International Journal of Engineering & Technology, 7 (2.7) (2018) 5-8

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Dual and joint estimation for speech enhancement

V. Gopi Tilak[1]*, S. Koteswara Rao[2]

[1]M. Tech student, ECE Department, KLEF (Deemed to be University), Guntur
[2]Professor, ECE Department, KLEF (Deemed to be University), Guntur
**Corresponding author E-mail: gopitilak7@gmail.com*

Abstract

Maintaining good quality and intelligibility of speech is the primary constraint in mobile communications. The present work is on the enhancement of speech under the consideration of additive white and colored noise environments using Kalman filter. Dual and Joint estimation techniques were applied and the quality of speech is analyzed through the signal to noise ratio. The techniques were applied in both ideal and practical cases for two different speech samples.

Keywords: *Dual Estimation; Joint Estimation; Intelligibility; Kalman Filter; Signal To Noise Ratio*

1. Introduction

Different coding techniques are specified for speech coding and analysis. The substantial performance of speech coders accomplished for noise-free speech made the system unusable in case of additive noise effects, particularly for colored noise conditions [1]-[3]. General noise cancellation algorithms like adaptive least squares where the convergence is high but the application needs to be done by calculating the perfect correlation of noise frame by frame [4]. The spectral subtraction method, the speech and noise parts in a noisy speech is selected based on the energy levels and this method is inappropriate for speech confined in music or very low energy levels of noise[4]-[7]. Another constraint of spectral subtraction method is that the speech is a stationary signal, but it is proved that speech is quasi-stationary for short periods of time. [8] The Weiner filter approach where the filter works under the assumption of the jointly wide sense stationary property of speech state space model and measurement model and another assumption is the observations are already present.

Initially, the Kalman filter applied for speech enhancement in the ideal case, where the parameters are estimated from ideal speech sample [9]. But the practical condition is the speech sample is contaminated with noise. The present research on applying dual and joint estimation schemes of Kalman filter for noisy speech contaminated in white noise and colored noise respectively [10]-[11]. The iterative filtering scheme introduced for parameter estimation from a noisy speech signal [11]-[12]. The dual estimation scheme applied for speech and parameter estimation for speech corrupted in white noise, the joint scheme applied for speech contaminated in colour noise which in turns the augmented canonical form as the vector multiplication of speech and noise. The prior estimates of parameters for filter initiation are estimated using MMSE estimators for ideal and practical cases for both white and colored noise conditions. The performance analysis is presented in graphs and the signal to noise ratio.

2. Speech Enhancement

2.1. Dual Estimation

Dual estimation scheme can be formulated by assuming AR model to the speech signal and giving a dynamic model to the AR parameters. The time-varying speech signal observed by the microphone is $z(t) = x(t) + v(t)$, where $v(t)$ is an additive background noise and $x(t)$ is sampled speech signal realized as

$$\boldsymbol{x(t)} = -\sum_{k=1}^{p} \boldsymbol{a_k(t)x(t-1)} + \boldsymbol{J_x(t)l_x(t)} \quad (1)$$

Where $a_1(t), a_2(t) \dots a_p(t)$ are the time varying AR coefficients, $l_x(t)$ is a normalized white noise with zero mean and unit covariance, $J_x(t)$ Indicates the innovation gain. Now, the observation model observed by the microphone with an additive white Gaussian noise with zero mean and variance ${J_v}^2$. The augmented innovation gain and measurement model vectors are defined as $\boldsymbol{J_x}^T(t) = [J_x(t)\ 0 \dots 0]$ and ${H_x}^T = [1\ 0 \dots 0]$. Then the Euclidean representation is given by

$$\boldsymbol{x}_p(t) = C_x(t)\boldsymbol{x}_p(t-1) + \boldsymbol{J}_x l_x(t) \quad (2)$$
$$Y(t) = {H_x}^T \boldsymbol{x}_p(t) + v(t)$$

Where $\boldsymbol{x_p}^T(t) = [x(t)\ x(t-1) \dots x(t-p)]$ and the state transition matrix $C_x(t)$ is

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

$$C_x(t) = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & & 0 \\ \vdots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & & \ddots & \ddots & \ddots & 1 \\ -a_p(t) & -a_{p-1}(t) & -a_{p-2}(t) & \cdots & -a_2(t) & -a_1(t) \end{bmatrix} \quad (3)$$

$H_x = J_x = [0\ 0\ 0 \ldots 0\ 1]^T$

The Parameter state equations are given by

$$\boldsymbol{a}(t) = C_a\boldsymbol{a}(t-1) + \boldsymbol{l}_a(t) \quad (4)$$
$$y(t) = \boldsymbol{H}_a{}^T(t)\boldsymbol{a}(t) + \boldsymbol{J}_x(t)l_x(t) + v(t)$$

Where the parameter vector is $\boldsymbol{a}^T(t) = [a_1(t)\ a_2(t) \ldots\ a_p(t)]$ the innovation vector with respective covariance $Q_a(t) = E\{\boldsymbol{l}_a(t)\boldsymbol{l}_a{}^T(t)\}$ and the measurement model is $\boldsymbol{H}_a{}^T(t) = [x(t-1)\ x(t-2) \ldots x(t-p)]$ and $c_a(t) = I_{p\text{x}p}$ or very close to it. State dynamic noise and measurement noises are assumed to be uncorrelated.

The Kalman filter implementation for speech as follows
Prediction:

$$\hat{x}_{\boldsymbol{p}}(t/t-1) = C_x\hat{x}_{\boldsymbol{p}}(t-1/t-1) \quad (5)$$
$$P(t/t-1) = C_xP(t-1/t-1){C_x}^T + \boldsymbol{J}_x\boldsymbol{J}_x{}^T$$

Gain:

$$K(t) = \frac{P(t/t-1)\boldsymbol{H}_x}{\boldsymbol{H}_x{}^TP(t/t-1)\boldsymbol{H}_x + R_x} \quad (6)$$

Correction:

$$\hat{x}_{\boldsymbol{p}}(t/t) = \hat{x}_{\boldsymbol{p}}(t/t-1) + K(t)[Y(t) - \boldsymbol{H}_x{}^T\hat{x}_{\boldsymbol{p}}(t/t-1)] \quad (7)$$
$$P(t/t) = P(t/t-1) - K(t)[\boldsymbol{H}_x{}^TP(t/t-1)\boldsymbol{H}_x + R_x]K^T(t)$$

Kalman filter for parameter is
Prediction:

$$\hat{\boldsymbol{a}}(t/t-1) = C_a\hat{\boldsymbol{a}}(t-1/t-1) \quad (8)$$
$$P_a(t/t-1) = C_aP(t-1/t-1){C_a}^T + \boldsymbol{J}_a\boldsymbol{J}_a{}^T$$

Gain:

$$K_a(t) = \frac{P_a(t/t-1)\boldsymbol{H}_a}{\boldsymbol{H}_a{}^TP_a(t/t-1)\boldsymbol{H}_a + R_x + R_a} \quad (9)$$

Correction:

$$\hat{\boldsymbol{a}}(t/t) = \hat{\boldsymbol{a}}(t/t-1) + K_a(t)[y(t) - \boldsymbol{H}_a{}^T\hat{\boldsymbol{a}}(t/t-1)] \quad (10)$$
$$P_a(t/t) == P_a(t/t-1) + K_a(t)[\boldsymbol{H}_a{}^TP_a(t/t-1)\boldsymbol{H}_a + R_x + R_a]K_a^T(t)$$

In each time point, the speech signal gives the estimated AR parameters and the current parameter estimate gives the speech estimate. The Kalman filter gives best optimal causal MMSE results which include the required speech signal x(t) under the assumption that the signal and parameters of noise are known.

2.2 Joint Estimation

Joint estimation applied for nonlinear state estimation of speech where the speech and noise are modeled as stochastic processes and the augmented state vector comprises the multiplication of both speech and noise. Here, the noise is assumed to be a wide sense stationary colored noise modeled in auto-regression is

$$w(t) = \sum_{i=1}^{p} b_iw(t-i) + \gamma(t) \quad (11)$$

Where $\gamma(t)$ is white Gaussian sequence with zero mean and covariance ${\sigma_w}^2$. The state space model and measurement model for colored noise are given by

$$W(t) = C_ww(t-1) + J_w\,\gamma(t) \quad (12)$$
$$y_n(t) = H_w^TW(t)$$

Where $W(t) = [w(t-p+1)w(t-p+2) \ldots w(t)]^T$ and $H_w^T = J_w^T = [0\ 0 \ldots 0\ 1]$. And the noise covariance matrix is

$$C_w(t) = \begin{bmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & & 0 \\ \vdots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & & \ddots & \ddots & \ddots & 1 \\ -b_p(t) & -b_{p-1}(t) & -b_{p-2}(t) & \cdots & -b_2(t) & -b_1(t) \end{bmatrix} \quad (13)$$

The augmented state vector for speech with colored noise is given as a vector as

$$\bar{X}(t) = \bar{C}\bar{X}(t-1) + \bar{J}\bar{l}(t) \quad (14)$$
$$Y(t) = \bar{H}^T\bar{X}(t)$$

Where $\bar{X}(t) = [x(t); W(t)]$, $\bar{l}(t) = [l(t);\ \gamma(t)]$ and the state transition vector is

$$\bar{C} = \begin{bmatrix} C_x & 0 \\ 0 & C_w \end{bmatrix},$$

$$\bar{J} = \begin{bmatrix} J_x & 0 \\ 0 & J_w \end{bmatrix} \text{ and}$$

$$\bar{H}^T = [H_s^T\ \ H_w^T].$$

The correlation between the colored noise and the white noise sequences is given by $\bar{\boldsymbol{J}} \cong E\left[\bar{l}(t)\bar{l}^T(t)\right] = \begin{bmatrix} {\sigma_w}^2 & 0 \\ 0 & {\sigma_\gamma}^2 \end{bmatrix}$. Here the measurement model noise covariance is zero, i.e. R=0.

Now, the Kalman filter applied as follows

$$\hat{\bar{X}}(t) = \bar{C}\hat{\bar{X}}(t-1) \quad (15)$$

The Gain and updated covariance are given by

$$K(t) = P(t/t-1)\bar{H}/[\bar{H}P(t/t-1)\bar{H}^T] \quad (16)$$
$$P(t/t-1) = \bar{C}P(t-1/t-1)\bar{C}^T + j\boldsymbol{J}j^T \quad (17)$$
$$P(t/t) = [I - K(t)\bar{H}^T]\,P(t/t-1) \quad (18)$$

The speech estimate is

$$\hat{x}(t) = [H^T\ \ 0]\hat{\bar{X}}(t) \quad (19)$$

Quality measure is taken on the basis of signal to noise ratio averaged over all the segments is given by

$$\text{SNR} = 10\log_{10}\frac{\frac{1}{N}\sum_{t=0}^{N}x^2(t)}{\frac{1}{N}\sum_{t=0}^{N}[x(t)-\hat{x}(t)]^2} \quad (20)$$

Here, the total number of samples in the utterance is represented by N.

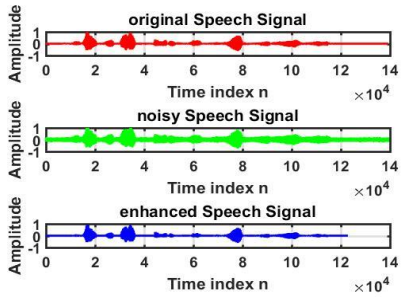
3. Experimental Results

Two speech samples are taken which are male and female voices with different signal to noise ratios. For the estimation procedure, the AR parameters are estimated through an MMSE estimator which gives optimal estimation results and the estimation in two cases. First, for the ideal case, the parameters are estimated through noise-free speech sample and in another hand, the parameters estimated for noisy measurements in both white and colored noise conditions.

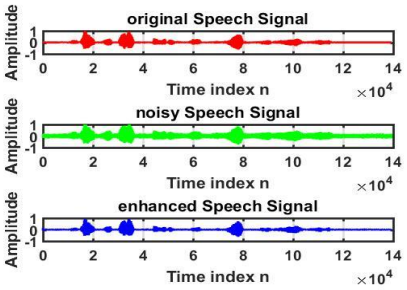
Based on the sampling frequency, the best estimation followed by dividing the entire signal into frames and for each frame, estimation is carried out. The present frame doesn't count the end results of the previous frame for its initiation. Multiple iterations are carried out for each frame. Generally less than 5 iterations enough for obtaining better estimates, 3 iterations are preferred for present work [12].

Sentence 1: Male voice "She had your dark suit in greasy wash water all year".

Sentence 2: Female voice "Which tea party did baker go to?"

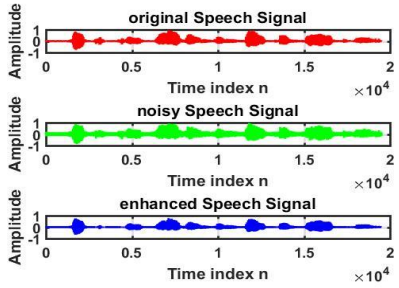


1(a) Ideal case

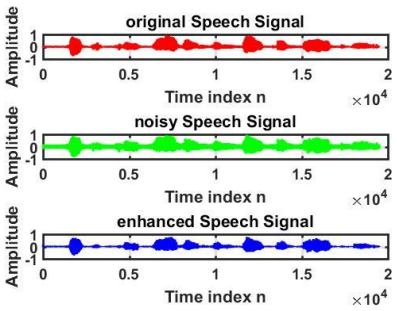


1(a) Practical case

Fig 1. Dual estimation of sentence 1 in 1(a) Ideal and 1(b) Practical cases respectively for white noise condition.

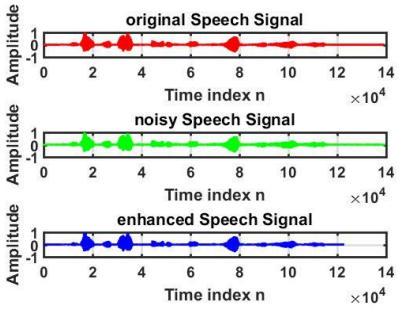


2(a) Ideal case

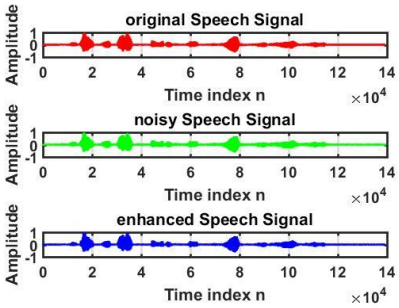


2(b) Practical case

Fig 2. Dual estimation of sentence 2 in 2(a) Ideal and 2(b) Practical cases respectively for colored noise condition.

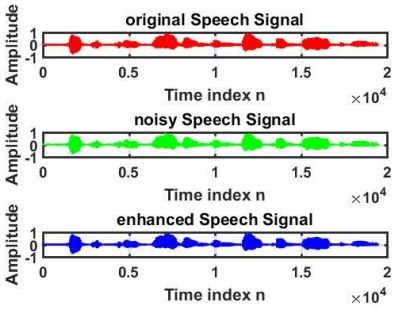


3(a) Ideal case

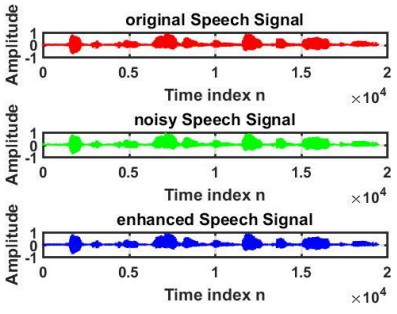


3(b) Practical case

Fig 3. Joint estimation of sentence 1 in 3(a) Ideal and 3(b) Practical cases respectively for white noise condition.



4(a) Ideal case



4(b) Practical case

Fig 4. Joint estimation of sentence 2 in 4(a) Ideal and 4(b) Practical cases respectively for colored noise condition.

The quality measures of the enhanced speech given in below tabular forms.

Table 1: Signal to noise ratios for enhanced speech using dual estimation.

	Practical	Ideal
Sentence 1	14.5450	17.7183
Sentence 2	6.7404	7.8263

Table 2: Signal to noise ratio for enhanced speech using joint estimation.

	Practical	Ideal
Sentence 1	9.0356	12.2341
Sentence 2	1.7660	2.1687

4. Conclusion

It is observed that the iterative Kalman filter with dual and joint estimation for speech in noise contaminated (both white and colored noise) cases gives better signal to noise ratio in ideal case than that of practical case for a single microphone speech enhancement problem. The future work extended to adaptive Kalman filter and Non-linear Kalman filter algorithms applied for single microphone speech enhancement and dual microphone speech dereverberation.

References

[1] M. R. Sambur and N. **S.** Jayant, "LPC analysis/synthesis from speech inputs containing quantization noise or additive white noise," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, pp. 488- 494, Dec. 1976.

[2] C. F. Teacher and D. Coulter, "Performance of LPC vocoders in a noisy environment," in Proc. IEEE Int. Con$ Acoust., Speech, Signal Processing (Washington, DC), Apr. 2-4, 1979, pp. 216-219.

[3] J. D. Gibson, "Theory of LPC analysis for distorted inputs," in Proc. Eighteenth Annu. Allerton Con$ Commun. , Contr., Computing**,** (Monticello, IL), Oct. 8-10. 1980, p. 955.

[4] Vinayshankar Somalara Nataraj, Athulya M. S., Sathidevi Puthumangalathu Savithri, 'Single Channel Speech Enhancement Using Adaptive Filtering and Best Correlating Noise Identification', 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE).

[5] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Tans. Acoust., Speech, Signal Processing, vol. 27 Issue 2, pp. 113-120, April 1979.

[6] Shambhu Shankar Bharti, Manish Gupta, and Suneeta Agarwal, "A New Spectral Subtraction Method for Speech Enhancement using Adaptive Noise Estimation", 2016 3rd International Conference on Recent Advances in Information Technology (RAIT),pp. 128-132, 2016.

[7] Ching-Ta Lu, Kun-Fu Tsen, Yung-Yue Chen, Ling-Ling Wang, and Chung-Lin Leita, "Speech ENhancement Using Spectral Subtraction Algorithm with Over-Subtraction and Reservation Factors Adapted by Harmonic Properties", 2016 International Conference on Applied System Innovation (ICASI),pp. 1-5, 2016.

[8] Monson H. Hayes, "Statistical digital signal processing and modelling", John Wiley and Sons Inc., publishers.

[9] K. K. Paliwal and Anjan Basu,"A speech enhancement method based on Kalman filtering", acoustics, speech, and signal processing, ieee international conference on icassp '87.

[10] Sharon Gannot and Marc Moonen,"On the application of the unscented Kalman filter to speech enhancement", International Workshop on Acoustic Echo and Noise Control (IWAENC2003), Sept. 2003, Kyoto, Japan.

[11] Jerry D. Gibson, Boneung Koo and Steven D. Gray,"Filtering of colored noise for speech enhancement and coding", IEEE Transactions on signal processing. vol. **39.** no. ***8.*** august 1991.

[12] Sharon Gannot, David Burshtein and Ehud Weinstein, 'Iterative and sequential Kalman filter-based speech enhancement algorithms", IEEE Transactions on speech and audio processing, vol. 6,no. 4, July 1998.