

GENERAL KALMAN FILTER & SPEECH ENHANCEMENT FOR SPEAKER IDENTIFICATION

Vijay Kiran Battula and Appala Naidu Gottapu

Department of ECE, JNTUK-UCEV, Vizianagaram, Andhra Pradesh, India

ABSTRACT

Presence of noise increases the dimension of the information. A noise suppression algorithm is developed with an idea of combining the General Kalman Filter and Estimate Maximization (EM) frame work. This combination is helpful and effective in identifying noise characteristics of an acoustic environment. Recursion between Estimate step and Maximization step enabled the algorithm to deal any model of noise. The same Speech enhancement procedure is applied in the pre-processing stage of a conventional Speaker identification method. Due to the non-stationary nature of noise and speech adaptive algorithms are required. Algorithm is first applied for Speech enhancement problem and then extended to using it in the pre-processing step of the Speaker identification. The present work is compared in terms of significant metrics with existing and popular algorithms and results show that the developed algorithm is dominant over them.

KEYWORDS

Speech processing, Speech enhancement, Speaker identification, General Kalman filter and EM algorithm

1. INTRODUCTION

The two major applications of the speech processing namely speech enhancement and speaker identification are inter related in terms of the core techniques used for performing them. Speech enhancement directly resembles the pre-processing stage of the speaker identification procedure in most of the cases. The work depicts the relation and significance of developing an algorithm to address the core areas of their respective step wise procedure is the area of interest.

Speech and noise in a natural environment always finds in a combined form, Speech is every time degraded by the noise. General assumption is to always find an addition of Gaussian noise to the speech information signal. This noise varies in its characteristics for different acoustic environments. The idea is to develop a robust algorithm to deal a noise of any environment that is degrading speech. Elimination of noise could be even possible through the transducers like microphone used to record the speech information and convert them to the electrical equivalent representation. There are techniques developed based on the number of microphones to be used

to identify the noise components in the signal. The present work is based on Single microphone speech enhancement.

Noise in the problem is first modelled as an Auto Regressive process. This performed as in [3]. Estimate Maximization frame work is studied which involves speech enhancement and as well as parameter estimation. General Kalman Filter(GKF) is a time domain algorithm first studied for using it system identification problem like echo cancellation in [4]. In Echo cancellation echo path estimation is performed using GKF. For extending the General Kalman filter to the speech enhancement it has to be given with at least one among the knowledge of two characteristics of speech or noise. EM frame work helps in providing the noise characteristics.

The time domain implementation of the Kalman filter is difficult and not dominant. G.Enzner through his studies proposed the effective implementation of Kalman filter in frequency domain [2]. But GKF is time domain implemented version of Kalman filter and hence draws the attention. Speaker Recognition is a part behavioural characteristic of the Biometrics. The two major classifications are Speaker verification and Speaker identification namely. Speaker identification can be either text dependent or text independent. In this paper a text independent speaker identification method is considered with a pre-processing stage of developed algorithm.

2. DEFINITION OF THE PROBLEM AND RELATED CONCEPTS

$x(n)$ is a noisy speech signal, $s(n)$ and $v(n)$ are its two components.

$$x(n) = s(n) + v(n)$$

Clean speech signal is represented with $s(n)$ and it obtained by supressing additive noise $v(n)$ in $x(n)$.

2.1. Aim

Single microphone is used to record the degraded speech. The developed algorithm is aimed to use in studying then noise characteristics and suppression of the noise components from the signal. Thus to provide the speech quality enhancement and intelligibility.

2.2. General Kalman Filter

In [4] General Kalman filter is implemented with two estimation problems. The first is a hidden Markov modelled impulse response coefficient estimation and the second is the estimation of desired response with knowledge of the above estimated coefficients and far end signal. In this work the problem is redefined to fit for a speech enhancement problem. The similar coefficient estimation is equivalent to the change affecting the noise characteristics of the acoustic environment under consideration. And these are updated for find new noise model from previous effects as recursive process by applying estimate maximization frame work.

2.3. Estimate Maximization frame work

There are two steps in this process namely estimation step and maximization step. It is a recursive procedure. And the ending if the recursion is predefined or left to the intelligibility of the algorithm. General Kalman filter provides estimation of the required entity in estimation step. These parameters involved in the estimation are identified and updated in the maximization or parameter estimation step and again fed to estimation step. This repeats as an iterative process until the clean entity of interest is obtained.

2.4. Proposed Methods

The Figure 1 represent the block diagram of combined speech enhancement for speaker identification, the first half of the diagram illustrates the process flow of speech enhancement as follows

- i. Noise degraded speech information is collected through single microphone.
- ii. This is given to the speech enhancement stage.
- iii. The first step of speech enhancement contains the segmentation of speech signal satisfying the stationary condition of the signal.
- iv. Thus frames are extracted from the source information and each frame is now forwarded for noise suppression individually.
- v. The AR model of the noise of the acoustic environment is acquired prior to the above steps and state space formulation is done to it.
- vi. The estimation process starts by giving noise estimate and present speech frame to General Kalman filter to find optimum output.
- vii. The parameters are updated in maximization step and again given to GKF for new estimation
- viii. This Process continues until clean speech frame is obtained for a fixed number of iterations.
- ix. Finally all the frames are concatenated to obtain clean speech.

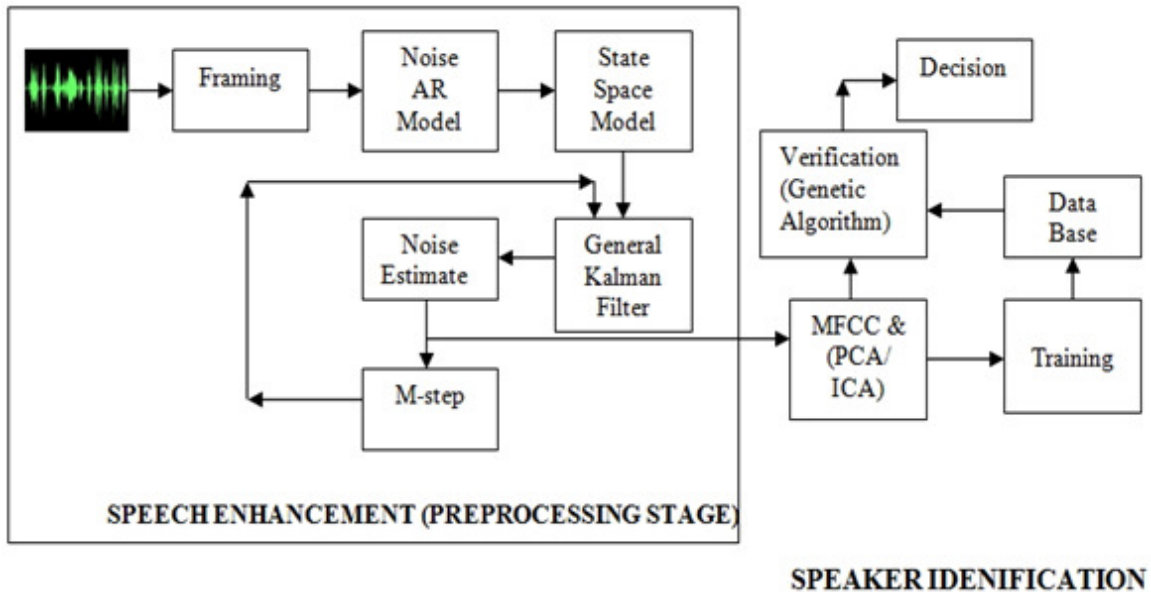


Figure 1. Proposed Speech Enhancement for Speaker identification

Figure 1 represents the proposed method for Speaker identification

- i. The initial stage of speaker identifications always be the speech enhancement stage. The output of the Speech enhancement stage is a clean speech.
- ii. Training phase and Verification are two stages of identification. In training phase speech features of the concerned group are collected and stored in data base.
- iii. In verification stage they are retrieved to compare with the features of persons speech waiting for authentication.
- iv. If the matching criteria satisfies authentication is provided otherwise rejected.
- v. In this paper Eigen features based speech features are extracted in feature extraction stage. MFCC range of frequencies is under consideration.
- vi. Dimensionality reduction techniques like principal component analysis (PCA) and Independent Component analysis (ICA) used for reducing the feature space.
- vii. Genetic Algorithm is used for verification procedure to provide optimum results and the cost function of it is defined below.

$$\sum_{m=1}^L (\max |R(m)|)^2 + \lambda \sum_{m=1}^{L-1} \sum_{n=m+1}^L (\max |P(m,n)|)^2 \quad (1)$$

R=Auto correlation (Correlation between each speech feature itself), P=Cross Correlation (Correlation between speech feature with other speech feature),L=Length of each speech feature, Weighing factor λ

3. ALGORITHM

3.1. Auto Regressive modelling of noise

Let us consider the problem in a discrete time index of n ,

$$x(n) = s(n) + v(n) \tag{2}$$

Noise and speech are non-stationary. Hence, any one of them has to be known in order to go for formulation. Consider the speech absent periods to study noise characteristics,

$$x(n) = v_0(n) \tag{3}$$

Where $v_0(n)$ is signal collected by recording device when speech is absent and use as initial signal vector of noise. Noise is modelled as stochastic AR process:

$$v(n) = -\sum_{q=1}^k \lambda_q v(n-k) + \frac{g_v}{2} u_v(n) \tag{4}$$

$\lambda_1, \lambda_2, \dots, \lambda_k$ represents AR parameters of noise process and g_v represents power level. $u_v(n)$ is normalized (zero-mean unit variance), white Gaussian noise.

3.2. E-step

AR model of noise process is converted into state space formulation

$$v_k^T(n) = [v(n-k+1), v(n-k+2), \dots, v(n)]$$

$$\Phi_v = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ -\lambda_k & -\lambda_{k-1} & -\lambda_{k-2} & -\lambda_{k-3} & \dots & -\lambda_2 & -\lambda_1 \end{bmatrix}$$

is noise transition matrix and k-

dimensional vectors $g_v^T = [0 \ 0 \ \dots \ g_v]$. The below equation provides estimate of noise

$$v_k(n) = \Phi_v v_k(n-1) + g_v(n)u_v(n) \quad (5)$$

$$x(n) = s(n) + h_v^T(n)v_k(n) \quad (6)$$

$$s(n) = x(n) - h_v^T(n)v_k(n) \quad (7)$$

3.3. M-step

Parameters for estimate and AR model, GKF algorithm procedural steps results in finding the minimal error and optimal estimate. The following are the related equations

Initialize with $\hat{h}_v(n) = 0$ and $R_\mu(0) = \mathcal{E}I_q$ where \mathcal{E} is small positive constant

$$\hat{h}_v(n) = \hat{h}_v(n-1) + w(n) \quad (8)$$

$$x(n) = s(n) + h_v^T(n)v_k(n) \quad (9)$$

$$e(n) = x(n) - \hat{v}(n) \quad (10)$$

$$R_m(n) = R_\mu(n-1) + \sigma_w^2(n)I_q \quad (11)$$

$$R_e(n) = v_k^T(n)R_m(n)v_k(n) + \sigma_v^2(n)I_p \quad (12)$$

$$K(n) = R_m(n)v_k(n)R_e^{-1}(n) \quad (13)$$

$$e(n) = x(n) - v_k^T(n)\hat{h}_v(n-1) \quad (14)$$

$$\hat{h}_v(n) = \hat{h}_v(n-1) + K(n)e(n) \quad (15)$$

$$R_\mu(n) = [I_q - K(n)v_k^T(n)]R_m(n) \quad (16)$$

Where $R_m(n)$ is priori misalignment, $R_\mu(n)$ is posteriori misalignment correlation matrix, $K(n)$ is Kalman gain, I_q is identity matrix, $\sigma_w^2(n)$ is variance of $w(n)$, $R_e(n)$ is priori error vector correlation matrix and $e(n)$ is error between signal and estimated.

Let θ be the vector of unknown parameters and given as vector $\theta^T = [\lambda^T \quad g_v]$

$$Z = v_q(n-1)v_q^T(n-1) \quad (17)$$

$$Y = v_q(n-1)v(n) \quad (18)$$

Where updated parameters are defined as

$$\hat{\lambda}^{(q+1)} = -\sum E[Z] \sum Y \quad (19)$$

$$\hat{g}_v^{(q+1)} = \sum \left[Z + \left(\hat{\lambda}^{(q+1)} \right)^T E[Y] \right]. \quad (20)$$

4. RESULTS

4.1. Simulation and Metrics

NOIZEUS speech corpus contains noisy speech signals. These are directly used to test the performance of the proposed algorithm. The similar experiments are conducted with popular algorithms and are used for comparison. Peak signal to Noise ratio and Mean square error are used to evaluate performance for Speech enhancement. Identification rate and elapsed time for a fixed number of generations are compared for Speaker identification. Simulations are performed in MATLAB platform and respective analysis is presented below.

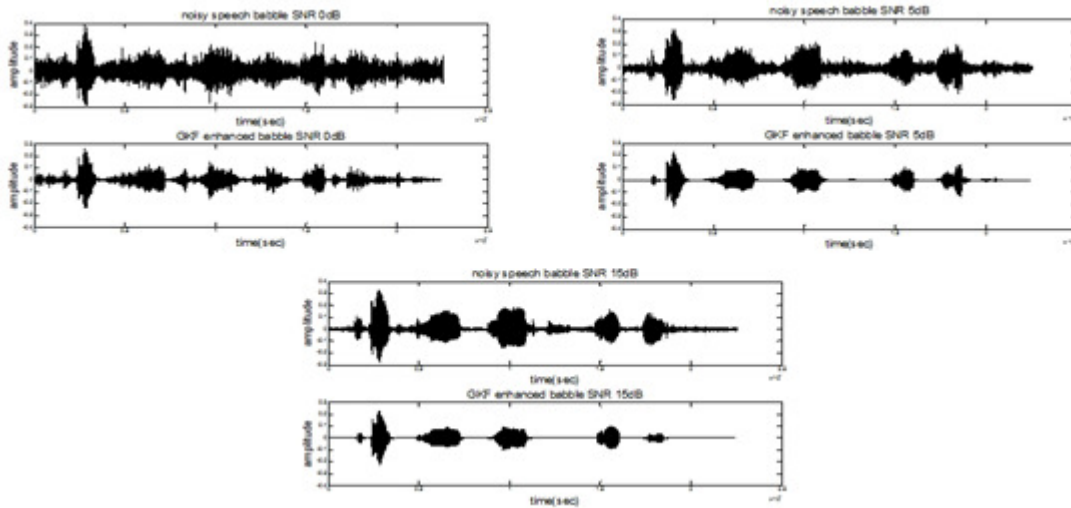


Figure 2. Noisy and GKF enhanced babble noise type of SNR 0dB, 5dB, 15dB

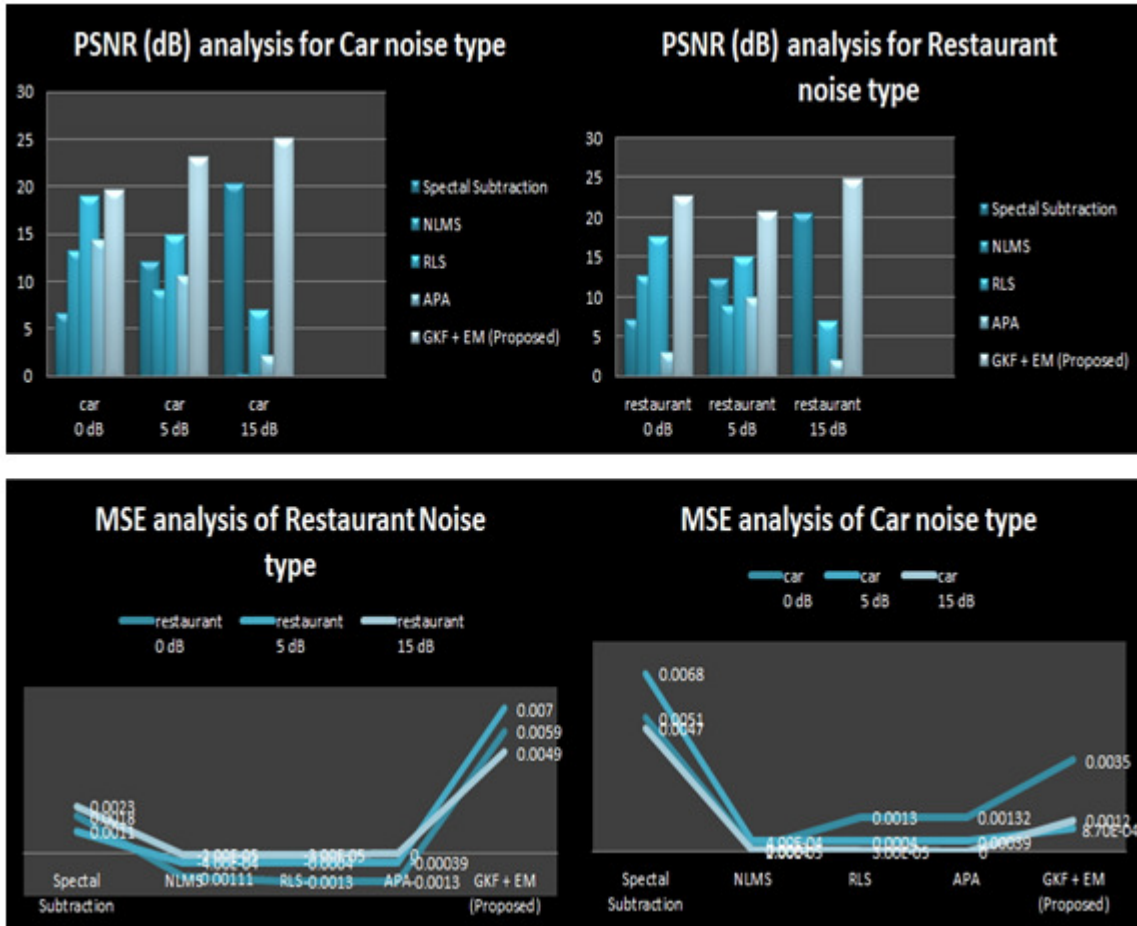


Figure 3. PSNR and MSE comparison of SS, NLMS, RLS, APA and GKF (Speech Enhancement)

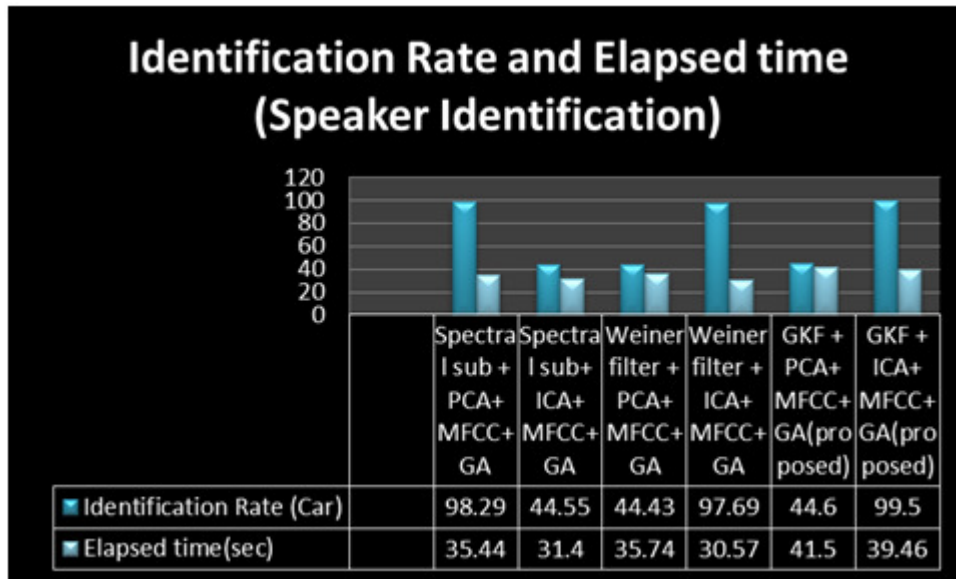


Figure 4. Identification rate comparison of different processing method combinations (Speaker Identification)

5. CONCLUSIONS

General Kalman filter resembles Affine projection algorithm (APA) for certain considerations but the combination with Estimate Maximization is more efficient for GKF. Figure 2 presents the MATLAB simulated results of babble noise type of various SNR levels. Figure 3 is the analysis of proposed algorithm, Normalized least mean square algorithm, Recursive least squares, APA and Spectral subtraction for Speech enhancement applications. Proposed algorithm has given better results in terms of Mean square error and Peak signal to noise ratio. Figure 4 clearly shows the dimensionality reduction effect on the speaker identification system with change in filtration scheme. The respective identification rates and elapsed time of different methods and combination under considerations presented in figure 4 shows GKF is optimum. This work could be further extended to real time and then performance has to analyzed to know its compatibility.

REFERENCES

- [1] R. E. Kalman, "A new approach to linear filtering and prediction problems," J. Basic Eng., vol.82, pp. 35–45, Mar. 1960.
- [2] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," Signal Process., vol.86, pp. 1140–1156, 2006.
- [3] Jae Lim; A. Oppenheim, "All-pole modeling of degraded speech," IEEE Transactions on Acoustics, Speech, and Signal Processing Year: 1978, Volume: 26, Issue: 3
- [4] C. Paleologu, J. Benesty, and S. Ciochina, "Study of the general Kalman filter for echo cancellation," IEEE Trans. Audio, Speech, Language Processing, vol. 21, pp. 1539–1549, Aug. 2013.

- [5] S. E. Bou-Ghazale and K. Assaleh, "A robust endpoint detection of speech for noisy environments with application to automatic speech recognition," in Proc. ICASSP2002, vol. 4, pp. 3808–3811, 2002.
- [6] S. Gannot, "Speech processing utilizing the Kalman filter," IEEE Instrum. Meas. Mag., vol. 15, no. 3, pp. 10–14, Jun. 2012.
- [7] T. Kishore Kumar, Siva Prasad Nandyala, Ravi Bolimera," Speech Enhancement using Spectral Subtraction, Affine Projection Algorithms and Classical Adaptive Filters ," International Conference On Advances In Electronics, Electrical And Computer Science Engineering - EEC 2012
- [8] Sadaoki Furui, "50 Years of Progress in Speech and Speaker Recognition Research," Ecti Transactions on computer and Information Technology, Vol.1, No.2, November 2005
- [9] Omar Daoud, Abdel-Rahman Al-Qawasmi and Khaled daqrouq, "Modified PCA Speaker Identification Based System Using Wavelet Transform and Neural Networks," International Journal of Recent Trends in Engineering, Vol 2, No. 5, November 2009.
- [10] Reynolds, D.A., "Experimental evaluation of features for robust speaker identification," IEEE Transactions on SAP, Vol. 2, pp. 639-643, 2008.
- [11] K. Koteswara rao, G. Appala Naidu, "Eigen features based speaker identification model using Genetic algorithm," VSPICE-2k15 National conference, JNTU KAKINADA

AUTHORS

B.Vijay Kiran received B.Tech degree from JNTUK and currently pursuing M.Tech degree in UCEV Vizianagaram, JNT University Kakinada and Vizianagaram.



G. Appala Naidu received both B.Tech and M.Tech degrees from JNTUH. He is currently working in University College of Engineering Vizianagaram, JNT University Kakinada, and Vizianagaram.

