International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Sound Noise Reduction Based on Deep Neural Networks

[*1]Bathini Pravalika, [2]P Preethi Grace, [3]Komuravelly Srimanya

[*1]Assistant Professor, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, Telangana, India

[*2,3]Student, Department of CSE, Bhoj Reddy Engineering College for Women, Hyderabad, Telangana, India

ARTICLEINFO

Article History:

Accepted: 01 June 2023
Published: 05 June 2023

Publication Issue
Volume 10, Issue 3
May-June-2023

Page Number
349-354

ABSTRACT

Audio transmittance is a generation that is now rapidly growing as a connectivity option for everyone around the world, demanding to experience the frictionless transfer of audio messages. Audio transmittance has a wide range of capabilities compared to other connectivity technologies. But we are living in the noisy world, hence while transmitting audio signal; we don't only transmit audio, different types of noise gets transmitted with our audio signal as well which will lead to an unclear communication The basic purpose of this model is specifically focused on detecting and restoring noisy audio signals which consists various background noise. The removal of noise from the audio signal will enhance the information carrying capacity of the signal during audio communication. For the removal of noise from audio signal, a stacked Long Short Term Memory (LSTM) model is proposed. 'Edinburgh DataShare' dataset has been used to train the model. During the evaluation of model, the Huber loss of 0.0205 has been evaluated in 50 epochs which shows that the LSTM network was successfully implemented for noise removal of audio signal. Hence on the basis of result, we can conclude that that Stacked LSTM network works well in noise removal of audio signals.

Keywords: De-Noising, Noise Removal Of Audio Signal, Long Short Term Memory (LSTM), Recurrent Neural Network.

I. INTRODUCTION

The modern society has reached a point where the daily life of every individual is unimaginable without electronic gadgets. From checking today's weather to staying touch with friends and families; the modern gadgets do it all. A modern man is surrounded by smart phone, smart watches, smart TVs, he's surrounded by possibilities. The use of these gadgets may be for the purpose of entertainment, education, health, advertisements yet they pose an important role in the life of every individual in today's world. Communication is the process of exchanging information from sender to receiver via different mediums like newspapers, magazines, different electronic gadgets, etc. We can transmit information on different forms like textual, graphical, audio, visual, etc. Audio communication is non- verbal

communication i.e. it is the type of framework based on hearing. In audio communication, recipient perceives information transmitted by the sender on the basis of what he/she have heard. Human can hear the sounds in a frequency range from about 20 Hz to 20 kHz[1]. During the transmission of audio signals, not only the original signal but certain noises like background noise (birds chirping, vehicles beep, wind, vacuum cleaner sound, etc.) are also transmitted to the receiver, which will lead to an uncertain communication. Receiver may either do not perceive any information or may perceive vague information from the noisy audio signals. In order to achieve a clear audio signal transmission, noise removal procedure must be implemented. Noise removal is the removal of noise from the audio signal.

Elimination of noise enhances the quality of the audio signal as de-noised audio signals carry more information than a noisy audio signal. De-noised audio signals enhance the listening and learning experience as we can concentrate de-noised audio than a noisy audio signal. Additionally, de-noised signals offers better communication where receiver can receive the desired information since the audio signal is not distorted by the noise. Deep learning is a method that uses multiple layers of neural networks to process and compute on a large amount of data which works based on ANN (inspired by the function and working of the human brain).Long Short Term Memory Networks (LSTMs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), etc. are some popular algorithms used for deep learning applications. LSTMs (Long Short-Term Memory) are a type of recurrent neural network commonly used for tasks such as text-to-speech or natural language processing. LSTM can update each time when new data is fed through the network hence are capable of learning long term dependencies. When multiple LSTM layers are used, they are called stacked LSTM which perform best in time series problems. So, LSTM is well suited for the removal of noisy audio signal which can provide better evaluation for de-noising a noisy audio signal.

II.RELATED WORK

A regression-based speech enhancement model using multiple deep neural networks (DNNs) concluded that the model achieved good improvement while evaluating via different objective measures. And during a subjective evaluation, this model was preferred by 76.35% to other conventional techniques [2]. The redundant Convolutional Encoder-Decoder (RCED) model demonstrated that using CNN with few parameters can give a better result and are applicable on embedded systems like hearing aids [3]. For the development of text-tospeech (TTS) system two approaches are used where first uses Mel cepstral (MCEP) coefficients, aperiodicity values and fundamental frequency to train TTS acoustic models and second approach we uses RNN to train the model using only the MCEP coefficients which was extracted from the magnitude spectrum. It was concluded that the second approach i.e. use of RNN provided better result than the first approach [4].LSTM model's performance was concluded to be the best in both the test set and the real case scenario to DNN, CNN, LSTM and BLSTM models. The noise reduction effects were reduced from 31.23% to 25.89% on the Xiaomi speakertest set during the demonstration [5].Hybrid model validates a noise suppression approach by combining DSPbased techniques with deep learning. Since it is hard to tune noise suppression using deep learning only, so they added DSP based techniques as well. They accept as true that this hybrid technique can be easily extended to residual echo suppression. Likewise, it can be used in microphone for post filtering purpose [6].Generative adversarial networks (GAN) operated to boost up the speed for de-noising waveform chunks. They have suggested further experiments to compare SEGAN with other competitive approaches [7].Variant U-Net architecture provides a better

performance to tackle a de-noising and de-reverberation problem with a singlestage framework. To achieve a real-time inference, an optimization strategy for U-Net significantly reduces the computational overhead up to 88.9\% compared to the naıve version [8]. A real-time speech enhancement approach based on a compact recurrent neural network was proposed where the model demonstrated how to control the amount of speech distortion with fixed-weighted and SNR (Signal to Noise Ratio)-weighted coefficients in the loss function. This paper explores various characteristics of training a RNN model and focuses on a RNN that enhances short-time speech on a single-frame-in, single-frame-out basis, a framework adopted by most classical signal processing methods [9]. Deep learning algorithms are able to suppress the noise from noisy audio signal but there is distortion on the speech too. A model has been proposed to address this issue by suppressing the noise in first phase using LSTM network and resorting original speech using convolutional encoder-decoder network. The model shows the improvement of PESQ by about 0.1 points [10].

III.PROPOSED SYSTEM

As a result, researchers created RNN variations that use gates to tackle this problem. Gates are operations that can learn how to add or delete information from a concealed state. Long Short-Term Memory is the neural network that employs these gates (LSTM). Two separation cores with two LSTM layers are followed by a fully-connected (FC) layer and sigmoid activation to provide a mask output in the stacked dual-signal transformation LSTM network design. The mask predicted by the FC layer and the sigmoid activation is multiplied by the magnitude of the mixture and translated back to the time domain using the phase of the input mixture without recreating the waveform in the first separation core. To produce the feature representation, a 1D-Convlayer processes the frames from the first network. The feature representation is processed by a normalization layer before it is fed to the second separation core. The unnormalized version of the feature representation is multiplied by the expected mask of the second core. The result is sent into a 1D-Conv layer, which converts the estimated representation back to the time domain. Finally, the signal is rebuilt using an overlap and addition process.

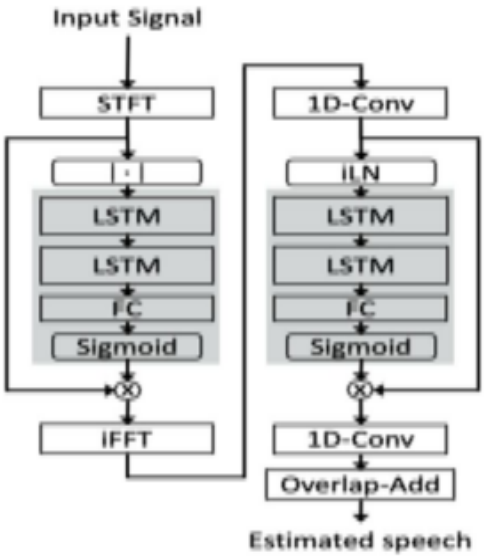

Fig 1 DTLN MODEL

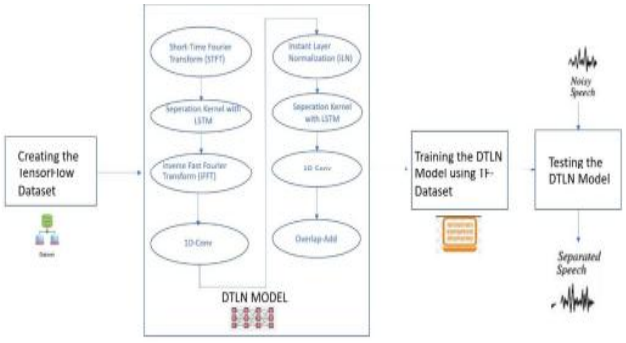

Fig 2 ARCHITECTURE

A. Create a Tensorflow Dataset

Class to create a Tensorflow dataset based on an iterator from a large scale audio dataset. This audio generator only supports single channel audio files.

i. It lists the data of the dataset and count the number of samples.

ii. Process the audio files

iii. create the dataset

B. DTLN Model

To create a mask output in the LSTM stacked double signal transform network architecture, two isolation cores with two LSTM layers are followed by a fully connected layer (FC) and sigmoid activation. The mix size is multiplied by the mask predicted by the FC layer and sigmoid activation, and the phase of the input mix is used to translate back to the time domain without recreating the waveform in the first separation kernel. A 1D-Conv layer processes frames from the first network to construct the feature representation. Before going to the second separation core, the feature representation is treated by a normalization layer. The predicted mask of the second kernel is multiplied by the unnormalized version of the feature representation. The estimated representation is then transmitted to a 1DConv layer, which transforms it back to the time domain. Finally, superimposition and addition are used to recreate the signal.

IV.RESULTS AND DISCUSSION

The experiments are run for 50 episodes with two LSTM layers and five time distributed dense layers. The learning rate is 0.001, dropout rate of 0.1. Adam optimizer is used for training the network. The activation function for all layers is taken as ReLU. Table 1 shows the parameters used to train the LSTM model.

Table 1: Model Parameters

SN	Parameter	Value
1	Epoch	50
2	Batch Size	64
3	Learning Rate	0.001
4	Dropout	0.1
5	Activation Function	ReLU
6	Optimizer	Adam

For the evaluation of the model, Huber loss evaluation measure is used. It is a combination of the mean squared error (MSE) function and the absolute value function which combines their properties. It is quadratic for small error and linear otherwise. It is identified by its delta parameters. Once the loss is less than or equal to delta, the quadratic function is used otherwise linear function is used for evaluating loss function. Huber loss is more robust to outliers than Mean Squared Error (MSE). Huber loss can be evaluated using the equation:

$$L_\delta = \begin{cases} \frac{1}{2}(y - f(x))^2, & \text{for } (y - f(x) \le \delta \\ \delta|y - f(x)| - \frac{1}{2}\delta^2, & \text{otherwise} \end{cases}$$

Figure 3: Model training and validation loss

As seen in the figure 3, over the epochs, it can be noted that the loss is decreasing rapidly at first few epochs and grows steady, but still decreases after that which shows that the model is learning well. The Huber loss is loss is 0.0205 which shows that the model has low validation loss. The spectrogram for input clean signal, noisy signal and predicted signal can be seen in figure below.

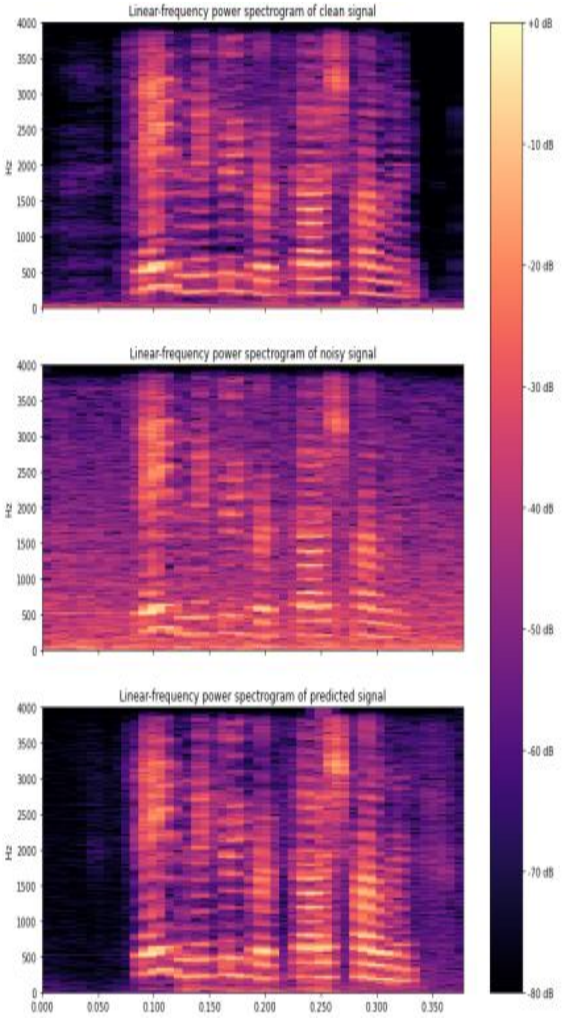

Figure 4 : Spectrogram for clean, noisy and predicted signal

II. CONCLUSION

This paper described a noise suppression method based on a stacked dual signal transformation LSTM network for real-time speech enhancement that had been trained on a large data set. In a stacked network approach, we were able to demonstrate the benefit of using two types of analysis and synthesis bases.

III. REFERENCES

[1]. S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactions on acoustics, speech, and signal processing, vol. 27, no. 2, pp. 113–120, 1979.

[2]. H. G. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition," in Proc. ICASSP, 1995, vol.1, pp. 153–156.

[3]. T. Gerkmann and R.C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 4, pp. 1383–1393, 2012.

[4]. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 33, no. 2, pp. 443–445, 1985.

[5]. A. Maas, Q.V. Le, T.M. O'Neil, O. Vinyals, P. Nguyen, and A.Y. Ng, "Recurrent neural networks for noise reduction in robust ASR," in Proc.INTERSPEECH, 2012.

[6]. D. Liu, P. Smaragdis, and M. Kim, "Experiments on deep learning for speech denoising," in Proc. Fifteenth Annual Conference of the International Speech Communication Association, 2014.

[7]. Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "A regression approach to speech enhancement based on deep neural networks," IEEE Transactions on Audio, Speech and Language Processing, vol. 23, no. 1, pp. 7–19, 2015.

[8]. A. Narayanan and D. Wang, "Ideal ratio mask estimation using deep neural networks for robust speech recognition," in Proc. ICASSP, 2013, pp. 7092–7096.

[9]. S. Mirsamadi and I. Tashev, "Causal speech enhancement combining data-driven learning and suppression rule estimation.," in Proc. INTERSPEECH, 2016, pp. 2870–2874.

[10]. https://grubbrr.com/the-importance-of-backgroundnoise-suppression-in-ai/

[11]. https://developer.nvidia.com/blog/nvidia-real-timenoise-suppression-deep-learning/

[12]. Jean-Marc Valin, " A Hybrid DSP / Deep Learning Approach to Real-Time Full-Band

Speech Enhancement," arXiv:1709.08243v3 [cs.SD] 31 May 2018

[13]. Urmila Shrawankar and Vilas Thakare, "Noise Estimation and Noise Removal Techniques for Speech Recognition in Adverse Environment,".

Cite this article as :

B Pravallika, P Preethi Grace, Komuravelly Srimanya, "Inspection of Transmission Line with Robot", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 10 Issue 3, pp. 349-354, May-June 2023.
Journal URL : https://ijsrset.com/IJSRSET23103118