# Speech Enhancement with Bionic Wavelet Transform and Recurrent Neural Network

Mourad TALBI*, Lotfi SALHI* , Wahid BARKOUTI* and Adnen CHERIF*

**Signal Processing Laboratory, Sciences Faculty of Tunis, 1060 Tunis, TUNISIA*

**mouradtalbi1969@yahoo.fr**

**adnen2fr@yahoo.fr**

**lotfi.salhi@laposte.net**

**barkouti@hotmail.fr**

**Abstract:** This paper deals with problem of speech enhancement using Bionic wavelet transform and recurrent neural network. Indeed this work describes a new method to remove additive background noise from noisy speech. The method can be divided into two stages, the first is the application of the Bionic wavelet transform to the speech signals and the second consists in applying an Elman neural network to find an optimal thresholding set to remove related noise wavelet coefficients. Simulation results show good performances of the proposed technique in comparison with respect to many other methods.

**Key words:** Speech enhancement, Bionic wavelet transform, Threshold, Spectral subtraction, Elman neural network.

## INTRODUCTION

In speech signal processing, the presence of background noise is a very important problem. In fact the noise existence can affects the performances of speech recognition, coding and synthesis. Generally speaking, there are different types of noise depending on the way of contamination. So we distinguish three forms of noise: convolutive, multiplicative and additive. We will be interested to the latter one. A noisy speech signal with an additive noise can be expressed as [VAN93]:

$$x(t) = s(t) + b(t) \tag{1}$$

where x(t), s(t) and b(t) represent respectively the noisy speech signal, the clean speech signal and the noise signal.

Hence the problem of speech enhancement consists in removing the noise signal and in improving the cleaned signal quality. In other word, the influence of the noise have to be reduced and ameliorating the quality of the speech signal by determining an estimate of the clean speech signal, $\hat{s}(t)$ which should be optimal and favoured by a human listener. In practice, Background noise removal is a very difficult task and it is generally followed by quality degradation. Traditional algorithms of speech enhancement include the Wiener filtering, spectral subtraction and denoising methods based on microphone array. The wavelet transform discriminates itself in non stationary signals analysis such as speech signal. A denoising method based on wavelet coefficients thresholding, has been introduced by Donoho [DON95]. This method has proved its efficiency in white noise reduction. Actually, several methods use the wavelet thresholding [SEO97], [BAH01], [SHE01], [CHA02]. Analysis in wavelet domain is based on the modelling of pre-post perceptual periphery; hence some efforts have been performed for employing this processing tool in speech denoising. This approach makes an association between the nonlinear filtering and multi-resolution analysis [XIA03, MAL89]. While the technique based on wavelet doesn't require a noise or speech model and can be applied to a broader signals class, just a general thresholding of the wavelet coefficients doesn't guarantee a good performance. The Bionic wavelet transform introduced by J.Yao and Y.T.Zhang[XIA03] provides a better concentration of the signal energy. Furthermore it gives better selectivity of time-frequency and this will be expected to yield far more efficient performance of thresholding. In this paper, we use a recurrent neural network for determining suitable thresholds to be employed to threshold the bionic wavelet coefficients. In fact, the success of the classical thresholding techniques is based on the suitable choice of the values of the thresholds to be employed. Our speech enhancement method is inspired from the technique of wavelet denoising of speech using neural networks for threshold selection. This technique is introduced by C.A. Medina & al[MED03]. It consists in applying the discrete wavelet transform to the noisy speech signal and applies a level dependent threshold for each band. These thresholds are determined by applying a neural network for each band. In this paper we first dealing with speech denoising techniques based on wavelets, second we are interesting in bionic wavelet transform and its employment in speech enhancement and third we deal with recurrent neural networks more specifically Elman neural network. Finally we present our proposed speech denoising technique and give some simulation results.

# 1. Wavelet based denoising approaches

The traditional techniques of signal denoising, used Fourier analysis. This is based on the fact that the noise is principally manifested as high frequency oscillations. Bearing this in mind, signals is decomposed into sinusoidal waveforms having different frequencies and low frequency components are only used when reconstructing the enhanced signal. The denoising techniques based on wavelet, suppose that the signal analysis at different resolutions might improve the true underlying signal separation from noise. Since the discrete wavelet transform (DWT) is orthogonal and linear, consequently when transforming white noise in time domain, we obtain a white noise in the wavelet domain. It also enables compact coding, since the wavelet coefficients of the details possess high absolute values only in the intervals of rapid time series change. These proprieties led Donoho and Johnstone [DON94] to suggest a thresholding denoising approach. The three following steps summarise this approach:

-Apply the discrete wavelet transform:

$$DWT(x)(j,k) = \int_{-\infty}^{+\infty} x(t)\cdot\psi^*\left(\frac{t-k\cdot 2^j}{2^j}\right)dt \qquad (2)$$

to noisy signal, $x(t)$ given by (1), where $b$ is a Gaussian white noise having $\sigma^2$ as a variance. In wavelets domain, we have:

$$W_x = W_s + W_b \qquad (3)$$

-Apply thresholding to the obtained wavelet coefficients.

-obtain the enhanced signal $\tilde{x}$ by applying the inverse transform $W^{-1}$ to the thresholded wavelets coefficients vector $Y_{TH}$ :

(4)

$$\tilde{x} = W^{-1}Y_{TH} \qquad (4)$$

The thresholding is non linear and generally is hard or soft. The thresholding denoising approach, is based on the fact that the energy of the clean signal is concentrated in small number of great wavelet coefficients although the noise contaminates all coefficients. For handling Gaussian white noise, Donoho has employed a universal threshold which is expressed as follow:

$$\lambda = \hat{\sigma}\sqrt{2\log(n)} \qquad (5)$$

where $n$ designates the noisy signal length and $\hat{\sigma}$ represents the estimate of the noise standard deviation, given by:

$$\hat{\sigma} = MAD / 0.6745 \qquad (6)$$

with the MAD is the absolute median estimated on the first scale. To handle a correlated noise, Johnstone and Silverman [JOH97] have suggested a level dependent threshold which is defined as:

$$thr_j = \hat{\sigma}_j\sqrt{2\cdot\log(n)} \qquad (7)$$

where $\hat{\sigma}_j = MAD_j / 0.6745$ and $MAD_j$ designates the absolute median estimated at scale *j*. Sungwook Chang, Y. Kwon, Sung-il Yang [CHA02] have proposed to employ a node dependent threshold. This threshold is applied to each node of the wavelet packet tree and is expressed as follow:

$$thr_{j,k} = \hat{\sigma}_{j,k}\sqrt{2\cdot\log(n)} \qquad (8)$$

with $\hat{\sigma}_{j,k} = MAD_{j,k}/0.6745$ and $MAD_{j,k}$ is the absolute median estimated at the scale *j* and subband *k*.

### 1.1. Threshold limitation

The wavelet based denoising method doesn't require a speech or a noise model and can be used to a large class of signals. Though a general wavelet coefficients thresholding doesn't ensure a good performance as obtained by bionic wavelet transform, BWT. The latter owns a better propriety of concentration of signal energy and time-frequency selectivity. This leads to an efficient thresholding performance [XIA03].

# 2. Bionic wavelet transform

J. Yao and Y.T. Zhang have proposed the bionic wavelet transform (BWT) as a new time-frequency technique and this by referring to the perceptual model

[YAO01]. The term "bionic" means that it is guided by an active biological mechanism [XIA03]. The BWT decomposition is both perceptually scaled and adaptive [MIC07]. The initial perceptual aspect of the transform comes from the logarithmic spacing of the baseline scale variables, which are designed to match basilar membrane spacing [MIC07]. Then, two adaptation factors control the time-support employed at each scale, based on a non-linear perceptual model of the auditory system [MIC07]. The basis for this transform is the Giguerre -Woodland non linear transmission line model of the auditory system [GIG93, GIG94], an active-feedback electro-acoustic model incorporing the auditory canal, middle ear, and cochclea[MIC07]. The model yields estimates of the time-varying acoustic compliance and resistance along the displaced basilar membrane, as a physiological acoustic mass function, cochlear frequency-position mapping, and feedback factors representing the active mechanisms of outer hair cells. The net result can be seen as a technique for the estimation of the time-varying quality factor $Q_{eq}$ of the cochlear filter banks as the input sound waveform function [MIC07]. The references [GIG94], [ZHE99] and Yao and Zhang [YAO01] give the complete details on the elements of this model. The BWT adaptive nature is insured by a time-varying linear factor $T(a,\tau)$ which represents the scaling of the cochlear filter bank quality factor $Q_{eq}$ at each scale over time [MIC07]. For each scale and time, the adaptation factor of BWT, $T(a,\tau)$, is calculated by employing the update equation[MIC07]:

$$T(a,\tau+\Delta\tau)=\frac{1}{\left(1-G_1\dfrac{C_s}{C_s+\left|X_{BWT}(a,\tau)\right|}\right)\left(1+G_2\left|\dfrac{\partial}{\partial t}X_{BWT}(a,\tau)\right|\right)} \qquad (9)$$

where $C_s$ is a constant ($C_s=0.8$)that represents non linear saturation effects in the cochlear model [MIC07, YAO01]. The quantities $G_1$ and $G_2$ are respectively, the active gain factor that represents the outer hair cell active resistance function and the active gain factor that represents the time-varying compliance of the Basilar membrane [MIC07]. Practically speaking, the partial derivative in (eq.9), can be approximated using the first difference of the previous points of the BWT at that scale [MIC07].

The quantity $X_{BWT}(a,\tau)$ represents the bionic wavelet transform of the signal $x(t)$. It is given by:

$$X_{BWT}(a,\tau)=\frac{1}{T(a,\tau)\sqrt{a}}\int x(t)\cdot\tilde{\varphi}^{*}\left(\frac{t-\tau}{a\cdot T(a,\tau)}\right)\cdot e^{-j\omega_0\left(\frac{t-\tau}{a}\right)}dt \qquad (10)$$

where $\tilde{\varphi}$ is the mother wavelet envelop so we have[7]:

$$\varphi(t)=\frac{1}{T(a,\tau)\sqrt{a}}\tilde{\varphi}\left(\frac{t}{T(a,\tau)}\right)\cdot\exp(j\omega_0 t) \qquad (11)$$

where $\omega_0$ is the base fundamental frequency of the unscaled mother wavelet. In practice $\omega_0$ is equals to 15165.4 for the human auditory system [YAO01]. The discretization of the scale $a$ is achieved by employing a pre-determined logarithmic spacing across the desired frequency range, so that at each scale, the canter frequency is expressed by[MIC07]:

$$\omega_m=\omega_0/(1.1623)^m \text{ with } m=0,1,2,... \qquad (12)$$

For this implementation, based on original work for cochlear implant coding [YAO02], coefficients at 22 scales, $m=7,...,22$, are computed employing numerical integration of the continues wavelet transform [MIC07]. These 22 scales are corresponding to centre frequencies logarithmically spaced from 225Hz to 5300Hz [MIC07].In the formula (eq.11), the role of first factor $T(a,\tau)$ multiplying $\sqrt{a}$ is to ensure that the energy remains the same for each mother wavelet. The role of second factor $T(a,\tau)$ is to adjust the envelop $\tilde{\varphi}(t)$ without adjusting the central frequency of $\varphi(t)$ [ XIA03]. Thus, the main difference between (BWT) and the continuous wavelet transform (CWT) is based on the fact that the time-frequency resolution achieved by (BWT) can be adjusted with adaptive manner not only by frequency variation of the signal but also by instantaneous amplitudes of this signal. That is the mother wavelet that makes adaptive the continuous wavelet transform, while the adaptive characteristic of the bionic wavelet transform, BWT, comes from the mechanism of active control in the human auditory model, which adjusts the mother wavelet associated to (BWT) according to the analyzed signal. Basically, the idea of the (BWT) is inspired from the fact that we need to make the mother wavelet envelop varying in time according the signal characteristics. The employed mother wavelet $\varphi$ in [XIA03] is a Morlet wavelet and it's envelop $\tilde{\varphi}$ is given by [MIC07]:

$$\tilde{\varphi}(t)=\exp\left[-\left(\frac{t}{T_0}\right)^2\right] \qquad (12)$$

where $T_0$ represents the initial time-support.

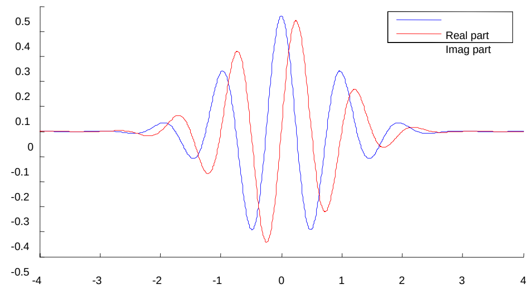


**Figure 1.** *Morlet wavelet.*

It can be shown [XIA03, YAO02] that the obtained BWT coefficients, $X_{BWT}(a,\tau)$ are derived by using the following formula [MIC07]:

$$X_{BWT}(a,\tau) = K(a,\tau) X_{WT}(a,\tau) \qquad (13)$$

with $K(a,\tau)$ satisfying:

$$K(a,\tau) = \frac{\sqrt{\pi}}{C} \frac{T_0}{\sqrt{1+T^2(a,\tau)}} \qquad (14)$$

where $C$ represents a normalizing constant calculated from the squared mother wavelet integral. This representation yields to an effective computational technique for calculating in direct manner, the BWT coefficients from those of the wavelet transform, WT without using the BWT definition given by Eq.(10). There are some key differences between the discretized CWT employing the Morlet wavelet, used for the BWT, and a filterbank based WPT employing an orthonormal wavelet, for example the Daubechies family, as employed for the comparative baseline technique. One is that the WPT is perfectly reconstructable, while the discretized CWT is an approximation whose exactness depending on the number and placement of frequency bands selected [MIC07]. To solve this problem, we propose in our work, to use 30 scales instead of 22 scales in the expression of the discretized CWT using the Morlet mother wavelet. This choice of the number of scales ($N = 30$) is done by simulation: it is suitable for a perfect reconstruction of the whole speech signals belonging to our Arabic speech signals database.

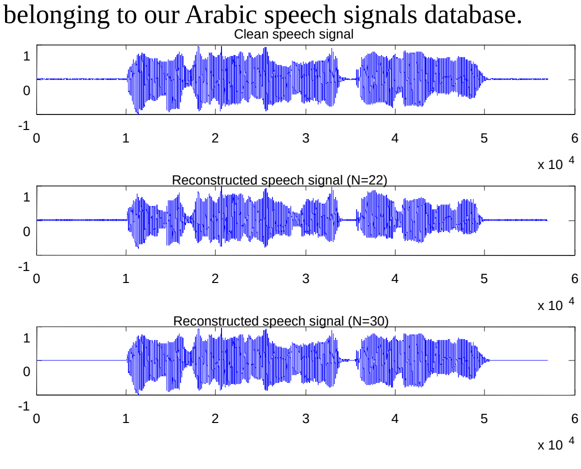


**Figure 2.** *Perfect speech signal reconstruction using 30 Scales.*

### 2.1. Denoising by BWT

The reference [XIA03], gives the principle of speech enhancement scheme based on bionic wavelet transform (BWT). This principle is illustrated in figure 14.

## 3. Neural net work and speech enhancement

There are many research works employing artificial neural networks (ANN) to perform nonlinear signal filtering for the reason to enhance signal and reduce noise. But due to the nonlinear nature, the majority of applications have to be developed for specific training and are data dependent [NAS04]. Consequently, when applying the neural network for nonlinear filtering, we have to be able to collect a samples training extensive set for the purpose of covering all possible situations and developing a neural network for the purpose of adapting to the given training set [NAS04]. Suppose we have a samples training set $\{u(i), y(i)\}$ where $u(i)$ represents the input vector and $y(i)$ designates the output one. The goal of function approximation is to make identification of a mapping $\varphi$ from $u$ to $y$ satisfaying $y = \varphi(u)$ [NAS04, SEO93] such that the expected sum of square approximation error $E\left\{|y - \varphi(u)|^2\right\}$ is minimized. The structures of neural networks such as MLP and radial basis networks constitute the good candidate algorithms for determining this function, $\varphi(u)$ [NAS04]. Neural enhancement methods permit to effectively reduce the musical noise effect because of the ability of neural networks to provide a signal smoother estimate [SHA04, WAN98]. Since the artificial neural networks (ANNs) are able to make the approximation of any non linear function, they are appropriate for non linear transformations commonly employed in speech feature extraction such as Mel frequency, log spectrum and cepstral coefficients (MFCCs) [SHA04]. Different research works have employed ANNs for speech enhancement [JON96], [TAM90], [SEO93]. Classical ANNs, in spite of their ability of generalisation, can't easily model the temporal behaviour of speech signal: the single way for addressing this issue consists at employing a windowed input of time-neighbouring features [SHA04]. On the other hand, the recurrent neural network, RNN has the ability to deal naturally with variable length of speech signal and can detect long term contextual effects over time, which can be helpful for a better speech enhancement [SHA04]. Hence the neural network needs dynamic properties (recurrent connections) so that it is able to respond to the temporal behaviours. As an example having these proprieties, we can mention the Elman neural network which is used in our speech enhancement technique. This choice is based on the fact that the Elman neural network has been successful for temporal association of speech signal moreover it can be easily trained with standard back-propagation (BP)[ SEO93, ELM88].

## 3. New proposed Speech enhancement technique

In this paper, we propose a new speech enhancement technique based on BWT by using the Elman neural network. This technique is inspired from the technique introduced by C.A. Medina & al [MED03]. It consists in applying the discrete wavelet transform to the noisy speech signal and then applies one neural network for each decomposition level for the purpose of estimating the appropriate value of threshold to be applied to that level. Our technique consists in applying the bionic wavelet transform to the noisy speech frame, *x* and to the clean speech one, *s*:

$$wtb = BWT(x) \tag{15}$$

$$wtc = BWT(s) \tag{16}$$

The frame length is equal to 512 with 256 as overlap. $wtb$ and $wtc$ are two matrixes having a size of $30 \times 512$; The application of the BWT is done by the previous mention modification ($N$=30). The Elman neural network is trained by a set of pairs $(P,T)$ where $P$ is the input of the Elman neural network and is equals to $wtb$. $T$ represents the target or the desired output of the Elman neural network. It is a matrix having the same dimension of $P$. Each coefficient of $T$, $T(i,j)$, $1 \le i \le 30, 1 \le j \le 512$, is an ideal threshold and is chosen to be:

$$T(i,j) = \begin{cases} sigma \cdot \dfrac{|wtb(i,j)|}{|wtc(i,j)|} & if \ |wtc(i,j)| > 0 \\ \sqrt{2\log(n)} & if \ not \end{cases} \tag{17}$$

where $n$ is the length of the frame and $sigma$ is the noise level. it is selected to be:

(14)

$$sigma = mad(|wtb(1,:)|)/0.6745 \tag{18}$$

The employed Elman neural network in this work is constituted by two layers, one hidden having four unities with 'tansig' as an activation function and one output having thirty unities with 'purelin' as an activation function. The figure 3 illustrates the general architecture of the Elman neural network [STE06].

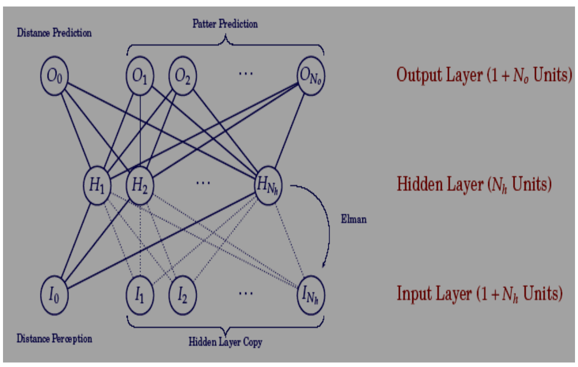


**Figure 3.** *General architecture of Elman neural network.*

The used backpropagation training is 'traingdx'. The performance function is 'mse'. Training the Elman neural networks is done for the purpose to generate a sequence of target vectors when this network is presented with a given sequence of input vectors. In learning phase of our Elman neural networks, we use a set of 114 speech signals taken from Timit database. In the phase of test we test we use 36 speech signals taken from our Arabic database and also from Timit database. The figure 4 illustrated a training example of our Elman neural network.

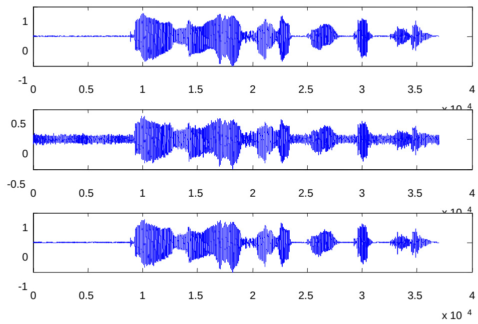

**Figure 4**. *An example of training of the Elman neural network.*

## 5. Results and evaluations

To illustrate the performance of the proposed enhancement techniques, we tested them in different in various noisy conditions, taken from Noisex-92 database: white Gaussian noise, F16 cockpit noise and Volvo car noise with different values of signal to noise ratio, SNR. These values are -5dB, 0dB, 5dB, 10dB and 15dB. To evaluate the performance of the speech enhancement algorithms, it is necessary to identify the differences and similarities in perceived quality and subjectively measured intelligibility. Speech quality is an indicator of the processed speech signal "naturalness". Speech signals intelligibility is the amount measure of speech information present in the signal that is responsible for covering what the speaker is saying. Tests of performance evaluation can be done by objective quality measures or subjective quality measures. First, an objective signal to noise ratio SNR measure is employed and then we make some listening tests.

### 5.1. Objective evaluation

Objective measures are based on mathematical comparison between the original and processed speech signals. The measure of the signal to noise ratio, SNR is one of the most extensively used. As the name suggests, it is computed as the ratio of the signal to noise powers in decibels:

$$SNR = 10 \cdot \log_{10}\left[\frac{\sum_n s^2(n)}{\sum_n [s(n) - \hat{s}(n)]^2}\right] \tag{19}$$

where $s$ and $\hat{s}$ are respectively the clean and the clean and the enhanced speech signals.

**Table 1.** *Case of wite noise.*

| SNR (dB) | Our proposed denoising technique | Denoising by BWT | Spectral subtraction technique |
|---|---|---|---|
| -5 | 6.6979 | 0.7719 | 4.4555 |
| 0 | 11.4061 | 4.1682 | 6.8233 |
| 5 | 14.4490 | 8.0418 | 10.3593 |
| 10 | 17.3100 | 11.8941 | 14.3707 |
| 15 | 19.2095 | 14.8939 | 16.7257 |

**Table 2.** *Case of Volvo noise.*

| SNR (dB) | Our proposed denoising technique | Denoising by BWT | Spectral subtraction technique |
|---|---|---|---|
| -5 | 9.2559 | 5.9133 | 6.9604 |
| 0 | 11.9527 | 9.6057 | 10.0438 |
| 5 | 14.5012 | 11.8789 | 13.1521 |
| 10 | 16.8791 | 13.9853 | 15.5135 |
| 15 | 18.8058 | 15.4581 | 17.1486 |

**Table 3.** *Case of F16 noise.*

| SNR (dB) | Our proposed denoising technique | Denoising by BWT | Spectral subtraction technique |
|---|---|---|---|
| -5 | 4.8200 | 1.5479 | 5.3599 |
| 0 | 8.1349 | 6.3499 | 8.3660 |
| 5 | 11.6963 | 11.3092 | 11.2194 |
| 10 | 16.2772 | 13.5167 | 14.7832 |
| 15 | 19.4656 | 16.0224 | 15.2019 |

### 5.2. Subjective evaluation

A subjective evaluation is done by making listening tests and computation of the recognition rate which expressed by:

$$\frac{number \ \ of \ \ the \ \ recognized \ \ words}{total \ \ number \ \ of \ \ the \ \ employed \ \ words} \times 100 \qquad (20)$$

In the listening tests, listeners are listening to the enhanced speech signals and say what is said for every pronounced sentence. and this for the three different cases: white noise, car noise and F16 noise. The figures 5, 6 and 7 illustrate the recognition rate vs the SNR.

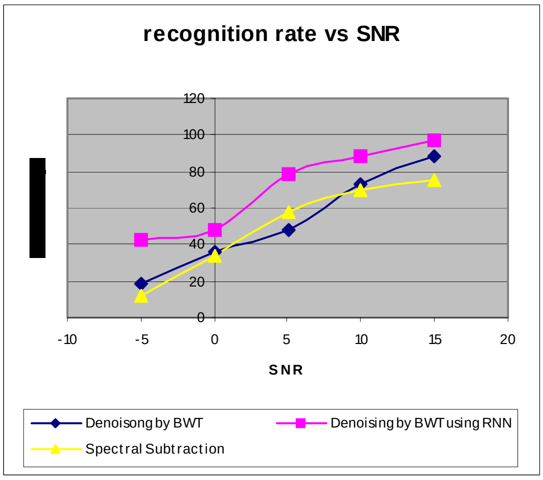


**Figure 5.** *Case of White noise.*

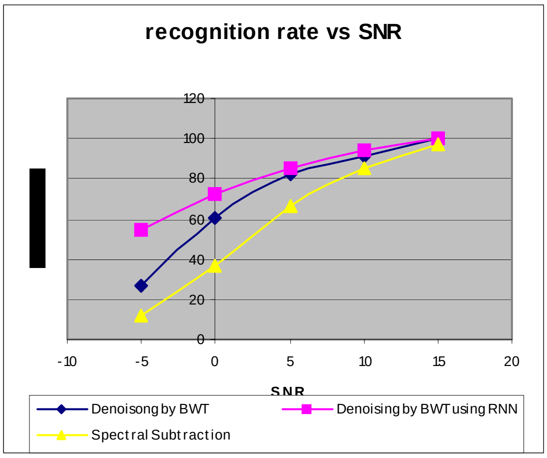


**Figure 6.** *Case of Volvo noise.*

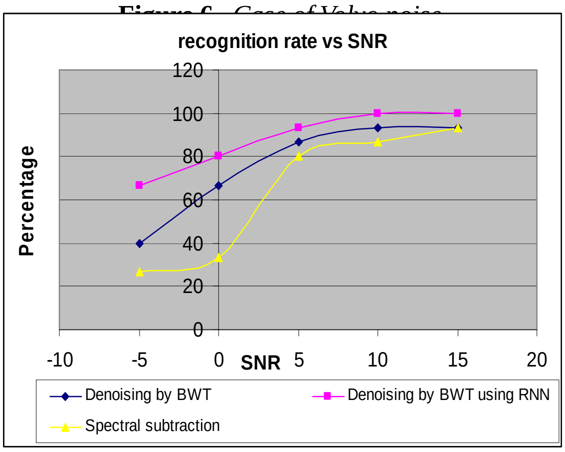


**Figure 7.** *Case of F16 noise.*

Figures 5, 6 and 7 illustrate the three curves representing the recognition rate vs the SNR in cases of white noise, car noise and F16 noise. This curves show clearly that listening tests are in favour of our proposed denoising method based on the Elman neural network and bionic wavelet transform. Our proposed technique presents the best scores when compared to the two denoising techniques based on BWT and spectral subtraction.

### 5.3. Speech signal representation

The figures, 7, 8 and 9 show the efficiency of our proposed method based on BWT and using the Elman neural networks. In fact a great amount of noise was suppressed while preserving the enhanced speech signal.

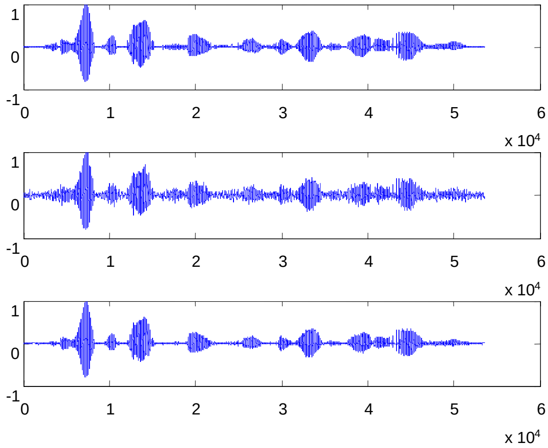


**Figure 8.** *Speech signal corrupted by Car noise with SNR=5dB.*

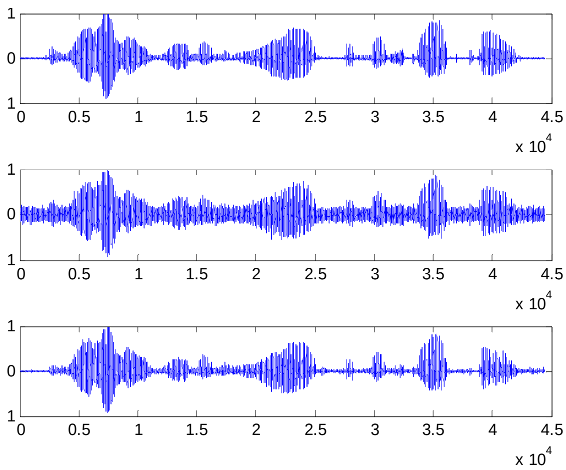


**Figure 9.** *Speech signal corrupted by White noise with SNR=5dB.*

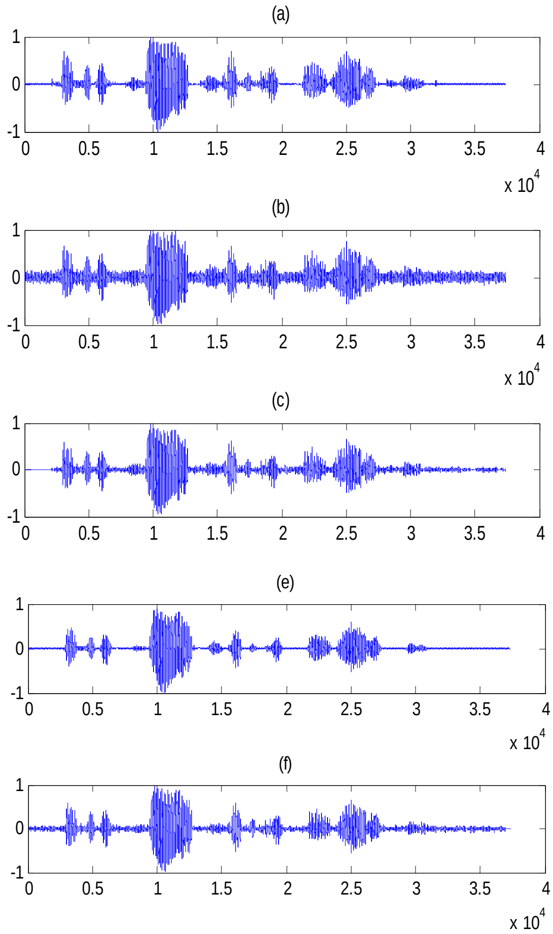


**Figure10.** *(a) Clean speech, (b) Speech signal corrupted by F16 noise with SNR=5dB, (c) Enhanced speech by our proposed technique, (e) Enhanced speech by spectral subtraction, (f) Enhanced speech by bionic wavelet transform.*

The figures 11, 12 and 13 represent respectively the spectrograms of clean speech, the noisy speech and the enhanced where the noisy speech signal is obtained by corrupting the clean speech by F16 noise with SNR equals to 5dB.

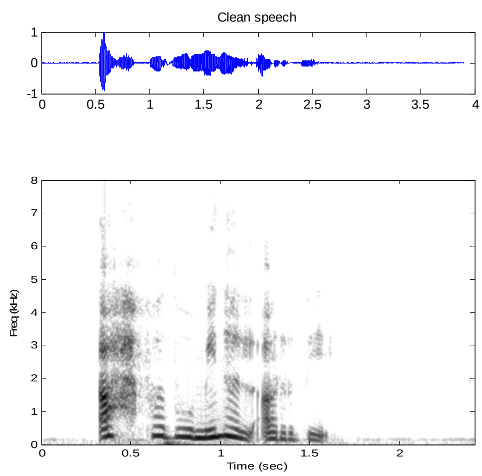


**Figure 11**. *Clean speech spectrogram.*

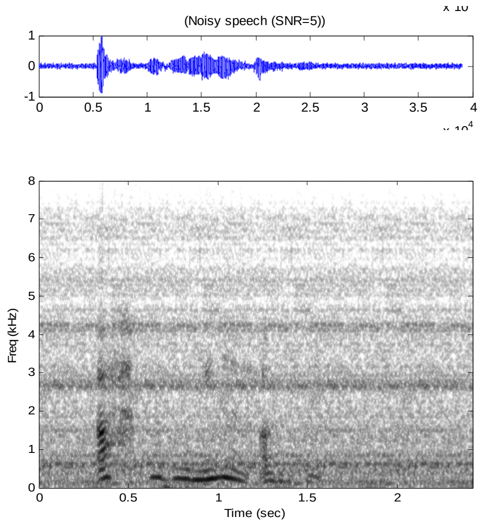


**Figure 12**. *Noisy speech spectrogram (SNR=5dB).*

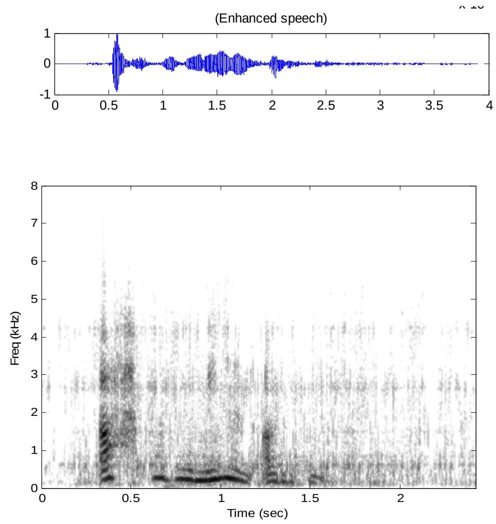


**Figure 13.** *Enhanced speech spectrogram.*

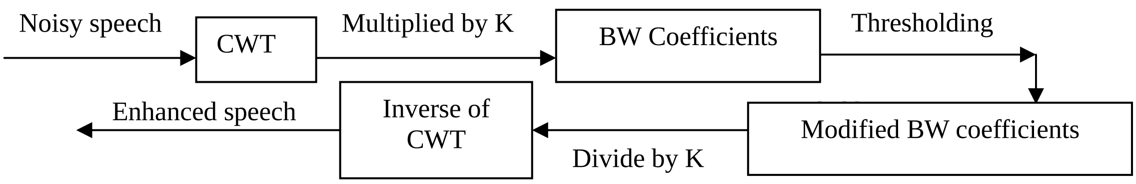


**Figure14.** *Speech enhancement approach by bionic wavelet transform, BWT.*

## 6. Conclusion

A new technique for speech enhancement using the bionic wavelet transform. For thresholding the bionic wavelet coefficient, we employ the Elman neural network. This is done for determining the set of convenient thresholds. Results obtained from SNR and listening tests show the performance of our proposed denoising method when compared with the two denoising techniques based on BWT and spectral subtraction.

## References

[VAN93] Van Compernolle, D. (1993). "Speech Enhancement for Applications in Communication and Recognition." Revue HF Tijdschrift (Special Issue: Speech Processing for Telecommunications) **1993**(1-2-3): 99-10.

[DON95] Donoho, D. L. (1995). "Denosing by soft thresholding." IEEE Trans. on Information Theory **41**(3): 613-627.

[SEO97] Seok, J. W. (1997). Speech enhancement with reduction of noise components in the wavelet domain. International conference on audio,speech and signal processing of IEEE.

[BAH01] Bahoura, M. a. J. R. (2001). "Wavelet speech enhancement based on the Teager energy operator." IEEE Signal Processing Letters **8**(1): 10-12.

[SHE01] Sheikhzadeh, H. (2001). An improved wavelet-based speech enhancement system. Eurospeech.

[CHA02] Chang, S., Y.Kwon (2002). Speech enhancement for non-stationary noise enviroment by adaptive wavelet packet. International conference on audio,speech and signal processing of IEEE.

[XIA03] Xiaolong Yuan (2003), B.S.E.E. A THESIS, « Auditory Model-based Bionic Wavelet Transform for speech enhancement. Electrical and computer engineering.

[MAL89] Mallat, S. (1989). "A theory for multiresolution signal decomposition." IEEE Trans. on Pattern Anal. Machine Intell.(11): 674-693.

[MED03] C.A. Medina, A. Alcaim et J.A. Apolinário, Jnr, ‘‘wavelet denoising of speech using neural networks for threshold selection’’, IEE 2003, pp. 1869-1871, 2003.

[DON94] D. Donoho et I. M. Johnstone. ‘‘Ideal spatial Adaptation via Wavelet Shrinkage’’ Biometrika, 41. pp. 425-455, 1994.

[JOH97] I. M. Johnstone and B. W. Silverman, “Wavelet threshold estimators for data with correlated noise”, *J. Roy. Statist. Soc. B*, vol. 59, pp. 319-351, 1997.

[YAO01] Yao, J., Y.T.Zhang (2001). "Bionic wavelet transform: a new time-frequency method based on an auditory model." IEEE Trans. on Biomedical Engineering 48(8): 856-863.

[MIC07] Michael T. Johnson[a,*] , Xiaolong Yuan[b], Yao Ren[a,1], ‘‘Speech signal enhancement through adaptive wavelet thresholding’’, in conference Elsevier 2007, pp 123-133.

[GIG93] Giguere, C. 1993. Speech processing using a wave digital filter model of the auditory periphery. Ph.D., University of Cambridge, UK.

[GIG94] Giguere, C., Woodland, PC., 1994. A computational model of the auditory periphery for speech and hearing research. J. Acoust. Soc. Amer. 95 (1), 331-342.

[ZHE99] Zheng, L., Zhang, Y., T., Yang, F.-S., Ye, D.-T., 1999. Synthesis and decompossition of transient-evoked otoacoustic emissions base don an active auditory model. IEEE Trans. Biomed. Eng. 46 (9), 1098-1106.

[YAO02] Yao, J., Zhang, Y.T., 2002, ‘‘The application of bionic wavelet transform to speech signal processing in cochlear implants using neural network simulations’’. IEEE Trans. Biomed. Eng. 49 (11), 1299-1309.

[NAS04]: Nasir Mahmood, Asif PN and S K Hasnain PN, ‘‘A Unified Approach Using Neural Networks Efficient Algorithms in Audio Signal Processing’’, in conference IEEE, INMIC 2004, pp 26-31.

[SEO93] Seokyong Moon, Jenq-Neng Hwang, ‘‘Coordinated training of noise removing networks’’, in ICASSP-93, vol. 1, pp 573-576.

[SHA04] Shala Parveen and Phil Green ‘‘Speech enhancement with missing data techniques using recurrent neural networks’’, ICASSP2004, pp 733-736.

[WAN98] Wan et al (1998). ‘‘Networks for speech enhancement.’’ In Handbook of neural networks for speech processing, Boston, USArtech House, Eds. Shigeru Katagiri, 1998, ISBN: 0-89006-954. p. 541-541.

[JON96] Jones, M. and Sridharan, S. (1996). ‘‘Improving the effectiveness of existing noise reduction techniques using neural networks.’’ In Proceedings of Fourth International Symposition on Signal Processing and its Applications, volume1, p. 387-388.

[TAM90] Tamura et al. (1990). ‘‘Improvements to the noise reduction neural networks.’’ Proc. ICASSP 1990, p. 825-828.

[ELM88] J. L. Elman, D. Zipser. Learning the hidden structure of speech. J. Acoustic Soc. Am., Vol. 83, No.4, pp. 1615-1626, April 1988.

[STE06] Stefano Zappacosta, Glanluca Baldassarre and Stefano Nolfi, ‘‘Elman Neural networks and integration for object recognition’’, July 2006, pp 1-8.