Indian Journal of Science and Technology, *Vol 9(35), DOI: 10.17485/ijst/2016/v9i35/100366, September 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Mean-Median based Noise Estimation Method using Spectral Subtraction for Speech Enhancement Technique

Bittu Kumar*

Department of Electronics Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad - 826004, Jharkhand, India; kumarbittu135@gmail.com

Abstract

Background/Objectives: This article proposes a new noise estimation method which is based on mean and median statistical tools. In this article, the proposed method i.e., Mean-Median based noise estimation method have applied in the spectral subtraction method for speech enhancement technique. **Methods/Statistical Analysis:** The MATLAB/SIMULINK platform is used for simulating the simulation model of proposed technique. For performance evaluation of proposed noise estimation, Perceptual Evaluation of Speech Quality (PESQ) Score and Simulation time are chosen. Different noisy speech signal and clean speech signal have taken as input signal for proposed model and for finding PESQ score, respectively. **Findings:** The advantage of our proposed noise estimation method, it does not require signal to noise ratio, Voice Activity Detector, or histograms. This paper is carried comparison results with Modified Cascaded Median (MCM) based noise estimation method and also simulation time for different corrupted speech files. **Application/Improvements:** As per expected, proposed technique is given better speech quality signal (in PESQ scores). As compare to MCM based technique, it takes less simulation time and also no memory storage requirement.

Keywords: Mean-Median based Noise Estimation Method, Modified Cascaded Median based Noise Estimation, Spectral Subtraction, Speech Enhancement

1. Introduction

In the application of speech processing, several systems such as mobile communication, multiparty teleconferencing and audio controlled systems are involved. Due to effect of different background noises present in our living environment, these systems are required effective techniques for recovering the desired speech signals. Speech enhancement is the plausible techniques in the field of speech processing for enhancing the noisy speech signals. Noise estimation and speech estimation are two major parts of that technique which are taken to improve the perceptual aspects of degraded speech signals. To obtain only the speech data from degraded speech signal, the first requirement is to separate out the noise using noise estimation method then recover the speech data from noisy speech signal with the help of estimated noise through speech estimation method.

Various speech estimation methods such as Spectral Subtraction (SS)[1,2], Minimum Mean Square Error Short Time Spectral Amplitude (MMSE STSA)[3], Wiener filtering[4] etc., have been investigated in different research articles. In these methods, spectral subtraction has received a lot of attention from researchers toward the development of effective enhancement techniques. Particularly, in single channel speech enhancement techniques, spectral subtraction *(it is categories of frequency domain speech enhancement technique)* is frequently used for obtaining real speech data by subtracting the estimated noise from noisy speech signals.

Different noise estimation methods have been made for speech enhancement techniques but most of the time, noise estimation method have used '*Voice Activity Detector*

**Author for correspondence*

(VAD)' in case of single channel speech enhancement technique or '*reference signal'(it is recorded by additional microphone in noisy environment)* in case dual channel speech enhancement techniques. For tracking of the noise, VAD use to detect the speech absence period. But it does not detect silence period for low Signal to Noise Ratio (SNR) noisy speech signal and also long speech segments. In dual channel speech enhancement techniques, for obtaining original speech signal, one additional microphone is needed for separate recording of background noises. If recorded noisy speech signal and noise are independent to each other than noise can't be separate from noisy speech signal using additional microphone.

In recent few decades, statistical based noise estimation methods are frequently applied to estimate the noise from noisy speech signal. It does not need VAD, prior knowledge of noise type and SNR. For an effective speech enhancement technique, noise can be estimated by effective noise estimation method. Several statistical based noise estimation methods[5–9] are reported by different researcher for speech enhancement techniques. In minimum statistics algorithm[5], the minimum of magnitude spectra of past few segment of noisy speech spectrum have taken for finding the noise spectrum of present segment. Minimum statistics algorithm fails, when amplitude of noise is equal to the minimum value of desired speech signal *(original speech signal)*. In[6], authors have observed that the 80–90% of total speech segments are noise spectra *(low energy frames)* and rest of frames are voiced speech spectra *(high energy frames)* in particular frequency bines. On the basis of this observation, quantile-based noise estimation method is used certain quantile values histogram of previous frame of noisy speech signals for selecting the noise spectrum. Due to use of sorting operation, it requires more calculation for finding samples of noise estimation at individual quantile values. In[7], author used 0.5 quantile i.e., *median* in cascaded connection and proposed Cascaded Median (CM) based noise estimation method which carries computation complexity. But it requires memory for storing the medians of noisy spectrum. In Dynamic Quantile Tracking (DQT) based noise estimation[8], noise spectrum get without sorting operation and storage requirement, in[8] did not more attention regarding speech quality. In speech enhancement techniques, principally we should focus on speech quality with affecting the speech intelligibly. In[9], Modified Cascaded Median (MCM) based noise estimation is reported. It needs less memory storage as well as gives superior speech quality compared to CM-based noise estimation.

In this article proposed a noise estimation method which is constructed with Mean and Median statistical tool. The Mean-Median (MM) based noise estimation method does not require any addition microphone for use of references signal and also does not use voice activity detector. It gives better simulated result in terms of simulation time and PESQ score. Performance of mean median based noise estimation is better in speech quality with less simulation time and without memory storage requirement.

Current article is organized as follows. In Section 2, describe about spectral subtraction with block diagram of speech enhancement technique. The description of proposed noise estimation method is present in Section 3. The result part of this article shows in Section 4. In the end, Section 5 concludes the whole article.

2. Spectral Subtraction

Spectral subtraction method is easy used to implement in the application of speech enhancement techniques. It is used to recover the real speech data from recorded speech data (degraded speech signal) which is recorded through only single channel microphone in noisy environment. The block diagram of speech enhancement technique using spectral subtraction is shown in Figure 1.

In spectral subtraction, the original speech signal is obtained just by subtracting the obtained noise from observed speech signal in the frequency domain and noise is estimated from our proposed or existing noise estimation method. Assume an observed speech signal $d(n)$ which is mixed of speech signal $c(n)$ and background noise $u(n)$. Also assume that $c(n)$and $u(n)$ are un-correlated.

$$d(n) = c(n) + u(n) \tag{1}$$

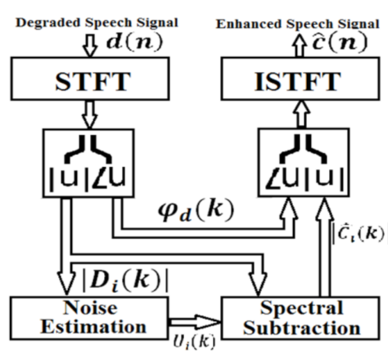

Figure 1. Block diagram of speech enhancement technique.

where ‘n’ is the discrete time index. $d(n)$ Passes through Short Time Fourier transform (STFT), after this process we find the noisy speech spectrum and also Equation (1) can be written as

$$D_i(k) = C_i(k) + U_i(k) \tag{2}$$

where $C_i(k), D_i(k)$ and $U_i(k)$ denote spectrum of speech signal, observed speech signal and unwanted noise, respectively. $D_i(k)$ can decompose into magnitude $\left(\left|D_i(k)\right|\right)$ and phase noisy spectrum $\varphi_d(k)$ as Equation (3) shows.

$$D_i(k) = \left|D_i(k)\right| e^{j\varphi_d(k)} \tag{3}$$

Noise spectrum $U_i(k)$ is obtained from the magnitude spectrum of observed speech signal by proposed noise estimation method. The obtained noise spectrum is updated during the periods on or after the frames, when the speech signal is available or absent in noisy (observed) spectrum.

For calculating the enhanced magnitude spectrum, $\left|\hat{C}_i(k)\right|$ according to the generalized expression (*which is showed in equation ‘4’*) of spectral subtraction, we subtract $U_i(k)$ from $\left|D_i(k)\right|$.

$$|\hat{C}_i(k)| = \begin{cases} b^{1/y} U_i(k), & if\ |D_i(k)| < (a+b)^{\frac{1}{y}} U_i(k) \\ [|D_i(k)|^y - a(U_i(k))^y]^{1/y}, & otherwise \end{cases} \tag{4}$$

where y, b and a express exponent factor, spectral floor and over subtraction factor, respectively. The angle of observed spectrum $\varphi_d(k)$ does not disturb the quality and intelligibility of the speech signal. To re-construct of enhanced signal $\hat{c}(n)$into time domain, we passed the enhanced spectrum $\hat{Y}_i(k)$ through Inverse Short Time Fourier Transform (ISTFT) method.

3. Proposed Noise Estimation and Compare with Existing Noise Estimation Method

Noise estimation which is name as Mean-Median (MM) based noise estimation, we are required only magnitude of segmented noisy speech spectrum for estimating the noise spectrum. It constructs of two basic statistical tools – mean and median, that are used for finding the threshold level of noise spectrum from noisy speech spectrum. The block diagram of MM-based noise estimation is shown in Figure 2. Basically, in this method first we separate the input spectrum (*magnitude of segmented noisy speech spectrum*) into sub frame, after that we have taken the square power of each sample value within sub frame. As compare to magnitude value, the squared value is more help for comparing the accurate differences between speech signal and noise. The median and maximum value of obtained squared value (*power of sample value*) is calculated within the sub frame. These median and maximum values are contributed for finding the threshold value by the mean of that median and maximum value (*average of median value and maximum value*). We compared the threshold value from noisy speech spectrum within sub frame, if the threshold value is greater than noisy spectrum then these noisy speech spectrum is consider as noise spectrum i.e., in noisy speech spectrum, speech is absent (*only noise is present in noisy spectrum*) within sub frame. If noisy spectrum is greater than threshold value then noise spectrum is threshold value. We can also conclude that the minimum value of threshold value and noisy speech spectrum within sub frame is considered as a noise spectrum for that sub frame. Here, we are using the magnitude spectral subtraction method for separating the speech signal from noisy speech signal. The obtained noise spectrum in power, it convert into magnitude form, we takes the square root of that obtained noise spectrum. The MM-based noise estimation method does not need memory storage requirement for finding noise spectrum. It implements with less complexity and less sorting operation.

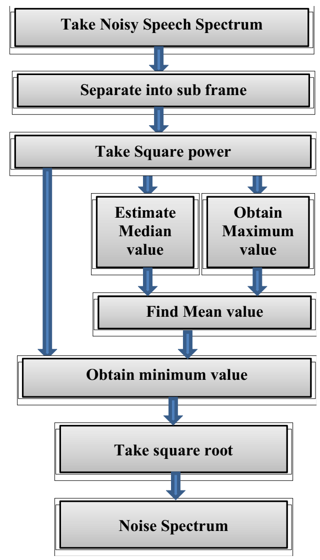

Figure 2. Block diagram of proposed noise estimation.

In[9], author shows that modified cascaded median based noise estimation method require $R \times f + (S-1) \times R$

samples memory storages (S = *total number of stages*, R = *number of frames in single stage*, f = *number of samples in a frame*), it reduces the storage $R \times (S - 1) \times (f - 1)$ samples as compared with original cascaded median based noise estimation method. Mean-Median based noise estimation method doesn't require memory storage; this is one of the best advantages of proposed noise estimation method.

4. Results

In this section it has evaluated the performance of Mean-Median based noise estimation method in simulation results which has obtained through MATLAB[10–13]/SIMULINK platform. Simulation results (*enhanced speech signals*) are obtained from Simulink model of speech enhancement[14–16] using proposed noise estimation (*Mean Median based noise estimation*) and also existing noise estimation method (*MCM-based noise estimation*). The obtained enhanced speech signals are used for finding the PESQ (Perceptual Evaluation of Speech Quality) Score of that enhanced speech file by help of original speech file. SpEAR (*Speech Enhancement Assessment Resource*) database[17] is taken for input speech signal (*noisy speech signal*) in our experiment. Before start our experiment, we have set the parameters values in Simulink model of each section according to Table 1.

4.1 Simulation Results

The performance of proposed noise estimation method have evaluated using simulation results in terms of PESQ Score[18] and also compared the results of MCM-based noise estimation method. For testing of mean median based noise estimation, we have chosen nine noisy speech file which have different in Signal to Noise Ratio (SNR). First we enhanced the noisy speech signal through simulation model of speech enhancement technique using proposed and existing noise estimation method, than find the PESQ Score of each enhanced speech signals by help of clean speech signal which are taken from SpEAR database. Figure 3 shows the plot of simulation results (*PESQ Score Vs SNR*) for noisy speech signal, existing noise estimation (*MCM-based noise estimation*) and proposed noise estimation method (*MM-based noise estimation*). It observed that from Figure 3, the PESQ Score lines for both proposed and existing noise estimation methods are almost overlapped to each other; the PESQ Score line of enhanced speech signal are far from PESQ Score line of noisy speech signal. We conclude that the speech quality of enhanced speech signals for MM-based noise estimation is better and as compared with speech quality of enhanced

Table 1. Name of parameters and its values

Name of Parameters	values
Sampling Frequency (Hz)	16000
Frame Size (*in samples*)	256
Sub-frame Size (*in samples*) (*for noise estimation*)	16
Overlap (%)	50
Exponent factor	1.0
Subtraction Factor	2.5
Spectral Floor Factor	0.002

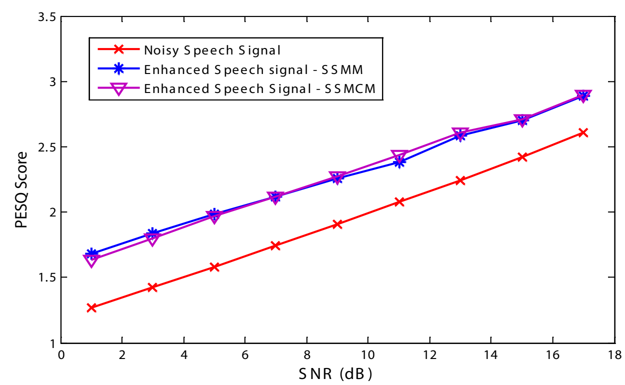

Figure 3. Comparison of Simulation results (PESQ Score Vs SNR); SSMM – Spectral Subtraction using Mean Median based noise estimation and SSMCM – Spectral Subtraction using Modified Cascaded Median based noise estimation.

Table 2. Average simulation time for MCM-based noise estimation and MM- based noise estimation

Sl No.	Name of speech file	No. of frame	Avg. Simulation time (MCM-based Noise Estimation)		Avg. Simulation time (MM- based Noise Estimation)	
			Total time (sec)	Time per frame(ms)	Total time (sec)	Time per frame(ms)
1	Vega_pinkr7_16.wav	1211	43.22	35.68	30.89	25.50
2	Bigtips_white3_16.wav	180	10.18	56.55	7.26	40.33
3	Butter_ white11_16. wav	157	9.26	58.9	6.49	41.33

speech signals for MCM-based noise estimation, both are almost same.

4.2 Simulation Time

Simulation time shows the time taken by running of simulation model. It is obtained through Simulink profiler that is enabled before run of model. We have taken three degraded speech signals which are different in SNR and noises. Table 2 shows the average simulation time for different speech file (*Vega_pinkr7_16.Wav, Bigtips_ white3_1 6.wav, Butter_w hite11_16.wav*) and different simulation model of speech enhancement techniques (*SSMCM- Spectral Subtraction using Modified Cascaded Median based noise estimation, SSMM- Spectral Subtraction using Mean Median based noise estimation*). In Table 2, the average simulation time (total time and also in time per frame) of MM-based noise estimation Simulink model is lesser, as compare to average simulation time of MCM-based noise estimation Simulink model for each different speech file.

4.3 Waveforms

Figure 4(a) (b) (c) (d) shows the waveform of different speech file- observed speech signal, speech signal, enhanced speech signal by SSMCM (*Spectral Subtraction using Modified Cascaded Median based noise estimation*) and enhanced speech signal by SSMM (*Spectral Subtraction using Mean Median based noise estimation*), respectively. From waveform, we can observed that the waveform of enhanced speech signal by proposed method

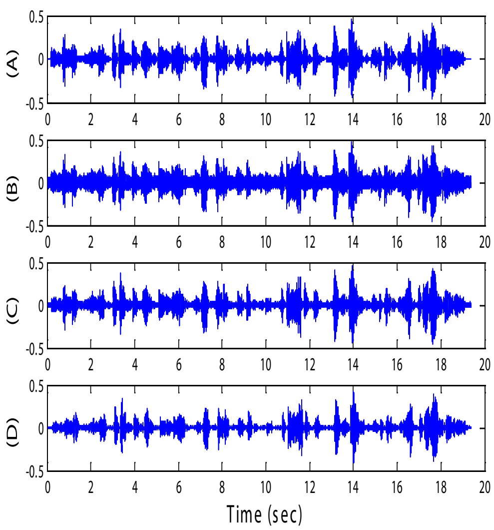

Figure 4. Waveform of. **(a)** Speech signal. **(b)** Noisy speech signal. **(c)** Enhanced speech signal by SSMCM. **(d)** Enhanced speech signal by SSMM.

is more close to the waveform of clean speech signal as compare to enhanced speech signal by SSMCM.

5. Conclusion

This paper has attained to new noise estimation made by mean and median statistical tools. The performance of mean median based noise estimation with different noisy speech files for SpEAR database is investigated. Meanwhile the result of proposed noise estimation is carried out in terms of average simulation time and PESQ Score. Also comparison results with modified cascaded median based noise estimation are presented. The mean median based noise estimation method gives superior outcome in terms of simulation time with same speech quality as compared to prior noise estimation method-Modified Cascaded Median based noise estimation and it doesn't need storage requirement.

6. References

1. Boll SF. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans Acoust, Speech, Signal Process. 1979; 27(2):113–20.
2. Loizou PC. Speech enhancement: Theory and practice. New York: CRC; 2007.
3. Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans Acoust, Speech, Signal Processing, 1984 Dec; ASSP-32:1109–21.
4. Hsu C-C, Lin T-E, Chen J-H, Chi T-S. Spectro-temporal subband wiener filter for speech enhancement. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Kyoto. 2012.
5. Martin R. Spectral subtraction based on minimum statistics. Proc Eur Signal Process Conf; 1994. p. 1182–5.
6. Stahl V, Fisher A, Bipus R. Quantile based noise estimation for spectral subtraction and Wiener filtering. Proc IEEE ICASSP; Istanbul, Turkey. 2000. p. 1875–8.
7. Waddi SK, Pandey PC, Tiwari N. Speech enhancement using spectral subtraction and cascaded-median based noise estimation for hearing impaired listeners. Proc 19th Nat Conf Commun (NCC 2013); Delhi, India. 2013.
8. Pandey PC, Tiwari N. Speech enhancement using noise estimation based on dynamic quantile tracking for hearing impaired listeners. IEEE Proc 21th National Conference on Communications (NCC 2015); Mumbai: IIT; 2015.
9. Kumar B. Spectral subtraction using modified cascaded median based noise estimation for speech enhancement. ACM, 6th International Conference on Computer and Communication (ICCCT-15); Allahabad, India: MNNIT. 2015 Sep 25–27.

10. Kim S, Jang B. Development of bellows design software using MATLAB. Indian Journal of Science and Technology. 2015 Apr; 8(S8). DOI: 10.17485/ijst/2015/v8iS8/70619.
11. Koochaki A. Teaching calculation of transformer equivalent circuit parameters using MATLAB/Simulink for undergraduate electric machinery courses. Indian Journal of Science and Technology. 2015 Aug; 8(17). DOI: 10.17485/ijst/2015/v8i17/59182.
12. Sivakumar P, Vinod B, Devi RSS, Rajkumar ERJ. Real-time task scheduling for distributed embedded system using MATLAB toolboxes. Indian Journal of Science and Technology. 2015 Jul; 8(15). DOI: 10.17485/ijst/2015/v8i15/55680.
13. Pradeep M, Kumar MS, Sathiskumar S, Raja SH. Interleave isolated boost converter as a front end converter for solar/fuel cell application to attain maximum voltage in MATLAB. Indian Journal of Science and Technology. 2016 Apr; 9(16). DOI: 10.17485/ijst/2016/v9i16/92234.
14. Ahmed SF, Memon AR, Azim ChF, Desa H. Global optimal solution for active noise control problem. Indian Journal of Science and Technology. 2011; 4(9):1015–20.
15. Ganesan, Vignesh, Manoharan S. Surround noise cancellation and speech enhancement using sub band filtering and spectral subtraction. Indian Journal of Science and Technology. 2015; 8(33).
16. Kumar KR, Bhavani L, Spandana P, Rishitha K, Ashitha G. Removal of real world noise in speech: comparision of various parameters using kalman and h-infinity filter algorithms. Indian Journal of Science and Technology. 2016; 9(30).
17. Speech Enhancement Assessment Resource (SpEAR) database. CSLU; Oregon, USA. Available from: www.cslu.ogi.edu/nsel/data/SpEAR_database.html
18. ITU. Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. ITU-T Rec; 2001. p. 862.