International Journal of Engineering Research and Development
eISSN : 2278-067X, pISSN : 2278-800X, www.ijerd.com
Volume 2, Issue 2 (July 2012), PP. 57-64

Different Techniques for the Enhancement of the Intelligibility of a Speech Signal

Pankaj Bactor[1], Anil Garg [2]
[1]*Department of ECE, M.M.E.C-M.M. University, Mullana-Ambala-Haryana-India*
[2]*Department of ECE, M.M.E.C-M.M. University, Mullana-Ambala-Haryana-India*

Abstract—Speech enhancement is a popular method for making ASR systems more robust. Spectral subtraction is performed by subtracting the average magnitude of the noise spectrum from the spectrum of the noisy speech to estimate the magnitude of the enhanced speech spectrum. In this paper, adaptive techniques have been explored for the enhancement of the speech such as ANC, ANFIS, DFNN, MDFNN and EDFNN. By using these techniques, it has been observed that the root mean square error (RMSE) and number of epochs are less and at the same time the membership functions are also less.

Keywords—ANC, ANFIS, DFNN, EDFNN, RMSE, SNR

I. INTRODUCTION

In this paper, a comparative performance analysis of single channel, dual-channel and multi-channel (using microphone arrays) speech enhancement techniques, with different types of noise at different SNR'S have been studied. Single channel spectral subtraction was originally designed to improve human speech intelligibility and many attempts have been made to maximized signal-to-noise Ratio (SNR) or minimized speech distortion. Speech enhancement refers to the improvement in the quality or intelligibility of a speech signal and the reversal of degradations that have corrupted it. Quality is a subjective measure which reflects on the pleasantness of the speech or on the amount of effort needed to understand the speech material. Intelligibility is an objective measure which signifies the amount of speech material correctly understood. The main objective of Speech Enhancement is to enhance the speech signal to obtain a clean signal with higher quality. Such system has been widely used in long distance telephony applications.

Artificial intelligence is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence. Fuzzy logic deals with reasoning that is approximate rather than fixed and exact. Fuzzy logic is a form of many-valued logic; it deals with reasoning that is approximate rather than fixed and exact. In contrast with traditional logic theory, where binary sets have two-valued logic: true or false, fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Artificial Neural Network includes training, learning and generalization. Training is the process by which these connection weights (W) are assigned. These weights determine the output of the neural network. The weights are adjusted based on how valid the neural network performed. This process is repeated until the validation error is within an acceptable limit. Training a network involves presenting input patterns in a way so that the system minimizes its error and improves its performance. The weights are adjusted based on how valid the neural network performed. This process is repeated until the validation error is within an acceptable limit. Learning is necessary when the information about inputs/ outputs is unknown or incomplete a priori. Learning process is based on comparison, between networks computed output and the correct expected output, generating 'error'. The 'error' generated is used to change network parameter that result improved performance. The NN possesses the capability to generalize. They can predict new outcomes from past trends. The NN is said to generalize well when it sensibly interpolates input patterns that are new to network. When input-output mapping computed by network is correct.

Some of them are as follows: SCSET (Single-Channel Speech Enhancement Techniques) – Spectral Subtraction Method, Spectral Subtraction with over subtraction model and Non-linear spectral subtraction, MCSET (Multi-Channel Speech Enhancement Techniques) – Adaptive Noise cancellation and Multisensory Beam forming, SSUAA (Spectral Subtraction Using Adaptive Averaging), NREWF (Noise Reduction using Enhanced Wiener Filtering), CSSFG (Cepstral Smoothing of Spectral Filter Gains), SSSTMD (Spectral Subtraction in the Short Time Modulation Domain), Adaptive Noise Cancellation In The Wavelet Domain (ANCWD), Speech Enhancement Using Adaptive Neuro-Fuzzy Filtering (SEANFF) and Adaptive Noise Cancellation Using Cascaded Correlation Neural Networks (ANCCCNN).

II. SINGLE CHANNEL SPEECH ENHANCEMENT TECHNIQUES

Single-channel speech enhancement techniques apply to situations in which .a unique acquisition channel is available. This may be imposed by the system used (as telephone-based applications) or by the availability of the desired signal (as prerecorded applications). When the noise process is stationary and speech activity can be detected, spectral subtraction (SS) is a direct way to enhance the noisy speech.

A. Spectral Subtraction Method: (SS)

The spectral subtraction method can lead to negative values, resulting from differences among the noise estimator and the actual noise frame. To cope with this problem, negative values must be set to zero, Producing Spectral spikes, well-known as “musical noise”. This effect causes an annoying perception of enhanced speech and, therefore, it must be corrected.

B. Spectral Subtraction with Oversubtraction Model: (SSOM)

SSOM procedure was introduced in order to compensate for the “musical noise” effect. It reduces the perception of musical noise.

C. Non-Linear Spectral Subtraction: (NSS)

NSS approach is based on combining two different ideas: i) The use of an extended noise and an over subtraction model ii) Non-linear implementation of the subtraction process, taking into account that the subtraction process must depend on the SNR of the frame, in order to apply less subtraction with high SNRs and vice versa.

III. MULTI CHANNEL SPEECH ENHANCEMENT TECHNIQUES

Multi-channel speech enhancement techniques take advantage of the availability of multiple signal input to our system, making possible the use of noise references in an adaptive noise cancellation device, the use of phase alignment to reject undesired noise components, or even the use of phase alignment and noise cancellation stages into a combined scheme.

A. Adaptive Noise Cancellation: (ANC)

Adaptive noise cancellation is a powerful speech enhancement technique based in the availability of an auxiliary channel, known as reference path, where a correlated sample or reference of the contaminating noise is present. This reference input will be filtered following an adaptive algorithm, in order to subtract the output of this filtering process from the main path, where noisy speech is present.

B. Multisensor Beamforming

Multisensor beamforming through microphone arrays, derived from radar and sonar applications, can be implemented in a variety of ways, being delay-and-sum beamforming the most direct approach. The underlaying idea of this scheme is based on the assumption that the contribution of the reflex ion is small, and that we know the direction of arrival of the desired signal. Then, through a correct alignment of the phase function in each sensor, the desired signal can be enhanced, rejecting all the noisy components not aligned in phase.

IV. SPECTRAL SUBTRACTION USING ADAPTIVE AVERAGING (SSUAA)

This method provides a noise reduction procedure which functions well with arbitrary frame lengths, gives low residual noise, high-quality speech, and low background noise artifacts, and introduces only a short delay. These are important properties when the noise reduction methods are integrated together with other speech enhancement methods and speech coders in real-time communication systems. The method reduces the variability of the gain function—in this case, a complex function—in two ways. First, the variance of the current block’s spectrum estimate is reduced using the Bartlett method by trading frequency resolution for variance reduction. Second, an adaptive averaging of the gain function is used which is dependent on the discrepancy between the estimated noise spectrum and the current input signal spectrum estimate.

V. NOISE REDUCTION USING ENHANCED WIENER FILTERING (NREWF)

The problem of noise reduction has attracted a considerable amount of research attention over the past several decades. Among the numerous techniques that were developed, the optimal Wiener filter can be considered as one of the most fundamental noise reduction approaches, which has been delineated in different forms and adopted in various applications. Although it is not a secret that the Wiener filter may cause some detrimental effects to the speech signal (appreciable or even significant degradation in quality or intelligibility), few efforts have been reported to show the inherent relationship between noise reduction and speech distortion. When no a priori knowledge is available, we can still achieve a better control of noise reduction and speech distortion by properly manipulating the Wiener filter, resulting in a suboptimal Wiener filter. In case that we have multiple microphone sensors, the multiple observations of the speech signal can be used to reduce noise with less or even no speech distortion.

VI. CEPSTRAL SMOOTHING OF SPECTRAL FILTER GAINS (CSSFG)

Cepstral smoothing is a useful amendment to speech enhancement filters operating in real noise environments. Annoying noise fluctuations are prevented even in the case of babble noise. As opposed to conventional methods, Cepstral smoothing allows for a selective smoothing of different spectral structures represented by the respective Cepstral coefficients. This makes the protection of the characteristics of speech possible while musical noise is suppressed. Cepstral smoothing preserves speech onsets, plosives, and quasi-stationary narrowband structures like voiced speech. The proposed recursive temporal smoothing is applied to higher Cepstral coefficients only, excluding those representing the pitch information. As the higher Cepstral coefficients describe the finer spectral structure of the Fourier spectrum, smoothing them along time prevents single coefficients of the filter function from changing excessively and independently of their neighboring bins, thus suppressing musical noise. The proposed Cepstral smoothing technique is very effective in nonstationary noise.

VII. SPECTRAL SUBTRACTION IN THE SHORT TIME MODULATION DOMAIN (SSSTMD)

Speech enhancement aims at improving the quality of noisy speech. This is normally accomplished by reducing the noise (in such a way that the residual noise is not annoying to the listener), while minimizing the speech distortion introduced during the enhancement process. In this technique, the modulation domain has been investigated as an alternative to the acoustic domain for speech enhancement. More specifically, it determines how competitive the modulation domain is for spectral subtraction as compared to the acoustic domain. For this purpose, the traditional analysis, modification, synthesis and framework to include modulation domain processing has been extended. Then it compensates the noisy modulation spectrum for additive noise distortion by applying the spectral subtraction algorithm in the modulation domain. Using an objective speech quality measure as well as formal subjective listening tests, it has been showed that the proposed method results in improved speech quality. Furthermore, the proposed method achieves better noise suppression than the MMSE method.

VIII. ADAPTIVE NOISE CANCELLATION IN THE WAVELET DOMAIN (ANCWD)

Adaptive filtering has been used for speech denoising in the time domain. During the last decade, wavelet transform has been developed for speech enhancement. In this paper we are proposing to use adaptive filtering in the Wavelet transform domain. We propose a hybrid method of using adaptive filters on the lower scales of a wavelet transformed speech together with the conventional methods (Thresholding, Spectral Subtraction, and Wiener filtering) on the higher scale coefficients. The results demonstrate that the suggested approach is computationally efficient and has a good performance.

IX. SPEECH ENHANCEMENT USING ADAPTIVE NEURO-FUZZY FILTERING (SEANFF)

It presents an adaptive neuro-fuzzy filtering scheme using the artificial neuro-fuzzy inference system (ANFIS) for noise reduction in speech. The measurable output noisy speech with 5dB SNR level is taken as the contaminated version of the interference to compare with the output data of the filter. This function returns the initial FIS structure that contains a set of fuzzy rules to cover the feature space. Finally, the ANFIS hybrid learning algorithm that combines the recursive least-squares estimation (RLSE) method and the back propagation gradient descent (BP/GD) is applied to determine the premise and the consequent parameters. After training, the ANFIS output (i.e. estimated interference) was determined. Then the estimated information signal is calculated as the difference between the measured signal and the estimated interference.

X. ADAPTIVE NOISE CANCELLATION USING CASCADED CORRELATION NEURAL NETWORKS (ANCCCNN)

The main objective of Speech Enhancement is to enhance the speech signal to obtain a clean signal with higher quality. Such system has been widely used in long distance telephony applications. A novel adaptive noise cancellation algorithm using cascaded correlation neural networks is described. In the proposed algorithm the objective is to filter out an interference component by identifying the non-linear model between a measurable noise source and the corresponding immeasurable interference. The cascaded correlation neural network algorithm has the powerful capabilities of learning and adaptation.

XI. WORKING ALGORITHM

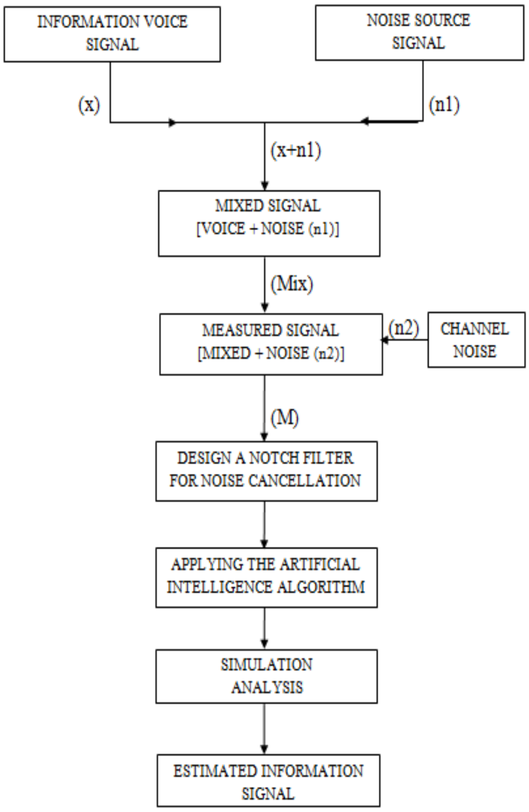

Fig- 1 Algorithm for SNR estimation

XII. SIMULATION STUDIES

MATLAB Simulations has been carried out to estimate the SNR. Various steps for estimation of the SNR are as follows:

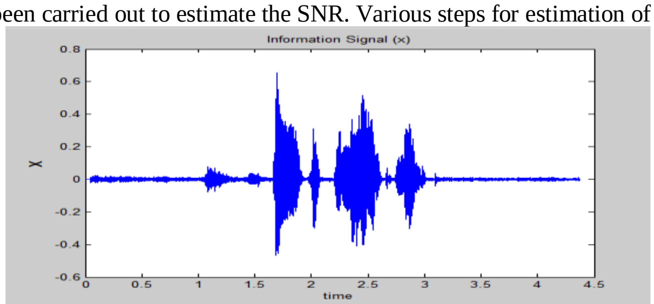

Fig- 2 Voice signal

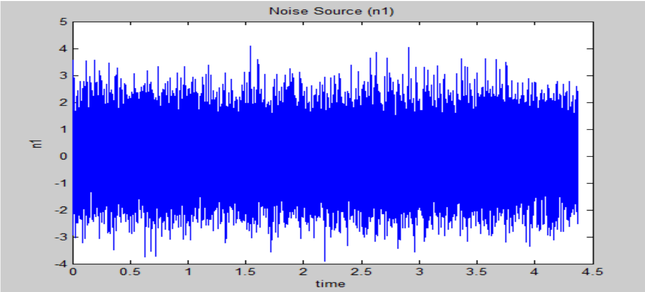

Fig- 3 Noise source signal

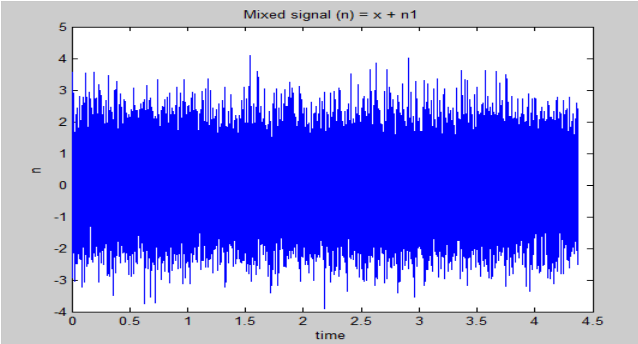

Fig- 4 Mixed signal

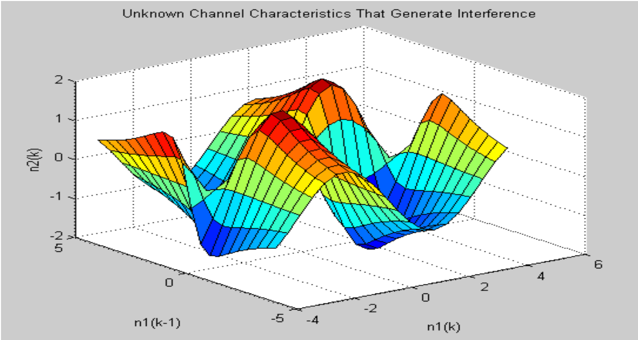

Fig- 5 Unknown channel characteristics

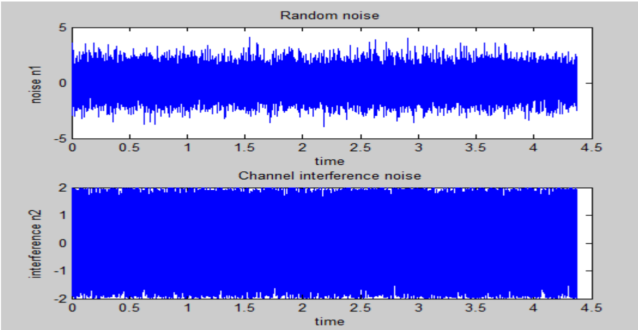

Fig- 6 Random and channel Noise

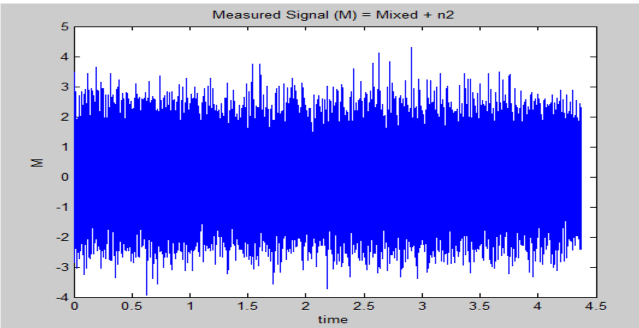

Fig- 7 Measured Signal

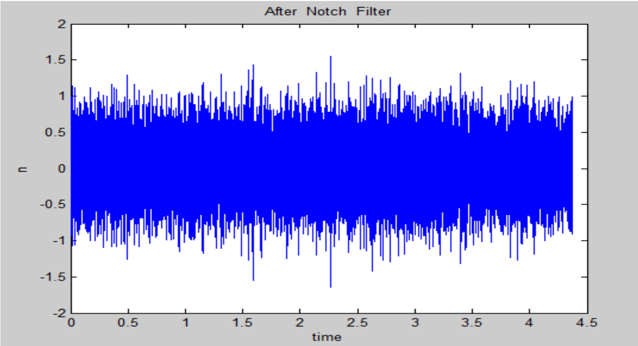

Fig- 8 After notch filter

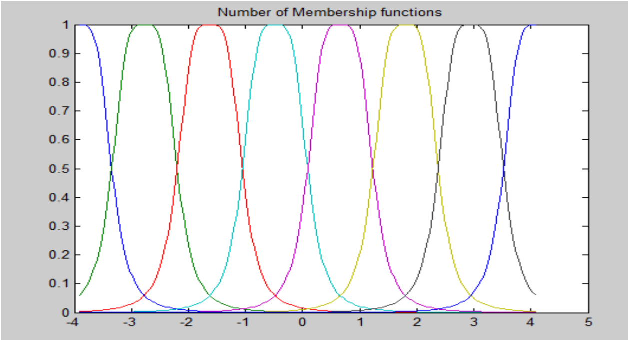

Fig- 9 Number of membership functions

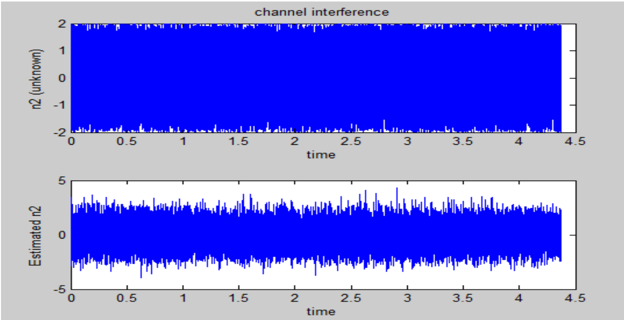

Fig- 10 Unknown and estimated Noise

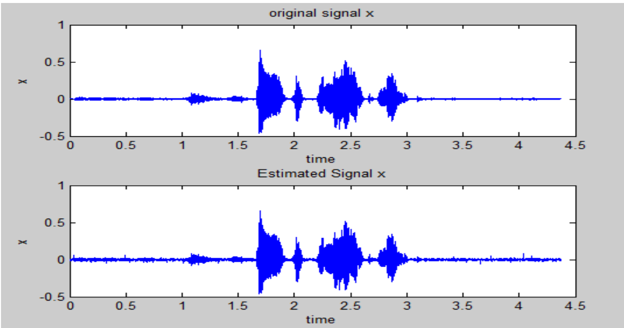

Fig- 11 original and estimated signal

XIII. RESULT ANALYSIS

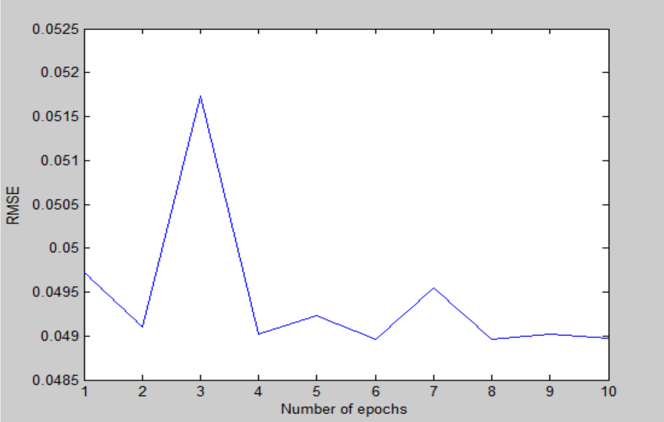

Fig- 12 RMSE v/s Number of epochs

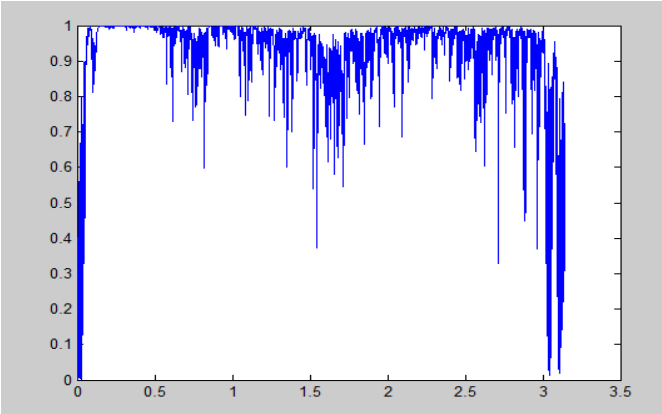

Fig- 13 Coherence between original and estimated signal

XIV. CONCLUSION

In this paper, some speech enhancement techniques are reviewed and after that it has been concluded that the fuzzy and neural algorithms used for minimizing noise from a set of sound files has given the best optimal results. The root mean square error (RMSE) and number of epochs are less and at the same time the membership functions are also less. Since a database consist of multiple voice files of one subject recorded with different words, the fuzzy tool require some fine tuning of parameters to arrive at best possible results where noise effect is minimum and RMSE is also close to zero and estimated SNR is very close to the original signal SNR. It can further be extended on images and videos. Therefore it is suggested that one should try to minimize noise using similar methodology for reducing noise form images and videos also as fuzzy algorithm has given us very promising results.

REFERENCES

[1]. Zhixin Chen "*Simulation of Spectral Subtraction Based Noise Reduction Method*" International Journal of Advanced Computer Science and Applications, Vol. 2, No.8, 2011.

[2]. Philipos C Loizou, Gibak Kim, "*Reasons why Current Speech Enhancement Algorithms do not Improve Speech Intelligibility and Suggested Solutions*", in IEEE Transactions on Audio, Speech and Language Processing, Vol. 19, NO. 1, January 2011.

[3]. Kuldip Paliwal, Belinda Schwerin, "*Single-channel speech enhancement using spectral subtraction in the short-time modulation domain*" in ELSEVIER 9 February 2010.

[4]. P. Loizou, Speech Enhancement: Theory and Practice. New York: Taylor & Francis, 2007.

[5]. J. Dheebal and A. Padmal "*Intelligent Adaptive Noise Cancellation using Cascaded Correlation Neural Networks*" IEEE-ICSCN 2007, MIT Campus, Anna University, Chennai, India. Feb. 22-24, 2007. pp. 178-182.

[6]. Colin Breithaupt, Timo Gerkmann, and Rainer Martin "*Cepstral Smoothing of Spectral Filter Gains for Speech Enhancement without Musical Noise*" IEEE SIGNAL PROCESSING LETTERS, *VOL. 14, NO. 12*, DECEMBER 2007.

[7]. Jingdong Chen, Jacob Benesty, Yiteng (Arden) Huang and Simon Doclo "*New Insights into the Noise Reduction Wiener Filter*". in IEEE Transactions on Audio, Speech, and Language Processing, *VOL.14, NO. 4*, JULY 2006.

[8]. Meng Joo Er, "Adaptive Noise Cancellation Using Enhanced Dynamic Fuzzy Neural Networks", IEEE Trans. Fuzzy Systems, Vol. 13, No. 3, June 2005, and pp: 331-342.

[9]. J. Thevaril et. al "*Speech Enhancement using Adaptive Neuro Fuzzy Filtering*", Proc. of 2005 Int. Sym. on Intelligent Signal Processing and Communication Systems, 13-16 December 2005, pp. 753 – 756.

[10]. Mohammad A Akhaee, Ali Ameri, "*Speech Enhancement by Adaptive Noise Cancellation in the Wavelet Domain*", in ICICS 0-7803-9282-5/05/$20.00 ©2005 IEEE.

[11]. Herald Gustafson, Sven Erik Nordholm and Ingvar Claesson "*Spectral Subtraction Using Reduced Delay Convolution and Adaptive Averaging*" in IEEE Transactions on Speech and Audio Processing, Vol. 9, No. 8, November 2001.

[12]. Bin Zhang "*Fuzzy System with Adaptive Rule base*", 900-903, IEEE International Fuzzy Systems Conference, 2001

[13]. C.F.Juang, X.-T.Lin, "*A recurrent self-organizing neural fuzzy inference network,*" IEEE Trans. Neural Networks, vol. 10, no. 4, pp. 828–845, Jul. 1999.

[14]. Javier Onega-Garcia and Joaquin Gonzalez "*Overview of Speech Enhancement Techniques for Automatic Speaker Recognition*" in Proc. of IEEE Int.Joint Conference, 1997

[15]. S. F. Boll, "*Suppression of acoustic noise in speech using spectral subtraction*," in IEEE Trans. Acoust Speech and Signal Processing, vol.-2, ASSP-27, no.2, pp-113-120, 1979