Scientific Journal of Impact Factor (SJIF): 4.14
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406
International Journal of Advance Engineering and Research Development
Volume 3, Issue 5, May -2016

Single Microphone Speech Enhancement Using General Kalman Filter and Estimate Maximization Frame work

Vijay Kiran Battula[1], Nagaraju Tatipudi[2]

[1]M. Tech Scholar, Department of Electronics and Communication Engineering, JNTUK-UCEV, Vizianagaram
[2]M. Tech Scholar, Department of Electronics and Communication Engineering, GMRIT, Rajam, Srikakulam

Abstract —*The adverse effects of noise from various sources on speech signal are aimed to minimize in every speech enhancement algorithm. Non-stationary behavior of both speech and noise becomes a challenging problem. In this paper we present a method for speech enhancement which uses General Kalman Filter .GKF is one of the forms derived from popular kalman filter mostly used in echo cancellation application. An Estimate maximization (EM)frame work accompanied with GKF deals effectively with noise and its effects. The simulation results show that proposed method is dominant over some popular algorithm's.*

Keywords-*General Kalman Filter, Speech Enhancement, Estimate maximization.*

I. INTRODUCTION

Speech processing has many applications, one such application is speech enhancement. Speech of a speaker is collected using a microphone can be done in various ways. Application requirement and method decides the use of one microphone or more. Speech enhancement deals with improving the quality of speech signal. Present work is related to single microphone speech enhancement. Noise characteristics of a one place differs from another. Similarly speech signal remains non stationary in most of the cases. The problem arises when separating two unknowns from each other. Many algorithms presented for speech enhancement are restricted to some modeled noise characteristics and does not satisfy every possible condition.

Hence, real time approach is very essential to propose any technique. An EM frame work is done to achieve both signal enhancement and parameter estimation. Speech can be degraded by addicted noise (or) by a competitive speaker. Performing Speech quality enhancement and as well as intelligibility improvement is indeed needed. Process the speech signal through any number of microphones to improve intelligibility in harsh conditions. The existing methods are of frequency domain and time domain methods. Spectral subtraction method is predominant one of frequency domain methods proposed by Boll[8] and it requires a very good speech activity detector. Time domain methods like minimizing of the square error between estimated and clean signal proposed by Widrow[7] is widely used for Speech enhancement. Optimal filtering like wiener filter also show a better result since an adaptive filter adjusts itself in accordance with the dynamic behaviour of input signal. Noise cancelling variation of optimal filtering is the area of interest. In the technique using two microphones, primary microphone which collects a noisy signal and secondary microphone detects a signal in correlation with one of the primary microphone picked components, reference microphone is provided is followed by adaption procedure, and its aim is to minimize adaption error this is done by providing delay to picked up signal. Methods related to periodicity like using a comb filter with several taps spaced pitch periods are developed. Speech production study had given rise to LPC(linear predictive coding) modelling of speech which involves Maximum A-posteriori estimate. In [6] iterative Estimate maximize algorithm for maximum likelihood is developed. We implement E step in time domain by Kalman smoother/Kalman filter and maximization of parameters is performed by log likelihood estimation.

A similar procedure [6] is followed using General Kalman filter proposed for echo cancellation by[4] in enhancing a speech signal along with EM frame work.

II. PROPOSED METHOD AND CONCEPTS

2.1. Problem statement

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

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

$s(n)$ represents the clean speech signal and it has to be separated from additive noise $v(n)$.

2.2. Speech enhancement

It is a process of increasing speech quality and intelligibility. A noisy speech signal is segmented, such that stationary condition holds for each frame of speech. Speech enhancement techniques are applied on each frame, until each frame is free from noise.

In most of the cases it is iterative process and has a fixed number of iterations and convergence criteria. VAD (voice activity detection) are used to detect noise characteristics in speech absent periods and thus help in separating speech from noise.

2.3. Proposed method

Speech signal is collected through single microphone. To observe noise characteristics we consider speech absent periods .General procedure of obtaining frames of speech signal is done and algorithm is applied in each frame and finally frames are concatenated .This noise is modeled as Auto Regressive process and an adaption procedure is applied to it. These noise characteristics are fed as initial input to GKF which falls under estimation step, and a recursive procedure takes place between estimation step and Maximization step for fixed number of iterations to get clean speech.Fig.1 below is the block diagram of the proposed method.

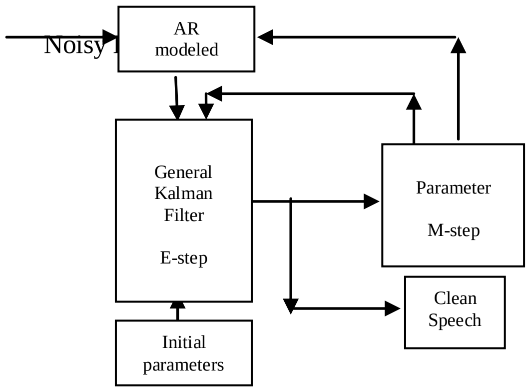

Figure 1. Block diagram of Speech enhancement(proposed)

III. FORMULATION

3.1. Auto Regressive modeling of noise

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

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

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,

$$x(n) = v_0(n) \tag{2}$$

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:

$$v(n) = -\sum_{q=1}^{k} \lambda_q v(n-k) + \frac{g_v}{2} u_v(n) \tag{3}$$

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

3.1.1. E-step

AR model of noise process is converted into state space formulation $v_k^{\,T}(n) = [v(n-k+1), v(n-k+2), \ldots\ldots, v(n)]$

$$\Phi_v = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots\cdots & 0 & 0 \\ 0. & 0 & 1 & 0 & \cdots\cdots & 0 & 0 \\ . & . & . & . & & . & . \\ . & . & . & . & \cdots\cdots & . & . \\ . & . & . & . & & . & . \\ . & . & . & . & \cdots\cdots & . & . \\ -\lambda_k & -\lambda_{k-1} & -\lambda_{k-2} & -\lambda_{k-3} & \cdots\cdots & -\lambda_2 & -\lambda_1 \end{bmatrix}$$ is noise transition matrix and k-dimensional vectors

k-dimensional vectors $g_v^T = [0 \quad 0 \quad \ldots\ldots \quad g_v]$.The below equation provides estimate of noise

$$v_k(n) = \Phi_v v_k(n-1) + g_v(n) u_v(n) \tag{4}$$

$$x(n) = s(n) + h_v^T(n) v_k(n) \tag{5}$$

$$s(n) = x(n) - h_v^T(n) v_k(n) \tag{6}$$

3.1.2. M-step

Parameters for estimate and AR model, GKF algorithm procedural steps results in finding the minimal error and optimal estimate. Our objective is to estimate with an adaptive filter $\hat{h}_v$ .The following are the related equations

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

$$\hat{h}_v(n) = \hat{h}_v(n-1) + w(n) \tag{7}$$

$$x(n) = s(n) + h_v^T(n) v_k(n) \tag{8}$$

$$e(n) = x(n) - \hat{v}(n) \tag{9}$$

$$R_m(n) = R_\mu(n-1) + \sigma_w^2(n) I_q \tag{10}$$

$$R_e(n) = v_k{}^T(n) R_m(n) v_k(n) + \sigma_v^2(n) I_p \tag{11}$$

$$K(n) = R_m(n) v_k(n) R^{-1}{}_e(n) \tag{12}$$

$$e(n) = x(n) - v_k{}^T(n) \hat{h}_v(n-1) \tag{13}$$

$$\hat{h}_v(n) = \hat{h}_v(n-1) + K(n) e(n) \tag{14}$$

$$R_\mu(n) = [I_q - K(n) v_k{}^T(n)] R_m(n) \tag{15}$$

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 $w(n)$ process noise, $R_e(n)$ 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 $\theta^T = \begin{bmatrix} \lambda^T & g_v \end{bmatrix}$

$$Z = v_q(n-1) v_q^T(n-1) \tag{16}$$

$$Y = v_q(n-1) v(n) \tag{17}$$

Where updated parameters are defined as

$$\hat{\lambda}^{(q+1)} = -\sum E[Z] \sum Y \tag{18}$$

$$\hat{g}_v^{(q+1)} = \sum \left[Z + \left(\hat{\lambda}^{(q+1)} \right)^T E[Y] \right]. \tag{19}$$

IV. RESULTS

Simulation is performed using MATLAB software. NOIZEUS, speech corpus is used for evaluation of method. The objective measures of speech quality assessment namely signal to noise ratio(SNR),mean square error(MSE), perceptual evaluation of speech quality (PESQ),Segmental SNR(SNRseg),log likelihood ratio((LLR) are calculated for proposed method and compared with existing methods.

Figure 2. GKF enhancement of SNR 5dB car noise type

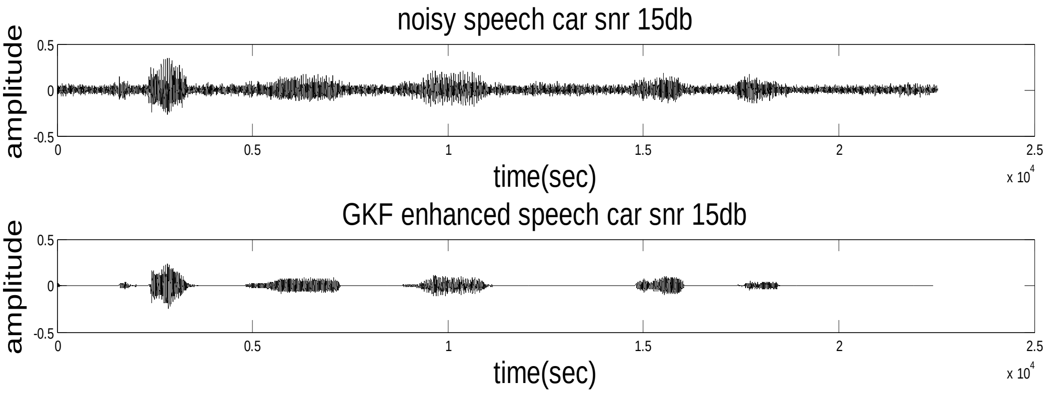

Figure 3. GKF enhancement of SNR 15dB car noise type

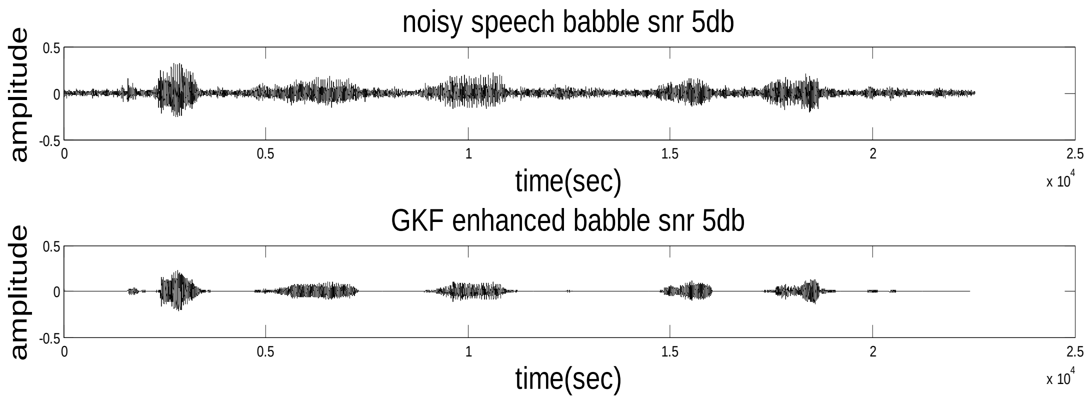

Figure 4. GKF enhancement of SNR 5dB babble noise type

Speech enhancement method	MSE	SNR
Spectral subtraction	0.0012	17.2197
Least mean square	0.2369	6.9123
Normalized LMS	0.0249	20.4030
Recursive Least squares	0.005-.001i	24.0+13.42i
Affine projection algorithm	0.0067	20.5682
Wiener filter	0.0011	21.2077
Kalman filter	0.0185	13.1920
GKF+EM(proposed)	0.0025	21.3466

Table 1. MSE and SNR measures comparison of Speech enhancement algorithms applied on SNR 15dB car noise type

Speech enhancement method	PESQ	LLR	SNRseg
Spectral subtraction	1.74318	1.349	-0.0430
Least mean square	0.9235	1.0846	-10.000
Normalized LMS	1.62133	1.1199	-8.764
Recursive Least squares	1.9089	1.1113	-9.7194
Affine projection algorithm	0.87084	0.8736	-3.464
Wiener filter	1.2563	1.36328	0.06796
Kalman filter	1.5814	1.3727	-9.10184
GKF+EM(proposed)	0.6116	1.5897	-0.8925

Table 2. PESQ,LLR and SNRseg measures comparison of Speech enhancement algorithms applied on SNR 15dB car

IV. CONCLUSIONS

The Figure.2 ,Figure.3 represent the enhanced speech of car noise type speech signal from NOIZEUS data base with SNR of 5 and 15 dB. Figure.4 represent babble noise type of SNR 5dB.The results from Table.1, Table.2 show that utilization General Kalman filter and EM frame for speech enhancement is efficient. The future work of this paper is to test the algorithm in real time environment by considering all possible conditions.

REFERENCES

[1] G. 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 ProcessingYear: 1978, Volume: 26, Issue: 3
[4] C. Paleologu, J. Benesty, and S. Ciochina, “Study of the general Kalman ˇfilter for echo cancellation,” IEEE Trans. Audio, Speech, LanguageProcessing, 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] B. Widrow, J.R. Glover Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H.Hearn, J.R. Zeider, E. Dong Jr., and R.C. Goodlin.”Adaptive Noise Cancelling: Principals and Applications.” Proceeding of the IEEE, 63(12):1692–1716, Dec 1975.
[8] S. F. Boll. “Suppression of acoustic noise in speech using spectral subtraction.”In Jae S. Lim, editor, Speech Enhancement, Alan V. Oppenheim series, pages 61–68. Prentice-hall, 1983.
[9] Yi Hu and Philipos C. Loizou, “Evaluation of Objective Quality Measures for Speech Enhancement.” IEEE Transactions on audio, Speech, and language processing, Vol. 16, No. 1, January 2008.