Multimedia and Video Communications (SSY150) Lab 1

Group 5: Haitham Babbili, Olalekan Peter Adare

17th April 2020

Task 1

1.1

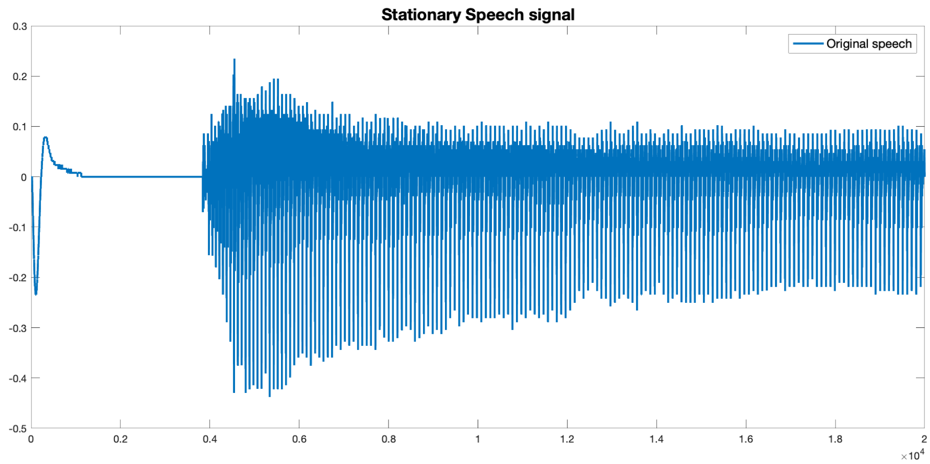

Figure 1: Graph showing stationary recorded sound waveform

Figure 1 show the wave form of stationary recorded speech signal

1.2

Block size = 300 ms
Sampling frequency, Fs =10 kHz
L = Fs $\times$ $Blocksize$
$L = 3000 samples$
a ̇j= 1 $\times$ 13,where j=1,2,..,p.The filter is of order p=12. This implies p +1 = 13 coefficients.
$Variance$ =1.628107205585200$\times 10^{-5}$

1.3

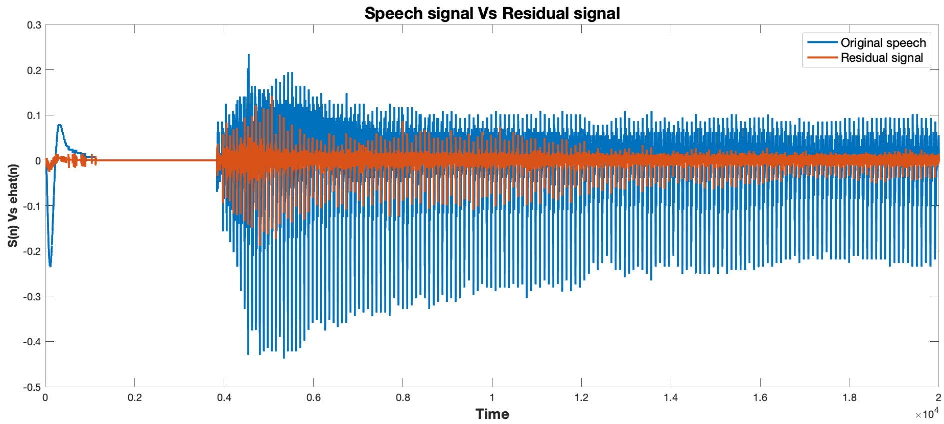

Figure 2: Original speech signal Vs the residual sequence ê(n).

Figure 2 show the original speech signal along with the residual sequence signal.

1.4

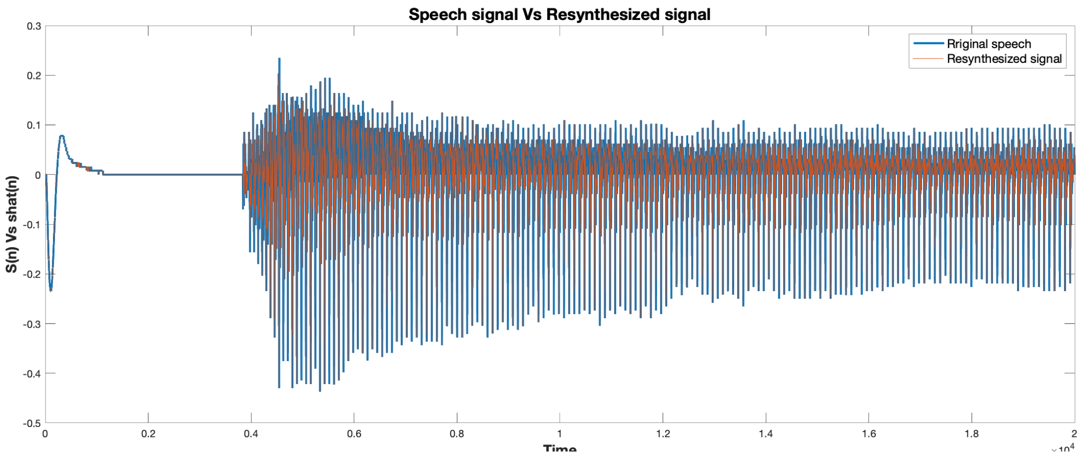

Figure 3: The original s(n) and the re-synthesized speech signal ŝ(n)

Figure 3 show that both s(n) and ŝ(n) are same since ŝ(n) synthesizing the speech signal s(n).
3-4 formants, with a filter order of 8 should be enough to re-construct the signal with no noise. In this lab, the order of the filter is 12, which implies 6 formants, so the signal construction will be perfect, more or less. This is why there seem to be no difference pronounced difference between the original and the re-synthesized signal.

Task 2

2.1

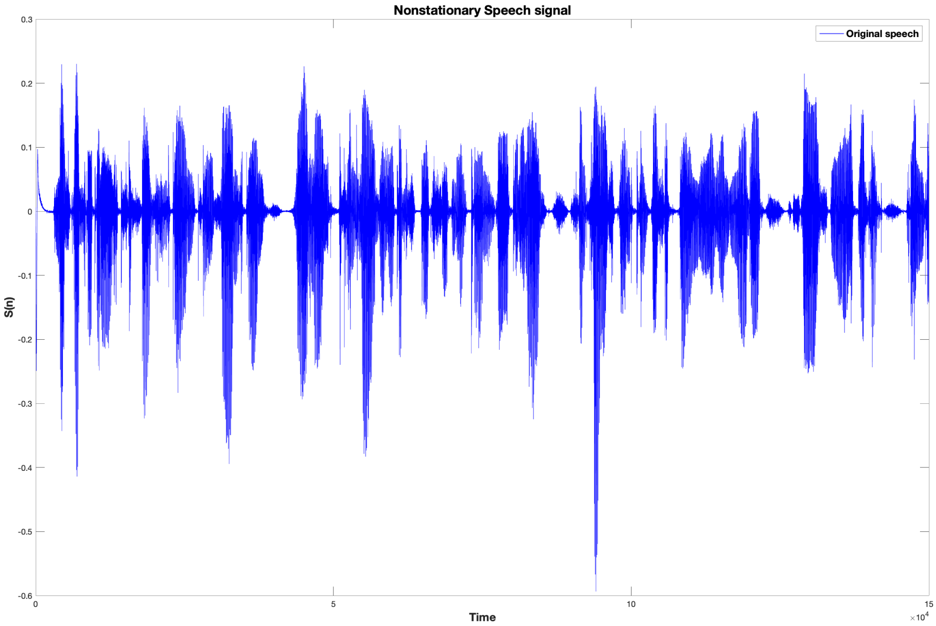

Figure 4: Graph Showing Non-stationary recorded sound waveform

2.2

Block size = 10 ms
L = Fs $\times$ $Blocksize$
$a.Thenumberofsamplesineachblock : L = 100$
$b.Thetotalnumberofblocks : TotalBlocks = 1500$
$c.Thesamplingrate : Fs = 10kHz$

2.3

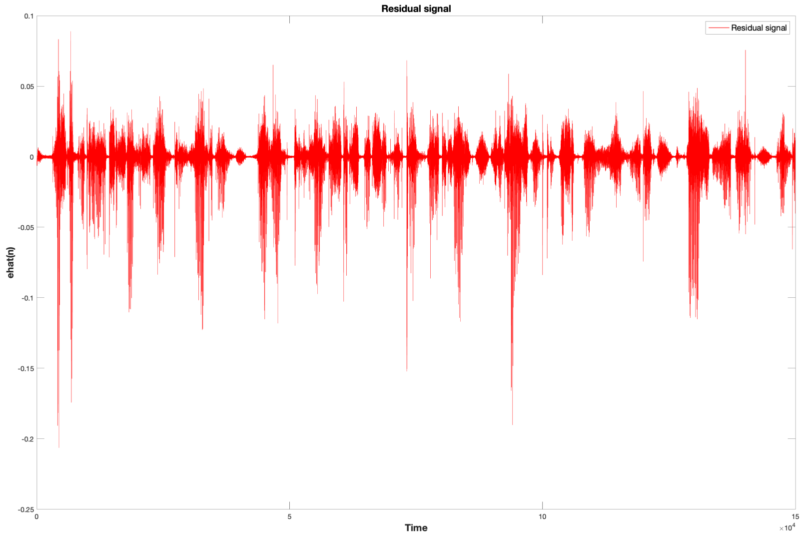

Figure 5: The entire residual sequence ê(n).

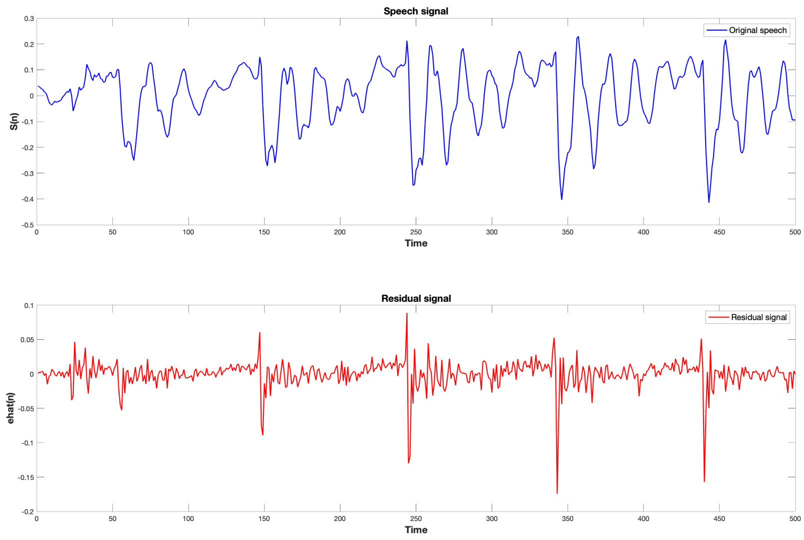

Figure 6: the residual sequence and the corresponding speech s(n) between (65) and (70)

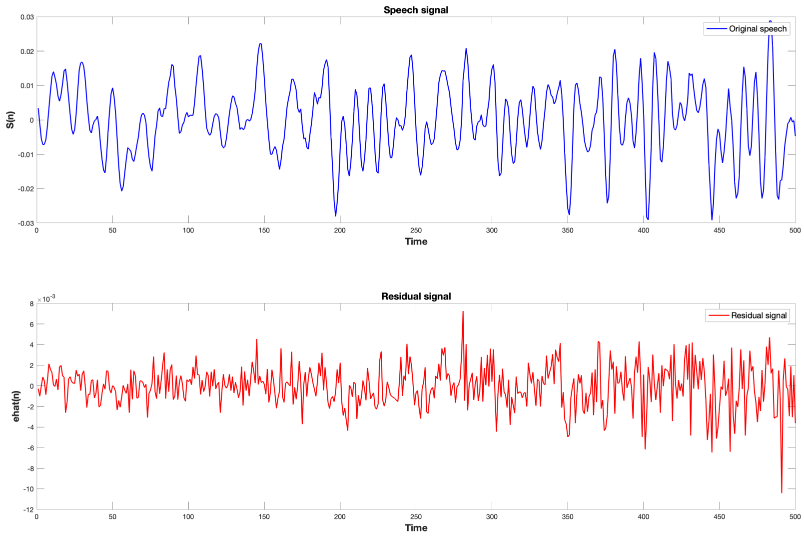

Figure 7: the residual sequence and the corresponding speech s(n) between (300) and (305)

Figure 6 shows the plot of the residual and original speech signal. It can be observed that pitch periodicity can be seen better from the original signal plot, than from the residual signal plot.
By manual counting, 6 Almost 19 samples in one pitch period.
By manual counting, 7 Almost 33 samples in one pitch period. The pitch period, or cycle, is about 97.

2.4

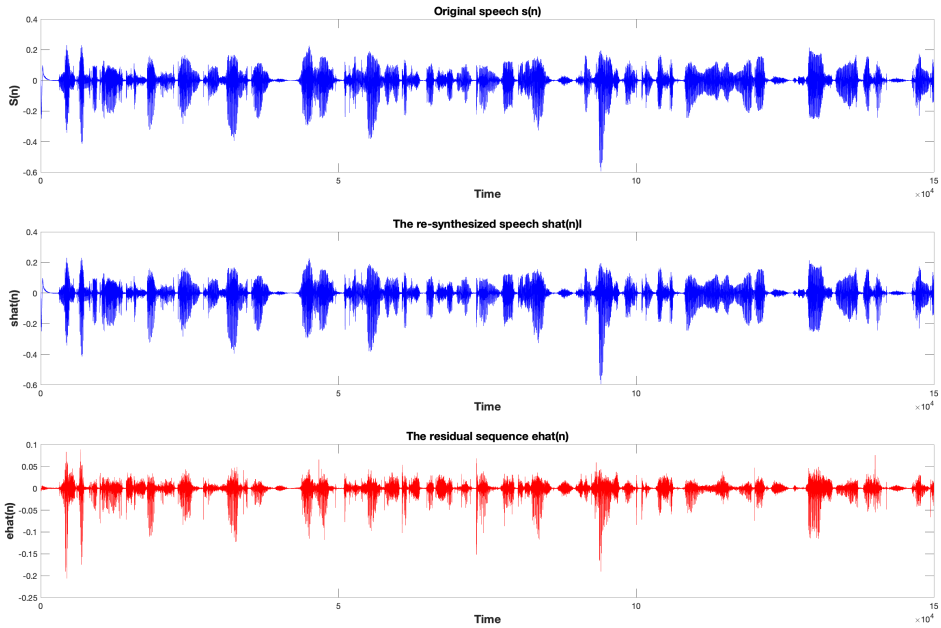

Figure 8: Comparing the original speech s(n), the re-synthesized speech ŝ(n), and the residual sequence ê(n)

2.4.a

By careful observation, the original speech, re-synthesized speech and the residual sequence are almost the same, but the residual sequence has a lower amplitude for the same duration of the observation.

2.4.b

After listening to the synthetic speech and the original speech, there seem to be no particular difference between them. This can be due to the fact that the modelled filter, with p= 12, is very high. The reconstruction loss cannot be heard by the human ear.

Task 3

3.1

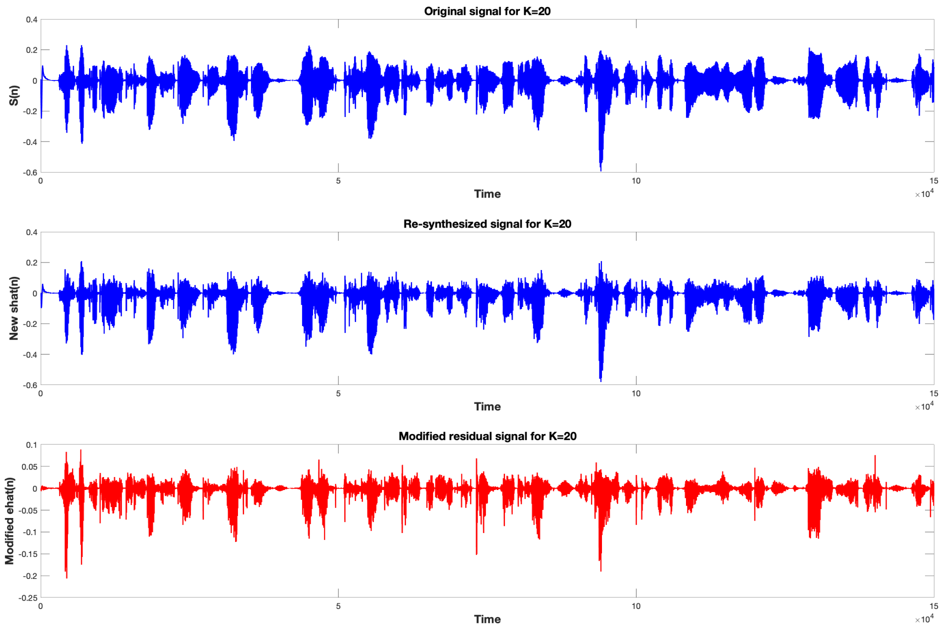

Figure 9: comparing between Original signal, synthesise signal and residual signal at K= 20

3.3

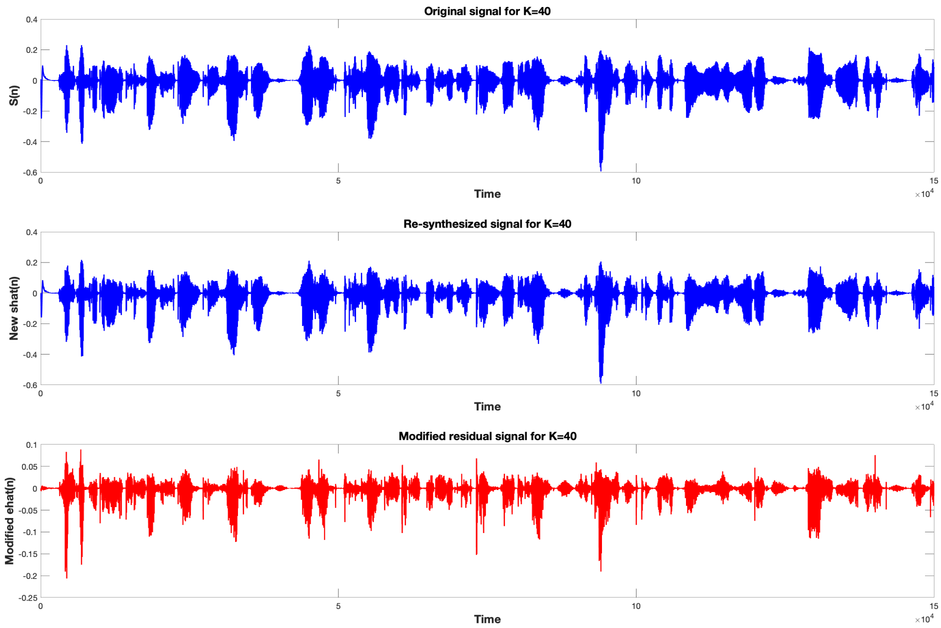

Figure 10: comparing between Original signal, synthesise signal and residual signal at K= 40

3.3.b

After listening to the original recorded speech and the re-synthesized speech, it can be noted that the original has less excitation, or residual voice and clearer, while the re-synthesized signal is louder and has a bit of noise distortions. The loudness shows a gain in the signal.

3.3.d

With K=20, the original signal is better and the residual signal has some noise distortions, or breakout. With K =40 the original signal is better and the residual signal also became better, with very little background noise. K corresponds to the number of residuals. K= 40 is thus a better case than K=20.

Task 4

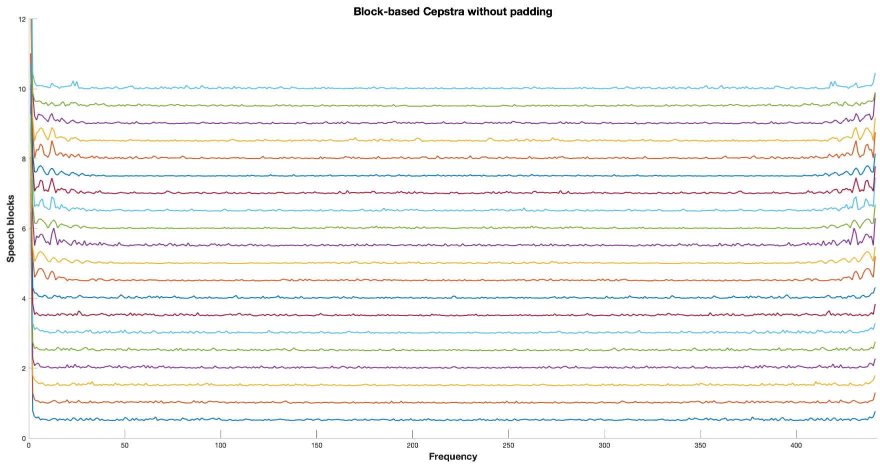

Figure 11: block-based cepstral without padding

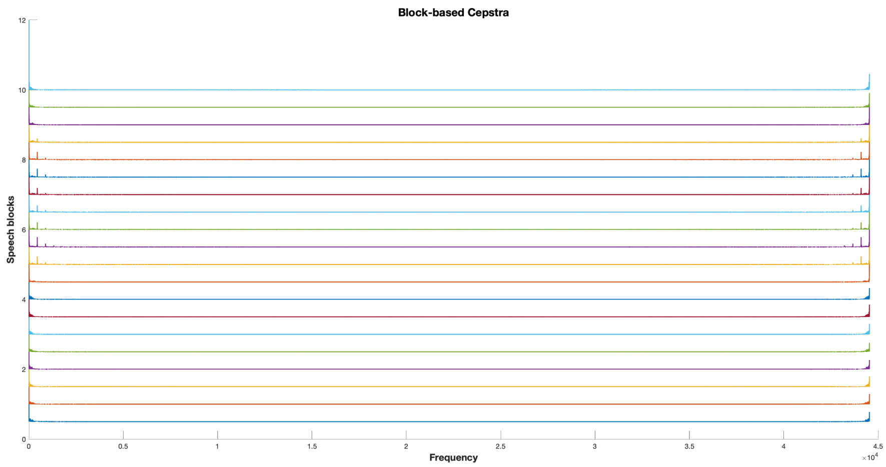

Figure 12: block-based cepstral

Task 5

5.1

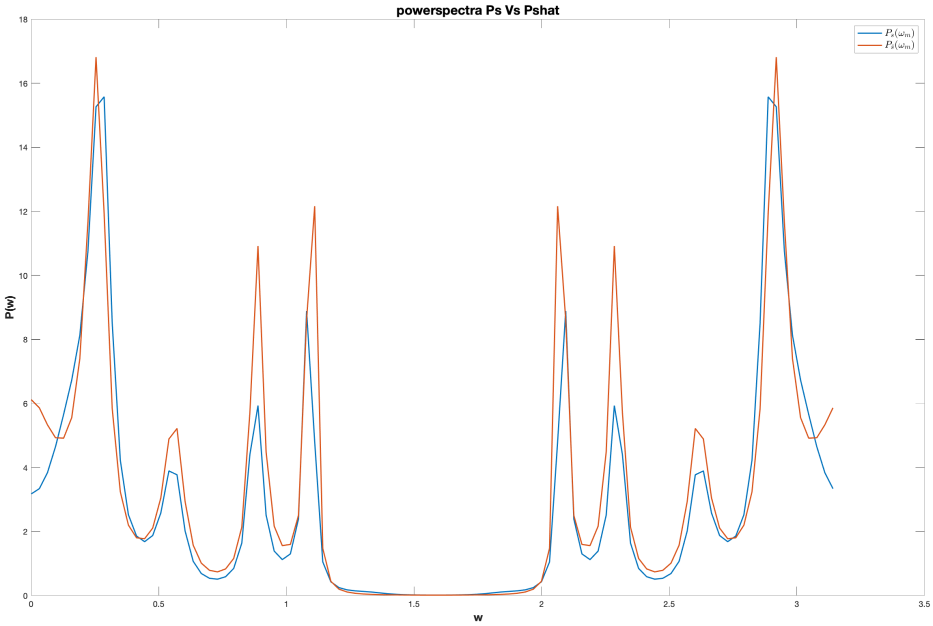

Figure 13: Power spectrum density for original and synthetic speech signal

5.2

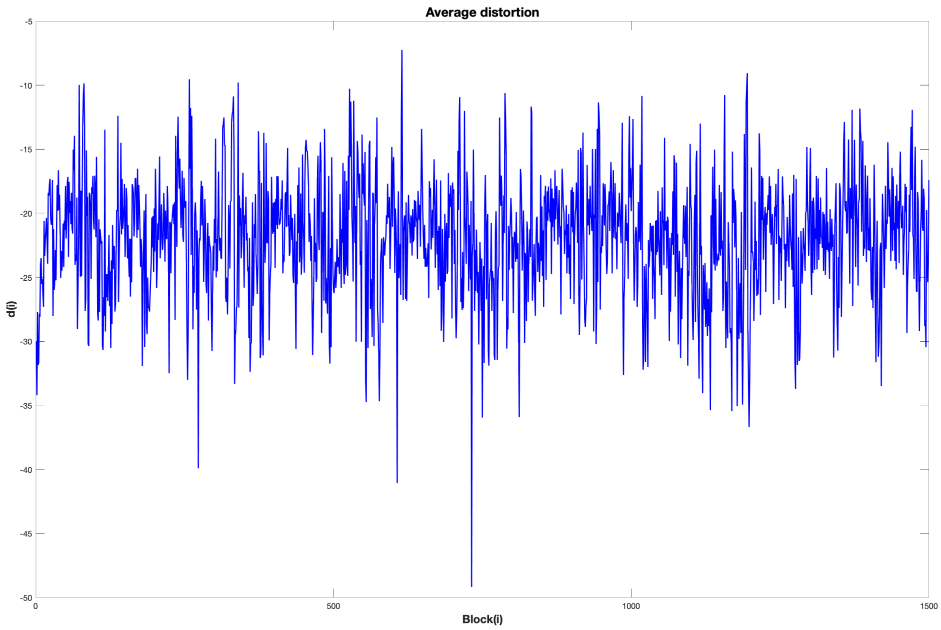

Figure 14: Average distortion

5.3

By careful observation, based on figure 13 , the synthetic speech signal power is more than the original speech power. There were 4 signal peaks noticeable from the plot. Also, they appear to be symmetric after 1.60,approximately, of the windowed. From figure 14, there is a high variation in the signal.
By varying the values of P to be lower than 12: for p=6, there were 2 peaks and the power of the original signal has more power than the synthesised signal. For p=8, 3 peaks became visible. P=12, 4 peaks became visible. P=14, 16, and 18, the were 4 dominant peaks, and the power level seems not to change much. Beyond p=18, that is 20 and above, smaller peaks apart from the dominant ones were visible. In all of this, the average distortion was relatively the same.
In general, it was observed that at beyond p=12, there seem to be no much visible improvements to the power of the signal, and the plot was shifted to the left. It is as if it has reached its peak, with relatively the same average peak.

Task 6

6.1

In the lab, LPC algorithm is used for the compression. Compression is achieved by segmenting the original signal and filtering, based on a specific order of the filter.

6.2

From Task 1, L= 3000 samples, at 10kHz. Given that each sample requires 8 bits for encoding.
10000 x 0.3 x 8 = 24000 bits
Therefore 24,000 bits will be required for encoding the uncompressed speech.
From Task 2, L= 100 samples, at 10kHz. Given that each sample requires 8 bits for encoding.
10000 x 0.01 x 8 = 800 bits
bit rate = sampling rate x bits per sample
10000 x 8 = 80,000 bits/second

6.3

The minimum parameters are: LPC filter order , LPC filter coefficients, variance, pitch period and gain.

6.4

Taken P= 10, as the minimum order of the filter, and then encode other parameters with 8bits each; pitch period, variance and gain.
(10 + 1 + 1 + 1) x 8 = 104 bits
This means segment size = 104 bits/segments The lower the number of filter coefficients, the lower the required bit rate, which leads to compression.

6.5

Uncompressed bit rate = 80,000bits/second
Compressed bit rate = segment size x segment rate
segment size = 104 bits/segments
Segment rate = sampling frequency x Block speech size
Segment rate = 10000 x 0.01 = 100 segments/second
Compressed bit rate = 104 x 100 = 10400 bits/second

$$\text{Compression Ratio} = \frac{UncompressedBitrate}{CompressedBitrate}$$

$$\text{Compression Ratio} = \frac{80000 bits/second}{104 bits/second} \approx 7.7$$

This can be described as 7:1 or 1:7 compression ratio.
This means there is a compression factor of 7

6.6

Non-parametric methods: Transform coding.