# ADAPTIVE KALMAN FILTER FOR SPEECH ENHANCEMENT FROM COLORED NOISE

M. Gabrea(*), E. Mandridake and M. Najim
Equipe Signal et Image, ENSERB
BP 99, 33 402 Talence, France
(*)Technical University of Timisoara
Electronics Faculty
V. Parvan 2, 1900 Timisoara, Romania
email: marcel@goelette.tsi.u-bordeaux.fr

## ABSTRACT

In the framework of speech enhancement we propose a new approach for signal recovering in colored noise based on adaptive Kalman filter. The approaches proposed in the past, in this context, operate in two steps: they first estimate the noises variances and the parameters of the signal and noise models and secondly estimate the speech signal. In this paper we propose a new parameters estimation method based on the EM (Expectation-Maximisation) algorithm.

## 1 INTRODUCTION

The problem we are dealing with is to restore a speech signal corrupted by a noise when only one noisy observation is available. Many approaches based on the Kalman filter [1-5] for speech enhancement have been reported in the literature.

In [1][2][3] and [5] the noise under a simplified assumption is considered as an white Gaussian process. But in [4] the noise is considered colored and modelled as an $AR$ process.

In [1] a time-adaptive algorithm is used to adaptively estimate the speech model parameters and the noise variance. The estimation method of the speech model parameters used in [4] is a suboptimal solution of the maximum likelihood argument.

In this paper we propose a new approach in the case of a speech signal corrupted by an additive colored noise. Signal and noise are modelled as $AR$ processes. The coefficients of the $AR$ processes and the $AR$ driving processes variances are estimated based on $EM$ (Expectation-Maximisation) algorithm. We extend for the colored noise the method proposed by Deriche [6] to estimate the $AR$ parameters of the signal corrupted by white noise.

This paper is organised as follows. We present in section 2 the speech enhancement approach based on the Kalman filter algorithm. The section 3 is concerned with the presentation of the estimation parameter techniques. In section 4 we provide experimental results and evaluate the performance of the proposed approach.

## 2 SPEECH ENHANCEMENT BASED ON THE KALMAN FILTERING

Let us consider an observed signal $z(n) = s(n) + b(n)$ where the speech signal $s(n)$ and the noise $b(n)$ are modelled respectively as a $p$ and $q$ order $AR$ processes generated respectively by $u(n)$ and $b(n)$, uncorrelated Gaussian zero mean white noises with variances $\sigma_u^2$ and $\sigma_v^2$.

This system can be represented by the following state-space model[4]:

$$\mathbf{x}(n+1) = \Phi\mathbf{x}(n) + \Gamma\mathbf{w}(n+1) \tag{1}$$

$$z(n) = s(n) + b(n) = \mathbf{H}\mathbf{x}(n) \tag{2}$$

where

$$\mathbf{x}(n) = [s(n-p+1), \cdots, s(n), b(n-q+1), \cdots, b(n)]^T \tag{3}$$

is the state-vector and

$$\mathbf{w}(n) = [u(n) \quad v(n)]^T \tag{4}$$

$$\Phi = \begin{bmatrix} \Phi_s & 0 \\ 0 & \Phi_b \end{bmatrix} \tag{5}$$

is the state-transition matrix. It is an extended matrix containing two submatrices $\Phi_s$ and $\Phi_b$ related to the model's parameter of the signal and the noise as follow:

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

$$\Phi_b = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ c_p & c_{p-1} & \cdots & c_1 \end{bmatrix} \tag{7}$$

$\Gamma$ and $\mathbf{H}$ are the input and the output matrices defined by :

$$\Gamma = \begin{bmatrix} 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{8}$$

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{9}$$

The standard Kalman filter provides the following updating state-vector estimation [7]:

$$\widehat{\mathbf{x}}(n+1/n) = \Phi\widehat{\mathbf{x}}(n/n-1) + \Phi\mathbf{K}(n)e(n) \tag{10}$$

$$e(n) = z(n) - \mathbf{H}\widehat{\mathbf{x}}(n/n-1) \tag{11}$$

where $e(n)$ is the innovation sequence and $\mathbf{K}(n)$ the Kalman gain.

$\Phi$ and $\mathbf{K}(n)$ are unknown and hence will be respectively estimated by $\widehat{\Phi}$ and $\widehat{\mathbf{K}}(n)$.

Then, the updating state vector estimation becomes:

$$\widehat{\mathbf{x}}(n+1/n) = \widehat{\Phi}\widehat{\mathbf{x}}(n/n-1) + \widehat{\Phi}\widehat{\mathbf{K}}(n)e(n) \tag{12}$$

The estimated speech signal is a $p$ component of the state-space vector :

$$\widehat{\mathbf{x}}(n/n) = \widehat{\Phi}\widehat{\mathbf{x}}(n/n-1) \tag{13}$$

In the Kalman filter literature the case where $z(n) = \mathbf{H}(n)\mathbf{x}(n)$ is called "noise free observation"[7]. In this case the application of the standard Kalman falls into a singular problem. Among the solutions proposed to overcome such a singularity we retain the coordinate transformation suggested in Maybeck [8].

## 3 PARAMETER ESTIMATION

Let use denote by $\mathbf{s}_N(n)$, $\mathbf{b}_N(n)$ and $\mathbf{z}_N(n)$ the $N$x1 vectors made of the samples of $s(n)$, $b(n)$ and $z(n)$. We will consider that $s(n)$ and $b(n)$ are statisticaly independent.

The "Expectation" step at iteration $m$ consists of computing the following functions:

$$\begin{aligned} Q_s &= E[\log\left\{p\left(\mathbf{s}_N(n); a_i, \sigma_u^2\right)\right\} \\ &\quad /z_N(n), a_i^{(m)}, \sigma_u^{2(m)}, c_j^{(m)}, \sigma_v^{2(m)}] \end{aligned} \tag{14}$$

$$\begin{aligned} Q_b &= E[\log\left\{p(\mathbf{b}_N(n); c_i, \sigma_v^2)\right\} \\ &\quad /\mathbf{z}_N(n), a_i^{(m)}, \sigma_u^{2(m)}, c_j^{(m)}, \sigma_v^{2(m)}] \end{aligned} \tag{15}$$

Next $Q_s$ and $Q_b$ are maximized with respect to $a_i, \sigma_u, c_j, \sigma_v$ to yield a new estimate $a_i^{(m+1)}, \sigma_u^{2(m+1)}, c_j^{(m+1)}, \sigma_v^{2(m+1)}$. This is the "Maximization" step.

The density functions of $\mathbf{s}_N(n)$ and $\mathbf{b}_N(n)$ are given by :

$$\begin{aligned} p\left(\mathbf{s}_N(n); a_i, \sigma_u^2\right) &= (2\pi)^{-N/2}\left|\mathbf{R}_s(a_i, \sigma_u^2)\right|^{-1/2} \\ &\exp\left\{-\frac{1}{2}\mathbf{s}_N^T(n)\mathbf{R}_s^{-1}(a_i, \sigma_u^2)\mathbf{s}_N(n)\right\} \end{aligned} \tag{16}$$

$$\begin{aligned} p\left(\mathbf{b}_N(n); c_i, \sigma_v^2\right) &= (2\pi)^{-N/2}\left|\mathbf{R}_b(c_i, \sigma_v^2)\right|^{-1/2} \\ &\exp\left\{-\frac{1}{2}\mathbf{b}_N^T(n)\mathbf{R}_b^{-1}(c_i, \sigma_v^2)\mathbf{b}_N(n)\right\} \end{aligned} \tag{17}$$

The conditional means and covariances of $\mathbf{s}_N(n)/\mathbf{z}_N(n)$ and $\mathbf{b}_N(n)/\mathbf{z}_N(n)$ are given by :

$$\begin{aligned} \mu_{s/z}(a_i, \sigma_u^2, c_j, \sigma_v^2) &= \mathbf{R}_s(a_i, \sigma_u^2) \\ &\left\{\mathbf{R}_s(a_i, \sigma_u^2) + \mathbf{R}_b(c_j, \sigma_v^2)\right\}^{-1}\mathbf{z}_N(n) \end{aligned} \tag{18}$$

$$\begin{aligned} \mu_{b/z}(a_i, \sigma_u^2, c_j, \sigma_v^2) &= \mathbf{R}_b(c_j, \sigma_v^2) \\ &\left\{\mathbf{R}_s(a_i, \sigma_u^2) + \mathbf{R}_b(c_j, \sigma_v^2)\right\}^{-1}\mathbf{z}_N(n) \end{aligned} \tag{19}$$

$$\begin{aligned} \mathbf{R}_{s/z}(a_i, \sigma_u^2, c_j, \sigma_v^2) &= \mathbf{R}_s(a_i, \sigma_u^2) + \\ \left\{\mathbf{R}_s(a_i, \sigma_u^2) + \mathbf{R}_b(c_j, \sigma_v^2)\right\}^{-1} &\mathbf{R}_s(a_i, \sigma_u^2) \end{aligned} \tag{20}$$

$$\begin{aligned} \mathbf{R}_{b/z}(a_i, \sigma_u^2, c_j, \sigma_v^2) &= \mathbf{R}_b(c_j, \sigma_v^2) + \\ \left\{\mathbf{R}_s(a_i, \sigma_u^2) + \mathbf{R}_b(c_j, \sigma_v^2)\right\}^{-1} &\mathbf{R}_b(c_j, \sigma_v^2) \end{aligned} \tag{21}$$

Using (16) - (21) we can rewrite $Q_s$ and $Q_b$ by neglecting the constants:

$$\begin{aligned} Q_s &= -\frac{1}{2}\log\left|\mathbf{R}_s(a_i, \sigma_u^2)\right| \\ &\quad -\frac{1}{2}Trace\left\{\mathbf{R}_s^{-1}(a_i, \sigma_u^2)\mathbf{G}_s^m\right\} \end{aligned} \tag{22}$$

$$\begin{aligned} Q_b &= -\frac{1}{2}\log\left|\mathbf{R}_b(c_j, \sigma_v^2)\right| \\ &\quad -\frac{1}{2}Trace\left\{\mathbf{R}_b^{-1}(c_i, \sigma_v^2)\mathbf{G}_b^m\right\} \end{aligned} \tag{23}$$

where:

$$\begin{aligned} \mathbf{G}_s^{(m)} &= \mathbf{R}_{s/z}(a_i^{(m)}, \sigma_u^{2(m)}, c_j^{(m)}, \sigma_v^{2(m)}) + \\ &\quad \mu_{s/z}^T(a_i^{(m)}, \sigma_u^{2(m)}, c_j^{(m)}, \sigma_v^{2(m)}) \\ &\quad \mu_{s/z}(a_i^{(m)}, \sigma_u^{2(m)}, c_j^{(m)}, \sigma_v^{2(m)}) \end{aligned} \tag{24}$$

$$\begin{aligned} \mathbf{G}_b^{(m)} &= \mathbf{R}_{b/z}(a_i^{(m)}, \sigma_u^{2(m)}, c_j^{(m)}, \sigma_v^{2(m)}) + \\ &\quad \mu_{b/z}^T(a_i^{(m)}, \sigma_u^{2(m)}, c_j^{(m)}, \sigma_v^{2(m)}) \\ &\quad \mu_{b/z}(a_i^{(m)}, \sigma_u^{2(m)}, c_j^{(m)}, \sigma_v^{2(m)}) \end{aligned} \tag{25}$$

## 4 SIMULATIONS AND RESULTS

The method has been exercised to natural speech signal corrupted by a noise captured in a car (database of the Matra Company). The order $p$ of the $AR$ process of speech signal has been fixed to 10.

An example of speech enhancement results, using this algorithm, is reported in the Table 2. A SNR improvement from 1.45dB to 8.01dB has been obtained when $p = 10$ and $q = 10$ (case 1) for an input SNR varying from -10 to 10dB. If we consider the car noise as an white process (case 2) the SNR improvement decreases from 0.83 to 2.14dB (depending on SNR of the input) comparatively with the case where the noise is considered colored.

| Input SNR(dB) | SNR improvement (dB) | |
|---|---|---|
| | case 1 | case 2 |
| -10 | 8.01 | 5.87 |
| -5 | 5.89 | 4.08 |
| 0 | 4.06 | 3.19 |
| 5 | 2.60 | 1.53 |
| 10 | 1.53 | 0.63 |

Table 1: Output SNR improvement for different input SNR

Figures 1, 2 and 3 represent respectively the noisy speech, the enhanced speech signal and the original speech signal with theirs spectrograms. For this example, $p = 10$, $q = 10$ and the SNR of the noisy speech signal is 0dB.

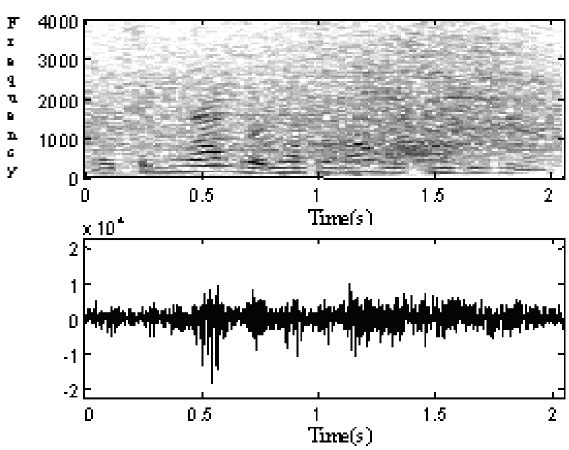


Figure 1: Noisy speech signal

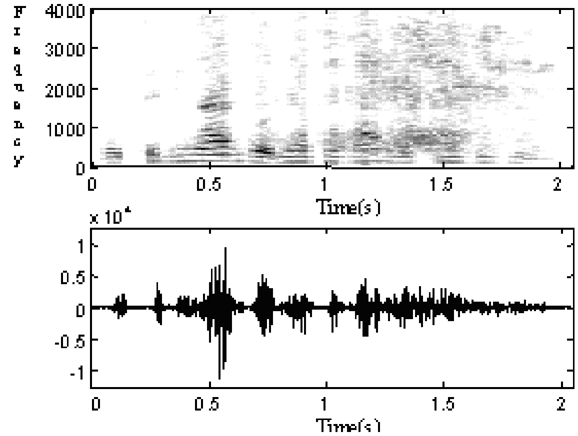


Figure 2: Enhanced speech signal

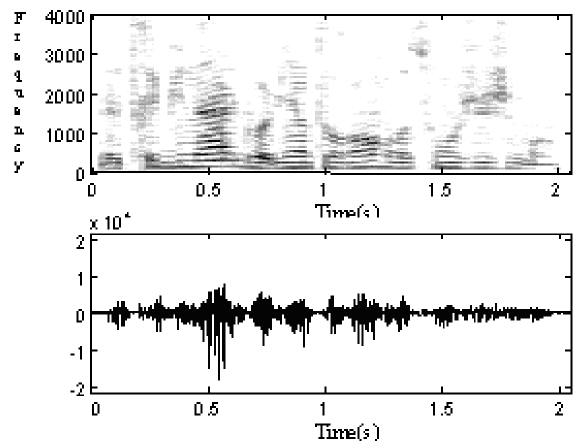


Figure 3: Original speech signal

**Acknowlegment**
M. Gabrea is grateful towards the Ministère des Affaires Etrangères (France).

## References

[1] A.V. Oppenheim, E. Weinstein, K.C. Zangi, M. Feder and D. Gauger, "Single-Sensor Active Noise Cancellation," IEEE Trans. on Speech and Audio Processing, Vol. 2,No. 2, pp. 285-290, Apr. 1994.

[2] K.K. Paliwal and A.Basu, "A Speech Enhancement Method Based on Kalman Filtering," ICASSP 87, pp. 177-180.

[3] H. Morikawa and H. Fujisaki, "Noise Reduction of Speech Signal by Adaptive Kalman Filtering," Special issue on Model Based in Signal Processing APII-AFCET-Journal, Ed. by M. Najim, Vol. 22, No. 1, pp. 53-68, 1988.

[4] J.D. Gibson, B. Koo and S.D. Gray, " Filtering of Colored Noise for Speech Enhancement and Coding," IEEE Trans. on Sig. Proc., Vol. 39, No.8, pp. 1732-1742, Aug. 1991.

[5] M. Gabrea, E. Mandridake and M. Najim, "A single sensor noise canceller based on the adaptive Kalman filter," EUSIPCO'96, Vol. II, pp. 979-982, Sep. 1996.

[6] M. Deriche, " AR Parameter Estimation from Noisy Data using the EM Algorithm ," ICASSP 94, vol. IV, pp. 69-73.

[7] M. Najim, Modelization and Identifications in Signal Processing, Masson, Paris 1988 (in French).

[8] P.S. Maybeck, Stochastic Models, Estimation, and Control, Volume 1, Academic Press, 1979.