Noise Reduction in Speech Signals Using Discrete-time Kalman Filters Combined with Wavelet Transforms

Ranieri Guimarães França, Jucelino Cardoso Marciano dos Santos, Leandro Aureliano da Silva,
Edna Lúcia Flôres, Rodrigo Pinto Lemos, and Gilberto Arantes Carrijo

***Abstract*—Noise reduction in speech signals is a growing area that encountered several applications like communication channel transmission, automatic speech recognition, telephony and hearing aids, among others. This paper introduces a technique for noise reduction in speech signals that combines both Discrete-time Kalman filtering and Wavelet transforms. While filtering provides noise reduction, Wavelets transforms allow minimizing spectral distortion. In order to assess the efficiency of this combination, we compared both the segmental signal-to-noise ratio and the Itakura-Saito distance at the input to their respective values at the output of the proposed system. Also, we compared the noise reduction performance of the proposed system to that of Kalman filtering and the combination of Wavelet transforms with Kalman filtering has shown satisfactory results.**

***Index Terms*—Kalman filter, Itakura-saito distance, noise suppression, speech signals, wavelets**

I. INTRODUCTION

Noise reduction in speech signals is a field of study devoted to recovering an original signal from its noise corrupted version. The noise can be white, filtered, impulsive or even other types of noise usually found in speech communication systems.

Over the past decades, the removal of this noise from speech signals has become an area of interest of several investigators around the world, since the presence of noise can significantly degrade the quality and intelligibility of these signals. In this sense, many studies have been conducted since the sixties, with the goal of developing algorithms for improving the quality of audio and speech signals [1], [3], [4], [8], [10] [11], [12] and [13]. Some techniques and methods gained greater prominence, among them: psychoacoustics, spectral subtraction, Wiener filter, Kalman and processed wavelet filters.

Each of these techniques has both favorable characteristics and technical challenges for application in noise reduction. In the case of Spectral Subtraction, efforts have been made to eliminate the musical noise generated by the result of the subtraction [14] - [16].

When using Wavelet transforms, the signal is divided into approximation and detail coefficients to which a threshold is applied for noise reduction. However, what values to adopt in thresholding as well as new kinds of thresholds still remain a matter under investigation [17] and [18].

Yu Shao-and Chip-Hong Chang [21] used a Kalman filter based on a wavelet filter bank for the enrichment of speech signals corrupted by noise. The adaptation of this filter in the Wavelet domain effectively reduced non-stationary noise. At the end, a perceptual weighting filter was applied to the Kalman filter output signal. The application of this last filter sought to take advantage of psychoacoustic model properties to improve speech intelligibility. It was observed that the human auditory model could be used both in the time and frequency domain. The developed system was able to reduce the noise in different environments for low degradation of the speech signal.

Dhivya and Justin [22] proposed a noise reduction technique based on a combination of wavelet and spectral subtraction. In this technique, spectral subtraction is applied to the approximation coefficients while the threshold is applied to the detail coefficients. They compared five Wavelet filters and found the best filter based on the signal-to-noise ratio. To check the performance of the proposed technique, they employed the signal-to-noise ratio, the correlation coefficient and the perceptual evolution of speech quality (PESQ).

Although the advances in these algorithms show how noise removal was satisfactory, they do not show how much they are able to minimize it for spectral distortion.

The purpose of this work is to combine discrete-time Kalman filtering with wavelet transform such that while the filter is used to reduce the noise, the transform is used to minimize the spectral distortion.

This paper is organized as follows: section II describes both Kalman filtering and Wavelet transforms; the proposed combination and its corresponding results are shown in sections III and IV respectively. Finally, section V brings the conclusions.

R. G. França is with Electrical Engineering Department, Universidade Federal de Uberlândia (UFU), Brasil. (corresponding author to provide phone: +55 034 99661-4919; e-mail: ranierig@gmail.com).

J. C. M. Santos was with Mathematics Department, Universidade Federal de Goiás (UFG), Brasil. He is now with Electrical Engineering Department, Universidade Federal de Uberlândia (UFU).

L. A. Silva was with Electrical Engineering Department, Universidade de Uberaba (UNIUBE), Brasil. He is now with Electrical Engineering Department, Universidade Federal de Uberlândia (UFU).

E. L. Flôres is with Electrical Engineering Department, Universidade Federal de Uberlândia (UFU), Brasil.

R. P. Lemos was with Electrical Engineering Department, Universidade Federal de Goiás (UFG), Brasil. He is now with Electrical Engineering Department, Universidade Federal de Uberlândia (UFU).

G. A. Carrijo is with Electrical Engineering Department, Universidade Federal de Uberlândia (UFU), Brasil.

ISBN: 978-988-14047-6-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

II. DESCRIPTION OF THE ALGORITHMS

A. Discrete Kalman Filtering (KF)

In the sixties, Rudolf Emil Kalman published his seminal paper entitled "*A New Approach to Linear Filtering and Prediction Problems*" in which he described a recursive solution to discrete-time linear filtering problem [1].

Since then, due to major advances in digital computing, Kalman filtering has become a very important tool in such diverse areas as navigation, processes monitoring, Economics and signal reconstruction from noisy samples.

In that paper, Kalman filtering is derived according to the heuristics proposed by Vaseghi [8], in which the speech signal is initially modeled by an auto-regressive (*AR*) process of order *P* (*AR*(*P*)) such that:

$$x(n) = \sum_{k=1}^{P} a_P(k)x(n-k) + w(n) \tag{1}$$

where $a_P(k)$ are the linear prediction coefficients of order P, $w(n)$ is the prediction error associated to the excitation of the source-filter model of speech production, and $x(n)$ is the speech signal.

As the acquisition of speech and audio signals occurs in the presence of some type of additive noise, this should be considered in the modeling process:

$$y(n) = x(n) + v(n) \tag{2}$$

where $y(n)$ is the noisy speech signal and $v(n)$ is the Gaussian additive noise.

Deriving a state-space representation to (1) and (2), we can rewrite them as:

$$\boldsymbol{x}(n) = \boldsymbol{A}(n-1)\boldsymbol{x}(n-1) + \boldsymbol{w}(n) \tag{3}$$

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

such that $\boldsymbol{x}(n)$ is the $P \times 1$ state vector at time n; $\boldsymbol{A}(n-1)$ is the $P \times P$ state transition matrix from time $n-1$ to n; $\boldsymbol{w}(n)$ is an the $P \times 1$ input excitation vector, modeled as white noise; $\boldsymbol{y}(n)$ is the $M \times 1$ observation vector; $\boldsymbol{H}(n)$ is a $M \times P$ channel distortion matrix; $\boldsymbol{v}(n)$ is a $M \times 1$ additive white noise vector [8].

According to Vaseghi [8], $\boldsymbol{w}(n)$ and $\boldsymbol{v}(n)$ are considered to be independent white noise processes such that:

$$E[\boldsymbol{v}(n)\boldsymbol{v}^T(k)] = \begin{cases} R(n); & k = n \\ 0; & k \neq n \end{cases} \tag{5}$$

$$E[\boldsymbol{w}(n)\boldsymbol{w}^T(k)] = \begin{cases} Q(n); & k = n \\ 0; & k \neq n \end{cases} \tag{6}$$

where $R(n)$ and $Q(n)$ are respectively the diagonal elements of the additive noise and prediction error covariance matrices.

The Kalman filter estimates a process by using a kind of feedback control: first, the filter estimates the process state at a given time, then the feedback is obtained in the form of a new measure.

According to Brown and Hwang [2] and Vaseghi [8], Kalman filter equations can be divided into time-update (prediction) and measurement-update (correction) equations. Time-update equations are given by:

$$\hat{\boldsymbol{x}}(n/n-1) = \boldsymbol{A}(n-1)\hat{\boldsymbol{x}}(n-1/n-1) \tag{7}$$

and measurement-update equations are given by:

$$\boldsymbol{K}(n) = \boldsymbol{P}(n/n-1)\boldsymbol{H}^T(n) \times [\boldsymbol{H}(n)\boldsymbol{P}(n/n-1)\boldsymbol{H}^T(n) + R(n)]^{-1} \tag{8}$$

$$\hat{\boldsymbol{x}}(n/n) = \hat{\boldsymbol{x}}(n/n-1) + \boldsymbol{K}(n)[\boldsymbol{y}(n) - \boldsymbol{H}(n)\hat{\boldsymbol{x}}(n/n-1)] \tag{9}$$

$$\boldsymbol{P}(n/n) = [\boldsymbol{I} - \boldsymbol{k}(n)\boldsymbol{H}(n)]\boldsymbol{P}(n/n-1) \tag{10}$$

where $\boldsymbol{P}(n/n)$ is the covariance matrix of the prediction error at time n; $\boldsymbol{K}(n)$ is the Kalman gain matrix, responsible for minimizing the diagonal elements of $\boldsymbol{P}(n)$, the covariance matrix of the estimation error, and $\hat{\boldsymbol{x}}(n/n)$ is the estimate at time n, given past observations $\boldsymbol{y}(n)$.

B. Wavelet Based Noise Reduction (WT)

The Wavelet transform of a signal $f(t)$ is defined as [5]:

$$Wf(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt \tag{11}$$

For a N-sample discrete signal, this integral can be approximated by a summation:

$$Wf(a,b) = \sum_{t=0}^{N-1} f(t)\psi_{a,b}(t) \tag{12}$$

The function $\psi_{a,b}(t)$, called Wavelet, is derived from a function $\psi(t)$ by making:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{13}$$

where b stands for wavelet position or translation and a is the scale parameter associated to the width of the time window.

There is a broad class of functions $\psi(t)$, called mother Wavelets, to be chosen: *Daubechies*, *symlets*, *coiflet*, etc [5].

Dubey and Gupta [20] compared two different wavelet families, Daubechies and Coiflets, for noise reduction in speech signals. This comparison was performed using cross-correlation to determine the best noise reduction setting. They used Coiflet wavelets with order 5 and Daubechies with orders 9 and 10. The speech signals resulting from the use of Daubechies 9 wavelets have sounded more pleasant.

ISBN: 978-988-14047-6-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

The WT allows successively decomposing the original signal into approximation and detail coefficients to form decomposition tree. Approximation coefficients A_m carry the low frequency information of the signal while detail coefficients D_m carry the high frequency content associated to the mother wavelet. In the present work, we adopted Daubechies 9 wavelets and proceeded to decompose the signals until the level $m = 3$.

The basic principle of noise reduction is choosing a threshold to select which coefficients will be kept in order to preserve the signal information while minimizing the noise level.

We found that detail coefficients at level 1 were those with higher noise content, such that we applied the threshold directly on them. This have been already stated by Duarte, Vieira Filho and Villarreal [19] who said that adding white noise to the original signal can improve the estimated signal by reducing more significantly the noise at high frequencies. Therefore, we have chosen applying the threshold to D_1, which is related to highest frequencies.

Among the thresholds found in the literature, we adopted Hard Thresholding [9] that replaces by zero the smaller coefficients, according to:

$$D_{1(n)} = \begin{cases} D_1(n), & D_1(n) \geq 0.3 \times \mathrm{Max}\left(D_1(n)\right) \\ 0, & otherwise \end{cases} \quad (14)$$

III. APPLICATION OF NOISE REDUCTION ALGORITHMS

A. Using Discrete-time Kalman Filtering

In order to evaluate the performance of the proposed algorithm, multiple speech signals have been previously recorded in ".wav" format at a sampling frequency of $22050\ Hz$, using $16\ bits\ per\ sample$. Next, the samples where normalized to lie in the range $[-1, 1]$ and added Gaussian noise. Then, the signal was segmented using Hamming windows with $512\ samples$ each and 50% overlapping.

Kalman filtering requires processing each block according to the algorithm in Fig. 1.

After the signal is processed, we calculate the segmental signal-to-noise ratio ($SNRseg$) an the Itakura-Saito distance to evaluate the performance of the algorithm.

B. Using Discrete-time Kalman Filtering combined with Wavelet Transform (Proposed Technique)

Applying the proposed technique requires the same initial steps described here in section III, item A. The only difference is that we apply the WT to the signal estimate provided by the Kalman filter. Fig. 2 resumes the proposed algorithm.

Kalman filtering used the same parameters that we adopted in the last section. The Wavelet transform algorithm performed signal decomposition until the level 3 using the *Daubechies 9* wavelet.

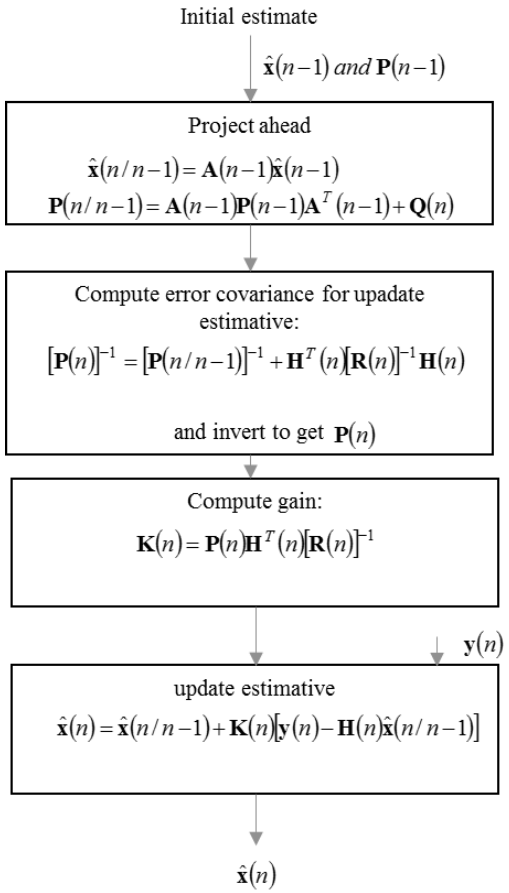

Fig. 1. Block diagram of Kalman filtering.

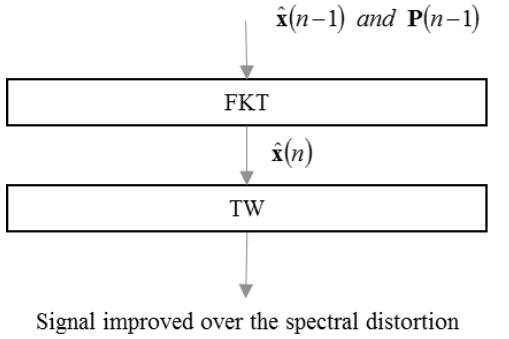

Fig. 2. Block diagram of Kalman filtering combined with wavelet transforms.

IV. RESULTS

In order to evaluate the performance of the proposed algorithm, we used different speech signals sampled at $22050\ Hz$ and using $16\ bits\ per\ sample$. Those samples where contaminated with additive white Gaussian noise and the resulting signal was segmented using Hamming windows with $512\ samples$ each and 50% overlapping. Both Kalman filtering and Wavelet transform algorithms were implemented using Matlab R2013b.

We employed the segmental signal-to-noise ratio ($SNRseg$) and the Itakura-Saito distance $\left(d(b,a)\right)$ to compare the performance of the proposed technique to that of the Kalman filter alone.

The segmental signal-to-noise ratio is an effective measure that can be computed used in short speech segments in order to balance the weights according to the signal strength of each segment. It can be calculated using [6]:

ISBN: 978-988-14047-6-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$SNRseg = \frac{10}{M}\sum_{j=0}^{M-1}\log_{10}\left[\sum_{n=mj-N+1}^{mj}\frac{x^2(n)}{[x(n)-\hat{x}(n)]^2}\right] \quad (15)$$

where mj represents the bounds of each of the M N-sized frames.

$SNRseg$ does not provide a significant performance measure when comparing signals with different spectra. However, distance measurements are quite sensitive to spectral variation. In those cases, Itakura-Saito distance provides better results and it can be calculated using linear prediction coefficients (LPC) [7]:

$$d(\boldsymbol{a},\boldsymbol{b}) = \log\left[\frac{\boldsymbol{aRa}^T}{\boldsymbol{bRb}^T}\right] \quad (16)$$

where $\boldsymbol{a}$ is the LPC vector of the original signal; $\boldsymbol{R}$ is the original signal correlation matrix and $\boldsymbol{b}$ is the LPC vector of the estimated signal.

When the result of (16) is close to zero, the spectra of the original and estimated signals are close to each other. If the result is exactly zero then the spectra are equal to each other.

In order to illustrate the efficiency of the proposed technique regarding noise reduction and spectral distortion, Figure 3 shows the application of the Kalman filtering to the utterance "elétrica" with input $SNR = 3\ dB$. The filtered signal allows noting a significant reduction in the noise level, especially during silence intervals. The resulting output SNR was $10\ dB$ and the Itakura-Saito distance was 0.4844.

Figure 4 shows the application of the proposed technique to the same utterance also with input $SNR = 3\ dB$. Again, the filtered signal allows noting a significant reduction in the noise level, especially during silence intervals. However, the resulting output SNR was $11\ dB$ and the Itakura-Saito distance was 0.2924. This indicates that the Kalman filtering combined with Wavelet transforms performed better than Kalman filtering alone.

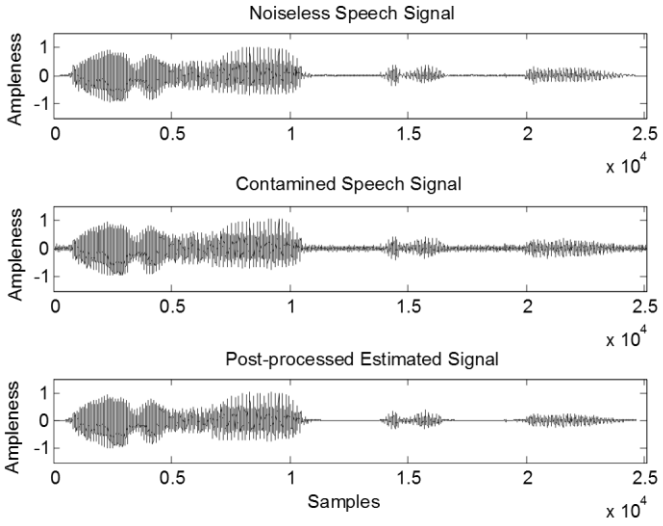

Fig. 3 – Processing using FKT.

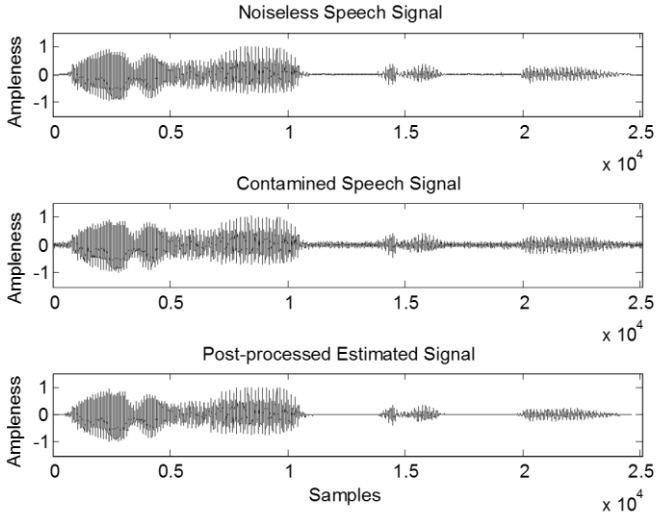

Fig. 4 – Processing using FKT combined with TW.

In Figures 5, 6 and 7 we extend this analysis to a set of 11 different utterances for input SNR 0 dB, 3 dB and 6 dB, respectively.

In almost all cases, the output SNR provided by the proposed technique remained larger than that provided by Kalman filtering alone. In the worst cases, we got a draw. Also, both techniques provided larger output SNR improvement for smaller SNR (0 dB).

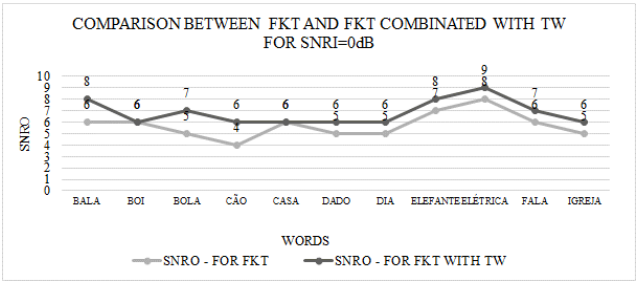

Fig. 5 – Comparison of the techniques FKT e FKT combined with TW in relation to $SNRO$ for $SNRI = 0\ dB$.

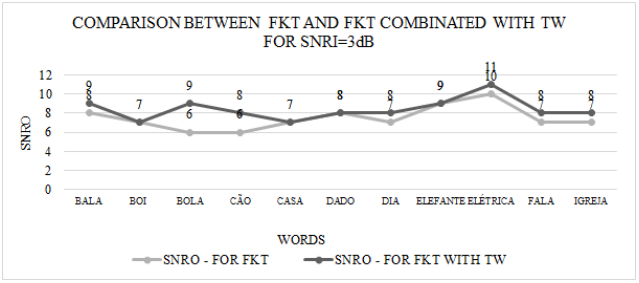

Fig. 6 – Comparison of the techniques FKT e FKT combined with TW in relation to $SNRO$ for $SNRI = 3\ dB$.

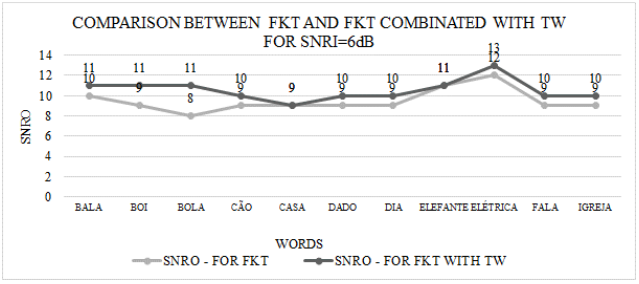

Fig. 7 – Comparison of the techniques FKT e FKT combined with TW in relation to $SNRO$ for $SNRI = 6\ dB$.

ISBN: 978-988-14047-6-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

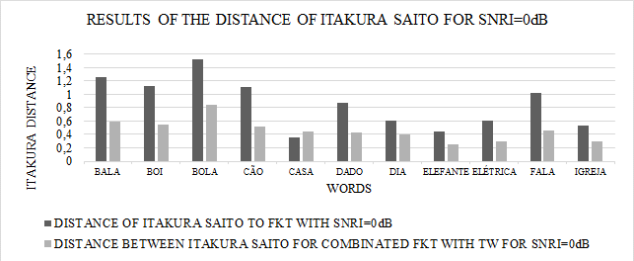

Fig. 8 – Comparison between the techniques of FKT e FKT combined with TW for $SNRI = 0\ dB$, using the distance of Itakura Saito.

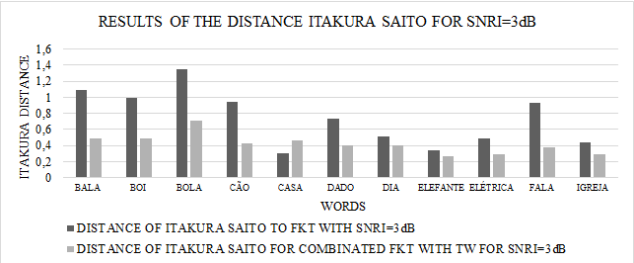

Fig. 9 – Comparison between the techniques FKT e combined FKT with TW for $SNRI = 3dB$, using the distance of Itakura Saito.

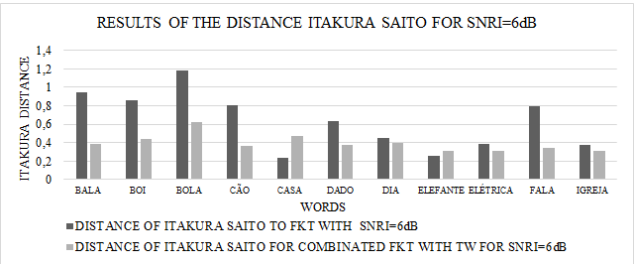

Fig. 10 – Comparison between the techniques of FKT and FKT combined with TW for $SNRI = 6\ dB$, using the distance of Itakura Saito.

An important issue in speech noise reduction is the ability to suppress/reduce the noise without producing large distortion on the estimated signal. In order to assess the level of distortion provided by the techniques under study, in Figures 8, 9 and 10, we show the Itakura-Saito distance between the estimated signal and the original signal for the same 11 utterances with input SNR $0\ dB$, $3\ dB$ and $6\ dB$, respectively.

Combining Kalman filtering with Wavelet transforms provided estimated signals closer to their respective original versions in 10 out of the 11 utterances under analysis.

In Figures 11 and 12, we compare the average behavior in terms of the input SNR, measured along the 11 utterances. Be it regarding the output SNR or the Itakura-Saito distance, the proposed technique performed better than using Kalman filtering alone, providing smaller spectral distortion.

Indeed, the use of Wavelet transforms practically halved the Itakura-Saito distance provided by the Kalman filtering.

V. Conclusion

This paper presented a technique which combines discrete-time Kalman Filter with Wavelet Transform for reducing spectral distortion in speech signal denoising. In order to assess the denoising effectiveness, we compared the input and output segmental signal-to-noise ratios. Also, we used Itakura-Saito distance to measure the amount of spectral distortion produced by each approach under study.

After several tests, combining Kalman filtering with Wavelets showed better results in reconstructing signals corrupted by white noise than using Kalman filtering alone, as well as practically halved spectral distortion.

Then, the proposed technique of noise reduction proved to be more reliable than the Kalman filtering alone for producing low spectral distortion.

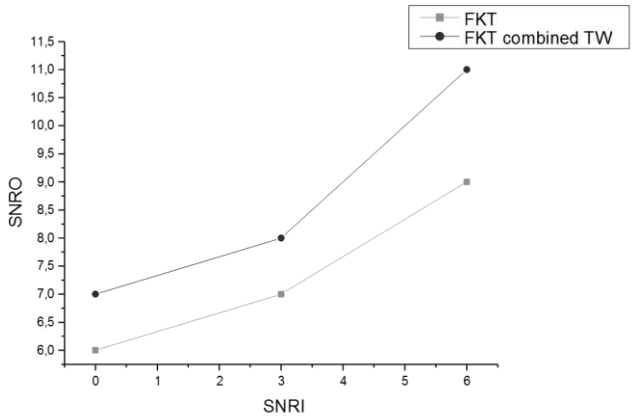

Fig. 11 – Average result of 11 words for $SNRI's$ of 0, 3 and 6 dB, using the technics of FKT e FKT with TW for $SNRO$.

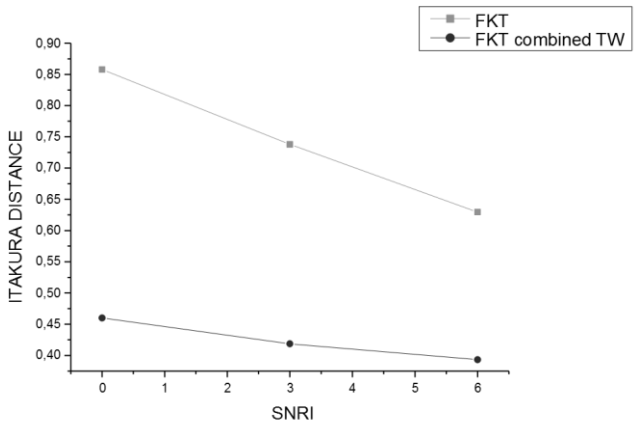

Fig. 12 – Average result of 11 words for $SNRI's$ of 0, 3 and 6 dB, using the technics of FKT and FKT with TW for $d(b,a)$.

References

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Transaction of the ASME-Journal of Basic Engineering, March 1960, pp. 35-45.

[2] R. G. Brown and P. Y. C. Hwang, "Introduction to Random Signals and Applied Kalman Filtering," John Wiley & Sons, Inc, 1997.

[3] Ma. N, Bouchard, M. and Goubran, R, "A Kalman Filter with a perceptual post-filter to enhance speech degraded by colored noise," Proceedings of IEEE. Conf. on Acousties. Speech and Proc. 2004.

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

[5] Misiti, M; Misiti, Y; Oppenheim, G.; Poggi, J. M; "Wavelet Toolbox User'S Guide," The Mathworks: Natick, 1996.

[6] J. R. Deller, J. G. Proakis and J. H. L. Hansen, "Discrete-Time Processing of Speech Signals," Prentice Hall, 1993.

[7] L. R. Rabinner and R. W.Schafer, "Digital processing of speech signals," Prentice- Hall, 1978.

[8] S. V. Vaseghi, "Advance Digital Signal Processing and Noise Reduction," John Wiley & Sons, 2000.

[9] G. Strang and T. Nguyen, "Wavelets and Filter Banks," Wellesley-Cambridge Press, 1996.

[10] N. Ma, B. Martin and G. A. Rafik, "Speech Enhancement Using a Masking Threshold Constrained Kalman Filter and Its Heuristic

Implementations," IEEE Transactions on Audio, Speech, And Language Processing, Vol. 14, No. 1, Jan, 2006.

[11] D. O'Shaughnessy, "Enhancing speech degraded by additive noise of interfering," IEEE Commun. Mag., vol. 27, No. 2, Feb. 1989, pp. 46-52.

[12] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactions on Audio, Speech, Signal Processing, Vol. ASSP – 27, No. 2, Apr. 1979 pp. 113-120.

[13] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," IEEE Trans. Speech Audio Processing, vol. ASSP-32, Dec. 1984, pp. 1109-1121.

[14] R. Miyazaki, H. Saruwatari, T. Inoue, Y. Takahashi, K. Shikano, K. Kondo, "Musical-noise-free speech enhancement based on optimized iterative spectral subtraction," IEEE Trans. Audio Speech Lang. Process., Jul. 2012, pp. 2080–2094.

[15] R. Miyazakia, H. Saruwataria, S. Nakamuraa, K. Shikanoa, K. Kondob, J. Blanchettec, M. Bouchardc. "Musical-noise-free blind speech extraction integrating microphone array and iterative spectral subtraction," Elsevier, Signal Processing., Vol. 102, Sept. 2014, pp. 226-239.

[16] C- Ta Lu, K-Fu Tseng and C- Tsung Chen, "Reduction of Musical Residual Noise Using Hybrid-Mean Filter," International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 6, No. 4, Aug, 2013.

[17] Rajeev Aggarwal, Jai Karan Singh, Vijay Kumar Gupta, Sanjay Rathore, Mukesh Tiwari and Anubhuti Khare. "Noise Reduction of Speech Signal using Wavelet Transform with Modified Universal Threshold," International Journal of Computer Applications, Apr, 2011, pp. 14-19.

[18] Li Ruwei, B. Changchun; X. Bingyin and J. Maoshen. "Speech enhancement using the combination of adaptive wavelet threshold and spectral subtraction based on wavelet packet decomposition," Signal Processing (ICSP), 2012 IEEE 11th International Conference on Vol. 1, 21-25 Oct. 2012, pp. 481-484.

[19] M. A. Q. Duarte, J. Vieira Filho e F. Villarreal, "Um novo Método de Redução de Ruído em Sinais de Voz Baseado em Wavelets," XXI Simpósio Brasileiro de Telecomunicações-SBRT04, 06-09 de Setembro de 2004, Belém-PA.

[20] K. Dubey and V. Gupta, "A Review on Speech Denoising Using Wavelet Techniques," International Journal of Engineering Trends and Technology (IJETT), Vol. 4 Issue 9, Sept, 2013.

[21] Yu-Shao and Chip-Hong Chang. "A Kalman filter based on wavelet filter-bank and psychoacoustic modeling for speech enhancement," Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on Island of Kos. 21-24 May 2006.

[22] R. Dhivya and J. Justin, "A Novel Speech Enhancement Technique," IJRET: International Journal of Research in Engineering and Technology, Vol. 3, Special Issue: 07, May 2014, pp. 98-102.

ISBN: 978-988-14047-6-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)