**Efficiency of Impulsive Noise Detection in Audio Recordings Using the Adaptive Filtering Method**

**Preprint 3464 (B1-4)**

Krzysztof Cisowski
Technical University of Gdansk
Gdansk, Poland

**Presented at the 94th Convention 1993 March 16–19 Berlin**



**AN AUDIO ENGINEERING SOCIETY PREPRINT**

"EFFICIENCY OF IMPULSIVE NOISE DETECTION IN AUDIO RECORDINGS USING THE ADAPTIVE FILTERING METHOD"

Krzysztof CISOWSKI

Control System Division, Faculty of Electronics,
Technical University of Gdańsk,
Majakowskiego 11/12, 80-952 Gdańsk, Poland

## 1. INTRODUCTION

The noise detector presented in this paper is the first part of the computer system for reconstruction of audio recordings (mainly archival gramophone recordings) which is currently under development. The problem of impulsive noise detection is difficult because of similarity of the time-freguency properties of both musical sounds and typical disturbances. First, based on some experimental evidence, it is argued that audio signals can be efficiently parametrized in terms of autoregresive coefficients (the AR model). Next, the idea of impulsive niose detection based on monitoring of one-step-ahead prediction errors yielded by the AR adaptive filter, is discused. Description of the computer system for impulsive noise detection is presented. Finally, results obtained for several kinds of audio sighals corrupted with the same artifical impulsive noise patterns are discused leading to several conclusions on efficiency of the proposed detection scheme.

## 2. PROPERTIES OF IMPULSIVE NOISE OF THE GRAMOPHONE RECORDINGS

The are two kinds of impulsive noise signals that are typical of the gramophone audio recordings. The first group consist of the elementary clicks with a lasting time of several tens of μs. They have almost invariant shape in the time domain and can be interpreted as the impulse response of the pick-up cartridge system excited by a short time noise impulse [1], [3]. The elementary clicks have several chracteristic features such as, a short rise time, short decay time and quick jumps between extreme values of amplitudes. The second group of noise patterns consists of complex clicks, i.e., clicks with a lasting time up to 1 ms for a contemporary gramophone recordings and up to several ms for old gramophone recordings. These clicks can be regarded as a convolution of the impulse response of the pick-up cartridge system and the disturbance signal [3]. Artifacts belonging to the first category can be attributed to the limitations of the pick-up cartridge system and are caused by mechanical resonances of the system consisting of the playing needle with cooperated masses and the gramophone recording surface [1], [illegible]. Since complex clicks consist of many elementary clicks (fig. 1c), they have rise and decay times comparable with those of elementary clicks (several tens of μs) [3]. In case of old gramophone recordings with severe damages of the record groove these times can be up to ten time longer [4]. As opposed to clicks typical audio signals have rise and decay times bigger than 30 ms. Only in special cases, for percussion, brass, pulled string and elecronic instruments, can a rise time be shorter (but not less) than 2 ms. The decay time due to dying away or artificial reverberation is also longer than 30 ms [3], [4]. The spectrum of clicks covers a wide range of frequencies (fig. 1b, 1d) starting (in case of old gramophone recordings) from 300 Hz [4]. The peaks of the click spectrum can appear at frequencies of several kHz occasionally exceeding 10 kHz. The spectrum of audio signals has also its maxima in the range of low and middle frequencies, i.e., at several kHz [4]. It can be noticed that the spectrum of clicks overlaps with that of musical sounds which makes the task of detection and removal of impulsive noise more difficult. When the stereophonic pick-up cartridge is used for playing-back monophonic or stereophonic gramophone recordings one can see that the signal of difference between the left and right channel has a higher level of clicks and lower level of the signal than each of the channels alone. This effect is straightforward consequence of the fact that noise signals observed in both channels are not correlated with each other while the corresponding audio signals (both in the monophonic or stereophonic case) are rather strongly correlated [1], [4].

## 3. SUBSTANTIATION OF MUSICAL SIGNALS AUTOREGRESIVE MODELING

The proposed method of impulsive noise detection is based on the assumption that the investigated audio signal can be modelled as a time-varying autoregresive (AR) process:

$$y_n = \sum_{i=1}^{p} a_i(n) * y_{n-i} + e_n \qquad n \in \langle 1, N \rangle \tag{1}$$

where $p$ is the model order, $a_1(n),\dots,a_p(n)$ denote time-varying autoregressive coefficients and $\{e_n\}$ is a white Gaussian noise. The method of exponentially weghted least squares is used to estimate the vector of autoregresive coefficients $\mathbf{a}(n) = [a_1(n),\dots,a_p(n)]^T$ from samples of the signal $\{y_n\}$. This amounts to chooseing $\hat{\mathbf{a}}(n)$ so as to minimize the function $\xi(\hat{\mathbf{a}},n)$ [2].

$$\xi(\hat{\mathbf{a}},n) = \sum_{i=p+1}^{n} \lambda^{n-i} * \left| y_i - \sum_{j=1}^{p} \hat{a}_j(n) * y_{i-j} \right|^2 \longmapsto \min \tag{2}$$

where $\lambda$, $0 < \lambda < 1$, is the so-cold exponential weighting (forgeting) factor. Based on the local model of the signal the sequence of prediction errors can be determined

$$\epsilon_n = y_n - \hat{y}_{n|n-1} = y_n - \sum_{i=1}^{p} \hat{a}_i(n-1) * y_{n-i} \tag{3}$$

Due to exponential weighting used in (2) the most recent modelling errors have stronger influence upon the current estimate than those observed in the remote past. As a consequence the adaptive filter can track slow variations in $\hat{\mathbf{a}}(n)$. The effective length of an exponential window $\tau_e$ is given by

$$\tau_e = \frac{1}{1 - \lambda} \tag{4}$$

Another useful measure of the window width used in practice is the equivalent number of observations $\tau_r$

$$\tau_r = \frac{1 + \lambda}{1 - \lambda} \tag{5}$$

which shows how many samples from the past have the real influence on current estimates of AR coefficients [5], [7]. The value of the forgetting factor $\lambda$ should be chosen in accordance with the speed of variation of AR coefficients. In practice this value is usually selected in an experimental way.

Most of musical instruments produce sounds with line spectrum. For this reason the typical musical signal which is a sum of sounds, corresponding to different instruments, has a spectrum with sharp maxima (fig. 3). On the other hand it is well-known [6] that the autoregresive modeling yelds good results when applied to signals with peaky spectrum. The second reason behind the choice of this representation is due to the availability of effective algorithms for AR coefficients estimation. From a large number of recursive least squares routines the least squares lattice (LSL) algorithm described in [2] was chosen. The algorithm has several advantages, namely:

- the obtained models are always stable (the modulus of reflection coefficients is always < 1);
- the algorithm is relatively insensitive to the choice of initial condition;
- the algorithm is numerically stable (numerical errors do not accumulate)

## 4. IDEA OF IMPULSIVE NOISE DETECTION

In the proposed method of an impulsive noise detection instead of analyzing the audio signal $\{y_n\}$ the prediction error signal $\{\epsilon_n\}$ is investigated. The scheme used to generate prediction errors (whitening filter) is shown in fig. 2. Its main part is an AR-based predictor which at each time instant $n$ determines the estimate $\hat{y}_{n|n-1}$ of the signal $y_n$. If the order of the model is chosen propriately (i. e. by means of analysing a number of the power density maxima - c.f. chapter 6) the prediction (innovation) sequence can be, with a small error, regarded as a squence of uncorrelated Gaussian variables. Moreover the variance of the prediction error is much smaller than the variance of the signal [2]. Since disturbances are not corelated with the signal they produce high-magnitude burst on the output of the whitening filter and hence are easily detectable.

### 4.1. PARAMETER ESTIMATION IN THE PRESENCE OF IMPULSIVE NIOSE

Estimation of autoregresive coefficients in the presence of impulsive noise should be preceded by outlier detection and removal. Then one of the available methods designed for identification of incomplete time series (time series with missing observations) should be applied. This complicated approach which can significantly improve the quality of the impulsive noise detection will be a subject of future research. In the present development a much simpler approach is taken exploiting the quick convergence property of the least squares adaptive filters. When samples of the audio signal corrupted with impulsive noise enter the adaptive filter the estimates of autoregresive coefficients become slightly perturbed. However, shortly after the disturbance dies away the estimates converge to their previous values (characterizing the unperturbed signal). Insignificant modification of the AR coefficients still gives a high level of the impulsive noise in the signal of prediction error so detection of a disturbance is possible.

## 4.2. DEFINITION OF IMPULSIVE NOISE DETECTION TRESHOLD

A very simple outlier detection rule was adopted: it was assumed that the observed sample of the signal was corrupted by impulsive noise if the corresponding prediction error exceeded a given multiple of its standard deviation $\sigma$ (when the model order is chosen properly, the prediction error signal can be regarded a white Gaussian noise). Two methods of estimating the variance $\sigma^2$ of the prediction error (for uncorrupted signal) were proposed. The exponentially weighted estimate of $\sigma^2$ is given by:

$$\hat{\sigma}_a(n)^2 = \frac{\sum_{i=0}^{n-1}(\epsilon_{n-i} - \hat{m}_a(n))^2 * \mu^i}{N_{ef}(n)} \tag{6}$$

$$\hat{m}_a(n) = \frac{\sum_{i=0}^{n-1}\epsilon_{n-i} * \mu^i}{N_{ef}(n)} \tag{7}$$

$$N_{ef}(n) = \sum_{i=0}^{n-1} \mu = \frac{1 - \mu^n}{1 - \mu} \tag{8}$$

where $\mu$, $0 < \mu < 1$, is forgetting factor (the value of $\mu$ is chosen independenly of a value of the forgeting factor $\lambda$ in (2)). Alternatively, the sliding window estimates can be used

$$\hat{\sigma}_b(n)^2 = \frac{\sum_{i=0}^{M-1}(\epsilon_{n-i} - \hat{m}_b(n))^2}{M} \tag{9}$$

$$\hat{m}_b(n) = \frac{\sum_{i=0}^{M-1} \epsilon_{n-i}}{M} \tag{10}$$

where $M$ is the length of the local (rectangular) observation window.

In the first approach the most recent samples of the prediction error have stronger influence upon the current variance estimate than those observed in the remote past. In order to make the estimation algorithm insensitive to outliers, the following modifications were introduced:

1° if at the time instant $n$ the absolute value of prediction error $\epsilon_n$ is not larger than $3*\hat{\sigma}_a(n-1)$ then variance $\hat{\sigma}_a(n)^2$ is calculated using (6)-(8),

2° in the opposite case the variance estimate is not updated.

The $3*\sigma$ rule is based on the assuption that the prediction error signal is a white Gaussian niose. Therefore the probability of "accidental" exceeding of the level $3*\sigma$ is very small (the probability of "false alarm" is in this case smaller than 0.003). The recursive algorithm for computing (6)-(8) (including the above mentioned modifications) takes the form:

1° if $|\epsilon_n| \leq 3*\hat{\sigma}_a(n-1)$

$$\hat{\sigma}_a(n)^2 = \frac{g_a(n)}{N_{ef}(n)} - \left(\frac{d_a(n)}{N_{ef}(n)}\right)^2 \tag{11}$$

where: $g_a(n) = \mu * g_a(n-1) + \epsilon_n^2$
$d_a(n) = \mu * d_a(n-1) + \epsilon_n$
$N_{ef}(n) = \mu * N_{ef}(n-1) + 1$

2° if $|\epsilon_n| > 3*\hat{\sigma}_a(n-1)$

$\hat{\sigma}_a(n)^2 = \hat{\sigma}_a(n-1)^2$
$g_a(n) = g_a(n-1)$
$d_a(n) = d_a(n-1)$
$N_{ef}(n) = N_{ef}(n-1)$

In order to initialize the algorithm one should set: $g_a(0) = d_a(0) = N_{ef}(0) = 0$.

The expression (9) can be also converted into the recursive form. Similarly as in the case of (11) modification based on the $3*\sigma$ rule is introduced to make the estimation algorithm insensitive to outliers. The following auxiliary vector

$\mathbf{c}=[c_1(n),\ldots,c_M(n)]^T$ is used to store the last $M$ samples "accepted" by the $3*\sigma$ rule. In order to initialize the algorithm properly one should set at instant $n=M$

$$c_i(M) = \epsilon_i \qquad i\in\langle 1,M\rangle \tag{12}$$

Similarly, the initial values of two auxiliary quantities $d_b(n)$ and $g_b(n)$ should be

$$d_b(M) = \sum_{i=1}^{M} \epsilon_i \tag{13}$$

$$g_b(M) = \sum_{i=1}^{M} \epsilon_i^2 \tag{14}$$

For $n = M$ the estimate is given by

$$\hat{\sigma}_b(M)^2 = \frac{g_b(M)}{M} - \left(\frac{d_b(M)}{M}\right)^2 \tag{15}$$

For $n>M$ the algorithm for determining the estimate $\hat{\sigma}_b(n)^2$ has form:

$$1^\circ \text{ if } |\epsilon_n| \leq 3*\sigma_b(n-1) \tag{16}$$

$$\hat{\sigma}_b(n)^2 = \frac{g_b(n)}{M} - \left(\frac{d_b(n)}{M}\right)^2$$

where: $g_b(n) = g_b(n-1) + \epsilon_n^2 - c_1(n-1)^2$

$d_b(n) = d_b(n-1) + \epsilon_n - c_1(n-1)$

$c_i(n) = c_{i+1}(n-1) \qquad i\in\langle 1,M-1\rangle$

$c_M(n) = \epsilon_n$

$2^\circ$ if $|\epsilon_n| > 3*\sigma_b(n-1)$

$\hat{\sigma}_b(n)^2 = \hat{\sigma}_b(n-1)$

$g_b(n) = g_b(n-1)$

$d_b(n) = d_b(n-1)$

$c_i(n) = c_i(n-1) \qquad i\in\langle 1,M\rangle$

As it was already said the outlier detection procedure is based on testing whether the currently observed prediction error exceeds a given multiple (usually equal to 3) of its standard deviation. The "zero-one" signal is then formed with "zeros" showing positions of undestorted samples and "ones" (numbers = 16384) indicating locations of outliers.

## 5. DESCRIPTION OF THE COMPUTER SYSTEM FOR THE IMPULSIVE NOISE DETECTION

The computer system is written in the BORLAND C++ language and can be used under the DOS operating system (version 3.31 and higher). In this paper only the most important system functions, needed for the impulsive niose detection, will be discussed. The system has main "menu" which allows the user to read 16-bit signal samples from the disk (in the future, the system will be equipped with AC/DC converter) to several independent buffers (max 32767 samples) and to write on a disk samples from the optional buffer. Samples placed in the buffers can be:

- multiplied or to divided by a given number,
- increased or decreased by a given number,
- multiplied or divided by samples from the other optional buffer,
- raised to the powers of: 2, 3, 1/2, 1/3.

Additionally one can:

- copy a fragment of a given buffer to another optional buffer,
- smooth samples by: a moving average, Hann, Hamming or Blackman windows, or by means of approximating polynomials,
- filter samples by the adaptive filter (the LSL algorithm);
- evaluate the mean standard deviation of signals by one of the two methods discussed above.

The system has the option of visualization of samples from two selected buffers. In particular it is possible to move the observation window across the data, adjust the scale of previewed fragments (linear scale), shift selected portions of the data between the buffers etc..

## 6. DESCRIPTION OF THE RESEARCH MATERIAL AND THE WAY OF INVESTIGATION CARING OVER

Two fragments of music recordings were chosen as a material for experimental studies. The first of them is a fragment of clasic composition for piano solo while the second one is a fragment of a composition for a baroque orchestra (with harpsichord). After removing the mean value both (noise-free) signals were sampled at the rate of 24 kHz (the cut-off frequency of the anti-aliasing filter $f_g$ = 10 kHz). The total number of samples collected for each fragment was 16384. Several "typical" artifical impulsive noise patterns were generated and added to noise-free signals. The following noise waveforms were used (c.f. figures 5b and 8b):

- a single impulse,
- a pair (one after another) of impulses of the opposite sign,
- a pair of impulses of the same sign separated by several sampling periods,
- a pair of impulses of the opposite sign separated by several sampling periods,

The order of autoregression was chosen after analysing a number of maxima of the averaged periodograms (fig. 3). The periodograms were calculated for those segments of noise-free signals to which the impulsive noise was later added (+-512 samples). The periodogram obtained for the fist segment shows only one significant maximum, which indicates that it is sufficient to put $\mathbf{p}$ = 2 (twice the number of dominant peaks of the power spectral density) [6]. In the case of the second segment the periodogram has 8 significant maxima which suggests that the AR model of order $\mathbf{p}$ = 16 should be used. Later on it was found out experimentally that the "compromise" choice of $\mathbf{p}$ = 10 yielded good results for most of the audio (music) signals under consideration. Figures 5 - 8 show the results obtained for the two test signals described above and different selections of design parameters such as $\lambda$ and $\mu$ (forgetting constants) or $\mathbf{M}$ (the width of the rectangular window). In all cases the horizontal axes show locations of the corresponding samples and the vertical axes - their normalized amplitudes.

## 7. CONCLUSIONS

The analysing of the obtained results shows, that the efficiency of the detector does not depend critically on the forgeting factor $\lambda$. Based on the experimental evidence the value of $\lambda$ = 0.95 can be recommended. Generally, if the value of $\lambda$ is too close to 1 (e.g. $\lambda$>0.99) the effects of the outlier can be seen at the output of the adaptive filter for a long time after such disturbance is "absorbed" (c.f. fig. 4). Similarly, if the value of $\lambda$ is too small (e.g. $\lambda<0.9$) the results of parameter estimation become too poor to guarantee satisfactory performance of the detector.

Even though the choice between the two proposed methods of estimating the prediction error variance is also not critical (see figures 5 and 7) the exponentially weighted estimator seems to be giving slightly better results (leading to a smaller number of false alarms). In this method the best selection of the value of the forgeting factor $\mu$ is the value of the forgeting factor $\lambda$. Likewise in the method with a rectangular window, one can get good results when the window width $M$ is equal to the equivalent number of observations $\tau_r$ (5).

Comparision of detection results obtained for two different audio signals corrupted by the same impulsive noise leads to the following conclusions:

- if disturbances are added to signals with a "rich" spectrum (fig. 3) it is more difficult to distinguish them from high frequency components of the signal (fig. 8).
- for efficient detection of impulsive noise one should use "$k*\sigma$" criterion with $k>3$ (i.e., $k=4$ or $k=5$); in particular the choice $k=4$ seems to be a good trade-off between sesitivity of the decision rule and the probability of generating false alarms (in all examples discussed above k was set to 4).

## 8. BIBLIOGRAPHY


[1] M.J. Carey, I. Buckner "A system for reducing impulsive noise on gramophone reproduction equipment", The Radio and Electronic Engineer, Vol.50, No.7, pp. 331-336, July 1980.

[2] S. Haykin "Adaptive filter theory", Prentice-Hall, Englewood Cliffs, New Jersey 07632, 1986.

[3] G.R. Knize, jr., D.W. Gravereaux "Automatic detection of impulse noise", JAES, Vol.21, No.3, April 1973, PP. 181-184.

[4] M. Królewski "An electronic circuit for removing impulse noise from analog program" 84th AES Convention, Preprint No 2571 (B-5), Paris, 1988.

[5] J. I. Makhoul, L. K. Cosell "Adaptive Lattice Analysis of Speech" IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-29, No. 3, June 1981, pp. 654-659.

[6] S. L. Marple Jr. "Digital spectral analysis" Prentice-Hall INC. Englewood Cliffs, New Jersey 07632, 1987.

[7] M. Niedźwiecki "On time and frequency characteristics of weighted least squares estimators applied to nonstationary system identification" Proceedings of 24th Conference on Decision and Control, Fl. Lauderdale, Fl - December 1985, pp. 225-230.

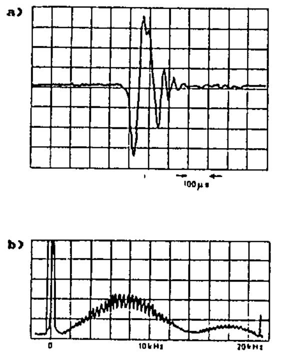


Fig. 1. Typical waveforms (fig. a,c) and spectra (fig. b,d) of an elementary click (fig. a,b) and a complex click (fig, c,d) for a Shure M75ED2 pick-up cartridge [1].

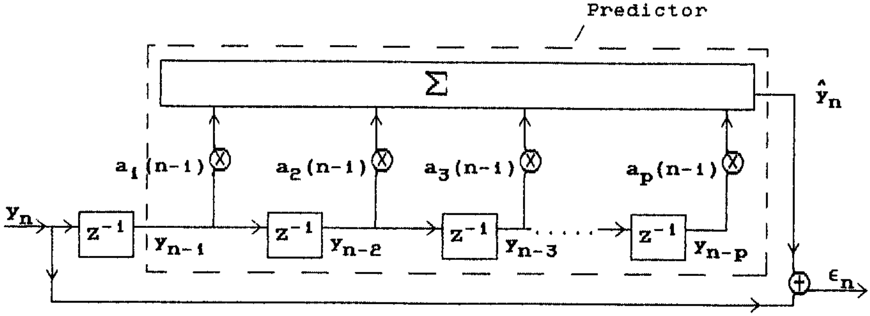


Fig. 2. Adaptive whitening filter [2].

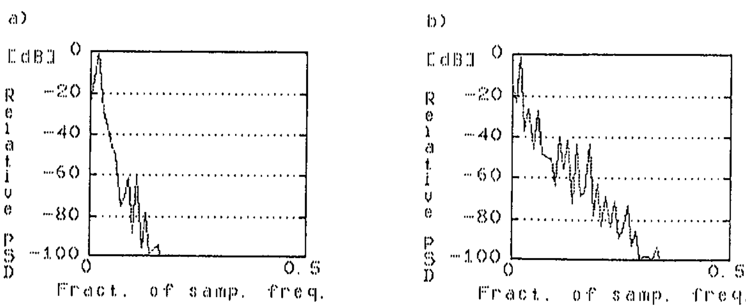


Fig. 3. Averaged periodograms of the investigated audio signals (1024 - point periodograms averaged over 10 adjacent frequency bins) [6]: a) piano solo (clasic composition), b) baroque orchestra.

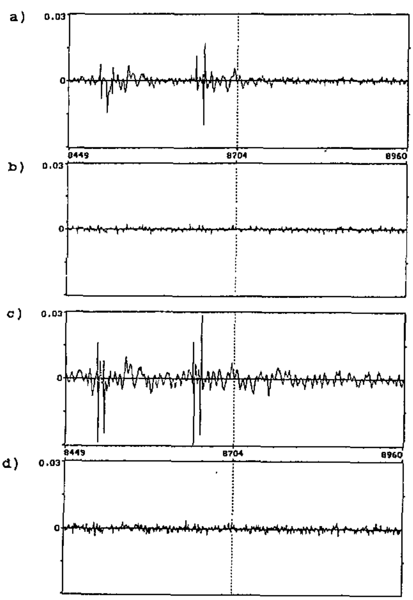


Fig. 4 Prediction errors obtained for the first test signal (piano) for two different values of the forgeting factor $\lambda$ (the order of AR model $p=10$): a) $\lambda = 0.95$, signal with disturbances, b) $\lambda = 0.95$, signal without disturbances, c) $\lambda = 0.99$, signal with disturbances, d) $\lambda = 0.99$, signal without disturbances.

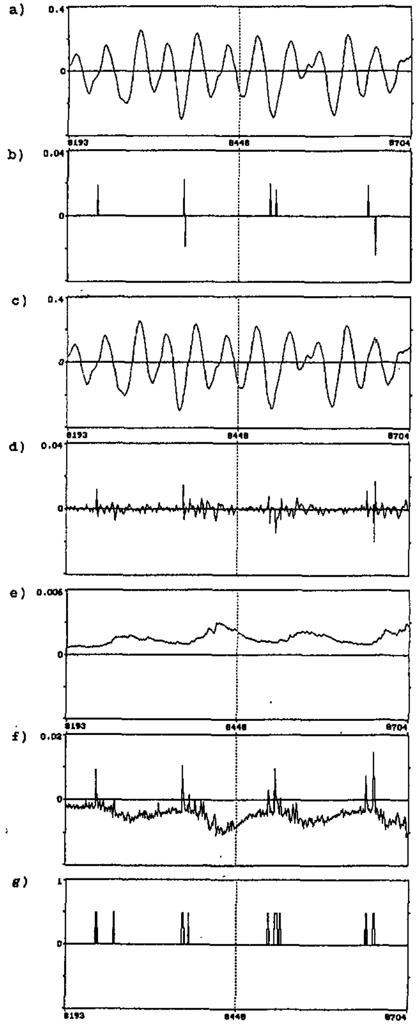


Fig. 5. Impulsive noise detection for the first test signal ($p=10$, $\lambda=0.95$, exponentially weighted estimation of standard deviation with $\mu=0.95$): a) signal without disturbance, b) the impulsive noise, c) signal after adding the impulsive noise, d) the prediction error signal observed at the output of the adaptive filter, e) the estimate of the mean standard deviation, f) difference of the absolute value of the prediction error and the multiple ($k=4$) of the standard deviation, g) the "zero-one" signal at the output of the detector.

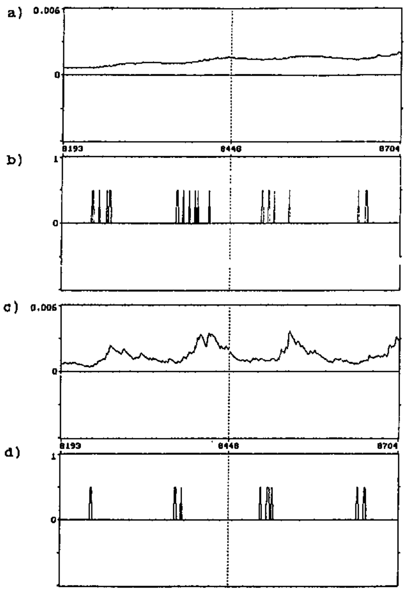


Fig. 6. Impulsive noise detection for the first test signal for two different values of the forgetting factor $\mu$ ($p=10$, $\lambda=0.95$): a) estimate of the standard deviation, $\mu=0.99$, b) the "zero-one" signal at the output of the detector for $\mu=0.99$, c) estimate of the standard deviation, $\mu=0.9$, d) the "zero-one" signal at the output of the detector for $\mu=0.9$.

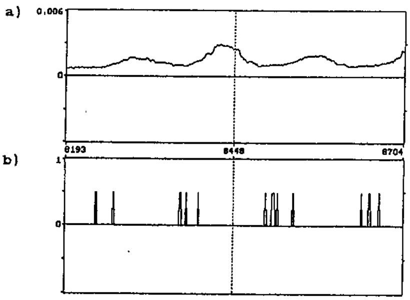


Fig. 7. Impulsive noise detection for the first test signal under sliding window ($M = 40$) estimation of the standard deviation of the prediction error ($p=10$, $\lambda=0.95$): a) estimate of the standard deviation, b) the "zero-one" signal observed at the output of the detector.

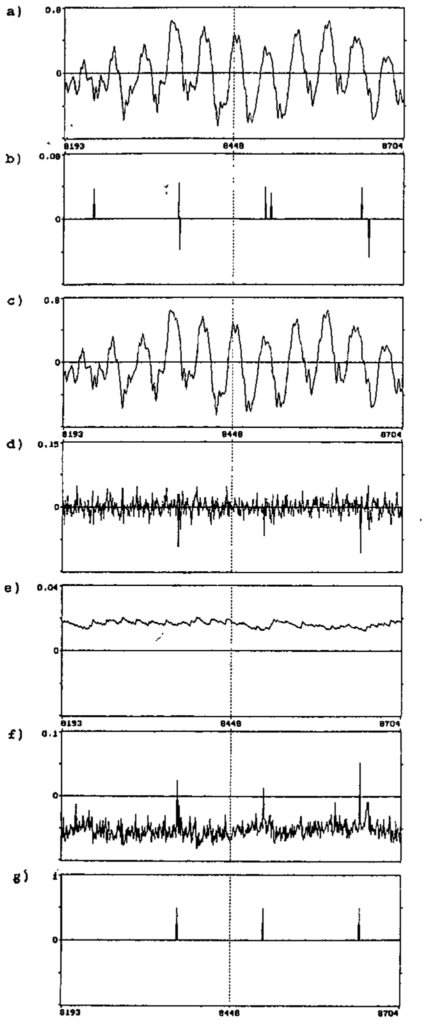


Fig. 8. Impulsive noise detection of the second test signal ($p=10$, $\lambda=0.95$, exponentially weighted estimation of the standard deviation with $\mu=0.95$): a) signal without disturbance, b) the impulsive noise, c) signal after adding the impulsive noise, d) the prediction error signal observed at the output of the adaptive filter, e) the estimate of the standard deviation, f) difference of the absolute value of the prediction error and the multiple ($k=4$) of the standard deviation, g) the "zero-one" signal at the output of the detector.