IJARIIT
INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume3, Issue4)
Available online at www.ijariit.com

Speech Enhancement Using Dual Transform-Normalized LMS Algorithm for Speech Recognition Application

Dr. D. Deepa
Professor, Department of ECE,
Bannari Amman Institute of Technology
ddeepa@bitsathy.ac.in

Dr. C. Poongodi
Professor, Department of ECE,
Bannari Amman Institute of Technology
poongodic@bitsathy.ac.in

Abstract: Speech enhancement is one of the important preprocessing techniques required for any speech processing applications. This work represents dual channel speech enhancement where the desired signal is available and the input noisy speech is enhanced based on the reference signal, this method of speech enhancement can be used in robots where the machine can recognize the comments given by a human. In this paper noisy and the desired speech signals are dual transformed using discrete cosine transform and Hadamard transform and applied to the adaptive filter using Normalized Least Mean Square algorithm. In NLMS the step size parameter is varying based on the input signal rather in LMS the step size is constant. This variable step size will lead to fast convergence of noisy speech towards desired speech and the enhanced signal gives better performance compared to conventional LMS algorithm. The performance analysis is done through various subjective and objective measures.

Keywords: Speech Enhancement, Adaptive Filtering, Normalized Least Mean Square Algorithm, Variable step size, Speech Recognition, Signal to Noise ratio, Mean Square Error, IS distance.

I. INTRODUCTION

Processing of Speech signal relates to the compression, enhancement or recognition of speech signals. Ease and speed of representing, storing, retrieving and processing speech data have contributed to the development of efficient speech processing techniques to address the issues related to speech. The main aim of speech enhancement technique [8] is to improve perceptual aspects of speech by reducing noise contents. This is important in a variety of contexts, like environments with interfering background noise in speech recognition systems, hands-free environment for cars, hearing aid devices etc. The other important use of speech enhancement is to improve the perceptual quality of speech in order to reduce listener's fatigue condition.

In dual channel system, one channel has the noisy signal that is to be processed and the other channel consists of the reference signal. The adaptive filter is used for this type of noise cancellation or speech enhancement. The ultimate goal of this work is to eliminate the additive noise present in the speech signal and restore the speech signal to its original form. In real time speech recognition application, if the input is noisy speech the system cannot recognize the signal due to clean desired signal, and also convergence rate is very important because it is real time application. In order to have fast convergence Normalized Least Mean Square algorithm is used.

Objective speech quality measures analyzed [15] in this work are given below:

i) Signal to Noise Ratio (SNR): SNR is the ratio of signal power to noise power expressed in decibels (dB)

ii) Itakuro - Saito (IS) distance measure: IS distance gives the information about the phase difference between the reference signal and the enhanced signal. Normally the IS distance value should be minimum for better enhancement.

iii) Perceptual Evaluation of Speech Quality (PESQ): PESQ measure [1] is one of the most commonly used measures to predict the subjective opinion score of a degraded or enhanced speech. It is to assess the quality of the enhanced speech signal. The PESQ score ranges from 1 to 5, with a higher score indicating better quality.

iv) Mean Square Error (MSE): Mean Square Error (MSE) metric is frequently used in signal processing and is defined as the average value of the square of the difference between clean and enhanced signal.

II. LITERATURE SURVEY

The works related to speech enhancement using adaptive filtering is studied and few important references are discussed below

Improved speech to noise ratio is obtained in the adaptive filtering method proposed by [9]. This method improves speech to noise ratios by eliminating spectral regions with intense noise samples. This mechanism was examined in static listening conditions with sensorineural hearing loss patients.

An adaptive subband noise removal algorithm is proposed by [11]. It performs binaural preprocessing of speech signals for a hearing aid. The Multi Microphone Subband Adaptive (MMSBA) signal processing scheme uses the Least Mean Square (LMS) algorithm in frequency limited subbands. Subband approach will perform better due to the separation of the signal based on its frequency the elimination of high-frequency noise will be better in adaptive algorithm. The results show that there were some distortion and a considerable amount of noise present in the output signal, which will lead to reduced intelligibility and create listener fatigue.

Two methods proposed in [13],[14] for dual channel, in the first method the noisy speech signal is transformed using DCT and processed using the adaptive algorithm, in another method the noisy speech is transformed using DFT and processed using the adaptive algorithm. This method has poor intelligibility.

III.P PROPOSED DUAL CHANNEL SPEECH ENHANCEMENT ALGORITHM

In this paper, a dual channel speech enhancement algorithm is proposed with the use of adaptive filter using Normalized least mean square (NLMS) algorithm. Adaptive filter [2] processes two input signals. One of the input signals is the noisy speech to be enhanced and the second input is the reference signal. In this work input, noisy speech and the reference signal are transformed from time domain to frequency domain using discrete cosine transform and Hadamard transform. This is to separate high-frequency noise from a noisy speech signal for speech enhancement process since analysis of the signal is effected in the frequency domain rather in the time domain.

Speech enhancement using NLMS algorithm with DCT and Hadamard transformation of input signals as a preprocessing technique is composed of four stages. Initially desired speech and noisy speech signals are transformed using Discrete Cosine Transform (DCT). Transformed signal is then processed using Hadamard transform, which aligns the DCT samples based on its frequency; this process is used for reducing high-frequency noise with adaptive step size in NLMS algorithm. In the third stage dual transformed signal is normalized using power normalization technique and finally, these signals are applied to an adaptive filter where NLMS algorithm is used for adaptation. Due to the arrangement of samples based on its frequency the enhancement is done promptly using NLMS algorithm.

A. Discrete Cosine Transform

Generally, transforms will convert time domain signal to frequency domain signal. Discrete Cosine Transform converts the input time domain speech signal into a frequency domain signal by representing it as coefficients. It consists of real valued components. With a smaller amount of coefficients, this will give a better approximation. It has strong energy compaction property which means that the signal information will be available in few low-frequency components.

B. Hadamard Transform

In the second stage, Hadamard transformation is applied to the cosine transformed noisy speech signal and the reference signal. The Hadamard transform is a square matrix of $2^m \times 2^m$, it is scaled by a normalization factor, which transforms 2^m real time domain numbers x(n) into 2^m real frequency domain numbers X(k). The 1 × 1 Hadamard transform H_0 is defined by the identity $H_0 = 1$, and then H_m is defined for $m > 0$

$$H_m = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{m-1} & H_{m-1} \\ H_{m-1} & -H_{m-1} \end{pmatrix} \tag{8}$$

Equation (8) specifies the Hadamard transform of order ‘m’, where the 1/√2 is a normalization value. Other than the normalization factor, the Hadamard matrices are made up entirely of 1 and −1. The same Hadamard Transform equation is used during the inverse process after adaptation.

C. Power Normalization

The dual transformed signal is then normalized by the square root of its power $p_k(i)$. The powers are estimated based on the sliding rectangular window.

The power normalized signal $v_k(i)$ is given in equation (9)

$$v_k(i) = \frac{u_k(i)}{\sqrt{p_k(i) + \xi}} \tag{9}$$

Where $p_k(i) = \beta\, p_{k-1}(i) + (1-\beta)\, u^2{}_k(i)$, β is the normalization constant between 0 to 1, the values are chosen based on the performance of the system and $p_{k-1}(i)$ is the power value of the previous sample. The small constant ξ is introduced to avoid numerical instabilities when $p_k(i)$ is close to zero. These processes of Hadamard transform and power normalization speed up the convergence of the adaptive weights of the filter. The output vector after power normalization is specified in equation (10).

$$v_k(i) = [\,\mathrm{v_k}(0), \mathrm{v_k}(1), \ldots.\ \ldots \mathrm{v_k}(\mathrm{i}-1)]^{\mathrm{T}} \tag{10}$$

Further, the dual transformed, power normalized speech samples are applied as inputs for the adaptive filter using NLMS algorithm

D. Adaptive Filtering

The adaptive filter is used for real time data processing, where filtering is done adaptively, i.e., weight updating of the filter is adaptive based on the input signals. Transform domain adaptive filters have been proposed by [5]. The key component in designing the adaptive filter is of different algorithms used for weight updating. In this work NLMS algorithm is used for weight updating and it has adaptive step size parameter based on input signal.

i) Normalized Least Mean Square (NLMS) Algorithm

The main drawback of conventional LMS algorithm is fixed step size parameter for every iteration [3]. Before adaptive filtering, the step size must be fixed. This requires a knowledge of the input signal prior to initiation of the adaptive filtering. The NLMS algorithm overcomes this issue by finding the step size value with respect to the input signal. This step size is proportional to the inverse of the total expected energy of the instantaneous values of the coefficients of the input vector x(n)[10]. Figure 1 shows the block diagram specifying the NLMS algorithm based adaptive filter (Simon Haykin 2002), (Monson H. Hayes 2005). The DCT-Hadamard transformed the noisy input signal x(n) and the error signal e(n) is given as inputs to the weight control mechanism. Weight update for the adaptive NLMS filter is done automatically based on the error signal obtained from the subtraction of estimated signal from adaptive filter and reference input. The process is continued with a number of iterations until the error signal becomes negligible or zero. Finally, the estimated signal from the adaptive filter is the required enhanced signal.

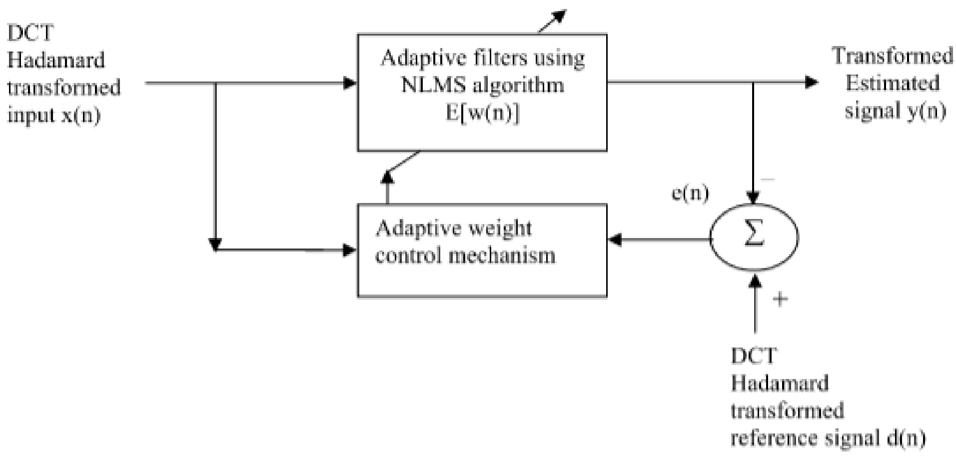

Fig 1: Block Diagram for proposed speech enhancement algorithm

The DCT and Hadamard transformed noise and desired speech signals are applied as inputs for adaptive filter, whose weights are adjusted using NLMS algorithm.

The weight vector in NLMS algorithm is denoted as $w_n(k)$. The update value of tap weight vector is given by the equation (11).

$$\mathrm{W_{n+1}} = \mathrm{W_n} + \mu(\mathrm{n})\,\mathrm{e}(\mathrm{n})\,\mathrm{x}^*(\mathrm{n}) \tag{11}$$

The update of the k^{th} coefficient is given as

$$\mathrm{w_{n+1}}(\mathrm{k}) = \mathrm{w_n} + \mu(\mathrm{n})\,\mathrm{e}(\mathrm{n})\,\mathrm{x}^*(\mathrm{n}-\mathrm{k}) \tag{12}$$

Simplicity in implementation, Stable and robust performance against different signal conditions and fast convergence is the main advantage of NLMS algorithm:

The computation procedure for NLMS algorithm for a p^{th} order is given in Table 1. In that, the step size or learning rate parameter is μ(n) which is updated based on the samples of input noisy speech signal and the weight vector value is initialized as 0. For every value of n, i.e. for every sample in the input signal the corresponding error signal, weight value and the output are calculated.

The weight (w_n) is updated based on the error signal and the previous weight value (w_{n-1}). This procedure is repeated for a number of iterations until the error value becomes negligible or close to 0.

TABLE I
SUMMARY OF NLMS ALGORITHM

Parameters:	p = filter order
	$\mu(n)$ = variable step size
Initialization:	$w_0=0$
Computation:	for n = 0, 1, 2...
	$y(n)= w_n^T x(n)$
	$e(n)=d(n)-y(n)$
	$\mu(n)=1/[x^T(n).x(n)]$
	$W_{n+1}=w_n+\mu(n)e(n)x^*(n)$

The output of the NLMS filter is then inverse transformed with Hadamard and IDCT to get the time domain signal, where the transformed signals are in the frequency domain.

IV. EXPERIMENTAL RESULTS

The experimental results of the proposed DCT-Hadamard-NLMS algorithm are given in terms of time domain and frequency domain plots further the performance measures are analysed [4]. For time domain representation amplitude versus time plot is taken. The further spectrogram is given to represent the time versus frequency relationship for the proposed method.

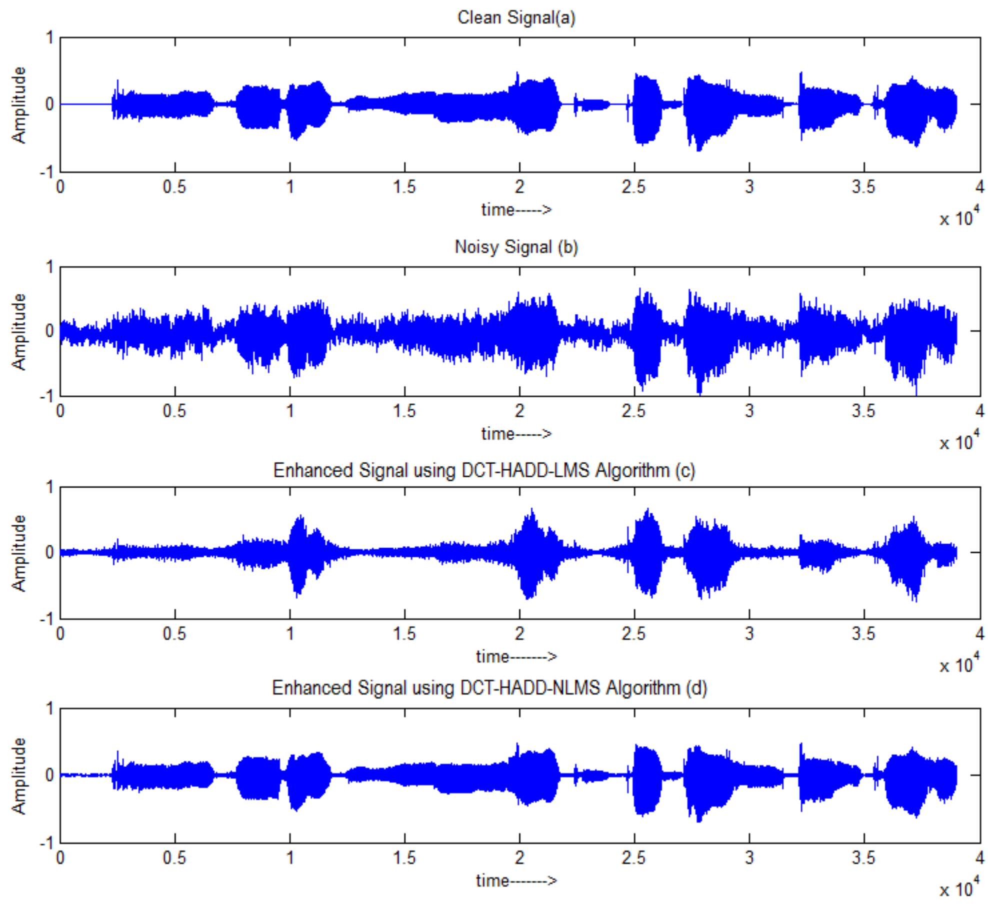

Fig 2 Time domain plot for pink stationary 0 dB noisy signal

Figure 2 shows the plots of the time domain results for pink stationary 0 dB signal of the proposed method. From the top, it is a noisy signal, clean signal, signal enhanced by DCT- HADD-LMS algorithm and signal enhanced by proposed method respectively. In the proposed method the signal is closure to the clean signal.

Figure 3 shows the spectrogram of the proposed method and compared with the conventional method. The spectral components of the conventional method have some noise components with reduced speech level and in the proposed method it is similar to the clean signals spectrogram.

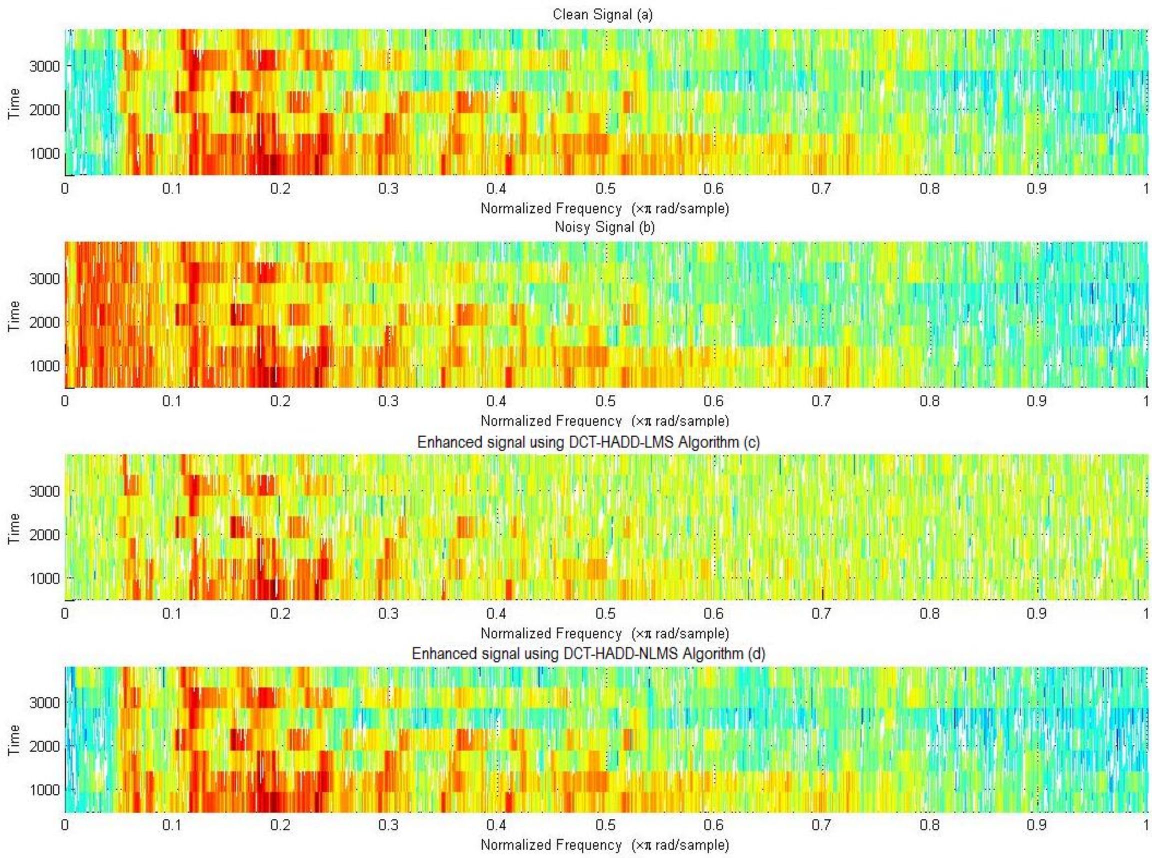

Fig 3 Spectrogram plot for pink stationary 0 dB noisy signal

The performances of the proposed dual channel speech enhancement algorithms are analyzed based on the objective and subjective quality measures.

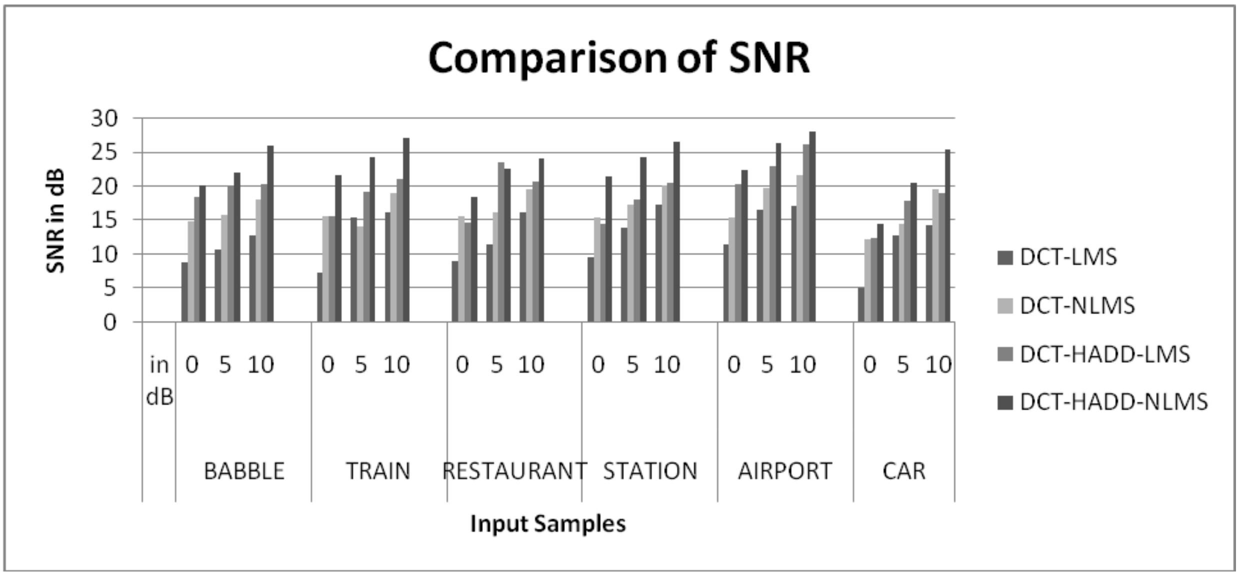

Fig 4 Comparison of SNR values

Figure 4 illustrates the SNR comparison of the proposed method with DCT-LMS, DCT-NLMS and DCT-HADD-LMS algorithms. From the results, it is identified that SNR value of DCT-HADD-NLMS is increased compared to all other methods due to adaptive step size and application of Hadamard transform in addition to the cosine transform. Table 2 gives the MSE values of the proposed dual channel speech enhancement algorithms for NOIZEUS database samples. It shows that the MSE values are reducing when the conventional adaptive algorithms are preprocessed with the Hadamard transform.

TABLE III
COMPARISON OF MEAN SQUARE ERROR VALUES OF PROPOSED DUAL CHANNEL ALGORITHMS

Test samples in the following noisy condition	Input SNR in dB	MSE	
		DCT -HADD-LMS	DCT-HADD-NLMS
Babble	0	0.0014	0.00042
	5	0.001	0.0002538
	10	0.030925	0.0080675
Train	0	0.0031	0.0017865
	5	0.096009	0.001149
	10	0.029822	0.004301
Airport	0	0.0215	0.007427
	5	0.034	0.003795
	10	0.0110	0.005101
Car	0	0.0037	0.000406
	5	0.00014	0.000192
	10	0.0094	0.00090
Exhibition	0	0.00035	0.000237
	5	0.00012	0.000019
	10	0.00069	0.000103

TABLE IIIII
COMPARISON OF PESQ VALUES OF PROPOSED DUAL CHANNEL ALGORITHMS FOR NOIZEUS DATABASE SAMPLES

Test samples in the following noisy conditions	Input SNR in dB	PESQ			
		DCT-LMS	DCT-RLS	DCT-HADD-LMS	DCT-HADD-RLS
Airport	0	1.9729	2.2502	2.4245	3.3336
	5	2.1271	2.4819	2.8346	3.5726
	10	2.4620	2.7638	3.3802	3.7996
Babble	0	1.7869	2.1572	2.2933	4.1857
	5	1.8264	2.4449	2.7819	3.8805
	10	2.2132	2.6509	3.4340	3.7253
Car	0	2.0660	2.2712	2.3887	3.5795
	5	2.1820	2.547	2.7442	3.5537
	10	2.3958	2.7107	3.4970	3.7647
Exhibition	0	2.0715	2.2986	2.1575	4.0765
	5	1.5186	2.7949	2.6397	3.3882
	10	2.1180	2.6118	3.3548	3.7250

Based on the PESQ values given in Table 3, it is found that the perceptual evaluation of subjective quality is well improved in DCT-Hadamard-NLMS method with the values in the range of 3.3 to 4.1. From the above results and plots, it is concluded that the performance of DCT-Hadamard-NMLS algorithm is better compared to the conventional and the DCT-Hadamard-LMS algorithm

CONCLUSIONS

In this paper dual channel speech enhancement algorithm is proposed and the performances of the adaptive filter using NLMS algorithm is discussed. It is identified by the proposed method that adaptive filter operated in frequency domain input signals perform well than that of the input signals in time domain. Compared to the conventional LMS algorithm with DCT and Hadamard transformed input signals, improvement in SNR is obtained in the proposed method for different types of noisy environment. The mean square error also reduced in the proposed method due to the adaptive step size parameter, and also the subjective measures obtained for quality improvement, which shows the clearance of speech signal after enhancement.

REFERENCES

[1]. A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "*Perceptual evaluation of speech quality (PESQ)-A new method for speech quality assessment of telephone networks and codecs*," in Proceedings of IEEE International Conference on Acoustics, speech, and signal processing vol. 2, pp. 749–752, 2001.

[2]. Dessouky, M.I, Diab, S.M., Abd El-Fattah, M.A. and Abd El-Samie, F.E., "Speech *Enhancement using an Adaptive Wiener Filtering* approach", Progress in Electromagnetics Research M, Vol. 4, pp. 167-184, 2008.

[3]. D. Deepa and Dr. A. Shanmugam (2010), "*Dual Channel Speech Enhancement Using Hadamard-LMS Algorithm with DCT Preprocessing Technique*" in International Journal of Engineering Science and Technology (IJEST), Vol. 2 (09) PP 4424 to 4430.

[4]. D. Deepa and Dr. A. Shanmugam (2009), "*Time And Frequency Domain Analysis Of Subband Spectral Subtraction Method Of Speech Enhancement Using Adaptive Noise Estimation Algorithm*" in International Journal of Engg. Research & Indu. Appls. (IJERIA)., ISSN 0974-1518, Vol.2, No. VII, PP 57-72.

[5]. Francosie Beaufays "*Transform domain adaptive filters: An analytical approach*", IEEE Transactions on Signal Processing, Vol. 43, No. 2, pp. 422-431, 1995.

[6]. L.R.Rabiner and R.W.Schafer, "*Digital Processing of Speech Signals*". Pearson Publications 2004.

[7]. Monson H. Hayes, "*Statistical Digital signal processing and modeling*", John Wiley & Sons, 2005.

[8]. Philipos C. Loizou, "*Speech Enhancement Theory and Practice*", CRC Press, Taylor& Francis, 2007.

[9]. Rankovic.C.M., "*Factors governing speech reception benefits of adaptive linear filtering for listeners with sensorineural loss*". Journal of the Acoustical Society of America, 103(2). pp. 1043-57, 1998.

[10]. Shaul Florian and Neil J Bershad. "A Weighted Normalized Frequency Domain LMS Adaptive Algorithm". IEEE Transactions on Acoustics speech and signal processing. Vol. 36, no. 7, 1998.

[11]. Shields, P.W. and Campbell, B.R. "*Improvements in the intelligibility of noisy reverberant speech using a binaural sub band adaptive noise-cancellation processing scheme*", Journal of the Acoustical Society of America, Vol. 110, No. 6, pp. 3232-3242, 2001

[12]. Simon Haykin, "*Adaptive Filter Theory*", Pearson Education Asia, 4th Edition, 2002.

[13]. Sunitha. S.L., and V. Udayashankara. "*Fast Factored DCT-LMS Speech Enhancement for Performance Enhancement of Digital Hearing Aid*". World Academy of Science, Engineering, and Technology, 10, 2005.

[14]. Sunitha S L and Dr.V Udayashankara," *DFT-LMS Speech Enhancement Technique for Sensorineural loss Patients*". Journal of Bioinformatics India, Vol 3, Jan- March 2005.

[15]. Yi Hu and Philipos C. Loizou, "*Evaluation of Objective Quality Measures for Speech Enhancement*", IEEE Transactions on Audio, Speech and Language Processing, Vol. 16.No 1 PP. 229-238, 2008.

