Military Technical College
Kobry El-Kobbah,
Cairo, Egypt

7th International Conference
on Electrical Engineering
ICEENG 2010

The Use of Biorthogonal Wavelets in Speech Enhancement

By

Ibrahim N. Abu-Isbeih*   Khaled Daqrouq*

Abstract:

This paper presents a new speech enhancement method. This method based on wavelet filters via multistage convolution with Reverse Biorthogonal Wavelets (RBW) in the high and low pass band frequency parts of the speech signal. In this method a speech signal is decomposed into two parts; a high-pass and a low-pass, the noise is then removed in each band individually in different stages via wavelet filters. The proposed method provides good results where it does not cut the speech information, which occurs when utilizing conventional thresholding. The proposed method is tested and the objective evaluation is used to compare the results with the other used methods. The new proposed method shows superiority over Donoho and Johnstone thresholding method and Birge-Massart thresholding strategy method.

Keywords:

Wavelet Filters, Speech Enhancement and Biorthogonal Wavelets

---

* Department of Communication & Electronics Engineering,
P. O. Box 1, Philadelphia University (19392), Jordan.

1. Introduction:

The principal aspiration of speech enhancement is to improve the quality and intelligibility of speech signal, as perceived by human hearing process. Speech enhancement is an essential procedure within the field of speech and signal processing, which impacts on many computers based speech and speaker recognition, coding and mobile communications. The quality of such applications is decidedly dependent on how much the noise is eliminated.

There exist a large variety of algorithms addressing the speech enhancement problem, such as spectral subtraction, Wiener filtering, Ephraim Malah filtering, hidden Markov modeling, or signal subspace methods [1-7]. A non Gaussian model based on Ephraim–Malah filter was evolved. This model is implemented by spectral amplitude estimation based on the generalized Gamma distribution (GCD) of speech and MAP estimator [8]. Md. Kamrul Hasan [9] presented an improved thresholding technique for speech enhancement in the discrete cosine transform (DCT) domain, where the signal-bias-compensated noise level was used as the threshold parameter. Speech classification into voiced and silent frames is essential in many speech processing applications, as well as, segmentation of voiced speech into individual pitch epochs is necessary in several high quality speech synthesis and coding techniques. Veprek P. and Michael Scordilis [10] introduced criteria for measuring the performance of automatic procedures performing this task against manually segmented and labeled data, where five basic pitch determination algorithms (SIFT, comb filter energy maximization, optimal temporal similarity and dyadic wavelet transform) were evaluated. A new pitch determination method based on Hilbert-Huang Transform (HHT) was presented in [11]. Qinghua Huang *et* al. [12] proposed a Variational Bayesian learning approach for speech modeling and enhancement. They used time-varying autoregressive process to model clean speech signal and used Variational Bayesian learning to estimate the model parameters. The majorities of these methods deal with short-time spectral attenuation of the noisy effect and are capable to eliminate background noise powerfully but distorting artifacts remain in the enhanced speech signal. These artifacts are recognized as ‘‘musical noise’’ due to their tonal spectrum.

The idea of the wavelet started with the Gabor Transform [13]. Later on, the subject of multi-scale signal decomposition has been tried by applied mathematicians for a number of years. The papers of mathematicians Mallat [14,15] and Daubechies [16,17] directed the attention of signal processing researchers in the theory of wavelet transforms, as well as its engineering applications. These papers established the theory of multirate filter banks basing on wavelet transforms. The idea of the noise removing by wavelet transform started early in 90's, particularly basing on the singularity information analysis [18] and the thresholding of the wavelet coefficients [19].

Mallat and Hwang [18] proposed an iterative algorithm to remove the noise via proving

that the modulus maxima of the wavelet coefficients give a comprehensive representation of the signal. Donoho and Johnstone [19-22] proposed a well-known universal wavelet threshold to remove White Gaussian Noise.
This paper presents a wavelet filters enhancement method (WFEM) via multistage convolution by Reverse Biorthogonal Wavelets in high and low pass bands speech signal parts of frequencies.

2. Wavelet Transform Thresholds and Reverse Biorthogonal Wavelets:

In literature there are so many algorithms, which utilize different thresholds. Generally, these algorithms can be summarized in the following steps: decomposing the signal by wavelet transform, thresholding remaining signal and finally, reconstructing the clean signal by Inverse Wavelet Transform (IWT).
Soft thresholding function was expressed as follows [20-22]:

$$T_S(\lambda, w_k) = \begin{cases} \mathrm{sgn}(w_k)(|w_k| - \lambda) & \text{if } |w_k| > \lambda \\ 0 & \text{if } |w_k| \le \lambda \end{cases} \tag{1}$$

Where w_k is the wavelet coefficient and λ is the universal threshold for WT proposed by the same authors

$$\lambda = \sigma \sqrt{2 \log (N)} \tag{2}$$

Where $\sigma = MAD/0.6745$ is the noise level, MAD is the absolute of median estimated on first scale, and N is the length of the enhanced (de-noised) signal. For Wavelet Packets Transform the threshold calculated by

$$\lambda = \sigma \sqrt{2 \log (N \log_2 N)} \tag{3}$$

Johnstone and Silverman [23] investigated the correlated noise situation to define a level-dependent threshold

$$\lambda_j = \sigma_j \sqrt{2 \log (N)} \tag{4}$$

Where the noise level is $\sigma_j = MAD_j/0.6745$, and MAD_j is the absolute of median estimated on the scale j.
Birgé, L. and P. Massart [24] proposed a level-dependent threshold based on Birge-Massart strategy, which can be explained by the following sequent concepts:
$[C, L]$ is the wavelet structure of the decomposed signal to be enhanced (de-noised), at level $j = length(L) - 2$. α and M are real numbers greater than 1. T is a vector of length j; $T(i)$ contains the threshold for level i. N_{KEEP} is a vector of length j; $N_{KEEP}(i)$ contains the number of coefficients to be kept at level i.
The strategy definition:
1- For level $j+1$, everything is kept.

2- For level i from 1 to j, the ni largest coefficients are kept with $ni = M \ (j+2-i)^{\alpha}$. Typically α =3 for de-noising. Recommended values for M are from $L(1)$ to $2 * L(1)$.The scale-adapted threshold was suggested in [25]. For a given sub band k, the matching threshold is defined by

$$\lambda_k = \sigma_k \sqrt{2\log(N)} \tag{5}$$

Where the noise level is $\sigma_k = MAD_k / 0.6745$, MAD_k is the absolute of median estimated on the sub band k.

In this paper, Reverse Biorthogonal Wavelets RBW shown in Fig.1 are used. This family is generated from the biorthogonal wavelet father ϕ and mother ψ [26]. RBW are compactly supported biorthogonal spline wavelets for which symmetry and precise reconstruction are probable with FIR filters. Has arbitrary number of vanishing moments and arbitrary regularity. It is well known in the subband filtering region that symmetry and exact reconstruction are incompatible if the same FIR filters are used for reconstruction and decomposition, then two filters should be used.

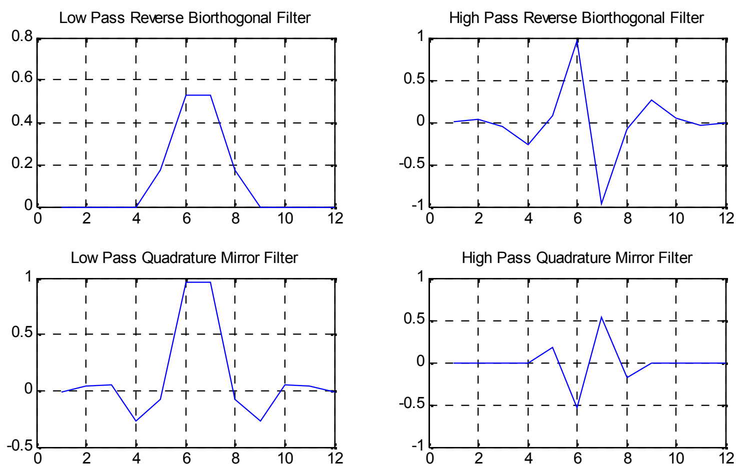

*Figure (1): **Reverse Biorthogonal Wavelets***

3. Proposed Method:

For many signals, the low-frequency content is the most important part. It is what gives the signal its identity. The high-frequency content, on the other hand, imparts flavor or nuance. If you remove the high-frequency components, the voice sounds different, but you can still tell what's being said. However, if you remove enough of the low-frequency components, you hear gibberish. But for high quality enhancement the high and low parts should be filtered carefully. In this work wavelet filter based speech enhancement method is presented. The method contains multistage wavelet filtration based on convolution with Reverse Biorthogonal Wavelets. The proposed method is based on filtration the low frequency and high frequency parts separately, without thresholding (cutting) the values, which leads to losing the essential speech information.

For wavelet filters, we start with the scaling function ϕ If w_n is the coefficient of the linear combination given by:

$$\phi(\frac{x}{2}) = 2^{1/2} \sum_n w_n \phi(x-n) \tag{6}$$

Where w_n is given by [14,15]:

$$w_n = \frac{1}{2^{1/2}} \int \phi\left(\frac{x}{2}\right) \phi(x-n)dx \tag{7}$$

Clearly if ϕ is compactly supported, the sequence w_n is finite and can be viewed as a filter. The filter w_n (scaling filter) is a low-pass Finite Impulse Response (FIR) filter, of length $2N$. A low digital filter's output $y(k)$ is interrelated to its input $s(k)$ by convolution with its impulse response $w(k)$.

$$y(k) = w(k) * s(k) = \sum_{\tau=-\infty}^{\infty} w(k-\tau)s(\tau) \tag{8}$$

In general, the z-transform $Y(z)$ of a digital filter's output $y(n)$ is related to the z-transform $X(z)$ of the input by

$$Y(z) = W(z)S(z) = \frac{b(1) + b(2)z^{-1} + \cdots + b(n+1)z^{-n}}{a(1) + a(2)z^{-1} + \cdots + a(m+1)z^{-m}} X(z) \tag{9}$$

Here $W(z)$ is the filter's transfer function., Where the constants $b(i)$ and $a(i)$ are the filter coefficients and the order of the filter is the maximum of n and m. Therefore $y(k)$ is the low pass output signal. To accomplish better final results two addition low pass filters are applied

$$y_1(k) = w(k) * y(k) \quad \text{and} \quad y_2(k) = w(k) * y_1(k) \tag{10}$$

Now for high pass frequency filtration, high pass wavelet filter must be generated form mother wavelet

$$\psi\left(\frac{x}{2}\right) = 2^{1/2} \sum_n h_n \psi(x-n) \tag{11}$$

$$h_n = \frac{1}{2^{1/2}} \int \psi\left(\frac{x}{2}\right) \psi(x-n)dx \tag{12}$$

A high pass digital filter's output $u(k)$ is related to its input $s(k)$ by convolution with its impulse response $h(k)$.

$$u(k) = h(k) * s(k) = \sum_{\tau=-\infty}^{\infty} h(k-\tau)s(\tau) \tag{13}$$

$u(k)$ is the high pass output signal. To accomplish better final results two addition low pass filters are applied

$$u_1(k) = w(k) * u(k) \quad \text{and} \quad u_2(k) = w(k) * u_1(k) \tag{14}$$

$y_2(k)$ and $u_2(k)$ present clean low frequency part and high frequency part of the speech signal $s(k)$, respectively. The length of these signals is nearly equal to the length of $s(k)$ signal. Therefore, decimation operation must be implemented before giving to reconstruction process. The length of each filter is equal to $2N$. If n is the length of s, the signals $y(k)$ and $u(k)$ are of length $n+2N-1$. So it's very important to eliminate these redundant samples from the beginning and from the end of each convolution results to guarantee optimal s signal reconstruction. In reconstruction as shown in Fig.2, quadrature mirror filters are used

$$h_q(k) = (-1)^k \, w_q(2N+1-k) \quad \text{for } k = 1, 2, ..., 2N \tag{15}$$

afterwards, the clean speech signal is accomplished by zero-padding operation as will as summation of convolution the low and high parts $y_2(k)$ and $u_2(k)$ with quadrature mirror filters $w_q(k)$ and $h_q(k)$, respectively

$$\tilde{s}(k) = \sum_{\tau=-\infty}^{\infty} y_2(\tau) w_q(k-\tau) + \sum_{\tau=-\infty}^{\infty} u_2(\tau) h_q(k-\tau) \tag{16}$$

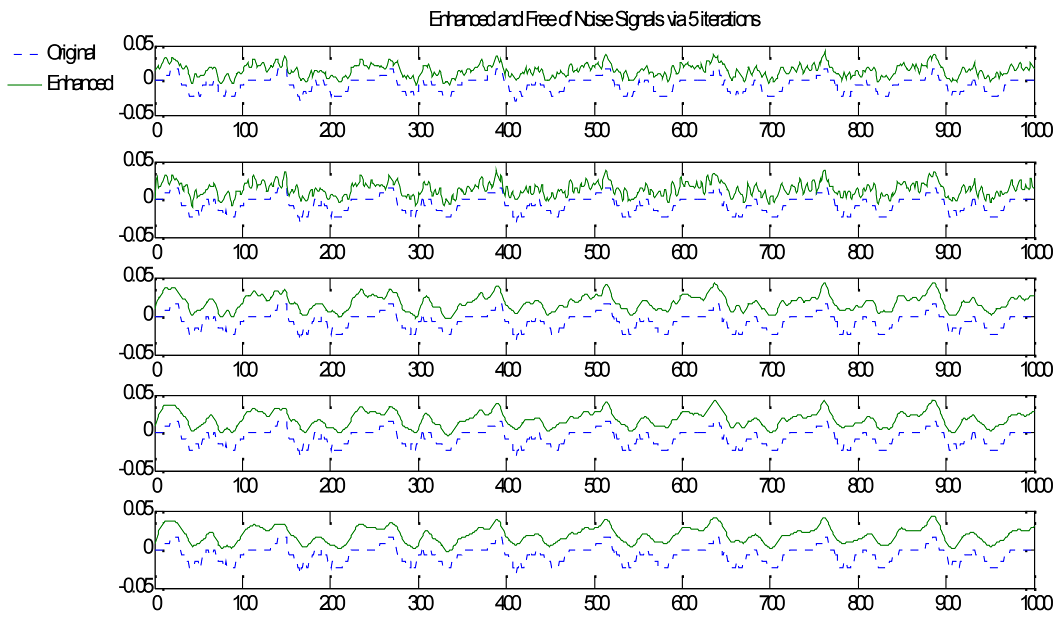

Figure (2)**:** Enhanced signal of SNR -3.7631 dB, by WFEM and clean signal via iteration 1, 3, 5, 7 and 11, with improved SNR: 4.3603 dB, 3.3839 dB, 5.1238 dB, 5.5211 dB and 5.7370 dB, respectively.

The reconstruction error is calculated by the difference of the two victors s and $\tilde{s}$

$$e = s - \tilde{s} \tag{17}$$

After 5 iterations the highest Signal-to-Noise ratio is accomplishing where

$$SNR = 10\log 10 \frac{\sum \tilde{s}^2}{\sum (s-\tilde{s})^2} \tag{18}$$

and the minimum Mean Square Error is given by:

$$MSE = E[(s-\tilde{s})^2] \tag{19}$$

The proposed method presents new approach of speech signal enhancement by using wavelet filters particularly RBW. As we mentioned above the method based on filtration the low frequency and high frequency parts separately, without thresholding (cutting) the values, which leads to losing the essential speech information.

4. Objective Evaluation:

Different methods are used for speech enhancement systems evaluation. All of these methods based on extracting an original signal to enhanced signal ratio measure or distance measure. The most popular measure, which gives a measure of the signal power improvement related to the noise power is SNR and segmental SNR (segSNR). From spectral domain evaluation algorithm we can mention Weighted–Slope Spectral distance (WSS) [27]:

$$d_{WSS} = \frac{1}{M}\sum_{m=0}^{M-1} \frac{\sum_{j=1}^{K} W(J,M)(s_C(J,M) - s_P(J,M))^2}{\sum_{j=1}^{K} W(J,M)} \tag{20}$$

Where $W(j,m)$ is the weight placed on jth frequency band, K is the number of bands and M is the number of frames in the signal. $s_C(j,m)$ and $s_p(j,m)$ spectral are slope of the clean and enhanced signals, respectively. In [28] K is proposed as 25.

Cepstrum distance has been used in [28,29] as a difference of original signal cepstrum and enhanced signal cepstrum

$$d_{CEP}(\vec{C}_C, \vec{C}_P) = \frac{10}{\log 10}\sqrt{2\sum_{k=1}^{p}(C_C(k) - C_P(k))^2} \tag{21}$$

where $\vec{C}_C$ and $\vec{C}_P$ are original signal cepstrum and enhanced signal cepstrum vectors, respectively.

In literature, LPC-based objective measures have been utilized such log-likelihood ratio (LLR) [30]:

$$d_{LLR}(\vec{a}_P, \vec{a}_C) = \log\left(\frac{\vec{a}_P \mathrm{R}_C \vec{a}_P^T}{\vec{a}_C \mathrm{R}_C \vec{a}_C^T}\right) \tag{22}$$

where $\vec{a}_C$ and $\vec{a}_P$ are LPC vectors of the original and enhanced signals, respectively

and R_C is autocorrelation of original signal.
In [28] composite evaluation is proposed, which was obtained as a correlation between objective and subjective evaluation by using two merits correlation coefficient and standard deviation.
In this paper a new evaluation measure is proposed by Continuous Wavelet Transform (CWT). This measure is obtained by calculated the differences between CWT of original signal and enhanced signal over three levels, low, medium and high. And then average of standards deviation is obtained

$$d_{CWT} = \frac{\sum_{j}^{J} \sqrt{E\left[(C_j - \overline{C}_j)^2\right]}}{3} \quad \text{for } j = 5, 10 \text{ and } 15 \tag{23}$$

Where $C_j = CWT_j(s) - CWT_j(\tilde{s})$ and $\overline{C}$ is the mean value. The level determination as 5, 10 and 15 is according to the sampling frequency of the speech signal. These levels present low, medium and high pass bands of the signal frequency. So that, the utilizing this measure helps studying the difference between filtered and clean signals via three bands, instead of whole signal overlapped bands.

5. Results and Discussion

Tested speech signals were recorded via PC-sound card, with sampling frequency of 16000 Hz, over about 2 sec. time duration. Each speaker recorded Arabic expression "besme allah Alrahman Alraheem" that means in English "In the Name of God" that was recorded one time by the speaker. The speaker recorded 26 utterances. 4 females 18 males got a part in utterances recording. The recording process was provided in normal university office conditions.
The experimental part of this research is introduced by utilizing several objective measures such as d_{CWT}, modified Cepstrum distance

$$Md_{CEP}(\vec{C}_C, \vec{C}_P) = \log 10 \sqrt{2\sum_{k=1}^{P}(C_C(k) - C_P(k))^2} \tag{24}$$

and the modified LPC-based log-likelihood ratio Md_{LLR}

$$Md_{LLR} = \left| \mathrm{Re}\left(\log\left(\frac{\sum_{n}^{N} a_s(n) R_s}{\sum_{n}^{N} a_{\tilde{s}}(n) R_{\tilde{s}}} \right) \right) \right| \tag{25}$$

where $a_s(n)$ and $a_{\tilde{s}}(n)$ are the LPC of the original and enhanced signals, respectively. R_s, $R_{\tilde{s}}$ are autocorrelation of original and enhanced signals. The modification is done to make the two measures more suitable for our research. Correlation coefficient and MSE are also used.
The used objective evaluation based on White Gaussian Noise (WGN).

The Gaussian Density function is given by

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{\frac{-(x-\mu_X)^2}{2\sigma_X{}^2}} \tag{26}$$

Where μ is the mean value and σ is the standard deviation of the random variable X. The results are obtained via different SNR levels and shown in Table (1). The proposed method WFEM is compared with two wavelet conventional thresholding methods. The first method is Donoho and Johnstone thresholding method (DTM) in (Eq.2) and the second method is Birge-Massart thresholding strategy method (BMSM) presented in section 2. The soft (S) and hard (H) thresholding are utilized. DTM and BMSM are used in the optimal condition related to the used parameters according to literature presented in section 2 and experimental observations. Table 1 presents the SNR of the enhanced signal by three methods WFEM, DTM (S and H) and BMSM (S and H). The SNR of corrupted signal utilized in the experiment ranges from -5.6 dB to 14.08 dB. Proposed method is tested over fifteen iterations, where maximum SNR is taken for each corrupted SNR level trail. WFEM improves the SNR practically 10dB (-5.628 to 5.065). The results illustrate that WFEM is superior in SNR improvement.

Table (1): *Improved SNR by WFEM, DTM and BMSM with WGN*

Corrupted SNR dB	WFEM SNR dB	DTM (S) SNR dB	BMSM (S) SNR dB	DTM (H) SNR dB	BMEM (H) SNR dB
14.088	19.561	17.257	15.691	17.780	16.552
7.183	13.58	12.589	17.839	12.867	9.426
5.300	12.036	10.987	12.371	9.963	7.100
1.584	9.237	7.793	10.673	8.255	5.309
0.462	8.741	7.117	7.525	6.468	4.077
-1.804	6.873	5.300	6.731	5.448	3.270
-2.945	6.296	4.653	4.950	4.396	2.198
-4.866	5.358	3.043	4.336	4.312	2.393
-5.055	5.492	2.781	2.801	4.100	2.14
-5.628	5.065	3.081	2.672	3.016	1.212

Five different objective measures are used, the correlation coefficient ρ, MSE, Md_{CEP}, d_{CWT} and Md_{LLR}. The results are shown in Table (2). WFEM is tested over 15 iterations, where maximum ρ and minimum of MSE, Md_{CEP}, d_{CWT} and Md_{LLR} are taken for each corrupted SNR level trial. The result of each method is taken as an average of the results of the 10 SNR levels from -5.6 dB to 14.08 dB. The objective measure results indicate the superiority of WFEM in White Gaussian Noise case as shown in Fig. 3.

Table (2): *Objective evaluation with WGN*

Objective Measure	WFEM	DTm (S)	BMsm (S)	DTm (H)	BMsm (H)
ρ	0.915	0.863	0.852	0.853	0.757
MSE	0.0011	0.0030	0.0015	0.0031	0.0023
MdCEP	0.314	0.341	0.349	0.307	0.504
dCWT	0.056	0.069	0.070	0.073	0.067
MLLR	0.721	2.6205	0.646	2.139	0.886

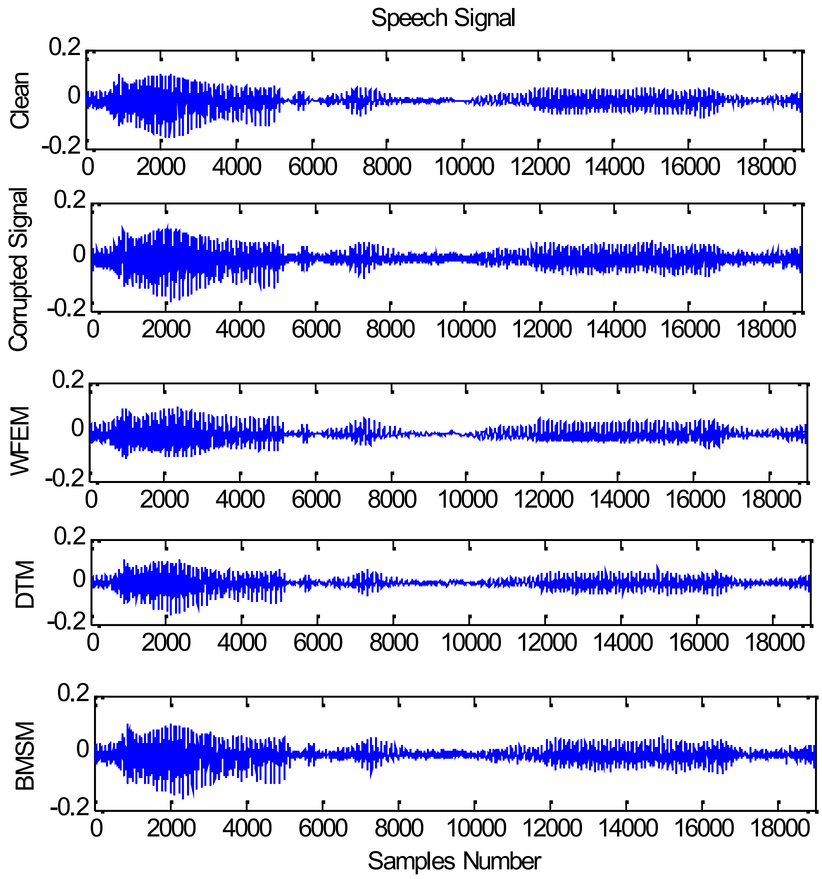

Figure (3): Enhanced speech signal using WFEM, DTM and BMSM with WGN (SNR=7dB)

6. Conclusions:

In this Paper, a new wavelet filters speech enhancement method (WFEM) is proposed. The proposed method in this study depends on two steps: Filtration using Reverse Biorthogonal Wavelets filters, and reconstruction the clean signal by quadrature mirror of Reverse Biorthogonal Wavelets Filters. The method improves the SNR in some cases about 15 dB. The new method is compared with Donoho and Johnstone thresholding method and Birge-Massart thresholding strategy method. The proposed method is tested by objective measures via Gaussian noise probability distribution functions. The new method shows superiority over Donoho and Johnstone thresholding method and Birge-Massart thresholding strategy method.

References:

[1] Berouti, M., Schwartz, R., Makhoul, J., 1979. Enhancement of speech corrupted by acoustic noise. *IEEE ICASSP 1*, pp. 208–211.
[2] Ghanbari, Y., Karami, M., 2004. Spectral subtraction in the wavelet domain for speech enhancement. *Internat. J. Software Inf. Technol. (IJSIT) 1 (1)*, 26–30.
[3] Kamath, S., Loizou, P., 2002. A Multi-band spectral subtraction method for enhancing speech corrupted by colored noise. *Proceedings of International Conference on Acoustics, Speech and Signal Processing*.
[4] Deller, J.R., Hansen, J.H.L., Proakis, J.G., 2000. Discrete-Time Processing of Speech Signals, second ed. IEEE Press, New York.
[5] Ephraim, Y., Malah, D., 1984. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process. ASSP-32 (6), 1109–1121.
[6] Sameti, Hmm-based strategies for enhancement of speech signals embedded in nonstationary noise. IEEE Trans. Acoust. Speech Signal Process. v6. 445-455,1998.
[7] Klein, M., Kabal, P., 2002. Signal subspace speech enhancement with perceptual post-filtering. Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. (ICASSP) 1, 537–540.
[8] Dat, T.H., Takeda, K., Itakura, F., 2005. Generalized gamma modeling of speech and its online estimation for speech enhancement. In: Proc. ICASSP-2005.
[9] Md. Kamrul Hasan, Reducing signal-bias from MAD estimated noise level for DCT speech enhancement, Signal Processing 84 (2004) 151 – 162
[10] P. Veprek, Michael S. Scordilis. Analysis, enhancement and evaluation of five pitch determination techniques, Speech Communication 37 (2002) 249–270.
[11] Hai Huang and Jiaqiang Pan, Uniform and warped low delay filter-banks for speech enhancement, Signal Processing 86 (2006) 792–803.

[12] Qinghua Huang, Jie Yang, and Shoushui Wei,Variational Bayesian learning for speech modeling and enhancement, Signal Processing 87 (2007) 2026–2035.
[13] Gabor, D., 1946. Theory of communications. J. Inst. Electr. Eng. Lond. 93, 429–457.
[14] Mallat S., A theory for multiresolution signal decomposition, IEEE Trans. PAMI 11 (7) (1989) 674–693a.
[15] Mallat S., Multifrequency channel decompositions of images and wavelet models, IEEE Trans. Acoust. Speech Signal Process. 37 (12) (1989) 2091–2110 b.
[16] Cohen A., Daubechies I., Feauveau J.C, Biorthogonal bases of compactly supported wavelets, Commun. Pur. Appl. Math. 45 (1992) 485–560.
[17] Daubechies I., Orthonormal bases of compactly supported wavelets, Commun. Pur. Appl. Math. 41 (11) (1988) 909–996.
[18] Mallat, S., Hwang, W., 1992. Singularity detection and processing with wavelets. IEEE Trans. Inform. Theory 38, 617– 643.
[19] Donoho, D., 1993. Nonlinear wavelet methods for recovering signals, images, and densities from indirect and noisy data. Proc. Symp. Appl. Math. 47, 173–205.
[20] Donoho, D., Johnstone, I., 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425–455.
[21] Donoho, D.L., 1995. Denoising by soft thresholding. IEEE Trans. Inform. Theory 41 (3), 613–627.
[22] Donoho, D.L., Johnstone, I.M., 1995. Adapting to unknown smoothness via wavelet shrinkage. J. Amer. Statist. Assoc. 90, 1200–1224.
[23] Johnstone, I.M., Silverman, B.W., 1997. Wavelet threshold estimators for data with correlated noise. J. Roy. Statist. Soc., Ser. B (Gen.) 59, 319–351.
[24] Birgé, L.; P. Massart (1997), "From model selection to adaptive estimation," in D. Pollard (ed), Festchrift for L. Le Cam, Springer, pp. 55-88.
[25] Mohammed Bahoura, Jean Rouat , Wavelet speech enhancement based on time–scale adaptation, Speech Communication 48 (2006) 1620–1637.
[26] Daubechies, I. (1992), Ten lectures on wavelets, CBMS-NSF conference series in applied mathematics. SIAM Ed.
[27] Klatt D. , "Prediction of perceived phonetic distance from critical band spectra," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1982, vol. 7, pp. 1278–1281.
[28] Hu Y. and Loizou P., Evaluation of Objective Quality Measuresfor Speech Enhancement, IEEE Tran. on Audio, Speech, and Language Processing, Vol. 16, N0. 1, Jan. 2008
[29] Kitawaki N., Nagabuchi H., and Itoh K., "Objective quality evaluation for low bit-rate speech coding systems," *IEEE J. Sel. Areas Commun.*, vol. 6, no. 2, pp. 262–273, Mar. 1988.

[30] Hansen and J. Pellom B, "An effective quality evaluation protocol for speech enhancement algorithms," in *Proc. Int. Conf. Spoken Lang. Process.*, 1998, vol. 7, pp. 2819–282.