ISSN: 2230-9926

Available online at http://www.journalijdr.com

IJDR

International Journal of Development Research
Vol. 12, Issue, 06, pp. 57109-57113, June, 2022
https://doi.org/10.37118/ijdr.24870.06.2022

RESEARCH ARTICLE OPEN ACCESS

TRANSFORMED WAVELET APPLIED IN VOICE SIGNALS FOR NOISE REDUCTION

Leandro Aureliano da Silva[1], Felipe Santos Moreira[2], Antonio Carlos Lemos Júnior[3], Cleiton Silvano Goulart[4], Lidiana Mendes Sousa[5], Viviane Beatriz do Nascimento[6], RogérioBernardes Andrade[7] and Gilberto Felipe Fernandes[8]

[1]Doutor em Engenharia Elétrica, Faculdade de Talentos Humanos – FACTHUS. [2]Doutor em Engenharia Química, Faculdade de Talentos Humanos – FACTHUS. [3]Mestre emInovação Tecnológica, Faculdade de Talentos Humanos – FACTHUS. [4]Mestre em Física, Faculdade de Talentos Humanos – FACTHUS. [5]Mestre em Engenharia Elétrica, Faculdade de Talentos Humanos – FACTHUS. [6]Especialista emEngenharia de Segurança do Trabalho, Faculdade de Talentos Humanos – FACTHUS. [7]Mestre em Inovação Tecnológica, Faculdade de Talentos Humanos – FACTHUS. [8]Mestre em Engenharia Mecânica. Instituição: Faculdade de Talentos Humanos – FACTHUS

ARTICLE INFO

Article History:
Received 11th March, 2022
Received in revised form
18th April, 2022
Accepted 07th May, 2022
Published online 28th June, 2022

Key Words:
Wavelet Transform, Noise Reduction, Spectral Subtraction and Speech Signals.

Corresponding author:
Leandro Aureliano da Silva

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.

Citation: ***Leandro Aureliano da Silva, Felipe Santos Moreira, Antonio Carlos Lemos Júnior, Cleiton Silvano Goulart et al.*** "Transformed Wavelet Applied in Voice Signals for Noise Reduction*", International Journal of Development Research*, 12, (06), 57109-57113.

INTRODUCTION

In many situations involving voice and/or audio transmission, the presence of additive noise can degrade the quality and intelligibility of signals. Much research is being carried out in this area, and, consequently, techniques to improve voice processing have emerged with the aim of eliminating or at least reducing the intensity of additive noise. Some techniques stand out, among them: spectral subtraction (Berouti, 1979; Vaseghi, 2000), Wiener filters (Vaseghi, 2000), adaptive filters (Vaseghi, 2000; Brown, 1997), neural networks (Daqrouq, 2009; Ishwarya, 2012) among others. However, this article presents the use of the Wavelet transform (6for noise reduction, comparing it with the power spectral subtraction. Comparisons were performed having as parameters: distance from Itakura Saito and the segmented signal-to-noise ratio. This article has been divided into four parts: the introduction (section I), described above; the materials and methods in section II, demonstrating the operation of both algorithms in reducing noise in voice signals. The section III, responsible for the presentation of the results, proved the efficiency of the proposed techniques and finally, in section IV, there is the conclusion of this work.

MATERIALS AND METHODS

Noise Reduction Algorithm using Wavelet: Let f(t) be a continuous time signal. The Wavelet transform of this is defined by Equation 1 (Misiti, 1996):

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

For a discrete signal of N points, the above integral can be approximated by a summation, such that:

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

The function $\psi_{a,b}(t)$, Wavelet, is derived from a function $\psi(t)$ through the following transformation:

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

wherein: "b" represents position or translation of the Wavelet and "a" called scale parameter, which is associated with the window width (Misiti, 1996). There is a wide range of choice for function $\psi(t)$, called "mother Wavelet", amongst them: Daubechies, symlets, coiflet, among other (Strang, 1996). Let the voice signal *y(n)* be contaminated by an additive noise such that:

$$y(n)=x(n)+v(n) \tag{4}$$

wherein: *x(n)* is the noiseless speech signal and *v(n)* is the Gaussian additive noise. The basic principle of noise reduction in the *y(n)* signal using Wavelet transform consists of performing decompositions on the original signal into approximation and detail coefficients, generating the decomposition tree. In this article, decomposition was used up to the level (m = 3). The approximation coefficients (A_m) bring the low frequency information associated with the adopted Wavelet, in the case of coiflet support (Strang, 1996). The detail ones (D_m) bring high frequency information. Thus, the basic idea is to choose which coefficients will be kept to preserve the information and, in which coefficients a threshold will be applied, whose objective is to eliminate or reduce the noise intensity. One of the thresholds used in the literature is Hard Thresholding, which consists of replacing coefficients smaller than the threshold by zero. In this article, from several tests, the following threshold was adopted (Strang, 1996):

$$D_1(n)=\begin{cases} D_1(n) & ,if\ D_1(n)\ \geq 0{,}3\ Máx\left(D_1(n)\right) \\ 0 & ,otherwise \end{cases} \tag{5}$$

After the cutting process, the inverse Wavelet transform is used to obtain the noise-free signal after processing.

Power Spectral Subtraction: Given $\left|Y\left(e^{j\omega_k}\right)\right|^2$ as the signal power spectrum contaminated by noise, $\left|\mu\left(e^{j\omega_k}\right)\right|^2$ as the average of the noise power spectrum evaluated in silence stretches (1, 2), the power spectral subtraction is given by:

$$\left|\hat{S}\left(e^{j\omega_k}\right)\right|^2=\left|Y\left(e^{j\omega_k}\right)\right|^2-\alpha\left|\mu\left(e^{j\omega_k}\right)\right|^2 \tag{6}$$

wherein: $\left|\hat{s}\left(e^{j\omega_k}\right)\right|^2$ is an estimate of the noiseless signal power spectrum. The parameter α controls how much noise is subtracted from the contaminated signal. Due to the random nature of noise, spectral subtraction can generate negative values, which will decrease the signal-to-noise ratio (SNR). To overcome this problem, a rectification given in (1).

$$\left|\hat{X}(e^{j\omega_k})\right|^2=\begin{cases} \left|\hat{X}(e^{j\omega_k})\right|^2, & if\ \left|\hat{X}(e^{j\omega_k})\right|^2>\beta\left|\mu(e^{j\omega_k})\right|^2 \\ \beta\left|\mu(e^{j\omega_k})\right|^2, & otherwise \end{cases} \tag{7}$$

wherein: $0\leq\beta<<1$ is the minimum spectral limit. The parameter α depends on the input SNR and can be calculated as (1):

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

wherein: α_0 has its value equal to 4 (1).

Once processed, the estimated time-domain signal is obtained, using the IDFT together with the phase of the noise contaminated signal, as shown below:

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

wherein: $\omega_k=\frac{2\pi}{N}k$ is the discrete frequency of the transform and $\theta_Y\left(e^{j\omega_k}\right)$ is the phase of the signal contaminated by noise.

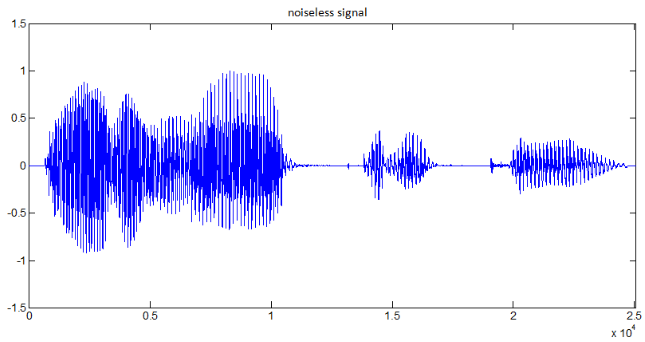

Figure 1. Electric word noiseless signal.

RESULTS AND DISCUSSION

For the evaluation of the algorithms presented in the previous sections, different voice signals were used, sampled at a rate of 22050 Hz with 16 bits. These signals were contaminated by white noise. The window applied in the Power Spectral Subtraction algorithm is the Hamming window with 512 samples and 50% overlap. The Wavelet used was the Coiflet support (Strang, 1996), being both algorithms developed in the environment Matlab R2013B. Based on observations made on the approximation and detail coefficients, it is noted that the detail coefficient 1 (D1) is the one with the highest noise intensity. Thus, the threshold described in Equation (5) will be applied directly to this coefficient. To evaluate and compare the results of the algorithms described above, the measures of segmented signal-to-noise ratio (SNRseg) and the Itakura-Saito distance were used (d(a,b)). The SNRseg is a more effective measure that can be calculated in short segments of the voice signal in order to balance the weights assigned to the segments of higher and lower signal strengths. This measure is calculated using Equation 10 (Deller, 1993):

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

wherein: mj represents the boundaries of each of the M frames of size N.

SNRseg does not provide a significant measure of performance when two signals differ in their spectra. However, distance measurements are sensitive to spectrum variations. In this case, the Itakura Saito distance gives better results and can be calculated using linear prediction parameters (LPC) (Rabinner, 1978).

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

wherein: "a" is the vector of LPC coefficients of the original signal; "R" is the autocorrelation matrix of the original signal and "b" is the vector of LPC coefficients of the estimated signal. In the first test, the voice signal was contaminated with white noise, obtaining a signal to noise input ratio (SNRI) of 3 dB. In this, the technique used was the Wavelet. The noiseless signal, the contaminated signal and the signal after processing are shown in Figures 1, 2 and 3 respectively.

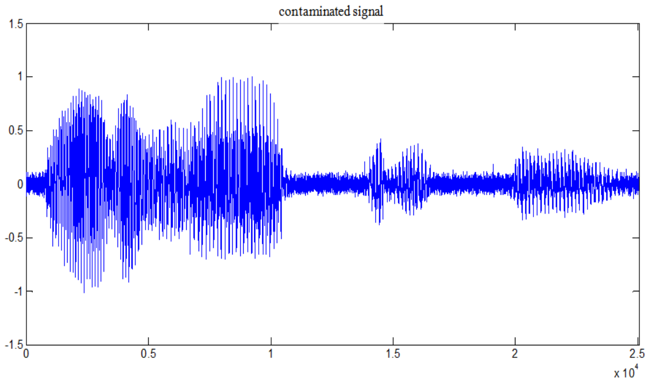

Figure 2. Signal contaminated by white noise

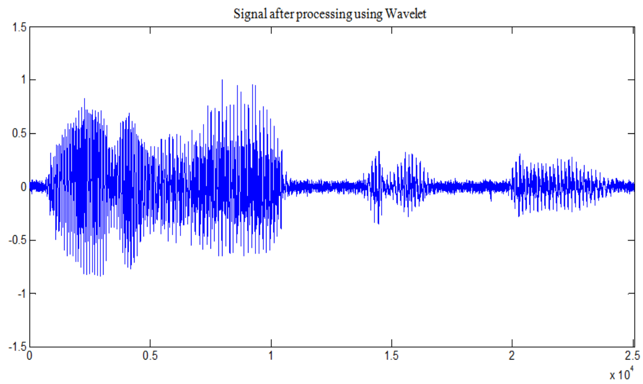

Figure 3. Signal after processing using Wavelet in noise reduction

Analyzing Figure 3 in relation to Figure 2, we can see a considerable reduction in noise, especially in moments of silence. This processing resulted in a segmented output signal-to-noise ratio (SNRO) of 5 dB and a spectral distortion measured by the Itakura Saito distance of 0.3495. Another parameter also used to evaluate the results from the processing is the spectrogram.

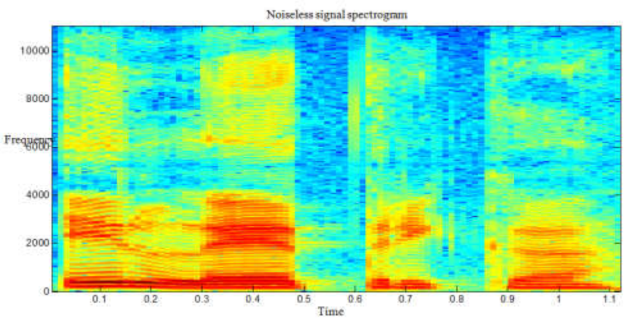

Figure 4. Electric word noiseless signal spectrogram

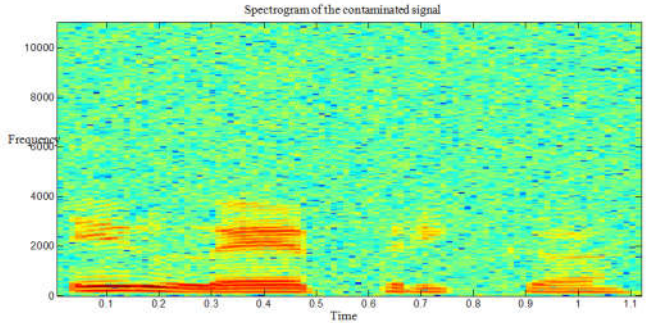

Figure 5. Signal spectrum contaminated with white noise

Figures 4, 5 and 6 correspond to the spectrograms of the noiseless signal, contaminated signal and estimated signal. According to the analysis of the three spectrograms, it can be seen in Figure 6 that in the frequencies between 2000 and 3000 Hz, the signal strength was highlighted by the Wavelet algorithm compared to the spectrogram of Figure 5. It can also be seen that although there is a reduction in noise intensity, residual noise remains.

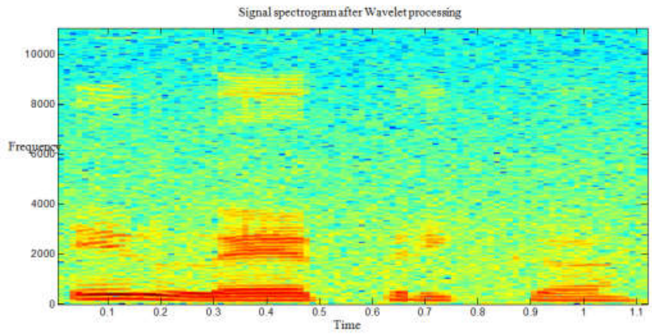

Figure 6. Signal spectrogram after processing using Wavelet

In the second procedure, power spectral subtraction was used, where the voice signal, once again, was contaminated by Gaussian noise, obtaining an SNRI of 3 dB. The processing results can be seen in Figures 7, 8 and 9.

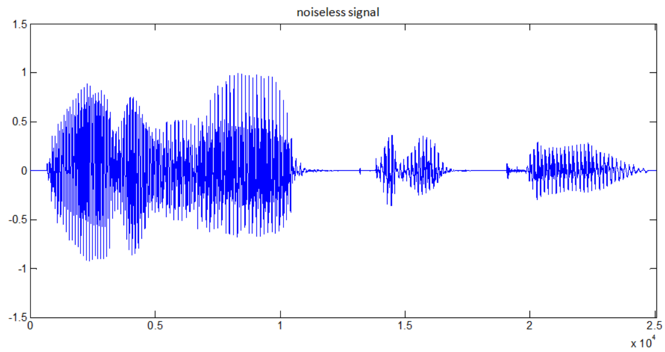

Figure 7. Electric word noiseless signal

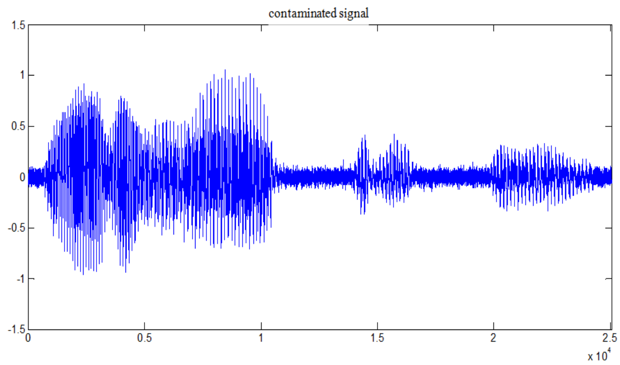

Figure 8. Signal contaminated by white noise

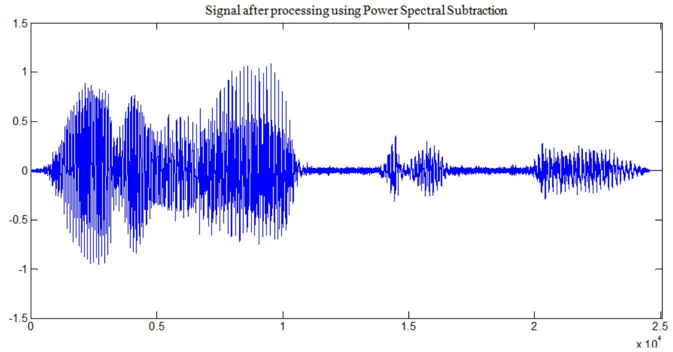

Figure 9. Signal after processing using Spectral Power Subtraction in noise reduction

It can be seen in Figure 9 in relation to Figure 8, a considerable reduction of noise mainly in moments of silence. This processing resulted in an SNRO of 9 dB and a spectral distortion measured by the Itakura Saito distance of 0.3780, a little higher in relation to the technique that uses Wavelet. Although this algorithm presents considerable noise reduction, after processing, a phenomenon known as musical noise caused by spectral subtraction appears, and its elimination is almost impossible. The spectrograms for the application of Spectral Subtraction can be seen in Figures 10, 11 and 12. To verify which of the techniques presented above causes less distortion in the signal reconstruction, Figure 13 shows a graph that relates the distance of Itakura Saito in relation to the SNRI for 0dB, 3dB and 6dB, using the word "electrical" as a test.

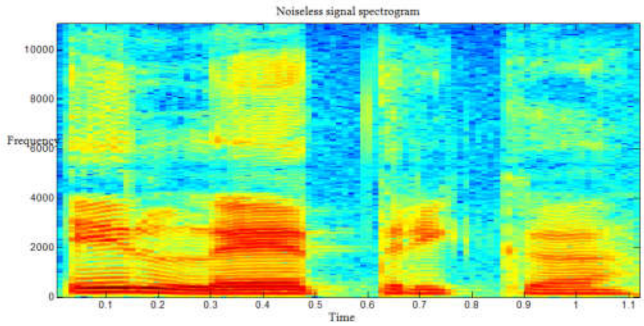

Figure 10. Signal spectrogram without contamination

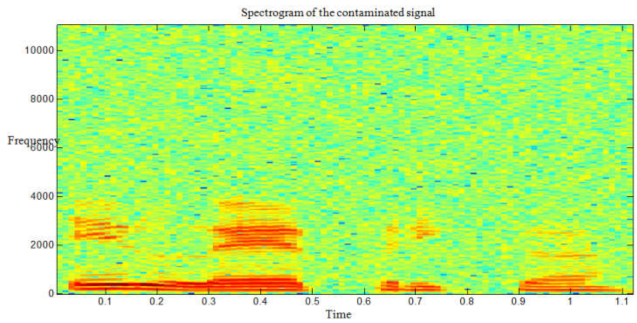

Figure 11. Spectrogram of the contaminated signal

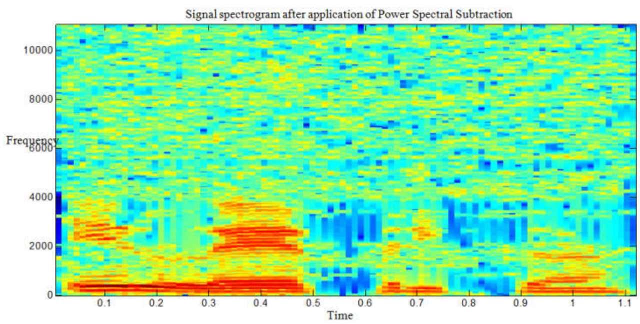

Figure 12. Processing spectrogram using Power Spectral Subtraction in noise reduction

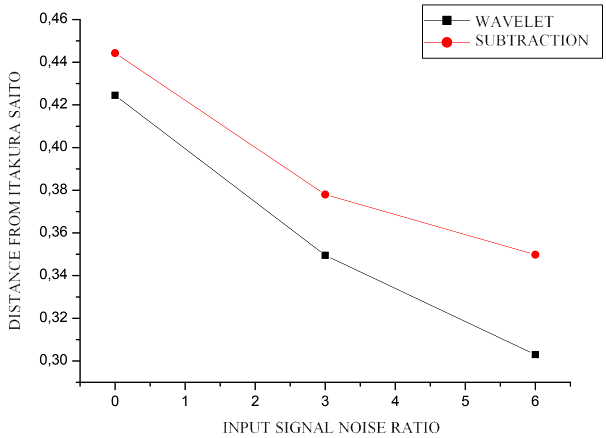

Figure 13. Distance from Itakura Saito as a function of the input signal-to-noise ratio (SNRI)

Analyzing the graph in Figure 13, it is clear that the technique using Wavelet was the one that presented the lowest spectral distortion. Continuing the simulations, it is also verified that the power spectral subtraction is the technique that presents the best results regarding SNRO x SNRI. Figure 14 illustrates the results for 0dB, 3dB and 6dB, taking the word "electrical" as a test. It should also be noted that in all simulations using Spectral Power Subtraction, the presence of musical noise is verified, which does not happen with Wavelet. To prove the Wavelet efficiency with respect to spectral distortion, Figure 15 shows the average result of 5 words, relating Itakura Saito distance with SNRI ratio of 0dB, 3dB and 6dB. To validate the SNRO x SNRI results, Figure 16 presents the average results for 5 words, taking the SNRI's of 0dB, 3dB and 6dB as a parameter. Comparisons are performed using Wavelet and Power Spectral Subtraction techniques.

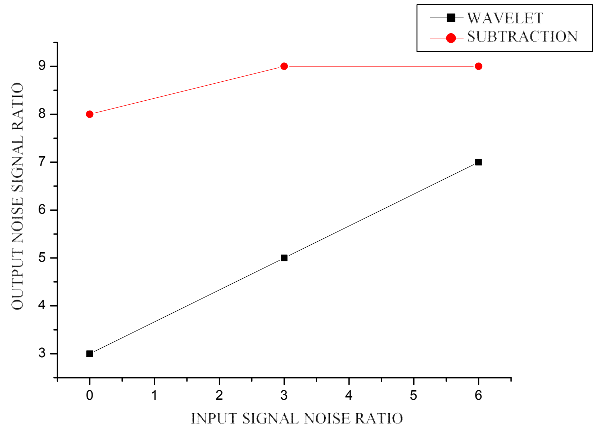

Figure 14. Output Signal Noise Ratio x Input Signal Noise Ratio.

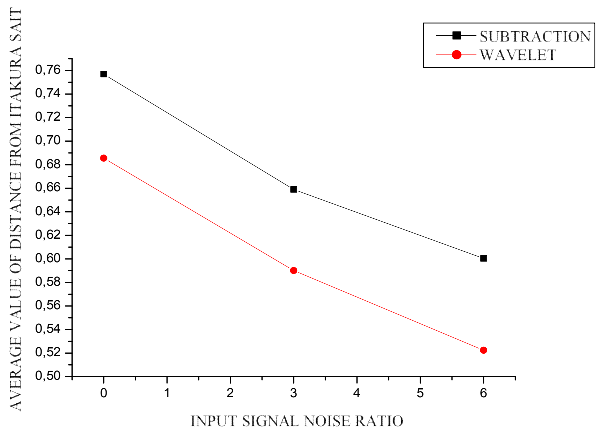

Figure 15. Distance from Itakura Saito as a function of the input signal-to-noise ratio (SNRI), for 5 words.

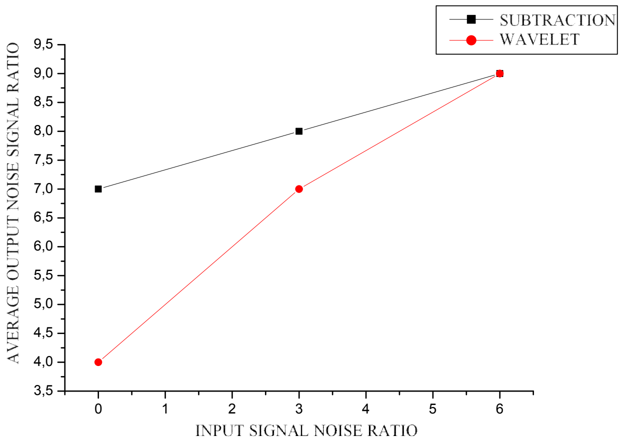

Figure 16. Average Output Signal Noise Ratio for 5 words x Input Signal Noise Ratio

From Figure 15, it can be seen that the Wavelet technique, compared with the spectral power subtraction, presents better results for the spectral distortion. For the output signal-to-noise ratio, the results are close to input signal-to-noise ratios around 6 dB as shown in Figure 16.

CONCLUSION

This article showed the study of two techniques for noise reduction in voice signals, the Wavelet transform and spectral power subtraction. To prove the efficiency of these techniques, Itakura Saito distance measurements and the segmented signal-noise ratio were used.

After tests carried out, it is concluded that the Wavelet technique presents better results in relation to spectral distortion and spectral power subtraction referring to the output signal to noise ratio for values of 0dB and 3dB. For values of input signal-to-noise ratio around 6dB, the output signal-to-noise ratio results are close.

REFERENCES

Berouti, M. R. Schwartz, J. Makhoul, “Enhancement of speech corrupted by acoustic noise”, Proc. IEEE Int. conf. Acoust., Speech, Signal Process., pp. 208-211, Apr 1979.

Vaseghi, S. V. Advance Digital Signal Processing and Noise Reduction, John Wiley & Sons, Nova Iorque, 2000.

Brown, R. G. P. Y. C. Hwang, Introduction to Random Signals and Applied Kalman Filtering, John Wiley & Sons, Nova Iorque, 1997.

Daqrouq, K. 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.

Ishwarya, M. V. “An Improved Online Tamil Character Recognition Using Neural Networks”, International Journal of Advanced Science and Technology (IJAST), vol. 42, pp. 1-10. May 2012.

Strang, G. T. Nguyen, Wavelets and Filter Banks, Wellesley-Cambridge Press, Massachusetts, 1996.

Misiti, M, Misiti, Y, Oppenheim, G, Poggi, J. M, Wavelet Toolbox User’S Guide, The Mathworks, Natick, 1996.

Rabinner, L. R. R. W.Schafer, Digital processing of speech signals, Prentice- Hall, Nova Jersey, 1978.

Deller, J. R. J. G. Proakis; J. H. L. Hansen, Discrete-Time Processing of Speech Signals, Prentice Hall, Nova Jersey, 1993.

****