DESIGN AND ANALYSIS OF SPEECH PROCESSING USING KALMAN FILTERING

[1]VINEELA MURIKIPUDI, [1]K.PHANI SRINIVAS [1]D.S.RAMKIRAN, [2]PROF.HABIBULLA KHAN, [1]G.MRUDULA, [1]K.SUDHAKAR BABU, [1]T.RAGHAVENDRA VISHNU

[1]Department of ECE, K L University, Guntur DT, AP, India
[2]Head of the department, Dept.of ECE, K L University, Guntur DT, AP, India
Email: vineela.murikipudi@gmail.com

ABSTRACT

Speech processing is used widely in every day's applications that most people take for granted, such as network wire lines, cellular telephony, telephony system and telephone answering machines. Due to its popularity and increasing of demand, engineers are trying various approaches of improving the process. One of the methods for improving the process is Kalman filtering. Kalman filtering now become a popular filtering technique for estimating and resolving redundant errors containing in speech. The objective of this paper is to generate a reconstructed output speech signal from the input signal involving the application of a Kalman filter estimation technique. In this paper, Kalman filter is used to estimate the parameters of the autoregressive (AR) process and represented in the state-space domain.

Keywords: *Kalman Filter, Noise Reduction, Speech Processing.*

1. INTRODUCTION

Speech is a form of communication in every day life. It is essential to know how we produce and perceive it and how speech technology may assist us in communication. Therefore in this project, we will be looking more into speech processing with the aid of an interesting technology known as the Kalman Filter [1-2]. One of the common adaptive filtering techniques that are applied to speech is the Wiener filter. This filter is capable of estimating errors however at only very slow computations. On the other hand, the Kalman filter suppresses this disadvantage.

As widely known to the world, Kalman filtering techniques are used on GPS (Global Positioning System) and INS (Inertial Navigation System). Nonetheless, they are not widely used for speech signal coding applications. The reason why Kalman filter is so popular in the field of radar tracking and navigating system is that it is an optimal estimator, which provides very accurate estimation of the position of either airborne objects or shipping vessels. Due to its accurate estimation characteristic, engineers are picturing the Kalman filter as a design tool for speech, whereby it can estimate and resolve errors that are contained in speech after passing through a distorted channel. Due to this motivating fact, there are many ways a Kalman filter can be tuned to suit engineering applications such as network telephony and even satellite phone conferencing. Knowing the fact that preserving information, which is contained in speech, is of extreme importance, the availability of signal filters such as the Kalman filter is of great importance [3-5].

2. KALMAN FILTER

The Kalman Filter is an estimator for what is called the "linear- quadratic problem", which focuses on estimating the instantaneous "state" of a linear dynamic system perturbed by white noise. Statistically, this estimator is optimal with respect to any quadratic function of estimation errors. It is a Recursive Data Processing Algorithm. The block diagram of typical Kalman filter application is shown in Figure 1 [6-7].

In practice, this Kalman Filter is one of the greater discoveries in the history of statistical estimation theory and possibly the greatest discovery in the twentieth century. It has enabled mankind to do many things that could not have been done without it, and it has become as indispensable as silicon in the makeup of many electronic systems.

In a more dynamic approach, controlling of complex dynamic systems such as continuous manufacturing processes, aircraft, ships are the most immediate applications of Kalman filter. In order to control a dynamic system, one needs to know what it is doing first. For these applications, it is not always possible or desirable to measure every variable that you want to control, and the Kalman filter provides a means for inferring the missing information from indirect (and noisy) measurements. Some amazing things that the Kalman filter can do is predicting the likely future courses of dynamic systems that people are not likely to control, such as the flow of rivers during flood, the trajectories of celestial bodies or the prices of traded commodities [8-10].

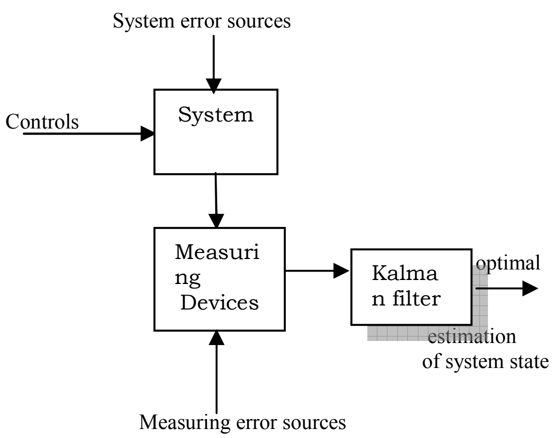

Figure 1. Block Diagram of typical Kalman Filter application

2.1. Process of estimation

The process commences with the addresses of a general problem of trying to estimate the state o f a discrete-time controlled process that is governed by a linear stochastic difference equation:

$$X_k = Ax_{k-1}+Bu_k+W_{k-1} \quad (1)$$

With a measurement $z \in R^m$ that is

$$Z_k = Hx_k+V_k \quad (2)$$

The random variables W_k and V_k represent the process and measurement noise (respectively).We assume that they are independent of each other, white, and with normal probability distributions

$$P(W) \sim N(O,Q) \quad (3)$$
$$P(V) \sim N(O,R) \quad (4)$$

Ideally, the process noise covariance Q and measurement noise covariance R matrices are assumed to be constant, however in practice, they might change with each time step or measurement. In the absence of either a driving function or process noise, the $n \times n$ Matrix A in the $n \times 1$ difference equation, relates the state at the previous time step K-1 to the state at the current step k In practice, change with each time step, however here it is assumed constant. The matrix B relates the optional control input $u \in R^I$ to the state X^H which relates the state to the measurement Z^k. In practice H might change with each time step or measurement; however we assume it is constant.

2.2. Computational origins of the filter

Let say we define $\hat{x}^- \in R^n$ (note:"super minus")to be our priori state estimate at step k,given knowledge of the process prior to step k and $\hat{x}_k \in R^n$ (note: with out the "super minus") to be our posteriori state estimate at step k,given measurement z_k.The priori and posteriori estimate errors can be defined as: [11-13]

$$e_k^- = x_k - \hat{x}^-_k \quad (5)$$

$$e_k = x_k - \hat{x}_k \quad (6)$$

The priori estimate error covariance is then

$$P_k^- = E[e_k^- e_k^{-T}] \quad (7)$$

And the posterior estimate error covariance is

$$P_k = E[e_k e_k^T] \quad (8)$$

After deriving the equations for the Kalman filter, the goal is to find an equation that computes a posteriori state estimate $\hat{x}_k$, as a linear combination of a priori estimate $\hat{x}_k^-$ and a weighted difference between an actual measurement Z_k and a measurement prediction $H\hat{x}_k^-$ as shown below in e q (9).Some justification for eq (9) is given in the section "The probabilistic Origins of the Filter"found below.

$$\hat{x}_k = \hat{x}_k^- + K(Z_k - H\hat{x}_k^-) \qquad (9)$$

The gain, K or otherwise known as the blending factor, minimizes the posteriori error covariance in equation 8 and is a $n \times m$ matrix in eq 9. This minimization can be accomplished by first substituting equation 9 into the above definition for e_k, after which substituting e_k into eq 10, performing the indicated expectations, taking the derivative of the trace of the result with respect to K, setting that result equal to zero, and then solving for K_k. One form of the resulting K that minimizes eq (8) is given by eq (10) as follow:

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

$$K_k = \frac{P_k^- H^T}{HP_k^- H^T + R} \qquad (10)$$

By looking at eq (10), it can be seen that as the measurement error covariance R approaches zero, the gain K weights the residual more heavily. Specifically,

$$\mathrm{Lim}_{R \to 0} K_k = H^{-1} \qquad (11)$$

On the other hand, as the priori estimate error covariance P_k^- approaches zero, the gain K will weight the residual less heavily. Specifically,

$$\mathrm{Lim}_{P_k \to 0} K_k = 0 \qquad (12)$$

Another way of describing the weighting by K is that as the measurement error covariance R approaches zero, the actual measurement Z_k will be "depended" on more and more, whereas the predicted measurement $H\hat{x}_k^-$ is depended on less and less. On the other hand, as the priori estimate error covariance P_k^- approaches zero, the actual measurement Z_k is depended on less and less, and the predicted measurement $H\hat{x}_k^-$ is depended on more and more [14-16].

2.3. Probabilistic Origins of the Filter

This section is a short section describing the justification as mentioned in the previous section for eq (9). This justification is rooted in the probability of a priori estimate $\hat{X}_k^-$ conditioned on all prior Z_k measurements (Bayes' rule). For now it is suffice to point out that the Kalman filter maintains the first two moments of the state distribution,

$$E[X_k] = \hat{X}_k \qquad (13)$$

$$E\left[(X_k - \hat{X}_k)(X_k - \hat{X}_k)^T\right] = P_k \qquad (14)$$

The posteriori state estimate of (9) reflects the mean (the first moment) of the state distribution it is normally distributed if the conditions of (2) and (3) are met. The posteriori estimate error covariance of (8) reflects the variance of the state distribution (the second non-central moment).In other words,

$$p(X_k | Z_k) \approx N(E[X_k], E[(X_k - \hat{X}_k)(X_k - \hat{X}_k)^T])$$

$$= N(\hat{X}_k, P_k) \qquad (15)$$

3. IMPLEMENTATION OF KALMAN FILTER TO SPEECH

From a statistical point of view, many signals such as speech exhibit large amounts of correlation. From the perspective of coding or filtering, this correlation can be put to good use. The all pole, or autoregressive (AR), signal model is often used for speech. The AR signal model is introduced as: [17-19]

$$y_k = \frac{1}{1 - \sum_{i=1}^{N} a_i Z^{-i}} W_k$$

(16)

Equation (3.1) can also be written in this form as shown below:

$$y_k = a_1 y_{k-1} + a_2 y_{k-2} + \ldots\ldots + a_N y_{k-N} + w_k \qquad (17)$$

where,

k = Number of iterations;

y_k = current input speech signal sample;
y_{k-N} = $(N-1)^{th}$ sample of speech signal;
a_N = N^{th} Kalman filter coefficient and
w_k = excitation sequence (white noise).
In order to apply Kalman filtering to the speech expression shown above, it must be expressed in state space form as:

$$H_k = X\, H_{k-1} + w_z \quad (18)$$

$$y_k = g\, H_k \quad (19)$$

Where

$$X = \begin{bmatrix} a_1 & a_2 & .... & a_{N-1} & a_N \\ 1 & 0 & .... & 0 & 0 \\ . & . & .... & . & . \\ 0 & 0 & .... & 1 & 0 \end{bmatrix}$$

$$H_k = \begin{pmatrix} y_k \\ y_{k-1} \\ y_{k-2} \\ . \\ . \\ y_{k-N+1} \end{pmatrix} \quad W_k = \begin{pmatrix} w_k \\ 0 \\ 0 \\ . \\ . \\ 0 \end{pmatrix} \quad (20)$$

$$g = [1 \quad 0 \quad .... \quad 0]$$

X is the system matrix, H_k consists of the series of speech samples; w_k is the excitation vector and g, the output vector. The reason of $(k-N+1)^{th}$ iteration is due to the state vector, H_k, consists of N samples, from the k^{th} iteration back to the$(k-N+1)^{th}$ iteration.

The above formulations are suitable for the Kalman filter. As mentioned in the previous chapter, the Kalman filter functions in a looping method. Referring to as a guide in implementing Kalman filter to speech, I denote the following steps within the loop of the filter. Define matrix H^T_{k-1} as the row vector [20-24] .

$$H^T_{k-1} = -[y_{k-1} \quad y_{k-2} \quad \dots \quad y_{k-N}] \quad (21)$$

$$Z_k = H^T_{k-1} X_k + w_k \quad (22)$$

where X_k will always be updated according to the number of iterations, k. Note that when the k = 0, the matrix H_{k-1} is unable to be determined. However, when the time z_k is detected, the value in matrix H_{k-1} is known. The above purpose is thus sufficient enough for defining the Kalman filter, which consists of:

$$X_k = [I - K_k H^T_{k-1}]\, X_{k-1} + K_k\, z_k \quad (23)$$

Where

$$I = \begin{pmatrix} 1 & 0 & .... & 0 & 0 \\ 0 & 1 & .... & 0 & 0 \\ . & . & .... & . & . \\ . & . & .... & . & . \\ 0 & 0 & .... & 1 & 0 \\ 0 & 0 & .... & 0 & 1 \end{pmatrix}$$

$$K_k = P_{k-1} H_{k-1} [H^T_{k-1} P_{k-1} H_{k-1} + R]^{-1} \quad (24)$$

Where K_k is the Kalman gain matrix, P_{k-1} is the a priori error covariance matrix, R is measurement noise covariance, and

$$P_k = P_{k-1} - P_{k-1} H_{k-1} [H^T_{k-1} P_{k-1} H_{k-1} + R]^{-1} . H^T_{k-1} P_{k-1} + Q \quad (25)$$

Where P_k is the a posteriori error covariance matrix;
And

$$Q = \begin{pmatrix} 1 & 0 & .... & 0 & 0 \\ 0 & 1 & .... & 0 & 0 \\ . & . & .... & . & . \\ . & . & .... & . & . \\ 0 & 0 & .... & 1 & 0 \\ 0 & 0 & .... & 0 & 1 \end{pmatrix}$$

Thereafter the reconstructed speech signal, Y_k

after Kalman filtering will be formed in a manner similar to eq (27).

$$Y_k = a_1 Y_{k-1} + a_2 Y_{k-2} + \ldots\ldots + a_N Y_{k-N} + w_k \qquad (26)$$

Since the value of y_k is the input at the beginning of the process, there will be no problem forming $H^T{}_{k-1}$.In that case a question rises, Y_k formed by the parameters w_k and $\{a_i\}^N{}_{i-1}$ are determined from application of the Kalman filter to the input speech signal Y_k That is in order to construct Y_k,we will need matrix X that contains the Kalman coefficients and the white noise, w_k which both are obtained from the estimation of the input signal. This information is enough to determine HH_{k-1}.

Where

$$HH_{k-1} = \begin{pmatrix} Y_{k-1} \\ Y_{k-2} \\ Y_{k-3} \\ \ldots \\ Y_{k-N+1} \end{pmatrix}$$

3.1. Cross Correlation

Cross Correlation is actually a method for measuring the similarity of two waveforms based upon the amount of common components contained within the two waveforms. The purpose of this technique is to compare the differences in the time frame of two signals. Before we move on to the results of cross correlation between the input speech signals and the output speech signals, let us briefly take a look at cross correlation.

Cross correlation of f_1 (t) and f_2 (t) is defined as:

$$C_{12} = \frac{\int_{t_1}^{t_2} f_1(t) f_2(t)\,dt}{\sqrt{\int_{t_1}^{t_2} f_1(t)^2\,dt \int_{t_1}^{t_2} f2(t)^2\,dt}} \qquad (28)$$

The magnitude of the integral in the numerator of (4.2) is an indication of the similarity of those two signals. If,

$$\int_{t_1}^{t_2} f_1(t) f_2(t)\,dt = 0 \qquad (29)$$

It means that the two signals will have no similarity over the time interval (t_1, t_2).

In general, the cross correlation function will be modified accordingly to

$$C_{12}(\tau) = \int_{-\infty}^{\infty} f_1(t) f_2(t-\tau)\,dt = 0 \qquad (30)$$

Whether Then, C_{12} will be a measurement of the similarity of two particular signals over an entire interval ($-\infty$, ∞), whereby τ is the time shift parameter. This integral determines f_2 is shifted in time relative to f_1.

4. RESULTS

The following results are obtained by setting 5 Kalman coefficients of a 5th order Kalman filter. -0.8,0.2, -0.6, o.7 and -0.4. Different coefficients at different iterations are shown in Table 1.

$$y_k = (-0.8) y_{k-1} + (0.2) y_{k-2} + (-0.6) y_{k-3} + (0.7) y_{k-4} + (0.4) y_{k-5} + w_k$$

Table 1. Different Coefficients at Different Iterations

Coefficient/ Iterations	1st Set	2nd Set	3rd Set	4th Set	5th Set
1-2000	-0.7	0.3	-0.6	0.7	-0.5
2001- 4000	-0.4	0.5	-0.2	0.5	-0.2
4001- 6000	-0.5	0.3	-0.5	0.4	-0.6
6001- 8000	-0.6	0.5	-0.4	0.2	-0.4
8001-10000	-0.4	0.6	-0.1	0.3	-0.2

4.1. Speech Samples of s180.od.

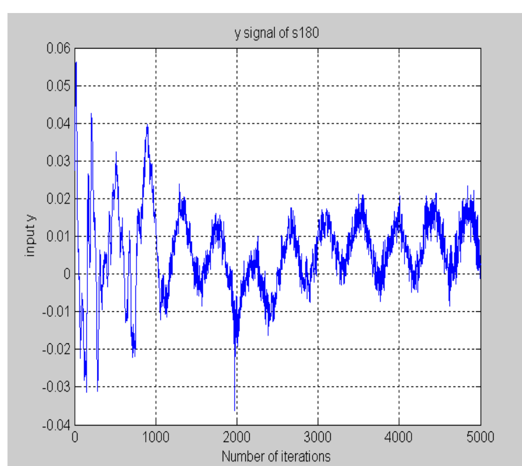

Figure 2. Input Signal of s180.od.

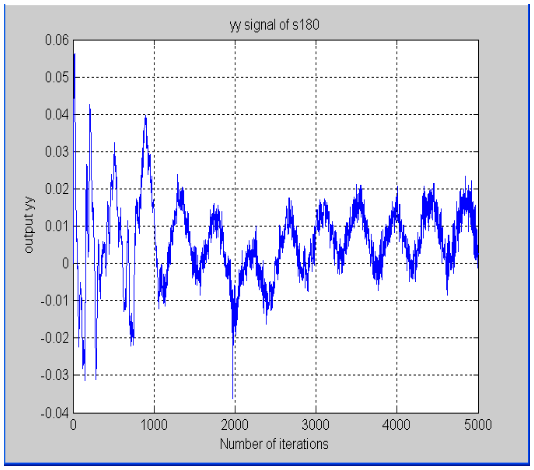

Figure 3. Output Signal of s180.od.

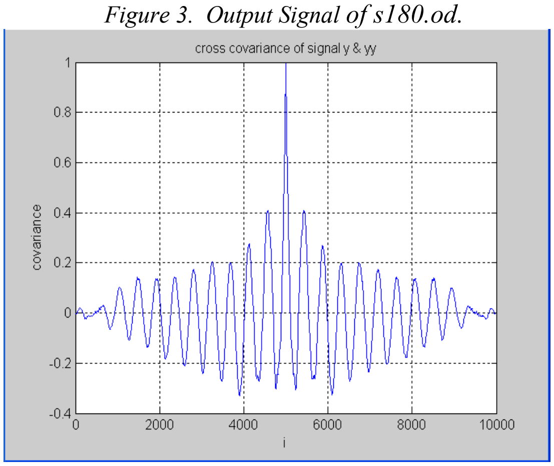

Figure 4. Cross Correlation of s180.od.

4.2. Speech Samples of s680.od.

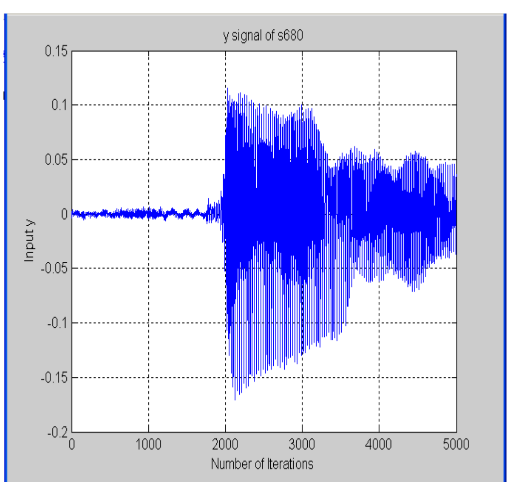

Figure 5. Input Signal of s680.od.

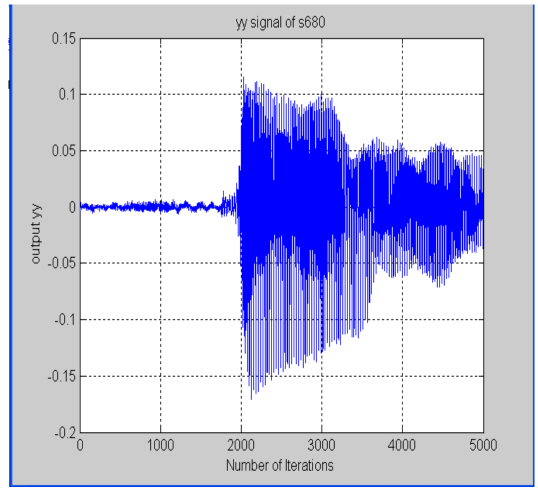

Figure 6. Output Signal of s680.od.

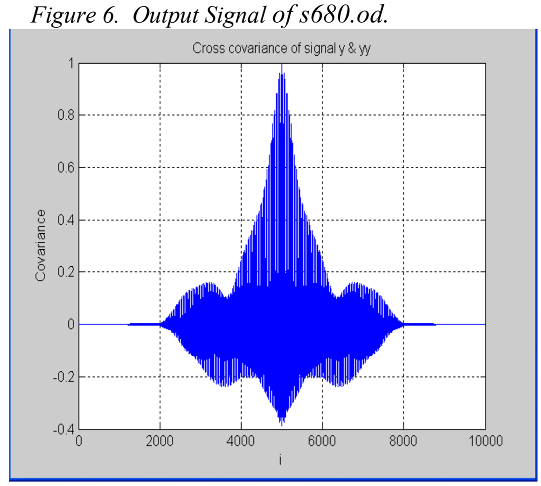

Figure 7. Cross Correlation of s680.od.

4.3. Speech Samples of s1180.od.

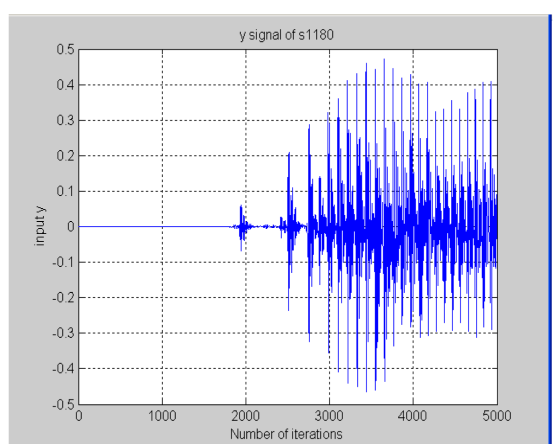

Figure 8. Input Signal of s1180.od.

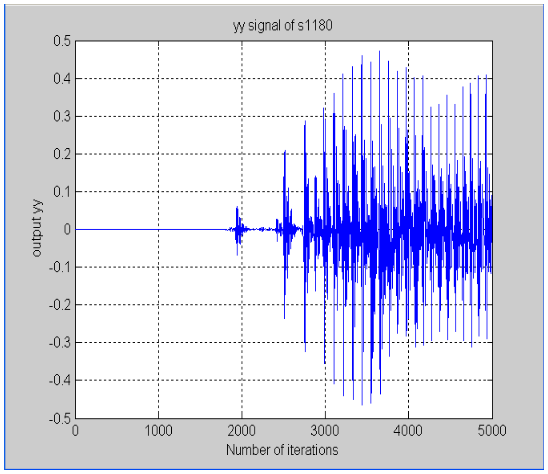

Figure 9. Output Signal of s1180.od.

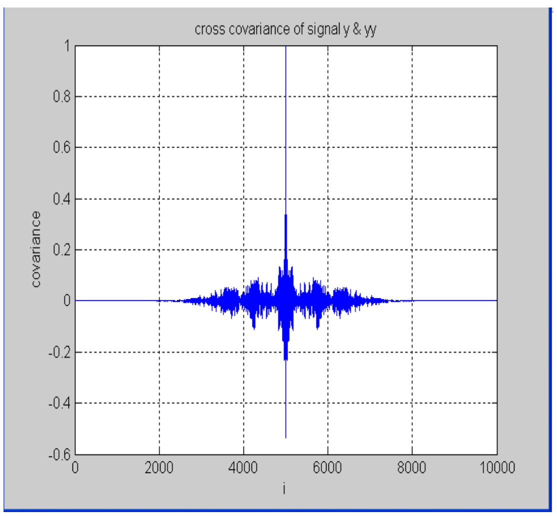

Figure 10. Cross Correlation of s1180.od.

5. CONCLUSIONS

In this paper, an implementation of employing Kalman filtering to speech processing had been developed. As has been previously mentioned, the purpose of this approach is to reconstruct an output speech signal by making use of the accurate estimating ability of the Kalman filter.

Furthermore, the results have also shown that Kalman filter could be tuned to provide optimal performance. With the introduction of tuning parameters Q & R, output speech signals can be obtained similar to the input speech signals. Additional testing on different orders of the Kalman filter when applied to speech had also been conducted.

As a consequence, parameter Q has to be tuned in order to meet the objective. However, parameter R is of superfluous to be tuned. Moreover, a test for cross correlation had also been conducted during this paper for measuring the similarity of the input and output speech signals. This test is of necessity for the reason that different signals are bound to be similar but not identical. By and large, this paper has been quite successful in terms of achieving the objectives. Consequently, perception on signal processing and Kalman filter had also been treasured throughout the process.

7. REFERENCES

[1] R. G. Brown and P. Y. C. Hwang, Introduction to Random Signals and Applied Kalman Filtering 2nd Edition, John Wiley & Sons, Inc., 1992.

[2] A. Gelb, Applied Optimal Estimation, MIT Press, Cambridge, MA, 1974.

[3] M. S. Grewal and A.P. Andrews, Kalman Filtering Theory and Practice, Upper Saddle River, NJ USA, Prentice Hall, 1993.

[4] Jacobs, O. L. R. Introduction to Control Theory 2nd Edition, Oxford University Press, 1993.

[5] Julier, Simon and Jeffrey Uhlman. "A General Method of Approximating Nonlinear Transformations of Probability Distributions," Robotics Research Group, Department of Engineering Science, University of Oxford, Nov 1995.

[6] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A New Approach for Filtering Nonlinear Systems", Proceedings of the 1995 American Control Conference, Seattle, Washington, pp 1628-32.

[7] Kalman, R. E. 1960. "A New Approach to Linear Filtering and Prediction Problems," Transaction of the ASME--Journal of Basic Engineering, pp. 35-45, March 1960.

[8] Lewis, Richard, Optimal Estimation with an Introduction to Stochastic Control Theory, John Wiley & Sons, Inc., 1986.

[9] Maybeck, Peter S. 1979. Stochastic Models, Estimation, and Control, Volume 1, Academic

Press, Inc.

[10] Sorenson, H. W. 1970. "Least-Squares estimation: from Gauss to Kalman,"IEEE Spectrum, vol. 7, pp. 63-68, July 1970.

[11] A. Gelb, J. F. Kasper, Jr., R. A. Nash, Jr., C. F. Price, and A. A. Sutherlandd, Jr.,Applied Optimal Estimation, MIT Press, Cambridge, MA 1974.

[12] M.S. Grewal and A.P.Andrews, Kalman Filtering Theory& Practice Using MATLAB 2nd edition,John Wiley&Sons, Canada, 2001, pp 1-12.

[13] M.S. Grewal and A.P. Andrews, Kalman Filtering Theory and Practice Using MATLAB 2nd edition, John Wiley & Sons, Canada, 2001, pp 16-17.

[14] R.G. Brown and P.Y.C. Hwang, Introduction to Random Signals & Applied Kalman Filtering with MATLAB EXERCISES & Solutions 3rd edition, John Wiley & Sons, Canada, 1997, pp 242-252.

[15] M.S. Grewal and A.P. Andrews, Kalman Filtering Theory and Practice UsingMATLAB 2nd eition, John Wiley & Sons, Canada, 2001, pp 15-17

[16] R. P. Ramachandran and R. Mammone, Modern Methods of Speech Processing, Kluwer Academic Publishers, Massachusetts, USA, 1994.

[17] A.N. Ince, Digital Speech Coding, Kluwer Academic Publishers, Massachusetts, USA, 1992.

[18] A.N. Ince, “Overview of Voice Communications and Speech Processing”,Digital Speech Coding, Kluwer Academic Publishers, Massachusetts, USA,1992, pp 2-41

[19] C. Wheddon & R. Linggard, “Low Bit Rate Speech”, Speech and Language Processing, Chapman and Hall, 1990, pp 31-33.

[20] R.V. Cox, “Current Methods of Speech Coding”, International Journal of High Speed Electronics and Systems, Vol. 8, No. 1, World Scientific Publishing Company, 1997, pp-13-68.

[21] S. Crisafulli, J.D. Mills, and R.R Bitmead, “Kalman Filtering Techniques in Speech Coding”. In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, San Francisco, March 1992.

[22] B.D.O. Anderson and J.B.Moore,“The Kalman Filter,” Optimal Filtering, Prentice Hall Inc., Englewood Cliffs, N.J., 1979, pp. 50-52.

[23] C. R. Watkins, “Practical Kalman Filtering in Signal Coding”, New Techniques in Signal Coding, ANU, Dec 1994.

[24] S. Saito and K. Nakata, “Digitization”, Fundamental of Speech Processing”, Academic Press, Japan, 1985, pp 10-11.