DEVELOPMENTS IN SPECTRAL SUBTRACTION FOR SPEECH ENHANCEMENT

Paurav Goel[1] , Anil Garg[2]

[1,2]Electronics and Communication Department,
M.M. University, Mullana (Ambala), Haryana , INDIA

Abstract

Speech enhancement aims to improve speech quality by using various techniques and algorithms. Over the past several years there has been considerable attention focused on the problem of enhancement of speech degraded by additive background noise. Background noise suppression has many applications. Using mobile in a noisy environment like in streets or in a car is an obvious application, removing the background noise when sending speech from the cockpit of an airplane to the ground or to the cabin. The spectral - subtractive algorithm is historically one of the first algorithms proposed for additive background noise and it has gone through many modifications with time. This is a review paper and its objective is to provide an overview of the variety of spectral subtraction techniques that have been proposed for enhancement of speech degraded by additive background noise during past decades . Section I gives the Introduction to Speech enhancement and explain basic Spectral Subtraction technique. Section II gives the various modified versions of spectral subtraction till date.

Keywords – speech enhancement, spectral subtraction, residual noise, musical noise.

INTRODUCTION

Speech signals from the uncontrolled environment may contain degradation components like additive background noise along with required speech components. This make the listening task difficult for a direct listener and gives poor performance in automatic speech processing tasks like speech recognition speaker identification, hearing aids, speech coders etc. The aim of speech enhancement is to improve the quality and intelligibility of degraded speech signal. Improving quality and intelligibility of speech signals, reduces listener's fatigue. Quality can be measured in terms of signal distortion but intelligibility and pleasantness are difficult to measure by any mathematical algorithm. In this study, a speech enhancement algorithm using spectral subtraction and its modified versions are presented.

BASICS OF SPECTRAL SUBTRACTION

Spectral Subtraction is a single channel speech enhancement technique. Single channel enhancement techniques apply to situations in which only one acquisition channel is available. These methods are interesting due to the simplicity in microphone installation but the major constraint of single channel methods is that there is no reference signal for the noise available. Therefore the power spectral density of the noise has to be estimated based on the available noisy speech signal only and this is what makes it a challenging task. In all single channel enhancement techniques, we assume the available speech signal model as

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

Where x(n) represents the pure speech signal, which is assumed to be a stationary signal whenever processing is done on a short time basis, d(n) is the uncorrelated additive noise and y(n) represents the degraded speech signal.

Spectral subtraction is based on the principle[1] that one can obtain an estimate of the clean signal spectrum by subtracting an estimate of the noise spectrum from the noisy speech spectrum. The noise spectrum can be estimated, and updated, during the periods when the signal is absent or when only noise is present. Assumption is noise is additive¸ 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. The noise corrupted input speech signal which is composed of the clean speech signal x(n) and the additive noise signal d(n) is shown in eq.(1) above. The above eq. can be given in Fourier domain as shown:

Y[w] = X[w] +D[w] **(2)** Y[w] can be expressed in terms of Magnitude and phase as

Y[w] = Y |(w)| e $^{j Ø y}$ **(3)** here |Y(w)| is the magnitude spectrum and Ø is the phase spectra of the corrupted noisy speech signal. Noise spectrum in terms of magnitude and phase spectra as

D[w] =| D[w] | e $^{j Ø y}$ **(4)**

We can estimate the clean speech signal simply by subtracting noise spectrum from noisy speech spectrum¸ in equation form

X̂ (w) = [|Y(w)|-|D(w)|] e $^{j Ø y}$ **(5)**

The magnitude of noise spectrum |D(w)| is unknown but can be replaced by its average value computed during non speech activity i.e. during speech pauses.

The spectral subtraction algorithm is computationally simple as it only involves a forward and inverse Fourier Transform. The basic block diagram of spectral subtraction method is shown below.

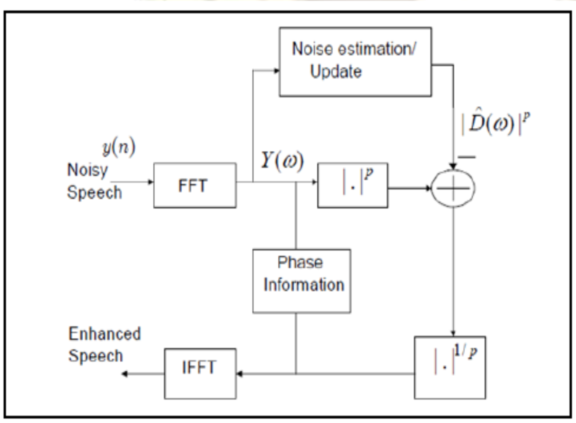

Fig.1 Block diagram of Spectral Subtraction technique.

Noisy speech signal is segmented and then windowed using Hamming Window[8]. Then Discrete Fourier transform (DFT) of segmented and windowed Noisy speech signal is taken. DFT of noisy signal is then given to noise estimation block and speech enhancement block. Noise estimation block estimate the noise during the pauses and find the noise spectrum. In most speech-enhancement algorithms, it is make assumed that an estimate of the noise spectrum is available. The noise estimate can have a major impact on the quality and intelligibility of the enhanced signal. If the noise estimate is too low, unwanted residual noise will be audible, if the noise estimate is too high, speech will be distorted.

From the above discussion it is clear that subtraction process needs to be done carefully to avoid any speech distortion. If too much is subtracted, then some speech information might be removed as well, while if too little is subtracted then much of the interfering noise remains. It is clear from equation(5) 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[1].

|Xe(ω)| = |Y(ω)| - | De(ω)|,
if |Y(ω)| > |De(ω)| else
= 0 **(6)**

SECTION II

The first method for Spectral subtraction was introduced in post 1970's. In past more then 30 years this method has been modified and new methods has been developed. This section gives the study of some of such methods beginning from the starting till date.

2.1. In 1979 Berouti [2] gave a Spectral Subtraction method, for enhancing speech corrupted by broadband noise. As discussed in Section 1, original method entails subtracting an estimate of the noise power spectrum from the speech power spectrum, setting negative differences to zero, recombining the new power spectrum with the original phase, and then reconstructing the time waveform. While this method reduces the broadband noise, it also usually introduces an annoying "musical noise"[11]. We have devised a method that eliminates this "musical noise" while further reducing the background noise. The method consists in subtracting an overestimate of the noise power spectrum, and preventing the resultant spectral components from going below a preset minimum level (spectral floor). The method can automatically adapt to a wide range of signal-to-noise ratios, as long as a reasonable estimate of the noise spectrum can be obtained. The technique can be described using equation below

$$|X_ej(\omega)|^2 = |Y_j(\omega)|^2 - |D_e(\omega)|^2, \quad \text{if } |Y_j(\omega)|^2 > (\alpha+\beta)|D_e(\omega)|^2$$

else

$$= \beta|D_e(\omega)|^2 \qquad (7)$$

Here $|X_ej(\omega)|$ denotes the enhanced spectrum estimated in frame j and $|D_e(\omega)|$ is the spectrum of the noise obtained during non speech activity.

With $\alpha \geq 1$ and D < $\beta \leq 1$. 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 α > 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 β > 0 compared to when β =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. In this way this method can adapt to various Signal to Noise ratios by adjusting the α and β and reduce the musical noise. The parameter values have to be set optimally so that the best enhancement performance can be achieved. First compute the average segmental SNR as a function of α for various values of β. Then find β optimally, by fixing α at unity. And then refine α by fixing β to its optimal value.[21]

2.2. In the same year 1979, S.F.Boll[3] also proposed method for removal of acoustic noise in speech. In this method a spectral estimator is used to compute the spectral error and then four methods are used to minimize the error. Speech, suitably low-pass filtered and digitized, is analyzed by windowing data from half-overlapped input data buffers. The magnitude spectra of the windowed data are calculated and the spectral noise bias calculated during non speech activity is subtracted off. Resulting negative amplitudes are then zeroed out. Secondary residual noise suppression is then applied. A time waveform is recalculated from the modified magnitude. This waveform is then overlap added to the previous data to generate the output speech.

Consider that a windowed noise signal *n(k)* has been added to a windowed speech signal ***s(k),*** with their sum denoted by x(k) **).**

Then $x(k) = s(k) + n(k).$ **(8)**

The spectral subtraction estimator response is given by

$$\hat{S}(e^{j\omega}) = [|X(e^{j\omega})| - \mu(e^{j\omega})]\, e^{j\theta_x(e^{j\omega})} \qquad (9)$$

Where $\mu(e^{jw})$ is the average value of $|N(e^{jw})|$ taken during non- speech activity given by

$$\mu(e^{j\omega}) = E\{|N(e^{j\omega})|\} \qquad (10)$$

The response can also be given as

$$\hat{S}(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega}) \qquad (11)$$

where

$$H(e^{j\omega}) = 1 - \frac{\mu(e^{j\omega})}{|X(e^{j\omega})|} \qquad (12)$$

denotes the spectral filter.

The spectral error is given by:

$$\epsilon(e^{j\omega}) = \hat{S}(e^{j\omega}) - S(e^{j\omega}) = N(e^{j\omega}) - \mu(e^{j\omega})\, e^{j\theta_x} \qquad (13)$$

This technique works to reduce the perceptual effects of this spectral error by following means- i) magnitude averaging; ii) half-wave rectification; iii) residual noise reduction iv) additional signal attenuation during non speech activity.

i) Magnitude averaging: The variance of the noise spectral estimate i.e the spectral is reduced by averaging over as many spectral magnitude sets as possible as shown-

$$\overline{|X(e^{j\omega})|} = \frac{1}{M}\sum_{i=0}^{M-1} |X_i(e^{j\omega})| \qquad (14)$$

Where $X_i(e^{jw})$ is the ith timed window transform of x(k). Now the estimate is given by-

$$S_A(e^{j\omega}) = [\overline{|X(e^{j\omega})|} - \mu(e^{j\omega})]\, e^{j\theta_x(e^{j\omega})} \qquad (15)$$

Now the spectral error is given as

$$e(e^{j\omega}) = S_A(e^{j\omega}) - S(e^{j\omega}) \cong \overline{|N|} - \mu \qquad \textbf{(16)}$$

It depicts that if the number of averages will be more lesser will be the error. But the number of averages is limited by the number of analysis windows which can be fit into the stationary speech time interval with a 256-point window being used.

ii) Half Wave rectification: For each frequency w where the noisy signal spectrum magnitude $X(e^{jw})$ is less than the average noise spectrum magnitude $\mu(e^{jw})$, the output is set to zero. This modification can be simply implemented by half-wave rectifying $H(e^{jw})$. The advantage of half rectification is that the noise floor is reduced by $\mu(e^{jw})$. The disadvantage of half rectification can exhibit itself in the situation where the sum of the noise plus speech at a frequency w is less than $\mu(e^{jw})$. **.** Then the speech information at that frequency is incorrectly removed, implying a possible decrease in intelligibility.

iii) Residual Noise Reduction: The noise that remains after the mean is removed can be suppressed or even removed by selecting the minimum magnitude value from the three adjacent analysis frames in each frequency bin where the current amplitude is less than the maximum noise residual measured during non speech activity.

iv) Additional signal attenuation during non speech activity: The final improvement in noise reduction is signal suppression during non speech activity. If speech activity is absent, then $\hat{S}(e^{jw})$ will consist of the noise residual which remains after half-wave rectification and minimum value selection. A measure for detecting the absence of speech is given by

$$T = 20 \log_{10} \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\hat{S}(e^{j\omega})}{\mu(e^{j\omega})} \right| d\omega \right]$$

If T was less than - 12 dB, the frame was classified as having no speech activity. If such situation arise the further attenuation of upto -30 db is carried out.

2.3. Nathalie Virag, in 1999[4] came up with a technique that proposed the introduction of human perception in enhancement process. This process models the certain aspects of auditory system in the enhancement process. This model uses the phenomenon of noise masking. It is related to the concept of critical band analysis, which is a central analysis mechanism in the inner ear. Auditory system is incapable of distinguishing two signals close in the time or frequency, a weak signal is made inaudible by a stronger signal occurring simultaneously. This method combines the generalized spectral subtraction with adaptive parameters and noise masking and gives the best tradeoff between the amount of noise reduction, the speech distortion and the level of residual noise in a perceptual sense. The proposed enhancement scheme composed of the following main steps:

1) Spectral decomposition.
2) Speech/noise detection and estimation of noise during speech pauses.
3)Calculation of the noise masking threshold[16][17]
4) Adaptation in time and frequency of the subtraction parameters and based on the noise masking threshold
5)Calculation of the enhanced speech spectral magnitude via parametric subtraction with adapted parameters
6) Inverse transform

First and last two steps can be realized using AMS framework discussed ahead. Calculation of Noise masking threshold is an important aspect of this method. The noise masking threshold is computed from the clean speech signal. However, in the proposed enhancement scheme, only the noisy signal is available. Therefore this threshold has to be estimated in noise The residual noise modifies the [18][19][20] tonality of the signal and the masking threshold is slightly different from the one obtained from the clean speech, especially for high frequencies. This is represented in Fig2.

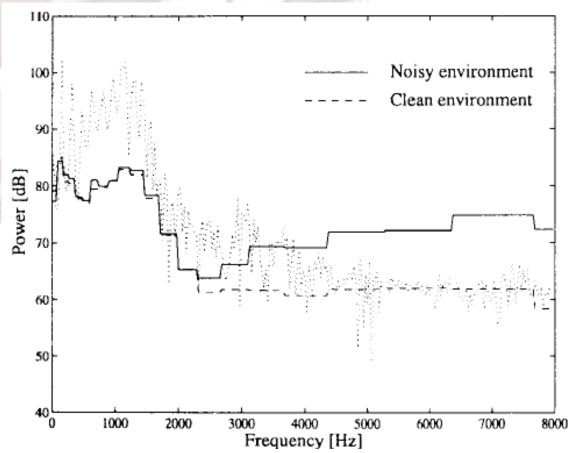

Fig.2 Example of Noise masking Threshold

The adaption of subtraction parameters(α,β) has been discussed earlier. Here the parameters are chosen in such a way that the residual noise stays below the masking threshold of the auditory system. This would ensure that the residual noise is masked, and remains inaudible. However, if the noise level increases, the masking threshold is too low to completely mask the residual noise without increasing the speech distortion, leading to a synthetic sound.

Therefore, the proposed adaptation is based on the following consideration: if the masking threshold is high, residual noise will be naturally masked and inaudible. Hence, there is no need to reduce it in order to keep distortion as low as possible. In this case, the subtraction parameters are kept to their minimal values. However, if the masking threshold is low, residual noise will be annoying to the human listener and it is necessary to reduce it. This is done by increasing the subtraction parameters(7). This is depicted in Fig3.

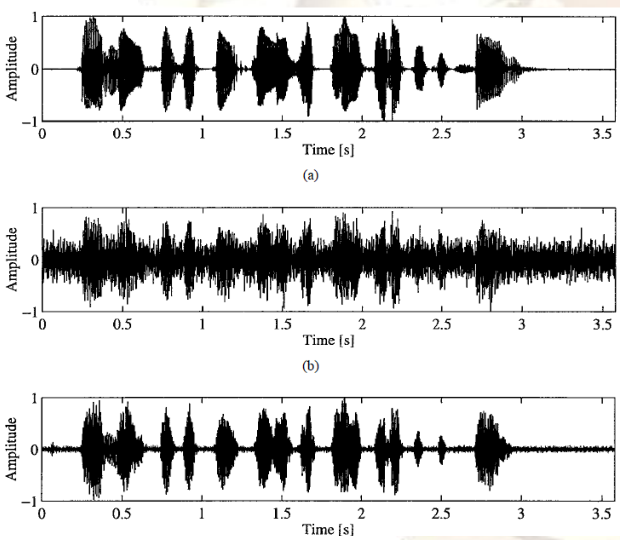

Fig.3 Example of adaption of parameters w.r.t to noise threshold.(a) clean speech signal (b) Noisy signal (c) Enhanced Signal

Most of the methods we have discussed yet are using the traditional AMS framework i.e. Analysis–Modification-Synthesis[5] framework.

Let us consider an additive noise model

$$x(n)= s(n) + d(n) \qquad (18)$$

In speech processing, the hamming window with 20–40 ms[8] duration is typically employed . Using STFT analysis we can represent Eq. (18) as

$$X(n,k) = S(n,k) + D(n,k). \qquad (19)$$

The block diagram in fig 4 of Acoustic AMS framework further explains the method.

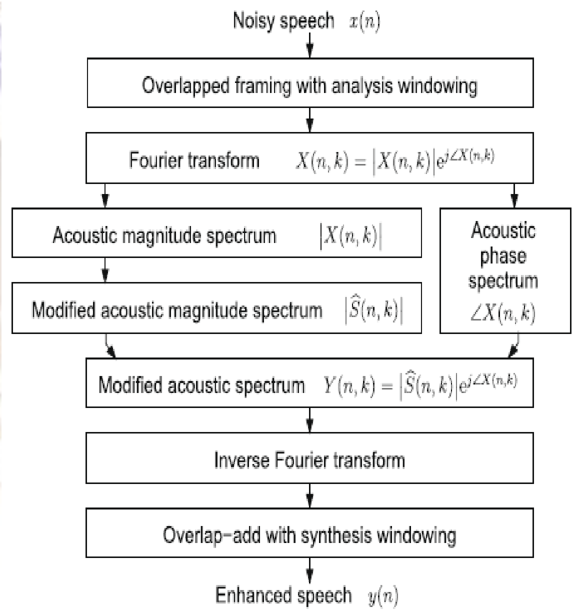

Fig.4 AMS framework

Here in the fig.4 |X(n,k)| denotes the acoustic magnitude spectrum and $\angle X(n,k)$ denotes the acoustic phase spectrum $|\tilde{S}(n,k)|$ denote the enhanced magnitude spectrum and Y(n,k) denote the modified spectrum which ia combination of enhanced magnitude specrum and the noisy phase spectrum .

2.4. K. Paliwal in 2010[10] came with a technique that compensated the additive noise distortion by applying spectral subtraction algorithm in modulation domain. He extended the traditional analysis-modification-synthesis framework to include modulation domain processing. This method uses the time trajectories of the short-time acoustic magnitude spectrum for the computation of the short-time modulation spectrum[14][15]. This speech enhancement method processes each frequency component of the acoustic magnitude spectra, obtained during the analysis stage of the acoustic AMS procedure frame-wise across time using a secondary (modulation)

AMS framework as shown below in fig5. Thus the modulation spectrum is computed using STFT analysis as follows

$$\mathcal{X}(\eta, k, m) = \sum_{l=-\infty}^{\infty} |X(l, k)| v(\eta - l) e^{-j2\pi m l/M} \quad \textbf{(20)}$$

where η is the acoustic frame number, k refers to the index of the discrete acoustic frequency, m refers to the index of the discrete modulation frequency, M is the modulation frame duration (in terms of acoustic frames) and v(η) is a modulation analysis window function.

|X(n,k,m)| represents the modulation magnitude spectrum which is replaced with |S῀(n,k,m)| as the estimate of clean modulation magnitude spectrum. Similarly as discussed before adaptive parameters(α, β) are used to detemine the oversubtraction factor and the spetral floor.ρ is the factor that determines thr spectrum domain.ρ=1 represents the magnitude spectrum domain and ρ=2 denotes the power spectrum domain(fig 1). This method results in improved speech quality and does not suffer from musical noise artifacts .

2.5. Yang Lu, Philipos C. Loizou in 2008[9] came with a geometric approach for spectral subtraction that addresses the shortcomings of the spectral subtraction algorithm. The traditional power spectral subtraction algorithm is computationally simple to implement but suffers from musical noise distortion. The derivation of the spectral subtraction equations is based on the assumption that the cross terms involving the phase difference between the clean and noise signals are zero[7]. The cross terms are assumed to be zero because the speech signal is uncorrelated with the interfering noise. The results of geometric approach tells the ffect of neglecting the cross terms on speech recognition performance.

In the power spectral subtraction[1] we come across an expression which gives the short-term power spectrum of the noisy speech given as

$$\begin{aligned} |Y(\omega_k)|^2 &= |X(\omega_k)|^2 + |D(\omega_k)|^2 + X(\omega_k) \cdot D^*(\omega_k) \\ &\quad + X^*(\omega_k) D(\omega_k) \\ &= |X(\omega_k)|^2 + |D(\omega_k)|^2 + 2|X(\omega_k)| \\ &\quad \cdot |D(\omega_k)| \cos(\theta_X(k) - \theta_D(k)). \end{aligned} \quad \textbf{(21)}$$

In traditional spectral subtraction, we assume cos($\theta_x(k) - \theta_D(k)$) to be zero as we consider the noise and speech to be uncorrelated i.e orthogonal to each other. As a result we are left with

$$|\mathbf{Y}(\mathbf{w_k})|^2 = |\mathbf{X}(\mathbf{w_k})|^2 + |\mathbf{D}(\mathbf{w_k})|^2 \quad \textbf{(22)}$$

Above eq. can be written as

$$|\hat{\mathbf{X}}(\mathbf{w_k})|^2 = \mathbf{H}^2(\mathbf{w_k})|\mathbf{Y}(\mathbf{w_k})|^2 \quad \textbf{(23)}$$

Where

$$H(\omega_k) = \sqrt{1 - \frac{|\hat{D}(\omega_k)|^2}{|Y(\omega_k)|^2}} = \sqrt{\frac{\gamma(k) - 1}{\gamma(k)}} \quad \textbf{(24)}$$

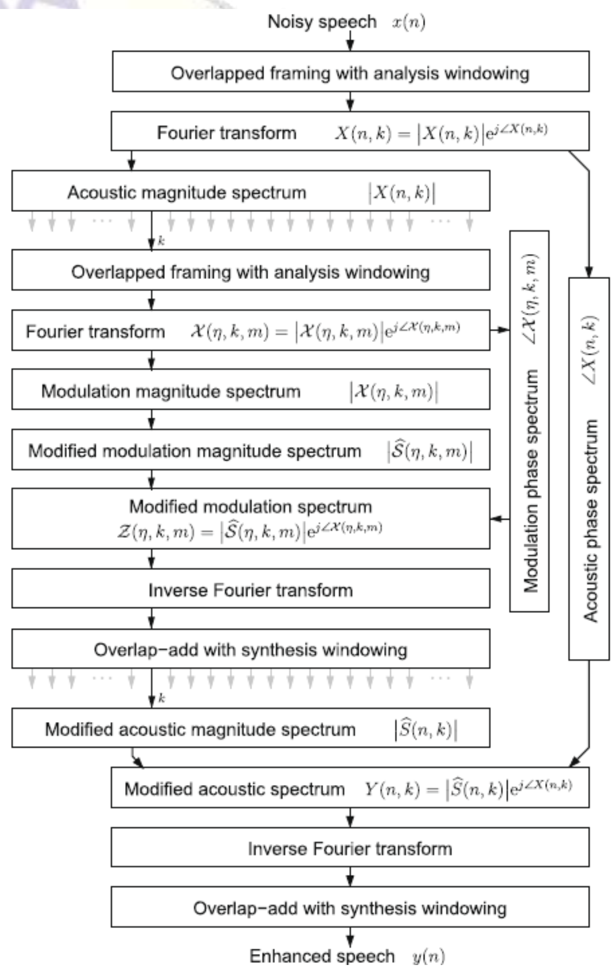

Fig.5 Block diagram of Spectral subtraction in modulation domain

Is the gain (or suppression) function when cross terms are zero.

But if we involve cross terms what effect it would have on speech enhancement. Consider the following expression

$$|Y(\omega_k)|^2 = |X(\omega_k)|^2 + |D(\omega_k)|^2 + \Delta Y(\omega_k) = |\widehat{Y}(\omega_k)|^2 + \Delta Y(\omega_k) \quad \textbf{(26)}$$

$\Delta Y(w_k)$ represents the cross terms.

The relative error introduced when neglecting the cross terms

$$\varepsilon(k) \triangleq \frac{\left| |Y(\omega_k)|^2 - |\widehat{Y}(\omega_k)|^2 \right|}{|Y(\omega_k)|^2} = \frac{|\Delta Y(\omega_k)|}{|Y(\omega_k)|^2} \quad \textbf{(27)}$$

The cross term error can be written in terms of SNR as

$$\varepsilon(k) = \left| \frac{2\sqrt{\xi(k)}\cos(\theta_X(k) - \theta_D(k))}{1 + \xi(k) + 2\sqrt{\xi(k)}\cos(\theta_X(k) - \theta_D(k))} \right| \quad \textbf{(28)}$$

The above expression helps in understanding the relation between cross terms and spectral SNR. It is depicted through following fig6. The fig 6. gives support to our consideration that that the cross terms are zero is not valid for spectral SNR as cross term error has significant values near 0 dB which is the region wherein most speech enhancement algorithms operate. Consequently, large estimation errors[15] can result from the approximation taken in traditional power spectrum subtraction i.e taking cross terms to be zero.

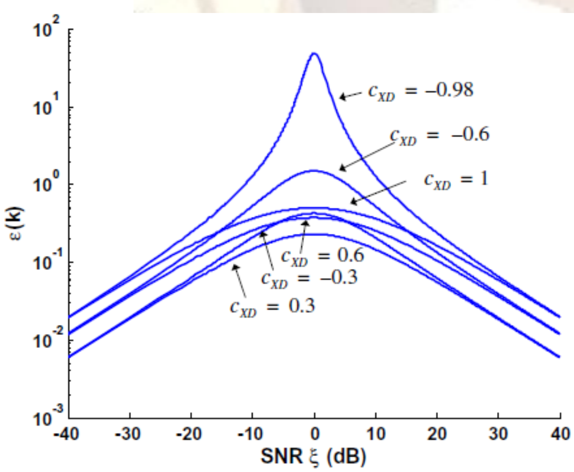

Fig.6 Plot between cross term error and SNR

The geometric approach which takes cross terms into consideration represents the $Y(w_k)$ geometrically in the complex plane as sum of two complex numbers $X(w_k)$ and $D(w_k)$.

Where a_X, a_Y, a_D and θ_X, θ_Y, θ_D are the magnitude and phase of Clean , noisy and noise spectra respectively in the polar form of speech model. Using this geometric approach we get the gain or suppression function as

$$H_{GA} = \frac{a_X}{a_Y} = \sqrt{\frac{1 - c_{YD}^2}{1 - c_{XD}^2}}. \quad \textbf{(29)}$$

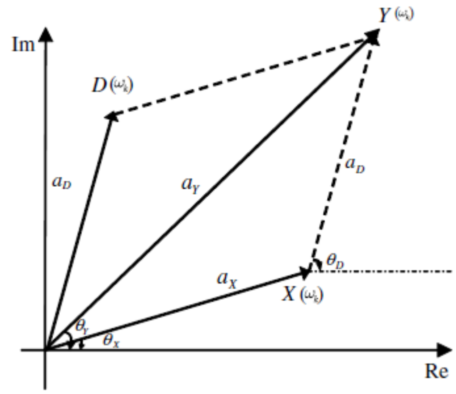

Fig.6 Geometric representation of speech model

where $C_{YD} = \cos(\theta_Y - \theta_D)$ and $C_{XD} = \cos(\theta_X - \theta_D)$.

The above gain function is always real and positive (i.e. $H_{GA} \geq 0$) since the terms C_{YD} and C_{XD} are bounded by one. Unlike the power spectral subtraction gain function which is always positive and smaller (or equal) than one, the above gain function can be larger than one if $|C_{YD}| < |C_{XD}|$

In the above gain function if cross terms are taken to be zero i.e. $C_{XD} = 0$, we get

$$c_{YD} = \frac{a_D}{a_Y} \quad \textbf{(30)}$$

Multiplication of the noisy signal by the suppression function given in Eq. (5) would not yield the clean signal magnitude spectrum even if we had access to the true noise magnitude spectrum (i.e., $|D(x_k)|$). In contrast, multiplication of the noisy magnitude spectrum (a_Y) by the suppression function given in Eq. (11) would yield exactly the clean signal magnitude spectrum (i.e., a_x).

GA algorithm shows that it performs significantly better than the traditional spectral subtraction algorithm. The above discussion tells that musical noise is absent when GA for spectral subtraction is used but possess some residual noise.

CONCLUSION

We have gone through number of spectral subtraction techniques and they give a good idea of not only the technique and what kind of work is been carried out on these techniques. Some suggested over-subtracting estimates of the noise spectrum and spectral flooring (rather than setting to zero) negative values (Berouti et al.1979). Yet, others suggested using a psychoacoustical model to adjust the over-subtraction parameters so as to render the residual noise inaudible (Virag, 1999).

Berouti and Boll's subtraction methods form the basis of spectral subtraction for the new methods. In most of the methods we have found that there remains a trade off between residual noise and the signal distortion[13]. Spectral subtraction suffers from a problem of introducing artifacts like musical noise while removing residual noise. As discussed earlier this noise is due to production of small peaks which have tone like nature while making the negative values zero after subtraction. Most of the research on spectral subtraction techniques now is concentrated on decreasing or removing this musical noise. As we have discussed in this paper the various techniques which aimed at reducing the musical noise. The methods like spectral subtraction using modulation domain and geometric approach for spectral subtraction has been successful to limit musical noise to a certain extent. When both objective and subjective tests were performed on modulation approach the results of these experiments show that the proposed method results in improved speech quality and it does not suffer from musical noise typically associated with spectral subtractive algorithms. These results indicate that the modulation domain processing is a useful alternative for spectral subtraction. Where as geometric approach gave better gain function than tradition subtraction technique which lead to removal of musical noise to a large extent. It has also been proved with various subjective tests.

REFERENCES

[1]Lim, J., Oppenheim, A., 1979. Enhancement and bandwidth compression of noisy speech. Proc. IEEE 67 (12), 1586–1604.

[2]Berouti,M. Schwartz,R. and Makhoul,J.," *Enhancement of Speech Corrupted by Acoustic Noise*", *Proc ICASSP 1979, pp208-211*,.

[3] Boll,S.F.," Suppression of Acoustic Noise in Speech using Spectral Subtraction", *IEEE Trans ASSP 27*(2):113-120, April 1979

[4]Virag, N., 1999. Single channel speech enhancement based on masking properties of the human auditory system. IEEE Trans. Speech Audio Process. 7 (2), 126–137.

[5] **J.** Allen, "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-25, pp. 235-239, June 1977.

[6] Kamath S. and Loizou P.(2002) "A multiband spectral subtraction method for enhancing speech corrupted by colored noise" *Proc .IEEE Intl. Conf. Acoustics, Speech, Signal Processing*

[7]Wang, D., Lim, J., 1982. The unimportance of phase in speech enhancement. IEEE Trans. Acoust. Speech Signal Process. 30 (ASSP-4), 679–681.

[8] Paliwal, K., Wo´ jcicki, K., 2008. Effect of analysis window duration on speech intelligibility. IEEE Signal Process. Lett. 15, 785–788.

[9]Yang Lu, Philipos C. Loizou,2008. A geometric approach to spectral subtraction. Speech Communication 50,453-466

[10]Kuldip Paliwal, Kamil Wo´jcicki , Belinda Schwerin, 2010. Single-channel speech enhancement using spectral subtraction in the short-time modulation domain. Speech communication 52,450-475

[11]Anuradha R. Fukane, Shashikant L. Sahare, 2011. Enhancement of Noisy Speech Signals for Hearing Aids**.** International Conference on Communication Systems and Network Technologies

[12] Evans, N., Mason, J., Liu, W., Fauve, B., 2006. An assessment on the fundamental limitations of spectral subtraction. In: Proc. IEEE Internat. Conf. on Acoustics, Speech, Signal Processing, Vol. I. pp. 145–148.

[13] Hu, Y., Loizou, P., 2007. Subjective evaluation and comparison of speech enhancement algorithms. Speech Communication 49, 588– 601.

[14] Falk, T., Stadler, S., Kleijn, W.B., Chan, W.-Y., 2007. Noise suppression based on extending a speech-

dominated modulation band. In: Proc. ISCA Conf. Internat. Speech Commun. Assoc. (INTERSPEECH). Antwerp, Belgium, pp. 970–973.

[15] Kim, D., 2004. A cue for objective speech quality estimation in temporal envelope representations. IEEE Signal Process. Lett. 11 (10), 849–852.

[16] J. D. Johnston, "Transform coding of audio signal using perceptual noise criteria," *IEEE J. Select. Areas Commun.*, vol. 6, pp. 314–323, Feb. 1988

[17] N. Virag, "Speech enhancement based on masking properties of the auditory system," in *Proc. IEEE ICASSP*, Detroit, MI, May 1995, pp. 796–799.

[18] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models.* Berlin, Germany: Springer-Verlag, 1990.

[19] M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *J. Acoust. Soc. Amer.*, vol. 66, pp. 1647–1652, Dec. 1979.

[20] D. Sinha and A. H. Tewfik, "Low bit rate transparent audio compression using adapted wavelets," *IEEE Trans. Signal Processing*, vol. 41, pp.3463–3479, Dec. 1993.

[21] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and projection, for robust recognition in cars," *Speech Commun.*, vol. 11, pp. 215–228, June 1992.