

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA**

Diego Marques do Carmo

**ALGORITMO ADAPTATIVO DE REDUÇÃO DE RUÍDO
E PRESERVAÇÃO DAS DIFERENÇAS DE NÍVEL
INTERAURAL PARA APARELHOS AUDITIVOS**

Florianópolis

2017

Diego Marques do Carmo

**ALGORITMO ADAPTATIVO DE REDUÇÃO DE RUÍDO
E PRESERVAÇÃO DAS DIFERENÇAS DE NÍVEL
INTERAURAL PARA APARELHOS AUDITIVOS**

Dissertação submetida ao Programa de
Pós-graduação em Engenharia Elétrica
da Universidade Federal de Santa Catarina
para a obtenção do Grau de Mestre
em Engenharia Elétrica.
Orientador: Prof. Dr. Márcio Hols-
bach Costa

Florianópolis

2017

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

do Carmo, Diego Marques

Algoritmo adaptativo de redução de ruído e preservação das diferenças de nível interaural para aparelhos auditivos / Diego Marques do Carmo ; orientador, Márcio Holsbach Costa, 2017.

110 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós Graduação em Engenharia Elétrica, Florianópolis, 2017.

Inclui referências.

1. Engenharia Elétrica. 2. Aparelhos auditivos binaurais. 3. algoritmo adaptativo. 4. redução de ruído. 5. preservação da ILD. I. Holsbach Costa, Márcio. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

ALGORITMO ADAPTATIVO DE REDUÇÃO DE RUÍDO E PRESERVAÇÃO DAS DIFERENÇAS DE NÍVEL INTERAURAL PARA APARELHOS AUDITIVOS

Diego Marques do Carmo

Esta Dissertação foi julgada aprovada para a obtenção do Título de “Mestre em Engenharia Elétrica”, e aprovada em sua forma final pelo Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina.

Florianópolis, 23 de fevereiro 2017.

Prof. Dr. Marcelo Lobo Heldwein
Coordenador

Prof. Dr. Márcio Holsbach Costa
Orientador

Banca Examinadora:

Eduardo Luiz Ortiz Batista, Dr.

Jefferson Luiz Brum Marques, Ph.D.

José Carlos Moreira Bermudez, Ph.D.

Stephan Paul, Dr.

A meus pais, Helino e Roseth, pelo apoio incondicional.

AGRADECIMENTOS

Agradeço, primeiramente, aos meus pais, Helino e Roseth, por acreditarem e incentivarem os meus sonhos. A vocês, toda minha admiração pelos anos de batalha que culminaram nesse trabalho.

Ao meu orientador, Prof. Márcio H. Costa, por me transmitir um pouco de seu conhecimento e sabedoria através de sua competente e amigável orientação.

A minha namorada, Maria Negri, pois um pouco de amor sempre ajuda na caminhada.

Aos meus colegas do LPDS, Sala da Pós-Graduação e de outros laboratórios pelo companheirismo e convivência essenciais no dia-a-dia. Em especial, aos amigos Fábio Itturiet e Ricardo Borsoi por fornecerem inúmeras vezes suas orelhas para opinar sobre áudios de ruído.

A todas que contribuíram com a escrita deste texto: Maria Negri, Jaisa Araújo, Rafaela Martins e Amália Honda.

E ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro.

Somewhere, someone is asking a question,
and I stand squinting at the classroom
with one hand cupped behind my ear,
trying to figure out where that voice is
coming from.

Tony Hoagland

Resumo

Aparelhos auditivos binaurais consistem em um par de aparelhos que compartilham e processam os sinais sonoros que recebem em seus microfones. Decorrente disto, estes dispositivos podem explorar pistas binaurais de localização das fontes acústicas. Isto permite o desenvolvimento de técnicas de redução de ruído para processar os sinais recebidos com o intuito de criarem uma percepção mais verossímil do cenário acústico para o usuário de aparelhos auditivos. Neste sentido, uma técnica bastante utilizada em aparelhos binaurais é o filtro de Wiener multicanal (MWF—Multichannel Wiener filter). Os coeficientes do MWF determinam um estimador linear para o sinal desejado de fala. Este estimador preserva as pistas binaurais de localização da fala e altera, contudo, as do ruído aditivo. Por esta razão, várias técnicas foram desenvolvidas com a finalidade de preservar as pistas binaurais do ruído. Algumas destas técnicas baseiam-se na extensão da função custo do erro quadrático médio através de um termo derivado a partir da definição matemática da pista de localização binaural denominada diferença de intensidade interaural (ILD—*Interaural Level Difference*). Entretanto, a determinação dos coeficientes ótimos que minimizam estas funções custo requer o uso de técnicas de otimização numérica, o que pode acarretar em tempos de processamento e/ou recursos computacionais inadmissíveis em aplicações em tempo real. Diante disto, este trabalho propõe um algoritmo adaptativo para redução de ruído e preservação da ILD do sinal de ruído. O método proposto determina os coeficientes dos filtros lineares de forma iterativa, o que representa uma mudança de paradigma e uma redução de custo computacional em relação à técnicas anteriores. Para os cenários acústicos avaliados, o algoritmo desenvolvido diminui em 6,96 dB a métrica de preservação da ILD do ruído, tendo um aumento de apenas 0,98 dB na métrica de preservação da ILD da fala em relação ao MWF. As métricas de qualidade de fala, redução de ruído e inteligibilidade se mantêm aproximadamente as mesmas no ouvido mais distante da fonte de ruído. Adicionalmente, resultados subjetivos mostram que o método proposto preserva a percepção da localização da fonte de ruído próximo de sua posição original.

Palavras-chave: Aparelhos binaurais, algoritmo adaptativo, redução de ruído, preservação da ILD.

Abstract

Binaural hearing aids are two devices which share and process the signals received in their microphones. Consequently, these devices can explore binaural localisation cues. This results in the development of noise reduction techniques which process the received signals to create a reliable perception of the acoustic scenario for the hearing impaired person. In this sense, the techniques based on the multichannel Wiener filter (MWF) are extensively used in applications of binaural hearing aids. The MWF coefficients produce a linear estimator of the desired speech component. This estimator preserves the speech binaural cues, and shifts the cues of the additive noise. Because of this, many techniques have been developed aiming to preserve the noise localisation cues. Some of that are based in the extension of the mean squared error cost function using a perceptually motivated term based on the mathematical expression of a binaural localization cue called interaural level difference (ILD). However, the technique uses numerical optimization approaches to calculate the coefficients of linear noise reduction filters. Considering a real time application, this may result in unacceptable processing time and/or computational resources. For this reason, this work presents an adaptive algorithm for both noise reduction and preservation of noise ILDs. The proposed method calculates de linear filters using recursive equation which results in a reduction of the computational cost with respect to the previews techniques. In the assessed acoustic scenario, the method leads to a reduction of the noise ILD preservation measurement of up to 6.96 dB, and to an increase of just 0.98 dB in speech ILD preservation measurement, with respect to the MWF. The measures of speech quality, noise reduction and intelligibility are kept approximately the same in the ear in the opposite side of the noise source. In addition, subjective experiments show the proposed technique preserves the noise position perception near of the original noise source.

Keywords: Binaural hering aids, adaptive algorithms, noise reduction, ILD preservation.

Lista de Abreviaturas e Siglas

AI	Articulation index Índice de articulação
BIF	Band-importance function Função de importância de banda
CEPSH	Comitê de ética em pesquisas com seres humanos
HRTF	Head related transfer function Função de transferência relacionada à cabeça
ILD	Interaural level difference Diferença de intensidade interaural
ITD	Interaural time difference Diferença de tempo interaural
ISNR	Intelligibility-weighted signal to noise
L	Left Esquerda
MAE	Mean absolute error Erro absoluto médio
MAC	Multiplica-acumula
MOS	Mean opinion score Pontuação média de opinião
MWF	Multichannel Wiener filter Filtro de Wiener multicanal
N	Noise Ruído
NP	Não-processado
PESQ	Perceptual evaluation of speech quality Avaliação perceptual da qualidade de fala
R	Right Direita
RIQ	Região inter-quartil
S	Speech Fala

STI	Speech-transmission index Índice de transmissão de fala
SNR	Signal to noise ratio Relação sinal ruído
STFT	Short-time Fourier transform Transformada de Fourier de tempo curto
VAD	Voice activity detector Detector de fala

Sumário

1	INTRODUÇÃO	19
1.1	Preâmbulo	19
1.2	Objetivos	21
1.3	Estrutura do Trabalho	21
2	FUNDAMENTAÇÃO TEÓRICA	23
2.1	Anatomia do Sistema Auditivo Humano	23
2.2	Sistema de Coordenadas Espaciais	26
2.3	Região de Campo Distante	26
2.4	Audição Espacial	28
2.5	Localização e Lateralização	28
2.6	Métricas de Desempenho	29
2.6.1	Relação sinal ruído	29
2.6.2	Avaliação perceptual da qualidade de fala	29
2.6.3	Ganho ponderado de inteligibilidade	30
2.6.4	Erro na diferença de intensidade interaural	31
2.6.5	Erro absoluto médio	32
2.6.6	Diagrama de caixas – <i>Boxplot</i>	32
2.6.7	Teste de Wilcoxon	33
2.7	Conclusões	34
3	ABORDAGEM PROPOSTA	35
3.1	Sistema de Processamento Binaural	35
3.2	Função Custo Binaural	39
3.3	Método Proposto	41
3.3.1	Equações recursivas	41
3.3.2	Estimação das matrizes de correlação	43
3.3.3	Análise do custo computacional	45
3.4	Conclusões	46
4	MATERIAIS E MÉTODOS	49
4.1	Cenários Acústicos Simulados	49
4.1.1	Funções de transferência	49
4.1.2	Sinais acústicos	50
4.1.3	Configuração do cenário acústico	51
4.2	Métricas de Desempenho	53

4.3	Determinação de Parâmetros	54
4.3.1	Estimação das matrizes de correlação	54
4.3.2	Coefficiente de redução de ruído	55
4.3.3	Número de iterações	56
4.4	Conclusões	57
5	RESULTADOS	61
5.1	Avaliação 1: Métricas Objetivas	61
5.1.1	Resultados e discussão	63
5.2	Avaliação 2: Teste Subjetivo	68
5.2.1	Aspectos éticos	68
5.2.2	Ambiente acústico simulado	69
5.2.3	Metodologia de avaliação	71
5.2.4	Resultados e discussão	74
5.3	Avaliação 3: Rastreamento de Fonte de Ruído	86
5.3.1	Resultados e discussão	88
5.4	Conclusão	92
6	CONCLUSÃO	95
6.1	Propostas para Continuação do Trabalho	96
	REFERÊNCIAS	97
	APÊNDICE A – Derivação da função custo da J_{ILD}	103
	APÊNDICE B – Determinação do gradiente da J_{ILD} ...	105
	APÊNDICE C – Estimativas das matrizes de correlação	107
	ANEXO A – Termo de consentimento livre e esclarecido	109

Capítulo 1

Introdução

A audição apresenta um papel de extrema importância no cotidiano de qualquer pessoa. Ouvir é um dos elementos que possibilita a comunicação com os outros, permitindo-nos socializar com amigos, estar alerta a perigos e trabalhar de forma efetiva. Assim sendo, o declínio ou a perda da capacidade auditiva representa uma diminuição na qualidade de vida. Aparelhos auditivos são uma possibilidade para compensar o problema de perda auditiva. Restaurar a capacidade auditiva envolve não apenas o entendimento da informação em um meio ruidoso, mas também a capacidade de localizar fontes sonoras. Por isso, no caso de necessidade de compensação da audição em ambas as orelhas, aparelhos auditivos binaurais (que permitem a troca de informação entre os aparelhos em cada uma das orelhas) têm também como objetivo preservar pistas sonoras utilizadas naturalmente pelo sistemas auditivo central humano para localizar fontes sonoras, a fim de fornecer às pessoas que os utilizam uma percepção espacial verossímil dos sons em um ambiente.

1.1 PREÂMBULO

Os relatos históricos sobre aparelhos auditivos iniciam a partir de dispositivos incipientes utilizados entre os séculos XVII e XIX. Esses aparelhos tinham a forma de corneta e eram feitos a partir de chifres de animais, metal ou madeira. A primeira grande modificação tecnológica nos aparelhos auditivos ocorreu em 1898 com a invenção do aparelho auditivo elétrico (BAUMAN, 2013). Diferentemente dos modelos atuais, utilizados fixos na orelha, os aparelhos elétricos tinham o microfone e a bateria presos ao corpo do usuário e o alto-falante era pressionado contra o pavilhão auditivo. Ao longo dos séculos XX e XXI, os aparelhos auditivos acompanharam o desenvolvimento da tecnologia em diferentes áreas. Isso resultou em modificações no seu formato, tamanho e na maneira de utilizá-los. Além disso, houve também o aperfeiçoamento

e o acréscimo de diferentes funcionalidades, como por exemplo, a preservação e a ênfase da capacidade de localização de fontes sonoras.

Um par de aparelhos auditivos pode ser classificado como 1) bilateral: par de aparelhos que operam de forma independente um do outro; 2) binaural: par de aparelhos que operam de forma conjunta no processamento de sinais, i.e., compartilham e processam os sinais sonoros que recebem. Contudo, Bogaert et al. (2006) mostraram que a localização de fontes sonoras é melhor realizada quando se desliga os aparelhos bilaterais. Os aparelhos binaurais foram desenvolvidos para compartilhar e processar os sinais recebidos por seus microfones permitindo-os explorar estratégias de preservação da sensação subjetiva original da posição espacial das fontes sonoras. Em seres humanos, as principais pistas espaciais utilizadas são: a diferença de tempo interaural (ITD – *Interaural time difference*) e a diferença de nível interaural (ILD – *Interaural Level Difference*) (DILLON, 2001).

Ao longo dos últimos anos, várias técnicas de redução de ruído têm sido propostas para aparelhos auditivos binaurais. Dentre as principais está o filtro de Wiener multicanal (MWF – *Multichannel Wiener Filter*). Apesar desta técnica reduzir o ruído aditivo de forma eficiente, ela acarreta distorções nas pistas de localização espacial dos sinais envolvidos. Em situações caracterizadas por uma relação sinal ruído positiva, a percepção da posição da fonte de ruído pode ser alterada, diminuindo a separação espacial original entre a fonte de fala de interesse e a fonte de ruído (CORNELIS et al., 2010). Todavia, como mostrado por Peissig e Kollmeier (1997) e Hawley, Litovsky e Culling (2004), a separação espacial entre as fontes de fala e ruído contribui para uma melhoria na inteligibilidade. Por isso, várias técnicas foram propostas (DOCLO et al., 2005; BOGAERT et al., 2007; KLASSEN et al., 2007; CORNELIS et al., 2010; COSTA; NAYLOR, 2014) com o objetivo de reduzir ruído e preservar a localização da fonte sonora, através da preservação das pistas de localização espacial.

Por isso, com o intuito de preservar a percepção relacionada à localização da fonte de ruído, Doclo et al. (2005) propuseram uma extensão na função custo que determina o MWF através de um termo adicional associado à preservação da ILD. Os resultados mostraram que, em comparação ao MWF convencional, a técnica proposta mantém o valor da ILD do ruído processado próximo de seu valor original e aumenta a relação sinal ruído (SNR – *Signal to Noise Ratio*) em até 1,3 dB. Posteriormente, Costa e Naylor (2014) apresentaram uma nova contribuição através de uma modificação no termo de preservação da ILD. A partir disso, em relação à técnica de Doclo et al. (2005), houve

um aumento de até 9,6 dB na SNR e 1 PESQ (métrica de qualidade de fala), assumindo uma mesma distorção de ILD em ambas as orelhas para as duas técnicas. Entretanto, o cálculo dos coeficientes do filtro de redução de ruído em ambas as técnicas requer o uso de métodos numéricos que podem acarretar tempos de processamento inadmissíveis em aplicações em tempo real.

1.2 OBJETIVOS

A partir do exposto na seção anterior, os objetivos deste trabalho são:

- Propor um algoritmo adaptativo de custo computacional controlado para reduzir ruído e preservar a localização espacial de sua fonte.
- Avaliar o algoritmo proposto em cenários acústicos simulados e verificar seu desempenho através de métricas objetivas e de experimentos subjetivos.

1.3 ESTRUTURA DO TRABALHO

Estruturou-se esta dissertação da seguinte forma: no Capítulo 1, apresentou-se de forma introdutória o tema abordado e elencou-se os objetivos deste trabalho. O Capítulo 2 apresenta os fundamentos teóricos para compreensão dos capítulos subsequentes. O Capítulo 3 descreve, inicialmente, os sinais e sistemas que caracterizam os cenários acústicos de interesse e o sistema de processamento binaural, e, por fim, apresenta-se o método proposto. O Capítulo 4 descreve os experimentos realizados para avaliação do desempenho do algoritmo adaptativo proposto. O Capítulo 5 apresenta e discute os resultados dos experimentos realizados. Por fim, o Capítulo 6 apresenta as conclusões do trabalho e algumas propostas de trabalhos futuros relacionados ao tema desenvolvido.

Capítulo 2

Fundamentação Teórica

Esse capítulo introduz e define termos que serão utilizados ao longo do texto, bem como apresenta conceitos e definições que servirão de suporte para o entendimento do problema a ser tratado, do método que será proposto e das avaliações de desempenho que serão realizadas.

2.1 ANATOMIA DO SISTEMA AUDITIVO HUMANO

As ondas sonoras geradas ao nosso redor são captadas, processadas e decodificadas pelo sistema auditivo humano, Figura 1. Primeiramente, o som chega à porção inicial do sistema auditivo denominada pavilhão externo ou pina, , ilustrado na Figura 2, e pelo meato acústico externo ou canal auditivo. As ondas sonoras, ao percorrerem as elevações e depressões da pina, sofrem reflexões e atenuações que, ao serem processadas pelo sistema auditivo central, fornecem informações sobre a localização da fonte geradora do som. Após passarem pela pina, as ondas sonoras percorrem o canal auditivo e chegam até a porção inicial da orelha média.

A orelha média é composto, dentre outras partes, pela membrana timpânica (ou tímpano) e por três ossículos denominados martelo, bigorna e estribo. A Figura 3 apresenta a porção final da orelha externa e a anatomia da orelha média. As ondas sonoras, no final do canal auditivo, atingem o tímpano, fazendo com que ele vibre. Essas vibrações são transferidas para a cadeia dos três ossículos (martelo, bigorna e estribo). A partir dessa vibração da cadeia ossicular, as ondas sonoras são transferidas para a orelha interna. Dentre as funções da orelha média está a transferência eficiente das ondas sonoras que chegam do ambiente externo para as porções mais internas do sistema auditivo.

Na orelha interna, as ondas sonoras percorrem um tubo em forma de espiral (cóclea)(ver Figura 1) preenchido com um fluido (endolinfa). Ao longo de todo o comprimento da cóclea, células receptoras, denominadas células ciliadas, transformam as vibrações mecânicas (ondas

Figura 1 – Sistema auditivo humano. Referência: Netter (2011).

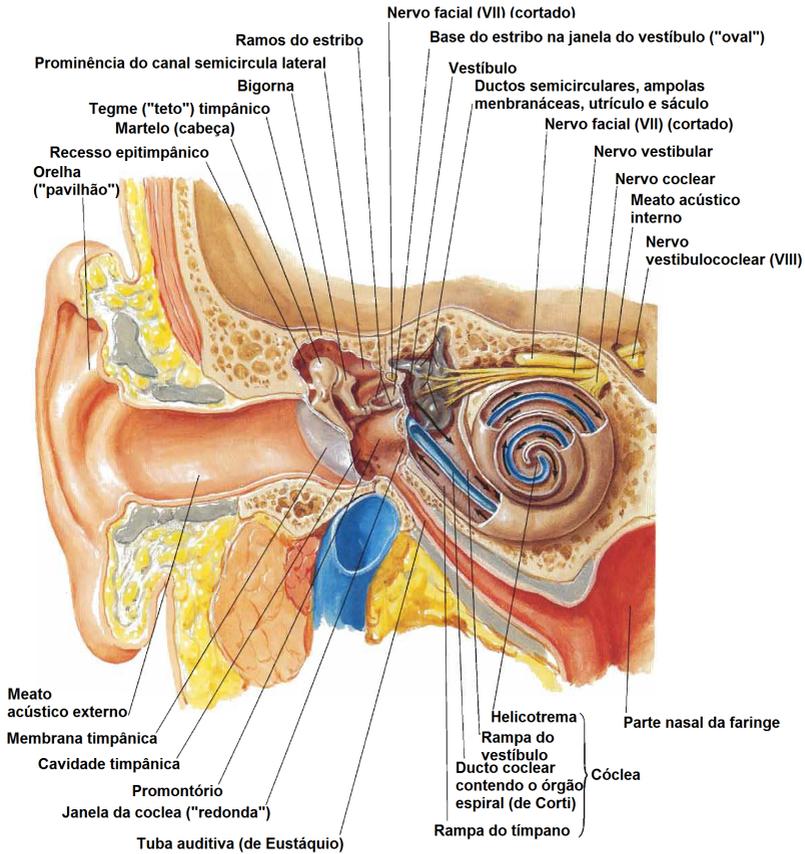
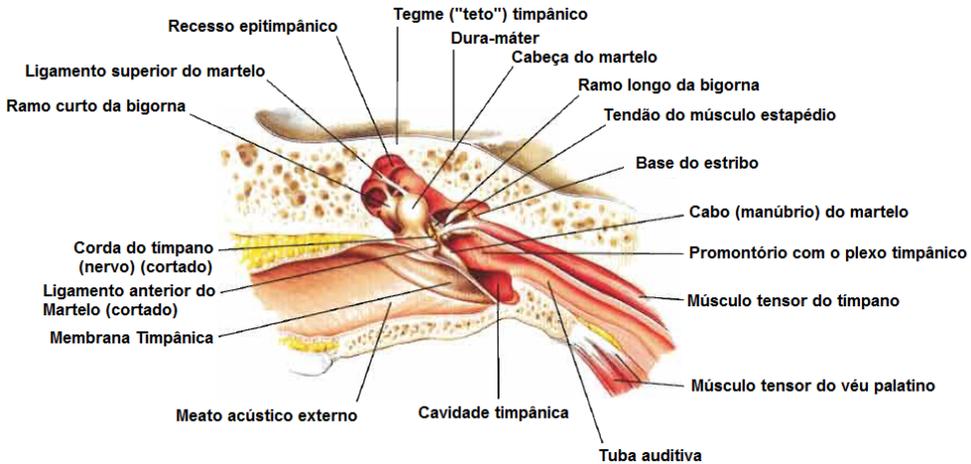


Figura 2 – Sistema auditivo humano. Referência: Netter (2011).



Figura 3 – Anatomia simplificada da orelha média: membrana timpânica, martelo, bigorna e estribo. Referência: Netter (2011).



sonoras), que se propagam através da endolinfa, em sinais elétricos, que são enviados ao cérebro pelo nervo auditivo. As células na porção inicial da cóclea, são responsáveis por captar vibrações mecânicas de alta frequência, enquanto que a porção apical ou final da cóclea, capta as vibrações mecânicas de baixa frequência.

2.2 SISTEMA DE COORDENADAS ESPACIAIS

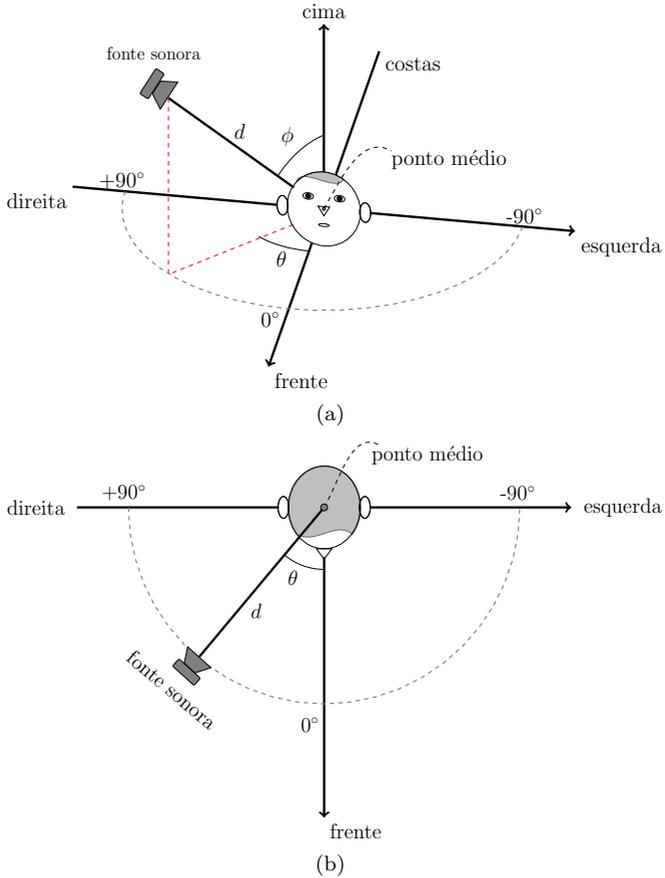
Convencionalmente, a posição de uma fonte sonora, tendo como referencial o ponto média da linha que conecta as entradas dos canais auditivos, é descrita através de coordenadas esféricas. Por isso, a posição da fonte é caracterizada pelos ângulos de azimute θ e de elevação ϕ e pela distância d da fonte sonora em relação ao ponto de referência, como ilustrado na Figura 4(a)¹. Nesse trabalho, considera-se que tais fontes sonoras são pontuais, estáticas e localizadas sobre o plano $\phi = 0$, denominado plano horizontal, apresentado na Figura 4(b), e, portanto, estão localizadas na mesma altura da linha que conecta a entrada dos canais auditivos do usuário de aparelho auditivo. O ponto médio dessa linha, utilizado como ponto de referência para a caracterização da posição das fontes acústicas, é destacado na Figura 4.

2.3 REGIÃO DE CAMPO DISTANTE

Uma fonte sonora pode se localizar próxima (na região de campo próximo) ou distante (na região de campo distante) do aparelho auditivo (DOCLO; MOONEN, 2003; STADLER; RABINOWITZ, 1993). Por caracterizarem várias situações reais de interesse e por resultar em facilidade de desenvolvimento, vários algoritmos de melhoramento de fala (CORNELIS et al., 2010; BAI; LIN, 2005; DOCLO; MOONEN, 2003; STADLER; RABINOWITZ, 1993) consideram que as fontes acústicas se localizam no campo distante do arranjo de microfones do aparelho auditivo. Isso possibilita considerar-se planas as frentes de onda sonora que chegam até os microfones. Uma regra utilizada para determinar se uma fonte sonora se localiza no campo distante de um arranjo de microfones

¹O sistema de coordenadas apresentado segue as convenções de ângulos utilizadas por Costa e Naylor (2014), Kayser et al. (2009), Doclo et al. (2009), Klasen et al. (2007)

Figura 4 – Localização espacial de uma fonte sonora em relação ao centro da cabeça de um humano: (a) azimute θ , elevação ϕ e distância d ; (b) plano horizontal, $\phi = 0$.



é dada pela seguinte expressão (DOCLO; MOONEN, 2003):

$$r > \frac{d_{\text{tot}}^2 f_s}{c}, \quad (2.1)$$

onde r é a posição radial da fonte sonora, d_{tot} é o comprimento total do arranjo de microfones, f_s é a frequência de amostragem e c é a velocidade do som, i.e., 340m/s.

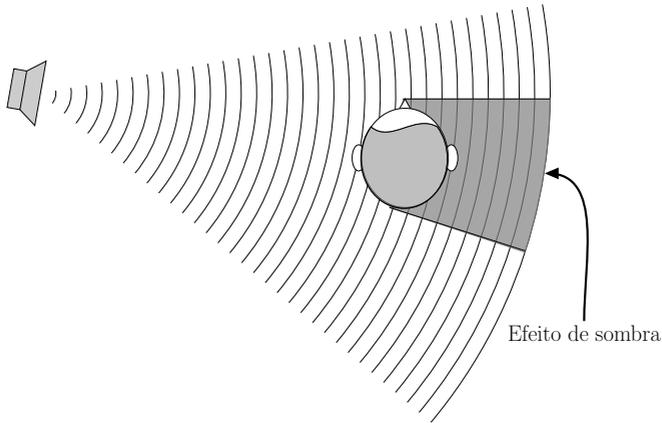
2.4 AUDIÇÃO ESPACIAL

De forma geral, a localização de uma fonte sonora pontual e estática, sobre o plano $\phi = 0$, é feita através de pistas de localização associadas às duas orelhas (pistas binaurais), quais sejam: diferença de tempo interaural (ITD) e diferença de nível interaural (ILD). A ITD é definida como a diferença de tempo de chegada, de uma mesma frente de onda, nas orelhas esquerda e direita. A distância entre as orelhas resulta numa diferença de tempo de chegada não nula para fontes sonoras que não se localizem imediatamente à frente ou atrás do ouvinte, i.e., $\theta = 0^\circ$ ou $\theta = 180^\circ$. A ILD resulta, principalmente, de atenuações acústicas da onda sonora causadas por refração e absorção na cabeça e no tronco, denominado de efeito de sombreamento, ilustrado na Figura 5. A ITD e a ILD são pistas predominantes em faixas de frequências diferente, a ITD é a pista de localização espacial predominante para sons com componentes em frequência abaixo 1,5 kHz, a ILD é predominante para sons com frequência acima de 1,5 kHz.

2.5 LOCALIZAÇÃO E LATERALIZAÇÃO

A localização de uma fonte sonora consiste na determinação sua posição em termos de sua direção e distância. Nesse caso, a localização é dita extracraniana, pois o ouvinte tem a percepção de que a fonte sonora se localiza externamente à sua cabeça. Ao utilizar fones de ouvido, contudo, um ouvinte frequentemente tem a percepção de que a fonte sonora se localiza dentro de sua cabeça. Isso decorre da diminuição das pistas espectrais que dependem fortemente das reflexões na pina (ausentes quando se utiliza fones de ouvido). Consequentemente, perde-se a percepção espacial de localização e passa-se a distinguir o lado e a distância aparente da fonte sonora em relação à linha mediana da cabeça, fenômeno denominado lateralização.

Figura 5 – Diferença de nível interaural (ILD) resultante de refrações e absorções da cabeça e tronco.



2.6 MÉTRICAS DE DESEMPENHO

Essa seção apresenta quatro métricas objetivas utilizadas para avaliação de desempenho de algoritmos de redução de ruído. Além disso, são apresentados também duas técnicas estatísticas utilizadas para a análise de dados provenientes de experimentos subjetivos.

2.6.1 Relação sinal ruído

A Relação Sinal Ruído (SNR – *Signal to Noise Ratio*) é uma métrica comumente utilizada para medir a qualidade de um sinal de fala. No domínio do tempo, ela pode ser dada, em dB, por:

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \left(\frac{P_f}{P_r} \right) \quad (2.2)$$

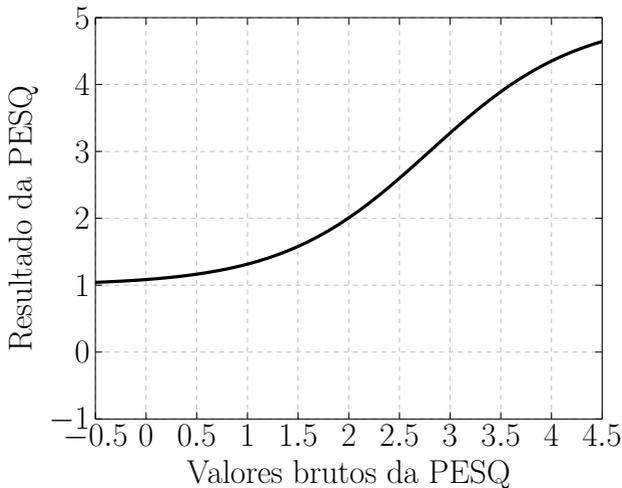
onde P_f é a potência do sinal de fala e P_r é a potência do sinal de ruído.

2.6.2 Avaliação perceptual da qualidade de fala

A avaliação perceptual da qualidade da fala (PESQ – *Perceptual Evaluation of Speech Quality*), padronizada como a recomendação ITU-

T P.862, é uma métrica objetiva que fornece informação da qualidade da fala em redes telefônicas e *codecs* de áudio. métrica pode ser utilizada em dois modos de operação: modo de banda estreita, utilizado para determinar a qualidade de sinais processados por sistemas telefônicos entre 300 e 3100 Hz; e modo de banda larga, utilizado para determinar a qualidade de sinais de fala processados por sistemas de comunicação que operam entre 50 e 7000 Hz. A PESQ no modo banda larga é determinada mapeando-se, através de uma função logística (Figura 6), a faixa de valores brutos da PESQ entre -0,5 e 4,5 para a faixa de valores entre 1 e 4,5 (ITU-T P.862.3, 2005). O resultado da PESQ, obtido nessa faixa de valores, tenta prever o valor da pontuação média de opinião (MOS-*Mean opinion score*), teste subjetivo de qualidade de fala.

Figura 6 – Função logística que determina os valores da PESQ no modo banda larga – mapeamento entre os valores brutos da métrica e os valores resultantes da utilização da métrica no modo banda larga.



2.6.3 Ganho ponderado de inteligibilidade

A maioria das métricas de inteligibilidade (SI-*Speech Inteligibility*) depende do cálculo da SNR em determinadas bandas de frequência e, portanto, é definida através da seguinte equação genérica (LOIZOU,

2013):

$$SI = \sum_{k=1}^K W_k \text{SNR}_k \quad (2.3)$$

onde K é o número de bandas de frequência do sinal analisado; W_k é a função de importância de banda (*BIF–Band-importance function*) que determina a contribuição relativa da banda de frequência k na inteligibilidade total da fala; e SNR_k é a SNR efetiva ou SNR aparente da banda k , cujo cálculo varia de acordo com a aplicação considerada.

O índice de articulação (*AI–Articulation index*) e o índice de transmissão de fala (*STI–Speech-Transmission Index*) são métricas muito utilizadas para caracterizar a SI. Em ambos casos, o cálculo é feito limitando-se a faixa dinâmica de valores da SNR em aproximadamente 30 dB. Além disso, o valor final do AI e do STI é dado entre 0 e 1. Tendo em vista essas restrições, Greenberg, Peterson e Zurek (1993) propuseram modificações no cálculo do AI e do STI de forma a evitar a limitação da faixa dinâmica da SNR. Como resultado, obtiveram a métrica denominada *intelligibility-weighted SNR* (ISNR), definida, em dB, como:

$$\text{ISNR} = \sum_{k=1}^K W_k \text{SNR}_k. \quad (2.4)$$

onde a ISNR é dada em dB e seu valor pode ser determinado ou através das etapas matemáticas da métrica AI ou da métrica STI, excluindo-se dos procedimentos matemáticos o estágio de restrição da faixa dinâmica da SNR. A partir de (2.4), Greenberg, Peterson e Zurek (1993) também propuseram a métrica denominada ganho de inteligibilidade do sistema ΔISNR , definida como:

$$\Delta\text{ISNR} = \text{ISNR}_{\text{out}} - \text{ISNR}_{\text{in}} \quad (2.5)$$

onde ISNR_{out} e ISNR_{in} são as ISNR calculadas, respectivamente, na saída e na entrada de um dado sistema de processamento.

2.6.4 Erro na diferença de intensidade interaural

A principal métrica de preservação da ILD é o erro na diferença de nível interaural, dada em dB (COSTA; NAYLOR, 2014). essa métrica é definida por:

$$\Delta\text{ILD} = 10 \log_{10} \text{ILD}_{\text{out}} - 10 \log_{10} \text{ILD}_{\text{in}} \quad (2.6)$$

onde ILD_{in} e ILD_{out} são respectivamente as ILDs do sinal na entrada e na saída do sistema de processamento. Essas duas ILDs são calculadas para frequências acima de 1,5 kHz.

2.6.5 Erro absoluto médio

Em experimentos subjetivos de determinação de azimute de localização de fontes sonoras, uma das métricas utilizadas para determinar a exatidão dos resultados é o erro absoluto médio (MAE–*Mean Absolute Error*), dado por (BOGAERT et al., 2008):

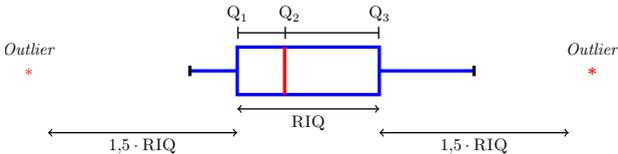
$$MEA = \frac{\sum_{i=1}^{N_a} |\theta_{OR} - \theta_{RP}|}{N_a} \quad (2.7)$$

onde N_a é o número de estímulos sonoros (fontes sonoras) que chegam até à posição do voluntário do experimento, θ_{OR} é o azimute original da fonte sonora e θ_{RP} é o azimute percebido pelo voluntário do experimento.

2.6.6 Diagrama de caixas – *Boxplot*

Diagrama de caixa ou diagrama de extremos e quartis é uma forma de representação de um conjunto de dados que fornece informações como: centro do conjunto de dados, dispersão, simetria e observações não usuais (*outliers*) (MONTGOMERY; RUNGER, 2011).

Figura 7 – Elementos do diagrama de caixas.



A Figura 7 apresenta os elementos que constituem um diagrama de caixa. O quartil inferior, a mediana e o quartil superior são representados respectivamente por Q_1 , Q_2 e Q_3 . A caixa representa a região inter-quartil (RIQ), i.e., $Q_3 - Q_1$. Dentro dela, a mediana, Q_2 , é representada através de uma linha paralela aos extremos da caixa. As linhas que se estendem a partir dos extremos da caixa são denominadas *whiskers*. O *whisker* inferior (linha à esquerda) inicia em Q_1 e

termina no menor valor do conjunto de dados que seja maior ou igual a $Q_1 - \text{RIQ} \cdot 1,5$. O *whisker* superior (linha à direita) inicia em Q_3 e termina no maior valor do conjunto de dados que seja menor ou igual a $Q_3 + \text{RIQ} \cdot 1,5$. Os valores localizados além dos *whiskers* são denominados valores discrepantes (*outliers*).

2.6.7 Teste de Wilcoxon

O teste de Wilcoxon (*Wilcoxon signed-rank test*) é um teste de hipóteses não paramétrico desenvolvido para se fazer inferências sobre a mediana resultante da diferença entre duas populações (GIBBONS; CHAKRABORTI, 2003). Para que esse teste seja utilizado é necessário que os dados sejam pareados e provenientes da mesma população.

Considerando a diferença entre as populações X e Y (i.e., $X - Y$) uma distribuição contínua e com mediana M_D , o teste de Wilcoxon verifica se M_D está centrada em 0 ou não. Essas afirmações são analisadas de maneira formal através do seguinte teste de hipóteses:

$$\begin{aligned} H_0 : M_D &= 0, \\ H_1 : M_D &\neq 0. \end{aligned} \tag{2.8}$$

O processo de decisão através de (2.8) pode resultar em dois tipos de erros:

- Erro tipo I: consiste em rejeitar a hipótese nula (H_0) quando ela é correta.
- Erro tipo II: consiste em não rejeitar a hipótese nula quando ela é falsa.

A estes erros estão associadas as probabilidades α_1 e β_2 . O valor α_1 determina a probabilidade do erro tipo I e é também denominado nível de significância do teste de hipóteses; o valor β_2 determina a probabilidade do erro tipo II. Uma maneira de informar o resultado de um teste de hipóteses é através do valor-P (p). Esse valor pode ser entendido como o risco de se rejeitar erroneamente H_0 considerando os dados observados. De forma prática, para se determinar o resultado de um teste de hipóteses basta comparar p com o valor de α_1 da seguinte forma (GIBBONS; CHAKRABORTI, 2003):

$$\begin{aligned} \text{Se } p &\leq \alpha_1, \text{ então rejeita-se } H_0, \\ \text{Se } p &> \alpha_1, \text{ então não se rejeita } H_0. \end{aligned}$$

2.7 CONCLUSÕES

Esse capítulo apresentou de forma breve alguns conceitos e tópicos que serão importantes durante o desenvolvimento desse trabalho.

Capítulo 3

Abordagem Proposta

Esse capítulo apresenta o método proposto para redução de ruído e preservação das ILDs dos sinais recebidos por um aparelho binaural. Na Seção 3.1 são apresentados os sinais e sistemas que caracterizam o cenário acústico e o sistema de processamento binaural. Na Seção 3.2 é detalhada a função custo binaural a partir da qual é desenvolvido o método proposto. Na Seção 3.3 é derivado o algoritmo adaptativo para redução de ruído e preservação das ILDs dos sinais recebidos pelo sistema de processamento binaural. Posteriormente nessa Seção, são apresentados os estimadores para as matrizes de correlação necessárias no método proposto e, em seguida, é feita uma análise do custo computacional médio para a determinação dos coeficientes do algoritmo adaptativo. Por fim, na Seção 3.4 são apresentadas as conclusões do capítulo.

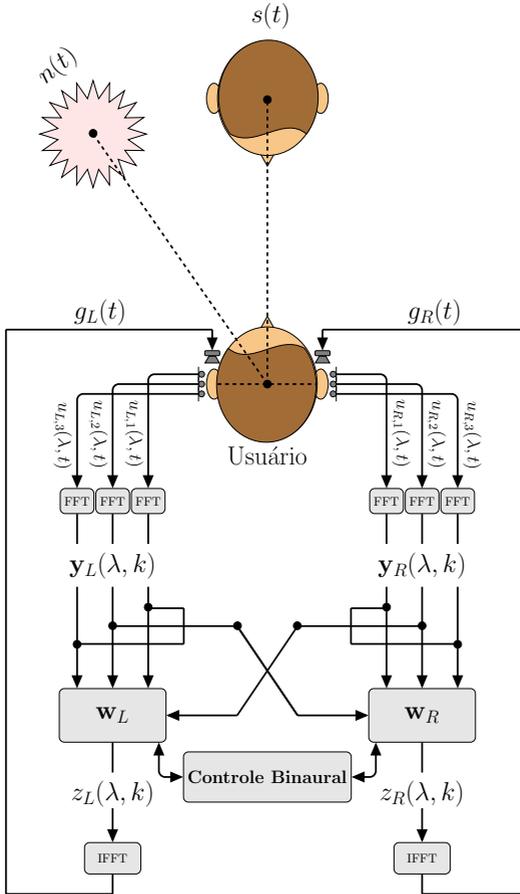
3.1 SISTEMA DE PROCESSAMENTO BINAURAL

A Figura 8 apresenta um usuário de aparelho auditivo binaural (sistema de processamento binaural) em um cenário acústico com duas fontes sonoras pontuais: uma fonte de fala, à frente, e uma fonte de ruído, à esquerda (*L - Left*), ambas localizadas no campo distante do usuário. O aparelho auditivo na orelha esquerda possui M_L microfones, e o aparelho auditivo na orelha direita (*R - Right*), M_R . A representação no tempo do sinal recebido no l -ésimo microfone da esquerda, denotado pelo subíndice $(\cdot)_{L,r}$, e no r -ésimo da direita, denotado por $(\cdot)_{R,r}$, é dada respectivamente por

$$\begin{aligned} u_{L,r}(t) &= s_{L,r}(t) + n_{L,r}(t), \\ u_{R,r}(t) &= s_{R,r}(t) + n_{R,r}(t), \end{aligned} \tag{3.1}$$

onde $l = 1, 2, \dots, M_L$; $r = 1, 2, \dots, M_R$; $s_i(t)$ é o sinal de fala desejado e $n_i(t)$ o ruído aditivo para $i \in \{(L, l), (R, r)\}$.

Figura 8 – Cenário acústico com fontes pontuais e sistema de processamento binaural de redução de ruído.



Os sinais em (3.1) são processados em segmentos de sinal denominados *frames*, cuja representação no domínio do tempo é definida por

$$\begin{aligned} u_{L,r}(\lambda, t) &= s_{L,r}(\lambda, t) + n_{L,r}(\lambda, t), \\ u_{R,r}(\lambda, t) &= s_{R,r}(\lambda, t) + n_{L,r}(\lambda, t), \end{aligned} \quad (3.2)$$

onde λ é o índice do *frame* de tamanho F . Decompondo-se os sinais em (3.2) em suas componentes de frequência, através de uma transformada de Fourier de tempo curto (STFT – *Short-Time Fourier Transform*)

de N bins, tem-se a seguinte representação em frequência:

$$\begin{aligned} y_{L,r}(\lambda, k) &= x_{L,r}(\lambda, k) + v_{L,r}(\lambda, k), \\ y_{R,r}(\lambda, k) &= x_{R,r}(\lambda, k) + v_{R,r}(\lambda, k), \end{aligned} \quad (3.3)$$

onde $y_i(\lambda, k)$, $x_i(\lambda, k)$ e $v_i(\lambda, k)$ são as STFTs dos sinais $u_i(\lambda, t)$, $s_i(\lambda, t)$ e $n_i(\lambda, t)$, respectivamente, e k é o índice (*bin*) de frequência. Para cada *frame* a ser processado, as amostras da frequência k são agrupadas nos vetores \mathbf{y}_L e \mathbf{y}_R definidos da seguinte forma:

$$\begin{aligned} \mathbf{y}_L(\lambda, k) &= \mathbf{x}_L(\lambda, k) + \mathbf{v}_L(\lambda, k), \\ \mathbf{y}_R(\lambda, k) &= \mathbf{x}_R(\lambda, k) + \mathbf{v}_R(\lambda, k), \end{aligned} \quad (3.4)$$

ou ainda, definidos da seguinte forma:

$$\begin{aligned} \mathbf{y}_L(\lambda, k) &= [y_{L,1}(\lambda, k) \ y_{L,2}(\lambda, k) \ \cdots \ y_{L,M_L}(\lambda, k)]^T, \\ \mathbf{y}_R(\lambda, k) &= [y_{R,1}(\lambda, k) \ y_{R,2}(\lambda, k) \ \cdots \ y_{R,M_R}(\lambda, k)]^T, \end{aligned} \quad (3.5)$$

sendo $\mathbf{y}_L(\lambda, k)$ e $\mathbf{y}_R(\lambda, k)$, respectivamente, de dimensões $M_L \times 1$ e $M_R \times 1$. Considerando que o sistema binaural possui um canal de comunicação *full duplex*¹ e sem restrições na taxa de *bits*, os vetores em (3.4) podem ser utilizados por ambos os aparelhos auditivos. Dessa forma, esses vetores são agrupados em um único vetor, obtendo-se o vetor de entrada no domínio da frequência

$$\mathbf{y}(\lambda, k) = [\mathbf{y}_L^T(\lambda, k) \ \mathbf{y}_R^T(\lambda, k)]^T, \quad (3.6)$$

que possui dimensão $M \times 1$, com $M = M_L + M_R$ e $(\cdot)^T$ é a operação de transposição. De forma semelhante, definem-se os vetores $\mathbf{x}(\lambda, k)$ e $\mathbf{v}(\lambda, k)$. Os sinais de saída em cada aparelho auditivo, $z_L(\lambda, k)$ e $z_R(\lambda, k)$, são obtidos através do produto interno entre $\mathbf{y}(\lambda, k)$ e os vetores de coeficientes $\mathbf{w}_L(\lambda, k)$ e $\mathbf{w}_R(\lambda, k)$:

$$\begin{aligned} z_L(\lambda, k) &= \mathbf{w}_L^H(\lambda, k) \mathbf{y}(\lambda, k) = z_L^x(\lambda, k) + z_L^v(\lambda, k), \\ z_R(\lambda, k) &= \mathbf{w}_R^H(\lambda, k) \mathbf{y}(\lambda, k) = z_R^x(\lambda, k) + z_R^v(\lambda, k), \end{aligned} \quad (3.7)$$

onde $(\cdot)^H$ representa o transposto hermitiano e $z_L^x(\lambda, k)$ e $z_R^x(\lambda, k)$ são

¹Um sistema *full duplex* permite comunicação entre os dois aparelhos auditivos de forma simultânea e em ambas as direções.

as componentes de fala na saída do filtro linear, dadas por:

$$\begin{aligned} z_L^x(\lambda, k) &= \mathbf{w}_L^H \mathbf{x}(\lambda, k), \\ z_R^x(\lambda, k) &= \mathbf{w}_R^H \mathbf{x}(\lambda, k), \end{aligned} \quad (3.8)$$

e $z_L^v(\lambda, k)$ e $z_R^v(\lambda, k)$ são as componentes de ruído na saída do filtro linear, dadas por:

$$\begin{aligned} z_L^v(\lambda, k) &= \mathbf{w}_L^H \mathbf{v}(\lambda, k), \\ z_R^v(\lambda, k) &= \mathbf{w}_R^H \mathbf{v}(\lambda, k). \end{aligned} \quad (3.9)$$

Os sinais determinados pelos vetores em (3.7) são transformados de volta para o domínio do tempo, através de uma rotina de *overlap-and-weight* (CROCHIERE, 1980) resultando nos sinais $g_L(t)$ e $g_R(t)$, que são enviados aos alto-falantes dos respectivos aparelhos auditivos. O bloco de controle binaural gerencia a troca de informações e parâmetros de controle entre os aparelhos auditivos.

Dentro do sistema binaural descrito, é importante definir-se os microfones de referência dos aparelhos auditivos. Para tanto, utilizam-se os vetores de seleção de microfones \mathbf{q}_L e \mathbf{q}_R , ambos de dimensão $M \times 1$. Nos dois vetores, as M_L posições iniciais são associadas aos microfones do lado esquerdo e as M_R finais são associadas aos microfones do lado direito. Usualmente, em aplicações binaurais e bilaterais, esses microfones são definidos como sendo os microfones frontais de ambos aparelhos. Dessa forma, o vetor do sinal do microfone frontal direito é indicado pelo vetor \mathbf{q}_R que apresenta o valor 1 em sua primeira posição e zero nas restantes. O sinal de referência do lado esquerdo (microfone frontal) apresenta valores zero nas primeiras M_R posições, seguidas pelo valor 1 e $(M_L - 1)$ zeros. Para o caso em que $M_L = 3$ e $M_R = 3$, i.e., $M = 6$, tem-se, portanto, $\mathbf{q}_L = [100000]^T$ e $\mathbf{q}_R = [000100]^T$. A partir desses vetores é possível determinar-se o sinal de entrada nos microfones de referência por meio de:

$$\begin{aligned} y_{L,\text{ref}}(\lambda, k) &= \mathbf{q}_L^T \mathbf{y}(\lambda, k), \\ y_{R,\text{ref}}(\lambda, k) &= \mathbf{q}_R^T \mathbf{y}(\lambda, k). \end{aligned} \quad (3.10)$$

Os vetores \mathbf{q}_L e \mathbf{q}_R e \mathbf{w}_L e \mathbf{w}_R podem ser concatenados para obter uma representação mais simplificada, de forma que

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_L \\ \mathbf{w}_R \end{bmatrix} \quad \text{e} \quad \mathbf{q} = \begin{bmatrix} \mathbf{q}_L \\ \mathbf{q}_R \end{bmatrix}, \quad (3.11)$$

onde \mathbf{q} e \mathbf{w} possuem dimensão $2M \times 1$.

3.2 FUNÇÃO CUSTO BINAURAL

Uma das estratégias mais utilizadas para determinação dos coeficientes \mathbf{w}_L e \mathbf{w}_R é o MWF. Contudo, o estimador definido por essa técnica distorce as pistas binaurais dos sinais processados (CORNELIS et al., 2010) e, portanto, modifica o cenário acústico percebido pelo usuário do aparelho binaural. De forma a evitar tal distorção e com o intuito de manter uma percepção mais verossímil do cenário acústico, Doclo et al. (2005) utilizaram como estratégia a adição de uma função custo auxiliar à função custo do MWF². Essa função adicional é baseada na minimização do erro entre as ILDs do ruído na entrada e na saída do sistema binaural. Posteriormente, a partir de uma alteração na função custo auxiliar proposta por Doclo et al. (2005), Costa e Naylor (2014) obtiveram uma função custo com maior simetria em relação a ambos os aparelhos, i.e., o custo associado à preservação das pistas acústicas das fontes sonoras localizadas em ângulos simétricos é o mesmo. Isto decorre de uma estimativa mais precisa da morfologia da ILD no domínio da frequência, como uma função do ângulo incidente. O objetivo principal das duas abordagens (DOCLO et al., 2005; COSTA; NAYLOR, 2014) é obter coeficientes \mathbf{w}_L e \mathbf{w}_R que permitam reduzir a potência do ruído aditivo assim como preservar a localização espacial original das fontes do cenário acústico. Para tanto, define-se a seguinte função custo (DOCLO et al., 2005; COSTA; NAYLOR, 2014):

$$J(\lambda, k) = J_{\text{MWF}}(\lambda, k) + \alpha J_{\text{ILD}}(\lambda, k), \quad (3.12)$$

cuja primeira parcela, $J_{\text{MWF}}(\lambda, k)$, é a função custo do MWF, definida como

$$J_{\text{MWF}}(\lambda, k) = E \left\{ \left\| \begin{bmatrix} x_{L,\text{ref}}(\lambda, k) - \mathbf{w}_L^H(\lambda, k) \mathbf{y}(\lambda, k) \\ x_{R,\text{ref}}(\lambda, k) - \mathbf{w}_R^H(\lambda, k) \mathbf{y}(\lambda, k) \end{bmatrix} \right\|^2 \right\}, \quad (3.13)$$

onde $E\{\cdot\}$ é o operador valor esperado e $\|\cdot\|^2$ é a norma euclidiana ao quadrado. A segunda parcela, $J_{\text{ILD}}(\lambda, k)$, é a função custo auxiliar para preservação da ILD do ruído, derivada a partir do erro na diferença de

²Com intuito de manter o padrão de notação adotado na literatura especializada, optou-se por se utilizar *função custo do MWF* e não *função custo do erro quadrático médio*.

intensidade interaural do ruído (ver Seção 2.6.4), dada por:

$$J_{\text{ILD}}(\lambda, k) = \left[10 \log_{10} \left(\frac{\text{ILD}_{\text{out}}(\lambda, k)}{\text{ILD}_{\text{in}}(\lambda, k)} \right) \right]^2 \quad (3.14)$$

em que a ILD de saída, ILD_{out} , e a ILD de entrada, ILD_{in} , são definidas como:

$$\begin{aligned} \text{ILD}_{\text{out}}(\lambda, k) &= \frac{E\{|z_{\text{L}}^v(\lambda, k)|^2\}}{E\{|z_{\text{R}}^v(\lambda, k)|^2\}}, \\ \text{ILD}_{\text{in}}(\lambda, k) &= \frac{E\{|v_{\text{L,ref}}(\lambda, k)|^2\}}{E\{|v_{\text{R,ref}}(\lambda, k)|^2\}}. \end{aligned} \quad (3.15)$$

Por fim, o fator α em (3.12) é uma constante de ponderação que determina a importância da $J_{\text{ILD}}(\lambda, k)$ na função custo global $J(\lambda, k)$. Por simplicidade e conveniência de notação, suprime-se a dependência de (λ, k) em todas as equações posteriores.

Manipulando-se (3.13), chega-se a:

$$\begin{aligned} J_{\text{MWF}} &= r_{x\text{L}} - \mathbf{w}_{\text{L}}^{\text{H}} \mathbf{r}_{x\text{L}} - \mathbf{r}_{x\text{L}}^{\text{H}} \mathbf{w}_{\text{L}} + \mathbf{w}_{\text{L}}^{\text{H}} \mathbf{R}_{\mathbf{y}} \mathbf{w}_{\text{L}} + \\ &\quad r_{x\text{R}} - \mathbf{w}_{\text{R}}^{\text{H}} \mathbf{r}_{x\text{R}} - \mathbf{r}_{x\text{R}}^{\text{H}} \mathbf{w}_{\text{R}} + \mathbf{w}_{\text{R}}^{\text{H}} \mathbf{R}_{\mathbf{y}} \mathbf{w}_{\text{R}}. \end{aligned} \quad (3.16)$$

A partir de (3.14) e efetuando-se as aproximações presentes no ANEXO A, resulta em (COSTA; NAYLOR, 2014):

$$J_{\text{ILD}} = \left(\frac{r_{v\text{R}} \mathbf{w}_{\text{L}}^{\text{H}} \mathbf{R}_{\mathbf{v}} \mathbf{w}_{\text{L}} - r_{v\text{L}} \mathbf{w}_{\text{R}}^{\text{H}} \mathbf{R}_{\mathbf{v}} \mathbf{w}_{\text{R}}}{r_{v\text{R}} \mathbf{w}_{\text{L}}^{\text{H}} \mathbf{R}_{\mathbf{v}} \mathbf{w}_{\text{L}} + r_{v\text{L}} \mathbf{w}_{\text{R}}^{\text{H}} \mathbf{R}_{\mathbf{v}} \mathbf{w}_{\text{R}}} \right)^2 \quad (3.17)$$

em que

$$\mathbf{R}_{\mathbf{y}} = E\{\mathbf{y}\mathbf{y}^{\text{H}}\}, \quad \mathbf{R}_{\mathbf{x}} = E\{\mathbf{x}\mathbf{x}^{\text{H}}\} \text{ e } \mathbf{R}_{\mathbf{v}} = E\{\mathbf{v}\mathbf{v}^{\text{H}}\} \quad (3.18)$$

são as matrizes de correlação do sinal de entrada, do sinal de fala e do ruído aditivo, respectivamente, cujas dimensões são todas iguais a $2M \times 2M$. Os vetores

$$\begin{aligned} \mathbf{r}_{x\text{L}} &= E\{x_{\text{L,ref}}^* \mathbf{x}\} = \mathbf{R}_{\mathbf{x}} \mathbf{q}_{\text{L}}, \\ \mathbf{r}_{x\text{R}} &= E\{x_{\text{R,ref}}^* \mathbf{x}\} = \mathbf{R}_{\mathbf{x}} \mathbf{q}_{\text{R}}, \end{aligned} \quad (3.19)$$

determinam a correlação entre o sinal de fala no microfone de referência

e o vetor do sinal de fala para o respectivo lado. Por fim, os escalares:

$$\begin{aligned}
 r_{xL} &= E\{|x_{L,\text{ref}}|^2\} = \mathbf{q}_L^T \mathbf{R}_x \mathbf{q}_L, \\
 r_{xR} &= E\{|x_{R,\text{ref}}|^2\} = \mathbf{q}_R^T \mathbf{R}_x \mathbf{q}_R, \\
 r_{vL} &= E\{|v_{L,\text{ref}}|^2\} = \mathbf{q}_L^T \mathbf{R}_v \mathbf{q}_L, \\
 r_{vR} &= E\{|v_{R,\text{ref}}|^2\} = \mathbf{q}_R^T \mathbf{R}_v \mathbf{q}_R,
 \end{aligned} \tag{3.20}$$

são a potência do sinal de fala nos microfones de referência dos lados esquerdo e direito, e a potência do ruído nos microfones de referência dos lados esquerdo e direito, respectivamente. Substituindo (3.16) e (3.17) em (3.12), obtém-se a função custo:

$$\begin{aligned}
 J &= r_{xL} - \mathbf{w}_L^H \mathbf{r}_{xL} - \mathbf{r}_{xL}^H \mathbf{w}_L + \mathbf{w}_L^H \mathbf{R}_y \mathbf{w}_L \\
 &\quad + r_{xR} - \mathbf{w}_R^H \mathbf{r}_{xR} - \mathbf{r}_{xR}^H \mathbf{w}_R + \mathbf{w}_R^H \mathbf{R}_y \mathbf{w}_R \\
 &\quad + \alpha \left(\frac{r_{vR} \mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L - r_{vL} \mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R}{r_{vR} \mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L + r_{vL} \mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R} \right)^2.
 \end{aligned} \tag{3.21}$$

A Equação (3.21) não é estritamente convexa em função de \mathbf{w}_L e \mathbf{w}_R , podendo apresentar múltiplos mínimos locais. A determinação dos coeficientes ótimos (que minimizam a função custo) é realizada através de métodos numéricos de otimização (DOCCLO et al., 2005; COSTA; NAYLOR, 2014).

3.3 MÉTODO PROPOSTO

Essa seção apresenta o método proposto para a estimação dos vetores de coeficientes \mathbf{w}_L e \mathbf{w}_R . Adicionalmente são propostos estimadores para as matrizes de correlação de fala e de ruído necessários para o algoritmo apresentado. Finalmente, é avaliado o custo computacional por iteração da técnica proposta.

3.3.1 Equações recursivas

Um procedimento iterativo para calcular os vetores de coeficientes \mathbf{w}_L e \mathbf{w}_R é obtido aplicando-se o método do gradiente descendente na função custo dada pela Equação (3.21). Para tanto, considera-se a seguinte equação recursiva para cada frequência k do *frame* λ (TA-

RIGHAT; SAYED, 2004):

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \beta \nabla_{\mathbf{w}^*} J(\mathbf{w}_i), \quad (3.22)$$

onde β é o passo de adaptação e i é um índice de recursão, que varia de 1 a N_i . Expandindo-se (3.22) em função dos vetores \mathbf{w}_L e \mathbf{w}_R , tem-se:

$$\begin{bmatrix} \mathbf{w}_{L_{i+1}} \\ \mathbf{w}_{R_{i+1}} \end{bmatrix} = \begin{bmatrix} \mathbf{w}_{L_i} \\ \mathbf{w}_{R_i} \end{bmatrix} - \beta \begin{bmatrix} \partial_{\mathbf{w}_L^*} J(\mathbf{w}_{L_i}) \\ \partial_{\mathbf{w}_R^*} J(\mathbf{w}_{R_i}) \end{bmatrix}. \quad (3.23)$$

Desacoplando (3.23) e expandindo os termos $\partial_{\mathbf{w}_L^*} J$ e $\partial_{\mathbf{w}_R^*} J$, obtém-se as seguintes equações recursivas:

$$\begin{aligned} \mathbf{w}_{L_{i+1}} &= \mathbf{w}_{L_i} - \beta \partial_{\mathbf{w}_L^*} J_{MWF}(\mathbf{w}_{L_i}) - \gamma \partial_{\mathbf{w}_L^*} J_{ILD}(\mathbf{w}_{L_i}), \\ \mathbf{w}_{R_{i+1}} &= \mathbf{w}_{R_i} - \beta \partial_{\mathbf{w}_R^*} J_{MWF}(\mathbf{w}_{R_i}) - \gamma \partial_{\mathbf{w}_R^*} J_{ILD}(\mathbf{w}_{R_i}), \end{aligned} \quad (3.24)$$

onde $\gamma = \alpha\beta$. Considerando $\mathbf{w}_L = \mathbf{w}_L^{\Re} + j\mathbf{w}_L^{\Im}$ e $\mathbf{w}_R = \mathbf{w}_R^{\Re} + j\mathbf{w}_R^{\Im}$, onde \mathbf{w}_L^{\Re} e \mathbf{w}_R^{\Re} são os vetores das componentes reais de \mathbf{w}_L e \mathbf{w}_R , respectivamente, e \mathbf{w}_L^{\Im} e \mathbf{w}_R^{\Im} são, respectivamente, os vetores das componentes imaginárias dos vetores \mathbf{w}_L e \mathbf{w}_R . As derivadas parciais de J_{MWF} em relação a $\mathbf{w}_{L_i}^*$ e $\mathbf{w}_{R_i}^*$ são definidas como (HJØRUNGNES; GESBERT, 2007):

$$\begin{aligned} \partial_{\mathbf{w}_L^*} J_{MWF}(\mathbf{w}_{L_i}) &= \frac{1}{2} \left[\partial_{\mathbf{w}_L^{\Re}} J_{MWF}(\mathbf{w}_{L_i}) + j \partial_{\mathbf{w}_L^{\Im}} J_{MWF}(\mathbf{w}_{L_i}) \right], \\ \partial_{\mathbf{w}_R^*} J_{MWF}(\mathbf{w}_{R_i}) &= \frac{1}{2} \left[\partial_{\mathbf{w}_R^{\Re}} J_{MWF}(\mathbf{w}_{R_i}) + j \partial_{\mathbf{w}_R^{\Im}} J_{MWF}(\mathbf{w}_{R_i}) \right]. \end{aligned} \quad (3.25)$$

A partir dos resultados apresentados em Brandwood (1983), resulta que:

$$\begin{aligned} \partial_{\mathbf{w}_L^*} J_{MWF}(\mathbf{w}_{L_i}) &= \mathbf{R}_y \mathbf{w}_{L_i} - \mathbf{r}_{xL}, \\ \partial_{\mathbf{w}_R^*} J_{MWF}(\mathbf{w}_{R_i}) &= \mathbf{R}_y \mathbf{w}_{R_i} - \mathbf{r}_{xR}. \end{aligned} \quad (3.26)$$

As derivadas parciais de J_{ILD} em relação à \mathbf{w}_L^* e \mathbf{w}_R^* são definidas de forma equivalente às derivadas em (3.25). O desenvolvimento matemático realizado para determiná-las é descrito no Apêndice B. Os

resultados obtidos são:

$$\begin{aligned} \partial_{\mathbf{w}_{L_i}^*} J_{\text{ILD}}(\mathbf{w}_{L_i}) &= 4 r_{vL} r_{vR} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i} \\ &\cdot \frac{r_{vR} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i} - r_{vL} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i}}{(r_{vR} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i} + r_{vL} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i})^3} \mathbf{R}_v \mathbf{w}_{L_i}, \\ \partial_{\mathbf{w}_{R_i}^*} J_{\text{ILD}}(\mathbf{w}_{L_i}) &= 4 r_{vR} r_{vL} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i} \\ &\cdot \frac{r_{vL} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i} - r_{vR} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i}}{(r_{vR} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i} + r_{vL} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i})^3} \mathbf{R}_v \mathbf{w}_{R_i}. \end{aligned} \quad (3.27)$$

Substituindo (3.26) e (3.27) em (3.24), obtém-se a equação de atualização do algoritmo adaptativo proposto para redução de ruído e preservação da ILD, dado por:

$$\begin{aligned} \mathbf{w}_{L_{i+1}} &= \mathbf{w}_{L_i} - \beta (\mathbf{R}_y \mathbf{w}_{L_i} - \mathbf{r}_{xL}) - \gamma 4 r_{vL} r_{vR} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i} \\ &\cdot \frac{r_{vR} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i} - r_{vL} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i}}{(r_{vR} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i} + r_{vL} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i})^3} \mathbf{R}_v \mathbf{w}_{L_i}, \\ \mathbf{w}_{R_{i+1}} &= \mathbf{w}_{R_i} - \beta (\mathbf{R}_y \mathbf{w}_{R_i} - \mathbf{r}_{xR}) - \gamma 4 r_{vR} r_{vL} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i} \\ &\cdot \frac{r_{vL} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i} - r_{vR} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i}}{(r_{vR} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i} + r_{vL} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i})^3} \mathbf{R}_v \mathbf{w}_{R_i}. \end{aligned} \quad (3.28)$$

Note $\gamma = 0$ leva a uma versão adaptativa do MWF.

3.3.2 Estimação das matrizes de correlação

O cálculo dos coeficientes \mathbf{w}_L e \mathbf{w}_R , definidos em (3.28), necessita da estimação das matrizes \mathbf{R}_y , \mathbf{R}_x e \mathbf{R}_v . Isto é feito com o auxílio de um detector de fala (VAD – *Voice activity detector*). Nos *frames* do sinal de entrada que contêm fala+ruído, o VAD apresenta valor 1, e as estimativas $\hat{\mathbf{R}}_y$, $\hat{\mathbf{R}}_x$ e $\hat{\mathbf{R}}_v$ são determinadas, em cada *frame* λ , a partir dos seguintes estimadores recursivos, para $i > 0$, (HURTADO; ANDERSON, 2012):

$$\begin{aligned} \hat{\mathbf{R}}_{v,i}(k) &= \hat{\mathbf{R}}_{v,i-1}(k), \\ \hat{\mathbf{R}}_{y,i}(k) &= \mu_y \hat{\mathbf{R}}_{y,i-1}(k) + (1 - \mu_y) \mathbf{y}_i(k) \mathbf{y}_i^H(k), \\ \hat{\mathbf{R}}_{x,i}(k) &= \mu_x \hat{\mathbf{R}}_{x,i-1}(k) + (1 - \mu_x) (\hat{\mathbf{R}}_{y,i}(k) - \hat{\mathbf{R}}_{v,i}(k)). \end{aligned} \quad (3.29)$$

Nos *frames* contendo apenas ruído, o VAD apresenta valor diferente de 1 e a estimativas das matrizes de correlação são determinados a partir de:

$$\begin{aligned}\hat{\mathbf{R}}_{x,i}(k) &= \hat{\mathbf{R}}_{x,i-1}(k), \\ \hat{\mathbf{R}}_{y,i}(k) &= \mu_y \hat{\mathbf{R}}_{y,i-1}(k) + (1 - \mu_y) \mathbf{y}_i(k) \mathbf{y}_i^H(k), \\ \hat{\mathbf{R}}_{v,i}(k) &= \mu_v \hat{\mathbf{R}}_{v,i-1}(k) + (1 - \mu_v) \mathbf{y}_i(k) \mathbf{y}_i^H(k).\end{aligned}\tag{3.30}$$

Em (3.29) e (3.30), os fatores μ_y , μ_x e μ_v , denominados fatores de esquecimento, apresentam valores entre 0 e 1 e determinam a ponderação entre as estimativas passadas (e.g., $\hat{\mathbf{R}}_{y,i-1}(k)$) e as inovações (e.g., $\mathbf{y}_i(k) \mathbf{y}_i^H(k)$) na estimativa atual (e.g., $\hat{\mathbf{R}}_{y,i}(k)$).

Os valores iniciais de $\hat{\mathbf{R}}_{y,i}$, $\hat{\mathbf{R}}_{x,i}$ e $\hat{\mathbf{R}}_{v,i}$ (i.e., $i = 0$) são determinados, a partir de um trecho de T_{VI} segundos iniciais do sinal de entrada, através dos seguintes filtros de resposta ao impulso finita (VALIN et al., 2004):

$$\begin{aligned}\hat{\mathbf{R}}_{y,0}(k) &= \frac{1}{N_y} \sum_{\lambda \in \lambda_y}^{N_y} \mathbf{y}(\lambda, k) \mathbf{y}^H(\lambda, k) \\ \hat{\mathbf{R}}_{v,0}(k) &= \frac{1}{N_v} \sum_{\lambda \in \lambda_v}^{N_v} \mathbf{y}(\lambda, k) \mathbf{y}^H(\lambda, k) \\ \hat{\mathbf{R}}_{x,0}(k) &= \hat{\mathbf{R}}_{y,0}(k) - \hat{\mathbf{R}}_{v,0}(k).\end{aligned}\tag{3.31}$$

onde N_y e N_v são, respectivamente, os números totais de *frames* de fala+ruído (λ_y) e *frames* de apenas ruído (λ_v) presentes nos T_{VI} segundos iniciais do sinal processado.

3.3.3 Análise do custo computacional

Para a determinação do custo computacional das equações recursivas, rearranjou-se (3.28) da seguinte forma:

$$\begin{aligned}
 \mathbf{w}_{L_{i+1}} &= (\mathbf{I} - \beta \mathbf{R}_y) \mathbf{w}_{L_i} - \beta \mathbf{r}_{xL} - \gamma 4 r_{vL} r_{vR} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i} \\
 &\quad \cdot \frac{r_{vR} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i} - r_{vL} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i}}{(r_{vR} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i} + r_{vL} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i})^3} \mathbf{R}_v \mathbf{w}_{L_i}, \\
 \mathbf{w}_{R_{i+1}} &= (\mathbf{I} - \beta \mathbf{R}_y) \mathbf{w}_{R_i} - \beta \mathbf{r}_{xR} - \gamma 4 r_{vR} r_{vL} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i} \\
 &\quad \cdot \frac{r_{vL} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i} - r_{vR} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i}}{(r_{vR} \mathbf{w}_{L_i}^H \mathbf{R}_v \mathbf{w}_{L_i} + r_{vL} \mathbf{w}_{R_i}^H \mathbf{R}_v \mathbf{w}_{R_i})^3} \mathbf{R}_v \mathbf{w}_{R_i}.
 \end{aligned} \tag{3.32}$$

O custo computacional do método proposto foi determinado a partir do número de operações multiplica-acumula (MAC) necessárias para calcular os coeficientes do filtro adaptativo a partir da Equação (3.28) utilizando os estimadores das matrizes de correlação em (3.29) e (3.30). Para tanto, o custo computacional de algumas operações elementares (Tabela 1) foi determinado a partir dos seguintes operandos: escalares a e $b \in \mathbb{R}$, c e $d \in \mathbb{C}$; vetores \mathbf{a} , \mathbf{b} , \mathbf{c} e $\mathbf{d} \in \mathbb{C}^{M \times 1}$; matrizes \mathbf{H} e $\mathbf{R} \in \mathbb{C}^{M \times M}$ onde $\mathbf{R} = \mathbf{R}^H$ e matriz identidade $\mathbf{I} \in \mathbb{R}^{M \times M}$. Para a inversão de um número complexo ($1/c$) é necessário uma operação de módulo de um número complexo e 2 divisões reais (ALT; LEEUWEN, 1981), uma divisão real é dada por 16 operações MAC (LU et al., 2005).

A Tabela 2 apresenta um pseudo-código para a estimação das matrizes \mathbf{R}_y , \mathbf{R}_x e \mathbf{R}_v e o número de operações MAC. Os trechos de ruído e fala+ruído são determinados a partir do valor da variável VAD: se VAD = 1, então o *frame* atual é composto por fala+ruído e aplica-se (3.29); se VAD = 0, então o *frame* atual é composto apenas por ruído e aplica-se (3.30). O valor da variável VAD é determinado em etapa anterior ao processamento binaural. O custo computacional total do processo de estimação, em função do número total de microfones ($M = M_L + M_R$) e do número total de frequências processadas ($N/2$), presente na última linha da Tabela 2, é determinado para os casos mutuamente exclusivos onde VAD = 0 e VAD = 1.

A Tabela 3, apresenta um pseudo-código do método proposto e, a partir dos resultados presentes na Tabela 1, o custo computacional das equações recursivas, em função do número de microfones do aparelho binaural (M) e do número de iterações (N_i). A partir dos resultados obtidos nas Tabelas 2 e 3, obtêm-se então o custo computacional do

processo de determinação dos coeficientes do método proposto para preservação da ILD do ruído, apresentado na Tabela 4, para os casos complementares: $VAD = 0$ e $VAD = 1$. O custo computacional médio foi determinado considerando que 25% do tempo total corresponde a trechos de fala (HU, 2008).

3.4 CONCLUSÕES

Esse capítulo apresentou uma nova técnica para estimação dos coeficientes do filtro de redução de ruído de um sistema binaural. Essa técnica é desenvolvida a partir da função custo do MWF, acrescida de um termo de preservação da ILD do ruído. Os coeficientes do método proposto são calculados a partir de equações recursivas e, portanto, não necessitam de métodos numéricos de otimização. Como resultado final foi apresentada uma análise de custo computacional do algoritmo adaptativo através do número de operações MAC.

Tabela 1 – Operações elementares e número de operações MAC.

Operação	MAC
ab	1
ac	2
cd	4
c^3	8
$\frac{1}{c}$	34
aa	$2M$
$a\mathbf{R}$	M^2
$\mathbf{a}^H\mathbf{b}$	$4M$
$\mathbf{a}\mathbf{b}^H$	$4M^2$
$\mathbf{a}\mathbf{a}^H$	$2M^2$
$\mathbf{H}\mathbf{a}$	$4M^2$
$\mathbf{I} + \mathbf{H}$	M
$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} + \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix}$	$4M$

Tabela 2 – Custo computacional, determinado a partir de operações MAC, necessário para se determinar as estimativas $\hat{\mathbf{R}}_y$, $\hat{\mathbf{R}}_x$ e $\hat{\mathbf{R}}_v$.

Pseudo-código	MAC
$\delta_y = 1 - \mu_y$	1
$\delta_v = 1 - \mu_v$	1
$\delta_x = 1 - \mu_x$	1
PARA $k = 1$ até $N/2$ FAÇA :	
$\mathbf{Y}(k) = \hat{\mathbf{y}}(k)\hat{\mathbf{y}}^H(k)$	$2M^2 \frac{N}{2}$
$\hat{\mathbf{R}}_y(k) = \mu_y \hat{\mathbf{R}}_y(k) + \delta_y \mathbf{Y}(k)$	$3M^2 \frac{N}{2}$
SE VAD = 0 ENTÃO :	
$\hat{\mathbf{R}}_v(k) = \mu_v \hat{\mathbf{R}}_v(k) + \delta_v \mathbf{Y}(k)$	$3M^2 \frac{N}{2}$
SENÃO SE VAD = 1 ENTÃO :	
$\hat{\mathbf{R}}_x(k) = \mu_x \hat{\mathbf{R}}_x(k) + \delta_x (\hat{\mathbf{R}}_y(k) - \hat{\mathbf{R}}_v(k))$	$4M^2 \frac{N}{2}$
FIM SE	
FIM PARA	
Total	VAD = 0: $8M^2 \frac{N}{2} + 3$ VAD = 1: $9M^2 \frac{N}{2} + 3$

Tabela 3 – Custo computacional das equações recursivas que definem os coeficientes do método proposto.

Algoritmo	MAC
PARA de $k = 0$ até $N/2$ FAÇA :	
$\mathbf{a}_L = \beta \mathbf{r}_{xL}$	$2M \frac{N}{2}$
$\mathbf{a}_R = \beta \mathbf{r}_{xR}$	$2M \frac{N}{2}$
$\mathbf{B} = \mathbf{I} - \beta \mathbf{R}_y$	$(M^2 + M) \frac{N}{2}$
$\mathbf{R}_{vL} = r_{vL} \mathbf{R}_v$	$M^2 \frac{N}{2}$
$\mathbf{R}_{vR} = r_{vR} \mathbf{R}_v$	$M^2 \frac{N}{2}$
$a_L = -\gamma 4r_{vL}$	$2 \frac{N}{2}$
$a_R = -\gamma 4r_{vR}$	$2 \frac{N}{2}$
PARA de $n = 1$ até N_i FAÇA :	
$\mathbf{b}_L = \mathbf{B} \mathbf{w}_L(k)$	$4M^2 N_i \frac{N}{2}$
$\mathbf{b}_R = \mathbf{B} \mathbf{w}_R(k)$	$4M^2 N_i \frac{N}{2}$
$\mathbf{c}_L = \mathbf{R}_{vL} \mathbf{w}_L(k)$	$4M^2 N_i \frac{N}{2}$
$\mathbf{c}_R = \mathbf{R}_{vR} \mathbf{w}_R(k)$	$4M^2 N_i \frac{N}{2}$
$b_L = \mathbf{w}_L^H \mathbf{c}_L$	$4M N_i \frac{N}{2}$
$b_R = \mathbf{w}_R^H \mathbf{c}_R$	$4M N_i \frac{N}{2}$
$c = \frac{1}{(b_L + b_R)^3}$	$43 N_i \frac{N}{2}$
$d = b_L - b_R$	$N_i \frac{N}{2}$
$\mathbf{d}_L = a_R b_R c d \mathbf{c}_L$	$(2M + 3) N_i \frac{N}{2}$
$\mathbf{d}_R = -a_L b_L c d \mathbf{c}_R$	$(2M + 3) N_i \frac{N}{2}$
$\begin{bmatrix} \mathbf{w}_L(k) \\ \mathbf{w}_R(k) \end{bmatrix} = \begin{bmatrix} \mathbf{a}_L \\ \mathbf{a}_R \end{bmatrix} + \begin{bmatrix} \mathbf{b}_L \\ \mathbf{b}_R \end{bmatrix} + \begin{bmatrix} \mathbf{d}_L \\ \mathbf{d}_R \end{bmatrix}$	$8M N_i \frac{N}{2}$
FIM PARA	
FIM PARA	
Total	$[(16M^2 + 20M + 50)N_i + (3M^2 + 4M + 4)] \frac{N}{2}$

Tabela 4 – Custo computacional médio do método proposto.

Tipo do <i>frame</i>	Custo computacional
VAD = 0	$[(16M^2 + 20M + 50)N_i + (11M^2 + 4M + 4)] \frac{N}{2} + 3$
VAD = 1	$[(16M^2 + 20M + 50)N_i + (12M^2 + 4M + 4)] \frac{N}{2} + 3$
Custo Médio	$[(16M^2 + 20M + 50)N_i + (10,7M^2 + 4M + 4)] \frac{N}{2} + 3$

Capítulo 4

Materiais e Métodos

Este capítulo apresenta os materiais e métodos que serão utilizados para avaliação de desempenho do método proposto. Na Seção 4.1 são apresentadas as características gerais dos cenários acústicos simulados assim como os elementos utilizados para criá-los (funções de transferência e sinais sonoros). Em virtude das definições de variáveis apresentadas no Capítulo 3, na Seção 4.2 redefinem-se algumas figuras de mérito apresentadas no Capítulo 2. Na Seção 4.3 são apresentados os experimentos e justificativas para escolha dos valores dos parâmetros utilizados para a avaliação de desempenho do método proposto. Finalizando, na Seção 4.4 são apresentadas as conclusões do capítulo.

4.1 CENÁRIOS ACÚSTICOS SIMULADOS

As subseções a seguir apresentam as características gerais dos cenários acústicos que foram utilizados na avaliação de desempenho do método proposto: Na Seção 4.1.1 são apresentadas as funções de transferência utilizadas para a criação destes cenários. Na Seção 4.1.2 são descritos os sinais acústicos que caracterizam as fontes sonoras de fala e ruído. Na Seção 4.1.3 são apresentadas características gerais compartilhadas por todos os cenários acústicos criados.

4.1.1 Funções de transferência

Para geração dos cenários acústicos, foram utilizadas funções de transferência (*HRTF-Head related transfer functions*) obtidas em uma sala anecoica. Essas HRTFs foram obtidas por Kayser et al. (2009) e as medições foram realizadas utilizando-se uma fonte sonora e um manequim (simulador cabeça e torço Brüel & Kjær tipo 4128-C (Kjær & Brüel, 2014)), ao qual acoplou-se dois aparelhos auditivos retroauriculares (esquerdo e direito) cada um com 3 microfones ($M_L = M_R = 3$),

assim como mostrado na Figura 9.

Figura 9 – Medições das funções de transferência em sala anecoica utilizando manequim cabeça e torço, à esquerda, e uma fonte sonora, à direita. Fonte: (KAYSER et al., 2009).



4.1.2 Sinais acústicos

Todos os cenários acústicos criados são caracterizados por duas fontes pontuais: uma fonte de fala (*S-Speech*) e uma fonte de ruído (*N-Noise*). O sinal emitido pela fonte de fala foi simulado através de um trecho de voz masculina¹ (ITU-T P.50, 1998) de 14,7 segundos de duração, ilustrado pela Figura 10. O sinal emitido pela fonte de ruído foi simulado utilizando-se dois tipos de ruído com características espectrais complementares. O primeiro tipo de ruído utilizado possui características espectrais e temporais semelhantes às da fala² (*speech-like noise*) com predominância de baixas frequências (DRESCHLER et al., 2001), denominado ICRA. O segundo tipo de ruído é caracterizado predominantemente por altas frequências tendo sido obtido pela filtragem passa-altas de um ruído branco gaussiano, dadas as seguintes especificações: frequência de bloqueio definida em 100 Hz, frequência de passagem de 1,5 kHz, 3 dB de atenuação máxima na banda de passagem e 20 dB de atenuação mínima na banda de rejeição. A Figura 11

¹Arquivo de áudio A_eng.m5.wav disponível para download em <https://www.itu.int/net/itu-t/sigdb/genaudio/AudioForm-g.aspx?val=1000050>

²ICRA 1: Ruído Gaussiano não modulado com espectro idealizado de um sinal de fala masculino.

apresenta as magnitudes dos periodogramas do sinal de fala e dos sinais de ruído descritos neste parágrafo.

Figura 10 – Representação do sinal de fala, no domínio do tempo, utilizado nos experimentos de avaliação do filtro adaptativo.

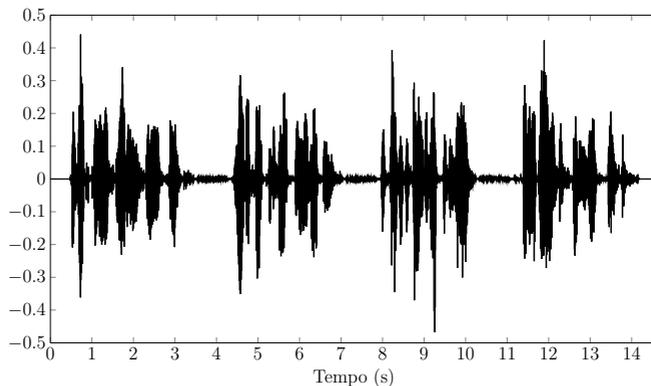
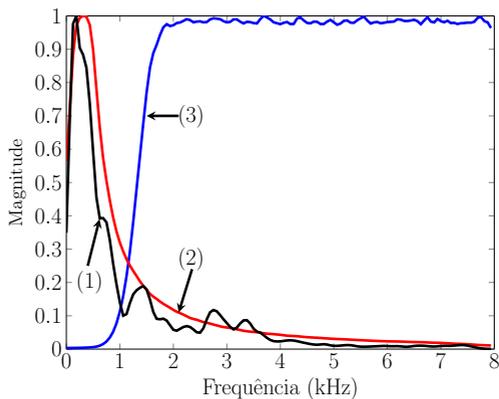


Figura 11 – Periodograma normalizado do sinal de fala (1), do ruído ICRA (2) e do ruído passa-altas (3).



4.1.3 Configuração do cenário acústico

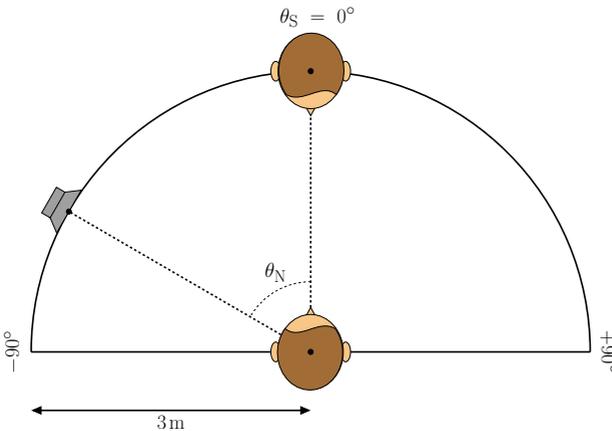
A Figura 12 apresenta as características gerais dos cenários acústicos utilizados nos experimentos de avaliação de desempenho. O ponto

médio da linha que conecta a entradas dos canais auditivos dos usuários de aparelhos auditivos é definido como ponto de referência. A fonte de fala é caracterizada pelo azimute θ_S e localiza-se exatamente a frente do ponto de referência, i.e., $\theta_S = 0^\circ$. A posição da fonte de ruído é caracterizada pelo azimute θ_N . Ambas as fontes possuem $\phi = 0^\circ$ (elevação nula) e estão localizadas a 3 metros do ponto de referência. Nessa configuração do cenário acústico, os ângulos negativos de azimute estão à esquerda da fonte de fala e, os positivos, estão à direita.

A criação dos cenários acústicos foi realizada através da convolução dos sinais (fala e ruído) com as respostas ao impulso associadas às respectivas HRTFs, descritas na Seção 4.1.1, relativas a uma distância de 3 metros, campo distante (ver Seção 2.3), elevação nula ($\phi = 0^\circ$) e azimute $\theta \in \{\theta_S, \theta_N\}$.

Na descrição dos resultados, os cenários acústicos criados foram caracterizados através da seguinte notação: S_0N_θ , onde S e N representam respectivamente as fontes de fala e de ruído e os subscritos representam o valor numérico do azimute dessas fontes. Como exemplo, o cenário acústico S_0N_{75} , possui a fonte de fala em 0° e a fonte de ruído em 75° .

Figura 12 – Características gerais dos cenários acústicos criados: elevação $\phi = 0^\circ$; fonte de fala, localizada em $\theta_S = 0^\circ$; fonte de ruído localizada no azimute θ_N ; ambas as fontes estão a 3 metros de distância do ponto de referência.



A frequência de amostragem utilizada foi 16 kHz. Os sinais foram transformados para o domínio da frequência através da STFT proposta em Crochiere (1980), utilizando-se o método de *overlap-and-add*

ponderado cujo critério de escolha dos coeficientes da janela de pesos é descrito por Smith (2011). Para a STFT, utilizou-se uma janela de análise de 128 pontos (8ms) e 50% de sobreposição (4ms). A SNR foi definida em 0 dB na orelha mais próxima da fonte de ruído em todos os cenários acústicos criados.

4.2 MÉTRICAS DE DESEMPENHO

Como consequência das definições de variáveis apresentadas no Capítulo 3, as figuras de mérito apresentadas na Seção 3.1 serão definidas de forma mais específica. A SNR de saída do sistema binaural nas orelhas esquerda (SNR_L) e direita (SNR_R) é definida como:

$$\begin{aligned}\text{SNR}_L &= \frac{E\{\|z_L^x\|^2\}}{E\{\|z_L^v\|^2\}} = \frac{\mathbf{w}_L^H \mathbf{R}_x \mathbf{w}_L}{\mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L}, \\ \text{SNR}_R &= \frac{E\{\|z_R^x\|^2\}}{E\{\|z_R^v\|^2\}} = \frac{\mathbf{w}_R^H \mathbf{R}_x \mathbf{w}_R}{\mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R}\end{aligned}\quad (4.1)$$

onde $z_L^x = \mathbf{w}_L^H \mathbf{x}$, $z_L^v = \mathbf{w}_L^H \mathbf{v}$ são as componentes de fala e ruído na saída do filtro de redução de ruído do lado esquerdo e $z_R^x = \mathbf{w}_R^H \mathbf{x}$, $z_R^v = \mathbf{w}_R^H \mathbf{v}$ são as componentes equivalentes para o lado direito. Os erros entre as diferença de níveis interaurais na entrada e na saída do sistema binaural, para o sinal de ruído ($\Delta\text{ILD}_{\text{No}}$) e para o sinal de fala ($\Delta\text{ILD}_{\text{Sp}}$), são dadas por:

$$\begin{aligned}\Delta\text{ILD}_{\text{No}} &= \frac{10}{K_F - K_I + 1} \sum_{K_I}^{K_F} \log_{10} \left(\frac{\sum_{\lambda} \mathbf{w}_L^H \mathbf{v} \mathbf{v}^H \mathbf{w}_L \sum_{\lambda} \mathbf{q}_R^T \mathbf{v} \mathbf{v}^H \mathbf{q}_R}{\sum_{\lambda} \mathbf{w}_R^H \mathbf{v} \mathbf{v}^H \mathbf{w}_R \sum_{\lambda} \mathbf{q}_L^T \mathbf{v} \mathbf{v}^H \mathbf{q}_L} \right), \\ \Delta\text{ILD}_{\text{Sp}} &= \frac{10}{K_F - K_I + 1} \sum_{K_I}^{K_F} \log_{10} \left(\frac{\sum_{\lambda} \mathbf{w}_L^H \mathbf{x} \mathbf{x}^H \mathbf{w}_L \sum_{\lambda} \mathbf{q}_R^T \mathbf{x} \mathbf{x}^H \mathbf{q}_R}{\sum_{\lambda} \mathbf{w}_R^H \mathbf{x} \mathbf{x}^H \mathbf{w}_R \sum_{\lambda} \mathbf{q}_L^T \mathbf{x} \mathbf{x}^H \mathbf{q}_L} \right).\end{aligned}\quad (4.2)$$

Ambas as métricas são calculadas de $K_I = 1.5 \text{ kHz}^3$ a $K_F = 8 \text{ kHz}$.

³Frequência na qual a diferença de tempo de chegada (ITD) das ondas sonoras de uma orelha em relação à outra se torna ambígua e a partir da qual as reflexões e refrações da cabeça e torso, que determinam a ILD, se tornam, portanto, importantes para a localização da fonte sonora (BLAUERT, 1997).

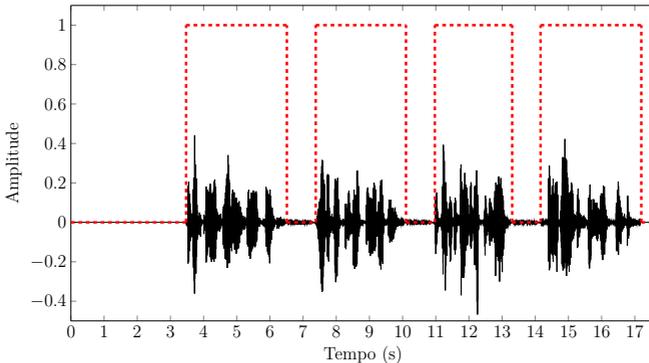
4.3 DETERMINAÇÃO DE PARÂMETROS

Esta seção apresenta os critérios e os experimentos utilizados para definição dos parâmetros necessários para se determinar as estimativas $\hat{\mathbf{R}}_y$, $\hat{\mathbf{R}}_x$ e $\hat{\mathbf{R}}_v$ e os coeficientes \mathbf{w}_L e \mathbf{w}_R . Os valores apresentados nessa seção serão utilizados nos experimentos de avaliação de desempenho do método proposto.

4.3.1 Estimação das matrizes de correlação

Para obter uma estimativa adequada dos valores iniciais das matrizes de correlação \mathbf{R}_y , \mathbf{R}_x e \mathbf{R}_v , acrescentou-se 3 segundos iniciais de silêncio ao sinal de fala apresentado na Seção 4.1.2. Além disso, tendo como objetivo analisar o limite superior de desempenho do método proposto, a determinação dos *frames* de fala+ruído e de ruído foi feita a partir de um VAD ideal. A Figura 13 apresenta o sinal do VAD ideal, determinado *a priori*, e o sinal de fala, contendo 3 segundos iniciais de silêncio. Como descrito na Seção 3.3.2, o sinal do VAD possui valor 1 na ocorrência de fala.

Figura 13 – Representação no tempo do sinal de fala (contínuo) e do sinal do VAD (tracejado).



Os valores iniciais de $\hat{\mathbf{R}}_{y,0}$, $\hat{\mathbf{R}}_{x,0}$ e $\hat{\mathbf{R}}_{v,0}$, definidos em (3.31), foram determinados a partir dos 6 segundos iniciais do sinal de entrada, i.e., $T_{VI} = 6$ segundos. Após este intervalo de tempo, iniciou-se a determinação das estimativas $\hat{\mathbf{R}}_{y,i}$, $\hat{\mathbf{R}}_{x,i}$ e $\hat{\mathbf{R}}_{v,i}$, a partir de (3.29) e (3.30).

Tendo em vista que técnicas baseadas no MWF necessitam de grandes janelas de tempo para estimação das matrizes de correlação (NHO, 2011) e que valores usuais de tempos de estimação estão entre 2 e 6 segundos (NHO, 2011; CORNELIS; MOONEN; WOUTERS, 2011; DOCLO et al., 2007; SPRIET; MOONEN; WOUTERS, 2005), neste trabalho utilizou-se coeficientes μ_x e μ_y iguais 0,999, valor associado a uma janela de tempo de 4 segundos⁴. Adicionalmente, utilizou-se $\mu_v = 0,995$, correspondente a uma janela de tempo de 0,8 segundos, tendo por objetivo fazer com o que o algoritmo adaptativo se adaptasse mais rapidamente à mudanças na posição da fonte de ruído.

4.3.2 Coeficiente de redução de ruído

A escolha do valor do passo de convergência do processo iterativo (β) foi feita utilizando-se $\gamma = 0$ em (3.28), i.e., a partir da versão adaptativa do filtro de Wiener. Para tanto, o cenário acústico virtual S_0N_{-60} foi gerado (Figura 14). Os sinais que caracterizam o cenário de interesse foram criados convolvendo-se os sinais de fala e ruído com as respostas ao impulso (transformada inversa de Fourier das HRTF) do caminho entre a posição da fonte sonora e os microfones dos aparelhos auditivos. O sinal de fala utilizado é o apresentado na Seção 4.1.2. Como sinal de ruído, utilizou-se os ruídos ICRA e passa-altas.

A escolha de β foi feita avaliando-se as seguintes métricas: 1) PESQ, 2) SNR, 3) Δ ISNR, 4) Δ ILD, para os sinais de fala (Δ ILD_{Sp}) e para os sinais de ruído (Δ ILD_{No}). A estimação das matrizes de correlação foi feita utilizando-se $T_{v1} = 6$ segundos (tem inicial do sinal de entrada) e coeficientes $\mu_y = 0,999$, $\mu_x = 0,999$ e $\mu_v = 0,995$. A estimação dos coeficientes dos filtros lineares foi feita utilizando-se: $N_i = 700$ e $\gamma = 0$. A partir disso, seis valores de β foram testados: $1 \cdot 10^{-3}$, $5 \cdot 10^{-3}$, $1 \cdot 10^{-2}$, $5 \cdot 10^{-2}$, $1 \cdot 10^{-1}$, $5 \cdot 10^{-1}$.

Os resultados das métricas de desempenho em função do passo de adaptação, para ambos os tipos de ruído, são apresentados na Figura 15. As curvas obtidas mostram que, para $\beta \geq 1 \cdot 10^{-1}$, todas as métricas apresentam valores não numéricos (valores ausentes nos gráficos). Desta forma, considerando o conjunto de parâmetros utilizados, adotou-se $\beta = 5 \cdot 10^{-2}$, nos experimentos realizados neste trabalho, valor no qual todas as métricas avaliadas apresentaram resultados numéricos. Este valor representa, portanto, metade do valor máximo

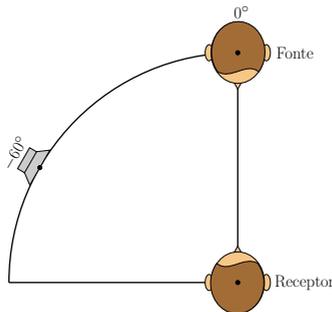
⁴O apêndice C apresenta uma equação para a determinação da janela de tempo a partir do valor do coeficiente de suavização.

de β associado à $\gamma = 0$.

4.3.3 Número de iterações

A escolha do número de iterações N_i foi realizada a partir do cenário acústico do experimento anterior (S_0N_{-60}). Os seguintes valores de N_i foram testados: 10, 50, 100, 200, 300, 400, 500, 600, 700. Para estimação das matrizes de correlação e estimação dos coeficientes do filtro linear utilizou-se: $\gamma = 0$, $\mu_x = \mu_y = 0,999$, $\mu_v = 0,995$ e $\beta = 5 \cdot 10^{-2}$. Os resultados do experimento, apresentadas na Figura 16, mostram que para $N_i \geq 100$ há pequenas variações em todas as métricas de desempenho, exceto para a ΔILD do ruído passa-altas. Contudo, experimentos de audição realizados pelo autor indicam que, a partir de $N_i = 200$, a versão adaptativa do filtro de Wiener desloca a posição aparente da fonte de ruído para a direção da fonte de fala, comportamento descrito teoricamente por Cornelis et al. (2010) e corroborado por experimentos subjetivos apresentados por Bogaert et al. (2008). Esse fato, indica a efetiva ação do processo de filtragem. Por isso, com o intuito de manter o custo computacional do método proposto o mais reduzido possível, adotou-se $N_i = 200$ em todos os experimentos apresentados nas partes subsequentes deste trabalho.

Figura 14 – Cenário acústico utilizado para escolha dos parâmetros β e N_i : fonte de fala em 0° , e fonte de ruído em -60° .



4.4 CONCLUSÕES

Este capítulo apresentou os materiais e métodos utilizados para verificar o desempenho do método proposto. Inicialmente, apresentou-se os elementos constituintes e as características dos cenários acústicos criados: funções de transferência, sinais de fala e ruído, azimute, elevação e distância das fontes. Em seguida, as métricas de desempenho, introduzidas na Seção 2.6, foram redefinidas a fim de ressaltar a forma de cálculo. Por fim, foram apresentados as justificativas e os experimentos utilizados para escolha: dos valores dos coeficientes de esquecimento, do tempo de determinação dos valores iniciais das matrizes de correlação, do coeficiente associado ao termo de redução de ruído e o número de iterações do algoritmo adaptativo.

Figura 15 – Resultados do experimento para a determinação do valor de β : PESQ, SNR, Δ ISNR, Δ ILD_{No} e Δ ILD_{Sp} obtidos a partir do cenário acústico SoN₆₀, criado com os ruídos ICRA (1) e passa-alta (2), em função de β .

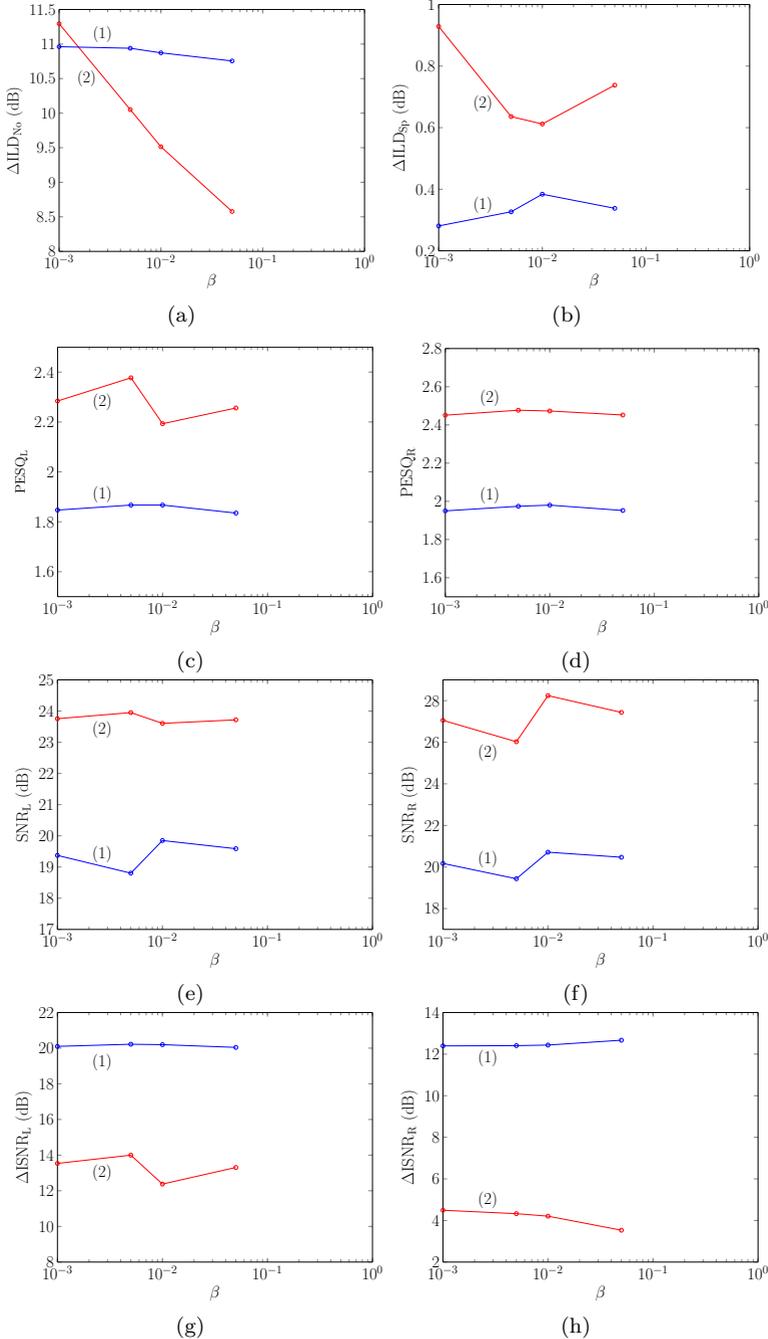
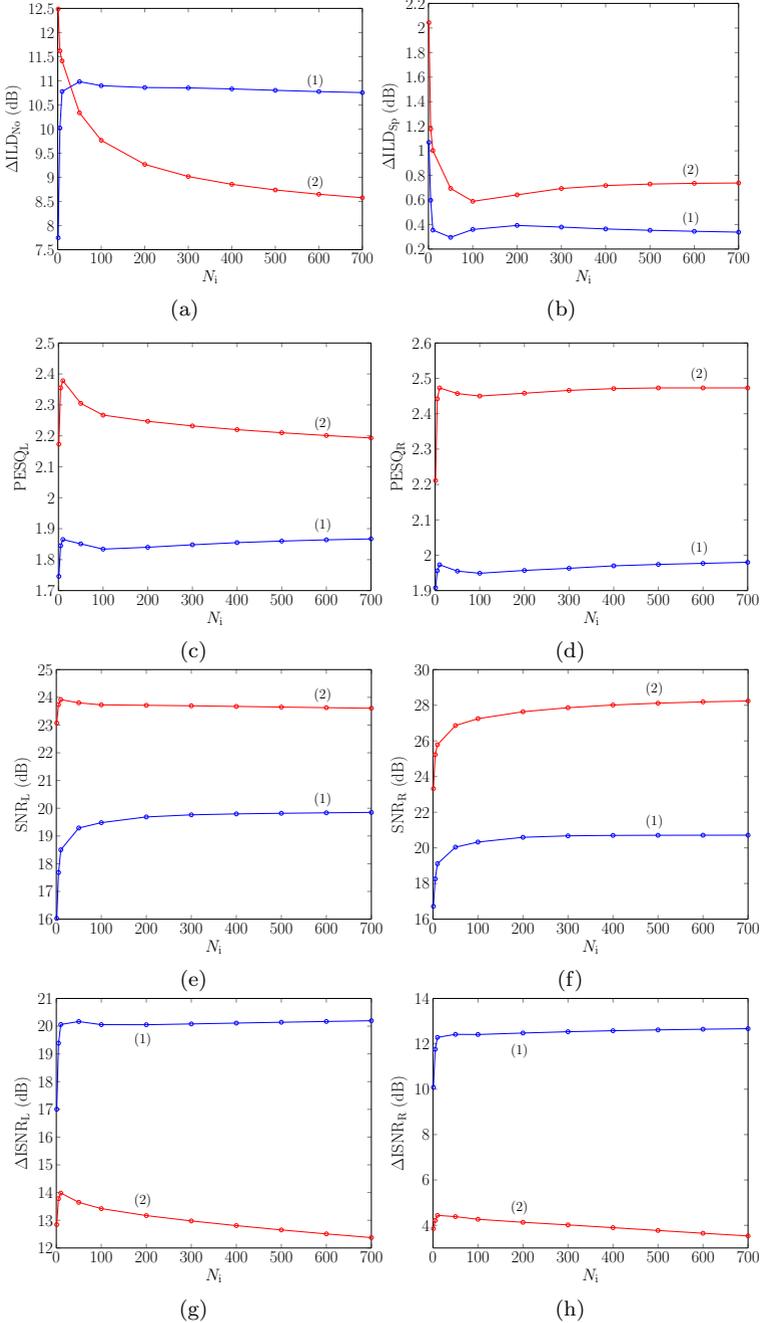


Figura 16 – Resultados do experimento para a determinação do valor de N_i : PESQ, SNR, ΔISNR , ΔILD_{N_o} e ΔILD_{S_p} obtidos a partir do cenário acústico S_0N_{60} , criado com os ruídos ICRA (1) e passa-alta (2), em função de N_i .



Capítulo 5

Resultados

Este capítulo apresenta três experimentos realizados para ilustrar a capacidade de lateralização do método proposto. As Seções 5.1, 5.2 e 5.3 descrevem os objetivos dos experimentos, a criação dos cenários acústicos e, por fim, apresentam os seus respectivos resultados e suas discussões. As características essenciais dos cenários acústicos em cada avaliação, o cálculo dos coeficientes do filtro adaptativo e o processo de estimação das matrizes de correlação foram apresentados no Capítulo 4.

Os resultados apresentados neste capítulo têm por finalidade ilustrar a capacidade de lateralização do método proposto, denominado MWF-ILD_{AF}. Todavia, com o intuito de fornecer um parâmetro de comparação com a abordagem clássica, conjuntamente aos resultados do MWF-ILD_{AF}, serão apresentados os resultados do filtro de Wiener adaptativo, que corresponde à versão sem preservação da ILD do método proposto ($\gamma = 0$), denominado de MWF_{AF}. Em todas as avaliações realizadas utilizou-se: $\beta = 5 \cdot 10^{-2}$, $N_i = 200$, $\mu_y = 0,999$, $\mu_x = 0,999$ e $\mu_v = 0,995$.

5.1 AVALIAÇÃO 1: MÉTRICAS OBJETIVAS

O objetivo da Avaliação 1 é verificar, de forma qualitativa e quantitativa, o desempenho do MWF-ILD_{AF} através de quatro métricas objetivas (ver Seções 2.6 e 4.2): 1) PESQ (qualidade de fala); 2) SNR (redução de ruído); 3) Δ ISNR (ganho de inteligibilidade do sistema binaural); 4) Δ ILD, métrica de preservação da ILD, que quantifica a variação entre a ILD da entrada e da saída do sistema binaural e é determinada para o sinal de fala (Δ ILD_{Sp}) e para o sinal de ruído (Δ ILD_{No}). Nesta avaliação, três cenários acústicos de interesse foram gerados: S₀N₋₆₀, S₀N₋₄₅ e S₀N₋₃₀, os quais estão sobrepostos na Figura 17. Os sinais gerados por esses cenários foram processados pelo MWF_{AF} ($\gamma = 0$) e pelo método proposto, utilizando valores de γ na

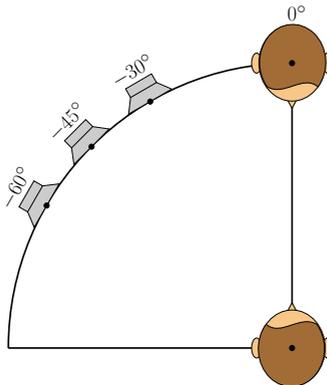
faixa entre $9 \cdot 10^{-11}$ e 10^2 , a fim de fornecer uma macro visão do comportamento das métricas avaliadas. Após esse processamento, as curvas de desempenho foram traçadas a partir dos resultados obtidos para cada valor de γ .

A avaliação qualitativa do desempenho foi realizada a partir do comportamento individual das curvas e das relações de compromisso determinadas a partir da comparações dessas curvas. A avaliação quantitativa foi feita comparando-se os resultados das métricas para um determinado valor de γ diferente de 0 ($MWF\text{-}ILD_{AF}$) com os resultados obtidos com $\gamma = 0$ (MWF_{AF}).

Nos três cenários acústicos, utilizou-se o sinal apresentado na Seção 4.3 como sinal emitido pela fonte de fala. Ele possui 17,7 segundos de duração, sendo 3 segundos iniciais de silêncio. Para simular o sinal emitido pela fonte de ruído, utilizou-se os ruídos ICRA e passa-altas com a mesma duração do sinal de fala. A partir desses sinais, a geração dos cenários acústicos seguiu a descrição apresentada na Seção 4.1.3. As métricas de desempenho foram medidas nos 4 segundos finais dos sinais gerados pelos cenários acústicos.

Considerando os três cenários de interesse e os dois tipos de ruído, seis cenários acústicos diferentes foram avaliados: três cenários acústicos criados com ruído ICRA, e três cenários acústicos criados com ruído passa-altas. A fim de fornecer informação sobre os sinais antes do processamento binaural, apresentam-se também a PESQ e a SNR dos sinais gerados pelos cenários acústicos antes do processamento pelas técnicas analisadas neste capítulo.

Figura 17 – Cenário acústico da Avaliação 1: fonte de fala em 0° , e fonte de ruído nos azimutes -60° , -45° e -30° .



5.1.1 Resultados e discussão

A Tabela 5 apresenta a $PESQ_L$, $PESQ_R$, SNR_L e SNR_R para os sinais antes do processamento binaural – sinais gerados pelos seis diferentes cenários acústicos. Os resultados na tabela indicam que, para ambos os tipos de ruído a qualidade percebida da fala é muito baixa em ambas as orelhas ($PESQ_L, PESQ_R \leq 1,45$). Adicionalmente, como apresentado na Seção 4.1.3, em todos os cenários acústicos a SNR na orelha ipsilateral à fonte de ruído (orelha esquerda) é igual a 0 dB, i.e., $SNR_L = 0$ dB. Os valores de SNR_R na orelha contralateral (orelha direita), em todos os cenários acústicos, resultam, principalmente, do efeito de sombra da cabeça, apresentado na Seção 2.4.

Tabela 5 – $PESQ_L$, $PESQ_R$, SNR_L e SNR_R para sinais gerados pelos três cenários acústicos, antes do processamento binaural.

Ruído ICRA			
Cenários Acústicos	S_0N_{-60}	S_0N_{-45}	S_0N_{-30}
$PESQ_L$	1,10	1,09	1,09
$PESQ_R$	1,22	1,22	1,18
SNR_L [dB]	0	0	0
SNR_R [dB]	3,79	3,72	3,10
Ruído Passa-Altas			
Cenários Acústicos	S_0N_{-60}	S_0N_{-45}	S_0N_{-30}
$PESQ_L$	1,07	1,07	1,07
$PESQ_R$	1,45	1,33	1,19
SNR_L [dB]	0	0	0
SNR_R [dB]	13,43	11,62	8,69

As Figuras 18 e 19 apresentam as métricas de desempenho em função de γ . As subfiguras (a) e (b) apresentam a variação da ΔILD_{No} e da ΔILD_{Sp} , respectivamente para os ruídos ICRA e passa-altas, nos três cenários acústicos avaliados. As subfiguras de (c) a (h) apresentam as curvas da $PESQ$, SNR e da $\Delta ISNR$ associadas às orelhas esquerda e direita para os ruídos ICRA e passa-altas. Os resultados obtidos mostram que para $\gamma > 10^{-6}$ (ruído ICRA) e $\gamma > 10^{-5}$ (ruído passa altas) há uma redução de ΔILD_{No} e aumento de ΔILD_{Sp} , em relação à condição $\gamma = 0$, para todas as curvas apresentadas. Estes resultados indicam que a ΔILD do ruído processado pelo MWF-ILD_{AF} é menor do que a ΔILD do ruído processado pelo MWF_{AF}, e, portanto, a sensação subjetiva da fonte de ruído provavelmente estará mais próxima

da posição original desta fonte. Como efeito colateral, há um aumento da ΔILD do sinal de fala processado pelo $\text{MWF-ILD}_{\text{AF}}$, quando comparada à ΔILD da fala processada pelo MWF_{AF} . Note, entretanto, que esse aumento da $\Delta\text{ILD}_{\text{Sp}}$ só é substancial para $\gamma > 1$. Adicionalmente, na orelha esquerda, a partir de $\gamma > 10^{-4}$ há uma diminuição, em relação ao MWF_{AF} ($\gamma = 0$), das métricas de qualidade de fala (PESQ), de redução de ruído (SNR) e de ganho de inteligibilidade do sistema (ΔISNR) em todos os cenários acústicos analisados. Este mesmo comportamento ocorre na orelha direita para $\gamma > 10^{-1}$. Portanto, os resultados apresentados nas Figuras 18 e 19 mostram que a variação de γ estabelece um compromisso entre a preservação da ILD do ruído e as demais métricas avaliadas.

A Tabela 6 apresenta uma avaliação quantitativa do desempenho do $\text{MWF-ILD}_{\text{AF}}$ e do desempenho do MWF_{AF} nos três cenários de interesse. Para o método proposto, utilizou-se $\gamma = 3 \cdot 10^{-2}$, para a avaliação com ruído ICRA, e $\gamma = 3,5 \cdot 10^{-2}$, para a avaliação com ruído passa-altas. Estes valores de γ foram escolhidos, pois resultam em valores aproximadamente iguais de SNR na contralateral, para ambos os métodos,. Esta situação estabelece a condição de idêntica redução de ruído na orelha que apresenta melhor relação sinal ruído para os sinais processados pelo MWF_{AF} e $\text{MWF-ILD}_{\text{AF}}$. Note que nesse caso os resultados da PESQ_{R} dos dois métodos são também aproximadamente iguais. A partir disso, comparando-se o desempenho do $\text{MWF-ILD}_{\text{AF}}$ em relação ao desempenho do MWF_{AF} , é possível afirmar que:

- Há uma diminuição de até 6,96 dB (S_0N_{-60}) na $\Delta\text{ILD}_{\text{No}}$ para a avaliação realizado com ruído ICRA, e de até 6,33 dB (S_0N_{-60}) para a avaliação realizado com ruído passa-altas.
- Há um aumento de até 0,98 dB (S_0N_{-60}) na $\Delta\text{ILD}_{\text{Sp}}$ para o ruído ICRA, e de até 0,86 dB (S_0N_{-60}) para o ruído passa-altas. Audições informais realizadas pelo autor indicam que o incremento apresentado não é suficiente para resultar em alteração da lateralização da fala.
- Na orelha ipsilateral à fonte de ruído, há uma uma redução de 0,26 PESQ, 2,99 dB de SNR e 5,87 dB de ΔISNR , para o ruído ICRA; e de 0,74 PESQ, 9,25 dB de SNR e 5,26 dB de ΔISNR , para o ruído passa-alta. Todos esses resultados foram obtidos no cenário acústico S_0N_{-60} .
- Na orelha contralateral à fonte de ruído: PESQ e SNR mantêm-se aproximadamente constantes em todos os casos analisados; e

o ΔISNR apresenta um aumento de até 2,63 dB (S_0N_{-60}) no caso do ruído ICRA e de até 1,86 dB (S_0N_{-45}) no caso do ruído passa-altas.

Portanto, o método proposto mantém a ΔILD do ruído mais próxima de seu valor original (antes do processamento binaural), se comparado ao MWF_{AF} . Isto resulta na preservação da lateralidade da fonte de ruído pelo $\text{MWF-ILD}_{\text{AF}}$, como será apresentado através de avaliações subjetivas e de testes de hipóteses. Adicionalmente, o $\text{MWF-ILD}_{\text{AF}}$ apresenta valores de $\Delta\text{ILD}_{\text{Sp}}$ inferiores a 1,48 dB, indicando que a fonte de fala se mantém próxima de sua posição original no azimute 0° . Em relação à orelha ipsilateral, há uma diminuição da qualidade de fala, da redução de ruído e do ganho de inteligibilidade do sistema binaural. Contudo, na orelha contralateral a qualidade de fala e a redução de ruído se mantêm aproximadamente constantes e há um aumento de até 2,63 dB no $\Delta\text{ISNR}_{\text{R}}$.

Figura 18 – $\Delta\text{ILD}_{\text{No}}$, $\Delta\text{ILD}_{\text{Sp}}$, PESQ, SNR e ΔISNR obtidos partir da avaliação com os cenários acústicos: (1) S_0N_{-60} , (2) S_0N_{-45} e (3) S_0N_{-30} , criados com ruído ICRA.

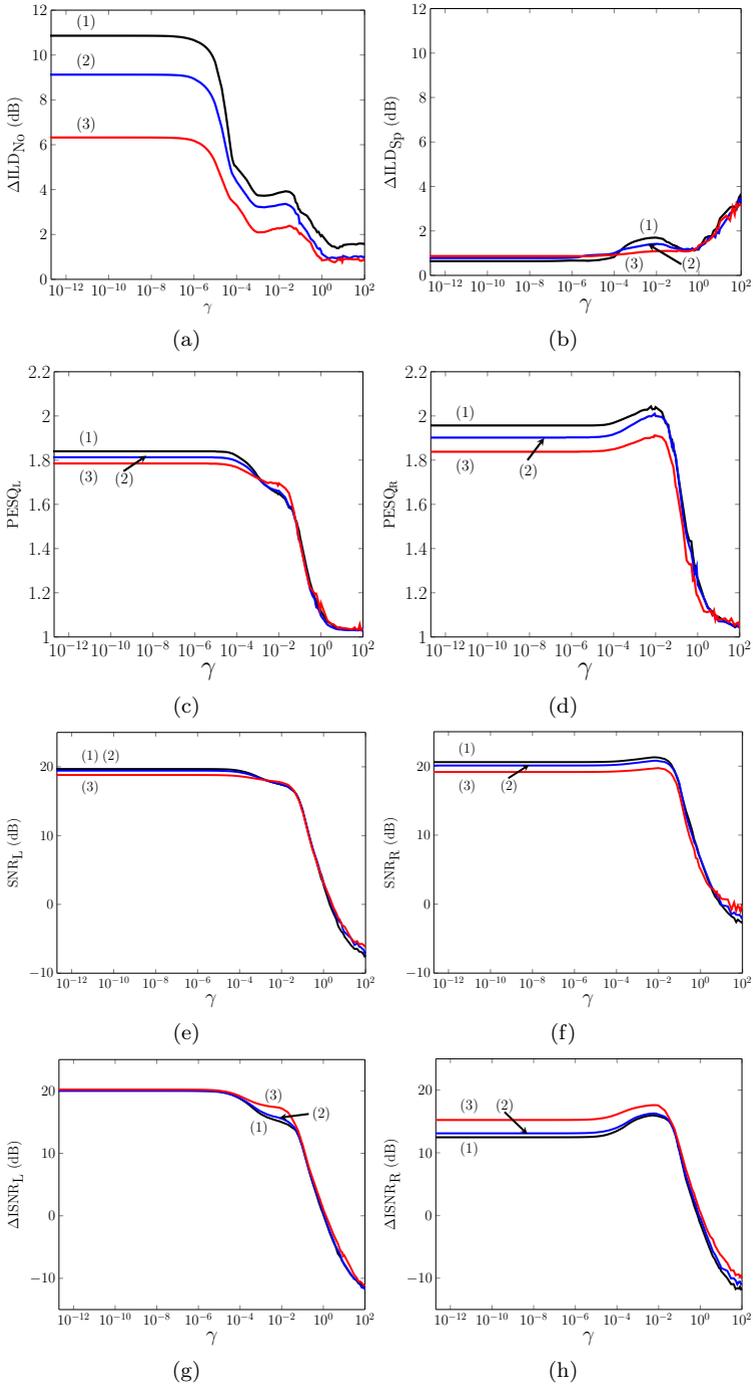


Figura 19 – $\Delta\text{ILD}_{\text{No}}$, $\Delta\text{ILD}_{\text{Sp}}$, PESQ, SNR e ΔISNR obtidos partir da avaliação com os cenários acústicos: (1) S_0N_{-60} , (2) S_0N_{-45} e (3) S_0N_{-30} , criados com ruído passa-altas.

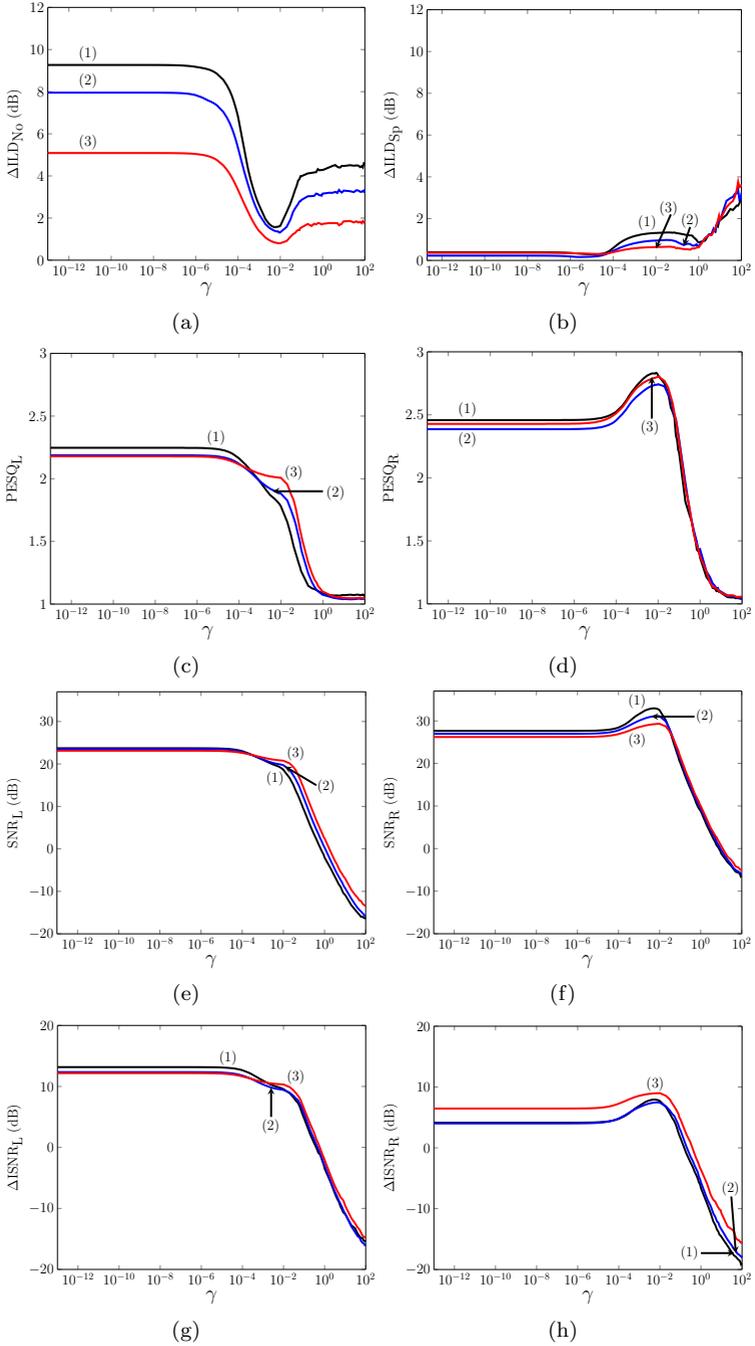


Tabela 6 – Comparação entre os métodos MWF_{AF} e $MWF-ILD_{AF}$ nos cenários acústicos criados com ruído ICRA e passa-altas.

Ruído ICRA						
Cenário Acústico	S_0N_{-60}		S_0N_{-45}		S_0N_{-30}	
Método	MWF_{AF}	$MWF-ILD_{AF}$	MWF_{AF}	$MWF-ILD_{AF}$	MWF_{AF}	$MWF-ILD_{AF}$
ΔILD_{No} [dB]	10,86	3,90	9,12	3,29	6,32	2,38
ΔILD_{Sp} [dB]	0,39	1,33	0,22	0,98	0,36	0,65
PESQ _L	1,84	1,58	1,81	1,59	1,78	1,65
PESQ _R	1,95	1,96	1,90	1,97	1,83	1,88
SNR _L [dB]	19,68	16,69	19,44	16,70	18,81	16,93
SNR _R [dB]	20,59	20,70	20,09	20,33	19,15	19,22
$\Delta ISNR_L$ [dB]	20,05	14,18	19,98	14,59	20,19	15,68
$\Delta ISNR_R$ [dB]	12,47	15,10	13,10	15,20	15,23	15,71

Ruído Passa-Altas						
Cenário Acústico	S_0N_{-60}		S_0N_{-45}		S_0N_{-30}	
Método	MWF_{AF}	$MWF-ILD_{AF}$	MWF_{AF}	$MWF-ILD_{AF}$	MWF_{AF}	$MWF-ILD_{AF}$
ΔILD_{No} [dB]	9,26	2,93	7,95	2,07	5,08	1,09
ΔILD_{Sp} [dB]	0,64	1,48	0,78	1,28	0,86	1,10
PESQ _L	2,24	1,50	2,18	1,72	2,17	1,84
PESQ _R	2,45	2,64	2,38	2,62	2,42	2,68
SNR _L [dB]	23,71	14,46	23,51	16,87	23,12	18,92
SNR _R [dB]	27,63	26,95	26,95	27,49	26,20	27,08
$\Delta ISNR_L$ [dB]	13,16	7,90	12,35	8,21	12,18	9,11
$\Delta ISNR_R$ [dB]	4,13	5,20	4,01	5,87	6,46	7,64

5.2 AVALIAÇÃO 2: TESTE SUBJETIVO

O propósito da Avaliação 2 é verificar a capacidade de lateralização do método proposto, a partir de uma avaliação subjetiva de determinação perceptual do azimute de fontes de ruído. As seções a seguir detalham os aspectos éticos, o ambiente simulado, a metodologia e os resultados obtidos no teste subjetivo de lateralização.

5.2.1 Aspectos éticos

Por envolver a participação de pessoas, a avaliação descrita nesta seção seguiu as normas descritas na Resolução Nº 466/12, do Conselho Nacional de Saúde (CSN), que determina que toda pesquisa que tenha participação de seres humanos deve respeitar aspectos éticos e ser submetida a um comitê de ética para avaliação de seu mérito (CNS

Nº 466/12, 2012). Por isso, a avaliação realizada com os voluntários foi enviada, sob a forma de um projeto de pesquisa, ao Comitê de Ética em Pesquisa com Seres Humanos (CEPSH) da Universidade Federal de Santa Catarina. Após a avaliação do projeto, o CEPSH emitiu parecer (número 1.291.583) favorável à realização do experimento. Tendo o parecer favorável do CEPSH, a aplicação do teste necessitou também da anuência formal dos voluntários através da assinatura do Termo de Consentimento Livre e Esclarecido (TCLE)(ANEXO A), que descreve, dentre outros aspectos, a razão pela qual a pesquisa foi realizada, os riscos presentes durante a avaliação e os direitos do voluntário ao participar da pesquisa. No teste conduzido, os 30 voluntários manifestaram concordância com a realização dos experimentos após lerem e assinarem o TCLE.

5.2.2 Ambiente acústico simulado

A avaliação para a determinação perceptual do azimute de fontes de ruído foi realizada considerando-se um ambiente acústico simulado constituído por 13 fontes sonoras virtuais, espaçadas 15° uma da outra e localizadas no plano horizontal. A partir de uma destas fontes, um sinal de ruído, do tipo ICRA ou passa-altas, é emitido e percorre o caminho acústico simulado entre a fonte sonora e os microfones do sistema de processamento binaural, apresentado na Seção 3.1. O sinal de ruído é filtrado pelo sistema de processamento e enviado para os fones de ouvido do voluntário da avaliação, que se localiza virtualmente a 3 metros distante de todas as fontes sonoras, como ilustrado na Figura 20.

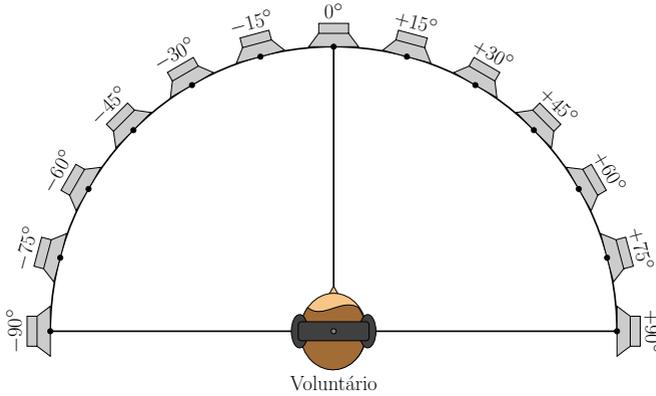
Considerando as fontes sonoras individualmente, o ambiente simulado é constituído de 13 cenários acústicos individuais, caracterizados pelos azimutes das fontes de ruído. Desta forma, o ruído emitido por uma das fontes é gerado envolvendo-se 17,7 segundos¹ dos ruídos ICRA e passa-altas com as respostas ao impulso associadas a cada um dos microfones do sistema binaural.

Os sinais utilizados durante as avaliações constituem um conjunto total de 38 ruídos divididos em dois grupos: I) grupo dos ruídos ICRA, contendo 19 sinais, e II) grupo de ruídos passa-altas, contendo também 19 sinais. Cada grupo individual é composto por dois tipos de sinais:

1. Sinais de ruído não processado (NP): ruído emitido por uma das

¹Tamanho do sinal de fala utilizado na Avaliação 1.

Figura 20 – Ambiente acústico virtual utilizado nas avaliações de lateralização.



13 fontes sonoras e filtrado com $\mathbf{w}_L(\lambda, k) = [1, 0, 0, 0, 0, 0]^T$ e $\mathbf{w}_R(\lambda, k) = [0, 0, 0, 1, 0, 0]^T$, filtros que transferem os sinais dos microfones de referência para a saída do sistema binaural, conectada ao fone de ouvido do voluntário. Cada um dos dois grupos de ruídos utilizados na avaliação de lateralização contém 13 sinais não processados.

2. Sinais de ruído processados: ruído emitido pelas fontes sonoras nos azimutes: -60° , -45° e -30° , e filtrado com $\mathbf{w}_L(\lambda, k)$ e $\mathbf{w}_R(\lambda, k)$ estimados a partir dos métodos MWF_{AF} e $\text{MWF-ILD}_{\text{AF}}$. Os coeficientes $\mathbf{w}_L(\lambda, k)$ e $\mathbf{w}_R(\lambda, k)$ utilizados na Avaliação 2 foram determinados (estimados e armazenados) durante a realização da Avaliação 1, assim como ilustrado na Figura 21. Cada um dos dois grupos de ruídos contém 6 sinais processados: 3 sinais de ruído processados por MWF_{AF} e 3 sinais de ruído processados por $\text{MWF-ILD}_{\text{AF}}$.

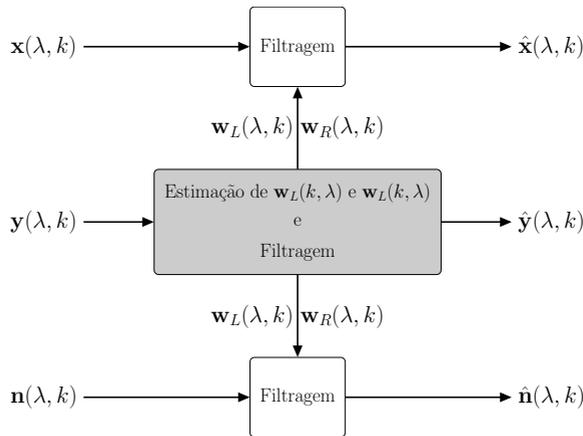
Portanto, dos 19 sinais em cada grupo, 13 são sinais não processados (NP) e 6 são sinais processados. A avaliação foi realizada reproduzindo isoladamente esses sinais de ruído, tendo como objetivo evitar a influência do fenômeno de mascaramento² pelo sinal de fala. Considerando os 17,7 segundos dos sinais de ruído gerados para as avaliações, apenas os 4 segundos finais foram apresentados aos voluntários da avaliação através de um fone de ouvido Sennheiser HD 202. Devido à

²Redução da percepção de um som de interesse devido a presença de outro som, denominado mascarador (GELFAND, 2010).

redução de potência, decorrente do processo de filtragem, os sinais de ambos canais foram amplificados por um mesmo fator de normalização, de forma a facilitar na identificação da localização de fonte sonora.

O grupo de voluntários da avaliação é composto de 30 pessoas divididas em dois grupos de 15 pessoas: o primeiro grupo é composto exclusivamente de homens, o segundo grupo é composto de 13 homens e 2 mulheres. Um tipo de ruído (ICRA ou passa-altas) foi utilizado na avaliação realizada com cada grupo.

Figura 21 – Geração dos sinais de ruído utilizados nas avaliações de localização intracraniana.



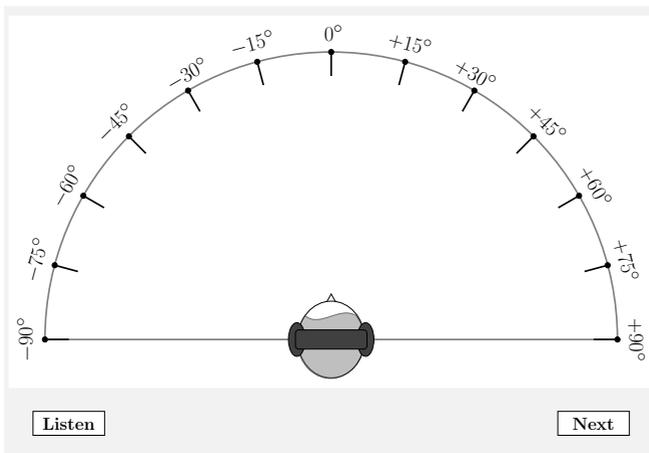
5.2.3 Metodologia de avaliação

Devido à facilidade de entendimento de uma avaliação de localização espacial de fonte sonora (localização extracraniana), a avaliação de lateralização (localização intracraniana) realizada foi conduzida como uma avaliação de determinação perceptual de azimuth, mesmo considerando a utilização de fones de ouvido pelos voluntários. Dessa forma, a avaliação foi dividida em três etapas, em cada uma delas uma *interface* gráfica foi utilizada para facilitar e agilizar o desenvolvimento das atividades. O objetivo de cada uma das três etapas, assim como o funcionamento das *interfaces* associadas a cada delas, é descrito a seguir:

1. **Etapa de Aprendizagem:** o objetivo desta etapa é ambien-

tar o participante da avaliação com o processo de determinação da localização (lateralização) das 13 posições do cenário acústico virtual. Para isso, a *interface* presente na Figura 22 foi apresentada ao voluntário. A partir dela, o voluntário foi orientado a pressionar o botão *Listen* para que ouvisse cada um dos 13 sinais não processados associados a cada um dos 13 azimutes. Após a apresentação de todos os sinais, o voluntário deveria pressionar o botão *next*, para prosseguir para a etapa seguinte da avaliação.

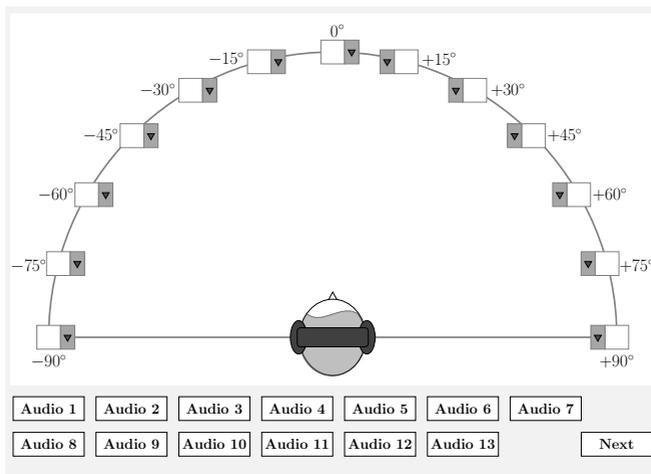
Figura 22 – Interface gráfica utilizada na etapa de aprendizagem da avaliação de lateralização.



- 2. Etapa de Treinamento:** o objetivo desta etapa é avaliar a habilidade do voluntário em identificar ruídos emitidos a partir das 13 posições apresentadas na etapa de aprendizagem. Para isso, a *interface* presente na Figura 23 foi apresentada ao participante. A partir dela, o participante deveria pressionar um dos 13 botões, nomeados: *Audio 1*, *Audio 2*, *Audio 3*, etc; e ouvir os ruídos associados a cada um deles. Um sinal de ruído não processado da etapa de aprendizagem foi associado a cada botão. Após ouvir o áudio, o participante deveria atribuir o número do áudio a um dos 13 azimutes apresentados. Cada áudio ouvido deveria necessariamente ser atribuído a um único azimute. Nesta etapa, os usuários puderam escutar mais de uma vez cada áudio, fazer comparações entre eles e modificar uma escolha já feita. Após atribuído a cada azimute um determinado áudio, o usuário deveria pressionar *Next*,

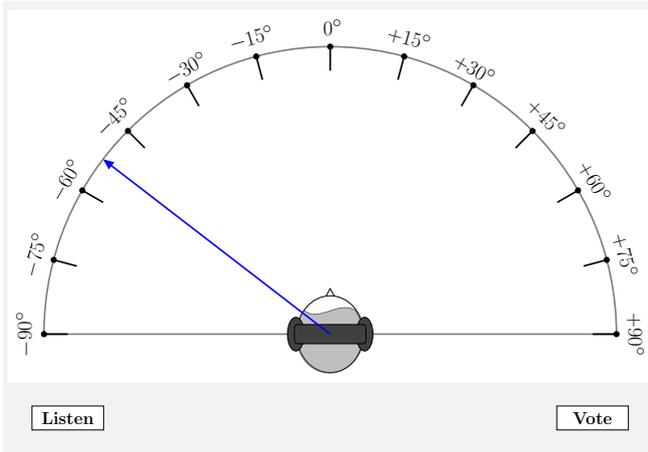
para prosseguir para a etapa de testes.

Figura 23 – Interface gráfica utilizada na etapa de treinamento.



3. **Etapa de Teste:** o objetivo desta etapa foi determinar o azimute percebido de um conjunto de 19 sinais: 13 sinais não processados, 3 sinais processados pelo MWF_{AF} e 3 sinais processados pelo $MWF-ILD_{AF}$. Para isso, a *interface* na Figura 24 foi utilizada. A partir dela, o participante deveria pressionar o botão *Listen* e escutar um dos 16 ruídos apresentados de forma aleatória. Em seguida, através da seta de seleção, deveria determinar o provável azimute da fonte geradora do som. Por fim, era necessário confirmar escolha do azimute pressionando o botão *Vote*. O azimute escolhido poderia ser qualquer um entre -90° e 90° , tendo em vista que as técnicas de processamento podem modificar o azimute original da fonte de ruído para um azimute diferente dos 13 inicialmente apresentados. A resolução de seleção de azimute foi de 1° . Devido à dificuldade desta etapa, foi permitido aos voluntários ouvir mais de uma vez cada um dos áudios, contudo, uma vez realizada a escolha do azimute, não foi permitida alteração.

Figura 24 – Interface gráfica utilizada na etapa de teste.



5.2.4 Resultados e discussão

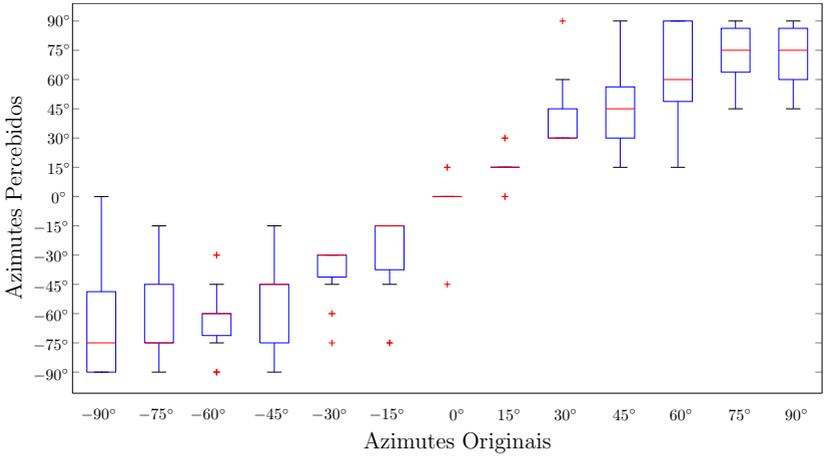
As Figuras 25 e 26 apresentam os diagramas de caixa das medidas de lateralização obtidas nas etapas de treinamento e teste utilizando-se os sinais não processados. A análise desses diagramas será realizada a partir dos resultados apresentados na Tabela 7, que apresenta a média, a mediana e o desvio padrão obtidos a partir dos conjuntos de votos representados pelos diagramas de caixa. Na tabela, a seguinte nomenclatura é apresentada:

- μ_{TR} , $\tilde{\mu}_{TR}$ e σ_{TR} são, respectivamente, os azimutes médio, mediano e o desvio padrão das medidas de percepção obtidas a partir dos sinais não processados na etapa de treinamento da avaliação de lateralização.
- μ_{TS} , $\tilde{\mu}_{TS}$ e σ_{TS} são, respectivamente, os azimutes médio, mediano e o desvio padrão das medidas de percepção obtidas a partir dos sinais não processados na etapa de teste da avaliação de lateralização.

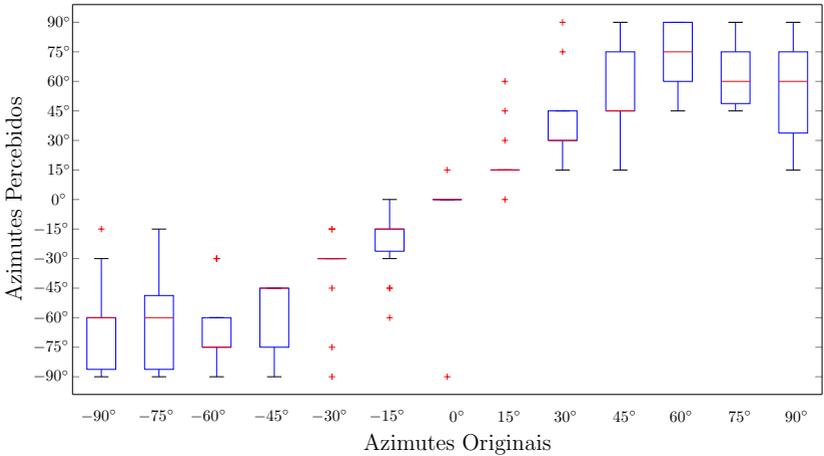
Os resultados na Tabela 7 mostram a capacidade dos voluntários de determinar a localização das fontes sonoras, tanto na etapa de treinamento quanto na etapa de teste, i.e., para os ruídos emitidos a partir das fontes nos azimutes negativos ($\theta_N < 0^\circ$), os azimutes médio e mediano percebidos localizam-se a esquerda da linha mediana da cabeça

($\mu_{TR}, \tilde{\mu}_{TR}, \mu_{TS}, \tilde{\mu}_{TS} < 0^\circ$) e para os ruído emitidos a partir dos azimutes positivos ($\theta_N > 0^\circ$), os azimutes percebidos se localizam à direita da linha mediana da cabeça ($\mu_{TR}, \tilde{\mu}_{TR}, \mu_{TS}, \tilde{\mu}_{TS} > 0^\circ$), para os dois tipos de ruídos.

Figura 25 – Diagrama de caixa dos votos obtidos a partir dos sinais de ruído não processados da etapa de treinamento: (a) ruído ICRA e (b) ruído Passa-Altas.



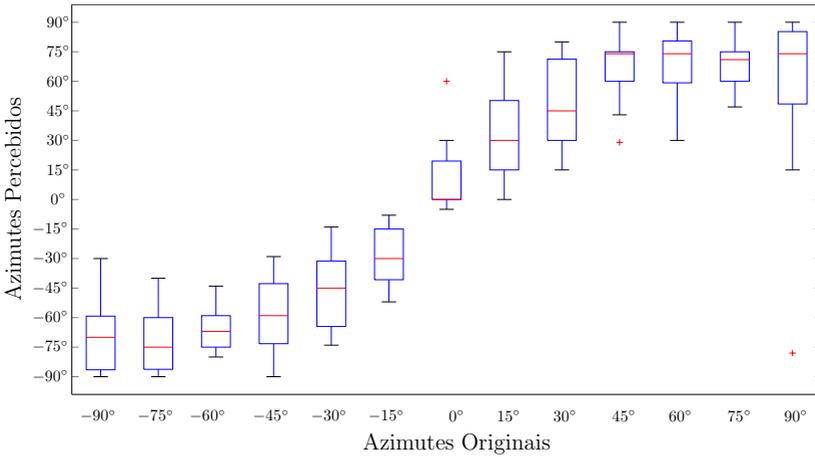
(a)



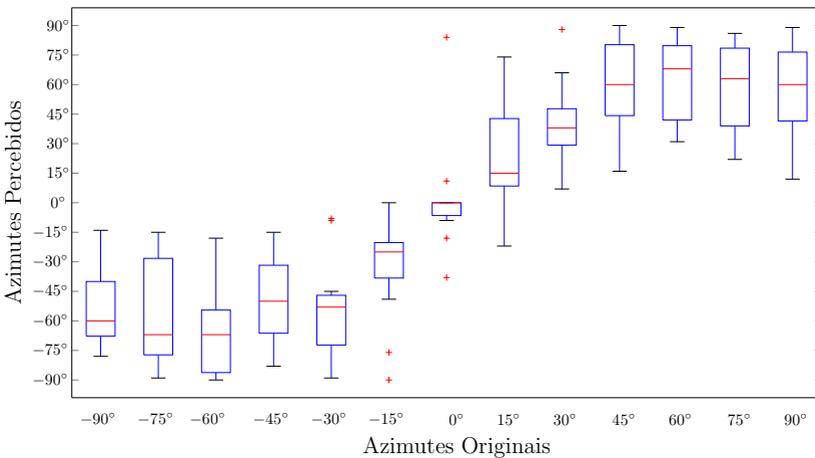
(b)

A Tabela 8 apresenta o MAE das etapas de treinamento (MAE_{TR}) e de teste (MAE_{TS}) para os dois grupos de participantes: grupo que realizou a avaliação com ruído ICRA e grupo que realizou a avaliação com o ruído passa-altas. Os resultados apresentados mostram um aumento

Figura 26 – Diagrama de caixa dos votos obtidos a partir dos sinais de ruído não processados da etapa de testes: (a) ruído ICRA e (b) ruído Passa-Altas.



(a)



(b)

do MAE na etapa de testes em comparação com a MAE na etapa de treinamento para todos os voluntários analisados, exceto para os casos dos voluntários 1, 8 e 12. Este comportamento se deve provavelmente à

Tabela 7 – Média (μ_p), mediana ($\tilde{\mu}_p$) e desvio padrão (σ_p) das medidas de lateralização obtidas a partir dos sinais não processados nas etapas de treinamento e de teste.

Ruído ICRA						
	Treinamento			Teste		
	μ_{TR}	$\tilde{\mu}_{TR}$	σ_{TR}	μ_{TS}	$\tilde{\mu}_{TS}$	σ_{TS}
-90°	-65°	-60°	23°	-53°	-60°	21°
-75°	-63°	-60°	23°	-56°	-67°	27°
-60°	-66°	-75°	22°	-67°	-67°	21°
-45°	-58°	-45°	18°	-51°	-50°	22°
-30°	-35°	-30°	21°	-55,26°	-53°	24°
-15°	-22°	-15°	16°	-32°	-25°	24°
0°	-5°	0°	24°	2°	0°	26°
15°	20°	15°	15°	26°	15°	28°
30°	39°	30°	20°	41°	38°	20°
45°	56°	45°	24°	59°	60°	24°
60°	76°	75°	16°	63°	68°	21°
75°	66°	60°	17°	59°	63°	21°
90°	57°	60°	23°	60°	60°	23°

Ruído Passa-Altas						
	Treinamento			Teste		
	μ_{TR}	$\tilde{\mu}_{TR}$	σ_{TR}	μ_{TS}	$\tilde{\mu}_{TS}$	σ_{TS}
-90°	-66°	-75°	31°	-69°	-70°	18°
-75°	-64°	-75°	24°	-72°	-75°	17°
-60°	-62°	-60°	19°	-66°	-67°	11°
-45°	-55°	-45°	24°	-58°	-59°	21°
-30°	-38°	-30°	15°	-48°	-45°	21°
-15°	-27°	-15°	23°	-29°	-30°	14°
0°	-1°	0°	14°	12°	0°	18°
15°	15°	15°	9°	32°	30°	23°
30°	42°	30°	18°	48°	45°	22°
45°	47°	45°	21°	67°	74°	18°
60°	65°	60°	24°	67°	74°	18°
75°	74°	75°	14°	68°	71°	14°
90°	70°	75°	14°	60°	74°	44°

maior dificuldade da etapa de testes (e.g., usuários não puderam fazer comparações entre áudios) se comparada com a etapa de treinamento.

Tabela 8 – Erro absoluto médio (MAE) dos 30 voluntários da avaliação nas etapas de Treinamento (MAE_{TR}) e de Teste (MAE_{TS}) obtido a partir dos sinais não processados.

Ruído ICRA			Ruído Passa-Altas		
	MAE_{TR}	MAE_{TS}		MAE_{TR}	MAE_{TS}
Voluntário 1	24°	17°	Voluntário 16	19°	20°
Voluntário 2	10°	19°	Voluntário 17	28°	25°
Voluntário 3	14°	15°	Voluntário 18	19°	25°
Voluntário 4	7°	11°	Voluntário 19	12°	16°
Voluntário 5	12°	23°	Voluntário 20	10°	23°
Voluntário 6	5°	19°	Voluntário 21	19°	21°
Voluntário 7	26°	34°	Voluntário 22	3°	20°
Voluntário 8	33°	17°	Voluntário 23	14°	23°
Voluntário 9	14°	16°	Voluntário 24	21°	27°
Voluntário 10	12°	13°	Voluntário 25	24°	29°
Voluntário 11	3°	27°	Voluntário 26	12°	18°
Voluntário 12	21°	16°	Voluntário 27	7°	13°
Voluntário 13	7°	18°	Voluntário 28	24°	33°
Voluntário 14	14°	21°	Voluntário 29	19°	30°
Voluntário 15	5°	12°	Voluntário 30	14°	17°

A Figura 27 apresenta os diagramas de caixa das medidas de lateralização associadas aos sinais de ruído não processado, sinais de ruído processados pelo MWF_{AF} e sinais de ruído processados pelo $MWF-ILD_{AF}$ para as fontes de ruído nos azimutes -60° , -45° e -30° . As subfiguras (a), (c) e (e) apresentam os diagramas das medidas obtidas para o grupo de sinais do ruído ICRA, e as subfiguras (b), (d) e (f), os diagramas para o grupo de sinais do ruído passa-altas. Em cada figura, os azimutes originais das fontes de fala (θ_S) e de ruído (θ_N) são destacados através de linhas tracejadas. A análise dos diagramas obtidos é feita a partir das Tabelas 9, 10 e 11.

A Tabela 9 apresenta estimativas do azimute médio, mediano e do desvio padrão determinados a partir dos votos dos voluntários. A nomenclatura utilizada é descrita por:

- μ_{NP} , $\tilde{\mu}_{NP}$ e σ_{NP} são, respectivamente, o azimute médio, o azimute mediano e o desvio padrão determinados a partir dos votos dos voluntários ao ouvirem os sinais não processados.

- μ_{MWF} , $\tilde{\mu}_{\text{MWF}}$ e σ_{MWF} , respectivamente, o azimute médio, o azimute mediano e o desvio padrão determinados a partir dos votos dos voluntários ao ouvirem os sinais processados pelo método MWF_{AF} .
- μ_{ILD} , $\tilde{\mu}_{\text{ILD}}$ e σ_{ILD} , respectivamente, o azimute médio, o azimute mediano e o desvio padrão determinados a partir dos votos dos voluntários ao ouvirem os sinais processados pelo método $\text{MWF-ILD}_{\text{AF}}$.

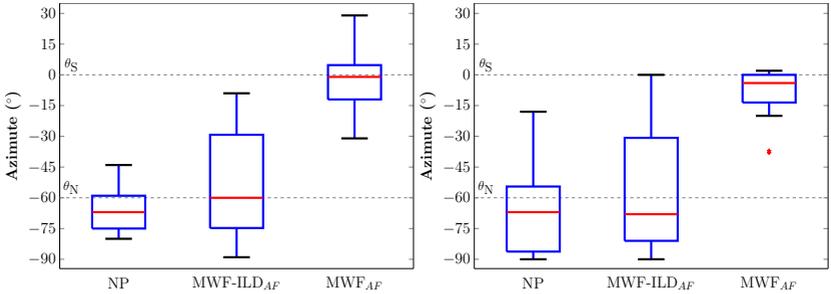
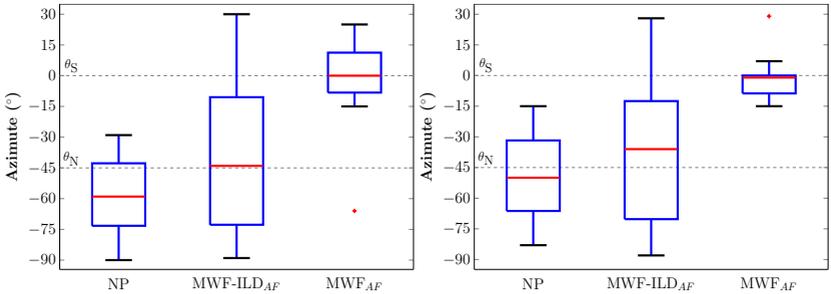
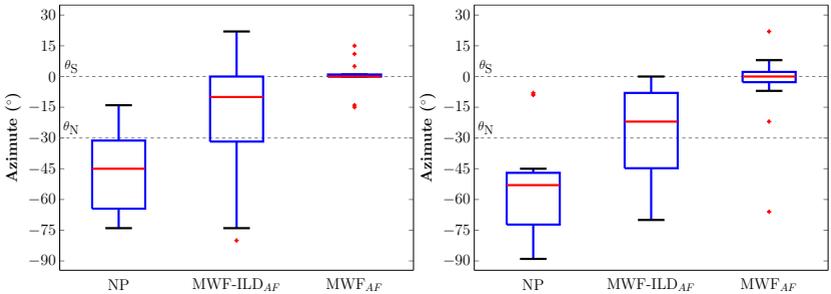
As estimativas na Tabela 9 indicam que, considerando todos os cenários acústicos apresentados na Figura 27, há uma visível separação entre o azimute médio e mediano determinados a partir das medidas de lateralização associadas ao ruído processado pelo MWF_{AF} ($0^\circ \leq \mu_{\text{NP}}$, $\tilde{\mu}_{\text{NP}} \leq -9^\circ$) e os azimutes médio e mediano determinados a partir das medidas de lateralização associadas ao ruído processado pelo $\text{MWF-ILD}_{\text{AF}}$ (μ_{ILD} , $\tilde{\mu}_{\text{ILD}} \geq -10^\circ$). Esta observação, corroborada por testes estatísticos que serão apresentados a seguir, destaca a capacidade de preservação da lateralização das fontes de ruído pelo método proposto.

Em relação ao desvio padrão dos resultados, há uma menor precisão nas medidas obtidas a partir do experimento com os ruídos processados pelo $\text{MWF-ILD}_{\text{AF}}$ ($25^\circ \leq \sigma_{\text{ILD}} \leq 37^\circ$) em comparação com as medidas obtidas com os ruídos processados pelo MWF_{AF} ($8^\circ \leq \sigma_{\text{ILD}} \leq 20^\circ$) e com os ruídos não processados ($11^\circ \leq \sigma_{\text{ILD}} \leq 24^\circ$). Isto indica uma maior dificuldade dos participantes em determinar o azimute da fonte sonora ouvindo os ruídos processados pelo $\text{MWF-ILD}_{\text{AF}}$. Considerando os desvios padrões das medidas obtidas a partir dos sinais não processados, há uma melhor precisão dos resultados quanto mais afastada está a fonte do plano mediano. Observando os desvios padrões obtidos a partir do ruído emitido pelas fontes de ruído em -60° e -45° e processados pelo $\text{MWF-ILD}_{\text{AF}}$, há uma melhor precisão dos resultados para a fonte de ruído em -60° .

Tabela 9 – Média (μ_p), mediana ($\tilde{\mu}_p$) e desvio padrão (σ_p) para os sinais não processados e processados pelos métodos MWF_{AF} , $MWF-ILD_{AF}$ nos cenários acústicos da Avaliação 2.

Ruído ICRA									
	NP			MWF_{AF}			$MWF-ILD_{AF}$		
	μ_{NP}	$\tilde{\mu}_{NP}$	σ_{NP}	μ_{MWF}	$\tilde{\mu}_{MWF}$	σ_{MWF}	μ_{ILD}	$\tilde{\mu}_{ILD}$	σ_{ILD}
S_0N_{-60}	-66°	-67°	11°	-4°	-1°	17°	-54°	-60°	28°
S_0N_{-45}	-58°	-59°	21°	-3°	0°	21°	-38°	-44°	37°
S_0N_{-30}	-48°	-45°	21°	1°	0°	8°	-18°	-10°	30°
Ruído Passa-Altas									
	NP			MWF_{AF}			$MWF-ILD_{AF}$		
	μ_{NP}	$\tilde{\mu}_{NP}$	σ_{NP}	μ_{MWF}	$\tilde{\mu}_{MWF}$	σ_{MWF}	μ_{ILD}	$\tilde{\mu}_{ILD}$	σ_{ILD}
S_0N_{-60}	-67°	-67°	21°	-9°	-4°	14°	-57°	-68°	30°
S_0N_{-45}	-51°	-50°	22°	-2°	-1°	11°	-38°	-36°	34°
S_0N_{-30}	-56°	-53°	24°	-5°	0°	20°	-28°	-22°	25°

Figura 27 – Diagrama de extremos e *quartis* determinados a partir dos votos dos voluntários da Avaliação 2 associados às fontes de ruído nos azimutes -60° , -45° e -30° : as subfiguras (a), (c) e (e) apresentam os diagramas dos votos obtidos com o conjunto de ruídos ICRA, as subfiguras (b), (d) e (f) apresentam os diagramas dos votos obtidos com os ruído do tipo passa-altas. Os azimutes de fala (θ_S) e de ruído (θ_N) são destacados através das linhas tracejadas nos gráficos.

(a) ICRA: S_0N_{-60} (b) Passa-altas: S_0N_{-60} (c) ICRA: S_0N_{-45} (d) Passa-altas: S_0N_{-45} (e) ICRA: S_0N_{-30} (f) Passa-altas: S_0N_{-30}

A Tabela 10 compara os azimutes percebidos pelos voluntários com os azimutes originais das fontes de fala e de ruído. A nomenclatura utilizada é descrita por:

- $|\theta_S - \mu_{\text{NP}}|$ e $|\theta_N - \tilde{\mu}_{\text{NP}}|$: diferença entre os azimutes percebidos sem o processamento binaural (ao ouvir os sinais não processados) e os azimutes originais das fontes de fala e de ruído.
- $|\theta_S - \mu_{\text{MWF}}|$ e $|\theta_N - \tilde{\mu}_{\text{MWF}}|$: diferença entre os azimutes percebidos com o processamento binaural (sinais processados pelo MWF_{AF}) e os azimutes originais das fontes de fala e de ruído.
- $|\theta_S - \mu_{\text{ILD}}|$ e $|\theta_N - \tilde{\mu}_{\text{ILD}}|$: diferença entre os azimutes percebidos com o processamento binaural (sinais processados pelo MWF-ILD_{AF}) e os azimutes originais das fontes de fala e de ruído.

As diferenças na Tabela 10 indicam que os azimutes médio e mediano associados ao ruído processado pelo MWF_{AF}, se localizam, em módulo, até 9° da fonte de fala ($|\mu_{\text{MWF}} - \theta_S|, |\tilde{\mu}_{\text{MWF}} - \theta_S| \leq 9^\circ$), ou seja, muito próximo do local de emissão do sinal. Esses resultados concordam com os obtidos através de experimentos subjetivos de localização de azimute realizados com ruídos processados pelo MWF (BOGAERT et al., 2008), e decorre do fato de o MWF_{AF} ser uma versão adaptativa do MWF, e, portanto, ter como uma de suas características o deslocamento da fonte de ruído em direção à fonte de fala (CORNELIS et al., 2010). No caso do método proposto, as médias e medianas dos azimutes percebidos, determinados após os participantes ouvirem os ruídos processados pelo MWF-ILD_{AF}, se localizam, em todos os cenários acústicos, no mínimo 10°, em módulo, distante da fonte de fala ($|\mu_{\text{ILD}} - \theta_S|, |\tilde{\mu}_{\text{ILD}} - \theta_S| \geq 10^\circ$). Em relação ao azimute original da fonte de ruído, o MWF-ILD_{AF} é mais preciso quanto mais próximo do azimute -60° se localiza a fonte de ruído ($|\mu_{\text{ILD}} - \theta_N|, |\tilde{\mu}_{\text{ILD}} - \theta_N| \leq 8^\circ$, para $\theta_N = -60^\circ$). Esse mesmo comportamento é verificado também ao se analisar os azimutes médio e mediano dos sinais não processados ($|\mu_{\text{NP}} - \theta_N|, |\tilde{\mu}_{\text{NP}} - \theta_N| \leq 7^\circ$, para $\theta_N = -60^\circ$).

A Tabela 11 compara os azimutes percebidos pelos voluntários sem o processamento (ao ouvirem os sinais não processados) e com o processamento dos métodos MWF_{AF} e MWF-ILD_{AF} (ao ouvirem os sinais de ruído processados pelo MWF_{AF} e MWF-ILD_{AF}). A nomenclatura utilizada é descrita por:

- $|\mu_{\text{NP}} - \mu_{\text{MWF}}|$ e $|\tilde{\mu}_{\text{NP}} - \tilde{\mu}_{\text{MWF}}|$: diferença, em módulo, entre as médias ou medianas obtidas a partir da avaliação com o ruído

sem processamento (μ_{NP} , $\tilde{\mu}_{NP}$) e com o ruído processado pelo MWF_{AF} (μ_{MWF} e $\tilde{\mu}_{MWF}$).

Tabela 10 – Diferença entre as percepções média e mediana obtidas com os ruídos sem processamento (μ_{MWF} , μ_{ILD} , $\tilde{\mu}_{MWF}$ e $\tilde{\mu}_{ILD}$) em relação o azimute original das fontes de fala (θ_S) e de ruído (θ_N).

Ruído ICRA				
NP				
	$ \theta_S - \mu_{NP} $	$ \theta_N - \mu_{NP} $	$ \theta_S - \tilde{\mu}_{NP} $	$ \theta_N - \tilde{\mu}_{NP} $
S ₀ N ₋₆₀	66°	6°	67°	7°
S ₀ N ₋₄₅	58°	13°	59°	14°
S ₀ N ₋₃₀	48°	18°	45°	15°
MWF_{AF}				
	$ \theta_S - \mu_{MWF} $	$ \theta_N - \mu_{MWF} $	$ \theta_S - \tilde{\mu}_{MWF} $	$ \theta_N - \tilde{\mu}_{MWF} $
S ₀ N ₋₆₀	4°	57°	1°	59°
S ₀ N ₋₄₅	3°	43°	0°	45°
S ₀ N ₋₃₀	1°	30°	0°	30°
$MWF-ILD_{AF}$				
	$ \theta_S - \mu_{ILD} $	$ \theta_N - \mu_{ILD} $	$ \theta_S - \tilde{\mu}_{ILD} $	$ \theta_N - \tilde{\mu}_{ILD} $
S ₀ N ₋₆₀	54°	7°	60°	0°
S ₀ N ₋₄₅	38°	8°	44°	1°
S ₀ N ₋₃₀	18°	13°	10°	20°
Ruído Passa-Altas				
NP				
	$ \theta_S - \mu_{NP} $	$ \theta_N - \mu_{NP} $	$ \theta_S - \tilde{\mu}_{NP} $	$ \theta_N - \tilde{\mu}_{NP} $
S ₀ N ₋₆₀	67°	7°	67°	7°
S ₀ N ₋₄₅	51°	6°	50°	5°
S ₀ N ₋₃₀	56°	26°	53°	23°
MWF_{AF}				
	$ \theta_S - \mu_{MWF} $	$ \theta_N - \mu_{MWF} $	$ \theta_S - \tilde{\mu}_{MWF} $	$ \theta_N - \tilde{\mu}_{MWF} $
S ₀ N ₋₆₀	9°	52°	4°	56°
S ₀ N ₋₄₅	2°	43°	1°	44°
S ₀ N ₋₃₀	5°	26°	0°	30°
$MWF-ILD_{AF}$				
	$ \theta_S - \mu_{ILD} $	$ \theta_N - \mu_{ILD} $	$ \theta_S - \tilde{\mu}_{ILD} $	$ \theta_N - \tilde{\mu}_{ILD} $
S ₀ N ₋₆₀	57°	3°	68°	8°
S ₀ N ₋₄₅	38°	8°	36°	9°
S ₀ N ₋₃₀	28°	3°	22°	8°

- $|\mu_{\text{NP}} - \mu_{\text{ILD}}|$ e $|\tilde{\mu}_{\text{NP}} - \tilde{\mu}_{\text{ILD}}|$: diferença, em módulo, entre as médias e medianas obtidas a partir sem processamento (μ_{NP} , $\tilde{\mu}_{\text{NP}}$) e com o processamento do MWF-ILD_{AF} (μ_{ILD} e $\tilde{\mu}_{\text{ILD}}$).

As diferenças na Tabela 11 mostram que, como consequência do deslocamento da fonte de ruído em direção à fonte de fala, as médias e medianas dos azimutes percebidos associados às fontes de ruído processadas pelo MWF_{AF} se localizam, no mínimo, 48° distantes dos azimutes percebidos sem a utilização do método de redução de ruído ($|\mu_{\text{NP}} - \mu_{\text{MWF}}|$ e $|\tilde{\mu}_{\text{NP}} - \tilde{\mu}_{\text{MWF}}| \geq 48^\circ$). Em relação aos resultados do método proposto, os azimutes percebidos pelos participantes apresentam maior exatidão quanto mais próximo a fonte de ruído se localiza do azimute -60° ($|\mu_{\text{NP}} - \mu_{\text{ILD}}|$, $|\tilde{\mu}_{\text{NP}} - \tilde{\mu}_{\text{ILD}}| \leq 12^\circ$).

Tabela 11 – Diferença entre as médias e medianas obtidas com o ruído sem processamento (μ_{NP} e $\tilde{\mu}_{\text{NP}}$) e as obtidas com o ruído processado pelas técnicas de redução de ruído (μ_{MWF} , μ_{ILD} , $\tilde{\mu}_{\text{MWF}}$ e $\tilde{\mu}_{\text{ILD}}$).

Ruído ICRA				
	NP e MWF _{AF}		NP e MWF-ILD _{AF}	
	$ \mu_{\text{NP}} - \mu_{\text{MWF}} $	$ \tilde{\mu}_{\text{NP}} - \tilde{\mu}_{\text{MWF}} $	$ \mu_{\text{NP}} - \mu_{\text{ILD}} $	$ \tilde{\mu}_{\text{NP}} - \tilde{\mu}_{\text{ILD}} $
S ₀ N ₋₆₀	62°	66°	12°	7°
S ₀ N ₋₄₅	56°	49°	20°	15°
S ₀ N ₋₃₀	48°	55°	31°	35°
Ruído Passa-Altas				
	NP e MWF _{AF}		NP e MWF-ILD _{AF}	
	$ \mu_{\text{NP}} - \mu_{\text{MWF}} $	$ \tilde{\mu}_{\text{NP}} - \tilde{\mu}_{\text{MWF}} $	$ \mu_{\text{NP}} - \mu_{\text{ILD}} $	$ \tilde{\mu}_{\text{NP}} - \tilde{\mu}_{\text{ILD}} $
S ₀ N ₋₆₀	59°	63°	11°	1°
S ₀ N ₋₄₅	49°	49°	13°	14°
S ₀ N ₋₃₀	52°	53°	28°	31°

A Tabela 12 apresenta o valor-P (p) obtido a partir do teste T de Wilcoxon (ver Seção 2.6.7) realizado em cada cenário acústico, utilizando nível de significância α_1 igual a 0,05, a partir dos seguintes pares de medidas: NP e MWF_{AF}, NP e MWF-ILD_{AF}, MWF_{AF} e MWF-ILD_{AF}. Comparando as medidas de lateralização obtidas a partir do experimento com os ruídos processados pelo MWF_{AF} com as medidas de lateralização obtidas a partir dos ruídos não processados e ruídos processado pelo MWF-ILD_{AF}, os testes de hipóteses mostram que as medidas associadas ao MWF_{AF} são estatisticamente diferentes das medidas associadas ao MWF-ILD_{AF} e aos ruídos não processados, i.e., há uma diferença estatística significativa ($p < 0,05$) entre os conjuntos de

votos obtidos a partir do ruído processado pelo MWF_{AF} e os conjuntos de votos associados ao $MWF-ILD_{AF}$ e às associadas aos ruídos não processados, em todos os cenários acústicos analisados. Os resultados do teste T realizado com os pares de medidas NP e $MWF-ILD_{AF}$, mostram que, para uma fonte de ruído nos azimutes -60° e -45° , não há diferença estatística entre os dois conjuntos de medidas ($p \geq 1,52 \cdot 10^{-1}$). Contudo, quando a fonte de ruído se localiza no azimute -30° , as medidas de lateralização associadas ao $MWF-ILD_{AF}$ são estatisticamente diferentes das medidas associadas aos sinais não processados ($p < 0,05$).

Tabela 12 – Valor-P obtido a partir do teste T de Wilcoxon dos seguintes pares de dados: NP e MWF_{AF} , NP e $MWF-ILD_{AF}$, MWF_{AF} e $MWF-ILD_{AF}$, em cada cenário acústico analisado.

Ruído ICRA			
Cenários Acústicos	S_0N_{-60}	S_0N_{-45}	S_0N_{-30}
NP e MWF_{AF}	$3,27 \cdot 10^{-6}$	$1,67 \cdot 10^{-5}$	$4,0 \cdot 10^{-6}$
NP e $MWF-ILD_{AF}$	$4,17 \cdot 10^{-1}$	$1,52 \cdot 10^{-1}$	$3,56 \cdot 10^{-3}$
MWF_{AF} e $MWF-ILD_{AF}$	$4,77 \cdot 10^{-5}$	$4,22 \cdot 10^{-3}$	$3,58 \cdot 10^{-2}$
Ruído Passa-Altas			
Cenários Acústicos	S_0N_{-60}	S_0N_{-45}	S_0N_{-30}
NP e MWF_{AF}	$5,82 \cdot 10^{-6}$	$3,98 \cdot 10^{-6}$	$2,83 \cdot 10^{-5}$
NP e $MWF-ILD_{AF}$	$4,42 \cdot 10^{-1}$	$3,19 \cdot 10^{-1}$	$1,34 \cdot 10^{-2}$
MWF_{AF} e $MWF-ILD_{AF}$	$1,38 \cdot 10^{-4}$	$1,27 \cdot 10^{-3}$	$3,83 \cdot 10^{-4}$

Os resultados apresentados nas Tabelas 9, 10, 11 e 12 mostram a capacidade de lateralização do método proposto nos três cenários acústicos da Avaliação 1: S_0N_{-60} , S_0N_{-45} e S_0N_{-30} . O $MWF-ILD_{AF}$ apresenta uma maior exatidão e precisão quanto mais distante a fonte de ruído se localiza da linha mediana. Adicionalmente, testes de hipóteses realizados mostram que as medidas de lateralização obtidas com o método proposto são estatisticamente diferentes das medidas relacionados ao MWF_{AF} e, também, não apresentam diferença estatística significativa em relação aos resultados obtidos com os sinais não processados, para uma fonte de ruído localizada em -60° e -45° . Porém, para uma fonte de ruído em -30° , as medidas de lateralização associadas ao $MWF-ILD_{AF}$ são estatisticamente diferentes das medidas obtidas com os ruídos não processados, embora, estas medidas sejam, também, estatisticamente diferentes das medidas associadas ao MWF_{AF} . Adicionalmente, os azimutes percebidos a partir da avaliação com a fonte de ruído em -30° cujo ruído emitido é processado pelo $MWF-ILD_{AF}$, se

localizam mais laterais que -10° . Todos estes resultados corroboram, portanto, a capacidade de lateralização da abordagem proposta.

5.3 AVALIAÇÃO 3: RASTREAMENTO DE FONTE DE RUÍDO

O objetivo da Avaliação 3 é verificar a capacidade de adaptação do método proposto dados os parâmetros utilizados nos avaliações 1 e 2. Para tanto, o cenário acústico ilustrado na Figura 28 foi gerado. Ele possui uma fonte de fala fixa no azimute de 0° , e uma fonte de ruído que modifica sua localização original, no azimute -60° , para uma nova localização, no azimute 60° , e vice-versa. Nessa avaliação, o sinal processado possui duração de 237,5 segundos. Durante esse tempo, o azimute da fonte de ruído se modifica duas vezes: a primeira, em 17,7 segundos, e, a segunda, em 127,6 segundos. Considerando estas modificações no azimute, o sinal processado pelo sistema binaural pode ser dividido em três segmentos:

- Segmento 0: compreende o intervalo de tempo, em segundos, entre $[0; 17,7]$, gerado a partir do cenário acústico S_0N_{-60} .
- Segmento 1: compreende o intervalo de tempo, em segundos, entre $]17,7; 127,6]$, gerado a partir do cenário acústico S_0N_{60} .
- Segmento 2: compreende o intervalo de tempo, em segundos, entre $]127,6; 237,5]$, gerado a partir do cenário acústico S_0N_{-60} .

Os sinais de fala utilizados para gerar os segmentos apresentados possuem duração diferentes: no segmento inicial (segmento 0), o sinal de fala possui duração de 17,7 segundos, sendo 3 segundos iniciais de silêncio e 14,7 segundos de sinal de fala (Seção 4.3); nos segmentos intermediário (Segmento 1) e final (Segmento 2), o sinal de fala utilizado possui duração de 109,9 segundos. Estes sinal de fala foi criado a partir da união de sete repetições do sinal de fala apresentado na Figura 29, que possui 1 segundo inicial de silêncio e 14,7 segundos de fala. Nos três segmentos, a SNR total da orelha ipsilateral à fonte de ruído foi definida em 0 dB. A estimação dos valores iniciais das matrizes de correlação ($\hat{\mathbf{R}}_{y,0}$, $\hat{\mathbf{R}}_{x,0}$ e $\hat{\mathbf{R}}_{v,0}$) foi feita a partir de 6 segundos iniciais do sinal processado pelo sistema binaural, ou seja, os 6 segundos iniciais do segmento inicial. Após esse período, inicia-se a filtragem do sinal.

A capacidade de rastreamento do azimute da fonte de ruído pelos métodos MWF_{AF} e $MWF-ILD_{AF}$, para os parâmetros definidos nas avaliações 1 e 2, foi analisada a partir do gráfico da ΔILD_{No} , calculada

para cada *frame* λ , e através da PESQ, SNR, Δ ISNR, Δ ILD_{No} e Δ ILD_{Sp} medidas nos 4 segundos finais de cada um dos três segmentos.

Figura 28 – Cenário acústico da Avaliação 3: fonte de fala em 0° e fonte de ruído cujo azimuth varia de -60° para 60° , em 17,7 segundos, e de 60° para -60° , em 127,6.

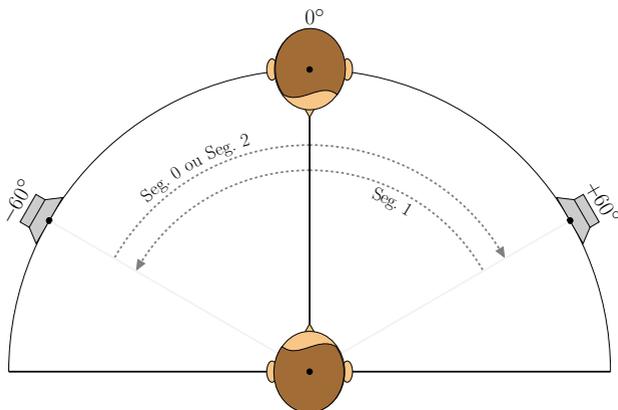
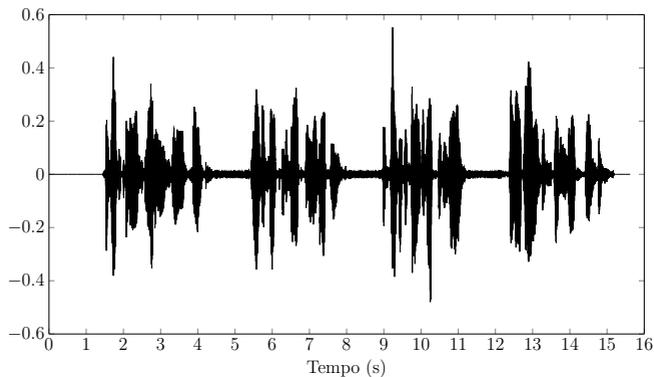


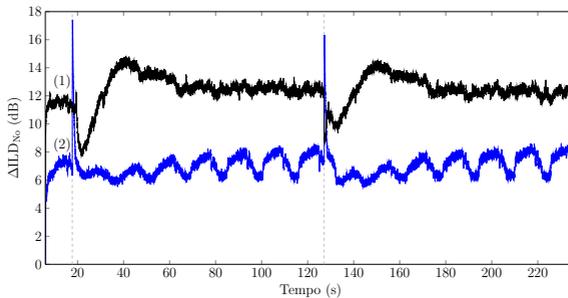
Figura 29 – Sinal de fala utilizado na geração dos segmentos 2 e 3.



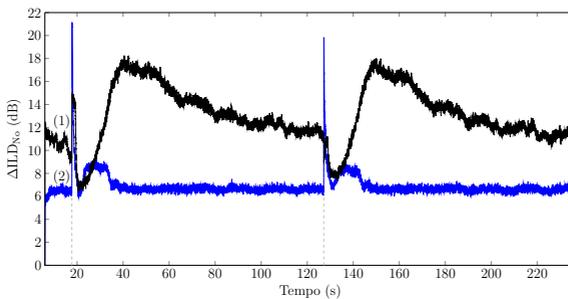
5.3.1 Resultados e discussão

A Figura 30^{3,4} apresenta o desempenho do $MWF-ILD_{AF}$ através da ΔILD_{No} em cada *frame* para o cenário acústico gerado com ruído ICRA (Figura 30(a)) e para o cenário gerado com o ruído passa-altas (Figura 30(b)). A avaliação da Figura 30 é feita a partir dos resultados da Tabela 13.

Figura 30 – Desempenho do (1) MWF_{AF} e (2) $MWF-ILD_{AF}$ para o cenário acústico com fonte de ruído que varia sua posição espacial. Cenários acústicos gerados com (a) ruído ICRA e (b) ruído passa-altas.



(a)



(b)

A Tabela 13 apresenta as quatro métricas de desempenho utilizadas para avaliar a capacidade de adaptação do método proposto em relação às novas condições apresentadas pela variação de cenário

³Curvas suavizadas através de um filtro média móvel de ordem 15.

⁴O valor médio do gráfico na Figura 30 é diferente do valor apresentado na Tabela 13 devido às características não lineares no cálculo da ΔILD_{No} .

acústico. Os resultados mostram que as métricas obtidas nos trechos iniciais (segmentos 0, S_0N_{-60}) para ambos os tipos de ruído concordam com as métricas obtidas a partir do cenário acústico S_0N_{-60} gerado com os ruídos ICRA e passa-altas na Avaliação 1 (Tabela 6). Desta forma, a sensação subjetiva de lateralização é a mesma descrita na Avaliação 2: os azimutes médio e mediano obtidos a partir do experimento com o MWF_{AF} se localizam próximo da fonte de fala ($8,8^\circ$ distante de θ_s), e os azimutes médio e mediano percebidos pelos voluntários após ouvirem os ruídos processados pelo $MWF-ILD_{AF}$ se localizam, para uma fonte de ruído em -60° de azimute, mais laterais que $-53,46^\circ$ (Tabela 9).

Analisando a ΔILD dos ruídos ICRA e passa-altas processados pelo MWF_{AF} , os resultados mostram que, em ambos os casos, após a modificação do azimute da fonte de ruído, a ΔILD_{No} dos 4 segundos finais dos segmentos intermediário (Segmento 1) e final (Segmento 2) se mantém superior à ΔILD_{No} dos 4 segundos finais do trecho inicial: no caso do ruído ICRA, a ΔILD_{No} é superior a 10,86 dB, e, no caso do ruído passa-altas, a ΔILD_{No} dos segmentos finais está acima de 9,86 dB. Considerando os resultados apresentados na Tabela 6 e na Figura 27, é possível concluir que os valores de ΔILD_{No} obtidos são consequência do deslocamento da fonte de ruído em direção à fonte de fala. Em relação à ΔILD do sinal de fala processado pelo MWF_{AF} , para o experimento com ambos os ruídos, os resultados mostram que a ΔILD_{Sp} nos trechos após a variação de azimute se mantém aproximadamente igual ou inferior a ΔILD_{Sp} do trecho antes da variação de azimute (aumento de até 0,02 dB). Isto mostra que, mesmo após a variação do azimute da fonte de ruído, o MWF_{AF} retorna aos níveis de ΔILD_{Sp} obtidos antes da modificação de azimute.

Considerando o experimento realizado com o $MWF-ILD_{AF}$, os resultados mostram que após a variação da posição da fonte de ruído, a ΔILD_{No} se mantém abaixo de 4,42 dB, nos experimentos com ambos os ruídos. Portanto, considerando os resultados nas Tabelas 6 e na Figura 27, é possível concluir que o método proposto mantém a capacidade de lateralização da fonte de ruído mesmo após a variação do azimute da fonte de ruído. Em relação à ΔILD_{Sp} pelo $MWF-ILD_{AF}$, para o experimento realizado com ambos os tipos de ruído, os resultados mostram uma redução de 0,17 dB na ΔILD_{Sp} , após a variação do azimute da fonte de ruído (de 1,48 dB, no trecho inicial, para 1,31 dB, no trecho final, no experimento com o ruído ICRA).

Considerando o cenário acústico criado com ruído ICRA e processado pelo $MWF-ILD_{AF}$, após a variação da posição da fonte de ruído, há uma aumento de até 0,52 dB na ΔILD_{No} (de 3,90 dB no segmento

inicial, para 4,42 dB no segmento final). No experimento realizado com ruído passa-altas, após a modificação de azimute, a $\Delta\text{ILD}_{\text{No}}$ do ruído processado pelo MWF-ILD_{AF} diminuiu em até 0,69 dB (de 2,93 dB, no segmento inicial, para 2,24 dB no segmento intermediário). Considerando os resultados nas Tabelas 6 e na Figura 27, conclui-se que os resultados da $\Delta\text{ILD}_{\text{No}}$ indicam a preservação da lateralização da fonte de ruído pelo método proposto, mesmo após a variação do azimute da fonte de ruído. Em relação à $\Delta\text{ILD}_{\text{Sp}}$ pelo MWF-ILD_{AF}, para o experimento realizado com ambos os tipos de ruído, os resultados mostram uma redução de 0,17 dB na $\Delta\text{ILD}_{\text{Sp}}$, após a variação da fonte de ruído (de 1,48 dB, no trecho inicial, para 1,31 dB, no trecho final, no experimento com o ruído ICRA).

A avaliação da PESQ na Tabela 13 será feita considerando que variações acima de 0,2 PESQ podem ser percebidas em avaliações subjetivas (SERVETTI; MARTIN, 2005). Além disso, todas as avaliações a seguir serão feitas levando em conta a orelha ipsilateral à fonte de ruído (orelha esquerda para fonte no azimute -60° e orelha direita para fonte de ruído no azimute 60°) e a orelha contralateral à fonte de ruído (orelha direita para fonte de ruído no azimute -60° e orelha esquerda para fonte de ruído no azimute 60°).

Para o experimento realizado com o MWF_{AF}, os resultados mostram que, após a variação de azimute, a qualidade de fala, no caso do ruído passa-altas, se mantém a mesma em ambas as orelhas (variação máxima de 0,09 PESQ) e, no caso do ruído ICRA, apresenta um aumento significativo de qualidade (aumento de até 0,43 PESQ). Para o experimento realizado com o MWF-ILD_{AF}, após a modificação da fonte de ruído, há um aumento da PESQ, para ambos os tipos de ruído e em ambas as orelhas: no caso do ruído ICRA, há aumentos de até 0,37 PESQ na orelha ipsilateral, e de até 0,52 PESQ na orelha contralateral à fonte de ruído; no caso do ruído passa-altas, há aumentos de até 0,28 PESQ (orelha ipsilateral) e 0,36 PESQ (orelha contralateral). Portanto, o método proposto, mesmo após a variação de azimute da fonte de ruído, apresentou níveis superiores de qualidade de fala, em relação ao momento anterior a variação de azimute, em ambas as orelhas.

Em relação à SNR, os resultados do experimento realizado com o ruído ICRA processado pelo MWF_{AF} mostram que, nos trechos finais, há um aumento de até 3,86 dB na SNR da orelha ipsilateral à fonte de ruído (de 16,68 dB, na orelha esquerda no trecho inicial, para 20,54 dB, na orelha direita no trecho final) e, aproximadamente, a mesma SNR na orelha contralateral (variação, em módulo, de 0,06 dB). No caso do ruído passa-altas, há uma redução tanto na orelha ipsilateral quanto na

orelha contralateral à fonte de ruído: na orelha ipsilateral, a SNR reduz de 23,71 dB no trecho inicial, para 22,25 obtido no trecho intermediário (redução de até 1,46 dB); na orelha contralateral, a SNR reduz de 27,63 dB no trecho inicial, para 24,36 dB no trecho intermediário (redução de 3,27 dB).

Para o experimento com o MWF-ILD_{AF}, considerando os resultados dos ruídos ICRA e passa-altas após a variação de azimute, há um aumento da SNR em ambas as orelhas: no caso do ruído ICRA, aumento de até 0,92 dB na orelha ipsilateral à fonte de ruído e aumento de até 1,56 dB na orelha contralateral à fonte de ruído; no caso do ruído passa-altas, há um aumento de até 2,30 dB na SNR da orelha ipsilateral e aumento de até 0,96 dB na SNR da orelha contralateral.

Considerando os resultados de SNR discutidos, conclui-se que, após a variação do azimute da fonte de ruído, a capacidade de redução de ruído do método proposto apresenta um aumento em relação ao momento anterior à variação (aumento de até 2,30 dB em ambas as orelhas) para ambos os tipos de ruídos avaliados. No caso do MWF_{AF}, após a variação de azimute, há uma diminuição da capacidade de redução de ruído em relação ao trecho inicial.

Analisando-se a Δ ISNR, os resultados do cenário acústico criado com ruído ICRA e processado pelo MWF_{AF} indicam que, após a variação do azimute da fonte de ruído, há uma redução, em ambas as orelhas, do ganho de inteligibilidade do sistema binaural: redução de até 3,86 dB na Δ ISNR da orelha ipsilateral à fonte de ruído, e redução de até 2,13 dB de Δ ISNR na orelha oposta à fonte de ruído. Considerando o cenário acústico criado com o ruído ICRA e processado pelo MWF-ILD_{AF}, os resultados mostram que a Δ ISNR se mantém, aproximadamente, a mesma em ambas as orelhas (aumento de 0,28 dB na orelha ipsilateral à fonte de ruído, e aumento de 0,46 na orelha contralateral). Portanto, considerando as variações apresentadas no Δ ISNR, os resultados do experimento com o ruído ICRA mostram que, após as transições do azimute da fonte de ruído, o MWF_{AF} apresenta aumentos na inteligibilidade do sistema binaural, enquanto que o método proposto mantém a mesma Δ ISNR em ambas as orelhas, mesmo após a transição de azimutes. Contudo, considerando os valores absolutos, o método proposto apresenta um desempenho melhor que MWF_{AF} na contralateral à fonte de ruído: Δ ISNR na orelha oposta superior a 15,10 dB, para o MWF-ILD_{AF}, e inferior a 12,47, para o MWF_{AF}.

Considerando a avaliação realizado com o ruído passa-altas processado pelo MWF_{AF}, os resultados mostram que, após a transição do azimute da fonte de ruído, há uma redução de até 4,78 dB na

Δ ISNR da orelha ipsilateral à fonte de ruído, e uma redução de 5,2 dB na Δ ISNR da orelha contralateral à fonte de ruído. Os resultados do experimento realizado com o ruído passa-altas processado pelo MWF-ILD_{AF} mostram que, nos trechos intermediário e final, há, também, uma redução da Δ ISNR em ambas as orelhas: redução de até 3,14 dB na Δ ISNR na orelha ipsilateral à fonte de ruído e redução de até 2,79 dB da Δ ISNR da orelha contralateral à fonte de ruído. Além disso, o método proposto apresenta ganhos positivos de Δ ISNR em ambas as orelhas. Portanto, considerando as variações de Δ ISNR e os valores absolutos apresentados, pode-se concluir que ambos os métodos apresentam redução da Δ ISNR de ambas as orelhas, contudo, o método proposto apresenta, mesmo com estas reduções, ganhos positivos de Δ ISNR, enquanto, que o MWF_{AF} apresenta ganhos negativos.

Os resultados apresentados nessa avaliação indicam que os parâmetros de projeto utilizados nas avaliações anteriores anteriores, embora não ótimos, possibilitam a adaptação do método proposto, de maneira a possibilitar o rastreamento de fontes sonoras, mantendo a preservação da percepção da localização original e melhorando o conforto acústico e a inteligibilidade.

5.4 CONCLUSÃO

Este capítulo apresentou três avaliações utilizadas para ilustrar a capacidade de lateralização do método proposto. Na Avaliação 1 avaliou-se o método proposto em três cenários acústicos, a saber: S_0N_{-60} , S_0N_{-45} e S_0N_{-30} , utilizando as seguintes métricas de desempenho: PESQ, SNR, Δ ISNR, Δ ILD_{No} e Δ ILD_{Sp}. Os resultados objetivos mostram que o MWF-ILD_{AF} reduz a Δ ILD_{No}, resultando na lateralização da fonte de ruído. Como consequência da preservação da ILD do ruído, há uma redução da qualidade de fala (PESQ), da redução de ruído (SNR) e da inteligibilidade (Δ ISNR) na orelha ipsilateral à fonte de ruído (esquerda), em relação ao desempenho do MWF_{AF}. Contudo, na orelha contralateral à fonte de ruído, o desempenho da PESQ, SNR e Δ ISNR são aproximadamente os mesmos apresentados pelo filtro de Wiener adaptativo.

Na Avaliação 2, avaliou-se a capacidade de adaptação do método proposto na preservação subjetiva da lateralização da fonte de ruído. Os votos dos participantes da avaliação mostraram que: o MWF_{AF}, como esperado, desloca a percepção da fonte de ruído em direção à fonte de fala, enquanto que o MWF-ILD_{AF} preserva a lateralidade da

fonte de ruído em todos os cenários acústicos analisados.

Na Avaliação 3, analisou-se a capacidade do método proposto de rastrear variações abruptas no azimute da fonte de ruído. Os resultados obtidos corroboram os resultados apresentados nas avaliações 1 e 2: o MWF_{AF} apresenta valores altos de ΔILD_{No} , deslocando, portanto, a fonte de ruído para a direção da fonte de fala, e o $MWF-ILD_{AF}$, comparado ao MWF_{AF} , apresenta valores mais baixos de ΔILD_{No} .

Tabela 13 – Cenário acústico gerado com ruído ICRA: desempenho do MWF_{AF} e $MWF-ILD_{AF}$ determinado a partir da PESQ, SNR, $\Delta ISNR$, ΔILD_{No} e ΔILD_{Sp} nos 4 segundos finais de cada trecho que compõe o sinal processado.

Ruído ICRA						
Método	MWF_{AF}			$MWF-ILD_{AF}$		
Segmento	0	1	2	0	1	2
ΔILD_{No} [dB]	10,86	11,91	11,84	3,90	4,39	4,42
ΔILD_{Sp} [dB]	0,39	0,19	0,13	1,33	1,17	1,20
PESQ _L	1,84	2,12	2,27	1,58	2,27	1,95
PESQ _R	1,95	2,07	2,29	1,96	1,83	2,48
SNR _L [dB]	16,68	20,54	20,54	16,69	21,96	17,61
SNR _R [dB]	20,59	20,10	20,65	20,70	17,57	22,26
$\Delta ISNR_L$ [dB]	20,05	10,34	19,74	14,18	15,38	14,46
$\Delta ISNR_R$ [dB]	12,47	19,28	10,61	15,10	14,37	15,56
Ruído Passa-Altas						
Método	MWF_{AF}			$MWF-ILD_{AF}$		
Segmento	0	1	2	0	1	2
ΔILD_{No} [dB]	9,26	10,67	10,33	2,93	2,24	2,25
ΔILD_{Sp} [dB]	0,64	0,66	0,51	1,48	1,36	1,31
PESQ _L	2,24	2,45	2,30	1,50	2,99	1,78
PESQ _R	2,45	2,27	2,54	2,64	1,77	3,00
SNR _L [dB]	23,71	24,36	22,57	14,46	27,86	16,76
SNR _R [dB]	27,63	22,25	24,97	26,95	16,59	27,91
$\Delta ISNR_L$ [dB]	13,16	-1,53	8,59	7,90	2,43	4,84
$\Delta ISNR_R$ [dB]	4,13	8,38	-1,07	5,20	4,76	2,41

Capítulo 6

Conclusão

Neste trabalho, apresentou-se um algoritmo adaptativo de custo computacional controlado para o cálculo dos coeficientes do filtro de redução de ruído de um sistema de processamento binaural. O objetivo desse algoritmo é reduzir a diferença entre a ILD do ruído na entrada do sistema binaural e a ILD do ruído na saída, resultando com isto, na preservação do cenário acústico.

Para fornecer um indicativo do custo computacional, determinou-se o número de operações MAC da abordagem proposta em função de parâmetros de projeto do método proposto. Adicionalmente, para avaliar a capacidade de lateralização da abordagem proposta, três avaliações foram realizadas utilizando cenários acústicos que emulam condições próximas das encontradas em aplicações práticas de aparelhos auditivos. Conjuntamente aos resultados do MWF-ILD_{AF}, foram apresentadas comparações com a forma adaptativa, sem preservação da ILD, do filtro de Wiener (MWF_{AF}). Os resultados obtidos mostram que, em cenários acústicos com uma fonte de ruído com posição fixa, o MWF-ILD_{AF} preserva a lateralização da fonte de ruído. Adicionalmente, em comparação ao MWF_{AF}, o MWF-ILD_{AF} apresenta aproximadamente os mesmos níveis de qualidade de fala e redução de ruído na orelha ipsilateral à fonte de ruído, contudo, na orelha contralateral à fonte de ruído, o método proposto apresenta reduções dessas métricas. Em relação ao ganho de inteligibilidade do sistema binaural, o MWF-ILD_{AF} mostra um aumento na Δ ISNR da orelha contralateral à fonte de ruído e uma redução dessa métrica na orelha ipsilateral à fonte de ruído, em comparação a Δ ISNR dos sinais processados pelo MWF_{AF}.

Em um terceiro cenário acústico com uma fonte de ruído cujo o azimute varia abruptamente, os resultados apresentados pelo MWF-ILD_{AF} corroboram a capacidade de manutenção da lateralização original da fonte de ruído, mesmo após a modificação do azimute da fonte. Adicionalmente, após a variação do azimute, o método proposto converge novamente para níveis iguais ou superiores de qualidade de fala

e redução de ruído, encontrados antes da variação do azimute, independente do tipo de ruído analisado. Em relação ao ganho de inteligibilidade do sistema binaural, o MWF-ILD_{AF} apresenta valores de Δ ISNR inferiores aos apresentados pelo MWF_{AF}, na orelha ipsilateral à fonte de ruído, contudo, na orelha contralateral à fonte de ruído, o método apresenta ganho de Δ ISNR de até 3,98 dB acima dos apresentados pelo MWF_{AF}.

6.1 PROPOSTAS PARA CONTINUAÇÃO DO TRABALHO

Algumas sugestões para a continuidade do trabalho são:

- **Estratégia para determinação automática do coeficiente de preservação da ILD (γ):** neste trabalho o coeficiente de preservação da ILD foi determinado levantando-se as curvas de desempenho do método proposto. Em um caso prático, esta abordagem pode ser inviável, por isto, o desenvolvimento de uma estratégia de determinação automática do valor de γ seria de grande utilidade.
- **Impacto da utilização de um VAD não ideal:** o algoritmo proposto foi testado com um VAD ideal a fim de evitar erros de classificação dos trechos de fala e trechos de fala+ruído. Contudo, seria de grande utilidade avaliar o desempenho do método proposto utilizando um método de VAD não ideal. Através disto, seria possível verificar o impacto dos erros de classificação deste tipo de detector de fala no desempenho do método proposto.
- **Avaliação do método proposto em cenários acústicos reais:** os cenários acústicos simulados utilizados neste trabalho apresentarem várias características semelhantes às presentes em cenários acústicos reais, contudo, a fim de validar os resultados obtidos nos cenários simulados seria de grande interesse testar o método apresentado, em condições controladas, utilizando cenários acústicos não simulados.

REFERÊNCIAS

- ALT, H.; LEEUWEN van. The complexity of basic complex operations. **Computing**, v. 27, p. 205–215, 1981.
- BAI, M.; LIN, C. Microphone array signal processing with application in three-dimensional spatial hearing. **Journal of the Acoustical Society of America**, v. 117, n. 4, p. 2112–2121, 2005.
- BAUMAN, N. The hearing aids of yesteryear: A brief history of hearing aids from then to now. 2013.
- BLAUERT, J. **Spatial hearing: the psychoacoustics of human sound localization**. [S.l.]: MIT Press, 1997.
- BOGAERT, T. et al. Binaural cue preservation for hearing aids using an interaural transfer function multichannel wiener filter. **IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)**, p. 565–568, 2007.
- BOGAERT, T. et al. Horizontal localization with bilateral hearing aids: without is better than with. **Journal of the Acoustical Society of America**, v. 19, n. 1, p. 515–526, 2006.
- BOGAERT, T. Van den et al. The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids. **Journal of the Acoustical Society of America**, v. 124, n. 1, p. 484–497, 2008.
- BRANDWOOD, D. A complex gradient operator and its application in adaptive array theory. **IEE Proceedings H - Microwaves, Optics and Antennas**, v. 130, n. 1, p. 11–16, 1983.
- CNS Nº 466/12. **Conselho Nacional de Saúde – Resolução Nº466/12**. [S.l.], 2012.
- CORNELIS, B. et al. Theoretical analysis of binaural multimicrophone noise reduction techniques. **IEEE Transactions on Audio, Speech e Language Processing**, v. 2, n. 18, p. 342–355, 2010.
- CORNELIS, B.; MOONEN, M.; WOUTERS, J. Binaural voice activity detection for MWF-based noise reduction in binaural hearing aids.

19.º European Signal Processing Conference (EUSIPCO), Barcelona, Espanha, p. 486–490, Aug. 2011.

COSTA, M.; NAYLOR, A. ILD preservation in the multichannel Wiener filter for binaural hearing aid applications. **IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)**, Lisboa, Portugal, p. 636–640, Sept. 2014.

CROCHIERE, R. A weighted overlap-add method of short-time Fourier analysis/synthesis. **IEEE Transactions on Acoustics, Speech and Signal Processing**, v. 28, n. 1, p. 99–102, 1980.

DILLON, H. **Hearing Aids**. [S.l.]: Boomerang press, 2001.

DOCLO, S. et al. Extension of the multi-channel wiener filter with localisation cues for noise reduction binaural hearing aids. **Inter. Workshop on Acoustic Echo and Noise Control (IWAENC)**, p. 221–224, Sept. 2005.

DOCLO, S.; MOONEN, M. Design of far-field and near-field broadband beamformers using eigenfilters. **Signal Processing Elsevier**, v. 83, n. 12, p. 2641–2673, Dec. 2003.

DOCLO, S. et al. Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids. **IEEE Transactions on Audio, Speech e Language Processing**, v. 17, n. 1, p. 38–51, Jan. 2009.

DOCLO, S. et al. Frequency-domain criterion for the speech distortion weighted multichannel wiener filter for robust noise reduction. **Speech Communication**, v. 49, n. 7-8, p. 636–656, 2007.

DRESCHLER, W. et al. ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. **Audiology**, v. 40, p. 148–157, 2001.

EPHRAIM, Y.; MALAH, D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. **IEEE Transactions on Acoustics, Speech and Signal Processing**, v. 32, n. 6, p. 1109–1121, dec 1984.

GELFAND, S. **An Introduction to Psychological and Physiological Acoustics**. 5.º ed. Nova York, Basel, EUA: Marcel Dekker, Inc., 2010.

GEORGE, E. B.; SMITH, M. J. T. Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model. **IEEE Transaction on Speech and Audio Processing**, v. 5, n. 5, p. 389–406, 1997.

GIBBONS, J.; CHAKRABORTI, S. **Non-parametric statistical inference**. 4.^o. ed. Nova York, Basel, EUA: Marcel Dekker, Inc., 2003.

GRADSHTEYN, I. S.; RYZHIK, I. M. **Table of integrals, series, and products**. 7.^o. ed. [S.l.]: Elsevier/Academic Press, Amsterdam, 2007.

GREENBERG, J.; PETERSON, P.; ZUREK, P. Intelligibility-weighted measures of speech-to-interference ratio and speech system performance. **Journal of the Acoustical Society of America**, v. 94, n. 5, p. 3009–3010, Nov. 1993.

HAWLEY, M.; LITOVSKY, R.; CULLING, J. The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. **Journal of the Acoustical Society of America**, v. 115, n. 2, p. 833–843, Feb. 2004.

HJØRUNGNES, A.; GESBERT, D. Complex-valued matrix differentiation: techniques and key results. **IEEE Transactions on Signal Processing**, v. 55, n. 6, June 2007.

HU, G. Segregation of unvoiced speech from nonspeech interference. **Journal of the Acoustical Society of America**, v. 2, n. 124, p. 1306–1319, 2008.

HUANG, H.; PAN, J. Speech pitch determination based on Hilbert-Huang transform. **Signal Processing**, v. 86, n. 4, p. 792–803, 2006.

HURTADO, J.; ANDERSON, D. Practical MWF-based noise-reduction methods for binaural hearing aids. **XVII Symposium of image, signal processing, and artificial vision (STSIVA)**, p. 8, 2012.

ITU-T P.50. **Appendix I: Test signals**. [S.l.], Feb. 1998.

ITU-T P.862.3. **Application Guide for Objective Quality Measurement Based on Recommendations P.862, P.862.1 and P.862.2**. [S.l.], Nov. 2005.

KAYSER, H. et al. Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. **EURASIP J. Adv. Signal Proc.**, v. 2009, p. 10, 2009.

Kjær & Bruel. **Product Data, Head and Torso Simulator – Type 4128-C and 4128-D**. [S.l.], 2014.

KLASEN, T. et al. Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues. **IEEE Transactions on Signal Processing**, v. 55, n. 4, p. 1579–1585, 2007.

LOIZOU, P. **Speech enhancement: theory and practice**. 2°. ed. Boca Raton, FL, USA: CRC Press, Inc., 2013.

LU, Y. et al. Enhancing echo cancellation via estimation of delay. **IEEE Transactions Signal Processing**, v. 53, n. 11, p. 4159–4168, Nov. 2005.

MONTGOMERY, D.; RUNGER, G. **Applied statistics and probability for engineers**. 5°. ed. [S.l.]: John Wiley and Sons, 2011.

NETTER, F. H. **Atlas de Anatomia Humana**. 5. ed. [S.l.]: Elsevier, 2011.

NHO, K. **Digital signal processing algorithms for noise reduction, dynamic range compression, and feedback cancellation in hearing aids**. Tese (Doutorado) — Katholieke Universiteit Leuven, July 2011.

PALIWAL, K. K.; SCHWERIN, B.; WÓJCICKI, K. K. Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator. **Speech Communication**, v. 54, n. 2, p. 282–305, 2012.

PEISSIG, J.; KOLLMEIER, B. Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. **Journal of the Acoustical Society of America**, v. 101, n. 3, p. 1660–1670, 1997.

SERVETTI, A.; MARTIN, J. D. Error tolerant MAC extension for speech communications over 802.11 WLANs. **IEEE Vehicular Technology Conference**, Dec. 2005.

SMITH, J. **Choice of WOLA Window**. 2011. https://ccrma.stanford.edu/~jos/sasp/Choice_WOLA_Window.html. Acesso: 2015-03.

SPIEGEL, M. **Mathematical handbook of formulas and tables**. Nova York, St Louis, São Francisco: McGraw Hill, 1999.

SPRIET, A.; MOONEN, M.; WOUTERS, J. Robustness analysis of multichannel wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications. **IEEE Transaction on Speech and audio Processing**, v. 13, n. 4, July 2005.

STADLER, R.; RABINOWITZ, W. M. On the potential of fixed arrays for hearing aids. **Journal of the Acoustical Society of America**, v. 93, n. 4, p. 2317–2317, 1993.

STETZLER, T. et al. **Low power real-time programmable DSP development platform for digital hearing aids**. [S.l.], Apr. 2000.

TARIGHAT, A.; SAYED, A. Least mean-phase adaptive filters with applications to communications systems. **Proceedings of IEEE Signal Processing Letters**, v. 11, n. 2, p. 220–223, Feb. 2004.

VALIN, J. et al. Localization of simultaneous moving sound sources for mobile robot using a frequency-domain steered beamformer approach. **IEEE International Conference on Robotics and Automation (ICRA)**, abs/1602.08629, Apr. 2004.

Apêndice A

Derivação da função custo da J_{ILD}

O erro na diferença de intensidade interaural do ruído é apresentado em (2.6) e definido, para cada frequência k do *frame* λ , como:

$$\Delta ILD(\lambda, k) = 10 \log_{10} ILD_{out}(\lambda, k) - 10 \log_{10} ILD_{in}(\lambda, k) \quad (A.1)$$

onde:

$$\begin{aligned} ILD_{out}(\lambda, k) &= \frac{E\{\|z_L^v(\lambda, k)\|^2\}}{E\{\|z_R^v(\lambda, k)\|^2\}} = \frac{\mathbf{w}_L^H E\{\mathbf{v}\mathbf{v}^H\} \mathbf{w}_L}{\mathbf{w}_R^H E\{\mathbf{v}\mathbf{v}^H\} \mathbf{w}_R} = \frac{\mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L}{\mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R}, \\ ILD_{in}(\lambda, k) &= \frac{E\{\|v_{L,ref}(\lambda, k)\|^2\}}{E\{\|v_{R,ref}(\lambda, k)\|^2\}} = \frac{\mathbf{q}_L^T E\{\mathbf{v}\mathbf{v}^H\} \mathbf{q}_L}{\mathbf{q}_R^T E\{\mathbf{v}\mathbf{v}^H\} \mathbf{q}_R} = \frac{\mathbf{q}_L^T \mathbf{R}_v \mathbf{q}_L}{\mathbf{q}_R^T \mathbf{R}_v \mathbf{q}_R}. \end{aligned} \quad (A.2)$$

Elevando-se (A.1) ao quadrado e aplicando-se a propriedade de soma do logaritmo, obtém-se a definição, sem aproximações, da função custo de preservação da ILD do ruído, presente em (3.14), e dada por:

$$J_{ILD}(\lambda, k) = \left[10 \times \log_{10} \left(\frac{ILD_{out}(\lambda, k)}{ILD_{in}(\lambda, k)} \right) \right]^2 \quad (A.3)$$

Substituindo-se (A.2) em (A.3), resulta em:

$$J_{ILD}(\lambda, k) = \left[10 \times \log_{10} \left(\frac{\mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L \mathbf{q}_R^T \mathbf{R}_v \mathbf{q}_R}{\mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R \mathbf{q}_L^T \mathbf{R}_v \mathbf{q}_L} \right) \right]^2. \quad (A.4)$$

Aproximando o logaritmo em (A.4) pelo termo de primeira ordem da expansão em série de potências da função tangente hiperbólica inversa (GRADSHTEYN; RYZHIK, 2007):

$$\log_{10}(x) \approx \frac{2}{\log_e(10)} \left(\frac{x-1}{x+1} \right), \quad (A.5)$$

e desconsiderando o fator de ponderação resultante deste processo, obtém-se a função custo aproximada de preservação da ILD do ruído, definida como:

$$J_{ILD}(\lambda, k) = \left(\frac{\mathbf{q}_R^T \mathbf{R}_v \mathbf{q}_R \mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L - \mathbf{q}_L^T \mathbf{R}_v \mathbf{q}_L \mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R}{\mathbf{q}_R^T \mathbf{R}_v \mathbf{q}_R \mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L + \mathbf{q}_L^T \mathbf{R}_v \mathbf{q}_L \mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R} \right)^2 \quad (A.6)$$

Apêndice B

Determinação do gradiente da J_{ILD}

O objetivo deste apêndice é determinar as derivadas $\partial_{\mathbf{w}_L^*} J_{ILD}$ e $\partial_{\mathbf{w}_R^*} J_{ILD}$. Para tanto, redefine-se a J_{ILD} , apresentada em (3.17), em função das variáveis f_c e g_c , o que resulta em:

$$J_{ILD} = \left(\frac{f_c}{g_c} \right)^2, \quad (\text{B.1})$$

onde:

$$\begin{aligned} f_c &= r_{vR} \mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L - r_{vL} \mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R, \\ g_c &= r_{vR} \mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L + r_{vL} \mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R. \end{aligned} \quad (\text{B.2})$$

Definindo-se $\partial_{\mathbf{w}_L^*} J_{ILD}$, também em função f_c e g_c , obtém-se (SPIEGEL, 1999):

$$\partial_{\mathbf{w}_L^*} J_{ILD} = 2f_c \frac{g_c \partial_{\mathbf{w}_L^*} f_c - f_c \partial_{\mathbf{w}_L^*} g_c}{g_c^3} \quad (\text{B.3})$$

onde:

$$\begin{aligned} \partial_{\mathbf{w}_L^*} f_c &= r_{vR} \mathbf{R}_v \mathbf{w}_L, \\ \partial_{\mathbf{w}_L^*} g_c &= r_{vR} \mathbf{R}_v \mathbf{w}_L. \end{aligned} \quad (\text{B.4})$$

Como $\partial_{\mathbf{w}_L^*} f_c = \partial_{\mathbf{w}_L^*} g_c$, logo (B.3) é dado por:

$$\partial_{\mathbf{w}_L^*} J_{ILD} = 2(g_c - f_c) \frac{f_c}{g_c^3} \partial_{\mathbf{w}_L^*} f_c \quad (\text{B.5})$$

onde:

$$g_c - f_c = 2r_{vL} \mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R. \quad (\text{B.6})$$

Portanto, substituindo (B.2), (B.4) e (B.6) em (B.5), obtém-se a derivada parcial da J_{ILD} em relação ao vetor \mathbf{w}_L^* , dada por:

$$\partial_{\mathbf{w}_L^*} J_{ILD} = 4r_{vL} r_{vR} \mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R \frac{r_{vR} \mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L - r_{vL} \mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R}{(r_{vR} \mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L + r_{vL} \mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R)^3} \mathbf{R}_v \mathbf{w}_L. \quad (\text{B.7})$$

Definindo-se agora $\partial_{\mathbf{w}_R^*} J_{ILD}$ em função f_c e g_c , obtém-se:

$$\partial_{\mathbf{w}_R^*} J_{ILD} = 2f_c \frac{g_c \partial_{\mathbf{w}_R^*} f_c - f_c \partial_{\mathbf{w}_R^*} g_c}{g_c^3} \quad (\text{B.8})$$

onde:

$$\begin{aligned} \partial_{\mathbf{w}_R^*} f_c &= -r_{vL} \mathbf{R}_v \mathbf{w}_R, \\ \partial_{\mathbf{w}_R^*} g_c &= r_{vL} \mathbf{R}_v \mathbf{w}_R. \end{aligned} \quad (\text{B.9})$$

Como $\partial_{f_c} \mathbf{w}_L^* = -\partial_{g_c} \mathbf{w}_L^*$, então (B.8) é dada por:

$$\partial_{J_{ILD}} \mathbf{w}_R^* = -2(g_c + f_c) \frac{f_c}{g_c^3} \partial_{\mathbf{w}_R^*} f_c \quad (\text{B.10})$$

onde:

$$g_c + f_c = 2r_{vR} \mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L. \quad (\text{B.11})$$

Portanto, substituindo-se (B.2), (B.9) e (B.11) em (B.10), tem-se a derivada parcial da J_{ILD} em relação \mathbf{w}_R^* , dada por:

$$\partial_{\mathbf{w}_R^*} J_{ILD} = 4r_{vR} r_{vL} \mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L \frac{r_{vL} \mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R - r_{vR} \mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L}{(r_{vR} \mathbf{w}_L^H \mathbf{R}_v \mathbf{w}_L + r_{vL} \mathbf{w}_R^H \mathbf{R}_v \mathbf{w}_R)^3} \mathbf{R}_v \mathbf{w}_R. \quad (\text{B.12})$$

Apêndice C

Estimativas das matrizes de correlação

Várias abordagens de redução de ruído que utilizam múltiplos canais (Cornelis, Moonen e Wouters (2011), Nho (2011), Spriet, Moonen e Wouters (2005)) necessitam da estimação das matrizes de correlação dos sinais recebidos pelo sistema binaural. Uma maneira de se realizar isso é através do seguinte filtro de resposta ao impulso infinita:

$$y(n) = \mu y(n-1) + (1-\mu)x(n), \quad (\text{C.1})$$

onde $y(n)$ representa a estimativa atual da matriz de correlação, $y(n-1)$ e $x(n)$ são respectivamente a informação passada e a inovação utilizadas para se determinar $y(n)$ e μ é denominado fator de esquecimento ou coeficiente de suavização, e determina a importância das informações $y(n-1)$ e $x(n)$ na estimativa atual de $y(n)$.

Para determinar-se a faixa de valores na qual μ é válido, é necessário analisar a equação (C.1). Para tanto, considere a transformada Z do sistema descrito por essa equação:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1-\mu}{1-\mu z^{-1}}. \quad (\text{C.2})$$

Analisando (C.2), verifica-se que o sistema possui um polo em $z = \mu$. Portanto, para que o sistema seja estável, é necessário que $|\mu| < 1$. De forma prática, utiliza-se μ na faixa entre $[0;1]$. Além da estabilidade do sistema em (C.1), μ está ligado ao número de amostras necessárias para que uma informação atual perca 37% de importância. Considerando isso e a resposta ao impulso, $h(n)$, do sistema em (C.2), dada por:

$$h(n) = (1-\mu)\mu^n, \quad (\text{C.3})$$

podemos definir $n = n_s$ como sendo o número de amostras necessárias para que o valor inicial da resposta ao impulso, $(1-\mu)$, decaia a $1/e \approx 0,37$ ¹ de seu valor inicial. Portanto:

$$(1-\mu) \frac{1}{e} = (1-\mu) \mu^{n_s}. \quad (\text{C.4})$$

¹O valor e é o número de Neper

Aplicando o logaritmo natural em ambos os lados da equação anterior, temos que (STETZLER et al., 2000):

$$\mu = 1 - \frac{1}{n_s} \quad (\text{C.5})$$

Em aplicações de melhoramento de fala, a escolha de n_s deve levar em consideração o tempo de quase-estacionaridade do sinal de fala, valores típicos utilizados estão entre 20ms e 40ms (Ephraim e Malah (1984), George e Smith (1997), Huang e Pan (2006), Paliwal, Schwerin e Wójcicki (2012), Loizou (2013)).

ANEXO A

Termo de consentimento livre e esclarecido

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA E ELETRÔNICA

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Prezado voluntário,

Você está sendo convidado a participar da Pesquisa “Desenvolvimento de métodos para redução de ruído em aparelhos auditivos e implantes cocleares”, sob a responsabilidade do pesquisador Prof. Dr. Márcio Holsbach Costa do Departamento de Engenharia Elétrica e Eletrônica da Universidade Federal de Santa Catarina.

Objetivo: Esse trabalho pretende avaliar o desempenho de métodos de redução de ruído para uso em aparelhos auditivos.

Metodologia: Sua participação é voluntária e consiste em ouvir sons, utilizando fones de ouvido ou moldes auriculares e, em sequência, responder a perguntas sobre as características percebidas, como qualidade, inteligibilidade, conforto acústico e direção (ângulo) de chegada. O tempo total do procedimento é de cerca de cinquenta (50) minutos.

Benefícios: Os benefícios esperados envolvem a produção de conhecimento científico relacionado ao desenvolvimento de novos métodos de redução de ruído para aparelhos auditivos.

Desconfortos e riscos: Os possíveis riscos ou desconfortos decorrentes de sua participação na pesquisa são: o desconforto acústico (devido a um volume sonoro elevado), o desconforto mecânico (devido ao uso de moldes auriculares ou fones de ouvido) e a fadiga (devido ao tempo de realização do procedimento). O desconforto acústico é evitado através do controle (aumento/diminuição) de volume que poderá ser alterado a qualquer momento. O desconforto mecânico é evitado pelo reposicionamento ou retirada, a qualquer momento, dos moldes auriculares ou fones de ouvido. A fadiga é minimizada pela limitação do tempo total do procedimento e pela possibilidade de interrupção e descanso a qualquer momento.

Outras informações: O Sr(a) receberá respostas e esclarecimentos a todas às suas perguntas e dúvidas sobre os procedimentos

realizados e assuntos relacionados a essa pesquisa por meio do contato com o pesquisador, que assume o compromisso de proporcionar informações atualizadas sobre o estudo. O Sr(a) não terá nenhuma remuneração ou despesa em decorrência dessa pesquisa e tem o direito de retirar seu consentimento a qualquer momento, seja antes ou após a coleta dos dados, sem qualquer represália ou prejuízo. O pesquisador declara que cumprirá as exigências contidas na Resolução CNS 466/2012, que o sigilo e a privacidade dos participantes serão garantidos durante todas as etapas da pesquisa, inclusive na divulgação dos resultados e que os participantes terão direito ao ressarcimento de eventuais despesas e indenização diante de eventuais danos produzidos por essa pesquisa. Em caso de necessidade de contato, o endereço do pesquisador responsável é: Departamento de Engenharia Elétrica e Eletrônica, Centro Tecnológico, Bloco E, terceiro andar, Universidade Federal de Santa Catarina, Campus Universitário João David Ferreira Lima, bairro Trindade, CEP 88.040-900, Florianópolis-SC, telefone **(48) 3721-2260**, e-mail: **costa@eel.ufsc.br**. O endereço de do Comitê de Ética em Pesquisa da Universidade Federal de Santa Catarina é: Prédio Reitoria II, Rua Desembargador Vitor Lima, número 222, sala 401, Trindade, Florianópolis-SC, CEP 88.040-400, telefone **(48) 3721-6094**, e-mail: **cep.propesq@contato.ufsc.br**.

CONSENTIMENTO PÓS-INFORMADO

Eu, _____, portador (a) do RG/CPF _____ concordo em participar dessa pesquisa, bem como com a utilização dos dados coletados, desde que seja mantido o sigilo de minha identificação, conforme as normas do Comitê de Ética em Pesquisa com Seres Humanos. A minha participação é voluntária podendo ser suspensa a qualquer momento. Pelo presente consentimento, declaro que fui esclarecido(a) sobre a pesquisa a ser realizada, de forma detalhada, livre de qualquer constrangimento e obrigação, e que recebi uma cópia desse termo, assinada pelos pesquisadores.

Florianópolis, _____ de _____ de 20____.

Assinatura do Participante

Assinatura do Pesquisador Responsável
Prof. Dr. Márcio Holsbach Costa
RG: 1031584426 SJTC/RS