A Brief Review on Advancements in Kalman Filtering and Phase Based Modulation Domain Speech Enhancement

V.Srinivasarao, Umesh Ghanekar

***Abstract:** In the present paper, Kalman Filter based speech enhancement algorithms have been studied. Starting from the basic Kalman Filter approach to enhance signal to noise ratio against conventional wiener filtering, to the recent modulation domain Kalman filtering that is based on phase of the speech signal for tracking the phase of speech along with logarithmic spectra of noise as well as speech, the improvisation has been carefully observed and presented here. In the modified Kalman filtering algorithm, for the reconstruction of speech signal, speech phase posterior is utilized for developing an improved phase spectrum of the speech. Kalman filter is operated in two steps, one is to model temporally correlating inter frames of speech and logarithmic spectral-amplitudes of noise, where as the second models their nonlinear relations, assuming speech and noise will get added in complex STFT domain. This method is assessed using speech intelligibility and quality metrics, over a range of SNR values with various types of noise. The performance measures highlighted the consistent enhancement in quality of speech over conventional algorithms used for speech enhancement.*

***Keywords:** Kalman Filter, Modulation Domain, Speech Enhancement and Speech Phase.*

I. INTRODUCTION

In the environments which are noisy and non stationary, speech enhancement is a challenging task. Kalman filter (written as KF in short at few places in this paper) is a linear MMSE estimator in the time domain, where the estimation of the enhanced speech signal is carried-out, sample by sample recursively. In the environments that are non-stationary, Kalman filter is used for estimation of both phase and magnitude of speech signal and is termed as a joint estimator. Modulation domain is used in modeling the inter frame correlations temporally rather than independently [1],[2].

Recently modulation domain is being used as an alternate to auditory domain for improving the speech signal. The auditory spectrum is the STFT of a signal and modulation domain is the sequential trajectories of amplitude spectrum throughout the frequency range. Kalman filter is an optimal recursive algorithm for data processing. It is mathematically defined by a set of equations which are computationally efficient for the estimation of a process through minimizing the mean square error [6]. It is a recursive process and does not require storing the previous values. Kalman filter operates through a prediction and correction process.

Revised Manuscript Received on June 05, 2019.

V.Srinivasarao, Research Scholar, ECE Department, National Institute of Technology, Kurukshetra, Haryana, India.

Umesh Ghanekar,Professor, ECE Department, National Institute of Technology, Kurukshetra, Haryana, India.

II. LITERATURE REVIEW

K.K.Paliwal and A.Basu,[3] proposed speech enhancement using KF which performs better than the wiener filtering at that time. Speech can be expressed using an autoregressive process which is an output of a system having a noise sequence as the input. This can be represented in the form of a state space model as:

$$A(n)= \varphi\, A(n-m)+ G\, v(n) \qquad (1)$$

Here A is a state vector, φ is state transition matrix and G is input matrix respectively.The observation procedure for a noisy signal y(n) = s(n)+u(n) is expressed as:

$$Y(n)=H\, A(n) +u(n) \qquad (2)$$

where H is called as observation matrix.

Noises v(n) and u(n) are white and uncorrelated Gaussian random processes with mean zero.

When Kalman filter is applied to these state and observation equations, MMSE of the observations {y(n)}can be obtained and this estimate is denoted by A^(n/n) and covariance error matrix is C(n/n).

A^(n/n-m) is the one step prediction estimate of A(n) and C(n/n-m) is the covariance matrix of error. Using this notation, the algorithm for Kalman filtering is given by the recursive relations:

A^(n/n)=A^(n/n-m) +N(n)[y(n)-H^(n/n-m)]

A^(n/nl) = φA^(n-m/n-m), with A^(0/0)=Ao

C(n/n) = [I-N(n)H] C(n/n-1)

Where

$$N(n) = C(n/n-m)\, H^T[H\, C(n/n-m)H^T+R(n)]^{-1}$$
$$C(n/n-m) = \varphi\, C(n-m/n-m)\, \varphi^T +G\, Q(n)G^T$$

The two step procedure that describes Kalman filtering in speech enhancement is as follows:

(1) Autoregressive coefficients {d1,d2,……. ……dp} and variances of noise Q and R of each stationary speech segment are estimated.
(2) Using these values of the estimated parameters,

apply Kalman filtering recursive algorithm. The end element of the state vector A(n) = [s(n-p+l) . . . . s(n)], i.e., dp^n)=s^(n) provide the estimate of speech signal s(n).

As an extension to this a method of Kalman filtering with delay was also introduced which delays the computation of S^(n) until (n+p-1) and hence this estimate is called as delayed Kalman filter estimate.

Thomas Esch and Peter Vary [4] presented a frequency domain modified Kalman Filter for enhancing the speech in the single channel mode. It is a two step procedure. Firstly, by obtaining the estimate of current speech coefficients, previously enhanced DFT coefficients are exploited. In the second step, by the use of three spectral estimators, the first prediction gain is updated including conventional Kalman filter gain. The objective evaluation measures show the betterment of the proposed technique over conventional Kalman filter [9].

Stephen So, K. K. Paliwal [1], investigated the Kalman filter in modulation domain (MDKF) and the performance comparison was done with other speech enhancement techniques that are time domain and acoustic domain. MDKF is a linear adaptive MMSE estimator which employs the models of temporal changes in speech and noise magnitude spectrum. Kalman filter is the most suitable filter for processing in modulation domain, because role of phase is crucial in modulation domain. Subjective and objective results highlight the performance of MDKF when compared to other time domain and acoustic domain techniques.

General noisy signal model

y(m) = s(m)+u(m) (3)

Here y(m), s(m) are noisy and clean speech signals, where as u(m) denotes the noise signal with zero mean. Speech signal is treated as a quasi stationary signal and is generally analyzed frame by frame with STFT analysis. The STFT of noisy speech y(m) is denoted by Y(m,k), where k is discrete frequency variable, N is duration of the frame. Similarly, the STFTs of clean speech and noise are S (m,k) and U(m,k) respectively. Using STFT analysis, Y (m,k) is written as Y(m,k) = S(m,k)+U(m,k). Now Kalman filter is applied in modulation domain. The magnitude spectrum can be viewd as a combination of N modulating signals spread over the time. Yu Wang and Mike Brookes [7], proposed a modulation domain KF based speech enhancement algorithm operating in time-frequency domain. The Kalman filter here in this enhancement algorithm will combine Gaussian mixture and colored noise models of residual noise [9]. The performance of this algorithm, on the TIMIT set is evaluated and demonstrate the consistent improvement in the performance over both the baseline enhancer and Kalman filter post-processor [8].

T.Mellahi and R.Hamdi [5], presented an iterative Kalman filtering method in which the estimation of linear prediction model parameters, was obtained from corrupted speech signal. Estimates of variance of excitation and LPC accuracy dictate the performance of Kalman filter. Speech analysis with Linear Prediction Coefficients is noise sensitive. This problem may be surpassed with the application of LPC-based formant augmentation method. In this, logarithmic amplitude spectrum of LPC model is modified and the new LPCs are revalued that are to be applied on KF. These new enhanced Linear Prediction Coefficients are good indicators of the performance of the Kalman filter.

Jingxian Tu and Youshen Xia [14], proposed an algorithm for speech enhancement using Kalman filtering in time domain for distributed multichannel case, in colored noise environment. In comparison to conventional algorithms, this algorithm has less computational complexity, better noise reduction, low signal distortion and higher speech intelligibility.

Here segmental SNR, log likelihood ratio, PESQ and STOI are the metrics used for objective evaluation[13].

III. PHASE BASED MODULATION DOMAIN SPEECH ENHANCEMENT USING KALMAN FILTERING

In this an algorithm for enhancing the speech in modulation-domain using Kalman filtering was performed with the help of circular statistics to track the phase of the speech, along with logarithmic-spectra of speech signal and noise [10], [11]. In this algorithm, during reconstruction of speech signal, speech phase posterior was used through creation of improved spectrum of speech phase. The KF prediction step, separately represents the inter-frame correlation of the noise and speech logarithmic spectral amplitudes and the KF update step represents their non-linear relations. Here the assumption is that in complex STFT domain, the noise and speech gets added [12],[15]. Here phase-sensitive speech enhancement is done using modulation-domain Kalman filtering that shows significant improvement in both the objective and subjective evaluation metrics [16].

The degraded speech signal, ~y(m), is given by ~y(m) = ~x(m)+~v(m), where ~x(m) and ~v(m) are the pure speech and noise signals respectively, and m is the index of time in discrete domain. Applying Short Time Fourier Transform, one can obtain Yt(k) = Xt(k) +Vt(k) where t is index of time frame and k is frequency index. In complex Short Time Fourier Transform domain, the equation Y = S + N implies $|Y|\, e^{j\theta} = |X|\, e^{j\varphi} + |V|\, e^{j\psi}$ where θ,φ and ψ are respectively, the phases of degraded speech, pure speech and noise. The STFT logarithmic-spectral magnitudes of the degraded, pure speech and noise signal are denoted respectively by, $y = \log_{10}|Y|$, $s = \log_{10}|S|$ and $n = \log_{10}|N|$.

The algorithm of this enhancement method is depicted in the following figure [15]. The important part of this algorithm is the Kalman Filter (KF) highlighted with a dotted line in the flow chart. The Kalman filter tracks posterior distributions of logarithmic spectra of both the noise and speech as well as of the speech phase. Instead of tracking speech phase, it is appropriate to follow the complex exponential quantity exp(jφ), as it removes the discontinuity at ±π. Initially, the degraded speech signal is transformed by using STFT into frequency domain. In this algorithm, spectral amplitudes and phase of the degraded speech are separately considered. In the prediction step, autoregressive models for exponential phase function and logarithmic spectra of noise and phase are estimated. The linear KF prediction makes use of these results to predict the KF state. Finally, the generation of STFT coefficients is done by combining logarithmic spectrum of estimated speech and phase from the Kalman Filter state vector. The time domain enhanced speech is reconstructed by taking inverse The input and output wave forms for different types of noise in varied SNR levels are shown in the figure below.

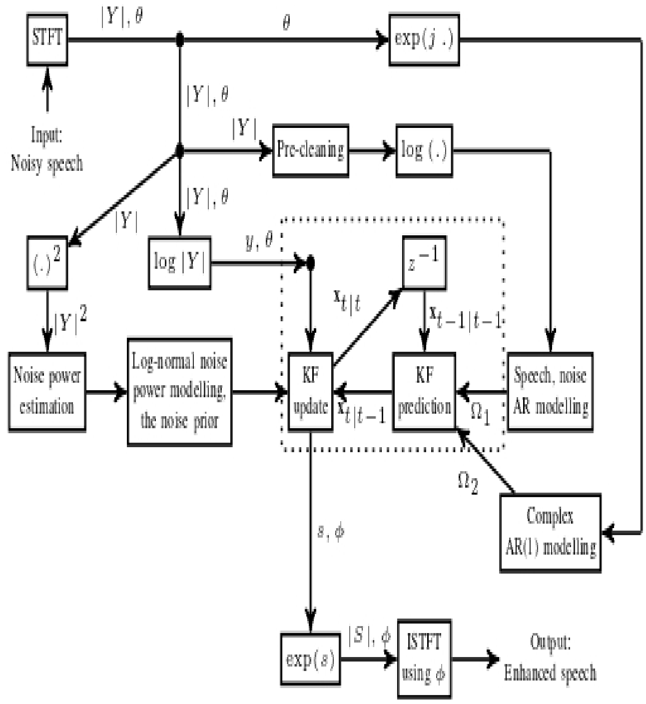
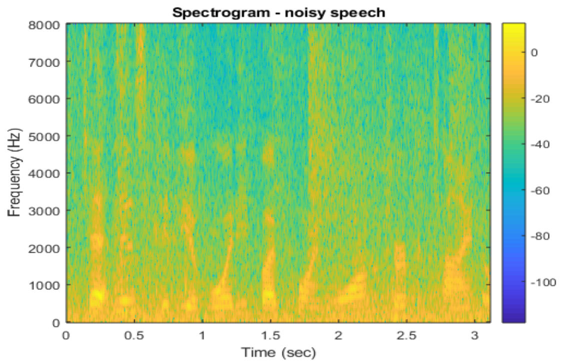


Figure1: Phase based algorithm flow chart



For SNR= -5dB, Babble noise:
PESQ:2.897

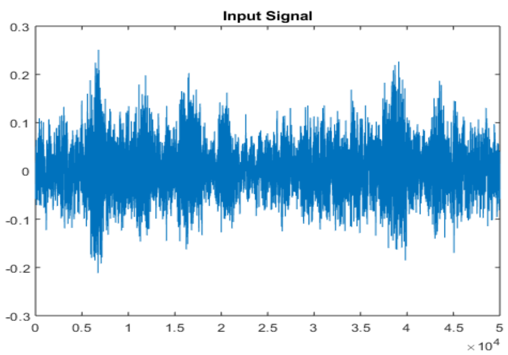


Figure 4: Noisy input and enhanced output speech signals for Babble noise -5dB SNR case.

STFT.

IV. RESULTS

For SNR: 0dB and factory noise:
PESQ:4.569

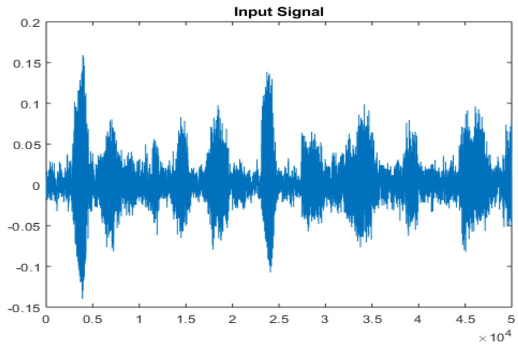


Figure 2: Noisy input and enhanced output speech signals for factory noise 0dB SNR case.

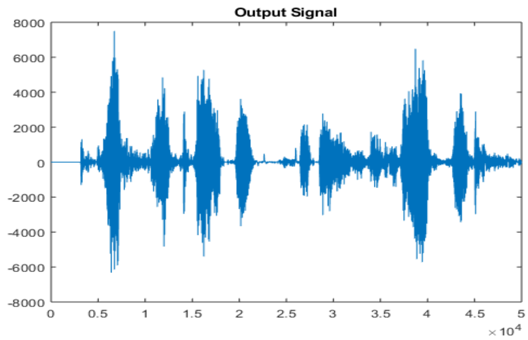
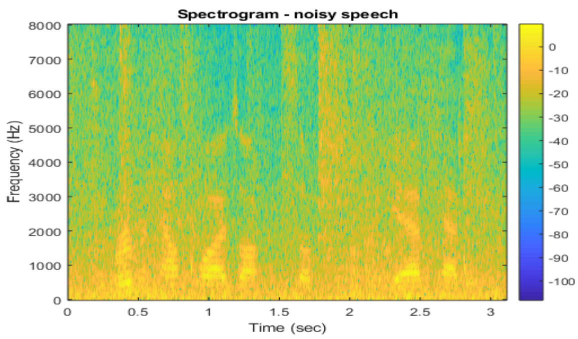


Figure 3: Spectrograms of Noisy input and enhanced output speech signals



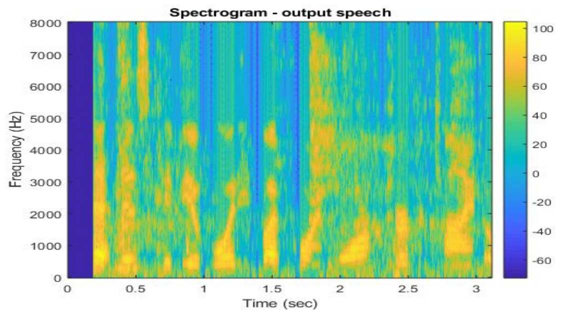
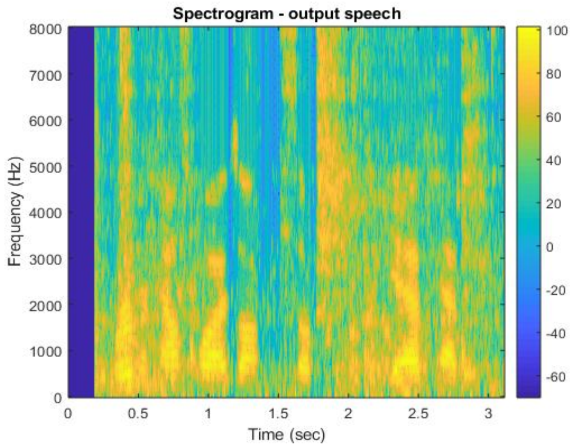
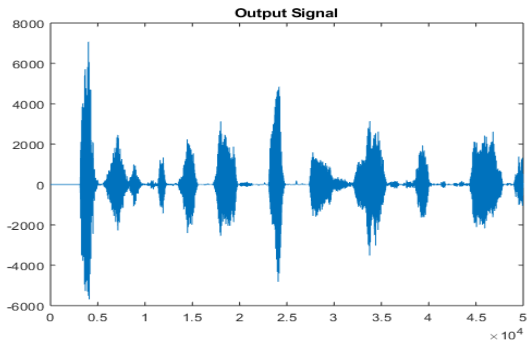


Figure 5: Spectrograms of Noisy input and enhanced



V.CONCLUSION

In this paper, few Kalman filtering methods that are improved gradually over the years in the time, frequency and modulation domains are presented. However, Kalman filtering is still the widely used technique for speech quality and intelligibility improvement. In this work an improved version of Kalman filtering is presented considering the



phase estimate of signal in modulation domain and it proved effective in terms of SNR and other performance metrics in both subjective as well as objective tests.

REFERENCES

1. K. K. Paliwal and S. So, "Modulation-domain Kalman filtering for single-channel speech enhancement," Elsevier Journal of Speech Communication, volume. 53, no. 6, pp. 818-829, July 2011.
2. K. K. Paliwal, S. So and K. K. Wójcicki, "Single-channel speech enhancement using Kalman filtering in the modulation domain," in the Conference Proceedings by International Speech Communication Association, Makuhari, Sept. 2010.
3. K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, volume 12, pp. 177–180, 1987.
4. T. Esch and P. Vary, "Speech enhancement using a modified Kalman filter based on complex linear prediction and supergaussian priors," Proceedings of IEEE International Conference on Audio and Speech Signal Processing, pp. 4877-4880, Las Vegas, April 2008.
5. T.Mellahi and R.Hamdi," LPC-based formant enhancement method in Kalman filtering for speech enhancement", Elsevier International Journal of Electronics and Communications, volume 69, pp 545-554, 2015.
6. T. Esch and P. Vary, "Model-based speech enhancement using SNR dependent MMSE estimation," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4652-4655, Prague, May 2011.
7. Y. Wang and M. Brookes, "Speech enhancement using a modulation domain Kalman filter post-processor with a Gaussian mixture noise model," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 7024-7028, Florence, May 2014.
8. Y. Wang and M. Brookes, "Speech enhancement using a robust Kalman filter post-processor in the modulation domain," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, May 2013.
9. Y.Wang and M.Brookes, " Speech enhancement using an MMSE spectral amplitude estimator based on a modulation Kalman filter with a gamma prior", in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Shanghai, March 2016.
10. N. Dionelis and M. Brookes, "Modulation-domain speech enhancement using a Kalman filter with a Bayesian update of speech and noise in the log-spectral domain," in Proceedings of IEEE Hands-free Speech Communication and Microphone Arrays, California, March 2017.
11. N. Dionelis and M. Brookes, "Speech enhancement using modulation-domain Kalman filtering with active speech level normalized log-spectrum global priors," in Proceedings of European Signal Processing Conference, Kos Islands, August 2017.
12. P. Vary and T. Esch, "Exploiting temporal correlation of speech and noise magnitudes using a modified Kalman filter for speech enhancement," in Proceedings of ITG Conference on Voice Communication, Aachen, Oct. 2008.
13. Y. Hu and P.C. Loizou, "Evaluation of objective quality measures for speech enhancement", IEEE Transactions on Audio, Speech, and Language Processing, volume 16, no. 1, January 2008.
14. Jingxian Tu, Youshen Xia, " Effective Kalman Filtering Algorithm for distributed Multi channel speech enhyancement" Elsevier journal of neuro computing, June 2017.
15. M. Brookes and N.Dionelis, "Phase aware single channel speech enhancement with modulation domain kalman filtering," IEEE Transactions on Audio, Speech and Language Processing, volume 26, no. 5, pp. 937-950, 2018.
16. B. Champagne M. Parchami, and W.P. Zhu, "Speech dereverberation using weighted prediction error with correlated inter-frame speech components," Elsevier Journal of Speech Communication, volume 87, pp. 49-57, 2017.