Performance Analysis of Noise Filters and Speech Enhancement Techniques in Adverse Mixed Noisy Environment for HCI

Article · October 2012

CITATIONS
5

READS
286

2 authors:

Urmila Shrawankar
G H Raisoni College of Engineering, Nagpur, (MS) India
174 PUBLICATIONS **358** CITATIONS
SEE PROFILE

V. M. Thakare
249 PUBLICATIONS **718** CITATIONS
SEE PROFILE

International Journal of Research and Reviews in Computer Science (IJRRCS)
Vol. 3, No. 5, October 2012, ISSN: 2079-2557

www.sciacademypublisher.com/journals/index.php/IJRRCS

Performance Analysis of Noise Filters and Speech Enhancement Techniques in Adverse Mixed Noisy Environment for HCI

Urmila Shrawankar[1] and Vilas Thakare[2]

[1]G H Raisoni College of Engineering, Nagpur, India
[2]PG Dept. of Computer Science, SGB Amravati University, Amravati, India

Email: urmila@ieee.org

Abstract – Mixed real world noisy environment is the major adverse condition for speech recognition systems. Due to variety of noise and their levels, it is very difficult to identify an appropriate filter or speech signal enhancement technique a prior, and it directly affects word recognition accuracy. The aim of this paper is to implement and test the performance of various filters and speech signal enhancement techniques for different category of noise. In order to illustrate the analysis of enhancement techniques, four categories of noise are considered and tested for the performance of a system. Three categories of noise (NOIZEUS, NOISEX, Random Gaussian white noise) are added artificially to the speech sample and speech recorded at real world mixed noisy environment is considered as the fourth category of noise. The performance has been measured using objective measures by computing SNR improvement test as well as subjective quality evaluation is done using an informal listening test, spectrogram and waveform observation. This experimental study contributes performance improvement of Automatic Speech Recognition (ASR) in a wide range of unexpected mixed noisy environments and results improvement for HCI.

Keywords – ASR for Real world Mixed Noisy Environment, Informal Subjective Test, Objective Measures, Speech Enhancement, Speech User Interface (SUI)

1. Introduction

Robust speech recognition in adverse conditions is one of the most challenging areas of speech recognition [9]. This so-called robustness problem not only leads to a significant degradation in performance but also hampers the fast commercialization of speech recognition applications.

The gap between human and machine recognition remains large, especially in adverse natural environments. Error rate of automatic speech recognition system depends on two conditions. The first is, if noise type is known and second is training and testing environment matches. In fact there are countless different kinds of noise in real world environment. Training-test mismatch always occurs when mixed noise is involved. Therefore the performance of a system is totally dependent on the effective noise reduction and enhancement algorithms [2].

The current Speech Recognition Systems have been developed using two categories of enhancement techniques. The first is Optimal Filtering, for example, spectral subtraction, Wiener Filtering, Kalman Filtering, or subspace decomposition [25, 27, 34, 36, 38, 41] and the other one is Optimal Estimation, for example, minimum mean-square error, or maximum a posteriori, estimators [32, 35, 37, 39, 40].

In optimal filtering, no specific knowledge about the speech is assumed, except that is independent of the noise; like, spectral subtraction or Wiener filtering. In optimal estimation, a priori knowledge of the probability distribution of the speech is assumed and this is used to derive the estimators, for example, minimum mean-square error (MMSE), maximum a posteriori (MAP), or perceptually weighted Bayesian spectral estimators [2].

The statistical approaches require pre-specified parametric models for both the signal and the noise. The model parameters are obtained by maximizing the likelihood of the training samples of the clean signals using expectation–maximization (EM) algorithm. Since the true model for speech remains unknown, a variety of statistical models have been proposed. Short-time spectral amplitude (STSA) estimator and log-spectral amplitude estimator (LSAE) [7] assume that the spectral co-efficients of both signal and noise obey Gaussian distribution. Their difference is that STSA minimizes the mean square error (MMSE) of the spectral amplitude while the LSAE uses the MMSE estimator of the log-spectra. LSAE is more appropriate because log-spectrum is believed more suitable for speech processing.

A data-driven approach to a priori signal-to-noise ratio (SNR) plays an important role in many speech enhancement algorithms [18]. The most widespread approach to determine the a priori SNR estimates is the decision-directed (DD) estimator of Ephraim and Malah [12]. As an alternative, Cohen [12] proposed a non-causal a priori SNR estimator. It may be used with a wide range of speech enhancement techniques, such as, Minimum Mean Square Error (MMSE) (log) spectral amplitude estimator, the super Gaussian joint maximum a posteriori (JMAP) estimator, or the Wiener filter. Wiener filter and both the short-time spectral amplitude (STSA) and the log-spectral amplitude (LSA) estimator are

proposed by Ephraim and Malah [3].

The goal of speech enhancement is to remove the noise while preserving the clean speech as much as possible.

Subjective evaluation of speech enhancement algorithms is further complicated by the fact that the quality of enhanced speech has both signal and noise distortion components, and it is not clear as to whether listeners base their quality judgments on the signal distortion, noise distortion or both [5,6].

Performance of speech recognition systems is dependant on all major levels of the system like feature extraction, feature enhancement, speech modeling, training & testing and adoption. Since this work is focusing on signal enhancement part only, the performance evaluation is presented for noise filters and enhancement algorithms only.

Performance of the system is tested for four categories of noise at four signal-to-noise ratio levels and all the considered combination of enhancement techniques by three classes of speech enhancement algorithms. The noisy signals are enhanced using two categories of techniques like traditional noise filters, and speech signal enhancement algorithms.

The analysis of performance is done using two measures i.e. objective and subjective test. The SNR test is considered as an objective measure while the subjective quality evaluation is done using an informal listening test, spectrogram as well as waveform observation. The listening test is performed by normal hearing persons.

The main objective of the present study is to report on the evaluation of conventional as well as different objective and subjective measures that could be used to predict overall speech quality and speech/noise distortions introduced by representative speech enhancement algorithms [20,23] from various classes [5].

The paper is organized as section 2 for overview of the system, section 3 explains Speech Enhancement Techniques, and section 4 gives in detail of proposed work and Results section 5 for Performance Analysis and finally section 6 for concluding remarks of the study.

2. The System Overview

The system assumes that the speech signal is corrupted with additive background environment noise, refer fig. 1.

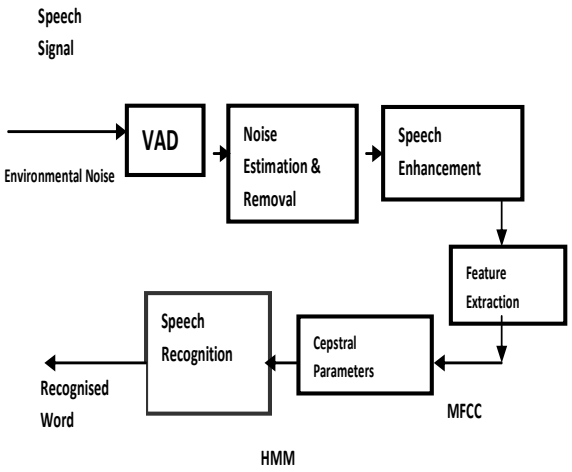


Figure 1. The System Overview.

2.1. Voiced, Unvoiced Or Silence (VUS) Detection :

The presence of speech in the input audio signal can be detected by implementing Voice Activity Detection (VAD) techniques [19]. The decision whether audio is voiced, unvoiced or silence (VUS) is based on the different features extracted from speech signals. The extracted features like Energy, Zero-Crossing Rate ZCR, Average Magnitude Difference Function (AMDF), spectral entropy, Teager Energy operator etc. help in making the decision. The detection of speech presence can be calculated by detecting the beginning and end-point of an utterance. This two point detection algorithm is based on the measure of the signal, zero crossing rate and short-time energy.

2.2. Pre-processing

The entire speech recognition system is broadly divided into front-end processing for speech feature extraction and back-end processing for HMM decoding. First, the noise is detected and then it applies the filters to remove it. The enhancement methods improve the quality of distorted signals. Most of the efforts are focused on speech signal enhancement techniques, which is the way to compensate for the additive unknown background noise effects. The methods to compensate for noise can be implemented at the front-end or the back-end or both. The front-end processing method is to suppress the noise and get more robust parameters, while back-end processing is to compensate for noise and adapt the parameters inside the HMM system.

2.3. Feature Extraction

The feature extraction step also plays an important role in performance of ASR. This system uses Mel Frequency Cepstral Coefficients (MFCCs) and is the most suitable option in current development, since it can obtain more accurate results of speech recognition under a no-noise condition. [9]. MFCCs are coefficients that represent audio. They are derived from a type of cepstral representation of the audio clip. The frequency bands are positioned logarithmically (on the mel scale) which approximates the human auditory system's response more closely than the linearly-spaced frequency bands obtained directly from the FFT or DCT. This can allow for better processing of data

2.4. Training & Testing

Extracted Cepstral parameters passed to Recognition System for Training & Testing, The Hidden Markov model (HMM) is used by this system. A statistical based HMM is a successful model and plays an important role in speech recognition system. The developed HMM with gain adaptation uses for the speech enhancement and for the recognition of clean and noisy speech. In contrast to the frequency-domain models, the density of log-spectral amplitudes is modeled by a Gaussian mixture model (GMM) with parameters trained on the clean signals [7].

The major difficulty in this area is the unknown nature of additive background noise and due to the non-stationery and unknown nature precise methods for both speech signal and noise are unavailable, thus speech enhancement problem remains unsolved [17].

3. Speech Enhancement Techniques

3.1. Speech Enhancement Approaches

Current speech processing algorithms can roughly be divided into three domains, spectral subtraction, sub-space analysis and filtering algorithms:

- Spectral Subtraction: Spectral Subtraction algorithms operate in the spectral domain by removing, from each spectral band, that amount of energy which corresponds to the noise contribution. While spectral subtraction is effective in estimating the spectral magnitude of the speech signal, the phase of the original signal is not retained, which produces a clearly audible distortion known as "ringing".

Spectral subtraction estimator

$$\tilde{S}(e^{j\omega}) = [|X(e^{j\omega})| - \mu(e^{j\omega})]e^{j\theta_x(e^{j\omega})} \quad (1)$$

- Signal Sub-Space : Sub-Space analysis operates in the autocorrelation domain, where the speech and noise components can be assumed to be orthogonal, whereby their contributions can be radialy separated. Unfortunately, finding the orthogonal components is computationally expensive. Moreover, the orthogonality assumption is difficult to motivate.

$$r = \tilde{y} - y = (Hy + Hw) - y = (H - I)y + Hw$$

$$\Rightarrow r = r_y + r_w \quad (2)$$

where,

$$r_y = (H - I)y$$

Represents signal distortion and

$$r_w = Hw$$

Represents the residual noise

- Filtering Techniques: Filtering algorithms are time-domain methods that attempt to either remove the noise component (Wiener filtering) or estimate the noise and speech components by a filtering approach.
- A Priori / Decision-Directed (DD) and A Priori Signal-To-Noise Ratio (SNR) Estimator : This estimator is called a "decision-directed" [3] type estimator, because it is updated based on the previous frame's amplitude estimate. As it can be seen from the equation that the first term comes from the amplitude estimator of the previous frame while the second term is an ML estimate determined from the a posteriori SNR.
- The motivation for using an ML approach is that ML estimation can estimate an unknown parameter of a given PDF without any prior assumptions on the parameter. This estimator maximizes the joint conditional PDF of noisy spectral amplitude, given clean signal variance k and noise variance.
- The computation of spectral weighting rules in speech enhancement is often driven by the a posteriori and a priori signal-to-noise ratio (SNR) [12]. However, the performance of most weighting rules is dominantly determined by the a priori SNR, while the a posteriori SNR acts merely as a correction parameter in case of low a priori SNR [12]. With a look ahead of a few frames it is capable of discriminating between speech onsets and irregularities in the a posteriori SNR corresponding to noise only, resulting in less transient distortion and less musical tones.

The decision-directed a priori SNR estimator of Ephraim and Malah [30,32] is given by

$$\hat{\xi}_{k,D-D}(n) = \alpha \frac{\hat{A}_k^2(n-1)}{\lambda_d(k,n-1)} + (1-\alpha)P[\gamma_k(n)-1] \quad (3)$$

a posteriori SNR of the kth spectral component in the nth analysis frame, respectively.

The P function is given by:

$$P[\gamma_k(n)-1] \triangleq \begin{cases} \gamma_k(n)-1 & \gamma_k(n)-1 \geq 0 \\ 0 & otherwise \end{cases} \quad (4)$$

- Maximum A Posteriori Estimator: A maximum a posteriori probability (MAP) estimate is a mode of the posterior distribution. The MAP can be used to obtain a point estimate of an unobserved quantity on the basis of empirical data. It is closely related to Fisher's method of maximum likelihood (ML), but employs an augmented optimization objective which incorporates a prior distribution over the quantity one wants to estimate.

Maximum Likelihood Estimate of θ.

$$\hat{\theta}_{ML}(x) = \arg\max_{\theta} f(x|\theta) \quad (5)$$

$$\hat{\theta}_{MAP}(x) = \arg\max_{\theta} \frac{f(x|\theta)\,g(\theta)}{\int_{\Theta} f(x|\theta')\,g(\theta')\,d\theta'} = \arg\max_{\theta} f(x|\theta)\,g(\theta). \quad (6)$$

- Wiener Filter: The goal of the Wiener filter [29] is to filter out noise that has corrupted a signal. It is based on a statistical approach.

Typical filters are designed for a desired frequency response. However, the design of the Wiener filter takes a different approach. One is assumed to have knowledge of the spectral properties of the original signal and the noise, and one seeks the linear time-invariant filter whose output would come as close to the original signal as possible. Wiener filters are characterized by the following:

a) Assumption: signal and (additive) noise are stationary linear stochastic processes with known spectral characteristics or known autocorrelation and cross-correlation
b) Requirement: the filter must be physically realizable/causal (this requirement can be dropped, resulting in a non-causal solution)
c) Performance criterion: minimum mean-square error (MMSE)

$$SNR_{prio}(f_k) \triangleq \frac{E\{|S_k|^2\}}{E\{|B_k|^2\}} \quad (7)$$

$$SNR_{post}(f_k) \triangleq \frac{|X_k|^2}{E\{|B_k|^2\}} \quad (8)$$

This filter is frequently used in the process of de-convolution; for this application.

3.2. Speech Enhancement Algorithms:

This section explains the algorithms and the techniques like,

- The Generalized Spectral Subtraction Method by Boh Lim Sim [25]: In this method an optimized estimator based on spectral subtraction assumptions is derived. This estimator uses estimates of a priori and a posteriori SNR in its gain function. Also a flooring function is incorporated which uses a function to floor very small values of amplitude estimate. (Hasan04) Decision-Directed method is used for estimation of A priori SNR.
- Spectral Subtraction based on Berouti [26]: This method uses Nonlinear Spectral Subtraction based on a power spectral subtraction with adjusting subtraction factor. The adjustment is done according to local a postriori SNR.
- Spectral Subtraction based on Boll [27]: This Spectral Subtraction method is based on Amplitude spectral subtraction including Magnitude Averaging and Residual noise Reduction.
- Multi-Band Spectral subtraction postriori by Kamath [28]: Multi-band Spectral subtraction based on adjusting subtraction factor. The subtraction with the adjustment is according to local a postriori SNR and the frequency band.
- A priori SNR using Decision-Directed by Wiener-Scalart [29]: This is the Wiener filter based on tracking a priori SNR using Decision-Directed method. In this method it is assumed that SNRpost=SNRprior +1. based on this the Wiener Filter can be adapted to a model like Ephraims model in which a gain function which is a function of a priori SNR and a priori SNR is being tracked using Decision Directed method.
- Posteriori SNR by Ephraim and Malah [30, 32]: Estimation is done on a priori i.e. the earlier frames and a posteriori from the later frames, select 0 for a posteriori SNR estimation and 1 for a priori.
- MMSE log-spectral Estimator by Cohen [31]:It is Denoising methods using MMSE log-spectral Estimator
- MMSE-STSA by Ephraim [32]: It is a Spectral Amplitude Minimum Mean Square Error Method. Short time Spectral Amplitude Minimum Mean Square Error Method for Denoising noisy speech. Signal is the input noisy speech. The output is the restored estimate of clean speech.

4. Proposed Work

The experimental evaluation is done using MatLab software Ver R10 [42]. A software is prepared to record samples, testing samples for voice / unvoiced / silence, filter noise, to enhance the quality of signal, extract the features from speech signal, normalize the feature vector, train the machine to recognize spoken isolated words (digits 0-9) and finally test the accuracy.

4.1. Steps

- Step I: Samples Collection : Recording: The experiment starts from recording speech samples (words) with sampling rate of 8 kHz and time duration of each word sample is 3 Sec. because of particularities of human recognition system and due to possible noise. Recorded speech samples stored using WAV file format under Windows platform. Samples are recorded from 25 speakers (male and female) and five utterances of ten isolated words (number digits 0-9) for each word for every speaker. Recording is done in glass cabin for clean environment and outside for natural environment (Mixed environment).
- Step II: Adding Noise: Four categories of noise are considered.

a) The original NOIZEUS database [44] is used and a selection of different real-world noise that have been artificially added to the speech at SNRs of 15dB, 10dB, 5dB, 0dB and -5dB. These noisy signals have been recorded at different places like Airport, Car, Exhibition, Restaurant, Station, Street, Train etc.

b) NOISEX database [43] is used for unknown SNR level (NOISEX sample files do not mention SNR level) noise of different type like Bubble, Buccaneer, Engine, Factory, Hfchannel, Machinegun, Pink, Volvo, White etc.

c) Random Gaussian White Noise is used to corrupt clean or unknown natural noisy sample.

d) Real Environment (Natural) unknown noise (mixed environment) samples are collected from different speakers at various outdoor places or locations, open environment.

- Step III: Voice Activity Detector (VAD)

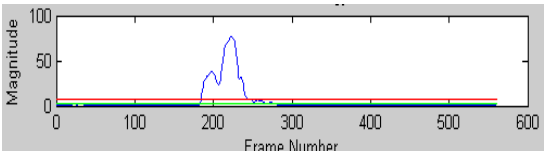
a) Windowing and Framing: A variable size framing (between 10 ms and 30 ms), windowing and overlap is used to construct the matrix. More specifically, the signal was divided into n-ms frames with m% overlap between frames. The samples in the n-ms frame were used to construct a Toeplitz covariance matrix. The n-ms frames were further subdivided into l-ms frames with m% overlap. The noisy data in each l-ms frame were enhanced using the same eigenvector matrix derived from the Toeplitz covariance matrix.

b) Voice / Unvoiced/ Silence (VUS) separation: The detection of the speech presence is calculated by detecting the beginning and end-point of an utterance using VAD. This two point detection algorithm is based on measures of the signal, zero crossing rate and short-time energy.

4.2. Short Time Energy and Zero Crossing Rate :

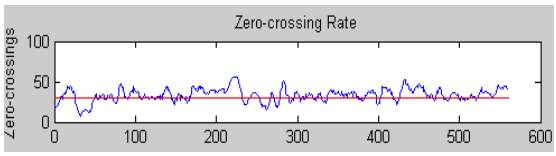
4.2.1. Short Time Energy

$$E(n) = \sum_{i=-50}^{50} |s(n+i)|, \quad (9)$$

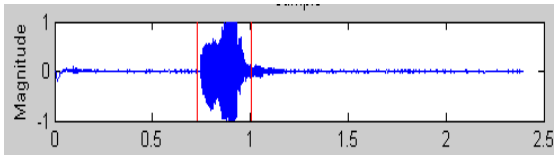


4.2.2. Zero Crossing Rate

$$IZCT = \text{MIN}(IF, \overline{IZC} + 2\sigma_{IZC}) \quad (10)$$



Separation of Voiced part of the signal



Sample for word 'Seven'

At the time of separating voiced part from the signal 10 ms is kept on both the sides, assuming that this part of the signal is unvoiced.

This voiced part is further sent to noise filtering and enhancing.

- Step IV: Noise Cancellation: Under this Pre-emphasis step, Filters are implemented to reduce or filter the noise. This category of filters is implemented for removing the noise from speech signals that are corrupted due to additive background noise.

In this experiment four fundamental traditional filters FIR like high-pass, low-pass, band-pass, and band-stop filters are tested. These filters are used for different frequency ranges.

- Step V: Enhancement: Speech enhancement algorithms attempt to recover a clean speech signal from a degraded signal containing additive noise. The evaluation of performance measures [5] performed using four category of noisy environments and eight speech enhancement algorithms encompassing different classes such as spectral subtractive (multiband spectral subtraction, and spectral subtraction using reduced delay convolution and adaptive averaging), statistical-model-based (MMSE, log-MMSE, and log-MMSE under signal presence uncertainty) and Wiener-filtering type algorithms (the a priori SNR estimation based method, the audible-noise suppression method [14,24] are considered and tested for performance.
- Step VI: Performance Evaluation: Following categories of methods and Corpus are considered for evaluation [1, 4, 5, 6, 8, 10, 11, 15, 21, 22]
  - 4 Types of Traditional Fundamental Filters
    1) High-pass filter
    2) Low-pass filter
    3) Band-pass filter
    4) Band-shop filter
  - 8 Algorithms of 3 Class of Enhancement Techniques
    1) Spectral-Subtractive algorithms
    2) Generalized Spectral Subtraction - Boh Lim Sim
    3) Spectral Subtraction - Boll
    4) Spectral Subtraction - Berouti
    5) Multi-band Spectral subtraction postriori - Kamath
    6) Wiener Filtering
    7) Wiener Filter a priori SNR - Wiener-Scalart
    8) Statistical-Model based algorithms
    9) Malah posteriori SNR, Ephraim and Malah
    10) MMSE-STSA - Ephraim
    11) log MMSE Spectral Estimator - Cohen
  - Noise Corpus :
    - NOIZEUS :
    1) SNR level : 0, 5, 10, 15 dB
    2) Noise Type : Airport, Babble, Car, Exhibition, restaurant, street, station, train and clean
    - NOISEX :
    1) SNR level : unknown
    2) Noise Type : Babble, buccaneer, factory, machinegun, Volvo, pink and white.
    - Random Gaussian White Noise
    1) SNR level : 0, 5, 10, 15 dB
    - Real Environment Unknown Natural Noise
  - Subjective and Objective Performance [1, 4, 5, 6, 8, 10, 11, 15]: The performance measure is done using two performance measures i.e. objective and subjective measures.
  - Objective analysis (SNR improvement test): A common approach for evaluating the quality of an estimated source signal is to compute the Signal-to-Noise Ratio (SNR) [16] between the energy of the reference, i.e., the clean target signal, and that of the distortion, Spectrogram and Waveform.
  - The SNR improvement test is considered as an objective measure by calculating the SNR and is compared with SNR before implementing filter; spectrogram as well as waveform plotted and clarity was observed after implementing the filter or enhancement algorithm.
  - Subjective analysis (Listening Test): The subjective quality evaluation [15] is done by using a listening test. The listening test performed by normal hearing persons and the following parameters were observed.
    - Overall quality (Intelligibility, Fidelity, Suppression etc)
    - Musical noise salience , musical noise or other artifacts
    - Preference
  - Listening Test: Subsequent Informal listening tests conducted for subjective evaluation. This test is a qualitative test. Ten volunteers were requested to evaluate the performance of the speech enhancement methods that were implemented in this project. The

listeners gave their decisions on an individual basis. Ten speech samples were considered, each (digit) isolated word sample for every listener. First all the samples were numbered and played in the same order in which it was enhanced. The listeners ranked the methods based on the intelligibility and quality of the enhanced speech. On the basis of this test following observations have been noted down, listed in sec VI.

5. Results

The signal was filtered using all approaches given in tables. The SNR, waveform and spectrogram are obtained for the clean, noisy and enhanced signals. SNR test results are given in tables 1-7. The SNR, waveform and spectrogram are obtained for all categories. Some limited samples of results are shown here because of limitation of page space.

Table 1. Noise type : noizeus – babble

SN	Filter	SNR 0	SNR 5	SNR 10	SNR 15
		0.0972	0.3504	0.1955	0.1880
	Fundamental Filters				
1	Low Pass	3.508	4.967	2.216	2.916
2	High Pass	0.265	0.152	0.558	0.558
3	Band Pass Filter	3.266	2.2265	3.0564	1.0564
4	Band Stop	4.5302	5.3384	5.1215	6.2393
	Enhancement Algorithms				
1	GSS Boh Lim Sim 98	5.1772	5.2237	5.2135	5.2117
2	SS Boll 79	6.5483	7.0273	8.1161	9.0364
3	SS Berouti 79	16.2654	18.7906	18.6512	18.4430
4	MBSS Postriori Kamath 2002	14.4385	14.8384	14.7106	14.7091
5	a Priori SNR Wiener-Scalart 96	29.1508	29.7968	29.9176	29.9281
6	Posteriori SNR Ephraim and Malah, 1985	2.4035	1.9091	1.9410	1.9615
7	MMSE-STSA - Ephraim 1984	0.0834	0.6980	0.6355	0.6216
8	logMMSE - Cohen 2004	14.7495	20.4443	21.5972	20.8799

Table 2. Noise type : noizeus – exhibition

SN	Filter	SNR 0	SNR 5	SNR 10	SNR 15
	SNR Before Filter	0.4204	0.2918	0.1539	0.0703
	Fundamental Filters				
1	Low Pass	2.905	3.679	3.916	4.017
2	High Pass	0.365	0.753	1.657	1.9873
3	Band Pass Filter	2.672	3.164	3.473	3.9564
4	Band Stop	1.5021	1.8421	2.111	2.933
	Enhancement Algorithms				
1	GSS Boh Lim Sim	5.1966	5.2066	5.2087	5.2112
2	SS Boll	5.8522	9.0390	8.1514	9.2238
3	SS Berouti	16.8687	18.4533	18.6128	18.5924
4	MBSS Postriori Kamath	13.9955	14.6413	14.6405	14.6378
5	a Priori SNR Wiener-Scalart	28.4426	29.4745	29.4160	29.8698
6	Posteriori SNR Ephraim and Malah	2.0676	1.9651	1.9900	2.0559
7	MMSE-STSA – Ephraim	0.5000	0.5738	0.6489	0.6223
8	logMMSE	7.1912	20.4801	20.8561	21.8277

Table 3. Noise type : noizeus – street

SN		SNR 0	SNR 5	SNR 10	SNR 15
	SNR Before Filter	0.0205	0.4419	0.1037	0.0683
	Fundamental Filters				
1	Low Pass	1.500	1.927	2.634	2.982
2	High Pass	0.151	0.223	0.382	0.376
3	Band Pass Filter	4.236	2.932	3.428	2.732
4	Band Stop	3.254	3.829	4.578	3.924
	Enhancement Algorithms				
1	GSS Boh Lim Sim	5.1896	5.2076	5.1942	5.2082
2	SS Boll	5.7470	9.4323	7.1052	8.6803
3	SS Berouti	19.1040	18.4605	18.8893	18.3354
4	MBSS Postriori Kamath	14.0818	14.8176	14.7278	14.5878
5	a Priori SNR Wiener-Scalart	28.8038	29.4848	29.4191	29.7691
6	Posteriori SNR Ephraim and Malah	2.0499	1.9289	2.1246	1.9872
7	MMSE-STSA – Ephraim	0.5081	0.6741	0.6382	0.6133
8	logMMSE	7.4778	22.3592	20.5288	20.9700

Table 4. Noise type : noizeus – station

	Filter	SNR 0	SNR 5	SNR 10	SNR 15
	SNR Before Filter	0.1048	0.7281	0.3118	0.1274
	Fundamental Filters				
1	Low Pass	2.732	2.573	3.692	3.884
2	High Pass	0.213	0.306	0.297	0.383
3	Band Pass Filter	3.618	2.154	1.835	1.265
4	Band Stop	4.2364	3.736	4.921	4.389
	Enhancement Algorithms				
1	GSS Boh Lim Sim	5.2145	5.2238	5.2204	5.2089
2	SS Boll	6.5083	7.7683	8.6582	9.4984
3	SS Berouti	17.7971	18.4065	18.4885	18.6216
4	MBSS Postriori Kamath	14.5829	14.6864	14.6877	14.6232
5	a Priori SNR Wiener-Scalart	28.8869	29.1152	29.5275	29.7025
6	Posteriori SNR Ephraim and Malah	1.9077	2.5514	2.2550	1.9559
7	MMSE-STSA – Ephraim	0.6593	0.6619	0.5840	0.6268
8	logMMSE	6.2490	19.5363	20.0484	22.6036

Table 5. Noise type : noisex

	Filter	Babble	Factory	Pink	White
	SNR Before Filter	3.9630	4.5259	3.0263	-0.6483
	Fundamental Filters				
1	Low Pass	2.3557	5.6353	5.4994	421.84
2	High Pass	5.238	3.9694	3.0791	8.194
3	Band Pass Filter	3.0195	2.0902	2.6657	3.129
4	Band Stop	4.0989	4.7347	5.1726	5.2075
	Enhancement Algorithms				
1	GSS Boh Lim Sim	5.1827	5.1972	5.1897	5.1829
2	SS Boll	6.3951	6.6336	5.7754	6.5067
3	SS Berouti	19.4219	18.7426	15.9158	17.5463
4	MBSS Postriori Kamath	14.5183	14.3871	14.2512	14.5695
5	a Priori SNR Wiener-Scalart	29.0730	29.0662	29.2441	29.0898
6	Posteriori SNR Ephraim and Malah	2.0374	1.8025	2.1127	2.2042
7	MMSE-STSA –Ephraim	0.4027	0.0756	0.7274	0.6317
8	logMMSE	16.1936	18.2053	17.9909	17.0864

Table 6. Noise type : random gaussian white noise

SN	Filter	SNR 0	SNR 5	SNR 10	SNR 15
	SNR Before Filter	0.7551	1.5952	1.5952	0.4743
	Fundamental Filters				
1	Low Pass	9.78	10.606	10.198	9.9862
2	High Pass	2.7898	1.0284	3.0512	4.0872
3	Band Pass Filter	3.0479	0.4432	2.9132	1.2646
4	Band Stop	6.1372	7.9092	8.5639	8.2703
	Enhancement Algorithms				
1	GSS Boh Lim Sim	5.1389	5.2727	5.2727	5.2124
2	SS Boll	4.8109	4.4126	4.4126	5.6284
3	SS Berouti	17.5979	14.4433	14.4433	18.8380
4	MBSS Postriori Kamath	14.2342	13.2152	13.2152	14.4707
5	a Priori SNR Wiener-Scalart	27.7916	28.2276	28.2276	29.2637
6	Posteriori SNR Ephraim and Malah	2.2319	2.0404	2.0404	2.1764
7	MMSE-STSA – Ephraim	1.2573	1.9958	2.2958	2.4160
8	logMMSE	5.6379	0.1456	0.1456	15.9172

Table 7. Noise type : real environment unknown natural noise (mixed noise)

SN	Filter	Case I (location 1)	Case II (location 2)	Case III (location 3)	Case IV (location 4)
	SNR Before Filter	0.1034	0.8625	0.0153	0.1041
	Fundamental Filters				
1	Low Pass	5.338	1.4784	1.7923	3.7483
2	High Pass	6.5848	9.918	8.231	5.212
3	Band Pass Filter	2.1939	1.286	1.7185	1.1569
4	Band Stop	3.8308	4.2499	5.3498	3.5554
	Enhancement Algorithms				
1	GSS Boh Lim Sim	5.2104	2.5278	5.0115	4.3457
2	SS Boll	9.2597	29.3589	42.8410	8.7250
3	SS Berouti	18.5282	21.8965	27.8153	23.0776
4	MBSS Postriori Kamath	14.6648	14.1669	25.1645	23.5776
5	a Priori SNR Wiener-Scalart	29.9913	29.6483	30.3530	29.9431
6	Posteriori SNR Ephraim and Malah	1.9380	6.7726	1.6448	0.9184
7	MMSE-STSA – Ephraim	0.6129	0.0607	0.9229	0.4526
8	logMMSE	22.1824	15.9609	17.2468	20.3818

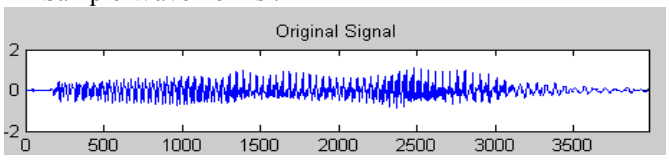
6. Performance Analysis

Several experiments have been carried out and evaluation is done

6.1. A. Objective Performance Evaluation:

- The SNR improvement results are shown in table 1-7.
- In Fig 2, we can observe the clarity in the speech waveform after implementing the filter than the original waveform.

Sample Wave Forms :



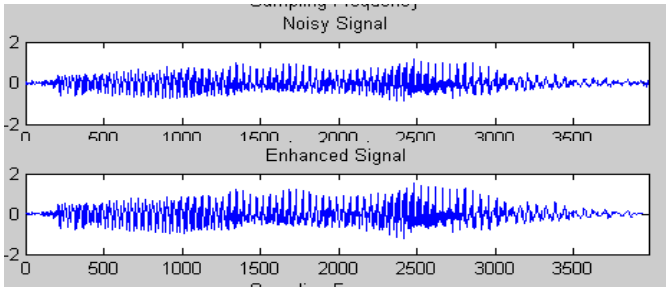


Figure 2. Wave Form for Wiener-Scalart 96 Algorithm For Noise Type : Babble noise at SNR 0

- Fig 3 shows improvement of clarity in the spectrogram. Red color indicates the SNR level. The spectrogram before filter was scattered and after filtering is very compact.
- We have studied all three results but due to space limitation, Waveform and Spectrogram only a sample result is given here.
- The SNR test shows the fundamental filters like low pass, high pass, band pass and band shop filter reduced noise but was unable to remove complete noise and distortion from the speech signal.
- In some cases fundamental filters have been effective, but not in all cases.
- None of the traditional Fundamental Filter indusial, gives good result in case of Real Environment Unknown Natural Noise (Mixed Noise)
- In real environment, noise category is unknown, on that time it is difficult to apply filter as per frequency range, only band pass filter that is the combination of law pass and high pass filter can be used.
- All Spectral Subtraction methods show improved performance.
- The multi-band spectral subtraction algorithm performed good as well as the statistical-model based algorithms in nearly all conditions
- Out of the four spectral-subtractive algorithms tested, the Priori SNR Wiener performed consistently the best across all conditions, in terms of overall quality.
- Out of all Spectral subtraction algorithms Priori SNR Wiener is the best followed by SS Berouti, MBSS Postriori Kamath. GSS Boh Lim Sim could not show any performance. SS Boll showed average performance.

C. Sample Spectrogram:

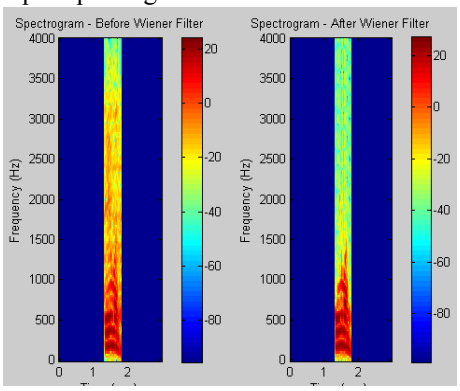


Figure 3. Spectrogram for Wiener-Scalart 96 Algorithm for Noise Type : Babble noise at SNR 0

- We have studied all three results but due to space limitation, Waveform and Spectrogram only a sample result is given here.
- The SNR test shows the fundamental filters like low pass, high pass, band pass and band shop filter reduced noise but was unable to remove complete noise and distortion from the speech signal.
- In some cases fundamental filters have been effective, but not in all cases.
- None of the traditional Fundamental Filter indusial, gives good result in case of Real Environment Unknown Natural Noise (Mixed Noise)
- In real environment, noise category is unknown, on that time it is difficult to apply filter as per frequency range, only band pass filter that is the combination of law pass and high pass filter can be used.
- All Spectral Subtraction methods show improved performance.
- The multi-band spectral subtraction algorithm performed good as well as the statistical-model based algorithms in nearly all conditions
- Out of the four spectral-subtractive algorithms tested, the Priori SNR Wiener performed consistently the best across all conditions, in terms of overall quality.
- Out of all Spectral subtraction algorithms Priori SNR Wiener is the best followed by SS Berouti, MBSS Postriori Kamath. GSS Boh Lim Sim could not show any performance. SS Boll showed average performance.
- Wiener filtering, assuming known signal and noise spectra, gives performance in all most all conditions
- The algorithms based on Statistical methods working well in almost all conditions.
- Out of the three statistical-model based algorithms examined, the log-MMSE algorithms performed the best.
- MMSE-STSA -Ephraim does not show performance.
- log MMSE method also produced consistently the lowest signal distortion compared to the statistical-model based methods, however it suffers from high noise distortion.
- Overall, the statistical-model based methods performed the best across all conditions, followed by the multi-band spectral subtraction method .
- In terms of overall quality and speech distortion, the following algorithms performed the best: Priori SNR Wiener-Scalart, MBSS Postriori Kamath, and SS Berouti.

6.2. B. Subjective Performance Evaluation: Informal Listening Tests

- The clarity improvement observed in case of Wiener filter.
- Statistical methods improved performance of signal and good quality observed.
- Distortion is not found much more after enhancement; all listeners could recognize all words.
- None of the methods could not remove complete (noticeable) noise in case of Random Gaussian White Noise
- Acceptable performance gives all the methods in case of Natural Environmental Noise (mixed noise) also; listener could not recognize the type and category of noise after implementation of enhancement method.
- Reverberation or echo effect is not noticed by any listener in enhanced word sample.

7. Discussion and Conclusion

- The key aim of this project is to compare the performance of speech enhancement methods that have been implemented. Since neither quality nor intelligibility can be measured well mathematically, quantitative as well as qualitative tests are required.
- Various tests are carried out to evaluate the performance of Speech Enhancement methods implemented: SNR Improvement Test as a Quantitative Test and Listening Test as a Qualitative Test
- Imposing constraints from human speech production model, variation within a person and among the persons as well as other speech characteristics produce better signal spectrum estimation and hence improve performance
- None of the method alone removes complete noise or distortion or both.
- Speech Signal enhancement is required at every step of ASR model to achieve the best performance of the system because neither performance improves in one step nor with any one of the algorithm alone.
- Multiple sources of noise are not considered here, however, a simple adaptive filter may be implemented for this problem.
- As this study is concentrated on adverse noisy environment only, variation in speech due to human factors are not considered here, normalization methods may help in this matter.
- Environment Adaptation model for training in real nature environment for mixed noise is required especially when unexpected noise occurs at the testing time.
- Performance in terms of word recognition accuracy rate by computer is not predictable though all listeners could recognize all words.
- Speech enhancement at back-end and front-end together is important to improve performance in human to machine interaction.
- Since the aim of this study especially for environmental Unknown Natural Noise (Mixed Noise), Priori SNR Wiener, MBSS Postriori Kamath, SS Berout and in some cases logMMSE – Cohen are the best solutions.
- The combinations of algorithms may give better performance than the indusial. This is the further part of our study.

References

[1] Valentin Emiya, et.al.., Subjective and Objective Quality Assessment of Audio Source Separation, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 19, No. 7, September 2011

[2] Ji Ming, et.al., A Corpus-Based Approach to Speech Enhancement From Nonstationary Noise, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 19, No. 4, May 2011
[3] Colin Breithaupt and Rainer Martin, Analysis of the Decision-Directed SNR Estimator for Speech Enhancement With Respect to Low-SNR and Transient Conditions, IEEE Transactions On Audio, Speech, And Language Processing Vol 19, No 2, Feburary 2011
[4] Ma, J., Hu, Y., Loizou, P., Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. J. Acoust. Soc. Amer. 125, 3387–3405. ,2009
[5] Yi Hu and Philipos C. Loizou, Evaluation of Objective Quality Measures for Speech Enhancement, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 16, No. 1, January 2008
[6] Yi Hu and Philipos C. Loizou, Subjective Comparison Of Speech Enhancement Algorithms, IEEE Transactions On Audio, Speech, And Language Processing, 2006,
[7] Jiucang Hao, et.al, Te-Won Lee, and Terrence J. Sejnowski, Speech Enhancement, Gain, and Noise Spectrum Adaptation Using Approximate Bayesian Estimation, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 17, No. 1, January 2009
[8] Dong Yu, Jinyu Li, and Li Deng, Calibration of Confidence Measures in Speech Recognition, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 19, No. 8, November 2011
[9] Weizhong et.al, Using Noise Reduction And Spectral Emphasis, ASRU 2003
[10] Jianfen Maa, Philipos C. Loizou, SNR loss: A new objective measure for predicting the intelligibility of noise-suppressed speech, Vol. 53, No. 3, March 2011
[11] Thierry Etame, et.al, Catherine Quinquis, Laetitia Gros, and Gérard Faucon, Towards a New Reference Impairment System in the Subjective Evaluation of Speech Codecs, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 19, No. 5, July 2011
[12] Suhadi Suhadi, et.al, A Data-Driven Approach to A Priori SNR Estimation, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 19, No. 1, January 2011
[13] Richard C. Hendriks, Richard Heusdens and Jesper Jensen, Forward-Backward Decision Directed Approach For Speech Enhancement, pg 109- 112
[14] Brady N. M. Laska, Miodrag Bolic´, and Rafik A. Goubran, Particle Filter Enhancement of Speech Spectral Amplitudes, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 18, No. 8, November 2010
[15] ITU-T P.832, Methods for objective and subjective assessment of quality (05/2000)
[16] Yao Ren, Michael T. Johnson, An Improved SNR Estimator For Speech Enhancement, ICASSP 2008, pg 4901- 4904
[17] Loizou, P., Kim, G., Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. IEEE Trans. Acoust. Speech Signal Process. 19, 47–56., 2011
[18] Ulpu Remes, et.al, Robust Automatic Speech Recognition Using Acoustic Model Adaptation Prior To Missing Feature Reconstruction, 17th EUSIPCO 2009 Glasgow, Scotland, August 24-28, 2009
[19] J. Ramírez, J. M. Górriz and J. C. Segura, Voice Activity Detection. Fundamentals and Speech Recognition System Robustness, I-Tech, Vienna, Austria, June 2007
[20] Loizou, P., Speech Enhancement: Theory and Practice. CRC Press LLC, Boca Raton, Florida.,2007
[21] Michael Cowling, and Renate Sitte, Analysis of Speech Recognition Techniques for use in a Non-Speech Sound Recognition System, 2004, pg 16-20
[22] Weizhong Zhu, et.al, Using Noise Reduction And Spectral Emphasis Techniques To Improve ASR Performance In Noisy Conditions IEEE Conf. Automatic Speech Recognition and Understanding, Dec-2003, 357-362,
[23] Lu, Y., Loizou, P.,Speech enhancement by combining statistical estimators of speech and noise. Proc. IEEE Internat. Conf. Acoust. Speech Signal Process., 4754–4757., 2010
[24] Huijun Ding, et.al, Over-Attenuated Components Regeneration for Speech Enhancement, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 18, No. 8, November 2010
[25] Boh Lim Sim, A parametric formulation of the generalized spectral subtraction method, Speech and Audio Processing, IEEE Transactions on , Volume: 6 Issue: 4 , pp: 328 – 337, July 1998
[26] M. Berouti, et.al l, “Enhancement of speech corrupted by acoustic noise,” Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 208-211, Apr. 1979.
[27] S.F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust., Speech, Signal Process., vol.27, pp. 113-120, Apr. 1979.
[28] S. D. Kamath and Philipos C. Loizou, A Multi-Band Spectral Subtraction Method For Enhancing Speech Corrupted By Colored Noise, 2002
[29] Breithaupt, C., Martin, R., Analysis of the decision-directed SNR estimator for speech enhancement with respect to low-SNR and transient conditions. IEEE Trans. Audio Speech Lang. Process. 19, 277–289., 2010
[30] Ephraim, Y., Malah, D., Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process. 33, 443–445.,1985
[31] Cohen, Modeling speech signals in the time-frequency domain using GARCH, Signal Processing 84(12) 2453–2459., 2004
[32] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square short-time spectral amplitude estimator," IEEE Trans. Acoustic. Speech Signal Process., vol. 32, pp. 1109-1121, 1984.
[33] R.J. McAulay, M.L. Malpass, Speech enhancement using a soft-decision noise suppression "lter, IEEE Trans. Acoust. Speech Signal Process. ASSP-28, April 1980.
[34] Jensen, J., Heusdens, R., Improved subspace-based single-channel speech enhancement using generalized super-Gaussian priors. IEEE Transactions on Audio, Speech, and Language Processing 15, 862–872., 2007
[35] Lotter, T., Vary, P., Speech enhancement by MAP spectral amplitude estimation using a super-gaussian speech model. EURASIP J. Appl. Signal Proc. 7, 1110–1126, 2005.
[36] S. Gannot, D. Burshtein and E. Weinstein, “Iterative and sequential Kalman-filter-based speech enhancement algorithms,” IEEE Transactions on Speech and Audio Processing, volume 6, no. 4, pp. 373-385, July 1998
[37] H. Sameti, H. Sheikhzadeh, Li Deng, and R. Brennan, HMM-based Strategies for Enhancement of Speech Signals Embedded in Nonstationary Noise, in IEEE Trans. on Speech and Audio Processing, vol. 6, no. 5, pp. 445-455, January 1998
[38] Y. Ephraim and H. L. Van Trees, A signal subspace approach for speech enhancement, IEEE Trans. Speech and Audio Processing, vol. 3, pp. 251-266, July 1995
[39] Y. Ephraim, “Statistical model based speech enhancement systems,” IEEE Proc., vol. 80, pp. 1526-1555, Oct. 1992.
[40] Y. Ephraim, “A Bayesian estimation approach for speech enhancement using hidden Markov models,” IEEE Trans. Signal Processing, vol. SP-40, pp. 725-735, April 1992.
[41] R McAulay, M Malpass , Speech enhancement using a soft-decision noise suppression filter in IEEE Transactions on Acoustics Speech and Signal Processing (1980)
[42] MathWorks - MATLAB and Simulink for Technical Computing, www.mathworks.com/
[43] www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html
[44] www.utdallas.edu/~loizou/speech/noizeu