

Uma Máscara Tempo-frequência Conformável Assimétrica para Redução de Ruído em Aparelhos Auditivos

Rivael Pigatto, Bruno C. Bispo e Márcio H. Costa

Resumo— As máscaras tempo-frequência são uma das principais estratégias para redução de ruído em aparelhos auditivos. As máscaras encontradas na literatura possuem curvas de ganho, em função da razão sinal-ruído em decibel, simétricas em relação ao meio ganho. Este trabalho apresenta uma primeira proposta de máscara paramétrica conformável assimétrica. A máscara proposta é capaz de modelar as do estado-da-arte. O seu desempenho é avaliado em termos de qualidade sonora perceptual e inteligibilidade de sinais de fala utilizando métricas objetivas e vários tipos de ruído. Os resultados indicam que a maior maleabilidade da máscara proposta permite um leve aumento na qualidade e inteligibilidade da fala.

Palavras-Chave— Máscaras tempo-frequência, máscara assimétrica, redução de ruído, aparelhos auditivos.

Abstract— Time-frequency masks are one of the main strategies for noise reduction in hearing aids. The masks found in the literature have gain curves, as a function of the signal-to-noise ratio in decibel, symmetrical with respect to the half gain. This work presents a first proposal of asymmetric conformable parametric mask. The proposed mask is capable of modeling the state-of-the-art ones. Its performance is evaluated in terms of perceptual sound quality and intelligibility of speech signals using objective metrics and various types of noise. The results indicate that the greater malleability of the proposed mask achieves a slight increase in speech quality and intelligibility.

Keywords— Time-frequency masks, asymmetric mask, noise reduction, hearing aids.

I. INTRODUÇÃO

Segundo a Organização Mundial da Saúde, no ano de 2019, 466 milhões de pessoas, mais de 5% da população mundial, tinham perda auditiva incapacitante. Estima-se que, em 2050, serão mais de 900 milhões de pessoas [1]. E de acordo com o último censo realizado no Brasil, no ano de 2010, 9,7 milhões de cidadãos brasileiros, ou 5,1% da população nacional, apresentavam algum tipo de limitação auditiva [2].

A deficiência auditiva pode acarretar isolamento social, dificuldades profissionais e risco à segurança pessoal, tendo como consequência uma redução na qualidade de vida do indivíduo [3]. A sua compensação depende do tipo e da quantificação do grau da perda auditiva, podendo ser realizada de diversas formas, entre elas, com a utilização de aparelho auditivo ou implante coclear [3].

Os ouvintes com deficiência auditiva apresentam maior dificuldade de inteligibilidade de fala na presença de ruído [4], [5]. Apesar de aumentar a audibilidade de sinais de fala,

a capacidade dos aparelhos auditivos modernos em melhorar a inteligibilidade de fala é bastante limitada em ambiente ruidoso [4], [5]. Segundo [6], a inteligibilidade média em aparelhos auditivos é de aproximadamente 90% em condições de silêncio. Mas o desempenho cai para próximo de 65 e 10% com razões sinal-ruído (SNR) de 10 e 0 dB, respectivamente.

Similar diminuição de inteligibilidade de fala na presença de ruído ocorre em implantes cocleares. Segundo [7], a inteligibilidade média de sentenças em condição de silêncio é de aproximadamente 75%. Mas, para uma razão sinal-ruído (SNR) de 5 dB, está entre 19 e 35% dependendo do tipo de ruído e da idade do usuário. E, conforme [8], a inteligibilidade média é de aproximadamente 80, 50, 40 e 20% para condição de silêncio e SNRs de 10, 5 e 0 dB, respectivamente.

Uma das principais estratégias de redução de ruído em aparelhos auditivos faz uso de máscaras tempo-frequência [9]. A cada unidade tempo-frequência, um fator de atenuação, referido como ganho, é aplicado ao sinal de fala contaminado. Essas máscaras são caracterizadas por uma curva de ganho em função da SNR associada à unidade tempo-frequência [10]. As principais máscaras encontradas na literatura são a máscara binária [11], a máscara de Wiener [12] e suas versões restrita [12] e paramétrica [13], e a máscara paramétrica conformável [14]. Em todas elas, a curva de ganho, em função da SNR em decibel, é simétrica em relação ao meio ganho.

Com o intuito de investigar o potencial de máscaras assimétricas, este trabalho apresenta uma primeira proposta de máscara paramétrica conformável assimétrica para redução de ruído em aparelhos auditivos e avalia o seu desempenho em termos de qualidade sonora e inteligibilidade de fala. O artigo está organizado da seguinte maneira: na Seção II, as principais máscaras encontradas na literatura são apresentadas e a máscara assimétrica é proposta; a Seção III descreve os dois experimentos realizados; a Seção IV apresenta e discute os resultados obtidos; por fim, a Seção V conclui o artigo.

II. MÁSCARAS TEMPO-FREQUÊNCIA

Sejam o sinal de fala de interesse e o ruído ambiente denotados por $x(n)$ e $v(n)$, respectivamente. O sinal de fala contaminado é definido por $y(n) = x(n) + v(n)$. Assume-se que $x(n)$ e $v(n)$ são não-observáveis e descorrelacionados entre si. Neste caso, a transformada de Fourier de tempo curto (STFT) de $y(n)$ é definida como

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

onde $X(k, \lambda)$ e $V(k, \lambda)$ são as STFTs de $x(n)$ e $v(n)$, respectivamente, k é o bin frequencial e λ é o índice da janela temporal.

A técnica de mascaramento tempo-frequência para supressão de ruído consiste em, a cada unidade tempo-frequência $\{k, \lambda\}$, multiplicar $Y(k, \lambda)$ por uma máscara de ganhos $M(k, \lambda)$, resultando numa estimativa de $X(k, \lambda)$ dada por [12]

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

No domínio do tempo, a estimativa $\hat{x}(n)$ do sinal de fala de interesse é reconstruída utilizando a STFT inversa e uma estratégia de sobreposição e soma (*overlap-and-add*) [15].

As máscaras podem ser definidas utilizando critérios objetivos ou heurísticas. Em geral, $0 \leq M(k, \lambda) \leq 1$ e $M(k, \lambda)$ é uma função da razão sinal-ruído associada a λ -ésima janela e k -ésimo bin frequencial, a qual é definida como

$$\xi(k, \lambda) = \frac{S_x(k, \lambda)}{S_v(k, \lambda)}, \quad (3)$$

onde $S_x(k, \lambda) = E\{|X(k, \lambda)|^2\}$ e $S_v(k, \lambda) = E\{|V(k, \lambda)|^2\}$ são as densidades espectrais de potência de $x(n)$ e $v(n)$, respectivamente, e $E\{\cdot\}$ é o operador valor esperado. A SNR na escala decibel é dada por $\xi_{dB}(k, \lambda) = 10 \log_{10} \xi(k, \lambda)$.

A. Máscara Binária

A máscara tempo-frequência mais simples é a máscara binária (BM), a qual é definida como [11]

$$M(k, \lambda) = \begin{cases} 1, & \xi(k, \lambda) \geq \xi_0, \\ 0, & \xi(k, \lambda) < \xi_0, \end{cases} \quad (4)$$

onde ξ_0 é uma constante geralmente igual a 0 dB. Devido à descontinuidade, a BM é classificada como uma máscara dura. Por outro lado, máscaras suaves são caracterizadas por funções de ganho com transição suave entre seus valores extremos.

B. Máscara de Wiener

A principal máscara suave é a máscara de Wiener (WM), a qual é definida como [12]

$$M(k, \lambda) = \frac{\xi(k, \lambda)}{\xi(k, \lambda) + 1}. \quad (5)$$

A WM é o filtro ótimo que minimiza, no domínio da frequência, o erro quadrático médio da estimativa do sinal de fala de interesse, isto é, $MSE(k, \lambda) = E\{|\hat{X}(k, \lambda) - X(k, \lambda)|^2\}$.

C. Máscara Restrita de Wiener

Uma suavização na WM resulta na máscara restrita de Wiener (WR) definida como [12]

$$M(k, \lambda) = \frac{\sqrt{\xi(k, \lambda)}}{\sqrt{\xi(k, \lambda)} + 1}. \quad (6)$$

A WR é o filtro que iguala as distorções relativas de $v(n)$ e $x(n)$.

D. Máscara Paramétrica de Wiener

A máscara paramétrica de Wiener (WP) é uma extensão heurística da WM realizada com a introdução de dois parâmetros na função de ganho, sendo definida como [13]

$$M(k, \lambda) = \left(\frac{\xi(k, \lambda)}{\xi(k, \lambda) + \eta} \right)^\beta. \quad (7)$$

Quando $\beta = \eta = 1$, a WP equivale a WM.

E. Máscara Paramétrica Conformável

Em [14], uma máscara paramétrica conformável (CM) foi proposta como uma extensão heurística da WP e definida como

$$M(k, \lambda) = \frac{\xi^\gamma(k, \lambda)}{\xi^\gamma(k, \lambda) + \mu^\gamma}, \quad (8)$$

onde $\mu > 0$ e $\gamma > 0,5$ são parâmetros que permitem o ajuste do deslocamento lateral e da inclinação da máscara, respectivamente. Recentemente, em [9], demonstrou-se que a CM é o filtro ótimo que minimiza uma função custo dada pela soma ponderada de potências das distorções de $x(n)$ e $v(n)$.

A CM apresenta uma maior maleabilidade na função de ganho em relação ao WP, permitindo seu ajuste inclusive na forma de máscaras duras. Para $\mu = \gamma = 1$, a CM equivale à WM. Para $\mu = 0,5$ e $\gamma = 1$, ela equivale à WR. E para $\mu = 10 \log_{10} \xi_0$ e $\gamma \rightarrow \infty$, ela equivale à BM.

F. Proposta de Máscara Paramétrica Conformável Assimétrica

Todas as máscaras do estado-da-arte, descritas nas subseções anteriores, possuem uma característica em comum: simetria em relação ao ganho 0,5. Com o objetivo de remover essa restrição e investigar o potencial de máscaras assimétricas, este trabalho propõe uma máscara paramétrica conformável assimétrica (ACM) definida como

$$M(k, \lambda) = \begin{cases} \frac{1}{1 + e^{-\beta_1 [\xi_{dB}(k, \lambda) - \xi_0]}}, & \xi_{dB}(k, \lambda) < \xi_0, \\ \frac{1}{1 + e^{-\beta_2 [\xi_{dB}(k, \lambda) - \xi_0]}}, & \xi_{dB}(k, \lambda) \geq \xi_0, \end{cases} \quad (9)$$

onde os parâmetros positivos β_1 e β_2 controlam o comportamento exponencial da máscara para ganhos inferiores e superiores a 0,5, respectivamente, e ξ_0 representa o deslocamento lateral (em dB) da curva. Para $\beta_1 = \beta_2 = 0,1 \gamma \ln 10$ e $\xi_0 = 10^{0,1\mu}$, a máscara equivale à CM. Para $\beta_1 = \beta_2 = 0,1 \ln 10$ e $\xi_0 = 0$, ela equivale à WM. Para $\beta_1 = \beta_2 = 0,1 \ln \sqrt{10}$ e $\xi_0 = 0$, ela equivale à WR. E para $\beta_1 = \beta_2 \rightarrow \infty$, ela equivale à BM. A Figura 1 ilustra a máscara proposta para dois conjuntos de parâmetros juntamente com as máscaras BM, WM e WR.

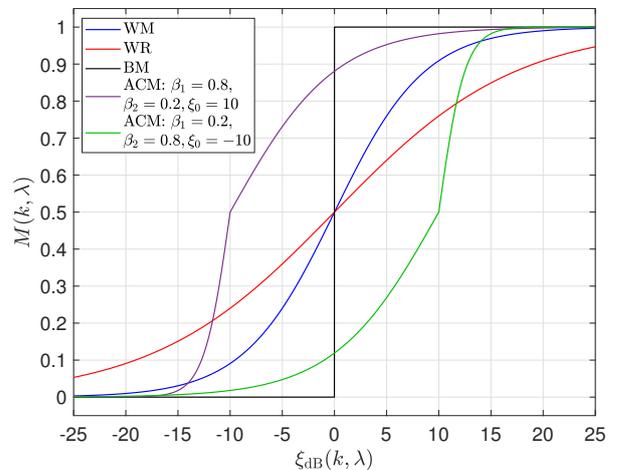


Fig. 1. Ilustração das curvas de ganho das seguintes máscaras: WM (azul), WR (vermelho), BM (preto) e duas configurações da ACM (verde e violeta).

III. CONFIGURAÇÃO DAS SIMULAÇÕES

Essa seção descreve a configuração dos dois experimentos realizados para avaliar a capacidade da máscara ACM na redução de ruído. O primeiro investiga a qualidade perceptual de $\hat{x}(n)$. O segundo examina a inteligibilidade de $\hat{x}(n)$.

A. Sinais de Fala e Ruído

São utilizados 10 sinais de fala gravados por uma mulher, do banco de dados descrito em [16], amostrados em 16 kHz. Eles foram contaminados com 5 tipos de ruído: ruído de refeitório (*cafeteria babble noise*) [12]; ruído dentro de vagão de trem em movimento (*inside train noise*) [12]; ruído de fala-sobre-fala (*speech-in-speech*) usando o sinal ISTS [17]; ruído com espectro de fala (*speech-shaped noise*) usando o sinal ICRA [18]; e ruído branco Gaussiano. A contaminação foi realizada com 3 níveis de SNR: -10 dB, 0 dB e 5 dB.

B. Aplicação das Máscaras

As STFTs $Y(k, \lambda)$, $X(k, \lambda)$ e $V(k, \lambda)$ são calculadas utilizando uma janela de Hamming com duração de 20 ms, sobreposição de 50% e uma transformada discreta de Fourier (DFT) de 320 pontos. Para cada janela λ , uma estimativa de $\xi(k, \lambda)$ é obtida como

$$\hat{\xi}(k, \lambda) = \frac{|X(k, \lambda)|^2}{|V(k, \lambda)|^2}, \quad k = 1, 2, \dots, 320, \quad (10)$$

e utilizada para computar a curva de ganho $M(k, \lambda)$ das diferentes máscaras. A curva $M(k, \lambda)$ é por sua vez aplicada a $Y(k, \lambda)$ conforme (2), obtendo a estimativa $\hat{X}(k, \lambda)$. Por fim, a estimativa $\hat{x}(n)$ do sinal de fala limpo é construída utilizando as STFTs inversas de $\hat{X}(k, \lambda)$ e *overlap-and-add*.

Três máscaras são aplicadas: WM, CM e ACM. Na ACM, o parâmetro β_2 foi analisado no intervalo $[0,01; 0,25]$ com passos de 0,01 e no intervalo $[0,3; 1]$ com passo de 0,05. Os parâmetros β_1 e ξ_0 foram analisados nos intervalos $[0,01; 0,25]$ e $[-15; 15]$ dB com passos de 0,01 e 1 dB, respectivamente. A máscara CM foi implementada através da ACM com $\beta_1 = \beta_2$, com intervalo descrito de β_2 e ξ_0 .

C. Métricas de Avaliação

a) *W-PESQ*: O *W-PESQ* (*Wideband Perceptual Evaluation of Speech Quality*) é um algoritmo para avaliação objetiva da qualidade de sinais de fala amostrados a 16kHz [19], [20]. Ele compara representações psicoacústicas de um sinal de fala possivelmente degradado e sua correspondente referência não corrompida [21]. A pontuação bruta do *W-PESQ* pode ser mapeada para a escala 1-5 da opinião média (MOS), resultando na pontuação MOS-LQO [22]. A correspondência entre a escala de cinco pontos e a classificação da categoria de degradação (DCR) é mostrada na Tabela I. No entanto, o máximo MOS-LQO fornecida pelo *W-PESQ* é 4,644 quando os sinais de referência e degradados são idênticos.

Neste trabalho, o *W-PESQ* é utilizado para avaliar o desempenho das máscaras em relação à qualidade sonora dos sinais processados. Para isso, os sinais $x(n)$ e $\hat{x}(n)$ são utilizados como os sinais de referência e degradado, respectivamente.

TABELA I
CORRESPONDÊNCIA ENTRE PONTUAÇÃO E DCR.

Pontuação	DCR - Qualidade da fala
5	Degração inaudível
4	Degração inaudível, mas não incômoda
3	Degração pouco incômoda
2	Degração incômoda
1	Degração muito incômoda

b) *NCM*: O *NCM* (*Normalized Covariance Metric*) é uma métrica para avaliação objetiva da inteligibilidade de sinais de fala [12], [23]. Ela é baseada na covariância entre os envelopes temporais de sub-bandas de um sinal de fala possivelmente degradado e sua correspondente referência não corrompida [12]. A sua pontuação varia entre 0 e 1, onde valores mais altos indicam maior inteligibilidade [12].

Neste trabalho, o *NCM* é utilizado para analisar o desempenho das máscaras em relação à inteligibilidade dos sinais processados. Para isso, os sinais $x(n)$ e $\hat{x}(n)$ são utilizados como os sinais de referência e degradado, respectivamente.

IV. RESULTADOS

Esta seção apresenta os resultados dos dois experimentos realizados. Os sinais, métricas de avaliação e procedimentos descritos na Seção III são empregados. Os resultados são apresentados e analisados de forma global, considerando todos os tipos de ruídos, para cada SNR.

A. Experimento de Maximização da Qualidade

Neste experimento, buscou-se por configurações das máscaras CM e ACM que resultam em qualidade perceptual superior à WM. Para cada conjunto de resultados, testes de hipótese de distribuição normal dos valores de MOS-LQO foram aplicadas e a hipótese não foi rejeitada. Assumindo então dados com distribuição normal, média, mediana e variância do MOS-LQO são utilizadas como parâmetros de desempenho das máscaras. E, devido a diferenças não muito significativas na média, somente as configurações que resultam num incremento na média sem aumento na variância são consideradas com desempenho superior à WM.

As três configurações das máscaras CM e ACM que fornecem os maiores incrementos na média do MOS-LQO são apresentadas na Tabela II. Observa-se que a ACM obtém desempenho superior à WM em todas as SNRs avaliadas. A CM, por outro lado, não supera a WM para SNR = -10 dB. Observa-se também que, para aumentar a qualidade perceptual em relação à WM, a curva de ganho de ambas as máscaras desloca-se para a direita ($\xi_0 > 0$), fazendo $M(k, \lambda) = 0,5$ ocorrer em $\xi_{dB}(k, \lambda)$ maiores. E β_1 e β_2 tendem a diminuir com o aumento da SNR, levando a curvas mais suaves.

Além disso, nota-se que a ACM apresenta $8 \text{ dB} \leq \xi_0 \leq 10 \text{ dB}$. E que $\beta_2 > \beta_1$ para SNR = -10 dB, indicando a necessidade de uma curva mais abrupta nos ganhos acima de 0,5 neste caso. A Figura 2 mostra a melhor configuração da ACM para as três SNRs, ilustrando as observações acima.

TABELA II
AS TRÊS MELHORES CONFIGURAÇÕES DAS MÁSCARAS
ACM E CM EM TERMOS DE MOS-LQO.

Config. das máscaras	SNR	ACM			ACM = CM	
		β_1	β_2	ξ_0 (dB)	$\beta_1 = \beta_2$	ξ_0 (dB)
1°	-10 dB	0,17	0,8	8	—	—
	0 dB	0,18	0,13	8	0,2	6
	5 dB	0,14	0,06	10	0,14	8
2°	-10 dB	0,16	0,85	9	—	—
	0 dB	0,17	0,8	10	0,19	7
	5 dB	0,14	0,08	9	0,13	9
3°	-10 dB	0,17	0,85	8	—	—
	0 dB	0,17	0,08	10	0,18	8
	5 dB	0,14	0,08	9	0,15	8

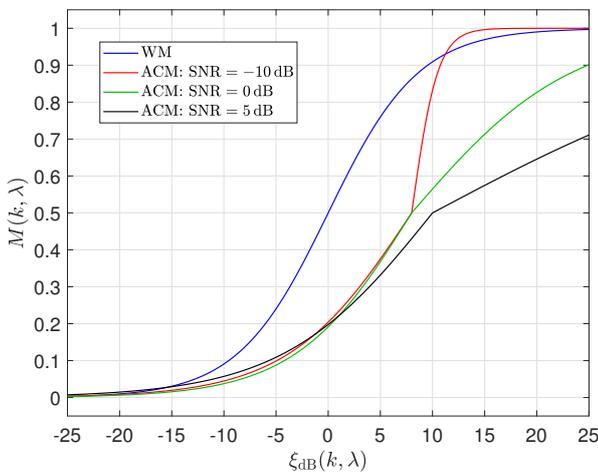


Fig. 2. Curvas de ganho da melhor configuração da ACM em termos de qualidade para as SNRs: -10 dB (vermelho), 0 dB (verde), 5 dB (preto).

Média, mediana e variância do MOS-LQO obtidas pela primeira configuração das máscaras ACM e CM, para cada SNR, são exibidas na Tabela III. De forma geral, a maior maleabilidade da ACM permite uma maior qualidade perceptual de $\hat{x}(n)$. Na média, a maior vantagem da ACM sobre a WM foi de 5,71% na SNR = 5 dB. E sobre a CM foi de 0,62% na SNR = 0 dB. Embora pequena, a vantagem da ACM sobre a CM se torna relevante ao observar que somente a ACM conseguiu superar a WM para SNR = -10 dB.

B. Experimento de Maximização da Inteligibilidade

Neste experimento, buscou-se por configurações das máscaras CM e ACM que resultam em inteligibilidade superior à WM. Para cada conjunto de resultados, testes de hipótese de distribuição normal dos valores de NCM foram aplicadas e a hipótese não foi rejeitada. Assumindo então dados com distribuição normal, média, mediana e variância do NCM são utilizadas como parâmetros de desempenho das máscaras. E devido a diferenças não muito significativas na média, somente as configurações que resultam num incremento na média sem aumento na variância são consideradas com desempenho superior à WM.

TABELA III
VALORES ESTATÍSTICOS DO MOS-LQO OBTIDOS PELA MÁSCARA WM E
PELA MELHOR CONFIGURAÇÃO DAS MÁSCARAS CM E ACM.

SNR	Estatística	WM	ACM	CM
-10 dB	Média	1,2338	1,2367	—
	Mediana	1,2179	1,2475	—
	Variância	0,0071	0,0070	—
0 dB	Média	1,5574	1,6050	1,5951
	Mediana	1,5203	1,5877	1,5672
	Variância	0,0225	0,0225	0,0224
5 dB	Média	1,7990	1,9018	1,8948
	Mediana	1,7417	1,8861	1,8784
	Variância	0,0365	0,0358	0,0358

TABELA IV
AS TRÊS MELHORES CONFIGURAÇÕES DAS MÁSCARAS
ACM E CM EM TERMOS DE NCM.

	SNR	ACM			ACM = CM	
		β_1	β_2	ξ_0 (dB)	$\beta_1 = \beta_2$	ξ_0 (dB)
1°	-10 dB	0,15	0,23	-3	0,16	-3
	0 dB	0,15	0,3	-1	0,19	-2
	5 dB	0,16	0,25	0	0,2	-1
2°	-10 dB	0,15	0,23	-3	0,15	-3
	0 dB	0,15	0,25	-1	0,2	-2
	5 dB	0,17	0,25	0	0,19	-1
3°	-10 dB	0,15	0,24	-3	0,17	-4
	0 dB	0,15	0,24	-1	0,17	-1
	5 dB	0,17	0,24	0	0,21	-1

As três configurações das máscaras CM e ACM que fornecem os maiores incrementos na média do NCM são apresentadas na Tabela IV. Diferentemente do primeiro experimento, tanto a ACM quanto a CM obtêm desempenho superior a WM em todas as SNRs avaliadas. E observa-se que, para aumentar a inteligibilidade em relação à WM, as curvas apresentam comportamentos opostos aos obtidos no aumento da qualidade perceptual. A curva de ganho das máscaras desloca-se para a esquerda ($\xi_0 \leq 0$), fazendo $M(k, \lambda) = 0,5$ ocorrer em $\xi_{dB}(k, \lambda)$ menores. E β_1 e β_2 tendem a aumentar com o aumento da SNR, embora essa tendência não seja muito evidente na ACM, levando a curvas mais abruptas.

Além disso, nota-se que a ACM apresenta $-3 \text{ dB} \leq \xi_0 \leq 0 \text{ dB}$, aumentando com o aumento da SNR. E que $\beta_2 > \beta_1$ para todas as SNRs, indicando a necessidade constante de uma curva mais abrupta nos ganhos acima de 0,5. E as variações em β_1 e β_2 em função da SNR são geralmente pequenas. A Figura 3 mostra a melhor configuração da ACM para as três SNRs, ilustrando as observações acima.

Média, mediana e variância do NCM obtidas pela primeira configuração das máscaras ACM e CM, para cada SNR, são exibidas na Tabela V. De forma geral, a maior maleabilidade da ACM permite uma maior inteligibilidade de $\hat{x}(n)$. Na média, a maior vantagem da ACM sobre a WM e CM foi de 4,42% e 0,12%, respectivamente, na SNR = -10 dB. O ganho de desempenho sobre a CM proporcionado pela ACM foi menor na inteligibilidade que na qualidade perceptual.

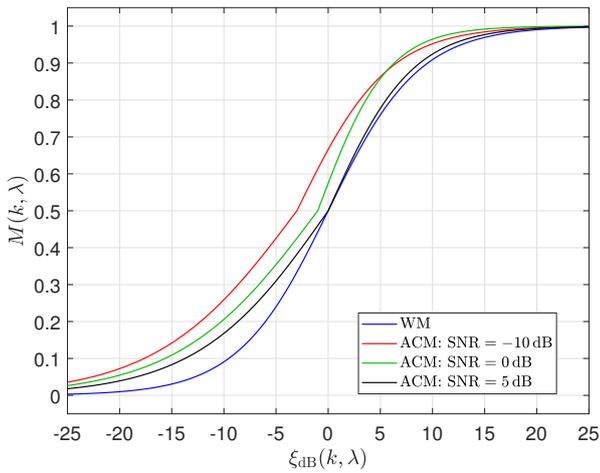


Fig. 3. Curvas de ganho da melhor configuração da ACM em termos de inteligibilidade para as SNRs: -10 dB (vermelho), 0 dB (verde), 5 dB (preto).

TABELA V

VALORES ESTATÍSTICOS DO NCM OBTIDOS PELA MÁSCARA WM E PELA MELHOR CONFIGURAÇÃO DAS MÁSCARAS CM E ACM.

SNR	Estatística	WF	ACM	CM
-10 dB	Média	0,8160	0,8521	0,8511
	Mediana	0,8132	0,8527	0,8507
	Variância	0,0017	0,0014	0,0013
0 dB	Média	0,9239	0,9296	0,929
	Mediana	0,9278	0,93	0,9302
	Variância	5,941e-4	5,7677e-4	5,9156e-4
5 dB	Média	0,9541	0,9549	0,9547
	Mediana	0,9597	0,9569	0,9569
	Variância	3,2933e-4	3,1170e-4	3,2313e-4

V. CONCLUSÕES

Este trabalho apresentou uma primeira proposta de máscara paramétrica conformável assimétrica para redução de ruído em aparelhos auditivos e implantes cocleares. A máscara proposta é capaz de modelar as do estado-da-arte, nomeadamente, máscara binária, máscara de Wiener e suas versões restrita e paramétrica, e a máscara paramétrica conformável.

O desempenho da máscara proposta foi avaliado separadamente em termos de qualidade sonora perceptual e inteligibilidade de sinais de fala. Os resultados indicaram que a maior maleabilidade da máscara proposta permite um leve aumento na qualidade e inteligibilidade da fala. Mas o comportamento da curva de ganho para maximizar individualmente esses aspectos da fala são opostos, indicando que a melhoria conjunta deles é um desafio a ser investigado no futuro.

REFERÊNCIAS

- [1] World Health Organization, “Deafness and hearing loss,” <https://www.who.int/en/news-room/fact-sheets/detail/deafness-and-hearing-loss>, 2019.
- [2] Instituto Brasileiro de Geografia e Estatística, “Censo demográfico brasileiro,” 2010.
- [3] T. Zahnert, “The differential diagnosis of hearing loss,” *Deutsches Ärzteblatt International*, vol. 108, no. 25, pp. 433–444, June 2011.
- [4] H. Dillon, *Hearing Aids*, 2nd ed. Boomerang Press, 2001.

- [5] B. C. J. Moore, *Cochlear hearing loss: physiological, psychological and technical issues*, 2nd ed. John Wiley & Sons, 2007.
- [6] J. M. Kates, K. H. Arehart, M. C. Anderson, R. K. Muralimanohar, and L. O. H. Jr, “Using objective metrics to measure hearing aid performance,” *Ear and Hear*, vol. 39, no. 6, pp. 1165–1175, 2018.
- [7] A. Hast, L. Schlücker, F. Digeser, T. Liebscher, and U. Hoppe, “Speech perception of elderly cochlear implant users under different noise conditions,” *Otology & Neurology*, vol. 36, no. 10, pp. 1638–1643, December 2015.
- [8] F. Bergeron and 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*, vol. 133, no. 1, pp. S4–S6, June 2016.
- [9] R. A. Chiea, M. H. Costa, and G. Barrault, “New insights on the optimality of parameterized wiener filters for speech enhancement applications,” *Speech Communication*, vol. 109, 2019.
- [10] —, “Uma comparação entre máscaras tempo-frequência para redução de ruído em implantes cocleares,” in *Proceedings of XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, Petrópolis, Brazil, October 2019.
- [11] D. Wang and G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Wiley-IEEE Press, 2006.
- [12] P. C. Loizou, *Speech enhancement: Theory and Practice*, 2nd ed. CRC Press, 2013.
- [13] J. S. Lim and A. V. Oppenheim, “Enhancement and bandwidth compression of noisy speech,” *Proceedings of the IEEE*, vol. 67, no. 12, pp. 1586–1604, 1979.
- [14] M. Fontaine, A. Liutkus, L. Girin, and R. Badeau, “Explaining the parameterized wiener filter with alpha-stable processes,” in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, October 2017, pp. 51–55.
- [15] R. Crochiere, “A weighted overlap-add method of short-time fourier analysis/synthesis,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 99–102, 1980.
- [16] C. A. Ynoguti, “Reconhecimento de fala contínua usando modelos ocultos de markov,” Ph.D. dissertation, Universidade Estadual de Campinas Faculdade de Engenharia Elétrica e de Computação Departamento de Comunicações, 1999.
- [17] I. Holube, S. Fredelake, M. Vlaming, and B. Kollmeier, “Development and analysis of an international speech test signal (ists),” *International Journal of Audiology*, vol. 49, no. 12, p. 891–903, 2010.
- [18] W. A. Dreschler, H. Verschuure, C. Ludvigsen, and S. Westermann, “Icra noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment,” *Audiology*, vol. 40, no. 3, p. 148–157, 2001.
- [19] ITU-T P.862, “Perceptual evaluation of speech quality (PESQ): objective method for end-to-end speech quality assessment of narrow band telephone networks and speech codecs,” International Telecommunications Union, Geneva, Switzerland 2001.
- [20] ITU-T P.862.2, “Wideband extension to recommendation P.862 for the assessment of wideband telephone networks and speech codecs,” International Telecommunications Union, Geneva, Switzerland 2005.
- [21] B. C. Bispo, P. A. A. Esquef, L. W. P. Biscainho, A. A. de Lima, F. P. Freeland, R. A. de Jesus, A. Said, B. Lee, R. W. Schafer, and T. Kaller, “EW-PESQ: A quality assessment method for speech signals sampled at 48 kHz,” *Journal of the Audio Engineering Society*, vol. 58, no. 4, pp. 251–268, April 2010.
- [22] ITU-T P.800.1, “Mean opinion score (MOS) terminology,” International Telecommunications Union, Geneva, Switzerland 2006.
- [23] I. Holube and B. Kollmeier, “Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model,” *The Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1703–1716, September 1996.