

Proposta de um parâmetro robusto para redução de ruído e preservação espacial em aparelhos auditivos biauriculares

Diego M. do Carmo, Ricardo A. Borsoi, Márcio H. Costa

Resumo—O filtro de Wiener multicanal (MWF) tem sido amplamente utilizado para redução de ruído em aparelhos auditivos biauriculares. Neste trabalho é apresentada uma proposta de parâmetro de ponderação para funções custo aumentadas, baseadas no método MWF com preservação da percepção espacial. De forma a demonstrar seu desempenho, é realizada uma análise matemática, suportada por simulações computacionais do método “MWF com preservação da função de transferência interauricular”. Os resultados indicam que, em contraposição ao uso de um parâmetro fixo, a estratégia dinâmica proposta diminui a sensibilidade do ponto de operação projetado ao efeito Lombard.

Palavras-Chave—Aparelhos auditivos, pistas biauriculares, redução de ruído.

Abstract—The multichannel Wiener filter (MWF) has been widely used for noise reduction in binaural hearing aids. This work presents a design method for the weighting parameter employed in augmented cost functions based on the MWF method with spatial preservation. To demonstrate its performance, a mathematical analysis is presented, which is supported by computer simulations of the “MWF with preservation of the interauricular transfer function”. Results indicate that, in contrast to the use of a fixed parameter, the proposed dynamic strategy decreases the sensitivity of the projected operating point to the Lombard effect.

Keywords—Hearing aids, binaural cues, noise reduction.

I. INTRODUÇÃO

O filtro multicanal de Wiener (*multichannel Wiener Filter*—MWF) e suas variações estão entre as abordagens mais estudadas para redução de ruído em aparelhos auditivos biauriculares. Considerando os sinais captados pelos microfones do aparelho auditivo como resultado da contaminação da fala por ruído aditivo, o MWF minimiza o erro quadrático médio entre a fala e o sinal contaminado filtrado. Dessa forma, a potência do ruído é reduzida, as distorções na fala são minimizadas e as informações de localização da fonte desejada são preservadas. No entanto, como efeito adverso, o MWF distorce as pistas biauriculares do ruído, fazendo com que a localização percebida do ruído seja a mesma da fonte associada à fala [11]. De forma a controlar o efeito indesejado que o MWF gera sobre as pistas de localização do ruído residual, métodos recentes de redução de ruído incorporam, à função custo original do MWF, um termo adicional de penalização sobre as distorções

das pistas acústicas do ruído [1]–[7]. Esse termo é definido por uma função custo que penaliza soluções de redução de ruído que distorcem as pistas biauriculares da fonte interferente. A composição das duas funções custo é ponderada por um parâmetro que determina o compromisso entre os objetivos de redução de ruído, associado à função custo do MWF, e de preservação da percepção espacial do ruído interferente residual, associado à função custo auxiliar.

Uma rigorosa revisão do estado da arte demonstra que os parâmetros de ponderação propostos na literatura da área são do tipo estático (fixos) [1]–[7] e, portanto, apresentam desempenho adequado apenas para sinais estacionários e condições invariantes no tempo. No entanto, sinais não-estacionários podem levar a uma degradação significativa do desempenho dos métodos utilizados, levando à perda do compromisso desejado entre redução de ruído e preservação da percepção espacial do ruído. Tais condições são intrínsecas à comunicação vocal, na qual mudanças imprevisíveis na potência do ruído ambiente podem levar a variações na potência da fala. Variações essas, que são necessárias para que o falante ajuste a razão sinal ruído de forma a manter a inteligibilidade da conversa. Esse fenômeno é conhecido como efeito Lombard [9]–[11].

Os termos de penalização espacial propostos até o momento são invariantes às modificações na potência do sinal captado [1]–[7]. Entretanto, a função custo do MWF não apresenta essa propriedade de invariância; como resultado, o equilíbrio entre redução de ruído e preservação das pistas espaciais da fonte interferente pode ser significativamente afetado por variações na potência absoluta dos sinais recebidos (ocasionadas, por exemplo, pelo efeito Lombard). Essa observação motiva a proposição de um parâmetro de ponderação dinâmico para manter o ponto de operação desejado. Essa abordagem evitaria a necessidade de ajustes manuais repetitivos dos parâmetros de controle do aparelho auditivo para acomodar variações no cenário acústico.

A partir do exposto, este trabalho apresenta uma proposta de estrutura e um método para o ajuste automático do parâmetro de ponderação de técnicas biauriculares de redução de ruído com preservação espacial, a fim de garantir bons níveis adequados de redução de ruído juntamente com a preservação das informações de localização da fonte interferente em situações em que ocorrem variações na potência do sinal de entrada. No conhecimento dos autores, essa é a primeira contribuição apresentada na literatura da área. O método proposto pode ser aplicado diretamente nas funções custo mais utilizadas para preservação espacial em aplicações com aparelhos auditivos,

Diego M. do Carmo, Ricardo A. Borsoi e Márcio H. Costa Universidade Federal de Santa Catarina. E-mails: diego.carmoh@gmail.com, raborsoi@gmail.com, costa@eel.ufsc.br. Este trabalho foi parcialmente financiado por CAPES e CNPq, projeto 315020/2018-0.

tais como em [1]–[7], que também têm por objetivo manter o compromisso entre níveis adequados de redução de ruído e preservação de informação espacial.

O método proposto foi desenvolvido para cenários acústicos compostos por uma fonte de fala e um sinal de interferência, cujo campo acústico pode ser do tipo difuso [4], ou gerado por uma fonte pontual [1][2][6]. Uma análise teórica é apresentada para esclarecer os mecanismos pelos quais variações de potência do sinal contaminado de entrada impactam o desempenho dos métodos biauriculares baseados no MWF. Em seguida, são apresentadas propriedades matemáticas que devem ser satisfeitas pela função custo aumentada do MWF para que ela apresente invariância de ponto de operação com respeito a variações de potência do sinal de entrada. Essas observações motivam o projeto de um parâmetro de ponderação dinâmico que confere robustez a variações de potência. Simulações computacionais ilustram o desempenho do método proposto. Ao longo deste trabalho, variáveis escalares são representadas por letras minúsculas em itálico, constantes por letras maiúsculas em itálico, vetores por letras minúsculas em negrito e matrizes por letras maiúsculas em negrito.

II. SINAIS E SISTEMA

Esta seção descreve os estágios de aquisição e processamento de sinais de um aparelho auditivo biauricular.

Considerando uma fala humana sujeita à influência de um ruído interferente e um dado segmento fixo de tempo (λ), a representação na frequência (ω) do sinal captado pelo microfone $m \in \{1, 2, \dots, M\}$ no aparelho auditivo indexado por $\ell \in \{L, R\}$ é dada por $y_{\ell,m}(\omega, \lambda) = x_{\ell,m}(\omega, \lambda) + v_{\ell,m}(\omega, \lambda)$, em que $x_{\ell,m}(\omega, \lambda)$ é o componente da fala; $v_{\ell,m}(\omega, \lambda)$ é o componente de ruído; e $M = M_L + M_R$ é o número total de microfones no sistema biauricular, sendo M_L o número de microfones no aparelho auditivo da esquerda (L), e M_R no da direita (R). Por simplicidade, os índices ω e λ serão omitidos no restante do texto a partir deste ponto.

O conjunto de amostras contaminadas a cada instante é representado pelo vetor $\mathbf{y} \in \mathbb{C}^M$, conhecido por ambos os aparelhos auditivos devido a um canal de comunicação *full duplex*; definido como $\mathbf{y} = [\mathbf{y}_L^T \mathbf{y}_R^T]^T$, em que $\mathbf{y}_\ell = [y_{\ell,1}, y_{\ell,2}, \dots, y_{\ell,M_\ell}]^T$ é o vetor de amostras no lado ℓ . O vetor \mathbf{y} pode ser decomposto da seguinte forma:

$$\mathbf{y} = \mathbf{x} + \mathbf{v}, \quad (1)$$

em que \mathbf{x} e \mathbf{v} são definidos de maneira similar a \mathbf{y} .

As matrizes de coerência do sinal contaminado ($\Phi_{\mathbf{y}}$), da fala ($\Phi_{\mathbf{x}}$) e do ruído ($\Phi_{\mathbf{v}}$) são genericamente definidas da seguinte forma: $\Phi_{\mathbf{d}} = \mathbb{E}\{\mathbf{d}\mathbf{d}^H\}$, em que $\mathbf{d} \in \{\mathbf{y}, \mathbf{x}, \mathbf{v}\}$, e $\mathbb{E}\{\cdot\}$ representa o valor esperado em relação à \mathbf{x} e \mathbf{v} . Define-se também as matrizes $\Phi_{\mathbf{y}\mathbf{y}}$ e $\Phi_{\mathbf{x}\mathbf{x}}$ dadas por:

$$\Phi_{\mathbf{y}\mathbf{y}} = \begin{bmatrix} \Phi_{\mathbf{y}} & \mathbf{0}_M \\ \mathbf{0}_M & \Phi_{\mathbf{y}} \end{bmatrix} \quad \text{e} \quad \Phi_{\mathbf{x}\mathbf{x}} = \begin{bmatrix} \Phi_{\mathbf{x}} & \mathbf{0}_M \\ \mathbf{0}_M & \Phi_{\mathbf{x}} \end{bmatrix}, \quad (2)$$

em que $\mathbf{0}_M$ é uma matriz quadrada de ordem M composta por zeros.

Em geral, o microfone frontal de cada aparelho auditivo é utilizado como referência. O sinal nesse microfone é dado

por $\mathbf{q}_\ell^T \mathbf{y}_\ell = y_\ell = x_\ell + v_\ell$, em que \mathbf{q}_ℓ é o vetor de seleção da amostra de referência na posição associada ao microfone frontal do lado ℓ e zero nas demais posições.

O sinal enviado aos alto-falantes dos aparelhos auditivos, no domínio da frequência, é dado por $z_\ell = \mathbf{w}_\ell^H \mathbf{y}$, em que $\mathbf{w}_\ell \in \mathbb{C}^M$ representa o filtro biauricular de redução de ruído no lado ℓ . O filtro de redução de ruído e o vetor de seleção do sistema biauricular são dados respectivamente por:

$$\mathbf{w} = [\mathbf{w}_L^T \mathbf{w}_R^T]^T, \quad \mathbf{q} = [\mathbf{q}_L^T \mathbf{q}_R^T]^T. \quad (3)$$

III. PRESERVAÇÃO DE INFORMAÇÃO ESPACIAL EM APARELHOS AUDITIVOS BIAURICULARES

O MWF tem sido amplamente estudado no contexto de redução de ruído para aplicações em aparelhos auditivos biauriculares. Ele é baseado na minimização da função custo J_{MWF} , dada por:

$$J_{\text{MWF}} = \mathbb{E}\{|x_L - \mathbf{w}_L \mathbf{y}|^2\} + \mathbb{E}\{|x_R - \mathbf{w}_R \mathbf{y}|^2\}, \quad (4)$$

em que $|\cdot|$ é a operação módulo. Manipulando-se (4), considerando \mathbf{x} e \mathbf{v} não correlacionados, e utilizando (2) e (3), obtém-se:

$$J_{\text{MWF}}(\mathbf{w}) = \mathbf{w}^H \Phi_{\mathbf{y}\mathbf{y}} \mathbf{w} - \mathbf{w}^H \mathbf{p}_{\mathbf{x}\mathbf{x}} - \mathbf{p}_{\mathbf{x}\mathbf{x}}^H \mathbf{w} + p_{\mathbf{x}\mathbf{x}}, \quad (5)$$

em que o vetor $\mathbf{p}_{\mathbf{x}\mathbf{x}} = \Phi_{\mathbf{x}\mathbf{x}} \mathbf{q}$ contém as correlações entre o componente de fala no microfone de referência e todos os componentes de fala captados, i.e., $\mathbb{E}\{\mathbf{x}\mathbf{x}_\ell^*\}$, e $p_{\mathbf{x}\mathbf{x}} = \mathbf{q}^T \Phi_{\mathbf{x}\mathbf{x}} \mathbf{q} = \mathbb{E}\{|x_L|^2\} + \mathbb{E}\{|x_R|^2\}$ é a potência total da fala nos microfones de referência.

A fim de penalizar soluções que distorçam as pistas biauriculares da fonte de ruído interferente, adiciona-se à função custo original do MWF uma função custo auxiliar que visa impedir soluções \mathbf{w}_L e \mathbf{w}_R que não preservam as informações espaciais da fonte interferente presentes no componente de ruído [1]–[7]. A função custo resultante é dada por:

$$J_{\text{T}} = J_{\text{MWF}} + \sum_{i=1}^I \alpha_i J_i, \quad (6)$$

em que $\alpha_i \geq 0$ são os parâmetros de ponderação que determinam o equilíbrio (esforço de minimização) entre os distintos termos de J_{T} , e I é o número total de funções custo auxiliares. Em geral, a função custo J_i é definida como:

$$J_i = \mathbb{E}\{|BM_{\text{ou}} - BM_{\text{in}}|\}, \quad (7)$$

em que $BM \in \{ITD, ILD, ITF, IC\}$ significa “medida biauricular”; e os subscritos “in” e “out” referem-se às medidas na entrada e na saída dos aparelhos auditivos. As definições matemáticas das medidas biauriculares, que caracterizam a percepção espacial dos campos acústicos, podem ser encontradas nos artigos [1]–[7]. Em geral, os valores de BM_{ou} e BM_{in} são definidos a partir de uma razão entre estimativas das pistas espaciais biauriculares.

A função custo em (6) permite estabelecer um compromisso desejado entre redução de ruído (devido à minimização da J_{MWF}) e a preservação da informação espacial da fonte interferente (devido ao termo J_i). As técnicas propostas na literatura estabelecem um valor fixo para α_i , geralmente projetado de maneira empírica [1]–[7].

IV. DESCRIÇÃO DO PROBLEMA E MÉTODO PROPOSTO

Esta seção descreve matematicamente o problema a ser resolvido (Seção A) e apresenta uma solução utilizando o conceito de homogeneidade de uma função (Seção B). Por fim, uma descrição da metodologia proposta para o projeto do parâmetro α_i é apresentada (Seção C).

A. Descrição do problema

Para demonstrar o efeito que um valor fixo para α_i tem sobre os termos da função custo em (6), os vetores de fala e de ruído em (1) são descritos por:

$$\mathbf{x} = g\tilde{\mathbf{x}}_{\text{SNR}}, \quad \mathbf{v} = g\tilde{\mathbf{v}}, \quad (8)$$

em que $g^2 > 0$ é a potência média do ruído nos microfones, dada por:

$$g^2 = \mathbb{E}\{\|\mathbf{v}\|^2\}; \quad (9)$$

o vetor $\tilde{\mathbf{v}} = \mathbf{v}/\mathbb{E}\{\|\mathbf{v}\|^2\}^{1/2}$ é o vetor normalizado de ruído; $\tilde{\mathbf{x}}_{\text{SNR}} = \text{SNR}_{\text{in}}^{1/2}\tilde{\mathbf{x}}$, em que $\tilde{\mathbf{x}} = \mathbf{x}/\mathbb{E}\{\|\mathbf{x}\|^2\}^{1/2}$ é o vetor normalizado da fala; e SNR_{in} é a razão sinal-ruído (SNR) global de entrada, dada por

$$\text{SNR}_{\text{in}} = \mathbb{E}\{\|\mathbf{x}\|^2\} / g^2. \quad (10)$$

Usando (8) em (1), resulta em $\mathbf{y} = g(\tilde{\mathbf{x}}_{\text{SNR}} + \tilde{\mathbf{v}})$. Note que \mathbf{y} é modelado como um ganho aplicado sobre a soma ponderada de dois sinais normalizados. O parâmetro g é nomeado aqui como *ganho de Lombard* [9]–[11], uma vez que ele pode ser associado ao esforço de fala necessário para manter a mesma SNR original em condições variáveis de potência do ruído aditivo.

Assumindo uma SNR fixa, o modelo apresentado em (9) estabelece um modelo simplificado para o efeito Lombard, uma vez que não considera outras modificações espectrais [12]. Ainda assim, esse modelo nos permite estudar o impacto das variações da potência dos sinais de entrada no desempenho de métodos de redução de ruído baseados em (6), o que pode ser feito considerando variações homogêneas de potência, i.e., com SNR fixa, ou não-homogêneas, i.e., com variações de SNR.

Substituindo (8) em (6), J_{MWF} pode ser apresentada de maneira compacta como:

$$J_{\text{MWF}} = g^2 \left[\text{SNR}_{\text{in}} \|\mathbf{Q} - \mathbf{W}\|_{\Phi_{\tilde{\mathbf{x}}}^{1/2}}^2 + \|\mathbf{W}\|_{\Phi_{\tilde{\mathbf{v}}}^{1/2}}^2 \right], \quad (11)$$

em que $\mathbf{Q} = [\mathbf{q}_L \mathbf{q}_R]^\top$, $\mathbf{W} = [\mathbf{w}_L \mathbf{w}_R]^\top$, $\Phi_{\tilde{\mathbf{x}}} = \mathbb{E}\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\}$ e $\Phi_{\tilde{\mathbf{v}}} = \mathbb{E}\{\tilde{\mathbf{v}}\tilde{\mathbf{v}}^H\}$; e $\|\mathbf{x}\|_{\mathbf{A}} = \|\mathbf{Ax}\|$ para matrizes \mathbf{A} de dimensões compatíveis.

Como as pistas biauriculares utilizadas na literatura são definidas a partir da razão entre estimativas de pistas biauriculares, é possível verificar que a definição geral de J_i não é afetada nem pelo ganho g , nem por variações na SNR_{in} .

B. Homogeneidade

Um conceito útil para compreender a influência do ganho de Lombard no desempenho de métodos baseados em (6) é o grau de homogeneidade [13], definido como segue:

Definição 1: Uma função $f : \mathbb{C}^M \cdot \mathbb{C}^M \rightarrow \mathbb{R}$ é chamada *positivamente homogênea* de grau N (ou N -homogênea), com respeito a ambos $\mathbf{a}, \mathbf{b} \in \mathbb{C}^M$, para $N \in \mathbb{R}$, se $f(c\mathbf{a}, c\mathbf{b}) = c^N f(\mathbf{a}, \mathbf{b})$ para todo $c > 0$.

Como o mínimo de uma função não varia se ela é multiplicada por um escalar positivo, o mínimo de funções N -homogêneas é invariante a modificações positivas em seus argumentos, i.e., $\min f(c\mathbf{a}, c\mathbf{b}) = \min c^N f(\mathbf{a}, \mathbf{b})$. Essa propriedade é útil no estudo do impacto do efeito Lombard no desempenho de métodos de redução biauriculares de ruído baseados no MWF, bem como no projeto dos parâmetros α_i .

Aplicando a Definição 1 em (11), tem-se:

$$J_{\text{MWF}}(g\tilde{\mathbf{x}}_{\text{SNR}}, g\tilde{\mathbf{v}}) = g^2 J_{\text{MWF}}(\tilde{\mathbf{x}}_{\text{SNR}}, \tilde{\mathbf{v}}), \quad (12)$$

ou seja, a função custo em (6) é 2-homogênea em relação a $\tilde{\mathbf{x}}_{\text{SNR}}$ e $\tilde{\mathbf{v}}$. Fazendo o mesmo com as funções de preservação J_i utilizadas na literatura da área [1]–[7], pode-se verificar que

$$J_i(g\tilde{\mathbf{x}}_{\text{SNR}}, g\tilde{\mathbf{v}}) = J_i(g\tilde{\mathbf{v}}) = J_i(\tilde{\mathbf{v}}), \quad (13)$$

ou seja, elas são funções 0-homogêneas com relação a $\tilde{\mathbf{x}}_{\text{SNR}}$ e $\tilde{\mathbf{v}}$. Logo, quando assume-se o uso de parâmetros α_i fixos, a função custo J_T não é homogênea, uma vez que é a soma de termos com grau de homogeneidade diferentes. Consequentemente, o mínimo de J_T é afetado tanto por g quanto pela SNR de entrada.

A partir do exposto, neste trabalho propõe-se a utilização de parâmetros α_i dinâmicos, de forma a propiciar maior robustez ao compromisso estabelecido entre redução de ruído e preservação da percepção espacial da fonte interferente em relação ao efeito Lombard e à variações na SNR de entrada. Mais especificamente, propõe-se o projeto de α_i definido através de um constante multiplicada pelo ganho de Lombard ao quadrado, o que resulta em:

$$\alpha_i = \beta \cdot g^2, \quad (14)$$

em que β é uma constante que define o ponto de operação da função custo J_T , sendo associado ao compromisso desejado entre redução de ruído e preservação das pistas biauriculares. Usando (14) em (6), pode-se mostrar que:

$$\begin{aligned} J_T(g\tilde{\mathbf{x}}_{\text{SNR}}, g\tilde{\mathbf{v}}) &= g^2 \left[J_{\text{MWF}}(\tilde{\mathbf{x}}_{\text{SNR}}, \tilde{\mathbf{v}}) + \beta \sum_{i=1}^I J_i(\tilde{\mathbf{v}}) \right] \\ &= g^2 J_T(\tilde{\mathbf{x}}_{\text{SNR}}, \tilde{\mathbf{v}}). \end{aligned} \quad (15)$$

A equação (15) mostra que com a escolha proposta para o parâmetro α_i , a função custo J_T se torna 2-homogênea com relação a $\tilde{\mathbf{x}}_{\text{SNR}}$ e $\tilde{\mathbf{v}}$, garantindo que variações conjuntas (homogêneas) nas potências da fala e do ruído não afetem o desempenho do método.

C. Robustez a variações de SNR

Apesar do parâmetro em (14) conferir robustez à variações no ganho de Lombard, J_T ainda não é homogênea com relação a variações na SNR de entrada. Para investigar o impacto dessa variável sobre J_T , o processo de otimização da função custo analisada é descrito a partir das equações (11) e (15) como:

$$g^2 \cdot \min_{\mathbf{W}} \left\{ \text{SNR}_{\text{in}} \|\mathbf{Q} - \mathbf{W}\|_{\Phi_{\tilde{\mathbf{x}}}^{1/2}}^2 + \|\mathbf{W}\|_{\Phi_{\tilde{\mathbf{v}}}^{1/2}}^2 + \beta \sum_{i=1}^I J_i(\tilde{\mathbf{v}}) \right\}. \quad (16)$$

Analisando-se (16), nota-se que quando a SNR de entrada (SNR_{in}) aumenta, a contribuição do termo $\|\mathbf{Q} - \mathbf{W}\|_{\Phi_x}^{1/2}$ em J_T também aumenta. Logo, as soluções \mathbf{w}_L e \mathbf{w}_R se aproximam de \mathbf{q}_L e \mathbf{q}_R , respectivamente, i.e., de filtros passa tudo, que preservam as pistas biauriculares da fonte interferente. Essa suposição é condizente com a solução do filtro de Wiener apresentada em [14]. Portanto, o aumento da SNR de entrada tende a tornar as pistas biauriculares do sinal processado mais próximas das pistas do sinal original, o que evita erros nas informações biauriculares e, conseqüentemente, permite uma percepção espacial mais precisa da fonte interferente.

D. Projeto do parâmetro de ponderação

Considerando as observações apresentadas na seção IV.C, o projeto do parâmetro β se torna simples. Baseando-se na estrutura definida em (14), que provê robustez ao ganho de Lombard, e definindo uma SNR inicial mínima de projeto, β pode ser escolhido como o menor valor tal que um nível mínimo de preservação das pistas biauriculares seja atingido pelas soluções \mathbf{w}_L e \mathbf{w}_R , i.e., de forma a resultar em erros na ILD e ITD menores que um determinado limiar superior. Esse processo pode ser realizado por um algoritmo iterativo para escolha de β , como por exemplo o apresentado em [4].

V. EXPERIMENTOS

Para ilustrar o desempenho do método proposto, foram realizadas simulações numéricas utilizando três técnicas de processamento: o MWF, definido pela minimização de (5); o MWF com preservação da função de transferência interauricular (MWF-ITF), definido pelas equações (5) e (6), para $I = 1$. α fixo, e as aproximações definidas em [1]; e a técnica proposta, denominada MWF-ITF-R, definida da mesma maneira que o MWF-ITF, contudo, utilizando α apresentado em (14) e a metodologia proposta na Seção IV-D. Sub-índices foram adicionados para se referir à fala (S), ruído (N) e orelhas esquerda (L) e direita (R).

1) *Cenário acústico*: O cenário acústico considerado é composto por fontes pontuais de fala e de interferência, e foi simulado usando respostas ao impulso (HRIR, *head related impulse response*) obtidas em uma câmara anecoica usando um manequim Brüel & Kjær tipo 4128-C sobre o qual foi acoplado um par de aparelhos auditivos retroauriculares [15]. Os aparelhos auditivos utilizados possuíam 3 microfones cada, resultando em $M_L + M_R = M = 6$. A fonte da fala foi assumida a uma distância de 0,8 m do manequim, com azimute de chegada de 0° ; a fonte interferente foi assumida com azimute de chegada de -60° (lado esquerdo) e distante 3,0 m do manequim. As HRIRs dos caminhos acústicos entre as fontes de fala e de interferência até os microfones dos aparelhos auditivos são representadas por $HRIR_S$ e $HRIR_N$, respectivamente.

2) *Sinais de fala e ruído*: Foram utilizados 32 trechos de fala obtidos do banco de dados em [16]. Como sinal da fonte interferente, utilizou-se um conjunto de 32 segmentos não sobrepostos de um ruído artificial com propriedades espectro-temporais semelhantes às da voz humana (sinal ICRA tipo I) [17]. A frequência de amostragem utilizada foi $f_s = 16$

kHz. Os componentes de fala e de interferência foram gerados convoluindo-se individualmente os sinais fala e de interferência com as HRIR associados aos azimutes de chegada das fontes sonoras. O sinal contaminado foi gerado somando-se os componentes de fala e interferência de maneira que a SNR global de entrada (SNR_{in}) fosse igual a 0 dB [8], o que resultou em uma SNR de -3 dB no ouvido esquerdo e de 2 dB no ouvido direito.

3) *Processamento*: O sinal contaminado foi processado em janelas de 128 amostras (ou 8 ms). Essas amostras foram ponderadas por uma janela de análise, definidas pela raiz quadrada da janela de Hanning [18], e a representação tempo-frequência do resultado foi obtida utilizando a transformada de Fourier de tempo curto com $K = 256$ pontos e preenchimento por zeros. A sobreposição entre trechos de amostras adjacentes foi de 50%. As matrizes de coerência Φ_v e Φ_y foram calculadas diretamente dos sinais de interferência e do sinal contaminado (simulando o uso de um detector de fala ideal), enquanto que a matriz de coerência da fala foi aproximada por $\Phi_x = \Phi_y - \Phi_v$. Para todos os métodos, os coeficientes \mathbf{w}_L e \mathbf{w}_R foram obtidos utilizando o algoritmo de otimização descrito em [19].

4) *Crítérios objetivos*: Os resultados da simulação foram avaliados de acordo com os erros na ILD (ΔILD_N) e na ITD (ΔITD_N) do ruído [20][21]. Adicionalmente, também são apresentados resultados associados à diferença entre as SNRs de entrada e saída na esquerda e direita, representadas por ΔSNR_L e ΔSNR_R , que medem a variação de SNR em cada orelha.

VI. RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados obtidos a partir de experimentos de simulação computacional e respectiva discussão.

A. Robustez a variações do ganho de Lombard

A Figura 1 apresenta a ΔILD_N e ΔITD_N para sinais processados pelo MWF, MWF-ITF e MWF-ITF-R em função do ganho quadrático médio de Lombard (\bar{g}^2). Observa-se que o desempenho do MWF-ITF convencional é significativamente afetado por variações no ganho Lombard, como consequência dos diferentes graus de homogeneidade dos termos que definem a sua função custo. Verifica-se também que o MWF e o MWF-ITF-R são aproximadamente invariantes a mudanças no \bar{g}^2 . Esse comportamento é esperado para o MWF visto que a função custo que o define é originalmente 2-homogênea. No caso do método proposto, a robustez a variações de \bar{g}^2 é decorrente do comportamento dinâmico de α_i definido em (14). Nas condições analisadas, o MWF apresenta os maiores valores de ΔILD_N e ΔITD_N , enquanto o MWF-ITF-R apresenta a menor distorção em toda faixa de valores de \bar{g}^2 . Este mesmo comportamento é verificado para ΔSNR_L e para ΔSNR_R . No lado esquerdo, o MWF é constante em $\Delta SNR_L = 27,9$ dB, enquanto que o MWF-ITF-R é aproximadamente constante em $\Delta SNR_L = 25,5$ dB; o MWF-ITF transita do MWF-ITF-R ao MWF à medida que \bar{g}^2 aumenta. Portanto, a técnica proposta apresenta um compromisso entre redução de ruído e preservação das pistas biauriculares.

B. Robustez a variações de SNR

A Fig. 2 apresenta a ΔILD_N e ΔITD_N em função da SNR_{in} . Os métodos MWF-ITF e MWF-ITF-R apresentam o mesmo desempenho em $SNR_{in} = -5$ dB. Pode-se observar que os resultados de ΔILD_N e ΔITD_N são afetados pela variação de SNR_{in} . Note-se entretanto, que as pistas biauriculares do ruído processado pelo método proposto sofrem variações significativas apenas no intervalo -50 dB $< SNR_{in} < -5$ dB, ou seja, apenas para situações de SNR_{in} inferiores a SNR_{in} utilizada no projeto do parâmetro α ($SNR_{in} = 0$ dB). Por outro lado, para o MWF-ITF convencional essa variação ocorre no intervalo -5 dB $< SNR_{in} < 60$ dB. Como resultado, o método proposto apresenta melhor preservação espacial na faixa mais comum de conversação para usuários de aparelhos auditivos (2 dB $< SNR_{in} < 14$ dB) [8].

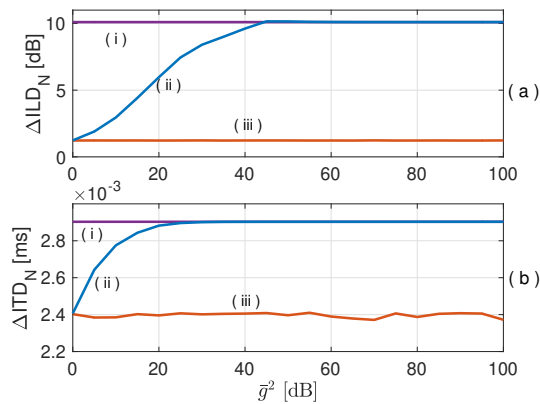


Fig. 1. Erros médios nas pistas biauriculares com relação a \bar{g}^2 em dB para o (i) MWF (roxo), (ii) MWF-ITF (azul), e (iii) MWF-ITF-R (vermelho). (a) ΔILD_N ; e (b) ΔITD_N .

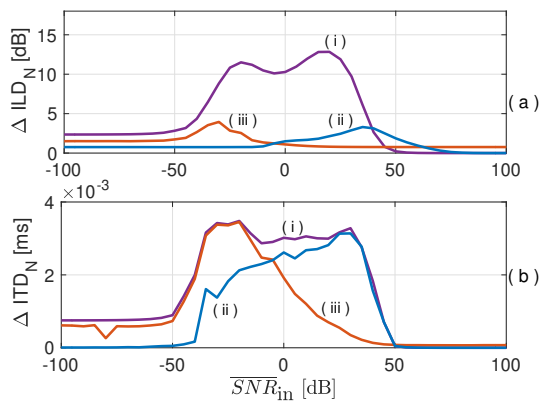


Fig. 2. Erros médios nas pistas biauriculares com relação a SNR_{in} para o (i) MWF (roxo), (ii) MWF-ITF (azul), e (iii) MWF-ITF-R (vermelho). (a) ΔILD_N ; e (b) ΔITD_N .

VII. CONCLUSÕES

Este trabalho propôs uma metodologia para projeto de um parâmetro dinâmico de ponderação em métodos de redução de ruído baseados no filtro de Wiener, com preservação do cenário acústico. Como resultado, há um aumento de robustez na preservação das pistas biauriculares às variações na potência dos sinais observados (efeito Lombard) e na SNR de entrada,

garantindo que o ponto de operação desejado seja mantido. Os resultados apresentados são de especial interesse para a área de aparelhos auditivos.

REFERÊNCIAS

- [1] B. Cornelis, S. Doclo, T. Bogaert, M. Moonen, e J. Wouters, “Theoretical analysis of binaural multimicrophone noise reduction techniques”, *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 2, pp. 342–355, 2009.
- [2] J. Werner e M. H. Costa, “A noise-reduction method with coherence enhancement for binaural hearing aids”, *J. Commun. Inf. Syst.*, vol. 35, no. 1, pp. 338–348, 2020.
- [3] F. P. Itturriet e M. H. Costa, “Perceptually relevant preservation of interaural time differences in binaural hearing aids”, *IEEE Trans. Audio Speech Lang. Process.*, vol. 27, no. 4, pp. 753–764, 2019.
- [4] D. Marquardt e S. Doclo, “Interaural coherence preservation for binaural noise reduction using partial noise estimation and spectral postfiltering”, *IEEE Trans. Audio Speech Lang. Process.*, vol. 26, no. 7, pp. 1261–1274, 2018.
- [5] V.H. Marquardt e S. Doclo, “Interaural coherence preservation in multi-channel Wiener filtering-based noise reduction for binaural hearing aids”, *IEEE Trans. Audio Speech Lang. Process.*, vol. 23, no. 12, pp. 2162–2176, 2015.
- [6] S. Doclo, R. Dong, T.J. Klasan, J. Wouters, S. Haykin, e M. Moonen, “Extension of the multi-channel Wiener filter with localisation cues for noise reduction in binaural hearing aids”, em *Proc. IWAENC*, 2005, pp. 221–224.
- [7] M. H. Costa e P. A. Naylor, “ILD preservation in the multichannel Wiener filter for binaural hearing aid applications”, em *Proc. EUSIPCO*, 2014, pp. 636–640.
- [8] Y.-H. Wu, E. Stangl, O. Chipara, S. S. Hasan, A. Welhaven, e J. Oleson, “Characteristics of real-world signal-to-noise ratios and speech listening situations of older adults with mild-to-moderate hearing loss”, *Ear and Hear.*, vol. 39, no. 2, p. 293, 2018.
- [9] H. Brumm e S. A. Zollinger, “The evolution of the Lombard effect: 100 years of psychoacoustic research”, *Behaviour*, vol. 148, no. 11–13, pp. 1173–1198, 2011.
- [10] J.-C. Junqua, “The Lombard reflex and its role on human listeners and automatic speech recognizers”, *The Journal of the Acoustical Society of America*, vol. 93, no. 1, pp. 510–524, 1993.
- [11] R. Marxer, J. Barker, N. Alghamdi, e S. Maddock, “The impact of the Lombard effect on audio and visual speech recognition systems”, *Speech Commun.*, vol. 100, pp. 58–68, 2018.
- [12] J. H. Hansen, “Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition”, *Speech Commun.*, vol. 20, no. 1–2, pp. 151–173, 1996.
- [13] F. Brickell, “A theorem on homogeneous functions”, *J. London Math. Soc.*, vol. 1, no. 1, pp. 325–329, 1967.
- [14] J. Chen, J. Benesty, Y. Huang, e S. Doclo, “New insights into the noise reduction Wiener filter”, *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 4, pp. 1218–1234, 2006.
- [15] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, e B. Kollmeier, “Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses”, *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–10, 2009.
- [16] G. E. Henter *et al.*, “Repeated harvard sentence prompts corpus version 0.5”, 2014, disponível em: <https://datashare.is.ed.ac.uk/handle/10283/561?show=full>.
- [17] W. A. Dreschler, H. Verschuure, C. Ludvigsen, e S. Westermann, “ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment”, *Audiol.*, vol. 40, no. 3, pp. 148–157, 2001.
- [18] R. Crochiere, “A weighted overlap-add method of short-time Fourier analysis/synthesis”, *IEEE Trans. Audio Speech Lang. Process.*, vol. 28, no. 1, pp. 99–102, 28.
- [19] E. A. Habets e P. A. Naylor, “An online quasi-Newton algorithm for blind SIMO identification”, em *Proc. ICASSP*, 2010, pp. 2662–2665.
- [20] D. Marquardt, E. Hadad, S. Gannot, S. Doclo, “Theoretical analysis of linearly constrained multi-channel Wiener filtering algorithms for combined noise reduction and binaural cue preservation in binaural hearing aids”, *IEEE Trans. Audio Speech Lang. Process.*, vol. 23, no. 12, pp. 2384–2397, 2015.
- [21] T. Bogaert, S. Doclo, M. Moonen, e J. Wouters, “Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter”, em *Proc. ICASSP*, 2005, pp. 221–224.