Speech Enhancement Using Nonlinear Kalman Filtering

T. Namratha, B. Indra Kiran Reddy, M. V. Deepak Chand Reddy, and P. Sudheesh

Abstract Speech enhancement is today a growing necessity for a wide range of applications in which the noise-free speech signal is important and necessary for the processing to be continued. The main purpose of these speech enhancement techniques is on a higher level is to remove noise from the speech signal. The reverberation component in the corrupted speech signal is also removed using the auto-regressive techniques for better performances. In this paper, unscented Kalman filtering which is an adaptive algorithm is proposed that executes both denoising and dereverberation of the speech recorded in adverse conditions. The algorithm relies on the parameter such as mean and covariance of the state spaces created and updating the concerned measurements to provide the optimal denoised and dereverberated signal. This proposed algorithm is assessed with regard to quality of speech, intelligibility of speech and performance metrics like the figure of merit and cross correlation and is also compared with other denoising and dereverberation techniques. The trial outputs on executing the algorithm using the noisy reverberant speech exhibit the adequacy of the proposed adaptive enhancement algorithm.

Keywords Speech enhancement · Adaptive algorithm · Denoising · Dereverberation · Kalman filtering · Auto-regressive filtering · Unscented Kalman filter · Parameter estimation · Unscented transform · Parameter estimation

1 Introduction

These days, innovation is truly advancing with enormous demand, and the interest for speech enhancement frameworks is clear. Speech improvement in uproarious reverberant conditions, for the audience, is hard and testing. The speech signal is corrupted by the noise and resonation when caught utilizing an inaccessible mouthpiece [1]. A room impulse response will incorporate segments at long postponements, subsequently coming about in resonation and echoes. Reverberation is considered to be a

T. Namratha · B. Indra Kiran Reddy · M. V. Deepak Chand Reddy · P. Sudheesh (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: p_sudheesh@cb.amrita.edu

G. Ranganathan et al. (eds.), *Inventive Communication and Computational Technologies*, Lecture Notes in Networks and Systems 311,
https://doi.org/10.1007/978-981-16-5529-6_45

convolutive distortion that actuates big haul correlation between successive observations and can be very time-taking with a resonation time [2]. Noise and reverberation can be stationary or non-stationary and inconveniently affect both discourse quality and discourse comprehensibility [2]. Different techniques have been introduced on speech enhancement.

The Kalman filtering is one of them and is a good and dependable speech improvement algorithm. It utilizes the minimum mean square error wisely [3]. Nonetheless, admittance to clean speech and added substance commotion data for the state-space model boundaries for the greater part of the traditional KF-based speech enhancement techniques is needed. In particular, the linear prediction coefficients and the additive noise variance estimation, which is unrealistic in practical speaking to get the noisy speech [4, 5]. Also, the authors in [6] proposed that the fundamental cycle of noise reduction calculation is Kalman filtering. The underlying incentive for KF is dictated by ASS. To get higher exactness, the following calculation is proposed. From the outset, the power spectrum of clean speech is assessed from the spectrum by the KF algorithm. At that point, the acquired power spectrum is filled in for initial value, and Kalman filter calculation is rehashed. On doing this calculation, we acquired greater precision of decrease in noise. It can be repeated at 1.5–2.0 occasion times of constant by taking the noisy speech signal as an input, and fast Fourier transform (FFT) was done to get power spectrum. Using adaptive spectral subtraction (ASS), we get estimates of power spectrum, i.e., noise signal power subtracted is subtracted from mixed signal spectrum [7].

2 Related Works

As per the work done in [8], first a noisy speech signal is given as input, and this input speech signal is assumed as stationary during each frame and processed using three algorithms, which are spectral subtraction, Wiener filter and Kalman filters, and the work suggests that the spectral subtraction can be used only for stationary signals and real-time signals are non-stationary. The Wiener filter is also suitable for stationary signals but denies working on musical noise. To oversee these boundaries, the paper suggests Kalman filtering. When talking about the UKF algorithm, it was first proposed in [9, 10]. In [11], the work proposes that most approaches use the stationary AWGN assumption, but the same of colored noise is believed to be more useful for speech denoising and speech dereverberation. The Kalman filter, because of its flexibility, is widely used for signal enhancement. Kalman filter has a considerable amount of numerical complexity while dealing with colored noise. Moreover, Kalman filtering is a model-based adaptive method, where speech as well as noise is modeled as AR processes. Thus, a major issue in Kalman filtering is the estimation of the AR parameters in the presence of noise. The traditional algorithm utilizes the EM technique to repeatedly calculate the AR boundaries. Unfortunately, its computational complexity is high. The method used in our work is built on spectral subtraction for estimation of AR parameters of clean signal and corresponding noise

[12]. It is computationally efficient and can be easily implemented. The mathematical model for the algorithm of the state-space model and Kalman filter equations was formulated, and the obtained results were compared to the WF method [13, 14].

The work proposed by the authors in [15] is the computer-based algorithms which are generally used for controlling and monitoring a computer where human, digital and analog interactions occur. The cyber-physical systems (CPS) scheme is used in many areas due to its easily available and connectivity features and also offers large amount of storage and computing resources. However, the limitation of this scheme is its large energy consumption. As in [16], spectral subtraction method is applied in the estimation of parameters, musical noise appears in the enhanced speech. To acquire a Kalman filter output with better audible quality, a conceptual post-filter is set at the output of the Kalman filter to decrease the musical noise level. The perceptual filter minimizes signal distortion while constraining the noise spectrum.

3 Methodology

3.1 Flow Process

In the time domain, the distorted speech, $d_k(t)$, is given by $d_k\ (t) = C_k\ (t) * r_k\ (t) + n_k(k)$ where C$_k$ (t) is the clean speech component, r$_k(t)$ is the reverberant speech component, and $n_k\ (t)$ is the noise [2]. The time frame index is represented as k. The algorithm holds each time frame bit on its own. In the limits of the algorithm, k is introduced as a variable in the equations that involve multiple time frames [2]. Figure 1 explains the flow process of the algorithm.

The clean speech which is downloaded from the database is processed and is reverberated using the reverb parameters and convolution. The output of the first block in Fig. 1 is the reverberated speech with some given delay, and the magnitude of the speech changes according to the coefficient of reverberation taken [17]. The approach in Eq. 1 is used to do the reverberation process as

$$O(n) = I(n) + aO(n - d) \tag{1}$$

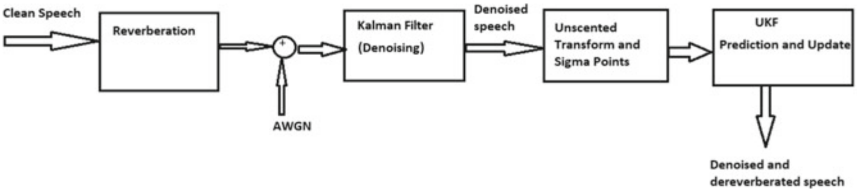

Fig. 1 Flow process

where $i(n)$ is the input audio signal, O (n) is the output (echoed) audio signal, d is the echo delay (in samples), and alpha is the coefficient that governs the amount of echo fed back. Then, the reverberated signal is then added with a certain amount of additive white Gaussian noise as shown in Fig. 1. Here, we have the corrupted speech signal that needs to be denoised and de-reverberated.

The corrupted speech is then taken as k reduced time frames or into k smaller time frames that are of a specific period which are called the state spaces. For this process of converting the clear speech signal to state spaces, we use three different windows. They are the rectangular window, the hamming window and the Gaussian window [12]. The proposed algorithm treats each time frame or the state space on its own. Firstly, as in the third block of Fig. 1, each of these frames then undergo the unscented transform in which the sigma points of the first state space are calculated. Then, the statistical mean and covariance of the present state are calculated. Then, the two main steps of the algorithm, the time update and the measurement update steps, are done for the first state space. Being an auto-regressive algorithm, the same is applied to all the k state spaces, i.e., the set of time update equations and measurement update equations given in the following Sect. 3.2 are implemented. The detailed equations to the above algorithm are also mentioned in the Sect. 3.2.

3.2 Unscented Kalman Filtering

3.2.1 Unscented Transform

The unscented transform (UT) is a method for estimating the mean and covariance of RV that goes through a nonlinear transformation [3, 18]. Take into consideration the propagation a RV $\boldsymbol{x}$ into a function $\mathbf{y} = f(x)$. Consider $\overline{x}$ is the mean, and P_X is the covariance of RV x.

Figure 2 explains the steps in the unscented transform step in Fig. 1. The $\overline{x}$ and P_x depicted in Fig. 2 are the mean the covariance of the random variable x, respectively,

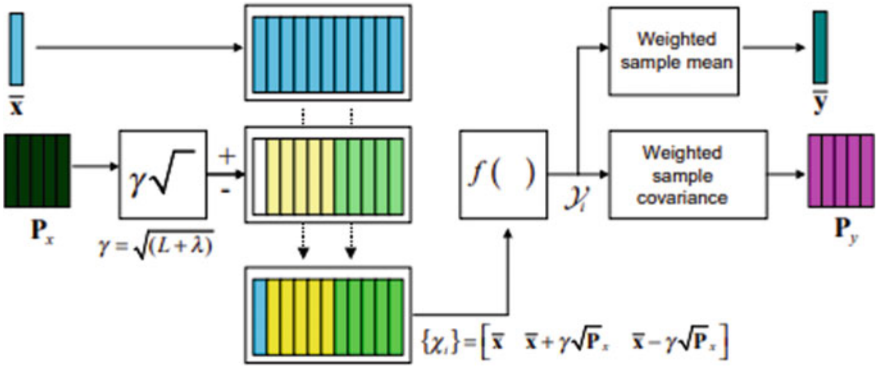
$\overline{\mathbf{x}}$
$\mathbf{P}_x$
$\gamma\sqrt{\ }$
$\gamma=\sqrt{(L+\lambda)}$
+
-
$f(\ )$
$\mathcal{Y}_i$
Weighted sample mean
Weighted sample covariance
$\overline{\mathbf{y}}$
$\mathbf{P}_y$
$\{\chi_i\}=[\overline{\mathbf{x}}\quad \overline{\mathbf{x}}+\gamma\sqrt{\mathbf{P}_x}\quad \overline{\mathbf{x}}-\gamma\sqrt{\mathbf{P}_x}]$

Fig. 2 Diagram of UT

then the sigma points are calculated which are then propagated through non-linear function. Then, the weighted sample mean and weighted sample covariance are calculated for further process [19].

To evaluate the mean and variance of ***y*****,** we initiate a matrix X_i of $2L + 1$ sigma vector X_i, relating to the following Eqs. 2–4 as shown in Fig. 2.

$$X_0 = \overline{x} \tag{2}$$

$$X_i = \overline{x} + \left(\sqrt{P_{\boldsymbol{X}}(L + \lambda}\right)_i, i = 1, \ldots, L \tag{3}$$

$$X_i = \overline{x} + \left(\sqrt{P_{\boldsymbol{X}}(L + \lambda}\right)_i, i = L + 1, \ldots, 2L \tag{4}$$

where $\lambda = \alpha^2(L + k) - L$. α is a coefficient that governs the sigma point spread around $\overline{x}$ and is generally set to a positive minor value (e.g., $1 \le \alpha \le 1e - 4$). k is a constant that is generally equal to 0 or 3-L and β is used for integration [20]. The initial information of the distribution of $\boldsymbol{x}$ (for Gaussian distribution $\beta = 2$ is ideal), $\left(\sqrt{P_X(L + \lambda}\right)_i$ is the ith column of the square root of the matrix. These sigma vectors undergo transition throughout as in Eq. 5,

$$y_i = f(X_i)i = 0, 1, 2, \ldots 2L \tag{5}$$

And using Eqs. 5–10, the weighted sample mean and covariance of the posterior sigma points are used to approximate the mean and covariance of $\boldsymbol{y}$ [21],

$$\overline{\boldsymbol{y}} \approx \sum_{i=0}^{2L} W_i^{(m)} y_i \tag{6}$$

$$P_y = \sum_{i=0}^{2L} W_i^{(c)} \{y_i - \overline{\boldsymbol{y}}\}\{y_i - \overline{\boldsymbol{y}}\}^T \tag{7}$$

With weights W_i are

$$W_0^{(m)} = \lambda/(\mathrm{L} + \lambda) \tag{8}$$

$$W_0^{(c)} = \lambda/(\mathrm{L} + \lambda) + \left(1 - \alpha^2 + \beta\right) \tag{9}$$

$$W_i^{(m)} = W_i^{(c)} = 1/\{2(\mathrm{L} + \lambda)\} \tag{10}$$

A diagram representing the steps in unscented transform is depicted in Fig. 1. Consider that, it varies considerably from the Monte-Carlo sampling methods that need more sample and orders of magnitude to propagate through a precise distribution of state [22, 23]. The illusionary simple way through with the UT leads to an approximation that are nearly equal to the third order of Gaussian inputs for all nonlinearities [14, 24]. For non-gaussian inputs, approximation is reduced precisely to 1st or 2nd order and the selection of α and β with the exactness of third order and other higher order moments are found.

3.2.2 Unscented Kalman Filter Equations

The UKF is a clear augmentation of the UT to the recurring assessment, when the state RV is reclassified due to the addition of the original state and noise variables: $x_k^a = \left[x_k^T V_k^T n_k^T\right]$. The UT sigma point choosing scheme (in Eq. 4) is put in to the new state random variable to determine the respective sigma matrix, X_k^a [2]. Then, the equations are initialized as shown in Eqs. 11–14. So, however, no conspicuous computation of Jacobians is important to execute this calculation. Moreover, the general number of calculations is a similar request as the EKF.

Initialize with

$$\hat{X}_0 = \mathbb{E}[X_0] \tag{11}$$

$$P_0 = \mathbb{E}\left[\left(X_0 - \hat{X}_0\right)\left(X_0 - \hat{X}_0\right)^T\right] \tag{12}$$

$$\hat{X}_0^a = \mathbb{E}\left[X^a\right] = \left[\,\hat{X}_0^T\ 0\ 0\,\right] \tag{13}$$

$$P_0^a = \mathbb{E}\left[\left(X_0^a - \hat{X}_0^a\right)\left(X_0^a - \hat{X}_0^a\right)^T\right] = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & R^v & 0 \\ 0 & 0 & R^n \end{bmatrix} \tag{14}$$

Calculation of sigma points:

$$X_{k-1}^a = \left[\hat{X}_{k-1}^a \hat{X}_{k-1}^a + \gamma\sqrt{P_{k-1}^a}\hat{X}_{k-1}^a - \gamma\sqrt{P_{k-1}^a}\right] \tag{15}$$

The time update equations are given from Eq. 16–20:

$$X_{k|k-1}^x = F\left[X_{k-1}^x, u_{k-1}, X_{k-1}^v\right] \tag{16}$$

$$\hat{X}_k^- = \sum_{i=0}^{2L} W_i^{(m)} X_{i,k|k-1}^x \tag{17}$$

$$P_k^- = \sum_{i=0}^{2L} W_i^{(c)} \Big[X^-_{i,k|k-1} - \hat{X}_k^- \Big] \Big[X^-_{i,k|k-1} - \hat{X}_k^- \Big]^T \tag{18}$$

$$y_{k|k-1} = H[X^x_{k|k-1}, X^n_{k-1}] \tag{19}$$

$$\hat{y}_k^- = \sum_{i=0}^{2L} W_i^{(m)} y_{i,k|k-1} \tag{20}$$

The measurement update equation is from Eqs. 21–25:

$$P_{\hat{y}_k \bar{y}_k} = \sum_{i=0}^{2L} W_i^{(c)} [y_{i,k|k-1} - \hat{y}_k^-][y_{i,k|k-1} - \hat{y}_k^-]^T \tag{21}$$

$$P_{x_k y_k} = \sum_{i=0}^{2L} W_i^{(c)} [x_{i,k|k-1} - \hat{x}_k^-][y_{i,k|k-1} - \hat{y}_k^-]^T \tag{22}$$

$$K_k = P_{x_k y_k} P^{-1}_{\hat{y}_k \bar{y}_k} \tag{23}$$

$$\hat{x}_k = \hat{x}_k^- + K_k \big(y_k - \hat{y}_k^- \big) \tag{24}$$

$$P_k = P_k^- - K_k P_{\hat{y}_k \bar{y}_k} K_k^T \tag{25}$$

where $x^a = \left[x^T v^T n^T \right]$, $X^a = \left[(X^x)^T (X^v)^T (X^n)^T \right]^T$, $\gamma = \sqrt{(\mathrm{L} + \lambda)}$, where $R^v is$ the process noise variance, R^v is the measurement noise covariance, and W_i are the weights that are calculated in Eq. 4. The measurement is then updated in each time frame of the speech taken [2]. Then, all the time frames are then augmented to get back the denoised and dereverberated clean processed speech.

4 Experimental Results

In this section, the simulation results we obtained from the approach detailed in the above section, i.e., the UKF algorithm are discussed. There were few .wav files on which we performed the algorithm under various windowed processing like the rectangular window, hamming window and the Gaussian window. The results we obtained are plotted as wave forms. There are two wave files on which this algorithm was performed. Let the names be Speech A.wav and Speech B.wav. The SNR was precalculated for later use in the comparisons. The waves were then reverberated, and the observation noise was added to both. The observation noise added to all the wave forms is additive white Gaussian noise (AWGN).

After processing the two waveforms through the algorithm and getting the results, we calculated the parameters such as the figure of merit and the correlation between SNR of the processed output and the precalculated SNR of the clean speech for three different number of iterations in the algorithm. A table is given below with the particular details of the figure of merit and correlation for the above two wave forms.

Table1 shows analysis of the performance metrics FOM and correlation between the input and the output of the algorithm proposed.

Table 2 shows the comparison between the SNR values for the different windows—rectangular, hamming and the Gaussian windows used for chopping and the number of iterations performed on both speech A and speech B.

5 Conclusion

In this project, speech enhancement technique using Kalman filtering has been implemented. The objective was to design an effective method to process a noise invaded and reverberated speech in adverse environments. We were able to perform the denoising and dereverberation on the corrupted speech. The proposed algorithm can be used in the cases of nonlinear systems, where in most of the algorithms, this is not possible. Also, this algorithm is time-efficient. So, it can be used for mediocre length speeches. Here, the proposed algorithm, unscented Kalman filtering, uses three windows—rectangular, hamming and Gaussian for the chopping of the signal before processing, and from Table 1, the results significantly differ from each window for every iteration. The performance is slightly increasing with the increasing number of iterations in any window up to a certain number of iterations. Then, there is fall in both the performance metrics—figure of merit and the correlation taken in this report. This is due to the repeated denoising and dereverberation, which causes a damage to the intelligibility of the desired output. Then, Table 2 compares the SNRs of the outputs of different windows under different number of iterations.

Table 1 Performance metrics

Number of iterations	Parameter	Feared_or_respected.wav			Movie-05.wav		
	window	Rectangular	Hamming	Gaussian	Rectangular	Hamming	Gaussian
15	Figure of merit	0.824	0.837	0.831	0.506	0.398	0.5564
	Correlation	0.457	0.443	0.448	0.461	0.432	0.4560
30	Figure of merit	0.822	0.824	0.827	0.498	0.401	0.7104
	Correlation	0.480	0.463	0.472	0.478	0.461	0.4703
50	Figure of merit	0.833	0.830	0.844	0.432	0.421	0.5503
	Correlation	0.475	0.459	0.472	0.468	0.460	0.4561

Table2 SNR comparison

		Speech A.wav			Speech B.wav		
Window	Iterations	15	30	50	15	30	50
Rectangular		18.15	18.23	17.8	7.4	7.75	7.6
Hamming		18.44	18.56	18.2	5.82	5.83	5.78
Gaussian		18.31	18.42	18.1	8.13	8.23	8.09

References

1. M. Mosallaei, Performance evaluation of instrumentation sensor network design using a data reconciliation technique based on the unscented Kalman filter, in *2007 IEEE Conference on Emerging Technologies & Factory Automation (EFTA 2007),* 09/2007.
2. N. Dionelis, M. Brookes, Modulation-Domain Kalman filtering for monaural blind speech denoising and dereverberation. IEEE/ACM Trans Audio, Speech, Lang Process **27**(4), 799–814 (April 2019). https://doi.org/10.1109/TASLP.2019.2894909.
3. A.UmaMageswari, J. Joseph Ignatious ,R. Vinodha, A comparitive study of Kalman Filter, Extended Kalman Filter And Unscented Kalman Filter For Harmonic Analysis of the non-stationary signals International Journal of Scientific & Engineering Research (2012)
4. M. Fujimoto, Y. Ariki, Noisy speech recognition using noise reduction method based on Kalman filter, in *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference,* vol. 3. Pp. 1727–1730. https://doi.org/10.1109/ICASSP.2000.862085
5. M.G. Muthukrishnan, P. Sudheesh, M. Jayakumar, Channel estimation for a high mobility MIMO system using Particle filter, in *İnternational Conference on Recent Trends in Information Technology* (2016), pp.197–207
6. F. Asano, S. Hayamizu, T. Yamada, S. Nakamura, Speech enhancement based on the subspace method. IEEE trans. Speech Audio Proc. **8**(5), 97–507 (2000)
7. M.A.A. El-Fattah, M.I. Dessouky et al., Speech enhancement with an adaptive Wiener filter. Int. J. Speech Technol. **17**, 53–64 (2014)
8. S.F.Boll, Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoustics, Speech Sig. Proc. **ASSP-27**(2), 113–120,1979
9. B. Cornelis, M. Moonen, J. Wouters, Performance analysis of multichannel wiener filter-based noise reduction in hearing aids under second order statistics estimation errors. IEEE Trans. Audio Speech Lang. Process. **19**(5), 1368–1381 (2011)
10. R.E. Kalman, A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng. **82**(D), 35–45 (1960)
11. B.L. Sim, Y.C. Tong, J. Chang, C.T. Tan, A parametric formulation of the generalized spectral subtraction method. IEEE Trans. Speech Audio Proc. **6**(4), 328–337 (1998)
12. S.J. Julier, J.K. Uhlmann, H. A New Extension of Kalman filter to Nonlinear systems, *ın proc ofAreoSense*: *The 11th int. symp. on Areospace/Defence sensing. Simulaton and controls* (1997)
13. K.K. Paliwal, A. Basu, A speech enhancement method based on kalman filtering, ın *Proc. ICASSP,* vol 12 (1987)
14. S. Braun, E.A.P. Habets, Linear prediction based online dereverberation and noise reduction using alternating Kalman filters, in *IEEE/ACMTrans. Audio, Speech, Lang. Process.*, vol. 26, no. 6 (2018), pp. 1119–1129
15. S. Haoxiang Wang, Smys, secure and optimized cloud-based cyber-physical systems with memory-aware scheduling scheme. J Trends Comput. Sci. Smart Technol. (TCSST) **2**(03), 141–147 (2020)
16. V.R. Balaji, Maheswaran S, M. Rajesh Babu, M. Kowsigan, Prabhu E., Venkatachalam K, *Combining statistical models using modified spectral subtraction method for embedded system Microprocessors and Microsystems*, vol 73(2020),102957, ISSN 0141-9331

17. J. Wei, L. Du, Z. Yan, H, Zeng, *Improved Kalman Filter-Based Speech Enhancement* (2003)
18. E.A. Wan, R. Van Der Merwe, The unscented Kalman filter for nonlinear estimation, in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)* (2000)
19. A. Suraj, S, A, G., S, S. Chakravarthy, R. Ramnathan, Routing in wireless sensor network based on swarm ıntelligence, in *3rd International Conference on Trends in Electronics and Informatics* (2019)*,* pp. 215–217
20. T. S. Kavya, T. Peng, Y.M. Jang, E. Tsogtbaatar, S.B. Cho, Face Tracking Using Unscented Kalman Filter, in *2020 International Conference on Electronics, Information, and Communication (ICEIC)* (2020)
21. R. Van Der Merwe, The unscented Kalman filter for nonlinear estimation, in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing Communications and Control Symposium (Cat No 00EX373) ASSPCC-00* (2000)
22. B.S. Atal, Speech analysis and synthesis by linear prediction of the speech wave. J. Acoust. Soc. Am. **47**(1), 65 (1970)
23. J. Ramnarayan, J.P. Anita, P. Sudheesh, Estimation and Tracking of a Ballistic Target Using Sequential Importance Sampling Method. Commun. Comput. Inf. Sci. **746**, 387–398 (2017)
24. R.G. Reddy, R. Ramnathan, An Empirical study on MAC layer in IEEE 802.11p/WAVE based Vehicular ad hoc Networks. Procedia Comput Sci **143**, 720–727 (2018)