Speech enhancement using sub-band wiener filter with pitch synchronous analysis

Conference Paper · August 2013

DOI: 10.1109/ICACCI.2013.6637140

2 authors:

Sunnydayal Vanambathina
University of Crete
19 PUBLICATIONS **74** CITATIONS

SEE PROFILE

T. Kishore Kumar
National Institute of Technology, Warangal
44 PUBLICATIONS **390** CITATIONS

SEE PROFILE

Speech Enhancement using Sub-band Wiener Filter with Pitch Synchronous Analysis

V.Sunnydayal,T.Kishore Kumar
Dept. Of Electronics and Communications Engineering
National Institute of Technolgy Warangal, Warangal,India
Email: sunnydayal45@gmail.com, kishorefr@gmail.com

***Abstract*- This paper introduces a novel speech enhancement system based on sub-band wiener filter with pitch synchronous analysis. The perceptual filter bank provides a good auditory representation, good perceptual quality of speech. Sub-band wiener filter based Pitch synchronous analysis reduces drawbacks of fixed window shift. To increase the inter frame similarities the shift of analysis window is based on pitch period. The pitch is extracted by using clipping level method. Further, to reduce noise each sub-band is wiener filtered using a priori SNR with adaptive parameter. Wavelets transform based wiener filter approach works well than DCT based approach. Objective (SNR, segSNR, LLR, LSD, IS,WSS) experiments prove that the new speech enhancement system is capable of significant noise reduction.**

***Keywords*—Speech enhancement; Speech Processing; Wavelet packet transform.**

I. INTRODUCTION

Speech enhancement is an important technology in speech signal process fields. Speech enhancement algorithms are used to remove background noises [1][2].Speech enhancement methods can be divided in to categories based on domain of operations namely time domain, frequency domain and time-frequency domain. Time domain method includes subspace approach[3], frequency domain approach includes spectral subtraction, MMSE[4], wiener filtering[5] and time-frequency domain methods involve wavelets [6][7].

A new sub-band approach to the Wiener filter method is developed that reduces background noise. In this approach each sub-band is wiener filtered using a priori SNR with adaptive parameter.

Wavelet Packet method is widely used in speech enhancement because of good enhancement effect. This method analyzes in time domain and frequency domain, with time frequency localization compared to Fourier analysis. The selection of wavelet function is also very flexible. Wavelet analysis represents a variable-sized windowing. At lower frequencies parameters are calculated over a greater length of time and have higher frequency resolution. On the other hand, at higher frequencies parameters are calculated with high time resolution but poor frequency resolution. Wavelet packet transform (WPT) is decomposes signal in to sub-bands. Each sub band is robust to local variations. Each band reflect the over all spectral shape .

Pitch synchronous analysis is used in prosody modification system [8] , speech Analysis/synthesis system [9] and speech recognition system [10].The resulted pitch synchronous signal can be applied to different purposes such as to control the value of the synthesized pitch. The main aim of pitch synchronous is to reduce the discontinuities associated with windowing. Standard window –shift without phase representation shows higher variation in the transform coefficient magnitudes. This impacts negatively on the inter-frame techniques such as the decision-directed approach [4] for the estimation of *a priori* SNR.

The remainder of this paper is organized as follows. In Section II, a detailed description of the sub-band wiener filter with pitch synchronous analysis is given. Section III give a brief description wavelet packet transform (wpt). In Section IV wiener filter with adaptive controller is discussed.. Section V Pitch synchronization, In Section VI demonstrates the experimental results and Section VII concludes the paper.

II. SUB-BAND WIENER FILTER WITH PITCH SYNCHRONOUS ANALYSIS

Noisy speech is filtered by noise reduction algorithm and voiced/unvoiced decision is made. If it contains voiced signal, the time-shift will be changed to one pitch period. Otherwise, the time-shift is of original fixed value. In this paper analysis window shift is not fixed. The proposed method is shown in Fig.1. The analysis window is a rectangular window because this window is not going to suppress more dominant values at the starting point in contrast to other smooth window functions.

978-1-4673-6217-7/13/$31.00 ©2013 IEEE

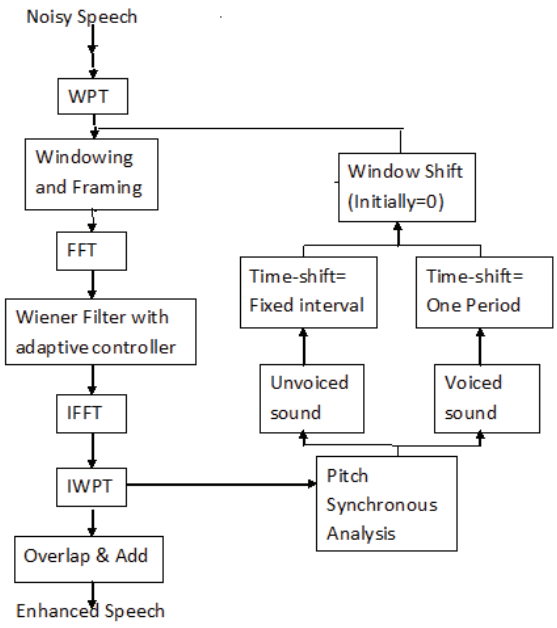

Fig. 1. Block diagram of proposed method

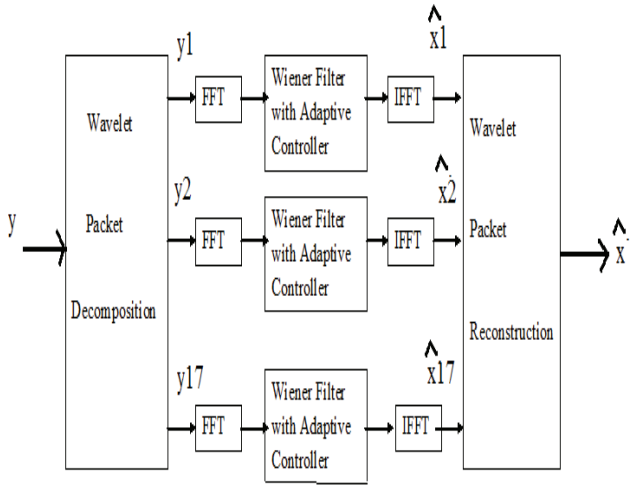

Fig. 2. Sub-band wiener filtering

We apply the wavelet packet transform (WPT) to the input to produce 2^J sub band wavelet packets, where J is the number of levels for WPT. Figure 2 shows the wavelet packet decomposition of noisy speech signal. Then a posterior and a priori segmental signal-to-noise ratio (SSNR) is computed for each sub band. Wiener filtering is used for enhancement of frequency bands of bark scaled filter bank. Finally, the inverse wavelet transform synthesizes the enhanced speech.Enhanced speech is obtained by overlap and add method.

III. WAVELET PACKET TRANSFORM (WPT)

Wavelet transform is used as a powerful tool in noise reduction. Discrete orthogonal wavelet packets are used as a time-domain transparent analysis–synthesis system with the capability of perceptually tuned multi resolution time-spectral decomposition. The signal is decomposed in to approximation (A) and detail (D) parts and are obtained by using a high-pass filter and a low-pass filter. In this paper implemented with the Daubechies family wavelet. Using a five-level DWPT , a frequency resolution of 125 Hz can be achieved as shown in Fig.3. This corresponds to a maximal frame duration of 4 ms. As the number of stages WPT increases Daubechies filters gives us best frequency selectivity. The analysis window length at higher frequencies is limited to 5–10 ms , whereas they can spread up to 100 ms at lower frequencies. The duration of the analysis windows or basis functions is given by

$$W=(L-1)(F_j-1)+1 \quad (1)$$

The frame length at stage j is given by

$$F_{j=2}{}^{(j+1)} \quad (2)$$

At higher frequencies (j=2), the window length is W_{min} =64 samples, At lower frequencies (j=5) the window length is W_{max}=658 samples.

Window duration controls the parameters calculation. The control of rate at which the parameters track the dynamics of the signal depends on frame duration and window duration.

Critical bands of human auditory system are shown in the Table I [11]. The input speech signal is decomposed into sub band signals using a wavelet packet decomposition tree. The input signal is band limited to 4 kHz and sampled at 8 kHz. A Wavelet packet filter-bank is designed and implemented such that 18 sub bands obtained closely match with the critical bands of the human auditory system given in table I.

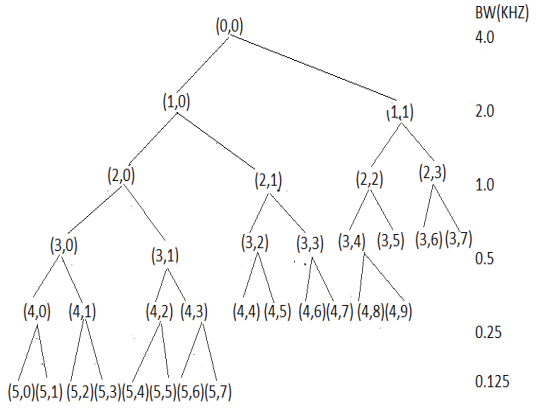

Fig. 3.Wavelet packet tree and its sub bands of 4KHz input signal

Table I CRITICAL BANDS OF HUMAN AUDITORY SYSTEM.

Band NO	Centre Frequency (Hz)	Bandwidth (Hz)
1	50	0-100
2	150	100-200
3	250	200-300
4	350	300-400
5	450	400-510
6	570	510-630
7	700	630-770
8	840	770-920
9	1000	920-1080
10	1175	1080-1270
11	1370	1270-1480
12	1600	1480-1720
13	1850	1720-2000
14	2150	2000-2320
15	2500	2320-2700
16	2900	2700-3150
17	3400	3150-3700
18	4000	3700-4400

IV. *WIENER FILTER WITH ADAPTIVE CONTROLLER*

The novelty of this paper is implementation of wavelet packet transform, which divides the input signal into subbands, each subband is wiener filtered and apriori SNR is estimated from each subband where as in case of DCT apriori SNR is estimated frame to frame from input noisy speech without considering subbands **.**Wiener filter is the optimal filter in finding mean square error [12]**.**The estimation of MSE depends on a priori SNR. The a priori SNR can be computed by decision- directed approach. [4]**.** Let the noisy speech, clean speech and noise signal be denoted as y,x and d respectively.

$$y(n) = x(n) + d(n) \tag{3}$$

Noisy signal y(n) is framed and , wavelet packet transform is applied to each input frame and the corresponding coefficients are $Y_k^i(m)$ where is m the time frame index , k is the frequency index and i is the sub band number.

Let $\zeta_k^i(m)$, $A_k^i(m)$, $\lambda_k^i(m)$ and $\gamma_k^i(m)$ denote a priori SNR, the Amplitude, noise variance and posteriori SNR respectively of corresponding k^{th} spectral component in m^{th} analysis frame of i^{th} subband..The estimated a priori $\hat{\xi}_k^i(m)$ can be expressed as

$$\hat{\xi}_k^i(m) = a\frac{\hat{X}^{i}{}_k^2(m-1)}{\lambda_d^i(k,m-1)} + (1-a)\max[\frac{Y_k^2(m)}{\lambda_d(k,m)} - 1,0] \tag{4}$$

Where $\hat{X}^{i}{}_k^2(m-1)$ is the estimated clean speech in the previous frame, and $\lambda_d^i(k,m)$ is the noise variance and is equals to the expectation of the power magnitude of the noise signal $E[D^{i}{}_k^2(m)]$. Adapting **a** for better estimation of apriori SNR in WPT domain leads to improved version of DCT[1] and wiener filter. From Decision direct approach (4) apriori SNR in terms of subbands can be expressed as

$$\hat{\xi}_k^i(m) = a_k^i(m)\hat{\xi}_k^i(m-1) + (1-a_k^i(m))\max[\,\gamma_k^i(\mathrm{m}) - 1,0] \tag{5}$$

Where $a_k^i(m)$ is an adaptive version of **a** and

$$\hat{\xi}_k^i(m-1) = a\frac{\left|\hat{X}^{i}{}_k^2(m-1)\right|}{\lambda_d^i(k,m-1)}$$

The parameter **a** is used to set a proportion of contributions from the previous frames to the current estimate.

If any phase discontinuity occurs , it destroys the phase information. Adaptive **a** for direct decision approach using DCT for estimation of the *a priori* SNR is discussed in [1]. The proposed method also follows same approach in estimation of a priori SNR for each critical band individually. The proposed method is modified with WPT to get improved version of wiener filter.

The derivation for optimal $a_k^i(m)$ is given in [1].

The value of optimal $a_k^i(m)$ is given by

$$a_k^i(m) = \frac{2a_k{}^{i^2}(m) + 4a_k^i(m)}{\left(\hat{a}_k^i(m-1) - a_k^i(m)\right)^2 + 2a^{i^2}{}_k(m) + 4a_k^i(m)}$$

$$\cong \frac{1}{1 + \frac{1}{2}\left[\frac{\left(\hat{a}_k^i(m-1) - a_k^i(m)\right)}{1 + a_k^i(m)}\right]^2} \tag{6}$$

If the SNR changes slowly, the value $a_k^i(m)$ is close to one, If SNR changes sharply the value of $a_k^i(m)$ is very small. The proposed algorithm is better estimates of apriori SNR when compared with existing DCT [1] approach and causes improvement in overall SNR.

V. PITCH SYNCHRONIZATION

The output of Wiener filtered speech is given by

$$\hat{X}_k^i(m) = \frac{\hat{\xi}_k^i(m)}{\hat{\xi}_k^i(m)+1} Y_k^i(m) \qquad (7)$$

where $\hat{X}_k^i(m)$ represents the enhanced speech frame of *i* th sub-band. The final enhanced speech signal is reconstructed by using the standard overlap- and-add method. This enhanced speech is utilized for pitch detection to obtain a more accurate estimation to implement the adaptive time shift analysis (ATSA) algorithm.

A. Autocorrelation

The time domain autocorrelation method is common to detect the pitch period. Pitch period is detected by using clipping level. For better pitch period estimation clipping is applied to each speech segment.The speech signal is processed by a three-level centre clipper and the correlation function is computed over a range of pitch periods. Autocorrelation function is calculated by

$$\Phi(\tau) = \sum_{n=0}^{N-1} \hat{x}(n)\hat{x}(n+\tau) \qquad (8)$$

Where $\hat{x}(n)$ is speech signal, τ is lag number, n is time for discrete signal. This method is robustness against noise may be useful. A distinct peak is defined as 0.5 times of $\Phi(0)$. A distinct peak in the range of 50 Hz to 500 Hz implies the presence of voiced. If no distinct peak is found, then it is a silence or unvoiced frame. The pitch period is extracted from voiced frames, and is used as the analysis window shift. Window length should be atleast twice as long as pitch period. The final enhanced speech is obtained by overlap add process. Because of adaptive window shifting , this process is different from the original process. The weighting function records all the windows frame by frame and calculates the net weighting function. The weighting function can be calculated from the current and the previous frames and hence can be performed in real time. Thereafter, the enhanced speech has to be normalized by the weighting function.

VI. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

The proposed algorithm has been tested on the spoken English sentence chosen from the TIMIT database. Noisy speech signals are obtained by corrupting the clean speech with white, pink and F16 noises. Five SNR levels, including −5, 0, 5, 10, 15 dB, are used to evaluate the performance of a speech enhancement system. The sentence is about 10s with the sampling rate of 8 kHz and spoken by a male (M) and female (F) speakers. The used wavelet was ‘‘db4’’.The overlap between speech analysis frames is 75%, it gives better resulting speech compared with 50%. To test the performance of proposed speech enhancement system**,** the objective quality measurement tests, signal to-noise ratio (SNR), segmental signal-to-noise ratio (segSNR) , Log likelihood ratio (LLR), Log spectral distance (LSD), Itakura Saito Distortion Measure (IS) and Wide Spectral Slope measure (WSS) were used[13][14].

A.Signal-To-Noise Ratio

$$SNR(dB) = 10\log 10 \frac{\sum_{n=0}^{N-1} x^2(n)}{\sum_{n=0}^{N-1}\left[\hat{x}(n)-x(n)\right]^2}$$

where *x(n)* denotes original signal, $\hat{x}(n)$ denotes enhanced signal and *N* denotes number of samples in original speech signal.

B. Segmental Signal-To-Noise Ratio

The frame based segmental SNR is a reasonable measure of speech quality. It is formed by averaging frame level estimates as follows

$$segSNR = \frac{10}{M}\sum_{m=0}^{M-1} \log 10 \frac{\sum_{n=Lm}^{Lm+L-1} x^2(n)}{\sum_{n=Lm}^{Lm+L-1}\left[\hat{x}(n)-x(n)\right]^2}$$

where *L* is the frame length , and *M* the number of frames

C. Log Likelihood Ratio (LLR**)**

LLR measure is defined as

$$d_{LLR}\left(\vec{a}_p, \vec{a}_c\right) = \log\left(\frac{\vec{a}_p R_c \vec{a}_p^T}{\vec{a}_c R_c \vec{a}_c^T}\right)$$

Where $\vec{a}_c$ is the LPC vector of original signal frame , $\vec{a}_p$ is the LPC vector of enhanced speech frame and R_c is the autocorrelation matrix of original speech signal.

D. Log Spectral Distance (LSD)

Log spectral distance is defined as

$$LSD = \frac{1}{L}\sum_{i=1}^{L}\sqrt{\frac{1}{\frac{k}{2}+1}\sum_{k=0}^{K/2}\left(10\log|\hat{s}(k)| - 10\log|s(k)|\right)^2}$$

Where s(k) is the clean speech frame and $\hat{s}(k)$ is the enhanced speech frame of K- point DFT calculation, and L is the number of frames.LSD measure is of frequency domain. Smaller the LSD, closer the shape of log spectrum of clean speech.

E. Itakura Saito Distortion Measure (IS)

Itakura Saito Distortion Measure ratio is given by

$$d_{IS}\left(\vec{a}_p,\vec{a}_c\right) = \frac{\sigma_c^2}{\sigma_p^2}\left(\frac{\vec{a}_p R_c \vec{a}_p^T}{\vec{a}_c R_c \vec{a}_c^T}\right) + \log\left(\frac{\sigma_c^2}{\sigma_p^2}\right) - 1$$

Where $\vec{a}_c$ is the original clean speech frame with linear prediction vector , $\vec{a}_p$ is the processed speech frame with linear prediction vector , σ_c is LPC gain of clean speech frame and σ_p is the LPC gain of processed speech frame.

F. Wide Spectral Slope measure (WSS)

Wide Spectral Slope measure is defined by

$$d_{WSS} = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{K} W(j,m)\left(S_c(j,m) - S_p(j,m)\right)^2}{\sum_{j=1}^{K} W(j,m)}$$

Where W(j,m) are the weights j is the jth frequency m is the frame number.Sc(j,m) is the spectral slope for jth frequency at frame m of clean speech signal and Sp(j,m) is the spectral slope for jth frequency at frame m of processed speech signal. K is the number of bands. Simulation results are shown in fig.4 and fig 5.
Table II shows comparison results of SNR and seg SNR
Table III shows comparison results of LLR and LSD.
Table IV shows comparison results of IS and WSS.

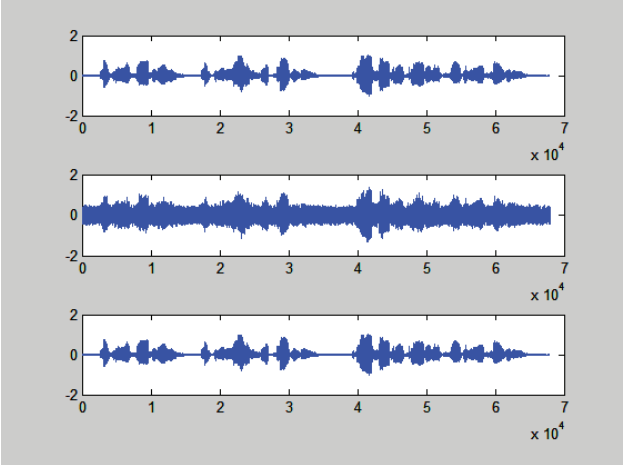

Fig. 4. (a) Clean speech (Female) (b) Noisy(White Noise, 5 dB) speech (c) Enhanced speech(Proposed method)

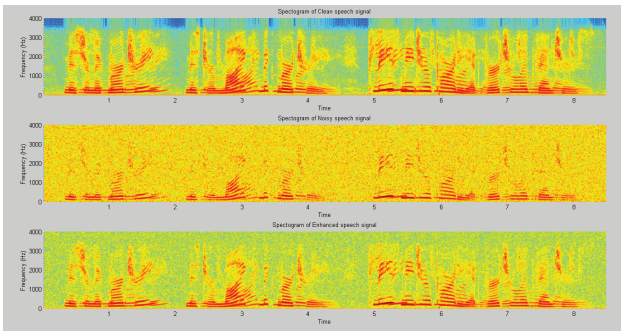

Fig. 5. Spectrogram of (a) Clean speech (Female) (b) Noisy speech(White Noise, 5dB) (c) Enhanced speech(Proposed method)

TABLE II COMPARISON OF SNR AND SEGSNR IMPROVEMENT FOR THE ENHANCED SPEECH IN VARIOUS NOISES

		SNR				**Seg SNR**			
Noise Type	**SNR (dB)**	**DCT [1] (M))**	**Proposed (M)**	**DCT [1] (F)**	**Proposed(F)**	**DCT [1] (M)**	**Proposed (M)**	**DCT [1] (F)**	**Proposed (F)**
White	-5	3.62	**3.97**	3.01	**3.94**	1.21	**1.91**	1.00	**1.31**
	0	5.87	**6.02**	4.98	**5.71**	3.48	**4.36**	2.53	**3.55**
	5	8.53	**9.21**	7.51	**8.25**	5.39	**6.19**	3.86	**4.17**
	10	11.7	**12.5**	9.53	**10.1**	7.12	**8.15**	6.52	**7.31**
	15	14.0	**15.9**	12.3	**13.6**	10.6	**11.7**	8.31	**9.19**
Pink	-5	2.92	**3.56**	1.42	**2.35**	1.21	**2.56**	0.97	**1.32**
	0	5.29	**6.25**	4.55	**5.37**	3.57	**4.21**	2.85	**3.41**
	5	8.15	**9.10**	6.93	**7.72**	5.42	**6.72**	4.21	**5.01**
	10	11.6	**12.1**	10.1	**12.3**	7.12	**8.13**	6.32	**7.46**
	15	13.9	**14.7**	12.3	**13.7**	9.60	**10.1**	8.73	**9.13**
F16	-5	2.53	**3.21**	2.11	**2.96**	1.85	**2.53**	1.58	**2.21**
	0	4.72	**5.52**	3.92	**5.12**	3.69	**4.31**	3.02	**3.91**
	5	6.71	**7.42**	5.62	**6.48**	6.02	**7.65**	4.84	**5.62**
	10	7.94	**8.41**	6.93	**7.94**	7.11	**8.81**	6.97	**7.84**
	15	9.24	**10.1**	8.11	**9.52**	8.93	**9.71**	8.15	**9.41**

TABLE III COMPARISON OF LLR AND LSD IMPROVEMENT FOR THE ENHANCED SPEECH IN VARIOUS NOISES

		LLR				LSD			
Noise Type	**SNR (dB)**	**DCT [1] (M)**	**Proposed (M)**	**DCT [1] (F)**	**Proposed (F)**	**DCT [1] (M)**	**Proposed (M)**	**DCT [1] (F)**	**Proposed (F)**
White	-5	1.92	**1.45**	2.21	**2.11**	2.12	**1.73**	2.41	**2.17**
	0	1.79	**1.12**	1.94	**1.75**	2.01	**1.55**	2.15	**1.82**
	5	1.31	**1.05**	1.62	**1.43**	1.92	**1.42**	1.98	**1.63**
	10	0.96	**0.81**	1.27	**1.02**	1.68	**1.51**	1.83	**1.59**
	15	0.74	**0.64**	1.03	**0.84**	1.24	**1.12**	1.31	**1.04**
Pink	-5	1.57	**1.29**	2.31	**1.96**	2.96	**2.64**	2.95	**2.71**
	0	1.23	**0.11**	1.83	**1.54**	2.53	**2.41**	2.74	**2.49**
	5	0.92	**0.83**	1.52	**1.35**	2.10	**1.83**	2.40	**2.19**
	10	0.71	**0.63**	1.26	**0.95**	1.85	**1.69**	2.15	**1.96**
	15	0.49	**0.35**	0.95	**0.74**	1.53	**1.37**	1.79	**1.51**
F16	-5	1.77	**1.59**	1.77	**1.72**	2.71	**2.53**	2.92	**2.71**
	0	1.60	**1.45**	1.81	**1.53**	2.42	**2.20**	2.71	**2.43**
	5	1.31	**1.20**	1.58	**1.32**	2.15	**2.01**	2.42	**2.15**
	10	1.01	**0.91**	1.31	**1.15**	1.84	**1.63**	2.19	**2.05**
	15	0.82	**0.73**	1.01	**0.81**	1.51	**1.31**	1.91	**1.72**

TABLE IV COMPARISON OF IS AND WSS IMPROVEMENT FOR THE ENHANCED SPEECH IN VARIOUS NOISES

		IS				WSS			
Noise Type	**SNR (dB)**	**DCT [1] (M)**	**Proposed (M)**	**DCT [1] (F)**	**Proposed (F)**	**DCT [1] (M)**	**Proposed (M)**	**DCT [1] (F)**	**Proposed (F)**
White	-5	2.12	**1.72**	2.43	**2.17**	86.3	**82.1**	89.6	**81.8**
	0	1.63	**1.37**	2.10	**1.86**	61.9	**55.8**	68.5	**62.1**
	5	1.21	**0.81**	1.85	**1.53**	45.5	**37.5**	53.8	**49.7**
	10	0.97	**0.63**	1.32	**1.19**	31.2	**27.5**	40.6	**36.6**
	15	0.68	**0.47**	0.96	**0.89**	21.7	**19.2**	30.5	**27.5**
Pink	-5	1.94	**1.52**	2.53	**2.36**	77.1	**71.3**	79.5	**74.2**
	0	1.12	**0.92**	1.96	**1.68**	54.7	**46.9**	57.3	**52.8**
	5	0.86	**0.58**	1.36	**1.20**	35.6	**28.7**	40.2	**37.1**
	10	0.61	**0.52**	0.94	**0.83**	29.4	**27.4**	34.8	**29.7**
	15	0.32	**0.21**	0.67	**0.53**	21.6	**18.7**	27.6	**23.8**
F16	-5	2.53	**2.28**	2.83	**2.52**	80.4	**76.8**	83.2	**78.5**
	0	2.17	**1.83**	2.35	**2.09**	64.2	**60.5**	69.4	**64.4**
	5	1.63	**1.32**	1.95	**1.64**	43.1	**39.8**	49.8	**43.4**
	10	1.17	**0.97**	1.41	**1.09**	32.7	**27.4**	38.1	**32.9**
	15	0.85	**0.64**	1.14	**0.85**	24.1	**19.7**	27.4	**23.1**

In summary, the paper proposes a speech enhancement algorithm based on wavelet packets decomposition and Wiener filter. The proposed method proved to achieve better results than DCT based method [1]. WPT approach gives good quality and noise reduction. Proposed method yields improvement in SNR, seg SNR, LLR, LSD, WSS, IS compared with existing [1] method. Tables II, III, IV have provided comparative analysis of speech signal showing improvement.

VII. CONCLUSION

In this paper , a new speech enhancement system is proposed, sub band wiener filter with pitch synchronous analysis. Overlapping of frames while segmenting speech signal causes change in coefficients from frame to frame because of non-ideal analysis window positions. To reduce this pitch synchronous analysis is used. The analysis window shift is based on pitch period . Each sub band is wiener filtered and a priori SNR is estimated for each sub band. Adaptive parameter **a** is used in estimating a priori SNR perceptual wavelet transform provides an efficient auditory representation. Objective measures SNR, segmental SNR,WSS,IS,LSD,LLR are utilized to evaluate the proposed system. The analysis window shift, WPT, adaptive parameter **a** are integrated in to a speech enhancement system which produces good quality enhanced speech and gives the better noise reduction. The results shows that significant improvements can be obtained by the proposed techniques compared with DCT approach .

REFERENCES

[1] Huijun Ding, Ing Yann Soon, and Chai Kiat Yeo," A DCT-Based Speech Enhancement System With Pitch Synchronous Analysis" IEEE Transactions on audio, speech, and language processing, vol. 19, no. 8, november 2011

[2] P. C. Loizou, *Speech Enhancement: Theory and Practice*. BocaRaton, FL: CRC Press, 2007.

[3] D. O'Shaughnessy, Speech Enhancement: Theory and Practice, IEEE Press, NY,2000.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[5] I. Almajai, B. Milner, "Visually derived wiener filters for speech Enhancement",IEEE Trans. Audio, Speech, Lang. Process. 19, 2011

[6] Y. Hu, P. Loizou, Speech enhancement based on wavelet thresholding the multitaper spectrum, IEEE Trans. Speech Audio Process. 12 (2004) 59–67.

[7] Y. Ghanbari, M.R.K. Mollaei, A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets, Speech Commun. 48 (2006) 927–940.

[8] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Commun., vol. 9, no. 5-6, pp. 453–467, 1990.

[9] Y. Medan and E. Yair, "Pitch synchronous spectral analysis scheme for voiced speech," IEEE Trans. Acoust., Speech, Signal Process., vol. 37, no. 9, pp. 1321–1328, Sep. 1989

[10] M. Hunt and C. Lefebvre, "Speech recognition using an auditory model with pitch-synchronous analysis," in *Proc. ICASSP*, 1987, vol. 12, pp. 813–816.

[11] E.ambikairajah, A.G. Davis, W.T..K. wong, "Auditory masking and MPEG-1 audio compression" Electronics and communication engineering journal, August 1997.

[12] I. Y. Soon, S. N.Koh, and C. K.Yeo, "Noisy speech enhancement using discrete cosine transform," Speech Commun., vol. 24, pp1998

[13] Yi Hu and Philipos C. Loizou, Senior Member, IEEE, "Evaluation of Objective Quality Measures for Speech Enhancement*"* IEEE transactions on audio, speech, and language processing, vol. 16, no. 1, january 2008.

[14] J.Hansen and B. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," in Proc. Int. Conf. Spoken Lang.Process., 1998, vol. 7, pp. 2819–2822.