

Uma Comparação entre Máscaras Tempo-frequência para Redução de Ruído em Implantes Cocleares

Rafael Attili Chiea, Márcio Holsbach Costa e Guillaume Barrault

Resumo—Técnicas de filtragem no domínio tempo-frequência têm sido utilizadas com relativo sucesso para a redução de ruído em implantes cocleares. Entre as estratégias mais investigadas destacam-se a máscara Binária e o filtro de Wiener, que, entretanto, apresentam limitada capacidade de conformação. Para superar esse problema, alternativas como o filtro paramétrico de Wiener e variações foram propostas. Este trabalho apresenta uma comparação de desempenho entre quatro máscaras tempo-frequência em termos de inteligibilidade e qualidade da fala. Simulações numéricas foram realizadas e critérios objetivos correlacionados à percepção psicoacústica de usuários de implantes cocleares foram analisados. Os resultados indicam que as máscaras paramétricas são mais adequadas para esse tipo de aplicação, apesar de ainda não existirem estratégias bem definidas para a determinação de seus parâmetros.

Palavras-Chave — Máscaras tempo-frequência, Implantes cocleares, Filtro de Wiener, Máscara Binária.

Abstract— Time-frequency filtering techniques have been used for noise reduction in cochlear implants with relative success. The Binary Mask and the Wiener Filter are among the most applied strategies; however, they have limited conformation capability. In order to overcome this issue, alternatives have been proposed, such as the parametric Wiener Filter and variations. This work presents a performance comparison between four time-frequency masks in terms of speech intelligibility and quality. Numerical simulations were performed and objective measures correlated to the psychoacoustic perception of cochlear-implant users' were analyzed. Results indicate that parametric masks are more appropriate to this specific application, even though there are not well defined methods to determine parameters' values.

Keywords — Time-Frequency masks, Cochlear implants, Wiener Filter, Binary Mask.

I. INTRODUÇÃO

Implantes cocleares (IC) são próteses auditivas implantáveis que permitem a restauração da audição por pessoas com perda auditiva severa a profunda ou surdez total [1]. Atualmente, são consideradas como as interfaces máquina-cérebro mais bem-sucedidas [2].

Diferentemente dos aparelhos auditivos, que têm como saída um sinal acústico, os implantes cocleares estimulam diretamente o nervo auditivo com corrente elétrica aplicada através de eletrodos posicionados no interior da cóclea [1] [2].

Apesar das enormes limitações na resolução dos sons transmitidos ao usuário do implante, a inteligibilidade da fala se aproxima da encontrada em ouvintes normais. Essa observação pode ser verificada em uma série de testes de audição [3]. Entretanto, na presença de ruído e reverberação, esse desempenho diminui significativamente. Segundo [4], em condições de silêncio absoluto a inteligibilidade das palavras

em uma conversação normal é de aproximadamente 80%, mas diminui para cerca de 20% para uma razão sinal-ruído (SNR) de 5 dB. Em [5], experimentos psicoacústicos resultaram em inteligibilidade média de 80%, 60%, 40% e 10%, para as condições de silêncio e SNRs de 10 dB, 5 dB, e 0 dB, respectivamente. Desta forma, a incorporação de técnicas de redução de ruído na cadeia de processamento do implante coclear é importante para permitir a conversação em ambientes ruidosos.

Uma das principais formas de se realizar supressão de ruído em implantes cocleares é a utilização de máscaras tempo-frequência [6] [7]. Entre as principais máscaras tempo-frequência apresentadas na literatura científica, destacam-se a Máscara Binária (BM) [8], o Filtro de Wiener (WF) [9] e o Filtro Paramétrico de Wiener (WP) [10]. Esse último foi desenvolvido como um aprimoramento heurístico do WF através da introdução de dois parâmetros que permitem o ajuste da função de supressão.

Recentemente, foi demonstrado que uma máscara paramétrica alternativa ao WF, denominada de Máscara conformável (CM) é a solução ótima de uma função custo que relaciona distorção de fala com redução de ruído [11]. Este trabalho expande os resultados apresentados em [11] introduzindo critérios objetivos de inteligibilidade adequados para aplicações em implantes cocleares.

Na seção II é realizada uma revisão sobre máscaras tempo-frequência e na seção III são apresentados os critérios objetivos de inteligibilidade e qualidade, como também os métodos para simulação e avaliação. A seção IV apresenta os resultados obtidos e a seção V uma discussão à luz das requisições dos requerimentos de sistemas para implante coclear. Finalizando, a seção VI apresenta as conclusões deste trabalho.

II. MÁSCARAS TEMPO-FREQUÊNCIA

Seja um sinal de fala de interesse $x[n]$ contaminado por um ruído aditivo $v[n]$ descrito por $y[n] = x[n] + v[n]$. Considera-se que $x[n]$ e $v[n]$ são individualmente não-observáveis, além de decorrelacionados entre si. Tomando-se a Transformada de Fourier de tempo curto (STFT) de N -bins de $y[n]$, tem-se:

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

em que k e λ são, respectivamente, os índices da banda de frequência (*bin*) e da janela de tempo; e $Y(k, \lambda)$, $X(k, \lambda)$ e $V(k, \lambda)$ são respectivamente as STFTs de $y[n]$, $x[n]$ e $v[n]$.

A técnica de mascaramento tempo-frequência para supressão do ruído consiste na multiplicação (a cada unidade tempo-frequência (k, λ)) do sinal contaminado $Y(k, \lambda)$ por uma máscara de ganhos $M(k, \lambda)$, resultando em uma estimativa do sinal de fala $\hat{X}(k, \lambda)$, ou seja,

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

Rafael Attili Chiea, e Márcio Holsbach Costa Universidade Federal de Santa Catarina (UFSC), e-mails: rafaechiea@gmail.com, costa@eel.ufsc.br. Guillaume Barrault, Wavetech, e-mail: guillaume@wavetech-st.com. Este trabalho foi parcialmente financiado por CAPES e CNPq, projetos 304867/2015-2 e 315020/2018-0.

Após o processo de filtragem, o sinal estimado é reconstituído no domínio do tempo através da transformada de Fourier (TF) inversa e de uma estratégia de sobreposição e soma (*overlap-and-add*) [12].

Diferentes abordagens podem ser utilizadas para a definição da máscara de supressão $M(k, \lambda)$, como, por exemplo, um critério objetivo, ou uma abordagem heurística. Em geral, as máscaras possuem valores no intervalo $[0,1]$, definidos em função da razão sinal-ruído (SNR) *a priori* associada a cada banda de frequência k e janela temporal λ . A SNR é definida por $\xi(k, \lambda) = \sigma_x^2(k, \lambda) / \sigma_v^2(k, \lambda)$, em que $\sigma_x^2(k, \lambda) = E\{|X(k, \lambda)|^2\}$ e $\sigma_v^2(k, \lambda) = E\{|V(k, \lambda)|^2\}$ são, respectivamente, as densidades espectrais de potência da fala e do ruído; $E\{\cdot\}$ é o valor esperado e $|\cdot|$ representa o módulo. As três máscaras mais utilizadas são descritas a seguir.

A. Máscara Binária

A função de supressão mais simples é a Máscara Binária (BM), inicialmente definida em [8]:

$$M(k, \lambda) = B(k, \lambda) = \begin{cases} 1 & : \xi(k, \lambda) \geq \mu_0 \\ 0 & : \xi(k, \lambda) < \mu_0 \end{cases}, \quad (3)$$

em que, μ_0 é um limiar de decisão, geralmente igual a 0 dB. Devido à descontinuidade, a BM é classificada como uma máscara *dura*.

B. Filtro de Wiener

Máscaras *suaves* são caracterizadas por funções de supressão com transição suave entre os valores extremos $[0,1]$. A máscara suave mais popular é o Filtro de Wiener (WF) [9], cuja função de ganho é dada por:

$$M(k, \lambda) = W(k, \lambda) = \frac{\xi(k, \lambda)}{\xi(k, \lambda) + 1}. \quad (4)$$

O WF é a solução ótima que minimiza o erro quadrático médio (MSE) entre as STFTs dos sinais desejado e estimado, $MSE = E\{|\hat{X}(k, \lambda) - X(k, \lambda)|^2\}$.

C. Filtro paramétrico de Wiener

O filtro paramétrico de Wiener (WP) é caracterizado por uma função de supressão definida heurísticamente [10], com a introdução de dois parâmetros β e η no WF convencional, permitindo o seu ajuste. O WP é dado por:

$$M(k, \lambda) = W_p(k, \lambda) = \left(\frac{\xi(k, \lambda)}{\xi(k, \lambda) + \eta} \right)^\beta \quad (5)$$

Para o caso particular em que $\beta = 1$ e $\eta = 1$ o WP é equivalente ao WF.

D. Máscara paramétrica conformável

Uma forma alternativa de máscara paramétrica conformável (CM) é dada por:

$$H(k, \lambda) = \frac{\xi(k, \lambda)^\gamma}{\xi(k, \lambda)^\gamma + \mu^\gamma}, \quad (6)$$

em que $\mu > 0$ e $\gamma > \frac{1}{2}$ são parâmetros que permitem o ajuste, respectivamente, do limiar e da inclinação da transição entre os valores extremos. A Equação (6) foi descrita originalmente em [13], de forma alternativa à Equação (5) através de um raciocínio puramente heurístico. Entretanto, em [11], foi demonstrado que a CM é a solução ótima que minimiza uma função custo dada pela soma ponderada de potências das

distorções da fala e do ruído.

A CM tem como vantagem uma maior maleabilidade da função de supressão em relação ao WP, permitindo seu ajuste inclusive na forma de máscaras duras [11]. Para $\mu = 1$ e $\gamma = 1$, a CM equivale ao WF, enquanto que para $\mu = \mu_0$ e $\gamma \rightarrow \infty$, a CM tende à BM.

III. SIMULAÇÕES NUMÉRICAS

Simulações computacionais foram realizadas para avaliar o desempenho da máscara CM em comparação com as máscaras BM (com $\mu_0 = 0$ dB), WF e WP.

Todas as 720 frases das listas foneticamente balanceadas do *IEEE corpus* [14] foram contaminadas com ruído aditivo para 3 valores diferentes de SNR: -10 dB, 0 dB e 5 dB.

Foram utilizados dois tipos diferentes de ruído: um áudio gravado em um refeitório (*cafeteria babble noise*) [9] e um áudio gravado dentro de um vagão de trem em movimento (*inside train noise*) [9]. Esses sinais possuem longa duração, tendo sido recortados aleatoriamente para contaminação aditiva artificial da fala.

Os sinais contaminados foram divididos em dois grupos: um conjunto de treinamento e um conjunto de teste. O conjunto de treinamento consistiu de 6 subgrupos, um para cada tipo de ruído e SNR estudados, contendo 100 arquivos de áudio cada um. Por sua vez, o conjunto de teste foi composto por 6 subgrupos com 620 áudios cada.

Todos os sinais foram amostrados a 16 kHz e foram transformados para o domínio frequência usando a STFT, com 512 pontos, janela de Hanning de 20ms e sobreposição de 50%. Após o processamento por cada uma das máscaras, os sinais resultantes foram transformados de volta para o domínio do tempo através do método *weighted-overlap-and-add* [12]. Os sinais de fala e ruído foram processados separadamente de forma a calcular os ganhos das máscaras ideais.

Dois critérios objetivos foram utilizados para a avaliação de desempenho: (a) *wideband Perceptual Evaluation of Speech Quality* (WPESQ), para avaliação da qualidade [15]; e a (b) *Speech-to-Reverberation Modulation Energy Ratio – Cochlear Implant* (SRMR-CI), para avaliação de inteligibilidade [16]. A SRMR-CI foi especificamente desenvolvida para apresentar alta correlação com a inteligibilidade de portadores de implantes cocleares em experimentos psicoacústicos e pode ser mapeada em valores de inteligibilidade da seguinte maneira [16]:

$$I_{\%} = \frac{1}{1 + e^{a_1 \text{SRMR-CI}_{norm} + a_2}} \times 100\%, \quad (7)$$

em que $I_{\%}$ é o percentual de inteligibilidade; $a_1 = 12,1742$ e $a_2 = -7,4535$ são coeficientes de ajuste [17]; $\text{SRMR-CI}_{norm} = \text{SRMR-CI} / \text{SRMR-CI}_{clean}$; e SRMR-CI_{clean} é o valor da métrica para a fala original (sem contaminação).

Valores de inteligibilidade e qualidade referentes ao processamento com a BM, WF, WP e CM foram calculados e comparados entre si usando análise de variância (ANOVA), com nível de significância especificado em 0,05. Na sequência foi realizada uma análise de comparação múltipla utilizando-se o teste de Tukey, um teste *post hoc* adequado ao cenário deste estudo [18], que permite a comparação pareada.

Para cada condição de SNR e tipo de ruído, o conjunto de treinamento foi utilizado para a obtenção dos melhores parâmetros (γ e μ para a CM; β e η para o WP), de acordo com um critério objetivo: (1) maximização da métrica de qualidade

(WPESQ); (2) maximização da métrica de inteligibilidade (I_0); ou (3) maximização de uma combinação arbitrária das métricas de qualidade e inteligibilidade.

Os melhores conjuntos de parâmetros para cada um dos sinais contaminados foram obtidos dentro de uma grade arbitrária de possibilidades. Para a CM, a grade de valores de parâmetros aborda todas as combinações de μ entre -60 dB a 60 dB com passos de 5 dB, e γ variando de 0,5 a 1 em 6 passos, como também de 1,25 a 100 em 6 passos igualmente espaçados em escala logarítmica, resultando em 300 combinações diferentes. A grade de parâmetros inclui o WF ($\mu = 1$ e $\gamma = 1$) e uma aproximação da BM ($\mu = 1$ e $\gamma = 100$). O mesmo procedimento foi realizado com o WP, com β variando de 0,2 a 1 e de 1,25 a 400 em, respectivamente, 4 e 6 passos igualmente espaçados em escala logarítmica. O parâmetro η foi variado entre -35 dB e 25 dB em passos de 2,5 dB, totalizando 250 máscaras.

Para cada tipo de ruído e SNR, utilizando-se os melhores conjuntos de parâmetros, foi calculada a mediana dos resultados referentes a cada subgrupo. Em sequência, esse conjunto de medianas foi mapeado para os valores mais próximos nas respectivas grades. Os parâmetros resultantes desse mapeamento foram utilizados na avaliação do conjunto de teste.

IV. RESULTADOS

A seguir são apresentados os resultados relativos aos três experimentos realizados: maximização de qualidade; maximização de inteligibilidade; e maximização da combinação arbitrária de qualidade e inteligibilidade.

A. Experimento I – Maximização de qualidade

Nesse experimento, foram utilizados os parâmetros que maximizam a qualidade dos sinais processados pelas máscaras CM e WP de acordo com o critério objetivo WPESQ. Os melhores conjuntos de parâmetros obtidos para cada tipo de ruído e SNR são apresentados na Tabela I, enquanto que os valores médios de WPESQ obtidos por cada máscara com o conjunto de sinais de teste são apresentados na Tabela II.

TABELA I – PARÂMETROS DAS MÁSCARAS WP E CM QUE MAXIMIZAM A MÉTRICA WPESQ.

Ruído	SNR	WP		CM	
		β	η [dB]	γ	μ [dB]
cafeteria babble	-10 dB	0,72	7,5	0,66	10
	0 dB	0,53	5	0,66	5
	5 dB	0,53	5	0,66	5
inside train	-10 dB	0,53	15	0,57	25
	0 dB	0,53	12,5	0,50	20
	5 dB	0,53	10	0,57	15

A Fig. 1 apresenta diagramas de caixa bidimensionais com os desempenhos dos métodos analisados (BM, WF, WP e CM) no conjunto de teste, assumindo ruído *cafeteria babble* e uma SNR = -10 dB. O eixo horizontal quantifica a qualidade em unidades WPESQ e o vertical o percentual de inteligibilidade. Nesse tipo de apresentação, o desempenho geral aumenta com a distância da origem. Os quadrados indicam o primeiro e terceiro quartis, enquanto que os sinais de cruz (+), na área de destaque, representam as medianas das métricas de qualidade e inteligibilidade. Valores discrepantes (*outliers*) são omitidos

por clareza. Todas as distribuições obtidas (BM, WF, WP e CM) para SNR = -10 dB apresentam diferença estatística (Tabela II).

TABELA II – VALORES MÉDIOS DE WPESQ OBTIDOS POR CADA MÁSCARA NO CONJUNTO DE TESTE. OS PARÂMETROS DAS MÁSCARAS WP E CM FORAM OTIMIZADOS DE ACORDO COM O CRITÉRIO WPESQ. PARA CADA LINHA, OS MAIORES VALORES DE WPESQ ESTÃO EM NEGRITO. O SÍMBOLO § INDICA QUE OS VALORES NÃO SÃO ESTATISTICAMENTE DISTINTOS ($P > 0,05$).

Ruído	SNR	BM	WF	WP	CM
cafeteria Babble	-10 dB	1,14	1,36§	1,37§	1,41
	0 dB	1,58	2,00	2,06§	2,08§
	5 dB	2,07	2,57	2,63§	2,65§
inside train	-10 dB	1,18	1,42	1,55§	1,56§
	0 dB	1,56	1,98	2,12§	2,14§
	5 dB	2,04	2,49	2,59§	2,62§

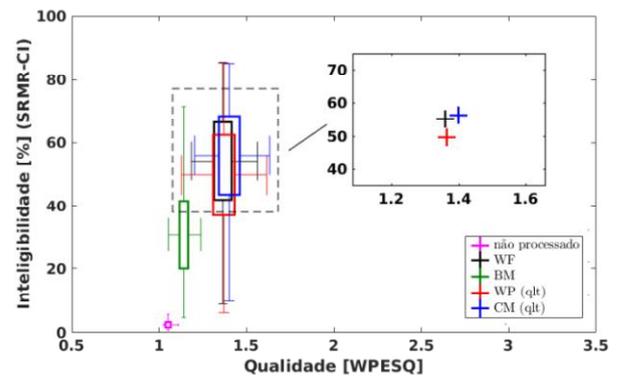


Fig. 1 – Diagramas de caixa bidimensionais de qualidade (WPESQ) e inteligibilidade (I_0) para: conjunto de teste, ruído *cafeteria babble* e SNR = -10 dB. Os parâmetros das máscaras WP e CM foram otimizados para qualidade. BM em verde, WF em preto, WP em vermelho, CM em azul e não-processado em magenta.

TABELA III – PARÂMETROS DAS MÁSCARAS WP E CM QUE MAXIMIZAM I_0

Ruído	SNR	WP		CM	
		β	η [dB]	γ	μ [dB]
cafeteria babble	-10 dB	0,53	2,5	0,66	0
	0 dB	0,72	2,5	0,76	5
	5 dB	1,25	-2,5	0,87	5
inside train	-10 dB	0,53	12,5	0,57	20
	0 dB	0,53	10	0,66	10
	5 dB	0,72	5	0,66	5

TABELA IV – VALORES MÉDIOS DE I_0 OBTIDOS POR CADA MÁSCARA NO CONJUNTO DE TESTE. OS PARÂMETROS DAS MÁSCARAS WP E CM FORAM OTIMIZADOS DE ACORDO COM O CRITÉRIO DE INTELIGIBILIDADE. PARA CADA LINHA, OS MAIORES VALORES DE INTELIGIBILIDADE ESTÃO EM NEGRITO. O SÍMBOLO § INDICA QUE OS VALORES NÃO SÃO ESTATISTICAMENTE DISTINTOS ($P > 0,05$).

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

B. Experimento II – Maximização de inteligibilidade

Nesse segundo experimento, foram utilizados os parâmetros que maximizam a inteligibilidade dos sinais processados das máscaras CM e WP de acordo com o critério objetivo I_0 . Os melhores conjuntos de parâmetros obtidos para cada tipo de ruído e SNR são apresentados na Tabela III, enquanto que os valores médios de I_0 obtidos por cada máscara com o conjunto de sinais de teste são apresentados na Tabela IV. Os diagramas de caixa na Fig. 2 são referentes a um caso particular da Tabela IV: ruído *cafeteria babble* com SNR = -10 dB.

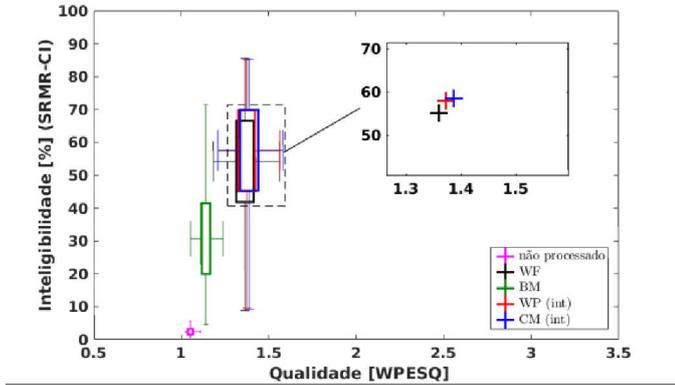


Fig. 2 – Diagramas de caixas bidimensionais de qualidade (WPESQ) e inteligibilidade (I_0) para: conjunto de teste, ruído *cafeteria babble* e SNR = -10 dB. Os parâmetros das máscaras WP e CM foram otimizados para inteligibilidade. BM em verde, WF em preto, WP em vermelho, CM em azul e não-processado em magenta.

C. Experimento III – Maximização conjunta de inteligibilidade e qualidade

Diferentemente dos experimentos anteriores, neste, os parâmetros das máscaras WP e CM são otimizados de acordo com uma métrica arbitrária de desempenho global, considerando-se tanto os aspectos de qualidade quanto de inteligibilidade. O desempenho global é calculado como sendo a distância quadrática normalizada, dada por:

$$d_p = \left(\frac{I_0}{100} \right)^2 + \left(\frac{\text{PESQ}}{5} \right)^2, \quad (8)$$

em que I_0 é calculado de acordo com a equação (7). Vale ressaltar que não há nenhuma evidência preliminar de que haja correlação da equação (8) com a satisfação global humana.

Os parâmetros das máscaras WP e CM que maximizam a equação (8) são apresentados na Tabela V. Por sua vez, as Tabelas VI e VII mostram, respectivamente, as médias de I_0 e WPESQ obtidas com esses parâmetros.

A Fig. 3 apresenta diagramas de caixa para o caso de ruído *cafeteria babble* com SNR = -10 dB.

V. DISCUSSÃO

A análise global dos resultados obtidos para as SNRs de 0 dB e 5 dB indica um desempenho aproximadamente similar para as três máscaras suaves analisadas. Esse achado faz sentido uma vez que sinais com pequenos níveis de contaminação (maior SNR) resultam em máscaras com ganhos semelhantes e próximos da unidade. A máscara binária consistentemente apresentou índices de qualidade e inteligibilidade menores que as demais. Com a redução da SNR, as funções de supressão operam mais próximas da região de transição, evidenciando as diferenças.

TABELA V – PARÂMETROS DAS MÁSCARAS WP E CM QUE MAXIMIZAM O CRITÉRIO DE DESEMPENHO GLOBAL.

Ruído	SNR	WP		CM	
		β	η [dB]	γ	μ [dB]
<i>cafeteria babble</i>	-10 dB	0,53	0	0,66	0
	0 dB	0,72	5	0,66	5
	5 dB	0,53	5	0,66	5
<i>inside train</i>	-10 dB	0,53	10	0,57	20
	0 dB	0,53	10	0,57	15
	5 dB	0,53	5	0,57	10

TABELA VI – VALORES MÉDIOS DE I_0 OBTIDOS POR CADA MÁSCARA NO CONJUNTO DE TESTE. OS PARÂMETROS DAS MÁSCARAS WP E CM FORAM OTIMIZADOS DE ACORDO COM O CRITÉRIO DE DESEMPENHO GLOBAL. PARA CADA LINHA, OS MAIORES VALORES DE I_0 ESTÃO EM NEGRITO. O SÍMBOLO § INDICA QUE OS VALORES NÃO SÃO ESTATISTICAMENTE DISTINTOS ($p > 0,05$).

Ruído	SNR	WF	BM	CM	PW
<i>cafeteria babble</i>	-10 dB	52,55	31,47	56,91 §	56,41 §
	0 dB	78,39 §	71,48	78,77 §	78,77 §
	5 dB	84,93 §	83,23	84,87 §	84,93 §
<i>inside train</i>	-10 dB	58,34	36,50	65,03 §	62,89 §
	0 dB	78,22 §	69,16	79,69 §	79,47 §
	5 dB	84,22 §	81,04	84,54 §	84,60 §

TABELA VII – VALORES MÉDIOS DE WPESQ OBTIDOS POR CADA MÁSCARA NO CONJUNTO DE TESTE. OS PARÂMETROS DAS MÁSCARAS WP E CM FORAM OTIMIZADOS DE ACORDO COM O CRITÉRIO GLOBAL DE DESEMPENHO. PARA CADA LINHA, OS MAIORES VALORES DE WPESQ ESTÃO EM NEGRITO. O SÍMBOLO § INDICA QUE OS VALORES NÃO SÃO ESTATISTICAMENTE DISTINTOS ($p > 0,05$).

Ruído	SNR	WF	BM	CM	PW
<i>cafeteria babble</i>	-10 dB	1,364	1,141	1,396 §	1,400 §
	0 dB	1,998	1,578	2,078 §	2,024 §
	5 dB	2,565	2,068	2,650 §	2,631 §
<i>inside train</i>	-10 dB	1,418	1,178	1,538	1,502
	0 dB	1,984	1,557	2,135	2,092
	5 dB	2,487	2,041	2,600 §	2,564 §

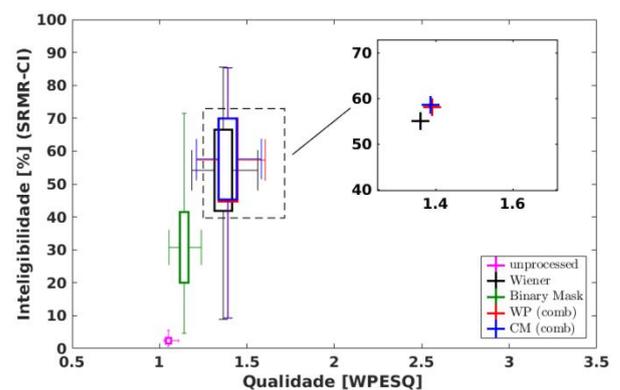


Fig. 3 – Diagramas de caixa bidimensionais de qualidade (WPESQ) e inteligibilidade (I_0) para: conjunto de teste, ruído *cafeteria babble* e SNR = -10dB. Os parâmetros das máscaras WP e CM foram otimizados para desempenho global. BM em verde, WF em preto, WP em vermelho, CM em azul e não-processado em magenta.

No Experimento I (maximização da qualidade), as máscaras WP e CM apresentam melhor desempenho em comparação às máscaras BM e WF, para todos os tipos de ruído e SNR estudados. O desempenho da máscara WP é estatisticamente equivalente ($p > 0,05$) ao da CM em todos os

casos, exceto para *cafeteria babble* com SNR = -10 dB. Esse último caso é ilustrado na Fig. 1, na qual pode-se verificar que também a inteligibilidade obtida com a CM é maior que com a WP. Entretanto, apesar do teste de Tukey indicar diferença significativa da WPESQ entre WP e CM, o desempenho dessas duas máscaras pode ser considerado similar, visto que a diferença é menor que a mínima variação perceptível (JND - *Just Noticeable Difference*) por seres humanos (WPESQ = 0,2) [19] [20].

No experimento II (maximização da inteligibilidade), os desempenhos associados à WP e à CM são similares, sendo significativamente diferentes da WF apenas para SNR = -10 dB.

O experimento III (maximização do critério de desempenho global) indica novamente similaridade (em relação à JND) entre WP e CM (Tabela VI). Para valores não-negativos de SNR, o desempenho do WF equivale estatisticamente aos das máscaras paramétricas.

De forma geral, verifica-se que a máscara CM apresenta desempenho no mínimo equivalente ao da máscara WP, frequentemente utilizada em aplicações associadas à redução de ruído em aparelhos auditivos e implantes cocleares [7] [21] [22] [23] [24]. Entretanto, a máscara CM apresenta diferenciais interessantes em relação à WP, como: o fato de ser a solução ótima de uma função custo objetiva, com clara interpretação física dos processos de distorção da fala e redução de ruído; e possuir maior flexibilidade em sua conformação, permitindo uma maior gama de funções de supressão suaves e duras [11]. Essa última característica é especialmente desejável em aplicações para implantes cocleares, devido à grande variabilidade nas preferências individuais dos usuários [6] [7].

VI. CONCLUSÃO

Neste trabalho foi realizada uma comparação de desempenho entre máscaras tempo-frequência para redução de ruído em implantes cocleares. Simulações numéricas utilizando critérios objetivos associados à inteligibilidade e à qualidade da fala percebida por usuários de implante coclear indicaram maior desempenho de estratégias paramétricas suaves em relação às máscaras consideradas duras e/ou não-configuráveis. Em especial, a máscara paramétrica conformável (CM) apresenta interessante potencial para implantes cocleares, em decorrência de sua configurabilidade, capacidade de redução de ruído e suporte teórico. O principal fator limitante para seu uso, porém, ainda é a falta de estratégias bem definidas para o cálculo de seus parâmetros.

REFERÊNCIAS

[1] J. Wouters et al., "Sound processing for better coding of monaural and binaural cues in auditory prostheses", *Proceeding of the IEEE*, v. 101, n. 9, pp. 1986-1997, 2013.

[2] J. Wouters, H. J. McDermott, T. Francart, "Sound coding in cochlear implants", *IEEE Signal Processing Magazine*, v. 32, n. February, pp. 67-80, 2015.

[3] B. S. Wilson, M. F. Dorman, "The surprising performance of present-day cochlear implants", *IEEE Transactions on Biomedical Engineering*, v. 54, n. 6, pp. 969-972, 2007.

[4] A. Hast et al., "Speech perception of elderly cochlear implant users

under different noise conditions", *Otology and Neurology*, v. 36, n. 10, pp. 1638-1643, 2015.

[5] F. Bergeron, M. Hotton, "Perception in noise with the Digisonic SP cochlear implant: Clinical trial of Saphyr processor's upgraded signal processing", *European Annals of Otorhinolaryngology, Head and Neck Diseases*, pp. 3-5, 2016.

[6] R. Koning, N. Madhu, J. Wouters, "Ideal time-frequency masking algorithms lead to different speech intelligibility and quality in normal-hearing and cochlear implant listeners", *IEEE Transactions on Biomedical Engineering*, v. 62, n. 1, pp. 331-341, 2015.

[7] S. J. Mauger, P. W. Dawson, A. A. Hersbach, "Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction", *The Journal of the Acoustical Society of America*, v. 131, n. 1, pp. 327-336, 2012.

[8] D. Wang, G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Wiley/IEEE Press, 2006.

[9] P. C. Loizou, *Speech enhancement - Theory and practice 2nd*. CRC Press, 2013.

[10] J. S. Lim, A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech", *Proceedings of the IEEE*, v. 67, n. 12, pp. 1586-1604, 1979.

[11] R. A. Chiea, M. H. Costa, G. Barrault, "New insights on the optimality of parameterized Wiener filters for speech enhancement applications" *Speech Communication*, n. 109, pp. 46-54, 2019.

[12] R. E. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 28, n. 1, pp. 99-102, 1980.

[13] M. Fontaine et al., "Explaining the parameterized Wiener filter with alpha-stable processes", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2017, pp. 51-55.

[14] IEEE recommended practice for speech quality measurements, *IEEE Transactions on Audio and Electroacoustics*, v.17 no. 3, pp. 225-246, 1969.

[15] International Telecommunication Union, P.862.2 Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs, 2007.

[16] J. F. Santos, T. H. Falk, "Updating the SRMR-CI metric for improved intelligibility prediction for cochlear implant users", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 22, n. 12, pp. 2197-2206, 2014.

[17] J. F. Santos et al., "A non-intrusive objective speech intelligibility metric tailored for cochlear implant users in complex listening environments", Université du Québec, M.Sc. Thesis, 2013.

[18] M. L. McHugh, "Multiple comparison analysis testing in ANOVA", *Biochemia Medica*, v. 21, n. 2, pp. 203-209, 2011.

[19] K. W. Wilson et al., "Speech denoising using nonnegative matrix factorization with priors", *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 4029-4032.

[20] A. Servetti, J.C. Martin, "Error tolerant MAC extension for speech communications over 802.11 WLANs", *IEEE Vehicular Technology Conference*, v. 4, 2005, pp. 2330-2334.

[21] W. He, N. Zheng, Q. Meng, "The effect of gain thresholds on speech intelligibility for statistical model based noise reduction for cochlear implants: A simulation based verification", *International Symposium on Chinese Spoken Language Processing*, 2017, pp 1-4.

[22] T. Goehring et al., "Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users", *Hearing Research*, v. 344, pp. 1-12, 2016.

[23] O. Hazrati et al., "Simultaneous suppression of noise and reverberation in cochlear implants using a ratio masking strategy", *The Journal of the Acoustical Society of America*, v. 134, n. 5, pp. 3759-65, 2013.

[24] P. W. Dawson, S. J. Mauger, A. A. Hersbach, "Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus® cochlear implant recipients", *Ear and Hearing*, v. 32, n. 3, pp. 382-390, 2011.