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. Little attention is currently paid to "modern" electro-acoustic music restoration. On the other hand, since the tapes are exposed to deterioration (often the performances are recorded on semi-professional media and preserved in a precarious state) the preservation/restoration problem is significant. The main issue is to rely on good signal models to perform a high quality separation of signal from noise (to be attenuated).

Do to the fact that existent devices and developed methodologies are oriented to classical and popular music restoration, few results are available in the field of electro-acoustic music. The types of degradation common in electro-acoustic tapes can be broadly classified into localized and global degradations [D. Schueller, (1991)]. The former are finite duration defects which occur at random in the waveform and include clicks, scratches, clipping, etc. 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. In this work, we employ an algorithm whose objective is, in principle, to simultaneously solve the problems of filtering/parameter tracking/elimination of the outliers ("clicks") by using the Extended Kalman Filter (EKF) theory, as proposed by [M. Niedzwiecki, (1996)]. In particular the algorithm in [M. Niedzwiecki and K. Cisowski, (1996)] 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.

At medium/high signal-to-noise (SNR) ratios, the performance of such filter is superior to that of other standard methods like Spectral Attenuation, nevertheless its plain use does not guarantee the best results. One reason for this is that the nonstationarity of audio signals leads to errors in parameter and tracking noise filtering, especially during fast transients.

In order to achieve maximum performance from the EKF, it is essential to optimize its implementation. For this purpose, to cope with the non stationary nature of the audio signal, we used two properly combined EKF filters (forward and backward), and introduced a bootstrapping procedure for model tracking. The careful combination of the proposed techniques and an accurate choice of some critical parameters, allows one to improve the performance of the EKF algorithm. Different audio examples validate the effectiveness of the presented procedure.

2. PROBLEM STATEMENT

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

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

driven by the gaussian zero-mean white noise sequence \( e(t) \) with variance \( \sigma_e^2 \).

The time evolution of the time varying coefficients \( a_i(t) \) is modelled by the random walk model

\[
a_i(t+1) = a_i(t) + w_i(t) \quad (i = 1, \ldots, p)
\]

with \( w_i(t) \) zero mean gaussian white processes of variance \( \sigma_w^2 \), mutually uncorrelated, i.e., \( \text{E}[w_i(t) w_j(t')]=0 \) 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 \( \nu(t) \) (independent of \( e(t) \) and \( w_i(t) \)), so that the available signal \( y(t) \) can be written as

\[
y(t) = s(t)+z(t)+\nu(t)
\]

The noise \( \nu(t) \) is assumed gaussian zero-mean white noise (see later for relaxing this hypothesis) of variance \( \sigma_\nu^2 \), while \( y(t) \) is assumed gaussian zero-mean noise with variance \( \sigma_y^2 = \infty \) if a click is present, or \( \sigma_y^2 = 0 \) 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 \( \{\ldots, y(t-1), y(t+1), \ldots\} \).

In [M. Niedzwiecki and K. Cisowski, (1996)] it is shown that under the hypothesis made, the problem of recovering the signal \( s(t) \) based on the noisy measurements \( Y(t)=\{y(t), y(t-1), \ldots\} \),
\( y(1) \) can be optimally handled by the Extended Kalman Filter (EKF). To this purpose it is convenient to represent signal \( s(t) \) in (1) in the non-minimal state-space form

\[
x_{q}(t+1) = A_{q}(t)x_{q}(t) + b_{q}e(t)
\]

where \( x_{q}(t) = [s(t), \ldots, s(t-p+1)]^{T} \), \( q \geq p \), is the signal vector, \( a_{q}^{T}(t) = [a_{q1}(t), \ldots, a_{qT}(t)]^{T} \) is the vector of the AR model coefficients, \( b_{q}^{T}(t) = [1, 0, \ldots, 0]^{T} \), and \( A_{q}(t) \) is the companion matrix associated with the extended parameter vector \( a_{q}^{T}(t) \) and is obtained linearizing (5).


based on the theory of extended Kalman filter (EKF) [M. Niedzwiecki. (1997)] a clever reduced complexity split EKF algorithm is presented that was used in the actual experiments known in advance about \( q \).

The problem of estimating the model parameters \( a_{q}(t) \) and the noise-free signal \( s(t) \) is reduced to a nonlinear filtering problem in the state space. A (suboptimal) solution to the problem can be accordingly [B.D.O. Anderson, J.B. Moore, (1979)]. Such a combined treatment is obtained by combining the unknown AR model coefficients and the signal vector in a \( p + q \) “state vector” \( x^{T}(t) = [s(t), a_{q}^{T}(t)] \) and is obtained linearizing (5) as

\[
\begin{align*}
x(t+1) &= f[x(t)] + u(t) \\
y(t) &= c^{T}x(t) + \xi(t)
\end{align*}
\]

where

\[
f[x(t)] = \begin{bmatrix} A_{q}(t) & 0 \\ 0 & 1 \end{bmatrix}x(t), \quad u(t) = \begin{bmatrix} b_{q}c(t) \\ w(t) \end{bmatrix}
\]

with \( w^{T}(t) = [w_{1}(t), \ldots, w_{q}(t)] \) and \( \xi(t) = z(t) + v(t) \), \( c^{T} = [b_{q}^{T}, 0] = [1, 0, \ldots, 0] \).

The EKF can be started with the values

\[
\Omega = \frac{1}{\sigma_{i}^{2}} \begin{bmatrix} b_{q}b_{q}^{T} & 0 \\ 0 & 0 \end{bmatrix}
\]

with \( \xi = \frac{\sigma_{i}^{2}}{\sigma_{q}^{2}} \). The EKF equations become for the prediction step:

\[
\begin{align*}
\hat{x}(t|t-1) &= f[\hat{x}(t-1|t-1)] \\
\Sigma(t|t-1) &= F(t-1)\Sigma(t-1|t-1)F^{T}(t-1) + \Omega
\end{align*}
\]

and for the update step:

\[
\begin{align*}
\hat{x}(t|t) &= \hat{x}(t|t-1) + L(t)(y(t) - \hat{y}(t|t-1)) \\
\Sigma(t|t) &= (I - L(t)L^{T})\Sigma(t|t-1)
\end{align*}
\]

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

\[
e(t) = y(t) - c^{T}\hat{x}(t|t-1) = y(t) - \hat{y}(t|t-1) \quad \text{is the prediction error (Kalman filter innovation)} \quad \text{and} \quad L(t) = \text{the Kalman gain, whose value depends from the click indicator function } \hat{d}(t):
\]

\[
L(t) = \begin{cases} \Sigma(t|t-1)e & \text{if } \hat{d}(t) = 0 \\ c^{T}\Sigma(t|t-1)e + k(t) & \text{if } \hat{d}(t) \neq 0 \\ 0 & \end{cases}
\]

\[
k(t) = \frac{\sigma_{i}^{2}}{\sigma_{q}^{2}}
\]

The EKF can be started with the values

\[
\hat{z}(0|0) = 0, \quad \Sigma(0|0) = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{i}^{2} \end{bmatrix}
\]

with \( \delta \) a large positive constant (\(-100\)) to account that nothing is known in advance about \( a(0) \).

The corresponding algorithm has a complexity \( O(pq^{2}) \). In [M. Niedzwiecki. (1997)] a clever reduced complexity split EKF algorithm is presented that was used in the actual experiments referred to in the following.

It can be noticed that it is not difficult to drop the hypothesis of a white noise \( z(t) \). In case of colored noise \( z(t) \) it suffices to model it as an AR process and to increase the state dimension accordingly [B.D.O. Anderson, J.B. Moore, (1979)]. Such a provision was found quite effective for the noise reduction of some old vinyl and tape records.

### 2.1 Click detection

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

\[
\hat{d}(t) = \begin{cases} 0 & \text{if } \{ e(t) \} \leq \mu \hat{z}(t) \\ 1 & \text{if } \{ e(t) \} > \mu \hat{z}(t) \end{cases}
\]

where \( \mu \) is a large positive constant (\(-100\)) to account that nothing is known in advance about \( a(0) \).

The corresponding algorithm has a complexity \( O(pq^{2}) \). In [M. Niedzwiecki. (1997)] a clever reduced complexity split EKF algorithm is presented that was used in the actual experiments referred to in the following.

It can be noticed that it is not difficult to drop the hypothesis of a white noise \( z(t) \). In case of colored noise \( z(t) \) it suffices to model it as an AR process and to increase the state dimension accordingly [B.D.O. Anderson, J.B. Moore, (1979)]. Such a provision was found quite effective for the noise reduction of some old vinyl and tape records.
Impulsive noise detection (in practice with \( s \)) can be noticed that starting the algorithm from scratch, i.e., with the initial conditions given by (10), implies an initial transient of the innovation corresponding to the “corrected” signal becomes unstable AR-model the parameter update is skipped. The nonstationary nature of the audio signals poses a challenging problem to EKF: it has both to filter the noisy components and to continuously track the signal model parameters. As a consequence of signal nonstationarity, errors in parameter tracking are small in high SNR regions but can severely affect signal filtering during fast transients and in regions with a low SNR. A provision that improves the algorithm performance, is given by the use of two properly combined EKF’s operating forward and backward on the signal.

The comparison of the prediction error variances \((\hat{\sigma}_{e1}^2(t)\) and \(\hat{\sigma}_{e2}^2(t)\)) shows that the two variances are not equal. Moreover, there is a trend in their behavior: on the average the forward variance \(\hat{\sigma}_{e1}^2(t)\) is higher than the backward one \(\hat{\sigma}_{e2}^2(t)\) at the beginning and lower at the end, i.e., tracking improves as time goes on at least on a small time scale (few seconds). Moreover a finer scrutiny shows that, quite often, during attacks the backward filter has a lower innovation variance. This fact can be justified noticing that such filter traverses such transitions going from a high SNR region, where parameter tracking and noise filtering are less critical, to a low SNR region, where the contrary happens.

A similar behavior is exhibited by the residual errors \(s(t) = \hat{s}(t) - \hat{z}(t)\) and \(\bar{s}(t) - \hat{\bar{z}}(t)\). Since the two filters give different signal estimates, \(\hat{s}(t)\) and \(\hat{\bar{z}}(t)\), on the lack of a better procedure we found it effective to combine them according to

\[
\hat{s}(t) = \frac{\hat{\sigma}_{e1}^2(t)}{\hat{\sigma}_{e1}^2(t) + \hat{\sigma}_{e2}^2(t)} \hat{z}(t) + \frac{\hat{\sigma}_{e2}^2(t)}{\hat{\sigma}_{e1}^2(t) + \hat{\sigma}_{e2}^2(t)} \hat{\bar{z}}(t)
\]

Notice that the performance of the procedure improves, if the procedure is started from an high SNR region instead than from the beginning and the end of the chosen excerpt: details are given in [A. Bari, (1999)].

With such a provision it is possible, in the authors experience, to effectively remove broadband noise for (average) SNR ≥ 30 dB. In addition, since we have two different click detectors, the effectiveness in removing impulsive disturbances is also improved.

4. EXPERIMENTAL RESULTS

To evaluate the performance of the EKF algorithm it is necessary to identify the noise on the input recording. As a preliminary step, different noise free recordings were intentionally corrupted with different scratches and broadband noises. The noise free recordings were selected such as to cover a wide range of musical genre (Electronic, Jazz, Western Classical, Pop); the stimuli were played with different instruments (piano solo, orchestra, voice, and so forth) in order to analyze the effectiveness of the EKF related to the different physical characteristics of audio signal (bandwidth, transients, timbre). This helped us to gain some insight into the method and confidence on the choice of parameters \(\xi, p, \mu\) and \(\sigma_c\), which determine the ultimate performance of the algorithm. In addition, the starting value of \(\sigma_0^2\) has to be estimated. We found it useful to estimate \(\sigma_0^2(0)\) via Youle-Walker equations (assuming a
The parameter $\zeta = \frac{\sigma^2}{\sigma^2}$ should be chosen in accordance with the degree of nonstationarity of the signal at hand. We found a constant value $\zeta = 10^{-6}$, as in [M. Niedzwiecki and K. Cisowski (1996)], adequate in most examples.

As for parameter $\mu$, a small $\mu$ value (say 4) 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 it preferable to use a high $\mu$ value (i.e., $\mu = 4-6$) and to iterate the de-clicking process decreasing the values of $\mu$, after careful listening to the filtered signal $\hat{y}(\ell)$ and to the difference signal $y(\ell) - \hat{y}(\ell)$.

To evaluate the white noise reduction performance, the amount of “white” noise added to the “clean” recordings was varied. The SNR $\gamma$, of the output signal produced by the smoothing algorithm was measured and related to the SNR $\gamma_i$ of the input signal. It was found that the equation:

$$\text{SNR}_{\gamma} = \text{SNR} + q$$

well represents the measured values for $20 \leq \text{SNR} \leq 40$ dB. The $m$ and $q$ values depend on the particular piece: representative values for the typical values $p = 12$ are $m = 0.8$ and $q = 12$ that provide an average SNR gain of about 6 dB for $\text{SNR} = 30$ dB, the minimum average SNR $\text{SNR}_{\gamma}$ value at which we found the algorithm to give fine results.

Another index used in order to evaluate the performance of the algorithm is the Spectral Distance (SpD) measure, related to the harmonic components of the signals:

$$\text{SpD} = \frac{1}{10} \log 10 \left\{ \frac{1}{2 \pi} \int_{0}^{2 \pi} \log \frac{S_i(\theta)}{S_o(\theta)} \, d\theta \right\} \quad (\text{dB})$$

where $S_i(\theta)$ and $S_o(\theta)$ are the (Welch) periodograms of input and output signals respectively.

The experience gained led us to the following protocol. The recording is segmented into frames up to say, $10 \, \text{s}$ long and each frame is restored according to a three step procedure. At each step the algorithm is tuned to obtained a particular restoration effect. To this purpose, it is worth noticing a nice property of the EKF: it can be forced to perform only click detection (setting $\sigma^2 = 0$) or only noise filtering (setting a very high $\mu$ value). In addition, in noise filtering it is possible to modulate the “true” $\sigma^2$ value in order to make the filter more effective against noise in medium/low SNR regions.

The three step procedure used was:

1. de-clicking, to remove the most significant clicks:

$$\hat{x}(\ell) = \left( \hat{x}(\ell) + \hat{z}(\ell) \right) / 2$$

2. noise only removal: $\hat{x}(\ell)$ is obtained according to (13);

3. de-clicking, to remove the residual small clicks:

$$\hat{x}(\ell) = \left( \hat{x}(\ell) + \hat{z}(\ell) \right) / 2$$

where step 3) can be skipped.

The protocol was exercised on electronic-music recordings supplied by the RAI. As an example, let us consider the restoration of an excerpt taken from B. Maderna, Notturno. The piece was restored according to the parameters describe above. Fig. 1 shows the comparison among the (Welch) periodograms of the original (noisy) and restored recordings (limited to $0 \leq f = F_s/2 = 22 \, \text{kHz}$).

From the figure it can be appreciated that the restored version spectrum follows that of the original up to frequencies at which the white noise power density equals that of the clean recording.

In addition, differently from what would be obtained by a simple low-pass filter or by spectral subtraction, the restored version “follows” the original spectrum even for higher frequencies. This property results to be perceptively important and appreciated by experienced listeners.

![Fig. 1: Comparison between the (Welch) periodograms of the original and restored recordings (limited to $0 \leq f = F_s/2 = 22 \, \text{kHz}$).](image)

5. REFERENCES


