BACKGROUND AND RESIDUAL NOISE REDUCTION FOR SPEECH ENHANCEMENT: SPECTRAL SUBTRACTION AND MEDIAN FILTER APPROACH

[1]Disha Desai
[1]Post Graduate Student, Department of Electronics & Communication,
Sarvajanik College of Engineering & Technology, Surat, Gujarat, India
Email:[1]disha.ec20@gmail.com

Abstract - **Analysis of speech signal is used to illustrate parts of speech that comprise momentary the speech component and develop an algorithm to extract those components. In noisy environment in which channel is bit noisy, speaker and listener may experience several annoying noises which causes significant degradation in efficiency of communication system which makes communication bit difficult. It may happen that speaker has to speak much louder and listener has to give more concentration in listening process to maintain fruitful communication. Spectral subtraction method is normally used for background noise removal but there will be still some residual noise is present in this enhanced speech that needs to be suppressed for efficient communication system. Median filter can be used for second stage enhancement which removes residual noise from already enhanced speech in first stage.**

Keywords- Segmental SNR(SNR), Perceptual evaluation of speech quality(PESQ), Signal to Noise Ratio (SNR)

I. INTRODUCTION

Speech enhancement is one of the major parts of any communication block. Analysis of speech signal is used to illustrate parts of speech that comprise momentary the speech component and develop an algorithm to extract those components. Many speech enhancement techniques remove background noise very efficiently from noisy speech. But the musical effect of residual noise appears in the enhanced speech.

In simple speech enhancement, only background noise can be removed from noisy speech. Output of this speech enhancement is called as a pre-processed speech which contains some amount of noise after enhancement. So quality of speech is degraded and cannot get proper desired speech signal. This noise can be called as 'Musical Residual Noise'. This effect of musical noise is appearing due to some spectral patches of residual noise randomly appearing or disappearing in successive frames and over the neighboring sub bands. The majority of the speech enhancement system suffers from this musical residual noise which significantly degrades the perception quality of pre-processed speech. This type of residual noise is infuriating to human ear. If this residual noise is too prominent in nature, then it is more disturbing than the interference before speech enhancement. So for all these reasons it is required to remove residual noise for better speech quality.

II. SPEECH ENHANCEMENT

There are mainly two stages to reduce residual noise. First stage contains background noise reduction using speech enhancement algorithm. Output of this stage is called as pre-processed speech which contains some amount of residual noise after enhancement. And second stage is used to reduce the residual noise from pre-processed speech. Residual noise is reduced using post-processed algorithm for improved speech quality. Output of second stage contains enhanced speech quality with reduction of residual noise.

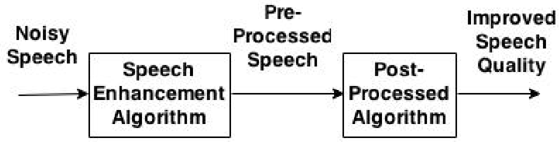


Figure 1 Process for Reduction of Residual Noise from Speech

A) First stage speech enhancement

First stage speech enhancement contains background noise reduction. Different approaches are used to remove background noise (e.g. Spectral subtraction, Wiener filtering, Wavelet denoising)[1]. Spectral subtraction is used to remove background noise from noisy speech.

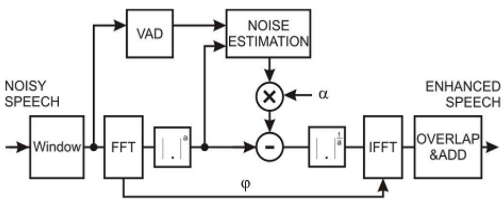


Figure 2 Spectral Subtraction Method [5]

Spectral subtraction is one of the noise reduction techniques that try to approximate the short-time spectrum of an additive noise which is used to corrupt a speech signal. In spectral subtraction, an approximation of the short-time spectrum of original speech is obtained by difference between estimated noise spectrum and spectrum of noisy speech. Estimation of original speech is obtained by combining estimated spectrum with the phase of noisy speech.

These processes can be used to improve the signal-to-noise ratio by attenuating the short-time spectrum when the signal-to-noise ratio is relatively low and there will be no attenuation in short time spectrum when SNR is high. Various spectral subtraction methods are decided by estimation method of short time spectrum of AWGN.

B) Second Stage Speech Enhancement

Enhanced speech after first stage may contain residual noise. Zero crossing detector is used to identify noisy portion and speech portion available in first stage enhanced speech. Number of zero crossing in a given signal tells about ZCR. More number of zero crossing shows that there are too quick changes in signal and it may contain high frequency information. On another hand, less number of zero crossing shows changes are gradual and it can be understood that signal may contain low frequency information. Frequency contents of a signal can be fetched out indirectly from parameter ZCR. Low ZCR value indicates voiced signal and high ZCR value indicates unvoiced signal.

For speech signal ZCR is given by,

$$z(n) = \frac{1}{2N}\sum_{m=0}^{N-1} s(m).\,w(n-m) \qquad (2.1)$$

Musical residual noise can be removed by a two-stage speech enhancement approach. Two-step noise reduction algorithm contains Two-Step-Decision-Directed approach. Basic goal of this method is to detect musical tone and attenuate the tone for the enhanced speech signal. When the spectral distance changes largely among the neighbors, the reference spectrum is notified as isolated spectrum which can be further adapted by the median filter. Conversely, the reference is declared as harmonic spectrum if the spectral distance changes easily among the neighbors. Now no more processing is done on this spectrum to preserve speech quality and kept it as it is. The central spectrum is terms as reference spectrum. Initially the optimum motion direction of the reference spectrum is detected. Neighboring directions are selected. The decision rule is used to find the minimum spectral distance between those directions. The spectral-distance measure is given by

$$d^{(i)}(m,\omega) = \sum_{\Delta m}\sum_{\Delta\omega} \frac{[|X(m+\Delta m,\ \omega+\Delta w| - |X(m,\omega)|]^2}{|X(m,\omega)|^2} \qquad (2.2)$$

The minimum spectral-distance measure is given in equation 2.2 and it is confirmed as the motion direction for the reference spectrum. Based on the value of spectral distance for the optimum direction, the decision is taken if the reference spectrum is speech spectrum or not. The optimum spectral distance measure is given as

$$d^{(i*)}(m,\omega) = \min\left\{d^{(i)}(m,\omega), 1 \le i \le 4\right\} \qquad (2.3)$$

If the reference spectrum is an isolated musical tone, the spectrum varies seriously over the neighbors of the reference spectrum. The value of the spectral-distance is large on the optimum direction. Hence, optimum spectral-distance is used to classify the reference spectrum is either speech or musical tone. The decision rule can be given by

$$d^{(i*)}(m, \omega) > \Theta(m, \omega) \text{ : residual noise}$$
$$d^{(i*)}(m, \omega) < \Theta(m, \omega) \text{ : speech} \quad (2.4)$$

If musical tone is detected in reference spectrum then it is replaced by the weighted median value of the window what is analyzed. The modified spectrum is given as

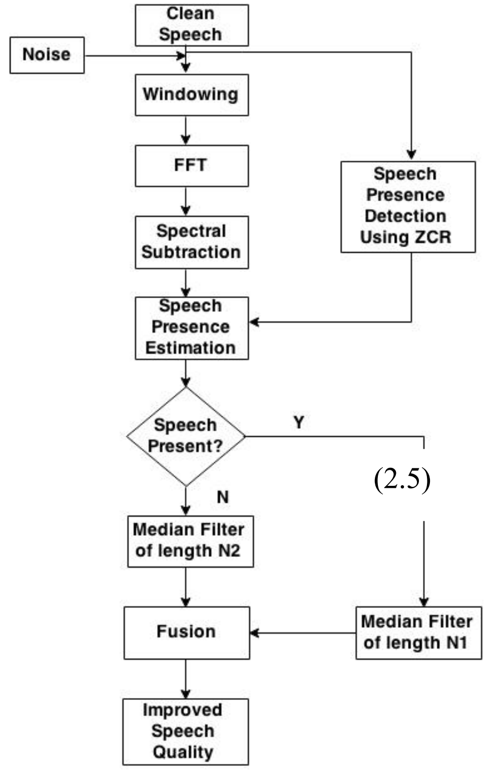
$$\hat{S}(m, \omega) = \gamma(m, \omega).\widetilde{S}(m, \omega) + [1 - \gamma(m, \omega)].M(m, \omega) \quad (2.5)$$

Where, Replacing flag and modified spectrum are denoted as $\gamma(m,\omega)$ and $M(m,\omega)$ respectively. Reference spectrum is modified if replacing flag shows zero value. Unity value of replacing flag is set when reference spectrum is classified as speech. Fluctuation of random spectral peaks can be diminished by weighted median filter is used that modifies the isolated reference spectrum. Best direction of those different directions decides the weight of reference spectrum. The weighted median filter is given as

$$M(m, \omega) = \text{median}\{\widetilde{S}(m + \Delta m, \omega + \Delta\omega)\} \quad (2.6)$$

Block and directional median (BDM) filter[3] that is post processing which reduces musical effect of residual noise presented in first stage enhanced speech. Directional median filtering is employed to slightly decrease musical effect of residual noise in speech like spectrum and on other hand vowel spectrum which have strong harmonic spectrum is well maintained. Then the quality of post-processed speech is assured. Spectral changes in noise dominant area are reduced by median filtering which smoothens spectral peaks of musical noise. Based on the speech presence probability, both preprocessed and post-processed spectra are integrated.

III. IMPLEMENTED WORK



(2.5)

Figure.3 Flow chart of Proposed Work

IV. SIMULATION RESULTS

Simulation is carried out on NOIZEUS speech database with airport, car and babble noise for different input SNR values. SNR, SSNR and PESQ are used for performance measurement.

(2.6)

Software used:	
Speech Database:	NOIZEUS (A Noisy Speech Corpus for Evaluation of Speech Enhancement Algorithm)

Table 1 SNR Results in dB for Female and Male Voice for AWGN Noise

Male/ Female	Noisy Speech	Spectral Subtraction	Proposed Method
Female	-5.0173	2.1476	9.7314
	-0.0724	7.0833	8.3779
	0.9971	8.0141	8.2852
Male	-5.0187	0.7415	10.2264
	0.0048	5.5907	8.2924
	2.0292	7.6293	8.1469

Table 2 SNR Results in dB for Male Voice with Different Noise

Type of Noise	Noisy Speech (dB)	Spectral Subtraction	Proposed Method
Airport	-0.6669	3.5642	8.3453
	4.3331	8.1861	8.6017
Babble	-0.6668	3.6147	8.6259
	4.3331	7.9708	8.6115
Car	-0.6668	3.4560	10.0459
	4.3332	8.2772	8.2313

Table 1 PESQ Results for Male Voice for Different Noise

Type of Noise	SNR of Noisy Speech	PESQ	
		Spectral Subtraction	Proposed Method
Airport	-0.6669	2.5553	4.3154
	4.3331	2.9150	4.1222
Babble	-0.6668	2.5645	4.1817
	4.3331	2.9718	4.0309
Car	-0.6668	2.4582	3.9549
	4.3332	2.8447	4.2076

Table 2 SSNR Results for Male Voice for Different Noise

Type of Noise	Noisy Speech	Spectral Subtraction	Proposed Method
Airport	-4.4125	-1.8151	0.1610
	-1.8234	1.7360	2.1605
Babble	-4.3520	-1.4235	0.4358
	-1.4040	2.4222	2.7495
Car	-4.4421	-2.3841	-1.7550
	-1.5324	1.7572	2.2691

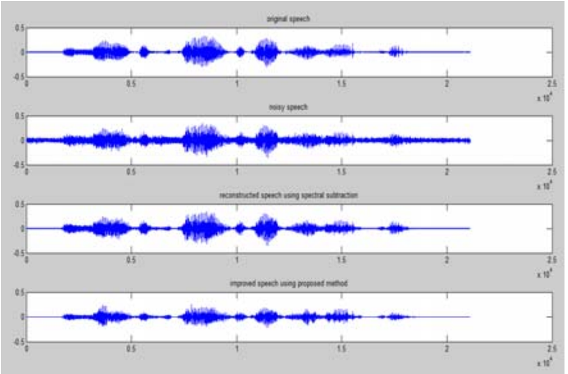


Figure. 4 For Male Voice (a) Clean Speech (b) Noisy Speech with Airport Noise at 5 dB (c) Reconstructed Speech using Spectral Subtraction Method (d) Improved Speech Quality using Proposed Method

V. CONCLUSION

Plenty of noises are always present during ongoing communication in practical world and those noises are main culprit for degradation of efficiency of communication system. Elimination or suppression of these noises is essential for fruitful communication. Major work is carried out over here is removal of background noise and residual noise which is still present after first stage of noise removal process. Spectral subtraction method is used for removal of background noise. Although output of this method contains some part of residual noise which is degrading the quality of speech. Removal of this residual noise is necessary. For that, first detection of speech part and noise part is done. Here ZCR approach is used to differentiate between speech part and noisy part. Median filter is used in last stage enhancement.

Experiments are carried out on free available NOIZEUS speech database with different noises (e.g. car, babble, airport, AWGN, F16, factory and machine). Performance of algorithm is measured using parameters like SNR, SSNR and PESQ. It has been noted that application of spectral subtraction method removes noise from noisy speech but not in significant way. Second stage speech enhancement (using median filter) improves speech quality which can be seen from comparison of SNR, SSNR and PESQ of first stage and second stage of enhancement.

It is observed that after second stage enhancement, output SNR increases around 4 to 5 dB from first stage enhanced SNR values for 0 dB (lower SNR) input SNR with airport, babble and car noises. This SNR increment is around 1dB when input SNR is around 5 dB.

REFERENCES

[1] Marwa A. Abd El-Fattah, Moawad I. Dessouky, Alaa M. Abbas, Salaheldin M. Diab, El Sayed M. El-Rabaie, Waleed Al-Nuaimy, Saleh A. Alshebeili, Fathi E. Abd El-samie, “Speech Enhancement with an Adaptive Wiener Filter,” Springer Journal on Int J Speech Technol, 17, PP 53-64, August 2013.

[2] Ching-Ta Lu, Kun-Fu Tseng, Chih-Tsung Chen, “Reduction of Residual Noise Using Directional Median Filter,” IEEE Conference on Computer Science and Automation Engineering, Vol. 3, PP 475-479, June 2011.

[3] Ching-Ta Lu, "Reduction of Musical Residual Noise using Block-and-Directional Median Filter Adapted by Harmonic Properties," Elsevier Journal on Speech Communication Vol. 58, PP 35-48, March 2014.

[4] Thomas Esch, Peter Vary, "Efficient Musical Noise Suppression for Speech Enhancement Systems", IEEE International Conference on Acoustics, Speech and Signal Processing, PP 4409-4412, April 2009.

[5] Ekaterina Verteletskaya and Boris Simak, "Noise Reduction Based on Modified Spectral Subtraction Method", in IJCS, February 2011.

[6] 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, January 2008.

[7] Thomas F. Quatieri, "Discrete-Time Speech Signal Processing: Principles and Practice", New Jersey: Prentice Hall, 2001.