Dereverberation of speech signals using weighted prediction error

Debapriya Tula
S20170010039

Problem Description:

We aim at a statistical model-based dereverberation approach where we try to estimate the late reverberation in a speech signal captured by distant microphones without prior knowledge of the room impulse response, i.e., the transfer function between the speaker and the microphone. Late reverberations are dense, unaffected by small variations in the arrival time of the speech signal to the microphone, unlike the early reverb parts which look sparse so that their frequency characteristics are greatly affected. We try to estimate the amount of past signal contained in the present signal using a technique called Delayed Linear Prediction(DLP).

Existing solutions and background:

There have been quite a lot of approaches to this problem, like estimating the room transfer function using a microphone array and estimating the directions of arrival (DOA) and enhancing the signal component coming from source by controlling the directivity of the microphone. The drawback is it requires a large number of microphones to obtain sufficient directivity gain.
The next idea was finding the inverse characteristics of reverberation(called inverse filtering). However, it does not work so well when the reverberation characteristics change, e.g., due to small positional differences of the speaker or the microphone.
Multichannel linear prediction: The source signal is assumed to be independent and identically distributed(iid). It can estimate the channels without any prior knowledge of the characteristics of the signals. The drawback of this approach is that a speech signal is not an iid sequence and this method cannot thus be directly applied to speech signals.

Methodology:

The method we have used is known as the weighted prediction error(WPE) which is an improvisation over the multichannel linear prediction mentioned earlier by dividing the speech signal into small time frames and resolving it in the frequency domain.
We cannot use Fourier Transformation right away due to the varying nature of frequency of

speech signal with time. So we divide the signal into frames and apply FFT to it using STFT. To this, we can then apply WPE to each frame which estimates the amount of past signal in the present (a Gaussian function with certain mean and variance is estimated) based on a prediction order. After this, we can subtract the estimation for a time frame and henceforth. With DLP, the reverberation can be divided into two parts, viz, early and late reverb. It can be shown that DLP can suppress the late reverb effectively without significantly distorting the short time correlations of the speech, with the assumption that speech is stationary. With the use of time-varying speech characteristics with multichannel linear prediction, the reverbs have been reduced to a significant extent.

Results:

We have recorded voice samples on our phones(non-reverberant), added reverb to them using a software known 'Audacity', fed this reverberant signal through the algorithm and evaluated the reverb-free reconstructed signal for similarity with the original reverb-free signal. We are finding mean absolute error((1/m) * (| original_signal[i] - reconstructed_signal[i] |) for all i in m) at each discrete point and finding the percentage error.

We see that there is around 65-80 % accuracy in the results.

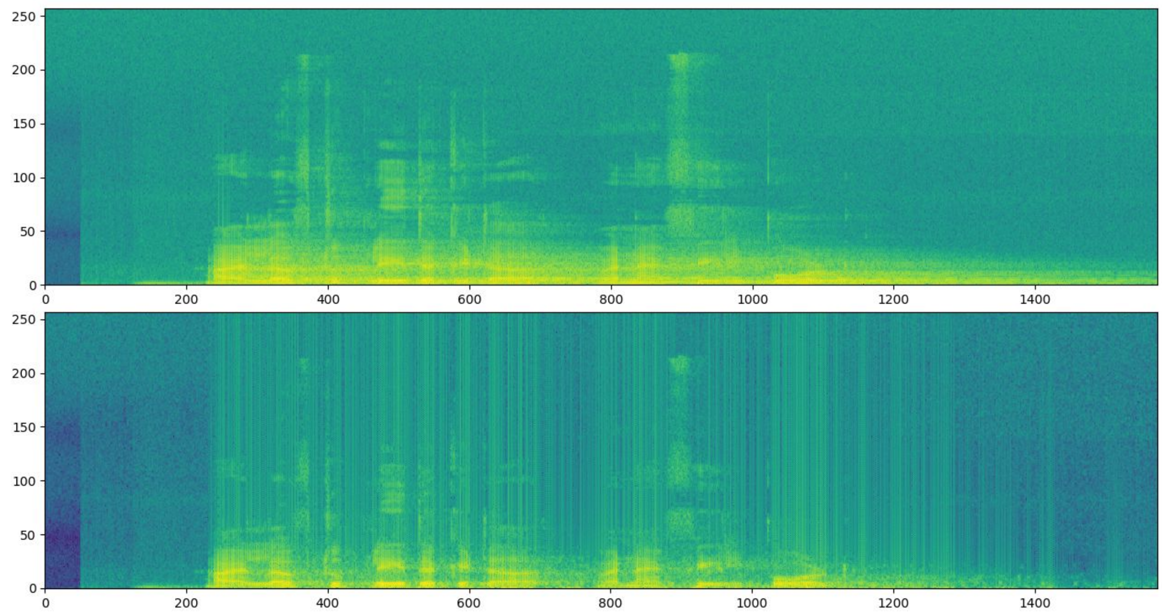
In the following graphs, the signal plotted in the top part of the image is the original signal containing reverb and the signal below that is the one with the reverb removed. It is a frequency(kHz) vs time(sec) graph.

The values of 10 samples

1) Sample 1

Settings:
Input channel: 2
Output channel: 2
Prediction order: 200

Processing...
Done!
Total time: 140.724320
Write to file: wav_out/Rec01_out.wav.
Percentage of accuracy(mean absolute error): 75.81862502880888

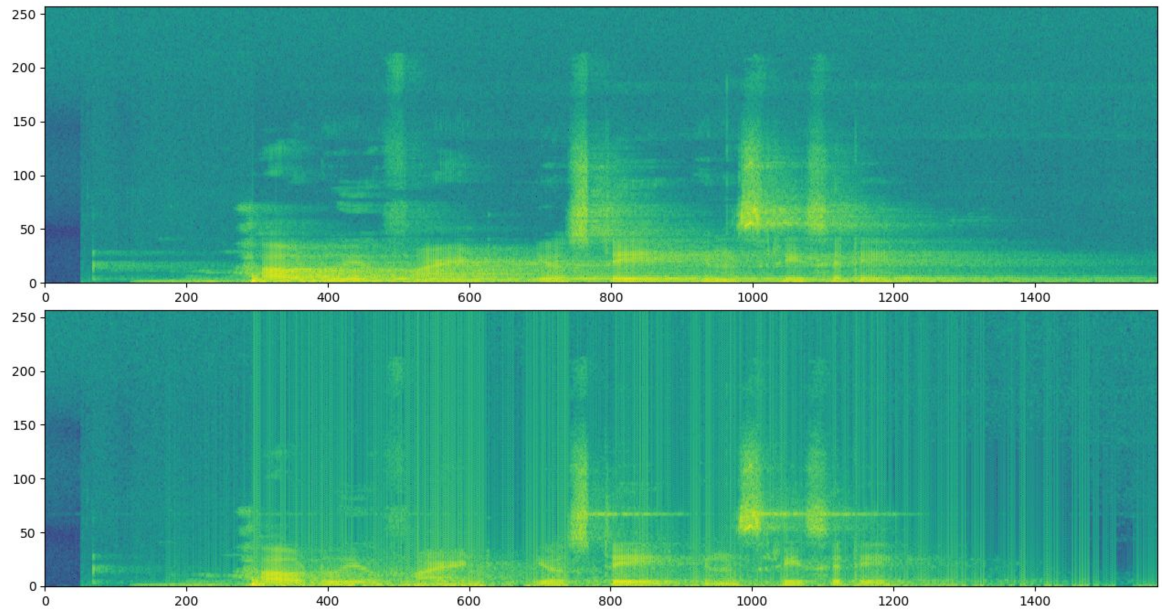


2) Sample 2

Settings:
Input channel: 2
Output channel: 2
Prediction order: 200

Processing...
Done!
Total time: 154.068527

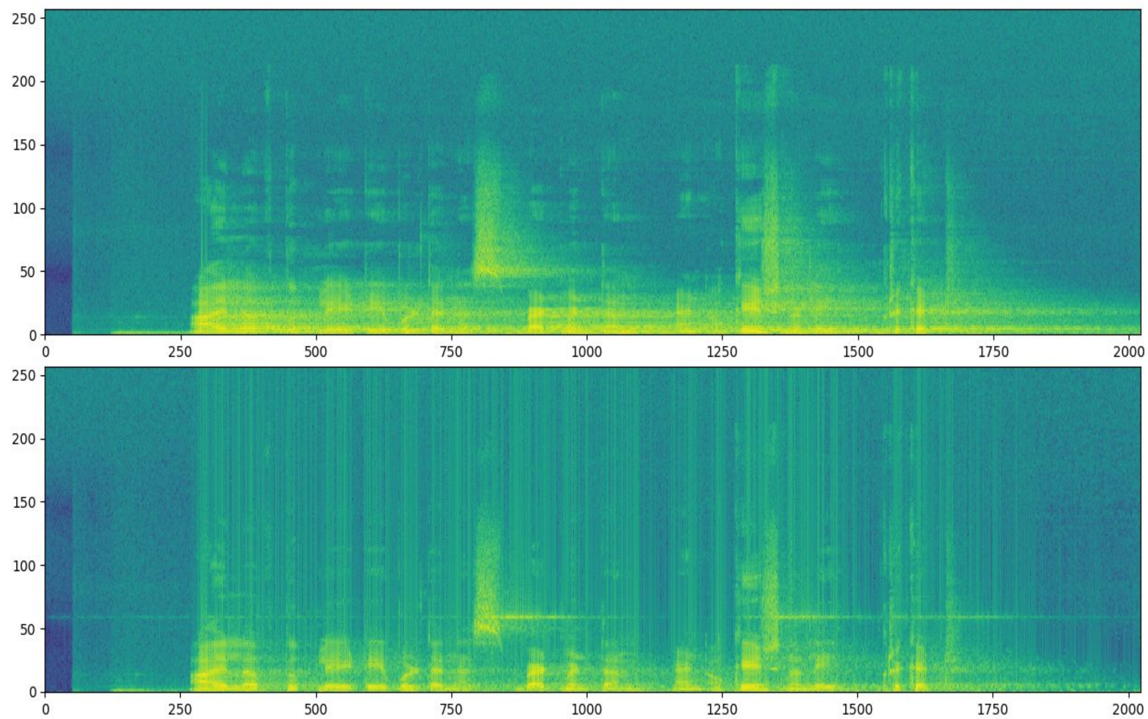
Write to file: wav_out/Rec02_out.wav.
Percentage of accuracy(mean absolute error): 76.50449169813385



3) Sample 3

Settings:
Input channel: 2
Output channel: 2
Prediction order: 200

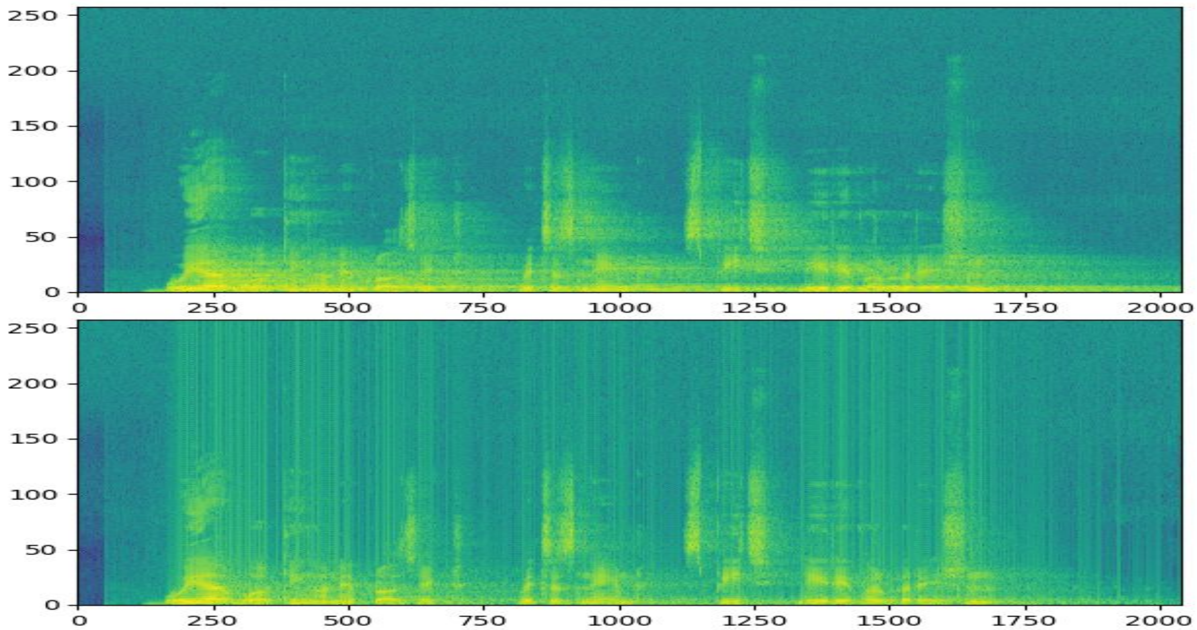
Processing...
Done!
Total time: 194.020392
Write to file: wav_out/Rec03_out.wav.
Percentage of accuracy(mean absolute error): 76.58281324992166



4) Sample 4

Settings:
Input channel: 2
Output channel: 2
Prediction order: 200

Processing...
Done!
Total time: 180.692888
Write to file: wav_out/Rec04_out.wav.
Percentage of accuracy(mean absolute error): 76.41074831483027

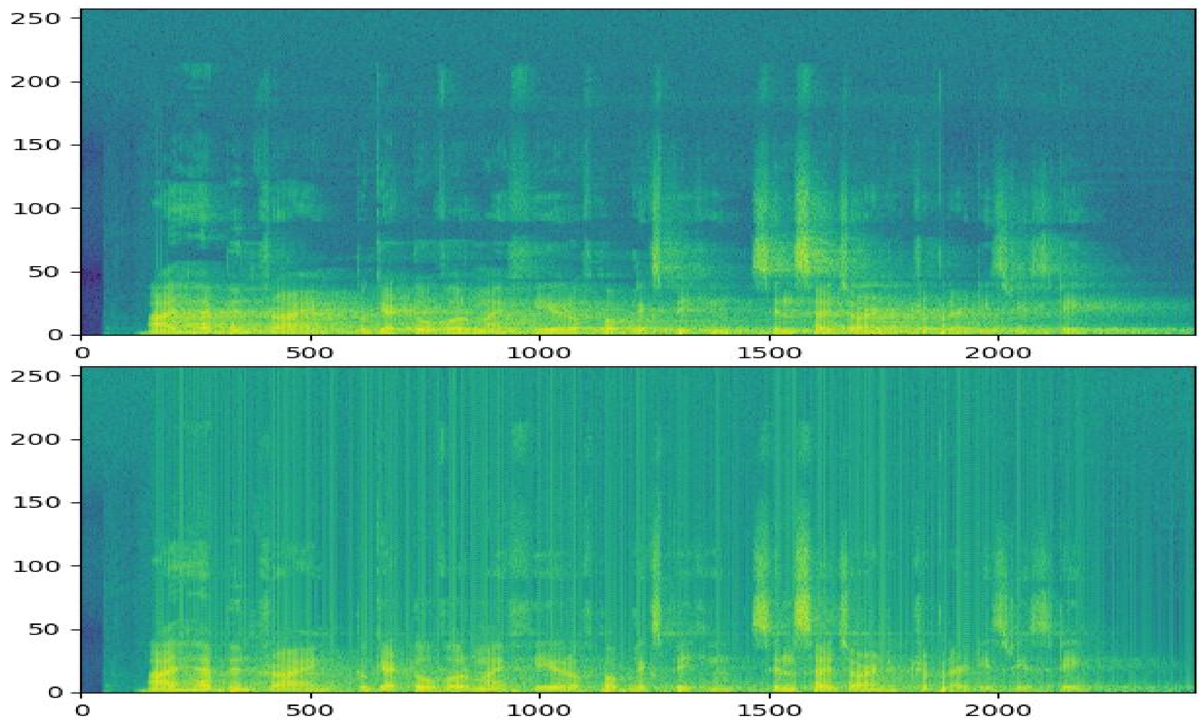


5) Sample 5

Settings:
Input channel: 2
Output channel: 2
Prediction order: 200

Processing...
Done!
Total time: 240.187127

Write to file: wav_out/Rec05_out.wav.
Percentage of accuracy(mean absolute error): 75.88644782207997

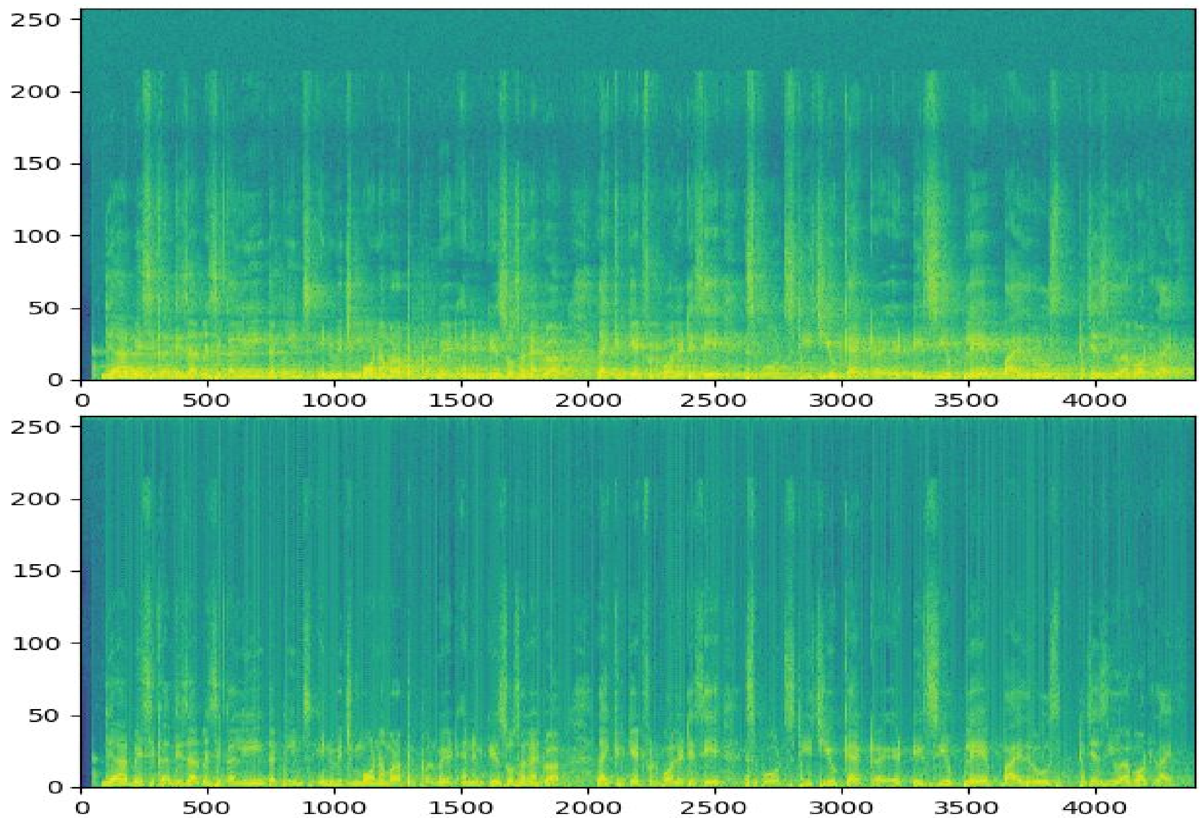


6) Sample 6

Settings:
Input channel: 2
Output channel: 2
Prediction order: 200

Processing...
Done!
Total time: 634.087554

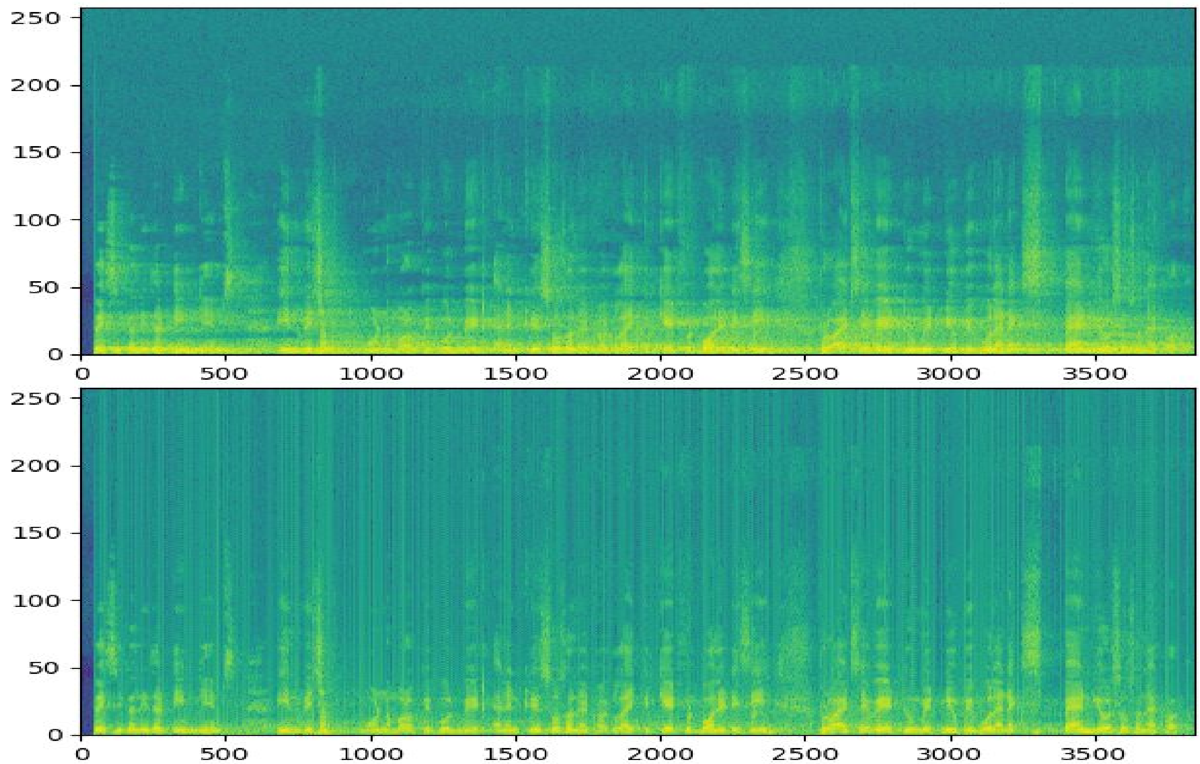
Write to file: wav_out/Rec06_out.wav.
Percentage of accuracy(mean absolute error): 76.2680479592358



7) Sample 7

Settings:
Input channel: 2
Output channel: 2
Prediction order: 200

Processing...
Done!
Total time: 786.825162
Write to file: wav_out/Rec07_out.wav.
Percentage of accuracy(mean absolute error): 76.572782648106

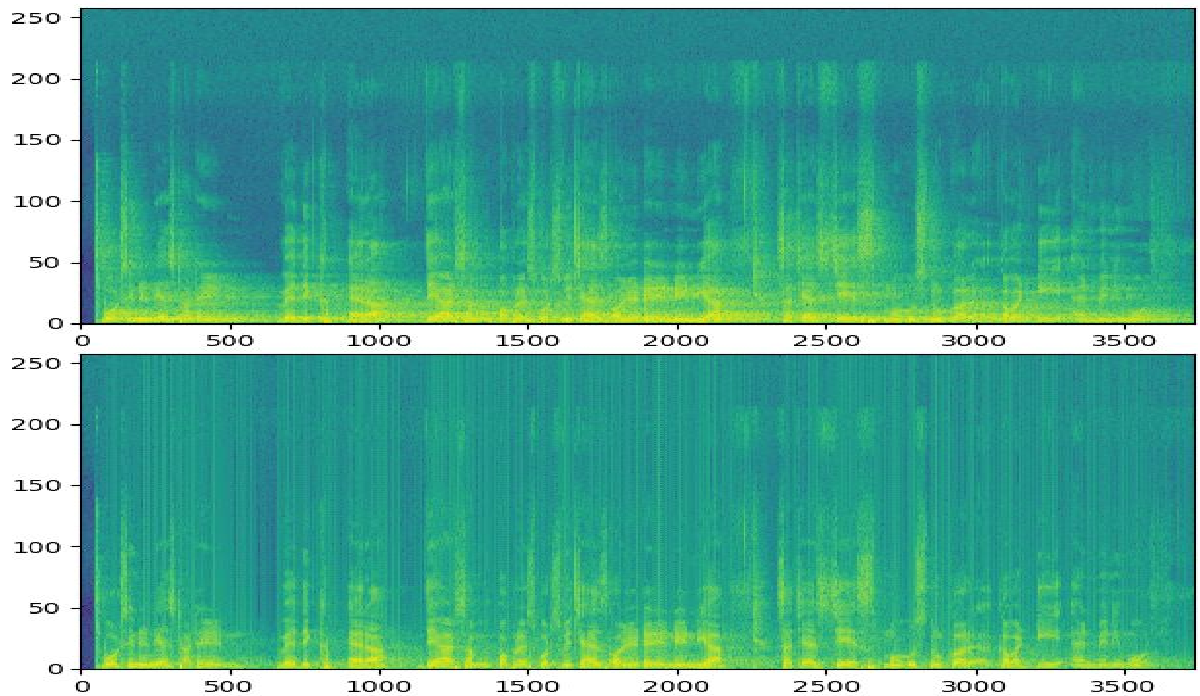


8) Sample 8

Settings:
Input channel: 2
Output channel: 2
Prediction order: 200

Processing...
Done!
Total time: 553.615700

Write to file: wav_out/Rec08_out.wav.
Percentage of accuracy(mean absolute error): 76.5603348399462

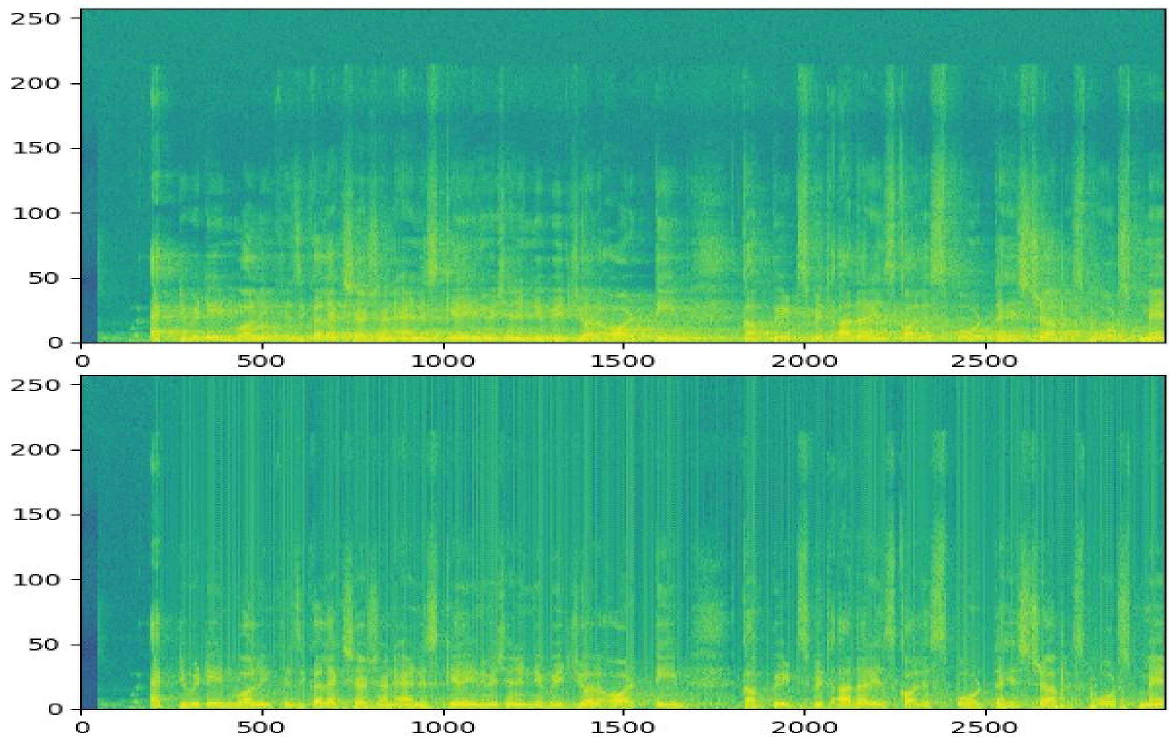


9) Sample 9

Settings:
Input channel: 2
Output channel: 2
Prediction order: 200

Processing...
Done!
Total time: 339.417050

Write to file: wav_out/Rec09_out.wav.
Percentage of accuracy(mean absolute error): 76.49482891477783

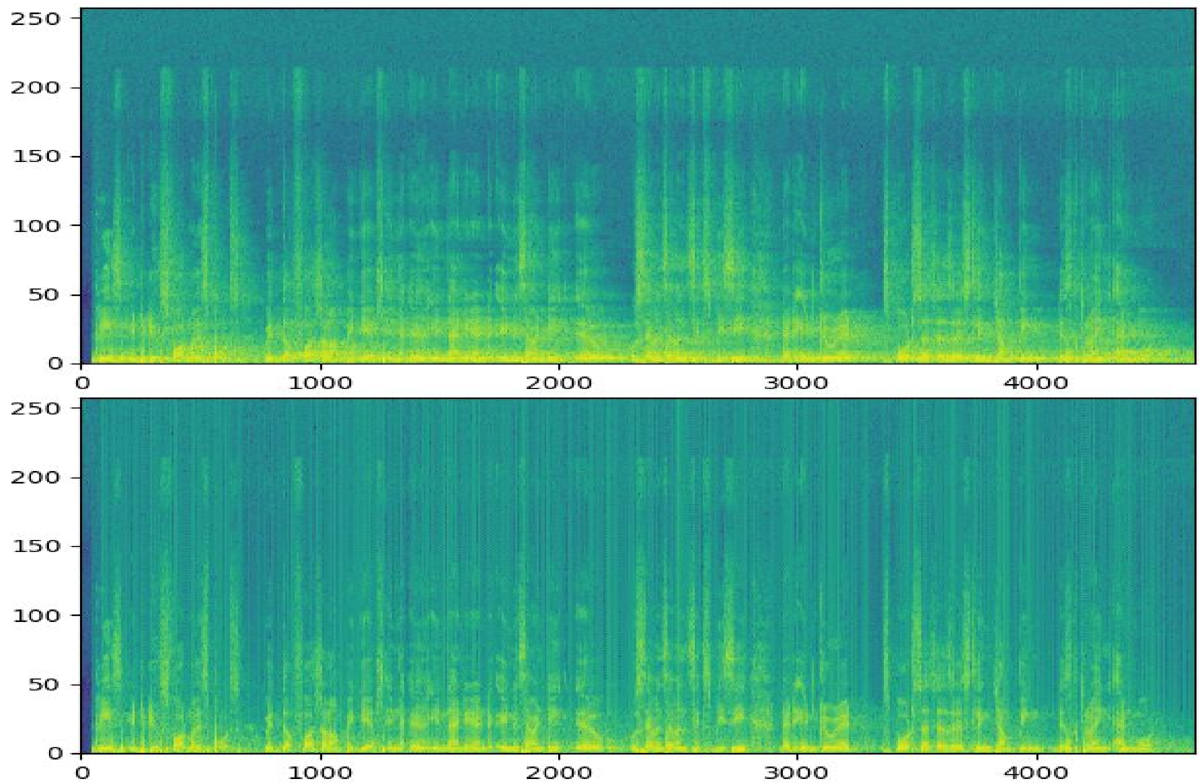


10) Sample 10

Settings:
Input channel: 2
Output channel: 2
Prediction order: 200

Processing...
Done!
Total time: 742.545817

Write to file: wav_out/Rec10_out.wav.
Percentage of accuracy(mean absolute error): 76.60359807521385

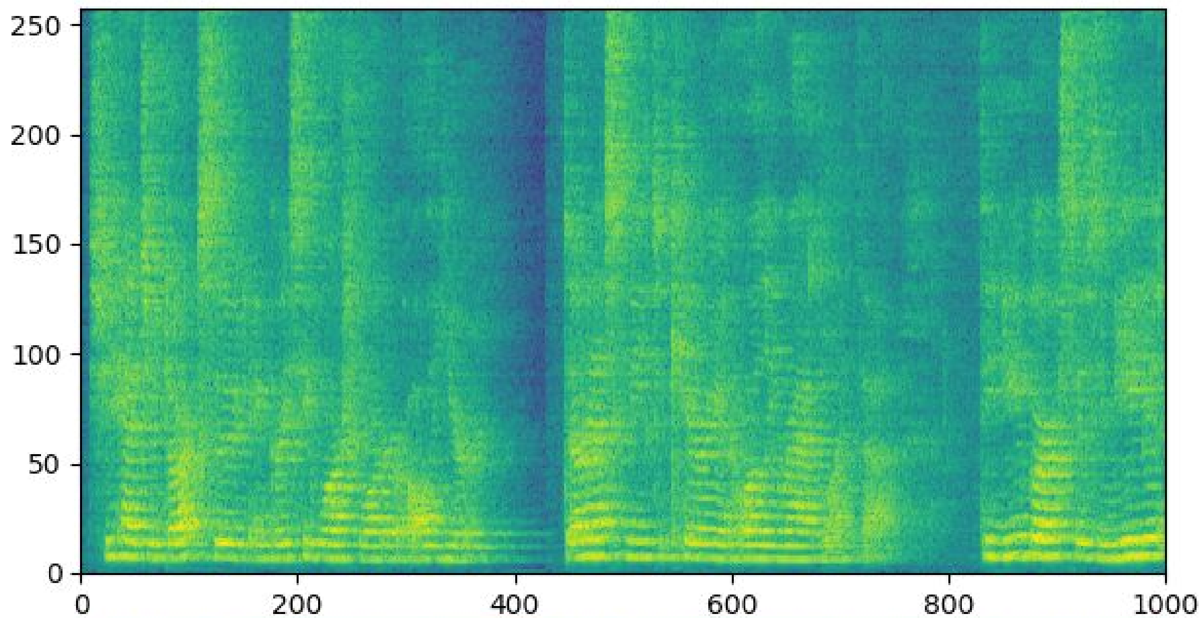


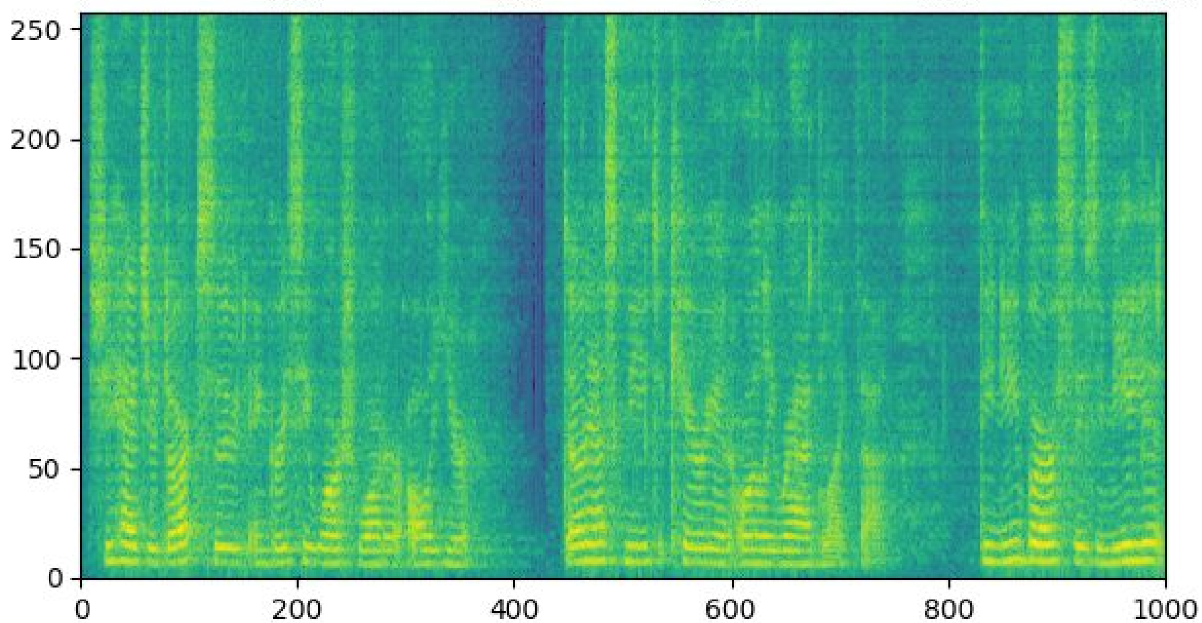
11) Sample 11

Settings:
Input channel: 4
Output channel: 3
Prediction order: 30

Processing...
Done!
Total time: 12.999754

Write to file: wav_out/sample_4ch_out.wav.
Percentage of accuracy(mean absolute error): 86.52402802947645


250
200
150
100
50
0
0
200
400
600
800
1000


250
200
150
100
50
0
0
200
400
600
800
1000

Analysis and Discussion:

We saw that the late reverbs were removed to a good extent when the number of input channels was at least three. We had recorded samples on our phones with 2 input channels only, which is why the reverb was not removed to a great extent.

For speech signals with at least 3 input channels, the outputs were much better. Even with less dependency on the previous time-frames, the results were much better in the last sample(4 input channels, 3 output channels and a prediction order of 30)

We also observed that setting the prediction order according to the RTF of the room brought great changes even when the number of input channels was low. This is something we hadn't observed previously. The prediction order was entered manually. It is the number of past time-frames on which the present frame depends, and this parameter is dependent on how long the late reverbs are. This is a problem that could not be fixed owing to analyzing the signal first for a suitable prediction order and then analyzing it again. There do exist techniques like model order weighing which estimate a good prediction order for a given reverberated signal.

The original samples:
https://drive.google.com/open?id=1mxsxvJh5tOVXfv7PR1jh4R1wMqK3ok33

The original samples with reverb added to them
https://drive.google.com/open?id=1jSDQ4H8f5S-89GXYcn_U6KDj-OjrAe-g

The dereverberated signals
https://drive.google.com/open?id=1SbnUwlBKH3R4V1E7mwKOnQaIrrRxA43G

References:

https://ieeexplore.ieee.org/document/5547558
https://www.youtube.com/watch?v=Nr5fwEEInRQ
https://www.sciencedirect.com/science/article/abs/pii/0167639396000118
https://ieeexplore.ieee.org/document/859140
https://ieeexplore.ieee.org/document/6854590

Article we went through for finding a suitable prediction order

https://ieeexplore.ieee.org/document/790651