Speech Enhancement and Recognition using Kalman Filter Modified via Radial Basis Function

Mario Barnard
Department of Electrical and Computer Engineering
Oakland University
Rochester, MI, USA
Email: mmbarna2 [AT] oakland.edu

Amr S. Mahmoud
Department of Electrical and Computer Engineering
Oakland University
Rochester, MI, USA
Email: amahmoud [AT] oakland.edu

Farag M. Lagnf
Department of Electrical and Computer Engineering
Oakland University
Rochester, MI, USA
Email: flagnf [AT] oakland.edu

Mohamed Zohdy
Department of Electrical and Computer Engineering
Oakland University
Rochester, MI, USA
Email: zohdyma [AT] oakland.edu

Abstract—**In this paper, a Radial Basis Function-based Kalman filter has been utilized to perform speech enhancement of an audio signal. Moreover, in order to accomplish speech recognition, correlation after detecting signal envelop has been applied. Based on the simulation result, it shows that using the radial basis function-based Kalman filter (non-linear functions to estimate Q parameter) should lead to obtain better results.**

Index Terms— Kalman Filter, Radial Basis Function, Speech Recognition

1. INTRODUCTION

Over the past decade, there has been enormous research in the area of speech recognition since all technologies adopted it in order to make people's lives much easier. The use of Kalman filter in the speech enhancement process to further recognize the voice input to the system and hence the ability to compare the input with an array of data, allowing the system to specifically recognize the phrase. However, speech is mostly accompanied with noise whether this noise is significant or not is determined by the environment in which the measurement is captured. The Kalman filter is used to enhance the signal input through noise evaluation (Q) and cancellation [1].

Furthermore, to allow for maximum efficiency and system high performance sparse sampling is applied (2). The noise covariance (Q) is then assumed based upon the nonlinear radial basis function. In order to contribute to the systems efficiency, the values of the covariance is changed non-linearly over the process for better performance over least time, in comparison to using a linear function to generate different covariance values over the process which would increase the time needed for higher performance.

2. MEASURED DATA

Voice signals such as "Hello", "Estimation", and "Oakland" have been recorded to utilize them in this paper. Each of those signals represents three-second sensor measurements obtained from three different people and different sensors (different microphones). After applying the Kalman filter to enhance these signals from noise, 9 voice signals would be stored in a database. As a result, speech recognition would be accomplished by matching a test voice signal with the database. Graphically, to illustrate those voice signals, the "Estimation" signal of three people has been depicted in time and frequency domain as shown in Figures (1) and (2). Figures (3) and (4) represent Hello, Estimation, and Oakland, which are voice signals obtained by another person in order to examine the system for speech recognition purposes. Those phrases will be identified if the sound waves are primarily auto-correlated. It would double the data amplitude and, hence, the phrase would be identified where the input data was noisy or disrupted. The data would be enhanced via the Kalman filter to match the array of data base. Each sensor measurement of a word "Estimation" has been depicted in the time and frequency domain with background noise as shown in the following Figures (1) and (2).
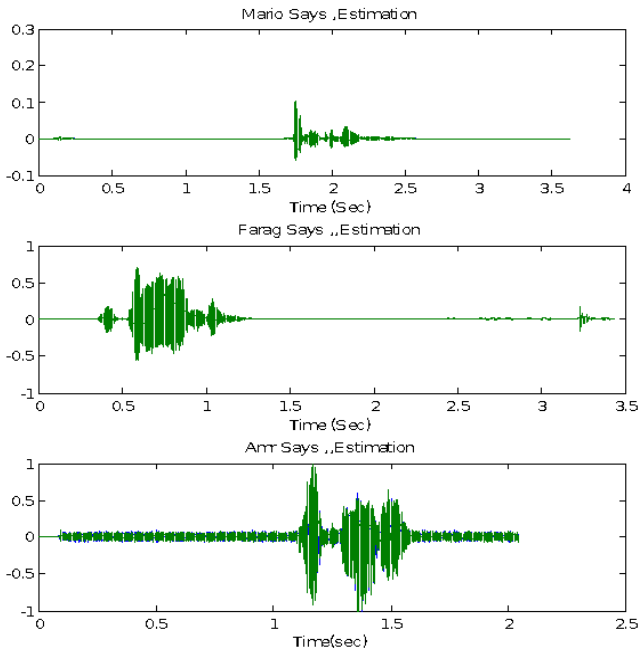
Figure 1: Time Domain of a Voice Signal "Estimation" obtained by three different sensors (Mics) and people
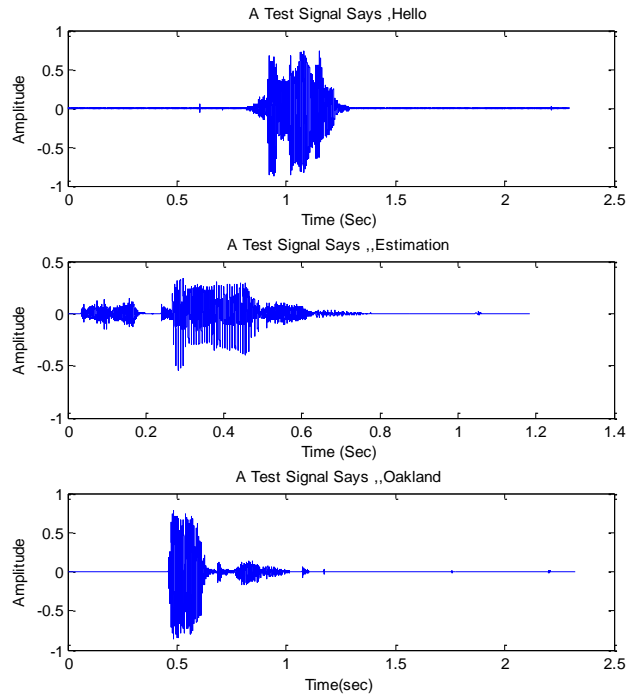

Figure3: Time Domain of Voice Test Signals,"Hello","Estimation", and "Oakland" obtained by a Person
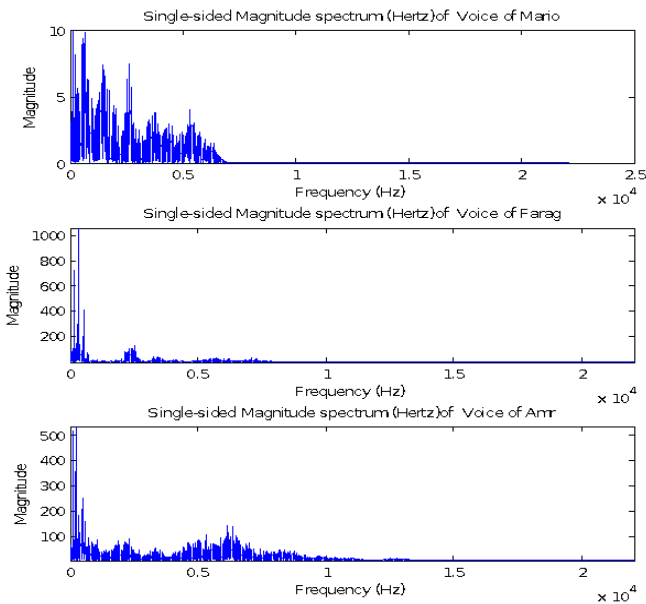

Figure 2: Frequency Domain of a Voice Signal "Estimation" obtained by three different Sensors and People
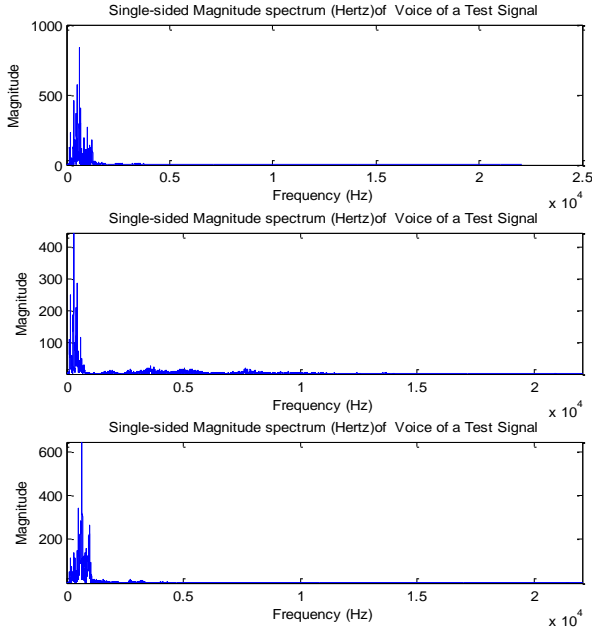
Figure 4: Frequency Domain of Voice Test Signals"Hello',"Estimation", and "Oakland" obtained by a Person

3. RADIAL BASIS FUNCTION (RBF)

RBF is a radially symmetric function is around some point x_m called function's center of dataset as illustrated in equation (1).

$$\sum_{m=1}^{N} w_m \exp(-\gamma \|x_n - x_m\|^2) = y_n \quad (1)$$

Where w_n represents the weights of the centers x_m of dataset (x_n, y_n), while γ affects the shape of radial basis function [7].

A. Calculating Weight of RFB

$$\underbrace{\begin{bmatrix} \exp(-\gamma\|x_1-x_1\|^2) & \dots & \exp(-\gamma\|x_1-x_N\|^2) \\ \exp(-\gamma\|x_2-x_1\|^2) & \cdots & \exp(-\gamma\|x_2-x_N\|^2) \\ \vdots & \vdots & \vdots \\ \exp(-\gamma\|x_N-x_1\|^2) & \dots & \exp(-\gamma\|x_N-x_N\|^2) \end{bmatrix}}_{\phi} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} (2)$$

In the equation (2), the w vector can be found by the following equation (2) if **ϕ** is invertible:

$$W_{n\times 1} = \phi^{-1}{}_{n\times n} Y_{n\times 1} \quad (2)$$

4. KALMAN FILTER

The Kalman filter is one of the most important optimal filters to estimate the state of a dynamic system from a series of noise measurements and resolve an error which is contained in speech, after passing it through a distorted channel. For this reason, it was found that the Kalman filter could be used to solve and suppress the very slow computations of the Wiener filter, which is also one of the common adaptive filtering techniques applied to speech [2].

$$X(k) = \varphi x(k-1) + Gu(k) \quad (3)$$

where the dimension of X(k) matrix is the (p×1) state vector matrix, while the dimension φ is the (p×p) state transition matrix that uses LPCs calculated from noisy speech according to 1.8, G is the (p×1) input matrix and u(k) is the noise corrupted input signal at the kth instant. When speech is noise corrupted, the output y(k) is given as:

$$y(k) = x(k) + w(k) \quad (4)$$

Where w(k) is the measurement noise, a zero-mean Gaussian noise with variance σ^2_w.

In vector form, this equation may be written as
$$y(k) = Hx(k) + w(k) \quad (5)$$

Where H is the (1×p) observation matrix given by
H =(0 0 · · · 0 1)

The application of Kalman filter for speech enhancement is divided into two phases which are [3]:

1. Estimation of AR coefficients $a_1,\dots$, and a_p, Q process noise, and measurement noise R. The Kalman filter requires information about the expected noise (Q) in the sensor measurements, and this information is provided through the Gaussian Radial Basis Function (RFB)

2. Implementing Kalman filter algorithm based on those estimations.

5. SIGNAL ENVELOP

Each audio signal has two aspects such as voiced data and silent data. By eliminating the silent data of the audio signal, it would save time in the computational process during speech recognition. As a result, envelope detection was applied to all audio signals in order to keep voiced data. The results are demonstrated below in figures (5) and (6).
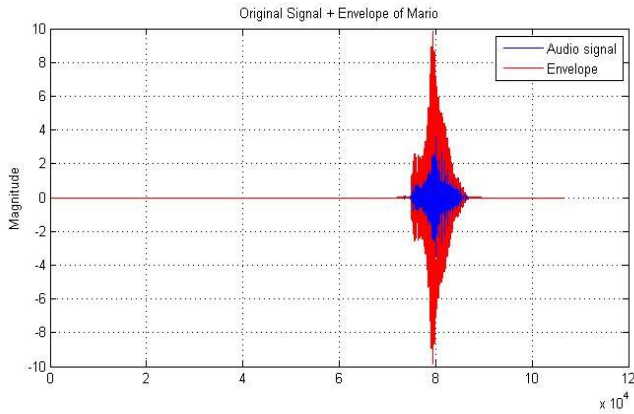

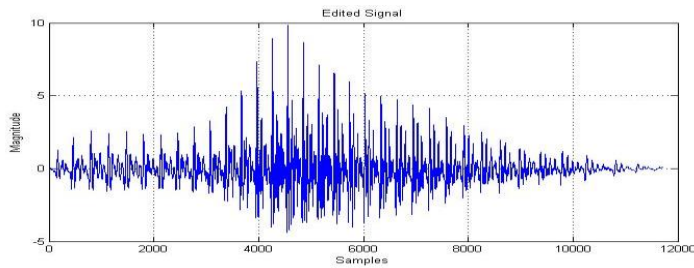*Figure 5: Envelope Detection of an audio signal (voice and silence)*

Figure 6: Voiced data of the audio signal

6. SIMULATION RESULTS

After adding noise to the audio signal and applying the Kalman filter modified by RBF, it is clear from Figure (7) that the cancelation of noise is not that effective.
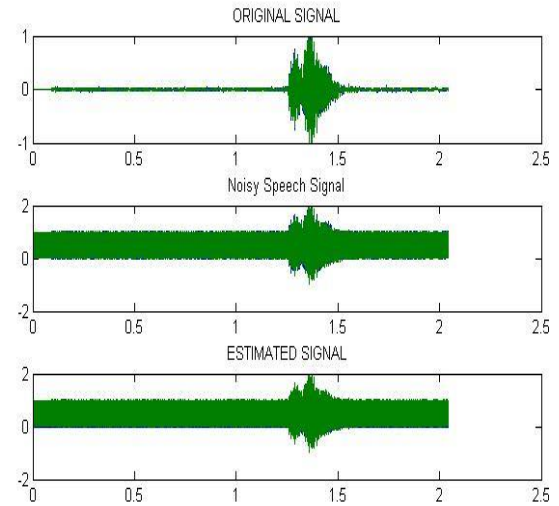

Figure 7: Results of implementing RBF-based Kalman Filter

7. CONCLUSION

In this paper, the Kalman filter modified by the radial basis function technique (the non-linear function instead of the linear function indicated in [1] to obtain the Q process noise is proposed for the speech enhancement. Its performance has found no positive effects on the results since many parameters of that technique such as weights w and Gamma γ needs to be estimated to obtain decent results.

REFERENCES

[1] M. A. Zohdy, Aftab Ali Khan, Paul Benedict "Fused multi-sensor data using a Kalman filter modified with interval probability support" American Control Conference, June 1995

[2] Yeh Huann Goh , Paramesran Raveendran, Yann Ling Goh "Robust speech recognition system using bidirectional Kalman filter" IET Signal Processing, April 2015.

[3] K.K. Paliwal, A. Basu, "A speech enhancement method based on kalman filtering"

[4] O. Das, "Kalman Filter in Speech Enhancement", Jadavpur University, April 2016.X. Gao, X. Zhong, D. You, and S, "Kalman filtering compensated by radial basis function neural network for seam tracking of laser welding," IEEE TRANS, vol. 21, no. 5, Septemper 2013.

[5] R. Yokota, L.A. Barba, M. G. Knepley "A parallel O(N) algorithm for radial basis function interpolation with Gaussians", Computer Methods in Applied Mechanics and Engineering p. 1793–1804, 2010.

[6] M. Mongillo, "Choosing Basis Functions and Shape Parameters forRadial Basis Function Methods" , 2011.

[7] M. Berouti, R. Schwartz and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," Proc. of the IEEE Conference on Acoustics, Speech and Signal Processing, April 1979, pp. 208-211.

[8] M. Gabrea, "Adaptive Kalman filteringbased speech enhancement algorithm," in Proc. of Canadian Conference on Electrical and Computer Engineering, Federicton, NewBrunswick, vol.1, 2001, pp. 521-526.

[9] M. Gabrea, "Robust adaptive Kalman filtering-based speech enhancement algorithm," in Proc. ICASSP'04, 2004, pp. 301-304.