A Modernized Speech Enrichment Method using Intuitive Weighting Factor

M. Venkata Rao
Assistant Professor
Department of ECE
ASIST, Paritala, A.P.

G. Charan Teja
B.Tech. student
Department of ECE
ASIST, Paritala, A.P.

M. Gopi Chand
B.Tech. student
Department of ECE
ASIST, Paritala, A.P.

D. Lakshmi Sudha
B.Tech. student
Department of ECE
ASIST, Paritala, A.P.

Ch. V. L. Suvarchala
B.Tech. student
Department of ECE
ASIST, Paritala, A.P.

K. Anvesh
B.Tech. student
Department of ECE
ASIST, Paritala, A.P.

Abstarct:- **This paper deals with acoustic noise upshot from intuitive speech enrichment type methods and especially wiener filtering. Even if intuitive speech enrichment methods act prominent than the non-intuitive methods, most of them still recrudescence annoying vestigial acoustic noise. This is due to the fact that if only noise above the noise masking threshold is filtered then noise below the noise masking threshold can become hearable if its maskers are filtered. It can disturb the execution of emotional speech improvement method that process hearable noise only. In order to overthrown this difficulty here schemed a new speech enrichment approach. Its ambition is to shape up the peculiarity of the enhanced speech signal supplied by intuitive wiener filtering by ruling the latter via a second filter taken into consideration as a psycho acoustically instigated weighting determinant. The simulation demonstrates that the accomplishment is made better distinguished to other intuitive speech Enhancement methods. Here we are completely take off background noise from speech signal. It is used in mobile applications.**

Keywords: Acousticnoise, intuitive weighting factor, peculiarity, residual noise.

INTRODUCTION:

The fair of speech enrichment process is to sharpen the feature and clearness of speech in a noise atmosphere. Many ways are proposed like subtractive type [1-4], Perceptual Wiener filtering algorithms. In these spectral subtraction and Wiener filtering algorithms are broadly used because these are having low computational complication and powerful enforcement. In these algorithms, such methods again contain residual noise known as acoustical noise. This type of noise is quite causing irritation. For decreasing the effect of musical noise, several solutions have been proposed. Some involve rectifying frameworks of spectral subtraction so as to offer more flexibility as in [2] and [3].Other such as proposed in

[4], are based on signal subspace approaches. Despite the forcefulness of these approaches to improve the signal to noise ratio (SNR), the problem of eliminating acoustic noise is still a challenge to many analyzers.. In the last few decades the interpolation of psychoacoustic draws attention an extensive bargain of interest. The main aim is to increase the intuitive characteristic of the improved signal. In [3], a psychoacoustic model is used to control the parameters of the spectral subtraction in order to find the best trade of between noise reduction and speech distortion. For musical noise in hearable, the uninterrupted estimator proposed in [5] includes the masking substances of the human auditory system. In the covering threshold and transitional signal, which is slightly de noised and free of musical noise are used to detect tuneful tones generated by the spectral subtraction methods. This detection can be used by a post processing aimed at reducing the detected tones. These intuitive speech enrichment systems reduced the noise but introduce some unwanted inaccurate to the enrichment signal. When this inaccurate estimated speech signal is applied to the acknowledged systems their performance degrades extremely.

The basic idea of the recommended method is to eradicate, intuitively significant noise modules from the noisy signal, so that the unsoiled speech modules are not affected by processing. In addition, the technique requires very little a priori information of the features of the noise. In the present paper, we recommended to control the intuitive wiener filtering by psycho acoustically inspired filter that can be regarded as weighting factor. The purpose is to minimize the perception of musical noise without degrading the clarity of the enriched speech.

2. STANDARD SPEECH ENRICHMENT METHOD

Let the noisy signal can be expressed as

$$y(n) = s(n) + d(n) , \quad\quad (1)$$

where $x(n)$ is the original clean speech signal and $d(n)$

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCACSPV - 2016 Conference Proceedings

is the additive random noise signal, uncorrelated with the original signal. Taking DFT to the observed signal gives

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

Where $m = 1,2,...,M$ is the frame index, $k = 1,2,....,K$ 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 modules 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)

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

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

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

Where $\xi(m,k)$ is an priory SNR, which is defined as

$$\xi(m,k) = \frac{\Gamma_s(m,k)}{\Gamma_d(m,k)}. \qquad (5)$$

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

$\Gamma_s(m,k) = E\left\{|S(m,k)|^2\right\}$ represents the estimated noise power spectrum and clean speech power spectrum, respectively. A posteriori estimation is given by

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

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

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

Where $V(m,k) = \gamma(m,k) - 1$, $P[x] = x$ if $x \geq 0$ and $P[x] = 0$ otherwise.

The noise suppression gain function is chosen as the Wiener filter similar to [13]

3. PERCEPTUAL SPEECH ENHANCEMENT

Although the Weiner filtering reduces the level of acoustic noise, it does not eliminate it [15]. Musical noise exists and intuitively annoying. In an effort to make the residual noise intuitively inaudible, many intuitive speech enrichment methods have been proposed which integrates the auditory masking properties [2-9]. In these methods residual noise is shaped according to an estimate of the signal masking threshold [9, 13]. Figure 1 describes the complete block diagram of the proposed speech enrichment technique
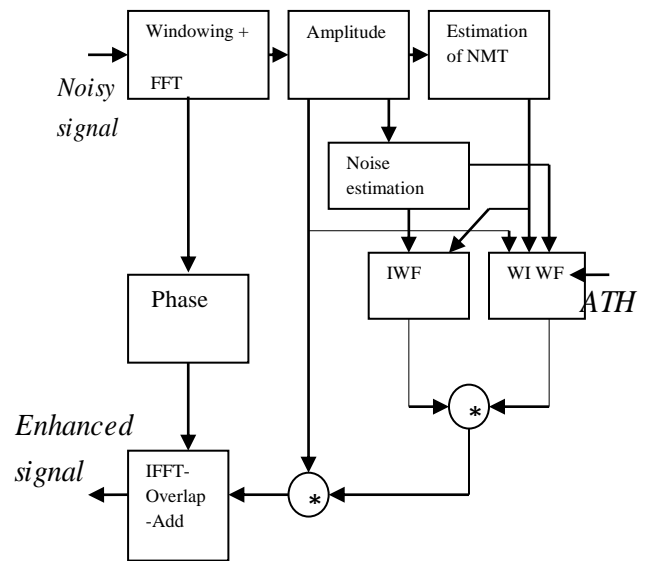

Figure1. Block diagram of the proposed speech enrichment method

3.1 Gain of Intuitive Wiener filter (PWF)

The Intuitive Wiener filter (PWF) gain function $H_1(m,k)$ is calculated based cost function, J which is defined as

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

Substituting (2) and (3) in (9) results to

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

$$= d_i + r_i \qquad (9)$$

Where

$$d_i = (H_1(m,k) - 1)^2 E\left[|S(m,k)|^2\right] \qquad \text{And}$$

$$r_i = H_1^2(m,k)E\left[|D(m,k)|^2\right] \qquad \text{represents speech}$$

distortion energy and residual noise energy.

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCACSPV - 2016 Conference Proceedings

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

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

By including the above constraint and substituting $\Gamma_d(m,k) = E\{|D(m,k)|^2\}$ and $\Gamma_s(m,k) = E\{|S(m,k)|^2\}$ in (9) the cost function will become as

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

The desired intuitive modification of Wiener is obtained by differentiating J w.r.t $H_1(m,k)$ and equating to zero. The obtained perceptually defined Wiener filter gain function is given by

$$H_1(m,k) = \frac{\Gamma_s(m,k)}{\Gamma_s(m,k) + \max(\Gamma_d(m,k)-T(m,k),0)} \qquad (12)$$

By multiplying and dividing equation (12) with $\Gamma_d(m,k)$, $H_1(m,k)$ will become as

$$H_1(m,k) = \frac{\hat{\xi}(m,k)}{\hat{\xi}(m,k) + \frac{\max(\Gamma_d(m,k)-T(m,k),0)}{\Gamma_d(m,k)}} \qquad (13)$$

$T(m,k)$ is noise masking threshold which is estimated based on[16] noisy speech spectrum. A priori SNR and noise power spectrum were estimated using the two-step a priori SNR estimator proposed in [15] and weighted noise estimation method proposed in[17],respectively.

3.2 WEIGHTED PWF

Although intuitive speech enrichment methods perform better than the non-intuitive methods, most of them still return trying residual musical noise. Enriched speech signal obtained using above mentioned intuitive 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 act of intuitive speech enrichment method that processes audible noise only.

In order to overcome this drawback we propose to weight the intuitive Wiener filters using a psychoacoustic ally motivated weighting filter. Psychoacoustic ally motivated weighting filter is given by

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

Where $ATH(m,k)$ is the absolute threshold of hearing. This weighting factor is used to weight the intuitive wiener filter. The gain function of the $H_2(m,k)$ of the proposed weighted intuitive .Wiener filter is given by

$$H_2 = H_1(m,k)W(m,k) \qquad (16)$$

4. SIMULATION RESULTS

To evaluate and compare the performance of the proposed scheme of speech enrichment, simulations are carried out with the NOIZEUS, A noisy speech amount for estimation of speech enrichment systems, database [18]. The noisy database contains 30 IEEE sentences (produced by three male and three female 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 dB, 5 dB, 10 dB and 15 DB. In this estimation 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 enrichment method are the segmental SNR and PESQ measures [19]. It is well known that the segmental SNR is more accurate in indicating the speech distortion than the overall SNR. The higher value of the segmental SNR point out the weaker speech distortion. The higher PESQ score indicates better perceived quality of the proposed signal [19]. The performance of the proposed method is compared with Wiener filter and intuitive Wiener filter.

The simulation results are summarized in Table 1 and Table 2. The proposed method leads to better de noising quality for chronological 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. Figure2. Represents the spectrograms of the clean speech signal, noisy signal and enriched speech signals.

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCACSPV - 2016 Conference Proceedings

Table.1 Segmental SNR values of Enriched signals

Noise Type	Input SNR (dB)	WF	PWF	Proposed Method
Babble	0	1.231	1.114	1.527
	6	1.758	1.856	1.936
	12	2.056	2.298	2.502
	18	2.131	2.674	2.818
Car	0	1.171	1.493	1.834
	6	1.712	1.967	2.217
	12	2.121	2.681	2.418
	18	2.271	2.675	3.227
Train	0	1.512	1.582	1.831
	6	1.713	1.786	2.233
	12	2.158	2.221	2.579
	18	2.110	2.230	2.814
Airport	0	1.654	1.651	1.859
	6	1.924	1.967	2.342
	12	2.520	2.451	2.638
	18	2.429	2.759	2.815
Street	0	1.736	1.827	1.917
	6	1.691	1.901	1.968
	12	2.220	2.311	2.492
	18	2.481	2.735	2.783

Table.2 PESQ values of the enriched signals

Noise Type	Input SNR (dB)	WF	PWF	Proposed Method
Babble	0	-4.49	-0.56	0.18
	6	-1.29	0.04	0.39
	12	0.04	0.65	2.21
	18	0.85	2.75	4.11
Car	0	-3.83	-0.21	0.91
	6	-1.55	0.57	1.30
	12	0.72	0.77	2.11
	18	0.75	2.38	3.89
Train	0	-3.25	-0.46	0.25
	6	-0.76	0.40	0.47
	12	-0.29	0.81	2.29
	18	0.72	2.67	4.1
Airport	0	-4.31	-0.21	0.21
	6	-2.47	0.11	0.48
	12	-0.04	0.17	1.12
	18	0.67	1.91	3.70
Street	0	-2.78	-0.11	0.11
	6	-2.03	0.64	0.80
	12	0.75	131	2.76
	18	0.81	2.30	3.48

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
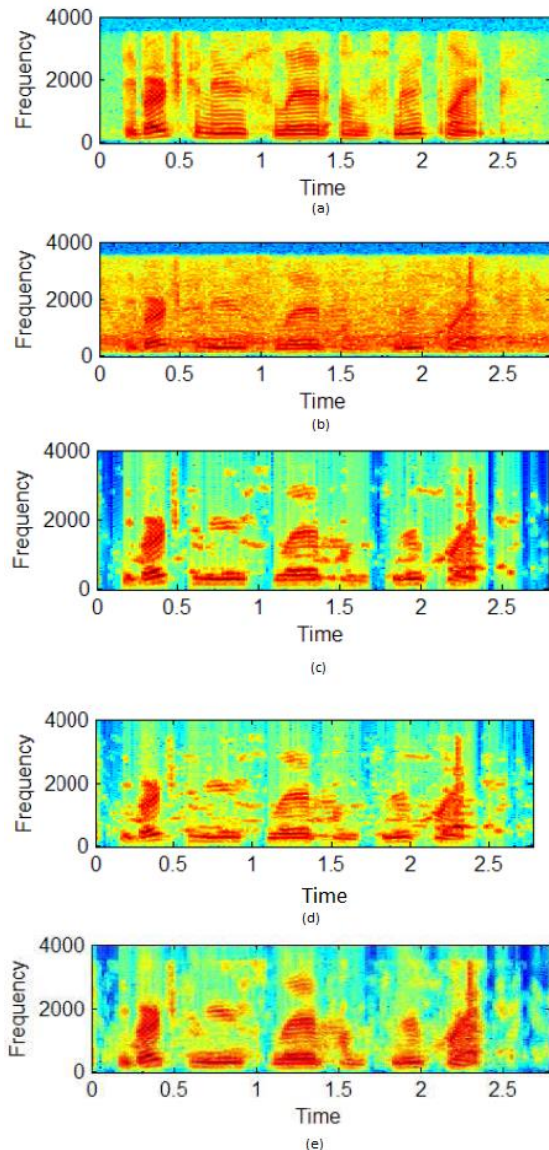**NCACSPV - 2016 Conference Proceedings**

Figure2. speech spectrogram,(a)original clean signal,(b) noisy signal(babble noise SNR=5dB),(c)enriched signal using Wiener filter(d)enriched signal using PWF,(e)enriched signal using Weighted IWF

5. CONCLUSION

In this paper, an effective approach for suppressing musical noise presented after wiener filtering has been introduced. Based on the intuitive properties of the human auditory system, a weighting factor emphasizes the de noising process when noise is intuitively insignificant and prevents that residual noise components might become audible in the absence of adjacent maskers. When the speech signal is additively corrupted by babble noise and car noise objective measure results showed the improvement brought by the recommended method in comparison to some recent filtering techniques of the same type.

6. REFERENCES

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

[2] Y. Hu and P. Loizou, "Incorporating a psychoacoustic model in frequency domain speech enhancement," *IEEE Signal Processing Letters*, vol. 11(2), pp. 270–273, 2004.

[3] T. Lee and Kaisheng Yao, "Speech enhancement by perceptual filter with sequential noise parameter estimation," Proc. of ICASSP, vol. I, pp. 693–696, 2004.

[4] Md. Jahangir Alam, Sid-Ahmed Selouani, Douglas O'Shaughnessy and S. Ben Jebara, "Speech enhancement using a Wiener denoising technique and musical noise reduction" in the Proceeding of *INTERSPEECH'08*, Brisbane, Australia, pp. 407-410, September 2008.

[5] Amehraye, D. Pastor, and A. Tamtaoui, "Perceptual improvement of Wiener filtering." Proc. of ICASSP, pp. 2081–2084, 2008.

[6] Md. Jahangir Alam, Douglas O'Shaughnessy and Sid-Ahmed Selouani, "Speech enhancement based on novel two-step *a priori* SN estimators," in the Proceeding of *INTERSPEECH'08*, Brisbane, Australia, pp. 565-568, September 2008.

[7] http://www.utdallas.edu/~loizou/speech/noizeus/

[8] Yi Hu and Philips 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.