ISSN: 2454-9940

# INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

ISSN 2454-9940
www.ijasem.org
INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
Vol 16, Issue 1, 2022

# Real-Time Speech Enhancement Through Fast Adaptive Kalman Filtering and Spectral Weighting

***Mr.Neralla Vivek***

*Assistant Professor, Department of CSE,*
*Malla Reddy College of Engineering for Women.,*
*Maisammaguda., Medchal., TS, India*

## Abstract:

***Voice augmentation is a method used to boost a speech signal that has been distorted by background noise. Many types of filters, such as the conventional, Fast Adaptive, weighted, etc., are used for this purpose. Traditional Kalman filters need non-adaptive processes like determining the parameters of an AR (auto-regressive) model and performing an inverse matrix operation. In the present study, we suggest the use of an adaptive Kalman filter. Includes a perceptual weighting filter, which gets rid of the need for matrix operations to simplify the computations. Depending on the human auditory system's masking features, adaptive filters use perceptual weighting to automatically adjust the noise estimate based on what they've seen. The simulation findings demonstrate an increase in performance above the status quo.***

## 1. Introduction

Rapid advances in multimedia communication and related applications have made speech enhancement a popular topic for academic study in recent years. Due to the presence of Noise in the background makes it difficult to understand what is being spoken. Suppressing this kind of background noise and enhancing the perceived quality and intelligibility of speech is the job of noise reduction or speech enhancement algorithms. The noise is unpredictable, and the complexity of human speech makes it tough to eliminate. There is often a compromise between how much noise is reduced and how much speech distortion is added when using various noise reduction methods. In the field of voice enhancement, a number of methods have been presented for this specific goal, including the spectral subtraction approach, the wiener filter, the Kalman filter, and the weighted filter. The success of these methods relies on the clarity and accuracy of the processed speech input. Most methods work at increasing the voice signal-to-noise ratio.

**Kalman Filter:**

The Kalman filter is a recursive prediction filter that is based \son the utilization of state space approaches and recursive algorithms. It makes forecasts about the condition of a living system. Such a dynamic some noise, often regarded as White noise, may disrupt a system. The Kalman filter incorporates perturbed measurements that are relevant to the state and may help enhance the estimated state. Thus the Kalman filter consists of two steps:

(1) The forecast

Secondly, the amendment in the first stage, the dynamic model is used to make a prediction about the state. The second stage involves adjusting the model with the observations in order to reduce the estimator's error covariance. It's the best possible estimate in this regard. This process is carried out

For each time step, with the previous time steps state serving as the beginning value. It is because of its recursive nature that the Kalman filter is referred to as a recursive filter. Calculating LPC in an AR (auto-regressive) model is a prerequisite for noise reduction in many alternative approaches as well as standard ones. Although it is demonstrated how to implement a simple Kalman filtering method without generating LPC coefficients in (3) and (4), this approach still includes a matrix inversion operation and a considerable amount of redundant input and is hence non-adaptive.

Better voice enhancement than the standard technique was developed by combining the adaptive Kalman filtering algorithm with a perceptual weighting filter. Initial state vector Z value (n) Perceptual weighting filter was regularly updated in the adaptive algorithm, and human auditory features are the source of this information. Forgetfulness factor in (4) and (5) automatically adjusts the estimation of environmental noise based on observation data, making it possible for the algorithm to capture the true noise in real time without human intervention.

### 2. Kalman Filtering Algorithm

**Conventional Kalman Filtering Method:**

White noise, colour noise, and other sounds were only some of the options. Specifically, we assume that the speech signal was fed by the all-linear, pole-zero output of white noise. It's a looping procedure. In a q-step AR model (one that anticipates a system's output based on the outcome of the previous iteration), a purely speech signal may be created by

ISSN 2454-9940
www.ijasem.org
INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
Vol 16, Issue 1, 2022

$$S(n) = \sum_{i=1}^{q} a_i(n) \times s(n-i) + w(n) \quad \ldots\ldots\ldots(1)$$

To clarify (1), we have the LPC coefficient of the AR model, air (n), and white Gaussian noise, w (n), with zero mean and variance u2. Speech signal, denoted as S (n). The zero-mean, variance-squared pure speech signal S (n) is degraded by the additive observation noise v (n) (i.e. environmental noise). The resulting garbled audio signal, y (n), may be written as

$$y(n)=S(n)+v(n) \ldots\ldots\ldots\ldots(2)$$

Assuming that v 2 is a constant, we used a method of quiet segmentation to reduce complexity. Both (1) and (2) have the potential to be written out as a set of equations, state and observation, provided by

**[State equation]**

$$m(n)=F(n)m(n-1)+G\,w(n) \ldots\ldots(3)$$

**[Observation equation]**

$$y(n)=H\,m(n)+v(n) \ldots\ldots(4)$$

H: Observation vector.

F (n) is the q q transition matrix expressed as

$$F(n)=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \ldots\ldots(5)$$

F (n) =bunch of speech signals.

The traditional Kalman filter made it simple to make estimates and keep an eye on the speech signal by using LPC coefficients. The above method only required 50% as much computing time. Changes are made to both the transition matrix F and the observation matrix H. The term "defined as" has been defined by them.

$$F=H=\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 1 & \cdots & 0 \end{bmatrix} \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

A single candidate speech signal for amplification out of a set of candidate signals is sometimes referred to as a "q 1 state vector."

**[State equation]**

$$Z(n)=F \times M(n-1)+Q(n) \ldots\ldots\ldots\ldots\ldots (7)$$

**[Observation equation]**

$$Y(n)=H \times Z(n)+R(n) \ldots\ldots\ldots\ldots (8)$$

The speech signal constituted the state equation, while the speech signal and additive noise constituted the observation equation (3). The Kalman filter iteratively updates the calculation of the system's state vector using the latest observation data as input. The calculation of the LPC coefficient may be avoided using the recursive estimate of the Kalman filtering technique shown below, provided that the noise variance v 2 is known.

**The Conventional Method Procedure**

[Initialization]

$Z(0/0)=0,\ P(0/0)=I(\text{identity matrix})$

$B_v(n)=\delta^2_v \qquad G=[1\ 0\ldots\ldots\ldots0],$

$$B_s(n)[i,j]=\begin{cases} E(Y(n) \times Y(n)) - \delta v\hat{}2(i,j=1) \\ 0 \qquad \text{otherwise} \end{cases}$$

[iteration]

$$P(n/n-1)=F \times P(n-1/n-1) \times F^T+G \times B_s(n) \times G^T \ldots(9)$$

$$K(n)=P(n/n-1) \times G^T / G \times P(n/n-1) \times G^T+B_v(n).(10)$$

$$Z(n/n-1)=F \times Z(n-1/n-1) \ldots\ldots\ldots\ldots\ldots(11)$$

$$Z(n/n)=Z(n/n-1)+K(y(n)-G \times Z(n/n-1)) \ldots\ldots(12)$$

$$P(n/n)=(I-K(n) \times G) \times P(n/n-1) \ldots\ldots\ldots\ldots\ldots\ldots(13)$$

$$S(n)=K(n) \times y(n) \ldots\ldots\ldots(14)$$

**Adaptive Kalman Filtering Algorithm**

**Classical Adaptive Filters:**

In adaptive filters, if the error is not zero, an update algorithm will readjust the filter parameters such that the actual output matches the intended output. One kind of adaptive filter Figure 1 depicts a typical setup.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 16, Issue 1, 2022

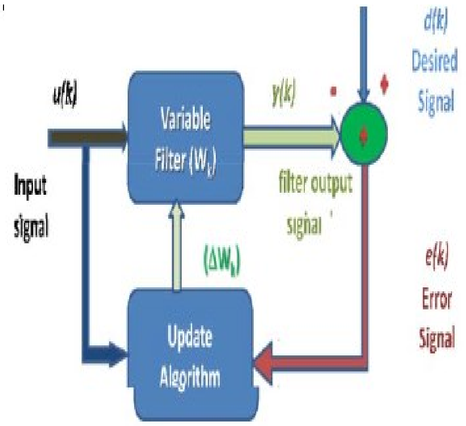



**Figure 1:** General adaptive Filter Configuration Algorithm

Because of the ever-evolving nature of the noise's source, it is essential that noise models be regularly updated to reflect the situation. As a result of shifts in the adaptive Kalman filtering technique may modify ambient noise by regularly revising the noise floor. The quick Adaptive Kalman filtering method estimates process and measurement noise online using the measured value and the filtered value, monitoring noise changes in real time to adjust filter parameters and enhance filtering performance. In the adaptive technique, we may adjust the threshold by which we judge whether or not the current speech frame contains noise. There are essentially two phases to it: a) Increasing the threshold U and b) increasing the ambient noise variance by (n).

**Updating the variance of the environmental noise by:**

$$B_v(n)=(1\text{-}d)\times B_v(n)+d\times B_u(n)\text{.............}(15)$$

Under the present estimate, d is the loss factor that reduces the filtering memory's duration and boosts the importance of fresh observation. Increasing the importance of fresh information in the estimate, while progressively forgetting the previous information.

The formula for (4), where it appears, is

$$d=1\text{-}b/1\text{-}b^{t+1}\text{.................}\ (16)$$

Where b is a forgetting factor (0b1), where b typically is in the range 0.95–0.99 for the purposes of this article, b is assumed to be 0.99. The threshold U is compared to the current speech frame Bu (n) before eq (15) is applied. The algorithm will recalculate the noise variance if Bu (n) is less than U, indicating that the current speech frame may be treated as noise. Since we do not have information on the variance of the background noise, Bu (n) cannot be used in lieu of by (n). Our mistake is minimized by the constant d.

**Updating the threshold by:**

$$U=(1\text{-}d)\times U+d\times B_u(n)\text{............}(17)$$

Since the updating threshold U is not constrained by the constraint Bu (n) U and is just impacted by Bu, noise will be considerable when mistakes are big (n) Foregoing Equation Implementation to boost the signal-to-noise ratio (SNR) of the speech frames, as specified by (17); we impose the following additional restriction:

$\delta^2_r$: variance of the pure speech signals,
$\delta^2_x$: variance of the input noise speech signals,
$\delta^2_v$: variance of background noise.

We mainly calculate two SNRs and compare between them. According to (6),

1. $SNR_1(n)$: SNR for the current speech frame
   $SNR_1(n)=10\times\log_{10}((\delta^2_r(n)\text{-}\delta^2_v(n))/\delta^2_v(n))\ldots(18)$
2. $SNR_0(n)$: SNR for the whole speech frame
   $SNR_0(n)=10\times\log_{10}((\delta^2_r\text{-}\ \delta^2_v(n))/\ \delta^2_v(n))\ldots.(19)$

In [18] and [19], n is the number speech frames, higher accuracy may be reached by updating δ2v. If SNR1 (n) ≤ SNR0 (n), or SNR0 (n) <0, then speech frame is noise and therefore these frames will follow the second constraint (Bu (n) ≤ U). However, if SNR1 (n) is bigger than SNR0 (n), the noise estimation will be reduced to prevent harming the speech signals. This dampening may be calculated using Eq. (7).

$$B_v(n)=B_v(n)/1.2\text{...................}(20)$$

**Perceptual Weighting Filter Algorithm:**

Perceptual Weighting filter process generally leads in improvement in the speech performance and it is based on linear prediction (LP) co-efficient that describes the shorter voice signal correlation. There are several applications for weighting filters, including the evaluation of electrical noise on telephone lines and the assessment of noise as experienced via the acoustic response of various instruments.

$$W(Z)=\frac{A(Z)}{A(\frac{Z}{\gamma})}=(1\text{-}\textstyle\sum_{i=1}^{p}a_iZ^{i})/(1\text{-}\sum_{i=1}^{p}a_i\gamma^{i}Z^{-i})\text{......}(21)$$

If air is an LP coefficient, then A (Z) is the p-the order LP analysis filter. The filter coefficient for this filter is computed using linear predictive

ISSN 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

www.ijasem.org

Vol 16, Issue 1, 2022

analysis in (8). Furthermore, is a sensory weighting that doesn't shift the fundamental formant frequency but does enlarge the formant spectrum. In particular, the widening of frequency f given by

$$\delta_f = (f_s/\pi)\ln\gamma \ldots\ldots\ldots\ldots\ldots\ldots\ldots(22)$$

$f_s$ : sampling frequency in hertz.

To compensate, the weighting filter minimizes the significance of the format structure and maximizes the significance of the format troughs in the voice signal. As a consequence, the formant area, where spectral masking causes the auditory system to make a bigger matching mistake those systems that are less likely to suffer from quantization errors Error. Through a series of subjective hearing experiments, we've determined that 0.9 is the best value for 8KHZ sampling.

## 3. Simulation Results

Using MATLAB, we examine the differences between the standard, the quick filtering method, and the perceptual weighting filter. Analysis of Adaptive Method and Perceptual Weighting Filter's Efficiency:
Two different variations of the noisy speech were used as signal samples for the simulations. Both the male and female voice signals have been tainted by ambient noise. Under the conditions of SNRin=6.50[dB] for the female signal and SNRin=3.71[dB] for the male signal, the filtering efficiency SNRout is shown in Table 1. When speech signals are damaged by white noise, the adaptive method's SNRout is greater than that of the no adaptive approach, as seen in the table below. When the speech signal is damaged by white noise, it is obvious that the adaptive technique can produce greater performance noise suppression capabilities than the no adaptive method

**Table 1:** Snout Result for the Noisy Speech Signal with
White Noise

| $SNR_{in}[dB]$ | | $SNR_{out}[dB]$ | | |
|---|---|---|---|---|
| | | Non-adaptive | Adaptive | Perceptual weighting |
| Female | 6.50 | 8.89 | 13.29 | 19.40 |
| Male | 3.71 | 4.49 | 5.39 | 9.04 |

**Table 2:** Different filtering methods comparisons of MSE for male and female speech signal

| Speech Signal | Kalman Filter | Adaptive Kalman Filter | Perceptual Weighting Filter |
|---|---|---|---|
| Male | 0.431 | 0.045 | 0.002 |
| Female | 0.324 | 0.032 | 0.001 |

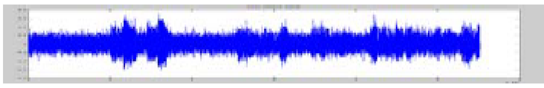
**Table 3:** Different filtering methods comparisons of CPU time for male and female speech signal

| Speech Signal | Kalman Filter | Adaptive Kalman filter | Perceptual Weighting Filter |
|---|---|---|---|
| Male | 9.602 sec | 5.701 sec | 3.562 sec |
| Female | 8.490 sec | 3.324 sec | 2.826 sec |

The suggested technique is shown to be simpler and to provide superior filter efficiency while considerably reduced running time in Tables 1, Table 2, and Table 3, respectively. The indication of speaking

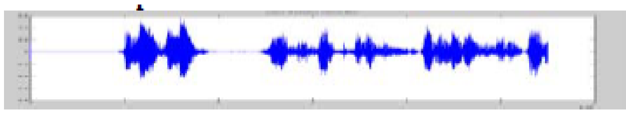
**The Filtering Results for the Male Speech with Noise:**

**Clean Speech Signal**



**Noisy Speech Signal**

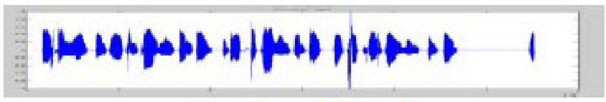
**Output of Conventional Kalman Filter**



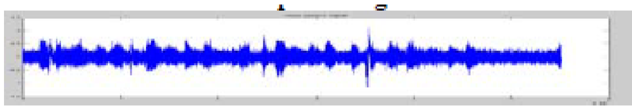
**Output of Adaptive Kalman Filter**

**Output of Weighted Filter**

**The Filtering Results for the Female Speech with Noise:**

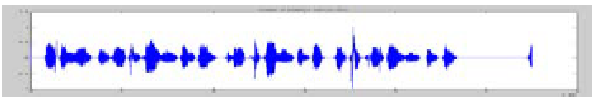


**Clean Speech Signal**

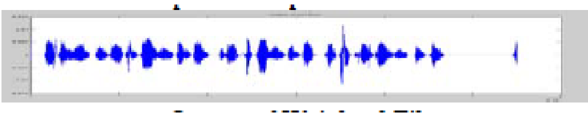


**Noisy Speech Signal**

**Output of Conventional Kalman Filter**



**Output of Adaptive Kalman Filter**



**Output of Weighted Filter**

## 4. Conclusion

Using a rapid adaptive Kalman filtering algorithm and a perceptual weighting filter, the authors of "A novel approach for voice enhancement" are able to reduce the amount of noise in recorded speech. Matrix operations with the aid of the coefficient factor, and the features of the human ear. The suggested approach has been proved to have reasonable efficacy by numerical findings and subjective assessment results.

Compared to other methods, the suggested method's Snout is greater when the speech signals are damaged by white noise and it takes less time to execute thanks to the use of two-state multiplications in each step. After comparing the suggested technique to the state-of-the-art methods, we found that the proposed method was less complicated and achieved better noise suppression while maintaining or improving the voice signal's quality.

Using the results of this publication as a starting point, we will refine the adaptive algorithm to provide a more precise evaluation of background noise in future research.

## References

*[1] Quanshen Mai, Dongzhi He, Yibin Hou, Zhangqin Huang, A Fast Adaptive Kalman Filtering Algorithm For Speech Enhancement, Aug.2011.*

*[2] ZHANG Xiu-zhen, FU Xu-hui, WANG Xia, improved Kalman filter method for speech enhancement. Computer Application, Vol.28, pp.363-365, Dec.2008*

*[3] Nari Tanabe, Toshiniro Furukawa, Hideaki Matsue and Shigeo Tsujii. Kalman Filter for Robust Noise suppression in White and colored Noise. IEEE International symposium on circuits and Systems, 2008.1172-1175.*

*[4] Nari Tanabe,Toshiniro Furukawa, Shigeo Tsujii. Fast noise Suppression Algorithm with Kalman Filter Theory. Second International Symposium on Universal Communication, 2008, 411-415.*

*[5] WU Chung-ling, HAN Chong-Zhao.Square-Root Quadrature Filter. Acta Electronica silica, Vol.37, No.5, pp.987-992, May. 2009.*

*[6] SU Wan-xin,HUANG Chun-mei,LIU Pei-wei,MA Ming-long Application of adaptive Kalman filter technique in initial alignment of inertial navigation system. Journal of Chinese Inertial Technology, Vol.18, No.1, pp.44-47, Feb.2010.*

*[7] GAO Yu,ZHANG Jian-qiu. Kalman Filter with Wavelet-Based unknown Measurement Noise Estimation and Its Application for Information Fusion. Acta Electronica Sinica, Vol.35, No.1,pp .108- 111,Jan.2007.*

*[8] XIE-Hua. Adaptive Speech Enhancement Base on Discrete cosine Transform in High Noise Environmental. Harbin Engineering University, 2006.*