PERFORMANCE OF THE EXTENDED KALMAN FILTER FOR RESTORATION OF AUDIO DOCUMENTS

G. De Poli, G.A. Mian, G. Re
Dipartimento di Elettronica e Informatica
Via Gradenigo 6/a, 35100 Padova (IT)

Abstract

The problem of removing impulsive and background (white or coloured) noise from audio recordings is considered. The algorithm used simultaneously solves the problems of wideband noise filtering, signal parameter tracking and impulsive noise elimination by using the Extended Kalman Filter theory (EKF), as proposed by M. Niedzwiecki and K. Cisowski [5, 6]. Results, obtained with the proposed method for significant cases, are presented. Moreover, features and performance of the method are compared with other existing techniques.

1 Introduction

The introduction of high quality digital media, combined with an increasing awareness of the historical importance of “audio heritage”, has led to a growing requirement for the preservation and restoration of old recordings [7]. In this work we present some results on the restoration of magnetic tapes and vinyl records, carried out within the project ” Beni Culturali ” of the CNR [1], whose aim is the preservation and fruition of all Italian cultural assets.

The types of degradation common in audio sources can be broadly classified into localized and global degradations [3]. The former are finite duration defects which occour at random in the waveform and include clicks, scratches, clipping, ... (in the sequel they will be simply referenced to as ”clicks”). The latter affect all the audio recording and include background noise (perceived as “hiss”), wow, flutter and some types of linear and nonlinear distortion.

In this context, we consider the problem of the reduction of impulsive and background noise from audio signals. This task is usually carried out using different methods for detection/restoration of impulsive noise and for broadband noise reduction [3, 8].

In this work we employ an algorithm whose objective is to simultaneously solve the problems of filtering/parameter tracking/elimination of the outliers (“clicks”) by using the Extended Kalman Filter theory (EKF), as proposed by M. Niedzwiecki and K. Cisowski [4, 5, 6]. In particular the algorithm in [6] can be interpreted as the nonlinear combination of two Kalman filters: the first is used to follow the slow variations of the signal time–varying AR model parameters, while the second takes part in the reduction of background and impulsive noise.

2 Problem statement

Let the audio signal $s(t)$, $t = 1, 2, \cdots$, be modelled by a p order *time varying* autoregressive (AR) model

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

driven by the gaussian zero–mean white noise sequence $e(t)$ with variance σ_e^2.
The time evolution of the time varying coefficients $a_i(t)$ is modelled by the random walk model

$$a_i(t+1) = a_i(t) + w_i(t) \qquad (2)$$

with $w_i(t)$ zero–mean gaussian white processes of variance σ_w^2 mutually uncorrelated, i.e., $E[w_i(t)w_j(t)] = 0$ for $i \neq j$, and independent of $e(t)$. Moreover, let us assume that the original signal $s(t)$ is corrupted by a mixture of a broadband noise $z(t)$ and impulsive noise $v(t)$ (independent of $e(t)$ and $w_i(t)$), so that the available signal $y(t)$ can be written as

$$y(t) = s(t) + z(t) + v(t). \qquad (3)$$

The noise $z(t)$ is assumed gaussian zero–mean white noise (see later for relaxing this hypothesis) of variance σ_z^2, while $v(t)$ is assumed gaussian zero-mean noise with $\sigma_v^2(t) = \infty$, if a click is present, or $\sigma_v^2(t) = 0$, otherwise. As a consequence, if a click is revealed at time t, the corresponding sample $y(t)$ must be discarded since it not bears information on $s(t)$ and $s(t)$ must be recovered from $\{\cdots, y(t-1), y(t+1), \cdots\}$.

In [5] it is shown that under the hypothesis made, the problem of recovering the signal $s(t)$ based on the noisy measurements $\mathbf{Y}(t) = \{y(t), y(t-1), \cdots, y(1)\}$ can be optimally handled by the extended Kalman filter (EKF). To this purpose it is convenient to represent signal $s(t)$ in eqn. (1) in the non–minimal state space form

$$\mathbf{s}_q(t+1) = \mathbf{A}_q[\mathbf{a}_p(t)]\mathbf{s}_q(t) + \mathbf{b}_q e(t) \qquad (4)$$

where $\mathbf{s}_q(t) = [s(t), ..., s(t-p), \cdots, s(t-q+1)]^T$, $q \geq p$, is the signal vector, $\mathbf{a}_p^T(t) = [a_1(t), \cdots, a_p(t)]^T$

is the vector of the AR model coefficients, $\mathbf{b}_q^T = [1, 0, \cdots, 0]$, and $\mathbf{A}_q(t)$ is the companion matrix associated with the extended parameter vector $\mathbf{a}_q^T(t) = [\mathbf{a}_p^T(t), \mathbf{0}_{q-p}^T]$. The provision of a nonminimal state–space description: $q > p$ will allow one for two–sided reconstruction of up to $q - p$ samples corrupted by impulsive noise.

Notice that to remove noise an accurate signal model is needed and to obtain a reliable signal model the signal should be noiseless. The problems of filtering and parameter tracking are strictly tied and are to be jointly solved. The solution to their combined treatment is obtained by combining the unknown AR model coefficients and the signal vector in a $p + q$ "state vector" $\mathbf{x}^T(t) = [\mathbf{s}_q^T(t), \mathbf{a}_p^T(t)]$ and by rewriting (1–3) as

$$\left\{ \begin{array}{l} \mathbf{x}(t+1) = f[\mathbf{x}(t)] + \mathbf{u}(t) \\ y(t) = \mathbf{c}^T\mathbf{x}(t) + \zeta(t) \end{array} \right. \tag{5}$$

where

$$f[\mathbf{x}(t)] = \left[\begin{array}{cc} \mathbf{A}_q(t) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \end{array} \right] \cdot \mathbf{x}(t), \quad \mathbf{u}(t) = \left[\begin{array}{c} \mathbf{b}_q e(t) \\ \mathbf{w}(t) \end{array} \right]$$

with $\mathbf{w}^T(t) = [w_1(t), \cdots, w_p(t)]$ and

$$\zeta(t) = z(t) + v(t), \quad \mathbf{c}^T = [\mathbf{b}_q^T, 0^T] = [1, 0, \cdots, 0].$$

The problem of estimating the model parameters $\mathbf{a}_p(t)$ and the noisefree signal $\mathbf{s}(t)$ is reduced to a nonlinear filtering problem in the state space. A (suboptimal) solution to the problem can be based on the theory of extended Kalman filter (EKF)[5, 6] and is obtained linearizing (5).
Let us denote with $\hat{\mathbf{x}}(t|t)$ the estimate of the state at time t from the measurements $y(\tau) : \tau \leq t$ and with $\hat{\mathbf{x}}(t|t-1)$ the state prediction at time t from the measurements $y(\tau) : \tau \leq t-1$. Let $\mathbf{F}(t)$ denote the state transition matrix of the linearized system

$$\mathbf{F}(t) = \frac{\partial f[\mathbf{x}]}{\partial \mathbf{x}} |_{\mathbf{x}=\hat{\mathbf{x}}(t|t)} = \left[\begin{array}{cc} \mathbf{A}_q(t|t) & \begin{array}{c} \hat{\mathbf{s}}_p^T(t|t) \\ \mathbf{0_{q-1\times p}} \end{array} \\ \mathbf{0_{p\times q}} & \mathbf{I}_p \end{array} \right] \tag{6}$$

where $\hat{\mathbf{x}}^T(t|t) = [\hat{\mathbf{s}}_q^T(t|t), \hat{\mathbf{a}}_p^T(t|t)]$ is the filtered state trajectory given by the EKF algorithm, $\mathbf{A}_q(t|t) = \mathbf{A}_q[\hat{\mathbf{a}}_p(t|t)]$ and $\hat{\mathbf{s}}_p(t|t)$ is the vector made up with the first p components of $\hat{\mathbf{s}}_q(t|t)$. Moreover, let:

$$\Omega = cov[\mathbf{u}(t)]/\sigma_e^2 = \left[\begin{array}{cc} \mathbf{b}_q\mathbf{b}_q^T & 0 \\ 0 & \xi\mathbf{I}_p \end{array} \right] \tag{7}$$

with $\xi = \sigma_w^2/\sigma_e^2$.

The EKF equations become for the **prediction** step:

$$\left\{ \begin{array}{ll} \hat{\mathbf{x}}(t|t-1) & = f[\hat{\mathbf{x}}(t-1|t-1)] \\ \Sigma(t|t-1) & = \mathbf{F}(t-1)\Sigma(t-1|t-1)\mathbf{F}^T(t-1) + \Omega \end{array} \right. \tag{8}$$

and for the **update** step:

$$\left\{ \begin{array}{ll} \hat{\mathbf{x}}(t|t) & = \hat{\mathbf{x}}(t|t-1) + L(t)\varepsilon(t) \\ \Sigma(t|t) & = (I_{p+q} - L(t)\mathbf{c}^T)\Sigma(t|t-1) \end{array} \right. \tag{9}$$

where $\Sigma(t|t) = E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t))(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t))^T]$ is the state estimation error covariance and $\Sigma(t|t-1) = E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t-1))(\mathbf{x}(t) - \hat{\mathbf{x}}(t|t-1))^T]$ is the state prediction error covariance. Moreover in (9)

$$\varepsilon(t) = y(t) - \mathbf{c}^T\hat{\mathbf{x}}(t|t-1) = y(t) - \hat{s}(t|t-1)$$

is the *prediction error* (Kalman filter innovation) and $L(t)$ is the Kalman gain, whose value depends from the click indicator function $\hat{d}(t)$:

$$L(t) = \left\{ \begin{array}{ll} \frac{\Sigma(t|t-1)c}{c^T\Sigma(t|t-1)c+k(t)} & if \quad \hat{d}(t) = 0 \\ 0 & if \quad \hat{d}(t) = 1 \end{array} \right.$$

and $k(t) = \sigma_z^2/\sigma_e^2(t)$.

The EKF can be started with the values

$$\hat{x}(0|0) = 0, \quad \Sigma(0|0) = \left[\begin{array}{cc} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \delta I_p \end{array} \right]$$

with δ a large positive constant (~ 100) to account that nothing is known in advance about $\mathbf{a}(0)$.

The corresponding algorithm has a complexity $O((p+q)^2)$. In [6] a reduced complexity split EKF algorithm is presented and it was used in the actual experiments referred to in the next section.
In addition, it can be noticed that it is not difficult to drop the hypothesis of a white noise $z(t)$. In case of coloured noise $z(t)$ it suffices to model it as an AR process and to increase the state dimension accordingly [2]. Such a provision was found quite effective for the noise reduction of some old vinyl records.

2.1 Click detection

The detection of clicks is based on the value assumed at each t by the prediction error

$$\hat{d}(t) = \left\{ \begin{array}{lll} 0 & if & |\varepsilon(t)| \leq m\hat{\sigma}_\varepsilon(t) \\ 1 & if & |\varepsilon(t)| > m\hat{\sigma}_\varepsilon(t) \end{array} \right. \tag{10}$$

In (10)

$$\hat{\sigma}_\varepsilon^2(t) = \eta(t)\hat{\sigma}_e^2(t) \;\text{ with }\; \eta(t) = \mathbf{c}^T\Sigma(t|t-1)\mathbf{c} + k(t)$$

is the estimated innovation variance, m is the parameter determining the threshold for impulsive noise detection (in practice $m = 3 \div 5$) and $\hat{\sigma}_e^2(t)$ is the *local* estimate of the model input noise $e(t)$ variance

$$\hat{\sigma}_e^2(t) = \left\{ \begin{array}{ll} \lambda\hat{\sigma}_e^2(t-1) + (1-\lambda)\frac{\varepsilon^2(t)}{\eta(t)} & if \quad \hat{d}(t) = 0 \\ \hat{\sigma}_e^2(t-1) & if \quad \hat{d}(t) = 1 \end{array} \right. \tag{11}$$

In (11) $0 < \lambda < 1$ determines the adaptation speed. In actual experimentation we used $\lambda = 0.98$, except in the case of signals with fast dynamics as in the guitar case, where a smaller value (0.7) was used.
Fig. 1 shows a segment taken from an old 78 rpm gramophone disc and the corresponding innovation ($p = 12$), which takes on greater values in corrispondence with signal discontinuities.

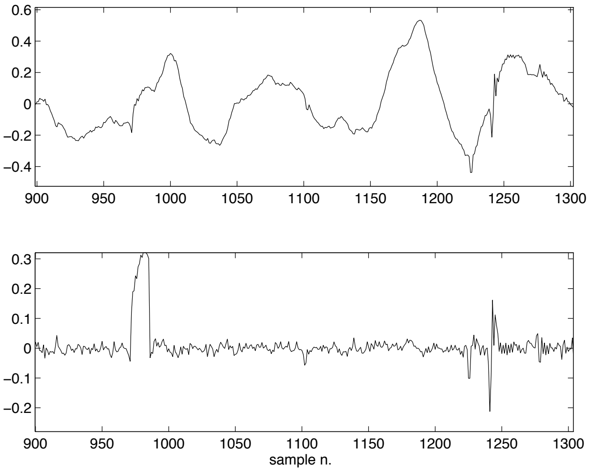

Figure 1: Click detection: noisy signal (top) and corresponding innovation (bottom).

2.2 Smoothing and reconstruction

It can be noticed that $\hat{\mathbf{s}}(t|t) = [\hat{s}(t|t), \cdots, \hat{s}(t-q+1|t)]^T$ represents the optimal (mean square) smoothed estimate of $s(t), \cdots, s(t-q+1)$ given $\mathbf{Y}(t)$, i.e., all the measurements available up to time t. To make full use of the available information, it is convenient to use, at time t, $\hat{s}(t-q+1|t)$ as an estimate of $s(t-q+1)$, i.e., it is convenient to introduce a delay of q samples.

As a result, for signal smoothing it is enough to use $q = p$. In presence of clicks it can be shown that, for a p order AR process, a block consisting of at least p “good” future successive samples is needed for good reconstruction [4]: for a group of n successive samples corrupted by a click, a value $q \geq p + n$ is required.

This consideration can be exploited to derive a variable order EKF [6], which usually uses $q = p$ and, in presence of clicks, increases q until the filter innovation corresponding to the ”corrected ” signal becomes “white” noise or q does not reach a predetermined threshold. Thus the length of the replaced signal is incremented until this condition is true. During the interpolation step the order of the filter is temporarily increased, in order to allow for a better estimation, and both past and ”future” measurements are employed, so as to carry out a ” forward – backward ” interpolation. Such a provision offers a significant computational reduction over the use of a fixed large q value.

3 Experimental results

To evaluate the performance of the EKF algorithm it is necessary to identify noise on the input recording. As a preliminary step, we used computer generated noise (white/coloured and impulsive) and added it to some test CD quality recordings, supposed to be “noise free”. This helped us to gain some insight into the method and confidence on the choice of parameters ξ, $k(0)$, m and p, which determine the ultimate performance of the algorithm.

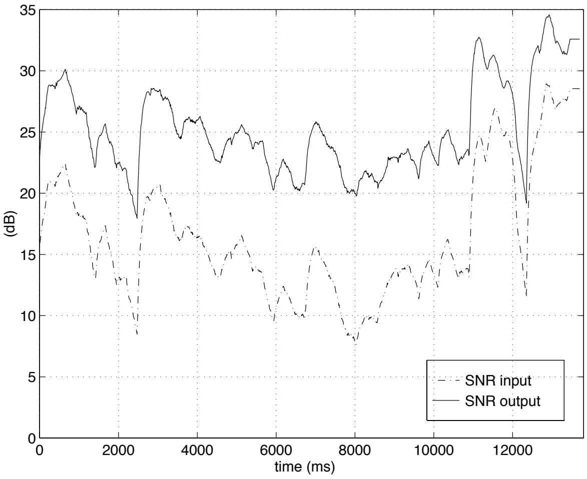

Figure 2: Example of segmental SNR improvement

The parameter $\xi = \sigma_w^2/\sigma_e^2$ (see (2) and (7)) should be chosen in accordance with the degree of nonstationarity of the signal at hand. We found a constant value $\xi \simeq 10^{-4}$ adequate in most examples, the most noticeable exception beeing an old Segovia excerpt. The fast guitar attacks required a greater value $\xi \simeq 10^{-2}$. In the future it is planned to use a time–varying value $\xi(t)$ for ξ.

The parameter $k(0)$ allows one to obtain an initial estimate of $\sigma_e^2(0) = \sigma_v^2/k(0)$ and to start the recursive estimation of $\hat{\sigma}_e^2(t)$ via (11) (σ_v^2 can be measured during silences). Its value was found not critical and in most cases we used $k(0) \simeq 2$.

As for parameter m, a small m value, say $2\div3$, allows one to detect small clicks but introduces many false alarms. This gives rise to the substitution of many samples that would be better dealt with by the EKF smoother. As a rule of thumb, we found preferable to use a high m value (i.e., $m = 4\div5$) and, in any case, to iterate the declicking process starting from, e.g., $m = 5$ and forcing a high $k(t)$ value during the first iteration/s to reduce smoothing effects accumulation.

To evaluate the white noise reduction performance, controlled amounts of “white” noise were added to “clean” recordings. The SNR_o of the output signal produced by the smoothing algorithm was measured and related to the SNR_i input signal. It was found that equation

$$SNR_o \simeq 12 + 0.8\, SNR_i \quad (dB)$$

well represents the measured values for $0 \leq SNR_i \leq$ 40 dB and $p \geq 10$, i.e., the algorithm provides an average SNR improvement of about 10 dB.

Fig. 2 reports the segmental SNR_i and SNR_o *vs* time for a 20 dB overall SNR_i (the segmental SNRs were computed every 10 ms on a 20 ms window).

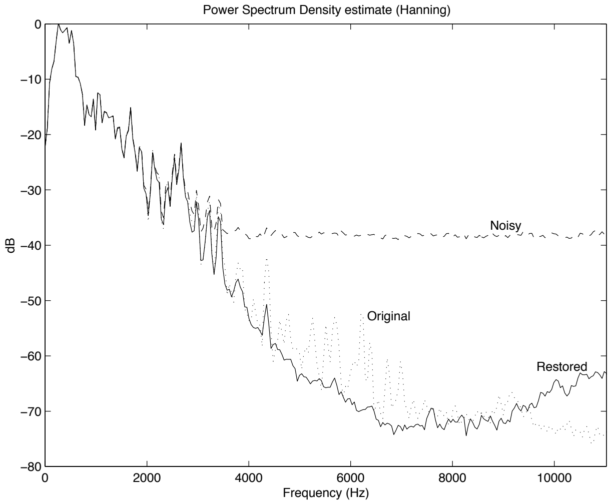

Figure 3: Power spectrum density estimate.

From the figure it is apparent that the SNR gain is approximately uniform (actually it is greater in lower SNR_i regions).
Fig. 3 gives (limitedly to $0 \div 11$ kHz) the (Welch) power spectrum estimate of a 5 s long Schubert piano piece taken from a CD, the one corresponding to its noisy version ($SNR_i = 20$ dB) and that corresponding to the restored version. From the figure it can be appreciated that the restored version spectrum strictly follows that of the original up to about 3 kHz, i.e., up to frequencies at which the white noise power density equals the one of the clean recording. In addition, differently from what would be obtained by a simple low-pass filter (with cutoff at 3 kHz) or by spectral subtraction, beyond 3 kHz the restored version “follows” the original spectrum. This property results to be perceptively important and appreciated by experienced listeners.

In order to evaluate the role of predictor order p, the algorithm was tested on artificially degraded recordings ($SNR_i = 20$ dB) with p varying between 2 and 30. The general conclusion was that the SNR gain ($SNR_o - SNR_i$) quickly increases up to $p = 8 \div 10$ and then it remains approximately constant. (The value $p = 12$ was used in all the prsented figures).

Finally, Fig. 4 shows a segment taken from a noisy piano recording (kindly supplied by S. Godsill) and its restored version. The proposed method seems to have dealth properly with clicks and wideband noise.

Acknowledgments

This work has been carried out with the finantial support of the C.N.R.–”Progetto Finalizzato Beni Culturali”.

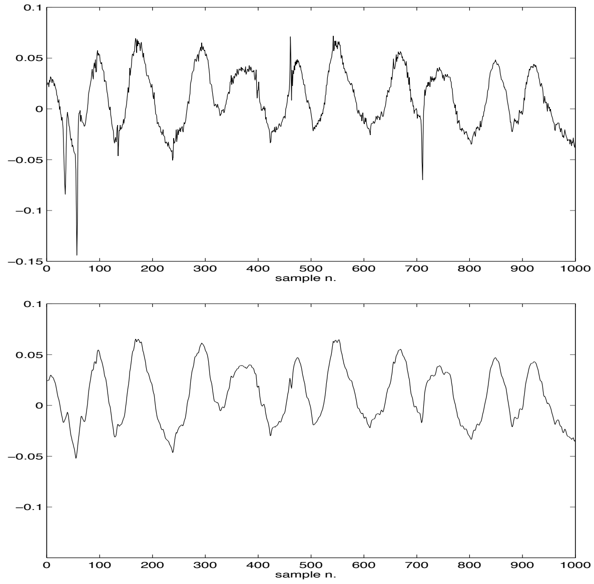

Figure 4: Segment of noisy audio (top) and its restored version (bottom).

References

[1] G. Adamo, G.B. Debiasi, G. De Poli, P. Giua, G.A. Mian, M.C. Sotgiu, A. Vidolin, “*Problemi di conservazione e restauro di archivi sonori*”, in Atti Convegno A.I.A., Perugia, 1997, pp. 106-113.

[2] B.D.O.Anderson, J.B.Moore, “*Optimal filtering*”, Prentice-Hall, 1979.

[3] S. Godsill, P. Rayner, O. Cappé, “*Digital audio restoration*”, in M. Kahrs and K. Brandenburg (Eds), “Applications of digital signal processing to audio and acoustics”, Kluwer, 1998.

[4] M. Niedźwiecki, ”*Steady-state and parameter tracking proprieties of self-tuning minimun variance regulators*”, Automatica, pp. 597-602, 1989.

[5] M. Niedźwiecki and K. Cisowski, ”*Adaptive scheme for elimination od broadband noise and impulsive distubances from AR and ARMA signals*”, IEEE Trans. Signal Processing, vol. 44, March 1996.

[6] M. Niedźwiecki, “*Identification of time-varying processes in the presence of measurement noise and outliers*”, Proc. 11th IFAC Symp. System Identification, 1765–1768, Tokyo, 1997.

[7] D. Schueller, ”The ethics of preservation, restoration and re–issues of historical sound”, J. Audio Eng. Soc., 39, 12, pp. 1014-1016, Dec 1991.

[8] R.Veldhuis, “*Restoration of lost samples in digital signals*”, Prentice-Hall, 1990