Journal of Engineering and Applied Sciences 2 (8): 1326-1331, 2007
ISSN: 1816-949X

Bionic Wavelet Transform Based Speech Enhancement

Mourad Talbi and Adnen Cherif
Laboratory of Signal Processing, Science Faculty of Tunis, 1060 Tunis

Abstract: In this study, we are interested in speech enhancement technique using Bionic Wavelet Transform (BWT). This technique is compared with conventional denoising methods such as those based on Discrete Wavelet Packet Transform (DWPT) using spectral entropy and spectral subtraction. This comparison is based on signal to noise ratio, SNR and listening tests. In experimental part, the speech signals are corrupted with several types of noises such as White and car ones with different values of SNR. The performance of the bionic wavelet transform has been shown by the obtained results.

Key words: Bionic wavelet transform, speech enhancement, spectral entropy, DWPT, SNR

INTRODUCTION

In speech processing, one of most important problem is the existence of background noise. This noise is generally additive, multiplicative and convolutive. In this research we are interested in additive noise (Van, 1993):

$$y(t) = s(t) + n(t) \quad (1)$$

So we have to reduce the influence of n(t) and enhance the quality of speech signal by searching an optimal estimate ŝ(t) preferred by a human listener. Denoising speech signal has forever been non easy problem for researchers. A background noise eliminating is almost unfeasible and speech distortion is expected. Conventional algorithms of speech denoising comprise Wiener filtering, spectral subtraction and techniques based on microphone array. The wavelet transform differentiates itself in non-stationary signals analysis such as speech. Donoho (1995) introduced a denoising technique based on wavelet shrinkage. This technique has proved its efficiency in denoising signals contaminated by additive white noise. Recently, many efforts employing wavelet shrinkage, have been made by Seok (1997), Bahoura (2001), Sheikhzadeh (2001) and Chang (2002). Wavelet analysis lies on the modelling of pre-post perceptual periphery, that's why some efforts have been made for employing this processing tool in speech signal enhancement. This approach associated multi-resolution with non-linear filtering. In this study, we are interested in speech denoising based on bionic wavelet transform BWT. This approach was introduced by Xiaolong Yuan. A comparison will be and we will make a comparison between three speech enhancement techniques. One is based on BWT, the second is based on the DWPT using the entropy spectral (Sungwook *et al.*, 2000) and the third is based on the spectral subtraction.

DENOISING BY WAVELETS

Classical signals denoising techniques used Fourier analysis under the hypothesis that the noise is manifested principally as high-frequency oscillations. Bearing this in mind, a signal is decomposed into sinusoidal waveforms of different frequencies and only low-frequency components are left in the denoised signal. The wavelet based enhancement signal, supposes that the signal analysis at different resolutions might ameliorate the separation of the true underlying signal from noise.

Wavelet multi-resolution analysis: The wavelet multi-resolution analysis (Mallat, 1998, 1989) lies on the scaling function $\phi(t)$ and the corresponding mother wavelet $\psi(t)$. The scaling function and the mother wavelet are localized both in time and frequency domain. By shifting and dilatation of the wavelet mother and scaling function, we obtain:

$$\phi_{j,k} = 2^{-j/2}\phi(2^{-j}t - k) \quad (2)$$

$$\psi_{j,k} = 2^{-j/2}\psi(2^{-j}t - k) \quad (3)$$

Where j designates the scale or corresponding resolution of the functions and k localises the functions in time. On each scale j, the functions $\phi_{j,k}(t)$, $k \in Z$ and $\psi_{j,k}(t)$, $k \in Z$ onstitute an orthonormal

Corresponding Author: Mourad Talbi, Laboratory of Signal Processing, Science Faculty of Tunis, 1060 Tunis

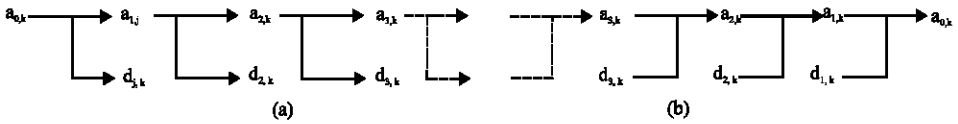

Fig. 1: Mallat pyramidal scheme of decomposition (a) and reconstruction (b)

basis in the spaces of the square integrable functions L^2 (IR) (Mallat, 1989, 1998). Let a signal $x(t) \in L^2$ (IR), we have:

$$x(t) = \sum_{n\in Z} a_{J,k}\phi_{J,k}(t) + \sum_{j\le J}\sum_{k\in Z} d_{j,k}\psi_{j,k}(t) \quad (4)$$

Where the first term represents the approximation on the scale j and the second term the details on the scale j and all finer scales. In the wavelet multi-resolution analysis, the wavelet coefficients $a_{j,k}$ of the approximation and the wavelet coefficients $d_{j,k}$ of the details on adjacent scales are related by the decomposition:

$$a_{j,k} = \sum_{n\in Z} h^*_{n-2k} a_{j-1,n} \quad (5)$$

$$d_{j,k} = \sum_{n\in Z} g^*_{n-2k} a_{j-1,n} \quad (6)$$

as well as by the reconstruction:

$$a_{j-1,k} = \sum_{n\in Z} (h_{k-2n} a_{j,n} + g_{k-2n} d_{j,n}) \quad (7)$$

Where * designates the complex conjugates. The pyramidal schemes of decomposition and reconstruction (Mallat, 1989, 1998) are illustrated in Fig. 1:

The coefficients h_n and g_n employed in decomposition and reconstruction formulae are:

$$h_n = \int_{-\infty}^{+\infty} \phi(t)\phi_{-1,n}(t)dt \quad (8)$$

$$g_n = (-1)^n h^*_{1-n} \quad (9)$$

h_n and g_n are named conjugate mirror filters.

Wavelet denoising: The Discrete Wavelet Transform (DWT) is linear and orthogonal, thus transforming white noise noise in time space to white noise in the space of the wavelet coefficients. 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 (1994) to propose thresholding denoising approach. This approach is summarized as follow:

- Apply the discrete wavelet transform to noisy signal:

$$y = x + b \quad (10)$$

Where b is a Gaussian white noise having σ^2 as a variance and x is the clean signal. In wavelets domain, we have:

$$W_y = W_x + W_b \quad (11)$$

- The searched signal is obtained by applying the inverse transform W^{-1} to the thresholded wavelets coefficients vector Y_{TH}:

$$x = W^{-1}Y_{TH} \quad (12)$$

The thresholding is non linear and generally is soft or hard. This approach of thresholding based denoising, based on the fact that the clean signal energy is concentrated in small number of great wavelets coefficients however the noise contaminates all coefficients.

For handling gaussian white noise, Donoho (1995) had employed an universal threshold which is expressed as follow:

$$thr = \sigma \cdot \sqrt{2\log(N)} \quad (13)$$

Where σ is the noise standard deviation and N is the noisy signal length. This standard deviation is defined as:

$$\sigma = MAD/0.6745 \quad (14)$$

With MAD is the absolute median estimated in the coarsest scale of wavelet tree.

For handling a correlated noise, Johnstone and Silverman (1997) had proposed a level dependent threshold which is defined as:

$$thr_j = \sigma_j \sqrt{2\cdot\log(N)} \quad (15)$$

With $\sigma_j = MAD_j/0.6745$ and MAD_j designate the absolute median estimated at scale *j*.

Sungwook *et al.* (2000) had proposed to employ a node dependent threshold. This threshold is applied to each node of the wavelet packet tree and is expressed as follow:

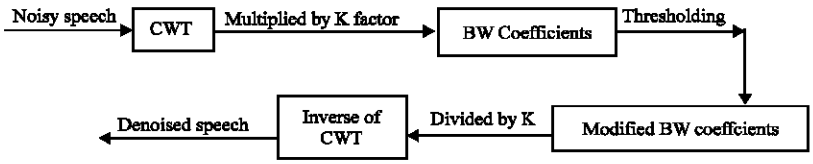

Fig. 2: The approach of denoising by bionic wavelet transform

$$thr_{j,k} = \sigma_{j,k}\sqrt{2\cdot\log(N)} \quad (16)$$

Where

$$\sigma_{j,k} = MAD_{j,k} / 0.6745 \text{ and } MAD_{j,k}$$

is the absolute median estimated at the scale *j* and subband *k*.

Thresholding limitation: The wavelet based denoising technique doesn't require a speech or a noise model and can be utilized to a large class of signals. However, a general thresholding of wavelets coefficients doesn't ensure a good performance as obtained by bionic wavelet transform introduced in Xiaolong. The latter owns a better propriety of signal energy concentration and time-frequency selectivity. This leads an efficient thresholding performance.

BIUNIQUE BIONIC WAVELET TRANSFORM

Yao and Zhang (2001) had proposed the Bionic Wavelet Transform (BWT) as a new time-frequency technique and this by referring to the perceptual model. The term "bionic" means that is guided by an active biological mechanism.

The difference between (BWT) and the Continuous Wavelet Transform (CWT) consists in 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. This is the mother wavelet that makes adaptive the transform (CWT). However, the mechanism of active control in the human auditory model that adjusts the mother wavelet associated to (BWT) according 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. This thing is ensured by introducing a time varying function T, in the mother wavelet expression:

$$\varphi(t) = \frac{1}{T\sqrt{a}}\tilde{\varphi}\left(\frac{t}{T}\right)\cdot\exp(j\omega_0 t) \quad (10)$$

Where *a* is the scale facto rand $\tilde{\varphi}(t)$ designate the envelop of φ.

In this expression, the role of first factor T multiplying $\sqrt{a}$ is to ensure that the energy remains the same for each mother wavelet. The role of second factor T is to adjust the envelop

$\tilde{\varphi}(t)$ without adjusting the central frequency of $\varphi(t)$. When applying the Classical Wavelet Transform (CWT) with the new adaptive mother wavelet on a signal x(t), we obtain the Bionic Wavelet Transform (BWT):

$$BWT_x(a,\tau) = \frac{1}{T\sqrt{a}} \int x(t)\,\tilde{\varphi}^*\left(\frac{t-\tau}{Ta}\right)\exp\left(-j\omega_0\left(\frac{t-\tau}{a}\right)\right)dt \quad (11)$$

Denoising by bionic wavelet transform: Xiaolong gives the speech signal denoising algorithm using the bionic wavelet transform. The Fig. 2 illustrates this algorithm:

Thus, for obtaining the bionic wavelet coefficients are obtained by multiplying those of Continue Wavelet Transform (CWT), by a K factor defined in Xiaolong.

IMPLEMENTATION

In this sutdy we make a comparison between denoising techniques using the Bionic Wavelet Transform (BWT), the Discrete Wavelet Packet Transform (DWPT) and the spectral subtraction. In denoising method based on (BWT), we use a level depending threshold which is defined by Eq. 15 and the employed thresholding is soft. The level depending threshold is also used in the denoising method based on the (DWPT). In this method, we use spectral entropy for estimating level noise instead of Median Absolute Deviation (MAD).We also eliminate some coefficients which are considered belonging to noise. The remained wavelet packet coefficients are thresholded. Elimination of these coefficients is based on the fact that the speech signal can be classified as voiced, unvoiced and silence (Jafer and Mahdi, 2003). Voiced speech is quasi-periodic in the time domain and harmonically structured in the frequency range, less than 1kHz, whereas the energy of unvoiced speech is usually concentrated at the high end of frequency scale ($\geq$3kHz). If we want to get a discrimination of the voiced and unvoiced sounds we must derive benefit from the information contained in

those bands where the voiced sound or the unvoiced sound is dominant compared with the other sounds. It is known that most of the speech signal power is contained around the first formant. The statistical results for many vowels of adult males and females indicates that the first formant frequency doesn't exceed 1kHz and doesn't below 100Hz approximately. In addition, pitch frequency lies in normal speech between 80 and 500Hz.

RESULTS AND DISCUSSION

In this research, we use ten Arabic sentences sampled at 16 kHz and pronounced by female voice. These sentences are corrupted by Volvo and White noises with signal to noise ratio varies from -5-15dB.

Signal to noise ratio computation: In the following table, we give the values of the signal to noise ratio after denoising:

Table 1: Speech signal corrupted by Volvo noise

SNR	Denoising by DWPT	Denoising by BWT	Denoising by spectral subtraction
-5	0.5003	6.9059	7.9914
0	5.2792	10.2984	10.8758
5	9.5202	12.2924	12.8477
10	12.7577	13.0935	14.8147
15	14.4355	13.3456	15.9747

Table 2: Speech corrupted by white noise

SNR	Denoising by DWPT	Denoising by BWT	Denoising by spectral subtraction
-5	-2.3771	0.0347	8.7618
0	2.5971	3.9630	11.9686
5	7.2187	8.5728	14.1584
10	11.1970	11.9589	15.3463
15	13.5665	13.1728	16.4229

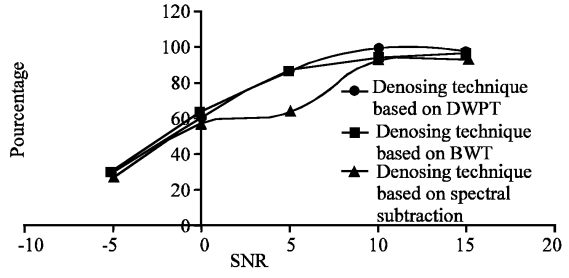

Fig. 3: Female voice corrupted by Volvo noise

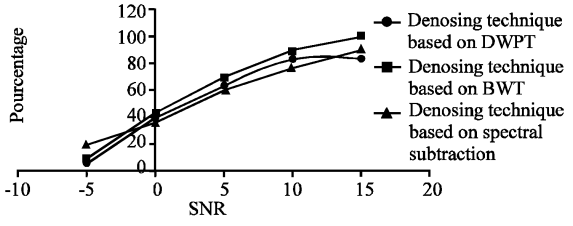

Fig. 4: Female voice corrupted by white noise

Table 1 and 2 show that the results obtained by using the denoising technique based on spectral subtraction, are better than those obtained by the two denoising methods based on DWPT and BWT. denoising method based on spectral subtraction improves the signal to noise ratio for all values taken by SNR (-5-15dB). However, in case of female voice, the two denoising techniques based on BWT and DWPT improve the signal to noise ratio when the SNR varies from -5-10dB and we have the opposite when the SNR is equals to 15dB.

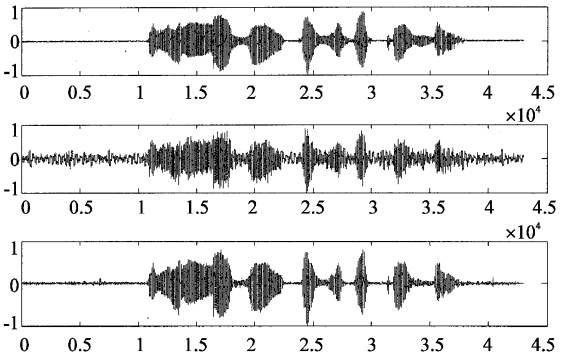

Fig. 5: (BWT) based denoising speech signal corrupted by Volvo noise (SNR=5dB)

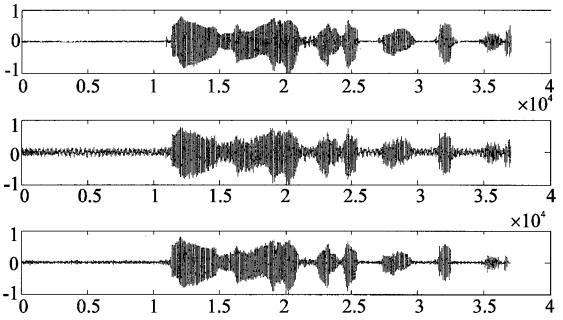

Fig. 6: (BWT) based denoising speech signal corrupted by factory noise (SNR=10dB)

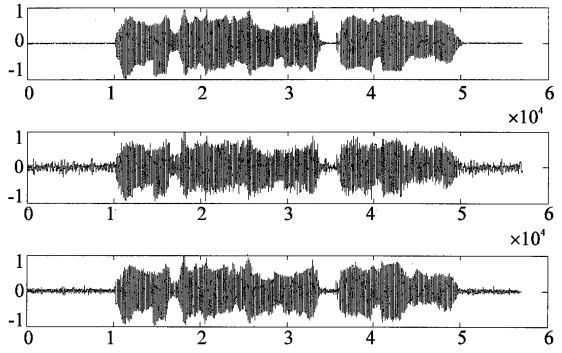

Fig. 7: (DWPT) based denoising speech signal corrupted by Volvo noise (SNR=10dB)

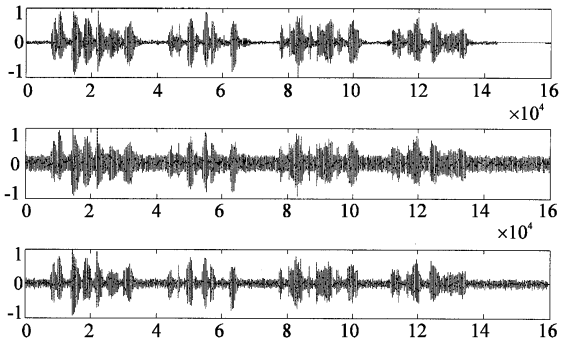

Fig. 8: (DWPT) based denoising speech signal corrupted by F16 noise (SNR=5dB)

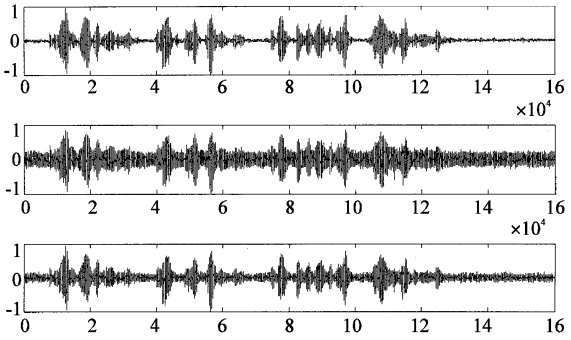

Fig. 9: (BWT) based denoising speech signal corrupted by White noise (SNR=5dB)

Listen tests: Figure 3 and 4 illustrates the variation of recognition percentage vs the signal to noise ratio SNR.

The curves of the Fig. 3 and 4, show that the two denoising techniques based on BWT and (DWPT) are better than the spectral subtraction based denoising method. Thus the spectral subtraction introduces more degradation on the signal after enhancement. These curves show also that the denoising technique employing BWT is practically better than the denoising method based on the DWPT using spectral entropy.

Time representation: Figure 5-9 represent the clean, the noisy and enhanced signals.

The figures illustrate the time representation of the clean speech, the noisy speech and the enhanced speech. These figures shows the superiority of the denoising technique based on BWT. This technique is very efficient in speech denoising. In fact an important amount of noise was suppressed from non speech segments and the noise is reduced from speech ones while preserving the majority of the speech signal. We remark that there is a little difference between the enhanced speech signal and the clean one. Thus there isn't any sever distortion of the denoised signal.

CONCLUSION

In this study, a denoising method based on Bionic Wavelet Transform (BWT), has been developed under Matlab and compared with conventional thresholding technique using spectral entropy and spectral subtraction based denoising technique. The results obtained from the computation of the signal to noise ratio, SNR show that the three denoising techniques improve the SNR. The computation of the SNR show in this case that the technique based on spectral subtraction is the best. BWT based denoising method is better the DWPT based denoising one. When speaking about intelligibility of the enhanced speech signal, the listening tests show that the two denoising techniques based on BWT and DWPT are better than the spectral subtraction based denoising one.

REFERENCES

Bahoura, M.A., 2001. Wavelet speech enhancement based on the Teager energy operator. IEEE. Signal Processing Lett., 8: 10-12.

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

Donoho, D. and I.M. Johnstone, 1994. Ideal spatial Adaptation via Wavelet Shrinkage. Biometrika, 41: 425-455.

Donoho, D.L., 1995. Denosing by soft thresholding. IEEE. Trans. Inform. Theory, 41: 613-627.

E. Jafer and A.E. Mahdi, 2003. Wavelet-based Voiced/ Unvoiced classification algorithm. 4th EURASIP Conf., Croatia, 2: 667-672.

Johnstone, I.M. and B.W. Silverman, 1997. Wavelet threshold estimators for data with correlated noise. J. Roy. Statist. Soc. B., 59: 319-351.

Mallat, S., 1989. A theory for multiresolution signal decomposition. IEEE TRans. Pattern Anal. Machine Intell., 11: 674-693.

Mallat, S., 1998. A wavelet tour of signal processing. New York, Academic Press.

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.

Sheikhzadeh, H., 2001. An improved wavelet-based speech enhancement system. Eurospeech.

Sungwook Chang, Y. Kwon, Sung-il Yang and I-jae Kim, 2000. Speech enhancement for non-stationary noise environment by adaptive wavelet packet IEEE.Tans., pp: 561-564.

Van Compernolle, D., 1993. Speech Enhancement for Applications in Communication and Recognition. Revue HF Tijdschrift (Special Issue: Speech Processing for Telecommunications), pp: 99-10.

Yao, J. and Y.T. Zhang, 2001. Bionic wavelet transform: A new time-frequency method based on an auditory model. IEEE. Trans. Biomed. Eng., 48: 856-863.