Speech Enhancement Using Machine learning

CHINNI SARAT BHARGAVA

Problem Statement

- Speech Enhancement is the task of removing the noise present in the speech signal and maintaining the intelligibility.
- Signal Processing Based Approaches
  - Wiener Filter
  - Spectral Subtraction
- In this project we try to solve this problem using Neural Networks

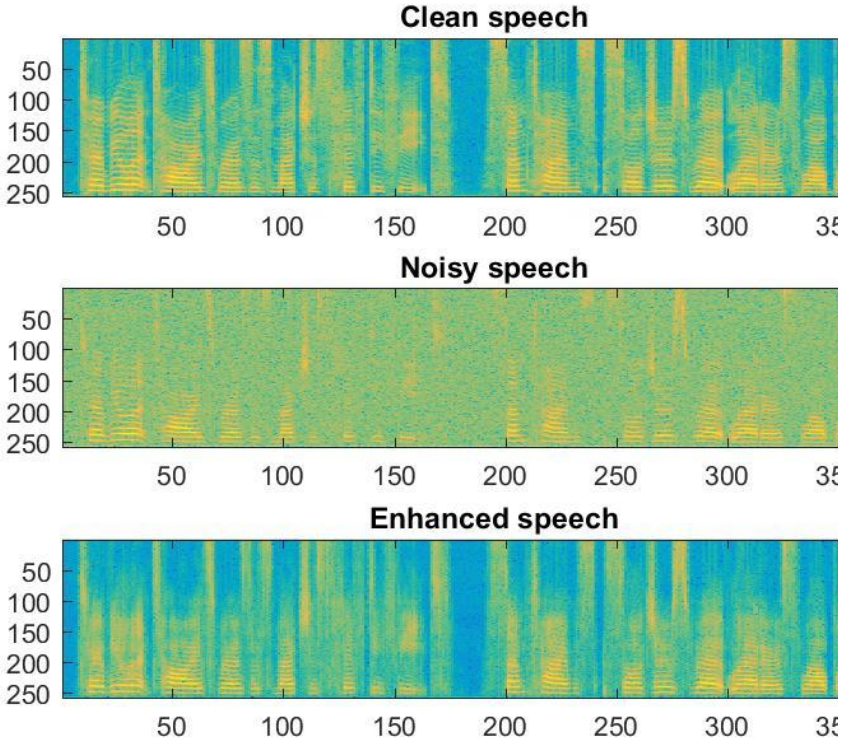
Data Set and Features used

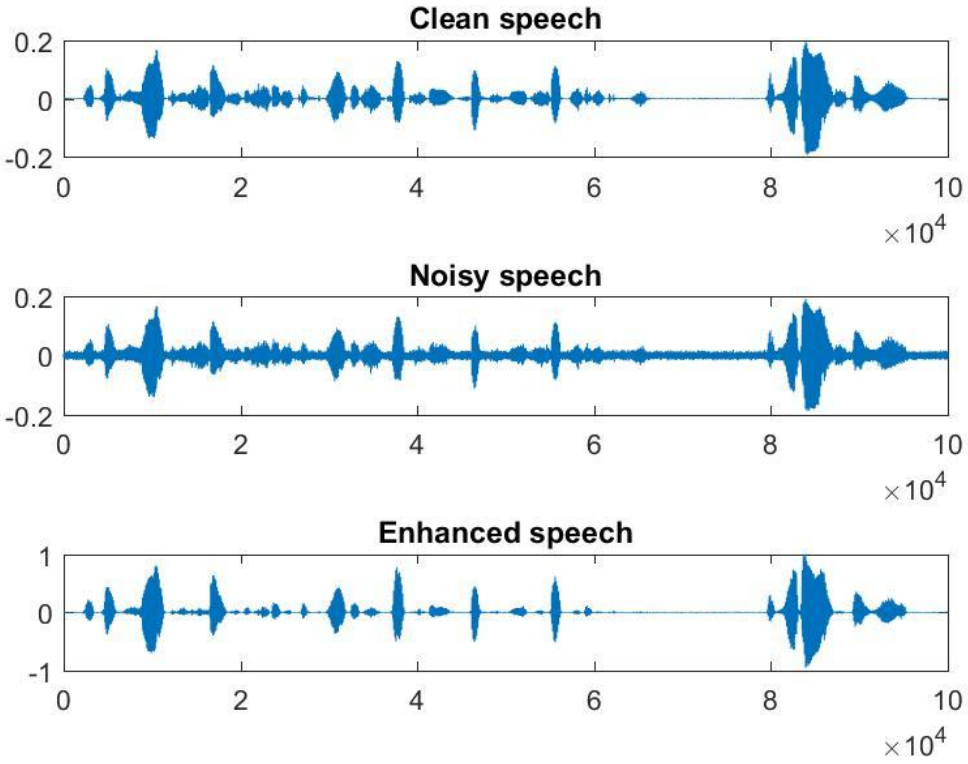
- The data set used for clean speech is TIMIT.
- Two types of Noises have been used in this project
  - White Noise using SNR.
  - 100 different types of noises with unknown SNR.
    - Example Noises are babel noise, car noise etc.
- Features used are log magnitude STFT with window length of 32ms and hop size of 16ms.
-

Direct Regression

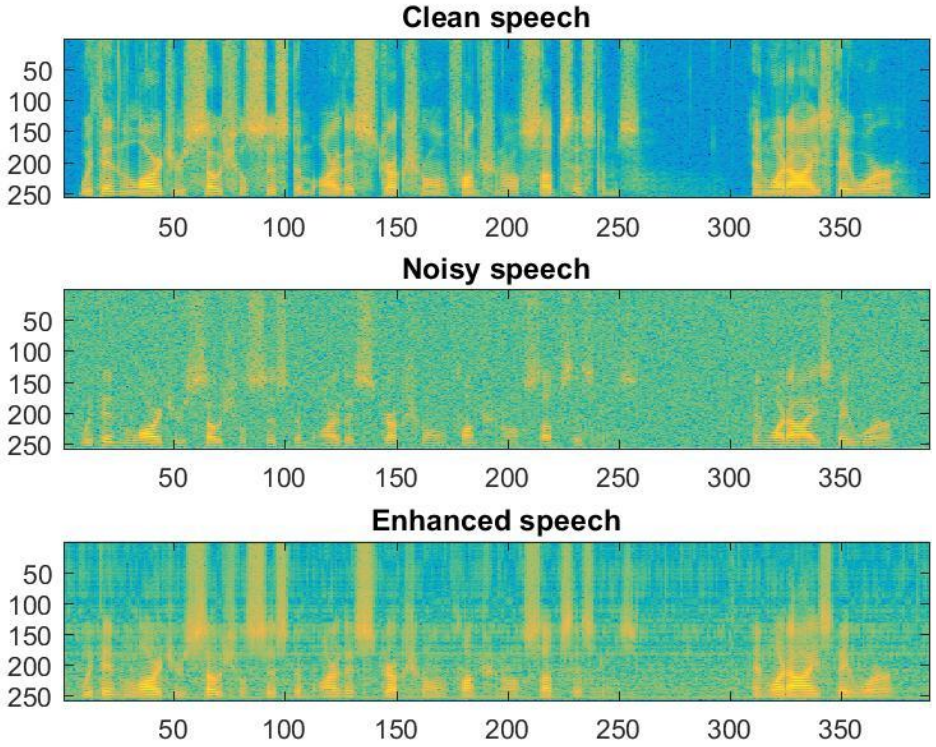
- Three Different Architectures have been used for direct regression.
- Feed Forward Neural Network with 257x1000x257 is the architecture used.
- Hidden layers has Tanh and Output layer has linear activation function.
- Recurrent Neural Network with 257x1000x257 is the architecture used.
- Feed Forward Neural Network with a context of past and future 3 frames are concatenated and used as a feature
- The network architecture is 1799x1000x1000x1000x257 is the architecture used.

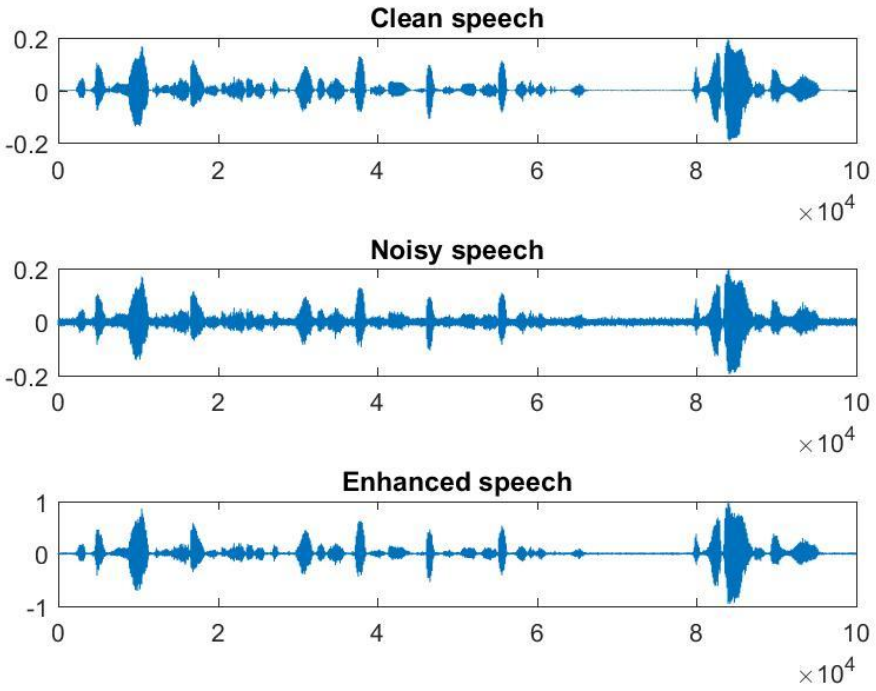
MLP


Clean speech
Noisy speech
Enhanced speech
50
100
150
200
250
300

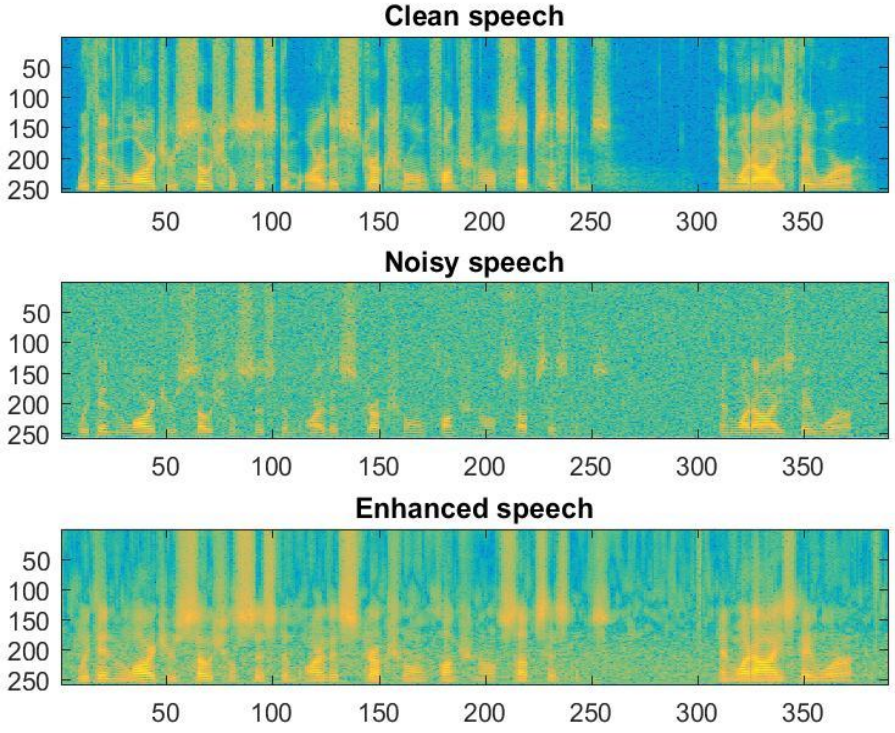

Clean speech
Noisy speech
Enhanced speech
0.2
0
-0.2
1
-1
0
2
4
6
8
10
×10^4

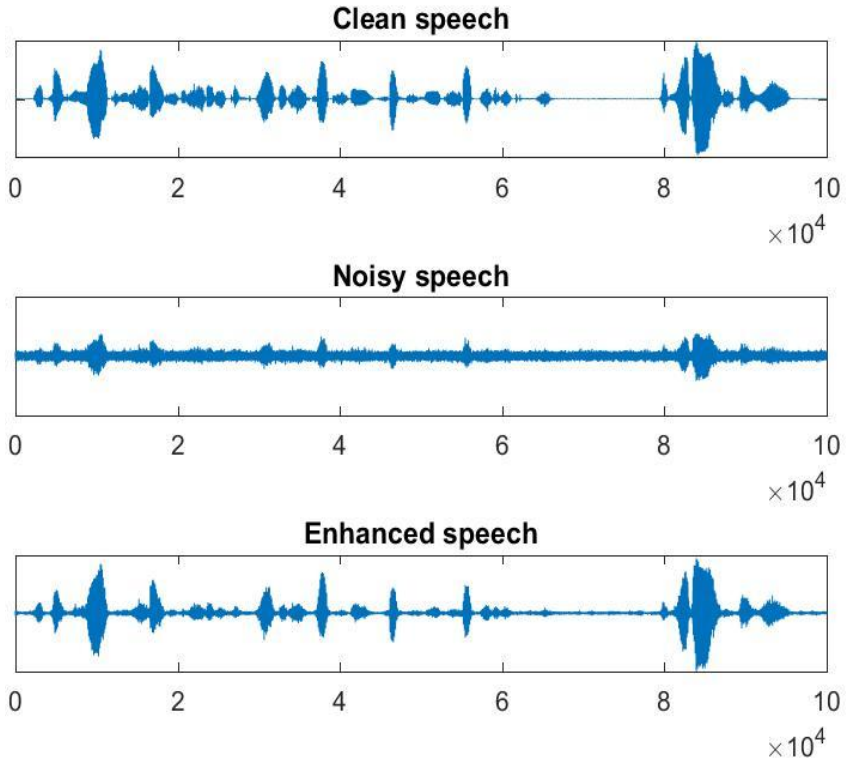
MLP with preamp filter



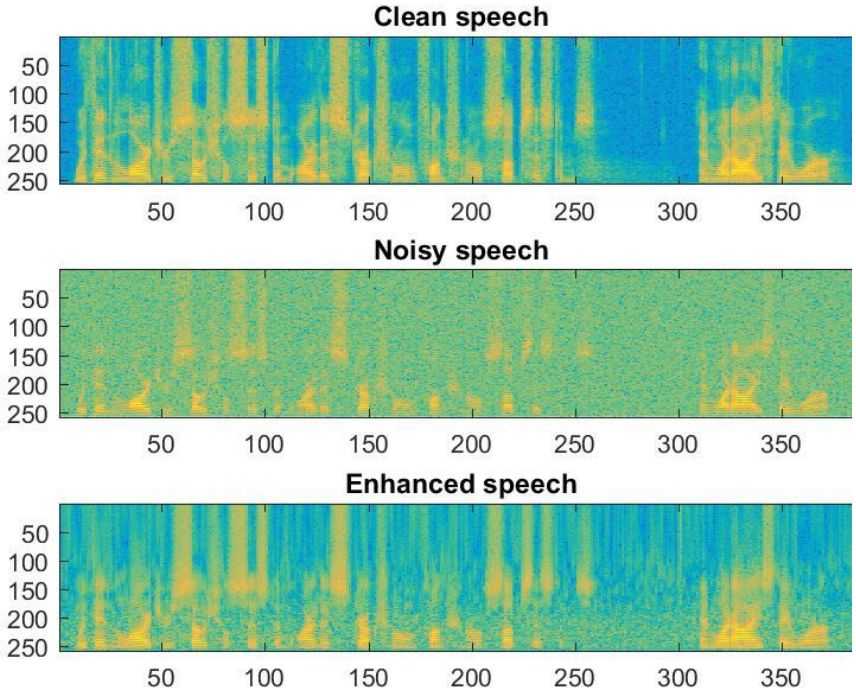


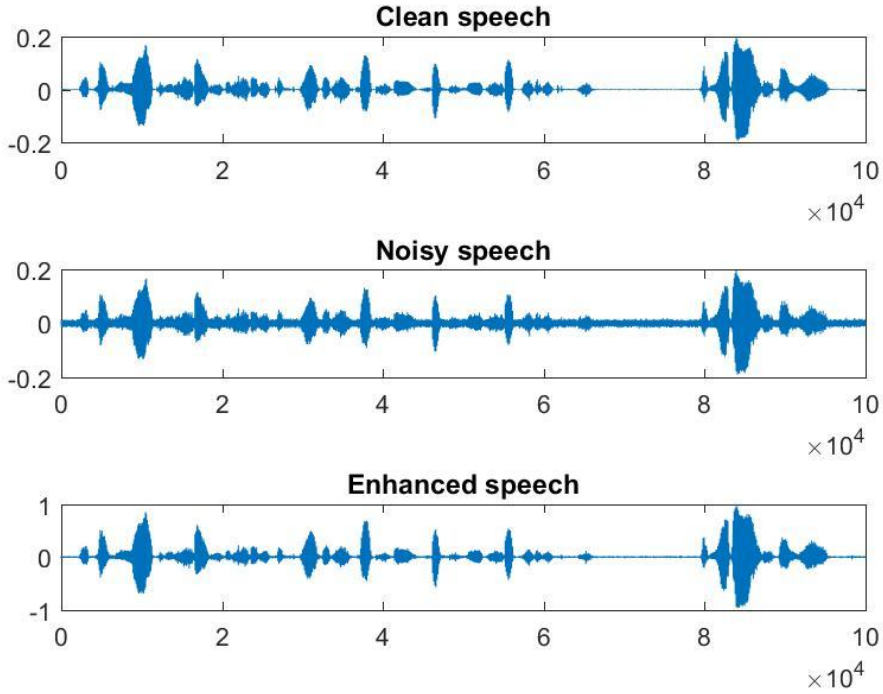
MLP with context





RNN


Clean speech
Noisy speech
Enhanced speech
50
100
150
200
250
300
350


Clean speech
Noisy speech
Enhanced speech
0.2
0
-0.2
1
-1
0
2
4
6
8
10
×10^4

Results

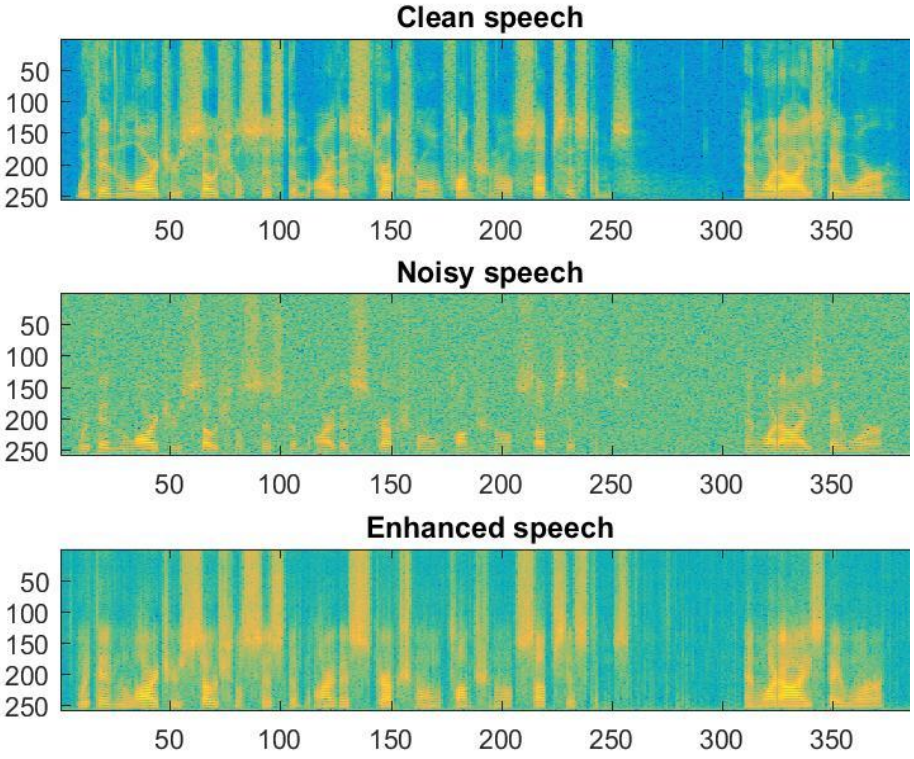
- Results using direct regression and white noise.

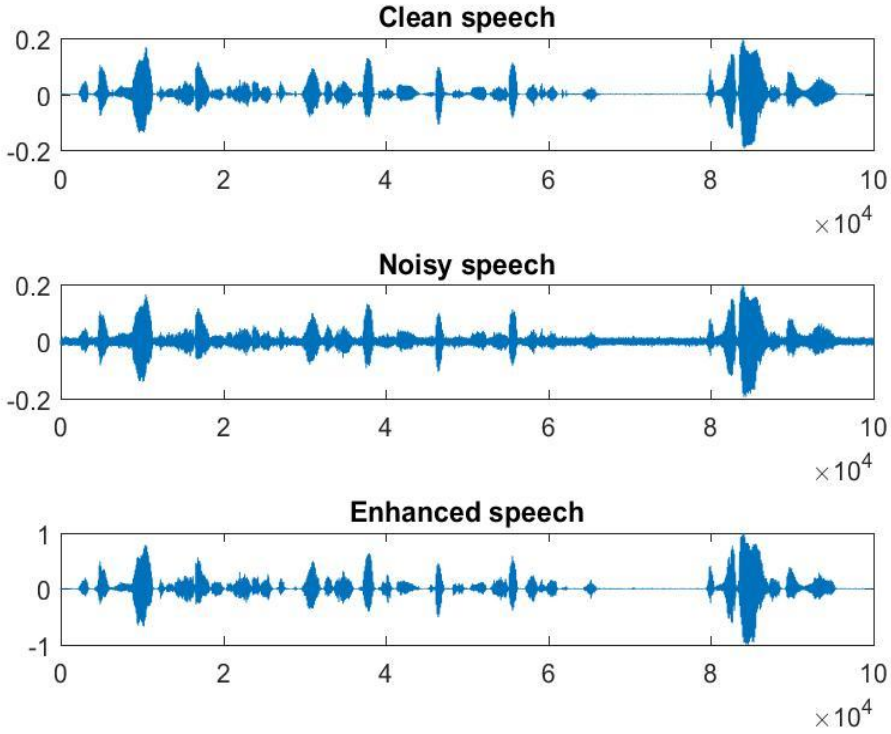
Direct Regression	Test Set PESQ Score (Est/Noisy)
MLP	2.828/2.203
MLP with preamp filter	2.646/2.213
MLP with context	2.911/2.221
RNN	2.828/2.203

Suppression Rule Based Estimation

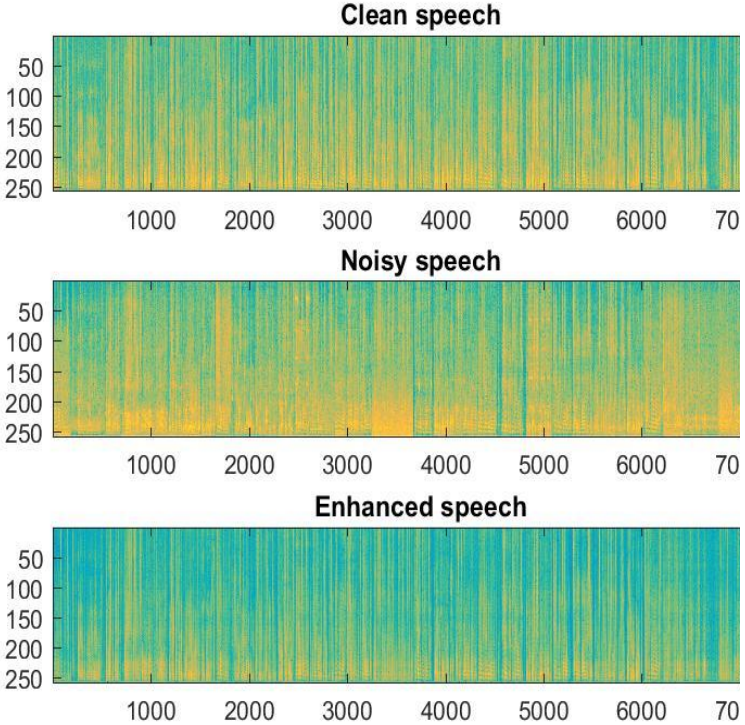
- Here we try to estimate the ratio of clean to noisy speech Magnitude spectrum instead of directly estimating the clean speech magnitude spectrum
- Similar to above approach we have trained using MLP, RNN, MLP with a context.
- The architectures are as follows
- MLP with 257x1000x257 is the architecture and RNN with also the same architecture
- MLP with a context of past 3 and future 3 frames and having an architecture of 1799x1000x1000x1000x257.

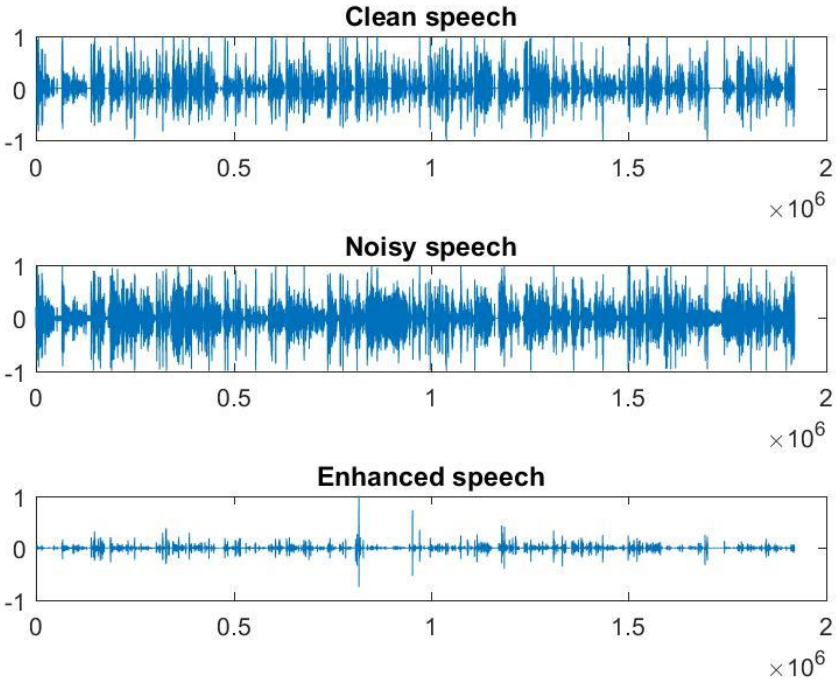
MLP with white noise



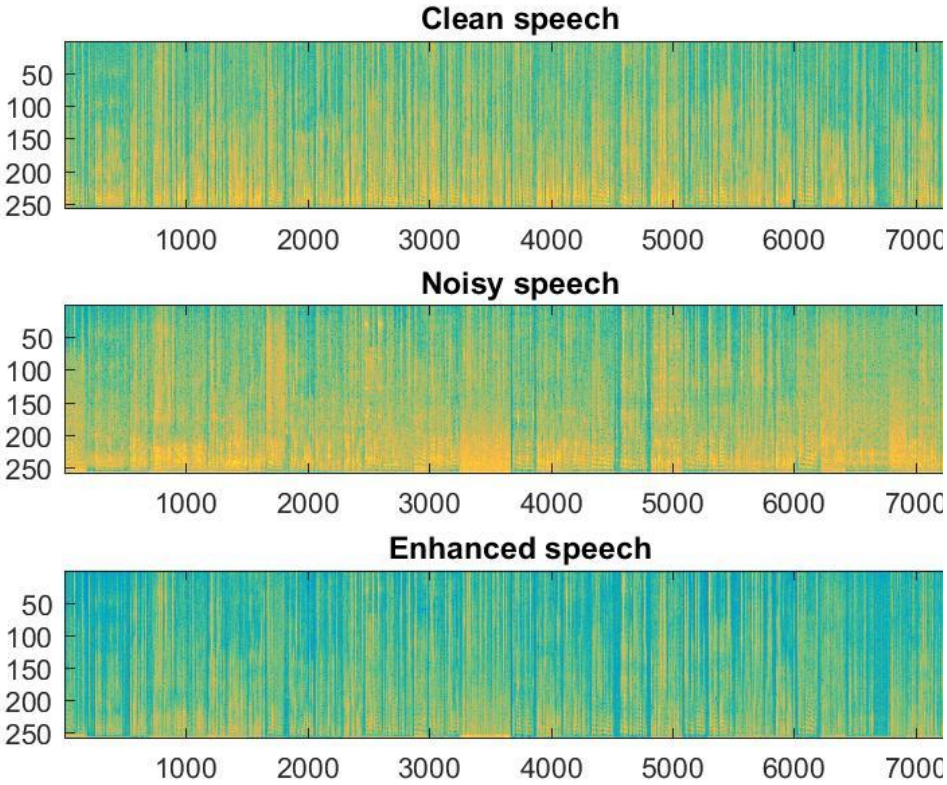


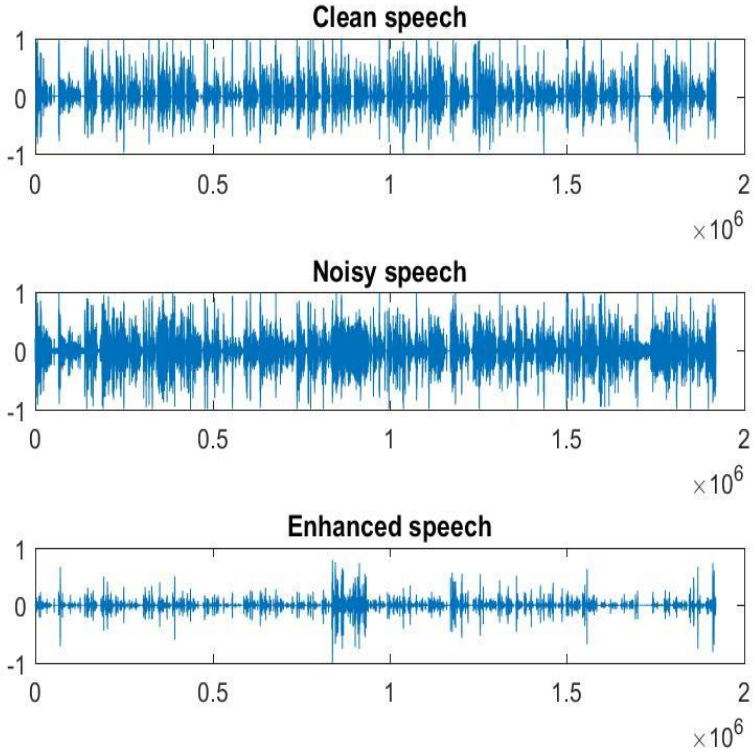
MLP with context



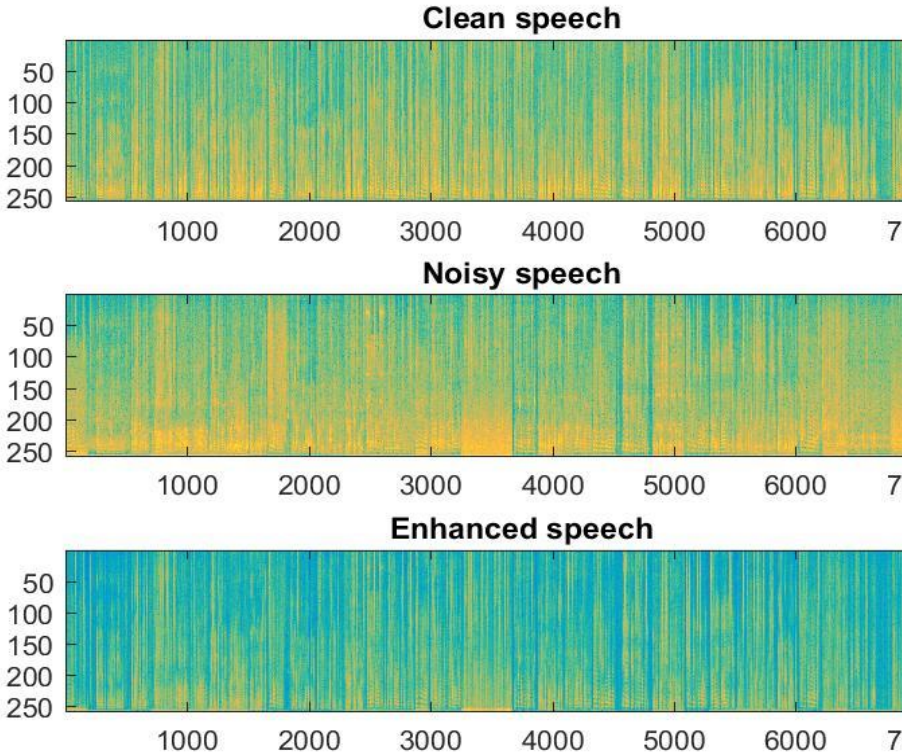


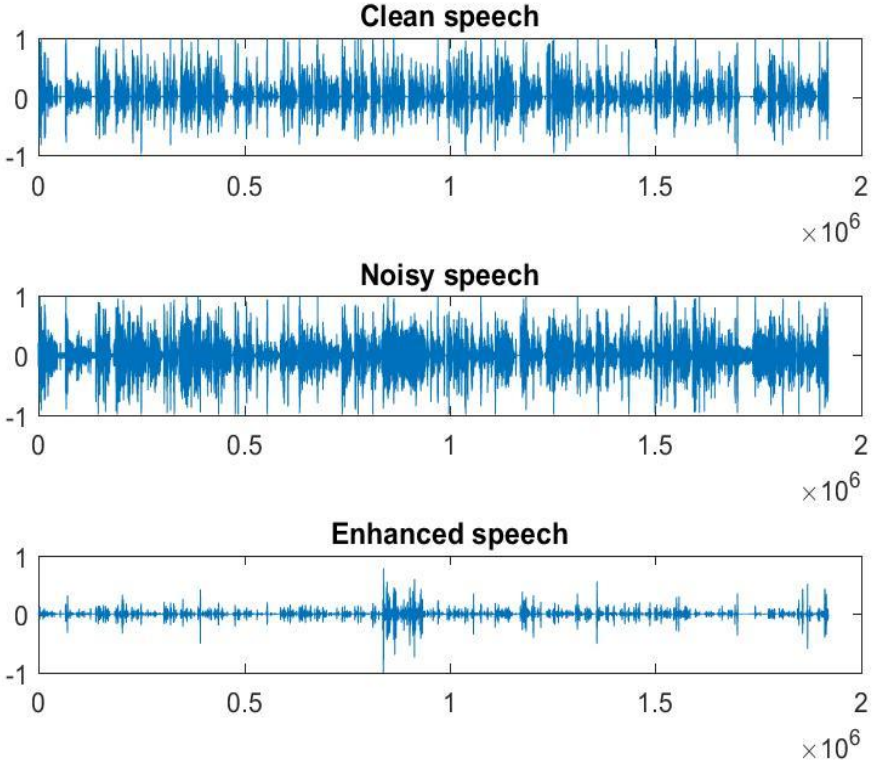
RNN


Clean speech
Noisy speech
Enhanced speech
50
100
150
200
250
1000
2000
3000
4000
5000
6000
7000


Clean speech
Noisy speech
Enhanced speech
1
0
-1
0
0.5
1
1.5
2
×10^6

BRNN


Clean speech
50
100
150
200
250
1000
2000
3000
4000
5000
6000
7
Noisy speech
50
100
150
200
250
1000
2000
3000
4000
5000
6000
7
Enhanced speech
50
100
150
200
250
1000
2000
3000
4000
5000
6000
7


Clean speech
1
0
-1
0
0.5
1
1.5
2
×10^6
Noisy speech
1
0
-1
0
0.5
1
1.5
2
×10^6
Enhanced speech
1
0
-1
0
0.5
1
1.5
2
×10^6

Results

- Results using Suppression rule based estimation.

Suppression Rule Based Estimation	Train Set PESQ (Est/Noisy)	Test Set PESQ (Est/Noisy)
MLP with context	2.233/1.8017	2.126/1.965
RNN	1.928/1.8017	2.005/1.965
BRNN	2.052/1.8017	2.142/1.965