

Redução de Ruído em Implantes Cocleares: Preservar a Magnitude Espectral ou a Envoltória Temporal?

Rafael A. Chiea, Luciana M. X. de Souza e Márcio H. Costa

Resumo— Sistemas de redução de ruído são parte importante de implantes cocleares. Este trabalho apresenta uma comparação entre duas técnicas de redução de ruído baseadas nos princípios de estimação da magnitude espectral e da envoltória temporal. Simulações numéricas, utilizando um critério objetivo não-intrusivo especificamente projetado para a estimação da inteligibilidade percentual em usuários de implantes cocleares, indicam que o método baseado na envoltória temporal resulta em desempenho significativamente superior, podendo chegar a uma diferença de até 10% para $SNR \leq 0$ dB em relação ao método baseado na estimação da magnitude.

Palavras-Chave— *Implante coclear, Redução de ruído, Inteligibilidade da fala.*

Abstract— Noise reduction systems are important part of cochlear implants. This work presents a comparison between two noise reduction techniques based on the principles of spectral-magnitude and temporal-envelope estimations. Computational simulations, using a non-intrusive objective-criteria specifically developed for percentual intelligibility estimation in cochlear implant users, indicate that the method based on the temporal envelope presents a significant higher performance, providing up to 10% increase for $SNR \leq 0$ dB as compared to the spectral magnitude estimation method.

Keywords— *Cochlear implant, Noise reduction, Speech intelligibility.*

I. INTRODUÇÃO

Implantes cocleares (IC) são as mais bem-sucedidas interfaces cérebro-máquina desenvolvidas até os dias de hoje [1]. São próteses auditivas que possibilitam a percepção de sons complexos, propiciando elevada inteligibilidade da fala humana por pessoas com perda auditiva severa a profunda ou mesmo com surdez total [2]. Diferentemente dos aparelhos auditivos, que têm como saída um sinal acústico, os implantes cocleares estimulam diretamente o nervo auditivo com corrente elétrica aplicada através de um conjunto de 16 a 24 eletrodos posicionados no interior da cóclea [1] [2].

Apesar das limitações na resolução temporal e espectral dos estímulos transmitidos ao usuário do IC [1], a inteligibilidade da fala se aproxima da encontrada em ouvintes normais. Essa observação pode ser verificada em uma série de testes de audição [3]. Entretanto, na presença de ruído e reverberação, esse desempenho cai significativamente [4] [5], o que motiva a utilização de técnicas de redução de ruído. Dentre as principais técnicas utilizadas em ICs destacam-se as máscaras tempo-frequência, particularmente a máscara binária (MB) [6] e o filtro de Wiener (FW) e suas variantes (paramétrico e restrito) [7]. Apesar de resultarem em ganhos substanciais de inteligibilidade para usuários de IC [8], essas máscaras não foram desenvolvidas especificamente para esse tipo de dispositivo, e não levam em

consideração as particularidades dessa aplicação em seu desenvolvimento.

As estratégias de codificação utilizadas na maior parte dos implantes cocleares atuais são baseadas em *vocoders* [1] [9]. Nesse tipo de estratégia, o sinal captado pelo microfone de aquisição é processado por um estágio de análise, gerando distintas saídas associadas a diferentes bandas de frequência. A envoltória temporal (envelope) do sinal resultante em cada canal é calculada, e utilizada para modular a amplitude de um trem de pulsos, geralmente com taxa constante, que aciona uma fonte de corrente elétrica aplicada ao respectivo eletrodo de estimulação. Como consequência, a informação de fase, também chamada de informação temporal fina (ou ainda, de estrutura fina), não é utilizada, acarretando restrições nas informações disponibilizadas ao usuário.

O estágio de análise pode ser implementado de duas maneiras: 1) no domínio tempo, com o uso de banco de filtros – geralmente do tipo *gammatone* – baseados no sistema auditivo humano [10] [11]; ou 2) no domínio frequência, utilizando a transformada de Fourier de tempo curto (STFT, do inglês *Short Time Fourier Transform*) [7].

Em [12] foi demonstrado que não há significativo impacto de desempenho na inteligibilidade de indivíduos implantados em relação ao uso de arquiteturas baseadas em bancos de filtros ou na STFT. Entretanto, em [13] verificou-se que, em comparação com a STFT, o banco de filtros *gammatone* apresenta um melhor potencial para separação de tons harmônicos complexos. Adicionalmente, bancos de filtros do tipo *gammatone* podem ser implementados com menor atraso que sistemas baseados na STFT, para uma mesma resolução em baixas frequências [13].

Em [14] foi realizada uma comparação de desempenho entre as máscaras tempo-frequência do tipo MB, FW e o estimador de magnitude espectral baseado no mínimo erro quadrático médio (MMSE, do inglês *Minimum Mean Square Error*) [15]. A máscara MMSE baseia-se na minimização do erro quadrático médio entre a magnitude do espectro da fala estimada e desejada. Foi verificado que para usuários de IC o método MMSE resulta em níveis de inteligibilidade maiores que os obtidos pela MB e pelo FW. Esse resultado corrobora a hipótese apresentada de que métodos baseados na aproximação da magnitude da fala (caso do MMSE) são mais apropriados para aplicações de IC, visto que a informação da estrutura fina da fala é limitada ou descartada pela forma de codificação utilizada. Todas as três máscaras foram implementadas no domínio frequência através da STFT.

Em [16], assumindo uma arquitetura de redução de ruído implementada no domínio do tempo (banco de filtros) e utilizando a hipótese de que as distorções na estrutura fina são irrelevantes à inteligibilidade em aplicações de IC, os autores propuseram um método de redução de ruído baseado na minimização do erro quadrático médio entre o quadrado das

Rafael Attili Chiea, Luciana Menezes Xavier de Souza e Márcio Holsbach Costa, Universidade Federal de Santa Catarina, e-mails: rafaelchiea@gmail.com, lu_mxs@yahoo.com.br, costa@eel.ufsc.br. Este trabalho foi parcialmente financiado por CNPq, projetos 315020/2018-0, 142341/2019-3.

envoltórias da fala estimada e desejada. O método é aplicado independentemente a cada canal de saída do banco de filtros e, portanto, pode ser considerado como uma máscara tempo-frequência. Simulações numéricas e um experimento psicoacústico piloto indicaram que este método resulta em maior inteligibilidade em comparação ao filtro de Wiener implementado no domínio do tempo [17].

Neste trabalho, é apresentada uma comparação entre o método proposto em [16] (aqui denominado de C2F) e o método MMSE [15]. O desempenho comparativo entre ambos os métodos é avaliado através de extensivas simulações computacionais utilizando uma medida objetiva de inteligibilidade com alta correlação com experimentos psicoacústicos envolvendo usuários de IC.

Na Seção II deste trabalho são descritos os métodos avaliados (MMSE e C2F), enquanto que na Seção III é detalhado o método de comparação. A Seção IV apresenta os resultados e discussão. Finalmente, a Seção V apresenta as conclusões.

II. REDUÇÃO DE RUIDO

A fala contaminada por ruído aditivo é representada na forma $x(n) = s(n) + v(n)$, em que $s(n)$ é o sinal de fala desejado, $v(n)$ é o ruído e n é o tempo discreto. Assume-se que ambos os sinais $s(n)$ e $v(n)$ são não-observáveis individualmente e descorrelacionados entre si.

A. Estimador de Magnitude MMSE

O MMSE é uma técnica pouco explorada em aplicações de redução de ruído em IC [14]. É um método bayesiano que aproxima a magnitude espectral do sinal estimado à do sinal de fala desejado [7] [15].

Sejam $S(\lambda, k)$, $V(\lambda, k)$ e $X(\lambda, k)$ a representação no domínio frequência de $s(n)$, $v(n)$ e $x(n)$ obtida via STFT, para cada trecho de tempo λ e bin de frequência k , o estimador de magnitude MMSE tem como objetivo a minimização da seguinte função custo

$$J_{\text{MMSE}_k} = E \left\{ [\hat{S}_k(\lambda) - S_k(\lambda)]^2 \right\}, \quad (1)$$

em que $S_k(\lambda)$ é a magnitude do sinal de fala limpo na frequência k e trecho λ ; $\hat{S}_k(\lambda) = c_k(\lambda)X_k(\lambda)$ é a magnitude espectral estimada da fala; $X_k(\lambda)$ é a magnitude do sinal de fala contaminado; $c_k(\lambda)$ é a função de supressão da máscara tempo-frequência MMSE; e $E\{\cdot\}$ é o valor esperado em relação a λ . Considerando que $S(\lambda, k)$, $V(\lambda, k)$ e $X(\lambda, k)$ têm distribuição gaussiana¹ e que os coeficientes da transformada de Fourier são descorrelacionados, a minimização de (1) resulta em [15]:

$$c_k(\lambda) = \frac{\sqrt{\pi} \sqrt{v_k(\lambda)}}{2 \gamma_k(\lambda)} \exp^{-v_k(\lambda)/2} \times \left[[1 + v_k(\lambda)] I_0 \left(\frac{v_k(\lambda)}{2} \right) + v_k I_1 \left(\frac{v_k(\lambda)}{2} \right) \right]. \quad (2)$$

em que $I_0(\cdot)$ e $I_1(\cdot)$ são as funções modificadas de Bessel, de ordem zero e um, respectivamente, enquanto que $v_k(\lambda)$ é definido como:

$$v_k(\lambda) = \frac{\xi_k(\lambda)}{1 + \xi_k(\lambda)} \gamma_k(\lambda); \quad (3)$$

em que $\xi_k(\lambda) = E\{S_k^2(\lambda)\}/E\{V_k^2(\lambda)\}$ é a razão sinal ruído (SNR) *a priori*; e $\gamma_k(\lambda) = E\{X_k^2(\lambda)\}/E\{V_k^2(\lambda)\}$ é a SNR *a posteriori* [15].

B. Estimador de Envelope

O método C2F tem como objetivo minimizar o erro quadrático médio entre o quadrado das envoltórias da fala estimada e desejada [16]. Nele, a fala contaminada $x(n)$ é processada por um banco de filtros com $k = 1, \dots, M$ canais, resultando em $x_k(n) = s_k(n) + v_k(n)$. A envoltória temporal de cada um desses sinais é dada por [18]:

$$x_k^{\text{env}}(n) = |x_{ak}(n)| = \sqrt{x_k^2(n) + (H\{x_k(n)\})^2}, \quad (4)$$

em que $H\{\cdot\}$ é o operador da transformada de Hilbert; $x_{ak}(n)$ é o sinal analítico de $x_k(n)$; e $x_k^{\text{env}}(n)$ é a envoltória temporal instantânea do sinal $x_k(n)$.

O método C2F tem como objetivo minimizar a seguinte função custo:

$$J_{\text{C2F}_k} = E \left\{ |s_{ak}(n)|^2 - |\hat{s}_{ak}(n)|^2 \right\}, \quad (5)$$

em que

$$\hat{s}_{ak}(n) = g_k(n)x_{ak}(n); \quad (6)$$

$g_k(n)$ é o coeficiente do filtro de estimação no canal k ; e $\hat{s}_{ak}(n)$ é a estimativa do sinal analítico de $s_k(n)$.

Igualando-se a zero a derivada de (5) em relação a $g_k(n)$, e assumindo-se que $s_k(n)$ e $v_k(n)$ tenham distribuição gaussiana (devido ao teorema do limite central), chega-se em

$$g_k(n) = \sqrt{\frac{2\zeta_k^2(n) + \zeta_k(n)}{2\zeta_k^2(n) + 4\zeta_k(n) + 2}}, \quad (7)$$

em que $\zeta_k(n) = E\{s_k^2(n)\}/E\{v_k^2(n)\}$ é a SNR no canal k ; e $E\{s_k^2(n)\}$ e $E\{v_k^2(n)\}$ são, respectivamente, as variâncias da fala e do ruído no canal k . O coeficiente $g_k(n)$ é recalculado a cada quadro temporal, funcionando de maneira análoga a uma máscara tempo-frequência.

III. SIMULAÇÕES NUMÉRICAS

Nesta seção é apresentado o protocolo de comparação entre os métodos MMSE e C2F. Para tanto, foram realizadas simulações computacionais utilizando o critério objetivo não-intrusivo *Speech-to-Reverberation Modulation Energy Ratio for Cochlear Implants* (SRMR-CI) [19]. Essa métrica foi especificamente desenvolvida para apresentar alta correlação com a inteligibilidade de usuários de implantes cocleares em experimentos psicoacústicos. O SRMR-CI pode ser mapeado em valores de inteligibilidade da seguinte maneira [19]:

$$I_{\%} = \frac{1}{1 + e^{a_1 \text{SRMR-CI}_{\text{norm}} + a_2}} \times 88,92\%, \quad (8)$$

em que $I_{\%}$ é o percentual de inteligibilidade; $a_1 = -12,1742$ e $a_2 = +7,4535$ são coeficientes de ajuste [20]; e $\text{SRMR-CI}_{\text{norm}}$ é

¹ Suposição aceitável para janelas temporais de duração da ordem de algumas dezenas de milissegundos [9][27].

calculada por:

$$SRMR - CI_{\text{norm}} = \frac{SRMR - CI_t}{SRMR - CI_s}, \quad (9)$$

em que $SRMR_t$ é a métrica do sinal em teste e $SRMR_s$ é o valor da métrica para o sinal de fala limpo.

Foram utilizados como sinais de fala todas as 720 frases do conjunto de listas foneticamente balanceadas do IEEE *corpus* [21]. Como ruído aditivo utilizou-se um ruído de balbuciação, formado artificialmente pela concatenação de quatro sequências do sinal *International Speech Test Signal* (ISTS) [22], que foram aleatoriamente deslocadas no tempo. As SNRs utilizadas foram: -20 dB, -15 dB, -10 dB, -5 dB, 0 dB, 10 dB e 20 dB.

A SNR foi calculada utilizando os sinais $s(n)$ e $v(n)$ separadamente para evitar erros de estimação em $\xi_k(\lambda)$, $\gamma_k(\lambda)$ e $\nu_k(n)$. Portanto, os resultados correspondem à situação de máximo desempenho. Em aplicações reais, as SNRs devem ser estimadas, o que pode ser feito, por exemplo, através do método de decisão direta e suas variações [7] [23].

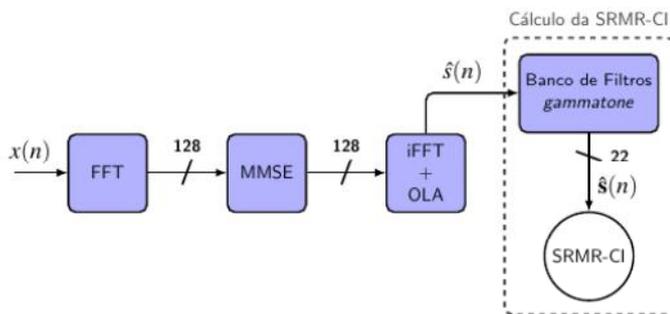


Fig. 1. Diagrama de blocos referente ao método MMSE. O sinal captado pelo microfone é processado por uma FFT de 128 pontos. Para cada *bin* é aplicado o método MMSE para redução do ruído e, em seguida o sinal estimado de fala é reconstruído utilizando iFFT+OLA. Esse sinal é finalmente utilizado para o cálculo da métrica SRMR-CI.

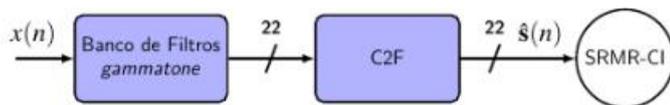


Fig. 2. Diagrama de blocos referente ao método C2F. O sinal captado pelo microfone é processado por um banco de filtros do tipo *gammatone* de 22 canais. Em cada canal é aplicado o método C2F, e são obtidos os 22 envelopes temporais da fala estimada, que, por sua vez, são utilizados para o cálculo da métrica SRMR-CI.

A. Estimador de Magnitude (MMSE)

A Figura 1 apresenta o diagrama de blocos referente ao processamento da fala contaminada, realizado no domínio da frequência, via o estimador de magnitude MMSE.

O sinal $s(n)$ foi transformado para o domínio frequência usando a STFT, com 128 pontos, janela de Hanning de 20 ms e sobreposição de 50%. A função de supressão utilizada foi a equação (2).

As magnitudes espectrais estimadas da fala, a cada quadro temporal, foram então, transformados de volta para o domínio do tempo através da transformada inversa de Fourier (iFFT) e do método *weighted-overlap-and-add* (OLA) [24]. Nesse processo, foi utilizada a fase da fala contaminada em cada canal. Finalmente, a inteligibilidade foi estimada através do critério SRMR-CI. Para isso, o sinal processado $\hat{s}(n)$ foi separado em

bandas de frequência por um banco de filtros do tipo *gammatone* [10] com 22 canais, resultando no conjunto de sinais $\hat{s}(n) = [\hat{s}_1(n), \dots, \hat{s}_{22}(n)]^T$, utilizado para o cálculo da métrica SRMR-CI.

B. Estimador de Envelope (C2F)

A Figura 2 apresenta o diagrama de blocos referente ao método C2F.

O sinal contaminado $x(n)$ foi processado por um banco de filtros do tipo *gammatone* [10] com 22 canais. Os sinais resultantes de cada saída do banco de filtros foram tratados de forma independente. A cada trecho de 1 ms foram calculadas as SNRs *a priori* para cada canal k e os respectivos ganhos $g_k(n)$ ideais. Em seguida, a função de supressão foi aplicada aos respectivos trechos do sinal contaminado $x_k(n)$, resultando no sinal estimado da fala $\hat{s}_k(n)$ no canal k .

A métrica SRMR-CI, originalmente, é calculada a partir de um sinal de áudio monocanal. De forma a compatibilizar as implementações de ambos os métodos, o SRMR-CI foi dividido em dois blocos de processamento. O primeiro consiste na separação do sinal de entrada em bandas de frequência através de um banco de filtros *gammatone* de 22 canais (Figura 1), enquanto o segundo é associado ao cálculo da métrica propriamente dita (Figura 2). Assim, o SRMR-CI é calculado a partir do conjunto de sinais processados em cada banda k , $\hat{s}(n)$. Essa alteração não influencia o valor resultante da métrica.

IV. RESULTADOS E DISCUSSÃO

Os resultados obtidos nas simulações numéricas previamente descritas são apresentados na Figura 3, no formato de diagramas de caixa. Em cada caixa, o círculo central indica a mediana; os retângulos representam o primeiro e terceiro quartis, e as linhas indicam os valores máximo e mínimo. Os resultados de inteligibilidade percentual em função da SNR são apresentados em azul para o método MMSE, em vermelho para o método C2F e em preto para a fala contaminada (não processada).

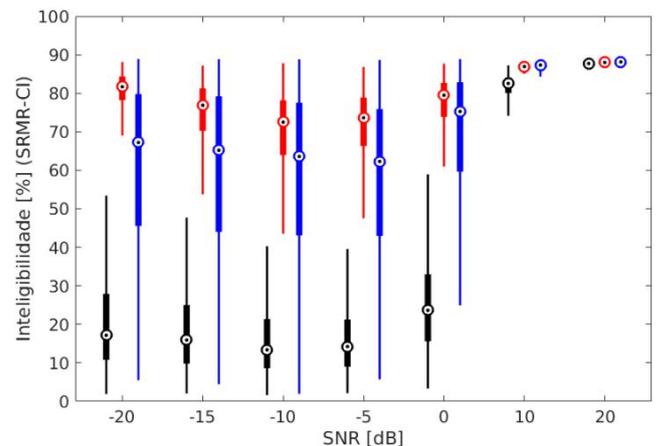


Fig. 3. Diagramas de caixa com os resultados de inteligibilidade percentual das simulações numéricas, em função do valor de SNR de entrada. (a) sinal contaminado original (não processado) (preto); (b) sinal contaminado processado pelo método C2F (vermelho); (c) sinal contaminado processado pelo método MMSE (azul). Em cada caixa, o círculo indica a mediana, enquanto que os retângulos representam o primeiro e terceiro quartis; as linhas indicam os valores máximo e mínimo.

Conforme a Figura 3, a fala contaminada original apresenta uma inteligibilidade percentual acima de 80% para

$SNR \geq 10$ dB. Entretanto, para $SNR \leq 0$ dB a inteligibilidade decresce significativamente, ficando abaixo de 20% para $SNR \leq -5$ dB. Esses resultados simulados corroboram os obtidos em experimentos psicoacústicos reais realizados em [4] e [5].

A Figura 3 ainda mostra que após o processamento pelos métodos de redução de ruído analisados, os níveis de inteligibilidade aumentam consideravelmente para $SNR \leq 0$ dB, apresentando medianas acima de 60%. Esses resultados suportam as observações anteriores [14] [16] de que a utilização dessas técnicas pode ter impacto significativo na compreensão de fala por usuários de IC em ambientes ruidosos.

A Tabela I contém as medianas das distribuições dos resultados obtidos com a utilização do MMSE e do C2F.

TABELA I. MEDIANAS DOS RESULTADOS OBTIDOS COM CADA TÉCNICA DE REDUÇÃO DE RÚIDO EM PERCENTUAL DE INTELIGIBILIDADE.

Método	SNR [dB]						
	-20	-15	-10	-5	0	10	20
C2F	81,7	76,9	72,5	73,6	79,5	86,9	88,0
MMSE	67,2	65,2	63,6	62,3	75,3	87,3	88,1

Para cada cenário de SNR testado, as distribuições dos sinais processados pelos métodos MMSE e C2F foram comparadas estatisticamente entre si usando o teste de Wilcoxon [25] com significância $p < 0,001$. Em quase todos os cenários, o teste indica que as distribuições não podem ser consideradas equivalentes, contudo, para as SNRs de 10 dB e 20 dB o valor obtido é relativamente próximo, o que faz sentido, uma vez que os ganhos de ambas as técnicas para $SNR \gg 0$ dB é aproximadamente 1.

O MMSE resultou em medianas maiores que o C2F apenas para SNRs positivas, nas quais as diferenças, apesar de estatisticamente significativas, foram menores que 0,5%. Além disso, nessa faixa de SNR a fala contaminada (não-processada) apresenta alta inteligibilidade original, de forma que a utilização de um método de redução de ruído não é necessária. Por outro lado, para $SNR \leq 0$ dB, o C2F resulta em níveis de inteligibilidade consistentemente maiores que para o MMSE. Nessa faixa, as medianas têm diferença de aproximadamente 10% em favor do método C2F.

Apesar de haver uma justificativa teórica que sustenta a utilização do MMSE para ICs, baseada na não utilização da informação de fase nas estratégias de estimulação usadas pelo dispositivo [26], os resultados indicam que a abordagem C2F é mais adequada para essa aplicação. A diferença entre os resultados obtidos com cada uma dessas técnicas está associada às diferenças entre as respectivas funções custo (equações (1) e (5)). O MMSE estima a magnitude espectral a cada bin de frequência, ou seja, estabelece parte da estrutura fina da fala como elemento central de interesse. Por sua vez, o C2F estima as envoltórias temporais da fala, sendo essa a informação efetivamente utilizada pelo usuário do IC para compreensão da fala.

Vale notar que a para a avaliação do método MMSE é

necessária a reconstrução do áudio usando a iFFT, na qual é utilizada a informação de fase do sinal ruidoso. Da mesma forma, no processamento com o C2F, a informação de fase contida em cada canal é a mesma do sinal ruidoso, uma vez que os ganhos aplicados nesse processamento não alteram a fase do sinal. De qualquer forma, isso não deve afetar o resultado final, uma vez que a informação de fase dos sinais não é utilizada no cálculo da métrica SRMR-CI.

Um fator que deve ser considerado neste trabalho é a utilização de estimadores ideais de SNR. Dessa forma, os resultados refletem o melhor desempenho possível com cada técnica. O impacto de erros de estimação não é abordado neste trabalho. Adicionalmente, os resultados obtidos pela métrica SRMR-CI têm correlação comprovada com a inteligibilidade de usuários de IC apenas para $SNR \geq -5$ dB [19] [20]. Neste trabalho, a utilização da SRMR-CI foi arbitrariamente extrapolada para valores de SNR inferiores. Tal extrapolação parece válida ao considerar-se que os sinais pós-processados apresentam uma SNR maior que 0 dB, resultando em níveis de inteligibilidade maiores que os obtidos com sinais não-processados com SNR igual à 0dB (Fig. 3). Além disso, métricas objetivas, como a utilizada neste trabalho, fornecem apenas uma noção qualitativa do impacto de técnicas de redução de ruído na inteligibilidade de usuários de IC. Desta forma, trabalhos futuros devem envolver a realização de experimentos psicoacústicos para corroborar as observações obtidas nas simulações realizadas.

V. CONCLUSÃO

Neste trabalho foram comparadas duas técnicas de redução de ruído para aplicações em implantes cocleares. Os métodos avaliados são baseados na estimação da magnitude espectral (MMSE) e da envoltória temporal (C2F) da fala de interesse. O método C2F justifica-se pelo fato de que a informação transmitida ao usuário do implante é restrita ao envelope do sinal de fala, negligenciando a estrutura fina. Simulações numéricas foram utilizadas para avaliar a inteligibilidade percentual da fala contaminada por ruído de balbuciação aditivo utilizando o critério objetivo não intrusivo SRMR-CI. Resultados indicaram que sinais processados pelo método C2F apresentam níveis mais elevados de inteligibilidade em comparação ao método MMSE. Essa diferença pode chegar a cerca de 10% para $SNR \leq 0$ dB. Esses resultados preliminares motivam a busca por novos métodos de redução de ruído projetados especificamente para aplicações em implantes cocleares. Futuramente, esses achados devem ser corroborados por experimentos psicoacústicos com usuários de implantes cocleares.

AGRADECIMENTOS

Os autores gostariam de agradecer à CAPES, FINEP e CNPq pelo suporte financeiro, e à Wavetech-ST.

REFERÊNCIAS

- [1] J. Wouters, H. J. McDermott, T. Francart, "Sound coding in cochlear implants", *IEEE Signal Processing Magazine*, v. 32, n. 2, pp. 67-80, 2015.
- [2] J. Wouters, S. Doclo, R. Koning, T. Francart, "Sound processing for better coding of monaural and binaural cues in auditory prostheses", *Proceedings of the IEEE*, v. 101, n. 9, pp. 1986-1997, 2013.
- [3] B. S. Wilson, M. F. Dorman, "The surprising performance of present-day cochlear implants", *IEEE Transactions on Biomedical Engineering*, v. 54, n. 6, pp. 969-972, 2007.
- [4] A. Hast, L. Schlucker, F. Digeser, T. Liebscher, Ulrich, Hoppe. "Speech perception of elderly cochlear implant users under different noise conditions", *Otology and Neurology*, v. 36, n. 10, pp. 1638-1643, 2015.

- [5] F. Bergeron, M. Hotton, "Perception in noise with the Digisonic SP cochlear implant: Clinical trial of Saphyr processor's upgraded signal processing", *European Annals of Otorhinolaryngology, Head and Neck Diseases*, pp. 3-5, 2016.
- [6] D. Wang, G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*, Wiley/IEEE Press, 2006.
- [7] P. C. Loizou, *Speech Enhancement - Theory and Practice*, 2nd ed., CRC Press, 2013.
- [8] R. Koning, N. Madhu, J. Wouters, "Ideal time-frequency masking algorithms lead to different speech intelligibility and quality in normal-hearing and cochlear implant listeners", *IEEE Transactions on Biomedical Engineering*, v. 62, n. 1, pp. 331-341, 2015.
- [9] P. C. Loizou, "Speech processing in vocoder-centric cochlear implants", *Cochlear and Brainstem Implants*, v. 64, pp. 109-143, 2006.
- [10] P. I. M. Johannesma, "The pre-response stimulus ensemble of neurons in the cochlear nucleus", *Symposium on Hearing Theory*, pp. 58-69, 1972.
- [11] S. Cosentino, T. H. Falk, D. McAlphine, "Cochlear implant filterbank design and optimization: A simulation study", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 22, n. 2, pp. 347-353, 2014.
- [12] M. A. M. Jong, J. J. Briaire, J. H. M. Frinjns, "Take-home trial comparing fast Fourier transformation-based and filter bank-based cochlear implant speech coding strategies", *BioMed Research International*, pp. 1-7, 2017.
- [13] S. Tabibi, A. Kegel, W. K. Lai, N. Dillier, "Investigating the use of a gammatone filterbank for a cochlear implant coding strategy", *Journal of Neuroscience Methods*, v. 277, pp. 63-74, 2016.
- [14] G. L. Mourão, M. M. C. Pinheiro, M. H. Costa, S. Paul, B. H. P. Murta, "Análise do impacto de métodos de redução de ruído na inteligibilidade da fala em implantes cocleares", *Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, pp. 65-69, 2018.
- [15] Y. Ephraim, D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator", *IEEE Transactions on Acoustics, speech, and Signal Processing*, v. ASSP-32, n. 6, pp. 1109-1121, 1984.
- [16] R. A. Chiea, . H. Costa, M. M. C. Pinheiro, J. A. Cordioli, "Um estimador de envoltória temporal para redução de ruído em implantes cocleares", *Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, pp. 1-5, 2019.
- [17] J. Chen, J. Benesty, Y. Huang, S. Doclo, "New insights into the noise reduction Wiener filter", *IEEE Transactions on Audio, Speech, and Language Processing*, v. 14, n. 4, pp. 1218-1234, 2006.
- [18] J. Justice, "Analytic signal processing in music computation", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 27, n. 6, pp. 670-684, 1979.
- [19] J. F. Santos, T. H. Falk, "Updating the SRMR-CI metric for improved intelligibility prediction for cochlear implant users", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 22, n. 12, pp. 2197-2206, 2014.
- [20] J. F. Santos, S. Cosentino, O. Hazrati, P. C. Loizou, T. H. Falk, "A non-intrusive objective speech intelligibility metric tailored for cochlear implant users in complex listening environments", *Universit  du Qu bec, M.Sc. Thesis*, 2013.
- [21] IEEE recommended practice for speech quality measurements, *IEEE Transactions on Audio and Electroacoustics*, v.17 no. 3, pp. 225-246, 1969.
- [22] I. Holube, S. Fredelake, M. Vlaming, B. Kollmeier, "Development and analysis of an international speech test signal (ISTS)," *International Journal of Audiology*, v. 49, n. 12, pp. 891-903, 2010.
- [23] Y. Hu, P. C. Loizou, N. Li, K. Kasturi, "Use of a sigmoidal-shaped function for noise attenuation in cochlear implants", *The Journal of the Acoustical Society of America*, v. 122, n. 4, pp. 128-134, 2007.
- [24] R. E. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 28, n. 1, pp. 99-102, 1980.
- [25] F. Wilcoxon, "Individual comparisons of grouped data by ranking methods", *Journal of Economic Entomology*, vol. 39, no. 6, p. 269, 1946.
- [26] G. L. Mourão, M. H. Costa, S. Paul, B. "Speech intelligibility for cochlear implant users with the MMSE noise-reduction time-frequency mask", *Biomedical Signal Processing and Control*, v.60, p.101982, 2020.
- [27] J. Jensen, I. Batina, R.C. Hendriks, R. Heusdens "A study of the distribution of time-domain speech samples and discrete Fourier coefficients" *Proceedings of SPS-DARTS*, p. 155-158, 2005.