METHODS FOR DETECTING IMPULSIVE NOISE IN SPEECH AND AUDIO SIGNALS

Ismo Kauppinen

University of Turku, Department of Physics, FIN-20014 Turku, Finland
iska@utu.fi

Abstract: In this paper computationally efficient methods for detecting non-Gaussian impulsive noise in digital speech and audio signals are presented. The aim of the detection is to find the errors without false detections in the case of e.g. percussive sounds in music signal or stop-consonants in speech signal. Various methods for computing a detection signal and a threshold curve are studied and tested. The detection can be applied in real time to a digital data stream.

1. INTRODUCTION

Disturbing impulsive noise occur in analog transmission media such as communications, telephony, and radio broadcast caused by electromagnetic disturbances and atmospheric noise. Impulsive type noise result also from physical damages to the storage media e.g. degraded gramophone recordings.

The restoration of the corrupted signal can be divided in two stages: detection and correction. For impulsive type of degradation the signal processing should be applied only to the corrupted samples to avoid unnecessary processing of noiseless samples. The exact locations of the damaged samples must be known prior to applying any signal restoration procedure.

Two different types of approaches for detecting impulsive errors in audio signals are presented in the literature: threshold based [1]-[3] and statistical model based [4]-[7]. The threshold based approaches are accomplished by first processing the noisy signals to enhance the detectability of noise pulses. Linear prediction is commonly used for the enhancement by transforming to the excitation domain of the linear prediction model. The statistical methods apply statistical models to both signal and noise. A survey of several statistical methods of detecting abrupt changes in signals is given in [8]. Also neural networks have been used to detect abrupt changes in signals [9]. Adaptive median hybrid filters have been used to detect impulsive type interference in speech signals [10].

In this paper we present new threshold based block-processing methods for finding the locations of audible noise pulses in audio signals with low computational complexity. These methods can be used as a front end for missing data interpolation algorithm.

2. THE DETECTION

When the clean signal x_n is corrupted by additive impulsive noise u_n, the corrupted signal model is given by

$$y_n = x_n + u_n. \tag{1}$$

The detection of impulsive noise bursts is carried out by forming a detection signal with peaks indicating the locations of the disturbances. The peaks are judged to represent impulsive noise by an adaptive threshold curve.

2.1. The derivative method

Since the characteristic of an impulsive error is a sudden fast change, one way to form the detection signal is to observe the difference between successive samples, i.e. the discrete derivative of the signal. The derivative d_n of the discrete signal y_n can be obtained, according to the definition of the derivative, by dividing the difference of successive samples by the sampling interval Δt, i.e.

$$d_n = \frac{y_{n+1} - y_n}{\Delta t}. \tag{2}$$

This gives the derivative of the signal in a middle point between the two samples. To form a detection signal with peaks at the positions of the noise pulses in the signal, an absolute value of the derivative signal must be taken. To further enhance the detectability of the peaks, the differentiation can be applied several times. In practice it has been found that the fourth derivative is accurate enough to find the smallest audible errors in the signal. The derivative detection signal is given by

$$g_n^{\mathrm{DER}} = \frac{|y_{n-2} - 4y_{n-1} + 6y_n - 4y_{n+1} + y_{n+2}|}{(\Delta t)^4}, \tag{3}$$

which is obtained by applying Eq. (2) four times in succession, compensating the shift resulting from the differentiation, and finally taking an absolute value (see Fig. 1). Because the amplitude envelope of the detection signal follows the changes of the temporary loudness in the original signal, the use of a constant threshold level would not give satisfying results. Percussive instruments usually make the amplitude envelope of the detection signal rise (see Fig. 2). An adaptive threshold can be obtained by averaging the detection signal around each sample point and multiplying by a suitable threshold scaling factor:

$$v_n^{\mathrm{AVG}} = \frac{k}{2i+1} \sum_{m=n-i}^{n+i} g_m^{\mathrm{DER}}, \tag{4}$$

where k is the threshold scaling factor and the length of the average is $2i + 1$. The threshold parameter k is set experimentally for each signal. The length of the average must be relatively long for the average value to remain low enough at a position of a peak in the detection signal.

0-7803-7503-3/02/$17.00 ©2002 IEEE

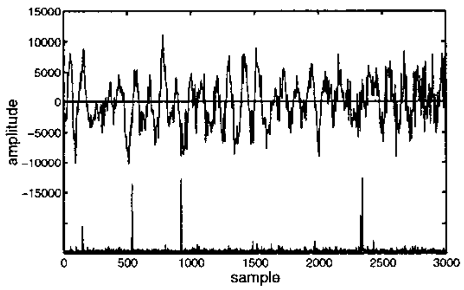

Fig. 1. The derivative detection signal formed by taking the absolute value of the fourth derivative of the signal.

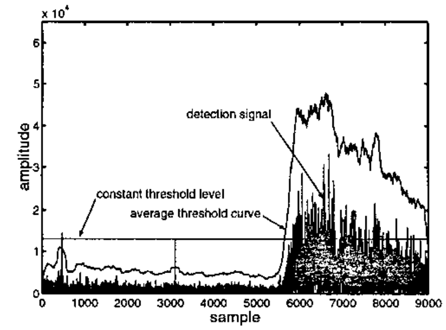

Fig. 2. The derivative detection signal with constant and average threshold functions. The music signal contains two impulsive errors, and a snare drum hit.

2.2. The RMS method

One way to detect sudden changes in the signal is to observe the RMS (Root Mean Square) value of the signal given by

$$r_n = \sqrt{\frac{1}{2i+1}\sum_{m=n-i}^{n+i} y_m^2}, \tag{5}$$

where $2i+1$ is the RMS length. The RMS as such is not a very good detection signal. However, significant improvement can be achieved by taking the absolute value of its fourth derivative. The RMS detection signal

$$g_n^{\mathrm{RMS}} = \frac{|r_{n-2} - 4r_{n-1} + 6r_n - 4r_{n+1} + r_{n+2}|}{(\Delta t)^4} \tag{6}$$

gives two peaks for each disturbance, since it measures changes in the loudness of the signal. The first peak results from the sudden increase of loudness, and the second peak from the loudness decreasing. The disturbance is located in the middle of the two peaks. The distance between the two peaks is the RMS length plus the width of the error (see Fig 3). Problems with this method will occur, if the distance between impulsive errors is smaller than the RMS length. In this case, the start and the end peaks of the errors can get mixed up. The same type of threshold curve described in the derivative method can be applied to the RMS detection signal.

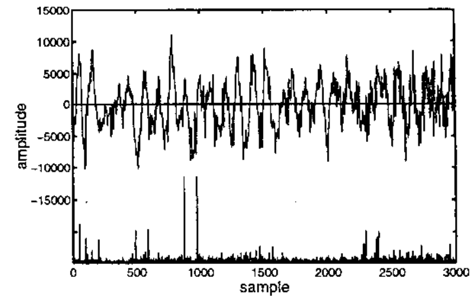

Fig. 3. The RMS detection signal plotted beneath the same music signal as in Fig. 1. The RMS length was 101.

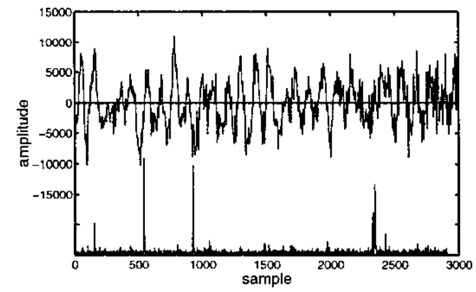

Fig. 4. The AR detection signal ($p = 10$) plotted beneath the same music signal as in Fig. 1 and 3.

2.3. AR-model based approach

A Gaussian Auto-Regressive (AR) model is assumed for the underlying signal

$$x_n = -\sum_{k=1}^{p} a_k x_{n-k} + \sigma_n, \tag{7}$$

where a_k are the AR model coefficients and σ_n is noise-like signal ideally uncorrelated and statistically independent of x_n. Residual signal e_n can be obtained by estimating the model parameters a_k and passing the corrupted signal y_n through an inverse filter given by

$$H(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}, \tag{8}$$

where most of the stationary signal components are reduced, leaving the impulsive errors more conspicuous. To model the audio signal properly, the order p should be very large (around 1000) [11]. However, for the detection of impulsive disturbances a much lower order (around 10) will give good results. The detection signal based on the residual of the AR model is given by (see Fig. 4)

$$g_n^{\mathrm{AR}} = \frac{|e_{n-2} - 4e_{n-1} + 6e_n - 4e_{n+1} + e_{n+2}|}{(\Delta t)^4}. \tag{9}$$

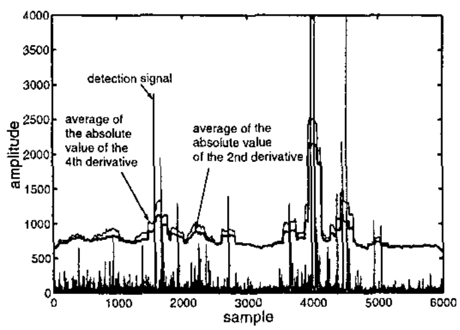

Fig. 5. Better adaptivity is obtained, if the threshold curve is formed by using the second derivative.

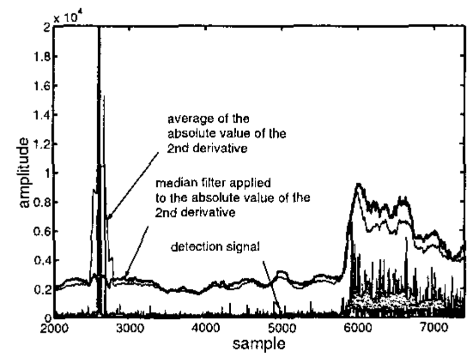

Fig. 6. By using the median threshold curve, the threshold level can be set still much higher than if the threshold curve was obtained directly from the second derivative.

3. IMPROVING ADAPTIVITY

3.1. Using the average of the lower order derivatives as the threshold curve

Improved adaptivity is obtained if the threshold curve is calculated from lower than the fourth order derivative of the signal. In lower order derivatives the peaks are weaker. As a result, the average will remain lower at the positions of the peaks and a higher threshold level can be chosen. Hence, smaller peaks that are not caused by impulsive noise, are not causing false detections. Because of the dispersive nature of the odd order derivatives, even order derivatives should be used. When the detection signal is formed by using the fourth derivative, the second derivative should be used to obtain the threshold curve. The next step would be to use the zeroth derivative (i.e. the signal itself), but this has been proven not to give good results. In Fig. 5, the average threshold curve obtained from the second derivative is compared to the average threshold obtained from the fourth derivative.

3.2. The median filter

The median filter is a nonlinear signal enhancement technique for smoothing of signals. The median filter computes the pointwise median of the signal. The median of a set is defined as the middlemost value of an ordered table of the set values. Our aim is to use the median filter for obtaining an adaptive threshold curve by applying the median filter to the detection signal. The median threshold curve is given by

$$v_n^{\mathrm{MEDIAN}} = k \cdot \mathrm{median}(w_n), \qquad (10)$$

where k is a scaling factor, w_n is a subset of the detection signal sequence given by

$$w_n = \{g_{n-i}, \dots, g_{n-1}, g_n, g_{n+1}, \dots, g_{n+i}\}, \qquad (11)$$

and $2i+1$ is the length of the median filter.

If the length of the median filter is longer than the size of the impulsive error (i.e. approximately the length of the peak width in the detection signal), then the peak will not cause the median filtered threshold curve to rise at the position of the peak. Transients in the music signal result in a group of peaks in the detection signal, but the length of the peak group is usually much longer than the section corrupted by impulsive noise burst, and the threshold curve will rise at the position of the transient without causing false detection.

The median filter gives better adaptivity to the threshold curve than the average. This is demonstrated in Fig. 6. The music signal contains a large click and a starting trumpet sound at the end of the file, which can be seen as higher amplitude envelope at the end of the detection signal. The median filter does not rise at the position of a peak in the detection signal, but rises rapidly at the beginning of the trumpet sound. Now the threshold level could be set higher without missing the click in the signal and being well overhead to the transients.

4. EXPERIMENTS

Three different methods presented in this paper and a similar type of method presented in the literature [1]-[2], were compared for performance by applying them to four different test signals with approximately 0.6-1.5 % of the signal samples corrupted by impulsive noise. The original signals were made available to obtain missing and false detection rates. The test signals were selected to challenge the detection algorithm. Test signal 1 (drums) contains playing of a basic drum set. Test signal 2 (git) contains electric guitar solo with rapidly changing notes and transients. Test signal 3 (speech) contains male speaking and test signal 4 (sax) contains playing of a saxophone. The detection signal for the first method, called DER, was obtained by Eq. (3). In the second method, called RMS, the detection signal was obtained by Eq. (6). In the third method, called AR, the detection signal is obtained by Eq. (9). In all the above methods the threshold curve is obtained by applying the median filter to the second derivative. In the fourth method, presented in papers [1]-[2], the detection signal is formed by filtering the residual of the AR model with a match filter, which is a filter matched to the AR inverse filter coefficients. In this method, here called MATCH, the threshold curve is the power estimate of the residual.

The length of the RMS in all the tests was 81, the size of the median filter was 201, and the data was processed in blocks of 2000 samples. The threshold level for DER, RMS, and AR was adjusted by multiplying with a threshold scaling factor k and in the MATCH method the threshold level was adjusted by adding a threshold offset a. The threshold parameter for each signal was optimized by experimental observations.

Tables 2. and 3. show the results of comparative missing and false detection tests. In terms of the audio quality of the restored signal, a missing detection results in much higher quality loss than a false detection. When missing detection occurs, the click remains in the restored signal, but in the case of false detection the quality loss is depended on the performance of the interpolation algorithm. The AR method gives clearly the best results. Also relatively good results are obtained from the DER method. The RMS method performs poorly on the guitar solo. This is due to the undesired modulation effect in the threshold curve which results from the modulation of the RMS signal.

Table 1. Missing detection rate comparison.

Signal	DER%	RMS%	AR%	MATCH%
drums	19,4	41,9	6,5	41,9
git	7,7	65,4	0,0	34,6
speech	14,3	49,0	8,2	42,9
sax	13,6	22,7	0,0	54,5

Table 2. False detection rate comparison.

Signal	DER%	RMS%	AR%	MATCH%
drums	16,1	9,7	9,7	74,2
git	26,9	15,4	15,4	11,5
speech	16,3	10,2	4,1	22,5
sax	4,5	9,0	0,0	45,5

5. CONCLUSION

In this paper methods for detecting impulsive noise pulses in speech and audio signals were studied. The detection is carried out by forming a detection signal (with peaks indicating the error locations) and an adaptive threshold curve. Three different methods for computing the detection signal were presented, and also various methods for obtaining the threshold curve were presented. The performance of three methods presented in this paper and a method, presented earlier in the literature, were compared by a comparative missing and false detection tests for corrupted speech and music signals.

Different combinations of the detection signal and the threshold curve can be used depending on the input signal. In general, for music signals the best results can be achieved by using the residual detection signal and the median threshold curve. However, the median filter is computationally more complex than the average.

These detection methods can also be applied to other digital signals e.g. image signals, communication channels, and EKG.

6. REFERENCES

[1] S. V. Vaseghi and R. Frayling-Cork, "Restoration of Old Gramophone Recordings," *J. Audio Eng. Soc.*, vol. 40, No. 10, 1992.

[2] S. V. Vaseghi and P. J. W. Rayner, "Detection and Suppression of Impulsive Noise in Speech Communication Systems," *IEEE Proc., communications, speech, and vision*, vol. 137, Pt 1, No. 1, 1990, pp. 38-46.

[3] P. A. A. Esquef, L. W. P. Biscainho, P. S. R. Diniz, and F. P. Freeland, "A Double-Threshold-Based Approach to Impulsive Noise Detection in Audio Signals," *Proc. EUSIPCO*, 2000, pp. 2041-2044.

[4] S. V. Vaseghi and P. J. W. Rayner, "The Effects of Non-Stationary Signal Characteristics on the Performance of Adaptive Audio Restoration Systems", *Proc. IEEE ICASSP*, 1989, pp. 377-380.

[5] S. J. Godsill and P. J. W. Rayner, "A Bayesian Approach to the Restoration of Degraded Audio Signals," *IEEE Transactions on speech and audio processing*, vol. 3, No. 4, 1995, pp. 267-278.

[6] A. F. M. Smith, "A Bayesian Approach to Inference About a Change-Point in a Sequence of Random Variables," *Biometrica*, vol. 62, No. 2, 1975, pp. 407-416.

[7] S. J. Godsill and P. J. W. Rayner, "A Bayesian Approach to the Detection and Correction of Error Bursts in Audio Signals," *Proc. IEEE ICASSP*, vol. 2, 1992, pp. 261-264.

[8] M. Basseville, "Detecting Changes in Signals and Systems - A Survey," *Automatica*, vol. 24, No. 3, 1988, pp. 309-326.

[9] C. L. Fancourt and J. C. Principe, "On the use of Neural Networks in the Generalized Likelihood Ratio Test for Detecting Abrupt Changes in Signals," *Proc. IJCNN*, 2000.

[10] A. Nieminen, P. Heinonen, and Y. Neuvo, "Suppression and Detection of Impulsive type Interference Using Adaptive Median Hybrid Filters," *Proc. IEEE ICASSP*, 1987, pp. 117-120.

[11] I. Kauppinen, J. Kauppinen, and P. Saarinen, "A Method for long extrapolation of audio signals", J. Audio Eng. Soc., vol. 49, No. 12, December 2001, pp. 1167-1180.