Improvement in Speech Quality by Modified Codebook Based Technique

Lalita Devi
M.Tech Scholar
Computer Science & Engg. Department
R. P. Inderaprastha Institute of Technology, Kurukshetra

Er. Parmeet Kaur
Asst. Professor
Computer Science &Engg. Department
R. P. Inderaprastha Institute of Technology, Kurukshetra

Abstract- **In this work, it presents an improved universal codebook-based speech enhancement framework that relies on a single codebook to encode both speech and noise components. The aim of the noise reduction algorithms is to estimate the clean speech signal from the noisy recordings in order to improve the quality and intelligibility of the enhanced signal. Given the good performance of deep learning in signal representation, a deep auto encoder will be employed in this work for accurately modelling the clean speech spectrum. Due to this, it proposes a method for speech enhancement using codebook based method iteratively. In this work, it provides speech enhancement under different noisy conditions. The speech power spectrum will be measured greatly for different types of speech sound. It shows the performance comparison of proposed method with actual methods and proposed method shows better improvement in ΔILD, ΔIPD values.**

***Keywords*-Speech Processing, Speech Enhancement, etc.**

I. INTRODUCTION

The use of systems involving speech-based communication technology is now ubiquitous; such systems include mobile phones, hearing aids and video-conferencing technology. The perceived quality, and in more severe cases the intelligibility, of the speech signal in these systems is reduced when they are used under the adverse noise conditions encountered in real environments such as offices, crowded public spaces, or railway stations [1].

A speech signal consists of three classes of sounds. They are voiced, fricative and plosive sounds. Voiced sounds are caused by excitation of the vocal tract with quasi-periodic pulses of airflow. Fricative sounds are formed by constricting the vocal tract and passing air through it, causing turbulence those results in a noise-like sound. Plosive sounds are created by closing up the vocal tract, building up air behind it then suddenly releasing it. A speech signal can be considered as a linear composite of the above three classes of sound, each of these sounds are stationary and remain fairly constant over intervals of the order of 30 to 40 ms [2].

Vowels are made by a voiced sound with no choking in the vocal tract. Estimation of a spotless discourse signal from an uproarious chronicle is a run of the mill signal estimation task. Be that as it may, due to the non-stationary of the discourse and a large portion of the functional commotion signals, and furthermore because of the significance of the issue, noteworthy measure of research has been committed to this difficult undertaking. Single-channel discourse improvement calculations for example utilize the worldly and ghostly data of discourse and commotion to plan the estimator. For this situation, just the loud recording got from a solitary amplifier is given while the clamour type, speaker character or speaker sexual orientation is typically not known.Figure2: Speech Enhancement System with Corrupted Noise [2]

Multichannel or multi amplifier commotion decrease frameworks, use the transient and phantom data just as the spatial data to appraise an ideal discourse signal from the given boisterous chronicles Consonants, be that as it may, can be begun by a voiced or an unvoiced sound and are delegated:

•Stops: which happen when the wind stream is blocked and all of a sudden discharged.
•Nasals: created when the air is halted in the oral pit yet not through the nasal cavity.
•Approximants: created when there is a tightening yet not limited enough to bring about choppiness.
•Fricatives: a thin narrowing in the vocal tract bringing about a tempestuous wind stream.

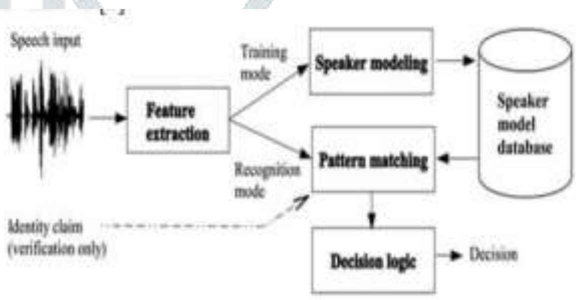

Fig 1: Process of Speech Recognition [1]

The principle commotion estimation approaches utilized Voice Activity Detector (VAD) estimators to distinguish clamour just interims. The commotion could be then determined by a fleeting normal during the discourse nonappearances utilizing an averaging time-steady that relies upon the expected stationary of the clamour. The premise of this methodology is that over a given time-interim there will be delays in the discourse in each recurrence band and thus the base estimation of the uproarious discourse range inside a recurrence band will relate to the commotion control. The clamour control range can likewise be determined by utilizing a Minimum Mean Squared Error (MMSE) estimator. A MMSE estimator was utilized to limit the intensity of the distinction work between the assessed and the genuine clamour control range.

1. Characteristics and Estimation of Noise Signals

Noise, in contrast to speech, can originate from any kind of source and have any spectral and temporal characteristics. There are, however, some common assumptions made about the noise when approaching the speech enhancement problem [2]:

- The power spectrum of noise is more stationary than that of speech, and
- Speech and noise are statistically independent.

Many speech enhancement techniques require an estimation of the noise power spectrum, or, equivalently, the SNR at each time-frequency bin. The accuracy of the noise estimation technique has a major impact on both the quality and intelligibility performance of the processed speech.

The paper is ordered as follows. In section II, it represents description of speech recording chain system. In Section III, It defines proposed work related to speech enhancement. Section IV describes the results of system. Finally, conclusion is explained in Section V.

II. SPEECH RECORDING CHAIN

To represent the section of a discourse signal from talker to audience, a common single channel discourse recording chain is appeared. The ideal discourse signal goes through a convolutive acoustic channel before arriving at the receiver, where it is joined with sound from other acoustic sources in nature and it is transduced into the electronic area. The discourse sign can get debased by further added substance clamor just as by conceivable non-direct mutilation inside the electronic area. It is helpful to group discourse signal debasements into the accompanying three classes which contrast in their causes and potential cures:

Additive foundation commotion that can emerge in either the electronic or acoustic spaces, albeit genuine sign corruption is regularly caused distinctly by acoustic clamor from undesirable sources in the earth;

- Convolutive impacts including reverberation and resonation; and
- Non-direct discourse contortion which may, for instance, be presented by adequacy restricting or cutting in the mouthpiece, enhancer or Coder-Decoder (CODEC).

In late decades a different scope of arrangements has been proposed to address these debasement impacts. Discourse improvement methods plan to re-establish ruined discourse flag by evacuating or making up for corruption without harming the discourse signal itself. The work in this theory is worried about the improvement of single channel discourse flag that have been ruined by levels of added substance commotion that are sufficiently high to influence the coherence of the discourse.

1. Time Delay Estimation

Time delay estimation is the issue of evaluating the time delay between got signals which have started from a similar source. The two beneficiaries are isolated from one another by a separation D. In signal Processing Time Delay Estimation (TDE) is a significant issue for assessing the distinction in appearance times of a sign got at two spatially isolated sensors/mouthpiece [6] within the sight of commotion. The issue has significant applications in differing regions as radar, sonar, seismology, signal insight, geophysics, ultrasonics', and correspondences heading discovering, source confinement and speed following and so forth. The time-postpone Estimation (TDE) issue can have the reasons or intention in the estimation of the deferral to have the best model guess of the framework for a programmed control application or for signal preparing applications like radar go estimation, heading of appearance estimation with cluster radio wires, signal averaging, etc. There exist different techniques for estimation of the time delay happening because of the gathering of the sign at two distinctive spatially isolated beneficiaries. In this proposition we have thought about the Cross-Correlation (CC) technique and Phase Transform (PHAT) strategies falling under the Generalized Cross-Correlation Method of Time Delay Estimation. Furthermore, the Average Square Difference Function (ASDF) and the versatile least mean square channel (LMS) technique for the estimation of time delay. The precision of these calculations is being looked at in this theory work.

2. Need For The Time Delay Estimation

Postponement, otherwise called time slack, emerges in physical, synthetic, organic and monetary frameworks, and during the time spent estimation and calculation. In signal handling, a period delay is otherwise called a period contrast of appearance between two flag; the estimation of such deferral emerges in submerged following applications, biomedicine, geophysics, cosmology, acoustics, seismology and media communications. Frequently in these applications, the time delay is assessed without different procedure parameters.

A. Acoustic Source Localization

The TDE based acoustic source restriction framework utilizes the arrangement of TDOA evaluations to ascertain the area of the acoustic source, for example the area of the source must be tended to, given the cluster geometry and relative TDOA gauges among various amplifier sets. A variety of N receivers is considered at some area. The primary amplifier is viewed as the reference, and set at the beginning of the facilitate framework. The separation between the source [7] and the ith amplifier is assessed .the distinction separation between the mouthpieces I and j from the source is processed. The thing that matters is generally named as the range contrast and is corresponding to the time postponement of appearance with the speed of sound, c.

B. Cognitive Radio Systems

In psychological radio frameworks auxiliary clients can use different scattered groups that are not utilized by essential clients. Subjective radios can be viewed as progressively skilled variants of programming characterized radios as in they have detecting, mindfulness, learning, adjustment, objective driven self-ruling tasks and re-configurability highlights, which encourage effective utilization of radio assets, for example, power and band width. The fundamental distinction between Time defer estimation in psychological radio frameworks and in customary frameworks is that a subjective radio framework can transmit and get over numerous scattered groups.

C. Speech Recognition

Discourse acknowledgment is a significant application using the time defer estimation. Discourse acknowledgment is the way toward changing over an acoustic sign, caught by a receiver or a phone, to a lot of words. The perceived words can be the conclusive outcomes, with respect to applications, for example, directions and control, information section, and report arrangement. They can likewise fill in as the contribution to facilitate semantic handling so as to accomplish discourse understanding. The discourse acknowledgment framework is for the most part influenced by the resonation and not influenced by the clamour. Furthermore, these resonation impacts can be decreased or dispensed with by thinking about the time postpone gauges.

III. PROPOSED WORK

Speech enhancement or noise reduction has been one of the main investigated problems in the speech community for a long time. The problem arises whenever a desired speech signal is degraded by some disturbing noise.

The noise can be additive or convolutive. Even this additive noise can reduce the quality and intelligibility of the speech signal considerably. Therefore, the aim of the noise reduction algorithms is to estimate the clean speech signal from the noisy recordings in order to improve the quality and intelligibility of the enhanced signal. Due to this, it proposes a method using modified codebook based speech enhancement signal to improve noise reduction.

In this work it study the time-defer estimation (TDE) issue, where we need to appraise the Time Delay 'D', for example the issue of assessing the time delay and the relationship work between two got signals is introduced [6]. A numerical model for the two sign is presented. We are keen on the estimation of the time-postpone that the sign endures because of the contrasting spatial areas of the unmistakable collector from the source.

A. System Model

It considers a multi-way condition where one source and two sensors are exhibited; the two sensors are situated at various good ways from a similar source. They got sign at the two amplifiers can be displayed as:

$$r_1(t) = s(t) + n_1(t), \qquad 0 \le t \le T$$

$$r_2(t) = s(t - D) + n_2(t)$$

Where r_1 (t) and r_2 (t) are the yields of the two mouthpieces that are isolated spatially, s(t) is the source signal, n_1 (t) and n_2 (t) are speaking to the added substance noises'', the perception interim, and 'D', the time delay between the two got signals. The sign and clamours are thought to be uncorrelated having zero-mean and Gaussian dissemination. Our goal is to appraise this 'D' and in this way the issue 'Time Delay Estimation'.

Since Time Delay Estimation is a significant system for distinguishing, restricting and following radiation sources. Due to its focal essentialness, exactness and accuracy are of basic significance to the TDE calculations. Since now there exist different techniques and calculations to appraise the time delay. Here we are thinking about the relative investigation of just four strategies for TDE, viz. the Cross-connection Function (CCF) strategy, the Phase Transform (PHAT) Method falling under the Generalized Cross-relationship technique and the Average square Difference Function (ASDF) technique and the versatile least mean square channel (LMS) strategies are talked about and lookedSince Time Delay Estimation is a significant system for distinguishing, restricting and following radiation sources. Due to its focal essentialness, exactness and accuracy are of basic significance to the TDE calculations. Since now there exist different techniques and calculations to appraise the time delay. Here we are thinking about the relative investigation of just four strategies for TDE, viz. the Cross-connection Function (CCF) strategy, the Phase Transform (PHAT) Method falling under the Generalized Cross-relationship technique and the Average square Difference Function (ASDF) technique and the versatile least mean square channel (LMS) strategies are talked about and looked.

B. Speech Enhancement

Most of the vitality in a discourse signal is amassed in the voiced interims. In the time-recurrence space, the greater part of the voiced discourse vitality is situated in few symphonious pinnacles that stay distinguishable even at poor SNRs. In this area, we propose a technique to appraise the discourse dynamic level at low SNRs from the vitality of the consonant tops during voiced interims. Understandability and charm are hard to quantify by any scientific calculation. In existing work, it works on speech enhancement based on atomic notebook-based probability system. It uses a wave signal for further processing and also use wiener filter for noise removal. The main problem in this work is the phase and level difference of signals under different SNR is high and also it does not work on live signal that affects the environment. Typically listening tests are utilized. Be that as it may, since masterminding listening tests might be costly, it has been generally examined how to

anticipate the after effects of listening tests. The focal techniques for upgrading discourse are the evacuation of foundation clamour, reverberation concealment and the procedure of falsely carrying certain frequencies into the discourse signal. Above all else, each discourse estimation performed in a regular habitat contains some measure of reverberation. Echoless discourse, estimated in a unique anechoic room, sounds dry and dull to human ear. By and large the foundation irregular clamour is included with the ideal discourse sign and structures an added substance blend which is grabbed by mouthpiece. It very well may be stationary or non-stationary, white or shaded and having no connection with wanted discourse signal.

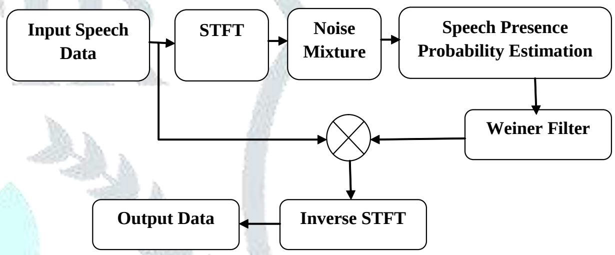

Figure 2: System Model based on Speech Codebook

The power standardization is performed over the whole term of the expression. On the off chance that the information signal was long enough to remember changes for the discourse dynamic level, the sign could be split into fragments to play out this stage. Most voiced discourse vitality is focused inside the major recurrence and its music. In this way, recognizing voiced discourse fragments and assessing their essential recurrence makes it conceivable to find high discourse vitality districts. The calculation gives a basic recurrence gauge at each time span, together with a likelihood of each time allotment containing voiced discourse. Recognizing time spans which contain sibilant telephones is significant for the conservation of occasional discourse vitality at high frequencies. Moreover, an estimation of the power range of the sibilant telephone would likewise help distinguishing the recurrence groups containing the majority of the sibilant discourse vitality.

C. Proposed Method

This work presents speech enhancement using phase codebook method. In this work, it uses the concept of phase and level compensation to improve speech amplitude level as shown in Fig 3.4. Also it works on live speech data that is recorded using microphone slot on system. It detects the high pitch data from live environment and then applies further enhancement steps for improving the system performance. Initially it assumes that all atoms are normalized in nature and is treated as eq. (1)

$$\sum W = 1 \qquad (1)$$

It examines a technique with any pre-preparing of commotion models. The main suspicion about the commotion is that it is not the same as the included discourse.

Consequently, the commotion estimation ends up being finding the parts which can't be enough spoken to by a very much characterized discourse model. Given the great execution of profound learning in signal portrayal, a profound auto encoder (DAE) is utilized for precisely demonstrating clean discourse range. In the upgrade arrange, an extra DAE is acquainted with speak to the lingering part acquired by subtracting the evaluated clean discourse range (by utilizing the pre-prepared DAE) from the uproarious discourse range. Discourse signals are spoken to by their spectro-worldly appropriation of acoustic vitality, a spectrogram. The model-based methodologies proposed in this work in the Mel scale extent spectrogram area, with the term greatness alluding to the square foundation of vitality in a period recurrence component. The cepstral highlights utilized in traditional frameworks depend on a (for the most part logarithmic) pressure of the extent esteems pursued by a de-corresponding cosine change. In this structure, in any case, we utilize the greatness esteems legitimately to disentangle the additivity of discourse and commotion. The extent spectrogram portraying a perfect discourse signal is a B × T dimensional network S (with B recurrence groups and T time spans).

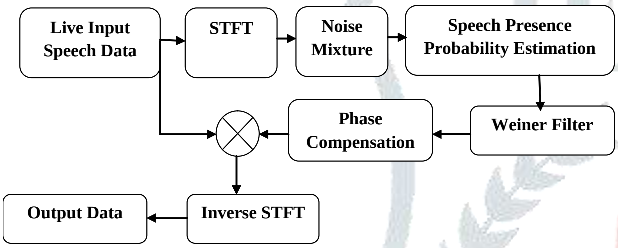

Figure 3: Proposed System Model based on Speech Codebook

Normalization is done by iteratively scaling each row and column so that its Euclidean norm equals unity. After normalization, the norms of the columns equal unity, and the norms of the rows are approximately equal. During decoding, each noisy speech segment y is scaled using the frequency band normalization applied to A. Because the magnitude of the exemplar activations in x can vary, arbitrary speech levels and SNRs can be matched.

IV. RESULTS & DISCUSSION

This work presents a method for speech enhancement using speech codebooks-based method. This method basically works on live speech data. It presents phase and level-based difference methods to enhance the speech by reducing its phase and level difference values. In the binary mask approach to speech enhancement, a binary-valued gain mask is applied to the speech in the time-frequency domain and the signal is then transformed back into the time-domain. This procedure is similar to that used in conventional approaches such as spectral subtraction or MMSE estimators except that, in the latter cases, a continuously variable gain function is applied. The proposed method uses two type of noises namely babble noise and ventilation noise with SNR variation value 0 dB, -5 dB and 5 dB.

1. *Recording of Live Data*

This step includes recording of live signal from user side. It includes any type of speech data normal, abnormal and background data information. It stores all signal data into database for further processing.

2. *Speech Enhancement by Proposed Approach*

This work presents speech enhancement using phase codebook method. In this work, it uses the concept of phase and level compensation to improve speech amplitude level. Also it works on live speech data that is recorded using microphone slot on system. It detects the high pitch data from live environment and then applies further enhancement steps for improving the system performance. Active noise suppression is a method in which the idea is to produce anti-noise into the listener's ear to cancel the noise. The delay must be kept very small to avoid producing more noise instead of cancelling the existing noise. First of all, every speech measurement performed in a natural environment contains some amount of echo. Echoless speech, measured in a special anechoic room, sounds dry and dull to human ear. In most cases the background random noise is added with the desired speech signal and forms an additive mixture which is picked up by microphone. The original and noisy speech are shown in fig 4. This figure shows the original signal representation with its noisy signal with and without filtering. The response with its estimated signal is shown.

Fig. 4: Original Signal with Noisy Filtering Response

The proposed speech enhancement is done by phase compensation method that helps to eliminate errors in signal that improves the amplitude of signal. The existing method shows speech enhancement response in which it removes noise from original signal and provides desired output with enhancement. But the proposed method shows amplitude spike up to 0.8 after enhancement and removal of noise from signal.

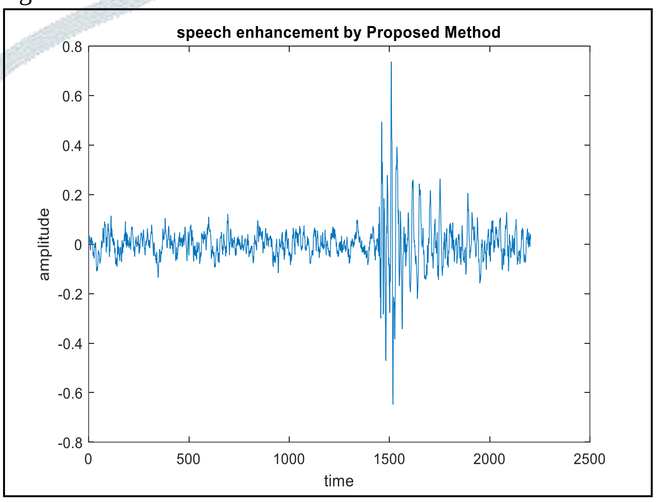

Fig. 5: Speech Enhancement of Signal by Proposed Method

1. *Response Evaluation at SNR=5 dB*

All signals speech and noise are evaluated at 16 KHz frequency value. These signals are processed with STFT with 1024 frame size. The use of Short time Fourier transform provides the generation of time varying components over time. It is used to determine sinusoidal frequency and phase of signal under varying SNR. The performance starts with SNR of 5 dB with babble Noise in system. The response of ILD, IPD and ICM are shown with noisy speech and clean speech after wiener filtering for SNR=5 dB in Fig 6.

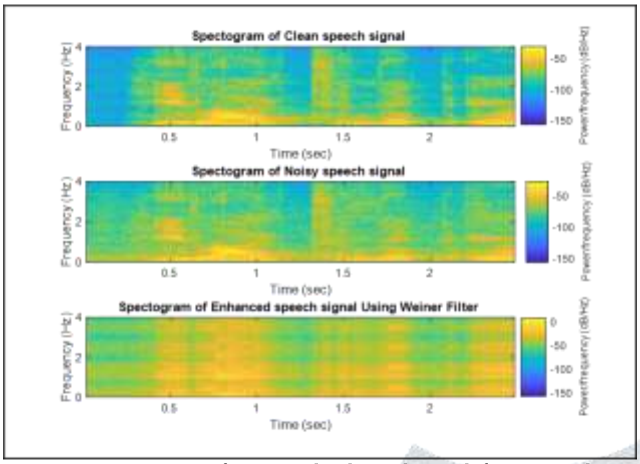

Fig. 6: Spectrogram of Noisy and Enhanced Speech for SNR=5 dB

2. Response Evaluation at SNR= 0 dB

It uses the concept of variation of SNR to next level of 0 dB. After reducing the SNR from 5 to 0 dB, it increases the value of ΔILD and ΔIPD. The increase in change in level and phase difference can decrease the performance of signal quality. The response of ILD, IPD and ICM are shown with noisy speech and clean speech after wiener filtering for SNR=0 dB in Fig 7.

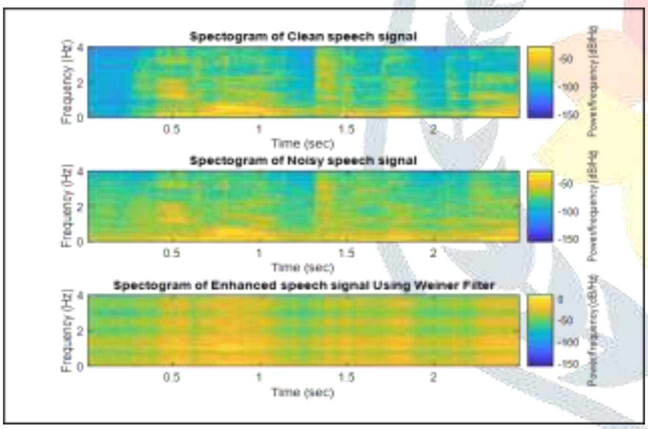

Fig. 7: Spectrogram of Noisy and Enhanced Speech for SNR=0 dB

3. Response Evaluation at SNR= -5 dB

It provides the evaluation response of signal at SNR=-5 dB . After reducing the SNR from 5 to 0 dB, it increases the value of ΔILD and ΔIPD. The increase in change in level and phase difference can decrease the performance of signal quality. The response of ILD, IPD and ICM are shown with noisy speech and clean speech after wiener filtering for SNR=0 dB in Fig 8.

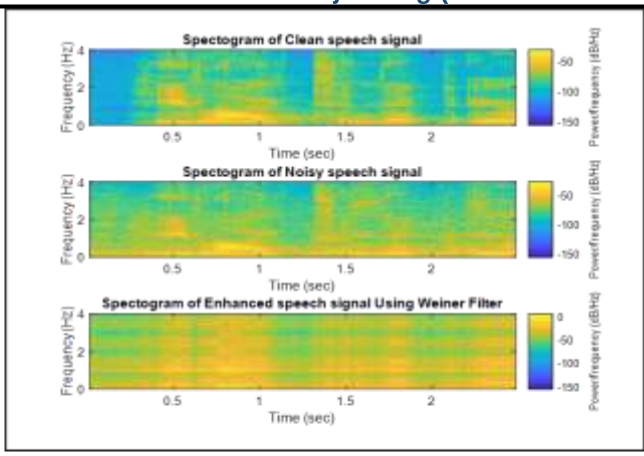

Fig. 8: Spectrogram of Noisy and Enhanced Speech for SNR=5 dB

The Interaural phase difference (IPD) is mainly the spatial data which is frequency dependent and mixture of input signals. It basically provides the difference of phase levels of signals. The Interaural Level Difference (ILD) is a function of frequency and provides the difference in level of signal. Table 2 and 3 show the performance parameters of proposed approach under babble noise and fan noise with their effective values. Table 4 shows the performance comparison of proposed method with actual methods and proposed method shows better improvement in ΔILD, ΔIPD. The change in level and phase difference is reduced by proposed method and better as compared to actual method.

Table 2: Performance Parameters by Proposed Approach for Babble Noise

Parameter	SNR=5 dB	SNR=0 dB	SNR=-5 dB
ΔILD	0.45	0.67	0.81
ΔIPD	0.03	0.12	0.19
Lambda	0.0048	0.0045	0.0037

Table 3: Performance Parameters by Proposed Approach for Fan Noise

Parameter	SNR=5 dB	SNR=0 dB	SNR=-5 dB
ΔILD	0.26	0.54	0.75
ΔIPD	0.13	0.04	0.023
Lambda	0.0042	0.0036	0.0027

Table 4: Performance Comparison of Proposed Approach with Existing Method

Babble Noise						
	SNR= 5 dB		SNR= 0 dB		SNR= -5 dB	
Approach	**ΔILD**	**ΔIPD**	**ΔILD**	**ΔIPD**	**ΔILD**	**ΔIPD**
Existing [23]	1.6	0.04	1.8	0.09	2.2	0.19
Proposed	0.45	0.03	0.67	0.12	0.81	0.19

Fan Noise						
	SNR= 5 dB		SNR= 0 dB		SNR= -5 dB	
Approach	**ΔILD**	**ΔIPD**	**ΔILD**	**ΔIPD**	**ΔILD**	**ΔIPD**
Existing [23]	1.84	0.13	2.08	0.08	0.98	0.10
Proposed	0.26	0.13	0.54	0.04	0.75	0.02

VI. CONCLUSION

This work presents the concept of speech enhancement with phase compensation method to improve accuracy of system. The original goal of binary mask estimation was to identify the regions where the SNR was higher than 0 dB. In addition, we have developed an algorithm for estimating the active level of a speech signal even when high levels of noise are present. The noise can be additive or convolutive. This work presents speech enhancement using phase codebook method. In this work, it uses the concept of phase and level compensation to improve speech amplitude level. Also it works on live speech data that is recorded using microphone slot on system. It detects the high pitch data from live environment and then applies further enhancement steps for improving the system performance. Therefore, the aim of the noise reduction algorithms is to estimate the clean speech signal from the noisy recordings in order to improve the quality and intelligibility of the enhanced signal. It provides the speech enhancement with noise reduction in the system. shows the performance comparison of proposed method with actual methods and proposed method shows better improvement in ΔILD, ΔIPD values.

REFFERENCES

[1] J. S. Lim and A. V. Oppenheim, (1979) "*Enhancement and bandwidth compression of noisy speech*," Proc. IEEE, vol. 67, no. 12, pp. 1586–1604.

[2] P. Vary and R. Martin, (2005) "*Digital Speech Transmission: Enhancement, Coding and Error Concealment*", Chichester, U.K.: Wiley.

[3] P. Loizou, (2005), *Speech Enhancement: Theory and Practice*. Boca Raton, FL: CRC.

[4] S. Doclo and M. Moonen, (2005) ‘‘*On the output SNR of the speech-distortion weighted multichannel Wiener filter*,’’ IEEE Signal Processing Letters, vol. 12, no. 12, pp. 809-811.

[5] Anuradha R. Fukane, L. Sahare, (2011) "*Enhancement of Noisy Speech Signals for Hearing Aids*", International Conference on Communication Systems and Network Technologies, 2011.

[6] E. Plourde, and B. Champagne, (2011) "*Multidimensional STSA Estimators for Speech Enhancement With Correlated Spectral Components*", IEEE Transactions On Signal Processing, Vol. 59, No. 7.

[7] Christian D. Sigg, Tomas Dikk, (2012) "*Speech Enhancement Using Generative Dictionary Learning*", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 20, No. 6.

[8] H. Veisi H. Sameti, (2012) "*Hidden-Markov-model-based voice activity detector with high speech detection rate for speech enhancement*", IET Signal Processing, pp. 01-06.

[9] Xia Yousheng, Huang Jianwen, (2014) "*Speech Enhancement Based on Combination of Wiener Filter and Subspace Filter*", IEEE, pp. 81-86.

[10] Zhang Jie, Xiaoqun Zhao, Jingyun Xu, (2014) "*Suitability of Speech Quality Evaluation Measures in Speech Enhancement*", IEEE pp. 102-108.

[11] Atsunori Ogawa, Keisuke Kinoshita, Takaaki Hori, (2014) "*Fast Segment Search For Corpus-Based Speech Enhancement Based On Speech Recognition Technology*", IEEE International Conference on Acoustic, Speech and Signal Processing.

[12] AN.SaiPrasanna, Iyer Chandrashekaran, (2014) "*Real Time Codebook Based Speech Enhancement with GPUs*", International Conference on Parallel, Distributed and Grid Computing, IEEE, pp. 1042-1048.

[13] Zawar Shah, Ather Suleman, Imdad Ullah, (2014) "*Effect of Transmission Opportunity and Frame Aggregation on VoIP Capacity over IEEE 802.11n WLANs*", IEEE pp. 256-262.

[14] Lee Ngee Tan, Abeer Alwan, (2014) "*Feature Enhancement Using Sparse Reference And Estimated Soft-Mask Exemplar-Pairs For Noisy Speech Recognition*", IEEE International Conference on Acoustic, Speech and Signal Processing, pp. 22-28.

[15] Seung Yun, Young-Jik Lee, and Sang-Hun Kim, (2014) "*Multilingual Speech-to-Speech Translation System for Mobile Consumer Devices*", IEEE Transactions on Consumer Electronics, Vol. 60, No. 3, pp. 232-238.

[16] Shoko Arakit, Tomoki Hayashi, (2015) "*Exploring Multi-Channel Features for Denoising-Auto-encoder-Based Speech Enhancement*", IEEE pp. 424-430.

[17] Zheng Gong and Youshen Xia, (2015) "*Two Speech Enhancement-Based Hearing Aid Systems and Comparative Study*", IEEE International Conference on Information Science and Technology, April 24-26 pp. 122-128.

[18] Feng Deng, Changchun Bao, (2015) "*Sparse Hidden Markov Models for Speech Enhancement in Non-Stationary Noise Environments*", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 23, No. 11.

[19] Pejman Mowlaee and Josef Kulmer, (2015) "*Harmonic Phase Estimation in Single-Channel Speech Enhancement Using Phase Decomposition and SNR Information*", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 23.

[20] Swati R. Pawar, Hemant kumar B. Mali, (2015) "*Implementation of Binary Masking Technique for Hearing Aid Application*", IEEE International Conference on Pervasive Computing, 2015.

[21] Meng Sun, Xiongwei Zhang, Hugo Van hamme, (2016) "*Unseen Noise Estimation Using Separable Deep Auto Encoder for Speech Enhancement*", IEEE Transactions On Audio, Speech, And Language Processing, Vol. 24, No. 1.

[22] Renjith S, Manju K G, (2017) " *Speech Based Emotion Recognition in Tamil and Telugu using LPCC and Hurst Parameters*", International Conference on circuits Power and Computing Technologies, pp. 4967-4973.

[23] Sean U. N. Wood, Johannes K. W. Stahl, (2019) "*Binaural Codebook-based Speech Enhancement with Atomic Speech Presence Probability*", IEEE/ACM Transactions On Audio, Speech, And Language Processing, pp. 01-12.