IJCTA, 9(8), 2016, pp. 3497-3504

Handling Impulsive Disturbances in Speech Recordings using Filtering Techniques

S. Selva Nidhyananthan*, M. Pandi Muneeswari and A.J. Ashmi Grace*****

ABSTRACT

This paper presents enhancement techniques of the noisy speech signal by eliminating impulse noise from noisy speech. Median filter, SDROM (Signal Dependent Rank Order Mean Filter), Adaptive filter, Wavelet filter, are the filters used to remove impulse noise from audio speech. The performance of the filtering techniques is measured with both subjective and objective measures. Subjective measure involves MOS (Mean Opinion Score). Objective measures involve Signal to Noise Ratio (SNR) and Segmental SNR (SNRseg). Experimental result reveals the supremacy of Rank Order Mean filter over other filters with an SNR of 42dB.

Index Terms: Impulse noise, Filtering techniques(Median, SDROM, Adaptive, wavelet).

1. INTRODUCTION

Most of the speech files from recordings are suffered from impulsive disturbances. Clicks, pops are the example of impulse noise. They are also generated by electromagnetic interferences, scratches on a recording disk. Mostly they have very short duration but they have high peak, they are called as On/Off noise also. Impulse noise is also harmful for hearing impairment. Speech signal is a measurable physical quantity.

Generally, it is a aperiodic signal but in practical, it is stationary for 20ms.

The most general approach to remove impulse noise is filtering technique. In this project, real time impulse noise and also mathematically simulated impulse noise for experimentation.

One of the classier filters is median filter[1]-[5]. It is a well known filter for impulse noise in Image Processing as well as in Signal Processing. It estimates the noise samples and replace with median value. It works fine but in case if we use large window size then most of the uncorrupted samples also replace with median value, so it becomes inefficient one, when the window size is large. This drawback is eliminated using Rank Order Mean Filtering technique [6].

In Wavelet filtering technique [7], the speech signal is divided into two sub-bands using wavelet transform having high frequency and low frequency bands. Most of the noise is present in the high frequency component. So in reconstruction, selected high frequency sub-bands are not considered, thus impulse noise is removed. The next approach is Adaptive filter [8]-[12]; here the noisy speech is passed to the unknown filter and also the Adaptive filter. Adaptive filter change its coefficients based on the error computed by LMS algorithm. This paper is organized as follows. Section II provides database description. Section III provides proposed method. Section IV includes preprocessing. Filtering techniques are described in section V. Section VI includes performance analysis and section VII provides conclusion.

* AP-Senior Grade Department of ECE Mepco Schlenk Engineering college, Sivakasi, India, *Email: nidhyan@mepcoeng.ac.in*

** UG Scholar Department of ECE Mepco Schlenk Engineering college, Sivakasi, India, *Email: pandimuneeswarim@gmail.com*

*** UG Scholar Department of ECE Mepco Schlenk Engineering college, Sivakasi, India, *Email: ashmigrace22@gmail.com*

2. DATABASE DESCRIPTION

One hundred samples of input speech from 50 speakers and 100 samples of impulse noise were recorded. They are all recorded in Gold Wave Platform with mono, 16 bit, PCM coding. They are all recorded with sampling frequency of 16000Hz in a noise free glass room environment.

3. SPEECH ENCHANCEMENT METHOD

Speech mixed with impulse noise gives noisy speech. Let the input speech is $x(n)$, the impulse noise is $\delta(n)$, then the noisy speech is given by,

$$X'(n) = x(n) + \delta(n) \tag{1}$$

Then the noisy speech is given to the filter section(Median/SDROM/Adaptive/Wavelet)The filtered section is evaluated by performance measures. It involves subjective and objective performance measures. The proposed speech enhancement method is given in Fig. 1.

4. PREPROCESSING

The preprocessing stage in speech is used in order to increase the efficiency of speech quality, and also to prepare for future processing. Commonly preprocessing includes Pre-emphasis and Signal Segmentation.

Pre-emphasis: In speech processing, the original signal usually has lower frequency energy, so we have to process the signal to emphasize higher frequency. Pre-emphasis is given by,

$$y(n) = X'(n) - \alpha X'(n-1) \tag{2}$$

Where $y(n)$ is pre-emphasized output, $X'(n)$ is the noisy input speech, α is the pre-emphasis parameter whose value is 0.97.

Signal Segmentation: Mostly speech is non-stationary one, but it is stationary for some limited duration, so that the signal is framed to n-number of frames. The speech signal is then blocked into frames of N number of samples, with adjacent frames. The second frame begins with N samples after first frame and

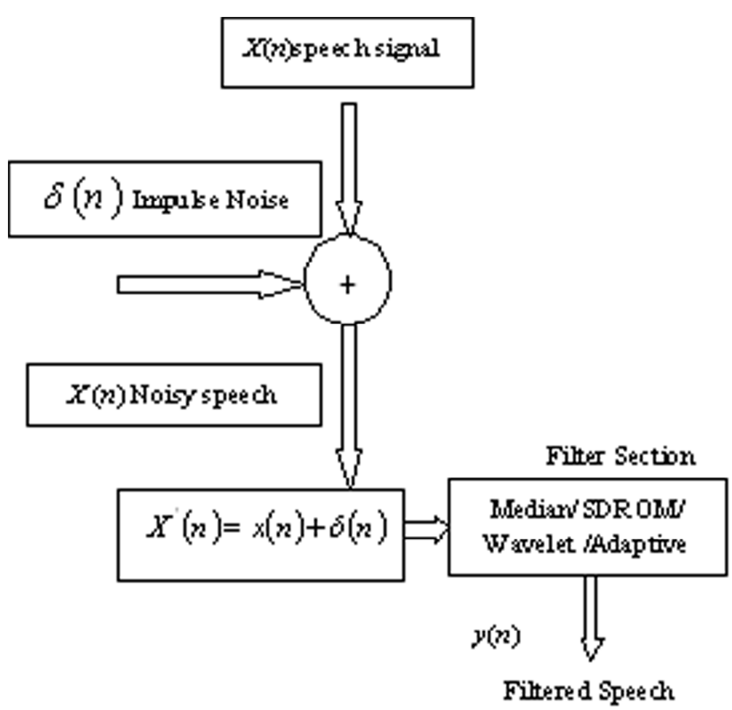

Figure 1: Proposed Speech Enhancement Method

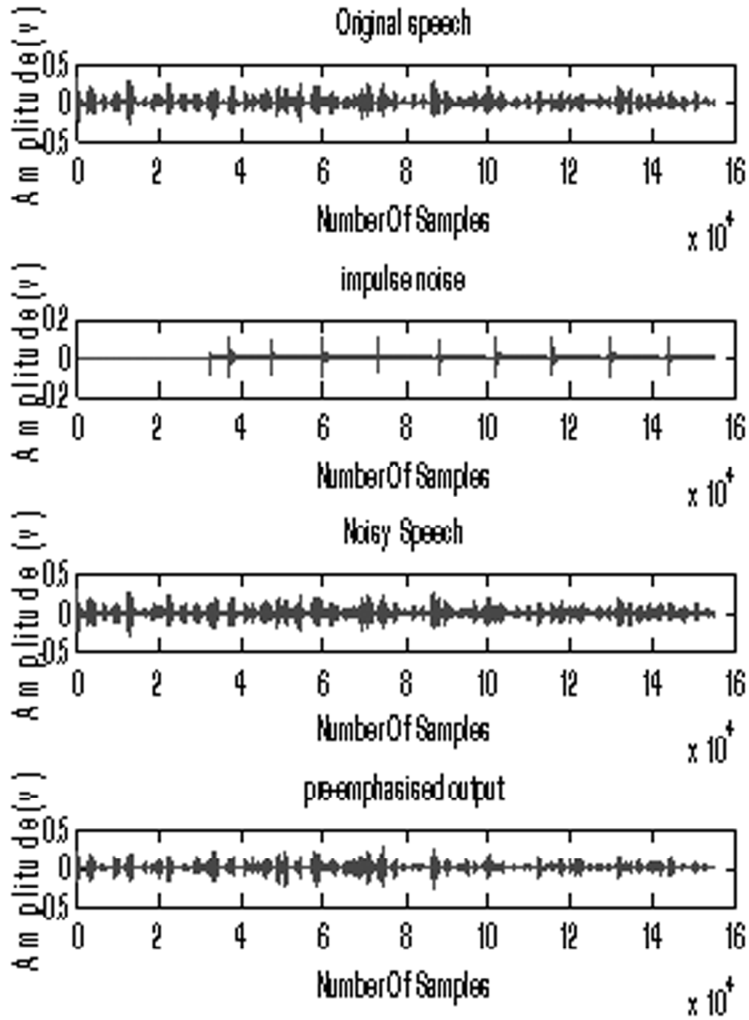

Figure 2: Pre-emphasized signal

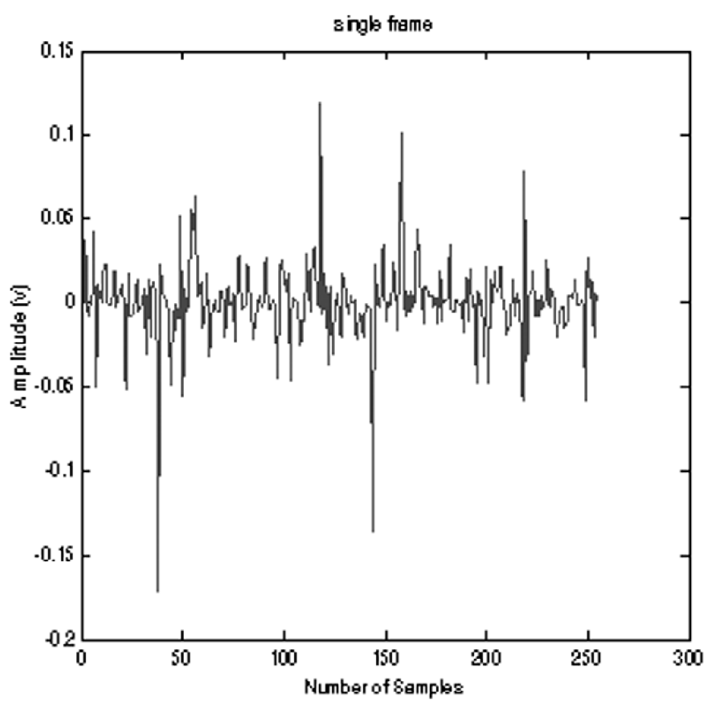

Figure 3: One Frame of Noisy Speech

overlaps by previous samples i.e. 50% of the samples are overlapped. In this paper we framed the input speech signal having 155384 samples into 607 frames; each frame has 256 samples, over that 50% of the sample will be overlapped for signal smoothing. The pre-emphasized speech signal & the single frame of the noisy speech is given in Fig. 2 and Fig. 3 respectively.

5. FILTERING

5.1. Median Filter

Median filters are particularly effective in the presence of impulse noise. Consider a sliding window size of five centered at $x(n)$. Sort the values and find the median position value. Replace $x(n)$ with computed median value & so on. We experimentally have done the median filter with various speeches with impulse noise in MATLAB. The Result is shown in Fig. 4.

5.2. SDROM (Signal Dependent Rank Order Mean Filter)

SDROM is one of the statistic filters to remove the noise impulses from audio & speech signals. It is a Detection-estimation approach. Here we used 1-D sliding window of size five centered at $x(n)$. $W(n)$ be the vector which excludes centered.

$$w(n) = [x(n-2),\ x(n-1),\ x(n+1),\ x(n+2)] \tag{3}$$

Then, the $W(n)$ are sorted and stored in $r(n)$,

$$r(n) = [r1(n),\ r2(n),\ r3(n),\ r4(n)] \tag{4}$$

where the elements in $r(n)$ are sorted by rank i.e. $r1(n) \le r2(n) \le r3(n) \le r4(n)$. Then compute rank order mean given by

$$\mu(n) = [r2(n) + r3(n)]/2 \tag{5}$$

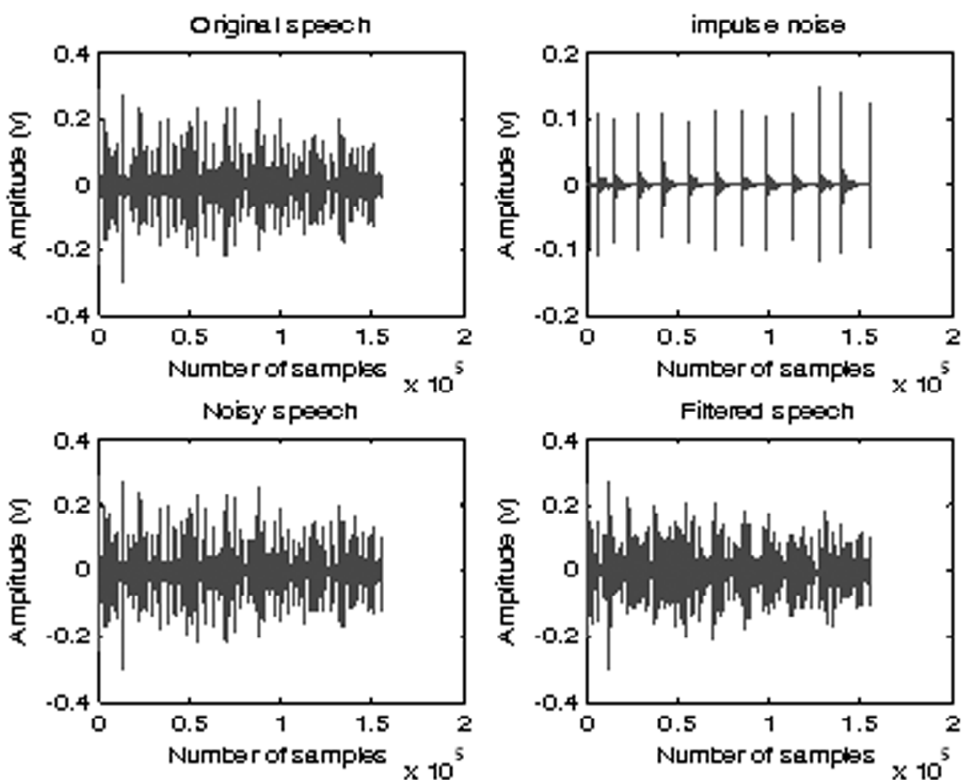

Figure 4: Result of Median filter for real time impulse noise

The rank order difference is given by

For i = 1, 2

$$\begin{cases} d_i(n) = \{r_i(n) - x(n)\} \\ d_i(n) = \{x(n) - r_{4-i}(n)\} \end{cases} \tag{6}$$

The algorithm computes $x(n)$ is noise impulse, if $d_i(n) > T_i$ for i = 1, 2 where T_i are threshold value. [0.15, 0.15] are the threshold used for 16-bit PCM. Every detected impulse noise is replaced by rank order mean $\mu(n)$.

The result is shown in Fig 5.

5.3. Wavelet Filter

Discrete Wavelet Transform decomposes the speech signal into mutually orthogonal set of wavelets. Here, Wavelet transform is done by choosing a wavelet at a decomposition level 2.

For first level, noisy speech(s) is decomposed into two sub-bands (approximation coefficients, detail coefficients) by choosing *Db4* wavelet. For the second level, the approximation coefficients are again split into two sub-bands followed by decimation by 2. In reconstruction high frequency component in the first stage of decomposition is eliminated after interpolation by 2. The experimental result of wavelet filter is shown in Fig 6.

5.4. Adaptive Filter

Adaptive filter change the filter coefficients to adapt to changing signal characteristic. It restores the desired speech signal by passing the input noisy speech through a FIR filter whose coefficients are calculated by minimizing the MSE between the desired and estimated signals.

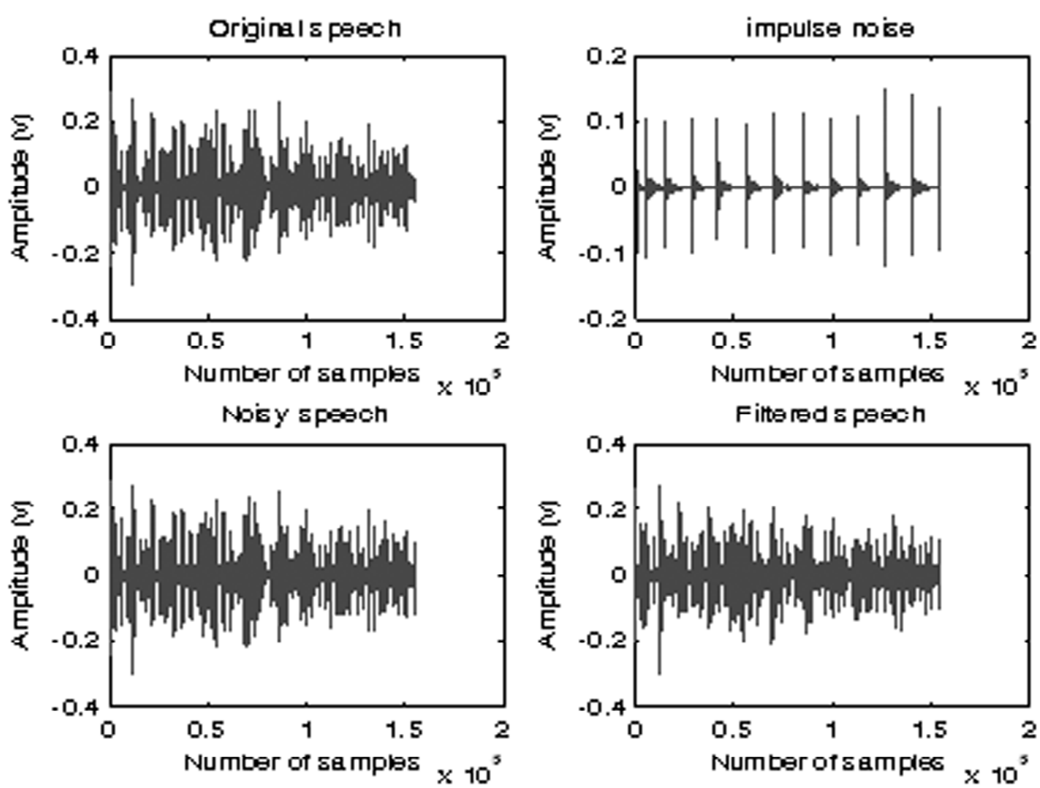

Figure 5: Result of SDROM for real time impulse noise.

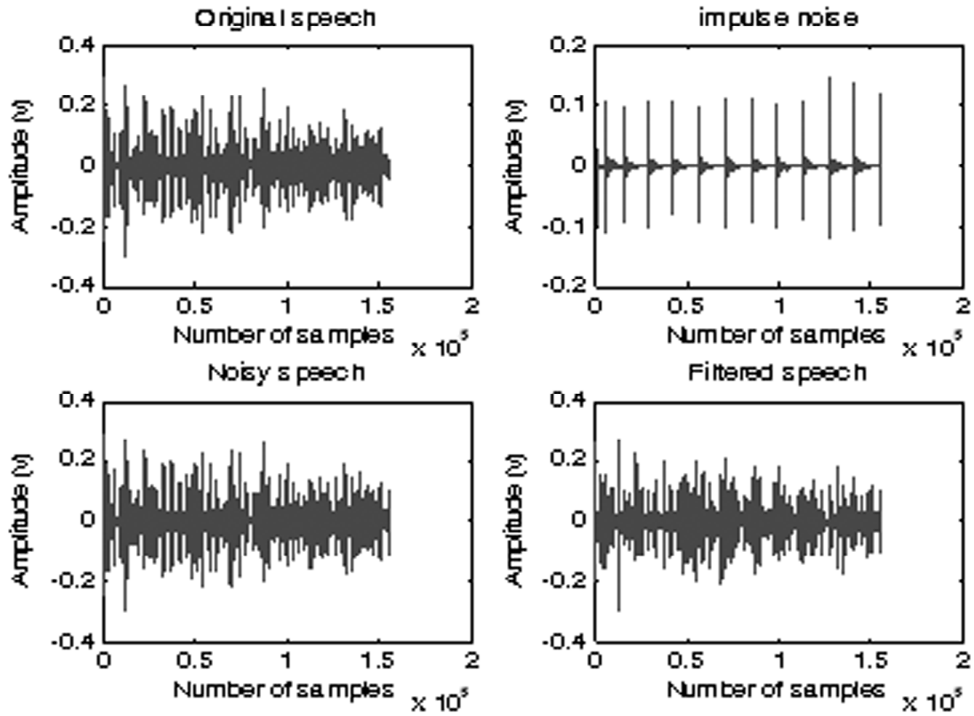

Figure 6: Result of Wavelet for real time impulse noise

The noisy speech $X'(n)$ is given by

$$X'(n) = x(n) + v(n) \tag{7}$$

Let $w(n)$ be the weight vector is given by

$$w(n) = [w_1(n), w_2(n), \; .... \; w_n(n)]^T \tag{8}$$

The filter output is given by

$$y(n) = w^T(n) X(n) \tag{9}$$

Adaptive filter adjust $w(n)$ at each instant according to the error $e(n)$ is

$$e(n) = d(n) - y(n) \tag{10}$$

The experimental result of adaptive filter is shown in Fig 7.

6. PERFORMANCE ANALYSIS

The performance of the filtering techniques is measured with both subjective and objective measures. Subjective measure involves MOS (Mean Opinion Score). Objective measures involve Signal to Noise Ratio (SNR) and Segmental SNR (SNRseg).

Signal to Noise Ratio (SNR):

$$SNR = 10\log_{10}(P_{SIGNAL} / P_{NOISE}) \tag{11}$$

$$SNR = 10\log 10 \frac{\sum_n s^2(n)}{\sum_n [s(n) - s^\wedge(n)]^2} \tag{12}$$

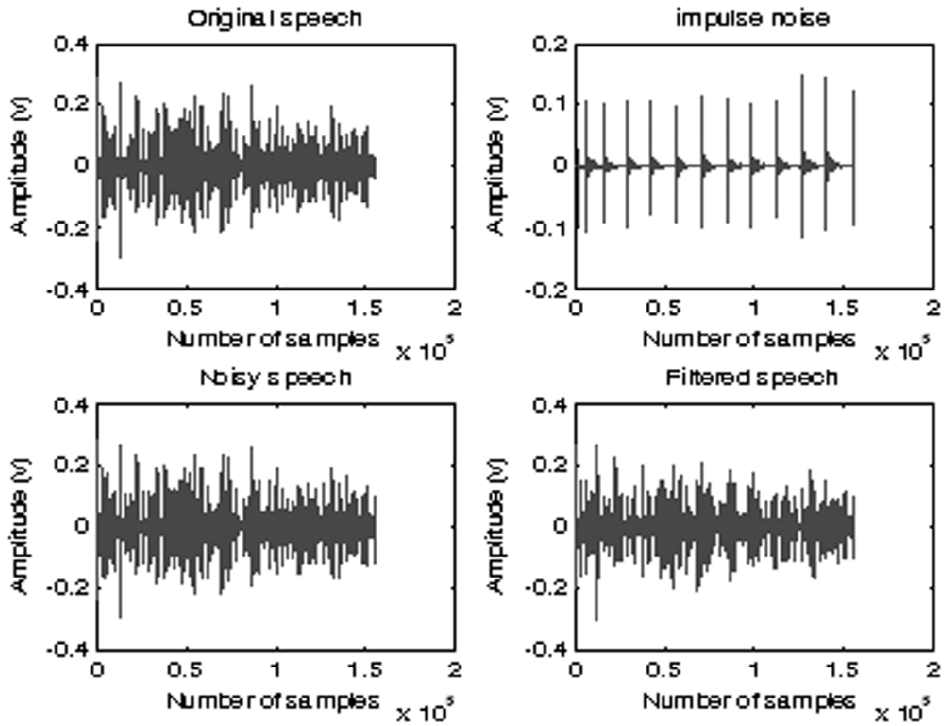

Figure 7: Result of Adaptive filter for real time impulse noise

Where $s\wedge(n)$ is the filtered speech,

$[s(n)-s\wedge(n)]$ is the residual noise after filtering.

Segmental SNR (SNRseg):

It can be obtained by measuring SNR over short frames & the result is averaged.

$$SNR = \frac{1}{M} 10 \log_{10} \frac{\sum_n s^2(n)}{\sum_n [s(n) - s\wedge(n)]^2} \tag{13}$$

Mean Opinion Score:

Subjective measures are based on the opinion of the listener. Mean Opinion Score (MOS)-is an opinion rating method used to access the degree of the quality of speech. MOS method rates the speech under the test on a five point scale where a listener's subjective impression is assigned a numerical value. The rating table is given below.

The SNR measure is experimentally found with simulated noise and real time noise.

Table 1
MOS Rating table

Rating	*Speech Quality*	*Level of distortion*
5	Excellent	Imperceptible
4	Good	Not Annoying
3	Fair	Slightly Annoying
2	Poor	Not objectionable
1	Unsatisfactory	Objectionable

Table 2
Performance analysis-objective

Filter	*SNR in dB (stimulated noise)*	*SNR in dB (real time impulse noise)*
Median	33.35	37.38
SDROM	40.47	45.46
Adaptive	34.2	34.73
Wavelet	38.67	42

Table 3
Performance analysis-subjective

Filter	*MOS Rating scale*
Median	2
SDROM	5
Adaptive	3
Wavelet	4

7. CONCLUSION

By various filtering techniques the impulse noise is eliminated from the noisy speech. Among these filters Rank Order Mean is the best filter to remove impulse noise & also it improves the quality of speech. The application is to preserve our heritage by eliminating noise in old recordings and also it prevent from hearing impairment. It is also used in Telecommunication to remove unwanted sharp sounds.

REFERENCES

[1] S. Akkoul, R. L‘ed‘ee, R. Leconge, and R. Harba, "A new adaptive switching median filter," IEEE Signal Processing, 2010 vol. 17, no. 6, pp. 587-590.

[2] Y. Liu, C. Liu, and D. Wang, "A 1D time-varying median filter for seismic random, spike-like noise elimination," Geophysics, 2009, vol. 74, no. 1, pp. 17-24.

[3] X. Kang, M. Stamm, A. Peng, and K. R. Liu, "Robust median filtering forensics using an autoregressive model," IEEE Transactions on Information Forensics and Security, 2013, vol. 8, no. 9, pp. 1456-1468.

[4] Y. Zhang, S. Li, S. Wang, and Y. Q. Shi, "Revealing the traces of median filtering using high-order local ternary patterns," IEEE Signal Processing Letters, 2014.vol. 21, no. 3, pp. 275-280

[5] N. Gallagher and G. Wise, "A theoretical analysis of the properties of median filters," IEEE Transactions on Acoustics, Speech and Signal Processing,, 1981,. vol. 29, no. 6.

[6] Charu Chandra, Michael S. Moore and Sanjit K.Mitra, "An Efficient Method for the removal of impulse noise from speech and audiosignals," IEEE Signal processing, 1998, vol. 4.

[7] Brani Vidakovic and Concha Bielza Lozoya, "On Time-Dependent Wavelet Denoising**",** IEEE Transaction on signal processing, September 1998 Vol. 9.

[8] B. Pšenicka and J. Zadak, "Digital filters matricial analysis," Acta Polytechnica, 1991, vol. III., no. 1, pp. 5-15.

[9] S. Haykin, Adaptive Filter Theory, 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1996.

[10] Lorenzo Pasquato and Izzet Kale, Adaptive IIR Filter Initialization via Hybrid FIR/IIR Adaptive Filter Combination, IEEE Transaction on Instrumentation and Measurement, 2001, Vol. 50, No. 6.

[11] I. Kale, J. Gryka, G. D. Cain, and B. Beliczynski, "FIR filter order reduction: Balanced model truncation and Hankel-Norm optimal approximation," IEEE Proc. Vis. Image Signal Processing, June 1994., vol. 141.

[12] H. Fan and M. Nayeri, "On reduced order identification: Revisiting 'On some system identification techniques for adaptive filtering'," IEEE Trans. Circuits Syst., Sept. 1990., vol. 37, pp. 1144-1151.