Speech quality frequency sharpener using Wavelet Transforms

G.Shiva Prasad Reddy, studying ECE at IARE, Hyderabad, INDIA, shivaprasadreddy390@gmail.com

Dr.S.China Venkateswarlu[2], Professor of ECE, IARE, Hyderabad, INDIA. c.venkateswarlu@iare.ac.in

Dr.V.Padmabha Reddy[3], Professor of ECE, IARE, Hyderabad, INDIA. C.venkateswarlu@iare.ac.in

Abstract:

Deep neural networks show significant improvement in enhancing the speech.It consists of multiple layers between input and output, it's function is similar to human neurons where the linkage is perfect within the structure .The paper describes the new method for enhancing the speech developed by convolutional and residual neural networks architecture and explores two criteria for loss function optimization: weighted and uniform progressive. Resnet introduce the skip connection which are built out of something called a residual blocks with this resnets we can minimize the vanishing gradient in signal processing with the using of this we can minimize the progressive loss function. This work carries out the evaluation on simulated and real speech samples with reverberation and added noise using REVERB and Voice Home datasets with this approach we can enhance the information up to 82 percent where is a efficient method for progressive enhancement

Keywords: Progressive loss function, Speech enhancement, ResNet, CNN

INTRODUCTION:

Speech is the only and the most basic form to communicate and to express. In most noisy areas or noisy environments , the speech signal is combined with noisy signals sending energy at the same time, which might be result into noise or distinct speech signals which is difficult to analyze. As a result, it is difficult to increase the quality of speech . Speech enhancement (SE) techniques have been used in communications systems, such as mobile communication and speech recognition, as well as hearing aids. The primary goal of SE algorithms is to enhance some distinguished characteristics of speech that have been distorted by noise. SE algorithms are used in hearing aids to clean the noise in the signal before enhancing by minimizing background noise, since hearing-impaired people have great difficulty communicating in surroundings with various types of noise In many cases, lowering background noise causes speech distortion, which affects speech intelligibility in loud circumstances. It is a subjective performance evaluation statistic since it reflects the individual tastes of listeners. The intelligibility is an objective metric since it provides the percentage of words that listeners can properly identify. Based on these two requirements, the significant difficulty in building an efficient SE algorithm for hearing aids is to improve overall speech quality and intelligibility by decreasing noise while introducing no apparent signal distortion. Many ideas and methods to SE have been researched and presented in recent years.

Weiss et al. pioneered spectrum subtraction (SS) methods in the correlation domain while L. Chen et al. proposed the spectral subtraction methodology in current hearing aids for real-time speech improvement. The method relies on a vocal activity detector (VAD) to estimate the noise spectrum when speech pauses (in quiet) and remove it from the noisy speech to calculate the clean speech .In each frame, the SS methods regularly generate a new sort of noise at random frequency positions. This sort of noise is known as musical noise, and it can be more unpleasant not only to the human ear but also to SE systems than the initial distortions.Harbach in employed a directional microphone and the Wiener filter approach based on previous SNR calculation to enhance voice quality. However, the all-pole spectrum of the voice signal may have abnormally strong peaks, resulting in a considerable reduction in speech quality. In the Minimum Mean Square Error (MMSE) method based on log-magnitude was proposed.The method works by minimizing the mean square error (MSE) of the log-magnitude spectra to find the coefficient. The results of this approach's tests revealed decreased levels of residual noise. Meanwhile, deep neural network (DNN) techniques have shown tremendous potential and interest in tackling SE issues .L. Ding et al. employed a DNN model for speech denoising,

for example. When faced with noisy speech inputs, the model predicts clean speech spectra without the need for RBM pre-training or sophisticated recurrent structures. To increase voice quality, X. Lu et al. introduced a regression model using the denoising autoencoder (DAE). S. Meng et al. developed a distinct deep autoencoder (SDAE) technique in that estimates the noisy and clean spectra by minimizing the overall reconstruction error of the noisy speech spectrum by changing the predicted clean speech signal. Lai et al. proposed employing a deep denoising autoencoder networks (DDAE) model in cochlear implant (CI) simulations to improve vocoded speech intelligibility.

The paper contains usage of two methodologies of DNN that is usage of CNN and RESNET topologies the architecture kept a constant number of channels along all the blocks of the DNN. The constant number of channels allowed the output reconstruction and a visualization probe at any internal block. The mandatory progressive signal reconstruction forced an incremental process of the SE that tended to improve the robustness of the model. Besides, this architecture uses a weighted composition of reconstruction errors by block to perform the loss function optimization. This way, each block makes partial reconstruction, and the next block has as input a previously enhanced representation of the signal.

1.2 LITERATURE SURVEY

Ying,L.H.-2018:

Speech is the only and the most basic form to communicate and to express. In most noisy areas or noisy environments , the speech signal is combined with noisy signals sending energy at the same time, which might be result into noise or distinct speech signals which is difficult to analyze. As a result, it is difficult to increase speech quality and makes worst speech intelligibility. Speech enhancement (SE) techniques have been used in high-tech communications systems, such as mobile communication and speech recognition, as well as hearing aids. The main aim of speech enhancement algorithms is to increase/enhance some characteristics of speech that have been destructed by noisy environment . SE algorithms are used in hearing aids (HA) to remove the noisy signal before amplification, since hearing-impaired people have great difficulty communicating in surroundings with various degrees and types of noise.

WHO-2019:

In many cases, lowering background noise causes speech distortion, which affects speech intelligibility in loud circumstances.[2]

According to a WHO estimate, more than 5% of the world's population – or 430 million individuals are noticed that facing hearing loss problems (432 million adults and 34 million children). Also, it is came to know that by 2050 nearly 700 million individuals – or twenty in every hundred people – would have to face hearing loss. Hearing loss that is 'disabling' is defined as hearing loss that is higher than 35 decibels (dB) in the better hearing ear. Almost 80% of patients with severe hearing loss in most of the old people . Hearing loss becomes more common as people get older; more than a quarter of persons over the age of 60 have high hearing loss.

Shifas, M.-2020

Convolutional neural network (CNN) modules are commonly utilised in the development of high-end voice enhancing neural models. However, the dimensionality limitation of the integrated convolutional kernels has reduced the feature extraction capacity of vanilla CNN modules, resulting in a failure to effectively simulate the noisy context information at the feature extraction step. Though they were resistant against a class of sounds whose spectral distribution can be completely represented by second order statistics, their performance against more structurally dispersed noises was not sufficient..

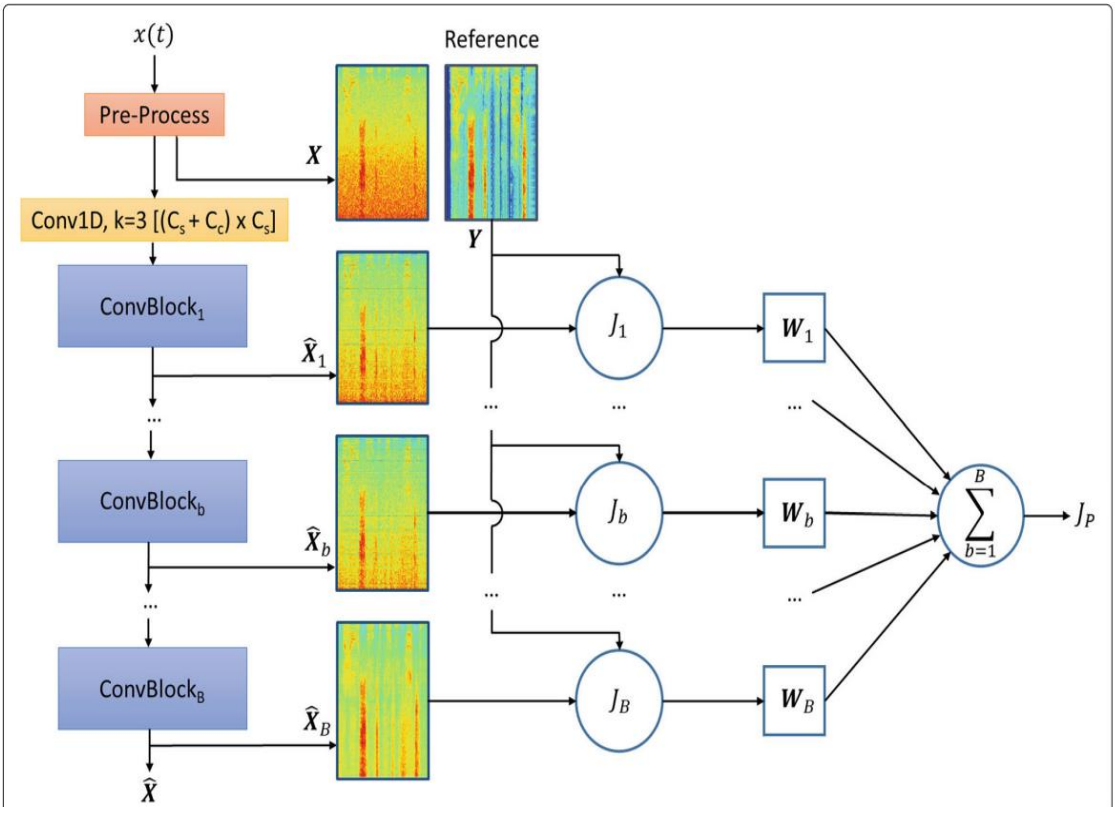
Lai, Y.-H.-2019:

Based on the findings of several objective-based SE methods, it is confirmed that deep learning-based SE is a very useful and better strategy for improving speech quality which was affected by noise for hearing aid users. Furthermore, multi-objective learning can increase the performance of the DDAE technique for the majority of hearing aid users. More particularly, the suggested SE approach (M-DDAE) has higher HASQI and HASPI scores.

2.1 EXISTING METHOD

Deep neural networks show significant improvement in enhancing the speech.It consists of multiple layers between input and output, it's function is similar to human neurons where the linkage is perfect within the structure .The paper describes the new method for enhancing the speech developed by convolutional and residual neural networks architecture and explores two criteria for loss function optimization: weighted and uniform progressive. Resnet introduce the skip connection which are built out of something called a residual blocks with this resnets we can minimize the vanishing gradient in signal processing with the using of this we can minimize the progressive loss function. This work carries out the evaluation on simulated and real speech samples with reverberation and added noise using REVERB and Voice Home datasets with this approach we can enhance the information up to 82 percent where is a efficient method for progressive enhancement .

2.1.1 BLOCK DIAGRAM



The input provided to the CNN, ResNet, P-CNN, and P-ResNet architectures consists of the logarithm of the magnitude of the 512-STFT of the corrupted signal, sampled at 16 kHz. The STFT is computed every 10 ms for a 25 ms sliding Hamming window. We also concatenate the Mel-Scaled Filter-bank and the MFCC as auxiliary inputs, with filter bank sizes 32, 50, and 100, every 10 ms. MFCC are computed using the discrete cosine transform (DCT)

without truncation. However, each frequency resolution has a different sliding Hamming window of 25 ms, 50 ms, and 75 ms respectively. These auxiliary features provide different frequency and temporal resolutions, which can benefit the speech enhancement process . Taking into account that the LSA dimension is 512, the overall input size is 876

2.2 PROBLEM IDENTIFICATION:

Using CNN and RESnet will increase the complexity and also need to deal with numerous calculations to process it takes more time to enhance the speech **.** The receiver of an in-the-ear hearing aid can get clogged with earwax and moisture. "There is also an occlusion effect with ITEs. Patients with mild hearing loss may not choose an ITE because the whole ear is plugged up. Everything is altered.

2.3 PROPOSED METHOD:

Speech Enhancement :

Speech enhancement main aim is to improve speech quality and decrease the noise which is added to the speech by using various algorithms and techniques . The objective of enhancement is improvement speech quality and/or overall perceptual quality of degraded speech signal using audio signal processing techniques.

Enhancing of speech degraded by noise is the most critical point of speech enhancement, and used for many applications such as mobile phones, VoIP, teleconferencing systems, speech recognition, and hearing aids

Algorithm:

1.We given an input signal to the matlab model

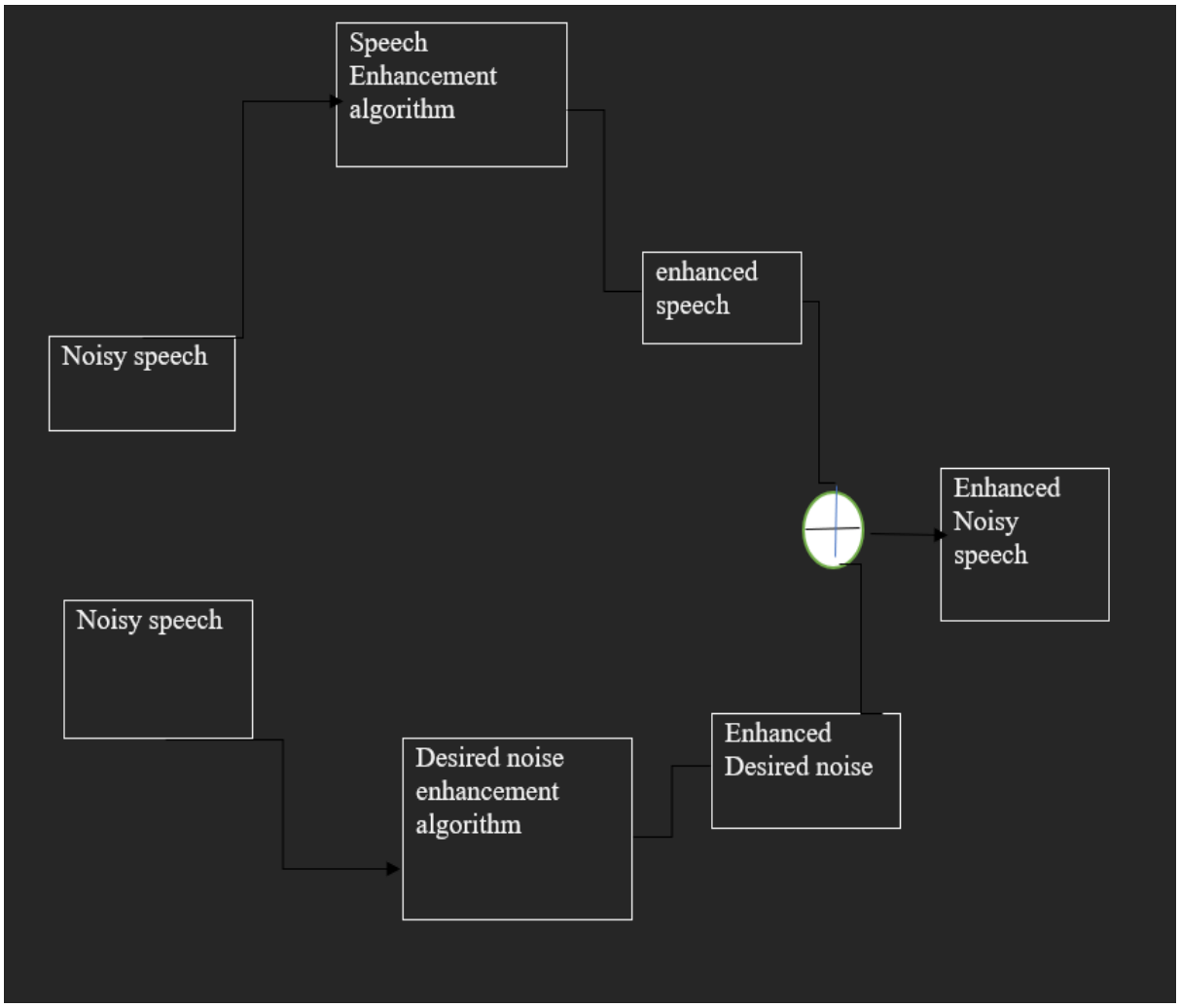
2.Then we add noise to the input signal because for this system the input signal is a clean signal,some noise is added to create a noisy signal and then to construct the original signal

3.We use wavelet filter to reduce the added noise to the system

4.And then we use frequency shifter to adjust the loss frequencies

5.In final the amplitude compression is used to improve the gain of the signal

2.3.1 BLOCK DIAGRAM:



In this we are going to add a noise which are of any kind like white,gaussian,pink etcc to the plain speech and with that the speech becomes noisy speech then we enhance the noisy speech signal and then we use wavelet filter to reduce the noise and use frequency adder to obtain the lost frequency of the signal thus how we gonna improve the speech in our system

3.1 SOFTWARE USED:

The project is done using Matlab of r2021a version

Matlab r2021a:

1.Matlab is a programming platform designed specifically for students and scientist to analyze and design system products

2.Matlab application is built around matlab based programming language

3. Common usage of the MATLAB application involves using the "Command Window" as an interactive mathematical shell or executing text files containing MATLAB code.

3.1.1 PRACTICAL SETUP:

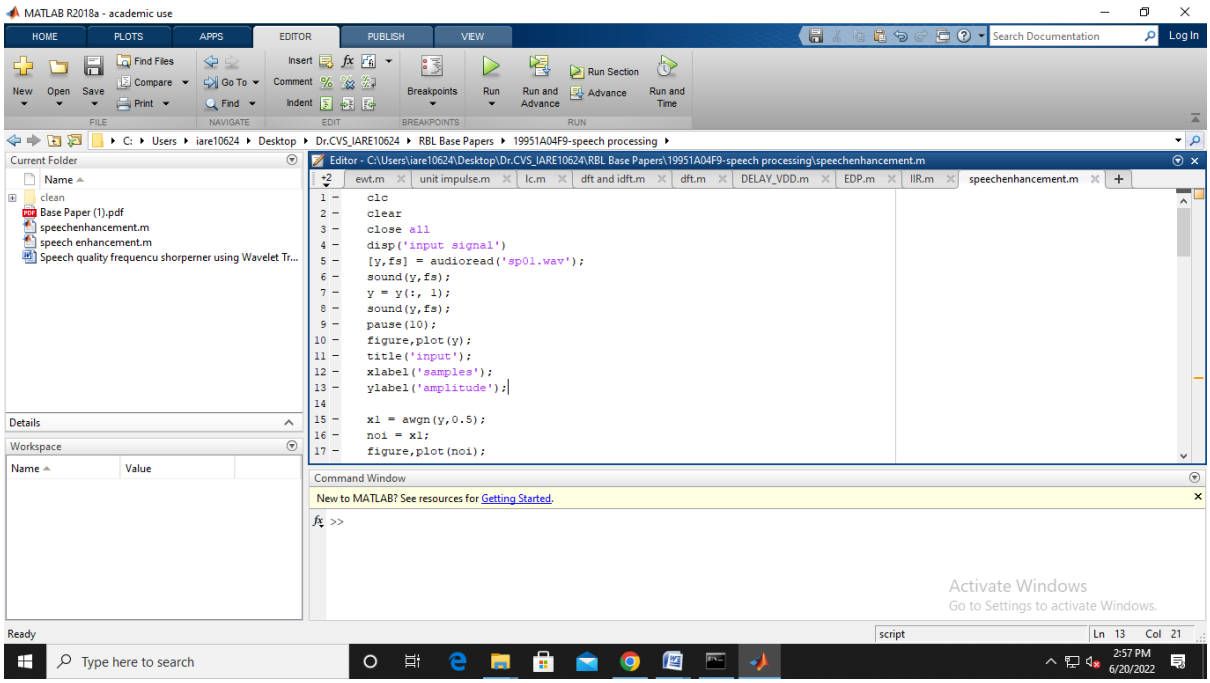


Figure 3.1: Practical setup using editor window in matlab

The above image represents the practical setup of the matlab. In the above image the editor window shows the code we are using for the speech enhancement and the left side of the image shows the files we are going to use in the code.

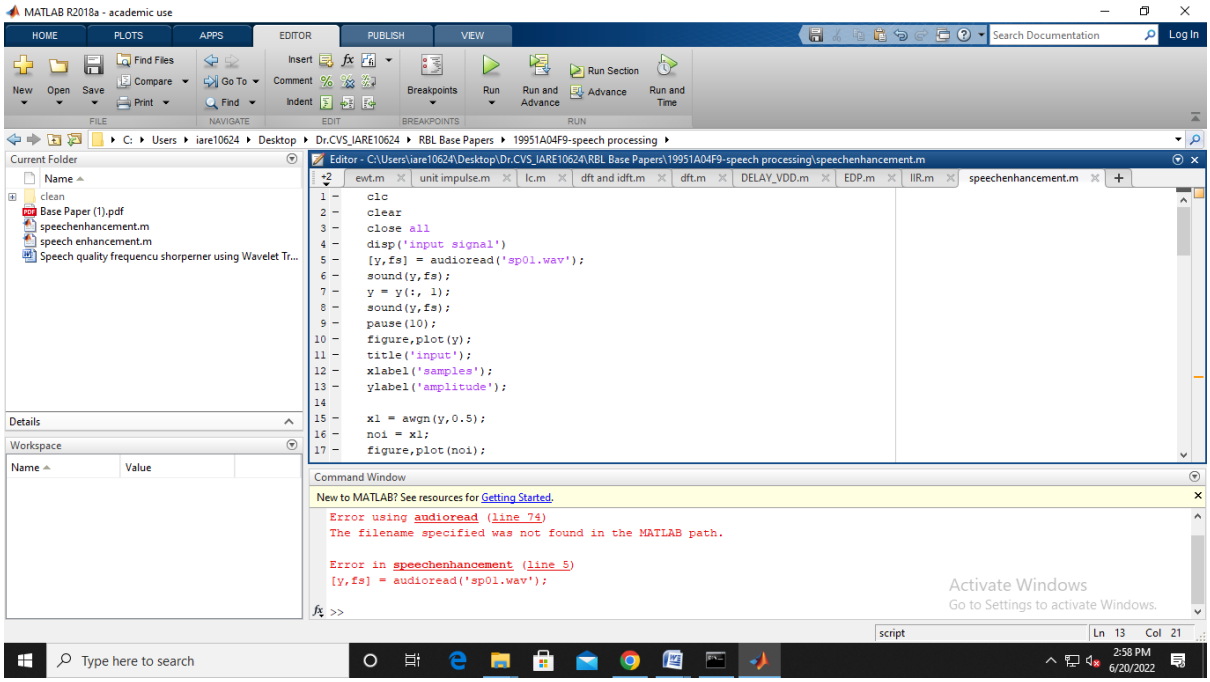


Figure 3.2: Matlab reading the given input

In the above image it represents that the code contains an audio input file which was taken from the files located at the left side of the image.

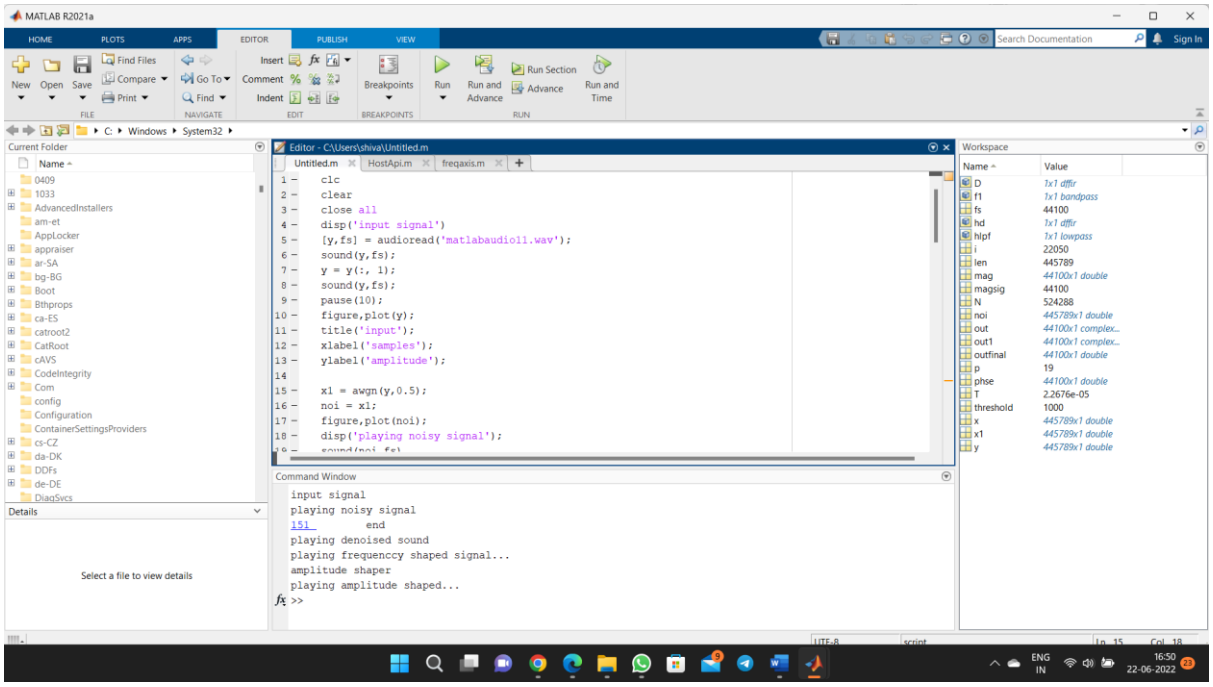


Figure 3.3: Matlab showing the outputs

The above image is about showing the results of the output in the command window and the calculations done in the workspace

4.1 RESULTS AND DISCUSSIONS:

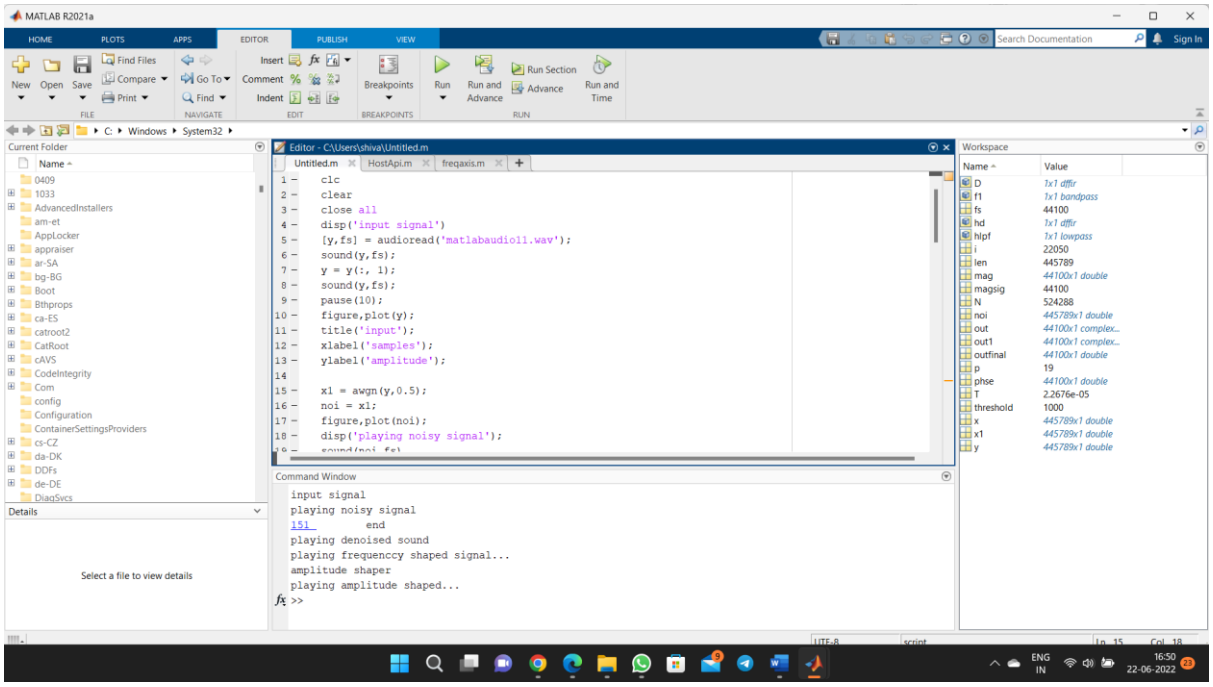


Figure 4.1: Simulation process started

The above image represents after entering the code in the editor and installing the required libraries we run or execute the code. At the bottom of the image a dialog box has opened which indicates the execution of the code and in the terminal section we can observe the output

4.1.1 CASE 1:

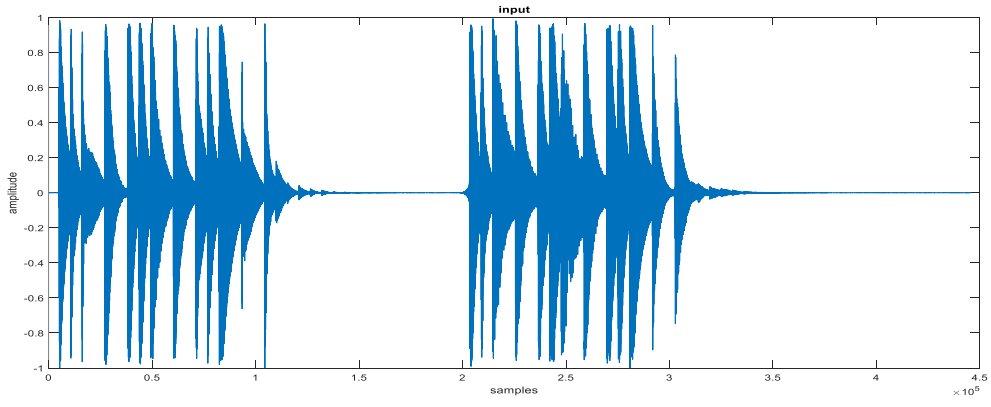


Figure 4.2: Input signal

The figure describes the waveform of the input signal which was read through audioread command in the matlab program. It is a clean speech signal

4.1.2 CASE 2:

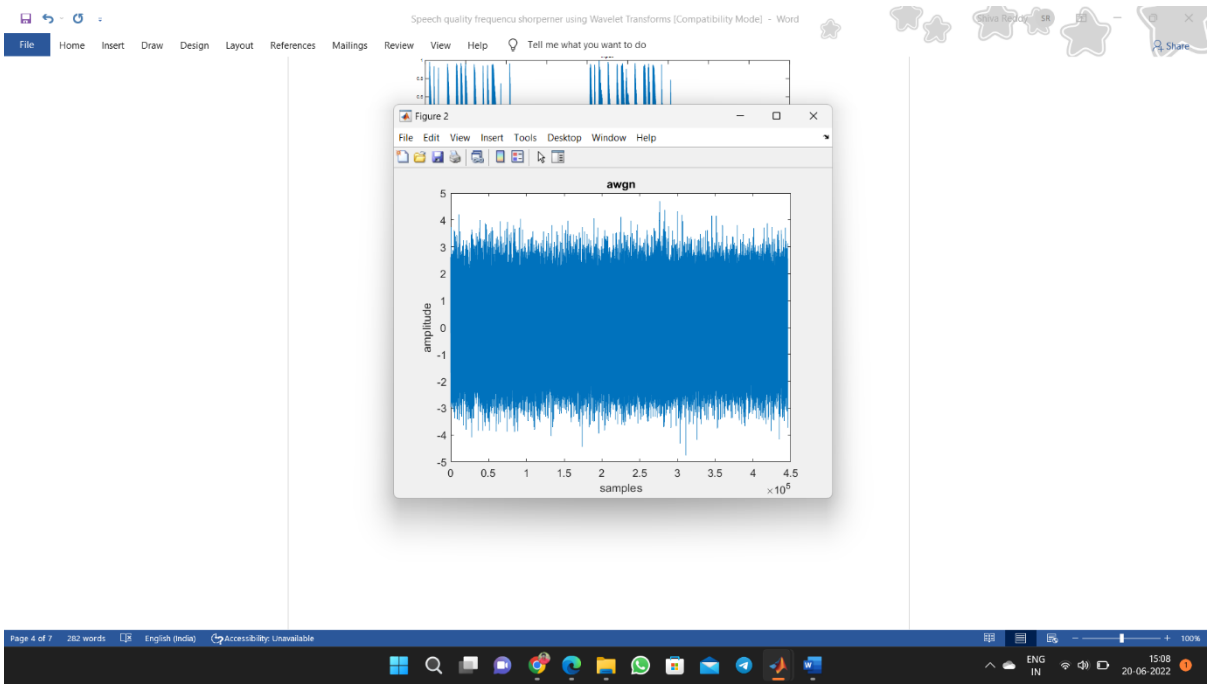


Figure 4.2: AWGN

The figure is about the noise of white gaussian noise which was added to the clean speech signal to form a noisy speech signal which is unclear to listen

4.1.3 CASE 3:

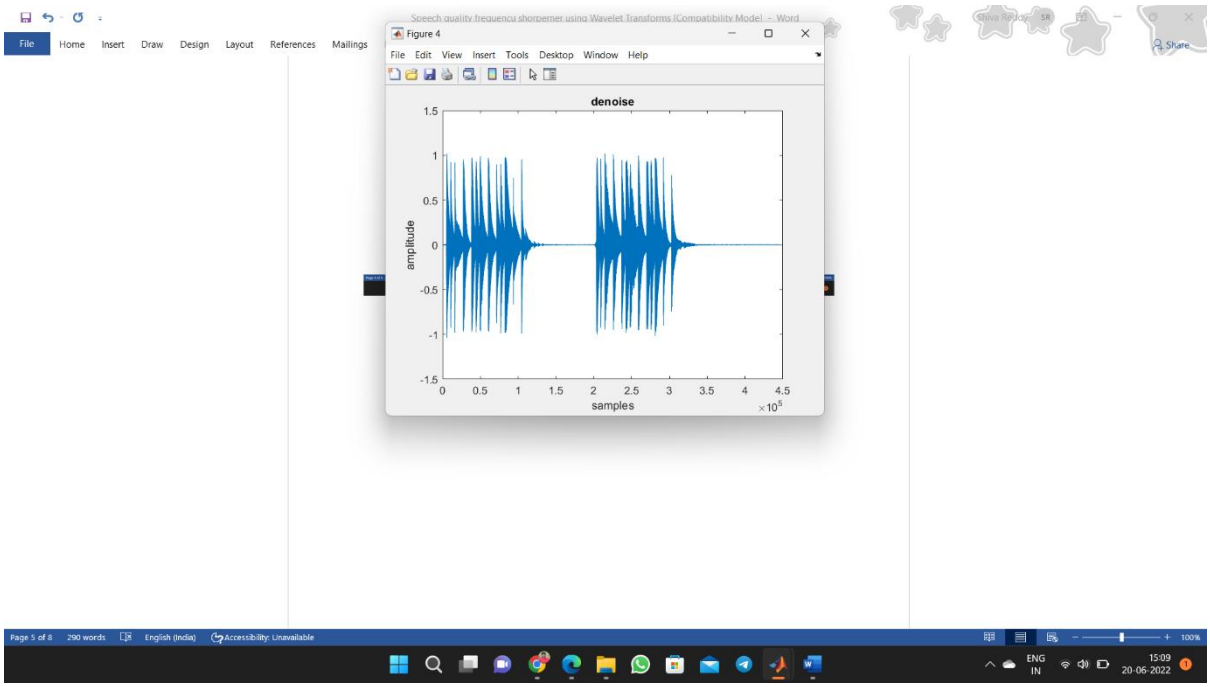


Figure 4.3: Denoise signal

The above figure is the result of the denoised signal where the added noise(AWGN) was removed from the noisy speech signal to get a denoised signal(clean speech signal) but this loss some of its frequency components while denoising.

4.1.4 CASE 4:

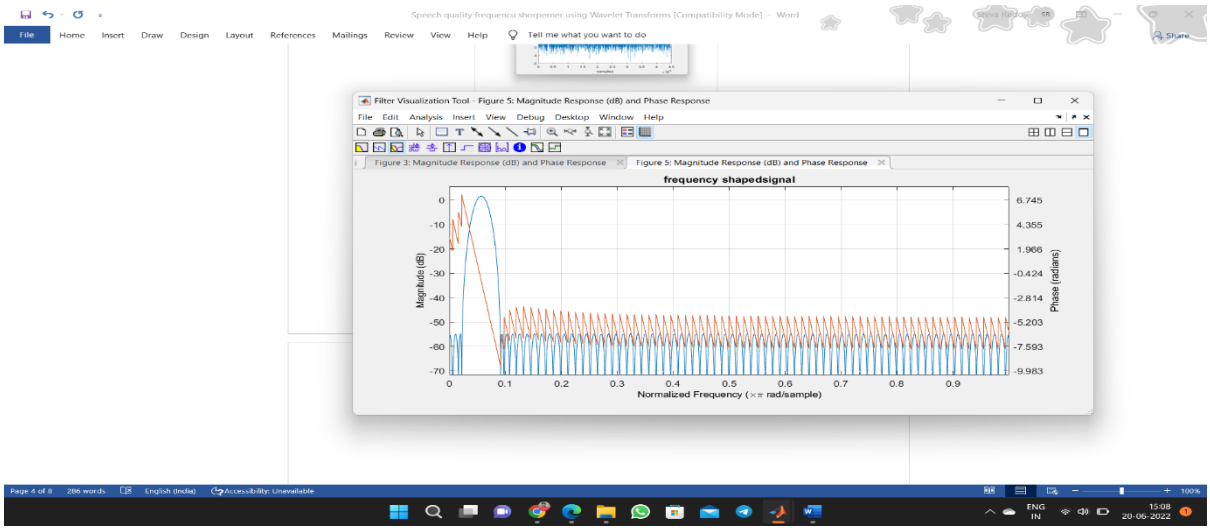


Figure 4.3: Frequency shaped signal

The figure is about the magnitude and phase response of the enhanced denoised signal. Where in this the lost frequency components of the denoised signal will be added back to it and thus its also increases the strength and magnitude of the signal.

4.1.5 CASE 5:

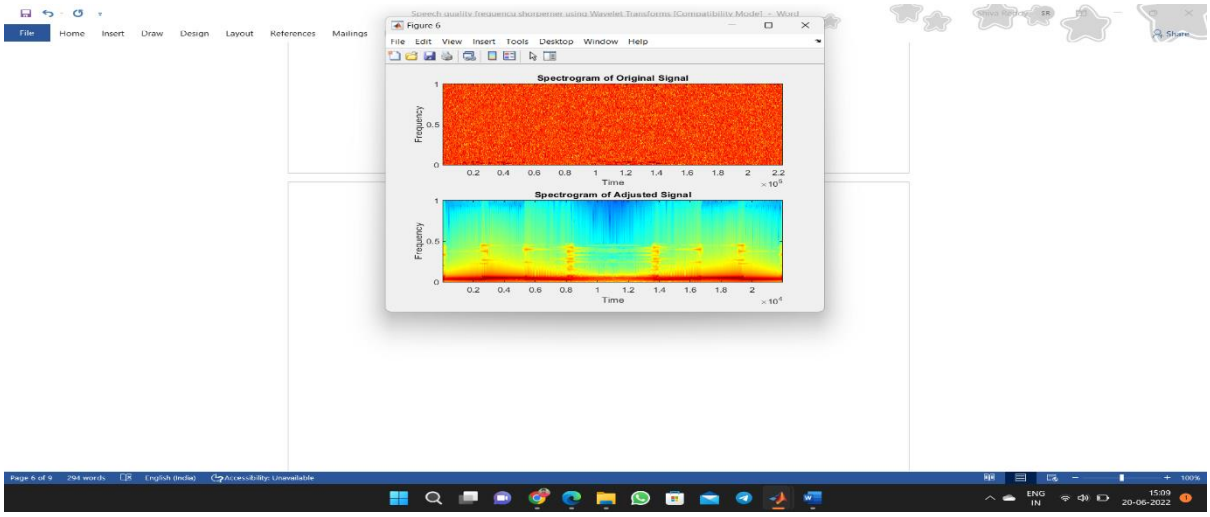


Figure 4.4: Spectrogram

Spectrogram is a visual way of representing the signal strength or “loudness” of a signal over time at various frequencies present in a particular waveform

5. CONCLUSION:

In this section we are discussing about the results we acquired during our process. In this algorithm we are doing to enhance the speech which is corrupted by noise so inorder to do obtain enhanced speech we are passing a normal speech and to that we are adding some noise like (bubble,pink,guassian etc) so in this case we are using additive white gaussian noise noise (awgn) which is a fundamental model used in data accumulation. Awgn is a continuous and uniform frequency spectrum over a specified band and has a equal power per hertz.Thus when we add the awgn noise to the normal speech then we obtain noisy speech in disturbed form then we let to pass this noisy speech to a filter to reduce the noise so in this we are using wavelet filter which is a command allows us to enhance or decrease the details in a certain spatial frequency domain and the wavelet transform translates the time amplitude of a signal in toa frequency representation of a signal and these wavelet coefficients help us to find the noise pattern in the signal. Now we use frequency shaper which helps us to restore the frequencies which were lost during reducing the noise and tries to improve the frequency to its original shape this filter is truly designed by ourselves and then at last the amplitude compression is used to improve the gain of the signal.

5.1 FUTURE SCOPE:

Future work will further study the progressive strategy on several deep de-noising autoencoder networks, each working on a small specific enhancement task and learning to handle a subset of the entire training set. To assess the execution of the approach, the hearing-aid speech perception list, the hearing aid sound quality record, and the perceptual evaluation of speech quality were utilized. Seven patients with sensor neural hearing loss audiograms were used to assess improvements in speech quality and intelligibility. We analyzed the suggested method's performance to that of individual de-noising autoencoder networks with three and five hidden layers. When compared to three and five layers i.e., existing method, the experimental findings demonstrated that the proposed technique produced superior quality and was more comprehensible. The proposed approach for improving the quality and intelligibility of speech is based on PESQ, HASQ, HASPI, segmental SNR, And also hearing aid sound quality index, speech quality, Intelligibility additional DNN architectures such as U-Net and GAN. We will also assess the performance of 2D-convolutions, as the core of convolutional blocks, and compare them with 1D-convolutions.

6 REFERENCES:

1.Lai, Y.-H.; Chen, F.; Wang, S.-S.; Lu, X.; Tsao, Y.; Lee, C.-H. A Deep Denoising Autoencoder Approach to Improving the Intelligibility of Vocoded Speech in Cochlear Implant Simulation. IEEE Trans. Biomed. Eng. 2017, 64, 1568–1578. [CrossRef] [PubMed]

2. Lai, Y.-H.; Tsao, Y.; Lu, X.; Chen, F.; Su, Y.-T.; Chen, K.-C.; Chen, Y.-H.; Chen, L.-C.; Li, L.P.-H.; Lee, C.-H. Deep Learning–Based Noise Reduction Approach to Improve Speech Intelligibility for Cochlear Implant Recipients. Ear Hear. 2018, 39, 795–809. [CrossRef] [PubMed]

3. Lai, Y.-H.; Zheng, W.-Z. Multi-objective learning based speech enhancement method to increase speech quality and intelligibility for hearing aid device users. Biomed. Signal Process. Control. 2019, 48, 35–45. [CrossRef]

4. Kim, M. Collaborative Deep Learning for Speech Enhancement: A Run-time Model Selection Method Using Autoencoders. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017.

7 AUTHOR BIOGRAPHY:

NAME: Gangasani Shiva Prasad Reddy

Roll No: 19951A04F9

DEPARTMENT: ECE

STUDENT OF INSTITUTE OF AERONAUTICAL ENGINEERING

EMAIL ID: shivaprasadreddy390@gmail.com

MOBILE NUMBER: 6304219989

NAME : DR. V PADMANABHA REDDY

FACULTY ID : IARE10622

PROFESSOR AT INSTITUTE OF AERONAUTICAL ENGINEERING

MAIL ID : v.padmanabhareddy@iare.ac.in

NAME : DR. S CHINA VENKATESWARLU

FACULTY ID : IARE10624

PROFESSOR AT INSTITUTE OF AERONAUTICAL ENGINEERING

MAIL ID : c.venkateswarlu@iare.ac.in

8 APPENDIX:

```
clc
clear
close all
disp('input signal')
[y,fs] = audioread('matlabaudio11.wav');
sound(y,fs);
y = y(:, 1);
sound(y,fs);
pause(10);
figure,plot(y);
title('input');
xlabel('samples');
ylabel('amplitude');

x1 = awgn(y,0.5);
noi = x1;
figure,plot(noi);
disp('playing noisy signal');
sound(noi,fs)
pause(10)
xlabel('samples');
ylabel('amplitude');
title('awgn');
pause(10)
%'Fp,Fst,Ap,Ast' (passband frequency, stopband frequency,
passband ripple, stopband attenuation)
hlpf =
fdesign.lowpass('Fp,Fst,Ap,Ast',10.0e3,10.5e3,0.5,50,fs);
D = design(hlpf);
freqz(D);
x = filter(D,y);
disp('playing denoised sound');
figure,plot(x);
title('denoise');
sound(x,fs);
xlabel('samples');
ylabel('amplitude');
pause(10)
% freq shaper using band pass
T = 1/fs;
len = length(x);
p = log2(len);
p = ceil(p);
N = 2^p;
f1 =
fdesign.bandpass('Fst1,Fp1,Fp2,Fst2,Ast1,Ap,Ast2',1000,2000,30
00,8000,60,2,60,2*fs);
hd = design(f1,'equiripple');
```

```
y = filter(hd,x);
freqz(hd);
y = y*100;
disp('playing frequenccy shaped signal...');
title('frequency shapedsignal');
sound(y,fs);
pause(10);
% amplitude shaper
disp('amplitude shaper')
out1=fft(y,fs);
phse=angle(out1);
mag=abs(out1)/N;
[magsig,~]=size(mag);
threshold=1000;
out=zeros(magsig,1);
for i=1:magsig/2

   if(mag(i)>threshold)
       mag(i)=threshold;mag(magsig-i)=threshold;

   end

   out(i)=mag(i)*exp(j*phse(i));
   out(magsig-i)=out(i);

end

outfinal=real(ifft(out))*10000;
disp('playing amplitude shaped...');
sound(outfinal,fs);
pause(10);

% load handel.mat
figure;
subplot(2,1,1);
specgram(noi);
title('Spectrogram of Original Signal');

subplot(2,1,2);
specgram(outfinal);
title('Spectrogram of Adjusted Signal');
```