Speech Enhancement Using Combination of Digital Audio effects with Kalman Filter

Ummidala Santosh Kumar

PG Scholar, Department of ECE
IEEE Student Member
GMR Institute of Technology, India
uskscientist@gmail.com

Dr. G. Manmadha Rao

Associate Professor, Department of ECE
GMR Institute of Technology
Rajam, Andhra Pradesh, India
manmadharao.g@gmrit.org

Abstract— **The term "Quality of Speech" in Speech Enhancement techniques is associated with Clarity and Intelligibility. Till now due to the variable nature and characteristics of noise with time and process to process, Speech Enhancement is a difficult problem in Noisy environment. In this paper, we proposed a method to improve the quality of speech based on combination of Digital Audio Effects with Improved Adaptive Kalman Filter when only corrupted speech is available. In this approach to enhance the Speech content in the Noisy speech signal, Digital audio effects are used. A Digital Expander generates an audio effect which operates on a low signal level and create more likely sound characteristics. And further, noise is removed by Auto Regressive modeled improved adaptive Kalman filter. The performance of the proposed method with additive color noise is found to be better compared to other spectral subtraction, wiener and Kalman filter methods in terms of Signal-to-Noise ratio and intelligibility.**

Keywords— Kalman filter; intelligibility; Digital audio effect; digital expander; Wiener filter; Spectral Subtraction..

I. INTRODUCTION

The primary objective of many Speech Enhancement algorithms is to improve the perceptual quality of extracting speech signal from noisy speech. Noise estimation is the major component in speech enhancement techniques, because better noise estimation gives a high quality of speech extraction. Till now, removing noise from noisy speech is challenging issue because spectral properties of non-stationary noise is very difficult to estimate and predict. Noise estimation is a careful issue in speech enhancement algorithms since if the noise power is more than speech power, then that speech content may be removed due to treating that as a noise. Due to the wide use of Speech processing in many applications like teleconferencing systems, speech recognition based security devices, biomedical signal processing, hearing aids, ATM machines and computers, Speech enhancement is a hot research area in signal processing and remains a challenging issue because of most of the cases only the noisy speech is available [1]. Over the past years, researchers have developed different types of efficient algorithms to improve the noisy speech even though still it poses a challenge to the researches because of characteristics of noise signal varies in a dramatic manner over time and application to application. There are many speech enhancement techniques are proposed using filtering approach by researchers last ten years such as spectral subtraction method, wiener filtering, Kalman filter method and so on.

Spectral subtraction is used for enhancing speech degraded by additive stationary background noise, but it is affected by musical noise and also it does not remove noise during the silence period [2]. In Wiener filter based speech enhancement method original speech signal is recovered by minimizing Mean Square Error (MSE) between the clean speech and the estimated signal [3]. Spectral and wiener filter based speech enhancement algorithms require the characteristics of clean speech. But in real time clean speech may not available in all the cases. From the literature study, we found that some of the techniques have been proposed to enhance the speech. In [4] using harmonic structure of speech signal, speech is recovered form noisy speech signal, in [5] sinusoidal model is adopted, in [6] MMSE estimator to enhance the speech was introduced by Ephraim. At first The advantage in the use of Kalman filter for speech enhancement was proposed by K.K Paliwal and A.Basu by using estimation of speech signal parameters from clean speech before it corrupted by white noise is proposed in [7]. And further extended to the random and colored noises [8]. In these methods a tradeoff should be maintained between SNR and intelligibility.

Later, many changes are made to the Kalman filter for better improvement, it does not meet the expectations and also complexity is more. In this paper with less complexity and better performance a new adaptive Kalman filter based method with the combination of nonlinear digital filter called digital expander is proposed to recover the speech signal from noisy speech. The additive noise is modeled as the AR process based on linear prediction coefficient estimation (LPC) in Kalman filtering algorithm [7]. In addition to coefficient estimation this paper solved problem of de-noising the random and colored noises. We considered an assumption that the colored noise is also an autoregressive process [8]. So we estimated its AR coefficients and variances by linear prediction estimation in the same way.

In this paper, to overcome above stated problem a new adaptive Kalman filter based method with preprocessing of a digital audio effecting technique called digital expander 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 coefficient

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.

In this paper the content is organized as follows. In Section II we mentioned the theoretical and mathematical description of proposed method. Section III is deals with Implementation and evaluation of the proposed method. Simulation results are placed at the end.

II. MATHEMATICAL DESCRIPTION

Digital Audio Effects:

Digital Audio Effects can be classified as Basic Filtering, Time Varying Filters, Delays, Modulators, Non-linear Processing, Spacial effects. Non-linear Digital Filters are characterized by creating harmonic and inharmonic frequency components which are not present in the original signal intentionally or unintentionally. In Dynamic Processing signal envelope is controlled to minimize harmonic distortion using compressors or limiters.

Digital expander is a signal limiter which minimizes the distortion in the speech. Expander operates at low signal levels to boost the dynamics of the signal and it is useful to create a more likely sound characteristic [9].

The signal x(n) is determined from the input with variable attack and release time data. The logarithm of this x(n) signal is compared with the threshold value. If the signal is above the threshold, then the difference is multiplied by the negative slope of the limiter LS. Then the output is applied to antilogarithm. The control factor $f(n)$ obtained is then smoothed with a first-order low pass filter. If the signal $x(n)$ lies below the threshold level, then the signal $f(n)$ is set to $f(n) = 1$. The delayed input $x(n - D1)$ is multiplied by the smoothed control factor $g(n)$ to give the output $y(n)$. The Figure (a) shows a digital expander block diagram. The logarithm of the signal $x_{RMS}(n)$ is taken and multiplied by 0.5. The value obtained is compared with two thresholds in order to determine the operating range of the static curve. If thresholds are crossed, the resulting difference is multiplied by the corresponding slope and antilogarithm of the result is taken. A first-order low-pass filter subsequently provides the attack and release time.
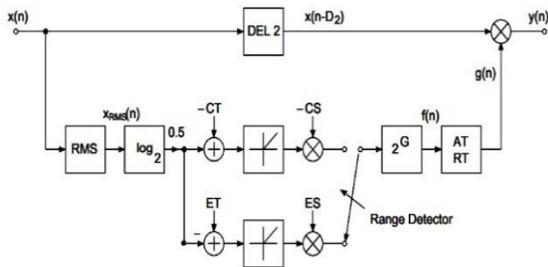

Figure (a): block diagram of Digital Expander

Kalman Filtering:

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. A Kalman filter is an estimation and updating process.

In this process both the speech signal and the additive noise signals are treated as $s(n)$ and $v(n)$ respectively and expressed in terms of pth order autoregressive model (AR) as follows

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

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

And Noisy speech can be expressed as

$$s(n) = x(n) + v(n) \qquad (3)$$

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

AR modeled speech signal can be expressed in State-space form shown below.

$$x(n + 1) = A(n)x(n) + (u(n), 0, \dots \dots \dots, 0)^T \qquad (4)$$

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

$$v(n + 1) = B(n)s(n) + (w(n), 0, \dots \dots \dots, 0)^T \qquad (6)$$

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

From above equations state vector $X(n)$ and driving noise vector $W(n)$ can be written as.

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

From Eq.(3) and (4)

$$X(n + 1) = F(n)X(n) + GW(n) \qquad (9)$$

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

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

$e_s = (1, 0, \ldots \ldots 0,)^T$ with d+1 dimension and e_v with q dimension. From[8] noise suppression can be done by calculating the variance, Kalman gain.

Estimation: state vector propagation, parameter covariance matrix propagation

Iterative Kalman filter time updating process is done by following equations

$$\hat{x}_k^- = A\hat{x}_{k-1} + B\hat{u}_{k-1} \qquad (10)$$

$$P_k^- = A\hat{P}_{k-1}A^T + Q \qquad (11)$$

Updating: compute Kalman gain, state vector update, parameter covariance matrix update

The coefficients in the above equations are updated every time frame by using following Discrete Kalman filter update equations

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \qquad (12)$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \qquad (13)$$

$$P_k = (1 - K_k H) P_k^- \qquad (14)$$

These parameters are updated for each iteration.

III. IMPLEMENTATION AND EXPERIMENTAL RESULTS

Figure (a) shows the block diagram for combination of Digital Audio Effect with modified adaptive Kalman filter based speech enhancement method. Based on it Matlab code is developed. A noisy speech is generated using a Clean Speech, which is taken from the Noizeus Database and random values (random noise or color noise) are added to the clean speech. Later it passed through a Digital Expander. And the output of Digital Expander is shown in Figure3. Digital Expander expanding factor value is set to 0.5. It is further applied to Iterative modified Kalman filter to suppress the noise. We set the Kalman filter AR model order to P=20. These 20 AR coefficients are updated for every time frame of 25ms duration which is chopped by 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 7. Real time noisy signals (NOIZEUS database) of 0dB, 5dB, 10dB and 15dB are considered for performance analysis, with Hanning window. We have observed and tabulated the results of basic Spectral Subtraction, Wiener Filter, Kalman filter methods and compared with Digital Audio Effect based Kalman filtering method. Compared to all these methods, proposed algorithm giving better improvement in terms of SNR as well as

intelligibility. The corresponding waveforms are shown below. Experimental results show that the proposed technique is effective for speech enhancement compare to conventional Kalman filter. Iterative Kalman filter and proposed method results and waveforms are placed below.
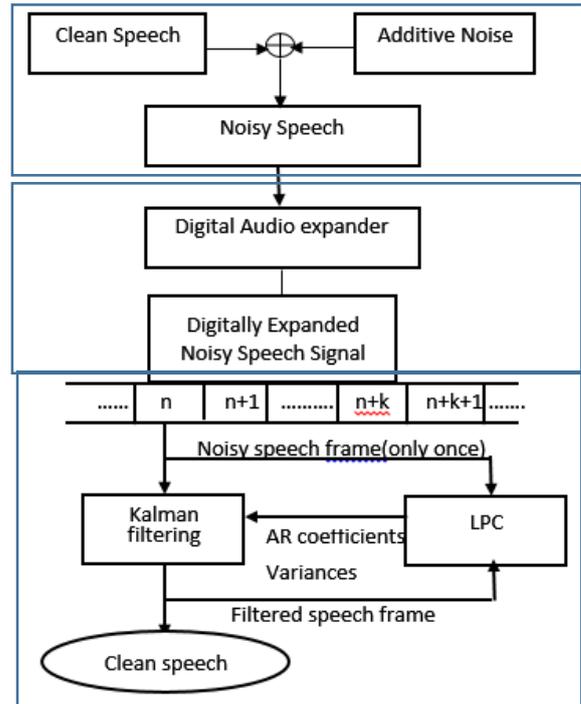

Figure (a): Block diagram for combination of Digital Audio Effect with modified adaptive Kalman filter

Figure (4) is output waveform of Iterative adaptive Kalman filter, Figure (5) is output of proposed method(combination of digital audio effects with Kalman filter), Figure(6) is comparison between clean speech and recovered speech using proposed method, Figure (7) is related to output Spectrogram waveforms of Kalman and Proposed method, Table (1) represent corresponding SNR ratios.
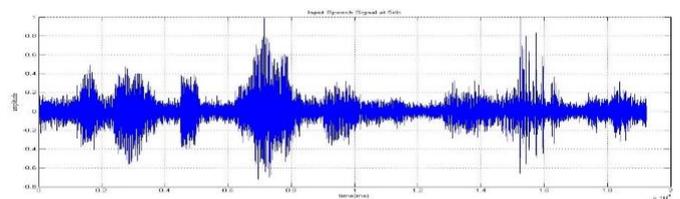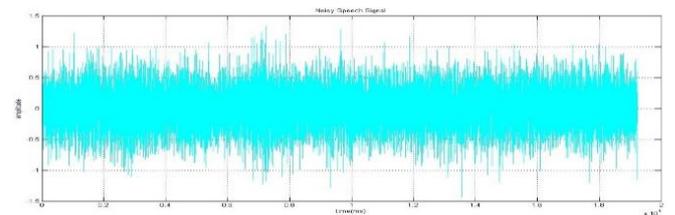

Figure 1: Input signal_5db
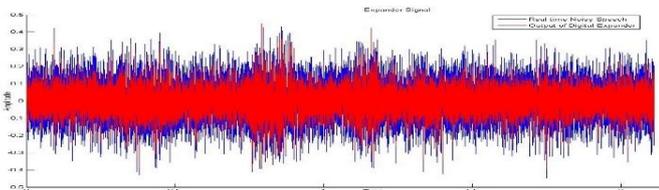

Figure2: Noisy Speech Signal

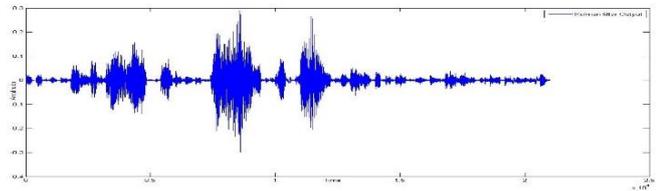Figure3: Noisy speech Vs Digital Expander output

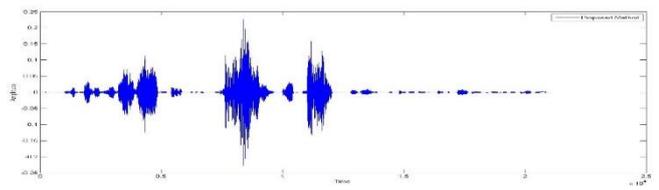Figure (4): Kalman Filter based Speech Enhancement method Output waveform

Figure (5): Combination of Digital Audio Effect with Kalman Filter based Speech Enhancement recovered Output
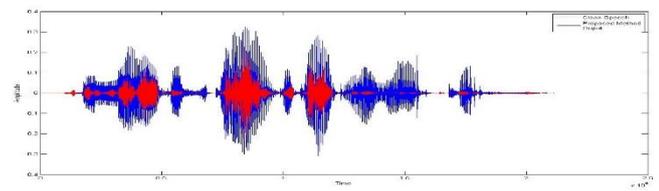

Figure (6): Proposed Method output (Kalman with Digital Expander) and clean speech waveforms
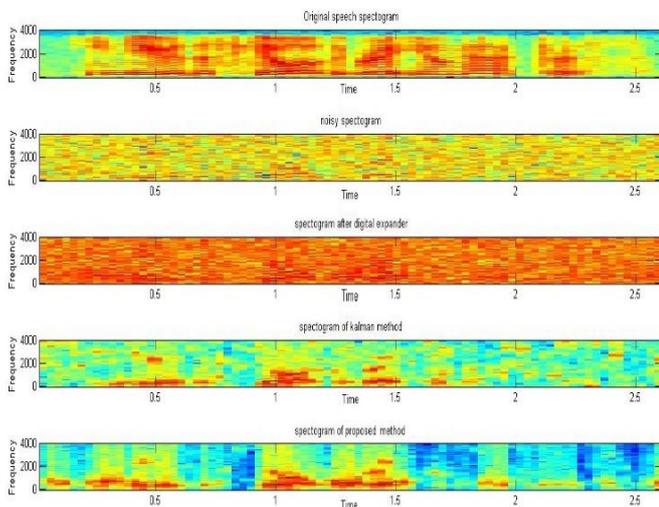

Figure (7): Spectrogram of clean & noisy speech signals, Kalman filter method, and proposed method

Noise/ Method	Spectral Subtraction	Wiener filter	Kalman filter	Proposed Method
Clean speech	-7.0125	-6.3115	-3.2859	-1.4122
0dB	-10.5875	-9.0075	-4.6595	-2.5242
5dB	-8.5654	-7.9854	-3.5105	-2.2545
10dB	-7.4850	-7.0985	-3.4589	-2.0587
15dB	-7.2106	-6.6002	-3.2857	-1.5600

Table1: Input-Output SNR Comparisons

IV. V. CONCLUSION

In the present study, an improved method for speech enhancement by combining Digital Audio Effecting techniques with improved Adaptive Kalman filter technique is proposed. In this paper, we discussed the drawbacks of basic methods such as speech enhancement with spectral subtraction and wiener filter methods. Even though other Kalman filter approach based speech enhancement methods are giving better results than a conventional Kalman filter, more complexity is involved.it leads to more time taken process. In this paper, we proposed a method to overcome the disadvantages of earlier methods in terms of performance and Speed. All these methods are simulated using MATLAB and input output SNR values of respective methods are compared. Performance of Proposed method is analyzed with different Input SNR noise level. It is observed that the proposed method gives better output SNR values and its performance is comparatively superior for both stationary and non-stationary signals.

References

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

[2] S.Boll,"Suppression of acoustic noise in speech using spectral subtraction," IEEE Transaction on Speech, Signal Process., volume. ASSP-27,no.2 pp.113-120,1979

[3] T.V.Sreenivas and P.kirnapure,"Codebook constrained Wiener filtering for speech enhancement", IEEE Trans. Speech and Audio processing,sep 1996.

[4] D.V.Anderson and M.A.Clements "Audio signal noise reduction using harmonic modeling," in Proc.IEEE Int. conf. Acoust. ICASSP,1999

[5] by J.Jensen and J.H.L Hansen,"Speech enhancement using a constrained iterative sinusoidal model,"IEEE Trans. Speech Audio Process., vol.9 ,2001.

[6] Y.Epharaim,"A minimum mean square error approach for speech enhancement,"in Proc. IEEE conf 1990.

[7] 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.

[8] B.Chabane and B.Daoued"On the use of Kalman Filter for enhancing speech corrupted by Colored Noise," WSEAS Trans. On signal processing 2008 Dec.

[9] "Digital Audio Signal Processing," second edition, by Udo Zolzer

[10] "Noizeus- A noisy speech corpus for evaluation of speech enhancement algorithms," http://ecs.utdallas.edu/loizou/speech/noizeus

[11] Hu.Y and Loizou.P,"Evaluation of objective quality measures for speech enhancement," IEEE Trans. On speech and audio processing, 2008