é do que a geração de sinais sonoros a serem ouvidos (geralmente através de fones de ouvido), os quais permitem criar a sensação auditiva de o ouvinte estar imerso no ambiente que foi simulado, ver Figura 1. Isso é muito útil, por exemplo, para projetar determinado ambiente, de forma que se possa prever o comportamento acústico da sala, mesmo sem ela existir. Há também o interesse em modelar uma sala existente, por exemplo, quando se deseja avaliar e otimizar a sua qualidade acústica.

Figura 1 - Sensação auditiva 3D



Fonte: O autor, 2017.

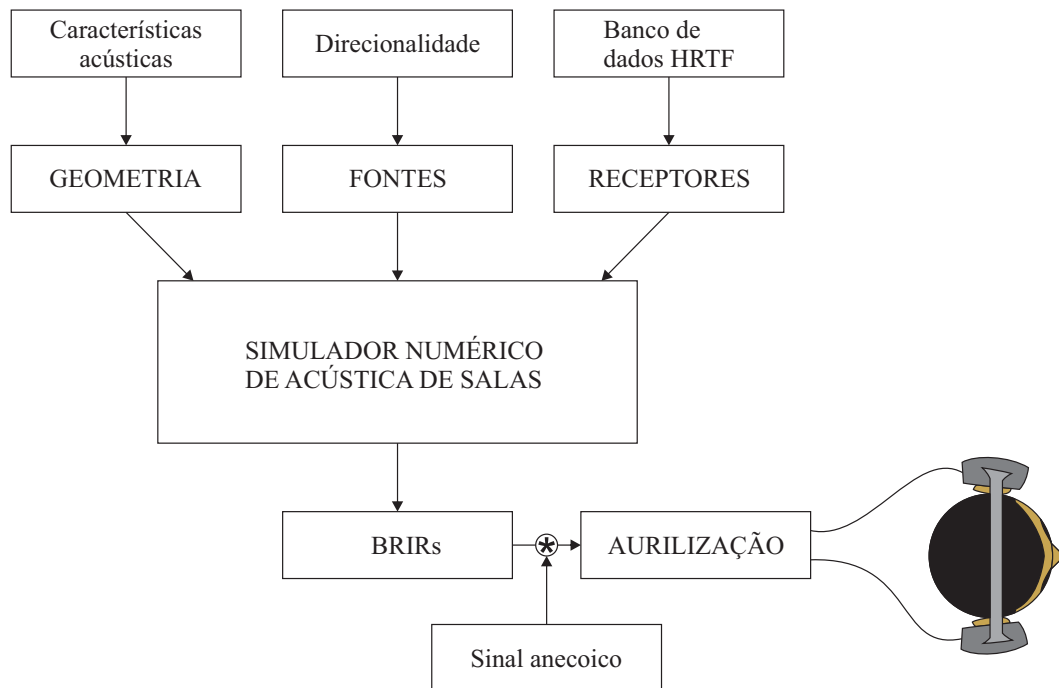
Para que um sistema de geração de aurilização apresente resultados satisfatórios

¹ O termo aurilização, utilizado neste trabalho, corresponde à tradução da expressão em inglês *auralization*, o qual foi denominado em 1993 por Kleiner et al. (1993). Uma outra tradução muito utilizada é auralização. Contudo, como observado em Tenenbaum (2017), o antepositivo *auri* está associado, segundo os léxicos, à audição.

é necessário que todos passos envolvidos sejam modelados com precisão. Tais passos são apresentados na Figura 2. Note que para obter a aurilização a partir de dados simulados é necessário dispor-se de um simulador de acústica de salas que forneça de forma adequada para um dado par fonte-receptor as respostas impulsivas biauriculares (BRIRs - na sigla em inglês, *Binaural Room Impulse Responses*), núcleo da aurilização (BLAUERT, 1997). Para tal fim, esse simulador necessita modelar as fontes, os receptores, as superfícies de contorno da sala com suas características acústicas e os diversos fenômenos intervenientes na propagação de ondas acústicas (RINDEL, 2000). Os algoritmos utilizados precisam ser capazes de simular os fenômenos, tais como reflexão, difusão (D'ANTONIO; COX, 1998a; D'ANTONIO; COX, 1998b), difração, absorção, transmissão etc. (RINDEL; CHRISTENSEN, 2003). Os fenômenos acústicos usualmente dependem da temperatura, da umidade relativa, da pressão atmosférica, da geometria da sala e das propriedades acústicas dos materiais, como coeficientes de absorção e espalhamento. Uma outra observação é que o receptor a ser modelado não é um microfone, mas sim uma pessoa. A modelagem desse receptor é feita por meio das funções direcionais conhecidas como respostas impulsivas associadas à cabeça (HRIR - na sigla em inglês, *Head Related Impulse Response*) ou das funções de transferência associadas à cabeça (HRTFs - na sigla em inglês, *Head Related Transfer Function*), estas dadas no domínio da frequência. São essas funções que conferem as características de filtragem resultante da inserção da cabeça humana no campo acústico, dependentes da direção de chegada das ondas sonoras e, além disso, elas diferem de pessoa para pessoa, pois o som que adentra no ouvido é função, também, da cabeça, do torso e do pavilhão auricular. Em outras palavras, pode-se dizer que as HRIRs (ou HRTFs) descrevem como uma onda sonora em uma determinada direção chega à entrada do canal auditivo, para cada ouvido. Assim, para cada direção de chegada de uma onda sonora existe uma par de HRIRs (uma para cada ouvido). As BRIRs são, então, obtidas a partir das informações das frentes de onda, obtidas na simulação acústica da sala, que atingem o receptor e das HRIRs. Por fim, na última etapa de um sistema de aurilização é realizada a convolução de um sinal monofônico e anecóico com as BRIRs. A saída obtida é, então, reproduzida utilizando-se fones de ouvido de boa qualidade e equalizados ou através de sistemas multicanais, tais como ambisonics.

Diante do que foi mencionado, dois elementos essenciais para a aurilização são a simulação da acústica da sala e a modelagem das HRTFs. Para a simulação da acústica da sala, como será explicado no Capítulo 1, são utilizados os métodos baseados na teoria ondulatoria, na acústica geométrica ou na acústica estatística. Cada metodologia tem as suas vantagens e desvantagens, tanto do ponto de vista da melhor representação do campo acústico, quanto do custo computacional.

Figura 2 - Esquema para gerar aurilização



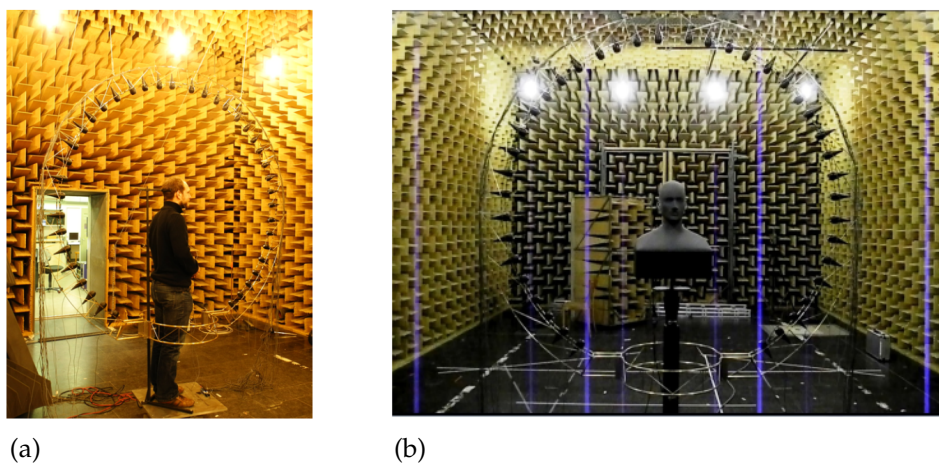
Fonte: Adaptado de TENENBAUM, 2016.

Há várias maneiras de se caracterizar as HRTFs. Uma opção seria através de medições (ZOTKIN et al., 2006). Nesse caso, essas funções são captadas por microfones instalados no ouvido do ouvinte que pode ser uma pessoa (Figura 3a) ou um manequim (Figura 3b). Para evitar a influência da sala, essas medições são realizadas em uma câmara anecóica, e para uma posterior aplicação, tais funções são armazenadas em um banco de dados discretos onde existe um par de funções para cada direção medida. Como essa função é contínua em todas as direções espaciais, logo, para encontrar as HRTFs nas posições não medidas, é necessário um método de interpolação, para evitar a degradação na qualidade do áudio. Esta pode ser feita no domínio do tempo ou da frequência. No primeiro caso obtém-se uma estimativa da HRIR, enquanto no segundo da HRTF (LANGENDIJK; BRONKHORST, 2000; FREELAND et al., 2004).

Outra maneira consiste na modelagem matemática. Na literatura podem ser encontradas diversas técnicas as quais estão fundamentadas na teoria de processamento de sinais, estatística, métodos numéricos, redes neurais artificiais etc. Alguns dos propósitos do desenvolvimento de metodologias que viabilizam a modelagem das HRTFs são:

- Evitar as medições (ou novas medições). Embora haja técnica de medições rápidas e precisas (RICHTER et al., 2016), ela necessita equipamentos especiais e requer pessoas especializadas;

Figura 3 - Medição das HRTFs



Legenda: (a) - Pessoa; (b) - cabeça artificial.

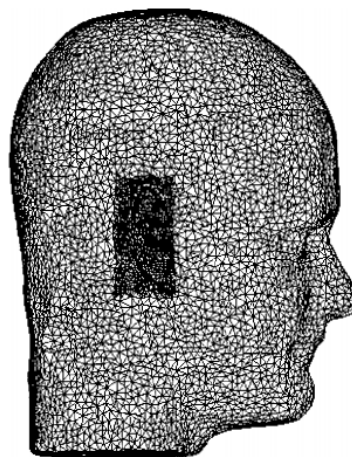
Fonte: MASIERO et al., 2011.

- Buscar solução mais acurada;
- Reduzir o custo computacional dos algoritmos de aurilização;
- Reduzir o tamanho do banco de dados de HRTFs.

O cálculo da HRTF baseada no método dos elementos de contorno (BEM - *Boundary Element Method*) parte da solução da equação de Helmholtz (GUMEROV; DURAISWAMI, 2005), que traduz o comportamento da onda sonora no domínio da frequência. Para isso, é necessária a discretização da cabeça, como mostra a Figura 4. O BEM corresponde à solução de um conjunto de equações, onde cada equação contém um componente para cada elemento na malha, e existe uma equação para cada elemento. Portanto, à medida que o tamanho da malha cresce, o tamanho do cálculo cresce como o quadrado do número de elementos. Para uma modelagem acurada, o método BEM requer uma amostragem muito densa e tempo de computação muito longo. A amostragem densa é necessária para se determinar a resposta em altas frequências (KATZ, 2001). Diante disso, diversos trabalhos foram publicados, tais como Algazi e Duda (2002), Otani e Ise (2006) e Gumerov et al. (2010), na tentativa de reduzir o custo computacional. Uma técnica numérica mais rápida que o BEM é o método das diferenças finitas no domínio do tempo (FDTD - *Finite-Difference Time-Domain*) (XIAO; LIU, 2003; PREPELITAA, 2016). Há também uma outra técnica baseada no princípio de superposição de ondas (KOOPMANN et al., 1989; WILLIAMS, 1999), a qual apresenta uma eficiência computacional superior ao BEM, proposta por Bai e Tsao (2006). O empecilho desses métodos é que não são acessíveis para qualquer usuário, uma vez que é necessário dispor de equipamentos especiais, como um *scanner* 3D e um sistema

de imagem por ressonância magnética para a modelagem da cabeça e do ouvido.

Figura 4 - Discretização da cabeça utilizada no BEM



Fonte: KATZ, 2001.

Sabendo que as HRTFs estão relacionadas com determinados parâmetros antropométricos, elas podem ser caracterizadas a partir de medidas geométricas (ALGAZI et al., 2001). Essas técnicas são denominadas como *métodos de regressão antropométrica*. Elas permitem modelar a HRTF baseada em um banco de dados de HRTFs previamente medidas e uma técnica de redução de dimensionalidade, como a *análise de componentes principais* (PCA - *Principal Components Analysis*) (KISTLER; WIGHTMAN, 1992; MIDDLEBROOKS; GREEN, 1992; CHEN et al., 1995; SCARPACI; COLBURN, 2005; NISHINO et al., 2007). Mas, devido à complexa relação entre as HRTFs e as múltiplas variáveis (direção, frequência, características antropométricas) os métodos lineares, como o PCA, não possuem a capacidade de representar tal relação. Para contornar essa dificuldade, técnicas como o *método tensor-SVD* (*Singular Value Decomposition*) (GRINDLAY; VASILESCU, 2007), *redes neurais* (HU et al., 2008; LI; HUANG, 2013), *algoritmos genéticos* (CHEUNG et al., 1998), *Isomap* (GRIJALVA et al., 2016), *Locally Linear Embedding* (LLE) (KAPRALOS; MEKUZ, 2007), *Support Vector Regression* (HUANG; FANG, 2009) têm sido utilizadas. Em Kapralos et al. (2008) faz-se um estudo comparativo entre o PCA, o LLE e o Isomap, onde conclui-se que o LLE e o Isomap apresentam desempenho melhor que o PCA. Cabe ressaltar, porém, que essas metodologias servem apenas para a obtenção da magnitude da HRTFs. Sendo a HRTF uma função complexa, essa é estimada, em geral, através da reconstrução de fase mínima (KISTLER; WIGHTMAN, 1992). Contudo, Møller et al. (1995) e Nam et al. (2008) mostram que essa aproximação não é satisfatória para todas as direções. Para evitar a aproximação da fase, Chen et al. (1995) propõem um modelo chamado de *spatial*

feature extraction and regularization model (SFER), o qual está embasado na *expansão de Karhunen-Loève*. Note-se que, embora esses métodos utilizem um banco de dados de HRTFs, eles evitam que novas medições sejam realizadas para novos indivíduos.

Uma outra técnica encontrada na literatura é a modelagem por polos e zeros. Essa técnica consiste em aproximar as HRTFs a partir de filtros FIR (*Finite Impulse Response*) ou IIR (*Infinite Impulse Response*) (BLOMMER; WAKEFIELD, 1997; HANEDA et al., 1999; LIU; HSIEH, 2001; KULKARNI; COLBURN, 2004). O principal objetivo, em geral, desses modelos é ajustar os coeficientes desses filtros de tal maneira que o erro de aproximação (medido através de uma norma) entre o modelo e a envoltória espectral seja o menor possível e, além disso, é requerido que a ordem do modelo também seja o menor possível, sem que as características de direcionalidade e conteúdo espectral gerem erros perceptíveis. Para minimizar esse erro, pode-se utilizar, por exemplo, uma técnica heurística, como algoritmos genéticos (DURANT; WAKEFIELD, 2002). Existem outros métodos que permitem tal ajuste dos coeficientes, como o *Prony*, o *Shanks* (CHEN; HSIEH, 2000) e as redes neurais artificiais (JENISON, 1995). Em (HACIHABIBOGLU et al., 2002) apresenta-se um método fundamentado nas *wavelets*, com o qual se consegue reduzir a ordem do filtro IIR. Um outro trabalho baseado nas *wavelets* é de Torres et al. (2004), o qual utiliza, além das transformadas *wavelets*, um conjunto de filtros esparsos. Nessa abordagem a HRIR é vista como um sistema FIR e a modelagem é realizada por um grupo de filtros esparsos, os quais processam os resultados das transformadas *wavelets*. Essa modelagem resulta em uma redução do número de coeficientes que implementam o modelo para aproximadamente 70% do tamanho original.

Uma outra alternativa às implementações tradicionais é a modelagem das HRTFs usando redes neurais artificiais. A solução de problemas por meio de RNAs é bastante atrativa, uma vez que as redes possuem a capacidade de aprender por meio de exemplos e de generalizar a informação e, portanto, obter um desempenho superior ao dos modelos convencionais (BISHOP, 2005). A generalização está associada à capacidade da rede responder corretamente a padrões que não fizeram parte do treinamento. Diante das diversas potencialidades das RNAs, pode-se encontrar uma gama de aplicações em diversas áreas, tais como reconhecimento de caracteres, reconhecimento de imagens, previsão do tempo, previsão financeira, mineração de dados etc. No que se refere às aplicações à acústica pode-se citar: síntese e reconhecimento de voz (RAHIM et al., 1992), predição de ruído e da qualidade do som em ambientes urbanos (GENARO et al., 2010; YU; KANG, 2005; YU; KANG, 2009), localização de fontes sonoras (STEINBERG et al., 1991) e predição de parâmetros de qualidade acústica (NANNARIELLO, 1999; NANNARIELLO et al., 2001).

Diferentemente dos trabalhos que foram citados anteriormente, onde as redes eram utilizadas apenas para auxiliar na estimativa das HRTFs, os trabalhos que serão citados utilizam como núcleo as redes neurais artificiais. A abordagem apresentada

em Jenison e Fissell (1996) utiliza a rede neural *von Mises Basis Function* (VMBF). Esta rede utiliza uma função de ativação denominada de *von Mises-Arnold-Fisher* (FISHER et al., 1987). A entrada da rede é composta de dois elementos, os quais representam o azimute e a elevação (ou seja, a direção). A saída possui seis neurônios, os quais correspondem às seis direções principais da magnitude da HRTFs extraída a partir do PCA. Para a reconstrução da HRTF utiliza-se a transformação linear inversa. Já a camada intermediária utiliza 25 neurônios. Para o treinamento foram utilizadas 450 HRTFs e o algoritmo de treinamento foi o gradiente decrescente. Embora, essa rede permita o cálculo da magnitude da HRTF para qualquer direção, a fase é perdida nessa metodologia. Além disso, ela não permite o cálculo para diferentes indivíduos (não-individualizadas).

Há uma outra metodologia, desenvolvida por Wu et al. (1998), que utiliza o mesmo tipo de rede utilizada em Jenison e Fissell (1996). A diferença é que, ao invés do cálculo da HRTF, a rede neural foca no cálculo da HRIR. Essa rede neural tem como entrada a direção em coordenadas esféricas, azimute e elevação, e a saída 10 componentes principais, os quais possuem características espaciais das HRIRs. Tal saída é então processada para obter a HRIR para uma determinada direção. Para gerar os padrões de treinamento foi utilizado o banco² de HRIRs medidos por Gardner e Martin (1995). Os padrões de saída foram obtidos usando a expansão de Karhunen-Loeve, para extrair as características espaciais das HRIRs. Dentre os 710 padrões gerados, 600 foram utilizados no treinamento e o restante para testes. Embora essas redes tenham sido projetadas para modelar a HRIR para qualquer direção, nenhum resultado foi apresentado por Wu. Além disso, não foi mencionada a quantidade de neurônios na camada intermediária. Vale ressaltar, que a técnica apresentada gera HRIRs não-individualizadas, uma vez que o banco de dados foi obtido de um manequim (KEMAR).

No entanto, a modelagem proposta por Hu et al. (2008) utiliza um conjunto de redes neurais artificiais do tipo *multilayer perceptron* (MLP) para estimar as HRTFs individualizadas. Cada rede neural é projetada para cada direção, tendo como entrada 8 parâmetros antropométricos e como saída 12 componentes principais (obtidos a partir do PCA) e a ITD (*interaural time difference*), ou seja, a saída da rede é composta de 13 neurônios. E dessas informações há um pós-processamento para obter a magnitude da HRTF. Finalmente, utilizando essa magnitude e a fase mínima, a HRTF é reconstruída. Todas as redes neurais possuem uma única camada intermediária com 16 neurônios. Os padrões de treinamento foram extraídos do banco de dados CIPIC (*Center for Image Processing and Integrated Computing*) da *California University* (ALGAZI et al., 2001). Esse banco possui as HRIRs de 45 indivíduos para 1250 direções e, além disso, as informações

² Disponível no site <http://sound.media.mit.edu/resources/KEMAR.html>

das medidas antropométricas. Contudo, foram escolhidas apenas as HRIRs de 37 indivíduos, sendo 13 de mulheres e 22 de homens, cujas idades variam de 18 a 63 anos, e 2 de manequins (KEMAR). Desses 37 “indivíduos”, 30 foram escolhidos para treinamento e 7 para testes. Como o banco de dados das HRTFs utilizado não é suficientemente discretizado, há a necessidade de um processo de interpolação, para a obtenção das HRTFs para as direções onde não há medições. Além disso, como ressalva dos próprios autores, em função da baixa diversidade de indivíduos no processo de treinamento, não se pode garantir que a rede terá a capacidade de generalização. Uma outra observação é a necessidade de averiguar outras técnicas de análise estatística multivariada, além do PCA, o qual foi utilizado.

Um segundo método de personalização de HRTFs é relatado em Li e Huang (2013), o qual segue a mesma linha do artigo do Hu et al. (2008). Sendo assim, a rede neural tem como entrada os parâmetros antropométricos e como saída a compressão da HRTF. No entanto, o número de elementos na entrada e na saída são diferentes. Na entrada foram utilizados 12 e na saída 20. Além dessa diferença, a maneira como foram obtidas essas variáveis se distinguem. Para a seleção das variáveis de entrada utilizaram-se o *Laplacian score* (HE et al., 2005) e a análise de correlação, e para a saída o *high-order singular value decomposition* (HOSVD) (LATHAUWER et al., 2000), uma generalização da decomposição de uma matriz em valores singulares (SVD). Por outro lado, foi utilizada a mesma quantidade de neurônios na camada intermediária, ou seja, 16 neurônios. Também, utilizou-se o mesmo banco de dados de HRTFs. Nesse trabalho é apresentada uma comparação entre a técnica utilizada em Hu et al. (2008), ou seja o PCA, e o HOSVD, de onde se conclui que o HOSVD tem uma melhor performance que o PCA. Outro resultado comparativo demonstra que a rede RBF possui resultados mais acurados que a rede MLP. No entanto, os autores ressaltam que há um desvio maior nos entalhes da HRTF.

Note-se que, até o momento, todas as técnicas apresentadas visam apenas a modelagem das HRTFs ou das HRIRs. Como o cálculo das BRIRs envolvem não só a modelagem das HRIRs (ou HRTFs), mas também, em uma etapa posterior, o cômputo da HRIR filtrada. Isso corresponde ao produto de convolução entre o raio sonoro que chega ao receptor e a HRIR da direção correspondente. A não redução do custo computacional dessa etapa torna o sistema de auralização extremamente lento, devido à grande quantidade de raios que chegam no receptor. Posto isso, Naranjo (2014) propôs uma metodologia que é capaz de modelar não só as HRIRs, mas também as HRIRs filtradas e a interpolação. Este procedimento está baseado nas redes neurais artificiais do tipo MLP. Para cobrir todo o espaço auditivo de recepção, uma estrutura com múltiplas redes foi utilizada, onde cada rede opera em uma região específica do espaço, denominada de *gomo*. Os detalhes dessa estrutura são comentados no Capítulo 4. Através dessa abordagem obteve-se um ganho computacional de 62%. No

entanto, foram observados erros de modelagem na região de alta frequência. Também, utilizou-se uma quantidade pequena de amostras das HRIRs, a saber 128 amostras, que corresponde a uma baixa resolução no domínio da frequência, como será discutido no Capítulo 4.

Este trabalho tem como objetivo principal desenvolver uma modelagem que permita recuperar as HRIRs filtradas com uma precisão superior à apresentada em Naranjo (2014), com um menor custo computacional. Então, foi desenvolvido neste trabalho um conjunto de redes neurais artificiais do tipo RBF (*Radial Basis Function*) (BROOMHEAD; LOWE, 1988) especializadas na geração de HRIRs filtradas. Cada rede está associada a uma direção de chegada e possui como entrada os componentes do espectro da frente de onda, em bandas de oitava, e, como saída, a HRIR filtrada. Para verificar a eficiência das redes, foram geradas as respostas impulsivas biauriculares para alguns pares fonte-receptor para uma das salas utilizadas no projeto de intercomparação de simuladores de acústica de salas, denominado *Round Robin 4* (RR4) (<http://rr.auralisation.net/downloads/>), promovido pelo *Audio Communication Group da Universidade Técnica de Berlim* (TU Berlin) e pelo *Institute of Technical Acoustics* (RWTH Aachen), Alemanha. As BRIRs construídas utilizando as redes neurais artificiais são então comparadas com as medidas.

Como a geração das BRIRs requer um simulador de acústica de salas, nesta tese utilizou-se o código computacional RAIOS (*Room Acoustics Integrated and Optimized Software*) (TENENBAUM et al., 2006a; TENENBAUM et al., 2006b). Esse simulador utiliza o *método do traçado de raios* para modelar o som direto e as reflexões especulares (KULOWSKI, 1984; EMBRECHTS, 1982) e o *método de transição de energia* para modelar as reflexões difusas (KURZINS; FRICKE, 1982; ALARCÃO et al., 2000). Desse modo, ao longo da pesquisa, foram necessárias as seguintes modificações principais do código computacional RAIOS:

- O ajuste e reparo de diversos *bugs* do código;
- A extensão do código para trabalhar com nove bandas de oitava (de 63 Hz a 16 kHz);
- A inclusão das tabelas de materiais com os coeficientes de absorção e espalhamento, por bandas de oitava;
- A implementação de um novo banco de dados das direcionalidades de fontes sonoras utilizadas no *Round Robin 4*;
- A implementação do cálculo das BRIRs;
- A inclusão da parte do cálculo das reflexões difusas (matrizes de transição) no cômputo das BRIRs.

Organização do trabalho

No Capítulo 1, intitulado “Acústica de salas”, são apresentados os fundamentos teóricos que embasam a acústica de salas. É feita uma breve descrição de alguns fenômenos da propagação do som, tais como a reflexão, absorção, difusão, atenuação, e como elas podem ser modeladas. Além disso, são apresentados alguns conceitos básicos de processamento de sinais.

No Capítulo 2, intitulado “Aurilização de salas”, são apresentados os passos para se gerar a aurilização a partir de dados simulados. Cada passo está descrito em uma seção, para uma melhor compreensão. A metodologia apresentada nesse capítulo, corresponde ao assim chamado método tradicional, enquanto a modelagem utilizando as redes neurais artificiais, objetivo central desta tese, será explicada no Capítulo 4.

No Capítulo 3, intitulado “Redes neurais artificiais”, são apresentados os principais conceitos de redes neurais artificiais. Além das redes de funções de base radiais, utilizadas na metodologia proposta nesta tese, são abordadas, também, as redes *perceptron* de múltiplas camadas. A descrição destas faz-se necessário, uma vez que, este trabalho foi motivado a partir da tese de Naranjo (2014), que as utiliza.

No Capítulo 4, intitulado “Modelagem de HRIRs filtradas com RNAs”, são apresentadas as etapas da modelagem de HRIRs filtradas utilizando as redes de funções de base radiais, e como essas redes são utilizadas na geração da aurilização. Também é apresentada brevemente a abordagem utilizada em Naranjo (2014), a fim de mostrar as principais dificuldades encontradas e como elas foram contornadas aqui.

No Capítulo 5, intitulado “Resultados do *Round Robin on Auralization*”, são apresentados os resultados obtidos utilizando as formulações apresentadas nos Capítulos 2 e 4, bem como os resultados experimentais.

Por fim, o último capítulo apresenta as conclusões e propostas para trabalhos futuros.

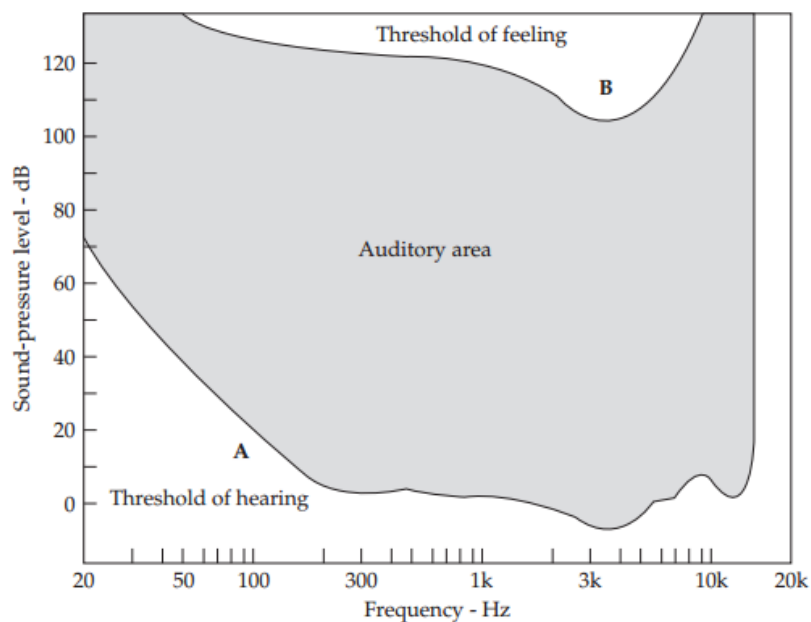
1 ACÚSTICA DE SALAS

Um dos objetivos deste capítulo é fornecer informações sobre o comportamento da propagação do som em salas e como suas propriedades podem ser quantificadas e analisadas. Além do mais, esses estudos formam uma base para o entendimento de um sistema de audição. Como neste trabalho as respostas impulsivas obtidas a partir de um simulador são confrontadas com as medidas, faz-se necessário apresentar alguns aspectos relativos às medições das respostas impulsivas.

1.1 Introdução

Diz-se que existe um som se uma perturbação propagada através de um material elástico provoca uma alteração na pressão e um deslocamento das partículas do material que podem ser detectadas por uma pessoa (BERANEK, 1954). Para que uma fonte sonora seja percebida pelo ouvido humano, esta deve vibrar na faixa de 20 Hz a 20 kHz. Isso corresponde a ondas sonoras com comprimento de onda entre 1,7 cm e 17 m, já que a velocidade dessas ondas é constante e igual a 343 m/s a uma temperatura de 20°C. Além disso, a sensibilidade do ouvido varia com a frequência da onda sonora. Isto significa que sons com o mesmo nível, porém com frequências diferentes, causarão distintas impressões subjetivas de intensidade. As curvas A e B da Figura 5 foram obtidas a partir de ouvintes treinados, os quais deveriam julgar quando um determinado tom puro (som numa única frequência), em uma dada frequência da faixa audível, com determinado nível sonoro se tornava audível (curva A) e desconfortável (curva B). A curva A é o limiar da audição. Essa curva mostra que o sistema auditivo é mais sensível em torno de 3 kHz, ou seja, tons puros com frequência nessa região estimulam a audição com níveis sonoros menores do que em outras frequências. Nessa faixa de maior sensibilidade, um nível de pressão sonora de 0 dB (20 μ Pa) é suficiente para estimular o sistema auditivo. Por outro lado, a curva B representa o limite de desconforto auditivo. Para frequências na faixa de maior sensibilidade, essa sensação é sentida em torno de 110 dB, e a partir de 120 dB para as baixas e altas frequências. Já a região compreendida entre as curvas A e B é a *área de audição*. Todos os sons que os seres humanos experimentam devem ser tais que a frequência e o nível se enquadrem nessa área auditiva. Um outro aspecto subjetivo sobre o som é que os seres humanos são capazes de perceber a direção de uma fonte sonora. Esta habilidade é possível devido à presença dos dois ouvidos, e que o córtex auditivo é capaz de captar diferenças de amplitude e tempo de chegada do som entre os ouvidos (RUMSEY, 2001) (mais detalhes serão explicados no Capítulo 2).

Figura 5 - Área de audição dos seres humanos



Legenda: (curva A) - limiar da audição humana; (curva B) - limiar do desconforto auditivo.

Fonte: EVEREST; POHLMANN, 2009.

A acústica de salas caracteriza o comportamento do som em determinado ambiente, como salas de aula, teatros, cinemas, salas de concertos, estúdios de gravação etc. O estudo do campo sonoro em uma sala requer a consideração de diversas variáveis, tais como a geometria da sala, a absorção acústica, as reflexões nas superfícies e as fontes sonoras. Sabendo que cada sala possui suas próprias características acústicas, então um dos objetivos da acústica de salas é projetar ambientes que satisfaçam a sua função acústica.

Por exemplo, as fábricas devem ter um nível de ruído aceitável (BISTAFA, 2011), salas de aula devem garantir a inteligibilidade, uma vez que a fonte sonora principal é a voz do professor e, além disso, a qualidade acústica adequada das salas de aulas está relacionada à performance dos alunos (BISTAFA; BRADLEY, 2000; SEEP, 2002; RONSSE, 2011; CHOI et al., 2014). As salas de concerto devem imprimir uma boa qualidade acústica à música, já que são ambientes destinados apenas à música; para isso, o projetista deve se preocupar com o volume da salas, com o formato do palco, com as posições dos refletores nos tetos e nas paredes laterais e, além disso, com o número de assentos (BERANEK, 2004; ANDO, 2012; BRANDÃO, 2016).

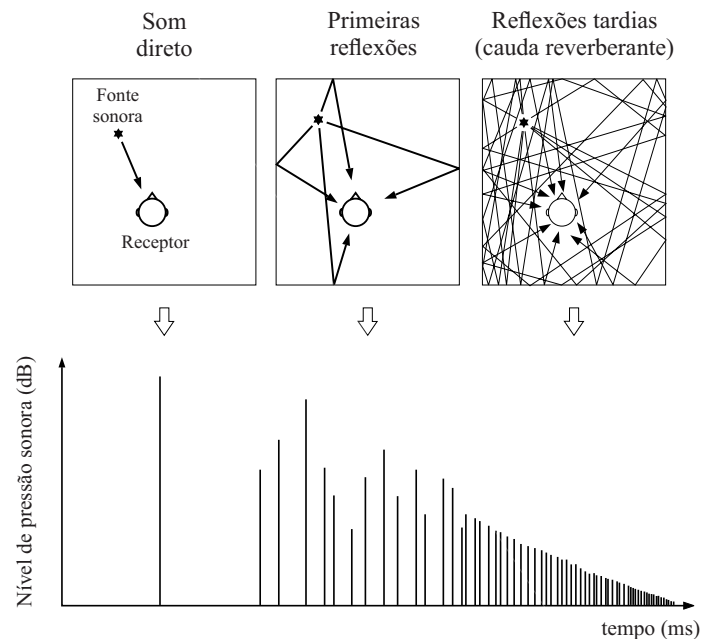
1.2 Campo sonoro

Um campo sonoro é uma região do espaço onde as ondas sonoras se propagam, ou seja, onde ocorrem as vibrações acústicas das partículas do meio elástico, sendo definido pela pressão sonora do meio e pela velocidade das partículas. Ele caracteriza a relação das ondas sonoras com o meio, que, por sua vez, depende da fonte sonora, da distância em relação à fonte e dos obstáculos encontrados na trajetória dessas ondas.

O som gerado em uma sala que chega ao receptor é a composição do som direto, ou seja, aquele que parte diretamente da fonte e chega ao receptor sem influência das superfícies, e as subsequentes reflexões sofridas pela onda sonora. A representação gráfica temporal da chegada das várias reflexões, acompanhadas do seu nível de energia correspondente, é chamada de *reflectograma*. A Figura 6 mostra a chegada de diferentes frentes de onda (representadas por raios sonoros) a um receptor juntamente com o reflectograma associado, com indicação do *som direto*, a zona das *primeiras reflexões* e a área de *reflexões tardias* (ou cauda reverberante). A zona das primeiras reflexões corresponde à região entre o som direto e 50 ms (para a análise da inteligibilidade da fala) ou 80 ms (para música) após aquele. Essas reflexões reforçam o som direto em termos de amplitude, ou seja, elas são percebidas pelo sistema auditivo como se estivessem integradas ao som direto e contribuem para a percepção espacial do som (direção, distância etc.). Por outro lado, as reflexões tardias são percebidas como uma extensão do som gerado pela fonte, cujo efeito é aumentar a sua duração. Essa persistência do som numa sala é a sua *reverberação*. A sua duração é descrita pelo *tempo de reverberação*, o que será abordado na Seção 1.5. Não se pode dizer que a reverberação é uma característica positiva ou negativa, pois depende dos fins a que se destina a sala. Por exemplo, em salas de aula e teatros uma curta reverberação é imprescindível já que a inteligibilidade é fundamental; em contrapartida, as salas de concerto requerem uma certa reverberação, de modo a fornecer suporte à música que nela é executada.

De modo geral, a energia do som direto é maior do que a das primeiras reflexões, uma vez que a distância percorrida por estas é maior e, além disso, existe uma perda de energia devido à incidência destas sobre as superfícies, as quais absorvem parte da energia radiada pela onda incidente. Observe que há um silêncio entre o tempo de chegada do som direto e a primeira reflexão. Esse intervalo de tempo é chamado de *intervalo de tempo de atraso inicial* (ITDG - *Initial Time Delay Gap*). Quando um receptor está distante da fonte e próximo a uma superfície refletora, a tendência é que o ITDG seja menor. Por outro lado, se o receptor está próximo à fonte e distante de uma superfície refletora, a tendência é que o ITDG seja maior. Note-se também que a energia das ondas refletidas que chegam posteriormente à primeira não é necessariamente menor, dado que a energia dessas dependem da energia absorvida pela superfície e não só da distância de propagação. Uma outra observação é que à medida que o tempo progride,

Figura 6 - Distribuição temporal da energia sonora para um determinado par fonte-receptor



Fonte: ISBERT, 1998.

maior é a quantidade de reflexões que chegam ao receptor. Em função disso, maior é a densidade temporal da cauda do reflectograma.

Discute-se, agora, como o som direto se relaciona com a reverberação em uma sala. O campo sonoro em torno da fonte na sala pode ser subdividida em dois campos (BISTAFA, 2011): próximo e afastado. O campo próximo é a região perto da fonte onde o nível de pressão sonora varia consideravelmente com uma pequena alteração da posição do receptor. A região se estende a uma distância menor que o comprimento de onda da menor frequência emitida pela fonte, ou do dobro da maior dimensão da fonte, a maior distância é a que prevalece. As medidas de pressão sonora nessa região devem ser evitadas, já que o nível de pressão tende a ser errático (BRANDÃO, 2016). O campo afastado, por sua vez, é dividido em campo livre e campo reverberante. A extensão desses campos depende das dimensões da sala e da absorção sonora de cada superfície; no campo livre o som direto é predominante, enquanto no campo reverberante o som refletido é predominante.

A base para a compreensão dos estudos relacionados aos problemas da acústica de salas decorre da representação de quatro dimensões: tempo, frequência, espaço e cognitiva. A variável tempo é importante, pois a compreensão de como a pressão sonora varia em relação à variável tempo determina como percebemos o som em uma

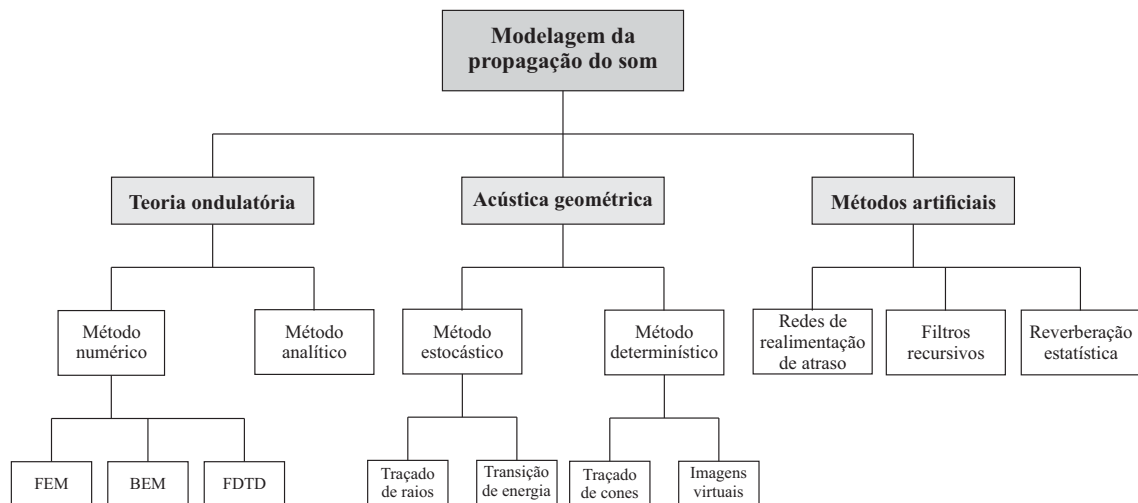
determinada sala. Já a frequência deve-se ao fato de que os domínios do tempo e frequência estão relacionados, sendo essa relação dada pela transformada de Fourier (ver Seção 1.6.1). Sabendo que a resposta temporal também é alterada com o tipo de revestimentos interiores, o tratamento acústico de uma sala exige o conhecimento de como os materiais se comportam em relação à frequência. Por exemplo, materiais porosos são excelentes absorventes de sons de alta frequência. Assim, se uma sala é tratada apenas com esse tipo de materiais, apresentará uma resposta em frequência pobre nos agudos. Um outro fator que faz com que a resposta temporal varie é a posição relativa entre a fonte e o receptor, posto que a mudança da posição da fonte ou do receptor altera as distâncias percorridas pelas ondas sonoras, a direção de chegada das frentes de onda ao receptor, as superfícies atingidas pelas ondas sonoras etc. Essa mudança afeta principalmente a parte inicial do reflectograma (som direto e as primeiras reflexões), uma vez que a quantidade reflexões é tão grande (alta densidade) na parte final que a mudança da fonte e do receptor não alteram substancialmente a cauda reverberante. Daí a relevância da variável espaço. Por fim, a dimensão cognitiva está associada ao modo como a experiência subjetiva do ouvinte é alterada com as dimensões tempo-frequência-espaço. Para isso, é necessário a definição de métricas para avaliar a experiência subjetiva. Foram então criados diversos parâmetros objetivos, os quais estão correlacionados com a experiência subjetiva, sendo a maioria definido direta ou indiretamente a partir da resposta impulsiva da sala. Contudo, alguns parâmetros são obtidos da resposta impulsiva biauricular. Esses são chamados de *parâmetros biauriculares*, enquanto os demais são designados como *parâmetros monoauriculares*. Como exemplos de parâmetros monoauriculares tem-se: tempo de reverberação, tempo de decaimento inicial (EDT - Early Decay Time), fator de clareza, definição, tempo central, fator de ganho e fração de energia lateral (LF - Lateral Energy Fraction). Como exemplo de parâmetro biauricular tem-se a correlação cruzada interauricular (IACC - Inter-Aural Cross-Correlation Coefficient) etc. Para detalhes sobre esses e outros parâmetros objetivos, o leitor pode consultar (BARRON; LEE, 1988; BRADLEY; SOULODRE, 1995; BERANEK, 2004; VORLÄNDER, 2008; BRANDÃO, 2016). Os conceitos de resposta impulsiva mono e biauricular são abordados na Seção 1.6.2. Como se pode observar, o conhecimento da resposta impulsiva é um dado fundamental para a avaliação da qualidade acústica de uma sala. Como a resposta impulsiva muda com a posição da fonte e do receptor, conseqüentemente esses parâmetros também serão alterados. É desejável, na maioria dos casos, que a sala tenha um campo acústico com propriedades semelhantes em todo o espaço, pois assim os parâmetros não terão uma variação significativa com a posição. Uma maneira de minimizar essas variações é utilizando difusores (D'ANTONIO; COX, 2000; COX; D'ANTONIO, 2004). O motivo pelo qual essas métricas são obtidas da resposta impulsiva (mono ou biauricular) será abordado na Seção 1.6. Vale ressaltar que um conjunto de parâmetros faz-se ne-

cessário para essa avaliação, já que um único parâmetro não pode quantificar todas as experiências subjetivas.

1.2.1 Modelagem do campo sonoro

O campo acústico pode ser modelado utilizando diferentes métodos. A Figura 7 ilustra a classificação dos métodos para simulação da propagação sonora. Sendo as duas principais abordagens baseadas: na *teoria ondulatória* e na *acústica geométrica* (SAVIOJA; SVENSSON, 2015).

Figura 7 - Classificação dos métodos para simulação da propagação sonora



Fonte: Adaptado de VORLÄNDER, 2008.

Na acústica ondulatória, busca-se a solução da *equação da onda* dada por

$$\nabla^2 p(\mathbf{r}, t) = \frac{1}{c^2} \frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2}, \quad (1)$$

onde $p(\mathbf{r}, t)$ é a pressão sonora em função da coordenada espacial $\mathbf{r} = (x, y, z)$ e da coordenada temporal t , c é a velocidade do som no meio e ∇^2 é o operador Laplaciano. A solução da Eq. (1) pode ser obtida por meio de um método analítico, para problemas simples, ou de um método numérico, para problemas complexos. A equação da onda é derivada a partir das leis da conservação da massa, da conservação da quantidade de movimento e da termodinâmica, cujos detalhes podem ser encontrados em Kinsler et al. (2000) e Bruneau (2006). A complexidade da solução da equação da onda depende da geometria, das condições de contorno e das condições iniciais do problema. Em geral, os problemas de interesse são complexos e a Eq. (1) não poderá ser

resolvida analiticamente, logo, deve-se recorrer a alguma técnica numérica, tais como o método dos elementos de contorno e o método dos elementos finitos. Entretanto, esses métodos tornam-se computacionalmente custosos à medida que a frequência de análise aumenta.

Por outro lado, na acústica geométrica todas as propriedades da teoria ondulatória são ignoradas, e considera-se a propagação sonora através de *raios sonoros* (em similaridade a raios luminosos), cuja propagação é retilínea, já que a velocidade da onda é considerada como constante se não houver gradientes de temperatura. Estas considerações são válidas para altas frequências, pois o comprimento de onda é menor que as dimensões da sala e dos objetos em seu interior (SCHUSTER; WAETZMANN, 1929). Como exemplo, pode-se destacar: o método de traçado de raios (KROKSTAD et al., 1968; KUTTRUFF, 1971), o método de imagens virtuais (ALLEN; BERKLEY, 1979; ZAGAR, 1983; BORISH, 1984; DANCE; SHIELD, 1997) e o método de traçado de cones ou pirâmides (LEWERS, 1993; FARINA, 1995; STEPHENSON, 1996). Embora as técnicas fundamentadas na acústica geométrica sejam mais rápidas, elas apresentam resultados menos acurados para baixas frequências, onde predominam os modos da sala.

Portanto, como a região de frequência da audição é ampla, é útil a divisão desse espectro para a aplicação onde cada método oferece vantagens. De acordo com o que foi mencionado dos métodos citados acima, pode-se utilizar a acústica ondulatória para análise das baixas frequências e a acústica geométrica para médias e altas frequências. Segundo Pelzer et al. (2011) essa abordagem híbrida fornece resultados bastante precisos quando comparados aos resultados medidos. No entanto, incertezas, que são inevitáveis na caracterização do comportamento acústico das fontes, dos receptores, dos materiais etc. do mundo real, fazem com que os resultados simulados sempre destoem, em maior ou menor medida, dos experimentais.

1.3 Interação onda-superfície

Quando o som se propaga em uma sala, ele atingirá diversas superfícies. E ao incidir sobre uma superfície, dois fenômenos ocorrem: absorção e reflexão. Com relação à absorção, esta se decompõe em duas parcelas: a energia que é dissipada e energia que é transmitida. A porcentagem da energia que é absorvida é dada pelo *coeficiente de absorção α (absorption coefficient)*, o qual é definido pela razão entre a energia sonora absorvida e a energia sonora incidente. Seu valor está relacionado com as propriedades físicas do material e, além disso, varia com a frequência e apresenta valores entre 0 (material totalmente reflexivo) e 1 (material totalmente absorvente). Já no caso da reflexão, ela pode ser decomposta em reflexão especular e difusa. A reflexão

especular ocorre quando o ângulo com reta normal à superfície da onda incidente é igual ao ângulo com reta normal à superfície da onda refletida estando todos no mesmo plano. Na reflexão difusa a onda incidente é refletida para todas as direções. Esse fato está relacionado com a dimensão finita e às irregularidades das superfícies. Essas reflexões também podem ocorrer em superfícies lisas, ou seja, sem irregularidades, desde que tenham uma impedância acústica não-uniforme (COX; D'ANTONIO, 2004). Para mensurar a quantidade de energia refletida de forma difusa utiliza-se o *coeficiente de espalhamento* δ (*scattering coefficient*). Esse coeficiente foi estabelecido no ano de 2000 por Vorländer e Mommertz (VORLÄNDER; MOMMERTZ, 2000) e possui uma grande importância para os *softwares* de acústica de salas que utilizam a acústica geométrica (DALENBACK et al., 1994; COX et al., 2006). Ele mede a relação entre a energia refletida de forma difusa com a energia total refletida pela superfície. E, assim como o coeficiente de absorção, o coeficiente de espalhamento também varia com a frequência e o tipo de material. A Figura 8 ilustra a quantidade de energia que será radiada em cada uma das formas. De acordo com a definição do coeficiente de espalhamento, este não especifica como as diversas superfícies espalharão o som no espaço. Uma maneira de quantificar o quanto de energia será espalhada em cada direção, consiste em utilizar a *lei do cosseno de Lambert* (NOSAL et al., 2004; VORLÄNDER, 2008). Essa direcionalidade é função somente do ângulo, β , de elevação em relação ao plano de reflexão (ver Figura 8), a qual é dada por

$$S_{\beta} = 2 \sin \beta. \quad (2)$$

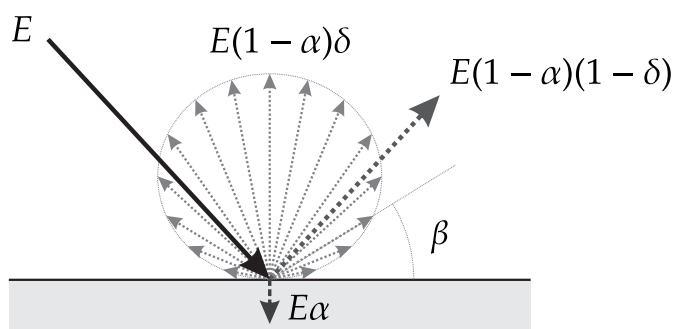
A técnica de medição do coeficiente de espalhamento pode ser consultada em Vorländer e Mommertz (2000).

Vale destacar que, há uma carência de dados quanto ao coeficiente de espalhamento dos diversos tipos de superfície utilizados em acústica de salas, já que é um parâmetro relativamente recente. A escassez desses valores tabelados faz com que o projetista estime tais coeficientes (BRANDÃO, 2016).

1.4 Absorção sonora

A onda sonora, ao se propagar, não só sofre perda energética na reflexão, como também pela absorção do ar. Existem vários fatores que participam da absorção do som no ar (KUTTRUFF, 2009). Uma das principais causas é provocada pela dissipação de energia no processo de relaxamento vibratório das moléculas de oxigênio e nitrogênio e, além disso, também pode ser originada pela combinação dos efeitos de viscosidade e condução de calor. A absorção de som no ar depende principalmente da distância,

Figura 8 - Distribuição da energia do raio incidente pela absorção, reflexão especular e espalhamento



Nota: Energia do raio incidente E , refletido especularmente $E(1-\alpha)(1-\delta)$ e difusamente $E(1-\alpha)\delta$, e absorvida $E\alpha$, onde α e δ são os coeficientes de absorção e espalhamento, respectivamente

Fonte: Adaptado de CAMILO, 2003.

temperatura e umidade (HARRIS, 1966). Expressões analíticas para a absorção no ar, também conhecida como atenuação do ar em função da umidade relativa, da temperatura e da frequência foram publicadas em Bass et al. (1972) e ISO 9613-1 (1993).

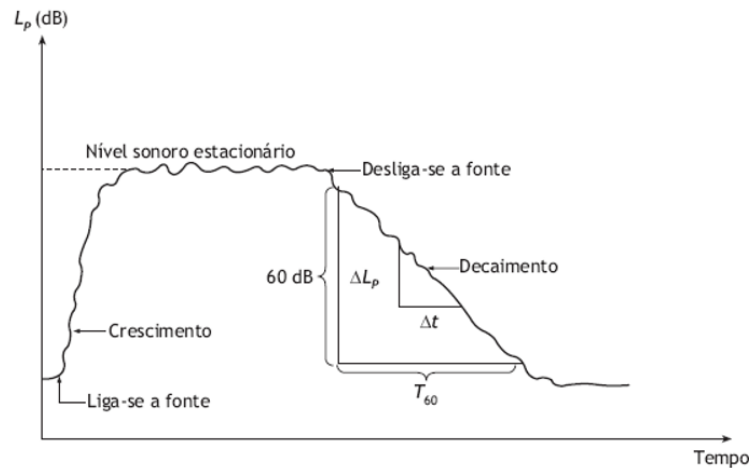
Na acústica geométrica a absorção do ar é levada em conta por meio do coeficiente de atenuação γ [m^{-1}] (VIAN; MARTIN, 1992). O fator da contribuição da absorção do ar no cálculo da energia do raio, após percorrer uma distância d desde a sua emissão pela fonte sonora, é dada por $\exp(-\gamma d)$, onde d é a distância percorrida, em metros.

1.5 Tempo de reverberação

Ao ligar uma fonte sonora em uma sala, a energia sonora aumenta até alcançar um estado de equilíbrio. Nessa situação, tem-se um nível sonoro estacionário, visto que a energia gerada pela fonte na sala está balanceada pela energia sonora absorvida pelas superfícies interiores da sala e todos os elementos nela contidos. Desligando a fonte sonora, o nível sonoro diminui e a taxa com que ocorre esse decaimento depende da absorção sonora e do volume da sala. Associado a esse decaimento, caracteriza-se o *tempo de reverberação* (T_{60}). Segundo a definição de Wallace C. Sabine (SABINE, 1922), físico norte-americano, considerado o “pai da acústica arquitetônica”, esse parâmetro corresponde ao intervalo de tempo necessário para se verificar um decaimento do nível sonoro de 60 dB (o qual corresponde a um decaimento de 1 milionésimo da energia no estado estacionário), após o desligamento da fonte sonora. Como o tempo de reverberação é função da absorção da sala, esse não é uniforme para toda gama de frequência, dado que cada material apresenta resposta diferente para cada frequência.

A Figura 9 ilustra o crescimento e o decaimento do nível sonoro em salas, com a indicação do tempo de reverberação.

Figura 9 - Crescimento e decaimento sonoro em salas



Fonte: BISTAFA, 2011.

O tempo de reverberação pode ser avaliado por relações matemáticas ou por meio de medições. A primeira expressão, formulada por Sabine, a qual foi obtida de forma empírica, e que posteriormente realizou uma análise teórica publicada em seu livro *Collected papers on acoustics* (SABINE, 2011) é dada por

$$T_{60}(f) = \frac{0,161V}{A\bar{\alpha}(f)} [s], \quad (3)$$

onde f é a frequência, V é o volume da sala em m^3 e $A\bar{\alpha}(f)$ é a absorção total, sendo $\bar{\alpha}(f)$ o coeficiente de absorção médio, o qual corresponde à média ponderada dos coeficientes de absorção $\alpha_i(f)$, cujos pesos são as áreas A_i das superfícies internas que compõem a sala, isto é,

$$\bar{\alpha}(f) = \frac{1}{A} \sum_i A_i \alpha_i(f), \quad (4)$$

onde $A = \sum_i A_i$ é somatório das áreas.

Cabe aqui fazer algumas observações quanto às Eqs. (3) e (4):

- Observe-se que o T_{60} é proporcional ao volume da sala e inversamente proporcional à quantidade de absorção. Sendo assim, o aumento do volume de uma sala provoca um aumento no T_{60} e, por outro lado, a inserção de materiais absorventes irá reduzir o T_{60} .

- A Eq. (3) negligencia a absorção do ar. Pode-se demonstrar, a partir da teoria estatística, que

$$T_{60}(f) = \frac{0,161V}{A\bar{\alpha}(f) + 4\gamma V} , \quad (5)$$

onde γ é o coeficiente de atenuação do ar, dado em $[\text{m}^{-1}]$. Essa equação expressa a fórmula de Sabine incluindo os efeitos da atenuação do ar.

- O desenvolvimento das Eqs. (3) e (5) assume que a absorção sonora é uniforme em toda a sala, isso fica implícito no $\bar{\alpha}$, dado que a Eq. (4) não informa como a absorção é distribuída na sala.

A partir dessas e outras questões, diversas fórmulas para o cálculo do tempo de reverberação foram desenvolvidas, tais como as fórmulas de Eyring (EYRING, 1930), Millington (MILLINGTON, 1932), Sette (SETTE, 1933), Fitzroy (FITZROY, 1958), Kuttruff (KUTTRUFF, 2009) etc. Diante dessa grande variedade, torna-se difícil a análise completa e a escolha da melhor fórmula. Segundo Brandão (2016), é recomendado o uso de pelo menos duas ou três fórmulas para a predição do T_{60} na fase de projeto. Atualmente, com a introdução dos simuladores, a estimação do tempo de reverberação em salas pode ser feita com muito maior precisão.

1.6 Processamento de sinais

A representação, a transformação e a manipulação de sinais e, além disso, a extração de informações que os sinais contêm, são assuntos tratados em processamento de sinais (OPPENHEIM; SCHAFER, 2010). Entende-se por sinal uma função que representa uma quantidade física, a qual pode ser medida e contém informações sobre o comportamento de um fenômeno. Comumente essa quantidade varia com o tempo, embora possa variar com outro parâmetro, como o espaço. O sinal de interesse deste trabalho é a pressão acústica, já que essa é a principal variável de estudo dos sistemas de aurilização de salas. Uma vez que esse tema exige o conhecimento de conceitos da área de processamento de sinais, esta seção tem como finalidade introduzir conceitos de processamento de sinais que serão fundamentais para o entendimento do assunto que será abordado no Capítulo 4.

1.6.1 Transformada de Fourier

Como foi discutido nas seções anteriores, a análise dos sinais no domínio da frequência é importante em acústica de salas e está ligado com o domínio do tempo pela transformada de Fourier. A transformada de Fourier é uma operação matemática aplicada a um sinal no domínio do tempo (ou espaço) para extrair seu conteúdo em frequência (ou número de onda).

Denomina-se $S(f)$ a *transformada de Fourier* de $s(t)$, sendo representada da seguinte forma

$$S(f) = \int_{-\infty}^{\infty} s(t) \exp(-j2\pi ft) dt, \quad (6)$$

e $s(t)$ a *transformada de Fourier inversa* de $S(f)$, a qual é dada por

$$s(t) = \int_{-\infty}^{\infty} S(f) \exp(j2\pi ft) df, \quad (7)$$

onde t [s] é o tempo, f [Hz] é a frequência, $j = \sqrt{-1}$ é a unidade imaginária e $\exp(\cdot)$ é a função exponencial natural. Pode-se indicar esse par de transformadas utilizando as seguintes notações:

$$S(f) = \mathfrak{F}\{s(t)\}; \quad (8)$$

$$s(t) = \mathfrak{F}^{-1}\{S(f)\}. \quad (9)$$

Como $S(f)$ é uma função complexa, ela pode ser representada na sua forma polar, isto é,

$$S(f) = |S(f)| \exp(j\angle S(f)), \quad (10)$$

onde $|S(f)|$ e $\angle S(f)$ são denominados o *espectro de magnitudes* e o *espectro de fases* de $s(t)$, respectivamente, dados por:

$$|S(f)| = \sqrt{(\Re\{S(f)\})^2 + (\Im\{S(f)\})^2}; \quad (11)$$

$$\angle S(f) = \arctan\left(\frac{\Im\{S(f)\}}{\Re\{S(f)\}}\right), \quad (12)$$

onde $\Re\{\cdot\}$ e $\Im\{\cdot\}$ representam a parte real e imaginária de $S(f)$, respectivamente.

A seguir, apresentam-se algumas propriedades da transformada de Fourier que serão utilizadas mais adiante.

A primeira propriedade é a *propriedade da simetria de conjugado*. Se $s(t)$ for um

sinal real, então, da Eq. (6), obtém-se que

$$S(-f) = \int_{-\infty}^{\infty} s(t) \exp(j2\pi ft) dt. \quad (13)$$

Daí,

$$S(-f) = \bar{S}(f), \quad (14)$$

em que $\bar{S}(f)$ é o conjugado complexo de $S(f)$, e, portanto,

$$|S(-f)| = |S(f)|; \quad (15)$$

$$\angle S(-f) = -\angle S(f). \quad (16)$$

Isto implica que o espectro de amplitude é uma função par e o espectro de fase é uma função ímpar de f .

Antes de enunciar a próxima propriedade, a propriedade de convolução, define-se primeiramente o conceito de convolução. Assim como qualquer operação algébrica sobre dois sinais, a convolução também é uma operação que quando aplicada sobre duas funções previamente conhecidas, gera uma terceira. Dados dois sinais s_1 e s_2 absolutamente integráveis, define-se a *convolução* entre s_1 e s_2 por

$$s_1(t) * s_2(t) = \int_{-\infty}^{\infty} s_1(\tau) s_2(t - \tau) d\tau, \quad (17)$$

onde o símbolo $*$ denota o produto de convolução. A partir dessa definição, pode-se deduzir que a convolução possui as seguintes propriedades: comutatividade, associatividade e distributividade. Essas três propriedades mostram que a convolução se comporta como a operação algébrica da multiplicação.

A *propriedade de convolução no tempo*, afirma que se dois sinais $s_1(t)$ e $s_2(t)$ são absolutamente integráveis, de modo que existam suas transformadas de Fourier, $S_1(f)$ e $S_2(f)$, respectivamente, e sendo $s(t)$ a convolução entre s_1 e s_2 , isto é, $s(t) = s_1(t) * s_2(t)$, então a sua transformada de Fourier é dada por $S(f) = S_1(f)S_2(f)$.

A transformada de uma função de grande importância em acústica de salas é a *função impulso unitário* (também conhecida como *distribuição delta de Dirac*) $\delta(t)$. Essa função generalizada é definida como

$$\begin{aligned} \delta(t) &= 0 \quad \forall t \neq 0 \\ \int_{-\infty}^{\infty} \delta(t) dt &= 1. \end{aligned} \quad (18)$$

Uma função ordinária que é nula em todos os instantes, exceto em um único ponto, deve ter a integral igual a zero, no sentido de integral de Riemann. Logo, $\delta(t)$ não é uma

função no sentido ordinário, mas sim uma *função generalizada*. Uma função generalizada é definida por seu efeito em outras funções em vez de seus valores em todo instante de tempo. Desse modo, a função impulso pode ser definida pela *propriedade da amostragem*, ou seja,

$$\int_{-\infty}^{\infty} \delta(t)\varphi(t) dt = \varphi(0), \quad (19)$$

onde $\varphi(t)$ é qualquer função regular contínua em $t = 0$ (conhecida como *função teste*). A Eq. (19) é uma expressão simbólica e não deve ser considerada como uma integral de Riemann ordinária. Assim, $\delta(t)$ é classificada como uma *distribuição* ou *função generalizada*, dado que uma função generalizada é definida por seu efeito sobre outras funções ao invés de seus valores em todo instante de tempo (LATHI, 2005).

De modo semelhante, a função impulso unitário retardada $\delta(t - \tau)$ pode ser definida como

$$\int_{-\infty}^{\infty} \delta(t - \tau)\varphi(t) dt = \varphi(\tau), \quad (20)$$

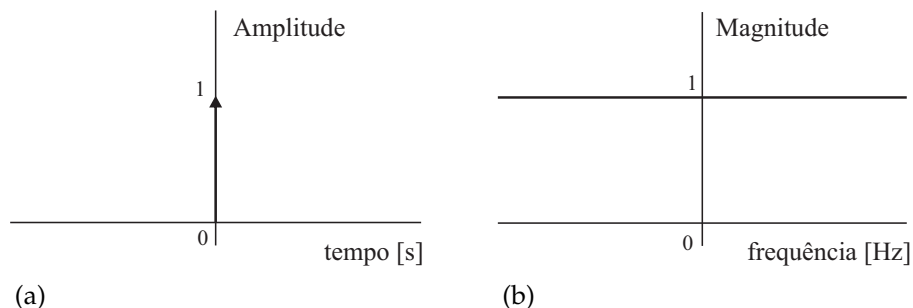
onde $\varphi(t)$ é qualquer função regular contínua em $t = \tau$.

A transformada de Fourier da função impulso unitário é

$$\mathfrak{F}\{\delta(t)\} = \int_{-\infty}^{\infty} \delta(t) \exp(-j2\pi ft) dt = \exp(-j2\pi f \cdot 0) = 1. \quad (21)$$

A interpretação da Eq. (21) é que essa função possui todas as componentes de frequência com igual magnitude. A Figura 10 ilustra a função impulso unitário nos domínios do tempo e frequência.

Figura 10 - Função impulso unitário



Legenda: (a) - Domínio do tempo; (b) - Domínio da frequência (magnitude).

Fonte: O autor, 2017.

Dado que, na prática, a análise de sinais é realizada no computador, logo, tanto $s(t)$ quanto $S(f)$ devem ser sinais discretos e ter comprimento finito. Sendo assim, esses

sinais precisam ser discretizados (amostrados) em uma determinada taxa e ter uma quantidade finita de amostras. Amostrar $s(t)$ significa gerar uma série temporal $s(t_n)$ ($:= s[n]$), onde $t_n = nT_s$, em que T_s é o período de amostragem, ou seja, o intervalo de tempo entre duas amostras. O período de amostragem é o inverso da frequência de amostragem f_s , isto é, $T_s = 1/f_s$. Segundo o *teorema da amostragem* (NYQUIST, 1928), um sinal que possui faixa de frequência limitada a f_m [Hz] pode ser completamente representado por amostras discretas, sem qualquer perda de informação, desde que a frequência de amostragem satisfaça o *critério de Nyquist*, $f_s \geq 2f_m$. Em outras palavras, esse teorema diz que a menor taxa de amostragem necessária para preservar a informação de $s(t)$ deve ser $2f_m$, conhecida como *frequência de Nyquist*. Sabendo que o ser humano é capaz de ouvir sons cujas frequências variam entre 20 Hz e 20 kHz, deve-se utilizar uma taxa de amostragem mínima de 40 kHz, para que todas as frequências audíveis possam ser registradas. Tipicamente utiliza-se a taxa de amostragem de 44,1 kHz ou 48 kHz.

O método numérico utilizado para calcular a transformada de Fourier para sinais discretos é a transformada discreta de Fourier. Sendo $s[n]$ uma sequência de comprimento finito N , isto é, $s[n]$ é nulo fora do intervalo $0 \leq n \leq N - 1$, define-se a *transformada discreta de Fourier* $S[m]$ como

$$S[m] = \sum_{n=0}^{N-1} s[n] \exp(-j2\pi nm/N), \quad m = 0, 1, \dots, N - 1, \quad (22)$$

e a *transformada inversa discreta de Fourier* como

$$s[n] = \frac{1}{N} \sum_{m=0}^{N-1} S[m] \exp(j2\pi nm/N), \quad n = 0, 1, \dots, N - 1. \quad (23)$$

Algumas importantes características da transformada discreta de Fourier são:

- Existe uma correspondência biunívoca entre $s[n]$ e $S[m]$;
- Esse par de transformadas é apropriado para o uso de computadores pois é discreta e de comprimento finito, tanto no domínio do tempo quanto no domínio da frequência;
- Há um algoritmo rápido, chamado de *transformada rápida de Fourier* (FFT - *fast Fourier transform*), o qual reduz drasticamente o número de cálculos necessários para executar a transformada discreta de Fourier. Existem algoritmos para a FFT que podem calcular para qualquer valor de N , no entanto, os mais rápidos são aqueles para N igual à potência de dois;
- Note que o valor de N na Eq. (22) não está fixado. Se $s[n]$ tiver um compri-

mento menor do que N , então pode-se fazer que $s[n]$ tenha comprimento N , simplesmente adicionando amostras de valor zero. Essa adição de amostras nulas é denominada de *preenchimento nulo*. Pode-se tirar proveito disso para escolher, de forma criteriosa, o valor de N que seja uma potência de 2, para assim utilizar o algoritmo FFT;

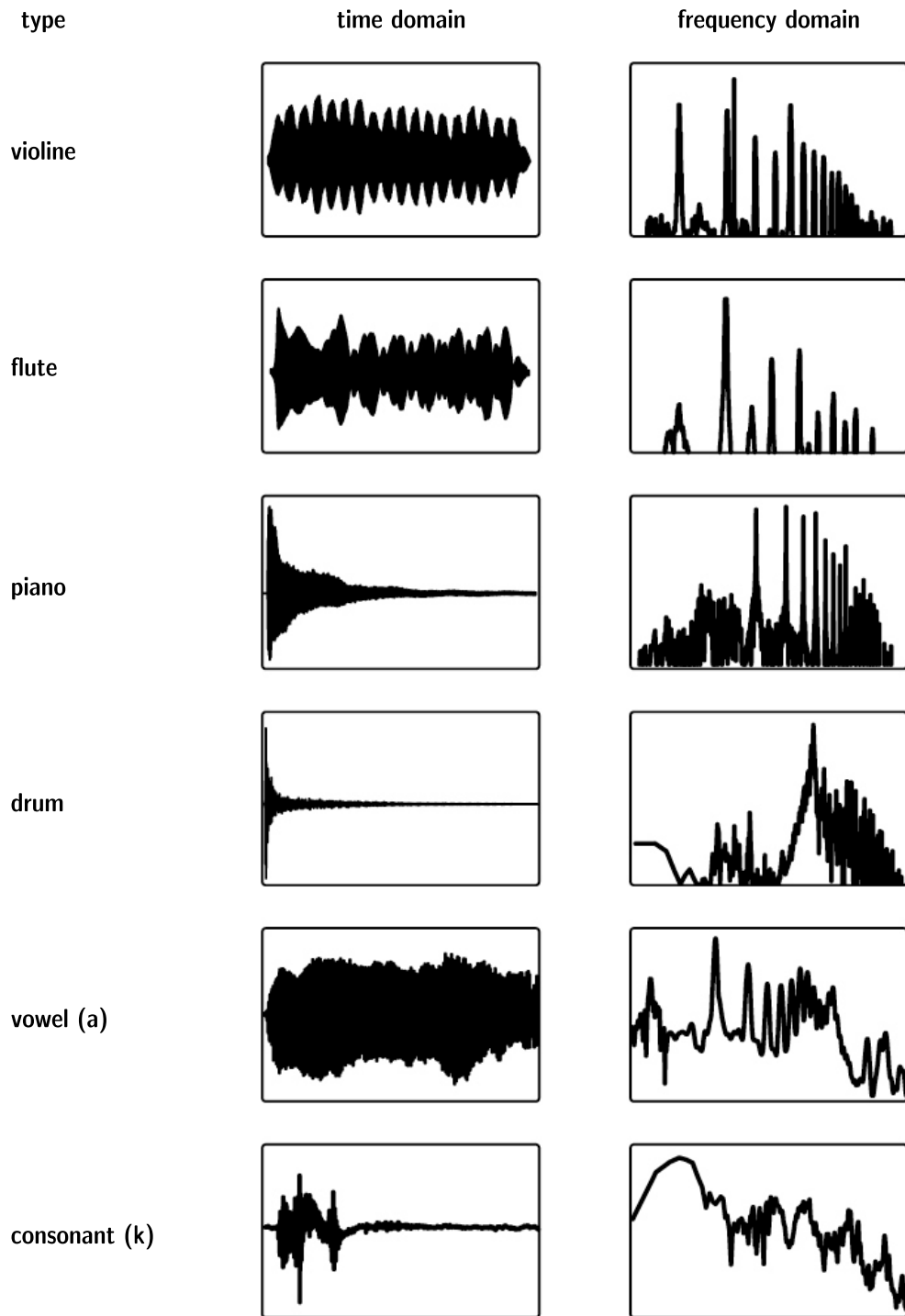
- A propriedade de simetria de conjugado, apresentada anteriormente, também se aplica à transformada discreta de Fourier. Como os sinais acústicos no domínio do tempo tratados neste trabalho são reais e de comprimento finito, e em virtude dessa propriedade, apenas a primeira metade do gráfico dos componentes de frequência será mostrado;
- No domínio de tempo discreto a convolução é dada por

$$s[n] = s_1[n] * s_2[n] = \sum_{k=-\infty}^{\infty} s_1[k]s_2[n - k]. \quad (24)$$

A explicação detalhada de toda a teoria da passagem do domínio contínuo para o discreto e vice-versa foge ao escopo deste trabalho, mas os detalhes podem ser encontrados em Oppenheim e SCHAFER (2010).

Para finalizar esta seção, na Figura 11 mostra-se o espectro de alguns sinais acústicos. Os quatro primeiros sinais mostram formas de onda de quatro instrumentos musicais, a saber, na ordem: violino, flauta, piano e tambor, enquanto o penúltimo corresponde ao som da vogal “a” e o último da consoante “k”. Observando os espectros dos instrumentos musicais de corda (violino) e sopro (flauta) pode-se notar a presença de tons puros (picos). Essas características são típicas desses instrumentos, já que produzem sons harmônicos. Sons harmônicos são os sons resultantes da superposição de vários tons puros (som em uma única frequência), sendo que essas frequências, chamadas de harmônicos, estão relacionadas com a frequência mais baixa, chamada de fundamental, através da multiplicidade inteira desta última. Por outro lado, têm-se que os instrumentos de percussão (tambor) geram sons não harmônicos. Analisando, agora, os sons das letras, observa-se as seguintes características: a vogal apresenta uma duração mais longa do que a consoante; existe um intervalo de silêncio na pronúncia da consoante “k”, essa é uma característica das consoantes oclusivas; há um decaimento de magnitude com o aumento da frequência tanto no espectro do som da vogal quanto da consoante. Outras análises sobre as características da voz podem ser encontradas, por exemplo, em Alcaim e Oliveira (2011).

Figura 11 - Exemplos de espectros de alguns sinais acústicos

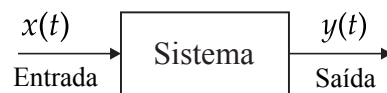


Fonte: VORLÄNDER, 2008.

1.6.2 Sistemas lineares e invariantes no tempo

Um sistema é algo que transforma um conjunto de sinais em outro conjunto de sinais. A Figura 12 ilustra o diagrama representando o sistema sendo excitado pelo sinal $x(t)$ e produzindo uma resposta $y(t)$. Um sistema linear satisfaz a propriedade de superposição, isto é, se y_1 e y_2 são as respectivas saídas às entradas x_1 e x_2 , então para todos os valores de constantes c_1 e c_2 , e para todo x_1 e x_2 , tem-se que $c_1y_1 + c_2y_2$ é a saída associada à entrada $c_1x_1 + c_2x_2$. Agora, um sistema é invariante no tempo quando suas características não variam com o tempo, ou seja, se a entrada for atrasada por τ segundos ($x(t - \tau)$), então a saída é a mesma anterior, mas atrasada também por τ segundos ($y(t - \tau)$).

Figura 12 - Representação de um sistema que produz uma resposta $y(t)$ baseada em uma entrada $x(t)$



Fonte: O autor, 2017.

Como o sistema formado por sala-fonte-receptor pode ser considerado um *sistema linear e invariante no tempo* (SLIT) (SHIN; HAMMOND, 2008; VORLÄNDER, 2008), desde que sejam satisfeitas determinadas condições, faz-se necessário apresentar algumas importantes propriedades do sistema dotado desses atributos. Os requisitos para o sistema sala-fonte-receptor seja um SLIT são: as características da sala não variem com o tempo, e que as características da fonte e do receptor sejam lineares (BRANDÃO, 2016).

A resposta ao impulso $h(t)$ (IR - *Impulse Response*) corresponde ao sinal de saída de um sistema quando o sinal de excitação é o impulso $\delta(t)$. Pode-se mostrar que a resposta de um SLIT é o produto de convolução de sua resposta impulsiva com a excitação, ou seja,

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau) d\tau. \quad (25)$$

Assim vê-se que um SLIT é completamente caracterizado por sua resposta ao impulso, e essa propriedade é válida em geral somente para os SLITs (OPPENHEIM; WILLSKY, 2010).

Usando o teorema da convolução, a Eq. (25) pode ser descrita no domínio da

frequência, isto é,

$$Y(f) = X(f)H(f), \quad (26)$$

onde $X(f)$, $Y(f)$ e $H(f)$ (esta última conhecida como função de transferência) são as transformadas de Fourier dos sinais $x(t)$, $y(t)$ e $h(t)$, respectivamente. Note que, a partir dessa equação é possível calcular $h(t)$, quando os sinais $x(t)$ e $y(t)$ são conhecidos, da seguinte maneira

$$h(t) = \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{y(t)\}}{\mathcal{F}\{x(t)\}} \right\}. \quad (27)$$

Esse método é conhecido como *deconvolução*.

Do ponto de vista teórico, o sinal delta de Dirac é o sinal ideal para excitar uma sala, dado que todas as componentes de frequência têm igual energia e, além disso, ele só existe no tempo $t = 0$. Como esse sinal não é fisicamente realizável, a resposta impulsiva da sala (RIR - *Room Impulse Response*), na prática, é uma função temporal da pressão sonora, resultante da excitação dessa sala por uma função que se aproxima ou que tenha as mesmas características da função delta de Dirac (VORLÄNDER; BIETZ, 1994). Diante disso, diversas técnicas de obtenção da resposta impulsiva foram desenvolvidas para contornar esse problema, por exemplo, varredura em frequência (*sweep sine*), *Maximum Length Sequences* (MLS) e deconvolução (análise por FFT). O detalhamento e a comparação dessas e outras técnicas podem ser vistas em Müller e Massarani (2001) e Stan et al. (2002). Para a medição da RIR é utilizada uma fonte sonora onidirecional, para excitar a sala usando uma das técnicas mencionadas anteriormente, e um microfone para captar a resposta. Quando é utilizada uma cabeça artificial (ver Figura 13), no lugar do microfone, o que se mede são as *respostas impulsivas biauriculares* (BRIR - *Binaural Room Impulse Response*). Esses registros temporais são obtidos a partir de dois microfones, localizados um em cada ouvido. Vale ressaltar que em algumas situações é desejável excitar a sala com uma fonte sonora que apresente uma direcionalidade que se aproxime, por exemplo, da do ser humano ou de algum instrumento musical (BRANDÃO, 2016). Além disso, as fontes devem ter potência suficiente para que a relação sinal-ruído seja grande em toda faixa de frequência de interesse. A relação sinal-ruído (SNR - *Signal-to-Noise Ratio*) é definida como a razão entre a potência do sinal e a potência do ruído. Daí, quanto mais alta for a SNR, menor será o efeito do ruído sobre a medição do sinal. Uma outra observação é que a escolha do sinal de excitação é importante para a obtenção de uma resposta impulsiva de boa qualidade. Em Melo (2012) mostra-se que a varredura em frequência é o método mais indicado, em virtude das seguintes vantagens em relação às outras técnicas:

- Possibilidade de eliminação da distorção harmônica da resposta impul-

siva que é causada pela não-linearidade da fonte sonora.

- Facilidade de introdução de pré-ênfases ao sinal. Denominam-se pré-ênfases as correções dos sinais dos equipamentos utilizados no levantamento da resposta impulsiva, como microfones e altofalantes, já que esses não possuem uma resposta plana. Uma vez que é desejável avaliar a função de transferência da sala sem a influência desses elementos.
- Possibilidade de criarem-se varreduras com qualquer conteúdo espectral.

É necessário frisar também que como a RIR é medida em pares de fonte-receptor, uma única RIR não pode caracterizar a sala como todo. Sendo assim, faz-se necessário a análise da variabilidade dos parâmetros ao longo da sala, sendo desejável que essas variações não sejam tão perceptíveis. É recomendada, para o cálculo dos parâmetros acústicos, a medição da RIR em diversas posições distintas. Diante disso, os sinais de excitação devem ser reproduzíveis ao longo do tempo e do espaço (BERKHOUT et al., 1980). Tais procedimentos que devem ser tomados na medição da resposta impulsiva podem ser vistos na norma ISO 3382 (1997).

Figura 13 - Exemplos de cabeças artificiais de adulto de diversos fabricantes



Fonte: FEDTKE, 2007.

Para efeito de exemplo de medições de RIR e BRIR, considere a Figura 14 que ilustra o aparato de medição das respostas impulsivas, as quais são mostradas nas Figuras 15 e 16. Essas medições estão disponíveis no site do *Institute of Communication Systems and Data Processing* da *RWTH Aachen University* (<http://www.ind.rwth-aachen.de/AIR>). Na Figura 15a tem-se a representação da resposta impulsiva medida com um microfone na mesma posição que se encontra a cabeça artificial, cujo espectro de magnitude é mostrado na Figura 15b. A Figura 16 representa a resposta impulsiva biauricular e os seus respectivos espectros de magnitude. À primeira vista, pode-se, equivocadamente, concluir que não há diferença entre os gráficos da RIR (Figura 15a) e das BRIRs (Figuras 16a e 16b). Embora os gráficos possuam uma certa semelhança, as

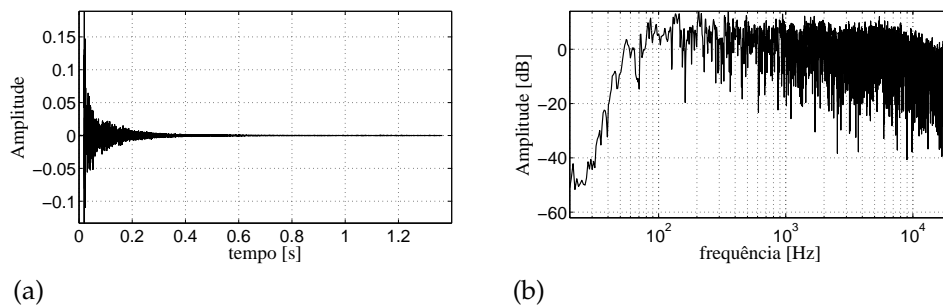
BRIRs carregam as informações associadas à presença da cabeça e da parte superior do torso, já que tais partes do corpo humano funcionam como filtros acústicos naturais. A dissemelhança das respostas é melhor evidenciada quando os respectivos espectros são comparados. No Capítulo 2 serão apresentadas com mais detalhes essas e outras características da audição humana. Também será explicada a determinação das BRIRs através de um modelo computacional.

Figura 14 - Aparato de medição de respostas impulsivas



Fonte: JEUB, 2011.

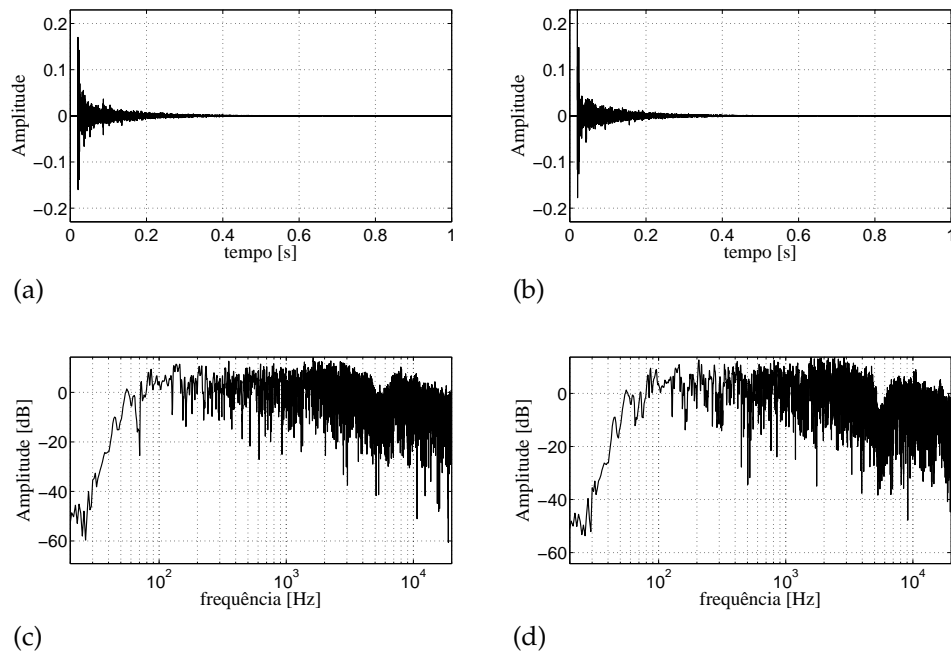
Figura 15 - Resposta impulsiva medida na mesma posição da cabeça artificial da sala mostrada na Figura 14



Legenda: (a) - Domínio do tempo; (b) - Domínio da frequência (magnitude).

Fonte: O autor, 2017.

Figura 16 - Respostas impulsivas biauriculares de acordo com as configurações mostradas na Figura 14



Legenda: (a) - BRIR esquerda - domínio do tempo; (b) - BRIR direita - domínio do tempo; (c) - BRIR esquerda - domínio da frequência (magnitude); (d) - BRIR direita - domínio da frequência (magnitude).

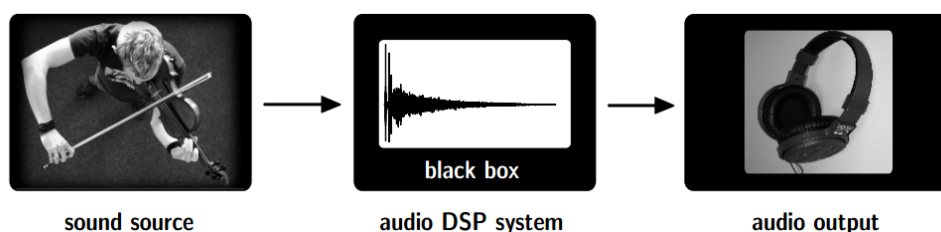
Fonte: O autor, 2017.

Uma vez obtidas as RIRs ou as BRIRs calculam-se, então, os parâmetros objetivos para uma avaliação da sala em questão. Esses parâmetros normalmente são analisados em bandas de oitava ou de terço de oitava. Para isso, as respostas impulsivas são processadas com um banco de filtros passa-banda, cujas frequências centrais são padronizadas tendo, em geral, a frequência de 1 kHz como referência e seus múltiplos e submúltiplos como frequências centrais das demais bandas.

2 AURILIZAÇÃO DE SALAS

Conforme mencionado no capítulo Introdução, a aurilização pode ser obtida a partir de dados medidos ou simulados. A aurilização obtida a partir de dados medidos consiste em medir as BRIRs, *in loco*, e posteriormente convoluí-las com um som anecóico e monofônico. A Figura 17 sintetiza o processo de implementação e geração de aurilização. Em Melo (2012) é discorrido minuciosamente sobre o processamento de sinais para a geração de aurilização a partir de sinais medidos. A técnica de geração de aurilização utilizando dados simulados, objeto de estudo deste trabalho, será abordado neste capítulo.

Figura 17 - Diagrama de blocos de geração de aurilização.



Nota: A fonte sonora (sinal de entrada) é processada por um sistema de processamento de sinais de áudio, e o sinal de saída é, então, reproduzido a partir de fones de ouvido.

Fonte: VORLÄNDER, 2008.

O modelo computacional de acústica de salas que possui como saída a aurilização demanda quatro passos: no primeiro passo é feita a modelagem da geometria da sala, da fonte sonora, do receptor; no segundo passo um algoritmo fundamentado na teoria acústica é responsável pela geração das BRIRs; no terceiro passo é feita uma gravação de um som anecóico e monofônico; e finalmente no quarto e último passo as BRIRs são convoluídas com o sinal anecóico e o sinal resultante é reproduzido utilizando fones de ouvido. Todos esses passos serão descritos com mais detalhes nas próximas seções.

A Figura 18 mostra a janela principal do código computacional RAIOS. Abaixo da barra de ferramentas tem-se a área de trabalho (*Edit and View*), cujo espaço é dedicado para a criação dos projetos. A janela, *Data Entry*, que fica à direita da área de trabalho, permite que o usuário forneça os dados de entrada. Essa janela é composta de quatro abas, a saber:

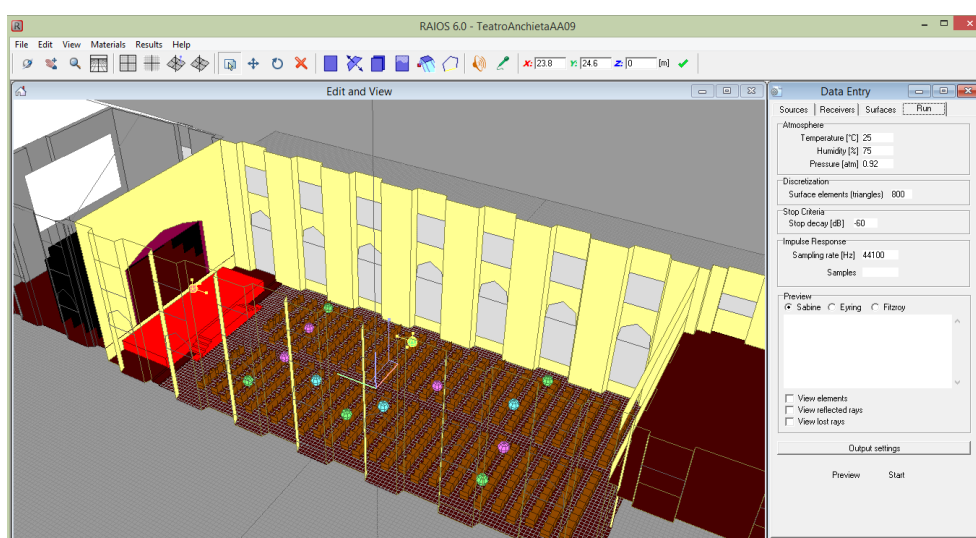
- *Sources*: contém as informações da fonte, tais como a posição, a potência,

a orientação e o número de raios emitidos;

- *Receivers*: contém as informações do receptor, tais como a posição e a orientação;
- *Surfaces*: contém as informações dos materiais das superfícies, tais como o coeficiente de absorção e espalhamento;
- *Run*: contém as informações das condições atmosféricas, da discretização das superfícies e do critério de parada. Como pode ser observado na Figura 18, os parâmetros escolhidos para as condições atmosféricas são: a temperatura, a umidade e a pressão. Esses parâmetros afetam os resultados da simulação porque a velocidade da onda depende deles, mas principalmente porque a atenuação do ar depende fortemente desses dados, especialmente em bandas de frequência média e alta. O outro parâmetro dessa aba é a discretização, a qual estabelece o número (inteiro) de subdivisões triangulares de superfícies, necessário para o refinamento do cálculo de reflexões difusas. O terceiro importante parâmetro destacado é o *stop decay*, em dB, que indica a atenuação máxima de um raio acústico, antes que ele seja descartado do cálculo. Essa condição é explicada em detalhes na Seção 2.3.

Para um melhor entendimento, os detalhes de cada aba serão abordados ao longo das seções seguintes.

Figura 18 - Janela principal do RAIOS



Fonte: O autor, 2017.

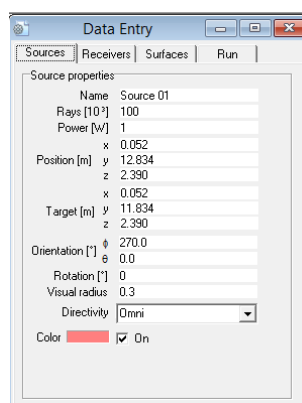
2.1 Fontes sonoras

Um conjunto de informações das fontes sonoras são necessárias para a determinação da distribuição temporal da energia sonora para o sistema sala-fonte-receptor, a saber:

- A sua posição na sala em relação à origem do sistema de coordenadas, uma vez que uma mudança na posição resulta numa alteração da resposta ao impulso do sistema sala-fonte-receptor;
- A sua potência sonora, dada em Watts;
- A sua direcionalidade, que indica como a energia sonora é distribuída espacialmente. Quando a fonte sonora não apresenta direções preferenciais de radiação, diz-se que a fonte é *onidirecional*; caso contrário, a fonte é *direcional*. As fontes sonoras deixam de ser onidirecionais pelos seguintes motivos (BISTAFA, 2011): por não apresentarem forma esférica, e porque a amplitude e fase das vibrações de suas diferentes superfícies não são uniformes. Sendo assim, para fontes sonoras direcionais, a potência sonora não é suficiente para caracterizar acusticamente a fonte. Há, portanto, a necessidade do conhecimento da sua direcionalidade. A orientação do eixo principal de radiação usualmente corresponde à direção em que a fonte radia mais energia sonora. Assim, se a fonte utilizada é direcional, deve-se informar os ângulos de azimute e elevação formados pelo eixo principal de radiação em relação ao sistema de coordenadas da sala.

A Figura 19 mostra a aba *Sources* da janela *Data Entry* do código RAIOS, onde essas propriedades são as entradas da fonte.

Figura 19 - Aba *Sources* da janela *Data Entry* do código RAIOS



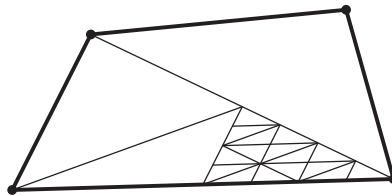
Fonte: O autor, 2017.

2.2 Geometria da sala: superfícies de contorno da sala

A modelagem de uma superfície pode ser feita aproximando as superfícies que compõem a sala por um conjunto de polígonos planos, como mostra a Figura 18.

No código computacional RAIOS, apresentado na Seção 2.3, em cada polígono é feita uma subdivisão em triângulos, como ilustrado na Figura 20, para a modelagem do fenômeno de difusão, onde cada elemento triangular que compõe a sala possui as seguintes informações:

Figura 20 - Triangularização da superfície



Fonte: CAMILO, 2003.

- Das coordenadas dos seus vértices V_1 , V_2 e V_3 , as quais definirão o vetor normal à superfície. A partir desse vetor e do vetor raio incidente pode-se determinar a direção do raio refletido specularmente, já que essa reflexão obedece à *lei de Snell*, isto é, o raio incidente \mathbf{r}_i , o raio refletido \mathbf{r}_r e a normal à superfície de reflexão \mathbf{n} são coplanares, e o ângulo de incidência θ_i é igual ao ângulo de reflexão θ_r , como mostra a Figura 21a. Daí, a nova direção do raio após a reflexão é dada pela seguinte relação vetorial

$$\mathbf{r}_r = \mathbf{r}_i - 2(\mathbf{r}_i \cdot \mathbf{n})\mathbf{n}. \quad (28)$$

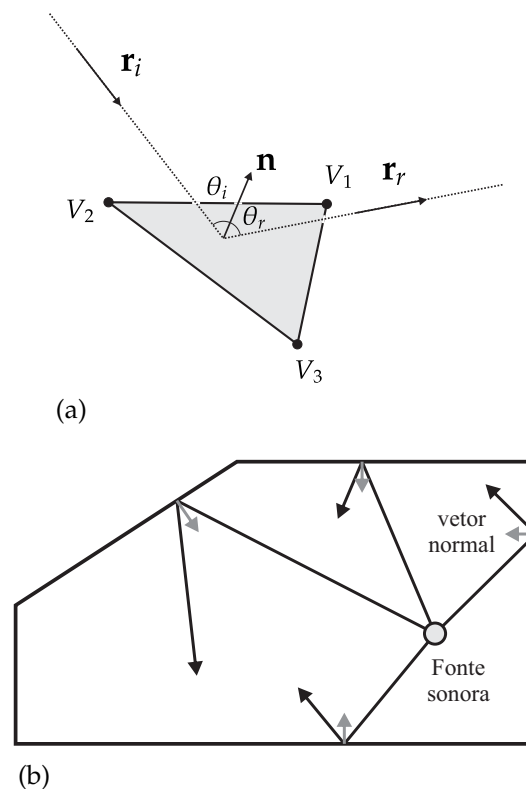
Uma observação importante com respeito ao vetor normal, é que ele deve apontar sempre para o interior da sala, como mostra a Figura 21b. A visibilidade de um elemento triangular para um determinado raio é verificada a partir das seguintes condições:

- Checar o produto escalar entre os vetores \mathbf{r}_i e \mathbf{n} . Se o resultado desse produto é positivo, então o elemento triangular não é visível ao raio. Caso contrário, ele pode ser visível.
- Na hipótese da primeira condição ser satisfeita, verificar se nenhum plano intercepta o segmento de reta $\overline{P_0P_1}$, onde

P_0 é um ponto de emissão do raio \mathbf{r}_i que pode ser de uma superfície (ponto da superfície em que ocorreu a última reflexão) ou de uma fonte sonora e P_1 é um ponto pertencente ao elemento triangular (ver Figura 22). Caso não exista, o elemento triangular é visível.

- Dos coeficientes de absorção α e de espalhamento δ do material de que é feita a superfície. Esses dados são necessários para o cômputo da energia do raio que é refletido de maneira especular e difusa. A Figura 23 mostra a aba *Surfaces* da janela *Data Entry* do código RAIOS, onde esses dados são as entradas das superfícies.

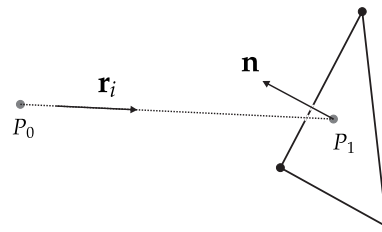
Figura 21 - Orientação do vetor normal ao elemento triangular deve estar sempre apontada para o interior da sala



Legenda: (a) - Representação dos vetores que descrevem o comportamento da reflexão especular, a saber, o vetor raio incidente \mathbf{r}_i , o vetor raio refletido \mathbf{r}_r e o vetor normal \mathbf{n} ; (b) - Orientação do vetor normal e a representação em 2D, para facilitar a compreensão, da propagação dos raios acústicos e da sua reflexão especular.

Fonte: O autor, 2017.

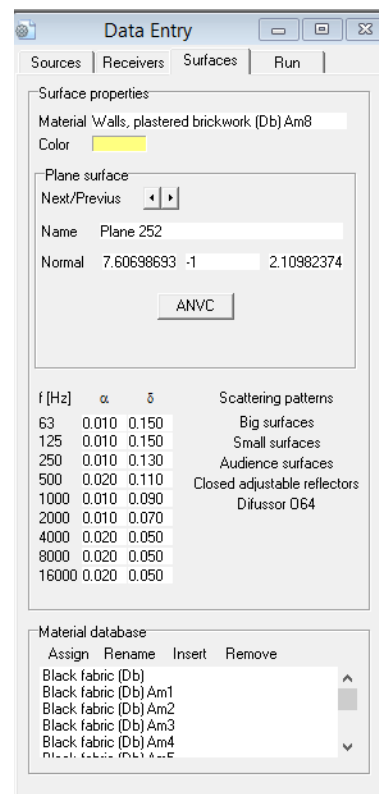
Figura 22 - Visibilidade de um elemento triangular ao raio acústico



Nota: P_0 é um ponto de emissão do raio \mathbf{r}_i que pode ser de uma superfície ou de uma fonte, P_1 é um ponto pertencente ao elemento triangular e \mathbf{n} é o vetor normal.

Fonte: O autor, 2017.

Figura 23 - Aba *Surfaces* da janela *Data Entry* do código RAIOS



Fonte: O autor, 2017.

2.3 Campo sonoro

O código computacional RAIOS, que foi utilizado neste trabalho, utiliza um método híbrido para gerar a resposta impulsiva (CAMILO, 2003): o método de *traçado de raios* e o método de *transição de energia*. Faz-se necessário a hibridização, uma vez que cada técnica tende a favorecer a modelagem de uma determinada parte do fenômeno de propagação sonora.

2.3.1 Método do traçado de raios

No método do traçado de raios, assume-se que a onda sonora é constituída por raios acústicos, onde a energia inicial de cada raio, emitida pela fonte sonora, é igual à energia total da fonte dividida pelo número de raios, caso a fonte seja onidirecional. Cada raio carrega as seguintes informações: o espectro de potência em nove bandas de oitava (63 Hz - 16 kHz); a direção de propagação; e o tempo de percurso. O raio, ao se propagar em linha reta, vai sofrendo alterações no nível de energia em seu espectro a cada reflexão, devido à absorção sonora das superfícies, e, além disso, por meio dos efeitos de dissipação, causados pela viscosidade do ar. A energia do raio após m reflexões é calculada pela expressão (CAMILO, 2003)

$$E_m = \frac{E_F}{N_R} D_{\phi\theta} e^{-\gamma d} \prod_{i=1}^m (1 - \alpha_i)(1 - \delta_i), \quad (29)$$

onde E_F é a energia total emitida pela fonte, N_R é o número de raios, $D_{\phi\theta}$ é a direcionalidade da fonte, γ é o coeficiente de absorção do ar, d é a distância, α é o coeficiente de absorção dos materiais das superfícies e δ é o coeficiente de espalhamento dos materiais das superfícies.

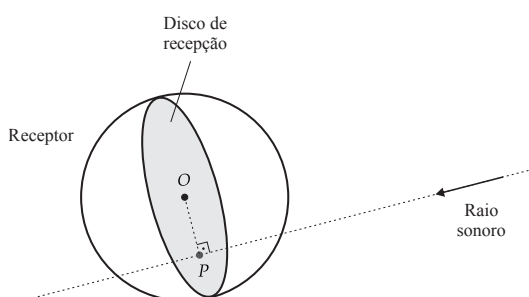
Uma vez que o raio é emitido pela fonte, alguns testes são verificados durante a sua propagação, a saber:

- **Teste de detecção do receptor.** Nesse teste verifica-se se o raio intercepta o receptor; para isso traça-se uma reta perpendicular à trajetória retilínea do raio, ver Figura 24. Se a distância entre o centro do disco receptor e ponto de intercessão dessas retas for menor que o raio do disco receptor, então o raio interceptou-o. À medida que os raios interceptam o receptor, são armazenadas as seguintes informações: o espectro de potência em nove bandas de oitava, a direção de propagação, e o tempo de percurso, para posterior processamento.
- **Teste de detecção do elemento triangular.** Nesse teste é averiguado se

o raio intercepta algum elemento triangular. A descrição desse teste foi apresentada na Seção. 2.2.

- **Teste de aniquilação do raio.** Sabe-se que o raio, após ser emitido pela fonte, sofrerá infinitas reflexões. Sendo assim, é necessário um teste para determinar se o raio deve continuar ou não a sua propagação. Há várias maneiras para isso. Uma seria limitar a existência do raio a partir da definição de um tempo máximo de propagação. Uma outra maneira seria através do estabelecimento de um valor de energia mínima, ou seja, se a energia do raio atingir um valor abaixo de um valor mínimo preestabelecido, o raio é aniquilado. Assim que o raio é “parado”, um outro raio é emitido até que todos os raios sejam emitidos.

Figura 24 - Teste de detecção para verificar se o raio acústico intercepta o receptor, cuja posição corresponde ao ponto O



Fonte: O autor, 2017.

Observe que a recepção dos raios mencionada aqui foi utilizando um receptor esférico (similar a um microfone onidirecional). Para modelar uma cabeça receptora, ver-se-á na Seção 2.4 que é necessário modelar a difração da onda sonora na cabeça, no torso e no pavilhão auricular, já que esses causam uma modificação no espectro do raio antes de alcançar o canal auditivo. A caracterização desse tipo de receptor está fundamentada na chamada *função de transferência relacionada à cabeça*, HRTF.

2.3.2 Método das transições de energia

Para modelar a difusão do som é utilizado o método das transições de energia (KRUZINS; FRICKE, 1982; ALARCÃO et al., 2000). O cômputo das reflexões difusas é feita depois do processamento das reflexões especulares, ou seja, após a execução do método de traçado de raios, pois nessa fase, quando ocorrem as reflexões especulares, é armazenada a energia difusa em uma matriz, denominada *matriz de transição* E_D

(Eq. (30)), onde as linhas representam os elementos triangulares (ver Seção 2.2) e as colunas expressam a discretização do tempo.

$$\mathbf{E}_D = \begin{bmatrix} E_{1,1} & E_{1,2} & E_{1,3} & \dots & E_{1,m-2} & E_{1,m-1} & E_{1,m} \\ E_{2,1} & E_{2,2} & E_{2,3} & \dots & E_{2,m-2} & E_{2,m-1} & E_{2,m} \\ E_{3,1} & E_{3,2} & E_{3,3} & \dots & E_{3,m-2} & E_{3,m-1} & E_{3,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ E_{n-2,1} & E_{n-2,2} & E_{n-2,3} & \dots & E_{n-2,m-2} & E_{n-2,m-1} & E_{n-2,m} \\ E_{n-1,1} & E_{n-1,2} & E_{n-1,3} & \dots & E_{n-1,m-2} & E_{n-1,m-1} & E_{n-1,m} \\ E_{n,1} & E_{n,2} & E_{n,3} & \dots & E_{n,m-2} & E_{n,m-1} & E_{n,m} \end{bmatrix}, \quad (30)$$

onde n é o número de elementos triangulares de todas as superfícies interiores à sala e m é um número inteiro igual ao tempo de reverberação aproximado, T_{60} , o qual é calculado previamente em milissegundos. Nessa primeira etapa, a cada elemento i onde ocorreu a incidência do raio, será acrescentada uma parcela da energia desse raio, da seguinte forma

$$E_{i,t} \leftarrow E_{i,t} + E_m(1 - \alpha_i)\delta_i, \quad (31)$$

onde t é o tempo de chegada do raio no elemento. Ao final dessa etapa obtém-se a matriz inicial, de onde calculam-se as transições de energia entre os elementos e os receptores “visíveis”. A transição de energia difusa de um elemento i , refletida no instante t , para outro elemento j , recebida no instante $t + \tau$, dada por

$$E_{j,t+\tau} \leftarrow E_{j,t+\tau} + E_{i,t}\Psi_{ij}, \quad (32)$$

sendo que $\tau = d_{ij}/c$ é o tempo que a energia leva para percorrer a distância d_{ij} entre os centros geométricos dos elementos i e j , com velocidade de propagação c , e

$$\Psi_{ij} = \frac{\Omega_{ij}}{2\pi} S_\beta \exp(\gamma d_{ij}), \quad (33)$$

onde Ω_{ij} é o ângulo sólido do elemento j em relação ao elemento i , 2π é a área de uma superfície semiesférica de raio unitário, S_β é dada pela Eq. (2), γ é o coeficiente de atenuação do ar. Vale ressaltar que nessa segunda etapa não se considera que a energia refletida difusamente venha a sofrer novas reflexões especulares. Uma outra observação é que durante a troca de energia difusa entre os elementos triangulares, há também a troca de energia entre cada elemento e o receptor.

Por fim, cabe aqui ressaltar algumas vantagens e desvantagens desses dois métodos que modelam o campo acústico. O método do traçado de raios possui as seguintes vantagens: controle da precisão dos resultados através do número de raios

e decaimento e apresenta boa precisão na parte inicial da resposta impulsiva. As desvantagens desse método são: necessidade de grande número de raios para se calcular a parte intermediária e final da resposta impulsiva e não considera os modos da sala. Agora, com relação o método das transições de energia suas principais vantagens são: apresenta boa estimativa para as partes intermediária e final da resposta impulsiva e é um bom modelo simplificado para difusão. Por outro lado, tem como desvantagens: dá bons resultados apenas para salas medianamente difusas; o cálculo é pouco preciso da parte inicial da resposta impulsiva; possui baixa discretização temporal da resposta impulsiva, já que uma alta discretização exige uma alta capacidade de armazenamento das informações e de processamento, pois maior será o número de colunas da matriz dada em (30).

2.4 Receptores

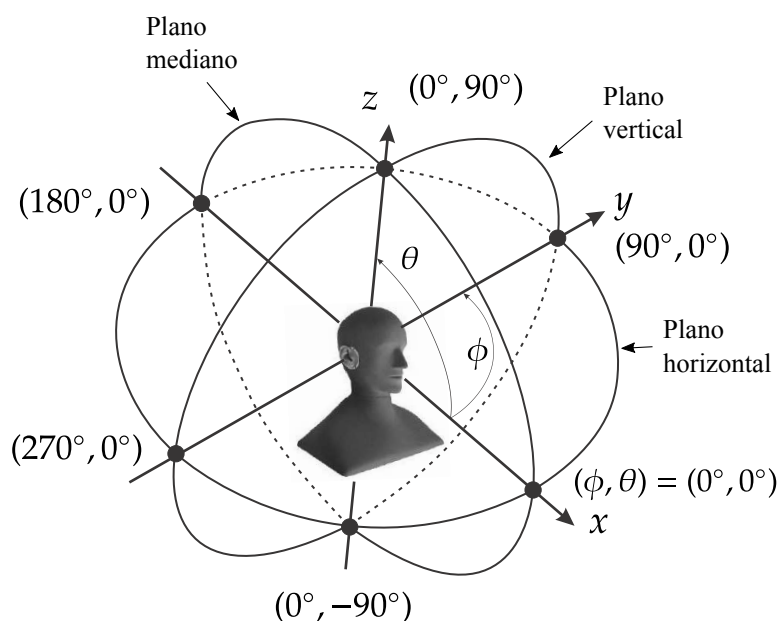
Dois tipos de receptores são geralmente usados nos software que usam a acústica geométrica para a modelagem da propagação da onda acústica: os receptores pontuais e os receptores biauriculares (cabeças).

Os receptores pontuais são semelhantes aos microfones onidirecionais, os quais medem a pressão sonora sem fazer qualquer discriminação da direção de chegada dos raios sonoros que os interceptam. Assim, a única informação necessária é sua posição na sala.

A modelagem do receptor biauricular, semelhante à cabeça humana, é fundamental quando se deseja fazer a aurilização em um determinado ponto da sala. Tal como as fontes direcionais, é necessário o conhecimento de algumas informações, para uma correta modelagem desse receptor, que são:

- A sua posição em relação à origem do sistema de coordenadas da sala, posto que uma mudança da sua posição faz com que a resposta ao impulso do sistema sala-fonte-receptor se altere;
- O estabelecimento de um sistema de coordenadas, fixo ao receptor, e, a partir deste, definir a orientação do receptor para determinar a direção do raio sonoro que o atinge, segundo os ângulos de azimute ϕ e de elevação θ (ver Figura 25). A orientação da cabeça é importante, pois o som que intercepta o receptor e alcança o canal auditivo sofre alterações devido a cabeça, o torso e o pavilhão auricular, e essa filtragem é dependente, também, da direção de incidência do som e da frequência. Essas modificações podem ser caracterizadas pelas chamadas respostas impulsivas relacionadas à cabeça (HRIRs), uma para cada ouvido, ou seja, para

Figura 25 - Sistema de coordenadas fixo ao receptor



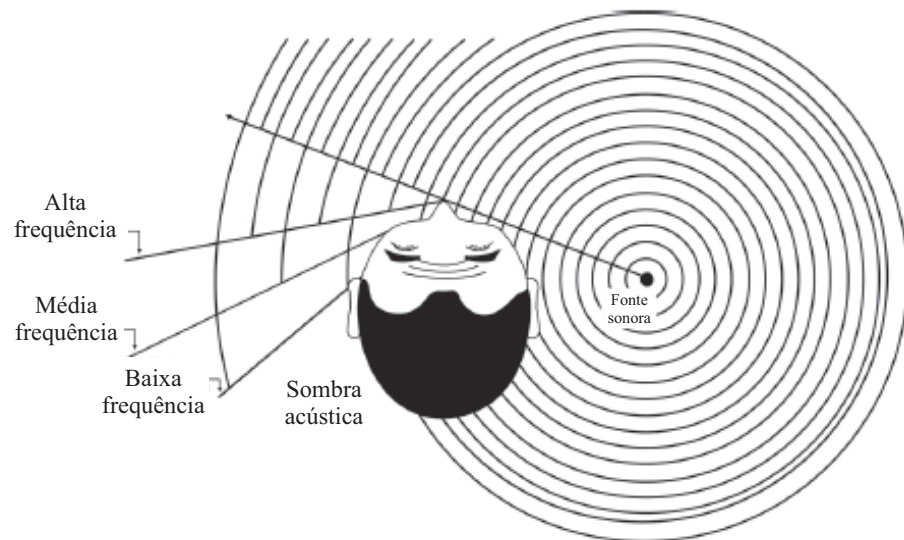
Fonte: O autor, 2017.

cada direção de onde chega o raio no receptor existe um par de HRIRs. As HRIRs, quando expressas no domínio da frequência, são chamadas de funções de transferência relacionadas à cabeça (HRTFs) (SHAW, 1982), sendo obtidas por meio da transformada de Fourier da HRIRs. Como a HRIR é um sinal real, pela propriedade da simetria de conjugado, a magnitude da HRTF possui simetria par. Diante disso, somente metade do gráfico da HRTF será apresentada. Uma outra observação é que para cada direção, na verdade, tem-se um par de HRTFs, uma para cada ouvido. Embora a informação da origem do som possa ser explicada pela HRTF em uma única orelha, a localização da fonte sonora envolve a escuta com os dois ouvidos (BISTAFA, 2011). Segundo Paul (2009), os primeiros estudos que procuravam aclarar o funcionamento da audição biauricular devem-se a Wells (WELLS, 1792) e Venturi (VENTURI, 1796) e em seguida, no século XIX e no início do século XX, após esses trabalhos pioneiros, diversos outros pesquisadores continuaram o trabalho na área, tais como Wheatstone, Dove, Seebeck, Steinhauser, Thompson e Lord Rayleigh (RAYLEIGH, 1894). Eles chegaram à conclusão que a presença de duas orelhas é fundamental para a localização e percepção de distância das fontes. Lord Rayleigh verificou que a cabeça gerava uma “sombra” acústica no ouvido mais afastado da fonte sonora, e que a extensão dessa sombra depende da frequência (ver Figura 26). Ele constatou, então, que

a sombra causa uma diferença nos níveis de pressão sonora interauricular e que esse, por sua vez, era o fator dominante para a localização lateral de sinais de alta frequência, enquanto, para os sinais de baixa frequência, a diferença temporal interauricular é o fator dominante. O atraso é devido à distância entre os ouvidos, e que, em função disso, o ouvido mais afastado da fonte recebe o som mais atrasado que o ouvido mais próximo. Esses dois fatores, diferenças de nível interauriculares (ILD - Interaural Level Differences) e diferenças temporais interauriculares (ITD - Interaural Time Differences), conhecido como a *Teoria Duplex* (MACPHERSON; MIDDLEBROOKS, 2002), estão envolvidos na nossa percepção de localização das fontes sonoras. É em torno de 1,5 kHz que essas duas características podem ser divididas em relação à sua relevância na localização, pois abaixo dessa frequência o comprimento de onda é maior que as dimensões da cabeça e, portanto, a cabeça não representa um obstáculo à onda sonora. Diante disso, a ILD torna-se muito pequena e, conseqüentemente, insuficiente para explicar o efeito de localização. Assim, nessa faixa, o sistema auditivo consegue melhor localizar a fonte a partir da ITD. Por outro lado, acima de 1,5 kHz, sons nessa faixa são facilmente localizados pelo efeito da diferença de intensidade entre os ouvidos, já que as dimensões da cabeça são maiores que o comprimento de onda e, por conseguinte, a cabeça constitui uma interferência à chegada do som no ouvido mais distante da fonte. Além disso, vale ressaltar também que esses fatores só são cabíveis quando o som é proveniente de uma única direção fixa. Conforme descrito na Seção 1.2, em recintos fechados, o som que o ouvido percebe é a composição do som direto e as subseqüentes reflexões sofridas pela onda sonora nas superfícies do interior desse recinto. No entanto, o som que intercepta primeiro o receptor é o principal determinante da percepção de direção da fonte (HAAS, 1972), visto que, as reflexões precoces são integradas com o som direto, ou seja, elas não são escutadas como sons isolados, mas sim como um reforço ao som direto (HARTMANN, 1983). Por outro lado, dependendo do nível sonoro das reflexões tardias, o sistema auditivo irá percebê-las como ecos.

Para exemplificar os fatos acima destacados, considere a Figura 27, a qual mostra as HRIRs (Figura 27a) para ambos os ouvidos e as correspondentes magnitudes (Figura 27b) e atrasos de fases (Figura 27c) das HRTFs, para a direção $\phi = 45^\circ$ e $\theta = 0^\circ$, sendo assim, o ouvido esquerdo está mais próximo da fonte sonora. O atraso de fase,

Figura 26 - Sombra acústica gerada pela fenômeno de difração na cabeça



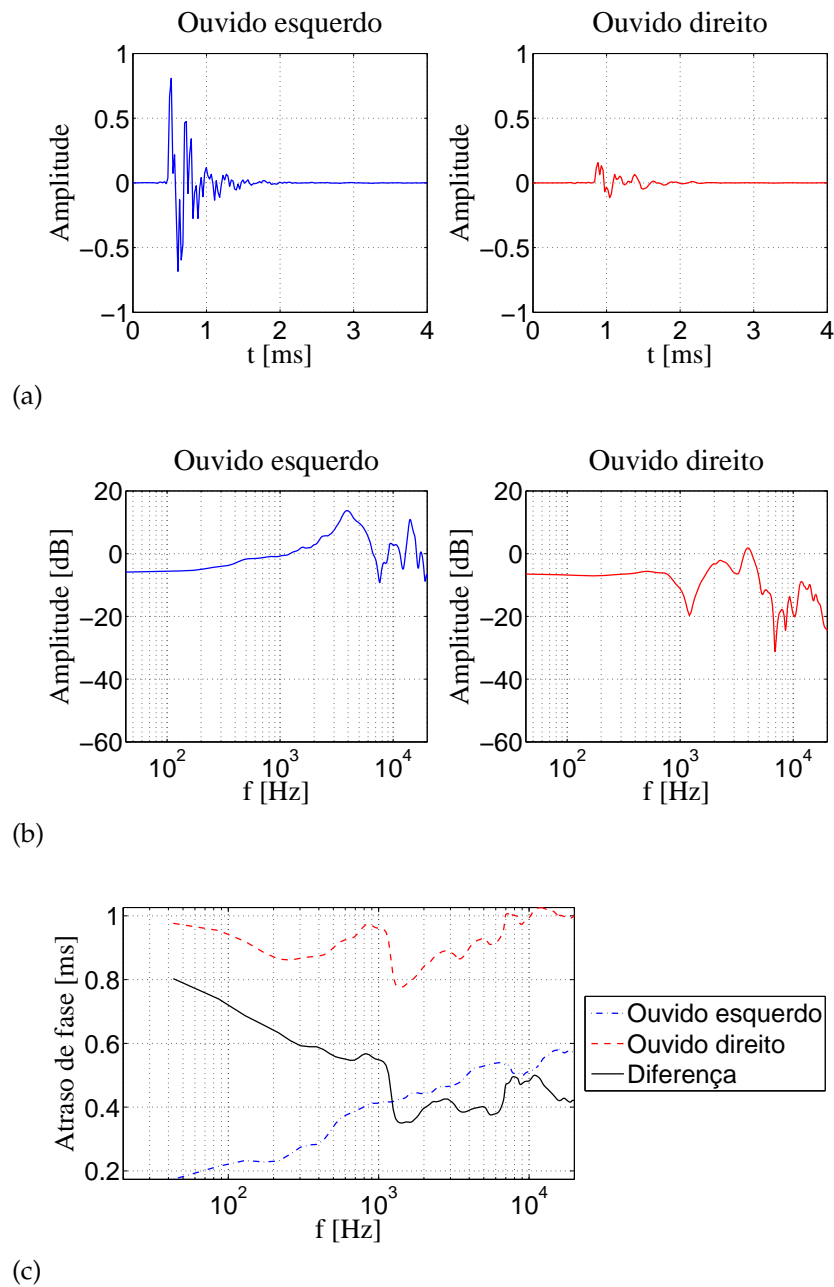
Fonte: BISTAFA, 2011.

mencionado anteriormente, é definido como (BOMHARDT, 2017)

$$\tau_p = \frac{-\angle(\text{HRTF})}{2\pi f}, \quad (34)$$

onde $\angle(\text{HRTF})$ é a fase da HRTF em rad e f é a frequência de análise em Hz. Da Figura 27a pode-se notar que a intensidade do som que atinge o ouvido esquerdo é mais intenso do que o som que chega ao ouvido direito, essa diferença sendo causada pelos efeitos da difração e da absorção devido à presença da cabeça e, além disso, o som que chega ao ouvido direito é mais atrasado, devido à maior distância da fonte. Na Figura 27b observa-se que a magnitude das componentes de alta frequência são mais atenuadas para os sons que chegam ao ouvido direito. Nessa mesma figura é possível notar que a magnitude é maior na faixa de frequências entre 1 kHz e 5 kHz, já que esta é a região de maior sensibilidade do ouvido humano (ver Figura 5). Na curva da diferença do atraso de fase mostrada na Figura 27c verifica-se que a diferença é maior para as frequências mais baixas, isso deve-se ao fato de que a difração de ondas em torno da cabeça faz com que as ondas de baixa frequência se movam mais lentamente do que as ondas de alta frequência (KENDALL, 1995; CONSTAN; HARTMANN, 2003).

Figura 27 - HRIRs e HRTFs para $\phi = 45^\circ$ e $\theta = 0^\circ$



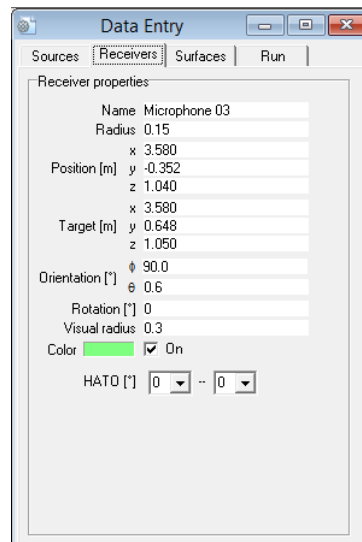
Legenda: (a) - HRIRs; (b) - HRTFs (magnitude); (c) - HRTFs (atraso de fase).

Fonte: O autor, 2017.

Diante disso, para a modelagem do receptor biauricular é necessária a informação da HRTF para cada direção do raio que intercepta o receptor. Para isso, podem-se utilizar técnicas que permitam a sua modelagem ou um banco de dados, conforme descrito na Introdução. A modelagem utilizada neste trabalho será apresentada no Capítulo 4.

Por fim, a Figura 28 mostra a aba *Receivers* da janela *Data Entry* do código RAIOS, onde pode-se destacar os principais dados de entrada que são a posição e a orientação.

Figura 28 - Aba *Receivers* da janela *Data Entry* do código RAIOS



Fonte: O autor, 2017.

2.5 Método tradicional de geração das respostas impulsivas biauriculares da sala

Considera-se, aqui, que o sistema sala-fonte-receptor seja linear e invariante no tempo (SLIT), uma vez que as posições das fontes e dos receptores são fixos, a temperatura e a umidade relativa do ar mantêm-se constantes e não há correntes de ar (XIANG; BLAUERT, 1993). Logo, esse sistema pode ser completamente caracterizado pela sua resposta impulsiva $h(t)$ de maneira que a resposta $y(t)$ desse sistema a um sinal $x(t)$ qualquer seja a convolução entre $x(t)$ e $h(t)$ (KUTTRUFF, 2009).

Para o processo de geração das respostas impulsivas biauriculares (BRIRs), $b_E(t)$ e $b_D(t)$, onde os subscritos E e D referem-se aos ouvidos esquerdo e direito, respectivamente, e t é o tempo, são necessárias as seguintes informações:

- para cada raio acústico que chega ao receptor, seu espectro de potência (E_n), o tempo de chegada (τ_n) e os ângulos de chegada (ϕ_n – azimute e θ_n – elevação), sendo o índice n o número do raio;
- as HRIRs, dos dois ouvidos, $h^E(\theta_n, \phi_n, t)$ e $h^D(\theta_n, \phi_n, t)$, da cabeça que está sendo simulada. Denota-se aqui as HRTFs correspondentes às HRIRs como $H^E(\theta_n, \phi_n, f)$ e $H^D(\theta_n, \phi_n, f)$.

As BRIRs, neste trabalho, foram obtidas a partir dos seguintes procedimentos:

1. Primeiramente são gerados os raios acústicos que chegam ao receptor, obtidos pelo código computacional RAIOS, para um determinado par fonte-receptor da sala que está sendo simulada;

2. Com base nos ângulos (elevação e azimute) de chegada dos raios sonoros determina-se qual das HRIRs será utilizada. Como é muito improvável que a frente de onda atinja o receptor em uma direção onde existe uma função medida, tem-se duas possibilidades para obter a HRIR para essa direção: a primeira, seria um processo de interpolação, e a outra, consiste em utilizar a HRIR medida que se encontra mais próxima. Neste trabalho, utiliza-se a segunda opção, já que o banco de HRIRs utilizado é suficientemente discretizado (discretização de 1°) tanto no azimute quanto na elevação.
3. Em seguida, o espectro de energia E_n é utilizado para modificar as HRIRs de uma dada direção de chegada do raio (θ_n, ϕ_n) . Essas modificações são dadas por:

$$\begin{aligned}\bar{h}^E(\theta_n, \phi_n, t) &= \mathfrak{F}^{-1} \left\{ E_n(f) \cdot H^E(\theta_n, \phi_n, f) \right\} ; \\ \bar{h}^D(\theta_n, \phi_n, t) &= \mathfrak{F}^{-1} \left\{ E_n(f) \cdot H^D(\theta_n, \phi_n, f) \right\} ,\end{aligned}\tag{35}$$

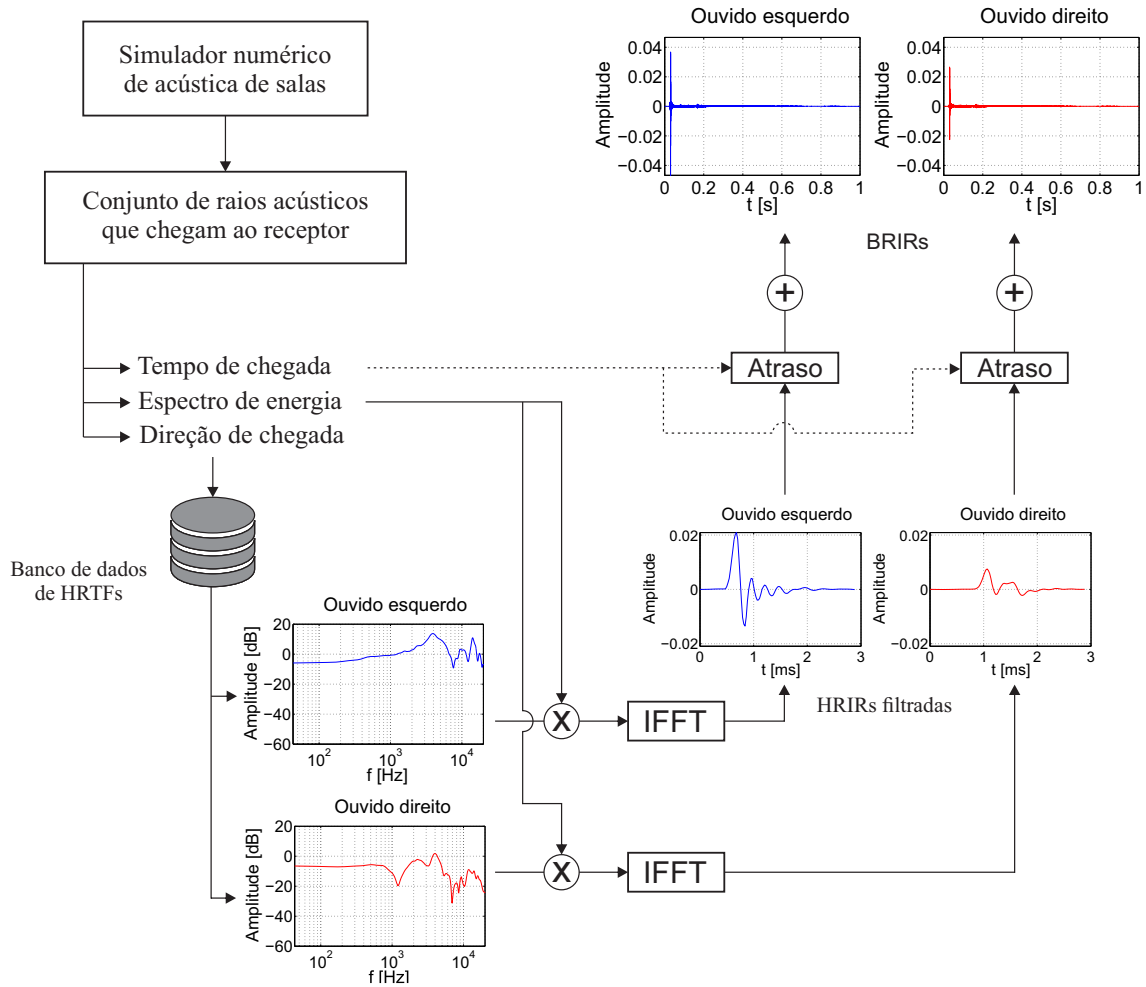
para $n = 1, 2, \dots, N$, onde N é o número de raios que chegam ao receptor e \mathfrak{F}^{-1} é a transformada inversa de Fourier.

4. Quando as HRIRs modificadas \bar{h}^E e \bar{h}^D estão finalmente calculadas, estas são atrasadas de acordo com os valores de tempo de chegada τ_n de cada raio, cuja soma resultará no par de respostas impulsivas biauriculares da sala:

$$\begin{aligned}b_E(t) &= \sum_{n=1}^N \bar{h}^E(\theta_n, \phi_n, t - \tau_n) ; \\ b_D(t) &= \sum_{n=1}^N \bar{h}^D(\theta_n, \phi_n, t - \tau_n) .\end{aligned}\tag{36}$$

A Figura 29 ilustra de forma esquemática a geração das BRIRs utilizando o procedimento descrito acima. Note que o processo para obter cada par de HRIRs modificadas (Eqs. (35)) é repetido para cada raio que chega ao receptor. Como o número destes é da ordem de 10^5 , o cômputo das HRIRs modificadas pode ser computacionalmente muito custoso. Por isso, neste trabalho temos como objetivo apresentar uma outra abordagem na modelagem das HRIRs modificadas, via *redes neurais artificiais*.

Figura 29 - Esquema de geração das BRIRs via método tradicional



Fonte: O autor, 2017.

2.6 Aurilização

Uma vez calculadas as BRIRs, pode-se gerar a aurilização por meio da convolução de $b_E(t)$ e $b_D(t)$ com um sinal anecóico $s(t)$ (sinal gravado em câmara anecóica), isto é,

$$\begin{aligned} y_E(t) &= b_E(t) * s(t); \\ y_D(t) &= b_D(t) * s(t). \end{aligned} \quad (37)$$

Esses sinais são, então, reproduzidos para um ouvinte a partir de fones de ouvido devidamente equalizados. Daí, a escuta desses sinais geram a sensação de estar imerso na sala simulada, na posição do receptor.

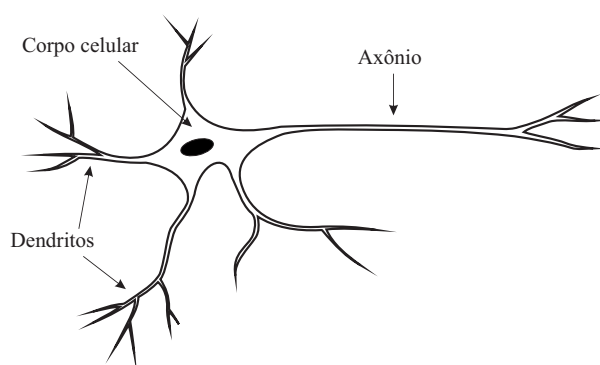
3 REDES NEURAIS ARTIFICIAIS

Neste capítulo são abordados os aspectos mais relevantes das redes neurais artificiais (RNAs) para a compreensão da proposta reportada nesta tese, uma vez que o tema é extenso. Além disso, são discutidos os conceitos fundamentais das redes *perceptron* de múltiplas camadas e de funções de base radiais. A discussão de ambas redes é necessária para o discernimento entre a metodologia apresentada neste trabalho e a desenvolvida em Naranjo (2014).

3.1 As redes neurais biológicas

As redes neurais artificiais (RNAs) são baseadas no paradigma do processamento do cérebro humano, que possui em torno de 86 bilhões de neurônios (AZEVEDO et al., 2009). O neurônio é constituído pelas seguintes partes principais: *dendritos*, *corpo celular* e *axônio*, como ilustra a Figura 30. Muito resumidamente, o dendrito é incumbido da condução dos estímulos vindos de outros neurônios até o corpo celular. O corpo celular, por sua vez, tem a função de coletar e combinar as informações oriundas de outros neurônios e novas informações são geradas. Por fim, o axônio é responsável por transmitir essas novas informações a outros neurônios. Através das *sinapses* os neurônios se comunicam, formando a rede neural biológica. Essa rede possibilita uma extraordinária capacidade de armazenamento e processamento de informações.

Figura 30 - Principais componentes do neurônio biológico



Fonte: O autor, 2017.

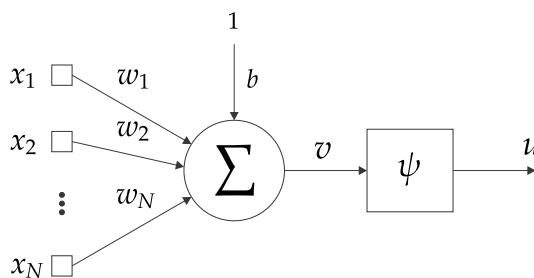
A rede neural artificial constitui um sistema de processamento de informações baseado em modelos matemáticos simplificados dos neurônios biológicos cujo processo

de aprendizagem ocorre por experiência. O conhecimento obtido pela rede através dos exemplos é armazenado na forma de pesos sinápticos das conexões, os quais serão ajustados com o objetivo de tomar decisões corretas a partir de novas entradas, ou seja, a rede tem a capacidade de *generalizar* a informação aprendida. O processo de ajuste de pesos sinápticos é realizado pelo algoritmo de aprendizado. As redes neurais artificiais são ferramentas úteis para resolver muitos tipos de problemas. Esses problemas podem ser, por exemplo, de classificação, de agrupamento, de otimização, de aproximação ou de previsão. Existem várias redes neurais disponíveis para cada tipo de problema. Para usá-las de forma eficaz, é importante compreender as características, pontos fortes e limitações de cada uma.

3.2 Neurônio artificial

As redes neurais artificiais possuem em sua estrutura um conjunto de unidades processadoras, chamadas de neurônios artificiais, os quais são modelos bem simplificados dos neurônios biológicos. O modelo geral dos neurônios (ver Figura 31), utilizado nas diferentes arquiteturas de redes neurais artificiais, possui em sua estrutura os seguintes elementos:

Figura 31 - Representação de um neurônio artificial



Fonte: O autor, 2017.

- *Entradas* (x_1, x_2, \dots, x_N): o modelo é composto de terminais de entrada (equivalentes aos dendritos), os quais recebem os sinais oriundos do meio externo;
- *Pesos sinápticos* (w_1, w_2, \dots, w_N): esses valores servem para ponderar os sinais de entrada. Os pesos permitem quantificar a pertinência de cada variável de entrada;

- *Combinador linear* (Σ): seu papel é realizar a soma ponderada das entradas com os seus respectivos pesos sinápticos como fatores de ponderação;
- *Limiar de ativação* ou *bias* (b): tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação;
- *Potencial de ativação* (v): é a soma do resultado do combinador linear com o limiar de ativação. O valor obtido determina se o neurônio produzirá um potencial excitatório ou inibitório;
- *Função de ativação* (ψ): sua finalidade é restringir a amplitude da saída de um neurônio;
- *Saída* (u): é o valor fornecido pelo neurônio em relação a um determinado conjunto de sinais de entrada.

As expressões matemáticas que relacionam a entrada e a saída dos neurônios são dadas pelas seguintes equações:

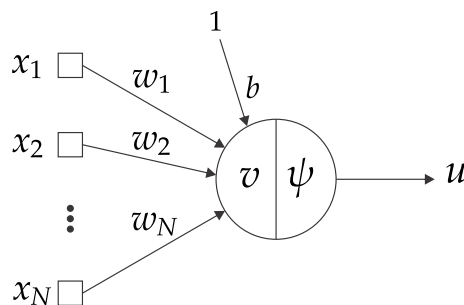
$$v = \sum_{j=1}^N (w_j x_j) + b \quad (38)$$

e

$$u = \psi(v) . \quad (39)$$

Com o objetivo de simplificar a representação do neurônio artificial ilustrado na Figura 31, neste trabalho o neurônio será representado como ilustra a Figura 32.

Figura 32 - Modelo de um neurônio artificial utilizado neste trabalho



Fonte: O autor, 2017.

Embora esse modelo tenha muitas simplificações em relação ao neurônio biológico, um conjunto desses neurônios interconectados, formando uma rede neural, é capaz de solucionar problemas de alta complexidade.

3.2.1 Funções de ativação

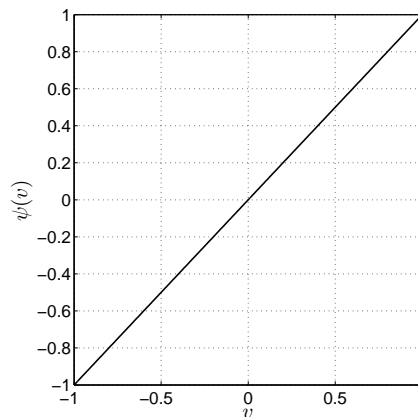
As funções de ativação mais comumente utilizadas são descritas a seguir.

- *Linear*: esse tipo de função de ativação, mostrada na Figura 33, é definida por

$$\psi(v) = v. \quad (40)$$

Pode ser vista como um combinador linear, já que a saída dessa função é o próprio sinal de entrada, que é a soma ponderada entre as entradas do neurônio e os seus pesos sinápticos (ver Eq. (38)).

Figura 33 - Função de ativação linear



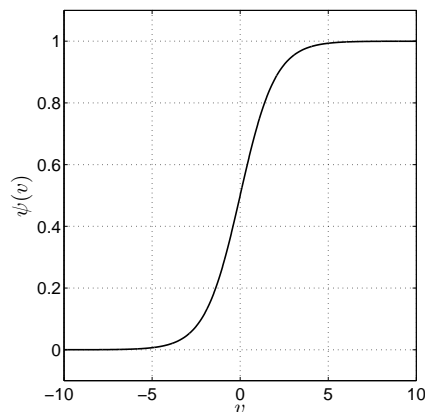
Fonte: O autor, 2017.

- *Logística*: esse tipo de função de ativação se assemelha a uma curva em “S”, mostrada na Figura 34, definida por

$$\psi(v) = \frac{1}{1 + \exp(-v)}. \quad (41)$$

Essa função é limitada no intervalo entre 0 e 1, estritamente crescente e diferenciável em todo seu domínio. A diferenciabilidade é uma característica importante na teoria de redes neurais, como será visto na Seção 3.5.1.

Figura 34 - Função de ativação logística



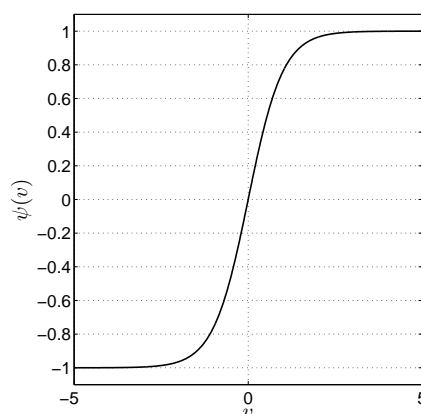
Fonte: O autor, 2017.

- *Tangente hiperbólica*: para esse tipo de função de ativação, mostrada na Figura 35, tem-se

$$\psi(v) = \tanh(v) = \frac{1 - \exp(-v)}{1 + \exp(-v)}. \quad (42)$$

Diferentemente da função logística, a função tangente hiperbólica é limitada no intervalo -1 e 1 , e antissimétrica em relação à origem.

Figura 35 - Função de ativação tangente hiperbólica



Fonte: O autor, 2017.

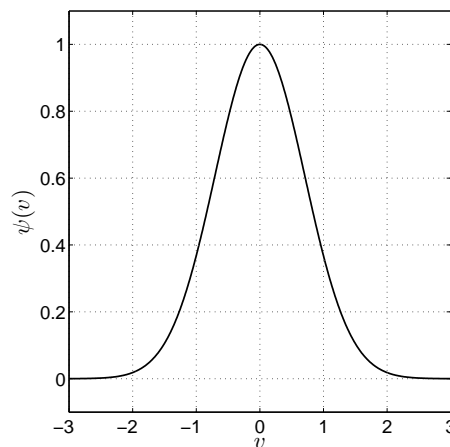
Tanto a função logística quanto a tangente hiperbólica pertencem à classe das funções denominadas de sigmoidais. O nome “sigmóide” vem da forma em S do seu gráfico.

- *Gaussiana*: para esse tipo de função de ativação, mostrada na Figura 36, tem-se

$$\psi(v) = \exp(-v^2), \quad (43)$$

sendo $v = \frac{(x - \mu)}{\sqrt{2}\sigma}$, onde μ é o centro e σ é a largura, cujos efeitos sobre a função são, respectivamente, translação horizontal e escalamento horizontal (compressão / dilatação).

Figura 36 - Função de ativação gaussiana



Fonte: O autor, 2017.

A seleção do tipo de função depende do problema a ser tratado. Para os casos onde as saídas desejadas sejam contínuas, normalmente utilizam-se as funções lineares na camada de saída. Agora, caso as saídas sejam discretas (problemas de classificação) usam-se as funções não-lineares tanto na camada intermediária quanto na camada de saída. As redes de retro propagação geralmente utilizam a função tangente hiperbólica por sua propriedade de ser diferenciável. As redes neurais do tipo Funções de Base Radiais utilizam neurônios com funções de ativação radiais. Essas funções representam uma classe cujo valor diminui ou aumenta em relação à distância de um ponto central, como a função gaussiana, como será mostrado na Seção 3.6.

3.3 Arquitetura

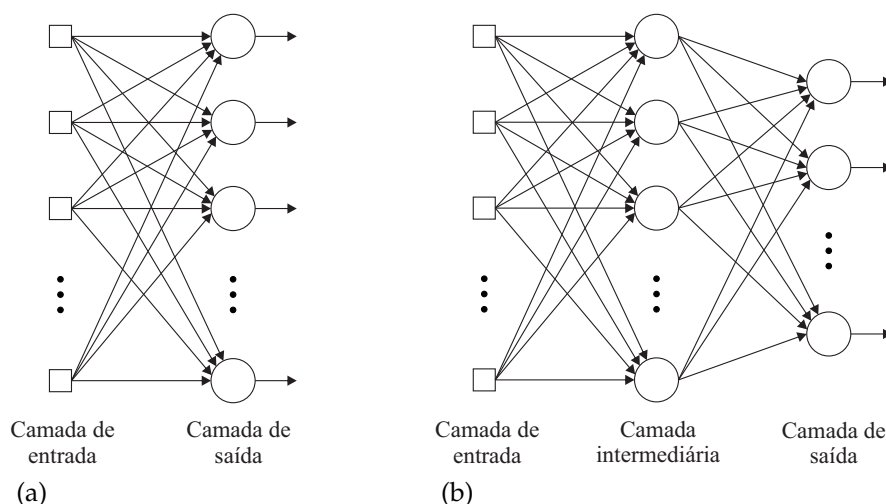
A arquitetura das redes neurais artificiais descreve a maneira como os neurônios estão conectados uns em relação aos outros. Existem dois tipos de arquiteturas, descritas em seguida.

1. Redes neurais não-recorrentes: não possuem conexões de realimentação, ou seja, não existem conexões ligando um neurônio de uma camada a outro neurônio de uma camada anterior nem a um neurônio da mesma camada. Assim, o sinal de entrada se propaga através da rede, camada a camada, até alcançar a saída. Duas importantes redes que apresentam essa arquitetura são:

- Redes progressivas de uma única camada (*single-layer feed-forward neural networks*). Essas redes possuem apenas uma camada.
- Redes progressivas multicamadas (*multi-layer feed-forward neural networks*). Essas redes possuem uma ou mais camadas intermediárias. A adição de camadas intermediárias é relevante quando o tamanho da camada de entrada é grande (HAYKIN, 1999). Nessas redes os neurônios em cada camada têm como suas entradas apenas os sinais de saída da camada antecedente.

Note que as unidades da camada de entrada são apenas unidades de armazenamento que passam os sinais sem modificá-los. Diante disso, a rede mostrada na Figura 37a é designada como uma rede de uma camada e a da Figura 37b como rede de duas camadas.

Figura 37 - Redes neurais não-recorrentes

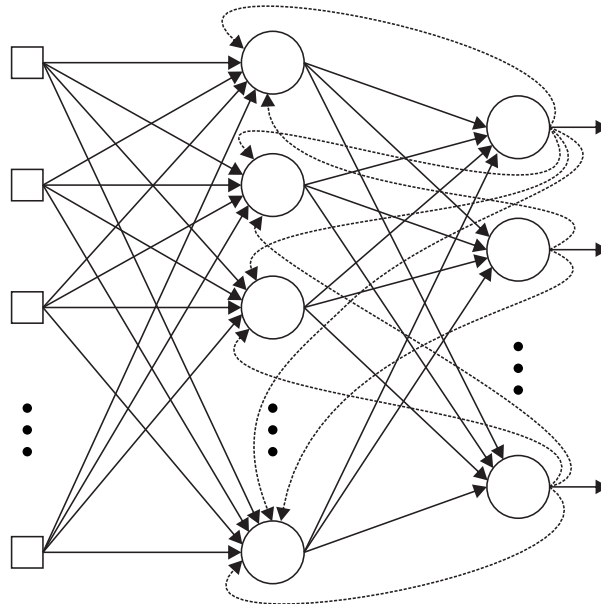


Legenda: (a) - Rede progressiva de uma camada; (b) - Rede progressiva de duas camadas.

Fonte: O autor, 2017.

2. Redes neurais recorrentes: nessas redes os sinais de saída da rede retornam como entrada das camadas intermediárias. Portanto, as saídas são determinadas pelas entradas atuais e pelas saídas anteriores. A Figura 38 ilustra a arquitetura de uma rede neural recorrente.

Figura 38 - Rede neural recorrente



Fonte: O autor, 2017.

3.4 Aprendizado

Uma das propriedades mais importantes das redes neurais artificiais é certamente a habilidade de aprender por meio de exemplos. Segundo Haykin (1999) “Aprendizagem é o processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimulação pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira pela qual a modificação dos parâmetros ocorre”.

O processo de aprendizado é feito mediante o ajuste dos pesos sinápticos e bias, utilizando um *algoritmo de aprendizagem*. De um modo geral, esse ajuste dos pesos sinápticos é obtido a partir de um procedimento iterativo dado pela expressão

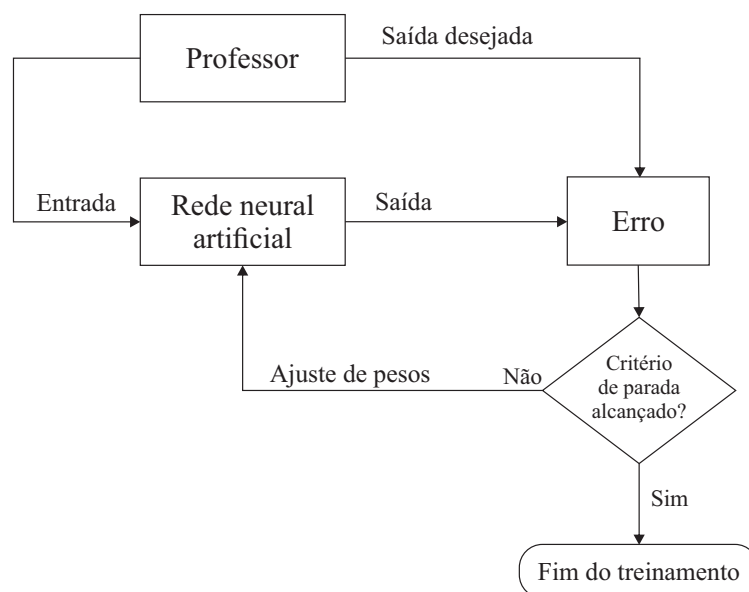
$$\mathbf{w}(\lambda + 1) = \mathbf{w}(\lambda) + \Delta\mathbf{w}(\lambda), \quad (44)$$

onde $\mathbf{w}(\lambda)$ e $\mathbf{w}(\lambda + 1)$ são os vetores de pesos nos instantes λ e $\lambda + 1$, e $\Delta\mathbf{w}(\lambda)$ é o ajuste (correção) dos pesos. Os diferentes algoritmos de aprendizado se diferenciam uns dos

outros pelo modo como é realizado o cálculo do ajuste dos pesos. Além disso, esses algoritmos podem ser agrupados em dois principais paradigmas de aprendizagem, descritos a seguir.

1. Aprendizado supervisionado. Nesse método, é necessário o conhecimento de amostras (exemplos) dos sinais de entrada da rede e das suas respectivas saídas desejadas. Esse conjunto de pares de entrada e saída forma o conjunto de *padrões de treinamento*. No aprendizado supervisionado existe um “professor” que tem o papel de examinar a saída da rede a cada sinal de entrada. A Figura 39 mostra a representação esquemática do aprendizado supervisionado. As etapas desse treinamento são: o padrão de entrada é apresentado a rede; a saída corrente é comparada com a saída desejada; a partir do erro encontrado os pesos são ajustados de acordo com um algoritmo que minimize esse erro. O algoritmo mais relevante é o de retropropagação de erros (*error back-propagation*), o qual será descrito na Seção 3.5.

Figura 39 - Aprendizado supervisionado



Fonte: O autor, 2017.

2. Aprendizado não-supervisionado. Nesse método, por sua vez, é necessário apenas os padrões de entrada. O processo desse treinamento é extrair as características dos padrões de treinamento, agrupando-os de acordo com suas similaridades. Os modelos mais conhecidos são os mapas auto-organizáveis (SOM - *Self-Organizing Maps*) e os modelos teoria da ressonância adaptativa (ART - *Adaptive Resonance Theory*). Para mais

detalhes, o leitor pode consultar Kohonen (1982) e Carpenter e Grossberg (1988).

3.5 Redes perceptron de múltiplas camadas

As redes neurais artificiais do tipo perceptron de múltiplas camadas (MLP - *Multilayer Perceptron*) são redes progressivas multicamadas, cujos pesos são ajustados utilizando um algoritmo de aprendizagem supervisionado. O desenvolvimento de uma rede MLP envolve a determinação de:

1. o número de nós na camada de entrada;
2. o número de neurônios na camada de saída;
3. o número de camadas intermediárias;
4. o número de neurônios em cada camada intermediária;
5. a função de ativação dos neurônios em cada camada;
6. os pesos sinápticos que interconectam os neurônios;
7. os padrões de treinamento.

A especificação do número de neurônios em cada camada é notável para sua performance, pois determina a capacidade da rede para solucionar problemas com determinada complexidade. O número de nós na entrada e o número de neurônios na camada de saída dependem do problema em questão a ser resolvido. Os itens (3) e (4) determinam a complexidade da rede neural e, infelizmente, *não há regras para tal especificação*. O que se faz, na prática, é estimar de forma empírica, mas buscando o dimensionamento mínimo que satisfaça os requisitos do problema. Redes superdimensionadas podem gerar os chamados erros de *overfitting*, enquanto as redes subdimensionadas podem causar *underfitting*. Na primeira situação, a rede memoriza os padrões de treinamento de tal forma que perde a capacidade de generalização. Já na segunda situação, a rede não consegue aprender, devido ao baixo número de neurônios. Vale ressaltar que os erros de *overfitting* e *underfitting* também estão relacionados ao tempo de treinamento da rede, como será analisado em mais detalhes adiante.

Cumprir notar que, na maioria dos problemas práticos, necessita-se de uma ou, no máximo, duas camadas intermediárias. A utilização de um número maior do que o necessário pode-se causar problemas na convergência global da rede, dado que as camadas intermediárias são levadas em conta na estimativa do erro na fase de treinamento.

As funções de ativação geralmente utilizadas nas camadas intermediárias, são as não-lineares, sigmóide, e na camada de saída, as lineares. A utilização de apenas funções lineares resultaria em uma rede de uma única camada, visto que transformações lineares sucessivas resultam em uma única transformação linear.

Para a determinação dos pesos sinápticos a rede MLP utiliza o algoritmo de aprendizado supervisionado retro-propagação (*backpropagation algorithm*). Esse treinamento ocorre em duas fases.

- Fase *forward*: um vetor de entrada (um padrão de treinamento) é apresentado à rede e se propaga camada por camada para produzir uma saída. Nessa fase todos os pesos sinápticos são fixos e a saída da rede é comparada com a saída desejada (alvo). Assim, nessa etapa, o objetivo é obter o erro de saída após a propagação do vetor de entrada por todas as camadas.
- Fase *backward*: os pesos sinápticos nessa fase são todos ajustados com a regra de aprendizado por correção de erro. A saída da rede, determinada na fase *forward*, é subtraída da saída desejada para produzir um *signal de erro*. Esse sinal é transmitido para trás através da rede no caminho oposto ao das conexões sinápticas. Dessa forma, os pesos são ajustados de tal maneira que a resposta obtida pela rede se aproxime da saída desejada.

Além da correta determinação dos parâmetros da rede, um dos fatores que influencia a capacidade de generalização é a determinação dos padrões de treinamento, que consiste na adequada escolha do conjunto de treinamento. Embora, existam fórmulas para estimar o tamanho da amostra de treinamento suficiente para uma boa generalização, elas normalmente não preveem a quantidade realmente necessária (HAYKIN, 1999).

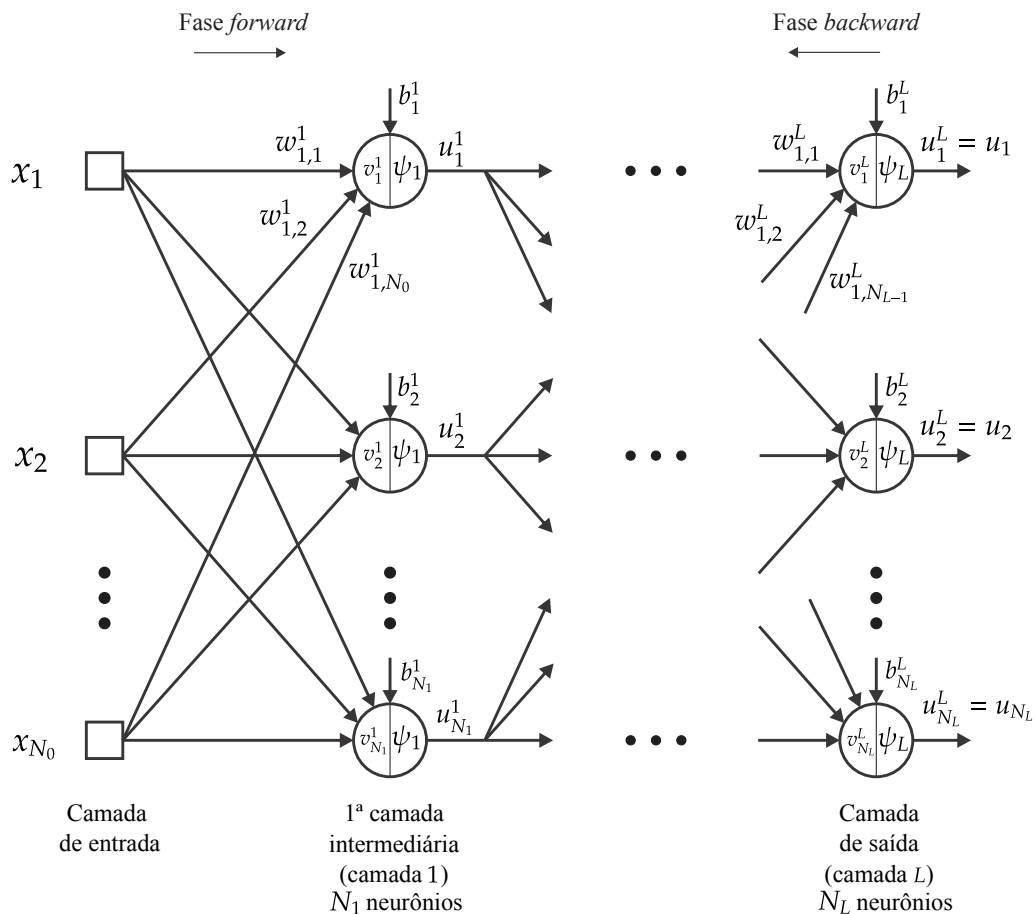
3.5.1 Algoritmo *backpropagation*

Para uma melhor compreensão do algoritmo *backpropagation*, faz-se necessário a definição de algumas variáveis e parâmetros. Primeiramente, seja $\{(\mathbf{x}(p), \mathbf{y}(p))\}_{p=1}^P$ o conjunto de padrões de treinamento, sendo $\mathbf{x}(p)$ é o p -ésimo padrão de entrada, $\mathbf{y}(p)$ o p -ésimo padrão de saída desejado e P é o número de padrões de treinamento. A Figura 40 ilustra uma rede MLP, com L camadas, juntamente com as variáveis que auxiliarão na dedução do algoritmo. Nessa figura utilizaram-se as seguintes convenções:

- Função de ativação (ψ_l): l é o índice da camada.

- Pesos sinápticos ($w_{i,j}^l$): l é o índice da camada, i é o índice do neurônio da l -ésima camada e j é o índice do componente do vetor de entrada do neurônio.
- Bias (b_i^l): l é o índice da camada, i é o índice do neurônio da l -ésima camada.
- Potencial de ativação (v_i^l): l é o índice da camada, i é o índice do neurônio da l -ésima camada.
- Saída do neurônio (u_i^l): l é o índice da camada, i é o índice do neurônio da l -ésima camada.

Figura 40 - Notações para a dedução do algoritmo *backpropagation*



Fonte: O autor, 2017.

Como a fase *forward* corresponde à obtenção do erro de saída da rede, calcula-se, após a apresentação do vetor de entrada, a saída de todos os neurônios camada a camada até a geração da saída da rede. A saída dos neurônios de cada camada é dada

por:

$$u_i^1 = \psi_1(v_i^1), \text{ onde } v_i^1 = \sum_{j=1}^{N_0} w_{i,j}^1 x_j + b_i^1, \text{ para } i = 1, 2, \dots, N_1; \quad (45)$$

$$u_i^2 = \psi_2(v_i^2), \text{ onde } v_i^2 = \sum_{j=1}^{N_1} w_{i,j}^2 u_j^1 + b_i^2, \text{ para } i = 1, 2, \dots, N_2; \quad (46)$$

⋮

$$u_i^L = \psi_L(v_i^L), \text{ onde } v_i^L = \sum_{j=1}^{N_{L-1}} w_{i,j}^L u_j^{L-1} + b_i^L, \text{ para } i = 1, 2, \dots, N_L. \quad (47)$$

Assim, a expressão geral para o cálculo da saída dos neurônios de cada camada é:

$$\begin{cases} u_i^0 = x_i, \text{ para } i = 1, 2, \dots, N_0; \\ u_i^{l+1} = \psi_{l+1}(v_i^{l+1}), \text{ onde } v_i^{l+1} = \sum_{j=1}^{N_l} w_{i,j}^{l+1} u_j^l + b_i^{l+1}, \text{ para } i = 1, 2, \dots, N_{l+1}, \end{cases} \quad (48)$$

sendo $l = 0, 1, 2, \dots, L - 1$. As saídas dos neurônios da última camada são considerados como a saída da rede e serão representados por

$$u_i = u_i^L, \text{ para } i = 1, 2, \dots, N_L. \quad (49)$$

Uma vez calculada a saída de todos os neurônios da camada de saída, o próximo passo é calcular o sinal de erro, isto é,

$$e_i(p) = y_i(p) - u_i(p), \text{ para } i = 1, 2, \dots, N_L, \quad (50)$$

onde $y_i(p)$ é a saída desejada.

Na fase *backward*, os parâmetros livres (pesos e bias) são ajustados a partir da minimização de uma *função de custo* baseada no sinal de erro, de tal maneira que a saída da rede se aproxime da saída desejada. Uma função de custo comumente utilizada é a *soma dos erros quadráticos*

$$C(p) = \sum_{i=1}^{N_L} e_i^2(p) = \sum_{i=1}^{N_L} (y_i(p) - u_i(p))^2. \quad (51)$$

Note que soma dos erros quadráticos é uma função de todos os parâmetros livres da rede e, como consequência, define uma *superfície de erro* sobre o espaço desses parâmetros. Além disso, como essa função é dependente da função de ativação, que na maioria das aplicações são não-lineares, tal superfície terá mínimos locais e globais. Os métodos de otimização que serão abordados neste trabalho utilizam o gradiente do

erro.

Uma outra função custo para avaliar o desempenho do algoritmo *backpropagation* é o erro médio quadrático

$$\bar{C} = \frac{1}{P} \sum_{p=1}^P C(p). \quad (52)$$

A minimização dessa função custo é o objetivo do processo de treinamento.

Antes de se apresentar o algoritmo *backpropagation*, vale ressaltar que existem dois modos de treinamento (HAYKIN, 2009): *modo padrão-a-padrão (on-line)* e *modo por lote (off-line ou batch)*. No primeiro modo, os pesos são atualizados após a apresentação de cada padrão. Desse modo, o algoritmo tende a levar a rede a aprender melhor o último padrão apresentado. Por isso, é necessário embaralhar aleatoriamente a ordem de apresentação dos padrões. As principais vantagens dessa abordagem são: requerem menos memória, é mais rápido e menos susceptível a mínimos locais. Por outro lado, tem como desvantagens: tornar a rede em geral instável e requerer o controle da taxa de aprendizado. Enquanto, no segundo modo, os pesos são ajustados só depois da apresentação de todos os padrões. Sendo assim, cada ajuste leva em consideração o sinal de erro de todas as amostras de treinamento, e tem como vantagens a estimativa mais precisa do vetor gradiente e o fato de ser estável; mas tem como desvantagem ser mais lento.

Primeiramente, consideremos o primeiro modo. Logo, os pesos serão ajustados a partir da soma dos erros quadráticos ($C(p)$) e o desempenho global da rede será avaliada de acordo com o erro médio quadrático (\bar{C}).

Para ajustar os pesos sinápticos e bias o algoritmo *backpropagation* utiliza o método do *gradiente descendente*. A ideia é que o vetor de pesos seja ajustado a partir da direção contrária ao gradiente do erro, ou seja,

$$w_{i,j}^l(\lambda + 1) = w_{i,j}^l(\lambda) - \alpha \frac{\partial C}{\partial w_{i,j}^l}; \quad (53)$$

$$b_i^l(\lambda + 1) = b_i^l(\lambda) - \alpha \frac{\partial C}{\partial b_i^l}, \quad (54)$$

onde α é a *taxa de aprendizagem*. Para determinar $w_{i,j}^l(\lambda + 1)$ e $b_i^l(\lambda + 1)$ precisam ser determinados, respectivamente, os gradientes $\partial C / \partial w_{i,j}^l$ e $\partial C / \partial b_i^l$. Aplicando a regra da cadeia, obtém-se:

$$\frac{\partial C}{\partial w_{i,j}^l} = \frac{\partial C}{\partial v_i^l} \frac{\partial v_i^l}{\partial w_{i,j}^l}; \quad (55)$$

$$\frac{\partial C}{\partial b_i^l} = \frac{\partial C}{\partial v_i^l} \frac{\partial v_i^l}{\partial b_i^l}. \quad (56)$$

O segundo termo de ambas as equações pode ser facilmente computado, já que o potencial de ativação v_i^l é uma função explícita dos pesos e bias. Assim,

$$\frac{\partial v_i^l}{\partial w_{i,j}^l} = \frac{\partial}{\partial w_{i,j}^l} \left(\sum_{k=1}^{N_{l-1}} w_{i,k}^l u_k^{l-1} + b_i^l \right) = \frac{\partial (w_{i,j}^l u_j^{l-1})}{\partial w_{i,j}^l} = u_j^{l-1}; \quad (57)$$

$$\frac{\partial v_i^l}{\partial b_i^l} = \frac{\partial}{\partial b_i^l} \left(\sum_{k=1}^{N_{l-1}} w_{i,k}^l u_k^{l-1} + b_i^l \right) = \frac{\partial b_i^l}{\partial b_i^l} = 1. \quad (58)$$

Substituindo (57) em (55) e (58) em (56), obtêm-se:

$$\frac{\partial \mathcal{C}}{\partial w_{i,j}^l} = \delta_i^l u_j^{l-1}; \quad (59)$$

$$\frac{\partial \mathcal{C}}{\partial b_i^l} = \delta_i^l, \quad (60)$$

onde δ_i^l é definido como o gradiente local em relação ao i -ésimo neurônio da l -ésima camada, sendo definido como

$$\delta_i^l \equiv \frac{\partial \mathcal{C}}{\partial v_i^l}. \quad (61)$$

Logo, as Eqs. (53) e (54) podem ser reescritas da seguinte forma:

$$w_{i,j}^l(\lambda + 1) = w_{i,j}^l(\lambda) - \alpha \delta_i^l u_j^{l-1}; \quad (62)$$

$$b_i^l(\lambda + 1) = b_i^l(\lambda) - \alpha \delta_i^l. \quad (63)$$

Observe, das Eqs. (62) e (63), que o termo fundamental para o ajuste dos pesos e das bias é o cálculo dos δ 's. Para a camada de saída $l = L$, tem-se

$$\delta_i^L = \frac{\partial \mathcal{C}}{\partial v_i^L} = \frac{\partial}{\partial v_i^L} \left(\sum_{k=1}^{N_L} (y_k(p) - u_k(p))^2 \right) = -2(y_i(p) - u_i(p)) \frac{\partial u_i}{\partial v_i^L}. \quad (64)$$

Como

$$\frac{\partial u_i}{\partial v_i^L} = \frac{\partial u_i^L}{\partial v_i^L} = \frac{\partial \psi_L(v_i^L)}{\partial v_i^L} = \psi'_L(v_i^L), \quad (65)$$

pode-se escrever a Eq. (64) na forma

$$\delta_i^L = -2(y_i(p) - u_i(p)) \psi'_L(v_i^L), \quad (66)$$

onde $\psi'(\cdot)$ denota a derivada da função de ativação em relação ao argumento.

Agora, como as camadas intermediárias ($l = 1, 2, \dots, L - 1$) não têm acesso

direto aos valores das saídas desejadas da rede, os pesos da camada intermediária l serão ajustados a partir das estimativas dos erros de saída gerados pelos neurônios da camada $l + 1$, uma vez que estes já foram anteriormente ajustados. Desse modo, os ajustes dos pesos de uma determinada camada intermediária só poderá ser realizada após o ajuste dos pesos de todas as camadas posteriores a ela.

Diante disso, para desenvolver a expressão para o cálculo dos δ 's dos neurônios das camadas intermediárias, aplica-se a regra da cadeia na expressão (61),

$$\delta_i^l = \sum_{k=1}^{N_{l+1}} \frac{\partial \mathcal{C}}{\partial v_k^{l+1}} \frac{\partial v_k^{l+1}}{\partial v_i^l} = \sum_{k=1}^{N_{l+1}} \delta_k^{l+1} \frac{\partial v_k^{l+1}}{\partial v_i^l}. \quad (67)$$

Sabendo que $v_k^{l+1} = \sum_{m=1}^{N_l} w_{k,m}^{l+1} u_m^l + b_k^{l+1}$, a Eq. (67) pode ser reescrita como

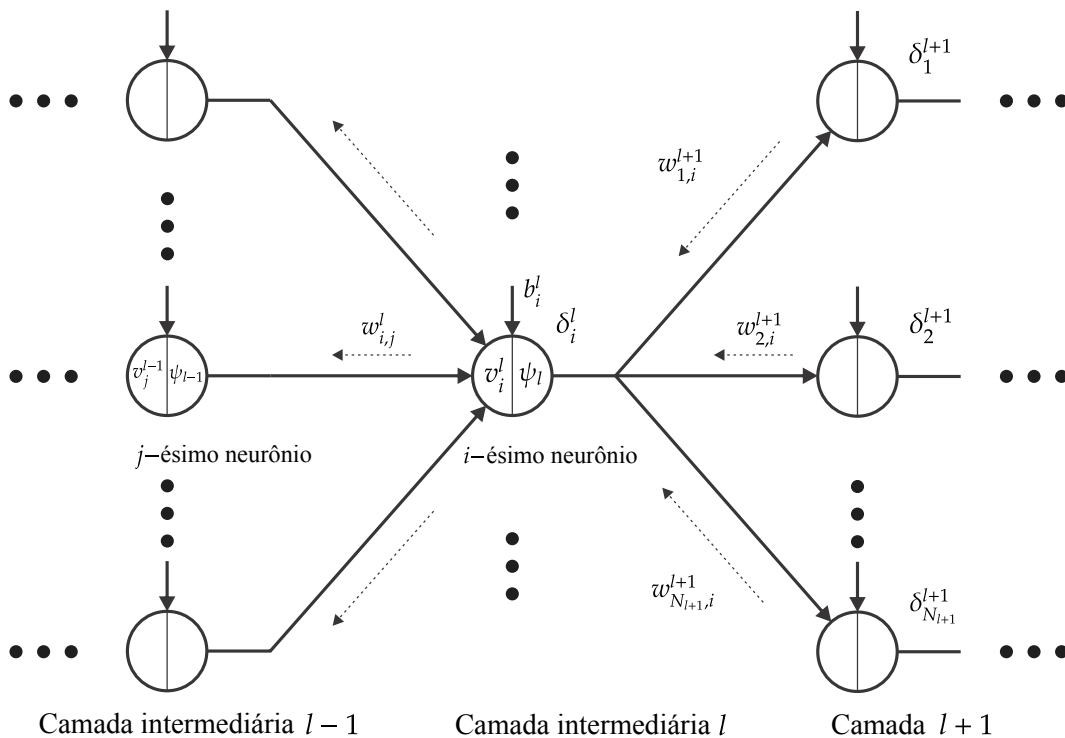
$$\delta_i^l = \sum_{k=1}^{N_{l+1}} \delta_k^{l+1} w_{k,i}^{l+1} \frac{\partial u_i^l}{\partial v_i^l} = \sum_{k=1}^{N_{l+1}} \delta_k^{l+1} w_{k,i}^{l+1} \frac{\partial \psi_l(v_i^l)}{\partial v_i^l} = \psi_l'(v_i^l) \sum_{k=1}^{N_{l+1}} \delta_k^{l+1} w_{k,i}^{l+1}. \quad (68)$$

Cabe salientar, que a Eq. (68) mostra que o cálculo dos δ 's das camadas intermediárias são obtidos a partir da soma dos δ 's da camada seguinte, ponderada pelos pesos sinápticos, os quais já foram previamente ajustados. A Figura 41 apresenta os parâmetros envolvidos para o cálculo do δ do i -ésimo neurônio da camada intermediária l .

Apresenta-se sucintamente a seguir a fase treinamento utilizando o algoritmo *backpropagation*, para a aprendizagem usando padrão-por-padrão.

- 1 Inicializar os parâmetros livres com valores aleatórios no intervalo $[-0,5; 0,5]$ ou $[-1, 1]$.
- 2 Informar a taxa de aprendizado α .
- 3 Especificar a precisão requerida e o número máximo de iterações $\lambda_{\text{máx}}$.
- 4 Inicializar o contador de iterações $\lambda = 0$.
- 5 Apresentar o vetor $\mathbf{x}(p)$ pertencente ao conjunto de treinamento.
- 6 Fase *forward*:
 - 6.1 Calcular as saídas dos neurônios utilizando a Eq. (48);
 - 6.2 Calcular o sinal de erro com a Eq. (50).
- 7 Fase *backward*:
 - 7.1 Calcular o gradiente local utilizando as Eqs. (66) e (68);

Figura 41 - Ajuste dos pesos sinápticos da l -ésima camada intermediária



Fonte: O autor, 2017.

7.2 Atualizar os parâmetros livres utilizando as Eqs. (62) e (63).

- 8 Se todos os vetores do conjunto de treinamento foram apresentados, ir para o passo 9. Caso contrário, voltar para o passo 5.
- 9 Embaralhar aleatoriamente a ordem dos vetores do conjunto de treinamento.
- 10 Determinar o erro médio quadrático com a Eq. (52).
- 11 Se o erro médio quadrático for menor que a precisão requerida ou o número de iterações for maior do que o número máximo de iterações, então armazenar os parâmetros livres da rede. Caso contrário, incrementar o contador λ ($\lambda = \lambda + 1$) e voltar ao passo 5.

Após o término do treinamento da rede MLP, pode-se utilizar a rede para estimar as saídas para novos padrões de entrada. A *fase de execução* pode ser descrita pelo seguinte algoritmo.

- 1 Fornecer um novo padrão de entrada.
- 2 Carregar os parâmetros livres, ajustados durante a fase de treinamento.
- 3 Propagar o vetor de entrada até a camada de saída utilizando a Eq. (48).

A implementação do modo de treinamento por lote requer o cálculo do gradiente do erro médio quadrático para o ajuste dos pesos. Desse modo, calculando a derivada parcial da Eq. (52) em relação $w_{i,j}^l$, obtém-se

$$\frac{\partial \bar{C}}{\partial w_{i,j}^l} = \frac{1}{P} \sum_{p=1}^P \frac{\partial C(p)}{\partial w_{i,j}^l}, \quad (69)$$

sendo $\partial C(p)/\partial w_{i,j}^l$ calculado da mesma forma descrita anteriormente. Logo, os procedimentos de ajuste dos parâmetros livres são obtidos a partir das seguintes equações:

$$w_{i,j}^l(\lambda + 1) = w_{i,j}^l(\lambda) - \frac{\alpha}{P} \sum_{p=1}^P \delta_i^l(p) u_j^{l-1}(p); \quad (70)$$

$$b_i^l(\lambda + 1) = b_i^l(\lambda) - \frac{\alpha}{P} \sum_{p=1}^P \delta_i^l(p), \quad (71)$$

onde $\delta_i^l(p)$ é obtido a partir das Eqs. (66) e (68).

Apresenta-se sucintamente agora o funcionamento da fase treinamento utilizando o algoritmo *backpropagation*, para a aprendizagem usando lote.

- 1 Inicializar os parâmetros livres com valores aleatórios no intervalo $[-0,5; 0,5]$ ou $[-1, 1]$;
- 2 Informar a taxa de aprendizado α .
- 3 Especificar a precisão requerida e o número máximo de iterações $\lambda_{\text{máx}}$.
- 4 Inicializar o contador de iterações $\lambda = 0$.
- 5 Apresentar o vetor $\mathbf{x}(p)$ pertencente ao conjunto de treinamento.
- 6 Fase *forward*:
 - 6.1 Calcular as saídas dos neurônios utilizando a Eq. (48);
 - 6.2 Calcular o sinal de erro com a Eq. (50).
- 7 Se todos os vetores do conjunto de treinamento foram apresentados, ir para o passo 8. Caso contrário, voltar para o passo 5;
- 8 Fase *backward*:
 - 8.1 Calcular o gradiente local utilizando as Eqs. (66) e (68);
 - 8.2 Atualizar os parâmetros livres utilizando as Eqs. (70) e (71).
- 9 Determinar o erro quadrático médio com a Eq. (52).

10 Se o erro quadrático médio for menor que a precisão requerida ou o número de iterações for maior que o número máximo de iterações, então armazenar os parâmetros livres da rede. Caso contrário, incrementar o contador λ ($\lambda = \lambda + 1$) e voltar ao passo 5.

A performance de cada modo de treinamento depende do problema. Segundo Moller (1993) o modo padrão-a-padrão apresenta melhor desempenho em problemas que possuem um grande número de informações redundantes e as superfícies de erro pouco complexas, enquanto o modo por lote tem um melhor funcionamento em problemas que não apresentam essas características.

Embora o algoritmo *backpropagation* seja um dos mais importantes no desenvolvimento das redes neurais, sua taxa de convergência é muito lenta. Em função disso, diversas técnicas foram desenvolvidas com o objetivo de acelerar a convergência e melhorar o desempenho do treinamento. Pode-se destacar algumas variações: *backpropagation* com *momentum* (RUMELHART; MCCLELLAND, 1986), Quickprop (FAHLMAN, 1988), Newton (BATTITI, 1992), Rprop (RIEDMILLER, 1994), Levenberg-Marquardt (HAGAN; MENHAJ, 1994), sendo este último o mais utilizado e aceito como o algoritmo mais eficiente para solucionar diversos problemas, sendo discutido na seção seguinte.

3.5.2 Algoritmo Levenberg-Marquardt

O algoritmo de treinamento Levenberg-Marquardt (LM) ajusta os parâmetros livres a partir da minimização a função custo $\bar{C}(\mathbf{\Omega})$ dada pela Eq. (52). Definindo o vetor de erro como

$$\boldsymbol{\varepsilon}^T = [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_N] = [e_1(1) \ e_2(1) \ \dots \ e_{N_L}(1) \ \dots \ e_1(P) \ e_2(P) \ \dots \ e_{N_L}(P)] , \quad (72)$$

e substituindo a Eq. (51) na Eq. (52), tem-se

$$\bar{C}(\mathbf{\Omega}) = \frac{1}{P} \sum_{p=1}^P \sum_{i=1}^{N_L} e_i^2(p) = \frac{1}{P} \sum_{j=1}^N \varepsilon_j^2 = \frac{\boldsymbol{\varepsilon}^T \cdot \boldsymbol{\varepsilon}}{P} , \quad (73)$$

onde $N = P \cdot N_L$.

Nesse método, o ajuste dos parâmetros livres da rede MLP é dado por (MARQUARDT, 1963; HAGAN; MENHAJ, 1994)

$$\mathbf{\Omega}(\lambda + 1) = \mathbf{\Omega}(\lambda) - [\mathcal{H}(\mathbf{\Omega}(\lambda))]^{-1} \mathbf{J}^T(\mathbf{\Omega}(\lambda)) \boldsymbol{\varepsilon}(\mathbf{\Omega}(\lambda)) , \quad (74)$$

onde

$$\mathbf{J}(\boldsymbol{\Omega}) = \begin{bmatrix} \frac{\partial \varepsilon_1(\boldsymbol{\Omega})}{\partial \omega_1} & \frac{\partial \varepsilon_1(\boldsymbol{\Omega})}{\partial \omega_2} & \cdots & \frac{\partial \varepsilon_1(\boldsymbol{\Omega})}{\partial \omega_M} \\ \frac{\partial \varepsilon_2(\boldsymbol{\Omega})}{\partial \omega_1} & \frac{\partial \varepsilon_2(\boldsymbol{\Omega})}{\partial \omega_2} & \cdots & \frac{\partial \varepsilon_2(\boldsymbol{\Omega})}{\partial \omega_M} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \varepsilon_N(\boldsymbol{\Omega})}{\partial \omega_1} & \frac{\partial \varepsilon_N(\boldsymbol{\Omega})}{\partial \omega_2} & \cdots & \frac{\partial \varepsilon_N(\boldsymbol{\Omega})}{\partial \omega_M} \end{bmatrix} \quad (75)$$

é a matriz Jacobiana,

$$\mathcal{H}(\boldsymbol{\Omega}(\lambda)) = \mathbf{J}^T(\boldsymbol{\Omega}(\lambda)) \mathbf{J}(\boldsymbol{\Omega}(\lambda)) + \xi(\lambda) \mathbf{I} \quad (76)$$

é a matriz Hessiana, sendo \mathbf{I} a matriz identidade e ξ o fator de amortecimento, o qual varia a cada iteração, e

$$\boldsymbol{\Omega}^T = [\omega_1 \ \omega_2 \ \dots \ \omega_M] = [w_{1,1}^1 \ w_{1,2}^1 \ \dots \ w_{N_1, N_0}^1 \ b_1^1 \ \dots \ b_{N_1}^1 \ \dots \ w_{N_L, 1}^L \ \dots \ w_{N_L, N_{L-1}}^L \ b_{N_L}^L] \quad (77)$$

o vetor que armazena todos os parâmetros livres da rede, onde $M = (N_0 + 1)N_1 + (N_1 + 1)N_2 + \dots + (N_L + 1)N_{L-1}$.

Observe que, na prática, não se inverte a matriz $\mathcal{H}(\boldsymbol{\Omega}(\lambda))$. Em vez disso resolve-se o sistema

$$\mathcal{H}(\boldsymbol{\Omega}(\lambda)) \Delta \boldsymbol{\Omega}(\lambda) = -\mathbf{J}^T(\boldsymbol{\Omega}(\lambda)) \varepsilon(\boldsymbol{\Omega}(\lambda)) \quad (78)$$

e, uma vez calculado $\Delta \boldsymbol{\Omega}(\lambda)$, pode-se utilizá-lo para determinar a nova estimativa

$$\boldsymbol{\Omega}(\lambda + 1) = \boldsymbol{\Omega}(\lambda) + \Delta \boldsymbol{\Omega}(\lambda). \quad (79)$$

Além disso, percebe-se que o passo fundamental para a implementação do algoritmo LM é a determinação da matriz Jacobiana $\mathbf{J}(\boldsymbol{\Omega}(\lambda))$.

Os elementos da matriz Jacobiana podem ser determinados de modo semelhante ao algoritmo *backpropagation*. Para isso, decompõe-se essa matriz como uma matriz bloco da seguinte maneira

$$\mathbf{J}(\boldsymbol{\Omega}) = [\mathbf{J}_1(\boldsymbol{\Omega}_1) \ \mathbf{J}_2(\boldsymbol{\Omega}_2) \ \dots \ \mathbf{J}_L(\boldsymbol{\Omega}_L)], \quad (80)$$

onde cada bloco é dado por

$$\mathbf{J}_l(\boldsymbol{\Omega}_l) = \begin{bmatrix} \frac{\partial \varepsilon_1}{\partial w_{1,1}^l} & \frac{\partial \varepsilon_1}{\partial w_{1,2}^l} & \cdots & \frac{\partial \varepsilon_1}{\partial w_{N_l, N_l-1}^l} & \frac{\partial \varepsilon_1}{\partial b_{N_l}^l} \\ \frac{\partial \varepsilon_2}{\partial w_{1,1}^l} & \frac{\partial \varepsilon_2}{\partial w_{1,2}^l} & \cdots & \frac{\partial \varepsilon_2}{\partial w_{N_l, N_l-1}^l} & \frac{\partial \varepsilon_2}{\partial b_{N_l}^l} \\ \vdots & \vdots & & \vdots & \vdots \\ \frac{\partial \varepsilon_N}{\partial w_{1,1}^l} & \frac{\partial \varepsilon_N}{\partial w_{1,2}^l} & \cdots & \frac{\partial \varepsilon_N}{\partial w_{N_l, N_l-1}^l} & \frac{\partial \varepsilon_N}{\partial b_{N_l}^l} \end{bmatrix}, \quad \text{para } l = 1, 2, \dots, L \quad (81)$$

sendo

$$\boldsymbol{\Omega}_l^T = [w_{1,1}^l \ w_{1,2}^l \ \dots \ w_{1, N_l-1}^l \ b_1^l \ \dots \ w_{N_l,1}^l \ w_{N_l,2}^l \ \dots \ w_{N_l, N_l-1}^l \ b_{N_l}^l] \quad (82)$$

o vetor que armazena os parâmetros livres da l -ésima camada da rede neural.

Vamos agora determinar os elementos da matriz (81). Iniciamos aplicando a regra da cadeia a esses elementos:

$$\frac{\partial \varepsilon_n}{\partial w_{i,j}^l} = \frac{\partial e_k(p)}{\partial w_{i,j}^l} = \frac{\partial e_k(p)}{\partial v_i^l(p)} \frac{\partial v_i^l(p)}{\partial w_{i,j}^l} = \bar{\delta}_{i,n}^l u_j^{l-1}(p); \quad (83)$$

$$\frac{\partial \varepsilon_n}{\partial b_i^l} = \frac{\partial e_k(p)}{\partial b_i^l} = \frac{\partial e_k(p)}{\partial v_i^l(p)} \frac{\partial v_i^l(p)}{\partial b_i^l} = \bar{\delta}_{i,n}^l, \quad (84)$$

sendo

$$\bar{\delta}_{i,n}^l \equiv \frac{\partial \varepsilon_n}{\partial v_i^l(p)} = \frac{\partial e_k(p)}{\partial v_i^l(p)}, \quad (85)$$

onde, da Eq. (77), $n = k + (p - 1)N_L$, para $k = 1, 2, \dots, N_L$ e $p = 1, 2, \dots, P$.

Assim como ocorre no algoritmo *backpropagation*, as derivadas dos erros do algoritmo LM dependem também do cálculo dos δ 's, sendo nesse algoritmo denotado por $\bar{\delta}$. Para a camada de saída $l = L$, tem-se

$$\bar{\delta}_{i,n}^L = \frac{\partial \varepsilon_n}{\partial v_i^L(p)} = \frac{\partial e_k(p)}{\partial v_i^L(p)} = \frac{\partial}{\partial v_i^L(p)} (y_k(p) - u_k(p)) = -\frac{\partial u_k(p)}{\partial v_i^L(p)} = \begin{cases} -\psi'_L(v_i^L(p)) & \text{se } i = k \\ 0 & \text{se } i \neq k \end{cases}. \quad (86)$$

A expressão para o cálculo dos $\bar{\delta}$'s para as camadas intermediárias é igual ao que foi desenvolvido no algoritmo *backpropagation*, isto é,

$$\bar{\delta}_{i,n}^l = \psi'_i(v_i^l(p)) \sum_{j=1}^{N_{l+1}} \bar{\delta}_{j,n}^{l+1} w_{ji}^{l+1}, \quad \text{para } l = 1, 2, \dots, L - 1. \quad (87)$$

Finalmente, apresentam-se os passos para a execução do algoritmo LM.

- 1 Fornecer um conjunto de treinamento $\{(\mathbf{x}(p), \mathbf{y}(p))\}_{p=1}^P$.
- 2 Inicializar o contador de iterações $\lambda = 0$.
- 3 Inicializar os parâmetros livres $\mathbf{\Omega}(0)$.
- 4 Inicializar fator de amortecimento $\xi(0)$.
- 5 Fornecer a precisão desejada ϵ e número máximo de iterações.
- 6 Fase *forward*:
 - 6.1 Calcular as saídas dos neurônios utilizando a Eq. (48).
- 7 Fase *backward*:
 - 7.1 Calcular os blocos da matriz Jacobiana:
 - 7.1.1 Calcular $\mathbf{J}_L(\mathbf{\Omega}_L)$ utilizando as Eqs. (83), (84) e (86);
 - 7.1.2 Calcular $\mathbf{J}_l(\mathbf{\Omega}_l)$, para $l = L - 1, \dots, 2, 1$, utilizando as Eqs. (83), (84) e (87).
 - 7.2 Determinar $\Delta\mathbf{\Omega}(\lambda)$ a partir da resolução do sistema (78);
 - 7.3 Calcular $\mathbf{\Omega}(\lambda + 1)$ com a Eq. (79).
- 8 Com os parâmetros livres $\mathbf{\Omega}(\lambda + 1)$ ajustados, calcular as saídas dos neurônios utilizando a Eq. (48).
- 9 Calcular o erro médio quadrático de acordo com a Eq. (73).
- 10 Avaliar o critério de parada.
 - 10.1 Se $\bar{C}(\mathbf{\Omega}) < \epsilon$ ou o número de iterações λ for maior que o máximo permitido, então armazenar os parâmetros livres. Caso contrário, incrementar o contador λ , ajustar o fator de amortecimento ξ e voltar para o passo 7.

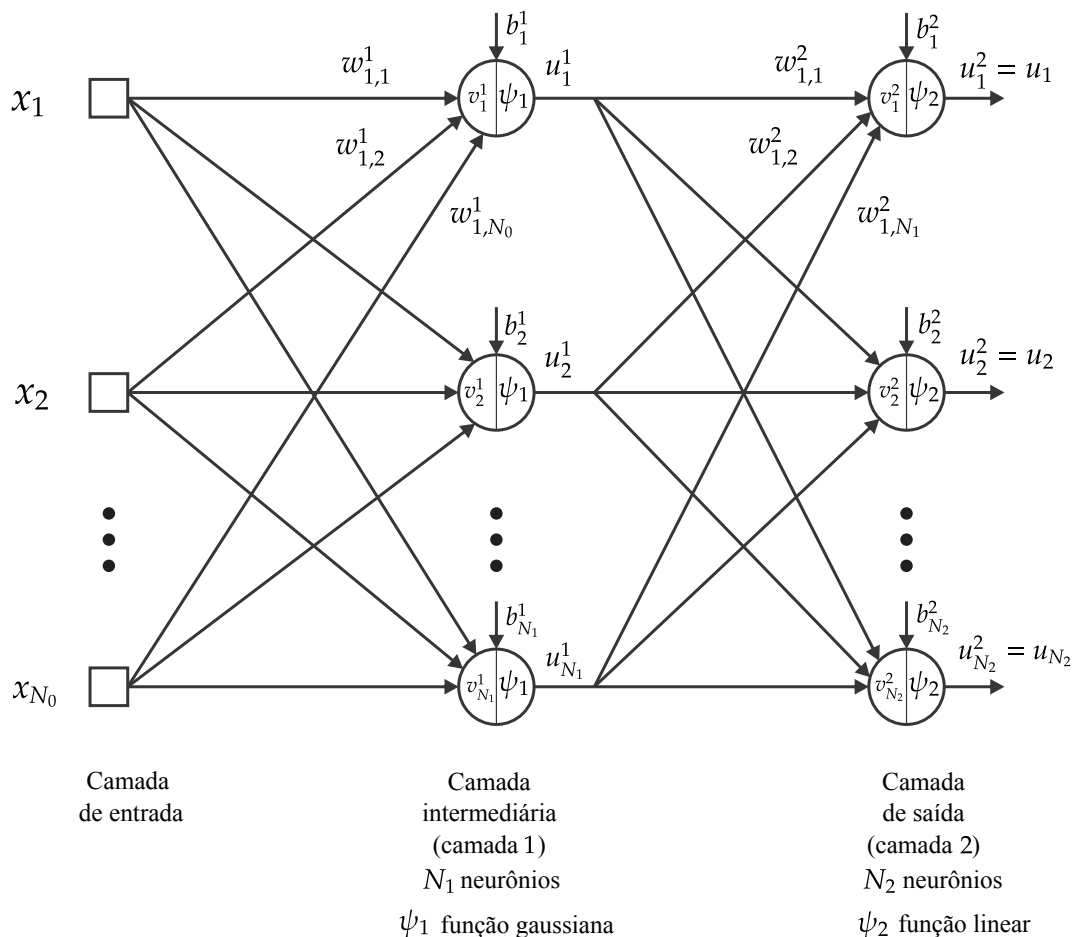
O ajuste do fator de amortecimento, de acordo a proposta de Marquardt, ocorre da seguinte maneira: se o valor $\bar{C}(\mathbf{\Omega})$ é reduzido de uma iteração para outra, então o valor de ξ também é reduzido $\xi(\lambda + 1) = \xi(\lambda)/c_1$, onde c_1 é uma constante arbitrária maior que um. Caso contrário $\xi(\lambda + 1) = c_2\xi(\lambda)$, onde c_2 é uma constante arbitrária maior que um.

Por meio dessas modificações, o algoritmo LM consegue um treinamento mais rápido que o *backpropagation* (HAGAN; MENHAJ, 1994), apresentado na Seção 3.5.1. Entretanto, problemas de convergência podem ocorrer quando a matriz Jacobiana for mal-condicionada (MCKEOWN et al., 1997). Outra limitação refere-se ao tempo de treinamento, pois quando a quantidade de padrões aumenta, o processo de treinamento tende a ser mais lento.

3.6 As redes de Funções de Base Radiais

As redes de Funções de Base Radiais (RBF - *Radial Basis Function*) são redes com apenas uma camada intermediária cujos neurônios utilizam funções de base radiais, enquanto na camada de saída é utilizada a função linear (BROOMHEAD; LOWE, 1988). A Figura 42 ilustra a arquitetura de uma rede RBF, juntamente com as notações que serão utilizadas. As notações são iguais às utilizadas na Seção 3.5, com algumas interpretações diferentes dos parâmetros livres dos neurônios da camada intermediária. A rede é constituída de N_0 nós na camada de entrada, N_1 neurônios na camada intermediária e N_2 neurônios na camada de saída.

Figura 42 - Arquitetura de uma rede neural do tipo *Radial Basis Function*



Fonte: O autor, 2017.

As funções de base radiais pertencem a uma classe de funções cuja resposta cresce ou decresce monotonicamente em relação à distância de um ponto central. O centro e a taxa de crescimento ou decrescimento são parâmetros a serem ajustados na fase de treinamento. A função de base radial que decresce monotonicamente mais comumente

utilizada é a função gaussiana, e que será utilizada neste trabalho. Assim, a saída do j -ésimo neurônio da camada intermediária, para uma entrada multidimensional, é dado por

$$u_j^1 = \psi_1(v_j^1) = \exp\left(-\left(v_j^1\right)^2\right), \quad (88)$$

com

$$v_j^1 = \frac{\|\mathbf{x} - \boldsymbol{\mu}_j\|}{\sqrt{2}\sigma_j}, \quad (89)$$

onde

$$\boldsymbol{\mu}_j = [w_{j,1}^1 \ w_{j,2}^1 \ \dots \ w_{j,N_0}^1]^T \quad (90)$$

é o vetor centro,

$$\sigma_j = \frac{1}{\sqrt{2}b_j^1} \quad (91)$$

é a largura da função gaussiana e o operador $\|\cdot\|$ é a norma euclidiana.

Considerando as Eqs. (90) e (91), a Eq. (89) pode ser reescrita em termos dos pesos sinápticos e da bias da seguinte forma

$$v_j^1 = b_j^1 \sqrt{\sum_{n=1}^{N_0} (x_n - w_{j,n}^1)^2}. \quad (92)$$

Diante do exposto acima, pode-se observar duas diferenças no cômputo do potencial de ativação v_j^1 dos neurônios da camada intermediária das redes RBF em relação as redes MLP com duas camadas. A primeira é que ao invés de calcular a soma ponderada das entradas, determina-se a distância entre o vetor de entrada e o vetor centro. A segunda é que a *bias* é multiplicada e não somada.

Já a saída da rede é obtida da mesma forma das redes MLP, ou seja, a resposta gerada pelo i -ésimo neurônio de saída da rede, é dada por

$$u_i = u_i^2 = \psi_2(v_i^2) = v_i^2 = \sum_{j=1}^{N_1} w_{i,j}^2 u_j^1 + b_i^2, \quad (93)$$

onde $w_{i,j}^2$ é o peso sináptico que conecta o i -ésimo neurônio da camada de saída ao j -ésimo neurônio da camada intermediária.

O desenvolvimento de uma rede RBF compreende a determinação:

1. do número de nós na camada de entrada (N_0);
2. do número de neurônios na camada de saída (N_2);
3. do número de neurônios na camada intermediária (N_1);
4. dos parâmetros das funções de base radial: vetor centro μ e a largura σ ;
5. dos pesos sinápticos que interconectam os neurônios (w);
6. dos padrões de treinamento $\{(\mathbf{x}(p), \mathbf{y}(p))\}_{p=1}^P$, onde $\mathbf{x}(p)$ é o p -ésimo padrão de entrada, $\mathbf{y}(p)$ é o p -ésimo padrão de saída desejado e P é o número de padrões de treinamento.

As dimensões dos vetores de entrada e saída da rede dependem da complexidade do problema em questão, ou seja, dependem das informações disponíveis e do que se quer obter, e se todas essas são necessárias e suficientes para a resolução do problema.

Pode-se determinar o número de neurônios da camada intermediária como sendo igual ao número de padrões de entrada. Assim, a rede mapeia com exatidão todos os vetores de entrada com as saídas desejadas. Contudo, esse mapeamento exato é indesejável, pois pode levar ao *overfitting*. Uma outra maneira de operar consiste em utilizar um número de neurônios dessa camada menor que a quantidade de padrões de treinamento. Uma quantidade muito pequena também não é muito desejável, pois a rede gastará muito tempo tentando encontrar a melhor representação.

Diferentes algoritmos de treinamento podem ser usados para o ajuste dos parâmetros livres (pesos, centros, larguras) das redes RBF. A maioria desses métodos possuem duas fases. Na primeira fase, o número de funções radiais e seus parâmetros são determinados. Esses parâmetros são ajustados por algoritmos não-supervisionados. Enquanto na segunda fase são ajustados os pesos dos neurônios de saída.

O treinamento de uma rede RBF é geralmente mais rápido do que o das redes MLP, uma vez que os métodos de aprendizagem para o ajuste dos parâmetros das funções radiais e dos pesos sinápticos podem ser encadeados sequencialmente (MOODY; DARKEN, 1989; MULGREW, 1996), viabilizando um treinamento otimizado devido à divisão de tarefas.

Neste trabalho, os parâmetros foram calculados da seguinte maneira (LIPPMANN, 1989; CHEN et al., 1992): primeiramente os centros são obtidos utilizando o algoritmo não-supervisionado *K-means* (MACQUEEN, 1967; DUDA; HART, 1973; MOODY; DARKEN, 1989). Uma vez que os centros tenham sido calculados, determinam-se a partir destes as larguras. E, por fim, depois de definidos os parâmetros das funções radiais, os parâmetros livres da camada de saída são computados utilizando os mesmos procedimentos que foram usados para a camada de saída das redes MLP (HAYKIN, 1999). Note que esse processo de aprendizagem híbrido acelera o processo treinamento,

uma vez que os parâmetros livres não são estimados simultaneamente. Diferentemente dessa rede, a rede MLP ajusta todos os parâmetros livres ao mesmo tempo.

3.6.1 Primeira fase do treinamento

O algoritmo *K-means* é uma técnica de *clustering* (agrupamento) que particiona o conjunto dos padrões de entrada em N_1 conjuntos disjuntos \mathcal{S}_j contendo K_j vetores cada. Esse particionamento é realizado da seguinte forma: inicialmente são selecionados aleatoriamente N_1 vetores de entrada para serem os centros de N_1 conjuntos, e os outros vetores de entrada são agrupados ao conjunto que tiver a menor distância do centro mais próximo. Após essa etapa, os centros são recalculados através da média dos vetores pertencentes ao conjunto, na forma

$$\mu_j = \frac{1}{K_j} \sum_{\mathbf{x}(p) \in \mathcal{S}_j} \mathbf{x}(p), \quad (94)$$

para $j = 1, 2, \dots, N_1$. Na sequência, os vetores de entrada são atribuídos ao centro mais próximo

$$\mathcal{S}_j = \{ \mathbf{x}(p) : \|\mathbf{x}(p) - \mu_j\| \leq \|\mathbf{x}(p) - \mu_i\|, \forall i \neq j, i = 1, 2, \dots, N_1 \}. \quad (95)$$

Essa atualização dos centros e dos vetores associados ao centro mais próximo é feita até o algoritmo convergir, ou seja, quando os centros não mais precisarem ser alterados. Em outras palavras, a convergência do algoritmo ocorre quando a soma dos quadrados das distâncias entre os vetores de entrada e o centro do conjunto \mathcal{S}_j ao qual eles pertencem,

ou seja, $\sum_{j=1}^{N_1} \sum_{\mathbf{x}(p) \in \mathcal{S}_j} \|\mathbf{x}(p) - \mu_j\|^2$, atinge o seu valor mínimo.

Assim, ao final da convergência desse algoritmo toma-se o centro de cada conjunto como o centro das funções radiais.

Infelizmente, não existe nenhum método formal para a determinação do número N_1 de centros (que corresponde ao número de neurônios na camada intermediária). Assim, esse número é definido por tentativa e erro.

Com os centros já calculados, determina-se a largura de cada função radial σ_j através da média sobre todas as distâncias euclidianas entre o centro do conjunto \mathcal{S}_i e o centro de \mathcal{S}_j , ou seja,

$$\sigma_j = \frac{1}{N_1} \sum_{\substack{i=1 \\ i \neq j}}^{N_1} \|\mu_j - \mu_i\|. \quad (96)$$

3.6.2 Segunda fase do treinamento

Após o ajuste dos parâmetros livres da camada intermediária, o próximo passo é calcular os pesos sinápticos e as bias dos neurônios da camada de saída. Para isso são utilizados os mesmos procedimentos usados para o ajuste dos parâmetros livres da última camada da rede MLP, descritos na Seção 3.5.1. Portanto, em contraste com a primeira fase do treinamento da rede RBF, a segunda fase utiliza um processo de aprendizado supervisionado.

Utilizando as notações apresentadas na Figura 42 e as Eqs. (62), (63) e (66), o ajuste dos pesos sinápticos e das bias é dado por:

$$w_{i,j}^2(\lambda + 1) = w_{i,j}^2(\lambda) - \alpha \delta_i^2 u_j^1; \quad (97)$$

$$b_i^2(\lambda + 1) = b_i^2(\lambda) - \alpha \delta_i^2, \quad (98)$$

para $i = 1, 2, \dots, N_2$ e $j = 1, 2, \dots, N_1$, onde

$$\delta_i^2 = -2(y_i(p) - u_i(p)). \quad (99)$$

Substituindo as Eqs. (88) e (99) nas Eqs. (97) e (98), obtém-se:

$$w_{i,j}^2(\lambda + 1) = w_{i,j}^2(\lambda) + 2\alpha(y_i(p) - u_i(p)) \exp\left(-\left(b_j^1\right)^2 \sum_{n=1}^{N_0} (x_n(p) - w_{j,n}^1)^2\right); \quad (100)$$

$$b_i^2(\lambda + 1) = b_i^2(\lambda) + 2\alpha(y_i(p) - u_i(p)). \quad (101)$$

Note-se que essas equações de ajuste são para uma implementação do modo de treinamento padrão-a-padrão. Para uma aprendizagem por lote, deve-se utilizar as Eqs. (70) e (71) e também as Eqs. (99) e (88). Logo,

$$w_{i,j}^2(\lambda + 1) = w_{i,j}^2(\lambda) + \frac{2\alpha}{P} \sum_{p=1}^P (y_i(p) - u_i(p)) \exp\left(-\left(b_j^1\right)^2 \sum_{n=1}^{N_0} (x_n(p) - w_{j,n}^1)^2\right); \quad (102)$$

$$b_i^2(\lambda + 1) = b_i^2(\lambda) + \frac{2\alpha}{P} \sum_{p=1}^P (y_i(p) - u_i(p)), \quad (103)$$

para $i = 1, 2, \dots, N_2$ e $j = 1, 2, \dots, N_1$.

O algoritmo abaixo descreve a fase de treinamento da rede RBF, para a aprendizagem por lote:

1 Primeira fase

1.1 Ajustar os centros utilizando o algoritmo *k-means*, conforme descrito na Seção 3.6.1;

1.2 Ajustar as larguras utilizando a Eq. (96);

2 Segunda fase

- 2.1 Inicializar os parâmetros livres dos neurônios da camada de saída;
- 2.2 Informar a taxa de aprendizado α ;
- 2.3 Especificar a precisão requerida e o número máximo de iterações $\lambda_{\text{máx}}$;
- 2.4 Inicializar o contador de iterações $\lambda = 0$;
- 2.5 Apresentar o vetor $x(p)$ pertencente ao conjunto de treinamento;
- 2.6 Calcular a saída da rede utilizando as Eqs. (88) e (93);
- 2.7 Se todos os vetores do conjunto de treinamento foram apresentados, ir para o passo 2.9. Caso contrário, voltar para o passo 2.5;
- 2.8 Atualizar os parâmetros livres utilizando as Eqs. (102) e (103);
- 2.9 Determinar o erro médio quadrático com a Eq. (52);
- 2.10 Se o erro médio quadrático for menor que a precisão requerida ou o número de iterações for maior que o número máximo de iterações, então armazenar os parâmetros livres da rede. Caso contrário, incrementar o contador λ ($\lambda = \lambda + 1$) e voltar ao passo 2.5;

Uma vez finalizadas as duas fases do treinamento, a rede é capaz de produzir saídas com um erro aceitável, para novos estímulos. A seguir, apresenta-se o algoritmo da fase de execução:

- 1 Carregar os parâmetros livres ajustados na fase de treinamento;
- 2 Fornecer um novo padrão de entrada;
- 3 Calcular a saída dos neurônios da camada intermediária com a Eq. (88);
- 4 Calcular a saída dos neurônios da camada de saída utilizando a Eq. (93).

4 MODELAGEM DE HRIRS FILTRADAS COM RNAS

Conforme descrito na Seção 2.5, para a geração das BRIRs é necessário calcular as HRIRs filtradas de acordo com a Eq. (35) que, lembremos, é equivalente a fazer o produto de convolução de cada raio que atinge o receptor com as HRIRs da direção de chegada do raio. Isso é computacionalmente muito custoso, já que a quantidade de vezes que esse cálculo deve ser realizado é da ordem de 10^5 , que corresponde à quantidade média de raios sonoros que chegam ao receptor para uma fonte emitindo 10^6 raios (TENENBAUM et al., 2006a). Desse modo, o propósito da presente tese é desenvolver uma metodologia que permita substituir esse cálculo tradicional, utilizando as redes neurais artificiais. O método desenvolvido utilizou as redes RBF, o qual apresenta um desempenho computacional superior ao método tradicional e com uma precisão satisfatória. Além disso, trabalhos anteriores relacionados à modelagem de HRIRs, alguns apresentados na revisão bibliográfica deste trabalho, mostram que a rede RBF apresenta resultados mais acurados do que a rede MLP.

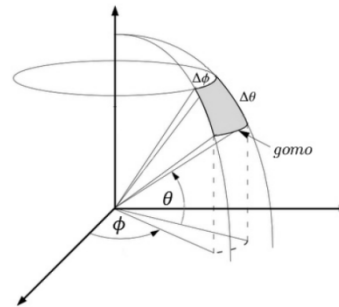
Neste capítulo apresenta-se a modelagem das redes RBF, descrita no Capítulo 3, sendo o método tradicional utilizado para validar o modelo proposto. Faz-se, também, um breve resumo da metodologia formulada por Naranjo (2014), para mostrar as diferenças das redes empregadas neste trabalho. Deve-se salientar que as outras metodologias citadas na Introdução não serão discutidos neste capítulo, pelo fato de modelarem apenas as HRIRs/HTRFs, ou seja, não fazem a modificação da HRIR para um raio sonoro com um espectro qualquer.

4.1 Modelagem utilizando as redes MLP

A modelagem das HRIRs filtradas baseada nas redes MLP formulada por Naranjo (2014) utiliza uma estrutura com múltiplas redes, sendo que cada rede neural é responsável por uma região do espaço, denominada *gomo*, como ilustra a Figura 43. A entrada de cada rede é composta por 11 elementos. Dois deles representam a direção de incidência das frentes de onda, dada pelos ângulos de azimute e elevação. Os outros nove representam o espectro de energia por banda de oitava. A saída possui 128 neurônios, onde cada um representa uma amostra temporal das HRIRs filtradas. A topologia que apresentou a melhor relação custo-benefício foi a rede com duas camadas intermediárias, sendo a primeira com sete neurônios e a segunda com três neurônios.

Como visto no Capítulo 3, as redes MLP utilizam o algoritmo supervisionado para ajustar os parâmetros livres. Conseqüentemente, é necessário definir um conjunto de pares de amostras dos sinais de entrada e das suas respectivas saídas. Para a entrada,

Figura 43 - Área de operação da rede (gomo)



Fonte: NARANJO, 2014.

os espectros dos raios foram gerados valores aleatórios no intervalo $(0, 1)$, enquanto que, para a geração das saídas, utilizou-se o método tradicional de geração das HRIRs filtradas. Então, é necessário um banco de dados de HRIRs previamente medidas. O banco de dados utilizado por Naranjo foi composto de 720 direções (GARDNER; MARTIN, 1995). Pode-se observar na Tabela 1 que esse banco não é suficientemente discretizado. Como é muito improvável que a frente de onda atinja o receptor na direção onde existe uma função medida, é necessário então, nesse caso, um procedimento de interpolação para evitar efeitos audíveis não desejados, como a descontinuidade do som gerado (*clicks*) (WERSENYI, 2009). Decorre daí a necessidade da rede operar em pequenos gomos, para cobrir todo o espaço auditivo. Portanto, essas redes não só fazem a modificação espectral, como substituem o processo de interpolação das HRIRs. Para treinar a rede utilizou-se o algoritmo Levenberg-Marquardt, descrito na Seção 3.5.2.

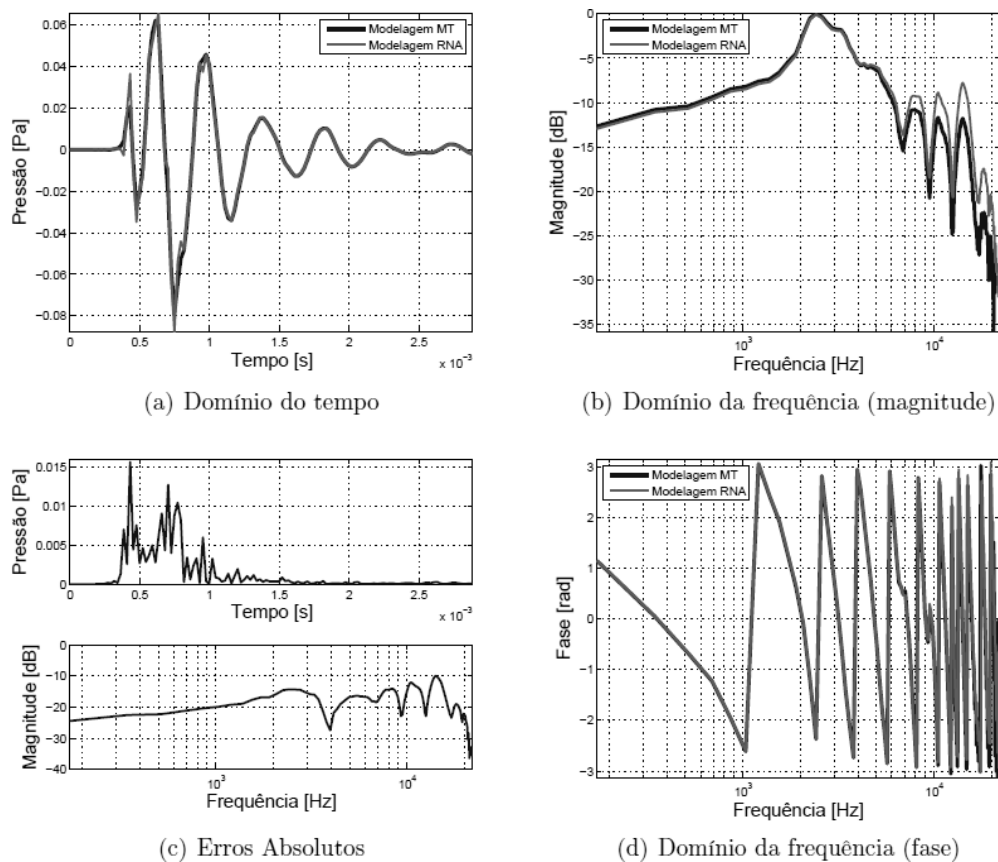
Tabela 1 - Número de medições e incrementos de azimute para cada elevação.

Elevação	Número de medições	Incremento de azimute
-40	56	6,43
-30	60	6
-20	72	5
-10	72	5
0	72	5
0	72	5
10	72	5
20	72	5
30	60	6
40	56	6,43
50	45	8
60	36	10
70	24	15
80	12	30
90	1	x

Fonte: GARDNER; MARTIN, 1995.

Com esta metodologia, uma vez que a rede esteja treinada, o custo computacional foi significativamente reduzido (ganho de 62% em relação ao método tradicional), porém os resultados ainda não são satisfatórios para as componentes de alta frequência, como mostra a Figura 44. Os resultados podem ser vistos em Naranjo (2014). Assim, dando continuidade a esse estudo, nesta tese, desenvolveram-se novas redes que permitam não só sanar esse problema, como também modelar com precisão as HRIRs filtradas para qualquer região do espaço auditivo de recepção, dado que essas redes foram treinadas a partir de um banco de dados de HRIR suficientemente discretizado.

Figura 44 - Resultados gerados pelas abordagens RNA-MLP e MT



Nota: Correspondentes ao ouvido direito, obtidos para um raio com direção dada pelos parâmetros $\theta = -15,8^\circ$ e $\phi = 352,9^\circ$. Em (c), indicam-se os erros absolutos.

Fonte: NARANJO et al., 2014.

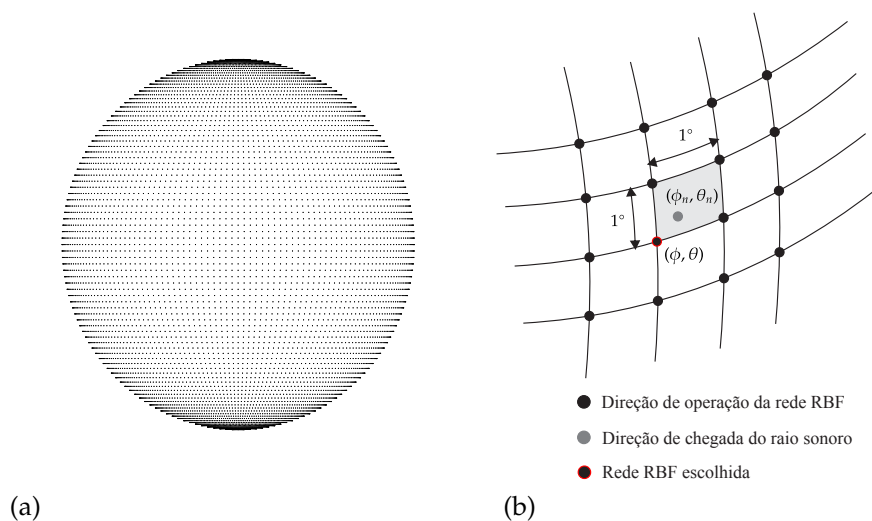
4.2 Modelagem utilizando as redes RBF

Em contraste à modelagem apresentada na Seção 4.1, as novas redes não operam em gomos, mas em cada direção definida pelos ângulos de azimute e elevação (Figura 45a), sendo estes, números inteiros, pertencentes ao intervalo $[0^\circ; 359^\circ]$ e $[-90^\circ; 90^\circ]$, respectivamente, isto é,

$$\mathcal{D} = \{(\phi, \theta) \in \mathbb{Z}^2; 0^\circ \leq \phi \leq 359^\circ \wedge -90^\circ \leq \theta \leq 90^\circ\}. \quad (104)$$

Logo, tem-se um total de 64442 redes para mapear todo o espaço auditivo de recepção. Caso uma frente de onda atinja o receptor em uma direção cujos ângulos (ϕ_n, θ_n) não correspondam aos valores requeridos, efetua-se, então, um arredondamento destes, ou seja, $(\lfloor \phi_n \rfloor, \lfloor \theta_n \rfloor) \in \mathcal{D}$, como ilustra a Figura 45b. Essa aproximação é aceitável; de fato, sendo o mínimo ângulo audível (MAA - *Minimum Audible Angle*) o paradigma utilizado para avaliar a capacidade humana de localizar fontes sonoras quando ocorre uma variação na posição do estímulo sonoro (MILLS, 1958), resultados experimentais demonstram que, em média, a acuidade humana de localização é maior na direção frontal, com o MAA atingindo um mínimo de 1° . Em outras regiões, tal acuidade diminui e, conseqüentemente, tem-se um MAA maior (BLAUERT, 1997). Por estes motivos não é necessário o processo de interpolação e nem que as redes operem em gomos.

Figura 45 - Direções do espaço auditivo onde existe uma rede RBF e a escolha desta para uma determinada frente de onda



Legenda: (a) - Espaço auditivo de recepção; (b) - Detalhe da escolha da rede RBF.

Fonte: O autor, 2017.

4.2.1 Topologia da rede RBF

O número de nós na camada de entrada e o número de neurônios na camada de saída da rede RBF foram definidos da seguinte maneira:

- **Camada de entrada:** Sabendo que a filtragem de uma dada HRIR depende do espectro do raio sonoro, define-se, então, a quantidade de nós igual ao número de bandas de oitava do raio, ou seja, $N_0 = 9$.
- **Camada de saída:** Como o espectro de energia dos raios é dado em bandas de oitava, então essas bandas são divididas em bandas centrais, em que cada frequência central é o dobro da anterior, sendo a largura de cada banda Δf_b , de frequência central f_c , dada por $\Delta f_b = f_s - f_i$, onde $f_i = f_c / \sqrt{2}$ (frequência de corte inferior da banda) e $f_s = \sqrt{2}f_c$ (frequência de corte superior da banda) (RANDALL, 1977). Vale ressaltar, que as nove bandas são necessárias para englobar toda a *faixa de áudio* de interesse (BISTAFA, 2011). Embora a informação relevante esteja nas primeiras 128 amostras (TORRES et al., 2004; NARANJO, 2014) da HRIR (pois a partir dessas amostras os valores tendem a zero), essa quantidade não é suficiente para o cálculo da HRIR filtrada. De fato, sendo a frequência de amostragem de 44,1 kHz e o número de amostras da HRIR de 128, não existe nenhuma amostra da HRTF pertencente às bandas de 63 e 125 Hz, como mostra a Tabela 2, e, conseqüentemente, não haverá um processo de filtragem da HRTF para essas frequências. Logo, 128 amostras é uma baixa resolução no domínio da frequência. O mesmo acontece para 256 amostras, só que apenas para a banda de 63 Hz. Diante disso, para uma correta filtragem em todas as bandas é necessário pelo menos 512 amostras, como pode-se ver na Tabela 2. Repare que o número de amostras apresentada nessa tabela é uma potência de dois, pois o método tradicional de geração das HRIRs filtradas, Eq (35), utiliza a FFT. Em virtude do que foi exposto, analisa-se agora o efeito da quantidade de amostras no cálculo da HRIR filtrada, mostrando alguns exemplos que comparam as HRIRs filtradas calculadas utilizando 128, 256 e 512 amostras. Os exemplos apresentados correspondem aos sinais registrados pelo ouvido esquerdo (não é necessária a análise do ouvido direito, já que a cabeça é considerada como simétrica) para as seguintes direções de chegada das frentes de onda: $(\phi = 2^\circ, \theta = 12^\circ)$ (frontal), $(\phi = 87^\circ, \theta = 79^\circ)$ (lateral esquerda), $(\phi = 191^\circ, \theta = 67^\circ)$ (traseira) e $(\phi = 274^\circ, \theta = -3^\circ)$ (lateral direita). Nas Figuras 46-49, à esquerda, são apresentadas HRIRs filtradas usando 128, 256 e 512 amostras e, à direita, são apresentadas as comparações duas a

Tabela 2 - Número de amostras da HRTF em cada banda de oitava, para a HRIR com 128, 256 e 512 amostras.

f_i (Hz)	f_c (Hz)	f_s (Hz)	Amostras		
			128	256	512
44	63	88	0	0	1
88	125	177	0	1	1
177	250	355	1	1	2
355	500	710	1	2	4
710	1000	1420	2	4	8
1420	2000	2840	4	8	16
2840	4000	5680	8	16	33
5680	8000	11360	16	33	66
11360	16000	22720	32	63	125

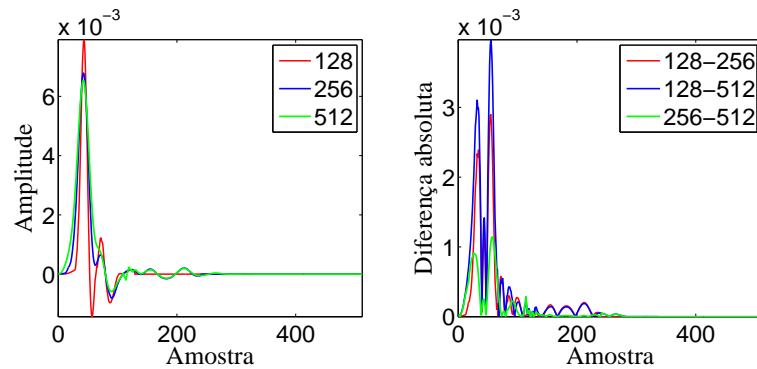
Fonte: O autor, 2017.

duas. Como pode-se observar, para todas as direções, 128 amostras é uma quantidade insuficiente para representar a HRIR filtrada, já que há uma filtragem imprecisa e, além disso, tem-se uma perda de informação na cauda devido ao truncamento. O mesmo ocorre com 256 amostras. Por outro lado, o uso de 512 amostras mostra-se suficiente. Note que, nesse último caso, depois de calculada a HRIR filtrada não é necessário levar em conta todas amostras, já que após uma determinada amostra todos os valores são praticamente nulos. Após o cálculo das HRIRs filtradas para diversas direções, verificou-se que os dados relevantes encontram-se, aproximadamente, até a 300^a amostra. Assim, os padrões de saída foram determinados da seguinte maneira: primeiramente calculam-se as HRIRs filtradas utilizando a Eq (35) e considerando as HRTFs com 512 amostras e, em seguida, essas HRIRs filtradas são truncadas na 300^a amostra. Daí, definiu-se, então, a quantidade de neurônios da camada de saída $N_2 = 300$. Uma última observação é que essas HRIRs filtradas que formam os padrões de saída contêm os atrasos de medição e os atrasos relativos às diferenças interauriculares.

Já o número de neurônios da camada intermediária N_1 é discutida posteriormente na, Seção 4.3, onde se avalia o custo computacional das redes projetadas, já que essa quantidade é definida a partir da melhor relação custo-benefício, ou seja, seleciona-se a rede com a menor quantidade de neurônios nessa camada com erros gerados considerados aceitáveis e custo computacional baixo em relação ao método tradicional.

Antes de finalizar esta seção, cumpre destacar que a direção de chegada dos raios não corresponde a um dado de entrada, pois esta serve apenas para selecionar

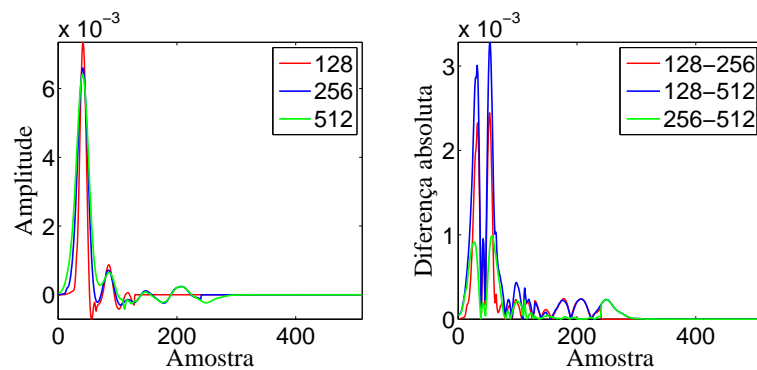
Figura 46 - Comparação das HRIRs filtradas para ($\phi = 2^\circ$ e $\theta = 12^\circ$)



Legenda: (à esquerda) - usando 128, 256 e 512 amostras; (à direita) - e a diferença absoluta entre elas, duas a duas.

Fonte: O autor, 2017.

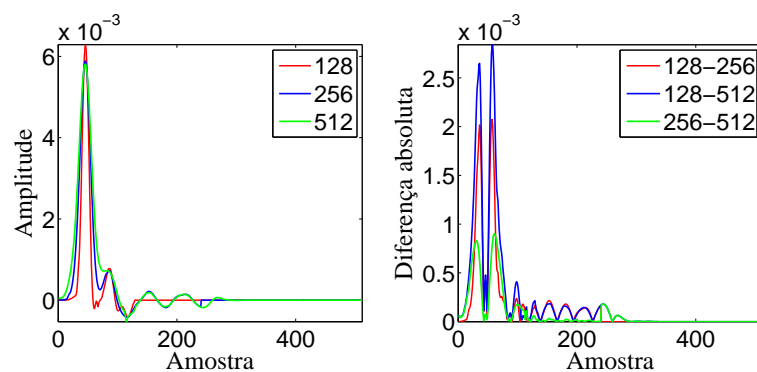
Figura 47 - Comparação das HRIRs filtradas para ($\phi = 87^\circ$ e $\theta = 79^\circ$)



Legenda: (à esquerda) - usando 128, 256 e 512 amostras; (à direita) - e a diferença absoluta entre elas, duas a duas.

Fonte: O autor, 2017.

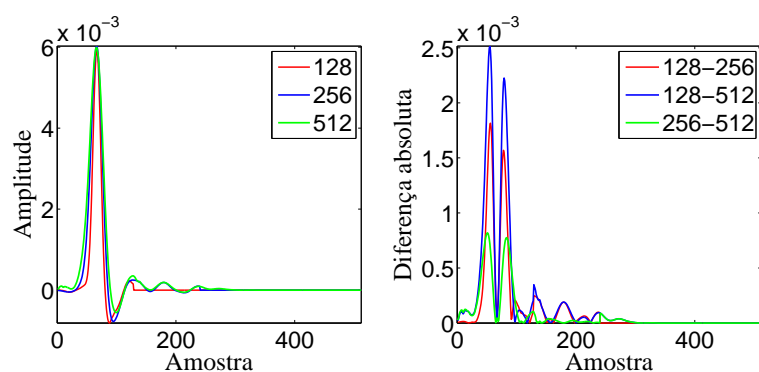
Figura 48 - Comparação das HRIRs filtradas para ($\phi = 191^\circ$ e $\theta = 67^\circ$)



Legenda: (à esquerda) - usando 128, 256 e 512 amostras; (à direita) - e a diferença absoluta entre elas, duas a duas.

Fonte: O autor, 2017.

Figura 49 - Comparação das HRIRs filtradas para ($\phi = 274^\circ$ e $\theta = -3^\circ$)



Legenda: (à esquerda) - usando 128, 256 e 512 amostras; (à direita) - e a diferença absoluta entre elas, duas a duas.

Fonte: O autor, 2017.

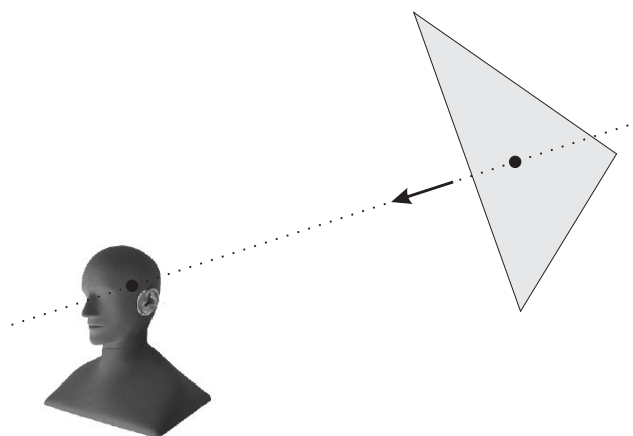
a rede. Uma outra observação importante é que como foi considerado que a cabeça do receptor biauricular é simétrica, foram projetadas as redes apenas para o ouvido esquerdo. Para computar as HRIRs filtradas do ouvido direito para uma determinada direção (ϕ, θ), seleciona-se a rede do ouvido esquerdo da direção $(360 - \phi, \theta)$.

4.2.2 Ajuste das reflexões difusas para a aplicação do método das redes RBF

Observe que as reflexões difusas calculadas pelo método de transições de energia (TENENBAUM et al., 2006a) não podem ser aplicadas às redes RBF projetadas, já que não trabalha com raios acústicos, mas com troca de energia entre as fontes, as superfícies e os receptores de acordo com o ângulo sólido (ver Seção 2.3). Sendo assim, alguns ajustes são necessários para tal aplicação. Para incluir as reflexões difusas no cálculo, deve-se tratar a troca de energia entre a superfície e o receptor por meio de raios, ou seja, além da informação do espectro de energia, o tempo e a direção de chegada também são requeridos.

Para incluir o tempo de chegada, basta armazenar o tempo em que o receptor recebeu tal energia da fonte ou de um determinado elemento triangular. Agora, para incluir a direção de chegada, armazenam-se as direções dos vetores diretores da reta que passa pelo centro geométrico de cada elemento triangular, visível ao receptor, e o centro da cabeça, como mostra a Figura 50. Daí, têm-se os dados necessários para incluir, além das reflexões especulares, as reflexões difusas no cálculo das BRIRs via redes RBF.

Figura 50 - Direção de chegada do raio acústico oriundo das reflexões difusas



Fonte: O autor, 2017.

4.2.3 Padrões de treinamento

Como o ajuste dos pesos é feito utilizando um algoritmo supervisionado, é necessário, então, definir um conjunto de pares de entrada (espectro de energia dos raios) e saída (HRIR filtrada). Note-se que, uma vez definidos os espectros dos raios que farão parte do treinamento, calculam-se os alvos utilizando a Eq. (35). Sendo assim, devem-se escolher os espectros de modo que a rede consiga uma boa generalização, isto é, bom desempenho com padrões *não* vistos no treinamento. Como a energia em cada banda do raio que chega ao receptor pode ter diferentes níveis, pois este depende das características da fonte, da sala, das condições ambientais, do número de reflexões sofridas etc., o processo de seleção dos espectros que farão parte do treinamento torna-se difícil. Neste trabalho, não foi utilizada a ideia adotada em Naranjo (2014), ou seja, de gerar os espectros de maneira aleatória, pois o conjunto gerado pode não conter as características desejadas para um treinamento eficaz. Para contornar esse problema e melhorar a capacidade de generalização, foram adotadas as estratégias a seguir.

- O conjunto dos padrões de treinamento é formado por espectros de energia dos raios gerados de simulações de salas com diferentes características. Para uma maior variabilidade de espectros dos raios sonoros, utilizou-se diversos tipos de materiais nas superfícies de contorno, salas com diferentes volumes e diferentes posições para as fontes e os receptores. Sabendo que a quantidade de raios é muito grande em cada simulação, uma seleção foi feita via algoritmo *K-means*, descrito na Seção 3.6.1. Ao final da execução desse algoritmo o conjunto de raios é particionado em uma certa quantidade de subconjuntos, onde cada subconjunto possui um

centróide, o qual determina quais são os elementos que fazem parte do subconjunto. Esses centróides, que correspondem ao espectro de energia do raio, serão os padrões escolhidos em cada simulação. Assim, depois de todas as simulações terem sido realizadas, têm-se os padrões de entrada. Cabe observar que a ideia de utilizarmos esse algoritmo, é que os raios pertencentes a um determinado *cluster* devem ser mais “similares” entre si do que em relação a raios pertencentes a outros *clusters*. Daí, pode-se escolher apenas o centróide de cada *cluster* para representá-los. Diante disso, à medida que se aumenta o número de simulações, tem-se a maximização do conteúdo de informações, e, portanto, maior será a capacidade de generalização da rede (LECUN, 1993).

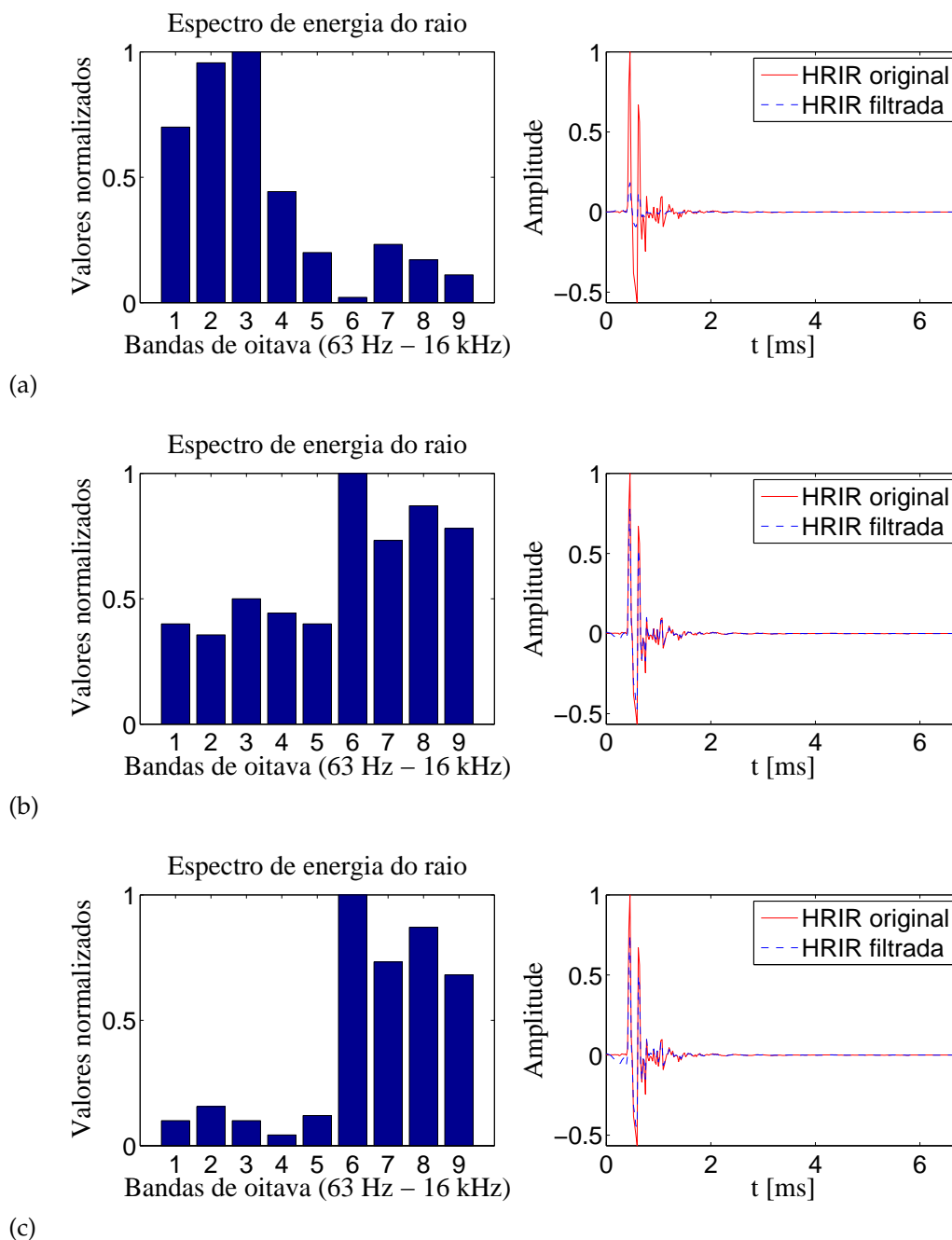
- Como visto na Seção 4.2.1, a quantidade de amostras da HRTF é maior nas bandas de frequências mais altas. Portanto, como o espectro de energia dos raios acústicos, que são emitidos da fonte, sofre modificações até atingir o receptor devido à viscosidade do ar e à absorção das superfícies e sendo, em geral, a absorção maior nas altas frequências, os espectros dos raios terão maior atenuação nessas frequências, e, conseqüentemente, maior será a modificação da HRTF. Como exemplo desses efeitos, considere a Figura 51, onde são mostradas três comparações de uma mesma HRIR com duas HRIRs filtradas, sendo a primeira obtida a partir de um raio com maior atenuação nas altas frequências e as outras duas nas baixas frequências. A diferença da Figura 51c para a Figura 51b está apenas na maior atenuação nas baixas frequências. Observe que, mesmo com uma maior atenuação, a HRIR filtrada apresentou poucas modificações. Em virtude do que foi mencionado, padrões de treinamento com tais características semelhantes à Figura 51a são primordiais para que a rede alcance uma boa generalização.
- Os padrões de entrada selecionados são, então, normalizados de modo que os seus valores fiquem no intervalo [0; 1]. Para tal fim, cada componente do conjunto dos vetores $E(p)$ é normalizado através da transformação

$$E_i^N(p) = \frac{E_i(p)}{E_{max}(p)}, \quad \text{para } i = 1, 2, \dots, 9, \quad (105)$$

onde o sobrescrito N refere-se a normalizado e $E_{max}(p) = \max_{1 \leq i \leq 9} \{E_i(p)\}$.

- Já os padrões de saída foram normalizados de modo que os seus valores

Figura 51 - Comparação entre a HRIR original com a filtrada (à direita), sendo que uma foi obtida a partir de um raio com maior atenuação nas altas frequências e as outras nas baixas frequências



Legenda: (a) - Raio com maior atenuação nas altas frequências; (b) - Raio com maior atenuação nas baixas frequências; (c) - Raio, também, com maior atenuação nas baixas frequências, com a diferença de uma maior atenuação nessas frequências em relação ao anterior.

Fonte: O autor, 2017.

situem-se no intervalo $[-1; 1]$, utilizando a seguinte transformação

$$\bar{h}_i^N(p) = \frac{\bar{h}_i(p)}{h_{max}}, \quad \text{para } i = 1, 2, \dots, 300, \quad (106)$$

onde o sobrescrito N refere-se a normalizado e $h_{max}(\phi, \theta) = \max_{1 \leq i \leq 300} \{h_i(\phi, \theta)\}$ é o valor máximo absoluto da HRIR, $h_i(\phi, \theta)$, para a direção (ϕ, θ) .

- Uma vez gerados os padrões, esses são divididos, de forma aleatória, em dois conjuntos disjuntos: dois terços dos padrões gerados para treino e o restante para validação. O motivo para essa divisão é validar o modelo em um conjunto de dados distintos daqueles que foram utilizados para ajustar os parâmetros livres. A aplicação dessa técnica evita que ocorra o *overtraining* (treinamento excessivo) da rede, pois a rede treinada em excesso aprende vigorosamente os dados de treinamento e, portanto, apresenta péssima capacidade de generalizar tal conhecimento (HAYKIN, 1999).

4.2.4 Algoritmo modificado para o ajuste dos parâmetros livres

Para ajuste dos parâmetros livres da camada intermediária utilizou-se o algoritmo descrito na Seção 3.6.1. Já para os da camada de saída utilizou-se uma abordagem com algumas modificações do algoritmo apresentado na Seção 3.6.2.

Tais modificações foram feitas por causa de possíveis valores que as amostras das HRIRs filtradas podem assumir. Como foi visto na Figura 51, para uma mesma direção, as HRIRs filtradas podem sofrer severas modificações na sua forma dependendo do espectro do raio. Sendo assim, a HRIR filtrada pode apresentar baixos ou altos níveis de energia. Agora, como o algoritmo de ajuste dos parâmetros livres da camada de saída utiliza o erro médio quadrático para avaliar o desempenho da rede, esses erros podem ser mascarados pelo baixo valor absoluto (ver Figura 51a). O erro relativo também não é uma boa opção, já que os erros relativos podem ser altos, mas isso não significando que haja um erro importante entre o medido e o simulado. Portanto, neste trabalho, para contornar essas limitações, utiliza-se o *coeficiente de correlação cruzada normalizada* (NCCC - *normalized cross-correlation coefficient*) (POLLOW; BEHLER, 2009; KLEIN et al., 2014; XIE et al., 2015), o qual indica o grau de similaridade entre a saída da rede e a

saída desejada (HRIR filtrada obtida pelo método tradicional) sendo dado por

$$\text{NCCC}(p) = \frac{\sum_{j=1}^{N_2} u_j^2(p) y_j(p)}{\sqrt{\left\{ \sum_{j=1}^{N_2} [u_j^2(p)]^2 \right\} \left\{ \sum_{j=1}^{N_2} [y_j(p)]^2 \right\}}}, \quad (107)$$

onde $u_j^2(p)$ é a saída do j -ésimo neurônio da camada de saída da rede, $y_j(p)$ é o valor desejado e p é o número do padrão. Os valores do coeficiente variam entre -1 e 1 , sendo que o valor 1 indica que os sinais são completamente idênticos. Daí, a função custo para avaliar o desempenho da rede, considerando o modo de treinamento por lote, agora será dado por

$$\bar{C} = \frac{1}{P} \sum_{p=1}^P \text{NCCC}(p), \quad (108)$$

onde p é o número do padrão e P é o número de padrões.

Vamos agora deduzir as equações de ajuste dos parâmetros livres da última camada. Vale ressaltar que as variáveis e os parâmetros auxiliares que norteiam a dedução são os mesmos definidos no Capítulo 3. Observe que precisamos apenas recalcular as expressões para os δ^2 's. Partindo da Eq. (61) e utilizando a Eq. (108), tem-se

$$\delta_i^2 = \frac{\partial \bar{C}}{\partial v_i^2} = \frac{\partial}{\partial v_i^2} \left[\frac{1}{P} \sum_{p=1}^P \text{NCCC}(p) \right] = \frac{1}{P} \sum_{p=1}^P \frac{\partial}{\partial v_i^2} [\text{NCCC}(p)]. \quad (109)$$

Substituindo a Eq. (107) na Eq. (109) e aplicando a regra da cadeia, obtém-se

$$\delta_i^2 = \frac{1}{P} \sum_{p=1}^P \frac{\left[\sum_{j=1}^{N_2} u_j^2(p) \right] y_i(p) - \left[\sum_{j=1}^{N_2} u_j^2(p) y_j(p) \right] u_i^2(p)}{\sqrt{\left\{ \sum_{j=1}^{N_2} [u_j^2(p)]^2 \right\}^3 \left\{ \sum_{j=1}^{N_2} [y_j(p)]^2 \right\}}} \frac{\partial u_i^2(p)}{\partial v_i^2}. \quad (110)$$

Como a função de ativação dos neurônios da camada de saída é linear, então $u_i^2(p) = v_i^2$

e, portanto,

$$\delta_i^2 = \frac{1}{P} \sum_{p=1}^P \frac{\left[\sum_{j=1}^{N_2} u_j^2(p) \right] y_i(p) - \left[\sum_{j=1}^{N_2} u_j^2(p) y_j(p) \right] u_i^2(p)}{\sqrt{\left\{ \sum_{j=1}^{N_2} [u_j^2(p)]^2 \right\}^3 \left\{ \sum_{j=1}^{N_2} [y_j(p)]^2 \right\}}}. \quad (111)$$

Assim, substituindo a Eq. (111) nas Eqs. (97) e (98), obtém-se as equações de atualização dos parâmetros livres.

4.3 Custo computacional

Para comparar a eficiência do método tradicional (MT) com o da rede RBF, foi feita uma comparação quanto ao número de operações aritméticas que cada técnica requer. Sabendo que o tempo computacional gasto para efetuar uma operação de multiplicação/divisão é equivalente ao da soma/subtração (CHAPRA; CANALE, 2010), para os computadores modernos, o cômputo do número total de operações é simplesmente a soma do número de operações básicas, sem ponderação.

4.3.1 Custo computacional do método tradicional de geração das HRIRs filtradas

O número de operações que o método tradicional de geração das HRIRs filtradas (Eq. 35) realiza é igual à soma de duas parcelas, descritas a seguir.

- A primeira corresponde ao número de multiplicações entre o espectro do raio e a HRTF. Sendo a HRTF um complexo com ℓ amostras e o espectro por bandas de oitava do raio dado por números reais, o número de operações da primeira parcela é $\ell/2$ (multiplicações). Observe que são necessárias apenas $\ell/2$ multiplicações, em razão da simetria da HRTF, conforme discutido no Capítulo 2.
- A segunda corresponde ao número de operações para o cálculo da transformada inversa de Fourier desse produto. Como a transformada inversa rápida de Fourier (IFFT) envolve $(\ell/2) \log_2 \ell$ multiplicações complexas ($2\ell \log_2 \ell$ multiplicações reais) e $\ell \log_2 \ell$ somas complexas ($3\ell \log_2 \ell$ somas reais) (PROAKIS; MANOLAKIS, 1996; OPPENHEIM; SCHAFER, 2010), o total de operações dessa parcela é $5\ell \log_2 \ell$.

Assim, o número total de somas \mathcal{S}_{MT} e multiplicações \mathcal{M}_{MT} que o método tradicional efetua é, respectivamente,

$$\mathcal{S}_{\text{MT}} = 3\ell \log_2 \ell; \quad (112)$$

$$\mathcal{M}_{\text{MT}} = \frac{\ell}{2} + 2\ell \log_2 \ell, \quad (113)$$

e, portanto o número total de operações \mathcal{T}_{MT} é

$$\mathcal{T}_{\text{MT}} = \frac{\ell}{2} + 5\ell \log_2 \ell. \quad (114)$$

A Tabela 3 apresenta o número de operações que o método tradicional executa para a HRTF com $\ell = 512$ amostras.

Tabela 3 - Número de operações que o método tradicional executa para a HRTF com 512 amostras.

\mathcal{S}_{MT}	\mathcal{M}_{MT}	\mathcal{T}_{MT}
13824	9472	23296

Fonte: O autor, 2017.

4.3.2 Custo computacional da fase de execução da rede RBF

O número de operações da rede RBF durante a fase de execução é dada pela soma de duas parcelas.

- A primeira, está associada à soma do número de operações realizadas para calcular a saída da função de ativação, dada pelas Eqs. (88) e (92), de cada neurônio da camada intermediária. Para obter a saída do j -ésimo neurônio da camada intermediária u_j^1 são necessárias as operações listadas na Tabela 4. A primeira linha corresponde ao cálculo do expoente da exponencial e a segunda ao valor da exponencial. O cálculo desta última deve-se ao fato de que qualquer função de ativação pode ser representada por uma série de Taylor truncada em N termos, e cujo valor pode ser obtido através da regra de Horner (CONTE; BOOR, 1980; STOER; BULIRSCH, 2000), (para mais detalhes ver Anexo A). Logo, sendo N_1 o número de neurônios nessa camada, o número de somas e produtos para a primeira parcela é $N_1 (2N_0 + N - 1)$ e $N_1 (N_0 + N + 1)$, respectivamente.
- A segunda corresponde à soma do número de operações para calcular a

Tabela 4 - Número de operações para calcular a saída do j -ésimo neurônio da camada intermediária.

Operação	Somas	Multiplicações
$-(v_j^1)^2 = -(b_j^1)^2 \sum_{n=1}^{N_0} (x_n - w_{j,n}^1)^2$	$2N_0 - 1$	$N_0 + 1$
$u_j^1 = \exp\left(\frac{1}{(v_j^1)^2}\right)$	N	$N + 1$

Fonte: O autor, 2017.

saída de cada neurônio da camada de saída, dada pela Eq. (93). Sendo N_2 o número de neurônios dessa camada, o número de somas e produtos da segunda parcela é N_1N_2 e N_1N_2 .

Além dessas operações, é necessário levar em conta o custo para normalizar as entradas e a posterior multiplicação da saída da rede pelo fator que foi utilizado para normalizar a entrada. Esses procedimentos requerem $N_0 + N_2$ multiplicações. Pode-se, então, concluir que o total de somas e produtos para gerar a saída da rede é:

$$\mathcal{S}_{\text{RBF}} = N_1 (2N_0 + N_2 + N - 1) ; \quad (115)$$

$$\mathcal{M}_{\text{RBF}} = N_1 (N_0 + N_2 + N + 1) + N_0 + N_2, \quad (116)$$

e o total de operações será

$$\mathcal{T}_{\text{RBF}} = N_1 (3N_0 + 2N_2 + 2N) + N_0 + N_2. \quad (117)$$

Considerando $N_0 = 9$, $N_2 = 300$, definidos na Seção 4.2.1, e $N = 35$ (ver Anexo A), tem-se o seguinte número de operações em função de N_1 conforme mostra a Tabela 5.

Tabela 5 - Número de operações que o método da rede RBF requer.

\mathcal{S}_{RBF}	\mathcal{M}_{RBF}	\mathcal{T}_{RBF}
$352N_1$	$345N_1 + 309$	$697N_1 + 309$

Fonte: O autor, 2017.

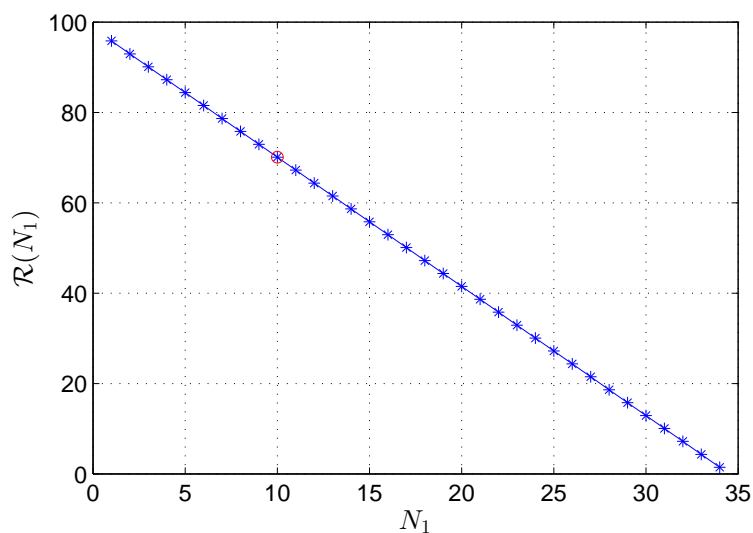
Analisa-se, agora, os possíveis valores de N_1 de tal maneira que o número de operações da rede seja menor do que o custo do método tradicional e, simultaneamente, o erro associado ao uso da rede fique dentro de uma margem considerada aceitável. Para determinar o valor ótimo de N_1 , inicialmente, compara-se o custo entre os métodos tradicional e o da rede RBF. Essa comparação é feita através do cálculo da redução

percentual, isto é,

$$\mathcal{R}(N_1) = \frac{\mathcal{T}_{\text{MT}} - \mathcal{T}_{\text{RBF}}}{\mathcal{T}_{\text{MT}}} \times 100\% = \frac{22987 - 697N_1}{23296} \times 100\%. \quad (118)$$

A Figura 52 mostra a redução percentual do número de operações usando a rede RBF em função do número de neurônios N_1 . A partir do gráfico tem-se que o valor mínimo de N_1 é 1 e o máximo 34. Com base nessa restrição foram desenvolvidas e testadas 34 redes para cada uma das 64442 direções. Verificou-se que a rede com $N_1 = 10$ foi a que apresentou a melhor relação custo-benefício, apresentando uma redução do custo computacional de aproximadamente 70%. Para $N_1 < 10$ há uma alta redução no custo computacional, mas os resultados não são suficientemente acurados. Por outro lado, para $N_1 > 10$ há um alto custo computacional, contudo os resultados são um pouco mais acurados. Embora os resultados sejam mais condizentes para essas quantidades de neurônios, os erros gerados pela rede com $N_1 = 10$ são aceitáveis, já que a avaliação de todos padrões de treinamento obtiveram o NCCC próximo de um.

Figura 52 - Redução percentual do número de operações usando a rede RBF em função do número de neurônios na camada intermediária



Fonte: O autor, 2017.

Cumprir notar que o custo computacional do método tradicional foi calculado considerando a HRTF com 512 amostras, pois essa quantidade é necessária para obter a mesma acurácia que o método das redes RBF.

Em função do que foi exposto, não se modelou as redes para operarem em gomos, visto que a dimensão do vetor de entrada aumentaria para onze, onde esses dois

elementos a mais corresponderiam à direção de chegada do raio, e, conseqüentemente, aumentaria a quantidade de padrões a serem aprendidas e o número de neurônios da camada intermediária, para obter resultados com a mesma acurácia. Portanto, maior será o custo computacional, tornando as redes desvantajosas em relação ao método tradicional.

Como uma última observação, as redes neurais foram programadas em linguagem C++ e foram treinadas em três máquinas Intel Core i7-3520M CPU @2.90 GHz com 8 GB de memória. O treinamento de cada rede levou cerca de 2 min. Como foram treinadas 64442 redes, levou-se, então, cerca de 30 dias para concluir todo o treinamento.

4.4 Resultados dos testes de algumas redes RBF

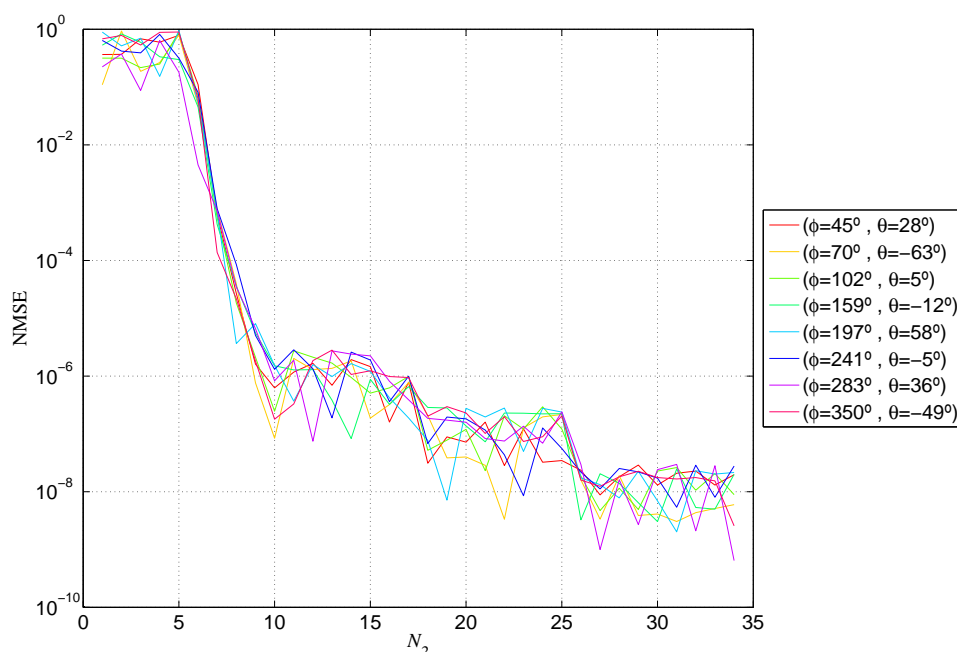
Uma vez que há uma infinidade de raios acústicos com diferentes espectros de energia e como a quantidade de redes a serem testadas é muito grande, nesta seção são apresentados resultados de oito redes, para o ouvido esquerdo, apenas para averiguar a capacidade de generalização da rede, para as seguintes direções: $(\phi = 45^\circ, \theta = 28^\circ)$, $(\phi = 70^\circ, \theta = -63^\circ)$, $(\phi = 102^\circ, \theta = 5^\circ)$, $(\phi = 159^\circ, \theta = -12^\circ)$, $(\phi = 197^\circ, \theta = 58^\circ)$, $(\phi = 241^\circ, \theta = -5^\circ)$, $(\phi = 283^\circ, \theta = 36^\circ)$ e $(\phi = 350^\circ, \theta = -49^\circ)$. Note-se que cada uma dessas direções pertencem a um determinado octante em torno do centro da cabeça.

Antes de apresentar esses resultados, convém mostrar a performance das redes RBF em função do número de neurônios da camada intermediária. A Figura 53 mostra a evolução do erro médio quadrático normalizado (NMSE - *Normalized Mean Square Error*) dado por

$$\text{NMSE} = \frac{1}{N_2 P} \sum_{p=1}^P \sum_{i=1}^{N_2} \left[\frac{y_i(p) - u_i(p)}{y_i(p)} \right]^2, \quad (119)$$

o qual foi alcançado pela rede na fase de treinamento, à medida que o número de neurônios da camada intermediária aumenta. Nesse gráfico pode ser observado que o desempenho das redes para as direções destacadas possuem um comportamento semelhante, pois ao passo que o número de neurônios da camada intermediária aumenta, o NMSE diminui. Vale ressaltar que esse mesmo comportamento é também semelhante para outras direções. Utilizando dez neurônios da camada intermediária, conforme discutido na Seção 4.3.2 foi a quantidade que apresentou a melhor relação custo-benefício, o NMSE foi cerca de 10^{-6} para todas as direções onde as redes atuam.

Figura 53 - Performance da rede RBF em função do número de neurônios na camada intermediária para diversas direções



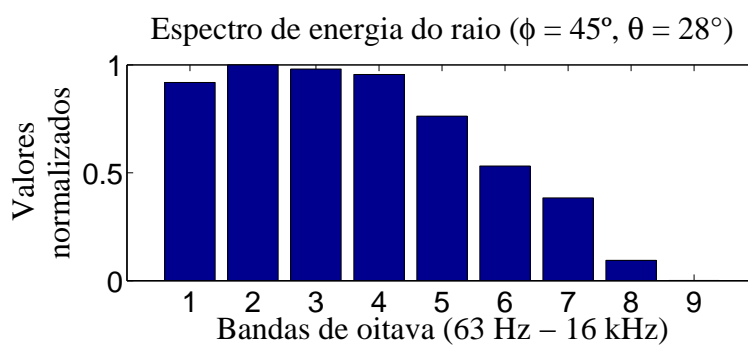
Nota: Para as direções $(\phi = 45^\circ, \theta = 28^\circ)$, $(\phi = 70^\circ, \theta = -63^\circ)$, $(\phi = 102^\circ, \theta = 5^\circ)$, $(\phi = 159^\circ, \theta = -12^\circ)$, $(\phi = 197^\circ, \theta = 58^\circ)$, $(\phi = 241^\circ, \theta = -5^\circ)$, $(\phi = 283^\circ, \theta = 36^\circ)$ e $(\phi = 350^\circ, \theta = -49^\circ)$.

Fonte: O autor, 2017.

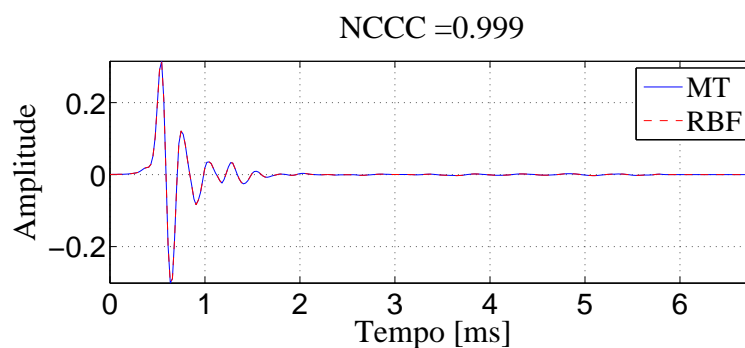
Esses resultados são confrontados com os do método tradicional e avaliados a partir do coeficiente de correlação cruzada normalizada (NCCC), dado pela Eq. (107). As Figuras 54-61, mostram a comparação das HRIRs filtradas e a magnitude e fase das correspondentes HRTFs filtradas e, além disso, os valores dos NCCCs. Como pode ser observado, para todas as direções, os valores de NCCC estão bem próximos de 1, o que mostra que há uma boa concordância nos resultados. Embora os resultados das Figuras 58, 60 e 61 apresentem um pequeno desvio na fase na região de alta frequência, esses erros não são observáveis, pois o ouvido humano é relativamente insensível à distorção de fase, principalmente na região de alta frequência (HEINZ, 1994; LATHI, 2005). Outros resultados podem ser vistos em Tenenbaum et al. (2018).

Como pode-se notar, os resultados utilizando as redes RBF são mais precisos do que aqueles apresentados na Seção 4.1. Vale ressaltar, que foi feita a tentativa de modelagem utilizando as redes MLP com as mesmas entradas e saídas da rede RBF e, além disso, adotando estratégias semelhantes às utilizadas na rede RBF. Essas redes foram abandonadas pelos seguintes motivos: necessitam mais de uma camada intermediária

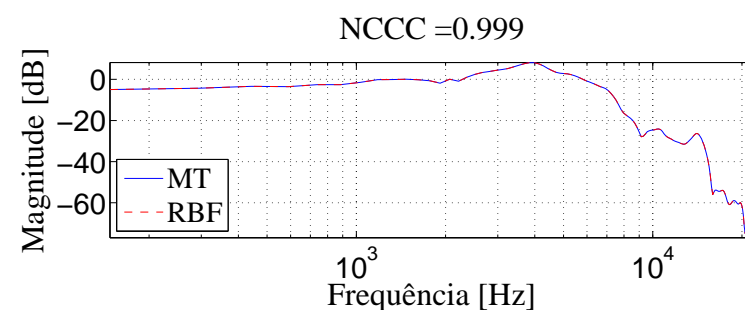
Figura 54 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 45^\circ$ e $\theta = 28^\circ$, utilizando os métodos tradicional e da rede RBF



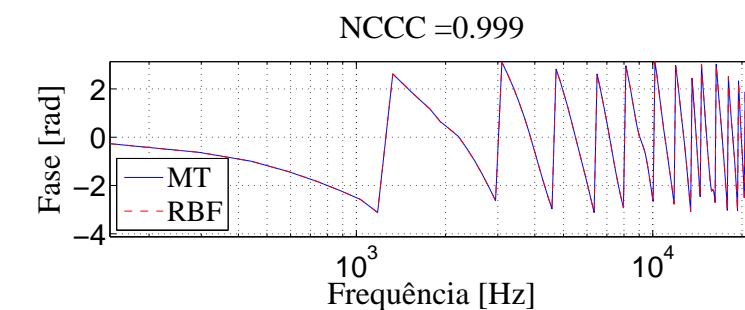
(a)



(b)



(c)

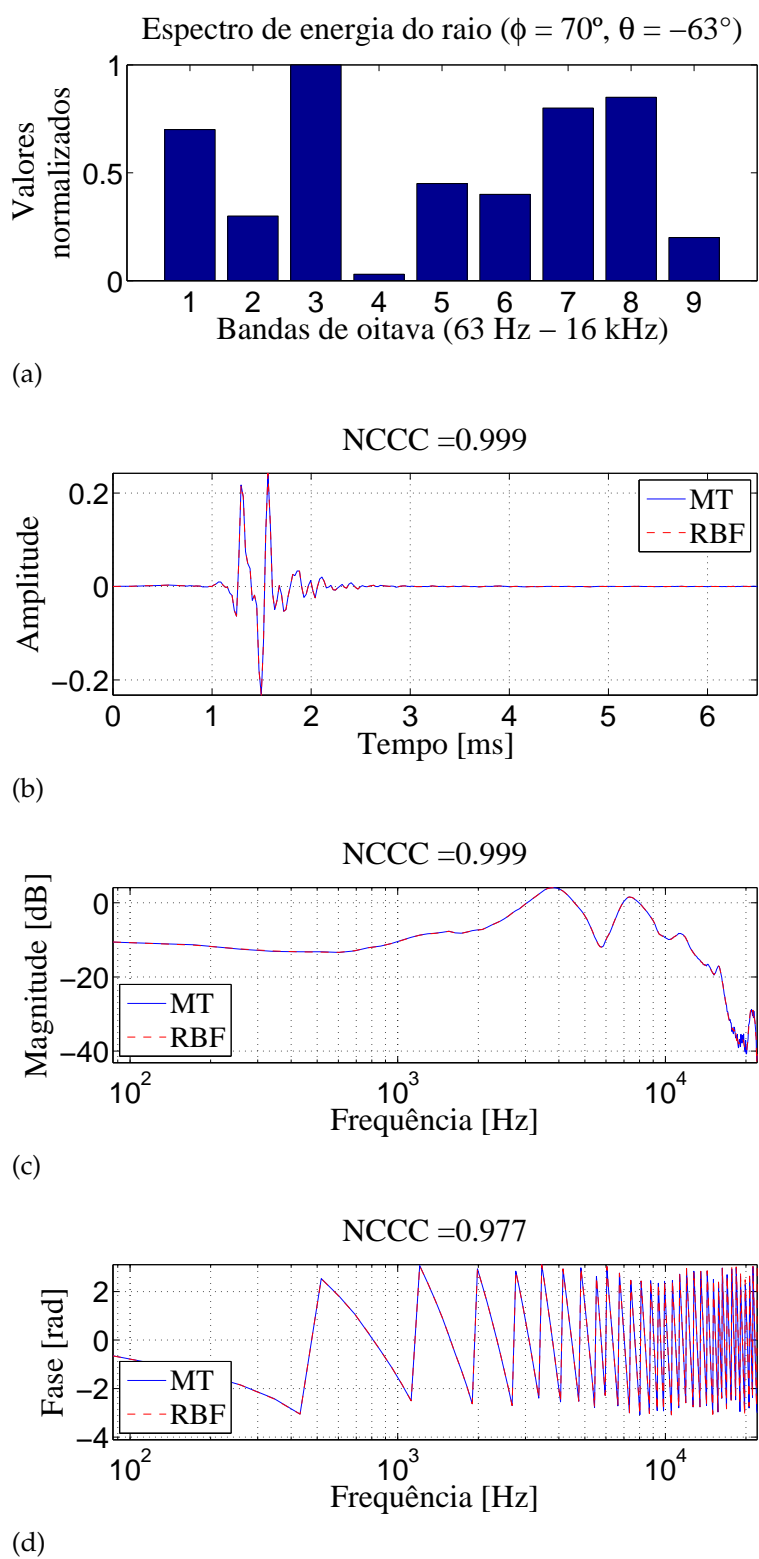


(d)

Legenda: (a) - Espectro de energia normalizado do raio; (b) - HRIR filtrada; (c) - Magnitude da HRTF filtrada; (d) - Fase da HRTF filtrada.

Fonte: O autor, 2017.

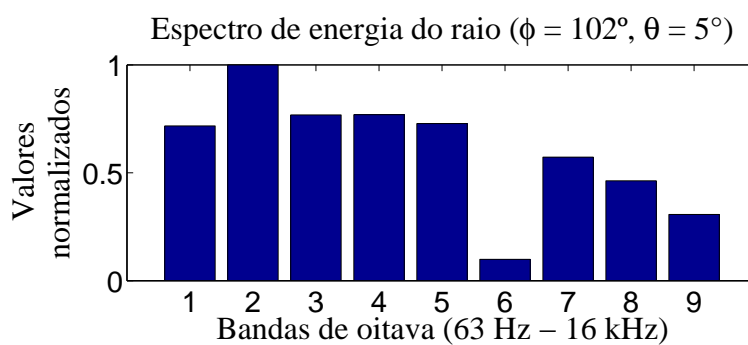
Figura 55 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 70^\circ$ e $\theta = -63^\circ$, utilizando os métodos tradicional e da rede RBF



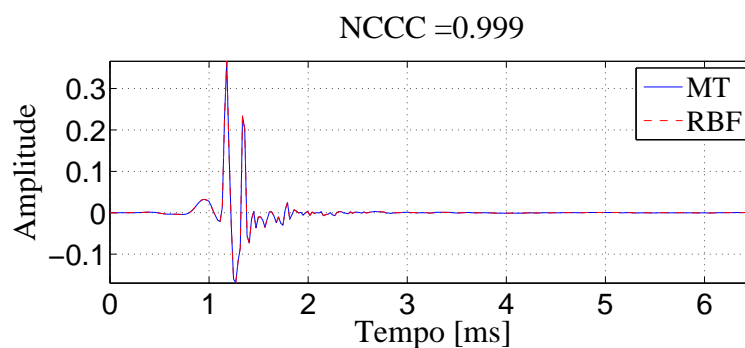
Legenda: (a) - Espectro de energia normalizado do raio; (b) - HRIR filtrada; (c) - Magnitude da HRTF filtrada; (d) - Fase da HRTF filtrada.

Fonte: O autor, 2017.

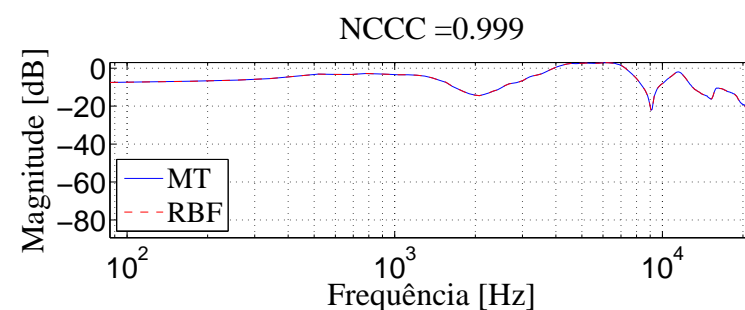
Figura 56 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 102^\circ$ e $\theta = 5^\circ$, utilizando os métodos tradicional e da rede RBF



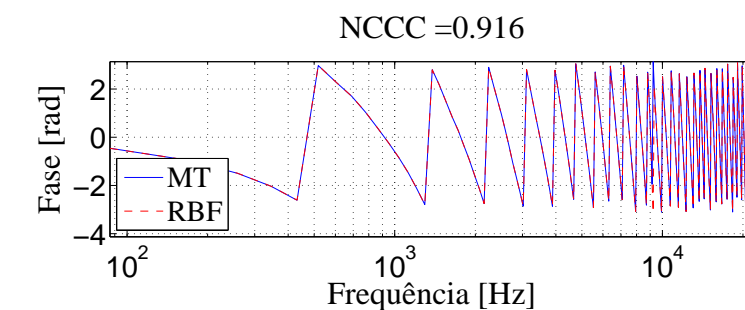
(a)



(b)



(c)

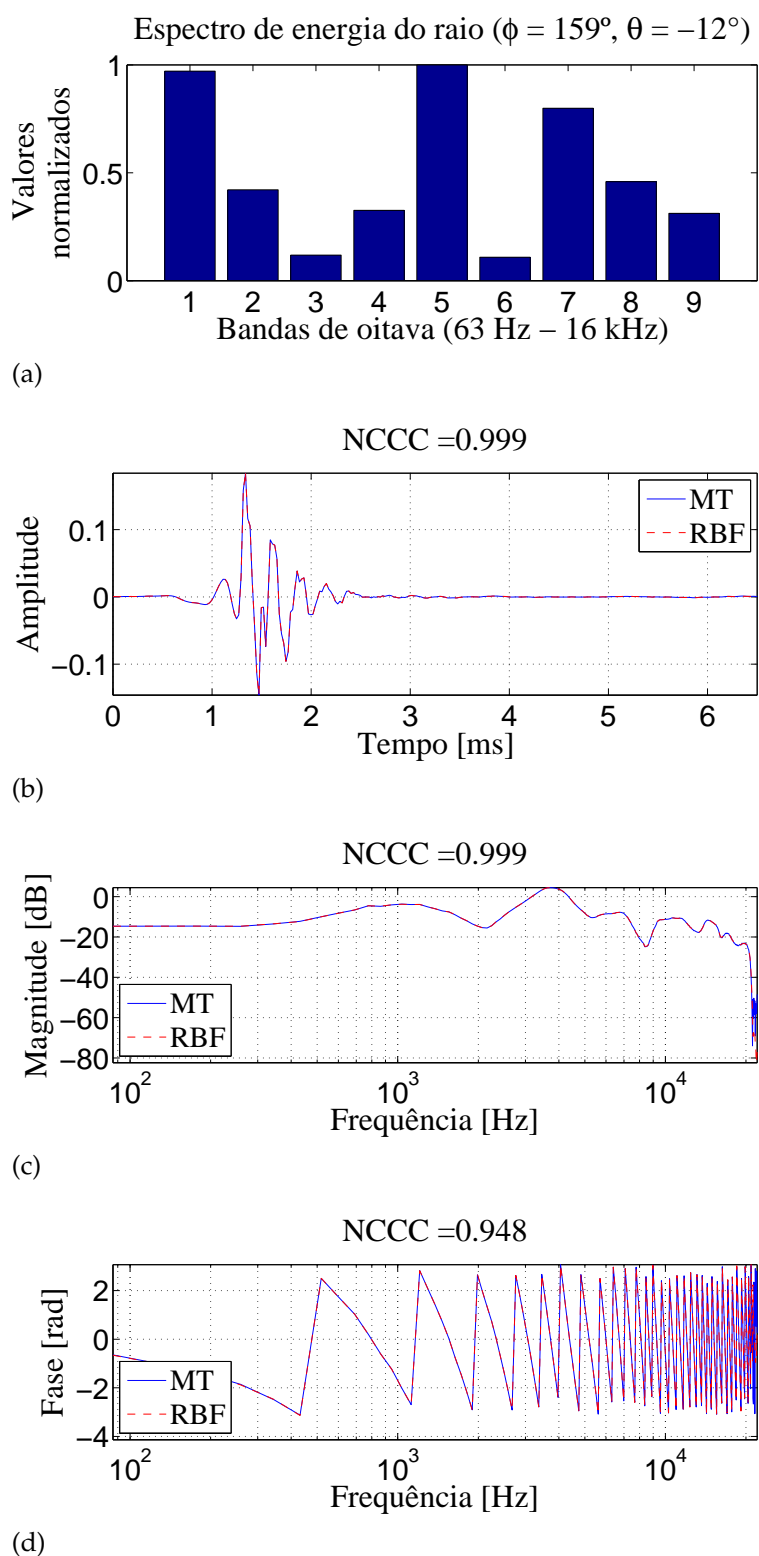


(d)

Legenda: (a) - Espectro de energia normalizado do raio; (b) - HRIR filtrada; (c) - Magnitude da HRTF filtrada; (d) - Fase da HRTF filtrada.

Fonte: O autor, 2017.

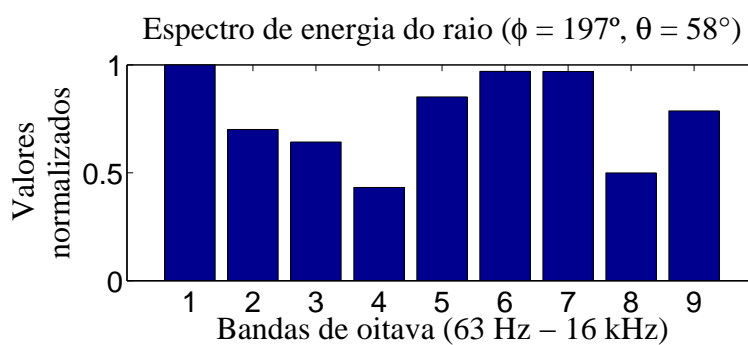
Figura 57 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 159^\circ$ e $\theta = -12^\circ$, utilizando os métodos tradicional e da rede RBF



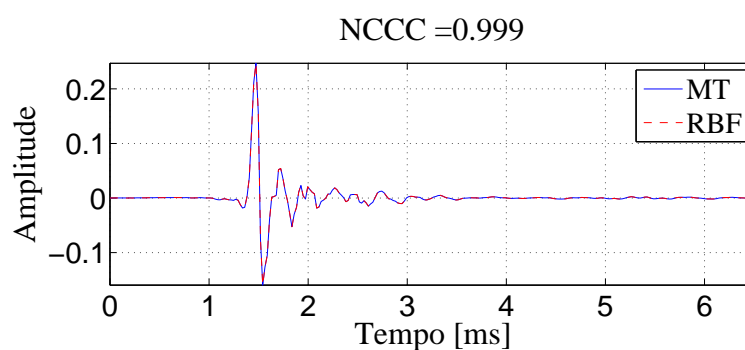
Legenda: (a) - Espectro de energia normalizado do raio; (b) - HRIR filtrada; (c) - Magnitude da HRTF filtrada; (d) - Fase da HRTF filtrada.

Fonte: O autor, 2017.

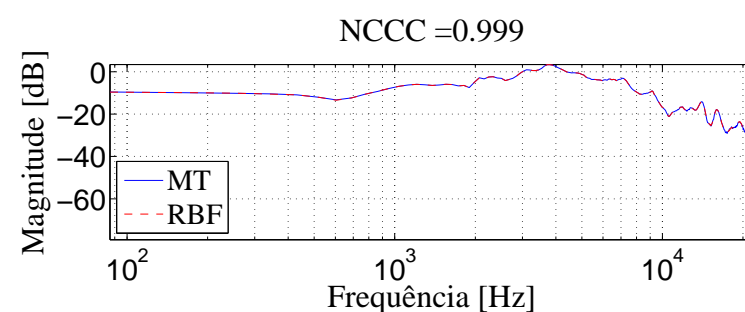
Figura 58 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 197^\circ$ e $\theta = 58^\circ$, utilizando os métodos tradicional e da rede RBF



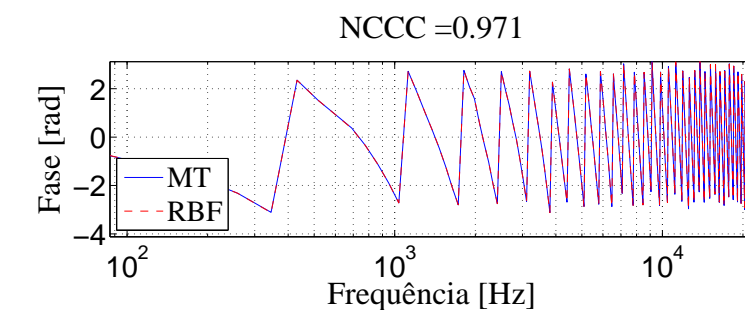
(a)



(b)



(c)

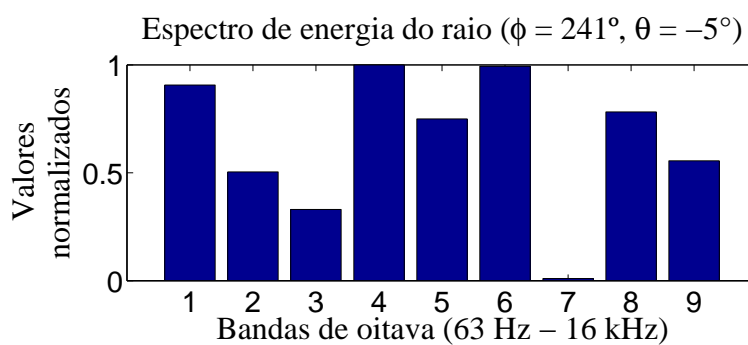


(d)

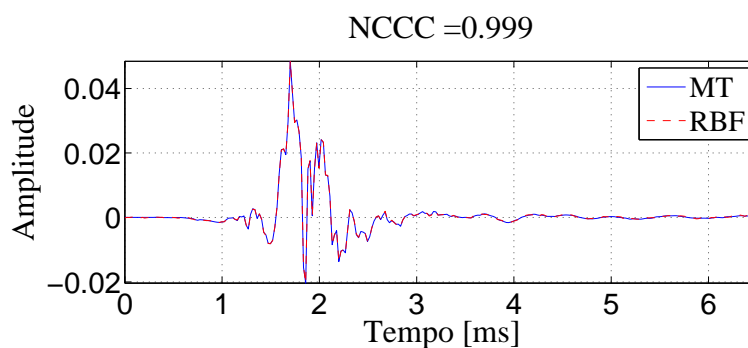
Legenda: (a) - Espectro de energia normalizado do raio; (b) - HRIR filtrada; (c) - Magnitude da HRTF filtrada; (d) - Fase da HRTF filtrada.

Fonte: O autor, 2017.

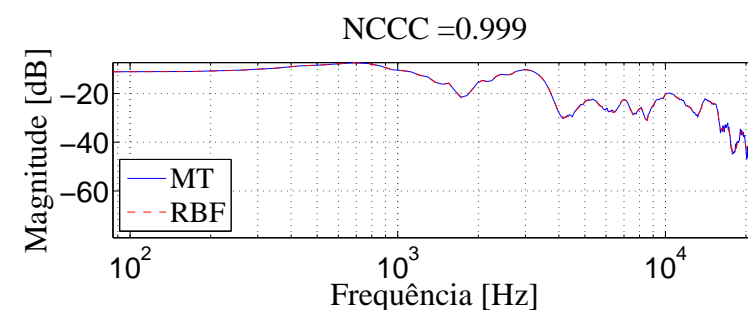
Figura 59 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 241^\circ$ e $\theta = -5^\circ$, utilizando os métodos tradicional e da rede RBF



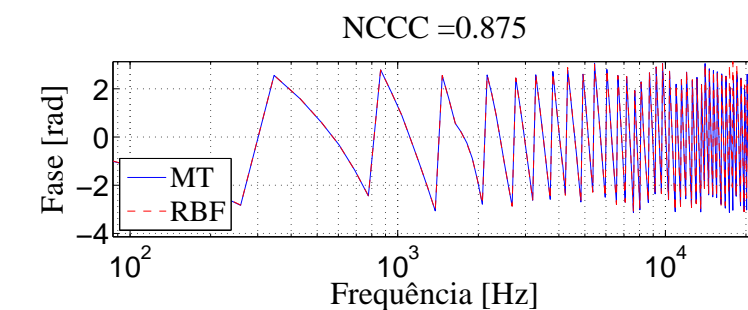
(a)



(b)



(c)

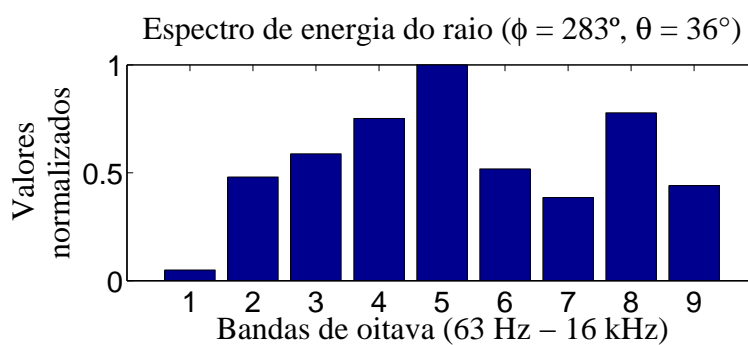


(d)

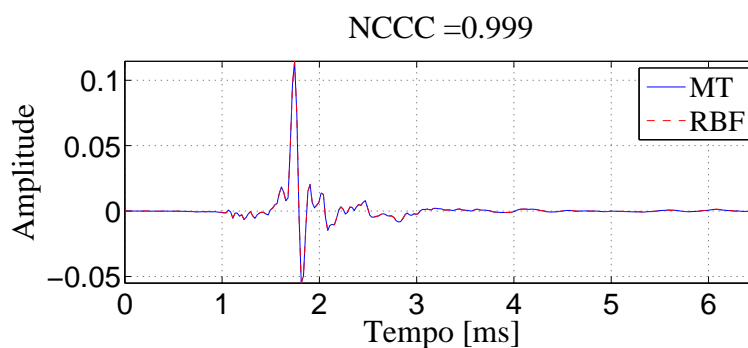
Legenda: (a) - Espectro de energia normalizado do raio; (b) - HRIR filtrada; (c) - Magnitude da HRTF filtrada; (d) - Fase da HRTF filtrada.

Fonte: O autor, 2017.

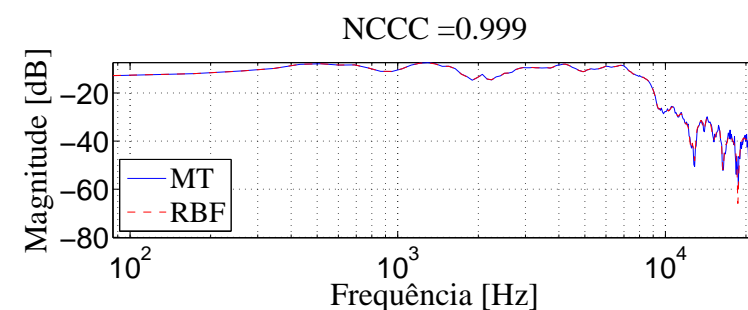
Figura 60 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 283^\circ$ e $\theta = 36^\circ$, utilizando os métodos tradicional e da rede RBF



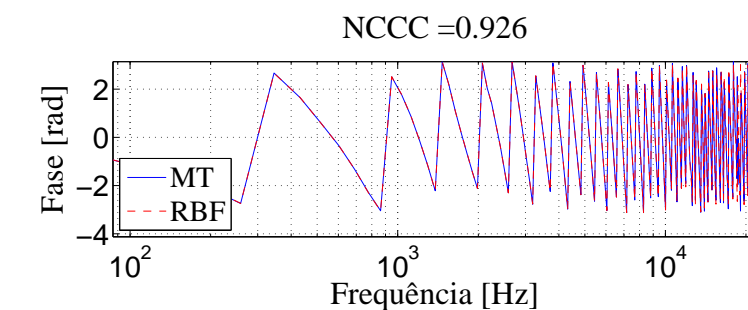
(a)



(b)



(c)

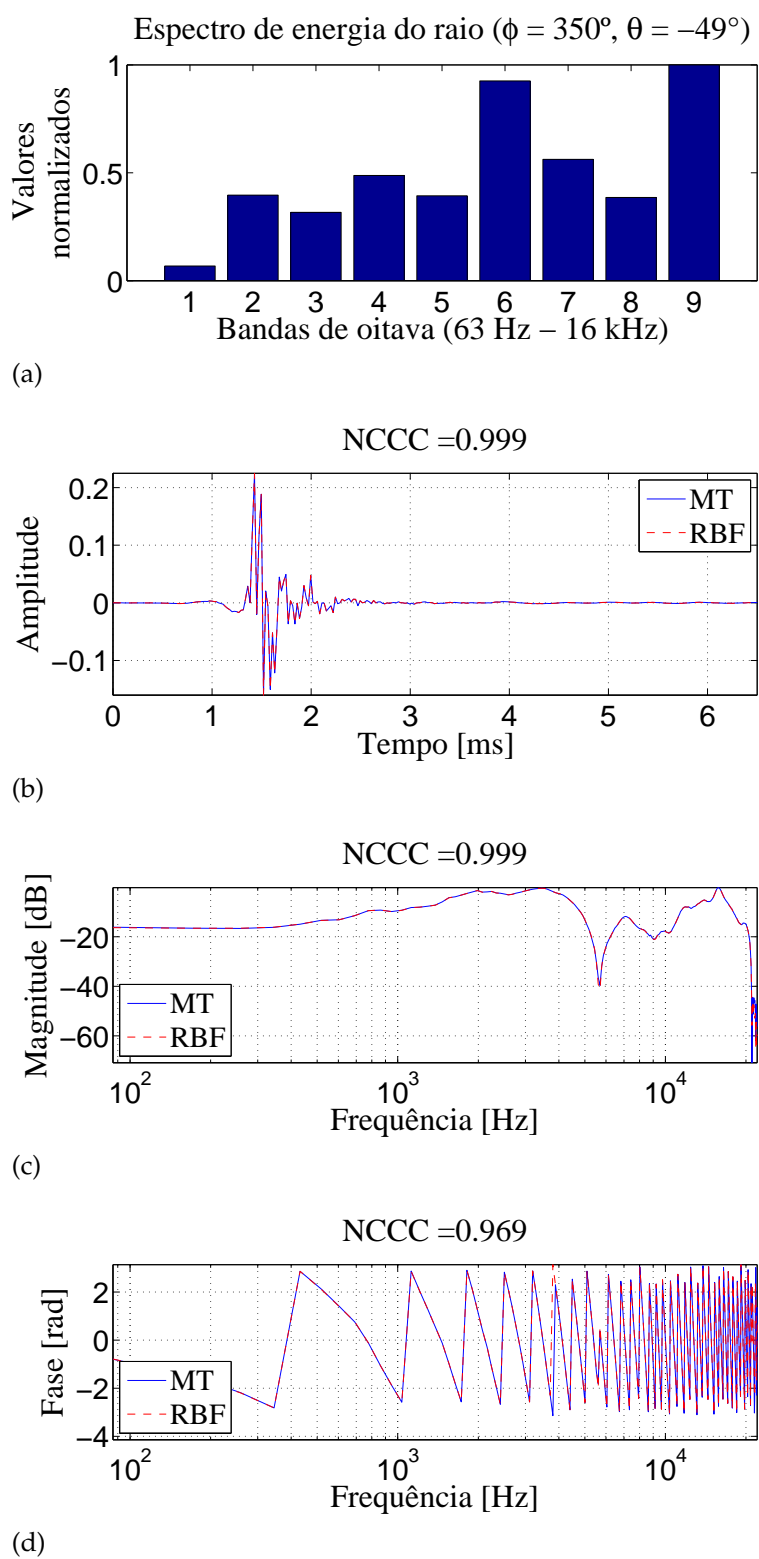


(d)

Legenda: (a) - Espectro de energia normalizado do raio; (b) - HRIR filtrada; (c) - Magnitude da HRTF filtrada; (d) - Fase da HRTF filtrada.

Fonte: O autor, 2017.

Figura 61 - Comparação das HRIRs e HRTFs filtradas do ouvido esquerdo, para $\phi = 350^\circ$ e $\theta = -49^\circ$, utilizando os métodos tradicional e da rede RBF



Legenda: (a) - Espectro de energia normalizado do raio; (b) - HRIR filtrada; (c) - Magnitude da HRTF filtrada; (d) - Fase da HRTF filtrada.

Fonte: O autor, 2017.

e muitos neurônios para essas camadas de modo a obter uma boa generalização e, conseqüentemente, torna-se desvantajoso do ponto de vista do custo computacional, em relação ao método tradicional. Além disso, o processo de treinamento é muito lento. Portanto, para o problema abordado nesta tese, as redes RBF são bem mais vantajosas do que as redes MLP.

4.5 Construção das BRIRs utilizando as redes RBF

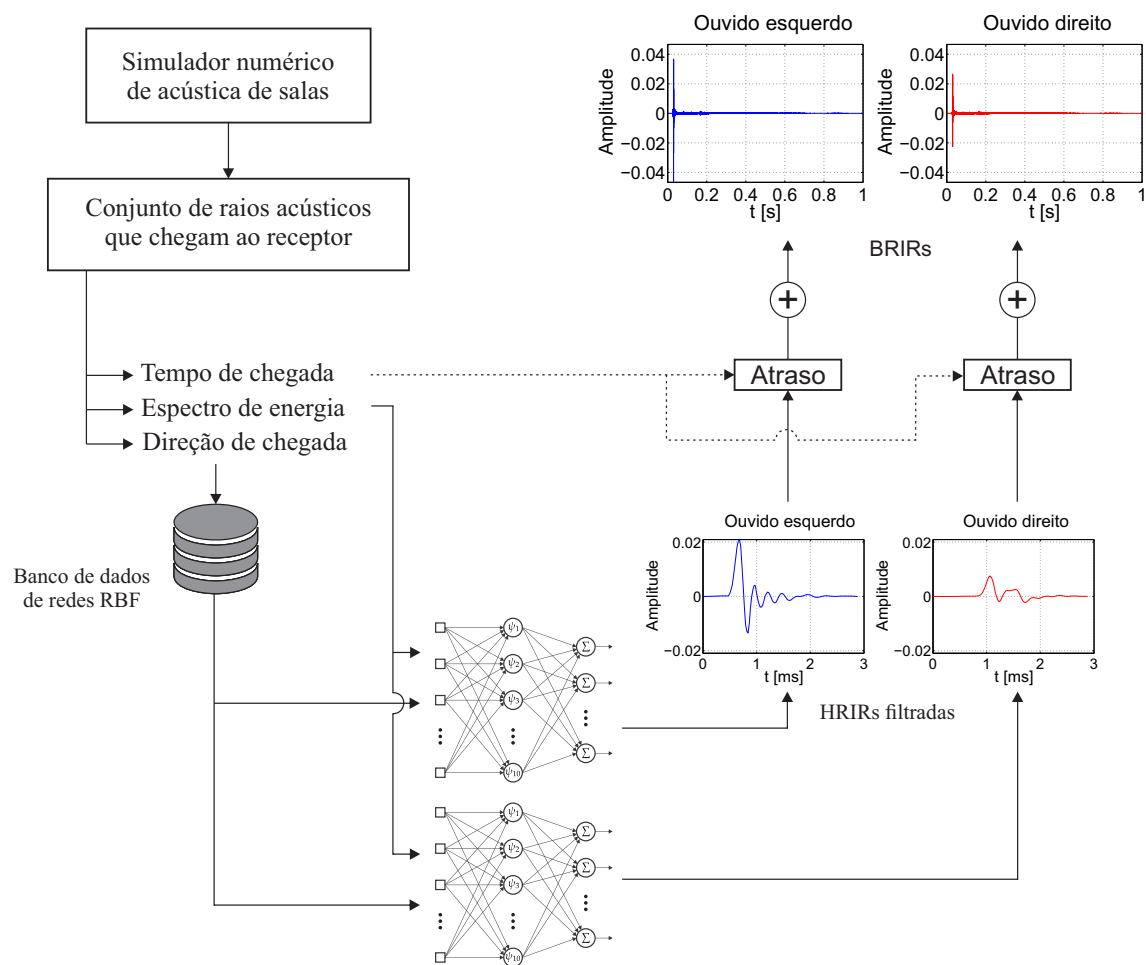
A geração das BRIRs utilizando as redes RBF segue as seguintes etapas:

1. Simular a sala para obter as informações dos raios que chegam ao receptor, ou seja, a direção (ϕ, θ) , o tempo de chegada τ e o espectro de energia por bandas de oitava \mathbf{E} ;
2. Escolher a rede a partir da direção de chegada do raio;
3. A rede escolhida, então, determina a HRIR filtrada (\bar{h}^E, \bar{h}^D) , fornecendo como entrada da rede o espectro de energia \mathbf{E} ;
4. Finalmente, a HRIR filtrada é atrasada pelo tempo de chegada do raio, como descrito na Eq. (36). Esse processo é repetido até que todos os raios sejam computados. E, uma vez finalizado, todas essas HRIRs filtradas e atrasadas são somadas para compor as BRIRs.

A Figura 62 ilustra de forma esquemática a geração das BRIRs utilizando as redes RBF, conforme descrito acima.

No capítulo seguinte serão apresentados os resultados das BRIRs para diversos pares fonte-receptor de várias salas de modo a avaliar o desempenho da metodologia apresentada neste trabalho.

Figura 62 - Esquema de geração das BRIRs via redes RBF



Fonte: O autor, 2017.

5 RESULTADOS DO ROUND ROBIN ON AURALIZATION

Neste capítulo apresenta-se a comparação entre as respostas impulsivas biarticulares medidas pelo grupo de pesquisa SEACEN³ (*Simulation und Evaluation Akustischer Umgebungen*), na primeira fase do *Round Robin on Auralization* (RR4), e os resultados obtidos pelo código computacional RAIOS com a aplicação do modelo das redes neurais proposto neste trabalho.

5.1 *Round Robin on Auralization*

Round Robin é um projeto de intercomparação internacional de simuladores de acústica de salas, cujo propósito é avaliar o desempenho dos *software* participantes. Além disso, é de grande importância para os desenvolvedores para o aperfeiçoamento de seus programas. Já houve quatro edições: a primeira ocorreu no ano de 1993 (VORLÄNDER, 1995), a segunda em 1999 (BORK, 2000), a terceira em 2002 (BORK, 2005a; BORK, 2005b) e a quarta em 2016 (<http://rr.auralisation.net>). Nas três primeiras edições, os softwares foram avaliados a partir do cálculo dos parâmetros de qualidade acústica. Já na última edição o enfoque foi a auralização. Para isso, foram fornecidos os dados geométricos (geometria da sala) de nove salas, os coeficientes de absorção e espalhamento de todas as superfícies, a direcionalidade das fontes sonoras, e as HRIRs da cabeça artificial utilizada. Para a avaliação foram solicitados o cálculo de 157 RIRs e 814 BRIRs. Sabendo que a quantidade de resultados requeridos é muito grande, nesta tese serão apresentados apenas os resultados de três salas, cujas características são descritas na seção seguinte.

5.2 Dados de entrada para a geração das BRIRs

Nesta seção descrevem-se apenas três das nove salas utilizadas na última edição do *round robin* e que serão objeto de estudo desta pesquisa. A primeira sala foi denominada de *Seminar room* (Figura 63), a segunda de *Chamber music hall* (Figura 64) e a terceira de *Auditorium* (Figura 65). Comparando as salas pelos seus volumes, pode-se considerar que primeira é uma sala pequena (144 m³), a segunda é média (3340 m³) e a terceira é grande (8490 m³). Nas Tabelas 6-8 estão contidos os valores dos coeficientes de

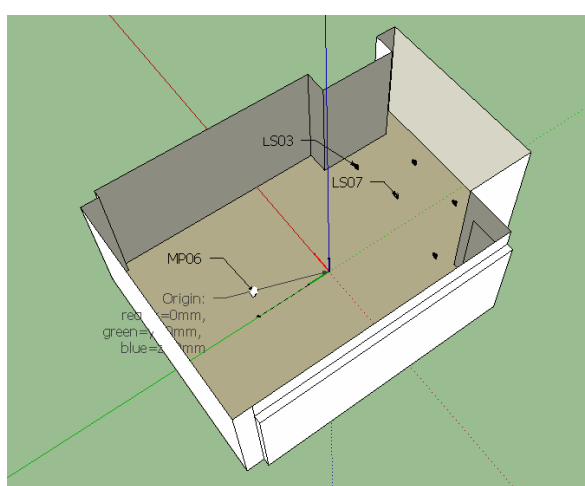
³ <http://www.seacen.tu-berlin.de/seacen/>

absorção e espalhamento utilizados em cada sala, todos eles fornecidos pelo SEACEN. Nas Figuras 63b, 64c e 65c estão destacados os pares de fonte-receptor (LS-MP) onde serão calculadas as BRIRs. As posições das fontes e dos receptores são apresentados na Tabela 9. A Tabela 10 fornece a direção do eixo principal de radiação das fontes. Esse dado é importante já que a fonte utilizada é direcional. Essa tabela também indica a orientação da cabeça. Essa informação é relevante pois as BRIRs variam, não só com relação à posição relativa entre fonte e receptor, mas também com a orientação da cabeça. A partir dessas informações, pode-se notar que o eixo principal de radiação da fonte LS7 está contido no plano mediano da cabeça MP6 em todas os casos. E a fonte LS3 sempre à esquerda desse plano. Como todas as salas apresentam uma razoável simetria em relação ao plano mediano da cabeça MP6, é de se esperar que a BRIR do ouvido esquerdo seja similar ao do ouvido direito para o par LS7-MP6. Por outro lado, para o par LS3-MP6 a BRIR do ouvido esquerdo terá amplitude maior que a do ouvido direito.

Figura 63 - Sala do *Round Robin 4* denominada *Seminar room* (sala pequena)



(a)



(b)

Legenda: (a) - Sala real; (b) - Posições das fontes e do receptor.

Fonte: O autor, 2017.

Tabela 6 - Coeficientes de absorção (α) e espalhamento (δ) em bandas de oitava utilizadas nas superfícies da sala *Seminar room*.

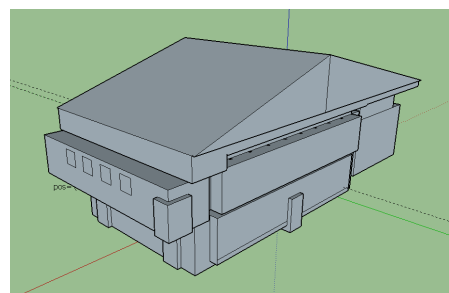
		63 Hz	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz
<i>Ceiling</i>	α	0,004	0,048	0,096	0,056	0,056	0,055	0,099	0,111	0,158
	δ	0,014	0,021	0,031	0,047	0,070	0,331	0,582	0,624	0,668
<i>Concrete</i>	α	0,041	0,049	0,069	0,066	0,068	0,069	0,069	0,090	0,119
	δ	0,006	0,008	0,011	0,014	0,018	0,024	0,031	0,040	0,189
<i>Floor</i>	α	0,046	0,041	0,084	0,082	0,075	0,072	0,068	0,072	0,094
	δ	0,003	0,004	0,005	0,007	0,009	0,012	0,015	0,020	0,095
<i>Plaster</i>	α	0,040	0,019	0,046	0,046	0,051	0,056	0,057	0,062	0,117
	δ	0,009	0,012	0,016	0,021	0,027	0,035	0,046	0,060	0,284
<i>Window</i>	α	0,119	0,101	0,067	0,057	0,066	0,155	0,087	0,116	0,121
	δ	0,005	0,006	0,008	0,010	0,014	0,018	0,023	0,030	0,142

Fonte: O autor, 2017.

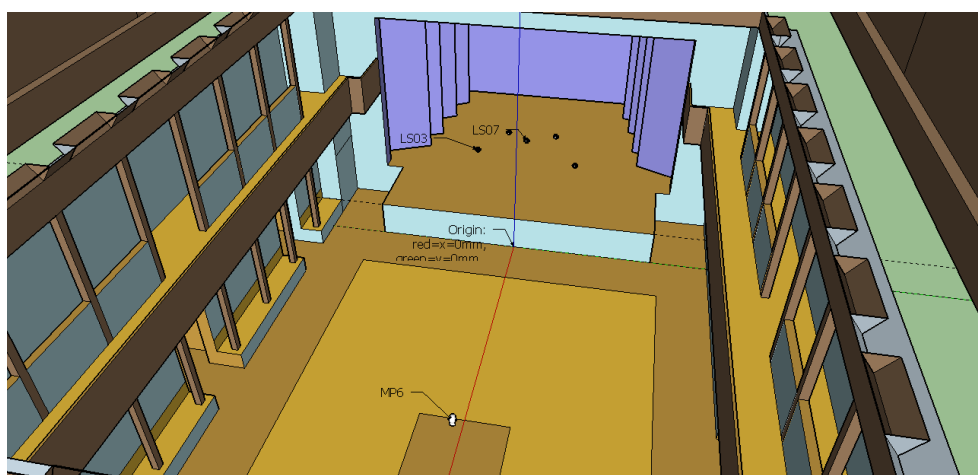
Figura 64 - Sala do *Round Robin 4* denominada *Chamber music hall* (sala média)



(a)



(b)



(c)

Legenda: (a) - Sala real; (b) - Vista externa em perspectiva; (c) - Posições das fontes e do receptor.

Fonte: O autor, 2017.

Tabela 7 - Coeficientes de absorção (α) e espalhamento (δ) em bandas de oitava utilizadas nas superfícies da sala *Chamber music hall*.

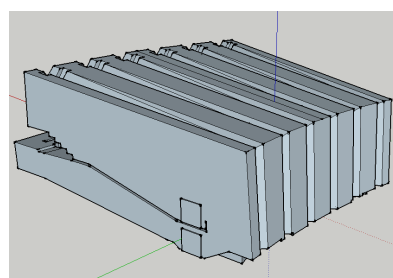
		63 Hz	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz
<i>Ceiling</i>	α	0,050	0,051	0,056	0,092	0,126	0,155	0,211	0,204	0,185
	δ	0,007	0,009	0,013	0,019	0,027	0,038	0,177	0,319	0,354
<i>Floor</i>	α	0,088	0,038	0,054	0,052	0,061	0,062	0,070	0,111	0,143
	δ	0,003	0,005	0,008	0,013	0,042	0,110	0,128	0,136	0,144
<i>Plaster</i>	α	0,071	0,091	0,092	0,106	0,122	0,138	0,159	0,171	0,168
	δ	0,003	0,004	0,006	0,008	0,010	0,014	0,018	0,056	0,165
<i>Seating</i>	α	0,128	0,139	0,241	0,328	0,395	0,665	0,363	0,581	0,736
	δ	0,047	0,151	0,574	0,739	0,763	0,789	0,815	0,842	0,870
<i>Stage panels</i>	α	0,034	0,026	0,056	0,069	0,089	0,101	0,123	0,141	0,153
	δ	0,002	0,002	0,003	0,003	0,005	0,006	0,008	0,010	0,047
<i>Structured Plaster</i>	α	0,070	0,081	0,081	0,122	0,172	0,197	0,159	0,171	0,168
	δ	0,012	0,018	0,027	0,040	0,060	0,284	0,499	0,535	0,573

Fonte: O autor, 2017.

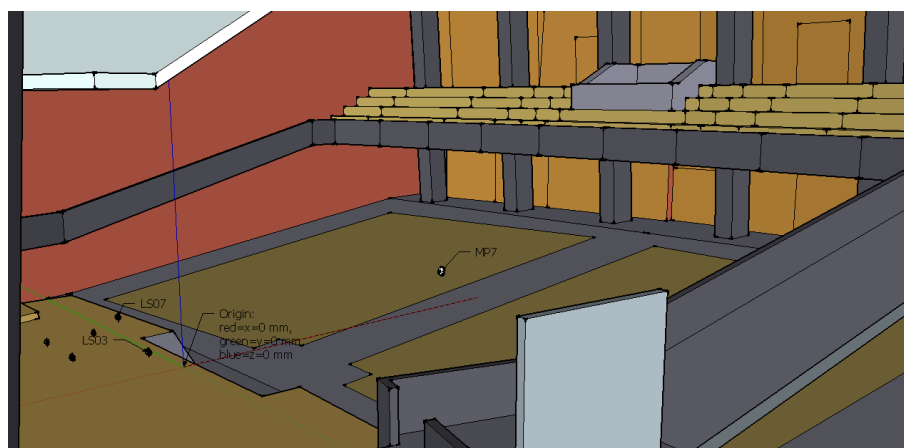
Figura 65 - Sala do *Round Robin 4* denominada *Auditorium* (sala grande)



(a)



(b)



(c)

Legenda: (a) - Sala real; (b) - Vista externa em perspectiva; (c) - Posição das fontes e dos receptores.

Fonte: O autor, 2017.

Tabela 8 - Coeficientes de absorção (α) e espalhamento (δ) em bandas de oitava utilizadas nas superfícies da sala *Auditorium*.

		63 Hz	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz
Brickwall	α	0,154	0,136	0,091	0,213	0,106	0,382	0,267	0,502	0,495
	δ	0,010	0,015	0,022	0,033	0,050	0,237	0,416	0,446	0,477
Concrete	α	0,039	0,058	0,065	0,076	0,090	0,068	0,086	0,158	0,143
	δ	0,003	0,005	0,006	0,009	0,012	0,016	0,021	0,099	0,192
Linoleum	α	0,139	0,079	0,078	0,077	0,088	0,203	0,158	0,149	0,169
	δ	0,005	0,007	0,009	0,012	0,016	0,022	0,030	0,142	0,274
Parquet	α	0,075	0,041	0,039	0,037	0,066	0,138	0,179	0,173	0,170
	δ	0,003	0,004	0,006	0,007	0,010	0,013	0,017	0,034	0,145
Seating	α	0,153	0,066	0,224	0,312	0,404	0,503	0,608	0,698	0,803
	δ	0,029	0,062	0,225	0,581	0,667	0,692	0,718	0,745	0,772
White panels	α	0,072	0,066	0,064	0,061	0,119	0,150	0,143	0,126	0,162
	δ	0,003	0,004	0,006	0,007	0,010	0,013	0,017	0,034	0,145
Wood panels	α	0,068	0,051	0,139	0,170	0,166	0,201	0,219	0,237	0,318
	δ	0,011	0,018	0,027	0,042	0,093	0,356	0,469	0,499	0,531

Fonte: O autor, 2017.

Tabela 9 - Posições das fontes e dos receptores nas salas *Seminar room*, *Chamber music hall* e *Auditorium* no sistema de coordenadas cartesianas de cada sala.

Posição [mm]	<i>Seminar room</i>			<i>Chamber music hall</i>			<i>Auditorium</i>		
	LS3	LS7	MP6	LS3	LS7	MP6	LS3	LS7	MP6
<i>x</i>	1498	0	0	-2020	-2020	7840	-1500	-1500	8500
<i>y</i>	-1974	-1974	2026	-1500	0	0	-1500	0	0
<i>z</i>	1026	1500	1230	2030	2500	1230	1030	1500	1090

Fonte: O autor, 2017.

Tabela 10 - Orientação do eixo principal de radiação das fontes e das cabeças das salas *Seminar room*, *Chamber music hall* e *Auditorium*.

Direção [°]	<i>Seminar room</i>			<i>Chamber music hall</i>			<i>Auditorium</i>		
	LS3	LS7	MP6	LS3	LS7	MP6	LS3	LS7	MP6
Azimute (ϕ)	98,7	90	270	8,7	0	180	8,5	0	180
Elevação (θ)	0	0	0	0	0	0	0	0	0

Fonte: O autor, 2017.

5.3 Comparação dos resultados numéricos com os medidos

Nesta seção são apresentados resultados com o objetivo de averiguar a capacidade do código RAIOS em gerar as BRIRs. Para isso, o modelo apresentado nesta tese será confrontado com os resultados medidos. Vale observar, contudo, que os resultados encontrados dependem fortemente dos algoritmos de cálculo do código computacional RAIOS.

As BRIRs das Figuras 66-71 foram normalizadas ao intervalo $[-1, 1]$. Além disso, para comparar as BRIRs simuladas com as medidas utilizaram-se os parâmetros de qualidade acústica (PQA), tais como T_{30} , D_{50} , C_{80} e T_s , ver Tabelas 11-16. O tempo de reverberação, T_{30} , já foi explicado no Capítulo 1. O parâmetro adimensional *definição*, D_{50} , relaciona a energia sonora recebida nos 50 ms com a energia sonora total (THIELE, 1953), ou seja,

$$D_{50} = \frac{\int_0^{50\text{ms}} p^2(t) dt}{\int_0^{\infty} p^2(t) dt} \times 100\%, \quad (120)$$

onde $p^2(t)$ é a resposta impulsiva quadrática, e t é o tempo. O parâmetro *clareza*, C_{80} , dado em dB, relaciona a energia sonora recebida nos primeiros 80 ms do sinal recebido com a energia sonora recebida dos 80 ms até o final da reverberação (REICHARDT et al., 1974), ou seja,

$$C_{80} = 10 \log \frac{\int_0^{80\text{ms}} p^2(t) dt}{\int_{80\text{ms}}^{\infty} p^2(t) dt}. \quad (121)$$

Por fim, o parâmetro *tempo central*, T_s , é definido como o instante onde a energia recebida antes deste ponto é igual à energia recebida depois deste ponto (KÜRER, 1971), ou seja,

$$T_s = \frac{\int_0^{\infty} t \cdot p^2(t) dt}{\int_0^{\infty} p^2(t) dt}. \quad (122)$$

Embora, esses parâmetros sejam utilizados apenas para as RIRs, pode-se utilizá-los para verificar o balanço energético das respostas impulsivas biauriculares. É claro que a avaliação não é a mesma, já que, diferentemente das RIRs, as BRIRs contêm em sua estrutura as modificações geradas pela presença da cabeça. Em cada uma dessas tabelas mostram os PQA calculados para as componentes esquerda e direita das BRIRs medidas e simuladas (e avaliadas pelo erro relativo absoluto), para cada sala e para cada par de fonte-receptor.

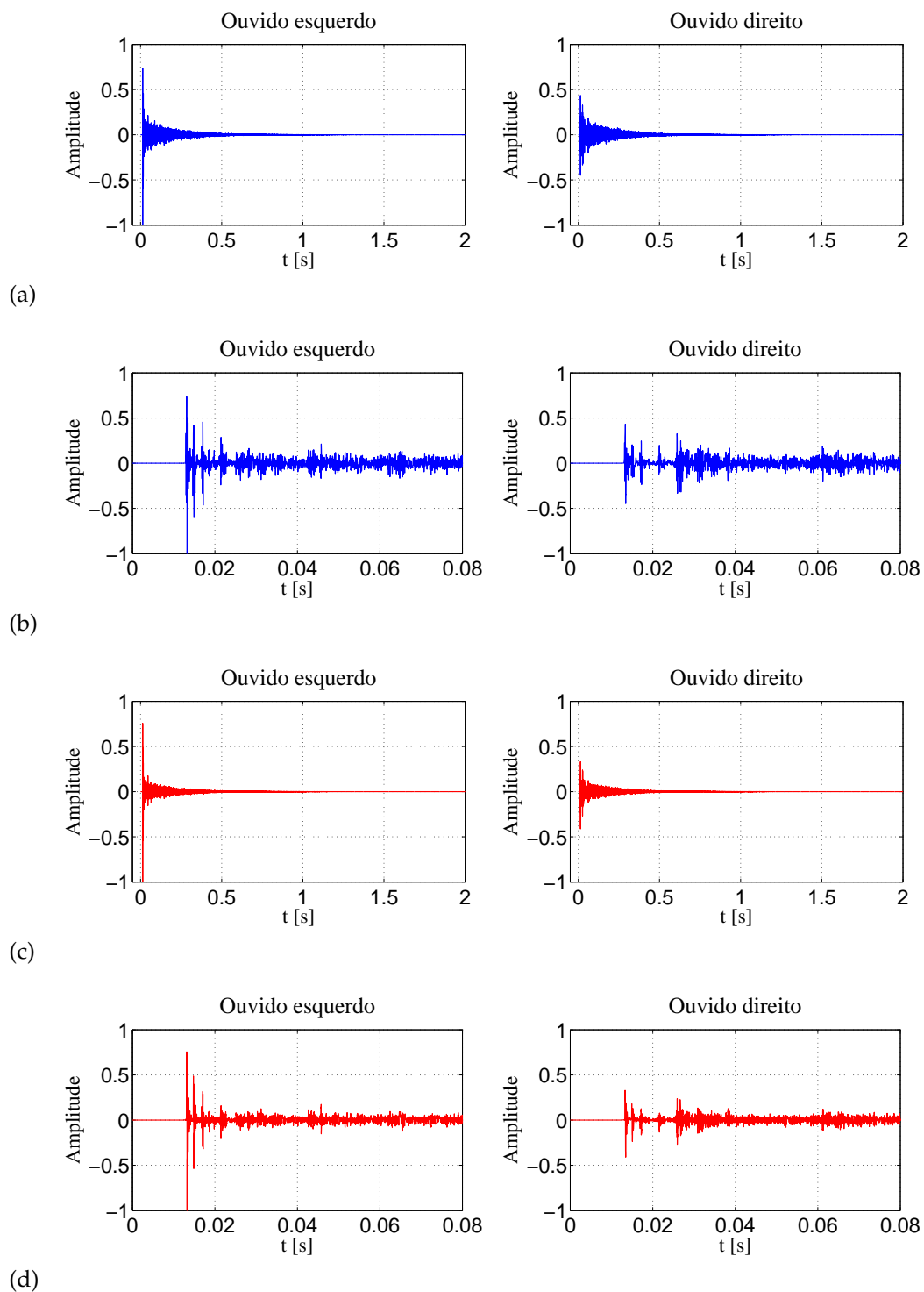
A Figura 66 mostra BRIRs medidas e simuladas para o par LS3-MP6 da sala

Seminar room. Embora não haja uma concordância pico a pico, pode-se observar uma excelente concordância entre os envelopes das BRIRs. A concordância pico a pico não é possível, já que sempre haverá uma incerteza nos dados de entrada do simulador, principalmente em relação às medições dos coeficientes de absorção e de espalhamento. Já a Figura 67 mostra BRIRs medidas e simuladas para o par LS7-MP6 dessa mesma sala. Note-se que a mudança da posição da fonte não afeta a qualidade do resultado, ou seja, ainda há uma boa concordância entre os envelopes. As Tabelas 11 e 12 mostram que para ambos os pares fonte-receptor há uma boa concordância entre os PQA medidos e simulados.

A Figura 68 mostra BRIRs medidas e simuladas para o par LS3-MP6 da sala *Chamber music hall*. Note que, neste caso, há um pequeno desvio nos envelopes das BRIRs medidas e simuladas. Essa discordância é, muito provavelmente, uma consequência do aumento da distância entre a fonte e o receptor, e, além disso, devido ao aumento no volume da sala, já que esse aumento diminui a probabilidade do raio interceptar o receptor. Sendo assim, maior é a quantidade de raios “perdidos”. Mudando a posição da fonte para LS7 pode-se observar os mesmos efeitos comentados anteriormente, como mostra a Figura 69. Embora haja uma boa aderência entre o T_{30} e T_s medidos com os simulados, há um desvio maior no D_{50} e no C_{80} , para ambos os pares fonte-receptor, como mostram as Tabelas 13 e 14. Esses maiores desvios são causados pela “falta” de energia na cauda reverberante devido a poucos raios sonoros chegando ao receptor neste intervalo de tempo.

Finalmente apresentam-se os resultados da última sala, ou seja, o *Auditorium*. Primeiramente, mostram-se BRIRs medidas e simuladas para o par LS3-MP6, ver Figura 70. Apesar da intensidade entre os ouvidos estarem coerentes, isto é, a intensidade da BRIR esquerda ser maior que a do ouvido direito, observa-se que o pico máximo e mínimo do simulado divergem do medido. Além disso, as BRIRs simuladas têm um decaimento energético mais rápido do que as medidas. Esse fato pode ter ocorrido devido ao aumento da distância da fonte ao receptor e do volume da sala, como aconteceu com os resultados da sala *Chamber music hall*. Note que essas discordâncias também ocorrem com o par LS7-MP6, como pode ser observado na Figura 71. Como consequência os PQA também apresentam uma piora significativa nos resultados, ver Tabelas 15 e 16.

Figura 66 - Comparação das BRIRs medidas e simuladas para o par LS3-MP6 da sala *Seminar room*

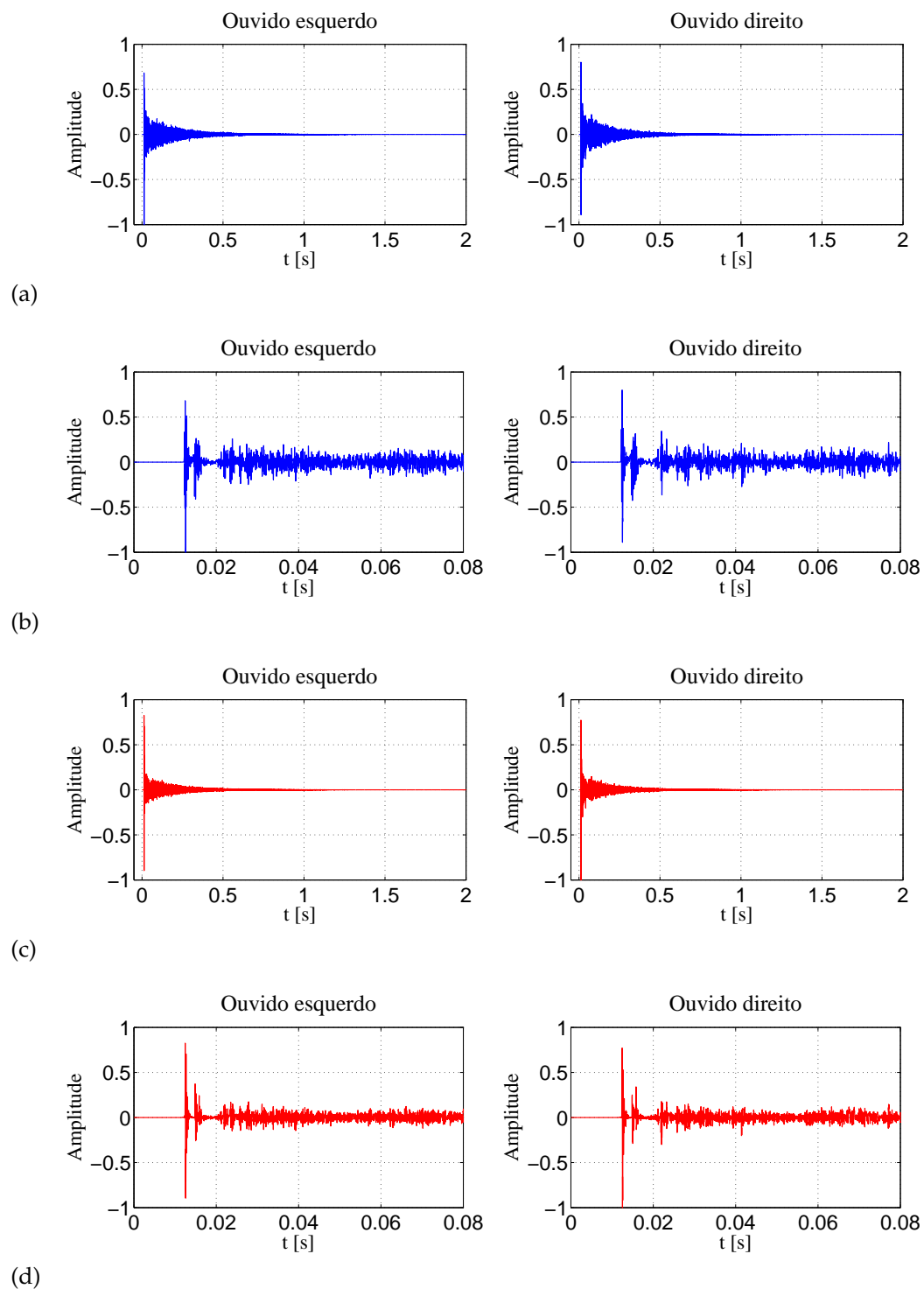


Legenda: (a) - Medidas; (b) - Medidas - detalhamento no intervalo de 0 a 80 ms;

(c) - Simuladas; (d) - Simuladas - detalhamento no intervalo de 0 a 80 ms.

Fonte: O autor, 2017.

Figura 67 - Comparação das BRIRs medidas e simuladas para o par LS7-MP6 da sala *Seminar room*

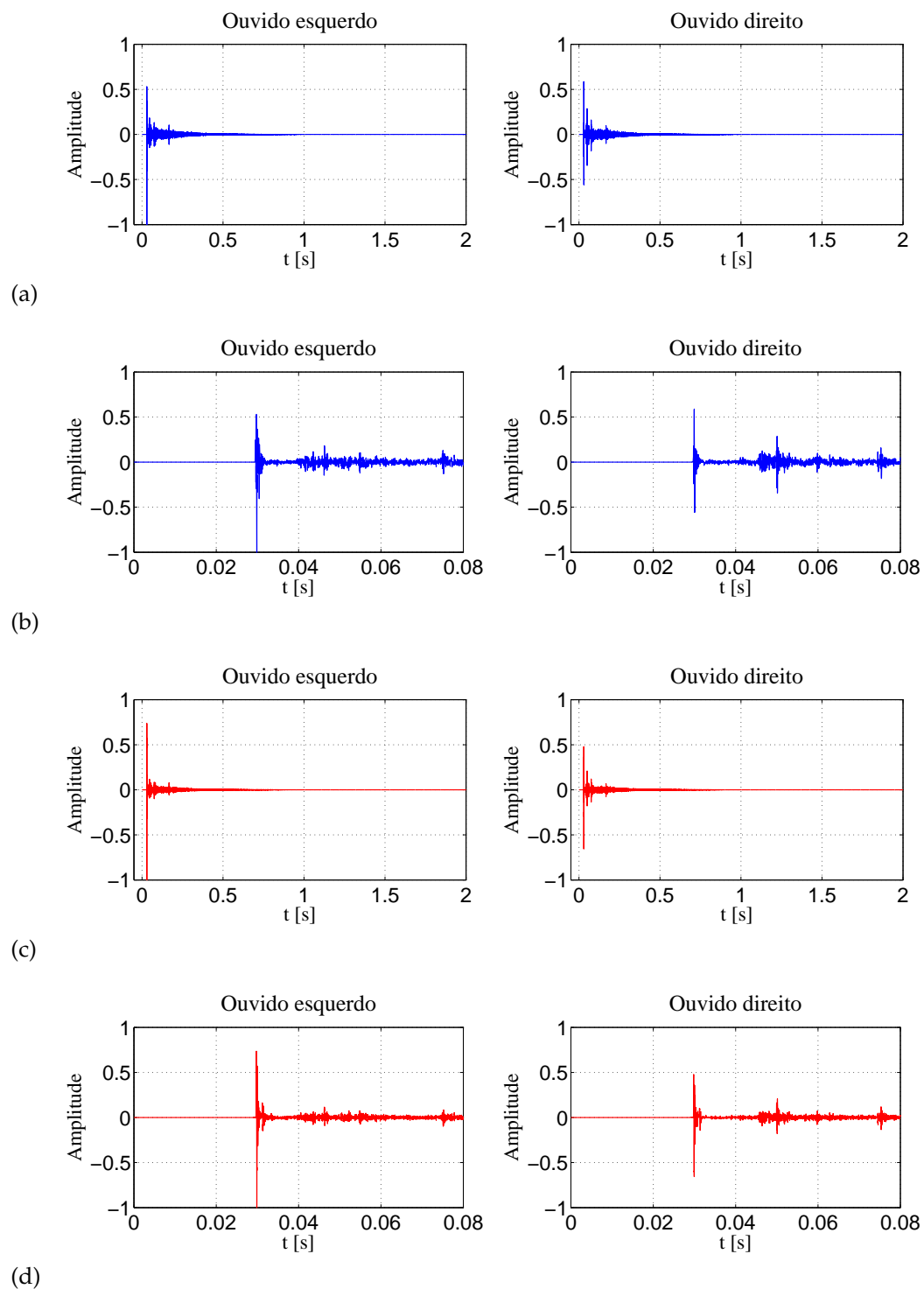


Legenda: (a) - Medidas; (b) - Medidas - detalhamento no intervalo de 0 a 80 ms;

(c) - Simuladas; (d) - Simuladas - detalhamento no intervalo de 0 a 80 ms.

Fonte: O autor, 2017.

Figura 68 - Comparação das BRIRs medidas e simuladas para o par LS3-MP6 da sala *Chamber music hall*

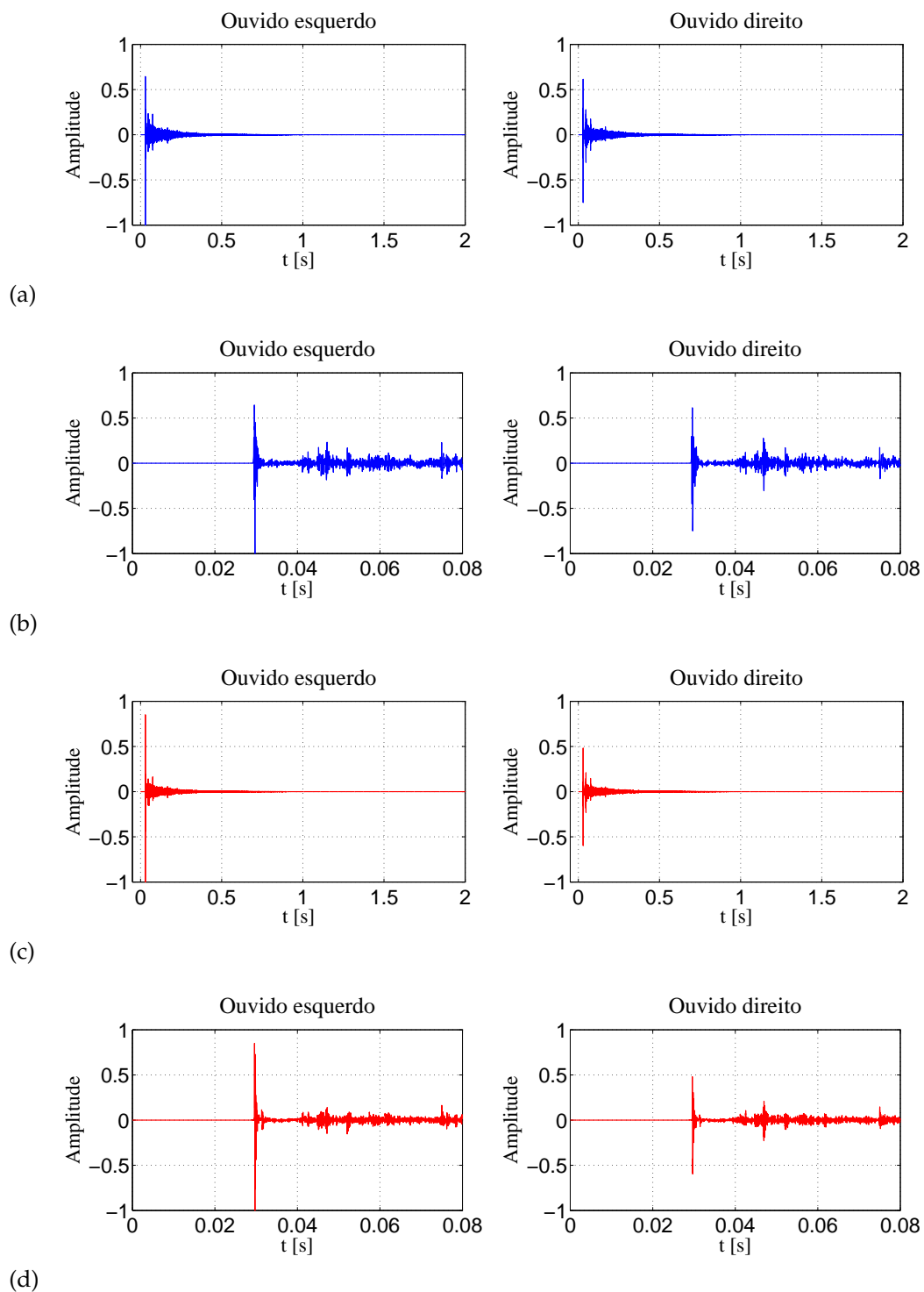


Legenda: (a) - Medidas; (b) - Medidas - detalhamento no intervalo de 0 a 80 ms;

(c) - Simuladas; (d) - Simuladas - detalhamento no intervalo de 0 a 80 ms.

Fonte: O autor, 2017.

Figura 69 - Comparação das BRIRs medidas e simuladas para o par LS7-MP6 da sala *Chamber music hall*

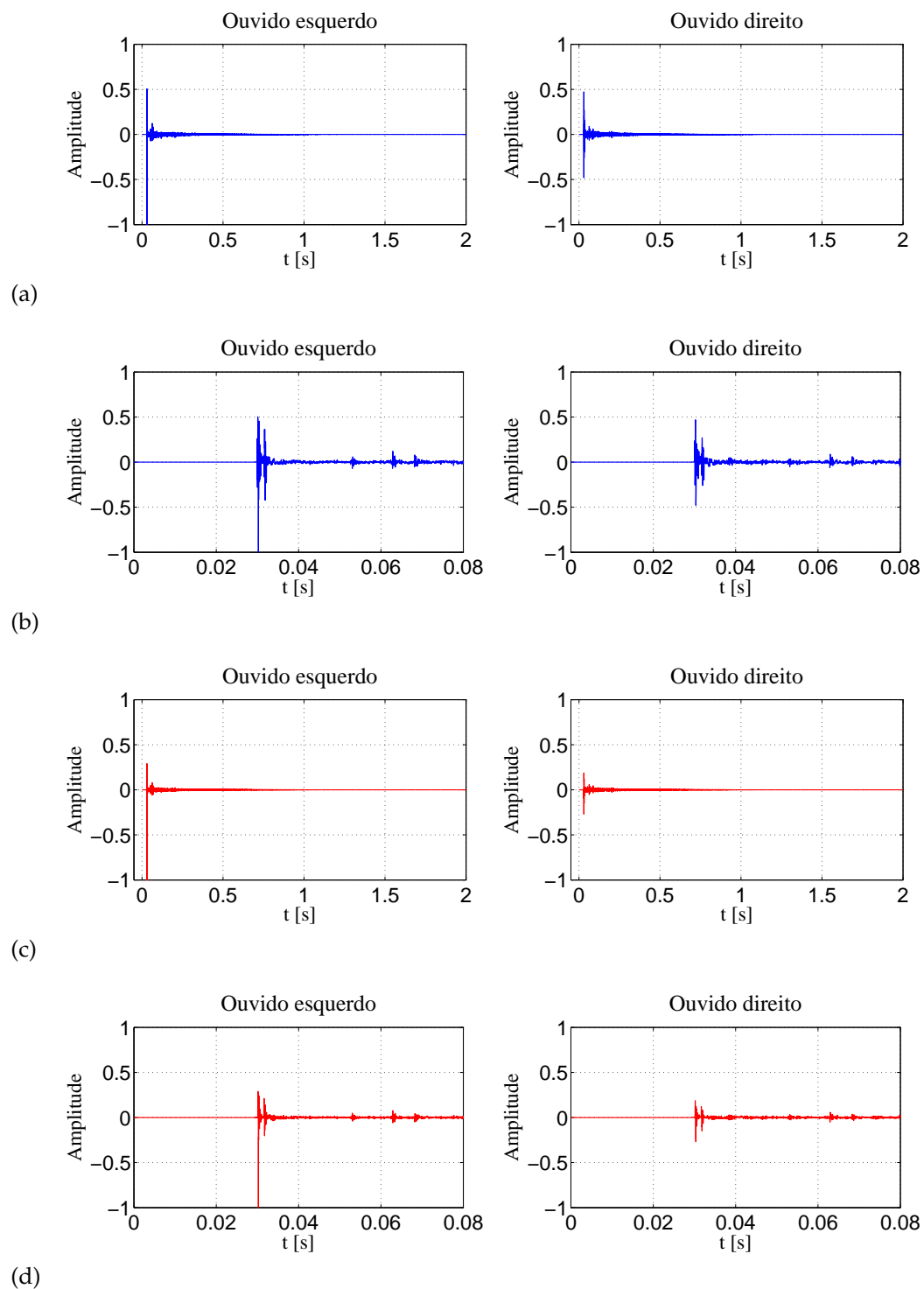


Legenda: (a) - Medidas; (b) - Medidas - detalhamento no intervalo de 0 a 80 ms;

(c) - Simuladas; (d) - Simuladas - detalhamento no intervalo de 0 a 80 ms.

Fonte: O autor, 2017.

Figura 70 - Comparação das BRIRs medidas e simuladas para o par LS3-MP6 da sala *Auditorium*

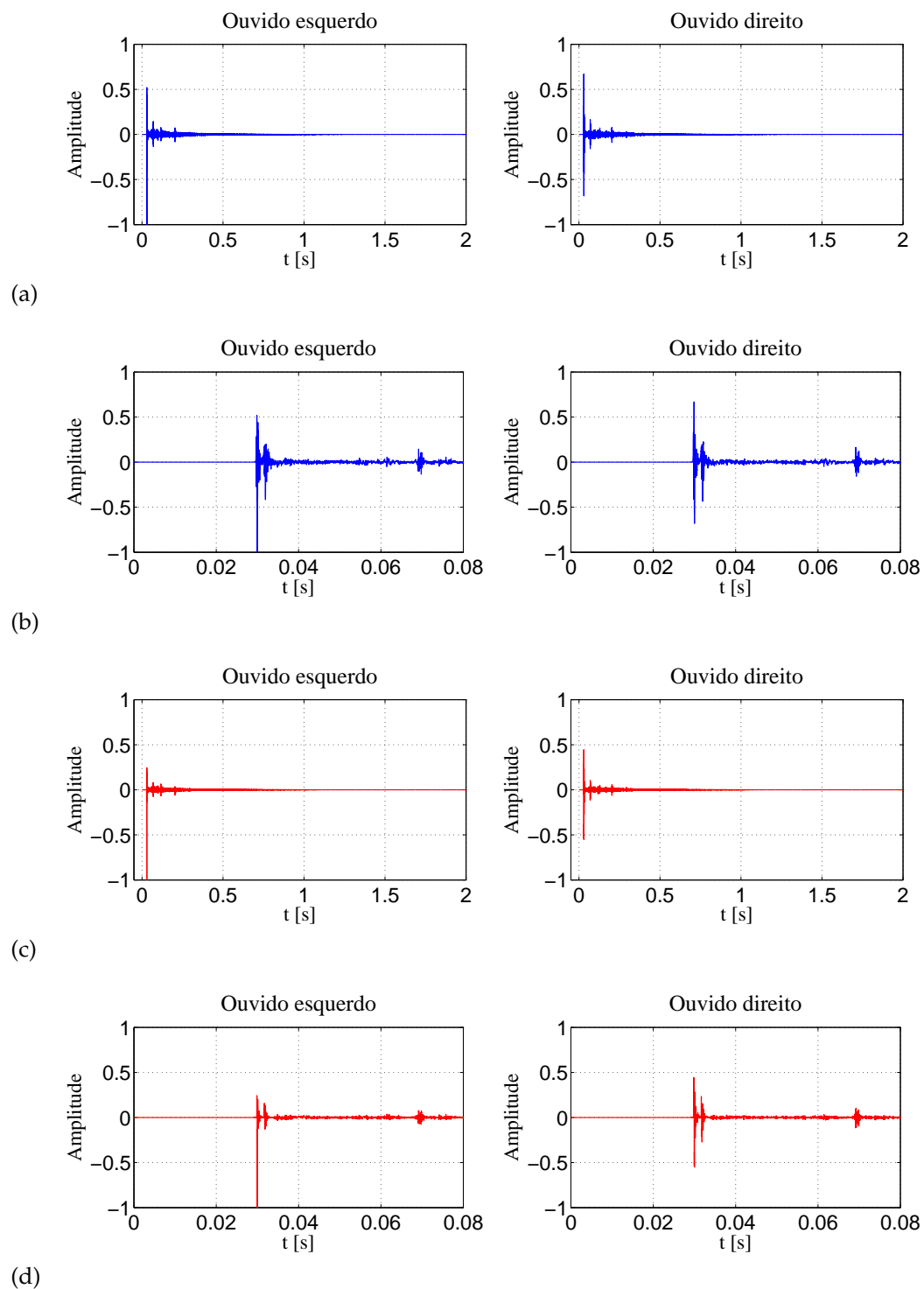


Legenda: (a) - Medidas; (b) - Medidas - detalhamento no intervalo de 0 a 80 ms;

(c) - Simuladas; (d) - Simuladas - detalhamento no intervalo de 0 a 80 ms.

Fonte: O autor, 2017.

Figura 71 - Comparação das BRIRs medidas e simuladas para o par LS7-MP6 da sala *Auditorium*



Legenda: (a) - Medidas; (b) - Medidas - detalhamento no intervalo de 0 a 80 ms;

(c) - Simuladas; (d) - Simuladas - detalhamento no intervalo de 0 a 80 ms.

Fonte: O autor, 2017.

Tabela 11 - Comparação do parâmetros de qualidade acústica das BRIRs medidas e simuladas, sala *Seminar room*, par LS3-MP6.

PQA	Ouvido esquerdo			Ouvido direito		
	Medido	Simulado	Erro relativo [%]	Medido	Simulado	Erro relativo [%]
T_{30} [s]	1,45	1,46	0,7	1,48	1,49	0,7
D_{50} [%]	54,46	60,81	11,7	39,42	40,72	3,3
C_{80} [dB]	2,86	3,82	33,6	0,94	1,14	21,3
T_s [s]	0,08	0,07	12,5	0,10	0,08	20

Fonte: O autor, 2017.

Tabela 12 - Comparação do parâmetros de qualidade acústica das BRIRs medidas e simuladas, sala *Seminar room*, par LS7-MP6.

PQA	Ouvido esquerdo			Ouvido direito		
	Medido	Simulado	Erro relativo [%]	Medido	Simulado	Erro relativo [%]
T_{30} [s]	1,49	1,48	0,7	1,49	1,48	0,7
D_{50} [%]	44,68	51,14	14,5	44,42	50,14	12,9
C_{80} [dB]	1,50	2,34	56,0	1,43	2,24	56,6
T_s [s]	0,09	0,1	11,1	0,09	0,1	11,1

Fonte: O autor, 2017.

Tabela 13 - Comparação do parâmetros de qualidade acústica das BRIRs medidas e simuladas, sala *Chamber music hall*, par LS3-MP6.

PQA	Ouvido esquerdo			Ouvido direito		
	Medido	Simulado	Erro relativo [%]	Medido	Simulado	Erro relativo [%]
T_{30} [s]	1,23	1,21	1,6	1,23	1,20	2,4
D_{50} [%]	56,92	72,24	26,9	40,43	57,78	42,9
C_{80} [dB]	3,70	6,22	68,1	2,82	4,90	73,8
T_s [s]	0,08	0,06	25,0	0,09	0,07	22,2

Fonte: O autor, 2017.

Tabela 14 - Comparação do parâmetros de qualidade acústica das BRIRs medidas e simuladas, sala *Chamber music hall*, par LS7-MP6.

PQA	Ouvido esquerdo			Ouvido direito		
	Medido	Simulado	Erro relativo [%]	Medido	Simulado	Erro relativo [%]
T_{30} [s]	1,19	1,17	1,7	1,23	1,24	0,8
D_{50} [%]	51,20	65,04	27,0	56,67	49,66	12,4
C_{80} [dB]	3,49	5,38	54,2	4,14	3,26	21,3
T_s [s]	0,08	0,06	25,0	0,08	0,06	25,0

Fonte: O autor, 2017.

Tabela 15 - Comparação do parâmetros de qualidade acústica das BRIRs medidas e simuladas, sala *Auditorium*, par LS3-MP6.

PQA	Ouvido esquerdo			Ouvido direito		
	Medido	Simulado	Erro relativo [%]	Medido	Simulado	Erro relativo [%]
T_{30} [s]	1,57	1,64	4,5	1,68	1,75	4,2
D_{50} [%]	84,36	98,58	16,9	76,22	54,11	29,0
C_{80} [dB]	9,17	7,90	13,8	6,51	2,62	59,8
T_s [s]	0,05	0,07	40,0	0,06	0,09	50,0

Fonte: O autor, 2017.

Tabela 16 - Comparação do parâmetros de qualidade acústica das BRIRs medidas e simuladas, sala *Auditorium*, par LS7-MP6.

PQA	Ouvido esquerdo			Ouvido direito		
	Medido	Simulado	Erro relativo [%]	Medido	Simulado	Erro relativo [%]
T_{30} [s]	1,66	1,70	2,4	1,65	1,67	1,2
D_{50} [%]	75,30	64,28	14,6	72,83	87,76	20,5
C_{80} [dB]	6,57	4,57	30,4	6,01	5,04	16,1
T_s [s]	0,06	0,08	33,3	0,07	0,09	28,5

Fonte: O autor, 2017.

CONCLUSÕES E CONTINUIDADE DO TRABALHO

Embora haja modelos de geração de aurilização que permitam previsões precisas e realistas em aplicações complexas, ainda possuem um custo computacional muito elevado, motivo pelo qual leva o projetista a procurar uma nova metodologia e/ou simplificações que não comprometam o resultado. Neste trabalho, foi apresentada uma nova técnica para a geração das respostas impulsivas biauriculares (BRIRs) com intuito de reduzir o seu custo computacional. Para isso foi formulado um novo procedimento de cálculo das HRIRs filtradas. Além disso, outro objetivo foi recuperar com precisão os componentes de alta frequência dessas HRTFs filtradas. Para isso, foram projetadas um conjunto de redes neurais artificiais do tipo RBF. As redes do tipo *multilayer perceptron* (MLP) não foram utilizadas, uma vez que a convergência no processo de treinamento é muito lenta e é necessário uma quantidade maior de neurônios para a obter o resultado com a mesma precisão da rede RBF, o que implicaria no aumento do custo computacional. Embora a técnica das redes neurais para a modelagem das HRIRs filtradas não seja um conceito novo, as técnicas desenvolvidas aqui trazem ideias ainda não exploradas em outros trabalhos.

Como se dispunha um banco de HRTFs suficientemente discretizado, com resolução espacial de 1° tanto no azimute quanto na elevação, foi modelada uma rede para cada direção onde há uma HRTF medida, o que permitiu reduzir a complexidade da rede neural, pois não é necessária a interpolação. A rede que apresentou a melhor relação custo-benefício foi utilizando dez neurônios na camada intermediária, tornando possível uma redução de 70% do número de operações aritméticas em relação ao método tradicional, que utiliza convoluções, reduzindo, portanto, o custo computacional do processo de geração da aurilização. O conjunto de dados formado, com as estratégias apresentadas neste trabalho, permitiu contornar as dificuldades encontradas em Naranjo (2014). Esses padrões de treinamento foram obtidos a partir da seleção, empregando o algoritmo de agrupamento *K-means*, dos raios gerados da simulação de diversas salas, com diferentes características acústicas, utilizando o código computacional RAIOS. Além disso, a função custo utilizada no algoritmo tradicional de treinamento das redes RBF foi alterada, a qual contribuiu para um bom desempenho da rede neural.

Para avaliar o modelo proposto, primeiramente, foram feitas várias comparações entre a modelagem das HRIRs filtradas utilizando as redes neurais e o método tradicional. Os resultados mostraram que as redes projetadas são capazes de reconstruir as HRIRs por meio da modificação espectral. Um número bem significativo de direções foram testadas — dentre as quais oito dessas direções foram apresentadas. Porém em todas as direções testadas o coeficiente de correlação comprovou a exatidão do modelo.

Na sequência, é simulada a geração das BRIRs, de três salas, denominadas como *Seminar room*, *Chamber music hall* e *Auditorium*, utilizadas no *Round Robin on Auralization* (RR4). Essas salas escolhidas estão em ordem crescente no volume e no grau de complexidade. Dois pares de fonte-receptor foram selecionados para o cálculo das BRIRs. Para o cálculo das BRIRs simuladas foram utilizados o simulador RAIOS com o modelo das redes RBF. Como o RAIOS utiliza dois algoritmos para modelar a propagação da onda acústica, a saber o método traçado de raios e o método de transição de energia, foi necessária a modificação deste último para a aplicação às redes projetadas. Desse modo, as BRIRs são construídas, não só a partir das reflexões especulares, mas também das reflexões difusas. Os resultados foram então avaliados a partir dos parâmetros de qualidade acústica (PQA). Embora, esses PQA sejam utilizados apenas para as RIRs, eles se prestam para verificar o balanço energético das componentes esquerda e direita das BRIRs. Assim, a partir dos PQA pode-se concluir que o modelo apresenta resultados confiáveis para as salas *Seminar room* e *Chamber music hall*. Contudo, os da sala *Auditorium* não foram tão satisfatórios. Logo, os resultados obtidos no presente trabalho revelam que à medida que o volume da sala aumenta, os resultados tendem a ser piores. Isso pode ser comprovado pelo crescimento dos erros nos parâmetros de qualidade acústica. Por fim, vale ressaltar que, embora, a técnica apresentada tenha se mostrada bastante promissora, principalmente do ponto de vista de uma implementação eficiente em um sistema de auralização, é necessário uma análise nos algoritmos que o RAIOS utiliza na modelagem da propagação da onda acústica, uma vez que erros no decaimento energético e na modelagem do som direto, principalmente, em salas grandes foram observados nas BRIRs simuladas. Isso será alvo de investigações futuras.

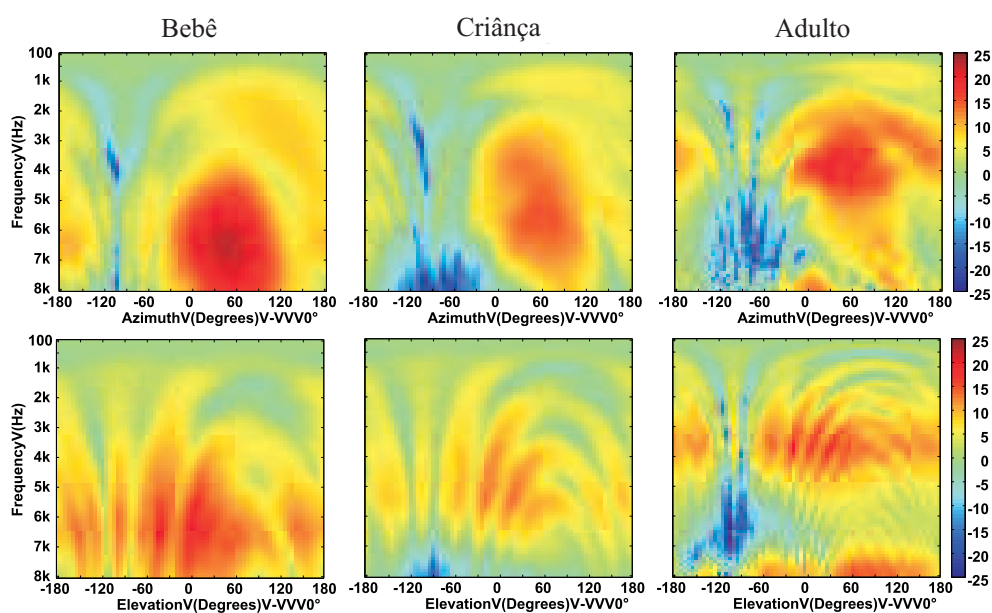
Os próximos passos na evolução deste trabalho estão listados a seguir.

- Analisar o código fonte dos algoritmos utilizados na modelagem da propagação da onda acústica do código computacional RAIOS, no intuito de verificar o motivo de o decaimento energético das BRIRs simuladas apresentarem-se mais acentuadas do que as medidas, particularmente para salas de grande volume.
- Estudar a possibilidade de reduzir ainda mais o tempo de resposta do método com técnicas de computação paralela.
- Considerar a possibilidade de incrementar outros elementos na entrada da rede, tais como as características antropométricas do ouvinte de maneira que o sistema possa fornecer resultados personalizados.
- Desenvolver uma técnica de validação objetiva de processos de auralização.
- Analisar as similaridades das redes, desenvolvidas neste trabalho, próximas com o objetivo de reduzir a sua quantidade.

- Outra importante contribuição que se pode buscar é o desenvolvimento de um sistema de aurilização utilizando HRTFs de cabeças infantis. Sabendo que as HRTFs mudam com a idade (FELS, 2008), as redes desenvolvidas aqui são úteis apenas para adultos. A Figura 72 mostra a comparação entre as HRTFs de um bebê, uma criança e um adulto no plano horizontal e mediano. Note-se que, em torno de 3 kHz, corresponde à região de maior sensibilidade de um adulto, enquanto de um bebê ou de uma criança é de 6 kHz.

O principal propósito desse estudo consiste na aurilização de salas de aula. Como a aurilização permite prever a acústica da sala mesmo que ela não exista, isso permite ao projetista analisar e otimizar a qualidade acústica. Uma vez que, a qualidade acústica de uma sala de aula é o fator preponderante na qualidade e eficácia do aprendizado.

Figura 72 - Comparação entre as HRTFs de um bebê, uma criança e um adulto no plano horizontal e mediano



Fonte: FELS, 2008.

REFERÊNCIAS

- ALARCÃO, D.; BENTO COELHO, J.L.; TENENBAUM, R.A. On modeling of room acoustics by a sound energy transition approach. In: EEA SYMPOSIUM ON ARCHITECTURAL ACOUSTICS. *Proceedings of EEA Symposium on Architectural Acoustics*. [S.l.], 2000.
- ALCAIM, A.; OLIVEIRA, C.A.S. *Fundamentos do processamento de sinais de voz e imagem*. 1. ed. Rio de Janeiro: Interciência, 2011.
- ALGAZI, V.R.; DUDA, R.O. Approximating the head-related transfer function using simple geometric models of the head and torso. *Journal of Acoustical Society of America*, v. 112, n. 5, p. 2053, 2002.
- ALGAZI, V.R. et al. The CIPIC HRTF database. In: IEEE WORKSHOP ON APPLICATIONS OF SIGNAL PROCESSING TO AUDIO AND ACOUSTICS. *Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New York: IEEE, 2001. p. 99–102.
- ALLEN, J.B.; BERKLEY, D.A. Image method for efficiently simulating small-room acoustics. *Journal of Acoustical Society of America*, v. 65, p. 943, 1979.
- ANDO, Y. *Concert Hall Acoustics*. Berlin: Springer-Verlag, 2012. (Springer Series in Electronics and Photonics).
- AZEVEDO, F.A.C. et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, v. 513, n. 5, p. 532–541, 2009.
- BAI, M.R.; TSAO, T.C. Numerical modeling of head-related transfer functions using the boundary source representation. *J. Vib. Acoust.*, v. 128, p. 594–603, 2006.
- BARRON, M.; LEE, L.-J. Energy relations in concert auditoriums. *Journal of Acoustical Society of America*, v. 84, p. 618, 1988.
- BASS, H.; BAUER, H.-J.; EVANS, L. Atmospheric absorption of sound: Analytical expressions. *Journal of Acoustical Society of America*, v. 52, p. 821–825, 1972.
- BATTITI, R. First- and second-order methods for learning: Between steepest descent and newton's method. *Neural Computation*, v. 4, n. 2, p. 141–166, mar. 1992.
- BERANEK, L.L. *Acoustics*. New York: McGraw-Hill, 1954.
- _____. *Concert hall and opera houses: music, acoustics, and architecture*. 2. ed. New York: Springer-Berlag, 2004.
- BERKHOUT, A.J.; VRIES, D.; BOONE, M.M. A new method to acquire impulse responses in concert halls. *J. Acoust. Soc. Am.*, v. 68, n. 1, p. 179–183, 1980.
- BISHOP, C. *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press, 2005.

BISTAFA, S.R. *Acústica aplicada ao controle do ruído*. 2. ed. São Paulo: Blücher, 2011. 380 p.

BISTAFA, S.R.; BRADLEY, J.S. Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics. *Journal of Acoustical Society of America*, v. 107, n. 2, p. 861–875, 2000.

BLAUERT, J. *Spatial Hearing*. Cambridge: The MIT Press, 1997.

BLOMMER, M.A.; WAKEFIELD, G.H. Pole-zero approximations for head-related transfer functions using a logarithmic error criterion. *IEEE Transactions on Speech and Audio Processing*, v. 5, n. 3, p. 278–287, 1997.

BOMHARDT, R. *Anthropometric Individualization of Head-Related Transfer Functions Analysis and Modeling*. Berlin: Logos Verlag Berlin, 2017. (Aachener Beiträge zur Akustik).

BORISH, J. Extension of the image model to arbitrary polyhedra. *Journal of Acoustical Society of America*, v. 75, p. 1827, 1984.

BORK, I. A comparison of room simulation software - the 2nd round robin on room acoustical computer simulation. *Acta Acustica United with Acustica*, v. 84, p. 943, 2000.

_____. Report on the 3rd round robin on room acoustical computer simulation - part i: Measurements. *Acta Acustica United with Acustica*, v. 91, n. 4, p. 740, 2005a.

_____. Report on the 3rd round robin on room acoustical computer simulation - part ii: Calculations. *Acta Acustica United with Acustica*, v. 91, n. 4, p. 753–763, 2005b.

BRADLEY, J.; SOULODRE, G.A. Objective measures of listener envelopment. *Journal of Acoustical Society of America*, v. 98, p. 2590, 1995.

BRANDÃO, E. *Acústica de salas: projeto e modelagem*. 1. ed. São Paulo: Blücher, 2016.

BROOMHEAD, D.S.; LOWE, D. Multivariable functional interpolation and adaptive networks. *Complex Systems*, v. 2, p. 321–355, 1988.

BRUNEAU, M. *Fundamentals of Acoustics*. London: Wiley-ISTE, 2006.

CAMILO, T.S. *Método Híbrido para Simulação Numérica de Acústica de Salas: Combinação dos Métodos de Traçado de Raios e Transição de Energia*. 2003. Dissertação (Mestrado em Engenharia Mecânica) — Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2003.

CARPENTER, G.A.; GROSSBERG, S. The ART of adaptive pattern recognition by a self-organizing neural network. *Computer*, n. 21, p. 77–88, 1988.

CHAPRA, S.C.; CANALE, R.P. *Numerical Methods for Engineers*. 6. ed. New York: McGraw-Hill, 2010.

CHEN, J.; VEEN, B.D.V.; HECOX, K.E. A spatial feature extraction and regularization model for the head-related transfer function. *Journal of Acoustical Society of America*, v. 97, n. 1, p. 439–452, 1995.

- CHEN, M.C.; HSIEH, S.F. Common acoustical-poles/zeros modeling for 3D sound processing. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*. Istanbul: IEEE, 2000.
- CHEN, S.; MULDREW, B.; MCLAUGHLIN, S. Adaptive bayesian feedback equalizer based on a radial basis function network. *IEEE International Conference on Communications*, Chicago, v. 3, p. 1267–1271, 1992.
- CHEUNG, N.M.; TRAUTMANN, S.; HORNER, A. Head-related transfer function modeling in 3-d sound systems with genetic algorithms. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*. Seattle: IEEE, 1998. p. 3529–3532.
- CHOI, S. et al. Indoor environmental quality of classrooms and student outcomes: A path analysis approach. *Journal of Learning Spaces*, v. 2, n. 2, p. 1–14, 2014.
- CONSTAN, Z.A.; HARTMANN, W.M. On the detection of dispersion in the head-related transfer function. *Journal of Acoustical Society of America*, v. 114, n. 2, p. 998–1008, 2003.
- CONTE, S.D.; BOOR, C. de. *Elementary numerical analysis: An algorithmic approach*. 3. ed. New York: McGraw-Hill, 1980.
- COX, T.J. et al. A tutorial on scattering and diffusion coefficients for room acoustic surfaces. *Acta Acustica united with Acustica*, v. 92, n. 1, p. 1–15, 2006.
- COX, T.J.; D'ANTONIO, P. *Acoustic absorbers and diffusers: Theory, design and application*. 1. ed. New York: Spon, 2004.
- DALENBACK, B.; KLEINER, M.; SVENSSON, P. A macroscopic view of diffuse reflection. *J. Audio Eng. Soc*, v. 42, p. 793–807, 1994.
- DANCE, S.M.; SHIELD, B.M. The complete image-source method for the prediction of sound distribution in non-diffuse enclosed spaces. *Journal of Sound and Vibration*, v. 201, n. 4, p. 473–489, 1997.
- D'ANTONIO, P.; COX, T.J. Two decades of sound diffuser design and development part 1: Application and design. *Journal of Audio Engineering Society*, v. 46, n. 11, p. 955–976, 1998a.
- _____. Two decades of sound diffuser design and development part 2: Prediction measurement and characterization. *Journal of Audio Engineering Society*, v. 46, n. 11, p. 1075–1091, 1998b.
- _____. Diffuser application in rooms. *Applied Acoustics*, v. 60, p. 113–142, 2000.
- DUDA, R.O.; HART, P.E. *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- DURANT, E.A.; WAKEFIELD, G.H. Efficient model fitting using a genetic algorithm: pole-zero approximations of hrtfs. *IEEE Transactions on Speech and Audio Processing*, v. 10, n. 1, p. 18–27, 2002.

- EMBRECHTS, J.J. Randomly traced sound ray techniques. *Acustica*, v. 51, p. 285–295, 1982.
- EVEREST, F.A.; POHLMANN, K.C. *Master Handbook of Acoustics*. 5. ed. New York: McGraw-Hill, 2009.
- EYRING, C.F. Reverberation time in “dead” rooms. *Journal of Acoustical Society of America*, v. 1, p. 168, 1930.
- FAHLMAN, S.E. *An Empirical Study of Learning Speed in Backpropagation Networks*. Technical Report, Carnegie Mellon University, Pittsburgh, 1988. 17 p.
- FARINA, A. RAMSETE - a new pyramid tracer for medium and large scale acoustic problems. In: EURONOISE. *Proceedings of the euronoise*. Lyon, France, 1995.
- FEDTKE, T. Kunstkopftechnik - eine bestandsaufnahme. *Acta Acustica/Acustica*, v. 93, n. 1, 2007.
- FELS, J. *From children to adults: how binaural cues and ear canal impedances grow*. Tese (Tese de Doutorado) — RWTH Aachen University, Aachen, Germany, 2008.
- FISHER, N.I.; LEWIS, T.; EMBLETON, B.J.J. *Statistical Analysis of Spherical Data*. New York: Cambridge University Press, 1987.
- FITZROY, D. Reverberation formula which seems to be more accurate with nonuniform distribution of absorption. *Journal of Acoustical Society of America*, v. 30, n. 7, p. 679, 1958.
- FREELAND, F.P.; BISCAINHO, L.W.P.; DINIZ, P.S.R. Interpositional transfer function for 3d-sound generation. *J. Audio Eng. Soc.*, v. 52, n. 9, p. 915–930, 2004.
- GARDNER, B.; MARTIN, K. HRTF measurements of a kemar dummy-head microphone. *Journal of Acoustical Society of America*, v. 97, n. 6, p. 3907–3908, 1995.
- GENARO, N. et al. A neural network based model for urban noise prediction. *J. Acoust. Soc. Am.*, v. 128, n. 4, p. 1738–1746, 2010.
- GRIJALVA, F.; MARTINI, L.; FLORENCIO, D. A manifold learning approach for personalizing hrtfs from anthropometric features. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 24, n. 3, p. 559–570, 2016.
- GRINDLAY, G.; VASILESCU, M.A.O. A multilinear (tensor) framework for HRTF analysis and synthesis. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*. Honolulu, Hawaii: IEEE, 2007. v. 1, p. 161–164.
- GUMEROV, N.A.; DURAI SWAMI, R. *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*. [S.l.]: Elsevier Science, 2005. (Elsevier Series in Electromagnetism).
- GUMEROV, N.A. et al. Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. *Journal of Acoustical Society of America*, v. 127, n. 1, p. 370, 2010.
- HAAS, H. The influence of a single echo on the audibility of speech. *J. Audio Eng. Soc.*, v. 20, n. 2, p. 146–159, 1972.

HACIHABIBOGLU, H.; GUNEL, B.; MURTAGH, F. Wavelet-based spectral smoothing for head-related transfer function filter design. In: INTERNATIONAL CONFERENCE: VIRTUAL, SYNTHETIC, AND ENTERTAINMENT AUDIO, 22. *Proceedings of the 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*. [S.l.]: Audio Engineering Society Conference, 2002.

HAGAN, M.T.; MENHAJ, M.B. Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks*, v. 5, n. 6, p. 989–993, 1994.

HANEDA, Y. et al. Common-acoustical-pole and zero modeling of head-related transfer functions. *IEEE Transactions on Speech and Audio Processing*, v. 7, n. 2, p. 188–196, 1999.

HARRIS, C.M. Absorption of sound in air versus humidity and temperature. *Journal of Acoustical Society of America*, v. 40, p. 148, 1966.

HARTMANN, W.M. Localization of sound in rooms. *Journal of Acoustical Society of America*, v. 74, n. 5, p. 1380–1391, 1983.

HAYKIN, S. *Neural Networks*. 2. ed. New Jersey: Prentice Hall, 1999.

_____. *Neural Networks and Learning Machines*. 3. ed. New Jersey: Prentice Hall, 2009.

HE, X.; CAI, D.; PARTHA, N. Laplacian score for feature selection. In: INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS. *Proceedings of the 18th International Conference on Neural Information Processing Systems*. Vancouver, Canada, 2005. p. 507–514.

HEINZ, R. *Development and evaluation of computer-assisted methods for binaural room simulation*. Tese (Dissertação de Mestrado) — RWTH Aachen University, Aachen, Germany, 1994.

HU, H. et al. HRTF personalization based on artificial neural network in individual virtual auditory space. *Applied Acoustics*, v. 69, n. 2, p. 163–172, 2008.

HUANG, Q.; FANG, Y. Modeling personalized head-related impulse response using support vector regression. *J. Shanghai Univ.*, v. 13, n. 6, p. 428–432, 2009.

ISBERT, A.C. *Diseño acústico de espacios arquitectónicos*. 1. ed. Barcelona: UPC, 1998. (Politecnos: Àrea d'arquitectura i urbanisme).

ISO 3382. *Acoustics - measurement of of the reverberation time of rooms with reference to the other acoustical parameters*. Geneva, Switzerland, 1997.

ISO 9613-1. *Acoustic - Attenuation of sound during propagation outdoors - part 1: Calculation of the absorption of sound by the atmosphere*. Geneva, Switzerland, 1993.

JENISON, R.L. A spherical basis function neural network for pole-zero modeling of head-related transfer functions. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*. New Paltz, NY: IEEE, 1995.

JENISON, R.L.; FISSELL, K. A spherical basis function neural network for modeling auditory space. *Neural Computation*, v. 8, p. 115–128, 1996.

JEUB, M. *Aachen Impulse Response Database*. 2011. (<https://www.iks.rwth-aachen.de/en/research/tools-downloads/databases/aachen-impulse-response-database/>). Accessed on: 01-11-2017.

KAPRALOS, B.; MEKUZ, N. Application of dimensionality reduction techniques to HRTFs for interactive virtual environments. In: INTERNATIONAL CONFERENCE ON ADV. COMPUT. ENTERTAIN. TECHNOL. *Proceedings of the International Conference on Adv. Comput. Entertain. Technol.* Salzburg, Austria, 2007. p. 256–257.

KAPRALOS, B. et al. Dimensionality reduced HRTFs: a comparative study. In: INTERNATIONAL CONFERENCE ON ADV. COMPUT. ENTERTAIN. TECHNOL. *Proceedings of the International Conference on Adv. Comput. Entertain. Technol.* Yokohama, Japan, 2008. p. 59–62.

KATZ, B.F.G. Boundary element method calculation of individual head-related transfer function. i. rigid model calculation. *Journal of Acoustical Society of America*, v. 110, n. 5, p. 2440, 2001.

KENDALL, G.S. A 3-d sound primer: directional hearing and stereo reproduction. *Computer Music Journal*, v. 19, n. 4, p. 23–46, 1995.

KINSLER, L.E. et al. *Fundamentals of Acoustics*. 4. ed. New York: John Wiley & Sons, 2000.

KISTLER, D.J.; WIGHTMAN, F.L. A model of hrtfs based on principal component analysis and minimum-phase reconstruction. *Journal of Acoustical Society of America*, v. 91, n. 3, p. 1637–1647, 1992.

KLEIN, J.; POLLOW, M.; VORLÄNDER, M. Optimized spherical sound source for auralization with arbitrary source directivity. In: EAA JOINT SYMPOSIUM ON AURALIZATION AND AMBISONICS. *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*. Berlin: Universitätsverlag der TU Berlin, 2014. p. 56–61.

KLEINER, M.; DALENBÄCK, B.I.; SVESSON, P. Auralization - an overview. *Journal of Audio Engineering Society*, v. 41, p. 861, 1993.

KOHONEN, T. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, n. 43, 1982.

KOOPMANN, G.H.; SONG, L.; FAHNLIN, J.B. A method for computing acoustic fields based on the principle of wave superposition. *Journal of Acoustical Society of America*, v. 86, p. 2433, 1989.

KROKSTAD, A.; STROM, S.; SORSDAL, S. Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, v. 8, n. 1, p. 118–125, 1968.

KRUZINS, E.; FRICKE, F. The prediction of sound fields in non-diffuse spaces by a “random walk” approach. *Journal of Sound and Vibration*, v. 81, n. 4, p. 549–564, 1982.

KULKARNI, A.; COLBURN, H.S. Infinite-impulse-response models of the head-related transfer function. *Journal of Acoustical Society of America*, v. 115, n. 4, p. 1714–1728, 2004.

KULOWSKI, A. Algorithmic representation of the ray tracing technique. *Applied Acoustics*, v. 18, p. 449–469, 1984.

KÜRER, R. Einfaches messverfahren zur bestimmung der schwerpunktzeit raumakustischer impulsantworten (a simple measuring procedure for determining the center time of room acoustical impulse responses). In: INTERNATIONAL CONGRESS ON ACOUSTICS. *Proceedings of the 7th International Congress on Acoustics*. Budapest, 1971.

KURZINS, E.; FRICKE, F.R. The prediction of sound fields in non-diffuse spaces by random walk approach. *Journal of Sound and Vibration*, v. 81, n. 4, p. 549–564, 1982.

KUTTRUFF, H. Simulated reverberation curves in rectangular rooms with diffuse sound fields. *Acustica*, v. 25, n. 6, p. 333–342, 1971.

_____. *Room Acoustics*. 5. ed. London: Spon Press, 2009.

LANGENDIJK, E.H.; BRONKHORST, A.W. Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *Journal of Acoustical Society of America*, v. 107, n. 1, p. 528–537, 2000.

LATHAUWER, L.; MOOR, B.; VANDEWALLE, J. A multilinear singular value decomposition. *SIAM Journal of Matrix Analysis and Applications*, v. 21, n. 4, p. 1253–1278, 2000.

LATHI, B.P. *Linear Systems and Signals*. 2. ed. New York: Oxford University Press, 2005.

LECUN, Y. *Efficient Learning and Second-order Methods*. Denver, 1993. A Tutorial at NIPS 93.

LEWERS, T. A combined beam tracing and radiant exchange computer model of room acoustics. *Applied Acoustics*, v. 38, p. 161–178, 1993.

LI, L.; HUANG, Q. HRTF personalization modeling based on rbf neural network. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*. Vancouver, Canada: IEEE, 2013. p. 3707–3710.

LIPPMANN, R.P. Pattern classification using neural networks. *IEEE Communications Magazine*, v. 27, p. 47–64, 1989.

LIU, C.J.; HSIEH, S.F. Common-acoustic-poles/zeros approximation of head-related transfer functions. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*. Salt Lake City, UT: IEEE, 2001.

MACPHERSON, E.A.; MIDDLEBROOKS, J.C. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *Journal of Acoustical Society of America*, v. 11, n. 5, p. 2219–2236, 2002.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. [S.l.], 1967. p. 281–297.

- MARQUARDT, D.W. An algorithm for least squares estimation of non-linear parameters. *Journal of the Society for Industrial and Applied Mathematics*, v. 11, p. 431–441, 1963.
- MASIERO, M.; POLLOW, M.; FELS, J. Design of a fast broadband individual head-related transfer function measurement system. In: *Forum Acusticum Aalborg*. Denmark: [s.n.], 2011.
- MCKEOWN, J.J.; STELLA, F.; HALL, G. Some numerical aspects of the training problem for feed-forward neural nets. *Neural Networks*, v. 10, n. 8, p. 1455–1463, 1997.
- MELO, V.S.G. *Avaliação de inteligibilidade em salas de aula do ensino fundamental a partir de respostas impulsivas biauriculares obtidas com cabeça artificial de dimensões infantis*. 2012. Tese (Doutorado em Engenharia Mecânica) — Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2012.
- MIDDLEBROOKS, J.C.; GREEN, D.M. Observations on a principal components analysis of head-related transfer functions. *Journal of Acoustical Society of America*, v. 92, p. 597, 1992.
- MILLINGTON, G. A modified formula for reverberation. *Journal of Acoustical Society of America*, v. 4, p. 69, 1932.
- MILLS, A.W. On the minimum audible angle. *Journal of Acoustical Society of America*, v. 30, p. 237–246, 1958.
- MØLLER, H. et al. Head-related transfer functions of human subjects. *J. Audio Eng. Soc.*, v. 43, n. 5, p. 300–321, 1995.
- MOLLER, M. Supervised learning on large redundant training sets. *International Journal of Neural System*, v. 4, n. 1, p. 15–25, 1993.
- MOODY, J.; DARKEN, C.J. Fast learning in networks of locally-tuned processing units. *Neural Computation*, v. 1, p. 281–294, 1989.
- MULGREW, B. Applying radial basis functions. *IEEE Signal Processing Magazine*, p. 50–65, 1996.
- MÜLLER, S.; MASSARANI, P. Transfer-function measurement with sweeps. *Journal of Audio Engineering Society*, v. 49, n. 6, p. 443–471, 2001.
- NAM, J.; KOLAR, M.A.; ABEL, J.S. On the minimum-phase nature of head-related transfer functions. In: *AES. the 125th Audio Engineering Society Convention*. San Francisco, CA, USA: AES, 2008.
- NANNARIELLO, J. The prediction of reverberation time using neural network analysis. *Applied Acoustics*, v. 58, p. 305–325, 1999.
- NANNARIELLO, J.; HODGSON, M.; FRICKE, F.R. Neural network predictions of speech levels in university classrooms. *Applied Acoustics*, v. 62, p. 749–767, 2001.
- NARANJO, J.F.L. *Inteligência computacional aplicada na geração de respostas impulsivas biauriculares e em auralização de salas*. 2014. 128 p. Tese (Doutorado em Modelagem Computacional) — Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2014.

NARANJO, J.F.L.; TENENBAUM, R.A.; TORRES, J.C.B. Realidade virtual acústica: a abordagem das redes neurais artificiais. In: XXV Encontro da Sociedade Brasileira de Acústica. Campinas, 2014. p. 360–367.

NISHINO, T. et al. Estimation of hrtfs on the horizontal plane using physical features. *Applied Acoustics*, v. 68, n. 8, p. 897–908, 2007.

NOSAL, E.-M.; HODGSON, M.; ASHDOWN, I. Improved algorithms and methods for room sound-field prediction by acoustical radiosity in arbitrary polyhedral rooms. *Journal of Acoustical Society of America*, v. 116, n. 2, p. 970–980, 2004.

NYQUIST, H. Certain topics in telegraph transmission theory. *Transactions AIEE*, v. 47, p. 617–644, 1928.

OPPENHEIM, A.V.; SCHAFER, R.W. *Discrete-Time Signal Processing*. 3. ed. New Jersey: Prentice-Hall, 2010.

OPPENHEIM, A.V.; WILLSKY, A.S. *Sinais e sistemas*. 2. ed. São Paulo: Prentice-Hall, 2010.

OTANI, M.; ISE, S. Fast calculation system specialized for head-related transfer function based on boundary element method. *Journal of Acoustical Society of America*, v. 119, n. 5, p. 2589, 2006.

PAUL, S. Binaural recording technology: A historical review and possible future developments. *Acta Acustica united with Acustica*, v. 95, p. 767–788, 2009.

PELZER, S.; ARETZ, M.; VORLÄNDER, M. Quality assessment of room acoustic simulation tools by comparing binaural measurements and simulations in an optimized test scenario. In: *Forum Acusticum Aalborg*. Denmark: [s.n.], 2011.

POLLOW, M.; BEHLER, G.K. Variable directivity for platonic sound sources based on spherical harmonics optimization. *Acta Acustica united with Acustica*, v. 95, n. 6, p. 1082–1092, 2009.

PREPELITAA, S. Influence of voxelization on finite difference time domain simulations of head-related transfer functions. *Journal of Acoustical Society of America*, v. 139, p. 2489, 2016.

PROAKIS, J.G.; MANOLAKIS, D.G. *Digital Signal Processing: Principles, algorithms, and applications*. 3. ed. New Jersey: Prentice-Hall, 1996.

RAHIM, M.G. et al. On the use of neural networks in articulatory speech synthesis. *J. Acoust. Soc. Am.*, v. 93, n. 2, p. 1109–1121, 1992.

RANDALL, R.B. *Application of B & K Equipment to Frequency analysis*. 2. ed. Naerum, Denmark: Bruel & Kjaer, 1977. 239 p.

RAYLEIGH, J.W.S.B. *The Theory of Sound*. [S.l.]: Macmillan, 1894. v. 1. (The Theory of Sound, v. 1).

REICHARDT, W.; Abdel Alim, O.; SCHMIDT, W. Abhängigkeit der grenzen zwischen brauchbarer und unbrauchbarer durchsichtigkeit von der art des musikmotives, der nachhallzeit und der nachhalleinsatzzeit. *Applied Acoustics*, v. 7, n. 4, p. 243–264, 1974.

RICHTER, J.G.; BEHLER, G.; FELLS, J. Evaluation of a fast HRTF measurement system. In: INTERNATIONAL AES CONVENTION. *Proceedings of the 140th International AES Convention*. Paris, France: AES, 2016.

RIEDMILLER, M. *Rprop - Description and Implementation Details*. Technical Report, University of Karlsruhe, Karlsruhe, 1994. 2 p.

RINDEL, J.H. The use computer modeling in room acoustics. *Journal of Vibroengineering*, v. 4, n. 3, p. 41–72, 2000.

RINDEL, J.H.; CHRISTENSEN, C.L. Room acoustic simulation and auralization - how close can we get to the real room. *WESPAC 8*, 2003.

RONSSSE, L.M. *Investigations of the relationships between unoccupied classroom acoustical conditions and elementary student achievement*. Tese (Architectural Engineering) — University of Nebraska, Lincoln, 2011.

RUMELHART, D.E.; MCCLELLAND, J.L. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*. Cambridge: The MIT Press, 1986. 567 p.

RUMSEY, F. *Spatial Audio*. 2. ed. Oxford: Focal Press, 2001.

SABINE, W. *Collected papers on acoustics*. Cambridge: Harvard university press, 1922.

SABINE, W.C. *Collected Papers on Acoustics*. [S.l.]: Nabu Press, 2011.

SAVIOJA, L.; SVENSSON, U.P. Overview of geometrical room acoustic modeling techniques. *Journal of Acoustical Society of America*, v. 138, n. 2, p. 708–730, 2015.

SCARPACI, J.W.; COLBURN, H.S. Principal components analysis interpolation of HRTFs using locally chosen basis functions. In: ACOUSTICAL SOCIETY OF AMERICA. *Proceedings of the 149th Meeting: Acoustical Society of America*. Vancouver, 2005.

SCHUSTER, K.; WAETZMANN, E. Über den nachhall in geschlossenen räumen ("about the reverberation in closed rooms"). *Annalen der physik*, v. 393, n. 5, p. 671–695, 1929.

SEEP, B. Acústica das salas de aula: um recurso para criar ambientes de aprendizado com condições desejáveis de audibilidade. *Acústica e Vibrações*, v. 29, 2002.

SETTE, W.J. A new reverberation time formula. *Journal of Acoustical Society of America*, v. 4, n. 3, p. 193, 1933.

SHAW, E.A.G. External ear response and sound localization. In: GATEHOUSE, R.W. (Ed.). *Localization of Sound: Theory and Applications*. Groton CT: Amphora Press, 1982. p. 30–41.

SHIN, K.; HAMMOND, J. *Fundamentals of signal processing for sound and vibration engineers*. Chichester: John Wiley & Sons, 2008.

STAN, G.B.; EMBRECHTS, J.J.; ARCHAMBEAU, D. Comparison of different impulse response measurement techniques. *Journal of Audio Engineering Society*, v. 50, n. 4, p. 249–262, 2002.

STEINBERG, B.Z. et al. A neural network approach to source localization. *J. Acoust. Soc. Am.*, v. 90, n. 4, p. 2081–2090, 1991.

STEPHENSON, U.M. Quantized pyramidal beam tracing - a new algorithm for room acoustics and noise immission prognosis. *Acta Acustica*, v. 82, p. 517–525, 1996.

STOER, J.; BULIRSCH. *Introduction to numerical analysis*. 3. ed. New York: Springer-Verlag, 2000.

TENENBAUM, R.A. Realidade virtual acústica. Disponível em <http://www.lidav.iprj.uerj.br/cursos>. 2016.

_____. Inteligência artificial aplicada à geração de realidade virtual acústica. Palestra proferida no Centro de Tecnologia da Universidade Federal de Santa Maria, 2017.

TENENBAUM, R.A. et al. Hybrid method for numerical simulation of room acoustic: Part 1 - theoretical and numerical aspects. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, v. 29, n. 2, p. 211–221, 2006a.

_____. Hybrid method for numerical simulation of room acoustic: Part 2 - validation of the computational code RAIOS 3. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, v. 29, n. 2, p. 223–231, 2006b.

_____. Auralization generated by modeling HRIRs with artificial neural networks and its validation using articulation tests. *Applied Acoustics*, v. 130, p. 260–269, 2018.

THIELE, R. Richtungsverteilung und zeitfolge der schallruckwurfe in raumen. *Acustica*, v. 3, p. 291–302, 1953.

TORRES, J.C.B; PETRAGLIA, M.R.; TENENBAUM, R.A. An efficient wavelet-based HRTF model for auralization. *Acustica united with Acta Acustica*, v. 90, n. 1, p. 108–120, 2004.

VENTURI, J.B. Considérations sur la connaissance de l'étendue que nous donne le sens de l'ouïe. *Journal des Sciences, des Lettres et des Arts*, v. 3, p. 29–37, 1796.

VIAN, J.-P; MARTIN, J. Binaural room acoustics simulation: Practical uses and applications. *Applied Acoustics*, v. 36, p. 293–305, 1992.

VORLÄNDER, M. International round robin on room acoustical computer simulation. In: INTERNATIONAL CONGRESS ON ACOUSTICS. *Proceedings of the 15th International Congress on Acoustics*. Trondheim, 1995.

_____. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Berlin: Springer-Verlag, 2008.

VORLÄNDER, M.; BIETZ, H. Comparison of methods for measuring reverberation time. *Acta Acustica*, v. 80, n. 3, p. 205–215, 1994.

VORLÄNDER, M.; MOMMERTZ, E. Definition and measurement of random-incidence scattering coefficients. *Applied Acoustics*, v. 60, p. 187–199, 2000.

WELLS, W.C. *An Essay upon Single Vision with two Eyes: together with Experiments and Observations on several other Subjects in Optics*. London: Cadell, 1792.

WERSENYI, G. Effect of emulated head-tracking for reducing localization errors in virtual audio simulation. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 17, n. 2, p. 247–252, 2009.

WILLIAMS, E.G. *Fourier Acoustics: Sound Radiation and Near field Acoustical Holography*. New York: Academic Press, 1999.

WU, Z. et al. Neural network model of binaural hearing based on spatial feature extraction of the head related transfer function. In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY. *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Hong Kong, China, 1998. v. 20, n. 3, p. 1109–1112.

XIANG, N.; BLAUERT, J. Binaural scale modeling for auralization and prediction of acoustics in auditoria. *Applied Acoustics*, v. 38, p. 267–290, 1993.

XIAO, T.; LIU, Q.H. Finite difference computation of head-related transfer function for human hearing. *Journal of Acoustical Society of America*, v. 113, n. 5, p. 2434, 2003.

XIE, B.; ZHONG, X.; HE, N. Typical data and cluster analysis on head-related transfer functions. *Applied Acoustics*, v. 94, p. 1–13, 2015.

YU, L.; KANG, J. Neural network analysis of soundscape in urban open spaces. In: ACOUSTICAL SOCIETY OF AMERICA. *The 149th Meeting of the Acoustical Society of America*. [S.l.], 2005.

_____. Modeling subjective evaluation of soundscape quality in urban open spaces: An artificial neural network approach. *J. Acoust. Soc. Am.*, v. 126, n. 3, p. 1163–1174, 2009.

ZAGAR, L.E. The use of the image source method for modeling room acoustics. *Journal of Acoustical Society of America*, v. 74, p. 1914, 1983.

ZOTKIN, D.N. et al. Fast head-related transfer function measurement via reciprocity. *Journal of Acoustical Society of America*, v. 120, n. 4, p. 2202–2215, 2006.

APÊNDICE – Cálculo da função exponencial

Dado que a função de ativação dos neurônios da camada intermediária das redes RBF é do tipo gaussiana (Eq. (43)), a qual é dada pela função exponencial (função transcendental), esta pode ser obtida a partir de uma série de Taylor truncada. À vista disso, neste apêndice, discute-se a quantidade de termos necessária da tal série, de modo que o erro cometido com essa truncagem seja desprezível. Também é apresentado um algoritmo eficiente para calcular essa série de Taylor truncada.

A.1 Truncamento da expansão em série de Taylor da exponencial

A fim de examinar a quantidade necessária de termos da série, considere o seguinte teorema, o qual permite estimar o tamanho do erro gerado após o truncamento da série de Taylor.

Teorema A.1.1. *Sejam*

$$\Psi_n(v) = \sum_{i=0}^n \frac{\psi^{(i)}(0)}{i!} x^i \quad (123)$$

o n -ésimo polinômio de Taylor de ψ centrado em $v = 0$, e

$$\psi(v) = \sum_{i=0}^{\infty} \frac{\psi^{(i)}(0)}{i!} x^i \quad (124)$$

a série de Taylor de ψ em torno de $v = 0$. Suponha que a derivada $\psi^{(n+1)}(v)$ exista e seja contínua. Se L é uma constante tal que $|\psi^{(n+1)}(v)| \leq L$ para todo v entre 0 e v , então

$$|\psi(v) - \Psi_n(v)| \leq L \frac{|v|^{n+1}}{(n+1)!}. \quad (125)$$

A série de Taylor de $\exp(v^2)$ é obtida substituindo v por $-v^2$ na série de Taylor de $\exp(v)$, isto é,

$$\psi(v) = \exp(-v^2) = \sum_{i=0}^{\infty} \frac{(-1)^i v^{2i}}{i!} = 1 - v^2 + \frac{v^4}{2!} - \frac{v^6}{3!} + \frac{v^8}{4!} - \dots \quad (126)$$

Uma vez que a expansão de Taylor de $\exp(v)$ é válida para todo v , essa expansão também é válida para todo v . Sendo assim, o valor absoluto da i -ésima derivada de $\psi(v)$ é limitada para todo v .

A partir do Teorema A.1.1, tem-se que, mantendo n fixo, o erro da estimativa de $\psi(v)$ aumenta à medida que se afasta de $v = 0$. Portanto, para diminuir o erro desses valores de v é necessário aumentar o número de termos, n , da série. Agora, como os valores de v relevantes estão no intervalo $[-3, 3]$, pois os valores fora desse intervalo são praticamente nulos (ver Figura 36), basta analisar o erro para $v = 3$ (ou $v = -3$). Fazendo $n = 70$ (ou seja, considerando os 36 primeiros termos da série), tem-se $\max_{0 \leq v \leq 3} \{|\psi^{(71)}(v)|\} \approx 1.3648 \times 10^{61}$. Portanto, pode-se aplicar a estimativa do erro com $L = 1.3648 \times 10^{61}$ para obter

$$|\psi(v) - \Psi_{70}(v)| \leq 1.3648 \times 10^{61} \frac{|3|^{71}}{71!} \approx 1.2051 \times 10^{-7} < 10^{-6} \quad (127)$$

Portanto, a avaliação dos valores de $\psi(v)$ para $v \in [-3, 3]$ utilizando o polinômio de Taylor de ordem 70 (Eq. (128)) obtém erros menores do que 10^{-6} .

$$\Psi_{70}(v) = \sum_{i=0}^{35} \frac{(-1)^i v^{2i}}{i!} \quad (128)$$

A.2 Método eficiente para o cálculo da série de Taylor

Nesta seção apresenta-se o algoritmo Horner que permite calcular o polinômio de Taylor de maneira otimizada. Primeiramente, observe que a Eq. (123) pode ser reescrita como

$$\Psi_n(v) = a_0 + v(a_1 + v(a_2 + \dots + v(a_{n-1} + a_n v) \dots)), \quad (129)$$

onde $a_i = \psi^{(i)}(0)/i!$, para $i = 1, 2, \dots, n$. Logo, $\Psi_n(v)$ pode ser obtido a partir da seguinte relação de recursão

$$b_{i-1} = a_{i-1} + b_i v, \quad (130)$$

com $b_n = a_n$, para $1 \leq i \leq n$, de modo que $\Psi_n(v) = b_0$. De fato, substituindo iterativamente b_i na Eq. (129), tem-se

$$\begin{aligned} \Psi_n(v) &= a_0 + v(a_1 + v(a_2 + \dots + v(a_{n-1} + b_n v) \dots)) \\ &= a_0 + v(a_1 + v(a_2 + \dots + v(b_{n-1}) \dots)) \\ &\quad \vdots \\ &= a_0 + b_0 v \\ &= b_0 \end{aligned} \quad (131)$$

Esse algoritmo requer n somas e n multiplicações, caso todos os coeficientes sejam não-nulos. Como o polinômio de Taylor relativo à função de ativação possui apenas os coeficientes pares não-nulos, a Eq. (128) pode ser reescrita da seguinte forma

$$\Psi_{70}(v) = a_0 + v^2(a_2 + v^2(a_4 + \dots + v^2(a_{68} + a_{70}v^2)\dots)), \quad (132)$$

onde $a_i = \frac{(-1)^i}{i!}$. A avaliação desse polinômio aplicando a regra de Horner efetua 35 somas e 36 multiplicações.