SPEECH ENHANCEMENT BASED ON A ROBUST ADAPTIVE KALMAN FILTER

Marcel Gabrea

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

ABSTRACT

This paper deals with the problem of Adaptive Noise Cancellation (ANC) when only a corrupted speech signal with an additive Gaussian white noise 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. All the approaches based on the Kalman filter proposed in the past, in this context, operate in two steps: they first estimate the noise variances and the parameters of the signal model and secondly estimate the speech signal. In this paper the estimation of the noise variances is made by reformulating and adapting the Mehra approach. 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. The aim is to retrieve the desired speech signal from the noisy observations. The approaches based on the Kalman filter reported in the literature [1] [2] [3] [4] [5] [6] differ essentially one from the other by the choice of the algorithm used to estimate the parameters of such a model, the models adopted for the speech signal and the additive noise.

In [1][2] [4] [5] and [6] the noise under a simplified assumption is considered as a white Gaussian process, but in [3] the noise is considered colored and modelled as an AR process. The speech signal is modelled as an AR process except in [4] where it is modelled as an ARMA process.

In this approach the signal is modelled as an AR process. 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 so-called discrimination function, developed for automatic detection of abrupt changes in stationarity of signal. The estimation of the driving noise variance and of the additive noise variance are handled after a preliminary Kalman filtering. 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 organised as follows. In Section II we present the speech enhancement approach based on the Kalman filter algorithm. Section III is concerned with the presentation of the estimation of the AR parameters and the process variances. The simulation results are the subject of Section IV.

2. NOISY SPEECH MODEL AND KALMAN FILTERING

The speech signal $s(n)$ is modeled as a pth-order order AR process

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

$$y(n) = s(n) + v(n) \qquad (2)$$

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

This system can be represented by the following state-space model

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

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

where

1. the sequences $u(n)$ and $v(n)$ are uncorrelated Gaussian white noise sequences with the zero means and the variances σ_u^2 and σ_v^2
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

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

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 [7] provides the updating state-vector estimator equations

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

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

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

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

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

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

where

1. $\hat{\mathbf{x}}(n/n-1)$ is the minimum mean-square estimate of the state vector $\mathbf{x}(n)$ given the past $n-1$ 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. *e*(*n*) is the innovation sequence
8. **K**(*n*) is the Kalman gain

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

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

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. [8]. The estimation of driving noise variance σ_u^2 and of additive noise σ_v^2 was derived using the property of the innovation sequence, obtained after a preliminary Kalman filtering with an initial gain by reformulating and adapting the approach proposed in control by R. K. Mehra [9].

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

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

where

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

The robust recursive least square approach estimates the vector $\hat{\theta}(n)$ by minimising the M-estimation criterion [8]

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

where

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

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

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

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

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

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

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 non-stationary 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 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

$$\begin{aligned} D(i,N) &= L(i-N+1, i+N) - L(i-N+1, i) \\ &- L(i+1, i+N) \end{aligned} \quad (20)$$

where

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

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

Let the predicted state-error vector $\tilde{\mathbf{x}}(n/n-1) = \mathbf{x}(n) - \hat{\mathbf{x}}(n/n-1)$, $\mathbf{P}(n/n-1) = E[\tilde{\mathbf{x}}(n/n-1)\tilde{\mathbf{x}}^T(n/n-1)]$ the predicted state-error correlation matrix and $r_{ee}(k) = E[e(n)e(n-k)]$ the autocorrelation of the innovation process $e(n)$. Using the standard Kalman filter equations and the state-space model equations, the innovation autocorrelation function is obtained:

$$r_{ee}(k) = \begin{cases} T_1(n)\left[T_2(n) - T_3(n)\right] & k > 0 \\ T_4(n) & k = 0 \end{cases} \quad (22)$$

where:

$$T_1(n) = \mathbf{H}\left\{\prod_{i=1}^{k-1}\mathbf{F}[\mathbf{I} - \mathbf{K}(n-i)\mathbf{H}]\right\}\mathbf{F}$$

$$T_2(n) = \mathbf{P}(n-k/n-k-1)\mathbf{H}^T$$

$$T_3(n) = \mathbf{K}(n-k)[\mathbf{H}\mathbf{P}(n-k/n-k-1)\mathbf{H}^T + \sigma_v^2]$$

$$T_4(n) = \mathbf{H}\mathbf{P}(n/n-1)\mathbf{H}^T + \sigma_v^2$$

It is known that in the optimal case the innovation process $e(n)$ is orthogonal to all past observations $y(1), y(2), \cdots$, $y(n-1)$ and consists of a sequence of random variables that are orthogonal to each other, as shown by $r_{ee}(k) = 0$ for $k > 0$ [10]. If a suboptimal gain $\mathbf{K}_0$ and the estimation of transition matrix $\hat{\mathbf{F}}$ were used, the innovation sequence in general is not a white process and $r_{ee}(k) \neq 0$ for $k > 0$. In the steady-state, using the suboptimal gain $\mathbf{K}_0$, $\mathbf{P}(n-k/n-k-1) \simeq \mathbf{P}_0$ and the innovation autocorrelation function $r_{ee}^0(k)$ is:

$$r_{ee}^0(0) = \mathbf{H}\mathbf{P}_0\mathbf{H}^T + \sigma_v^2 \quad (23)$$

$$r_{ee}^0(k) = \mathbf{H}[\hat{\mathbf{F}}(\mathbf{I} - \mathbf{K}_0\mathbf{H})]^{k-1}\hat{\mathbf{F}}[\mathbf{P}_0\mathbf{H}^T - \mathbf{K}_0 r_{ee}^0(0)] \quad k > 0 \quad (24)$$

It is very difficult to estimate the predicted state-error correlation matrix $\mathbf{P}_0$ in terms of $\hat{\mathbf{F}}$, $r_{ee}^0(k)$ and $\mathbf{K}_0$, but using the innovation autocorrelation function (23)(24) it is easy to estimate $\mathbf{P}_0\mathbf{H}^T$, a linear combination of their column. Using (24) the following relationship for $\mathbf{P}_0\mathbf{H}^T$ is obtained [9]:

$$\mathbf{P}_0\mathbf{H}^T = \mathbf{K}_0 r_{ee}^0(0) + \begin{bmatrix} \mathbf{H}\hat{\mathbf{F}} \\ \mathbf{H}[\hat{\mathbf{F}}(\mathbf{I} - \mathbf{K}_0\mathbf{H})]^1\hat{\mathbf{F}} \\ \vdots \\ \mathbf{H}[\hat{\mathbf{F}}(\mathbf{I} - \mathbf{K}_0\mathbf{H})]^{p-1}\hat{\mathbf{F}} \end{bmatrix}^{-1} \begin{bmatrix} r_{ee}^0(1) \\ r_{ee}^0(2) \\ \vdots \\ r_{ee}^0(p) \end{bmatrix} \quad (25)$$

Using $r_{ee}^0(0)$ given by (23), the expression of additive noise variance σ_v^2 is:

$$\sigma_v^2 = r_{ee}^0(0) - \mathbf{H}\mathbf{P}_0\mathbf{H}^T \quad (26)$$

3.3. Estimation of Driving Process Statistics

The estimation of driving noise variance is based on the equation of the update of the state-error covariance matrix. However, we need to reformulate this expression in a way that it becomes convenient for computation of the driving noise variance. We need this reformulation because we can easily estimate $\mathbf{P}(n/n-1)\mathbf{H}^T$ (in our case $\mathbf{P}_0\mathbf{H}^T$) and not $\mathbf{P}(n/n-1)$. Using the Riccati equation [7] the update of the state-error covariance matrix can be rewritten by:

$$\begin{aligned} \mathbf{P}(n/n-1) = \\ & \mathbf{F}\mathbf{P}(n-1/n-2)\mathbf{F}^T + \mathbf{G}\mathbf{G}^T\sigma_u^2 + \\ & \mathbf{F}\mathbf{K}(n-1)\mathbf{K}^T(n-1)r_{ee}(0)\mathbf{F}^T - \\ & \mathbf{F}\mathbf{K}(n-1)\mathbf{H}\mathbf{P}(n-1/n-2)\mathbf{F}^T - \\ & \mathbf{F}\mathbf{P}(n-1/n-2)\mathbf{H}^T\mathbf{K}(n-1)\mathbf{F}^T \end{aligned} \quad (27)$$

With the suboptimal gain $\mathbf{K}_0$ and the estimation of transition matrix $\hat{\mathbf{F}}$ in the steady-state, the equation (27) is given by:

$$\mathbf{P}_0 = \hat{\mathbf{F}}\mathbf{P}_0\hat{\mathbf{F}}^T + \mathbf{G}\mathbf{G}^T\sigma_u^2 + \mathbf{V} \tag{28}$$

where

$$\mathbf{V} = \hat{\mathbf{F}}[\mathbf{K}_0\mathbf{K}_0^T r_{ee}^0(0) - \mathbf{K}_0\mathbf{H}\mathbf{P}_0 - \mathbf{P}_0\mathbf{H}^T\mathbf{K}_0]\hat{\mathbf{F}}^T$$

Now, substituting $\mathbf{P}_0$ in the right side of (28), after $k-1$ substitutions the following equation is obtained:

$$\begin{aligned}\mathbf{P}_0 = \hat{\mathbf{F}}^k\mathbf{P}_0(\hat{\mathbf{F}}^k)^T &+ \sum_{i=0}^{k-1}\hat{\mathbf{F}}^i\mathbf{G}\mathbf{G}^T(\hat{\mathbf{F}}^i)^T\sigma_u^2 \\ &+ \sum_{i=0}^{k-1}\hat{\mathbf{F}}^i\mathbf{V}(\hat{\mathbf{F}}^i)^T \end{aligned} \tag{29}$$

To reformulate (29) in terms of $\mathbf{P}_0\mathbf{H}^T$ we premultiply with $\mathbf{H}$ and postmultiply with $(\hat{\mathbf{F}}^{-k})^T\mathbf{H}^T$. Using the symmetry property of $\mathbf{P}_0$ and $\mathbf{G} = \mathbf{H}^T$, the following expression of driving noise variance is obtained:

$$\begin{aligned}\sigma_u^2 &= \frac{(\mathbf{P}_0\mathbf{H}^T)^T(\hat{\mathbf{F}}^{-k})^T\mathbf{H}^T - \mathbf{H}\hat{\mathbf{F}}^k\mathbf{P}_0\mathbf{H}^T}{\sum_{i=0}^{k-1}\mathbf{H}\hat{\mathbf{F}}^i\mathbf{H}^T\mathbf{H}(\hat{\mathbf{F}}^{i-k})^T\mathbf{H}^T} \\ &- \frac{\sum_{i=0}^{k-1}\mathbf{H}\hat{\mathbf{F}}^i\mathbf{V}(\hat{\mathbf{F}}^{i-k})^T\mathbf{H}^T}{\sum_{i=0}^{k-1}\mathbf{H}\hat{\mathbf{F}}^i\mathbf{H}^T\mathbf{H}(\hat{\mathbf{F}}^{i-k})^T\mathbf{H}^T}\end{aligned} \tag{30}$$

We can see that the denominator of equation (30) is a combination of terms $\mathbf{H}\hat{\mathbf{F}}^i$, $(0 \leq i \leq k-1)$, and $\mathbf{H}(\hat{\mathbf{F}}^{-j})^T\mathbf{H}^T$, $(1 \leq j \leq k)$. If we pay attention to the particular structure of $\hat{\mathbf{F}}$ and $\mathbf{H}$ we remark that $\mathbf{H}(\hat{\mathbf{F}}^{-j})^T\mathbf{H}^T$ is null for $1 \leq j \leq p-1$. This is the reason why the modification of the equation (28) was needed and a value of k greater or equal to p will be chosen.

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. 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.

For input SNR between -5 and 15 dB the proposed method provides better results than three previously proposed methods by the author [2] [6] [5] and Gibson's algorithm [3]. Gibson's algorithm [3], 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.

	Out SNR				
In SNR (dB)	[3] (dB)	[2] (dB)	[6] (dB)	[5] (dB)	prop (dB)
-5.00	2.46	-2.52	-1.46	2.61	2.82
0.00	4.57	2.61	2.65	4.95	5.17
5.00	7.96	6.83	7.08	8.52	8.73
10.00	11.92	10.95	11.46	12.71	13.08
15.00	16.00	15.08	15.34	16.86	17.21

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

5. REFERENCES

[1] E. Grivel, M. Gabrea, and M. Najim, "Speech Enhancement as a Realisation Issue," *Signal Processing*, vol. 82, pp. 1963–1978, Dec. 2002.

[2] 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.

[3] 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.

[4] H. Morikawa and H. Fujisaki, "Noise reduction of speech signal by adaptive kalman filtering," *Special issue in Signal Processing APII-AFCET*, vol. 22, pp. 53–68, 1988.

[5] M. Gabrea, "Robust adaptive kalman filtering-based speech enhancement algorithm," in *Proc. ICASSP'04*, 2004.

[6] M. Gabrea and D. O'Shaugnessy, "Speech signal recovery in white noise using an adaptive kalman filter," in *Proc. EUSIPCO'00*, 2000, pp. 159–162.

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

[8] B. D. Kovacevic, M. M. Milosavljevic, and M. Dj. Veinovic, "Robust recursive ar speech analysis," *Signal Processing*, vol. 44, pp. 125–138, 1995.

[9] R. K. Mehra, "On the identification of variances and adaptive kalman filtering," *IEEE Trans. Automatic Control*, vol. AC-15, pp. 175–184, Apr. 1970.

[10] T. Kailath, "An innovations approach to least-squares estimation, part i: Linear filtering in additive white noise," *IEEE Tran. Automatic Control*, vol. AC-13, pp. 646–655, Dec. 1968.