

# Study on the Selection of Specific Filters for Enhancement of Recorded Speech for Speaker Identification

P. V. Jiju<sup>1</sup>, C. P. Singh<sup>2</sup> and Dr. R. M. Sharma\*<sup>3</sup>

<sup>1</sup>Research Associate, Central Forensic Science Laboratory, Directorate of Forensic Science, Ministry of Home Affairs, Govt. of India, Plot No-2, Sector-36A, Chandigarh, India; <sup>2</sup>Assistant Director, Physics Division, Forensic Science Laboratory, Govt. of NCT of Delhi, India; <sup>3</sup>Department of Forensic Science, Punjabi University, Patiala, India

**Abstract:** Understanding the noise characteristics for finding appropriate filtering technique/s so as to obtain sufficiently clear speech samples for Speaker Identification, is one of the challenging tasks in Forensic Acoustics. Speaker's idiosyncratic speech should not be affected when the noise reduction is carried out; otherwise, Speaker Identification becomes highly erroneous. We have collected fifty noisy speech samples reported to be recorded in different modes from actual crime cases received in the laboratory. The samples are analyzed after subjecting to various filtering techniques and compared with the clear speech mixed with the noise collected from non-speech portion. Distortion levels on the speech are studied at various stages of application of filters in terms of SNR and Speaker Specific Information. Retaining the Speaker Specific Information as primary concern of our study, the limitation of filtering techniques depending on the characteristic and intensity level of noise is worked out for noisy speech samples. Subsequently a statistical study is also conducted. Listening tests were conducted to ensure that the perceptual features of the original noisy speech are preserved while applying filters. This work demonstrates the efficiency of Noise reduction filters in improving SNR and their controlled applications for preserving Speaker dependent features depending on the various noise characteristics embedded on speech samples.

**Keywords:** Distortion level, Speaker specific information, Intelligibility.

## 1. INTRODUCTION

Audio Forensics has a challenging history of enhancement problems of speech samples received for examination. It is observed that out of the total speech samples received for Speaker Identification in the Laboratory, a large number of recordings requires enhancement. Speech is a non-linear time series represented in terms of complex number. Hence separating noise from noisy speech in spectral domain results into countless solutions.

The main objective of a Noise Cancellation system is to obtain a clear signal with higher quality of speech signal. The presence of noise in speech signals can create higher degree of mismatch in performance of speech processing systems used for Speaker Identification as well as Speech Recognition. Inappropriate filtering of noise corresponds to extracting features of noise together with the actual speech signal during the feature extraction process. However, the desired parametric representation carries a high amount of error rate. The presence of broadband noise and a very low SNR deteriorate the intelligibility of most of the recorded speech samples.

Speaker's idiosyncratic speech is affected when the noise reduction is carried out. Thus the Speaker Identification

becomes highly erroneous. To reduce or eliminate different form of noises from a recorded speech, forensic audio expert use variety of filters. The study refers to examination of noise characteristics for finding appropriate filtering technique(s) to improve SNR and to preserve speaker dependent features so as to obtain sufficiently clear speech samples for single channel recordings. The paper insight into the efficiency of Noise reduction filters in improving SNR by preserving speaker dependent parameter.

## 2. METHODS

Fifty noisy speech exemplars of 20 seconds duration are collected from the exhibits of actual crime cases received in the Laboratory. The speech exemplars recorded directly on tape, through telephone channel and mobile phone channel were found to be embedded with different kinds of noise. Seventeen direct (hereafter referred as Noisy Speech-I), seventeen telephonic (hereafter referred as Noisy Speech-II) and sixteen mobile (hereafter referred as Noisy Speech-III) recordings have been studied.

According to the "Golden rule" of enhancement, no audio signals are removed or attenuated that decreases the speech intelligibility, even slightly [1]. The noisy observation  $y(m)$  can be modeled as

$$y(m) = b(m)*x(m) + n(m) \text{ -----(1)}$$

where,  $x(m)$  is the signal to be detected,  $n(m)$  is the noise and  $b(m)$  is a binary - valued state sequence, such that  $b(m) = 1$  indicates the presence of the signal  $x(m)$  and  $b(m) = 0$  indicates that the signal is absent. If the signal  $x(m)$  has a known shape, then a correlator or a matched filter can be

\*Address correspondence to this author at the Department of Forensic Science, Punjabi University, Patiala, India; Tel: 91-175-3046271; Fax: 91-175-2283073; E-mail: rmsforensics@yahoo.co.in

used to detect the signal. In speech enhancement, quality of the speech signal,  $x(m)$  is increased by reducing or suppressing the noisy part,  $n(m)$ .

“Fig. (1)” shows the configuration of a matched filter followed by a threshold comparator for detecting signals in noise. Suitable filters are identified for filtering the noise based on the information obtained from their waveform statistics and frequency analysis.

The samples are digitized with a sampling rate of 22050Hz with 16-bit quantization. They are classified into Noisy Speech-I, Noisy Speech-II and Noisy Speech-III. The classified exemplars are subjected to undergo application of various filtering techniques. The filters applied are Band-pass, Noise Gate, Noise Reduction, Digital Equalizer and Parametric Equalizer.

Simulated speech samples are created by collecting non-speech regions containing noise from the sample and mixed with reference speech signal recorded in ideal condition. A comparative study of systematic noise filtering is carried out by comparing the distortion level and SNR of the filtered speech signal with reference speech signal, a standard method adopted else where [2, 3]. The SNR of simulated speech samples before and after applying the filters are also compared and studied. From the resultant signal, speech is reclaimed to a maximum extent by nullifying the effect of embedded noise with the help of specific filter(s). This in turn helped to decide the degree of quality of speech signal retrieved for Speaker Identification under noisy conditions in terms of SNR and perceptual features.

Thus, various groups of noise embedded with the speech samples are subjected for suitable filtering technique(s) for efficient noise reduction. Then a study has been conducted on the effect of filter application on SNR and speaker dependent features of samples.

As perceptual features are very much important in the case of Forensic Speaker Identification, listening tests are conducted to ensure that the perceptual features of the original noisy speech are preserved while applying filters. For this purpose, 10 listeners of age group of 25-30 were trained with the original noisy speech until they are familiar to follow the perceptual features of each speaker. The trained listeners are then subjected to filtered speech samples. Finally, an opinion is made based on the method of Mean Opinion Score (MOS); the most widely accepted method for speech quality evaluation and a simplest subjective measure for the assessment that gives an overall opinion of the performance. The standard and possible set of score for MOS is presented in the Table 1.

**Table 1. Mean Opinion Score Definition**

Score	Opinion	Impairment Scale
1	Unsatisfactory	Annoying and unacceptable
2	Poor	Annoying
3	Fair	Slightly annoying
4	Good	Perceptible but acceptable
5	Excellent	Imperceptible

In the experiment, Blackman windowing technique is used with a frame length of 20 milliseconds. SNR of both the original as well as simulated samples are measured before and after the filter application. The energy contour of the input and output signals are studied and compared for measuring the improvement in SNR for each filter application.

$$\text{SNR (dB)} = 10 \log_{10}(P_{\text{signal}}/P_{\text{noise}}) = 20 \log_{10}(A_{\text{signal}}/A_{\text{noise}}) \quad (2)$$

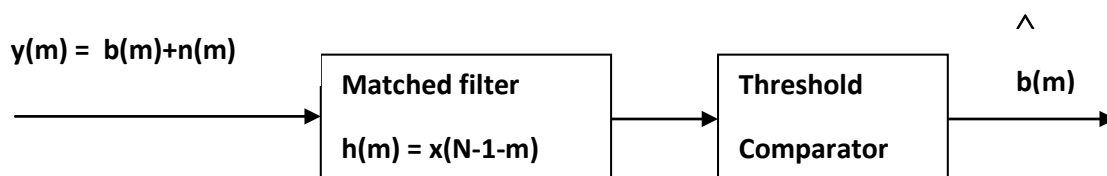
A decision on the processed samples is made by estimating SNR with help of energy spectrum and Mean Opinion Score (MOS) test.

### 3. RESULTS AND DISCUSSIONS

By the controlled application of various filters a balance in both the perceptual as well as measuring qualities of speech signal are achieved. Based on the spectrum analysis it is checked and established that the speaker dependent features useful for Speaker Identification are preserved during the process of noise reduction.

In our experiment, all the fifty original samples and their corresponding simulated noisy samples are subjected for Band-pass, Noise gate, Noise reduction, Digital equalizer and Parametric equalizer. In order to preserve maximum speaker specific information, when Band-pass filter is applied, frequency range has been chosen as per the classification of samples (Noisy Speech-I, Noisy Speech-II and Noisy Speech-III). In the original noisy samples, 47% of noisy samples of Noisy Speech-I, made significant improvement in SNR while Speaker dependent features have been unaltered. In case of Noisy Speech-II and Noisy Speech-III, there were 70% and 18.75% improvement (SNR & Speaker dependent features being preserved) of samples respectively. Out of the total 50 samples 46% of them made significant improvement (given in Table 2).

In the simulated noisy speech samples, for Noisy Speech-I, 47% of noisy samples made significant improvement (SNR, Speaker dependent features). In case of Noisy Speech-II, improved (SNR, Speaker dependent features)



**Fig. (1).** Configuration of a matched filter followed by a threshold comparator for detection of signals in noise.

**Table 2. Effect of Various Filters Upon Various Modes of Recordings of Original Speech Samples**

Filters	Improvement in Percentage			Overall Improvement in Percentage
	Noisy Speech-I	Noisy Speech-II	Noisy Speech-III	
Band-pass	47.00%	70.00%	18.75%	46%
Noise gate	41.20%	59.00%	68.75%	56%
Noise reduction	82.35%	76.50%	68.75%	76%
Digital equalizer	29.40%	17.60%	25.00%	24%
Parametric equalizer	29.40%	29.40%	37.50%	32%

version of 53% of Noisy Speech-II was obtained. The improvement (SNR, Speaker dependent features) in Noisy Speech-III speech samples was 12.50% when subjected to Band-pass filtering (given in Table 3). Out of the total 50 samples 38% of them made significant improvement.

As this filtering technique is flexible it does not affect the desired frequency band and hence it is of great value for Speaker Identification purposes. The promiscuous speech exemplars in which the noise falls in the broadband, this technique is found to be unsuitable. The experiment with exclusively noisy exemplars of corresponding speech exemplars stands as the ground for this.

Application of Noise Gate is effective only in removing the background hiss from quiet parts (i.e. any place where the speaker paused or stopped talking). Results from our experiment have also emphasized this fact. 41.20% of the Noisy Speech-I, 59.00% of the Noisy Speech-II and 68.75% of the Noisy Speech-III noisy speech samples became useful for the analysis by the application of this technique and there was an overall improvement (SNR) of 56% among the original noisy samples (given in Table 2). In the case of simulated noisy speech samples, 29.40% of the Noisy Speech-I, 29.40% of the Noisy Speech-II and 37.50% of the Noisy Speech-III noisy speech samples became useful for the analysis and there was an overall improvement of 32% (given in Table 3). Noise Gate is not efficiently applicable if the speech region is not embedded by the characteristics of the noise in the gap and it does not remove background hiss

from louder parts of the selection -yet another disadvantage of this filtering technique.

Noise Reduction Filter has been widely accepted for removing the background hiss, power hum or random interference. 82.35% of Noisy Speech-I, 76.50% of Noisy Speech-II and 68.75% of Noisy Speech-III has become useful (improved SNR and with unaltered Speaker dependent features) for the analysis by the application of this technique. A total of 76% of 50 exemplars are turned to be good enough after applying this filtering technique (given in Table-2) whereas for the simulated samples, a total improvement (in SNR and with unaltered Speaker dependent features) of 58% and 70.60%, 53% and 50% for Noisy Speech-I, Noisy Speech-II and Noisy Speech-III respectively were achieved (given in Table 3). This result has been verified with the corresponding noise only samples by spectrum analysis.

Digital Equalizer is applied to all the 50 exemplars. A combination of different frequencies has been found out for each exemplars that can reduce the effect of noise and thus, to increase the intelligibility of the interested portion. A final decision on the equalizer settings is made by perceptual means and subsequently by spectrographic study. There was 29.40%, 17.60% and 25.00% improvement (SNR) in the Noisy Speech-I, Noisy Speech-II and Noisy Speech-III respectively. The technique of Digital Equalizer, have made improvement (SNR) of 24% of the total samples (given in Table 2). When it applied to the simulated noisy speech

**Table 3. Effect of Various Filters Upon Various Modes of Recordings of Simulated Speech Samples**

Filters	Improvement in Percentage			Overall Improvement in Percentage
	Noisy Speech-I	Noisy Speech-II	Noisy Speech-III	
Band-pass	47%	53%	12.50%	38%
Noise gate	29.40%	29.40%	37.50%	32%
Noise reduction	70.60%	53%	50%	58%
Digital equalizer	35.30%	23.50%	6.25%	22%
Parametric equalizer	23.50%	41.20%	12.50%	26%

samples, a total improvement (SNR) of 22% was observed and there was an improvement of 35.30%, 23.50 and 6.25% in Noisy Speech-I, Noisy Speech-II and Noisy Speech-III (given in Table 3). Application of Digital Equalizer affects the relative amplitude of the speech signal, which in turn degrades the speaker specific information.

The Parametric Equalizer is a flexible tool for reducing or enhancing ranges of frequencies. The study of 50 noisy speech exemplars suggests that more the number of bands used more promising the results are. This is because of the fact that frequency resolution increases when number of bands increases. Noisy Speech-I, Noisy Speech-II and Noisy Speech-III have made individual improvement (SNR) of 29.40%, 29.40% and 37.50% respectively. There was an overall improvement of 32% of the total 50 samples when Parametric equalizer is applied to the original noisy speech samples (given in Table 2). When it is applied to the simulated noisy speech samples, there was an improvement (SNR) of 23.50%, 41.20% and 12.50% in Noisy Speech-I, Noisy Speech-II and Noisy Speech-III respectively. A total of 26% was qualified for Speaker Identification by the use of this technique (given in Table 3).

Corresponding filters have been identified based on the contributed enhancement upon noisy speech signal, which amount to the fruitful Speaker Identification test. Table-2 & Table-3 explains how each filter is efficient in improving the quality of speech signal recorded in different modes (Noisy Speech-I, Noisy Speech-II, Noisy Speech-III).

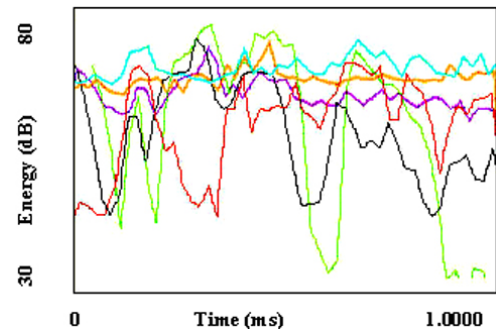
It is insured that the Speaker-Specific Information are preserved during the process of noise reduction by various filters. A comparative study upon the performance of each filter is conducted and specific filters are identified for speech enhancement for the noisy speech samples.

The subjective and objective qualities of the speech signal are affected when filters are applied. For this reason, controlled applications of filters are carried out to maintain the subjective quality of the speech to preserve the speaker dependent features for Speaker Identification. Application of different filters gave improvement in SNR value of the noisy speech signal at different rates.

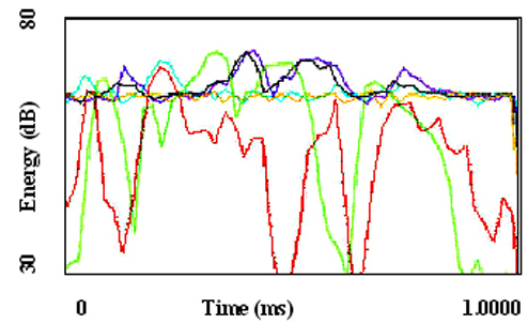
Among the filters applied, Noise reduction filter was most successful in improving the Signal to Noise Ratio (SNR) as well as preserving Speaker dependent features of the speech signal, than any other filters when applied to noisy speech samples and corresponding simulated samples. With the help of energy spectrum, level of the embedded noise in the recorded speech is estimated [4, 5]. A comparative study is conducted for evaluating the SNR improvement by the application of Noise reduction filter by studying energy contour of Noisy Speech-I, Noisy Speech-II, Noisy Speech-III, corresponding embedded noise signal and their corresponding simulated samples.

When the original noisy speech samples were subjected to Noise reduction filter, an average improvement of 14.82 dB, 22.55 dB and 16.61 dB in SNR by Noisy Speech-I, Noisy Speech-II and Noisy Speech-III respectively was observed. The Simulated speech samples when subjected to Noise reduction filter, Noisy Speech-I, Noisy Speech-II and Noisy Speech-III produced an average improvement in SNR by 19.38 dB, 14.75 dB and 16.23 dB respectively given

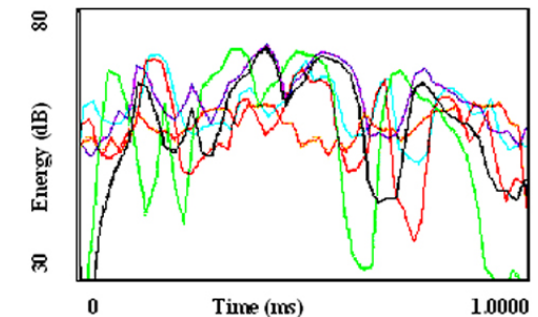
(Figs. (2, 3 & 4) and Tables 4 & 5). A maximum reduction in noise embedded in speech signal is observed by the application of this filtering technique.



**Fig. (2).** Energy Contour of Original noisy speech signal  
Corresponding embedded noise signal  
Original noisy speech signal after-applying noise reduction filter  
Specimen speech signal  
Simulated speech signal  
Simulated speech signal after Applying noise reduction filter



**Fig. (3).** Energy Contour of Original noisy speech signal  
Corresponding embedded noise signal  
Original noisy speech signal after-applying noise reduction filter  
Specimen speech signal  
Simulated speech signal  
Simulated speech signal after applying noise reduction filter



**Fig. (4).** Energy Contour of Original noisy speech signal  
Corresponding embedded noise signal  
Original noisy speech signal after-applying noise reduction filter  
Specimen speech signal  
Simulated speech signal  
Simulated speech signal after-applying noise reduction

**Table 4. SNR Improvement by the Application of Noise Reduction Filters Upon Original Noisy Speech**

Mode of Recording	SNR Improvement in dB	Average Improvement in dB
Noisy Speech-I	14.82	18
Noisy Speech-II	22.55	
Noisy Speech-III	16.61	

**Table 5. SNR Improvement by the Application of Noise Reduction Filters Upon Simulated Speech Samples**

Mode of Recording	SNR Improvement in dB	Average improvement in dB
Noisy Speech-I	19.38	16.79
Noisy Speech-II	14.75	
Noisy Speech-III	16.23	

With the help of Computerized Speech Laboratory (CSL), comparative study (wideband, narrowband and LPC analysis) on noisy speech samples before and after the application of filters has been conducted and speaker dependent features are found to be consistent.

#### 4. CONCLUSION

Noise Reduction technique for enhancing the noisy speech recorded in any of the three modes of recordings (direct, telephonic and mobile); proves to be efficient in terms of improving the SNR and preserving the perceptual and acoustic features of the speech signal. Wideband, Narrowband and LPC analysis of both the original and simulated samples before and after filtering have shown that

controlled application of Noise reduction filter can remove the embedded noise in the recorded speech to a maximum extent by preserving the Speaker Specific Information useful for Speaker Identification. Application of Noise reduction filter improved the SNR by 18 dB and 16.79 dB for original and simulated samples respectively. The Mean Opinion Score (MOS) test conducted gave human judgment over subjective measure of the enhanced speech quality. The degree of efficiency of Noise reduction filter in enhancing noisy speech for Forensic Speaker Identification is evaluated by conducting statistical study for samples in three modes of recordings and their corresponding simulated samples are demonstrated in this study. Experiment performed by Digital Equalizer on noisy speech revealed increasing variability on relative amplitude.

#### 5. ACKNOWLEDGEMENTS

The authors are highly grateful to the Director, Central Forensic Science Laboratory, Directorate of Forensic Science, Ministry of Home Affairs, Govt. of India, for providing constant support and facilities without which the study would not be possible.

#### REFERENCES

- [1] Koenig, B. E. Enhancement of forensic audio recordings. *J. Audio Eng. Soc.*, **1988**, 36(11), 884-893.
- [2] Wan, E. A.; Nelson, A. T. *Removal of noise from speech using the dual EKF algorithm 0-7803-4428-6/98* © Proceedings of IEEE International Conference, **1998**, pp. 381-384.
- [3] Muhmoudi, D.; Drygujlo, A. *Combined Wiener and Coherence filtering in wavelet domain for microphone array speech enhancement, 0-7803-4428-6/98* © Proceedings of IEEE International Conference, **1998**, pp. 385-388.
- [4] Martin, R. *An efficient algorithm to estimate the instantaneous SNR of speech signals*, International Proceedings of Eurospeech, ESCA, **1993**, pp. 1093-1096.
- [5] Martin, R. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Trans. Speech Audio Process.*, **2001**, pp. 504-512.

Received: February 13, 2009

Revised: February 21, 2009

Accepted: February 23, 2009

© Jiju et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.