# A Fast Adaptive Kalman Filtering Algorithm for Speech Enhancement Under Stationary Noise Environment

**C.N. Prabhavathi and K.M. Ravikumar**

**Abstract** Kalman Filtering is one of the time domain speech-enhancement techniques. The conventional Kalman filtering technique involves more number of matrix operations. The complexity of matrix operations is reduced by fast adaptive Kalman Filtering technique. The proposed method of fast adaptive filtering technique is simple and gives best results for stationary noises. From the simulation results, it is seen that the proposed method of Kalman filtering is more effective in obtaining the clean speech signal. The performance of this filter is compared with the conventional method with respect to the signal to noise ratio and the execution time.

**Keywords** Speech enhancement · SNR · Execution time

## 1 Introduction

In voice communication, speech signals can be contaminated by environmental noise-like babble noise, mall noise, traffic noise, etc., and as a result the communication quality will be affected with less intelligibility. The quality of such speech signals can be improved using speech-enhancement techniques. One such technique is the Kalman filtering technique. There are many applications based on Kalman filtering algorithm as seen in [1–6]. These method models noisy speech signals in terms of state space equation and observation equation. The computational complexity is high as it needs the calculation of Linear prediction coding (LPC) coefficient and inverse matrix calculations [7].

C.N. Prabhavathi (✉)
Department of ECE, CERSSE, Hebbal Campus, Jain University, Bangalore
Karnataka, India
e-mail: prabhacngowda@gmail.com

K.M. Ravikumar
Department of ECE, SJCIT, VTU (Research Guide Jain University),
Chickballapur, Karnataka, India
e-mail: kmravikumar@rediffmail.com



S.S. Dash et al. (eds.), *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, Advances in Intelligent Systems and Computing 394,
DOI 10.1007/978-81-322-2656-7_2

The Kalman Filtering algorithm can be categorized as follows:

1. Conventional (Matrix) Kalman Filtering algorithm.
2. An improved filtering algorithm.
3. Modified fast adaptive Kalman filtering algorithm.

A brief of above algorithm is as follows:

Conventional Kalman filtering method is purely a matrix method, which involves matrix multiplications and matrix inversions. This method contains large number of redundant data and it is less adaptive. An improved filtering method reduces some matrix operations, and hence reduces the complexity of the conventional method. The proposed method which is the modified fast adaptive filtering is less complex and eliminates most of the matrix operations. This algorithm adapts for any type of environmental noise and works better with stationary type of background noises [8]. It has been observed that the intelligibility of speech signal is much better in the proposed method than with the conventional method.

This paper is organized as follows: Sect. 2 describes conventional filtering algorithm, improved filtering algorithm, and modified fast adaptive filtering algorithm. Section 3 provides the implementation steps with the results and conclusion is drawn in Sect. 4.

## 2 Fast Adaptive Kalman Filter Algorithm

### *2.1 Conventional Kalman Filtering Algorithm*

The conventional method involves matrix operations and matrices and is created for covariance, variance, Kalman coefficient, Kalman gain, noise coefficient, etc. [7].

The noisy speech signal $y(n)$ is given by

$$y(n) = s(n) + v(n) \tag{1}$$

where $s(n)$ is a clean speech signal and $v(n)$ is a noise signal.

The state equation and observation equation in matrix form is expressed as

[State equation]

$$X(n) = F(n) \times x(n-1) + Gw(n) \tag{2}$$

[Observation equation]

$$Y(n) = Hx(n) + v(n)$$

where $F(n)$ is a L × L transition matrix, $G$ is the input vector and $H$ is the observation vector.

The recursion equation of conventional Kalman filtering algorithm is as given below. The following set of equations provides the procedure for the implementation.

[Initialization]

$$\begin{aligned} X(0|0) &= 0, P(0|0) = I \\ R_v(n) &= \delta_v^2, G = [10\ldots0] \\ R_s(n)[I,j] &= \begin{cases} E(y(n) \times y(n) - \delta_v^2, (i,j) & 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{3}$$

[Iteration]

$$P\left(\frac{n}{n-1}\right) = F \times P\left(\frac{n-1}{n-1}\right)F^T + G \times R_s(n) \times G^T \tag{4}$$

$$K(n) = P\left(\frac{n}{n-1}\right)G^T/(G \times P\left(\frac{n}{n-1}\right)G^T + R_v(n) \tag{5}$$

$$X\left(\frac{n}{n-1}\right) = F \times X\left(\frac{n-1}{n-1}\right) \tag{6}$$

$$X\left(\frac{n}{n}\right) = X\left(\frac{n}{n-1}\right) + K \times X\left(y(n) - G \times X\left(\frac{n}{n-1}\right)\right) \tag{7}$$

$$P\left(\frac{n}{n}\right) = (I - K(n) \times G) \times P\left(\frac{n}{n-1}\right) \tag{8}$$

where $X$: State Vector, $P$: Parity matrix, $R_v$: Variance of noise (matrix), $R_s$: variance of noisy speech (matrix), $K$: Kalman Gain matrix, $G$: Input vector

Thus in the implementation of conventional method, each of the above parameters are realized using separate matrices which are then used to perform matrix operations including matrix inversions which are extremely time consuming.

## 2.2 Improved Filtering Algorithm

During the whole process of filtering, only the value of noisy speech $y(n)$ is useful. Hence in improved filtering algorithm very few matrix operations are done. This has much lesser execution time when compared to conventional method. Here even though the enhancement efficiency is lesser, the quality of required speech signal is not compromised. It is much more basic and simpler to understand than the conventional method. Here the matrix inversion is totally avoided [7].

Algorithm for the improved filtering method is given in the following steps:
[Initialization]

$$S(0) = 0,\ \ R_v = \delta_v^2,$$
$$R_s(n) = E(y(n) \times y(n)) - \delta_v^2.$$

[Iteration]

$$K(n) = \frac{R_s(n)}{(R_s(n) + R_v(n)} \tag{9}$$

$$S(n) = K(n) \times y(n) \tag{10}$$

## 2.3 Modified Fast Adaptive Kalman Filtering Algorithm

It is observed that in conventional filtering, any changes in the environmental noise due to the surrounding environment, it could be difficult to update the estimation of noise regularly. Hence it is required to estimate the noise regularly and obtain the filter algorithm which adapts to environmental noise change. This adaptation of filter is the fast adaptive Kalman filter which is useful in speech enhancement.

In order to update the estimation of noise constantly, a threshold level has to be set to determine whether current speech frame is noise or not. The environmental noise is updated with the following steps (a) update the variance of environmental noise. (b) Obtaining the SNR.

(a) Update the variance of environmental noise by [7],

$$R_v(n) = (1-d) \times R_v(n) + d \times R_u(n), \tag{11}$$

In "(11)," $d$ is the loss factor and it is given by

$$d = \frac{1-b}{(1-b^{t+1})} \tag{12}$$

b is a constant and it is assumed as 0.99.

Here before updating the variance of environmental noise, the variance of current speech frame $R_u(n)$ is compared with the threshold $U$ "(13)." If $R_u(n)$ is less than or equal to $U$, the current speech frame can be considered as noise and then the algorithm will reestimate the noise variance.

$$U = (1-d) \times U + d \times R_u(n) \tag{13}$$

(b) Obtaining the SNR

When the noise is very large "(13)," the updating threshold cannot be used directly and hence the SNRs are calculated by determining the variance of pure speech signal, variance of input degraded speech signal, and variance of background noise. The two SNRs are calculated, one for current speech frame (SNRI($n$)) and another for whole speech signal (SNRo($n$)) and are compared. According to [6]

$$SNR_I(n) = 10 \times \log_{10}(\frac{\delta_r^2(n) - \delta_v^2(n)}{\delta_v^2(n)}) \tag{14}$$

$$SNR_o(n) = 10 \times \log_{10}\left(\frac{\delta_x^2(n) - \delta_v^2(n)}{\delta_v^2(n)}\right) \tag{15}$$

In "(14)" and "(15)," $n$ is the number of speech frames. The speech frame is noise when $\mathrm{SNR_I}(n)$ is ≤SNRo($n$). If SNRI($n$) is >SNRo($n$), the noise estimation will be attenuated to avoid damaging of speech signals. According to [9], noise attenuation can be expressed as

$$R_v(n) = \frac{R_v(n)}{1.2} \tag{16}$$

Updating Amplitude Threshold

Along with the above two steps, an extra condition is set by determining an amplitude threshold for the noisy speech signal. This condition is sufficient to have the updating of background noise constantly. Here an amplitude threshold $Z$ of 0.008 is set for the noise samples taken for simulation. Here all the sample values less than 0.008 has been considered as only noise and been processed such that their enhanced amplitude is 0. The implementation steps are described with the following set of equations:

Let

$$Y(n) = S(n) + v(n)$$
$$R_s(n) = \delta_x^2(n)$$
$$R_v(n) = \delta_v^2(n)$$

[Initialization]

$$S(0) = 0, \;\; R_v(1) = \delta_v^2(1)$$

[Iteration]
If

$$SNR_I(n) \leq SNR_o(n) \,||\, SNR_o(n) < 0$$

then
If $R_s(n) < Z$
Then

$$\delta_r^2(n) = 0$$

Else

$$R_s(n) = \delta_x^2(n)$$

End

$$K(n) = \frac{R_s(n)}{R_s(n) + R_v(n)}$$
$$S(n) = K(n)Xy(n)$$

where $R_s(n) =$ Variance of noisy speech, $R_v(n) =$ Variance of noise, $\delta_r^2(n) =$ variance of speech frame, $K(n)$ = Kalman gain, $S(n)$ = Enhanced speech.

## 3 Implementation Steps and Results

The prerequisite for the implementation of the algorithm is the creation of the database for the noise and the speech. The noise database is created by considering various acoustic environmental noises. For the implementation purpose, the database includes stationary noises like white noise, traffic noise, canteen noise, shop mall noise, etc. All these noises are recorded using a microphone and an i-phone for duration of 60 s [8].

Similarly, a separate speech database is created by recording a sentence, which is taken from a TSP speech database. This database includes hundreds of sentences of which a sentence "A king ruled the state in early days" is recorded by various age groups. The age group includes the ages <12 years, between 12 and19, 20 and 29, 30 and 40, and above 40. All the age groups include a male and a female. The above sentence is recorded in a closed silence room.

The purpose of creating separate database is that, in this paper, the degraded speech is assumed to be additive, and hence the noise and speech are added separately to obtain a degraded or corrupted noisy speech signal [8]. The simulation is done for all the groups mentioned and the modified fast adaptive algorithm works better for the various age groups.

The performance results are categorized into three different sections: (i) In the form of waveform (ii) Compared with respect to SNR values and (iii) Compared with respect to execution time.

## 3.1 Performance Results in the Form of Waveform

Here the performance is compared in the form of time waveform for all the three methods for the noises like traffic noise and shop mall noise. The waveform is as shown in Figs. 1, 2, 3, 4, 5, 6, 7, and 8. Each figure has the waveform for clean speech signal, noisy speech signal, and enhanced speech signal.

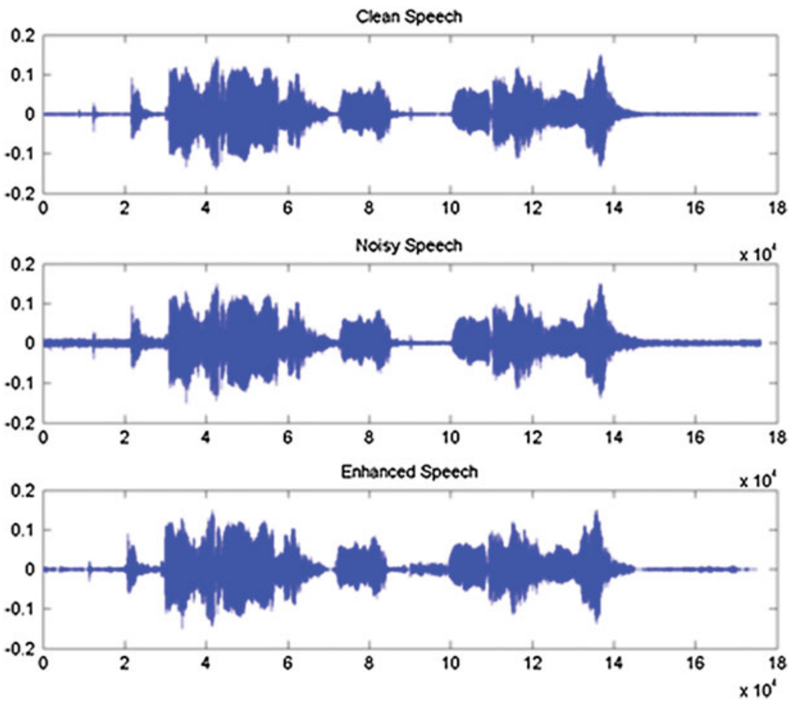


**Fig. 1** Filtering results for male voice under traffic noise

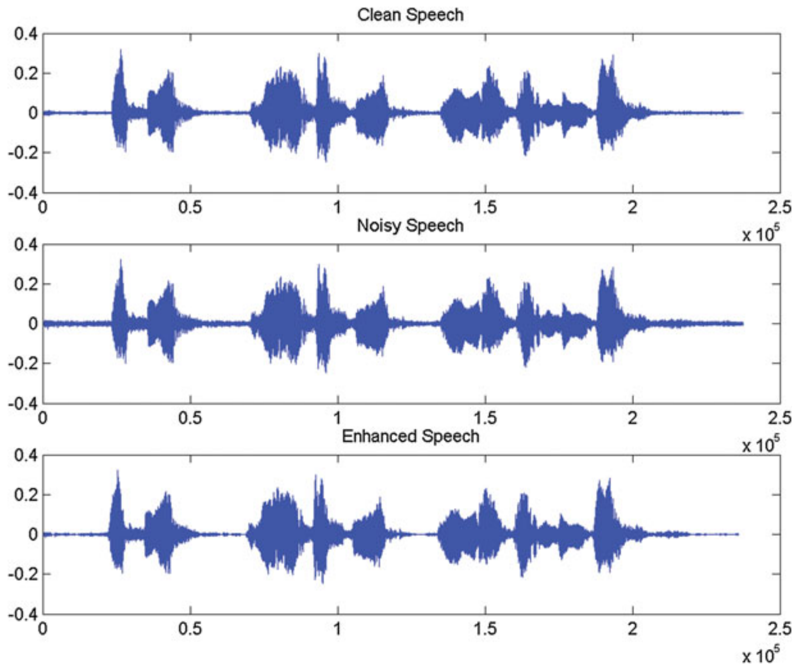


**Fig. 2** Filtering results for female voice under traffic noise

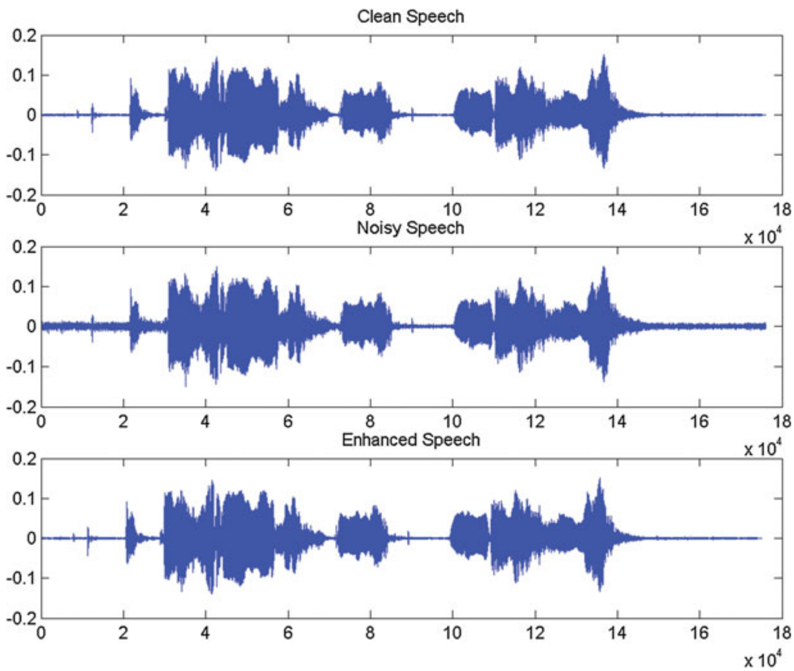


**Fig. 3** Filtering results for male voice under traffic noise

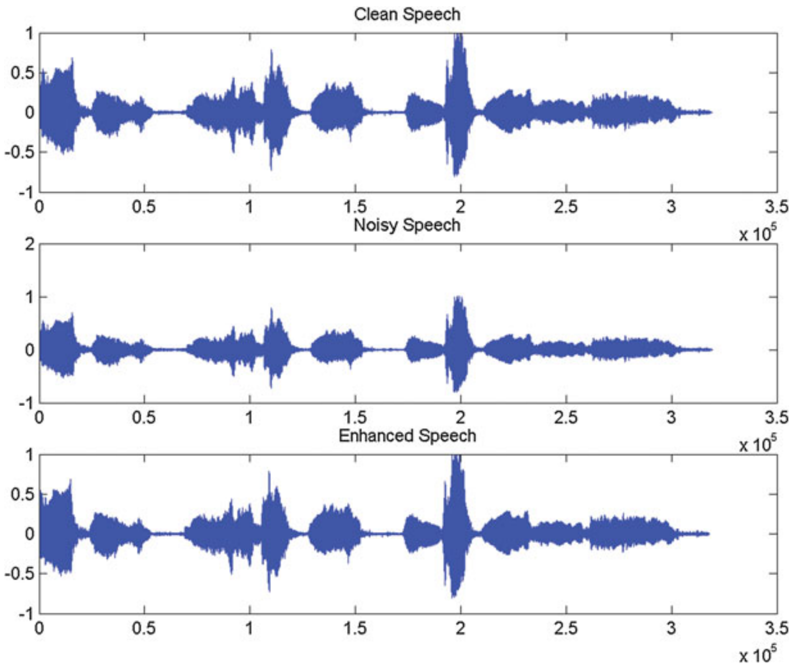


**Fig. 4** Filtering results for female voice under traffic noise

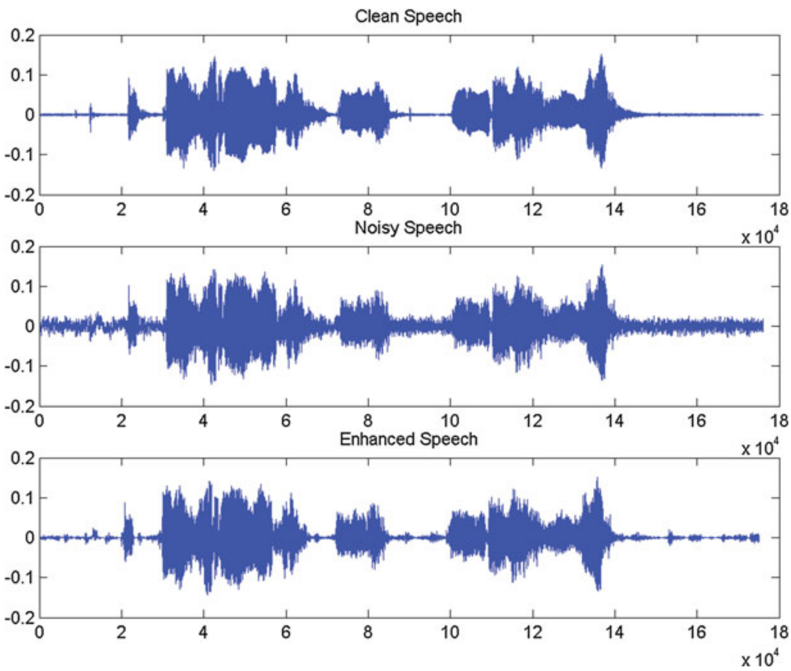


**Fig. 5** Filtering results for male voice under shop mall noise

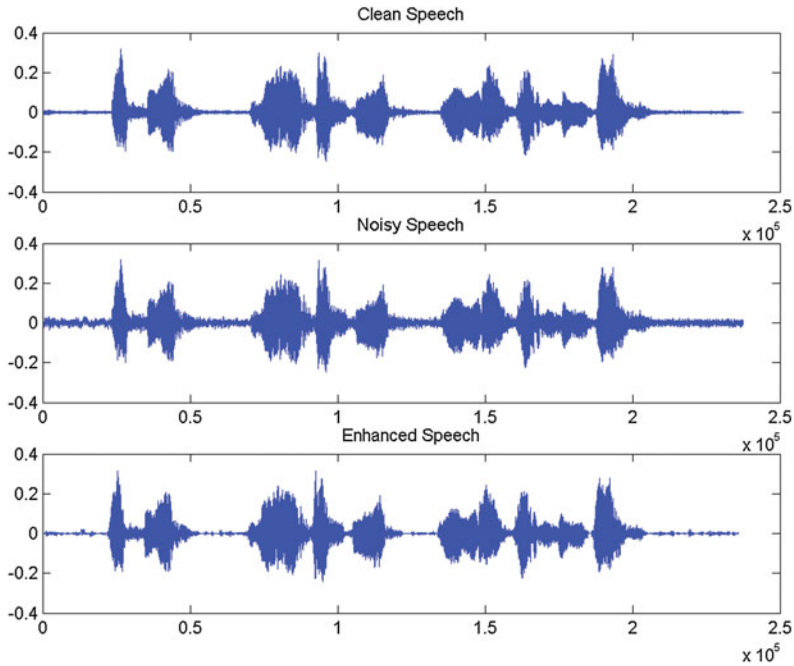


**Fig. 6** Filtering results for female voice under shop mall noise

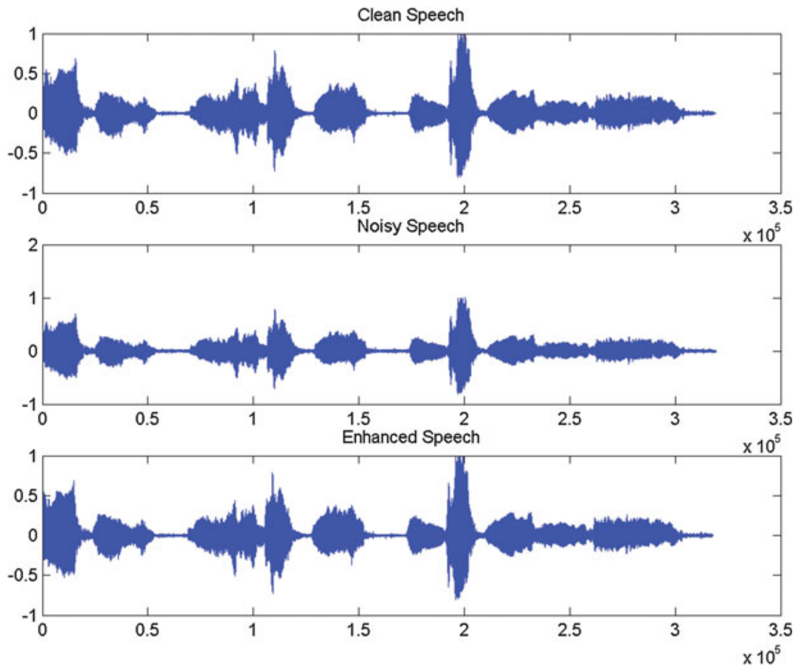


**Fig. 7** Filtering results for male voice under shop mall noise

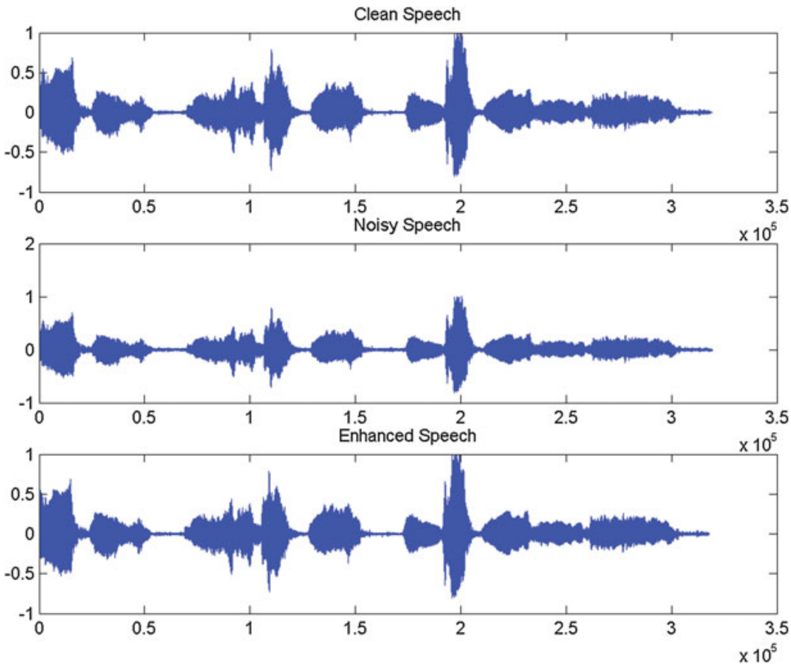


**Fig. 8** Filtering results for female voice under shop mall noise

### 3.1.1 Traffic Noise

**A. Conventional method**

From the observation of all the three signal waveforms, the intelligibility of the enhanced speech signal is same as clean speech signal and it is much better in fast adaptive algorithm than the conventional and adaptive filter algorithm (see Figs. 1 and 2).

**B. Modified Fast adaptive Kalman Filter**

See Figs. 3 and 4.

### 3.1.2 Shopmall Noise

**A. Conventional method**

See Figs. 5 and 6.

**B. Modified Fast adaptive Kalman Filter**

In all the above figures, the results of modified fast adaptive filtering method have better enhanced waveform which resembles the original clean signal. Here the simulation results are shown for male speaker and a female speaker with an average

**Table 1** Comparison table for SNR

| Sl. No | Methods | SNR(in)db | SNR(out)db |
|---|---|---|---|
| 1 | Conventional method | −10.2081 | −3.288 |
| 2 | Adaptive filtering method | −11.8536 | −3.2067 |
| 3 | Modified fast adaptive filtering method | −10.2171 | −3.2054 |

**Table 2** Comparison table for execution time

| Methods | Execution time (s) |
|---|---|
| Conventional method | 29.9 |
| Adaptive filtering method | 0.055 |
| Modified fast adaptive filtering method | 1.3068 |

age of 35 years. The simulation results for adaptive filtering resemble almost the conventional method, and hence the plots for the adaptive filtering are not shown.

The results are also compared with respect to the SNRs as described below (see Figs. 7 and 8).

### *3.2 Performance Results with Respect to SNR Values*

Here the input SNR of noisy speech and output SNR of enhanced speech is tabulated in Table 1 for all the three methods.

Here the SNR values are not normalized. The results show that there is an improvement in the SNR (out) of the modified fast adaptive filtering method when compared with the conventional method and adaptive filtering method.

### *3.3 Performance Results with Respect to Execution Time*

The execution time of all the three methods is tabulated in Table 2. The execution time for adaptive filter is less and it is more for conventional method. Hence the adaptive filtering algorithm is faster than the conventional method and fast adaptive filtering algorithm.

## 4 Conclusions

This paper has presented Kalman filtering with conventional method, adaptive filtering method and fast adaptive filtering method for stationary noise. The listening test and the simulation results show that fast adaptive Kalman filtering is more efficient in obtaining the clean speech signal. The SNR values obtained for fast adaptive filtering is better than the other two methods. But the execution time required for the fast adaptive filter is more than the adaptive filtering technique and lesser than the conventional method. Hence the proposed modified fast adaptive filtering is more efficient.

## References


1. Zhang X, Fu X, Wang X. Improved filter method for speech enhancement. Comput Appl. 2008; 28:363–5.
2. Tanabe N., Furukawa T., Matsue H, Tsujii S. Fast noise suppression algorithm with Kalman filter theory. In: Second International Symposium on universal communication, 2008, pp. 411–5.
3. Tanabe N, Furukawa T, Matsue H, Tsujii S. Kalman Filter for robust noise suppression in white and colored noise. In: IEEE International Symposium on Circuits and Systems, 2008.
4. Wu C, Han C. Square—root quadrature filter. Acta Electronica Sinica. 2009;37(5):987–92.
5. Su W, Huang C, Liu P, Ma M. Application of adaptive Kalman filter technique in initial alignment of inertial navigation system. J. Chin. Inertial Technol. 2010;18 1:44–47.
6. Gao Y, Zhang J. Kalman filter with wavelet based unknown measurement noise estimation and its application for information fusion. Acta Electronica Sinica 2007;35 1:108–111.
7. Mai Q, He D, Hou Y, Huang Z. A fast adaptive Kalman filtering algorithm for speech enhancement. In: IEEE International Conference on Automatic Science and Engineering Trieste, Italy, 2011.
8. Prabhavathi CN, Ravikumar KM. Acoustic noise classification and characterisation using statistical properties. IJETAE, 2014;4 6:638.
9. Hua X. Adaptive speech enhancement based on discrete cosine transform in high noise environment. Harbin: Harbin Engineering University; 2006.