Preface

Speech is the most extensively used mode of communication among human. If recorded properly it is also counted as one of the prime physical evidences by the court of law all over the world whether it is a call of extortion, threatening, bribery, kidnapping or an act of terrorism. Noisy or corrupt audio recordings loose its value as evidence if the Judge and Jury cannot understand the text of conversation which demands noise-corrupted speech to be made recognizable and hence the process of noise cancellation/reduction is necessary in such cases for Speech Recognition and Speaker Identification.

The task of noise reduction in Forensic case examination is in two folds, either restricted filtering for Speaker Identification or Intelligibility Enhancement (Speech Recognition) by noise subtraction for Court’s review. Speech Recognition is the process of recognizing what is spoken in the context (questioned sample) by the speaker rather than who is spoken. But in the case of Speaker Identification, the expert try to find out who has uttered the given text by comparing the voice in the questioned sample with the given sample voice of the suspect(s). Understanding the characteristics of embedded noise structure associated with the recorded speech samples received for Forensic Speaker Identification is one of the challenging tasks to be tackled up by the experts in the field.

Considering human/machine interfaces as a major area of applications, it is obvious that the signal becomes more challenging as the acoustic environment becomes more complex and hostile. Unnecessary disturbance within the frequency band of the speech in a communication channel or in other words any undesired sound is generally termed as “Noise”. Based on the means by which it is produced, noise has been generally partitioned into Additive and Convolutional. Additive noise has been considered as those produced from energy-radiating source and can be further classified into time-correlated and time uncorrelated. Audio signals are usually a blend of correlated and uncorrelated signal components. In fact, a correlated signal is predictable and that uncorrelated signal is random. Convolutional is a product of the transfer function of the environment or the electrical i.e. transfer function of the communication channel as Tonal noises, Broadband noise, Banded noises etc.

The tedious part of most restoration jobs is eradicating continuous broadband noise, like tape hiss. The filters provided for this usually produce only modest results with default settings, and learning how to set each parameter takes a lot of experimentation. The performance of systems for Speaker Identification can be quite good with clear speech than with noisy speech or
degraded speech. Considering human/machine interfaces as a major area of application, it is obvious that the signal becomes more challenging as the acoustic environment becomes more complex and hostile.

It is quite possible that analysis of information on tape recordings constitutes one of the single most powerful tools criminal investigators have at their command. Indeed, it is without question that it could be exploited to a far greater extent than it is currently and that, with somewhat improved knowledge and techniques, the amount of information that could be captured, and ultimately utilized could be expanded by a substantial factor. That is, even present technology and procedures would permit a considerably more effective use of tape recordings if they were properly utilized. It is conceded that only rather primitive techniques presently are available that permit enhancement of the intelligibility of speech recorded on tape in the presence of noise and/or other distorting conditions.

In the proposed study, noisy speech samples received in actual crime cases are characterized in order to classify according to their various parametric values with Signal Processing Filters for Speech Recognition and Speaker Identification. On the basis of the characterization and the class of noise associated with Forensic speech samples, customized set of Signal Processing Filters are categorically identified for each specific set of noise. Statistical distribution of noise on the crime samples is also studied based on the applied Signal Processing Filters’ limitation for Speech Recognition and Speaker Identification.