International Journal of Emerging Engineering Research and Technology
Volume 2, Issue 2, May 2014, PP 96-100

# Speech Enhancement in Terms of Intelligibility Using Modified Multiband Spectral Subtraction

**Suraj P. Patil , Uday P Mithapelli**
Dept. of E & Tc, KJEI's Trinity College of Engineering & Research,
Pune, India

**Abstract:** *Speech signals from the uncontrolled environments contain required speech components along with the degradation components. The degraded speech gives poor performance in automatic speech processing tasks like speech and speaker recognition and is also uncomfortable for human listening. The degraded speech therefore needs to be processed for the enhancement of speech components. Hence it is much more important to obtain the noise free Speech signal. In this paper; we propose a hybrid approach that combines the concepts of multiband spectral subtraction with amplification & attenuation distortion effect on enhanced speech to remove the real time noise. The noisy signal is divided into four bands using band pass filters & each band is applied to standard spectral subtraction algorithms. The signals coming from these are then compared with original speech spectra and apply the amplification constraints on the signal. Then undistorted enhanced signal is reconstructed by overlap & adding the filtered IFFT of these. The method is applied to Speech signals from NOIZEOUS noisy speech corpus database.*

**Keywords:** *multiband spectral subtraction, speech enhancement, speech intelligibility, NOIZEOUS: noisy speech corpus database*.

## 1. Introduction

The problem of understanding the effect of noise on the spectrum of speech is very important in speech processing. Since, such knowledge could help us design better noise reduction algorithms that could potentially help speech processing applications to work well & also hearing-impaired listeners to understanding speech in noise[1]. For accurate working a clean (noise free) speech signal is always required. Speech signal can be corrupted by background noise at different SNR levels starting from 0dB to 10dB [2]. Some of the reasons why speech degradation in terms of intelligibility in spectral subtraction algorithms are [3]:

• do not have a good estimate of the background noise spectrum,

• Voice activity detector is unable to accurately track the spectrum of non-stationary noise.

• utilizes a gain function that not correlate with intelligibility of speech

• In MSE measures pay no attention to positive or negative differences between the estimated and clean spectra.

Steven F. Boll [4] proposed a stand-alone noise suppression algorithm is presented for reducing the spectral effects of acoustically added noise in speech. It suppresses stationary noise from speech by subtracting the spectral noise bias calculated during nonspeech activity. Secondary procedures are then applied to attenuate the residual noise left after subtraction. The algorithm was designed with assumption that PSD & priori SNR is known. The approach is successful in removing noise however some distortion in waveform is observed. Pascal Scalart et. al [5] proposed a enhancement technique for noisy speech signals based on posteriori as well as a priori SNR. Boh Lim Sim et. al. in [6] proposed a generalized Spectral Subtraction method based on Parametric Formulation which are predominantly non-statistical and employing statistical optimization using an MMSE criterion. From results it seems that the two parametric estimators improved noise suppression performances under stationary white Gaussian noise and semi-stationary Jeep noise. Y. Ephraim &.D. Malah [7] proposed derive a short-time spectral amplitude (STSA) estimator for speech signals which minimizes the mean square error of the log-spectra. Fourier expansion coefficients of the speech process, as well as of the noise process, are modelled as statistically independent Gaussian random variables. A priori SNR is calculated using Decision direct approach. This method is somewhat successful to achieve much lower residual noise level without affecting the speech. Israel Cohen in [8] proposed a Speech

Enhancement based on logMMSE spectral subtraction using a Non-causal a Priori SNR Estimator. In contrast to the decision directed estimator of Ephraim and Malah, the noncausal estimator is capable of discriminating between speech onsets and noise irregularities. Onsets of speech are better preserved, while a further reduction of musical noise is achieved. Here memory requirements & computations are somewhat increased but more noise suppression is achieved

## 2. Spectral Subtraction

Spectral subtraction is a method for restoration of the power spectrum or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum. The noise spectrum is usually estimated, and updated, from the periods when the signal is absent and only the noise is present. The assumption is that the noise is a stationary or a slowly varying process, and that the noise spectrum does not change significantly in between the update periods. For restoration of time-domain signals, an estimate of the instantaneous magnitude spectrum is combined with the phase of the noisy signal, and then transformed via an inverse discrete Fourier transform to the time domain. In terms of computational complexity, spectral subtraction is relatively inexpensive.

There are two types of Spectral subtraction: 1] Magnitude Spectral Subtraction; and 2] Power Spectral Subtraction. Generalize equation which describes spectral subtraction can be expressed as

$$|\hat{X}(k)|^{a} = |Y(k)|^{a} - E\left[|\hat{D}(k)|^{a}\right]$$

Where $|\hat{X}(k)|^{a}$ is estimated signal spectra which computed by subtracting $|\hat{D}(k)|$ i. e noise spectra from noisy signal spectra $|Y(k)|$ and a is power exponent which decides type of spectral subtraction

### 2.1 Multiband Spectral Subtraction

Almost all implementations and variations of the basic spectral subtraction techniques advocates subtraction of the noise spectrum estimate over the entire speech spectrum. But real world noise is mostly colored and does not affect the speech signal uniformly over the entire spectrum. The multi-band spectral subtraction approach which takes into account the fact that noise affects the speech spectrum differently at various frequencies [9].

Thus, here we divide spectrum of 0 to 4KHz into four equally spaced bands each of 1KHz bandwidth. After splitting the band we processed each band using traditional algorithms and processed output added together to form a enhanced signal. The concept behind splitting entire band in four bands is come in mind after seeing the SNR of each band separately; we observe that noise is more dominant in high frequency regions than low frequency. So instead of using one gain function over a entire spectra we split the spectra and apply different gain functions on each separately.

So, estimate of clean speech spectrum in the ith band is obtained by

$$|\hat{X}_i(k)|^{a} = |Y_i(k)|^{a} - E\left[|\hat{D}_i(k)|^{a}\right] \qquad b_i \le k \le e_i$$

Where bi and ei are the beginning and ending frequency bins of the ith band.

## 3. Amplification & Attenuation Distortions

According to Philipos C. Loizou & Gibak Kim [3] instead of time-domain SNR measure, the frequency domain version of the segmental SNR measure has been shown to correlate highly with both speech quality and speech intelligibility so as per there results we refer to this measure as the signal-to-residual spectrum measure SNRESI. The advantages in computing the SNRESI measure in the frequency domain include use of 1)critical-band frequency spacing for proper modelling of the frequency selectivity of normal-hearing listeners; 2)perceptually motivated weighting functions which can be applied to individual bands. Thus it is the combination of these two attractive features in the computation of the SNRESI measure that contributes to its high correlation with speech intelligibility. According to [3] SNRESI is given as $SNR_{ESI}(k) = \frac{X^2(k)}{(X(k) - \hat{X}(k))^2}$ Where X(k) denotes the clean magnitude spectrum and X(k) denotes the estimated magnitude spectrum. Dividing both numerator and denominator by D2(k), where D(k) denotes the noise magnitude spectrum, we get

$$SNR_{ESI}(k) = \frac{SNR_{ESI}(k)}{(\sqrt{SNR(k)} - \sqrt{SNR_{ESI}(k)})^2}$$

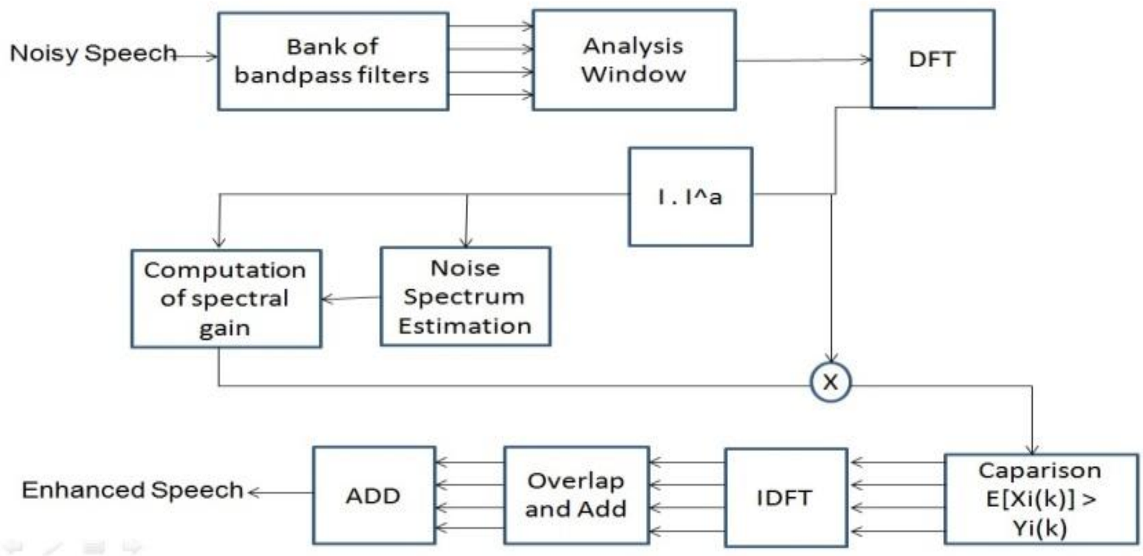
According to [3] important insights about the contributions of the two distortions to SNRESI, effect of attenuation distortion & amplification distortions in region {X(k)<X^ (k)≤2.X(k)} are negligible. While region {X^(k) ≥2.X(k)} amplification distortion perceptually it not acceptable to change the amplitude of estimated spectra & clean spectra more than 6.02dB.

## 4. Proposed approach

In our proposed approach by considering effect these distortions we just replace the components of estimated spectra in region{X^(k) ≥2.X(k)}; by original spectra. Figure1 Shows block diagram of the proposed method, consists of 6 stages. In the first stage we split the noisy speech spectra into different frequency bands by passing it from bank of bandpass filter. In the second stage, the signals are windowed and the magnitude spectrum is estimated using the DFT. In the third stage we apply the traditional speech enhancement techniques to remove the noise in each frequency band separately. In the fourth stage we compare the estimated signal spectra with noisy signal spectra for each frequency & if we found amplification distortion more than 6dB then we replace that frequency bin with original bin. In fifth stage time signal for each band is obtained by using the noisy phase information and taking the IFFT. Lastly, the modified frequency bands are recombined. Here we are using five different speech enhancement techniques to verify the effect of multiband spectral subtraction on estimated speech.

The sentence from NOIZEOUS: noisy speech corpus database [10], “Here the measure pick among the rider to break away the American tradition of real fem” uttered by female speaker is used to evaluate the proposed multi-band spectral subtraction approach. We are using a store sentence at 0dB SNR with .wav extension. Figure 2 shows time domain representation of clean, noisy & enhanced speech. The simulation is performed in MATLAB 7.8.0.

*Figure 1: Proposed modified spectral subtraction method*



## 5. Objective Measures for Performance Evaluation

### 5.1 Segmental SNR

SegmentalSNR is calculated by performing the summation over smaller frames of the speech signal. This method has a higher correlation to subjective results as compared to the global SNR method [11].

$$SNR_{seg} = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\cdot\sum_{i=Nm}^{Nm+N-1}\left(\frac{\sum_{i=0}^{N} X^2(k)}{\sum_{i=0}^{N}(X(k)-\hat{X}(k))^2}\right)$$

### 5.2 Log Likelihood Ratio

LLR assess the difference between the spectral envelopes, as computed by the LPC model, of the noise-free and processed signals[11] & defined as:

$$d_{LLR}(\vec{a}_p,\vec{a}_c) = \log\left(\frac{\vec{a}_p R_c \vec{a}_p^T}{\vec{a}_c R_c \vec{a}_c^T}\right)$$

Where $\vec{a}_c$ & $\vec{a}_p$ are the LPC vector of the clean & processed enhanced speech signal, and Rc is autocorrelation matrix of the clean speech signal.

### 5.3 Weighted Spectral Slope

The WSS distance measure computes the weighted difference between the spectral slopes in each frequency band. The spectral slope is obtained as the difference between adjacent spectral magnitudes in decibels [11].

### 5.4 Perceptual Evaluation of Speech Quality (PESQ)

The PESQ produces a score between 1.0 and 4.5, with high values indicating better quality. The original clean and degraded signals are first level equalized to a standard listening level and filtered by a filter with response similar to that of a standard telephone handset. The signals are time aligned to correct for time delays, and then processed through an auditory transform to obtain the loudness spectra. The difference in loudness between the original and degraded signals is computed and averaged over time and frequency to produce the prediction of subjective quality rating [12].

## 6. Results

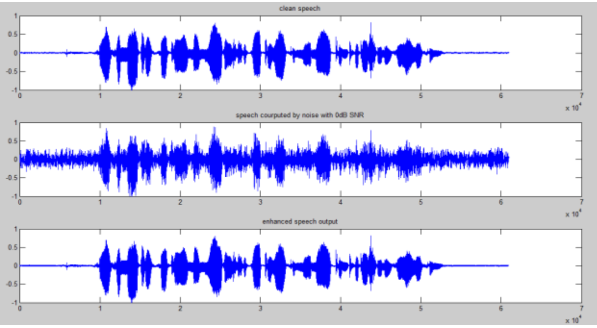
**Table 1.** *Results showing improvement in Enhanced signal by proposed method than that of standard spectral subtraction methods*

| Algorithm / Parameter | Spectral Subtraction boll | logMMSE (DDA) | logMMSE (non-causal) | Spectral Subtraction (priori SNR) | Parametric subtraction |
|---|---|---|---|---|---|
| LLR (Standard) | 1.736074 | 0.703127 | 0.655166 | 0.549810 | 0.741035 |
| (Multiband) | 0.982447 | 0.761247 | 1.573940 | 2.112646 | 0.923098 |
| SNRseg (Standard) | -4.043803 | 2.103497 | 2.886964 | 0.155983 | 2.213682 |
| (Multiband) | -6.519383 | -3.732374 | -2.994761 | -1.721276 | -2.786061 |
| WSS (Standard) | 42.813775 | 45.515834 | 46.026846 | 47.849768 | 46.589966 |
| (Multiband) | 50.723854 | 41.133676 | 32.158672 | 39.666836 | 40.905510 |
| PESQ (Standard) | 2.358964 | 2.645212 | 2.678357 | 2.661533 | 2.714321 |
| (Multiband) | 2.707843 | 2.701989 | 2.418640 | 2.317895 | 2.534113 |

*Figure 2: Denoising of Speech Signal*

## 7. Conclusion

In this paper, a hybrid approach based on the




concepts of multiband spectral subtraction with amplification & attenuation distortion effect on enhanced speech to remove the real time noise is presented. The results confirm the improvements in enhanced speech terms of intelligibility & naturalness with less number of computations in the proposed method.

## References

[1] Fukane, A.R. ; Sahare, S.L., “Enhancement of Noisy Speech Signals for Hearing Aids”, Communication Systems and Network Technologies (CSNT), International Conference on Digital Object Identifier.2011.

[2] Veisi, H. ; Sameti, H. , “Hidden-Markov-model-based voice activity detector with high speech detection rate for speech enhancement” Signal Processing, IET 2012.,Volume: 6 Page(s): 54 - 63

[3] Philipos C. Loizou and Gibak Kim, “Reasons why Current Speech-Enhancement Algorithms do not Improve Speech Intelligibility and Suggested Solutions”, IEEE Trans. on Speech and Audio Proc, Vol. 19, No. 1, January 2011.

[4] Steven boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction”, IEEE Transactions On Acoustics, Speech, And Signal Processing, Vol. Assp-27, No. 2, April 1979.

[5] Pascal Scalart, Jozue Vieira Filho, “Speech Enhancement Based On A Priori Signal To Noise Estimation”, IEEE Transactions On Acoustics, Speech, And Signal Processing,1996.

[6] Boh Lim Sim, Yit Chow Tong, Joseph S. Chang, and Chin Tuan Tan, “A Parametric Formulation of the Generalized Spectral Subtraction Method, IEEE Transactions On Speech And Audio Processing, Vol. 6, No. 4, July 1998

[7] Y. Ephraim & D. Malah, “Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator”, IEEE Transactions On Speech And Audio Processing, Vol.Assp-33,No-2,April 1985.

[8] Israel Cohen, “Speech Enhancement Using a Noncausal A Priori SNR Estimator”, IEEE Signal Processing Letters, Vol. 11, No. 9, September 2004.

[9] S. Kamath and P. Loizou,,“A multi-band spectral subtraction method for enhancing speech corrupted by colored noise,” Proc. ICASSP, Orlando, FL, 2002.

[10] NOIZEOUS: noisy speech corpus: www.utdallas.edu/~loizou/speech/noizeus/

[11] Yi Hu and Philipos C. Loizou, “Evaluation of Objective Quality Measures for Speech Enhancement”, IEEE Trans. Speech Audio Procs., vol. 16, no. 1, January 2008.

[12] Al-Akhras, M. ; Daqrouq, K. ; Al-Qawasmi, A.R., “Perceptual evaluation of speech enhancement Systems”, Signals and Devices (SSD), 7th International Multi-Conference on Digital Object Identifier 2010 , Page(s): 1 – 6.

## Author’s Biography

**Suraj Patil** received Bachelors of Engineering degree in 2010, from the University of Pune, Department of Electronics & Telecommunication Engineering. He received his MTech degree in 2013, from Vishwakarma Institute of Technology, Pune India.

Currently he is working as Assistant Professor in KJEI’s Trinity College of Engineering & Research, Pune, India. His main interests lie in speech synthesis, coding, and recognition.

**Uday P Mithapelli** received Bachelors of Engineering degree in 2010, from the University of Pune, Department of Electronics & Telecommunication Engineering. He received his MTech degree in 2013, from Vishwakarma Institute of Technology, Pune India.

Currently he is working as Assistant Professor in KJEI’s Trinity College of Engineering & Research, Pune, India. His main interests lie in speech synthesis, coding, and recognition