

Redução de Ruído na Radiocomunicação de Alta Frequência através de Máscaras Tempo-Frequência Aprimoradas por Correlação Cruzada

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

Resumo—Este artigo investiga o impacto da premissa de ortogonalidade estatística na redução de ruído por máscaras tempo-frequência na radiocomunicação de alta frequência (HF). Além das clássicas máscara de Wiener (WM) e raiz quadrada de Wiener (SRW), um modelo existente (EWM) e um modelo proposto (PEWM) da WM aprimorada por correlação cruzada são avaliados. Avaliações objetivas indicaram a superioridade dos modelos aprimorados na qualidade e inteligibilidade da fala processada. A EWM exibiu o melhor desempenho, atingindo valores próximos ao topo de escala das métricas em todos os níveis de razão sinal-ruído. Os resultados sugerem que a capacidade da SRW em preservar a potência e o envelope temporal da fala prevalece sobre o melhor desempenho da WM em estimar a forma de onda da fala, mas que a superioridade da EWM em estimar a forma de onda predomina sobre a capacidade da PEWM em preservar a potência da fala. Os resultados indicam que o desenvolvimento de métodos robustos para estimação da densidade espectral de potência cruzada é um caminho promissor para aprimorar a redução de ruído na radiocomunicação de HF.

Palavras-Chave—Radiocomunicação de alta frequência, redução de ruído, máscara tempo-frequência, filtro de Wiener, correlação cruzada.

Abstract—This paper investigates the impact of the statistical orthogonality assumption on noise reduction using time-frequency masks in high-frequency (HF) radiocommunication. In addition to the classical Wiener mask (WM) and square-root Wiener mask (SRW), an existing model (EWM) and a proposed model (PEWM) of the cross-correlation-enhanced WM are evaluated. Objective evaluations indicated the superiority of the enhanced models regarding the quality and intelligibility of the processed speech. The EWM exhibited the best performance, reaching values near the upper bound of the metrics across all SNR levels. The results suggest that the capability of SRW to preserve the speech power and temporal envelope prevails over the better performance of WM in estimating speech waveform, but that the superiority of EWM in estimating speech waveform heavily predominates over the capability of PEWM to preserve speech power. The findings indicate that developing robust methods for cross-power density spectrum estimation represents a promising path to improve HF noise reduction.

Keywords—High-frequency radio communication, noise reduction, time-frequency mask, Wiener filter, cross-correlation.

I. INTRODUÇÃO

A radiocomunicação em alta frequência (HF) é um modelo de comunicação sem fio que viabiliza comunicações globais e além do horizonte por meio da refração ionosférica, sendo crucial para aplicações comerciais, militares e de emergência em

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.

regiões isoladas [1]. No entanto, a operação na banda HF (3–30 MHz) é desafiadora devido ao ambiente altamente ruidoso e à complexidade na seleção de frequências utilizáveis [2].

A qualidade e a inteligibilidade da fala são severamente afetadas por interferências, ruídos acústicos e ruídos eletromagnéticos. As características temporais do ruído eletromagnético determinam os requisitos de potência de transmissão e a estratégia de detecção do sinal [3]. As fontes de ruído atuam predominantemente de forma aditiva e multiplicativa [4]. O ruído aditivo surge de ruídos intrínsecos ao receptor e interferências externas, enquanto o multiplicativo de fenômenos que afetam a propagação das ondas, principalmente o desvanecimento [4].

Embora a filtragem passa-banda seja convencionalmente utilizada para atenuar componentes fora da faixa de interesse na banda passante [5], essa abordagem isolada costuma ser insuficiente. Consequentemente, torna-se imprescindível a aplicação de métodos complementares de processamento de sinais em banda base para a efetiva recuperação da fala [6].

Uma das principais estratégias para redução de ruído em fala utiliza máscaras tempo-frequência (MTFs). Essa técnica se baseia na multiplicação, em cada unidade tempo-frequência (T-F), da fala contaminada por um fator de atenuação [7]. Esses ganhos, projetados para atenuar as bandas de frequência dominadas pelo ruído, são tipicamente derivados da razão sinal-ruído (SNR) local. O mapeamento da SNR para o ganho determina a função de atenuação que define a máscara [7].

Recentemente, em [8], o desempenho de MTFs clássicas — máscara binária (BM), máscara de Wiener (WM) e máscara raiz quadrada de Wiener (SRW) — na redução de ruído na radiocomunicação de HF foi avaliado. As funções de atenuação da WM e SRW são derivadas assumindo ortogonalidade estatística entre fala e ruído. Versões da WM sem essa suposição foram propostas em [9], [10] para a redução de ruído branco, ruído de burburinho, e ruído com espectro de fala, apresentando desempenhos superiores à WM.

O presente trabalho tem como objetivo avaliar o impacto da premissa de ortogonalidade na redução de ruído HF por MTFs. Para tanto, modelos aprimorados da WM e SRW, incorporando a correlação cruzada, são empregados. Este trabalho está organizado da seguinte maneira: a Seção II formula o problema e apresenta a teoria das MTFs, incluindo a proposição de uma MTF aprimorada por correlação cruzada; 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 formula o problema da contaminação por ruído na radiocomunicação de HF, introduz a técnica de masca-

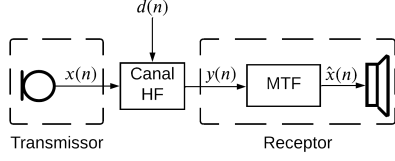


Fig. 1: Modelo do sistema de comunicação HF.

ramento T-F para redução de ruído, e descreve duas MTFs clássicas e suas versões aprimoradas por correlação cruzada.

A. Formulação do Problema

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

Como em [8], são considerados um canal de transmissão sem desvanecimento e condições de propagação estáveis ao longo do tempo. Dessa forma, considera-se que os sistemas são invariantes no tempo, em uma determinada janela de análise, e contaminação puramente por ruído aditivo. Assim, o sinal em banda base no receptor é definido como $y(n) = x(n) + d(n)$.

A transformada de Fourier de tempo curto (STFT) de $y(n)$ na janela temporal l é dada por

$$Y_l(k) = \sum_{n=0}^{N-1} y(n + lM)h(n)e^{-j(2\pi/K)kn}, \quad (1)$$

onde $k = 0, \dots, K-1$ é o bin de frequência, $h(n)$ é uma janela de análise com comprimento N , e M é o tamanho do deslocamento da janela.

Como $y(n) = x(n) + d(n)$, $Y_l(k)$ pode ser escrito como

$$Y_l(k) = X_l(k) + D_l(k), \quad (2)$$

onde $X_l(k)$ e $D_l(k)$ são as STFTs de $x(n)$ e $d(n)$ na janela temporal l , respectivamente.

B. Máscaras Tempo-Frequência

A técnica de redução de ruído por mascaramento T-F tem como objetivo atenuar as faixas de frequência da fala contaminado dominadas pelo ruído. Isso é realizado ao multiplicar, a cada unidade T-F $\{k, l\}$, o espectro $Y_l(k)$ por um fator $M_l(k)$, resultando em uma estimativa de $X_l(k)$ dada por [7]

$$\hat{X}_l(k) = Y_l(k)M_l(k). \quad (3)$$

A fala estimada $\hat{x}(n)$ é então construída usando as STFTs inversas de $\hat{X}_l(k)$ e uma estratégia de sobreposição-e-soma.

O princípio fundamental dessa técnica é que, embora a SNR local permaneça inalterada por essa operação de escala, a SNR global pode ser significativamente aprimorada se a fala e o ruído estiverem concentrados em regiões T-F distintas.

As máscaras podem ser derivadas utilizando critérios objetivos ou heurísticas. Entre as objetivas, destacam-se as MTFs derivadas a partir da minimização, em cada unidade T-F, da função custo de erro quadrático médio (MSE) dada por

$$J_l(k) = E\{|X_l(k) - \hat{X}_l(k)|^2\}. \quad (4)$$

A seguir, três MTFs existentes na literatura — duas clássicas e uma aprimorada por correlação cruzada — são descritas, e uma aprimorada e com preservação de potência é proposta.

1) *Máscara de Wiener*: A solução clássica desse problema de otimização é a WM. Ela assume que $P_{x,d,l}(k) = 0$, ou seja, que $X_l(k)$ e $D_l(k)$ são variáveis aleatórias ortogonais. Essa suposição estatística é equivalente a $X_l(k)$ e $D_l(k)$ serem de média nula e decorrelacionadas linearmente.

Sua função de atenuação é real e definida como [7]

$$M_l(k) = \frac{P_{xx,l}(k)}{P_{xx,l}(k) + P_{dd,l}(k)} = \frac{\xi_l(k)}{\xi_l(k) + 1}, \quad (5)$$

onde $\xi_l(k) = P_{xx,l}(k)/P_{dd,l}(k)$ é a SNR *a priori* associada a l -ésima janela e k -ésimo bin, $P_{xx,l}(k) = E\{|X_l(k)|^2\}$ e $P_{dd,l}(k) = E\{|D_l(k)|^2\}$ são as densidades espectrais de potência (PSD, do inglês *power spectral density*) na janela l da fala e ruído, respectivamente, $P_{x,d,l}(k) = E\{X_l(k)D_l^*(k)\}$ é a PSD cruzada entre fala e ruído na janela l .

2) *Máscara Raiz de Wiener*: A SRW é uma versão menos agressiva da WM, derivada originalmente de forma heurística, com função de atenuação dada por [7]

$$M_l(k) = \sqrt{\frac{P_{xx,l}(k)}{P_{xx,l}(k) + P_{dd,l}(k)}} = \sqrt{\frac{\xi_l(k)}{\xi_l(k) + 1}}. \quad (6)$$

Recentemente, demonstrou-se que a SRW é a MTF que minimiza (4) sob a restrição de preservação da potência da fala [11], assumindo $P_{x,d,l}(k) = 0$. Devido à restrição, o valor de $J_l(k)$ obtido pela SRW é maior ou igual ao da WM. Também demonstrou-se que, quando $P_{x,d,l}(k) = 0$, a SRW é um estimador quase-ótimo do envelope temporal da fala [12].

Como as propriedades de preservação de potência e envelope da fala estão relacionadas à inteligibilidade, a SRW tem se consolidado como uma escolha preferencial em aplicações de fala, particularmente em arquiteturas de aprendizado de máquina, nas quais fornece resultados que frequentemente superam os alcançados por métodos baseados na WM [12].

3) *Máscara de Wiener Aprimorada*: A máscara de Wiener aprimorada (EWM, do inglês *enhanced WM*) é a MTF que minimiza (4) sem a suposição de que $P_{x,d,l}(k) = 0$. Sua função de atenuação é dada por

$$M_l(k) = \frac{P_{xx,l}(k) + P_{x,d,l}(k)}{P_{xx,l}(k) + P_{dd,l}(k) + 2\text{Re}\{P_{x,d,l}(k)\}}. \quad (7)$$

A EWM é o estimador MMSE (do inglês *minimum MSE*) de (4). Nota-se que ela é uma MTF de valor complexo, com $|M_l(k)| \leq 1$, que se torna a WM quando $P_{x,d,l}(k) = 0$.

4) *Máscara de Wiener Aprimorada com Preservação de Potência*: A máscara de Wiener aprimorada com preservação de potência (PEWM, do inglês *power-constrained EWM*) é a MTF que minimiza (4) sob a restrição de preservação da potência da fala, mas sem a suposição de que $P_{x,d,l}(k) = 0$.

Pode-se demonstrar, a partir da derivação da SRW por multiplicador de Lagrange em [11], que sua função de atenuação é

$$M_l(k) = \sqrt{\frac{P_{xx,l}(k)}{P_{xx,l}(k) + P_{dd,l}(k) + 2\text{Re}\{P_{x,d,l}(k)\}}}. \quad (8)$$

Nota-se que a PEWM é uma MTF de valor real quando $-2\text{Re}\{P_{x,d,l}(k)\} \leq P_{xx,l}(k) + P_{dd,l}(k)$ e de valor imaginário caso contrário, que se torna a SRW quando $\text{Re}\{P_{x,d,l}(k)\} = 0$.

III. CONFIGURAÇÃO DAS SIMULAÇÕES

Esta seção descreve a configuração das simulações realizadas para avaliar o desempenho das MTF clássicas e aprimoradas na redução de ruído HF em sinais de fala.

A. Sinais

a) *Ruído*: A base de dados Ham Radio [13] foi usada para obter trechos de ruídos de aplicações em HF. Essa base possui sinais recebidos por estações de rádio definido por *software*, armazenados a uma taxa de amostragem de 8 kHz.

Destes sinais, 750 trechos sem fala foram extraídos para formar um conjunto de sinais $d(n)$ representando trechos reais de ruído proveniente do canal de comunicação em HF.

b) *Sinais de Fala*: A base TIMIT [14] foi utilizada para gerar os sinais de fala. Ela contém 6.300 áudios, com uma sentença cada, gravados a uma taxa de amostragem de 16 kHz. Cada locutor, de oito regiões dialetais dos EUA, pronunciou dez sentenças. Os áudios foram reamostrados para 8 kHz.

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

c) *Fala Contaminada*: Os sinais de fala contaminada $y(n)$ foram criados pela soma de trechos de mesmo tamanho de $x(n)$ e $d(n)$. As potências de $d(n)$ foram manipuladas de forma a obter SNRs globais em $y(n)$ iguais a $\{-20, -15, -10, -5, 0, 5\}$ dB, que representam condições tipicamente encontradas no canal HF [14]. Isso resultou em 750 conjuntos $\{x(n), d(n), y(n)\}$, sendo 125 para cada nível de SNR. Os sinais de cada conjunto foram normalizados pelo maior valor absoluto do seu $y(n)$ correspondente.

B. Mascaramento Tempo-Frequência

As STFTs $Y_l(k)$, $X_l(k)$ e $D_l(k)$ foram calculadas utilizando uma janela de Hamming com duração de 32 ms, sobreposição de 75% e uma FFT de 512 pontos. Para cada janela l , estimativas instantâneas das PSDs foram obtidas como $\hat{P}_{xx,l}(k) = |X_l(k)|^2$, $\hat{P}_{dd,l}(k) = |D_l(k)|^2$ e $\hat{P}_{xd,l}(k) = X_l(k)D_l^*(k)$, e utilizadas para calcular a função de atenuação $M_l(k)$ das máscaras. Em seguida, $M_l(k)$ foi aplicada a $Y_l(k)$ conforme (3), gerando $\hat{X}_l(k)$. Por fim, a estimativa $\hat{x}(n)$ foi construída utilizando as STFTs inversas de $\hat{X}_l(k)$ e sobreposição-e-soma.

A aproximação das PSDs por valores instantâneos, presumindo conhecimento prévio dos sinais de fala e ruído, é empregada em uma parcela significativa da literatura existente sobre aprimoramento de fala. Essa metodologia resulta nos coeficientes exatos das MTFs, chamadas neste caso de ideais, permitindo identificar seus limites de desempenho. Essa abordagem foi utilizada, por exemplo, em [15], [16], [17], [18].

C. Métricas de Avaliação

1) *Erro Quadrático Médio*: A redução de ruído proporcionada pelas MTFs foi avaliada através do MSE entre as formas de onda da fala desejada e da fala processada, definido como

$$E_T = \frac{1}{N} \sum_{n=0}^{N-1} [x(n) - \hat{x}(n)]^2, \quad (9)$$

onde N denota o comprimento dos sinais.

2) *PESQ*: O PESQ (*Perceptual Evaluation of Speech Quality*) é um algoritmo padronizado para avaliação objetiva da qualidade de sinais de fala amostrados a 8 kHz [19]. Ele compara representações psicoacústicas de um sinal de fala degradado e sua correspondente referência não-corrompida [20].

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*) [21]. Mas o valor máximo fornecido pelo PESQ é 4,549, quando o sinal degradado é idêntico ao de referência.

Neste trabalho, o algoritmo PESQ foi utilizado para avaliar o desempenho das máscaras quanto à qualidade sonora de $\hat{x}(n)$. Para isso, os sinais $x(n)$ e $\hat{x}(n)$ foram utilizados como os sinais de referência e degradado, respectivamente.

3) *STOI*: O STOI (*Short-Time Objective Intelligibility*), proposto em [22], é uma métrica intrusiva amplamente estabelecida para avaliar a inteligibilidade da fala [23]. Ela avalia a correlação entre envelopes de energia de tempo curto de um sinal de fala degradado e de sua correspondente referência não-corrompida. Sua pontuação varia de 0 a 1, onde a unidade corresponde à preservação da inteligibilidade da referência.

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

IV. RESULTADOS E DISCUSSÃO

Esta seção apresenta e discute os resultados de redução de ruído HF proporcionados pelas MTFs clássicas e aprimoradas.

A. Termo Cruzado

A magnitude de $\hat{P}_{xx,l}(k)$, $\hat{P}_{dd,l}(k)$ e $\hat{P}_{xd,l}(k)$, para a janela $l = 254$, um par de sinais $x(n)$ e $d(n)$ aleatoriamente escolhidos e SNR de 5 dB, são exibidos na Figura 2. Observa-se que $\hat{P}_{xd,l}(k)$ possui, em grande parte do espectro de frequências, valores na mesma ordem de grandeza de $\hat{P}_{xx,l}(k)$ ou $\hat{P}_{dd,l}(k)$.

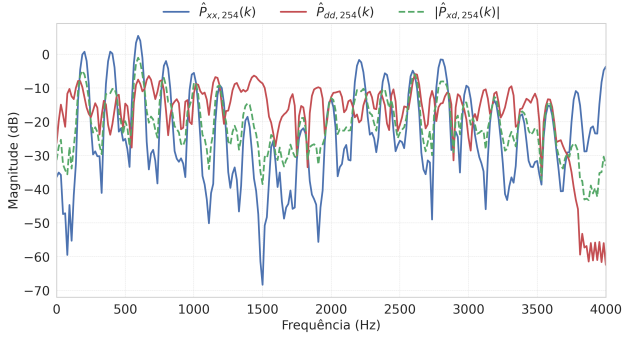
Portanto, o termo cruzado instantâneo $\hat{P}_{xd,l}(k)$ possui valores não desprezíveis no caso de ruído HF, pelo menos para as SNRs típicas desse canal. A possibilidade desse fenômeno ocorrer para alguma classe de ruído é discutida em [7, p. 7]. Ao negligenciar o termo cruzado, a WM e SRW não produzem os coeficientes que minimizam (4) na respectiva janela l . Essa inacurácia, que ocorre ao longo das janelas, prejudica a redução do ruído HF, como será mostrado a seguir.

B. Erro Quadrático Médio

Diagramas de caixa dos valores de E_T , obtidos pelas MTFs para os diferentes níveis de SNRs, são mostrados na Figura 3. Valores médios e de desvio padrão, inclusive do caso sem processamento (—), são apresentados na Tabela I.

Análises estatísticas foram realizadas em cada nível de SNR. O teste de Shapiro-Wilk foi aplicado a cada distribuição de MSE. A hipótese nula de normalidade foi consistentemente rejeitada. Então, o teste não-paramétrico de Friedman foi aplicado para identificar diferenças globais de desempenho entre as MTFs. Como variações significativas foram detectadas, o teste post-hoc de Nemenyi foi utilizado para comparações aos pares. Com um nível de significância de $\alpha = 0,05$, não foram encontradas diferenças significativas entre: PEWM e SRW em todas as SNRs; EWM e WM em -20 dB; PEWM e WM para $\text{SNR} \geq -10$ dB; SRW e WM para $\text{SNR} \geq -5$ dB.

Os resultados confirmam que a EWM apresenta o melhor desempenho quanto ao E_T em todos os níveis de SNR, indicando uma melhor estimação da forma de onda da fala $x(n)$. Devido à inclusão da restrição de preservação de potência, a


 Fig. 2: Exemplo de $\hat{P}_{xx,254}(k)$, $\hat{P}_{dd,254}(k)$ e $|\hat{P}_{xd,254}(k)|$.

PEWM proporciona um E_T maior que a EWM, enquanto a SRW fornece um E_T maior que a WM, conforme esperado.

A superioridade da EWM frente à WM aumenta com a SNR. Em -20 dB, elas apresentaram desempenhos similares. Em -15 dB, a EWM proporcionou uma redução no E_T de 20,7%. Na SNR de 5 dB, a redução atingiu 83,7%. Isso sugere que, em termos de E_T , o efeito de incluir o termo cruzado é inversamente proporcional ao nível de contaminação.

Comportamento semelhante ocorreu entre a PEWM e a SRW para $SNR \geq -5$ dB. Ressalta-se que, para $SNR \leq -10$ dB, a PEWM não superou a SRW. O uso de valores instantâneos do termo cruzado garante que a PEWM preserve a potência da fala de forma mais eficaz do que a SRW. No entanto, o desempenho inferior da SRW sobre a WM e da PEWM face à EWM indica que uma maior preservação de potência da fala não produz obrigatoriamente um menor E_T .

C. Qualidade e Inteligibilidade

Diagramas de caixa das pontuações MOS-LQO e STOI, obtidos pelas MTFs estudadas e sem processamento (—), para os diferentes níveis de SNRs são mostrados na Figura 4. Valores médios e de desvio padrão são apresentados na Tabela II.

Análises estatísticas foram realizadas em cada nível de SNR. O teste de Shapiro-Wilk foi aplicado a cada distribuição de MOS-LQO e STOI. A hipótese nula de normalidade das distribuições foi consistentemente rejeitada. Então, o teste não-paramétrico de Friedman foi aplicado para identificar diferenças globais de desempenho entre as MTFs. Como variações significativas foram detectadas, o teste post-hoc de Nemenyi foi utilizado para comparações aos pares. Com um nível de significância de $\alpha = 0,05$, não foram encontradas diferenças significativas entre: PEWM e SRW no MOS-LQO em todas as SNRs e na STOI em 0 dB; SRW e WM no STOI para $SNR \geq -5$ dB; PEWM e WM na STOI na SNR de 0 dB.

Os resultados indicam que a EWM apresenta consistentemente o melhor desempenho, atingindo o topo da escala da STOI e valores de MOS-LQO próximos ao topo de escala do PESQ, em todos os níveis de SNR. A EWM superou significativamente a WM inclusive na SNR de -20 dB, condição na qual elas apresentaram desempenhos similares no E_T . Isso indica que, na média, uma redução de E_T não implica necessariamente em melhoria de qualidade e inteligibilidade.

A PEWM proporcionou índices de inteligibilidade levemente superiores à SRW, exceto na SNR de 0 dB. Apesar de não haver diferença significativa, ela obteve valores médios de MOS-LQO superiores aos da SRW em todas as SNRs. Por outro lado, comparada a WM, a SRW alcançou índices

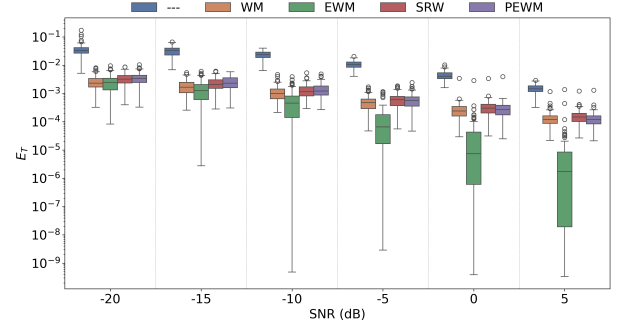

 Fig. 3: Distribuição de E_T para cada MTF e cada SNR.

 TABELA I: Resultados médios ($\times 10^{-4}$) de E_T .

Máscara	SNR (dB)					
	-20	-15	-10	-5	0	5
—	356 ± 205	321 ± 117	236 ± 73,3	109 ± 31,2	44,1 ± 15,7	15,0 ± 5,19
WM	26,4 ± 14,5	19,3 ± 11,2	11,7 ± 7,41	5,11 ± 3,00	2,83 ± 3,11	1,41 ± 1,20
SRW	34,4 ± 16,1	24,0 ± 12,4	14,1 ± 8,11	6,30 ± 3,52	3,44 ± 3,17	1,67 ± 1,25
EWM	26,3 ± 16,6	15,3 ± 12,3	6,43 ± 6,93	1,32 ± 1,99	0,57 ± 0,26	0,23 ± 0,14
PEWM	36,5 ± 17,5	26,0 ± 12,8	14,6 ± 8,53	6,17 ± 3,75	3,12 ± 3,55	1,39 ± 1,27

superiores de MOS-LQO em todas as SNRs e de STOI para $SNR \leq -10$ dB, apesar de proporcionar valores maiores de E_T .

A superioridade da SRW sobre a WM na qualidade e inteligibilidade está de acordo com estudos realizados com outros tipos de ruído [11], [12]. Isso sugere que a capacidade da SRW em preservar a potência e o envelope temporal da fala [11], fatores importantes para a sua percepção, prevalece sobre o melhor desempenho da WM na métrica E_T . Por outro lado, sugerem também que a superioridade substancial da EWM no erro temporal E_T tem predominância sobre a capacidade da PEWM em preservar a potência da fala.

Os resultados apresentados indicam que a incorporação do termo cruzado melhora o desempenho das MTFs, especialmente da WM, na redução de ruído HF. Ressalta-se que esses resultados são imunes a imprecisões nas estimativas das PSDs, uma vez que as funções $M_l(k)$ foram calculadas usando seus valores instantâneos. Portanto, eles provêm de um cenário ideal e constituem o limite superior de desempenho das MTFs.

No entanto, os resultados sugerem que o desenvolvimento de métodos robustos de estimação das PSDs é um caminho promissor para elevar os níveis de redução de ruído e aprimorar a qualidade e inteligibilidade da fala recuperada, sendo uma alternativa às pesquisas recentes por novas funções custos.

V. CONCLUSÕES

Este artigo investigou o impacto da premissa de ortogonalidade estatística na redução de ruído por MTFs na radiocomunicação de HF. Modelos aprimorados da WM, incorporando a correlação cruzada, foram propostos e avaliados.

Avaliações objetivas indicaram a superioridade dos modelos aprimorados. A EWM apresentou o melhor desempenho, atingindo o topo da escala da STOI e valores de MOS-LQO próximos ao topo de escala do PESQ, em todos os níveis de SNR. Ela superou significativamente a WM inclusive na SNR de -20 dB, condição na qual elas apresentaram desempenhos similares quanto ao erro quadrático médio da fala estimada.

Os resultados sugerem que a capacidade da SRW em preservar a potência e o envelope temporal da fala prevalece sobre o melhor desempenho da WM em estimar a forma de onda da

TABELA II: Resultados médios de MOS-LQO e STOI.

Métrica	Máscara	SNR (dB)					
		-20	-15	-10	-5	0	5
MOS-LQO	—	1,26 ± 0,26	1,23 ± 0,11	1,26 ± 0,11	1,33 ± 0,14	1,42 ± 0,17	1,56 ± 0,15
	WM	2,49 ± 0,22	2,65 ± 0,21	2,92 ± 0,20	3,23 ± 0,20	3,50 ± 0,17	3,74 ± 0,13
	SRW	3,05 ± 0,27	3,15 ± 0,27	3,37 ± 0,22	3,60 ± 0,19	3,82 ± 0,14	3,96 ± 0,11
	EWM	4,53 ± 0,17	4,55 ± 0,00	4,53 ± 0,14	4,55 ± 0,00	4,53 ± 0,19	4,53 ± 0,09
	PEWM	3,15 ± 0,30	3,27 ± 0,28	3,49 ± 0,21	3,70 ± 0,18	3,89 ± 0,15	4,01 ± 0,11
STOI	—	0,34 ± 0,05	0,38 ± 0,05	0,45 ± 0,05	0,56 ± 0,07	0,68 ± 0,07	0,78 ± 0,05
	WM	0,83 ± 0,02	0,85 ± 0,03	0,87 ± 0,02	0,90 ± 0,02	0,93 ± 0,02	0,95 ± 0,01
	SRW	0,88 ± 0,02	0,88 ± 0,02	0,89 ± 0,01	0,91 ± 0,02	0,94 ± 0,02	0,95 ± 0,01
	EWM	1,00 ± 0,01	1,00 ± 0,00	1,00 ± 0,01	1,00 ± 0,00	1,00 ± 0,01	1,00 ± 0,00
	PEWM	0,89 ± 0,02	0,90 ± 0,02	0,91 ± 0,01	0,93 ± 0,02	0,95 ± 0,01	0,97 ± 0,01

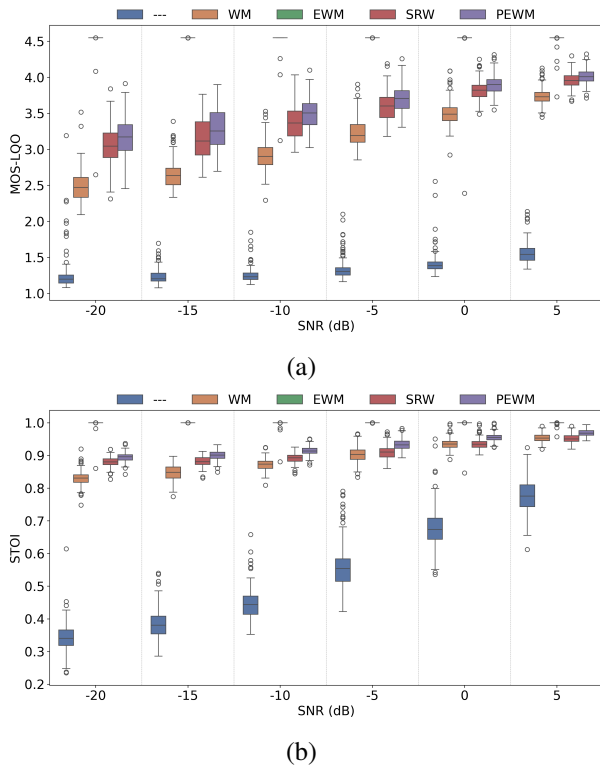


Fig. 4: Distribuição das pontuações: (a) MOS-LQO; (b) STOI.

fala. Por outro lado, sugerem também que a superioridade da EW em estimar a forma de onda da fala predominância sobre a capacidade da PEWM em preservar a potência da fala.

Portanto, conclui-se que a incorporação do termo cruzado na função de atenuação melhora o desempenho das MTFs na redução de ruído HF. Essa descoberta sugere que o desenvolvimento de métodos robustos para estimação das PSDs é um caminho promissor para aprimorar a redução de ruído HF.

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, “Wideband steps up to fill the gap,” Abr. 2020, Acessado: 2023-12-21. [Online]. Disponível: <http://wireless.nmsu.edu/hf/papers/signalWBHF.pdf>
- [3] N. M. Maslin, *HF Communications: A Systems Approach*. CRC Press, 2017.
- [4] S. R. Saunders e A. A. Aragón-Zavala, *Antennas and Propagation for Wireless Communication Systems*. John Wiley & Sons, 2007.
- [5] D. Tse e P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [6] Z. Zhang, Y. Shi, G. Jia, e J. Yang, “The comparison of denoising methods based on air-ground speech of civil aviation,” em *Biometric Recognition Chinese Conference*, Tianjin, China, 2015, pp. 480–487.
- [7] P. C. Loizou, *Speech enhancement: Theory and Practice*, 2a ed. CRC Press, 2013.
- [8] E. S. Milesi, M. H. Costa, e B. C. Bispo, “Redução de ruído na radio-comunicação de alta frequência através de máscaras tempo-frequência,” em *XLII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, Belém, Brasil, Out. 2024.
- [9] Y. Hu, M. Bhatnagar, e P. Loizou, “A cross-correlation technique for enhancing speech corrupted with correlated noise,” em *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, EUA, Maio 2001, pp. 673–676.
- [10] L. Akter e M. Hasan, “Crosscorrelation compensated Wiener filter for speech enhancement,” em *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, França, 2006, pp. 457–460.
- [11] P. H. Gubert, B. C. Bispo, e M. H. Costa, “On the optimality of the square-root Wiener time-frequency mask for noise reduction in cochlear implants,” *Biomedical Signal Processing and Control*, vol. 101, p. 107237, Mar. 2025.
- [12] —, “A theoretical framework for envelope-based time-frequency masks in cochlear implant noise reduction applications,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 33, pp. 3665–3677, 2025.
- [13] J. Heitkaemper, J. Schmalenstroer, J. Ullmann, V. Ion, e R. Haeb-Umbach, “A database for research on detection and enhancement of speech transmitted over HF links,” em *14th ITG Conference on Speech Communication*, 2021, pp. 1–5.
- [14] L. F. Lamel, R. H. Kassel, e S. Seneff, “Speech database development: Design and analysis of the acoustic-phonetic corpus,” em *Proc. Speech Input/Output Assessment and Speech Databases*, Veneza, Itália, 1989.
- [15] Y. Hu e P. C. Loizou, “Environment-specific noise suppression for improved speech intelligibility by cochlear implant users,” *Journal of the Acoustical Society of America*, vol. 127, no. 6, pp. 3689–3695, 2010.
- [16] R. A. Chiea, M. H. Costa, e G. Barrault, “New insights on the optimality of parameterized wiener filters for speech enhancement applications,” *Speech Communication*, vol. 109, pp. 46–54, Maio 2019.
- [17] R. A. Chiea, M. H. Costa, e J. A. Cordioli, “An optimal envelope-based noise reduction method for cochlear implants: an upper bound performance investigation,” *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 29, pp. 1729–1739, 2021.
- [18] R. Pigatto, B. C. Bispo, e M. H. Costa, “Uma máscara tempo-frequência conformável assimétrica para redução de ruído em aparelhos auditivos,” em *XXXVIII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, Florianópolis, Brasil, Nov. 2020.
- [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, Genebra, Suíça 2001.
- [20] B. C. Bispo, et. al, “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, 2010.
- [21] ITU-T P.800.1, “Mean opinion score (MOS) terminology,” International Telecommunications Union, Genebra, Suíça 2006.
- [22] C. H. Taal, R. C. Hendriks, R. Heusdens, e 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.
- [23] T. H. Falk, et. al, “Objective quality and intelligibility prediction for users of assistive listening devices: Advantages and limitations of existing tools,” *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 114–125, Mar. 2015.