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 a corrupted speech signal with an additive Gaussian white 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 optimal Kalman gain estimation approach proposed in the control literature by Carew and Belanger.

1. 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.

Various approaches based on the Kalman filter are presented 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.

Gibson et al. [2] have proposed a method that provides a sub-optimal 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).

Gannot et al. [3] have proposed the use of 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.

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.

Lee and Jung [5] 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.

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.

In this paper the signal is modeled as an AR process and a Kalman filter based-method is proposed by reformulating and adapting the approach proposed for control applications by Carew and Belanger [7]. This method 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.

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 a VAD is not required. Another advantage of this algorithm compared to the one, similar in structure, presented in [8], is the superiority in terms of computational load. A filtering step is not required in the optimal Kalman gain estimation.

This paper is organized as follows. In Section II we present the speech enhancement approach based on the Kal-man filter algorithm. Section III is concerned with the estimation of AR parameters and optimal Kalman gain. 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:

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

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

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

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

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

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

where:

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

2. $\mathbf{x}(n)$ is the $p \times 1$ state vector

$$\mathbf{x}(n) = [s(n-p+1) \cdots s(n)]^T \tag{5}$$

3. $\mathbf{F}$ is the $p \times p$ transition matrix

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ a_p & a_{p-1} & a_{p-2} & \cdots & a_1 \end{bmatrix} \tag{6}$$

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

$$\mathbf{H} = \mathbf{G}^T = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \tag{7}$$

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

$$e(n) = y(n) - \mathbf{H}\hat{\mathbf{x}}(n/n-1) \tag{8}$$

$$\hat{\mathbf{x}}(n/n) = \hat{\mathbf{x}}(n/n-1) + \mathbf{K}(n)e(n) \tag{9}$$

$$\hat{\mathbf{x}}(n+1/n) = \mathbf{F}\hat{\mathbf{x}}(n/n) \tag{10}$$

where $\hat{\mathbf{x}}(n/n-1)$ is the minimum mean-square estimate of the state vector $\mathbf{x}(n)$ given the past observations $y(1), \ldots, y(n-1)$, $\hat{\mathbf{x}}(n/n)$ is the filtered estimate of the state vector $\mathbf{x}(n)$, $e(n)$ is the innovation sequence and $\mathbf{K}(n)$ 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 noise variances σ_u^2 and σ_v^2 are needed to compute the Kalman gain $\mathbf{K}(n)$. However, the transition matrix and the Kalman gain are unknown and hence must be estimated. The parameter estimation (the transition matrix and the optimal Kalman gain) 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 the modified Yule-Walker equations. The estimation of the optimal Kalman gain is derived using the property of the innovation sequence, obtained after a preliminary Kalman filtering with an initial gain.

3.1. Estimation of the Transition Matrix

In our approach, getting $\mathbf{F}$ requires the AR parameter estimation. This issue being outside the scope of the present paper we propose to estimate the AR parameters from modified Yule-Walker equations [11], even if this approach may sometimes lead to unsatisfactory performances, especially for wideband signals [12]:

$$\begin{bmatrix} \hat{a}_p \\ \vdots \\ \hat{a}_1 \end{bmatrix} = \begin{bmatrix} r_{yy}(1) & \cdots & r_{yy}(p) \\ \vdots & \ddots & \vdots \\ r_{yy}(p+l) & \cdots & r_{yy}(2p+l-1) \end{bmatrix}^{\dagger} \times$$
$$\times \begin{bmatrix} r_{yy}(p+1) \\ \vdots \\ r_{yy}(2p+l) \end{bmatrix} \tag{12}$$

where $r_{yy}(k) = E[y(n)y(n-k)]$ denotes the observation autocorrelation function, $E[.]$ denotes the expectation, $[.]^{\dagger}$ denotes the pseudoinverse operator and $l \geq 0$.

3.2. Estimation of the Optimal Kalman Gain

It is known that in the optimal case the innovation process $e(n)$ is orthogonal to all past observations $y(1)$, $\ldots$, $y(n-1)$ and it consists of a sequence of random variables that are orthogonal to each other. In this case the autocorrelation of the innovation process $r_{ee}(k) = E[e(n)e(n-k)]$ is zero for $k > 0$ [13].

Let a sub-optimal Kalman gain $\mathbf{K}^*$, $\mathbf{x}^*(n/n-1)$ the estimate of the state vector $\mathbf{x}(n)$ given the past observation $y(1)$, $\ldots$, $y(n-1)$ and $e^*(n) = y(n) - \mathbf{H}\mathbf{x}^*(n/n-1)$ the innovation sequence obtained using the sub-optimal Kalman gain $\mathbf{K}^*$. In this case the innovation sequence $e^*(n)$ is not a white process and $r^*_{ee}(k) = E[e^*(n)e^*(n-k)]$ is not zero for $k > 0$. Let define the difference estimate vector $\tilde{\mathbf{x}}(n/n-1)$ obtained by difference between the estimates of $\mathbf{x}(n)$ using the optimal and respectively the sub-optimal Kalman gain and $\mathbf{M}(n/n-1) = E[\tilde{\mathbf{x}}(n/n-1)\tilde{\mathbf{x}}^T(n/n-1)]$ the difference estimate correlation matrix.

In the steady-state $\mathbf{K}(n) \simeq \mathbf{K}$ and $\mathbf{M}(n/n-1) \simeq \mathbf{M}$. Using the estimation transition matrix $\hat{\mathbf{F}}$, the standard Kalman filter equations (8)-(10) and the state space model equations (3)(4), the innovation autocorrelation function $r^*_{ee}(k)$ is computed as:

$$
\begin{aligned}
r^*_{ee}(k) &= \mathbf{H}[\hat{\mathbf{F}}(\mathbf{I}-\mathbf{K}^*\mathbf{H})]^{k-1}\hat{\mathbf{F}} \times \\
&\times \;\; [(\mathbf{I}-\mathbf{K}^*\mathbf{H})\mathbf{M}\mathbf{H}^T + (\mathbf{K}-\mathbf{K}^*)r_{ee}(0)] \\
&\qquad k > 0 \qquad (13)
\end{aligned}
$$

$$
r^*_{ee}(0) = \mathbf{H}\mathbf{M}\mathbf{H}^T + r_{ee}(0) \qquad (14)
$$

Using the equation (13), the optimal steady-state Kalman gain is obtained as:

$$
\mathbf{K} = \mathbf{K}^* - (\mathbf{I}-\mathbf{K}^*\mathbf{H})\mathbf{M}\mathbf{H}^T/r_{ee}(0) +
\begin{bmatrix} \mathbf{H}\hat{\mathbf{F}} \\ \mathbf{H}[\hat{\mathbf{F}}(\mathbf{I}-\mathbf{K}^*\mathbf{H})]^1\hat{\mathbf{F}} \\ \vdots \\ \mathbf{H}[\hat{\mathbf{F}}(\mathbf{I}-\mathbf{K}^*\mathbf{H})]^{p-1}\hat{\mathbf{F}} \end{bmatrix}^{-1}
\begin{bmatrix} r^*_{ee}(1) \\ r^*_{ee}(2) \\ \vdots \\ r^*_{ee}(p) \end{bmatrix} / r_{ee}(0) \qquad (15)
$$

Using the steady-state Riccati equation [9] the difference estimate correlation matrix is obtained as:

$$
\begin{aligned}
\mathbf{M} &= \hat{\mathbf{F}}(\mathbf{I}-\mathbf{K}^*\mathbf{H})\mathbf{M}(\mathbf{I}-\mathbf{K}^*\mathbf{H})^T\hat{\mathbf{F}}^T \\
&+ \;\; r_{ee}(0)(\mathbf{K}-\mathbf{K}^*)(\mathbf{K}-\mathbf{K}^*)^T \qquad (16)
\end{aligned}
$$

Carew and Belanger [7] have proposed an iterative method to solve the equations (14)(15)(16) in terms of $r_{ee}(0)$, $\mathbf{K}$ and $\mathbf{M}$. Adapting this method in our case, in the first iteration we start with $\mathbf{M}^{(0)}$ the initial value of $\mathbf{M}$, compute the first estimates of innovation autocorrelation $r^{(0)}_{ee}(0)$ using (14) and compute the first estimate of the optimal Kalman gain $\mathbf{K}^{(0)}$ using (15). After i iterations the estimates of $r_{ee}(0)$, $\mathbf{K}$ and $\mathbf{M}$ are:

$$
r^{(i)}_{ee}(0) = r^*_{ee}(0) - \mathbf{H}\mathbf{M}^{(0)}\mathbf{H}^T \qquad (17)
$$

$$
\begin{aligned}
\mathbf{K}^{(i)} &= \mathbf{K}^* - (\mathbf{I}-\mathbf{K}^*\mathbf{H})\mathbf{M}\mathbf{H}^T/r_{ee}(0) + \\
&+ \begin{bmatrix} \mathbf{H}\hat{\mathbf{F}} \\ \mathbf{H}[\hat{\mathbf{F}}(\mathbf{I}-\mathbf{K}^*\mathbf{H})]^1\hat{\mathbf{F}} \\ \vdots \\ \mathbf{H}[\hat{\mathbf{F}}(\mathbf{I}-\mathbf{K}^*\mathbf{H})]^{p-1}\hat{\mathbf{F}} \end{bmatrix}^{-1} \times \\
&\times \begin{bmatrix} r^*_{ee}(1) \\ r^*_{ee}(2) \\ \vdots \\ r^*_{ee}(p) \end{bmatrix} / r_{ee}(0) \qquad (18)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{M}^{(i+1)} &= \hat{\mathbf{F}}(\mathbf{I}-\mathbf{K}^*\mathbf{H})\mathbf{M}^{(i)}(\mathbf{I}-\mathbf{K}^*\mathbf{H})^T\hat{\mathbf{F}}^T + \\
&+ \;\; r^{(i)}_{ee}(0)(\mathbf{K}-\mathbf{K}^*)(\mathbf{K}-\mathbf{K}^*)^T \qquad (19)
\end{aligned}
$$

4. 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.

Input SNR (dB)	Output SNR Mean (dB)	Std (dB)	Max (dB)
-5.00	2.73	0.58	4.42
0.00	5.63	0.31	7.23
5.00	9.70	0.23	11.21
10.00	12.53	0.17	13.63
15.00	16.96	0.08	17.23

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

The approach was also tested using a speech signal and an additive Gaussian white 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.

Figure 1 represents, respectively, the time signal of the noise-free speech, the noisy speech and the enhanced speech. For this example, the SNR of the noisy speech signal is 0 dB.

Compared to the method similar in structure previously proposed by the author in [8] and to the Gibson's algorithm [2], the proposed method provides increases in SNR, as well as improved speech quality and intelligibility for input SNR between -5 and 15 dB. Gib-

Input SNR (dB)	Output SNR [2] (dB)	[8] (dB)	proposed (dB)
-5.00	2.46	-2.52	2.48
0.00	4.57	2.61	4.72
5.00	7.96	6.83	8.29
10.00	11.92	10.95	12.31
15.00	16.00	15.08	16.47

Table 2: OUTPUT SNR FOR AN INPUT SPEECH SIGNAL PLUS WHITE NOISE

son'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.

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] 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.

[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] E. Grivel, M. Gabrea, and M. Najim, "Subspace State Space Model Identification for Speech Enhancement," in *Proc. ICASSP'99*, pp. 781–784.

[5] 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.

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

[7] B. Carew and P. R. Belanger, "Identification of Optimum Filter Steady-State Gain for Systems with Unknown Noise Covariances," *IEEE Tran. Automatic Control*, vol. AC-18, pp. 582–587, Dec. 1973.

[8] 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.

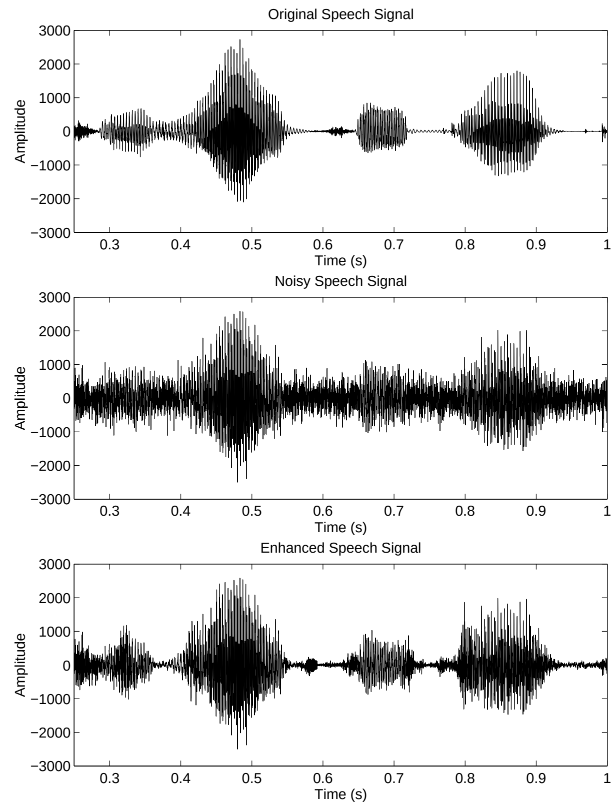

Figure 1: Example of speech signal enhancement (Input SNR = 0 dB)

[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] S. M. Kay, *Modern Spectral Estimation: Theory and Applications*, NJ:Prentice-Hall, Englewood Cliffs, 1988.

[12] K. K. Paliwal, "A Noise-compensated Long Correlation Matching Method for AR Spectral Estimation of Noisy Signals," in *Proc. ICASSP'86*, pp. 1369–1372.

[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.