M Venkatrao * et al.
(IJITR) INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND RESEARCH
Volume No.5, Issue No.6, October - November 2017, 7626-7630.

A New Speech Enhancement Technique Using Perceptual Wiener Filter

M.VENKATRAO
Asst.Prof, Department of ECE, Amrita Sai Institute of Science and Technology.

K. TIRUMALA RAO
Asst.Prof, Department of ECE, Amrita Sai Institute of Science and Technology.

G. SIVA KUMAR
Asst.Prof, Department of ECE, Amrita Sai Institute of Science and Technology.

Abstract- **This paper deals with musical noise result from perceptual speech enhancement type algorithms and especially wiener filtering. Although perceptual speech enhancement methods perform better than the non perceptual methods, most of them still return annoying residual musical 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 audible if its maskers are filtered. It can affect the performance of perceptual speech enhancement method that process audible noise only. In order to overcome this drawback here proposed a new speech enhancement technique. It aims to improve the quality of the enhanced speech signal provided by perceptual wiener filtering by controlling the latter via a second filter regarded as a psychoacoustically motivated weighting factor. The simulation results shows that the performance is improved compared to other perceptual speech enhancement methods**

I. INTRODUCTION

The objective of speech enhancement process is 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 like subtractive type [1-4], Perceptual Wiener filtering algorithms. Among them spectral subtraction and the Wiener filtering algorithms are widely used because of their low computational complexity and impressive performance. In these algorithms, Such methods return residual noise known as musical noise. This type of noise is quite annoying. In order to reduce the effect of musical noise, several solutions have been proposed. Some involve adjusting parameters 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 effectiveness of these techniques to improve the signal to noise ratio (SNR), the problem of eliminating or reducing musical noise is still a challenge to many researchers. In the last few decades the introduction of psychoacoustic models has attracted a great deal of interest. The objective is to improve the perceptual quality of the enhanced 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. To make musical noise inaudible, the linear estimator proposed in [5] incorporates the masking properties of the human auditory system. In [6], the masking threshold and intermediate signal, which is slightly denoised and free of musical noise, are used to detect musical tones generated by the spectral subtraction methods. This detection can be used by a post-processing aimed at reducing the detected tones. These perceptual speech enhancement systems reduce the musical noise but introduce some undesired distortion to the enhanced speech signal. When this distorted estimated speech signal is applied to the recognition systems their performance degrades drastically.

The basic idea of the proposed method is to remove, perceptually significant noise components from the noisy signal, so that the clean speech components 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 propose to control the perceptual wiener filtering by psychoacoustically motivated filter that can be regarded as weighting factor. The purpose is to minimize the perception of musical noise without degrading the clarity of the enhanced speech.

II. STANDARD SPEECH ENHANCEMENT TECHNIQUE

Let the noisy signal can be expressed as

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

Where $x(n)$ is the original clean speech signal and $d(n)$ 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 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 eqation (3)

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

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

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

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

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

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

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

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

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

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]

III. PERCEPTUAL SPEECH ENHANCEMENT

Although the Wiener filtering reduces the level of musical noise, it does not eliminate it [15]. 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 [2-9]. In these methods residual noise is shaped according to an estimate of the signal masking threshold [9, 13]. Figure 1 depicts the complete block diagram of the proposed speech enhancement method.

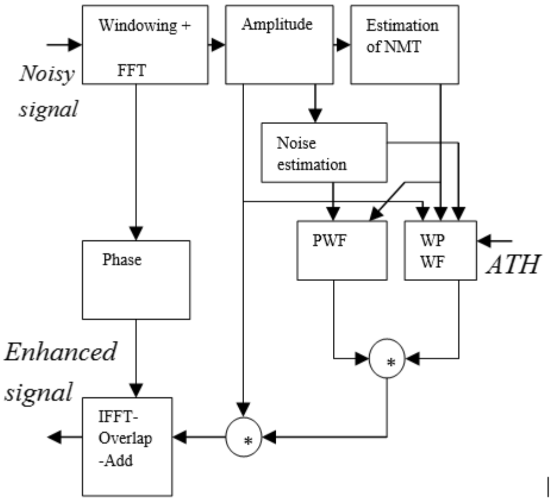

Figure1. Block diagram of the proposed speech enhancement method

3.1 Gain of Perceptual Wiener filter (PWF)

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

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

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

$$= E\left\{\left|(H_1(m,k)-1)S(m,k) + H_1(m,k)D(m,k)\right|^2\right\}$$
$$= d_i + r_i \quad (9)$$

Where

$d_i = (H_1(m,k)-1)^2 E\left[|S(m,k)|^2\right]$ and $r_i = H_1^2(m,k)E\left[|D(m,k)|^2\right]$ represents speech distortion energy and residual noise energy.

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

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

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

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

The desired perceptual 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

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

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

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

$T(m,k)$ 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 perceptual speech enhancement methods perform better than the non-perceptual methods, most of them still return annoying residual musical 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 motivated weighting filter. Psychoacoustically motivated weighting filter is given by

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

Where $ATH(m,k)$ is the absolute threshold of hearing. This weighting 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

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

IV. SIMULATION RESULTS

To evaluate and compare the performance of the proposed scheme of speech enhancement, simulations are carried out with the NOIZEUS, A noisy speech corpus for evaluation of speech enhancement algorithms, 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 evaluation 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 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 indicates 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 perceptual Wiener filter.

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. Figure2. Represents the spectrograms of the clean speech signal, noisy signal and enhanced speech signals.

Table.1 Segmental SNR values of Enhanced Signals

Noise Type	Input SNR (dB)	WF	PWF	Proposed method
Babble	0	-4.59	-0.61	0.22
	5	-1.39	0.01	0.32
	10	0.02	0.65	2.14
	15	0.75	2.71	3.97
Car	0	-3.93	-0.24	0.85
	5	-1.65	0.52	1.20
	10	0.69	0.70	2.37
	15	0.72	2.31	3.81
Train	0	-3.45	-0.49	0.15
	5	-0.86	0.38	0.43
	10	-0.39	0.77	2.20
	15	0.75	2.62	3.5

Airport	0	-4.37	-0.24	0.19
	5	-2.57	0.15	0.43
	10	-0.06	0.14	1.09
	15	0.75	1.88	3.65
Street	0	-2.88	-0.15	0.08
	5	-2.13	0.61	0.73
	10	0.69	1.20	2.70
	15	0.77	2.25	3.42

Table.2 PESQ values of the enhanced signals

Noise Type	Input SNR (dB)	WF	PWF	Proposed method
Babble	0	1.221	0.952	1.427
	5	1.728	1.750	1.836
	10	2.034	2.276	2.402
	15	2.127	2.609	2.718
Car	0	1.165	1.439	1.734
	5	1.694	1.697	2.107
	10	1.921	2.168	2.318
	15	2.265	2.645	3.127
Train	0	1.450	1.482	1.731
	5	1.680	1.715	2.133
	10	2.009	2.096	2.479
	15	2.040	2.032	2.714
Airport	0	1.472	1.561	1.759
	5	1.492	1.769	2.242
	10	2.025	2.413	2.538
	15	2.249	2.579	2.715
Street	0	1.636	1.782	1.817
	5	1.679	1.857	1.968
	10	2.119	2.260	2.392
	15	2.380	2.573	2.683

V. CONCLUSION

In this paper, an effective approach for suppressing musical noise presented after wiener filtering has been introduced. 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. When the speech signal is additively corrupted by babble noise and car noise objective measure results showed the improvement brought by the proposed method in comparison to some recent filtering techniques of the same type.

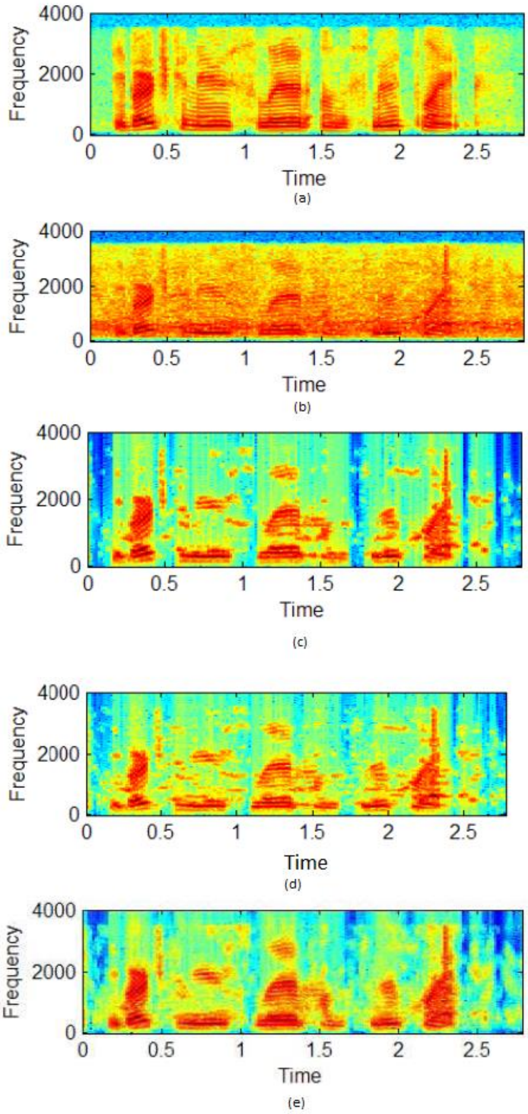

Figure2. speech spectrogram,(a)original clean signal,(b) noisy signal(babble noise SNR=5dB),(c)enhanced signal using Wiener filter(d)enhanced signal using PWF,(e)enhanced signal using Weighted PWF

VI. REFERENCES

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

[3] N.Virag, "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] Y. Ephraim 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] 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.

[6] F. Jabloun and B. Champagne, "Incorporating the human hearing properties in the signal subspace approach for speech enhancement," *IEEE Trans. Speech and Audio Processing*,vol. 11, pp. 700–708, 2003.

[7] Y.M. Cheng and D. O'Shaughnessy, "Speech enhancement based conceptually on auditory evidence," IEEE Trans. Signal Processing, vol.39, no.9, pp.1943–1954, 1991.

[8] D. Tsoukalas, M. Paraskevas, and J. Mourjopoulos, "Speech enhancement using psychoacoustic criteria," IEEE ICASSP, pp.359–362, Minneapolis, MN, 1993.

[9] Y. Hu and P.C. Loizou, "A perceptually motivated approach for speech enhancement," *IEEE Trans. Speech Audio Processing*, pp. 457-465. Sept. 2003.

[10] L. Lin, W. H. Holmes and E. Ambikairajah, "Speech denoising using perceptual modification of Wiener filtering," *IEE Electronic Letters*, vol. 38, pp. 1486–1487, Nov 2002.

[11] P. Scalart C. Beaugeant, V. Turbin and A. Gilloire, "New optimal filtering approaches for hands-free telecommunication terminals, "*Signal Processing*, vol. 64 (15), pp. 33–47, Jan 1998.

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

[13] 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.

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

[15] 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.

[16] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," IEEE on Selected Areas in Comm., vol. 6, pp. 314–323, February1988.

[17] M. Kato, A. Sugiyama and M. Serizawa, "Noise suppression with high speech quality based on weighted noise estimation and MMSESTSA," IEICE Trans. Fundamentals, vol. E85-A, no.7, pp. 1710-1718, July 2002.

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

[19] 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.

AUTHOR's PROFILE

M.VENKATRAO, Asst.Prof, Department of ECE, Amrita Sai institute of science and technology

K.TIRUMALARAO, Asst.Prof, Department of ECE, Amrita Sai institute of science and technology

G.SIVA KUMAR, Asst.Prof, Department of ECE, Amrita Sai institute of science and technology

