

Implementação em tempo real de um sistema de redução de ruído binaural com preservação da função de transferência interaural

Arthur D. Reys, Diego M. Carmo, Gabriel Souza, Márcio H. Costa

Resumo—Esse trabalho apresenta a implementação em tempo real de um algoritmo adaptativo para redução de ruído em aparelhos auditivos com preservação de pistas binaurais. O método do gradiente estocástico é aplicado à função custo do filtro de Wiener multicanal com preservação da função de transferência interaural e implementado em um processador digital de sinais de uso geral. Comparações com simulações realizadas em Matlab, utilizando detectores de fala ideal e real, são apresentadas. Critérios objetivos de qualidade, conforto acústico e preservação do cenário acústico espacial, assim como experimentos psicoacústicos realizados com voluntários, indicam desempenho compatível com as requisições encontradas em aplicações de auxílio à audição.

Palavras-Chave— *Aparelhos auditivos, redução de ruído, MWF, ITF, pistas binaurais.*

Abstract—This work presents the real time implementation of a noise reduction adaptive algorithm for binaural hearing aids. The stochastic gradient method was applied to the multichannel Wiener filter with interaural transfer function and implemented in a general purpose digital signal processor. Comparisons with Matlab simulations, using a real and an ideal voice activity detector are presented. Quality, comfort and spatial preservation objective criteria, as well as psychoacoustic experiments with volunteers, indicate performance results compatible with those required by hearing aid applications.

Keywords— *Hearing aids, noise reduction, MWF, ITF, binaural cues.*

I. INTRODUÇÃO

O envelhecimento da população, a proliferação de fontes sonoras artificiais e a exposição prolongada a níveis inadequados de potência acústica têm resultado em um aumento crescente na demanda por equipamentos relacionados à tecnologia assistiva auditiva a cada ano. Aparelhos auditivos têm sido amplamente utilizados para a compensação de limitações na capacidade auditiva. Nos casos em que a perda é bilateral, muitas vezes são utilizados aparelhos binaurais, em virtude da possibilidade de compartilhamento de informações. Métodos de redução de ruído são parte importante desses aparelhos, aprimorando a inteligibilidade da fala e o conforto acústico do usuário [1].

Técnicas de redução de ruído binaural, além de proporcionar conforto acústico ao usuário, devem preservar as características espaciais das fontes sonoras presentes em uma cena acústica. Em cenários com fontes sonoras pontuais, a localização dessas fontes no plano horizontal é feita a partir de pistas binaurais. A diferença de tempo interaural (ITD –

Interaural Time Difference) é a pista binaural primária de localização sonora nos mamíferos, sendo dominante até 1500 Hz. A diferença de nível interaural (ILD – *Interaural Level Difference*) [2] é a pista binaural secundária de localização e é dominante nas frequências superiores a 1500 Hz [3] [4]. Essas pistas estão relacionadas à disposição espacial das fontes de interesse e de ruído, afetando diretamente a percepção da localização dessas fontes acústicas.

Dentre o conjunto de técnicas para redução de ruído binaural, o filtro de Wiener Multicanal (MWF – *Multichannel Wiener Filter*) é uma das mais utilizadas. Por ser uma técnica multicanal, esta técnica proporciona uma significativa redução de ruído ao mesmo tempo em que preserva as pistas acústicas da fonte de interesse (fala). No entanto, as informações binaurais das demais fontes acústicas (ruído) são alteradas. Em cenários com uma fonte sonora interferente pontual, a posição percebida é deslocada para a posição da fonte de fala (geralmente em frente ao usuário), o que pode acarretar desconforto, confusão ou mesmo situações de perigo.

Em [5], foi apresentada uma proposta de método de redução de ruído visando a preservação das pistas binaurais (ITD e ILD) da (única) fonte interferente de ruído 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 do ruído captado pelos microfones e a do ruído processado. Como grande desvantagem, essa técnica apresenta alto custo computacional em função de se apresentar como um problema de otimização não convexa.

A capacidade de processamento de aparelhos auditivos binaurais é crescente, porém ainda bastante limitada. Assim, é imprescindível que o método de redução de ruído empregado tenha baixo custo computacional e suficiente desempenho. Além disso, não pode haver atraso perceptível entre a captação dos sinais pelos microfones e a saída de áudio processado para o usuário.

Em [6], foi apresentado um algoritmo adaptativo para redução de ruído e preservação da função de transferência do sinal interferente. Resultados obtidos indicam a capacidade do algoritmo em, concomitantemente, reduzir ruído e preservar as pistas espaciais tanto da fala como do ruído.

Este trabalho trata da implementação em tempo real, em

Arthur Reys, Diego Carmo e Márcio Costa, Universidade Federal de Santa Catarina, Gabriel Souza, Softenge Automação Industrial. E-mails: arthur.reys@gmail.com, diego.carmoh@gmail.com, gabrisou28@gmail.com, costa@eel.ufsc.br. Este trabalho foi parcialmente financiado por CAPES e CNPq, projetos 304867/2015-2 e 315020/2018-0.

um processador digital de sinais, do método adaptativo de redução de ruído MWF-ITF_{SD} para aparelhos auditivos binaurais, apresentado originalmente em [6].

A divisão das seções foi feita da seguinte forma: na seção II é apresentada uma descrição do processamento de sinais em aparelhos auditivos binaurais. A seção III descreve o filtro de Wiener multicanal com preservação da ITF e, na seção IV, é descrito o método utilizado para a sua implementação em tempo real. Na seção V são apresentados os materiais e métodos utilizados para análise de desempenho e, na seção VI, os resultados obtidos e sua discussão. Por fim, a seção VII apresenta as conclusões desse trabalho.

Ao longo desse texto, letras maiúsculas e minúsculas em negrito representam, respectivamente, matrizes e vetores, enquanto que letras em itálico são escalares.

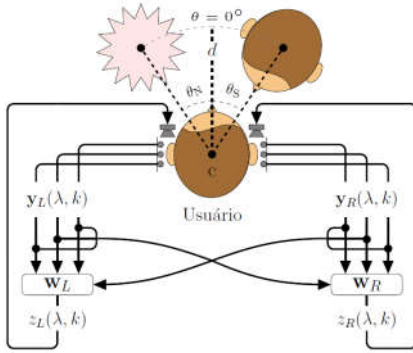


Fig. 1: Sistema de processamento binaural.

II. PROCESSAMENTO BINAURAL

O esquema de aquisição e processamento de sinais de um aparelho auditivo binaural é apresentado na Figura 1. O aparelho auditivo da esquerda (L) possui M_L microfones, e o da direita (R) M_R microfones. O conjunto total de microfones do sistema é dado por $M = M_L + M_R$. Os sinais captados pelo conjunto de microfones são armazenados em trechos (*frames*) para processamento. O vetor de amostras do sinal de entrada é transformado para domínio da frequência através da Transformada de Fourier de Tempo Curto (STFT – *Short Time Fourier Transform*). Para cada *frame* λ e frequência k , tem-se:

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

em que $y_{l,m}(\lambda, k)$ é o sinal captado pelo microfone $m = \{1, \dots, M_l\}$ do aparelho auditivo no lado $l = \{L, R\}$. O vetor $\mathbf{y}(\lambda, k)$ tem dimensão $M \times 1$ e é acessível por ambos os aparelhos através de um canal de comunicação sem fio, *full duplex*. Este vetor pode ser descrito por $\mathbf{y}(\lambda, k) = \mathbf{x}(\lambda, k) + \mathbf{v}(\lambda, k)$, em que $\mathbf{x}(\lambda, k)$ é o sinal recebido de fala e $\mathbf{v}(\lambda, k)$ é o ruído aditivo recebido.

Assumindo-se $\mathbf{x}(\lambda, k)$ e $\mathbf{v}(\lambda, k)$ como vetores aleatórios estacionários de média zero, definem-se as seguintes matrizes de coerência:

$$\begin{aligned} \Phi_x(k) &= E\{\mathbf{x}(\lambda, k)\mathbf{x}^H(\lambda, k)\} \\ \Phi_v(k) &= E\{\mathbf{v}(\lambda, k)\mathbf{v}^H(\lambda, k)\}, \\ \Phi_y(k) &= E\{\mathbf{y}(\lambda, k)\mathbf{y}^H(\lambda, k)\} \end{aligned} \quad (2)$$

em que $E\{\cdot\}$ é o valor esperado estatístico em função de λ , $(\cdot)^H$

é o transposto hermitiano, Φ_x é a matriz de coerência do sinal de fala, Φ_v é a matriz de coerência do ruído, Φ_y é a matriz de coerência do sinal de entrada. Assumindo-se independência entre os sinais de fala e ruído, tem-se que $\Phi_y = \Phi_x + \Phi_v$.

De forma geral, o microfone frontal de cada aparelho auditivo é designado como microfone de referência. O sinal recebido por esse microfone é dado por:

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

em que m_l é o índice correspondente ao microfone de referência e \mathbf{q}_l é um vetor com valor 1 na posição m_l e 0 nas demais posições.

O sinal processado no domínio da frequência é dado por:

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

em que $\mathbf{w}_l(k)$ representa um par de filtros de redução de ruído, $\mathbf{w}_L(k)$ e $\mathbf{w}_R(k)$, cada qual correspondente a um dos aparelhos auditivos. O sinal em (4) é transformado para o domínio do tempo utilizando-se a transformada inversa de Fourier e um método de reconstrução do tipo *overlap-and-add*, sendo, por fim, enviado aos alto-falantes do aparelho auditivo.

O conjunto dos coeficientes de ambos os aparelhos auditivos pode ser descrito por $\mathbf{w}(k) = [\mathbf{w}_L^T(k) \mathbf{w}_R^T(k)]^T$, com dimensão $2M \times 1$.

III. FILTRO DE WIENER MULTICANAL COM PRESERVAÇÃO DE PISTAS BINAURAIS

A redução de ruído binaural para aplicações em aparelhos auditivos apresenta duas funções importantes: reduzir ruído e preservar o cenário acústico na forma mais próxima possível do original.

Um conjunto bastante utilizado de técnicas de redução de ruído binaural é baseado no MWF, filtro que determina um estimador do sinal de fala recebido nos microfones de referência, i.e. x_{L,m_L} e x_{R,m_R} , através da minimização da seguinte função custo:

$$J_{MWF}(k) = E \left\{ \left\| \begin{bmatrix} x_{L,m_L}(\lambda, k) - \mathbf{w}_L^H(k) \mathbf{y}(\lambda, k) \\ x_{R,m_R}(\lambda, k) - \mathbf{w}_R^H(k) \mathbf{y}(\lambda, k) \end{bmatrix} \right\|^2 \right\}. \quad (5)$$

Os sinais resultantes da filtragem de Wiener multicanal apresentam as pistas acústicas do sinal de fala perfeitamente preservadas. Contudo, para o caso em que o cenário acústico é composto por uma fonte pontual de ruído, o MWF distorce as pistas acústicas dessa fonte fazendo com que sua posição percebida seja a mesma da fala.

Em [8], foi apresentada uma função custo auxiliar para regularização do MWF, baseada no erro quadrático médio entre a ITF do ruído captado pelos microfones (ITF_{in}) e a ITF do ruído processado (ITF_{ou}), dada por

$$\begin{aligned} J_{ITF}(k) &= E \left\{ \left\| ITF_{ou}(\lambda, k) - ITF_{in}(\lambda, k) \right\|^2 \right\} \\ &= E \left\{ \left\| \frac{\mathbf{w}_L^H(k) \mathbf{v}(\lambda, k)}{\mathbf{w}_R^H(k) \mathbf{v}(\lambda, k)} - \frac{\mathbf{q}_L^T \mathbf{v}(\lambda, k)}{\mathbf{q}_R^T \mathbf{v}(\lambda, k)} \right\|^2 \right\}. \end{aligned} \quad (6)$$

Para se obter de forma conjunta a redução de ruído e a preservação do cenário acústico, utiliza-se uma função custo

aumentada, dada por:

$$J_T(k) = J_{MWF}(k) + \alpha(k)J_{ITF}(k), \quad (7)$$

em que $\alpha(k)$ realiza a ponderação entre os termos de $J_T(k)$. Neste trabalho $\alpha(k) = \alpha$ é considerada constante para todos os *bins*. Omitindo-se (λ, k) e (k) , por questões de espaço, e assumindo-se descorrelação entre fala e ruído, a equação (5) resulta em:

$$J_{MWF} = P_x - \mathbf{w}^H \boldsymbol{\Phi}_x - \boldsymbol{\Phi}_x^H \mathbf{w} + \mathbf{w}^H \boldsymbol{\Phi}_{yy} \mathbf{w}, \quad (8)$$

em que

$$P_x = E\{|x_{L,m_L}|^2\} + E\{|x_{L,m_L}|^2\}, \quad (9)$$

$$\boldsymbol{\Phi}_x = \begin{bmatrix} \boldsymbol{\Phi}_x \mathbf{q}_L \\ \boldsymbol{\Phi}_x \mathbf{q}_R \end{bmatrix}, \quad \boldsymbol{\Phi}_{yy} = \begin{bmatrix} \boldsymbol{\Phi}_y & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_y \end{bmatrix},$$

sendo $\mathbf{0}$ uma matriz de zeros de dimensão $2M \times 2M$. Manipulando-se (6) e realizando-se algumas aproximações [8] chega-se em:

$$J_{ITF} = \frac{\mathbf{w}^H \boldsymbol{\Phi}_{vr} \mathbf{w}}{\mathbf{w}^H \boldsymbol{\Phi}_{vr} \mathbf{w}}, \quad (10)$$

em que

$$\boldsymbol{\Phi}_{vr} = \begin{bmatrix} \boldsymbol{\Phi}_v & -ITF_{in}^* \boldsymbol{\Phi}_v \\ -ITF_{in} \boldsymbol{\Phi}_v & |ITF_{in}|^2 \boldsymbol{\Phi}_v \end{bmatrix} \text{ e } \boldsymbol{\Phi}_{vr} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_v \end{bmatrix}. \quad (11)$$

em que $(\cdot)^*$ é a operação de conjugado complexo.

IV. IMPLEMENTAÇÃO RECURSIVA DO FILTRO MWF-ITF

Em [6] é apresentado um algoritmo recursivo com custo computacional controlado para otimização da função custo em (7). Como o objetivo primário de J_T é a redução de ruído, assume-se que a constante α deve possuir um valor relativamente pequeno, de modo que a convexidade de J_{MWF} é assumida como preponderante sobre J_{ITF} . Assim, é possível aplicar o método do gradiente descendente sobre a superfície de custo descrita em (7), de forma que [6]:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{\mu}{2} \nabla_{\mathbf{w}^*} J_T, \quad (12)$$

em que μ é uma constante de adaptação que controla a velocidade de convergência e a estabilidade do algoritmo. Substituindo-se (7) em (12), temos:

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \frac{\mu}{2} \nabla_{\mathbf{w}^*} J_{MWF} - \frac{\mu\alpha}{2} \nabla_{\mathbf{w}^*} J_{ITF}. \quad (13)$$

Os gradientes de J_{MWF} e J_{ITF} com relação a \mathbf{w}^* são dados, respectivamente, por [6]:

$$\nabla_{\mathbf{w}^*} J_{MWF} = 2(\boldsymbol{\Phi}_{yy} \mathbf{w} - \boldsymbol{\Phi}_x) \quad (14)$$

e

$$\nabla_{\mathbf{w}^*} J_{ITF} = 2 \frac{(\mathbf{w}^H \boldsymbol{\Phi}_{vr} \mathbf{w}) \boldsymbol{\Phi}_{vr} \mathbf{w} - (\mathbf{w}^H \boldsymbol{\Phi}_{vr} \mathbf{w}) \boldsymbol{\Phi}_{vr} \mathbf{w}}{(\mathbf{w}^H \boldsymbol{\Phi}_{vr} \mathbf{w})^2}. \quad (15)$$

Substituindo-se (14) e (15) em (13), obtém-se a seguinte equação de atualização dos coeficientes [6]:

$$\mathbf{w}_{n+1} = (\mathbf{I} - \beta \boldsymbol{\Phi}_{yy}) \mathbf{w}_n + \beta \boldsymbol{\Phi}_x + \gamma \left(\frac{(\mathbf{w}^H \boldsymbol{\Phi}_{vr} \mathbf{w}) \boldsymbol{\Phi}_{vr} \mathbf{w} - (\mathbf{w}^H \boldsymbol{\Phi}_{vr} \mathbf{w}) \boldsymbol{\Phi}_{vr} \mathbf{w}}{(\mathbf{w}^H \boldsymbol{\Phi}_{vr} \mathbf{w})^2} \right), \quad (16)$$

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

V. MATERIAIS E MÉTODOS

Essa seção apresenta os materiais e métodos utilizados para a implementação em tempo real do sistema de processamento binaural.

A. Hardware de Implementação

A implementação do algoritmo proposto foi realizada no processador digital de sinais (DSP – *Digital Signal Processing*) ADSP-21489, fabricado pela *Analog Devices* [10]. Para tanto, foi utilizada a placa de desenvolvimento ADSP-21489 EZ-KIT Lite [11], que realiza operações em ponto flutuante no formato 32/40 bits. O EZ-KIT é composto por dois processadores de até 450 MHz que operam em formato de instrução única. Possui 5 Mbits de memória RAM interna, múltiplos barramentos de dados e aceleradores independentes para realização da transformada rápida de Fourier e implementação de filtros digitais.

B. Sistema Desenvolvido

A Figura 2 apresenta o diagrama em blocos do sistema desenvolvido para operação em tempo real. Primeiramente, ocorre a aquisição das amostras do sinal de entrada, que compõem um *frame* de áudio, obtido a cada 4 ms. Em sequência, o sinal de entrada é transformado para o domínio da frequência através de uma STFT de 256 pontos. Utilizando informação fornecida por um detector de fala (VAD – *Voice Active Detector*), as matrizes de coerência são estimadas utilizando (17), (18) e (19). Estimadas essas matrizes, determinam-se os coeficientes do filtro MWF-ITF por meio de (16), enquanto que o processo de filtragem do sinal de entrada é realizado por meio de (4). Esses procedimentos geram uma estimativa do sinal de interesse no domínio da frequência. Essa estimativa é transformada para o domínio do tempo, constituindo o *frame* de sinal a ser enviado aos alto-falantes de saída.

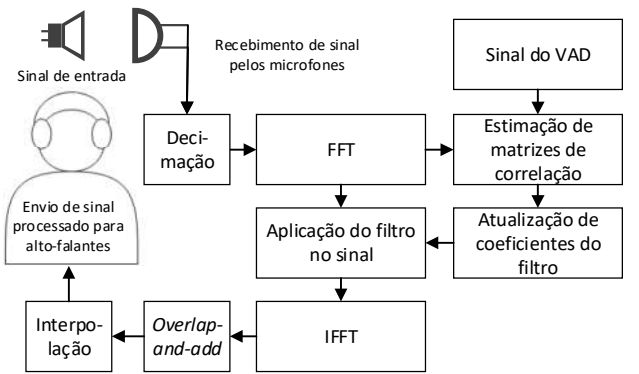


Fig. 2: Diagrama em blocos do sistema.

C. Detector de Fala

Como descrito na Seção IV-B, para a estimação das matrizes de coerência é necessário o uso de um VAD. Esse componente do sistema binaural discrimina os trechos contendo fala contaminada dos trechos que contêm apenas ruído. Dois tipos de VAD foram utilizados no sistema: um *VAD real*, representado pelo método em [19]; e um *VAD ideal*, obtido a partir da anotação manual do início e fim de cada trecho de fala, assumindo-se $\text{SNR} \rightarrow \infty$. A análise comparativa de ambas as situações permite inferir sobre como erros no VAD impactam no desempenho do sistema. Nos

experimentos apresentados nesse trabalho, ambos os VADs foram implementados *off-line* – o sinal do VAD foi calculado externamente ao DSP e sua informação foi carregada em memória previamente à execução do processamento. Contudo, em implementação posterior, o VAD descrito em [20] foi incorporado de forma *online* no sistema.

D. Parâmetros de projeto

As matrizes de coerência, necessárias para o cálculo da equação (16), são obtidas da seguinte forma:

$$\hat{\Phi}_y(\lambda, k) = \begin{cases} \eta_y \hat{\Phi}_y(\lambda-1, k) + (1-\eta_y) \mathbf{y}(\lambda, k) \mathbf{y}^H(\lambda, k), & VAD = 1 \\ \hat{\Phi}_y(\lambda-1, k), & VAD = 0 \end{cases} \quad (17)$$

$$\hat{\Phi}_v(\lambda, k) = \begin{cases} \hat{\Phi}_v(\lambda-1, k), & VAD = 1 \\ \eta_v \hat{\Phi}_v(\lambda-1, k) + (1-\eta_v) \mathbf{y}(\lambda, k) \mathbf{y}^H(\lambda, k), & VAD = 0 \end{cases} \quad (18)$$

$$\hat{\Phi}_x(\lambda, k) = \eta_x \hat{\Phi}_x(\lambda-1, k) + (1-\eta_x) (\hat{\Phi}_y(\lambda, k) - \hat{\Phi}_v(\lambda, k)) \quad (19)$$

Os fatores de esquecimento para a estimação de (17), (18) e (19) foram definidos como $\eta_y = \eta_v = \eta_x = 0,999$ [12], equivalentes a uma constante de tempo de 4 s.

O passo de adaptação foi definido como $\mu = 0,04$, estabelecendo um compromisso entre capacidade de rastreamento e desempenho em regime permanente. O valor de α foi definido arbitrariamente a fim de estabelecer uma redução máxima de 15% de SNR média (em ambas as orelhas, esquerda e direita) quando comparada à solução determinada pelo MWF. Assim, definiu-se $\alpha = 2,5 \times 10^{-3}$ e, por conseguinte, $\gamma = 10^{-4}$. O número de iterações da equação (16), por bin, foi definido arbitrariamente como $P = 20$.

E. Cenário acústico simulado

O cenário acústico simulado é similar ao apresentado na Figura 1 e consiste em uma fonte de fala localizada em $\theta_S = 0^\circ$ e uma fonte de ruído localizada em $\theta_N = -60^\circ$. Ambas as fontes apresentam elevação de 0° em relação ao eixo transversal e estão a 3 metros do ponto central (C), na região de campo distante do usuário de aparelhos auditivos.

O sinal analisado consiste em um período inicial de 3 s contendo apenas ruído, seguido por 14,7 s de fala contaminada. O ruído é do tipo ICRA (*International Collegium of Rehabilitative Audiology*) [13], e apresenta características espectro-temporais semelhantes às da fala humana. A frequência de amostragem dos sinais é de 16 kHz.

O cenário acústico descrito foi gerado convolvendo-se os sinais de fala e ruído com as funções de transferência (HRIR – *Head Related Transfer Functions*) que determinam o caminho acústico entre cada fonte e os microfones dos aparelhos auditivos. Foi utilizado um banco de HRIRs obtidas através de experimentos reais sob as seguintes condições [14]: câmara anecoica, manequim Bruel & Kjaer tipo 4128-C e um par de aparelhos auditivos retro-auriculares.

F. Métricas Objetivas

Para determinar a eficácia do processo de redução de ruído e da preservação perceptual da localização da fonte de ruído, foram utilizadas quatro métricas de avaliação: a variação da

avaliação perceptual da qualidade da fala ($\Delta WPESQ$) [15], que mede o incremento na qualidade da fala processada; a variação da relação sinal ruído (ΔSNR), que mede a redução global de ruído e determina o conforto acústico; a variação da diferença de tempo interaural (ΔITD) [16], calculada até 1,5 kHz, que mede a diferença entre as ITDs do sinal recebido (ITD_{in}) e do sinal processado (ITD_{ou}); e a diferença de nível interaural (ΔILD) [17], calculada para frequências acima de 1,5 kHz, que mede a diferença entre as ILDs do sinal antes (ILD_{in}) e após (ILD_{ou}) processamento.

O sinal contaminado recebido apresenta as seguintes métricas de referência: $PESQ_L = 1,1$, $PESQ_R = 1,3$, $SNR_L = -0,5$ dB e $SNR_R = 3,2$ dB.

VI. RESULTADOS E DISCUSSÃO

Essa seção apresenta os resultados da implementação do sistema de redução de ruído com preservação da localização espacial de uma fontes interferente.

A. Avaliação Objetiva

A Tabela I apresenta os resultados das métricas objetivas de avaliação de desempenho para o sinal processado, utilizando o *VAD ideal* e o *real*. A implementação do sistema foi realizada de duas formas: *off-line* no programa Matlab e em tempo real na placa de desenvolvimento, obtendo-se resultados muito próximos. Dessa forma, por questões de espaço, não são apresentados os resultados obtidos no Matlab. Os métodos comparados foram: o MWF adaptativo (MWF_{SD}), obtido com $\gamma = 0$ em (16), e o MWF-ITF adaptativo ($MWF-ITF_{SD}$) para $\gamma = 10^{-4}$.

As soluções obtidas pelo MWF_{SD} resultam em um aumento médio de SNR de 15,4 dB em ambas as orelhas. Entretanto, as métricas de localização do ruído apresentam valores elevados ($\Delta ILD_N = 7,3$ dB e $\Delta ITD_N = 0,5$), o que indica a alteração das pistas de localização acústica do ruído processado. Há pequena variação na ITD da fala para todas as técnicas analisadas. As soluções obtidas pelo $MWF-ITF_{SD}$ apresentam uma variação média da ILD da fala de $\Delta ILD_S = 4,59$. Apesar disso, experimentos psicoacústicos preliminares realizados pelos autores não apresentaram alterações na localização percebida da fala. O método $MWF-ITF_{SD}$ apresenta ganho médio de SNR de 13,8 dB, com variação média 4,25 dB na ΔILD_N . Já a ΔITD_N permanece inalterada, em todos os casos analisados.

TABELA I – MÉTRICAS OBJETIVAS PARA O SINAL PROCESSADO.

	VAD IDEAL		VAD REAL	
	MWF_{SD}	$MWF-ITF_{SD}$	MWF_{SD}	$MWF-ITF_{SD}$
$\Delta PESQ_L$	1,33	0,70	1,31	0,79
$\Delta PESQ_R$	1,31	1,11	1,30	1,13
ΔSNR_L	17,20	14,51	17,60	15,09
ΔSNR_R	13,62	13,03	14,03	13,34
ΔITD_S	0,01	0,01	0,00	0,01
ΔILD_S	0,58	5,66	0,40	3,52
ΔITD_N	0,54	0,50	0,53	0,51
ΔILD_N	7,26	3,12	7,88	5,44

B. Avaliação Subjetiva

De modo a avaliar os resultados objetivos na Seção VI-A,

foi realizado um experimento psicoacústico com 10 voluntários, conforme os procedimentos aprovados pelo Comitê de Ética em Pesquisa com Seres Humanos mediante certificado CEP/SH-UFSC número 49741615.2.0000.0121.

Cada participante ouviu aleatoriamente trechos de ruído processado e não processado, devendo indicar seu azimute de origem. As potências dos sinais foram normalizadas para padronização de volume. Os resultados são apresentados na Figura 3, sendo: (a) para *VAD ideal*; e (b) para *VAD real*. O ruído não processado é indicado por NP.

Para a situação em que é utilizado um *VAD ideal* (Fig. 3a), o ruído processado pelo MWF_{SD} é deslocado para a posição da fala, resultando em uma mediana próxima ao azimute de 0° e, portanto, em um erro médio de 60° . A diferença entre a localização percebida com o ruído processado pelo método $MWF-ITF_{SD}$ e com o ruído não processado é de $22,5^\circ$, apresentando pequeno espalhamento das amostras, com 9 das 10 amostras contidas em um intervalo de 30° .

O uso do VAD real reduz o desempenho da técnica proposta em relação ao VAD ideal, prejudicando a preservação espacial do ruído. Nesse caso, o erro entre o azimute percebido por meio do método proposto e o azimute percebido ouvindo-se o ruído não processado aumenta para 45° . O espalhamento é um pouco maior que para o caso do *VAD ideal*, com todas as amostras contidas em um intervalo de 60° .

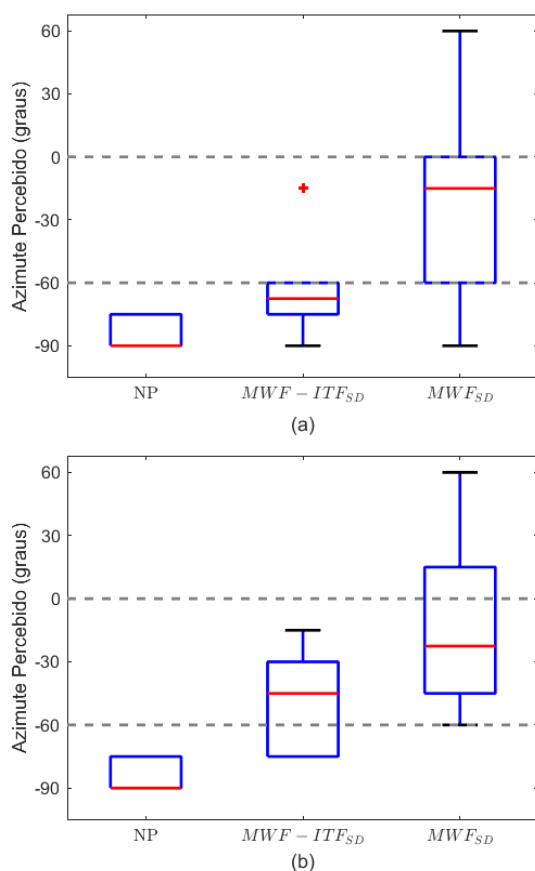


Fig. 3: Resultados dos experimentos psicoacústicos com voluntários: (a) VAD ideal; e (b) VAD real.

VII. CONCLUSÃO

Esse trabalho apresentou a implementação em tempo real,

em um DSP, de um método adaptativo de redução de ruído para aparelhos auditivos binaurais. Critérios objetivos de qualidade, conforto acústico e preservação do cenário acústico espacial, assim como experimentos psicoacústicos realizados com voluntários, indicam a viabilidade para redução de ruído e preservação espacial de fontes sonoras, apresentando desempenho compatível com as requisições encontradas em aplicações de auxílio à audição. Entretanto, o desempenho da técnica proposta apresenta sensibilidade ao tipo de detector de fala utilizado.

REFERÊNCIAS

- [1] M. H. Costa, "A complementary low-cost method for broadband noise reduction in hearing aids for medium to high SNR levels", *Comput. Biol. Med.*, vol. 46, pp. 29-41, 2014.
- [2] J. Middlebrooks, D. Green, "Sound localization by human listeners", *An. Rev. Psy.*, vol. 42, no. 1, pp. 135-159, 1991.
- [3] T. T. Sandel, D. C. Feddersen, L. A. Jeffress, "Localization of sound from single and paired sources", *J. Acoust. Soc. Am.*, vol. 27, n. 5, pp. 842-852, 1955.
- [4] A. W. Mills, "Lateralization of high frequency tones", *J. Acoust. Soc. Am.*, vol. 32, n. 1, pp. 132-134, 1960.
- [5] T. J. Klasen et al. "Binaural multichannel Wiener filtering for hearing-aids: preserving interaural time and level-differences", *Proc. IEEE Int. Conf. Acoustics, Speech, Lang. Process. (ICASSP)*, vol. 5, pp. 145-148, 2006.
- [6] D. M. do Carmo, F. P. Iturriet, M. H. Costa, "Redução de ruído binaural em tempo real com preservação da função de transferência interaural", *Anais do Simp. Bras. Telecom. (SBRT)*, pp. 438-442, 2018.
- [7] R. E. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis", *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-28, n. 1, pp. 99-102, 1980.
- [8] S. Doclo et al. "Speech distortion weighted multichannel Wiener filtering techniques for noise reduction", *Speech Enhancement*, Springer-Verlag, pp. 199-228, 2005.
- [9] T. J. Klasen et al. "Preservation of interaural time delay for binaural hearing aids through multi-channel Wiener filtering based noise reduction", *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 2932, 2005.
- [10] ADSP 21489, Analog Devices Datasheet.
- [11] ADSP 21489 EZ-Board Evaluation System Manual, Analog Devices.
- [12] J.I. Marin-Hurtado, D.V. Anderson, "Practical MWF-based noise reduction methods for binaural hearing aids", *Proc. STSIVA*, pp. 1-8, 2012.
- [13] 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", *Audiol.*, vol. 40, n. 3, pp. 148-157, 2001.
- [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] 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.
- [16] T. Bogaert et al. "Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter", *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 565-568, 2007.
- [17] Costa, M. H., Naylor, P. A. "ILD preservation in the multichannel wiener filter for binaural hearing aid applications", *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, pp. 636-640, 2014.
- [18] J.C. Servetti, Martin, "Error tolerant MAC extension for speech communications over 802.11 WLANs", *Proc. VTC*, vol. 4, pp. 2330-2334, 2005.
- [19] J. Sohn, N. S. Kim, W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Processing Letters*, vol 16, n. 1, pp. 1-3, 1999.
- [20] Marzinik, M., & Kollmeier, B. (2002). Speech pause detection for noise spectrum estimation by tracking power envelope dynamics. *IEEE Transactions on Speech and Audio Processing*, 10(2), 109-118.