International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.9, (2016), pp.317-324
http://dx.doi.org/10.14257/ijsip.2016.9.9.29

Speech Enhancement Using Iterative Kalman Filter with Time and Frequency Mask in Different Noisy Environment

G. Manmadha Rao[1] and Ummidala Santosh Kumar[2]

[1]Associate Professor, Department of ECE
[2]PG Scholar,Department of ECE, IEEE Student Member
GMR Institute of Technology
Rajam, AndhraPradesh, India
[1]manmadharao.g@gmrit.org, [2]uskscientist@gmail.com

Abstract

The main aim of the Speech Enhancement algorithms is to improve the Quality of speech. The Quality of speech is expressed in two parameters. One is clarity, and another is intelligibility. In this paper, we proposed a method to improve the quality of speech based on computationally efficient AR modeled Iterative Kalman Filter with time and frequency mask. This approach is based on reconstruction of noisy speech signals using Auto Regressive modeled Kalman filter and further to reduce artifact noise time and frequency mask is applied to the Kalman filter output. The results of the proposed method are found to be better compared to spectral subtraction, wiener filter and Kalman filter methods.

Keywords: *Kalman filter; intelligibility; Spectral Subtraction; Wiener filter; speech enhancement; white-colored noises*

1. Introduction

Speech is the primary mode of Communication among all human beings. It is very efficient and effective way of communication. The Speech processing is widely used in many applications like mobile phones, VOIP, Teleconferencing systems, voice enabled security devices, household appliances, speech recognition systems, hearing aids, biomedical signal processing, ATM machines and computers [1]. Various types of additive noise in real-world environments often corrupt speech. Unfortunately, the characteristics of this additive noise are difficult to estimate due to it has different characteristics in different environments. So speech enhancement is very much required. There are many speech enhancement techniques in stationary and non- stationary noisy environments. Some of them are Spectral Subtraction method, Wiener filtering, and Subspace based methods and so on. As a most fundamental technique Spectral subtraction is a method for restoration of the power spectrum or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum [2].

Spectral subtraction is one of the traditional methods used for enhancing speech degraded by additive stationary background noise, but a common problem for the Spectral Subtraction method is the characteristic of the residual noise called musical noise. Spectral subtraction also does not attenuate noise sufficiently during the silence period [3]. The Wiener filter is a linear filter employed to recover the original speech signal from the noisy signal by reducing the Mean Square Error (MSE) between the estimated signal and the original one with the help of transfer function [4]. However, after application of the algorithm, speech quality is improved, but the musical noise still influences the speech quality [5]. Paliwal and Basu have used an estimation of speech signal parameters for clean speech, before it gets disturbed by white noise [6]. A time-adaptive algorithm to

ISSN: 2005-4254 IJSIP

adaptively estimate the speech model parameters and noise variance is used by Oppenheim et al [7]. Expectation-Maximization algorithm iteratively estimate the Spectral parameters of speech and noise parameter have proposed by Gannot[8]. There are lots of changes made to the basic above stated algorithms by many authors, but all this doesn't meet the expectations.

In this paper a new adaptive or Iterative Kalman filter based method with post processing of time-frequency mask is proposed to recover the speech signal from a sequence (frame) of noisy speech signals and the additive noise is modeled as the AR process [9].this estimation of time-varying auto regressive (AR) speech model parameters are based on linear prediction coefficient estimation (LPC).in addition to coefficients estimation this paper solved problem of de-noising the colored noise. We made an assumption that the noise is also an autoregressive process [10]. So we estimated its AR coefficients and variances by LPC in the same way. After that time-frequency mask is applied as a post-filter to this Iterative Kalman filter.

The paper is organized as follows. In Section 2 we present the speech enhancement approach based on the Kalman filter algorithm and time-frequency mask mathematically. Section 3 is concerned with Implementation and evaluation of the proposed method. Simulation results are placed in Section 4.

2. Mathematical Description

Kalman filtering is one of the effective speech enhancement technique, in which speech signal is usually modeled as autoregressive (AR) model and represented in the state-space domain. Kalman filter based approaches proposed in the past, operate in two steps. First estimate the noise and the driving variances and parameters of the signal model, by using these parameters estimate the speech signal. Due to low SNR and the speech intelligibility degradation of Kalman filter based speech enhancement We proposed this method.

Iterative Kalman filter:

The speech signal $s(n)$ and the additive noise $v(n)$ are expressed in terms of p th order autoregressive (AR) model as below

$$s(n) = \sum_{i=1}^{p} a_j s(n-i) + u(i) \tag{1}$$

$$v(n) = \sum_{\mathrm{j=1}}^{\mathrm{p}} \mathrm{b_j x(n-j) + w(i)} \tag{2}$$

And Noisy speech can be expressed as

$$y(n) = s(n) + v(n) \tag{3}$$

Where $s(n)$ is the n th sample of the speech single, $v(n)$ is the n th sample of the additive noise, $y(n)$ is the n th sample of noisy speech. a_j And b_j or AR model parameters.

State-space form of above AR model is

$$s(n+1) \quad = \quad A(n)s(n) + (u(n), 0, \dots\dots\dots, 0)^T \tag{4}$$

$$A(n) = \begin{bmatrix} a_1(n) & \dots & a_p(n) & 0 & \dots & 0 & 0 \\ 1 & \dots & 0 & 0 & \dots & 0 & 0 \\ \vdots & \ddots & & & & & \vdots \\ 0 & & 1 & 0 & \dots & 0 & 0 \\ 0 & & 0 & 1 & \cdots & 0 & 0 \\ \vdots & & & & \ddots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{5}$$

$$v(n+1) \quad = \quad B(n)s(n) + (w(n), 0, \ldots\ldots\ldots., 0)^T \tag{6}$$

$$B(n) = \begin{bmatrix} b_1(n) & \ldots & b_{q-1}(n) & b_q(n) \\ 1 & \ldots & 0 & 0 \\ \vdots & \ddots & 0 & \vdots \\ 0 & \ldots & 1 & 0 \end{bmatrix} \tag{7}$$

From the above discussion the augmented state vector $X(n)$ and driving noise vector $W(n)$.

$$X(n) = \begin{pmatrix} s(n) \\ v(n) \end{pmatrix}, \quad W(n) = \begin{pmatrix} u(n) \\ w(n) \end{pmatrix} \tag{8}$$

From Eq.(3) and (4)

$$X(n+1) = F(n)X(n) + GW(n) \tag{9}$$

$$y(n) = C^T X(n)$$

Where $F(n) = \begin{pmatrix} A(n) & 0 \\ 0 & B(n) \end{pmatrix}$ $\quad G = \begin{pmatrix} e_s & 0 \\ 0 & e_v \end{pmatrix}$ $\quad C = \begin{pmatrix} e_s \\ e_v \end{pmatrix}$

$e_s = (1, 0, \ldots\ldots 0,)^T$ with d+1 dimension and e_v with q dimension.Now we can optimally suppresses the disturbing noise by calculating Kalman filter basic parameters such as variance and gain [8].

The process of Iterative Kalman filter [8] is in two steps. 1. Estimation: state vector propagation, parameter covariance matrix propagation and 2. Updating: compute Kalman gain, state vector update, parameter covariance matrix update.

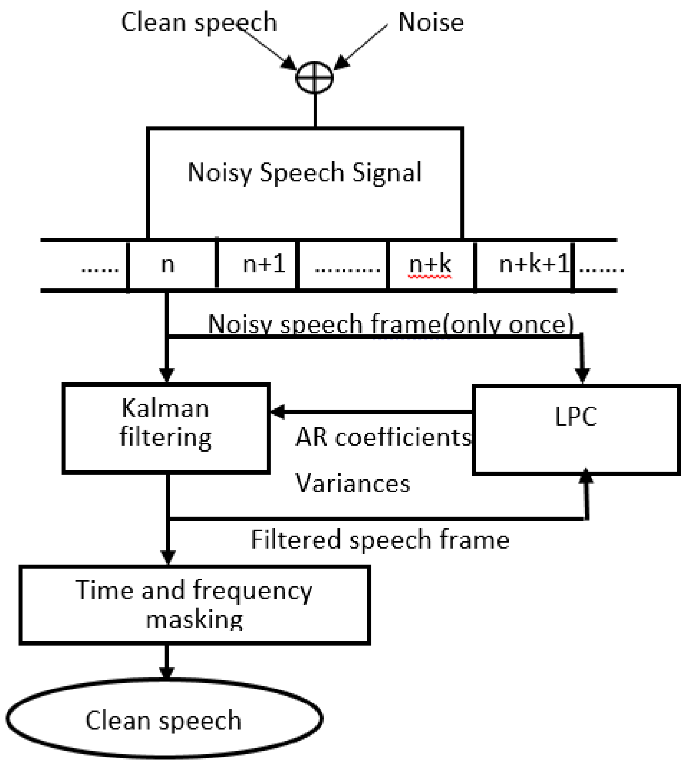

Figure 1. Block Diagram for Iterative Kalman Filter with Time and Frequency Masking

Time-Frequency making:

After applying noisy speech to the Iterative Kalman filter we got better results than spectral and wiener filter methods. It is completely removing background noise and producing clean speech for number of iteration in-between 3-5. But if the no. of iteration is more than 5 then more noise is reduced, but the intelligibility of speech is degraded. So to improve intelligibility of speech time-frequency mask with a weighting factor of 1/(1+gamma*diff (t, f)) is used as a post process to the iterative Kalman filter. Time-Frequency mask have been found useful by the various speech processing researchers to remove unwanted energy [11-12]. Here a mask of weights is applied to a time-frequency by multiplying the values in the cells by weights. Here time-frequency grid mean a spectrogram type representation.

The weights are set closer to 0 the more unwanted energy was judged to be in the cell.

To find the spectrogram cells where there is more energy after the IKF (Iterative Kalman Filter) than before the IKF processing, the code subtracts the spectrally normalized magnitude spectrogram of the input to the IKF (Iterative Kalman Filter) from that of the ITF output, setting negative values to zero.

$$diff(t,f) = max(0, outMagNormed(t,f) - inMagNormed(t,f) \quad (10)$$

The time - frequency mask is expressed as

$$mask(t,f) = 1/(1 + gamma * diff(t,f)) \quad (11)$$

The mask is then applied to the magnitude spectrogram (the non-spectrally-normalized magnitude spectrogram in the following fragment.

$$outmagMasked(t,f) = outMag(t,f) * mask(t,f) \quad (12)$$

Finally, $outmagMasked$ is combined with the phase spectrogram of the ITF output to create a complex DFT spectrogram, from which a time-domain waveform is calculated

3. Implementation and Evaluation of Proposed Method

The figure shows the flow of the proposed method. Matlab code is developed, where Kalman filter is applied to different real time noisy signals taken from Noizeus database. In this method first Noisy speech signal and noise signals are modeled by an AR model of order P=20. These 20 AR coefficients are updated for every analysis frame of 25ms duration which is obtained from chopped the noisy speech signal into 25ms duration with the help of Hanning window and analyzed using the linear prediction analysis method (LPC).

The additive measurement noise is assumed to be stationary during the each small frame. LPC coefficient estimation, order is taken as 13 for both noisy speech and noise signals. Number of iterations are set to be 4. If it is more than 4 it is removing more noise, but speech intelligibility is degrading, so it is set to be at 4. In my experiments with the ITF started with the way of calculating the weights that are currently in the Matlab code (i.e., as the weighting factor) and I made an adjustment increasing gamma from starting value of gamma=1 to 100. We found better performance at 100. So we considered gamma is 100. Perhaps a higher or lower value of gamma will be better for post-processing of other algorithms or other data.

4. Results

In this paper, we have observed and tabulated the results of Spectral Subtraction, Wiener Filter Kalman filter methods and compared with the Iterative Kalman filter with time and frequency mask method. Here different real time noise signals (From NOIZES database) of 0dB, 5dB , 10dB and 15 dB are considered to process using a Hanning window in above three stated algorithms. Compared to all these methods, proposed algorithm giving best results in SNR. Speech, Noise and Expected waveforms are shown below figure. Experimental results show that the proposed technique is effective for speech enhancement compare to conventional Kalman filter.

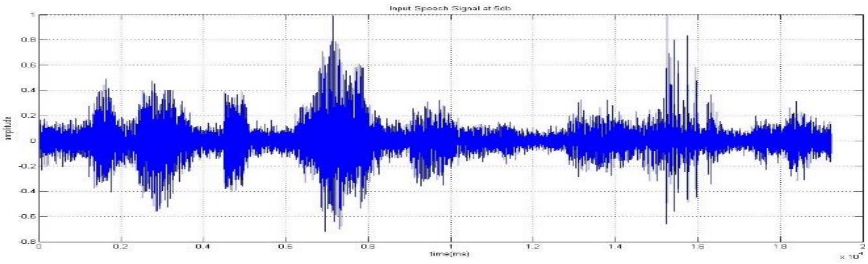

Figure 1. Input Signal_5db

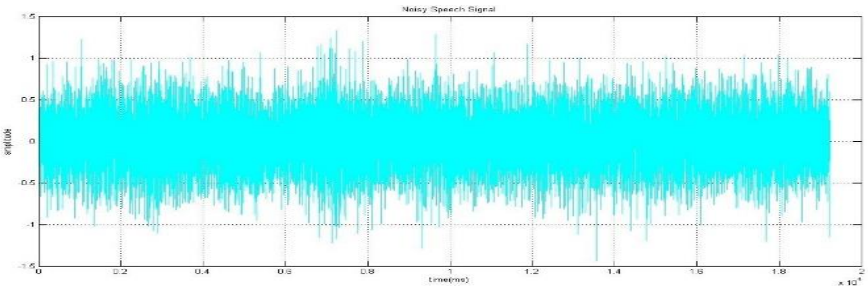

Figure 2. Noisy Speech Signal

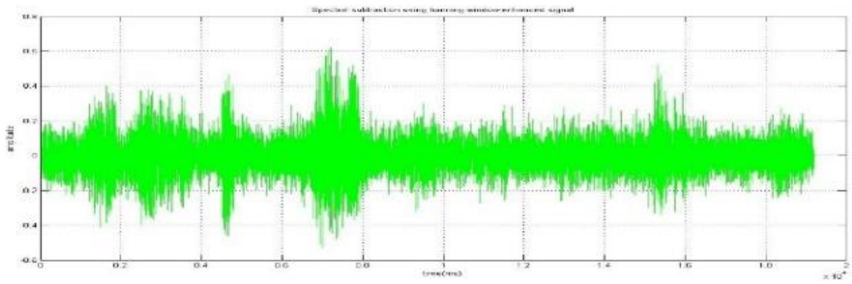

Figure 3. Enhanced Signal (at 5dB) Using a Spectral Subtraction Method

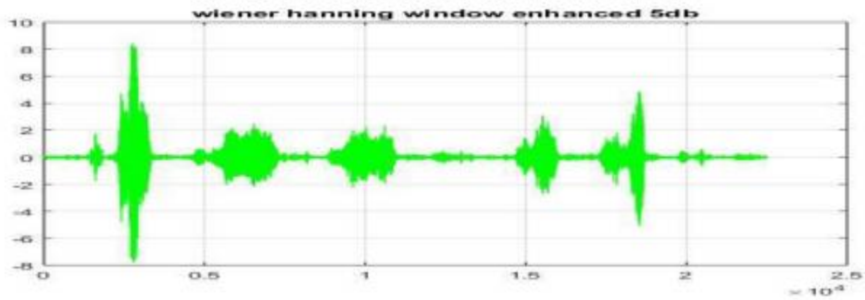

Figure 4. Enhanced Signal Using Wiener Filtering at 5dB

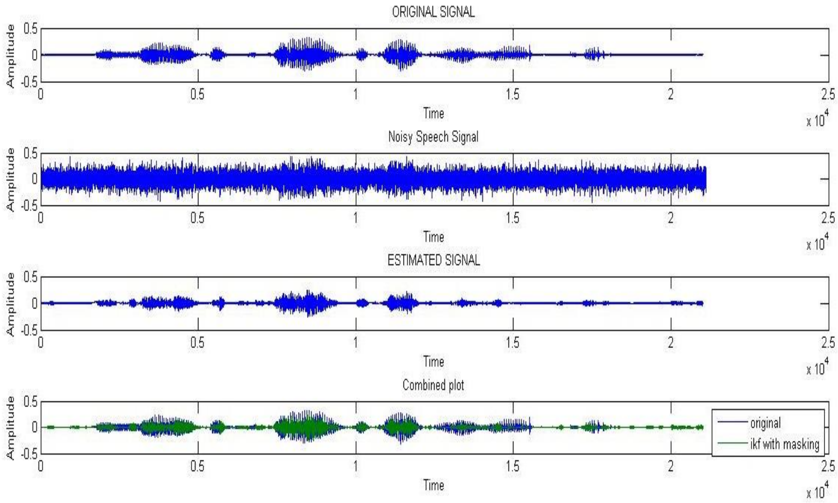

Figure 5. Iterative Kalman with Time-Frequency Mask Output Speech Wave Forms

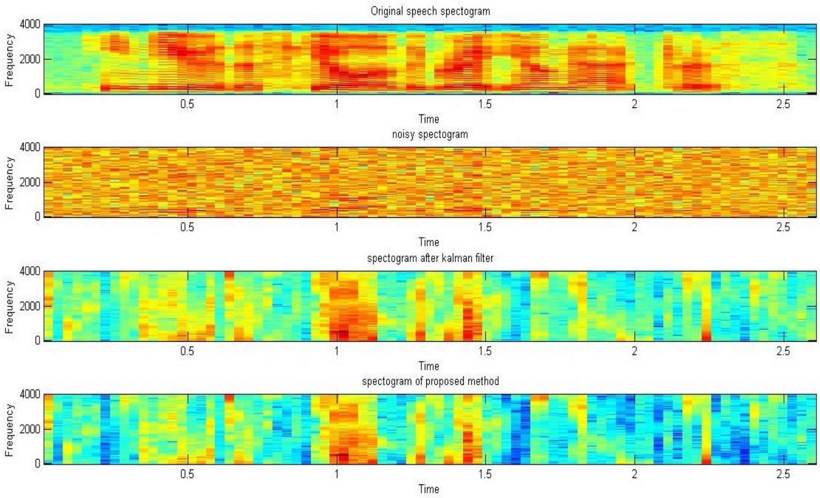

Figure 6. Spectrograms of Speech, Noise and Output Waveforms of Proposed Method

Table 1. Comparison of Input-Output SNR of Different Speech Enhancement Algorithms

OUTPUT SIGNAL TO NOISE RATIO RESULTS(white noise)				
method/noise level	clean speech	0dB	5db	10db
Spectral Subtraction	2.9245	1.8926	2.2142	2.6056
wiener filter	2.9325	1.8992	2.3523	2.7021
Iterative Kalman	2.6702	1.6521	2.3852	2.6595
proposed method	3.0288	1.8998	2.406	2.8091

Table 2. SNR Analysis (Color Noise) of Proposed Method in Different Noisy Environment

NOISE TYPE	INPUT SNR	SPECTRAL SUBTRACTION	KALMAN	PROPOSED METHOD
CLEAN SPEECH	Random	-1.9352	0.0228	0.0358
CAR NOISE	Car 15dB	-2.0055	0.0054	0.0065
	Car 10dB	-2.1407	0.0041	0.0043
	Car5db	-2.3057	0.0020	0.0018
	Car0dB	-2.4310	-0.0803	0.0009
BABBLE NOISE	Babble 15db	-2.0112	0.0032	0.0035
	Babble10db	-1.9310	-0.2047	0.0021
	Babble 5db	-2.6615	-0.2588	-0.1002
	Babble 0db	-2.3986	-0.3121	-0.2253
RESTAUTANT NOISE	Restaurant 15db	-1.9167	-0.0220	0.0025
	Restaurant 10db	-2.2211	-0.0250	-0.0098
	Restaurant 5db	-2.3174	-0.0746	-0.0958
	Restaurant 0db	-2.2415	-0.2174	-0.1995
STATION NOISE	Station 15db	-2.1123	-0.0208	0.0036
	Station 10db	-2.3563	-0.0198	-0.0075
	Station 5db	-2.3748	-0.0058	-0.1546
	Station 0db	-2.3974	-0.1092	-0.2584

5. Conclusion

In the present study, based on Adaptive Kalman filter, an improved method of Iterative Kalman filter with time and frequency mask is proposed. In this paper, we discussed the drawbacks of speech enhancement with spectral subtraction and wiener filter methods. Even though the conventional Kalman filter approach is giving better results than spectral and wiener but its SNR is low. In this paper, we implemented iterative Kalman filter with time and frequency mask which overcome the disadvantages of early two methods. All these methods are simulated and SNR values of respective methods are compared. It is observed that the proposed method giving better SNR values and its performance is comparatively superior for both stationary and non-stationary signals.

References

[1] J. Benesty, S. Makino, and J. Chen, Eds., "Speech Enhancement", Springer, Berlin, **(2005)**.

[2] Muhammad T. Sadiq,Noman Shabbir and Wlodek J.Kulesza, "Spectral Subtraction for Speech Enhancement in Modulation Domain," IJCSI International Journal of Comuputer science, vol.10, Issue 4, no. 2, July **(2013)**.

[3] Ganga Prasad, Surendar "A Review of Different Approaches of Spectral Subtraction Algorithms for Speech Enhancement" Current Research in Engineering, Science and Technology (CREST) Journals, vol. 1, no. 02, April **(2013)**, pp. 57-64

[4] A. Spriet, M. Moonen and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction", Signal Processing, vol. 84, no. 12, **(2004)**, pp.2367-2387.

[5] M. J. Alam, D , "Perceptual improvement of Wiener filtering employing a post-filter", Digital Signal Processing, vol. 21, no. 1, **(2010)**, pp. 54-65.

[6] K. K. Paliwal and A.Basu, "A speech enhancement method based on Kalman filtering, Proceedings of ICASSP'87, pp.177-180. Dallas, TX, USA, **(1987)**.

[7] A. V. Oppenheim, E. Weinstein, K. C. Zangi, M. Feder, and D. Gauger, "Single-Sensor Active Noise Cancellation," IEEE Trans. Speech and Audio Processing, vol. 2, **(1994)** Apr. , pp. 285–290.

[8] B. Schwartz, S. Gannot and Emanuel A. P. Habets,"Online speech dereverberation using Kalman filter and EM algorithm',IEE Transction on Audio,Spech and Language Processing ,2013
[9] T. Yoshioka, M. Miyoshi and T. Nakatani "Integrated speech enhancement method using noise suppression and dereverberation," IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 2, February **(2009)**, pp. 231–246.
[10] Jerry D.Gibson, "Filtering of colored noise for speech enhancement and coding" IEEE Trans. Signal processing, vol.39, no. 8, **(1991)** August.
[11] P. S. Whitehead, D. V. Anderson and M. A. Clements, "Adaptive, acoustic noise suppression for speech enhancement", ICME, **(2003)**.
[12] Post-processing of dereverberation/denoising algorithms to reduce artifact noise, using a time frequency mask (David Gelbart, 2004-2005)

Authors

Dr. G. Manmadha Rao, obtained his BE, ME and PhD degrees from Department of Electronics and Communication Engineering, AU College of Engineering, Andhra University, Visakhapatnam in 1998, 2003 and 2014 respectively. He worked in different industries for 11 years. He is in teaching profession for more than 13 years. Presently he is working as Associate Professor in the Department of Electronics and Communication Engineering, GMR Institute of Technology, Rajam, Srikakulam, India. He is coauthor for text books Pulse and Digital Circuits published by Pearson Education. His research interests in RADAR Signal Design and Processing, Speech Signal processing and VLSI. He has published more than 25 papers in various National, International Journals and conferences. He has guided more than 8 PG students and presently guiding 1 PhD student and 2 PG students.

Santosh Kumar Ummidala, received the Bachelors degree in Electronics and Communication Engineering from Sri Venkateswara college of Engineering and Technology,Andhrapradesh, India . He is currently pursuing M.Tech degree in the Department of Electronics and Communication Engineering (DECS) in GMR Institute of Technology from Andhrapradesh,India. His Research interests are Speech Enhancement,Speech Recognition techniques.

