International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 8, No. 4 (2015), pp. 49-58
http://dx.doi.org/10.14257/ijsip.2015.8.4.05

Overlapping Frame Approach to Estimate and Reduce Noise from Single Channel Speech

Harjeet Kaur and Dr. Rajneesh Talwar

Ph. D Research Scholar, Punjab Technical University, Jalandhar, Punjab, India
Principal, Chandigarh Group of Colleges, Technical Campus Jhanjeri, Punjab, India
mail2hkaur@gmal.com, rtphdguidance@gmail.com

Abstract

Speech enhancement is a long standing problem with various applications like telephone conversation and speech recognition. The corruption of speech due to presence of additive background noise causes severe difficulties in various communication environments. If the background noise is evolving more slowly than the speech, then the estimation of the noise during speech pauses is easier as compared to non stationary noise. If in case the Noise is varying rapidly then estimation is more difficult. This paper focuses on the class of single-channel noise reduction methods that are performed with frequency domain using short-time Fourier Transform (STFT). There are number of publications and implementations on noise reduction systems. But, there are still some issues in non- stationary noisy systems. This single-channel approach is more dominant and effective approach for practical systems. From last few years, algorithms have been proposed for this problem but most of them are worked on noisy signal in current frame. So in this paper we are trying to propose the new model using Wiener filter by using the concept of multi-frame approach with different window sizes and overlaps. The proposed method shows the results with its superiority.

Keywords: *Log Likelihood ratio (LLR), weighted spectral slope (WSS), non stationary, Signal to noise ratio, speech enhancement*

1. Introduction

Non stationary noise remains one of the biggest challenges for current state-of-the-art single-channel noise reduction schemes. The statistics of the background noise must, therefore, be only slowly time-variant. Moderately non-stationary noise can be tracked with these algorithms but the performance breaks down severely with increasing non-stationary status of the background noise [13]. Every speech communication and processing system suffers from the ubiquitous presence of additive noise, but today's widespread cellular phones and hands-free handsets are more likely to be used in acoustically adverse environments where background noise from different origins is loud and where the microphone may not be in close proximity to the speech source. The external disturbance degrades the perceptual quality of speech and will impair the speech intelligibility when the signal-to-noise ratio (SNR) comes down to a certain level. Noise reduction intends to suppress such additive noise for the purposes of speech enhancement. Noise reduction algorithms generally can enhance only the perceptual quality of speech when presented directly to a human listener with normal hearing, but may improve both speech quality and intelligibility when the enhanced speech goes through a voice communication channel before being played out [13] and/or for the hearing impaired [7]. So single-channel noise reduction (SCNR) has a large variety of applications including mobile phones, hearing aids, voice over Internet protocol (VoIP), just to name a few. The first SCNR system was developed over 45 years ago by Schroeder [20, 21]. Nowadays,

ISSN: 2005-4254 IJSIP

and the principle of Schroeder's system is the well-known Spectral Magnitude Subtraction method. This work, however, has not received much public attention, probably because it is a purely analog implementation and more importantly it was never published in journals or conferences outside of the Bell System. The interest in a digital form of the spectral subtraction technique was sparked by a 1974 paper by Weiss, Ashkenazy, and Parsons [23]. A few years later, Boll, in his often-cited paper [6], reintroduced the Spectral Subtraction method yet for the first time in the framework of digital Short-Time Fourier analysis. These early algorithms were all based on an intuitive and simple idea: the clean speech spectrum can be restored by subtracting the estimate of the noise spectrum from the noisy speech spectrum and the noise spectrum is estimated and updated during silent periods. Though practically effective, the Spectral Magnitude Subtraction approach is by no means optimal. It was thanks to the papers of [17, 13] that the Spectral Subtraction technique began being examined in the framework of optimal estimation theory. This treatment initiated the development of many new noise reduction algorithms in the last three decades. The Wiener filter that intends to directly recover the complex (amplitude and phase) spectrum (*i.e.,* the waveform in the time domain) of the clean speech [13, 17], and in contrast to those in which only the spectral amplitude of the clean speech is estimated while its phase is copied from the phase of the noisy signal. The spectral amplitude can be taken as the square root of a Maximum-Likelihood (ML) estimate of the clean speech's power spectrum. This leads to the spectral power subtraction method [17, 5] which is subtly different from the ML spectral amplitude estimator [17]. In addition to the classical approach of ML estimation, the Bayesian decision rule was found also very useful. Ephraim and Malah introduced a celebrated minimum mean square error (MMSE) estimator for spectral amplitude (MMSE-SA) in [10]. This original idea was later enriched by the MMSE estimator for log spectral amplitude (MMSE-LSA) [26] and other generalized Bayesian estimators [26-18], which minimize the posterior expectation of various distance measures between the actual and estimated speech spectral amplitude. Maximum a posteriori (MAP) is another important Bayesian decision rule based on which Wolfe and God sill developed an MAP Spectral Amplitude Estimator (MAP-SA) [24]. These calculations are practically reasonable but may not be strictly true. Alternatively a super-Gaussian model was suggested to be applied in combination with the MAP-SA approach in [16]. More complicated statistical speech models (*e.g.,* Hidden Markov Model) can also be used [9] but no close-form solution will be possibly deduced. While SCNR has been widely studied in the time domain and other transform domains too [3-4], the frequency-domain techniques are by far the most popular choice in practical systems for their simplicity and relative effectiveness. In this paper, we will focus only on this class of approaches. In spite of using the distinctive optimization rules (ML, MMSE, or MAP), spectral distance measures (linear versus log), and statistical models for speech [Gaussian, Super-Gaussian, or Hidden Markov Model (HMM)], the existing frequency-domain noise reduction algorithms have one feature in common: the solution is eventually expressed as a gain function applied to the Short-Time Fourier Transform (STFT) of the noisy signal in each frequency. This is due to a simplified formulation of the problem in which it has been implicitly assumed that the STFT of the current frame is uncorrelated with that in the neighboring frames. However, this is not accurate for speech enhancement since speech is a highly self-correlated signal. Consequently, by taking the inter frame correlation into account; we should be able to develop more sophisticated algorithms with hopefully better noise reduction results. In this case, when we estimate the STFT of the clean speech in the current frame, we use the STFTs of the noisy signal both in the current frame and the previous frames (with respect to the same frequency) [12]. This leads to a new model similar to a microphone array system: We have multiple noisy speech observations; their speech components are correlated while their noise components are presumably uncorrelated or correlated in a different way than speech components. As a result, the multichannel (here multi-frame) Wiener filter and the Minimum Variance Distortion less Response (MVDR) filter that were usually associated with microphone arrays will be developed for SCNR in this paper. It is well known that the gain functions of the existing frequency-domain SCNR algorithms cannot improve the narrowband SNR and full band noise reduction is achieved at a price of speech distortion. With the new algorithms developed in this paper, we will show that both the narrowband and full band SNRs can be improved. An early attempt at exploiting the inter-frame correlation of speech in sub-bands was reported in [25]. A simple first-order autoregressive (AR) model was used to describe the variation of speech and hence the Kalman filter was developed to estimate the clean speech signals in each sub band. The coefficients of the sub band AR models need to be estimated from the noisy microphone signal and their estimates are usually biased in practice. So this method is subject to errors from model misspecification. In a recent paper [19], it was also suggested that the inter-frame correlation of speech STFTs could be exploited and an iterative optimization scheme was proposed to improve the traditional frequency-domain Wiener filter. There are many algorithms for colour fidelity [1] which can generally be divided into three classes: first class includes approaches using low level image features

2. Problem Formulation

The noise reduction problem considered in this paper is one of recovering the desired signal (or clean speech)x(t), t being the time index, of zero mean from the noisy observation (microphone signal) [22].

$$y(t) = x(t) + s(t) \tag{1}$$

Where s(t) is the unwanted additive noise with zero-mean random process white or colored but uncorrelated with x(t). To simplify the development and analysis of the main ideas of this work, we further assume that all signals are Gaussian and wide sense stationary. Using the Short-Time Fourier Transform (STFT), [13] can be rewritten in the frequency domain as

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

where Y(k,m),X(k,m),and S(k,m)are the STFTs of y(t), x(t),and s(t), respectively, at frequency-bin k€{0,1,2………………,k-1} and time-frame m. Since x(t) and s(t) are uncorrelated by assumption, the variance of Y(k,m) is

$$\phi_y(k,m) = E[|Y(k,m)|^2] = \phi_x(k,m) + \phi_s(k,m) \tag{3}$$

where E[.] denotes mathematical expectation and $\phi_X(k,m)=E[|X(k,m)|^2]$ and $\phi s(k,m)=E[|S(k,m)|^2]$ are the variances of X(k,m) and S(k,m) respectively.

3. Wiener Filter in Frequency Domain

The Wiener filter is a popular technique that has been used in many signal enhancement methods. The basic principle of the Wiener filter is to obtain an estimate of the clean signal from the corrupted additive noise. This estimate is obtained by minimizing the Mean Square

Error (MSE) between the desired signal x(n) and the estimated signal $\hat{X}(n)$. The frequency domain solution to this optimization problem gives the following filter transfer function [18].

$$H(w) = P_{x(w)} / P_{x(w)} + P_{s(w)} \tag{4}$$

Where $Px(\omega)$ and $Ps(\omega)$ are the power spectral densities of the clean and the noisy signal, respectively. This formula can be derived considering the signal x and the noise s as uncorrelated and stationary signals. The SNR is

$$SNR = P_{x(w)} / \hat{P}_{s(w)} \tag{5}$$

This definition can be incorporated to the Wiener filter equation as follows:

$$H(w) = [1 + 1/SNR]^{-1} \tag{6}$$

The wiener filter gives fixed frequency response at all frequencies that considered as the limitation of the wiener filter and the requirement to estimate the power spectral density of the clean and noisy signal prior to filtering.

4. A New Linear Model for Speech Spectral Estimation

In the linear model, we try to estimate our desired signal, X(k,m), from the observation signal ,Y(k,m), by applying a complex gain to it [1].

$$\begin{aligned} \hat{X}(k,m) &= H^{*}(k,m)Y(k,m) \\ &= H^{*}(k,m)[X(k,m) + S(k,m)] \\ &= X_{fd}(k,m) + S_{rn}(k,m) \end{aligned} \tag{7}$$

where the superscript * denotes complex conjugation, $X_{fd}(k,m)= H^{*}(k,m) X(k,m)$ is the filtered desired signal and $S_{rn}(k,m)= H^{*}(k,m)S(k,m)$ is the residual noise. Using the Mean-Square Error (MSE) between the estimated and desired signals, we can easily derive the optimal Wiener gain, which is real and is given by [1].

$$H_{w}(k,m) = \phi_{x}(k,m) / \phi_{y}(k,m)$$

$$= 1 - \phi_{y}(k,m) / \phi_{y}(k,m) \tag{8}$$

As a result, the estimate of X(k,m) in the Wiener sense is [1].

$$\hat{X}_{w}(k,m) = H_{w}(k,m)Y_{w}(k,m) \tag{9}$$

In (4), we implicitly assumed that the observation signal at the current time-frame is uncorrelated with itself at the previous time-frames. Therefore, the interface correlation should be taken into account in the derivation of any noise reduction algorithms [21].

5. Performance Measures

In this section, we give some very useful measures that fit well with the linear model developed in this section, where the inter frame correlation is taken into account. We define the narrowband and full band input SNRs as

$$iSNR(k,m) = \phi_{x}(k,m) / \phi_{s}(k,m) \tag{10}$$

$$iSNR = \sum_{k=0}^{k-1} \phi_{x}(k,m) / \sum_{k=0}^{k-1} \phi_{x}(k,m)\phi_{s}(k,m) \tag{11}$$

$$iSNR \leq \sum_{k=0}^{k-1} iSNR(k,m) \tag{12}$$

5.1. Optimal filters

In this part, we derive three fundamental filters with the linear inter frame model and show how they are related to each other. We also show the relationship with all of them. For that, we need to derive first the MSE criterion and its relation with the MSE of speech distortion and residual interference-plus-noise. We define the narrowband error signal between the estimated and desired signals as

$$\begin{aligned}\varepsilon(k,m) &= X1(k,m) - X(k,m) \\ &= h^{H}(k,m)y(k,m) - X(k,m)\end{aligned} \tag{13}$$

$$\begin{aligned}\varepsilon(k,m) &= \varepsilon_d(k,m) + \varepsilon_r(k,m) \\ \varepsilon_d(k,m) &= \varepsilon_{fd}(k,m) - X(k,m)\end{aligned} \tag{14}$$

5.1.1. Wiener: The Wiener filter is easily derived by taking the gradient of the narrowband MSE, with respect to h^H (k,m) and equating the result to zero:

$$h_w(k,m) = \phi_y^{-1}(k,m)\phi_{yx}(k,m)i_1 \tag{15}$$

where ϕy(k,m)= E[y(k,m)y^H(k,m)] is the covariance matrix of y(k,m) and ϕyx(k,m)= E[y(k,m)x^H(k,m)] is the cross-correlation matrix between y(k,m) and x(k,m), but

$$\phi_{yx}(k,m)i_1 = \phi_x(k,m)\gamma_x^*(k,m) \tag{16}$$

$$h_w(k,m) = \phi_x(k,m)\phi_y^{-1}(k,m)\gamma_x^*(k,m) \tag{17}$$

The Wiener filter can also be written in this form

$$h_w(k,m) = \phi_y^{-1}(k,m)\phi_x(k,m)i_1 \tag{18}$$

$$= [1 - \phi_y^{-1}(k,m)\phi_s(k,m)]\,i_1 \tag{19}$$

Interestingly, the higher is the value of oSNR means to increase e number of inter frames, and less the distortions in the desired signal with the Wiener filter at frequency-bin k,

$$oSNR\ [h_w(k,m)] > iSNR\ (k,m) \tag{20}$$

6. Experimental Results

In this section, we present the experimental results of the frequency-domain algorithm that is a SCNR and that may use the concept of multiple STFT frames. Comparisons with the traditional single-frame Wiener filter will be used to study and validate the merits of exploiting inter frame correlations. Due to the limitation, the main focus is placed on showing the results of the new multi-frame Wiener filters.

6.1. Set up and Metrics

In our experiments, the microphone signal is artificially synthesized by adding prerecorded real-world noise to a clean speech signal. The clean speech signals were recorded from female and male speakers. Each speaker provided 2 to 4 minutes of conversational speech” that is a “story” about anything that came to his/her mind. All recordings were originally digitized at a sampling rate of 8 kHz with 16 bits per sample and down sampled to 4 KHz with alpha is 0.9 and min SNR is -10 and max. SNR is 35. In the experiments presented here, we consider only one male speaker. Each story was cut to have the same length of 6s and babble noise. The noise is fairly stationary but colored with an energy roll-off (approximately 12 dB per octave) towards high frequencies. The babble noise was recorded in the Mumbai railway station. It is not only colored but also non stationary with mixtures of nearly inaudible voices and sporadic cell phone rings. The noise level is adjusted according to that of the clean speech and a specified input SNR. In the following, if not explicitly stated otherwise, the noise is white Gaussian random noise and the speech source is the first male speaker. The full band output SNR and speech distortion measures are used in our experiments. Moreover, we will use the weighted spectral slope WSS for the measurement of objective speech quality and Log Likelihood Ratio LLR are calculated for each frame of the input speech.

6.2. Algorithm Implementation

The algorithms discussed and developed in this paper are all frequency-domain approaches. The STFT is implemented with the hamming window and the Fast Fourier Transforms (FFT).The max size of FFT is 512. The window size in the samples is set to be a power of 2. For the traditional single-frame Wiener filter, an overlap of 50% between neighboring windows is commonly used while for the proposed multi-frame Wiener filter we adopt an overlap of 64% to 75% to retain a higher inter-frame correlation. This analysis and synthesis procedure is nearly perfect in Mat lab, resulting in little distortion in the reconstructed signal if no manipulation is carried out to its frequency-domain representations.

6.2.1. Wiener Filters: We first show the performance of the traditional single-frame Wiener filter, which provides a benchmark for studying other noise reduction filters. Such a Wiener filter takes (k=256) (corresponding to 32 ms) and 64% overlapping windows. Figure 1 plots the results. Using a large (forgetting factor=1), we cannot capture the short-term variations of non stationary signals, but with a small value, the sample estimate of the signal variance has a large variation due to a limited number of data to do averaging. So the best performance is achieved. An interesting observation is that the oSNR reaches its peak when the forgetting factor of output is equal to forgetting factor of noise. The second experiment considers the where k=256 and overlap=75% and L go from 1 to 16, third experiment considers k=256 and overlap=50% and forth experiment considers k=64 and overlap=50%. As window size k affects the performance of the multi frame wiener filter. When k is small, FFT resolution is poor. In case k is increasing it will be helpful to improve the performance but with the increase of a value of k it corresponds to long gap in consecutive time frames as a result inter frame correlation is weaker. So from the analysis we conclude that as window size is increasing, speech distortion is increasing, same with increasing the no of frames speech distortion is increasing. Log Likelihood Ratio (LLR) is decreasing with the decrease of value of k, but weighted spectral slope is increasing. An interesting discovery is that the gain is greater for a low iSNR than for a high iSNR. Before we conclude this subsection, there is one thing that needs to be clarified and discussed that which set of performance measures we used for the above presented experiments. As a matter of fact, we used the conventional definitions.

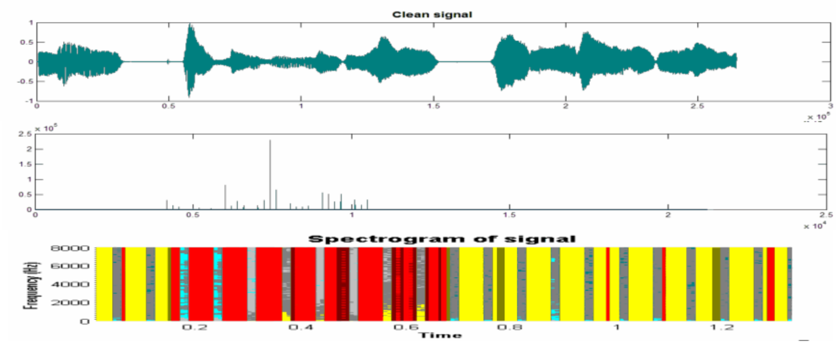

Figure 1. Signal Waveform with k=256 and Overlap=64%: (a) Clean Signal (b) Noisy Signal (C) Spectogram

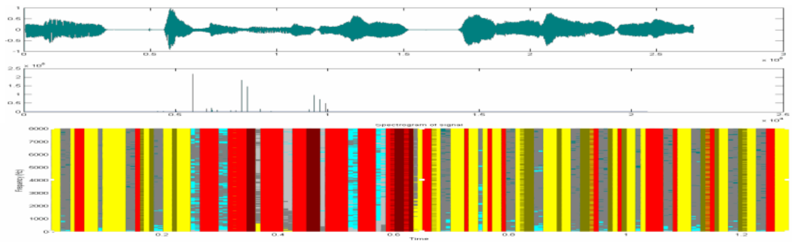

Figure 2. Signal Waveform with k=256 and Overlap=75%: (a) Clean Signal (b) Noisy Signal (C) Spectogram

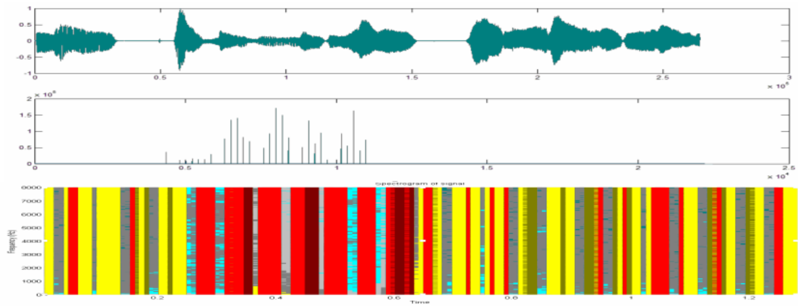

Figure 3. Signal Waveform with k=256 and Overlap=50%: (a) Clean Signal (b) Noisy Signal (C) Spectogram

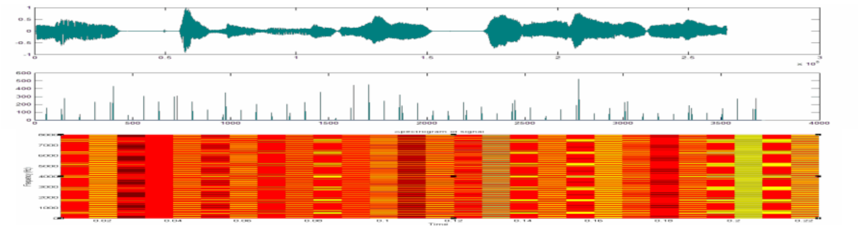

Figure 4. Signal Waveform with k=64 and Overlap=50%: (a) Clean Signal (b) Noisy Signal (C) Spectogram

Table 1. Comparison of LLR, oSNR, SNR seg, WSS for Different Values of k and Overlap

Input	LLR	oSNR	SNR Seg	WSS
k=256,O=50%	13.8596	-85.85	-10	161.057
k=256,O=64%	13.8596	-81.817	-10	121.34
k=256,O=75%	13.8596	-83.82	-10	120.744
k=64,O=50%	11.686	-45.74	-10	243.922

Table 2. Analysis with Noisy and Enhanced Speech Pattern for Input k=256 and Overlap=64%

Noise Type	SNR	oSNR	LLR	SNR Seg	WSS
Babble	0	-17.385899	2.246574	-9.123063	127.987540
	5	-14.929	2.23597	-8.46154	131.094126
	10	-13.395822	2.208571	-8.000444	129.676754
	15	-12.960638	2.188211	-7.937447	127.291182
exhibition	0	-18.33098	2.232366	-9.258967	131.220742

	5	-16.13787	2.239518	-8.951642	127.735754
	10	-13.14529	2.261398	-8.059477	130.771905
	15	-14.5642	2.246833	-8.253590	124.698888
Restaurant	0	-15.642186	2.137032	-8.640776	126.291438
	5	-15.1970	2.204341	-8.580691	125.626808
	10	-12.13749	2.239478	-8.000559	129.394346
	15	-14.49530	2.229242	-7.962446	131.093963
Street	0	-17.6881	2.311712	-9.325353	132.4266
	5	-14.67466	2.235174	-8.748175	132.346674
	10	-13.18220	2.206806	-8.23634	128.0524
	15	-15.7807	2.326610	-8.599206	130.364324
Car	0	-18.74221	2.236087	-8.98998	127.0998
	5	-16.147181	2.226145	-8.794045	131.165087
	10	-14.39887	2.567100	-8.67789	128.3245
	15	-13.846244	2.201041	-8.021956	129.667464

6.3. Objective Speech Quality Measure

The conducted research indicates that the output SNR and the speech distortion index provide a complete and insightful picture of the noise reduction performance. They are closely aligned with our perception of the quality of the enhanced signals in informal listening tests, using proper set of definitions, It has become clear that exploiting inter frame correlations is helpful to the Wiener filters, but it can give rise to arguments if we compare the performance of the Wiener filters using different sets of performance measure definitions. So for this task, we chose to use the WSS measure, which has been found to have higher correlations, than other widely known objective measures, with the subjective ratings of overall quality of enhanced speech signals (Lim and Oppenheim 2012). WSS ranges between 161.057 and 243.922 table shows the results for the different values of K and overlap. For the traditional single-frame Wiener filter (with a 64% overlap), we set according to the results presented in Figure 1. The multi-frame Wiener filter performs always better than the single-frame counterpart for all noise types. It is noted that the MVDR filter produces low speech distortion but high residual noise. When the input SNR is low (lower than 10 dB), the high level of the residual noise outweighs speech distortion in the PESQ measure such that the MVDR filter yields lower PESQ scores than the two Wiener filters. On the contrary, when the input SNR gets practically high, speech distortion becomes much easier to be perceived with lower residual noise in the background. Consequently, the MVDR filter has higher PESQ scores than the Wiener filters in those conditions.

7. Conclusion

In this paper, we presented an insightful analysis of the frequency-domain SCNR algorithms whose solutions are all finally expressed as gain functions applied to the spectrum of the noisy speech only in the current frame. We explained that this common feature is due to the disregard of the inter frame correlation, which may be strong for speech. By taking the inter frame correlation into account, we proposed a new linear model for speech spectral estimation and developed namely, the Wiener filters. It was proved that both the narrowband and full band output SNRs can be improved. Extensive simulation results were reported and clearly justified the advantage of exploiting the inter frame correlation for SCNR.

References

[1] J. Benesty, J. Chen, Y. Huang and I. Cohen, “Noise Reduction in Speech Processing”, Berlin, Germany, Springer-Verlag, **(2009)**.
[2] J. Benesty, J. Chen and Y. Huang, “On noise reduction in the Karhunen-Loeve expansion domain”, in Proc. IEEE ICASSP, **(2009)**, pp. 25–28.
[3] J. Benesty and J. Chen, “Optimal Time-Domain Noise Reduction Filters A Theoretical Study Berlin”, Germany: Springer.
[4] J. Benesty, J. Chen and Y. Huang, “Speech Enhancement in the Karhunen-Loeve Expansion Domain”, San Rafael, CA: Morgan & Claypool, **(2011)**.
[5] M. Berouti, Schwartz and J. Makhoul, “Enhancement of speech corrupted by acoustic noise”, in Proc. IEEE ICASSP, **(1979)**, pp. 208–211.
[6] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction”, IEEE Trans. Acoust., Speech, Signal Process., ASSP, vol. 27, no. 2, **(1979)**, pp. 113–120.
[7] V. Bray and M. Valente, “Can omni-directional hearing aids improve speech understanding in noise”, Audiol. Online, **(2011)**.
[8] I. Cohen, “Relaxed statistical model for speech enhancement and a priori SNR estimation”, IEEE Trans. Speech Audio Process, vol. 13, **(2005)**, pp. 870–881.
[9] Y. Ephraim, D. Malah and B. H. Juang, “On the application of hidden Markov models for enhancing noisy speech”, IEEE Trans. Acoust., Speech, Signal Process, vol. 37, no. 12, **(2011)**, pp. 1846–1856.
[10] Y. Ephraim and D. Malah, “Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator”, IEEE Trans. Acoust., Speech, Signal Process, vol. 32, no. 6, **(1984)**,pp. 1109–1121.
[11] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error log-spectral amplitude estimator”, IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP, vol. 33, no. 2, **(1985)**, pp. 443–445.
[12] Y. A. Huang and J. Benesty, “A multiframe approach to the frequency domain single channel noise reduction problem”, IEEE Trans. Audio, Speech and Language process., vol. ASSP, vol. 27,no. 2, **(1979)**, pp. 113–120.
[13] J. S. Lim and A. V. Oppenheim, “Enhancement and bandwidth compression of noisy speech”, roc. IEEE, vol. 20, no. 4, **(2012)**, pp. 1256–1269.
[14] P. C. Loizou, “Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum”, IEEE Trans. Speech Audio Process., vol. 13, no. 5, **(2005)**, pp. 857–869.
[15] P. Loizou, “Speech Enhancement: Theory and Practice”, Boca Raton, FL: CRC, **(2007)**.
[16] T. Lotter and P. Vary, “Noise reduction by maximum a posteriori spectral amplitude estimation with supergaussian speech modelling”, in Proc. Int. Workshop Acoust. Echo Noise control, **(2003)**, pp. 83–86.
[17] R. J. McAulay and M. L. Malpass, “Speech enhancement using a soft decision noise suppression filter”, IEEE Trans. Acoust., Speech, Signal Process., vol. 28, **(1980)**, pp. 137–145.
[18] E. Plourde and B. Champagne, “Generalized Bayesian estimators of the spectral amplitude for speech enhancement”, IEEE Signal Process.Lett., vol. 16, no. 6, **(2009)**, pp. 485–488.
[19] J. L. Roux, E. Vincent, Y. Mizuno, H. Kameoka, N. Ono and S. Sagayama, “Consistent Wiener filtering: Generalized time-frequency masking respecting spectrogram consistency”, in Proc. 9th Int. Conf. Latent Variable Anal. Signal Separat. (LVA/ICA), **(2010)**, pp. 89–96.
[20] M. R. Schroeder, “Apparatus for suppressing noise and distortion in communication signals”, U.S. Patent 3,180, vol. 936, **(1965)**.
[21] M. R. Schroeder, “Processing of communications signals to reduce effects of noise”, U.S. Patent 3,403, vol. 224, **(1968)**.
[22] P. Vary and R. Martin, “Digital Speech Transmission: Enhancement, Coding and Error Concealment”, Chichester, U.K.: Wiley, **(2006)**.
[23] M. R. Weiss, E. Aschkenasy and T. W. Parsons, “Processing speech signals to attenuate interference”, in Proc. IEEE Symp. Speech Recognition, **(1974)**, pp. 292–295.
[24] P. J. Wolfe and S. J. Godsill, “Efficient alternatives to the Ephraim-Malah suppression rule for audio signal enhancement”, EURASIP J. Appl. Signal Process, Special Iss.: Digital Audio for Multimedia Commun., **(2003)**, pp. 1043–1051.
[25] W. R. Wu and P. C. Chen, “Subband Kalman filtering for speech enhancement”, IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process, vol. 45, no. 8, **(1998)**, pp. 1072–1083.
[26] C. H. You, S. N. Koh and S. Rahardja, “M-order MMSE spectral amplitude estimation for speech enhancement”, IEEE Trans. Speech Audio Process, vol. 13, no. 2, **(2005)**, pp. 475–486.

Authors

Harjeet Kaur, obtained her Bachelor's degree in Electronics & Communication from Agra University. Then she obtained her Master's degree in Electronics & Communication and PhD (pursuing) in Electronics in Signal Processing from Punjab Technical University, Jalandhar, Punjab, India. Currently, she is an Assistant Professor at the Indira College of Engg. & Mgmt. Pune. Her specializations include Digital Signal Processing, Adaptive Digital Signal Processing, and Speech Processing.

Rajneesh Talwar, received the B.tech. Degree in 2001 from Aurangabad University, the M. tech. Degree from Thapar University, in 2002, and the Ph.D. Degree from the Thapar University in 2010. He has a Professional experience of more than ten years in teaching and research. He is currently work as a Principal at Chandigarh Group of Colleges, Landran His research interests are fiber optics, semiconductor devices, communication He has a U.S patent "FIBER OPTIC POINT TEMPERATURE SENSOR" to his credit, He is Editorial board member of International Journal of Engg. Science and Technology, Nigeria, was Reviewer of MAEJO International Journal of Engineering Science and Technology, Thailand and "Materials and Design", a ELSEVIER International Journal.

