# A DOUBLE-THRESHOLD-BASED APPROACH TO IMPULSIVE NOISE DETECTION IN AUDIO SIGNALS

*Paulo A. A. Esquef, Luiz W. P. Biscainho, Paulo S. R. Diniz and Fábio P. Freeland*
COPPE/PEE & EE/DEL, UFRJ,
CP 68504, CEP 21945-970, Rio de Janeiro, RJ, BRAZIL
Tel: +55 21 2605010; fax: +55 21 5900788
e-mail: {esquef,wagner,diniz,freeland}@lps.ufrj.br

## ABSTRACT

In this paper we review threshold-based algorithms for detection of impulsive noise in audio signals, and propose a new detection strategy that can increase their capability to detect impulsive disturbances without any remarkable increase in computational complexity. The main features addressed by this paper are the use of two thresholds, one for detection and the other for estimation of impulse durations; the merging of disturbances detected as adjacent, under certain circumstances; and the alteration of the threshold level during the re-processing procedure of a given signal block. Experimental results assessed by human subjects as well as objective measurements confirm the improvements attained by the proposed detection scheme.

## 1 INTRODUCTION

One of the most common problems associated with historical musical recordings is degradation of the audio signal by impulsive disturbances. Generically, the impulsive disturbances or clicks can be described as localized discontinuities of short duration (typically less than 1 ms) that randomly corrupt an underlying audio signal. In most cases it can be assumed that no more than 10% of the signal samples are affected by the impulsive noise degradation [1].

Digital processing techniques for suppression of clicks from discretized audio signals can be in general separated in two stages: Detection and Reconstruction. In the Detection part, clicks are located in time. The Reconstruction part is supposed to replace the corrupted samples previously detected by other better representative of the underlying audio content [1].

The final quality of the restored audio signal depends, of course, on the performance of both stages. However, in most cases, a simple Reconstruction algorithm like LSAR (Least Squares Autoregressive) Interpolation [1] achieves excellent results, since the Detection stage is carefully performed. The ideal target is to leave uncorrupted samples untouched and to reconstruct only the corrupted ones.

A block-processing threshold-based technique for impulsive noise detection (henceforth called conventional detection) was first proposed in [2]. In [1], it is described together with more general and accurate methods for detection/suppression of impulsive noise which apply statistical models to both signal and noise, such as Bayesian Detection and Markov-Chain Monte Carlo. However, the latter techniques improved accuracy implies higher computational cost, which may turn impractical on-line restoration using such methods. In this work, the conventional method is reexamined and a modified detection strategy is proposed in order to improve its performance, without any remarkable increase in computational complexity.

The remaining of the text is organized as follows. In Section 2, the conventional technique for impulsive noise detection is reviewed and some of its limitations are pointed out. In Section 3, a modified detection scheme is proposed. In Section 4, efficiency of the propositions is evaluated. Conclusions are drawn in Section 5.

## 2 CONVENTIONAL IMPULSIVE NOISE DETECTION

In the conventional method [1], the corrupted audio signal, $y(k)$, is segmented in blocks of $N$ samples. Each block is modeled as an order-$p$ autoregressive (AR) process, $x(k)$, additively corrupted by impulsive noise, $d(k)$, as follows:

$$y(k) = x(k) + d(k), \qquad (1)$$

where

$$x(k) = \sum_{j=1}^{p} a(j)x(k-j) + e(k), \quad k = p, p+1, \cdots, N-1 \qquad (2)$$

being $a(j)$ the model parameters and $e(k)$ the excitation signal associated to $x(k)$.

Disturbances detection is performed in the excitation signal, where clicks become more evidenced than in the signal itself [2]. To obtain the excitation, it is necessary first to estimate the AR parameters and then pass the corrupted signal through the inverse filter

$$I(z) = 1 - \sum_{j=1}^{p} a(j) z^{-j}. \qquad (3)$$

The detection criterion consists of taking as corrupted all those samples in the signal that correspond to excitation samples whose magnitudes exceed a threshold $\lambda$, computed as

$$\lambda = K\,\hat{\sigma}_e, \qquad (4)$$

where $\hat{\sigma}_e$ is the estimated value of the excitation standard-deviation [1].

The next step consists in replacing corrupted samples by an interpolation scheme [2, 1, 3].

The main limitation related to the above criterion is the difficult trade-off between missing and false detections, considering the non-stationarity of audio signals, the variety of disturbances in amplitude as well as in duration, and that the value of $K$ is to be held fixed along the entire signal processing. As consequence, some disturbances are not detected or detected with underestimated duration, leading to inappropriate corrections in the Reconstruction stage.

The ideas presented in the next Section intend to improve robustness of the conventional detection without any remarkable increase in computational complexity.

## 3 THE MODIFIED DETECTION SCHEME

The performance of the threshold-based detection method can be improved by changing the test for disturbance presence. The proposed detection strategy involves using two-thresholds, merging adjacent disturbances and altering the threshold level along a possible re-processing scheme in a given signal block.

### 3.1 Use of two thresholds

The immense variety of magnitudes and durations of the impulses and their spread in the excitation signal (caused by inverse filtering), where they are detected and located, turns the threshold-based detection a quite difficult job, as described in Section 2. In general, conventional detection can lead to underestimation of click durations and to bad reconstruction of the signal, resulting in perceptible artifacts. Simply decreasing the threshold value to minimize this effect would necessarily increase occurrences of false detections.

One way to obtain a better trade-off between false and missing detections is to reformulate the detection criterion. In the modified scheme, the original threshold is renamed Detection Threshold, $\lambda_D$, and a reduced one called Location Threshold and defined by

$$\lambda_L = b\,\lambda_D, \qquad 0 < b < 1, \qquad (5)$$

is used only to estimate disturbances locations (see Fig. 1).

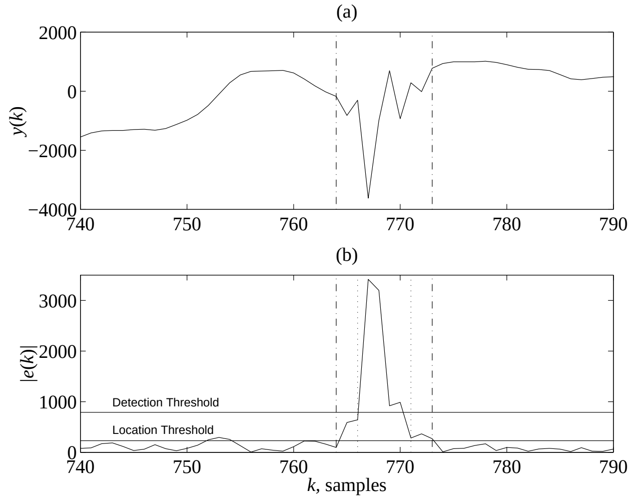


Figure 1: (a) Original noisy signal extract—a click can be viewed between the vertical dash-dot lines. (b) Detection and Location thresholds applied on the corresponding excitation magnitude—the vertical dotted lines indicate the click location by using only $\lambda_D$ while the dash-dot ones its location by using both $\lambda_D$ and $\lambda_L$.

In the two-threshold strategy, disturbances in a given block are detected from the one with the highest to the one with the lowest magnitude by following the steps below:

1. Select the sample with the highest magnitude exceeding $\lambda_D$ in the excitation signal;
2. Select the set of contiguous samples containing the previously selected one whose magnitude values are higher than $\lambda_L$;
3. Map the indexes of excitation samples previously selected to the corresponding signal indexes.
4. Set to zero the value of the selected excitation samples to allow detection of further clicks in the block;

Once all the disturbances in a given block are located, an algorithm for simultaneous reconstruction like the LSAR [1] method can be used.

### 3.2 Joining adjacent disturbances together

Inverse filtering of the noisy signal can lead to destructive or constructive interference inside a given disturbance in the excitation signal [1]. In the case of destructive interference, some intermediate samples of a relatively long disturbance can exhibit very low magnitudes. When it occurs, as illustrated in Fig. 2(b), the detection criterion, even using a two-threshold scheme, can find two or more disturbances separated by supposedly uncorrupted samples (which, in fact, are not), instead of only one disturbance.

We found out that an efficient way to solve this problem is forcefully join disturbances detected as adjacent

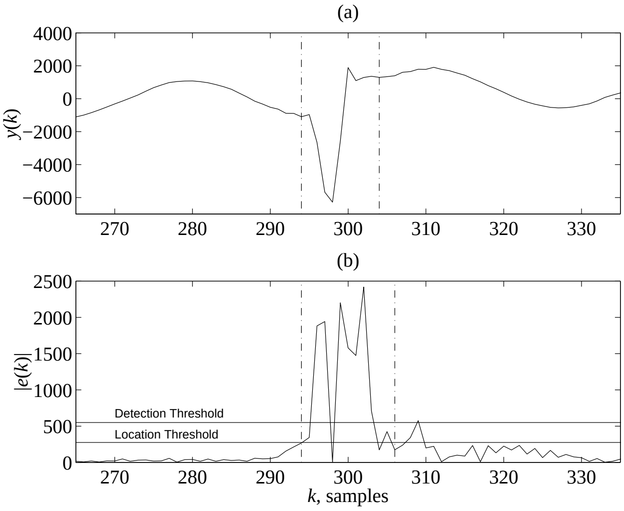


Figure 2: (a) Original noisy signal extract—a click can be viewed between the vertical dash-dot lines. (b) Detection and Location thresholds applied on the corresponding excitation magnitude—the vertical dash-dot lines indicate the click location by using both $\lambda_D$ and $\lambda_L$ and additionally joining adjacent clicks together when separated by 2 samples.

into one, when separated by less than a given number $n$ of samples. This implies a slight modification to the second step of the algorithm described in Section 3.1, which now turns to:

2. Examine the contiguous samples surrounding the previous selected sample, and delimit this particular disturbance by $n$ consecutive samples which magnitudes are lower than $\lambda_L$;

Typically, making $n$ equal to 2 or 3 samples (assuming a sample rate of 44100 Hz) is a suitable choice to prevent detection problems caused by destructive interferences.

### 3.3 Re-processing a given noisy signal block

In the presented block-based method, a single processing run may not suffice to detect all clicks in a block (especially those with lower magnitudes) and better detection performance can be attained by iterating processing. Of course, the re-processing strategy must include criteria to stop processing a given signal block and proceed to the subsequent one.

As the Reconstruction stage minimizes the excitation samples corresponding to the already suppressed disturbances, it can be expected that reconstructed portions of the signal will not be detected again in a new run of the Detection stage—under same processing parameters. Therefore, it seems reasonable to interrupt iterations when all excitation samples exhibit magnitudes below the Detection Threshold. However, there is no guarantee that this state will be reached in all blocks. In fact, low magnitudes disturbances may be detected with underestimated durations; attempts to restore them may not produce minimization of the corresponding excitation signal, so that they continue to be indefinitely detected. To overcome this we can properly decrease the Location Threshold depending on the performed number $i$ of iterations, as follows:

$$\lambda_{L_i} = b_i\,\lambda_D, \quad i = 1, 2, \ldots, i_{max}, \tag{6}$$

where

$$b_i = r^{\left\lfloor \frac{i-1}{f} \right\rfloor}\, b_1, \tag{7}$$

$r$ is a reduction factor, $f$ is a parameter that controls the decay rate of $b$ and $i_{max}$ is a maximum number of iterations, after which the algorithm proceeds to the subsequent block. Of course, the choice of parameters $i_{max}$, $f$ and $r$ is arbitrary and determined by experimental observations.

## 4 EXPERIMENTAL RESULTS

Comparative performance evaluation confronting conventional and modified detection methods was performed on a set of audio signals artificially corrupted by impulsive noise. Two classes of impulsive noise were used to artificially corrupt a set of audio signals. Class 1 noise contains clicks which corrupts approximately 0.6% of the audio signals samples, while Class 2 noise corrupts approximately 5%. Noise amplitudes were adjusted according to each audio signal power in order to make noise audible. The Reconstruction method used in all cases has been the LSAR as described in [1].

The modified detection, called MOD, employed the parameters shown in Table 1. In each block, the estimate $\hat{\sigma}_e$ was made proportional to the median value of the excitation samples magnitudes.

Table 1: Parameters used in modified detection method.

| $N$ | $p$ | $K$ | $b_1$ | $r$ | $n$ | $f$ | $g$ | $i_{max}$ |
|---|---|---|---|---|---|---|---|---|
| 1024 | 40 | 5 | 0.5 | 0.5 | 3 | 3 | 6 | 7 |

In fact, the conventional detection, called CONV, was implemented as a particular case of the MOD detection. In CONV, the use of only one Detection Threshold was attained by setting $b_1 = r = 1$ and the exclusion of any criterion to join adjacent clicks by setting $n = 1$.

In this work, originally uncorrupted signals were made available to allow the use of additional quantitative measurements, like missing and false detection ratios, as a way to confirm the subjective results obtained.

Define $n_o(k)$ as the corrupting noise added to each audio signal and $n_p(k)$ as the residual noise after processing, obtained as the difference between the restored signal and its uncorrupted version. Missing detection percentages can be taken as the ratio between non-zero samples in $n_o(k)$ that remain untouched in $n_p(k)$ and total non-zero samples in $n_o(k)$. False detection percentages can be taken as the ratio between zero samples in $n_o(k)$ that become nonzero in $n_p(k)$ and total zero samples in $n_o(k)$.

Table 2 shows comparative results of missing and false detection percentages obtained with MOD and CONV versions of the detection method[1]. Signals 1 to 4 were corrupted by Class 1 noise, while Signals 5 and 6 were corrupted by Class 2 noise. It must be emphasized that the parameter values were maintained the same to process all signals.

Table 2: Comparative measurements between conventional and proposed methods.

| | Missing Detection (%) | | False Detection (%) | |
|---|---|---|---|---|
| | MOD | CONV | MOD | CONV |
| Signal 1 | 1.56 | 16.41 | 3.26 | 1.17 |
| Signal 2 | 2.10 | 20.39 | 1.56 | 0.68 |
| Signal 3 | 1.56 | 10.71 | 1.83 | 0.87 |
| Signal 4 | 4.21 | 21.52 | 2.40 | 0.93 |
| Signal 5 | 1.63 | 14.86 | 6.65 | 3.54 |
| Signal 6 | 1.85 | 12.11 | 8.00 | 4.45 |

Perceptually, all restored signals obtained by using MOD were indisputably superior to those obtained by using CONV detection.

The more homogeneous subjective quality attained by MOD detection results for fixed values of the processing parameters indicates that the proposed detection method improve robustness of the method as to variations in signal and noise characteristics.

According to Table 2 MOD measurements against CONV ones indicate reduction by a factor between 5 and 10 in missing detection index—this fact justifies the best subjective results. It is also observed an increase by a factor between 2 and 3 in false detection index, which does not remarkably affect audio quality.

As a visual example, Fig. 3 confronts CONV and MOD detection after complete restoration procedures having been applied to the same signal shown before in Fig. 2(a).

Although the results in this section deal with artificially corrupted audio signals, restoration performed with CONV detection on naturally corrupted audio signals produce equally satisfactory perceptual results.

[1] Audio samples can be found at http://www.lps.ufrj.br/audio/

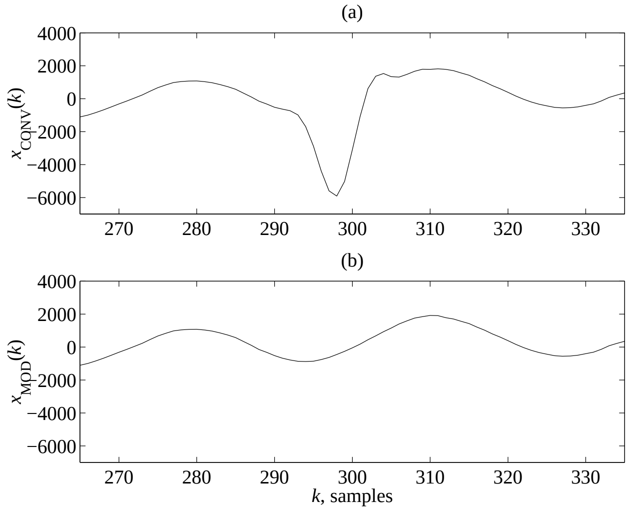


Figure 3: Restoration of the signal shown in Fig. 2: (a) Using CONV detection method. (b) Using MOD detection method.

## 5 CONCLUSIONS

This paper presented a new strategy to detect impusive noise in audio signals. In the proposed scheme the clicks are detected by a four-step algorithm that deals with two-thresholds. A simple rule to merge adjacents clicks is also incorporated to the scheme, as a way to prevent detection failure due to possible destructive interference on the excitation signal, where clicks are detected. Additionally, a block re-processing strategy is presented aiming to refine detection performance. Restoration applied to naturally or artificially corrupted real audio signals using the proposed detection method demonstrates an improvement in perceptual quality of the restored signals over those obtained by the conventional method. Quantitative measurements indicate a significant reduction of missing detections, justifying better subjective results, in spite of a slight increase in false detections. The modified detection scheme seems to be more robust to variations in audio and noise signals characteristics and does not imply any remarkable increase in computational complexity compared to the standard one.

## References


[1] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration.* London, UK: Springer-Verlag, 1998.

[2] S. V. Vaseghi, *Algorithms for Restoration of Archived Gramophone Recordings.* Ph.D. dissertation, University of Cambridge, Cambridge, UK, Feb. 1988.

[3] J. J. K. Ó. Ruanaidh and W. J. Fitzgerald, "Interpolation of missing samples for audio restoration," *Electronic Letters*, vol. 30, Apr. 1994.