



Noise Reduction by Beamforming in Cochlear Implants: a Time Domain Study

Luciana M X Souza, Marcio H Costa and Renata C Borges

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

October 6, 2022

Redução de ruído por conformação de feixe em implantes cocleares: Um estudo no domínio do tempo

L. M. X. Souza¹, M. H. Costa¹ e R. C. Borges²

¹Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, Florianópolis-SC, Brasil

²Departamento Acadêmico de Engenharia Elétrica, Universidade Tecnológica Federal do Paraná, Cornélio Procopio-PR, Brasil

Abstract— Cochlear implants are devices with the aim of restoring communication capability in cases of severe to profound hearing loss. For this purpose, acoustic signals are transformed into electrical stimuli and then directly applied to the cochlea through a set of electrodes. Despite its efficacy in restoring communication under optimal conditions, its performance is severely degraded due to additive noise. In this work, a comparison between two variations of the most common beamforming method for noise reduction in cochlear implants is performed. Computational simulations with the Minimum Variance Distortionless Response beamforming (MVDR), using noisy-speech and noise-only correlation matrices were performed, and results for intelligibility and output signal-to-interference-plus-noise ratio are presented. The results indicate that the MVDR beamformer based on the noise correlation matrix provides higher intelligibility and acoustic comfort. This finding justifies the choice of the MVDR cost function in cochlear implant applications.

Keywords— Cochlear implant, noise reduction, beamformer, speech processing.

I. INTRODUÇÃO

O sistema auditivo é de extrema importância para a comunicação entre seres humanos, sendo responsável pela captação e codificação da informação sonora. Dificuldades de comunicação em virtude da redução da capacidade de ouvir pode gerar prejuízos pessoais, sociais e econômicos [1]. As causas mais comuns da perda auditiva neurossensorial estão relacionadas à idade e à exposição ao ruído. Hoje, estima-se que cerca de 4,7% da população mundial viva com algum tipo de perda auditiva debilitante, podendo chegar a um total de 630 milhões de pessoas em 2030, e 900 milhões até o ano de 2050 [1].

Implantes cocleares (IC) constituem uma tecnologia de referência quando se trata de próteses neurais para compensação de perdas auditivas, pois apresentam alta taxa de sucesso na restauração da percepção sonora [2]. O IC estimula eletricamente as terminações do nervo auditivo por meio de eletrodos inseridos no interior da cóclea [4], restaurando a capacidade de comunicação em ambientes silenciosos. No entanto, a compreensão da fala continua sendo um grande

desafio em ambientes acusticamente complexos, devido principalmente à reverberação e ao ruído ambiente [5].

Técnicas de redução de ruído por conformação de feixe têm sido amplamente utilizadas para a redução de interferências decorrentes de fontes acústicas pontuais indesejadas. Para tanto, utilizam sinais captados por múltiplos microfones separados espacialmente, de forma a reduzir a interferência indesejada sem ocasionar distorções percebíveis no sinal de interesse. O conformador de feixe explora as diferenças de fase e magnitude dos sinais que chegam aos microfones para estabelecer diferentes padrões de direcionalidade. Dessa forma, estabelecem-se diferentes níveis de sensibilidade para distintos ângulos de chegada [6]. Essa seleção direcional pode reduzir drasticamente a quantidade de ruído se as fontes de interesse e interferentes estiverem separadas espacialmente [6].

Entre as diferentes estratégias de conformação de feixe que podem ser encontradas na literatura, as mais utilizadas em processamento de fala são o método de mínima variância com resposta sem distorção (*Minimum Variance Distortionless Response* - MVDR) e o de mínima variância com restrições lineares (*Linear Constrained Minimum Variance* - LCMV) [7]. Embora tenham sido amplamente estudadas em diferentes contextos, algumas perguntas permanecem em aberto na área de implantes cocleares. Em especial, quais as vantagens e desvantagens das duas principais formas de implementação do conformador MVDR com relação à função custo a ser minimizada: seja através da utilização da matriz de correlação da fala contaminada ou a do ruído global [7].

Neste artigo apresenta-se uma análise de desempenho das duas formas mais comuns de implementação do conformador MVDR para diferentes razões sinal-ruído e sinal-interferência, com o objetivo de verificar seu impacto na inteligibilidade da fala em aplicações de redução de ruído em implantes cocleares.

Este artigo é estruturado da seguinte forma: Na Seção II é apresentada a modelagem dos sinais e processamento da fala contaminada, enquanto que na Seção III é apresentada a derivação do conformador MVDR para ambas as formas de implementação abordadas. Na Seção IV são descritas as simulações realizadas e na Seção V apresentados e discutidos os resultados. Finalizando, a Seção 0 apresenta a conclusão do

trabalho.

Ao longo do texto, variáveis escalares são representadas por letras minúsculas em itálico, vetores por letras minúsculas em negrito, matrizes por letras maiúscula em negrito e os símbolos $(\cdot)^T$ e $(\cdot)^{-1}$ representam as operações de transposição e inversão, respectivamente.

II. CARACTERIZAÇÃO DOS SINAIS E MODELAGEM DA CADEIA DE PROCESSAMENTO EM IMPLANTES COCLEARES

A fala possui informações relevantes à inteligibilidade em uma faixa de frequências que varia entre 50 Hz e 8 kHz, sendo sua densidade espectral de potência concentrada preponderantemente abaixo de 4 kHz [8]. Do ponto de vista estocástico, embora seja um sinal não estacionário, suas estatísticas de segunda ordem podem ser consideradas aproximadamente constantes por períodos curtos (10 a 40 ms) de tempo [9]. Em geral, um grande conjunto de ruídos ambientais possui variações de suas estatísticas ainda mais lentas que as da fala [10].

A. Modelagem da fala contaminada

Implantes cocleares podem ser unilaterais ou bilaterais. Entretanto, mesmo no caso de um implante unilateral é possível a utilização de microfones na orelha contralateral, com o objetivo de aumentar a resolução espacial do campo acústico captado.

Os sinais acústicos captados por M microfones de aquisição são definidos da seguinte forma:

$$y_m(n) = x_m(n) + v_m(n) \quad (1)$$

em que $n = tf_s$ é o índice de tempo discreto; f_s é a frequência de amostragem; m é o índice do microfone; $y_m(n)$ é a fala contaminada no m -ésimo microfone de captação;

$$x_m(n) = \sum_{q=0}^{N_1-1} a_{m,q} s(n-q) \quad (2)$$

é a fala de interesse na posição do m -ésimo microfone; $s(n)$ é a fala de interesse no local de geração; $\{ a_{m,0} a_{m,1} \dots a_{m,N_1-1} \}$ são os coeficientes da resposta ao impulso do sistema acústico entre a fonte de fala e o m -ésimo microfone de captação;

$$v_m(n) = \sum_{p=1}^P f_{p,m}(n) + r_m(n) \quad (3)$$

é o ruído global na posição do m -ésimo microfone, composto por uma parcela de ruído não-correlacionado (representando ruído elétrico mais ruído difuso, $r_m(n)$) e a contribuição de P fontes pontuais interferentes, de forma que:

$$f_{p,m}(n) = \sum_{q=0}^{N_p-1} b_{p,m,q} i_p(n-q) \quad (4)$$

é o p -ésimo sinal interferente na posição do m -ésimo microfone; $\{ b_{p,m,0} b_{p,m,1} \dots b_{p,m,N_p-1} \}$ são os coeficientes da resposta ao impulso do sistema acústico entre a p -ésima fonte interferente e o m -ésimo microfone; e $i_p(n)$ é o sinal da p -ésima fonte pontual interferente no local de geração. Assume-se que os caminhos acústicos não variam significativamente dentro de uma determinada janela de observação.

Substituindo-se (2), (3) e (4) em (1), obtém-se:

$$y_m(n) = \sum_{q=0}^{N_1-1} a_{m,q} s(n-q) + \sum_{p=1}^P \sum_{q=0}^{N_p-1} b_{p,m,q} i_p(n-q) + r_m(n) \quad (5)$$

A equação (5) modela o sinal captado em cada um dos M microfones, compreendendo as influências da fala, sinais interferentes decorrentes de fontes pontuais, ruído elétrico dos microfones e ruído difuso. Adicionalmente são incluídas as informações referentes aos caminhos de propagação entre as fontes pontuais e microfones. Neste modelo as reflexões tardias do ambiente podem ser associadas à parcela de $r_m(n)$.

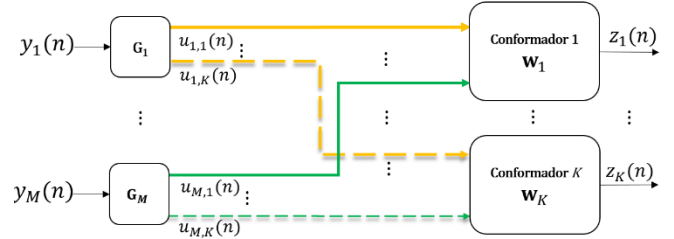


Fig. 1. Diagrama em blocos de um sistema de redução de ruído multimicrofone, baseado em conformação de feixe, para implante coclear. Assumem-se M microfones e K sub-bandas de estimulação.

B. Cadeia de processamento para redução de ruído em implantes cocleares

Sistemas de redução de ruído para implantes cocleares são caracterizados pela utilização de um banco de filtros para decomposição dos sinais captados em K sub-bandas (blocos G_i na Fig. 1). Considerando-se um conjunto de M microfones de captação são utilizados M bancos de filtros idênticos. Cada uma das K saídas de cada um dos bancos é obtida pela soma de convolução entre o sinal de entrada no m -ésimo microfone e a resposta ao impulso da k -ésima sub-banda, definida pelos coeficientes $\{ g_{k,0} g_{k,1} \dots g_{k,N-1} \}$. Dessa forma, a saída de cada uma das sub-bandas, para cada um dos microfones, é dada por:

$$u_{m,k}(n) = \sum_{l=0}^{N-1} g_{k,l} y_m(n-l) \quad (6)$$

em que N é o número de coeficientes da resposta ao impulso do banco de filtros. Assumindo-se uma janela de C amostras de observação para o conformador, temos:

$$\mathbf{u}_{m,k}(n) = \mathbf{G}_k \mathbf{y}_m(n) \quad (7)$$

em que $\mathbf{u}_{m,k}(n) = [u_{m,k}(n) \ u_{m,k}(n-1) \ \dots \ u_{m,k}(n-C+1)]^T$; $\mathbf{y}_m(n) = [y_m(n) \ y_m(n-1) \ \dots \ y_m(n-N-C+2)]^T$; e

$$\mathbf{G}_k = \begin{bmatrix} g_{k,1} & g_{k,2} & \dots & g_{k,N} & 0 & 0 & \dots & 0 \\ 0 & g_{k,1} & g_{k,2} & \dots & g_{k,N} & \dots & \dots & \dots \\ 0 & 0 & g_{k,1} & g_{k,2} & \dots & g_{k,N} & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad (8)$$

possui dimensão $C \times (N+C-1)$. Dessa forma, pode-se então definir a saída do conformador como:

$$z_k(n) = \sum_{m=1}^M \mathbf{w}_{m,k}^T(n) \mathbf{u}_{m,k}(n) = \sum_{m=1}^M \mathbf{w}_{m,k}^T(n) \mathbf{G}_k \mathbf{y}_m(n) \quad (9)$$

em que $\mathbf{w}_{m,k}(n) = [w_{m,k,0}(n) \ w_{m,k,1}(n) \ \dots \ w_{m,k,C-1}(n)]^T$ é o vetor de coeficientes do conformador referente a k -ésima sub-banda do m -ésimo microfone. A equação (9) pode ser expressa também sob a forma do seguinte produto interno entre dois vetores

$$z_k(n) = \mathbf{w}_k^T(n) \mathbf{u}_k(n) \quad (10)$$

em que $\mathbf{w}_k(n) = [\mathbf{w}_{1,k}^T(n) \ \mathbf{w}_{2,k}^T(n) \ \dots \ \mathbf{w}_{M,k}^T(n)]^T$ e $\mathbf{u}_k(n) = [\mathbf{u}_{1,k}^T(n) \ \mathbf{u}_{2,k}^T(n) \ \dots \ \mathbf{u}_{M,k}^T(n)]^T$ são vetores de dimensão $MC \times 1$, sendo, respectivamente, formados pelo empilhamento dos vetores $\mathbf{u}_{m,k}(n)$ e $\mathbf{w}_{m,k}(n)$ para $m = 1, 2, \dots, M$, de forma que

$$\mathbf{u}_k(n) = \mathbf{\Gamma}_k \mathbf{y}(n) \quad (11)$$

em que $\mathbf{y}(n) = [\mathbf{y}_1^T(n) \ \mathbf{y}_2^T(n) \ \dots \ \mathbf{y}_M^T(n)]^T$ possui dimensão $M(N+C-1) \times 1$ e

$$\mathbf{\Gamma}_k = \begin{bmatrix} \mathbf{G}_k & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_k & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G}_k \end{bmatrix} \quad (12)$$

é uma matriz bloco-diagonal de dimensão $MC \times M(N+C-1)$. Portanto, pode-se escrever a saída do conformador para cada uma das K sub-bandas como:

$$z_k(n) = \mathbf{w}_k^T(n) \mathbf{\Gamma}_k \mathbf{y}(n) \quad (13)$$

Na sequência uma estimativa da potência de $z_k(n)$ (para uma determinada janela de duração J) é calculada e então utilizada uma estratégia de estimulação [11]. Finalizando a cadeia de processamento, a potência de cada uma das sub-bandas selecionadas é transformada em um trem de pulsos de corrente para ser aplicado à cóclea.

III. CONFORMADOR DE FEIXE NO DOMÍNIO TEMPO

Substituindo-se (1) em (13) obtém-se:

$$z_k(n) = \mathbf{w}_k^T(n) \mathbf{\Gamma}_k \mathbf{x}(n) + \mathbf{w}_k^T(n) \mathbf{\Gamma}_k \mathbf{v}(n) \quad (14)$$

em que $\mathbf{x}(n) = [\mathbf{x}_1^T(n) \ \mathbf{x}_2^T(n) \ \dots \ \mathbf{x}_M^T(n)]^T$ e $\mathbf{v}(n) = [\mathbf{v}_1^T(n) \ \mathbf{v}_2^T(n) \ \dots \ \mathbf{v}_M^T(n)]^T$ possuem dimensão $M(N+C-1) \times 1$; e

$\mathbf{x}_m(n) = [x_m(n) \ x_m(n-1) \ \dots \ x_m(n-N-C+2)]^T$ e $\mathbf{v}_m(n) = [v_m(n) \ v_m(n-1) \ \dots \ v_m(n-N-C+2)]^T$ possuem dimensão $(N+C-1) \times 1$. Dessa forma, $z_k(n)$ é composto por duas parcelas: o resultado do processo de filtragem realizado pelo conformador sobre a fala e sobre o ruído. A partir de (14) pode-se calcular a potência total na saída do conformador, dada por

$$\begin{aligned} E\{z_k^2(n)\} &= \mathbf{w}_k^T(n) \mathbf{\Gamma}_k E\{\mathbf{x}(n) \mathbf{x}^T(n)\} \mathbf{\Gamma}_k^T \mathbf{w}_k(n) \\ &\quad + \mathbf{w}_k^T(n) \mathbf{\Gamma}_k E\{\mathbf{x}(n) \mathbf{v}^T(n)\} \mathbf{\Gamma}_k^T \mathbf{w}_k(n) \\ &\quad + \mathbf{w}_k^T(n) \mathbf{\Gamma}_k E\{\mathbf{v}(n) \mathbf{x}^T(n)\} \mathbf{\Gamma}_k^T \mathbf{w}_k(n) \\ &\quad + \mathbf{w}_k^T(n) \mathbf{\Gamma}_k E\{\mathbf{v}(n) \mathbf{v}^T(n)\} \mathbf{\Gamma}_k^T \mathbf{w}_k(n) \end{aligned} \quad (15)$$

em que $E\{\cdot\}$ é a operação de valor esperado.

Assumindo-se que fala e ruído são variáveis aleatórias não correlacionadas de média zero, tem-se que $E\{\mathbf{x}(n) \mathbf{v}^T(n)\} = E\{\mathbf{v}(n) \mathbf{x}^T(n)\} = \mathbf{0}$ e, portanto:

$$\begin{aligned} E\{z_k^2(n)\} &= \mathbf{w}_k^T(n) \mathbf{\Gamma}_k \mathbf{R}_{\mathbf{xx}}(n) \mathbf{\Gamma}_k^T \mathbf{w}_k(n) \\ &\quad + \mathbf{w}_k^T(n) \mathbf{\Gamma}_k \mathbf{R}_{\mathbf{vv}}(n) \mathbf{\Gamma}_k^T \mathbf{w}_k(n) \end{aligned} \quad (16)$$

em que $\mathbf{R}_{\mathbf{xx}}(n) = E\{\mathbf{x}(n) \mathbf{x}^T(n)\}$ e $\mathbf{R}_{\mathbf{vv}}(n) = E\{\mathbf{v}(n) \mathbf{v}^T(n)\}$

A partir de (16) pode-se definir duas possíveis funções objetivo: a minimização de $E\{z_k^2(n)\}$ propriamente dita ou apenas da parcela associada ao ruído global, de forma que:

$$\mathbf{w}_k^{\text{opt}}(n) = \arg.\min_{\mathbf{w}} \mathbf{w}^T \mathbf{\Gamma}_k \mathbf{R}_{\mathbf{vv}}(n) \mathbf{\Gamma}_k^T \mathbf{w} \quad (17)$$

em que $\mathbf{R}_{\mathbf{vv}}(n)$ pode ser definida como $\mathbf{R}_{\mathbf{vv}}(n)$ ou $\mathbf{R}_{\mathbf{yy}}(n) = \mathbf{R}_{\mathbf{xx}}(n) + \mathbf{R}_{\mathbf{vv}}(n)$.

A solução trivial para $\mathbf{R}_{\mathbf{vv}}(n) = \mathbf{R}_{\mathbf{yy}}(n)$ é $\mathbf{w}_k^{\text{opt}}(n) = \mathbf{0}$, enquanto que no caso de $\mathbf{R}_{\mathbf{vv}}(n) = \mathbf{R}_{\mathbf{vv}}(n)$ não há garantia de preservação da fala para a solução encontrada. Dessa forma, é necessária a inclusão de restrições ao processo de minimização para evitar distorções do sinal de interesse.

A. Restrições para preservação da fala

Assumindo-se a inexistência de ruído aditivo, a saída do conformador, definida em (14), é:

$$z_k^{\text{fala}}(n) = \mathbf{w}_k^T(n) \mathbf{\Gamma}_k \mathbf{x}(n) \quad (18)$$

Nessa situação, a saída ótima do conformador é dada por

$$z_k^{\text{fala opt}}(n) = x_{\text{REF},k}(n) = \mathbf{q}_{\text{REF}}^T \mathbf{\Gamma}_k \mathbf{x}(n) \quad (19)$$

em que o $x_{\text{REF},k}(n)$ é o sinal de fala captado pelo microfone de referência (geralmente definido como o microfone frontal, REF = 1) na banda k ; \mathbf{q}_{REF} é um vetor de dimensão $MC \times 1$ cujos elementos são nulos, com uma única exceção, a qual é igual à unidade (sendo REF o número do microfone de referência, o índice do valor não nulo de \mathbf{q}_{REF} é dado por $(\text{REF}-1) \cdot C + 1$).

Substituindo-se (19) em (18), obtém-se uma restrição instantânea necessária para evitar qualquer distorção da fala:

$$\mathbf{w}_k^T(n) \mathbf{\Gamma}_k \mathbf{x}(n) = \mathbf{q}_{\text{REF}}^T \mathbf{\Gamma}_k \mathbf{x}(n) \quad (20)$$

A restrição definida em (20) não pode ser utilizada visto que o sinal $x(n)$ não é observável. Sendo $x(n)$ uma variável aleatória, a igualdade descrita em (20) implica que os infinitos momentos estatísticos entre ambos os termos devem ser idênticos aos respectivos momentos do sinal de referência com ele mesmo. Portanto, em aplicações práticas, flexibiliza-se (20) de forma que apenas o momento cruzado de primeira ordem entre ambos sinais seja idêntico à potência do sinal no microfone de referência. Para tanto, multiplica-se $\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{x}(n)$ em ambos os lados de (20) e toma-se o valor esperado, de forma que:

$$\begin{aligned} E\{\mathbf{w}_k^T(n) \Gamma_k \mathbf{x}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}}\} \\ = E\{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{x}(n) \mathbf{x}^T(n) \Gamma_k^T \mathbf{q}_{\text{REF}}\} \end{aligned} \quad (21)$$

resultando em

$$\mathbf{w}_k^T(n) \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}}, \quad (22)$$

que configura uma restrição linear única para a preservação da fala. Incluindo-se a restrição definida em (22) em (17) chega-se em:

$$\mathbf{w}_k^{\text{opt}}(n) = \arg \min_{\mathbf{w}} \mathbf{w}^T \Gamma_k \mathbf{R}_{\psi\psi}(n) \Gamma_k^T \mathbf{w} \quad (23)$$

$$\text{s.t. } \mathbf{w}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} = \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}}$$

Outras restrições, que não a utilizada em (22), podem ser incluídas em (17) de forma a constituir diferentes compromissos entre redução de ruído e preservação da fala

B. Solução fechada para o conformador MVDR

Uma formulação fechada para o vetor de coeficientes ótimo do conformador MVDR pode ser encontrada através da técnica de multiplicadores de Lagrange. Utilizando esse método, definimos o Lagrangiano de (23) como

$$\begin{aligned} L(\mathbf{w}_k(n), \lambda_k(n)) = & \mathbf{w}_k^T(n) \Gamma_k \mathbf{R}_{\psi\psi}(n) \Gamma_k^T \mathbf{w}_k(n) \\ & + \lambda_k(n) \left[\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{w}_k(n) \right. \\ & \left. - \mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} \right] \end{aligned} \quad (24)$$

Calculando-se a derivada parcial de (24) em relação a $\mathbf{w}_k(n)$, e igualando o resultado a zero, chega-se em

$$2\Gamma_k \mathbf{R}_{\psi\psi}(n) \Gamma_k^T \mathbf{w}_k(n) + \lambda_k \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} = \mathbf{0} \quad (25)$$

Manipulando-se (25) obtém-se:

$$\begin{aligned} \mathbf{w}_k^{\text{opt}}(n) = & -\lambda_k^{\text{opt}}(n) [2\Gamma_k \mathbf{R}_{\psi\psi}(n) \Gamma_k^T]^{-1} \\ & \cdot \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} \end{aligned} \quad (26)$$

O vetor $\mathbf{w}_k^{\text{opt}}(n)$ obtido em (26) deve satisfazer a restrição imposta em (22), portanto:

$$\lambda_k^{\text{opt}}(n) = \frac{-2\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}}}{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T [\Gamma_k \mathbf{R}_{\psi\psi}(n) \Gamma_k^T]^{-1} \Gamma_k \mathbf{R}_{\text{xx}}(n) \Gamma_k^T \mathbf{q}_{\text{REF}}} \quad (27)$$

Substituindo-se (27) em (26) chega-se à uma fórmula fechada para cálculo do vetor de coeficientes do conformador MVDR:

$$\mathbf{w}_k^{\text{opt}}(n) = \frac{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}} \mathbf{R}_{\psi\psi}(n) \Gamma_k^T \mathbf{q}_{\text{REF}} (\Gamma_k \mathbf{R}_{\psi\psi} \Gamma_k^T)^{-1} \Gamma_k \mathbf{R}_{\text{xx}} \Gamma_k^T \mathbf{q}_{\text{REF}}}{\mathbf{q}_{\text{REF}}^T \Gamma_k \mathbf{R}_{\text{xx}} \mathbf{R}_{\psi\psi}(n) \Gamma_k^T (\Gamma_k \mathbf{R}_{\psi\psi}(n) \Gamma_k^T)^{-1} \Gamma_k \mathbf{R}_{\text{xx}} \mathbf{R}_{\psi\psi}(n) \Gamma_k^T \mathbf{q}_{\text{REF}}} \quad (28)$$

em que $\mathbf{R}_{\psi\psi}(n)$ pode ser definida como $\mathbf{R}_{\text{vv}}(n)$ ou $\mathbf{R}_{\text{yy}}(n)$ [7].

Note que para $\mathbf{R}_{\psi\psi}(n) = \mathbf{R}_{\text{yy}}(n)$ o vetor de coeficientes ótimo do conformador MVDR é idêntico à solução do filtro de Wiener multicanal a menos de um fator de escala dado em (27) [12].

IV. SIMULAÇÕES COMPUTACIONAIS

Nesta seção é apresentada a descrição das simulações computacionais realizadas para avaliar o desempenho das duas formas de implementação do conformador MVDR descritas para as condições logo a seguir de (28).

O cenário acústico utilizado é constituído por uma sala de escritório com tempo de reverberação de 300 ms. Os sinais captados nos microfones foram simulados através da convolução entre os sinais de fala e ruído e a resposta ao impulso do ambiente ‘Office 1’ descrito em [12].

A fonte de fala foi assumida no azimute de 0° (em frente ao usuário do implante) e a fonte pontual de interferência em -60° (lado esquerdo). Como sinal de fala foram utilizados 15 áudios do banco de dados IEEE *corpus* [14]. O sinal interferente constitui-se por ruído artificial do tipo ICRA-1 (*Artificial Speech-Like Noise*) [15]. O ruído difuso também é um ruído artificial, porém do tipo ruído branco Gaussiano (não correlacionado). A frequência de amostragem é 16 kHz.

Assumiu-se a existência de 6 microfones, 3 em cada lado da cabeça do usuário; janela de observação do conformador $C = 1$ (em função de limitação na disponibilidade de recursos computacionais em dispositivos comerciais); $N = 200$; SNR $\in \{0, 30\}$ dB; e $-10 \text{ dB} \leq \text{SIR} \leq 30 \text{ dB}$ em passos de 5 dB.

O banco de filtros utilizado é do tipo gammatone com $K = 22$ sub-bandas [16].

O critério objetivo utilizado para a estimação da inteligibilidade foi o *Speech to Reverberation Modulation Energy Ratio for Cochlear Implants* (SRMR-CI), desenvolvido especificamente para ICs [17]. O SRMR-CI é mapeado em valores de inteligibilidade percentual através da seguinte equação:

$$I_{\%} = 88,92(1 + e^{a_1 \text{SRMR-CI}_{\text{norm}} + a_2})^{-1} \quad (29)$$

em que $I_{\%}$ é o percentual de inteligibilidade; $a_1 = -12,1742$ e $a_2 = 7,4535$ são os coeficientes de ajuste. A $\text{SRMR-CI}_{\text{norm}}$ é calculada da seguinte forma:

$$\text{SRMR-CI}_{\text{norm}} = (\text{SRMR-CI}_s)^{-1} \text{SRMR-CI}, \quad (30)$$

em que $SRMR-CI_t$ é a métrica do sinal em teste e $SRMR-CI_s$ é o valor da métrica para o sinal de fala limpo.

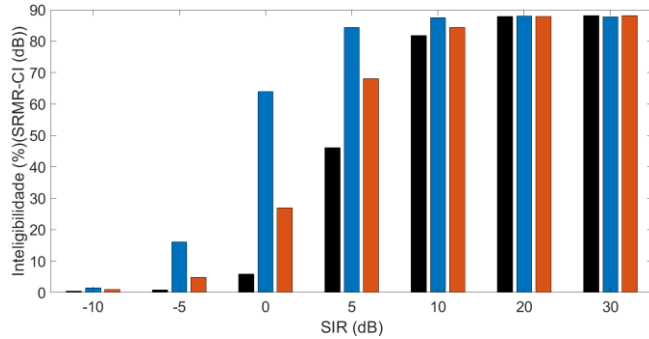


Fig. 2. Inteligibilidade percentual segundo o critério SRMR-CI em função da razão sinal-interferência (SIR) e SNR = 30 dB. Sinal contaminado (preto) e sinal contaminado processado pelo conformador MVDR, utilizando: (a) $R_{\Psi\Psi}(n) = R_{vv}(n)$ (azul) e (b) $R_{\Psi\Psi}(n) = R_{yy}(n)$ (laranja).

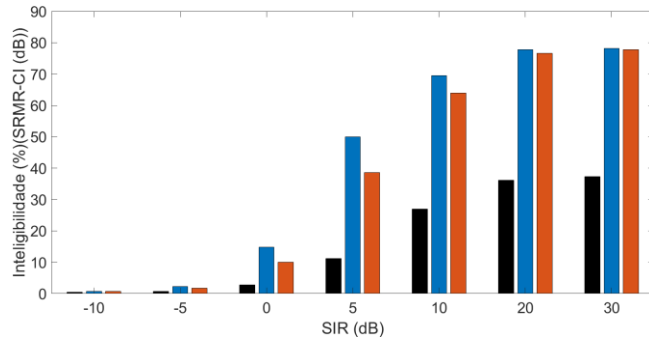


Fig. 3. Inteligibilidade percentual segundo o critério SRMR-CI em função da razão sinal-interferência (SIR) e SNR = 0 dB. Sinal contaminado (preto) e sinal contaminado processado pelo conformador MVDR, utilizando: (a) $R_{\Psi\Psi}(n) = R_{vv}(n)$ (azul) e (b) $R_{\Psi\Psi}(n) = R_{yy}(n)$ (laranja).

V. RESULTADOS

Na Fig. 2 são apresentados os resultados para SNR = 30 dB, que corresponde a uma situação em que o ruído global é basicamente decorrente da influência de uma única fonte pontual interferente, ou seja, com ruído não-correlacionado desprezível. Verifica-se claramente que a utilização de $R_{\Psi\Psi}(n) = R_{vv}(n)$ resulta em percentuais de inteligibilidade significativamente maiores. Considerando-se que para o estabelecimento adequado de conversação é necessário um percentual de inteligibilidade superior a 70% [16], o uso de $R_{\Psi\Psi}(n) = R_{vv}(n)$ permite a comunicação do usuário de implante coclear para $SIR \geq 5$ dB, enquanto que para $R_{\Psi\Psi}(n) = R_{yy}(n)$ é necessário $SIR \geq 10$ dB.

Na Fig. 3 a inteligibilidade percentual é avaliada para SNR = 0 dB, que corresponde a uma situação de intenso nível de ruído de fundo não correlacionado. Da mesma forma que para o caso anterior (SNR = 30 dB), a forma

$R_{\Psi\Psi}(n) = R_{vv}(n)$ apresenta níveis significativamente superiores de inteligibilidade em relação à $R_{\Psi\Psi}(n) = R_{yy}(n)$.

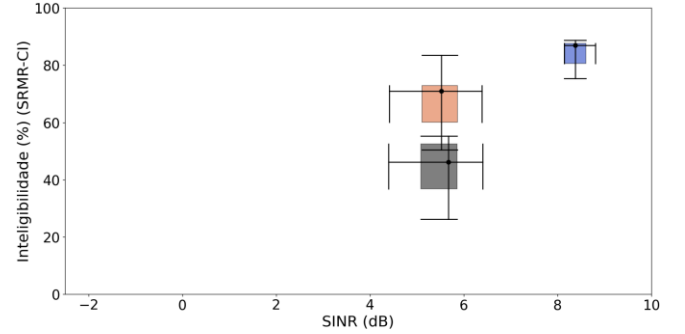


Fig. 4. Diagrama de caixas bidimensional para inteligibilidade percentual versus razão sinal-(interferência+ruído) de saída, para SNR = 30 dB e SIR = 5 dB. Sinal contaminado (preto) e sinal contaminado processado pelo conformador MVDR, utilizando: (a) $R_{\Psi\Psi}(n) = R_{vv}(n)$ (azul) e (b) $R_{\Psi\Psi}(n) = R_{yy}(n)$ (laranja).

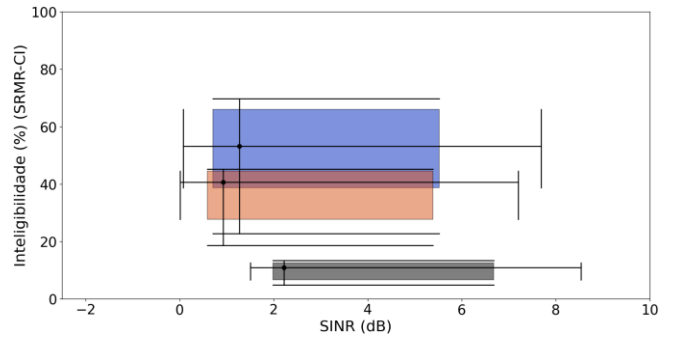


Fig. 5. Diagrama de caixas bidimensional para inteligibilidade percentual versus razão sinal-(interferência+ruído) de saída, para SNR = 0 dB e SIR = 5 dB. Sinal contaminado (preto) e sinal contaminado processado pelo conformador MVDR, utilizando: (a) $R_{\Psi\Psi}(n) = R_{vv}(n)$ (azul) e (b) $R_{\Psi\Psi}(n) = R_{yy}(n)$ (laranja).

A Fig. 4 e a Fig. 5 apresentam diagramas de caixa bidimensionais para a análise da relação entre inteligibilidade percentual e razão sinal-(interferência+ruído) (*Signal to Interference Plus Noise Ratio* - SINR) de saída, para $SIR = 5$ dB e $SNR \in \{0, 30\}$ dB. O desempenho global (inteligibilidade e conforto acústico) dos conformadores, pode ser estimado pela distância Euclidiana entre a origem do gráfico e a mediana do agrupamento. Desta forma, verifica-se que, consistentemente, a forma de implementação $R_{\Psi\Psi}(n) = R_{vv}(n)$ é a mais adequada para aplicações em implantes cocleares. Nota-se, entretanto, que à medida que o nível de ruído não correlacionado aumenta há uma maior variabilidade e consistente diminuição do valor absoluto das métricas; entretanto, mantendo as vantagens da forma $R_{\Psi\Psi}(n) = R_{vv}(n)$.

VI. DISCUSSÃO

O conformador de feixe MVDR possui duas formas de implementação, amplamente utilizadas na literatura, sendo elas a minimização da função custo baseada na matriz de correlação da fala-contaminada ($\mathbf{R}_{\Psi\Psi}(n)=\mathbf{R}_{yy}(n)=\mathbf{R}_{xx}(n)+\mathbf{R}_{vv}(n)$) [10, 18] e a baseada na matriz de correlação do ruído ($\mathbf{R}_{\Psi\Psi}(n)=\mathbf{R}_{vv}(n)$) [8, 12]. Entretanto, os trabalhos publicados não apresentam justificativa para sua escolha.

Os resultados obtidos no presente trabalho indicam que, para o caso de aplicações voltadas para implantes cocleares, a forma de implementação baseada na matriz de correlação do ruído global apresenta níveis de inteligibilidade significativamente maiores mesmo em situações de SNR muito baixa, como, por exemplo, SNR=0 dB.

VII. CONCLUSÕES

Este trabalho apresentou a comparação entre duas formas de implementação do conformador de feixes MVDR em sistemas de redução de ruído para implantes cocleares. Simulações computacionais baseadas em um critério objetivo de inteligibilidade e na razão sinal-(interferência+ruído) de saída indicam uma diferença significativa de desempenho, sendo que a forma baseada na minimização da potência do ruído global proporciona maior inteligibilidade e conforto acústico. Adicionalmente, verificou-se que a existência de níveis elevados de ruído de fundo não-correlacionado possuem impacto significativo no desempenho global do conformador.

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pelo CNPq (315020/2018-0, 302492/2021-6).

CONFLITOS DE INTERESSE

Os autores declaram que não há nenhum conflito de interesse.

REFERÊNCIAS

1. Einhorn R. (2017) Hearing aid technology for the 21st century: A proposal for universal wireless connectivity and improved sound quality IEEE Pulse 8(2):25-28.
2. Organização Mundial da Saúde. (2018, Feb). Global Estimates on Prevalence of Hearing Loss. Retirado de:

<https://apps.who.int/iris/bitstream/handle/10665/260336/9789241550260-eng.pdf>.

3. Ali H, Lobo A P, Loizou P C. (2013) Design and evaluation of a personal digital assistant-based research platform for cochlear implants. IEEE Trans. Biomed. Eng. 60:3060-3074.
4. Zeng F-G. (2004) Trends in cochlear implants. Trends Amplif. 8(1):1-34.
5. Lenarz M, Sonmez H, Joseph G, Buchner A, Lenarz T. (2012) Cochlear implant performance in geriatric patients. Laryngoscope 122:1361-1365.
6. Buechner A, Dyballa K H, Hehrmann P, Fredelake S, Lenarz T. (2014) Advanced beamformers for cochlear implant users: acute measurement of speech perception in challenging listening conditions. Plos One 9(4):1-9.
7. Ali R, Waterschoot T V, Moonen M. (2019) Integration of a priori and estimated constraints into an MVDR Beamformer for speech enhancement, IEEE/ACM Trans. Audio Speech Lang. Process. 27(12):2288-2300.
8. Marquardt D, Hadad E, Gannot S, Doclo S. (2014) Optimal binaural LCMV beamformer for combined noise reduction and binaural cue preservation. Int. Workshop Acoustic Signal Enhancement, pp. 288-292.
9. Loizou P C. (2013) Speech Enhancement: Theory and Practice. CRC Press.
10. Hadad E, Marquardt D, Doclo S, Gannot S. (2015) Theoretical analysis of binaural transfer function MVDR beamformers with interference cue preservation constraints. IEEE Trans. Audio Speech Lang. Process. 23:2449-2464.
11. Furlani A L, Costa, M H, Paul, S (2021) Métodos baseados em sorteio para seleção de canais de estimulação em implantes cocleares, Simp. Bras. Telecom. Process. Sinais pp. 1-5.
12. Benesty J, Chen J, Huang Y. (2008) Microphone Array Signal Processing. Berlin, Germany: Springer-Verlag.
13. Kayser H, Ewert S D, Anemuler J, Rohdenburg T, Hohmann V, Kollmeier B. (2009) Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. EURASIP J. Adv. Signal Process. 2009:1-10.
14. IEEE. (1969) Recommended practice for speech quality measurements. IEEE Trans. Audio Electroacoustics, 17(3):225-246.
15. Dreschler W A. (2001). ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. Audiol. 40(3):148-157.
16. Slaney, Malcolm. Auditory toolbox. Interval Research Corporation, Tech. Rep, v. 10, n. 1998.
17. Santos J F, Falk T H. (2014). Updating the SRMR-CI metric for improved intelligibility prediction for cochlear implant users. IEEE/ACM Trans. Audio Speech Lang. Process 22(12):2197-2206.
18. Brandstein M, Ward D. (2001). Microphone arrays: signal processing techniques and applications. Springer Science & Business Media.

Autora correspondente:

Autora: Luciana Menezes Xavier de Souza
Instituto: Universidade Federal de Santa Catarina
Cidade: Florianópolis - SC
País: Brasil
E-mail: lu_mxs@yahoo.com.br