EasyChair Preprint

№ 9063

Envelope Based Time-Frequency Mask for Noise Reduction in Cochlear Implants

Paulo Gubert, Márcio Holsbach Costa and Bruno Catarino Bispo

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

October 24, 2022

Máscara tempo-frequência baseada em envoltória para redução de ruído em implantes cocleares

P. H. Gubert, M. H. Costa e B. C. Bispo

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

Resumo — **Este trabalho apresenta uma máscara tempo-frequência para a redução de ruído em implantes cocleares. A função de atenuação proposta é obtida a partir da minimização do erro quadrático médio entre as envoltórias da fala e de sua estimativa, levando em consideração restrições de complexidade computacional. O desempenho do método proposto foi avaliado através de dois critérios objetivos de inteligibilidade, assumindo contaminação por ruído de balbuciação e diferentes razões sinal-ruído. Os resultados indicam que a máscara desenvolvida proporciona maior inteligibilidade ao usuário de implante coclear, em comparação às máscaras de Wiener e raiz de Wiener, ambas amplamente utilizadas na literatura.**

Palavras-chaves— **redução de ruído, máscara tempo-frequência, implante coclear, processamento da fala.**

I. Introdução

De acordo com a Organização Mundial da Saúde, em 2021, 1,5 bilhão de pessoas possuíam algum grau de perda auditiva, sendo 450 milhões delas com perda moderada ou severa no melhor ouvido [1]. Segundo o Instituto Brasileiro de Geografia e Estatística, estima-se que 2,3 milhões de brasileiros, ou 1,1% da população nacional, apresentam alguma limitação auditiva [2].

A deficiência auditiva pode gerar problemas pessoais (cognitivos, saúde mental), sociais (comunicação) e econômicos (perda de produtividade) [1]. A sua compensação depende do tipo e do grau da perda auditiva. No caso de perda severa ou profunda, aparelhos auditivos são incapazes de restaurar a capacidade de ouvir e o uso de implante coclear (IC) é recomendado.

Os ICs são dispositivos eletrônicos que estimulam eletricamente as fibras nervosas da cóclea através de eletrodos, proporcionando sensações capazes de serem interpretadas pelo cérebro como audição [3]. Apesar de apresentarem limitações na reprodução dos sons transmitidos ao usuário, a inteligibilidade da fala é de aproximadamente 80% em condições de silêncio, permitindo a conversação [4]. No entanto, seu desempenho diminui para cerca de 20% para uma razão sinal-ruído (SNR, do inglês *signal-to-noise ratio*) de 5 dB [4]. Desta forma, técnicas de redução de ruído são necessárias para permitir a comunicação em muitas situações cotidianas [5].

As máscaras tempo-frequência constituem uma importante classe de estratégias de redução de ruído. O procedimento de filtragem inicia pela decomposição da fala contaminada em sub-bandas. Em sequência, cada trecho é submetido a um fator de atenuação. O conjunto de fatores de atenuação, calculados em função da SNR, é denominado de curva de atenuação. No caso específico de IC, a envoltória do sinal processado em cada sub-banda é estimada, sendo então transformada em pulsos elétricos e aplicada à cóclea através de uma estratégia de estimulação [6]. Essas máscaras visam a atenuar os canais dominados pelo ruído, acarretando em aumento da SNR global quando fala e ruído ocupam bandas distintas em diferentes instantes de tempo.

As máscaras tempo-frequência mais populares são a máscara binária [7] e a máscara de Wiener e suas variantes [8]. Elas atuam sobre a estrutura temporal fina da fala contaminada, proporcionando redução substancial de ruído. No entanto, a percepção da fala por usuários de IC depende principalmente das informações da envoltória [9], motivo pelo qual os ICs ignoram a estrutura fina [10]. No conhecimento dos autores, até o momento, apenas um método de redução de ruído baseado na envoltória foi apresentado na literatura [5]. Contudo, possui como limitação o fato de assumir estimativas perfeitas da envoltória do sinal processado, as quais são inevitavelmente obtidas por meio de processamento não-causal.

Esse trabalho apresenta uma proposta de máscara tempo-frequência baseada na envoltória, para redução de ruído em ICs. Como contribuição adicional, considera-se a possibilidade de estimativas imperfeitas da envoltória, característica essa decorrente das limitações computacionais associadas aos ICs comerciais. Este trabalho está organizado da seguinte maneira: na Seção II, as principais máscaras encontradas na literatura são apresentadas; na Seção III, a nova máscara baseada na envoltória é proposta; a Seção IV descreve os experimentos realizados; a Seção V apresenta e discute os resultados obtidos; e, finalizando, a Seção VI conclui o artigo.

II. Máscaras tempo-frequência

Sejam a fala e o ruído denotados por $x(n)$ e $v(n)$, respectivamente. A fala contaminada é definida por $y(n) = x(n)+v(n)$. Assume-se que $x(n)$ e $v(n)$ são não-observáveis e descorrelacionados entre si.

Os sistemas monocanais de redução de ruído utilizam um banco de filtros ou a transformada de Fourier de curta duração para obter uma representação tempo-frequência do sinal de entrada em K sub-bandas.

Utilizando um banco de filtros com resposta ao impulso finita, a fala contaminada na k-ésima sub-banda é definida por

$$y_k(n) = \sum_{l=0}^{G-1} g_{k,l}\, y(n-l)\,, \tag{1}$$

em que $k = 1,2,...,K$ e $\mathbf{g}_k = [g_{k,0}\ g_{k,1}\ \ldots\ g_{k,G-1}]^{\mathrm{T}}$ é a resposta ao impulso do k-ésimo filtro com comprimento G.

A técnica de mascaramento tempo-frequência para supressão de ruído consiste em multiplicar $y_k(n)$ por um fator de atenuação $w_k(n)$, resultando numa estimativa de $x_k(n)$ dada por

$$\hat{x}_k(n) = w_k(n) y_k(n)\,. \tag{2}$$

No caso específico de ICs, estimativas das envoltórias de $\hat{x}_k(n)$ são utilizadas em conjunto com uma estratégia de estimulação multicanal, como a CIS (do inglês *Continuos Interleaved Sampling*).

As máscaras tempo-frequência podem ser definidas utilizando critérios objetivos ou heurísticos. Em geral, $0 \le w_k(n) \le 1$ e $w_k(n)$ é uma função da SNR associada à k-ésima sub-banda, a qual é definida como

$$\xi_k(n) = \frac{\mathbb{E}\{x_k^2(n)\}}{\mathbb{E}\{v_k^2(n)\}}\,, \tag{3}$$

em que $\mathbb{E}\{\cdot\}$ é o operador valor esperado.

A máscara de Wiener é definida como [11]

$$w_k(n) = \frac{\xi_k(n)}{\xi_k(n)+1}\,, \tag{4}$$

sendo o filtro ótimo que minimiza o erro quadrático médio entre a fala e sua estimativa. Ou seja, é o coeficiente que minimiza a função custo dada por

$$J_k(n) = \mathbb{E}\{[x_k(n) - w_k(n) y_k(n)]^2\}\,. \tag{5}$$

Outra máscara comumente utilizada na literatura e que apresenta bons resultados em implantes cocleares é a raiz de Wiener [11,12], a qual é definida como

$$rw_k(n) = \sqrt{\frac{\xi_k(n)}{\xi_k(n)+1}}\,. \tag{6}$$

A raiz de Wiener é a máscara que minimiza o erro entre as potências da fala e de sua estimativa.

III. Proposta de máscara tempo-frequência baseada na envoltória dos sinais

Essa seção apresenta a proposta de uma máscara tempo-frequência baseada na envoltória, para redução de ruído em ICs.

A. Estimação da Envoltória

A envoltória de um sinal é comumente estimada de duas maneiras: por retificação de onda completa seguida de filtragem passa-baixa, ou usando a transformada discreta de Hilbert (DHT). A DHT de um sinal é definida como o valor absoluto do sinal analítico. Portanto, a envoltória da fala contaminada na k-ésima sub-banda e no tempo discreto n é definida como

$$p_{yk}(n) = \left|y_{ak}(n)\right| = \left| y_k(n) + \mathrm{j}\,\tilde{y}_k(n) \right| \tag{7}$$

onde $y_{ak}(n)$ é o sinal analítico e $\tilde{y}_k(n)$ é a DHT de $y_k(n)$. Representação semelhante para a envoltória da fala pode ser obtida substituindo-se 'y' por 'x'.

No entanto, a resposta ao impulso associada à DHT possui duração infinita e é não-causal. Portanto, em implementações práticas, é necessário o uso de um atraso (Δ amostras) e a realização do truncamento da sua duração (M amostras). O atraso utilizado não pode ultrapassar o limite de 9 a 12,5 ms, de forma a evitar o prejuízos na leitura labial [13]. Assim, uma estimativa causal de $\tilde{y}_k(n)$ pode ser obtida como

$$\hat{\tilde{y}}_k(n-\Delta) = \mathbf{q}_\Delta^{\mathrm{T}} \mathbf{H} \mathbf{y}_k(n) \tag{8}$$

em que $\mathbf{q}_\Delta = [\,0\ \ldots\ 0\ 1\ 0\ \ldots\ 0\,]^{\mathrm{T}}$ (o único valor diferente de zero está na Δ-ésima entrada) e $\mathbf{y}_k(n) = [\,y_k(n)\ y_k(n-1)\ \ldots\ y_k(n-\Delta-M+1)\,]^{\mathrm{T}}$ são vetores coluna com dimensão $(\Delta+M)\times 1$; $M = 2\Delta+1$; e

$$\mathbf{H} = \frac{2}{\pi}\begin{bmatrix} 0 & 1/(1) & 0 & 1/(3) & \cdots \\ 1/(-1) & 0 & 1/(1) & 0 & \ddots \\ 0 & 1/(-1) & 0 & 1/(1) & \ddots \\ 1/(-3) & 0 & 1/(-1) & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{9}$$

é a matriz da DHT com dimensão $(\Delta+M)\times(\Delta+M)$.

Assim, uma estimativa causal da envoltória da fala contaminada é dada por

$$\hat{p}_{yk}(n-\Delta) = \left|\hat{y}_{ak}(n-\Delta)\right| = \left| y_k(n-\Delta) + \mathrm{j}\,\hat{\tilde{y}}_k(n-\Delta) \right|, \tag{10}$$

onde a fala contaminada, atrasada de Δ amostras, na k-ésima sub-banda, pode ser escrita na forma matricial como

$$y_k(n-\Delta) = \mathbf{q}_\Delta^{\mathrm{T}} \mathbf{y}_k(n) = \mathbf{q}_\Delta^{\mathrm{T}} \mathbf{G}_k \mathbf{y}(n) \tag{11}$$

onde $\mathbf{G}_k$ é uma matriz de dimensão $(\Delta+M)\times(\Delta+M+G+2)$ contendo os coeficentes do k-ésimo filtro do banco de filtros.

Portanto, uma aproximação para a envoltória do sinal $y_k(n-\Delta)$, sujeita a um atraso de Δ amostras, é dada por

$$\hat{p}_{yk}(n-\Delta)=\sqrt{[\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{y}(n)]^2+[\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{y}(n)]^2}\ . \quad (12)$$

A Fig. 1 apresenta as envoltórias estimadas de um sinal artificial, utilizando a equação (12), para dois valores distintos de Δ. Nota-se que a estimativa da envoltória se torna mais acurada com o aumento de Δ.

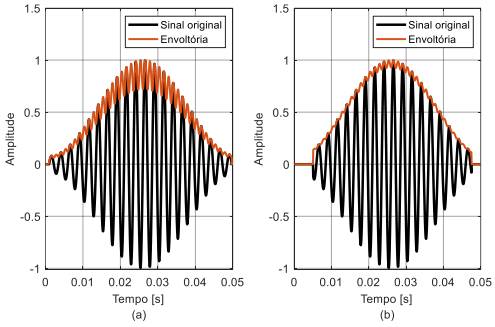

Fig. 1: Estimativas da envoltória (laranja) de um sinal artificial (preto) obtidas para $G = 1200$ e: (a) $\Delta = 5$ e $M = 11$; (b) $\Delta = 40$ e $M = 81$.

B. Proposta da nova máscara

Uma vez que a percepção de fala por usuários de IC depende principalmente das informações da envoltória, neste trabalho propõe-se uma máscara tempo-frequência $c_k(n)$ que minimize o erro quadrático médio entre a envoltória da fala e sua estimativa (a partir do sinal contaminado), de forma a minimizar a seguinte função custo

$$J_k(n)=\mathbb{E}\left\{\left[p_{xk}(n-\Delta)-c_k(n-\Delta)p_{yk}(n-\Delta)\right]^2\right\}. \quad (13)$$

A equação (13) pode ser expandida como

$$J_k(n)=\mathbb{E}\left\{p_{xk}^2(n-\Delta)+c_k^2(n-\Delta)p_{yk}^2(n-\Delta) - 2c_k(n-\Delta)p_{xk}(n-\Delta)p_{yk}(n-\Delta)\right\}. \quad (14)$$

Assumindo-se que $c_k(n-\Delta)$ é descorrelacionado de $p_{xk}(n-\Delta)$ e $p_{yk}(n-\Delta)$, a função custo se torna

$$J_k(n)=\mathbb{E}\{p_{xk}^2(n-\Delta)\}+c_k^2(n-\Delta)\mathbb{E}\{p_{yk}^2(n-\Delta)\} -2c_k(n-\Delta)\mathbb{E}\{p_{xk}(n-\Delta)p_{yk}(n-\Delta)\}. \quad (15)$$

A solução ótima de (15) é obtida derivando-a em função de $c_k(n-\Delta)$ e igualando-a a zero, resultando em

$$c_k(n-\Delta)=\frac{\mathbb{E}\{p_{xk}(n-\Delta)p_{yk}(n-\Delta)\}}{\mathbb{E}\{p_{yk}^2(n-\Delta)\}} \quad (16)$$

Substituindo $p_{yk}(n-\Delta)$ e $p_{xk}(n-\Delta)$ na equação (16) pelas suas respectivas estimativas fornecidas em (12), resulta em

$$\begin{aligned}
&\mathbb{E}\{p_{xk}(n-\Delta)p_{yk}(n-\Delta)\}\\
&=\mathbb{E}\Big\{\Big[\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{x}(n)\mathbf{v}^{\mathrm{T}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{v}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{v}(n)\mathbf{v}^{\mathrm{T}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{x}(n)\mathbf{v}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{v}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{v}(n)\mathbf{v}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{x}(n)\mathbf{v}^{\mathrm{T}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{v}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{v}(n)\mathbf{v}^{\mathrm{T}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{x}(n)\mathbf{v}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{v}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{v}(n)\mathbf{v}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\Big]^{1/2}\Big\}
\end{aligned} \quad (17)$$

A solução da equação (17) não é trivial e depende de aproximações. Assumindo que as variâncias das correlações da fala e do ruído são muito menores que as respectivas médias, de forma que $\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\cong\mathbb{E}\{\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\}=\mathbf{R}_{\mathbf{xx}}$, $\mathbf{v}(n)\mathbf{v}^{\mathrm{T}}(n)\cong\mathbb{E}\{\mathbf{v}(n)\mathbf{v}^{\mathrm{T}}(n)\}=\mathbf{R}_{\mathbf{vv}}$ e $\mathbf{x}(n)\mathbf{v}^{\mathrm{T}}(n)\cong\mathbb{E}\{\mathbf{x}(n)\mathbf{v}^{\mathrm{T}}(n)\}$, e considerando que $x(n)$ e $v(n)$ não são correlacionados e possuem média nula, de forma que $\mathbb{E}\{\mathbf{x}(n)\mathbf{v}^{\mathrm{T}}(n)\}=\mathbf{0}$, então (17) resulta em

$$\begin{aligned}
&\mathbb{E}\{p_{xk}(n-\Delta)p_{yk}(n-\Delta)\}\\
&\cong\mathbb{E}\Big\{\Big[\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\Big]^{1/2}\Big\}
\end{aligned} \quad (18)$$

e, portanto

$$\begin{aligned}
&\mathbb{E}\{p_{xk}(n-\Delta)p_{yk}(n-\Delta)\}\\
&=\Big[\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\\
&\quad+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta
\end{aligned}$$

$$+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k\mathbf{q}_\Delta$$
$$+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k\mathbf{q}_\Delta$$
$$+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta \quad .(19)$$
$$+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\Big]^{1/2}$$

Utilizando (12), o denominador da equação (16) se torna

$$\mathbb{E}\{p_{yk}^2(n-\Delta)\} \cong \mathbb{E}\Big\{\Big[\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{y}(n)\mathbf{y}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta + \mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{y}(n)\mathbf{y}^{\mathrm{T}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\Big]^2\Big\} \quad .(20)$$

Substituindo $\mathbf{y}(n) = \mathbf{x}(n)+\mathbf{v}(n)$ em (20) chega-se em

$$\mathbb{E}\{p_{yk}^2(n-\Delta)\} \cong \mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta + \mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta \; . \quad (21)$$
$$\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta + \mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta$$

Substituindo-se (19) e (21) em (16) obtém-se

$$c_k(n-\Delta) =$$
$$\Big[\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta$$
$$+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta$$
$$+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta$$
$$+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta$$
$$+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta \quad . (22)$$
$$+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta$$
$$+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta$$
$$+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\Big]^{1/2}$$
$$\cdot\Big[\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta + \mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta$$
$$+\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta + \mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{vv}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta\Big]^{-1}$$

Rearranjando-se (22) pode-se escrever

$$c_k(n-\Delta) = \sqrt{\frac{\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta + \mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{xx}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta}{\mathbf{q}_\Delta^{\mathrm{T}}\mathbf{G}_k\mathbf{R}_{\mathbf{yy}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{q}_\Delta + \mathbf{q}_\Delta^{\mathrm{T}}\mathbf{H}\mathbf{G}_k\mathbf{R}_{\mathbf{yy}}(n)\mathbf{G}_k^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{q}_\Delta}} \; . \quad (23)$$

em que $\mathbf{R}_{\mathbf{yy}} = \mathbb{E}\{\mathbf{y}(n)\mathbf{y}^{\mathrm{T}}(n)\}$.

Diferentemente das máscaras tempo-frequência baseadas na estrutura fina, como Wiener e raiz de Wiener, a função de atenuação da máscara proposta, definida em (23), não depende apenas da SNR, mas também da autocorrelação da fala e do ruído.

IV. Simulações computacionais

Nesta seção são apresentadas simulações computacionais com a finalidade de avaliar o desempenho da máscara proposta. No primeiro experimento investiga-se a validade das aproximações utilizadas e seu impacto na minimização da função custo definida em (13). No segundo, é realizada uma comparação entre a máscara proposta e as máscaras de Wiener e raiz de Wiener em termos de inteligibilidade, para diferentes SNR, utilizando dois critérios objetivos.

O banco de filtros utilizado é o mesmo empregado na métrica de inteligibilidade SRMR-CI, o qual contém $K = 22$ sub-bandas na escala ERB (do inglês *Equivalent Rectangular Bandwidth*) e filtros com com comprimento $G = 1200$ [14]. As envoltórias foram calculadas utilizando a DHT com $\Delta = 5$ e $M = 11$, de forma a emular a limitação de disponibilidade computacional intrínseca à aplicações embarcadas, como as em ICs.

A. Banco de dados

Foram utilizados 720 sinais de fala do banco de dados *IEEE corpus* [15]. O ruído de balbuciação presente em uma cafeteria com múltiplos locutores, disponível em [14], foi utlizado para contaminar os sinais de fala para SNR ∈ { −15, −12, −9, −6, −3, 0 } dB. Foi realizada contaminação aditiva artificial e os sinais foram reamostrados para a frequência de amostragem de 16 kHz.

B. Estimação dos Momentos de Segunda Ordem

As matrizes de correlação foram estimadas utilizando as seguintes relações

$$\hat{\mathbf{R}}_{\mathbf{yy}}(n) = \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{y}(n-i)\mathbf{y}^{\mathrm{T}}(n-i), \quad (24)$$

$$\hat{\mathbf{R}}_{\mathbf{vv}}(n) = \frac{1}{N-1}\sum_{i=0}^{N-1}\mathbf{v}(n-i)\mathbf{v}^{\mathrm{T}}(n-i) \quad (25)$$

e

$$\hat{\mathbf{R}}_{\mathbf{xx}}(n) = \hat{\mathbf{R}}_{\mathbf{yy}}(n) - \hat{\mathbf{R}}_{\mathbf{vv}}(n), \quad (26)$$

em que $\mathbf{y}(n) = [\, y(n)\; y(n-1)\; \ldots\; y(n-\Delta-M+1)\,]^{\mathrm{T}}$, $\mathbf{v}(n) = [\, v(n)\; v(n-1)\; \ldots\; v(n-\Delta-M+1)\,]^{\mathrm{T}}$ são respectivamente, os vetores de amostras da fala contaminada e ruído no instante de tempo n, e $N = 5$ é o número de vetores usados na estimação.

Em aplicações práticas, a identificação de períodos de somente ruído é realizada através de um detector de fala (VAD). Nas simulações realizadas, o ruído individualizado foi utilizado para a obtenção de $\mathbf{v}(n)$ com o objetivo de evitar erros do VAD, obtendo assim o limite máximo de desempenho.

C. Avaliação das Aproximações Utilizadas

A verificação da validade das aproximações utilizadas para a obtenção da equação (23) foi realizada através do erro

quadrático médio entre a envoltória da fala e a envoltória da fala contaminada processada, o qual é dado por

$$\varepsilon_k = \frac{1}{L}\sum_{i=1}^{L}\left[p_{xk}(i) - m_k(i)p_{yk}(i)\right]^2 . \qquad (27)$$

em que $m_k(n) \in \{ w_k(n), rw_k(n), c_k(n) \}$ representa a máscara avaliada (Wiener, raiz de Wiener e proposta) e L é o número total de amostras do sinal.

As envoltórias foram calculadas utilizando a transformada discreta de Hilbert causal, com $\Delta = 5$ e $M = 11$.

D. Métricas de inteligilidade

Para a avaliação da inteligibilidade, utilizou-se as métricas objetivas SRMR-CI e a STOI. Ambas possuem alta correlação com testes psicoacústicos com implantados [14].

A SRMR-CI foi modificada, retirando-se o banco de filtros e acoplando-se a estrutura de filtragem descrita na Seção II, na qual são aplicadas as diferentes máscaras. Seu resultado é convertido em porcentagem de inteligibilidade através da seguinte relação [14]

$$I = \frac{100}{1 + e^{a_1 \times \frac{SRMR-CI_p}{SRMR-CI_c} + a_2}} \ [\%] , \qquad (28)$$

em que $a_1 = -12{,}17$ e $a_2 = 7{,}45$ são parâmetros de ajuste; SRMR-CI_p é o resultado da métrica aplicada à fala contaminado processada, e SRMR-CI_c é o resultado da métrica aplicada à fala original.

A métrica STOI foi aplicada ao sinal reconstruído utilizando o vocoder tonal descrito em [16]. Sua pontuação varia entre 0 e 1, onde que valores mais elevados indicam maior inteligibilidade.

A diferença estatística entre os resultados obtidos com as diferentes máscaras foi avaliada através do teste de Friedman em conjunto com o teste de Dunn e ajuste de Bonferroni.

O nível de significância (α) dos testes é de 5%. A hipótese nula, correspondendo à igualdade das distribuições, é rejeitada quando a probabilidade de significância (valor-ρ) for menor ou igual ao nível de significância, ou seja, valor-$\rho \leq \alpha$.

V. Resultados e Discussão

Nesta seção são apresentados os resultados das simulações computacionais realizadas para avaliar o desempenho da máscara tempo-frequência proposta.

A. Erro Quadrático Médio

A Tabela 1 apresenta o erro quadrático médio entre as envoltórias da fala e de sua estimativa, para cada canal e SNR = −15 dB (limite inferior para SNR encontradas nas situações mais comuns de comunicação [17]). Os menores valores são apresentados em negrito. A máscara proposta apresenta o menor erro em todos os canais, indicando que as aproximações realizadas mantêm as características originais da função custo. Nota-se também que a máscara raiz de Wiener obtém resultados semelhantes aos da máscara proposta, a partir do décimo canal. Esse fato chama atenção uma vez que a máscara raiz de Wiener não foi proposta com esse objetivo. Entretanto, esse achado sustenta relatos apresentados em diversos trabalhos encontrados na literatura que, de forma empírica, indicam aumento considerável de inteligibilidade em ICs ao usar a máscara raiz de Wiener [12]. A máscara de Wiener apresenta o pior desempenho.

Tabela 1: Erro quadrático médio entre a envoltória da fala original e sua estimação a partir da fala contaminada para SNR = −15 dB. Em negrito é apresentado o menor valor de cada linha.

Canal	Wiener	Raiz de Wiener	Proposta
1	$4{,}5\cdot10^{-3}$	$2{,}9\cdot10^{-3}$	$\mathbf{2{,}6\cdot10^{-3}}$
2	$7{,}1\cdot10^{-3}$	$3{,}0\cdot10^{-3}$	$\mathbf{2{,}5\cdot10^{-3}}$
3	$1{,}2\cdot10^{-2}$	$6{,}6\cdot10^{-3}$	$\mathbf{5{,}5\cdot10^{-3}}$
4	$1{,}5\cdot10^{-2}$	$9{,}3\cdot10^{-3}$	$\mathbf{7{,}6\cdot10^{-3}}$
5	$1{,}4\cdot10^{-2}$	$7{,}9\cdot10^{-3}$	$\mathbf{6{,}0\cdot10^{-3}}$
6	$9{,}0\cdot10^{-3}$	$4{,}3\cdot10^{-3}$	$\mathbf{3{,}1\cdot10^{-3}}$
7	$7{,}2\cdot10^{-3}$	$3{,}3\cdot10^{-3}$	$\mathbf{2{,}6\cdot10^{-3}}$
8	$5{,}9\cdot10^{-3}$	$2{,}3\cdot10^{-3}$	$\mathbf{1{,}9\cdot10^{-3}}$
9	$5{,}9\cdot10^{-3}$	$2{,}8\cdot10^{-3}$	$\mathbf{2{,}6\cdot10^{-3}}$
10	$4{,}4\cdot10^{-3}$	$\mathbf{2{,}3\cdot10^{-3}}$	$\mathbf{2{,}3\cdot10^{-3}}$
11	$4{,}8\cdot10^{-3}$	$\mathbf{2{,}9\cdot10^{-3}}$	$\mathbf{2{,}9\cdot10^{-3}}$
12	$4{,}3\cdot10^{-3}$	$\mathbf{2{,}6\cdot10^{-3}}$	$\mathbf{2{,}6\cdot10^{-3}}$
13	$3{,}4\cdot10^{-3}$	$\mathbf{1{,}9\cdot10^{-3}}$	$\mathbf{1{,}9\cdot10^{-3}}$
14	$2{,}3\cdot10^{-3}$	$\mathbf{1{,}2\cdot10^{-3}}$	$\mathbf{1{,}2\cdot10^{-3}}$
15	$2{,}3\cdot10^{-3}$	$1{,}3\cdot10^{-3}$	$\mathbf{1{,}2\cdot10^{-3}}$
16	$1{,}9\cdot10^{-3}$	$\mathbf{1{,}1\cdot10^{-3}}$	$\mathbf{1{,}1\cdot10^{-3}}$
17	$1{,}6\cdot10^{-3}$	$\mathbf{1{,}0\cdot10^{-3}}$	$\mathbf{1{,}0\cdot10^{-3}}$
18	$1{,}5\cdot10^{-3}$	$\mathbf{9{,}6\cdot10^{-4}}$	$\mathbf{9{,}6\cdot10^{-4}}$
19	$1{,}0\cdot10^{-3}$	$\mathbf{6{,}8\cdot10^{-4}}$	$\mathbf{6{,}8\cdot10^{-4}}$
20	$1{,}0\cdot10^{-3}$	$\mathbf{7{,}2\cdot10^{-4}}$	$\mathbf{7{,}2\cdot10^{-4}}$
21	$6{,}0\cdot10^{-4}$	$3{,}2\cdot10^{-4}$	$\mathbf{3{,}1\cdot10^{-4}}$
22	$2{,}0\cdot10^{-4}$	$8{,}5\cdot10^{-5}$	$\mathbf{7{,}9\cdot10^{-5}}$

B. Inteligibilidade

A Fig. 2 apresenta resultados de inteligibilidade (na forma de diagramas de caixas) obtidos pela métrica SRMR-CI para os sinais processados pelas diferentes máscaras. Observa-se que a máscara proposta apresenta o melhor desempenho em todas as SNRs, apresentando aproximadamente 2% de aumento em relação à máscara raiz de Wiener. De forma oposta, a máscara de Wiener obtém a menor inteligibilidade. Os testes estatísticos aplicados indicam que todas distribuições ilustradas são estatisticamente diferentes.

A Fig. 3 ilustra os resultados de inteligibilidade obtidos pela métrica STOI. Observa-se novamente que a máscara

proposta e a máscara de Wiener apresentam o melhor e o pior desempenho em todos os casos, respectivamente. A máscara proposta apresenta um aumento médio de 4% em relação à máscara raiz de Wiener. Os testes estatísticos realizados indicam novamente que todas distribuições ilustradas na Fig. 3 são estatisticamente diferentes.

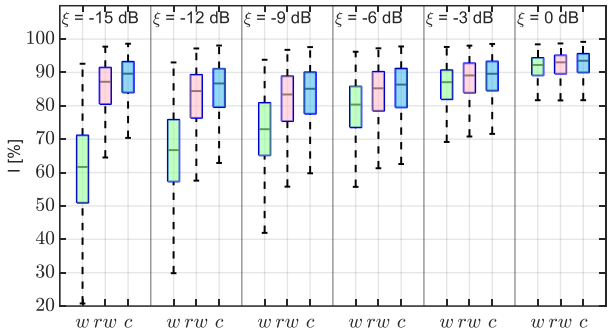

Fig. 2: Inteligibilidade percentual estimada pela métrica SRMR-CI.

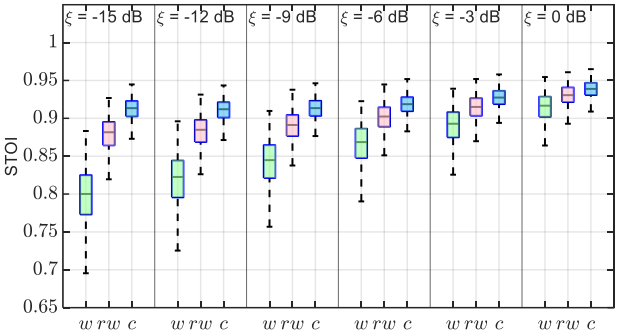

Fig. 3: Inteligibilidade segundo a métrica STOI.

VI. Conclusões

Este trabalho propôs uma máscara tempo-frequência para a redução de ruído em implantes cocleares. Essa máscara tem como objetivo minimizar o erro quadrático médio entre a envoltória da fala e sua estimativa, levando em consideração limitações computacionais nos processos de estimação. Simulações computacionais foram apresentados de forma a suportar as aproximações utilizadas e a demonstrar o aumento de inteligibilidade obtido em relação a outras máscaras utilizadas na literatura. Os resultados mostraram que a máscara proposta apresenta aumento de inteligibilidade de aproximadamente 2% em termos de SRMR-CI e 4% de STOI em relação à máscara raiz de Wiener.

Agradecimentos

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

Conflito de interesse

Os autores declaram que não há conflite de interesse.

Referências

1. WHO. (2021) World report on hearing. Geneva: World Health Organization.
2. IBGE. (2019) Pesquisa nacional de saúde 2019 : ciclos de vida do Brasil. Rio de Janeiro.
3. Wouters J, Mcdermott H J, Francart T. (2015) Sound coding in cochlear implants: From electric pulses to hearing. IEEE Signal Process. Mag. 32(2):67-80.
4. Hast A. et al. (2015) Speech perception of elderly cochlear implant users under different noise conditions. Otol. Neurotol. 36(10):1638-1643.
5. Chiea R A, Costa M H, Cordioli J A. (2021) An optimal envelope-based noise reduction method for cochlear implants: An upper bound performance investigation. IEEE/ACM Trans. Audio Speech Lang. Process. 29:1729-1739.
6. Tefili D et al. (2013) Implantes cocleares: Aspectos tecnológicos e papel socioeconômico. Rev. Bras. Eng. Biomed. 29(4):414-433.
7. Koning R et al. (2018) Perceptual and model-based evaluation of ideal time-frequency noise reduction in hearing-impaired listeners. IEEE Trans. Neural Syst. Rehabil. Eng. 26(3):687-697.
8. Chiea, R A, Costa M H, Barrault G. (2019) New insights on the optimality of parameterized Wiener filters for speech enhancement applications. Speech Commun. 109:46-54.
9. Moon I J, Hong S H. (2014) What is temporal fine structure and why is it important? Korean J. Audiol. 18(1):1.
10. Zeng F G et al. (2008) Cochlear implants: System design, integration, and evaluation. IEEE Rev. Biomed. Eng. 1:115-142.
11. Loizou P C (2013) Speech Enhancement - Theory and Practice. CRC Press.
12. Goehring T et al. (2019) Using recurrent neural networks to improve the perception of speech in non-stationary noise by people with cochlear implants. J. Acoust. Soc. Am. 146(1):705-718.
13. Zirn S et al. (2019) Reducing the device delay mismatch can improve sound localization inbimodal cochlear implant/hearing-aid users. Trends Hear. 23:1-13.
14. Falk T H et al. (2015) Objective Quality and Intelligibility Prediction for users of assistive listening devices: advantages and limitations of existing tools. IEEE Signal Process. Mag. 32(2):114-124.
15. IEEE. (1969) Recommended practice for speech quality measurements. IEEE Trans. Audio Electroacust.17(1):225-246.
16. Tseng R Y et al. (2020) A study of joint effect on denoising techniques and visual cues to improve speech intelligibility in cochlear implant simulation. IEEE Trans. Cogn. Develop. Syst. 13(4):984-994.
17. Smeds K, Wolters F, Rung M. (2015) Estimation of signal-to-noise ratios in realistic sound scenarios, J. Am. Acad. Audiol. 26(2):183-196.

Autor correspondente: Paulo Henrique Gubert
Instituição: Universidade Federal de Santa Catarina
Rua: Engenheiro Agronômico Andrei Cristian Ferreira
Cidade: Florianópolis
País: Brasil
E-mail: p.h.gubert@gmail.com