IJESMR

International Journal OF Engineering Sciences & Management Research

SPEECH SIGNAL ENHANCEMENT IN HEAVY NOISE INDUSTRIES

Motuma M. Abafogi and Askin Demirkol
Dept. of Electrical and Electronics Engineering Sakarya University, Turkey

KEYWORDS: Speech Enhancement, Kalman Filter, Least Mean Square, Spectral Subtraction, VAD

ABSTRACT
Audio signal in speech communication is corrupted by an additive random noise where there is a noisy environment. These noisy surroundings might be a moving car, train, factory, or a noisy telephone channel. Since noise is random and varying continuously, we need to estimate the noise at every instant to remove it from the desired signal. There are many schemes for noise reduction but the most efficient scheme to accomplish noise cancellation is to employ adaptive filters. In speech signal enhancement there are two categories of algorithms in which either a single microphone or multiple microphones are employed to clean up the noisy signal. In this paper, Matlab simulations of different adaptive algorithms and comparison of their performances for noise cancellation in a noisy environment, specifically industrial noise, is carried out. A comparative analysis of different adaptive filters in the presence of a single and dual microphone is provided. A robust voice activity detector (VAD) is incorporated in the single channel speech enhancement.

INTRODUCTION
Noisy environments usually constitute most part of our daily lives. The noise in automobiles, trains, or work areas like factories, which for the most part is unavoidable, makes communication between two individuals quite difficult. In order to improve the poor voice quality, speech enhancement becomes a necessity in these kinds of noisy environments. Some of the application areas of noise cancellation products are hearing aids, mobile phones, teleconferencing etc [1]. In this paper, the comparative analysis of some of the adaptive filtering techniques used in speech signal enhancement where a single and multiple input signals are available in a heavy noise industry is described.

SINGLE MICROPHONE NOISE CANCELLATION
The only available signal in a single microphone system is a noisy speech signal. In this kind of environment, the same microphone captures both voice and noise. From among the various techniques used to suppress the noise in the signal, the spectral subtraction technique employing the robust multiple observation likelihood ratio test (MO-LRT) voice activity detection (VAD) is used.

A. *Voice Activity Detection* (VAD)

To reduce the effect of noise a number of noise reduction techniques have been developed and these techniques often require a precise estimate of the noise statistics. The development of an efficient voice activity detection system is still an active field of research. From among the different VAD techniques mentioned in [2], the multiple observation likelihood ratio test (MO-LRT) which was described in [3] and [4] is employed so as to enhance the output signal quality of the single microphone.

For 2m+1 $\{y_{l-m},...,y_{l-1},y_l,...,y_{l+m}\}$ observation vectors, the MO-LRT is given as [4]:

$$l_{l,m}=\sum_{l-m}^{l+m}\ln\frac{p_{y_k/w_1}(y_k/w_1)}{p_{y_k/w_0}(y_k/w_0)} \qquad (1)$$

Where $p(w_i/y)$ is the posterior probability, l represents the frame being classified as speech (w1) or non-speech (w0) which is in fact the pure noise. The likelihood ratio test (LRT) can be computed recursively as follows.

IJESMR
International Journal OF Engineering Sciences & Management Research

$$l_{l+1,m} = l_{l,m} - \phi(l-m) + \phi(l+m+1) \qquad (2)$$

where ,

$$\phi(k) = \ln \frac{p_{y_k/w_1}\left(y_k / w_1\right)}{p_{y_k/w_0}\left(y_k / w_0\right)} \qquad (3)$$

With this the classification is made by comparing $l_{l,m}$ with the decision threshold η which is experimentally determined.

For $l_{l,m} \geq \eta$ frame contains speech and otherwise non-speech.

The graph below shows a typical output of a VAD system employing MO-LRT technique where a voice recorded in a noisy industrial environment is used as an input.

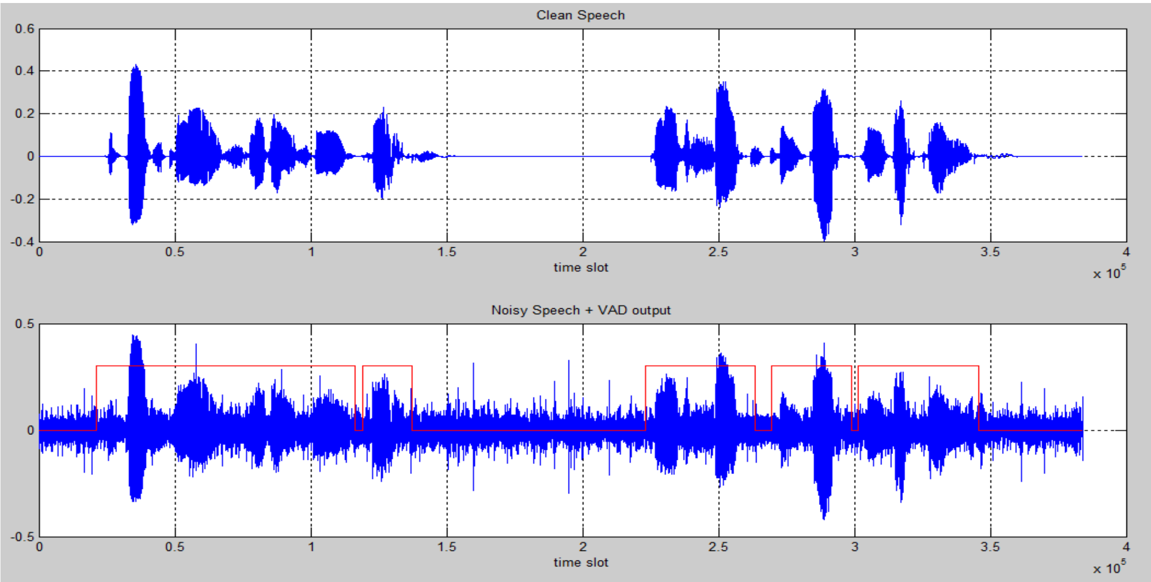

Figure 1: A typical output of a VAD system employing MO-LRT technique

A. Spectral Subtraction

Given a noisy speech signal y(m) as

$$y(m) = x(m) + n(m) \qquad (4)$$

where x(m) and n(m) are the clean speech and noise respectively, the frequency domain representation can be expressed as [1]

$$Y(k) = X(k) + N(k) \quad , k=0,1,\ldots,M-1 \qquad (5)$$

Where k denotes the discrete frequency variable and Y(k), X(k) and N(k) are the short time discrete Fourier transforms of the noisy speech, clean speech and noise respectively. The complex polar form representation is given as:

$$Y_k e^{j\theta_{Y_k}} = X_k e^{j\theta_{X_k}} + N_k e^{j\theta_{N_k}}, \quad k=0,1,\ldots,M-1 \quad (6)$$

IJESMR
International Journal OF Engineering Sciences & Management Research

where X_k, Y_k and N_k are the magnitudes of the frequency spectrums X(k), Y(k) and N(k) while θ_{Y_k}, θ_{X_k}, θ_{N_k} represent their corresponding phases.

Spectral subtraction is a technique used to restore the power or magnitude spectrum of a signal by subtracting an estimate of the average noise spectrum from the noisy signal spectrum [5]. The incoming signal y(m) is buffered and divided into segments of M samples length. These samples will be transformed into M spectral samples via discrete Fourier transform (DFT) after the samples are windowed using Hamming window. [5]

The frequency domain representation of the windowing operation will be

$$Y_w(f) = X_w(f) + N_w(f) \tag{7}$$

where *w* stands for windowed signals.

The spectral subtraction carried out in order to estimate the original clean speech spectrum is expressed as: [5][6][7]8]

$$\left|X_w(f)\right|^b = \left|Y_w(f)\right|^b - \alpha\left|N_w(f)\right|^b \tag{8}$$

Where *N(f)* is the time averaged noise spectra. The parameter α is 1 for full noise subtraction and greater than 1 for over subtraction while b is 1 for magnitude spectral subtraction and 2 for power spectral subtraction. In this paper the magnitude spectral subtraction with the MO-LRT voice activity detection technique described above is used. The block diagram below shows a simplified representation of the spectral subtraction operation.

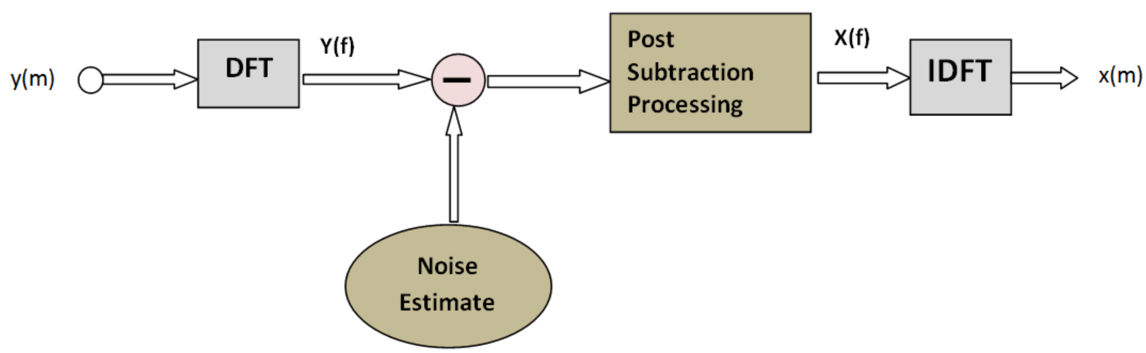

Figure 2: A simplified representation of a spectral subtraction technique [5]

In Figure 3, a voice recorded in a noisy industrial environment is used as an input to the developed MO-LRT based Spectral Subtraction method

IJESMR
International Journal OF Engineering Sciences & Management Research

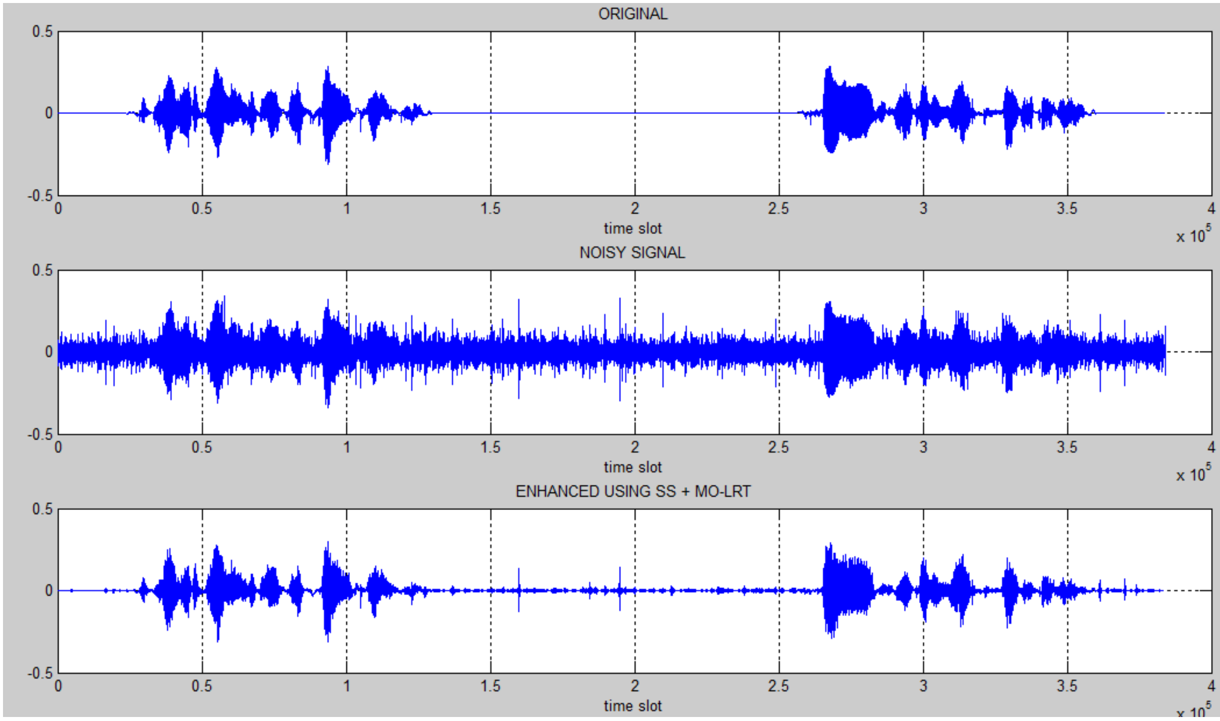

Figure 3: Output of a MO-LRT based SS technique

DUAL MICROPHONE NOISE CANCELLATION

In a situation where it is convenient to have more than one microphone a superior performance could be achieved as compared to a single microphone noise cancellation system. In a dual microphone system, the primary and reference microphones should be placed properly in such a way that the speech level at the primary microphone is higher than the level at the reference microphone. In this paper, the least mean square (LMS), Kalman filter and Kalman filter based LMS filter will be used for the case of a dual microphone.

A. *Least Mean Square* (LMS)

The least mean square algorithm is one of the most widely used and computationally simple noise reduction algorithms. The basic steps involved in LMS technique are [9][10]:

i. Computation of the output y(n) as:

$$y(n) = w^T(n)x(n) \qquad (9)$$

Where x(n) and w(n) are the input and weight vectors respectively

ii. Estimation of error e(n) as:

$$e(n) = d(n) - y(n) \qquad (10)$$

Where d(n) is the desired signal

iii. Update the weight vector as:

$$w(n+1) = w(n) + 2\mu e(n)x(n) \qquad (11)$$

B. Kalman Filter

The Kalman filter which was developed by R.E. Kalman in 1960 has a wide range of applications. It is based on the state space formulation of a discrete or continuous time systems. Kalman filtering technique using only a single microphone has been discussed in [11]. However in this paper, we consider a situation where there is an additional reference microphone. The one-dimensional speech signal is represented by the state space model of Kalman filter as:

IJESMR
International Journal OF Engineering Sciences & Management Research

$$x(k+1) = F(k)x(k) + n(k)$$
$$y(k) = H^T(k)x(k) + v(k) \qquad (12)$$

where $\mathbf{x(k)}$=[$x(k-L+1)$ $x(k-L+2)$ $x(k-L+3)$ ... $x(k)$]^T is the state vector ($x(k)$ is speech signal at time k)

F(k)=state transition matrix
n(k)=state noise
H(k)=observation matrix
V(k)=measurement noise

The autocorrelation of the state noise Q(k) and the measurement noise R(k) is expressed as:

$$Q(k) = E\left[n(k)n^T(k)\right]$$
$$R(k) = E\left[v(k)v^T(k)\right] \qquad (13)$$

C. Kalman Based LMS

Since the stability of the LMS algorithm depends on the step size μ, a more efficient version of LMS which is the normalized LMS is widely used. In this case the weight update in the LMS is modified as:

$$w(k+1) = w(k) + \frac{\alpha e(k)x(k)}{x^T(k)x(k) + q} \qquad (14)$$

Where 0<α<2 and q is small as compared to $x^T(k)x(k)$.

In the combined kalman based normalized LMS the weight is updated as [12]

$$w(k+1) = w(k) + \frac{e(k)x(k)}{P(k) + R(k)/\delta_w^2(k)} \qquad (15)$$

Where

$$\delta_w^2(k+1) = \delta_w^2(k)\left[1 - \frac{P(k)/N}{P(k) + R(k)/\delta_w^2(k)}\right] + Q(k) \qquad (16)$$

The figure below shows the enhancement of a noisy speech in an industrial environment using these various methods.

IJESMR
International Journal OF Engineering Sciences & Management Research

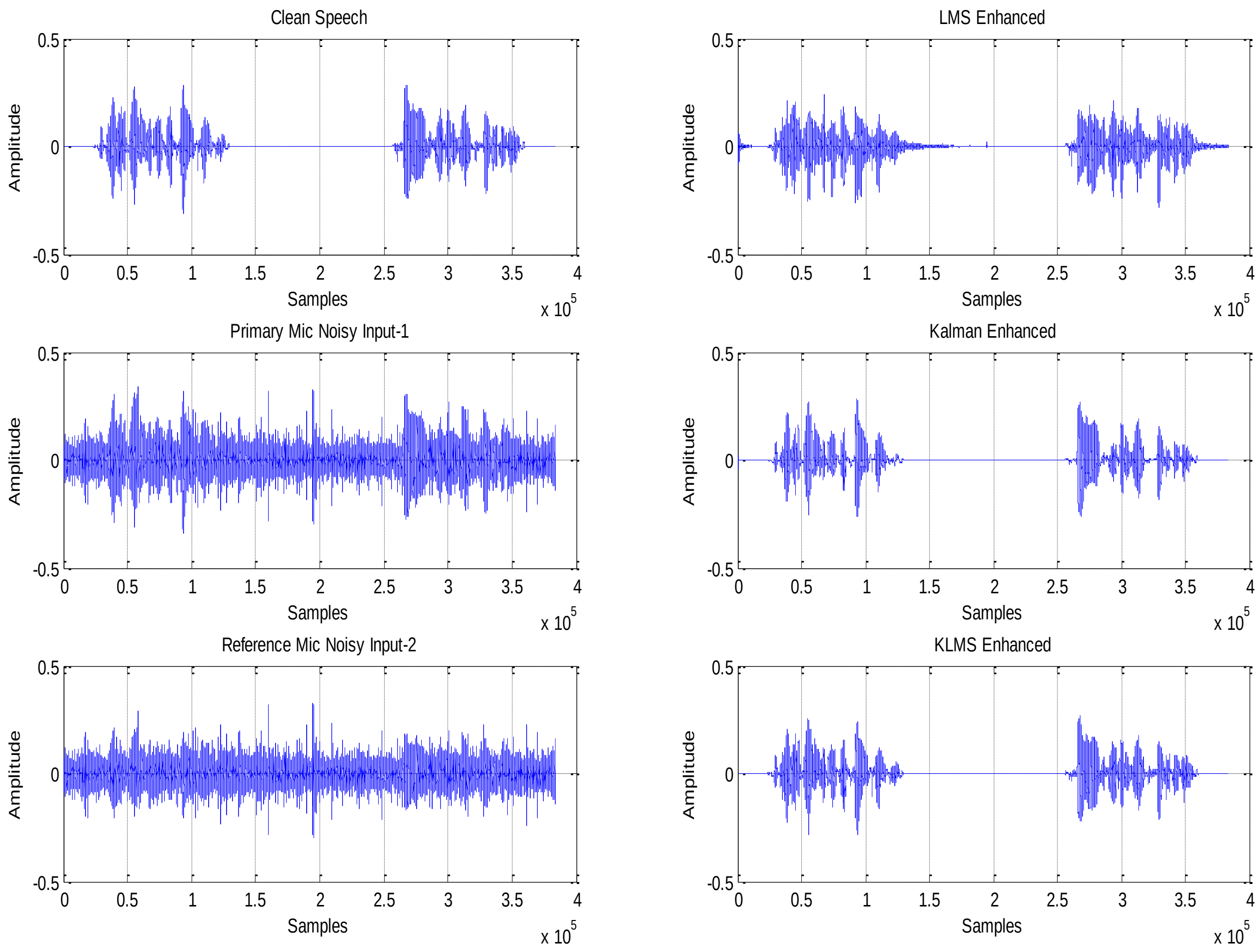

Figure 4: Output of the various methods

COMPARATIVE ANALYSIS

At this stage, we examine the performance of MO-LRT based spectral subtraction technique that employs a single microphone and the other adaptive filtering techniques mentioned above where there exists an additional reference microphone. From the mean square error (MSE) comparison given in Table 1, it is evident that in spite of a single microphone used in the MO-LRT based spectral subtraction technique the MSE obtained is very much closer to the dual microphone case. We can also notice from the table and the spectrogram in Figure 5 that the performance of LMS has improved when combined with the Kalman filter.

IJESMR

International Journal OF Engineering Sciences & Management Research

Algorithms	Mean Square Error (MSE)
Spectral Subtraction + MO-LRT	0.0000814
LMS	0.000809
Kalman Filter	0.0000491
KLMS	0.0000410

Table 1: MSE of the adaptive algorithms

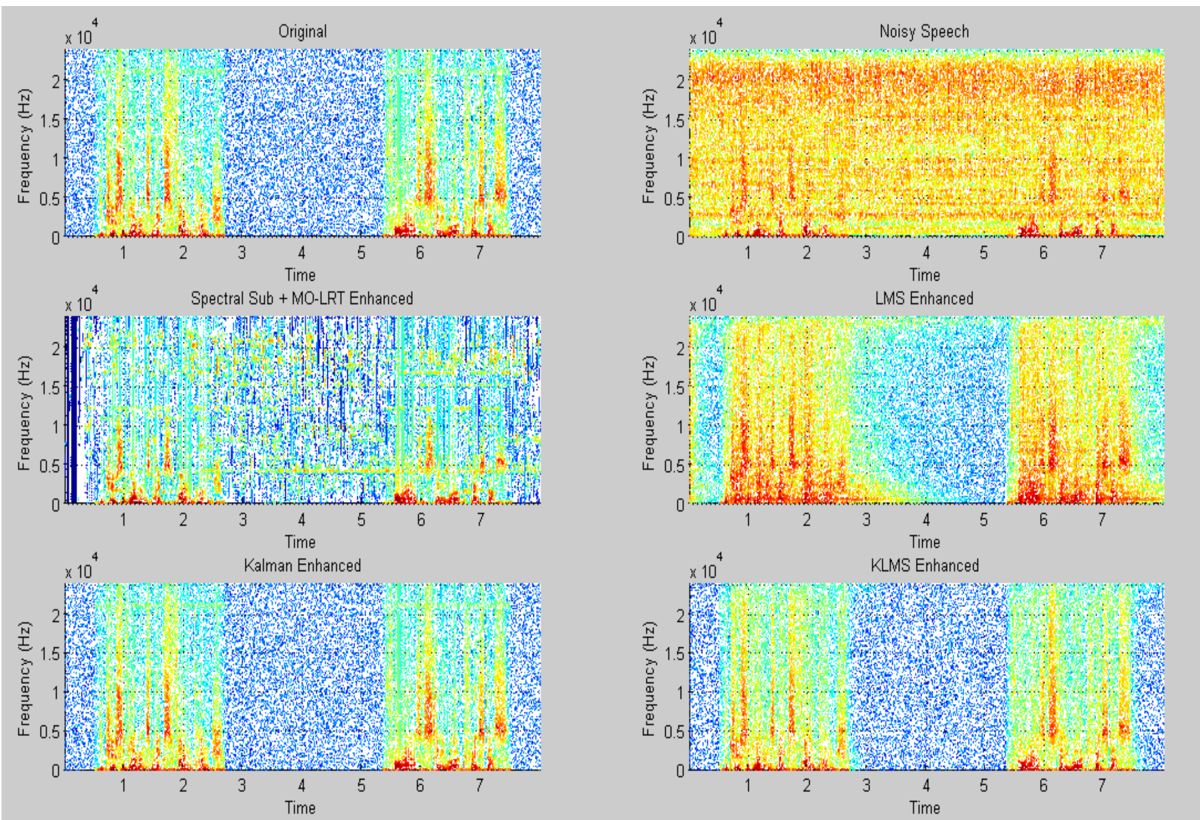

Figure 5: Spectrogram of the four adaptive filtering techniques

CONCLUSION

In this paper a comparative analysis of different adaptive filters has been conducted for the enhancement of a speech signal in heavy noise industries. A pretty good result has been obtained with the single microphone noise cancellation technique employing the MO-LRT based spectral subtraction technique. This is advantageous, because in many cases we usually only have a corrupted speech signal. With an additional reference microphone the performances of Kalman filter and Kalman based LMS are quite efficient.

IJESMR

International Journal OF Engineering Sciences & Management Research

REFERENCES

1. Saeed V. Vaseghi, ”*Multimedia Signal Processing: Theory and Appplications in Speech, Music and Communications*”, 2007 John Wiley & Sons Ltd.
2. Michael Grimm and Kristian Kroschel, “ *Robust Speech Recognition and Understanding*”, June 2007.
3. Lee Ngee Tan, Bengt J. Borgstrom and Abeer Alwan, “ *Voice Activity Detection Using Harmonic Frequency Components in Likelihood Ratio Test*”, ICASSP 2010.
4. Javier Ramirez, Jose C. Segura, “ *Stastical Voice Activity Detection Using a Multiple Observation Likelihood Ratio Test*”, IEEE Signal Processing Letters, VOL 12, No. 10, October 2005.
5. Saeed V. Vaseghi, “*Advanced Digital Signal Processing and Noise Reduction*”, John Wiley & Sons Ltd, 2000.
6. Poruba J.,”*Speech Enhancement Based on Non Linear Spectral Subtraction*”,IEEE Devices, Cicuits and Systems 2002.
7. Boll S.F., ”*Suppression of Acoustic Noise in Speech Using Spectral Subtraction*”, IEEE Trans ASSP 1979.
8. Kamath S. and Loizon P., “*A Multiband Spectral Subtraction Method for Enhancing Speech Corrupted by Colored Noise*”, Proc. IEEE conference on Acoustics, Speech, Signal Processing, 2002.
9. Sambur M.I., Nutley N.J., “*LMS Adaptive Filtering for Enhancing the Quality of Noisy Speech*”, IEEE Conference on ICASSP 1978.
10. Sayed A. Hadei, M. Lotfizad, “*A Family of Adaptive Filter Algorithms in Noise Cancellation for Speech Enhancement*”, IJCEE 2010.
11. Paliwal K.K., “*A Speech Enhancement Method Based on Kalman Filtering*”, IEEE Conference on ICASSP 1987.
12. Mahmoodzaadeh A., “*Speech Enhancement Using a Kalman Based Normalized LMS Algorithm*”, Telecommunications, 2008, IST symposium.