v.7, n.4, 2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

AVALIAÇÃO COMPARATIVA DE FAMÍLIAS WAVELET NA REDUÇÃO DE RUÍDO EM SINAIS DE VOZ CONTAMINADOS POR RUÍDO COLORIDO

COMPARATIVE EVALUATION OF WAVELET FAMILIES IN NOISE REDUCTION OF SPEECH SIGNALS CONTAMINATED BY COLORED NOISE

EVALUACIÓN COMPARATIVA DE FAMILIAS WAVELET EN LA REDUCCIÓN DE RUIDO EN SEÑALES DE VOZ CONTAMINADAS POR RUIDO COLOREADO

Leandro Aureliano da Silva[1], Eduardo Silva Vasconcelos[2], Luiz Fernando Ribeiro de Paiva[3], Welington Mrad Joaquim[4], Adriano Dawison de Lima[5], Edilberto Pereira Teixeira[6]

e747696
https://doi.org/10.47820/recima21.v7i4.7696

PUBLICADO: 04/2026

RESUMO

A presença de ruído em sinais de voz constitui um desafio relevante em aplicações de processamento digital de fala, como sistemas de comunicação, reconhecimento automático de fala e dispositivos de assistência auditiva. Nesse contexto, técnicas baseadas na transformada wavelet têm sido amplamente empregadas devido à sua capacidade de análise multirresolução, permitindo a decomposição do sinal em diferentes escalas de frequência. Este trabalho apresenta uma avaliação comparativa do desempenho de diferentes famílias de wavelets na redução de ruído em sinais de voz contaminados por ruído colorido. Foram analisadas wavelets das famílias Daubechies, Symlet e Coiflet, utilizando decomposição até o terceiro nível e limiarização baseada no método rigrsure (Estimativa de Risco Não Viesado de Stein). Os experimentos foram conduzidos com dez sinais de voz limpa combinados a dez realizações independentes de ruído colorido, totalizando cem condições experimentais. O desempenho foi avaliado por meio das métricas SegSNR, STOI e distância de Itakura–Saito baseada em LPC. Os resultados indicaram ganhos médios de aproximadamente 3 dB na SegSNR após o processo de redução de ruído. As wavelets Daubechies apresentaram maior ganho médio em SegSNR, enquanto as Coiflet, especialmente a coif4, demonstraram melhor desempenho global ao equilibrar redução de ruído e

[1] Doutor em Engenharia Elétrica pela UFU (2018) e mestre pela USP (2007), com especialização em Automação Industrial e graduação em Engenharia Elétrica. Atua em pesquisa, ensino e extensão nas áreas de IoT, automação, inteligência artificial e processamento de sinais.
[2] Doutor em Ciências pela UFU, mestre em Matemática (UFG) e em Educação Superior (UNITRI). Graduado em Matemática, Administração e Gestão do Agronegócio, com diversas especializações e MBA em Gestão Pública. Atua na educação desde 1990 no ensino médio e desde 2003 no ensino superior.
[3] Doutor em Educação pela UNIUBE e mestre em Ciência da Informação pela PUC-Campinas. Possui especializações em Avaliação na Educação Superior (UnB) e Análise de Sistemas (UNAERP), além de aperfeiçoamento em Educação a Distância. É tecnólogo em Processamento de Dados pela UNIUBE e atua como gestor de cursos de tecnologia.
[4] Doutorando em Educação pela UNIUBE e mestre em Ensino de Ciências e Matemática (PUC Minas), com ênfase em Física. Possui diversas especializações em educação, tecnologias e ensino de ciências, além de formação multidisciplinar em Computação, Gastronomia, Engenharia Ambiental, Pedagogia, Matemática e Física.
[5] Graduado em Licenciatura em Matemática pela Universidade de Uberaba (2004), com mestrado (2006) e doutorado (2009) em Energia na Agricultura pela UNESP. Atua como docente em tempo integral na Universidade de Uberaba. Possui experiência em Matemática Aplicada, com ênfase em Álgebra Linear, Geometria Analítica e Cálculo.
[6] Graduado, mestre e doutor em Engenharia Elétrica pela UNIFEI e UNICAMP. Possui ampla experiência na área, com ênfase em sistemas elétricos de potência e controle de processos. Atua em temas como controle multivariável, sistemas não lineares, lógica nebulosa e redes neurais.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

preservação da inteligibilidade. As Symlet, como a sym3, destacaram-se na preservação do envelope espectral. A análise estatística, baseada no teste de Friedman, confirmou diferenças significativas entre os métodos. Os resultados evidenciam que a escolha da wavelet ideal depende do objetivo da aplicação.

PALAVRAS-CHAVE: Processamento de sinais de voz. Transformada wavelet. Redução de ruído. Melhoria da fala. Ruído colorido.

ABSTRACT
The presence of noise in speech signals represents a significant challenge in digital speech processing applications, such as communication systems, automatic speech recognition, and hearing assistive devices. In this context, wavelet-based techniques have been widely used due to their multiresolution analysis capability, allowing signal decomposition across different frequency scales. This study presents a comparative evaluation of the performance of different wavelet families for noise reduction in speech signals contaminated by colored noise. Wavelets from the Daubechies, Symlet, and Coiflet families were analyzed using decomposition up to level three and thresholding based on the rigrsure method (Stein's Unbiased Risk Estimate). The experiments were conducted using ten clean speech signals combined with ten independent realizations of colored noise, resulting in one hundred experimental conditions. Performance was evaluated using SegSNR, STOI, and the Itakura–Saito distance based on Linear Predictive Coding (LPC). Results showed average gains of approximately 3 dB in SegSNR after denoising. Daubechies wavelets achieved the highest average SegSNR improvement, while Coiflet wavelets, particularly coif4, demonstrated the best overall performance by balancing noise reduction and speech intelligibility preservation. Symlet wavelets, such as sym3, showed superior performance in preserving the spectral envelope. Statistical analysis based on the Friedman test confirmed significant differences among the evaluated methods. The results indicate that the optimal wavelet choice depends on the specific application objectives.

KEYWORDS: *Speech processing. Wavelet transform. Speech enhancement; Colored noise. Wavelet thresholding.*

RESUMEN
La presencia de ruido en señales de voz representa un desafío significativo en diversas aplicaciones de procesamiento digital del habla, como sistemas de comunicación, reconocimiento automático de voz y dispositivos de asistencia auditiva. En este contexto, las técnicas basadas en transformadas wavelet han sido ampliamente utilizadas debido a su capacidad de análisis multirresolución, permitiendo descomponer la señal en diferentes escalas de frecuencia. Este trabajo presenta una evaluación comparativa del desempeño de diferentes familias de wavelets en la reducción de ruido en señales de voz contaminadas con ruido coloreado. Se analizaron wavelets de las familias Daubechies, Symlet y Coiflet, utilizando descomposición hasta el nivel tres y umbralización basada en el método rigrsure (Estimación de Riesgo No Sesgada de Stein). Los experimentos se realizaron con diez señales de voz limpia combinadas con diez realizaciones independientes de ruido coloreado, totalizando cien condiciones experimentales. El desempeño fue evaluado mediante las métricas SegSNR, STOI y la distancia de Itakura–Saito basada en Codificación Predictiva Lineal (LPC). Los resultados indicaron ganancias promedio de aproximadamente 3 dB en la SegSNR tras la reducción de ruido. Las wavelets Daubechies presentaron mayor ganancia promedio en SegSNR, mientras que las Coiflet, especialmente la coif4, mostraron el mejor desempeño global al equilibrar reducción de ruido y preservación de la inteligibilidad. Las Symlet, como la sym3, destacaron en la preservación del envolvente espectral. El análisis estadístico basado en la prueba de Friedman confirmó diferencias significativas entre los métodos. Los resultados indican que la elección de la wavelet óptima depende del objetivo de la aplicación.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

***PALABRAS CLAVE**: Procesamiento de señales de voz. Transformada wavelet. Reducción de ruido. Mejora del habla. Ruido coloreado.*

INTRODUÇÃO

A presença de ruído em sinais de voz representa um dos principais desafios em aplicações de processamento digital de fala, pois afeta diretamente a qualidade e a inteligibilidade do sinal em sistemas de comunicação, reconhecimento automático de fala e dispositivos de assistência auditiva. Em ambientes reais, o sinal captado pelos sensores acústicos frequentemente é contaminado por diferentes tipos de ruído provenientes do ambiente, o que pode comprometer significativamente a transmissão e a interpretação da informação acústica (Loizou, 2013).

Nos últimos anos, a área de redução de ruído em sinais de fala tem evoluído de forma significativa, impulsionada pela crescente demanda por sistemas de comunicação mais robustos e tecnologias baseadas em voz. Revisões recentes indicam que novas abordagens vêm sendo desenvolvidas com o objetivo de melhorar simultaneamente a qualidade e a inteligibilidade da fala, incorporando técnicas híbridas que combinam métodos clássicos de processamento de sinais com abordagens baseadas em aprendizado de máquina e análise multirresolução (Yousif *et al.,* 2025).

Nesse contexto, a transformada wavelet permanece como uma ferramenta relevante para análise e processamento de sinais de fala, principalmente devido à sua capacidade de representar simultaneamente informações nos domínios do tempo e da frequência. Essa característica permite analisar sinais não estacionários de forma mais eficiente, possibilitando a identificação e separação de componentes associadas ao sinal útil e ao ruído. Estudos recentes têm explorado o uso da decomposição wavelet em conjunto com arquiteturas modernas de processamento de sinais, demonstrando melhorias significativas na qualidade e na inteligibilidade da fala quando comparadas a métodos baseados exclusivamente em transformadas espectrais tradicionais (Wu *et al.,* 2025).

Além disso, novas abordagens têm investigado a integração entre técnicas de redução de ruído baseadas em wavelets e métodos híbridos de processamento, buscando melhorar o desempenho dos algoritmos em cenários com ruídos complexos ou não estacionários. Resultados recentes indicam que a combinação da transformada wavelet discreta com outros algoritmos de processamento pode aumentar significativamente a eficácia da redução de ruído, ao mesmo tempo em que preserva melhor as características perceptuais do sinal de fala (Iqbal *et al.,* 2025). Esses avanços reforçam a importância da investigação contínua sobre o desempenho de diferentes funções wavelet no contexto de algoritmos de redução de ruído.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

De forma mais específica, métodos baseados na transformada wavelet têm recebido crescente atenção na literatura devido à sua capacidade de realizar análise multirresolução do sinal, permitindo separar componentes associadas a diferentes escalas temporais e espectrais. Essa propriedade torna a transformada wavelet particularmente adequada para aplicações envolvendo sinais de voz, os quais apresentam natureza altamente não estacionária (Mallat, 2009).

Uma das aplicações mais relevantes da transformada wavelet em processamento de sinais é a técnica de limiarização de coeficientes wavelet. Esse método foi inicialmente proposto por Donoho e Johnstone (1995), que demonstraram que a atenuação de coeficientes associados ao ruído pode promover melhorias significativas na qualidade do sinal reconstruído. Desde então, diversos métodos de seleção de limiar foram desenvolvidos, entre os quais se destaca o método rigrsure (Estimativa de Risco Não Viesada de Stein), capaz de estimar automaticamente um limiar ótimo com base nas características estatísticas do sinal observado.

Outro aspecto importante em aplicações de redução de ruído em fala refere-se ao tipo de ruído considerado nos experimentos. Embora muitos estudos utilizem ruído branco aditivo como modelo de degradação, ambientes acústicos reais frequentemente apresentam ruídos coloridos, cuja densidade espectral de potência varia com a frequência. Esse tipo de ruído possui características mais complexas e pode impactar significativamente o desempenho dos algoritmos de redução de ruído em sinais de voz (Benesty; Chen; Huang, 2008).

Adicionalmente, diferentes famílias de wavelets apresentam propriedades matemáticas distintas, como número de momentos nulos, regularidade e comprimento dos filtros. Essas características influenciam diretamente a capacidade da transformada wavelet de representar diferentes estruturas do sinal, podendo afetar o desempenho dos métodos de redução de ruído. Entre as famílias de wavelets mais utilizadas na literatura destacam-se as wavelets Daubechies, Symlet e Coiflet, amplamente empregadas em aplicações de processamento de sinais (Daubechies, 1992).

Diante desse contexto, este trabalho tem como objetivo avaliar comparativamente o desempenho de diferentes famílias de wavelets na redução de ruído em sinais de voz contaminados por ruído colorido, utilizando a técnica de limiarização baseada no método rigrsure. O desempenho dos métodos foi avaliado por meio de métricas objetivas amplamente utilizadas na literatura, incluindo a relação sinal-ruído segmentada (SegSNR), a métrica de inteligibilidade de curto prazo - STOI e a distância de Itakura–Saito baseada em modelos LPC.

Para a realização dos experimentos, foram utilizadas múltiplas realizações de ruído colorido aplicadas a sinais de voz limpa, permitindo avaliar de forma robusta o comportamento das diferentes funções wavelet avaliadas e comparar seu desempenho em termos de redução de ruído, preservação da inteligibilidade da fala e manutenção das características espectrais do sinal.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

1. REFERENCIAL TEÓRICO

1.1. Redução de ruído em sinais de voz

A redução de ruído em sinais de voz tem como principal objetivo recuperar ou melhorar a qualidade de um sinal de fala degradado por ruído. Esse problema tem sido amplamente estudado nas áreas de processamento digital de sinais e engenharia acústica, devido à sua importância em aplicações como telecomunicações, sistemas de reconhecimento automático de fala e dispositivos auditivos (Loizou, 2013).

Entre as abordagens tradicionais utilizadas para redução de ruído destacam-se métodos baseados em subtração espectral, filtragem adaptativa e filtragem de Wiener. Apesar de sua ampla utilização, esses métodos apresentam limitações quando aplicados a cenários com ruído não estacionário ou com características espectrais complexas (Vaseghi, 2008).

1.2. Transformada Wavelet Discreta

A transformada wavelet discreta (DWT) constitui uma ferramenta poderosa para análise de sinais não estacionários, permitindo representar o sinal simultaneamente nos domínios do tempo e da frequência. Diferentemente da transformada de Fourier, que fornece apenas informações espectrais globais, a transformada wavelet permite analisar o sinal em diferentes escalas de resolução (Mallat, 2009).

Matematicamente, a transformada wavelet de um sinal pode ser expressa pela Equação 1 como:

$$W(a,b) = \int x(t)\psi_{a,b}(t)dt \qquad (1)$$

onde:
$x(t)$ representa o sinal analisado
a corresponde ao fator de escala
b representa a translação temporal
ψ é a função wavelet mãe.

A decomposição wavelet permite separar o sinal em coeficientes de aproximação e detalhe, correspondentes às componentes de baixa e alta frequência do sinal.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

1.3. Limiarização de coeficientes wavelet

Uma das aplicações mais importantes da transformada wavelet é a redução de ruído por limiarização de coeficientes, técnica proposta por Donoho e Johnstone (1995).

Nesse método, os coeficientes wavelet associados ao ruído são reduzidos ou eliminados através da aplicação de um limiar. Entre os principais tipos de limiarização destacam-se:

- limiarização dura *(hard thresholding)*
- limiarização suave *(soft thresholding)*

A limiarização suave é frequentemente utilizada em aplicações de processamento de fala por produzir reconstruções mais estáveis do sinal (Donoho; Johnstone, 1995).

1.4. Método de limiar rigrsure

O método “rigrsure” utiliza um estimador estatístico do erro quadrático médio para determinar automaticamente o limiar ideal para os coeficientes wavelet.

Esse método busca minimizar a função de risco associada à reconstrução do sinal, sendo particularmente eficaz quando a variância do ruído é desconhecida.

A função de risco associada ao estimador SURE pode ser expressa pela Equação 2 como (Donoho; Johnstone, 1995):

$$R(t) = \frac{1}{N}\left(N - 2k + \sum_{i=1}^{k} y_i + (N-k)y_k\right) \quad (2)$$

onde

yi representam os coeficientes wavelet ordenados.

1.5. Métricas de avaliação de qualidade de fala

A avaliação do desempenho de algoritmos de redução de ruído pode ser realizada por meio de métricas objetivas que estimam a qualidade ou inteligibilidade do sinal de voz.

Entre as métricas mais utilizadas destacam-se:

Relação sinal-ruído segmentada – SegSNR

A relação sinal-ruído segmentada mede o ganho de qualidade obtido após o processamento do sinal (Loizou, 2013).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Medida objetiva de inteligibilidade de curto prazo – STOI

A métrica STOI foi proposta por Taal *et al.* (2011) e avalia diretamente a inteligibilidade da fala em presença de ruído.

Distância de Itakura–Saito

A distância de Itakura–Saito é amplamente utilizada em análise de fala baseada em modelos LPC, sendo uma medida da discrepância entre dois espectros de potência (Itakura; Saito, 1970).

2. MÉTODOS

2.1. Base de dados de sinais de voz

Para a realização dos experimentos foi utilizado um conjunto de 10 sinais de voz limpa, gravados especificamente para este estudo. Os sinais correspondem às palavras isoladas: bala, casa, dado, elefante, elétrica, fala, igreja, janela, lata e maio.

As gravações foram realizadas em ambiente silencioso utilizando microfone de computador com taxa de amostragem de 16 kHz, resolução de 16 bits e canal monofônico, configuração amplamente utilizada em aplicações de processamento digital de fala (Loizou, 2013).

Após a aquisição, os sinais foram submetidos a um processo de normalização de amplitude, de modo a garantir comparabilidade entre os experimentos. Em seguida, os sinais foram utilizados como base para a geração das versões contaminadas por ruído colorido.

O uso de palavras isoladas permite avaliar de forma controlada o desempenho dos algoritmos de redução de ruído, mantendo variabilidade fonética suficiente para análise do comportamento dos métodos em diferentes estruturas espectrais da fala.

Embora o conjunto inicial de dados seja composto por apenas dez palavras isoladas, a robustez experimental do estudo foi ampliada por meio da geração de múltiplas realizações independentes de ruído para cada sinal de voz. Para cada palavra foram geradas dez instâncias distintas de ruído colorido, resultando em um total de 100 sinais degradados utilizados na avaliação dos algoritmos. Esse procedimento permite aumentar significativamente a variabilidade experimental e reduzir possíveis vieses associados a um conjunto de dados limitado, sendo uma estratégia amplamente utilizada em experimentos de processamento de sinais para validação estatística de métodos de redução de ruído.

2.2. Geração de ruído colorido

Para simular cenários realistas de degradação acústica, cada sinal de voz foi contaminado com ruído colorido aditivo. Diferentemente do ruído branco, o ruído colorido apresenta densidade

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

espectral de potência dependente da frequência, sendo mais representativo de ambientes acústicos reais (Vaseghi, 2008). A geração do ruído seguiu o modelo espectral descrito na (Equação 3):

$$S(f) \propto \frac{1}{f^{\beta}} \tag{3}$$

onde β controla a inclinação espectral do ruído e, de acordo com o seu valor, são gerados diferentes tipos de ruídos coloridos. Incluindo:

- ruído rosa (β = 1)
- ruído marrom (β = 2)
- ruído azul (β = −1)
- ruído violeta (β = −2)

Para cada um dos 10 sinais de voz, foram geradas 10 realizações independentes de ruído, totalizando 100 sinais contaminados, garantindo robustez estatística ao experimento.

2.3. Ajuste da relação sinal-ruído segmentada

A intensidade do ruído foi ajustada de forma a produzir uma relação sinal-ruído segmentada (SegSNR) de entrada de aproximadamente 3 dB.

A métrica SegSNR é amplamente utilizada em avaliação de algoritmos de redução de ruído em fala, pois considera a variabilidade temporal do sinal de voz, calculando a relação sinal-ruído em janelas curtas (Loizou, 2013).

A SegSNR é definida pela (Equação 4):

$$SegSNR = \frac{1}{M}\sum_{m=1}^{M} 10\log_{10}\left(\frac{\sum s_m^2(n)}{\sum (s_m(n) - \hat{s}_m(n))^2}\right) \tag{4}$$

onde:

$s_m(n)$ representa o sinal limpo no quadro m

$\hat{s}_{m(n)}$ representa o sinal processado

M corresponde ao número total de quadros analisados.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Nos experimentos realizados, foram utilizados quadros de 32 ms com sobreposição de 50%, configuração frequentemente utilizada em aplicações de análise de fala.

2.4. Redução de ruído baseada em wavelets

A redução de ruído foi realizada utilizando a transformada wavelet discreta (DWT), técnica amplamente empregada em processamento de sinais devido à sua capacidade de análise multirresolução (Mallat, 2009). Para tanto, foram avaliadas wavelets pertencentes às seguintes famílias:

- Daubechies
- Symlet
- Coiflet

As wavelets analisadas nos testes foram:

- db2 até db8
- sym2 até sym8
- coif1 até coif5

A decomposição wavelet foi realizada até o nível 3, permitindo separar o sinal em componentes de diferentes faixas de frequência.

O processo de redução de ruído seguiu as etapas:

1. Decomposição wavelet do sinal ruidoso;
2. Estimativa do nível de ruído nos coeficientes de detalhe;
3. Aplicação de limiarização nos coeficientes wavelet;
4. Reconstrução do sinal por transformada wavelet inversa.

A Figura 1 apresenta o *pipeline* (fluxo de processamento) geral do processo experimental adotado neste estudo. Inicialmente, os sinais de voz limpa são contaminados com ruído colorido. Em seguida, é aplicada a transformada wavelet discreta (DWT), seguida do processo de limiarização, utilizando o método rigrsure. Após a reconstrução do sinal por meio da transformada wavelet inversa (IDWT), são calculadas as métricas objetivas de avaliação, incluindo: SegSNR, STOI e distância de Itakura–Saito.

A implementação dos experimentos foi realizada em linguagem Python, utilizando bibliotecas científicas amplamente empregadas em processamento de sinais, incluindo NumPy, SciPy e PyWavelets. Essas bibliotecas permitem a realização eficiente da transformada wavelet discreta, bem como a manipulação e análise de sinais digitais. A escolha do ambiente Python

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

deve-se à sua ampla adoção na comunidade científica e à disponibilidade de ferramentas robustas para análise numérica e processamento de sinais.

2.5. Limiarização rigrsure

Para a eliminação dos coeficientes dominados por ruído foi utilizado o limiar rigrsure, baseado no método *Stein's Unbiased Risk Estimate* (SURE) (Donoho; Johnstone, 1995).

Esse método determina automaticamente um limiar ótimo minimizando a estimativa de risco entre o sinal original e o sinal reconstruído.

Após a determinação do limiar, foi aplicada limiarização suave (*soft thresholding*) nos coeficientes de detalhe.

Figura 1. Diagrama em blocos do *Pipeline* (fluxo de processamento) usado



Fonte: Elaborado pelos autores (2026).

2.6. Métricas de avaliação

O desempenho das diferentes wavelets foi avaliado por três métricas objetivas amplamente utilizadas na literatura de processamento de fala.

SegSNR de saída

A SegSNR de saída mede o ganho obtido após o processamento de redução de ruído.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

STOI

A métrica ***Short-Time Objective Intelligibility*** **(STOI)** avalia a inteligibilidade da fala após o processamento (TAAL *et al.*, 2011). Essa métrica varia entre 0 e 1, sendo valores próximos de 1 indicativos de maior inteligibilidade.

Distância de Itakura–Saito

A distância de Itakura–Saito foi utilizada para avaliar a distorção espectral do sinal processado. Essa métrica é baseada em modelos LPC (*Linear Predictive Coding*) e mede a discrepância entre os espectros do sinal limpo e do sinal reconstruído (ITAKURA; SAITO, 1970). Vale ressaltar que valores menores indicam maior fidelidade espectral.

2.7. Análise estatística

Para avaliar diferenças significativas entre as famílias wavelets foi utilizado o teste não paramétrico de Friedman, apropriado para experimentos com medidas repetidas (DEMŠAR, 2006).

Cada combinação entre sinal de voz e realização de ruído foi considerada uma repetição experimental independente, resultando em um total de 100 observações para cada wavelet analisada.

3. RESULTADOS E DISCUSSÃO

O Quadro 1 apresenta os valores médios obtidos para cada família wavelet considerando todas as métricas avaliadas ao longo dos experimentos realizados.

Observa-se inicialmente que a SegSNR de entrada se manteve praticamente constante, com valor médio aproximado de 3,27 dB para todas as configurações experimentais. Esse comportamento indica consistência no processo de geração do ruído colorido, garantindo que todos os métodos avaliados foram submetidos às mesmas condições de degradação do sinal.

Os resultados indicam que wavelets pertencentes à família Daubechies apresentaram, em geral, maior ganho na métrica SegSNR de saída. Entre as wavelets analisadas, a db8 apresentou o maior valor médio de SegSNR de saída, atingindo aproximadamente 6,33 dB, o que corresponde a um ganho médio de cerca de 3 dB em relação ao sinal ruidoso.

Esse comportamento pode ser explicado pelas propriedades matemáticas das wavelets de ordem mais elevada, que apresentam maior regularidade e filtros de suporte mais longo. Essas características favorecem uma melhor separação entre os componentes de sinal e de ruído no domínio wavelet, permitindo uma atenuação mais eficiente do ruído durante o processo de limiarização (Mallat, 2009).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Em termos de inteligibilidade da fala, avaliada pela métrica STOI, os valores observados situaram-se no intervalo entre 0,939 e 0,942, indicando que o processo de redução de ruído preservou adequadamente as características perceptuais do sinal de voz.

Entre as wavelets analisadas, a db6 apresentou o maior valor médio de STOI, com aproximadamente 0,942, sugerindo que wavelets de ordem intermediária podem proporcionar um equilíbrio favorável entre redução de ruído e preservação da inteligibilidade da fala.

A análise da distância de Itakura–Saito, apresentada também no Quadro 1, revelou um comportamento distinto das demais métricas avaliadas. Wavelets de menor ordem, como db2 e sym2, apresentaram os menores valores dessa métrica, indicando maior preservação do envelope espectral da fala após o processamento.

Esse resultado sugere que wavelets de menor ordem introduzem menor distorção espectral durante o processo de reconstrução do sinal, possivelmente devido ao menor comprimento dos filtros associados a essas funções wavelet.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

Quadro 1. Resultados médios das métricas de avaliação para cada wavelet.

family	wavelet	n	segsnr_in_mean	segsnr_out_mean	segsnr_out_std	delta_segsnr_mean	stoi_mean	stoi_std	is_mean	is_std
coif	coif1	100	3,267484247	5,756645532	3,601289379	2,489161285	0,938592496	0,058178	3,674794101	2,536715105
coif	coif2	100	3,267484247	6,225881268	4,1146719	2,958397021	0,940939266	0,060179	19,06371551	71,13329319
coif	coif3	100	3,267484247	6,298079201	4,262757525	3,030594954	0,941109746	0,059845	236,7951918	996,4186121
coif	coif4	100	3,267484247	6,323286338	4,270350481	3,05580209	0,941570184	0,060793	419,5654004	1330,732133
coif	coif5	100	3,267484247	6,309863801	4,314637106	3,042379553	0,941580909	0,06006	871,4356784	2548,012102
db	db2	100	3,267484247	5,766918524	3,568350138	2,499434277	0,938970924	0,058735	3,935258061	3,05359676
db	db3	100	3,267484247	6,090581034	3,96192045	2,823096787	0,940540092	0,058718	3,791597337	3,021336458
db	db4	100	3,267484247	6,134029185	4,058212905	2,866544938	0,941238601	0,059495	7,048851258	10,1571512
db	db5	100	3,267484247	6,213545734	4,182201801	2,946061486	0,941503899	0,059136	36,23211116	113,7101215
db	db6	100	3,267484247	6,313564038	4,245837084	3,046079791	0,942233529	0,059733	140,2699817	649,6376637
db	db7	100	3,267484247	6,154686595	4,155386786	2,887202347	0,941782875	0,05912	180,4009746	623,5135688
db	db8	100	3,267484247	6,331246166	4,296261841	3,063761919	0,941792136	0,060225	307,5368201	713,6717255
sym	sym2	100	3,267484247	5,766918524	3,568350138	2,499434277	0,938970924	0,058735	3,935258061	3,05359676
sym	sym3	100	3,267484247	6,090581034	3,96192045	2,823096787	0,940540092	0,058718	3,791597336	3,021336457
sym	sym4	100	3,267484247	6,14107802	4,068066557	2,873593773	0,940912374	0,058936	10,91171382	28,27143071
sym	sym5	100	3,267484247	6,163885625	4,086636916	2,896401377	0,940965989	0,059401	50,98575665	179,6919737
sym	sym6	100	3,267484247	6,242014634	4,197848911	2,974530387	0,940569529	0,060682	77,29341134	200,884765
sym	sym7	100	3,267484247	6,327206719	4,243735518	3,059722471	0,941796153	0,059452	138,7917755	447,7156277
sym	sym8	100	3,267484247	6,296479674	4,264310088	3,028995426	0,941496887	0,060894	307,5953501	832,8677135

Legenda: family – Família da função wavelet analisada (Daubechies, Symlet ou Coiflet); **wavelet** – Identificação específica da função wavelet utilizada na decomposição do sinal; **n** – Número total de amostras experimentais consideradas para cada wavelet; **segsnr_in_mean** – Valor médio da relação sinal-ruído segmentada (SegSNR) de entrada, antes da aplicação do algoritmo de redução de ruído; **segsnr_out_mean** – Valor médio da relação sinal-ruído segmentada após o processamento de redução de ruído; **segsnr_out_std** – Desvio padrão da SegSNR de saída, indicando a variabilidade dos resultados entre os experimentos; **delta_segsnr_mean** – Ganho médio de SegSNR obtido após o processo de denoising; **stoi_mean** – Valor médio da métrica STOI (*Short-Time Objective Intelligibility*), utilizada para avaliar a inteligibilidade da fala; **stoi_std** – Desvio padrão da métrica STOI; **is_mean** – Valor médio da distância de Itakura–Saito baseada em LPC, utilizada para avaliar a distorção espectral do sinal; **is_std** – Desvio padrão da distância de Itakura–Saito.Fonte: Elaborado pelos autores (2026).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

O Quadro 2 apresenta os resultados do teste de Friedman aplicado às métricas avaliadas.

Quadro 2. Resultados do teste de Friedman para comparação entre wavelets.

metric	friedman_stat	friedman_p	n_samples	n_wavelets
segsnr_out_db	325,7589474	2,36311E-58	100	19
stoi	229,4595789	1,19252E-38	100	19
itakura_saito_lpc	399,4301053	1,20304E-73	100	19

Fonte: Elaborado pelos autores (2026).

Os resultados do teste indicam diferenças estatisticamente significativas entre as wavelets avaliadas para todas as métricas consideradas, uma vez que os valores de p-value são significativamente menores que 0,05. Esse resultado confirma que a escolha da wavelet exerce influência direta no desempenho do algoritmo de redução de ruído.

Além disso, o elevado valor da estatística de Friedman observado para a métrica Itakura–Saito indica que as diferenças entre as wavelets são particularmente relevantes no que se refere à preservação das características espectrais do sinal de voz.

O Quadro 3 apresenta o ranking médio das wavelets considerando todas as métricas avaliadas.

Quadro 3. *Ranking* médio das wavelets de acordo com as métricas de avaliação

metric	best_wavelet	mean_rank
itakura_saito_lpc	sym3	6,12
segsnr_out_db	coif4	7,22
stoi	coif4	7,17

Fonte: Elaborado pelos autores (2026).

A partir da análise do ranking médio, observa-se que a wavelet coif4 apresentou o melhor desempenho médio tanto para a métrica de SegSNR de saída quanto para a métrica de inteligibilidade da fala STOI.

Esse resultado indica que essa wavelet apresentou maior capacidade de atenuação do ruído ao mesmo tempo em que preservou adequadamente as características perceptuais do sinal de voz.

Wavelets da família Coiflet apresentam propriedades matemáticas específicas, como maior número de momentos nulos e maior simetria aproximada, características que favorecem a representação eficiente de sinais suaves e com estrutura multiescala, como é o caso do sinal de voz (Mallat, 2009). Essas propriedades podem explicar o melhor desempenho observado para a wavelet coif4 nas métricas relacionadas à redução de ruído e inteligibilidade.

Por outro lado, a métrica de distância de Itakura–Saito baseada em LPC indicou melhor desempenho para a wavelet sym3, com ranking médio de 6,12. Essa métrica está diretamente

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

relacionada à preservação do envelope espectral da fala, sendo particularmente sensível a distorções introduzidas durante o processo de reconstrução do sinal.

A família Symlet apresenta propriedades de quase simetria e filtros relativamente curtos, o que pode contribuir para menor distorção espectral durante o processo de reconstrução do sinal após a limiarização dos coeficientes wavelet (Daubechies, 1992; Mallat, 2009).

De forma geral, os resultados indicam que diferentes wavelets podem apresentar vantagens dependendo da métrica considerada. Enquanto wavelets da família Coiflet, especialmente a coif4, apresentaram melhor desempenho em termos de redução de ruído e inteligibilidade, wavelets da família Symlet, como a sym3, demonstraram maior capacidade de preservação das características espectrais do sinal de voz.

Portanto, os resultados sugerem que a escolha da wavelet ideal depende do objetivo específico da aplicação. Em cenários em que a prioridade é a melhoria da relação sinal-ruído e da inteligibilidade da fala, a wavelet coif4 mostrou-se mais adequada. Por outro lado, quando a preservação do envelope espectral da fala é o critério principal, a wavelet sym3 pode representar uma escolha mais apropriada.

4. CONSIDERAÇÕES

Este trabalho apresentou uma análise comparativa do desempenho de diferentes famílias de wavelets na redução de ruído em sinais de voz contaminados por ruído colorido, utilizando a técnica de limiarização baseada no estimador rigrsure.

A utilização de múltiplas realizações independentes de ruído para cada sinal de voz permitiu ampliar significativamente o conjunto experimental, resultando em 100 condições de teste distintas. Esse procedimento contribui para maior robustez estatística dos resultados obtidos e aumenta a confiabilidade das conclusões apresentadas.

Os resultados experimentais demonstraram que a utilização da transformada wavelet discreta associada à limiarização rigrsure é capaz de promover melhorias significativas na qualidade do sinal de voz degradado, elevando a SegSNR média de aproximadamente 3,27 dB para valores superiores a 6 dB, dependendo da wavelet utilizada.

A análise das métricas objetivas indicou comportamentos distintos entre as famílias wavelet avaliadas. As wavelets da família Daubechies apresentaram maior ganho médio de SegSNR, destacando-se particularmente a wavelet db8, que apresentou o maior valor médio de SegSNR de saída. Em termos de inteligibilidade da fala, avaliada pela métrica STOI, os resultados indicaram desempenho superior para a wavelet db6, sugerindo que wavelets de ordem intermediária podem oferecer um equilíbrio adequado entre atenuação do ruído e preservação das características perceptuais da fala.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Por outro lado, a análise da distância de Itakura–Saito baseada em LPC indicou que wavelets de menor ordem, como sym3, apresentam maior capacidade de preservação do envelope espectral da fala, refletindo menor distorção espectral após o processo de reconstrução do sinal.

Os resultados do teste estatístico de Friedman confirmaram a existência de diferenças estatisticamente significativas entre as wavelets avaliadas para todas as métricas analisadas, demonstrando que a escolha da função wavelet exerce influência direta no desempenho do algoritmo de redução de ruído.

A análise do ranking médio das wavelets indicou que a wavelet coif4 apresentou o melhor desempenho global nas métricas relacionadas à redução de ruído e inteligibilidade da fala, enquanto a wavelet sym3 apresentou melhor desempenho na preservação das características espectrais do sinal.

De forma geral, os resultados obtidos indicam que não existe uma única wavelet universalmente superior para todas as métricas, sendo a escolha da wavelet dependente do objetivo específico da aplicação. Para aplicações em que a prioridade é a melhoria da relação sinal-ruído e da inteligibilidade da fala, a wavelet coif4 mostrou-se mais adequada. Por outro lado, em aplicações nas quais a preservação do envelope espectral da fala é mais relevante, a wavelet sym3 pode representar uma alternativa mais apropriada.

REFERÊNCIAS

BENESTY, J.; CHEN, J.; HUANG, Y. **Microphone Array Signal Processing**. Berlin: Springer, 2008.

DAUBECHIES, I. **Ten Lectures on Wavelets**. Philadelphia: Society for Industrial and Applied Mathematics (SIAM), 1992. Disponível em: https://epubs.siam.org/doi/book/10.1137/1.9781611970104.

DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. **Journal of Machine Learning Research**, v. 7, p. 1–30, 2006. Disponível em: https://jmlr.org/papers/v7/demsar06a.html. Acesso em: 13 mar. 2026.

DONOHO, D.; JOHNSTONE, I. Adapting to unknown smoothness via wavelet shrinkage. **Journal of the American Statistical Association**, v. 90, n. 432, p. 1200–1224, 1995. Disponível em: https://doi.org/10.1080/01621459.1995.10476626. Acesso em: 13 mar. 2026.

IQBAL, Y. *et al.* **A hybrid speech enhancement technique based on discrete wavelet transform and spectral subtraction**. [*S. l.: s. n.*], 2025. Disponível em: https://ieeexplore.ieee.org/document/10906581. Acesso em: 13 mar. 2026

ITAKURA, F.; SAITO, S. Analysis synthesis telephony based on the maximum likelihood method. *In:* **International Congress on Acoustics**, 1970. Disponível em: https://ieeexplore.ieee.org/document/1164317. Acesso em: 13 mar. 2026.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

LOIZOU, P. **Speech Enhancement:** Theory and Practice. 2. ed. Boca Raton: CRC Press, 2013.

MALLAT, S. **A Wavelet Tour of Signal Processing**: The Sparse Way. 3. ed. New York: Academic Press, 2009.

TAAL, C. *et al.* A short-time objective intelligibility measure for time-frequency weighted noisy speech. **IEEE Transactions on Audio, Speech and Language Processing**, v. 19, n. 7, p. 2125–2136, 2011. Disponível em: https://doi.org/10.1109/TASL.2011.2114881. Acesso em: 13 mar. 2026.

VASEGHI, S. **Advanced Digital Signal Processing and Noise Reduction**. 4. ed. Chichester: Wiley, 2008. Disponível em: https://doi.org/10.1002/9780470740156. Acesso em: 13 mar. 2026.

WU, Z. T. *et al.* Improving speech enhancement models using discrete wavelet transform features. **Electronics**, v. 14, n. 7, 2025. Disponível em: https://www.mdpi.com/2079-9292/14/7/1354. Acesso em: 13 mar. 2026.

YOUSIF, S. T. *et al*. Speech enhancement algorithms: a systematic literature review. **Algorithms**, v. 18, n. 5, 2025. Disponível em: https://www.mdpi.com/1999-4893/18/5/272. Acesso em: 13 mar. 2026.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.