Speech Enhancement using Spectral Subtraction

Dr. Shridhar.S.K
Electronics and communication
Basaveshwar Engineering college
Bagalkot, India.

Laxmi Doddimani
Electronics and communication
Basaveshwar Engineering college
Bagalkot, India.

Abdulkalam Hirekoppa
Electronics and communication
Basaveshwar Engineering college
Bagalkot, India.

Kenchappa Kodliwad
Electronics and communication
Basaveshwar Engineering college
Bagalkot, India.

Aishwarya Viraktamath
Electronics and communication
Basaveshwar Engineering college
Bagalkot, India

Abstract—**In real-time, the signal viewed from a distant point consists of information and noise which is due to the disturbance of the signal from other unwanted signals i.e. noise. The noise can be removed by using speech enhancement. The speech enhancement using spectral subtraction is a simple method to remove the noise. The objective of the project is to remove the background noise and to improve intelligibility.**
This paper aims to provide a comparison and simulation of the spectral subtraction-type algorithms viz. basic spectral subtraction, spectral subtraction using half-wave rectification, and power spectral subtraction. To test the performance of the algorithms, speech samples with different SNR levels are taken and tested. The NOISEX database is used in the algorithm. The proposed algorithm is evaluated using subjective and objective measures

Keywords: Speech Enhancement, spectral subtraction, SNR, Magnitude Spectrum, Phase Spectrum.

I. INTRODUCTION

Speech is one of the maximum essential methods of conversation among human to human and human to system in numerous fields through computerized speech reputation and speaker identification. The modern-day speech conversation structures are critically degraded because of numerous forms of noises which make the listening venture tough for an instantaneous listener and reason misguided switch of information [6]. It is a place where in progressed lot of innovation may be delivered in.
Speech signal needs to be noise free so that we applied basic spectral subtraction method. Its oldest approach for speech maybe negative so that check for non-negative magnitudes as per the half wave rectification. To improve this power spectral estimation is used. [4].
In speech processing using spectral subtraction method mainly we can listen the speech signal distortion and residual noise called musical noise in speech which is due to improper estimation of the noise spectrum [1]. Some part of noise will be reduced in half wave and power spectral subtraction based speech enhancement algorithm.

II. SPECTRAL SUBTRACTION

The incoming noisy speech signal is divide into frames of 160 samples. Each frame is Hamming windowed and then converted into the frequency domain using Fast Fourier Transform. Magnitude and Phase spectrum are calculated. It is an assumption that the noise is present in silence frames. So, the average of the first five frames is taken into consideration. This value is subtracted from all remaining frames of the signal. It is assumed that the signal is stationary

in each frame. After spectral subtraction, the magnitude spectrum is recombined with the phase of the noisy speech signal. We are unaware of the exact phrase of the noise signal so we subtract the magnitudes and leave the phase of the noisy signal unprocessed. It is converted again to the time domain. Each signal frame is then overlapped and added to the preceding and succeeding frames to form the final output.

III. METHOD

This approach operates within side the frequency area and makes the belief that the spectrum of the enter sign may be expressed because the sum of the speech spectrum (s[k]) and the noise spectrum (d[k]).

$$X[K]=S[K]+D[K] \qquad (1)$$

IV. OVERLAP-ADD PROCESSING

The overlap-upload approach is used to interrupt lengthy alerts into smaller segments for less difficult processing. FFT convolution makes use of the overlap-upload approach collectively with the Fast Fourier Transform, permitting alerts to be convolved via way of means of multiplying their frequency spectra. The overlap-upload approach is primarily based totally on the essential technique.
(1) Decompose the sign into easy additives,
(2) Manner every one of the additives in a few beneficial ways, and
(3) Recombine the processed additives into the very last sign.
The processing of the FFT sign ought to be finite so that non-stop time-area sign into overlapping chunks referred to as frames. To preserve the continuity of sign 50% of overlapped, FFT is carried out to every frame.

V. ESTIMATION OF NOISE SPECTRYUM

Noise Estimation cannot be calculated previously, it can be estimated at the beginning of speech speaker's lips are closed for around 2 seconds this can be used for estimating the noise. The average of the first five frames will be Noise estimation D[k]. We calculate the average noise estimation

VI. SUBTRACTING THE NOISE SPECTRUM

Phase information (P[k]) will not make the difference to ear, so keep the phase information as it is. In magnitude perform

noise estimation. Subtract the estimated noise from magnitude information

Not all the values are positive. Don't modify in basic spectral subtraction. While doing the half-wave will check for negative values and replace them with zeros. In the power spectral input power, magnitude is subtracted from the noise power estimation and check for negative values replace with zeros.

$$Y[k] = S[k] - D[k] \qquad (2)$$

Since we didn't calculate for phase information it can't be accurate

To transform the frequency domain to the time domain, the phase of the noisy signal is combined with the processed magnitude spectrum, and then the Inverse Fast Fourier transform is applied. By multiplying conjugates to speech signal Retain the real part because of conjugate-symmetric property and overlap-add processing to get the pure speech in the time domain.

VII. EVALUATION METHOD

In this chapter, we evaluated the performance of the proposed algorithm. The subjective and goal checks are calculated.
The proposed algorithm and graphs are implemented in MATLAB.

VIII. SUBJECTIVE LISTENING TEST

A test in which participants are required to give the rating to the given set of enhanced speech data of different noisy speech data for different SNR values. Rating is dependent on the auditory perception of participants. Rating is the opinion of an individual on the given enhanced speech data set.

Test Requirements:

Number of listeners:-

Listeners are also called "subjects". A certain minimum number of listeners should be invited to a listen since a large number of subjects increase the reliability of test outcomes.

Ratings	Meaning
1	Bad
2	Poor
3	Fair
4	Good
5	Excellent

Mean opinion score (MOS):-
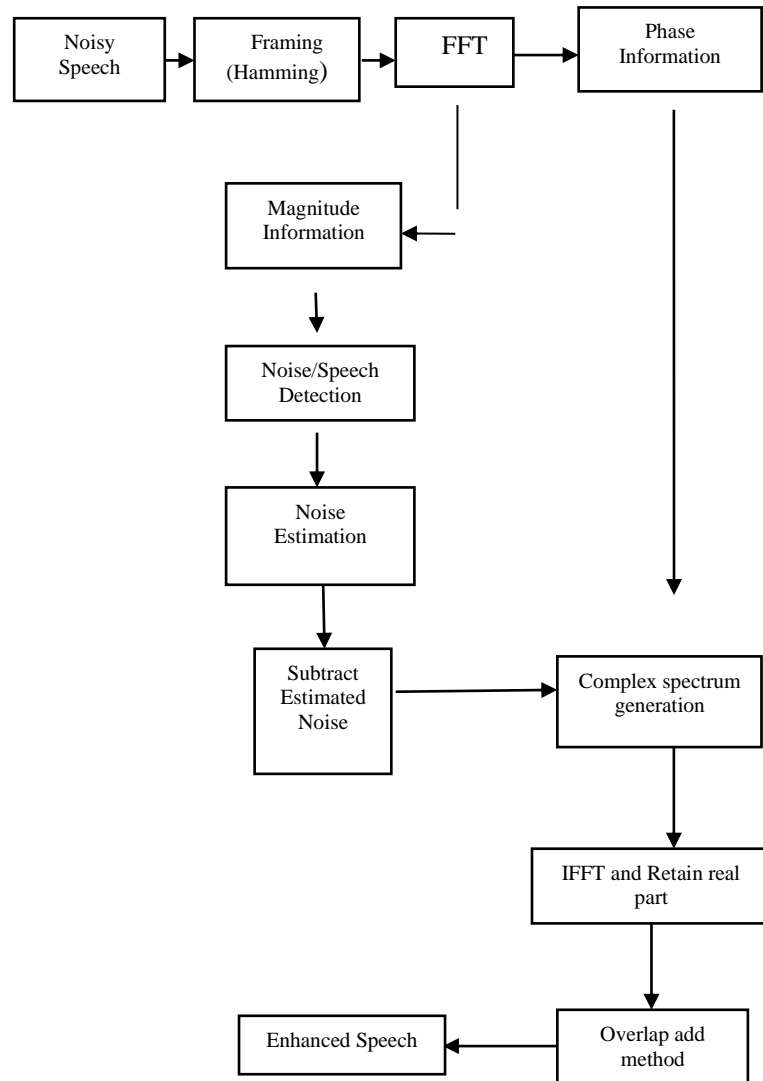It is a measure used in the domain of speech processing, representing the overall quality of a proposed algorithm

Fig .1   Block Diagram of Proposed Method

.

It is the arithmetic mean over all individuals "values on a predefined scale (ratings) that a subject assigns to his opinion of the performance of the proposed algorithm. Such scores are generally accumulated in a subjective exceptional assessment test. MOS is a commonly used measure for audio, and audiovisual quality evaluation. MOS depending on whether the score was obtained from audiovisual, conversational, listening, talking quality tests.

MOS specifies that "listener should be seated in a quiet room with room noise level must be below 30dbA with no dominant peaks in the spectrum.

Limitation of the subjective listening test:-
A careful subjective test can be tedious and time-consuming since the human perceptual domain is not entirely well understood until now.

IX.   OBJECTIVE QUALITY MEASURE

The fundamental method for evaluating speech quality is through the subjective listening test. Although subjective evaluation of speech enhancement algorithms is often

accurate and reliable performed under a well-suited environment but it is costly, time-consuming.

For that reason, much effort has been placed on developing objective measures that would predict the speech quality with high correlation.

Many goal speeches pleasant measures had been proposed with inside the beyond to expect the subjective pleasant of speech. Most of these measures, however, were developed to evaluate the distortion introduced by speech codes.

The most popular objective tests are as follows:-

1. Increment in segmental SNR
2. Log-Likelihood Ratio
3. Itakura Saito Distance

a.   Increment in Segmental SNR

It is also a time-domain objective measure. In any speech signal, energies are at non-stationary in nature which fluctuates randomly; as a result, an accurate SNR value for each segment of the frame is computed separately and combined to form segmental SNR. Thus the equation of segmental SNR is given in equation as:

$$SNRseg = \left(\frac{10}{I}\right) \sum_{i=0}^{I-1} \log_{10}\left(\frac{\sum_{n=N_i}^{N_i+N-1} X^2(n)}{\sum_{n=N_i}^{N_i+N-1}(X(n)-X^{\wedge}(n))2}\right)$$

Where

N= is the frame length, I= is the number of frames, x (n) = is the original noisy speech, x^ (n) = is the processed speech signal.

a.   Log Llikelihood Ratio (LLR):-

It is one of the LPC based measures of speech quality. The phase between the spectrum of clean speech and processed speech is calculated through LLR which gives the amount of distortion added during processing. The log-likelihood ratio test is calculated with respect to clean speech to enhanced speech. The equation is used to calculate LLR is shown in Equation as:

$$D_{LLR}(a_e, a_c) = \log_{10}\left(\frac{a_e\ R_c\ a_e^T}{a_c\ R_c\ a_c^T}\right)$$

Where

ae = LPC vector of the enhanced or processed speech signal.

ac = LPC vector of the clean speech signal.

Rc = is the autocorrelation matrix of the clean speech signal

Note:-LPC (Linear Predictive Coding):- The method used mostly in audio signal processing and speech processing for representing speech spectral envelope (envelop contains information of signal or speech). As less as the value we get, indicates the proposed method signal quality is best and lesser the speech distortions occur. For proper analysis, different SNRs at different noise Environments are performed. LLR value must be in the range of 1-2, indicates the proposed algorithm gives a good enhanced speech.

b.   Itakura Saito Distance (ISD):-

It is also one LPC-based measure of speech quality. The difference between enhanced speech signal and clean speech signal in terms of the corresponding spectral envelope is known as Itakura–Saito spectral distance.

Itakura Saito distance is also defined as the measure of the perceptual difference between the reference spectrum and test spectrum. The used to calculate the Itakura Saito Distance measure speech quality as:

$$D_{ISD}\ (a_e, a_c) = \frac{G_c}{G_e}\left(\frac{a_e\ R_c\ a_e^T}{a_c\ R_c\ a_c^T}\right) + \log_{10}\left(\frac{G_c}{G_e}\right) - 1$$

X.   EXPERMENTS RESULTS

We have used eight noise speeches as airport, babble, car, exhibition, restaurant, street, station, train noises at different SNR values from NOISEX database. Compared and simulated the spectral subtraction-type algorithms *viz.* basic spectral subtraction (SS), spectral subtraction using half-wave rectification (SS-H), and power spectral estimation (SS-P).
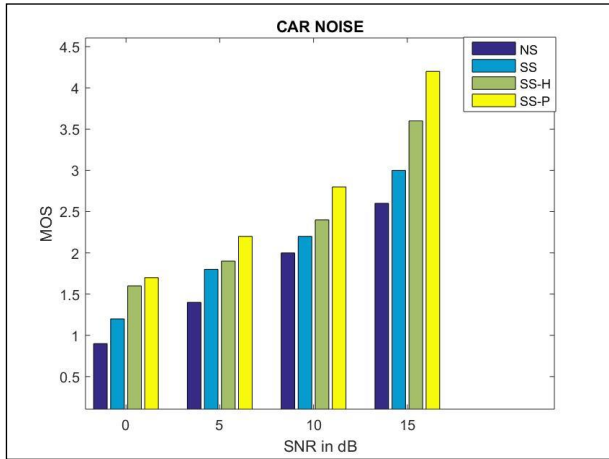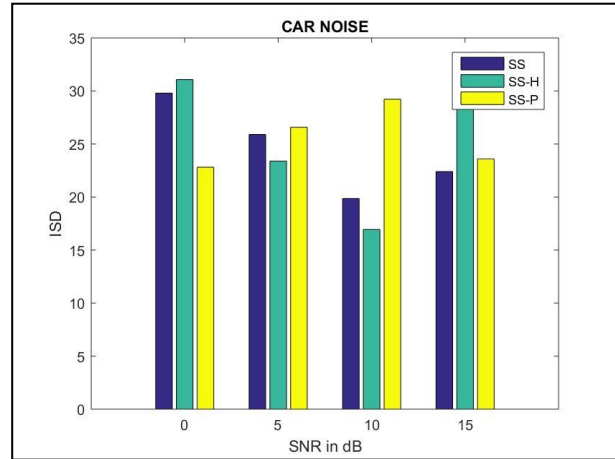
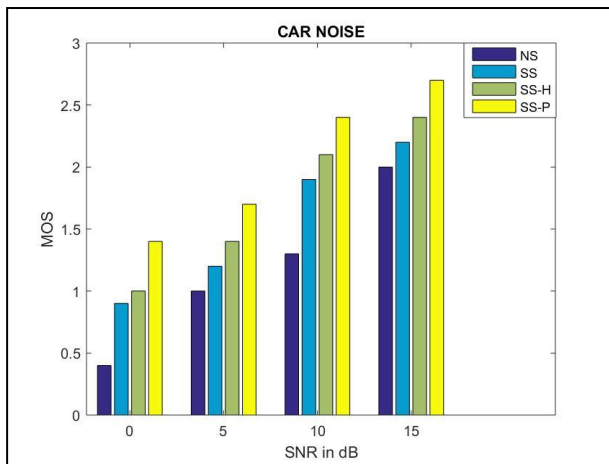Fig.2 Subjective Test for Normal Hearing

Fig.3 Subjective Test for Hearing Loss
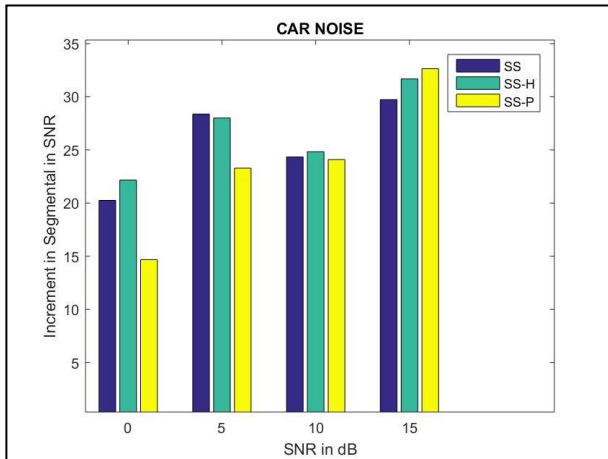

Fig.4 Increment in Segmental SNR

Fig.5 Itakura-saito distance for Car noise

X. CONCLUSION

In this paper, the speech improving approach primarily based totally at the spectral subtraction set of rules is introduced. It may be visible from the experimental outcomes that the proposed approach correctly reduces heritage noise in assessment with the generally used spectral subtraction kind set of rules. This approach makes the speech sign audible however together with heritage noise. Simulation results show that the set of regulations works successfully in reducing the historical past noise. The half-wave rectification approach has been exceptional for all environments of noise. The proposed algorithm exhibits best performance in the car, street, station noise speech and even better for the airport, babble, restaurant, and exhibition. This approach may be carried out in embedded structures associated with speech processing or communication-primarily based totally application.

ACKNOWLEDGEMENT:

REFERENCES

[1] Hamze Moazami Goodarzi and Saeed Seyedtabaii,"Speech Enhancement Using Spectral Subtraction Based on A Modified Noise Minimum Statistics Estimation" 5th International Joint Conference on INC, IMS and IDC.Aug. 2009,pp 25-27J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[2] S.Khadar Basha and Prem C. Pandey "Real-Time Enhancement of Eletrolaryangeal speech by spectral subtraction method" 18th National Conference on Communications, Kharagpur, 3-5 Feb-2012.

[3] Chunhe Yu, Long Su" Speech Enhancement Based on the Generalized Sidelobe Cancellation and Spectral Subtraction for Microphone Array"8th International Congress on image and signal processing,2015.

[4] Shambhu Shankar Bharti &others"A New Spectral Subtraction Method for Speech Enhancement using Adaptive Noise Estimation" 3rd Int'l Conf. on Recent Advances in Information Technology, 2016.

[5]].Chawdhury shahriar muzamulla" Noise Reduction Speech Signal Using Modified Spectral Sbutraction Technique"Eleventh international multi confrence on information processing,2015.

[6] Mariyadasu Mathe and Siva Prasad Nandyala" Speech Enhancement Using Kalman Filter for white, random and color noise"Int'l Conf. on devices, circuits and systems, march ,2012.

[7] Hilman Pardede ,Kalamullah Ramliand others "Speech Enhancement for Secure Communication Using Coupled Spectral Subtraction and Wiener Filter" , University of Indonesia, Jawa Barat -14-August,2019