

# **AVALIAÇÃO DA TÉCNICA DE SUBTRAÇÃO ESPECTRAL DE POTÊNCIA NA REDUÇÃO DE RUÍDO EM SINAIS DE VOZ: UM ESTUDO COMPARATIVO ENTRE RUÍDO ADITIVO BRANCO E COLORIDO**

Leandro Aureliano da Silva<sup>1</sup>; Eduardo Silva Vasconcelos<sup>2</sup>; Cleiton Silvano Goulart<sup>1</sup>; William Gigo<sup>1</sup>; Lúcio Rogério Júnior<sup>1</sup>; José Ricardo Gonçalves Manzan<sup>3</sup>; Rodrigo Pinto Lemos<sup>4</sup>

<sup>1</sup> Universidade de Uberaba - Uniube

<sup>2</sup> Instituto Federal Goiano - IFG

<sup>3</sup> Instituto Federal do Triângulo Mineiro - IFTM

<sup>4</sup> Universidade Federal de Goiás - UFG

Autor correspondente: [leandro.silva@uniube.br](mailto:leandro.silva@uniube.br)

## **RESUMO**

A subtração espectral de potência é uma técnica amplamente reconhecida no processamento de sinais, dedicada a reduzir a presença indesejada de ruídos, sendo um desafio crucial na análise de sinais de voz. Neste estudo, investigamos a aplicação desta abordagem em sinais de voz afetados por ruído aditivo branco e colorido, destacando a importância fundamental de diminuir o ruído para preservar a qualidade e a precisão das informações transmitidas pela voz humana. Para avaliar sua eficácia, empregamos métricas como a relação sinal/ruído segmentada e a distância de Itakura-Saito. Os resultados revelam uma melhoria significativa na qualidade dos sinais após o processamento, ressaltando assim a importância crítica da redução de ruído no contexto da análise e transmissão de sinais de voz.

Palavras-chave: Subtração Espectral, Enriquecimento de Voz, Supressão de ruído.

---



## ABSTRACT

Power spectral subtraction is a widely recognized technique in signal processing, dedicated to reducing the unwanted presence of noise, which poses a crucial challenge in speech signal analysis. In this study, we investigate the application of this approach to voice signals affected by white and colored additive noise, emphasizing the fundamental importance of noise reduction to preserve the quality and accuracy of information transmitted by human voice. To assess its effectiveness, we employ metrics such as segmented signal-to-noise ratio and Itakura-Saito distance. The results reveal a significant improvement in signal quality after processing, thus underscoring the critical importance of noise reduction in the context of speech signal analysis and transmission.

Keywords: Spectral Subtraction, Voice Enhancement, Noise Suppression.

## 1. INTRODUÇÃO

Nos últimos anos, houve uma profusão de algoritmos desenvolvidos com o intuito de aprimorar a qualidade dos sinais de fala. Em praticamente todas as situações de transmissão de sinais de fala, diversos fatores, como ruído ambiente, eco, perdas durante a codificação de enlace digital e interferências de outras fontes de sinal, podem prejudicar a comunicação Silva (2007). Esses elementos podem ter diferentes impactos no sinal, desde reduzir sua inteligibilidade até tornar-se a conversa ininteligível, além de afetar a eficiência de sistemas subsequentes, como reconhecimento ou codificação de voz. Portanto, diversos métodos foram desenvolvidos para aprimorar a qualidade dos sinais de fala, com o objetivo de realçar a inteligibilidade e a naturalidade da comunicação. Entre esses métodos, destacam-se a aplicação da wavelet na redução de sinais de voz, onde as pesquisas se beneficiam dos limiares utilizados após a decomposição do sinal em aproximação e detalhe Silva *et al.* (2014), e os filtros de Kalman no tempo discreto e frequência discreta, que se mostram eficientes, conforme o trabalho de Silva *et al.* (2018). Mais recentemente, o uso de redes de camada profunda tem ganhado destaque, como demonstrado no trabalho de Araújo (2020), que ressalta a crescente presença de Sistemas de Reconhecimento Automático de Voz (SRAV) em diversas áreas da vida cotidiana, incluindo aplicativos de pesquisa por voz, assistentes virtuais, automação residencial e veicular, entre outros. No entanto, em ambientes ruidosos, como aqueles com altos níveis de ruído e múltiplas fontes de interferência, a precisão e confiabilidade desses sistemas podem ser comprometidas. Por isso, o trabalho de Araújo (2020), investigou a eficiência de uma rede neural profunda na atenuação de ruído em sinais de voz e comparou os seus resultados com algoritmos tradicionais de redução de ruído, como o filtro de Wiener, entre outros.

Assim, o objetivo deste estudo é realizar uma análise comparativa do desempenho da subtração espectral de potência na presença de ruído aditivo

branco e colorido. Serão empregadas duas técnicas de medição tradicionais para verificar sua eficácia: a relação sinal/ruído segmentada e a distância de Itakura-Saito.

## 2. SUBTRAÇÃO ESPECTRAL DE POTÊNCIA

Para que se possa desenvolver o método, assume-se que o sinal contaminado por ruído seja dado por:

$$y(n) = x(n) + v(n) \quad (1)$$

em que  $y(n)$ ,  $x(n)$  e  $v(n)$  são, sinal contaminado por ruído, sinal livre de ruído e ruído aditivo, respectivamente.

Posteriormente,  $y(n)$  passa pela Transformada Discreta de Fourier (DFT), e o espectro resultante do sinal com a adição de ruído é elevado ao quadrado. Seguindo a metodologia proposta por Berouti *et al.* (1979), ocorre a subtração do espectro de potência do sinal contaminado pelo ruído pelo espectro de potência do ruído estimado em períodos de silêncio. Dessa forma, a subtração espectral de potência é executada conforme descrito a seguir:

$$|\hat{X}(e^{j\omega_k})|^2 = |Y(e^{j\omega_k})|^2 - |\mu(e^{j\omega_k})|^2 \quad (2)$$

em que  $|\hat{X}(e^{j\omega_k})|^2$  é o espectro de potência estimado do sinal livre de ruído,  $|Y(e^{j\omega_k})|^2$  é o espectro de potência do sinal degradado por ruído e  $|\mu(e^{j\omega_k})|^2$  é o valor médio do espectro de potência do ruído avaliado em trechos de silêncio.

Após a subtração espectral, podemos ter um fenômeno conhecido como ruído musical. Ele é causado devido a algumas limitações ou imprecisões no processo de estimativa e remoção do ruído. Para tanto, Berouti *et al.* (1979), modificou a Equação (2), com a inserção de um termo denominado fator de subtração. A Equação (3) mostra essa modificação:

$$|\hat{X}(e^{j\omega_k})|^2 = |Y(e^{j\omega_k})|^2 - \alpha |\mu(e^{j\omega_k})|^2 \quad (3)$$

em que  $\alpha$  é o fator de subtração. Este por sua vez, tem como finalidade o controle de redução de ruído, e é determinado empiricamente segundo Berouti *et al.* (1979) por:

$$\alpha = \alpha_0 - \frac{3}{20} SNR - 5dB \leq SNR \leq 20dB \quad (4)$$

em que  $\alpha_0$  é o valor de  $\alpha$  encontrado quando a relação sinal/ruído (SNR) for 0dB.

Outro passo significativo a se fazer na subtração espectral de potência é o de retificação. Nesse caso, é adicionado um limite espectral cuja função é não perder informações do sinal de baixa amplitude conforme relata Berouti *et al.* (1979). Esta mudança é dada pela Equação 5:

$$|\hat{X}(e^{j\omega_k})|^2 = \begin{cases} |\hat{X}(e^{j\omega_k})|^2, & \text{se } |\hat{X}(e^{j\omega_k})|^2 > \beta|\mu(e^{j\omega_k})|^2 \\ \beta|\mu(e^{j\omega_k})|^2, & \text{caso contrário} \end{cases} \quad (5)$$

em que  $\beta$  é o limite espectral, cuja função é mascarar o efeito das componentes do ruído musical. Segundo Berouti *et al.* (1979),  $\beta$  é encontrado a partir da relação sinal/ruído (SNR), podendo ter valores típicos que variam de 0,02 até 0,06 para uma  $SNR \leq 0\text{dB}$  e valores que estão entre 0,0005 até 0,02 para  $SNR > 0\text{dB}$ . Neste trabalho, o valor de  $\beta$ , foi fixado com o valor de 0,02.

Assim, o diagrama em blocos apresentado pela Figura 1, mostra as operações envolvidas no método de subtração espectral de potência.

Para restauração do sinal no domínio do tempo, o espectro de magnitude obtido após a raiz quadrada, gerando o sinal estimado  $|\hat{X}(e^{j\omega_k})|$  é combinado com a fase do sinal contaminado por ruído, e então passado para o domínio do tempo via transformada discreta de Fourier inversa (IDFT) como:

$$\hat{x}(n) = \sum_{k=0}^{N-1} |\hat{X}(e^{j\omega_k})| e^{j\theta_Y(e^{j\omega_k})} e^{-j\omega_k n} \quad (6)$$

em que  $w_k = \frac{2\pi}{N}k$  é a frequência discreta da transformada e  $\theta_Y(e^{j\omega_k})$  é a fase do sinal contaminado por ruído no domínio da frequência.

### 3. RESULTADOS E DISCUSSÃO

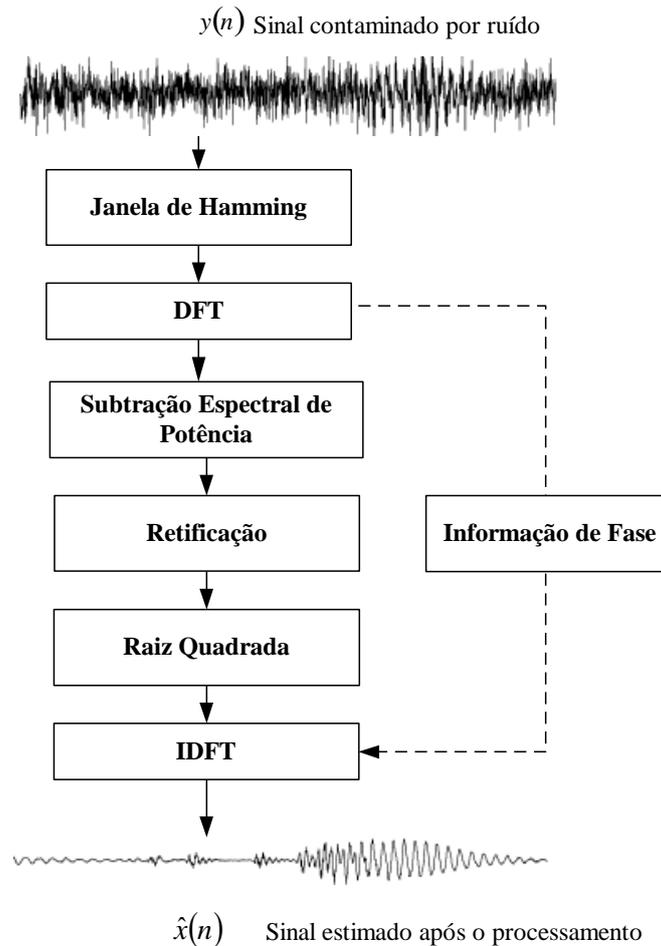
Para avaliar o algoritmo mencionado anteriormente, foram empregados diversos sinais de voz amostrados a uma taxa de 22050 Hz com resolução de 16 bits. Esses sinais foram deliberadamente contaminados por ruído branco e colorido. A janela aplicada no algoritmo de Subtração Espectral de Potência é a de Hamming com 512 amostras e com sobreposição de 50%. Para efetuar as simulações foi utilizado o software Matlab R2013B em um computador Core i7 com 8G de RAM.

Como medidas de desempenho, utilizou-se a relação sinal/ruído segmentada dada Deller; Proakis e Hansen (1993). Ela é dada por:

$$SNR_{seg} = \frac{10}{M} \sum_{j=0}^{M-1} \log_{10} \left[ \sum_{n=m_j-N+1}^{m_j} \frac{x^2(n)}{[x(n) - \hat{x}(n)]^2} \right] \quad (7)$$

em que:  $m_j$  representa o limites de cada um dos M quadros de tamanho N.

**Figura 1:** Diagrama em blocos da Subtração Espectral de Potência.



**Fonte:** Os autores.

A outra medida foi a distância de Itakura-Saito que foi usada para avaliar se o sinal processado sofreu distorções quando reconstruído em relação ao sinal original Silva, (2007). Essa medida é dada por:

$$d(a, b) = \log \left[ \frac{aRa^T}{bRb^T} \right] \quad (8)$$

em que: "a" é o vetor de coeficientes LPC do sinal original; "R" é a matriz de autocorrelação do sinal original e "b" é o vetor de coeficientes LPC do sinal estimado.

Cabe ressaltar que o valor da Distância de Itakura-Saito é uma métrica que mede a similaridade entre dois sinais, em que valores mais baixos indicam uma maior semelhança entre os sinais. No contexto da análise de sinais de voz, um valor baixo indica que o sinal reconstruído após a subtração de ruído é mais próximo do sinal de voz original.

Para comprovar de fato a eficiência da técnica proposta, objetivando mínima distorção espectral e redução de ruído na reconstrução do sinal, é apresentado o resultado de 25 palavras gravadas por diferentes pessoas e processadas pelo algoritmo de subtração espectral de potência com relação sinal ruído segmentada de entrada de 3dB e contaminadas com ruído branco. O Quadro 1, apresenta estes resultados:

**Quadro 1** – Resultados dos testes para as 25 palavras com SNRseg de entrada de 3 dB, usando a subtração espectral de potência, para o caso de o ruído ser branco.

<b>Relação Sinal/Ruído de Entrada de 3 dB, para a técnica de Subtração Espectral de Potência – Ruído Branco</b>		
<b>Palavras</b>	<b>Relação Sinal/Ruído de Saída em dB</b>	<b>Distância de Itakura-Saito</b>
Bala	9 dB	0,9793
Boi	7 dB	0,8925
Bola	9 dB	0,1938
Cão	10 dB	0,8510
Casa	5 dB	0,2109
Cola	9 dB	0,3160
Dado	7 dB	0,6304
Dia	8 dB	0,4304
Elefante	9 dB	0,2391
Elétrica	8 dB	0,3947
Fala	6 dB	0,8639
Igreja	8 dB	0,3812
Janela	8 dB	0,6015
Lã	9 dB	0,8425
Lata	11 dB	0,3991
Maio	10 dB	0,7600
Manga	8 dB	0,6848
Mão	8 dB	0,8785
Medo	7 dB	0,4964
Mel	10 dB	0,9687
Modem	10 dB	0,5478
Vovó	8 dB	0,7530
Tabela	8 dB	0,4003
Tia	9 dB	0,5146
Tio	9 dB	0,4429
Média	8 dB	0,5859

**Fonte:** Elaborado pelos autores.

Após ter apresentado os resultados com ruído branco, os sinais acima foram contaminados por ruído colorido. Para tanto, utilizou-se um filtro AR no ruído branco gerado anteriormente. Esse filtro foi de ordem 8 de acordo com Ma. Ning *et al.* 2004 conforme mostra a Equação 9:

$$\begin{aligned}
 v(n) = & -0.0851v(n-1) + 0.019126v(n-2) \\
 & + 0.0458v(n-3) + 0.0229v(n-4) \\
 & + 0.1057v(n-5) + 0.1553v(n-6) \\
 & - 0.132v(n-7) - 0.76v(n-8) + w(n)
 \end{aligned}
 \tag{9}$$

Os resultados para as 25 palavras com ruído colorido adicionado podem ser verificados no Quadro 2.

**Quadro 2** – Resultados dos testes para as 25 palavras com SNRseg de entrada de 3 dB, usando a subtração espectral de potência, para o caso de o ruído ser colorido.

Relação Sinal/Ruído de Entrada de 3 dB, para a técnica de Subtração Espectral de Potência – Ruído colorido		
Palavras	Relação Sinal/Ruído de Saída em dB	Distância de Itakura-Saito
Bala	9 dB	0,7404
Boi	7 dB	0,6290
Bola	9 dB	0,9234
Cão	10 dB	0,5415
Casa	5 dB	0,1661
Cola	9 dB	0,1022
Dado	7 dB	0,4260
Dia	8 dB	0,2425
Elefante	9 dB	0,1961
Elétrica	8 dB	0,2758
Fala	5 dB	0,6593
Igreja	8 dB	0,2230
Janela	7 dB	0,3319
Lã	8 dB	0,6559
Lata	11 dB	0,2540
Maior	10 dB	0,5505
Manga	8 dB	0,4943
Mão	8 dB	0,6513
Medo	6 dB	0,2986
Mel	9 dB	0,7943
Modem	9 dB	0,3515
Vovó	8 dB	0,4933
Tabela	8 dB	0,2138
Tia	8 dB	0,3394
Tio	9 dB	0,3280
Média	8 dB	0,4352

Fonte: Elaborado pelos autores.

Analisando os Quadros 1 e 2, percebe-se que a técnica de subtração espectral de potência foi eficaz em relação a relação sinal-ruído segmentada, tanto no sinal contaminado por ruído branco quanto colorido. No entanto, em relação à distorção espectral, observa-se que a técnica, quando aplicada ao ruído colorido, apresenta uma menor distorção espectral na média.

## 4. CONCLUSÕES

Este trabalho apresentou o estudo e a implementação do algoritmo de subtração espectral de potência com a finalidade de verificar o seu desempenho na presença de sinais de voz contaminados por ruídos aditivo branco e colorido. Para tanto, foi utilizado as medidas de relação sinal ruído segmentada e a distância de Itakura-Saito. Os testes foram realizados em 25 sinais de voz de diferentes pessoas com gravação de 16 bits e frequência de amostragem de 22050 Hz. Todos os sinais foram contaminados por ruído branco e colorido, sendo que os testes mostraram que a relação sinal ruído segmentada de saída são bem próximas. No entanto, verificou-se que a distorção espectral é menor quando o sinal é contaminado por ruído colorido, como pode ser verificado nos testes.

## REFERÊNCIAS

- ARAÚJO, Jacques Henrique Bessa. **Atenuação de ruído em sinais de voz utilizando redes neurais profundas**. 2020. Disponível em: <https://repositorio.ufc.br/handle/riufc/52548>. Acesso em: 6. abr. 2024.
- BEROUTI, Michael; SCHWARTZ, Richard; MAKHOUL, John. Enhancement of speech corrupted by acoustic noise. *In: ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 1979. p. 208-211.
- DELLER JR, John R.; PROAKIS, John G.; HANSEN, John H. **Discrete time processing of speech signals**. Prentice Hall PTR, 1993.
- MA, Ning; BOUCHARD, Martin; GOUBRAN, Rafik A. Perceptual Kalman filtering for speech enhancement in colored noise. *In: 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2004. p. I-717.
- SILVA, Leandro Aureliano da *et al.* Aplicação da transformada Wavelet na Redução de Ruído em Sinais de Voz. *In: XII Conferência de Estudos em Engenharia Elétrica CEEL*, 2014. Disponível em: [https://peteetricaufu.com.br/static/ceel/doc/artigos/artigos2014/ceel2014\\_artigo017\\_r01.pdf](https://peteetricaufu.com.br/static/ceel/doc/artigos/artigos2014/ceel2014_artigo017_r01.pdf).
- SILVA, Leandro Aureliano da *et al.* Comparative Study between the Discrete-Frequency Kalman Filtering and the Discrete-Time Kalman Filtering with Application in Noise Reduction in Speech Signals. **Journal of Electrical and Computer Engineering**, v. 2018, 2018.
- SILVA, Leandro Aureliano da. **Filtros de Kalman no tempo e frequência discretos combinados com subtração espectral**. 2007. Dissertação (Mestrado em Engenharia Elétrica), Universidade de São Paulo, 2007.