

Redução de ruído binaural em tempo real com preservação da função de transferência interaural

Diego M. do Carmo, Fábio P. Itturriet e Márcio H. Costa

Resumo—Esse trabalho apresenta uma nova técnica para redução de ruído em aparelhos auditivos binaurais. É proposta uma aproximação do filtro de Wiener multicanal com preservação da função de transferência interaural (MWF-ITF) via método do gradiente descendente. Como resultado, obtém-se um algoritmo com custo computacional controlado, que estabelece um compromisso entre redução de ruído e preservação das pistas acústicas. Uma análise de complexidade computacional em termos do número de operações multiplica e acumula é apresentada. Avaliações através de métricas objetivas indicam as capacidades de redução de ruído e preservação das pistas espaciais em um campo acústico direcional composto por uma fonte de fala e outra de ruído.

Palavras-Chave—Redução de ruído, aparelhos auditivos, binaural, MWF-ITF.

Abstract—This work presents a new noise reduction method for binaural hearing aids. It is proposed an approximation to the multichannel Wiener filter with preservation of the noise interaural transfer function (MWF-ITF) through the steepest descent method. As a result, we developed a controlled computational cost algorithm, which establishes a trade-off between noise reduction and binaural cue preservation. An analysis of the computational complexity based on multiply and accumulate is provided. Experiments using objective metrics exemplify the noise reduction and spacial cue preservation capacities in a directional acoustic field comprised by one speech and one noise sources.

Keywords—Noise reduction, hearing aids, binaural, MWF-ITF.

I. INTRODUÇÃO

Em grande número de situações cotidianas, seres humanos são submetidos a ambientes com elevados níveis de ruído acústico. Para se comunicar de maneira satisfatória em alguns desses cenários, é fundamental que as fontes sonoras mantenham a separação espacial existente entre elas. Dessa forma, mecanismos naturais de inibição espacial associados à direção da fonte indesejada entram em ação. Esse processo é realizado pelo cérebro utilizando informações extraídas do som que chega em cada uma das orelhas. Essas informações são conhecidas como pistas binaurais. Duas dessas pistas são a diferença de tempo interaural (ITD – *Interaural Time Difference*) e a diferença de nível interaural (ILD – *Interaural Level Difference*). De acordo com a teoria duplex, a ITD é a pista dominante nas frequências inferiores a 1,5 kHz, e a ILD é dominante nas frequências superiores a este valor [1] [2].

Diego M. do Carmo, Fábio P. Itturriet e Márcio H. Costa, Universidade Federal de Santa Catarina (UFSC), Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis-SC, Brasil, E-mails: diego.carmoh@gmail.com, fabio.itturriet@posgrad.ufsc.br e costa@eel.ufsc.br. Fábio Itturriet também é professor do Instituto Federal de Educação, Ciência e Tecnologia Sul-riograndense (IFSul), Charqueadas-RS, Brasil. Este trabalho foi parcialmente financiado pelo CNPq (162573/2014-6 e 304867/2015-2) e Capes (1755792).

Aparelhos auditivos são dispositivos utilizados por pessoas com perdas auditivas leves e moderadas para compensação de suas limitações. No caso de perdas bilaterais, aparelhos auditivos binaurais têm sido cada vez mais utilizados em decorrência de sua capacidade de compartilhamento de informações. Além disso, esses aparelhos contam com múltiplos microfones, o que aumenta a amostragem espacial e espectral das cenas acústicas [3].

Segundo [4], conversar em ambientes ruidosos é considerado como a pior situação enfrentada pelos usuários de aparelhos auditivos. Por isso, a utilização de algoritmos de redução de ruído se torna obrigatória. Para tanto, técnicas de filtragem espacial (*beamforming*) [5] [6] são bastante utilizadas, mas requerem um conhecimento prévio da localização e/ou das funções de transferência entre a fonte sonora e os microfones dos aparelhos auditivos. Quando essas considerações não são atendidas, tais técnicas apresentam significativa degradação de desempenho.

Outra técnica bastante utilizada para redução de ruído em aparelhos auditivos é o filtro de Wiener Multicanal (MWF). Esse método apresenta baixa distorção do sinal de interesse (fala) [7] [8], entretanto, ele modifica as pistas binaurais do ruído residual. Dessa forma, ocorre uma diferença entre a localização percebida da fonte de ruído e sua localização original. Como consequência, usuários de aparelhos auditivos reportam confusão espacial em função do descasamento entre pistas visuais e auditivas. Além de desconforto, essa confusão pode colocar os usuários em risco. Em [9], foi apresentado uma proposta de um método de redução de ruído visando a preservação das pistas binaurais do ruído (ITD e ILD) através da preservação da função de transferência interaural (ITF – *Interaural Transfer Function*). Essa proposta consiste na minimização de uma função custo, definida pela soma ponderada entre a função custo do MWF e uma função associada ao descasamento entre a ITF dos sinais recebidos e os processados. Essa abordagem não possui uma solução fechada para o cálculo dos coeficientes ótimos, por isso requer técnicas de otimização numérica que podem demandar um elevado tempo de computação, consumindo recursos nem sempre disponíveis no aparelho auditivo para processamento em tempo real.

Nesse trabalho é apresentada uma aproximação da técnica de MWF com preservação da ITF da fonte de ruído [9]. Essa proposta permite controlar o custo computacional envolvido, possibilitando sua implementação em tempo real.

II. SISTEMA DE PROCESSAMENTO BINAURAL

O sistema de processamento binaural, mostrado na Figura 1, é composto por M_L microfones no aparelho esquerdo (L) e

M_R microfones no aparelho direito (R). O vetor de amostras dos sinais captados pelos microfones, representado no domínio da Transformada de Fourier de tempo curto (STFT – *short time Fourier transform*), é dado por:

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

onde $y_{l,m}(\lambda, k)$ é o sinal sonoro captado pelo microfone $m = \{1, \dots, M_l\}$ do aparelho auditivo $l = \{L, R\}$; $\mathbf{y}(\lambda, k)$ é um vetor de dimensão $M \times 1$, sendo $M = M_L + M_R$; o parâmetro λ denota o índice de cada quadro de tempo e k denota o índice de frequência. O vetor $\mathbf{y}(\lambda, k)$ é acessível para ambos aparelhos, uma vez que compartilham informações através de um canal de comunicação sem fio do tipo *full duplex*. O vetor de amostras também pode ser definido como:

$$\mathbf{y}(\lambda, k) = \mathbf{x}(\lambda, k) + \mathbf{v}(\lambda, k), \quad (2)$$

em que $\mathbf{x}(\lambda, k)$ é o vetor de amostras do sinal de fala e $\mathbf{v}(\lambda, k)$ é o vetor de amostras do ruído aditivo, que representa os sinais indesejados presentes no cenário acústico. O microfone frontal de cada aparelho auditivo é designado como microfone de referência. Os sinais nesses microfones são definidos como:

$$y_{l,ref}(\lambda, k) = x_{l,ref}(\lambda, k) + v_{l,ref}(\lambda, k) = \mathbf{q}_l^T \mathbf{y}(\lambda, k), \quad (3)$$

onde *ref* refere-se ao índice do microfone de referência e \mathbf{q}_l é um vetor que possui valor 1 na posição deste microfone, i.e. $\mathbf{q}_l(ref) = 1$, e zero nas demais posições.

A saída do sistema de processamento é dada pelo produto interno entre o vetor de sinal de entrada $\mathbf{y}(\lambda, k)$ e os vetores complexos de pesos $\mathbf{w}_L(\lambda, k)$ e $\mathbf{w}_R(\lambda, k)$, de forma que:

$$\mathbf{z}_l(\lambda, k) = \mathbf{w}_l^H(\lambda, k) \mathbf{y}(\lambda, k), \quad (4)$$

em que $(\cdot)^H$ denota o transposto hermitiano. Os sinais no domínio tempo são reconstruídos através do método de *overlap and add*, a partir de $\mathbf{z}_L(\lambda, k)$ e $\mathbf{z}_R(\lambda, k)$. O conjunto de coeficientes do sistema de processamento é representado por $\mathbf{w} = [\mathbf{w}_L^T \mathbf{w}_R^T]^T$, com dimensão $2M \times 1$.

Por fim, definem-se as matrizes de correlação do sinal de entrada, $\mathbf{R}_y(k)$; de fala, $\mathbf{R}_x(k)$; e de ruído, $\mathbf{R}_v(k)$:

$$\begin{aligned} \mathbf{R}_y(k) &= E\{\mathbf{y}(\lambda, k) \mathbf{y}^H(\lambda, k)\}, \\ \mathbf{R}_x(k) &= E\{\mathbf{x}(\lambda, k) \mathbf{x}^H(\lambda, k)\}, \\ \mathbf{R}_v(k) &= E\{\mathbf{v}(\lambda, k) \mathbf{v}^H(\lambda, k)\}. \end{aligned} \quad (5)$$

III. FUNÇÃO CUSTO BINAURAL

Como mencionado anteriormente, o MWF altera as pistas binaurais do ruído aditivo e, conseqüentemente, a percepção da cena acústica original. A fim de se preservar a posição original do ruído, a seguinte função custo foi proposta em [8]:

$$J_T(k) = J_{MWF}(k) + \alpha(k) J_{ITF}(k). \quad (6)$$

onde o termo $J_{MWF}(k)$ denota a função custo do filtro multicanal de Wiener (MWF) e está associado à redução da potência do ruído. O termo $J_{ITF}(k)$ denota a função custo auxiliar, associada ao controle da ITF do ruído e conseqüente controle da ITD e da ILD. A constante $\alpha(k)$ é um fator de

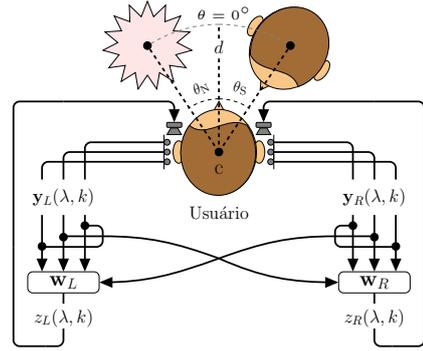


Fig. 1. Sistema de processamento binaural constituído por um par de aparelhos auditivos com canal de comunicação *full duplex*.

ponderação, responsável por controlar o compromisso entre redução de ruído e a preservação da pistas binaurais do ruído na função custo total $J_T(k)$. Para simplificação da notação, as dependências de (λ, k) e (k) serão omitidas nas equações subsequentes. A função custo do MWF é dada por [10]:

$$J_{MWF} = E \left\{ \left\| \begin{bmatrix} x_{L,ref} - \mathbf{w}_L^H \mathbf{y} \\ x_{R,ref} - \mathbf{w}_R^H \mathbf{y} \end{bmatrix} \right\|^2 \right\}, \quad (7)$$

em que $E\{\cdot\}$ é o operador valor esperado e $\|\cdot\|^2$ é a norma euclidiana ao quadrado. A função custo de preservação da ITF do ruído é dada por [11]:

$$J_{ITF} = E \left\{ \left\| \begin{bmatrix} \mathbf{w}_L^H \mathbf{v} \\ \mathbf{w}_R^H \mathbf{v} \end{bmatrix} - \text{ITF}_{in} \right\|^2 \right\}, \quad (8)$$

em que $\text{ITF}_{in} = \mathbf{q}_L^T \mathbf{R}_v \mathbf{q}_L / \mathbf{q}_R^T \mathbf{R}_v \mathbf{q}_R$. Assumindo-se estacionaridade dos sinais e independência entre as fontes de fala e de ruído, tem-se que (7) resulta em:

$$J_{MWF} = P - \mathbf{w}^H \mathbf{r} - \mathbf{r}^H \mathbf{w} + \mathbf{w}^H \mathbf{R}_{yy} \mathbf{w}, \quad (9)$$

em que

$$\begin{aligned} P &= E\{|x_{L,ref}|^2\} + E\{|x_{R,ref}|^2\}, \quad \mathbf{r} = \begin{bmatrix} \mathbf{R}_x \mathbf{q}_L \\ \mathbf{R}_x \mathbf{q}_R \end{bmatrix}, \\ \mathbf{R}_{yy} &= \begin{bmatrix} \mathbf{R}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_y \end{bmatrix}. \end{aligned} \quad (10)$$

Manipulando-se (8) sob as mesmas considerações anteriores, obtém-se:

$$J_{ITF} = \frac{\mathbf{w}^H \mathbf{R}_{vt} \mathbf{w}}{\mathbf{w}^H \mathbf{R}_{vr} \mathbf{w}}, \quad (11)$$

em que

$$\mathbf{R}_{vt} = \begin{bmatrix} \mathbf{R}_v & -\text{ITF}_{in}^* \mathbf{R}_v \\ -\text{ITF}_{in} \mathbf{R}_v & |\text{ITF}_{in}|^2 \mathbf{R}_v \end{bmatrix}, \quad \mathbf{R}_{vr} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_v \end{bmatrix}. \quad (12)$$

IV. TÉCNICA PROPOSTA

Embora as técnicas numéricas tradicionais para a minimização da função custo apresentada em (6) apresentem resultados satisfatórios para a obtenção de vetores ótimos \mathbf{w}_L e \mathbf{w}_R , elas são computacionalmente custosas. Sendo assim, podem inviabilizar a implementação em tempo real em aparelhos auditivos. De forma a superar esse problema, um algoritmo eficiente para redução e preservação da ILD do ruído foi proposto em [12]. Contudo, como a ILD é uma

pista preponderante apenas acima de 1,5 kHz, seu uso em baixas frequências contribui pouco para a preservação das pistas espaciais.

A. Equações Recursivas

Nessa seção é apresentada uma proposta de um algoritmo de tempo real com custo computacional controlável para a minimização da função custo descrita em (6).

Partindo-se do pressuposto que o objetivo primário da função custo J_T seja a redução de ruído, representada pela parcela J_{MWF} em (6) e, portanto, que a constante α não é muito elevada, a convexidade de J_{MWF} é assumida como preponderante sobre as características de J_{ITF} . Dessa forma, propõe-se o uso do método do gradiente descendente [13]:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{\mu}{2} \nabla_{\mathbf{w}^*} J_T, \quad (13)$$

em que $(\cdot)^*$ é a operação de conjugado complexo e μ é a constante de adaptação que controla a velocidade de convergência e a estabilidade do algoritmo. Substituindo-se (6) em (13), obtém-se:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{\mu}{2} \nabla_{\mathbf{w}^*} J_{MWF} - \frac{\mu\alpha}{2} \nabla_{\mathbf{w}^*} J_{ITF}. \quad (14)$$

Os gradientes de J_{MWF} e de J_{ITF} (Apêndice I) com relação a \mathbf{w}^* são dados por:

$$\begin{aligned} \nabla_{\mathbf{w}^*} J_{MWF} &= 2(\mathbf{R}_{yy} \mathbf{w} - \mathbf{r}), \\ \nabla_{\mathbf{w}^*} J_{ITF} &= 2 \frac{(\mathbf{w}^H \mathbf{R}_{vr} \mathbf{w}) \mathbf{R}_{vt} \mathbf{w} - (\mathbf{w}^H \mathbf{R}_{vt} \mathbf{w}) \mathbf{R}_{vr} \mathbf{w}}{(\mathbf{w}^H \mathbf{R}_{vr} \mathbf{w})^2}. \end{aligned} \quad (15)$$

Usando-se (15) em (14), chega-se a equação de atualização do algoritmo proposto:

$$\begin{aligned} \mathbf{w}_{n+1} &= (\mathbf{I} - \beta \mathbf{R}_{yy}) \mathbf{w}_n + \beta \mathbf{r} \\ &+ \gamma \left(\frac{(\mathbf{w}_n^H \mathbf{R}_{vt} \mathbf{w}_n) \mathbf{R}_{vr} \mathbf{w}_n - (\mathbf{w}_n^H \mathbf{R}_{vr} \mathbf{w}_n) \mathbf{R}_{vt} \mathbf{w}_n}{(\mathbf{w}_n^H \mathbf{R}_{vr} \mathbf{w}_n)^2} \right), \end{aligned} \quad (16)$$

em que $\beta = \mu$, $\gamma = \mu\alpha$ e \mathbf{I} é a matriz identidade de dimensão $2M \times 2M$.

B. Custo computacional

O custo computacional do algoritmo proposto é medido através de operações multiplica e acumula (MAC – *multiply and accumulate*). Considera-se que a inversão de um escalar complexo (1/c) requer o computo de uma 1 soma, 1 divisão e 2 multiplicações reais, e a inversão de um escalar real tem custo de 16 operações MAC [12]. As operações de conjugado, parte real de um número complexo e multiplicação por múltiplos de 2 não alteram o custo computacional. A Tabela I apresenta o pseudocódigo e o custo computacional em função do número total de microfones (M), de pontos da STFT (N) e do número de iterações (P). $\hat{\mathbf{R}}_y$, $\hat{\mathbf{R}}_x$ e $\hat{\mathbf{R}}_v$ representam estimativas das matrizes de correlação.

V. EXPERIMENTOS DE SIMULAÇÃO

Para ilustrar a capacidade do algoritmo proposto na redução de ruído e na preservação de pistas de localização, é apresentado um experimento com cenário acústico simulado.

TABELA I

PSEUDOCÓDIGO E CUSTO COMPUTACIONAL DO ALGORITMO PROPOSTO EM FUNÇÃO DO NÚMERO TOTAL DE MICROFONES (M), DE PONTOS DA STFT (N) E DO NÚMERO DE ITERAÇÕES (P).

	Algoritmo	MAC
	INICIALIZAÇÃO: $\beta, \gamma, \mathbf{w}, \hat{\mathbf{R}}_y, \hat{\mathbf{R}}_x, \hat{\mathbf{R}}_v$.	
ATUALIZAÇÃO	Para $k = 1, 2, \dots, \frac{N}{2}$	
	$ITF_{in}(k) = \frac{\mathbf{q}_L^H \hat{\mathbf{R}}_v(k) \mathbf{q}_R}{\mathbf{q}_R^H \hat{\mathbf{R}}_v(k) \mathbf{q}_R}$	18
	$\mathbf{a}_L = \beta \hat{\mathbf{R}}_x(k) \mathbf{q}_L$, $\mathbf{a}_R = \beta \hat{\mathbf{R}}_x(k) \mathbf{q}_R$	$4MN/2$
	$\mathbf{A} = \mathbf{I} - \beta \hat{\mathbf{R}}_y$	$(M^2 + M)N/2$
	Para $n = 1, 2, \dots, P$	
	$\mathbf{b}_L = \mathbf{A} \mathbf{w}_R(k)$, $\mathbf{b}_R = \mathbf{A} \mathbf{w}_R(k)$	$(12M^2 - 10M)PN/2$
	$\mathbf{c}_L = \hat{\mathbf{R}}_v \mathbf{w}_L(k)$, $\mathbf{c}_R = \hat{\mathbf{R}}_v \mathbf{w}_R(k)$	$(12M^2 - 10M)PN/2$
	$a = \mathbf{w}_L^H(k) \mathbf{c}_L$, $b = \mathbf{w}_R^H(k) \mathbf{c}_R(k)$	$8MPN/2$
	$c = \mathbf{w}_R^H \mathbf{c}_L$, $d = -\gamma/b^2$	$(4M + 19)P$
	$\mathbf{d}_L = d(b(\mathbf{c}_L - ITF_{in}^* \mathbf{c}_R))$	$(8M + 1)PN/2$
$e = 2cRe\{ITF_{in}\} - a$	$2PN/2$	
$\mathbf{d}_R = d(e \mathbf{c}_R - b ITF_{in} \mathbf{c}_L)$	$16MPN/2$	
$\begin{bmatrix} \mathbf{w}_L(k) \\ \mathbf{w}_R(k) \end{bmatrix} = \begin{bmatrix} \mathbf{b}_L \\ \mathbf{b}_R \end{bmatrix} + \begin{bmatrix} \mathbf{d}_L \\ \mathbf{d}_R \end{bmatrix}$	$4MPN/2$	
	Total	$[(12M^2 + 20M + 21)P + (M^2 + 5M + 18)]N/2$

A. Cenário Acústico Virtual

O cenário acústico considerado nesse trabalho é semelhante ao ilustrado na Figura 1 e composto por uma fonte de fala posicionada no azimute $\theta_S = 0^\circ$ e uma fonte de ruído no azimute $\theta_N = 60^\circ$. Ambas as fontes encontram-se a uma distância $d = 3$ m do ponto central (C) e com 0° de elevação em relação ao plano transversal. O sistema de processamento binaural conta com $M_L = M_R = 3$ microfones.

Para a criação do cenário acústico descrito, os sinais de fala e ruído foram individualmente processados por filtros digitais, cujas respostas ao impulso foram obtidas em um banco de dados de funções de transferência binaurais [14]. A configuração de obtenção dessas funções de transferência utilizou: câmara anecoica, manequim Bruel & Kjaer tipo 4128-C [15] e um par de aparelhos auditivos reatroauricular.

B. Sinais Utilizados

Como sinal de interesse, utilizou-se um trecho de 3 s iniciais de silêncio seguido por 14,7 s de fala masculina selecionada de [16]. Para caracterizar o sinal emitido pela fonte indesejada, utilizou-se um ruído com características temporais e espectrais semelhantes às da fala humana [17]. Os espectros de potência dos sinais de fala e de ruído são apresentados na Figura 2. A frequência de amostragem (f_s) é 16 kHz. Os sinais foram transformados para o domínio da frequência através da STFT com $N = 256$ pontos e reconstruídos de acordo com o algoritmo *weighted-overlap-add* descrito em [18]. Os sinais de fala e ruído filtrados foram artificialmente combinados de maneira que a SNR no ouvido mais próximo à fonte de ruído (ouvido esquerdo) fosse 0 dB.

C. Parâmetros do Algoritmo Adaptativo

Nos experimentos realizados, o passo de adaptação foi definido como $\mu = 0,04$. A estimação da matriz de correlação do sinal de entrada ($\hat{\mathbf{R}}_y$) foi feita nos períodos de fala contaminada por ruído. A estimação da matriz de correlação

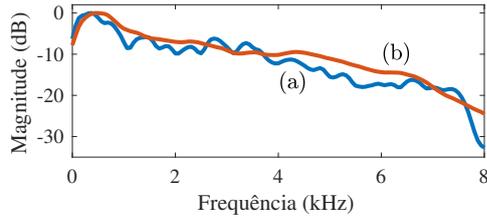


Fig. 2. Densidade espectral de potência do sinal de fala (a) e do ruído (b).

do ruído (\mathbf{R}_v) foi feita nos períodos contendo apenas ruído. O método utilizado para estimação dessas matrizes é apresentado em [19], utilizando-se coeficientes de esquecimento $\eta_y = \eta_x = \eta_v = 0,999$, equivalentes a uma constante de tempo de 4s. Utilizou-se um detector de atividade sonora (VAD – *Voice Activity Detector*) feito manualmente para definição dos períodos de fala e fala contaminada por ruído. Definiu-se o número de iterações como $P = 50$.

D. Métricas de Avaliação

Para a avaliação dos resultados apresentados, cinco métricas objetivas foram utilizadas. Duas delas referem-se à preservação das pistas acústicas espaciais, sendo elas: (1) a ΔILD , que mede a diferença entre as ILDs dos sinais na entrada e na saída do sistema de processamento (calculada apenas para frequências superiores a 1,5 kHz) [20] e a (2) ΔIPD ¹, que mede a variação entre as diferenças de fase interaurais (IPD – *Interaural Phase Difference*) na entrada e na saída do sistema de processamento (calculada apenas para frequências inferiores a 1,5 kHz). As demais métricas são: (3) variação da avaliação perceptual da qualidade da fala (ΔPESQ – *Wideband Perceptual evaluation of speech quality*) [21], que mede a qualidade percebida da fala; (4) a variação da relação sinal ruído entre o sinal processado e o sinal contaminado (ΔSNR), que mede a redução de ruído global; e (5) o ganho ponderado de inteligibilidade (ΔISNR) [22], que estima a variação na inteligibilidade do sinal de fala.

O sinal contaminado de entrada apresenta os seguintes valores de referência: $\text{PESQ}_L=1,1$, $\text{PESQ}_R=1,2$, $\text{ISNR}_L=0,9$ dB, $\text{ISNR}_R=10,3$ dB, $\text{SNR}_L=0$ dB e $\text{SNR}_R=3,1$ dB.

VI. RESULTADOS E DISCUSSÃO

A Figura 3 apresenta o impacto da variação do parâmetro α nas métricas relacionadas à orelha esquerda e à fonte de ruído. As métricas associadas à orelha direita e à fonte de fala foram suprimidas por questões de espaço.

Para a situação em que $\alpha \rightarrow 0$ (ponto mais à esquerda dos gráficos apresentados), os resultados obtidos correspondem ao processamento do sinal pela solução de Wiener isoladamente. Nessa situação, é claramente visível o aumento da relação sinal-ruído na orelha esquerda ($\Delta\text{SNR}_L=21,6$ dB) e o consequente aumento da qualidade ($\Delta\text{PESQ}_L=1,6$) e inteligibilidade ($\Delta\text{ISNR}_L=21,8$ dB) do sinal de fala processado. Comportamento semelhante é verificado na orelha direita. Entretanto, para essa mesma condição, é verificado um valor elevado nas métricas de localização ($\Delta\text{IPD}_N=0,3$ dB e $\Delta\text{ILD}_N=8,8$ dB),

¹O valor principal da IPD é dado por $\text{IPD} = \frac{2\pi f_s k}{N} \text{ITD}$.

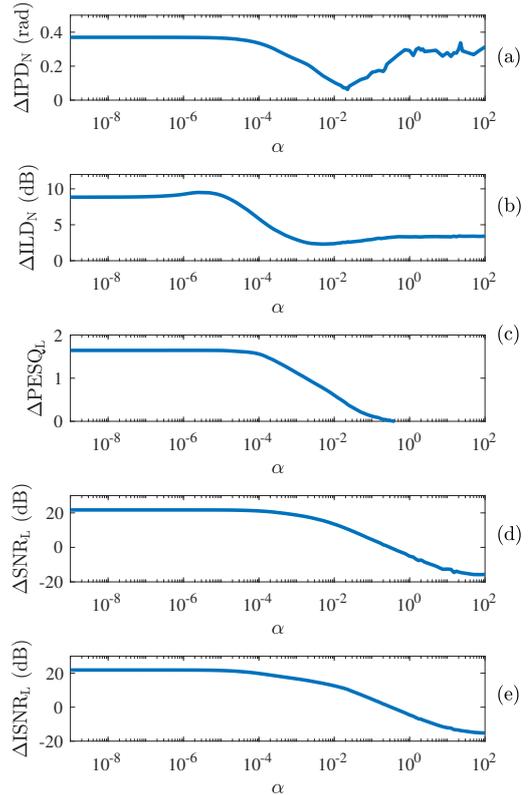


Fig. 3. Métricas objetivas em função do parâmetro α : (a) ΔIPD_N , (b) ΔILD_N , (c) PESQ_L , (d) ΔSNR_L e (e) ΔISNR_L .

o que indica uma alteração nas pistas acústicas espaciais do ruído processado. Situação semelhante foi verificada em [12], na qual experimentos subjetivos comprovaram uma alteração perceptual correlata da posição da fonte de ruído em voluntários. A medida que α cresce, e a J_{ITF} ganha importância dentro de J_T , os erros ΔILD e ΔIPD do ruído diminuem, principalmente na região $2,5 \times 10^{-4} < \alpha < 2,5 \times 10^{-2}$. Valores de α dentro dessa faixa resultam em percepção espacial associada à localização real da fonte de ruído². Em contrapartida, com o aumento de α há uma degradação significativa das demais métricas de qualidade/inteligibilidade. Adicionalmente, com o aumento de α verifica-se um aumento de ΔILD e ΔIPD da fala. Essas observações indicam a existência de um compromisso entre redução de ruído e preservação de suas pistas acústicas.

A Tabela II apresenta as métricas obtidas com sinal de fala processado por uma implementação em *batch*³ da equação (6) para: $\alpha = 0$, o que resulta no MWF convencional, e para $\alpha = 2,5 \times 10^{-3}$, o que resulta no MWF-ITF [11]. Também são apresentados os índices de qualidade para o algoritmo proposto utilizando os mesmos parâmetros ($\alpha = 0 \rightarrow \text{MWF}_{\text{SD}}$ e $\alpha = 2,5 \times 10^{-3} \rightarrow \text{MWF-ITF}_{\text{SD}}$).

Os resultados obtidos mostram que o algoritmo proposto (MWF-ITF_{SD}) apresenta uma diminuição da capacidade de

²Experimentos informais realizados pelos autores mostram que variações de localização das fontes sonoras são perceptíveis quando $\Delta\text{ILD} > 4$ dB

³Estimação a priori das matrizes de correlação utilizando o sinal completo e posterior filtragem do sinal contaminado com vetor de coeficientes fixo (processamento *off-line*) [24]

redução de ruído em relação à implementação em *batch*. Entretanto, ainda fornecendo um compromisso aceitável entre redução de ruído e preservação da ILD e da IPD. Note que ambas as implementações do MWF (*batch/online*) apresentam altos erros de ΔILD_N (>4 dB) e ΔIPD_N , indicando uma mudança no azimute percebido da fonte de ruído. O método proposto apresenta um aumento médio de 16,2 dB de SNR em ambas as orelhas, em relação ao sinal de fala contaminado, mantendo $\Delta ILD_N < 4$ dB. Adicionalmente, verifica-se um aumento acima de 1 PESQ em ambos os ouvidos, sendo que variações acima de 0,2 PESQ são perceptíveis em testes subjetivos [23].

TABELA II
MÉTRICAS OBJETIVAS PARA O SINAL DE FALA PROCESSADO.

	MWF	MWF-ITF	MWF _{SD}	MWF-ITF _{SD}
ΔIPD_S	0,5	0,1	0,3	0,1
ΔILD_S	0,1	3,8	0,2	4,4
ΔIPD_N	0	0	0	0
ΔILD_N	11,9	1,5	8,8	2,3
$\Delta PESQ_L$	2,2	1,6	1,6	0,9
$\Delta PESQ_R$	2,2	2,1	1,6	1,5
ΔSNR_L	26,8	20,9	21,6	17,1
ΔSNR_R	23,6	22,5	19,3	18,4
$\Delta ISNR_L$	27,8	18,9	21,8	15,2
$\Delta ISNR_R$	18,9	15,7	15	11,5

VII. CONCLUSÃO

Este trabalho apresentou a proposta de um algoritmo de redução de ruído para aparelhos auditivos binaurais baseado no método do gradiente estocástico. O objetivo é possibilitar o controle da complexidade computacional do algoritmo pelo projetista, mantendo um compromisso aceitável entre redução de ruído e preservação das pistas espaciais. O desempenho do algoritmo foi ilustrado através de um experimento em um cenário acústico simulado composto por uma fonte de fala e uma de ruído. Os resultados obtidos indicam a capacidade concomitante de redução de ruído e preservação de pistas espaciais.

REFERÊNCIAS

[1] T. T. Sandel, D. C. Feddersen e L. A. Jeffress, "Localization of sound from single and paired sources," *J. Acoust. Soc. Am.*, vol. 27, n. 5, pp. 842-852, 1955.

[2] A. W. Mills, "Lateralization of high frequency tones," *J. Acoust. Soc. Am.*, vol. 32, n. 1, pp. 132-134, 1960.

[3] D. P. Welker, J. E. Greenberg e P. M. Zurek, "Microphone-array hearing aids with binaural output-part ii: A two-microphone adaptive system," *IEEE Trans. Speech and Audio Processing*, vol. 5, n. 6, pp. 543-551, 1997.

[4] EHIMA, European Hearing Instrument Manufacturers Association, Hearing Aids improve Hearing-and a LOT more, Trends derived from the EuroTrak databases 2009-2015, [Online] Disponível em: <https://www.ehima.com/documents/>.

[5] B. V. Veen e K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, n. 2, pp. 4-24, 1988.

[6] L. J. Griffiths e C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. 30, n. 1, pp. 27-34, 1982.

[7] S. Doclo e M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Processing*, vol. 50, n. 9, pp. 2230-2244, 2002.

[8] S. Doclo, A. Spriet, J. Wouters e M. Moonen, "Speech Distortion Weighted Multichannel Wiener Filtering Techniques for Noise Reduction," in *Speech enhancement*, J. Benesty, S. Makino e J. Chen, Eds. Heidelberg, Berlin: Springer-Verlag, 2005, pp. 199-228.

[9] T. J. Klases, S. Doclo, T. Bogaert, M. Moonen e J. Wouters, "Binaural multichannel Wiener filtering for hearing-aids: preserving interaural time and level-differences," in *Proc. IEEE Int. Conf. Acoustics, Speech, Lang. Process.* vol. 5, pp. 145-148, 2006.

[10] T. J. Klases, M. Moonen, T. Bogaert e J. Wouters, "Preservation of interaural time delay for binaural hearing aids through multi-channel Wiener filtering based noise reduction," in *Proc. ICASSP*, pp. 29-32, 2005.

[11] S. Doclo, T. J. Klases, T. Van den Bogaert, J. Wouters e M. Moonen, "Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions," *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, 2006.

[12] D. M. do Carmo e M. H. Costa, "Online approximation of the multi-channel Wiener filter with preservation of interaural level difference for binaural hearing-aids," *Comput. Biol. Med.*, vol. 95, pp. 188-197, 2018.

[13] S.S. Haykin, *Adaptive Filter Theory*, 3^o ed., Prentice-Hall, 1996.

[14] H. Kayser, S.D. Ewert, J. Anemüller, 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 Proc.*, pp. 1-10, 2009.

[15] Product Data, *Head and Torso Simulator Types 4128-C and 4128-D*, Brüel & Kjær, 2014.

[16] ITU-T P.50, Telephone Transmission Quality, Telephone Installations, Local Line Networks: Objective Measuring Apparatus - Artificial Voices, Appendix I: test signals, 1998.

[17] W. A. Dreschler, H. Verschuure, C. Ludvigsen, S. Westermann, "ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment," *Audiology*, vol. 40, n. 3, pp. 148-157, 2001.

[18] R.E. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis," *IEEE Trans. Acoust. Speech Signal Process.* Vol. 28, n. 1, pp. 99-102, 1980.

[19] J.I. Marin-Hurtado, D.V. Anderson, "Practical MWF-based noise-reduction methods for binaural hearing aids," in *Proc. STSIVA*, pp. 1-8, 2012.

[20] Costa, M. H., Naylor, P. A. "ILD preservation in the multichannel wiener filter for binaural hearing aid applications," *IEEE Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*, pp. 636-640, 2014.

[21] ITU-T, P.862, Telephone Transmission Quality, Telephone Installations, Local Line Networks: Revised Annex a - Reference Implementations and Conformance Testing for ITU-T Recs P.862, P.862.1 and P.862.2, 2005.

[22] J. Greenberg, J. Peterson, P. Zurek, "Intelligibility weighted measures of speech to interference ratio and speech system performance," *J. Acoust. Soc. Am.*, vol. 94, n. 5, pp. 3009-3010, 1993.

[23] J.C. Servetti, Martin, Error tolerant MAC extension for speech communications over 802.11 WLANs, in *Proc. VTC*, vol. 4, pp. 2330-2334, 2005.

[24] T. J. Klases, T. Van den Bogaert, M. Moonen, J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Process.*, vol. 55, n. 4 pp. 1579-1585, 2007.

APÊNDICE I

GRADIENTE DA FUNÇÃO CUSTO J_{ITF}

A equação em (8) pode ser reescrita como:

$$J_{ITF} = \frac{\mathbf{w}^H \mathbf{R}_{vt} \mathbf{w}}{\mathbf{w}^H \mathbf{R}_{vr} \mathbf{w}} = \frac{f(\mathbf{w})}{g(\mathbf{w})}. \quad (17)$$

Logo, o gradiente da J_{ITF} em relação a \mathbf{w}^* é dado por:

$$\nabla_{\mathbf{w}^*} J_{ITF} = \frac{g(\mathbf{w}) \nabla_{\mathbf{w}^*} f(\mathbf{w}) - f(\mathbf{w}) \nabla_{\mathbf{w}^*} g(\mathbf{w})}{g(\mathbf{w})^2} \quad (18)$$

onde

$$\nabla_{\mathbf{w}^*} f = 2\mathbf{R}_{vt} \mathbf{w}, \quad \nabla_{\mathbf{w}^*} g = 2\mathbf{R}_{vr} \mathbf{w}. \quad (19)$$

Utilizando-se (17) e (19) em (18), obtém-se:

$$\nabla_{\mathbf{w}^*} J_{ITF} = 2 \frac{(\mathbf{w}^H \mathbf{R}_{vr} \mathbf{w}) \mathbf{R}_{vt} \mathbf{w} - (\mathbf{w}^H \mathbf{R}_{vt} \mathbf{w}) \mathbf{R}_{vr} \mathbf{w}}{(\mathbf{w}^H \mathbf{R}_{vr} \mathbf{w})^2} \quad (20)$$