

# Speech Enhancement in Noisy Environment using Kalman Filter

Erukonda Sravya<sup>1</sup>, Rakesh Ranjan<sup>2</sup>, Nitish J. Wadne<sup>3</sup>

<sup>1, 2</sup>Assistant professor, Dept. of ECE, CMR Engineering College, Hyderabad (India)

<sup>3</sup>PG Student, Dept. of E&TC, Vishwakarma Institute of Information Technology, Pune (India)

## ABSTRACT

The speech enhancement in noisy environment is a challenging research field with various applications. Noise removal plays a vital role in applications like telephone conversation, speech recognition, etc. 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, i.e., if the noise is more stationary than the speech, it is easy to estimate the noise during the pauses in speech. If the noise is varying rapidly then the estimation of noise in the noisy speech is more difficult. The main aim of this work is to investigate the enhancement of speech by applying Kalman filter. The performance of this filter is better as compared with traditional methods: it is found that it can give better results at the expense of execution speed and has good optimization.

**Keywords:** Adaptive Filter, Kalman Filter, NOZIEUS database, Non-Stationary Noise, Speech Enhancement, Windowing and Non-windowing technique.

## I. INTRODUCTION

Speech plays a vital role in our daily communication and also for human machine interfacing. Therefore, production and perception of speech have become an interesting part of the research since decades. But the quality and intelligibility of the speech are significantly degraded by the presence of background noise, which affects the ability in understanding other's speech, causes error in Human Machine Interfacing, etc. In this digital world, it's really hard for any signal in real-time environment to escape from noise. This hits us really hard when it comes to deliver a message from one place to another and there is a need for cleaning up or enhancing the message signal but at the same time, not giving up any intelligibility of the message (content, not just clarity). Since speech messages have been the mode of communication everywhere, need for speech enhancement is required whenever the signal comes in contact with the real-time environment.

Modelling of human speech production process helps in enhancing the speech. But, as speech is a highly non-stationary signal, it is difficult to model the human speech production process. Though speech is highly non-stationary signal, it is stationary for very short period of time [1, 3, 6]. Based on this fact, Classical speech enhancement techniques are considered for speech segment models for short time, but these short time models do not include the effects of the noise as noise has long term characteristics. On the other hand, such long-term characteristics are naturally taken care of in the autoregressive approach as speech signals are not modelled on a short-time basis but as a whole. The AR model is also known to be good for representing unvoiced speech [2].

However, it is not quite appropriate for voiced speech since voiced speech is often quite periodic in nature. This has motivated us to look into speech models which can satisfactorily describe both voiced and unvoiced speech, and allow for exploitation of the long-term characteristics of noise.

Speech enhancement is an area of speech processing where the goal is to improve the intelligibility and/or pleasantness of a speech signal. The most common approach in speech enhancement is noise removal, where we, by estimation of noise characteristics, can cancel noise components and retain only the clean speech signal. The basic problem with this approach is that if we remove those parts of the signal that resemble noise, we are also bounded to remove those parts of the speech signal that resemble noise. In other words, speech enhancement procedures, often inadvertently, also corrupt the speech signal when attempting to remove noise [6]. Algorithms must therefore compromise between effectiveness of noise removal and level of distortion in the speech signal. Current speech processing algorithms can roughly be divided into three domains: spectral subtraction, sub-space analysis and filtering algorithms.

Spectral subtraction algorithms operate in the spectral domain by removing, from each spectral band, that amount of energy which corresponds to the noise contribution. While spectral subtraction is effective in estimating the spectral magnitude of the speech signal, the phase of the original signal is not retained, which produces a clearly audible distortion known as “ringing”. Sub-space analysis operates in the autocorrelation domain, where the speech and noise components can be assumed to be orthogonal, whereby their contributions can be readily separated. Unfortunately, finding the orthogonal components is computationally expensive. Moreover, the orthogonality assumption is difficult to motivate. Finally, filtering algorithms are time-domain methods that attempt to either remove the noise component (Wiener filtering) or estimate the noise and speech components by a filtering approach (Kalman filtering).

To fulfil the objective of objective of speech enhancement was initially done by using Kalman Filter, but the results did not meet the requirement. So, we segregated the entire signal into small samples called windows by adopting different windowing techniques like rectangular windowing and Hamming windowing. We iterated the process for few times by updating the autoregressive filter coefficients after every repetition. Even though the process takes long time for a tiny speech signal data, the output can be compared with input for its similarity.

The paper is structured as follows: Section 2 describes the related for speech enhancement using different adaptive filters, Section 3 describes the Adaptive filters, sub-section 3.1 is all about Kalman Filters, Section 4 describes the proposed method along with code implementation, Section 5 presents the simulated results and finally conclusions is depicted in Section 6.

## **II. RELATED WORK**

Over the past few decades, a number of theoretical and effective algorithms have been developed by researchers to overcome the background noise problem and enhance the speech signal. However, suppressing the noise from the noisy speech still persists as a challenging part of signal processing [..So, the enhancement of speech in noisy environment has evolved as one of the major area of interest in the field of speech enhancement, which has both theoretical interest and practical importance.

There are a lot of approaches reported in the literature related to speech enhancement. For the last ten years, adaptive filters have been effective and popular approaches for the speech enhancement. The advantage of

Adaptive filters are used to detect time-varying potentials and also track the dynamic variations of the signals. The adaptive filter has a number of advantages which includes lower processing delay and better tracking of the trajectory of non-stationary signals. These are essential characteristics in applications such as echo cancellation, adaptive delay estimation, noise estimation, and channel equalization in mobile telephony, where low delay and fast tracking of time-varying processes and time-varying environments are important objectives.

Recent research on adaptive filter has focused on a non-linear approach of the signals. This approach has been justified due to the variation of signal-to-noise ratio across the speech spectrum. In those Kalman filter has some advantages as compared with other methods. Unlike white Gaussian noise (WGN), which has a flat spectrum, the spectrum of real-world noise is not flat. Thus, the noise signal does not affect the speech signal uniformly over the whole spectrum. Some frequencies are affected more adversely than the others. In multi-talker babble, for instance, the low frequencies, where most of the speech energy resides, are affected more than the high frequencies. Hence it becomes imperative to estimate a suitable factor that will subtract just the necessary amount of the noise spectrum from each frequency bin (ideally), to prevent destructive subtraction of the speech while removing most of the residual noise. Another factor that leads to variation in SNR in different frequency bands of speech corrupted with noise is the fact that noise has non-uniform effect on different vowels and consonants.

### **III. ADAPTIVE FILTERS**

Adaptive filters are filters that have self-designing algorithms implemented in it. The filter learns the initial input statistics and continuously track these for time variations. Thus the estimation can be done for the deterministic signals. The noise that is uncorrelated with the deterministic signal can be thus eliminated [6, 14, 15]. Recursive algorithms are used to design these self-designing adaptive filters and find application there is lack of knowledge a priori. Figure.4 shows an Adaptive Filter Structure. Here the input signal obtained from sensors is  $d(n)$ . It contains the desired signal  $d'(n)$  as well as the undesired signal  $u(n)$ .  $u(n)$  can be considered as the noise signal distorting the measured signal. Now, if the noise signal  $u(n)$  is known a priori, desired signal  $d'(n)$  can be extracted by subtracting noise signal  $u(n)$  from input signal  $d(n)$  which is corrupted with noise. It is difficult to obtain the entire noise signal, thus an estimated noise signal  $y(n)$  is used. Filters are used to estimate the noise signal  $y(n)$ , There is a linear relation between the measurable noise source  $x(n)$  and the noise signal  $y(n)$ . The estimated signal  $y(n)$  is then used to obtain the difference signal  $e(n)$  using which the estimation of the desired signal  $d'(n)$  is done. The closeness of estimated noise signal  $y(n)$  with the real noise signal  $u(n)$  determines the value of the desired signal. Adaptive filtering implements noise cancellation theory. Adaptive filter design can be classified into two parts, digital filter and adaptive algorithm implementation. Function of adaptive algorithm is to generate proper filter coefficient. General digital filters use fixed filter coefficients, this is not the case with adaptive filter, Here a change in filter coefficients is considered on the basis of input and output signal characteristics as well as the environmental changes. The adaptive filter has a number of advantages which includes lower processing delay and better tracking of the trajectory of non-stationary signals. These are essential characteristics in applications such as echo cancellation, adaptive delay estimation, noise estimation, and channel equalization in mobile telephony, where low delay and fast tracking of time-varying processes and time-varying environments are important objectives. Fig 1 shows the adaptive filter structure.

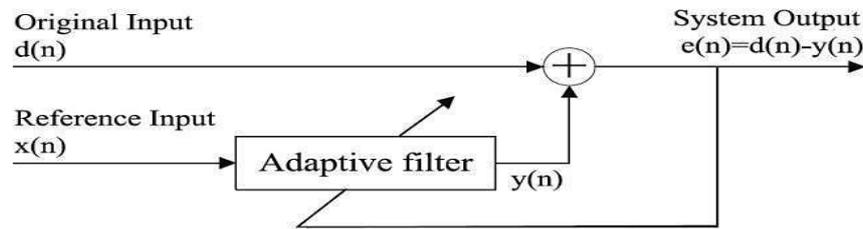


Fig. 1 Adaptive Filter Structure

### 3.1. Kalman Filter

Kalman filter is an optimal linear minimum mean-square-error state estimator for stochastic linear systems in a state form. Given with the model and, possibly noisy, measurements of inputs and outputs, it provides an optimal estimate of system states. If the noises involved are Gaussian, Kalman filter becomes an optimal mean-square-error estimator, i.e. not just among the linear estimators. Many formulations exist targeted for specific application. The main features of Kalman filtering are sequential operation model-based approach and possible non-stationary [6, 14].

The Kalman filter is a mathematical procedure which operates through a prediction and correction mechanism. Kalman filter combines all the available data measured, plus the knowledge of the system and the measurement devices, to produce an estimation of the desired variables in such a manner that the error is statistically minimized. The Kalman filter uses a system's dynamics model (i.e., physical laws of motion), known control inputs to that system, and measurements (such as from sensors) to form an estimate of the system's varying quantities (its state) that is better than the estimate obtained by using any one measurement alone. As such, it is a common sensor fusion algorithm. The use of Kalman Filter for speech enhancement in the form that is presented here was first introduced by Paliwal (1987). This method however is best suitable for reduction of white noise to comply with Kalman assumption. In deriving Kalman equations it normally assumed that the process noise (the additive noise that is observed in the observation vector) is uncorrelated and has a normal distribution [2, 13-15]. This assumption leads to whiteness character of this noise. There are, however, different methods developed to fit the Kalman approach to colored noises [10]. It is assumed that speech signal is stationary during each frame, that is, the AR model of speech remains the same across the segment. To fit the one-dimensional speech signal to the state space model of Kalman filter we introduce the state vector as:

$$x(n) = [x(n - p + 1)x(n - p + 2)x(n - p + 3) \dots x(n)]^T \quad (1)$$

Where  $x(k)$  is the speech signal at time  $n$ . Speech signal is contaminated by additive white noise  $v(n)$

Let  $s(n)$  and  $v(n)$  denote the clean speech and noise respectively. The observed noisy speech,  $x(n)$ , is given by-

$$x(n) = s(n) + v(n) \quad \text{Where, } n=1, 2, \dots \quad (2)$$

The clean speech signal and noise are modeled as AR pro-cesses



$$s(n) = \sum_{i=1}^p a_i s(n - i) + w(n) \quad (3)$$

$$v(n) = \sum_{i=1}^q b_i v(n - i) + u(n) \quad (4)$$

$$s(n) = F s(n - 1) + g w(n) \quad (5)$$

$$x(n) = C^T s(n) \quad (6)$$

Where,

$$\bar{s}(n) = \begin{bmatrix} s(n) \\ v(n) \end{bmatrix}, \quad \bar{w}(n) = \begin{bmatrix} w(n) \\ u(n) \end{bmatrix} \quad (7)$$

$$\bar{F} = \begin{bmatrix} F & 0 \\ 0 & F_v \end{bmatrix}, \bar{g} = \begin{bmatrix} g & 0 \\ 0 & g_v \end{bmatrix}, \bar{C} = \begin{bmatrix} C \\ C_v \end{bmatrix} \quad (8)$$

$$F = \begin{bmatrix} 0 & 1 & \dots & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \dots & 1 \\ a_p & a_{p-1} & \cdot & \dots & a_1 \end{bmatrix}, \quad F_v = \begin{bmatrix} 0 & 1 & \dots & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \dots & 1 \\ b_q & b_{q-1} & \cdot & \dots & b_1 \end{bmatrix} \quad (9)$$

$$s = [s(n - p + 1) \dots \dots \dots s(n - 1) s(n)]^T_{p \times 1} \quad (10)$$

$$v = [v(n - q + 1) \dots \dots \dots v(n - 1) v(n)]^T_{q \times 1} \quad (11)$$

$$g = C = [0 \dots \dots \dots 0 \ 1]^T_{p \times 1}, \quad g_v = C_v = [0 \dots \dots \dots 0 \ 1]^T_{q \times 1} \quad (12)$$

The estimate of the speech signal,  $\hat{s}(n)$ , can be obtained from the estimated state space of Kalman filtering by the below equation:

$$\hat{s}(n) = C_1^T \hat{s}(n|x_n), \quad C_1 = [C^T \ 0 \ \dots \ \dots \ 0]^T \quad (13)$$

The flow chart for Kalman filter algorithm is depicting in fig. 2.

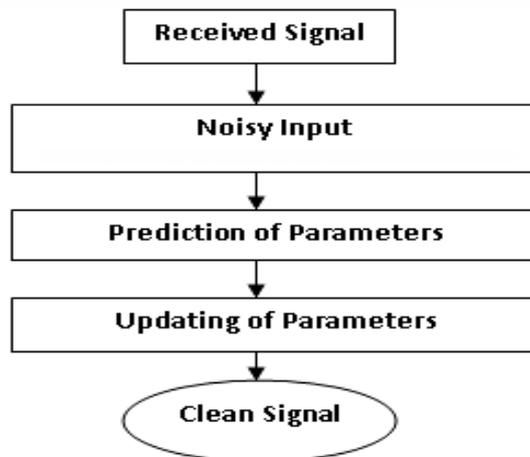


Fig.2 Flow chart of Kalman Filter

**IV. PROPOSED METHOD**

The main aim of the work is speech enhancement using Kalman filter. Initially, we have taken the audio input signal from NOZIEUS database which is implementing different noisy files and producing appropriate outputs respectively the signal that is used in this work is taken from the Noizous database [7]. The clean speech used in this work is a sentence pronounced by a male “Read verse out loud for pleasure.” (sp04.wav from NOZIEUS database). We have also taken a babble noise with SNR 10dB and calculated its LPC coefficients. Then we added babble noise with SNR of 10dB to the clean speech. This is used as the noisy speech which is given as the input to the Kalman as the data observed. As speech is not stationary for a long time we took small frames of speech by windowing. Here in this work, we observed the algorithm by taking different windowing techniques, Rectangular and Hamming. We took each frame length to be 240 samples. Now the segmented noisy speech is saved as a matrix where each row consists of the value of each window, where our each window is of 240 samples looping and taking one window at a time. We calculated the LPC coefficients of the original noisy speech signal and calculate the Kalman gain for each loop for updation of the next state. Looping is done as the past samples have an influence over the future samples. Finally after iterative process, the SNR of the output of the Kalman filter is calculated and compared with different techniques. Fig 3 shows the mechanism of Kalman filter in speech enhancement.

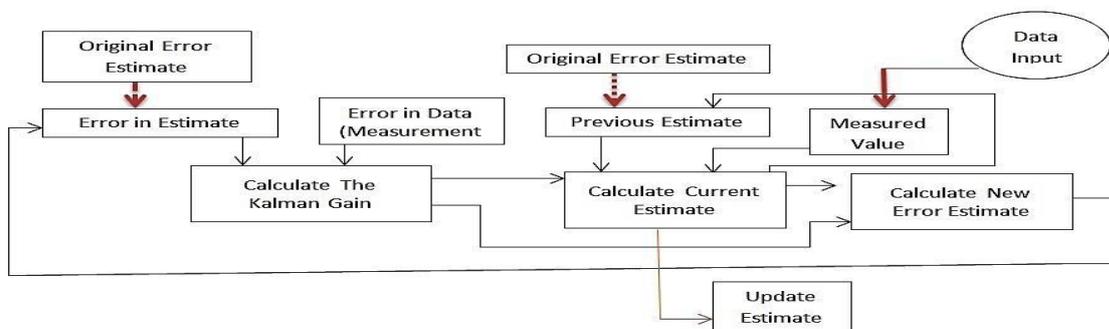


Fig. 3 Mechanism of Kalman filters in speech enhancement.

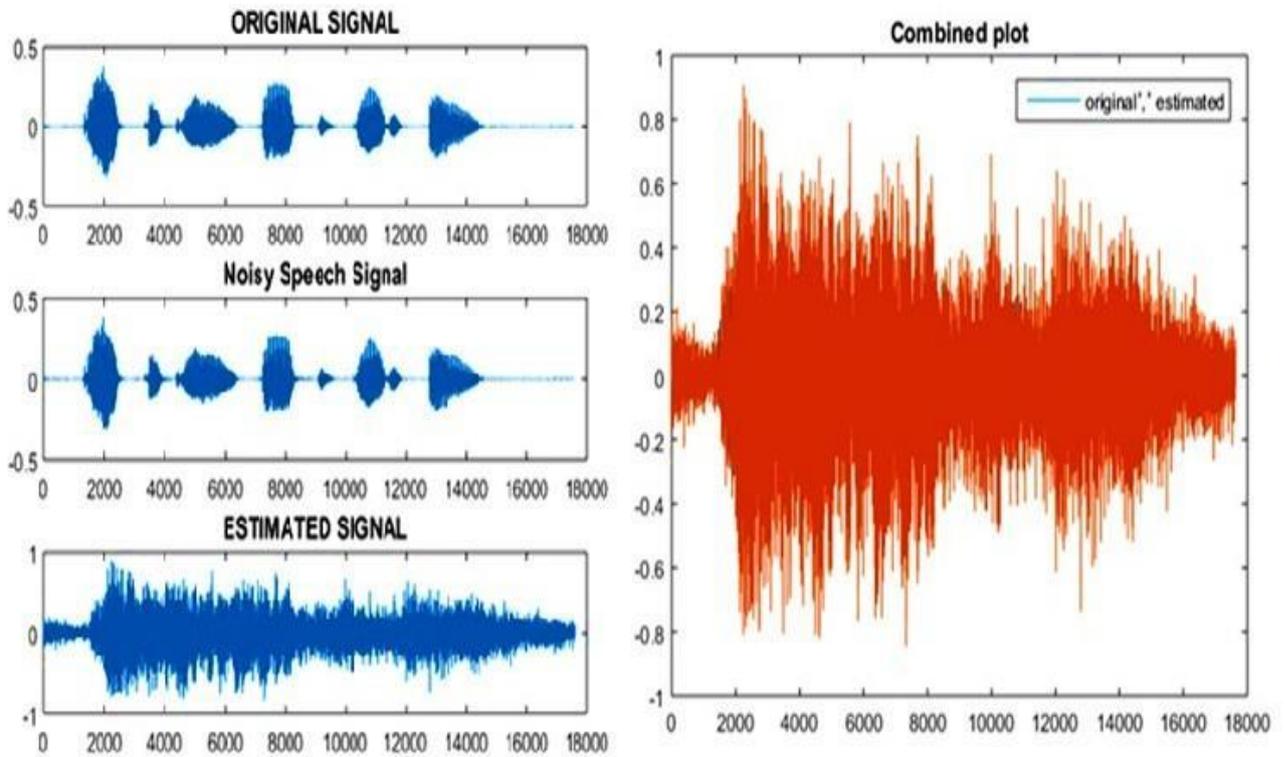
#### **4.1. Code Implementation:**

The code implementation follows the following steps:

1. First we have taken the audio input signal from NOZIEUS database which is implementing different noisy files and producing appropriate outputs respectively.
2. Then, we have given an input of speech containing noise.
3. An instruction to play the noisy speech with 0 SNR.
4. An instruction for noisy speech eradicating babble noise.
5. Then we can calculate the following data from the
  - Length of the input signal
  - Initialization of standard transition matrix
  - Transition matrix
  - Priori or posterior covariance matrix
  - Kalman gain
  - Kalman coefficient for  $yy$ .
  - Desired signal
  - Predicted state error
  - Estimated error sequence
  - Process noise covariance
  - Measurement noise covariance
  - Output of the signal

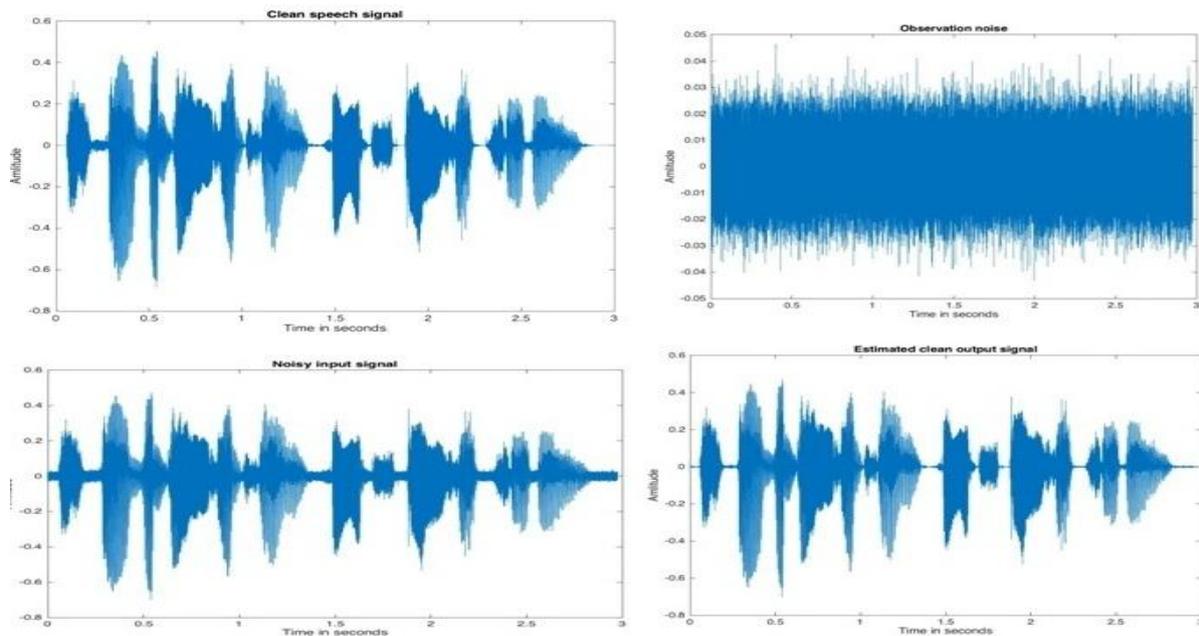
#### **V. SIMULATION RESULTS**

The noise removal from noisy speech signal was initially done by using Kalman Filter, but the results did not meet the requirement. So, we segregated the entire signal into small samples called windows by adopting different windowing techniques like rectangular windowing and Hamming windowing. We iterated the process for few times by updating the autoregressive filter coefficients after every repetition. Even though the process takes long time for a tiny speech signal data, the output can be compared with input for its similarity. Fig.4, 5 and 6 are depicting the simulation results under various situations like non-windowed processing, windowed processing (rectangular and hamming windowing).



**Fig. 4** Speech enhancement of entire signal by Kalman filter (non-windowing process)

In rectangular windowing process, we segregated the entire signal into small windows. After windowing, we estimated the output from noisy input signal (input + noise). This windowing process is better than Hamming process because of better intelligibility when compared to the hamming window signal output. This is evident from the plot as well.



**Fig. 5** Speech enhancement of sampled signal by Kalman filter (Rectangular windowing process)

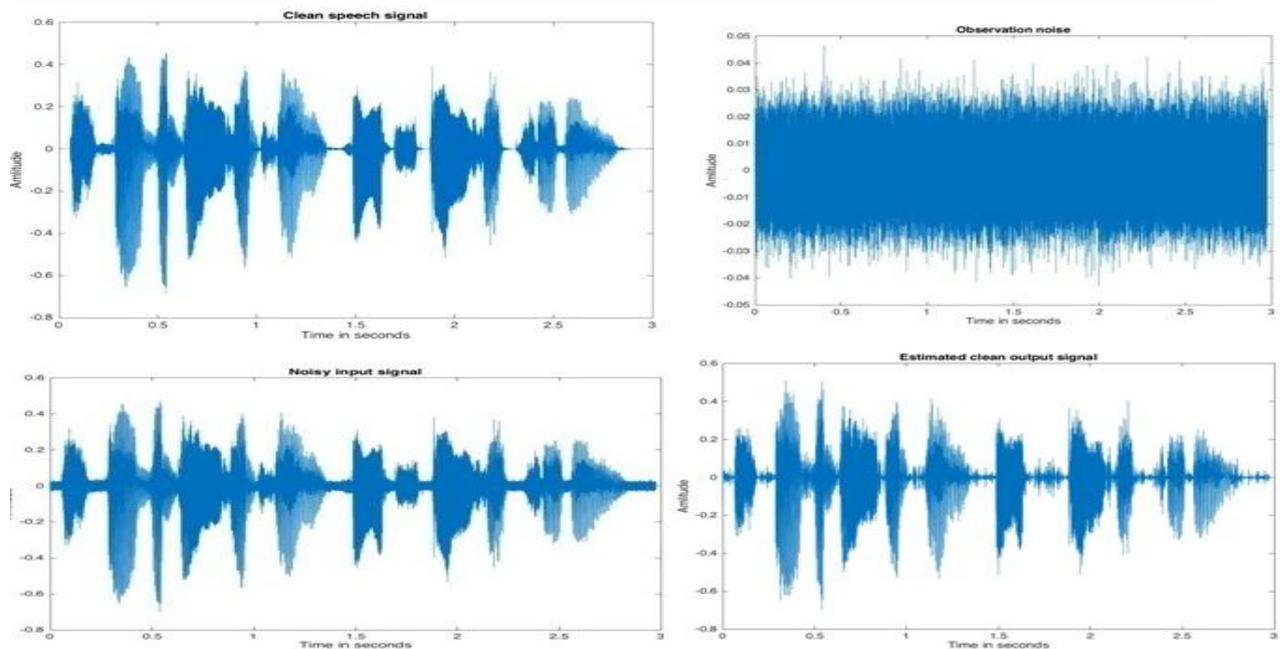


Fig. 6 Speech enhancement of sampled signal by Kalman filter (Hamming windowing process)

## VI. CONCLUSION

The enhancement of speech is important in various fields of communication and we have proposed kalman model which estimates the output, Kalman filter is implemented using NOZIEUS database. The windowing process model has some big matrices (P and A for instance), whose sizes are determined by choosing appropriate autoregressive filter order. The process is slow and it is not surprising given the number of matrix multiplications it has to do for every sample. The filter have some advantages compared to LMS, RLS and wiener, it works on real-time execution without storing observations or previous estimates, provides variance of the estimation error. The filter doesn't need any memory as it works in real time and is also good for stationary and non-stationary signals. Since the proposed algorithm has been enhanced the speech in efficient way. Hence, in future, it is expected to work better on music enhancement by using this algorithm. In case of low order of the AR (Autoregressive) model, the harmonic structure of music is often lost. Further advancement in this work is to test the algorithm with automatic order determination on music signals.

## REFERENCES

- [1] V. Abrol, P. Sharma and S. Buddhiraja, "Evaluating Performance of Compressed Sensing for Speech Signal" IEEE 3<sup>rd</sup> International Advance Computing Conference, 2013, pp.1159-1164.
- [2] U. Santosh Kumar and Dr. G. ManmadhaRao, "Speech Enhancement Using Combination of Digital Audio effects with Kalman Filter " in International conference on Signal Processing, Communication, Power and Embedded System (SCOPEs), 2016, pp- 1208-1211.
- [3] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," Proc. IEEE, vol. 67, pp. 1586-1604, Dec. 1979.

- [4] J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 197–210, Jun. 1978.
- [5] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 137–145, Apr. 1980.
- [6] J. Benesty, S. Makino, J. Chen: *Speech Enhancement*, Springer 2005
- [7] NOIZEUS: a noisy speech corpus for evaluation of speech enhancement algorithms
- [8] C. Plapous, C. Marro, P. Scalart, "Improved Signal-to-Noise Ratio Estimation for Speech Enhancement", *IEEE Trans. On Audio, Speech and Language Proc.*, vol. 14, no. 6, Nov., 2006
- [9] UtpalBhattacharjee&Pranab Das, "Performance Evaluation of Wiener filter and Kalman filter Combined with Spectral Subtraction in Speaker Verification System, *IJITEE*, ISSN: 2278-3075, vol. 2, Issue-2, January 2013.
- [10] V. Abrol, P. Sharma and S. Buddhiraja, "Evaluating Performance of Compressed Sensing for Speech Signal" *IEEE 3<sup>RD</sup> International Advance Computing Conference*, 2013, pp.1159-1164.
- [11] Emmanuel J. Candes and Michael B. Wakin, "An Introduction to Compressive Sampling", *IEEE Signal Processing Magazine*, March 2008
- [12] W.G. Yan, G.Y. Xiang and Z.X. Qun, "A signal Subspace Speech Enhancement method for Various Noises", *TELKOMNIKA*, vol.11, no. 2, pp. 726-735, Feb, 2013
- [13] Kalpana Naruka and Dr. O.P. Sahu "An Improved Speech Enhancement Approach based on Combination of Compressed Sensing and Kalman filter" in *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 2015
- [14] "Digital Audio Signal Processing," second edition, by Udo Zolzer
- [15] [http://dea.brunel.ac.uk/cmisp/Home\\_Esfandiar/KalmanTime.html](http://dea.brunel.ac.uk/cmisp/Home_Esfandiar/KalmanTime.html)