

Um novo Método para Dereverberação da Fala e Redução de Ruído com Preservação Espacial em Aparelhos Auditivos Binaurais

Johnny Werner e Márcio H. Costa

Resumo— O filtro de Wiener multicanal (MWF) e suas variações têm sido extensivamente utilizados em aparelhos auditivos. Sua principal desvantagem é a distorção das pistas binaurais do ruído residual, o que prejudica a localização de fontes sonoras. O MWF-IC(Φ_u) realiza conjuntamente os processos de dereverberação da fala e redução de ruído, mantendo a coerência interaural no caso de ruído difuso. Neste trabalho, é proposta uma nova variação do MWF-IC(Φ_u) com a finalidade de preservar as características espaciais originais de ruído aditivo proveniente de uma fonte acústica pontual. Cenário acústico este de grande interesse ao usuário de aparelho auditivo.

Palavras-Chave— *Aparelhos auditivos binaurais, redução de ruído, dereverberação da fala, preservação espacial, MWF.*

Abstract— The multichannel Wiener filter (MWF) and its versions have been extensively applied in hearing aids. Its major drawback is the binaural cue distortion of the residual noise, which impairs the user capacity for spatial localization of acoustic sources. The MWF-IC(Φ_u) is a previously developed technique, which aims for joint speech dereverberation and noise reduction, while preserving the original interaural coherence of diffuse noise fields. In this work, we propose a new variation of the MWF-IC(Φ_u), which preserves the original noise spatial characteristics due to a point acoustic source. This scenario is of great interest by hearing aid users.

Keywords— *Binaural hearing aids, noise reduction, speech dereverberation, spatial preservation, MWF.*

I. INTRODUÇÃO

Aparelhos auditivos são dispositivos projetados para compensar perdas de audição em seres humanos. Seu desempenho está associado a critérios de inteligibilidade, qualidade, conforto acústico e carga cognitiva [1]. Dessa forma, não apenas executam processos de amplificação e compressão da faixa dinâmica, mas também incorporam subsistemas para redução de ruído e melhoria da fala [2].

Atualmente, os aparelhos auditivos mais modernos são do tipo binaural, apresentando compartilhamento de sinais e parâmetros de controle entre os aparelhos de ambas as orelhas [3]. Essa característica permite a implementação de técnicas mais eficientes de tratamento de sinais, possibilitando a utilização de informação espacial adicional [2].

Dois grandes desafios atuais para melhoria do desempenho de aparelhos auditivos são o desenvolvimento de métodos de

redução de ruído e de dereverberação da fala. Níveis elevados de ruído e reverberação afetam significativamente a qualidade do som e a inteligibilidade [4], e podem interferir na satisfação do usuário [5], acarretando diminuição do uso do aparelho ao longo do dia.

Uma das técnicas mais utilizadas para redução de ruído em aparelhos auditivos é o filtro de Wiener multicanal (MWF – *Multichannel Wiener Filter*) [2]. Apesar de proporcionar aumento significativo da razão sinal-ruído (SNR – *Signal to Noise Ratio*), sua principal desvantagem é a distorção das pistas binaurais do ruído residual, o que pode levar a uma percepção equivocada do cenário acústico. Na tentativa de superar essa limitação, diferentes variações do MWF foram propostas. Em [6] foi apresentado o método MWF-IC, cujo objetivo é propiciar redução de ruído, preservando as características espaciais originais de campos acústicos difusos através do controle da coerência interaural (IC – *Interaural Coherence*) do ruído processado. Posteriormente, em [4], foi proposta uma variação do MWF-IC que estabelece uma estratégia conjunta para redução de ruído e dereverberação da fala, preservando as pistas acústicas originais no ruído residual quando o campo acústico do ruído é do tipo difuso. Sua maior limitação consiste no fato de que parte significativa dos cenários acústicos de interesse para usuários de aparelhos auditivos é caracterizada por ruído proveniente de fonte acústica pontual [7], como é o caso de alarmes sonoros ou sinalização.

Recentemente, em [8], foi demonstrado que o MWF-IC descrito em [6] também pode ser utilizado para preservar as pistas acústicas de diferença de tempo interaural (ITD – *Interaural Time Difference*) [2] de uma fonte de ruído pontual em campo livre. Em função desse achado abre-se um novo caminho para a busca de estratégias para redução de ruído e dereverberação da fala com preservação espacial do ruído decorrente de fontes acústicas pontuais.

Neste trabalho, seguindo as contribuições de [4] e [8], é proposta uma nova variação da técnica MWF-IC para dereverberação da fala e redução de ruído com preservação das características espaciais originais de uma fonte de ruído pontual, em aparelhos auditivos binaurais.

Ao longo desse texto, letras maiúsculas e minúsculas em negrito representam, respectivamente, matrizes e vetores. Escalares são apresentados em itálico.

II. MODELOS DE SINAIS E SISTEMA

Na configuração utilizada são adotados dois aparelhos auditivos binaurais, posicionados sobre as orelhas direita e esquerda, contendo M microfones cada um. O cenário

Johnny Werner, Universidade Tecnológica Federal do Paraná, Departamento Acadêmico de Elétrica, Pato Branco-PR, e-mail: werner@utfpr.edu.br; Márcio H. Costa, Universidade Federal de Santa Catarina, Departamento de Engenharia Elétrica e Eletrônica, Florianópolis-SC, e-mail: costa@eel.ufsc.br. Este trabalho foi parcialmente financiado pelo CNPq (315020/2018-0).

operacional pressupõe a existência de uma fonte acústica de interesse e uma fonte de ruído interferente. Assume-se que o cenário acústico é fixo ou lentamente variante em uma determinada janela de tempo. Os sinais sonoros adquiridos são transformados para o domínio da frequência através da transformada de Fourier de tempo curto (STFT – *Short-Time Fourier Transform*) com N bins. Assumindo-se uma frequência de amostragem de f_s amostras por segundo, para cada *frame* de tempo λ e cada frequência k , os sinais recebidos são empilhados no vetor $\mathbf{y}(\lambda, k)$ de dimensão $2M \times 1$ como

$$\mathbf{y}(\lambda, k) = [Y_{L,1}(\lambda, k) \dots Y_{L,M}(\lambda, k) Y_{R,1}(\lambda, k) \dots Y_{R,M}(\lambda, k)]^T, \quad (1)$$

em que $Y_{L,m}(\lambda, k)$ com $m = \{1, \dots, M\}$ são as STFTs dos sinais adquiridos nos M microfones do aparelho auditivo esquerdo (L – *Left*); $Y_{R,m}(\lambda, k)$ são os sinais no aparelho auditivo direito; e $(\cdot)^T$ é o operador de transposição. O campo sonoro é descrito por um modelo de sinal paramétrico [9], no qual o vetor de sinal do microfone é dado por

$$\mathbf{y}(\lambda, k) = \mathbf{a}_L(k)X_L(\lambda, k) + \mathbf{d}(\lambda, k) + \mathbf{v}(\lambda, k), \quad (2)$$

em que $X_L(\lambda, k)$ é a representação no domínio da frequência do componente de sinal desejado (fala) no microfone de referência do aparelho auditivo esquerdo; $\mathbf{a}_L(k) = [1 A_{L,2}(k)/A_{L,1}(k) \dots A_{L,M}(k)/A_{L,1}(k) A_{R,1}(k)/A_{L,1}(k) \dots A_{R,M}(k)/A_{L,1}(k)]^T$ é o vetor que contém as funções de transferência relativas (RTF – *Relative Transfer Function*) do sinal desejado no microfone de referência do aparelho auditivo esquerdo em relação a todos os $2M$ microfones; $\mathbf{d}(\lambda, k)$ é o componente de reverberação; e $\mathbf{v}(\lambda, k)$ é o ruído aditivo.

O componente $\mathbf{d}(\lambda, k)$ modela a reverberação tardia¹, que é assumida como não correlacionada com a fala desejada $X_L(\lambda, k)$. Para comprimentos típicos de janelas da STFT de 20 a 30 ms, todos os três componentes em (2) são assumidos como mutuamente não-correlacionados [9]. Assim, a matriz de densidade espectral de potência (PSD – *Power Spectral Density*) do sinal observado é dada por

$$\begin{aligned} \Phi_{\mathbf{y}}(\lambda, k) &= E\{\mathbf{y}(\lambda, k)\mathbf{y}^H(\lambda, k)\} \\ &= \varphi_{X_L}(\lambda, k)\mathbf{a}_L(k)\mathbf{a}_L^H(k) + \Phi_{\mathbf{d}}(\lambda, k) + \Phi_{\mathbf{v}}(\lambda, k) \end{aligned}, \quad (3)$$

em que $E\{\cdot\}$ indica o operador de valor esperado; $(\cdot)^H$ é o operador de transposição conjugada; $\varphi_{X_L}(\lambda, k) = E\{|X_L(\lambda, k)|^2\}$ é a PSD do sinal desejado no microfone de referência do aparelho esquerdo; $\Phi_{\mathbf{d}}(\lambda, k) = E\{\mathbf{d}(\lambda, k)\mathbf{d}^H(\lambda, k)\}$ é a matriz de PSD da reverberação tardia; e $\Phi_{\mathbf{v}}(\lambda, k) = E\{\mathbf{v}(\lambda, k)\mathbf{v}^H(\lambda, k)\}$ é a matriz de PSD do ruído. A matriz de PSD da reverberação tardia $\Phi_{\mathbf{d}}(\lambda, k)$ pode ser modelada como um campo sonoro espacialmente homogêneo e isotrópico com potência variante no tempo [4] [9]. Assim, $\Phi_{\mathbf{d}}(\lambda, k)$ pode ser descrita por uma matriz de coerência invariante no tempo $\Gamma_{\mathbf{d}}(k)$, que leva em consideração o somreamento da cabeça, a qual é ponderada pela PSD (variante no tempo) da reverberação tardia $\varphi_{\mathbf{d}}(\lambda, k)$ [10], ou seja,

$$\Phi_{\mathbf{d}}(\lambda, k) = \varphi_{\mathbf{d}}(\lambda, k)\Gamma_{\mathbf{d}}(k). \quad (4)$$

Pode-se descrever uma matriz de PSD do componente indesejado total como $\Phi_{\mathbf{u}}(\lambda, k)$, a qual é composta pela soma das matrizes de PSD dos componentes de reverberação tardia e de ruído, ou seja,

$$\Phi_{\mathbf{u}}(\lambda, k) = \varphi_{\mathbf{d}}(\lambda, k)\Gamma_{\mathbf{d}}(k) + \Phi_{\mathbf{v}}(\lambda, k). \quad (5)$$

Para uma notação mais simples, os índices de *frame* λ e frequência k serão omitidos, sempre que possível.

III. FILTRO DE WIENER MULTICANAL

A função custo do MWF binaural é definida como [2] [11]

$$J_W(\mathbf{w}) = E\left\{\left\|\begin{bmatrix} X_L - \mathbf{w}_L^H \mathbf{y} \\ X_R - \mathbf{w}_R^H \mathbf{y} \end{bmatrix}\right\|^2\right\}, \quad (6)$$

em que $\|\cdot\|^2$ é a norma euclidiana ao quadrado e \mathbf{w}_L e \mathbf{w}_R são, respectivamente, os vetores de coeficiente ótimos esquerdo e direito do MWF, ambos com dimensão $2M \times 1$. O vetor complexo de coeficientes empilhados \mathbf{w} , de dimensão $4M \times 1$, é definido como $\mathbf{w} = [\mathbf{w}_L^T \mathbf{w}_R^T]^T$. Manipular (6) leva a [8]

$$\begin{aligned} J_W(\mathbf{w}) &= \mathbf{q}_L^T \Phi_{\mathbf{x}} \mathbf{q}_L + \mathbf{q}_R^T \Phi_{\mathbf{x}} \mathbf{q}_R - \mathbf{q}_L^T \Phi_{\mathbf{x}} \mathbf{w}_L - \mathbf{q}_R^T \Phi_{\mathbf{x}} \mathbf{w}_R \\ &\quad - \mathbf{w}_L^H \Phi_{\mathbf{x}} \mathbf{q}_L - \mathbf{w}_R^H \Phi_{\mathbf{x}} \mathbf{q}_R + \mathbf{w}_L^H \Phi_{\mathbf{y}} \mathbf{w}_L + \mathbf{w}_R^H \Phi_{\mathbf{y}} \mathbf{w}_R \end{aligned}, \quad (7)$$

em que os vetores determinísticos \mathbf{q}_L e \mathbf{q}_R , ambos com dimensão $2M \times 1$, contêm 1 no elemento correspondente ao respectivo microfone de referência (esquerda / direita) e zeros nos demais. As matrizes de PSD da fala desejada (sem reverberação) $\Phi_{\mathbf{x}} = \varphi_{X_L} \mathbf{a}_L \mathbf{a}_L^H$ e do sinal observado $\Phi_{\mathbf{y}}$ são assumidas hermitianas positivo-definidas.

A. MWF-IC($\Phi_{\mathbf{u}}$)

Em [4], foi proposto o uso do MWF-IC para dereverberação da fala e redução de ruído, com preservação da característica dispersiva de campos sonoros difusos. Isso foi alcançado adicionando-se um termo extra à função custo do MWF, responsável por minimizar a diferença entre as ICs do componente indesejado (ruído e reverberação) na entrada e saída do aparelho auditivo. Sua função custo foi definida como

$$J_C^u(\mathbf{w}) = J_W(\mathbf{w}) + \alpha J_{IC}^u(\mathbf{w}), \quad (8)$$

em que α é um fator que pondera a importância da preservação da localização espacial em comparação ao esforço de redução de ruído/dereverberação. O segundo termo de (8) é dado por

$$J_{IC}^u(\mathbf{w}) = |IC_{out}^u - IC_{in}^u|^2, \quad (9)$$

em que a IC do componente indesejado na saída é definida por

$$IC_{out}^u = \frac{\mathbf{w}_L^H \Phi_{\mathbf{u}} \mathbf{w}_R}{\sqrt{(\mathbf{w}_L^H \Phi_{\mathbf{u}} \mathbf{w}_L)(\mathbf{w}_R^H \Phi_{\mathbf{u}} \mathbf{w}_R)}}, \quad (10)$$

e a IC do componente indesejado na entrada é dada por

$$IC_{in}^u = \frac{\mathbf{q}_L^T \Phi_{\mathbf{u}} \mathbf{q}_R}{\sqrt{(\mathbf{q}_L^T \Phi_{\mathbf{u}} \mathbf{q}_L)(\mathbf{q}_R^T \Phi_{\mathbf{u}} \mathbf{q}_R)}}. \quad (11)$$

No decorrer deste trabalho esse método será referido como MWF-IC($\Phi_{\mathbf{u}}$). Sua principal limitação é que a preservação do cenário acústico espacial é garantida apenas para o caso de ruído difuso.

IV. MÉTODO PROPOSTO

O MWF-IC($\Phi_{\mathbf{u}}$) proposto em [4] baseia-se na preservação da coerência interaural original dos componentes associados à reverberação tardia e ao ruído difuso.

¹ Para sinais de fala, a separação entre as reflexões iniciais e tardias é tipicamente definida em torno de 50 ms após o som do caminho direto.

Recentemente, foi demonstrado em [8] que minimizar a diferença entre a IC de entrada e a IC de saída do ruído aditivo proveniente de uma fonte acústica pontual e processado pelo MWF binaural corresponde a minimizar a diferença de ITDs do ruído entre entrada e saída. Dessa forma, mostrou-se ser possível utilizar o MWF-IC para preservar a informação espacial de uma fonte de ruído pontual em campo livre.

Com base nas contribuições de [4] e [8], propõe-se uma nova variação da técnica MWF-IC que, além de proporcionar conjuntamente a dereverberação da fala e a redução de ruído através da técnica de MWF, também permite preservar a percepção espacial original de uma fonte de ruído pontual. Este método é denominado MWF-IC(Φ_v).

O MWF-IC(Φ_v) é baseado no mesmo raciocínio do MWF-IC(Φ_u), porém ignorando a preservação das características espaciais da reverberação tardia. Sua função custo é dada por

$$J_C^v(\mathbf{w}) = J_W(\mathbf{w}) + \alpha J_{IC}^v(\mathbf{w}), \quad (12)$$

em que α é definido de maneira semelhante a (8). O termo auxiliar é dado por

$$J_{IC}^v(\mathbf{w}) = |IC_{out}^v - IC_{in}^v|^2, \quad (13)$$

em que a diferença entre as ICs de entrada e saída do componente indesejado é definida de forma semelhante a (10) e (11), porém, substituindo-se Φ_u por Φ_v .

V. CONFIGURAÇÃO EXPERIMENTAL

A análise do método proposto foi realizada através de simulações computacionais utilizando critérios objetivos de desempenho. Comparações entre as técnicas MWF binaural convencional [2], MWF-IC(Φ_u) [4] e MWF-IC(Φ_v) foram realizadas em dois ambientes acústicos distintos, cada um composto por uma fonte pontual de interesse e outra de ruído. Para tanto, foram utilizadas funções de transferência relacionadas à cabeça (HRTF – *Head-Related Transfer Functions*) obtidas a partir de experimentos com um manequim [12]. Foram utilizados aparelhos auditivos do tipo *behind-the-ear*, equipados com três microfones cada ($M_L = M_R = 3$), em ambas as orelhas. Respostas ao impulso (HRIR – *Head Related Impulse Response*) para dois ambientes acústicos reverberantes distintos foram utilizadas [12]: *Cafeteria* ($T_{60} \cong 800$ ms) e *Office I* ($T_{60} \cong 300$ ms).

A partir das HRIRs de [12], a fonte de fala foi posicionada em frente ao manequim (azimute $\theta_s = 0^\circ$), caracterizando a situação mais frequente durante conversação [13]. Os sinais de fala foram selecionados de [14], sendo consideradas 10 frases pronunciadas por locutores diferentes (5 homens e 5 mulheres). Cada frase possui cerca de 3 segundos de duração.

O ruído utilizado nas simulações foi obtido do banco de dados fornecido pelo *International Collegium for Rehabilitative Audiology* (ICRA) [15], desenvolvido especialmente para testes de aparelhos auditivos e avaliações psicoacústicas. Foi utilizado o ruído *track 7* que apresenta características espectrais e temporais semelhantes às do som de seis locutores falando simultaneamente. O ruído foi processado por 4 HRIRs diferentes, criando 4 cenários acústicos distintos: S_0N_{-60} e S_0N_{60} para o cenário *Office I*, e S_0N_{-90} e S_0N_{90} para o cenário *Cafeteria*². Os azimutes negativos correspondem ao lado esquerdo do plano sagital do manequim, enquanto os

azimutes positivos correspondem ao lado direito. Diferentes trechos processados do ruído *track 7* foram adicionados a cada sinal de fala para emular diferentes realizações de sinais contaminados.

Para os cenários S_0N_{-60} e S_0N_{60} , tanto a fonte de fala quanto a fonte de ruído foram posicionadas a uma distância de 100 cm do manequim, sem elevação. Para o cenário S_0N_{-90} , a fonte de fala foi colocada a uma distância de 162 cm, levemente elevada, e a fonte de ruído foi colocada a uma distância de 102 cm, sem elevação, ambas em relação ao manequim. Para o cenário S_0N_{90} , a fonte de fala foi posicionada a uma distância de 102 cm, sem elevação, e a fonte de ruído, a uma distância de 162 cm, ligeiramente elevada.

A SNR do sinal contaminado foi definida na orelha mais próxima da fonte de ruído (chamada *pior orelha*). Segundo [13], a situação mais comum à qual adultos com perda auditiva são expostos compreende SNRs de 2 a 14 dB. Assim, foram adotadas duas SNRs de entrada dentro desse intervalo, além de uma SNR abaixo e outra acima dessa faixa: 0, 5, 14 e 25 dB.

A frequência de amostragem foi ajustada para $f_s = 16$ kHz e os sinais de entrada foram transformados para o domínio da frequência através de uma STFT de $N = 1024$ bins, com janela de análise de 512 amostras (equivalente a 32 ms), *zero padding* e 50% de *overlap*. Os sinais transformados no domínio STFT foram reconstruídos pelo método *weighted overlap-and-add* [16].

Assume-se que os vetores de RTF da fala $\mathbf{a}_L(k)$ e $\mathbf{a}_R(k)$ e a matriz de PSD do ruído Φ_v são conhecidos [9]. Os vetores de RTF da fala foram obtidos diretamente da HRIR janelada contendo apenas o som direto e a reverberação inicial, que compreende os primeiros 50 ms após o som direto. A matriz de PSD do ruído foi calculada a partir da HRIR e das SNRs utilizadas. Na prática, tanto as RTFs [17] quanto Φ_v [18] devem ser estimadas. O impacto no desempenho devido a erros de estimação das RTFs e de Φ_v não foi analisado neste estudo.

A PSD da reverberação tardia φ_d e a matriz de coerência Γ_d foram obtidas diretamente do componente de reverberação \mathbf{d} , da mesma forma descrita em [9]. A representação no domínio do tempo da reverberação tardia foi obtida pela convolução do sinal de fala original com as HRIRs janeladas iniciando a 50 ms após o som de caminho direto. A PSD da reverberação tardia é dada por [9]

$$\varphi_d(\lambda, k) = \frac{\mathbf{d}^H(\lambda, k)\mathbf{d}(\lambda, k)}{2M}. \quad (14)$$

O (i, j) -ésimo elemento da matriz de coerência da reverberação tardia é calculado por [9]

$$\Gamma_d^{(i, j)}(k) = \frac{\sum_{\lambda=1}^N D_i(\lambda, k)D_j^*(\lambda, k)}{\sqrt{\left(\sum_{\lambda=1}^N |D_i(\lambda, k)|^2\right)\left(\sum_{\lambda=1}^N |D_j(\lambda, k)|^2\right)}}, \quad (15)$$

em que $D_m(\lambda, k)$ é o m -ésimo elemento de $\mathbf{d}(\lambda, k)$.

Na prática, a matriz de coerência Γ_d pode ser calculada *a priori* através da geometria conhecida do arranjo de microfones, enquanto a PSD da reverberação tardia φ_d deve ser estimada (ver [9]). O impacto no desempenho do método proposto devido a erros de estimação de Γ_d e φ_d não foi abordado neste estudo.

Cinco critérios objetivos foram utilizados para avaliar o desempenho dos métodos comparados. Tais critérios avaliam

² $S_{\theta_s}N_{\theta_v}$ significa fonte de fala (S) posicionada em θ_s graus de azimute e fonte de ruído (N) em θ_v graus de azimute, em relação ao centro da cabeça.

as diferenças entre entrada e saída, sendo: (i) variação da razão sinal ruído (ΔSNR), que mede a redução de ruído [8]; (ii) variação do PESQ (ΔPESQ), que mede a qualidade geral do sinal de fala [19]; (iii) variação da *speech-to-reverberation-modulation energy ratio* (ΔSRMR), que quantifica a reverberação percebida [20]; (iv) variação do erro quadrático médio da coerência (ΔMSC), que mede a variação da coerência [6]; e (v) a variação da diferença de tempo interaural (ΔITD) do ruído, calculado até 1,5 kHz [8], que mede a preservação da ITD [11]. Pistas binaurais da fala não são avaliadas, visto que naturalmente são preservadas pela técnica MWF [2].

Os coeficientes ótimos foram obtidos pela minimização das funções custo (8) e (12) através do método de otimização *quasi-newtoniano* Broyden-Fletcher-Goldfarb-Shanno [21]. O fator de ponderação α foi mantido fixo.

VI. RESULTADOS

Os resultados apresentados nas Figs. 1-3 referem-se à *pior orelha* (direita) nos cenários S_0N_{60} e S_0N_{90} com SNR de entrada de 5 dB. Eles representam a média de 10 realizações, uma vez que cada uma das 10 frases foi contaminada com um trecho distinto do ruído ICRA. Os resultados de S_0N_{60} e S_0N_{90} foram omitidos por restrição de espaço, bem como os resultados da *melhor orelha* de S_0N_{60} e S_0N_{90} , uma vez que apresentam comportamento semelhante.

A Fig. 1 apresenta o ΔSNR em função do fator de ponderação α para o MWF-IC(Φ_u) e o MWF-IC(Φ_v). Os platôs no lado esquerdo extremo correspondem ao ΔSNR obtido com o MWF convencional ($\alpha \rightarrow 0$). Claramente, para as técnicas MWF-IC(Φ_u) e MWF-IC(Φ_v), a ΔSNR tende a diminuir com o aumento do fator de ponderação.

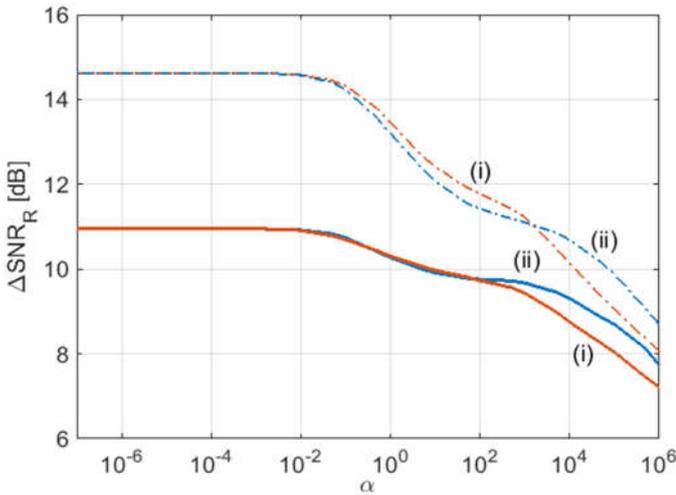


Fig. 1. ΔSNR para SNR de entrada de 5 dB na *pior orelha* (R – Right) em função de α : (i) MWF-IC(Φ_u) (laranja), e (ii) MWF-IC(Φ_v) (azul). Curvas contínuas: S_0N_{60} ; curvas traço-ponto: S_0N_{90} .

De maneira semelhante, os valores de ΔPESQ e ΔSRMR são apresentados, respectivamente, nas Figs. 2-3. Da mesma forma que o ΔSNR , o ΔPESQ diminui com o aumento de α . Por outro lado, o ΔSRMR se mostra menos susceptível a variações de α , decaindo com menor intensidade conforme o esforço relativo à preservação espacial aumenta.

As Figs. 1-3, obtidas a partir de uma SNR de entrada de 5 dB são representativas para as demais SNRs. Como resultado, verifica-se que o conforto acústico (SNR), qualidade da fala (PESQ) e dereverberação (SRMR) obtidos a partir dos

métodos MWF-IC(Φ_u) e MWF-IC(Φ_v) são aproximadamente os mesmos.

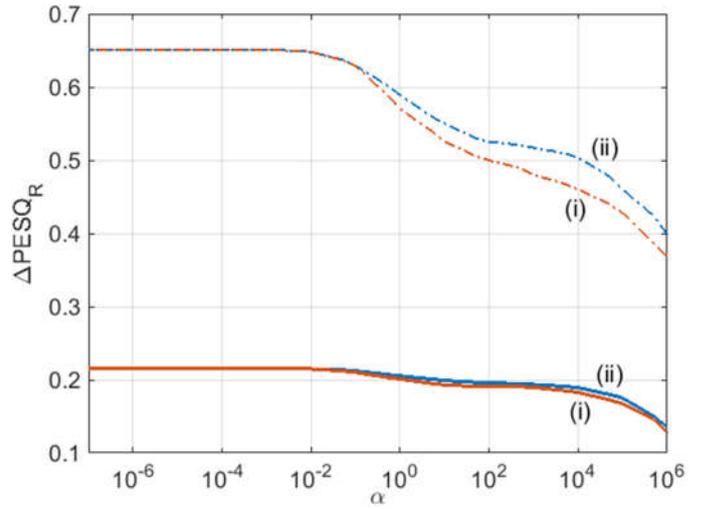


Fig. 2. ΔPESQ para SNR de entrada de 5 dB na *pior orelha* (R – Right) em função de α : (i) MWF-IC(Φ_u) (laranja), e (ii) MWF-IC(Φ_v) (azul). Curvas contínuas: S_0N_{60} ; curvas traço-ponto: S_0N_{90} .

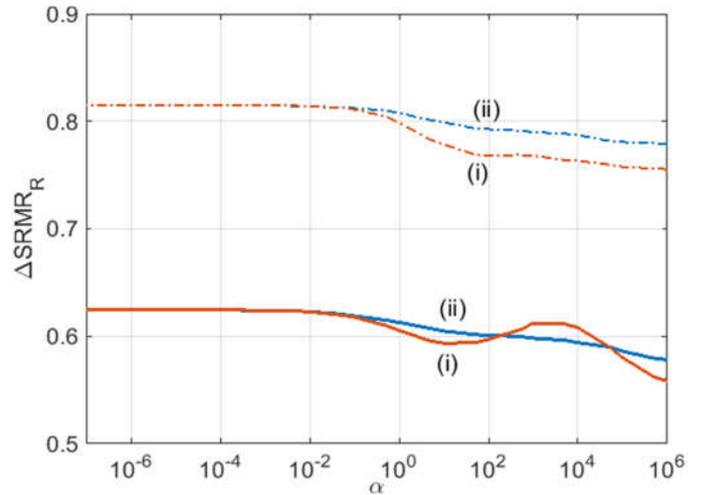


Fig. 3. ΔSRMR para SNR de entrada de 5 dB na *pior orelha* (R – Right) em função de α : (i) MWF-IC(Φ_u) (laranja), e (ii) MWF-IC(Φ_v) (azul). Curvas contínuas: S_0N_{60} ; curvas traço-ponto: S_0N_{90} .

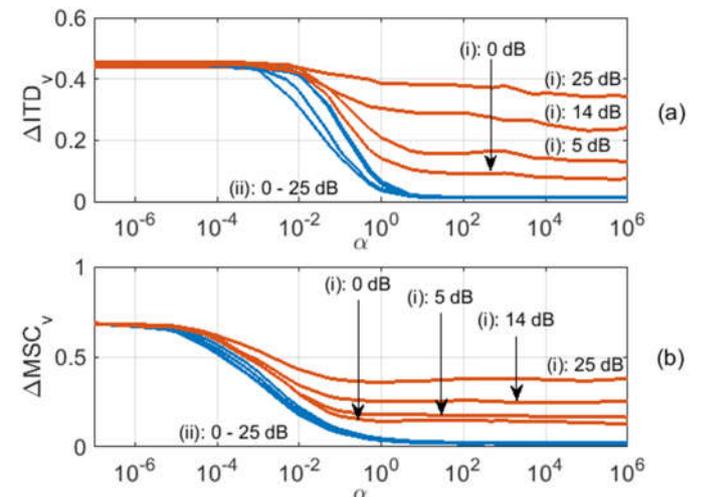


Fig. 4. ΔITD (a) e ΔMSC (b) do componente de ruído no cenário S_0N_{60} . As SNRs de entrada são: 0, 5, 14 e 25 dB. As curvas laranjas referem-se ao MWF-IC(Φ_u), e curvas em azul ao MWF-IC(Φ_v).

A Fig. 4 apresenta o ΔITD e o ΔMSC do ruído aditivo em função do parâmetro α para SNRs de entrada de $\{0, 5, 14, 25\}$ dB no cenário S_0N_{60} . Observe que, independentemente do valor de α , o MWF-IC(Φ_u) não resulta em ΔITD e ΔMSC próximos de zero, indicando alteração nas pistas acústicas binaurais do ruído residual. Por outro lado, o MWF-IC(Φ_v) proporciona uma redução significativa de ambas as métricas quando α aumenta, restaurando a diferença de tempo interaural original ($\Delta ITD \rightarrow 0$) e a coerência interaural ($\Delta MSC \rightarrow 0$) do ruído residual, ambos determinantes para a espacialização da fonte pontual [8].

VII. DISCUSSÃO

O principal objetivo desse trabalho foi vincular e expandir as contribuições apresentadas em [4] e [8] de forma a propor uma técnica de processamento de sinais para aparelhos auditivos binaurais que proporcione: (i) dereverberação da fala, (ii) redução de ruído, e (iii) preservação das características espaciais de ruído aditivo proveniente de fontes de ruído pontuais.

Para avaliar e comparar o desempenho do método proposto quanto ao conforto acústico (redução de ruído), qualidade da fala e dereverberação foram utilizadas, respectivamente, as medidas objetivas ΔSNR , $\Delta PESQ$ e $\Delta SRMR$. Os resultados indicam que tanto o MWF-IC(Φ_u) quanto o MWF-IC(Φ_v) apresentam aproximadamente o mesmo desempenho, independentemente do valor do fator de ponderação.

Por outro lado, os critérios relacionados à espacialização do ruído (ΔMSC e ΔITD) apresentados na Fig. 4 indicam que a técnica MWF-IC(Φ_u) proposta em [4] não preserva totalmente a coerência interaural nem a diferença de tempo interaural do ruído original. Essas distorções são particularmente enfatizadas para SNRs de entrada mais elevadas. O MWF-IC(Φ_u) permite diminuir apenas parcialmente os erros de ITD e MSC para valores de α extremamente elevados. Nessa situação, a SNR de saída se torna significativamente deteriorada. A técnica proposta MWF-IC(Φ_v) diminui consistentemente tanto a ΔITD quanto a ΔMSC para valores de α tais que o desempenho de ΔSNR , $\Delta PESQ$ e $\Delta SRMR$ são positivos.

Experimentos de audição preliminares realizados pelos autores indicam que tanto o MWF-IC(Φ_u) quanto o MWF-IC(Φ_v) preservam a cena acústica tanto na existência de uma fonte de ruído pontual quanto na presença de ruído difuso. No entanto, nota-se que uma melhor percepção das características originais do ruído (no caso difuso) e localização (no caso pontual) são obtidas através do sinal processado pelo método MWF-IC(Φ_v) proposto. Um experimento psicoacústico completo com voluntários está sendo realizado neste momento.

Adicionalmente, as diversas simulações realizadas no software Matlab[®] 2018, em um computador pessoal *desktop* com processador Intel[®] Core i7-4790 e 3,9 GHz de *clock*, indicaram que o processo de cálculo dos coeficientes (\mathbf{w}_L e \mathbf{w}_R) do MWF-IC(Φ_v) é cerca de 2,7 vezes mais rápido que o do MWF-IC(Φ_u). Esse resultado, embora tenha de ser investigado com mais profundidade, indica que a função custo proposta pode apresentar melhores características de otimização.

VIII. CONCLUSÕES

Esse trabalho apresentou a proposta de uma nova variação da técnica MWF-IC para aparelhos auditivos binaurais que permite estabelecer um compromisso entre dereverberação da fala e redução de ruído em função da preservação espacial da

fonte pontual de ruído. Quando comparada ao método originalmente apresentado na literatura, critérios objetivos indicam aproximadamente o mesmo desempenho em relação à melhoria de conforto acústico, qualidade da fala processada e dereverberação. Entretanto, apresenta como diferencial uma melhor capacidade de preservação das pistas binaurais associadas ao ruído original.

REFERÊNCIAS

- [1] H. Dillon, *Hearing Aids*, Thieme, 2001.
- [2] S. Doclo, S. Gannot, M. Moonen e A. Spriet, "Acoustic beamforming for hearing aid application", em *Handbook on Array Processing and Sensor Networks*. Wiley, 2010, pp. 269–302.
- [3] V. Hamacher *et al.*, "Binaural signal processing in hearing aids: Technologies and algorithms", *Advances in Digital Speech Transmission*, Wiley, pp. 401–429, 2008.
- [4] S. Braun *et al.*, "Multichannel dereverberation for hearing aids with interaural coherence preservation", em *International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2014, pp. 124–128.
- [5] L. Knudsen *et al.*, "Factors influencing help seeking, hearing aid uptake, hearing aid use and satisfaction with hearing aids: A review of the literature", *Trends in amplification*, vol. 14, no. 3, pp. 127–154, 2010.
- [6] D. Marquardt, V. Hohmann e S. Doclo, "Coherence preservation in multichannel Wiener filtering based noise reduction for binaural hearing aids", em *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 8648–8652.
- [7] K. Chung, "Challenges and recent developments in hearing aids - speech understanding in noise, microphone technologies and noise reduction algorithms", *Trends in Amplification*, vol. 8, no. 3, pp. 83–124, 2004.
- [8] F. Itturri e M. Costa, "Perceptually relevant preservation of interaural time differences in binaural hearing aids", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 4, pp. 753–764, 2019.
- [9] S. Braun *et al.*, "Evaluation and comparison of late reverberation power spectral density estimators", *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 26, no. 6, pp. 1052–1067, 2018.
- [10] O. Thiergart, M. Taseska e E. Habets, "An informed parametric spatial filter based on instantaneous direction-of-arrival estimates", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 2182–2196, 2014.
- [11] T. Bogaert *et al.*, "Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter" em *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2007, pp. 565–568.
- [12] H. Kayser *et al.*, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses", *EURASIP Journal on Advances in Signal Processing*, v. 6, 2009.
- [13] Y. Wu *et al.*, "Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss", *Ear and Hearing*, vol. 39, no. 2, pp. 293–304, 2018.
- [14] Recommendation ITU-T P.50, Telephone Transmission Quality, Telephone Installations, Local Line Networks: Objective Measuring Apparatus—Artificial Voices, Appendix I: Test Signals, 1998.
- [15] W. Dreschler *et al.*, "ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment", *Audiology*, vol. 40, no. 3, pp. 148–157, 2001.
- [16] 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.
- [17] Z. Chen, G. Gokeda e Y. Yu, *Introduction to Direction-of-Arrival Estimation*, London, U.K.: Artech House, 2010.
- [18] T. Gerkmann e R. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay". *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1383–1393, 2012.
- [19] Y. Hu e P. Loizou, "Evaluation of objective quality measures for speech enhancement", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, pp. 229–238, 2008.
- [20] J. Santos, M. Senoussaoui e T. Falk, "An improved non-intrusive intelligibility metric for noisy and reverberant speech" em *International Workshop on Acoustic Signal Enhancement*, 2014, pp. 55–59.
- [21] J. Arora, *Introduction to Optimum Design*, 2nd ed. Elsevier, 2004.