

Conformador de feixe robusto MVDR baseado na otimização de desempenho do pior caso para aparelhos auditivos biauriculares

Wilmer Lobato e Márcio Costa

Resumo— Este trabalho apresenta um conformador de feixe robusto para aplicações em aparelhos auditivos biauriculares. É utilizado o método de otimização de desempenho do pior caso para obter aumento na robustez do conformador de mínima variância com resposta sem distorção em relação a erros em seus parâmetros. Simulações computacionais indicam um aumento na qualidade da fala em até 1,1 MOS-WPESQ e no conforto acústico em até 6,2 dB em termos de razão sinal-ruído (SNR). O método proposto é especialmente efetivo para SNRs de entrada entre 5 dB e 15 dB.

Palavras-Chave— *Aparelho auditivo, conformação de feixe, otimização de desempenho do pior caso, diferença de nível interaural.*

Abstract— This work presents a robust beamformer for binaural hearing aid applications. The worst-case performance optimization method was applied to the minimum variance distortionless response beamformer for increasing its robustness against parameter errors. Computational simulations indicate significant improvements of up to 1.1 MOS-WPESQ in speech quality and up to 6.2 dB in acoustic comfort, in terms of signal to noise ratio (SNR). The proposed method is especially effective for input SNRs from 5 dB to 15 dB.

Keywords— *Hearing aids, beamforming, worst-case performance optimization, interaural level difference.*

I. INTRODUÇÃO

O conformador de feixe de mínima variância com resposta sem distorção (MVDR, *minimum variance distortionless response*) é uma técnica de redução de ruído que visa a minimizar a potência total de saída, preservando a informação relativa à fonte de interesse [1]. Formulações do conformador MVDR no domínio da frequência para redução de ruído em aparelhos auditivos foram apresentadas em [2].

O projeto do conformador MVDR requer informação *a priori* sobre o cenário acústico como, por exemplo, a matriz de coerência do ruído e o vetor de direção (SV, *steering vector*) desejado [3]. Os erros de estimação desses parâmetros podem resultar em uma degradação significativa no desempenho do conformador. Algumas fontes de erro são: desajuste de ganho e fase dos microfones; calibração imperfeita do arranjo; e acoplamento de microfones [4]. Além disso, os erros no vetor de direção podem gerar o fenômeno de cancelamento de sinal, o qual degrada de forma significativa o desempenho do conformador, reduzindo a qualidade e a inteligibilidade da fala desejada [5].

Wilmer Lobato e Márcio Costa, Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, Florianópolis-SC, E-mails: wilmer.lobato@gmail.com; costa@eel.ufsc.br. Este trabalho foi financiado pelo CNPq, projetos 162490/2015-1 e 315020/2018-0.

Uma forma de lidar com a sensibilidade do conformador ao erro de estimação é projetar conformadores robustos. Nesse sentido, a otimização de desempenho do pior caso (WCO, *worst case optimization*) é um método robusto que minimiza a variância de saída impondo uma magnitude de resposta maior ou igual à unidade, dentro de uma região ao redor da localização esperada da fonte sonora desejada. Assumindo uma determinada forma geométrica para o erro no SV, a formulação teórica resultante pode ser transformada em um problema de minimização convexo. Em [6], foi usada a máxima norma quadrática do erro no SV desejado, modelando um volume hiperesférico. Esse enfoque permitiu obter a solução através de um problema de tipo cone de segunda ordem (SOCP, *second-order cone programming*), resolvido através de métodos de pontos interiores, com alta eficiência.

Este trabalho propõe a adaptação do método de otimização de desempenho do pior caso para um conformador de feixe biauricular, utilizando-se o método de mínima variância com resposta sem distorção. Esse conformador, nomeado de WCO-BMVDR apresenta comportamento robusto a erros de estimação no vetor de direção desejado. As principais contribuições deste trabalho são: (a) adequação do método proposto em [6] para aplicação em aparelhos auditivos biauriculares; (b) demonstração de que os parâmetros de controle do conformador proposto podem ser projetados através de estimativas da diferença de nível interaural; e (c) realização de simulações para demonstrar o aumento de desempenho do conformador proposto em termos de qualidade da fala e conforto acústico.

Ao longo do texto, letras minúsculas em itálico representam escalares. Letras minúsculas e maiúsculas em negrito denotam, respectivamente, vetores e matrizes. Os símbolos $\{\cdot\}^T$ e $\{\cdot\}^H$ são, respectivamente, o operador transposto e Hermitiano.

II. FORMULAÇÃO DO PROBLEMA

O contexto da aplicação considera um par de aparelhos auditivos biauriculares, compostos por dois arranjos de microfones lineares, um em cada ouvido. Os sinais recebidos pelos microfones são transformados no domínio da frequência através da transformada de Fourier de tempo curto (STFT, *short-time Fourier transform*) de K bins, resultando em:

$$y_{L,m}(\lambda, k) = s(\lambda, k)a_{L,m}(\lambda, k) + v(\lambda, k)b_{L,m}(\lambda, k), \quad (1)$$

em que $y_{L,m}(\lambda, k)$ é a STFT do sinal de fala contaminado para $m = 0, 1, \dots, M-1$ microfones no ouvido esquerdo (L , *left*), para cada índice de quadro de tempo λ e *bin* de frequência k ; $s(\lambda, k)$ e $v(\lambda, k)$ são, respectivamente, a STFT do sinal de fala desejada e do ruído; $a_{L,m}(\lambda, k)$ e $b_{L,m}(\lambda, k)$ são as funções de transferência acústicas (ATF, *acoustic transfer function*), respectivamente,

desde as fontes de fala e de ruído até o m -ésimo microfone do ouvido esquerdo. Os sinais recebidos podem ser expressos na forma vetorial como:

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

em que $\mathbf{y}_L(\lambda, k) = \mathbf{s}_L(\lambda, k) + \mathbf{v}_L(\lambda, k) = [y_{L,0}(\lambda, k) \dots y_{L,M-1}(\lambda, k)]^T$ é o vetor que contém a fala contaminada; e $\mathbf{a}_L(\lambda, k) = [a_{L,0}(\lambda, k) \dots a_{L,M-1}(\lambda, k)]^T$ e $\mathbf{b}_L(\lambda, k) = [b_{L,0}(\lambda, k) \dots b_{L,M-1}(\lambda, k)]^T$ são, respectivamente, os SVs da fala e ruído em relação ao ouvido esquerdo. Uma notação similar é aplicada para o ouvido direito (R, *right*). Define-se o vetor empilhado para a configuração biauricular como $\mathbf{y}(\lambda, k) = \mathbf{s}(\lambda, k) + \mathbf{v}(\lambda, k) = [\mathbf{y}_L^T(\lambda, k) \ \mathbf{y}_R^T(\lambda, k)]^T$; $\mathbf{a}(\lambda, k) = [\mathbf{a}_L^T(\lambda, k) \ \mathbf{a}_R^T(\lambda, k)]^T$, e $\mathbf{b}(\lambda, k) = [\mathbf{b}_L^T(\lambda, k) \ \mathbf{b}_R^T(\lambda, k)]^T$, em que $\mathbf{a}(\lambda, k)$ e $\mathbf{b}(\lambda, k)$ são os SVs empilhados da fala e do ruído, com dimensão $2M \times 1$.

Os sinais de saída (esquerdo e direito) do conformador biauricular são obtidos pela filtragem linear dos sinais de entrada:

$$\begin{aligned} \mathbf{z}_L(\lambda, k) &= \mathbf{w}_L^H(\lambda, k)\mathbf{y}(\lambda, k) \\ \mathbf{z}_R(\lambda, k) &= \mathbf{w}_R^H(\lambda, k)\mathbf{y}(\lambda, k) \end{aligned}, \quad (3)$$

em que $\mathbf{w}_L(\lambda, k)$ e $\mathbf{w}_R(\lambda, k)$ são os vetores de coeficientes do lado esquerdo e direito, projetados para reduzir o ruído e realçar a informação desejada [7].

A. Conformador BMVDR ideal

Adaptado para aparelhos auditivos biauriculares em [8], o conformador BMVDR minimiza a potência total do ruído, preservando os componentes de fala através de uma restrição linear. A sua formulação no domínio da frequência é definida pelo seguinte problema de otimização:

$$\begin{aligned} \min_{\mathbf{w}(\lambda, k)} \quad & \mathbf{w}^H(\lambda, k)\mathbf{\Phi}(\lambda, k)\mathbf{w}(\lambda, k) \\ \text{sujeito a} \quad & \mathbf{A}^H(\lambda, k)\mathbf{w}(\lambda, k) = \mathbf{1}_{2 \times 1} \end{aligned}, \quad (4)$$

$$\text{em que: } \mathbf{A}(\lambda, k) = \begin{bmatrix} \bar{\mathbf{a}}_L(\lambda, k) & \mathbf{0}_{2M \times 1} \\ \mathbf{0}_{2M \times 1} & \bar{\mathbf{a}}_R(\lambda, k) \end{bmatrix}; \quad (5)$$

$$\mathbf{\Phi}(\lambda, k) = \begin{bmatrix} \mathbf{\Phi}_{vv}(\lambda, k) & \mathbf{0}_{2M \times 2M} \\ \mathbf{0}_{2M \times 2M} & \mathbf{\Phi}_{vv}(\lambda, k) \end{bmatrix}; \quad (6)$$

onde $\mathbf{w}(\lambda, k) = [\mathbf{w}_L^T(\lambda, k) \ \mathbf{w}_R^T(\lambda, k)]^T$ é o vetor de coeficientes biauriculares; $\mathbf{\Phi}_{vv}(\lambda, k) = E\{\mathbf{v}(\lambda, k)\mathbf{v}^H(\lambda, k)\} = \varphi_{vv}(\lambda, k)\mathbf{b}(\lambda, k)\mathbf{b}^H(\lambda, k)$ é a matriz de coerência do ruído; $E\{\cdot\}$ é o operador de valor esperado em relação a λ ; $\varphi_{vv}(\lambda, k)$ é a densidade espectral de potência (PSD) do ruído [8]; $\mathbf{1}_{2 \times 1}$ é um vetor coluna unitário (preenchido com 1) com dimensão 2×1 ; $\mathbf{0}_{i \times j}$ é uma matriz de zeros com dimensão $i \times j$; $\bar{\mathbf{a}}_L(k)$ e $\bar{\mathbf{a}}_R(k)$ são, respectivamente, os SVs esquerdo e direito, construídos a partir das funções de transferência relativas (RTF, *relative transfer function*) em relação aos microfones de referência (frontais) $a_{L,0}(k)$ e $a_{R,0}(k)$, associados aos ouvidos esquerdo e direito. Em aplicações práticas, as RTF-SVs são mais fáceis de se estimar que as ATF-SVs [3].

A solução fechada do conformador BMVDR definido em (4) é obtida através da solução de [8] como:

$$\mathbf{w}_{otimo}(\lambda, k) = \frac{\mathbf{\Phi}^{-1}(\lambda, k)\mathbf{A}(\lambda, k)\mathbf{1}_{2 \times 1}}{\mathbf{A}^H(\lambda, k)\mathbf{\Phi}^{-1}(\lambda, k)\mathbf{A}(\lambda, k)}, \quad (7)$$

em que $[\mathbf{A}^H(\lambda, k)\mathbf{\Phi}^{-1}(\lambda, k)\mathbf{A}(\lambda, k)]^{-1}\mathbf{1}_{2 \times 1}$ fornece os fatores de escala dos filtros ótimos $\mathbf{w}_L(\lambda, k)$ e $\mathbf{w}_R(\lambda, k)$, os quais não afetam a razão sinal-ruído (SNR) de saída [9]. Essa técnica é denominada conformador BMVDR ideal (I-BMVDR), devido ao fato de considerar que $\mathbf{A}(\lambda, k)$ e $\mathbf{\Phi}(\lambda, k)$ são perfeitamente conhecidos.

B. Implementação prática do conformador BMVDR

Na prática, $\mathbf{\Phi}(\lambda, k)$ e $\mathbf{A}(\lambda, k)$ são desconhecidos e devem ser estimados. Em aplicações de aparelhos auditivos, as matrizes de coerência podem ser estimadas utilizando-se um detector de fala (VAD, *voice activity detector*). De acordo com [10], a matriz de coerência da fala $\mathbf{\Phi}_{ss}(\lambda, k) = \varphi_{ss}(\lambda, k)\mathbf{a}(\lambda, k)\mathbf{a}^H(\lambda, k)$, em que $\varphi_{ss}(\lambda, k)$ é a densidade espectral de potência da fala, pode ser estimada de forma recursiva através de $\hat{\mathbf{\Phi}}_{ss}(\lambda, k) = \eta_s \hat{\mathbf{\Phi}}_{ss}(\lambda-1, k) + (1-\eta_s) \times (\hat{\mathbf{\Phi}}_{yy}(\lambda, k) - \hat{\mathbf{\Phi}}_{vv}(\lambda, k))$, e $\hat{\mathbf{\Phi}}_{yy}(\lambda, k) = \eta_y \hat{\mathbf{\Phi}}_{yy}(\lambda-1, k) + (1-\eta_y) \mathbf{y}(\lambda, k) \mathbf{y}^H(\lambda, k)$. Durante os períodos não vozeados (VAD inativo) tem-se que $\hat{\mathbf{\Phi}}_{vv}(\lambda, k) = \eta_v \hat{\mathbf{\Phi}}_{vv}(\lambda-1, k) + (1-\eta_v) \mathbf{y}(\lambda, k) \mathbf{y}^H(\lambda, k)$, enquanto que durante os períodos vozeados (VAD ativo) utiliza-se $\hat{\mathbf{\Phi}}_{vv}(\lambda, k) = \hat{\mathbf{\Phi}}_{vv}(\lambda-1, k)$. Essa abordagem considera que existe independência estatística entre as fontes de fala e ruído, de forma que $\hat{\mathbf{\Phi}}_{yy}(\lambda, k) = \hat{\mathbf{\Phi}}_{ss}(\lambda, k) + \hat{\mathbf{\Phi}}_{vv}(\lambda, k)$ [8].

As RTF-SVs $\bar{\mathbf{a}}_L(\lambda, k)$ e $\bar{\mathbf{a}}_R(\lambda, k)$ podem ser estimadas através do método de mínima distorção, proposto em [11], o qual somente requer o conhecimento da matriz de coerência estimada da fala $\hat{\mathbf{\Phi}}_{ss}(\lambda, k)$. Nesse sentido, a implementação prática do BMVDR, chamada de conformador E-BMVDR (BMVDR estimado), utiliza estimativas $\hat{\mathbf{\Phi}}(\lambda, k)$ e $\hat{\mathbf{A}}(\lambda, k)$ ao invés de seus valores ideais $\mathbf{\Phi}(\lambda, k)$ e $\mathbf{A}(\lambda, k)$.

Segundo diversos estudos [6], o conformador BMVDR é altamente sensível a incertezas na matriz de coerência e no vetor de direção do sinal desejado. A acurácia na estimação de $\mathbf{\Phi}(\lambda, k)$ e $\mathbf{A}(\lambda, k)$ é principalmente determinada pelo desempenho do VAD e da SNR de entrada [9]. Nesse sentido, as perdas de desempenho devido aos erros de estimação são intrínsecas à implementação prática.

C. Método de otimização de desempenho do pior caso

O método de otimização de desempenho do pior caso é uma abordagem robusta baseada na minimização da variância de saída do conformador, a qual é restrita pela imposição de uma resposta de magnitude do arranjo maior que a unidade, dentro de uma região de incerteza desejada [6]. Sob essa abordagem, a região de incerteza é modelada geometricamente, e a solução é obtida através de um problema de otimização convexo [9]. O presente trabalho utiliza a formulação hipersférica apresentada em [6].

Dentro do contexto dos aparelhos auditivos biauriculares, o problema pode ser formulado como segue: Considere um ponto no espaço definido pelas RTFs: $\hat{\mathbf{c}}_L(\lambda, k) = \hat{\mathbf{a}}_L(\lambda, k) + \mathbf{e}_L(\lambda, k)$ e $\hat{\mathbf{c}}_R(\lambda, k) = \hat{\mathbf{a}}_R(\lambda, k) + \mathbf{e}_R(\lambda, k)$, $\hat{\mathbf{c}}(\lambda, k) = [\hat{\mathbf{c}}_L^T(\lambda, k) \ \hat{\mathbf{c}}_R^T(\lambda, k)]^T$, em que $\mathbf{e}_L(\lambda, k)$ e $\mathbf{e}_R(\lambda, k)$ são os vetores de erro. Considera-se que o vetor de direção desejado está alocado dentro de uma região de incerteza ao redor do vetor de direção estimado, de tal forma que $\|\mathbf{e}_L(\lambda, k)\| \leq \zeta_L(\lambda, k)$ e $\|\mathbf{e}_R(\lambda, k)\| \leq \zeta_R(\lambda, k)$, em que $\zeta_L(\lambda, k)$ e $\zeta_R(\lambda, k)$ são os valores limites das normas quadráticas [6]. Dessa forma, o conformador de feixe biauricular com otimização de

desempenho do pior caso é definido pelo seguinte problema de minimização (λ foi omitido por simplicidade):

$$\begin{aligned} \min_{\mathbf{w}_L(k), \mathbf{w}_R(k)} \quad & \mathbf{w}_L^H(k) \hat{\Phi}_{vw}(k) \mathbf{w}_L(k) + \mathbf{w}_R^H(k) \hat{\Phi}_{vw}(k) \mathbf{w}_R(k) \\ \text{sujeito a} \quad & |\mathbf{w}_L^H(k) \mathbf{c}_L(k)| \geq 1, |\mathbf{w}_R^H(k) \mathbf{c}_R(k)| \geq 1 \\ & \forall \mathbf{c}_L(k), \mathbf{c}_R(k) \in \Xi(\zeta_L(k), \zeta_R(k)) \end{aligned} \quad (8)$$

em que $\Xi(\zeta_L(k), \zeta_R(k)) = \{ \bar{\mathbf{c}}_L(k), \bar{\mathbf{c}}_R(k) \mid \bar{\mathbf{c}}_L(k) = \hat{\mathbf{a}}_L(k) + \mathbf{e}_L(k), \bar{\mathbf{c}}_R(k) = \hat{\mathbf{a}}_R(k) + \mathbf{e}_R(k), \|\mathbf{e}_L(k)\| \leq \zeta_L(k), \|\mathbf{e}_R(k)\| \leq \zeta_R(k) \}$. Transformando-se (8) em um programa de cone de segunda ordem (SOCP), obtém-se [6]:

$$\begin{aligned} \min_{\mathbf{w}_L(k), \mathbf{w}_R(k)} \quad & \mathbf{w}_L^H(k) \hat{\Phi}_{vw}(k) \mathbf{w}_L(k) + \mathbf{w}_R^H(k) \hat{\Phi}_{vw}(k) \mathbf{w}_R(k) \\ \text{sujeito a} \quad & \zeta_L(k) \|\mathbf{w}_L(k)\|_2 \leq \mathbf{w}_L^H(k) \hat{\mathbf{a}}_L(k) - 1 \\ & \zeta_R(k) \|\mathbf{w}_R(k)\|_2 \leq \mathbf{w}_R^H(k) \hat{\mathbf{a}}_R(k) - 1 \\ & \text{Im}\{\mathbf{w}_L^H(k) \hat{\mathbf{a}}_L(k)\} = 0 \\ & \text{Im}\{\mathbf{w}_R^H(k) \hat{\mathbf{a}}_R(k)\} = 0 \end{aligned} \quad (9)$$

A solução de (9) pode ser obtida de forma eficiente através do método de pontos interiores [6]. A principal limitação dessa formulação é o projeto dos parâmetros $\zeta_L(k)$ e $\zeta_R(k)$, os quais estabelecem um compromisso entre a distorção da fala e a redução de ruído. O projeto dos parâmetros de controle do conformador é um problema em aberto, o qual depende das características específicas de cada aplicação.

III. PROJETO DO CONFORMADOR WCO-BMVDR

O conformador WCO-BMVDR descrito na equação (9) requer a definição de dois parâmetros de projeto: $\zeta_L(k)$ e $\zeta_R(k)$. Esses parâmetros estabelecem um compromisso entre a redução de ruído e qualidade da fala. Neste trabalho, propõe-se a utilização da diferença de nível interaural (ILD, *interaural level difference*) para a configuração dos parâmetros do conformador. A ILD é uma pista binauricular relacionada à diferença entre a potência acústica em ambos os ouvidos [7]. Para uma fonte pontual no espaço, a ILD pode ser definida como $ILD_c(k) = |c_{L,0}(k)/c_{R,0}(k)|^2$ [7].

A. Projeto dos parâmetros $\zeta_L(k)$ e $\zeta_R(k)$

Considerando que os erros de estimação são pequenos, ζ_R é definido como uma porcentagem da ILD da fala estimada, $\zeta_R = \eta \cdot \hat{ILD}_s^{1/2}$ para $0 \leq \eta \leq \eta_{max}$, obedecendo a relação:

$$(1-\eta)^2 \hat{ILD}_s \leq ILD_c \leq (1+\eta)^2 \hat{ILD}_s \quad (10)$$

Uma relação semelhante é usada no projeto de ζ_L , através de um percentual da inversa da ILD da fala estimada, $\zeta_L = \gamma \cdot \hat{ILD}_s^{-1/2}$, para $0 \leq \gamma \leq \gamma_{max}$, na forma:

$$(1+\gamma)^{-2} \hat{ILD}_s \leq ILD_c \leq (1-\gamma)^{-2} \hat{ILD}_s \quad (11)$$

As equações (10) e (11) estabelecem que a região de incerteza será definida através de uma variação de azimute ao redor da posição estimada da fonte de interesse.

Assumindo-se, por simplicidade, o caso particular em que $\eta(k) = \gamma(k) = \delta(k)$, o conformador WCO-BMVDR proposto requer o projeto de um conjunto de K parâmetros $\delta(k)$. O parâmetro ótimo é aquele que maximiza a qualidade da fala e o

conforto acústico. Porém, não existe solução teórica para esse problema.

B. Diferença de nível interaural

De acordo com a literatura, a ILD estimada pode ser modelada como uma variável aleatória gaussiana ou t -student [12]. Em geral, a função de densidade de probabilidade (PDF, *probability density function*) da ILD pode ser influenciada pelos seguintes fatores: a SNR de entrada, o ângulo de azimute, o ângulo de elevação, a faixa de frequência e as características do ruído [13]. A Fig. 1 mostra o histograma da ILD, obtido através de experimentos de simulação utilizando sinais reais de fala contaminados por ruído artificial [14] para as frequências: 1 kHz, 2 kHz, e 4 kHz. Observa-se que a média das PDFs varia com a frequência, e nota-se o crescimento da variância quando a SNR decresce, corroborando resultados apresentados em [13].

Assumindo-se uma distribuição gaussiana para a ILD e considerando uma abordagem conservativa em relação à distorção da fala, o conjunto de incerteza pode ser definido através de $20 \log_{10}(1 + \delta_{con}(k)) = 3 \sigma_{ILD}(k)$, em que $\sigma_{ILD}(k)$ é uma estimativa do desvio padrão, de forma a incluir 99,7% das amostras de ILD da fala, resultando em:

$$\delta_{con}(k) = 10^{\frac{3 \hat{\sigma}_{ILD}(k)}{20}} - 1 \quad (12)$$

De forma a evitar a degradação de desempenho no processo de redução de ruído, propõe-se o uso de um parâmetro robusto restrito, definido como:

$$\delta_{res} = \alpha \times \text{median}\{\delta_{con}(k) \mid 0 \leq k \leq K-1\}, \quad (13)$$

em que $0 \leq \alpha \leq 1$ é o fator restritivo, definido empiricamente; e $\text{median}(\cdot)$ é a mediana do conjunto que contém os parâmetros conservadores $\delta_{con}(k)$, para todos os bins $0 \leq k \leq K-1$. Esse método heurístico permite não somente obter um compromisso entre qualidade de fala e redução de ruído, mas também permite obter um único parâmetro de robustez para todos os bins.

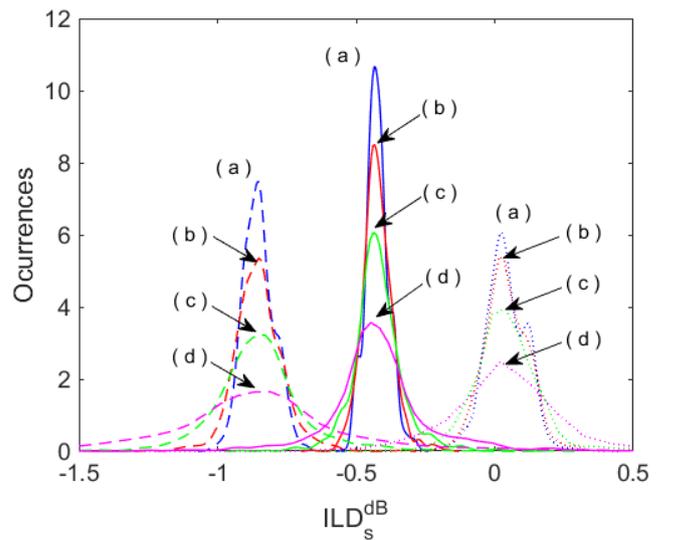


Fig. 1. Histograma da ILD obtido com 6 sinais de fala contaminadas com ruído ICRA-1 [14]. SNRs de entrada: (a) 20 dB (azul); (b) 15 dB (vermelho); (c) 10 dB (verde); e (d) 5 dB (magenta). Três frequências foram consideradas: 1 kHz (linhas tracejadas); 2 kHz (linhas sólidas); e 4 kHz (linhas pontilhadas).

IV. SIMULAÇÕES COMPUTACIONAIS

A. Descrição experimental

Com o intuito de realizar simulações computacionais, foi utilizado um banco de respostas impulsivas de diferentes cenários acústicos [15], o qual considera dois aparelhos auditivos de tipo BTE (*behind-the-ear*) colocados em cada ouvido de uma cabeça e torso artificiais de forma humana. Cada dispositivo possui $M = 3$ microfones com separação de 7,3 mm, os quais operam sob configuração biauricular. Foram consideradas duas fontes acústicas pontuais: uma fonte de fala alocada no azimute 0° (a frente do usuário) e uma fonte de ruído alocada em $+45^\circ$ (hemisfério direito), ambas com distâncias radiais de 0,8 m e ângulos de elevação de 0° . O ambiente acústico considerado foi uma câmara anecoica, com tempo de reverberação menor que 50 ms.

Um conjunto de 6 sinais de fala produzidas por um falante foi obtido de [16]. As frases têm duração média de 3,8 segundos. O tipo de ruído considerado é ICRA-1 [14]. Os sinais foram filtrados pelas respostas impulsivas obtidas em [15], considerando-se SNRs de -10 dB até 30 dB em passos de 5 dB, e 30 realizações com diferentes trechos de ruído.

B. Processamento de sinal

Os sinais de fala e ruído foram amostrados a $f_s = 16$ kHz. Foi aplicada a transformada de Fourier de tempo curto com $K = 256$ bins de frequência, usando uma janela de Hanning de 128 pontos, e 50% de sobreposição.

Foi utilizado o VAD descrito em [17]. Na ausência de fala é utilizado o método de covariância amostral apresentado em [5] para estimar a matriz de coerência do ruído. Na presença de fala a matriz de coerência da fala é estimada utilizando o método de subtração de covariâncias [18]. A estimação da matriz da coerência da fala pode ser utilizada para obter de forma direta os SVs esquerdo e direito, aplicando o método de mínima distorção baseado nos RTFs, apresentado em [11].

Foram realizadas comparações de desempenho entre o conformador BMVDR ideal (I-BMVDR), descrito em (7), que utiliza parâmetros ideais (matriz de covariância do ruído e os SVs da fala); o conformador BMVDR estimado (E-BMVDR), considerando a equação (7) e as estimativas destes parâmetros; e o conformador robusto (WCO-BMVDR) utilizando as mesmas estimativas para o conformador E-BMVDR e o parâmetro de robustez apresentado em (13).

Dois critérios objetivos foram utilizados para a avaliação de desempenho dos conformadores: (a) *Wideband Perceptual Evaluation of Speech Quality* (WPESQ), a qual utiliza a escala *Mean Opinion Score* (MOS), para avaliar a qualidade da fala; e a (b) razão sinal-ruído biauricular (média de ambas as orelhas), para avaliação de conforto acústico (redução de ruído). Os critérios MOS-WPESQ e SNR foram calculados para uma faixa de SNR de entrada de -10 dB a 30 dB em passos de 5 dB. Incrementos menores que 0,2 MOS-WPESQ e 3 dB de SNR são considerados não perceptíveis [19] [20].

C. Medidas objetivas versus SNR de entrada

A Fig. 2 mostra a comparação entre os conformadores I-BMVDR (azul), E-BMVDR (vermelho), e WCO-BMVDR para $\alpha = 0,10$ (magenta); $\alpha = 0,75$ (ciano); e $\alpha = 1,00$ (verde). Em preto são apresentadas as métricas relativas ao sinal não-

processado como referência de comparação. O conformador I-BMVDR incrementa enormemente ambas métricas objetivas em relação ao sinal não processado, resultando em um aumento de até 3,3 MOS-WPESQ e 43 dB. Por outro lado, a implementação prática convencional (conformador E-BMVDR) resulta em aumentos de até 2,0 MOS-WPESQ e 25 dB. Isso demonstra que apesar dos erros de estimação, o conformador E-BMVDR resulta em significativo aumento das métricas objetivas avaliadas em uma larga faixa de SNRs. Para valores elevados de α ($\geq 0,75$), o conformador WCO-BMVDR pode melhorar o desempenho em até 1,1 MOS-WPESQ e 6,2 dB de SNR em comparação com o conformador E-BMVDR. Esses incrementos são relevantes do ponto de vista psicoacústico [19] [20].

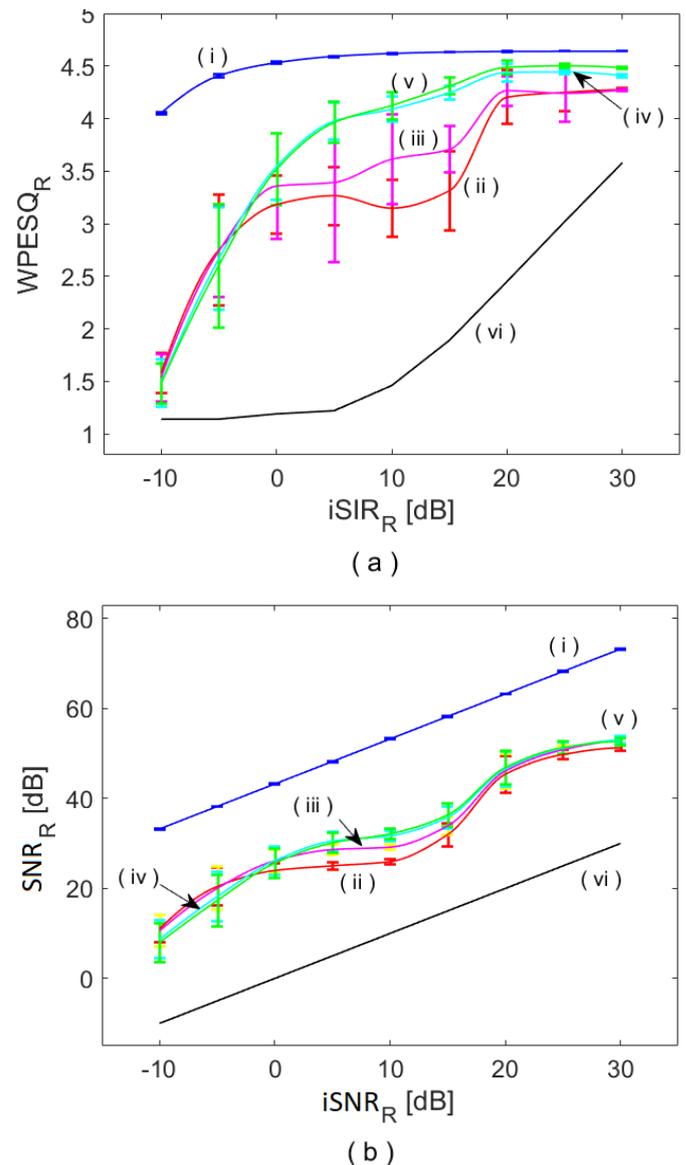


Fig. 2. Critérios objetivos MOS-WPESQ (a) e SNR (b) para os conformadores: (i) I-BMVDR (azul); (ii) E-BMVDR (vermelho); e WCO-BMVDR (iii) para $\alpha = 0,10$ (magenta), (iv) $\alpha = 0,75$ (ciano), e (v) $\alpha = 1,00$ (verde). Sinais não-processados em preto (vi).

V. DISCUSSÃO

A proposta de metodologia de projeto dos parâmetros de controle do conformador WCO-BMVDR, dentro da aplicação de aparelhos auditivos biauriculares, baseia-se na utilização da função de densidade de probabilidade aproximada da ILD da fala contaminada, obtida nos microfones de referência. Esse processo é suportado pelo fato de que as curvas da ILD, obtidas em aparelhos auditivos do tipo BTE, são similares às obtidas em microfones colocados internamente nas orelhas do usuário [21].

A Fig. 2 mostrou o comportamento da MOS-WPESQ e da SNR em função da SNR de entrada, demonstrando que o conformador WCO-BMVDR é uma técnica promissora para ser utilizada em aparelhos auditivos biauriculares, permitindo aumento de MOS-WPESQ e SNR para SNR de entrada (SNR_{in}) na faixa de $0 \text{ dB} < SNR_{in} < 15 \text{ dB}$, a qual é crucial em aplicações de aparelhos auditivos [7].

Embora os resultados preliminares do conformador proposto sejam promissores, há indícios de que o parâmetro de restrição (α) ótimo seja dependente das características espectrais do ruído de contaminação.

VI. CONCLUSÃO

O presente trabalho apresentou a proposta de um conformador de feixe robusto, chamado WCO-BMVDR, para aplicação em aparelhos auditivos biauriculares. O conformador WCO-BMVDR tem como objetivo aumentar o desempenho do conformador BMVDR convencional, que utiliza parâmetros estimados, através da inclusão de um par de restrições de desigualdade no problema de minimização original. Para atingir esse objetivo, foi proposto um método heurístico para projetar os seus parâmetros de controle, utilizando a função de densidade de probabilidade da diferença de nível interaural da fala contaminada, estimada a partir dos microfones de referência em ambas as orelhas. Os resultados apresentados indicam um aumento psicoacusticamente relevante da qualidade da fala e do conforto acústico em relação à implementação convencional do conformador de feixe biauricular MVDR.

REFERÊNCIAS

- [1] O. L. Frost, "An algorithm for linearly constrained adaptive array processing", *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926-935, 1972.
- [2] S. Gannot, D. Burshtein e E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech", *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614-1626, 2001.
- [3] E. Hadad, S. Doclo e S. Gannot, "The binaural LCMV beamformer and its performance analysis", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 543-558, 2016.
- [4] H. Chen, "Robustness analysis of nearfield subband beamformers in the presence of microphone gain and phase errors", *Digital Signal Processing: A Review Journal*, vol. 23, no. 5, pp. 1712-1719, 2013.
- [5] B. Cauchi, I. Kodrasi, R. Rehr, S. Gerlach, A. Jukic, T. Gerkmann, S. Doclo e S. Goetze, "Combination of MVDR beamforming and single-channel spectral processing for enhancing noisy and reverberant speech", *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 61, pp. 1-12, 2015.
- [6] S. A. Vorobyov, A. B. Gershman e Z. Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem", *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 313-324, 2003.
- [7] S. Doclo, S. Gannot, M. Moonen e S. Spriet, "Acoustic beamforming for hearing aid applications", in *Handbook on Array Processing and Sensor Networks*, John Wiley & Sons, 2008, pp. 269-302.
- [8] E. Hadad, D. Marquardt, S. Doclo e S. Gannot, "Theoretical analysis of binaural transfer function MVDR beamformers with interference cue preservation constraints", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2449-2464, 2015.
- [9] S. Gannot, E. Vincent, S. Markovich-Golan e A. Ozerov, "A consolidated perspective on multi-microphone speech enhancement and source separation", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 4, pp. 692-730, 2017.
- [10] D. M. Carmo e M. H. Costa, "Online approximation of the multichannel Wiener filter with preservation of interaural level difference for binaural hearing aids", *Computers in Biology and Medicine*, vol. 95, pp. 188-197, 2018.
- [11] M. Taseska e E. Habets, "Relative transfer function estimation exploiting instantaneous signals and the signal subspace", in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2015, pp. 404-408.
- [12] A. Deleforge e F. Forbes, "Rectified binaural ratio: a complex t-distributed feature for robust sound localization", in *Proceedings of European Signal Processing Conference (EUSIPCO)*, 2016, pp. 1257-1261.
- [13] J. Nix e V. Hohmann, "Sound source localization in real sound fields based on empirical statistics of interaural parameters", *The Journal of the Acoustical Society of America*, vol. 119, pp. 463-479, 2006.
- [14] 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", *Audiology*, vol. 40, pp. 148-157, 2001.
- [15] H. Kayser, S. D. Ewert, J. Anemuller, T. Rohdenburg, V. Hohmann e B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses", *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 1-10, 2009.
- [16] Y. Hu e P. Loizou, "Subject evaluation and comparison of speech enhancement algorithm", *Speech Communication*, vol. 49, pp. 588-601, 2007.
- [17] J. Sohn, N. S. Kim e W. Sung, "A statistical model-based voice activity detection", *IEEE Signal Processing Letters*, vol. 6, no. 1, pp. 1-3, 1999.
- [18] E. A. Habets e J. Benesty, "A perspective on frequency-domain beamformers in room acoustics", *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 3, pp. 947-960, 2012.
- [19] A. Servetti e J. C. D. Martin, "802.11 MAC protocol with selective error detection for speech transmission", in *Quality of Service in Multiservice IP Networks*, Springer, 2005, pp. 509-519.
- [20] D. McShefferty, W. M. Whitmer e M. A. Akeroyd, "The just meaningful difference in speech-to-noise ratio", *Trends in Hearing*, vol. 20, pp. 1-11, 2016.
- [21] H. G. Jones, A. Kan e R. Y. Litovsky, "The effect of microphone placement on interaural level differences and sound localization across the horizontal plane in bilateral cochlear implant users", *Ear Hearing*, vol. 37, no. 5, pp. 1-10, 2016.