|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Architecture | PESQ | ERLE | Remarks | Model Name |
| CNN 2D | 1.06 | 70.49 | Audio is heavily suppressed, hence high ERLE score | CNN2D\_1 |
| CNN 2D w Attention | 1.87 | 26.97 |  | CNN2D\_4 |
| CNN 1D | 1.40 | 28.08 |  | CNN1D\_1 |
| LSTM | 1.18 | 29.43 | 2 LSTM stacked layers | LSTM\_1 |
| LSTM | 1.33 | 29.43 | 3 LSTM stacked layers | LSTM\_2 |
| LSTM | 1.42 | 29.43 | Hidden dim changed from 256 to 1024 | LSTM\_3 |
| UNET | 2.22 | 25.52 |  | UNET\_2 |
| UNET | 1.06 | 25.64 | Random noise is spectrogram | UNET\_3 |
| UNET w LSTM bottleneck | 1.82 | 24.42 |  | UNET\_LSTM\_1 |
| UNET w LSTM and Attention | 1.30 | 25.21 | Attention on last layer | UNET\_LSTM\_2 |
| UNET w LSTM BN | 1.93 | 28.18 | Batch Normalize after every double convolution | UNET\_LSTM\_3 |
| Base Paper | 1.54 | 23.36 | InPalceCNN and FrequencyLSTM | Base paper |
| UNET Paper | 1.42 | 24.53 | UNET paper |  |
| UNET with Conformer | 1.56 | 25.38 | Conformer Block added to UNET w LSTM | UNET\_Confromer |

Acoustic Echo Cancellation

Notes:

* ERLE and PESQ are inversely impacted, when one increases the other decreases.
* UNETs considerably increase the PESQ score.
* High ERLE score is due to noise than actual audio, conclusion drawn from the respective spectrograms.