ROBUST ADAPTIVE KALMAN FILTER FOR SPEECH SIGNAL RECOVERY IN COLORED NOISE

Marcel Gabrea

`mgabrea@ele.etsmtl.ca`
École de Technologie Supérieure
Electrical Engineering Department
1100, Notre-Dame West, Montreal, Canada

ABSTRACT

This paper deals with the problem of speech enhancement when a corrupted speech signal with an additive colored noise is the only information available for processing. Kalman filtering is known as an effective speech enhancement technique, in which speech signal is usually modeled as autoregressive (AR) process and represented in the state-space domain. In the above context, all the Kalman filter-based approaches proposed in the past, operate in two steps: first, the noise and the signal parameters are estimated, and second, the speech signal is estimated by using Kalman filtering. In this paper a new sequential estimators are developed for sub-optimal adaptive estimation of the unknown a priori driving processes variances simultaneously with the system state. A robust recursive least-square algorithm with variable forgetting factor is used for the estimation of the speech AR parameters and a recursively least-squares lattice algorithm is used for the estimation of the noise AR parameters. The algorithm provides improved speech estimate at little computational expense.

1. INTRODUCTION

Kalman filtering is known as an effective speech enhancement technique, in which speech signal is usually modeled as autoregressive (AR) process and represented in the state-space domain. Many approaches using Kalman filtering have been referenced in the literature. They usually operate in two steps: first, the noise and the signal parameters are estimated, and second, the speech signal is estimated by using Kalman filtering. These approaches differ essentially one from the other by the choice of the algorithm used to estimate the parameters of such model, the models adopted for the speech signal and the additive noise.

In [1], [2] , [3] and [4] the noise under a simplified assumption is considered as an white Gaussian process, but in [5], [6] and [7] the noise is considered colored. Paliwal and Basu [1] have used estimates of the speech signal parameters from clean speech, before being contaminated by white noise. They then used a delayed version of Kalman filter in order to estimate the speech signal. In [2], Oppenheim et al. have used a time-adaptive algorithm to adaptively estimate the speech model parameters and the noise variance. Gannot et al. [7] have proposed the use of EM (Espectation-Maximisation) algorithm to iteratively estimate the spectral parameters of speech and noise parameters. The enhanced speech signal was obtained as a byproduct of the parameter estimation algorithm. In [6], the coefficients of the AR processes and the AR driving processes variances are estimated based on EM algorithm. Gabrea and O'Shaughnessy [4] have proposed estimating the noise and driving process variances using the property of the innovation sequence, obtained after a preliminary Kalman filtering with an initial gain.

In this paper the speech signal and the additive noise are modeled as the AR processes and a new adaptive Kalman filter based method is proposed to recover the speech signal from a sequence of the speech signal corrupted by an additive colored noise. The estimation of time-varying AR speech model parameters is based on robust recursive least square algorithm with variable forgetting factor. The variable forgetting factor is adapted to a nonstationary signal by a generalized likelihood ratio algorithm through so-called discrimination function, developed for automatic detection of abrupt changes in stationarity of signal. The sequential estimators are derived for sub-optimal adaptive estimation of the unknown a priori driving processes variances simultaneously with the system state by reformulating and adapting the classical approach used for control applications. A limited memory algorithm is developed for adaptive correction of the a priori statistics, which are intended to compensate for time varying model errors. The algorithm involves using the state corrections to estimate the driving processes variances and provides improved state estimates at little computational expense.

The paper is organized as follows. In Section II we present the speech enhancement approach based on the Kalman filter algorithm. Section III is concerned with the estimation of AR parameters and driving processes statistics. Simulation results are the subject of Section IV.

2. NOISY SPEECH MODEL AND KALMAN FILTERING

The speech signal $s(n)$ and the additive noise $v(n)$ are modeled as the pth-order order and qth-order AR processes:

$$s(n) = \sum_{i=1}^{p} a_i s(n-i) + u(n) \tag{1}$$

$$v(n) = \sum_{j=1}^{q} b_j v(n-j) + w(n) \tag{2}$$

$$y(n) = s(n) + v(n) \tag{3}$$

where $s(n)$ is the nth sample of the speech signal, $v(n)$ is the nth sample of the additive noise, $y(n)$ is the nth sample of the

observation, a_i is the *i*th AR speech model parameter and b_j is the *j*th AR noise model parameter.
This system can be represented by the following state-space model:

$$\mathbf{x}(n) = \mathbf{F}\mathbf{x}(n-1) + \mathbf{d}(n) \quad (4)$$

$$y(n) = \mathbf{H}\mathbf{x}(n) \quad (5)$$

where:

1. **x**(*n*) is the $(p+q) \times 1$ state vector
$$\mathbf{x}(n) = [s(n-p+1), \cdots, s(n), v(n-q+1), \cdots, v(n)]^T \quad (6)$$
2. **d**(*n*) is the $(p+q) \times 1$ vector
$$\mathbf{d}(n) = [0, \cdots, 0, u(n), 0, \cdots, 0, w(n)]^T \quad (7)$$
3. the sequences *u*(*n*) and *w*(*n*) are uncorrelated Gaussian white noise sequences with zero means and variances $\sigma_u^2(n)$ and $\sigma_w^2(n)$
4. **F** is the $(p+q) \times (p+q)$ transition matrix
$$\mathbf{F} = \begin{bmatrix} \mathbf{F_s} & 0 \\ 0 & \mathbf{F_v} \end{bmatrix} \quad (8)$$
$$\mathbf{F_s} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ a_p & a_{p-1} & a_{p-2} & \cdots & a_1 \end{bmatrix} \quad (9)$$
$$\mathbf{F_v} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ b_q & b_{q-1} & b_{q-2} & \cdots & b_1 \end{bmatrix} \quad (10)$$
5. **H** is the $1 \times (p+q)$ observation row vector
$$\mathbf{H} = [0, \cdots, 0, 1, 0, \cdots, 0, 1] \quad (11)$$

The standard Kalman filter [8] provides the updating state-vector estimator equations:

$$e(n) = y(n) - \mathbf{H}\hat{\mathbf{x}}(n/n-1) \quad (12)$$

$$\begin{aligned} \mathbf{K}(n) &= \mathbf{P}(n/n-1)\mathbf{H} \times \\ &\times [\mathbf{H}\mathbf{P}(n/n-1)\mathbf{H}^T]^{-1} \end{aligned} \quad (13)$$

$$\hat{\mathbf{x}}(n/n) = \hat{\mathbf{x}}(n/n-1) + \mathbf{K}(n)e(n) \quad (14)$$

$$\mathbf{P}(n/n) = [\mathbf{I} - \mathbf{K}(n)\mathbf{H}]\mathbf{P}(n/n-1) \quad (15)$$

$$\hat{\mathbf{x}}(n+1/n) = \mathbf{F}\hat{\mathbf{x}}(n/n) \quad (16)$$

$$\mathbf{P}(n+1/n) = \mathbf{F}\mathbf{P}(n/n)\mathbf{F}^T + \mathbf{Q}(n) \quad (17)$$

where:

1. $\hat{\mathbf{x}}(n/n-1)$ is the minimum mean-square estimate of the state vector $\mathbf{x}(n)$ given the past observations $y(1), \ldots, y(n-1)$
2. $\tilde{\mathbf{x}}(n/n-1) = \mathbf{x}(n) - \hat{\mathbf{x}}(n/n-1)$ is the predicted state-error vector
3. $\mathbf{P}(n/n-1) = E[\tilde{\mathbf{x}}(n/n-1)\tilde{\mathbf{x}}^T(n/n-1)]$ is the predicted state-error correlation matrix
4. $\hat{\mathbf{x}}(n/n)$ is the filtered estimate of the state vector $\mathbf{x}(n)$
5. $\tilde{\mathbf{x}}(n/n) = \mathbf{x}(n) - \hat{\mathbf{x}}(n/n)$ is the filtered state-error vector
6. $\mathbf{P}(n/n) = E[\tilde{\mathbf{x}}(n/n)\tilde{\mathbf{x}}^T(n/n)]$ is the filtered state-error correlation matrix
7. $\mathbf{Q}(n) = E[\mathbf{d}(n)\mathbf{d}^T(n)]$ is the driving processes correlation matrix
8. *e*(*n*) is the innovation sequence
9. **K**(*n*) is the Kalman gain

The estimated speech signal can be retrieved as the *p*th component of the state-vector estimator $\hat{\mathbf{x}}(n/n)$.

3. PARAMETER ESTIMATION

The estimation of the driving processes variances is derived under the assumption of the constant values over N samples by reformulating and adapting the approach proposed in control by Myers and Tapley [9]. The estimation of the transition matrix, which contains the AR models parameters, was made using a adaptation of the robust recursive least square algorithm with variable forgetting factor proposed by Milosavljevic et al. [10].

3.1. Estimation of Driving Processes Variances

The estimation of driving processes variances needed to compute the matrix $\mathbf{Q}(n)$ is derived under the assumption of the constant variance over N samples $u(n), u(n-1), \cdots, u(n-N+1)$ and $w(n), w(n-1), \cdots, w(n-N+1)$, respectively by reformulating and adapting the approach proposed in control by Myers and Tapley [9]. Using the state propagation equation (4) the samples of the driving process $u(n)$ are given by the equation:

$$u(n) = \mathbf{H}_1[\mathbf{x}(n) - \mathbf{F}\mathbf{x}(n-1)] \quad (18)$$

where $\mathbf{H}_1 = [0, \cdots, 0, 1, 0, \cdots, 0, 0]$. The true state vectors $\mathbf{x}(n)$ and $\mathbf{x}(n-1)$ are unknown, so $u(n)$ cannot be determined, but the approximation:

$$\alpha(n) = \mathbf{H}_1[\hat{\mathbf{x}}(n/n) - \hat{\mathbf{x}}(n/n-1)] \quad (19)$$

can be used [9]. The samples $\alpha(n)$ are assumed to be representative of $u(n)$ and can be considered independent and identically distributed. Based on the last N measurements the variance $\sigma_\alpha^2(n)$ is estimated [11]. The sample variance $\hat{\sigma}_\alpha^2(n)$ is obtained by:

$$\hat{\sigma}_\alpha^2(n) = \frac{1}{N}\sum_{i=0}^{N-1}[\alpha(n-i)]^2 \quad (20)$$

The analysis reduces to expanding $E\{[\alpha(n)]^2\}$ in term of $\sigma_u^2(n)$. We write $\alpha(n)$ in term of the filtered state-error vectors:

$$\alpha(n) = -\mathbf{H}_1\tilde{\mathbf{x}}(n/n) + \mathbf{H}_1\mathbf{F}\tilde{\mathbf{x}}(n-1/n-1) + u(n) \quad (21)$$

Since the filtered state-error vectors errors are not independent, the correlation are avoited by writing:

$$\alpha(n) + \mathbf{H}_1\tilde{\mathbf{x}}(n/n) = \mathbf{H}_1\mathbf{F}\tilde{\mathbf{x}}(n-1/n-1) + u(n) \quad (22)$$

The variance of this equation is:

$$E\{[\alpha(n)+\mathbf{H}_1\tilde{\mathbf{x}}(n/n)]^2\} = \mathbf{H}_1\mathbf{F}\mathbf{P}(n-1/n-1)\mathbf{F}^T\mathbf{H}_1^T+\sigma_u^2(n) \quad (23)$$

Now we develop $E\{[\alpha(n) + \mathbf{H}_1\tilde{\mathbf{x}}(n/n)]^2\}$ in term of $E\{[\alpha(n)]^2\}$ and of other computed terms in the Kalman filter:

$$\begin{aligned} E\{[\alpha(n) + \mathbf{H}_1\tilde{\mathbf{x}}(n/n)]^2\} &= E\{[\alpha(n)]^2\} \\ &+ 2E\{\alpha(n)\tilde{\mathbf{x}}^T(n/n)\mathbf{H}_1^T\} \\ &+ \mathbf{H}_1\mathbf{P}(n/n)\mathbf{H}_1^T \end{aligned} \quad (24)$$

Using the Kalman filter equations the filtered state-error vector can be rewriting as:

$$\tilde{\mathbf{x}}(n/n) = [\mathbf{I} - \mathbf{K}(n)\mathbf{H}]\tilde{\mathbf{x}}(n/n-1) - \mathbf{K}(n)v(n) \quad (25)$$

and the second term in (24) is:

$$\begin{aligned} E\{\alpha(n)\tilde{\mathbf{x}}^T(n/n)\mathbf{H}_1^T\} &= -\mathbf{H}_1\mathbf{P}(n/n)\mathbf{H}_1^T \\ &+ \mathbf{H}_1\mathbf{P}(n/n-1) \times \\ &\times [\mathbf{I} - \mathbf{K}(n)\mathbf{H}]^T\mathbf{H}_1^T \end{aligned} \quad (26)$$

By combining (23)(24) and (26) the resulting expression for is $E\{[\alpha(n)]^2\}$ is:

$$\begin{aligned} E\{[\alpha(n)]^2\} &= \mathbf{H}_1\mathbf{F}\mathbf{P}(n-1/n-1)\mathbf{F}^T\mathbf{H}_1^T \\ &+ \mathbf{H}_1\mathbf{P}(n/n)\mathbf{H}_1^T \\ &- 2\mathbf{H}_1\mathbf{P}(n/n-1)[\mathbf{I} - \mathbf{K}(n)\mathbf{H}]^T\mathbf{H}_1^T \\ &+ \sigma_u^2(n) \end{aligned} \quad (27)$$

Using (20) and (27) an unbiased estimator of $\sigma_u^2(n)$ is given by:

$$\begin{aligned} \hat{\sigma}_u^2(n) &= \frac{1}{N}\sum_{i=0}^{N-1}[\alpha(n-i)]^2 \\ &- \mathbf{H}_1\mathbf{F}\mathbf{P}(n-1/n-1)\mathbf{F}^T\mathbf{H}_1^T \\ &- \mathbf{H}_1\mathbf{P}(n/n)\mathbf{H}_1^T \\ &+ 2\mathbf{H}_1\mathbf{P}(n/n-1) \times \\ &\times [\mathbf{I} - \mathbf{K}(n)\mathbf{H}]^T\mathbf{H}_1^T \end{aligned} \quad (28)$$

The samples of the driving process $w(n)$ can be approximated by:

$$\beta(n) = \mathbf{H}_2[\hat{\mathbf{x}}(n/n) - \hat{\mathbf{x}}(n/n-1)] \quad (29)$$

where $\mathbf{H}_2 = [0, \cdots, 0, 0, 0, \cdots, 0, 1]$. Based on the last N measurements an unbiased estimator of $\sigma_w^2(n)$ is obtained by:

$$\begin{aligned} \hat{\sigma}_w^2(n) &= \frac{1}{N}\sum_{i=0}^{N-1}[\beta(n-i)]^2 \\ &- \mathbf{H}_2\mathbf{F}\mathbf{P}(n-1/n-1)\mathbf{F}^T\mathbf{H}_2^T \\ &- \mathbf{H}_2\mathbf{P}(n/n)\mathbf{H}_2^T \\ &+ 2\mathbf{H}_2\mathbf{P}(n/n-1) \times \\ &\times [\mathbf{I} - \mathbf{K}(n)\mathbf{H}]^T\mathbf{H}_2^T \end{aligned} \quad (30)$$

3.2. Estimation of the Transition Matrix

In our approach, getting $\mathbf{F}$ requires the AR parameters estimation. For a such purpose we estimate the transition matrix in two steps: first, we estimate the noise AR parameters during the silence period and second, the speech AR parameters.

The recursively least-squares lattice (RLSL) [12] algorithm is proposed for adaptive estimation of the noise AR parameters because it has a rate of convergence typically an order of magnitude faster than the least mean squares (LMS) algorithm used in [2] and it provide the best prediction in the sense of least-squarer error of the present value of the noise. The RLSL algorithm is in fact rewriting of the QR-decomposition-based least-squares lattice algorithm (QRD-LSL), which represent the most fundamental form of an order-recursive adaptive filter. This algorithm enjoys many of the properties of the QRD-LSL algorithm, namely, fast convergence, modularity, and an integral set of useful parameters and variables for signal processing applications.

A robust recursive least-square algorithm with variable forgetting factor is used for adaptive estimation of the speech AR parameters using the first *p* components of the state-vector estimator $\hat{\mathbf{x}}(n/n)$. The equation (1) can be rewritten in the form:

$$s(n) = \mathbf{x}_1^T(n-1)\theta(n) + u(n) \quad (31)$$

where

$$\theta(n) = \begin{bmatrix} a_p(n) & a_{p-1}(n) & \cdots & a_1(n) \end{bmatrix}^T \quad (32)$$

and

$$\mathbf{x}_1(n) = [s(n-p+1), \cdots, s(n)]^T \quad (33)$$

The robust recursive least square approach estimates the vector $\hat{\theta}(n)$ by minimising the M-estimation criterion [10]:

$$J_n = \frac{1}{n}\sum_{i=1}^{n}\lambda^{n-i}\rho\left[\epsilon^2(i)\right] \quad (34)$$

where

$$\psi(x) = \rho'(x) = min\left[\frac{|x|}{\sigma_u^2}, \frac{\Delta}{\sigma_u}\right] sgn(x) \quad (35)$$

is the Huber influence function and Δ is a chosen constant. The vector $\mathbf{x}_1(n)$ used in (31) is unknown but can be approximated by $\hat{\mathbf{x}}_1(n/n)$, the first *p* components of the state-vector estimator

$\hat{\mathbf{x}}(n/n)$. In this case the robust recursive least square approach gives the estimation equations:

$$\epsilon(i) = \mathbf{H}_1\hat{\mathbf{x}}(i/i) - \hat{\mathbf{x}}_1^T(i-1/i-1)\hat{\theta}(i-1) \tag{36}$$

$$\begin{aligned}\gamma(i) &= \lambda(i) \\ &+ \psi'\left[\epsilon(i)\right]\hat{\mathbf{x}}_1^T(i-1/i-1)\mathbf{R}(i-1)\hat{\mathbf{x}}_1(i-1/i-1)\end{aligned} \tag{37}$$

$$\mathbf{g}(i) = \frac{\mathbf{R}(i-1)\hat{\mathbf{x}}_1(i-1/i-1)}{\gamma(i)} \tag{38}$$

$$\mathbf{T}(i) = \mathbf{R}(i-1) - \mathbf{g}(i)\hat{\mathbf{x}}_1^T(i-1/i-1)\mathbf{R}(i-1)\psi'\left[\epsilon(i)\right] \tag{39}$$

$$\mathbf{R}(i) = \frac{\mathbf{T}(i)}{\lambda(i)} \tag{40}$$

$$\hat{\theta}(i) = \hat{\theta}(i-1) + \mathbf{R}(i)\hat{\mathbf{x}}_1(i-1/i-1)\psi\left[\epsilon(i)\right] \tag{41}$$

The forgetting factor $\lambda(i)$ is a data weighting factor that is used to weight recent data more heavily and thus to permit tracking slowly varying signal parameters. If a nonstationary signal is composed of stationary subsignals the estimation of the AR parameters can be given by using a forgetting factor varying between λ_{min} and λ_{max}. The modified generalized likelihood ratio algorithm [13] is used for the automatic detection of abrupt changes in stationarity of signal. This algorithm uses three models of the same structure and order, whose parameters are estimated on fixed length windows of signal. These windows are $[i-N+1, i]$, $[i+1, i+N]$ and $[i-N+1, i+N]$, and move one sample forward with each new sample. In the first step of this algorithm is calculated the discrimination function

$$D(i,N) = L(i-N+1, i+N) - L(i-N+1, i) - L(i+1, i+N) \tag{42}$$

where

$$L(a,b) = (b-a+1)ln\left[\frac{1}{b-a+1}\sum_{i=a}^{d}\epsilon^2(i)\right] \tag{43}$$

denotes the maximum of the logarithmic likelihood function. In the second step a strategy for choosing the variable forgetting factor is defined by letting $\lambda(i) = \lambda_{max}$ when $D = D_{min}$ and $\lambda(i) = \lambda_{min}$ when $D = D_{max}$, as well as by taking the linear interpolation between these values.

4. SIMULATION RESULTS

The approach was tested using a speech signal and an additive noise. The speech signals are sentences from the TIMIT database and the noise signals are the samples from the NOISEX database. Table 1 offers a comparison with others approaches, by showing averaged SNR gain based on 10 speech signals and 10 noise simulations for each speech signal.

Compared to the method similar in structure previously proposed by the author in [6] and to the Gibson's algorithm [5], the proposed method provides increases in SNR, as well as improved speech quality and intelligibility for input SNR between -5 and 15 dB. Gibson's algorithm needs two or three iterations to get the highest SNR gain and lead to computational requirements higher than those corresponding to the proposed approach.

	Output SNR		
Input SNR (dB)	[5] (dB)	[6] (dB)	proposed (dB)
-5.00	1.24	3.14	3.92
0.00	4.16	4.78	5.31
5.00	7.35	7.89	8.47
10.00	11.21	11.56	12.32
15.00	15.62	15.93	16.47

Table 1: OUTPUT SNR FOR AN INPUT SPEECH SIGNAL PLUS COLORED NOISE

5. REFERENCES

[1] K. K. Paliwal and A. Basu, "A Speech Enhancement Method Based on Kalman Filtering," in *Proc. ICASSP'87*, pp. 177–180.

[2] A. V. Oppenheim, E. Weinstein, K. C. Zangi, M. Feder, and D. Gauger, "Single-Sensor Active Noise Cancellation," *IEEE Trans. Speech and Audio Processing*, vol. 2, pp. 285–290, Apr. 1994.

[3] M. Gabrea, "Robust Adaptive Kalman Filtering-based Speech Enhancement Algorithm," in *Proc. ICASSP'04*, pp. 301–304.

[4] M. Gabrea and D. O'Shaugnessy, "Speech Signal Recovery in White Noise Using an Adaptive Kalman Filter," in *Proc. EUSIPCO'00*.

[5] J. D. Gibson, B. Koo, and S. D. Gray, "Filtering of Colored Noise for Speech Enhancement and Coding," *IEEE Trans. Signal Processing*, vol. 39, pp. 1732–1742, Aug. 1991.

[6] M. Gabrea, E. Mandridake, and M. Najim, "Adaptive Kalman Filter for Speech Enhancement from Colored Noise," in *Proc. EUSIPCO'98*, pp. 1457–1460.

[7] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and Sequentiel Kalman Filter-Based Speech Enhancement Algorithms," *IEEE Trans. Speech and Audio Processing*, vol. 6, pp. 373–385, July 1998.

[8] B. A. Anderson and J. B. Moore, *Optimal Filtering*, NJ:Prentice-Hall, Englewood Cliffs, 1979.

[9] K. A. Myers and B. D. Tapley, "Adaptive Sequential Estimation with Unknown Noise Statistics," *IEEE Tran. Automatic Control*, vol. AC-21, pp. 520–523, Aug. 1976.

[10] B. D. Kovacevic, M. M. Milosavljevic, and M. Dj. Veinovic, "Robust Recursive AR Speech Analysis," *Signal Processing*, vol. 44, pp. 125–138, 1995.

[11] J. F. Leathrum, "On Sequential Estimation of State Noise Variances," *IEEE Tran. Automatic Control*, vol. AC-26, pp. 745–746, June 1981.

[12] S. Haykin, *Adaptive Filter Theory*, NJ:Prentice-Hall, Upper Saddle River, 2002.

[13] M. Milosavljevic and I. Konvalinka, "The Modified Generalized Likelihood Ratio Algorithm for Automatic Detection of Abrupt Changes in Stationarity of Signals," in *Proc. ISS'88*.