Dual Particle-Number RBPF for Speech Enhancement

Seyed Farid Mousavipour, Saeed Seyedtabaii
Engineering Faculty, Shahed university
P.O. Box: 18155 159
Tehran, Iran
{Mousavipour,stabaii}@shahed.ac.ir

ABSTRACT: *In this paper, performance of a dual particle-number Rao-Blackwellized particle filter (RBPF) for speech enhancement is evaluated. Speech is corrupted by additive white, colored and real industrial noises that degrade its intelligibility. The performance indexes are: Quality of the processed speech scored by PESQ and computation time. The simulation results indicate that the RBPF methods outperform some well known Kalman based algorithms in the cost of more computation time. The weakness, then, is overcome by the dual particle-number RBPF that saves the quality of the processed speech while reduces remarkably the computation time.*

Keywords: Speech enhancement, Kalman filter, Particle filter, PESQ

Received: 18 July 2011, Revised 25 August 2011, Accepted 28 August 2011

1. Introduction

Processing of speech signal that has been degraded by additive background noise is of a great interest in variety of contexts. Speech enhancement aims at estimating clean speech, given noisy signal. Enhancement techniques are classified into single and multi channel categories. Single channel techniques are the most common real-time methods, since the second channel is not available in most of the applications, e.g. internet communication, speech recognition systems, and the case of speech-passing noise-cancelling headset.

Implementations of single channel systems are easy and less expensive than the multiple channel systems. However, because of the unavailability of the noise statistics, their algorithms are complicated and suffer from high computation cost and complexity.

Single channel speech enhancement algorithms can be roughly divided into three groups: spectral subtraction, sub-space analysis and filtering algorithms. The spectral subtraction approaches are straightforward and easy to implement [1]. However, they produce an audible distortion known as “ringing”. Sub-space analysis operates in the autocorrelation domain, where the speech and noise components can be assumed to be orthogonal, whereby their contributions can be readily separated. Unfortunately, estimating the orthogonal components consumes high computation. Moreover, the orthogonality assumption is difficult to motivate.

Filtering algorithms may be implemented in time-domain, frequency domain or even jointed domains. Their main attempts are removing the noise component (Wiener) or estimate the noise and speech components (Kalman, Particle). The original Kalman filter provides a minimum mean-squared error (MMSE) estimate of the clean speech if the assumed noise is Gaussian [2]. There have been numerous studies involving the enhancement of white noise corrupted speech [2, 3], however, in real world, colored noise assumption has been proved to be very effective for speech enhancement [4, 5]. In addition, for enhancing the performance, filtering methods have been equipped with various Expectation Maximization (EM) parameter estimation algorithms. Iterative Kalman [5], Kalman-EM-iterative (KEMI) and Kalman-gradient descent-sequential (KGDS) algorithm [6] are among them.

In case of non Gaussian noise or nonlinear model assumption, the particle filters may be implemented for speech enhancement. PF requires very few assumptions about the noise Power Density Function (PDF) compared to what Kalman filter family requires.

Filtering type approaches generally require AR model parameters of the degraded speech signal. AR model exploits the local correlations in a time series by forming the prediction of the current sample as a linear combination of the immediately preceding samples. The readily available choice may be to draw the noise statistics from silent periods and speech model parameters from noisy speech periods. One of the EM parameter estimation algorithms accompanied with RBPF has been described in [8]. In the algorithm presented in [7], speech and noise parameters are estimated by particle method and the state of the presumed linear Gaussian model is estimated by Kalman approach.

In this paper, dual particle-number RBPF is proposed that impressively reduces the computation cost and complexity. The results are assessed in comparison with the outcome of several well known Kalman and particle filter families of algorithms for speech enhancement such as: iterative Kalman filter, sequential and iterative Kalman filters, RBPF and constraint RBPF. White, colored and real industrial noise (drill) corrupted speech samples are adopted for assessment of the algorithm. Two indexes are considered: quality of the enhanced speech and computation time. The quality of the enhanced speech is exhibited by the PESQ score. The results show that in case of white and colored computer generated noise, the proposed method performance is similar to the original RBPF while the computation time has been dropped significantly. This is due to the provision embedded into the algorithm that assigns different particle number to the silent and the speech segments of the signal.

This paper is organized as follows. Speech and noise models are presented in section 2. Two variations of Kalman filter developed especially for speech enhancement are described in section 3. In section 4, PBRFs are detailed and in section 5 the proposed dual particle-number RBPF is elaborated. Section 6 evaluates the algorithm performance and finally conclusion comes in section 7.

2. Speech and Noise model

The autoregressive (AR) model is popular for audio signals. This model exploits the local correlation in a time series by forming the prediction of the current sample as a linear combination of the immediately preceding samples. The speech AR model is as follows:

$$S(n) = \sum_{k=1}^{p} a_k s(n-k) + w(n) \tag{1}$$

where s(n) is the clean speech signal, α is the AR parameters, p is the model order and $w(n)$ is a zero mean, white Gaussian excitation noise with variance σ_w. We may incorporate the more detailed voiced speech model in which the excitation process is composed of a weighted linear combination of an impulse train and a white noise sequence to represent voiced and unvoiced speech, respectively. However, this approach did not yield any significant performance improvements over the standard LPC modeling [6].

The source sequence is then contaminated by zero mean additive Gaussian noise $v(n)$, which is either white or colored but independent of $w(n)$.

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

The canonical state-space model of $s(n)$ and $y(n)$ are as follows,

$$x_s(n) = [s(n-p+1)\, s(n-p+2) ... s(n)]^T$$

$$x_s(n) = \begin{bmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_p & a_{p-1} & \cdots & a_1 \end{bmatrix} x_s(n-1) + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} w(n)$$

$$s(n) = [\,0 \;\dots\; 0 \;\; 1\,]\, x_s(n) \tag{3}$$

It can be written as:

$$x_s(n) = A_s\, x_s(n-1) + G_s\, w(n)$$

$$s(n) = C_s\, x_s(n)$$

$$y(n) = C_s\, x_s(n) + v(n) \tag{4}$$

$$v(n) = \sum_{\iota=1}^{m} b_\iota\; n(n-l) + u(n) \tag{5}$$

where m is the noise AR order and $u(n)$ is a zero mean Gaussian noise not correlated with $w(n)$. It is assumed that the noise is wide sense stationary and is adequately described by the AR (q) model.

$$X_n(n) = A_n\, x_n(n-1) + G_n\, u(n)$$

$$v(n) = C_n\, x_n(n) \tag{6}$$

By adjoining the states in (3) and (5) the augmented system has the form,

$$x_t(n) = A_t\, x_t(n-1) + G_t\, v(n)$$

$$y(n) = C_t\, x_t(n) \tag{7}$$

where $x_t(n) = [x_s(n)\; x_n(n)]'$ and the augmented matrices are:

$$A_t = \begin{bmatrix} A_s & 0 \\ 0 & A_n \end{bmatrix},\; G_t \begin{bmatrix} G_s & 0 \\ 0 & G_n \end{bmatrix}$$

$$C_t = [C_s \quad C_n],\quad Q_t = \begin{bmatrix} \delta_w^2 & 0 \\ 0 & \delta_u^2 \end{bmatrix} \tag{8}$$

3. Kalman Filter Type

Kalman filtering belongs to the group of parametric methods of filtering. It is classified into two main categories: time domain and frequency domain. Most studies focused on the time domain Kalman filter. In the time domain Kalman filtering of speech signal, signal is segmented into 20-40 ms frames. Then, AR coefficients are extracted for each frame. In a single channel system, estimation of speech and noise model parameters is drawn from noisy speech frames. Kalman filter requires measurement noise variance that has to be *a priori* known.

3.1 Iterative Kalman Filter in Speech Enhancement

In [5], the main method of iterative Kalman filter referred to as scalar Kalman filter is described. Noise is wide sense stationary and it is assumed to be adequately described by the AR(q) model (5). Augmented state space is obtained as (6). This is the so-called, noise-free measurements problem in the estimation literature. In case that (6) is linear and noise PDF is Gaussian, the optimal estimate of the states is given by the Kalman solution. The Kalman state and time update equations are as follows:

$$
\begin{aligned}
&x_t(n|n-1) = A_t\ x_t(n-1) \\
&x_t(n) = x_t(n|n-1) - K(n)[y(n) - C_t^T x_t(n|n-1)] \\
&P(n|n-1) = A_t P(n-1) A_t^T + G_t\ Q_t\ G_t^T \\
&K(n) = P(n|n-1)\ C_t^T [C_t\ P(n|n-1)\ C_t^T]^{-1} \\
&P(n) = [I - K(n)\ C_t]\ P(n|n-1)
\end{aligned} \tag{9}
$$

where *K(n)* is a Kalman gain vector, *P(n|n-1)* is *a priori* error covariance matrix and *P(n)* is an error covariance matrix. As equations show, knowledge of the noises' statistics is required for acceptable performance of the algorithm. It must be noted here that Kalman filter offers an optimal estimate when the system parameters are known, so that, it is important that the system matrices A_t, G_t and C_t and especially noise intensity, Q_t to be modeled as accurate as possible.

Iterative Kalman filter algorithm is initialized by segmenting the speech into (20-40 ms) frames, then, observation noise is estimated from the energy of silence frames within noisy speech signal. There are also other methods that may be used for the estimation of measurement noise required by Kalman approaches [15].

After first round of filtering, the enhanced speech obtained enters the second round of signal clean up. In each iteration observation and excitation noises STD's must be calculated. Output of the approach shows better SEG-SNR than what a single round filtering yields but the quality is unnatural. For improving naturalness, at the end of each iteration, a low amplitude white noise is added to the enhanced speech signal. This improves the intelligibility. 3 or 2 iterations are shown to be adequate for attaining optimum quality.

In [4], a similar algorithm is proposed where the extended Kalman filter algorithm is used for the estimation of the noise and clean speech statistics. The noise may also be non-stationary. In each frame, transition matrix (A_t) is calculated sample by sample as follows:

$$\hat{A}_t(n) = A_t(n - 1) - K(n)\ C_t \tag{10}$$

However, simulation results show no more improvement with respect to the previous method.

3.2 Iterative and Sequential Kalman Filter in Speech Enhancement

This algorithm has been proposed in [6] and is similar to [5]. In each frame augmented Kalman filtering is applied for the removal of noise from the degraded speech signal. Then, an EM method is followed to calculate more accurate parameters. Let γ to be the vector of the all unknown parameters in the extended model,

$$
\begin{aligned}
&\boldsymbol{\gamma} = [\alpha^T \quad g_s \quad \beta^T \quad g_v] \\
&\boldsymbol{\alpha} = [\alpha_p \quad \alpha_{p\text{-}1} \ \dots \ \alpha_1] \\
&\boldsymbol{\beta} = [\beta_p \quad \beta_{p\text{-}1} \ \dots \ \beta_1]
\end{aligned} \tag{11}
$$

The outcome of the EM is a new estimate, $\hat{\gamma}$ for all of the parameters involved. The EM procedure has two main steps: state estimation and parameter estimation.

State Estimation

State estimation is conducted as follows,

$$
\begin{aligned}
&\boldsymbol{\mu}(t|N) = \overline{\boldsymbol{x}(\mathrm{t})} \\
&\boldsymbol{p}(t|N) = \overline{\boldsymbol{x}(\mathrm{t})\,\boldsymbol{x}^T(\mathrm{t})} - \overline{\boldsymbol{x}(\mathrm{t})\,\boldsymbol{x}(\mathrm{t})}^T \\
&\boldsymbol{y} = [y(1)\ y(2) \dots y(N)]^T
\end{aligned} \tag{12}
$$

where $\boldsymbol{y}$ is the vector of the measured data in the current frame and $\mu(t|N)$ is the current state estimate based on $y(t)$, $t=1,\dots,N$. State estimation procedure is the same as in Kalman filter as it is described by (9). Parameter estimation based on the EM method is guaranteed to converge to the ML estimate of all unknown parameters. By each iteration, the likelihood of the estimate of the parameters is increased.

Parameter Estimation

In the next step, 2 main parameters are estimated: AR coefficients of the clean speech and the observation noise variance as follows,

$$\begin{aligned} \boldsymbol{S}(t-1) &= \boldsymbol{P}(t-1|t-1)\,\boldsymbol{F}^T(\boldsymbol{P}(t|t-1)^T \\ \hat{\alpha}^{(l-1)} &= -[\textstyle\sum_{t=1}^{N} \boldsymbol{s}_p(t-\overline{1)\,\boldsymbol{s}_p^T}\underline{(t}-1)]^{-1}\textstyle\sum_{t=1}^{N} \boldsymbol{s}_p(t-\overline{1)\,s}(t) \\ g_s^{(l-1)} &= \frac{1}{N}\textstyle\sum_{t=1}^{N}[\overline{s^2(t)} + (\hat{\alpha}^{(l+1)})^T \boldsymbol{s}_p(\overline{t-1})\,s(t)] \end{aligned} \tag{13}$$

Where $\boldsymbol{s}_p(t-\overline{1)\,\boldsymbol{s}_p}^T(t-1)$ is the upper left p x p sub-matrix $\boldsymbol{x}(t)\,\overline{\boldsymbol{x}^T}(t)$. Other such as $s^2\overline{(t)}$ is similarly extracted from $\boldsymbol{x}(t)\,\overline{\boldsymbol{x}^T}(t)$. In [6], AR coefficients of the colored noise is also obtained by similar equations. Since in this method the signal and noise parameter estimates are computed separately, the increase in computational complexity is unavoidable but is quite moderate. The iterative-batch EM algorithm requires the use of an analysis window over which the signal and noise statistics are to be wide sense stationary.

4. Rao-Blackwellized Particle Filter Type

If the linearity and Gaussian assumptions are not satisfied, alternative state estimation algorithm must be sought. Particle Filters (PF) constitute a family of solutions to the following general state estimation problem:

$$\begin{aligned} x_k &= f(x_{k-1}, w_k) \\ z_k &= g(x_k, v_k) \end{aligned} \tag{14}$$

Where x_k represents the state vector at instant k, z_k is the vector of observations, w_k and v_k are the process and the observation noises, f is a signal transition function and g is a measurement function (both assumed to be known). The goal is to estimate the state based on all available observations data up to time k. The PFs introduce an approximate recursive solution for very weak assumptions: f and g may be non-linear, v and w may be non-Gaussian, at the cost of a more computationally expensive implementation [8]. Computation cost is depended on the AR order and particle number and in some case lag smooth order. The AR coefficients order increases the matrix dimension. Each particle requires a KF procedure, therefore, If Particle number is increased, more computational is imposed. If the particle number exceeds its optimum quantity, it may cost divergence of the algorithm.

The sequential estimation method is based on Mounte-Carlo simulation, which can operate on the broadest range of state-space formulated problems [7, 9]. A general PF algorithm for speech enhancement has been presented in Table 1.

4.1 Low Cost RBPF Algorithm for Speech Enhancement

Some conditional dependencies between elements of the state vector can be analytically explicated and then there is no need to draw samples from the entire state space, which leads to RBPF. RBPFs are in practice mostly used when part of the state satisfies the linear- Gaussian condition. RBPF algorithm for speech enhancement is detailed in Table 2.

The RBPF procedure reduces the variance of the error estimates [16], while, the dimension of the part of the state on which the PF is running is smaller. Its consequences are surely computation efficiency with respect to the regular PF [5].

- *For every k* (sample), *do the following* :
- *For every $i \in$* {1, 2, ..., N} (*Particle*)

•*Draw*

$$X_{k,i} \sim q(X_k \mid X_{k-1}, Z_k)$$

•*And set*

$$X_{k,i} = \{X_{k,i} : X_{k-1,i}\}$$

• *compute the unnormalized weights*

$$(\hat{W}_{ki}) = W_{k-1,i} \frac{p(Z_k \mid X_{k-1}, Z_k - 1)\, p(X_{k,i} \mid X_{k-1,i}, Z_k - 1)}{q(X_{k,i} \mid X_{k-1,i}, Z_k}$$

•*Compute the normalized factor* $\sum_i \tilde{\omega}_{k,i}$

•*Obtain the normalized weights* $\omega_{k,i}$

•*resample particles*

Table 1. The general speech enhancement PF algorithm

- *For every k (sample), do the following:*
- *Run a PF on the sub-state* x_k
- *For every i (particle)*
- *update* ($x_k | X_{k-1,i}, Z_k$)

Table 2. RBPF speech enhancement algorithm

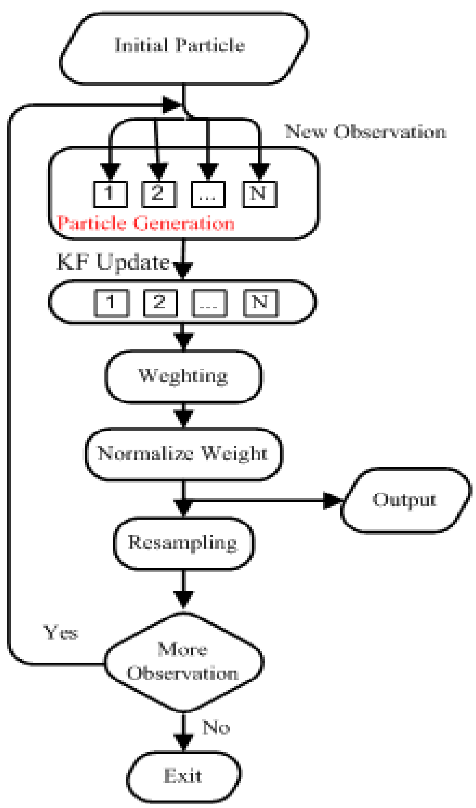


Figure 1. Main steps of the RBPF procedure

Figure 1 shows the RBPF procedure in details. In the initialization step, RBPF assigns random numbers to some variables such as clean speech AR coefficients, observation noise STD, excitation noise STD. The observation noise variance can be determined in a plenty of ways, however, here it is estimated once, in the initialization step. To each particle, it is assigned AR coefficients that are produced by a Gaussian random number generator. Each random number is multiplied by random walk coefficient to reduce the risk of instability. Inappropriate AR coefficients may cause algorithms instability, therefore, if the AR coefficients of a particle is unstable, stability check discards the coefficient and generates new AR ones.

In the beginning of any speech signal, some frames can certainly be assumed to be silent frames. By estimating the additive noise signal statistics in this step, improving the performance of the speech enhancement method is substantially mounted. In some methods such as KEM, the noisy speech AR coefficients are used, this impairs the estimation of the observation noise variance and clean speech characteristics.

The KF update step consists of the KF simple procedure that is applied to each particle. After updating the speech parameters such as transition matrix, *A* and observation noise STD, the important weighting procedure is followed that evaluates the likelihood criteria. The criteria depends on the observation noise STD and Kalman filter gain *K(n)*,

$$w(n) = \sqrt{\frac{1}{sp}} * e^{(-0.5*ep/Sp)} \qquad (15)$$

where *sp* represents the observation noise variance and *ep*(error) is the difference between the observed and the estimated signal, $x(n) - \hat{x}(n)$. Then the normalization is applied on each particle weight.

In the resampling step, particles with high weight score are kept and replicated and low weight particles are discarded.

Finally, the estimated results are stored. In case that a lag smooth scheme has been implemented, it is imposed on the results. The lag smooth scheme is useful for the reduction of unnatural variation in the processed speech.

4.2 A Constrained Sequential EM RBPF Algorithm

In [8], a Constrained Sequential EM (csEM) RBPF algorithm has been proposed to improve the accuracy of the speech enhancement algorithm. In the KEM method, an EM is applied to estimate the clean speech and colored noise parameters where there is not any constraint over the problem model. The csEM, on the other hand, estimates the model parameters recursively with constraints for some of the parameters. RBPF+csEM employs GAR model where the innovation sequence follows the generalized exponential distribution, which reflects the non-Gaussian characteristics. GAR model is a non-Gaussian extension of the AR model, where the same linear model (1) is used but the innovation vt is assumed to be drawn from the generalized exponential distribution with zero mean.

$$p(v; R, \beta) = \frac{R\beta^{1/R}}{2\Gamma(1/R)} e^{-\beta|v|^R} \qquad (16)$$

where Γ(.) is the gamma function, 1/β specifies width of the density, and R is the shaping parameter of distribution.

RBPF+csEM is similar to [7] with some changes. In this algorithm: 1) the speech or noise can be non-Gaussian (GAR model), 2) a Constrained sequential EM step added after KF update to improve the accuracy of the estimate of the parameters, 3) Augmented state applied is expected to remove colored noise perfectly, 4) the EM algorithm is an iterative method which finds local maxima of the log-likelihood function, 5) the E-step involves calculating the expected log-likelihood and M-step updates parameters that maximize the expected complete data log-likelihood.

The updating rules used in EM procedure are referred to as sequential Newton-Raphson EM. The outline of the sequential speech enhancement method is summarized in Table 3.

Initialization step:

At t =1,

- *Draw particles for clean speech from noisy speech.*
- *Set the initial model parameters for clean speech and noise.*
- *Initialize EM step.*

For t = 2,...

E step

- *Particle generation.*
- *Weight update and normalize.*
- *Resampling according to importance weights*
- *KF update.Calculation of expected score.*
- *Estimation of clean speech.*

M step

- *Parameter updating for clean speech and noise.*

Table 3. A constrained sequential EM RBPF speech enhancement algorithm

5. Dual Particle-Number RBPF in Speech Enhancement

To improve the performance and computation cost of the RBPF algorithms the following points should be especially regarded. In any speech signal, there are silent and spoken segments independent of language and accent. In standard speech database such as ITU dataset, 40% length of speech is silent. Silent frame model needs lower particles for accurate parameter estimation in case of RBPF. By assigning lower particles to the silent frames and higher particle number to the spoken frames, saving in the computation cost and consumed time is obtained. It also improves the convergence of the algorithm since discards the unemployed states and does not introduce more parameters than what is actually needed.

However, when noise is added to speech signal, silent portion resembles so much to the spoken section. Spoken frames can be discriminated by checking frame power and zero crossing. There are several indexes that may be used in this respect as follows: 1) Spoken frames have high power and low zero crossing against low power-high zero crossing silent frames, 2) Sudden changes in signal energy indicates the beginning or end to an impulsive noise, 3) The rate of change of the energy of the speech signal is limited by the inertia of human speech production system. It is widely accepted that the speech signal remains stationary within 5 to 10 ms segments, thus any quicker change in the speech signal can be attributed to noise, 4) Noise signal tends to be dominated by high frequency components and is much less autocorrelated than the speech signal.

Voice activity detectors employ such specifications to isolate voiced sections. Here, zeros-crossing and short time energy of signal is used for silent-spoken frames isolation.

After specifying silent-spoken frames, different particle number is assigned to each type of frames. If a spoken frame is processed and the next frame is silent, extra particles that are useless equalized to zero. For example, if for a spoken frame 1200 particles have been allocated and the number of particles for silent frames is 250, in this case 950 particles are nullified. This leads to decrease in computation cost and temporary memory.

6. Simulation and Results

For simulation purposes, ten noisy speech signals consist of 5 male and 5 female pieces of speeches are tested. The clean speech is from the NOIZEUS database sampled at 8 kHz. Noise types are white, computer generated colored noise and real industrial drill noise. The frame size is 200 samples, i.e. 25 ms frames and AR order *p*, is set to 12 for spoken frames.

Assessment of the quality of the processed speech signal is conducted by PESQ measure [13]. PESQ score is in a close match to the subjective tests score. Its minimum score is 0.5 and its highest point is 4.5, expressing the highest quality. In general case PESQ is sensitive to signal distortion and additive noise.

Algorithms that have been tried are:

1. IKF where the number of iterations is set to 2.
2. KEM
3. RBPF: Mustiere algorithm
4. csEM RBPF: Park algorithm

where the estimation of the observation noise variance is derived from silent frames.

The results have been depicted in Figure 2. In Figure 2(a), white Gaussian noise has been added to the signal for a wide range of SNR's from -5 to 5 db. The results indicate that the Mustiere algorithm supersedes the others. In Figure 2(b), the case of colored noise is investigated. Again under various SNR's speech enhancement are tested. What the results exhibit is that again the Mustiere algorithm is at the top while IKF is relatively close behind. In the third test that the noise is a real drill device noise, RBPF of Mustiere leads the others in PESQ score following by the PARK algorithm.

What is worth mentioning is that the RBPF based methods can perform well in speech enhancement, however, their computation cost is higher than their competitors. This is relieved by adopting dual particle-number strategy tested in this research. By incorporating the idea, the computation cost drops significantly as Table 4 vividly illustrates. The computation elapsed times are for a 25msec frame.

mode	AR order	Particle Number	Time(sec)	PESQ
Normal	12	1200	110	1.65
1	6	600	14	1.62
2	3	200	0.7	1.60

Table 4.The proposed RBPF performance merits

There is a plenty of room for improving the performance of the filtering types of speech enhancement as it has been depicted in Figure 3. Top belongs to the outcome of Kalman filtering when the noise and speech model parameters and statistics are assumed known. Reaching to the top needs further investigation in finding more accurate estimation of the model parameters from noisy speech signal.

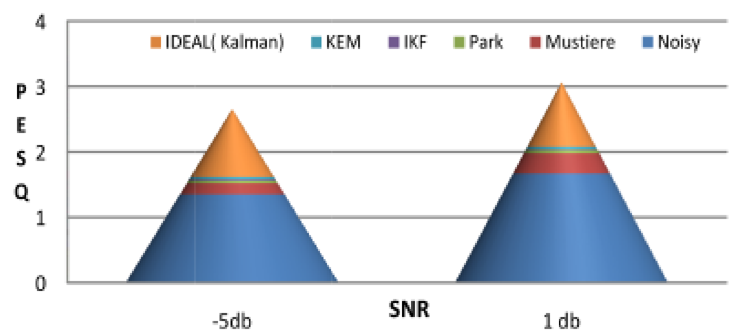
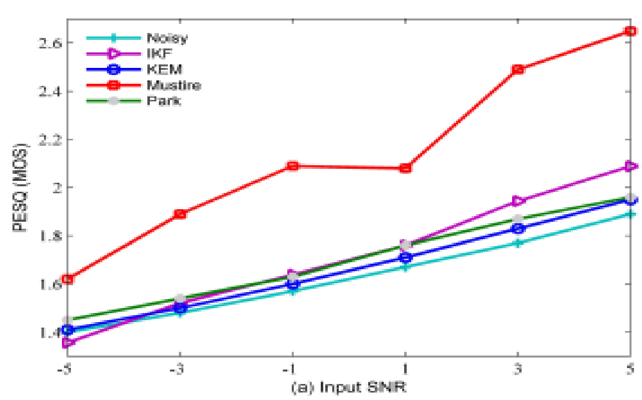
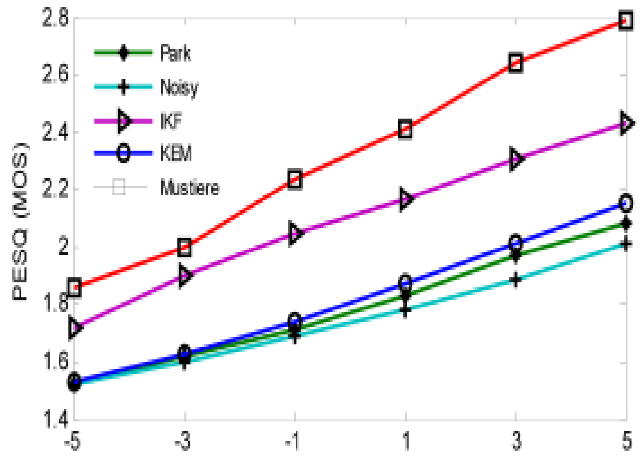


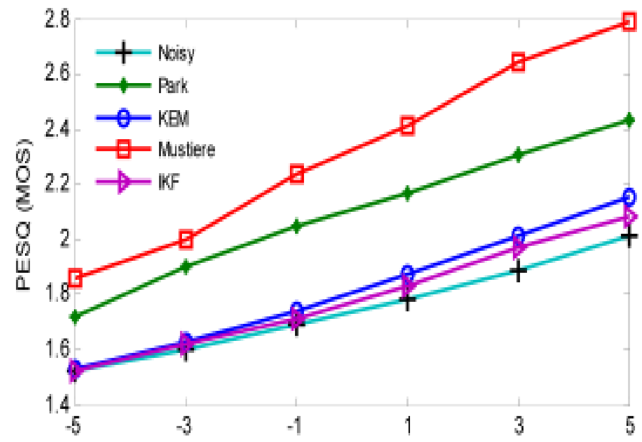
Figure 3. Top section exhibits the room for speech enhancement improvement; the ideal case is when the noise and speech model are exact



(a) Input SNR



(b) Input SNR



(c) Input SNR

Figure 2. The PESQ scores of various speech enhancement algorithms in case of (a) white noise, (b) colored noise, (c) drill noise

7. Conclusion

In this paper, the performance of several filtering algorithms for speech enhancement is evaluated. The algorithms either use 1) an approximate iterative and sequential EM or 2) RBPF for the model parameter estimation preceding the use of 1) Kalman filter or 2) RBPF for the state estimation and denoising of the speech signal. White, colored and real industrial noises are tested for evaluating the performance of the methods. The results show that RBPF methods render better quality than the IKF and KEM, while the RBPF computation cost exceeds IKF and KEM, substantially. Silent-spoken frames are treated differently in RBPF methods for the sake of computation saving in the proposed dual particle-number method. By the dual particle- number strategy for RBPF, its weakness is repaired while its performance remains intact.

8. Acknowledgement

This work has been partially supported by the research department of Shahed University.

References

[1] Martin, R. (1994). Spectral subtraction based on minimum statistics. *In*: Proc. of the Eur. Signal Processing Conf., p. 1182–1185.

[2] Paliwal, K. K., Basu, A. (1987). A speech enhancement method based on Kalman filtering. *In*: proc. Of the IEEE Int. Conf. Acoust., Speech, Signal Processing (Dallas, TX), Apr. 6-9, p. 177-180.

[3] Lim, J.S., Oppenheim, A.V. (1978). All-pole modeling of degraded speech, *IEEE Trans. Acoust., Speech, Signal Processing* 26, 197-210.

[4] Popescu, D.C., Zeljkovic, I. (1998). Kalman Filtering of Colored Noise for Speech Enhancement, *In*: proc. of IEEE CASSP'98, Seattle, USA , V.2, p. 997-1000.

[5] Gibson, J.D., Koo, B., Gray, S.D. (1991). Filtering of colored noise for speech enhancement and coding, *IEEE Trans. Acoust., Speech, Signal Processing* 39, 1732–1742.

[6] Gannot, S., Burshtein, D., Weinstein, E. (1998). Iterative and sequential kalman filter-based speech enhancement algorithms. *IEEE Transactions on Speech and Audio Processing* 6 (4) 373–385.

[7] Mustie're., F., Bouchard, M., Bolic, M. (2008). Low-cost modifications of Rao-Blackwellized particle filters for improved speech denoising. Signal Processing 88, 2678– 2692.

[8] Park, S., Choi, S. (2006). Rao_Blackwellized Particle Filter for Sequential Speech Enhancement. *In*: Proc. Of the International Joint Conference on Neural Networks, Vancouver, BC, Canada, July 16-21, p. 1254-1259.

[9] Ristic, B., Arulampalam, S., Gordon, N. (2004). Beyond the Kalman filter: particle filters for tracking applications. Artech House, London.

[10] Mustiere, F., Bolic´, M., Bouchard, M. (2007). Quality assessment of speech enhanced using particle filters. *In*: the proc. of the IEEE International Conference on Acoustics, Speech and Signal processing (ICASSP), Honolulu, USA, p. III-1197, III-1200.

[11] Ephraim, Y., Malah, D. (1985). Speech enhancement using a minimum mean square error log-spectral amplitude estimator. *IEEE Trans. Acoust., Speech, Signal Processing*, 33, 443–445.

[12] Ephraim, Y. (1992). A Bayesian estimation approach for speech enhancement using hidden Markov models. IEEE Trans. Signal Processing, 40, 725–735.

[13] Hu, Y., Loizou, P.C. (2006). Evaluation of objective measures for speech enhancement. in: proc. Of the International Conference on Spoken Language Processing, Pittsburg, USA, p. 229-338.

[14] Wei, Y., Yi, B. (2007). Unscented Kalman Filtering Based on a Voiced-Unvoiced Speech enhancement. IEEE Transactions on Audio, Speech and Language Processing, 15(4) p. 1280–1193.

[15] Doblinger, G. (1995). Computationally efficient speech enhancement by spectral minima tracking in subbands. In: proc. Of the Eurospeech, 2, p. 1513-1516.

[16] Doucet, A., de Freitas, J.F.G., Gordon, N. (Eds.) (2001). Sequential monte carlo methods in practice. Springer- Verlag, New York.

Author biographies

Farid Mousavipour received the B.S. degree in Biomedical Engineering and M.S. degrees in Electrical Engineering from Shahed University in 2007 and 2011, respectively. From 2011 he has been working for ITRC as a member of R&D Project Engineer. His research interest is speech processing, speech enhancement, voice codecs, and communication systems.

Saeed Seyedtabaii received the B.Sc. degree in Electronics from Sharif University of Technology, Tehran, Iran, in 1984, the M.Sc degree in electronics from Iran University of Science and Technology, Tehran, Iran, in 1987 and Ph.D degree in advanced control from DalTech, Halifax, Canada, in 1995.

Since 1998, he has been with the Engineering Faculty of Shahed University, Tehran, Iran, where he is currently an Assistant Professor. From 1981 to 1991, he was a researcher at the J. D. center of Sharif University, Tehran, Iran. His research interests are filter design, adaptive signal processing and control.