IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

MULTI-CHANNEL SPEECH DEREVERBERATION AND DENOISING USING DEEP LEARNING BASED ALGORITHM

[1]Prof.S.S.Lavhate,[2]Prof.R.N.Kadu [3]Snehal Balasaheb Shinde

[1]Assistant Professor, [2]Assistant Professor, [3]PG Student,

[1]Electronics Engineering, [2,3]Electronics and Telecommunication Engineering

[1,2,3]Pravara Rural Engineering College, Loni, Dist: Ahmednagar, Maharashtra,India

Abstract: Dereverberation of speech signals in a hands-free scenario by adaptive algorithms has been a research topic for several years now. Speech is more natural approach of gaining access to information, monitoring things and interacting. However, its intuitiveness is not convenient while interacting with computer. In real time situations, the speech quality is degraded by reverberation occurred in rooms and background noise. These effects causes combine corrupt the speech signal and overall speech level degraded resulting in listening issues in human beings moreover, for hearing-impaired persons. Neural network based speech dereverberation has achieved promising results in recent studies. This algorithm investigates deep learning based multi-channel speech dereverberation. For multi-channel processing, we first denoising the signals using symlet wavelet, and then LSTM for Dereverberation, which is expected to be a filtered version of target signals. The performance of the proposed algorithm is evaluated using performance evaluation metrics such Perceptual Evaluation of Speech Quality (PESQ) and LLR are used to validate our results.

Index Terms **- Speech Enhancement, PESQ, LLR, Dereverberation**

I. INTRODUCTION

Reverberation –

Reverberation, in acoustics, is a diligence of sound, or reverberation after a sound is created. Reverberation is made when a sound or sign is reflected making various reflections develop and afterward rot as the sound is consumed by the surfaces of items in the space - which could incorporate furnishings, individuals, and air. This is most observable when the sound source stops yet the reflections, their adequacy decline, until zero is reached. Reverberation is recurrence subordinate: the length of the rot, or reverberation time, gets unique thought in the building plan of spaces which need to have explicit reverberation times to accomplish ideal execution for their expected activity.In correlation with a particular reverberation, that is noticeable at least 50 to 100 ms after the past solid, reverberation is the event of reflections that show up in a grouping of not exactly roughly 50 ms. Over the long haul, the adequacy of the reflections slowly lessens to non-observable levels. Reverberation isn't restricted to indoor spaces as it exists in woods and other outside conditions where reflection exists.

Reverberation happens normally when an individual sings, talks, or plays an instrument acoustically in a lobby or execution space with sound-intelligent surfaces. Reverberation is applied falsely to utilizing reverb impacts, which mimic reverb through implies including closed quarters, vibrations sent through metal, and advanced handling. Despite the fact that reverberation can add effortlessness to recorded sound by adding a feeling of room, it can likewise lessen discourse clarity, particularly when clamor is likewise present. People with hearing misfortune, including clients of amplifiers, regularly report trouble in grasping discourse in reverberant, loud circumstances. Reverberation is likewise a huge wellspring of errors in programmed discourse acknowledgment.

Dereverberation -

Dereverberation is the process by which the effects of reverberation are removed from sound, after such reverberant sound has been picked up by microphones. Dereverberation is a subtopic of acoustic digital signal processing and is most commonly applied to speech but also has relevance in some aspects of music processing. Dereverberation of audio (speech or music) is a corresponding function to blind deconvolution of images, although the techniques used are usually very different. Reverberation itself is caused by sound reflections in a room (or other enclosed space) and is quantified by the room reverberation time and the direct-to-reverberant ratio. The effect of dereverberation is to increase the direct-to-reverberant ratio so that the sound is perceived as closer and clearer. A main application of dereverberation is in hands-free phones and desktop conferencing terminals because, in these cases, the microphones are not close to the source of sound – the talker's mouth – but at arm's length or further distance. As well as telecommunications, dereverberation is importantly applied in automatic speech recognition because speech recognizers are usually error-prone in reverberant scenarios.Dereverberation became established as a topic of scientific research in the years 2000 to 2005. Although a few notable early articles exist. The first scientific text book on the topic was published in 2010. A global scientific study sponsored by the IEEE Technical Committee for Audio and Acoustic Signal Processing took place in 2014. Three different approaches can be followed to perform dereverberation. In the first approach, reverberation is cancelled by exploiting a mathematical model of the acoustic system (or room) and, after estimation of the room acoustic model parameters, forming an estimate for the original signal. In the second approach, reverberation is suppressed by treating it as a type of (convolutional) noise and performing a de-noising process specifically adapted to reverberation. In the third approach, the original dereverberated signal is directly estimate from the microphone signals using, for example, a deep neural network machine learning approach or alternatively a multichannel linear filter. Examples of the most effective methods in the state-of-the art include approaches based on linear prediction.

II. RELATED WORK

Zhong-Qiu Wang and DeLiang Wang (2020) et.al proposed Deep Learning Based Target Cancellation for Speech Dereverberation. (1).Ziteng Wang, Yueyue Na, Biao Tian, Qiang Fu (2020) et.al proposed controllable multichannel speech dereverberation based on deep neural networks. (2).Siva Priyanka,T.Kishore Kumar(2019) et.al (3)Proposed GSC beamformer with LMS, NLMS, RLS. In this RLS gives better performance in comparison with GSC with LMS and NLMS. (3)

S.Siva Priyanka,T.Kishore Kumar(2019) et.al proposed The performance of the proposed GSC- FNLMS is compared with existing GSC-Least Mean Square (LMS) and Normalized LMS (NLMS) algorithms under several noisy conditions. The fast convergence and low complexity, FNLMS adaptive algorithm and GSC structure to have improved speech at the output. An enhanced speech with improved performance is achieved by GSC-FNLMS related to GSC-LMS, and GSC-NLMS algorithms.(4)S. Siva Priyanka, T.Kishore Kumar(2018) et.al proposed enhanced from directional noise using GSC with sub band feedback controller and diffuse noise is suppressed using zelinski-TSNR Postfilter. The proposed algorithm displays better results than existing methods in terms of PESQ, segmental SNR and LSD. (5)Yan Zhao et al.proposed a two-stage system for dereverberation and denoising. Two DNN sub-systems are utilized to perform denoising and dereverberation separately and then form a coherent system by joint optimization. The proposed approach improves objective metrics of speech intelligibility and quality significantly in a wide range of noisy and reverberant conditions.(6)

De Liang Wang et al.provided a comprehensive overview of the research on deep learning based supervised speech separation, speech enhancement (speech-non speech separation), speaker separation (multi-talker separation), and speech dereverberation, as well as multi-microphone techniques.(7).Yong Xu et al.Proposed DNN-based algorithm tends to achieve significant improvements in terms of various objective quality measures and Also improve the performance and make the enhanced speech less discontinuous.(8).Nicoleta Roman and John et al.Proposed Speech intelligibility in reverberation with ideal binary masking: Effects of early reflections and signal-to-noise ratio threshold," The Journal of the Acoustical Society of America, vol. 133, no. 3, pp. 1707–1717, 2013. (9) Bo Wu, Kehuang Li, Minglei Yang, and Chin-Hui Lee et al.Proposed A reverberation-time-aware approach to speech dereverberation based on deep neural networks," IEEE/ACM transactions on audio, speech, and language processing, vol. 25, no. 1, pp. 102–111, 2016.(10)

III. PROPOSED METHODOLOGY

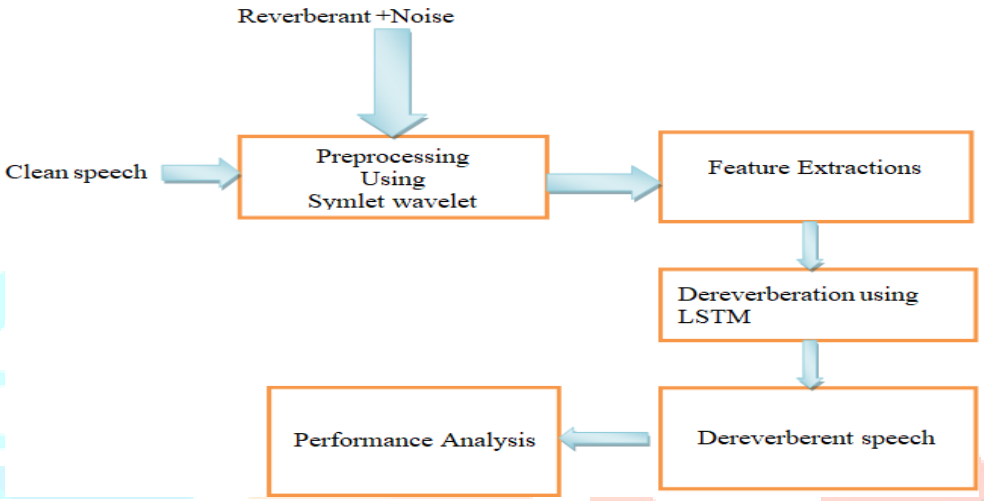


Fig.1.proposed Methodology of the System

Methodology description:

Architecture consists of the preprocessing of the signal using symlet wavelet decomposer, LSTM, Dereverberation, performance analysis blocks and the arrangements of block is exactly similar to the working of the system.

1. Preprocessing using Symlet wavelet:

The very first step in our model is reading of the speech and then preprocessing of the same. Here we are going give Noise, clean speech and reverberant speech as the input. Here speech we are giving is already have a Gaussian noise included in it. We have different types of input signal samples which include male as well as female voice samples. Basically we are going to read all the samples here. In the preprocessing we are going do is zero padding of the signals as every signal is not same in every perspective.We add zeros to the end of the input sequence so that the total number of samples is equal to the next higher power of two. Zero-padding a signal does not reveal more information about the spectrum, but it only interpolates between the frequencies bins that would occur when no zero-padding is applied. In particular, zero-padding does not increase the spectral resolution.Here we are using Symlet wavelet to reduce dimensionality and extract discriminating features from signals.

Here wavelet transform concentrates signal in a few large-magnitude wavelet coefficients. Wavelet coefficients which are small in value are typically noise and you can "shrink" those coefficients or remove them without affecting the signal or image quality. After that threshold the coefficients, we can also reconstruct the data using the inverse wavelet transform.This below plots shows how we get denoise signals by using wavelets.Because wavelets localize features in your data to different scales, we can preserve important signal features while removing noise.

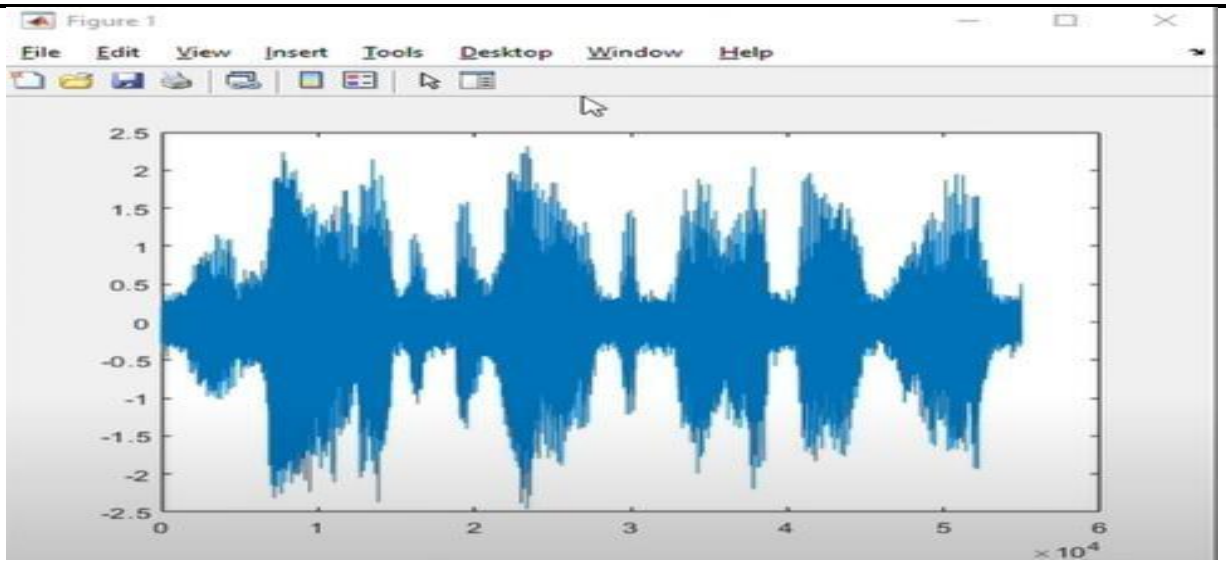


Fig.2.Signal before denoising

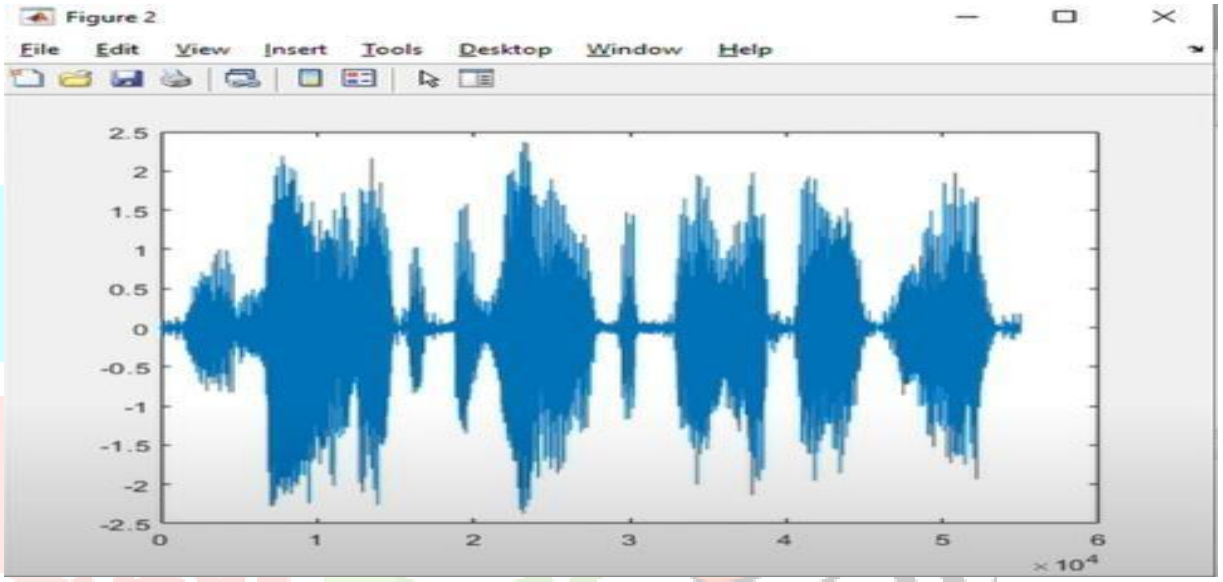


Fig.3.Denoise Audio

If we look at the the figure 1 there are more gaps there in noisy signal and the gap has been decreased in the second one after removing the noise.Also if we look at the end there is some noise at the of the signal and that have been removed after de noising of the signal.

2. **Features Extractions:** Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better results than applying machine learning directly to the raw data.feature extraction used here for to reduce the number of features in a dataset by creating new features from the existing ones (and then discarding the original features

Extractions of Features consist of extraction of features Audio power series, pitch, Formant frequency, Periodogram, Bandpower.

A. Pitch: Pitch is the fundamental period of the speech signal. It the perceptual correlate of fundamental frequency. It represents the vibration frequency of the vocal cords during the sound productions (like vowels, for example).

B. Formant Frequency: Formants are frequency peaks in the spectrum which have a high degree of energy. They are especially prominent in vowels. Each formant corresponds to a resonance in the vocal tract (roughly speaking, the spectrum has a formant every 1000 Hz). Formants can be considered as filters.Formant frequencies, in their acoustic definition, can be estimated from the frequency spectrum of the sound, using a spectrogram (in the figure) or a spectrum analyzer. However, to estimate the acoustic resonances of the vocal tract (i.e. the speech definition of formants) from a speech

recording, one can use linear predictive coding. An intermediate approach consists in extracting the spectral envelope by neutralizing the fundamental frequency,[8] and only then looking for local maxima in the spectral envelope.

*C. Periodogram:*The periodogram is a nonparametric estimate of the power spectral density (PSD) of a wide-sense stationary random process. The periodogram is the Fourier transform of the biased estimate of the autocorrelation sequence. For a signal xn sampled at fs samples per unit time, the periodogram is defined as.

That is, the periodogram is equal to the smoothed sample PSD. In the time domain, the autocorrelation function corresponding to the periodogram is Bartlett windowed.

*D. Band power:*p = band power(x) returns the average power in the input signal, x . If x is a matrix, then band power computes the average power in each column independently. example. p = band power(x , fs , freqrange) returns the average power in the frequency range, freqrange , specified as a two-element vector.

3. **Dereverberation using LSTM:** LSTM can be effectively trained to reduce the average error between the enhanced signal and the original clean signal by considering the effect of the long past time frames.Since the LSTM is free from the vanishing gradient problem, it can work even when there are very long delays. Another important advantage of the LSTM is that it can handle signals that have a mix of low and high frequency components due to the existence of three kinds of gate units. Therefore, the LSTM trained using multi-condition data is expected to be able to perform dereverberation adaptively whether the reverberation time of the test utterance is short or long.The core components of an LSTM network are a sequence input layer and an LSTM layer. A sequence input layer inputs sequence or time series data into the network. An LSTM layer learns long-term dependencies between time steps of sequence data.

4. **Dereverbent speech:**At the output we will get the speech which is dereverberent.It does not have any echo in it as well as it is clean one.

5. **Performance Analysis:**Once the data is downloaded, preprocess the downloaded data and extract features before training the DNN model.

Log likelihood ratio (LLR) - Linear predictive coding (LPC) based objective measurement. Smaller values indicate better quality.

PESQ (Perceptual Evaluation of Speech Quality):PESQ was developed to model subjective tests commonly used in telecommunications to assess the voice quality perceived by human beings. Consequently, it employs true voice samples as test signals. In order to characterize the listening quality as perceived by users, it is of paramount importance to load modern telecom equipment with speech-like signals. Many systems are optimized for speech and would respond in an unpredictable way to non-speech signals (e.g., tones, noise). Guidelines for proper applications of voice test samples are defined in the PESQ application guide contained in Recommendation.

6. **Dereverberation:** Dereverberation is the process by which the effects of reverberation are removed from sound, after such reverberant sound has been picked up by microphones. Dereverberation is a subtopic of acoustic digital signal processing and is most commonly applied to speech but also has relevance in some aspects of music processing. Dereverberation of audio (speech or music) is a corresponding function to blind deconvolution of images, although the techniques used are usually very different. Reverberation itself is caused by sound reflections in a room (or other enclosed space) and is quantified by the room reverberation time and the direct-to-reverberant ratio. The effect of dereverberation is to increase the direct-to-reverberant ratio so that the sound is perceived as closer and clearer.

IV. Experimentation

For experimentation and implementation of proposed technique we used the hardware and software for training and testing the model is as i7-7700k CPU and Nvidia 2060 GPU,Ram – 32 GB,HDD – 2 TB ,Windows 11 CUDA 11.2,cuDNN v8.4.5,TensorRT-
We have used 30 samples of male and female from 30 samples IEEE database has been used in this work to
Evaluate proposed approach. The reverberant speech signal is generated by convolving anechoic speech signal with Room Impulse Response (RIR). RIR was generated using image method.

white noise with -5db SNR is added in reverberant signal to make noisy and reverberant signal. Room dimensions were set to [6.1x5.3x2.7] m, reverberation time RT60 was set to 0.5 s, RIR were simulated at sampling rate of 16 kHz. The array of three micro-phones were used with inter distance between them is 4 cm. Test speech signal was processed frame by frame where each frame was 32 ms with 8 ms overlapped.

Furthermore four different source to array distances (1 to 4 m) were used for performance evaluation in proposed system. Input reverberant signal is considered as unprocessed signal and dereverberated output is considered as processed signal.

To show the performance improvement of the proposed model objective measures like PESQ , output SNR,LLR these parameters are calculated as follows. We will analyze results and plot the different analyzed data in given form.

a). SNR:

SNR or signal-to-noise ratio is the ratio between the desired information or the power of a signal and the undesired signal or the power of the background noise. SNR is defined as the ratio of signal power to the noise power, often expressed in decibels.

b). PESQ:

There are a number of methods available but this article is restricted to one of the more modern ones called PESQ (Perceptual Evaluation of Speech Quality). The PESQ Algorithm is designed to predict subjective opinion scores of a degraded audio sample. PESQ returns a score from 4.5 to -0.5, with higher scores indicating better quality.PESQ is designed to analyse specific parameters of audio, including time warping, variable delays, transcoding, and noise. It is primarily intended for applications in codec evaluation and network testing.

c). LLR

A ratio whose numerator and denominator comprise log likelihoods. It is used (in the log-likelihood test) to evaluate the goodness of fit of the null hypothesis and the alternative hypothesis in explaining sample data.

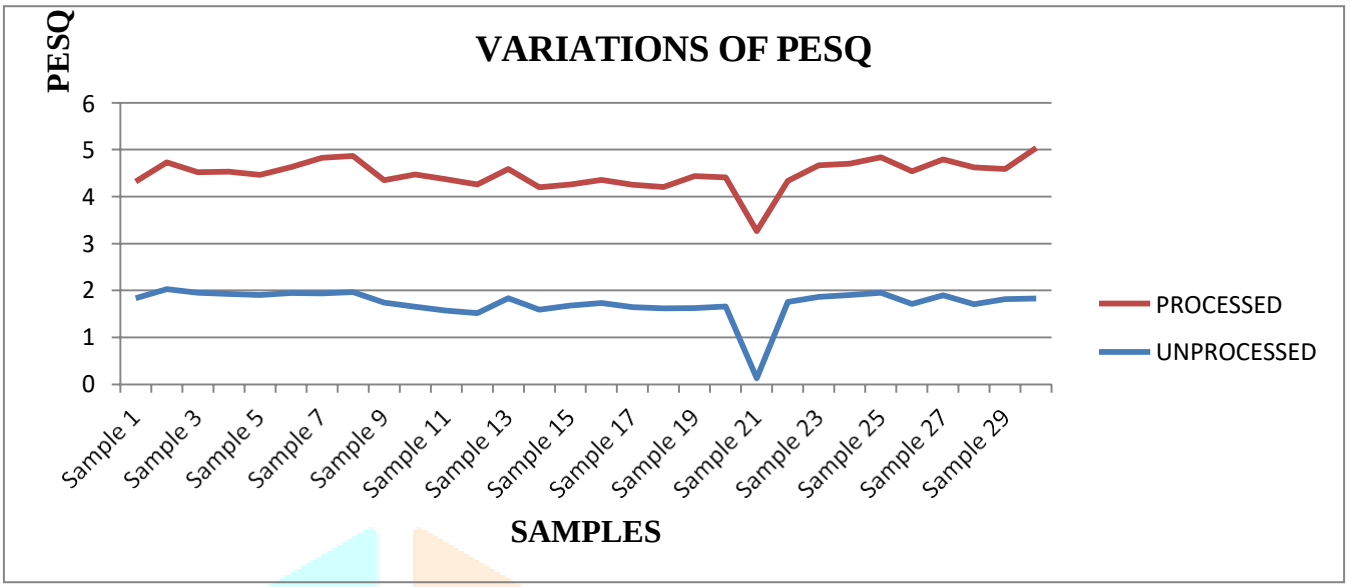
the likelihood-ratio test assesses the goodness of fit of two competing statistical models based on the ratio of their likelihoods, specifically one found by maximization over the entire parameter space and another found after imposing some constraint

V. RESULTS AND DISCUSSION

	PESQ		LLR	
Sample's	UNPROCESSED	PROCESSED	UNPROCESSED	PROCESSED
Sample 1	1.838148	2.480197	0.410906	0.321553
Sample 2	2.027186	2.701343	0.530829	0.402997
Sample 3	1.953089	2.565902	0.452539	0.3762
Sample 4	1.928741	2.603989	0.487084	0.424093
Sample 5	1.90396	2.55705	0.461706	0.405236
Sample 6	1.946673	2.683976	0.568606	0.436928
Sample 7	1.938128	2.885624	0.493656	0.375
Sample 8	1.970976	2.892193	0.526526	0.39771
Sample 9	1.743215	2.605468	0.574224	0.400609
Sample 10	1.65225	2.819748	0.529312	0.419733
Sample 11	1.572845	2.793427	0.621202	0.453213
Sample 12	1.51981	2.736882	0.537746	0.412703
Sample 13	1.84118	2.742936	0.52937	0.450057
Sample 14	1.596161	2.600962	0.471591	0.44649
Sample 15	1.6846	2.575632	0.609049	0.417222
Sample 16	1.738974	2.615991	0.490522	0.414045
Sample 17	1.644887	2.605521	0.467669	0.379936
Sample 18	1.622408	2.582112	0.488102	0.372766
Sample 19	1.626008	2.809245	0.526157	0.463051
Sample 20	1.660265	2.750238	0.495814	0.440127
Sample 21	0.136784	3.130019	0.530657	0.39509
Sample 22	1.757366	2.574675	0.461611	0.34346
Sample 23	1.867899	2.803233	0.490431	0.375411
Sample 24	1.905766	2.799281	0.559118	0.363583
Sample 25	1.952503	2.885163	0.557362	0.396254
Sample 26	1.714107	2.825795	0.48338	0.395968
Sample 27	1.903114	2.890719	0.497154	0.371058
Sample 28	1.707776	2.912522	0.47791	0.410226
Sample 29	1.820479	2.768581	0.525037	0.420037
Sample 30	1.830266	3.204764	0.417366	0.304694

Table 1:- Variations of PESQ, LLR for different samples

In Table 2, we directly evaluate the performance of the trained dereverberation models on Test II on th base paper. Our models obtain dramatically better performance than WPE, and WPE + Beamforming and also than the base paper which applies weighted DAS (WDAS) beamforming on the output of WPE, and WPE + DNN-Based MVDR.These comparisons show that the trained DNN models exhibit good generalization to novel reverberant and noisy conditions, and array configurations.

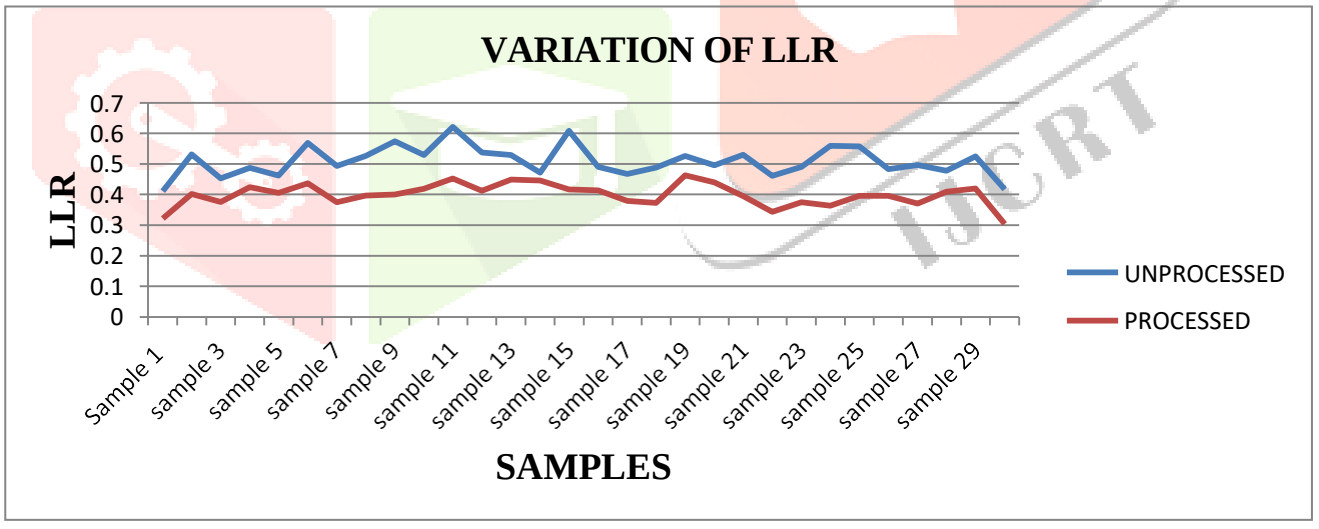


Plot 1: Variations of PESQ with different input samples for unprocessed and proposed (processed) data.

The above figure shows Variations of PESQ with different input samples for unprocessed and proposed data.

In above Plot Brown is showing processed data and purple is showing Unprocessed data.

From Plot it is clear that PESQ has increased for our proposed system than the unproposed system. If we see the graph the values of the PESQ has been slightly increased here.



Plot 2: Variations of LLR with different input samples for unprocessed and proposed (processed) data

The above figure shows Variations of LLR with different input samples for unprocessed and proposed data.

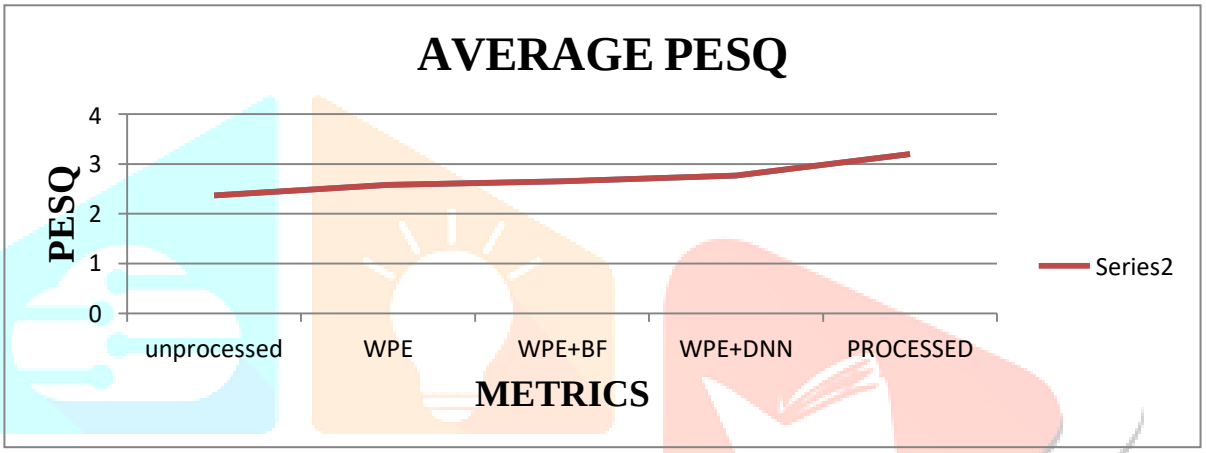
From Plot it is clear that LLR has decreased for our proposed system than the previous three methods. If we see the graph the values of the LLR has been slightly Decreased here

.In above Plot Brown is showing processed data and purple is showing unprocessed data.

SIMULATED DATA					
Metric	Unprocessed	WPE	WPE+BF	WPE+DNN	PROCESSED
PESQ	2.37	2.58	2.65	2.77	3.2
LLR	0.67	0.61	0.6	0.55	0.301471

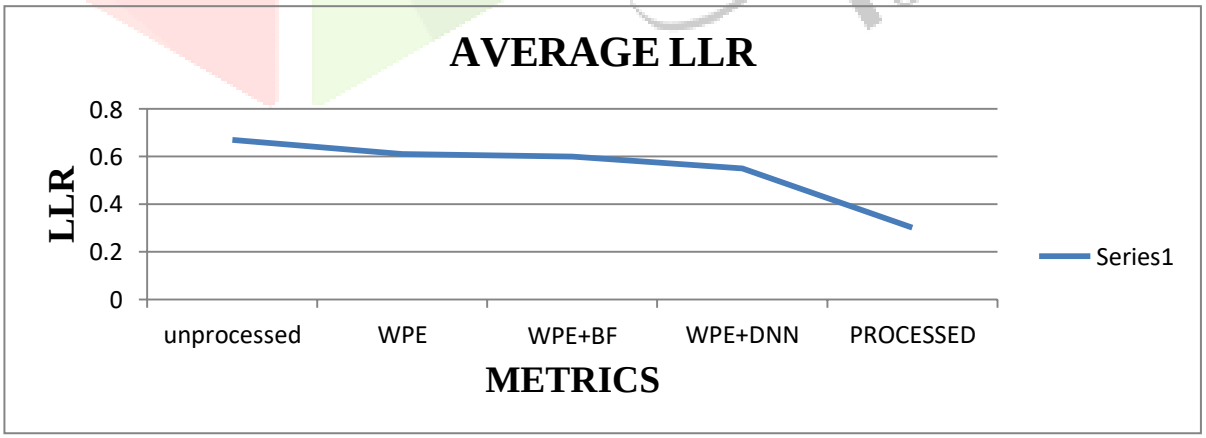
Table 2:- Average LLR, PESQ of different approaches

If we observe the above Table there are the values of average LLR, PESQ of different approaches. From above plot we can see the Variation of average LLR, PESQ of different approaches.



Plot 3: Variation of average PESQ of different approaches

From Plot it is clear that PESQ has increased for our proposed system than the previous three methods. If we see the graph the values of the PESQ has been slightly increased here.



Plot 4.Variation of average LLR of different approaches

From Plot it is clear that LLR has decreased for our proposed system than the previous three methods. If we see the graph the values of the LLR has been slightly Decreased here.

VI. CONCLUSION

We have proposed a Multi-channel speech dereverberation and denoising algorithm using Deep learning. We have used Symlet wavelet for denoising of the signal and LSTM for speech dereverberation.performance of the proposed algorithm is evaluated with the parameters PESQ,LLR. We have used the different samples of male and females speeches for model training. Our multi-channel dereverberation algorithm shows performance improvements over WPE, WPE+BF Model, WPE+DNN.

The evaluated parameters have more accuracy than the previous methods.PESQ value is increased by 20% however LLR has decreased by 43% here.

REFERENCES

1. Zhong-Qiu Wang and DeLiang Wang (2020) proposed Deep Learning Based Target Cancellation for Speech Dereverberation.
2. Ziteng Wang, Yueyue Na, Biao Tian, Qiang Fu (2021) proposed controllable multichannel speech dereverberation based on deep neural networks.
3. S. Siva Priyanka, T. Kishore Kumar (2019) ―GSC Beamforming using Different Adaptive Algorithms for Speech Enhancement‖ Institute of electrical and electronics engineers 45670.
4. S. Siva Priyanka, Kishore Kumar (2019) ―GSC Adaptive Beamforming Using Fast NLMS Algorithm for Speech Enhancement‖ Institute of electrical and electronics engineers 160-165, 2019.
5. S. Siva Priyanka, Kishore Kumar (2018) ―Adaptive Beamforming using Zelinski- TSNR Multichannel Postfilter for Speech Enhancement‖ Institute of electrical and electronics engineers 43488.
6. Y. Zhao, Z. Q. Wang, and D. Wang, "Two-stage deep learning for noisy-reverberant speech enhancement," IEEE/ACM Trans. Audio Speech Lang. Process., vol. 27, no. 1, pp. 53–62, Jan. 2019.
7. D. Wang and J. Chen, "Supervised speech separation based on deep learning: An overview," IEEE/ACM Transactions on Audio Speech and Language Processing, vol. 26, no. 10. Institute of Electrical and Electronics Engineers Inc., pp. 1702–1726, 01-Oct-2018.
8. Donald S Williamson and DeLiang Wang, "Timefrequency masking in the complex domain for speech dereverberation and denoising," IEEE/ACM transactions on audio, speech, and language processing, vol. 25, no. 7, pp. 1492–1501, 2017.
9. Feng Ni, Yi Zhou, Hongqing Liu (2019) ―A Robust GSC Beamforming Method for Speech Enhancement using Linear Microphone Array‖ Institute of electrical and electronics engineers,2019.
10. K.V.Sridhar, T Kishore kumar (2019) ―Performance Evaluation of CS Based Speech Enhancement using Adaptive and Sparse Dictionaries‖ Institute of electrical and electronics engineers 47735.