Simulation of Speech Denoising based on Voiced/ Unvoiced Decision by Using DWP

Basim Khalaf Jarullah, Izz Kadhum Abboud, Kareem Jabbar Tijil

Abstract— In this paper, a speech denoising system using wavelet packet thresholding algorithm is investigated. This method is based on thresholding the wavelet coefficients that depend on voiced/unvoiced detection. Wavelet threshold can be done by a standard deviation method for each frame by level dependent thresholding using semisoft threshold in the additive Gaussian noise channel. The results of simulation indicate that using Discrete Wavelet Packet (DWP) in speech denoising application provides a quality better than Discrete Wavelet Thresholding (DWT) and voiced/unvoiced detection enhanced the performance of the system.

Index Terms— DWP, DWT, Matlab.

I.INTRODUCTION

In many speech processing applications such as mobile communications, speech recognition and hearing aids, speech have to be processed in the presence of background noise. The performance of such applications is highly dependent on how much the noise is removed [1]. During the last decades, various approaches have been proposed to the problem such as spectral subtraction [2], wavelet based methods [1], hidden Markov modelling [3] and signal subspace methods [4].

The wavelet shrinkage is a powerful tool in denoising signal corrupted by noise. Speech denoising using wavelet transform is studied in [2,5].

Speech can be divided into numerous voiced and unvoiced regions. The energy of voiced regions is an order of magnitude larger than that of unvoiced regions. Therefore for additive white Gaussian noise, the SNRs of voiced regions are generally much higher than that of unvoiced regions; therefore, for a fixed threshold, enhancement in voice region is more effective than that in unvoiced region [6].

In this paper, the speech denoising system is introduced based on wavelet packet thresholding and voiced/unvoiced technique.

II. WAVELET PACKET TRANSFORM (WPT)

The wavelet Packet transform is a generalization of decomposition process that offers a richer range of probabilities of signal analysis. In wavelet analysis, a signal is split into an approximation and a detail. The approximation is then itself split into a second-level approximation and detail and the process are repeated. In wavelet packet analysis, the details as well as approximations can be split [1].

Manuscript received November 2013.

Basim Khalaf Jarullah, university of technology \Electrical and Electronic Engineering department. , Iraq.
Izz Kadhum Abboud, Mustansiriya University College of Engineering Computer And Software Engineering Department. , Iraq.
Kareem Jabbar Tijil, Institute of Technology Baghdad

In order to work directly with the wavelet transform coefficients, the relationship between the detailed coefficients at a given level in terms of those at previous level is used. In general, the discrete signal assumes the highest achievable approximation sequence, referred to as 0-th level scaling coefficients. The approximation and detail sequences at level j are given by [7]:

$$c_{j+1}(k) = \sum_m h_o(m-2k)\, c_j(m) \qquad (1)$$

And

$$d_{j+1}(k) = \sum_m h_1(m-2k)\, c_j(m) \qquad (2)$$

Equations (1) and (2) state that the approximation sequence at higher scale (lower level index), with the wavelet and scaling filters, $h_o(t)$ and $h_1(t)$ respectively, can be used to calculate the detail and approximation sequences at lower scales.

The scaling coefficients are related to wavelet coefficients by:

$$h_1(n) = (-1)^n h_o(N-n) \qquad (3)$$

Where N is a finite odd length of quadrature mirror filter.

Figure (1) shows 2-level wavelet packet decomposition tree. Let the function f(t) be a discretely sampled function. The decomposition of f(t) in the wavelet basis is done by recursive filtering with H_o and H_1 with down-sampling of a factor of two in each set. The coefficients $h_o(n)$ and $h_1(n)$, used to construct the set of scaling and wavelet basis, are low pass and high pass FIR filter coefficients respectively. $H_o=\{h_o(n)\}$ and $H_1=\{h_1(n)\}$. According to the Equation (3), H_1 is the reverse of H_o. [9] The symbol ↓2 is a down-sampled (Decimator) that takes a signal x(n) as input and produces an output of y(n)=x(2n), which means half of the data is discarded.

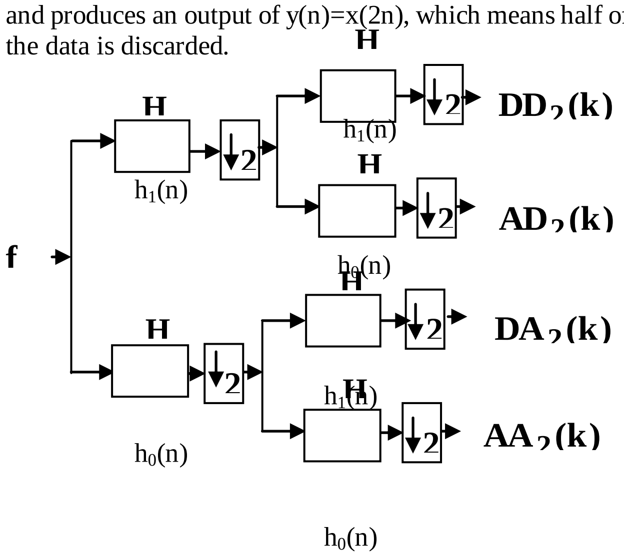

Figure (1) 2-Level wavelet packet decomposition tree

2-1 Wavelet Packet Reconstruction

The wavelet reconstruction process consists of upsampling and filtering. In the upsampling process, the input signal is stretched twice its original length and zeros are inserted in the even numbered samples. The inverse wavelet packet transform is illustrated in Figure (2).

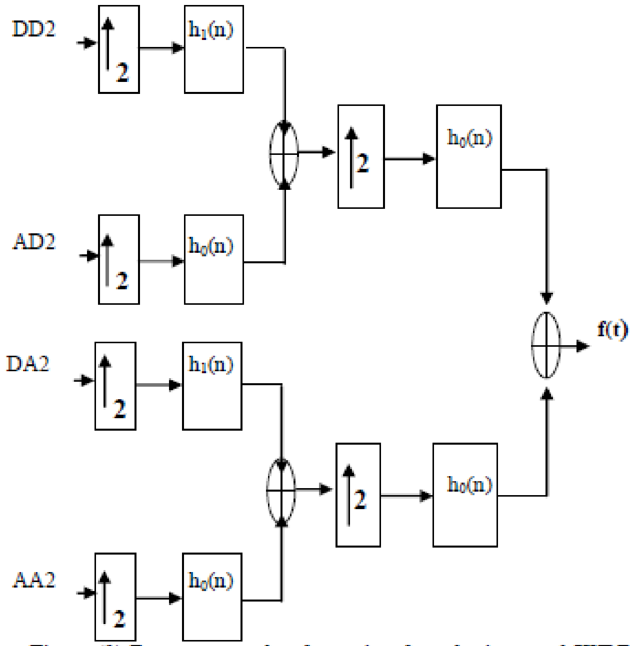

Figure (2) Two stage two-band tree signal synthesis tree of IWPT

Removing noise components by thresholding the wavelet coefficients is based on the fact that, the energy of speech signal is mostly concentrated in a small number of wavelet dimensions. So by thresholding the coefficients to set the smaller ones to zero, we can reduce the effect of noise components on speech signal components [1]. Let y be a finite length observation sequence of the signal x that is corrupted by zero-mean white Gaussian noise (n) [10]:

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

Let W(.) and W^{-1}(.) denote the forward and inverse wavelet transform operators. Let THR(., λ) denote the thresholding operator with threshold T. The practice of thresholding denoising consists of the following three steps:

1-Y=W(y) (5)

2- $\tilde{X}$=THR(Y,T) (6)

3-x=W^{-1}($\tilde{X}$) (7)

were $\tilde{X}$ representing the wavelet coefficients after thresholding. In this paper semisoft threshold is taken since it gives the best result as in study [11].

The semisoft-thresholding function is given by [11]:

$$THR(Y,T_1,T_2)=\begin{cases} 0 & |Y|\le T_1 \\ \operatorname{sgn}(Y)\dfrac{T_2(|Y|-T_1)}{T_2-T_1} & T_1<|Y|\le T_2 \\ Y & |Y|>T_2 \end{cases} \quad (8)$$

Where THR(Y, T_1, T_2) represents the output value after thresholding the wavelet coefficients and T_1 and T_2 denote lower and upper threshold respectively. The threshold value (T_1) is a node dependent threshold which is determined by [1]:

$$T_1=\sigma_{j,k}\sqrt{2\log(N_j.\log_2(N_j))} \quad (9)$$

$$\text{with } \sigma_{j,k}=\frac{MAD_{j,k}}{0.6745} \quad (10)$$

Where $MAD_{j,k}$ is median absolute deviation estimated on the scale j and sub-band k, and N_j is number of samples in scale j.

The threshold value (T_2) is given by [11]:

$$T_2=\sqrt{2}\,T_1 \quad (11)$$

3- Wavelet Packet thresholding system based on voiced/unvoiced decision

In order to prevent degradation of unvoiced region, first we must detect the unvoiced segments from the voiced ones. In this system, voiced/unvoiced detection is based on wavelet transform. Figure (3) shows the block diagram of the speech denoising algorithm.

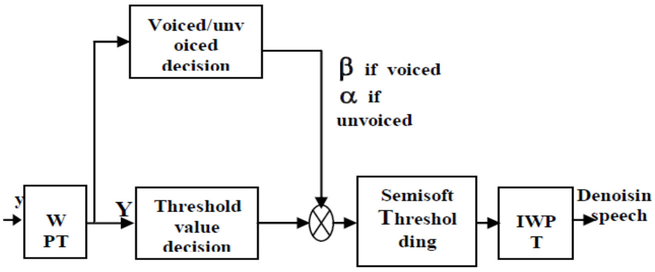

Figure (3) block diagram of speech denoising using level dependent threshold and voiced /unvoiced detection

The noisy speech signal is sectioned into frames (typical value of frame length is 256 samples). Then, the wavelet packet transform is taken from noisy speech, after that the threshold value decision is found using equations (9) and (11).

Noise energy is estimated from regions where the only noise is present. These regions can be detected using a voice activity detector (VAD). Using the noise energy estimated, a region can be classified into voiced / unvoiced by calculating the measure [6]:

$$\frac{ELS-ELN}{ES-EN}=\begin{cases} \text{voiced}, & >0.9 \\ \text{unvoiced}, & \text{otherwise} \end{cases} \quad (12)$$

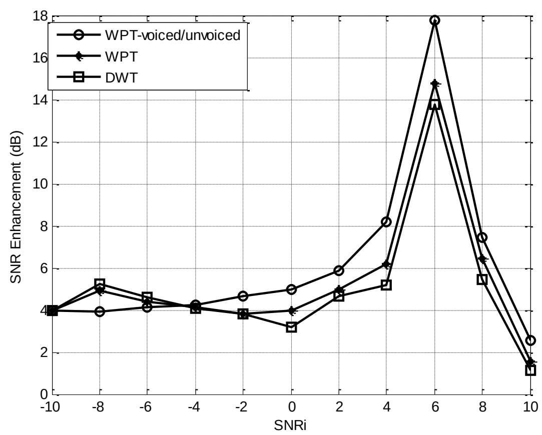

Figure (4) Comparison SNR enhancement results between WPT-voiced/unvoiced , WPT and DWT algorithms

Where the ELS, ELN, ES and EN are approximation energies of noisy speech, approximation energy of estimated noise, energy of noise speech and energy of the estimated noise in level 1 respectively.
Then the wavelet transform coefficients are filtered using semisoft threshold. Finally, the inverse wavelet packet transform is used to recover the clean speech signal. This procedure is repeated for each frame.

IV. OBJECTIVE MEASURES

The frequency signal-to-noise ratio (SNR) is the most widely used objective measures of speech quality. The SNR measure in frequency domain for the kth frame is defined by [11]:

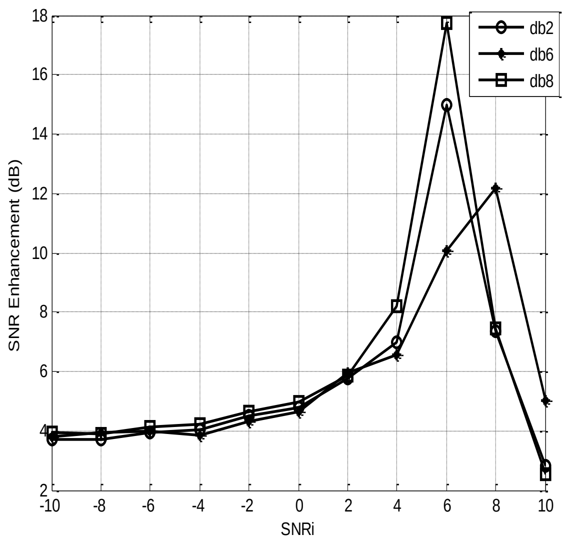

Figure (5) SNR enhancement results using semisoft-thresholding with different values of duabechies types

$$SNR_k = 10\log_{10}\frac{\sum_n |X_k(n)|^2}{\sum_n \left[X_k(n)-\tilde{X}_k(n)\right]^2}\ [\text{dB}] \qquad (13)$$

Where $X_k(n)$ is the DFT of the kth frame of the clean speech, and $\tilde{X}_k(n)$ is the DFT of the corresponding frame of the denoised speech signal. These SNR_k for different frames are averaged to give the overall SNR.
The SNR enhancement (in dB) is obtained by:

$$SNREnhancement = SNR_o - SNR_i \qquad (14)$$

Where SNR_o represents the output signal to noise ratio, and SNR_i represents the input signal to noise ratio.

V- SIMULATION RESULTS

Simulations of speech denoising using WPT over an AWGN channel was carried out. Various values of SNR_i from -10 dB to 10 dB is used for performance evaluation.
The sentence that is used in the recording is “one two three four five oh six”. The data is sampled at 8KHz using a computer sound blaster (in normal room conditions). The data samples are quantized into 16 bits.
The threshold value is modified based on the classification of analysis frames. If the analysis frame is classified as unvoiced frame, the threshold value is multiplied by constant α =2. If the voiced, the threshold value is multiplied by constant β =0.5.
Figure (4) shows Comparison of SNR enhancement results between WPT-voiced/ unvoiced, WPT and DWT algorithms
Figure (5) shows the SNR enhancement for WPT based on semisoft thresholding with daubechies wavelet of different length (db2, db4 , and db8).
Figure (6) shows the SNR enhancement for WPT based on db8 and different levels (2, 3,and 4).
Figure (7) shows the SNR enhancement for WPT at level 2 with different types of wavelet families (db 2, coiflet 2,and symmlet 2).

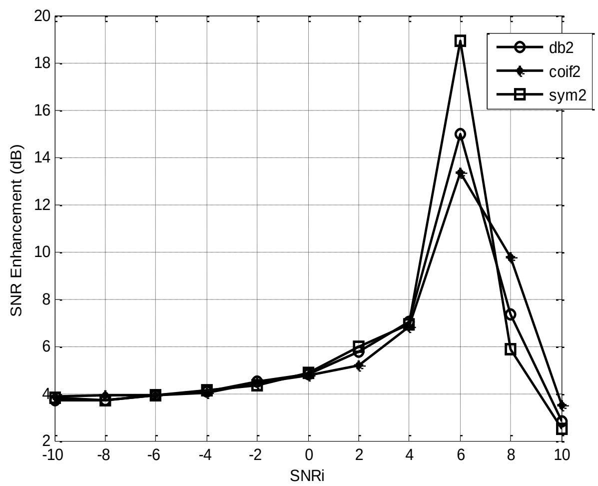

Figure (7) SNR enhancement results at level 2 with different types of wavelet families

VI- CONCLUSION

The following is a summary of the concluding remarks.
i- DWP gives better performance than DWT.
ii- DWP using voiced / unvoiced decision enhanced the performance of the WPT.
iii- The SNR enhancement is sensitive to changes in the wavelet families, length of filters and number of levels. Increasing the length of the filter does not necessarily enhance the performance of SNR. However, from the results db8 gives the best SNR enhancement. Symmlet family gives the best SNR enhancement when SNRi is about 6dB.

REFERENCES

[1]- Yasser Ghanbari and Mohammed Reza Karami,"A Modified Speech Enhancement System Based on the Thresholding of the wavelet Packets", 13th ICEE2005, Vol. 4, Zanjan, Iran, May 10-12, 2005.
[2]- Y. Ghanbari, M. Karami, B. Amelifard, "Improved Multi-band Spectral Subtraction Method for Speech Enhancement", Proc. Of the 6th IASTED Int. Conf. on Signal and Image Processing, USA, pp. 225-230, August 2004.
[3]- H. Sameti, H. Sheikhzadeh, Li Deng, R. L. Bernnan, "HMM-Based Strategies for Enhancement of Speech Signals Embedded in Nonstationary Noise", IEEE Transactions on Speech and Audio Processing, Vol. 6, No.5, September 1998.
[4]- M. Klein and P. Kabal, "Signal Subspace Speech Enhancement with Perceptual Post-filtering", Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, Orlando, FL, May 2002, pp. I-537-I-540.
[5]- Ahmed K. Hassan, “Speech Denoising Based Wavelet Transform”, M. Sc. Thesis , Mustansiria University, 2005.
[6]- Jong Kwan Lee and Chang D. Yoo, "Wavelet Speech Enhancement based on Voiced/ Unvoiced Decision", The 32nd International

Congress and Exposition on Noise Control Engineering Jeju International Convention Center, Seogwipo, Korea, August 25-28, 2003.
[7]- Yousef M. Hawwar, Ali M. Reza and Robert D. Turney , "Filtering (Denoising) in the Wavelet Transform Domain", University of Wisconsin-Milwaukee, Department of Electrical Engineering and Computer Science ,2000.
[8]- David L. Donoho and Iain M. Johnstone, "Adapting to Unknown Smoothness via Wavelet Shrinkage", Stanford Univesity, Division of Biostatistics, Technical Report No. 161, June 1993.
[9]- James F. Scholl, Jonathan R. Agre, Loren P. Clare, and Martin C. Gill, "A low Power Impulse Signal Classifier Using The Haar Wavelet Transform", Proc. SPIE, Sensors,C31, V.3577,pp. 136-145, 1999.
[10]- Xiaolong Yuan, B.S.E.E., "Auditory Model Based Bionic Wavelet Transform for Speech Enhancement", MSc. Thesis, Marquette University, May 2003.
[11]- Fadhil S. Hasan, "Performance of Discrete Wavelet Transform (DWT) based Speech denoising in Impulsive and Gaussian Noise", Engineering Journal in Mustansiriyah University ,2005.

Basim Khalaf Jarullah
Received his B.Sc. & M.Sc. degree in Electrical from the Faculty of Electrical and Electronic Engineering university of technology \Electrical and Electronic Engineering Department . His research interests include, Electronic circuit design, Wired & Wireless Network Administration.

Izz Kadhum Abboud
Received his B.Sc. in Electrical from the Faculty of Electrical and Electronic Engineering, University of Sarajevo in 1984, and received the M.Sc. degree in Electrical and Electronic engineering from Faculty of Electrical Engineering, University of Belgrade in 1986. Currently he is Assist Lecturer, Researcher and training supervisor, Dep. of Electronic in Institute of Technology Baghdad. His research interests include, Electronic circuit design .

Kareem Jabbar Tijil
Received his B.Sc. in Electrical from the Faculty of Electrical and Electronic Engineering, University of Sarajevo in 1984, and received the M.Sc. degree in Electrical and Electronic engineering from Faculty of Electrical Engineering, University of Belgrade in 1986. Currently he is Assist Lecturer, Researcher and training supervisor, Dep. of Electronic in Institute of Technology Baghdad. His research interests include, Electronic circuit design, Wired & Wireless Network Administration.