ISSN: 2229-6948 (ONLINE) ICTACT JOURNAL OF COMMUNICATION TECHNOLOGY, MARCH 2010, VOLUME: 01, ISSUE: 01

DOI: 10.21917/ijct.2010.0001

SPECTRAL SUBTRACTION METHOD OF SPEECH ENHANCEMENT USING ADAPTIVE ESTIMATION OF NOISE WITH PDE METHOD AS A PREPROCESSING TECHNIQUE

D. Deepa[1], A. Shanmugam[2]

Department of Electronics and Communication, Bannari Amman Institute of Technology, Sathyamangalam, India
E-mail: deepa_dhanaskodi@yahoo.co.in[1], dras@yahoo.co.in[2]

Abstract
Speech Enhancement can be used as a preprocessing technique in any of the speech communication systems. Speech communication involves a speaker, listener and communication device. The background noise present in the speech signal while transmitting has to be removed from the noisy speech signal to increase the signal intelligibility and to minimize the listener fatigue. The proposed approach is a speech enhancement method based on the spectral subtraction method, and the preprocessing is done by using partial differential equation. This method provides a greater degree of flexibility and control on the noise subtraction levels that reduces artifacts in the enhanced speech, resulting in improved speech quality and intelligibility. This method can be applied as a speech enhancement technique in Digital hearing aids, where sensor neural loss patients need 5dB to 10 dB higher SNR than normal hearing subjects.

Keywords
Spectral Subtraction, Partial differential equation, Adaptive noise estimation, Signal to Noise ratio, Mean Opinion score

1. INTRODUCTION

Speech enhancement is to improve the performance of speech communication systems in noisy environments. The corruption of speech due to presence of additive background noise cause severe difficulties in various communication environments. In single channel system, speech enhancement is a challenging one because reference noise signal will not be available for enhancement. The clean speech cannot be processed prior to being affected by the noise. This is one of the most difficult situations in speech enhancement for a single channel system. The conventional power spectral subtraction method for single channel speech enhancement substantially reduces the noise levels in the noisy speech but introduces residual (musical) noise.

The proposed method is a spectral subtraction method of speech enhancement with partial differential equation (PDE) method as a preprocessing technique. In this method first the Input noisy speech signal is enhanced using PDE by taking the adjacent samples and calculating the gradient, influencing coefficients and then the output of this process is applied to the input for spectral subtraction method. The noise estimate in spectral subtraction is updated by averaging the noise speech spectrum using a time and frequency dependant smoothing factor, which is adjusted based on signal presence probability in subbands. Signal presence is determined by computing the ratio of the noisy speech power spectrum to its local minimum, which is computed by averaging past values of the noisy speech power spectra with a look-ahead factor. This local minimum estimation algorithm adapts very quickly to highly non stationary noise environments [13]. Noise estimation algorithm in this method outperforms the standard power spectral subtraction method resulting in superior speech quality and largely reduced musical noise in single microphone system for both stationary and non stationary noise environments.

2. PARTIAL DIFFERENTIAL EQUATION TECHNIQUE

Speech affected by additive background noise can be enhanced by this method. First step in speech enhancement using PDE is to obtain the gradient (g) of each sample in noisy speech signal using the samples before and after the current sample.

$$g_f = S(x-\Delta x,t) - S(x,t)$$
$$g_b = S(x+\Delta x,t) - S(x,t) \quad (1)$$

Where, S(x,t) is the noisy speech signal, Δx is the sampling rate.

After the gradient is calculated the influencing coefficients in each directions of the current sample are computed.

$$IC_f = \frac{1}{1+\left(\frac{g_f}{k}\right)^2}$$
$$IC_b = \frac{1}{1+\left(\frac{g_b}{k}\right)^2} \quad (2)$$

Where, IC_f is the Forward influencing coefficient and IC_b is the backward influencing coefficient. In the equation (2) 'k' is constant value between 5 and 100.

From the above calculated influencing coefficients and gradients the speech signal is enhanced using

$$S(x,t+\Delta t) = S(x,t) + \Delta t(g_f IC_f + g_b IC_b) \quad (3)$$

In the above equations S(x, t) is the noisy speech signal, Δt is a coefficient between 0.1 to 0.3 representing the step of noise reduction in each iteration. The output of the signal is again processed by applying into the algorithm at the next iteration to gradually reduce the noise.

3. PROPOSED SPECTRAL SUBTRACTION METHOD

Spectral subtraction method is a well known noise reduction method based on the STSA estimation technique. In this proposed approach the output of PDE technique is applied as an

input for the spectral subtraction method. The basic power spectral subtraction technique, as proposed by Boll [1], is popular due to its simple underlying concept and its effectiveness in enhancing speech degraded by additive noise. The basic principle in this method is to subtract the magnitude spectrum of noise d(n) from the noisy speech y(n).

$$y(n) = s(n) + d(n) \quad (4)$$

Where s(n) is the clean speech. The noise is assumed to be uncorrelated and additive to the speech signal. The estimate of the noise is measured during silence or non-speech activity in the signal. The power spectrum of the noisy signal can be written as:

$$|Y(k)|^2 = |S(k)|^2 + |D(k)|^2 \quad (5)$$

Since the noise spectrum D(k) cannot be directly obtained, a time-average of the power spectrum $|\hat{D}(k)|$ is calculated during a period of silence, an estimate of the modified speech spectrum can be given as:

$$|\hat{S}(k)|^2 = |Y(k)|^2 - \alpha |\hat{D}(k)|^2 \quad (6)$$

The effectiveness of noise removal is dependent on accurate spectral estimate of noise signal. The better noise estimate gives lesser residual noise content in the resultant spectrum. The modified spectrum may contain some negative values due to the errors in the estimated noise spectrum. These values are rectified using half-wave rectification (set to zero) or full wave rectification (set to its absolute value). This can lead to further distortions in the resulting time signal. The accurate estimate of noise will overcome those drawbacks.

4. NOISE ESTIMATION

Noise estimation plays an important role in this work of speech enhancement. For an efficient noise estimation algorithm the resultant signal estimation will have great accuracy. Most of the noise estimation algorithms can be classified in to two classes. The first class is based on updating the noise estimate by tracking the silence regions of speech and other class is based on updating noise estimate using the histogram of the noisy speech power spectrum. The proposed algorithm comes under the first class.

4.1. EXISTING ALGORITHMS FOR NOISE ESTIMATION

Several noise-estimation algorithms have been proposed for speech enhancement applications [2] [3] [4] [5] [6] [10] [11] [13]. Here the basic two methods are taken before giving the proposed algorithm. Martin (2001) proposed a method for estimating the noise spectrum based on tracking the minimum of the noisy speech over a finite window. As the minimum is typically smaller than the mean, unbiased estimates of noise spectrum were computed by introducing a bias factor based on the statistics of the minimum estimates. The main drawback of this method [11] is that it takes slightly more than the duration of the minimum-search window to update the noise spectrum when the noise floor increases abruptly. Cohen proposed a minima controlled recursive algorithm (MCRA) which updates the noise estimate by tracking the noise-only regions of the noisy speech spectrum. These regions are found by comparing the ratio of the noisy speech to the local minimum against a threshold. The noise estimate, however, lags by at most twice that window length when the noise spectrum increases abruptly. In the improved MCRA approach [5], a different method was used to track the noise-only regions of the spectrum based on the estimated speech-presence probability. This probability, however, is also controlled by the minima, and therefore the algorithm incurs roughly the same delay as the MCRA algorithm for increasing noise levels.

In summary the main drawback of most of the noise estimation algorithms is that they are either slow in tracking sudden increases of noise power or that they are over estimating the noise energy resulting in speech distortion.

4.2. PROPOSED ADAPTIVE NOISE-ESTIMATION ALGORITHM

The smoothed power spectrum of noisy speech is computed using the following first-order recursive equation (9):

$$P(\alpha,k) = gP(\alpha-1,k) + (1-g)|Y(\alpha,k)|^2 \quad (7)$$

where P(α,k) is the smoothed power spectrum, α is the frame index, k is the frequency index, |Y(α,K)| 2 is the short-time power spectrum of noisy speech and g is a smoothing constant. The proposed algorithm is summarized in the following steps.

4.2.1 Classification of Speech Present and Speech Absent Frames:

In any speech sentence there are pauses between words which do not contain any speech; those frames will contain only background noise. The noise estimate can be updated by tracking those noise only frames [13].

To identify those frames , a simple procedure is used which calculates the ratio of noisy speech power spectrum to the noise power spectrum at 3 different frequency bands in each frame correspond to the frequency bins of 1 KHz, 3KHz and and the sampling frequency respectively. If all the three ratios are smaller than the threshold that frame is concluded as a noise only frame, otherwise , if any one or all the ratios are greater than threshold that frame is considered as speech present frame. The noise estimate is updated in speech absent frames with a constant smoothing factor. In speech present frames the noise is updated by tracking the local minimum of noisy speech and the deciding speech presence in each frequency bin separately using the ration of noisy speech power to its local minimum.

4.2.2. Minimum of Noisy Speech:

Various methods [10], [11] were proposed for tracking the minimum of the noisy speech power spectrum over a fixed search window length. These methods were sensitive to outliers and also the noise update was dependent on the length of the minimum-search window. A different non-linear rule is used in our method for tracking the minimum of the noisy speech by continuously averaging past spectral values [2]. In this algorithm if the value of the noisy speech spectrum in the present frequency bin is greater than the minimum value of previous frequency bin then the minimum value is updated, else the previous value is maintained as it is.

4.2.3. Detection of Speech-Presence Frames:

The approach taken to determine speech presence in each frequency bin is similar to the method used in [4]. Let the ratio of noisy speech power spectrum and its local minimum be defined as

$$S_r(\alpha,k) = P(\alpha,k) \Big/ P_{min}(\alpha,k) \quad (8)$$

This ratio is compared with a frequency dependent threshold, and if the ratio is found to be greater than the threshold, it is taken as a speech-present frequency bin else it is taken as a speech-absent frequency bin. This is based on the principle that the power spectrum of noisy speech will be nearly equal to its local minimum when speech is absent. Hence the smaller the ratio is in (6), the higher the probability that it will be a noise-only region and vice versa. Note that in [4], a fixed threshold was used in place of threshold.

From the above rule, the speech-presence probability, P(α,K), is updated. Using the following first-order recursion:

$$P(\alpha,k) = aP(\alpha -1,k) + (1- a\ ) I(\alpha,k) \quad (9)$$

where a is a smoothing constant. Note that the above recursion implicitly exploits the correlation for speech presence in adjacent frames. This may result in slight overestimate of the noise spectrum but will not likely have much effect on the enhanced speech.

4.2.4. Frequency-Dependent Smoothing Constants:

Using the above speech-presence probability estimate, we compute the time–frequency dependent smoothing factor as follows [4].

$$a(\alpha,K) = d\ + (\ 1- d)\ P(\alpha,k) \quad (10)$$

where d is a constant. Note that a(α,K) takes values in the range of $d \leq a(\alpha,K) \leq 1$.

4.2.5. Updating Noise Spectrum Estimate:

Finally, after computing the frequency-dependent smoothing factor a(α,k), the noise spectrum estimate is updated as

$$D(\alpha,k) = a(\alpha,k)\ D(\alpha -1,k) + (1- a(\alpha,k)\ )\ |Y(\alpha,K)|^2 \quad (11)$$

where D(α,k) is the estimate of the noise power spectrum [4]. Hence, the overall algorithm can be summarized as follows. After classifying the frequency bins into speech present/absent, we update the speech- presence probability and then use this probability to update the time– frequency dependent smoothing factor. Finally the noise spectrum estimate is updated using the time–frequency dependent smoothing factor.

This estimated noise is then subtracted from the input noisy speech signal to get an estimate of clean speech (Enhanced speech).

5. OBJECTIVE MEASURES FOR PERFORMANCE EVALUATION

Objective measures [15] are based on a mathematical comparison of the original and processed speech signals. The majority of objective quality measures quantify speech quality in terms of a numerical distance measure or a model of the perception of speech quality by the human auditory system. It is desired that the objective measures be consistent with the judgment of the human perception of speech [15]. However, it has been seen that the correlation between the results obtained by objective measures are not highly correlated with those obtained by subjective measures. The signal-to-noise ratio (SNR) is the most widely used objective measure.

5.1 SIGNAL-TO-NOISE RATIO (SNR)

The SNR is a popular method to measure speech quality. As the name suggests, it is a calculated as the ratio of the signal to noise power in decibels. If the summation is performed over the whole signal length, the operation is called *global SNR*. [15]

$$SNR_{dB} = 10\log_{10}\left(\frac{\sum_n S^2(n)}{\sum_n [S(n) - \hat{S}(n)]^2}\right) \quad (12)$$

5.2 MEAN OPINION SCORE MEASURE (MOS)

The mean opinion score (MOS) [15] provides a numerical measure of the quality of human speech. The scheme uses subjective tests (opinionated scores) that are mathematically averaged to obtain a quantitative indicator of the system performance. To determine MOS, a number of listeners rate the quality of test sentences read aloud over the communications circuit by male and female speakers. A listener gives each sentence a rating as follows:

(1) Bad (2) Poor (3) Fair (4) Good (5) Excellent.

The MOS is the arithmetic mean of all the individual scores, and can range from 1 (worst) to 5 (best).

6. EXPERIMENTAL RESULTS

Test samples are taken from SpEAR (Speech Enhancement Assessment Resource) database of CSLU (Center for Spoken Language Understanding).

Table 1: Comparison of SNR obtained in proposed method with basic methods and existing algorithms

Type of Noisy signal	SNR obtained in DEKF method	SNR for Spectral Subtraction method	SNR for PDE method	SNR for Proposed method
White Stationary 0dB Noisy Signal	7.6	9.54	9.82	10.15
Pink Stationary 0dB Noisy signal	5.5	3.83	10.32	13.56
Car Phone Noisy signal	--	11.42	16.29	18.89
Cellular Noisy signal	5.39	4.41	14.26	16.78
Colored cockpit noisy signal	--	10.24	23.47	25.63
Colored Factory Noisy signal	--	7.67	33.28	39.87

Factory phone noisy signal	--	16.25	30.54	32.58
White bursting 3dB noisy signal	10.07	9.48	5.65	12.49

Table 2: Mean Opinion Score obtained for proposed method

Type of Noisy signal	MOS
White stationary Noisy signal	4.2
Pink Stationary noisy signal	3.25
White bursting signal	3.2
Factory phone noise signal	4.1
Cockpit noise signal	3.7
Car noise signal	3.5
Cellular Noise signal	3.1

6.1 TIME DOMAIN RESULTS

Sample 1:

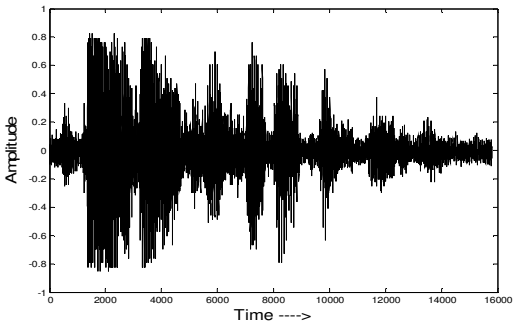

Fig.1 a: Cellular Clean Signal

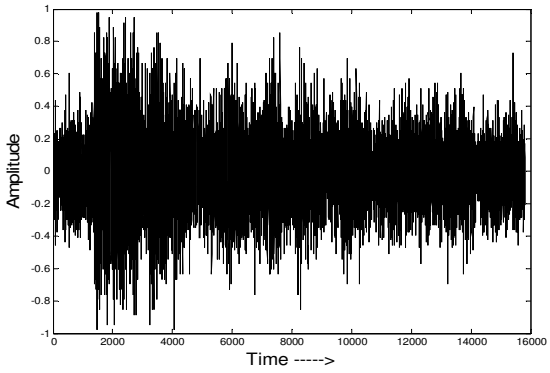

Fig.1b: Noisy Cellular Signal

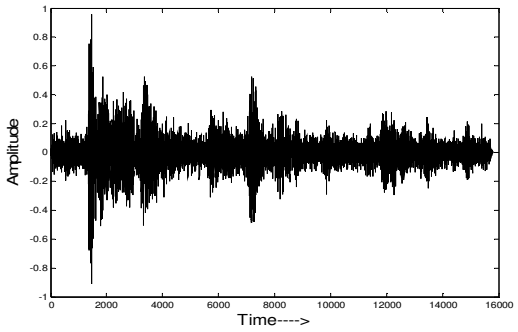

Fig.1c: Enhanced Cellular Signal

Sample 2:

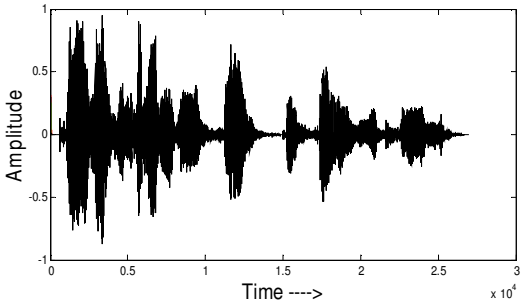

Fig.2.a. Colored F16 cockpit Clean speech

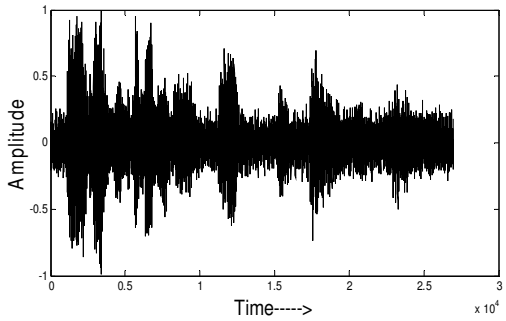

Fig.2.b. Colored F16 cockpit Noisy speech

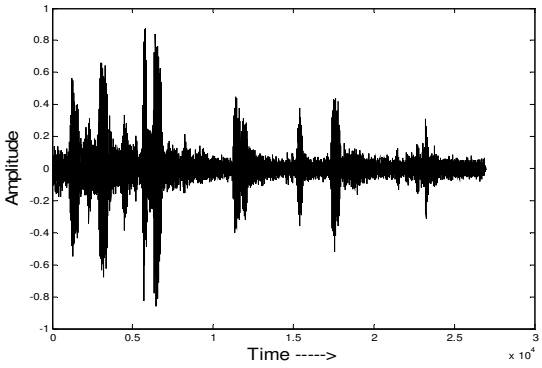

Fig.2.c. Enhanced Colored F16 cockpit speech

6.2 FREQUENCY DOMAIN ANALYSIS

Power Spectral Density of Reference Clean signal, Noisy Signal and the Enhanced Signals were obtained and compared.

Inference for Fig3: Power spectrum density of the enhanced signal is close to the power spectrum magnitude of clean signal.

Inference for Fig 4: Power spectrum magnitude of the enhanced signal is close to the power spectrum magnitude of clean signal in all the frequency range and the high frequency noise is removed.

Inference for Fig 5: Initial Noise segments are very much reduced in this method for colored factory noisy signal and the spectrum is close to the clean signal.

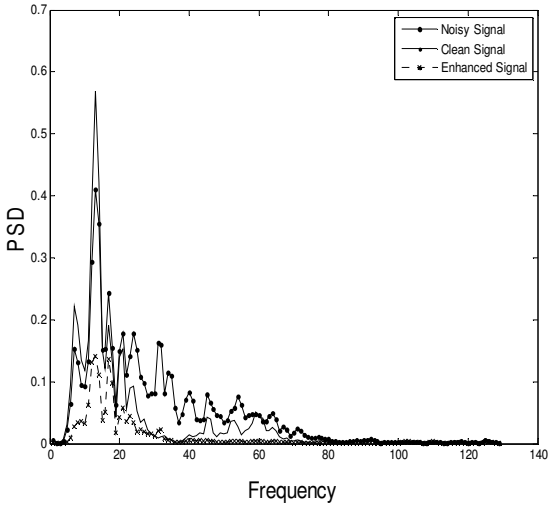

Fig.3. PSD plot of Cellular noisy, Clean and Enhanced signal

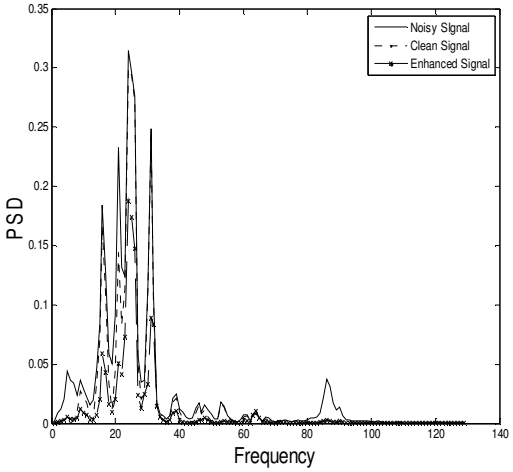

Fig.4. PSD plots of Colored F16 cockpit Noisy, Clean and Enhanced signal

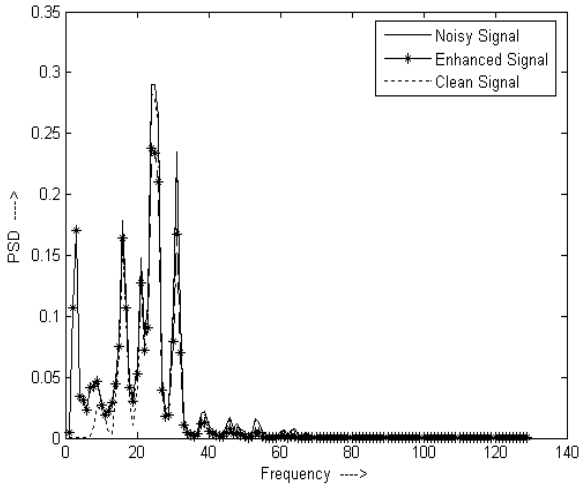

Fig.5. PSD plots of Colored factory Noisy, clean and Enhanced signal

7. CONCLUSION

The proposed spectral subtraction method with partial differential equation technique improves the speech quality by increasing the signal to noise ratio. This method provides a definite improvement over the conventional power spectral subtraction method. The added computational complexity of the algorithm is minimal and it adapts with non stationary noise environments. Further this method can be adapted with multi band spectral subtraction method to improve the performance. This method can be applied as a speech enhancement technique in Digital hearing aids, where sensorineural loss patients need 5dB to 10 dB higher SNR than normal hearing subjects.

REFERENCES

[1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, Signal Processing* vol.27, pp. 113-120, Apr. 1979.

[2] Doblinger, G., 1995. Computationally efficient speech enhancement by spectral minima tracking in subbands. Proceedings Eurospeech 2, 1513–1516.

[3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-term spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, Signal Processing* vol.ASSP-32, No.6, pp. 1109-1121, Dec.1984.

[4] Cohen, I., 2002. Noise estimation by minima controlled recursive averaging for robust speech enhancement. IEEE Signal Processing. Lett. 9 (1), 12–15.

[5] Cohen, I., 2003. Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. IEEE Transactions on Speech Audio Processing. 11 (5), 466–475.

[6] Hirsch, H., Ehrlicher, C., 1995. Noise estimation techniques for robust speech recognition. Proc. IEEE International Conference on Acoustics, Speech Signal Processing 153–156.

[7] Hu, Y., Loizou, P., 2004. Speech enhancement based on wavelet thresholding the multitaper spectrum. IEEE Transactions on Speech Audio Processing. 12 (1), 59–67.

[8] Lin, L., Holmes, W., Ambikairajah, E., 2003. Subband noise estimation for speech enhancement using a perceptual Wiener filter. Proceedings. IEEE International Conference on Acoustics, Speech Signal Processing I, 80–83.

[9] Malah, D., Cox, R., Accardi, A., 1999. Tracking speech-presence uncertainty to improve speech enhancement in non-stationary environments. Proc. IEEE International Conference on Acoustics, Speech Signal Processing 789–792.

[10] Martin, R., 1994. Spectral subtraction based on minimum statistics. Proceedings Euro. Signal Processing, 1182–1185.

[11] Martin, R., 2001. Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Transactions on Speech Audio Processing 9 (5), 504–512.

[12] Nilsson, M., Soli, S., Sullivan, J., 1994. Development of hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. Journal on Acoustics. Soc. Amer. 95(2), 1085–1099.

[13] Rangachari, S., Loizou, P., Hu, Y., 2004. A noise estimation algorithm with rapid adaptation for highly nonstationary environments. Proceedings IEEE International Conference on Acoustics, Speech Signal Processing. I, 305–308.

[14] Sohn, J., Kim, N., 1999. Statistical model-based voice activity detection. IEEE Signal Processing. Lett. 6(1), 1–3.

[15] S. Quackenbush, T. Barnwell, and M. Clements, "Objective Measures for Speech Quality Testing," Prentice-Hall, 1988.

[16] H. Hassanpour and E. Nadernejad," image enhancement using diffusion Equation", IEEE International conference on Signal processing and its applications, 2007.