

Different Approaches of Spectral Subtraction method for Enhancing the Speech Signal in Noisy Environments

Anuradha R. Fukane, Shashikant L. Sahare

Abstract—Enhancement of speech signal degraded by additive background noise has received more attention over the past decade, due to wide range of applications and limitations of the available methods. Main objective of speech enhancement is to improve the perceptual aspects of speech such as overall quality, intelligibility and degree of listener fatigue. Among the all available methods the spectral subtraction algorithm is the historically one of the first algorithm, proposed for background noise reduction. The greatest asset of Spectral Subtraction Algorithm lies in its simplicity. The simple subtraction process comes at a price. More papers have been written describing variations of this algorithm that minimizes the shortcomings of the basic method than other algorithms. In this paper we present the review of basic spectral subtraction Algorithm, a short coming of basic spectral subtraction Algorithm, different modified approaches of Spectral Subtraction Algorithms such as Spectral Subtraction with over subtraction factor, Non linear Spectral Subtraction, Multiband Spectral Subtraction, Minimum mean square Error Spectral Subtraction, Selective Spectral Subtraction, Spectral Subtraction based on perceptual properties that minimizes the shortcomings of the basic method, then performance evaluation of various modified spectral subtraction Algorithms, and conclusion.

Index Terms— speech enhancement; additive noise; Spectral Subtraction; intelligibility; Discrete Fourier Transform, vad.

1 INTRODUCTION

Speech signals from the uncontrolled environment may contain degradation components along with required speech components. The degradation components include background noise, speech from other speakers etc. Speech signal degraded by additive noise, this make the listening task difficult for a direct listener, gives poor performance in automatic speech processing tasks like speech recognition speaker identification, hearing aids, speech coders etc. The degraded speech therefore needs to be processed for the enhancement of speech components. The aim of speech enhancement is to improve the quality and intelligibility of degraded speech signal. Main objective of speech enhancement is to improve the perceptual aspects of speech such as overall quality, intelligibility and degree of listener fatigue. Improving quality and intelligibility of speech signals reduces listener's fatigue; improve the performance of hearing aids, cockpit communication, videoconferencing, speech coders and many other speech systems. Quality can be measured in terms of signal distortion but intelligibility and pleasantness are difficult to measure by any mathematical algorithm. Perceptual quality and intelligibility are two measures of speech signals and which are not co-related. In this study a speech signal enhancement using basic spec

tral subtraction and modified versions of spectral subtraction methods such as Spectral Subtraction with over subtraction, Non linear Spectral Subtraction, Multiband Spectral Subtraction, MMSE Spectral Subtraction, Selective Spectral Subtraction, Spectral Subtraction based on perceptual properties has been explained in detail with their performance evaluation.

2 METHODOLOGIES

2.1 Basic spectral subtraction algorithm

The speech enhancement algorithms based on theory from signal processing. The spectral - subtractive algorithm is historically one of the first algorithms proposed for noise reduction [4]. Simple and easy to implement it is based on the principle that one can estimate and update the noise spectrum when speech signal is not present and subtract it from the noisy speech signal to obtain clean speech signal spectrum[7]. Assumption is noise is additive and its spectrum does not change with time, means noise is stationary or it's slowly time varying signal. Whose spectrum does not change significantly between the updating periods. Let $y(n)$ be the noise corrupted input speech signal, is composed of the clean speech signal $x(n)$ and the additive noise signal $d(n)$. In mathematical equation form one can

$$y(n) = x(n) + d(n) \quad (1)$$

- Anuradha R. Fukane is currently pursuing master's degree program in signal processing in Electronics and Telecommunication Engg. branch in Cummins College of Engg. For Women Pune, in Pune University, Maharashtra, India. E-mail: anuraj110@rediffmail.com
- Shashikant L. Sahare is currently working as Asst. Professor in Cummins College of Engg. For Women in Pune University, Maharashtra, India. E-mail: shashikantsahare@rediffmail.com

Many of speech enhancement algorithms operates in the Discrete Fourier Transform (DFT) domain [3] assume that the real and imaginary part of the clean speech DFT coefficients can be modeled by different speech enhancement algorithms. In Fourier domain, we can write $y(n)$ as

$$Y[w] = x[w] + D[w]. \quad (2)$$

$Y[w]$ can be expressed in terms of Magnitude and Phase as

$$Y[w] = |Y(w)| e^{j\phi_y}$$

Where $|Y(w)|$ is the magnitude spectrum and ϕ_y is the phase spectra of the corrupted noisy speech signal. Noise spectrum in terms of magnitude and phase spectra is

$$D[w] = |D[w]| e^{j\phi_y}$$

The Magnitude of noise spectrum $|D(w)|$ is unknown but can be replaced by its average value or estimated noise $|D_e(w)|$ computed during non speech activity that is during speech pauses. The noise phase is replaced by the noisy speech phase ϕ_y that does not affect speech intelligibility [4]. We can estimate the clean speech signal simply by subtracting noise spectrum from noisy speech spectrum in equation form

$$X_e(w) = [|Y(w)| - |D_e(w)|] e^{j\phi_y} \quad (3)$$

Where $X_e(w)$ is estimated clean speech signal. Many spectral subtractive algorithms are there depending on the parameters to be subtracted such as Magnitude spectral subtraction, Power spectral subtraction, Autocorrelation subtraction. The estimation of clean speech Magnitude signal spectrum is

$$X_e[w] = |Y[w]| - |D_e[w]|$$

Similarly for Power spectrum subtraction is

$$X_e[w]^2 = |Y[w]|^2 - |D_e[w]|^2 \quad (4)$$

The enhanced speech signal is finally obtained by computing the inverse Fourier Transform of the estimated clean speech $|X_e[w]|$ for magnitude. Spectrum subtractions and $|X_e[w]|^2$ for power spectrum subtraction, using the phase of the noisy speech signal. The more general version of the spectral subtraction algorithms is

$$X_e[\omega]^p = |Y[\omega]|^p - |D_e[\omega]|^p \quad (5)$$

Where P is the power exponent, the general form of the spectral subtraction, when $p=1$ yielding the magnitude spectral subtraction algorithm and $p=2$ yielding the power spectral subtraction algorithm. The general form of the spectral subtraction algorithm is shown in figure 1. [4]

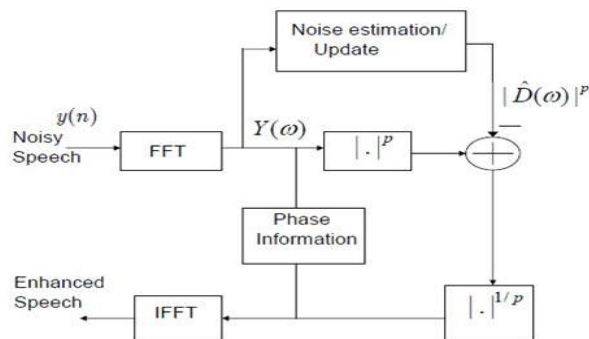


Figure1-The general form of the spectral subtraction algorithm [4]

2.2 Shortcomings of S. S. Algorithm

The subtraction process needs to be done carefully to avoid any speech distortion. If too little is subtracted then much of the interfering noise remains, if too much is subtracted then some speech information might be removed [1]. It is clear that spectral subtraction method can lead to negative values, resulting from differences among the estimated noise and actual noise frame. Simple solution is set the negative values to zero, to ensure a non-negative magnitude spectrum. This non-linear processing of the negative values called negative rectification or half-wave rectification [4]. This ensures a non-negative magnitude spectrum given by equation (6)

$$|X_e(\omega)| = |Y(\omega)| - |D_e(\omega)|, \text{ if } |Y(\omega)| > |D_e(\omega)| \\ = 0 \text{ else} \quad (6)$$

This non-linear processing of the negative values creates small, isolated peaks in the spectrum occurring at random frequency locations in each frame. Converted in the time-domain, these peaks sound like tones with frequencies that change randomly from frame to frame. That is, tones that are turned on and off at the analysis frame rate (every 20 to 30 ms). This new type of noise introduced by the half-wave rectification process has been described as warbling and of tonal quality, and is commonly referred to in the literature as "musical noise." Minor shortcoming of the spectral subtraction Algorithm is the use of noisy phase that produces a roughness in the quality of the synthesized speech [4]. Estimating the phase of the clean speech is a difficult task and greatly increases the complexity of the enhancement algorithm. The phases of the noise corrupted signal are not enhanced, because the presence of noise in the phase information does not contribute much to the degradation of speech quality [6]. The distortion due to noisy phase information is not very significant compared to that of the Magnitude spectrum especially for high SNRs. Combating musical noise is much more critical than finding methods to preserve the original phase. Due to that reason, much effort has been fo-

cused on finding methods to reduce musical noise which are explained in next section

2.3 Spectral Subtraction with over subtraction

Modifications made to the original spectral subtraction method are subtracting an over estimate of the noise power spectrum and preventing the resultant spectrum from going below a preset minimum level (spectral floor). This modifications lead to minimizing the perception of the narrow spectral peaks by decreasing the spectral excursions and thus lower the musical noise effect. Berouti [5] has taken a different approach that does not require access to future information. This Method consists of subtracting an overestimate of the noise power spectrum and presenting the resultant spectral components from going below a preset minimum spectral floor value. This algorithm is given in equation (7), where $|X_{ej}(\omega)|$ denotes the enhanced spectrum estimated in frame i and $|D_e(\omega)|$ is the spectrum of the noise obtained during non-speech activity

$$\begin{aligned} |X_{ej}(\omega)|^2 &= |Y_j(\omega)|^2 - |D_e(\omega)|^2 \\ &\quad \text{if } |Y_j(\omega)|^2 > (\alpha + \beta) |D_e(\omega)|^2 \\ &= \beta |D_e(\omega)|^2 \quad \text{else} \end{aligned} \quad (7)$$

With $\alpha \geq 1$ and $0 < \beta \leq 1$.

Where α is over subtraction factor and β is the spectral floor parameter. Parameter β controls the amount of residual noise and the amount of perceived Musical noise. If β is too small, the musical noise will became audible but the residual noise will be reduced. If β is too large, then the residual noise will be audible but the musical issues related to spectral subtraction reduces. Parameter α affects the amount of speech spectral distortion. If α is too large then resulting signal will be severely distorted and intelligibility may suffer. If α is too small noise remains in enhanced speech signal. When $\alpha > 1$, the subtraction can remove all of the broadband noise by eliminating most of wide peaks. But the deep valleys surrounding the peaks still remain in the spectrum [1]. The valleys between peaks are no longer deep when $\beta > 0$ compared to when $\beta = 0$ [4] Berouti found that speech processed by equation (7) had less musical noise. Experimental results showed that for best noise reduction with the least amount of musical noise, α should be smaller for high SNR frames and large for low SNR frames. The parameter α varies from frame to frame according to Burouti [5] as given below

$$\alpha = \alpha_0 - 3/20 \text{ SNR} \quad -5 \text{ dB} < \text{SNR} \leq 20 \text{ dB}$$

Where α_0 is the desired value of α at 0 dB SNR is the short time SNR estimate in each frame. It is an a posteriori estimate of the SNR computed based on the ratio of the noisy speech power to the estimated noise power. Berouti [5] determine the optimum values of α and β . For high

noise levels (SNR = -5dB), the suggested β is in the range of 0.02 to 0.06 and for lower noise levels (SNR > 0dB), β in the range 0.005 to 0.02. The parameter α suggested by Berouti [5] is in the range of 3 to 6. The influence of α also investigated by others Martin[4,15] suggest the range of α should lie between 1.3 and 2 for Low SNR conditions for high SNR conditions subtraction factor α less than one was suggested.

2.4 Non-linear Spectral Subtraction (NSS)

The NSS proposed by [8] Lockwood and Boudy. NSS is basically a modification of the method suggested in [5] by making the over subtraction factor frequency dependent and the subtraction process non-linear. In case of NSS assumption is that noise does not affects all spectral components equally. Certain types of noise may affect the low frequency region of the spectrum more than high frequency region. This suggests the use of a frequency dependent subtraction factor for different types of noise. Due to frequency dependent subtraction factor, subtraction process becomes nonlinear. Larger values are subtracted at frequencies with low SNR levels and smaller values are subtracted at frequencies with high SNR levels. The subtraction rule used in the NSS algorithm has the following form.

$$\begin{aligned} |X_e(\omega)| &= |Y(\omega)| - \alpha(\omega) N(\omega) \quad \text{if} \\ &\quad |Y(\omega)| > \alpha(\omega) N(\omega) + \beta |D_e(\omega)| \quad \text{else} \\ &= \beta |Y(\omega)| \end{aligned} \quad (8)$$

Where β is the spectral floor set to 0.1 in [8] $|Y(\omega)|$ and $|D_e(\omega)|$ are the smoothed estimates of noisy speech and noise respectively, $\alpha(\omega)$ is a frequency dependent subtraction factor and $N(\omega)$ is a non-linear function of the noise spectrum where

$$N(\omega) = \text{Max}(|D_e(\omega)|) \quad (9)$$

The $N(\omega)$ term is obtained by computing the maximum of the noise magnitude spectra $|D_e(\omega)|$ over the part 40 frames [4]. The $\alpha(\omega)$ given in [8] as

$$\alpha(\omega) = 1/r + p(\omega) \quad (10)$$

Where r is a scaling factor and $P(\omega)$ is the square root of the posteriori SNR estimate given as

$$P(\omega) = |Y(\omega)| / |D_e(\omega)| \quad (11)$$

The NSS algorithm was successfully used in [8] as a pre-processor to enhance the performance of speech recognition systems in noise.

2.5 Multiband Spectral Subtraction (MBSS)

In MBSS approach [9,4] the speech spectrum is divided into N overlapping bands and spectral subtraction is per-

formed independently in each band. The processes of splitting the speech signal into different bands can be performed either in the time domain by using band pass filters or in the frequency domain by using appropriate windows. The estimate of the clean speech spectrum in the i th band is obtained by [9].

$$|X_{ei}(\omega_k)|^2 = |Y_i(\omega_k)|^2 - \alpha_i \delta_i |D_i(\omega_k)|^2 \quad (12)$$

$$b_i < \omega_k < e_i$$

Where $\omega_k = 2\pi k / N$, $k = 0, 1 \dots N - 1$ are the discrete frequencies $|D_{ei}(\omega_k)|^2$ is the estimated noise power spectrum obtained during speech absent segment, α_i is the over subtraction factor of the i th band and δ_i is an additional band. Subtraction factor can be individually set for each frequency band to customize the noise removal processor b_i and e_i are the beginning and ending frequency bins of the i th frequency band. The band specific over subtraction factor is a function of the segmented SNR _{i} of the i th frequency band and is computed as follows [4]

$$\alpha_i = \begin{cases} 4.75 & \text{SNR}_i < -5 \\ 3/20 (\text{SNR}_i) & -5 < \text{SNR}_i < 20 \\ 1 & \text{SNR}_i > 20 \end{cases}$$

The values for δ_i are set to

$$\delta_i = \begin{cases} 1 & f_i < 1 \text{ KHz} \\ 2.5 & 1 \text{ KHz} < f_i < (F_s / 2) - 2 \text{ KHz} \\ 1.5 & f_i > (F_s / 2) - 2 \text{ KHz} \end{cases}$$

Where f_i is the upper frequency of the i th band and F_s is the sampling frequency in Hz. The main difference between the MB and the NSS algorithm is in the estimation of the over subtraction factors. The MB approach estimates one subtraction factor for each frequency band, whereas the NSS algorithm estimates one subtraction factor for each frequency bin [4]

2.6 MMSE Spectral Subtraction Algorithm

Minimum Mean Square Error (MMSE) Spectral subtraction Algorithm is proposed by Sim [11]. A method for optimally selecting the subtractive parameters in the mean error sense [17,18]. Consider a general version of the spectral subtraction algorithm

$$|X(\omega)|^P = \alpha_p(\omega) |Y(\omega)|^P - \alpha_p(\omega) |D_e(\omega)|^P \quad (13)$$

Where $\alpha_p(\omega)$ and $\alpha_p(\omega)$ are the parameters of interest. P is the power exponent and $|D_e(\omega)|$ is the average noise spectrum obtained during non speech activity. The parameter $\alpha_p(\omega)$ can be determined by minimizing the mean square error spectrum

$$e_p(\omega) = |X_p(\omega)|^P - |X_e(\omega)|^P \quad (14)$$

Where $|X_p(\omega)|$ is the clean speech spectrum, assuring an ideal spectral subtraction model and $|X_e(\omega)|$ is enhanced speech. Here assumption is that noisy speech spectrum consists of the sum of two independent spectra

the $|X_p(\omega)|^P$ spectrum and the true noise spectrum $|D_e(\omega)|^P$. Where P is constant, considering $P = 1$ and processing equation (13) by minimizing the mean square error of the error spectrum giving equation (14) with respect to $\alpha_p(\omega)$ and $\alpha_p(\omega)$, we get the following optimal subtractive parameters [4].

$$\alpha_p(\omega) = \xi^P(\omega) / (1 + \xi^P(\omega)) \quad (15)$$

$$\alpha_p(\omega) = \alpha_p(\omega) [1 - \xi^{-P/2}(\omega)] \quad (16)$$

Where

$$\xi(\omega) = E[|X_p(\omega)|^2] / E[|D_e(\omega)|^2] \quad (17)$$

2.7 Selective Spectral Subtraction Algorithm

All previously mentioned methods treated all speech segments equally, making no distinction between voiced and unvoiced segments. Due to the spectral differences between vowels and consonants [4] several researchers have proposed algorithms that treated the voiced and unvoiced segment differently. The resulting spectral subtractive algorithms were therefore selective for different classes of speech sounds [4]. The two band spectral subtraction algorithm was proposed in [13]. The incoming speech frame was first classified into voiced or unvoiced by comparing the energy of the noisy speech to a threshold. Voiced segments were then filtered into two bands, one above the determined cutoff frequency (high pass speech) and one below the determined cutoff frequency (low pass speech). Different algorithms were then used to enhance the low passed and high passed speech signals accordingly. The over subtraction algorithm was used for the low passed speech based on the short term FFT. The subtraction factor was set according to short term SNR as per [5]. For high passed voiced speech as well as for unvoiced speech, the spectral subtraction algorithm was employed with a different spectral estimator [4].

A dual excitation Model was proposed in [3] for speech enhancement. In the proposed approach, speech was decomposed into two independent components voiced and unvoiced components. Voiced component analysis was performed first by extracting the fundamental frequency and the harmonic amplitudes. The noisy estimates of the harmonic amplitudes were adjusted according to some rule to account for any noise that might have leaked to the harmonics. Following that the unvoiced component spectrum was computed by subtracting the voiced spectrum from the noisy speech spectrum. Then a two pass system, which included a modified Wiener Filter, was used to enhance the unvoiced spectrum. Finally the enhanced speech consists of the sum of the enhanced voiced and unvoiced components. Treating voiced and unvoiced segments differently can bring about substantial improvements in performance [4]. The major challenge with such algorithms is making accurate

and reliable voiced, unvoiced decisions particularly at low SNR conditions.

2.8 Spectral Subtraction based on perceptual properties

In the preceding methods, the subtractive parameters were computed experimentally, based on short term SNR levels [5] or obtained optimally in a mean square error sense [11]. No perceptual properties of the auditory system have been considered. An algorithm proposed by Virag [14] that incorporates psycho acoustical properties of speech signal, in the spectral subtraction process. The main objective of this algorithm is to remove the residual noise perceptually inaudible and improve the intelligibility of enhanced speech by taking into account the properties of the human auditory system [4]. Method proposed by Virag [14] was based on idea that, if the estimated masking threshold at a particular frequency is low, the residual noise level might be above. The threshold and will therefore be audible. The subtraction parameters should therefore attain their maximal values at that frequency. Similarly, if the masking threshold level is high at a certain frequency, the residual noise will be masked and will be inaudible. The subtraction parameters should attain their minimal values at that frequency. The subtraction parameters α & β are given as

$$\begin{aligned}\alpha(\omega) &= F_a [\alpha_{\min}, \alpha_{\max}, T(\omega)] \\ \beta(\omega) &= F_b [\beta_{\min}, \beta_{\max}, T(\omega)]\end{aligned}\quad (18)$$

Where $T(\omega)$ was the masking threshold, α_{\min} and α_{\max} were set to 1 and 6 respectively and spectral floor constants β_{\min} & β_{\max} , were set to 0 and 0.02 respectively in [4]. The $F_a(\omega)$ function had the following boundary conditions

$$\begin{aligned}F_a(\omega) &= a_{\max} && \text{if } T(\omega) = T(\omega)_{\min} \\ &= a_{\min} && \text{if } T(\omega) = T(\omega)_{\max}\end{aligned}\quad (19)$$

Where $T(\omega)_{\min}$ and $T(\omega)_{\max}$ are the minimal and maximum values of masking thresholds estimated in each frame. Similarly the function $F_b(\omega)$ was computed using β_{\min} and β_{\max} as boundary conditions. The main advantage of Virag's approach lies in the use of noise masking thresholds $T(\omega)$ rather than SNR levels for adjusting the parameters $\alpha(\omega)$ and $\beta(\omega)$. The masking thresholds $T(\omega)$ provide a smoother evolution from frame to frame than the SNR. This algorithm requires accurate computation of the masking threshold.

3 PERFORMANCE OF SPECTRAL SUBTRACTION ALGORITHMS

The spectral subtraction algorithm was evaluated in many studies, primarily using objective measures such as SNR improvement and spectral distances and then subjective listening tests. The intelligibility and speech quality measures reflect the true performance of speech en-

hancement [4] algorithms in realistic scenarios. Ideally, the SS algorithm should improve both intelligibility and quality of speech in noise. Results from the literature were mentioned as follows.

Boll[5] performed intelligibility and quality measurement tests using the Diagnostic Rhyme Test (DRT). Result indicated that SS did not decrease speech intelligibility but improved speech quality particularly in the area of pleasantness and inconspicuousness of the background noise. Lim [4] evaluated the intelligibility of nonsense sentences in white noise at -5, 0, and +5dB SNR processed by a generalized SS algorithm (eqa. No.5). the intelligibility of processed speech was evaluated for varies power exponents P ranging from P = 0.25 to P = 2. Results indicated that SS algorithm did not degrade speech intelligibility except when P = 0.25. Kang and Fransen [4] evaluated the quality of noise processed by the SS algorithm and then fed to a 2400 bps LPC recorder. Here SS algorithm was used as a pre-processor to reduce the input noise level. The Diagnostic Acceptability Measure (DAM) test [19] was used to evaluate the speech quality of ten sets of noisy sentences, recorded actual military platforms containing helicopter, tank, and jeep noise results indicated that SS algorithm improved the quality of speech. The largest improvement in speech quality was noted for relatively stationary noise sources [4, 2]. The NSS algorithm was successfully used in [8] as a pre-processor to enhance the performance of speech recognition systems in noisy environment. The performance of the multiband spectral subtraction algorithm [9] was evaluated by Hu Y. and Loizou [2, 19] using formal subjective listening tests conducted according to ITU-T P.835 [20]. The ITU T P.835 methodology is designed to evaluate the speech quality along with three dimensions signal distortion, noise distortion and overall quality. Results indicated that the MBSS algorithm performed the best consistently across all noise conditions, [4] in terms of overall quality. In terms of noise distortion the MBSS algorithms performed well, except in 5dB train and 10dB street conditions. The algorithm proposed by Virag was evaluated in [14] using objective measures and subjective tests, and found better quality than the NSS and standard SS algorithms. The low energy segments of speech are the first to be lost in the subtraction process; particularly when over subtraction is used. Overall most studies confirmed that the SS algorithm improves speech quality but not speech intelligibility.

4 CONCLUSION

Various spectral subtraction algorithms proposed for speech enhancement were described in above sections. These algorithms are computationally simple to implement as they involve a forward and an inverse Fourier transform. The simple subtraction processing comes at a price. The subtraction of the noise spectra from the noisy spectrum introduces a distortion in the signal known as

Musical noise [4]. We presented different techniques that mitigated the Musical noise distortion. Different variations of spectral subtraction were developed over the years. The most common variation involved the use of an over subtraction factor that controlled to some amount of speech spectral distortion caused by subtraction process. Use of spectral floor parameter prevents the resultant spectral components from going below a preset minimum value. The spectral floor value controlled the amount of remaining residual noise and the amount of musical noise [4]. Different methods were proposed for computing the over subtraction factor based on different criteria that included linear [5] and nonlinear functions [8] of the spectral SNR of individual frequency bins or bands [9] and psychoacoustic masking threshold [14]. Evaluation of spectral subtractive algorithms revealed that these algorithms [4] improve speech quality and not affect much more on intelligibility of speech signals.

ACKNOWLEDGMENT

Mrs. Anuradha R. Fukane wishes to thank Dr. Bhide S. D. and Dr. Madhuri Khambete for their valuable guidance and support.

REFERENCES

- [1] Yi Hu and Philipos C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms" *IEEE Trans. Speech Audio Proc.* 2007;49(7): 588-601.
- [2] Gustafsson H., Nordhohm S, Claesson I(2001) "Spectral subtraction using reduced delay convolution and adaptive averaging" . . *IEEE. Trans. Speech Audio Process*,9(8), 799-805.
- [3] Kim W, Kang S, and ko H.(2000) "Spectral subtraction based on phonetic dependancy and masking effects" *IEEE. Proc.vision image signal process*, 147(5),pp423-427
- [4] Phillips C Loizou "Speech enhancement theory and practice" 1st ed. Boca Raton, FL.: CRC, 2007. Releases Taylor & Francis
- [5] Berouti,M. Schwartz,R. and Makhoul,J.,"Enhancement of Speech Corrupted by Acoustic Noise", *Proc ICASSP 1979*, pp208-211,.
- [6] Paliwal K. and Alsteris L.(2005), "On usefulness of STFT phase spectrum in human listening tests " *Speech Commmun.*45(2),153-170
- [7] Boll,S.F.,"Suppression of Acoustic Noise in Speech using Spectral Subtraction", *IEEE Trans ASSP* 27(2):113-120, April 1979
- [8] Lockwood, P. and Boudy,J.,"Experiments with a Nonlinear Spectral Subtractor (NSS), Hidden Markov Models and the projection, for robust speech recognition in cars", *Speech Communication*, 11, pp215-228, Elsevier 1992.
- [9] Kamath S. and Loizou P.(2002) "A multiband spectral subtraction methode for enhancing speech currupted by colored noise" *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing*
- [10] Hu Y., Bhatnager M. Loizou P.(2001) "A crosscorellation technique for enhancing speech currupted with correlated noise" . *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing*1.pp 673-676
- [11] Sim B, Tong Y, chang J., Tan C.(1998) 'A parametric formulation of the generalized spectral subtraction method' *IEEE. Trans. Speech Audio Process*,6(4), 328-337
- [12] Hardwick J., Yoo C. and Lim J (1998) "speech enhancement using dual exitation model" *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing 2*, pp 367-370
- [13] He C, and Zweig G. (1999) "Adaptive two band spectral subtraction with multiwindow spectral estimation" *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing ,2*, pp 793-796
- [14] Virag, N., (1999). "Single channel speech enhancement based on masking properties of the human auditory system". *IEEE. Trans. Speech Audio Process*,7(3), 126-137..
- [15] Lebart K, Boucher J M,(2001)" A New method based on spectral subtraction for speech enhancement" *Acta acustica, Acustica vol. 87* pp359-366..
- [16] R. Martin, "Spectral Subtraction Based on Minimum Statistics," in *Proc. Euro. Signal Processing Conf. (EUSIPCO)*, pp. 1182-1185, 1994
- [17] Martin, R(2002) "Speech Enhancement Using MMSE Short Time Spectral Estimation with Gamma Distributed Speech Priors,"in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 1, pp. 253-256, 2002
- [18] Epraim Y. and malah D "Speech Enhancement Using minimum mean squire error shorttime spectral amplitude estimator" *IEEE, Trans. on Audio, Speech, signal pross.*vol 6(4)pp 328-337)
- [19] Yi Hu and Philipos C. Loizou, Senior Member, IEEE "Evaluation of Objective Quality Measures for Speech Enhancement" *IEEE, Trans. on Audio, Speech, and Language pross.*vol 16, (2008)
- [20] ITU-T(2003) "subjective test Methodology for evaluating speech communication system that include noise supression algorithn." *ITU-T recommendation p.835*