



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

Rafael Attili Chiea

**Máscaras tempo-frequência para a redução de ruído aditivo em implantes
cocleares**

Florianópolis
2019

Rafael Attili Chiea

**Máscaras tempo-frequência para a redução de ruído aditivo em implantes
cocleares**

Tese submetida ao Programa de Pós-Graduação
em Engenharia Elétrica da Universidade Federal de
Santa Catarina para a obtenção do título de doutor
em Engenharia Elétrica.

Orientador: Prof. Márcio Holsbach Costa, Dr.

Coorientador: Guillaume Barrault, Dr.

Florianópolis

2019

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

Chiea, Rafael Attili

Máscaras tempo-frequência para a redução de ruído aditivo em implantes cocleares / Rafael Attili Chiea ; orientador, Márcio Holsbach Costa, coorientador, Guillaume Barrault, 2019.

168 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2019.

Inclui referências.

1. Engenharia Elétrica. 2. Implante coclear. 3. Redução de ruído. 4. Enfatização de fala. 5. Máscaras tempo frequência. I. Costa, Márcio Holsbach. II. Barrault, Guillaume. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. IV. Título.

Rafael Attili Chiea

**Máscaras tempo-frequência para a redução de ruído aditivo em implantes
cocleares**

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca
examinadora composta pelos seguintes membros:

Prof. Romis Ribeiro de Faissol Attux, Dr.
Universidade Estadual de Campinas

Prof. Júlio Apolinário Cordioli, Dr.
Universidade Federal de Santa Catarina

Prof. Jefferson Luiz Brum Marques, Dr.
Universidade Federal de Santa Catarina

Prof. Bruno Catarino Bispo, Dr.
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi
julgado adequado para obtenção do título de doutor em Engenharia Elétrica.

Prof. Bartolomeu Ferreira Uchôa Filho, Dr.
Coordenador do Programa

Prof. Márcio Holsbach Costa, Dr.
Orientador

Florianópolis, 26 de setembro de 2019.

A Myrtille e Samuel.

AGRADECIMENTOS

Agradeço a todos que participaram direta ou indiretamente deste trabalho: minha esposa Myrtille e meu filho Samuel; meus pais, Sandra, Marcos; professor Márcio Costa, Guillaume Barrault, membros da banca; participantes do projeto IC; João dos Santos do MuSAELab; Hussnain Ali e prof. Hansen da UTD; colegas do LPDS; colegas do PPGEEL; colegas do LVA; prof^a. Madalena Pinheiro e suas alunas da fonoaudiologia e usuários de implante coclear participantes dos experimentos.

*"Work is love made visible."
(Khalil Gibran, 1923)*

RESUMO

Implantes cocleares (IC) são dispositivos que, a partir da estimulação elétrica do nervo auditivo, permitem a restituição parcial da audição em indivíduos com surdez profunda. Apesar de fornecerem uma informação limitada em resolução tanto no tempo quanto na frequência, seus usuários chegam a atingir índices de cerca de 80% de inteligibilidade da fala. Entretanto, esse desempenho cai significativamente na presença de ruído, o que caracteriza a maior parte dos cenários acústicos quotidianos. Técnicas de processamento de sinais para a redução de ruído se apresentam como uma alternativa para melhorar a percepção acústica de usuários de implante coclear. As principais técnicas propostas para redução de ruído em implantes cocleares consistem de máscaras tempo-frequência, destacando-se a máscara binária (BM), o filtro de Wiener (WF) e suas variantes (paramétrico e restrito). Neste trabalho, uma nova teoria unificada de máscaras tempo-frequência é apresentada. A partir do ajuste de dois parâmetros, diferentes funções de supressão podem ser realizadas, dentre as quais, algumas máscaras bem estabelecidas, tais como a máscara binária e o filtro de Wiener. Uma vantagem adicional da teoria proposta é que as máscaras derivadas por esse método são de alguma forma ótimas, diferentemente do que acontece com algumas propostas empíricas, como o filtro de Wiener paramétrico (WP). Além disso, a máscara proposta pode ser ajustada de maneira mais abrangente que o WP. Simulações numéricas extensivas mostram que a máscara proposta e a WP podem trazer melhorias na percepção de fala por usuários de IC em ambientes ruidosos. Entretanto, o desenvolvimento dessas máscaras não leva em conta características específicas do dispositivo. A maior parte dos ICs apresenta ao usuário apenas a informação de envelope temporal do sinal, ignorando totalmente a informação de fase. Nesse contexto, um novo filtro no domínio do tempo é proposto de forma a estimar o envelope de cada subbanda da fala. Simulações numéricas indicam que o filtro proposto leva a estimativas melhores do envelope em relação ao WF. Resultados de experimentos psicoacústicos tanto com normouvintes usando um simulador de IC, quanto com usuários de IC, indicam que o estimador de envelope proposto leva a maiores valores de inteligibilidade em relação ao WF, sobretudo para sinais com $SNR < -5\text{dB}$.

Palavras-chave: Implante coclear. Redução de ruído. Enfatização de fala. Máscaras tempo-frequência.

ABSTRACT

Cochlear implants (IC) are devices that partially restore hearing in subjects with severe deafness, this occurs through electrical stimulation of the auditory nerve. Even though the provided information is limited, due to poor time and frequency resolution, cochlear implant users may score up to 80% in speech intelligibility experiments. However, this performance is significantly reduced in presence of noise, which is the case in most everyday acoustic scenarios. Noise reduction techniques are generally applied to enhance acoustic perception by cochlear implant users. The main proposed techniques consist of time-frequency masks, such as the binary mask (BM), and the Wiener filter (WF) and its variations (parametric and constrained). In this work, a new unified theory for time-frequency masks is presented. By setting two parameters, different suppression functions may be realized, comprising well-established masks, such as the BM and the WF. Another advantage of the proposed theory is that the masks derived from it are somehow optimal, differently from heuristic masks such as the parametric Wiener filter (WP). Besides, the proposed mask can be adjusted within a wider range of suppression functions than the WP. Extensive numerical simulations show that the proposed mask and the WP may provide benefits to IC users perception in noisy environments. Nevertheless, those masks do not take into account specific IC characteristics. Most IC devices present only the signal's temporal envelope information to the user, regardless of phase information. Thus, a new time-domain filter is proposed in order to estimate the speech signal's temporal envelope. Numerical simulations show that this second proposed filter leads to better estimates of the speech envelope, compared to the WF. Psychoacoustical experiments with normal hearing subjects using an IC simulator, as well as with actual IC users indicate that the proposed envelope estimator leads to better intelligibility results when compared with the WF, mainly for signals corrupted at $\text{SNR} < -5\text{dB}$.

Keywords: Cochlear implant. Noise reduction. Speech enhancement. Time-frequency masks.

LISTA DE FIGURAS

Figura 1 – Anatomia da orelha humana (adaptado de OpenStax (2016)).	23
Figura 2 – Distribuição das frequências, em Hertz, percebidas de acordo com a posição de estimulação na cóclea (LOIZOU, 1999a).	23
Figura 3 – Ilustração do implante coclear.	24
Figura 4 – Diagrama de blocos da estratégia CIS – adaptado de Zeng <i>et al.</i> (2008).	26
Figura 5 – Sincronização dos pulsos de corrente – adaptado de Loizou (1999b).	27
Figura 6 – Diagrama de blocos de estratégias tipo "N of M-- adaptado de Zeng <i>et al.</i> (2008).	28
Figura 7 – Número de publicações sobre implante coclear por ano.	31
Figura 8 – Número cumulativo de publicações sobre redução de ruído para implantes cocleares por ano.	31
Figura 9 – Afiliação dos principais autores na área.	32
Figura 10 – Ganho de uma máscara binária em função na SNR do canal k	38
Figura 11 – Ganho de uma máscara de Wiener em função da SNR do canal k	40
Figura 12 – Efeito da variação do parâmetro β no filtro de Wiener paramétrico.	41
Figura 13 – Ganho de uma máscara de Wiener restrito em função da SNR do canal k	42
Figura 14 – Diagramas de blocos da decomposição do sinal de entrada em múltiplos canais no processamento de estratégias de estimulação baseadas na CIS.	44
Figura 15 – Pesos aplicados à cada bin de frequência de forma a se calcular o envelope associado a cada canal. Adaptado de (UTD, 2010)	45
Figura 16 – Diagrama de blocos do processamento com redução de ruído no domínio da frequência.	45
Figura 17 – Resposta em frequência do banco de filtros tipo <i>gammatone</i> de 22 canais, baseado em Slaney (1993)	46
Figura 18 – Relação entre γ e δ	54
Figura 19 – Efeito da variação dos parâmetros γ e μ em $H(k)$	55
Figura 20 – Erro normalizado da máscara conformável proposta em relação à máscara binária em função de γ	56
Figura 21 – Sobreposição das funções de supressão da CM proposta e do WP que mais se aproxima à primeira. Em (a), a CM tem $\gamma = 0,5$ e $\mu = 5$ dB. Em (b), a CM tem $\gamma = 100$ e $\mu = -10$ dB.	59
Figura 22 – Sobreposição das funções de supressão do WP e da CM proposta que mais se aproxima ao primeiro. Em (a), o WP tem $\beta = 0,5$ e $\eta = 10$ dB. Em (b), o WP tem $\beta = 100$ e $\eta = -20$ dB.	60

Figura 23 – Efeito dos parâmetros β no WP (a), e γ na CM proposta(b).	61
Figura 24 – Espectrogramas dos ruídos usados nos experimentos.	63
Figura 25 – Diagramas de caixa bidimensionais de qualidade (PESQ) e inteligibilidade (NCM). Os parâmetros da CM e e WP foram otimizados para qualidade.	67
Figura 26 – Diagramas de caixa bidimensionais de qualidade (PESQ) e inteligibilidade (SRMR-CI). Os parâmetros da CM e e WP foram otimizados para qualidade.	68
Figura 27 – Diagramas de caixa bidimensionais de qualidade (PESQ) e inteligibilidade (NCM). Os parâmetros da CM e e WP foram otimizados pela métrica NCM.	71
Figura 28 – Diagramas de caixa bidimensionais de qualidade (PESQ) e inteligibilidade (SRMR-CI). Os parâmetros da CM e WP foram otimizados de acordo com o critério SRMR-CI.	73
Figura 29 – Diagramas de caixa bidimensionais de qualidade (PESQ) e inteligibilidade (NCM). Valores médios de PESQ obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com a equação (63) com NCM.	77
Figura 30 – Diagramas de caixa bidimensionais de qualidade (PESQ) e inteligibilidade (SRMR-CI). Os parâmetros da CM e WP foram otimizados de acordo com a equação (63) com SRMR-CI.	79
Figura 31 – Função de supressão WP preferida por 10 usuários de IC diferentes. Adaptado de (MAUGER; DAWSON; HERBACH, 2012).	82
Figura 32 – Diagrama de blocos do processamento do implante coclear no domínio do tempo com redução de ruído.	84
Figura 33 – Relações entre momentos de segunda e quarta ordens do sinal de saída de cada canal k do banco de filtros, quando na entrada deste é colocado um processo laplaciano.	91
Figura 34 – Relações entre momentos de segunda e quarta ordens do sinal de saída de cada canal k do banco de filtros, quando na entrada deste é colocado um processo gaussiano.	91
Figura 35 – Função de supressão proposta (i), como função da SNR local, ξ , em dB.	105
Figura 36 – Erro relativo ϵ_{rel} entre os sinais estimados e real e, também, entre os envelopes dos sinais, ϵ_{rel}^{env}	108
Figura 37 – Diagramas de caixa da inteligibilidade calculada a partir da métrica SRMR-CI em função da SNR.	109
Figura 38 – Fasores $Y(k)$, $X(k)$ e $V(k)$, para uma mistura arbitrária a $SNR = -3dB$	110

Figura 39 – Ganho do filtro proposto (i) em comparação com o do filtro de Wiener (ii) e o de uma aproximação do estimador de magnitude MMSE (iii).	111
Figura 40 – Funcionamento da plataforma de pesquisa em ICs da UTD	113
Figura 41 – Diagrama de blocos do processamento de sinais para os experimentos psicoacústicos.	116
Figura 42 – Interface gráfica para a realização do Experimento I.	119
Figura 43 – Resultados individuais do Experimento I com normouvintes.	120
Figura 44 – Diagrama de caixa dos resultados do Experimento I com normouvintes, agrupando todos os voluntários.	121
Figura 45 – Resultados individuais do Experimento I com usuários de implante coclear.	122
Figura 46 – Diagrama de caixa dos resultados do Experimento I com usuários de implante coclear, agrupando todos os voluntários.	123
Figura 47 – Interface gráfica para a realização do Experimento II.	124
Figura 48 – Resultados do Experimento II com normouvintes.	125
Figura 49 – Resultados do Experimento II com usuários de IC.	128
Figura 50 – Fotografia da plataforma de pesquisas em ICs	156
Figura 51 – Fotografia do <i>hardware</i> da plataforma de pesquisas em ICs	156

LISTA DE TABELAS

Tabela 1 – Relações entre $H(k)$ e máscaras conhecidas de acordo com os parâmetros γ e μ	55
Tabela 2 – Pares de parâmetros β e η [dB] que mais aproximam o filtro de Wiener paramétrico à máscara proposta com parâmetros γ e μ	57
Tabela 3 – Raiz quadrada do erro quadrático médio obtido após a otimização dos parâmetros β e η	58
Tabela 4 – Pares de parâmetros γ e μ [dB] que mais aproximam a CM ao WP com parâmetros β e η	58
Tabela 5 – Raiz quadrada do erro quadrático médio obtido após a otimização dos parâmetros γ e μ	59
Tabela 6 – Parâmetros da CM e WP obtidos com o conjunto de treinamento, buscando maximização da métrica de qualidade PESQ.	65
Tabela 7 – Valores médios de PESQ obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com o critério PESQ.	66
Tabela 8 – Parâmetros da CM e WP obtidos com o conjunto de treinamento, buscando maximização da métrica de qualidade NCM.	69
Tabela 9 – Valores médios de NCM obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com o critério NCM.	70
Tabela 10 – Parâmetros da CM e WP obtidos com o conjunto de treinamento, buscando maximização da métrica de qualidade SRMR-CI.	72
Tabela 11 – Valores médios de inteligibilidade (SRMR-CI) obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com o critério SRMR-CI.	72
Tabela 12 – Parâmetros da CM e WP obtidos com o conjunto de treinamento, buscando maximização da equação (63), considerando a NCM como métrica de inteligibilidade.	74
Tabela 13 – Valores médios de NCM obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com a equação (63) com NCM.	75
Tabela 14 – Valores médios de PESQ obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com a equação (63) com NCM.	76
Tabela 15 – Parâmetros da CM e WP obtidos com o conjunto de treinamento, buscando maximização da equação (63), considerando a SRMR-CI como métrica de inteligibilidade.	78

Tabela 16 – Valores médios de inteligibilidade ($I_{\%}$) obtidos, com a métrica SRMR-CI, por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com a equação (63) com SRMR-CI.	78
Tabela 17 – Valores médios de PESQ obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com a equação (63) com SRMR-CI.	79
Tabela 18 – Dados dos voluntários normouvintes que participaram dos experimentos psicoacústicos.	115
Tabela 19 – Dados dos usuários de implante coclear que participaram dos experimentos psicoacústicos.	115
Tabela 20 – Valores de r_{pc} da análise de Bland-Altman e p do teste de Wilcoxon para cada participante normouvinte e cada comparação feita no Experimento II.	127
Tabela 21 – Valores de r_{pc} da análise de Bland-Altman e p do teste de Wilcoxon para cada participante usuário de IC e cada comparação feita no Experimento II.	127
Tabela 22 – Características principais do <i>codec</i> da plataforma.	158

LISTA DE ABREVIATURAS E SIGLAS

ACE	Codificador de combinação avançada (<i>Advanced combination encoder</i>)
ADC	Conversor analógico-digital (<i>Analog-to-Digital Converter</i>)
aGT	Limiar aparente de ganho (<i>Aparent gain threshold</i>)
ANOVA	Análise de variâncias (<i>Analisis of variance</i>)
BM	Máscara binária
BTE	Processador atrás da orelha (<i>Behind the ear</i>)
CI	Circuito integrado (<i>Integrated circuit</i>)
CIS	Amostragem contínua sequencial (<i>Continuous interleaved sample</i>)
CM	Máscara conformável proposta
dB	deciBel
DNS	Sistema de nome de domínio (<i>Domain name system</i>)
EAS	Estimulação elétrico-acústica (<i>Electric-acoustical stimulation</i>)
FB	Banco de Filtros (<i>Filterbank</i>)
FFT	Transformada rápida de Fourier (<i>Fast Fourier Transform</i>)
FIFO	Primeiro que entra é o primeiro que sai (<i>First in first out</i>)
FPGA	Arranjo de portas programável em campo (<i>Flexible programmable gate array</i>)
IBM	Máscara binária ideal (<i>Ideal binary mask</i>)
IBGE	Instituto brasileiro de geografia e estatística
IC	Implante coclear (<i>Cochlear implant</i>)
ICRA	Colégio internacional de audiologia reabilitatória (<i>International collegium for rehabilitative audiology</i>)
IEEE	Instituto de engenheiros eletricitistas e eletrônicos (<i>Institute of electrical and electronics engineers</i>)
IFFT	Transformada inversa rápida de Fourier (<i>Inverse fast Fourier transform</i>)

ILD	Diferença de volume interaural (<i>Interaural loudness difference</i>)
ISTS	Sinal de teste de fala internacional (<i>International speech test signal</i>)
IRM	Máscara de reverberação ideal (<i>Ideal reverberation mask</i>)
ITD	Diferença de tempo interaural (<i>Interaural Time Difference</i>)
ITU	União internacional de telecomunicações (<i>International telecommunication union</i>)
JND	Mínima variação perceptível (<i>Just noticeable difference</i>)
KU	Universidade católica da Lovaina, Bélgica (<i>Katholieke universiteit Leuven</i>)
LDO	Regulador <i>low-dropout</i>
MCU	Unidade microcontroladora (<i>Microcontroller unit</i>)
MIPS	Microprocessador sem estágios intertravados de pipeline (<i>Microprocessor without interlocked pipeline stages</i>)
MMSE	Mínimo erro quadrático médio (<i>Minimum mean square error</i>)
MPEAK	Estratégia de codificação de fala <i>multipeak</i> (<i>Multipeak speech-coding strategy</i>)
NCM	Medida de covariância normalizada (<i>Normalized covariance measure</i>)
opACE	Codificador de combinação avançada ótimo (<i>Optimal advanced combination encoder</i>)
OMS	Organização mundial da saúde
PESQ	Avaliação perceptual da qualidade da fala (<i>Perceptual evaluation of speech quality</i>)
PMC	Circuito de gerenciamento de energia (<i>Power management circuitry</i>)
pps	Pulsos por segundo
RF	Rádio frequência
RMSE	Raiz quadrada do erro quadrático médio (<i>Root mean square error</i>)
SNR	Relação sinal-ruído (<i>Signal-to-noise ratio</i>)

SPEAK	Pico espectral (<i>Spectral peak</i>)
SPI	Interface serial periférica (<i>Serial peripheral interface</i>)
SRMR	Relação de energia de modulação entre fala e reverberação (<i>Speech to reverberation modulation energy ratio</i>)
SRMR-CI	Relação de energia de modulação entre fala e reverberação para implantes cocleares
SRR	Razão sinal-reverberação (<i>Signal to reverberation ratio</i>)
STFT	Transformada de Fourier de tempo curto (<i>Short-time Fourier transform</i>)
SUS	Sistema único de saúde
T_{60}	Tempo de reverberação 60dB
TCP	Protocolo de controle de transmissão (<i>Transmission control protocol</i>)
DHT	Transformada discreta de Hilbert (<i>Discrete Hilbert transform</i>)
TFS	Estrutura temporal fina (<i>Temporal fine structure</i>)
UART	Receptor-transmissor universal assíncrono (<i>Universal asynchronous receiver-transmitter</i>)
UDP	Protocolo de datagrama de usuário (<i>User datagram protocol</i>)
UFSC	Universidade Federal de Santa Catarina
UHCP	Protocolo dinâmico de configuração de <i>host</i> (<i>Dynamic host configuration protocol</i>)
USB	Barramento serial universal (<i>Universal serial bus</i>)
UTD	Universidade de Texas em Dallas (<i>University of Texas at Dallas</i>)
VAD	Detector de fala (<i>Voice activity detector</i>)
WF	Filtro de Wiener
WP	Filtro paramétrico de Wiener
WSS	Processo estacionário no sentido amplo (<i>wide sense stationary</i>)

LISTA DE SÍMBOLOS

$()^T$	Operador transposição
$()^H$	Operador Hermitiano
$()^*$	Operador conjugado
\mathbf{a}	Vetor coluna
\mathbf{A}	Matriz
a_{GT}	Limiar aparente de transição
a_z	Quadrado do envelope
B	Máscara Binária
C	Critério de seleção da máscara binária
d_V	Densidade espectral de potência da distorção relativa ao ruído
d_X	Densidade espectral de potência da distorção relativa à fala
e	Base do logaritmo natural
g_k	Filtro de estimação do envelope temporal
g_{k0}	Filtro de estimação do envelope temporal com 1 coeficiente
H	Máscara conformável proposta
$\mathcal{H}\{\}$	Transformada de Hilbert
\tilde{h}	Resposta ao impulso da Transformada de Hilbert
H	Resposta em frequência da Transformada de Hilbert
\mathbf{H}	Matriz transformada discreta de Hilbert
\mathbf{I}	Matriz identidade
j	$\sqrt{-1}$
J	Função custo
M	Máscara tempo-frequência genérica
P_X	Densidade espectral de potência do sinal x

ξ	Relação sinal-ruído local em uma determinada célula tempo-frequência
\mathbb{R}	Conjunto dos números reais
$R_{z,w}$	Matriz de correlação cruzada entre os sinais z e w
R_z	Matriz de autocorrelação do sinal z
$r_{z,w}$	Sequência de de correlação cruzada entre os sinais z e w
r_z	Sequência de autocorrelação do sinal z
T	Limiar de decisão para a máscara binária
v	Sinal de ruído no domínio do tempo
V	Sinal de ruído no domínio da frequência
W	Filtro de Wiener
W_c	Filtro de Wiener restrito
W_p	Filtro de Wiener parametrizado
w_k	Filtro de Wiener no domínio do tempo
w_k	Filtro de Wiener no domínio do tempo com 1 coeficiente
x	Sinal de fala no domínio do tempo
X	Sinal de fala no domínio da frequência
\hat{X}	Estimativa do sinal de fala no domínio da frequência
y	Sinal de fala contaminada por ruído no domínio do tempo
Y	Sinal de fala contaminada por ruído no domínio da frequência
z_a	Sinal analítico
\tilde{z}	Transformada de Hilbert do sinal z
\mathbb{Z}	Conjunto dos números inteiros
β	Parâmetro do filtro de Wiener paramétrico
γ_k	Relação sinal-ruído <i>a posteriori</i> no <i>bin</i> de frequência k
δ	Parâmetro relacionado à intensificação das densidades espectrais de potência

ϵ^{norm}	Erro RMSE normalizado entre duas curvas contínuas
ϵ_{y_1, y_2}	Erro RMSE entre duas curvas contínuas
ϵ_g	Erro quadrático médio entre o sinal desejado e o estimado usando o filtro proposto g_k
ϵ_w	Erro quadrático médio entre o sinal desejado e o estimado usando o filtro de Wiener w_k
ϵ_g^{env}	Erro quadrático médio entre o envelope do sinal desejado e o envelope do sinal estimado usando o filtro proposto g_k
ϵ_w^{env}	Erro quadrático médio entre o envelope do sinal desejado e o envelope do sinal estimado usando o filtro proposto w_k
ϵ_{rel}	Erro quadrático médio do sinal estimado pelo filtro proposto g_k relativo ao erro quadrático médio do sinal estimado pelo filtro de Wiener w_k
ϵ_g^{env}	Erro quadrático médio do envelope do sinal estimado usando o filtro proposto g_k relativo ao erro quadrático médio do envelope do sinal estimado usando o filtro de Wiener w_k
ϵ	Erro de estimação da fala
ϵ_V	Distorção relativa ao ruído
ϵ_X	Distorção relativa à fala
η	Parâmetro do filtro de Wiener paramétrico
μ	Parâmetro da máscara conformável proposta
μ_0	Limiar de decisão da máscara binária sendo a relação sinal-ruído local o critério de decisão da máscara binária
μ_i	Média
ρ	Parâmetro representativo da ponderação entre as parcelas de distorção de fala e ruído
σ_i	Desvio padrão
$\sigma_V^2(k)$	Variância do sinal V na frequência k
$\sigma_X^2(k)$	Variância do sinal X na frequência k
σ_z^2	Variância da variável aleatória z

SUMÁRIO

1	INTRODUÇÃO	22
1.1	AUDIÇÃO NORMAL	22
1.2	IMPLANTES COCLEARES	24
1.3	ESTRATÉGIAS DE ESTIMULAÇÃO	25
1.3.1	CIS	26
1.3.2	“N of M”	27
1.4	DESEMPENHO E POSSÍVEIS MELHORAMENTOS	28
1.5	MOTIVAÇÃO	30
1.6	JUSTIFICATIVA	32
1.7	OBJETIVOS	33
1.7.1	Objetivo Geral	33
1.7.2	Objetivos Específicos	33
1.8	ORGANIZAÇÃO DO DOCUMENTO	33
2	TÉCNICAS DE REDUÇÃO DE RUÍDO PARA IMPLANTES COCLEARES	35
2.1	MÁSCARAS TEMPO-FREQUÊNCIA	35
2.2	MÁSCARA BINÁRIA	37
2.3	FILTRO DE WIENER E SUAS VARIANTES	40
2.4	IMPLEMENTAÇÃO DE TÉCNICAS DE REDUÇÃO DE RUÍDO E IMPLANTES COCLEARES	44
2.5	CONSIDERAÇÕES FINAIS	46
3	UMA MÁSCARA TEMPO-FREQUÊNCIA CONFORMÁVEL	48
3.1	MÁSCARA DE SUPRESSÃO CONFORMÁVEL PROPOSTA	48
3.2	COMPARAÇÃO COM O FILTRO DE WIENER PARAMÉTRICO	56
3.3	SIMULAÇÕES NUMÉRICAS	62
3.3.1	Experimento I – Maximização de qualidade	65
3.3.2	Experimento II – Maximização de inteligibilidade	67
3.3.3	Experimento III – Maximização conjunta de qualidade e inteligibilidade	72
3.4	DISCUSSÃO	80
3.5	CONSIDERAÇÕES FINAIS	83
4	UM FILTRO QUE MINIMIZA O ERRO DO ENVELOPE QUADRÁTICO	84
4.1	DESENVOLVIMENTO MATEMÁTICO	84
4.2	MÉTODO PROPOSTO	90
4.2.1	Características dos sinais	90
4.2.2	Formulação do filtro proposto	92
4.2.3	Filtro de um coeficiente	101

4.3	SIMULAÇÕES NUMÉRICAS	105
4.3.1	Metodologia	106
4.3.2	Resultados e Discussão	107
4.4	EXPERIMENTOS PSICOACÚSTICOS	112
4.4.1	Interface de pesquisa em implantes cocleares da UTD	112
4.4.2	Sujeitos	114
4.4.3	Processamento de sinais	114
4.4.4	Experimento I – Inteligibilidade	118
4.4.5	Experimento II – Qualidade	123
4.5	DISCUSSÃO E CONSIDERAÇÕES FINAIS	129
5	EPÍLOGO	131
5.1	DISCUSSÃO GERAL	131
5.2	CONCLUSÃO	133
5.3	PUBLICAÇÕES	134
	REFERÊNCIAS	135
	APÊNDICE A – DIFERENCIAÇÃO DE $z^H A z$, $z^H A z z^H B z$ E $(z^H A z)^2$	149
	APÊNDICE B – CÁLCULO DAS MÉTRICAS OBJETIVAS DE QUALIDADE E INTELIGIBILIDADE DA FALA	152
	APÊNDICE C – PLATAFORMA DE PESQUISA EM IMPLANTES COCLEARES - UNIVERSITY OF TEXAS AT DALLAS	155
	APÊNDICE D – MAPA PADRÃO DE USUÁRIO DE IC	159
	ANEXO A – TERMO DE CONSENTIMENTO PARA NORMOUVIN- TES	162
	ANEXO B – TERMO DE CONSENTIMENTO PARA USUÁRIOS DE IMPLANTE COCLEAR	165

1 INTRODUÇÃO

Implantes cocleares (IC) são as mais bem sucedidas interfaces máquina cérebro desenvolvidas até os dias de hoje (WOUTERS; MCDERMOTT; FRANCA, 2015). São próteses auditivas que permitem a percepção de sons por pessoas com perda auditiva severa a profunda ou surdez total (WOUTERS *et al.*, 2013). Diferentemente dos aparelhos auditivos, que têm como saída um sinal acústico, os implantes cocleares estimulam diretamente o nervo auditivo com uma corrente elétrica aplicada através de eletrodos posicionados no interior da cóclea.

Este primeiro capítulo é composto por uma breve apresentação da audição normal no ser humano, seguida da descrição do funcionamento do implante coclear, e do detalhamento das principais estratégias de codificação do som em pulsos elétricos usadas em aparelhos comerciais atualmente. Na sequência são apresentadas a motivação, a justificativa e os objetivos deste trabalho.

1.1 AUDIÇÃO NORMAL

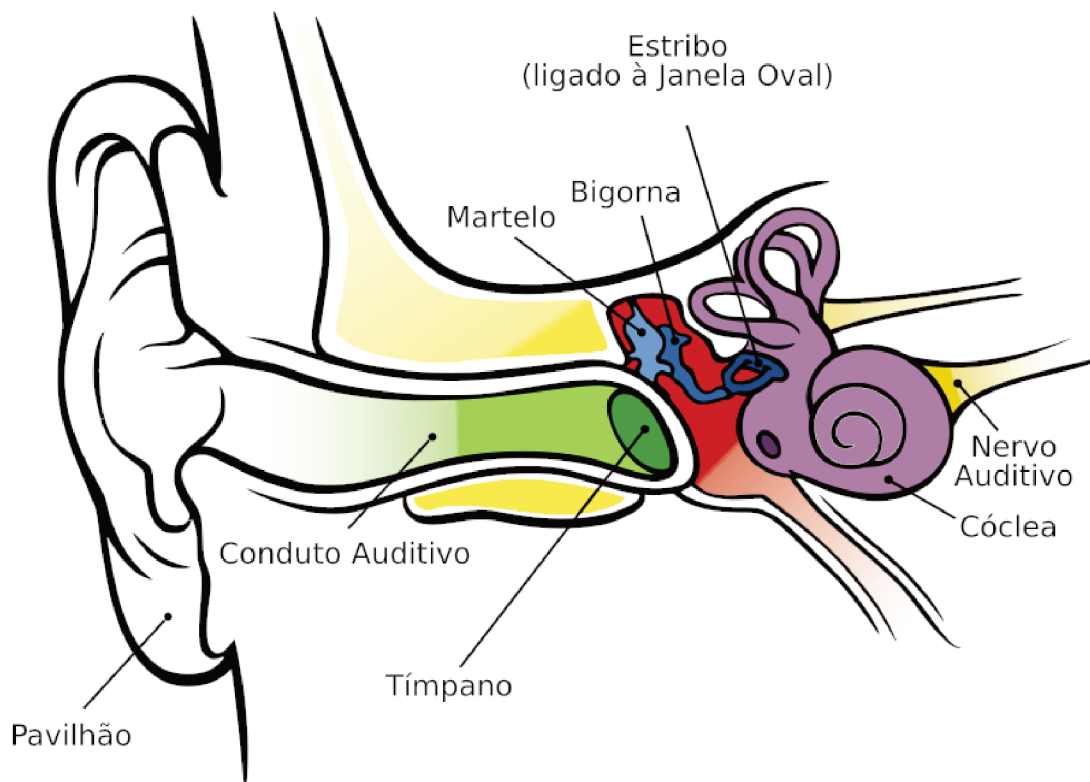
A percepção de sons pelos seres humanos se dá pelo sistema auditivo, um mecanismo que codifica ondas mecânicas em sequências de potenciais de ação que são transmitidos até o cérebro onde são interpretados.

A orelha humana é dividida em três partes, como ilustrado na Figura 1: a **orelha externa**, composta pelo pavilhão e o conduto auditivo até a parte externa da membrana timpânica; a **orelha média**, onde se encontra o conjunto de ossos formado por martelo, bigorna e estribo; e a **orelha interna**, que contém a cóclea.

No processo de audição normal humana, as ondas mecânicas do som penetram o canal auditivo e chegam à membrana timpânica, fazendo-a vibrar. Um conjunto de ossículos – martelo, bigorna e estribo – transmite a vibração da membrana timpânica à janela oval da cóclea, uma estrutura em forma de caracol. A cóclea hospeda o órgão sensitivo da audição, o órgão de Corti, que se conecta com o nervo auditivo. Esse órgão fica sobre uma membrana (basilar), que se estende pelo caracol e vibra seletivamente, de forma que a posição da máxima amplitude da vibração depende da frequência da onda sonora (CLARK, 2003). O máximo deslocamento da membrana é maior na região da base (próxima à orelha média) para frequências altas, e maior no ápice para baixas frequências.

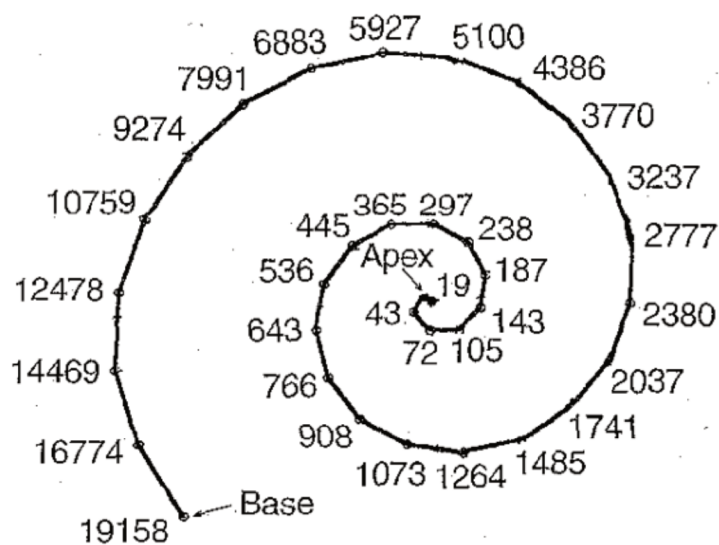
O movimento da membrana aciona o órgão de Corti, que faz a transdução da energia mecânica em pulsos elétricos. Esses pulsos são transmitidos pelo nervo auditivo ao cérebro, onde são interpretados como informação acústica. Tanto a membrana basilar quanto as terminações do nervo auditivo apresentam uma organização tonotópica ao longo da cóclea, de tal forma que a estimulação da inervação do ápice da cóclea resulta em uma percepção de tonalidade baixa, e a estimulação da base leva a

Figura 1 – Anatomia da orelha humana. (adaptado de OpenStax (2016)).



uma percepção de tonalidade alta (ZENG; POPPER; FAY, 2004). O mapeamento de frequência sonora em posição na membrana é mostrado na Figura 2.

Figura 2 – Distribuição das frequências, em Hertz, percebidas de acordo com a posição de estimulação na cóclea (LOIZOU, 1999a).



A perda parcial ou total da audição é denominada deficiência auditiva. Essa perda pode ter diversas causas, dentre as quais, congênita, envelhecimento, exposição a ruído, infecções, traumas físicos, medicamentos e agentes ototóxicos (AS-

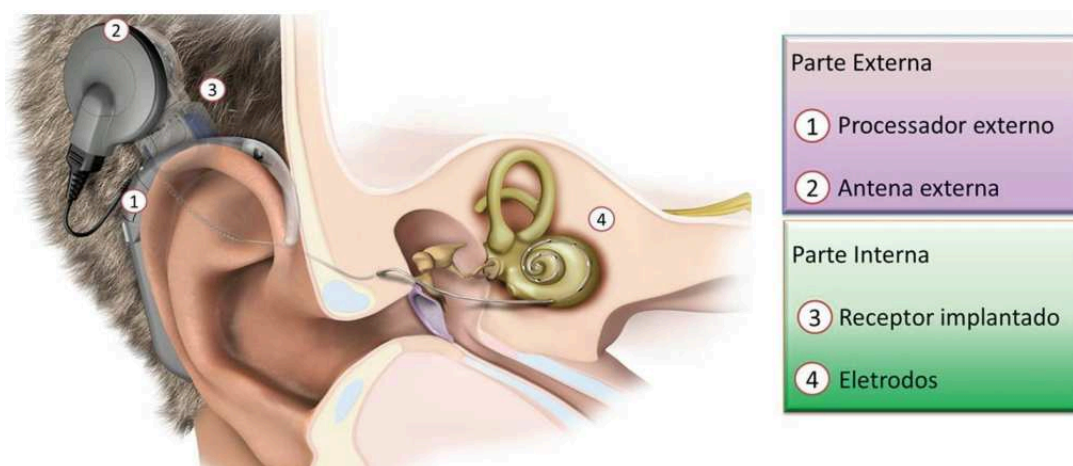
SOCIATION, 2019). A deficiência auditiva pode ser classificada por grau (de leve a profunda); por lateralidade (uma ou ambas as orelhas); ou por tipo (condutiva, sensório-neural, mista ou central) (SISTEMA DE CONSELHOS DE FONOAUDIOLOGIA, 2017). Tratamentos atuais para perda auditiva sensório-neural contam com a utilização de aparelhos auditivos acústicos, dispositivos implantáveis na orelha média e implantes cocleares (NEIL; RYUGO, 2012).

1.2 IMPLANTES COCLEARES

Implantes cocleares (IC) são próteses auditivas que restituem parcialmente a capacidade de perceber sons e compreender fala em pessoas com perda auditiva severa ou profunda (SHAPIRO; BRADHAM, 2012), desde que o nervo auditivo de pelo menos uma das orelhas ainda responda a estímulos elétricos.

Na Figura 3 vê-se uma ilustração do implante coclear comercial genérico, que é composto de uma parte externa e uma parte interna, não havendo conexão física entre elas. A parte externa consiste em um dispositivo (geralmente do tipo BTE – *behind the ear*) contendo um ou mais microfones; um processador de sinais, e uma antena de rádio-frequência (RF). Na parte implantada, encontram-se uma segunda antena de RF, o circuito de gerenciamento e de fontes de corrente, e um conjunto de eletrodos (de 16 a 24, dependendo do fabricante) inserido no interior cóclea de forma a estimular diretamente o nervo auditivo com pulsos de corrente elétrica.

Figura 3 – Ilustração do implante coclear. (1) Processador *behind the ear* com microfones; (2) antena transmissora; (3) antena receptora e circuito de fontes de corrente; (4) *array* de eletrodos (TEFILI *et al.*, 2013)



Na audição através do implante coclear, o processamento natural, realizado pela orelha externa, ossículos da orelha média e órgão de Corti, é ignorado e “substituído” por um processamento e estimulação artificiais.

O som é captado por um ou mais microfones, cujos sinais são processados geralmente por técnicas de formação de feixe (KOKKINAKIS *et al.*, 2012). O sinal resul-

tante é codificado pelo processador em sequências de pulsos que serão aplicadas nos diferentes eletrodos, de acordo com uma determinada estratégia de estimulação, como será mais detalhado na seção 1.3. O circuito da parte interna, ao receber a informação das sequências de pulsos, controla as fontes de corrente aplicando estímulos através dos eletrodos.

A antena externa é posicionada em alinhamento com a interna através de um ímã e se comunica com a parte implantada através de rádio-frequência. Durante o funcionamento do aparelho, o elemento externo envia as sequências de pulsos atribuídas a cada eletrodo, como também energia para alimentar os circuitos implantados (ZENG *et al.*, 2008). Por outro lado, a parte interna pode enviar informações sobre a impedância dos eletrodos no processo de ajuste do dispositivo, realizado no momento da implantação e em posteriores consultas ao fonoaudiólogo. Um outro eletrodo é geralmente colocado fora da cóclea e funciona como potencial de referência (terra).

Implantações bilaterais – implantes em ambos os lados sem comunicação entre si – já são uma realidade, e também há um crescente interesse de pesquisa em implantes biauriculares – dois implantes que se comunicam (DIETZ; MCALPINE, 2015; KOKKINAKIS; LOIZOU, 2010; LOPEZ-POVEDA *et al.*, 2016; MIRZAHASANLOO *et al.*, 2013). Entretanto, os implantes ainda são, atualmente, em sua grande maioria, unilaterais. Desta forma, o foco deste trabalho está nesse último caso.

1.3 ESTRATÉGIAS DE ESTIMULAÇÃO

As estratégias de estimulação visam a codificar o sinal de áudio de forma a otimizar a percepção sonora do usuário de implante coclear.

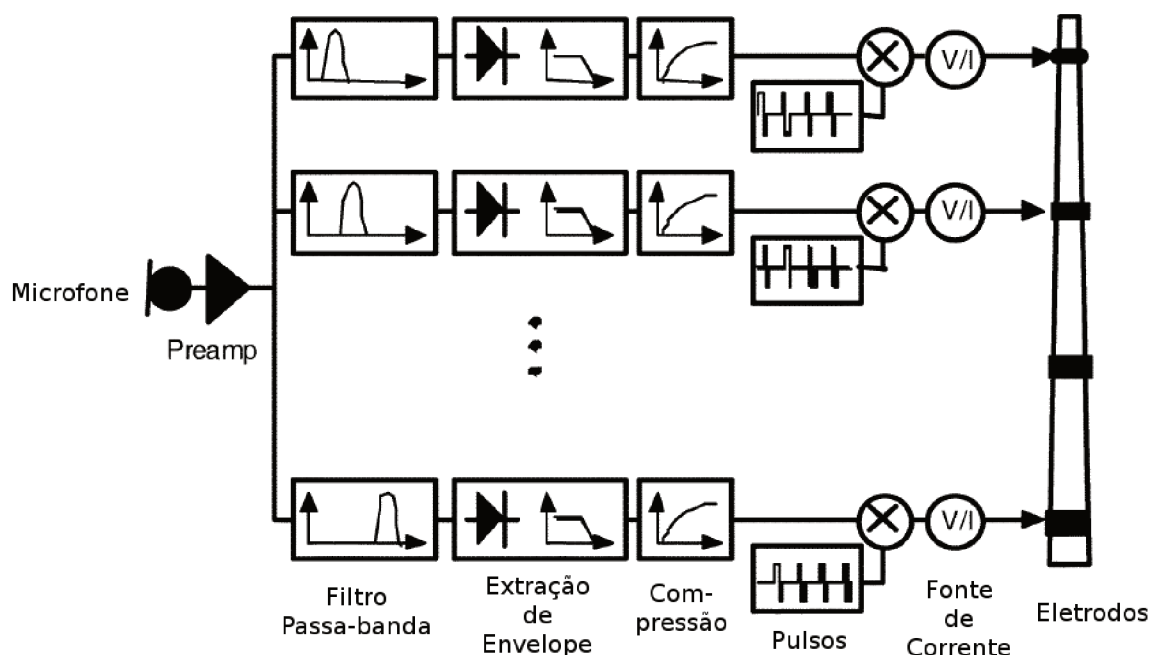
Os primeiros implantes eram monocanais, ou seja, possuíam apenas um eletrodo (LOIZOU, 1999b). Assim, não se aproveitavam da distribuição espacial tonotópica da cóclea, e criavam uma sensação auditiva bastante limitada, obtendo consequentemente, um desempenho ruim (menor que 5% (ZENG *et al.*, 2008)) na inteligibilidade da fala. Apesar disso, conseguiam transmitir alguma informação do envelope do sinal (LOIZOU, 1999b).

Quando do aparecimento do implante coclear multicanal esse quadro começou a mudar. As primeiras estratégias (por exemplo, F0/F2, F0/F1/F2, MPEAK) para esse novo tipo de implante buscavam enviar características do sinal, tais como as formantes, ao usuário (LOIZOU, 1999b). Entretanto, foi a introdução da estratégia *Continuous Interleaved Sample* (CIS) (WILSON, B. S. *et al.*, 1993) que fez o desempenho de inteligibilidade da fala dos usuários melhorar substancialmente, tornando esta técnica um marco no avanço dos implantes cocleares. Ainda hoje, esta estratégia ou variações dela são as mais usadas nos implantes comerciais.

1.3.1 CIS

O diagrama de blocos do processamento realizado na estratégia CIS é mostrado na Figura 4.

Figura 4 – Diagrama de blocos da estratégia CIS – adaptado de Zeng *et al.* (2008).

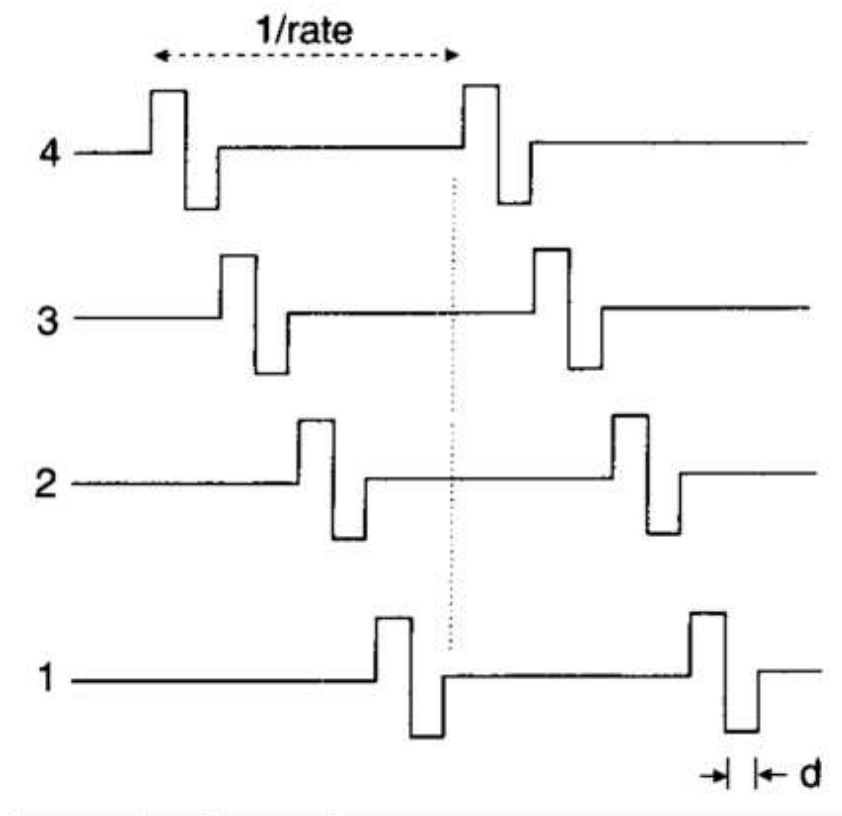


Nesta estratégia, o sinal de áudio é inicialmente dividido em sub-bandas de frequência, através de um banco de filtros. Na sequência, cada canal é submetido ao processamento descrito a seguir.

A envoltória do sinal é detectada através da transformada de Hilbert ou de um sistema de retificação e filtragem passa-baixas. O envelope é então submetido a uma operação de compressão logarítmica para mapear as amplitudes acústicas em uma faixa dinâmica de correntes (ZENGE *et al.*, 2002). O sinal resultante modula a amplitude de um trem de pulsos de largura e frequência fixas para estimulação elétrica. No caso da estratégia CIS, a taxa de estimulação por canal é maior que 800pps (pulsos por segundo), em contraste com suas predecessoras que raramente superavam 300pps (WILSON, B. S. *et al.*, 1993).

Cada canal (sub-banda) é associado a um dos eletrodos do *array*, e os trens de pulsos de corrente dos canais são sincronizados de tal forma que a estimulação aconteça em ciclos de varredura um a um, em sequência, do primeiro ao último eletrodo. Dois ou mais eletrodos nunca são acionados simultaneamente. A Figura 5 mostra esta configuração de pulsos no tempo para um modelo reduzido de apenas 4 eletrodos. A taxa de estimulação é denominada “*rate*”, e a largura dos pulsos, *d*. Os pulsos são sempre bifásicos, para que não ocorra um desequilíbrio, e consequente acúmulo, de cargas no interior da cóclea.

Figura 5 – Sincronização dos pulsos de corrente – adaptado de Loizou (1999b).



Os principais fatores responsáveis pelo sucesso da estratégia CIS são: o uso de estimulação não simultânea, minimizando a interferência inter-canal, e o uso de taxa de estimulação alta em relação às estratégias baseadas em formantes, favorecendo a representação de variações rápidas dos envelopes dos sinais acústicos.

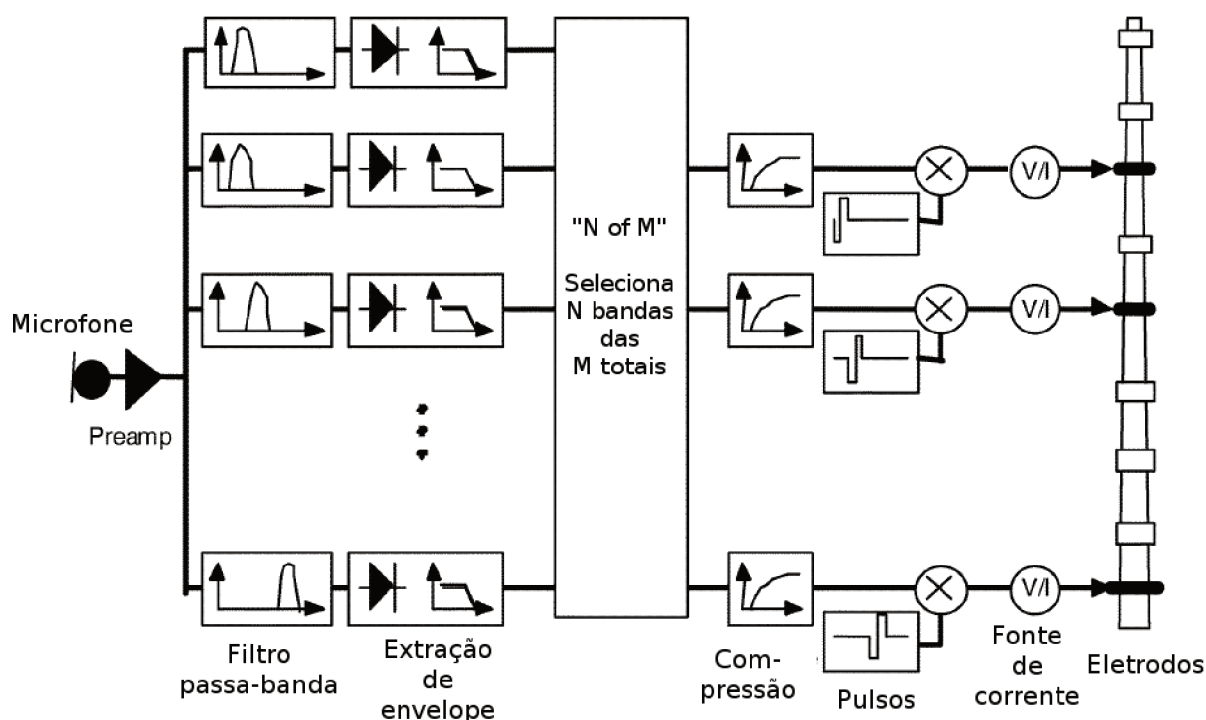
Vale notar que com a estratégia CIS, apenas a informação do envelope do sinal acústico é apresentada ao usuário de implante. Toda a informação de fase, ou estrutura temporal fina (TFS – *temporal fine structure*), do sinal é ignorada.

1.3.2 “N of M”

As estratégias do tipo “N of M” têm o processamento similar ao da CIS acrescido de um bloco seletor de canais (Figura 6).

O sinal é decomposto por um banco de filtros de M canais, mas apenas N são acionados em cada ciclo de estimulação. Geralmente são selecionados os canais com maior amplitude. Isso é o que acontece nas estratégias SPEAK, ACE e MP3000, presentes nos implantes da CochlearTM. Na primeira, o número de canais a serem ativados no ciclo varia de acordo com o sinal acústico (SOMEK *et al.*, 2006), enquanto na segunda, esse número é fixo (WILSON; DORMAN, 2008b). A estratégia MP3000, por sua vez, conta com uma seleção de canais baseada no efeito de mascaramento espectral (WOUTERS; MCDERMOTT; FRANCAERT, 2015).

Figura 6 – Diagrama de blocos de estratégias tipo "N of M-- adaptado de Zeng *et al.* (2008).



Hu e Loizou (2008) propõem a seleção dos canais nos quais a SNR (a priori) seja maior que 0dB, ou seja, dos canais em que há predominância de fala sobre o ruído. Os resultados apresentados indicam que esse critério de seleção pode restabelecer a inteligibilidade da fala em cenários acústicos de SNR global baixa (-10dB). Essa estratégia foi posteriormente denominada *opACE* (*optimal ACE*) (HU; LOIZOU, 2010).

1.4 DESEMPENHO E POSSÍVEIS MELHORAMENTOS

Tendo em vista o processamento e a informação apresentada a um usuário de implante coclear, algumas diferenças fundamentais em relação à audição normal são observadas. A resolução espectral é menor pois, como o número de eletrodos é limitado, apenas cerca de 20 regiões da cóclea podem ser estimuladas. Consequentemente, o número de sensações de *pitch* é consideravelmente reduzido em relação à audição normal. Além disso, o meio líquido no interior da cóclea é condutor, ocorrendo assim um espalhamento de corrente, o que faz com que ocorram interferências entre os canais, ou seja, a estimulação atinge uma área sensível relativamente grande.

A taxa de estimulação nas estratégias CIS e suas variações é constante, e apenas a informação de envelope do áudio é apresentada, ignorando-se completamente a informação temporal fina dos sinais. Desta forma, a resolução temporal também é reduzida.

Uma outra limitação é a falta de direcionalidade. Com implantes unilaterais, o

usuário não recebe pistas espaciais (e.g. diferença de nível interaural, ILD, e diferença de tempo interaural, ITD). Com implantação bilateral, algumas dessas pistas são percebidas, melhorando a inteligibilidade de fala e a localização de fontes (WILSON; DORMAN, 2008a). A transmissão dessas pistas pode ser, teoricamente, aprimorada com processamento biauricular. Entretanto, até o momento as estratégias propostas não aportaram benefícios significantes (WOUTERS *et al.*, 2013).

Apesar da limitação das informações efetivamente apresentadas aos implantados, o desempenho de inteligibilidade de fala de alguns usuários de implante coclear unilateral, atualmente, se aproxima ao de ouvintes normais, em uma série de testes de audição (WILSON; DORMAN, 2007). Entretanto, na presença de ruído e reverberação, esse desempenho cai significativamente. Em Hast *et al.* (2015), para uma relação sinal-ruído de +5dB, a inteligibilidade cai para cerca de 20%; quando em silêncio é de aproximadamente 80%. Já Bergeron e Hotton (2016) mostram que a inteligibilidade média é de 80%, 60%, 40% e 10%, para as condições de silêncio, e SNRs de +10dB, +5dB, e 0dB, respectivamente.

Dois caminhos se apresentam para o melhoramento dos resultados dos implantes unilaterais através do processamento de sinais. O primeiro consiste na criação de novas estratégias de estimulação, que apresentem uma informação mais qualificada para o usuário, incorporando de alguma forma pistas sobre a estrutura fina do sinal (CHEN; ZHANG, 2008, 2011; VERMEIRE; PUNTE; VAN DE HEYNING, 2010). Outra possibilidade seria o acionamento dos eletrodos em combinações, de forma a melhor concentrar a corrente elétrica. Desta forma, estimula-se com maior precisão regiões da cóclea e, conseqüentemente, melhora-se a resolução espectral (GEORGE *et al.*, 2015; PADILLA; LANDSBERGER, 2016; SRINIVASAN *et al.*, 2013). Trabalhos ainda estão sendo realizados com este intuito, mas, por enquanto, os resultados não são conclusivos ou trazem benefícios limitados (WOUTERS; MCDERMOTT; FRANCA, 2015).

A outra abordagem é a de se introduzir um bloco de redução de ruído ao processamento das estratégias já estabelecidas. A ideia é utilizar técnicas de redução de ruído, de forma a que o estímulo elétrico proporcionado ao usuário do implante represente um sinal de fala com uma melhor relação sinal-ruído (SNR) que o captado pelo microfone. Apesar de muitas dessas técnicas, na prática, não aportarem aumento significativo da inteligibilidade em pessoas com audição normal (HU; LOIZOU, 2007), quando usadas em implantes cocleares, podem levar a ganhos substanciais de inteligibilidade (KONING; MADHU; WOUTERS, 2015). Assim, resultados de estudos de percepção realizados com pessoas com audição normal não podem ser generalizados para implantados.

É nesta última abordagem que se baseia a presente tese.

1.5 MOTIVAÇÃO

Segundo a Organização Mundial da Saúde (OMS), em 2012, o número de pessoas com deficiência auditiva no mundo ultrapassava 360 milhões (TEFILI *et al.*, 2013). O Censo 2010, realizado pelo Instituto Brasileiro de Geografia e Estatística, contabilizou 9,8 milhões de brasileiros com deficiência auditiva, o que representava 5,2% da população brasileira. Deste total, 2,6 milhões eram surdos e 7,2 milhões apresentavam grande dificuldade para ouvir (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2012).

Estima-se que o número de pessoas, no mundo, que receberam um implante coclear até 2016 seja por volta de 500 mil, com a perspectiva que atinja 1 milhão até 2020, considerando-se o seu crescimento exponencial (WILSON, 2017). No Brasil, o governo federal, através do Sistema Único de Saúde (SUS), provê, de forma gratuita, a implantação desses dispositivos. Segundo a informação disponível no site do SUS¹, foram 7340 implantes cocleares realizados pelo SUS entre janeiro de 2008 e janeiro de 2019. A média aproximada é de 660 procedimentos por ano nesse período. Além disso, com o envelhecimento da população mundial, o número de pessoas com perda auditiva, e, conseqüentemente, o de possíveis candidatos a receberem um implante, tende a aumentar no futuro.

Segundo o *scopus.com*², o número de publicações sobre implantes cocleares também apresenta crescimento (ZENG; POPPER; FAY, 2004). A Figura 7 mostra o número de publicações sobre implantes cocleares ao longo dos anos até o ano de 2018 (o último completo até a escrita deste texto).

As publicações nas áreas de engenharia representam cerca de 7,2% do total, sendo que 4,5% destes (0,32% do total) contemplam o estudo de redução de ruído³. As figuras 8 e 9 mostram, respectivamente, o número cumulativo de publicações sobre redução de ruído para implantes cocleares por ano e a afiliação dos principais autores na área. Pode-se ver que a taxa de publicações aumentou significativamente nos últimos 10 anos, e que os principais grupos de pesquisas na área são o da Universidade de Texas em Dallas (UTD), nos Estados Unidos da América, e da Universidade Católica de Lovaina (KU Leuven), na Bélgica.

Como indicado anteriormente, diminuir a influência negativa que o ruído aditivo exerce sobre o nível de inteligibilidade de portadores de implantes cocleares poderá trazer melhoras significativas em sua qualidade de vida. Uma das maneiras mais pro-

¹ Acesso a <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sih/cnv/qiuf.def> em 25/03/2019.

² Pesquisa em *scopus.com* pelos termos TITLE-ABS-KEY ((cochlear AND implant*) em 26/03/2019.

³ Pesquisa em *scopus.com* pelos termos TITLE-ABS-KEY ((cochlear AND implant*) AND ("noise reduction" OR "noise suppression" OR "reverberation reduction" OR "reverberation suppression" OR "speech enhancement")) AND (LIMIT-TO (SUBJAREA , "ENGI")) em 25/03/2019.

Figura 7 – Número de publicações sobre implante coclear por ano. Os dados foram obtidos através de pesquisa na base de dados scopus . com, procurando pelo termo cochlear AND implant*.

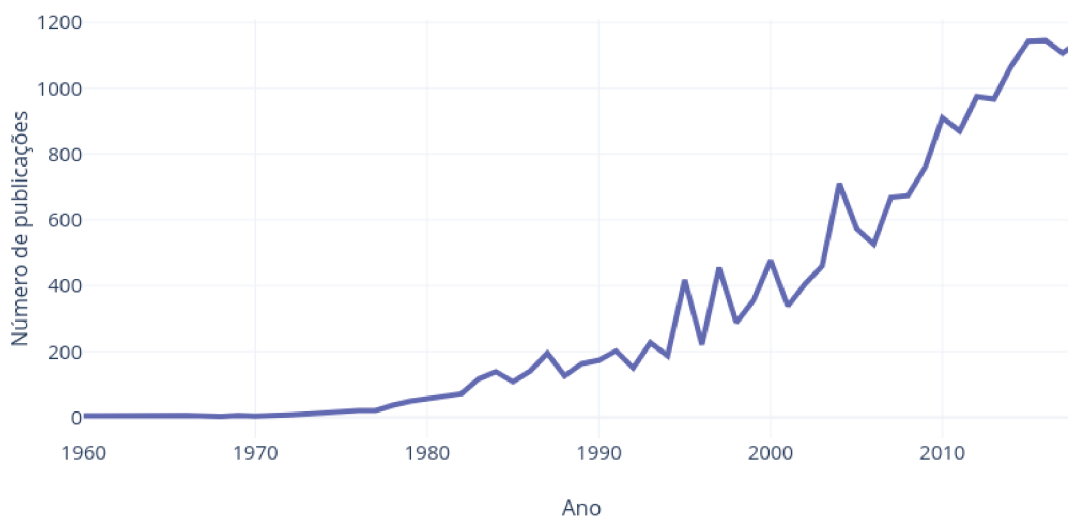


Figura 8 – Número cumulativo de publicações sobre redução de ruído para implantes cocleares por ano.

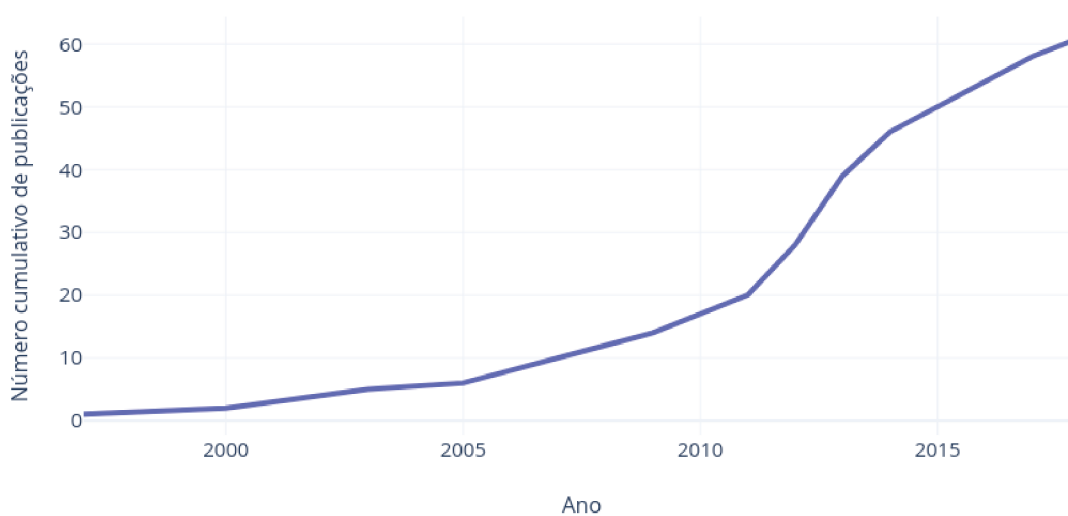
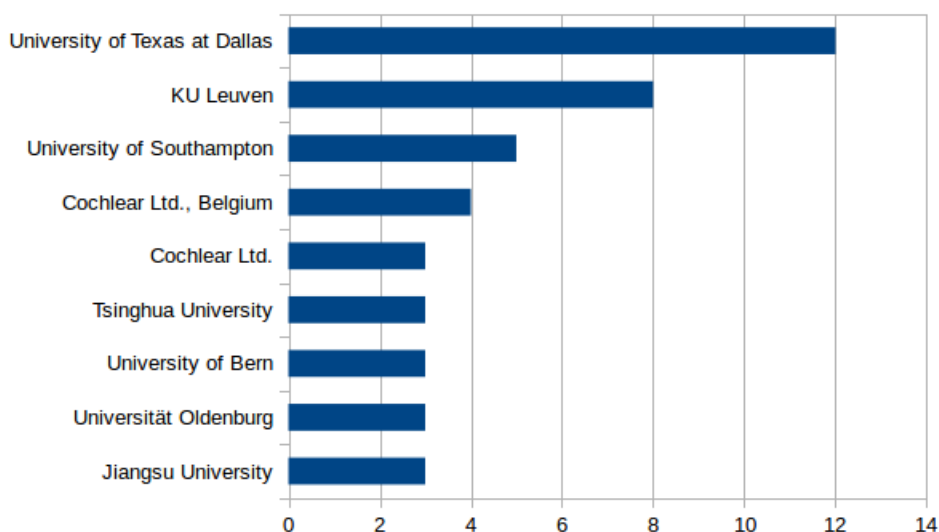


Figura 9 – Afiliação dos principais autores na área.



missoras de atingir esse objetivo é através de técnicas de redução de ruído adaptadas para as especificidades desta aplicação.

Em suma, o número de usuários de implantes cocleares está aumentando e com perspectiva de continuar essa tendência. Além disso, há um interesse igualmente crescente no estudo de implantes cocleares e um problema posto para a área de processamento de sinais: a redução de ruído para o aprimoramento da percepção acústica dos implantados. Além disso, há espaço para a inserção na comunidade científica nesse domínio, pois esta ainda é pequena relativamente às possibilidades de pesquisa nesse campo.

1.6 JUSTIFICATIVA

Apesar da limitada representação de sons por um implante coclear, o desempenho de seus usuários com respeito à inteligibilidade da fala é surpreendentemente elevado, e comparável ao de pessoas com audição normal. Entretanto, o desempenho dos usuários cai significativamente quando na presença de ruído. Essa é uma situação frequente no cotidiano, seja em ambientes externos, com o som de motores, vento, máquinas, como também em ambientes internos, como, por exemplo, restaurantes ou salões de eventos com muitas pessoas conversando em paralelo.

Um caminho promissor para se atenuar os efeitos negativos do ruído na inteligibilidade da fala consiste no uso de técnicas de redução de ruído através do processamento do sinal de áudio. A aplicação destas soluções em aparelhos já implantados pode ser feita através da reprogramação dos processadores, não sendo necessárias intervenções cirúrgicas.

1.7 OBJETIVOS

Nesta seção são destacados os objetivos geral e específicos do trabalho.

1.7.1 Objetivo Geral

O objetivo geral deste trabalho é o estudo e o desenvolvimento de técnicas de redução de ruído dedicadas a implantes cocleares, considerando as especificidades desses dispositivos e visando a melhora da compreensão de fala dos implantados na presença de ruído aditivo.

1.7.2 Objetivos Específicos

Os objetivos específicos do trabalho são:

- Levantamento das técnicas atualmente utilizadas para a redução de ruído em ICs;
- Fazer a crítica dos métodos atuais de redução de ruído em ICs, propondo uma alternativa;
- Desenvolvimento de um método de redução de ruído especificamente para ICs, considerando suas características particulares;
- Realizar a avaliação dos métodos propostos através de simulações numéricas e métricas objetivas;
- Conceber e realizar experimentos psicoacústicos, tanto com normouvintes quanto com usuários de IC, para a avaliação da redução de ruído.

1.8 ORGANIZAÇÃO DO DOCUMENTO

Este documento está dividido em capítulos. Neste primeiro capítulo encontra-se uma introdução a implantes cocleares, com foco no processamento de sinais. Foram descritas as estratégias mais usadas de estimulação elétrica do nervo auditivo. Apresentou-se a motivação, justificativa e objetivos do trabalho.

Na sequência, o capítulo 2 apresenta a descrição de máscaras tempo-frequência, a principal abordagem utilizada para redução de ruído em ICs. Na sequência encontra-se uma revisão das principais máscaras tempo-frequência utilizadas em implantes cocleares.

Uma técnica de redução de ruído através de uma nova máscara tempo-frequência conformável é proposta no capítulo 3. Simulações numéricas extensas foram realizadas utilizando métricas objetivas relacionadas à percepção de qualidade e inteligibilidade da fala tanto por normouvintes quanto por usuários de IC. Os

resultados e o suporte teórico indicam a viabilidade da utilização da máscara proposta e suas vantagens em relação às mais frequentemente encontradas na literatura.

No capítulo 4, é desenvolvido um filtro baseado especificamente para ICs, nos quais a informação de fase (ou TFS) do sinal é desconsiderada. Esse filtro é a solução ótima que minimiza o erro quadrático médio entre o envelope temporal da fala e sua estimação. Seu potencial para a aplicação em implantes cocleares é comprovado através de simulações numéricas e experimentos psicoacústicos.

Finalmente, no capítulo 5 encontra-se uma discussão geral à luz dos resultados obtidos e a conclusão do trabalho com a enumeração das contribuições aportadas pela tese.

2 TÉCNICAS DE REDUÇÃO DE RUÍDO PARA IMPLANTES COCLEARES

Como visto no capítulo anterior, o reconhecimento de fala de portadores de implantes cocleares é consideravelmente reduzido na presença de ruído ambiente. Esse problema pode ser contornado através de técnicas de processamento de sinais para a redução de ruído. No caso específico de implantes cocleares unilaterais, as propostas se dividem em duas abordagens principais: formação de feixe acústico (*beamforming* acústico) e redução de ruído de canal único (*single-channel*).

Na primeira, faz-se uso de dois (ou mais, se disponíveis) microfones, cujos sinais são combinados de forma a atenuar o som vindo da direção da fonte de ruído e/ou destacar a fala de interesse (BÜCHNER *et al.*, 2014; CHEN; GONG, 2012, 2013; HERSBACH *et al.*, 2013; KOKKINAKIS; COX, 2014; KORDUS *et al.*, 2015). Esta abordagem não será tratada neste trabalho.

Com relação à redução de ruído de canal único, destacam-se as máscaras tempo-frequência, que são detalhadas a seguir.

2.1 MÁSCARAS TEMPO-FREQUÊNCIA

Técnicas de aprimoramento da fala baseadas em máscaras tempo-frequência têm sido estudadas por muitos anos com o objetivo de melhorar a inteligibilidade da fala, redução de ruído e conforto acústico (LOIZOU, 2013). Nesse contexto, aparelhos auditivos são uma importante área de aplicação dessas técnicas, que ainda necessitam de aprimoramentos, de forma a aumentar a utilização diária de tais aparelhos por seus portadores (HAMACHER *et al.*, 2005).

Nas aplicações em aparelhos auditivos, o sinal captado é comumente decomposto em múltiplas bandas de frequência e então é processado de acordo com uma dada regra de supressão. Em geral, a função de supressão consiste de um vetor de ganhos (também chamado de máscara), que é ajustado para cada banda em cada janela temporal de análise (*frame*), de forma a minimizar uma determinada função custo, muitas vezes associada a um compromisso entre redução de ruído e distorção da fala.

Devido às características do aparelho, usuários de ICs estão acostumados a ouvir com informações acústicas limitadas em comparação a indivíduos com audição normal. Assim, é justificável o uso de técnicas que permitam uma maior atenuação do ruído mesmo que isso implique na introdução de maiores níveis de distorção no sinal de fala¹.

Duas técnicas muito utilizadas para a implementação de tais máscaras são o fil-

¹ Por exemplo, se uma determinada técnica de redução de ruído resulta em uma distorção de fase do sinal de fala, o usuário de IC não perceberá essa distorção, uma vez que, devido ao processamento de sinais do IC (visto no capítulo 1), a informação de fase não lhe é apresentada de qualquer maneira.

tro de Wiener (WF) (LOIZOU, 2013) e a Máscara Binária (BM) (WANG; BROWN, 2006). Ambas as técnicas já se mostraram bem sucedidas na melhora da inteligibilidade da fala por portadores de ICs, sendo o WF preferido em relação à BM. Esta preferência é, porém, muito menos proeminente que no caso de pessoas com audição normal (KONING; MADHU; WOUTERS, 2015).

O problema é matematicamente descrito da seguinte forma. Consideremos que um sinal de fala contaminado é definido por $y[n] = x[n] + v[n]$, em que $x[n]$ é o sinal de fala desejado, e $v[n]$ é o ruído. Ambos $x[n]$ e $v[n]$ são considerados não-observáveis individualmente, além de descorrelacionados entre si². Tomando-se a Transformada de Fourier de tempo curto (STFT) de N -bins para uma janela temporal finita de $y[n]$, tem-se:

$$Y(k, \lambda) = X(k, \lambda) + V(k, \lambda), \quad (1)$$

em que k e λ são, respectivamente, os índices da banda de frequência e da janela de tempo.

A abordagem tempo-frequência de enfatização da fala consiste em, a cada janela temporal, multiplicar o sinal ruidoso $Y(k, \lambda)$ por uma máscara de ganho $M(k, \lambda)$, para gerar uma estimativa do sinal de fala desejado $\hat{X}(k, \lambda)$ (em cada bin k), como explicitado na equação (2).

$$\hat{X}(k, \lambda) = M(k, \lambda) Y(k, \lambda) \quad (2)$$

O sinal de fala estimado, no domínio do tempo, é obtido por

$$\hat{x}(n) = \text{IFT}_{\text{OLA}}\{|\hat{X}(k, \lambda)| e^{j\angle Y(k, \lambda)}\}, \quad (3)$$

em que $\text{IFT}_{\text{OLA}}\{\cdot\}$ é a transformada inversa de Fourier (associada a uma estratégia *overlap-and-add* (CROCHIERE, 1980)) e $\angle Y(k, \lambda)$ é a fase do sinal contaminado.

Existem diferentes abordagens para se definir a máscara de supressão $M(k, \lambda)$. Em geral, as máscaras são baseadas em função da relação sinal-ruído (SNR) *a priori* associada a cada banda de frequência k em uma janela temporal λ , que é dada por:

$$\xi(k, \lambda) = \frac{P_X(k, \lambda)}{P_V(k, \lambda)}, \quad (4)$$

$P_X(k, \lambda)$ e $P_V(k, \lambda)$ são as funções de densidade espectral do sinal de fala limpo e do ruído, respectivamente.

Vale ressaltar que, na utilização dessas técnicas em dispositivos reais, os sinais $x[n]$ e $v[n]$ são desconhecidos e não é possível medir a real SNR *a priori*. Assim,

² Neste trabalho apenas será considerado o ruído aditivo. A reverberação é usualmente modelada como a convolução do sinal de fala com a resposta ao impulso da sala (NAYLOR *et al.*, 2010 apud HAZRATI *et al.*, 2013), e não é considerada neste trabalho.

$\xi(k, \lambda)$ deve ser estimado, o que pode ser feito aplicando-se, por exemplo, o método de decisão direta ou suas variações (ALAM; O'SHAUGHNESSY; SELOUANI, 2008; EPHRAIM; MALAH, 1984), ou, como no caso específico de alguns implantes cocleares, da análise da função de coerência entre os sinais captados por dois microfones próximos (YOUSEFIAN; LOIZOU, 2012).

Entretanto, em condições controladas, tendo-se conhecimento dos sinais $x[n]$ e $v[n]$, é comum a utilização dos valores exatos de $\xi(k, \lambda)$. Isso permite a avaliação da máxima performance das máscaras. As máscaras calculadas dessa maneira exata são chamadas de *ideais*.

As principais máscaras utilizadas em ICs são descritas a seguir. O índice λ será omitido daqui em diante para que a notação fique mais enxuta.

2.2 MÁSCARA BINÁRIA

A mais simples função de supressão é a máscara binária (BM), definida pela seguinte equação:

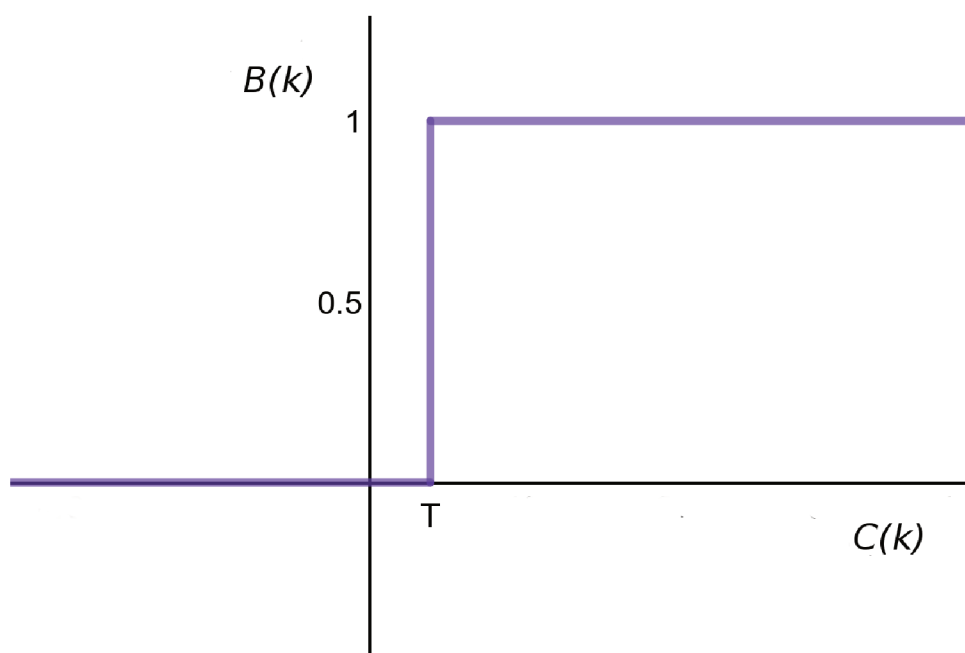
$$M(k) = B(k) = \begin{cases} 1 & : C(k) \geq T \\ 0 & : C(k) < T, \end{cases} \quad (5)$$

em que C é um determinado critério de seleção, geralmente definido como sendo a relação sinal ruído na frequência k , e T é o limiar de decisão, usualmente definido como $\text{SNR}_k = 0\text{dB}$. Devido à descontinuidade em seus ganhos (binários), esta abordagem é classificada como uma máscara dura (*hard mask*). A Figura 10 mostra a função de ganho da máscara binária. Nela pode-se ver a diferença na transição de ganho entre 0 e 1, que justifica essa classificação.

Algoritmos de máscara binária têm sido utilizados mais em aplicações de reconhecimento automático da fala do que em redução de ruído (LOIZOU, 2013). Em Cooke *et al.* (2001), sua utilização é baseada no fato de que no processamento auditivo natural, pessoas conseguem lidar com a falta de informação, isto é, mesmo sendo apresentados sinais incompletos, é possível compreender a fala. Assim, um sistema que utiliza apenas informações consistentes do sinal (trechos com predominância de fala sobre o ruído), ignorando-se a parte complementar, funcionaria de maneira robusta. Desta forma, a função de ganho da máscara binária não foi originalmente derivada como uma solução ótima que minimiza uma função custo, mas sim, uma solução heurística baseada em critérios psicoacústicos.

O critério de seleção mais frequente é a relação sinal-ruído (SNR) *a priori* ($\xi(k)$), definida na equação (4). Considerando esse critério de seleção, $T = \mu_0$ é um limiar de decisão, geralmente ajustado para $\mu_0 = 0\text{dB}$. Nesse caso, apenas os *bins* com predominância de sinal de interesse no *frame* λ são preservados, sendo a parte

Figura 10 – Ganho de uma máscara binária em função na SNR do canal k . A transição entre 0 e 1 acontece de forma abrupta a partir de um determinado valor do critério $C(k) = T$, por isso a máscara é classificada como dura.



complementar do sinal totalmente suprimida.

A máscara binária calculada com valores exatos de $\xi(k)$ e que usa esse critério de seleção é denominada *ideal* (IBM - *ideal binary mask*). Respeitadas algumas condições, é a escolha ótima entre todas as máscaras binárias (LI; WANG, 2009). Lu e Loizou (2011) mostraram que a IBM é equivalente ao estimador máximo *a posteriori* (MAP) do quadrado da magnitude espectral (X_k^2) do sinal alvo, considerando-se as partes real e imaginária de $X(k)$ e $V(k)$ como variáveis gaussianas independentes.

A IBM foi usada para redução de ruído para pessoas com audição normal, mostrando-se eficaz para o aumento de inteligibilidade da fala na presença de voz competidora (usando valores de μ_0 entre -10 dB e 0 dB (BRUNGART *et al.*, 2006), e -20 dB $< \mu_0 < +5$ dB (LI; LOIZOU, 2008b)). Para essas pessoas, a máscara binária é a única técnica que resulta em melhoras significativas de inteligibilidade (HU; LOIZOU, 2007; LOIZOU, 2013). A IBM se apresenta como uma técnica robusta. A inteligibilidade de frases contaminadas com ruído com características estatísticas semelhantes às da fala pode ser praticamente restituída usando-se a IBM com bancos de filtros de 16 canais ou mais (LI; LOIZOU, 2008a; WANG *et al.*, 2008). Além disso, com até 60% de falsos negativos³ a inteligibilidade ainda se manteve acima de 90%, enquanto que com apenas 20% de falsos positivos⁴, por outro lado, registrou-se uma queda de 20% em inteligibilidade, resultando em 80% de acertos (LI; LOIZOU, 2008b).

Em contrapartida, em testes com usuários de implante coclear, falsos positivos

³ falsos negativos são erros em que a máscara binária tem valor 0, quando idealmente deveria ser 1.

⁴ falsos positivos são erros em que a máscara binária tem valor 1, quando idealmente deveria ser 0.

se mostraram tão danosos à inteligibilidade quanto falsos negativos (KRESSNER *et al.*, 2016). Kressner *et al.* (2016) mostram que implantados são menos tolerantes que pessoas com audição normal a erros de estimação da máscara quando esses erros ocorrem de forma estruturada (em *clusters*) – como é o caso quando se usam algoritmos de estimação realistas.

Para implantes cocleares, Hu e Loizou (2008) propuseram uma estratégia de estimulação do tipo *N-of-M* em que o critério da seleção de canais é a relação sinal-ruído do canal em cada *frame* temporal. Apenas os canais com $SNR > 0$ dB são selecionados para estimulação, contrariamente a estratégias como a ACE, em que os canais cujos sinais possuem os maiores valores de envelope são selecionados. A técnica levou a melhoras significativas no desempenho de inteligibilidade com implantados, mesmo em situações com até 25% de erros na seleção de canais.

Uma técnica similar, denominada IRM (*Ideal Reverberation Mask*), foi proposta por Kokkinakis, Hazrati e Loizou (2011), em que o critério de seleção de canais é a relação sinal-reverberação (SRR), definida pela equação 6, representada em dB.

$$SRR = 10 \log_{10} \frac{|X(k,t)|^2}{|Y(k,t)|^2} \quad (6)$$

Vale ressaltar que, em aplicações práticas, $|X(k,t)|$ é desconhecido e a SRR deve ser estimada (THIERGART; DEL GALDO; HABETS, 2012).

Para sinais corrompidos com reverberação ($RT_{60} = 1$ s)⁵, a inteligibilidade foi aprimorada de 20% para 85% (experimentos com usuários de ICs), usando um limiar de $SRR = -5$ dB. Entretanto, a dificuldade em se estimar a SRR a partir do sinal com reverberação se apresenta como obstáculo para a implementação dessa estratégia em situações reais.

Motivados pela relação inversamente proporcional entre a reverberação e o momento de quarta ordem do sinal de fala (*kurtosis*), um método foi proposto para estimar uma máscara binária para reverberação usando apenas o sinal contaminado (HAZRATI; LEE; LOIZOU, 2013). Ganhos significativos de inteligibilidade foram observados em condições de reverberação $RT_{60} = 0,6$ s e $RT_{60} = 0,8$ s, mesmo ainda sendo consideravelmente menores que os obtidos com a máscara ideal (IRM).

Apesar de trazerem resultados promissores, as máscaras binárias podem não ser adequadas em cenários acústicos realísticos. Isso porque sons ambientes potencialmente importantes (sirenes, alarmes, entre outros) podem ser completamente omitidos do usuário de IC, quando o ganho aplicado é 0. Uma maneira de contornar isso é fazer uma transição de ganho mais suave entre os valores extremos (HU; LOIZOU, 2008).

⁵ A quantidade de reverberação é representada pelo tempo de reverberação (RT_{60}), que é definido como o tempo transcorrido entre a emissão de uma onda sonora e o momento em que seu nível decai de 60 dB, o que corresponde a uma diminuição de ordem 10^6 W/m² de intensidade sonora.

A seguir algumas máscaras suaves já utilizadas em pesquisas sobre implantes cocleares são descritas.

2.3 FILTRO DE WIENER E SUAS VARIANTES

Entre as máscaras consideradas suaves (*soft masks*), pode-se destacar o filtro de Wiener (WF) e suas variantes, o filtro de Wiener paramétrico e o filtro de Wiener restrito.

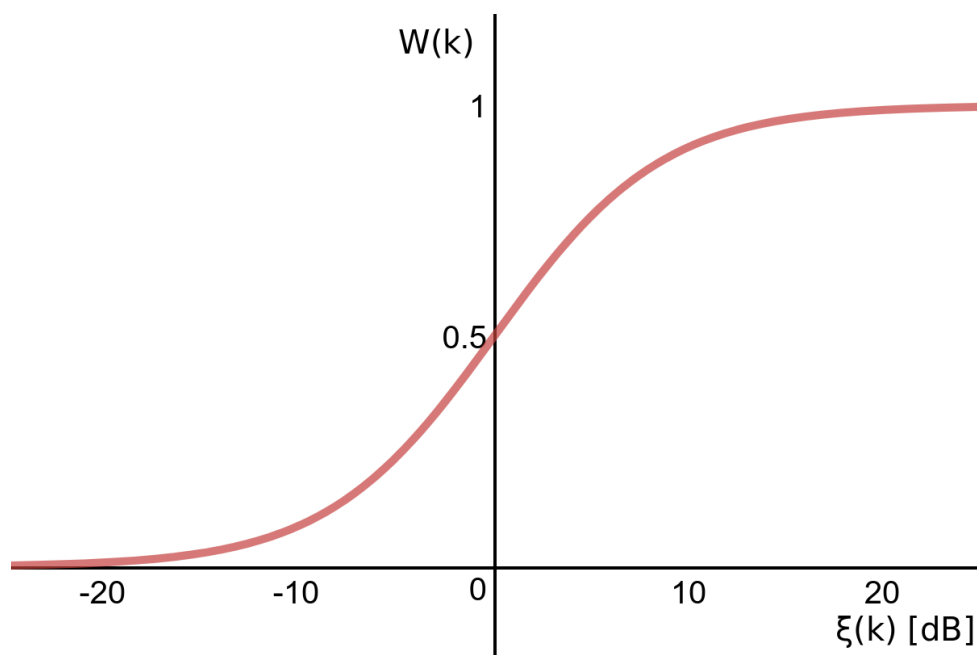
O filtro de Wiener no domínio STFT é descrito pela equação a seguir (LOIZOU, 2013):

$$M(k) = W(k) = \frac{\xi(k)}{\xi(k) + 1}, \quad (7)$$

em que $\xi(k)$ é a SNR a priori definida pela equação (4).

O gráfico da função de supressão é mostrado na Figura 11.

Figura 11 – Ganho de uma máscara de Wiener em função da SNR do canal k . A transição entre 0 e 1 acontece de forma gradativa com o aumento da SNR, por isso a máscara é classificada como suave.



A máscara de Wiener é a solução ótima no sentido do mínimo erro quadrático médio (MMSE), ou seja, é a máscara de supressão que minimiza a seguinte função custo (LOIZOU, 2013):

$$J = E \left\{ |\hat{X}(k) - X(k)|^2 \right\}. \quad (8)$$

Apesar de ser uma solução ótima (no sentido MMSE), sua utilização introduz distorções indesejadas na percepção do sinal acústico, tal como ruído musical (SILVA;

BERMUDEZ, 2011). Assim, algumas variações foram propostas, entre as quais o filtro de Wiener paramétrico e o filtro de Wiener restrito.

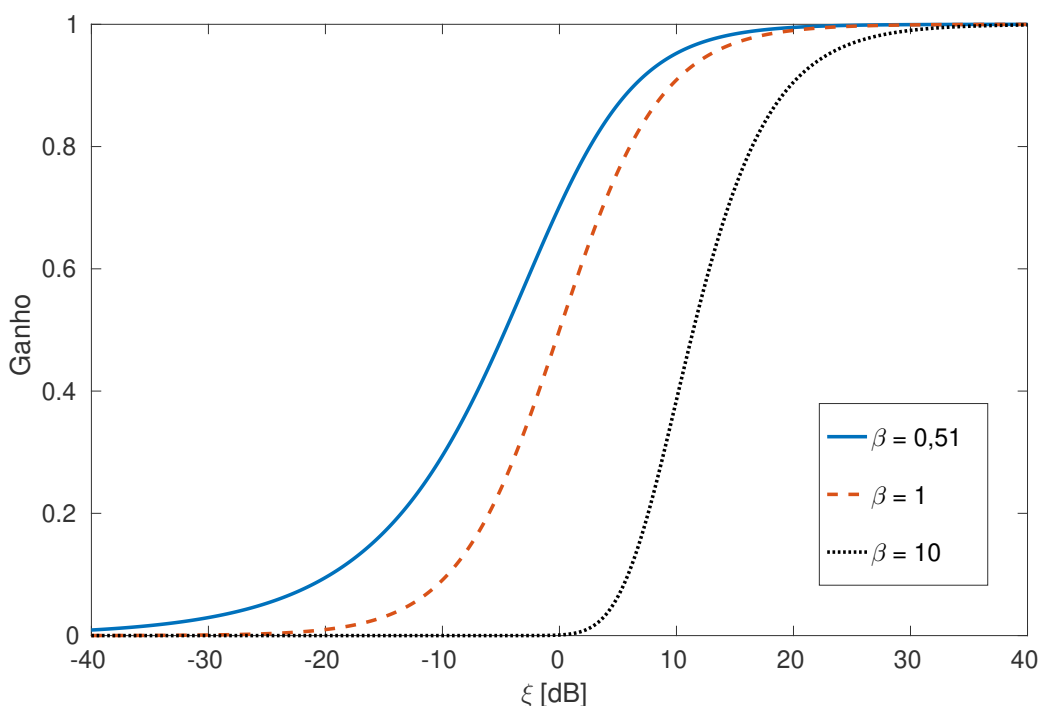
O filtro de Wiener paramétrico (WP), dado pela equação (9) (LIM; OPPENHEIM, 1979), é uma generalização do filtro de Wiener, com a introdução de parâmetros, β e η que permitem uma maior maleabilidade da função de acordo com a conveniência.

$$M(k) = W_p(k) = \left(\frac{\xi(k)}{\xi(k) + \eta} \right)^\beta \quad (9)$$

Pode-se verificar que para $\beta = 1$ e $\eta = 1$, obtém-se o filtro de Wiener (LOIZOU, 2013).

O parâmetro η é associado ao limiar de transição da função de ganho, enquanto β está associado ao quão abrupta é essa transição. Entretanto, a modificação do parâmetro β também altera o limiar (MAUGER; DAWSON; HERSBACH, 2012) (como mostrado na Figura 12), o que pode dificultar o projeto do filtro. Além disso, uma configuração genérica dos parâmetros β e η não tem propriedades ótimas associadas (BENESTY; SONDEHI; HUANG, 2008, p.857).

Figura 12 – Efeito da variação do parâmetro β no filtro de Wiener paramétrico. Todas as curvas têm parâmetro $\eta = 0$ dB



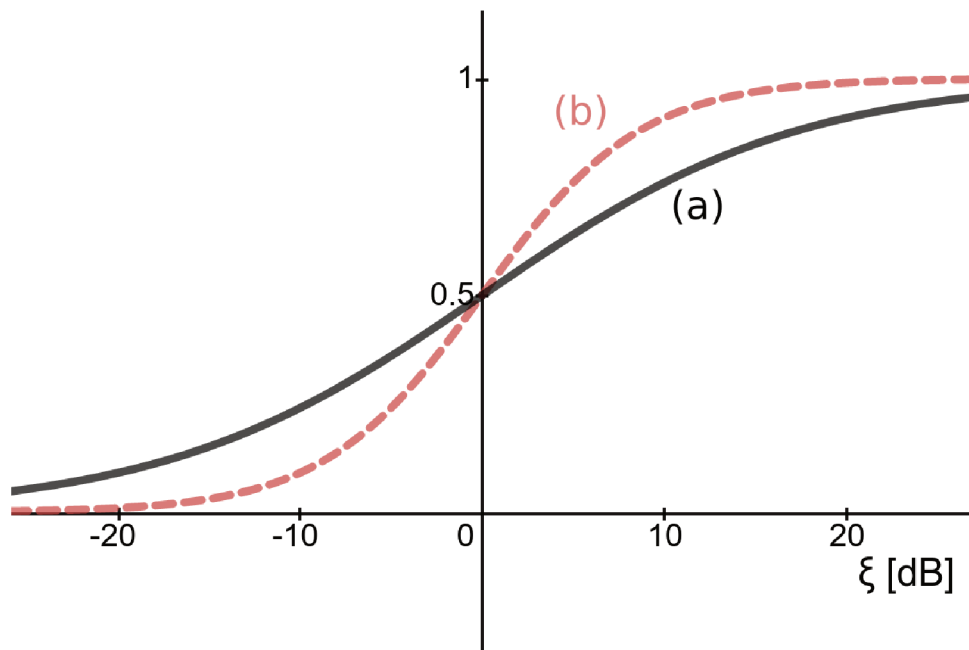
Os parâmetros β e η podem ser ajustados de acordo com diferentes critérios. Em Chehresa e Savoji (2009), η e β são funções de estatísticas de ordem superior, respectivamente, *skewness* e *kurtosis* de modelos dos sinais de fala e ruído. Já em Silva e Bermudez (2011), η é fixo e igual a 1, enquanto o valor de β se adapta a cada *frame* como uma função do *pitch*.

O filtro de Wiener restrito (*constrained Wiener filter*), por sua vez, é aquele que iguala a distorção relativa ao ruído com a distorção relativa à fala (LOIZOU, 2013), conceitos que serão definidos na próxima seção, sendo definido pela equação a seguir.

$$M(k) = W_c(k) = \frac{\sqrt{\xi(k)}}{\sqrt{\xi(k)+1}} \quad (10)$$

Esse filtro tem uma transição mais suave que o filtro de Wiener, o que é ilustrado na Figura 13. Isso pode reduzir distorções introduzidas pela atenuação excessiva de componentes com baixa SNR (CHEN; LOIZOU, 2010).

Figura 13 – Em (a), ganho de uma máscara de Wiener restrito, equação (10), em função da SNR do canal k . Para efeito de comparação, em (b) é apresentado o ganho do filtro de Wiener, equação (7)



Dentre as máscaras suaves propostas para implantes cocleares, o filtro de Wiener paramétrico é o mais referenciado, sendo seus parâmetros determinados de diferentes formas em cada estudo.

Uma função de ganho com formato sigmoidal dada pela equação (11) foi proposta por Yi Hu *et al.* (2007):

$$M_{Hu}(k) = \exp\left(-\frac{2}{\xi(k)}\right). \quad (11)$$

Essa função não é derivada como uma solução ótima, e foi aparentemente escolhida arbitrariamente por seu formato sigmoidal quando o gráfico é traçado com o eixo das abcissas representando ξ em dB. Vale notar que a equação (11) é um caso particular do filtro paramétrico de Wiener (MAUGER; DAWSON; HERBACH, 2012)⁶.

⁶ A equação (11) pode ser reescrita da seguinte maneira: $\exp\left(-\frac{2}{\xi(k)}\right) = \lim_{n \rightarrow \infty} \left(\frac{\xi(k)}{\xi(k) + 2/n}\right)^n$.

O uso dessa máscara resultou em melhoras de inteligibilidade de 10 a 15 pontos percentuais. Um desempenho melhor foi observado em relação a outros métodos propostos anteriormente, tais como o algoritmo de subespaço apresentado em Loizou, Lobo e Hu (2005) e Toledo, Loizou e Lobo (2003) e o método de subtração espectral proposto por Yang e Fu (2005). Estas abordagens aparecem apenas nos estudos citados e não foram encontradas novas referências a elas. Por isso, não são detalhadas aqui.

Em Mauger, Dawson e Hersbach (2012), o filtro de Wiener paramétrico com uma série de valores de η e β foi testado subjetivamente com implantados, e uma região de preferência foi obtida. Esses valores ($\beta = 1,2$ e $\eta = 2,92$) são usados em um filtro (MAUGER; ARORA; DAWSON, 2012) cujo desempenho, para usuários de IC, foi significativamente superior em comparação a um filtro projetado para pessoas com audição normal, para ruído de balbuciação (*babble*) e ruído ponderado por fala (SWN – *speech-weighted noise*).

Hazrati *et al.* (2013) usam arbitrariamente $\beta = 1,5$ e $\eta = 2$ e $\xi(k)$ estimado pela relação sinal-distorção (SDR), de forma a que os efeitos de reverberação também sejam atenuados. Essa máscara suave levou a aumentos significativos de inteligibilidade da fala em implantados.

Em He, Zheng e Meng (2017), β é fixado em 1 e $\eta = 10^{\frac{aGT}{10}}$ é calculado como uma função do limiar aparente de transição (aGT), definido como a SNR para a qual o ganho é igual a 0,5. Verificou-se que o melhor valor de η é diferente para cada indivíduo.

Há algoritmos de redução de ruído comerciais, mas não há muita informação disponível sobre detalhes de seu funcionamento. O sistema de redução de ruído *VoiceTrack*[®] da *Neurelec/Oticon* é baseado em uma implementação do filtro paramétrico de Wiener (ARSLAN, 2006) e mostrou ganhos significativos de inteligibilidade e qualidade (GUEVARA *et al.*, 2016), sobretudo após um mês de utilização (BERGERON; HOTTON, 2016). O algoritmo *ClearVoice*[™], usado pela *Advanced Bionics*, também usa algum tipo de máscara tempo-frequência, e apresenta melhora na tolerância ao ruído, mas nenhum acréscimo significativo na inteligibilidade (DINGEMANSE; GOEDGEBURE, 2014), o que é alcançado quando utilizado em conjunto com um algoritmo de direcionamento de feixe acústico (*beamforming*) (HEHRMANN *et al.*, 2012).

As principais técnicas de redução de ruído já propostas para implantes cocleares foram apresentadas, sendo elas máscaras tempo-frequência duras ou suaves. Quanto à inteligibilidade, o uso da máscara binária ou de uma máscara suave (filtro de Wiener) resultou em desempenho similar para implantados (KONING; MADHU; WOUTERS, 2015). Os resultados, entretanto, mostram uma possível pequena diferença a favor da máscara suave para sinais com degradação elevada (SNR = -25dB). Qualitativamente, máscaras suaves foram preferidas por usuários de implante, mas de maneira menos

proeminente que por pessoas de audição normal (MADHU *et al.*, 2013).

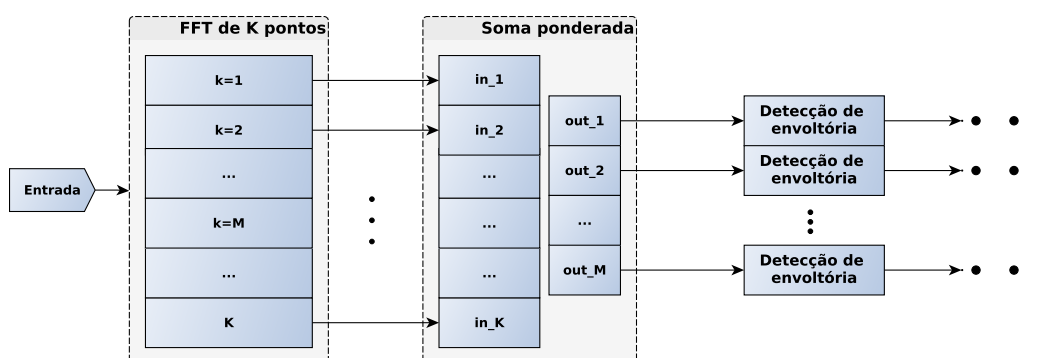
A seguir, discutiremos como essas técnicas podem ser inseridas no processamento das estratégias de estimulação baseadas na CIS.

2.4 IMPLEMENTAÇÃO DE TÉCNICAS DE REDUÇÃO DE RUÍDO E IMPLANTES COCLEARES

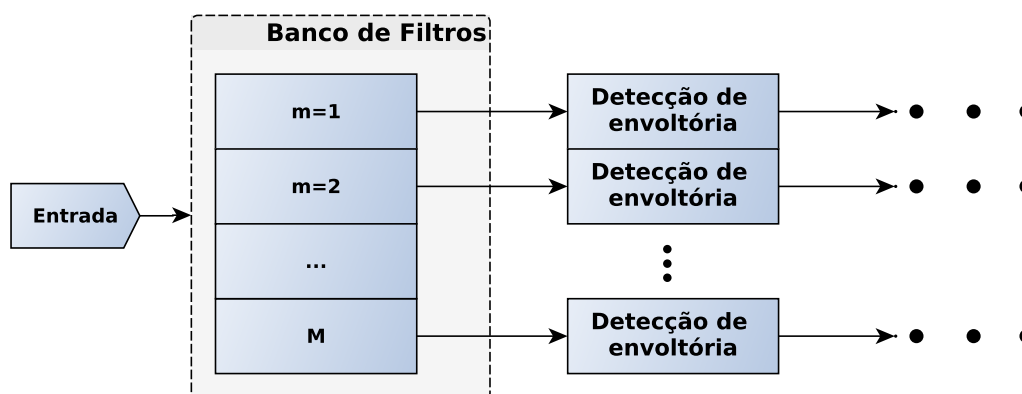
O primeiro passo do processamento de ICs baseados na estratégia CIS, como visto no capítulo 1, é a decomposição do sinal em diversas faixas de frequência. Isso pode ser realizado através da transformada de Fourier de tempo curto (STFT), como ilustrado na Figura 14a, ou utilizando-se um banco de filtros, Figura 14b, (LOIZOU, 2013; WOUTERS; MCDERMOTT; FRANCA, 2015).

Figura 14 – Diagramas de blocos da decomposição do sinal de entrada em múltiplos canais no processamento de estratégias de estimulação baseadas na CIS. Em (a), o sistema com utilização da transformada de Fourier (UTD, 2010; WOUTERS; MCDERMOTT; FRANCA, 2015). Em (b), a abordagem com a utilização de um banco de filtros (KOKKINAKIS *et al.*, 2012; ZENG; POPPER; FAY, 2004).

(a) Sistema usando a Transformada de Fourier



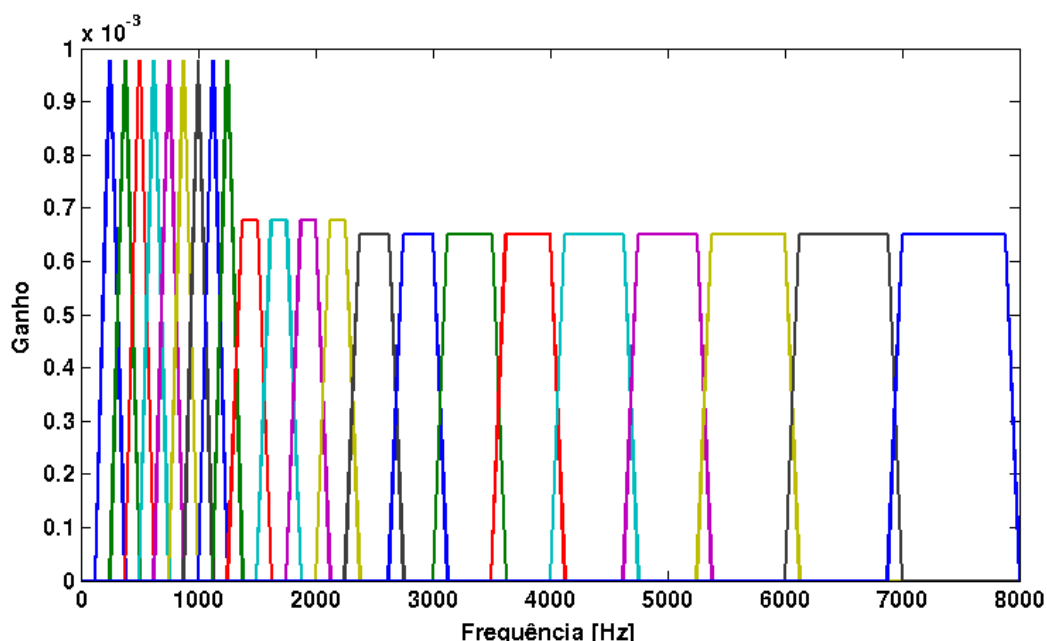
(b) Sistema usando Banco de Filtros



No primeiro caso, para cada quadro (*frame*) temporal, faz-se a transformada

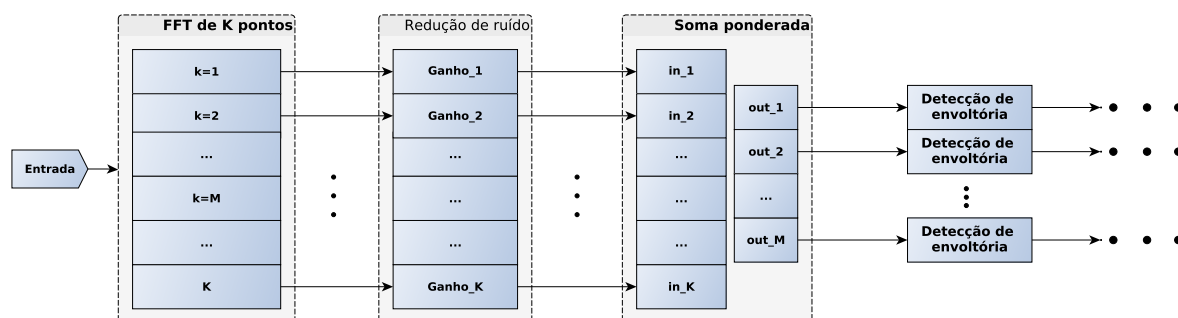
discreta de Fourier (DFT) de K pontos do sinal de entrada, e em seguida, o vetor de magnitudes espectrais é multiplicado por uma matriz $K \times M$ de pesos, em que M é o número de eletrodos, e o valor na saída é uma estimativa do envelope do sinal na respectiva faixa de frequência. Os pesos dessa matriz são mostrados pela Figura 15.

Figura 15 – Pesos aplicados à cada bin de frequência de forma a se calcular o envelope associado a cada canal. Adaptado de (UTD, 2010)



As técnicas de redução de ruído vistas neste capítulo, excetuando-se a proposta por Hu e Loizou (2008), são desenvolvidas para o processamento no domínio da frequência. Sua inserção no diagrama de blocos que descreve a cadeia de processamento de um implante coclear que usa a transformada de Fourier é mostrada na Figura 16.

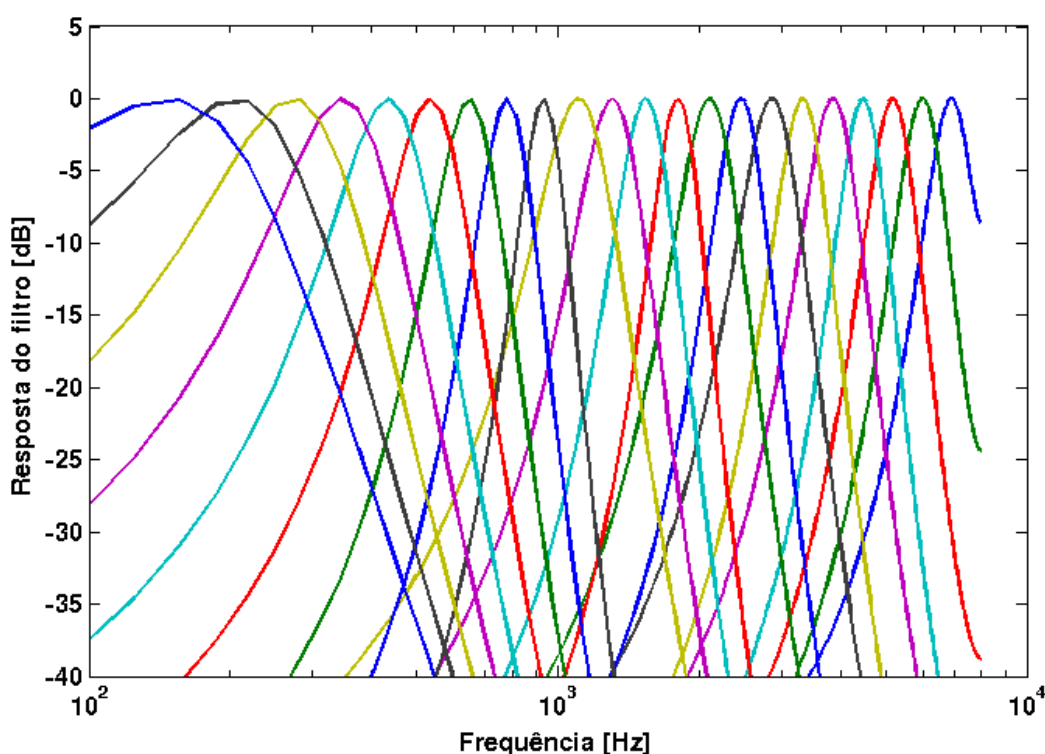
Figura 16 – Diagrama de blocos do processamento com redução de ruído no domínio da frequência.



Já no segundo método, o sinal passa por um banco com M filtros passa-banda e a envoltória em cada saída é detectada através de retificação e filtragem passa-baixas, ou da transformada de Hilbert. O banco de filtros é geralmente do tipo *gammatone*.

Esses filtros, frequentemente usados em modelagem da audição, foram introduzidos por Johannesma (1972), como um modelo da resposta do núcleo da cóclea. Sua resposta ao impulso é obtida pela modulação de um tom com envelope no formato da distribuição Gamma (COSENTINO *et al.*, 2014). A resposta em frequência de um banco de filtros *gammatone* de 22 canais pode ser vista na Figura 17.

Figura 17 – Resposta em frequência do banco de filtros tipo *gammatone* de 22 canais, baseado em Slaney (1993)



Para os sistemas que realizam o processamento como o mostrado na Figura 14b, uma filtragem no domínio do tempo seria mais apropriada. Esse caso será explorado no capítulo 4.

2.5 CONSIDERAÇÕES FINAIS

As principais técnicas de redução de ruído para implantes cocleares consistem de máscaras tempo-frequência, dentre as quais se destacam a máscara binária e o filtro de Wiener com suas variantes (restrito e paramétrico). Ganhos de inteligibilidade e qualidade foram observados, mas não há consenso sobre qual tipo de máscara apresentaria melhores resultados para usuários de implantes cocleares.

Muitas propostas são baseadas no filtro de Wiener paramétrico, pois seus parâmetros permitem uma maior flexibilidade de projeto. Entretanto, para valores arbitrários dos parâmetros, o filtro não está associado a nenhuma condição ótima. Além disso, os

parâmetros não controlam características da curva de supressão independentemente, o que pode dificultar o projeto.

No próximo capítulo uma nova teoria unificada de máscaras tempo-frequência desenvolvida nesta tese é apresentada. Ela engloba diversas máscaras encontradas na literatura, incluindo a máscara binária e o filtro de Wiener, como casos particulares de um mesmo problema de otimização. A máscara resultante possui dois parâmetros que aportam flexibilidade no projeto, assim como o filtro de Wiener paramétrico, mas com a vantagem de que seus parâmetros controlam independentemente o limiar de transição e a “dureza” da máscara. Além disso, a máscara proposta é a solução ótima que minimiza uma função custo envolvendo as distorções relativas à fala e ao ruído.

Outro ponto a ser destacado é que as máscaras utilizadas em ICs encontradas na literatura são formulações para aprimoramento da fala em geral, e não consideram as especificidades da aplicação em seu desenvolvimento. No capítulo 4, é desenvolvido um novo filtro para redução de ruído pensado especificamente para aplicação em ICs. O filtro proposto leva em conta a perda da informação de fase (ou TFS) na cadeia de processamento do IC e minimiza o erro quadrático médio entre o quadrado do envelope do sinal estimado e o quadrado do envelope da fala em cada sub-banda.

3 UMA MÁSCARA TEMPO-FREQUÊNCIA CONFORMÁVEL

No capítulo 2, foi mostrado que máscaras tempo-frequência têm sido regularmente estudadas para realizar a redução de ruído em implantes cocleares. Nas próximas seções, uma máscara de supressão conformável (CM) para redução de ruído é apresentada. Ela gera toda uma família de regras de supressão, que compreendem a BM e o WF e outras máscaras já estabelecidas, possuindo uma relação intuitiva entre os parâmetros de ajuste e o formato da função de supressão. Para ilustrar sua aplicabilidade e desempenho, um conjunto particular de máscaras tempo-frequência foi avaliado através de medidas objetivas relacionadas à qualidade e inteligibilidade da fala para portadores de implantes cocleares.

3.1 MÁSCARA DE SUPRESSÃO CONFORMÁVEL PROPOSTA

As funções de supressão mais utilizadas para aprimoramento da fala em implantes cocleares são, sem dúvida, a máscara binária e o filtro de Wiener (e suas variações). Trabalhos anteriores investigaram o desempenho dessas máscaras para usuários de implantes cocleares (KONING; MADHU; WOUTERS, 2015; MADHU *et al.*, 2013), mostrando melhora significativa na qualidade e inteligibilidade da fala em comparação com sinal ruidoso não-processado. Entretanto, as abordagens são consideravelmente diferentes em razão do formato suave/duro das funções de supressão. A seguir, é proposta uma nova função custo conformável para a derivação de máscaras de supressão que guarda uma relação intuitiva entre seu formato e seus parâmetros de projeto.

Consideremos um sinal ruidoso, $y[n]$, composto pelo sinal de fala, $x[n]$, e ruído aditivo, $v[n]$. Aplicando a STFT, os sinais são representados no domínio tempo-frequência por $Y(k, \lambda) = X(k, \lambda) + V(k, \lambda)$, em que k e λ são, respectivamente, os índices da banda de frequência e da janela de tempo. Assumindo-se um estimador linear do sinal de fala, $\hat{X}(k, \lambda) = H(k, \lambda)Y(k, \lambda)$, o erro de estimação é dado por:

$$\begin{aligned}\varepsilon(k, \lambda) &= \hat{X}(k, \lambda) - X(k, \lambda) \\ \varepsilon(k, \lambda) &= H(k, \lambda)Y(k, \lambda) - X(k, \lambda) \\ \varepsilon(k, \lambda) &= H(k, \lambda)(X(k, \lambda) + V(k, \lambda)) - X(k, \lambda) \\ \varepsilon(k, \lambda) &= \underbrace{(H(k, \lambda) - 1)X(k, \lambda)}_{\varepsilon_X(k, \lambda)} + \underbrace{H(k, \lambda)V(k, \lambda)}_{\varepsilon_V(k, \lambda)},\end{aligned}\tag{12}$$

em que $\varepsilon_X(k, \lambda)$ é a distorção relativa à fala, $\varepsilon_V(k, \lambda)$ é a distorção relativa ao ruído, no domínio da frequência (LOIZOU, 2013), e $H(k, \lambda)$ é um ganho complexo. Assumindo-se

que tanto a fala, quanto o ruído sejam sinais estacionários, em uma dada janela temporal, as densidades espectrais de potência de ambos $\varepsilon_X(k, \lambda)$ e $\varepsilon_V(k, \lambda)$ são definidas de acordo com

$$d_X(H) = E \left\{ |\varepsilon_X(k, \lambda)|^2 \right\} = |H(k, \lambda) - 1|^2 P_X(k, \lambda) \quad (13)$$

$$d_V(H) = E \left\{ |\varepsilon_V(k, \lambda)|^2 \right\} = |H(k, \lambda)|^2 P_V(k, \lambda). \quad (14)$$

Dadas estas definições, a função custo conformável proposta é definida como

$$J(H) = (d_X)^\delta + \rho (d_V)^\delta, \quad (15)$$

em que $\rho \in \mathbb{R}_{>0}$ é um fator de peso relativo entre as parcelas associadas a d_X e d_V , e $\delta \in \mathbb{R}_{>0}$ é um parâmetro relacionado à potência dos espectros de x e v .

Como $d_X \propto |X(k, \lambda)|^2$ e $d_V \propto |V(k, \lambda)|^2$, a função custo proposta pode ser vista, no contexto de processos α -estáveis, como uma soma ponderada dos α -espectrogramas¹ da fala e do ruído, com $\alpha = 2\delta$. É comum assumir que a soma dos α -espectrogramas de fontes independentes é igual ao α -espectrograma da mistura, pelo menos para $\alpha \in (0, 2]$ (LIUTKUS; BADEAU, 2015). Em Liutkus e Badeau (2015), é mostrado que a utilização de α -espectrogramas na modelização e separação de fontes é um procedimento ótimo se as fontes não forem entendidas como processos estacionários no sentido amplo (*wide sense stationary* – WSS), mas sim como processos α -estáveis harmonizáveis² localmente estacionários, uma classe de processos α -estáveis (NIKIAS; SHAO, 1995 apud LIUTKUS; BADEAU, 2015). Estudos mostraram que tais distribuições são frequentemente modelos melhores para sinais de áudio quando comparadas com distribuições gaussianas, devido à sua capacidade de lidar com amostras com grande desvio da média (LIUTKUS; BADEAU, 2015).

Vale notar que as funções $(d_X(H))^{\frac{1}{2}}$ e $(d_V(H))^{\frac{1}{2}}$ são normas e toda norma é uma função convexa (BOYD; VANDENBERGHE, 2004). Além disso, se uma função $g(x)$ é convexa, então $g^p(x)$ também é convexa se $p \geq 1$ (BOYD; VANDENBERGHE, 2004). Desta forma, tanto $\left[(d_X(H))^{\frac{1}{2}} \right]^{2\delta}$, quanto $\left[(d_V(H))^{\frac{1}{2}} \right]^{2\delta}$ são funções convexas para $\delta \geq 1/2$.

Ademais, a soma ponderada, desde que com coeficientes positivos, de funções convexas não altera sua convexidade (BOYD; VANDENBERGHE, 2004). Por conseguinte, a função custo $J(H)$ apresentada na equação (15) é convexa. A convexidade

¹ Espectrogramas são a representação do módulo ou da potência (módulo ao quadrado) do espectro a cada quadro temporal. α -espectrogramas são uma generalização na qual a intensidade em cada unidade tempo-frequência é o módulo do espectro elevado a uma potência $\alpha \in (0, 2]$.

² Um processo α -estável cujo espectro também é estacionário e α -estável é dito α -estável harmonizável (LIUTKUS; BADEAU, 2015).

é uma propriedade desejável, pois uma função convexa possui apenas um ponto de mínimo, o mínimo global (BOYD; VANDENBERGHE, 2004).

Segundo Messerschmitt (2006), os pontos estacionários (mínimos e máximos) de uma função real $f(z)$, diferenciável, de uma variável complexa z são os que respeitam a seguinte condição:

$$\frac{\partial f(z)}{\partial z^*} = 0, \quad (16)$$

em que z^* representa o complexo conjugado de z .

A função $J(H)$ (equação (15)) não é diferenciável em todo o domínio, especificamente para $\delta \in \left(0, \frac{1}{2}\right]$. Desta forma, considerando $\delta \in \left(\frac{1}{2}, \infty\right)$, o valor de H relativo ao mínimo global de $J(H)$ (equação (15)) está entre as soluções da equação a seguir. O índice λ será omitido daqui em diante para que a notação fique mais enxuta.

$$\frac{\partial J}{\partial H^*(k)} = 0 \quad (17)$$

$$\frac{\partial}{\partial H^*(k)} \left\{ |(H(k)-1)|^{2\delta} P_X(k)^\delta + |H(k)|^{2\delta} \rho P_V(k)^\delta \right\} = 0 \quad (18)$$

$$\underbrace{\frac{\partial}{\partial H^*(k)} \left\{ |(H(k)-1)|^{2\delta} \right\} P_X(k)^\delta}_I + \underbrace{\frac{\partial}{\partial H^*(k)} \left\{ |H(k)|^{2\delta} \right\} \rho P_V(k)^\delta}_{II} = 0 \quad (19)$$

Com o objetivo de deixar mais claro o desenvolvimento de (19), as expressões I e II são desenvolvidas separadamente. São utilizadas as derivadas de Wirtinger (KREUTZ-DELGADO, 2009), e a regra da cadeia (KREYSZIG, 2011).

A derivada descrita pela expressão I é dada por:

$$\frac{\partial}{\partial H^*(k)} |(H(k)-1)|^{2\delta} = \frac{\partial}{\partial H^*(k)} [(H(k)-1)(H(k)-1)^*]^\delta \quad (20)$$

$$= (H(k)-1)^\delta \delta ((H(k)-1))^{\delta-1} \quad (21)$$

$$= \delta \frac{[(H(k)-1)(H^*(k)-1)]^\delta}{(H^*(k)-1)}. \quad (22)$$

A derivada descrita pela expressão II é dada por:

$$\frac{\partial}{\partial H^*(k)} |H(k)|^{2\delta} = \frac{\partial}{\partial H^*(k)} [H(k)^\delta H^*(k)^\delta] \quad (23)$$

$$= (H(k))^\delta \delta H^*(k)^{\delta-1} \quad (24)$$

$$= \delta \frac{[H(k)H^*(k)]^\delta}{H^*(k)}. \quad (25)$$

Substituindo as equações (22) e (25) na equação (19), obtém-se:

$$\delta \frac{[(H(k)-1)(H^*(k)-1)]^\delta}{(H^*(k)-1)} P_X(k)^\delta + \delta \frac{[H(k)H^*(k)]^\delta}{H^*(k)} \rho P_V(k)^\delta = 0 \quad (26)$$

$$\delta \frac{[(H(k)-1)(H^*(k)-1)]^\delta}{(H^*(k)-1)} P_X(k)^\delta = -\delta \frac{[H(k)H^*(k)]^\delta}{H^*(k)} \rho P_V(k)^\delta \quad (27)$$

$$\frac{[(H(k)-1)(H^*(k)-1)]^\delta}{(H^*(k)-1)} \frac{H^*(k)}{[H(k)H^*(k)]^\delta} = -\rho \frac{P_V(k)^\delta}{P_X(k)^\delta} \quad (28)$$

$$\left[\frac{H(k)-1}{H(k)} \right]^\delta \left[\frac{H^*(k)-1}{H^*(k)} \right]^{\delta-1} = -\rho \left(\frac{P_V(k)}{P_X(k)} \right)^\delta \quad (29)$$

$$\left[1 - \frac{1}{H(k)} \right]^\delta \left[1 - \frac{1}{H^*(k)} \right]^{\delta-1} = -\rho \left(\frac{P_V(k)}{P_X(k)} \right)^\delta. \quad (30)$$

Fazendo uma substituição de variável tal que

$$Z := 1 - \frac{1}{H(k)}, \quad (31)$$

e considerando $\rho = \mu^\delta$, $\mu \in \mathbb{R}_{>0}$ tem-se:

$$Z^\delta (Z^*)^{\delta-1} = - \left(\frac{\mu P_V(k)}{P_X(k)} \right)^\delta. \quad (32)$$

Uma potência genérica c de um número complexo z tem valores múltiplos, que podem ser representados usando a definição generalizada de logaritmo complexo. Considerando um número complexo, em notação polar, $z = r e^{j\theta}$, uma potência generalizada de z é dada pela equação abaixo.

$$z^c = e^{c \ln z} \quad (33)$$

Nesta equação, $\ln z$ é o logaritmo generalizado de z , dado por $\ln z = \text{Ln } z + j2\pi n$, com $n \in \mathbb{Z}$. $\text{Ln } z = \text{Ln } |z| + j \arg z = \text{Ln } r + j\theta$ se refere ao valor principal do logaritmo (KREYSZIG, 2011). Assim, podemos reescrever a equação 33 da seguinte maneira:

$$z^c = e^{c(\ln r + j\theta + j2\pi n)} \quad (34)$$

$$z^c = r^c e^{j(c\theta + 2\pi n c)}. \quad (35)$$

Considerando $Z = |Z| e^{j\phi_Z}$, e usando o resultado da equação (35) na equação (32), obtêm-se o seguinte desenvolvimento:

$$|Z|^\delta e^{j(\delta\phi_Z + 2\pi p\delta)} |Z|^{\delta-1} e^{j(-\phi_Z + 2\pi q)(\delta-1)} = - \left(\frac{\mu P_V(k)}{P_X(k)} \right)^\delta \quad p, q \in \mathbb{Z} \quad (36)$$

$$|Z|^{2\delta-1} e^{j(\phi_Z\delta + 2\pi p\delta - \phi_Z\delta + 2\pi q\delta + \phi_Z - 2\pi q)} = e^{j(\pi + 2\pi m)} \left| \frac{\mu P_V(k)}{P_X(k)} \right|^\delta e^{j(2\pi l\delta)} \quad p, q, m, l \in \mathbb{Z} \quad (37)$$

$$|Z|^{2\delta-1} e^{j(\phi_Z + 2\pi(p+q)\delta)} e^{-j2\pi q} = \left| \frac{\mu P_V(k)}{P_X(k)} \right|^\delta e^{j(\pi + 2\pi m + 2\pi l\delta)} \quad p, q, m, l \in \mathbb{Z} \quad (38)$$

$$|Z|^{2\delta-1} e^{j(\phi_Z + 2\pi s\delta)} = \left| \frac{\mu P_V(k)}{P_X(k)} \right|^\delta e^{j(\pi + 2\pi m + 2\pi l\delta)} \quad s = p + q \Rightarrow s \in \mathbb{Z}, m, l \in \mathbb{Z}. \quad (39)$$

Para manter a identidade entre os dois lados da equação (39), as condições de igualdade de módulo e igualdade de fase devem ser respeitadas.

A igualdade de módulo é dada por

$$|Z|^{2\delta-1} = \left| \frac{\mu P_V(k)}{P_X(k)} \right|^\delta. \quad (40)$$

A igualdade de fase, por sua vez, é dada por

$$e^{j(\phi_Z + 2\pi s\delta)} = e^{j(\pi + 2\pi m + 2\pi l\delta)} \quad s, m, l \in \mathbb{Z} \quad (41)$$

$$\phi_Z + 2\pi s\delta = \pi + 2\pi m + 2\pi l\delta \quad s, m, l \in \mathbb{Z}. \quad (42)$$

Comparando-se os dois lados da igualdade, verifica-se que

$$s = l \quad (43)$$

$$\phi_Z = \pi + 2\pi m \quad m \in \mathbb{Z}. \quad (44)$$

Segundo a equação (44), $e^{j\phi_Z} = -1$. Assim, pode-se reescrever a equação (31) como segue.

$$|Z| e^{j\phi z} = 1 - \frac{1}{H(k)} \quad (45)$$

$$|Z| = -\left(1 - \frac{1}{H(k)}\right) \quad (46)$$

$$|Z| = \left(\frac{1}{H(k)} - 1\right) \quad (47)$$

Substituindo a equação (47) em (40), tem-se:

$$\left(\frac{1}{H(k)} - 1\right)^{2\delta-1} = \left|\frac{\mu P_V(k)}{P_X(k)}\right|^\delta \quad (48)$$

$$\frac{1}{H(k)} - 1 = \left|\frac{\mu P_V(k)}{P_X(k)}\right|^{\frac{\delta}{2\delta-1}} \quad (49)$$

$$\frac{1}{H(k)} = 1 + \left|\frac{\mu P_V(k)}{P_X(k)}\right|^{\frac{\delta}{2\delta-1}} \quad (50)$$

$$H(k) = \left(1 + \left|\frac{\mu P_V(k)}{P_X(k)}\right|^{\frac{\delta}{2\delta-1}}\right)^{-1} \quad (51)$$

Definindo a SNR local, $\xi(k) := \frac{P_X(k)}{P_V(k)}$, obtém-se $H(k)$:

$$H(k) = \left(1 + \left|\frac{\mu}{\xi(k)}\right|^{\frac{\delta}{2\delta-1}}\right)^{-1} \quad (52)$$

$$H(k) = \frac{1}{1 + \left|\frac{\mu}{\xi(k)}\right|^{\frac{\delta}{2\delta-1}}} \quad (53)$$

$$H(k) = \frac{|\xi(k)|^{\frac{\delta}{2\delta-1}}}{|\xi(k)|^{\frac{\delta}{2\delta-1}} + |\mu|^{\frac{\delta}{2\delta-1}}} \quad (54)$$

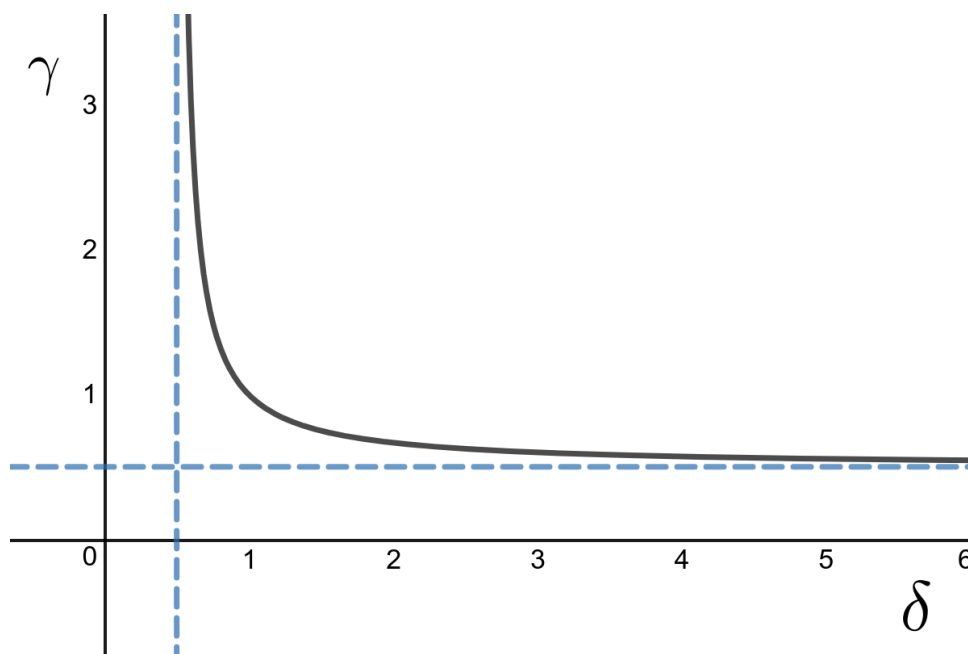
Como $\mu \in \mathbb{R}_{>0}$ e $\xi(k) \in \mathbb{R}_{\geq 0}$, pode-se omitir o operador de valor absoluto, e chega-se à máscara ajustável proposta, descrita pela equação a seguir:

$$H(k) = \frac{\xi^\gamma(k)}{\xi^\gamma(k) + \mu^\gamma}, \quad (55)$$

em que $\mu = \rho^{1/\delta}$ e $\gamma = \delta/(2\delta-1)$. A máscara conformável proposta $H(k)$ é chamada de CM no decorrer deste texto. Deve-se atentar para o fato de que para estrita correspondên-

cia com a equação (15) é necessário que $\mu \in \mathbb{R}_{>0}$, e $\gamma \in (1/2, \infty)$. A Figura 18 mostra a relação entre δ e γ no intervalo de validade de δ .

Figura 18 – Relação entre γ e δ . As linhas pontilhadas em $\delta = 1/2$ e $\gamma = 1/2$ indicam o limite de cada parâmetro quando o outro tende a ∞ .



Pode-se notar que $\delta = 1/2$ e $\gamma = 1/2$ são assíntotas. δ e γ são ambos positivos apenas para valores maiores que $1/2$. Vale lembrar que a equação (55) expressa o valor de $H(k)$ que é um mínimo global para a função custo $J(H)$ (equação (15)), e é válida para qualquer valor de $\delta \in (1/2, \infty)$.

A equação (55) já havia sido sugerida por (FONTAINE *et al.*, 2017), mas como uma proposição heurística. Aqui, foi demonstrado que ela é, de fato, a solução ótima que minimiza a função custo definida pela equação (15).

Simulações numéricas foram realizadas de forma a verificar o comportamento da função custo $J(H)$ (equação (15)) para diversos valores de δ . Nessas simulações, verifica-se que, para os casos em que $\delta \in (0, 1/2]$, a função $J(H)$ (equação (15)) apresenta dois mínimos locais: um em $H = 0$, e o outro em $H = 1$. O mínimo global é dependente de $\xi(k)$, com comportamento idêntico ao de uma máscara binária (equação (5)), sendo o critério de seleção $C(k) = \xi(k)$ e o valor de limiar de transição igual a $T(k) = \mu$.

Isto posto, conjectura-se que a máscara binária é a solução ótima de $J(H)$ (equação (15)) para $\delta \in (0, 1/2]$. A prova para essa conjectura deverá ser encontrada no futuro.

O impacto dos parâmetros μ e γ no formato da máscara $H(k)$, em função de $\xi(k)$ é facilmente verificado na Figura 19, na qual pode-se ver claramente que γ está diretamente relacionado ao quão abrupta é a inclinação da transição da máscara

resultante, enquanto μ controla o ponto de máxima derivada da função de supressão (limiar de transição). Ajustando-se os parâmetros γ and μ , a máscara conformável proposta $H(k)$ define uma família de máscaras sigmoidais que incluem o WF e a BM, como mostra a Tabela 1.

Figura 19 – Efeito da variação dos parâmetros γ e μ em $H(k)$. O parâmetro μ controla o deslocamento lateral, enquanto γ está associado à suavidade de transição. Para $\gamma = 1$ e $\mu = 0\text{dB}$, a máscara é igual ao filtro de Wiener e para $\gamma \geq 100$, ela se comporta praticamente como uma máscara binária.

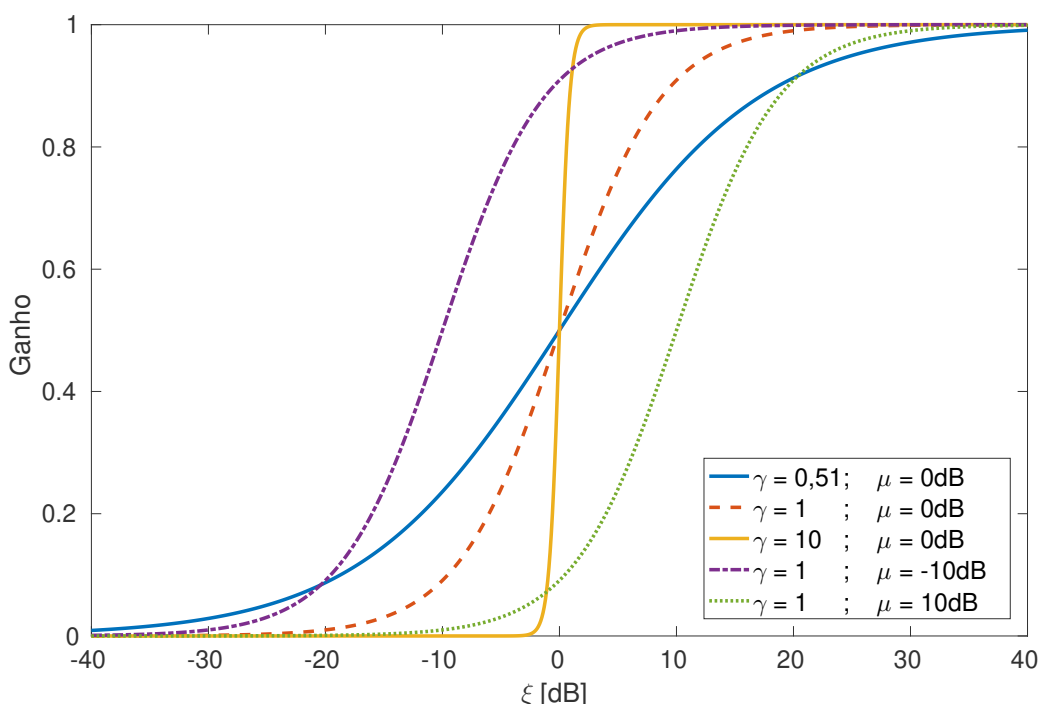


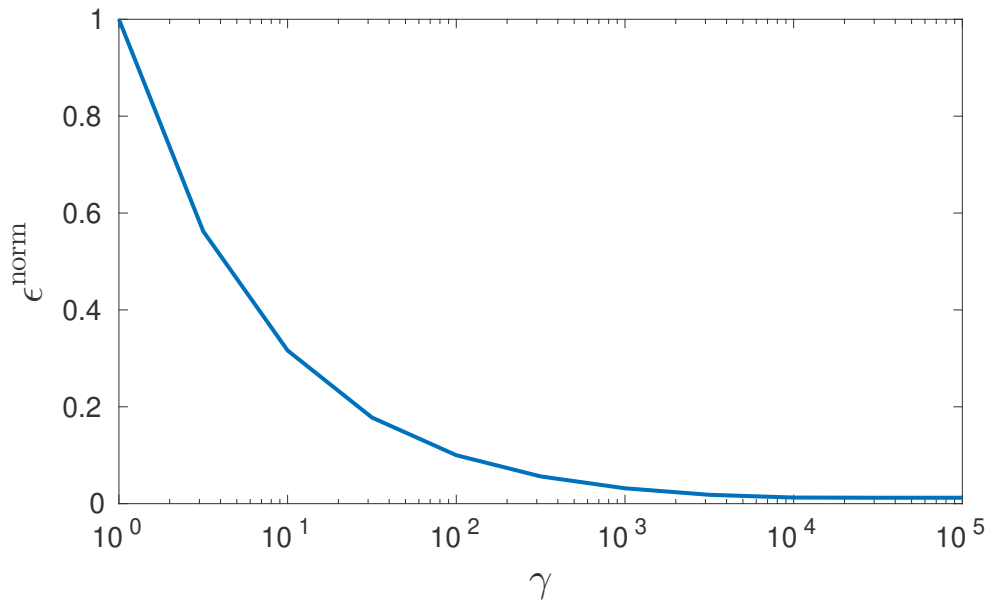
Tabela 1 – Relações entre $H(k)$ e máscaras conhecidas de acordo com os parâmetros γ e μ .

γ	μ	Efeito em $H(k)$
$\gamma \rightarrow \infty$	$\mu = \mu_0$	$H(k) \rightarrow \text{BM, eq. (5)}$
$\gamma = 1$	$\mu = \mu_0$	$H(k) \equiv \text{spectrum over-subtraction}^\dagger$
$\gamma = 1$	$\mu = 0\text{dB}$	$H(k) \equiv \text{WF, eq. (7)}$
$\gamma \rightarrow \frac{1}{2}_+$	$\mu = 0\text{dB}$	$H(k) \rightarrow \text{WF restrito, eq.(10)}$

[†] Hu e Loizou (2003)

Uma forma de se medir a diferença entre duas curvas, $y_1(x)$ e $y_2(x)$, é calcular

Figura 20 – Erro normalizado da máscara conformável proposta em relação à máscara binária em função de γ .



a raiz do erro quadrático (*root square error* – RSE) entre elas, dada por:

$$\epsilon_{y_1, y_2} = \sqrt{\int_{-\infty}^{+\infty} (y_2(x) - y_1(x))^2 dx}. \quad (56)$$

Define-se, então, ϵ^{norm} como sendo o erro entre a máscara conformável derivada neste trabalho $H(\xi)$ e a máscara binária $B(\xi)$, normalizado pelo erro entre o filtro de Wiener $W(\xi)$ e a máscara binária de acordo com a equação a seguir:

$$\epsilon^{\text{norm}} = \frac{\epsilon_{H,B}}{\epsilon_{W,B}}. \quad (57)$$

A Figura 20 mostra a variação de ϵ^{norm} em função do parâmetro γ . Para $\gamma = 100$, o erro da CM em relação à BM é 10% do erro entre o WF e a BM, $\epsilon_{W,B} = 1,295$, valor usado como referência. Quando $\gamma = 1000$, o erro cai para cerca de 3% da referência. Quando $\gamma \rightarrow +\infty$, a transição da CM é a mais abrupta possível, o erro tende a 0 e, conseqüentemente, a CM tende para a BM.

3.2 COMPARAÇÃO COM O FILTRO DE WIENER PARAMÉTRICO

Para se ajustar a máscara conformável (eq. (55)) ao filtro de Wiener paramétrico (WP) da equação (9), a seguinte condição deve ser respeitada:

$$(\eta + \xi)^\beta = \mu^\gamma + \xi^\gamma. \quad (58)$$

Forçando-se $\beta = \gamma$, pode-se escrever

$$\mu = [(\eta + \xi)^\gamma - \xi^\gamma]^\frac{1}{\gamma}. \quad (59)$$

Sendo $\gamma \in (1/2, \infty)$ e $\xi \in \mathbb{R}_{\geq 0}$ e considerando que $\eta \in \mathbb{R}_{>0}$, então μ é real³.

É possível provar que μ é real independentemente dos valores de η , γ e ξ . Por definição, $\eta > 0$, $\gamma > 1/2$ e $\xi \geq 0$. Logo, $(\eta + \xi) > \xi \geq 0$. Além disso, pode-se verificar que, para um expoente $c > 0$, a função $f(x) = x^c$ é constantemente crescente:

$$\frac{df(x)}{dx} = cx^{c-1} > 0, \quad (60)$$

e conseqüentemente $(\eta + \xi)^\gamma \geq \xi^\gamma$. Isso garante que $\mu \in \mathbb{R}_{>0}$, para quaisquer valores de γ e η possíveis. Assim, a máscara proposta pode ser ajustada para ter equivalência ao WP somente se o ajuste de μ for dependente de ξ .

As equações não podem ser igualadas de forma a que todos os parâmetros sejam fixos e independentes de ξ . Contudo, pode-se verificar o quão próximas são através de simulações numéricas. Para isso, um programa computacional foi criado de forma a encontrar os parâmetros β e η do filtro paramétrico de Wiener que minimizem o RSE (equação (56))⁴, utilizando o algoritmo *Nelder-Mead simplex* (LAGARIAS *et al.*, 1998), em relação a uma determinada máscara proposta de parâmetros γ e μ .

Os parâmetros β e η que mais aproximam o filtro de Wiener paramétrico à máscara proposta são mostrados na Tabela 2

Tabela 2 – Pares de parâmetros β e η [dB] que mais aproximam o filtro de Wiener paramétrico à máscara proposta com parâmetros γ e μ .

μ [dB]	γ			
	0,5	1	10	100
-10	0,30/+0,9	1,0/-10,0	1083,0/-42,0	1247,1/-42,6
-5	0,30/+5,9	1,0/-5,0	776,2/-35,5	1262,6/-37,6
0	0,30/+10,9	1,0/0,0	1836,6/-34,3	1294,5/-32,7
+5	0,31/+15,9	1,0/+5,0	1460,1/-28,3	1161,6/-27,2
+10	0,31/+20,9	1,0/+10,0	1350,5/-22,9	1558,9/-23,5

A Tabela 3 apresenta a raiz quadrada do erro quadrático (RSE) obtido após a otimização dos parâmetros β e η . Vale lembrar que o RSE entre a máscara binária e o filtro de Wiener é de 1,295, como referência.

³ Não foi encontrada nenhuma limitação explícita para os parâmetros do filtro paramétrico de Wiener. Entretanto, é sensato considerar $\beta, \eta \in \mathbb{R}_{>0}$ para garantir que a máscara seja real e limitada.

⁴ O RSE foi calculado aproximando a integral na equação (56) através do método trapezoidal para $-60\text{dB} \leq \xi \leq +60\text{dB}$ em passos de 0,001 dB.

Tabela 3 – Raiz quadrada do erro quadrático médio obtido após a otimização dos parâmetros β e η .

μ [dB]	γ			
	0,5	1	10	100
-10	0,290	0,000	0,900	1,073
-5	0,291	0,000	0,900	1,073
0	0,291	0,000	0,899	1,073
5	0,291	0,000	0,899	1,073
10	0,291	0,000	0,899	1,073

As máscaras se igualam quando $\gamma = 1$ (filtro de Wiener), mas quanto mais esse parâmetro se distancia desse valor, o erro mínimo entre as máscaras é maior. Isso mostra que, apesar de haver uma intersecção entre os possíveis filtros de Wiener paramétricos e os filtros propostos, os universos de máscaras das duas abordagens são substancialmente diferentes, sobretudo quando a CM se comporta como uma máscara dura ($\gamma \gg 1$). Essa afirmação é ilustrada pela Figura 21, que mostra exemplos de funções de supressão da CM sobrepostas às do WP que melhor as aproximam.

Em um segundo momento foi realizado o experimento inverso, no qual se procura pelos parâmetros γ e μ da CM que minimizem o RSE (equação (56)) em relação a um determinado WP definido por β e η . Os resultados dessa segunda simulação são apresentados na Tabela 4, que mostra os parâmetros da CM que melhor aproximam o WP, e na Tabela 5, que traz o RSE resultante.

Tabela 4 – Pares de parâmetros γ e μ [dB] que mais aproximam a CM ao WP com parâmetros β e η .

η [dB]	β			
	0,5	1	10	100
-20	0,74/-25,3	1,0/-20,0	1,4/-8,2	1,4/+2,0
-10	0,74/-15,3	1,0/-10,0	1,4/+1,8	1,4/+12,0
0	0,74/-5,3	1,0/0,0	1,4/+11,8	1,4/+22,0
+10	0,74/+4,7	1,0/+10,0	1,4/+21,8	1,4/+32,0
+20	0,74/+14,7	1,0/+20,0	1,4/+31,8	1,4/+42,0

O máximo RSE na Tabela 5 é 0,144, cerca de 13% do maior erro da Tabela 3, ou seja, 1,073. Isso indica que a CM aproxima o WP fixado melhor que o inverso, o

Figura 21 – Sobreposição das funções de supressão da CM proposta e do WP que mais se aproxima à primeira. Em (a), a CM tem $\gamma = 0,5$ e $\mu = 5$ dB. Em (b), a CM tem $\gamma = 100$ e $\mu = -10$ dB.

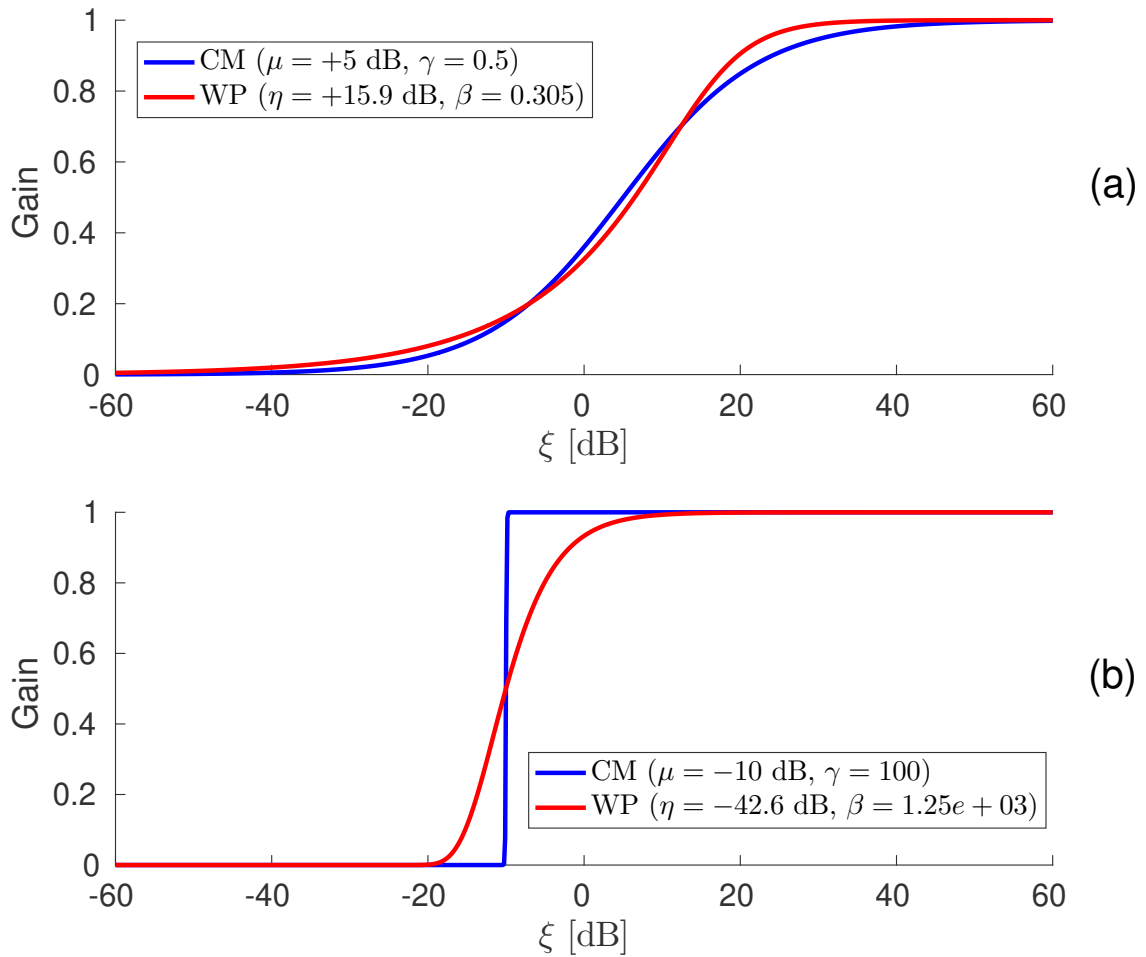
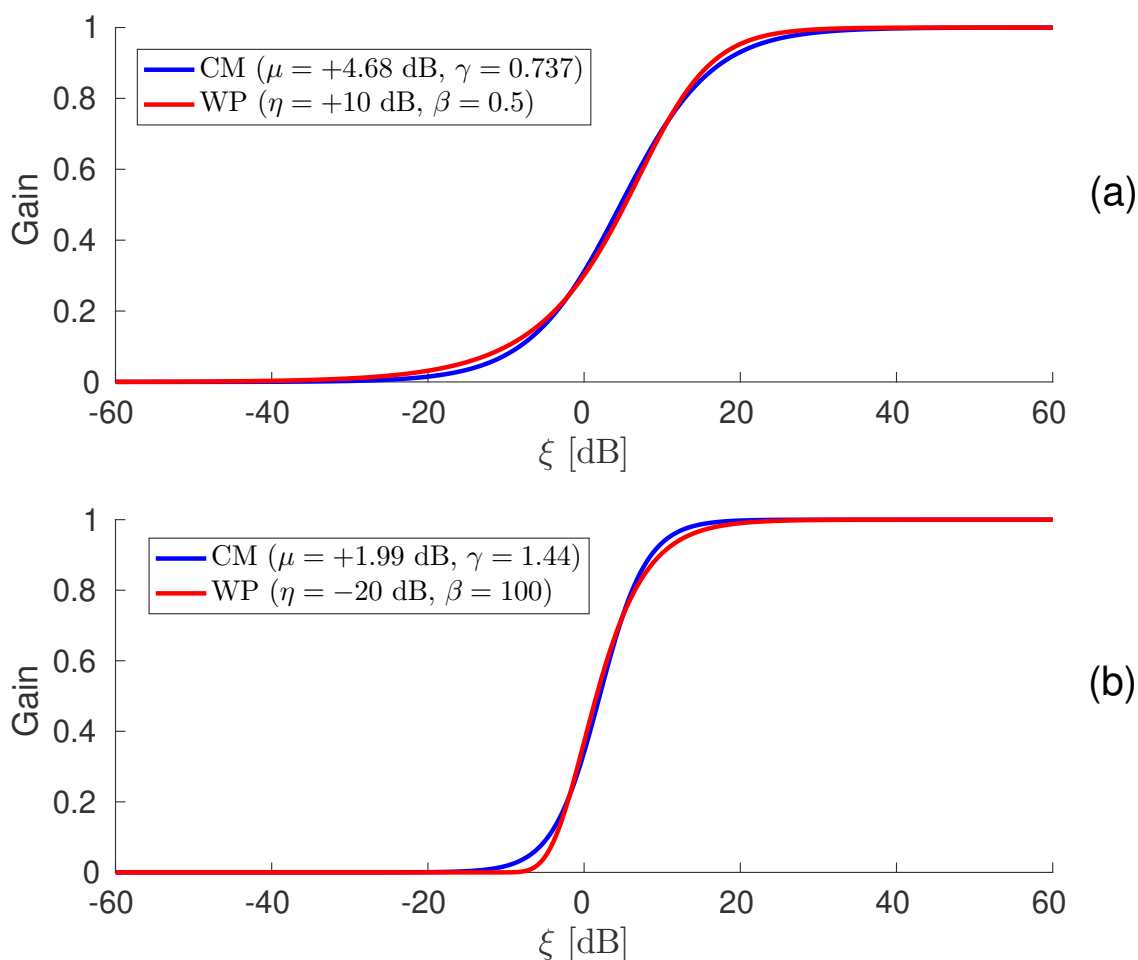


Tabela 5 – Raiz quadrada do erro quadrático médio obtido após a otimização dos parâmetros γ e μ .

η [dB]	β			
	0,5	1	10	100
-20	0,132	0,000	0,131	0,144
-10	0,133	0,000	0,131	0,144
0	0,133	0,000	0,131	0,144
+10	0,133	0,000	0,131	0,144
+20	0,133	0,000	0,131	0,143

que demonstra uma maior flexibilidade da primeira em relação ao último. A Figura 22 ilustra como CMs aproximam WPs diversos.

Figura 22 – Sobreposição das funções de supressão do WP e da CM proposta que mais se aproxima ao primeiro. Em (a), o WP tem $\beta = 0,5$ e $\eta = 10$ dB. Em (b), o WP tem $\beta = 100$ e $\eta = -20$ dB.

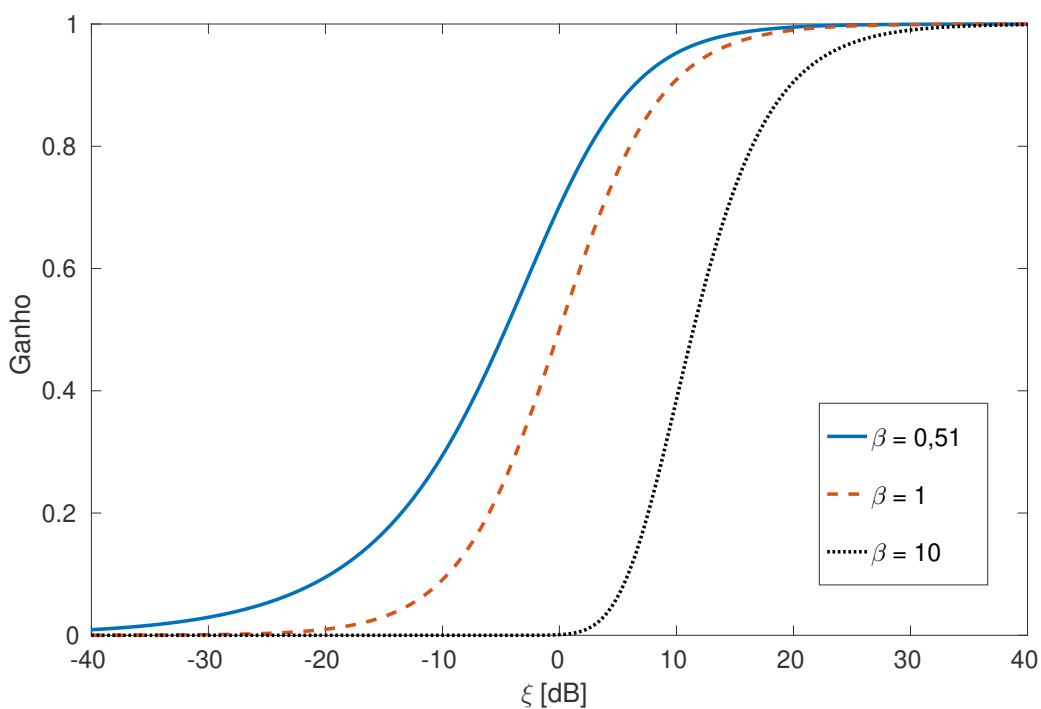


As duas funções de ganho, CM e WP, possuem dois parâmetros que permitem modelar o formato de conformação das máscaras. Uma vantagem da máscara conformável proposta é que seus parâmetros permitem o ajuste independente do deslocamento lateral, associado com o limiar, e a suavidade da transição, enquanto no WP, variações em β levam a mudanças em ambas as características, como mostrado na Figura 23.

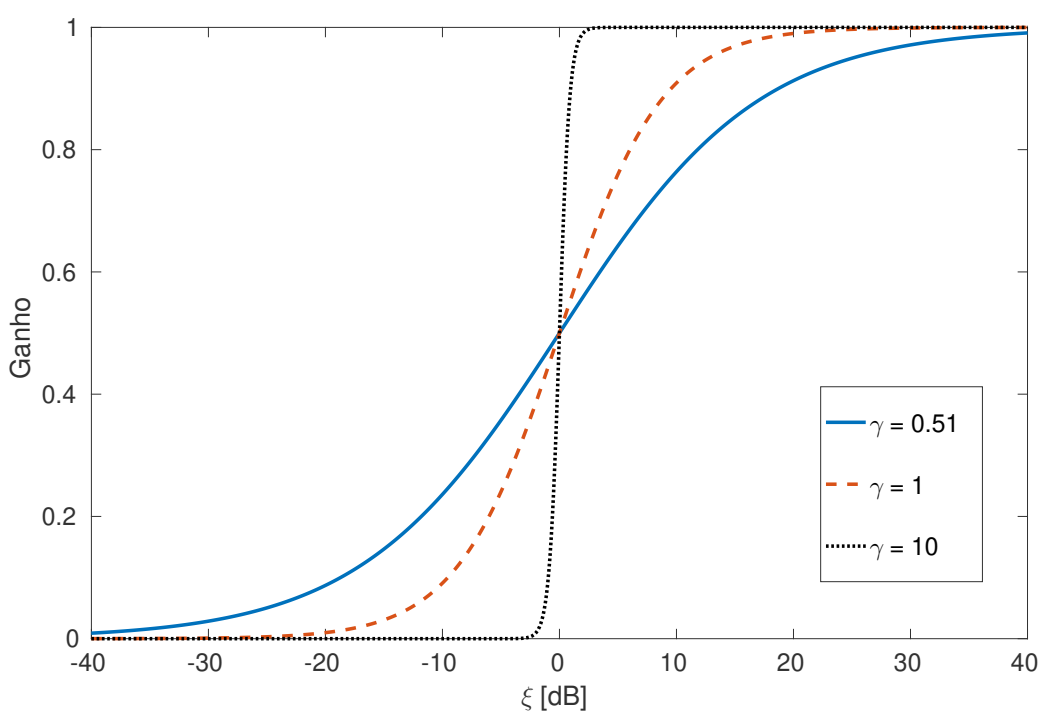
Além disso, o filtro de Wiener parametrizado é uma solução heurística (BENESTY; SONDHI; HUANG, 2008, p.857), contrariamente à máscara conformável proposta que é derivada como solução ótima que minimiza uma função custo.

Figura 23 – Efeito dos parâmetros β no WP (a), e γ na CM proposta (b). Em (a) tanto a inclinação da curva quanto o deslocamento lateral são afetados. Nos gráficos, $\eta = 0\text{dB}$ e $\mu = 0\text{dB}$.

(a) WF Paramétrico



(b) CM proposta



3.3 SIMULAÇÕES NUMÉRICAS

Simulações numéricas foram realizadas de forma a avaliar o impacto da máscara conformável proposta em comparação a outras na inteligibilidade e na qualidade da fala.

Foram utilizados cinco tipos de ruído diferentes: um áudio gravado em um refeitório durante a refeição (*cafeteria babble noise*) (LOIZOU, 2013); um áudio gravado dentro de um vagão de trem em movimento (*inside train noise*) (LOIZOU, 2013); ruído de fala-sobre-fala (*speech-in-speech*) usando o sinal ISTS (HOLUBE *et al.*, 2010); ruído com espectro de fala (*speech-shaped noise*) usando o sinal ICRA (DRESCHLER *et al.*, 2001); e ruído branco Gaussiano. Os espectrogramas dos ruídos são mostrados na Figura 24.

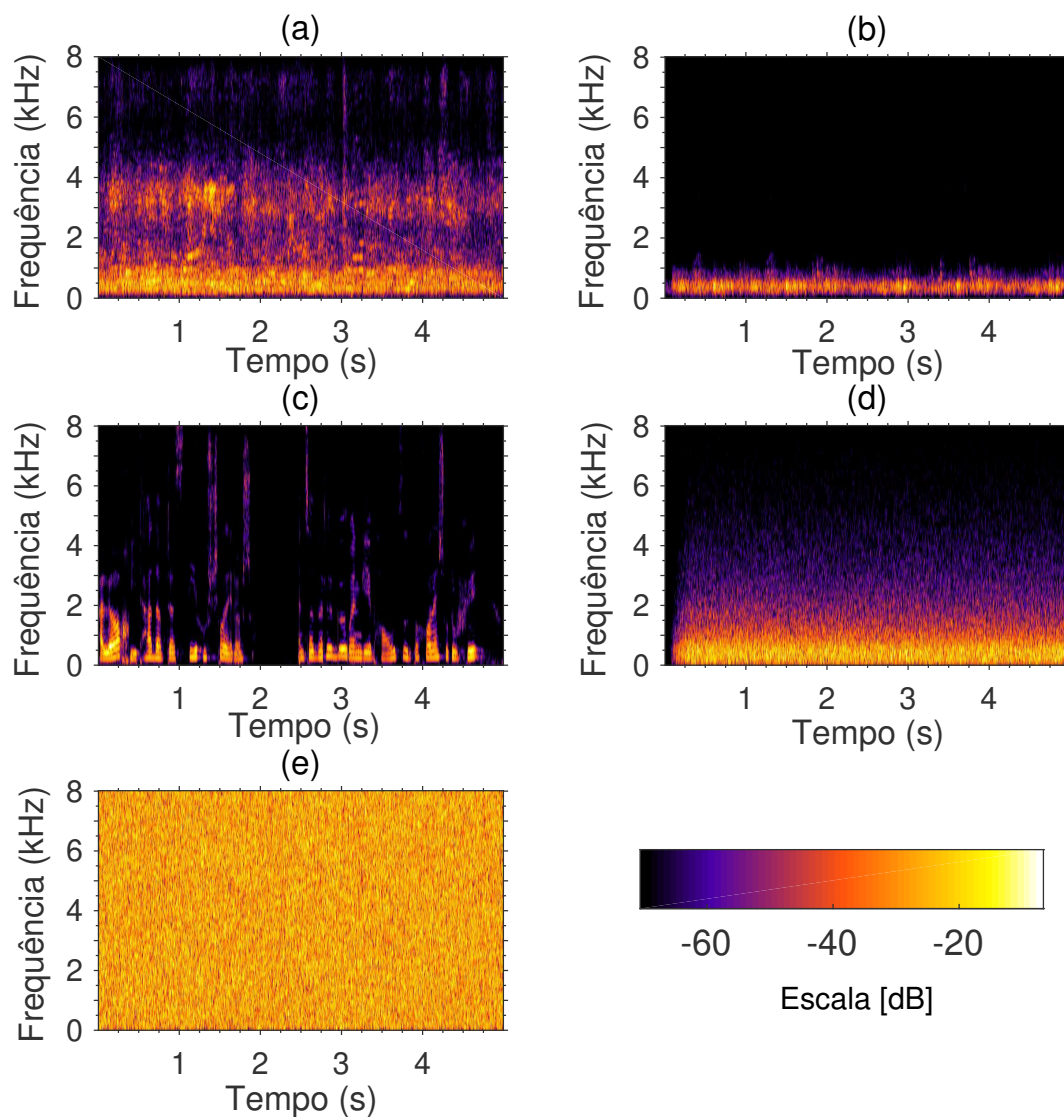
As 720 frases balanceadas do *IEEE corpus* (IEEE, 1969) foram contaminadas com ruído aditivo em 3 valores diferentes de SNR -10 dB, 0 dB, e $+5$ dB, no intuito de representar, respectivamente, uma situação de predominância de ruído, igualdade de potência entre fala e ruído, e ligeira predominância da fala. Isso resultou em um total de 10800 sinais ruidosos. Desses, 1500 foram usados para determinação dos parâmetros da CM e do WP, e os 9300 restantes foram utilizados para comparar as máscaras proposta (CM), binária (BM) (com $\mu_0 = 0$ dB), de Wiener (WF) e parametrizada de Wiener (WP), além do sinal não processado.

Os sinais ruidosos foram divididos em dois grupos: um conjunto de treinamento e um conjunto de teste. O conjunto de treinamento consistia de 15 subgrupos, um para cada tipo de ruído e SNR estudados, contendo 100 áudios cada um. Por sua vez, o conjunto de teste também conteve 15 subgrupos com 620 áudios cada.

Todos os sinais foram amostrados a 16 kHz e foram transformados para o domínio de tempo-frequência usando a STFT de 512 pontos, a janela de Hanning de 20 ms e sobreposição de 50%. Após o processamento por uma das máscaras, os sinais foram retransformados para o domínio do tempo através do método de sobreposição-e-soma (*overlap-and-add*) (CROCHIERE, 1980). Os sinais limpo e de ruído foram processados paralelamente de forma a calcular os ganhos das máscaras ideais.

O desempenho das máscaras tempo-frequência foi avaliado em termos de qualidade usando a métrica *wideband Perceptual Evaluation of Speech Quality* (PESQ) (INTERNATIONAL TELECOMMUNICATION UNION, 2007). Para a avaliação em termos de inteligibilidade utilizou-se primeiramente a métrica *Normalized Covariance Metric* (NCM) (HOLUBE; KOLLMEIER, 1996). De forma a avaliar a utilização da CM em implantes cocleares, em um segundo momento, foi utilizada a métrica *Speech-to-Reverberation Modulation Energy Ratio – Cochlear Implants* (SRMR-CI) (SANTOS; FALK, 2014) para se aferir a inteligibilidade. A SRMR-CI foi especificamente desenvolvida para apresentar alta correlação com a inteligibilidade de portadores de implantes cocleares em experimentos subjetivos. A SRMR-CI pode ser mapeada em valores de

Figura 24 – Espectrogramas dos ruídos usados nos experimentos. (a) *cafeteria babble noise* (LOIZOU, 2013); (b) *inside train noise* (LOIZOU, 2013); (c) *speech-in-speech* ISTS (HOLUBE *et al.*, 2010); (d) *speech-shaped noise* ICRA (DRESCHLER *et al.*, 2001); (e) ruído branco Gaussiano. A escala de cores é em dB.



inteligibilidade da seguinte maneira (SANTOS; FALK, 2014):

$$I_{\%} = \frac{1}{1 + e^{-(a_1 \text{SRMR-CI}_{norm} + a_2)}} \times 100\%, \quad (61)$$

em que $I_{\%}$ é o percentual de inteligibilidade; $a_1 = 12,1742$ e $a_2 = -7,4535$ são coeficientes de ajuste (SANTOS, 2013) e

$$\text{SRMR-CI}_{norm} = \frac{\text{SRMR-CI}}{\text{SRMR-CI}_{clean}}. \quad (62)$$

Por sua vez, SRMR-CI_{clean} é o valor da métrica calculada para o sinal sob teste e SRMR-CI_{clean} é o valor da métrica para o sinal de fala limpo.

O apêndice B traz uma definição mais detalhada das métricas utilizadas.

Os experimentos utilizando a SRMR-CI foram feitos apenas com os ruídos *cafeteria babble* e *inside train*. Eles apresentam situações realistas, respectivamente, de um ruído com modulação e de um ruído estacionário.

Valores de inteligibilidade e qualidade referentes ao processamento com a CM, BM, WF e WP foram calculados e comparados entre si usando-se análise de variância (ANOVA), com nível de significância especificado em 0,05. A seguir foi realizada uma análise de comparação múltipla utilizando-se o teste de *Tukey*, um teste *post hoc* adequado ao cenário deste estudo (MCHUGH, 2011), que permite a comparação pareada.

Para cada condição de SNR e tipo de ruído, o conjunto de treinamento foi utilizado para a obtenção dos melhores parâmetros (γ e μ para a CM, e β e η para o WP), de acordo com um critério objetivo. Três critérios foram utilizados: maximização da métrica de qualidade (PESQ); maximização da métrica de inteligibilidade (NCM ou SRMR-CI); e maximização de uma combinação das métricas de qualidade e inteligibilidade. Inicialmente, foram obtidos os melhores conjuntos de parâmetros para cada um dos sinais ruidosos, dentro de uma grade de possibilidades. Para a CM, a grade de valores de parâmetros continha todas as combinações de μ , entre -60 dB e 60 dB com passo de 5 dB, e γ variando de $0,5$ a 1 em 6 passos e de $1,25$ a 100 em 6 passos igualmente espaçados em escala logarítmica, resultando em 300 combinações diferentes. A grade de parâmetros inclui o WF ($\gamma = \mu = 1$) e uma aproximação da BM ($\gamma = 100$, $\mu = 1$). O mesmo procedimento foi realizado com o WP, com β variando de $0,2$ a 1 e de $1,25$ a 400 em 4 e 6 , respectivamente, passos igualmente espaçados em escala logarítmica, e η entre -35 dB e 25 dB em passos de $2,5$ dB, totalizando 250 máscaras.

Após a obtenção dos 1500 melhores conjuntos de parâmetros, foi calculada a mediana dos 100 resultados referentes a cada subgrupo. Esse conjunto de medianas foi mapeado para os valores mais próximos nas respectivas grades. Os parâmetros resultantes desse mapeamento, para cada SNR e tipo de ruído, foram utilizados na avaliação do conjunto de áudios de teste. A seguir são apresentadas três variações do

experimento: I – maximização de qualidade, II – maximização de inteligibilidade e III – maximização conjunta de qualidade e inteligibilidade).

3.3.1 Experimento I – Maximização de qualidade

No experimento I, o método descrito acima foi utilizado na busca dos parâmetros da CM e do WP que maximizam a métrica de qualidade, PESQ, independentemente do impacto na métrica de inteligibilidade (NCM ou SRMR-CI). Os melhores conjuntos de parâmetros obtidos para cada tipo de ruído e SNR são mostrados na Tabela 6.

Tabela 6 – Parâmetros da CM e WP obtidos com o conjunto de treinamento, buscando maximização da métrica de qualidade PESQ.

Noise	SNR	CM		WP	
		γ	μ [dB]	β	η [dB]
Cafeteria Babble	-10 dB	0,66	10,0	0,72	7,5
	0 dB	0,66	5,0	0,53	5,0
	+5 dB	0,66	5,0	0,53	5,0
Inside Train	-10 dB	0,57	25,0	0,53	15,0
	0 dB	0,50	20,0	0,53	12,5
	+5 dB	0,57	15,0	0,53	10,0
ISTS	-10 dB	0,57	15,0	0,53	12,5
	0 dB	0,57	15,0	0,53	10,0
	+5 dB	0,57	15,0	0,53	10,0
White	-10 dB	0,66	5,0	0,53	7,5
	0 dB	0,66	5,0	0,53	5,0
	+5 dB	0,66	0,0	0,53	5,0
ICRA	-10 dB	0,66	10,0	0,72	7,5
	0 dB	0,66	15,0	0,53	7,5
	+5 dB	0,57	10,0	0,53	10,0

Os valores médios de PESQ obtidos por cada máscara com o conjunto de áudios de teste são apresentados na Tabela 7.

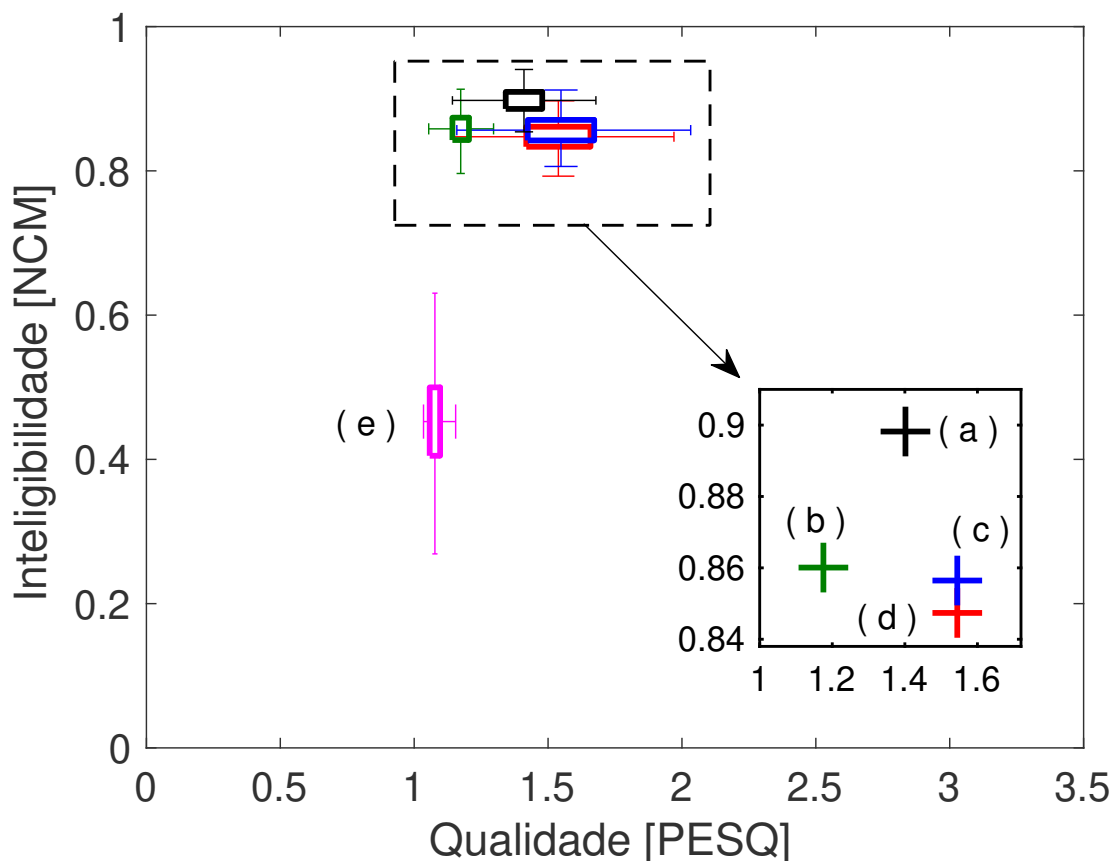
Tabela 7 – Valores médios de PESQ obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com o critério PESQ. Para cada linha, os maiores valores de PESQ estão em negrito, enquanto que o símbolo § indica que os valores não são estatisticamente distintos ($p > 0,05$).

Noise	SNR	WF	BM	CM	WP
Cafeteria Babble	-10 dB	1,364§	1,141	1,411	1,375§
	0 dB	1,998	1,578	2,078§	2,058§
	+5 dB	2,565	2,068	2,650§	2,629§
Inside Train	-10 dB	1,417	1,178	1,560§	1,552§
	0 dB	1,984	1,557	2,140§	2,115§
	+5 dB	2,487	2,041	2,624	2,588
ISTS	-10 dB	1,659	1,376	1,773	1,740
	0 dB	2,366	1,986	2,469	2,446
	+5 dB	2,800	2,430	2,907§	2,880§
White	-10 dB	1,633	1,245	1,740§	1,743§
	0 dB	2,557	1,778	2,814§	2,774§
	+5 dB	3,138	2,284	3,374§	3,346§
ICRA	-10 dB	1,344§	1,127	1,387	1,359§
	0 dB	1,812	1,471	1,903	1,884
	+5 dB	2,246	1,790	2,386§	2,390§

A Figura 25 mostra os diagramas de caixa (*boxplots*) bidimensionais que relacionam o desempenho, com o conjunto de áudios de teste, em qualidade, em unidades PESQ, no eixo horizontal, e inteligibilidade em NCM (a unidade corresponde a 100%), no eixo vertical, para os sinais de fala não-processado, e processado com WF, BM, CM e WP para uma $SNR = -10$ dB. Os parâmetros da CM e WP foram obtidos pela maximização de PESQ. Nesse tipo de apresentação, o desempenho geral aumenta com a distância da origem. Os quadrados indicam o primeiro e terceiro quartis, enquanto que os sinais de cruz (+), na região de destaque, representam as medianas das métricas de qualidade e inteligibilidade. Valores discrepantes (*outliers*) são omitidos por clareza.

O gráfico para a mesma situação, com a métrica SRMR-CI como medida de inteligibilidade para implantados, é mostrado na Figura 26.

Figura 25 – Diagramas de caixa bidimensionais de qualidade (PESQ) e inteligibilidade (NCM) para o conjunto de teste e ruído de *dentro do trem* a $SNR = -10\text{dB}$. Os parâmetros da CM e e WP foram otimizados para qualidade. (a) WF em preto; (b) BM em verde; (c) CM em azul; (d) WP em vermelho; (e) não processado em magenta.



3.3.2 Experimento II – Maximização de inteligibilidade

No segundo experimento, o mesmo procedimento para encontrar os melhores parâmetros foi aplicado à CM e ao WP, mas com o objetivo de maximizar a métrica de inteligibilidade, independentemente dos impactos na métrica de qualidade.

Em um primeiro momento, buscou-se pelo máximo valor da métrica NCM. A Tabela 8 mostra os melhores parâmetros, considerando a NCM, obtidos para cada tipo de ruído e SNR, usando o grupo de treinamento.

A Tabela 9 traz os valores médios de NCM obtidos com o grupo de teste para cada máscara, SNR e tipo de ruído.

Figura 26 – Diagramas de caixa bidimensionais de qualidade (PESQ) e inteligibilidade (SRMR-CI) para o conjunto de teste e ruído de *dentro do trem* a $SNR = -10\text{dB}$. Os parâmetros da CM e e WP foram otimizados para qualidade. (a) WF em preto; (b) BM em verde; (c) CM em azul; (d) WP em vermelho; (e) não processado em magenta.

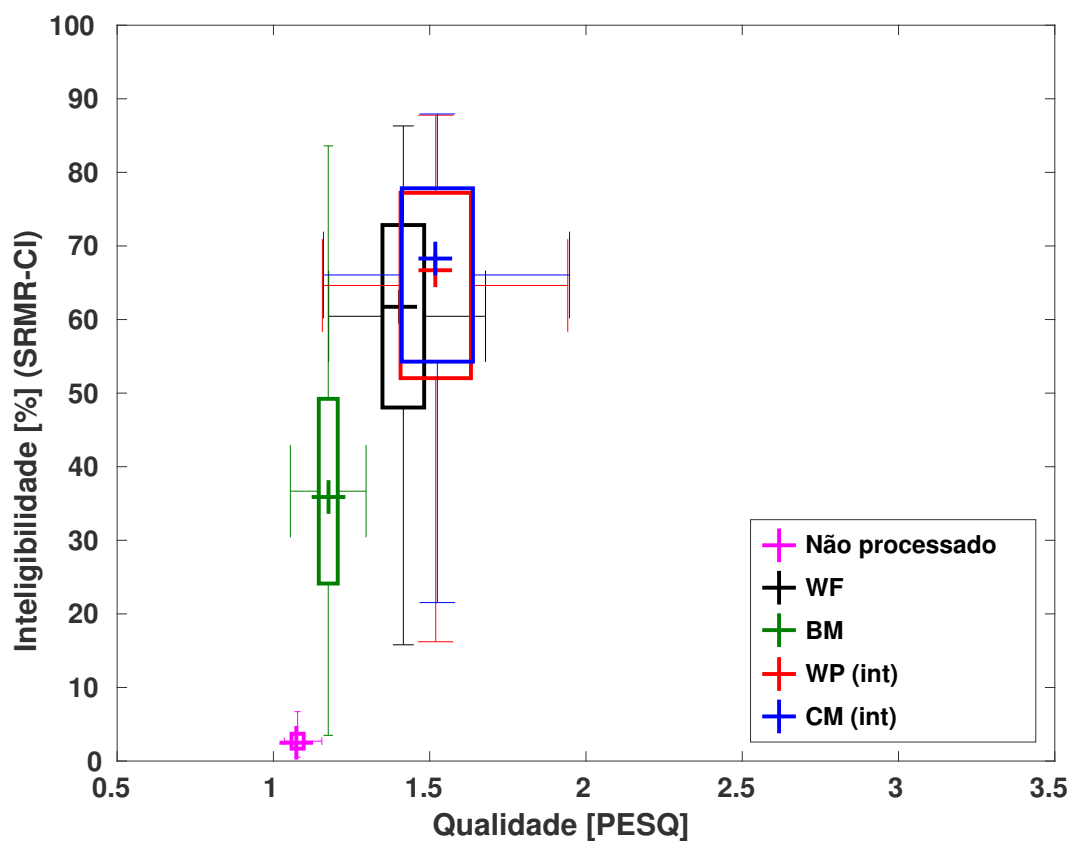


Tabela 8 – Parâmetros da CM e WP obtidos com o conjunto de treinamento, buscando maximização da métrica de qualidade NCM.

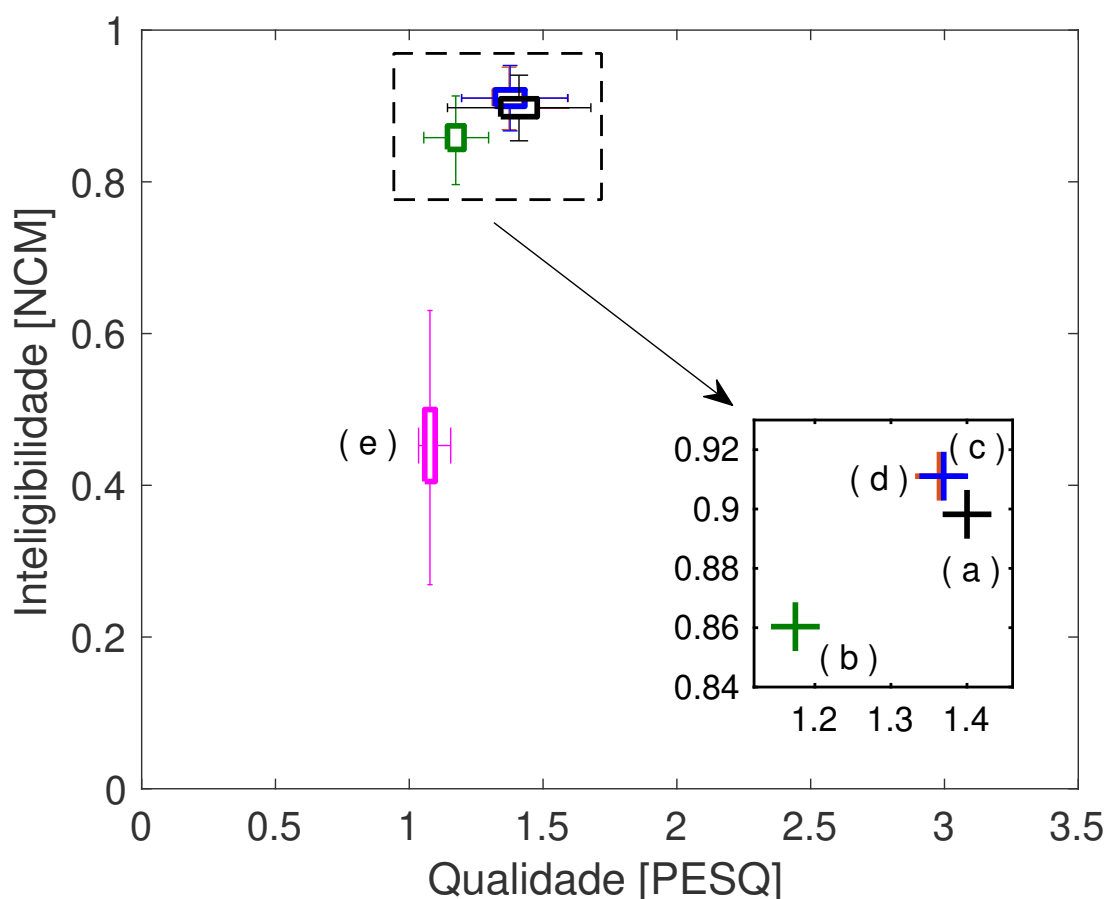
Noise	SNR	CM		WP	
		γ	μ [dB]	β	η [dB]
Cafeteria Babble	-10 dB	0,66	-5,0	0,53	0,0
	0 dB	0,76	-5,0	0,53	0,0
	+5 dB	0,76	-5,0	0,72	-2,5
Inside Train	-10 dB	0,76	-5,0	0,53	2,5
	0 dB	0,87	-10,0	0,72	-2,5
	+5 dB	0,87	-10,0	1,25	-5,0
ISTS	-10 dB	0,76	-5,0	0,72	0,0
	0 dB	0,87	-5,0	1,00	-2,5
	+5 dB	1,00	-10,0	39,81	-20,0
White	-10 dB	0,57	-5,0	0,53	0,0
	0 dB	0,66	-5,0	0,53	0,0
	+5 dB	0,66	-5,0	0,53	0,0
ICRA	-10 dB	0,66	-5,0	0,53	0,0
	0 dB	0,76	-5,0	0,72	-2,5
	+5 dB	0,87	-5,0	1,00	-2,5

Tabela 9 – Valores médios de NCM obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com o critério NCM. Para cada linha, os maiores valores de NCM estão em negrito, enquanto que o símbolo § indica que os valores não são estatisticamente distintos ($p > 0,05$).

Noise	SNR	WF	BM	CM	WP
Cafeteria Babble	-10 dB	0,832	0,733	0,881§	0,879§
	0 dB	0,948	0,908	0,967§	0,966§
	+5 dB	0,984	0,967	0,992§	0,992§
Inside Train	-10 dB	0,897	0,858	0,911§	0,910§
	0 dB	0,963	0,950	0,971§	0,971§
	+5 dB	0,989	0,984	0,992§	0,992§
ISTS	-10 dB	0,908	0,872	0,930§	0,927§
	0 dB	0,973	0,960	0,982§	0,981§
	+5 dB	0,992	0,987	0,994§	0,994§
White	-10 dB	0,857	0,733	0,940§	0,940§
	0 dB	0,960	0,902	0,991§	0,991§
	+5 dB	0,981	0,945	0,996§	0,996§
ICRA	-10 dB	0,851	0,758	0,889§	0,888§
	0 dB	0,958	0,932	0,967§	0,967§
	+5 dB	0,990	0,983	0,993§	0,993§

A Figura 27 mostra os diagramas de caixa (*boxplots*) bidimensionais que relacionam o desempenho, com o conjunto de áudios de teste, em qualidade, em unidades PESQ, no eixo horizontal, e inteligibilidade em NCM (a unidade corresponde a 100%), no eixo vertical, para os sinais de fala não-processado, e processado com WF, BM, CM e WP para uma $SNR = -10$ dB. Os parâmetros da CM e WP foram obtidos pela maximização de NCM.

Figura 27 – Diagramas de caixa bidimensionais de qualidade (PESQ) e inteligibilidade (NCM) para o conjunto de teste e ruído de *dentro do trem* a $SNR = -10$ dB. Os parâmetros da CM e WP foram otimizados pela métrica NCM. (a) WF em preto; (b) BM em verde; (c) CM em azul; (d) WP em vermelho; (e) não processado em magenta.



Quando a métrica SRMR-CI é considerada como medida de inteligibilidade, os melhores parâmetros para a CM e o WP são apresentados na Tabela 10.

A Tabela 11 apresenta os valores médios de SRMR-CI obtidos com o grupo de teste para cada máscara, SNR e tipo de ruído. Os diagramas de caixa referentes a um caso particular dessa tabela, ruído de *dentro do trem* com $SNR = -10$ dB, são mostrados na Figura 28.

Tabela 10 – Parâmetros da CM e WP obtidos com o conjunto de treinamento, buscando maximização da métrica de qualidade SRMR-CI.

Noise	SNR	CM		WP	
		γ	μ [dB]	β	η [dB]
Cafeteria Babble	-10 dB	0,66	0	0,53	2,5
	0 dB	0,76	5	0,72	2,5
	+5 dB	0,87	5	1,25	-2,5
Inside Train	-10 dB	0,57	20	0,53	12,5
	0 dB	0,66	10	0,53	10
	+5 dB	0,66	5	0,72	5

Tabela 11 – Valores médios de inteligibilidade (SRMR-CI) obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com o critério SRMR-CI. Para cada linha, os maiores valores de inteligibilidade estão em negrito, enquanto que o símbolo § indica que os valores não são estatisticamente distintos ($p > 0,05$).

Noise	SNR	WF	BM	CM	WP
Cafeteria Babble	-10 dB	52,55	31,47	56,91§	56,25§
	0 dB	78,39§	71,48	78,90§	78,95§
	+5 dB	84,93§	83,23	84,95§	85,05§
Inside Train	-10 dB	58,34	36,50	65,03§	63,81§
	0 dB	78,22§	69,16	79,75§	79,16§
	+5 dB	84,22§	81,04	84,62§	84,51§

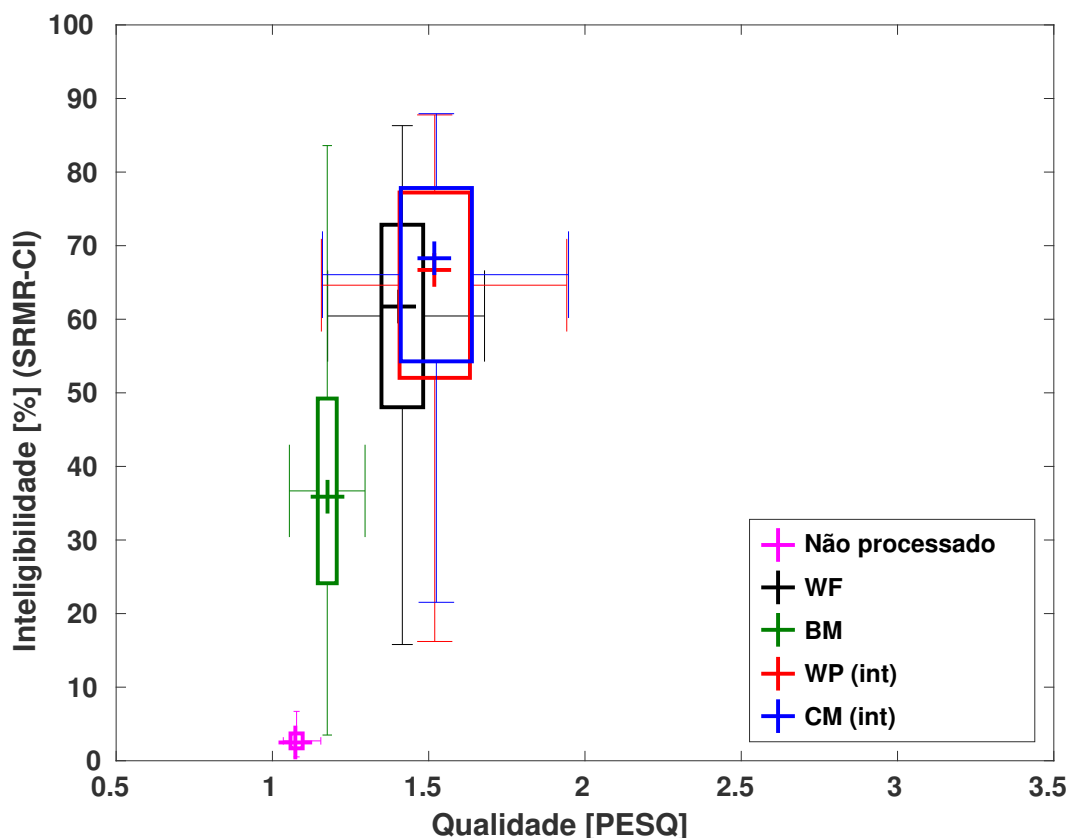
3.3.3 Experimento III – Maximização conjunta de qualidade e inteligibilidade

Diferentemente dos experimentos anteriores, neste, os parâmetros da CM e da WP são otimizados de acordo com a performance geral, considerando-se tanto a métrica de qualidade quanto a de inteligibilidade. A performance geral foi calculada como sendo a distância quadrática, dada por:

$$d = \left(\frac{\text{IM}}{\max(\text{IM})} \right)^2 + \left(\frac{\text{PESQ}}{5} \right)^2, \quad (63)$$

em que IM é a métrica de inteligibilidade, que pode ser NCM ou $I_{\%}$ (equação (61)). Os valores são normalizados, para que as escalas sejam equivalentes. Vale ressaltar que a equação (63) é uma combinação artificial de métricas objetivas com o objetivo

Figura 28 – Diagramas de caixa bidimensionais de qualidade (PESQ) e inteligibilidade (SRMR-CI) para o conjunto de teste e ruído de *dentro do trem* a $SNR = -10$ dB. Os parâmetros da CM e WP foram otimizados de acordo com o critério SRMR-CI. (a) WF em preto; (b) BM em verde; (c) CM em azul; (d) WP em vermelho; (e) não processado em magenta.



de estabelecer uma função custo simples e efetiva para a maximização conjunta de qualidade e inteligibilidade. Ela representa, em um espaço bidimensional, o quadrado da distância Euclidiana em relação à origem. Não há nenhuma evidência da sua correlação com a percepção humana.

Os parâmetros da CM e WP que maximizam a equação (63) com a métrica NCM estão dispostos na Tabela 12.

As Tabelas 13 e 14 mostram, respectivamente, as médias de NCM e PESQ obtidos com esses parâmetros. A Figura 29 traz os diagramas de caixa para o caso de ruído *babble* a $SNR = -10$ dB.

Tabela 12 – Parâmetros da CM e WP obtidos com o conjunto de treinamento, buscando maximização da equação (63), considerando a NCM como métrica de inteligibilidade.

Noise	SNR	CM		WP	
		γ	μ [dB]	β	η [dB]
Cafeteria Babble	-10 dB	0,66	-5,0	0,53	0,0
	0 dB	0,66	-5,0	0,72	0,0
	+5 dB	0,76	0,0	0,72	0,0
Inside Train	-10 dB	0,66	0,0	0,53	0,0
	0 dB	0,66	0,0	0,53	5,0
	+5 dB	0,66	0,0	0,53	5,0
ISTS	-10 dB	0,76	-5,0	0,72	0,0
	0 dB	0,76	0,0	0,72	0,0
	+5 dB	0,76	0,0	0,53	5,0
White	-10 dB	0,57	-5,0	0,53	0,0
	0 dB	0,66	0,0	0,53	5,0
	+5 dB	0,66	0,0	0,53	5,0
ICRA	-10 dB	0,66	-5,0	0,53	0,0
	0 dB	0,76	-5,0	0,72	0,0
	+5 dB	0,66	5,0	0,53	5,0

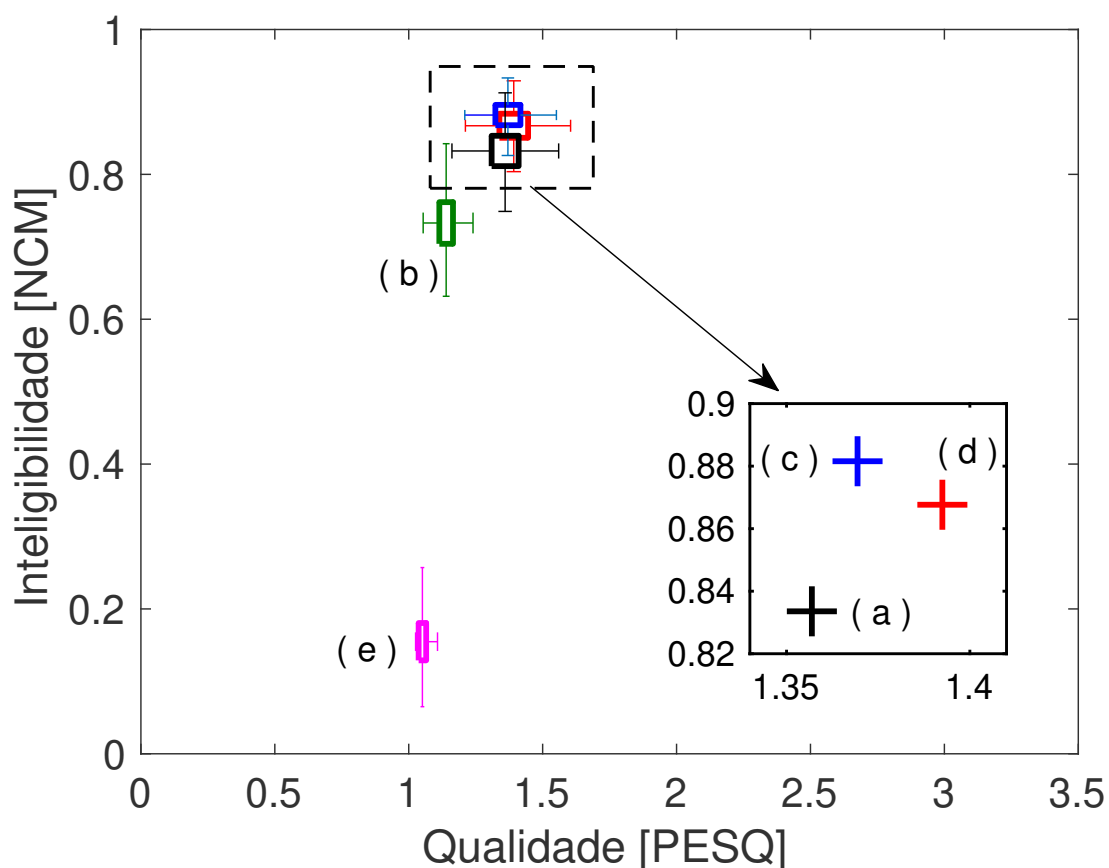
Tabela 13 – Valores médios de NCM obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com a equação (63) com NCM. Para cada linha, os maiores valores de NCM estão em negrito, enquanto que os símbolos § e * indicam que os valores não são estatisticamente distintos ($p > 0,05$).

Noise	SNR	WF	BM	CM	WP
Cafeteria Babble	-10 dB	0,733	0,832	0,881	0,867
	0 dB	0,908	0,948	0,967	0,955
	+5 dB	0,967	0,984	0,989	0,987
Inside Train	-10 dB	0,858	0,897	0,910§	0,908§
	0 dB	0,950	0,963	0,968	0,959
	+5 dB	0,984*	0,989§	0,990§	0,984*
ISTS	-10 dB	0,908	0,872	0,930	0,915
	0 dB	0,973§	0,960	0,976§	0,974§
	+5 dB	0,992§	0,987	0,992§	0,985
White	-10 dB	0,857	0,733	0,940	0,905
	0 dB	0,960	0,902	0,988	0,969
	+5 dB	0,981	0,945	0,995	0,985
ICRA	-10 dB	0,851	0,758	0,889	0,878
	0 dB	0,958	0,932	0,967	0,961
	+5 dB	0,990	0,983	0,988§	0,987§

Tabela 14 – Valores médios de PESQ obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com a equação (63) com NCM. Para cada linha, os maiores valores de NCM estão em negrito, enquanto que o símbolo § indica que os valores não são estatisticamente distintos ($p > 0,05$).

Noise	SNR	WF	BM	CM	WP
Cafeteria Babble	-10 dB	1,141	1,364§	1,376§	1,400
	0 dB	1,578	1,998§	1,999§	2,043
	+5 dB	2,068	2,565	2,633§	2,611§
Inside Train	-10 dB	1,178	1,417§	1,409§	1,427§
	0 dB	1,557	1,984§	2,010§	2,068
	+5 dB	2,041	2,487	2,518	2,564
ISTS	-10 dB	1,659§	1,376	1,653§	1,689§
	0 dB	2,366§	1,986	2,400§	2,396§
	+5 dB	2,800§	2,430	2,835§	2,876
White	-10 dB	1,633§	1,245	1,622§	1,715
	0 dB	2,557	1,778	2,781§	2,755§
	+5 dB	3,138	2,284	3,374	3,308
ICRA	-10 dB	1,344§	1,127	1,363§	1,381
	0 dB	1,812§	1,471	1,762§	1,837§
	+5 dB	2,246	1,790	2,362§	2,368§

Figura 29 – Diagramas de caixa bidimensionais de qualidade (PESQ) e inteligibilidade (NCM) para o conjunto de teste e ruído *babble* a $SNR = -10$ dB. Valores médios de PESQ obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com a equação (63) com NCM. (a) WF em preto; (b) BM em verde; (c) CM em azul; (d) WP em vermelho; (e) não processado em magenta.



Os parâmetros da CM e WP que maximizam a equação (63), com $IM = I_{\%}$, ou seja, quando a SRMR-CI foi usada como métrica de inteligibilidade, são os mostrados na Tabela 15.

As Tabelas 16 e 17 mostram, respectivamente, os valores de inteligibilidade ($I_{\%}$) e PESQ obtidos considerando CM e WP ajustadas de acordo com a equação (63), com SRMR-CI como métrica de inteligibilidade. A Figura 30 ilustra o caso de ruído *babble* a $SNR = -10$ dB com diagramas de caixa.

Tabela 15 – Parâmetros da CM e WP obtidos com o conjunto de treinamento, buscando maximização da equação (63), considerando a SRMR-CI como métrica de inteligibilidade.

Noise	SNR	CM		WP	
		γ	μ [dB]	β	η [dB]
Cafeteria Babble	-10 dB	0,66	0	0,53	0
	0 dB	0,66	5	0,72	5
	+5 dB	0,66	5	0,53	5
Inside Train	-10 dB	0,57	20	0,53	10
	0 dB	0,57	15	0,53	10
	+5 dB	0,57	10	0,53	5

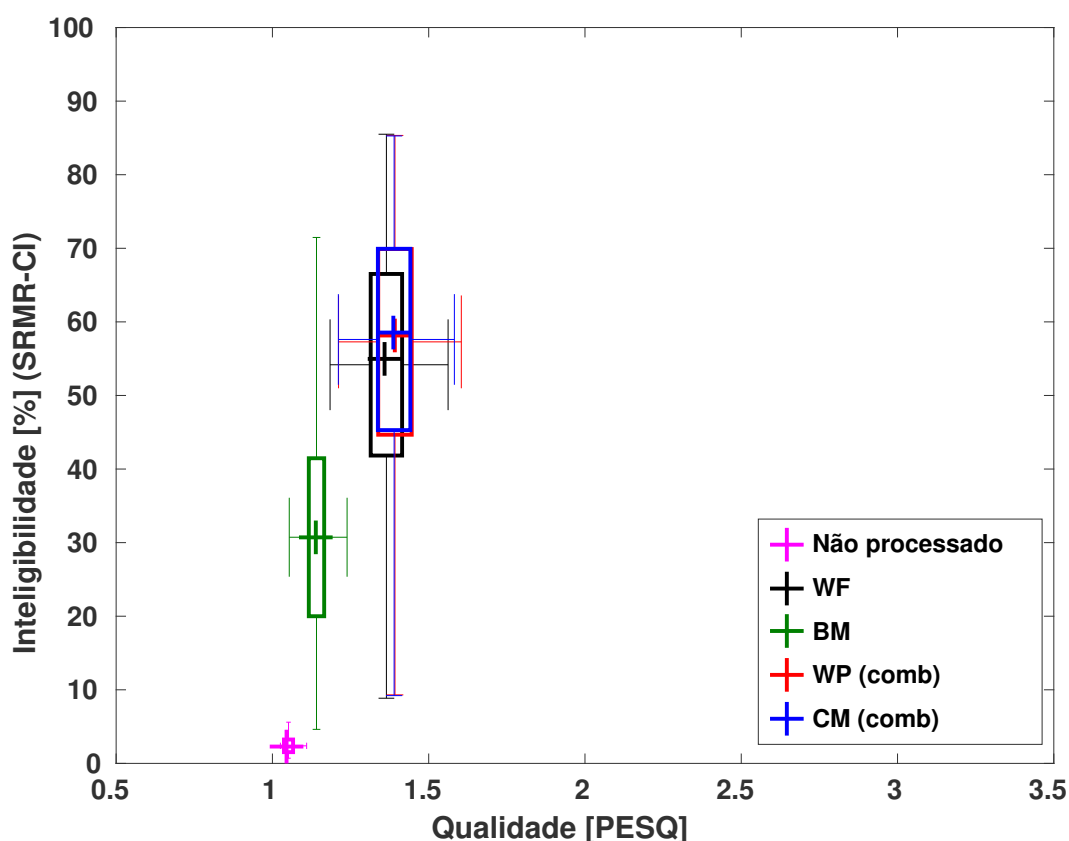
Tabela 16 – Valores médios de inteligibilidade ($I_{\%}$) obtidos, com a métrica SRMR-CI, por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com a equação (63) com SRMR-CI. Para cada linha, os maiores valores de inteligibilidade estão em negrito, enquanto que o símbolo § indica que os valores não são estatisticamente distintos ($p > 0,05$).

Noise	SNR	WF	BM	CM	WP
Cafeteria Babble	-10 dB	52,55	31,47	56,91§	56,41§
	0 dB	78,39§	71,48	78,77§	78,77§
	+5 dB	84,93§	83,23	84,87§	84,93§
Inside Train	-10 dB	58,34	36,50	65,03§	62,89§
	0 dB	78,22§	69,16	79,69§	79,47§
	+5 dB	84,22§	81,04	84,54§	84,60§

Tabela 17 – Valores médios de PESQ obtidos por cada máscara com o conjunto de teste. Os parâmetros da CM e WP foram otimizados de acordo com a equação (63) com SRMR-CI. Para cada linha, os maiores valores de PESQ estão em negrito, enquanto que o símbolo § indica que os valores não são estatisticamente distintos ($p > 0,05$).

Noise	SNR	WF	BM	CM	PW
Cafeteria Babble	-10 dB	1,364	1,141	1,396§	1,400§
	0 dB	1,998	1,578	2,078§	2,024§
	+5 dB	2,565	2,068	2,650§	2,631§
Inside Train	-10 dB	1,418	1,178	1,538	1,502
	0 dB	1,984	1,547	2,135	2,092
	+5 dB	2,487	2,041	2,600§	2,564§

Figura 30 – Diagramas de caixa bidimensionais de qualidade (PESQ) e inteligibilidade (SRMR-CI) para o conjunto de teste e ruído *babble* a $SNR = -10$ dB. Os parâmetros da CM e e WP foram otimizados de acordo com a equação (63) com SRMR-CI. (a) WF em preto; (b) BM em verde; (c) CM em azul; (d) WP em vermelho; (e) não processado em magenta.



3.4 DISCUSSÃO

A função de supressão proposta (CM), equação (55), foi obtida a partir de uma função custo que estabelece um compromisso entre redução de ruído e distorção da fala. Sua convexidade foi demonstrada provando que o ponto de derivada nula corresponde ao mínimo global. Seus parâmetros provêm dois graus de liberdade para o ajuste da função de supressão. O parâmetro μ está associado com o limiar de decisão, ou seja, $H(\xi = \mu) = 0,5$. Por sua vez, o parâmetro γ controla a inclinação da curva em relação a ξ . Diferentemente do amplamente conhecido WP, cada parâmetro da CM controla uma característica da função de supressão independentemente. Essa é uma característica interessante para aplicações que suportam adaptação automática da função de supressão baseada em variações do cenário acústico (JEONG; JEONG; OH, 2009). Esse é o caso de sistemas de aprimoramento de fala para aplicações embarcadas em tempo-real como aparelhos auditivos e implantes cocleares.

Outro achado importante com relação à conformabilidade da CM (equação (55)) é que, apesar de sua similaridade com o WP (equação (9)), essas máscaras não se igualam perfeitamente para valores genéricos dos parâmetros. Entretanto, a CM possibilita boas aproximações ao WP para uma grande variação de parâmetros β e η (Tabelas 4 e 5). O inverso não é verdadeiro (Tabelas 2 e 3), especialmente quanto à CM com características mais “duras”, ou seja, com $\gamma \gg 1$. Assim, pode-se dizer que a variedade de máscaras oferecida pela CM é mais abrangente que aquela possibilitada pelo WP.

No Experimento I, quando as máscaras paramétricas (CM e WP) são ajustadas de forma a maximizar a métrica de qualidade PESQ (seção 3.3.1), a performance em qualidade da CM é a melhor, em relação às máscaras BM e WF, para todos os tipos de ruído e SNR estudados. O desempenho do WP nesse mesmo critério também é maior que o da BM e WF, e é estatisticamente equivalente ($p > 0,05$) ao da CM em alguns, mas não em todos os casos. Quanto menor a SNR, mais discrepantes são os resultados em favor das máscaras paramétricas em relação ao WF e à BM. Isso é válido para todos os experimentos.

Vale notar que no Experimento I, os parâmetros da CM e do WP são projetados independentemente do impacto nas métricas de inteligibilidade. Assim, é possível que a melhora na qualidade leve a uma diminuição na métrica de inteligibilidade, em relação à obtida com o WF, como visto na Figura 25. Quando é usada a métrica SRMR-CI, a inteligibilidade com as máscaras paramétricas se mantém maior que a obtida com o WF.

Considerando o Experimento II (seção 3.3.2), a CM e a WP apresentam performance similar ($p > 0,05$), e levam aos melhores resultados de inteligibilidade, em relação à BM e ao WF, tanto usando a NCM (Tabela 9), quanto a SRMR-CI (Tabela 11). Nesse último caso, quando a SNR é não-negativa, a performance da WF é estatística-

mente similar às das máscaras paramétricas.

Os Experimentos I e II indicam que através da alteração dos parâmetros da CM ou do WP é possível ajustar a performance em relação ao compromisso entre inteligibilidade e qualidade. Isso mostra a versatilidade dessas máscaras e seu potencial de ajuste às preferências individuais do ouvinte.

Tendo em vista o Experimento III (seção 3.3.3), a CM é a máscara que obtém a melhor performance na métrica NCM. Isso é válido para todos os tipos de ruído e SNR, exceto para ruído ICRA a $SNR = +5\text{dB}$ (Tabela 13). Entretanto, para esse último caso, a diferença de NCM entre a obtida com a CM e o melhor desempenho, WF, é de apenas 0,2%. Por sua vez, o WP leva a melhores valores de PESQ, inclusive em comparação com a CM, na maior parte dos casos estudados (Tabela 14). Entretanto, o desempenho dessas duas máscaras pode ser considerado similar, tendo em conta que a mínima variação perceptível (JND - *Just Noticeable Difference*) de PESQ é 0,2 (SERVETTI; MARTIN, 2005; WILSON, K. W. *et al.*, 2008).

O experimento III usando a métrica SRMR-CI indica um desempenho similar em inteligibilidade com a CM e WP (Tabela 16). Para valores não-negativos de SNR, o desempenho do WF se equivale estatisticamente aos das máscaras paramétricas. Considerando-se a JND de PESQ, pode-se considerar que o desempenho em qualidade da CM e da WP também é equivalente.

Em todos os experimentos o desempenho da BM foi significativamente pior que o das máscaras suaves WF, CM e WP. Isso está de acordo com resultados experimentais apresentados em Koning, Madhu e Wouters (2015), Koning *et al.* (2018) e Madhu *et al.* (2013). Além disso, os resultados obtidos com a métrica SRMR-CI indicam que a CM pode trazer benefícios significativos à percepção de fala por usuários de IC, sobretudo para cenários com valores negativos de SNR.

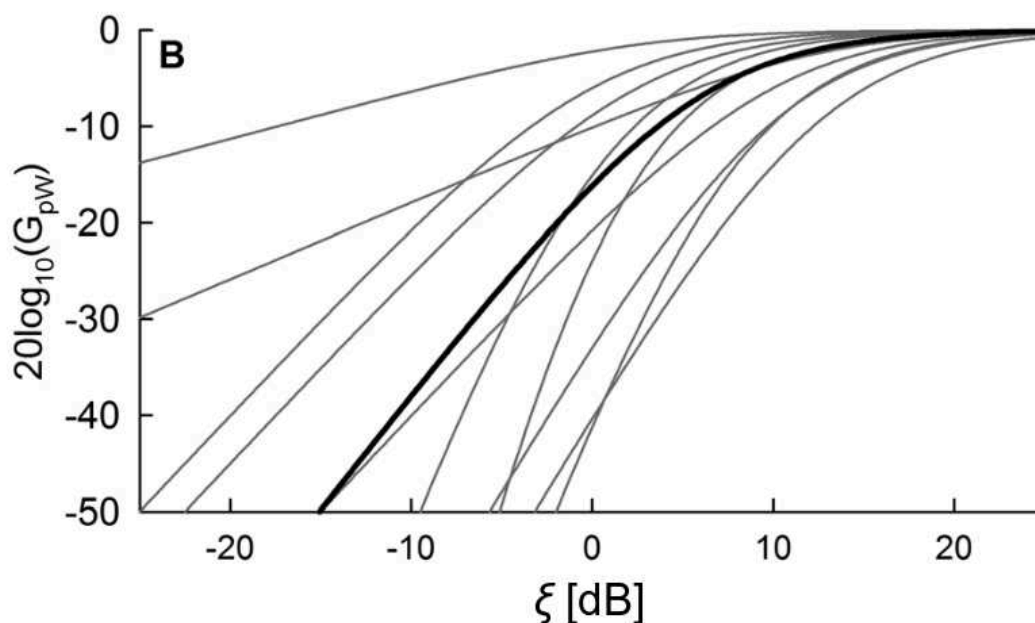
Os resultados das simulações mostram que, para todas as condições analisadas, o valor ótimo do parâmetro γ é menor que 1, que corresponde a máscaras mais “suaves” em comparação com o WF (Tabelas 6, 8, 10, 12 e 15). Pode-se observar também que, pela Tabela 6, os valores de PESQ são maximizados quando $\mu > 0\text{dB}$, enquanto que os valores de NCM são máximos para $\mu \leq 0\text{dB}$.

Apesar de, em geral, normouvintes apresentarem uma preferência por máscaras suaves, a BM tem relevância significativa. Experimentos com deficientes auditivos mostram uma grande variabilidade inter-indivíduos no desempenho dessas máscaras (KONING *et al.*, 2018; KONING; MADHU; WOUTERS, 2015; MAUGER; DAWSON; HERSBACH, 2012), sendo que, para algumas condições, alguns indivíduos preferem máscaras mais duras. Assim, uma máscara mais flexível como a CM (em relação à WP) deve ser uma escolha mais adequada para aparelhos auditivos e implantes cocleares.

A Figura 31, reproduzida de Mauger, Dawson e Hersbach (2012), ilustra a variabilidade na preferência pessoal de usuários de IC dentre máscaras WP com diferentes

parâmetros (β, η) .

Figura 31 – Função de supressão WP preferida por 10 usuários de IC diferentes. A linha em destaque representa a média das 10. Adaptado de (MAUGER; DAWSON; HERSBACH, 2012).



A literatura científica na área mostra que o ajuste ótimo do conjunto de parâmetros de uma máscara tempo-frequência não é uma tarefa trivial, dependendo da aplicação (aparelhos auditivos, reconhecimento automático de fala), cenário acústico, SNR, além da preferência individual subjetiva. Considerando-se a WP, uma máscara bem conhecida pela comunidade, apesar de esforços prévios, ainda não há uma maneira estabelecida para se fazer o projeto de parâmetros. Bolner e Hersbach (2018) sugerem que a inclinação da função de supressão seja inversamente proporcional à SNR do sinal. Outros trabalhos propõem o projeto dos parâmetros de acordo com estatísticas de maior ordem dos sinais de entrada (CHEHRESA; SAVOJI, 2009), do *pitch* (SILVA; BERMUDEZ, 2011) ou da preferência pessoal média (MAUGER; ARORA; DAWSON, 2012). Daniel *et al.* (2013) apresentam uma metodologia para a obtenção de uma curva de supressão individualizada, não atrelada a nenhuma função parametrizada.

Considerando-se $\gamma \rightarrow +\infty$, ou $\delta \rightarrow 1/2$, a função custo apresentada em (15) pode ser interpretada como a soma ponderada da raiz quadrada das distorções relacionadas à fala e ao ruído. Essa observação suporta a otimalidade da BM nesse contexto. A otimalidade da BM já foi abordada anteriormente. Em Li e Wang (2009), são apresentados argumentos que suportam a otimalidade da máscara binária ideal (com $\mu = 0$ dB) em relação a todas as máscaras binárias. Além disso, Lu e Loizou (2011) demonstraram que, assumindo que o quadrado da magnitude do espectro da mistura é igual à

soma dos quadrados das magnitudes espectrais da fala e do ruído, a BM é o estimador Máximo a Posteriori (MAP) do quadrado da magnitude espectral da fala. Entretanto, uma prova teórica da otimalidade da BM no contexto de um problema de minimização de uma função custo nunca havia sido demonstrada anteriormente.

Finalmente, a reunião das máscaras analisadas sob uma mesma estrutura teórica resultou em uma abordagem enriquecedora que traz um novo ponto de vista, novos significados e entendimento global sobre um grupo de técnicas previamente estabelecidas de maneira desconexa entre si (TSAO; LAI, 2016).

3.5 CONSIDERAÇÕES FINAIS

Neste capítulo foi demonstrado que uma versão até então considerada heurística do filtro de Wiener parametrizado, como sugerido em Fontaine *et al.* (2017), é, de fato, a máscara tempo-frequência ótima que minimiza um compromisso entre as distorções relativas à fala e ao ruído. A função custo associada é definida como uma soma ponderada de potências das magnitudes espectrais das distorções relativas à fala e ruído. A máscara resultante permite, através do projeto de seus parâmetros, o ajuste da inclinação e viés (deslocamento horizontal) da função de supressão em relação à SNR local, ξ , de maneira independente e intuitiva. Isso é uma característica desejável para aplicações que visam ao ajuste automático da função de supressão de acordo com mudanças no cenário acústico. Esse é o caso de sistemas de aprimoramento de voz em tempo-real em aplicações como aparelhos auditivos e implantes cocleares.

Simulações numéricas extensivas indicam que a máscara tempo-frequência aqui proposta, CM, tem performance melhor que o WF, principalmente para condições com SNR negativa, e pelo menos similar ao conhecido e heurísticamente desenvolvido WP. Além da otimalidade, outra vantagem da CM em relação à WP é que esta aproxima máscaras duras com maior precisão. Essa característica se mostra importante considerando a preferência por máscaras duras em determinados cenários acústicos, demonstrada em experimentos com deficientes auditivos.

Este trabalho provê uma estrutura teórica unificada para a derivação e interpretação da otimalidade de uma família de máscaras tempo-frequência que inclui máscaras bem-conhecidas como o WF e a BM, assim como algumas de suas variações. Por outro lado, apesar de apresentar potencial para um aumento da inteligibilidade para utilizadores de implante coclear, em comparação com o filtro de Wiener, a CM não tem em sua concepção uma ligação direta com as características de processamento do implante coclear. No próximo capítulo, um filtro desenvolvido especificamente para essa aplicação é apresentado.

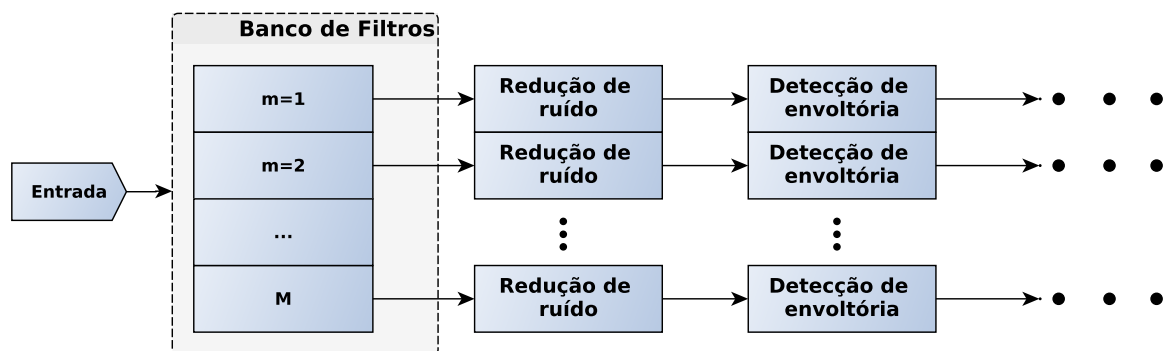
4 UM FILTRO QUE MINIMIZA O ERRO DO ENVELOPE QUADRÁTICO

As técnicas vistas até aqui se aplicam originalmente em uma estrutura na qual a decomposição do sinal em faixas de frequência é feita através da transformada de Fourier (Figura 16, capítulo 2). Essas técnicas podem ser utilizadas em implantes cocleares, mas nenhuma delas, mesmo a desenvolvida no capítulo 3, foi concebida pensando-se nas características específicas desta aplicação.

Neste capítulo, é desenvolvido um método de redução de ruído considerando algumas dessas características, principalmente a perda da informação de fase, como visto no capítulo 1. Tendo em vista que a inteligibilidade de normouvintes e de usuários de IC é consideravelmente diferente para sinais processados por uma mesma técnica (KONING; MADHU; WOUTERS, 2015), espera-se que um método desenvolvido especificamente para ICs leve a melhores índices de inteligibilidade.

A filtragem é realizada no domínio do tempo, tendo em vista uma arquitetura em que essa decomposição é feita com um banco de filtros. Esse filtro visa a minimizar, em cada canal de saída do banco de filtros, o erro do envelope do sinal processado em relação ao envelope temporal do sinal de interesse. A Figura 32 mostra como o bloco de redução de ruído se insere no sistema de processamento do implante.

Figura 32 – Diagrama de blocos do processamento do implante coclear no domínio do tempo com redução de ruído.



Nesse contexto, o sinal na saída de cada canal do banco de filtros tem sua envoltória detectada. Essa detecção pode se dar pela retificação seguida de filtragem passa-baixas, ou utilizando a transformada de Hilbert do sinal (WOUTERS; MCDERMOTT; FRANCA, 2015). Neste capítulo, utiliza-se a última opção, de forma a aproveitar as propriedades da transformada de Hilbert, descritas a seguir.

4.1 DESENVOLVIMENTO MATEMÁTICO

Na sequência deste capítulo, escalares são representadas com letra minúscula em itálico (*a*); vetores coluna, por letra minúscula em negrito e itálico (***a***); e matrizes,

Utilizando essa representação, define-se a transformada de Hilbert de um vetor \mathbf{z} como sendo

$$\tilde{\mathbf{z}}[n] = \mathbf{H}\mathbf{z}[n], \quad (68)$$

em que $\tilde{\mathbf{z}}[n] = [\tilde{z}[n], \tilde{z}[n-1], \dots, \tilde{z}[n-N]]^T$, e $\mathbf{z}[n] = [z[n], z[n-1], \dots, z[n-N]]^T$.

Com o auxílio da transformada de Hilbert, pode-se escrever o sinal em sua forma analítica. O sinal analítico é uma representação complexa de um sinal tal que seu espectro não contenha componentes em frequências negativas. O sinal analítico é definido como:

$$z_a[n] = z[n] + j\tilde{z}[n] = Z_a[n] e^{j\phi_z[n]}, \quad (69)$$

em que z_a é o sinal analítico e \tilde{z} é a transformada de Hilbert de z .

Esta representação explicita a fase instantânea, dada por

$$\phi_z[n] = \angle z_a[n] = \arctan\left(\frac{\tilde{z}[n]}{z[n]}\right), \quad (70)$$

e o envelope temporal instantâneo, calculado da seguinte maneira

$$Z_a[n] = |z_a[n]| = \sqrt{z^2[n] + \tilde{z}^2[n]}. \quad (71)$$

Por conseguinte, o quadrado do envelope, grafado como $a_z[n]$, é definido por

$$a_z[n] = |z_a[n]|^2 = z_a[n] z_a^*[n]. \quad (72)$$

O sinal analítico em forma vetorial, utilizando a equação (68), é dado por

$$\mathbf{z}_a = (\mathbf{I} + j\mathbf{H}) \mathbf{z}, \quad (73)$$

em que \mathbf{I} é a matriz identidade e \mathbf{H} é a matriz representando a transformada de Hilbert.

Outra definição útil para a sequência deste capítulo é a de matriz de correlação cruzada entre dois sinais z e w , que é representada por

$$\mathbf{R}_{\mathbf{z}, \mathbf{w}} = E\{\mathbf{z}\mathbf{w}^H\}, \quad (74)$$

em que o operador $E\{\cdot\}$ representa o valor esperado.

Quando houver apenas um subscrito, está-se referindo à matriz de autocorrelação,

$$\mathbf{R}_{\mathbf{z}} = E\{\mathbf{z}\mathbf{z}^H\}. \quad (75)$$

Cada elemento dessas matrizes é representado, respectivamente, por $r_{z,w}[l] = E\{z[n]w^*[n+l]\}$ e $r_z[l] = E\{z[n]z^*[n+l]\}$.

Algumas propriedades úteis para o desenvolvimento deste capítulo são elencadas a seguir.

Propriedade 1. Identidade de multiplicação de vetores.

$$\mathbf{a}^H \mathbf{b} = \left[(\mathbf{a}^H \mathbf{b})^H \right]^H \quad (76)$$

$$= \left[\mathbf{b}^H \mathbf{a} \right]^H \quad (77)$$

$$= \left[(\mathbf{b}^H \mathbf{a})^* \right]^T \quad (78)$$

$$= \left[\mathbf{b}^T \mathbf{a}^* \right]^T \quad (79)$$

Como $\mathbf{b}^T \mathbf{a}^*$ é um escalar,

$$\mathbf{a}^H \mathbf{b} = \mathbf{b}^T \mathbf{a}^*. \quad (80)$$

Propriedade 2. A sequência de autocorrelação da transformada de Hilbert é derivada a seguir. \tilde{z} é uma transformação linear (equação (64)) do sinal z . Desta forma, a densidade espectral de potência de \tilde{z} é dada por Kay (2006):

$$P_{\tilde{z}}(\omega) = |\mathbb{H}(\omega)|^2 P_z(\omega) = P_z(\omega). \quad (81)$$

Sendo a sequência de autocorrelação de um sinal a transformada inversa de Fourier da densidade espectral de potência do sinal, a sequência de autocorrelação de \tilde{z} também é igual à de z .

$$r_{\tilde{z}}[l] = r_z[l] \quad (82)$$

Propriedade 3. A correlação cruzada entre um sinal e sua transformada de Hilbert.

$$r_{z, \tilde{z}}[l] = E \{ z[n] \tilde{z}^*[n+l] \} \quad (83)$$

$$= E \{ z[n-l] (\tilde{h}^*[n] * z^*[n]) \} \quad (84)$$

$$= E \left\{ z[n-l] \sum_{m=-\infty}^{\infty} \tilde{h}^*[m] z^*[n-m] \right\} \quad (85)$$

$$= \sum_{m=-\infty}^{\infty} \tilde{h}^*[m] E \{ z[n-l] z^*[n-m] \} \quad (86)$$

$$= \sum_{m=-\infty}^{\infty} \tilde{h}^*[m] r_z[l-m] \quad (87)$$

A equação (87) representa a convolução entre $\tilde{h}^*[l]$ e $r_z[l]$. Além disso, segundo a equação (65), $\tilde{h}[l] \in \mathbb{R}$. Assim,

$$r_{z,\tilde{z}}[l] = \tilde{h}[l] * r_z[l]. \quad (88)$$

Analogamente, tem-se:

$$r_{\tilde{z},z}[l] = -\tilde{h}[l] * r_z[l]. \quad (89)$$

Isso leva às seguintes identidades:

$$r_{z,\tilde{z}}[l] = -r_{\tilde{z},z}[l], \quad (90)$$

$$r_{z,\tilde{z}}[0] = r_{\tilde{z},z}[0] = 0. \quad (91)$$

Assim, $z[n]$ e $\tilde{z}[n]$ são descorrelacionados, enquanto que $z[n+l]$ e $\tilde{z}[n]$ podem ter correlação não nula.

Propriedade 4. Autocorrelação do sinal analítico.

$$r_{za}[l] = E\{z_a[n]z_a^*[n+l]\} \quad (92)$$

$$= E\{(z[n] + j\tilde{z}[n])(z[n+l] - j\tilde{z}[n+l])\} \quad (93)$$

$$= E\{z[n]z[n+l] + j\tilde{z}[n]z[n+l] - jz[n]\tilde{z}[n+l] + \tilde{z}[n]\tilde{z}[n+l]\} \quad (94)$$

$$= E\{z[n]z[n+l]\} + jE\{\tilde{z}[n]z[n+l]\} - jE\{z[n]\tilde{z}[n+l]\} + E\{\tilde{z}[n]\tilde{z}[n+l]\} \quad (95)$$

$$= r_z[l] + jr_{\tilde{z},z}[l] - jr_{z,\tilde{z}}[l] + r_{\tilde{z}}[l] \quad (96)$$

Utilizando as propriedades 2 e 3:

$$\begin{aligned} r_{za}[l] &= 2(r_z[l] + jr_{\tilde{z},z}[l]) \\ &= 2(r_z[l] - j\tilde{h}[l] * r_z[l]) \end{aligned} \quad (97)$$

Em forma matricial, a partir da equação (73), tem-se

$$R_{za} = E \{ z_a z_a^H \} \quad (98)$$

$$= E \{ (I + jH) z z^H (I + jH)^H \} \quad (99)$$

$$= (I + jH) E \{ z z^H \} (I - jH^T) \quad (100)$$

$$= (I + jH) R_z (I - jH^T) \quad (101)$$

$$= (R_z + jHR_z) (I - jH^T) \quad (102)$$

$$= R_z + jHR_z - jR_z H^T + HR_z H^T \quad (103)$$

$$= R_z + jR_{\tilde{z},z} - jR_{z,\tilde{z}} + R_{\tilde{z}}. \quad (104)$$

Utilizando as propriedades 2 e 3:

$$R_{za} = R_z + jR_{\tilde{z},z} + jR_{\tilde{z},z} + R_z \quad (105)$$

$$= 2(R_z - jHR_z) \quad (106)$$

$$= 2(I - jH)R_z. \quad (107)$$

Propriedade 5. Valor esperado de $z_a[n] z_a[n+l]$ é nulo.

$$E \{ z_a[n] z_a[n+l] \} = E \{ (z[n] + j\tilde{z}[n]) (z[n+l] + j\tilde{z}[n+l]) \} \quad (108)$$

$$= E \{ z[n] z[n+l] + j\tilde{z}[n] z[n+l] + jz[n] \tilde{z}[n+l] + j^2 \tilde{z}[n] \tilde{z}[n+l] \} \quad (109)$$

$$= E \{ z[n] z[n+l] \} + jE \{ \tilde{z}[n] z[n+l] \} + jE \{ z[n] \tilde{z}[n+l] \} - E \{ \tilde{z}[n] \tilde{z}[n+l] \} \quad (110)$$

$$= r_z[l] + jr_{\tilde{z},z}[l] + jr_{z,\tilde{z}}[l] - r_{\tilde{z}}[l] \quad (111)$$

$$= r_z[l] + jr_{\tilde{z},z}[l] - jr_{\tilde{z},z}[l] - r_z[l] \quad (112)$$

$$= 0 \quad (113)$$

Em forma matricial, pode-se escrever $E \{ z_a z_a^T \} = \mathbf{0}$.

4.2 MÉTODO PROPOSTO

Nesta seção são apresentadas algumas características dos sinais na saída de cada sub-banda k e em seguida é feita a derivação do filtro proposto.

4.2.1 Características dos sinais

Seja $y[n]$ o sinal na entrada do banco de filtros tal que

$$y[n] = x[n] + v[n], \quad (114)$$

em que $x[n]$ é o sinal de fala e $v[n]$, ruído aditivo.

O sinal de fala, $x[n]$ pode ser considerado estacionário localmente, ou seja, quando analisado em trechos de 10 a 40 ms (EPHRAIM; MALAH, 1984; LOIZOU, 2013). O ruído $v[n]$ também é considerado estacionário localmente. Supõe-se que ambos os sinais, $x[n]$ e $v[n]$, são independentes entre si e têm média zero (LOIZOU, 2013; LU; LOIZOU, 2011).

Pode-se considerar o ruído $v[n]$ como tendo distribuição gaussiana, sendo a soma de diversas fontes menos a de interesse. Se as amostras do sinal na entrada do banco de filtros lineares têm distribuição gaussiana, as amostras dos sinais em cada canal do banco de filtros (lineares) também têm distribuição gaussiana (MANOLAKIS; INGLE; KOGON, 2005).

Também é possível considerar que a fala $x[n]$ tenha distribuição gaussiana, para trechos de até 2,5 ms (GAZOR; MEMBER; ZHANG, 2003). Já para trechos de 5 ms a 50 ms, a distribuição laplaciana é a mais apropriada para representar amostras do sinal de fala (GAZOR; MEMBER; ZHANG, 2003; JENSEN *et al.*, 2005).

Caso o sinal na entrada do banco de filtros tenha distribuição laplaciana, e que $f_k[n]$ seja a resposta ao impulso do filtro do canal k , pelo teorema do limite central, as amostras dos sinais em cada canal têm distribuição gaussiana, uma vez que $x_k[n] = f_k[n] * x[n] = \sum_{l=-\infty}^{+\infty} f_k[l] x[n-l]$. A Figura 33 mostra relações entre os momentos de segunda e quarta ordens do sinal de saída de cada canal k do banco de filtros, quando na entrada deste é colocado um processo laplaciano x .

A título de comparação, as relações entre os momentos de segunda e quarta ordens do sinal de saída de cada canal k do banco de filtros, quando na entrada deste é colocado um processo gaussiano x , estão ilustradas na Figura 34.

Tendo em vista a similaridade entre as Figuras 33 e 34, em cada canal k do filtro, tem-se $y_k[n] = x_k[n] + v_k[n]$, em que tanto $x_k[n]$ quanto $v_k[n]$ podem ser considerados como variáveis aleatórias gaussianas.

Dado um processo gaussiano $z[n]$ com média zero e variância σ_z^2 , têm-se as seguintes identidades (PAPOULIS; PILLAI, 2002).

Figura 33 – Relações entre momentos de segunda e quarta ordens do sinal de saída de cada canal k do banco de filtros, quando na entrada deste é colocado um processo laplaciano $x[n]$. Gráfico A mostra a relação entre os momentos de x_k e de sua transformada de Hilbert, \tilde{x}_k ; o gráfico B ilustra a *kurtosis* de x_k e de \tilde{x}_k ; e o gráfico C, a razão entre o valor esperado de $(x_k \tilde{x}_k)^2$ e as variâncias de x_k e \tilde{x}_k .

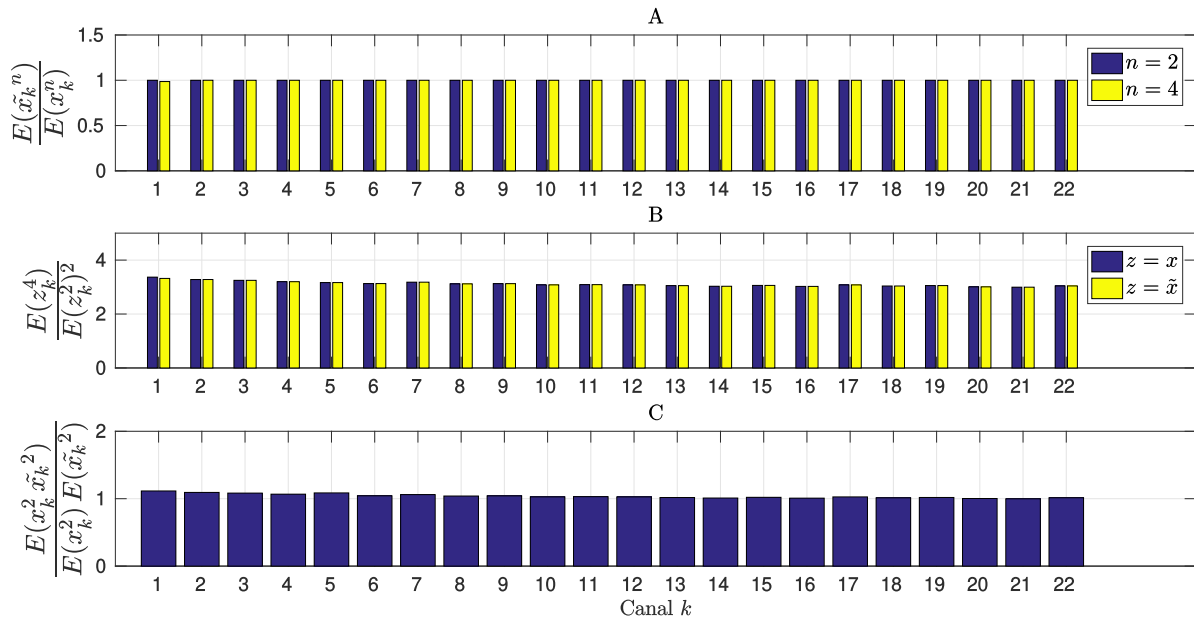
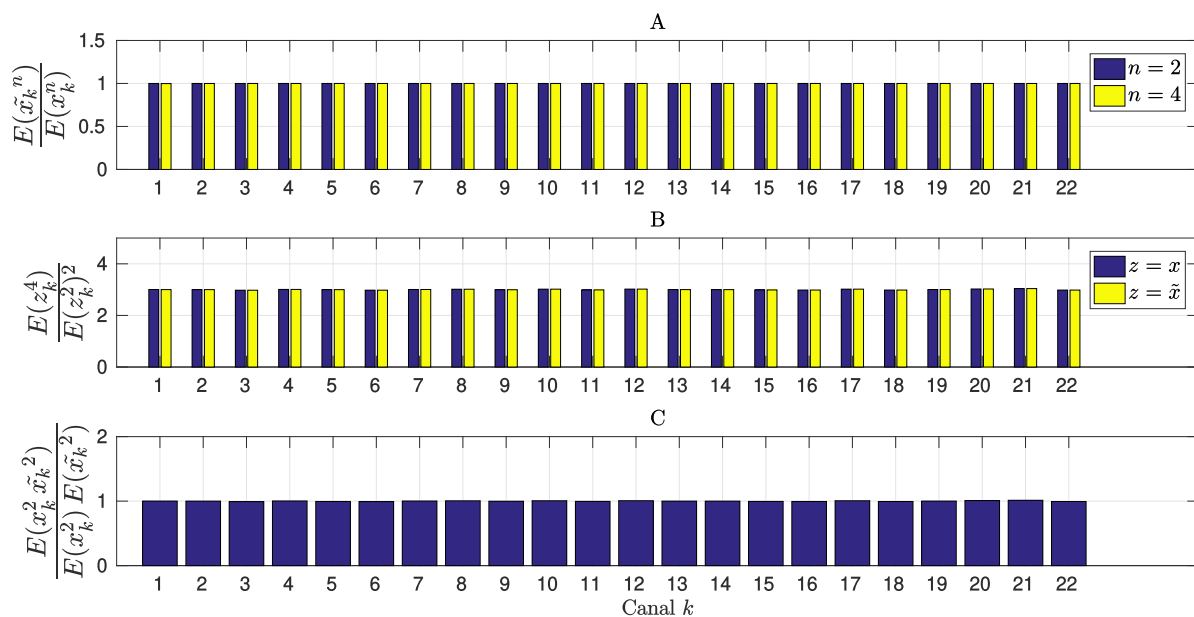


Figura 34 – Relações entre momentos de segunda e quarta ordens do sinal de saída de cada canal k do banco de filtros, quando na entrada deste é colocado um processo gaussiano $x[n]$. Gráfico A mostra a relação entre os momentos de x_k e de sua transformada de Hilbert, \tilde{x}_k ; o gráfico B ilustra a *kurtosis* de x_k e de \tilde{x}_k ; e o gráfico C, a razão entre o valor esperado de $(x_k \tilde{x}_k)^2$ e as variâncias de x_k e \tilde{x}_k .



$$E \{ z^2[n] \} = \sigma_z^2 \quad (115)$$

$$E \{ \tilde{z}^2[n] \} = \sigma_z^2 \quad (116)$$

$$E \{ z^4[n] \} = 3\sigma_z^4 \quad (117)$$

Os resultados mostrados na Figura 34 corroboram as equações (115 – 117). Além disso, permitem as seguintes suposições.

$$E \{ \tilde{z}^4[n] \} = 3\sigma_z^4 \quad (118)$$

$$E \{ \tilde{z}^2[n] z^2[n] \} = E \{ \tilde{z}^2[n] \} E \{ z^2[n] \} = \sigma_z^4 \quad (119)$$

4.2.2 Formulação do filtro proposto

No desenvolvimento do filtro proposto, duas características de ICs usando a técnica de estimulação CIS (ver capítulo 1.3) ou suas variações são consideradas: a informação de fase do sinal de entrada é perdida; o processamento em cada sub-banda de frequência é realizado de forma independente.

Desta forma, propõe-se um filtro \mathbf{g}_k que minimize o erro entre os envelopes temporais do sinal em sua saída e do sinal de interesse x_k , dado por $\sqrt{a_{xk}[n]}$, na k -ésima sub-banda. Essa formulação desconsidera a informação de fase, ou estrutura temporal fina, do sinal de entrada. a_{xk} é calculado de acordo com a equação (72) e os filtros em cada canal k são independentes entre si.

O filtro é representado por \mathbf{g}_k e permite a estimação do sinal analítico da fala $\hat{x}_{ak}[n]$ a partir de um vetor com amostras do sinal analítico da mistura $\mathbf{y}_{ak}[n]$, de forma que:

$$\hat{x}_{ak}[n] = \mathbf{g}_k^H[n] \mathbf{y}_{ak}[n]. \quad (120)$$

Como medida de qualidade da informação estimada do sinal desejado em cada canal, define-se aqui o erro entre os quadrados dos envelopes do sinal estimado e do sinal desejado:

$$\varepsilon = |\hat{x}_{ak}[n]|^2 - |x_{ak}[n]|^2. \quad (121)$$

Substituindo a equação (120) na equação (121), tem-se:

$$\varepsilon = \left| \mathbf{g}_k^H[n] \mathbf{y}_{ak}[n] \right|^2 - |x_{ak}[n]|^2. \quad (122)$$

O vetor sinal analítico $y_{ak}[n]$ pode ser calculado com o uso da transformação linear dada pela matriz $\mathbf{B} = \mathbf{I} + j\mathbf{H}$, como mostrado na equação (73), de forma que

$$y_{ak}[n] = \mathbf{B}y_k[n]. \quad (123)$$

Analogamente, $x_{ak}[n]$ pode ser representado de acordo com

$$x_{ak}[n] = \mathbf{e}^T \mathbf{B}x_k[n], \quad (124)$$

sendo $\mathbf{e} = [1 \ 0 \ 0 \ 0 \ \dots]^T$.

Pode-se, então, reescrever a equação (122) da seguinte maneira. O índice $[n]$ é omitido por clareza a partir deste ponto.

$$\varepsilon = \left| \mathbf{g}_k^H \mathbf{B}y_k \right|^2 - \left| \mathbf{e}^T \mathbf{B}x_k \right|^2. \quad (125)$$

A partir da função erro apresentada em (125), define-se seu valor quadrático médio como função custo a ser minimizada:

$$J = E \left\{ \varepsilon^2 \right\}. \quad (126)$$

Substituindo-se (125) em (126):

$$\begin{aligned} J &= E \left\{ \left(\left| \mathbf{g}_k^H \mathbf{B}y_k \right|^2 - \left| \mathbf{e}^T \mathbf{B}x_k \right|^2 \right)^2 \right\} \\ &= E \left\{ \left(\mathbf{g}_k^H \mathbf{B}y_k y_k^H \mathbf{B}^H \mathbf{g}_k - \mathbf{e}^T \mathbf{B}x_k x_k^H \mathbf{B}^H \mathbf{e} \right)^2 \right\} \\ &= E \left\{ \left(\mathbf{g}_k^H \mathbf{B}y_k y_k^H \mathbf{B}^H \mathbf{g}_k \right)^2 \right\} - 2E \left\{ \mathbf{g}_k^H \mathbf{B}y_k y_k^H \mathbf{B}^H \mathbf{g}_k \mathbf{e}^T \mathbf{B}x_k x_k^H \mathbf{B}^H \mathbf{e} \right\} \\ &\quad + E \left\{ \left(\mathbf{e}^T \mathbf{B}x_k x_k^H \mathbf{B}^H \mathbf{e} \right)^2 \right\}. \end{aligned} \quad (127)$$

Lembrando que $x_k[n]$ e $v_k[n]$ e são considerados como tendo média zero e sendo independentes entre si e, conseqüentemente, descorrelacionados (KAY, 2006), e sendo $\mathbf{G}_k = \mathbf{g}_k \mathbf{g}_k^H$, cada parcela da expressão pode ser escrita da seguinte maneira.

1. O primeiro termo da equação (127) é determinado da seguinte forma:

$$\begin{aligned}
& E \left\{ (\mathbf{g}_k^H \mathbf{B} \mathbf{y}_k \mathbf{y}_k^H \mathbf{B}^H \mathbf{g}_k)^2 \right\} \\
&= E \left\{ \left(\mathbf{g}_k^H \mathbf{y}_{ak} \mathbf{y}_{ak}^H \mathbf{g}_k \right)^2 \right\} \\
&= E \left\{ \left(\mathbf{g}_k^H (\mathbf{x}_{ak} + \mathbf{v}_{ak}) (\mathbf{x}_{ak} + \mathbf{v}_{ak})^H \mathbf{g}_k \right)^2 \right\} \\
&= E \left\{ \mathbf{g}_k^H (\mathbf{x}_{ak} \mathbf{x}_{ak}^H + \mathbf{x}_{ak} \mathbf{v}_{ak}^H + \mathbf{v}_{ak} \mathbf{x}_{ak}^H + \mathbf{v}_{ak} \mathbf{v}_{ak}^H) \right. \\
&\quad \left. \mathbf{g}_k \mathbf{g}_k^H (\mathbf{x}_{ak} \mathbf{x}_{ak}^H + \mathbf{x}_{ak} \mathbf{v}_{ak}^H + \mathbf{v}_{ak} \mathbf{x}_{ak}^H + \mathbf{v}_{ak} \mathbf{v}_{ak}^H) \mathbf{g}_k \right\} \\
&= E \left\{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + E \left\{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \right\}. \tag{128}
\end{aligned}$$

Em seguida, são avaliadas separadamente as parcelas de (128).

$$\begin{aligned}
\text{a) } E \left\{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \right\} \\
= \mathbf{g}_k^H E \left\{ \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \right\} \mathbf{g}_k \tag{129}
\end{aligned}$$

$$\begin{aligned}
\text{b) } E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
&= E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \} E \{ \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
&= \mathbf{g}_k^H E \{ \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \} \underbrace{E \{ \mathbf{v}_{ak}^H \}}_{=0} \mathbf{g}_k \\
&= 0
\end{aligned} \tag{130}$$

$$\begin{aligned}
\text{c) } E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
&= E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \} \\
&= E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \mathbf{x}_{ak}^H \mathbf{g}_k \} E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \} \\
&= \mathbf{g}_k^H E \{ \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \mathbf{x}_{ak}^H \} \mathbf{g}_k \mathbf{g}_k^H \underbrace{E \{ \mathbf{v}_{ak} \}}_{=0} \\
&= 0
\end{aligned} \tag{131}$$

$$\begin{aligned}
\text{d) } E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
&= \mathbf{g}_k^H E \{ \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \} \mathbf{g}_k \\
&= \mathbf{g}_k^H E \{ \mathbf{x}_{ak} \mathbf{x}_{ak}^H \} \mathbf{G}_k E \{ \mathbf{v}_{ak} \mathbf{v}_{ak}^H \} \mathbf{g}_k \\
&= \mathbf{g}_k^H \mathbf{R}_{xak} \mathbf{G}_k \mathbf{R}_{vak} \mathbf{g}_k
\end{aligned} \tag{132}$$

$$\begin{aligned}
\text{e) } E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
&= E \{ \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
&= E \{ \mathbf{v}_{ak}^H \mathbf{g}_k \} E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
&= \underbrace{E \{ \mathbf{v}_{ak}^H \}}_{=0} \mathbf{g}_k \mathbf{g}_k^H E \{ \mathbf{x}_{ak} \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \} \mathbf{g}_k \\
&= 0
\end{aligned} \tag{133}$$

$$\begin{aligned}
\text{f) } E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
&= E \{ \mathbf{v}_{ak}^H \mathbf{g}_k \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{g}_k^H \mathbf{x}_{ak} \} \\
&= E \{ \mathbf{g}_k^T \mathbf{v}_{ak}^* \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^T \mathbf{g}_k^* \}
\end{aligned} \tag{134}$$

Usando a propriedade 5:

$$\begin{aligned}
 & E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
 &= \mathbf{g}_k^T \underbrace{E \{ \mathbf{v}_{ak}^* \mathbf{v}_{ak}^H \}}_{=0} \mathbf{G}_k \underbrace{E \{ \mathbf{x}_{ak} \mathbf{x}_{ak}^T \}}_{=0} \mathbf{g}_k^* \\
 &= 0
 \end{aligned} \tag{135}$$

$$\begin{aligned}
 \text{g) } & E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
 &= E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
 &= E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
 &= \mathbf{g}_k^H E \{ \mathbf{x}_{ak} \mathbf{x}_{ak}^H \} \mathbf{G}_k E \{ \mathbf{v}_{ak} \mathbf{v}_{ak}^H \} \mathbf{g}_k \\
 &= \mathbf{g}_k^H \mathbf{R}_{\mathbf{x}_{ak}} \mathbf{G}_k \mathbf{R}_{\mathbf{v}_{ak}} \mathbf{g}_k
 \end{aligned} \tag{136}$$

$$\begin{aligned}
 \text{h) } & E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
 &= E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \} \\
 &= E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \mathbf{v}_{ak}^H \mathbf{g}_k \} E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \} \\
 &= \mathbf{g}_k^H \underbrace{E \{ \mathbf{v}_{ak} \}}_{=0} \mathbf{g}_k^H E \{ \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \mathbf{x}_{ak}^H \} \mathbf{g}_k \\
 &= 0
 \end{aligned} \tag{137}$$

$$\begin{aligned}
 \text{i) } & E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
 &= E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
 &= E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \} E \{ \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
 &= \mathbf{g}_k^H \underbrace{E \{ \mathbf{v}_{ak} \}}_{=0} \mathbf{g}_k^H E \{ \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \} \mathbf{g}_k \\
 &= 0
 \end{aligned} \tag{138}$$

$$\begin{aligned}
j) \quad & E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
&= E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
&= \mathbf{g}_k^H E \{ \mathbf{v}_{ak} \mathbf{v}_{ak}^H \} \mathbf{G}_k E \{ \mathbf{x}_{ak} \mathbf{x}_{ak}^H \} \mathbf{g}_k \\
&= \mathbf{g}_k^H \mathbf{R}_{vak} \mathbf{G}_k \mathbf{R}_{xak} \mathbf{g}_k \\
&= \mathbf{g}_k^H \mathbf{R}_{xak} \mathbf{G}_k \mathbf{R}_{vak} \mathbf{g}_k
\end{aligned} \tag{139}$$

$$\begin{aligned}
k) \quad & E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
&= E \{ \mathbf{x}_{ak}^H \mathbf{g}_k \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{g}_k^H \mathbf{v}_{ak} \} \\
&= E \{ \mathbf{g}_k^T \mathbf{x}_{ak}^* \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^T \mathbf{g}_k^* \}
\end{aligned} \tag{140}$$

Usando a propriedade 5:

$$\begin{aligned}
& E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
&= \mathbf{g}_k^T E \{ \underbrace{\mathbf{x}_{ak}^* \mathbf{x}_{ak}^H}_{=0} \} \mathbf{G}_k E \{ \underbrace{\mathbf{v}_{ak} \mathbf{v}_{ak}^T}_{=0} \} \mathbf{g}_k^* \\
&= 0.
\end{aligned} \tag{141}$$

$$\begin{aligned}
l) \quad & E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
&= E \{ \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
&= E \{ \mathbf{x}_{ak}^H \mathbf{g}_k \} E \{ \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
&= E \{ \underbrace{\mathbf{x}_{ak}^H}_{=0} \} \mathbf{g}_k E \{ \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \} \mathbf{g}_k \\
&= 0
\end{aligned} \tag{142}$$

$$\begin{aligned}
m) \quad & E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
&= \mathbf{g}_k^H E \{ \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \} \mathbf{g}_k \\
&= \mathbf{g}_k^H E \{ \mathbf{v}_{ak} \mathbf{v}_{ak}^H \} \mathbf{G}_k E \{ \mathbf{x}_{ak} \mathbf{x}_{ak}^H \} \mathbf{g}_k \\
&= \mathbf{g}_k^H \mathbf{R}_{vak} \mathbf{G}_k \mathbf{R}_{xak} \mathbf{g}_k \\
&= \mathbf{g}_k^H \mathbf{R}_{xak} \mathbf{G}_k \mathbf{R}_{vak} \mathbf{g}_k
\end{aligned} \tag{143}$$

$$\begin{aligned}
n) \ E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
&= E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \} \\
&= E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \mathbf{v}_{ak}^H \mathbf{g}_k \} E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \} \\
&= \mathbf{g}_k^H E \{ \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \mathbf{v}_{ak}^H \} \mathbf{g}_k \mathbf{g}_k^H \underbrace{E \{ \mathbf{x}_{ak} \}}_{=0} \\
&= 0
\end{aligned} \tag{144}$$

$$\begin{aligned}
o) \ E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
&= E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \} E \{ \mathbf{x}_{ak}^H \mathbf{g}_k \} \\
&= \mathbf{g}_k^H E \{ \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \} \underbrace{E \{ \mathbf{x}_{ak}^H \}}_{=0} \mathbf{g}_k \\
&= 0
\end{aligned} \tag{145}$$

$$\begin{aligned}
p) \ E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
&= \mathbf{g}_k^H E \{ \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \} \mathbf{g}_k
\end{aligned} \tag{146}$$

Utilizando os valores das equações (129)–(146), a primeira parcela da equação (127) torna-se:

$$\begin{aligned}
E \{ \mathbf{g}_k^H \mathbf{y}_{ak} \mathbf{y}_{ak}^H \mathbf{G}_k \mathbf{y}_{ak} \mathbf{y}_{ak}^H \mathbf{g}_k \} &= \mathbf{g}_k^H E \{ \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \} \mathbf{g}_k \\
&\quad + 4 \mathbf{g}_k^H \mathbf{R}_{xak} \mathbf{G}_k \mathbf{R}_{vak} \mathbf{g}_k \\
&\quad + \mathbf{g}_k^H E \{ \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \} \mathbf{g}_k.
\end{aligned} \tag{147}$$

2. O segundo termo da equação (127) é determinado da seguinte forma:

$$\begin{aligned}
E \{ \mathbf{g}_k^H \mathbf{B} \mathbf{y}_k \mathbf{y}_k^H \mathbf{B}^H \mathbf{g}_k \mathbf{e}^T \mathbf{B} \mathbf{x}_k \mathbf{x}_k^H \mathbf{B}^H \mathbf{e} \} \\
&= E \{ \mathbf{g}_k^H \mathbf{y}_{ak} \mathbf{y}_{ak}^H \mathbf{g}_k a_{xk} \} \\
&= E \{ \mathbf{g}_k^H (\mathbf{x}_{ak} + \mathbf{v}_{ak}) (\mathbf{x}_{ak}^H + \mathbf{v}_{ak}^H) \mathbf{g}_k a_{xk} \} \\
&= E \{ \mathbf{g}_k^H (\mathbf{x}_{ak} \mathbf{x}_{ak}^H + \mathbf{x}_{ak} \mathbf{v}_{ak}^H + \mathbf{v}_{ak} \mathbf{x}_{ak}^H + \mathbf{v}_{ak} \mathbf{v}_{ak}^H) \mathbf{g}_k a_{xk} \} \\
&= E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H a_{xk} \mathbf{g}_k \} + E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H a_{xk} \mathbf{g}_k \} \\
&\quad + E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H a_{xk} \mathbf{g}_k \} + E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H a_{xk} \mathbf{g}_k \}.
\end{aligned} \tag{148}$$

Cada parcela é avaliada a seguir.

$$\begin{aligned}
 \text{a) } E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H a_{xk} \mathbf{g}_k \} \\
 &= \mathbf{g}_k^H E \{ \mathbf{x}_{ak} \mathbf{x}_{ak}^H x_{ak} x_{ak}^* \} \mathbf{g}_k \\
 &= \mathbf{g}_k^H E \{ \mathbf{M} x_{ak} x_{ak}^* \} \mathbf{g}_k
 \end{aligned} \tag{149}$$

$$\mathbf{M} = \begin{bmatrix} x_{ak}[n] x_{ak}^*[n] & \dots & x_{ak}[n] x_{ak}^*[n-N+1] \\ \vdots & \ddots & \vdots \\ x_{ak}[n-N+1] x_{ak}^*[n] & \dots & x_{ak}[N-1] x_{ak}^*[N-1] \end{bmatrix}$$

$$\begin{aligned}
 \text{b) } E \{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{v}_{ak}^H a_{xk} \mathbf{g}_k \} \\
 &= E \{ \mathbf{g}_k^H \mathbf{x}_{ak} a_{xk} \mathbf{v}_{ak}^H \mathbf{g}_k \} \\
 &= \mathbf{g}_k^H E \{ \mathbf{x}_{ak} a_{xk} \} \underbrace{E \{ \mathbf{v}_{ak}^H \}}_{=0} \mathbf{g}_k \\
 &= 0
 \end{aligned} \tag{150}$$

$$\begin{aligned}
 \text{c) } E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H a_{xk} \mathbf{g}_k \} \\
 &= E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{x}_{ak}^H a_{xk} \mathbf{g}_k \} \\
 &= \mathbf{g}_k^H \underbrace{E \{ \mathbf{v}_{ak} \}}_{=0} E \{ \mathbf{x}_{ak}^H a_{xk} \} \mathbf{g}_k \\
 &= 0
 \end{aligned} \tag{151}$$

$$\begin{aligned}
 \text{d) } E \{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H a_{xk} \mathbf{g}_k \} \\
 &= \mathbf{g}_k^H E \{ \mathbf{v}_{ak} \mathbf{v}_{ak}^H \} E \{ a_{xk} \} \mathbf{g}_k \\
 &= \mathbf{g}_k^H \mathbf{R}_{vak} \mathbf{g}_k e^T \mathbf{R}_{xak} e
 \end{aligned} \tag{152}$$

Assim, a segunda parcela de (127) é dada por

$$\begin{aligned}
 2E \{ \mathbf{g}_k^H \mathbf{B} \mathbf{y}_k \mathbf{y}_k^H \mathbf{B}^H \mathbf{g}_k e^T \mathbf{B} \mathbf{x}_k \mathbf{x}_k^H \mathbf{B}^H e \} \\
 = 2 \mathbf{g}_k^H E \{ \mathbf{M} x_{ak} x_{ak}^* \} \mathbf{g}_k + \mathbf{g}_k^H \mathbf{R}_{vak} \mathbf{g}_k e^T \mathbf{R}_{xak} e.
 \end{aligned} \tag{153}$$

3. O terceiro termo da equação (127) é determinado da seguinte forma:

$$\begin{aligned}
E \left\{ \left(\mathbf{e}^T \mathbf{B} \mathbf{x}_k \mathbf{x}_k^H \mathbf{B}^H \mathbf{e} \right)^2 \right\} &= E \left\{ x_{ak} x_{ak}^* x_{ak} x_{ak}^* \right\} \\
&= E \left\{ (x_k + j\tilde{x}_k)(x_k - j\tilde{x}_k)(x_k + j\tilde{x}_k)(x_k - j\tilde{x}_k) \right\} \\
&= E \left\{ (x_k^2 + \tilde{x}_k^2)(x_k^2 + \tilde{x}_k^2) \right\} \\
&= E \left\{ x_k^4 + 2\tilde{x}_k^2 \tilde{x}_k^2 + \tilde{x}_k^4 \right\} \\
&= E \left\{ x_k^4 \right\} + 2E \left\{ x_k^2 \tilde{x}_k^2 \right\} + E \left\{ \tilde{x}_k^4 \right\}. \tag{154}
\end{aligned}$$

Substituindo as equações (147), (153) e (154) na equação (127):

$$\begin{aligned}
J &= \mathbf{g}_k^H E \left\{ \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \right\} \mathbf{g}_k + 4 \mathbf{g}_k^H \mathbf{R}_{xak} \mathbf{G}_k \mathbf{R}_{vak} \mathbf{g}_k \\
&\quad + \mathbf{g}_k^H E \left\{ \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \right\} \mathbf{g}_k - 2 \left(\mathbf{g}_k^H E \left\{ \mathbf{M} \mathbf{x}_{ak} \mathbf{x}_{ak}^* \right\} \mathbf{g}_k \right. \\
&\quad \left. + \mathbf{g}_k^H \mathbf{R}_{vak} \mathbf{g}_k \mathbf{e}^T \mathbf{R}_{xak} \mathbf{e} \right) + E \left\{ x_k^4[n] \right\} + 2E \left\{ x_k^2[n] \tilde{x}_k^2[n] \right\} \\
&\quad + E \left\{ \tilde{x}_k^4[n] \right\}. \tag{155}
\end{aligned}$$

O vetor \mathbf{g}_k que minimiza a função custo é tal que $\frac{\partial J}{\partial \mathbf{g}_k^*} = \mathbf{0}$ (HJØRUNGNES *et al.*, 2007; MESSERSCHMITT, 2006).

Para uma função $f(\mathbf{z}, \mathbf{z}^*)$ de um vetor complexo \mathbf{z} , de acordo com Hjørungnes *et al.* (2007),

$$\frac{\partial}{\partial \mathbf{z}^*} \mathbf{z}^H \mathbf{A} \mathbf{z} = \mathbf{z}^T \mathbf{A}^T, \tag{156}$$

em que \mathbf{A} é uma matriz genérica.

Além disso, considerando uma segunda matriz genérica \mathbf{B} ,

$$\frac{\partial}{\partial \mathbf{z}^*} \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^H \mathbf{B} \mathbf{z} = \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^T \mathbf{B}^T + \mathbf{z}^H \mathbf{B} \mathbf{z} \mathbf{z}^T \mathbf{A}^T. \tag{157}$$

Um caso particular da equação (157), em que $\mathbf{B} = \mathbf{A}^H$, com $\mathbf{A} = \mathbf{a}^H \mathbf{a}$, resulta em

$$\frac{\partial}{\partial \mathbf{z}^*} (\mathbf{z}^H \mathbf{A} \mathbf{z})^2 = 2 \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^T \mathbf{A}^T. \tag{158}$$

Utilizando-se as equações (156), (157) e (158) pode-se calcular $\frac{\partial J}{\partial \mathbf{g}_k^*}$. As derivações das equações (156) e (158) estão detalhadas no apêndice A.

Derivando-se a equação 155 e usando as equações (156) e (158), tem-se

$$\begin{aligned}
\frac{\partial J}{\partial \mathbf{g}_k^*} &= \frac{\partial}{\partial \mathbf{g}_k^*} \mathbb{E} \left\{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{G}_k \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \right\} \\
&\quad + \frac{\partial}{\partial \mathbf{g}_k^*} 4 \mathbf{g}_k^H \mathbf{R}_{xak} \mathbf{G}_k \mathbf{R}_{vak} \mathbf{g}_k \\
&\quad + \frac{\partial}{\partial \mathbf{g}_k^*} \mathbb{E} \left\{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{G}_k \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \right\} \\
&\quad - \frac{\partial}{\partial \mathbf{g}_k^*} 2 \mathbf{g}_k^H \mathbb{E} \left\{ \mathbf{M} \mathbf{x}_{ak} \mathbf{x}_{ak}^* \right\} \mathbf{g}_k - \frac{\partial}{\partial \mathbf{g}_k^*} 2 \mathbf{g}_k^H \mathbf{R}_{vak} \mathbf{g}_k \mathbf{e}^T \mathbf{R}_{xak} \mathbf{e} \\
&\quad + \frac{\partial}{\partial \mathbf{g}_k^*} \mathbb{E} \left\{ x_k^4[n] \right\} + \frac{\partial}{\partial \mathbf{g}_k^*} 2 \mathbb{E} \left\{ x_k^2[n] \tilde{x}_k^2[n] \right\} \\
&\quad + \frac{\partial}{\partial \mathbf{g}_k^*} \mathbb{E} \left\{ \tilde{x}_k^4[n] \right\} \\
\frac{\partial J}{\partial \mathbf{g}_k^*} &= 2 \mathbb{E} \left\{ \mathbf{g}_k^H \mathbf{x}_{ak} \mathbf{x}_{ak}^H \mathbf{g}_k \mathbf{g}_k^T \mathbf{x}_{ak}^* \mathbf{x}_{ak}^T \right\} + 4 \left(2 \mathbf{g}_k^H \mathbf{R}_{xak} \mathbf{g}_k \mathbf{g}_k^T \mathbf{R}_{vak} \right) \\
&\quad + 2 \mathbb{E} \left\{ \mathbf{g}_k^H \mathbf{v}_{ak} \mathbf{v}_{ak}^H \mathbf{g}_k \mathbf{g}_k^T \mathbf{v}_{ak}^* \mathbf{v}_{ak}^T \right\} - 2 \mathbf{g}^T \mathbb{E} \left\{ \mathbf{M} \mathbf{x}_{ak} \mathbf{x}_{ak}^* \right\} \\
&\quad - 2 \mathbf{g}^T \mathbf{R}_{vak}^T \mathbf{e}^T \mathbf{R}_{xak} \mathbf{e}. \tag{159}
\end{aligned}$$

Tendo em conta a equação (159), a solução para $\frac{\partial J}{\partial \mathbf{g}_k^*} = \mathbf{0}$ não é trivial. Quanto maior o número de coeficientes do filtro maior a complexidade de se encontrar \mathbf{g}_k que minimize $J(\mathbf{g}_k)$. Entretanto, o filtro de um coeficiente pode ser derivado mais facilmente.

4.2.3 Filtro de um coeficiente

Ao se reduzir a complexidade à determinação de apenas um coeficiente permite-se a derivação de uma fórmula para o cálculo do ganho g_k ótimo, grafado como g_{k0} . Nesta situação particular, o estimador é definido como $\hat{x}_{ak}[\lambda] = g_{k0} y_{ak}[\lambda]$ ¹. Vale notar que o filtro de um coeficiente na banda de frequência k é um ganho que é aplicado à mistura y_k a cada instante de tempo λ . Assim, o filtro de um coeficiente equivale a uma máscara tempo-frequência.

Reescrevendo a equação (159) para o filtro de um coeficiente e igualando a zero, obtêm-se:

$$\frac{\partial J}{\partial g_k^*} = 0 \tag{160}$$

¹ Aqui o índice de tempo é alterado de n para λ para melhor representar a equivalência com máscaras tempo-frequência. Além disso, em ICs com estratégias derivadas da CIS, a taxa de estimulação do eletrodo é da ordem de 1000 pulsos/s enquanto que a frequência de amostragem do sinal de áudio é geralmente 16kHz. Isso significa que o sinal apresentado ao usuário equivale a um envelope decimado em relação à frequência de amostragem original e, conseqüentemente, o envelope da fala não precisa ser estimado para todas amostras do sinal de áudio. A mudança de notação, portanto, também evidencia que, na aplicação de interesse, a filtragem acontece a uma frequência de amostragem (relativa a λ) diferente daquela do áudio original (associada a n).

$$\begin{aligned}
& 2g_{k0}^2 g_{k0}^* E \left\{ |x_{ak}[\lambda]|^4 \right\} + 2g_{k0}^2 g_{k0}^* E \left\{ |v_{ak}[\lambda]|^4 \right\} \\
& + 4g_{k0}^2 g_{k0}^* (2r_{xak}[0]r_{vak}[0]) - 2g_{k0} E \left\{ |x_{ak}[\lambda]|^4 \right\} \\
& - 2g_{k0} r_{xak}[0]r_{vak}[0] = 0
\end{aligned} \tag{161}$$

$$\begin{aligned}
& g_{k0} g_{k0}^* E \left\{ |x_{ak}[\lambda]|^4 \right\} + g_{k0} g_{k0}^* E \left\{ |v_{ak}[\lambda]|^4 \right\} \\
& + 4g g_{k0}^* (r_{xak}[0]r_{vak}[0]) - E \left\{ |x_{ak}[\lambda]|^4 \right\} - r_{xak}[0]r_{vak}[0] = 0
\end{aligned} \tag{162}$$

$$\begin{aligned}
& |g_{k0}|^2 E \left\{ |x_{ak}[\lambda]|^4 \right\} + |g_{k0}|^2 E \left\{ |v_{ak}[\lambda]|^4 \right\} \\
& + 4|g_{k0}|^2 (r_{xak}[0]r_{vak}[0]) - E \left\{ |x_{ak}[\lambda]|^4 \right\} \\
& - r_{xak}[0]r_{vak}[0] = 0
\end{aligned} \tag{163}$$

$$\begin{aligned}
& |g_{k0}|^2 \left(E \left\{ |x_{ak}[\lambda]|^4 \right\} + E \left\{ |v_{ak}[\lambda]|^4 \right\} + 4(r_{xak}[0]r_{vak}[0]) \right) \\
& - E \left\{ |x_{ak}[\lambda]|^4 \right\} - r_{xak}[0]r_{vak}[0] = 0.
\end{aligned} \tag{164}$$

Evidenciando-se o valor de g_{k0} , tem-se

$$|g_{k0}|^2 = \frac{E \left\{ |x_{ak}[\lambda]|^4 \right\} + r_{xak}[0]r_{vak}[0]}{E \left\{ |x_{ak}[\lambda]|^4 \right\} + E \left\{ |v_{ak}[\lambda]|^4 \right\} + 4(r_{xak}[0]r_{vak}[0])} \tag{165}$$

$$|g_{k0}| = \sqrt{\frac{E \left\{ |x_{ak}[\lambda]|^4 \right\} + r_{xak}[0]r_{vak}[0]}{E \left\{ |x_{ak}[\lambda]|^4 \right\} + E \left\{ |v_{ak}[\lambda]|^4 \right\} + 4(r_{xak}[0]r_{vak}[0])}}. \tag{166}$$

Os valores esperados de quarta ordem na equação (166) podem ser representados da seguinte maneira.

$$E \left\{ |x_{ak}[\lambda]|^4 \right\} = E \left\{ (x_{ak}[\lambda]x_{ak}^*[\lambda])^2 \right\} \tag{167}$$

$$= E \left\{ \left(x_k^2[\lambda] + \tilde{x}_k^2[\lambda] \right)^2 \right\} \tag{168}$$

$$= E \left\{ x_k^4[\lambda] + \tilde{x}_k^4[\lambda] + 2x_k^2[\lambda]\tilde{x}_k^2[\lambda] \right\} \tag{169}$$

$$= E \left\{ x_k^4[\lambda] \right\} + E \left\{ \tilde{x}_k^4[\lambda] \right\} + 2E \left\{ x_k^2[\lambda]\tilde{x}_k^2[\lambda] \right\} \tag{170}$$

Assumindo que $x_k[\lambda]$ e $v_k[\lambda]$ são processos gaussianos (ver Figuras 34 e 33) e substituindo as equações (115–119) na equação (170):

$$E \left\{ |x_{ak}[\lambda]|^4 \right\} = 3\sigma_x^4 + 3\sigma_x^4 + 2\sigma_x^4 = 8\sigma_x^4. \quad (171)$$

Analogamente,

$$E \left\{ |v_{ak}[\lambda]|^4 \right\} = 3\sigma_v^4 + 3\sigma_v^4 + 2\sigma_v^4 = 8\sigma_v^4. \quad (172)$$

Além disso, segundo as equações (91) e (97),

$$r_{xak}[0] = 2r_{xk}[0] = 2\sigma_{xk} \quad (173)$$

$$r_{vak}[0] = 2r_{vk}[0] = 2\sigma_{vk}. \quad (174)$$

Substituindo as equações de (171) a (174) na equação (166), obtém-se

$$|g_{k0}| = \sqrt{\frac{8\sigma_{xk}^4 + 4\sigma_{xk}^2\sigma_{vk}^2}{8\sigma_{xk}^4 + 8\sigma_{vk}^4 + 16\sigma_{xk}^2\sigma_{vk}^2}} \quad (175)$$

$$|g_{k0}| = \sqrt{\frac{8\frac{\sigma_{xk}^4}{\sigma_{vk}^4} + 4\frac{\sigma_{xk}^2}{\sigma_{vk}^2}}{8\frac{\sigma_{xk}^4}{\sigma_{vk}^4} + 8 + 16\frac{\sigma_{xk}^2}{\sigma_{vk}^2}}} \quad (176)$$

$$|g_{k0}| = \sqrt{\frac{2\xi_k^2 + \xi_k}{2\xi_k^2 + 4\xi_k + 2}}. \quad (177)$$

em que $\xi_k = \frac{\sigma_{xk}^2}{\sigma_{vk}^2}$, é a relação sinal-ruído local.

Vale notar que, no caso do filtro de um coeficiente, a função custo depende apenas do módulo de g_k .

$$\begin{aligned} J &= E \left\{ \left(|\hat{x}_{ak}[\lambda]|^2 - |x_{ak}[\lambda]|^2 \right)^2 \right\} \\ J &= E \left\{ \left(|g_k y_{ak}[\lambda]|^2 - |x_{ak}[\lambda]|^2 \right)^2 \right\} \\ J &= E \left\{ \left(|g_k|^2 |y_{ak}[\lambda]|^2 - |x_{ak}[\lambda]|^2 \right)^2 \right\} \end{aligned} \quad (178)$$

Assim, pode-se considerar o ganho como sendo real e positivo, de forma que $|g_{k0}| = g_{k0}$.

Para verificar se g_k é um ponto de mínimo, analisa-se a segunda derivada da equação 178 em relação a g_k . Utilizando a equação ((159)) para o filtro com um coeficiente real:

$$\frac{d^2 J}{dg_k^2} = \frac{d}{dg_k} \frac{dJ}{dg_k} \quad (179)$$

$$= \frac{d}{dg_k} \left[2g_k^3 E \left\{ |x_{ak}[\lambda]|^4 \right\} + 2g_k^3 E \left\{ |v_{ak}[\lambda]|^4 \right\} + 4g_k^3 (2r_{xak}[0]r_{vak}[0]) - 2g_k E \left\{ |x_{ak}[\lambda]|^4 \right\} - 2g_k r_{xak}[0]r_{vak}[0] \right] \quad (180)$$

$$= \frac{d}{dg_k} \left[2g_k^3 (E \left\{ |x_{ak}[\lambda]|^4 \right\} + E \left\{ |v_{ak}[\lambda]|^4 \right\} + 4(r_{xak}[0]r_{vak}[0])) - 2g_k (E \left\{ |x_{ak}[\lambda]|^4 \right\} + r_{xak}[0]r_{vak}[0]) \right] \quad (181)$$

$$= 6g_k^2 (E \left\{ |x_{ak}[\lambda]|^4 \right\} + E \left\{ |v_{ak}[\lambda]|^4 \right\} + 4(r_{xak}[0]r_{vak}[0])) - 2 (E \left\{ |x_{ak}[\lambda]|^4 \right\} + r_{xak}[0]r_{vak}[0]). \quad (182)$$

Substituindo (164) em (182):

$$\left. \frac{d^2 J}{dg_k^2} \right|_{g_k=g_{k0}} = 6 \frac{E \left\{ |x_{ak}[\lambda]|^4 \right\} + r_{xak}[0]r_{vak}[0]}{E \left\{ |x_{ak}[\lambda]|^4 \right\} + E \left\{ |v_{ak}[\lambda]|^4 \right\} + 4(r_{xak}[0]r_{vak}[0])} \times \left(E \left\{ |x_{ak}[\lambda]|^4 \right\} + E \left\{ |v_{ak}[\lambda]|^4 \right\} + 4(r_{xak}[0]r_{vak}[0]) \right) - 2 (E \left\{ |x_{ak}[\lambda]|^4 \right\} + r_{xak}[0]r_{vak}[0]) \quad (183)$$

$$= 6 (E \left\{ |x_{ak}[\lambda]|^4 \right\} + r_{xak}[0]r_{vak}[0]) - 2 (E \left\{ |x_{ak}[\lambda]|^4 \right\} + r_{xak}[0]r_{vak}[0]) \quad (184)$$

$$= 4 (E \left\{ |x_{ak}[\lambda]|^4 \right\} + r_{xak}[0]r_{vak}[0]). \quad (185)$$

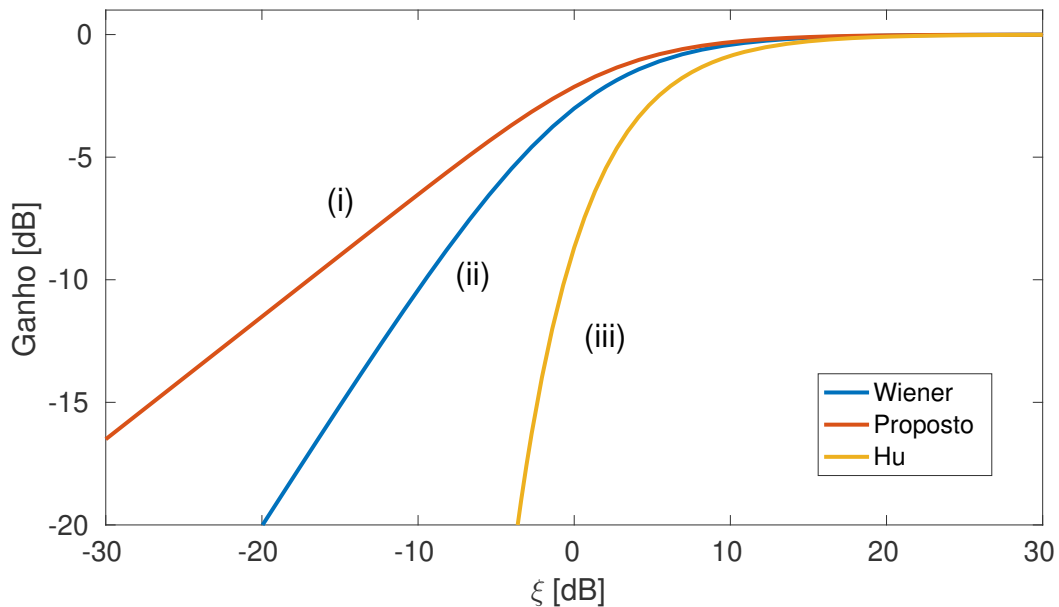
Com base na equação (185), e lembrando que $E \left\{ |x_{ak}[\lambda]|^4 \right\} \geq 0$ e $r_{xak}[0]r_{vak}[0] \geq 0$ pode-se verificar que

$$\left. \frac{d^2 J}{dg_k^2} \right|_{g_k=g_{k0}} \geq 0, \quad (186)$$

e, conseqüentemente, o ponto g_{k0} é um mínimo de J .²

A Figura 35 mostra a função de ganho do filtro proposto, g_{k0} , em função da SNR do canal. Para fins de comparação, também são mostradas na figura, a função de supressão do filtro de Wiener (CHEN *et al.*, 2006) (Equação (7)) e a proposta por Yi Hu *et al.* (2007) (Equação (11)). Esta última é uma máscara desenvolvida heurísticamente para a redução de ruído em implantes cocleares numa arquitetura similar à da máscara proposta neste capítulo da tese, na qual um ganho é aplicado ao envelope do sinal em cada canal, a cada quadro temporal.

Figura 35 – Função de supressão proposta (i), como função da SNR local, ξ , em dB. O filtro de Wiener (ii) e a função proposta por Yi Hu *et al.* (2007) (iii), também são mostradas para comparação.



4.3 SIMULAÇÕES NUMÉRICAS

Simulações numéricas foram realizadas como uma avaliação preliminar da possível aplicabilidade do estimador de envelope temporal para implantes cocleares.

² Na ausência de sinal de fala, $E\{|x_{ak}[\lambda]|^4\} = 0$ e $r_{xak}[0]r_{vak}[0] = 0$ e, conseqüentemente, $g_{k0} = 0$ e $\left.\frac{d^2J}{dg_k^2}\right|_{g_k=g_{k0}} = 0$. Esta condição não é suficiente para confirmar g_{k0} como um ponto de mínimo. Entretanto, nessas condições, pode-se verificar que $\left.\frac{d^3J}{dg_k^3}\right|_{g_k=g_{k0}} = 0$ e $\left.\frac{d^4J}{dg_k^4}\right|_{g_k=g_{k0}} > 0$. E esta condição permite afirmar que mesmo quando $g_k = g_{k0} = 0$, o filtro proposto corresponde a um ponto de mínimo.

4.3.1 Metodologia

Para avaliar o filtro proposto, foram usadas 720 frases foneticamente balanceadas do banco de dados IEEE *corpus* (IEEE, 1969).

Foi utilizado como interferência $v[n]$ um ruído de balbuciação formado artificialmente pela soma de quatro sequências do sinal *International Speech Test Signal* (ISTS – Sinal de Teste de Fala Internacional) (HOLUBE *et al.*, 2010) aleatoriamente deslocadas no tempo. O sinal ISTS é um sinal desenvolvido para a norma IEC 60118-15 referente a instrumentos auditivos e tem espectro e modulação temporal similares à da fala. As misturas utilizadas tiveram relações sinal-ruído de: -20dB , -15dB , -10dB , -5dB , 0dB , $+10\text{dB}$ e $+20\text{dB}$.

Os sinais de áudio foram introduzidos em um banco de filtros tipo *gammatone* (JOHANNESMA, 1972), o mesmo mostrado na Figura 14b, com 22 canais. O sinal em cada saída do banco de filtros foi tratado separada e independentemente.

Nas principais estratégias utilizadas em implantes cocleares, CIS e ACE, a taxa de estimulação em um canal é fixa com geralmente entre 900 e 1500 pulsos/s (RAO; RAMAIAH; KUMAR, 2012; WILSON; DORMAN, 2007; WOUTERS; MCDERMOTT; FRANCA, 2015; ZENG; POPPER; FAY, 2004). Assim, a cada trecho de 1 ms, foram calculados a SNR do canal ξ_k e o ganho g_k . Em seguida, o ganho g_k foi aplicado ao respectivo trecho do sinal ruidoso $y_k[n]$. O sinal resultante é denominado $\hat{x}_{kg}[n]$. Nesse experimento, tem-se acesso ao sinal de fala x_k e de ruído v_k separadamente, o que permite o cálculo do ganho g_k ideal.

Para fim de comparação, o sinal ruidoso também foi processado pelo filtro de Wiener ideal, w_k , resultando no sinal $\hat{x}_{kw}[n]$. O filtro de Wiener no domínio do tempo (CHEN *et al.*, 2006) é definido por

$$\mathbf{w}_k = \left[\mathbf{I} - \mathbf{R}_y^{-1} \mathbf{R}_v \right] \mathbf{u}_1, \quad (187)$$

em que \mathbf{I} é a matriz identidade e $\mathbf{u}_1^T = [1 \ 0 \ \dots \ 0]^T$.

O filtro de um coeficiente é, portanto, dado por

$$w_k = 1 - \frac{\sigma_{v_k}^2}{\sigma_{y_k}^2}, \quad (188)$$

em que σ^2 representa a variância do sinal. Rearranjando (188):

$$w_k = \frac{\sigma_{x_k}^2}{\sigma_{x_k}^2 + \sigma_{v_k}^2} \quad (189)$$

$$w_k = \frac{\xi_k}{\xi_k + 1}. \quad (190)$$

Com o objetivo de avaliar o desempenho dos filtros foi calculado o erro quadrático médio (MSE) entre os sinais limpo e processado, representado por ϵ_g , para a estimação realizada com o filtro g_k , e ϵ_w , para a estimação realizada com o filtro w_k . Também foi calculado o MSE entre os envelopes desses sinais, ϵ_g^{env} e ϵ_w^{env} .

$$\epsilon_g(k) = \frac{1}{N} \sum_{n=1}^N (\hat{x}_{kg}[n] - x_k[n])^2 \quad (191)$$

$$\epsilon_w(k) = \frac{1}{N} \sum_{n=1}^N (\hat{x}_{kw}[n] - x_k[n])^2 \quad (192)$$

$$\epsilon_g^{\text{env}}(k) = \frac{1}{N} \sum_{n=1}^N (|\hat{x}_{akg}[n]| - |x_{ak}[n]|)^2 \quad (193)$$

$$\epsilon_w^{\text{env}}(k) = \frac{1}{N} \sum_{n=1}^N (|\hat{x}_{akw}[n]| - |x_{ak}[n]|)^2 \quad (194)$$

Como o processamento em cada sub-banda k é independente das outras sub-bandas, cada frase do *corpus* resulta em 22 valores de erro para cada uma das equações (191) a (194), totalizando 15840 amostras.

Os erros relativos entre o processamento com o filtro proposto (g_k) e o filtro de Wiener (w_k) são dados pelas equações a seguir.

$$\epsilon_{\text{rel}} = 10 \log_{10} \frac{\epsilon_g}{\epsilon_w} \quad (195)$$

$$\epsilon_{\text{rel}}^{\text{env}} = 10 \log_{10} \frac{\epsilon_g^{\text{env}}}{\epsilon_w^{\text{env}}} \quad (196)$$

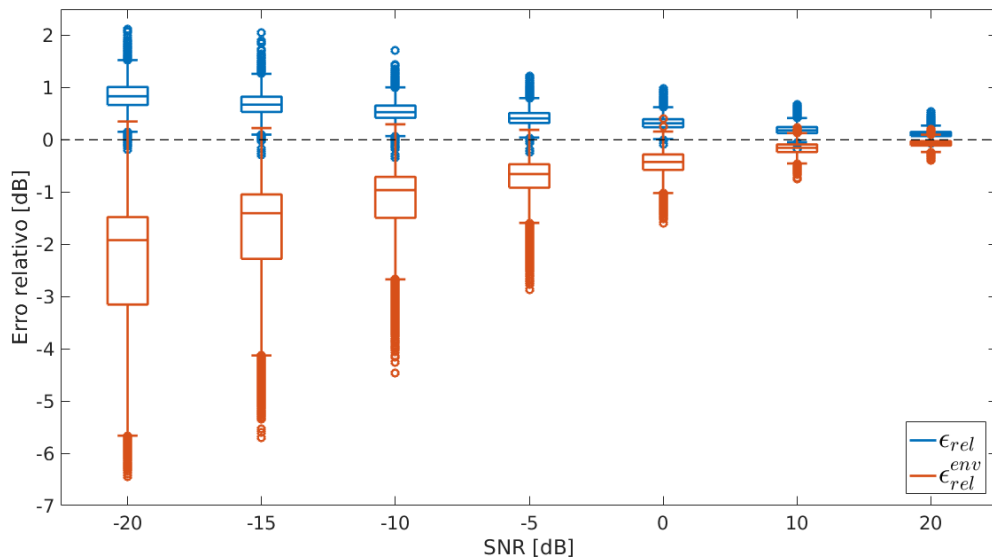
A métrica SRMR-CI normalizada também foi calculada para os sinais processados com o filtro proposto, o filtro de Wiener e pela função de supressão de Yi Hu *et al.* (2007), de forma a prever o impacto na avaliação subjetiva de inteligibilidade. Para tal, utilizou-se a equação (61) para o cálculo do percentual de inteligibilidade, da mesma maneira que no capítulo anterior.

4.3.2 Resultados e Discussão

A Figura 36 mostra os resultados das simulações com sinais do banco de falas balanceadas. Nela pode-se ver que, para misturas com predominância de fala (SNR > 0dB), tanto os erros relativos entre sinais quanto os erros relativos entre envelopes são próximos de 0dB, ou seja, o desempenho do filtro proposto é similar ao do filtro de

Wiener. Isto é coerente, uma vez que para relações sinal-ruído elevadas, os ganhos dos dois filtros são similares e próximos de 1.

Figura 36 – Erro relativo ϵ_{rel} entre os sinais estimados e real e, também, entre os envelopes dos sinais, ϵ_{rel}^{env} .



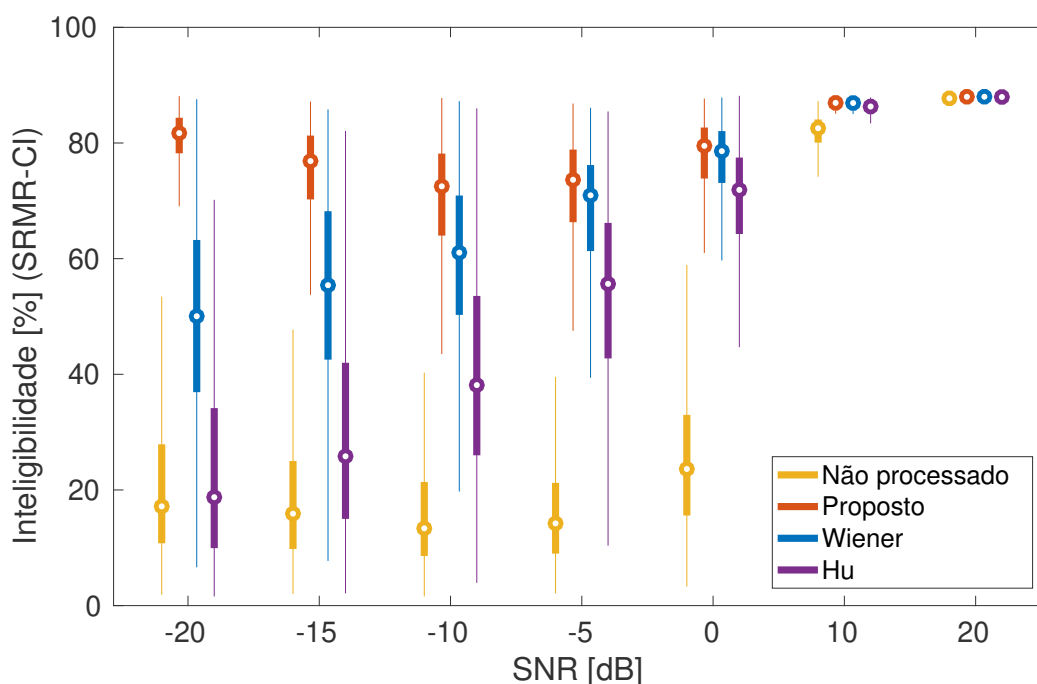
Por outro lado, quanto mais degradado o sinal de entrada, a utilização do filtro proposto resulta em um maior erro em relação ao sinal desejado, e menor erro em relação ao envelope desejado, em comparação ao de filtro Wiener. Esse resultado está em consonância com a teoria, dado que o filtro proposto é projetado para minimizar o erro quadrático médio entre os quadrados dos envelopes, enquanto que o filtro de Wiener minimiza o erro quadrático médio entre os sinais.

Os resultados de inteligibilidade estimada a partir da métrica SRMR-CI são mostrados em formato de diagramas de caixa na Figura 37. Os círculos indicam as medianas, as barras referem-se ao primeiro e terceiro quartis, e os bigodes aos limites superior e inferior. Valores discrepantes (*outliers*) são omitidos para uma melhor legibilidade.

O gráfico da Figura 37 mostra que, para o sinal não processado, a inteligibilidade decresce consideravelmente para valores $SNR \leq 0$. O processamento com técnicas de redução de ruído mitiga esse efeito, sendo que o desempenho do filtro proposto é o melhor dentre todos os testados, seguido pelo filtro de Wiener e o filtro de Hu, nessa ordem. Desta forma, os resultados indicam que o filtro proposto tem potencial para aumentar a inteligibilidade proporcionada pelo filtro de Wiener, ou a função de supressão de Hu, sobretudo para sinais com SNR negativa.

Vale notar que para SNRs de -15 dB e -20 dB, o valor da métrica SRMR-CI normalizada tanto para o sinal não processado quanto para o filtro proposto apresenta um aumento em relação ao seu valor em SNR de -10 dB. Isso pode ser um indicativo

Figura 37 – Diagramas de caixa da inteligibilidade calculada a partir da métrica SRMR-CI em função da SNR. Os círculos indicam as medianas, as barras o primeiro e terceiro quartis, e os bigodes os limites superior e inferior. Valores discrepantes (*outliers*) são omitidos. Da esquerda para a direita: sinais não processados (amarelo) e processados com o filtro proposto (laranja), com o filtro de Wiener (azul) e com a função de Hu (roxo).



de que a métrica não seja válida para sinais muito degradados ($SNR < -10$ dB). Efetivamente, em Santos e Falk (2014), a métrica é ajustada para valores experimentais obtidos com sinais de SNR entre -5 dB e $+10$ dB. Aqui esses limites foram extrapolados arbitrariamente, com o objetivo de verificar qualitativamente o comportamento das diferentes técnicas de redução de ruído. Isto posto, os resultados deste experimento, sobretudo para $SNR < 0$, motivam a realização de uma avaliação mais concreta e quantitativa do método proposto através de experimentos psicoacústicos, o que será abordado na seção 4.4.

A utilização do filtro de Wiener em um sinal ruidoso $y[n]$, leva a uma estimativa do sinal de fala $\hat{x}[n]$ com o menor erro quadrático médio em relação a $x[n]$. No domínio da frequência, tem-se:

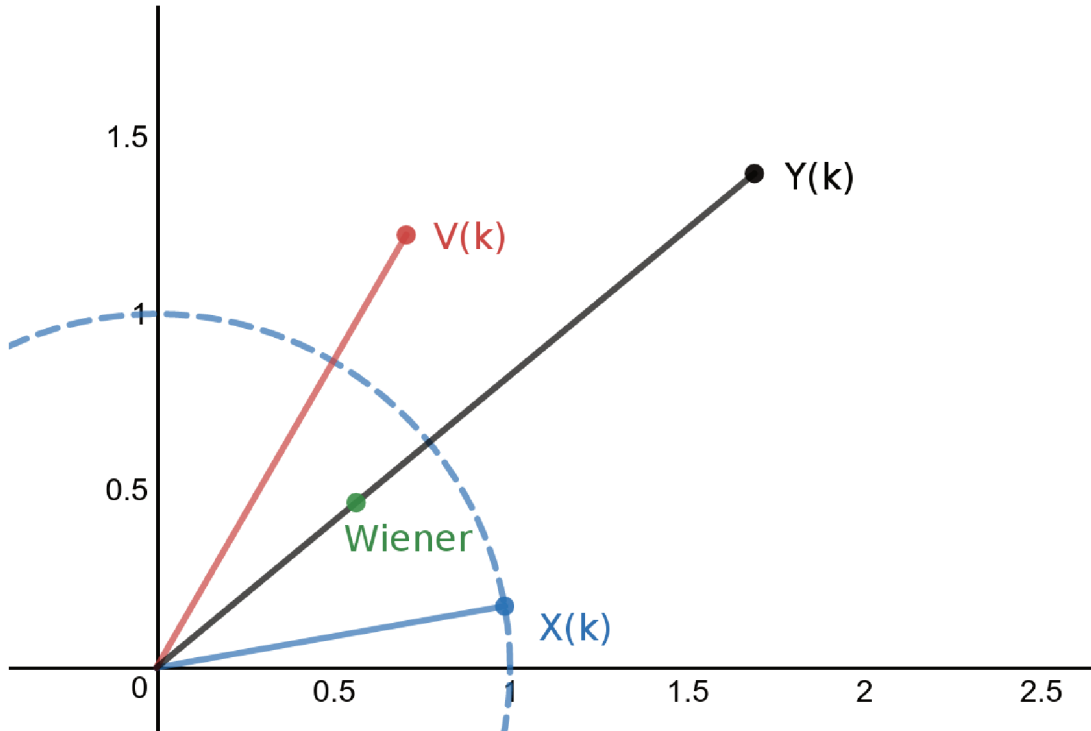
$$\hat{X}(k) = W(k) Y(k), \quad (197)$$

em que $\hat{X}(k)$ é a estimativa do espectro da fala, $Y(k)$ é o espectro do sinal ruidoso e $W(k) = \frac{\xi(k)}{\xi(k)+1}$ é a máscara de Wiener.

Como a máscara $W(k)$ é constituída apenas por números reais, apenas o módulo de $Y(k)$ é alterado, e a fase continua a mesma, ou seja ($\phi_{\hat{X}(k)} = \phi_{Y(k)}$). Desta

forma, a solução de mínimo erro quadrático médio não necessariamente resulta em $|\hat{X}(k)| = |X(k)|$, como ilustrado pela Figura 38.

Figura 38 – Fasores $Y(k)$, $X(k)$ e $V(k)$, para uma mistura arbitrária a $SNR = -3\text{dB}$. A estimativa de $X(k)$ usando o filtro de Wiener é mostrada em verde. Pode-se verificar que o filtro de Wiener não é necessariamente aquele que leva a melhor estimativa de $|X(k)|$.



O filtro g_{k0} , aqui proposto, se diferencia por fazer uma estimativa do envelope temporal do sinal $x_k[n]$, no domínio do tempo. No entanto, no domínio da frequência, encontram-se na literatura alguns estimadores de magnitude $|X(k)|$, geralmente derivados a partir de uma abordagem bayesiana (MMSE, log MMSE, e outras variações – ver Loizou (2013), capítulo 7, para maiores detalhes) .

Dentre esses estimadores de magnitude, tem-se como principal referência o MMSE proposto por Ephraim e Malah (1984), em cuja derivação considera-se que os coeficientes da transformada de Fourier da fala e do ruído têm distribuição gaussiana. A máscara resultante, para cada frequência k , é dada por

$$G_{\text{MMSE}}(\xi_k, v_k) = \frac{\sqrt{\pi}}{2} \frac{\sqrt{v_k}}{\gamma_k} \exp\left(-\frac{v_k}{2}\right) \left[(1 + v_k) I_0\left(\frac{v_k}{2}\right) + v_k I_1\left(\frac{v_k}{2}\right) \right], \quad (198)$$

em que v_k é definido por

$$v_k = \frac{\xi_k}{1 + \xi_k} \gamma_k, \quad (199)$$

e a SNR *a posteriori*, γ_k , e a SNR *a priori*, ξ_k , por

$$\gamma_k = \frac{|Y(k)|^2}{\sigma_V^2(k)} \quad (200)$$

$$\xi_k = \frac{\sigma_X^2(k)}{\sigma_V^2(k)}, \quad (201)$$

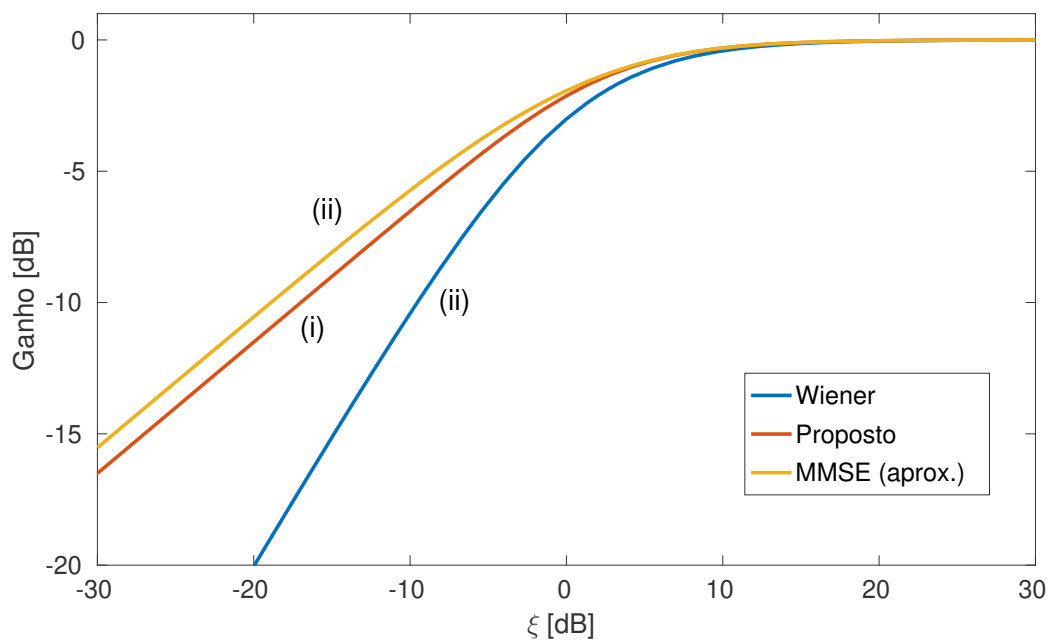
sendo $\sigma_X^2(k)$ e $\sigma_V^2(k)$, as variâncias da fala e do ruído, respectivamente.

O ganho do MMSE depende, assim da SNR *a priori*, e também da SNR *a posteriori*. Para comparar esse método no domínio da frequência com o proposto neste capítulo, faz-se a seguinte aproximação, para que fique dependente apenas de ξ_k .

$$\gamma_k \approx \xi_k + 1 \quad (202)$$

A Figura 39 mostra os ganhos do filtro de envelope do capítulo 4 com um coeficiente, da aproximação do MMSE e do filtro de Wiener.

Figura 39 – Ganho do filtro proposto (i) em comparação com o do filtro de Wiener (ii) e o de uma aproximação do estimador de magnitude MMSE (iii).



Os ganhos do filtro de envelope e da aproximação do MMSE são muito similares e se distinguem notavelmente do filtro de Wiener, com uma supressão menor para valores baixos de ξ_k .

Vale ressaltar que a definição de ξ_k , apesar de representar uma grandeza similar, não é a mesma para o filtro de envelope e para o MMSE. Para o primeiro, ξ_k

é a relação entre as variâncias da fala e do ruído em um sinal de banda estreita no domínio do tempo, tomado na saída k de um banco de filtros passa-faixa. Por outro lado, para o MMSE, ξ_k é a relação entre as variâncias dos coeficientes do *bin* k das transformadas de Fourier da fala e ruído.

De qualquer maneira, funções de supressão do ruído com uma transição suave, com ganhos mais elevados que o filtro de Wiener na região de $\xi_k < 0$ dB, levam, aparentemente, a melhores estimativas da magnitude ou envelope do sinal de fala $x_k[n]$. Isso é corroborado com os resultados do capítulo 3, em que as máscaras mais suaves (com parâmetro $\gamma < 0$ dB), obtiveram melhores desempenhos nos experimentos de inteligibilidade.

Os resultados apresentados nesta seção indicam que o filtro proposto deve resultar em uma melhor inteligibilidade de fala para usuários de IC, quando comparado com o filtro de Wiener e a função de supressão de Hu. Entretanto, o efetivo desempenho das funções de supressão testadas só pode ser aferido através de experimentos psicoacústicos.

4.4 EXPERIMENTOS PSICOACÚSTICOS

Como visto na seção anterior, o filtro proposto leva a menores erros de estimação do envelope que o filtro de Wiener. Ademais, os resultados com a métrica SRMR-CI corroboram com a hipótese de que a inteligibilidade de implantados será maior com uso do estimador de envelope em comparação com o filtro de Wiener.

Para testar essa hipótese, experimentos psicoacústicos foram realizados com normouvintes e usuários de implante cocleares. Dois experimentos, denominados Experimento I e Experimento II, foram projetados, com o objetivo de medir, respectivamente, a inteligibilidade e a qualidade da fala processada com o filtro proposto e com o filtro de Wiener. Os materiais e métodos utilizados são descritos nas próximas seções.

4.4.1 Interface de pesquisa em implantes cocleares da UTD

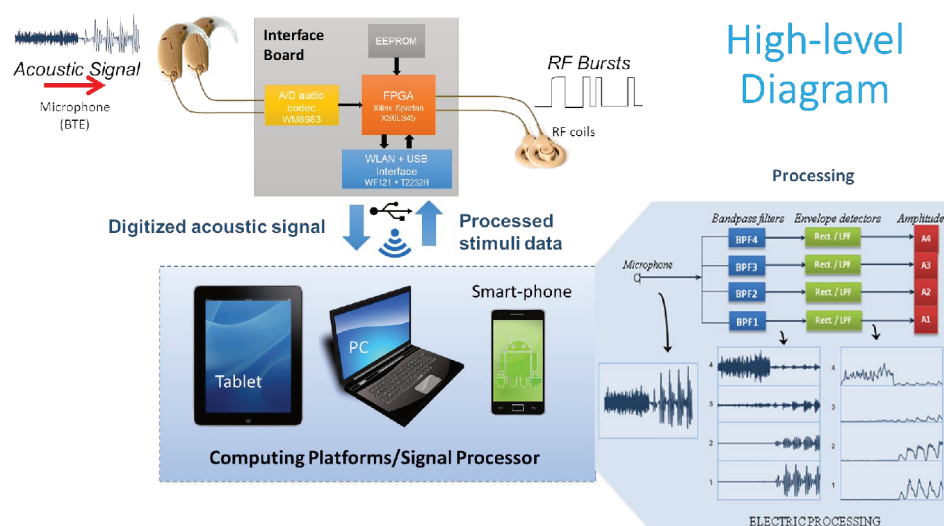
Implantes cocleares comerciais possuem código proprietário e permitem apenas o ajuste de poucos parâmetros para o conforto do implantado. As empresas fabricantes de ICs dispõem de interfaces de pesquisa³ para o desenvolvimento de seus produtos. Esses dispositivos se comunicam diretamente com a parte interna dos implantes, permitindo aos pesquisadores a implementação de técnicas de processamento de sinais e novas estratégias de estimulação. Contudo, essas interfaces não são disponíveis para a comunidade científica. São reservadas apenas a alguns poucos grupos de pesquisa

³ Interfaces de pesquisa são dispositivos que permitem a comunicação com a parte interna do implante. Desta forma possibilitam que sinais processados em outros ambientes (computador, telefone celular, entre outros) sejam usados para controlar os eletrodos já implantados, substituindo temporariamente o processador do usuário de IC.

diretamente ligados a uma ou mais dessas empresas. Além disso, impõem limitações consideráveis quanto à sua programação e utilização pelos cientistas (ALI *et al.*, 2013). Esta condição dificulta significativamente o desenvolvimento de novas tecnologias de processamento para ICs.

Neste contexto, pesquisadores da Universidade de Texas em Dalas (UTD) desenvolveram uma plataforma de pesquisas em implantes cocleares (ALI *et al.*, 2013) aberta. Um dispositivo que permite a comunicação direta com a parte interna do implante, através de uma antena de RF. O funcionamento da plataforma é ilustrado na Figura 40.

Figura 40 – Funcionamento da plataforma de pesquisa em implantes cocleares da UTD (ALI, 2016).



O sinal acústico é captado pelos microfones de um dispositivo tipo BTE sem as unidades de processamento de um aparelho auditivo, e em seguida, é condicionado e digitalizado. Uma FPGA faz o gerenciamento da informação dentro da interface. O sinal de entrada é direcionado, através de comunicação USB ou Wi-Fi, para uma unidade de processamento (computador, *tablet* ou *smartphone*), que pode ser programada com liberdade para realizar todo o processamento do sinal até a codificação em pulsos de corrente. A sequência de pulsos resultante é enviada de volta para a interface, onde é codificada de acordo com protocolos de comunicação do implante (dependente da fabricante) e transmitida, através da antena RF, à parte interna do implante, por fim estimulando a cóclea. Na versão atual, apenas o protocolo de comunicação com implantes da *Cochlear*[®] está implementado, mas existem planos para se incluir a comunicação com outros implantes no futuro. Para isso, bastará alterar a programação da FPGA, não sendo necessárias mudanças no *hardware*.

Com este dispositivo, pode-se realizar o processamento em tempo-real, em que o áudio captado pelo BTE é enviado à unidade de processamento em pacotes de

quadros (*frames*) de 8 ms. A interface pode igualmente ser utilizada em modo *bench-top*, quando a sequência de pulsos resultante de um áudio pré-processado é enviada ao usuário de implante coclear.

A plataforma de pesquisa possui duas entradas para BTEs e duas saídas para antenas RF, possibilitando a pesquisa para implantes bilaterais ou biauriculares. Além disso, existem duas portas de áudio, permitindo também estudos de estimulação eletroacústica (EAS). É possível implementar estratégias como aplicativos de celular, e, conseqüentemente, realizar experimentos do tipo *take-home trial*, em que o implantado usa o processamento a ser avaliado durante um período de tempo mais extenso (dias).

A programação é flexível e pode ser realizada em diversas linguagens, dentre as quais pode-se destacar MATLAB, C, C++, LabView e JAVA. Vale ressaltar que o único módulo fechado, ao qual não se tem acesso ao código e à reprogramação, é a FPGA. Isso se dá por motivo de segurança, pois esta programação impede que sejam aplicados ao paciente estímulos de corrente que lhe possam ser danosos (amplitudes elevadas, pulsos com desbalanceamento de cargas elétricas, por exemplo).

No final de do ano 2016, um exemplar da plataforma de pesquisas foi cedida pelos seus desenvolvedores da UTD ao grupo de pesquisas em ICs na UFSC⁴, estabelecendo-se um acordo de colaboração. Esta é a primeira unidade disponibilizada a um grupo de fora dos Estados Unidos da América. Detalhes técnicos sobre o dispositivo são apresentados no anexo C.

4.4.2 Sujeitos

Participaram dos experimentos seis normouvintes tendo declarado nunca terem apresentado problemas de audição ao longo da vida. Seus dados estão dispostos na Tabela 18.

Além disso, também participaram seis usuários de implante coclear, cujas informações pessoais estão colocadas na Tabela 19. Todos os voluntários usam implantes da marca Cochlear[®].

As pesquisas foram aprovadas no comitê de ética em pesquisa (CEP) sob os números CAAE: 56838816.7.0000.0121 / Parecer: 1.997.931, para usuários de implante coclear, e CAAE: 49741615.2.0000.0121 / Parecer: 1.291.583, para normouvintes.

4.4.3 Processamento de sinais

O conjunto de configurações individuais do implante é chamado de *mapa*. Ele contempla informações tais como a taxa de estimulação, a faixa de frequências e os

⁴ Grupo interdisciplinar em atividade desde o final de 2014, capitaneado por pesquisadores do Laboratório de Vibrações e Acústica (LVA) da engenharia mecânica. Conta com pesquisadores das áreas de engenharia mecânica, engenharia elétrica, fonoaudiologia e medicina, além das colaborações da empresa Wavetch Soluções Tecnológicas e da clínica de fonoaudiologia Otovida.

Tabela 18 – Dados dos voluntários normouvintes que participaram dos experimentos psicoacústicos.

Sujeito	Idade	Sexo
NH1	24	masculino
NH2	29	masculino
NH3	27	masculino
NH4	28	masculino
NH5	36	feminino
NH6	40	masculino
$\mu_j \pm \sigma_j$ §		30,7 ± 6

§ μ_j : média, σ_j : desvio padrão

Tabela 19 – Dados dos usuários de implante coclear que participaram dos experimentos psicoacústicos.

Sujeito	Sexo	Idade	# anos usando IC	Etiologia	Modelo do IC
IC1	Fem.	29	6	caxumba materna	CI422
IC2	Fem.	53	7	desconhecida	CI24RE
IC3	Fem.	32	4	desconhecida	CI422
IC4	Fem.	52	10	desconhecida	CI24RE
IC5	Masc.	50	5	TCE [†]	CI24RE
IC6	Fem.	39	3	surdez súbita	CI422
$\mu_j \pm \sigma_j$ §		43 ± 11	6 ± 2		

[†] Traumatismo Crânio-Encefálico.

§ μ_j : média, σ_j : desvio padrão

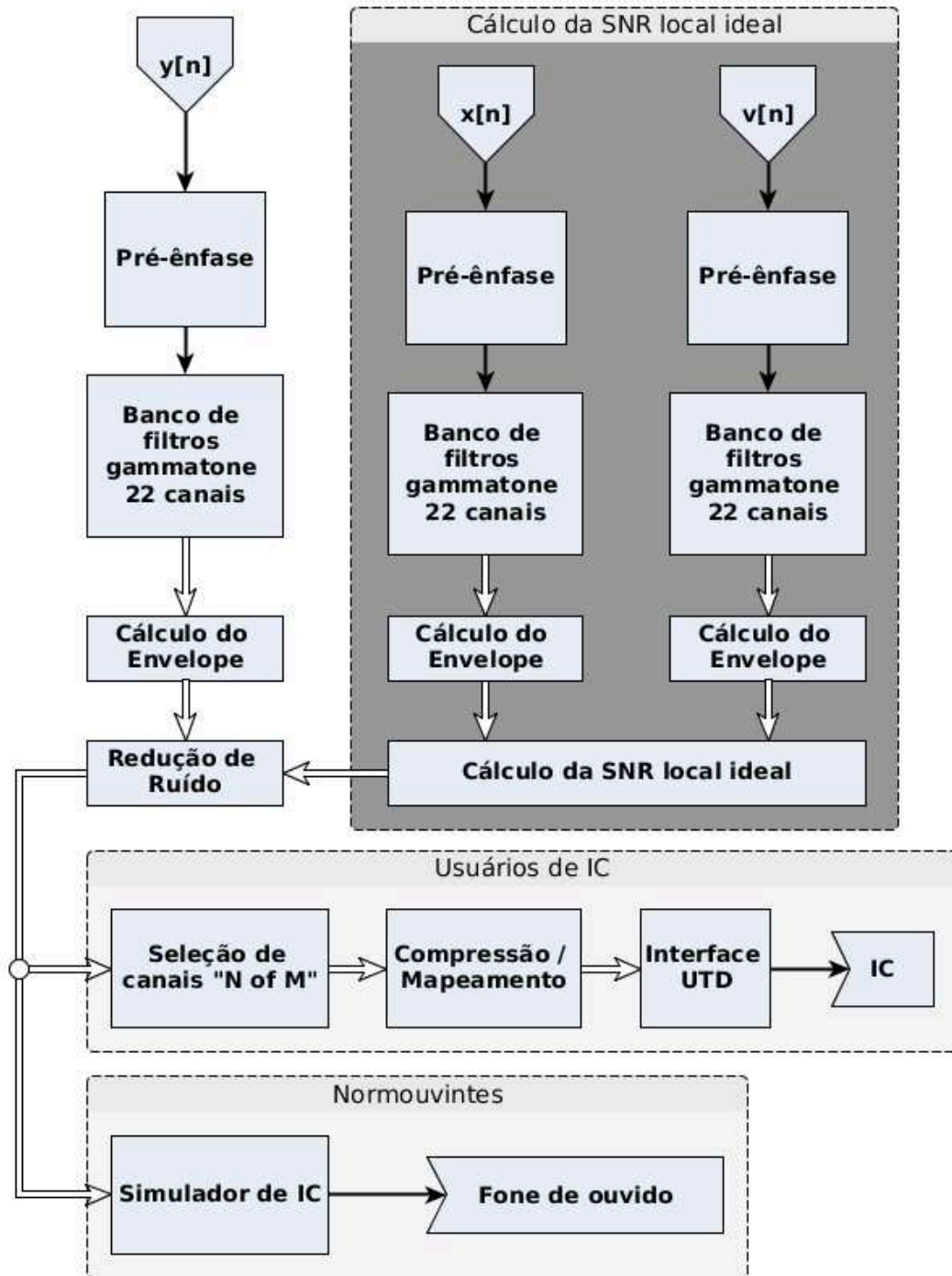
níveis limiares de estimulação (*T-level*) e de máxima corrente (*C-level*) de cada canal.

No início de cada sessão, o mapa do implante do voluntário sob avaliação foi carregado no sistema. No caso de normouvintes, um mapa padrão foi utilizado em um simulador do processamento do IC. Esse mapa padrão está detalhado no apêndice D.

A cadeia de processamento de sinais utilizada nos experimentos psicoacústicos é baseada na estratégia ACE, cujos códigos foram fornecidos juntamente da interface de pesquisas da Universidade do Texas em Dalas. O diagrama de blocos que ilustra esse processamento é mostrado na Figura 41. A frequência de amostragem dos sinais

é de 16kHz.

Figura 41 – Diagrama de blocos do processamento de sinais para os experimentos psicoacústicos. As setas negras de traço simples representam o tráfego de dados em série, enquanto que as brancas de traço duplo se referem ao tráfego em paralelo de M canais.



Primeiramente, o sinal $y[n]$ é processado por um filtro passa-altas, denominado

de filtro de pré-ênfase, cuja função de transferência é dada por (ALI, 2016):

$$H_{pe}(z) = \frac{0,5006 - 0,5006z^{-1}}{1 - 0,0012z^{-1}}. \quad (203)$$

Em seguida, o sinal é separado em sub-bandas através de um filtro *gammatone* de M canais, de acordo com o mapa do voluntário. O filtro *gammatone* utilizado possui, geralmente, $M = 22$ canais, com frequências centrais de 250,5Hz, 375,5Hz, 500,5Hz, 625,5Hz, 750,5Hz, 875,5Hz, 1000,5Hz, 1125,5Hz, 1250,5Hz, 1438Hz, 1688Hz, 1938Hz, 2188Hz, 2500,5Hz, 2875,5Hz, 3313Hz, 3813Hz, 4375,5Hz, 5000,5Hz, 5688Hz, 6500,5Hz, 7438Hz. Entretanto essa distribuição pode variar de acordo com o mapa do voluntário.

O processamento em cada canal ocorre de maneira independente entre si. O sinal na saída do banco de filtros tem sua envoltória temporal calculada utilizando-se a transformada de Hilbert de acordo com a equação (71). O envelope temporal é então dividido em blocos de tamanho e sobreposição (*overlap*) determinados de acordo com o mapa do voluntário. Cada bloco desses representa uma unidade tempo-frequência.

Paralelamente, o mesmo processamento é realizado com os sinais $x[n]$ e $v[n]$, de forma a se calcular ξ_k , a relação sinal-ruído ideal de cada unidade tempo-frequência. Assim, o ganho de redução de ruído é calculado pelas equações (177) ou (188).

Os envelopes do sinal $y_k[\lambda]$ são então multiplicados pelos ganhos de redução de ruído, resultando em envelopes estimados do sinal de fala $\hat{x}_k[\lambda]$. A partir desse ponto, o processamento segue caminhos diferentes para normouvintes e para implantados.

No primeiro caso, esses envelopes modificados são utilizados como entrada de um simulador de implantes cocleares adaptado de Bräcker e Schulte (2009). Esse simulador sintetiza um áudio considerando efeitos físicos da estimulação elétrica da cóclea, tais como o espalhamento de corrente na cóclea e o posicionamento dos eletrodos na cóclea. Esse simulador de implante coclear foi utilizado em estudos recentes (HU, H. *et al.*, 2018; LANGNER; JÜRGENS; JU, 2016; WILLIGES *et al.*, 2015). O áudio resultante é apresentado ao voluntário normouvinte através de fones de ouvido (Sennheiser modelo HD202).

No segundo caso, para usuários de implante coclear, os envelopes modificados passam por um bloco de seleção de canais, em que os N canais de maior amplitude, dentre um total de M , em cada *frame* temporal, são selecionados. Os sinais de envelope selecionados passam por um bloco de compressão logarítmica e em seguida são mapeados em uma escala de estimulação elétrica, determinada pelo mapa do voluntário. Na saída, têm-se o eletrodograma que é enviado ao implante através da interface de pesquisas em implantes cocleares (ALI *et al.*, 2013).

4.4.4 Experimento I – Inteligibilidade

No Experimento I foram usados como sinal de fala ($x[n]$) sinais de um *corpus* em português do Brasil (ALCAIM; SOLEWICZ; MORAES, 1992), que consiste em um conjunto de 20 listas foneticamente balanceadas, sendo 10 gravadas com voz masculina e outras 10 gravadas com voz feminina. Cada lista contém 10 frases.

O ruído ($v[n]$) de balbuciação baseado no sinal ISTS, descrito na seção 4.3.1 foi adicionado de forma a obter cinco condições de SNR global para avaliação, estando elas definidas entre -25dB e -5dB com passo de 5dB .

Sinais processados tanto com o filtro de Wiener (equação (188)) quanto com o estimador de envelopes proposto (equação (177)) são apresentados para os voluntários em todas as condições de SNR. Para cada uma dessas técnicas de redução de ruído e cada valor de SNR global, duas listas, uma com voz masculina e uma com voz feminina, são selecionadas aleatoriamente e sem reposição. Assim, em cada sessão completa do Experimento I é utilizado um total de 200 frases ($5\text{ SNR} \times 2\text{ listas} \times 10\text{ frases} \times 2\text{ métodos de redução de ruído}$).

Antes de iniciar-se o experimento, os voluntários, tanto normouvintes como implantados, ouviram frases de forma a permitir o ajuste do volume de maneira confortável. Essas frases eram sempre diferentes das utilizadas nos experimentos. Uma sessão do Experimento I tinha duração aproximada de 40 minutos, e os voluntários eram livres para interromper a sessão a qualquer momento para descanso. Todavia, todos os participantes optaram por realizar a sessão sem interrupções.

Para a realização do Experimento I, uma interface gráfica foi desenvolvida e é mostrada na Figura 42. O voluntário sob teste não teve acesso visual à interface gráfica durante o procedimento, sendo esta usada exclusivamente pelo pesquisador.

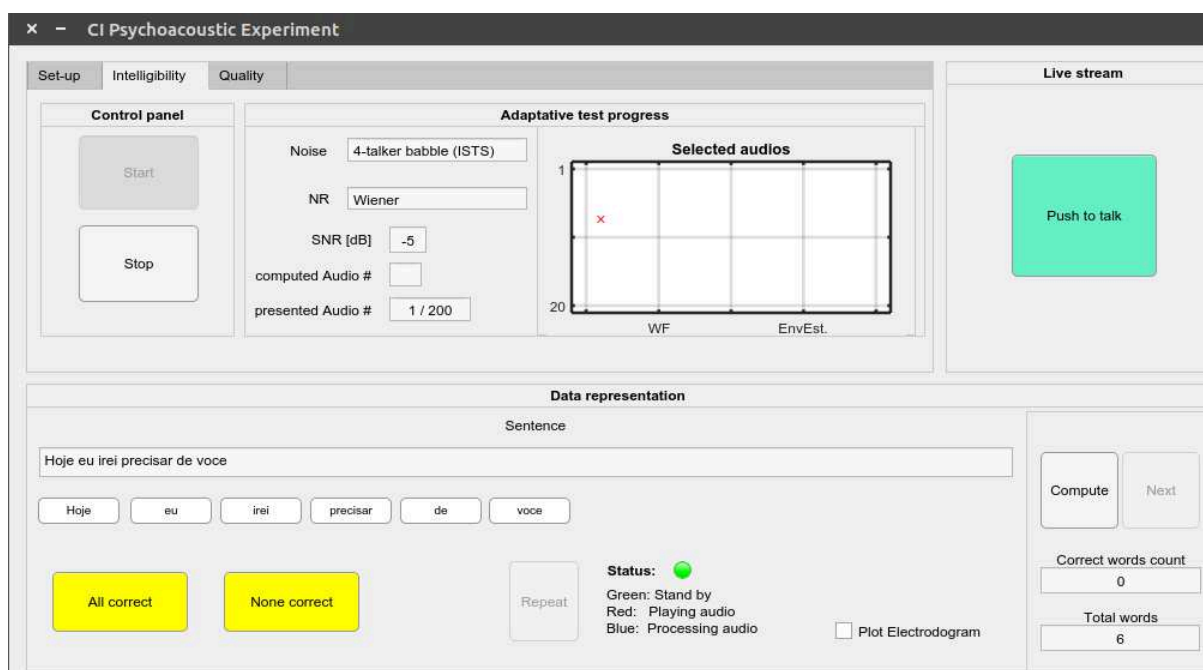
A cada iteração do procedimento, uma frase dentre o total de 200 era selecionada aleatoriamente, sem reposição. O voluntário pode ouvir cada frase apenas uma vez, e foi instruído a repetir em voz alta as palavras que havia compreendido. As palavras corretas foram marcadas e registradas pelo pesquisador através da interface gráfica. O processo foi repetido até que todas as frases tivessem sido reproduzidas.

A Figura 43 apresenta os resultados de inteligibilidade obtidos em cada lista (voz masculina e voz feminina) apresentada a cada voluntário normouvinte, referidos como NH_x .

Por sua vez, a Figura 44 mostra os diagramas de caixa reunindo os resultados individuais obtidos com normouvintes. Os círculos indicam as medianas, as barras o primeiro e terceiro quartis, e os bigodes os limites superior e inferior. Valores discrepantes (*outliers*) são representados por asteriscos (*). O sinal + indica os casos em que as medianas das duas distribuições (Wiener e proposto) não podem ser consideradas diferentes de acordo com o teste de Wilcoxon com significância $p < 0,05$.

Para todas as SNR utilizadas neste experimento, exceto para $\text{SNR} = -15\text{dB}$, os

Figura 42 – Interface gráfica para a realização do Experimento I. Apenas o pesquisador teve acesso a essa interface. As palavras corretas eram marcadas nos botões da parte inferior.



resultados do filtro proposto e do WF não foram significativamente diferentes. Vale ressaltar que o sinal na saída do simulador de IC é altamente distorcido, em relação ao áudio de fala sem presença de ruído. De fato, a inteligibilidade média desse sinal limpo é de aproximadamente 40%⁵. Isso posto, é compreensível que as diferenças entre o WF e o método proposto não sejam significativas.

Mesmo que haja alguma correlação entre resultados obtidos por normouvintes com simuladores de IC e aqueles obtidos com usuários de IC (STRYDOM *et al.*, 2011), há diferença significativa entre os resultados experimentais obtidos com esses dois grupos (KONING; MADHU; WOUTERS, 2015). Assim, os resultados com normouvintes apresentados devem ser considerados de maneira qualitativa, e não como tendo equivalência quantitativa com os obtidos em experimentos com usuários de IC.

Os resultados individuais do Experimento I realizado com usuários de IC, referidos como IC_x, são mostrados na Figura 45.

Os resultados na Figura 45 corroboram os relatos dos voluntários IC1, IC4 e IC5, apresentados ao final do experimento, sobre uma maior facilidade em se compreender vozes masculinas em relação à femininas. Em geral pode-se verificar que o método proposto resulta em melhores índices de inteligibilidade que os obtidos com o filtro de Wiener. A Figura 46, que evidencia essa observação, mostra os diagramas de caixa dos resultados agrupados de todos os usuários de IC. Analogamente à Figura

⁵ Medido com 4 normouvintes usando uma lista masculina e uma feminina. Os participantes desse teste, exceto um, não participaram dos Experimentos I e II.

Figura 43 – Resultados individuais do Experimento I com normouvintes. Em azul com marcadores vazados, o filtro de Wiener; em laranja com marcadores sólidos, o método proposto. As curvas contínuas com marcadores circulares se referem às listas com voz feminina, enquanto que as pontilhadas com marcadores triangulares, às listas com voz masculinas.

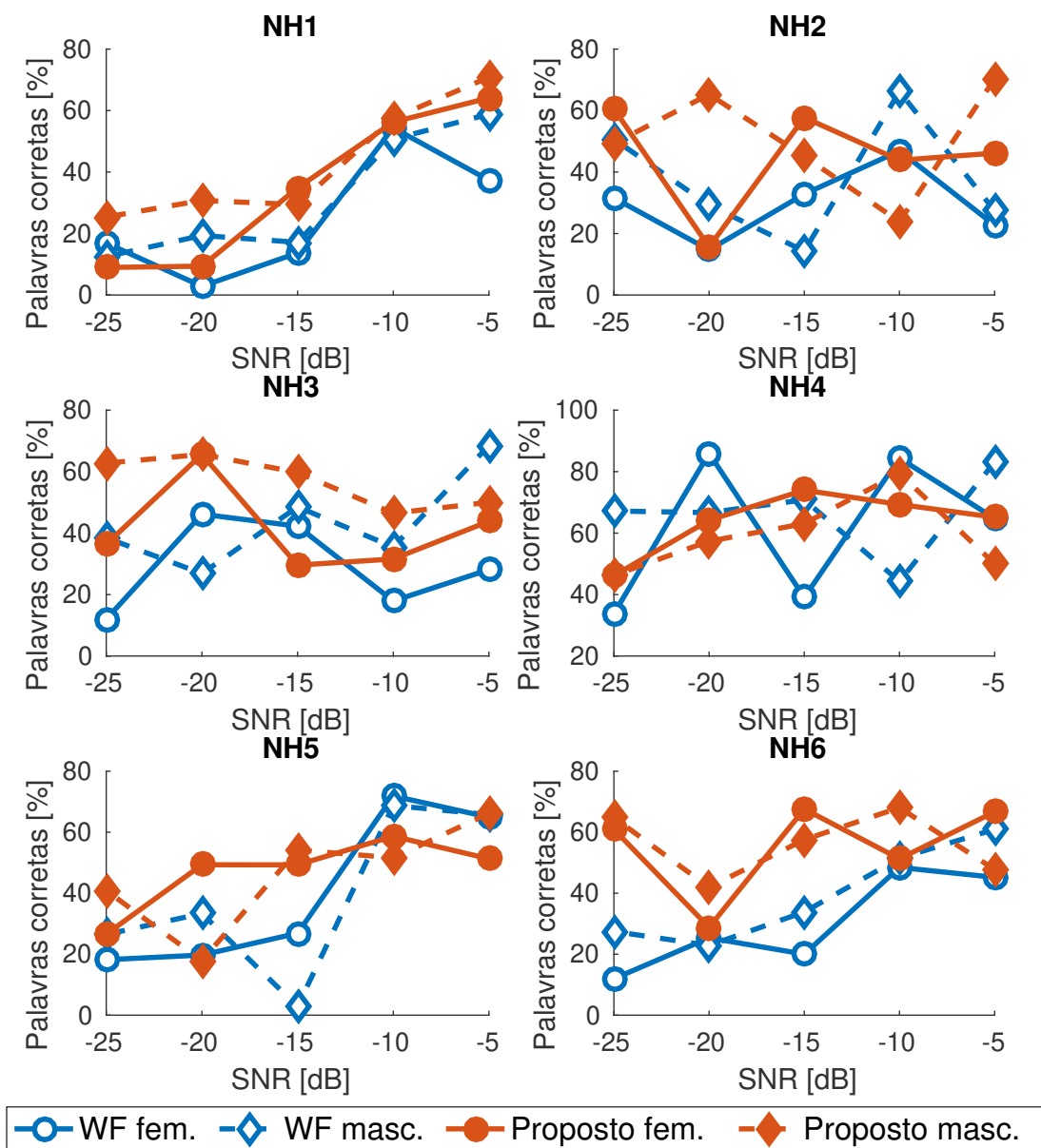
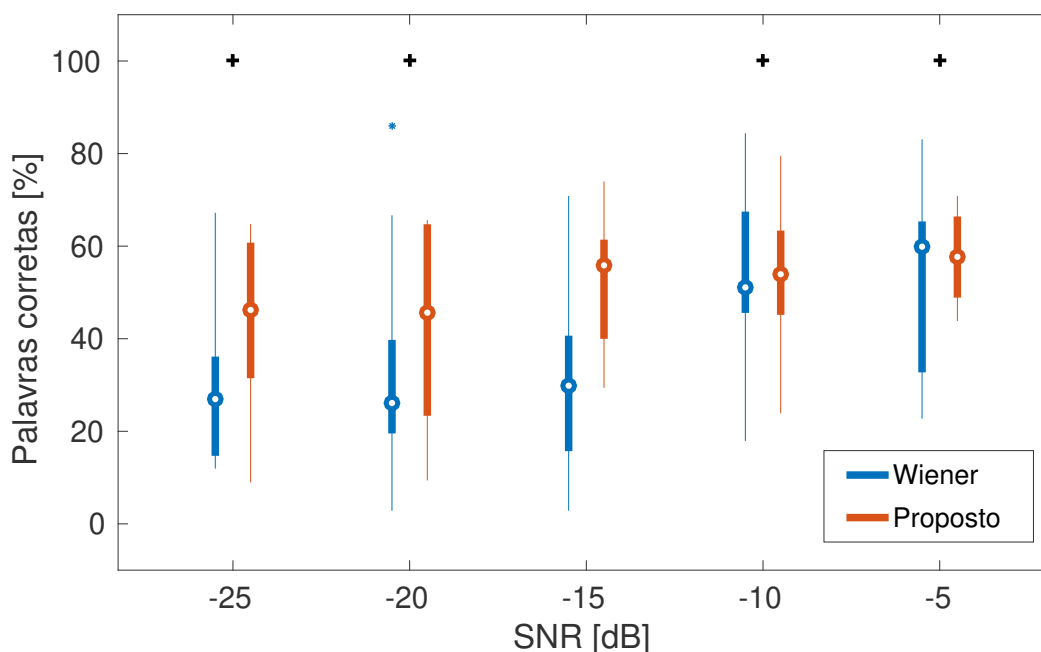


Figura 44 – Diagrama de caixa dos resultados do experimento I com normouvintes, agrupando todos os voluntários. Em azul, o filtro de Wiener; em laranja, o método proposto. Os círculos indicam as medianas, as barras o primeiro e terceiro quartis, e os bigodes os limites superior e inferior. Valores discrepantes (*outliers*) são representados por asteriscos (*). O sinal + indica que as medianas das duas distribuições (Wiener e proposto) não podem ser consideradas diferentes de acordo com o teste de Wilcoxon com $p < 0,05$.



44, o sinal + indica os casos em que as medianas das duas distribuições (Wiener e proposto) não podem ser consideradas diferentes de acordo com o teste de Wilcoxon com $p < 0,05$.

Na Figura 46 verifica-se que os resultados obtidos com o método proposto são maiores que os obtidos com o WF, sendo que para $SNR < -5$ dB, as medianas dos resultados são significativamente diferentes ($p < 0,05$), com uma vantagem clara do método proposto em relação ao WF. Nota-se que o desempenho no reconhecimento de frases processadas com o WF diminui com a redução da SNR, sendo praticamente nula para $SNR = -25$ dB. Por outro lado, quando os sinais são processados com o método proposto, a inteligibilidade mediana praticamente se mantém acima dos 70%.

Sinais processados com o filtro de Wiener foram excessivamente atenuados. Assim, para algumas unidades tempo-frequência, as estimativas da fala tiveram amplitudes baixas, ficando fora da faixa dinâmica determinada pelo mapa do usuário (T-level e C-level) e distorcendo o eletrograma. Os participantes relataram que as frases processadas com o filtro de Wiener tinham palavras cortadas, com degradação inversamente proporcional à SNR.

Por sua vez, o estimador de envelope proposto, que tem uma função de su-

Figura 45 – Resultados individuais do experimento I com usuários de implante coclear. Em azul com marcadores vazados, o filtro de Wiener; em laranja com marcadores sólidos, o método proposto. As curvas contínuas com marcadores circulares se referem às listas com voz feminina, enquanto que as pontilhadas com marcadores triangulares, às listas com voz masculinas.

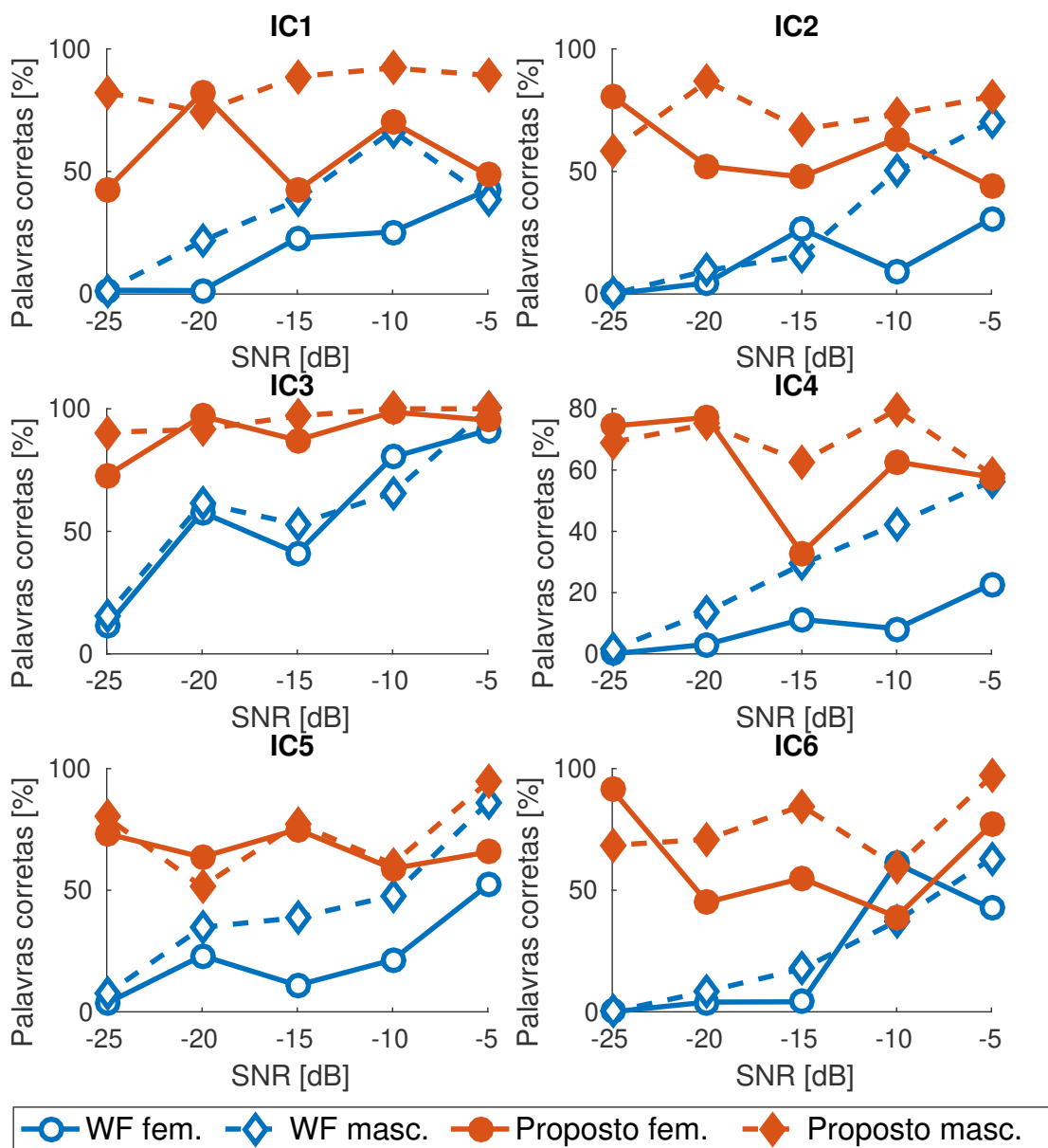
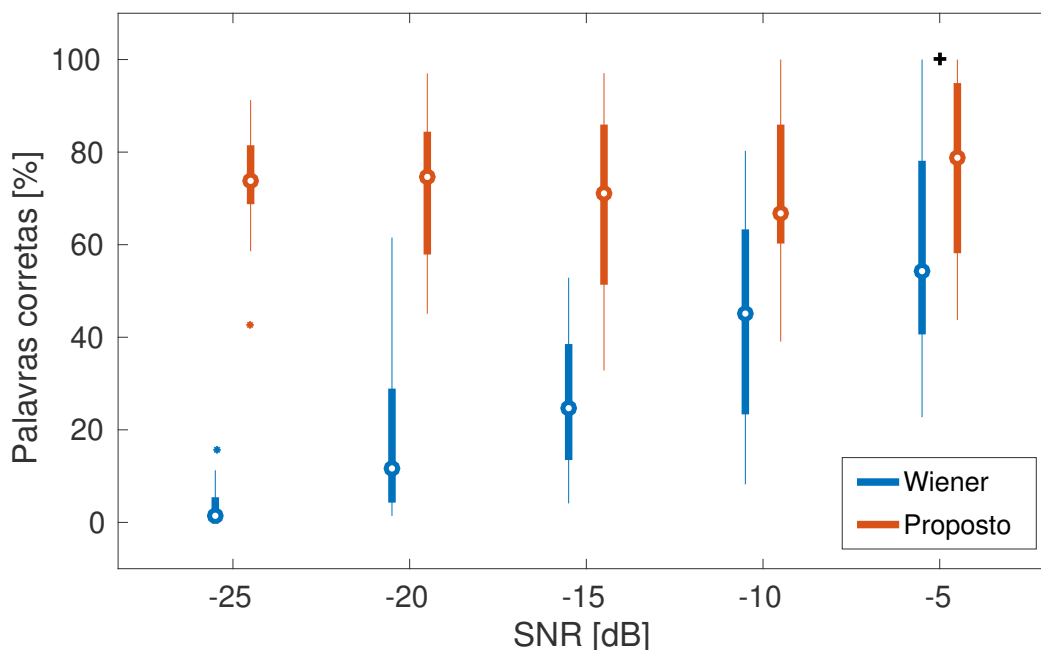


Figura 46 – Diagrama de caixa dos resultados do experimento I com usuários de implante coclear, agrupando todos os voluntários. Em azul, o filtro de Wiener; em laranja, o método proposto. Os círculos indicam as medianas, as barras o primeiro e terceiro quartis, e os bigodes os limites superior e inferior. Valores discrepantes (*outliers*) são representados por asteriscos (*). O sinal + indica que as medianas das duas distribuições (Wiener e proposto) não podem ser consideradas diferentes de acordo com o teste de Wilcoxon com $p < 0,05$.



pressão mais suave que a do filtro de Wiener, proporcionou melhores estimativas dos envelopes. A atenuação dos sinais foi proporcional à potência de ruído, de forma que, apesar da diminuição da SNR, as potências da fala em cada unidade tempo-frequência foram mantidas, conservando o eletrograma da fala.

4.4.5 Experimento II – Qualidade

Para o Experimento II, os sinais de fala são provenientes das 7 listas foneticamente balanceadas apresentadas em (COSTA; IORIO; ALBERNAZ, 2000). As listas foram gravadas com voz masculina. O mesmo ruído do experimento I foi utilizado.

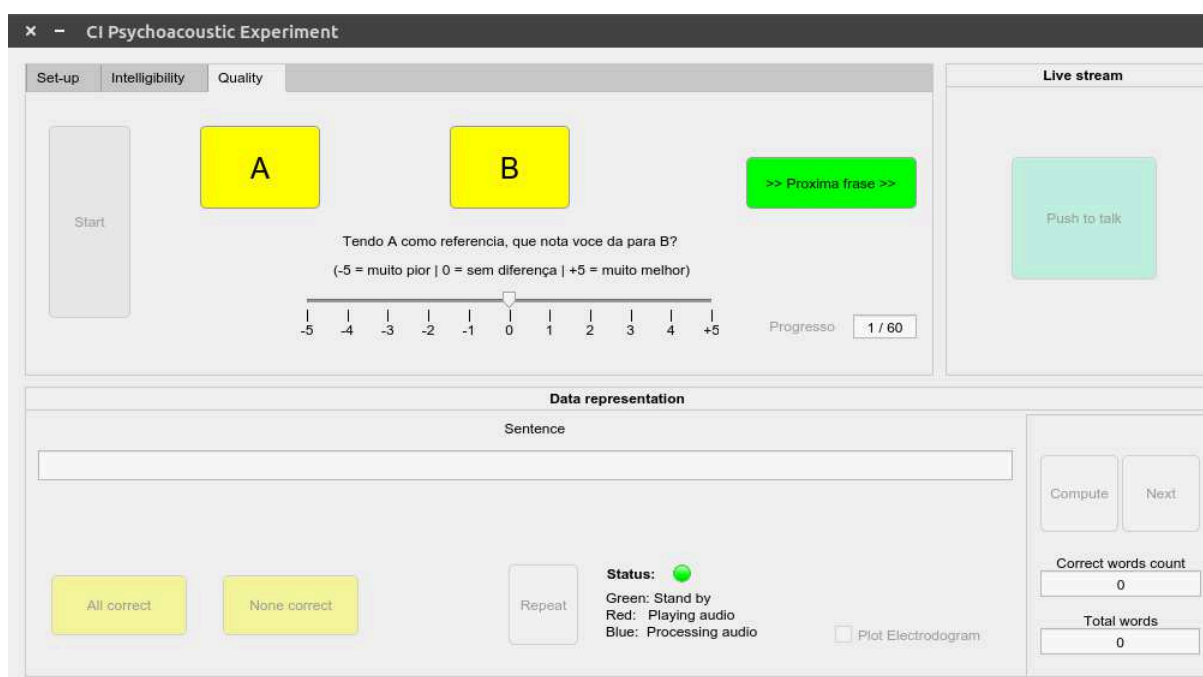
O experimento consistiu na avaliação pareada de áudios em 3 diferentes condições: sem ruído (limpo), com ruído aditivo a SNR de -10 dB e processados como descrito na seção anterior com o filtro de Wiener ou com o estimador de envelope proposto. Desta forma 3 comparações foram realizadas: limpo – Wiener; limpo – Proposto; Wiener – Proposto.

Para cada comparação uma lista com 10 frases foi aleatoriamente selecionada entre as sete disponíveis, sem reposição. O teste foi realizado duas vezes, de forma

a se verificar a confiabilidade dos voluntários. Assim, a sessão experimental completa continha 60 frases (3 comparações \times 10 frases \times 2 repetições).

Uma interface gráfica, mostrada na Figura 47, foi desenvolvida para a realização do experimento II.

Figura 47 – Interface gráfica para a realização do Experimento II. O ouvinte sob teste foi orientado a dar uma nota entre -5 e +5 ao áudio B com relação ao áudio A.



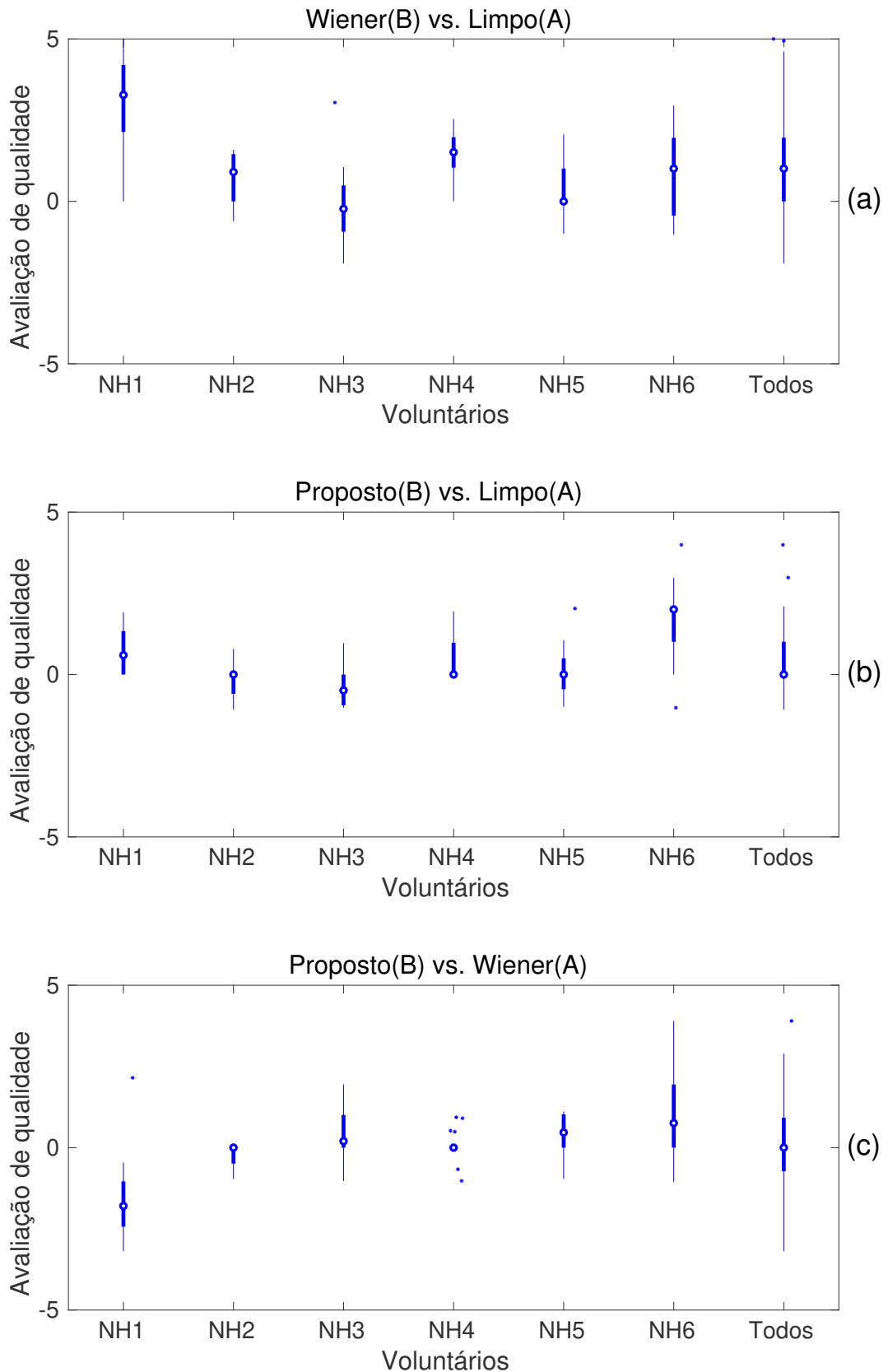
A cada iteração, uma frase entre as 60 era aleatoriamente selecionada e os dois áudios a serem comparados eram carregados. Os botões A e B iniciavam a reprodução dos áudios selecionados. Os participantes foram instruídos a, tomando o áudio em A como referência, dar uma nota entre -5 e +5 ao áudio B usando uma barra deslizante contínua. Notas negativas indicam que o áudio em B é pior que em A, enquanto que notas positivas indicam o oposto. Neste experimento, o voluntário pôde ouvir as frases quantas vezes achou necessário.

O Experimento II foi realizado após o Experimento I, com intervalo determinado pelo participante de maneira a evitar a fadiga. Uma sessão do Experimento II tinha duração aproximada de 20 minutos, e os voluntários eram livres para interromper a sessão a qualquer momento, mas todos os participantes optaram por realizar a sessão sem pausas.

Os resultados dos testes de qualidade para normouvintes são mostrados na Figura 48. Cada diagrama de caixa se refere a um voluntário, sendo que o último agrupa os dados de todos.

Considerando-se o conjunto das avaliações de todos os voluntários, as distribuições das notas são ligeiramente a favor dos sinais processados em relação ao sinal

Figura 48 – Resultados do Experimento II com normouvintes. Em (a), a comparação entre WF (B) e o sinal limpo (A); em (b), a comparação entre método proposto (B) e o sinal limpo (A); e em (c), a comparação entre WF (B) e o sinal limpo (A). Avaliações positivas são em favor do áudio em B, enquanto negativas refletem preferência ao áudio em A.



limpo, sem preferência entre os métodos de processamento, proposto ou Wiener. Em relação ao áudio limpo, o áudio resultante da simulação de IC é bastante distorcido, devido à perda da informação de fase, às interferências entre canais e à alteração das frequências centrais dos canais, introduzidos pelo simulador. Assim, o áudio simulado, mesmo na ausência de ruído, é incômodo para normouvintes. É possível que as atenuações de unidades tempo-frequência da fala devidas ao processamento com as funções de supressão, proposta ou Wiener, resultem em sinais com distorção reduzida aos ouvidos dos normouvintes, e isso justificaria essa leve predileção em favor dos sinais processados em relação à fala limpa.

A avaliação de qualidade é subjetiva e as preferências individuais são variadas. Por exemplo, o voluntário NH1 notavelmente prefere o processamento com o filtro de Wiener em relação ao método proposto ou mesmo o sinal limpo, enquanto que, por outro lado, NH6 tem uma preferência em favor ao método proposto. Entretanto a confiabilidade da avaliação de cada voluntário pode ser diferente. Neste experimento uma mesma lista de comparações foi apresentada duas vezes de forma idêntica, de forma a se avaliar a confiabilidade do julgamento dos voluntários. Isso pode ser estimado pela concordância entre os dois julgamentos de áudios idênticos.

Uma maneira de se quantificar a concordância é através do limite de concordância a 95%, com a análise de Bland-Altman (BARTLETT; FROST, 2008; BLAND; ALTMAN, 1986). Esse limite é representado pelo coeficiente de reprodutibilidade rpc , que representa o erro máximo esperado em 95% das próximas medidas feitas pelo voluntário (BARTLETT; FROST, 2008). Assim, quanto menor o rpc maior a concordância entre medidas realizadas pelo voluntário frase a frase. A análise de Bland-Altman apenas define os limites de concordância, mas não define se esses limites são aceitáveis ou não (GIAVARINA, 2015). Como a escala de avaliação varia entre -5 e 5, considera-se aqui aceitável, de maneira arbitrária, um $rpc < 2$.

Por outro lado, mesmo se a concordância entre medidas for baixa, pode ser que as distribuições das avaliações do teste e do reteste sejam similares. Assim, foi também realizado o teste de Wilcoxon para avaliar se as medianas dessas duas distribuições são significativamente diferentes ou não. Esse teste retorna o valor de significância p , tal que se $p < 0,05$ as distribuições podem ser consideradas diferentes, e caso contrário, não. Assim, quanto maior o valor de p , mais consistentes são as avaliações do usuário considerando-se toda a lista.

A Tabela 20 apresenta os valores de rpc da análise de Bland-Altman e p do teste de Wilcoxon para cada participante normouvinte e cada comparação feita no Experimento II.

Em muitos casos o rpc é grande (> 2), o que indica que na comparação frase a frase, não há concordância entre as avaliações do participante. Apesar disso, o teste de Wilcoxon não rejeita a hipótese de ambas as distribuições das repostas serem

Tabela 20 – Valores de rpc da análise de Bland-Altman e p do teste de Wilcoxon para cada participante normouvinte e cada comparação feita no Experimento II.

Voluntário	WF vs. Limpo		Proposto vs. Limpo		Proposto vs. WF	
	rpc	p	rpc	p	rpc	p
NH1	3,72	1,00	1,44	0,88	2,49	0,65
NH2	1,68	0,24	1,56	0,23	1,01	0,96
NH3	3,11	0,57	1,63	0,85	2,87	0,64
NH4	1,47	0,65	1,31	0,94	1,52	1,00
NH5	2,09	0,58	2,10	0,90	1,69	0,13
NH6	3,21	0,84	3,00	0,79	2,92	0,10

equivalentes. Assim, existe alguma consistência nas medidas e de certa forma, as avaliações dos voluntários são confiáveis.

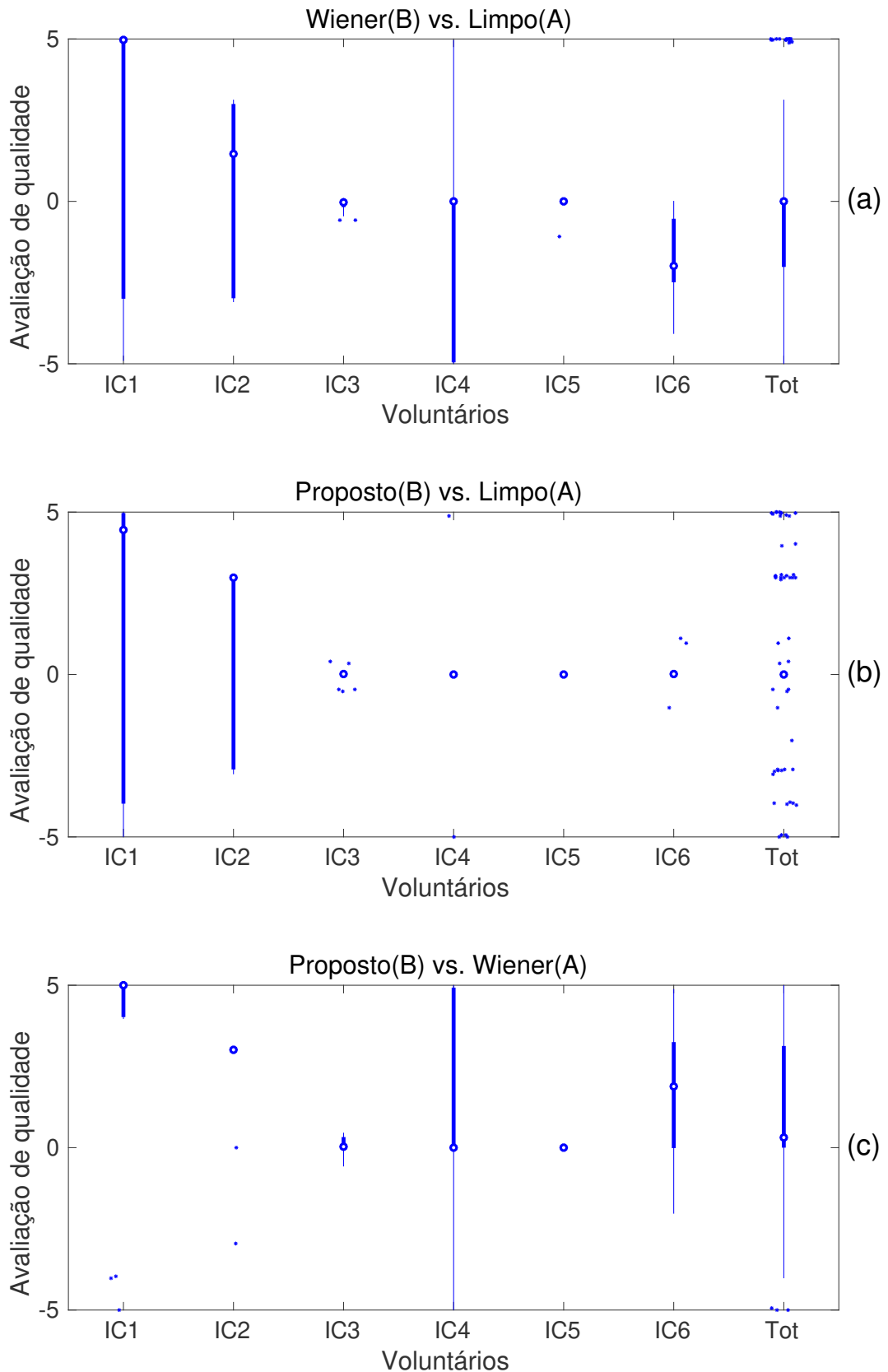
Os resultados do Experimento II com usuários de IC são apresentados na Figura 49 e os valores de rpc e p são mostrados na Tabela 21.

Tabela 21 – Valores de rpc da análise de Bland-Altman e p do teste de Wilcoxon para cada participante usuário de IC e cada comparação feita no Experimento II.

Voluntário	WF vs. Limpo		Proposto vs. Limpo		Proposto vs. WF	
	rpc	p	rpc	p	rpc	p
IC1	11,96	0,78	14,01	0,15	8,08	0,46
IC2	8,85	0,47	6,35	0,06	1,84	0,79
IC3	0,71	0,65	0,45	0,23	0,54	0,57
IC4	5,54	0,93	4,57	1,00	8,78	0,75
IC5	0,67	0,37	0,00	0,44	0,00	0,44
IC6	3,14	0,97	0,65	0,85	4,41	0,94

Assim como nos experimentos com normouvintes, as avaliações individuais dos usuários de IC são variadas. Os voluntários IC3 e IC5, cujas avaliações são as mais confiáveis em relação ao rpc , não identificaram praticamente nenhuma diferença entre os áudios limpo e processado por qualquer dos métodos. As medianas das avaliações de IC1 e IC2 indicam uma preferência de sinais processados em relação ao limpo, com o método proposto sendo considerado melhor que o de Wiener. As avaliações desses voluntários não são muito confiáveis ($rpc \gg 2$) para as comparações com sinal

Figura 49 – Resultados do Experimento II com usuários de IC. Em (a), a comparação entre WF (B) e o sinal limpo (A); em (b), a comparação entre método proposto (B) e o sinal limpo (A); e em (c), a comparação entre WF (B) e o sinal limpo (A). Avaliações positivas são em favor do áudio em B, enquanto negativas refletem preferência ao áudio em A.



limpo, mas as do IC2 são mais consistentes na comparação entre métodos de redução de ruído ($rpc = 1,84$). Para IC4 e IC6, há uma ligeira preferência do sinal limpo em relação ao filtro de Wiener, e do método proposto também em relação ao filtro de Wiener. Apesar de $rpc > 2$, tem-se $p > 0,75$ nessas comparações, o que atribui alguma confiabilidade aos avaliadores. IC4 e IC6 também não indicam diferença de qualidade entre o sinal limpo e o processado com método proposto, com alta confiabilidade em relação a rpc^6 e p .

Levando em conta o conjunto de todos os usuários, as medianas das avaliações nas três comparações é praticamente zero, com resultados distribuídos por toda a escala. Isso, somado à baixa confiabilidade nas avaliações dos voluntários, são fatores que não permitem afirmar que há diferenças significativas entre os sinais processados ou sinal limpo na percepção geral de usuários de IC em termos de qualidade. De qualquer forma, considerando-se apenas os voluntários com avaliações mais confiáveis, IC3 e IC5, pode-se afirmar que o processo ideal de redução de ruído (com cálculo exato dos ganhos) resulta em estímulos suficientemente próximos aos do sinal limpo, de forma a não introduzir distorções que reduzam a qualidade da percepção auditiva da fala.

4.5 DISCUSSÃO E CONSIDERAÇÕES FINAIS

Um novo filtro foi proposto com o objetivo de minimizar o erro de estimação do envelope temporal do sinal de fala. O método pode ser incorporado na cadeia de processamento de implantes cocleares, quando este é realizado no domínio de tempo, usando um banco de filtros.

Simulações numéricas indicam que apesar da estimação do sinal de fala ser pior que a realizada pelo filtro de Wiener, o erro associado à estimação do envelope desse sinal é consideravelmente menor, sobretudo para sinais submetidos a SNRs negativas. Isso era esperado visto que o filtro proposto minimiza o erro quadrático médio entre os envelopes quadráticos enquanto que o filtro de Wiener minimiza o erro quadrático médio entre os sinais. Além disso, o processamento com o novo método leva a melhores valores da métrica SRMR-CI, associada à inteligibilidade de fala em usuários de implantes cocleares.

Os resultados dos experimentos psicoacústicos de inteligibilidade, Experimento I, corroboram com os das simulações numéricas. Verifica-se que o método proposto leva a um melhor desempenho de inteligibilidade para usuários de IC que o WF, principalmente para baixos valores de SNR, menores que -5 dB. A Figura 46 resume os resultados que suportam essa afirmação. Verificou-se igualmente que o uso de méto-

⁶ Apesar de $rpc > 2$ nessa comparação para IC4, esse valor é alto devido a apenas dois outliers nos extremos da escala.

dos de redução de ruído, seja com o filtro de Wiener, seja com o método proposto, não leva a diferenças notáveis em termos de qualidade em relação ao sinal limpo.

Nos experimentos realizados neste trabalho, os ganhos do filtro proposto e do filtro de Wiener foram calculados de forma ideal, ou seja, com conhecimento *a priori* das potências do sinal e ruído em cada canal e *frame* de análise. Desta forma, os resultados obtidos representam o melhor desempenho possível de cada técnica.

Trabalhos futuros devem envolver a avaliação do desempenho do filtro em condições mais próximas das reais, em que os coeficientes dos filtros sejam calculados a partir de estimativas da relação sinal-ruído local.

5 EPÍLOGO

Neste capítulo final, faz-se uma discussão geral à luz dos resultados mostrados nos capítulos 3 e 4, seguida de um sumário da tese com destaque para as contribuições aportadas pelo trabalho.

5.1 DISCUSSÃO GERAL

Os resultados do Experimento I no capítulo 4 corroboram os indícios das simulações numéricas nesse mesmo capítulo, de que o método proposto proporciona mais inteligibilidade da fala que o filtro de Wiener, principalmente para valores de $SNR \leq -10dB$. Esse resultado é contrário à ideia de que métodos de redução de ruído mais agressivos seriam mais adequados para implantes cocleares (GOEHRING *et al.*, 2016; HERBACH *et al.*, 2012; KONING; MADHU; WOUTERS, 2015; MAUGER; ARORA; DAWSON, 2012; SERIZEL *et al.*, 2014). Segundo Chung, Nelson e Teske (2012), a justificativa para essa afirmação reside no fato de que a compressão realizada na cadeia de processamento do IC reduziria o efeito da redução de ruído. Isso porque o ruído atenuado seria amplificado pela compressão, enquanto que o sinal de fala, com maior amplitude, seria atenuado (CHUNG, 2007). Entretanto, essa visão pode não ser válida, uma vez que, a compressão no IC é realizada em cada canal independentemente. Além disso, como as informações de fase dos sinais são desconsideradas, em cada unidade tempo-frequência há apenas um valor, a soma das amplitudes de fala e ruído, sendo esses sinais indistinguíveis nesse ponto. A redução de ruído para IC pode ser considerada exitosa se a potência em cada unidade tempo-frequência for próxima à da fala, independentemente das contribuições do próprio sinal de fala e da interferência.

Uma outra justificativa é a de que usuários de IC são mais tolerantes à distorção introduzida pela redução de ruído do que normouvintes (QAZI *et al.*, 2012). Não é claro, entretanto, a que tipo de distorção os autores se referem. Como pode ser visto no capítulo 3, seção 3.1, considerando-se o ruído aditivo, a distorção introduzida pela redução de ruído pode ser associada a uma supressão insuficiente do ruído (Equação (14)), ou ligada à uma supressão excessiva da fala (Equação (13)).

Li e Loizou (2008b) mediram o efeito desses diferentes tipos de distorção quando do uso de máscaras binárias para normouvintes. Os autores verificaram que a distorção relacionada ao ruído tem maior impacto na degradação da inteligibilidade que a distorção de fala. Por outro lado, um estudo mais recente, com usuários de IC, conclui que ambos tipos de distorção são igualmente degradantes (KRESSNER *et al.*, 2016). Além disso, outros resultados indicam que usuários de IC seriam menos tolerantes que normouvintes a erros de estimação das máscaras binárias, na maioria das condições estudadas (KRESSNER *et al.*, 2016).

O fato é que usuários de IC estão habituados a informações com baixa resolução temporal e espectral. Desta forma, é de se esperar uma maior inteligibilidade de fala por usuários de IC, em relação a normouvintes, quando apresentados com sinais distorcidos. Contudo, atualmente, não há uma caracterização do que seria uma distorção aceitável, e, conseqüentemente, qual seria o impacto disso na escolha de um método de redução de ruído.

O próprio conceito de agressividade das máscaras não é objetivo. Mauger, Dawson e Hersbach (2012) discutem que comparações envolvendo o filtro paramétrico de Wiener (Equação (9)) relacionam a agressividade ao parâmetro β , enquanto que, no caso de máscaras binárias, essa agressividade está associada ao limiar de decisão. Os autores sugerem que a agressividade deve ser mensurada pelo Limiar Aparente de Ganho (aGT), que é o valor de SNR local para o qual a função de supressão cruza com 0,5. Por esse método, a agressividade não teria nenhuma relação com a suavidade da transição e, conseqüentemente, o filtro de Wiener e a máscara binária poderiam ter a mesma agressividade.

Tomando a máscara conformável derivada no capítulo 3 como referência, o aGT é equivalente ao parâmetro μ . De fato, a tabela 15 mostra que a métrica de inteligibilidade (SRMR-CI) foi maximizada quando o $\mu = aGT \geq 0$ dB. Contudo, a função de supressão do estimador de envelope do capítulo 4 (equação (177)) tem claramente $aGT < 0$ dB e resultou em melhores níveis de inteligibilidade que o filtro de Wiener ($\mu = aGT < 0$ dB). É possível que a diferença nos dois casos esteja relacionada à arquitetura de redução de ruído utilizada em cada um deles: enquanto as máscaras do capítulo 3 são aplicadas aos 256 *bins* da STFT do sinal contaminado, de acordo com a Figura 16; no capítulo 4 os ganhos de redução de ruído são aplicados nos 22 canais do banco de filtros, como ilustrado na Figura 14b. A influência das diferentes arquiteturas na inteligibilidade de usuários de IC deve ser investigada em trabalhos futuros.

Por outro lado, alguns autores consideram a máscara binária como sendo mais agressiva que abordagens mais suaves, como o filtro de Wiener, justamente devido à sua transição mais abrupta (KONING; MADHU; WOUTERS, 2015). Esse conceito confunde agressividade com a suavidade/dureza da máscara. A suavidade da máscara está associada ao parâmetro γ da máscara conformável do capítulo 3. Segundo a tabela 15, a inteligibilidade é maximizada quando $\gamma < 1$, ou seja, quando a máscara é mais suave que o filtro de Wiener. No capítulo 4, o estimador de envelope proposto, cujo desempenho em inteligibilidade se mostrou melhor que o do filtro de Wiener (Figuras 37 e 46), também tem uma função de supressão mais suave que a deste, como pode ser visto na Figura 35. Esses resultados indicam que, independentemente da arquitetura utilizada para a redução de ruído (FFT ou FB), máscaras mais suaves que o filtro de Wiener são recomendáveis para o aumento de inteligibilidade de usuários de IC, quando $SNR \leq -10$ dB.

5.2 CONCLUSÃO

Nesta tese foram apresentadas duas abordagens para a redução de ruído aditivo aplicáveis à implantes cocleares, cada uma considerando um diferente método de separação do sinal em bandas de frequência.

Em um primeiro caso, fazendo-se essa decomposição em frequências através da transformada de Fourier, foi derivada uma máscara conformável com dois parâmetros que permitem o ajuste da função de supressão. Simulações numéricas extensas mostram que o desempenho é, no mínimo, similar ao do filtro paramétrico de Wiener, com a vantagem de a máscara proposta ser uma função mais maleável e de ser a solução ótima da minimização de uma função custo envolvendo as distorções relacionadas à fala e ao ruído.

No segundo, em que o sinal é analisado através de um banco de filtros, considerando-se a perda da informação de fase, um estimador do envelope temporal foi projetado. Simulações numéricas comprovaram que o envelope do sinal desejado é melhor estimado pelo filtro proposto em relação ao filtro de Wiener. Além disso, avaliações psicacústicas com normouvintes e usuários de implante coclear mostraram um ganho significativo de inteligibilidade, principalmente em condições de baixa SNR.

Abaixo são listadas as contribuições aportadas pelos trabalhos realizados nesta tese.

1. Desenvolvimento de uma teoria de máscaras tempo-frequência parametrizadas baseada na otimização de uma função custo dependente das distorções relacionadas à fala e ao ruído. Ela engloba máscaras clássicas da literatura como o filtro de Wiener e a Máscara Binária (ambas amplamente sugeridas para utilização em implantes cocleares) como soluções particulares de um mesmo problema;
2. Demonstração inédita que a máscara binária pode ser entendida como a solução do problema de minimização de uma função custo;
3. Desenvolvimento de um estimador de envelope temporal a partir de características presentes nas principais estratégias de estimulação de implantes cocleares. Resultados de experimentos numéricos e psicoacústicos indicam um desempenho melhor em relação ao filtro de Wiener equivalente.
4. A contestação da noção frequente na literatura de que técnicas de redução de ruído mais agressivas seriam mais adequadas para implantes cocleares. As simulações numéricas e experimentos psicoacústicos apresentados nos capítulos 3 e 4 apontam para um melhor desempenho, em inteligibilidade, de máscaras mais suaves em relação à máscara binária e ao filtro de Wiener.

Além das contribuições listadas acima, vale ressaltar que este trabalho foi pioneiro no estudo e proposição de técnicas de redução de ruído para implantes cocleares no país. Assim sendo, trouxe conhecimento metodológico relevante ao grupo de pesquisa e aos colaboradores para realização de novos estudos nesse domínio, principalmente no que diz respeito à realização dos experimentos psicoacústicos com usuários de IC.

5.3 PUBLICAÇÕES

- CHIEA, Rafael Attili; COSTA, Márcio Holsbach; BARRAULT, Guillaume. *New insights on the optimality of parameterized Wiener filters for speech enhancement applications*. **Speech Communication**, v. 109, p. 46-54, 2019.
- CHIEA, Rafael Attili; COSTA, Márcio Holsbach; BARRAULT, Guillaume. *Uma Comparação entre Máscaras Tempo-frequência para Redução de Ruído em Implantes Cocleares*. **XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT)**, Petrópolis-RJ, 2019.
- CHIEA, Rafael Attili; COSTA, Márcio Holsbach; PINHEIRO, Maria Madalena; CORDIOLI, Júlio. *Um Estimador de Envelope Temporal para Redução de Ruído em Implantes Cocleares*. **XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT)**, Petrópolis-RJ, 2019.

REFERÊNCIAS

ALAM, Md Jahangir; O'SHAUGHNESSY, Douglas; SELOUANI, Sid Ahmed. Speech enhancement employing a sigmoid-type gain function with a modified a priori signal-to-noise ratio (SNR) estimator. *In: CANADIAN Conference on Electrical and Computer Engineering*. 2008. p. 631–635. DOI: 10.1109/CCECE.2008.4564612.

ALCAIM, Abraham; SOLEWICZ, José Alberto; MORAES, João Antônio. Frequência de ocorrência dos fonos e listas de frases foneticamente balanceadas no português falado no Rio de Janeiro. **Revista da Sociedade Brasileira de Telecomunicações (SBrT)**, v. 7, n. 1, p. 23–41, 1992. DOI: 10.14209/jcis.1992.2.

ALI, Hussnain. UT-Dallas Cochlear implant research interface: System level design. 2016.

ALI, Hussnain *et al.* Design and evaluation of a personal digital assistant-based research platform for cochlear implants. **IEEE Transactions on Biomedical Engineering**, v. 60, n. 11, p. 3060–3073, 2013. ISSN 00189294. DOI: 10.1109/TBME.2013.2262712.

ARSLAN, Levent M. Modified Wiener filtering. **Signal Processing**, Elsevier, v. 86, n. 2, p. 267–272, 2006. ISSN 01651684. DOI: 10.1016/j.sigpro.2005.05.021.

ASSOCIATION, American Speech-Language-Hearing. **Causes of Hearing Loss in Adults**. 2019. <https://www.asha.org/public/hearing/Causes-of-Hearing-Loss-in-Adults/>. Acessado em: 09/10/2019.

BARTLETT, J W; FROST, C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. **Ultrasound in Obstetrics and Gynecology**, v. 31, n. 4, p. 466–475, 2008. ISSN 09607692. DOI: 10.1002/uog.5256.

BENESTY, Jacob; SONDHI, Mohan M; HUANG, Yiteng. **Springer handbook of speech processing**. Springer Science & Business Media, 2008. p. 1176. ISBN 9783540491255. DOI: 10.1007/978-3-540-49127-9_19. arXiv: arXiv:1011.1669v3.

BERGERON, F; HOTTON, M. Perception in noise with the Digisonic SP cochlear implant: Clinical trial of Saphyr processor's upgraded signal processing. **European Annals of Otorhinolaryngology, Head and Neck Diseases**, Elsevier, p. 3–5, 2016. ISSN 18797296. DOI: 10.1016/j.anorl.2016.04.019.

BLAND, J Martin; ALTMAN, Douglas G. Statistical methods for assessing agreement between two methods of clinical measurement. **The Lancet**, v. 327, p. 307–310, 1986.

BOLNER, Federico; HERBACH, Adam A. **Body noise reduction in auditory prostheses**. v. 1. United States Patent e Trademark Office, 2018.

BOYD, Stephen; VANDENBERGHE, Lieven. **Convex optimization**. New York: Cambridge University Press, 2004. p. 701. ISBN 9780521833783.

BRÄCKER, Timo; SCHULTE, Michael. Simulation und Vergleich von Sprachkodierungsstrategien in Cochlea-Implantaten. **Zeitschrift der Audiologie/Audiological Acoustics**, January 2019, 2009.

BRUNGART, Douglas S *et al.* Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. **The Journal of the Acoustical Society of America**, v. 120, n. 6, p. 4007–4018, 2006. ISSN 00014966. DOI: 10.1121/1.2363929.

BÜCHNER, Andreas *et al.* Advanced beamformers for cochlear implant users: Acute measurement of speech perception in challenging listening conditions. **PLoS ONE**, v. 9, n. 4, 2014. ISSN 19326203. DOI: 10.1371/journal.pone.0095542.

CHEHRESA, S; SAVOJI, M H. Codebook constrained iterative and parametric wiener filter speech enhancement. *In: IEEE International Conference on Signal and Image Processing Applications*. 2009. p. 548–553. DOI: 10.1109/ICSIPA.2009.5478717.

CHEN, Fei; LOIZOU, Philipos C. Predicting the intelligibility of vocoded and wideband Mandarin Chinese. **The Journal of the Acoustical Society of America**, v. 129, n. 5, p. 3281–90, 2011. ISSN 1520-8524. DOI: 10.1121/1.3570957.

CHEN, Fei; LOIZOU, Philipos C. Speech enhancement using a frequency-specific composite Wiener function. *In: PROCEEDINGS of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 2010. p. 4726–4729. DOI: 10.1109/ICASSP.2010.5495166.

CHEN, Fei; ZHANG, Yuan-Ting. A novel temporal fine structure-based speech synthesis model for cochlear implant. **Signal Processing**, Elsevier, v. 88, n. 11, p. 2693–2699, 2008. ISSN 01651684. DOI: 10.1016/j.sigpro.2008.05.011.

CHEN, Fei; ZHANG, Yuan-Ting. Zerocrossing-based nonuniform sampling to deliver low-frequency fine structure cue for cochlear implant. **Digital Signal Processing**, Elsevier, v. 21, n. 3, p. 427–432, 2011. ISSN 10512004. DOI: 10.1016/j.dsp.2010.12.002.

CHEN, Jingdong *et al.* New insights into the noise reduction Wiener filter. **IEEE Transactions on Audio, Speech and Language Processing**, v. 14, n. 4, p. 1218–1234, 2006. ISSN 1558-7916. DOI: 10.1109/TSA.2005.860851.

CHEN, Yousheng; GONG, Qin. Broadband beamforming compensation algorithm in CI front-end acquisition. **Biomedical Engineering Online**, BioMed Central Ltd., v. 12, n. 1, p. 18, 2013. ISSN 1475-925X. DOI: 10.1186/1475-925X-12-18.

CHEN, Yousheng; GONG, Qin. Real-time spectrum estimation-based dual-channel speech-enhancement algorithm for cochlear implant. **Biomedical Engineering Online**, BioMed Central Ltd., v. 11, 2012. ISSN 1475925X. DOI: 10.1186/1475-925X-11-74.

CHUNG, King. Effective compression and noise reduction configurations for hearing protectors. **The Journal of the Acoustical Society of America**, v. 121, n. 2, p. 1090–1101, 2007. ISSN 0001-4966. DOI: 10.1121/1.2409859.

CHUNG, King; NELSON, Lance; TESKE, Melissa. Noise reduction technologies implemented in head-worn preprocessors for improving cochlear implant performance in reverberant noise fields. **Hearing research**, Elsevier, v. 291, p. 41–51, 2012. ISSN 1878-5891. DOI: 10.1016/j.heares.2012.06.003.

CLARK, Graeme. **Cochlear implants fundamentals and application**. 2003. p. 830. ISBN 0387955836.

COOKE, Martin *et al.* Robust automatic speech recognition with missing and unreliable acoustic data. **Speech Communication**, v. 34, n. 3, p. 267–285, 2001. ISSN 01676393. DOI: 10.1016/S0167-6393(00)00034-0.

COSENTINO, Stefano *et al.* Cochlear implant filterbank design and optimization: A simulation study. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, IEEE, v. 22, n. 2, p. 347–353, 2014. ISSN 2329-9290. DOI: 10.1109/TASLP.2013.2290502.

COSENTINO, Stefano *et al.* Towards objective measures of speech intelligibility for cochlear implant users in reverberant environments. *In: 11TH IEEE International Conference on Information Science, Signal Processing and their Applications*. 2012. p. 666–671.

COSTA, Maristela Julio; IORIO, Maria Cecilia Martinelli; ALBERNAZ, Pedro Luiz Mangabeira. Desenvolvimento de um teste para avaliar a habilidade de reconhecer a fala no silêncio e no ruído. **Pro Fono**, p. 9–16, 2000.

CROCHIERE, R E. 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. ISSN 00963518. DOI: 10.1109/TASSP.1980.1163353.

DANIEL, Adrien *et al.* An experimental framework for the derivation of perceptually-optimal noise suppression functions. *In: PROCEEDINGS of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 2013. p. 7800–7804. DOI: 10.1109/ICASSP.2013.6639182.

DIETZ, Mathias; MCALPINE, David. Advancing binaural cochlear implant technology. **Trends in Hearing**, v. 19, n. 0, p. 1–4, 2015. ISSN 2331-2165. DOI: 10.1177/2331216515623374.

DINGEMANSE, J Gertjan; GOEDGEBURE, André. Application of noise reduction algorithm ClearVoice in cochlear implant processing: Effects on noise tolerance and speech intelligibility in noise in relation to spectral resolution. **Ear & Hearing**, n. 2010, p. 1–11, 2014.

DRESCHLER, Wouter A *et al.* ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. **Audiology**, v. 40, n. 3, p. 148–157, 2001. ISSN 0020-6091. DOI: 10.3109/00206090109073110. arXiv: arXiv:1011.1669v3.

EPHRAIM, Yariv; MALAH, David. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. **IEEE Transactions on Acoustics, Speech and Signal Processing**, n. 6, p. 1109–1121, 1984. ISSN 0096-3518. DOI: 10.1109/TASSP.1984.1164453.

FALK, Tiago H. A non-Intrusive quality and intelligibility measure of reverberant and dereverberated speech. **IEEE Transactions on Audio, Speech and Language Processing**, v. 18, n. 7, p. 1766–1774, 2010. ISSN 1558-7916. DOI: 10.1109/TASL.2010.2052247.

FONTAINE, Mathieu *et al.* Explaining the parameterized Wiener filter with alpha-stable processes. *In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. 2017. p. 51–55. DOI: 10.1109/WASPAA.2017.8169993.

GAZOR, Saeed; MEMBER, Senior; ZHANG, Wei. Speech probability distribution. **IEEE Signal Processing Letters**, v. 10, n. 7, p. 204–207, 2003. ISSN 10709908. DOI: 10.1109/LSP.2003.813679.

GEORGE, Shefin S *et al.* Electrophysiological channel interactions using focused multipolar stimulation for cochlear implants. **Journal of Neural Engineering**, v. 12, n. 6, p. 066005, 2015. ISSN 1741-2560. DOI: 10.1088/1741-2560/12/6/066005.

GIAVARINA, Davide. Understanding Bland Altman analysis. **Biochemia Medica**, v. 25, n. 2, p. 141–151, 2015. ISSN 13300962. DOI: 10.11613/BM.2015.015.

GOEHRING, Tobias *et al.* Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users. **Hearing Research**, Elsevier, v. 344, November, p. 1–12, 2016. ISSN 18785891. DOI: 10.1016/j.heares.2016.11.012.

GOLD, B; OPPENHEIM, Alan V; RADER, C M. **Theory and implementation of the discrete Hilbert transform**. 1969. p. 235–250.

GUEVARA, Nicolas *et al.* The Voice Track multiband single-channel modified Wiener-filter noise reduction system for cochlear implants: patients' outcomes and subjective

appraisal. **International Journal of Audiology**, v. 2027, May, p. 1–8, 2016. ISSN 1499-2027. DOI: 10.3109/14992027.2016.1172267.

HAMACHER, Volkmar *et al.* Signal processing in high-end hearing aids: State of the art, challenges, and future trends. **EURASIP Journal on Advances in Signal Processing**, v. 2005, n. 18, p. 2915–2929, 2005. ISSN 1687-6172. DOI: 10.1155/ASP.2005.2915.

HAST, Anne *et al.* Speech perception of elderly cochlear implant users under different noise conditions. **Otology and Neurotology**, v. 36, n. 10, p. 1638–1643, 2015. ISSN 15374505. DOI: 10.1097/MAO.0000000000000883.

HAZRATI, Oldooz; LEE, Jaewook; LOIZOU, Philipos C. Blind binary masking for reverberation suppression in cochlear implants. **The Journal of the Acoustical Society of America**, v. 133, n. 3, p. 1607–1614, 2013. ISSN 1520-8524. DOI: 10.1121/1.4789891.

HAZRATI, Oldooz; LOIZOU, Philipos C. Tackling the Combined Effects of Reverberation and Masking Noise Using Ideal Channel Selection. **Journal of Speech, Language, and Hearing Research**, v. 55, n. 2, p. 500–510, 2012. ISSN 1092-4388. DOI: 10.1044/1092-4388(2011/11-0073).

HAZRATI, Oldooz *et al.* **Hearing assistive technologies: Challenges and opportunities**. 2016. Slide presentation at ISCA Interspeech.

HAZRATI, Oldooz *et al.* Simultaneous suppression of noise and reverberation in cochlear implants using a ratio masking strategy. **The Journal of the Acoustical Society of America**, v. 134, n. 5, p. 3759–65, 2013. ISSN 1520-8524. DOI: 10.1121/1.4823839.

HE, Wenzhi; ZHENG, Nengheng; MENG, Qinglin. The effect of gain thresholds on speech intelligibility for statistical model based noise reduction for cochlear implants: A simulation based verification. *In*: PROCEEDINGS of the 10th International Symposium on Chinese Spoken Language Processing. 2017. DOI: 10.1109/ISCSLP.2016.7918459.

HEHRMANN, Phillipp *et al.* Improved speech intelligibility with cochlear implants using state-of-the-art noise reduction algorithms. **ITG-Fachbericht**, v. 236, p. 26–28, 2012.

HERSBACH, Adam A *et al.* A beamformer post-filter for cochlear implant noise reduction. **The Journal of the Acoustical Society of America**, v. 133, n. 4, p. 2412–20, 2013. ISSN 1520-8524. DOI: 10.1121/1.4794391.

HERSBACH, Adam A *et al.* Combining directional microphone and single-channel noise reduction algorithms. **Ear & Hearing**, v. 33, n. 4, e13–e23, 2012. ISSN 0196-0202. DOI: 10.1097/AUD.0b013e31824b9e21.

HJØRUNGNES, Are *et al.* Complex-valued matrix differentiation : Techniques and key results. **IEEE Transactions on Signal Processing**, v. 55, n. 6, p. 2740–2746, 2007.

HOLUBE, Inga; KOLLMEIER, Birger. Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model. **The Journal of the Acoustical Society of America**, v. 100, n. 3, p. 1703–1716, 1996. ISSN 0001-4966. DOI: 10.1121/1.417354.

HOLUBE, Inga *et al.* Development and analysis of an international speech test signal (ISTS). **International Journal of Audiology**, v. 49, n. 12, p. 891–903, 2010. ISSN 14992027. DOI: 10.3109/14992027.2010.506889.

HU, Hongmei *et al.* Better-ear glimpsing with symmetrically-placed interferers in bilateral cochlear implant users. **The Journal of the Acoustical Society of America**, v. 143, n. 4, p. 2128–2141, 2018. ISSN 0001-4966. DOI: 10.1121/1.5030918.

HU, Yi; LOIZOU, Philipos C. A comparative intelligibility study of speech enhancement algorithms. *In: PROCEEDINGS of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 2007. DOI: 10.1109/ICASSP.2007.366974.

HU, Yi; LOIZOU, Philipos C. A new sound coding strategy for suppressing noise in cochlear implants. **The Journal of the Acoustical Society of America**, v. 124, n. 1, p. 498–509, 2008. ISSN 00014966. DOI: 10.1121/1.2924131.

HU, Yi; LOIZOU, Philipos C. A perceptually motivated approach for speech enhancement. **IEEE Transactions on Speech and Audio Processing**, v. 11, n. 5, p. 457–465, 2003. ISSN 1063-6676. DOI: 10.1109/TSA.2003.815936.

HU, Yi; LOIZOU, Philipos C. Environment-specific noise suppression for improved speech intelligibility by cochlear implant users. **The Journal of the Acoustical Society of America**, v. 127, n. 6, p. 3689–3695, 2010. ISSN 00014966. DOI: 10.1121/1.3365256.

HU, Yi *et al.* Use of a sigmoidal-shaped function for noise attenuation in cochlear implants. **The Journal of the Acoustical Society of America**, v. 122, n. 4, p. e1128–1134, 2007. ISSN 00014966. DOI: 10.1121/1.2772401.

IEEE. Recommended practice for speech quality measurements. **IEEE Transactions on Audio and Electroacoustics**, v. 17, n. 3, p. 225–246, 1969.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Censo demográfico 2010: Características gerais da população, religião e pessoas com deficiência, 2012.

INTERNATIONAL TELECOMMUNICATION UNION. **P.862: Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs**. v. 862. 2001. p. 862. DOI: 10.1109/ICASSP.2001.941023.

INTERNATIONAL TELECOMMUNICATION UNION. **P.862.2 Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs**. 2007. DOI: 10.1016/S1389-1286(03)00243-3.

JENSEN, Jesper *et al.* A study of the distribution of time-domain speech samples and discrete fourier coefficients delft. *In: PROCEEDINGS of SPS-DARTS*. 2005. p. 155–158.

JEONG, So Young; JEONG, Jae Hoon; OH, Kwang Cheol. Dominant speech enhancement based on SNR-adaptive soft mask filtering. *In: PROCEEDINGS of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 2009. p. 1317–1320. DOI: 10.1109/ICASSP.2009.4959834.

JOHANNESMA, Peter I M. The pre-response stimulus ensemble of neurons in the cochlear nucleus. *In: SYMPOSIUM on Hearing Theory*. 1972. p. 58–69.

KAK, Subhash. The number theoretic Hilbert transform. **Circuits, Systems, and Signal Processing**, v. 33, n. 8, p. 2539–2548, 2014. ISSN 15315878. DOI: 10.1007/s00034-014-9759-8.

KAY, Steven M. **Intuitive probability and random processes using MATLAB®**. 2006. p. 1–833. ISBN 9780387241579. DOI: 10.1007/b104645.

KOKKINAKIS, Kostas; COX, Casey. Reducing the impact of wind noise on cochlear implant processors with two microphones. **The Journal of the Acoustical Society of America**, v. 135, April 2014, e1219–e1225, 2014. ISSN 0001-4966. DOI: 10.1121/1.4871583.

KOKKINAKIS, Kostas; HAZRATI, Oldooz; LOIZOU, Philipos C. A channel-selection criterion for suppressing reverberation in cochlear implants. **The Journal of the Acoustical Society of America**, v. 129, n. 5, p. 3221–3232, 2011. ISSN 00014966. DOI: 10.1121/1.3559683.

KOKKINAKIS, Kostas; LOIZOU, Philipos C. Evaluation of objective measures for quality assessment of reverberant speech. *In: PROCEEDINGS of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 2011. p. 2420–2423. DOI: 10.1109/ICASSP.2011.5946972.

KOKKINAKIS, Kostas; LOIZOU, Philipos C. Multi-microphone adaptive noise reduction strategies for coordinated stimulation in bilateral cochlear implant devices. **The Journal of the Acoustical Society of America**, v. 127, n. 5, p. 3136–3144, 2010. ISSN 00014966. DOI: 10.1121/1.3372727.

KOKKINAKIS, Kostas *et al.* Single and multiple microphone noise reduction strategies in cochlear implants. **Trends in Amplification**, v. 16, n. 2, p. 102–116, 2012. ISSN 1084-7138. DOI: 10.1177/1084713812456906.

KONING, Raphael; MADHU, Nilesh; WOUTERS, Jan. Ideal time-frequency masking algorithms lead to different speech intelligibility and quality in normal-hearing and cochlear implant listeners. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 62, n. 1, p. 331–341, 2015. ISSN 0018-9294. DOI: 10.1109/TBME.2014.2351854.

KONING, Raphael *et al.* Perceptual and model-based evaluation of ideal time-frequency noise reduction in hearing-impaired listeners. **IEEE Trans. Neural Syst. Rehabil. Eng.**, v. 26, n. 3, p. 687–697, 2018. ISSN 15344320. DOI: 10.1109/TNSRE.2018.2794557.

KORDUS, Monika *et al.* An influence of directional microphones on the speech intelligibility and spatial perception by cochlear implant users. **Archives of Acoustics**, v. 40, n. 1, p. 81–92, 2015. ISSN 2300-262X. DOI: 10.1515/aoa-2015-0010.

KRESSNER, Abigail A *et al.* Cochlear implant speech intelligibility outcomes with structured and unstructured binary mask errors. **The Journal of the Acoustical Society of America**, v. 139, n. 2, p. 800–810, 2016. ISSN 0001-4966. DOI: 10.1121/1.4941567.

KREUTZ-DELGADO, Ken. The complex gradient operator and the CR-calculus. **arXiv**, p. 1–74, 2009. DOI: arXiv:0906.4835. arXiv: 0906.4835.

KREYSZIG, Erwin. **Advanced engineering mathematics**. 10th. New York, 2011. ISBN 978-0-470-45836-5 Printed.

LAGARIAS, Jeffrey C *et al.* Convergence properties of the Nelder–Mead simplex method in low dimensions. **SIAM Journal on Optimization**, v. 9, n. 1, p. 112–147, 1998. ISSN 1052-6234. DOI: 10.1137/S1052623496303470.

LANGNER, Florian; JÜRGENS, Tim; JU, Tim. Forward-masked frequency selectivity improvements in simulated and actual cochlear implant users using a preprocessing algorithm. **Trends in Hearing**, v. 20, p. 1–14, 2016. ISSN 23312165. DOI: 10.1177/2331216516659632.

LI, Ning; LOIZOU, Philipos C. Effect of spectral resolution on the intelligibility of ideal binary masked speech. **The Journal of the Acoustical Society of America**, v. 123, n. 4, e159–164, 2008a. ISSN 00014966. DOI: 10.1121/1.2884086.

LI, Ning; LOIZOU, Philipos C. Factors influencing intelligibility of ideal binary-masked speech: implications for noise reduction. **The Journal of the Acoustical Society of America**, v. 123, n. 3, p. 1673–1682, 2008b. ISSN 00014966. DOI: 10.1121/1.2832617.

LI, Yipeng; WANG, DeLiang. On the optimality of ideal binary time-frequency masks. **Speech Communication**, Elsevier, v. 51, n. 3, p. 230–239, 2009. ISSN 01676393. DOI: 10.1016/j.specom.2008.09.001.

LIM, Jae S; OPPENHEIM, Alan V. Enhancement and bandwidth compression of noisy speech. **Proceedings of the IEEE**, v. 67, n. 12, p. 1586–1604, 1979. ISSN 00189219. DOI: 10.1109/PROC.1979.11540.

LIUTKUS, Antoine; BADEAU, Roland. Generalized Wiener filtering with fractional power spectrograms. *In: PROCEEDINGS of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 2015. p. 266–270. DOI: 10.1109/ICASSP.2015.7177973.

LOIZOU, Philipos C. Introduction to cochlear implants. **IEEE Engineering in Medicine and Biology Magazine**, v. 15, n. 5, p. 32–42, 1999a. ISSN 0739-5175. DOI: 10.1109/51.740962.

LOIZOU, Philipos C. Signal-processing techniques for cochlear implants. **IEEE Engineering in Medicine and Biology Magazine**, IEEE, v. 18, n. 3, p. 34–46, 1999b. ISSN 0739-5175. DOI: 10.1109/51.765187.

LOIZOU, Philipos C. **Speech enhancement: Theory and practice**. 2nd editio. Boca Raton, FL: CRC Press, 2013. ISBN 9781466504226.

LOIZOU, Philipos C; LOBO, Arthur P; HU, Yi. Subspace algorithms for noise reduction in cochlear implants. **The Journal of the Acoustical Society of America**, v. 118, n. 5, p. 2791–2793, 2005. ISSN 00014966. DOI: 10.1121/1.2065847.

LOPEZ-POVEDA, Enrique A *et al.* A binaural cochlear implant sound coding strategy inspired by the contralateral medial olivocochlear reflex. **Ear & Hearing**, v. 37, n. 3, e138–48, 2016. ISSN 1538-4667. DOI: 10.1097/AUD.0000000000000273.

LU, Yang; LOIZOU, Philipos C. Estimators of the magnitude-squared spectrum and methods for incorporating SNR uncertainty. **IEEE Transactions on Audio, Speech and Language Processing**, v. 19, n. 5, p. 1123–1137, 2011. ISSN 15587924. DOI: 10.1109/TASL.2010.2082531.

MADHU, Nilesh *et al.* The potential for speech intelligibility improvement using the ideal binary mask and the ideal Wiener filter in single channel noise reduction systems: application to auditory prostheses. **IEEE Transactions on Audio, Speech and Language Processing**, v. 21, n. 1, p. 63–72, 2013. ISSN 1558-7916. DOI: 10.1109/TASL.2012.2213248.

MANOLAKIS, Dimitris G; INGLE, Vinay K; KOGON, Stephen M. **Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing**. 2005. ISBN 1580536107.

MAUGER, Stefan J; ARORA, Komal; DAWSON, Pam W. Cochlear implant optimized noise reduction. **Journal of Neural Engineering**, IOP Publishing LTD., v. 9, n. 6, p. 065007, 2012. ISSN 1741-2560. DOI: 10.1088/1741-2560/9/6/065007.

MAUGER, Stefan J; DAWSON, Pam W; HERSBACH, Adam A. Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction. **The Journal of the Acoustical Society of America**, v. 131, November 2011, p. 327, 2012. ISSN 00014966. DOI: 10.1121/1.3665990.

MCHUGH, Mary L. Multiple comparison analysis testing in ANOVA. **Biochemia Medica**, v. 21, n. 2, p. 203–209, 2011. ISSN 18467482. DOI: 10.11613/BM.2011.029.

MESSERSCHMITT, David G. **Stationary points of a real-valued function of a complex variable**. 2006. p. 1–10. Disponível em: <<http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-93.html>>.

MIRZAHASANLOO, Taher S *et al.* Environment-adaptive speech enhancement for bilateral cochlear implants using a single processor. **Speech Communication**, v. 55, n. 4, p. 523–534, 2013. ISSN 01676393. DOI: 10.1016/j.specom.2012.10.004.

NAYLOR, Patrick A *et al.* Models, measurement and evaluation. *In: SPEECH Dereverberation*. Springer, 2010. p. 21–56.

NEIL, Jahn N O; RYUGO, David K. Synaptic plasticity and auditory function. *In: UMAT, Cila (Ed.). Cochlear Implant Research Updates*. 2012. p. 167–192. ISBN 978-953-51-0582-4.

NIKIAS, Chrysostomos L; SHAO, Min. **Signal processing with alpha-stable distributions and applications**. Wiley-Interscience, 1995.

OPENSTAX, CNX. **Anatomy of the human ear**. 2016. https://commons.wikimedia.org/wiki/File:Figure_36_04_02.jpg.

PADILLA, Monica; LANDSBERGER, David M. Reduction in spread of excitation from current focusing at multiple cochlear locations in cochlear implant users. **Hearing Research**, Elsevier, 2016. ISSN 03785955. DOI: 10.1016/j.heares.2016.01.002.

PAPOULIS, Athanasios; PILLAI, S Unnikrishna. **Probability, random variables and stochastic processes**. 4th: McGraw-Hill Europe, 2002.

POULARIKAS, Alexander D. Hilbert transform. *In: POULARIKAS, Alexander D. (Ed.). The Handbook of Formulas and Tables for Signal Processing*. CRC Press, 1999. ISBN 978-3-540-64834-5.

QAZI, Obaid Ur Rehman *et al.* Speech understanding performance of cochlear implant subjects using time-frequency masking-based noise reduction. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 59, n. 5, p. 1364–73, 2012. ISSN 1558-2531. DOI: 10.1109/TBME.2012.2187650.

RAO, V Bhujanga; RAMAIAH, P Seetha; KUMAR, K Raja. A practical approach to speech processor for auditory prosthesis using DSP and FPGA. **ACM SIGBED Review**, v. 9, n. 1, p. 5–10, 2012. DOI: 10.1145/2188376.2188377.

SANTOS, João F. **A non-intrusive objective speech intelligibility metric tailored for cochlear implant users in complex listening environments**. 2013.

SANTOS, João F; FALK, Tiago H. Updating the SRMR-CI metric for improved intelligibility prediction for cochlear implant users. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, IEEE, v. 22, n. 12, p. 2197–2206, 2014. ISSN 2329-9290. DOI: 10.1109/TASLP.2014.2363788.

SANTOS, João F *et al*. Objective speech intelligibility measurement for cochlear implant users in complex listening environments. **Speech Communication**, v. 55, n. 7-8, p. 815–824, 2013. ISSN 01676393. DOI: 10.1016/j.specom.2013.04.001.

SANTOS, João F *et al*. Performance comparison of intrusive objective speech intelligibility and quality metrics for cochlear implant users. *In*: INTERSPEECH. 2012. p. 5–8.

SERIZEL, Romain *et al*. Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with application in cochlear implants. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, IEEE, v. 22, n. 4, p. 785–799, 2014. ISSN 2329-9290. DOI: 10.1109/TASLP.2014.2304240.

SERVETTI, Antonio; MARTIN, Juan Carlos de. Error tolerant MAC extension for speech communications over 802.11 WLANs. *In*: IEEE 61st Vehicular Technology Conference. IEEE, 2005. p. 2330–2334. DOI: 10.1109/VETECS.2005.1543751.

SHAPIRO, William H; BRADHAM, Tamala S. Cochlear implant programming. **Otolaryngologic clinics of North America**, Elsevier, v. 45, n. 1, p. 111–27, 2012. ISSN 1557-8259. DOI: 10.1016/j.otc.2011.08.020.

SILVA, Luiz Felipe da; BERMUDEZ, José C M. Speech enhancement using a frame adaptive gain function for Wiener filtering. *In*: 2. PROCEEDINGS of the IEEE Workshop on Statistical Signal Processing. 2011. p. 389–392. DOI: 10.1109/SSP.2011.5967712.

SISTEMA DE CONSELHOS DE FONOAUDIOLOGIA. **Guia de orientações na avaliação audiológica básica**. 2017. p. 7–30. ISBN 9780126434514.

SLANEY, Malcolm. **An efficient implementation of the auditory filter bank**. v. 1. 1993. p. 40. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.81.4406%7B%5C%7Drep=rep1%7B%5C%7Dtype=pdf>>.

SOMEK, Branko *et al*. Coding strategies for cochlear implants. **Automatika**, v. 47, p. 69–74, 2006.

SRINIVASAN, Arthi G *et al.* Improving speech perception in noise with current focusing in cochlear implant users. **Hearing research**, Elsevier, v. 299, p. 29–36, 2013. ISSN 1878-5891. DOI: 10.1016/j.heares.2013.02.004.

STRYDOM, Trudie *et al.* The performance of different synthesis signals in acoustic models of cochlear implants. **The Journal of the Acoustical Society of America**, v. 129, n. 2, p. 920–933, 2011. ISSN 0001-4966. DOI: 10.1121/1.3518760.

TAAL, Cees H; HENDRIKS, Richard C; HEUSDENS, Richard. Matching pursuit for channel selection in cochlear implants based on an intelligibility metric. *In: EUROPEAN Signal Processing Conference*. 2012. p. 504–508.

TEFILI, Diego *et al.* Cochlear implants: technological aspects and socioeconomic role. **Revista Brasileira de Engenharia Biomédica**, v. 29, n. 4, p. 414–433, 2013. ISSN 15173151. DOI: 10.4322/rbeb.2013.039.

THIERGART, Oliver; DEL GALDO, Giovanni; HABETS, Emanuël A P. Signal-to-reverberant ratio estimation based on the complex spatial coherence between omnidirectional microphones. *In: PROCEEDINGS of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 2012. p. 309–312. DOI: 10.1109/ICASSP.2012.6287878.

TOLEDO, F; LOIZOU, Philipos C; LOBO, Arthur P. Subspace and envelope subtraction algorithms for noise reduction in cochlear implants. *In: PROCEEDINGS of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2003. p. 2002–2005. DOI: 10.1109/IEMBS.2003.1280126.

TSAO, Yu; LAI, Ying Hui. Generalized maximum a posteriori spectral amplitude estimation for speech enhancement. **Speech Communication**, Elsevier, v. 76, p. 112–126, 2016. ISSN 01676393. DOI: 10.1016/j.specom.2015.10.003.

UTD, University of Texas at Dallas. **PDA based Research Platform for Cochlear Implant Speech Processing Strategies**: Technical Reference Manual. 2010.

VERMEIRE, Katrien; PUNTE, Andrea Kleine; VAN DE HEYNING, Paul. Better speech recognition in noise with the fine structure processing coding strategy. **Journal for Oto-Rhino-Laryngology, Head and Neck Surgery**, v. 72, n. 6, p. 305–11, 2010. ISSN 1423-0275. DOI: 10.1159/000319748.

WANG, DeLiang; BROWN, Guy J. **Computational auditory scene analysis: Principles, algorithms, and applications**. New York, NY: Wiley/IEEE Press, 2006. p. 395. ISBN 978-0-471-74109-1.

WANG, Deliang *et al.* Speech perception of noise with binary gains. **The Journal of the Acoustical Society of America**, v. 124, n. 4, p. 2303–2307, 2008. ISSN 1520-8524. DOI: 10.1121/1.2967865.

WILLIGES, Ben *et al.* Spatial release from masking in simulated cochlear implant users with and without access to low-frequency acoustic hearing. **Trends in Hearing**, v. 19, p. 1–14, 2015. ISSN 23312165. DOI: 10.1177/2331216515616940.

WILSON, Blake S. The modern cochlear implant. **IEEE Pulse**, v. 29, March/April, p. 29–32, 2017. ISSN 00306622. DOI: 10.3950/jibiinkoka.92.10sokai_1755.

WILSON, Blake S; DORMAN, Michael F. Cochlear implants: Current designs and future possibilities. **The Journal of Rehabilitation Research and Development**, v. 45, n. 5, p. 695–730, 2008a. ISSN 07487711. DOI: 10.1682/JRRD.2007.10.0173.

WILSON, Blake S; DORMAN, Michael F. Interfacing sensors with the nervous system: Lessons from the development and success of the cochlear implant. **IEEE Sensors Journal**, IEEE, v. 8, n. 1-2, p. 131–147, 2008b. ISSN 1530-437X. DOI: 10.1109/JSEN.2007.912917.

WILSON, Blake S; DORMAN, Michael F. The surprising performance of present-day cochlear implants. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 54, n. 6, p. 969–972, 2007. ISSN 0018-9294. DOI: 10.1109/TBME.2007.893505.

WILSON, Blake S *et al.* Design and evaluation of a continuous interleaved sampling (CIS) processing strategy for multichannel cochlear implants. **Journal of Rehabilitation Research and Development**, v. 30, n. 1, p. 110–116, 1993. ISSN 0007506X.

WILSON, Kevin W *et al.* Speech denoising using nonnegative matrix factorization with priors. *In: IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008. p. 4029–4032. DOI: 10.1109/ICASSP.2008.4518538.

WOUTERS, Jan; MCDERMOTT, Hugh J; FRANCAERT, Tom. Sound coding in cochlear implants: From electric pulses to hearing. **Signal Processing Magazine, IEEE**, v. 32, n. 2, p. 67–80, 2015. ISSN 1053-5888. DOI: 10.1109/MSP.2014.2371671.

WOUTERS, Jan *et al.* Sound processing for better coding of monaural and binaural cues in auditory prostheses. **Proceedings of the IEEE**, IEEE, v. 101, n. 9, p. 1986–1997, 2013. ISSN 0018-9219. DOI: 10.1109/JPROC.2013.2257635.

YANG, Li-Ping; FU, Qian-Jie. Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. **The Journal of the Acoustical Society of America**, v. 117, 3 Pt 1, p. 1001–1004, 2005. ISSN 00014966. DOI: 10.1121/1.1852873.

YOUSEFIAN, Nima; LOIZOU, Philipos C. A dual-microphone speech enhancement algorithm based on the coherence function. **IEEE Transactions on Audio, Speech and Language Processing**, IEEE, v. 20, n. 2, p. 599–609, 2012. ISSN 1558-7916. DOI: 10.1109/TASL.2011.2162406.

ZENG, Fan-Gang; POPPER, Arthur N; FAY, Richard R. **Cochlear implants: Auditory prostheses and electric hearing**. 2004. ISBN 9781441923462.

ZENG, Fan-Gang *et al.* Cochlear implants: system design, integration, and evaluation. **IEEE Reviews in Biomedical Engineering**, IEEE, v. 1, n. 600, p. 115–142, 2008. ISSN 1937-3333. DOI: 10.1109/RBME.2008.2008250.

ZENG, Fan-Gang *et al.* Speech dynamic range and its effect on cochlear implant performance. **The Journal of the Acoustical Society of America**, v. 111, n. 1, p. 377, 2002. ISSN 00014966. DOI: 10.1121/1.1423926.

APÊNDICE A – DIFERENCIAÇÃO DE $\mathbf{z}^H \mathbf{A} \mathbf{z}$, $\mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^H \mathbf{B} \mathbf{z}$ E $(\mathbf{z}^H \mathbf{A} \mathbf{z})^2$

A diferenciação a seguir é realizada de acordo com Hjørungnes *et al.* (2007).

Considerando-se o vetor $\mathbf{z} \in \mathbb{C}^N$ e \mathbf{A} e \mathbf{B} matrizes $N \times N$ constantes, o diferencial de $f(\mathbf{z}, \mathbf{z}^*)$ é dado pelos termos de primeira ordem de $\mathcal{S} = f(\mathbf{z} + d\mathbf{z}, \mathbf{z}^* + d\mathbf{z}^*) - f(\mathbf{z}, \mathbf{z}^*)$.

A.1 $f_1(\mathbf{z}, \mathbf{z}^*) = \mathbf{z}^H \mathbf{A} \mathbf{z}$

Para a primeira função analisada, tem-se:

$$\mathcal{S} = (\mathbf{z} + d\mathbf{z})^H \mathbf{A} (\mathbf{z} + d\mathbf{z}) - \mathbf{z}^H \mathbf{A} \mathbf{z} \quad (204)$$

$$= \mathbf{z}^H \mathbf{A} \mathbf{z} + d\mathbf{z}^H \mathbf{A} \mathbf{z} + \mathbf{z}^H \mathbf{A} d\mathbf{z} + d\mathbf{z}^H \mathbf{A} d\mathbf{z} - \mathbf{z}^H \mathbf{A} \mathbf{z} \quad (205)$$

$$= d\mathbf{z}^H \mathbf{A} \mathbf{z} + \mathbf{z}^H \mathbf{A} d\mathbf{z} + d\mathbf{z}^H \mathbf{A} d\mathbf{z}. \quad (206)$$

Selecionando-se apenas as parcelas de primeira ordem, obtém-se o diferencial de $df_1(\mathbf{z}, \mathbf{z}^*)$.

$$df_1(\mathbf{z}, \mathbf{z}^*) = d\mathbf{z}^H \mathbf{A} \mathbf{z} + \mathbf{z}^H \mathbf{A} d\mathbf{z} \quad (207)$$

As derivadas em relação à \mathbf{z} e \mathbf{z}^* são, respectivamente:

$$\frac{\partial f_1(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}} = \mathbf{z}^H \mathbf{A} \quad (208)$$

$$\frac{\partial f_1(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^*} = \mathbf{z}^T \mathbf{A}^T. \quad (209)$$

A.2 $f_2(\mathbf{z}, \mathbf{z}^*) = \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^H \mathbf{B} \mathbf{z}$

Para a segunda função analisada, tem-se:

$$\mathcal{S} = (\mathbf{z} + d\mathbf{z})^H \mathbf{A} (\mathbf{z} + d\mathbf{z}) (\mathbf{z} + d\mathbf{z})^H \mathbf{B} (\mathbf{z} + d\mathbf{z}) - f_2(\mathbf{z}, \mathbf{z}^*) = \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^H \mathbf{B} \mathbf{z} \quad (210)$$

$$= \left((\mathbf{z} + d\mathbf{z})^H \mathbf{A} (\mathbf{z} + d\mathbf{z}) \right) \left((\mathbf{z} + d\mathbf{z})^H \mathbf{B} (\mathbf{z} + d\mathbf{z}) \right) - \left(\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} \mathbf{z} \right) \quad (211)$$

$$= \left(\mathbf{z}^H \mathbf{A} \mathbf{z} + \mathbf{z}^H \mathbf{A} d\mathbf{z} + d\mathbf{z}^H \mathbf{A} \mathbf{z} + d\mathbf{z}^H \mathbf{A} d\mathbf{z} \right) \times \left(\mathbf{z}^H \mathbf{B} \mathbf{z} + \mathbf{z}^H \mathbf{B} d\mathbf{z} + d\mathbf{z}^H \mathbf{B} \mathbf{z} + d\mathbf{z}^H \mathbf{B} d\mathbf{z} \right) - \left(\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} \mathbf{z} \right) \quad (212)$$

$$= \left(\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} \mathbf{z} \right) + \left(\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} d\mathbf{z} \right) + \left(\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(d\mathbf{z}^H \mathbf{B} \mathbf{z} \right) + \left(\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(d\mathbf{z}^H \mathbf{B} d\mathbf{z} \right) + \left(\mathbf{z}^H \mathbf{A} d\mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} \mathbf{z} \right) + \left(\mathbf{z}^H \mathbf{A} d\mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} d\mathbf{z} \right) + \left(\mathbf{z}^H \mathbf{A} d\mathbf{z} \right) \left(d\mathbf{z}^H \mathbf{B} \mathbf{z} \right) + \left(\mathbf{z}^H \mathbf{A} d\mathbf{z} \right) \left(d\mathbf{z}^H \mathbf{B} d\mathbf{z} \right) + \left(d\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} \mathbf{z} \right) + \left(d\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} d\mathbf{z} \right) + \left(d\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(d\mathbf{z}^H \mathbf{B} \mathbf{z} \right) + \left(d\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(d\mathbf{z}^H \mathbf{B} d\mathbf{z} \right) + \left(d\mathbf{z}^H \mathbf{A} d\mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} \mathbf{z} \right) + \left(d\mathbf{z}^H \mathbf{A} d\mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} d\mathbf{z} \right) + \left(d\mathbf{z}^H \mathbf{A} d\mathbf{z} \right) \left(d\mathbf{z}^H \mathbf{B} \mathbf{z} \right) + \left(d\mathbf{z}^H \mathbf{A} d\mathbf{z} \right) \left(d\mathbf{z}^H \mathbf{B} d\mathbf{z} \right) - \left(\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} \mathbf{z} \right). \quad (213)$$

Selecionando-se apenas as parcelas de primeira ordem, obtém-se o diferencial de $df_2(\mathbf{z}, \mathbf{z}^*)$.

$$\begin{aligned} df(\mathbf{z}, \mathbf{z}^*) &= \left(\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} d\mathbf{z} \right) + \left(\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(d\mathbf{z}^H \mathbf{B} \mathbf{z} \right) \\ &\quad + \left(\mathbf{z}^H \mathbf{A} d\mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} \mathbf{z} \right) + \left(d\mathbf{z}^H \mathbf{A} d\mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} \mathbf{z} \right) \\ &= \left(\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(\mathbf{z}^H \mathbf{B} d\mathbf{z} \right) + \left(\mathbf{z}^H \mathbf{A} \mathbf{z} \right) \left(\mathbf{z}^T \mathbf{B}^T d\mathbf{z}^* \right) \\ &\quad + \left(\mathbf{z}^H \mathbf{B} \mathbf{z} \right) \left(\mathbf{z}^H \mathbf{A} d\mathbf{z} \right) + \left(\mathbf{z}^H \mathbf{B} \mathbf{z} \right) \left(\mathbf{z}^T \mathbf{A}^T d\mathbf{z}^* \right) \end{aligned} \quad (214)$$

As derivadas em relação à \mathbf{z} e \mathbf{z}^* são, respectivamente:

$$\frac{\partial f_2(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}} = \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^H \mathbf{B} + \mathbf{z}^H \mathbf{B} \mathbf{z} \mathbf{z}^H \mathbf{A} \quad (215)$$

$$\frac{\partial f_2(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^*} = \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^T \mathbf{B}^T + \mathbf{z}^H \mathbf{B} \mathbf{z} \mathbf{z}^T \mathbf{A}^T. \quad (216)$$

A.3 $f_3(\mathbf{z}, \mathbf{z}^*) = (\mathbf{z}^H \mathbf{A} \mathbf{z})^2$

A função $f_3(\mathbf{z}, \mathbf{z}^*) = (\mathbf{z}^H \mathbf{A} \mathbf{z})^2$ é um caso particular da $f_2(\mathbf{z}, \mathbf{z}^*) = \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^H \mathbf{B} \mathbf{z}$ em que a matriz $\mathbf{B} = \mathbf{A}^H$. Desta forma, as derivadas em relação à \mathbf{z} e \mathbf{z}^* são, respectivamente:

$$\frac{\partial f_3(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}} = \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^H \mathbf{A}^H + \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^H \mathbf{A} \quad (217)$$

$$\frac{\partial f_3(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^*} = \mathbf{z}^H \mathbf{A}^H \mathbf{z} \mathbf{z}^T \mathbf{A}^* + \mathbf{z}^H \mathbf{A}^H \mathbf{z} \mathbf{z}^T \mathbf{A}^T. \quad (218)$$

As equações (217) e (218) representam as derivadas da função $f_3(\mathbf{z}, \mathbf{z}^*)$ para uma matriz $\mathbf{A} N \times N$ qualquer.

Para o caso particular em que $\mathbf{A} = \mathbf{a} \mathbf{a}^H$, nota-se a função como $f_3^{(1)}(\mathbf{z}, \mathbf{z}^*)$, e tem-se:

$$\mathbf{A}^H = (\mathbf{a} \mathbf{a}^H)^H = \mathbf{a}^{HH} \mathbf{a}^H = \mathbf{a} \mathbf{a}^H = \mathbf{A} \quad (219)$$

$$\mathbf{A}^* = (\mathbf{a} \mathbf{a}^H)^* = \mathbf{a}^* \mathbf{a}^T \quad (220)$$

$$\mathbf{A}^T = (\mathbf{a} \mathbf{a}^H)^T = \mathbf{a}^* \mathbf{a}^T = \mathbf{A}^*. \quad (221)$$

Nesta situação específica, as equações (217) e (218) podem ser reescritas da seguinte forma:

$$\frac{\partial f_3^{(1)}(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}} = \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^H \mathbf{A} + \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^H \mathbf{A} = 2 \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^H \mathbf{A} \quad (222)$$

$$\frac{\partial f_3^{(1)}(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^*} = \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^T (\mathbf{A}^* + \mathbf{A}^T) = 2 \mathbf{z}^H \mathbf{A} \mathbf{z} \mathbf{z}^T \mathbf{A}^T. \quad (223)$$

APÊNDICE B – CÁLCULO DAS MÉTRICAS OBJETIVAS DE QUALIDADE E INTELIGIBILIDADE DA FALA

Neste apêndice são detalhadas as métricas associadas à qualidade (PESQ) e inteligibilidade (NCM e SRMR-CI) utilizadas nos capítulos 3, e 4.

B.1 PESQ

A métrica objetiva de qualidade mais utilizada em estudos de ICs é, sem dúvidas, a Avaliação Perceptual da Qualidade da Fala (*Perceptual Evaluation of Speech Quality*) (PESQ). É o método recomendado pela União Internacional de Telecomunicações (ITU) (INTERNATIONAL TELECOMMUNICATION UNION, 2001) e é provavelmente a métrica objetiva mais confiável para se prever a qualidade da fala (COSENTINO *et al.*, 2012). Foi desenvolvida para avaliar fala transmitida por redes de telecomunicações. Originalmente, definida para sinais limitados em banda (*narrowband*), em 2007 foi adaptada para sinais de banda larga (*wideband*) (INTERNATIONAL TELECOMMUNICATION UNION, 2007).

A métrica é baseada em um modelo sensorial da audição humana e é calculada através de uma combinação linear de dois fatores relacionados à distorção da fala: a perturbação média (D_{ind}), e perturbação média assimétrica (A_{ind}), de acordo com a equação 224 (KOKKINAKIS; LOIZOU, 2011). Uma descrição detalhada do cálculo da PESQ é encontrada em Loizou (2013).

$$PESQ = a_0 + a_1 D_{ind} + a_2 A_{ind} \quad (224)$$

A faixa de pontuação PESQ é de -0,5 a 4,5. Os coeficientes são $a_0 = 4,5$, $a_1 = -0,1$ e $a_2 = -0,0309$, resultados de otimização usando-se dados convencionais de telefonia (por exemplo, VoIP, wireless) (SANTOS *et al.*, 2012), de forma a haver boa correlação com resultados de avaliação subjetiva usando uma escala MOS (LOIZOU, 2013),

B.2 NCM

A métrica NCM é baseada na covariância entre o sinal de referência (limpo) e o sinal sob análise (CHEN; LOIZOU, 2011). O cálculo da métrica se dá da seguinte forma: o sinal é decomposto em 23 canais por um banco de filtros *gammatone*, e o envelope temporal de cada canal é calculado usando a transformada de Hilbert. Com a informação do coeficiente de correlação entre os envelopes do sinal de referência e do sinal sob análise, r_k , de cada canal k , estima-se a chamada *SNR aparente*, SNR_{app}

(SANTOS *et al.*, 2013).

$$\text{SNR}_{app}(k) = \left[10 \log_{10} \left(\frac{r_k}{1-r_k} \right) \right]_{-15,15} \quad (225)$$

O operador $[\]_{-15,15}$ limita a SNR_{app} no intervalo $[-15, +15]$. Em seguida, mapeia-se a SNR_{app} entre 0 e 1, obtendo o índice de transmissão (TI_k) de acordo com a equação (226).

$$\text{TI}_k = \frac{\text{SNR}_{app}(k) + 15}{30} \quad (226)$$

Finalmente, calcula-se a NCM como uma média ponderada das TIs. W_k é o peso relativo à importância da banda k .

$$\text{NCM} = \frac{1}{\sum_{k=1}^{23}} \sum_{k=1}^{23} W_k \text{TI}_k \quad (227)$$

Esta métrica mostrou boa correlação com a inteligibilidade de implantados (SANTOS *et al.*, 2013) e foi usada em Chen e Loizou (2011), Cosentino *et al.* (2014), Goehring *et al.* (2016), Hazrati e Loizou (2012) e Taal, Hendriks e Heusdens (2012).

B.3 SRMR-CI

A SRMR-CI é uma recente adaptação da métrica Relação de Energia de Modulação entre Fala e Reverberação (*Speech-to-reverberation modulation energy ratio*) (SRMR) (FALK, 2010) para o caso de implantes cocleares (SANTOS; FALK, 2014). Esta é a única métrica proposta que considera as especificidades desta aplicação.

O cálculo da SRMR-CI é realizado em quatro estágios. Primeiramente, o sinal de entrada $x(n)$ é decomposto em diferentes faixas de frequência, através de um banco de filtros de 22 canais, o mesmo usado nos implantes cocleares *Nucleus* da CochlearTM. Em seguida, os envelopes temporais $e_k(n)$ de cada saída do banco de filtros, $x_k(n)$ (com $k = 1, 2, \dots, 22$) são calculados usando a transformada de Hilbert. Toma-se a STFT dos envelopes temporais, obtendo-se a energia de modulação espectral para cada banda, $E_k(m, k)$, em que m e k são respectivamente os índices de quadro (*frame*) e frequência. Em um terceiro passo, o espectro de modulação é filtrado por um banco de 8 filtros com frequências centrais espaçadas logaritmicamente entre 4Hz e 64Hz. Finalmente, calcula-se a SRMR-CI como a razão entre a média da energia de modulação (por todos os quadros) das quatro primeiras bandas de modulação e a média da energia de modulação nas bandas restantes.

Maiores detalhes no cálculo da SRMR-CI são encontrados em Santos e Falk (2014), Santos *et al.* (2013) e Falk (2010).

É uma métrica não-intrusiva, ou seja, não necessita do sinal limpo de referência. Isso é uma vantagem, pois permite sua utilização no processamento em tempo-real de

dispositivos clínicos como informação para adaptação de algoritmos de aprimoramento da fala.

APÊNDICE C – PLATAFORMA DE PESQUISA EM IMPLANTES COCLEARES - UNIVERSITY OF TEXAS AT DALLAS

Neste apêndice serão detalhadas as características técnicas da plataforma de pesquisa em implantes cocleares desenvolvida e cedida pela UTD.

C.1 CARACTERÍSTICAS DA PLATAFORMA DE PESQUISA EM ICS

Os principais fatores que motivaram o desenvolvimento da plataforma são a falta de ferramentas de pesquisa para ICs e o fato dos processadores clínicos e comerciais serem de código proprietário e permitirem apenas a modificação de funcionalidades limitadas. Essa plataforma visa a possibilitar a implementação e avaliação de novas estratégias de processamento de sinais para ICs e a avaliação extensa e crônica de novas ideias de pesquisa (ALI *et al.*, 2013).

Essa plataforma tem flexibilidade para a estimulação uni e bilateral sincronizada (biauricular), além de permitir a estimulação exclusivamente elétrica ou a estimulação eletroacústica (EAS).

Dois modos operacionais são suportados: tempo-real e bancada (*bench-top*). O primeiro é similar a um processador clínico tipo portado no corpo (*body-worn*) para experimentos em campo livre, permitindo também testes em que os pacientes levam o dispositivo e usam o processamento investigado durante um determinado período (*take-home trials*). Nesta situação, um *tablet* ou *smartphone* funciona como processador (ver Figura 50). Por outro lado, o modo de bancada permite que a plataforma seja usada em modo “desconectado” (*offline*) para condução de experimentos em laboratório usando áudios pré-processados.

A plataforma apresenta flexibilidade na implementação de algoritmos de codificação do som como aplicativos, permitindo o uso de diversas linguagens de programação, tais como MATLAB, C, C++, C#, JAVA, LabView.

O dispositivo pode ser conectado a um computador, tablet ou smartphone via Wi-Fi ou USB.

C.2 HARDWARE

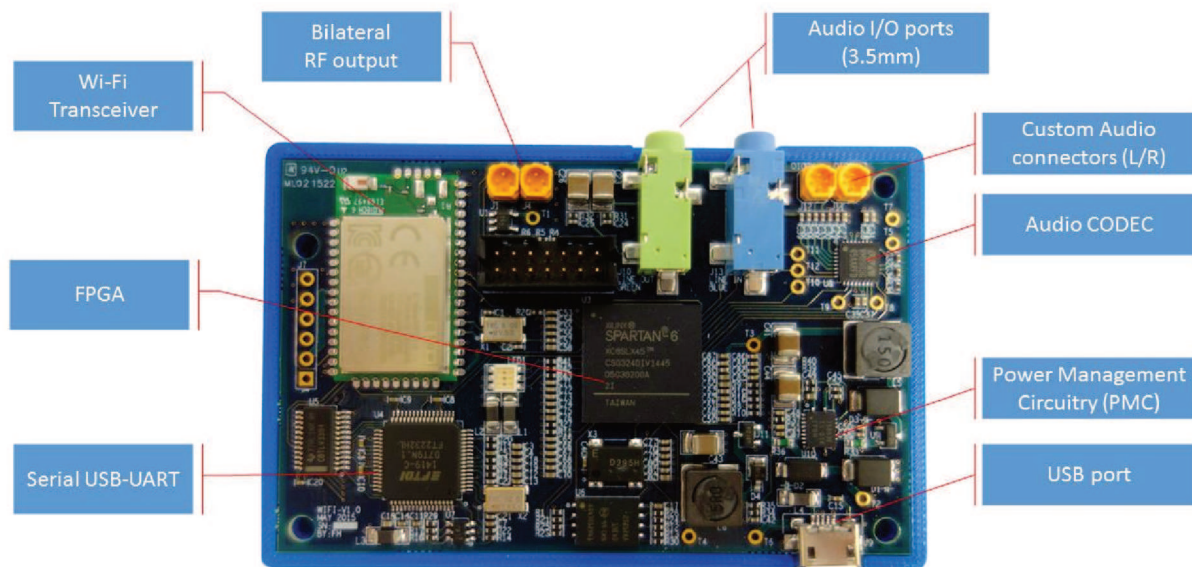
A Figura 51 mostra uma fotografia do *hardware* da plataforma de pesquisas em ICs. Os blocos constituintes estão destacados. O detalhamento desses blocos é realizado na sequência, sendo que as informações são obtidas de Ali (2016).

A alimentação da plataforma se dá via porta USB e é controlada pelo circuito de gerenciamento de energia (*Power Management Circuitry* – PMC). A plataforma dispõe de duas entradas de áudio customizadas *Custom Audio Connectors (L/R)*, nas quais são conectados os microfones dos processadores BTE (direito e esquerdo);

Figura 50 – Fotografia da plataforma de pesquisas em ICs conectada via Wi-Fi a um *smartphone*, que funciona como processador. A estratégia de codificação do som é programada como um aplicativo no telefone. Fonte: Hazrati *et al.* (2016)



Figura 51 – Fotografia do *hardware* da plataforma de pesquisas em ICs, com destaque para seus blocos funcionais. Fonte: Ali (2016)



duas portas de entrada/saída de áudio (*Audio I/O ports*) de 3,5mm (tipo P2) e dois conectores de saída para as antenas RF direita e esquerda (*Bilateral RF Outputs*). Os outros blocos são especificados a seguir.

C.2.1 FPGA

O processador central é um Arranjo de Portas Programável em Campo (FPGA) modelo XC6SLX45, da Xilinx. A FPGA controla o fluxo de dados no sistema. Ela é responsável pela captação sincronizada dos quadros (*frames*) de áudio pelo *codec*, pelo envio e recebimento de dados pela unidade de processamento (*tablet, smartphone* ou computador), por codificar os dados em RF, e pela transmissão sincronizada dos pulsos de estimulação para as antenas RF.

Os blocos de transmissão Wi-Fi e USB se comunicam com a FPGA usando o protocolo padrão UART, enquanto o *codec* de áudio usa o protocolo SPI para troca de dados. Além da codificação RF, o programa (*firmware*) da FPGA realiza verificações de segurança para evitar estimulações que possam causar danos aos implantados. A FPGA é programada em *Verilog* e é fechada, não sendo acessível aos pesquisadores.

C.2.2 Codec de áudio

O *codec* de áudio da *Wolfson Microelectronics* (WM8983) é um dispositivo de entrada e saída altamente integrado desenvolvido para sistemas de computação e comunicação embarcados. Ele integra pré-amplificadores para microfones diferenciais estéreo, e contém *drivers* para alto-falante, fone de ouvido e saída de linha diferencial ou estéreo. O processamento de sinais integrado (*on chip*) contém um equalizador de 5 bandas, um controle de ganho automático para entrada de microfone ou de linha através do ADC, assim como uma função limitadora puramente digital para gravação ou reprodução. Um filtro passa-altas programável no caminho do ADC é disponível para redução do ruído de vento e um filtro IIR com coeficientes programáveis pode ser usado como um filtro rejeita-faixa (*notch*) para supressão de ruído de frequência fixa. As principais características do *codec* são apresentadas na tabela a seguir.

C.2.3 Transceptor Wi-Fi

O transceptor da Bluegiga (WF121) é um módulo Wi-Fi *stand-alone* que possui um rádio 2,4GHz (IEEE802.11 b/n/g) completamente integrado, com potência de transmissão de +17dBm e sensibilidade de -97dBm.

O servidor conta com interfaces UART 20Mbps e USB *on-the-go*. Diversas interfaces periféricas flexíveis, tais como SPR, ADC, GPIO, Bluetooth, AIO e temporizadores, I2C, SPI e UART, Ethernet, podem ser conectadas diretamente ao módulo.

Tabela 22 – Características principais do *codec* da plataforma.

Parâmetro	Valor
ADC	4 canais
SNR	87dB
AGC	Sim
Interfaces	PCM, I ² S
Alimentação	1,8 – 3,6V
Encapsulamento	7 × 7mm 48-pin QFN

A interface conta com os seguintes protocolos TCP/IP e 802.11 MAC embarcados: IP, TCP, UDP, DHCP e protocolos DNS, protocolo de *host* BGAPI para utilização como modem. Linguagem de script BGScriptTM ou desenvolvimento C nativo para aplicações autônomas também são disponíveis.

O módulo possui um micro-controlador (MCU) de 32-bits embarcado com 80MHz, RAM de 128kB e memória *flash* de 512kB, e arquitetura MIPS.

C.2.4 Interface USB

O circuito da plataforma pode se comunicar diretamente com computadores ou outros dispositivos portáteis baseados em USB, através de um conector micro-USB. Para isso é utilizado o FT232H, um circuito integrado (CI) de USB 2.0 de alta velocidade (480Mb/s) para UART/FIFO.

C.2.5 Gerenciamento de Energia

O circuito funciona com uma bateria de 5V que se conecta à placa por uma porta micro-USB. O circuito de gerenciamento de energia possui dois componentes principais: TPS75003 e TPS73663. O TPS75003 é um CI de gerenciamento de uma fonte de alimentação integrada tripla para a FPGA Xilinx Spartan e regula níveis de tensão e corrente da parte digital da placa do circuito. O TPS73663 é um regulador *low-dropout* (LDO) e é usado para prover alimentação estável para a circuitaria analógica RF. O uso de dois reguladores separados assegura uma menor interferência cruzada entre os circuitos digital e analógico.

APÊNDICE D – MAPA PADRÃO DE USUÁRIO DE IC

```

1 % MAP file structure for CCI Platform
2
3 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
4 %
5 % Copyright: CRSS-CILab, UT-Dallas
6 %   Authors: Hussnain Ali
7 %     Date: 2015/09/28
8 %
9 % Modified by Rafael Chiea on 2018/12/10
10 % @ LPDS-UFSC, Florianopolis/SC
11 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
12
13
14 MAP.General.SubjectName      = 'Subject 01';      % Optional:
    Subject Name
15 MAP.General.SubjectID       = 'S01';                % Optional:
    Random Subject ID
16 MAP.General.MapTitle        = 'S01_ACE_900Hz';      % Optional: Map
    Title
17
18 MAP.General.NumberOfImplants = 2;                  % '1' for Unilateral and
    '2' for Bilateral
19 MAP.General.ImplantedEar     = 'Left';             % 'Left' for left side; '
    Right' for right side; 'Bilateral' for both sides
20 MAP.General.StimulateEars    = 'Left';             % 'Left' for left only; '
    Right' for right only; 'Both'|'Bilateral' for both sides; 'NULL
    ' for no stimulation
21
22 %% Left Ear Parameters
23 % remove this section if left side does not exist
24 MAP.Left.ImplantType         = 'CI24RE';           %Implant chip type, e
    .g., CI24RE(CS/CA), CI24R, CI24M, CI22M, ST
25 MAP.Left.SamplingFrequency   = 16000;             % Fixed
26 MAP.Left.NumberOfChannels    = 22;                % 22 fixed for
    implants from Cochlear Ltd.
27 MAP.Left.Strategy            = 'ACE';              % 'ACE' or 'CIS' or '
    Custom'
28 MAP.Left.Nmaxima             = 8;                  % Nmaxima 1 - 22 for
    n-of-m strategies

```



```

29 MAP.Left.StimulationMode      = 'MP1+2';          % Electrode
    Configuration/Stimulation mode e.g., MP1, MP1+2, BP1, BP1+2, CG
    ,.....etc.
30 MAP.Left.StimulationRate      = 1000;              % Stimulation rate
    per electrode in number of pulses per second (pps)
31 MAP.Left.PulseWidth           = 25;                % Pulse width in us
32 MAP.Left.IPG                  = 8;                % Inter-Phase Gap (
    IPG) fixed at 8us (could be variable in future)
33 MAP.Left.Sensitivity           = 2.3;              % Microphone
    Sensitivity (adjustable in GUI)
34 MAP.Left.Gain                  = 25;              % Global gain for
    envelopes in dB - standard is 25dB (adjustable in GUI)
35 MAP.Left.Volume                = 10;              % Volume Level on a
    scale of 0 to 10; 0 being lowest and 10 being highest (
    adjustable in GUI)
36 MAP.Left.Q                     = 20;              % Q-factor for the
    compression function
37 MAP.Left.BaseLevel             = 0.0156;          % Base Level
38 MAP.Left.SaturationLevel       = 0.5859;          % Saturation Level
39 MAP.Left.ChannelOrderType      = 'base-to-apex';  % Channel
    Stimulation Order type: 'base-to-apex' or 'apex-to-base'
40 MAP.Left.FrequencyTable        = 'Default';       % Frequency
    assignment for each band "Default" or "Custom"
41 MAP.Left.Window                 = 'Hanning';      % Window type
42 MAP.Left.El_CF1_CF2_THR_MCL_Gain = [
43   % El   F_Low   F_High  THR    MCL    Gain
44   22  188    313    100    200    0.0
45   21  313    438    100    200    0.0
46   20  438    563    100    200    0.0
47   19  563    688    100    200    0.0
48   18  688    813    100    200    0.0
49   17  813    938    100    200    0.0
50   16  938   1063    100    200    0.0
51   15 1063   1188    100    200    0.0
52   14 1188   1313    100    200    0.0
53   13 1313   1563    100    200    0.0
54   12 1563   1813    100    200    0.0
55   11 1813   2063    100    200    0.0
56   10 2063   2313    100    200    0.0
57   9  2313   2688    100    200    0.0
58   8  2688   3063    100    200    0.0
59   7  3063   3563    100    200    0.0

```

```

60     6     3563     4063     100     200     0.0
61     5     4063     4688     100     200     0.0
62     4     4688     5313     100     200     0.0
63     3     5313     6063     100     200     0.0
64     2     6063     6938     100     200     0.0
65     1     6938     7938     100     200     0.0
66     ];
67 % Number of active electrodes/bands
68 MAP.Left.NumberOfBands           = size(MAP.Left.
        El_CF1_CF2_THR_MCL_Gain, 1);
69 % Active Electrodes
70 MAP.Left.Electrodes              = MAP.Left.
        El_CF1_CF2_THR_MCL_Gain(:, 1);
71 % Low cut-off frequencies of filters
72 MAP.Left.LowerCutOffFrequencies = MAP.Left.
        El_CF1_CF2_THR_MCL_Gain(:, 2);
73 % Upper cut-off frequencies of filters
74 MAP.Left.UpperCutOffFrequencies = MAP.Left.
        El_CF1_CF2_THR_MCL_Gain(:, 3);
75 % Threshold Levels (THR)
76 MAP.Left.THR                    = MAP.Left.
        El_CF1_CF2_THR_MCL_Gain(:, 4);
77 % Maximum Comfort Levels (MCL)
78 MAP.Left.MCL                    = MAP.Left.
        El_CF1_CF2_THR_MCL_Gain(:, 5);
79 % Individual Band Gains (dB)
80 MAP.Left.BandGains              = MAP.Left.
        El_CF1_CF2_THR_MCL_Gain(:, 6);
81 % Optional: comments
82 MAP.Left.Comments               = '';

```

ANEXO A – TERMO DE CONSENTIMENTO PARA NORMOUVINTES



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

**ANEXO B – TERMO DE CONSENTIMENTO PARA USUÁRIOS DE IMPLANTE
COCLEAR**



UNIVERSIDADE FEDERAL DE SANTA CATARINA

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Você está sendo convidado(a) como voluntário(a) a participar da pesquisa intitulada **DESENVOLVIMENTO DE TRANSDUTOR IMPLANTÁVEL E ALGORITMOS AVANÇADOS PARA IMPLANTES COCLEARES COM OBJETIVO DE ADAPTAÇÃO À REALIDADE BRASILEIRA**. Esta pesquisa está elaborada de acordo com a Resolução 466/2012 conforme o CNS (Conselho Nacional da Saúde).

JUSTIFICATIVA E OBJETIVOS

Implantes cocleares vêm tendo sucesso na restauração da audição após perdas parciais ou mesmo profundas. Entretanto, apesar do sucesso, usuários de implante possuem dificuldades comunicativas em ambientes ruidosos. Assim, o motivo que nos leva a estudar algoritmos para implantes cocleares é avaliar como o usuário de implante percebe sons em ambientes ruidosos e determinar se novas técnicas de processamento dos sinais sonoros podem aumentar as capacidades comunicativas em situações barulhentas.

PROCEDIMENTOS

O procedimento de coleta de dados será da seguinte forma: Você utilizará um processador compatível com a configuração do seu Implante Coclear, configurado pelo profissional de fonoaudiologia, que o acompanhará durante os procedimentos de coleta de dados que lhe prestará toda a assistência necessária. O pesquisador testará o sistema previamente ao seu uso para garantir que os níveis de estimulação estão na faixa de segurança pré-estabelecida pelo seu mapeamento, levantado pelo profissional de fonoaudiologia durante seu acompanhamento pós-cirúrgico. Em seguida, o experimentador lhe apresentará alguns sons.

Serão realizadas perguntas a respeito das impressões de qualidade sonora dos sons ouvidos e perguntas sobre a identificação das palavras ou sílabas presentes nesses sons. Caso algum desconforto seja percebido, a avaliação será interrompida. Você responderá às perguntas em um programa em computador ou tablet disponibilizado no início das atividades. Durante os procedimentos de coleta de dados você estará sempre acompanhado por um dos pesquisadores capacitados para te auxiliar, que lhe prestará assistência integral e imediata ou acionará pessoal competente para isso, sem nenhum custo para você. Os pesquisadores estarão disponíveis para oferecer assistência pelo tempo que for necessário em caso de danos decorrentes da pesquisa.

DESCONFORTOS E RISCOS E BENEFÍCIOS

Existe um desconforto e risco mínimo para você se submeter ao procedimento de avaliação auditiva. Você pode sentir desconforto caso o estímulo enviado seja demasiado alto. O avaliador verificará os limites de estimulação levantados pelo profissional de fonoaudiologia durante seu acompanhamento pós-cirúrgico. Antes de se iniciar o experimento, iremos aumentar gradativamente os níveis de estimulação até níveis confortáveis, de modo a evitar qualquer aumento abrupto ou desconforto para você. Da mesma forma, estas avaliações oferecem riscos mínimos como cansaço mental na realização dos testes auditivos e/ou alterações na autoestima provocadas por tarefas que exigem evocação de memórias ou respostas verbais imediatas.

Ao sentir desconforto o indivíduo deverá avisar ao avaliador, pois o atendimento pode ser interrompido. O maior risco ao participante desta pesquisa, é a quebra de privacidade dos(as) mesmos(as), por isso os nomes deles(as) não serão utilizados, a identificação será por meio de números. Esse ensaio não traz benefícios diretos a você. Entretanto, espera-se que sua contribuição possa auxiliar os pesquisadores no

desenvolvimento de melhores estratégias de processamento com potencial de uso em implantes cocleares no futuro. Desta forma você pode ser beneficiado no futuro.

GARANTIA DE ESCLARECIMENTO, LIBERDADE DE RECUSA E GARANTIA DE SIGILO

Você será esclarecido(a) sobre a pesquisa em qualquer aspecto que desejar. Você é livre para recusar-se a participar, retirar seu consentimento ou interromper a participação a qualquer momento, sem ter que apresentar qualquer justificativa. A sua participação é voluntária e a recusa em participar não irá acarretar qualquer penalidade ou perda de benefícios. Os pesquisadores irão tratar a sua identidade com padrões profissionais de sigilo, de modo que os pesquisadores serão os únicos a ter acesso aos dados.

Os resultados da pesquisa poderão ser enviados para você caso haja interesse e permanecerão confidenciais, exceto por potenciais apresentações em encontros ou revistas, de modo que apenas os resultados como um todo, sem as informações pessoais sejam divulgados. Seu nome ou o material que indique a sua participação não será liberado sem a sua permissão. Você não será identificado(a) em nenhuma publicação que possa resultar deste estudo. Duas vias deste documento estão sendo rubricadas e assinadas por você e o pesquisador responsável, uma via deste termo de consentimento informado será arquivada e outra será fornecida a você. Guarde-a cuidadosamente pois este documento traz importantes informações de contato e garante os seus direitos como participante da pesquisa.

CUSTOS DA PARTICIPAÇÃO, RESSARCIMENTO E INDENIZAÇÃO POR EVENTUAIS DANOS

Será garantido o ressarcimento e cobertas as despesas tidas por você e seu acompanhante, para realização da pesquisa, conforme prevê a resolução CNS466/12 Item IV.3 (g). Caso vocês tenham algum custo com transporte ou alimentação o pesquisador responsável assegura cobrir estas despesas. Além de indenizar diante de eventuais danos que ocorram pela participação do mesmo decorrente da pesquisa conforme a resolução CNS466/12 Item IV.3 (h). Caso você se sinta constrangido ou tenha algum dano ao autorizar a sua participação nesta pesquisa, terá direito a indenização por danos eventuais.

OUTRAS INFORMAÇÕES

Em qualquer etapa do estudo, o senhor(a) poderá ter acesso aos profissionais envolvidos na pesquisa. O projeto está sendo coordenado pelo Prof. Julio Cordioli e os ensaios nos quais você participa são coordenados pelo Professor Dr. Eng. Stephan Paul e pela fonoaudióloga Luciana Berwanger Cigana, que podem ser encontrados no endereço: Laboratório de Vibrações e Acústica, Universidade Federal de Santa Catarina, Campus Trindade, telefone: (48) 3721-9278 ramal 207 ou pelos endereços eletrônicos: stephan.paul@ufsc.br e lbcigana@gmail.com.

Se o(a) senhor(a) tiver alguma consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa com Seres Humanos – CEPESH da Universidade Federal de Santa Catarina, que é constituído de pessoas que querem garantir os seus direitos como participante da pesquisa. Eles avaliam a forma como foi planejada a pesquisa, e se ela está acontecendo de forma ética. O contato pode ser feito pelo telefone (48)3721-6094, pelo email cep.propesq@contato.ufsc.br ou pelo endereço Reitoria II, R: Desembargador Vitor Lima, nº 222, 4º andar, sala 401 Trindade, 88040-400 – Florianópolis – SC de segunda à sexta das 10h às 12h e das 15h às 17h. Sua participação nesta pesquisa é de livre e espontânea vontade, sem nenhum custo e seu consentimento, declarado a partir do texto abaixo, poderá ser retirado a qualquer momento.

O pesquisador responsável, que também assina esse documento, compromete-se a conduzir a pesquisa de acordo com o que preconiza a Resolução 466/12 de 12/06/2012, que trata dos preceitos éticos e da

proteção aos participantes da pesquisa.

DECLARAÇÃO DO PARTICIPANTE:

Eu, _____, RG _____, Órgão Expeditor _____, li este documento (ou tive este documento lido por uma pessoa de confiança) e fui informada(o) dos objetivos, os procedimentos a serem realizados, seus desconfortos e riscos e as garantias de confidencialidade da pesquisa acima de maneira clara e detalhada e esclareci minhas dúvidas. Sei que em qualquer momento poderei solicitar novas informações e motivar minha decisão se assim o desejar. Ficou claro também que a minha participação é voluntária e isenta de despesas.

Concordo voluntariamente em autorizar minha participação neste estudo e poderei retirar o meu consentimento a qualquer momento, sem penalidades ou perda de qualquer benefício que eu possa ter adquirido, ou no meu atendimento neste Serviço.

Declaro que obtive de forma apropriada e voluntária o Consentimento Livre e Esclarecido deste paciente ou representante legal para a participação neste estudo. Em caso de dúvidas poderei entrar em contato com pesquisador principal: Prof. Dr. Júlio Apolinário Cordioli no telefone (48) 3721-7215 ou no email julio.cordioli@ufsc.br.

Declaro que concordo em participar desse estudo. Recebi uma via deste termo de consentimento livre e esclarecido e me foi dada a oportunidade de ler e esclarecer as minhas dúvidas.

Nome	Assinatura do Participante	Data
------	----------------------------	------

Nome	Assinatura do Pesquisador	Data
------	---------------------------	------
