ROBUST ADAPTIVE KALMAN FILTERING-BASED SPEECH ENHANCEMENT ALGORITHM

Marcel Gabrea

École de Technologie Supérieure Electrical Engineering Department 1100, Notre-Dame West, Montreal, Quebec, Canada H3C 1K3 mgabrea@ele.etsmtl.ca

ABSTRACT

This paper deals with the problem of speech enhancement when only a corrupted speech signal is available for processing. Kalman filtering is known as an effective speech enhancement technique, in which speech signal is usually modeled as autoregressive (AR) model and represented in the state-space domain. Various approaches based on the Kalman filter are presented in the literature. They usually operate in two steps: first, additive noise and driving process statistics and speech model parameters are estimated and second, the speech signal is estimated by using Kalman filtering. In this paper sequential estimators are used for suboptimal adaptive estimation of the unknown a priori driving process and additive noise statistics simultaneously with the system state. The estimation of time-varying AR signal model is based on robust recursive least-square algorithm with variable forgetting factor. The proposed algorithm provides improved state estimates at little computational expense.

1. INTRODUCTION

Speech enhancement using a single microphone system has become an active research area for audio signal enhancement. Various approaches based on the Kalman filter are presented in the literature. They usually operate in two steps: first, noise and driving process statistics and speech model parameters are estimated and second, the speech signal is estimated by using Kalman filtering. In fact these approaches differ only by the choice of the algorithm used to estimate model parameters and the choice of the models adopted for the speech signal and the additive noise.

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.

Gibson et al. [2] have proposed a method that provides a suboptimal solution, which is a simplified version of the Estimate-Maximize (EM) algorithm based on the maximum likelihood argument. However, noise variance was estimated during the silent period, which implies the use of Voice Activity Detector (VAD).

Gabrea et al. [3] have proposed a method that avoids the explicit estimation of noise and driving process variances by estimating the optimal Kalman gain. After a preliminary Kalman filtering with an initial sub-optimal gain, an iterative procedure is derived to estimate the optimal Kalman gain using the property of the innovation sequence. Grivel et al. [4] have suggested that the speech enhancement problem can be stated as a realization issue in the framework of identification. The state-space model was identified using a subspace non-iterative algorithm based on orthogonal projection.

In this paper the signal is modeled as an AR process and a Kalman filter based-method is proposed. The sequential estimators are derived for sub-optimal adaptive estimation of the unknown a priori driving process and additive noise statistics simultaneously with the system state by reformulating and adapting the classical approach used for control applications. First and secondorder moments of the additive noise and driving process are estimated based on state and residual noise samples generated in the Kalman filter algorithm. 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 innovation sequence to estimate the additive noise variance and the state corrections to estimate the driving process variance. The estimation of time-varying AR signal model 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 socalled discrimination function, developed for automatic detection of abrupt changes in stationarity of signal. The algorithm provides improved state estimates at little computational expense. A distinct advantage of the proposed algorithm is that a VAD is not required.

This 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, driving process and additive noise statistics. Simulation results are the subject of Section IV.

2. NOISY SPEECH MODEL AND KALMAN FILTERING

The speech signal s(n) is modeled as a *p*th-order order AR process

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

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

where s(n) is the *n*th sample of the speech signal, y(n) is the *n*th sample of the observation, and $a_i(n)$ is the *i*th AR parameter.

This system can be represented by the following state-space

model

$$\mathbf{x}(n) = \mathbf{F}(n)\mathbf{x}(n-1) + \mathbf{G}u(n)$$
(3)

$$y(n) = \mathbf{H}\mathbf{x}(n) + v(n) \tag{4}$$

where

- the sequences u(n) and v(n) are uncorrelated Gaussian white noise sequences with the means ū and v and the variances σ_u² and σ_v²
- 2. $\mathbf{x}(n)$ is the $p \times 1$ state vector

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

3. **F**(n) is the $p \times p$ transition matrix

$$\mathbf{F}(n) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ a_p(n) & a_{p-1}(n) & a_{p-2}(n) & \cdots & a_1(n) \end{bmatrix}$$

4. **G** and **H** are, respectively, the $p \times 1$ input vector and the $1 \times p$ observation row vector which is defined as follows

$$\mathbf{H} = \mathbf{G}^T = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

The standard Kalman filter [5] provides the updating statevector estimator equations

$$e(n) = y(n) - \mathbf{H}\hat{\mathbf{x}}(n/n-1) - \bar{v}$$
(5)

$$\mathbf{K}(n) = \mathbf{P}(n/n-1)\mathbf{H}^T \times \\ \times \left[\mathbf{H}\mathbf{P}(n/n-1)\mathbf{H}^T + \sigma_v^2\right]^{-1}$$
(6)

$$\hat{\mathbf{x}}(n/n) = \hat{\mathbf{x}}(n/n-1) + \mathbf{K}(n)e(n)$$
(7)

$$\mathbf{P}(n/n) = [\mathbf{I} - \mathbf{K}(n)\mathbf{H}]\mathbf{P}(n/n - 1)$$
(8)

$$\hat{\mathbf{x}}(n+1/n) = \mathbf{F}(n)\hat{\mathbf{x}}(n/n) + \mathbf{G}\bar{u}$$
 (9)

$$\mathbf{P}(n+1/n) = \mathbf{F}(n)\mathbf{P}(n/n)\mathbf{F}^{T}(n) + \mathbf{G}\mathbf{G}^{T}\sigma_{u}^{2} \quad (10)$$

where

- x̂(n/n − 1) is the minimum mean-square estimate of the state vector x(n) given the past n − 1 observations y(1), ..., y(n − 1)
- 2. $\mathbf{\tilde{x}}(n/n-1) = \mathbf{x}(n) \mathbf{\hat{x}}(n/n-1)$ is the predicted stateerror vector
- 3. $\mathbf{P}(n/n-1) = E[\mathbf{\tilde{x}}(n/n-1)\mathbf{\tilde{x}}^T(n/n-1)]$ is the predicted state-error correlation matrix
- 4. $\mathbf{\hat{x}}(n/n)$ is the filtered estimate of the state vector $\mathbf{x}(n)$
- 5. $\mathbf{\tilde{x}}(n/n) = \mathbf{x}(n) \mathbf{\hat{x}}(n/n)$ is the filtered state-error vector
- P(n/n) = E[\$\tilde{x}(n/n)\$\tilde{x}^T(n/n)]\$ is the filtered state-error correlation matrix

7. e(n) is the innovation sequence

8. $\mathbf{K}(n)$ is the Kalman gain

The estimated speech signal can be retrieved from the state-vector estimator

$$\hat{s}(n) = \mathbf{H}\hat{\mathbf{x}}(n/n) \tag{11}$$

The parameter estimation (the transition matrix and noise statistics) is presented in the next section.

3. PARAMETER ESTIMATION

The estimation of the transition matrix, which contains the AR speech model parameters, was made using a adaptation of the robust recursive least square algorithm with variable forgetting factor proposed by Milosavljevic et al. [6]. The estimation of the noise statistics 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 [7].

3.1. Estimation of the Transition Matrix

\$

In our approach, getting $\mathbf{F}(n)$ requires the AR parameter estimation. The equation (3) can be rewritten in the form

$$\mathbf{s}(n) = \mathbf{x}^{T}(n-1)\theta(n) + u(n) \tag{12}$$

where

$$\theta(n) = \begin{bmatrix} a_p(n) & a_{p-1}(n) & \cdots & a_1(n) \end{bmatrix}^T$$
(13)

The robust recursive least square approach estimates the vector $\hat{\theta}(n)$ by minimising the M-estimation criterion [6]

$$J_n = \frac{1}{n} \sum_{i=1}^n \lambda^{n-i} \rho\left[\epsilon^2(i)\right] \tag{14}$$

where

ų

$$\psi(x) = \rho'(x) = \min\left[\frac{|x|}{\sigma_u^2}, \frac{\Delta}{\sigma_u}\right] sgn(x)$$
(15)

is the Huber influence function and Δ is a chosen constant. The true state vector $\mathbf{x}(n)$ used in (12) is unknown but can be approximated by the state-vector estimator $\hat{\mathbf{x}}(n/n)$. In this case the robust recursive least square approach gives the estimation equations

$$\epsilon(i) = \mathbf{H}\hat{\mathbf{x}}(i/i) - \hat{\mathbf{x}}^{T}(i-1/i-1)\hat{\theta}(i-1)$$
(16)

$$\mathbf{g}(i) = \frac{\mathbf{Q}(i-1)\hat{\mathbf{x}}(i-1/i-1)}{\lambda(i) + \psi'\left[\epsilon(i)\right]\hat{\mathbf{x}}^{T}(i-1/i-1)\mathbf{Q}(i-1)\hat{\mathbf{x}}(i-1/i-1)}$$
(17)

$$\mathbf{Q}(i) = \frac{1}{\lambda(i)} \left[\mathbf{Q}(i-1) - \mathbf{g}(i) \hat{\mathbf{x}}^{T}(i-1/i-1) \mathbf{Q}(i-1) \psi'[\epsilon(i)] \right]$$
(18)

$$\hat{\theta}(i) = \hat{\theta}(i-1) + \mathbf{Q}(i)\hat{\mathbf{x}}(i-1/i-1)\psi[\epsilon(i)]$$
(19)

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 [8] 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

$$D(i,N) = L(i-N+1,i+N) - L(i-N+1,i) - L(i+1,i+N)$$
(20)

where

$$L(a,b) = (b-a+1)ln\left[\frac{1}{b-a+1}\sum_{i=a}^{d}\epsilon^{2}(i)\right]$$
(21)

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.

3.2. Estimation of Additive Noise Statistics

The estimation of additive noise statistics is derived under the assumption of the constant mean and variance over N samples v(n), $v(n-1), \dots, v(n-N+1)$. Using the equation (4) the samples of the additive noise are given by the equation

$$v(n) = y(n) - \mathbf{H}\mathbf{x}(n) \tag{22}$$

The true states vector $\mathbf{x}(n)$ is unknown, so v(n) cannot be determined, but the approximation

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

= $\mathbf{H}\tilde{\mathbf{x}}(n/n-1) + v(n)$ (23)

can be used [7]. The samples $\alpha(n)$ are assumed to be representative of v(n) and can be considered independent and identically distributed [7]. Based on the last N samples $\alpha(n), \alpha(n-1), \cdots, \alpha(n-N+1)$ the mean $\overline{\alpha}(n)$ and the variance $\sigma_{\alpha}^2(n)$ are estimated. An unbiased esimator for $\overline{\alpha}(n)$ is taken as the sample mean

$$\hat{\alpha}(n) = \frac{1}{N} \sum_{i=0}^{N-1} \alpha(n-i)$$
 (24)

and an unbiased estimator for $\sigma_{\alpha}^2(n)$ is obtained by

$$\hat{\sigma}_{\alpha}^{2}(n) = \frac{1}{N-1} \sum_{i=0}^{N-1} [\alpha(n-i) - \hat{\alpha}(n)]^{2}$$
(25)

The estimation of the additive noise mean is

$$\hat{v}(n) = \hat{\alpha}(n)$$
 (26)

If the samples $\alpha(n)$ are considered independent and identically distributed the expected value of $\hat{\sigma}_{\alpha}^2(n)$ is

$$E\{\hat{\sigma}_{\alpha}^{2}(n)\} = \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{HP}(n/n-1)\mathbf{H}^{T} + \sigma_{v}^{2}$$
(27)

Using (25) and (27) an unbiased estimator of $\sigma_v^2(n)$ is given by

$$\hat{\sigma}_{v}^{2} = \frac{1}{N-1} \{ \sum_{i=0}^{N-1} [\alpha(n-i) - \hat{\alpha}(n)]^{2} - \frac{N-1}{N} \mathbf{HP}(n-i/n-i-1)\mathbf{H}^{T} \}$$
(28)

3.3. Estimation of Driving Process Statistics

The estimation of driving process statistics is derived under the assumption of the constant mean and variance over N samples $u(n), u(n-1), \dots, u(n-N+1)$. Using the state propagation equation (3) the samples of the driving process are given by the equation:

$$u(n) = \mathbf{H}[\mathbf{x}(n) - \mathbf{F}\mathbf{x}(n-1)]$$
(29)

The true state vectors $\mathbf{x}(n)$ and $\mathbf{x}(n-1)$ are unknown, so u(n) cannot be determined, but the approximation

$$\beta(n) = \mathbf{H}[\hat{\mathbf{x}}(n/n) - \hat{\mathbf{x}}(n/n-1)]$$
(30)

can be used [7]. The samples $\beta(n)$ are assumed to be representative of u(n) and can be considered independent and identically distributed [7]. Based on the last N measurements the mean $\overline{\beta}(n)$ and the variance $\sigma_{\beta}^2(n)$ are estimated [9]. An unbiased estimator for $\overline{\beta}(n)$ is taken as the sample mean

$$\hat{\bar{\beta}}(n) = \frac{1}{N} \sum_{i=0}^{N-1} \beta(n-i)$$
 (31)

and an unbiased estimator for $\sigma_{\beta}^2(n)$ is obtained by

$$\hat{\sigma}_{\beta}^{2}(n) = \frac{1}{N-1} \sum_{i=0}^{N-1} [\beta(n-i) - \hat{\beta}(n)]^{2}$$
(32)

The esimation of the additive noise mean is

$$\hat{\bar{u}}(n) = \bar{\beta}(n)$$
 (33)

If the samples $\beta(n)$ are considered independent and identically distributed the expected value of $\hat{\sigma}_{\beta}^2(n)$ is

$$E\{\hat{\sigma}_{\beta}^{2}(n)\} = \frac{1}{N} \sum_{i=0}^{N-1} E\{[\beta(n-i)]^{2}\}$$
(34)

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

$$\beta(n) = -\mathbf{H}\tilde{\mathbf{x}}(n/n) + \mathbf{H}\mathbf{F}\tilde{\mathbf{x}}(n-1/n-1) + u(n) - \bar{u} \quad (35)$$

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

$$\beta(n) + \mathbf{H}\tilde{\mathbf{x}}(n/n) = \mathbf{H}\mathbf{F}\tilde{\mathbf{x}}(n-1/n-1) + u(n) - \bar{u} \quad (36)$$

The variance of this equation is

$$E\{[\beta(n) + \mathbf{H}\tilde{\mathbf{x}}(n/n)]^2\} = \mathbf{HFP}(n-1/n-1)\mathbf{F}^T\mathbf{H}^T + \sigma_u^2$$
(37)
Now we develop $E\{[\beta(n) - \mathbf{H}\tilde{\mathbf{x}}(n/n)]^2\}$ in term of $E\{[\beta(n)]^2\}$
and of other computed terms in the Kalman filter

$$E\{[\beta(n) + \mathbf{H}\tilde{\mathbf{x}}(n/n)]^2\} = E\{[\beta(n)]^2\} + 2E\{\beta(n)\tilde{\mathbf{x}}^T(n/n)\mathbf{H}^T\} + \mathbf{H}\mathbf{P}(n/n)\mathbf{H}^T$$
(38)

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

$$\tilde{\mathbf{x}}(n/n) = [\mathbf{I} - \mathbf{K}(n)\mathbf{H}]\tilde{\mathbf{x}}(n/n - 1) - \mathbf{K}(n)[v(n) - \bar{v}] \quad (39)$$

and the second term in (37) is

(

$$E\{\beta(n)\tilde{\mathbf{x}}^{T}(n/n)\mathbf{H}^{T}\} = -\mathbf{H}\mathbf{P}(n/n)\mathbf{H}^{T} + \mathbf{H}\mathbf{P}(n/n-1) \times \\ \times [\mathbf{I} - \mathbf{K}(n)\mathbf{H}]^{T}\mathbf{H}^{T} \quad (40)$$

Using (32) (34) and (37) an unbiased estimator of $\sigma_u^2(n)$ is given by

$$\hat{\boldsymbol{\tau}}_{u}^{2} = \frac{1}{N-1} \{ \sum_{i=0}^{N-1} [\beta(n-i) - \hat{\beta}(n)]^{2} \\ - \frac{N-1}{N} \mathbf{HFP}(n-i-1/n-i-1) \mathbf{F}^{T} \mathbf{H}^{T} \\ - \frac{N-1}{N} \mathbf{HP}(n-i/n-i) \mathbf{H}^{T} \\ + 2 \frac{N-1}{N} \mathbf{HP}(n-i/n-i-1) \times \\ \times [\mathbf{I} - \mathbf{K}(n-i) \mathbf{H}]^{T} \mathbf{H}^{T} \}$$
(41)

4. SIMULATION RESULTS

The approach was tested using a speech signal and an additive Gaussian white noise. The speech signals are sentences from the TIMIT database. Figure 1 represents, respectively, the noise-free speech signal, the noisy speech signal and the enhanced speech signal. For this example, the SNR of the noisy speech signal is 0 dB. 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.

		Output SNR		
Input SNR	[2]	[3]	[10]	proposed
(dB)	(dB)	(dB)	(dB)	(dB)
-5.00	2.46	-2.52	-1.46	2.61
0.00	4.57	2.61	2.65	4.95
5.00	7.96	6.83	7.08	8.52
10.00	11.92	10.95	11.46	12.71
15.00	16.00	15.08	15.34	16.86

 Table 1. Output SNR for an Input Speech Signal plus

 White Noise

For input SNR between -5 and 15 dB the proposed method provides better results than two previously proposed methods by the author [3][10] and Gibson's algorithm [2]. Gibson's algorithm [2], needs two to three iterations to get the highest SNR gain. Its computational requirements are higher, since a voice activity detector is required to determine silence periods.

5. REFERENCES

- K. K. Paliwal and A. Basu, "A speech enhancement method based on kalman filtering," in *Proc. ICASSP'87*, pp. 177– 180.
- [2] 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.



Fig. 1. Example of speech signal enhancement (Input SNR = 0 dB)

- [3] M. Gabrea, E. Grivel, and M. Najim, "A single microphone kalman filter-based noise canceller," *IEEE Signal Processing Lett.*, vol. 6, pp. 55–57, Mar. 1999.
- [4] E. Grivel, M. Gabrea, and M. Najim, "Speech Enhancement as a Realisation Issue," *Signal Processing*, vol. 82, pp. 1963– 1978, Dec. 2002.
- [5] B. A. Anderson and J. B. Moore, *Optimal Filtering*, NJ:Prentice-Hall, Englewood Cliffs, 1979.
- [6] B. D. Kovacevic, M. M. Milosavljevic, and M. Dj. Veinovic, "Robust recursive ar speech analysis," *Signal Processing*, vol. 44, pp. 125–138, 1995.
- [7] 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.
- [8] 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*.
- [9] J. F. Leathrum, "On sequential estimation of state noise variances," *IEEE Tran. Automatic Control*, vol. AC-26, pp. 745– 746, June 1981.
- [10] M. Gabrea and D. O'Shaugnessy, "Speech signal recovery in white noise using an adaptive kalman filter," in *Proc. EU-SIPCO'00*.