Two Parallel Filters with One-Dimensional Non-Local Mean Method for Noise Cancellation in Audio Signals

Ali S. Awad

Department of Engineering and Information Technology, Alazhar University, Gaza, Palestine

aawad@alumni.stevens.edu

Received 28/1/2018 *Accepted* 18/3/2018

Abstract:

Signal denoising is an important process and should be implemented before any advanced process on the audio signals. This paper proposes a two -stage detection method to detect the noisy samples effectively. In the first stage, two parallel filters are primarily implemented to determine the common noisy samples. In the second stage, intensity difference based method is employed in order to determine the truly noisy samples by filtering the common noisy samples. At the end, one-dimensional non local mean filter is deployed for estimating the truly noisy samples rather than estimating all the samples in the signal. Extensive simulation results show that the proposed method delivers superior results at high and low noise rates with respect to the other well-known denoising methods. Moreover, the suggested method is fast and consistent against the values of the suggested parameters.

Keywords: Impulse noise, Audio signal, Noise cancellation, Non-local mean

1 Introduction

Noise removal from the audio signal is very demanded in many applications such as; speech compression, signal processing in medical devices, mobile communication, speech enhancement, and wireless speech recognition. As a result, many algorithms are proposed in the literature for denoising the audio signal, some of them are median filtering-based approaches as those mentioned in [1-3]. The drawback of these filters is that they fail to differentiate between the noisy and the original samples, thus, many of the signal details will be lost. Spectral subtraction-based methods [4-5] are the most widely methods used in speech enhancement. In which, the enhanced speech is obtained by subtracting the estimated noise spectrum from the noisy speech spectrum. Nevertheless, these methods miscarry to provide a good estimation to the spectrum of the noisy signal. Other approaches are based on wavelet technique as described in [6-10]. In which the signal is decomposed into multi-levels and after then the noisy components is discarded. Many others are based on sparse representation [11-13] or mean square error [14-15]. In this paper, we are trying to remove impulse noise from corrupted audio signals efficiently by detecting and then filtering the noisy samples. Impulse noise which is one of the important and real noises that causes clicks, bursts or scratches in the audio signals [16-17]. This noise takes the same range of values between [-1,1] similar to the audio signal. Thus, it is difficult to differentiate between the original and the noisy values. To this end, two known filters which are the SD-ROM [18] and the ACWM [19] are used in the detection process. The principle of SD-ROM filter is based on calculating the Rank Order Mean, ROM, for the samples that surrounds the tested one. Then, the tested sample is considered noisy or not by comparing the ROM value with predefined thresholds. The principle of ACWM filter is based on calculating the differences between the tested sample and the outputs of center weighted median filters where each of them give different weight to the central sample. The SD_ROM is employed first for image restoration as described by the work presented in [20]. The

ACWM is used basically for image restoration. However, we intend to use it in this paper to restore noisy audio signals.

Two-stage detection filter is proposed, in this research, to detect the noisy samples spread in the audio signal. In the first detection stage, the SD-ROM and the ACWM filters are applied in parallel to the noisy audio samples. Then, the corresponding restored samples, attained by these methods, are compared to decide whether they are common noisy or original audio samples. If the corresponding samples have almost the same values they are considered original audio samples otherwise they are considered common noisy samples. The second stage is applied to filter the common noisy samples based on an intensity-based method. This stage helps deciding if the tested common noisy samples are truly noisy or not. Note that intensity-based approach is applied only on the common noisy samples but not all the audio samples. The reason for that is that we are trying to avoid any errors that would be produced if the detection process is implemented on all the samples in the corrupted sound. Eventually, one dimensional non local mean is applied to restore only the truly noisy samples not as the 2-D non-local mean stated in [21-23] which restore every value in the image. The advantage of this method is that several steps are used to detect the noisy samples; besides, a 1-D non-local mean is implemented to restore the noisy samples only. This in turns helps to keep the original sample intact and restore audio samples corrupted by noise rates greater than 30%.

The proposed paper is organized as the following: section 1 explains detection and restoration methods, section 2 presents simulation results. Finally, section 3 concludes this research.

2 Algorithm Description

2.1 Two-Stage Detection Process

In the two -stage detection process we are trying to determine the noisy samples that are spread in the corrupted audio signal for restoring them rather than restoring all the audio samples. To this end, we run ACWM and CD-ROM filters in parallel then compare their restored samples. More specifically, the samples that have almost similar values are considered original audio samples whereas the samples with different values are considered noisy and named as common noisy samples. Note that, it is preferable to compare between the restored samples rather than the detected noisy samples. The reason for that is that both filters have errors in detections. Another detection process is implemented over the common noisy samples in order to separate the samples that are most likely truly noisy. Finally, a 1-D non-local mean method is implemented to restore the truly noisy samples. For describing the first step in finding the truly noisy samples from a given noisy audio signal of length *N*, let the corrupted audio samples are denoted by S_{COR}, the restored samples obtained by ACWM filter are denoted by X_{ACWM} and the samples obtained by CD-ROM filter are denoted by $Y_{CD\text{-}ROM}$ as the following:

$$S_{COR} = (s_1\ \ s_2\ \ s_3\ \ s_4\ \ s_5 ......\ \ s_N) \qquad (1)$$

$$X_{ACWMF} = (x_1\ \ x_2\ \ x_3\ \ x_4\ \ x_5 ......\ \ x_N) \qquad (2)$$

$$Y_{CD\text{-}ROM} = (y_1\ \ y_2\ \ y_3\ \ y_4\ \ y_5 ......\ \ y_N) \qquad (3)$$

The corresponding samples in Eqs. 1 and 2 with similar values or very small intensity difference, ξ, are considered original otherwise they are considered common noisy samples. To this end, we calculate the intensity difference d_i between each two corresponding samples according to:

$$d_i = |x_i - y_i| \qquad i=1,2,3,....N \qquad (4)$$

if $d_i \le \xi$ then x_i ***and*** y_i considered original samples as $x_i = y_i = o_i$ otherwise they deemed as common noisy samples $x_i = y_i = n_i$.

Subtract the vectors in Eq, 2 and 3 and sort the attained intensity differences d_i's in ascending order, we will get the following:

$$sum\,|\,x_{ACWMF,i} - y_{CD\text{-}ROM,i}\,| = \sum_i d_i = \sum_j \xi_j + \sum_k \overline{\overline{d_k}} \qquad (5)$$

Where $\xi_j \le \xi$, $\overline{\overline{d}}_k > \xi$, $j=1:m-1$, and $k=m:N$. Take the logarithm of the inverse of each difference d_i to assess the performance of the proposed algorithm as:

$$\sum_i log_{10}(1/d_i) = \sum_j log_{10}(1/\xi_j) + \sum_k log_{10}(1/\overline{\overline{d_k}}) \qquad (6)$$

Let the outputs of the two filters are equals. Then

$$\sum_j log_{10}(1/\xi_j) = \infty \text{ and } \sum_k log_{10}(1/\overline{\overline{d_k}}) = 0 \qquad (7)$$

At very low noise rate the output samples mostly have similar values and with intensity differences d_i's less than ε. As a result $\sum_j log_{10}(1/\xi_j) >> \sum_k log_{10}(1/\overline{\overline{d_k}})$ and the performance of the proposed method will be clearly high. The reason is that the number of the common noisy samples is low and most of the samples are considered original. Another important point is that we should search about the maximum value of ε that would minimize eq. 6 but still delivers optimum performance as:

$$\underset{max\ \xi}{Min} \left\langle \sum_j log_{10}(1/\xi_j) + \sum_k log_{10}(1/\overline{\overline{d_k}}) \right\rangle \qquad (8)$$

Note that if the noise rate increases the noisy samples increases and the number of the intensity difference d_i which is greater than ε increases as well. In other words, at high noise rates we will get $\sum_j log_{10}(1/\xi_j) << \sum_k log_{10}(1/\overline{\overline{d_k}})$ and as a result the performance of the proposed filter is decreased. At the end, the proposed filter will fail to function when

$$\sum_k log_{10}(1/\overline{\overline{d_k}}) - \sum_j log_{10}(1/\xi_j) \approx \sum_k log_{10}(1/\overline{\overline{d_k}}) \qquad (9)$$

Now, define a new vector Z_{rest} , based on Eq. 4, to include two common noisy samples n_i and n_{i-2} , as an example, with the other original samples as:

$$Z_{rest} = (\ o_1\ o_2 \ldots n_{i-2}\ o_{i-1}\ n_i\ o_{i+1}\ o_{i+2}\ \ldots\ldots\ o_N) \qquad (10)$$

To increase the performance of the proposed algorithm, we implement the second detection method. In which, each common noisy sample n_i is detected one more time by comparing its intensity value with those of its surrounding samples in the vector S_{COR}. The length of the surrounding samples is determined based on the noise rate, in the sense that as the noise rate increases the length increases and vice versa. Let a noisy sample in the location i has the following surrounding samples attained from the vector S_{COR} as:

$$S_i=(s_{i-3}\ \ n_{i-2}\ \ s_{i-1}\ \ n_i\ \ s_{i+1}\ \ s_{i+2}\ \ s_{i+3}) \qquad (11)$$

if there is another noisy sample as n_{i-2} exits within the neighboring samples, it should be removed. Therefore, the result of the first norm $|S_i-n_i|$ will be as:

$$|S_i-n_i| =(| s_{i-3} -n_i|\ \ | s_{i-1}-n_i|\ \ | s_{i+1}-n_i|\ \ | s_{i+2}-n_i|\ \ | | s_{i+3}-n_i|) \qquad (12)$$

Now if there is a value in the vector $|S_i-n_i|$ somewhat small or less than a threshold value, then the noisy sample n_i is not strange and has a similar sample amongst its neighboring ones. In other words, the tested noisy sample n_i. is not a unique value. Thus, the sample n_i. is considered with minimum probability as original sample. Also, it is clear that as the number of the similar samples increases the originality of the tested noisy sample increases. This idea is explained by defining a function $f(j)$ as the following:

$$f(j)=\left(1-\left\lfloor\frac{|s_j-n_i|}{Th_1}\right\rfloor\right) \qquad j=i-k:i+k\ ,j\neq i \qquad (13)$$

$2k$ is the maximum length of the surrounding samples. Divide each element in the norm $| S_i-n_i|$ by a threshold value Th_1 and then take the round down function. Round down function means that any element greater than integer 1 is replaced by 1 and greater than 0 is replaced by 0. Thus $f(j)$ will take two elements either zero or 1. For example, let $k=2$ and $f(j)=\{0,1,0,0)$ that means the tested sample has one similar value and therefore considered as original sample o_i otherwise it will be flagged as F in the vector Z_{rest} as:

$$ni=\begin{cases} o_i & if\ \left(\sum_j f(j)\right)\geq 1 \\ F_i & else \end{cases} \qquad (14)$$

The sample that flagged as F will be restored by using a 1-D non-local mean method as shown in the following subsection.

2.2 One-Dimensional Non-Local Mean Filter

In this part, we are trying to find the most similar means of the surrounding patches with respect to the mean of the tested one. This is illustrated as the following. Let us assume that the output from the detection process as:

$$Z_{rest} =(o_1\ o_2\ o_3 \ldots F_8 \ldots o_{12}\ F_{13}\ o_{14} \ldots o_{N-1}\ o_N) \quad (15)$$

To restore the sample F_8, for example, we should define the search area A of a length L to be started at sample $a_1=8-w_1$ and ended at sample $a_2= 8+w_2$. For $w_1=w_2=7$, $a_1=1$, $a_2=15$ and $L=15$ as: :

$$A_8=(o_1\ o_2\ \ldots o_7\ F_8\ o_9\ o_{10} \ldots o_{12}\ F_{13}\ o_{14}\ o_{15}) \quad (16)$$

Let the length of the tested patch starts at sample $p_1=8-w_a$ and ends at sample $p_2=8+w_a$, for $w_a=2$, $p_1=6$ and $p_2=10$. Therefore, all the acquired patches will be as: tested patch **TP** $=(x_6\ x_7\ F_8\ x_9\ x_{10})$ with mean m_{Tp}. The other patches in the research areas will be as:

$$Tp1=(o_1\ o_2\ o_3\ o_4) ,Tp2=(o_2\ o_3\ o_4\ o_5),Tp3=(o_3\ o_4\ o_5\ o_6),$$
$$Tp4=(o_4\ o_5\ o_6\ o_7),Tp5=(o_5\ o_6\ o_7 F_8) \ldots TP12=(o_{12}\ F_{13}\ o_{14}\ o_{15})$$

to estimate the value of the noisy sample F_8 , the average of the most similar means of the different surrounding patches should be calculated. To this end, we calculate the mean of each patch m_{Tpi} , $i=1:12$. Note that F_8 and F_{13} are unknown samples and therefore they should be removed, and the mean is taking for the rest of the samples in the patch. The absolute differences D_i's between the mean of the tested patch m_{Tp} and the other means are attained as:

$$D_i=|m_{Tp}- m_{Tpi} | \quad (17)$$

. Let $m_{Tp1}\ m_{Tp2} \ldots m_{Tpk}$ be the most similar means to m_{Tp} , where their relevant difference D_i with respect to m_{Tp} is within a pre- defined threshold value Th_2 , i.e., $D_i \leq Th_2$. Finally, the estimated value will be as:

$$o_{est,8}= mean(m_{Tp1}\ m_{Tp2} \ldots m_{Tpk}\ m_{Tp}) \quad (18)$$

then we continue performing the same previous steps for estimating the other flagged noisy samples.

3 Simulation Results

In order to study the performance of the proposed filter, extensive experiments are implemented on different audio signals corrupted with random valued impulse noise at different noise rates. Many factors are used for evaluating the proposed filters performance such as; the a) visual quality of the restored samples compared with the original ones, b) Peak Signal to Noise Ratio, PSNR in decibel (dB), c) consumed time, d) output results against the values of the suggested parameters. In the simulated experiments, the visual quality, either for all the restored samples or for part of them, is shown in the following figures. But, PSNR is calculated for all the restored samples in the audio files. Note that the restoration results are improved gradually through the proposed filter; in the first step the common noisy samples are specified using ε=0.05 and then filtered to determine in more accurate way the truly noisy samples by using intensity difference-based method with threshold Th_1=20 and surrounding samples of length $k=2$.

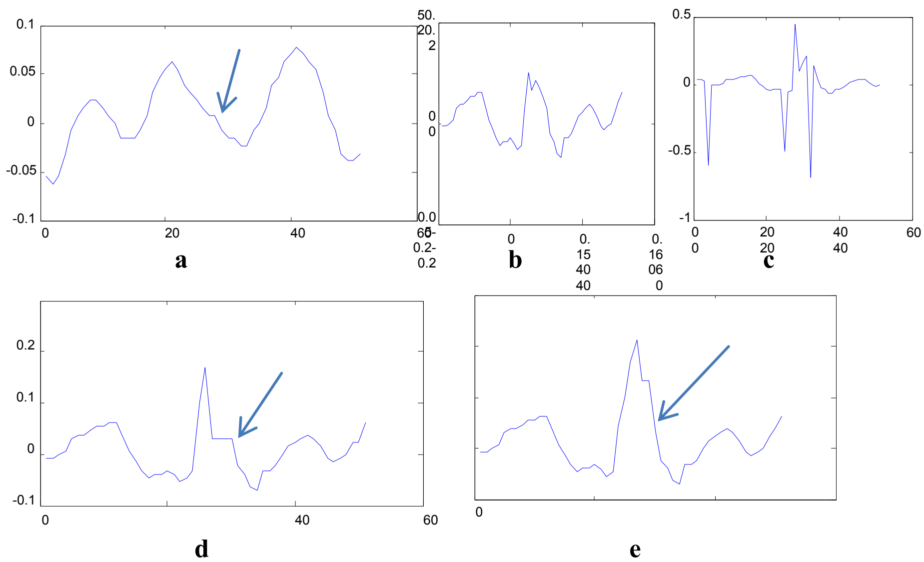

Fig.1 Restoration performance for 50 samples corrupted with 15% noise rate taken from the FlagRaising.wav sound file. A) original sound, b) corrupted version, c) CD-ROM filter,PSNR= 29.3813,d) ACWM filter ,PSNR= 31.6962,e) proposed method, PSNR= 33.6804 , Th_2=5

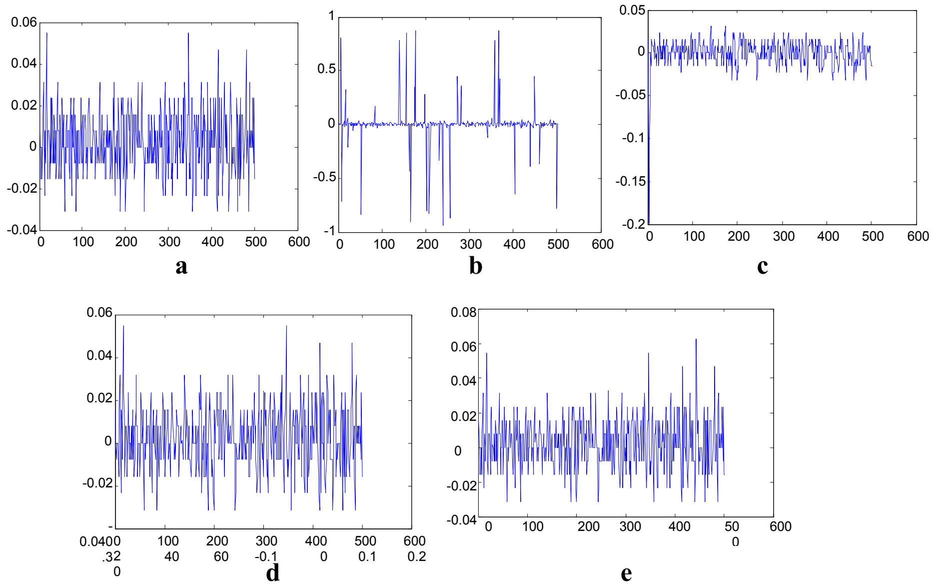

Fig.2 Restoration performance for 500 samples corrupted with 5% noise rate taken from the rain fall sound file. A) original sound, b) corrupted version, c) CD-ROM filter ,PSNR=33.4628\, d) ACWM filter ,PSNR=37.4474, e) proposed method, PSNR=39.27,$Th_2=5$

Finally, a 1-D non-local mean filter is applied on the truly noisy samples with tested patch length of $w_a=2$, search area A of length $L=14$, and threshold value Th_2 to determine the similar means. All the above variables produce almost the same outputs within a specified range as will be shown later. Fig 1, shows visibly, in a clear view, how the adopted algorithm performs better results than the other known algorithms, particularly where the arrow is pointed. It is clear that the proposed algorithm detects some errors found in the output of the ACWM filter, which in turns improve the final results. Fig.2 shows the restoration results for different filters materialized due to the restoration of 500 noisy samples taken from the rainfall.wav audio file. Low noise rate, 5% , is used in this experiment to illustrate clearly the improvement in the restored versions. It is clear that the proposed filter delivers the best performance compared to the others in terms of the visual quality of the samples amplitudes or in terms of PSNR.

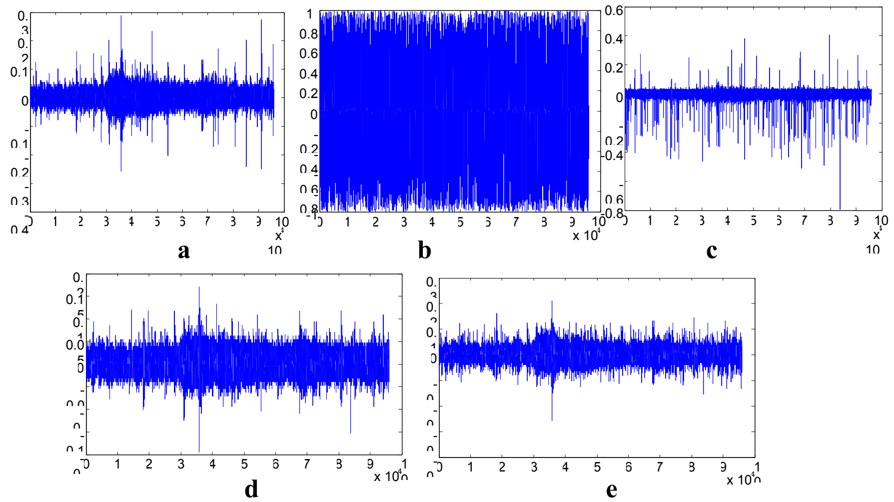

Fig.3 Restoration performance for the rain fall sound file corrupted with 10% noise rate. A) original sound, b) corrupted version, c) CD-ROM filter ,PSNR=31.8688 , d) ACWM filter ,PSNR=36.7103 ,e) proposed method, PSNR=38.1625 , Th_2=10

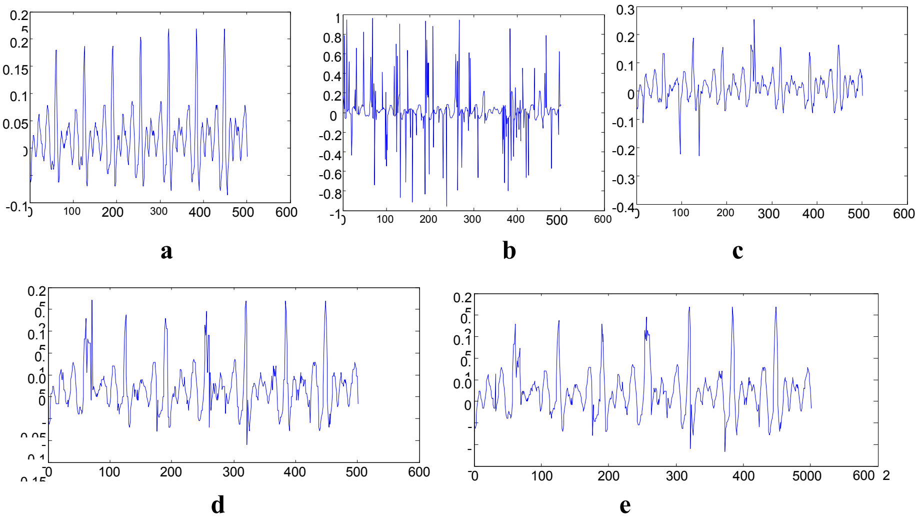

Fig.4 Restoration performance for 500 samples corrupted with 20% noise rate from the Flagraising.wav sound file. A) original sound, b) corrupted version, c) CD-ROM filter ,PSNR=27.4761, d) ACWM filter ,PSNR=30.1315 ,e) proposed method, PSNR=31. 577, Th_2=10

Also, it is obvious From Fig. 3 that the proposed method still delivers the best performance in restoring rainfall audio signal corrupted with higher noise rate of 10% impulse noise. Figure 4 is another example used to illustrate the restoration results for 500 samples taken from the Flag raising sounds corrupted with noise rate of 20% impulse noise. It is notable that the proposed filter still achieves the best results even at high noise rates. Figure 5 and 6 demonstrate the results in restoring two audio files corrupted with 30% and 40% impulse noise rates. An emergency and aircraft audio sounds are shown in fig 5 and fig.6 respectively. Even at high noise rates the proposed filter still provides the best performance. The reason is that two important method is implemented in the proposed filter. The first one is used for filtering the common noisy samples during the detection process and the second one is implementing a 1-D non-local mean in the restoration process.

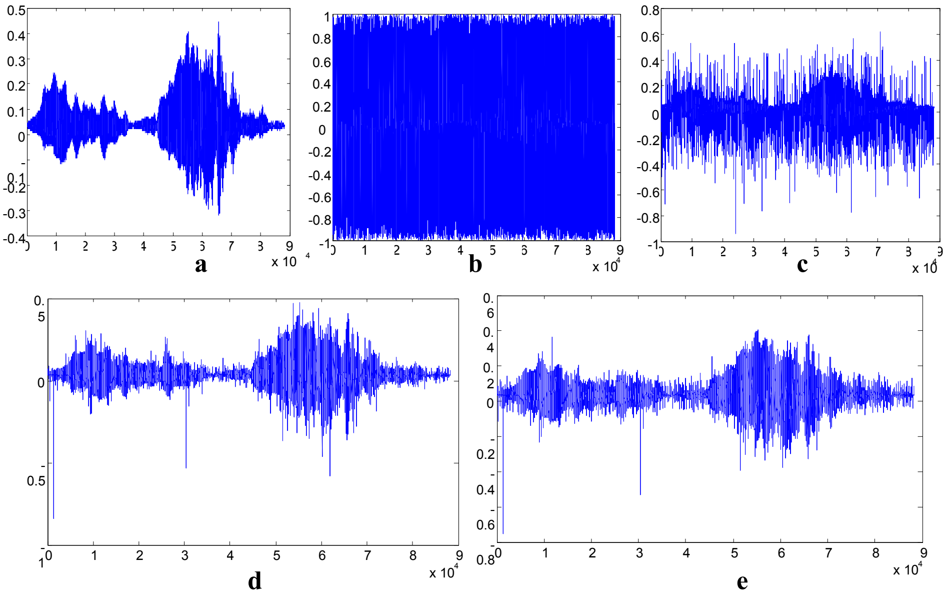

Fig.5 Restoration performance for an emergency sound file corrupted with 30% noise rate. A) original sound, b) corrupted version, c) CD-ROM filter ,PSNR=24.3212, d) ACWM filter ,PSNR=27.4941 ,e) proposed method, PSNR=28.3229,Th_2=10

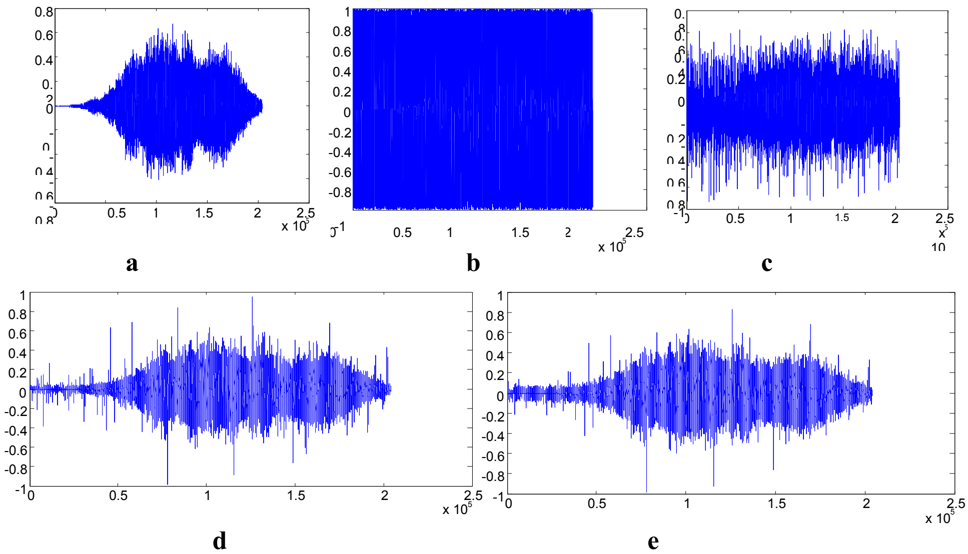

Fig.6 Restoration performance for an aircraft_landing.wav sound file corrupted with 40% noise rate. A) original sound, b) corrupted version, c) CD-ROM filter ,PSNR=20.1978 , d) ACWM filter ,PSNR=22.4439 ,e) proposed method, PSNR=22.7370 , Th_2=20

1-D non local mean restoration process helps delivering good results even at high noise rates because the mean value, in this case, is taken for a large number of similar samples. As the number of similar samples increases the mean becomes closer to the original value. Besides that, the effect of the noise is clearly disseminated. The number of similar samples are determined based on the parameter Th_2, which almost less than 10 for noise rate ≤ 30 and increases when the noise rate increases. Furthermore, the proposed method is fast since it consumes time between 11-30 seconds particularly for noise rates ≤ 30% but if the noise rate increases the consumed time will increase as well. CPU of 1.73 GHz, and RAM of 1 GB are used during the simulated experiments by using the MATLAB program. Table I illustrates the output results of the adopted method against the suggested parameters. Assuming that all the other parameters remain unchanged and we would like to find the values of the threshold Th_2

and the parameter *k* that would give optimum results. Th_2 represents the maximum intensity difference between the mean of the current patch and the means of the similar patches in the searching area *A*. But, the parameter *k* represents the length of the surrounding samples that may have at least one sample similar to the tested common noisy sample. It is clear from table 1 that the proposed method is almost solid against the suggested parameters.

TABLE 1: The effect of the parameters *k* and *Th2* in restoring different audio signals corrupted with impulse noise at 5% noise rate

Length of the surrounding samples k	**Th_2 to determine the number of similar means**							
	PSNR for Rain fall sound				**PSNR for Flag raising sound**			
	5	10	15	20	5	10	15	20
2	39.3	39.4	39.5	39.6	39	38.8	38.6	38.3
4	39.9	40	40	40.1	38.3	38.1	37.9	37.7
6	40.8	40.9	40.9	40.9	36.8	36.7	36.6	36.4
8	40.7	40.7	40.8	40.8	36.6	36.5	36.3	36.1

4 Conclusion

Two detection methods are proposed in this paper to detect and filter the noisy audio samples. The first one is implemented to detect the common noisy samples by differentiating between the outputs of two parallel filters. The second method is used to extract the truly noisy samples from the common ones. Eventually, one-dimensional non-local mean method is deployed for restoring only the noisy samples taken from the second detection technique rather than restoring all the samples of the corrupted sound. Simulation results proved that the proposed method is able to restore audio signals corrupted at different noisy rates, and to deliver better results than other known methods. Furthermore, the proposed method is fast and consistent against the proposed parameters.

References

1- A. Nieminen,"Suppression and detection of impulse type interference using adaptive median hybrid filters," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1987.

2- M.B.Jones, "Real-time speech enhancement using median filters," *in Proceedings of SST* 1994 , vol.1, pp.406–410,1994.

3- T. Kasparis and J. Lane , "Adaptive scratch noise filtering, " *IEEE Trans. Consumer Electronics*, Vol.39,No.4, pp.917-922,1993

4- S. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction, " *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. 27,no2, pp.113-120, Apr.1979

5- P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and projection, for robust recognition in cars," *Speech Commun*., vol. 11, pp. 215-28, 1992

6- Y. Hu, P. C. Loizou, "Speech enhancement based on wavelet thresholding the multiaper spectrum," *IEEE Transaction on Speech and Audio Processing*, vol. 12,no.1, pp. 59-67, January 2004.

7- S. Kadambe and G. Boudreaux-Bartels, "Application of the wavelet transform for pitch detection of speech signals," *IEEE Transactions on Information Theory*, vol. 38, pp. 917–924, 1992.

8- G. Yu, E. Bacry and S. Mallat, "Audio Signal Denoising with Complex Wavelets and Adaptive Block Attenuation", *IEEE International Conference on Acoustics, Speech and Signal Processing,* Vol. 3, 2007

9- M . Bahour, J. Rout, "Wavelet speech enhancement based on the Teager energy operator," *IEEE Signal Processing Letters*, vol. 8,no1, pp. 10-12, 2001

10- M. Nabih , EL-Sayed A. El-Dahshan, Ashraf H. Yahia, "Denoising of Heart Sound Signals Using Discrete Wavelet Transform, " *Circuits, Systems, and Signal Processing*, vol. 36, no.11, pp. 4482–4497,2017

11- B. Messaoud, M. A., and A. Bouzid, "Sparse Representations for Single Channel Speech Enhancement Based on Voiced/Unvoiced Classification," *Circuits, Systems, and Signal Processing*, Vol.36,No.5, PP.1912-1933, March 2017

12- A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval, M. D. Plumbley, "Audio Inpainting," *IEEE Transactions on Audio, Speech and Language Processing,* vol.20,no, 3, pp.922 – 932,. 2012

13- G. Bhattacharya and P. Depalle, "Sparse Denoising of Audio by Greedy Time-Frequency Shrinkage," *IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), 2014

14- Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoustic, Speech, Signal Processing, vol. ASSP*-32, pp. 1109- 1121,Dec. 1984

15- J. E. Porter and S. F. Boll, "Optimal estimators for spectral restoration of noisy speech," in Proc. IEEE , *Acoustic, Speech, Signal Processing*, , pp. 18A.2.1-l8A.2.4, Mar. 1984

16- P.A., Esquef, "Audio restoration", In: Havelock D., Kuwano S., Vorländer M. (eds) *Handbook of Signal Processing in Acoustics. Springer*, New York, NY , pp 773-784,2008

17- M. Niedźwiecki, M. Ciołek, " Localization of impulsive disturbances in audio signals using template matching," *Digital Signal Processing*, Elsevier, vol. 46, pp. 253-265 , Nov. 2015

18- C. Chandra, M. S. Moore and S. K. Mitra, "An efficient method for the removal of impulse noise from speech and audio signals," *IEEE Proc. on Circuits and Systems*, vol. 4, no. 8, pp. 206-208, June 1998

19- T. Chen and H. R. Wu. "Adaptive impulse detection using center-weighted median filters," *IEEE Signal Processing Letters*, vol.8, no.1, pp.1 -3, Jan. 2001

20- E. Abreu, M. Lightstone, S. Mitra, and K. Arakawa, "A new efficient approach for the removal of impulse noise from highly corrupted images," *IEEE Trans. on Image Processing*, vol. 5, no. 6, pp. 1012-1025, 1996

21- A. Buades , B. Coll , JM. Morel, "A non-local algorithm for image denoising," *IEEE international conference on computer vision and pattern recognition*, June 2005

22- T. Brox, O. Kleinschmidt, D. Cremers, "Efficient nonlocal means for denoising of textural patterns, " *IEEE Trans Image Process* , vol.17, no.7 pp. 1083–1092, *2008*

23- A. Buades , B. Coll , JM. Morel, "Nonlocal image and movie denoising". *Int J Comput Vision*, 2008,vol. 76,no.2, pp 123–139,2007