Redução de Ruído na Radiocomunicação de Alta Frequência através de Máscaras Tempo-Frequência

Erik S. Milesi, Márcio H. Costa e Bruno C. Bispo

***Resumo*— A radiocomunicação de alta frequência é uma importante forma de comunicação sem fio utilizada em aplicações comerciais e militares. Este estudo apresenta uma análise de desempenho de métodos de redução de ruído baseados em máscaras tempo-frequência em radiocomunicação de alta frequência. Foram avaliadas a máscara binária, de Wiener e raiz de Wiener. Simulações computacionais assumindo a estimação perfeita da razão sinal-ruído (SNR) e condições próximas às encontradas em comunicações marítimas, em termos de SNR e características do ruído aditivo, demonstraram aumento nos índices objetivos de inteligibilidade e qualidade. A máscara raiz de Wiener obteve resultados superiores, especialmente em baixa SNR.**

***Palavras-Chave*— Redução de ruído, radiocomunicação de alta frequência, máscara tempo-frequência, enfatização da fala.**

***Abstract*— High-frequency radio communication is an important form of wireless communication used in commercial and military applications. This study presents a performance analysis of time-frequency masking-based noise reduction methods in high-frequency radio communication. Binary, Wiener, and Wiener root masks were evaluated. Computer simulations assuming perfect signal-to-noise ratio (SNR) estimation and conditions close to those found in maritime communications, in terms of SNR and additive noise characteristics, demonstrated an increase in objective intelligibility and quality indices. The Wiener root mask achieved superior results, especially at low SNR.**

***Keywords*— Noise reduction, high-frequency radio communication, time-frequency mask, speech enhancement.**

I. Introdução

A radiocomunicação em alta frequência (HF, do inglês *high frequency*) é um importante modelo de comunicação sem fio, permitindo comunicações além do horizonte ou até globais, com alcance de milhares de quilômetros por meio da propagação de ondas celestes com refração ionosférica. O espectro HF corresponde à banda de 3 a 30 MHz e é atualmente utilizado para uma ampla variedade de aplicações comerciais e militares, por fornecer um meio de comunicação confiável para áreas isoladas ou remotas, onde serviços alternativos não estão disponíveis. Suas aplicações são variadas, incluindo comunicação rural em áreas remotas da Amazônia, comunicações de emergência em regiões de desastre, entre embaixadas, com aeronaves ou navios, e militares [1].

O rádio HF é valorizado pelo seu longo alcance, mas sua utilização não é tão simples como em outras bandas sem fio. O canal pode ser ruidoso e encontrar uma frequência utilizável pode ser desafiadora [2]. A qualidade e inteligibilidade da fala obtida a partir da faixa HF podem ser prejudicadas por interferências, ruído eletromagnético (geralmente de origem atmosférica, galáctica ou artificial), variando conforme a localização, horário e estação do ano, e ruído acústico. A interferência ocorre quando duas ou mais ondas se combinam para formar uma nova. O ruído eletromagnético surge de sinais que interferem na transmissão, enquanto o ruído acústico é uma forma de contaminação sonora gerada por atividades humanas e naturais nas imediações do microfone de captação [3].

O ruído eletromagnético proveniente de fontes solares e cósmicas estabelece um limite para a transmissão de informações na faixa HF. Esse limite é modificado pela radiação de fontes de ruído na troposfera, no ambiente terrestre e, especialmente, por fontes de rádio de origem humana. As características de longo prazo desse ruído afetam a potência necessária para transmissão, enquanto as características de curto prazo determinam como o sinal deve ser concebido e detectado para transmitir a informação desejada. Assim, o ruído é o fator decisivo que determina se o sinal é utilizável ou não para a transmissão de informações [3].

No canal sem fio, embora existam outras possibilidades, a forma de contaminação pelas fontes de ruído pode ser dividida basicamente em multiplicativa e aditiva. O ruído aditivo surge tanto devido às características intrínsecas do receptor (ruído térmico e de disparo), quanto a fontes externas (efeitos atmosféricos, radiação cósmica, e interferência de outros transmissores e aparelhos elétricos). O ruído multiplicativo tem origem em diversos processos que afetam as ondas transmitidas durante sua trajetória entre a antena transmissora e a receptora, principalmente o desvanecimento [4].

A redução de ruído em sistemas de transmissão sem fio é realizada na banda passante, utilizando filtros sintonizados do tipo passa-banda para atenuação de componentes fora da faixa de interesse [5]. Entretanto, em geral, essa estratégia não é suficiente e métodos de processamento do sinal na banda base são importantes para a recuperação da fala de interesse [6].

Uma das principais estratégias de redução de ruído em sinais de fala utiliza máscaras tempo-frequência (MTF) [7]. A cada unidade tempo-frequência, um fator de atenuação, referido como ganho, é aplicado ao sinal de fala contaminado. As MTFs são caracterizadas por uma curva de ganho em função da razão sinal-ruído (SNR, do inglês *signal-to-noise ratio*) [8]. Apesar da SNR em cada unidade de tempo-frequência não ser alterada, a SNR global pode ser aumentada se ruído e fala ocuparem bandas de frequência diferentes ao longo do tempo.

Este trabalho tem como objetivo investigar o desempenho das MTFs na redução de ruído aditivo em sinais de fala

Erik S. Milesi, Centro de Guerra Acústica e Eletrônica da Marinha, Niterói-RJ, Brasil, e Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, Florianópolis-SC, Brasil. Márcio H. Costa e Bruno C. Bispo, Departamento de Engenharia Elétrica e Eletrônica, Universidade Federal de Santa Catarina, Florianópolis-SC, Brasil. E-mails: milesi@marinha.mil.br, marcio.costa@ufsc.br, bruno.bispo@ufsc.br.

transmitidos em HF. Estimativas ideais da SNR *a priori* são utilizadas com o propósito de verificar o máximo desempenho alcançável pelas técnicas avaliadas. Para tanto, são utilizadas medidas objetivas de inteligibilidade e qualidade da fala.

Este trabalho está organizado da seguinte maneira: a Seção II formula o problema e apresenta a teoria das máscaras tempo-frequência; a Seção III descreve a configuração das simulações computacionais realizadas e os sinais empregados; na Seção IV são apresentados e discutidos os resultados obtidos; por fim, a Seção V conclui o estudo.

II. Metodologia

Esta seção apresenta o problema da contaminação por ruído na transmissão de fala em alta frequência e introduz a técnica de mascaramento tempo-frequência para redução de ruído, proporcionando a base teórica para a compreensão dos métodos e abordagens discutidos ao longo do estudo.

A. Formulação do Problema

O diagrama em blocos do problema de transmissão da fala em um canal sem fio de alta frequência é apresentado na Figura 1. O sinal de fala e o ruído são denotados por $s(n)$ e $v(n)$, respectivamente. Ambos são não-observáveis.

São considerados um canal de transmissão sem desvanecimento e condições de propagação relativamente estáveis ao longo do tempo. Dessa forma, considera-se que os sistemas são invariantes no tempo, em uma determinada janela de análise, e contaminação por ruído aditivo. Assim, o sinal de fala no receptor é definido como $y(n) = s(n) + v(n)$.

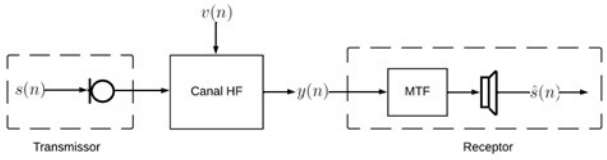

Fig. 1. Modelo do sistema de comunicação HF.

A transformada de Fourier de tempo curto (STFT, do inglês *short-time Fourier transform*) de $y(n)$ é dada por

$$Y(k,\lambda) = S(k,\lambda) + V(k,\lambda), \tag{1}$$

onde $S(k,\lambda)$ e $V(k,\lambda)$ são as STFTs de $s(n)$ e $v(n)$, respectivamente, λ denota o índice da janela de tempo e k representa o *bin* de frequência.

B. Máscaras Tempo-Frequência

A técnica de mascaramento tempo-frequência para supressão de ruído tem como objetivo atenuar as faixas de frequência do sinal de fala contaminado dominadas pelo ruído. Isso é realizado ao multiplicar, a cada unidade tempo-frequência $\{k,\lambda\}$, o sinal $Y(k,\lambda)$ por um fator $M(k,\lambda)$, resultando em uma estimativa de $S(k,\lambda)$ dada por [9]

$$\hat{S}(k,\lambda) = Y(k,\lambda)M(k,\lambda). \tag{2}$$

Apesar de não alterar a SNR em cada unidade tempo-frequência, a SNR global pode ser aumentada se ruído e fala ocuparem faixas de frequência distintas ao longo do tempo. A estimativa $\hat{s}(n)$ é construída utilizando as STFTs inversas de $\hat{S}(k,\lambda)$ e uma estratégia de sobreposição-e-soma [10].

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 SNR a priori associada a λ-ésima janela e k-ésimo bin, a qual é definida como

$$\xi(k,\lambda) = \frac{S_s(k,\lambda)}{S_v(k,\lambda)}, \tag{3}$$

onde $S_s(k,\lambda) = \mathrm{E}\left\{|S(k,\lambda)|^2\right\}$ e $S_v(k,\lambda) = \mathrm{E}\left\{|V(k,\lambda)|^2\right\}$ são as densidades espectrais de potência de $s(n)$ e $v(n)$, respectivamente, e $\mathrm{E}\{\cdot\}$ é o operador valor esperado.

Entre as MTFs encontradas na literatura, destacam-se a máscara binária (BM, do inglês *binary mask*) [11], a máscara de Wiener (WM, do inglês Wiener *mask*) [9] e a máscara raiz de Wiener (SRW, do inglês *square-root* WM) [9].

1) Máscara Binária: A BM é a MTF mais simples, com curva de ganho dada por [11]

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

onde ξ_0 é uma constante geralmente igual a 0 dB. Ela retém as componentes espectrais dominadas pela fala enquanto elimina as componentes dominadas pelo ruído.

2) Máscara de Wiener: A WM é a principal MTF encontrada na literatura e sua curva de ganho é definida por [9]

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

A WM é o filtro ótimo que minimiza, no domínio da frequência, o erro quadrático médio (MSE, do inglês *mean squared error*) entre o sinal de fala e o sinal de fala processado, ou seja, que minimiza a função de custo dada por

$$J(k,\lambda) = \mathrm{E}\{|\hat{S}(k,\lambda) - S(k,\lambda)|^2\}, \tag{6}$$

assumindo que $s(n)$ e $v(n)$ são independentes, ou pelo menos não-correlacionados, e têm média zero.

3) Máscara Raiz de Wiener: A SRW é uma versão suavizada da WM com curva de ganho dada por [9]

$$M(k,\lambda) = \sqrt{\frac{\xi(k,\lambda)}{\xi(k,\lambda)+1}}, \tag{7}$$

a qual é semelhante à curva de ganho do método de subtração espectral [9]. A SRW é o estimador ideal do espectro de potência da fala [9].

Nos últimos anos, a SRW tem sido amplamente utilizada em diversas aplicações de fala, mostrando resultados satisfatórios e frequentemente superando abordagens semelhantes que utilizaram a WM e a BM [12].

III. Configuração das Simulações

Esta seção descreve a configuração das simulações realizadas para avaliar o desempenho das máscaras BM, WM e SRW na redução de ruído em sinais de fala transmitidos por HF.

A. Ruído de Contaminação

A base de dados Ham Radio [13] foi utilizada para obter trechos de ruídos típicos em aplicações de HF. Essa base possui sinais recebidos por estações de rádios definidos por *software* (SDR, do inglês *software defined ratio*) da rede Kiwi. Os sinais de fala encontrados nessa base possuem níveis de SNRs variando entre -20 dB e 5 dB.

Os sinais são modulados em banda lateral única (SSB, do inglês *single sideband*), usando a banda lateral inferior (LSB, do inglês, *lower sideband*), com largura de banda de 2,7 kHz e frequências de portadoras de 7,05 a 7,053 MHz e de 3,6 a 3,62 MHz. A limitação em banda de 2,7 kHz atende às recomendações de transmissão da ITU. Embora originalmente amostrados a 12.001 Hz, os sinais armazenados na base de dados têm taxa de amostragem reduzida para 8 kHz.

A partir deste banco de dados, foram extraídos 750 trechos sem fala, de forma a obter amostras reais de ruído $v(n)$ proveniente do canal de comunicação em HF .

B. Sinais de Fala

A base de dados TIMIT [14] foi utilizada para gerar os sinais de fala. Essa base contém 6.300 áudios, com uma sentença cada, gravados a uma taxa de amostragem de 16 kHz. Cada locutor, de 8 grandes regiões dialetais dos EUA, pronunciou 10 sentenças. Os áudios foram reamostrados para 8 kHz.

Pares de áudios foram selecionados aleatoriamente, independentemente do gênero dos locutores, e concatenados. Intervalos de silêncio com duração de 1 s foram adicionados no início, entre as sentenças e no final de cada sinal concatenado. Um total de 750 sinais $s(n)$ foram assim gerados.

C. Fala Contaminada

Os sinais de fala contaminada $y(n)$ foram criados artificialmente pela soma de trechos de mesmo tamanho de $s(n)$ e $v(n)$. As potências de $v(n)$ foram manipuladas de forma a obter SNRs globais iguais a $\{-20, -15, -10, -5, 0, 5\}$ dB, resultando em 750 conjuntos de sinais $s(n)$, $v(n)$ e $y(n)$, sendo 125 para cada nível de SNR. Cada sinal $y(n)$ foi normalizado pelo seu maior valor absoluto. Este mesmo fator de normalização foi também utilizado para normalizar os respectvos sinais $s(n)$ e $v(n)$. Um conjunto de $s(n)$, $v(n)$ e $y(n)$ é ilustrado na Figura 2.

D. Filtragem por Máscara Tempo-Frequência

As STFTs $Y(k,\lambda)$, $X(k,\lambda)$ e $V(k,\lambda)$ foram calculadas utilizando uma janela de Hamming com duração de 32 ms, sobreposição de 75% e uma transformada discreta de Fourier (DFT) de 512 pontos. Para cada janela λ, uma estimativa de $\xi(k,\lambda)$ foi obtida como

$$\hat{\xi}(k,\lambda) = \frac{|S(k,\lambda)|^2}{|D(k,\lambda)|^2}, \quad k = 1, 2, ..., 512, \tag{8}$$

e utilizada para calcular a curva de ganho $M(k,\lambda)$ das diferentes máscaras. A curva $M(k,\lambda)$ foi por sua vez aplicada a $Y(k,\lambda)$ conforme (2), obtendo a estimativa $\hat{S}(k,\lambda)$. Por fim, a estimativa $\hat{s}(n)$ do sinal de fala foi construída utilizando as STFTs inversas de $\hat{S}(k,\lambda)$ e sobreposição-e-soma [10].

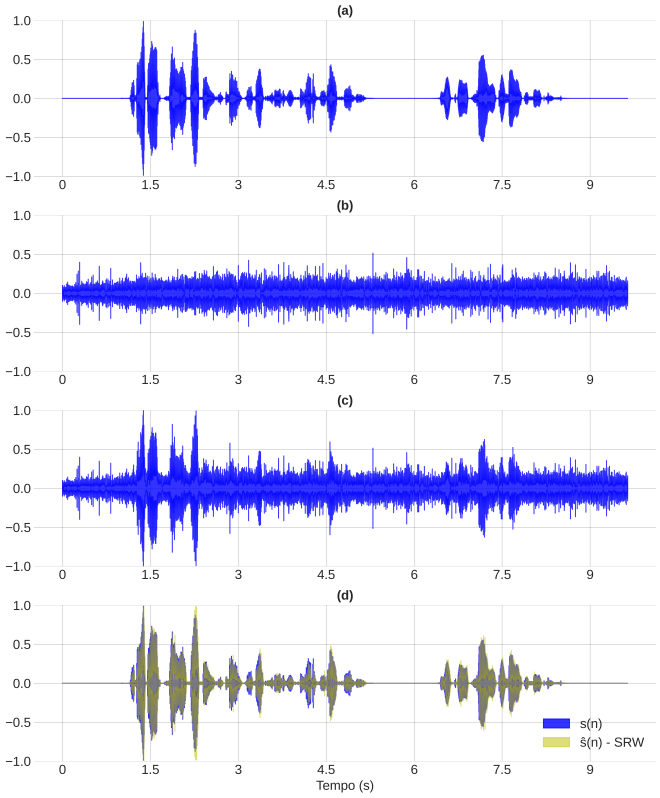

Fig. 2. Exemplo de sinais com SNR $= -10$ dB: (a) $s(n)$; (b) $v(n)$; (c) $y(n)$; (d) $\hat{s}(n)$ obtido pela SRW.

TABELA I
CORRESPONDÊNCIA ENTRE MOS-LQO E CATEGORIA DCR.

Pontuação	Categoria de Degradação
5	Inaudível
4	Audível, mas não incômoda
3	Pouco incômoda
2	Incômoda
1	Muito incômoda

E. Métricas de Avaliação

Duas métricas de desempenho foram utilizadas para avaliação da inteligibilidade e qualidade proporcionadas pelo processamento das máscaras tempo-frequência.

1) PESQ: O PESQ (*Perceptual Evaluation of Speech Quality*) é um algoritmo para avaliação objetiva da qualidade de sinais de fala amostrados a 8 kHz [15], [16]. Ele compara representações psicoacústicas de um sinal de fala possivelmente degradado e sua referência não corrompida [17].

A pontuação bruta do PESQ pode ser mapeada para a escala 1-5 da opinião média (MOS, do inglês *Mean Opinion Score*), resultando na pontuação MOS-LQO (*MOS-Listening Quality Objective*) [18]. A correspondência entre essa escala e a classificação da categoria de degradação (DCR, do inglês *Degradation Category Rating*) é mostrada na Tabela I. No entanto, o máximo MOS-LQO fornecido pelo PESQ é 4,644, quando os sinais de referência e degradados são idênticos.

Neste trabalho, o algoritmo PESQ foi utilizado para avaliar

o desempenho das máscaras quanto à qualidade sonora de $\hat{s}(n)$. Para isso, os sinais $s(n)$ e $\hat{s}(n)$ foram utilizados como os sinais de referência e degradado, respectivamente.

2) STOI: O STOI (*Short-Time Objective Intelligibility*) é uma métrica para avaliação objetiva da inteligibilidade de sinais de fala [19]. Ele é baseada em um coeficiente de correlação entre os envelopes temporais da fala limpa e degradada, utilizando segmentos sobrepostos de curta duração (384 ms). Fornece uma pontuação que varia de 0 a 1, onde valores mais elevados indicam uma maior inteligibilidade.

Neste trabalho, o algoritmo STOI foi utilizado para avaliar o desempenho das MTFs quanto à inteligibilidade de $\hat{s}(n)$. Para isso, $s(n)$ e $\hat{s}(n)$ foram utilizados como os sinais de referência e degradado, respectivamente.

IV. Resultados e Discussão

Esta seção apresenta e discute os resultados das simulações computacionais realizadas utilizando os sinais, métricas de avaliação e procedimentos descritos na Seção III.

Diagramas de caixas dos valores de MOS-LQO e STOI, obtidos pelas três MTFs estudadas e sem processamento (—), para os diferentes níveis de SNRs são mostrados na Figura 3 e 4, respectivamente. Os valores médios e o aumento percental, em relação ao caso sem processamento, de MOS-LQO e STOI obtidos são apresentados na Tabela II e na Figura 5. Observa-se que as três MTFs são capazes de proporcionar aumento de qualidade e inteligibilidade para as diferentes SNRs. Nas duas métricas e em todos os cenários, a máscara SRW apresentou o melhor desempenho, enquanto a BM o pior, mas sua vantagem sobre a WM diminui conforme a SNR aumenta.

Em relação à qualidade da fala, nota-se que o incremento proporcionado pelas máscaras tende a aumentar conforme a SNR aumenta. As máscaras WM e SRW destacaram-se ao proporcionarem incrementos médios no MOS-LQO superiores a 1,13 pontos para as SNRs de -20 e -15 dB, e superiores a 1,5 pontos para SNRs mais altas. Na SNR de -20 dB, cenário mais crítico, a superioridade da SRW frente às BM e WM foi de 71% e 15%, respectivamente. As melhorias na qualidade significam, segundo a Tabela I, que a degradação causada pelo ruído pode evoluir de muito para pouco incômoda.

Em relação à inteligibilidade, nota-se que o incremento proporcionado pelas máscaras tende a diminuir conforme a SNR aumenta, comportamento contrário ao observado na qualidade. As máscaras WM e SRW destacaram-se novamente, proporcionando incrementos médios superiores a 0,40 ponto no STOI, equivalente a mais de 100%, para as SNRs de -20 e -15 dB, e próximos a 0,10 ponto para as SNRs mais altas. Na SNR de -20 dB, cenário mais crítico, a superioridade da SRW frente às BM e WM foi de 34% e 9%, respectivamente

O desempenho superior da SRW na melhoria da inteligibilidade e qualidade da fala processada pode ser atribuído à sua capacidade de preservação do envelope temporal da fala [20], que é um fator importante para a sua percepção.

A partir desses resultados, verificou-se o potencial das MTFs para a redução de ruído na radiocomunicação em HF. Entretanto, trabalhos futuros deverão explorar o impacto de estimadores reais de SNR sobre o desempenho máximo dessa técnica. Apesar de trabalhos pregressos já terem explorado esse tipo de problema [21], as condições particulares da radiocomunicação de HF exigem uma análise aprofundada sobre a questão. Uma possibilidade é o uso de técnicas de aprendizado profundo para a estimação da SNR *a priori* [22].

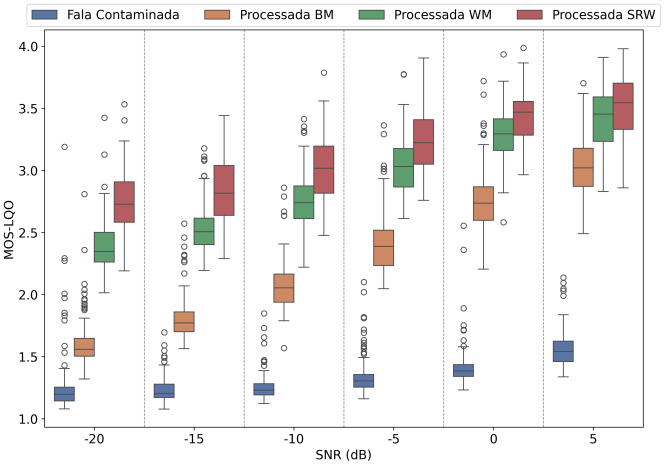

Fig. 3. Distribuição da pontuação MOS-LQO da fala contaminada e das falas processadas pelas máscaras BM, WM e SRW.

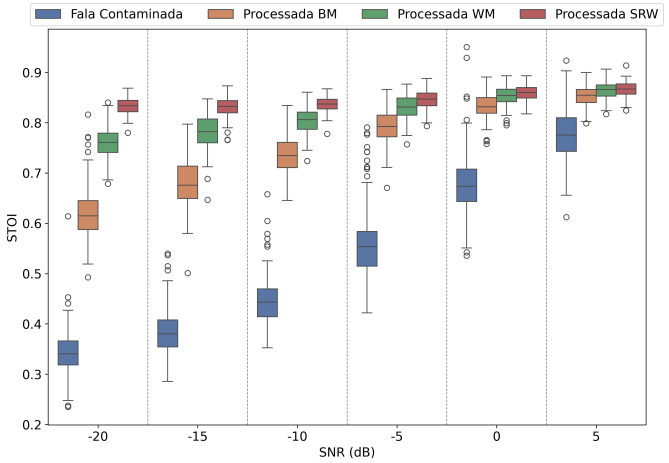

Fig. 4. Distribuição da pontuação do STOI da fala contaminada e das falas processadas pelas máscaras BM, WM e SRW.

TABELA II
Resultados médios de MOS-LQO e STOI.

Métrica	Máscara	SNR (dB)					
		-20	-15	-10	-5	0	5
MOS-LQO	—	1,26	1,23	1,26	1,33	1,42	1,56
	BM	1,60	1,82	2,07	2,41	2,76	3,03
	WM	2,39	2,54	2,76	3,04	3,28	3,42
	SRW	2,75	2,84	3,02	3,23	3,42	3,50
STOI	—	0,34	0,38	0,45	0,56	0,68	0,78
	BM	0,62	0,68	0,74	0,79	0,83	0,85
	WM	0,76	0,78	0,80	0,83	0,85	0,86
	SRW	0,83	0,83	0,84	0,85	0,86	0,87

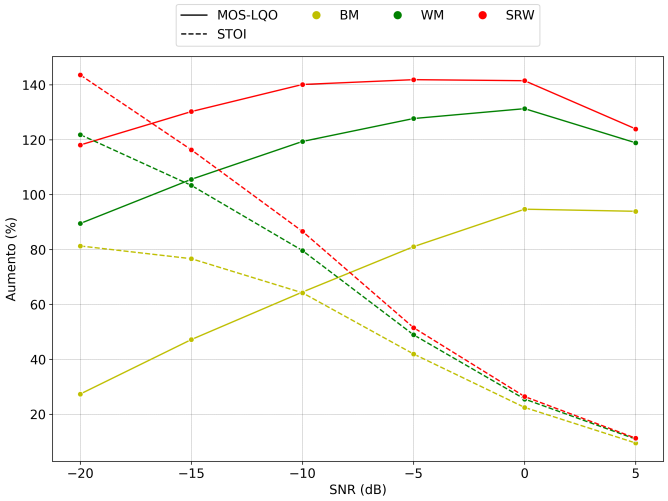

Fig. 5. Aumento percentual médio nas pontuações MOS-LQO e STOI.

V. Conclusões

Este estudo apresentou uma análise de desempenho de métodos de redução de ruído baseados em máscaras tempo-frequência em aplicações de radiocomunicação de alta frequência. Foram avaliadas as máscaras binária, de Wiener e raiz de Wiener. Simulações computacionais assumindo estimação perfeita da SNR local e condições próximas às encontradas em situações práticas, em termos de razão sinal-ruído e características do ruído aditivo, resultaram em aumentos nos índices objetivos de qualidade e inteligibilidade em relação ao sinal não processado. A máscara raiz de Wiener obteve resultados superiores, especialmente em condições de baixa SNR. Apesar das evidências promissoras, o impacto do uso de estimadores reais de SNR sobre o desempenho dessas técnicas deve ser avaliado com atenção, em especial no caso de baixas SNRs.

Referências

[1] J. Wang, G. Ding, and H. Wang, "HF communications: Past, present, and future," *China Communications*, vol. 15, no. 9, pp. 1–9, 2018.

[2] E. E. Johnson. (2020, April) Wideband steps up to fill the gap. Accessed: Dec. 21, 2023. [Online]. Available: http://wireless.nmsu.edu/hf/papers/signalWBHF.pdf

[3] N. M. Maslin, *HF Communications: A Systems Approach*. CRC Press, 2017.

[4] S. R. Saunders and A. A. Aragón-Zavala, *Antennas and Propagation for Wireless Communication Systems*. John Wiley & Sons, 2007.

[5] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[6] Z. Zhang, Y. Shi, G. Jia, and J. Yang, "The comparison of denoising methods based on air-ground speech of civil aviation," in *Proc. Biometric Recognition Chinese Conf.*, Tianjin, China, November 2015, pp. 480–487.

[7] 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, pp. 46–54, 2019.

[8] ——, "Uma comparação entre máscaras tempo-frequência para redução de ruído em implantes cocleares," *Proceedings of XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais, Petrópolis, Brazil*, 2019.

[9] P. C. Loizou, *Speech Enhancement: Theory and Practice*, 2nd ed. CRC Press, 2013.

[10] 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.

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

[12] Y. Wang, A. Narayanan, and D. Wang, "On training targets for supervised speech separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 1849–1858, 2014.

[13] J. Heitkaemper, J. Schmalenstroeer, J. Ullmann, V. Ion, and R. Haeb-Umbach, "A database for research on detection and enhancement of speech transmitted over HF links," in *Speech Communication; 14th ITG Conf.* VDE, 2021, pp. 1–5.

[14] L. F. Lamel, R. H. Kassel, and S. Seneff, "Speech database development: Design and analysis of the acoustic-phonetic corpus," in *Proc. Speech Input/Output Assessment and Speech Databases*, Venice, Italy, 1989.

[15] ITU-T, *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 Std., 2001.

[16] ——, *Wideband extension to recommendation P.862 for the assessment of wideband telephone networks and speech codecs*, International Telecommunications Union Std., 2005.

[17] 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.

[18] ITU-T, *Mean opinion score (MOS) terminology*, International Telecommunications Union Std., 2006.

[19] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, 2011.

[20] P. H. Gubert, M. H. Costa, and B. C. Bispo, "Máscara tempo-frequência baseada em envoltória para redução de ruído em implantes cocleares," in *IX Congresso Latino-Americano de Engenharia Biomédica e o XXVIII Congresso Brasileiro de Engenharia Biomédica*, Florianópolis, Brazil, Oct. 2022.

[21] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.

[22] Y. Xia and R. M. Stern, "A priori SNR estimation based on a recurrent neural network for robust speech enhancement," in *Proc. INTERSPEECH*, Hyderabad, India, 2018, pp. 3274–3278.