Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 51(65), Fascicola 2, 2006

Single Microphone Noise Canceller Based on a Robust Adaptive Kalman Filter

Marcel Gabrea[1]

Abstract – This paper deals with the problem of speech enhancement when a corrupted speech signal with an additive 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: they first estimate the noise and the driving variances and parameters of the signal model, then estimate the speech signal. This paper presents an alternative solution that does not require the explicit estimation of noise and driving process variances. This deals with a new formulation of the steady-state Kalman filter gain estimation based on the use of external description of systems. Unlike the conventional approaches, no suboptimal Kalman filter is needed here.

Keywords: speech enhancement, Kalman filtering, noise reduction.

I. INTRODUCTION

Speech enhancement using a single microphone system has become an active research area for audio signal enhancement. The aim is to minimize the effect of noise and to improve the performance in voice communication systems when input signals are corrupted by background noise.

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, noise and driving process variances 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.

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. [3] have proposed the use of the EM 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.

Lee and Jung [4] have developed a time-domain approach, with no a priory information, to enhance speech signals. The autoregressive-hidden filter model (AR-HFM) with gain contour was proposed for modeling the statistical characteristics of the speech signal. The EM algorithm was used for signal estimation and system identification. In the E-step, the signal was estimated using multiple Kalman filters with Markovian switching coefficient and the probability was computed using the Viterbi Algorithm (VA). In M-step, the gain contour and noise parameter were recursively updated by an adaptive algorithm.

Grivel et al. [5] 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.

Gabrea and O'Shaughnessy [6] 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.

The methods proposed in [7] and [8] avoid 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.

In this paper a quite different and simple approach to the estimation of the steady-state optimal Kalman filter gain based on the use of external description of systems is presented. This method avoids the explicit estimation of noise and driving process variances by estimating the optimal Kalman gain. Unlike the conventional approaches, no suboptimal Kalman filter is needed here. Thus, the divergence problem of the

[1] École de technologie supérieure, Département de génie électrique,
1100 Notre-Dame Ouest, H3C 1K3 Montréal, e-mail mgabrea@ele.etsmtl.ca

Kalman filter does not occur. The performance of this algorithm is compared to the one of alternative speech enhancement algorithms based on the Kalman filtering. A distinct advantage of the proposed algorithm is that no voice activity detector (VAD) is required to estimate noise variance. Another advantage of this algorithm compared to [7] and [8] is the superiority in terms of computational load. An iterative procedure is not required in the steady-state optimal Kalman gain estimation.

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 and optimal Kalman gain. Simulation results are the subject of Section IV.

II. NOISY SPEECH MODEL AND KALMAN FILTERING

The speech signal $s(n)$ is modeled as a p^{th} order AR process:

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

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

where $s(n)$ is the n^{th} sample of the speech signal, $y(n)$ is the n^{th} sample of the observation, $a_i(n)$ is the i^{th} AR parameter, $u(n)$ and $v(n)$ are uncorrelated Gaussian white noise sequences with zero means and the variances $\sigma_u^2(n)$ and $\sigma_v^2(n)$.

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

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

$$y(n) = \boldsymbol{H}\boldsymbol{x}(n) + v(n) \quad (4)$$

where:

1. $\boldsymbol{x}(n) = [s(n-p+1) \quad \cdots \quad s(n)]^T$ is the $p \times 1$ state vector
2. $\boldsymbol{F}(n)$ is the $p \times p$ transition matrix

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

3. $\boldsymbol{H} = \boldsymbol{G}^T = [0 \quad 0 \quad \cdots \quad 0 \quad 1]$ is the $1 \times p$ observation row vector and the input vector.

The standard Kalman filter [9][10] provides the updating state-vector estimator equations:

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

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

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

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

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

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

where:

1. $\hat{\boldsymbol{x}}(n/n-1)$ is the minimum mean-squares estimate of the state vector $\boldsymbol{x}(n)$ given the past observations $y(1), \cdots, y(n-1)$
2. $\tilde{\boldsymbol{x}}(n/n-1) = \boldsymbol{x}(n) - \hat{\boldsymbol{x}}(n/n-1)$ is the predicted state-error vector
3. $\boldsymbol{P}(n/n-1) = E\left[\tilde{\boldsymbol{x}}(n/n-1)\tilde{\boldsymbol{x}}^T(n/n-1)\right]$ is the predicted state-error correlation matrix
4. $\hat{\boldsymbol{x}}(n/n)$ is the filtered estimate of the state vector $\boldsymbol{x}(n)$
5. $\tilde{\boldsymbol{x}}(n/n) = \boldsymbol{x}(n) - \hat{\boldsymbol{x}}(n/n)$ is the filtered state-error vector
6. $\boldsymbol{P}(n/n) = E\left[\tilde{\boldsymbol{x}}(n/n)\tilde{\boldsymbol{x}}^T(n/n)\right]$ is the filtered state-error correlation matrix
7. $e(n)$ is the innovation sequence
8. $\boldsymbol{K}(n)$ is the Kalman gain

The estimated speech signal can be retrieved as the p^{th} component of the state-vector estimator $\hat{\boldsymbol{x}}(n/n)$. However, the transition matrix and the driving process statistics are unknowns and hence must be estimated. Here a quite different and simple approach to the estimation of the steady-state optimal Kalman filter gain based on the use of external description of systems is used. This method avoids the explicit estimation of noise and driving process variances by estimating the optimal Kalman gain. In this case the Kalman filter equations are:

$$e(n) = y(n) - \boldsymbol{H}\hat{\boldsymbol{x}}(n/n-1) \quad (11)$$

$$\hat{\boldsymbol{x}}(n+1/n) = \boldsymbol{F}(n)\hat{\boldsymbol{x}}(n/n-1) + \boldsymbol{F}(n)\boldsymbol{K}^{opt}e(n) \quad (12)$$

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

$$\hat{s}(n) = \boldsymbol{H}\hat{\boldsymbol{x}}(n/n-1) + \boldsymbol{HK}^{opt}e(n) \quad (13)$$

The parameter estimation (the transition matrix and the optimal Kalman gain) is presented in the next section.

III. 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. [11]. The estimation of the steady-state optimal Kalman filter gain is based on the external description of the systems.

A. *Estimation of the Transition Matrix*

In our approach, getting $\boldsymbol{F}$ requires the AR parameter estimation. The equation (1) can be rewritten in the form:

$$s(n) = \boldsymbol{x}^T(n-1)\theta(n) + u(n) \quad (14)$$

where:

$$\theta(n) = \begin{bmatrix} a_p(n) & a_{p-1}(n) & \cdots & a_1(n) \end{bmatrix}^T \quad (15)$$

The robust recursive least square approach estimates the vector $\hat{\theta}(n)$ by minimizing the M-estimation criterion [11]:

$$J(n) = \frac{1}{n}\sum_{i=1}^{n}\lambda^{n-i}\rho[\varepsilon^2(i)] \quad (16)$$

where:

$$\psi(x) = \rho'(x) = \min\left[\frac{|x|}{\sigma_u^2(n)}, \frac{\Delta}{\sigma_u(n)}\right] \quad (17)$$

is the Huber influence function and Δ is a chosen constant. The true state vector $\boldsymbol{x}(n)$ used in (14) is unknown but can be approximated by the state-vector estimator $\hat{\boldsymbol{x}}(n/n)$. In this case the robust recursive least square approach gives the estimation equations:

$$\varepsilon(i) = \boldsymbol{H}\hat{\boldsymbol{x}}(i/i) - \hat{\boldsymbol{x}}^T(i-1/i-1)\theta(i) \quad (18)$$

$$\boldsymbol{g}(i) = \frac{\boldsymbol{Q}(i-1)\hat{\boldsymbol{x}}(i-1/i-1)}{\lambda(i) + \psi'[\varepsilon(i)]\hat{\boldsymbol{x}}^T(i-1/i-1)\boldsymbol{Q}(i-1)\hat{\boldsymbol{x}}(i-1/i-1)} \quad (19)$$

$$\boldsymbol{Q}(i) = \frac{1}{\lambda(i)}\left[\boldsymbol{Q}(i-1) - \boldsymbol{g}(i)\hat{\boldsymbol{x}}^T(i-1/i-1)\boldsymbol{Q}(i-1)\psi'[\varepsilon(i)]\right] \quad (20)$$

$$\hat{\theta}(i) = \hat{\theta}(i-1) + \boldsymbol{Q}(i)\hat{\boldsymbol{x}}(i-1/i-1)\psi[\varepsilon(i) \quad (21)$$

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 $\lambda_{\min}$ and $\lambda_{\max}$. The modified generalized likelihood ratio algorithm [12] 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 $[i\text{-}N+1, i]$, $[i+1, i+N]$ and $[i\text{-}N+1, i+N]$, and move one sample forward with each new sample. In the first step of this algorithm is calculated the discrimination function:

$$\begin{aligned} D(i,N) &= L(i-N+1, i+N) - L(i-N+1, i) \\ &\quad - L(i+1, i+N) \end{aligned} \quad (22)$$

where:

$$L(a,b) = (b-a+1)\ln\left[\frac{1}{b-a+1}\sum_{i=a}^{b}\varepsilon^2(i)\right] \quad (23)$$

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.

B. *Steady-State Optimal Kalman Gain Estimation*

The Kalman filter always requires the knowledge of noise variances. When they are unknown, we must estimate them with some methods or we must estimate the steady-state optimal Kalman filter gain $\boldsymbol{K}^{opt} = \lim_{n\to\infty}\boldsymbol{K}(n)$ directly from the output data. Let $f(z) = z^m + \alpha_1 z^{m-1} + \cdots + \alpha_m$ be the minimal polynomial of the matrix $\boldsymbol{F}$ with $f(\boldsymbol{F}) = 0$.
From (11) and (12) in the steady-state:

$$y(n-m+i) = \boldsymbol{HF}^{i}\hat{\boldsymbol{x}}(n-m/n-m-1) + \sum_{j=0}^{i-1}\boldsymbol{HF}^{i-j}\boldsymbol{K}^{opt}e(n-m+i) + e(n-m+i) \quad (24)$$

and multiplying (24) by α_{m-i} ($\alpha_0 = 1$) and summing for $i = 0, 1, \cdots, m$ we obtain:

$$\sum_{i=0}^{m}\alpha_{m-i}y(n-m+i) = \sum_{i=0}^{m}\alpha_{m-i}\sum_{j=0}^{i-1}\boldsymbol{HF}^{i-j}\boldsymbol{K}^{opt}e(n-m+j) + \sum_{i=0}^{m}\alpha_{m-i}e(n-m+i) \quad (25)$$

or:

$$y(n) + \sum_{i=1}^{m}\alpha_i y(n-i) = e(n) + \sum_{i=0}^{m}\beta_i e(n-i) \quad (26)$$

where:

$$\beta_i = \alpha_i + \sum_{j=0}^{i-1}\alpha_j\boldsymbol{HF}^{i-j}\boldsymbol{K}^{opt},\ i = 1, \cdots, m \quad (27)$$

We can obtain the optimal gain $\boldsymbol{K}^{opt}$ by solving (26) with the knowledge of β_i for $i = 1, \cdots, m$. Define:

$$\xi(n) = e(n) + \sum_{i=1}^{m}\beta_i e(n-i) \quad (28)$$

It is known that in the optimal case the innovation process $e(n)$ is orthogonal to all past observations $y(1), \cdots, y(n-1)$ and it consists of a sequence of random variables that are orthogonal to each other. In this case the autocorrelation of the innovation process $r_e(k) = E[e(n)e(n-k)]$ is zero for $k > 0$ [13]. From (28) for $k = 0, 1, \cdots, m$ we obtain $r_\xi(k)$ the autocorrelation of $\xi(n)$, $r_\xi(k) = E[\xi(n)\xi(n-k)]$ as:

$$r_\xi(k) = r_e(0)\sum_{i=k}^{m}\beta_i\beta_{i-k},\ \beta_0 = 1 \quad (29)$$

The equations (28) can be solved for β_i, $i = 1, \cdots, m$ and $r_e(0)$ by using the estimate of the autocorrelation $\hat{r}_\xi(k)$:

$$r_\xi(k) = \frac{1}{N}\sum_{i=1}^{N}\xi(i)\xi(i-k) \quad (30)$$

where N is the sample size and is given by :

$$\xi(n) = y(n) + \sum_{i=1}^{m}\alpha_i y(n-i) \quad (31)$$

Now from (27) the estimate of the optimal gain $\hat{\boldsymbol{K}}^{opt}$ is given by:

$$\hat{\boldsymbol{K}}^{opt} = \begin{bmatrix} \boldsymbol{HF} \\ \vdots \\ \sum_{j=0}^{m-1}\alpha_j\boldsymbol{HF}^{m-j} \end{bmatrix}^{\dagger} \begin{bmatrix} \beta_1 - \alpha_1 \\ \vdots \\ \beta_m - \alpha_m \end{bmatrix} \quad (32)$$

IV. SIMULATION RESULTS

The proposed method was first tested using an AR signal that offers a good approximation of the spectral envelope of a speech signal and an additive Gaussian white noise. In the experiment, 256 samples of the AR signal were generated. In Table 1 we present the mean value, the standard deviation and the maximum value based on 1000 simulations.

Table 1

Input SNR (dB)	Output SNR (dB)		
	Mean	Std	Max
-5.00	2.93	0.48	4.46
0.00	5.72	0.29	7.33
5.00	9.82	0.21	11.27
10.00	12.71	0.15	13.72
15.00	17.08	0.07	17.31

The approach was also tested using a speech signal and additive noise. The speech signals are sentences from the TIMIT database. Table 2 offers a comparison with others approaches, by showing averaged SNR gain based on 10 speech signals and 10 noise simulations for each speech signal. Figures 2, 3 and 4 represent, respectively, the time signal followed by the spectrogram of the free-noise speech, the noisy speech and the enhanced speech. For this example, the SNR of the noisy speech signal is 0 dB.

Table 2

Input SNR (dB)	Output SNR (dB)			
	[14]	[7]	[8]	Prop.
-5.00	2.46	-2.52	2.48	2.56
0.00	4.57	2.61	4.72	4.88
5.00	7.96	6.83	8.29	8.37
10.00	11.92	10.95	12.31	12.48
15.00	16.00	15.08	16.47	16.76

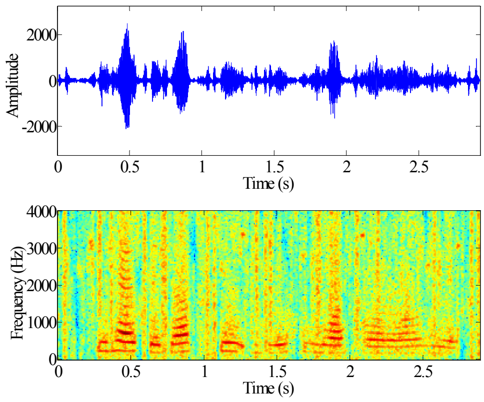

Fig. 1: Noise-free speech signal

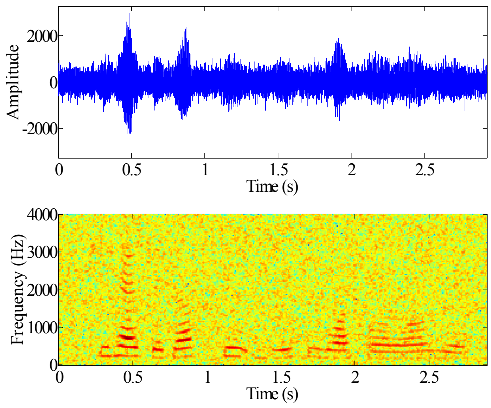

Fig. 2: Noisy speech signal

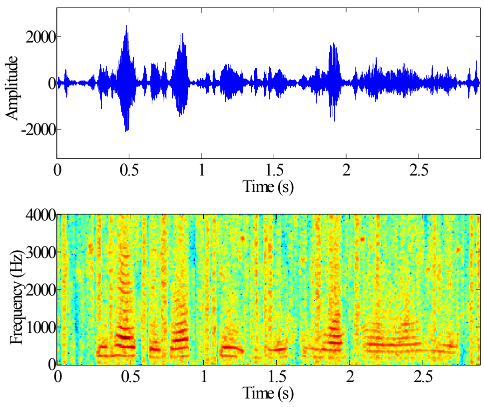

Fig. 3: Enhanced speech signal

Compared to the methods similar in structure previously proposed by the author in [7] and in [8] and to the Gibson's algorithm [14], 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. It uses a voice activity detector to determine silence periods. The above factors lead to computational requirements higher than those corresponding to the proposed approach.

REFERENCES

[1] K. K. Paliwal and A. Basu, "A Speech Enhancement Method Based on Kalman Filtering," in *Proc. ICASSP'87*, pp. 177-180, 1988.

[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] 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.

[4] K. Y. Lee and S. Jung, "Time-Domain Approach Using Multiple Kalman Filters and EM Algorithm to Speech Enhancement with Nonstationary Noise", *IEEE Trans. Speech and Audio Processing*, vol.8, pp. 282-291, May 2000.

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

[6] M. Gabrea and D. O'Shaugnessy, "Speech Signal Recovery in White Noise Using an Adaptive Kalman Filter," in *Proc. EUSIPCO'00*, 2000.

[7] M. Gabrea, E. Grivel and M. Najim, "Single Microphone Kalman Filter-Based Noise Canceller", *IEEE Signal Processing Lett.*, vol. 6, pp. 55-57, Mar. 1999.

[8] M. Gabrea, "Adaptive Kalman Filtering-Based Speech Enhancement Algorithm", in *Proc. IWAENC'01*, pp. 207-210, 2001.

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

[10] M. Najim, Modelization and Identification in Signal Processing, France:Masson, Paris, 1988.

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

[12] 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*, 1988.

[13] 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.

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