

APLICAÇÃO DA TRANSFORMADA WAVELET NA REDUÇÃO DE RUÍDO EM SINAIS DE VOZ

Leandro Aureliano da Silva¹, Jucelino Cardoso Marciano¹, Edna Lúcia Flôres², Milena Bueno Pereira Carneiro³, Shigueo Nomura³

¹Faculdade de Talentos Humanos – FACTHUS, Uberaba – MG

^{1,2,3}Universidade Federal de Uberlândia, Faculdade de Engenharia Elétrica, Uberlândia – MG

²Universidade Federal de Goiás, Goiânia – Go

lasilva@facthus.edu.br, edna@ufu.br, milena@eletrica.ufu.br, shigueonomura@felt.ufu.br, jucelinocardoso@gmail.com

Resumo - O presente estudo apresenta uma técnica para redução de ruído em sinais de voz que utiliza-se da transformada Wavelet comparando-a com subtração espectral de potência. Para verificação do desempenho deste trabalho, utilizou-se a relação sinal ruído segmentada e a distância de Itakura Saito. Após análise dos resultados extraídos, verificou-se que a técnica, usando a transformada Wavelet, apresentou menores distorções espectrais e em alguns casos, a melhor relação sinal ruído.

Palavras-Chave - Transformada Wavelet, Supressão de Ruído, Subtração Espectral e Sinais de Voz.

APPLICATION OF WAVELET TRANSFORM IN NOISE REDUCTION IN SPEECH SIGNALS

Abstract - This study presents a technique for noise reduction in speech signals adapting the wavelet transform by comparing it with the power spectral subtraction. To validate the results, segmented signal noise and Itakura Saito distance relationship were used. After analyzing the obtained results, we have verified that the technique based on Wavelet transform showed lower spectral distortion and, in some cases, a better signal to noise ratio.

Keywords – Wavelet Transform, Noise Reduction, Spectral Subtraction and Speech Signals.

I. INTRODUÇÃO

Em muitas situações envolvendo a transmissão de voz e/ou áudio, a presença de ruído aditivo pode degradar a qualidade e a inteligibilidade dos sinais. Muitas pesquisas estão sendo realizadas nessa área, e, conseqüentemente, técnicas de melhoria no processamento de voz têm surgido com o objetivo de eliminar ou pelo menos reduzir a intensidade do ruído aditivo. Algumas técnicas se destacam, dentre elas: subtração espectral [1] e [2], filtros de Wiener

[2], filtros adaptativos [2] e [3], redes neurais [4] e [5] dentre outras. Contudo, este artigo apresenta o uso da transformada Wavelet [6] para redução de ruído, comparando-a com a subtração espectral de potência. As comparações foram realizadas tendo como parâmetros: distância de Itakura Saito e a relação sinal ruído segmentada.

Este artigo foi dividido em quatro partes: a introdução (seção I), descrita acima; a descrição dos algoritmos na seção II, demonstrando o funcionamento de ambos os algoritmos na redução de ruído nos sinais de voz. Já a seção III responsável pela apresentação dos resultados, comprovaram a eficiência das técnicas propostas e por fim na seção IV, tem-se a conclusão deste trabalho.

II. DESCRIÇÃO DOS ALGORITMOS

A - Algoritmo de Redução de Ruído usando Wavelet

Seja $f(t)$ um sinal de tempo contínuo. A transformada Wavelet deste é definida [7]:

$$Wf(a, b) = \int_{-\infty}^{\infty} f(t) \psi_{a,b}(t) dt \quad (1)$$

Para um sinal discreto de N pontos, a integral acima pode ser aproximada por um somatório, tal que:

$$Wf(a, b) = \sum_{t=0}^{N-1} f(t) \psi_{a,b}(t) \quad (2)$$

A função $\psi_{a,b}(t)$, chamada Wavelet, é derivada de uma função $\psi(t)$ através da seguinte transformação:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (3)$$

em que: “b” representa posição ou translação da Wavelet e “a” chamado parâmetro de escala, o qual está associado à largura da janela [7].

Há uma ampla gama de escolha para função $\psi(t)$, denominada “Wavelet mãe”, dentre elas: Daubechies, symlets, coiflet, entre outras [6].

Seja o sinal de voz $y(n)$ contaminado por um ruído aditivo de tal forma que:

$$y(n) = x(n) + v(n) \quad (4)$$



XII CEEL – ISSN 2178-8308
13 a 17 de Outubro de 2014
Universidade Federal de Uberlândia – UFU
Uberlândia – Minas Gerais – Brasil

em que: $x(n)$ é o sinal de voz sem ruído e $v(n)$ é o ruído aditivo gaussiano.

O princípio básico da redução de ruído no sinal $y(n)$ utilizando transformada Wavelet consiste em realizar decomposições no sinal original em coeficientes de aproximação e detalhe, gerando a árvore de decomposição. Nesse artigo, usou-se a decomposição até o nível ($m = 3$). Os coeficientes de aproximação (A_m) trazem as informações de baixa frequência associadas a Wavelet adotada, no caso *coiflet* suporte (6). Já os de detalhe (D_m) trazem as informações de alta frequência. Assim, a ideia básica é escolher quais coeficientes serão mantidos para preservar a informação e, em quais coeficientes serão aplicados um limiar, cujo objetivo é eliminar ou reduzir a intensidade do ruído.

Um dos limiares utilizados na literatura é o *Hard Thresholding*, que consiste em substituir os coeficientes menores que o limiar por zero. Neste artigo, a partir de vários testes, adotou-se o seguinte limiar [6]:

$$D_1(n) = \begin{cases} D_1(n) & , \text{ se } D_1(n) \geq 0,3 \times \text{Máx}(D_1(n)) \\ 0 & , \text{ caso contrário} \end{cases} \quad (5)$$

Após o processo de corte, a transformada de Wavelet inversa, é utilizada para obter o sinal livre de ruído após o processamento.

B - Subtração Espectral de Potência (SEP)

Dado $|Y(e^{j\omega_k})|^2$ como o espectro de potência do sinal contaminado por ruído, $|\mu(e^{j\omega_k})|^2$ como a média do espectro de potência do ruído avaliado em trechos de silêncio [1, 2], a subtração espectral de potência é dada por:

$$|\hat{s}(e^{j\omega_k})|^2 = |Y(e^{j\omega_k})|^2 - \alpha |\mu(e^{j\omega_k})|^2 \quad (6)$$

em que: $|\hat{s}(e^{j\omega_k})|^2$ é uma estimativa do espectro de potência do sinal sem ruído. Já o parâmetro α controla o quanto de ruído que é subtraído do sinal contaminado.

Devido à natureza aleatória do ruído, a subtração espectral pode gerar valores negativos, que vão decrementar a relação sinal ruído (SNR). Para suplantar este problema é usada uma retificação dada em [1].

$$|\hat{x}(e^{j\omega_k})|^2 = \begin{cases} |\hat{s}(e^{j\omega_k})|^2 & , \text{ se } |\hat{s}(e^{j\omega_k})|^2 > \beta |\mu(e^{j\omega_k})|^2 \\ \beta |\mu(e^{j\omega_k})|^2 & , \text{ caso contrário} \end{cases} \quad (7)$$

em que: $0 \leq \beta \ll 1$ é o limite espectral mínimo. Já o parâmetro α depende da SNR de entrada e pode ser calculado como [1]:

$$\alpha = \alpha_o - \frac{3}{20} \text{SNR} \quad -5\text{dB} \leq \text{SNR} \leq 20 \quad (8)$$

em que: α_o tem seu valor igual a 4 [1].

Uma vez processado, o sinal estimado no domínio do tempo é obtido, utilizando-se a *IDFT* juntamente com a fase do sinal contaminado por ruído, conforme mostrado abaixo:

$$\hat{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \left| \hat{X}(e^{j\omega_k}) \right| e^{j\theta_Y(e^{j\omega_k})} e^{-j\omega_k n} \quad (9)$$

em que: $\omega_k = \frac{2\pi}{N}k$ é a frequência discreta da transformada e

$\theta_Y(e^{j\omega_k})$ é a fase do sinal contaminado por ruído.

III. RESULTADOS

Para a avaliação dos algoritmos apresentados nas seções anteriores, foram utilizados diferentes sinais de voz, amostrados numa taxa de 22050 Hz com 16 bits. Esses sinais foram contaminados por ruído branco. A janela aplicada no algoritmo de Subtração Espectral de Potência é a de Hamming com 512 amostras e com sobreposição de 50%.

A Wavelet utilizada foi a *Coiflet* suporte (6), sendo ambos os algoritmos desenvolvidos no ambiente Matlab R2013B.

Partindo de observações realizadas nos coeficientes de aproximação e detalhe, nota-se que o coeficiente de detalhe 1 (D_1) é o que possui maior intensidade de ruído. Dessa forma, o limiar descrito na Equação (5), será aplicado diretamente nesse coeficiente.

Para avaliar e comparar os resultados, dos algoritmos descritos acima, foram utilizadas as medidas de relação sinal ruído segmentada (SNRseg) e a distância de Itakura-Saito ($d(a,b)$).

A SNRseg é uma medida mais efetiva que pode ser calculada em segmentos curtos de sinal de voz de forma a equilibrar os pesos atribuídos aos trechos de maior e menor intensidades do sinal. Essa medida é calculada, utilizando-se a Equação 10 a seguir [9]:

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

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

A SNRseg não fornece uma medida significativa de desempenho quando dois sinais diferem em seus espectros. Entretanto, as medidas de distância são sensíveis às variações de espectro. Nesse caso, a distância de Itakura Saito fornece melhores resultados e pode ser calculada utilizando os parâmetros de predição linear (LPC) conforme abaixo [8].

$$d(a,b) = \log \left[\frac{aR a^T}{bR b^T} \right] \quad (11)$$

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.

No primeiro teste, o sinal de voz foi contaminado com ruído branco, obtendo uma relação sinal ruído de entrada (SNRI) de 3 dB. Neste, a técnica utilizada foi a de Wavelet.

O sinal sem ruído, o sinal contaminado e o sinal após o processamento são mostrados respectivamente nas Figuras 1, 2 e 3.

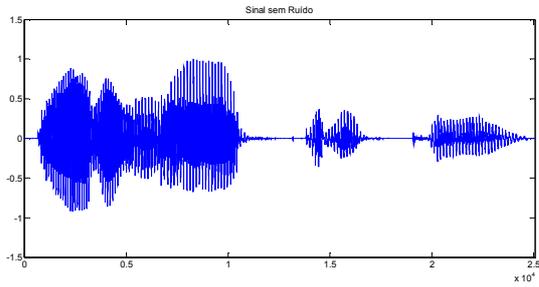


Fig. 1. Sinal sem ruído da palavra elétrica

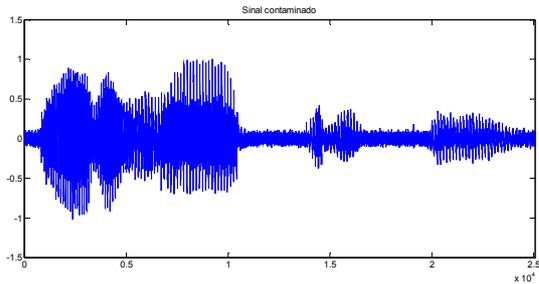


Fig. 2. Sinal contaminado por ruído branco

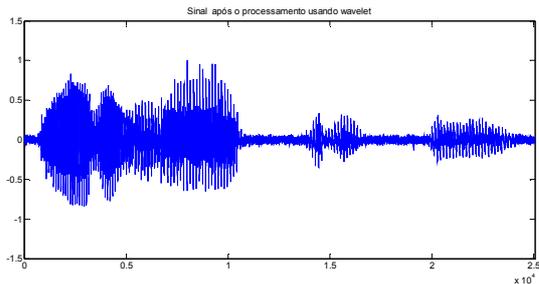


Fig. 3. Sinal após o processamento usando Wavelet na redução do ruído.

Analisando a Figura 3 em relação à Figura 2, percebe-se uma considerável redução de ruído principalmente em instantes de silêncio. Esse processamento resultou em uma relação sinal ruído segmentada de saída (SNRO) de 5 dB e uma distorção espectral medida pela distância de Itakura Saito de 0,3495.

Outro parâmetro utilizado também para avaliar os resultados oriundos do processamento é o espectrograma. As Figuras 4, 5 e 6 correspondem aos espectrogramas do sinal sem ruído, sinal contaminado e sinal estimado.

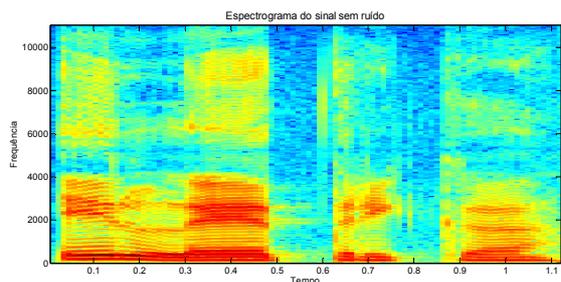


Fig. 4. Espectrograma do sinal sem ruído da palavra elétrica.

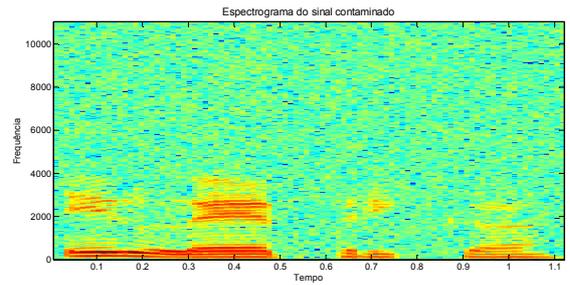


Fig. 5. Espectrograma do sinal contaminado com ruído branco.

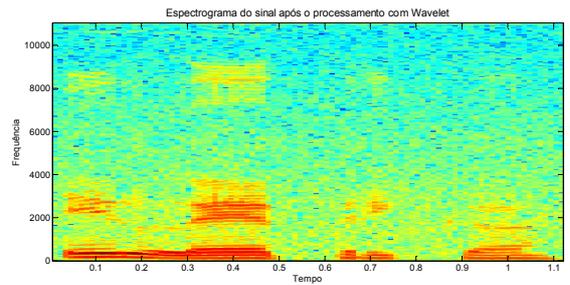


Fig. 6. Espectrograma do sinal após o processamento usando Wavelet.

Conforme análise dos três espectrogramas, percebe-se na Figura 6 que nas frequências entre 2000 e 3000 Hz, que a intensidade do sinal foi ressaltada pelo algoritmo de Wavelet se comparado com o espectrograma da Figura 5. Pode-se perceber também que embora haja uma redução da intensidade do ruído, um ruído residual permanece.

No segundo procedimento, usou-se a subtração espectral de potência, onde, o sinal de voz, mais uma vez, foi contaminado por ruído gaussiano, obtendo uma SNRI de 3 dB.

Os resultados do processamento podem ser visualizados nas Figuras 7,8 e 9.

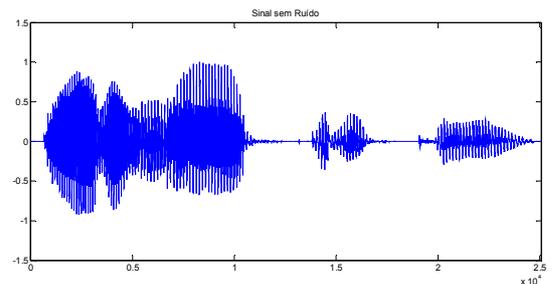


Fig. 7. Sinal sem ruído da palavra elétrica

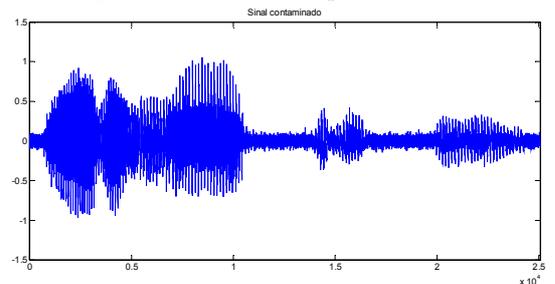


Fig. 8. Sinal contaminado por ruído branco.

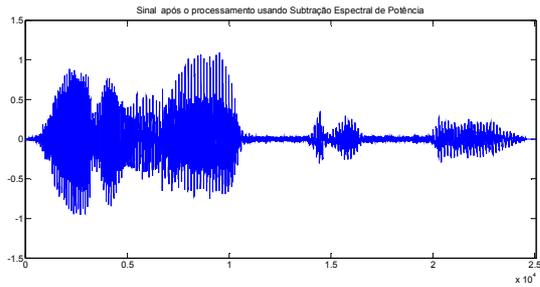


Fig. 9. Sinal após o processamento usando Subtração Espectral de Potência na redução do ruído.

Percebe-se na Figura 9 em relação à Figura 8, uma considerável redução de ruído principalmente em instantes de silêncio. Esse processamento resultou em SNRO de 9 dB e uma distorção espectral medida pela distância de Itakura Saito de 0,3780, um pouco maior em relação a técnica que utiliza-se Wavelet.

Apesar desse algoritmo apresentar uma redução de ruído considerável, após o processamento, surge um fenômeno conhecido como ruído musical causado pela subtração espectral, sendo sua eliminação quase impossível.

Os espectrogramas para a aplicação da Subtração Espectral podem ser visualizados nas Figuras 10, 11 e 12.

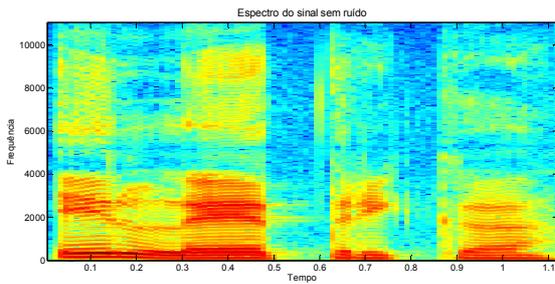


Fig. 10. Espectrograma do sinal sem contaminação.

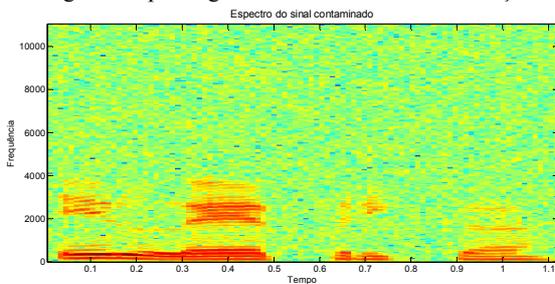


Fig. 11. Espectrograma do sinal contaminado.

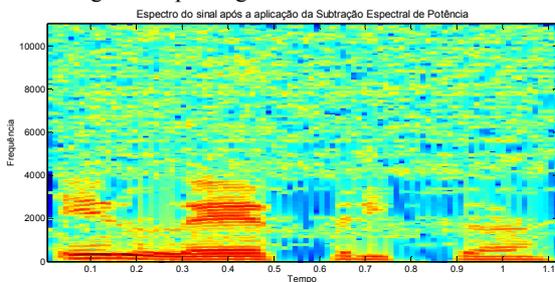


Fig. 12. Espectrograma do processamento usando Subtração Espectral de Potência na redução do ruído.

Para verificar qual das técnicas apresentadas acima provocam menor distorção na reconstrução do sinal, é

mostrado na Figura 13, um gráfico que relaciona a distância de Itakura Saito com relação à SNRI para 0dB, 3dB e 6dB, usando como teste a palavra “elétrica”.

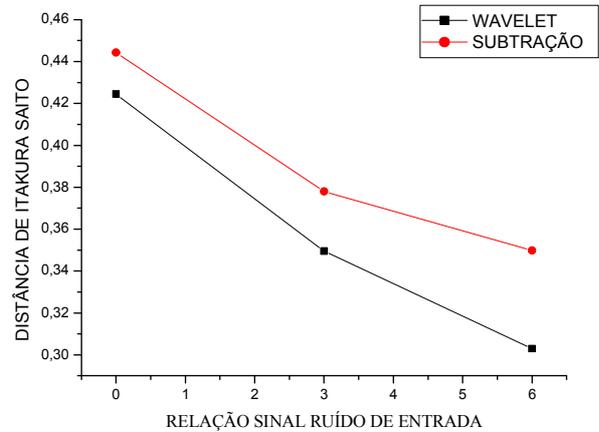


Fig. 13. Distância de Itakura Saito em função da relação sinal ruído de entrada (SNRI).

Analisando-se o gráfico da Figura 13, nota-se claramente que a técnica utilizando Wavelet, foi a que apresentou a menor distorção espectral.

Continuando as simulações, verifica-se também que a subtração espectral de potência é a técnica que apresenta melhores resultados referentes a SNRO x SNRI. A Figura 14 ilustra os resultados para 0dB, 3dB e 6dB, tomando a palavra “elétrica” como teste.

Cabe salientar também que em todas as simulações utilizando Subtração Espectral de Potência, verifica-se a presença do ruído musical, o que não acontece com Wavelet.

Para comprovar a eficiência de Wavelet com relação à distorção espectral, a Figura 15 mostra o resultado médio de 5 palavras, relacionando distância de Itakura Saito com relação SNRI de 0, 3 e 6 dB.

Para validar os resultados SNRO x SNRI, a Figura 16 apresenta os resultados médios para 5 palavras, tomando como parâmetro as SNRI's de 0, 3 e 6dB. As comparações são realizadas usando as técnicas de Wavelet e subtração espectral de potência.

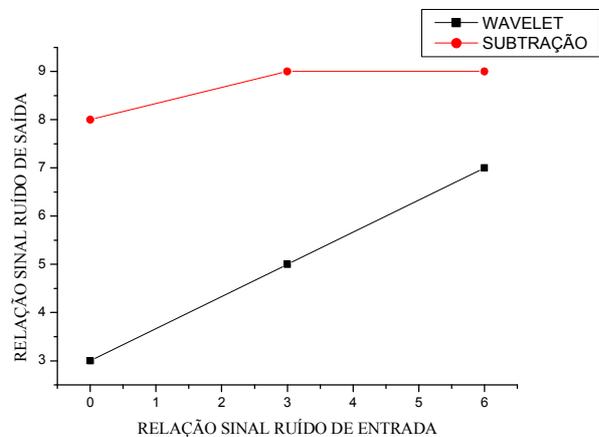


Fig. 14. Relação Sinal Ruído de Saída x Relação Sinal Ruído de Entrada.

REFERÊNCIAS

- [1] M. Berouti, R. Schwartz, J. Makhoul, "Enhancement of speech corrupted by acoustic noise", *Proc. IEEE Int. conf. Acoust., Speech, Signal Process.*, pp. 208-211, Apr 1979.
- [2] S. V. Vaseghi, *Advance Digital Signal Processing and Noise Reduction*, John Wiley & Sons, Nova Iorque, 2000.
- [3] R. G. Brown, P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons, Nova Iorque, 1997.
- [4] K. Daqrouq, I. N. Abu-Isbeih, M. Alfauri, "Speech signal enhancement using neural network and wavelet transform", *Proceedings of the 6th International Multi-Conference on Systems, Signals and Devices*, Djerba, Tunisia, pp. 1-6, 2009.
- [5] M. V. Ishwarya, "An Improved Online Tamil Character Recognition Using Neural Networks", *International Journal of Advanced Science and Technology (IJAST)*, vol. 42, pp. 1-10. May 2012.
- [6] G. Strang, T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Massachusetts, 1996.
- [7] Misiti, M, Misiti, Y, Oppenheim, G, Poggi, J. M, *Wavelet Toolbox User'S Guide*, The Mathworks, Natick, 1996.
- [8] L. R. Rabiner, R. W. Schafer, *Digital processing of speech signals*, Prentice- Hall, Nova Jersey, 1978.
- [9] J. R. Deller, J. G. Proakis; J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Prentice Hall, Nova Jersey, 1993.

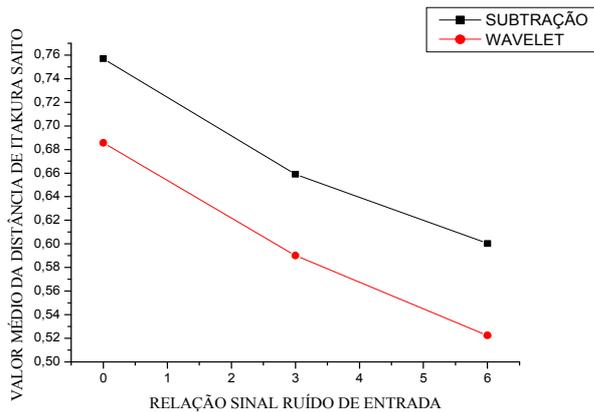


Fig. 15. Distância de Itakura Saito em função da relação sinal ruído de entrada (SNRI), para 5 palavras.

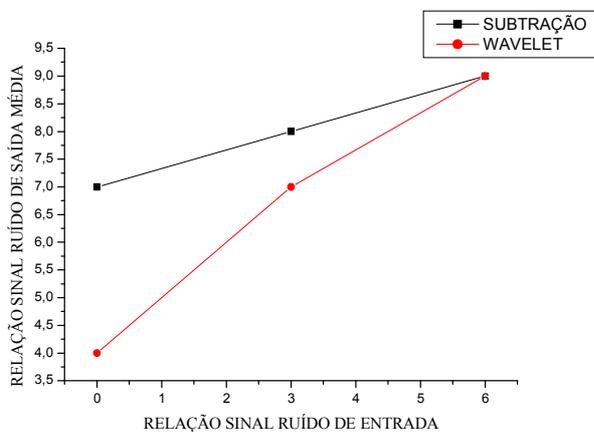


Fig. 16. Relação Sinal Ruído de Saída Média para 5 palavras x Relação Sinal Ruído de Entrada.

A partir da Figura 15, verifica-se que a técnica de Wavelet comparada com a subtração espectral de potência, apresenta melhores resultados para a distorção espectral.

Para a relação sinal ruído de saída, os resultados ficam próximos para relações sinal ruído de entrada em torno de 6 dB conforme ilustra a Figura 16.

IV. CONCLUSÃO

Este artigo mostrou o estudo de duas técnicas para a redução de ruído em sinais de voz, a transformada Wavelet e subtração espectral de potência. Para a comprovação da eficiência destas técnicas utilizaram-se as medidas de distância de Itakura Saito e a relação sinal ruído segmentada. Após testes realizados, conclui-se que a técnica de Wavelet apresenta melhores resultados em relação a distorção espectral e a subtração espectral de potência referente à relação sinal ruído de saída para valores de 0dB e 3dB. Já para valores de relação sinal ruído de entrada em torno de 6dB, os resultados da relação sinal ruído de saída ficam próximos.