A Modified Perceptual Constrained Spectral Weighting Technique For Speech Enhancement

Gowder Praveena Hiriyan[1], **A.Indhumathi**[2]

[1](M.Phil Scholar , SNS Rajalakshmi College Of Arts and Science/ Bharathiar University, India)

[2](Asst.professor, MCA Department, SNS Rajalakshmi College Of Arts and Science/ Bharathiar University,India)

ABSTRACT: *Presently Speech Communication becomes active area in signal processing. Many approaches are developed previously for enhancing of speech. Perceptual speech enhancement methods perform better than the non perceptual methods, but most of them still return annoying residual musical noise. When noise above the noise masking threshold is filtered then noise below the noise masking threshold can become audible if its maskers are filtered is the main reason for residual noise. This affect the performance of perceptual speech enhancement method that process audible noise only. To overcome this drawback here proposed a new speech enhancement technique by modifying the Perceptual Wiener filter. The simulation results shows that the performance of this method which is improved when compared to other perceptual speech enhancement methods.*

Keywords *- Signal to Noise Ratio (SNR), Perceptual Speech Enhancement, Perceptual Wiener filter (PWF), Wiener Filter, Perceptual Evaluation of Speech Quality Measure.*

I. INTRODUCTION

Improving the performance of speech communication systems became an interesting area in signal processing. It is applied to improve the quality and intelligibility of speech in noisy environments. The problem has been widely discussed over the years. Many approaches have been proposed (Ephraim and Malah (1984), R. Schwartz et al (1979), Virag (1999) and Ephraim and Van (1995)). The enhancement process aims to improve the speeches over all quality; to increase the speech intelligibility in order to reduce the listener fatigue, ambiguity etc depending on specific application.

To address the three issues in speech enhancement objectives (Ephraim, 1992): (a) the improvement of the perceptual quality of noisy speech, (b) the immunization of speech encoders against input noise(Atal and Gibson et al (1991)), and (c) the improvement of the performance of speech recognition systems in the presence of noise(Vaseghi , Milner and Logan, Robinson ,1997).

Speech enhancement has applications in a wide variety of speech communication contexts where the quality or the intelligibility of speech has been degraded by the presence of background noise. For example, cellular radio telephone systems are plagued not only by background noise but also by channel noise. Public telephones suffer from environmental disturbances of their location. Air-ground communication systems are corrupted with cockpit noise. Moreover the hearing impaired needs a rise of between 2.5 and 12 dB signal-to-noise ratio to achieve similar speech discrimination capabilities to those of normal hearing (Sameti, 1994). These problems call for the use of speech enhancement.

Wiener formulated the continuous-time, least mean square error, estimation issues in his classic work on interpolation extrapolation and smoothing some of time series. The extension of the Wiener theory from continuous time to discrete time is simple, and of more practical use for implementation on digital signal processors.

The spectral subtraction technique (Boll 1979) is one of the most effective for our situation. It operates by making an estimate of the spectral magnitude during periods of no speech and subtracting this spectral estimate of the noise is from the subsequent speech spectral magnitude.

Speech enhancement algorithms introduce different distortion and distortion affecting the speech signal itself and distortion affecting the background noise. Consequently, just some of the objective methodologies offers affordable results. Studies have been conducted to evaluate suitability of different objective methods for the quality assessment of speech enhanced by speech enhancement algorithm.

II. LITERATURE SURVEY

Several algorithms have been studied in the past decennium to combine noise reduction with the preservation of binaural localization cues. First, Wittkop and Hohmann (2003) proposed a method based on computational auditory scene analysis in which the input signal is split into different frequency bands. By comparing the estimated binaural properties, such as the coherence, of each frequency band with the expected properties of the signal component (typically it is assumed that the signal component arrives from the frontal area with ITD and ILD values close to 0 μs and 0 dB), these frequencies are either enhanced or attenuated. By applying identical gains to the left and the right hearing aid, binaural cues should be preserved. However, spectral enhancement. Artifacts such as "musical noise" will typically occur. Moreover, localization performance when using this technique was never evaluated .

HMMs have long been used as a reliable statistical model for speech as it can model the nom stationary nature of speech by transition between different states. A large number of states can be used to represent different spectral prototypes of speech. As mentioned earlier a state dependent probability density can be chosen to be a mixture of Gaussian probability densities. An advantage of such representations is that we get finer models of speech data (Ephraim 1992). In the case of speech recognition, a separate left-right model is used to characterize the temporal structure of every speech unit which may be a phoneme or a word (Wittkop and Hohmann 2003).

In general, the goal of the Wiener filter is to filter noise corrupting a desired signal. By exploitation the second-order statistical properties of the specified speech signal and the noise, the optimal filter or Wiener filter will be calculated. It generates associate degree output signal that estimates the desired signal in a very minimum mean square error sense. In contrast with a standard beamformer, it can do so without any prior assumption on the angle of arrival of the signal. In (Doclo and Moonen 2002), it was shown that a MWF can be used for monaural hearing aid applications. Later on, this approach was extended to a two-channel hearing aid configuration in which one or more contra lateral microphone signals can be added. One amongst the most advantages of a MWF is that it inherently preserves the interaural cues of the estimated speech component. This was mathematically proven in the work of (Doclo et al 2006). However, it was also proven that the interaural cues of the noise component are distorted into those of the speech component.

The Minimum Mean Square Error (MMSE) filter by (Ephraim and Malah 1984) is an important milestone. In these estimation type approaches, the changed coefficients are filtered in each short-time frame and attenuated independently of their intra-frame neighboring coefficients as well as inter-frame neighboring coefficients. Though, some correlation does exist among completely different time frames and this is often exploited by many researchers to some extent (Boll 1979, Soon and Koh 2003). This is apparent in many recent works which view speech as a 2D time–frequency signal, especially in the form of a spectrogram. Evans has applied morphological filtering on the spectrogram in (Evans et al 2002) using opening operator based on erosion and dilation which is borrowed from digital image processing tools, and has obtained improved results. However, this algorithm emphasizes more on 2D processing without exploiting the characteristics of the speech spectrogram, resulting in the attenuation of the speech content together with the noise.

III. METHODOLOGY

3.1 Speech Perception

Speech perception is the process by which the sounds of language are heard, interpreted and understood read betweeen lines. The study of speech perception is closely linked to the fields of acoustics and in linguistics and cognitive psychology and perception in psychology. Research in speech perception seeks to grasp however human listeners recognize speech sounds and use this information to grasp speech. Speech perception research has applications in building laptop systems that can recognize speech, in betterment of speech recognition for hearing and languages-impaired listeners, and in foreign-language teaching.

The human perception is quite different from the way many computer programs work. We do not analyze the sound range in complicated mathematical ways. The brain is able to, in a very successful way, distinguish interesting sound from noise. This noise could be of many kinds, it could be anything from noise of a big engine to a man speaking Danish. Almost any kind of noise can be nearly ignored by our brain, enabling us to perceive the important information. Briefly it is recognized that speech can be represented by finite number of sounds called phonemes.

The mechanism in the brain that detects the acoustic features leading to the meaning of the message. One approach is high level motor theory model of speech perception which states that acoustic feature map back to articulatory features. For example formants of vowel are detected or estimated and then unconsciously interpreted as place of articulation used in the vowel production.

The research and application of speech perception must deal with several problems which result from what has been termed the need of invariance. As was instructed above, reliable constant relations between a phone of a language and its acoustic manifestation in speech are difficult to find. There are various reasons for this:

Context-induced variation. Phonetic environment affects the acoustic properties of speech sounds. For example, /u/ in English is fronted once surrounded by coronal constants.[6] Or, the VOT values marking the boundary between voiced and voiceless plosives are distinct for labial, alveolar and velar plosives and they shift under stress or depending on the position within a syllable.[7]

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

3.2 Standard Speech Enhancement Technique

Let the noisy signal can be shown as above equation

Where $s(n)$ is that the original clean speech signal and $d(n)$ is the additive random noise signal, unrelated with the original signal.

$$Y(m,k) = S(m,k) + D(m,k) \qquad (2)$$

Taking DFT to the observed signal gives

Where $m = 1,2,\ldots,M$ is the frame index, $k = 1,2,\ldots\ldots,K$ is the frequency bin index, M is the total number of frames and K is the frame length, $Y(m,k), S(m,k)$ and $D(m,k)$ represent the short time spectral components of the $y(n), S(n)$ and (n) , respectively. Clean speech spectrum $\hat{S}(m,k)$ is obtained by multiplying noisy speech spectrum with filter gain function as given in equation (3)

$$\hat{S}(m,k) = \mathrm{H}(\mathrm{m},\mathrm{k})\mathrm{Y}(\mathrm{m},\mathrm{k}) \qquad (3)$$

Where $H(m,k)$ is the noise suppression filter gain function (conventional Wiener Filter (WF)), which is derived according to MMSE estimator and $H(m,k)$ is given by

$$H_1(m,k) = \frac{\xi(\mathrm{m},\mathrm{k})}{1 + \xi(\mathrm{m},\mathrm{k})} \qquad (4)$$

Where $\xi(\mathrm{m},\mathrm{k})$ is an apriori SNR, which is defined as

$$\xi(m,k) = \frac{\Gamma_s(m,k)}{\Gamma_d(m,k)}$$

$$\Gamma_d(\mathrm{m},\mathrm{k}) = \mathrm{E}\,\{|\,D(m,k)\,|^2\} \qquad (5)$$

$$\Gamma_s(\mathrm{m},\mathrm{k}) = \mathrm{E}\,\{|\,S(m,k)\,|^2\}$$

represents the estimated noise power spectrum and clean speech power spectrum, respectively. A posteriori estimation is given by

An estimate of $\hat{\xi}(\mathrm{m},\mathrm{k})$ of $\xi(\mathrm{m},\mathrm{k})$is given by the well known decision directed approach [9] and is expressed as

$$\gamma(m,k) = \frac{|\,Y(m,k)\,|^2}{\Gamma_d(\mathrm{m},\mathrm{k})} \qquad (6)$$

$$\hat{\xi}(\mathrm{m},\mathrm{k}) = \propto \frac{|\mathrm{H}(\mathrm{m}-1,\mathrm{k})\mathrm{Y}(\mathrm{m}-1,\mathrm{k})|^2}{\Gamma_d} + (1-\propto)\mathrm{P}[\mathrm{V}(\mathrm{m},\mathrm{k})] \qquad (7)$$

3.3.1 Gain of Modified Perceptual Wiener filter (MPWF)

The Modified perceptual Wiener filter (M PWF) gain function $H_1(m,k)$is calculated primarily based cost function, J which is stated as

$$J = \left[\left|\hat{S}(m,k) - S(m,k)\right|^2 \right] \qquad (8)$$

Substituting (3.2) and (3.3) in (3.9) results to

$$= E\{|\,(H_1(m,k) - 1)S(m,k) + H_1(m,k)D(m,k)|^2\} \qquad (9)$$

$$= d_i + \eta_i$$

Where

$d_i = (H_1(m,k) - 1)^2 E\,\lfloor|S(m,k)|^2\rfloor$ and

$\eta_i = H_i^2(m,k)E\,\lfloor|D(m,k)|^2\rfloor$

where r_i represents speech distortion energy and residual noise energy

To make this residual noise voiceless, the residual noise should be less than the auditory masking threshold, $\mathrm{T}(\mathrm{m},\mathrm{k})$. This constraint is given by

$$\eta_i \leq T(m,k) \qquad (10)$$

By including the above constraints and substituting

$\Gamma_d(\mathrm{m},\mathrm{k}) = \mathrm{E}\,\{|D(m,k)|^2\}$ and

$\Gamma_s(\mathrm{m},\mathrm{k}) = \mathrm{E}\,\{|S(m,k)|^2\}$

The generalized Wiener filter referred to in $\mathrm{d_i}$ and $\mathrm{r_i}$ results if we minimize the modified error criterion

$$J = (H_1(m, k) - 1)^2T_S(m, k)+H_i^2(m, k)\{max[T_d(m, k)-T(m, k)),0]\} \qquad (11)$$

$$J_i^{GW} = d_i + \mu\eta_i$$

Where μ is an arbitrary constant that allows a trade-off between signal distortion and noise: if μ is large the noise is reduced, but there is greater signal distortion.

The desired perceptual modification of the generalized Wiener filter is obtained by modifying the criterion further to allow for the auditory masking phenomenon:

Where 'η' is an arbitrary parameter that adds another degree of freedom to the solution. It is usually chosen to be less than 1.

$$J_i^{MPGM} = (H_1(m,k) - 1)^2\Gamma_s(m,k) + \mu H_i^2(m,k)\,\{\max[(\Gamma_d(m,k) - \eta T(m,k)),0]\} \qquad (12)$$

The noise is included in this perceptual criterion only if it exceeds the masking threshold (as modified by q). Furthermore, the noise is weighted into the criterion only by the amount that it actually exceeds this threshold.

The obtained perceptually stated Wiener filter gain function is given by

$$\mathrm{H_1(m,k)} = \frac{\Gamma_s(\mathrm{m},\mathrm{k})}{\Gamma_s(\mathrm{m},\mathrm{k}) + \max[(\Gamma_d(\mathrm{m},\mathrm{k}) - \eta\mathrm{T}(\mathrm{m},\mathrm{k})),0]} \qquad (13)$$

By multiplying and dividing equation with $\Gamma_d(\mathrm{m},\mathrm{k})$, $\mathrm{H_1(m,k)}$ will become as

$\mathrm{T}(\mathrm{m},\mathrm{k})$ is noise masking threshold which is estimated based on[56] noisy speech spectrum. A priori SNR and noise power spectrum were estimated using the two-step a priori SNR estimator proposed and weighted noise estimation method proposed respectively.

$$H_1(m,k) = \frac{\xi(\mathrm{m},\mathrm{k})}{\xi(\mathrm{m},\mathrm{k}) + \frac{\max[(\Gamma_d(m,k) - \eta T(m,k)),0]}{\Gamma_d(\mathrm{m},\mathrm{k})}} \qquad (14)$$

3.3.2 Weighted Pwf

Although perceptual speech enhancement methods perform better than the non-perceptual methods, most of them still return annoying residual noise. Enhanced speech signal obtained using above mentioned perceptual Wiener filter still contains some residual noise due to the fact that only noise above the noise masking threshold is filtered and noise below the noise masking threshold is remain. It can affect the performance of perceptual speech enhancement method that processes audible noise only. In order to overcome this drawback we propose to weight the perceptual Wiener filters using a psychoacoustically inspired weighting filter. Psychoacoustically inspired weighting filter is given by

$$W(m,k) = \begin{cases} H(m,k), if\ ATH\ (m,k) < \Gamma_d \leq T(m,k) \\ 1, otherwise \end{cases} \tag{15}$$

Where $ATH(m,k)$ is the perfect threshold of hearing. This coefficient factor is used to weight the perceptual wiener filter. The gain function of the $H_2(m,k)$of the proposed weighted perceptual Wiener filter is given by Using the equation d_i and r_i of that the speech distortion d_i is always smaller than achieved with the Wiener solution (is. if masking is not allowed for). Similarly, the noise residual r_i is always larger than with the Wiener solution, but the difference will be less audible due to masking.

$$H_2 = H_1(m,k)W(m,k) \tag{16}$$

3.3 perceptual speech enhancement

Although the Wiener filtering reduces the level of musical noise, it does not eliminate it. Musical noise exists and perceptually annoying. In an effort to make the residual noise perceptually inaudible, many perceptual speech enhancement methods have been proposed which incorporates the auditory masking properties. In these methods residual noise is shaped according to an estimate o the signal masking threshold. Figure 1 depicts the complete block diagram of the proposed speech enhancement method.

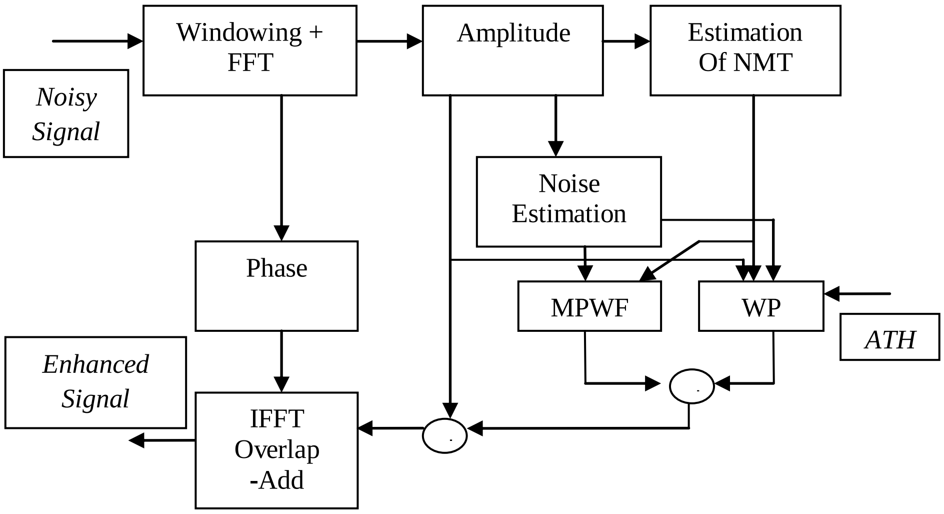

Figure 1. Block diagram of the proposed speech Enhancement method

IV. EXPERIMENTAL RESULT

To stimulate this proposed approach and compare the performance of the proposed scheme of speech enhancement, simulations are conceded with the NOIZEUS, A noisy speech collection for evaluation of speech enhancement algorithms, database (Yi and Loizou 2008). The noisy database contains 30 IEEE sentences (produced by three female and three male speakers) corrupted by eight different real world noises at different SNRs.

Speech signals were degraded with different types of noise at global SNR levels of 0(zero) dB, 5(five) dB, 10(ten) dB and 15(fifteen) dB. In this assessment only five noises are considered those are babble, car, train, airport and street noise. The objective quality measures used for the evaluation of the proposed speech enhancement method are the segmental SNR and perceptual evaluation of speech quality (PESQ) measures (Muni Kumar et al 2012). It is well known that the segmental SNR is more accurate in indicating the speech distortion than the overall SNR. The top value of the segmental SNR indicates the weaker speech distortion. The higher PESQ score indicates better perceived quality of the proposed signal.

The performance of the projected method is compared with existing method (Muni Kumar et al 2012) and perceptual Wiener filter.

Experimental results are taken using MATLAB. MATLAB (Matrix Laboratory) is used for the computation of the numerical analysis and is considered as a fourth-generation programming language. It is a viable Matrix Laboratory package which functions as an interactive programming environment.

Hence, for the present research, MATLAB has been taken into consideration and the techniques have been implemented by using MATLAB normal.

4.1 Simulation results

The resulted values of proposed MPWF are displayed in table1. This table compares the result of proposed method with existing PWF and traditional PWF at global SNR levels of 0(zero) dB, 5(five) dB, 10(ten) dB and 15 (fifteen)dB.

TABLE 3.1

SEGMENTAL SNR VALUES OF ENHANCED SPEECH SIGNALS

Noise Type	Input SNR(dB)	PWF	Existing PWF	Proposed MPWF
Babble	0	-0.16	0.22	0.28
	5	0.01	0.32	0.39
	10	0.65	2.14	2.19
	15	2.17	3.97	4.08
Car	0	-0.24	0.85	1.08
	5	0.52	1.2	1.67
	10	0.7	2.37	2.87
	15	2.13	3.81	3.89
Train	0	-0.49	0.15	0.42
	5	0.38	0.43	0.67
	10	0.77	2.2	2.52
	15	2.62	3.5	3.9
Airport	0	-0.24	0.19	0.24
	5	0.15	0.43	0.78
	10	0.14	1.09	1.19
	15	1.88	3.65	3.84
Street	0	-0.15	0.08	0.54
	5	0.61	0.73	0.98
	10	1.2	2.7	2.82
	15	2.25	3.42	3.69

From the table we can see that the proposed modified Perceptual Wiener filter produces better SNR values for all the input SNR signals when compared with existing method and traditional method. It produces better result for babble, car, train, airport and street noise.

TABLE 3.2

PESQ VALUES OF THE ENHANCED SIGNALS

Noise Type	Input SNR(dB)	PWF	Existing PWF	Proposed MPWF
Babble	0	0.95	1.42	1.54
	5	1.75	1.83	1.9
	10	2.27	2.4	2.45
	15	2.6	2.71	2.76
Car	0	1.43	1.73	1.79
	5	1.69	2.1	2.19
	10	2.16	2.31	2.42
	15	2.64	3.12	3.21
Train	0	1.48	1.73	1.82
	5	1.71	2.13	2.19
	10	2.09	2.47	2.54
	15	2.03	2.71	2.79
Airport	0	1.56	1.75	1.82
	5	1.76	2.24	2.31
	10	2.41	2.53	2.62
	15	2.57	2.71	2.78
Street	0	1.78	1.81	1.89
	5	1.85	1.96	2.06
	10	2.26	2.39	2.48
	15	2.57	2.68	2.71

The proposed PESQ measures were better than the conventional measures for all 5 types of noises when compared with existing PWF and traditional PWF.

The simulation results are summarized in Table 1 and Table 2. The proposed method leads to better denoising quality for temporal and the better improvements are obtained for the high noise level. The time-frequency distribution of speech signals provides more accurate information about the residual noise and speech distortion than the corresponding time domain wave forms. We compared the spectrograms for each of the method and confirmed a reduction of the residual noise and speech distortion.

V. CONCLUSION

A method of speech enhancement for suppressing musical noise is carried here. Based on the perceptual properties of the human auditory system, a weighting factor accentuates the denoising process when noise is perceptually insignificant and prevents that residual noise components might become audible in the absence of adjacent maskers. In this work modified Perceptual Wiener filter method is introduced. The unique feature of this method is that the subband gain calculation exploits the masking properties. Enhanced speech of good perceptual quality is obtained in both coloured and non-stationary noise. The performance of this approach is measured using performance metrics such as Segmental SNR and PESQ. And for evaluation five noises are considered those are babble, car, train, airport and street noise by these metrics result is obtained which conforms that this Modified PWF is efficiently suits for enhancing speech under noise environment also. There are many parameters available for further investigation. That can be implementing with fine tuning this proposed approach for future enhancement of this approach.

REFERENCES

[1] . Ephraim Y and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Processing,vol. ASSP-32, pp. 1109– 1121, Dec 1984.

[2] Schwartz R. M. Berouti and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," Proc. of ICASSP, 1979, vol. I, pp. 208–211

[3] Virag N, "Single channel speech enhancement based on masking properties of the human auditory system," IEEE Trans. Speech and Audio Processing, vol. 7, pp. 126–137, 1999.

[4] Ephraim Y and H.L. Van Trees, "A signal subspace approach for speech enhancement," IEEE Trans. Speech and Audio Processing, vol. 3, pp. 251–266, 1995

[5] Ephraim Y. "Statistical model based speech enhancement systems," Proc. IEEE, vol. 80, pp. 1526-1555, Oct. 1992.

[6] Atal B.S., Cuperman V and A.Gersho. Advances in Speech Coding. Kluwer Academic Publishers, 1991.

[7] Vaseghi S.V and B.P. Milner. "Noise Compensation Methods for Hidden Markov Model Speech Recognition in Adverse Environments", IEEE Transactions on Speech and Audio Signal Processing, 5(1):11-21, Jan. 1997.

[8] H-Sameti. Model-Based Approaches to Speech Enhancement: Stationary State and Nonstationary-State HMMs. PhD thesis, University of Waterloo, Department of Electrical Engineering, 1994.

[9] Boll S.F. (1979). "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. Acoust. Speech Signal Process. 27, 113–120.

[10] Wittkop T and Hohmann V (2003) "Strategy selective noise reduction for binaural digital hearing aids," Speech Commun. 39, 111–138

[11] Ephraim Y. "Statistical model based speech enhancement systems," Proc. IEEE, vol. 80, pp. 1526-1555, Oct. 1992.

[12] Doclo S and Moonen M. (2002). "GSVD-based optimal filtering forsingle and multi-microphone speech enhancement," IEEE Trans. Signal Process. 50, 2230–2244.

[13] Doclo S, Klasen T. J, Van den Bogaert T, Wouters J and Moonen M (2006). "Theoretical analysis of binaural cue preservation using multichannel Wiener filtering and interaural transfer functions," in Proceeding International Workshop on Acoustic Echo and Noise Control _IWAENC_, Paris, France, pp. 1–4.

[14] Evans N.W.D, Mason J.S, Roach M.J. (2002). "Noise compensation using spectrogram morphological filtering", In: Proc. 4th IASTED Internat. Conf. Signal Image Process, pp. 157–161.

[15] Yi Hu and Philipos C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement," IEEE Trans. on Audio, Speech and Language Processing, vol. 16, no. 1, pp. 229- 238, January 2008.

[16] Muni Kumar T, M.B.Rama Murthy , Ch.V.Rama Rao , K.Srinivasa Rao,' A New Speech Enhancement Technique Using Perceptual Constrained Spectral Weighting Factors', International Journal of Electronics Signals and Systems (IJESS), ISSN No. 2231- 5969, Volume-1, Issue-2, 2012