Journal of Signal and Information Processing, **2011, 2, 117-124**
doi:10.4236/jsip.2011.22016 Published Online May 2011 (http://www.SciRP.org/journal/jsip)

Speech Enhancement Using Cross-Correlation Compensated Multi-Band Wiener Filter Combined with Harmonic Regeneration

Venkata Rama Rao[1], Rama Murthy[2], K. Srinivasa Rao[3]

[1]Deptment of ECE, Gudlavalleru Engineering College, Gudlavalleru, India; [2]Jayaprakash Narayan College of Engineering, Dharmapur, Mahabubnagar, India; [3]Principal, TRR College of Engineering, Pathancheru, India.
Email: chvramaraogec@gmail.com, mbrmurthy@gmail.com, principaltrr@gmail.com

Received December 6th, 2010; revised April 25th, 2011; accepted April 29th,2011.

ABSTRACT

The speech signal in general is corrupted by noise and the noise signal does not affect the speech signal uniformly over the entire spectrum. An improved Wiener filtering method is proposed in this paper for reducing background noise from speech signal in colored noise environments. In view of nonlinear variation of human ear sensibility in frequency spectrum, nonlinear multi-band Bark scale frequency spacing approach is used. The cross-correlation between the speech and noise signal is considered in the proposed method to reduce colored noise. To overcome harmonic distortion introduced in enhanced speech, in the proposed method regenerate the suppressed harmonics are regenerated. Objective and subjective tests were carried out to demonstrate improvement in the perceptual quality of speeches by the proposed technique.

Keywords*: *Speech Enhancement, Wiener Filter, Critical Band and Speech Harmonics

1. Introduction

In many speech communication systems, recognition of speech signal from a corrupted speech signal with background noise is a challenging task especially at low SNR (signal to noise ratio) values. Speech quality and intelligibility might significantly deteriorate in the presence of background noise, especially when the speech signal is subjected to In many speech communication systems, background noise in corrupted speech is a challenging task especially at low SNR (signal noise ratio) values. Speech quality and intelligibility might significantly deteriorate in the presence of background noise, especially when the speech signal is subject to subsequent processing, such as automatic speech recognition and speech coding. Due to use of automatic speech processing systems in a variety of real world applications, speech enhancement has become an important topic of research. Several speech enhancement systems are available in the literature [1-4]. The enhancement of noise corrupted speech signal can be done using the Wiener filtering technique [5,6], spectral subtraction method [7] or Kalman filtering technique. The power spectral subtraction and the Wiener filtering algorithms are widely used because of their low computational complexity and impressive performance.

In general, in these algorithms the enhanced speech spectrum is obtained by subtracting an estimated noise spectrum from noisy speech spectrum or by multiplying the noisy spectrum with a gain function. Let the noisy speech, clean speech and noise signals are denoted by $y(n)$, $x(n)$ and $d(n)$ respectively in time domain. If it is assumed that noise is additive, then $y(n)$ can be expressed as:

$$y(n) = x(n) + d(n) \tag{1}$$

applying the Fast Fourier transform (FFT) to (1),at the m^{th} frame and k^{th} frequency bin, $y(n)$ can be represented as:

$$Y(m,k) = X(m,k) + D(m,k) \tag{2}$$

where $Y(m,k)$, $X(m,k)$ and $D(m,k)$ are the noisy speech, clean speech and noise signals FFT coefficients. An estimate of the clean speech component denoted as $X(m,k)$ can be obtained by multiplying with filter gain function $W(m,k)$ as given in (3)

$$\hat{X}(m,k) = W(m,k) \cdot Y(m,k) \tag{3}$$

The phase of the noisy speech is kept unchanged since it is assumed that the phase distortion is not perceived by the human ear. It is well-known the frequency resolution of human's hearing is non-uniform and usually described by critical bands or bark scale. The real-world noise does not affect the speech signal uniformly over the whole spectrum therefore; multiplying with a constant factor of noise spectrum over the whole range may remove speech also.

A new multi-band approach to the Wiener filter method that reduces colour noise is developed. The method uses a different weighting factor for each frequency sub-band. The factor includes cross-correlation components between clean speech and noise signal also. Enhanced speech quality can be improved in perceptual sense using non-linear Bark-scaled frequency spacing based on the fact that human ear sensibility varies nonlinearly in frequency spectrum.

In most spoken languages, voiced sounds represent a large amount (around 80%) of the pronounced sounds. In the classic short-time suppression techniques some harmonics are considered as noise only components and are consequently suppressed by the noise reduction process. This is one major limitation of those methods. To overcome this limitation, a method, called regeneration of suppressed harmonics that takes into account the harmonic characteristic of speech, is proposed. In this approach, the output signal of classic noise reduction technique is further processed to create an artificial signal where in the missing harmonics are automatically regenerated. This artificial signal is used to refine the apriori SNR used to compute a spectral gain.

2. Multi-Band Wiener Filter

In real environments, noise spectrum is not uniform for all the frequencies. For example, in the case of engine noise the most of noise energy is concentrated in low frequency. The human ear sensibility varies nonlinear in frequency spectrum. The principle of psychoacoustics [8,9] suggests that a spectral gain may be shared among adjacent high frequency components. A commonly used scale for signifying the critical bands is the Bark scale that divides the audible frequency range of 16 KHz into 24 abutting bands. **Figure 1** illustrates the relationship between the frequency in hertz and the critical band rate in Bark. An approximate analytical expression to describe the conversion from linear frequency *f*, into the critical band number *b* (in Bark) is:

$$b(f) = 13\arctan(0.76f) + 3.5\arctan\left[\left(\frac{f}{7.5}\right)^2\right]$$

In the frequency range from 0-8 KHz, there are 18

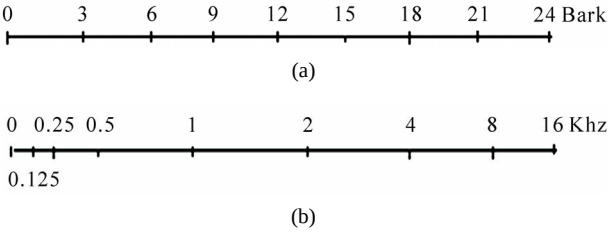

Figure 1. (a) Critical band rate and (b) Frequency.

critical bands. Therefore the spectral Wiener filter was modified for a critical band analysis to obtain the power spectral density on a Bark scale k:

$$Y_i(b) = \sum_{\omega(i)} |Y(k)|^2, \; i = 1, 2 \cdots K \tag{4}$$

where i is the critical band number, K = 18 is the total number of critical bands and $\omega(i)$ is the frequency index depending on the lower and upper frequency boundary of the critical band i.

The sub-band Wiener filter is derived according to the minimum mean square error (MMSE) criterion between the ideal and estimated sub-band speech signals in each of the sub-band. Its cost function in one sub-band is defined as

$$\varepsilon_i = E\left[\left(\hat{S}_i(k) - S_i(k)\right)^2\right] \tag{5}$$

where ε_i represents the expectation operator. $\hat{S}_i(k)$ and $S_i(k)$ denote the estimated and ideal sub-band speech signals in the i^{th} sub-band respectively.

In each sub-band, the noise suppression is performed by multiplying the Wiener filter gain to the sub-band noisy speech as:

$$\hat{S}_i(k) = G_i Y_i(k) \tag{6}$$

By substituting (6) in (5) and simplifying (5) we will get ε_i as

$$\begin{aligned}\varepsilon_i = {} & (G_i - 1)^2 E\left[S_i^2(k)\right] + G_i^2 E\left[D_i^2(k)\right] \\ & + 2(G_i - 1) G_i E\left[S_i(k) D_i(k)\right]\end{aligned} \tag{7}$$

2.1. Conventional Wiener Filter

In conventional Wiener filter assumed that $S_i(k)$ and $D_i(k)$ are zero mean and uncorrelated in each sub-band and (7) can be simplified to be

$$\varepsilon_i = (G_i - 1)^2 E\left[S_i^2(k)\right] + G_i^2 E\left[D_i^2(k)\right] \tag{8}$$

By setting the differentiation of (8) w.r.t weighting factor G_i to zero and the weighting factor G_i can be derived to be

$$G_i = \frac{E\left[S_i^2(k)\right]}{E\left[S_i^2(k)\right]+E\left[D_i^2(k)\right]} = \frac{\sigma_{S_i}^2}{\sigma_{S_i}^2+\sigma_{D_i}^2} = \frac{\sigma_{S_i}^2}{\sigma_{Y_i}^2} \quad (9)$$

where $\sigma_{S_i}^2$, $\sigma_{D_i}^2$ and $\sigma_{Y_i}^2$ represent the variance of the sub-band clean speech, noise and noisy speech in the i^{th} sub-band respectively.

2.2. Crosscorrelation Compensated Wiener Filter

The autocorrelation sequences of one frame of a clean speech, together with the background and noisy version of the same speech signal are shown in **Figure 2**. The autocorrelation sequence of noisy speech signal is not exactly equal to the sum of the autocorrelations of the noise and clean speech signals. This indicates the existence of the crosscorrelation between clean speech signal and noise signal [10].

Therefore, we cannot neglect the crosscorrelation between $S_i(k)$ and $D_i(k)$. Then by differentiating (7) w.r.t G_i and equating to zero and simplifying, we get

$$G_i = \frac{E\left[S_i^2(k)\right]+E\left[S_i(k)D_i(k)\right]}{E\left[S_i^2(k)\right]+E\left[D_i^2(k)\right]+2E\left[S_i(k)D_i(k)\right]}. \quad (10)$$

Since we have access only corrupt signal $Y_i(k)$ but not $S_i(k)$ it is not possible to estimate the cross correlation terms between $S_i(k)$ and $D_i(k)$. Hence, instead of calculating the cross term between $S_i(k)$ and $D_i(k)$, we estimate the crosscorrelation between $Y_i(k)$ and $D_i(k)$. Then

$$E\left[Y_i(k)D_i(k)\right] = E\left[\left(S_i(k)+D_i(k)\right)\cdot D_i(k)\right]$$
$$E\left[S_i(k)D_i(k)\right] = E\left[\left(Y_i(k)D_i(k)\right)\right]-E\left[D_i^2(k)\right] \quad (11)$$

By considering the crosscorrelation between $Y_i(k)$ and $D_i(k)$,

$$E\left[Y_i(k)D_i(k)\right] = \delta Y_i(k)D_i(k) \quad (12)$$

where δ is the crosscorrelation coefficient [9] for estimating the correlation between noisy speech signal and noise in a sub-band. By substituting (11) and (12) in (10), filter gain G_i can be obtained as

$$G_i = \frac{E\left[S_i^2(k)\right]+\delta Y_i(k)D_i(k)-E\left[D_i^2(k)\right]}{E\left[Y_i^2\right]}$$
$$G_i = \frac{\xi_i(k)+\delta\dfrac{Y_i(k)D_i(k)}{E\left[D_i^2(k)\right]}-1}{\gamma_i(k)} \quad (13)$$

where $\xi_i(k) = \dfrac{E\left[|S_i(k)|^2\right]}{E\left[|D_i(k)|^2\right]}$ and $\gamma_i(k) = \dfrac{E\left[|Y_i(k)|^2\right]}{E\left[|D_i(k)|^2\right]}$ represent the apriori SNR and the aposteriori SNR in the i^{th} sub-band respectively.

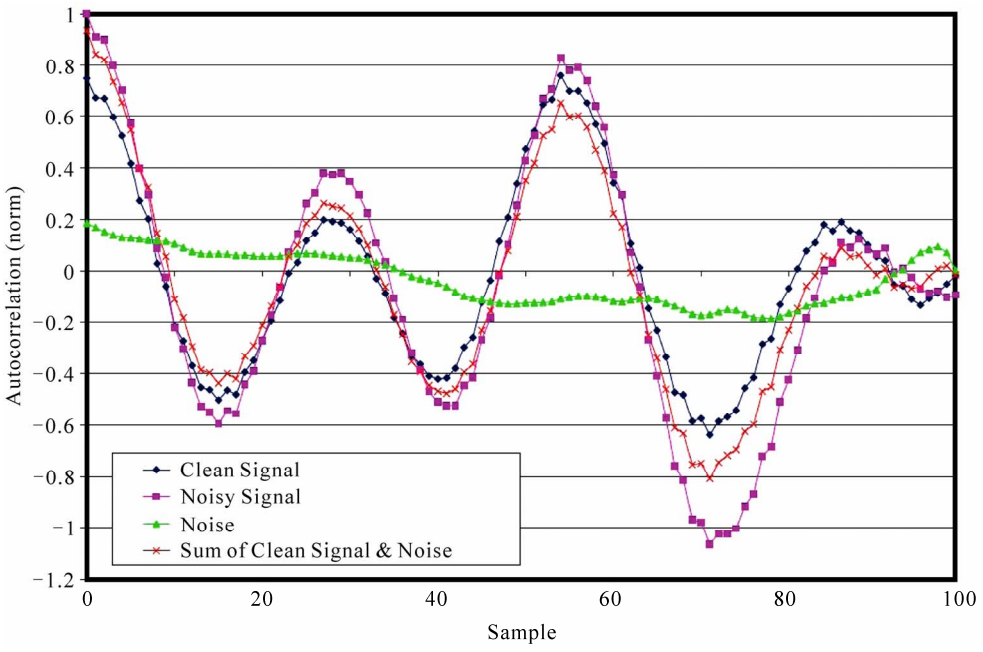

Figure 2. Autocorrelation sequences of clean speech, noisy speech, noise and sum of the clean speech and noise signals.

3. Regeneration of Suppressed Harmonics

The output signal $\hat{S}_i(k)$ or $\hat{s}(t)$ in time domain, obtained by the multiband Wiener filter presented in the previous section still suffers from distortions. This is inherent to the estimation errors introduced by the noise spectrum estimation since it is very difficult to get reliable instantaneous estimates in single channel noise reduction techniques. Since 80% of the pronounced sounds are voiced in average, the distortions generally turnout to be harmonic distortion. Indeed, some harmonics are considered as noise only components and are suppressed. For that reason, we propose to process the distorted signal to create a fully harmonic signal where all the missing harmonics are regenerated. This signal will then be used to compute a spectral gain able to preserve the speech harmonics. This will be called the speech harmonic regeneration step and can be used to improve the results of any noise reduction technique and not only the multiband Wiener filter.

A simple and efficient way to restore speech harmonics consists of applying a nonlinear function NL (e.g., absolute value, minimum or maximum relative to threshold, etc.) to the time signal enhanced in a first procedure with a classic noise reduction technique. Then, the artificially restored signal is obtained by

$$s_{harm}(t) = NL\left(\hat{s}(t)\right) \tag{14}$$

In this work, half wave rectification is used as a nonlinear function and applied to the signal. As a consequence, this signal cannot be used directly as clean speech estimation. Nevertheless, it contains very useful information that can be exploited to refine the apriori SNR.

$$\xi_i^{harm}(k) = \frac{\beta(k)E\left[\left|\hat{S}_i(k)\right|^2\right] + \left(1-\beta(k)\right)E\left[\left|s_{ham}(k)\right|^2\right]}{E\left[\left|D_i(k)\right|^2\right]} \tag{15}$$

The $\beta(k)$ parameter is used to control the mixing level of $E\left[\left|\hat{S}_i(k)\right|^2\right]$ and $E\left[\left|S_{harm}(k)\right|^2\right]$. The properties of this parameter are:

- when the estimation of $\hat{S}_i(k)$ provided by the multiband Wiener filter algorithm is reliable, the harmonic regeneration process is not needed and $\beta(k)$ should be equal to 1.
- when the estimation of $\hat{S}_i(k)$ provided by the multiband Wiener filter algorithm is unreliable, the harmonic regeneration process is required to correct the estimation and $\beta(k)$ should be equal to 0 (or any other constant value depending on the chosen nonlinear function).

The $\beta(k)$ parameter can be chosen constant to realize a compromise between the two estimators $\hat{S}_i(k)$ and $S_{harm}(k)$. In present work, we propose to choose $\beta(k) = G_i$ to match above properties. And the apriori SNR is refined which is used to compute a new spectral gain [11-13]

$$G_i^{harm} = \frac{\xi_i^{harm}(k) + \delta\dfrac{Y_i(k)D_i(k)}{E\left[D_i^2(k)\right]} - 1}{\gamma_i(k)} \tag{16}$$

4. Results and Discussion

To evaluate and compare the performance of the proposed method, simulations are carried out with the NOIZEUS [14], a database widely used in testing speech enhancement algorithms. 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 seven types of noise at global SNR levels of 0 dB, 5 dB, 10 dB and 15 dB. The noises were airport, car, babble, train and street noises. The objective quality measures used for the evaluation of the proposed method are the segmental SNR and noise reduction (NR) values. 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 and NR values indicates the weaker speech distortions and better perceived quality of the processed speech signal [15]. The performance of the proposed method is compared with Wiener filter and multi-band Wiener filter.

Table 1 shows the segmental SNR improvement with segment size equal to 256 for various noise levels. The performance of the proposed method almost outperforms that of the Wiener filter and multi-band Wiener filter.

Table 2 demonstrates the comparison of NR values. It reveals that the proposed method benefits low speech distortion and retains the residual noise at an acceptable level. The timing waveforms of the enhanced speech are demonstrated in **Figure 3**. Clean speech signal is corrupted by airport noise at 0 dB SNR. It shows that proposed method can efficiently remove the background noise.

Figure 4 shows the comparison of spectrograms. The background noise can be efficiently removed by the proposed method. It is evident from listening tests that the proposed method efficiently reduces the background noise with less speech distortion.

Table 1. Segmental SNR in the enhanced speech in various noise environments.

Type of noise and SNR (dB)	Wiener filter	Multi-band Wiener	Proposed method
Airport-0	–4.37	–2.39	–1.76
Airport-5	–2.57	0.67	0.87
Airport-10	–0.06	0.43	0.46
Airport-15	1.88	3.13	3.43
Babble-0	–4.59	–1.91	–1.14
Babble-5	–1.39	0.05	0.36
Babble-10	0.03	2.36	2.48
Babble-15	2.71	3.06	3.97
Car-0	–3.93	–1.02	–1.28
Car-5	–1.65	1.69	1.75
Car-10	0.68	2.40	2.60
Car-15	2.31	2.71	2.98
Street-0	–2.88	–1.97	–1.39
Street-5	–2.13	–0.29	–0.11
Street-10	1.20	2.42	2.54
Street-15	2.25	2.48	2.88
Train-0	–3.45	–2.13	–1.77
Train-5	–0.86	0.93	1.90
Train-10	–0.39	1.69	2.86
Train-15	2.62	2.57	3.57
Restaurant-0	–5.49	–3.44	–3.05
Restaurant-5	–3.61	–0.15	0.21
Restaurant-10	–0.49	1.28	1.56
Restaurant-15	1.80	2.47	2.68
Station-0	–3.62	0.51	0.89
Station-5	–1.93	1.18	1.57
Station-10	0.95	2.39	2.89
Station-15	2.72	2.86	3.52

Table 2. Noise reduction values in various noise environments.

Type of noise and SNR (dB)	Wiener filter	Multi-band Wiener	Proposed method
Airport-0	–4.37	25.00	26.05
Airport-5	–2.57	25.98	26.86
Airport-10	–0.06	24.01	24.25
Airport-15	1.88	26.22	26.92
Babble-0	–4.59	25.9	–1.14
Babble-5	–1.39	25.85	26.08
Babble-10	0.03	26.60	26.69
Babble-15	2.71	26.12	26.46
Car-0	–3.93	26.30	26.56
Car-5	–1.65	27.69	27.75
Car-10	0.68	26.89	27.03
Car-15	2.31	25.81	25.89
Street-0	–2.88	25.34	25.92
Street-5	–2.13	25.36	25.55
Street-10	1.20	25.86	26.86
Street-15	2.25	25.24	25.67
Train-0	–3.45	25.63	25.99
Train-5	–0.86	26.64	26.89
Train-10	–0.39	25.96	26.10
Train-15	2.62	25.44	25.56
Restaurant-0	–5.49	23.50	23.88
Restaurant-5	–3.61	25.42	25.79
Restaurant-10	–0.49	25.61	25.89
Restaurant-15	1.80	25.42	25.56
Station-0	–3.62	28.37	28.78
Station-5	–1.93	27.11	27.67
Station-10	0.95	26.76	26.98
Station-15	2.72	25.87	26.02

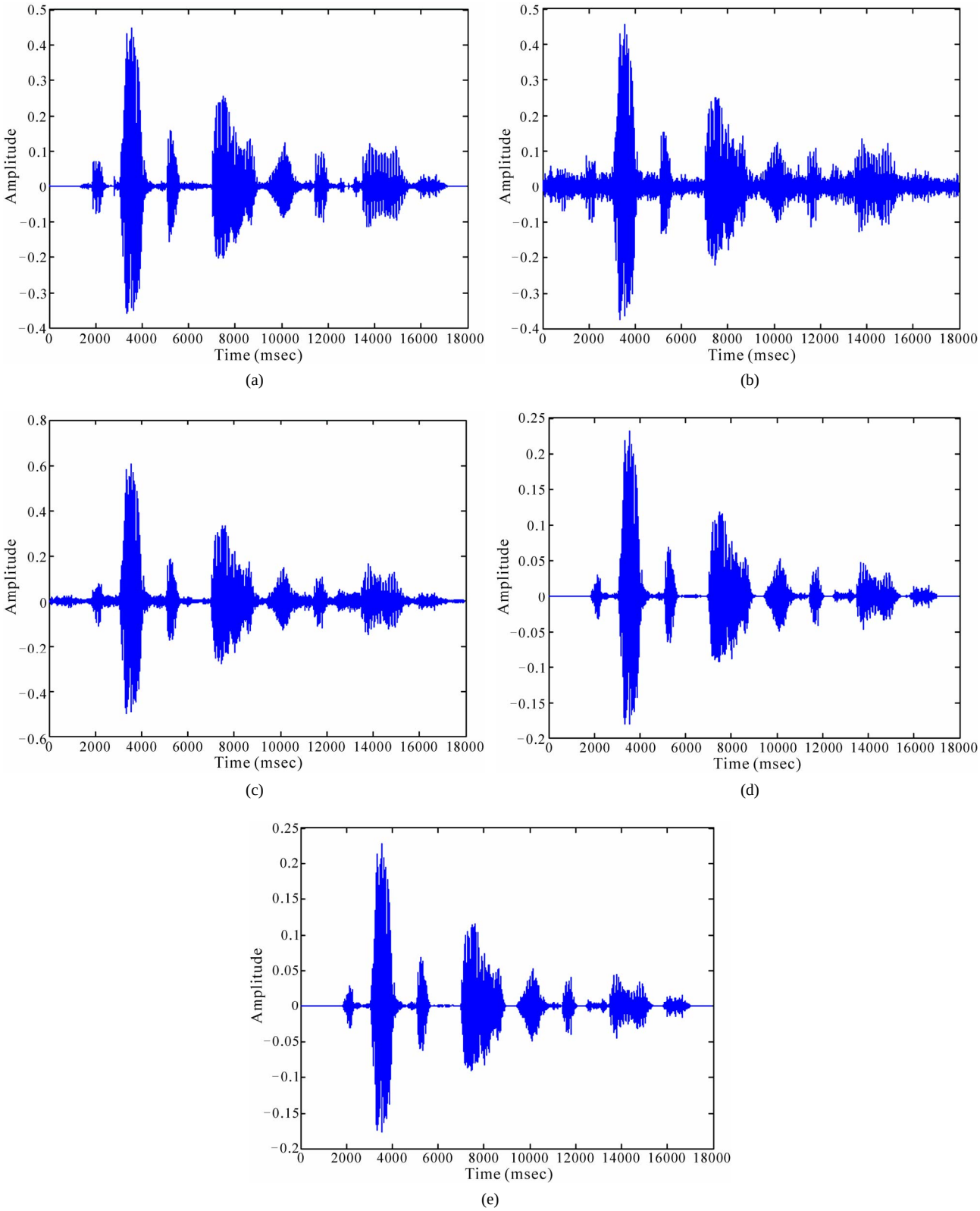

Figure 3. Timing waveforms of (a) the clean speech (b) noisy speech corrupted and the enhanced speech using (c) Wiener filter (d) Multi-band Wiener filter and (e) the proposed method.

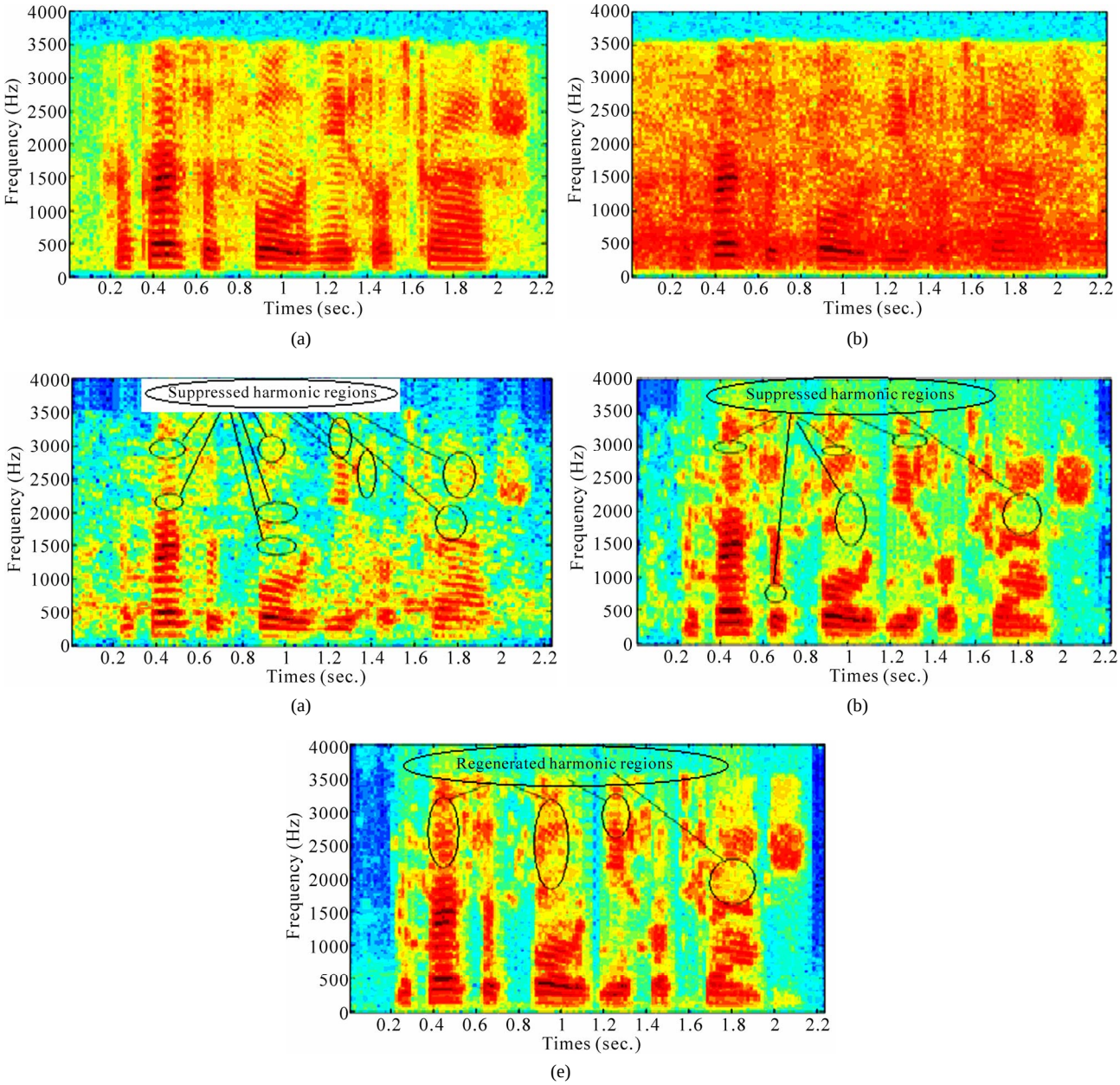

Figure 4. Spectrograms of (a) the clean speech (b) noisy speech corrupted and the enhanced speech using (c) Wiener filter (d) Multi-band Wiener filter and (e) the proposed method.

5. Conclusions

This paper presents an improved Wiener filtering method that takes into account the non-uniform effect of colored noise on the spectrum of speech. Proposed method includes the cross correlation terms between the clean speech and noise. Multi-band Wiener filtering method reduces residual musical tones that appear in enhanced speech for Wiener filtering. A noise reduction technique based on the principle of harmonic regeneration is also proposed. Classic techniques, including the Multi-band wiener, suffer from harmonic distortions when the SNR is low. This is mainly due to estimation errors introduced by the noise PSD estimator. To solve this problem, non-linearity is used to regenerate the degraded harmonics of the distorted signal in efficient way.

The resulting artificial signal is used to refine the apriori SNR which is then used to compute a spectral gain that preserves speech harmonics, and hence avoids distortions. Results are given in terms of segmental SNR

and noise reduction values. All these results demonstrate the good performance of the proposed method.

REFERENCES

[1] Y. Ephraim, "Statistical-Model-Based Speech Enhancement Systems," *Proceedings of IEEE*, Vol. 80, No. 10, 1992, pp. 1526-1555. doi:10.1109/5.168664

[2] J. R. Deller, H. G. Proakis and J. H. L. Hansen, "Discrete-Time Processing of Speech Signals," Macmillan, New York, 1993.

[3] S. V. Vaseghi, "Advanced Digital Signal Processing and Noise Reduction," 2nd Edition, John Wiley & Sons ltd., Chichester, 2000. doi:10.1002/0470841621

[4] Y. Gui and H. K. Kwan, "Adaptive Subband Wiener Filtering for Speech Enhancement Using Critical-Band Gammatone Filterbank," 48*th Midwest Symposium on Circuits and Systems*, Vol. 1, 7-10 August 2005, pp. 732-735.

[5] J. S. Lim and A. V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech," *Proceedings of IEEE*, Vol. 67, No. 12, 1979, pp. 1586-1604. doi:10.1109/PROC.1979.11540

[6] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean Square Error Short-Time Spectral Amplitude Estimator," *IEEE Transaction on Speech Audio Processing*, Vol. 32, No. 6, 1984, pp. 1109-1121.

[7] S. F. Boll, "Suppression of Acoutics Noise in Speech Using Spectral Subtraction," *IEEE Transaction on Acoustics, Speech, Signal Processing*, Vol. 27, No. 2, 1979, pp. 113-120. doi:10.1109/TASSP.1979.1163209

[8] E. Zwicker and H. Fastl, "Psychoacoustics," Springer Verlag, Berlin, 1990.

[9] K. A. Sheela, Ch. V. R. Rao, K. S. Prasad and A. V. N. Tilak, "A New Noise Reduction Pre-Processor for Mobile Voice Communication Using Perceptually Weighted Spectral Subtraction Method," 3*rd International Conference on Mobile Ubiquitous and Pervasive Computing*, VIT University, 16-19 December 2006.

[10] G. Farahani, S. M. Ahadi and M. M. Homayounpoor, "Robust Feature Extraction of Speech via Noise Reduction in Autocorrelation Domain," Lecture Notes in Computer Science 4105, Springer-Verlag, 2006, pp. 466-473.

[11] P. Scalart and J. V. Filho, "Speech Enhancement Based On a Priori Signal to Noise Estimation," *Proceedings of IEEE International Conference on Acoustics, Speech Signal Processing*, Atlanta, Vol. 2, May 1996, pp. 629-632. doi:10.1109/ICASSP.1996.543199

[12] J. E. Porter and S. F. Boll, "Optimal Estimators for Spectral Restoration of Noisy Speech," *Proceedings of IEEE International Conference on Acoustics, Speech Signal Processing*, Vol. 9, March 1984, pp. 53-56.

[13] I. Cohen, "Optimal Speech Enhancement under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator," *IEEE Signal Processing Letters*, Vol. 9, No. 4, 2002, pp. 113-116. doi:10.1109/97.1001645

[14] A Noisy Speech Corpus for Evaluation of Speech Enhancement Algorithms, 2011. http://www.utdallas.edu/~loizou/speech/noizeus/

[15] Y. Hu and P. C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement," *IEEE Transaction on Audio, Speech and Language Processing*, Vol. 16, No. 1, 2008, pp. 229-238. doi:10.1109/TASL.2007.911054

