International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 3, Issue 5, May 2016)

Noise Estimation in Single Channel Speech Enhancement Using FFT

Appannagouda
M.Tech (DCN)
Dept. of Electronics & Communication
PACE, Mangalore Karnataka India

Chandana B R
Assistant professor
Dept. of Electronics & Communication
PACE, Mangalore Karnataka India

Abdullah Gubbi
Professor & HOD
Dept. of Electronics & Communication
PACE, Mangalore Karnataka India
Hod_ece@pace.edu.in

***Abstract*—Conventional speech enhancement methods typically utilize the noisy phase spectrum for signal reconstruction. This letter presents a novel method to estimate the clean speech phase spectrum, given the noisy speech observation in single-channel speech enhancement. The proposed method relies on the phase decomposition of the instantaneous noisy phase spectrum followed by temporal smoothing in order to reduce the large variance of noisy phase, and consequently reconstructs an enhanced instantaneous phase spectrum for signal reconstruction. The effectiveness of the proposed method is evaluated in two ways: phase enhancement-only and by quantifying the additional improvement on top of the conventional amplitude enhancement scheme where noisy phase is often used in signal reconstruction. The instrumental metrics predict a consistent improvement in perceived speech quality and speech intelligibility when the noisy phase is enhanced using the proposed phase estimation method.**

***Index Terms*— Noise estimation, MATLAB, Signaling system, etc**

I. INTRODUCTION

Speech is exceptionally fundamental route for people to pass on data. The principle target of Speech is correspondence. Speech can be characterized as the reaction of vocal track to one or more excitation signal. Colossal measure of information transmission is extremely troublesome both regarding transmission and capacity. Speech Compression is a strategy to change over human speech into an encoded structure in a manner that it can later be decoded to get back the first flag. Pressure is essentially to expel repetition between close examples and between nearby cycles. Real goal of speech pressure is to speak to motion with lesser number of bits. The diminishment of information ought to be done in a manner that there is satisfactory loss of value.

In numerous speech applications a speech enhancement pre-processor is required to expand the heartiness of the general framework against foundation clamor. To this end, past strategies for the most part concentrate on determining estimators of the spotless speech phantom abundancy given the uproarious speech while the loud stage has been regularly specifically utilized for reproduction of the improved signal. The lower branch in Fig. 1 demonstrates the piece graph for the routine speech upgrade made out of an abundance adjustment stage took after by a blend stage where the boisterous stage range is commonly utilized unaltered to remake the improved signal. A wide range of commotion concealment rules have been proposed to channel the loud ghostly abundance. The concealment standards are elements of from the earlier and a posteriori SNRs evaluated from phantom adequacy and commotion power ghastly thickness. These techniques are either information driven where preparing information is misused as earlier learning (environment, or client advanced), or depend on a more broad earlier learning identified with likelihood thickness capacities. In both gatherings the uproarious stage has been commonly used in signal remaking.

II. PROBLEM STATEMENT

The upgrade of the speech signal is comparing with the physiological and behavioral qualities of the speaker. The individual speech may talk quick, moderate, louder, and smooth with differing speed the real reasons that expand the trouble of verification of voice/speech is the variability of the voice properties of the speaker as one voice can't be straightly contrasted and other. In our framework it is made to remember this kind of variability's and adjust to these variability's. Furthermore, we are upgrading the speech signal progressively environment by considering the natural clamor and varieties.

Contribution of the project

Continuously speech upgrade framework, we utilized diverse sorts of speech enhancement methods for the enhancing the speech signals with diminished commotion reason. Keeping in mind the end goal to do as such the voice database of various speech is taken into the thought than we have separated the voice elements of those speakers.

The encircling calculations are utilized to independent speech signal into equivalent length and windowing method is to gather all the isolated or confined flag and applying the hamming window strategy to every edge for the further results. Once the windowing strategy got finished we have figured the productivity of every calculation and actualized the best technique for enhancement of speech signal. In the wake of expelling the clamor, the speech signal will be opened up for some moment of time as for the season of speech signal showed up.

IJEEE

International Journal of Ethics in Engineering & Management Education

Website: www.ijeee.in (ISSN: 2348-4748, Volume 3, Issue 5, May 2016)

III. LITERATURE SURVEY

Authors: Sriram Srinivasan, Member, IEEE, Jonas Samuelsson, and W. Bastiaan Kleijn, Fellow, IEEE, in this paper, we watch that a Bayesian least mean squared blunder approach for the joint estimation of the transient indicator parameters of speech and commotion, from the loud perception. We utilize prepared codebooks of speech and commotion direct prescient coefficients to show the form of earlier data required by the Bayesian plan. As opposed to current Bayesian estimation approaches that consider the excitation changes as a feature of from the earlier data, in the proposed strategy they are figured online for every brief timeframe portion, in view of the current perception. Thusly, the technique performs well in nonstationary clamor conditions. The subsequent appraisals of the speech and clamor spectra can be utilized as a part of a Wiener channel or any best in class speech upgrade framework. We create both memory less (utilizing data from the present casing alone) and memory-based (utilizing data from the present and past edges) estimators. Estimation of elements of the transient indicator parameters is likewise tended to, specifically one that prompts the base mean squared mistake assessment of the perfect speech signal. Tests show that the plan proposed in this paper performs essentially superior to anything contending strategies.

The MMSE appraisal of clean speech given the loud speech is acquired as a weighted total of MMSE estimators relating to every condition of the HMM for the spotless signal. Be that as it may, the HMM-based frameworks regard the excitation change as a feature of the earlier data.

The MMSE gauge additionally regards the excitation difference as a feature from the earlier data. To represent the subsequent confound in the level of the increase of the spotless speech model amid preparing and testing, the HMM strategies more often than exclude pick up adjustment. Correspondingly, there is addition adjustment in the commotion demonstrate as well. For the speech model and models relating to stationary commotion, a general addition change in time is adequate. However to successfully manage nonstationary commotion, the increase modification should be performed either on a casing by-edge premise or at a rate not slower than the rate at which the clamor insights change.

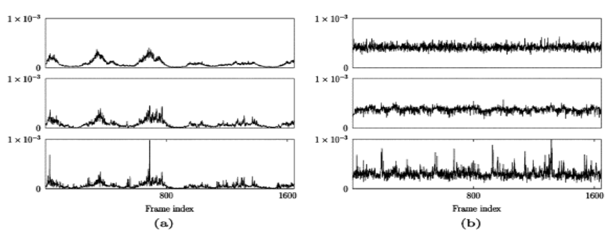

Figure 1: Shows plot of the true and estimated noise excitation variances with and without memory. (a) Highway noise. (b) White noise.

In each figure, the top plot corresponds to the true values of the excitation variances, the middle plot to memory-based estimates and the bottom plot to memory less estimates.

Both types of increase adjustment rely on an assessment of the commotion measurements, got from the perception. Thus, the execution of these strategies is restricted by the execution of the hidden commotion estimation calculations in nonstationary situations.

Authors: Nasser Mohammadiha*, Student Member, IEEE, Paris Smaragdis, Member, IEEE, Arne Leijon, Member, IEEE, Diminishing the impedance commotion in a monaural boisterous speech signal has been a testing undertaking for a long time. Contrasted with customary unsupervised speech upgrade techniques, e.g., Wiener separating managed methodologies, for example, calculations in view of Hidden Markov Models (HMM), lead to higher-quality improved speech signals. In any case, the primary down to earth trouble of these methodologies is that for every commotion sort a model is required to be prepared from the earlier. In this paper, we research another class of directed speech de noising calculations utilizing nonnegative matrix factorization (NMF). We propose a novel speech upgrade technique that depends on a Bayesian detailing of NMF (BNMF). To bypass the jumble issue between the preparation and testing stages, we propose two arrangements. To start with, we utilize a HMM in mix with (BNMF-HMM) to determine a minimum mean square error (MMSE) estimator for the speech signal with no data about the fundamental clamor sort. Second, we recommend a plan to take in the required clamor BNMF model on the web, which is then used to build up an unsupervised speech enhancement framework.

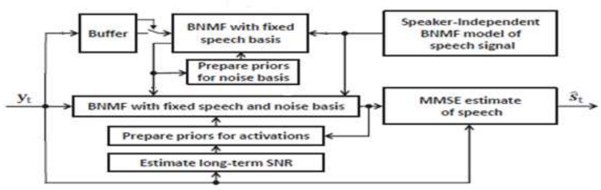

Figure 2: Block diagram representation of BNMF.

Yt and ˆst are the brief span ghostly amplitudes of the loud and upgraded speech flags responsively, at time allotment t. Objective of the "Get ready priors" boxes is to recursively reduce the earlier disseminations, which will be likewise talked about in III C.

Broad examinations are completed to explore the execution of the proposed strategies under various conditions. Additionally, we look at the execution of the created calculations with best in class speech enhancement plans utilizing different target measures. Our reproductions demonstrate that the proposed BNMF-based strategies beat the contending calculations significantly.

Authors: Jan S. Erkelens, Richard C. Hendriks, Richard Heusdens, and Jesper Jensen, This paper considers methods for single-channel speech upgrade taking into account the discrete Fourier Transform (DFT). In particular, we determine minimum mean square error (MMSE) estimators of speech DFT coefficient sizes and additionally of complex-esteemed DFT coefficients in light of two classes of summed up gamma dispersions, under an added substance Gaussian commotion presumption. The subsequent summed up DFT greatness estimator has as an uncommon case the current plan in view of a Rayleigh speech earlier, while the complex DFT estimators sum up existing plans in light of Gaussian, Laplacian, and Gamma speech priors. Broad reenactment explores different avenues regarding speech signals corrupted by different added substance clamor sources confirm that noteworthy enhancements are conceivable with the later estimators in view of super-Gaussian priors. The expansion in perceptual assessment of speech quality (PESQ) over the uproarious signals is around 0.5 focuses for road commotion and around 1 point for repetitive sound, free of information signal to-noise ratio (SNR). The suppositions made for determining the complex DFT estimators are less precise than those for the greatness estimators, prompting a higher most extreme achievable speech quality with the size estimators.

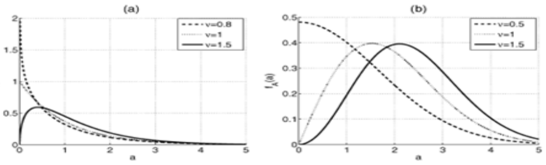

Figure 3: Magnitude estimation at priory state

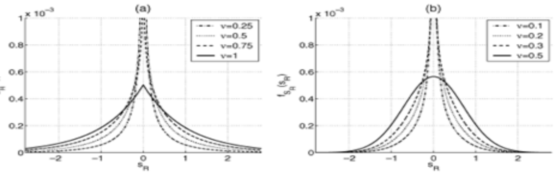

Figure 4: Magnitude estimation at posterior state

This paper considered DFT-based method for single channel speech enhancement. In the initial segment, we expanded existing MMSE estimators of the greatness estimators for DFT-based clamor concealment. The ideal estimators are found under an uneven summed up Gamma appropriation, which takes as extraordinary cases (diverse parameter settings) all priors utilized as a part of known clamor concealment plots in this way. Determining the MMSE estimators includes combination of (weighted) Bessel capacities. Keeping in mind the end goal to discover diagnostic arrangements, for some parameter settings approximations was essential. Eventually, we consolidated two sorts of Bessel capacity approximations utilizing a basic twofold choice between the two. We appeared by PC reproductions that the estimator in this manner acquired is near the definite MMSE estimator for all SNR conditions. The displayed estimators lead to enhanced execution contrasted with the concealment guideline proposed by Ephraim and Malah. Moreover, the most extreme conceivable execution is marginally superior to that of best in class MAP adequacy estimators. The second part of the paper managed MMSE estimators of complex DFT coefficients by determining two classes of estimators taking into account summed up gamma earlier pdfs.

Authors: Timo Gerkmann, Martin Krawczyk, and Robert Rehr. Speech Signal Processing, Faculty V, University of Oldenburg, 26111 Oldenburg, Germany, In this paper we audit the inspiration for ignoring stage estimation previously, and why late distributions infer that the estimation of the spotless speech stage might be helpful all things considered. Further, we introduce a calculation for indiscriminately evaluating the spotless speech phantom stage from the loud perception and demonstrate that the use of this stage gauge enhances the anticipated speech quality. Single channel speech upgrade portrays the change of a ruined speech signal caught with one amplifier in an uproarious domain, or at the yield of a multichannel speech enhancement calculation.

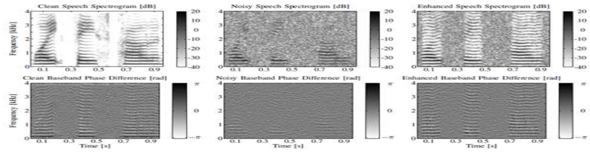
Figure 5: Amplitude spectra of clean (left), noisy (middle) and enhanced (right) speech signals are presented in upper row, together with the corresponding base band phase difference from segment to segment in the lower row.

Single channel speech enhancement is especially troublesome when the commotion is nonstationary, (for example, activity clamor), or even speech like (as prattle clamor). As portable speech specialized gadgets are regularly utilized in situations with nonstationary clamor, late research concentrates on making the calculations more vigorous in these commotion conditions. Speech upgrade calculations as a rule include a change of the boisterous speech into a phantom space to take into consideration a simpler partition amongst speech and clamor. A run of the mill and productive applicant is the brief span Fourier change (STFT) area. There, speech is sectioned into short portions of around 10-30ms, weighted with a decreased spectral investigation window and changed to the Fourier area.

IV. Proposed Method

Past strategies to diminish the clamor are wiener separating, unearthly subtraction, MMSE ghostly plentifulness estimation, and so forth. Stage decay should be possible by, to be specific discrete wavelet transform (DWT), discrete wavelet parcel transform (DWPT), and short time Fourier transform (STFT). Cutting edge wavelet de noising strategies have been effectively connected to picture clamor lessening. Notwithstanding it has not yet been broadly used to explain the speech signal clamor decrease issue, as couple of distributions in wavelet in contrast with colossal STFT papers. Since both Fourier and wavelet changes are straight and clamors are added substance, the FFT arrangements ought to relevant to the wavelet area.

The inspiration to utilize wavelet as a conceivable option for speech commotion lessening is to investigate better approaches to diminish computational multifaceted nature and to accomplish better clamor decrease execution. Firstly, in light of the fact that the wavelet change may not require covered windows because of the limitation property, the same channel could prepare less information. Furthermore, wavelet channel does not compare to time space convolution, so that move invariant is not protected. Be that as it may, the Fourier area channels can at present be stretched out to the wavelet

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 3, Issue 5, May 2016)

space, since they are determined by measurable properties of unearthly parts. The Martin least insights commotion estimator, the Wiener, the unearthly subtraction, the Wolfe-Godsill, and the Ephraim-Malah channels can be stretched out in the wavelet space too. These channels are like the cutting edge delicate, hard, or contracting edge strategies for wavelet de noising that both work on ghastly size and hold the indication of wavelet change coefficients (which identical to the stage in FFT). Thirdly, there are a wide range of wavelets and different wavelet change mixes.

The task of spectral subtraction is to furnish the vocoder analyzer with a cushion of commotion stifled speech in a period interim which is not exactly the support length time as well as sufficiently short to permit the analyzer to register and transmit the vocoder channel parameters. This interfacing limitation imposes certain conditions on the execution.

The calculation ought to utilize the same support size as the analyzer. Expecting a solitary processor it must register the commotion stifled speech in the time left over after the analyzer computations. It must supply the handled speech with least defer. Notwithstanding the fundamental clamor concealment systems, it must screen the signal to commotion environment and upgrade the normal clamor predisposition range if essential.

Data Segmentation

Cradle lengths of speech pressure analyzers come in all sizes. This methodology, be that as it may, prompts two operational bargains. In the first place, if the support is not a force of two then zeros must be attached before changing. Second, if cradle lengths are to be coordinated, with least postpone, then no covering (and along these lines no windowing) is permitted. The impact of cushioning with zeros essentially implies lower productivity (less focuses handled per FFT). It has a beneficial outcome of diminishing the measure of fleeting associating because of ghastly adjustment. No cover of time windows copies the preparing speed. The conceivable unfavorable impact of having no time window comprises of impelling discontinuities at the support limits.

Reconstituted waveforms from progressive examination cradles won't as a matter of course concur at the limit. Truth is told, in listening to the handled Speech entering the vocoder, a low level yet particular Clicking sound can at times be listened.

Frequency Analysis

The DFT of every information window is taken and the size is processed. Since genuine information is being changed, two information windows can be changed utilizing one FFT. The FFT size is set equivalent to the window size of 256. Growth with zeros was not joined. As effectively noted by Allen phantom change took after by reverse changing can mutilate the time waveform because of transient associating created by round convolution with the time reaction of the adjustment. Enlarging the info time waveform with zeros before spectral adjustment will minimize this associating. Tries different things with and without increase utilizing the speech came about as a part of unimportant contrasts, and along these lines enlargement was not fused. At last, since genuine information is dissected, change symmetries were exploited to lessen stockpiling necessities basically into equal parts.

Magnitude averaging

As was portrayed in the past area, the difference of the commotion ghostly gauge is lessened by averaging over whatever number spectral greatness sets as could be allowed. In any case the non stationary as far as possible the aggregate time interim accessible for neighborhood averaging. The quantity of midpoints is constrained by the quantity of investigation windows which are fit into the stationary Speech time interim. The decision of window length and averaging interim must tradeoff between clashing prerequisites. For satisfactory unearthly determination a window length more prominent than double the normal biggest pitch time frame is required with a 256-point window being utilized. For least clamor difference countless are required for averaging. At long last, for adequate time determination a slender examination interim is required. A sensible tradeoff between fluctuation diminishment and time determination gives off an impression of being three midpoints. These outcomes in a compelling examination time interim of 20 ms.

Noise Estimation

The spectral subtraction strategy requires an evaluation at every recurrence canister of the normal estimation of commotion size range PN =E {INI}. This assessment is gotten by averaging the signal size range 1x1 amid non speech action. Evaluating pN in this way puts certain requirements while actualizing the strategy. On the off chance that the commotion stays stationary amid the ensuing speech action, then an underlying startup or alignment time of clamor just flag is required. Amid this period (on the request of 33% of a second) an appraisal of pN can be registered. In the event that the clamor environment is non stationary, then another evaluation of pN must be computed before predisposition expulsion every time the commotion ranges changes. Since the evaluation is figured utilizing the commotion just flag amid non speech action, a voice switch is required. At the point when the voice switch is off, a normal clamor range can be recomputed. On the off chance that the clamor size range is changing quicker than an evaluation of it can be processed, then time averaging to gauge pN can't be utilized. In like manner, if the normal estimation of the commotion range changes after an evaluation of it has been figured, then clamor decrease through inclination evacuation will be less compelling or even.

Noise suppression during non speech activity

The last change in commotion lessening is signal concealment amid non speech movement. As was examined, parity must be kept up between the extent and qualities of the commotion that is seen amid speech movement and the clamor that is seen amid speech nonappearance. A viable speech movement identifier was characterized utilizing spectra produced by the unearthly subtraction calculation. This indicator required the determination of an edge flagging nonappearance of speech action. This edge was exactly resolved to guarantee that exclusive flags certainly comprising of foundation clamor would be lessened.

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 3, Issue 5, May 2016)

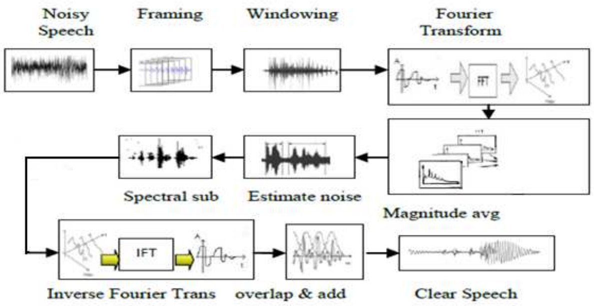

Figure 6: Block diagram of proposed method using Fourier Transform

Commotion diminishment channels are defined for both Fourier and wavelet spaces. Tests utilizing genuine commotion speech information have demonstrated that the Fourier change, the wavelet bundle change, and the wavelet change are the greatest, second, and last individually when all is said in done SNR sense. The wavelet bundle change can accomplish less twisting and is better for high SNR signals.

FFT Flow Chart

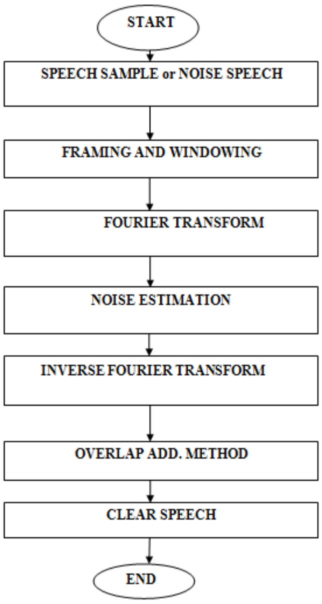

Figure 7: FFT flow chart

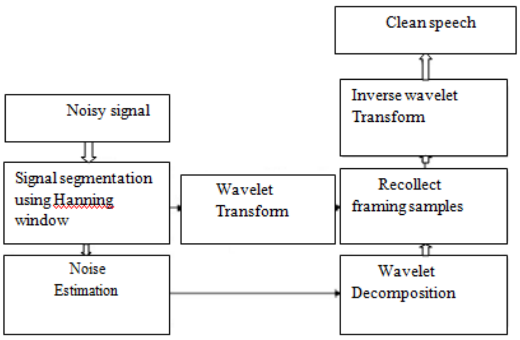

Figure 8: Block diagram of wavelet decomposition

By contrasting and the DWT, it breaks down the signal at various recurrence groups with various resolutions by disintegrating the signal into a coarse estimation and subtle element data. DWT utilizes two arrangements of capacities, called scaling capacities and wavelet capacities, which are connected with low pass and high pass channels, separately. The deterioration of the signal into various recurrence groups is essentially acquired by progressive high pass and low relax area signal. The first flag x[n] is initially gone through a half band high pass channel g[n] and a low pass channel h[n]. After the separating, half of the examples can be dispensed with as indicated by the Nyquist's standard, since the signal now has a most noteworthy recurrence of p/2 radians rather than p. The signal can accordingly be sub examined by 2, essentially by disposing of each other example.

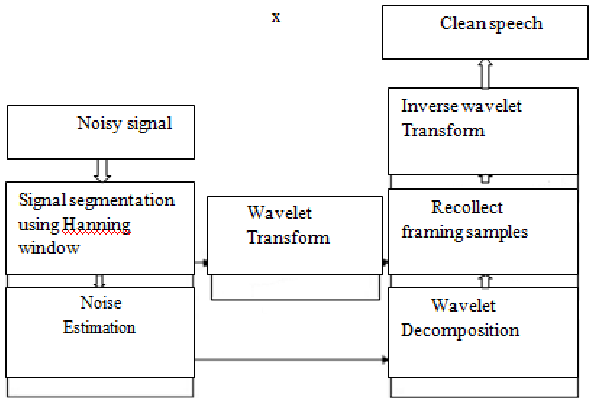

Figure 8: Block diagram of wavelet decomposition

DWT Flow Chart

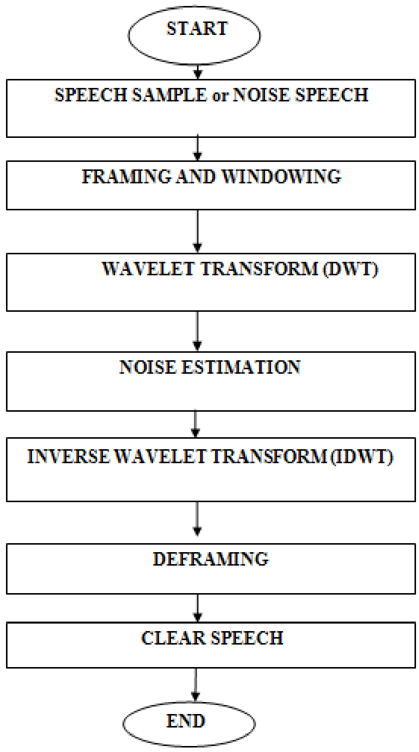

Figure 9: DWT Flow Chart

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 3, Issue 5, May 2016)

Implementation in MATLAB

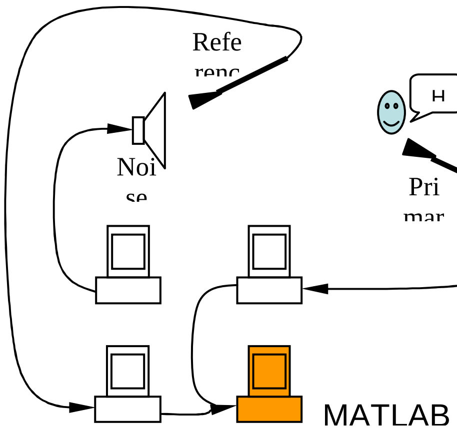

Figure 10: Implementation in MATLAB

Demonstrates the trial setup we utilized for the clamor cancelation executed in MATLAB. For every PC has stand out arrangement of sound info and yield, we need to utilize a few PCs in this usage to drive diverse gadgets. The clamor source is a speaker which consistently plays the background noise by the tone generator in the PC. The mouthpiece just adjacent to the speaker grabs the commotion as clamor reference information. The other amplifier remains around 0.5 meter away which grabs a man's speech signal in the uproarious foundation as the essential information. These two inputs are joined into a stereo link and sent to a PC. The MATLAB program in this PC records a specific time of essential information and reference include at the same time, then applies the versatile clamor cancelation calculation on them to make the de noised yield. We can think about the yield and the essential info recorded in MATLAB after the project is over. The accompanying picture demonstrates the essential information, the reference commotion information and the separated yield. We can obviously see the lessening of the commotion segment in the yield contrasted with that in the essential information.

V. RESULTS

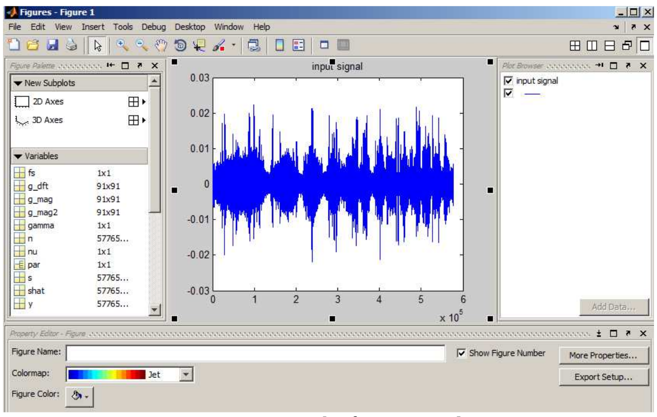

Figure 11: Graph of input signal.

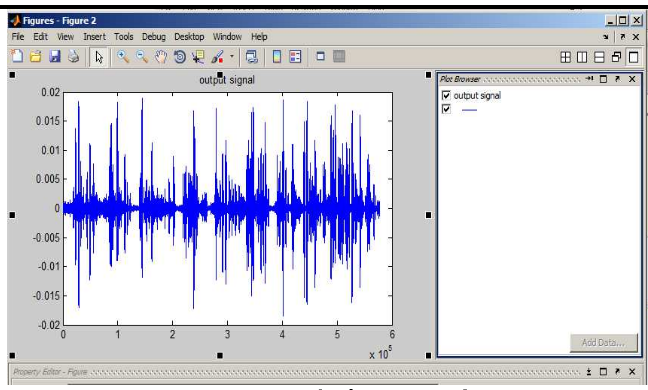

Figure 12: Graph of output signal

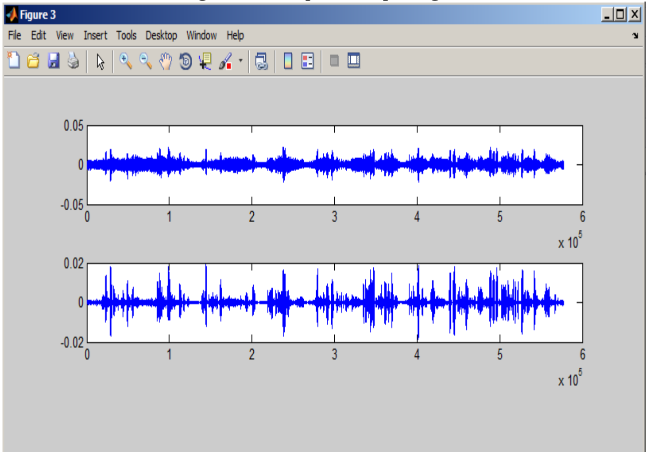

Figure 13: Sub Plot of Input and Output signal

Spectrogram Results

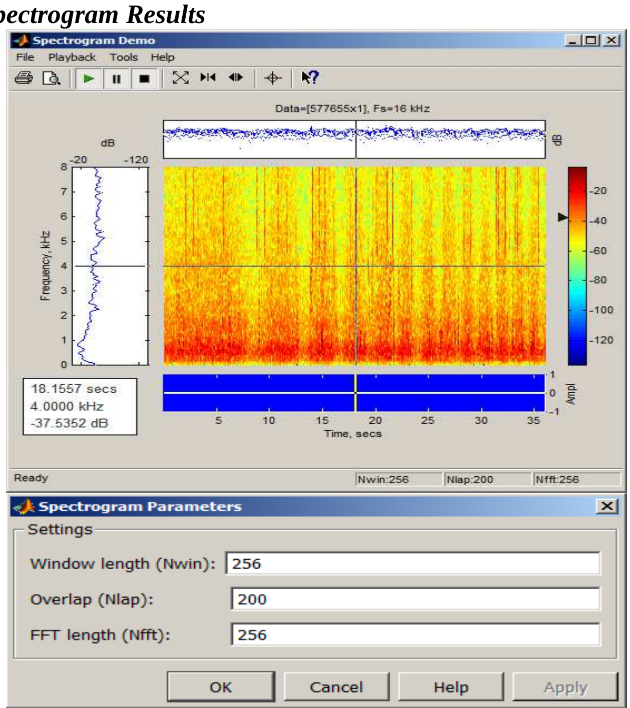

Figure 14: Spectrogram result of input signal

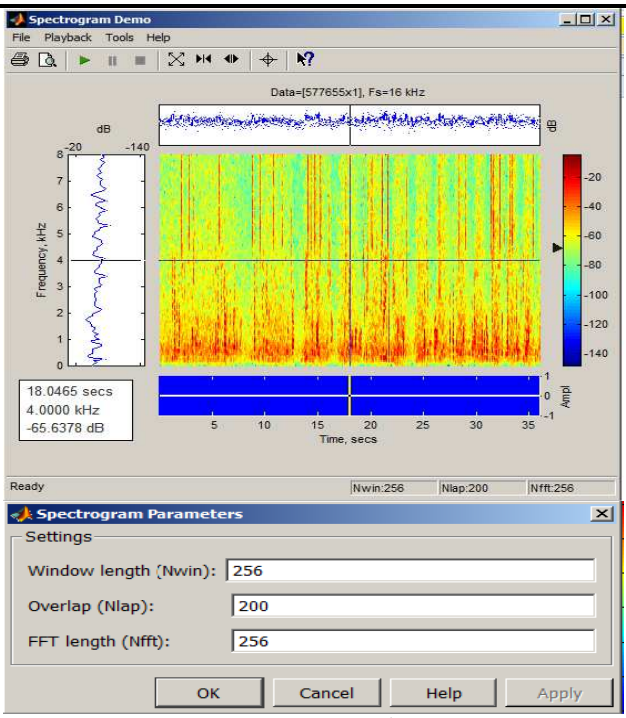

Figure 15: Spectrogram result of output signal

VI. APPLICATIONS

1. Hearing aids.
2. Speech enhancement for mobile telephony.
   - Modification of original signal.
   - Voice activity detection.
   - Continues speech enhancement.
3. Robust speech recognition.
   Influence of speech enhancement on recognition results.
4. Clean speech.
   - These signals were recorded in stopped cars without the effects of speakers stress. These signals were divided into three groups:
   - Isolated words with/without pauses; small vocabulary with less than 20 words.
   - Isolated sentences.
   - Continuous speech.
5. Car noise under different driving conditions.
   - Stationary: Driving on highways.
   - Non stationary: Slow changes; slow acceleration.
   - Non stationary in quick changes: Shifting gears, direction indicators, and fast accelerations.
6. Speech picked during driving in noisy environment

Advantages & Disadvantages

*1) **Advantages.***

1. Flexible to change parameters.
2. Simplicity in implementation.
3. Stable and robust performance against different signal conditions.
4. It takes less computational time.

*2) **Disadvantages.***

1) Background noise.
   a. Amplification of background noise is undesirable.
2) Echo.
   a. In a reverberant place, late multipath arrivals become the interference of direct first arrival signal.

VII. CONCLUSION

The outcome acquired by utilizing the Fourier Transform and stage estimation by the ghastly subtraction is much considerable for speech upgrade. In any case it is impractical to expel the whole clamor content present in the speech range because of the murmuring impact and shaded commotion present in the default signal.

As the quantity of centroid expands the execution of the framework increments however it debases the computational effectiveness, next stride is FFT it is extremely straightforward and productive for calculation, rather than information test MMSE and different systems, test expression have distorted concerning the reference signal after that minimum commotion estimation was taken for speech quality support the precision got in the wake of utilizing this strategy is around half since it has not taken the vocal tract data of a specific speaker it has attempted just to adjust the 2 vectors proficiently in time area.

At long last we reason that for speech upgrade progressively we have attempted parcel of calculations and strategies and actualized the best at last we got the precision around 75% yet in uproarious environment the exactness may get diminished because of clamor in highlight, we can enhance the precision by utilizing distinctive strategy and calculations which will decrease the ecological commotion and enhance the productivity.

REFERENCES

[1] R. C. Hendriks, T. Gerkmann, and J. Jensen, “DFT-Domain Based Single Microphone Noise Reduction for Speech Enhancement,” in *Synthesis Lectures on Speech and Audio Processing*. San Rafael, CA, USA: Morgan & Claypool, 2013.

[2] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, “Codebook-based Bayesian speech enhancement for nonstationary environments,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 2, pp. 441–452, Feb. 2007.

[3] T. Rosenkranz and H. Puder, “Integrating recursive minimum tracking and codebook-based noise estimation for improved reduction of non-stationary noise,” *Signal Process.*, vol. 92, no. 3, pp. 767–779, 2012.

[4] N. Mohammadiha, P. Smaragdis, and A. Leijon, “Supervised and un-supervised speech enhancement using nonnegative matrix factoriza-tion,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 10, pp. 2140–2151, Oct. 2013.

[5] C. Breithaupt, T. Gerkmann, and R. Martin, “Cepstral smoothing of spectral filter gains for speech enhancement without musical noise,” *IEEE Signal Process. Lett.* vol. 14, no. 12, pp. 1036–1039, Dec. 2007.

[6] C. Breithaupt, M. Krawczyk, and R. Martin, “Parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech,” in *Proc. ICASSP*, Mar. 2008, pp. 4037–4040.

[7] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, “Minimum mean-square error estimation of discrete Fourier coefficients with generalized Gamma priors,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 6, pp. 1741–1752, Aug. 2007.

[8] P. Mowlaee, R. Saeidi, and Y. Stylianou, "Phase importance in speech processing applications," in *Proc. 15th Int. Conf. Spoken Language Processing*, 2014, pp. 1623–1627.

[9] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," in *Proc. IEEE*, May 1981, vol. 69, no. 5, pp. 529–541.

[10] D. Wang and J. Lim, "The unimportance of phase in speech enhancement," *IEEE Trans.* Acoust., *Speech, Signal Process.*, vol. 30, no. 4, pp. 679–681, 1982.