FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 322-328 | e-ISSN: 2347-470X

Speech Enhancement with Background Noise Suppression in Various Data Corpus Using Bi-LSTM Algorithm

Vinothkumar G[1] and Manoj Kumar D[2]
[1]*Research Scholar, Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu, India, vinothkg@srmist.edu.in*
[2]*Assistant Professor, Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu, India, manojkud1@srmist.edu.in*

*__Correspondence:__ Vinothkumar G, vinothkg@srmist.edu.in

ABSTRACT- Noise reduction is one of the crucial procedures in today's teleconferencing scenarios. The signal-to-noise ratio (SNR) is a paramount factor considered for reducing the Bit error rate (BER). Minimizing the BER will result in the increase of SNR which improves the reliability and performance of the communication system. The microphone is the primary audio input device that captures the input signal, as the input signal is carried away it gets interfered with white noise and phase noise. Thus, the output signal is the combination of the input signal and reverberation noise. Our idea is to minimize the interfering noise thus improving the SNR. To achieve this, we develop a real-time speech-enhancing method that utilizes an enhanced recurrent neural network with Bidirectional Long Short Term Memory (Bi-LSTM). One LSTM in this sequence processing framework accepts the input in the forward direction, whereas the other LSTM takes it in the opposite direction, making up the Bi-LSTM. Considering Bi-LSTM, it takes fewer tensor operations which makes it quicker and more efficient. The Bi-LSTM is trained in real-time using various noise signals. The trained system is utilized to provide an unaltered signal by reducing the noise signal, thus making the proposed system comparable to other noise-suppressing systems. The STOI and PESQ metrics demonstrate a rise of approximately 0.5% to 14.8% and 1.77% to 29.8%, respectively, in contrast to the existing algorithms across various sound types and different input signal-to-noise ratio (SNR) levels.

Keywords: RNN, Bi-LSTM, SNR, Speech Enhancement, Background Noise, DNN.

ARTICLE INFORMATION
Author(s): Vinothkumar G and Manoj Kumar D;
Received: 14/10/2023; **Accepted:** 10/03/2024; **Published:** 30/03/2024
E- ISSN: 2347-470X
Paper Id: IJEER231013
Citation: 10.37391/IJEER.120144
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-12/ijeer-120144.html

Publisher's Note: FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

In today's rapidly evolving digital era, communication is crucial. There are so many areas that require effective communication to exchange important information. For example, there will be a huge catastrophic failure in the air, if there is even slight miscommunication in the airport. It shows the need for effective communication. There are major factors that affect communication, which include unwanted background noise or distortion. This may create a great impact during a crucial time of communication. These problems are cleared by the crucial technique in signal processing called speech enhancement [1]. It generally concentrates on enhancing both the clarity and the quality of the audio signals used for communication. This is achieved by reducing the unwanted background noise or distortion. It seeks to enhance the audio signal quality to provide clear communication between speakers and listeners [2], [3]. The speech enhancement technique is performed by several steps, which contain filtering, noise reduction, pre-processing, and waveform reconstruction. The tool that helps to achieve all the steps is Machine learning. Machine learning is one of the important aspects of the modern era, which contains millions of algorithms for specific needs. In this study, we utilize the modified Recurrent Neural Network (RNN) algorithm to remove background noises [4]. We are using this to model the temporal dependencies of the audio or speech signals and noise suppression models to meet our requirements. This technique is based on the recent development in the DNN which has performed speech enhancement for decades.

RNN algorithm uses two stages to perform the task of noise suppression, which includes the Training and Enhancement stages. The former contains the Noise speech training set, which possesses all the background noises and acts as a training data set. Whereas, the latter is responsible for the enhancement of the audio signals. In conclusion, this technique has the potential to enhance the quality of the audio signals and have effective communication in a noisy environment, where it can have a great impact on the majority of the fields [5].

The fundamental block diagram for deep learning-based voice improvement is shown in *Figure 1*. In this, the noisy speech signal is initially applied to the feature extraction block. Consequently, the feature extraction block receives the noisy speech input first. Based on mean and standard deviation parameters, the aforementioned block is used for extracting speech characteristics from noisy speeches.

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 322-328 | e-ISSN: 2347-470X

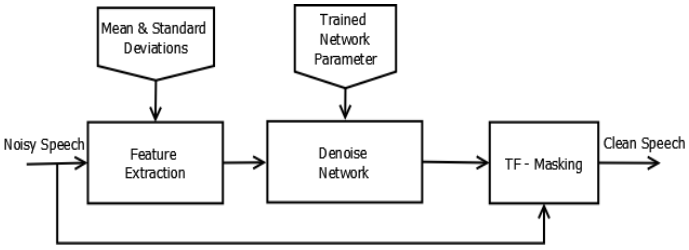

Figure 1: Basic Block Diagram for Noise Suppression with Speech Enhancement

Mean and Standardization (MS), which is first applied to the noisy speech input, is carried out to generate the normalized feature vector. From the reference of feature extraction machine learning algorithm will train the clean speech data from noisy datasets. Initially, the algorithm is trained by speech features from the dataset. The Denoise network is primarily employed to enhance voice quality by removing the background and distracting noises from noisy speeches. The Denoise network output is applied to the Time-Frequency masking technique with reference to noisy speech data. The estimated de-noised speech spectrum is then obtained by applying these masks in the consequent masking block.

1.1. Research Objectives

To develop a unique deep-learning algorithm to enhance voice quality in the context of background noises, the following research goals were addressed in this work.

1. The suggested technique enhances Perceptual Evaluation of Speech Quality (PESQ)) when compared with state-of-the-art Bidirectional-LSTM.
2. Improves speech quality parameter of Short-Time Objective Intelligibility (STOI).
3. Used for undetectable and extremely non-stationary noise types, such as voice interference.
4. To suppress various kinds of noise, for example, causal and non-causal types of noise.
5. Speech enhancement in multiple-noise Corpus.
6. It regularly surpasses all benchmark approaches and could enhance intelligibility in low-SNR situations.

2. LITERATURE REVIEW

In recent years, there have been more studies on how to improve speech. Jannu et al. [6] analyzed various speech enhancement methods and how deep neural networks (DNN) function. Background, ambient, and reverberant noise often impede speech interactions. There are a number of techniques for processing that may be employed to improve speech, including the short-time Fourier transform (STFT), short-time autocorrelation (STAC), and short-time energy (STE). Features like the Mel-Frequency Cepstral Coefficients (MFCCs), Logarithmic Power Spectrum (LPS), and Gamma-tone Frequency Cepstral Coefficients (GFCCs) may be added to a DNN to reduce speech distortion. Because it creates models utilizing a large amount of training data and assesses the effectiveness of the improved speech using specific performance parameters, DNN is crucial for speech enhancement. Numerous speech enhancement techniques have been looked at in this topic recently. Schroter et al. [7] proposed a wideband spectrogram-focused deep filtering noise reduction method based on DNN. Or to put it another way, the author predicted the complex filter coefficients that will be linearly fed to the congested band. They demonstrated that the various filter sizes outperform a sophisticated ratio mask by comparing them on the time and frequency axis. To effectively use the deep filtering method, the research additionally utilized a frequency response loss that works on a per-frequency-band method.

Several potential speech recognition and augmentation methods were described by Hepsiba et al. [8]. RNN and DNN were trained to conduct spectral masking, and other techniques such as ResLSTM, MOGA, and DCCRN were carried out to figure out the magnitude spectrograms of the impaired speech. Spectral subtraction, Wiener and Kalman filtering, MMSE estimation, phase spectrum compensation, ME2E, binaural codebook-based speech augmentation, VAD, ANC methods, as well as beamforming, are among the computational methods it employs to recognize speech. For the system that recognizes speech to better understand speech, noise must be removed. A causal attention mechanism and an encoder-decoder LSTM were used by Peracha et al. [9]. The introduction of a dynamical-weighted (DW) loss function enhances model learning by varying the weight loss values. According to the study, the recommended approach gradually enhanced speech conciseness, clearness, and noise reduction. The LSTM-based estimated suppression time-frequency mask outperformed the previous model for unknown noise types in the causal processing mode.

Huang et al. [10] examined speech improvement which is a cutting-edge masking-based LSTM method that replaced conventional unsupervised algorithms like the log-MMSE. Each method differs in its characteristics and limitations, making it difficult for it to constantly successfully handle noise. In order to create the AGM, which acts as a reliable learning objective for the proposed model, the study determined the optimal ratio mask from the trainer model integrated into the log-MMSE approach. Pandey Ashutosh & DeLiang Wang [11] proposed an Attentive recurrent network (ARN), also referred to as an SA-RNN, for enhancing cross-corpus generalization and time-domain voice augmentation. ARN is built on RNNs with SA and FF blocks. They examined ARN on a dissimilar corpus with non-stationary sounds and low SNR levels. The experimental findings demonstrated that the ARN model greatly outperformed competing time-domain speech improvement techniques like RNNs and dual-path ARNs.

A method employing the f-16, white, and babbling noises was recommended by Shukla et al. [12]. To show the effectiveness of the proposed method, PESQ, STOI, and SDR associated with various versions of OMP-based CS algorithms were employed. The findings from the study showed that the

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 322-328 | e-ISSN: 2347-470X

proposed approach outperformed the different OMP-based CS algorithm versions for speech parameter improvement by a maximum of 30% to 50%. Saleem et al. [13] proposed a residual connection-based BiGRU augmented KF model for improving and detecting speech. Driving noise variances and linear prediction coefficients (LPCs), which model both clear speech and noise signals as AR processes, make up the factors of the anticipated model. Noise variances were obtained using recurrent neural networks that have been trained to estimate line spectrum frequencies (LSFs) to reduce the variance concerning the expected and modeled autoregressive spectra of the noisy speech. Noisy speech is treated to enhanced KF with the estimated factors to reduce background noise and improve the clarity of the speech, comprehensibility, and bit error rates when it comes to replicating long-term dependencies.

In relation to objective sound quality measures, speech recognition accuracy, and model complexity, Yu, Meng, et al. [14] presented a framework that surpassed the cutting-edge F-T-LSTM hybrid echo cancellation and speech enhancement technique. They illustrated how an all-encompassing front-end speech improvement system may be constructed by combining this model with speaker embedding for increased target speech enhancement. Additionally, a branch is established for the automated gain control (AGC) task. Vinothkumar & Phani Kumar Polasi [15] combined the Sign-Sign LMS and the MPNLMS with a hybrid approach in order to enhance voice quality while lowering sparse noise. The increased SNR output segmentation for the SS-MPNLMS spans around 50% and above of the MSE. In comparison with the algorithms created for various sounds with various SNRs, PSNR also increased by 2.47% to 44.4%. Hence, it will be able to increase voice quality while they work, reduce noise, and preserve their hearing abilities with the help of the software that is being recommended.

3. BLOCK DIAGRAM

Bidirectional recurrent neural networks are essentially just two separate RNNs together. This framework enables these networks to access information about the sequence at every step, either forward or backward. When employing bidirectional, the supplied inputs will be processed in two distinct manners: one from the present to the future and the other from the future to the present. This technique differs from unilateral approaches in that it retains future information in the LSTM which travels backward by combining the two hidden states, which allows us to store information from both the present and the future at any given moment.

Figure 2 depicts the necessary stages to have an enhanced audio signal by using machine learning algorithms. The block diagram involves two stages namely, the training and enhancement stages. The former consists of dual stages, which are the Noisy Speech training sets and the Clean speech training sets. The input named "Noisy Speech training set" consists of the unwanted noise audio signal which is needed to be eliminated from the raw audio signal. The other input "Clear Speech training set" is the required audio signal which needs to be enhanced by the recurrent Neural Network (RNN) algorithm [16]. Every step involves the same procedure known as "Feature Extraction," which involves turning unfiltered raw data into numerical attributes that may be used while retaining the integrity of the original data set's contents. The inputs are subsequently fed into the Enhancement stage's recurrent Neural Network - LSTM intermediate process [17].

The Enhancement stage is the main process in the recurrent Neural Network algorithm to have a clear enhanced audio signal. The Enhancement stage consists of one input which is the "Noisy Speech training set", which is denoted by Y(m). The noisy speech training set then undergoes two kinds of processes, they are feature extraction and pre-processes. Pre-processes refers to the modification or manipulation of a data set before it is used to get a better and enhanced performance [18]. The result was then combined and represented as Y^{LPS} and sent to Recurrent Neural Network - LSTM training, which is the output obtained from the training stage. The output of the RNN-LSTM training is denoted as X^{LPS}. The output is then fed to the waveform reconstruction step. The waveform reconstruction step contains two inputs one is from the RNN-LSTM training, X^{LPS} and another output from the pre-processes step which is represented as $<Y^{R}$. Waveform reconstruction is a type of technique where it performs and simulates some unpredictable events which is developed by some devices under test. The output of the waveform reconstruction step is then fed to the enhanced speech. It is the final step to get a desired and enhanced audio signal from the noisy training data set.

The output is denoted as the X^{Y}. Feature extraction is a small step in the enhancement stage which undergoes lots of steps inside. First, the noisy speech training set is sent to the stage of framing. Framing is a process of separating the audio signals in different frames [19]. The output from the framing is then fed into Discrete Fourier Transform (DFT).

The output from the DFT is then separated into two steps. One is the square of the product of the parallel process and another input is sent to the comparison operator. The final output of the first step is Y^(LPS). And, the final output of the second step is < YR. Waveform reconstruction is a type of step in the enhancement stage [20]. The output from the RNN-LSTM training XLPS is fed into the Exp (), which corresponds to the process of exponential the input.

The output of Exp ( ) is then fed into the Inverse Discrete Fourier Transform (IDFT). There are two inputs for the IDFT [21]. One from the Exp ( ) and another one from the feature extraction process, $<Y^{R}$. The output of IDFT is fed to overlap-add and the final output of waveform reconstruction is $X^{\wedge R}$ [22] - [25].

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 322-328 | e-ISSN: 2347-470X

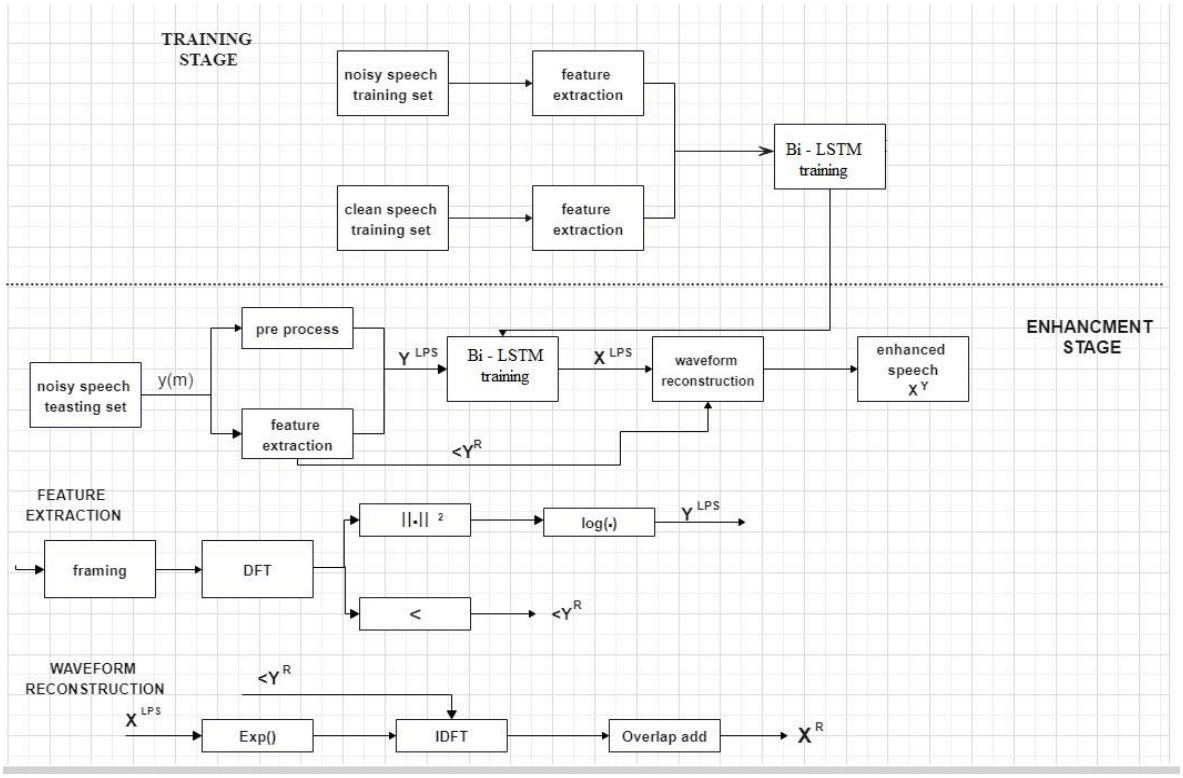

Figure 2: Block diagram of Proposed Model

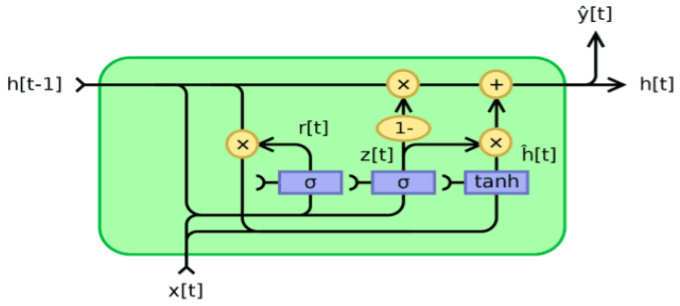

Figure 3: LSTM – Module

Figure 3 shows a mono-directional LSTM-based RNN network structure and the formulas below are used to run the required data and form the parameters to operate the algorithm for suppression of noise.

The current state is: $h_t = f(h_{t-1}, x_t)$ (1)

Activation function *(*tanh*)*:

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t) \quad (2)$$

Output: $y_t = W_{hy}, h_t)$ (3)

The working model of the LSTM algorithm is shown in *figure 3*. Given an input vector sequence $\{x_1,...,x_{t-1}, x_t, x_{t+1},...,x_T\}$, the hidden state at the time t, h_t is computed as,

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad (4)$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (5)$$

$$g_t = Tanh(W_{gx}x_t + W_{gh}h_{t-1} + b_g) \quad (6)$$

$$o_t = \sigma(W_{ox}x_t + W_{io}h_{t-1} + b_o) \quad (7)$$

$$c_t = f_t \otimes c_{t-1} + i \otimes g_t \quad (8)$$

$$h_t = o_t \otimes Tanh(c_t) \quad (9)$$

$$\sigma(s) = \frac{1}{1+e^{-s}} \quad (10)$$

$$Tanh(s) = \frac{e^{s} - e^{-s}}{1+e^{-s}} \quad (11)$$

where, at time *t*, x_t, g_t, and c_t correspond to the input, block input, and memory (cell) state, respectively. Additionally, the gates known as input gate, disregard gate, and output gate, respectively, are i_t, f_t, and o_t. Trainable weights and Biases are

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 322-328 | e-ISSN: 2347-470X

indicated by *W*'s and *b*'s. Since RNN is utilized with Bi-LSTM, it was seen from experimentation that this combination provides better results compared to other algorithms. *Figure 4* shows the Bi-LSTM-based RNN algorithm. The input layer passes to both forward and backward directions through the activation layer. By using this manner, we observed quality results compared with unidirectional-based LSTM.

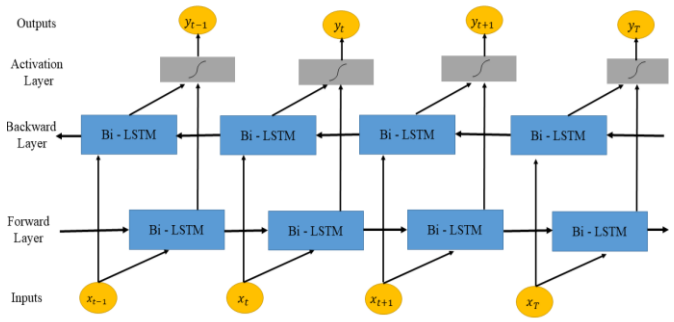

Figure 4: Bi-LSTM-based RNN algorithm

4. SIMULATION RESULTS

The time-domain plot as shown in *Figure 5* is compared with the algorithm's performance of noisy speech of babble noise with 10 dB SNR input. Bi-LSTM performance appears to be nearly identical to clean expression. In addition, the current speech signal quality has been compared, where it provides a better speech signal quality that is similar to the initial clean speech.

The Spectrogram plot as shown in *Figure 6* is compared with the algorithm's performance of noisy speech of babble noise with 10 dB SNR input. Bi-LSTM performance appears to be nearly identical to clean expression. In addition, the current speech signal quality has been compared where it provides better speech signal quality that is similar to the initial clean speech.

For testing the proposed algorithm performance, we have taken some existing algorithms like DCCRN, RNN-IRM, RNN-TCS, DCN, DPARN, and ARN. *Figures 7 and 8* show the STOI level for Babble and Cafeteria noise suppression of different datasets. The graph shows the STOI range of the same noise in different corpus. For -5dB and -2dB noise level consideration, we observed improved results in the proposed system compared with existing algorithms. Likewise, *Figures 9 and 10* show another parameter for speech enhancement of PESQ. Both STOI and PESQ will acquire better performance from the proposed algorithm.

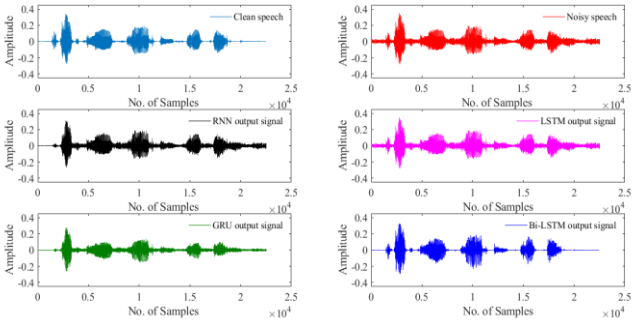

Figure 5: Time domain response of Clean speech, Noisy Speech, Current and Proposed algorithm output

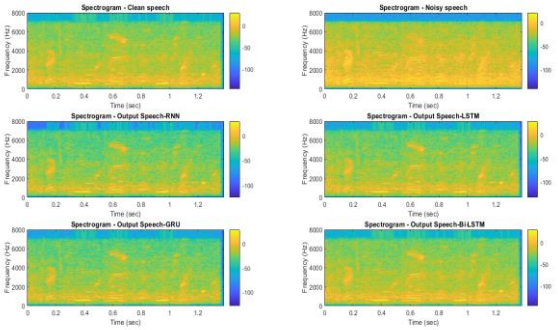

Figure 6: Spectrogram of Clean speech, Noisy Speech, Current and Proposed algorithm outputs

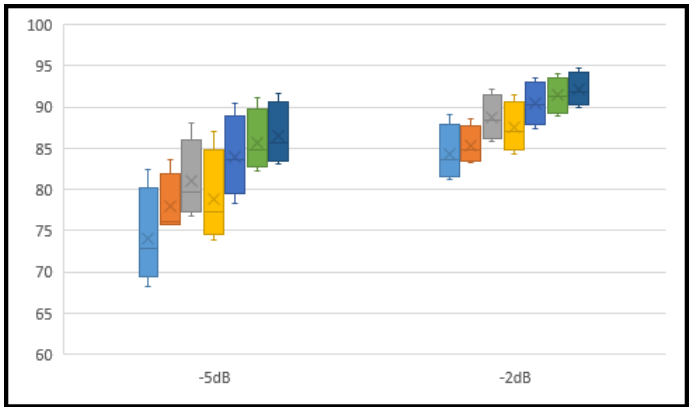

Figure 7. STOI (%) for Babble noise of different datasets

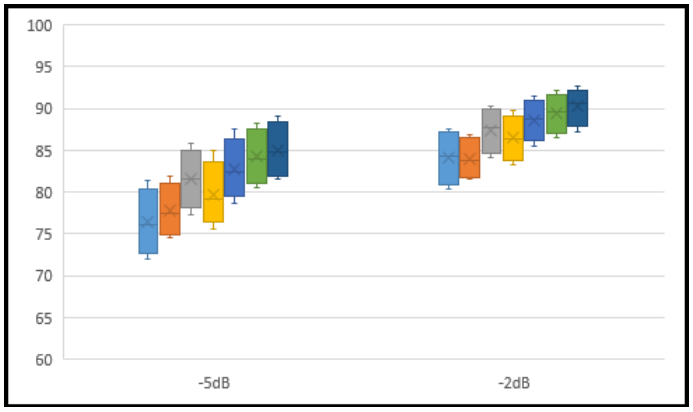

Figure 8. STOI (%) for Cafeteria noise of different datasets

FOREX Publication
Open Access | Rapid and quality publishing
International Journal of Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 322-328 | e-ISSN: 2347-470X

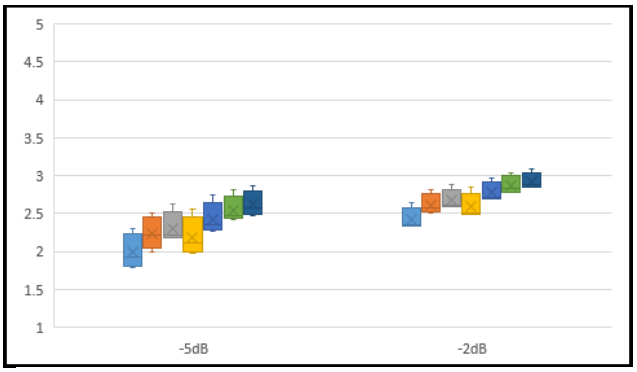

Figure 9. PESQ for Babble noise of different datasets

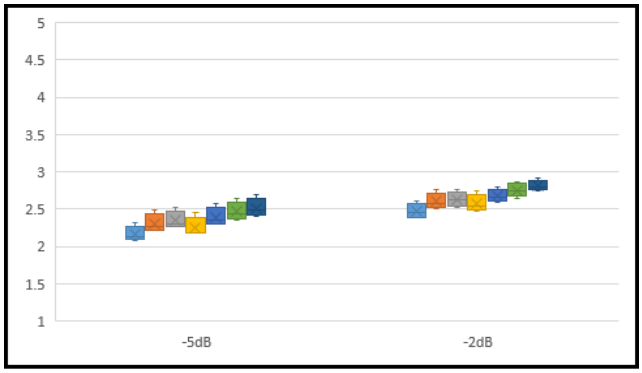

Figure 10. PESQ for Cafeteria noise of different datasets

Table 1: **Bi-LSTM Architecture hyperparameters**

Hyperparameter	Epoch	Interpolate method	Train data	Validation data	Test data	optimizer	Learning rate	Size of hidden layer
Bi-LSTM	80	Linear	60%	20%	20%	ADAM	0.001	128

Table 2: Performance analysis of STOI and PESQ of Current Algorithms vs. proposed Algorithm

Test noise		Babble								Cafeteria							
Test corpus		WSJ		TIMT		IEEE Male		IEEE Female		WSJ		TIMT		IEEE Male		IEEE Female	
Test SNR		-5dB	-2dB	-5dB	-2dB	-5dB	-2dB	-5dB	-2dB	-5dB	-2dB	-5dB	-2dB	-5dB	-2dB	-5dB	-2dB
STOI	**DCCRN**	82.5	89	73.1	82.5	68.3	81.3	72.5	84.6	81.4	87.6	74.8	82.6	72	80.3	77.4	86
	RNN-IRM	83.7	88.5	76.3	83.3	75.7	84.1	76	85.6	81.9	86.9	76.3	82.3	74.5	81.5	78.8	85.3
	RNN-TCS	88.1	92.2	79.3	87.5	76.7	85.8	80	89.2	85.8	90.3	80.4	86.6	77.3	84.1	82.6	88.7
	DCN	87.1	91.5	77.9	86.5	73.9	84.3	76.6	87.7	84.9	89.7	78.7	85.4	75.6	83.3	79.7	87.4
	DPARN	90.5	93.6	82.9	89.6	78.4	87.4	84.2	91.1	87.5	91.4	81.8	88	78.7	85.5	83.2	89.5
	ARN	91.1	94.1	83.9	90.6	82.3	88.9	85.6	92	88.3	92.1	82.7	88.6	80.6	86.6	85.3	90.5
	RNN-BI-LSTM	**91.7**	**94.8**	**84.3**	**91.1**	**83.1**	**90.0**	**87.1**	**92.7**	**89.1**	**92.6**	**83.1**	**90.1**	**81.5**	**87.2**	**86.4**	**91.1**
PESQ	**DCCRN**	2.31	2.65	1.99	2.38	1.86	2.33	1.79	2.33	2.32	2.61	2.12	2.39	2.08	2.4	2.14	2.5
	RNN-IRM	2.51	2.82	2.27	2.6	2.15	2.54	2	2.51	2.49	2.76	2.31	2.57	2.21	2.51	2.22	2.57
	RNN-TCS	2.63	2.89	2.22	2.59	2.2	2.59	2.18	2.62	2.52	2.76	2.26	2.53	2.27	2.59	2.34	2.65
	DCN	2.56	2.85	2.14	2.5	2.09	2.5	1.97	2.49	2.46	2.74	2.19	2.48	2.19	2.53	2.18	2.57
	DPARN	2.75	2.97	2.35	2.69	2.27	2.69	2.34	2.75	2.57	2.79	2.3	2.59	2.36	2.66	2.33	2.66
	ARN	2.82	3.04	2.43	2.78	2.45	2.79	2.48	2.86	2.64	2.87	2.36	2.65	2.43	2.73	2.45	2.76
	RNN-BI-LSTM	**2.87**	**3.09**	**2.48**	**2.85**	**2.59**	**2.85**	**2.55**	**2.91**	**2.69**	**2.91**	**2.40**	**2.75**	**2.48**	**2.79**	**2.49**	**2.80**

Table 1 shows the hyperparameters used for Bi LSTM architecture. *Table 2* shows the Babble and cafeteria noises with the decibel range from -5 to -2dB. By interfering with this, we can convey that the Recurrent Neural Network with Bi Grated Recurrent Unit was effective compared to that of other algorithms. The overall STOI value of the Bi LSTM algorithm comes around 84.3% to 94.8% whereas the other algorithm is in the range between 75.7 to 94.1%. By observing this, we can convey that RNN Bi LSTM is one of the effective algorithms for speech enhancement. These are the sound outputs wherein the first data is about clean speech, where a noise input is introduced, and then the noise overlaps with clean speech. By using different types of algorithms and comparing them with the RNN BI-LSTM algorithm we get the results accordingly and by comparing each result with the clean speech we can infer which algorithm has performed efficiently.

4.1. Discussions on the Results

Based on the experimental studies and findings, the following interferences were drawn:

1. Speech quality was well-preserved after the Bi directional LSTM and suppressed the noise level, with increased PESQ and STOI and reduced MSE.
2. The Proposed algorithm will offer better performance for various dataset corpus like WSJ, TIMIT, IEEE Male, and IEEE Female.
3. The Proposed algorithm will offer better performance for various noises like babble, cafeteria, etc., with different noise levels.

4. The bi-LSTM algorithm is more suitable for suppressing the following kinds of causal, non-causal, stationary, and non-stationary kinds of noises.
5. The main advantage of the proposed system is in producing stable results at all times with a minimum bit error rate.

5. CONCLUSION AND FUTURE WORKS

This paper suggests an innovative speech enhancement strategy by using the RNN-based Bi-LSTM model for enhancing the received signal quality in noisy and reverberant conditions. The RNN model learns the mapping functions between the input signal and the noise signal. The infused signal which is the combination of the clean and the noise signals is taken as input to the Bi-LSTM network. The RNN-Bi-LSTM model is designed to estimate and reconstruct the original waveform which is the clean signal. According to the experimental findings, the proposed models performed better in different contexts when compared to the existing models. The proposed model also showed robustness and adaptability. Hence, it could be decided that the RNN-Bi-LSTM framework shows impressive speech enhancement capability and brings in intelligibility and quality, likewise, we could perceive that the RNN-Bi-LSTM algorithm's effectiveness with the output sound after noise reduction is much closer to the clear noise. For the proposed and existing algorithms, simulations were conducted using a variety of sounds with various SNR input levels (-5dB, -2dB). The STOI and PESQ metrics demonstrate a rise of approximately 0.5% to 14.8% and 1.77% to 29.8%, respectively, in comparison with the existing algorithms over various sound types and different input SNR levels.

According to the findings, it could be observed that the proposed algorithm maintains consistency and also produces enhanced speech signals with minimal noise. In the future, in order to prevent hearing loss and improve speech quality, electronic safety hearing equipment for those who operate in hazardous sound-producing environments will be designed. Moreover, in the future, the complexity analysis of the proposed Bi-LSTM model will be performed, in cases of non-stationary noise in real-time applications. Similarly, by evaluating the performances under different other noise types with different SNR levels, the proposed model can be further enhanced to validate the effectiveness, noise cancellation, and clarity of the overall speech enhancement process.

REFERENCES

[1]. Loizou, P.C. Speech Enhancement: Theory and Practice; CRC Press: New York, NY, USA, 2013.

[2]. Xu, Yong, et al. "A regression approach to speech enhancement based on deep neural networks." IEEE/ACM Transactions on Audio, Speech, and Language Processing 23.1 (2014): 7-19.

[3]. Kumar, Anurag, and Dinei Florencio. "Speech enhancement in multiple-noise conditions using deep neural networks." arXiv preprint arXiv:1605.02427 (2016).

[4]. Park, Se Rim, and Jinwon Lee. "A fully convolutional neural network for speech enhancement." arXiv preprint arXiv:1609.07132 (2016).

[5]. Pandey, Ashutosh, and DeLiang Wang. "TCNN: Temporal convolutional neural network for real-time speech enhancement in the time domain." ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019.

[6]. Jannu, Chaitanya, and Sunny Dayal Vanambathina. "An Overview of Speech Enhancement Based on Deep Learning Techniques." International Journal of Image and Graphics (2023): 2550001.

[7]. Schroter, Hendrik, et al. "Low latency speech enhancement for hearing aids using deep filtering." IEEE/ACM Transactions on Audio, Speech, and Language Processing 30 (2022): 2716-2728.

[8]. Hepsiba, D., R. Vinotha, and L. D. Vijay Anand. "Speech Enhancement and Recognition Using Deep Learning Algorithms: A Review." Computational Vision and Bio-Inspired Computing: Proceedings of ICCVBIC 2022 (2023): 259-268.

[9]. Peracha, Fahad Khalil, et al. "Causal speech enhancement using dynamical-weighted loss and attention encoder-decoder recurrent neural network." Plos one 18.5 (2023): e0285629.

[10]. Huang, Ping, and Yafeng Wu. "Teacher-Student Training Approach Using an Adaptive Gain Mask for LSTM-Based Speech Enhancement in the Airborne Noise Environment." Chinese Journal of Electronics 32.4 (2023): 882-895.

[11]. Pandey, Ashutosh, and DeLiang Wang. "Self-attending RNN for speech enhancement to improve cross-corpus generalization." IEEE/ACM Transactions on Audio, Speech, and Language Processing 30 (2022): 1374-1385.

[12]. Shukla, Vasundhara, and Preety D. Swami. "Sparse Signal Recovery through Long Short-Term Memory Networks for Compressive Sensing-Based Speech Enhancement." Electronics 12.14 (2023): 3097.

[13]. Saleem, Nasir, et al. "Deepresgru: residual gated recurrent neural network-augmented kalman filtering for speech enhancement and recognition." Knowledge-Based Systems 238 (2022): 107914.

[14]. Yu, Meng, et al. "NeuralEcho: A self-attentive recurrent neural network for unified acoustic echo suppression and speech enhancement." arXiv preprint arXiv:2205.10401 (2022).

[15]. G. Vinothkumar and P. Phani Kumar Polasi “Filter performance of sparse noise for controlling the occurrence of noise-induced hearing loss using hybrid algorithm “ AIP Conference Proceedings 2405, 030013 (2022); https://doi.org/10.1063/5.0072454 Published Online: 05 April 2022.

[16]. Hasannezhad, Mojtaba, et al. "An integrated CNN-LSTM framework for complex ratio mask estimation in speech enhancement." 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, 2020.

[17]. Song, Zhendong, et al. "Hybrid dilated and recursive recurrent convolution network for time-domain speech enhancement." Applied Sciences 12.7 (2022): 3461.

[18]. Wang, Youming, et al. "Speech enhancement from fused features based on deep neural network and LSTM network." EURASIP Journal on Advances in Signal Processing 2021 (2021): 1-19.

[19]. Abdulbaqi, Jalal, et al. "Residual recurrent neural network for speech enhancement." ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020.

[20]. Vuong, Tyler, Yangyang Xia, and Richard M. Stern. "A modulation-domain loss for neural-network-based real-time speech enhancement." ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021.

[21]. Hasannezhad, Mojtaba, et al. "PACDNN: A phase-aware composite deep neural network for speech enhancement." Speech Communication 136 (2022): 1-13.

[22]. Cui, Xingyue, Zhe Chen, and Fuliang Yin. "Speech enhancement based on simple recurrent unit network." Applied Acoustics 157 (2020): 107019.

[23]. Abdulbaqi, Jalal, Yue Gu, and Ivan Marsic. "RHR-Net: A residual hourglass recurrent neural network for speech enhancement." arXiv preprint arXiv:1904.07294 (2019).

[24]. Peng, Kaibei, et al. "A Speech Enhancement Method Using Attention Mechanism and LSTM." 2021 3rd International Conference on Industrial Artificial Intelligence (IAI). IEEE, 2021.

[25]. Valin, Jean-Marc. "A hybrid DSP/deep learning approach to real-time full-band speech enhancement." 2018 IEEE 20th international workshop on multimedia signal processing (MMSP). IEEE, 2018.