Question 1

To begin with, we acquired 5 voice samples of a male individual and another 5 samples of a female individual making the five vowel sounds - "a", "e", "i", "o", "u", respectively. In order to get the dataset, we used the "Voice Recorder" software program that was installed by default from Microsoft, as our operating system is Microsoft Windows 10 and the microphone used was the default microphone from the Sony MDR-XB450AP On-Ear Headset.

Question 2

The following plots present the time-domain signals of the different five vowel sounds of a male and a female individual, respectively.

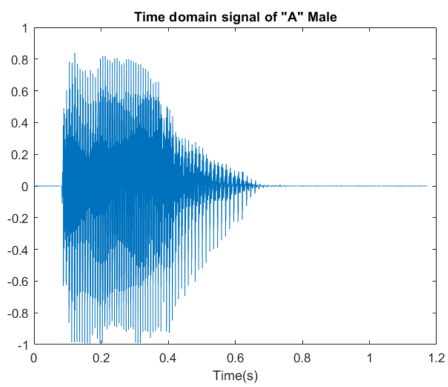

Figure 1. A-Male

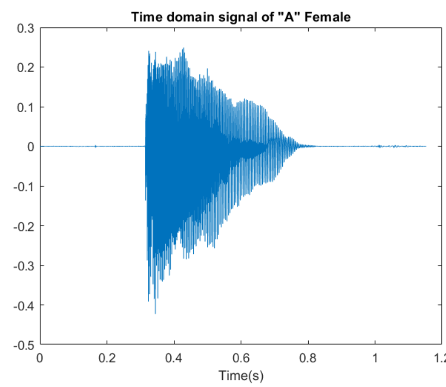

Figure 2. A-Female

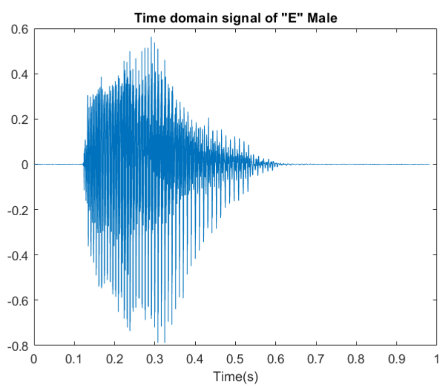

Figure 3. E-Male

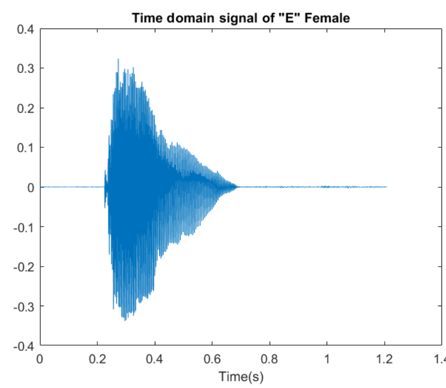

Figure 4. E-Female

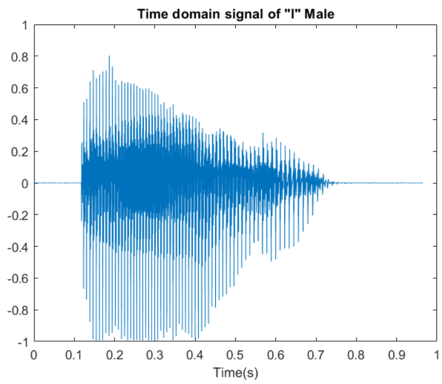

FIGURE 5. I-Male

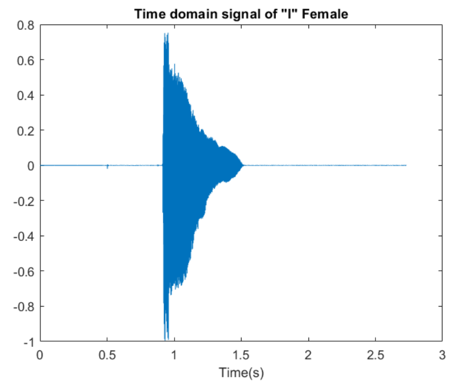

FIGURE 6. I-Female

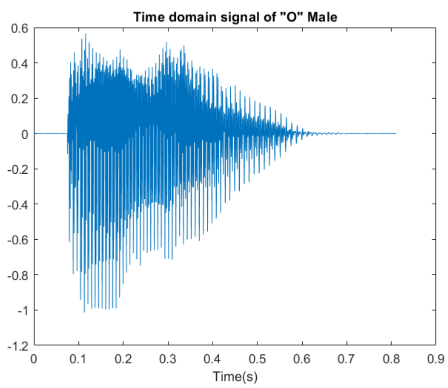

FIGURE 7. O-Male

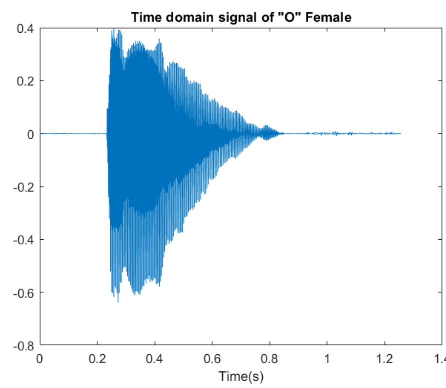

FIGURE 8. O-Female

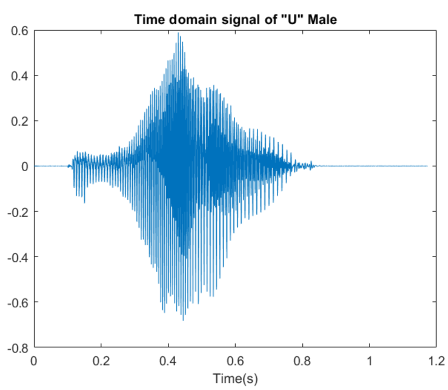

FIGURE 9. U-Male

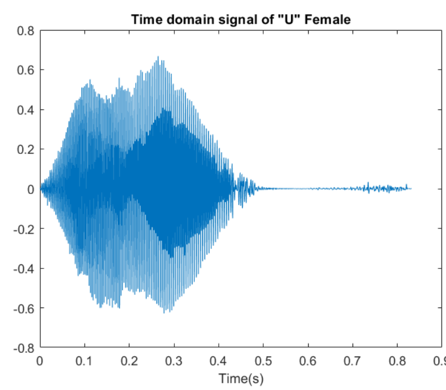

FIGURE 10. U-Female

Human speech is produced by vocal organs. The main energy source is the lungs with the diaphragm. When speaking, the air flow is forced through the glottis between the vocal cords and the larynx to the three main cavities of the vocal tract, the pharynx and the oral and nasal cavities. From the oral and nasal cavities the air flow exits through the nose and mouth, respectively. The V-shaped opening between the vocal cords, called the glottis, is the most important sound source in the vocal system. The vocal cords may act in several different ways during speech. The most important function is to modulate the air flow by rapidly opening and closing, causing buzzing sound from which vowels and voiced consonants are produced. The fundamental frequency of vibration depends on the mass and tension and is about 110 Hz and 200 Hz with men, women, respectively (Source).

In general, women speak at a higher pitch, about an octave higher than men. The fundamental frequency (f0) of the vocal fold vibration determines the perceived pitch of a voice. An adult woman's average range is from 165 to 255 Hz, while a man's is 85 to 155 Hz.
Pitch period is defined as the time interval between two consecutive voiced excitation cycles i.e. the distance in time from one peak to the next peak.
The fundamental frequency is calculated as: 1/(Pitch period).
Fundamental frequency (f0) estimation, also referred to as pitch detection, has been a popular research topic for many years, for speech classification problems.
Cepstral analysis shown provides a way for the pitch estimation, as the cepstrum of voiced speech intervals has strong peaks corresponding to the pitch period.
Given that a woman's fundamental frequency is higher compared to a man, we can hypothesize that the pitch period is going to be smaller, resulting to more peaks in the cepstrum domain in each and every different vowel sound.
We obtained this info from the following paper: (**Gender classification using pitch and formants, Kumar et al., 2011**).
Hence, we got the following plots of the cepstrum domain:
In order to get the following plots, we used the rceps function from the Signal Processing Toolbox.
We tried using the cceps function but the peaks in the cepstrum domain weren't visible at all.

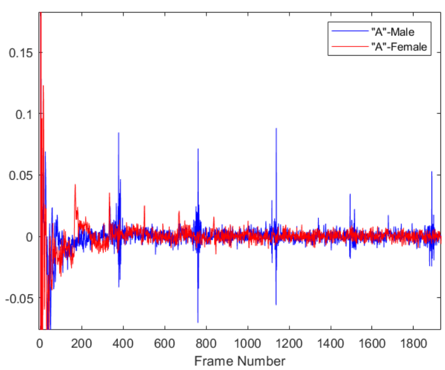

FIGURE 11. A-Female vs A-Male

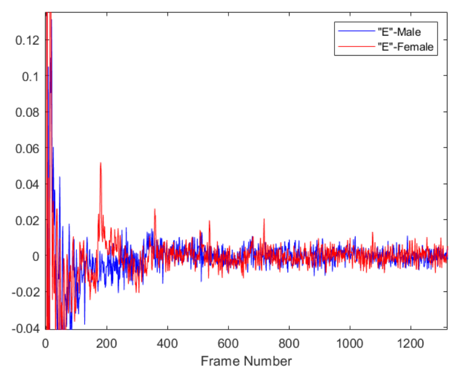

FIGURE 12. E-Female vs E-Male

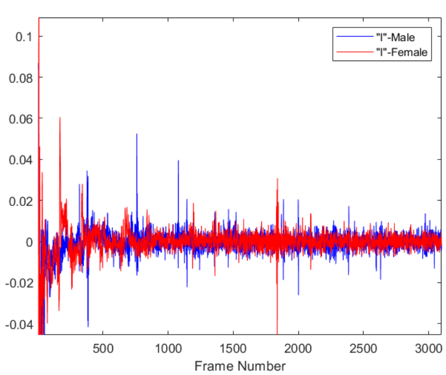

FIGURE 13. I-Female vs I-Male

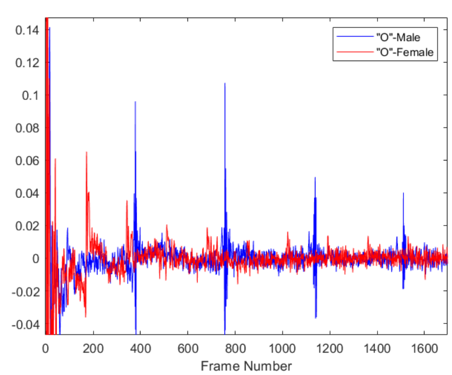

FIGURE 14. O-Female vs O-Male

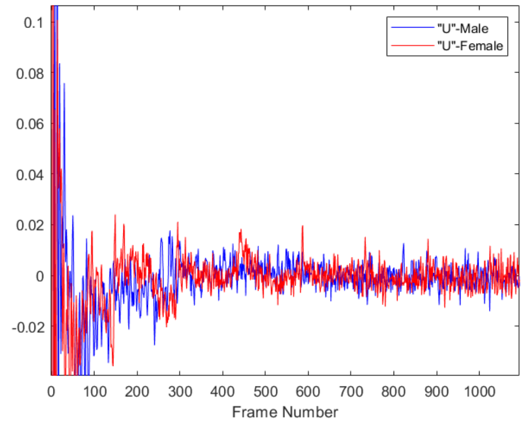

FIGURE 15. U-Female vs U-Male

Our hypothesis is confirmed, as we can clearly see that in each and every vowel the female voice results in more peaks in the cepstrum domain compared to that of the male voice samples.

Question 3

In this query, we are asked to lifter the cepstrum domain signals. The first step in order to do this is to design a window in order to remove the transfer function dependency. The length of the window has to be different for each audio file, depending on the pitch period of each file, respectively.
This length is approximately 2 or 3 times bigger than the pitch period.
The following plot presents the estimation of the pitch period of the vowel "a" of a male voice sample.

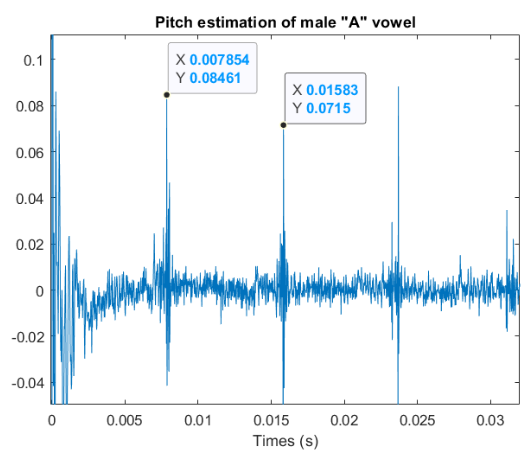

FIGURE 16. Pitch period estimation of males's "a" vowel

We can observe that the pitch period is approximately 11 ms. Thus, we should multiply the speech waveform by a Hamming window, whose length is approximately 2 or 3 times the pitch period.
We will follow the same procedure for each and every voice sample in order to estimate the different pitch periods.
Thus, the length of each window will change every time, depending on the pitch period of each voice sample.

The following plot presents the estimation of the pitch period of the vowel "a" of a female voice sample.

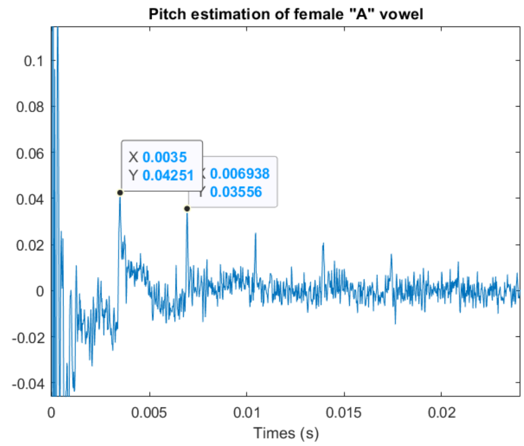

FIGURE 17. Pitch period estimation of females's "a" vowel

We can observe that the pitch period is approximately 3,5 ms. Thus, we should multiply the speech waveform by a Hamming window, whose length is approximately 2 or 3 times the pitch period.

The following plot presents the estimation of the pitch period of the vowel "e" of a male voice sample.

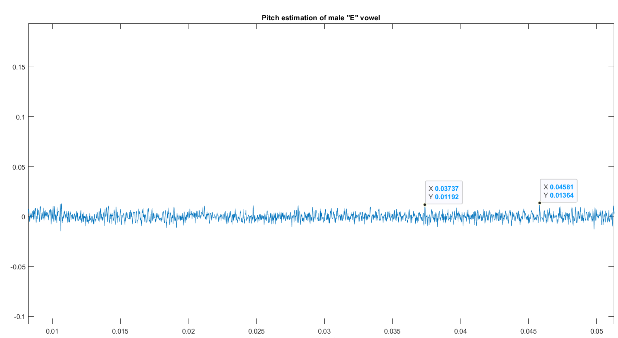

FIGURE 18. Pitch period estimation of males's "e" vowel

We can observe that the pitch period is approximately 8,4 ms, although for this particular vowel the pitch is not clearly visible. Thus, we should multiply the speech waveform by a Hamming window, whose length is approximately 2 or 3 times the pitch period.

The following plot presents the estimation of the pitch period of the vowel "e" of a female voice sample.

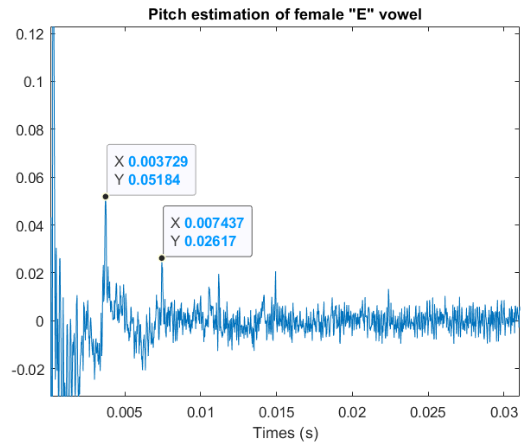

FIGURE 19. Pitch period estimation of females's "e" vowel

We can observe that the pitch period is approximately 3,7 ms. Thus, we should multiply the speech waveform by a Hamming window, whose length is approximately 2 or 3 times the pitch period.

The following plot presents the estimation of the pitch period of the vowel "i" of a male voice sample.

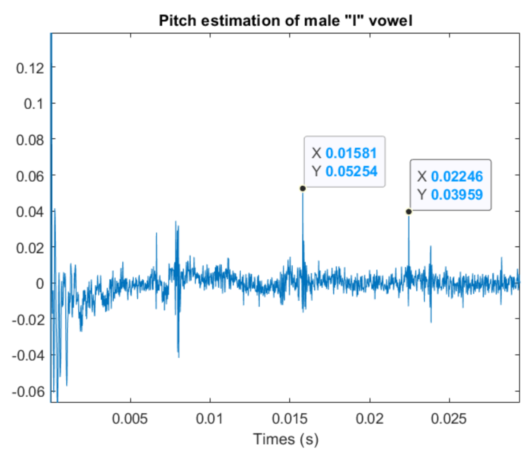

FIGURE 20. Pitch period estimation of males's "i" vowel

We can observe that the pitch period is approximately 6,6 ms. Thus, we should multiply the speech waveform by a Hamming window, whose length is approximately 2 or 3 times the pitch period.

The following plot presents the estimation of the pitch period of the vowel "i" of a female voice sample.

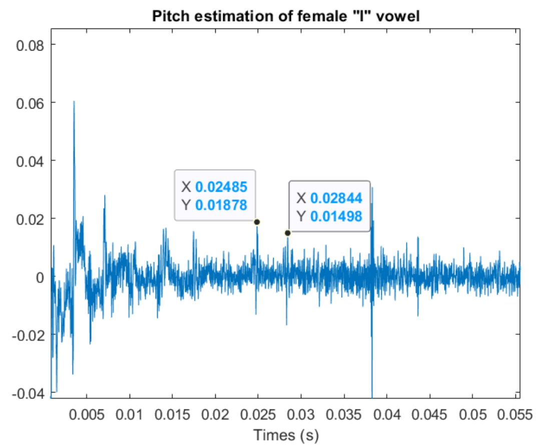

FIGURE 21. Pitch period estimation of females's "i" vowel

We can observe that the pitch period is approximately 3,6 ms. Thus, we should multiply the speech waveform by a Hamming window, whose length is approximately 2 or 3 times the pitch period.

The following plot presents the estimation of the pitch period of the vowel "o" of a male voice sample.

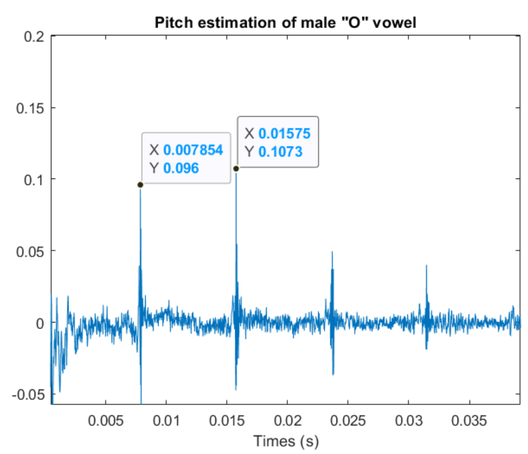

FIGURE 22. Pitch period estimation of males's "o" vowel

We can observe that the pitch period is approximately 7,9 ms. Thus, we should multiply the speech waveform by a Hamming window, whose length is approximately 2 or 3 times the pitch period.

The following plot presents the estimation of the pitch period of the vowel "o" of a female voice sample.

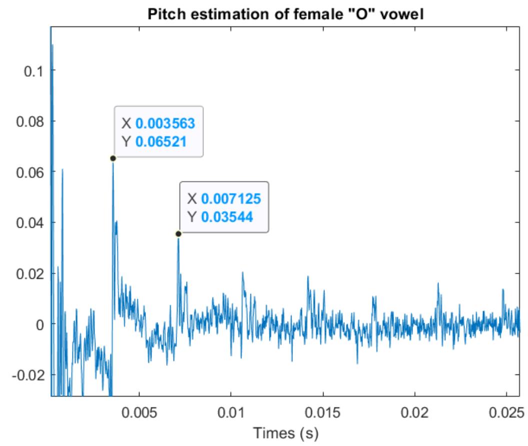

FIGURE 23. Pitch period estimation of females's "o" vowel

We can observe that the pitch period is approximately 3,5 ms. Thus, we should multiply the speech waveform by a Hamming window, whose length is approximately 2 or 3 times the pitch period.

The following plot presents the estimation of the pitch period of the vowel "u" of a male voice sample.

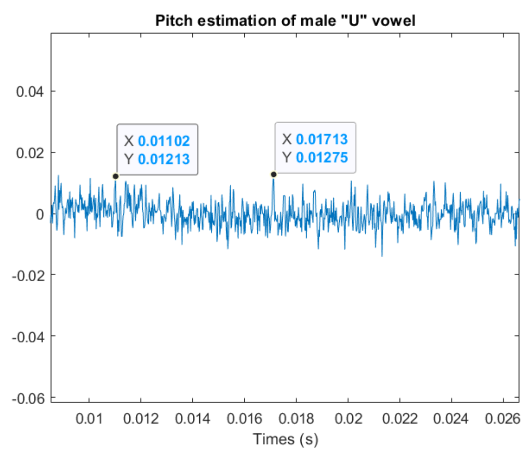

FIGURE 24. Pitch period estimation of males's "u" vowel

We can observe that the pitch period is approximately 6,1 ms, although for this particular vowel the pitch is not clearly visible. Thus, we should multiply the speech waveform by a Hamming window, whose length is approximately 2 or 3 times the pitch period.

The following plot presents the estimation of the pitch period of the vowel "u" of a female voice sample.

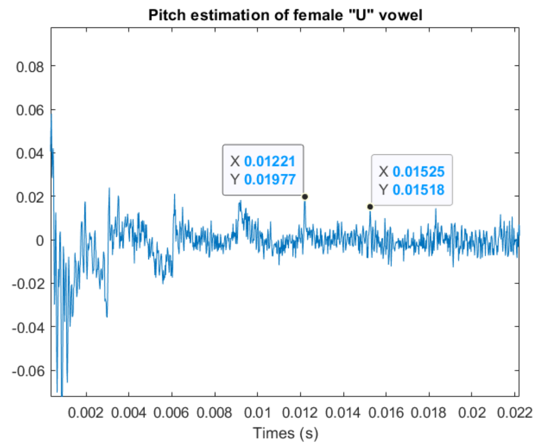

FIGURE 25. Pitch period estimation of females's "u" vowel

We can observe that the pitch period is approximately 3 ms. Thus, we should multiply the speech waveform by a Hamming window, whose length is approximately 2 or 3 times the pitch period.

We showed the procedure of calculating the pitch period of each and every voice sample in order to highlight that each and every different voice sample will be multiplied by a Hamming window of different length. This length tends to be larger for the male's voice samples compared to the female's samples.
After windowing the speech signals, we computed the cepstrum in order to lifter and remove the transfer function dependency.
Then, we computed the deconvolved time-domain signal using the icceps function from the Signal Processing Toolbox.
The following plot compares the original vs the deconvolved signal of the vowel "a" of a male voice sample.
We followed the same procedure for each and every vowel.

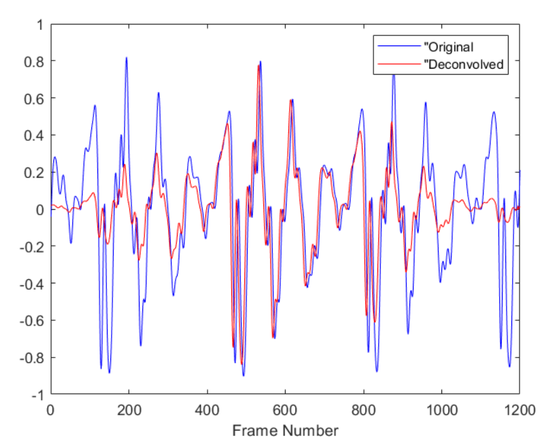

FIGURE 26. Original vs deconvolved signal of males's "a" vowel

Question 4

In the latter question we deconvolved the time domain in order to remove the transfer function dependency.
Here, we tried to synthesize back the voice sample by convolving the transfer function with the decolvolved signal.
More specifically, we synthesized back the voice sample of male's "a" vowel and we got moderate results.

Matlab code

Follow the homework assignments on github.