



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

Luciana Menezes Xavier de Souza

Redução de Ruído em Implantes Cocleares Utilizando Arranjo de Microfones

Florianópolis
2024

Luciana Menezes Xavier de Souza

Redução de Ruído em Implantes Cocleares Utilizando Arranjo de Microfones

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

Orientador: Prof. Márcio Holsbach Costa, Dr.
Coorientadora: Profa. Renata Coelho Borges, Dra.

Florianópolis

2024

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Menezes Xavier de Souza, Luciana
Redução de Ruído em Implantes Cocleares Utilizando
Arranjo de Microfones / Luciana Menezes Xavier de
Souza; orientador, Márcio Holsbach Costa,
coorientador, Renata Coelho Borges, 2024.
129 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-
Graduação em Engenharia Elétrica, Florianópolis,
2024.

Inclui referências.

1. Engenharia Elétrica. 2. Processamento da fala.
3. Redução de ruído. 4. Implante coclear. 5. Arranjo
de microfones. I. Holsbach Costa, Márcio. II. Coelho
Borges, Renata. III. Universidade Federal de Santa
Catarina. Programa de Pós-Graduação em Engenharia
Elétrica. IV. Título.

Luciana Menezes Xavier de Souza

Redução de Ruído em Implantes Cocleares Utilizando Arranjo de Microfones

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

Prof. Cássio Guimarães Lopes, Dr.
Universidade de São Paulo

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

Prof. Jefferson Luiz Brum Marques, 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 doutora em Engenharia Elétrica.

Prof. Telles Brunelli Lazzarin, Dr.
Coordenador do Programa

Prof. Márcio Holsbach Costa, Dr.
Orientador

Florianópolis, 28 de junho de 2024.

À minha família.

AGRADECIMENTOS

Agradeço primeiramente à Deus pela minha vida e por ter me proporcionado chegar até aqui.

Aproveito a oportunidade para agradecer à minha mãe Elizabethe, que sempre foi a minha fortaleza e meu incentivo, por todo amor e dedicação.

Também quero agradecer àqueles que não estão mais aqui, mas que iniciaram essa trajetória comigo; ao meu pai Márcio (*in memoriam*) por todo seu amor e por me mostrar como a eletrônica era interessante; e ao meu avô Manoel (*in memoriam*) por me mostrar que a vida deve ser vivida com graça e alegria. Agradeço por cada minuto desfrutado ao lado de vocês que sempre estarão em meu coração.

Agradeço também aos meus irmãos Mariana e Tiago vocês sempre me apoiaram e sempre foram meu espelho, amo muito vocês.

Quero agradecer ao professor Márcio que me orientou com muita dedicação, por todo aprendizado e pela convivência, a partir da qual sempre aprendi muito. Também agradeço à professora Renata por nossas conversas e empenho em sua orientação, mesmo distante.

Aos meus amigos e colegas do LPDS, MAGLAB, LABPOST e IEB pelas conversas e por tornarem essa passagem pelo doutorado mais divertida.

Agradeço à banca examinadora pela disponibilidade e competência em avaliar este trabalho.

Às fonoaudiólogas Francine e Nicole agradeço por sempre serem solícitas e me auxiliarem na preparação dos pacientes para os experimentos psicoacústicos.

Também quero agradecer à Universidade Federal de Santa Catarina, ao programa de Pós-graduação em Engenharia Elétrica e ao CNPQ por todo auxílio e infraestrutura.

Agradeço também a todos aqueles que passaram por minha vida e impactaram de alguma forma o modo como percebo e me relaciono com o mundo. Durante todos esses anos de aprendizado só posso dizer que o senso crítico, o pensamento científico e sobretudo a empatia que aprendi com vocês geraram o mais valioso tesouro que tenho, e por isso sou eternamente grata.

“It ain’t about how hard you hit. It’s about how hard you can get hit and keep moving forward. How much you can take and keep moving forward. That’s how winning is done!”

Rocky Balboa

RESUMO

O implante coclear (IC) é um dispositivo que tem como objetivo a restauração parcial da audição de indivíduos com perdas auditivas neurosensoriais de grau severo a profundo. Esse dispositivo utiliza técnicas de processamento de sinais para converter os sons captados pelos microfones em pulsos elétricos, que são aplicados no interior da cóclea por um conjunto de eletrodos. Embora seja capaz de restaurar a capacidade de comunicação em ambientes sem ruído, a existência de interferências sonoras pode prejudicar significativamente a inteligibilidade da fala percebida. De forma a minimizar esse efeito, métodos de redução de ruído são elementos fundamentais na cadeia de processamento do implante coclear. Entre os métodos utilizados, estão aqueles baseados em arranjos de microfones. Uma das formas de implementação é por meio do conformador de feixes de mínima variância com resposta sem distorção. Embora tenha sido utilizada com significativo sucesso em aplicações convencionais de melhoria da fala, como também em aparelhos auditivos, sua forma convencional não é ideal para implantes cocleares, uma vez que a inteligibilidade da fala para usuários desses dispositivos depende unicamente das características da envoltória dos sinais e não de sua estrutura fina. Diante desta constatação, este trabalho propõe uma estratégia de redução de ruído baseada em arranjo de microfones com restrição da envoltória temporal. Conjuntamente, é avaliado o impacto da latência decorrente do processo de estimação da envoltória e seu desempenho. A partir dos resultados obtidos, é derivado um caso particular do método proposto, no qual é possível obter uma solução analítica semi-fechada, resultando em menor latência e complexidade computacional. Simulações computacionais com critérios objetivos e experimentos psicoacústicos com voluntários normouvintes e implantados indicaram aumento significativo de inteligibilidade para o método proposto, quando comparado com técnicas clássicas encontradas na literatura.

Palavras-chave: Arranjo de microfones; implante coclear; redução de ruído; processamento da fala.

ABSTRACT

The cochlear implant (CI) is a device that aims to partially restore the hearing of individuals with severe to profound sensorineural hearing losses. This device uses signal processing techniques to convert sounds captured by microphones into electrical pulses, which are applied inside the cochlea by a set of electrodes. Although it is capable of restoring the ability to communicate in noise-free environments, the existence of sound interference can significantly impair the intelligibility of perceived speech. In order to minimize this effect, noise reduction methods are fundamental elements in the cochlear implant processing chain. Among the methods used are those based on microphone arrays. One of the ways of implementation is through the beamformer minimum variance distortionless response. Although it has been used with significant success in conventional speech improvement applications, as well as in hearing aids, its conventional form is not ideal for cochlear implants, since speech intelligibility for users of these devices depends solely on the characteristics of the signal envelope and not from its fine structure. Given this finding, this work proposes a noise reduction strategy based on a microphone arrangement with temporal envelope restrictions. Together, the impact of latency resulting from the envelope estimation process and its performance is evaluated. From the results obtained, a particular case of the proposed method is derived, in which it is possible to obtain a semi-closed analytical solution, resulting in lower latency and computational complexity. Computational simulations with objective criteria and psychoacoustic experiments with normal-hearing and implanted volunteers indicated a significant increase in intelligibility for the proposed method, when compared to classic techniques found in the literature.

Keywords: Microphone array; cochlear implants; noise reduction; speech processing.

LISTA DE FIGURAS

Figura 1 – Número de publicações sobre implante coclear por ano.....	17
Figura 2 – Quantidade total de publicações sobre implante coclear entre os anos de 1976 até 2023 por país.....	18
Figura 3 – Quantidade de publicações sobre redução de ruído para implantes cocleares por ano.....	18
Figura 4 – Representação da anatomia da orelha humana.	24
Figura 5 – Distribuição de frequências ao longo da membrana basilar da cóclea humana. (a) Cóclea humana sendo desenrolada com aproximação da distribuição de frequências em Hertz; (b) Cóclea desenrolada com apresentação da base coclear e do ápice coclear.....	25
Figura 6 – Representação da captação até a percepção do som: (1º) som entra pelo sistema auditivo; (2º) ossículos se movem e acionam a cóclea; (3º) células ciliadas são acionadas; (4º) as células ciliadas enviam as informações do som ao cérebro.....	26
Figura 7 – Arquitetura do implante coclear.....	29
Figura 8 – Vista frontal do implante coclear implantado em humano.....	30
Figura 9 – Diagrama de blocos funcional de um implante coclear moderno.	32
Figura 10 – Representação da fonação da palavra “Muito”: (a) domínio tempo; (b) domínio tempo-frequência (espectrograma).....	34
Figura 11 – Representação do sinal contaminado sendo enviado ao banco de filtros do tipo gammatone com K sub-bandas de saída. A entrada do banco de filtros é $y_m(n)$ para o m -ésimo microfone de captação.....	36
Figura 12 – Estrutura de funcionamento da plataforma de pesquisa da UTD.....	37
Figura 13 – Classificação dos conformadores de feixes fixos e adaptativos.....	40
Figura 14 – Composição do sinal captado pelo m -ésimo microfone.....	43

Figura 15 – Aquisição do campo acústico com captação simultânea em ambas as orelhas. Neste modelo considera-se a fala desejada $s(t)$, o ruído interferente proveniente de uma fonte pontual $i_p(t)$ e o ruído difuso $r(t)$	44
Figura 16 – Disposição dos microfones de captação assumindo que o implante esteja na orelha direita.....	44
Figura 17 – Processamento do IC para M microfones e K sub-bandas de estimulação.	45
Figura 18 – Cadeia de processamento de sinal para ICs com múltiplos microfones. M microfones e K sub-bandas de estimulação.....	48
Figura 19 – Estrutura interna do arranjo de microfones para cada sub-banda k e atraso ρ	49
Figura 20 – Estimação da envoltória dos sinais de saída do arranjo de microfones. 56	
Figura 21 – Diagrama de blocos do estimador de envoltória utilizando a transformada de Hilbert. Processamento no domínio tempo, amostra a amostra.	57
Figura 22 – (a) Sinal artificial (azul) e a estimativa de sua envoltória (vermelho) para $\Delta=5$; (vermelho); (b) Corte de (a) para as 500 primeiras amostras.	60
Figura 23 – (a) Sinal artificial (azul) e a estimativa de sua envoltória (vermelho) para $\Delta=100$; (vermelho); (b) Corte de (a) para as 500 primeiras amostras.	60
Figura 24 – Esquemático dos softwares desenvolvidos.....	73
Figura 25 – Resposta de magnitude do banco de filtros tipo gammatone de 22 sub-bandas.....	75
Figura 26 – Truncamento da resposta ao impulso para as 22 sub-bandas do filtro gammatone.	76
Figura 27 – Interface gráfica para a realização do experimento psicoacústico com IC.	80
Figura 28 – $SINR_{dB}^{out}$ em função da SIR de entrada, utilizando o vocoder tonal para a síntese do sinal. Fala contaminada (cinza) e processada por: CFV (azul), CFP (verde) e CFE (magenta).....	84
Figura 29 – Inteligibilidade percentual baseada na SRMR-CI. Fala contaminada (cinza) e processada por: CFV (azul), CFP (verde) e CFE (magenta).....	85
Figura 30 – Inteligibilidade STOI. Fala contaminada (cinza) e processada por: CFV (azul), CFP (verde) e CFE (magenta).....	86
Figura 31 – Inteligibilidade NCM. Fala contaminada (cinza) e processada por: CFV (azul), CFP (verde) e CFE (magenta).....	87

Figura 32 – Diagrama de caixas bidimensional para inteligibilidade percentual × STOI. Fala contaminada (cinza) e processada por: CFV (azul), CFP (verde) e CFE (magenta).	88
Figura 33 – Diagrama de caixas bidimensional para inteligibilidade percentual × NCM. Fala contaminada (cinza) e processada por: CFV (azul), CFP (verde) e CFE (magenta).	89
Figura 34 – Resultado do teste psicoacústico com normouvintes. Comparação do CFV (azul), CFP (verde) e CFE (magenta).....	90
Figura 35 – Resultado do teste psicoacústico com um usuário de implante coclear. Comparação entre CFV (azul), CFP (verde) e CFE (magenta).	91

LISTA DE QUADROS

Quadro 1 – Classificação do grau da perda auditiva	28
Quadro 2 – Pseudocódigo para a solução fechada apresentada em (45).	55
Quadro 3 – Pseudocódigo para a solução com restrição da envoltória temporal Equação (73).....	65
Quadro 4 – Pseudocódigo para a solução semi-fechada.....	72
Quadro 5 – Lista geral de softwares utilizados.....	74

LISTA DE TABELAS

Tabela 1 – Predomínio de perda auditiva incapacitante em todo o mundo.....	20
Tabela 2 – Parâmetros para avaliação das metodologias baseadas na restrição de potência.....	81
Tabela 3 – SINR de saída obtida a partir da solução da Equação (79) utilizando método numérico e o procedimento semi-fechado proposto no Quadro 4. SNR=30 dB.....	82
Tabela 4 – Inteligibilidade percentual de saída, via SRMR-CI, obtida a partir da solução da Equação (79) utilizando método numérico e o procedimento semi-fechado proposto no Quadro 4. SNR=30 dB.....	83
Tabela 5 – STOI obtida a partir da solução da Equação (79) utilizando método numérico e o procedimento semi-fechado proposto no Quadro 4. SNR=30 dB.....	83
Tabela 6 – Tempo médio para o cálculo dos filtros ótimos.	92

LISTA DE ABREVIATURAS E SIGLAS

ACE	Codificador de Combinação Avançada (<i>Advanced Combination Encoder</i>)
ANSI	Instituto Nacional Americano de Padrões (<i>American National Standards Institute</i>)
BTE	Atrás da Orelha (<i>Behind the Ear</i>)
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CIS	Amostra Intercalada Contínua (<i>Continuous Interleaved Sample</i>)
CFE	Método de Redução de Ruído com Restrição da Envoltória Temporal
CFP	Método de Redução de Ruído com Restrição da Potência
CFV	Método de Redução de Ruído Baseado no Conformador de Feixes MVDR
dB	Decibéis
DSP	Processamento Digital de Sinais (<i>Digital Signal Processing</i>)
FIR	Resposta Finita ao Impulso (<i>Finite Impulse Response</i>)
FPGA	Arranjo de Porta Programável em Campo (<i>Field Programmable Gate Array</i>)
GSC	Cancelador Generalizado de Lóbulo Lateral (<i>Generalized Sidelobe Canceller</i>)
HRIR	Resposta ao Impulso Relacionada à Cabeça (<i>Head Related Impulse Response</i>)
Hz	Hertz
IC	Implante Coclear
ICRA	Colégio Internacional de Audiologia Reabilitatória (<i>International Collegium for Rehabilitative Audiology</i>)
IIR	Resposta Infinita ao Impulso (<i>Infinite Impulse Response</i>)
LCMV	Mínima Variância com Restrições Lineares (<i>Linearly Constrained Minimum Variance</i>)
MMSE	Mínimo Erro Quadrático Médio (<i>Minimum Mean Square Error</i>)
MSE	Erro Quadrático Médio (<i>Mean Square Error</i>)
MVDR	Mínima Variância com Resposta Sem Distorção (<i>Minimum Variance Distortionless Response</i>)
NCM	Medida de Covariância Normalizada (<i>Normalized Covariance Measure</i>)
OMS	Organização Mundial de Saúde

RF	Radiofrequência
SINR	Razão Sinal-Ruído mais Interferência (<i>Signal to Interference plus Noise Ratio</i>)
SIR	Razão Sinal-Interferência (<i>Signal-to-Interference Ratio</i>)
SPL	Nível de Pressão Sonora (<i>Sound Pressure Level</i>)
SNR	Razão Sinal-Ruído (<i>Signal-to-Noise Ratio</i>)
SRMR-CI	Razão de Energia de Modulação entre Fala e Reverberação para Implantes Cocleares (<i>Speech to Reverberation Modulation Energy Ratio</i>)
STOI	Medida de Inteligibilidade Objetiva de Tempo Curto (<i>Short Term Objective Intelligibility</i>)
TCLE	Termo de Consentimento Livre e Esclarecido
VAD	Detector de Atividade Vocal (<i>Voice Activity Detector</i>)

LISTA DE SÍMBOLOS

Δ	Atraso de processamento
a_m	Coefficientes da resposta ao impulso do sistema acústico entre a fonte de interesse (fala) e o m -ésimo microfone de captação
$b_{p,m}$	Coefficientes das respostas ao impulso do sistema acústico entre a p -ésima fonte interferente e o m -ésimo microfone de captação
C	Número de coeficientes de observação do arranjo de microfones
$E\{\cdot\}$	Valor esperado
\hat{e}_k	Envoltória instantânea do sinal de entrada para a k -ésima sub-banda
$f_{p,m}$	p -ésima fonte interferente no m -ésimo microfone
f_s	Frequência de amostragem
$g_{k,N}$	Coefficiente do banco de filtros para a k -ésima sub-banda e N -ésima amostra
h_i	Transformada de Hilbert
i_p	Sinal da p -ésima fonte interferente no local de geração
k	Índice da sub-banda do banco de filtros
m	Índice do número do microfone
M	Quantidade total de microfones
N	Quantidade de coeficientes de cada sub-banda do banco de filtros
n	Índice de tempo discreto
p	Índice do número de fontes interferentes
REF	Microfone de referência
r_m	Ruído difuso no m -ésimo microfone
s	Sinal da fonte de interesse no local onde o sinal é gerado
t	Índice de tempo contínuo
$u_{m,k}$	Sinal filtrado pelo banco de filtros para o m -ésimo microfone e k -ésima sub-banda
v_m	Sinal de ruído no m -ésimo microfone
$w_{m,k,c}$	Arranjo de microfones de feixes no m -ésimo microfone, k -ésima sub-banda e c -ésima amostra
x_m	Sinal de interesse captado pelo m -ésimo microfone
$x_{\text{REF},k}$	Sinal de interesse captado pelo microfone de referência na k -ésima sub-banda
y_m	Sinal disponível no m -ésimo microfone
$y_{m,k}$	Sinal na saída do k -ésimo banco de filtros do m -ésimo microfone
z_{ak}	Sinal analítico aproximado do sinal de entrada filtrado na k -ésima sub-banda

z_k	Sinal filtrado na k -ésima sub-banda
\tilde{z}_k	Transformada de Hilbert do sinal filtrado na k -ésima sub-banda
τ	Tamanho da aproximação da transformada de Hilbert

Vetores

\mathbf{q}_Δ	Vetor seleção do Δ -ésimo elemento
\mathbf{q}_{REF}	Vetor de seleção do sinal de referência
\mathbf{u}_k	Vetor do sinal de fala filtrada pelo banco de filtros na k -ésima sub-banda
$\mathbf{u}_{m,k}$	Vetor do sinal de fala filtrada pelo banco de filtros no m -ésimo microfone e k -ésima sub-banda
\mathbf{v}	Vetor do sinal de ruído para os M microfones
\mathbf{v}_m	Vetor do sinal do ruído no m -ésimo microfone
\mathbf{w}_k	Vetor do arranjo de microfones na k -ésima sub-banda
$\mathbf{w}_{m,k}$	Vetor do arranjo de microfones no m -ésimo microfone e k -ésima sub-banda
\mathbf{x}	Vetor do sinal de interesse para os M microfones
\mathbf{x}_m	Vetor do sinal do sinal de interesse no m -ésimo microfone
\mathbf{y}	Vetor do sinal disponível para os M microfones
\mathbf{y}_m	Vetor do sinal disponível no m -ésimo microfone
$\tilde{\mathbf{y}}$	Vetor do sinal disponível na k -ésima sub-banda com atraso
\mathbf{z}_k	Vetor do sinal filtrado na k -ésima sub-banda com atraso
$\tilde{\mathbf{z}}_k$	Vetor do sinal filtrado na k -ésima sub-banda com atraso e tamanho da transformada de Hilbert

Matrizes

$\mathbf{0}$	Matriz de zeros
\mathbf{G}	Matriz do banco de filtros para a k -ésima sub-banda
\mathbf{H}	Matriz da transformada de Hilbert
$\mathbf{R}_{\mathbf{v}\mathbf{v}}$	Matriz de autocorrelação do ruído para os M microfones
$\mathbf{R}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}$	Matriz de autocorrelação do ruído com atraso para os M microfones

- \mathbf{R}_{xx}** Matriz de autocorrelação do sinal de interesse para os M microfones
- $\mathbf{R}_{\bar{x}\bar{x}}$** Matriz de autocorrelação do sinal de interesse com atraso para os M microfones
- $\mathbf{\Gamma}_k$** Matriz bloco-diagonal do banco de filtros para a k -ésima sub-banda
- \mathbf{W}_k** Matriz bloco-diagonal dos coeficientes do arranjo de microfones para a k -ésima sub-banda
- \mathbf{D}_k** Matriz de $\mathbf{\Gamma}_k$ para a k -ésima sub-banda

SUMÁRIO

1	INTRODUÇÃO	15
1.1	MOTIVAÇÃO	16
1.2	JUSTIFICATIVA	18
1.3	OBJETIVOS	21
1.3.1	Objetivo geral	21
1.3.2	Objetivos específicos	21
1.8	ORGANIZAÇÃO DO DOCUMENTO	21
2	FUNDAMENTAÇÃO TEÓRICA	23
2.1	ANATOMIA DO SISTEMA AUDITIVO HUMANO	23
2.1.1	Tipos de perdas auditivas	26
2.2	IMPLANTE COCLEAR	29
2.2.1	Arquitetura do implante	31
2.2.2	Estratégias de estimulação	32
2.2.2.1	<i>Continuous interleaved sample</i>	33
2.2.2.2	<i>Advanced combination encoder</i>	33
2.3	CARACTERIZAÇÃO DO SINAL	34
2.3.1	Fala humana	34
2.3.2	Ruído acústico	35
2.4	DECOMPOSIÇÃO EM SUB-BANDAS	35
2.5	INTERFACE DE PESQUISA EM IMPLANTES COCLEARES	36
3	CONFORMAÇÃO DE FEIXES	39
3.1	REDUÇÃO DE RUÍDO BASEADA EM ARRANJO DE MICROFONES PARA APLICAÇÃO EM IC	41

3.2	O MODELO DO SINAL	42
3.3	MODELAGEM DA CADEIA DE PROCESSAMENTO DO IMPLANTE COCLEAR UTILIZANDO ARRANJO DE MICROFONES.....	45
3.4	MODELAGEM DO MÉTODO DE REDUÇÃO DE RUÍDO NO DOMÍNIO DO TEMPO.....	49
3.5	ESTIMAÇÃO DOS COEFICIENTES ÓTIMOS BASEADO NO CONFORMADOR MVDR	52
4	ESTIMAÇÃO DA ENVOLTÓRIA TEMPORAL.....	56
4.1	EXEMPLIFICAÇÃO DA ESTIMAÇÃO DA ENVOLTÓRIA DOS SINAIS... ..	59
5	PROPOSTA.....	62
5.1	FILTRO DE REDUÇÃO DE RUÍDO MULTICANAL BASEADO NA RESTRIÇÃO DO ENVELOPE TEMPORAL NO DOMÍNIO DO TEMPO ..	62
5.2	ARRANJO DE MICROFONES COM RESTRIÇÃO ASSOCIADA À POTÊNCIA DO SINAL DESEJADO	65
5.3	SOLUÇÃO ANALÍTICA SEMI-FECHADA PARA OTIMIZAÇÃO DO ARRANJO DE MICROFONES COM RESTRIÇÃO DE POTÊNCIA.....	67
5.3.1	Problema de autovalores generalizados	69
6	MATERIAIS E MÉTODOS.....	73
6.1	SINAIS E SISTEMAS	74
6.2	CENÁRIO ACÚSTICO.....	76
6.3	ESTIMAÇÃO DAS MATRIZES DE CORRELAÇÃO	77
6.4	MEDIDAS OBJETIVAS DE DESEMPENHO	77
6.5	EXPERIMENTOS PSICOACÚSTICOS	78
6.5.1	Experimento psicoacústico com normouvintes	78
6.5.2	Experimento psicoacústico com implantados cocleares	78
7	RESULTADOS	81
7.1	VERIFICAÇÃO DA SOLUÇÃO ANALÍTICA PARA A PROPOSTA COM RESTRIÇÃO DE POTÊNCIA	81
7.2	COMPARAÇÕES ENTRE MÉTODOS DE REDUÇÃO DE RUÍDO	83

7.3	EXPERIMENTOS PSICOACÚSTICOS	90
7.4	COMPLEXIDADE COMPUTACIONAL.....	91
8	CONCLUSÃO.....	93
8.1	PUBLICAÇÕES	94
8.2	TRABALHOS FUTUROS	95
	APÊNDICE A	105
	APÊNDICE C – MÉTRICAS DE INTELIGIBILIDADE.....	111
	C.1 SRMR-CI.....	111
	C.2 STOI	112
	C.3 NCM.....	113
	ANEXO A – TCLE.....	115

1 INTRODUÇÃO

O sistema auditivo desempenha um papel fundamental na comunicação entre os seres humanos, sendo responsável pela captação e codificação da informação sonora. Essas informações são processadas pelo cérebro, possibilitando a interpretação da fala e auxiliando no processo de orientação espacial do indivíduo (Blauert, 2013).

De acordo com a Organização Mundial de Saúde (OMS), a audição normal é caracterizada pela capacidade de ouvir sons com pelo menos 25 decibéis de nível de pressão sonora (dB SPL) ou até sons mais baixos. Assim, quando uma pessoa ouve menos do que o limiar indicado se considera que ela possui algum tipo de perda auditiva (OMS, 2017). Por outro lado, os padrões estabelecidos pela *American National Standards Institute* (ANSI) definem a audição normal como a habilidade do indivíduo de detectar sons com intensidade de pelo menos 20 dB SPL, desta forma, chama-se de deficiente auditivo aquele que apresenta diferença significativa em relação a esse padrão (ANSI, 1989).

Para a ANSI, qualquer pessoa que tenha perdido aproximadamente 41 dB SPL em relação à audição normal é considerada deficiente auditivo. Levando em conta esses limiares, estima-se que cerca de 4,7% da população mundial viva com algum tipo de perda auditiva debilitante (OMS, 2017). As perdas auditivas são classificadas de leve a profunda e estão relacionadas com a intensidade sonora que um indivíduo pode escutar. As causas mais comuns da perda auditiva neurossensorial estão relacionadas à idade e à exposição ao ruído. Portanto, é de fundamental importância conhecer o sistema auditivo humano, sua anatomia e os possíveis tipos de perdas auditivas, com o objetivo de compreender o funcionamento desse sistema, e então desenvolver técnicas para compensação da deficiência auditiva.

Após estabelecer o contexto das perdas auditivas e sua relevância, este estudo visa contribuir significativamente para o desenvolvimento de técnicas avançadas de compensação da deficiência auditiva. As contribuições incluem o desenvolvimento de uma teoria de filtragem multimicrofones que visa preservar a envoltória temporal da fala, adaptando o arranjo de microfones da técnica clássica de mínima variância com resposta sem distorção para restrições específicas de potência. Além disso, são explorados os impactos da latência decorrente do processo de

estimação da envoltória e a diminuição do tempo de processamento. As simulações numéricas e experimentos psicoacústicos realizados demonstram melhorias significativas em termos de inteligibilidade, comparadas às abordagens tradicionais encontradas na literatura.

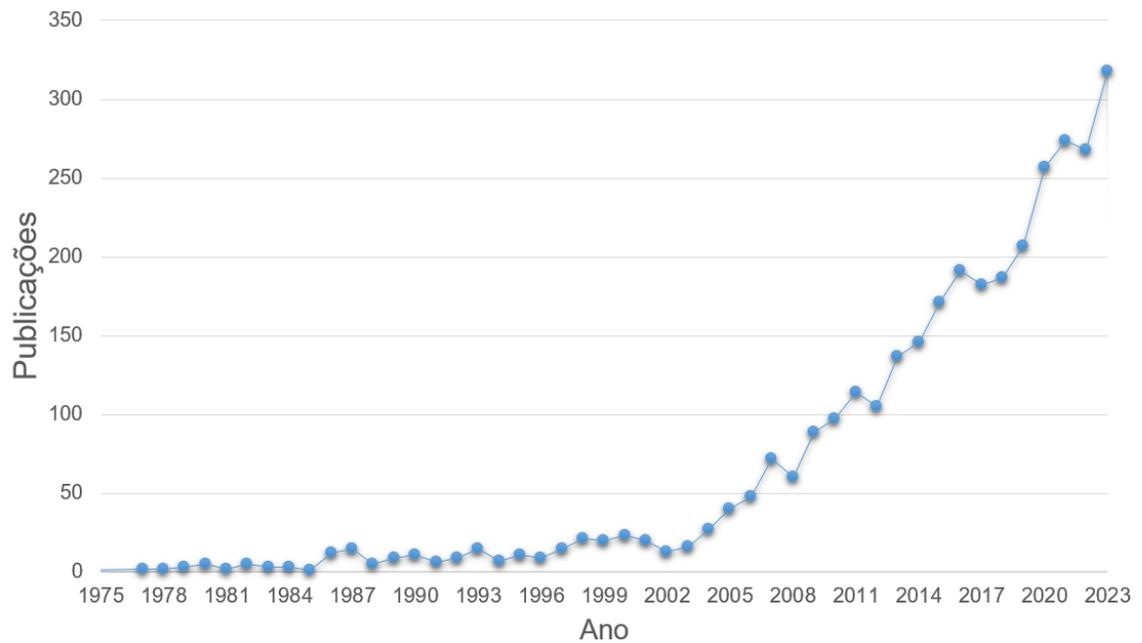
1.1 MOTIVAÇÃO

Segundo Zeng, Popper e Fay (2011), o número de publicações científicas a respeito de implantes cocleares tem apresentado um considerável aumento desde 2001. Para corroborar essa tendência foi realizada uma análise bibliométrica atualizada das publicações relacionadas a esse tema na base de dados *Scopus*. Esta base foi escolhida devido ao seu reconhecimento científico e à acessibilidade aos pesquisadores via portal da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). A análise realizada incluiu trabalhos publicados até dezembro do ano de 2023.

A Figura 1 apresenta o número de publicações a respeito do tema implantes cocleares ao longo dos anos. A busca foi realizada a partir da palavra-chave: implante coclear e limitada à área de engenharia e a artigos científicos ("cochlear implant" AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "ENGI"))). É possível verificar que o número de publicações aumentou significativamente nos últimos 15 anos. No ano de 2023 foram publicados 318 artigos relacionados à área.

As universidades que mais apresentam artigos sobre o tema estão localizadas nos Estados Unidos da América, na China e na Alemanha. A Figura 2 apresenta a quantidade total de publicações realizadas por país acerca do assunto "implante coclear" desde o ano de 1976.

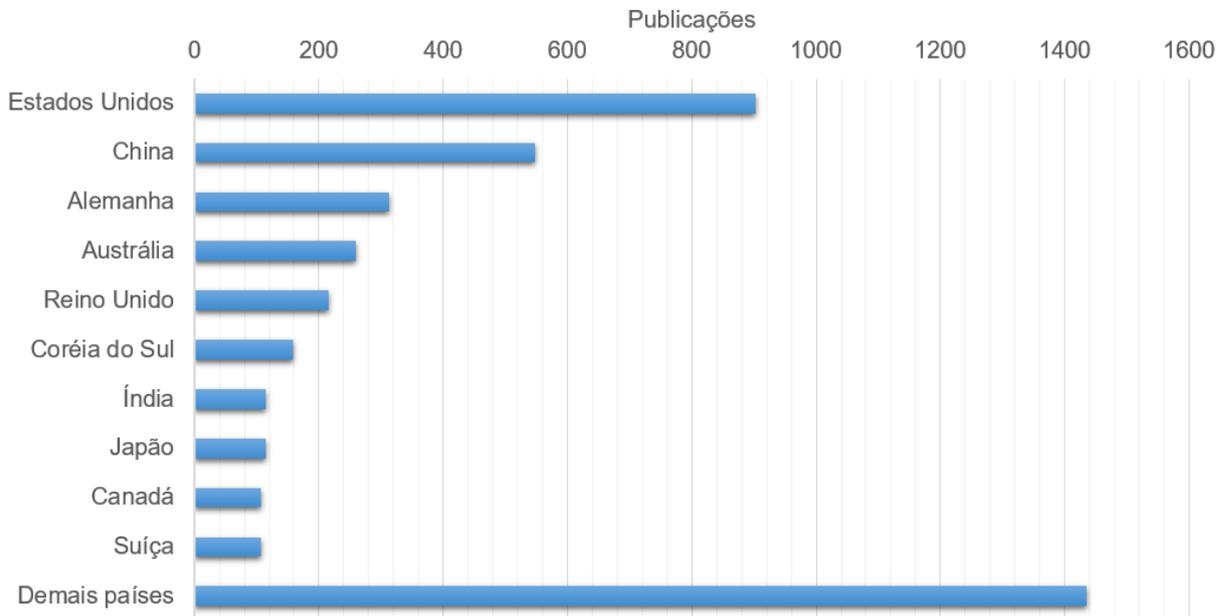
Figura 1 – Número de publicações sobre implante coclear por ano.



Fonte: Autoria própria.

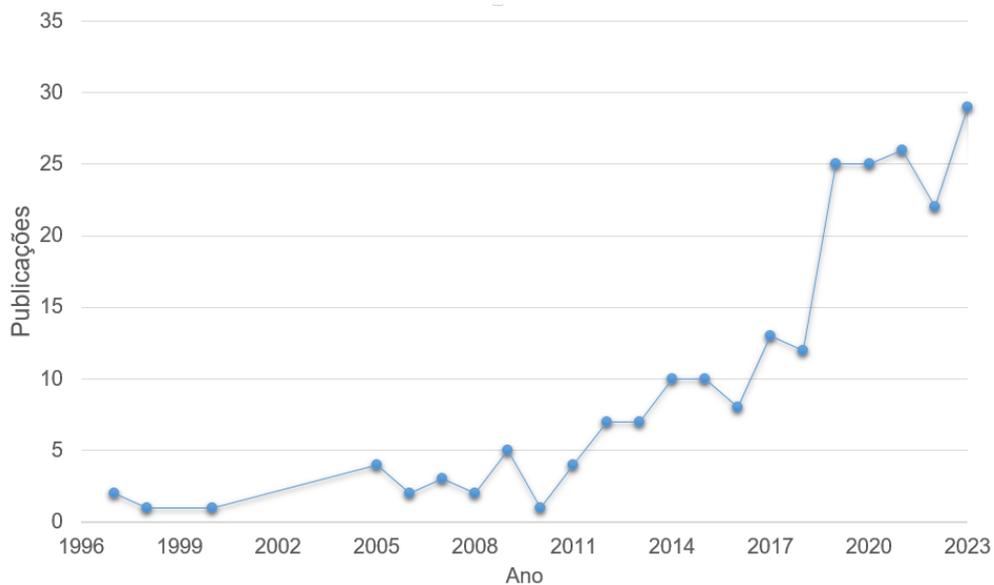
Refinando a busca com as palavras-chaves: implantes cocleares e redução de ruído ("cochlear implant" AND "noise reduction " AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "ENGI"))), obtém-se a Figura 3. Após essa análise mais aprofundada, foi possível identificar que existe uma tendência de crescimento no estudo de métodos de redução de ruído para implantes cocleares, e que desde 2011 a relevância nessa área tem aumentado. O principal grupo de pesquisa atuando nesse assunto está sediado na Universidade do Texas em Dallas nos Estados Unidos, seguido por pesquisas desenvolvidas na Universidade de Ciências e Tecnologia do Sul, na China, e na Universidade Católica de Louvain, situada na Bélgica.

Figura 2 – Quantidade total de publicações sobre implante coclear entre os anos de 1976 até 2023 por país.



Fonte: Autoria própria.

Figura 3 – Quantidade de publicações sobre redução de ruído para implantes cocleares por ano.



Fonte: Autoria própria.

1.2 JUSTIFICATIVA

Segundo dados da OMS, estima-se que cerca de 2.250.000 pessoas tenham algum grau de deficiência auditiva no Brasil, o que correspondia, em 2006, a 1,5% da

população. Tendo em vista essas informações, uma análise do Ministério da Saúde revelou que 15% dessa população possuía grande dificuldade de ouvir e 82% alguma dificuldade permanente (Brasil, 2016).

Em 2005, 278 milhões de pessoas no mundo viviam com deficiências auditivas de grau moderado a profundo, sendo que 80% dessas pessoas viviam em países com economia em desenvolvimento. Em um estudo mais recente da OMS, intitulado de “Estimativas globais sobre prevalência de perda auditiva” (*Global estimates on prevalence of hearing loss*, 2018), aponta-se que:

- 466 milhões de pessoas no mundo vivem com deficiência auditiva incapacitante, ou seja, 6,1% da população mundial tem perda auditiva superior a 40 dB SPL;
- 432 milhões (93% do total) são adultos, em que 242 milhões são homens e 190 milhões são mulheres; e 34 milhões são crianças (7% desse total);
- Um terço das pessoas com mais de 65 anos sofre de perda auditiva incapacitante;
- A previsão estimada é que o número de pessoas com deficiência auditiva incapacitante aumente ao longo dos anos. A estimativa mostra que até 2030, 630 milhões de pessoas terão algum tipo de perda e esse número crescerá para 900 milhões até o ano de 2050 (OMS, 2018).

Em 2019 o Instituto Brasileiro de Geografia e Estatística lançou o documento intitulado como “Pesquisa Nacional de Saúde” que aponta que pessoas na faixa etária de 2 anos ou mais de idade correspondem a 1,1% (ou 2,3 milhões) de pessoas com deficiência auditiva. À medida que a idade avança a quantidade de pessoas com limitações auditivas também avança, desta forma, identificou-se que existe 4,3% (1,5 milhões) de pessoas com deficiência auditiva na faixa etária dos 60 anos ou mais. As demais idades não foram apresentadas nesse relatório (PNS, 2021).

No “Relatório mundial sobre audição” (*World report on hearing*), desenvolvido pela OMS e publicado em 2022, a OMS estima que as previsões são ainda mais devastadoras que as de 2018 (OMS, 2023):

- Estima-se que até 2050 cerca de 2,5 bilhões de pessoas terão algum grau de perda auditiva;
- 700 milhões precisarão de reabilitação auditiva, o que inclui aparelhos e implantes auditivos (OMS, 2023).

A Tabela 1 apresenta o número de pessoas, em milhões, e em termos percentuais, que possuem algum tipo de perda auditiva incapacitante, em todas as idades, de acordo com as regiões que habitam. As populações que mais sofrem são as que estão em desenvolvimento econômico, pois o custo dos aparelhos auditivos e principalmente dos implantes cocleares ainda é significativamente alto. Segundo a OMS, metade de todos esses casos poderia ter sido prevenido se políticas públicas tivessem atingido esses cidadãos. Portanto, investigar e ter ferramentas capazes de gerar desenvolvimento científico e tecnológico nesta área é de interesse mundial.

A busca por sistemas de baixo custo, especialmente para os implantes cocleares, deve levar em conta restrições práticas, como o tamanho do implante, o consumo de energia e a capacidade computacional do sistema. Assim, se faz necessário e existe uma grande demanda por algoritmos eficientes para processamento de sinais. Logo, é imprescindível que novos métodos e técnicas sejam propostos e estudados afim de melhorar a qualidade de vida de toda uma população (OMS, 2018).

Tabela 1 – Predomínio de perda auditiva incapacitante em todo o mundo.

Região	Ambos os sexos	
	Milhões	Prevalência (%)
Países de alta renda	46,02	4,57
Europa Central / Oriental e Ásia Central	34,57	8,36
África Subsaariana	49,66	4,55
Oriente Médio e Norte da África	16,55	3,17
Sul da Ásia	131,67	7,37
Ásia-Pacífico	47,04	6,90
América Latina e Caribe	40,19	6,18
Leste da Ásia	100,76	6,85
Mundo	466,46	6,12

Fonte: Adaptado de OMS (2018).

1.3 OBJETIVOS

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

1.3.1 Objetivo geral

O objetivo geral deste trabalho é estudar, analisar e desenvolver técnicas de redução de ruído multicanal para implantes cocleares utilizando arranjo de microfones. Estas técnicas serão desenvolvidas para proporcionar a melhoria da compreensão da fala pelos implantados cocleares.

1.3.2 Objetivos específicos

Os objetivos específicos deste trabalho são:

- Estudar o problema de redução de ruído em implantes cocleares;
- Desenvolver métodos de redução de ruído para implantes cocleares em cenários acústicos sujeitos a múltiplas fontes sonoras;
- Avaliar os métodos desenvolvidos por meio de simulações numéricas baseadas em critérios objetivos e experimentos psicoacústicos.

1.8 ORGANIZAÇÃO DO DOCUMENTO

Este documento está dividido em capítulos. O primeiro capítulo compreende a introdução, onde são apresentadas a motivação, a justificativa e os objetivos do trabalho.

Na sequência, o Capítulo 2 apresenta a fundamentação teórica, com a descrição da anatomia do sistema auditivo humano e os tipos de perdas auditivas. Nesse capítulo, também é explicado o funcionamento do implante coclear e sua arquitetura.

O Capítulo 3 trata sobre redução de ruído baseada em arranjos de microfones para ICs, em especial sobre conformadores de feixes, bem como a caracterização do

sinal e a modelagem do processamento. No Capítulo 4 são tratados aspectos relativos à estimativa da envoltória temporal.

No Capítulo 5 apresenta-se a proposta principal desta tese e alguns resultados preliminares.

No Capítulo 6 apresentam-se os resultados e discussões. No capítulo 7 é realizada a conclusão e, por fim, as possibilidades de desenvolvimentos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo está organizado da seguinte forma: inicialmente é apresentada uma descrição do sistema auditivo e implantes cocleares, seguida da modelagem dos sinais e sistemas de interesse.

Ao longo deste trabalho é utilizada a seguinte notação: letras minúsculas em itálico representam variáveis escalares; letras minúsculas e maiúsculas em negrito descrevem, respectivamente, vetores e matrizes. Sinais são representados por uma letra minúscula em itálico seguido das variáveis independentes entre parênteses. As dimensões de matrizes e vetores são apresentadas na forma $L \times C$ em que L é o número de linhas e C é o número de colunas.

2.1 ANATOMIA DO SISTEMA AUDITIVO HUMANO

O som é uma onda de pressão que se propaga no meio, podendo vibrar de forma lenta ou rápida. Os sons graves estão associados a baixas frequências (abaixo de 300 Hz), enquanto os sons agudos abrangem frequências mais elevadas (acima de 5 kHz) (Gelfand, 2017). A orelha humana, representada na Figura 4, é responsável por receber o som, condicioná-lo e transformar as vibrações acústicas em sinais elétricos. A transformação do sinal mecânico para sinal elétrico é feita de forma que impulsos elétricos associados possam ser interpretados pelo cérebro (Gelfand, 2017).

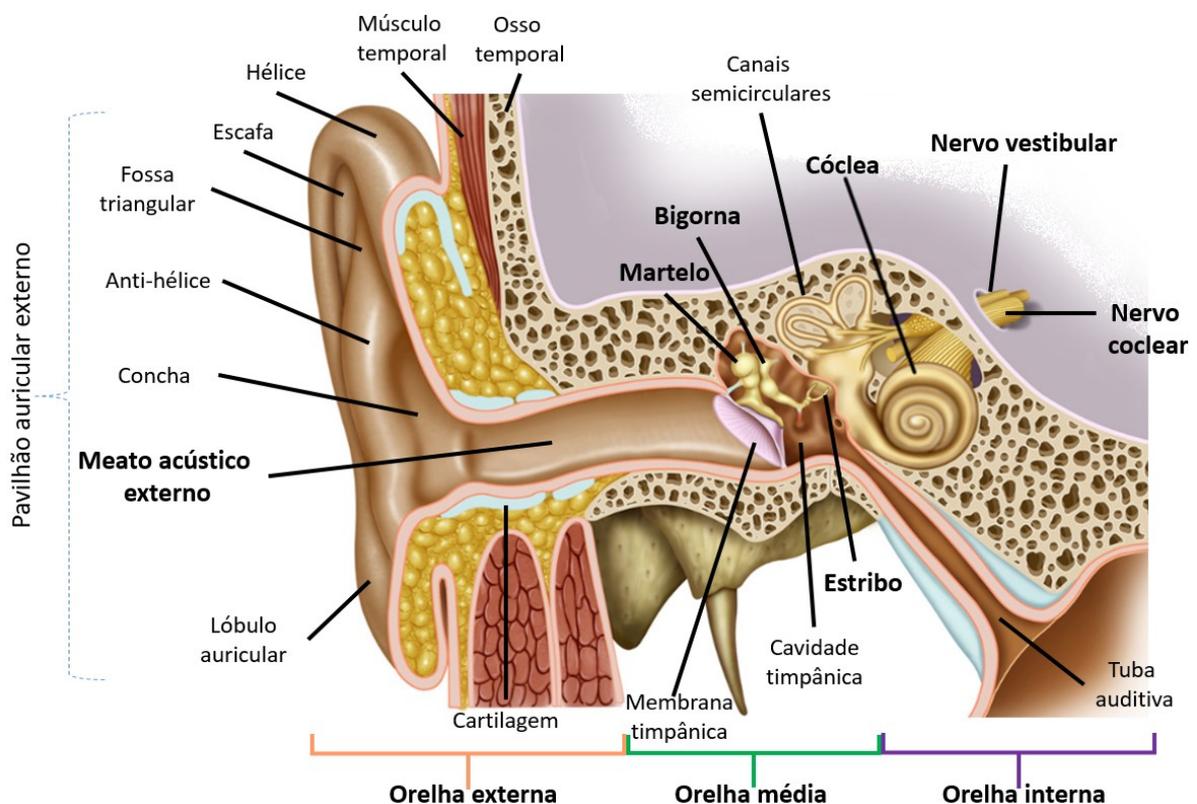
A orelha possui três partes principais que estão divididas em: orelha externa, média e interna. A orelha externa, por onde é feita a captação do som e o seu encaminhamento até a membrana timpânica, é formada por toda parte visível da orelha, que consiste no pavilhão auricular externo até o canal auditivo externo (meato acústico externo), apresentando um comprimento de até 2 cm (Yost, 2007).

A orelha média é formada pela membrana timpânica e por três ossículos, o martelo, o estribo e a bigorna. Essa região da orelha é uma pequena cavidade que tem aproximadamente 2 cm^3 de volume. Sua principal função é propiciar o casamento de impedâncias que é o condicionamento eficiente do som para a orelha interna. Assim, quando um estímulo acústico ocorre o som é recebido e enviado até a membrana timpânica, por onde é feita a transmissão do sinal sonoro até a orelha interna. Essa transmissão pode ocorrer de três maneiras: pela condução óssea; por

mudanças na pressão do ar na cavidade da orelha média; e/ou através da cadeia de ossículos da cavidade da orelha média, sendo a última maneira a mais importante (Yost, 2007).

Na orelha interna encontram-se o sistema vestibular e a cóclea. O sistema vestibular é caracterizado pelos canais semicirculares, sendo responsável pelo equilíbrio humano. É na cóclea que a energia mecânica é transformada em sinais elétricos para serem então enviados ao cérebro.

Figura 4 – Representação da anatomia da orelha humana.



Fonte: Adaptado de Texas Ear Center (2019).

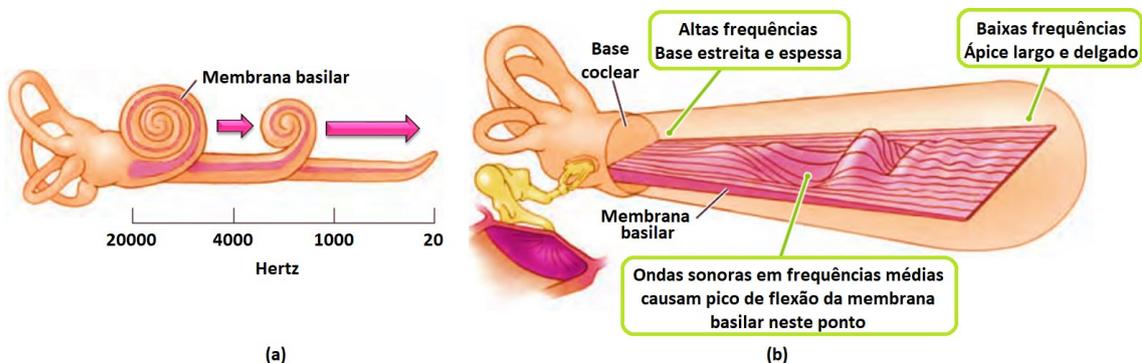
A cóclea é estruturada de modo a responder a diferentes frequências sonoras de acordo com a localização ao longo de sua extensão. Desta maneira, os sons agudos, que possuem as frequências mais elevadas, atuam sobre a base da membrana basilar, enquanto as frequências mais baixas agem sobre a região apical. Esse arranjo de frequências é denominado distribuição espacial tonotópica¹ da cóclea.

¹ Pertencente ao arranjo espacial de onde o som é percebido, transmitido ou recebido.

Na Figura 5 (a) é apresentada a distribuição de frequências ao longo da membrana basilar da cóclea humana desenrolada, onde se nota aumento progressivo na largura da membrana basilar desde a base (aproximadamente 20 kHz) até o ápice (20 Hz) (Dallos, Popper e Fay, 1996). Sabe-se ainda que a base da membrana basilar é espessa e o ápice apresenta característica oposta sendo delgado.

As vibrações dos ossículos mostrados ao lado da membrana desenrolada (Figura 5 (b)), geram ondas que viajam através do fluido coclear. Assim, as frequências de onda mais rápidas causam picos máximos de deslocamento perto da base da membrana basilar, enquanto as frequências de onda mais lentas resultam em picos de deslocamento máximo perto do ápice da membrana (Kolb, Whishaw e Teskey, 2001).

Figura 5 – Distribuição de frequências ao longo da membrana basilar da cóclea humana. (a) Cóclea humana sendo desenrolada com aproximação da distribuição de frequências em Hertz; (b) Cóclea desenrolada com apresentação da base coclear e do ápice coclear.



Fonte: Adaptado de Kolb, Whishaw, Teskey (2001).

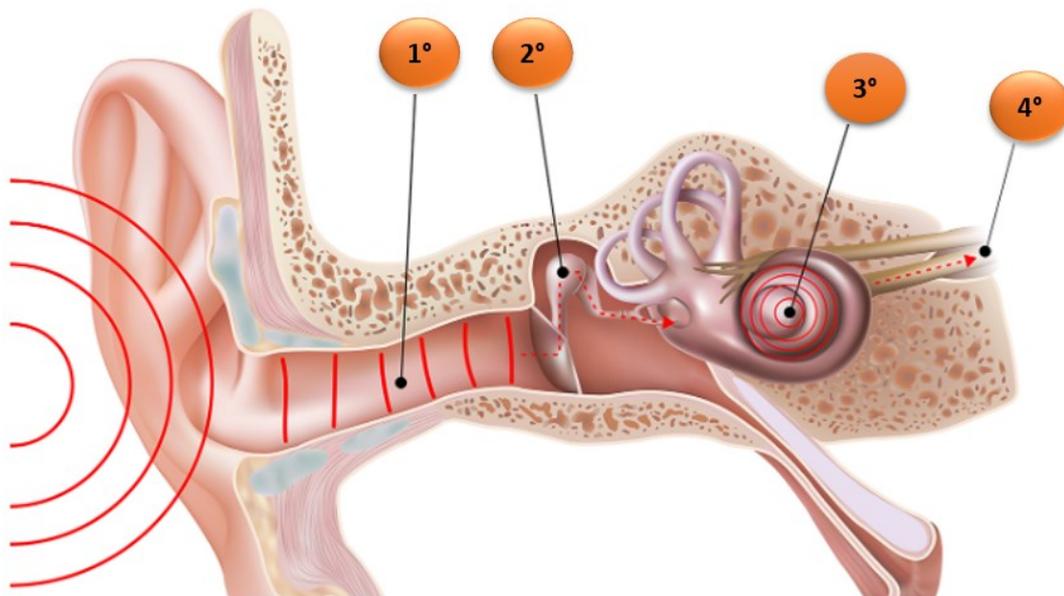
Na Figura 6 é ilustrado o processo de audição desde a captação até as ligações que enviam os sinais ao cérebro. Esse processo envolve quatro etapas:

- Etapa 1: O som é captado pelo canal auditivo e chega ao tímpano, fazendo-o vibrar;
- Etapa 2: O tímpano aciona os ossículos que se movem e estimulam a cóclea;
- Etapa 3: Com a força da pressão, a membrana basilar movimenta-se e o líquido dentro dela (endolinfa) também. Essa movimentação ativa

pequenas células sensoriais, chamadas de células ciliadas, que estão presentes em seu interior. Cada região da cóclea possui sensibilidade associada a uma determinada faixa de frequências. É dessa forma que se consegue diferenciar sons graves de agudos (Yost, 2007);

- Etapa 4: Ao receber esse estímulo, as células ciliadas enviam a informação do som para o cérebro. Todo o processo dura em torno de milésimos de segundos.

Figura 6 – Representação da captação até a percepção do som: (1º) som entra pelo sistema auditivo; (2º) ossículos se movem e acionam a cóclea; (3º) células ciliadas são acionadas; (4º) as células ciliadas enviam as informações do som ao cérebro.



Fonte: Adaptado de Texas Ear Center (2019).

2.1.1 Tipos de perdas auditivas

A perda auditiva é o resultado de danos a uma ou mais partes do ouvido externo, médio ou interno (ASHA, 2019). A gravidade e o tipo da perda auditiva dependem do problema que a causou. As três categorias de perdas existentes estão relacionadas com a região e a parte lesionada do sistema auditivo, são elas:

- Perda condutiva: ocorre na orelha externa ou média por obstrução do canal. As causas mais comuns são o acúmulo de cerume,

deformidades no canal auricular, infecções no ouvido médio, ruptura da membrana timpânica e alterações nos ossículos da orelha média (Nieman e Oh, 2020). Esse tipo de perda pode ser tratada com medicamentos e intervenções cirúrgicas (Fernandez, 2007);

- Perda neurossensorial: ocorre na orelha interna e advém de danos nas células ciliares da cóclea. Esse tipo de perda é considerado como moderado a grave. Ocorre com maior frequência em adultos, sendo, normalmente, irreversível (Nieman e Oh, 2020). Pode ter como causas a idade, infecções virais, tumores ou ainda sons de elevada magnitude;
- Perda central: há a alteração do processamento da informação sonora no cérebro, nesse caso o sistema nervoso central apresenta alguma falha relacionada com o sistema auditivo (Fernandez, 2007).

Os testes para diagnosticar a perda auditiva podem incluir exame físico para identificar problemas estruturais e possíveis inflamações no ouvido. Além disso, o médico pode realizar o teste do sussurro, no qual o paciente cobre um dos ouvidos para que seja observado o quão bem ele ouve as palavras faladas em vários volumes. Outra possibilidade é o teste de diapasão, em que se utiliza um instrumento metálico de duas pontas que produz som ao ser tocado, desta maneira, pode ajudar a detectar a perda auditiva, além de revelar em que parte do ouvido o dano ocorreu. Por fim, podem-se utilizar testes audiométricos (Bogle, 2022).

A avaliação do grau da perda auditiva é realizada pelo exame conhecido como audiometria, que deve ser conduzido por médicos ou fonoaudiólogos. Esse exame é realizado de maneira simples e não invasiva, necessitando apenas da participação do paciente (Lieberman *et al.*, 2016). Durante o procedimento, o paciente utiliza fones de ouvido que emitem sons e palavras direcionadas para cada um dos ouvidos. Cada tom é repetido em vários níveis até que se identifique o limiar auditivo (Bogle, 2022).

As perdas auditivas são classificadas pela intensidade sonora mínima que um indivíduo é capaz de escutar. Pessoas com audição saudável têm capacidade de ouvir sons de 0 a pelo menos 25 dB SPL. Os graus de perda auditiva são classificados da seguinte forma:

- Leve (26 a 40 dB SPL): A pessoa não percebe que não está ouvindo direito, assim aumenta o tom de voz para uma simples conversa. Com esse grau, conversas em ambientes com muito ruído se tornam difíceis;
- Moderada (41 a 60 dB SPL): o som da fala quase não será percebido. Nesse grau, a pessoa não escuta tons naturais de voz e busca ajuda médica para resolver o problema;
- Severa (61 a 80 dB SPL): nenhum som de fala é audível em nível normal de intensidade de voz;
- Profunda (acima de 81 dB SPL): impede que a pessoa escute a maioria dos sons, assim, apenas alguns sons graves de alta magnitude podem ser audíveis (Liberman *et al.*, 2016).

Na literatura são encontradas diferentes quantificações para o grau de perda auditiva. Segundo o Conselho de Fonoaudiologia (2017), a classificação mais utilizada é a dada pela OMS. Nessa escala, a perda auditiva é classificada a partir do nível médio de audição, que é obtido pela média aritmética do nível de audibilidade nas frequências de 500 Hz, 1 kHz, 2 kHz e 4 kHz. No Quadro 1 é apresentada a classificação dos graus da perda auditiva em dB SPL para adultos e crianças.

Uma parte significativa dos danos que o sistema auditivo sofre pode ser tratada com medicamentos, cirurgias, aparelhos auditivos, próteses de condução óssea e implantes cocleares. Essas são algumas das possíveis formas de auxiliar e melhorar a qualidade de vida dos deficientes auditivos. Em casos em que a perda auditiva é severa ou profunda apenas o implante coclear poderá auxiliar os indivíduos. Assim, na próxima seção discute-se sobre esse dispositivo, bem como o seu funcionamento e as partes que o compõe.

Quadro 1 – Classificação do grau da perda auditiva

Grau de perda auditiva	Média aritmética da perda nas frequências de 500 Hz, 1 kHz, 2 kHz e 4 kHz.	
	Crianças	Adultos
Audição normal	0 - 15 dB SPL	0 - 25 dB SPL
Leve	16 - 30 dB SPL	26 - 40 dB SPL
Moderada	31 - 60 dB SPL	41 - 60 dB SPL
Severa	61 - 80 dB SPL	61 - 80 dB SPL
Profunda	> 81 dB SPL	> 81 dB SPL

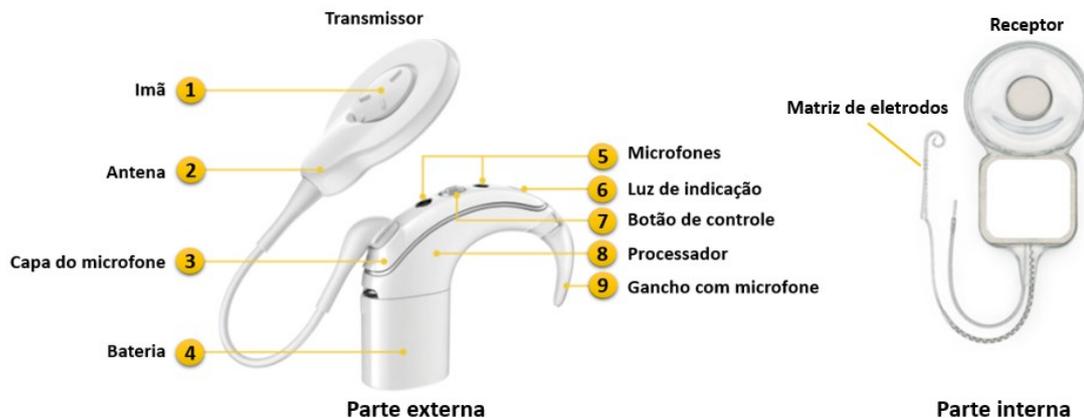
Fonte: Adaptado conselho de fonoaudiologia (2017).

2.2 IMPLANTE COCLEAR

O implante coclear (IC) é uma prótese auditiva que restaura de maneira parcial a capacidade de percepção sonora em pessoas com perda auditiva neurossensorial severa ou profunda (Shapiro; Bradham, 2012). O implante só tem efeito se o nervo auditivo de pelo menos uma das orelhas do indivíduo estiver preservado e responder a estímulos elétricos.

O IC é composto por duas partes, a externa e a interna, como apresentado na Figura 7. Existem várias marcas e modelos de ICs, sendo a arquitetura em questão baseada no implante *Nucleus 7* da empresa Cochlear® Limited. A parte externa, possui um ou mais microfones, um processador de sinais e uma antena transmissora por radiofrequência (RF). A unidade interna, que é cirurgicamente implantada, engloba um receptor, uma unidade de estimulação e um conjunto de eletrodos que são posicionados no interior da cóclea (Brito Neto, 2001).

Figura 7 – Arquitetura do implante coclear.



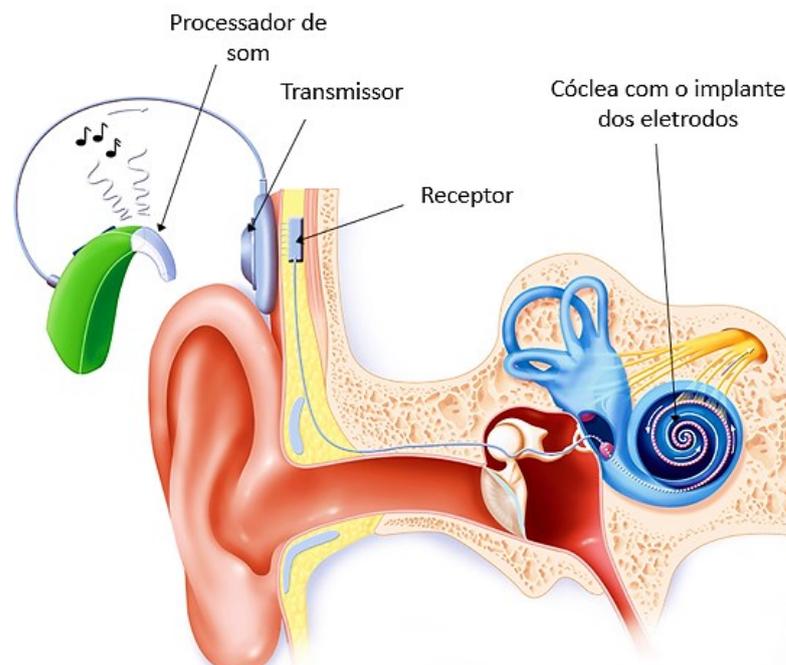
Fonte: Adaptado de Cochlear Limited (2020).

A Figura 8 ilustra o diagrama da vista frontal do IC implantado e algumas partes que o compõe. Existem etapas para a captação, processamento e estimulação da cóclea, que são descritas na sequência:

- Etapa 1: o processador externo, posicionado atrás da orelha (*Behind the Ear* – BTE) ou junto ao corpo, captura o som e o transforma para a forma digital. A alimentação é realizada por uma bateria;
- Etapa 2: o processador externo conforma os sinais captados de acordo com um mapa de compensação definido pelo fonoaudiólogo

- (Greenberg *et al.*, 2008); na sequência, codifica e transmite o sinal resultante por meio de uma antena de transmissão via radiofrequência;
- Etapa 3: o receptor, que fica implantado de forma subcutânea e próximo a orelha, possui outra antena de RF que recebe o sinal. O processador interno decodifica esse sinal e o transforma em pulsos de corrente elétrica, que são aplicados através dos eletrodos posicionados na cóclea. A matriz de eletrodos possui, geralmente, entre 16 e 24 eletrodos, dependendo do fabricante;
 - Etapa 4: os eletrodos do implante estimulam o nervo auditivo da cóclea, que, por sua vez, envia os impulsos elétricos para o cérebro, onde são interpretados (Arndt, 1999).

Figura 8 – Vista frontal do implante coclear implantado em humano.



Fonte: Adaptado de NIDCD (2018).

Os implantes cocleares podem ser divididos em dois tipos: os monocanais e os multicanais. O IC monocanal é um tipo de implante que estimula o nervo auditivo por meio do campo eletromagnético criado por dois eletrodos: um implantado no interior da cóclea e o outro fica localizado fora dela (Arndt, 1999; Zirn *et al.*, 2019). Essa forma é conhecida como estimulação monopolar e permite que a corrente elétrica se espalhe por uma grande área, ocorrendo assim o estímulo de uma grande

região neural com a utilização de uma corrente elétrica de baixa magnitude. No entanto, devido à limitação na precisão do estímulo e na distribuição da corrente, o que acaba diminuindo o desempenho do método, o seu uso tem diminuído drasticamente (Greenberg *et al.*, 2008). Os ICs multicanais foram desenvolvidos com o intuito de utilizar uma forma de estimulação que proporcionasse um campo eletromagnético mais localizado. Para isso, leva-se em consideração as propriedades tonotópicas da cóclea, na qual são inseridos eletrodos em diferentes posições (Brito Neto, 2001).

Compreendendo as características temporais e espectrais típicas da fala e do ruído, pode-se projetar um sistema para a compensação da audição, como também para redução de ruído. Em ICs que utilizam sistemas com dois ou mais microfones é possível utilizar métodos de filtragem espacial (Luo; Pavlovic; Nehorai, 2002).

É importante destacar que existem implantes bilaterais, que são realizados em ambas as orelhas e que atuam de forma independente. Além desses, há os ICs biauriculares que são implantados em ambas as orelhas e se comunicam entre si, realizando o compartilhamento de informações, o que pode melhorar o processamento e a percepção da espacialidade das fontes sonoras. Embora o número de pesquisas que tratam a respeito dos ICs biauriculares esteja crescendo (Dietz; Mcalpine, 2015; Kokkinakis; Loizou, 2010; Lopez *et al.*, 2016), essa ainda é uma tecnologia pouco acessível em relação ao custo financeiro.

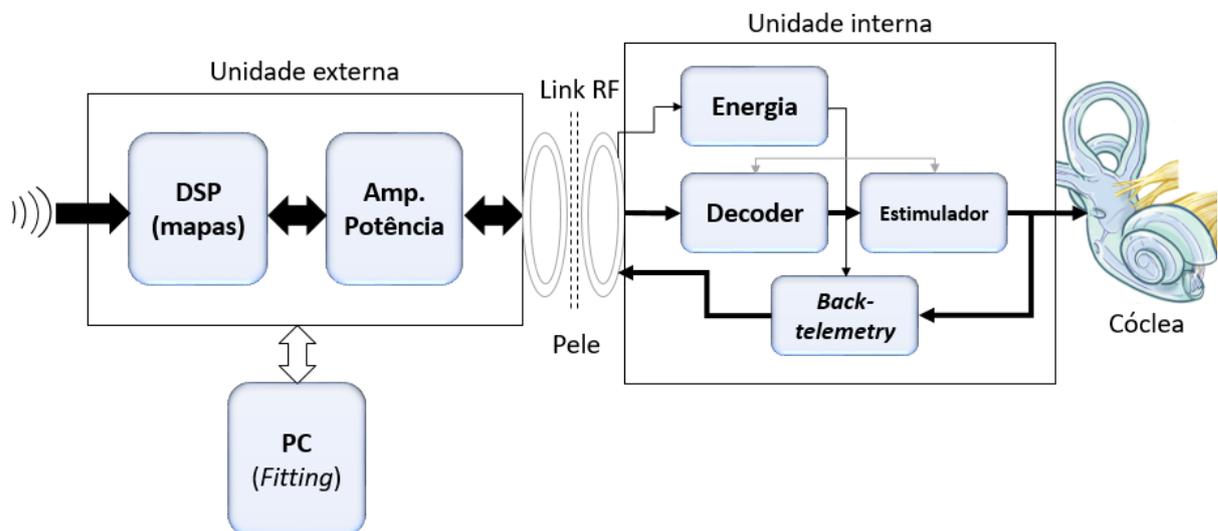
2.2.1 Arquitetura do implante

Os implantes cocleares mais simples apresentam a arquitetura ilustrada na Figura 9, e compreendem uma unidade externa e uma interna. A unidade externa, também conhecida como processador de voz, consiste em uma unidade de processamento digital de sinais (*Digital Signal Processing* - DSP), um amplificador de potência e um transmissor RF. O DSP é o cérebro do implante coclear, e tem como funções a conversão analógica-digital; a amplificação, a filtragem e conformação do sinal de interesse; e a codificação necessária para sua transmissão à unidade interna. Adicionalmente, o DSP possui em sua memória o mapa de configuração com as informações específicas do implantado. Os mapas e demais parâmetros de

processamento são definidos e modificados por um programa de ajuste (*fitting*) utilizado pelo fonoaudiólogo (Zeng *et al.*, 2008).

A unidade interna compreende um receptor RF e um estimulador, ambos hermeticamente selados. Como a unidade interna não tem bateria, o estimulador deve primeiro extrair energia do sinal de RF. Após a sua energização, o estimulador realiza a decodificação do sinal captado e a aplicação das correntes elétricas nos respectivos eletrodos. Vale lembrar que os sistemas modernos contêm um laço de realimentação que ajuda a monitorar a atividade elétrica e neural (Clark *et al.*, 1987).

Figura 9 – Diagrama de blocos funcional de um implante coclear moderno.



Fonte: Adaptado de Zeng *et al.* (2008).

2.2.2 Estratégias de estimulação

As estratégias de estimulação são empregadas para codificar o sinal de interesse visando otimizar a percepção sonora pelo usuário do IC. Nos primeiros implantes monocanais não havia um aproveitamento da distribuição espacial tonotópica da cóclea, o que resultava em um baixo desempenho em termos de inteligibilidade (Zeng *et al.*, 2008).

Os implantes multicanais trouxeram avanços significativos na questão de inteligibilidade (Taal *et al.*, 2011; Einhorn, 2017; Cosentino *et al.*, 2014), com parte desse desempenho atribuído às estratégias de estimulação. Entre as estratégias mais conhecidas estão a *Continuous Interleaved Sample* (CIS) (Wilson *et al.*, 1993) e a *Advanced Combination Encoder* (ACE) (Vandali *et al.* 2000).

2.2.2.1 *Continuous interleaved sample*

A estratégia de codificação por amostra intercalada contínua (*Continuous Interleaved Sample* - CIS) é uma das principais técnicas de codificação, responsável por um grande avanço na área de ICs, pois possibilitou um aumento considerável na capacidade de percepção sonora dos usuários (Zeng, 2004). A estratégia CIS atenua o problema de interferência entre as sub-bandas próximas por meio de estimulação assíncrona e intercalada. Assim, a interferência causada pelo espalhamento do campo elétrico, devido à proximidade dos eletrodos de estimulação e das características condutivas do meio, é quase totalmente eliminada (Wilson *et al.*, 1993).

Essa estratégia se baseia na aplicação de trens de pulsos elétricos com espaçamento temporal nos eletrodos. Assim, somente um pulso elétrico é entregue em cada instante de tempo, eliminando a sobreposição dos estímulos que ocorrem nas diferentes sub-bandas de estimulação (Wouters; Mcdermott; Francart, 2015).

2.2.2.2 *Advanced combination encoder*

Na década de 1990, surgiram as estratégias do tipo “*N-of-M*”, que selecionam um número reduzido de sub-bandas, geralmente entre 8 e 12 em cada ciclo de estimulação, de um total de $M > N$ sub-bandas de processamento. Nesse contexto, somente N sub-bandas são utilizadas em cada ciclo de estimulação. Esse tipo de estratégia visa reduzir a interação entre as sub-bandas de estimulação e também diminuir o consumo de energia da bateria do IC (Vandali *et al.* 2000).

As etapas de processamento de sinais na estratégia ACE são semelhantes às realizadas pela estratégia CIS, porém a diferença se encontra na inclusão da etapa de seleção das sub-bandas. Nesta etapa, onde são selecionadas as N sub-bandas com maior magnitude para o processamento (Nogueira *et al.*, 2005). É importante destacar que esse tipo de estratégia é a mais utilizada por usuários de IC da marca Cochlear® (Wouters; Mcdermott; Francart, 2015).

2.3 CARACTERIZAÇÃO DO SINAL

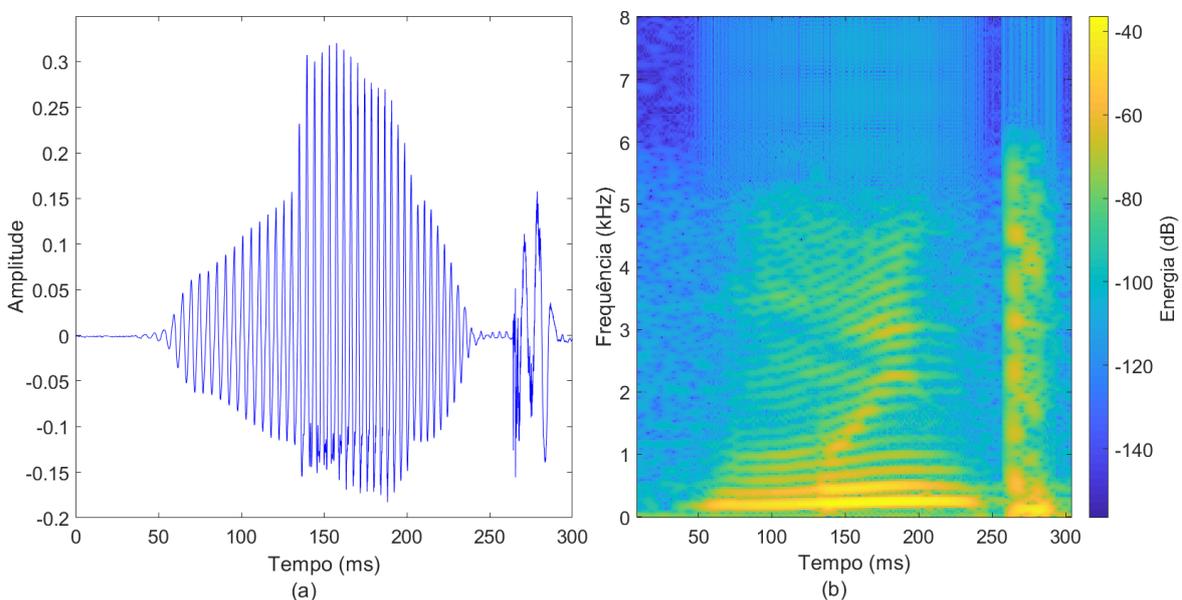
Nessa seção, são apresentadas a composição e caracterização espectral do sinal de fala e ruído de contaminação.

2.3.1 Fala humana

A fala possui informações cruciais para a inteligibilidade em uma ampla faixa de frequências, variando entre 50 Hz e 8 kHz, sendo que a maior parte da densidade espectral de potência está concentrada abaixo de 4 kHz (Marquardt *et al.*, 2014). Embora sejam consideradas não estacionárias, do ponto de vista estocástico, suas estatísticas de segunda ordem podem ser tratadas como estacionárias por períodos curtos de tempo, em geral, de 10 a 40 ms (Loizou, 2013).

Na Figura 10 é apresentada a representação da palavra “Muito” no domínio tempo (Figura 10(a)) e tempo-frequência (Figura 10(b)), através de seu espectrograma. Na Figura 10(a), as regiões de fala e silêncio são facilmente identificáveis, com magnitudes elevadas separadas por pausas. Por outro lado, observa-se claramente na Figura 10(b) uma baixa concentração de energia acima de 6 kHz.

Figura 10 – Representação da fonação da palavra “Muito”: (a) domínio tempo; (b) domínio tempo-frequência (espectrograma).



Fonte: Autoria própria.

2.3.2 Ruído acústico

O ruído sonoro está presente em quase todos os momentos da vida humana, abrangendo qualquer informação sonora que não seja de interesse primário em um dado instante (Hadad *et al.*, 2015). Em termos de comunicação, o ruído é geralmente caracterizado por uma densidade espectral de potência que muda lentamente ao longo do tempo, ao contrário dos sinais de fala, que são mais dinâmicos (Marquardt, 2015). Essa distinção foi muito estudada e explorada em pesquisas sobre redução de ruído, como em Marquardt *et al.* (2014), Hadad *et al.* (2015) e Hadad, Doclo e Gannot (2016).

Para reduzir o ruído é fundamental entender suas características temporais e espectrais, principalmente por que se busca distinguir a fala do ruído (Loizou, 2013). Assim, é necessário saber quais tipos de ruído estão no ambiente acústico de interesse. Em geral, o ruído aditivo captado em cada um dos m microfones, $v_m(n)$, é representado pela soma de sinais com diferentes origens, excluindo a fonte de interesse. Neste trabalho considera-se que o ruído aditivo é composto por duas partes: o ruído interferente, é representado por $f_{p,m}(n)$, decorrente do sinal proveniente da p -ésima fonte interferente captado pelo m -ésimo microfone, e o ruído de fundo, $r_m(n)$, captado pelo m -ésimo microfone, que pode ser de origem elétrica ou do ambiente (difuso).

A combinação ou soma dos sinais de fala e ruído aditivo resulta em um terceiro sinal, que recebe o nome de sinal contaminado $y_m(n)$, que será captado pelos M microfones.

2.4 DECOMPOSIÇÃO EM SUB-BANDAS

Existem diversas formas para realizar a decomposição do sinal em bandas. Duas possibilidades são o uso da transformada de Fourier de curto termo (*Short-time Fourier Transform* - STFT) e o uso de um conjunto de filtros passa-banda. No segundo caso, a forma mais utilizada em implantes cocleares é o filtro gammatone (Tabibi *et al.*, 2017).

O filtro gammatone é um filtro linear descrito por uma resposta ao impulso definida pelo produto de uma distribuição gama e um tom senoidal

(Johannesma,1972). Ele é comumente utilizado para simular o movimento da membrana basilar dentro da cóclea em função do tempo, onde a saída de cada filtro modela a resposta em frequência da membrana basilar. Desenvolvido de forma a reproduzir fisiologicamente a estrutura do estágio de processamento auditivo periférico, esse filtro foi popularizado por Johannesma em 1972 ao modelar o movimento da membrana basilar em gatos (Johannesma, 1972).

O uso do filtro gammatone resulta em um caso particular de banco de filtros em que as frequências centrais são distribuídas em proporção à largura de banda (Glasberg; Moore, 1990).

Neste trabalho o sinal contaminado, $y_m(n)$, adquirido por cada um dos M microfones, é processado por um banco de filtros, sendo decomposto em diferentes bandas de frequência. Os sinais de saída desse banco de filtros são denominados de $u_{m,k}(n)$ em que k corresponde ao índice da sub-banda e m ao microfone, conforme ilustrado na Figura 11.

Figura 11 – Representação do sinal contaminado sendo enviado ao banco de filtros do tipo gammatone com K sub-bandas de saída. A entrada do banco de filtros é $y_m(n)$ para o m -ésimo microfone de captação.



Fonte: Autoria própria.

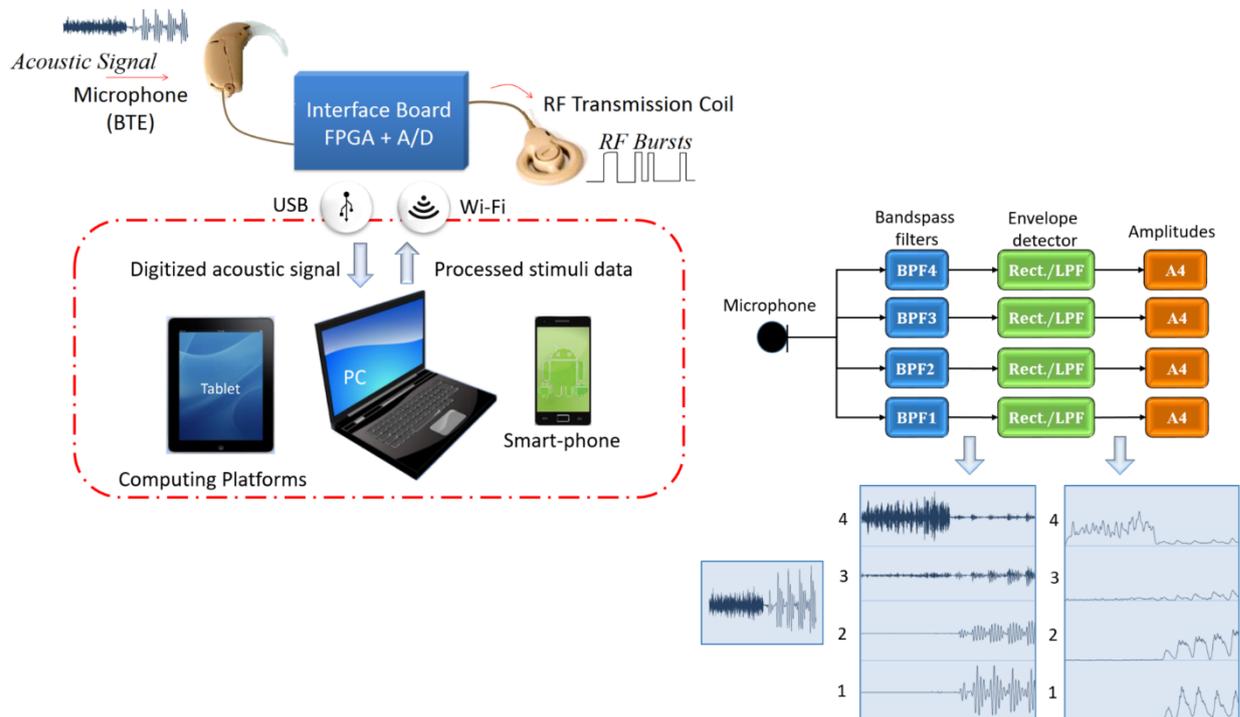
2.5 INTERFACE DE PESQUISA EM IMPLANTES COCLEARES

Experimentos com voluntários usuários de ICs podem ser realizados através de um sistema chamado de interface de pesquisa. Essa interface possibilita a comunicação direta com a parte interna do implante, permitindo que sinais processados em dispositivos externos, como computadores e smartphones, sejam utilizados na estimulação dos eletrodos cirurgicamente implantados no usuário. Isso permite a substituição temporária do processador externo do IC, facilitando a realização de experimentos e testes. No entanto, tais interfaces não estão disponíveis para a comunidade científica e são reservadas a poucos grupos de pesquisa que

estão diretamente ligados aos fabricantes de ICs. Essa condição dificulta o desenvolvimento científico, principalmente o surgimento de novas tecnologias não-proprietárias (Ali *et al.*, 2013).

Nessa circunstância, pesquisadores da Universidade do Texas em Dallas (*University of Texas at Dallas – UTD*) desenvolveram uma plataforma de pesquisa aberta para implantes cocleares (Ali; Lobo; Loizou, 2013), que permite a comunicação direta com a parte interna do implante por meio de uma antena de RF. O funcionamento da plataforma é ilustrado na Figura 12. O sinal acústico a ser enviado ao implantado pode ser captado diretamente por microfones instalados em um molde do tipo BTE, ou ainda obtido através de arquivos de áudio pré-gravados. Um arranjo de porta programável em campo (*Field Programmable Gate Array - FPGA*) é o responsável pelo fluxo da informação dentro da interface.

Figura 12 – Estrutura de funcionamento da plataforma de pesquisa da UTD.



Fonte: Adaptado de Ali H. (2016).

O sinal de áudio é direcionado por meio de comunicação via Wi-Fi ou USB para a unidade de processamento, que pode ser um computador, tablet ou até mesmo um smartphone. Essa unidade pode ser programada e realiza todo o processamento do sinal até a codificação em pulsos de corrente elétrica. A sequência de pulsos

resultante é então enviada novamente para a interface, onde é codificada de acordo com os protocolos de comunicação do implante, que pode variar de acordo com o fabricante. A transmissão do sinal é feita por meio de uma antena de RF até a parte interna do implante, que estimulará a cóclea.

A versão cedida pela UTD para o desenvolvimento desta pesquisa atende apenas o protocolo de comunicação com implantes cocleares da Cochlear®. Existem planos para incluir outras marcas de implante no futuro, o que irá acarretar em modificações apenas na programação do FPGA.

A plataforma de pesquisa atual possui duas entradas para moldes BTEs e duas antenas de RF, o que possibilita a pesquisa com implantes bilaterais e biauriculares. Além disso, oferece funcionalidades que permitem o desenvolvimento de estudos com estimulação eletroacústica (EAS), uma vez que conta com duas portas de áudio disponíveis no sistema. A plataforma também suporta a implementação de estratégias utilizando aplicativos de celular e a realização de experimentos do tipo *take-home trial*, nos quais o implantado utiliza o processamento em teste por um período de tempo prolongado, que pode durar dias.

A interface apresenta uma flexibilidade em relação à programação, pois suporta diversas linguagens, incluindo C, C++, JAVA, LabView e MatLab. No entanto, é importante destacar que o único módulo sem acesso à reprogramação é a FPGA. Essa restrição se dá por motivos de segurança, garantindo que os estímulos de corrente elétrica não sejam alterados, o que poderia causar danos ao paciente com implante coclear. A modificação desses estímulos é restrita aos fonoaudiólogos, que geralmente possuem conhecimento sobre a tecnologia utilizada em cada fabricante.

A unidade da interface disponível em nosso laboratório foi originalmente cedida ao Laboratório de Vibrações e Acústica da UFSC no final de 2016 pelos desenvolvedores da UTD, por meio de um acordo de colaboração estabelecido entre as duas instituições. É importante destacar que esse dispositivo foi o primeiro a ser cedido a um grupo fora dos Estados Unidos da América.

3 CONFORMAÇÃO DE FEIXES

Os *beamformers*, também conhecidos como conformadores de feixes, são filtros espaciais projetados para produzir uma região cônica de captação preferencial de feixes. Essa formação é projetada com a finalidade de enfatizar a captação dos sinais que são recebidos de uma determinada direção de chegada. Assim, o sinal desejado é captado e os demais sinais provenientes de direções indesejadas são atenuados (Van Veen; Buckley, 1988).

Inicialmente propostos para aplicações em telecomunicações, os conformadores de feixes são atualmente empregados em várias outras áreas como astronomia, acústica, sismologia, imagens médicas e no aprimoramento da fala, entre outras (Gershman *et al.*, 2003).

Esse tipo de filtro espacial é classificado de duas formas: conformadores de feixe fixos; e conformadores de feixe adaptativos. Os conformadores de feixe fixos e adaptativos são divididos conforme apresentado na Figura 13, de acordo com Lobato (2020).

Os conformadores de feixe fixos ou conformadores independentes de dados, são filtros espaciais projetados para propiciar uma resposta com ganho de magnitude unitária na direção desejada e atenuação nas demais direções. A implementação desse tipo de filtro é baseada em operações de soma e atraso (Van Veen; Buckley, 1988). Em geral, os conformadores de feixe fixos são utilizados quando a direção do alvo já é conhecida *a priori*, como em aplicações para celulares e carros. Nesses casos, o conformador fixo é projetado para focar na fonte alvo enquanto minimiza o ruído e a reverberação que chegam de outras direções (Gannot *et al.*, 2017). Exemplos de conformadores fixos (Figura 13) incluem os do tipo atraso e soma, que calculam a média dos sinais dos microfones atrasados (Van Veen; Buckley, 1988); o conformador de feixes superdiretivo, que maximiza o ganho da matriz de microfones para um campo de ruído difuso (Doclo *et al.*, 2007); e o conformador de feixes invariante em frequência, onde o padrão de feixes da matriz não tem nenhuma dependência da frequência para sinais de banda larga (Ward; Kennedy; Williamson, 1995).

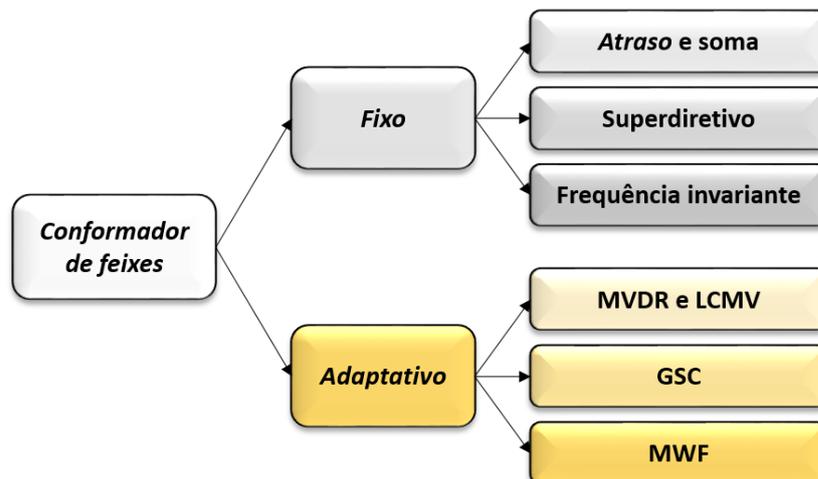
Por outro lado, têm-se os conformadores de feixe adaptativos. Eles recebem esse nome porque se auto ajustam, utilizando estatísticas dos sinais recebidos no

arranjo de microfones. Esse processo de otimização visa minimizar o ruído e enfatizar os sinais desejados (Van Veen; Buckley, 1988).

Um conformador de feixe adaptativo muito utilizado é o de mínima variância com resposta sem distorção (*Minimum Variance Distortionless Response - MVDR*) que tem como objetivo minimizar a potência do ruído na saída e preservar as componentes de interesse sem apresentar distorção. Assim, ele remove a máxima quantidade de ruído sem distorcer a fala de forma significativa. Para isso, utiliza-se um processo de minimização com restrição linear (Benesty; Chen; Huang, 2008; Bourgeois; Minker, 2009). Vale ressaltar que o MVDR é caracterizado por apresentar apenas uma restrição no processo de otimização.

Em cenários mais complexos, onde existe mais de uma fonte de interferência, é necessário um controle maior sobre o padrão do feixe. Nessas situações, é comum utilizar o conformador de mínima variância com restrições lineares (*Linearly Constrained Minimum Variance - LCMV*), que é uma extensão do MVDR para lidar com múltiplas restrições (Frost, 1972; Markovich; Gannot; Cohen, 2009; Resende; Bermudez, 2012).

Figura 13 – Classificação dos conformadores de feixes fixos e adaptativos.



Fonte: Autoria própria.

Além dos conformadores de feixe adaptativos mencionados, encontramos na literatura o cancelador generalizado de lóbulo lateral (*Generalized Sidelobe Canceller - GSC*) (Griffiths; Jim, 1982), que propõe um desacoplamento do filtro espacial em duas componentes. A primeira associada a uma da restrição fixa, que independe dos dados, e a segunda componente dependente diretamente dos sinais

captados (Buckley, 1987). Por fim, outra estratégia bastante utilizada é o filtro de Wiener multicanal (*Multichannel Wiener Filter* - MWF), que reduz a estimativa do erro quadrático médio entre o componente desejado e o sinal processado (Doclo; Moonen, 2002).

Considerando-se um cenário acústico com apenas uma fonte de interferência, é comum observar na literatura a utilização do método de mínimo erro quadrático médio (*Minimum Mean Square Error* - MMSE), que visa minimizar a variância do erro entre a magnitude da saída e do sinal desejado (Van Veen; Buckley, 1998).

O MVDR tem sido amplamente aplicado no domínio da frequência. As aplicações mais comuns são a redução de ruído, dereverberação, ênfase da fala e preservação de pistas acústicas. Muitas dessas aplicações estão associadas a aparelhos auditivos e sistemas para telecomunicações (Sun *et al.*, 2018; Ali, Van Waterschoot, Moonen, 2019; Malaver, Costa, 2020; Zedan *et al.*, 2021). Neste trabalho utiliza-se a estrutura e funcionamento do conformador de feixes MVDR em um arranjo de microfones associado a uma aplicação de implante coclear. Neste caso, o problema abordado é limitado à existência de apenas uma restrição, como será apresentado na sequência.

3.1 REDUÇÃO DE RUÍDO BASEADA EM ARRANJO DE MICROFONES PARA APLICAÇÃO EM IC

A redução de inteligibilidade da fala, especialmente quando há contaminação por ruído, é um problema recorrente para usuários de IC. No caso específico de implantes cocleares, as abordagens para redução de ruído se dividem em redução de ruído de canal único e multicanal.

Para o primeiro caso destacam-se as máscaras tempo-frequência, com três técnicas amplamente utilizadas: máscara binária (Wang; Brown, 2006), o filtro de Wiener (Loizou, 2013) e a máscara raiz de Wiener (Goehring *et al.*, 2019). No segundo caso, são utilizados dois ou mais microfones, cujos sinais são combinados de forma a ressaltar a fala de interesse e atenuar o sinal proveniente da direção da fonte de ruído (Büchner *et al.*, 2014; Chen; Gong, 2013; Hersbach *et al.*, 2013; Kordus *et al.*, 2015).

Este trabalho trata do uso de arranjos de microfones para a redução de ruído em aplicações que envolvem implantes cocleares.

3.2 O MODELO DO SINAL

No cenário em investigação considera-se um campo acústico monitorado por um conjunto de M microfones. O campo acústico pode ser constituído por emissões de várias fontes sonoras pontuais de ruído interferente, bem como por ruído de fundo, o qual pode ter origem acústica (campo difuso) ou elétrica, e pela fala na fonte $s(t)$. A fala de interesse que chega em cada um dos microfones é representada por $x_m(t)$ e os ruídos interferentes por $f_{p,m}(t)$ para $p = 1, 2, \dots, P$, em que p indica a fonte, P é o número de fontes interferentes e t é a variável de tempo contínuo.

O sinal captado em cada um dos microfones, em que $m \in \{1, 2, \dots, M\}$, é definido como (Figura 14):

$$y_m(t) = x_m(t) + v_m(t), \quad (1)$$

em que t é o índice de tempo contínuo e $v_m(t)$ é o ruído aditivo dado por:

$$v_m(t) = \sum_{p=1}^P f_{p,m}(t) + r_m(t), \quad (2)$$

sendo $r_m(t)$ o ruído de fundo (elétrico e/ou ambiental difuso). As contribuições da fala e das fontes interferentes em cada um dos microfones são definidas como:

$$x_m(t) = a_m(t) * s(t), \quad (3)$$

e

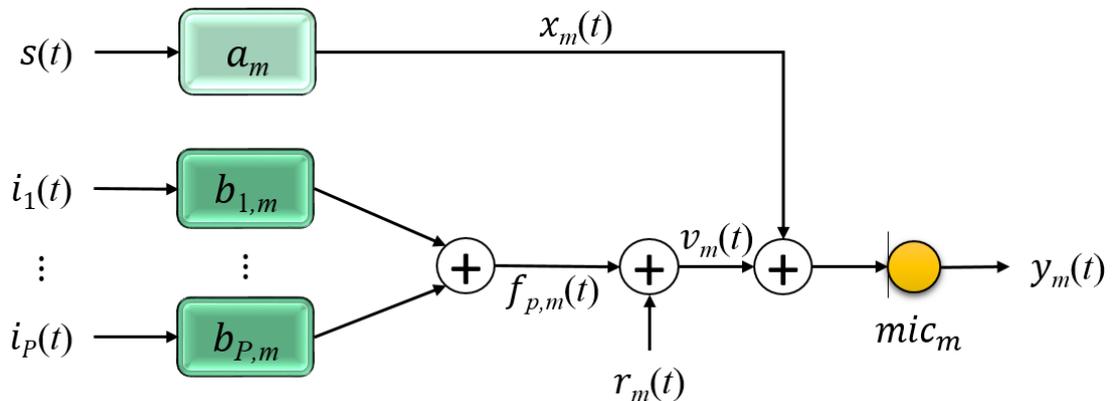
$$f_{p,m}(t) = b_{p,m}(t) * i_p(t), \quad (4)$$

em que $*$ é o operador de convolução e $a_m(t)$ e $b_{p,m}(t)$ são as respostas ao impulso associadas aos caminhos de propagação entre as fontes pontuais e os M microfones. As respostas ao impulso são conhecidas como “resposta ao impulso relacionada à cabeça” (*Head Related Impulse Response* - HRIR); $i_p(t)$ é o sinal da p -ésima fonte interferente.

A representação do problema em questão é apresentada na Figura 15. Assume-se que o IC está implantado em uma das orelhas e terá, nesta estrutura, $M/2$ microfones de captação disponíveis. Na orelha contralateral é colocado um molde,

semelhante a um aparelho auditivo, com outros $M/2$ microfones, cuja função é apenas a captação do som. Os sinais captados na orelha contralateral são enviados (geralmente por tecnologia sem fio) para o processador posicionado na orelha do implante (ipsilateral²), conforme ilustrado na Figura 16.

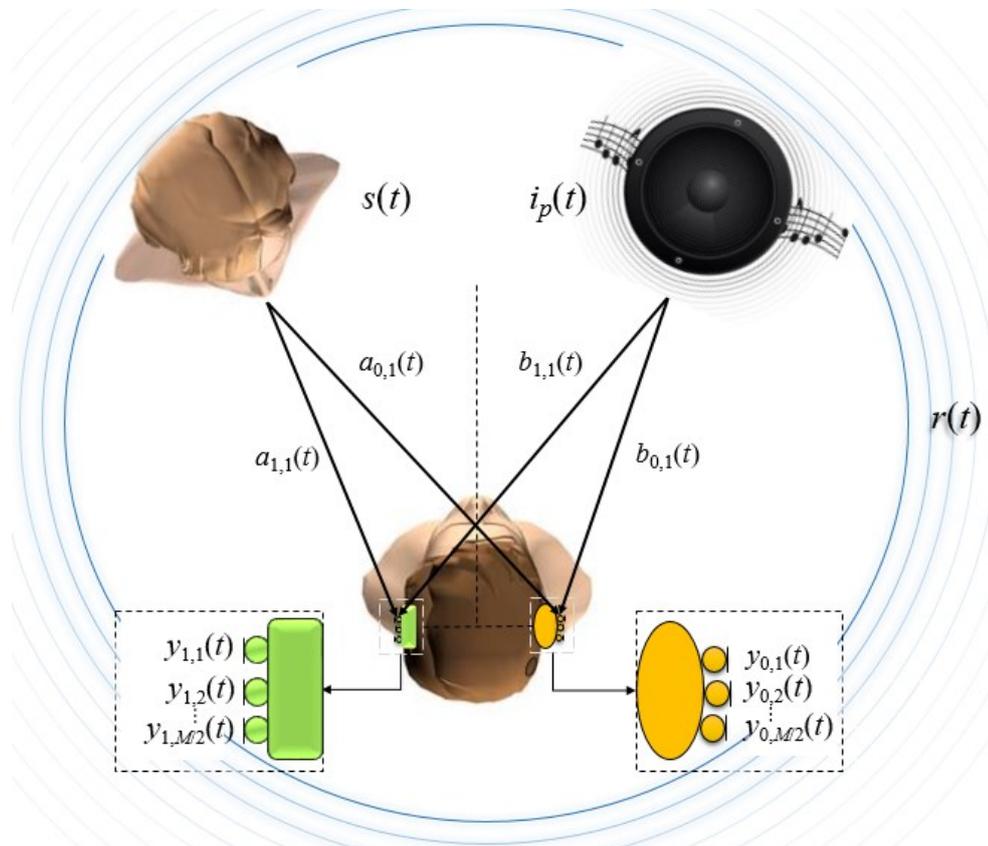
Figura 14 – Composição do sinal captado pelo m -ésimo microfone.



Fonte: Autoria própria.

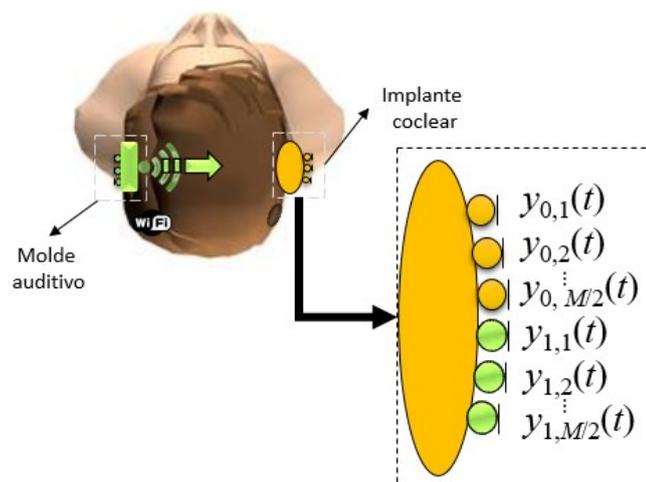
² Que está no mesmo lado do corpo.

Figura 15 – Aquisição do campo acústico com captação simultânea em ambas as orelhas. Neste modelo considera-se a fala desejada $s(t)$, o ruído interferente proveniente de uma fonte pontual $i_p(t)$ e o ruído difuso $r(t)$.



Fonte: Autoria própria.

Figura 16 – Disposição dos microfones de captação assumindo que o implante esteja na orelha direita.

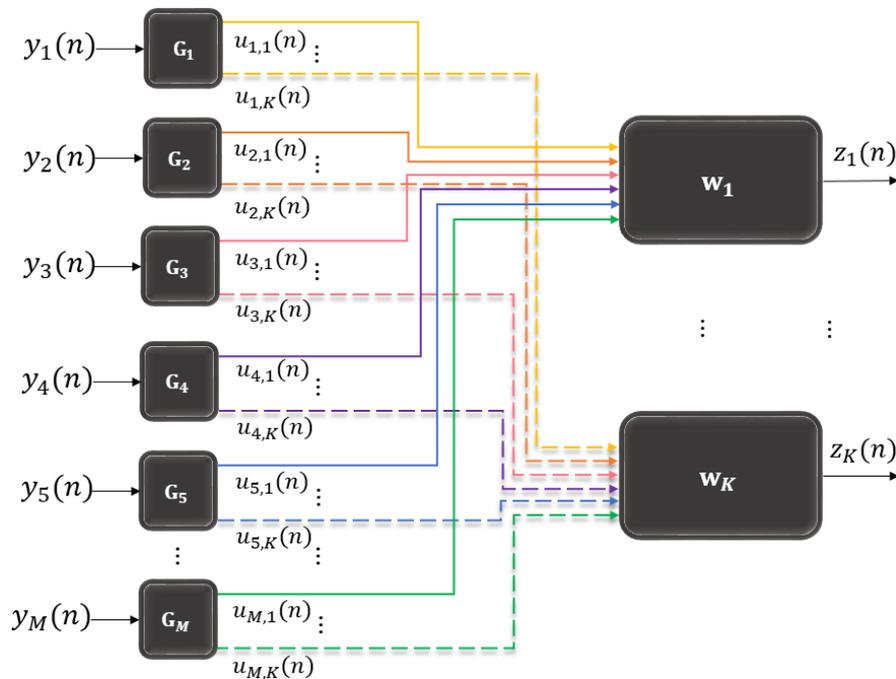


Fonte: Autoria própria.

3.3 MODELAGEM DA CADEIA DE PROCESSAMENTO DO IMPLANTE COCLEAR UTILIZANDO ARRANJO DE MICROFONES

O processo se inicia com a amostragem dos sinais captados pelos microfones e processamento por bancos de filtros, visando obter as componentes em sub-bandas, conforme representado na Figura 17 pelos blocos G_i . Esse processo separa cada um dos sinais captados em sub-bandas. Para facilitar a representação, os sinais captados são reescritos no domínio do tempo discreto.

Figura 17 – Processamento do IC para M microfones e K sub-bandas de estimulação.



Fonte: Autoria própria.

O sinal de entrada em cada um dos M bancos de filtros é definido por (1):

$$y_m(n) = x_m(n) + v_m(n). \quad (5)$$

Os sinais envolvidos em (5) são descritos por:

$$x_m(n) = \sum_{q=0}^{N_1-1} a_{m,q} s(n-q), \quad (6)$$

$$v_m(n) = \sum_{p=1}^P f_{p,m}(n) + r_m(n), \quad (7)$$

e

$$f_{p,m}(n) = \sum_{q=0}^{N_p-1} b_{p,m,q} i_p(n-q), \quad (8)$$

em que $\{ a_{m,0} a_{m,1} \dots a_{m,N_1-1} \}$ são os coeficientes da resposta ao impulso do sistema acústico entre a fonte de interesse (fala) e o m -ésimo microfone de captação, sendo N_1 o número de coeficientes; $\{ b_{p,m,0} b_{p,m,1} \dots b_{p,m,N_p-1} \}$ são os coeficientes das respostas ao impulso do sistema acústico entre a p -ésima fonte interferente e o m -ésimo microfone de captação, sendo N_p o número de coeficientes; $x_m(n)$ são as amostras do sinal de interesse captadas pelo m -ésimo microfone; $s(n)$ é o sinal da fonte de interesse no local onde o sinal é gerado; $f_{p,m}(n)$ são as amostras do p -ésimo sinal interferente no m -ésimo microfone; e $i_p(n)$ é o sinal da p -ésima fonte interferente no local de geração. Substituindo-se (6), (7) e (8) em (5), obtém-se:

$$y_m(n) = \sum_{q=0}^{N_1-1} a_{m,q} s(n-q) + \sum_{p=1}^P \sum_{q=0}^{N_p-1} b_{p,m,q} i_p(n-q) + r_m(n). \quad (9)$$

Considerando que os M bancos de filtros sejam idênticos, cada uma das K saídas é obtida pela soma de convolução entre o sinal de entrada do m -ésimo microfone e a resposta ao impulso da k -ésima sub-banda. Esta resposta ao impulso é definida pelos coeficientes $\{ g_{k,0} g_{k,1} \dots g_{k,N-1} \}$, no caso de aproximação por um filtro de resposta finita ao impulso, em que N representa o número de amostras com magnitude não desprezível (Slaney, 1998). Dessa forma, a saída de cada uma das sub-bandas do banco de filtros, para cada um dos microfones, é dada por:

$$u_{m,k}(n) = \sum_{l=0}^{N-1} g_{k,l} y_m(n-l). \quad (10)$$

em que $u_{m,k}(n)$ é a saída da k -ésima sub-banda do m -ésimo microfone.

Assumindo-se uma janela de C amostras de observação para o conformador, temos:

$$\mathbf{u}_{m,k}(n) = \mathbf{G}_k \mathbf{y}_m(n), \quad (11)$$

em que $\mathbf{u}_{m,k}(n) = [u_{m,k}(n) u_{m,k}(n-1) \dots u_{m,k}(n-C+1)]^T$ possui dimensão $C \times 1$; $\mathbf{y}_m(n) = [y_m(n) y_m(n-1) \dots y_m(n-N-C+2)]^T$ possui dimensão $(N+C-1) \times 1$; e

$$\mathbf{G}_k = \begin{bmatrix} \mathbf{g}_{k,0} & \mathbf{g}_{k,1} & \cdots & & & & & & \\ 0 & \mathbf{g}_{k,0} & \mathbf{g}_{k,1} & \ddots & & & & & \\ 0 & 0 & \mathbf{g}_{k,0} & \mathbf{g}_{k,1} & \ddots & & & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & & & & \end{bmatrix}, \quad (12)$$

possui dimensão $C \times (N+C-1)$. Dessa forma, pode-se então definir a saída do arranjo de microfones a cada instante de tempo $z_k(n)$ (Figura 19), como:

$$z_k(n) = \sum_{m=1}^M \mathbf{u}_{m,k}^T(n) \mathbf{w}_{m,k} = \sum_{m=1}^M \mathbf{y}_m^T(n) \mathbf{G}_k^T \mathbf{w}_{m,k}, \quad (13)$$

em que $\mathbf{w}_{m,k} = [w_{m,k,0} \ w_{m,k,1} \ \dots \ w_{m,k,C-1}]^T$ possui dimensão $C \times 1$. A Equação (13) pode ser expressa também sob a forma mais conveniente de um único produto interno, dado por:

$$z_k(n) = \mathbf{u}_k^T(n) \mathbf{w}_k, \quad (14)$$

em que $\mathbf{w}_k = [\mathbf{w}_{1,k}^T \ \mathbf{w}_{2,k}^T \ \dots \ \mathbf{w}_{M,k}^T]^T$ e $\mathbf{u}_k(n) = [\mathbf{u}_{1,k}^T(n) \ \mathbf{u}_{2,k}^T(n) \ \dots \ \mathbf{u}_{M,k}^T(n)]^T$ possuem dimensão $MC \times 1$ e são, respectivamente, formados a partir do empilhamento dos vetores $\mathbf{u}_{m,k}$ e $\mathbf{w}_{m,k}$ para $m = 1, 2, \dots, M$. Portanto, o vetor $\mathbf{u}_k(n)$ é definido como:

$$\mathbf{u}_k(n) = \mathbf{\Gamma}_k \mathbf{y}(n), \quad (15)$$

em que $\mathbf{y}(n) = [\mathbf{y}_1^T(n) \ \mathbf{y}_2^T(n) \ \dots \ \mathbf{y}_M^T(n)]^T$ possui dimensão $M(N+C-1) \times 1$ e

$$\mathbf{\Gamma}_k = \begin{bmatrix} \mathbf{G}_k & \mathbf{0} & \cdots & & \\ \mathbf{0} & \mathbf{G}_k & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_k & \end{bmatrix}, \quad (16)$$

é uma matriz bloco-diagonal de dimensão $MC \times M(N+C-1)$.

Portanto:

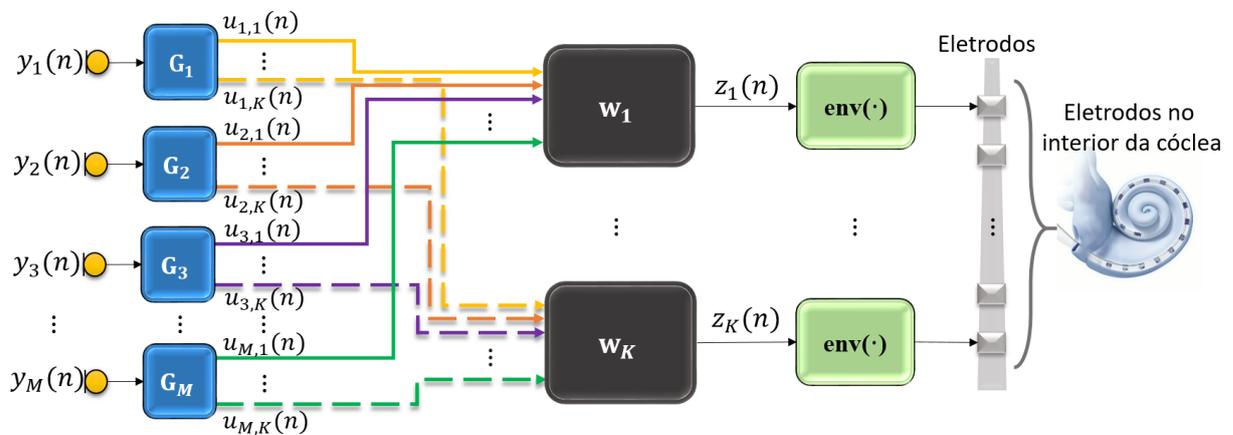
$$z_k(n) = \mathbf{y}^T(n) \mathbf{\Gamma}_k^T \mathbf{w}_k. \quad (17)$$

A Figura 18 rerepresenta a cadeia de processamento de sinais do IC, destacando a topologia de redução de ruído multicanal. Neste contexto, são considerados M bancos de filtros idênticos, cada um com K sub-bandas de

estimulação. Os sinais provenientes desta etapa são filtrados pelos arranjos de microfones, w_k , e geram o sinal representado na Equação (17) e na Figura 18 por $z_k(n)$.

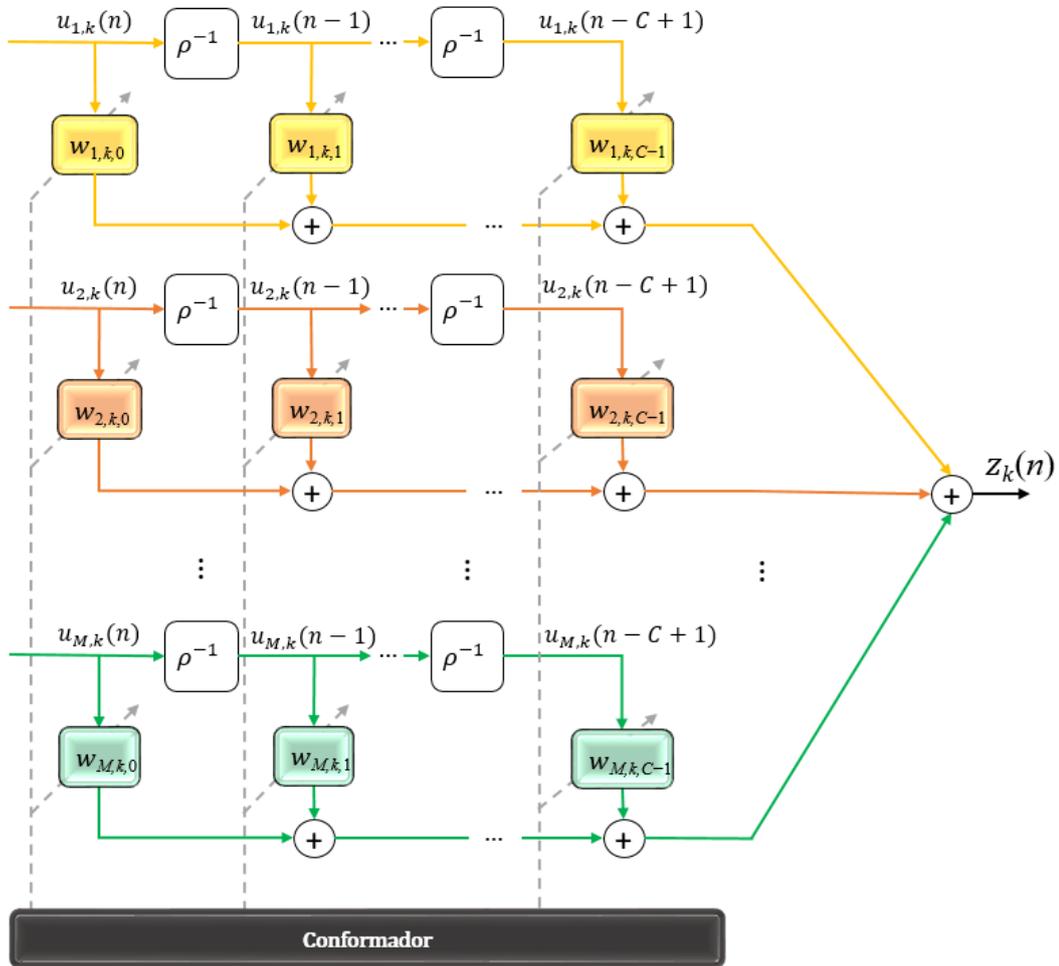
Após o processo de conformação, é realizada a estimativa da envoltória temporal em cada sub-banda. Em sequência, é aplicada uma estratégia de codificação e então a informação é transformada em pulsos de corrente elétrica para estimular diferentes locais da cóclea (Doclo; Moonen, 2002).

Figura 18 – Cadeia de processamento de sinal para ICs com múltiplos microfones. M microfones e K sub-bandas de estimulação.



Fonte: Autoria própria.

Figura 19 – Estrutura interna do arranjo de microfones para cada sub-banda k e atraso ρ .



Fonte: Autoria própria.

3.4 MODELAGEM DO MÉTODO DE REDUÇÃO DE RUÍDO NO DOMÍNIO DO TEMPO

Substituindo-se (5) em (17) obtém-se:

$$z_k(n) = [\mathbf{x}(n) + \mathbf{v}(n)]^T \mathbf{\Gamma}_k^T \mathbf{w}_k = \mathbf{x}^T(n) \mathbf{\Gamma}_k^T \mathbf{w}_k + \mathbf{v}^T(n) \mathbf{\Gamma}_k^T \mathbf{w}_k, \quad (18)$$

em que $\mathbf{x}(n) = [x_1^T(n) \ x_2^T(n) \ \dots \ x_M^T(n)]^T$ e $\mathbf{v}(n) = [v_1^T(n) \ v_2^T(n) \ \dots \ v_M^T(n)]^T$ possuem dimensão $M(N+C-1) \times 1$; e $\mathbf{x}_m(n) = [x_m(n) \ x_m(n-1) \ \dots \ x_m(n-N-C+2)]^T$ e $\mathbf{v}_m(n) = [v_m(n) \ v_m(n-1) \ \dots \ v_m(n-N-C+2)]^T$ possuem dimensão $(N+C-1) \times 1$. Portanto, $z_k(n)$ é composto

por duas parcelas, ou seja, o resultado do processo de filtragem sobre a fala e sobre o ruído.

A potência total na saída é dada por:

$$E\{z_k^2(n)\} = E\left\{\left[\mathbf{x}^T(n)\Gamma_k^T \mathbf{w}_k + \mathbf{v}^T(n)\Gamma_k^T \mathbf{w}_k\right]^2\right\}, \quad (19)$$

em que:

$$E\{z_k^2(n)\} = \mathbf{w}_k^T \Gamma_k E\{\mathbf{x}(n)\mathbf{x}^T(n)\} \Gamma_k^T \mathbf{w}_k + \mathbf{w}_k^T \Gamma_k E\{\mathbf{x}(n)\mathbf{v}^T(n)\} \Gamma_k^T \mathbf{w}_k + \mathbf{w}_k^T \Gamma_k E\{\mathbf{v}(n)\mathbf{x}^T(n)\} \Gamma_k^T \mathbf{w}_k + \mathbf{w}_k^T \Gamma_k E\{\mathbf{v}(n)\mathbf{v}^T(n)\} \Gamma_k^T \mathbf{w}_k. \quad (20)$$

Considera-se que a fala e o ruído são variáveis aleatórias independentes de média zero. Logo,

$$E\{\mathbf{x}(n)\mathbf{v}^T(n)\} = E\{\mathbf{v}(n)\mathbf{x}^T(n)\} = \mathbf{0}. \quad (21)$$

Assim, substituindo-se (21) em (20):

$$E\{z_k^2(n)\} = \mathbf{w}_k^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{w}_k + \mathbf{w}_k^T \Gamma_k \mathbf{R}_{vv} \Gamma_k^T \mathbf{w}_k, \quad (22)$$

em que

$$\mathbf{R}_{xx} = E\{\mathbf{x}(n)\mathbf{x}^T(n)\}, \quad \mathbf{R}_{vv} = E\{\mathbf{v}(n)\mathbf{v}^T(n)\}. \quad (23)$$

Considerando $\mathbf{R}_{yy} = \mathbf{R}_{xx} + \mathbf{R}_{vv}$ e substituindo-se essa relação na Equação (22) obtém-se:

$$E\{z_k^2(n)\} = \mathbf{w}_k^T \Gamma_k \mathbf{R}_{yy} \Gamma_k^T \mathbf{w}_k. \quad (24)$$

Na definição original do conformador de feixes, os coeficientes \mathbf{w} são selecionados com o objetivo de minimizar a potência do sinal na saída do sistema, $E\{z_k^2(n)\}$. Para essa situação, o conformador de feixes do tipo MVDR e o filtro de Wiener multicanal são idênticos a menos de um escalar (Benesty; Souden; Chen, 2013), conforme detalhado no Apêndice A. No entanto, diversos estudos optam por definir a função custo a ser minimizada como uma função de \mathbf{R}_{vv} , argumentando que o ruído possui uma menor variabilidade estatística ao longo do tempo em relação à fala e, portanto, suas estimativas são mais precisas (Benesty; Souden; Chen, 2013).

No contexto deste trabalho, como proposta de função custo a ser minimizada, será adotada a seguinte expressão:

$$J_k(n) = \mathbf{w}_k^T \Gamma_k [\alpha \mathbf{R}_{xx} + \mathbf{R}_{vv}] \Gamma_k^T \mathbf{w}_k, \quad (25)$$

dessa forma, quando $\alpha = 1$ a função custo depende de \mathbf{R}_{yy} e quando $\alpha = 0$ tem-se apenas \mathbf{R}_{vv} .

A solução trivial do processo de minimização de $J_k(n)$ é $\mathbf{w}_k^{\text{opt}} = \mathbf{0}$. Dessa forma, como o objetivo primário é a estimação da fala, é necessário incluir restrições no processo de minimização para evitar que o sinal de interesse também seja cancelado. Uma possibilidade é restringir o conjunto de soluções àquelas que não afetem o sinal de interesse, ou seja, utilizando (18):

$$z_k^{\text{fala}}(n) = \mathbf{x}^T(n) \Gamma_k^T \mathbf{w}_k = x_{\text{REF},k}(n). \quad (26)$$

O sinal $x_{\text{REF},k}(n)$ é o sinal de fala captado pelo microfone de referência (geralmente o microfone frontal, $m = 1$) na sub-banda k . A partir de (15) tem-se que:

$$x_{\text{REF},k}(n) = \mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}}, \quad (27)$$

em que \mathbf{q}_{REF} é um vetor de dimensão $MC \times 1$ cujos elementos são nulos, com uma única exceção, a qual é igual à unidade. Para $\text{REF} = 1, 2, \dots, M$, a posição do valor não nulo é $(\text{REF}-1) \cdot C + 1$. Assim, substituindo-se (27) em (26) chega-se em:

$$\mathbf{x}^T(n) \Gamma_k^T \mathbf{w}_k = \mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}}. \quad (28)$$

O vetor de sinal desejado $\mathbf{x}(n)$ não é conhecido e, portanto, não pode ser usado diretamente. Entretanto, a utilização de um detector de fala permite a discriminação de períodos de fala contaminada e apenas ruído. Assumindo que o ruído é aditivo e que tanto a fala quanto o ruído são aproximadamente estacionários dentro de uma determinada janela de tempo, é possível obter estimativas da matriz de correlação da fala através de $\mathbf{R}_{xx}(n) = \mathbf{R}_{yy}(n) - \mathbf{R}_{vv}(n)$. Portanto, pré-multiplicando (28) por $\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{x}(n)$

$$\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{x}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{w}_k = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{x}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}}. \quad (29)$$

e tomando-se o valor esperado de (29), obtém-se:

$$E\{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{x}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{w}_k\} = E\{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{x}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}}\}. \quad (30)$$

O que resulta em

$$\mathbf{q}_{\text{REF}}^T \Gamma_k E\{\mathbf{x}(n) \mathbf{x}^T(n)\} \Gamma_k^T \mathbf{w}_k = \mathbf{q}_{\text{REF}}^T \Gamma_k E\{\mathbf{x}(n) \mathbf{x}^T(n)\} \Gamma_k^T \mathbf{q}_{\text{REF}}, \quad (31)$$

chegando então a:

$$\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}} \Gamma_k^T \mathbf{w}_k = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}} \Gamma_k^T \mathbf{q}_{\text{REF}}. \quad (32)$$

As Equações (30) a (32) estabelecem que a correlação cruzada entre o sinal de saída (considerando apenas o sinal desejado na entrada do sistema) e o desejado, em cada sub-banda, deve ser idêntica à autocorrelação (potência) do sinal desejado.

Utilizando (25) e (32) pode-se definir um processo de otimização para a obtenção dos coeficientes ótimos do arranjo de microfones baseado no conformador de feixes MVDR, no domínio tempo, de tal forma que:

$$\begin{aligned} \mathbf{w}_k^{\text{opt}} &= \arg.\min_{\mathbf{w}_k} \alpha \mathbf{w}_k^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{w}_k + \mathbf{w}_k^T \Gamma_k \mathbf{R}_{\text{vv}}(n) \Gamma_k^T \mathbf{w}_k \\ \text{s.t. } & \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{w}_k = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} \end{aligned} \quad (33)$$

Uma outra abordagem para o arranjo de microfones, baseada no conformador de feixes ótimo, é apresentada no Apêndice B, a qual é equivalente ao procedimento apresentado por Benesty, Souden e Chen (2013) (Benesty; Souden; Chen, 2013). O procedimento matemático para se obter a solução fechada para o problema descrito em (33) é apresentado na próxima seção.

3.5 ESTIMAÇÃO DOS COEFICIENTES ÓTIMOS BASEADO NO CONFORMADOR MVDR

A partir do processo de minimização definido em (33), busca-se encontrar o vetor de coeficientes ótimos que solucione o problema de otimização. Uma formulação fechada baseada na obtenção dos coeficientes ótimos do conformador MVDR descrito em (33) pode ser obtida através da técnica de multiplicadores de Lagrange. Definindo

o vetor K -dimensional de multiplicadores de Lagrange como λ_k , em que $K = 1$, visto que há apenas uma restrição no sistema (Resende, 1996), temos que:

$$L(\mathbf{w}_k, \lambda_k) = \alpha \mathbf{w}_k^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{w}_k + \mathbf{w}_k^T \Gamma_k \mathbf{R}_{vv} \Gamma_k^T \mathbf{w}_k + \lambda_k [\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{w}_k - \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}}] \quad (34)$$

Calculando a derivada parcial de (34) em relação a \mathbf{w}_k , obtém-se:

$$\frac{\partial L(\mathbf{w}_k, \lambda_k)}{\partial \mathbf{w}_k} = \alpha \left[\Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{w}_k + (\Gamma_k \mathbf{R}_{xx} \Gamma_k^T)^T \mathbf{w}_k \right] + \Gamma_k \mathbf{R}_{vv} \Gamma_k^T \mathbf{w}_k + (\Gamma_k \mathbf{R}_{vv} \Gamma_k^T)^T \mathbf{w}_k + \lambda_k \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}} \quad (35)$$

Sabendo que as matrizes de correlação são simétricas, pode-se reescrever (35) como:

$$\frac{\partial L(\mathbf{w}_k, \lambda_k)}{\partial \mathbf{w}_k} = 2\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{w}_k + 2\Gamma_k \mathbf{R}_{vv} \Gamma_k^T \mathbf{w}_k + \lambda_k \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}} \quad (36)$$

Igualando (36) a zero (Luenberger, 1969) chega-se em:

$$2\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{w}_k + 2\Gamma_k \mathbf{R}_{vv} \Gamma_k^T \mathbf{w}_k + \lambda_k \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}} = \mathbf{0} \quad (37)$$

Manipulando (37), obtém-se:

$$\mathbf{w}_k = -\left(2\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T + 2\Gamma_k \mathbf{R}_{vv} \Gamma_k^T\right)^{-1} \lambda_k \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}} \quad (38)$$

Portanto, o arranjo de microfones baseado no conformador ótimo é dado por:

$$\mathbf{w}_k^{\text{opt}} = -\lambda_k^{\text{opt}} \left(2\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T + 2\Gamma_k \mathbf{R}_{vv} \Gamma_k^T\right)^{-1} \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}} \quad (39)$$

O vetor $\mathbf{w}_k^{\text{opt}}$ obtido em (39) deve satisfazer a restrição imposta em (33). Assim, substituindo-se $\mathbf{w}_k^{\text{opt}}$ em (33) obtém-se:

$$\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{w}_k^{\text{opt}} = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}} \quad (40)$$

Na sequência, substituindo (39) em (40):

$$\begin{aligned} -\lambda_k^{\text{opt}} \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \left(2\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T + 2\Gamma_k \mathbf{R}_{vv} \Gamma_k^T\right)^{-1} \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}} \\ = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}} \end{aligned} \quad (41)$$

Reorganizando (41):

$$\lambda_k^{\text{opt}} = -\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}} \cdot \left[\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \left(2\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T + 2\Gamma_k \mathbf{R}_{vv} \Gamma_k^T \right)^{-1} \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}} \right]^{-1}. \quad (42)$$

Substituindo-se (42) em (39) chega-se em:

$$\mathbf{w}_k^{\text{opt}} = \frac{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}}}{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \left(2\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T + 2\Gamma_k \mathbf{R}_{vv} \Gamma_k^T \right)^{-1} \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}}} \cdot \frac{1}{\left(2\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T + 2\Gamma_k \mathbf{R}_{vv} \Gamma_k^T \right)^{-1} \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}}}. \quad (43)$$

Rearranjando (43),

$$\mathbf{w}_k^{\text{opt}} = \frac{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}}}{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \left[2 \left(\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T + \Gamma_k \mathbf{R}_{vv} \Gamma_k^T \right) \right]^{-1} \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}}} \cdot \frac{1}{\left[2 \left(\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T + \Gamma_k \mathbf{R}_{vv} \Gamma_k^T \right) \right]^{-1} \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}}}. \quad (44)$$

Simplificando (44), chega-se finalmente em (45) que é a solução fechada para o problema de minimização apresentado em (33).

$$\mathbf{w}_k^{\text{opt}} = \frac{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}}}{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \left(\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T + \Gamma_k \mathbf{R}_{vv} \Gamma_k^T \right)^{-1} \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}} \left(\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T + \Gamma_k \mathbf{R}_{vv} \Gamma_k^T \right)^{-1} \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}}}, \quad (45)$$

em que $\mathbf{w}_k^{\text{opt}}$ tem dimensão $MC \times 1$.

O Quadro 2 apresenta o pseudocódigo para a obtenção da solução fechada apresentada na Equação (45).

Quadro 2 – Pseudocódigo para a solução fechada apresentada em (45).

Entradas: $\alpha, \mathbf{x}, \mathbf{v}, \mathbf{G}_k$

Saídas: $\mathbf{w}_k^{\text{opt}}$

- Calcule Γ_k de acordo com (16)
- Calcule \mathbf{R}_{xx} e \mathbf{R}_{vv} de acordo com (23)
- Determine \mathbf{q}_{REF} de acordo com (27)
- Calcule $\mathbf{w}_k^{\text{opt}}$ de acordo com (45)

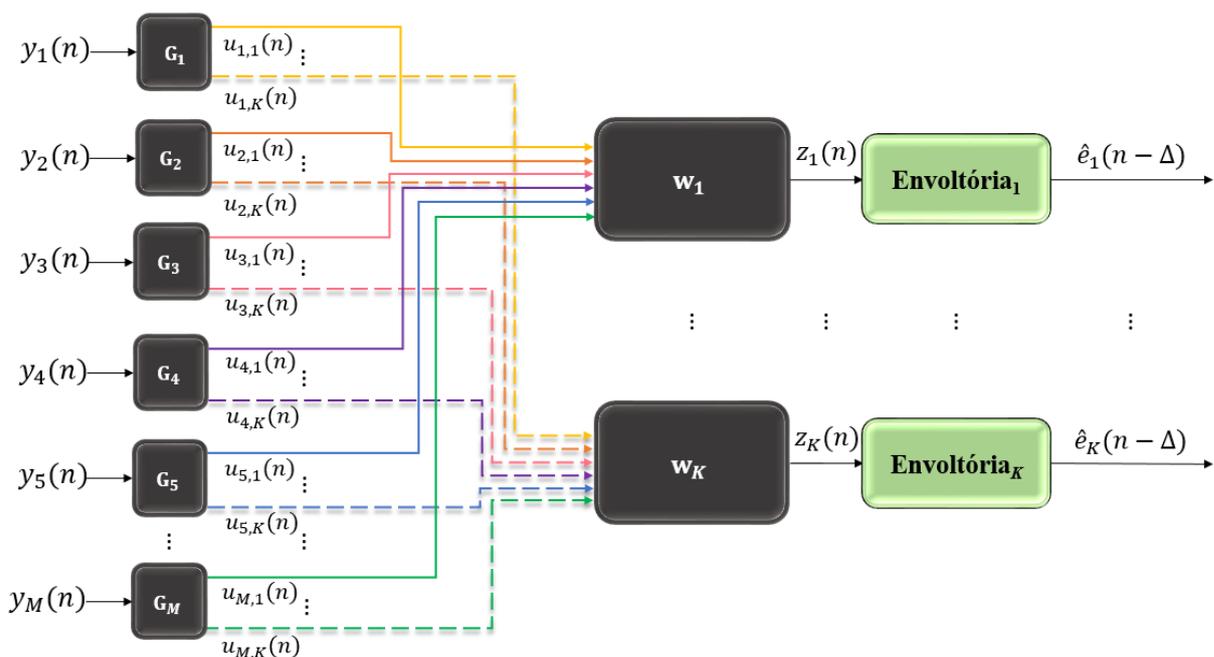
Retorna $\mathbf{w}_k^{\text{opt}}$

4 ESTIMAÇÃO DA ENVOLTÓRIA TEMPORAL

Neste capítulo, são feitas considerações acerca da estimativa da envoltória de um sinal no domínio do tempo discreto. Para isso, é utilizada a arquitetura associada à aplicação de implantes cocleares, onde o sinal obtido na saída de cada um dos K arranjos de microfones é denominado de $z_k(n)$, conforme mostrado na Figura 20. A estimativa da envoltória do sinal $z_k(n)$ é denotada por $\hat{e}_k(n)$. Os detalhes sobre o bloco de estimação da envoltória são apresentados na Figura 21.

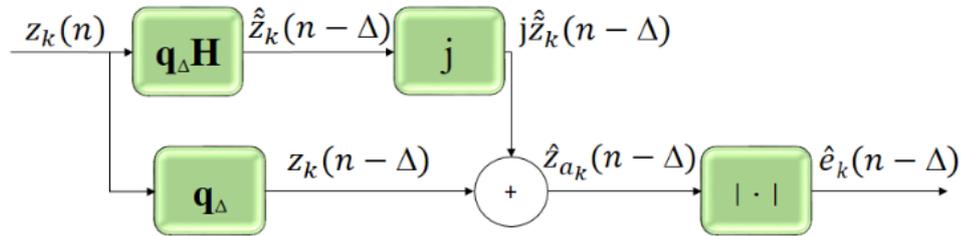
De acordo com a Figura 21, cada um dos sinais $z_k(n)$ é processado pela transformada de Hilbert truncada (τ e Δ finitos) resultando em $\tilde{z}_k(n)$. Isso é necessário visto que implantes cocleares são sistemas de processamento em tempo real com capacidade computacional limitada (Zeng *et al.*, 2008). O bloco q_Δ representa um atraso de Δ amostras; j é a unidade imaginária; $\hat{z}_{ak}(n)$ é uma aproximação do sinal analítico de entrada; $\hat{e}_k(n)$ é uma aproximação da envoltória do sinal de entrada; Δ é o atraso do processamento e τ é o tamanho do filtro de Hilbert, em que $\tau=2\Delta+1$ (Kak, 2014).

Figura 20 – Estimação da envoltória dos sinais de saída do arranjo de microfones.



Fonte: Autoria própria.

Figura 21 – Diagrama de blocos do estimador de envoltória utilizando a transformada de Hilbert. Processamento no domínio tempo, amostra a amostra.



Fonte: Autoria própria.

A envoltória do sinal $\hat{z}_k(n)$ é dada por:

$$e_k(n) = |\hat{z}_k(n)|, \quad (46)$$

em que o sinal analítico $\hat{z}_{ak}(n)$ é obtido como (Poularikas, 2018):

$$\hat{z}_{ak}(n) = \hat{z}_k(n) + j\tilde{\hat{z}}_k(n). \quad (47)$$

O sinal $\hat{z}_k(n-\Delta)$ pode ser escrito como:

$$\hat{z}_k(n-\Delta) = \mathbf{q}_\Delta^T \mathbf{z}_k(n), \quad (48)$$

em que $\mathbf{q}_\Delta = [0 \dots 0 \ 1 \ 0 \dots 0]^T$ possui dimensão $(\Delta+\tau) \times 1$, sendo composto por zeros, a menos do Δ -ésimo elemento, que tem valor unitário. O vetor $\mathbf{z}_k(n) = [\hat{z}_k(n) \ \hat{z}_k(n-1) \dots \hat{z}_k(n-\Delta-\tau-1)]^T$ possui dimensão $(\Delta+\tau) \times 1$.

Partindo da Equação (17) pode-se escrever:

$$\mathbf{z}_k(n) = \begin{bmatrix} \hat{z}_k(n) \\ \hat{z}_k(n-1) \\ \vdots \\ \hat{z}_k(n-\Delta-\tau-1) \end{bmatrix} = \begin{bmatrix} \mathbf{w}_k^T \mathbf{\Gamma}_k \mathbf{y}(n) \\ \mathbf{w}_k^T \mathbf{\Gamma}_k \mathbf{y}(n-1) \\ \vdots \\ \mathbf{w}_k^T \mathbf{\Gamma}_k \mathbf{y}(n-\Delta-\tau-1) \end{bmatrix}. \quad (49)$$

Separando os elementos de (49) na forma de um produto matricial, pode-se escrever:

$$\mathbf{z}_k(n) = \begin{bmatrix} \mathbf{w}_k^T & \mathbf{0} & \dots & \Gamma_k & \mathbf{0} & \dots & \mathbf{y}(n) \\ \mathbf{0} & \mathbf{w}_k^T & \dots & \mathbf{0} & \Gamma_k & \dots & \mathbf{y}(n-1) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{y}(n-\Delta-\tau-1) \end{bmatrix}, \quad (50)$$

e, portanto:

$$\hat{\mathbf{z}}_k(n-\Delta) = \mathbf{q}_\Delta^T \mathbf{z}_k(n) = \mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{y}}(n), \quad (51)$$

em que a matriz \mathbf{W}_k possui dimensão $MC(\Delta+\tau) \times (\Delta+\tau)$, sendo composta por repetições de \mathbf{w}_k em sua diagonal principal; a matriz \mathbf{D}_k com dimensão $MC(\Delta+\tau) \times M[N+C-1(\Delta+\tau)]$ representa o processo de decomposição dos sinais de entrada pelo filtro gammatone na k -ésima sub-banda; e o vetor $\bar{\mathbf{y}}(n)$, com dimensão $M[N+C-1(\Delta+\tau)] \times 1$, contém o sinal de fala contaminada presente na entrada do sistema.

Para obter $\hat{\mathbf{z}}_k(n-\Delta)$ pode-se realizar um processo semelhante ao utilizado em (48)-(51). Define-se a estimativa do vetor do sinal de saída do arranjo de microfones sujeito à transformação de Hilbert como:

$$\tilde{\mathbf{z}}_k(n), \quad (52)$$

em que a matriz de Hilbert (\mathbf{H}) possui dimensão $(\Delta+\tau) \times (\Delta+\tau)$ e é definida como (Poularikas, 2018):

$$\mathbf{H} = \frac{2}{\pi} \begin{bmatrix} 0 & 1/(1) & 0 & 1/(3) & \dots \\ 1/(-1) & 0 & 1/(1) & 0 & \ddots \\ 0 & 1/(-1) & 0 & 1/(1) & \ddots \\ 1/(-3) & 0 & 1/(-1) & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}, \quad (53)$$

temos então:

$$\tilde{\mathbf{z}}_k(n) = \mathbf{q}_\Delta^T \mathbf{H} \mathbf{z}_k(n). \quad (54)$$

Substituindo-se (51) em (54) chega-se a:

$$\tilde{\mathbf{z}}_k(n) = \mathbf{q}_\Delta^T \mathbf{H} \mathbf{z}_k(n) = \mathbf{q}_\Delta^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{y}}(n). \quad (55)$$

Assim, substituindo-se (51) e (55) em (47) e o resultado em (46), obtém-se:

$$\hat{e}_k(n-\Delta) = \left| \mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{y}}(n) + j \mathbf{q}_\Delta^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{y}}(n) \right|. \quad (56)$$

Desenvolvendo (56),

$$\hat{e}_k(n-\Delta) = \sqrt{\left(\mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{y}}(n) \right)^2 + \left(\mathbf{q}_\Delta^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{y}}(n) \right)^2}. \quad (57)$$

O sinal calculado em (57) representa a estimativa instantânea da envoltória temporal do sinal filtrado pelo arranjo de microfones, na k -ésima sub-banda e no instante $n-\Delta$.

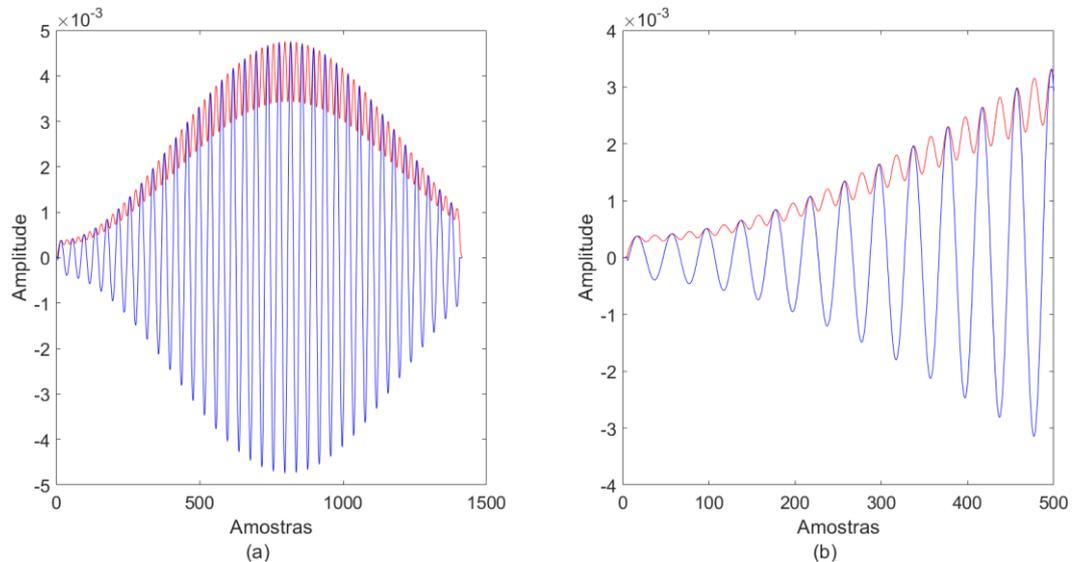
4.1 EXEMPLIFICAÇÃO DA ESTIMAÇÃO DA ENVOLTÓRIA DOS SINAIS

A estimativa da envoltória temporal de um sinal depende diretamente da dimensão da matriz \mathbf{H} , que neste contexto é formada de acordo com os parâmetros Δ e τ .

Na Figura 22 e na Figura 23 são apresentadas duas estimativas da envoltória de um sinal produzido artificialmente. O sinal artificial, com 1600 amostras, foi produzido a partir de uma janela Hamming multiplicada por um sinal senoidal com frequência fundamental de 400 Hz. Nas figuras são apresentados os resultados obtidos para a envoltória temporal utilizando $\tau=2\Delta+1$. O sinal em vermelho representa a envoltória estimada para diferentes valores de Δ e em azul o sinal artificial.

A Figura 22(a) apresenta a envoltória, em vermelho e espiculado, para os parâmetros de cálculo $\Delta=5$ e $\tau=11$. Na Figura 22(b) tem-se um recorte das 500 amostras iniciais da Figura 22(a), deixando evidente a envoltória espiculada sobreposta ao sinal em azul.

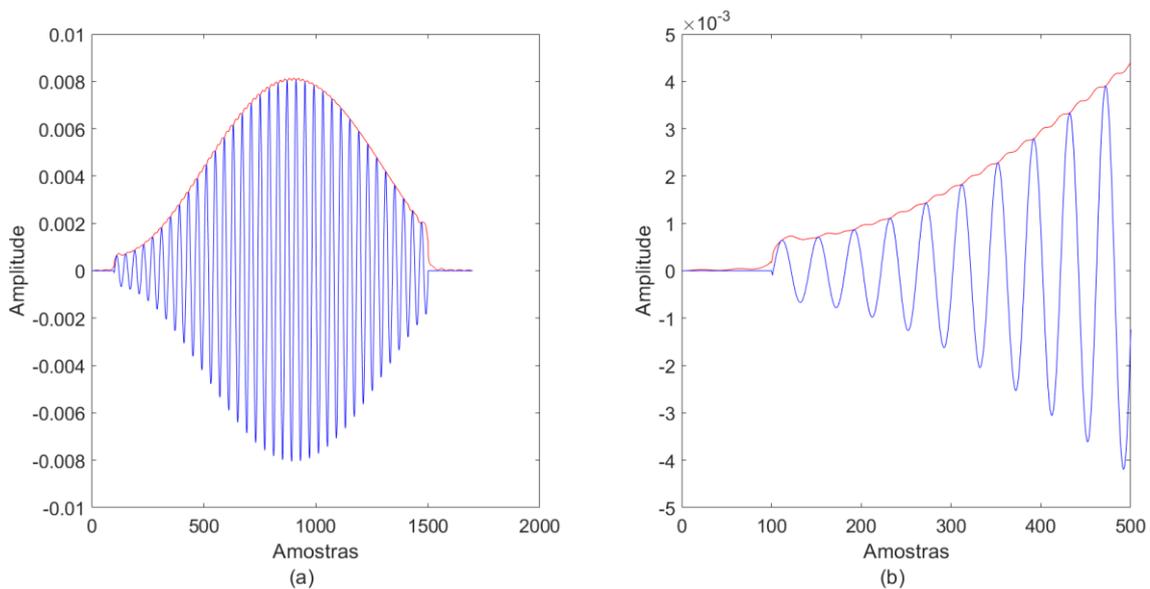
Figura 22 – (a) Sinal artificial (azul) e a estimativa de sua envoltória (vermelho) para $\Delta=5$; (vermelho); (b) Corte de (a) para as 500 primeiras amostras.



Fonte: Autoria própria.

Para a Figura 23, foi fixado o valor de $\Delta=100$. A envoltória temporal, em vermelho e pouco espiculada, e o sinal artificial (azul) são apresentados na Figura 23(a); enquanto na Figura 23(b) é mostrado um recorte das primeiras 500 amostras da Figura 23(a), onde fica evidente que a envoltória não está totalmente alisada.

Figura 23 – (a) Sinal artificial (azul) e a estimativa de sua envoltória (vermelho) para $\Delta=100$; (vermelho); (b) Corte de (a) para as 500 primeiras amostras.



Fonte: Autoria própria.

A partir desses resultados foi possível verificar que a acurácia da estimativa da envoltória instantânea $\hat{e}_k(n-\Delta)$ é significativamente influenciada pela escolha dos parâmetros τ e Δ , o que implica diretamente na complexidade computacional para a resolução da Equação (57).

5 PROPOSTA

No Capítulo 3 foi apresentada uma estratégia para redução de ruído em implantes cocleares utilizando o arranjo de microfones baseado no conformador de feixes MVDR com a restrição de que a correlação cruzada entre o sinal desejado e a saída do arranjo de microfones para o sinal desejado seja igual à potência do sinal desejado. Neste capítulo, propõe-se a modificação da restrição de forma a levar em consideração a preservação da informação da envoltória temporal da fala.

5.1 FILTRO DE REDUÇÃO DE RUÍDO MULTICANAL BASEADO NA RESTRIÇÃO DO ENVELOPE TEMPORAL NO DOMÍNIO DO TEMPO

Partindo da arquitetura de processamento do implante coclear para M microfones e K sub-bandas, definida na Figura 17, tem-se a função custo a ser minimizada, tal que:

$$J_k(n) = \mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k [\alpha \mathbf{R}_{\mathbf{x}\mathbf{x}}(n) + \mathbf{R}_{\mathbf{v}\mathbf{v}}(n)] \mathbf{D}_k^T \mathbf{W}_k \mathbf{q}_\Delta. \quad (58)$$

De forma a evitar a solução trivial para (58) propõe-se que a restrição imposta, em cada sub-banda, seja definida de forma que a envoltória do sinal de saída obtida pela transformação de Hilbert para um determinado Δ finito seja idêntica à envoltória perfeita do sinal desejado (obtida pela transformação de Hilbert para $\Delta \rightarrow \infty$), de forma que:

$$\text{env}_\Delta(z_k^{\text{fala}}(n)) = \text{env}(x_{\text{REF},k}(n)), \quad (59)$$

em que $z_k^{\text{fala}}(n)$ é a saída do arranjo de microfones assumindo a inexistência de ruído; $x_{\text{REF},k}(n)$ é o sinal desejado, correspondendo à fala captada pelo microfone de referência; e $\text{env}(\cdot)$ é o operador de cálculo da envoltória.

De acordo com (46), o envelope de $z_k^{\text{fala}}(n)$ é obtido pelo módulo do seu sinal analítico. Conforme (47) pode-se escrever:

$$z_{ak}^{\text{fala}}(n) = z_k^{\text{fala}}(n) + \tilde{j}. \quad (60)$$

A partir de (51) temos que

$$z_k^{\text{fala}}(n-\Delta) = \mathbf{q}_{\Delta}^T \mathbf{z}_k^{\text{fala}}(n) = \mathbf{q}_{\Delta}^T \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{x}}(n). \quad (61)$$

A partir de (55) obtém-se:

$$\tilde{z}_k^{\text{fala}}(n) = \mathbf{q}_{\Delta}^T \mathbf{H} \mathbf{z}_k^{\text{fala}}(n) = \mathbf{q}_{\Delta}^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{x}}(n), \quad (62)$$

em que $\mathbf{q}_{\Delta} = [0 \dots 0 \ 1 \ 0 \dots 0]^T$ possui dimensão $(\Delta+\tau) \times 1$, sendo composto por zeros, com exceção do Δ -ésimo elemento que tem valor unitário. A matriz \mathbf{W}_k possui dimensão $[MC \cdot (\Delta+\tau)] \times (\Delta+\tau)$, sendo composta por repetições de \mathbf{w}_k em sua diagonal principal; a matriz \mathbf{D}_k com dimensão $[MC(\Delta+\tau)] \times [M(N+C-1)(\Delta+\tau)]$ representa o processo de decomposição dos sinais de entrada pelo banco de filtros na k -ésima sub-banda; e o vetor $\bar{\mathbf{x}}(n)$ com dimensão $[M(N+C-1)(\Delta+\tau)] \times 1$, em que $\bar{\mathbf{x}}(n)$ é dado por:

$$\bar{\mathbf{x}}(n) = \begin{bmatrix} \mathbf{x}(n) \\ \mathbf{x}(n-1) \\ \vdots \\ \mathbf{x}(n-\Delta-\tau-1) \end{bmatrix}. \quad (63)$$

Para obter $x_{\text{REF},k}(n-\Delta)$ tem-se que:

$$x_{\text{REF},k}(n-\Delta) = \mathbf{q}_{\text{REF},k}^T \mathbf{D}_k \bar{\mathbf{x}}(n), \quad (64)$$

sendo

$$x_{\text{REF},k}(n-\Delta) = \begin{bmatrix} 0 & \dots & \dots & \mathbf{\Gamma}_k & \mathbf{0} & \dots & \mathbf{x}(n) \\ & & & \mathbf{0} & \mathbf{\Gamma}_k & \dots & \mathbf{x}(n-1) \\ & & & \vdots & \vdots & \ddots & \vdots \\ & & & \mathbf{0} & \mathbf{0} & \dots & \mathbf{x}(n-\Delta-\tau+1) \end{bmatrix}, \quad (65)$$

em que $\mathbf{q}_{\text{REF},k}$ é um vetor com dimensão $MC(\Delta+\tau) \times 1$ que é composto por zeros, no qual apenas o Δ -ésimo elemento é unitário, correspondendo à respectiva posição do microfone de referência, sendo $\text{REF} = 1, 2, \dots, M$.

Substituindo (61), (62), (64) em (46) e os resultados em (59) obtém-se:

$$\left(\mathbf{q}_{\Delta}^T \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{x}}(n) \right)^2 + \left(\mathbf{q}_{\Delta}^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{x}}(n) \right)^2 = \left(\mathbf{q}_{\text{REF},k}^T \mathbf{D}_k \bar{\mathbf{x}}(n) \right)^2 + \left(\mathbf{q}_{\text{REF},k}^T \mathbf{D}_k \bar{\mathbf{x}}(n) \right)^2, \quad (66)$$

simplificando,

$$\left(\mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{x}}(n)\right)^2 + \left(\mathbf{q}_\Delta^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{x}}(n)\right)^2 = 2\left(\mathbf{q}_{\text{REF}\Delta}^T \mathbf{D}_k \bar{\mathbf{x}}(n)\right)^2. \quad (67)$$

Visto que a fala original não é conhecida, a Equação (67) não pode ser determinada. Uma aproximação possível seria utilizar uma restrição associada ao momento de primeira ordem. Portanto, tomando-se o valor esperado em ambos os lados de (67) tem-se:

$$E\left\{\left(\mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{x}}(n)\right)^2\right\} + E\left\{\left(\mathbf{q}_\Delta^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{x}}(n)\right)^2\right\} = 2E\left\{\left(\mathbf{q}_{\text{REF}\Delta}^T \mathbf{D}_k \bar{\mathbf{x}}(n)\right)^2\right\}. \quad (68)$$

Os termos de (68) são determinados a seguir:

- Termo 1:

$$\begin{aligned} E\left\{\left(\mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{x}}(n)\right)^2\right\} &= E\left\{\mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{x}}(n) \bar{\mathbf{x}}^T(n) \mathbf{D}_k^T \mathbf{W}_k \mathbf{q}_\Delta\right\} \\ &= \mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k E\left\{\bar{\mathbf{x}}(n) \bar{\mathbf{x}}^T(n)\right\} \mathbf{D}_k^T \mathbf{W}_k \mathbf{q}_\Delta. \\ &= \mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k \mathbf{R}_{\bar{\mathbf{x}}\bar{\mathbf{x}}} \mathbf{D}_k^T \mathbf{W}_k \mathbf{q}_\Delta \end{aligned} \quad (69)$$

- Termo 2:

$$\begin{aligned} E\left\{\left(\mathbf{q}_\Delta^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{x}}(n)\right)^2\right\} &= E\left\{\mathbf{q}_\Delta^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \bar{\mathbf{x}}(n) \bar{\mathbf{x}}^T(n) \mathbf{D}_k^T \mathbf{W}_k \mathbf{H}^T \mathbf{q}_\Delta\right\} \\ &= \mathbf{q}_\Delta^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k E\left\{\bar{\mathbf{x}}(n) \bar{\mathbf{x}}^T(n)\right\} \mathbf{D}_k^T \mathbf{W}_k \mathbf{H}^T \mathbf{q}_\Delta. \\ &= \mathbf{q}_\Delta^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \mathbf{R}_{\bar{\mathbf{x}}\bar{\mathbf{x}}} \mathbf{D}_k^T \mathbf{W}_k \mathbf{H}^T \mathbf{q}_\Delta \end{aligned} \quad (70)$$

- Termo 3:

$$\begin{aligned} E\left\{\left(\mathbf{q}_{\text{REF}\Delta}^T \mathbf{D}_k \bar{\mathbf{x}}(n)\right)^2\right\} &= E\left\{\mathbf{q}_{\text{REF}\Delta}^T \mathbf{D}_k \bar{\mathbf{x}}(n) \bar{\mathbf{x}}^T(n) \mathbf{D}_k^T \mathbf{q}_{\text{REF}\Delta}\right\} \\ &= \mathbf{q}_{\text{REF}\Delta}^T \mathbf{D}_k E\left\{\bar{\mathbf{x}}(n) \bar{\mathbf{x}}^T(n)\right\} \mathbf{D}_k^T \mathbf{q}_{\text{REF}\Delta}. \\ &= \mathbf{q}_{\text{REF}\Delta}^T \mathbf{D}_k \mathbf{R}_{\bar{\mathbf{x}}\bar{\mathbf{x}}} \mathbf{D}_k^T \mathbf{q}_{\text{REF}\Delta} \end{aligned} \quad (71)$$

Substituindo (69), (70) e (71) em (68) obtém-se:

$$\begin{aligned} \mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k \mathbf{R}_{\bar{\mathbf{x}}\bar{\mathbf{x}}} \mathbf{D}_k^T \mathbf{W}_k \mathbf{q}_\Delta + \mathbf{q}_\Delta^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \mathbf{R}_{\bar{\mathbf{x}}\bar{\mathbf{x}}} \mathbf{D}_k^T \mathbf{W}_k \mathbf{H}^T \mathbf{q}_\Delta \\ = 2\mathbf{q}_{\text{REF}\Delta}^T \mathbf{D}_k \mathbf{R}_{\bar{\mathbf{x}}\bar{\mathbf{x}}} \mathbf{D}_k^T \mathbf{q}_{\text{REF}\Delta} \end{aligned} \quad (72)$$

Dessa forma, os coeficientes ótimos da proposta para o processamento do arranjo de microfones com restrição de envoltória temporal podem ser obtidos minimizando a Equação (58) sujeita à restrição descrita em (72), resultando em:

$$\begin{aligned} \mathbf{w}_k^{\text{opt}} &= \arg.\min_{\mathbf{w}_k} \mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k [\alpha \mathbf{R}_{\bar{x}\bar{x}} + \mathbf{R}_{\bar{v}\bar{v}}] \mathbf{D}_k^T \mathbf{W}_k \mathbf{q}_\Delta \\ \text{s.t. } &\mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k \mathbf{R}_{\bar{x}\bar{x}} \mathbf{D}_k^T \mathbf{W}_k \mathbf{q}_\Delta + \mathbf{q}_\Delta^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \mathbf{R}_{\bar{x}\bar{x}} \mathbf{D}_k^T \mathbf{W}_k \mathbf{H}^T \mathbf{q}_\Delta \cdot \\ &= 2 \mathbf{q}_{\text{REF}\Delta}^T \mathbf{D}_k \mathbf{R}_{\bar{x}\bar{x}} \mathbf{D}_k^T \mathbf{q}_{\text{REF}\Delta} \end{aligned} \quad (73)$$

O Quadro 3 apresenta o pseudocódigo para obter a solução do processo de otimização apresentado em (73).

Quadro 3 – Pseudocódigo para a solução com restrição da envoltória temporal Equação (73).

<p>Entradas: $\alpha, \mathbf{R}_{\bar{x}\bar{x}}, \mathbf{R}_{\bar{v}\bar{v}}, \Gamma_k, \mathbf{w}_k, \mathbf{q}_\Delta, \mathbf{q}_{\text{REF}\Delta}$</p> <p>Saídas: $\mathbf{w}_k^{\text{opt}}$</p>
<ul style="list-style-type: none"> • Calcule a matriz \mathbf{W}_k de acordo com (50) • Calcule a matriz \mathbf{D}_k de acordo com (50) • Calcule $\mathbf{R}_{\bar{x}\bar{x}}$ utilizando $\mathbf{R}_{\bar{x}\bar{x}}$ • Calcule $\mathbf{R}_{\bar{v}\bar{v}}$ utilizando $\mathbf{R}_{\bar{v}\bar{v}}$ • Calcule \mathbf{H} de acordo com (53) • Calcule a restrição como em (72) • Obtenha (73) para as k sub-bandas
<p>Retorna $\mathbf{w}_k^{\text{opt}}$</p>

5.2 ARRANJO DE MICROFONES COM RESTRIÇÃO ASSOCIADA À POTÊNCIA DO SINAL DESEJADO

A determinação de uma solução fechada para o problema de otimização descrito em (73) não é trivial, o que resulta na necessidade de utilização de algoritmos numéricos para a sua solução. Uma simplificação possível do problema descrito em

(72) seria o caso em que $\Delta=0$. Analisando-se individualmente os três termos que compõem (72), Equações (69)-(71) tem-se que:

- Termo 1:

$$\mathbf{q}_\Delta^T \mathbf{W}_k^T \mathbf{D}_k \mathbf{R}_{\bar{x}\bar{x}} \mathbf{D}_k^T \mathbf{W}_k \mathbf{q}_\Delta \xrightarrow{\Delta=0} \mathbf{w}_k^T \mathbf{\Gamma}_k \mathbf{R}_{xx} \mathbf{\Gamma}_k^T \mathbf{w}_k, \quad (74)$$

sendo que \mathbf{q}_Δ se torna um escalar unitário, por isso não aparece na Equação (74). A matriz \mathbf{W}_k , que possui dimensão $MC(\Delta+\tau) \times (\Delta+\tau)$, se torna o vetor \mathbf{w}_k de dimensão $MC \times 1$, uma vez que $\tau=2\Delta+1$. A matriz \mathbf{D}_k , com dimensão $MC(\Delta+\tau) \times M(N+C-1)(\Delta+\tau)$, se torna uma matriz menor conhecida como $\mathbf{\Gamma}_k$ com dimensão $MC \times M(N+C-1)$. Por fim, a matriz $\mathbf{R}_{\bar{x}\bar{x}}$, com dimensão $M(N+C-1)(\Delta_1+\tau_1) \times M(N+C-1)(\Delta_1+\tau_1)$, se torna a matriz \mathbf{R}_{xx} com dimensão $M(N+C-1) \times M(N+C-1)$.

- Termo 2:

$$\mathbf{q}_\Delta^T \mathbf{H} \mathbf{W}_k^T \mathbf{D}_k \mathbf{R}_{\bar{x}\bar{x}} \mathbf{D}_k^T \mathbf{W}_k \mathbf{H}^T \mathbf{q}_\Delta \xrightarrow{\Delta=0} 0. \quad (75)$$

Neste caso, a matriz de transformação de Hilbert, \mathbf{H} , representada em (53), que possui dimensão $(\Delta+\tau) \times (\Delta+\tau)$ se torna um escalar com valor nulo para $\Delta=0$.

- Termo 3:

$$2\mathbf{q}_{\text{REFA}}^T \mathbf{D}_k \mathbf{R}_{\bar{x}\bar{x}} \mathbf{D}_k^T \mathbf{q}_{\text{REFA}} \xrightarrow{\Delta=0} 2\mathbf{q}_{\text{REF}}^T \mathbf{\Gamma}_k \mathbf{R}_{xx} \mathbf{\Gamma}_k^T \mathbf{q}_{\text{REF}}, \quad (76)$$

em que \mathbf{q}_{REFA} , com dimensão $MC \cdot (\Delta+\tau) \times 1$ se torna \mathbf{q}_{REF} , com dimensão $MC \times 1$, para $\Delta=0$.

Por fim, substituindo (74), (75) e (76) em (72) chega-se em:

$$\mathbf{w}_k^T \mathbf{\Gamma}_k \mathbf{R}_{xx} \mathbf{\Gamma}_k^T \mathbf{w}_k + 0 \cdot \mathbf{w}_k^T \mathbf{\Gamma}_k \mathbf{R}_{xx} \mathbf{\Gamma}_k^T \mathbf{w}_k = 2\mathbf{q}_{\text{REF}}^T \mathbf{\Gamma}_k \mathbf{R}_{xx} \mathbf{\Gamma}_k^T \mathbf{q}_{\text{REF}}. \quad (77)$$

Com as devidas simplificações:

$$\mathbf{w}_k^T \mathbf{\Gamma}_k \mathbf{R}_{xx} \mathbf{\Gamma}_k^T \mathbf{w}_k = 2\mathbf{q}_{\text{REF}}^T \mathbf{\Gamma}_k \mathbf{R}_{xx} \mathbf{\Gamma}_k^T \mathbf{q}_{\text{REF}}. \quad (78)$$

Substituindo-se (78) em (73), o problema de minimização do ruído reduz-se

a:

$$\begin{aligned} \mathbf{w}_k^{\text{opt}} &= \arg.\min_{\mathbf{w}_k} \alpha \mathbf{w}_k^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{w}_k + \mathbf{w}_k^T \Gamma_k \mathbf{R}_{vv} \Gamma_k^T \mathbf{w}_k \\ \text{s.t. } \mathbf{w}_k^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{w}_k &= 2\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}} \end{aligned} \quad (79)$$

Nota-se que em (79) a restrição resultante estabelece que a potência na saída em cada sub-banda do sistema de redução de ruído deve ser igual à potência da fala de interesse. Surpreendentemente, uma extensa revisão da literatura indica que esse tipo de abordagem, preservação da potência do sinal de fala, ainda não foi utilizado na área de arranjos de microfones em aplicações de IC (Frost, 1972; Gannot *et al.*, 2017; Ward; Kennedy; Williamson, 1995; Bourgeois; Minker, 2009; Markovich; Gannot; Cohen, 2009; Buckley, 1987).

5.3 SOLUÇÃO ANALÍTICA SEMI-FECHADA PARA OTIMIZAÇÃO DO ARRANJO DE MICROFONES COM RESTRIÇÃO DE POTÊNCIA

Nesta seção é apresentada uma solução analítica semi-fechada para o problema de otimização descrito em (79).

Apesar do uso de algoritmos numéricos permitir a obtenção de uma solução para (79), há a necessidade de um esforço computacional elevado, pela existência de uma restrição de igualdade e a possibilidade de muitos graus de liberdade em função do número de coeficientes do filtro \mathbf{w}_k . Para evitar esse tipo de problema, pode-se recorrer à conversão de (79) em um problema de autovalores generalizados, uma vez que a função custo e a restrição são homogêneas. Dessa forma, a resolução do problema é simplificada e, conseqüentemente, o custo computacional necessário (Golub; Van Loan, 2013).

Para resolver o problema de otimização utilizando autovalores generalizados é necessário reescrevê-lo. Assim, para simplificar o desenvolvimento matemático vamos considerar:

$$\mathbf{A} = \alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T + \Gamma_k \mathbf{R}_{vv} \Gamma_k^T, \quad (80)$$

e,

$$\mathbf{B} = \frac{1}{2\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}}} \Gamma_k \mathbf{R}_{xx} \Gamma_k^T. \quad (81)$$

Logo:

$$\begin{aligned} \mathbf{w}_k^{\text{opt}} &= \arg.\min_{\mathbf{w}_k} \mathbf{w}_k^T \mathbf{A} \mathbf{w}_k \\ \text{s.t. } &\mathbf{w}_k^T \mathbf{B} \mathbf{w}_k = 1 \end{aligned} \quad (82)$$

Em 2019, Ghojogh, Karray e Crowley descrevem algumas alternativas para a resolução de problemas com autovalores e autovetores generalizados (Ghojogh; Karray; Crowley, 2019). Um caminho possível é a utilização do Lagrangeano, tal que:

$$L(\mathbf{w}_k, \lambda) = \mathbf{w}_k^T \mathbf{A} \mathbf{w}_k - \lambda (\mathbf{w}_k^T \mathbf{B} \mathbf{w}_k - 1), \quad (83)$$

em que $\lambda \in \mathbb{R}$ é o multiplicador de Lagrange.

A derivada do Lagrangeano em relação a \mathbf{w}_k é dada por:

$$\frac{\partial L}{\partial \mathbf{w}_k} = \frac{\partial L}{\partial \mathbf{w}_k} \mathbf{w}_k^T \mathbf{A} \mathbf{w}_k - \lambda \frac{\partial L}{\partial \mathbf{w}_k} (\mathbf{w}_k^T \mathbf{B} \mathbf{w}_k - 1). \quad (84)$$

Analisando-se as propriedades das derivadas tem-se que:

$$\frac{\partial \mathbf{x}^T \mathbf{C} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{C} + \mathbf{C}^T) \mathbf{x}. \quad (85)$$

neste caso o vetor \mathbf{x} e a matriz \mathbf{C} apenas auxiliam na exemplificação da solução. Se \mathbf{C} é simétrica, tem-se:

$$\frac{\partial \mathbf{x}^T \mathbf{C} \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{C} \mathbf{x}. \quad (86)$$

Uma vez que \mathbf{A} e \mathbf{B} são matrizes simétricas, pode-se aplicar as propriedades apresentadas em (85) e (86) em (84) e obter:

$$\frac{\partial L}{\partial \mathbf{w}_k} = 2\mathbf{A} \mathbf{w}_k - 2\lambda \mathbf{B} \mathbf{w}_k. \quad (87)$$

Igualando-se a derivada do Lagrangeano (87) a um vetor de zeros, tem-se que:

$$2\mathbf{A} \mathbf{w}_k - 2\lambda \mathbf{B} \mathbf{w}_k = \mathbf{0} \quad (88)$$

O problema em (88) pode ser representado como:

$$\mathbf{A} \mathbf{w}_k - \lambda \mathbf{B} \mathbf{w}_k = \mathbf{0}, \quad (89)$$

logo,

$$\mathbf{A} \mathbf{w}_k = \lambda \mathbf{B} \mathbf{w}_k. \quad (90)$$

Substituindo \mathbf{A} e \mathbf{B} em (90), tem-se:

$$\left(\alpha \Gamma_k \mathbf{R}_{xx} \Gamma_k^T + \Gamma_k \mathbf{R}_{vv} \Gamma_k^T \right) \mathbf{w}_k = \lambda \left(\frac{1}{2 \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \mathbf{q}_{\text{REF}}} \Gamma_k \mathbf{R}_{xx} \Gamma_k^T \right) \mathbf{w}_k. \quad (91)$$

O problema de minimização apresentado em (79) é equivalente ao problema de autovalores generalizado em (91). Desta maneira, a solução ótima $\mathbf{w}_k^{\text{opt}}$ do problema de minimização (79) é um autovetor de (91), e o multiplicador de Lagrange λ será o autovalor correspondente. Portanto, ao resolver (91), encontra-se a solução ótima do problema. Deve-se considerar que o autovetor desejado é aquele associado ao menor autovalor, pois trata-se de um problema de minimização.

5.3.1 Problema de autovalores generalizados

Os problemas com autovalores e autovetores generalizados desempenham um papel crucial em vários ramos da ciência, abrangendo áreas como física, estatística, matemática e até mesmo na aprendizagem de máquina. Muitos problemas de otimização encontram solução na generalização (Ghojogh; Karray; Crowley, 2019). Neste caso, a resolução pode ser obtida pela generalização do autovalor, que é reescrito na forma matricial para a resolução do problema apresentado em (90).

O problema de autovalor generalizado (Golub; Van Loan, 2012) para duas matrizes simétricas $\mathbf{A} \in \mathbb{R}^{d \times d}$ e $\mathbf{B} \in \mathbb{R}^{d \times d}$ é definido como:

$$\mathbf{A} \mathbf{w}_{k,i} = \lambda_i \mathbf{B} \mathbf{w}_{k,i}, \forall_i \in \{1, \dots, d\}, \quad (92)$$

em que i varia de 1 a d , sendo d a posição do autovetor.

Na forma matricial (Wang, 2015):

$$\mathbf{A} \Phi_k = \mathbf{B} \Phi_k \Lambda, \quad (93)$$

em que as colunas de $\mathbb{R}^{d \times d} \ni \Phi_k = [\mathbf{w}_{k,1}, \dots, \mathbf{w}_{k,d}]$ são os autovetores e os elementos da diagonal de $\mathbb{R}^{d \times d} \ni \Lambda = \text{diag}([\lambda_1, \dots, \lambda_d]^T)$ são os autovalores. Note que $\mathbf{w}_{k,d} \in \mathbb{R}^d$ e $\lambda_d \in \mathbb{R}$.

Uma solução pode ser obtida ao considerar a Equação (93). Existe um método que resolve o problema de autovalores generalizados proposto por Wang (2015) e adaptado para o problema em questão. Considerando o problema de autovalor para \mathbf{B} :

$$\mathbf{B}\Phi_{k,B} = \Phi_{k,B}\Lambda_B, \quad (94)$$

em que $\Phi_{k,B}$ e Λ_B são autovetores e autovalores matriciais de \mathbf{B} , respectivamente, então:

$$\begin{aligned} \mathbf{B}\Phi_{k,B} = \Phi_{k,B}\Lambda_B &\Rightarrow \Phi_{k,B}^{-1}\mathbf{B}\Phi_{k,B} = \underbrace{\Phi_{k,B}^{-1}\Phi_{k,B}}_{\mathbf{I}} \Lambda_B = \Lambda_B \\ &\stackrel{(a)}{\Rightarrow} \Phi_{k,B}^T \mathbf{B}\Phi_{k,B} = \Lambda_B \end{aligned}, \quad (95)$$

em que (a) decorre do fato de que $\Phi_{k,B}$ é uma matriz ortogonal (e suas colunas são ortonormais), assim $\Phi_{k,B}^{-1} = \Phi_{k,B}^T$. Multiplicando $\Lambda_B^{-1/2}$ em ambos os lados da Equação (95), chega-se em:

$$\begin{aligned} \Lambda_B^{-1/2} \Phi_{k,B}^T \mathbf{B}\Phi_{k,B} \Lambda_B^{-1/2} - \Lambda_B^{-1/2} \Lambda_B \Lambda_B^{-1/2} &= \mathbf{I} \\ \Rightarrow \check{\Phi}_{k,B} \check{\Phi}_{k,B} - \mathbf{I} &= \mathbf{0} \end{aligned}, \quad (96)$$

em que,

$$\check{\Phi}_{k,B} = \Phi_{k,B} \Lambda_B^{-1/2}. \quad (97)$$

A matriz $\Lambda_B^{-1/2}$, na Equação (97), é singular. Assim pode-se utilizar uma aproximação para $\check{\Phi}_{k,B} \approx \Phi_{k,B}(\Lambda_B^{-1/2} + \varepsilon \mathbf{I})^{-1}$, em que ε é um número pequeno e positivo que é o valor necessário para fazer com que $\Lambda_B^{-1/2}$ tenha o rank completo.

Desse modo, pode-se definir $\check{\check{\mathbf{A}}}$ como:

$$\check{\check{\mathbf{A}}} = \check{\check{\Phi}}_{k,B} \check{\check{\Phi}}_{k,B}, \quad (98)$$

que pode ser escrito como:

$$\check{\check{\mathbf{A}}} = (\check{\check{\Phi}}_{k,B} \check{\check{\Phi}}_{k,B}) + \check{\check{\Phi}}_{k,B} \check{\check{\Phi}}_{k,B} - \check{\check{\mathbf{A}}}, \quad (99)$$

em que (a) denota simetria.

O problema de autovalor para $\check{\check{\mathbf{A}}}$ é:

$$\tilde{\mathbf{A}} \tilde{\Phi}_{k,A} = \Phi_{k,A} \Lambda_A, \quad (100)$$

em que $\Phi_{k,A}$ e Λ_A são os autovetores e autovalores da matriz $\tilde{\mathbf{A}}$, respectivamente. Multiplicando o lado esquerdo da Equação (100) por $\Phi_{k,A}^{-1}$:

$$\Phi_{k,A}^{-1} \tilde{\mathbf{A}} \tilde{\Phi}_{k,A} = \underbrace{\Phi_{k,A}^{-1} \Phi_{k,A}}_{\mathbf{I}} \Lambda_A = \Lambda_A, \quad (101)$$

em que (a) decorre do fato de que $\Phi_{k,A}$ é uma matriz ortogonal (cujas colunas são ortonormais) e, portanto, $\Phi_{k,A}^{-1} = \Phi_{k,A}^T$. A matriz $\Phi_{k,A}$ é ortogonal e, portanto, $\tilde{\mathbf{A}}$ é simétrica, uma vez que os autovetores são ortogonais e ortonormais. A Equação (101) diagonaliza a matriz $\tilde{\mathbf{A}}$ (Wang, 2015).

Pode-se conectar a Equação (98) à Equação (101) de tal forma que:

$$\begin{aligned} \Phi_{k,A}^T \tilde{\Phi}_{k,B} \tilde{\mathbf{A}} \tilde{\Phi}_{k,B} \Phi_{k,A} &= \Lambda_A \\ \Rightarrow \Phi_{k,A}^T \Lambda_{k,B}^{-1/2} \Phi_{k,B}^T \mathbf{A} \Phi_{k,B} \Lambda_{k,B}^{-1/2} \Phi_{k,A} &= \Lambda_A, \\ \Rightarrow \Phi_k^T \mathbf{A} \Phi_k &= \Lambda_A \end{aligned} \quad (102)$$

sendo,

$$\Phi_k = \tilde{\Phi}_{k,B} \Phi_{k,A} = \Phi_{k,B} \Lambda_{k,B}^{-1/2} \Phi_{k,A}, \quad (103)$$

em que Φ_k também diagonaliza \mathbf{B} porque a matriz \mathbf{I} é diagonal:

$$\begin{aligned} \Phi_k^T \mathbf{B} \Phi_k &= (\Phi_{k,B} \Lambda_{k,B}^{-1/2} \Phi_{k,A})^T \mathbf{B} (\Phi_{k,B} \Lambda_{k,B}^{-1/2} \Phi_{k,A}) \\ &= \Phi_{k,A}^T \Lambda_{k,B}^{-1/2} (\Phi_{k,B}^T \mathbf{B} \Phi_{k,B}) \Lambda_{k,B}^{-1/2} \Phi_{k,A} \\ &= \Phi_{k,A}^T \underbrace{\Lambda^{-1/2} \mathbf{A} \Lambda^{-1/2}}_{\mathbf{I}} \Phi_{k,A} = \Phi_{k,A}^T \Phi_{k,A} \\ &\stackrel{(a)}{=} \Phi_{k,A}^{-1} \Phi_{k,A} = \mathbf{I} \end{aligned} \quad (104)$$

O termo (a) sobre a igualdade é devido à condição de ortogonalidade da matriz $\Phi_{k,A}$. Das Equações (103) e (104), tem-se:

$$\begin{aligned} \Phi_k^T \mathbf{B} \Phi_k = \mathbf{I} &\Rightarrow \Phi_k^T \mathbf{B} \Phi_k \Lambda_A = \Lambda_A \\ &\Rightarrow \Phi_k^T \mathbf{B} \Phi_k \Lambda_A = \Phi_k^T \mathbf{A} \Phi_k, \\ &\stackrel{(a)}{\Rightarrow} \mathbf{B} \Phi_k \Lambda_A = \mathbf{A} \Phi_k \end{aligned} \quad (105)$$

em que (a) decorre de $\Phi_k \neq \mathbf{0}$ (Golub; Van Loan, 2012).

Comparando as Equações (93) e (105), observa-se que:

$$\Lambda_A = \Lambda, \quad (106)$$

sendo uma matriz diagonal Λ composta pelos autovalores:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & & \\ 0 & \lambda_2 & \cdots & & \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & & \end{bmatrix} \quad (107)$$

O menor autovalor λ_i apresentado na i coluna, indica o autovetor $w_{k,i}$ que será utilizado. A matriz Φ_k corresponde a todos os autovetores, porém apenas a coluna na posição i será utilizada no processo de minimização, então:

$$\Phi_k = \begin{bmatrix} w_{k,1,1} & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ w_{k,1,MC} & \cdots & c \end{bmatrix}, \quad (108)$$

e, neste caso w_k , tem dimensão $MC \times 1$ e é proveniente de uma das colunas i de Φ_k .

Um pseudocódigo para este procedimento é detalhado no Quadro 4.

Quadro 4 – Pseudocódigo para a solução semi-fechada.

Entradas: $\alpha, R_{xx}, R_{vv}, \Gamma_k$

Saídas: w_k

- Calcule \mathbf{A} de acordo com (80)
- Calcule \mathbf{B} de acordo com (81)
- Determine $\Phi_{k,B}$ e Λ_B (94)
- Calcule $\tilde{\Lambda}$ de acordo com (98)
- Determine $\Phi_{\tilde{\Lambda}}$ e $\Lambda_{\tilde{\Lambda}}$ de acordo com (100)
- Calcule Φ_k de acordo com (103)
- Calcule Λ_A de acordo com a Equação (105)
- Determine o índice “ i ” do menor Λ_A
- $w_k \leftarrow$ coluna “ i ” de Φ_k

Retorna w_k

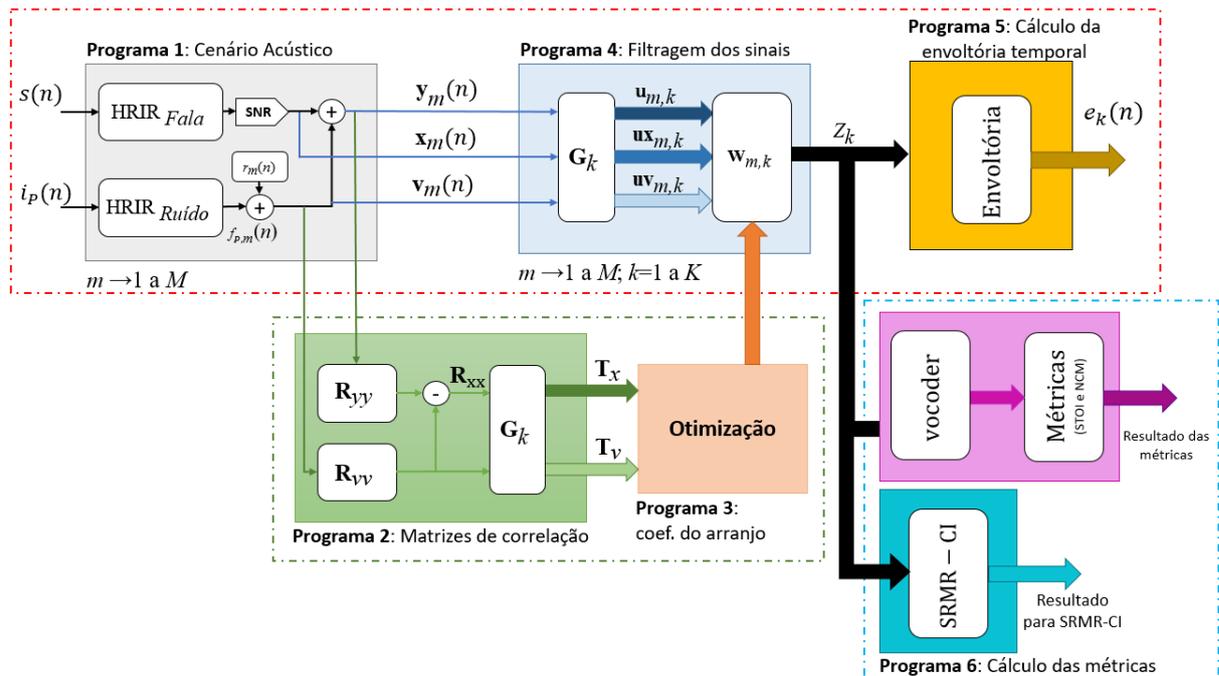
6 MATERIAIS E MÉTODOS

Neste capítulo, são descritos os materiais e métodos empregados neste estudo. Apresentam-se os softwares utilizados e as características do cenário acústico, além de detalhar a estimativa das matrizes de correlação. Também são descritos os procedimentos para a realização dos experimentos numéricos e psicoacústicos, com o objetivo de fornecer informações suficientes para a replicação da pesquisa.

O processamento e análise de dados foi feito utilizando o software MatLab e um computador Dell G15 com processador AMD Ryzen 5600H com Radeon Graphics, 3301 Mhz, 6 núcleos e 12 processadores lógicos com placa de vídeo GeForce RTX 4060.

A Figura 25 apresenta um diagrama em blocos da arquitetura utilizada para a implementação dos métodos sob comparação, enquanto que o Quadro 5 descreve sucintamente cada um dos blocos contidos nela.

Figura 24 – Esquemático dos softwares desenvolvidos.



Fonte: Autoria própria.

Quadro 5 - Lista geral de softwares utilizados.

Programa	Etapa	Entrada	Saída
1	Cenário Acústico	fala, ruído interferente e de fundo, HRIR _x , HRIR _v , SNR	$\mathbf{y}_m(n)$, $\mathbf{x}_m(n)$ e $\mathbf{v}_m(n)$
2	Cálculo das matrizes de correlação	$\mathbf{y}_m(n)$ e $\mathbf{v}_m(n)$	\mathbf{R}_{xx} , \mathbf{R}_{vv} e \mathbf{R}_{yy}
3	Cálculo dos coeficientes do arranjo de microfones	filtragem de \mathbf{T}_x e \mathbf{T}_v	\mathbf{w}_k
4	Filtragem dos sinais	$\mathbf{y}_m(n)$, $\mathbf{x}_m(n)$ e $\mathbf{v}_m(n)$, \mathbf{G}_k , \mathbf{w}_k	z_k
5	Cálculo da envoltória temporal dos sinais	z_k	e_k
6	Métricas	NCM, STOI, SRMR-CI e SINR	z_k e resultado das métricas objetivas

Fonte: Autoria própria.

6.1 SINAIS E SISTEMAS

O ruído interferente escolhido é do tipo ICRA-1 (*Artificial Speech-Like Noise*), introduzido por Dreschler *et al.* (2001) e desenvolvido pelo *International Collegium for Rehabilitative Audiology* como um conjunto de sinais de ruído artificial para testes audiológicos inspirados em experimentos psicofísicos (Dreschler *et al.*, 2001). O ruído ICRA-1 é amplamente utilizado na literatura, como nos trabalhos apresentados por Wouters e Berghie (2001), Foo *et al.* (2007), entre outros. Quanto ao ruído de fundo, é empregado um ruído branco com potência ajustada de acordo com a razão sinal ruído desejada de entrada. A frequência de amostragem é de 16 kHz.

Para a verificação da funcionalidade do algoritmo descrito no Quadro 4 foram utilizados áudios em inglês do banco de dados IEEE *corpus* (IEEE, 1969), enquanto que para as comparações entre os diversos métodos de redução de ruído foram utilizados 30 áudios de fala humana foneticamente balanceados em português do Brasil provenientes do banco de dados Alcaim (Alcaim; Solewicz; Moraes, 1992).

Para os experimentos, considerou-se um total de $M = 6$ microfones e o banco de filtros do tipo gammatone com 22 sub-bandas descrito em Glasberg e Moore (1990). A Figura 25 apresenta as respectivas respostas em magnitude das 22 sub-bandas. Os filtros utilizados são do tipo resposta infinita ao impulso (IIR) (Patterson *et al.*, 1992). Portanto, para que suas características possam ser incluídas no modelo utilizado (que considera filtros com resposta finita ao impulso) é necessário truncar as respostas ao impulso após decaimento substancial. Neste trabalho, definiu-se o ponto de truncamento na amostra 200 (indicado pela linha vertical tracejada em vermelho na Figura 26), o que equivale à manutenção de mais de 99% da energia total da resposta ao impulso original.

Figura 25 – Resposta de magnitude do banco de filtros tipo gammatone de 22 sub-bandas.

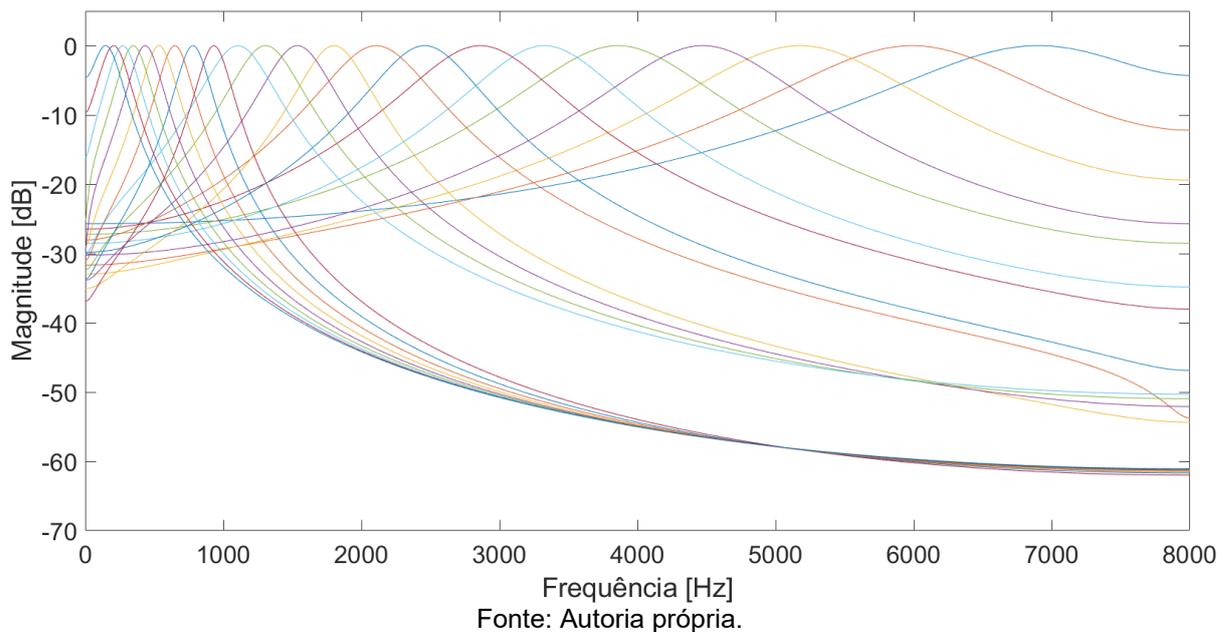
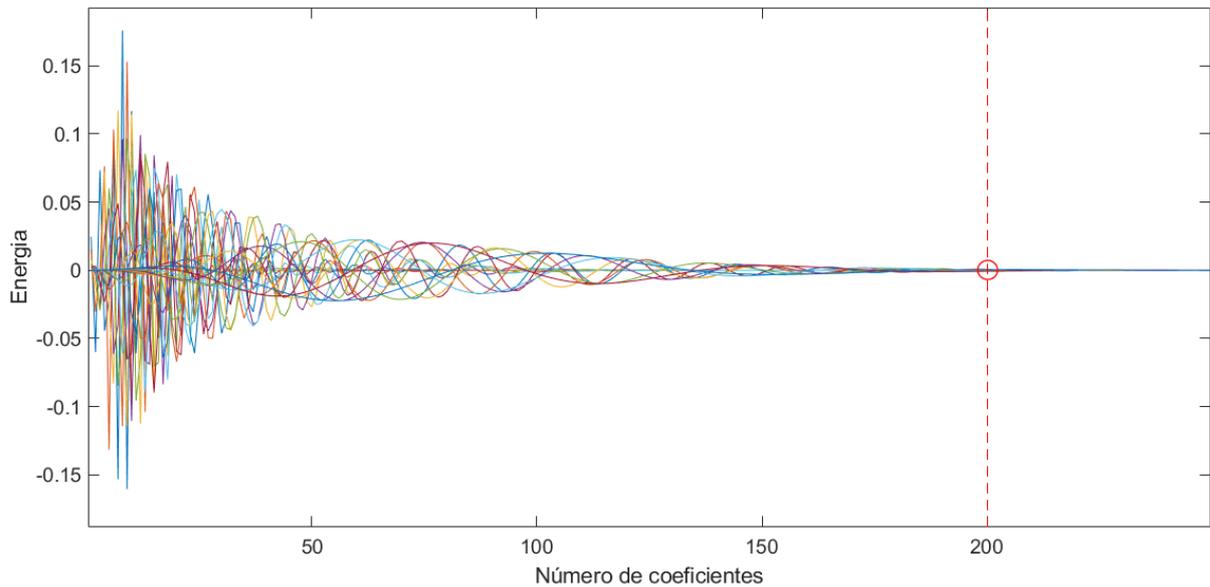


Figura 26 – Truncamento da resposta ao impulso para as 22 sub-bandas do filtro gammatone.



Fonte: Autoria própria.

6.2 CENÁRIO ACÚSTICO

O cenário acústico considera a propagação dos sinais em uma sala de escritório com tempo de reverberação (R_T) de 300 ms. As simulações foram realizadas com respostas ao impulso do tipo HRIR obtidas a partir de um banco de dados (ambiente 'Office I' em Kayser *et al.*, 2009) no qual foi utilizado um manequim (simulador de torso e cabeça - Bruel & Kjaer type 4128-C) com um aparelho auditivo de três microfones em cada orelha.

Arbitrou-se a fonte de fala no azimute de 0° (em frente à linha de visada do manequim) a uma distância de 3 metros e a fonte de ruído em -60° (lado esquerdo) a uma distância de 3 metros.

Os parâmetros de simulação utilizados foram: $C = 1$, $N = 200$, $\Delta = \{0, 40\}$, SNR (*Signal-to-Noise Ratio*) com valor de 30 dB, e SIR (*Interference Signal Ratio*) entre -5 dB e 5 dB, variando em passos de 5 dB.

6.3 ESTIMAÇÃO DAS MATRIZES DE CORRELAÇÃO

A implementação prática do arranjo de microfones baseado no MVDR emprega estatísticas de segunda ordem em vez de valores instantâneos (Gershman, 2013). Apesar da indisponibilidade de acesso individual às amostras de fala e ruído, as matrizes de correlação desses sinais geralmente podem ser estimadas com um detector de fala (*Voice Activity Detector - VAD*) (Loizou, 2013). O VAD permite a discriminação entre fala contaminada e períodos somente com ruído. Considerando que o ruído é aditivo e assumindo que a fala e o ruído são aproximadamente estacionários dentro de uma determinada janela de tempo, estimativas da matriz de correlação de fala podem ser obtidas por $E\{\mathbf{x}(n)\mathbf{x}^T(n)\} = \mathbf{R}_{xx}(n) = \mathbf{R}_{yy}(n) - \mathbf{R}_{vv}(n)$; em que $\mathbf{R}_{yy}(n) = E\{\mathbf{y}(n)\mathbf{y}^T(n)\}$ e $\mathbf{R}_{vv}(n) = E\{\mathbf{v}(n)\mathbf{v}^T(n)\}$ são calculados diretamente a partir do sinal captado sinalizado pelo VAD.

De forma a minimizar erros decorrentes da estimação das matrizes de correlação, as simulações foram realizadas utilizando o método “*Batch*”, que significa que as matrizes de correlação de fala contaminada e somente ruído são calculadas *a-priori*, utilizando os respectivos sinais de forma isolada e completa. Na sequência, o sinal é processado através de um filtro ótimo com coeficientes constantes para cada faixa de frequência (bin k). Os resultados obtidos podem ser considerados como o limite superior de desempenho da técnica. Essa estratégia é amplamente utilizada na literatura, como por exemplo em Marquart, Hohman e Doclo (2013), Cornelis, Moonen e Wouters (2011) e Zhang e Li (2021).

6.4 MEDIDAS OBJETIVAS DE DESEMPENHO

O critério de desempenho denominado de razão sinal ruído mais interferência (*Signal to Noise Plus Interference Ratio – SINR*) de saída (Zhang *et al.*, 2015), em dB, é obtido por meio do cálculo da potência dos sinais, de forma que:

$$\text{SINR}_{dB}^{\text{out}} = 10 \log_{10} \left[\frac{\frac{1}{N} E \left\{ \left| \sum_{k=1}^{k=22} \mathbf{y}^T(n) \mathbf{\Gamma}_k^T \mathbf{w}_k^{\text{opt}} \right|^2 \right\}}{\frac{1}{N} E \left\{ \left| \sum_{k=1}^{k=22} \mathbf{v}^T(n) \mathbf{\Gamma}_k^T \mathbf{w}_k^{\text{opt}} \right|^2 \right\}} \right] \quad (109)$$

em que N é o comprimento dos sinais.

A inteligibilidade é estimada através de três critérios: a razão de energia de modulação entre fala e reverberação para implantes cocleares (*Speech to Reverberation Modulation Energy Ratio for Cochlear Implants* – SRMR-CI) (Santos e Falk, 2014) (Apêndice C), a medida de inteligibilidade objetiva de tempo curto (*Short Term Objective Intelligibility* - STOI) (Taal *et al.*, 2011) (Apêndice C) e a medida de covariância normalizada (*Normalized Covariance Measure* - NCM) (Apêndice C).

Para o cálculo das métricas SINR, STOI e NCM, os sinais das sub-bandas de saída foram previamente submetidos a um vocoder tonal (Loizou, 2006) para a reconstrução de um sinal em banda larga. No caso da SRMR-CI, os sinais das 22 sub-bandas foram utilizados em substituição ao banco de filtros de entrada da métrica.

6.5 EXPERIMENTOS PSICOACÚSTICOS

Nesta subseção são descritos os procedimentos relacionados aos testes psicoacústicos conduzidos para avaliar a inteligibilidade em indivíduos normouvintes³ e implantados cocleares.

6.5.1 Experimento psicoacústico com normouvintes

Experimentos psicoacústicos foram realizados com 10 indivíduos com idade média de 31 ± 2 anos sem queixas prévias quanto a limitações auditivas. Trinta (30) frases foram processadas pelo sistema descrito na Figura 25 e então sinais de banda larga foram sintetizadas usando um vocoder tonal (Loizou, 2006). Essas frases foram apresentadas uma a uma para os normouvintes. Após ouvirem a sentença, os voluntários deveriam repeti-la de acordo com o que conseguiam compreender. Dessa forma, foi possível contabilizar a inteligibilidade obtida para cada SIR e o percentual de acertos.

6.5.2 Experimento psicoacústico com implantados cocleares

Para a realização dos experimentos com implantados foi utilizada uma interface gráfica aprimorada a partir do trabalho desenvolvido por Chiea em 2019. A

³ São pessoas com audição considerada normal, que não apresentam perdas auditivas significativas.

Figura 27 apresenta a nova interface, dividida em partes. A primeira delas é o checklist, que tem como objetivo garantir que nenhuma das etapas essenciais para a adequada execução do teste seja omitida. Assim, as etapas incluem explicar o teste ao paciente, apresentar o Termo de Consentimento Livre e Esclarecido (TCLE) (ANEXO A – TCLE), preencher e carregar o mapa do implantado (*map*), ajustar o volume para o IC e, por fim, verificar se as demais configurações foram definidas adequadamente.

O *map* inclui informações como o tipo de implante, a frequência de amostragem, o número de sub-bandas, estratégia de estimulação, configuração dos eletrodos, largura de pulso, volume do IC e sensibilidade. Para cada uma das K sub-bandas são definidas as faixas de frequência para cada eletrodo. Neste trabalho, essas informações foram fornecidas por fonoaudiólogas do Hospital Universitário Polydoro Ernani de São Thiago em Florianópolis-SC. Com esses dados, o pesquisador preenche um arquivo para carregar as informações do IC no software da interface, garantindo que o paciente seja capaz de ouvir adequadamente tudo o que lhe será apresentado. O voluntário sob teste não tem acesso visual à interface gráfica durante o procedimento, sendo utilizado apenas pelo pesquisador.

A cada iteração do procedimento, uma frase previamente gravada é apresentada ao voluntário, que é instruído a repetir em voz alta o que entendeu. Cada uma das frases é reproduzida uma única vez. As palavras corretas são marcadas no bloco *sentence* e registradas no bloco *verify*. Ao fim do teste os resultados obtidos são gravados e enviados para uma pasta de resultados.

Figura 27 – Interface gráfica para a realização do experimento psicoacústico com IC.

The interface is organized into several functional areas:

- Check-list:** Includes a 'Name' field (containing 'teste') and a list of checkboxes: 'Explain the text', 'Term of consent', 'Load Map', 'Adjust Volume', and 'Verify Scenario Configuration'. An 'OK' button is at the bottom.
- Parameters:** Contains 'Processing Type' (Real Time, Off-line), 'Noise Type' (Babble, Wind, Car, White), 'Gain' (a dial from 0 to 50), and 'Noise reduction technique' (C1, RW, C2, RTC2F). It also has 'Manual' and 'Stimulation Buffer' (Board, Proposed) options.
- Files characteristics:** Fields for 'ID', '# Computed audios', 'Cycle', 'ATR', 'Nchan', 'Técnica', and 'SNR'.
- State:** A status indicator with a green circle and text: 'Green: stand-by', 'Blue: Processing', and 'Red: Playing audio'.
- Sentence:** A text input field and a set of letter buttons (C, I, M, O, B, L, E, F).
- COM:** A dropdown menu and an 'Update' button.
- Test Panel:** Buttons for 'Repeat', 'Test Audio', 'Count', and 'Exit'.
- Verify:** Buttons for 'All Correct', 'Count', and 'Next', along with input fields for 'Total words', 'Correct words count', and 'Percentage'.
- Speak with the subject:** A large 'Speak' button and a smaller 'Repeat' button.

Footer text: Partnership: UFSC and UT-Dallas; Research-board: CCI-Mobile; Researcher crew: LPDS, LVA and UFSC fonoaudiology.

Fonte: Autoria própria.

Os experimentos psicoacústicos com implantados contaram com três voluntários. A primeira paciente foi uma mulher de 65 anos com IC unilateral na orelha direita e que fazia uso do dispositivo há 7 anos. O segundo paciente foi um homem de 55 anos com IC unilateral implantado na orelha direita e que fazia uso do IC há 8 anos. A terceira paciente foi uma mulher de 51 anos com IC alocado na orelha esquerda. A etiologia da perda auditiva dos três voluntários é desconhecida.

Os implantes utilizados pelos três pacientes são do modelo Nucleus CI422, com 22 sub-bandas ativas e frequência de estimulação de 1200 pulsos por segundo. O experimento consistiu na apresentação de 30 frases com SIR de -5 dB, 0 dB e 5 dB. Para essa avaliação, foi utilizada uma interface de pesquisa voltada para implantes cocleares conhecida como CCI *mobile*. Este sistema foi desenvolvido por pesquisadores da UTD e permite a comunicação direta com a parte interna do implante através de uma antena RF (Ali, 2016).

7 RESULTADOS

Neste capítulo são apresentados os resultados das simulações e experimentos psicoacústicos descritos no Capítulo 6.

Na sequência são apresentados os resultados obtidos a partir da aplicação de técnicas de redução de ruído multicanal baseadas em arranjo de microfones. O objetivo é analisar e comparar o desempenho das abordagens propostas neste estudo com as metodologias encontradas na literatura.

Os métodos de redução de ruído sob avaliação são: (i) o arranjo de microfones baseado no conformador de feixes MVDR, Equação (45) com $\alpha=0$ (CFV) baseado na literatura e apresentado no Quadro 2; (ii) o método de redução de ruído com restrição da potência, Equação (79) com $\alpha=0$, $\Delta=0$ e solução através do pseudocódigo apresentado no Quadro 4 (CFP) método proposto; e (iii) o método de redução de ruído com restrição da envoltória temporal, Equação (73) com $\alpha=0$ (CFE) método também proposto e apresentado no Quadro 3. Como elemento de referência também é apresentado o resultado para o sinal sem processamento.

7.1 VERIFICAÇÃO DA SOLUÇÃO ANALÍTICA PARA A PROPOSTA COM RESTRIÇÃO DE POTÊNCIA

Nesta seção é feita a verificação do procedimento apresentado no Quadro 4 em comparação ao uso de algoritmos de otimização numérica para resolver o problema proposto em (79), neste caso, foi utilizado o método do gradiente descendente para minimizar a função de otimização. Os resultados são comparados em termos das métricas SINR, SRMR-CI e STOI. A Tabela 2 apresenta os parâmetros utilizados. As Tabelas 3, 4 e 5 apresentam resultados na média, após 20 realizações.

Tabela 2 – Parâmetros para avaliação das metodologias baseadas na restrição de potência.

Parâmetros	<i>M</i>	<i>N</i>	<i>C</i>	<i>K</i>	SNR(dB)	SIR(dB)
	6	200	1	22	30	-10 a 10

Na Tabela 3 são apresentados resultados obtidos para a SINR em dB. Os resultados mostram que a solução semi-fechada apresenta melhores resultados, com um menor desvio padrão para a variação de SIR de -10 dB a 10 dB. Essas informações são evidências da validade do procedimento proposto.

Tabela 3 – SINR de saída obtida a partir da solução da Equação (79) utilizando método numérico e o procedimento semi-fechado proposto no Quadro 4. SNR=30 dB.

SIR (dB) \ Método	Método Numérico	Método Proposto
-10	2,3 ± 0,9	2,8 ± 0,6
-5	4,8 ± 1,3	5,5 ± 1,0
0	8,1 ± 2,9	9,7 ± 3,1
5	11,9 ± 1,8	13,1 ± 1,5
10	14,7 ± 1,3	16,5 ± 1,5

A análise de inteligibilidade obtida pelas métricas SRMR-CI percentual (Tabela 4) e STOI (Tabela 5), revela um aumento conforme a SIR aumenta, o que é esperado, uma vez que há menos ruído aditivo no sistema. Notavelmente, para todas as SIRs, o procedimento proposto de solução semi-fechada apresenta resultados semelhantes em termos de inteligibilidade. Isso se deve, provavelmente, a uma qualidade do sinal de saída muito semelhante (Tabela 3) para ambos os métodos de solução da Equação (79) e ao mapeamento não-linear associado ao critério SRMR-CI.

Quanto ao tempo de processamento das duas técnicas, constatou-se que a solução obtida pelo método numérico requer, em média, 593,24 segundos enquanto o método proposto demanda cerca de 332,64 segundos, ou seja, 56% do tempo do procedimento numérico.

Tabela 4 – Inteligibilidade percentual de saída, via SRMR-CI, obtida a partir da solução da Equação (79) utilizando método numérico e o procedimento semi-fechado proposto no Quadro 4. SNR=30 dB.

Método SIR (dB)	Método Numérico	Método Proposto
-10	10,8 ± 7	10,8 ± 7
-5	54,3 ± 11,8	54,3 ± 11,8
0	82 ± 2,8	82 ± 2,8
5	86,4 ± 1,4	86,4 ± 1,4
10	87,0 ± 1,3	87,0 ± 1,3

Tabela 5 – STOI obtida a partir da solução da Equação (79) utilizando método numérico e o procedimento semi-fechado proposto no Quadro 4. SNR=30 dB.

Método SIR (dB)	Método Numérico	Método Proposto
-10	0,569 ± 0,038	0,4977 ± 0,041
-5	0,598 ± 0,049	0,623 ± 0,040
0	0,696 ± 0,035	0,722 ± 0,031
5	0,775 ± 0,027	0,783 ± 0,027
10	0,828 ± 0,023	0,828 ± 0,025

7.2 COMPARAÇÕES ENTRE MÉTODOS DE REDUÇÃO DE RUÍDO

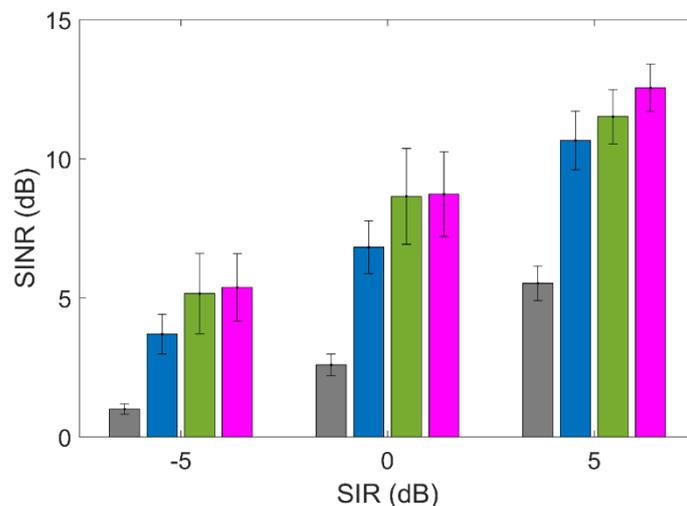
Nesta seção são apresentados resultados comparativos para as seguintes técnicas. (i) o conformador de feixes MVDR, Equação (45) com $\alpha=0$ (CFV) (em azul); (ii) o método de redução de ruído com restrição da potência, Equação (79) com $\alpha=0$, $\Delta = 0$ e solução através do procedimento apresentado no Quadro 4 (CFP) (verde) (proposto); e (iii) o método de redução de ruído com restrição da envoltória temporal, Equação (73) com $\alpha=0$, $\Delta = 40$ (CFE) (magenta) (proposto). Como elemento de referência também é apresentado o resultado para o sinal sem processamento (cinza). São utilizados gráficos de barras onde o ápice da barra identifica a média obtida e as hastes na cor preta informam o desvio padrão. A SNR de entrada utilizada foi de 30 dB.

A diferença estatística entre os conjuntos de dados foi determinada pelo teste de Kruskal-Wallis, H_0 : todas as distribuições são iguais, no nível de significância $p > 0,05$, para verificar quais pares diferiam. Como H_0 foi rejeitada, foi aplicado um teste

de comparações múltiplas utilizando o método Dunn-Bonferroni, H_0 : pares possuindo distribuições idênticas em nível de significância $p > 0,05$ (Corder; Foreman, 2014).

Na Figura 28 são apresentados os resultados para a $SINR_{dB}^{out}$, onde é possível identificar que os métodos CFP e CFE proporcionam uma maior redução da interferência/ruído para todos os valores de SIR avaliados. Esse resultado corrobora a expectativa inicial, uma vez que CFP e CFE não impõem restrições de fase ao vetor de coeficientes, aumentando as possibilidades para a minimização da função custo. Estatisticamente, os resultados obtidos mostram que, na média, CFP e CFE são similares, porém CFE apresenta uma ligeira vantagem em relação ao CFP, que exibe uma maior variabilidade. Para todas as SIRs analisadas, os métodos propostos apresentam resultados superiores ao método CFV.

Figura 28 – $SINR_{dB}^{out}$ em função da SIR de entrada, utilizando o vocoder tonal para a síntese do sinal. Fala contaminada (cinza) e processada por: CFV (azul), CFP (verde) e CFE (magenta).



Fonte: Autoria própria.

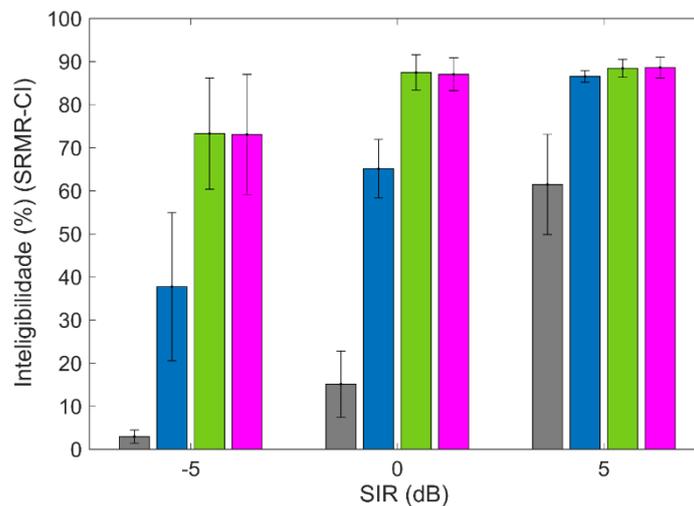
As estimativas de inteligibilidade para os métodos avaliados de acordo com os critérios objetivos SRMR-CI, STOI e NCM são apresentadas nas Figura 29, Figura 30 e Figura 31, respectivamente.

A Figura 29 ilustra os resultados obtidos para a inteligibilidade utilizando a métrica SRMR-CI percentual. Para o SIR de -5 dB, situação de elevada contaminação por ruído aditivo. Os valores de inteligibilidade para CFE e CFP são superiores a 70% em média, valor este considerado suficiente para a manutenção da comunicação

(Volberg *et al.*, 2006; Santos *et al.*, 2013). Esses valores são significativamente maiores do que os encontrados pela metodologia clássica CFV. Além disso, nota-se que as metodologias propostas conseguem reduzir significativamente o ruído em comparação ao sinal contaminado.

Para SIR de 0 dB, CFP e CFE apresentam resultados semelhantes e superiores ao CFV. Para SIR de 5 dB, os níveis de contaminação por ruído aditivo são menores, o que permite que a metodologia clássica também alcance resultados acima de 70% de inteligibilidade, mas não que supere as metodologias propostas.

Figura 29 – Inteligibilidade percentual baseada na SRMR-CI. Fala contaminada (cinza) e processada por: CFV (azul), CFP (verde) e CFE (magenta).



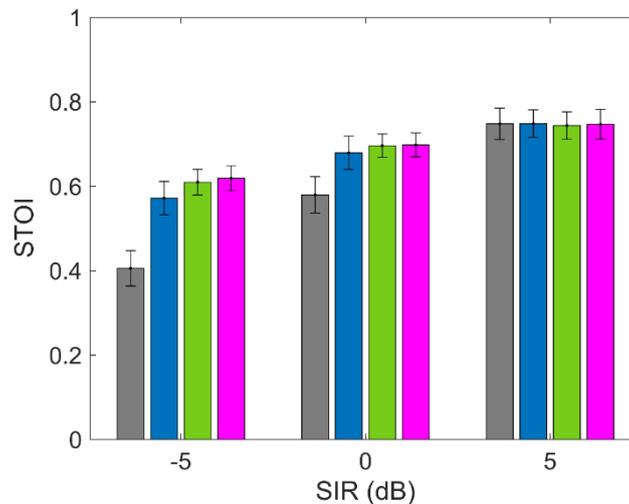
Fonte: Autoria própria.

Os testes estatísticos indicam que o CFP e o CFE não apresentam diferenças estatísticas significativas entre si. No entanto, o desempenho do CFV é significativamente inferior ao CFP e CFE para SIR < 5 dB. O CFV é inadequado para comunicação em SIR = -5 dB, pois requer pelo menos 10 dB adicionais de SIR para atingir 70% de inteligibilidade.

Ao analisar a métrica SRMR-CI percentual, especialmente sua construção metodológica, verifica-se que ela carece de informações mais detalhadas sobre a estrutura fina dos sinais analisados pelo sistema de estimulação do IC. Isso pode limitar a capacidade de diferenciação do método CFE em comparação ao método CFP.

Os resultados obtidos pela métrica STOI são apresentados na Figura 30. Observa-se que, para SIRs de -5 a 0 dB, a metodologia CFE apresenta, em média, os melhores níveis de inteligibilidade quando comparada com CFV, CFP e o sinal contaminado. É importante ressaltar que, para a métrica STOI, considera-se que uma conversação adequada, ou seja, a pessoa consegue compreender o que é dito, em valores próximos ou superiores a $0,7$ (Volberg *et al.*, 2014). Este nível só é obtido na SIR de 0 dB para CFE e CFP. Para SIR de 5 dB todas as metodologias tendem a apresentar valores de $STOI \geq 0,7$, pois o sistema quase não apresenta ruído nesse cenário. Em uma análise estatística dos resultados da métrica STOI é possível identificar que não há diferenças estatísticas significativas entre CFP e CFE para todos os valores de SIRs analisados.

Figura 30 – Inteligibilidade STOI. Fala contaminada (cinza) e processada por: CFV (azul), CFP (verde) e CFE (magenta).



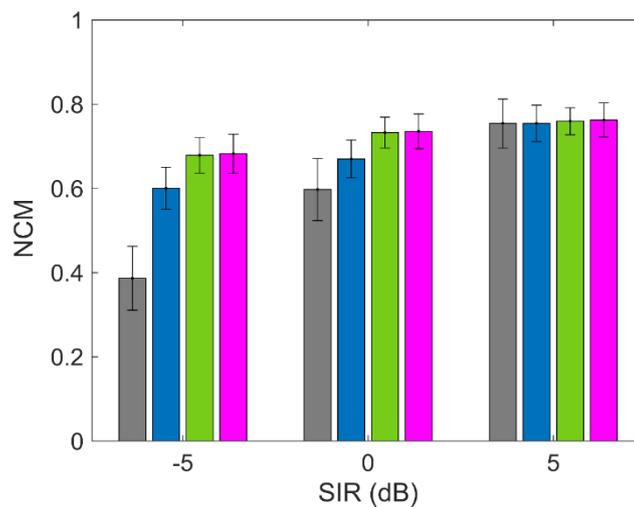
Fonte: Autoria própria.

Os resultados obtidos com a métrica NCM são mostrados na Figura 31. Essa métrica também avalia a inteligibilidade e considera que para valores acima de $0,7$ é possível estabelecer conversação (Goldsworthy; Greenberg, 2004). Para SIR de -5 a 5 dB, a metodologia CFE apresentou os melhores resultados em média. Em geral, os valores são estatisticamente iguais aos obtidos por CFP e ambos são melhores que a metodologia clássica, principalmente para SIRs de -5 e 0 dB, onde o sistema possui mais ruído aditivo. Isso indica que as metodologias propostas possuem um ganho

considerável na região onde o IC mais precisa de melhorias em termos de inteligibilidade, que é quando há mais ruído no ambiente. Os resultados do NCM corroboram as evidências apresentadas pelo SRMR-CI.

As Figuras 32 e 33 apresentam diagramas de caixas bidimensionais que exploram a relação entre inteligibilidade percentual, conforme medida pelo SRMR-CI, e pelo STOI, seguido pelo NCM. Essas análises foram conduzidas considerando sinais de entrada com SIR = -5 dB, escolhida por ser a situação com maior presença de ruído. Nos diagramas, os círculos preenchidos na cor preta indicam a mediana, enquanto que as barras se referem ao primeiro e terceiro quartis. Os valores discrepantes (*outliers*) foram omitidos para facilitar a interpretação dos resultados.

Figura 31 – Inteligibilidade NCM. Fala contaminada (cinza) e processada por: CFV (azul), CFP (verde) e CFE (magenta).



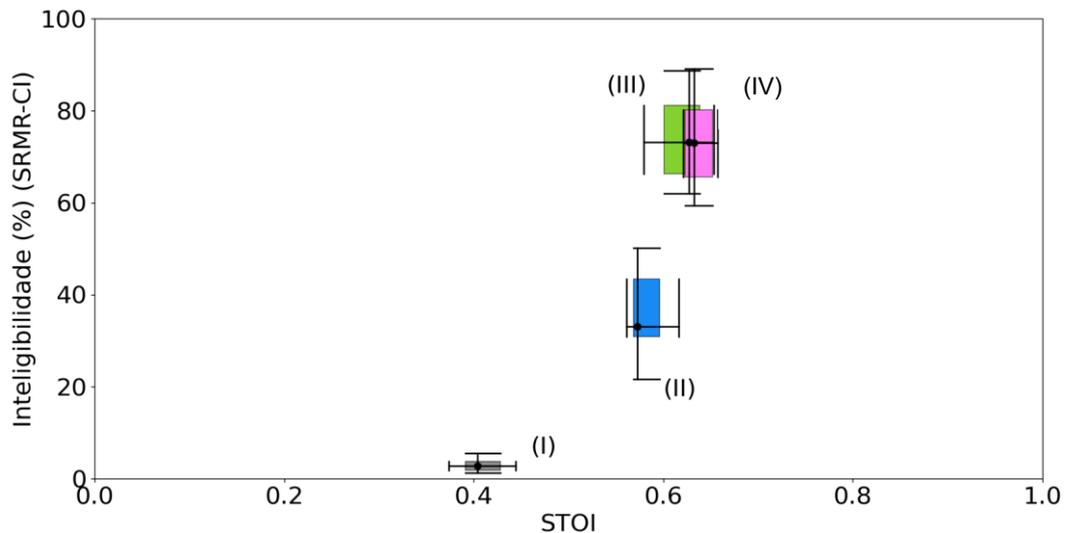
Fonte: Autoria própria.

Na Figura 32 o diagrama de caixas bidimensional apresenta os resultados obtidos para SNR = 30 dB, SIR = -5 dB e SINR = -5 dB. Neste contexto, observa-se que o sinal de fala contaminada (I) exibe valores percentuais de inteligibilidade muito pequenos (SRMR-CI <10% e STOI <0,5). Para o método CFV (II), a inteligibilidade percentual foi superior a 20%, com um STOI de 0,58. Para o método CFP (III), a inteligibilidade percentual ficou em torno de 73%, enquanto o STOI ficou abaixo de 0,64. Por outro lado, para o método CFE (IV), a inteligibilidade percentual foi de 72,5%

e o STOI foi superior ao obtido pelo CFP, chegando a uma média de 0,64. Isso sugere que a metodologia CFE é mais efetiva quando comparada com o CFP.

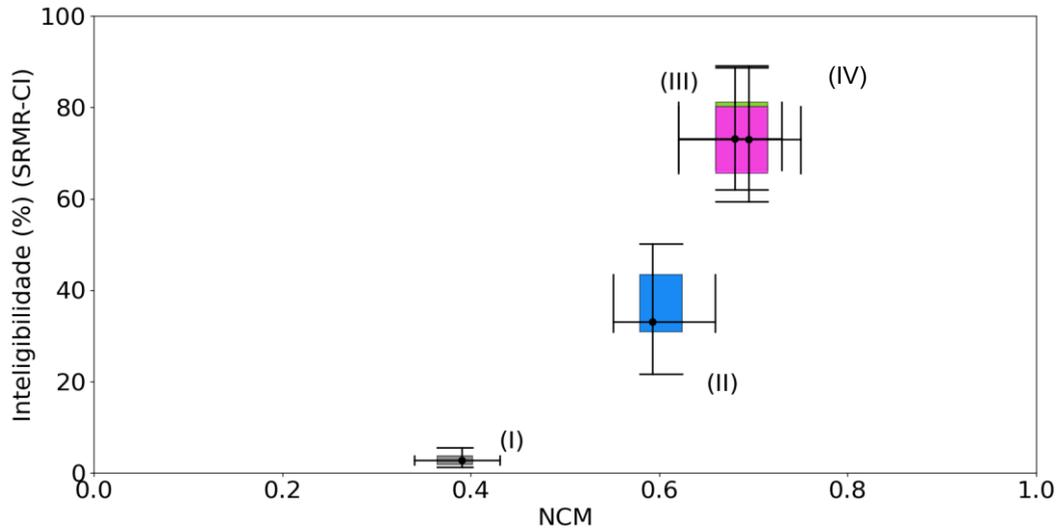
A avaliação utilizando o diagrama de caixas bidimensional para a comparação do SRMR-CI em relação ao NCM é apresentada na Figura 33. Nesta representação fica evidente que as metodologias propostas apresentam níveis de inteligibilidade maiores. Em média, o método CFE demonstra uma melhoria se comparado com o CFP.

Figura 32 – Diagrama de caixas bidimensional para inteligibilidade percentual × STOI. Fala contaminada (cinza) e processada por: CFV (azul), CFP (verde) e CFE (magenta).



Fonte: Autoria própria.

Figura 33 – Diagrama de caixas bidimensional para inteligibilidade percentual \times NCM. Fala contaminada (cinza) e processada por: CFV (azul), CFP (verde) e CFE (magenta).



Fonte: Autoria própria.

Com os resultados apresentados na Figura 29 até a Figura 33 notou-se que para a metodologia CFE, que se baseia na restrição do envelope temporal, os resultados obtidos são estatisticamente semelhantes ao CFP. Investigando o motivo de não ser possível observar uma melhoria na estimativa feita utilizando a envoltória temporal no CFE, observou-se que esse resultado pode ser consequência da cadeia de processamento interna dos critérios objetivos SRMR-CI, NCM e STOI, cuja análise evidenciou processos de filtragem passa-baixas que podem interferir na investigação proposta. Assim, uma aproximação do envelope do sinal pelo seu valor absoluto seria mais precisa para sub-bandas de frequência mais altas. Como resultado, estas métricas objetivas podem não ser adequadas para realizar investigações sobre a precisão do envelope do sinal.

A partir do exposto, sugere-se o uso do CFP para reduzir ruído em ICs. Em relação ao custo computacional, o CFE requer um valor Δ para calcular a envoltória temporal, e quanto maior esse valor, maior o tempo de processamento. Por outro lado, o CFP foi desenvolvido usando uma solução semi-fechada. Assim, pode-se dizer que o CFP apresenta uma solução mais rápida para o problema e desta forma resulta em um interessante compromisso entre desempenho de redução de ruído e custo computacional em ICs.

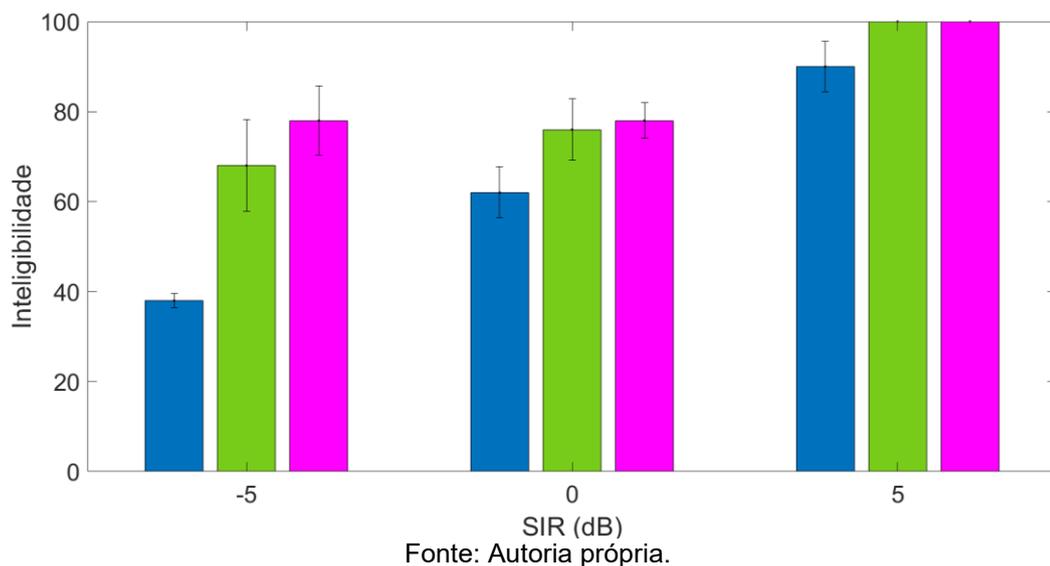
7.3 EXPERIMENTOS PSICOACÚSTICOS

Nesta seção são apresentados resultados de dois experimentos psicoacústicos, sendo o primeiro com indivíduos normouvintes e o segundo com implantados cocleares.

Considera-se que o percentual máximo (100%) é inteligibilidade perfeita, ou seja, o paciente foi capaz de repetir todas as palavras da frase que lhe foi apresentada, e 0% significa que não houve compreensão da fala.

Os resultados do primeiro experimento com normouvintes são apresentados na Figura 34. Nota-se que as metodologias CFE e CFP se destacam em termos de inteligibilidade para todos os pacientes quando comparadas ao método CFV. Tal fato corrobora os resultados obtidos pelas métricas objetivas. Para SIR = -5 dB, o método CFE teve melhor desempenho.

Figura 34 – Resultado do teste psicoacústico com normouvintes. Comparação do CFV (azul), CFP (verde) e CFE (magenta).

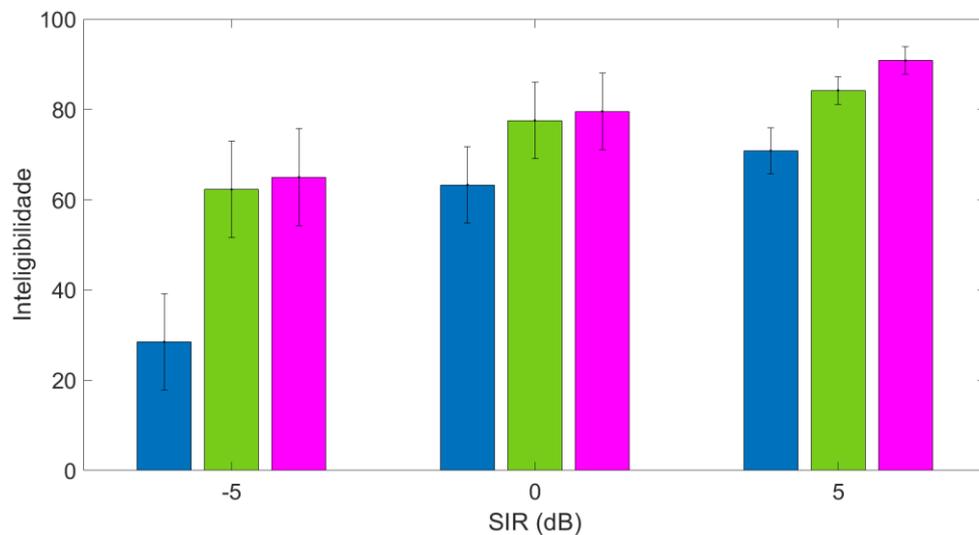


Para o experimento com pacientes implantados, os resultados obtidos são apresentados na Figura 35. Para SIRs de -5 a 5 dB, as metodologias propostas, CFE e CFP apresentam resultados significativamente maiores que a metodologia utilizada na literatura (CFV). Para a situação de maior contaminação, SIR = -5dB, o método CFE apresenta melhor desempenho do que o método CFP, sendo que esse efeito se

repete para $SIR = \{0, 5\}$ dB. Esse resultado aponta para uma melhoria na percepção do implantado quando considerada a envoltória temporal.

A partir dos resultados dos testes psicoacústicos tem-se fortes evidências das vantagens do uso da envoltória temporal em métodos de redução de ruído para aplicações em IC, corroborando os resultados de trabalhos anteriores (Chiea, Costa, Cordioli, 2021).

Figura 35 – Resultado do teste psicoacústico com usuários de implante coclear. Comparação entre CFV (azul), CFP (verde) e CFE (magenta).



Fonte: Autoria própria.

7.4 COMPLEXIDADE COMPUTACIONAL

Por fim, a Tabela 6 apresenta uma comparação entre os tempos médios de cálculo do CFV, CFP (pelo procedimento semi-fechado) e CFE. O método semi-fechado para o cálculo dos coeficientes do CFP é mais rápido que a solução de forma fechada CFV e a solução por métodos numéricos do CFE. Os resultados apresentados indicam que o CFP fornece um compromisso interessante entre ganho de inteligibilidade e custo computacional reduzido, importantes para aplicações que envolvem implantes cocleares.

Tabela 6 - Tempo médio para o cálculo dos filtros ótimos.

	CFV	CFP	CFE
Tempo [s]	845,94	332,64	21601,20

8 CONCLUSÃO

Este trabalho apresenta um método multimicrofone baseado em envoltória temporal para redução de ruído em implantes cocleares. Para tanto, é empregada uma abordagem baseada em formação de feixe com variância mínima e resposta sem distorção na qual é utilizada uma restrição de equivalência entre os momentos estatísticos de segunda ordem das envoltórias da fala original e processada. Um caso particular do método proposto resulta em um MVDR com restrição baseada em potência que possui uma solução semi-fechada.

Com base no problema de minimização da função custo associada ao arranjo de microfones buscou-se reduzir o ruído nos implantes cocleares alterando a restrição do problema, dando origem às metodologias CFE e CFP. O método CFE é baseado na preservação da envoltória temporal do sinal de fala que aumenta sua precisão conforme Δ aumenta. Para o caso particular de manutenção das características da potência da fala, denominado CFP, é possível obter uma solução semi-fechada. Ambas abordagens, quando comparadas com o método clássico da literatura, apresentam ganhos significativos de inteligibilidade.

Os resultados obtidos, principalmente com o método CFE, deixam claro que há ganhos de inteligibilidade nas SIRs abaixo de 0 dB, ou seja, o CFE foi capaz de aumentar a inteligibilidade mesmo em condições de muito baixa razão sinal interferência. As simulações usando critérios objetivos e os experimentos psicoacústicos com normouvintes e usuários de IC forneceram evidências adicionais da adequação da proposta, generalizando resultados publicados anteriormente para um método de redução de ruído baseado em microfone único.

A seguir são listadas as contribuições realizadas nesta tese.

- Desenvolvimento de uma teoria de filtragem multimicrofones baseada na minimização de uma função custo com restrição baseada na preservação da envoltória temporal da fala. Esse procedimento é semelhante ao MVDR clássico para o caso particular do problema com restrição de potência;

- Demonstração do efeito do parâmetro Δ na estimativa da envoltória e seu impacto no desempenho do método de redução de ruído;
- Desenvolvimento da solução particular da função custo com restrição baseada na preservação da potência da fala, resultando em uma solução semi-fechada com menor tempo de processamento;
- Realização de simulações numéricas e experimentos psicoacústicos que apontam para um melhor desempenho das técnicas propostas em termos de inteligibilidade, em relação à técnica clássica da literatura.

Além dessas contribuições, vale ressaltar que este trabalho foi pioneiro no estudo e proposição de técnicas de redução de ruído para implantes cocleares multimicrofones no Brasil. Assim sendo, trouxe conhecimento metodológico relevante aos pesquisadores, principalmente no que diz respeito à realização dos experimentos psicoacústicos com usuários de IC.

8.1 PUBLICAÇÕES

Em periódico científico:

SOUZA, Luciana M. X. De; COSTA, Márcio H; BORGES, Renata C. Envelope-Based Multichannel Noise Reduction for Cochlear Implant Applications. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 32, p. 1873-1884, 2024.

Em conferências científicas:

CHIEA, Rafael A.; SOUZA, Luciana M. X. de; COSTA, Márcio H. Redução de Ruído em Implantes Cocleares: Preservar a Magnitude Espectral ou a Envoltória Temporal?. XXXVIII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais. 2020.

SOUZA, Luciana M. X. De; COSTA, Márcio H; BORGES, Renata C. Redução de Ruído por Conformação de Feixe em Implantes Cocleares: Um Estudo no Domínio do Tempo. XXVIII Congresso Brasileiro de Engenharia Biomédica. 2022.

SOUZA, Luciana M. X. De; COSTA, Márcio H; BORGES, Renata C. Redução de Ruído em Implantes Cocleares Utilizando Arranjos de Microfones. XLI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais. 2023.

SOUZA, Luciana M. X. De; COSTA, Márcio H; BORGES, Renata C. Redução de Ruído em Implantes Cocleares baseado em Arranjo de Microfones com Preservação da Envoltória da Fala. XXIX Congresso Brasileiro de Engenharia Biomédica. 2024.

8.2 TRABALHOS FUTUROS

Este trabalho buscou não só explorar e descrever algumas questões relativas ao tema, como também levantar questões adicionais a serem estudadas posteriormente. Sendo assim, na sequência são apresentadas possíveis considerações para estudos posteriores. Acredita-se que os seguintes pontos são interessantes de serem explorados:

- Desenvolver uma solução semi-definida baseada na extração da envoltória temporal dos sinais para múltiplos microfones e N sub-bandas de estimulação;
- Estudar e desenvolver métricas objetivas para a análise da inteligibilidade quando avaliada pela envoltória temporal no domínio do tempo;
- Analisar o efeito da diminuição do número de sub-bandas na inteligibilidade do sinal processado.

REFERÊNCIAS

- ALCAIM, A.; SOLEWICZ, J. A.; MORAES, J. A. Frequência de ocorrência dos fones 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.
- ALI, H. UT-Dallas Cochlear implant research interface: System level design. 2016.
- ALI, H.; LOBO, A. P.; LOIZOU, P. C. Design and evaluation of a personal digital assistant-based research platform for cochlear implants. *IEEE Transactions on Biomedical Engineering*, v. 60, n. 11, pp. 3060–3073, 2013.
- ALI, R.; VAN WATERSCHOOT, T.; MOONEN, M. Integration of a priori and estimated constraints into an MVDR beamformer for speech enhancement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 27, n. 12, p. 2288-2300, 2019.
- ANSI. American National Standard Institute. American National Standard Specification for Audiometers (ANSI S3.6). New York: ANSI. 1989.
- ARNDT, P. L. Within subject comparison of advanced coding strategies in the Nucleus 24 cochlear implant. In: 1999 Conference on Implantable Auditory Prostheses, Asilomar, CA. 1999.
- ASHA. American Speech-Language-Hearing Association. Causes of Hearing Loss in Adults. 2019. Disponível em: <www.asha.org/public/hearing/causes-of-hearing-loss-in-adults/>. Acesso 19 de Out. 2020.
- BENESTY, J.; CHEN, J.; HUANG, Y. *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag. 2008.
- BENESTY, J.; SOUDEN, M.; CHEN, J. A perspective on multichannel noise reduction in the time domain. *Applied acoustics*. v.74, n.3, p.343-355. 2013.
- BLAUERT, Jens. *The technology of binaural listening*. Springer Science & Business Media, 2013.
- BOGLE, J. M. *Mayo clinic on hearing & balance: hear better, improve your balance, enjoy life*. 3ª edição. Mayo foundation for medical education and research. USA. 2022.
- BOURGEOIS, J.; MINKER, W. *Time-domain beamforming and blind source separation: speech input in the car environment*. Boston, MA: Springer US, 2009.
- BRASIL. Ministério da Saúde. Secretaria de Assistência à Saúde. *Norma Operacional Básica de Assistência à Saúde*. Brasília, 2016. Disponível em: <<https://aps.saude.gov.br/>>. Acesso em 29 jan. 2020.

BRITO NETO, R. V. Estudo dos resultados auditivos e da qualidade de vida em pacientes com implante coclear multicanal. Universidade de São Paulo, São Paulo, 2001.

BÜCHNER, A.; DYBALLA, K. H.; HEHRMANN, P.; FREDELAKE, S.; LENARZ, T. Advanced beamformers for cochlear implant users: Acute measurement of speech perception in challenging listening conditions. PLoS ONE, v. 9, n. 4, 2014.

BUCKLEY, K. M. Spatial/Spectral Filtering with Linearly Constrained Minimum Variance Beamformers. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1987.

CHEN, Y.; GONG, Q. Broadband beamforming compensation algorithm in CI front-end acquisition. Biomedical Engineering Online, BioMed Central Ltd., v. 12, n. 1, p. 18, 2013.

CHIEA, R. A.; COSTA, M. H.; CORDIOLI, J. A. An optimal envelope-based noise reduction method for cochlear implants: An upper bound performance investigation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, v. 29, p. 1729-1739, 2021.

CHIEA, R. A.; SOUZA, L. M. X. de; COSTA, M. H. Redução de Ruído em Implantes Cocleares: Preservar a Magnitude Espectral ou a Envoltória Temporal?. XXXVIII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais. 2020.

CHIEA, R. A. Máscaras Tempo-frequência para a redução de ruído aditivo em implantes cocleares. 2019. Tese (Doutorado em Engenharia Elétrica) – Universidade Federal de Santa Catarina, Florianópolis, 2019.

CLARK, G. M.; BLAMEY, P. J.; BROWN, A. M.; GUSBY, P. A.; DOWELL, R. C.; FRANZ, B. K.; WEBB, R. L. The University of Melbourne-Nucleus multi-electrode cochlear implant. Advances in oto-rhino-laryngology. v. 38, p. V-IX, 1, 1987.

COCHLEAR LIMITED. Nucleus 7 cochlear implant. 2020. Disponível em: <<https://www.cochlear.com/br/pt/home/ongoing-care-and-support/device-support/nucleus-family/nucleus-7>>. Acesso em: Dez. de 2020.

CONSELHO DE FONOAUDIOLOGIA. Guia de Orientações na Avaliação Audiológica Básica. Abr. 2017. Disponível em: <https://www.fonoaudiologia.org.br/cffa/index.php/guias-e-manuais/>. Acesso em: 10 nov. 2022.

CORDER, G. W.; FOREMAN, D. I. Nonparametric statistics: A step-by-step approach. John Wiley & Sons, 2014.

CORNELIS, B.; MOONEN, M.; WOUTERS, J. Performance analysis of multichannel Wiener filter-based noise reduction in hearing aids under second order statistics estimation errors. IEEE Transactions on Audio, Speech, and Language Processing, v. 19, n. 5, p. 1368-1381, 2010.

COSENTINO, S.; FALK, T. H.; MCALPINE, D.; MARQUARDT, T. 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.

DALLOS, P.; POPPER, A. N.; FAY, R. R.; *The cochlea*. Springer Science. v.8. 1996.

DIETZ, M.; MCALPINE, D. *Advancing Binaural Cochlear Implant Technology Trends in hearing*, 2015.

DOCLO, S.; GANNOT, S. Superdirective beamforming robust against microphone mismatch. *IEEE Transactions on Audio, Speech and Language Processing*, v. 15, n. 2, p. 617–631. 2007.

DOCLO, S.; MOONEN, M. GSVD-based optimal filtering for single and multi-microphone speech enhancement. *IEEE Transactions on signal processing*, v. 50, n. 9, p. 2230-2244, 2002.

DRESCHLER, W. A.; VERSCHUURE, H.; LUDVIGSEN, C.; WESTERMANN, S. ICRA Noises: Artificial Noise Signals with Speech-like Spectral and Temporal Properties for Hearing Instrument Assessment: Ruidos ICRA: Señales de ruido artificial con espectro similar al habla y propiedades temporales para pruebas de instrumentos auditivos. *Audiology*, v. 40, n. 3, p. 148-157, 2001.

EINHORN, R. Hearing aid technology for the 21st Century: A proposal for universal wireless connectivity and improved sound quality *IEEE Pulse*, 2017.

FALK, T. 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.

FERNANDEZ, H. M. Vision and hearing impairments. In: *Fundamentals of Geriatric Medicine*. New York: Springer. p.143-164. 2007.

FOO, C; RUDNER, M.; RÖNNBERG, J.; LUNNER, T. Recognition of Speech in Noise with New Hearing Instrument Compression Release Settings Requires Explicit Cognitive Storage and Processing Capacity. *Journal of the American Academy of Audiology*, v. 18, n. 7, p. 618–631, 2007.

FROST, O. L. An Algorithm for Linearly Constrained Adaptive Array Processing. *Proceedings of the IEEE*, v. 60, n. 8. 1972.

GANNOT, S.; VINCENT, E.; MARKOVICH-GOLAN, S.; OZEROV, A.. A Consolidated Perspective on Multi-Microphone Speech Enhancement and Source Separation. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, v. 25, n. 4, p. 692–730. 2017.

GELFAND, S. A. *Hearing: An introduction to psychological and physiological acoustics*. CRC Press, 2017.

GERSHMAN, A. B.; LUO, Z. Q.; SHAHBAZPANAHI, S.; VOROBYOV, S. A. Robust adaptive beamforming using worst-case performance optimization. *Conference*

Record of the Asilomar Conference on Signals, Systems and Computers. v. 2, p. 1353-1357. 2003.

GHOJOGH, B.; KARRAY, F.; CROWLEY, M. Eigenvalue and generalized eigenvalue problems: Tutorial. arXiv preprint arXiv:1903.11240, 2019.

GLASBERG, B. R.; MOORE, B. C. J. Derivation of auditory filter shapes from notched-noise data. *Hearing research*, v. 47, n. 1-2, p. 103-138, 1990.

GOEHRING, T.; KESHAVARZI, M.; CARLYON, R. P.; MOORE, B. C. Using recurrent neural networks to improve the perception of speech in non-stationary noise by people with cochlear implants. *The Journal of the Acoustical Society of America*, v. 146, n. 1, p. 705-718, 2019.

GOLDSWORTHY, Ray L.; GREENBERG, Julie E. Analysis of speech-based speech transmission index methods with implications for nonlinear operations. v. 116, n. 6, p. 3679–3689, 2004.

GOLUB, G. H.; VAN LOAN, C. F. *Matrix computations*. Johns Hopkins University JHU press, 2013.

GREENBERG, S.; AINSWORTH, A. W.; POPPER, N. A.; FAY, R. R. *Speech processing in the auditory system*. Springer handbook of auditory research. v.18. 2008.

GRIFFITHS, L.; JIM, C. An alternative approach to linearly constrained adaptive beamforming. *IEEE Transactions on Antennas and Propagation*, v. 30, n. 1, p. 27–34, 1982.

HADAD, E.; MARQUARDT, D.; DOCLO, S.; GANNOT, S. Theoretical Analysis of Binaural Transfer Function MVDR Beamformers with Interference Cue Preservation Constraints. *IEEE Transactions On Audio Speech and Language Processing*, v. 23, n. 12, p. 2449–2464, 2015.

HADAD, E.; DOCLO, S.; GANNOT, S. The Binaural LCMV Beamformer and its Performance Analysis. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 24, n. 3, p. 543–558, 2016.

HERSBACH, A. A.; GRAYDEN, D. B.; FALLON, J. B.; MCDERMOTT, H. J. 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.

HOLUBE, I.; KOLLMEIER, B. 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, set. 1996.

IEEE. Recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, v. 17, n. 3, p. 225–246, 1969.

JOHANNESMA, P. I. M. The pre-response stimulus ensemble of neurons in cochlear nucleus. In: *Symposium on Hearing Theory*. p. 58-59. 1972.

KAYSER, H.; EWERT, S. D.; ANEMÜLLER, J.; ROHDENBURG, T.; HOHMANN, V.; KOLLMEIER, B. Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. *EURASIP Journal on advances in signal processing*, v. 2009, p. 1-10, 2009.

KAK, S. The number theoretic Hilbert transform. *Circuits, Systems, and Signal Processing*, 2014.

KOKKINAKIS, K.; LOIZOU, P. C. Multi-microphone adaptive noise reduction strategies for coordinated stimulation in bilateral cochlear implant devices. *The Journal of the Acoustical Society of America*, 2010.

KAY, S. M. *Intuitive probability and random processes using MATLAB®*. p.1–833. 2006.

KOLB, B.; WHISHAW, I. Q.; TESKEY, G. C. *An introduction to brain and behavior*. New York: Worth, 2001.

KORDUS, M.; TYLER, R. S.; JAN, Ž. E. R. A.; OLESON, J. J. 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.

LIBERMAN, M. C.; EPSTEIN, M. J.; CLEVELAND, S. S.; WANG, H.; MAISON, S. F. Toward a differential diagnosis of hidden hearing loss in humans. *PLoS ONE*, 2016.

LOBATO, W. J. Worst-case optimization robust-MVDR beamformer for stereo noise reduction in hearing aids. Tese (doutorado) - Universidade Federal de Santa Catarina - Florianópolis. p.144, 2020.

LOIZOU, P. C. *Speech Enhancement: Theory and Practice*. 2. ed. Boca Raton, FL: CRC Press, 2013b. ISBN 9781466504226. 2013.

LOIZOU, P. C. Speech processing in vocoder-centric cochlear implants. *Adv. Otorhinolaryngol*, v. 64, p. 109-143, 2006.

LOPEZ, E. A. P.; MARTIN, E. A.; STOHL, J. S.; WOLFORD, R. D.; SCHATZER, R; WILSON, B.S. A binaural cochlear implant sound coding strategy inspired by the contralateral medial olivocochlear reflex. *Ear and hearing*, v. 37, n. 3, p. e138, 2016.

LUENBERGER, D. G. *Optimization by Vector Space Methods*. John Wiley & Sons, Inc; New York. 1969.

LUO, F. L.; PAVLOVIC, C.; NEHORAI, A. Adaptive null-forming scheme in digital hearing aids. *IEEE Transactions on signal processing*, v. 50, n. 7, p. 1583-1590, 2002.

MALAVAR, W. J. L.; COSTA, M. H. Worst-Case-Optimization Robust-MVDR beamformer for stereo noise reduction in hearing aids. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 28, p. 2224-2237, 2020.

MARKOVICH, S.; GANNOT, S.; COHEN, I. Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals. *IEEE Transactions on Audio, Speech and Language Processing*, 2009.

MARQUARDT, D.; HADAD, E.; GANNOT, S.; & DOCLO, S. Optimal binaural LCMV beamformers for combined noise reduction and binaural cue preservation. In: 2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC). IEEE, p. 288-292. 2014.

MARQUARDT, D.; HOHMANN, V.; DOCLO, S. Coherence preservation in multi-channel Wiener filtering based noise reduction for binaural hearing aids. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, p. 8648-8652. 2013.

MARQUARDT, D. Development and evaluation of psychoacoustically motivated binaural noise reduction and cue preservation techniques. S. 202. Doctoral Thesis – University of Oldenburg. 2015.

NIDCD. National Institute on Deafness and Other Communication Disorders. Cochlear Implant. Governo da Austrália. 2018. Disponível em: <<https://www.healthdirect.gov.au/cochlear-implant>>. Acesso em: 20 de Out. 2020.

NIEMAN, C. L.; OH, E. S. Hearing loss. *Annals of internal medicine*, v. 173, n. 11, p. ITC81-ITC96, 2020.

NOGUEIRA, W.; BUCHNER, A.; LENARZ, T.; EDLER, B. A psychoacoustic N-of-M-type speech coding strategy for cochlear implants. *EURASIP Journal on Applied Signal Processing*, England, v. 2005, n. 18, p. 3044–3059, 2005.

OMS. Global Estimates on Prevalence of Hearing Loss. Disponível em: <<https://apps.who.int/iris/bitstream/handle/10665/260336/9789241550260-eng.pdf>>. 2018. Acesso em: 20 de Mar. 2019.

OMS. 1.1 Billion People May Have Hearing Loss. Disponível em: <<https://nacoesunidas.org/oms-1.1BillionPeopleMayHaveHearingLoss>>. 2017. Acesso em: 19 de Mar. 2019.

OMS. World Report on hearing. Disponível: <<https://apps.who.int/iris/bitstream/handle/10665/339906/9789240021501-eng.pdf>>. 2022. Acesso em: 19 de Jan. 2023.

PATTERSON, R. D.; ROBINSON, K.; HOLDSWORTH, J.; MCKEOWN, D.; ZHANG, C.; ALLERHAND, M. Complex sounds and auditory images. *Auditory fisiologia e percepção*, Proc. 9º Simpósio Internacional de Audição. Pergamon, Oxford, pp. 123-177. 1992.

PNS. Pesquisa Nacional de Saúde 2019: Ciclos de vida. Instituto Brasileiro de Geografia e Estatística (IBGE), Coordenação de Trabalho e Rendimento. - Rio de Janeiro: IBGE, pp. 139. 2021. Disponível em: <https://www.pns.icict.fiocruz.br/wp-content/uploads/2021/12/liv101846.pdf>. Acesso em: 20 de mar. de 2024.

POULARIKAS, A. D. The Handbook of Formulas and Tables for Signal Processing. Chapter Hilbert transform. CRC Press, 2018.

RESENDE, L. S.; BERMUDEZ, J. C. M. A linearly-constrained approach for filtered-x Wiener filtering. In: 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO). IEEE, p. 16-20. 2012.

RESENDE, L. S. Algoritmos recursivos de mínimos quadrados para processamento Espacial/Temporal com restrições lineares: Aplicação em Antenas Adaptativas. 136 f. Tese de doutorado. Universidade Estadual de Campinas (UNICAMP). Campinas – SP. 1996.

SANTOS, J. F.; COSENTINO, S.; HAZRATI, O.; LOIZOU, P. C.; FALK, T. H. Objective speech intelligibility measurement for cochlear implant users in complex listening environments. *Speech Communication*, v. 55, n. 7-8, p. 815–824, 2013.

SANTOS, J. F.; FALK, T. 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.

SHAPIRO, W. H.; BRADHAM, T. S. Cochlear implant programming. *Otolaryngologic Clinics of North America*, v. 45, n. 1, p. 111-127, 2012.

SLANEY, M. Auditory toolbox. Interval Research Corporation, Tech. Rep, v. 10, n. 1998, p. 1194, 1998.

SUN, Z.; LI, Y.; JIANG, H.; CHEN, F.; & WANG, Z. A MVDR-MWF combined algorithm for binaural hearing aid system. In: 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS). IEEE, p. 1-4. 2018.

TAAL, C. H.; HENDRIKS, R. C.; HEUSDENS, R.; JENSEN, J. An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 19, n. 7, p. 2125–2136, 2011.

TABIBI, S.; KEGEL, A.; LAI, W. K.; DILLIER, N. investigating the use of a Gammatone filterbank for a cochlear implant coding strategy. *Journal of Neuroscience Methods*, v. 277, p. 63–74, fev. 2017.

TEXAS EAR CENTER. Anatomy of your ear. Disponível em: <<http://www.texasearcenter.com/disorders-of-the-ear#main>>. 2019. Acesso em: 20 de Mar. 2019.

VAN VEEN, B. D.; BUCKLEY, K. M. Beamforming: A Versatile Approach to Spatial Filtering. *IEEE ASSP Magazine*, 1988.

VANDALI, A. E.; WHITFORD, L. A.; PLANT, K. L.; CLARK, G. M. Speech perception as a function of electrical stimulation rate: using the Nucleus 24 cochlear implant system. *Ear and Hearing*, v. 21, n. 6, p. 608–624, 2000.

VOLBERG, L.; KULKA, M.; SUST, C. A.; LAZARUS, H. Speech intelligibility and the subjective assessment of speech quality in near real communication conditions. *Acta acustica united with acustica*, v. 92, n. 3, p. 406-416, 2006.

WANG, D.; BROWN, G. J. Computational auditory scene analysis: Principles, algorithms, and applications. New York, NY: Wiley/IEEE Press, p. 395. 2006.

WANG, R. Generalized eigenvalue problem. Disponível em: <http://fourier.eng.hmc.edu/e161/lectures/algebra/node7.html>, 2015. Acesso: Abril de 2023.

WARD, D.; KENNEDY, R.; WILLIAMSON, R. Theory and design of broadband sensor arrays with frequency invariant far field beam patterns. *The Journal of the Acoustical Society of America*, v. 97, n. 2, p. 1023–1034. 1995.

WILSON, B. S.; FINLEY, C. C.; LAWSON, D. T.; WOLFORD, R. D.; ZERBI, M. Design and evaluation of a continuous interleaved sampling (CIS) processing strategy for multichannel cochlear implants. *Journal of Rehabilitation Research and Development*, v. 30, p. 110-110. 1993.

WOUTERS, J.; VANDEN BERGHE, J. Speech recognition in noise for cochlear implantees with a two microphone monaural adaptive noise reduction system. *Ear & Hearing*, v. 22, n. 5, p. 420–430, 2001.

WOUTERS, J.; MCDERMOTT, H. J.; FRANCAERT, T. Sound coding in cochlear implants: From electric pulses to hearing. v. 32, n. 2, p. 67–80, mar. 2015.

YOST, W. A. *Fundamentals of Hearing - An Introduction*. 5. Ed. Chicago: Academic Press, 2007.

ZEDAN, A.; JÜRGENS, T.; WILLIGES, B.; KOLLMEIER, B.; WIEBE, K.; GALINDO, J.; WESARG, T. Speech Intelligibility and Spatial Release from Masking Improvements Using Spatial Noise Reduction Algorithms in Bimodal Cochlear Implant Users. *Trends in Hearing*, v. 25, p. 23312165211005931, 2021.

ZENG, F. G.; REBSCHER, S.; HARRISON, W.; SUN, X.; FENG, H. Cochlear implants: system design, integration, and evaluation. *IEEE reviews in biomedical engineering*, v. 1, p. 115-142, 2008.

ZENG, F. G.; POPPER, A. N.; FAY, R. R. *Auditory prostheses: New horizons*. Springer Science & Business Media, 2011.

ZHANG, G.; LI, Q.; ZHANG, Q.; QIN, J.; YANG, L. Signal-to-interference-plus-noise ratio-based multi-relay beamforming for multi-user multiple-input multiple-output cognitive relay networks with interference from primary network. *IET Communications*, v. 9, n. 2, p. 227-238, 2015.

ZHANG, J.; LI, C. Quantization-aware binaural MWF based noise reduction incorporating external wireless devices. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 29, p. 3118-3131, 2021.

ZIRN, S.; ANGERMEIER, J.; ARNDT, S.; ASCHENDORFF, A.; WESARG, T. Reducing the device delay mismatch can improve sound localization in bimodal cochlear implant/hearing-aid users. *Trends in Hearing*, v. 23, 2019.

APÊNDICE A – DERIVAÇÃO DO FILTRO DE WIENER MULTICANAL

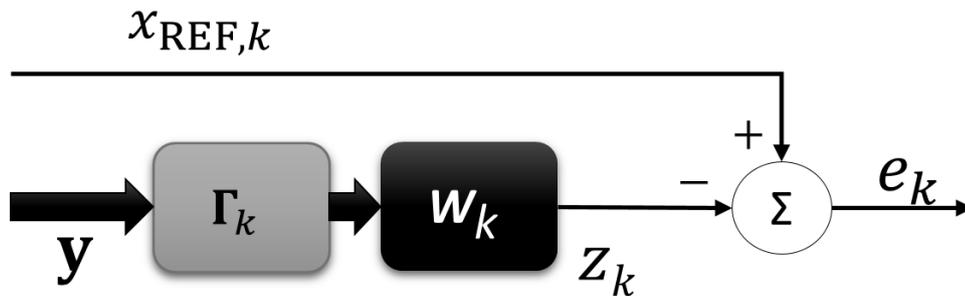
Neste apêndice é apresentada a derivação do filtro de Wiener multicanal no domínio do tempo (Doclo e Moonen, 2002).

O problema de filtragem considerado é apresentado na Figura A.1. O sinal $x_{\text{REF},k}(n)$ é o sinal desejado (fala captada pelo microfone de referência na sub-banda k).

O sinal de erro, que deve ser minimizado é dado por:

$$e_k(n) = x_{\text{REF},k}(n) - z_k(n). \quad (\text{A.1})$$

Figura A.1 – Problema de filtragem ótima utilizando a resposta desejada $x_{\text{REF},k}$.



Fonte: Autoria própria.

Definindo-se a função custo como o valor quadrático médio do sinal de erro tem-se que:

$$J_{\text{MSE}_k}(\mathbf{w}_k) = E\{e_k^2(n)\}, \quad (\text{A.2})$$

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

Substituindo (A.1) em (A.2), tem-se:

$$J_{\text{MSE}_k}(\mathbf{w}_k) = E\{[x_{\text{REF},k}(n) - z_k(n)]^2\}. \quad (\text{A.3})$$

Substituindo-se (17) e (27) em (A.3) chega-se em:

$$J_{\text{MSE}_k}(\mathbf{w}_k) = E\{[\mathbf{x}^T(n)\Gamma_k^T \mathbf{q}_{\text{REF}} - \mathbf{y}^T(n)\Gamma_k^T \mathbf{w}_k]^2\}. \quad (\text{A.4})$$

Sendo,

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n), \quad (\text{A.5})$$

e substituindo (A.5) em (A.4) obtém-se:

$$\begin{aligned} J_{\text{MSE}_k}(\mathbf{w}_k) &= E \left\{ [\mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}} - [\mathbf{x}(n) + \mathbf{v}(n)]^T \Gamma_k^T \mathbf{w}_k]^2 \right\} \\ &= E \left\{ [\mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}} - \mathbf{x}^T(n) \Gamma_k^T \mathbf{w}_k - \mathbf{v}^T(n) \Gamma_k^T \mathbf{w}_k]^2 \right\}. \end{aligned} \quad (\text{A.6})$$

Desenvolvendo a Equação (A.6), chega-se em:

$$\begin{aligned} J_{\text{MSE}_k}(\mathbf{w}_k) &= E \left\{ \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{x}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}} \right\} - E \left\{ \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{x}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{w}_k \right\} \\ &\quad - E \left\{ \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{x}(n) \mathbf{v}^T(n) \Gamma_k^T \mathbf{w}_k \right\} - E \left\{ \mathbf{w}_k^T \Gamma_k \mathbf{x}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}} \right\} \\ &\quad + E \left\{ \mathbf{w}_k^T \Gamma_k \mathbf{x}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{w}_k \right\} + E \left\{ \mathbf{w}_k^T \Gamma_k \mathbf{x}(n) \mathbf{v}^T(n) \Gamma_k^T \mathbf{w}_k \right\} \\ &\quad - E \left\{ \mathbf{w}_k^T \Gamma_k \mathbf{v}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}} \right\} + E \left\{ \mathbf{w}_k^T \Gamma_k \mathbf{v}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{w}_k \right\} \\ &\quad + E \left\{ \mathbf{w}_k^T \Gamma_k \mathbf{v}(n) \mathbf{v}^T(n) \Gamma_k^T \mathbf{w}_k \right\} \end{aligned} \quad (\text{A.7})$$

De forma que:

$$\begin{aligned} J_{\text{MSE}_k}(\mathbf{w}_k) &= \mathbf{q}_{\text{REF}}^T \Gamma_k E \left\{ \mathbf{x}(n) \mathbf{x}^T(n) \right\} \Gamma_k^T \mathbf{q}_{\text{REF}} - \mathbf{q}_{\text{REF}}^T \Gamma_k E \left\{ \mathbf{x}(n) \mathbf{x}^T(n) \right\} \Gamma_k^T \mathbf{w}_k \\ &\quad - \mathbf{q}_{\text{REF}}^T \Gamma_k E \left\{ \mathbf{x}(n) \mathbf{v}^T(n) \right\} \Gamma_k^T \mathbf{w}_k - \mathbf{w}_k^T \Gamma_k E \left\{ \mathbf{x}(n) \mathbf{x}^T(n) \right\} \Gamma_k^T \mathbf{q}_{\text{REF}} \\ &\quad + \mathbf{w}_k^T \Gamma_k E \left\{ \mathbf{x}(n) \mathbf{x}^T(n) \right\} \Gamma_k^T \mathbf{w}_k + \mathbf{w}_k^T \Gamma_k E \left\{ \mathbf{x}(n) \mathbf{v}^T(n) \right\} \Gamma_k^T \mathbf{w}_k \\ &\quad - \mathbf{w}_k^T \Gamma_k E \left\{ \mathbf{v}(n) \mathbf{x}^T(n) \right\} \Gamma_k^T \mathbf{q}_{\text{REF}} + \mathbf{w}_k^T \Gamma_k E \left\{ \mathbf{v}(n) \mathbf{x}^T(n) \right\} \Gamma_k^T \mathbf{w}_k \\ &\quad + \mathbf{w}_k^T \Gamma_k E \left\{ \mathbf{v}(n) \mathbf{v}^T(n) \right\} \Gamma_k^T \mathbf{w}_k \end{aligned} \quad (\text{A.8})$$

Considera-se que fala e o ruído são variáveis aleatórias independentes de média zero e substituindo-se (21) e (23), chega-se em:

$$\begin{aligned} J_{\text{MSE}_k}(\mathbf{w}_k) &= \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} - \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{w}_k \\ &\quad - \mathbf{w}_k^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} + \mathbf{w}_k^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{w}_k + \mathbf{w}_k^T \Gamma_k \mathbf{R}_{\text{vv}}(n) \Gamma_k^T \mathbf{w}_k \end{aligned} \quad (\text{A.9})$$

O filtro ótimo ou filtro multidimensional de Wiener pode ser encontrado definindo-se a derivada de $\partial J_{\text{MSE}_k} / \partial \mathbf{w}_k = \mathbf{0}$.

Assim, derivando-se a Equação (A.9), obtém-se:

$$\begin{aligned}
\frac{\partial J_{\text{MSE}_k}}{\partial \mathbf{w}_k} &= \frac{\partial}{\partial \mathbf{w}_k} \left[\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} \right] - \frac{\partial}{\partial \mathbf{w}_k} \left[\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{w}_k \right] \\
&\quad - \frac{\partial}{\partial \mathbf{w}_k} \left[\mathbf{w}_k^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} \right] + \frac{\partial}{\partial \mathbf{w}_k} \left[\mathbf{w}_k^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{w}_k \right] \quad (\text{A.10}) \\
&\quad + \frac{\partial}{\partial \mathbf{w}_k} \left[\mathbf{w}_k^T \Gamma_k \mathbf{R}_{\text{vv}}(n) \Gamma_k^T \mathbf{w}_k \right]
\end{aligned}$$

Igualando-se (A.10) ao vetor de zeros:

$$\begin{aligned}
& - \left(\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \right)^T - \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} \\
& \quad + 2 \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{w}_k + 2 \Gamma_k \mathbf{R}_{\text{vv}}(n) \Gamma_k^T \mathbf{w}_k^{\text{Wiener}} = \mathbf{0}, \quad (\text{A.11})
\end{aligned}$$

em que $\mathbf{0}$ tem dimensão $1 \times MC$, e reorganizando a Equação (A.11) chega-se em:

$$\left(2 \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T + 2 \Gamma_k \mathbf{R}_{\text{vv}}(n) \Gamma_k^T \right) \mathbf{w}_k^{\text{Wiener}} = \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} + \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}}. \quad (\text{A.12})$$

De acordo com as propriedades apresentadas em (23), pode-se reescrever (A.12) como:

$$\left(2 \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T + 2 \Gamma_k \mathbf{R}_{\text{vv}}(n) \Gamma_k^T \right) \mathbf{w}_k^{\text{Wiener}} = 2 \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}}, \quad (\text{A.13})$$

logo,

$$\begin{aligned}
& \left(2 \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T + 2 \Gamma_k \mathbf{R}_{\text{vv}}(n) \Gamma_k^T \right)^{-1} \left(2 \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T + 2 \Gamma_k \mathbf{R}_{\text{vv}}(n) \Gamma_k^T \right) \mathbf{w}_k^{\text{Wiener}} \\
& \quad = \left(2 \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T + 2 \Gamma_k \mathbf{R}_{\text{vv}}(n) \Gamma_k^T \right)^{-1} 2 \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} \quad (\text{A.14})
\end{aligned}$$

Reorganizando-se (A.14) tem-se:

$$\mathbf{w}_k^{\text{Wiener}} = \left(\Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T + \Gamma_k \mathbf{R}_{\text{vv}}(n) \Gamma_k^T \right)^{-1} \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}}. \quad (\text{A.15})$$

com dimensão $MC \times 1$.

APÊNDICE B – DERIVAÇÃO MVDR

Neste apêndice é apresentada a derivação do MVDR no domínio do tempo baseada no trabalho de Benesty, Souden e Chen (2013).

Seja,

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n), \quad (\text{B.1})$$

em que $\mathbf{x}(n)$ é o vetor do sinal de interesse (fala) na posição dos microfones de captação, com dimensão $M(N+C-1) \times 1$; \mathbf{A} é uma matriz com os coeficientes das respostas ao impulso dos caminhos de propagação entre a fonte de interesse e os microfones de captação, com dimensão $M(N+C-1) \times M(N_1+N+C-2)$; e $\mathbf{s}(n)$ é o vetor sinal de interesse na posição original da fonte com dimensão $M(N_1+N+C-2) \times 1$. O que pode ser expresso como:

$$\begin{bmatrix} x_1(n) \\ x_1(n-1) \\ \vdots \\ x_1(n-N-C+2) \\ x_2(n) \\ x_2(n-1) \\ \vdots \\ x_2(n-N-C+2) \\ \vdots \\ x_M(n) \\ x_M(n-1) \\ \vdots \\ x_M(n-N-C+1) \end{bmatrix} = \begin{bmatrix} a_{1,0} & a_{1,1} & \cdots & & & & \\ 0 & a_{1,0} & a_{1,1} & \cdots & & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ a_{2,0} & a_{2,2} & \cdots & & & & \\ 0 & a_{2,0} & a_{2,1} & \cdots & & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ a_{M,0} & a_{M,1} & \cdots & & & & \\ 0 & a_{M,0} & a_{M,1} & \cdots & & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \end{bmatrix} \begin{bmatrix} s(n) \\ s(n-1) \\ \vdots \\ s_{1-N-C+2} \end{bmatrix}. \quad (\text{B.2})$$

A restrição do conformador de feixes é dada por:

$$\mathbf{w}_k^T \mathbf{\Gamma}_k \mathbf{x}(n) = x_1(n), \quad (\text{B.3})$$

em que $\mathbf{\Gamma}_k$ é uma matriz bloco-diagonal de dimensão $MC \times M(N+C-1)$ com os coeficientes do banco de filtros.

O sinal $x_1(n)$ pode ser expresso como:

$$x_1(n) = \mathbf{q}_{\text{REF}}^T \mathbf{\Gamma}_k \mathbf{x}(n). \quad (\text{B.4})$$

Substituindo-se (B.4) em (B.3):

$$\mathbf{w}_k^T \Gamma_k \mathbf{x}(n) = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{x}(n). \quad (\text{B.5})$$

Multiplicando-se ambos os lados de (B.5) por (B.4) transposto:

$$\mathbf{w}_k^T \Gamma_k \mathbf{x}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}} = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{x}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}}. \quad (\text{B.6})$$

Substituindo-se (B.1) em (B.6):

$$\mathbf{w}_k^T \Gamma_k \mathbf{A} \mathbf{s}(n) \mathbf{s}^T(n) \mathbf{A}^T \Gamma_k^T \mathbf{q}_{\text{REF}} = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{A} \mathbf{s}(n) \mathbf{s}^T(n) \mathbf{A}^T \Gamma_k^T \mathbf{q}_{\text{REF}}. \quad (\text{B.7})$$

Tomando-se o valor esperado de (B.7):

$$E\{\mathbf{w}_k^T \Gamma_k \mathbf{A} \mathbf{s}(n) \mathbf{s}^T(n) \mathbf{A}^T \Gamma_k^T \mathbf{q}_{\text{REF}}\} = E\{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{A} \mathbf{s}(n) \mathbf{s}^T(n) \mathbf{A}^T \Gamma_k^T \mathbf{q}_{\text{REF}}\}, \quad (\text{B.8})$$

$$\mathbf{w}_k^T \Gamma_k \mathbf{A} E\{\mathbf{s}(n) \mathbf{s}^T(n)\} \mathbf{A}^T \Gamma_k^T \mathbf{q}_{\text{REF}} = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{A} E\{\mathbf{s}(n) \mathbf{s}^T(n)\} \mathbf{A}^T \Gamma_k^T \mathbf{q}_{\text{REF}}, \quad (\text{B.9})$$

$$\mathbf{w}_k^T \Gamma_k \mathbf{A} \mathbf{R}_{\text{ss}}(n) \mathbf{A}^T \Gamma_k^T \mathbf{q}_{\text{REF}} = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{A} \mathbf{R}_{\text{ss}}(n) \mathbf{A}^T \Gamma_k^T \mathbf{q}_{\text{REF}}, \quad (\text{B.10})$$

em que

$$\mathbf{R}_{\text{ss}}(n) = E\{\mathbf{s}(n) \mathbf{s}^T(n)\}, \quad (\text{B.11})$$

é a matriz de correlação do sinal de interesse.

Definindo-se:

$$\tilde{\Gamma} = \frac{1}{\sigma_s^2(n)} \mathbf{R}_{\text{ss}}(n), \quad (\text{B.12})$$

em que

$$\sigma_s^2(n) = E\{s^2(n)\}. \quad (\text{B.13})$$

Substituindo-se (B.12) em (B.10), tem-se:

$$\mathbf{w}_k^T \Gamma_k \mathbf{A} \sigma_s^2(n) \tilde{\Gamma} \mathbf{q}_{\text{REF}} = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{A} \tilde{\Gamma} \mathbf{q}_{\text{REF}}, \quad (\text{B.14})$$

$$\mathbf{w}_k^T \Gamma_k \mathbf{A} \tilde{\Gamma} \mathbf{q}_{\text{REF}} = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{A} \tilde{\Gamma} \mathbf{q}_{\text{REF}}. \quad (\text{B.15})$$

E rearranjando-se (B.15) obtém-se:

$$\mathbf{w}_k^T \left(\frac{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{A} \tilde{\mathbf{I}}}{\Gamma_k^T \mathbf{q}_{\text{REF}}} \Gamma_k \mathbf{A} \tilde{\mathbf{I}} \Gamma_k^T \mathbf{q}_{\text{REF}} \right) = 1, \quad (\text{B.16})$$

em que a parcela no interior dos parênteses em (B.16) é o equivalente no domínio tempo do *steering vector* (Benesty; Souden; Chen, 2013).

O conformador MVDR pode então ser definido como:

$$\begin{aligned} \mathbf{w}_k^{\text{opt}} &= \arg \min_{\mathbf{w}_k} \mathbf{w}_k^T \Gamma_k \mathbf{R}_{\text{vv}}(n) \Gamma_k^T \mathbf{w}_k \\ \text{s.t. } & \mathbf{w}_k^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q} = \mathbf{q}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q} \end{aligned} \quad (\text{B.17})$$

APÊNDICE C – MÉTRICAS DE INTELIGIBILIDADE

Neste apêndice apresentam-se as métricas SRMR-CI e STOI que foram utilizadas para a estimação de inteligibilidade.

C.1 SRMR-CI

A SRMR-CI baseia-se na métrica da razão de energia de modulação entre a fala e reverberação (*Speech to reverberation modulation energy ratio* - SRMR) (Falk, 2010). A SRMR foi adaptada para levar em consideração as especificidades dos implantados cocleares e acabou sendo conhecida como SRMR-CI (Santos; Falk, 2014).

A SRMR-CI é calculada em quatro etapas. Na primeira, o sinal de entrada $x(n)$ é submetido a um banco de filtros de 22 canais, o mesmo utilizado em implantes cocleares do dispositivo *Nucleus* da CochlearTM. Em sequência, calcula-se a envoltória temporal para cada saída do banco de filtros $x_k(n)$, utilizando-se a transformada de Hilbert, obtendo-se o sinal $e_k(n)$. As envoltórias são submetidas à transformada de Fourier de curto termo (*Short Time Fourier Transform* – STFT), de forma a obter-se a energia de modulação espectral para cada banda $\check{E}_k(p,k)$ em que p e k são os índices de quadro (frame) e frequência.

Na terceira etapa, o espectro de modulação é filtrado por um banco de filtros de 8 sub-bandas com frequências centrais espaçadas logaritmicamente entre 4 Hz e 64 Hz.

Por fim, 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.

Considera-se que essa métrica é não intrusiva, isso quer dizer, que ela não necessita do sinal de fala de referência para comparação. Isso é uma vantagem, pois permite sua utilização em tempo real em dispositivos clínicos, principalmente naqueles que precisam de informação para a adaptação do algoritmo de aprimoramento da fala.

O SRMR-CI pode ser mapeado em valores de inteligibilidade da seguinte maneira:

$$I_{\%} = \frac{1}{1 + e^{a_1 \text{SRMR-CI}_{norm} + a_2}} \times 88,92\%, \quad (\text{C.1})$$

em que $I_{\%}$ é o percentual de inteligibilidade; e $a_1 = -12,1742$ e $a_2 = 7,4535$ são os coeficientes de ajuste (Santos; Falk, 2014). A SRMR-CI_{norm} é obtida:

$$\text{SRMR-CI}_{norm} = \frac{\text{SRMR-CI}_t}{\text{SRMR-CI}_s}, \quad (\text{C.2})$$

em que SRMR-CI_t é a métrica do sinal em teste e SRMR-CI_s é o valor da métrica para o sinal de fala.

C.2 STOI

A medida de inteligibilidade objetiva de tempo curto (*Short Term Objective Intelligibility* - STOI) é uma métrica que quantifica a inteligibilidade da fala. Essa métrica é considerada intrusiva, isso quer dizer que é necessário conhecer o sinal de fala original $x(n)$ para estimar a inteligibilidade do sinal processado $z(n)$ (Taal *et al.*, 2011).

A métrica funciona de seguinte forma: inicialmente retiram-se os intervalos de tempo com ausência da fala, ou seja, os intervalos com silêncio são extraídos. A extração dessas regiões é feita removendo as partes do sinal com 40 dB abaixo do valor de máxima energia do sinal original. Na sequência, ambos os sinais são decompostos em 15 bandas com frequências centrais distribuídas entre 150 Hz até 4,3 kHz. A partir disto, as envoltórias dos sinais são extraídas.

O próximo passo é a comparação dos sinais, para isso utiliza-se a envoltória do sinal limpo $e_x(\lambda, m)$ e a envoltória do sinal processado $e_z(\lambda, m)$, essa comparação é feita em blocos, através da obtenção dos coeficientes de correlação. Pode-se representar um bloco temporal da envoltória da fala como:

$$\mathbf{e}_x(\lambda, m) = [e_x(\lambda - E + 1, m) \quad e_x(\lambda - E + 2, m) \quad \cdots \quad m]^T, \quad (\text{C.3})$$

em que E controla o tamanho da janela em análise. Para o sinal processado realiza-se o mesmo processo:

$$\mathbf{e}_z(\lambda, m) = [e_z(\lambda - E + 1, m) \quad e_z(\lambda - E + 2, m) \quad \cdots \quad m]^T. \quad (\text{C.4})$$

Os coeficientes de correlação entre os dois vetores são obtidos da seguinte forma:

$$r(\lambda, m) = \frac{[\mathbf{e}_x(\lambda, m) - \boldsymbol{\mu}_{e_x}(m)]^T \cdot [\bar{\mathbf{e}}_z(\lambda, m) - \boldsymbol{\mu}_{\bar{\mathbf{e}}_z}(m)]}{\|\mathbf{e}_x(\lambda, m) - \boldsymbol{\mu}_{e_x}(m)\| \cdot \|\bar{\mathbf{e}}_z(\lambda, m) - \boldsymbol{\mu}_{\bar{\mathbf{e}}_z}(m)\|}, \quad (\text{C.5})$$

em que $\bar{\mathbf{e}}_z$ é o vetor \mathbf{e}_z normalizado e limitado em amplitude para compensar as diferenças globais que não interferem na inteligibilidade (Taal *et al.*, 2010).

Por fim, pode-se definir a STOI pela média entre todos os coeficientes de correlação, tal que:

$$\text{STOI} = \frac{1}{JB} \sum_{\lambda, m} r(\lambda, m), \quad (\text{C.6})$$

em que J representa o número total de blocos utilizados e B é o número de bandas.

C.3 NCM

A medida de covariância normalizada (*Normalized Covariance Measure* - NCM) realiza em sua primeira etapa a filtragem do sinal em $M=22$ bandas de frequência (Holube; Kollmeier, 1996; Goldsworthy; Greenberg, 2004). Na sequência calcula-se a envoltória temporal de cada sub-banda utilizando a transformada de Hilbert e reamostra-se o sinal para $f_s=100$ Hz.

A covariância entre as envoltórias do sinal de referência (fala original) $e_x(\lambda, m)$ e do sinal cuja inteligibilidade deve ser quantificada $e_y(\lambda, m)$ é calculada para cada uma das M sub-bandas como:

$$r(m) = \frac{\sum_{\lambda} [e_x(\lambda, m) - \mu_x(m)] [e_y(\lambda, m) - \mu_y(m)]}{\sqrt{\sum_{\lambda} [e_x(\lambda, m) - \mu_x(m)]^2} \sqrt{\sum_{\lambda} [e_y(\lambda, m) - \mu_y(m)]^2}}, \quad (\text{C.6})$$

em que $\mu_x(m)$ e $\mu_y(m)$ são as médias de $e_x(\lambda, m)$ e $e_y(\lambda, m)$ nas respectivas m -ésimas sub-bandas. Os valores resultantes para $r(m)$ pertencem ao intervalo $[-1, 1]$.

Na sequência é calculada a SNR aparente em dB para cada sub-banda, tal que:

$$\overline{\text{SNR}}(m) = 10 \log_{10} \left(\frac{r(m)^2}{1 - r(m)^2} \right), \quad (\text{C.7})$$

cujo resultado é limitado ao intervalo $[-15, 15]$ dB.

O índice de transmissão em cada sub-banda é calculado por um mapeamento linear que resulta em valores entre 0 e 1 utilizando a seguinte equação:

$$O(m) = \frac{\overline{\text{SNR}}(m) + 15}{30}. \quad (\text{C.7})$$

A NCM é obtida pelo cálculo da média ponderada entre os índices de transmissão de todas as sub-bandas, tal que:

$$\text{NCM} = \frac{1}{\sum_{m=1}^M U(m)} \cdot \sum_{m=1}^M U(m) \cdot O(m), \quad (\text{C.7})$$

em que $U(m)$ são os pesos de importância dados para cada uma das M sub-bandas.

ANEXO A – TCLE

Neste anexo é apresentado o termo de consentimento aprovado para a realização dos testes psicoacústicos.



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

Campus Universitário – Trindade – Florianópolis-SC – CEP 88040-900
Fone: (48) 3721-9422

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Prezado participante

Gostaríamos de convidá-lo a participar de nossa pesquisa, cujo título é **“Estudos sobre Métodos de Processamento de Sinais para Aparelhos Auditivos e Implantes Cocleares”**.

O objetivo desse documento é informar sobre o referido trabalho para obter uma autorização por escrito, referente à vossa participação na pesquisa que está sendo realizada.

As informações contidas neste termo foram fornecidas por um participante do projeto de pesquisa em questão, sob supervisão do professor Dr. Márcio Holsbach Costa do Departamento de Engenharia Elétrica e Eletrônica, do Professor Dr. Stephan Paul do Departamento de Engenharia Mecânica e da Professora Dra. Maria Madalena Canina Pinheiro do Departamento de Fonoaudiologia da Universidade Federal de Santa Catarina.

Objetivo: Este trabalho tem como objetivo o estudo de métodos de

processamento de sinais para aparelhos auditivos e implantes cocleares.

Método: Sua participação é voluntária e consiste em ouvir sons e, em sequência, responder a perguntas sobre características percebidas, como qualidade, inteligibilidade, conforto acústico e espacialidade (direção de chegada). Os sons podem ser emitidos por um sistema de alto-falantes ou conjunto de fones de ouvido. Caso utilize aparelho auditivo a emissão do som poderá ser realizada através de fone embutido em molde auditivo ou por cabo ligado a uma entrada apropriada de seu dispositivo. Caso utilize implante coclear, poderá ser utilizada interface de pesquisa CCI-Mobile (<https://personal.utdallas.edu/~hussnain.ali/CCIMobile.html>), dispensando o processador externo do seu dispositivo. O tempo total do procedimento é de cerca de cinquenta (50) minutos. Ao mesmo tempo em que você escuta o som você utilizará óculos especiais para monitoração do diâmetro das pupilas de seus olhos. O objetivo é determinar o seu grau de concentração durante o experimento. Dependendo do tipo de experimento poderá ser solicitada a realização de um procedimento conhecido por audiometria, ou similar, realizado por profissionais capacitados, para avaliação da sua condição auditiva.

Benefícios: A participação neste estudo não trará benefício direto para o participante. É esperada a produção de conhecimento científico, a médio e longo prazo, que pode servir para a melhoria e proposição de novos métodos de processamento de sinais para aparelhos auditivos e implantes cocleares.

Desconfortos e riscos: Os possíveis riscos ou desconfortos decorrentes de sua participação na pesquisa são: o desconforto acústico (devido ao volume sonoro elevado), o desconforto mecânico (devido ao uso de moldes auriculares, fones de ouvido, cabo ou interface para implante coclear), a fadiga (devido ao tempo de realização do procedimento) e a claustrofobia (quando o procedimento for realizado em ambiente fechado). O desconforto acústico é evitado através do controle (diminuição) de intensidade sonora que poderá ser realizado a qualquer momento. O desconforto mecânico é evitado pelo reposicionamento ou retirada, a qualquer momento, dos moldes auriculares, fones de ouvido, cabo de ligação ou interface para implante coclear. A fadiga é minimizada pela limitação do tempo total do procedimento e pela possibilidade de pausa a

qualquer momento ou de interrupção precoce. O sentimento de claustrofobia é evitado pela saída do ambiente fechado (quando utilizado) a qualquer momento.

Outras informações: Não haverá ônus ou bônus financeiro em nenhum momento decorrente da participação nessa pesquisa. Você receberá respostas e esclarecimentos a todas as perguntas e dúvidas sobre os procedimentos realizados e assuntos relacionados a essa pesquisa por meio de contato com o pesquisador, que assume o compromisso de proporcionar informações atualizadas sobre o estudo. 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. Os participantes terão garantia ao ressarcimento de eventuais despesas, como também garantia à 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: **marcio.costa@ufsc.br**. O endereço do Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal de Santa Catarina é: Prédio Reitoria II, Rua Desembargador Vitor Lima, número 222, sala 401, Trindade, Florianópolis-SC, CEP 88040-400, telefone **(48) 3721-6094**, e-mail: **cep.propesq@contato.ufsc.br**.

O Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal de Santa Catarina é um órgão colegiado interdisciplinar, deliberativo, consultivo e educativo, vinculado à Universidade Federal de Santa Catarina, mas independente na tomada de decisões, criado para defender os interesses dos participantes da pesquisa em sua integridade e dignidade e para contribuir no desenvolvimento da pesquisa dentro de padrões éticos.

CONSENTIMENTO PÓS-INFORMADO

Eu, _____, portador(a) do CPF

_____, concordo em participar da pesquisa “Estudo sobre Métodos de Processamento de Sinais para Aparelhos Auditivos e Implantes Cocleares”, desde que seja mantido o sigilo de minha identificação, conforme as normas do Comitê de Ética em Pesquisa com Seres Humanos dessa Universidade. A minha participação é voluntária podendo ser cancelada 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 via desse termo, assinada pelo pesquisador.

Florianópolis, ____ de _____ de 20__.

Assinatura do participante

Assinatura do Pesquisador
Principal/Responsável Prof. Dr. Márcio
Holsbach Costa
RG: 5.808.610 SSP/SC

Elaborado com base na Resolução 466/2012 do CNS.