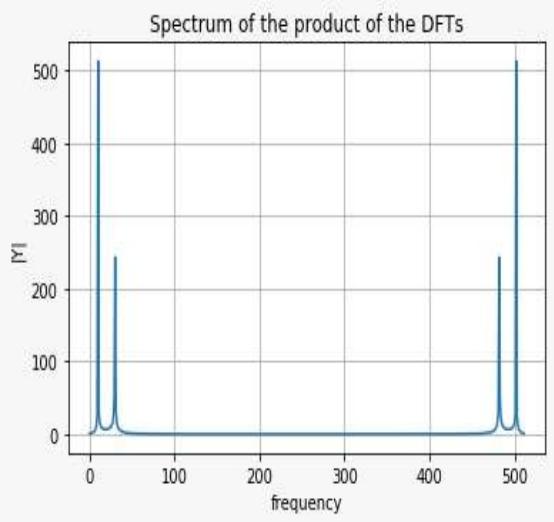
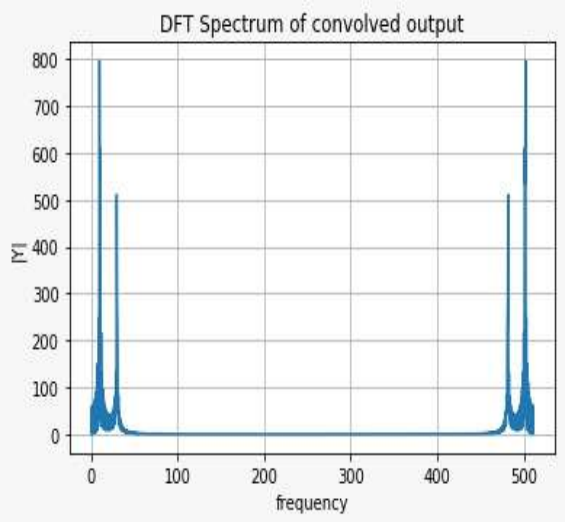
Assignment 2 – Speech Technology

By Abhishek Sekar EE18B067

& Keshav Rao MM18B021

Question 1:

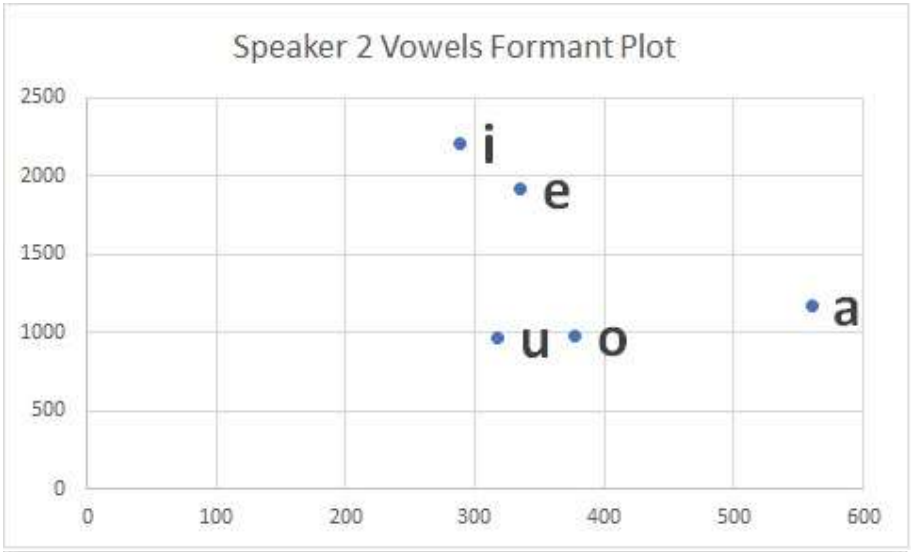


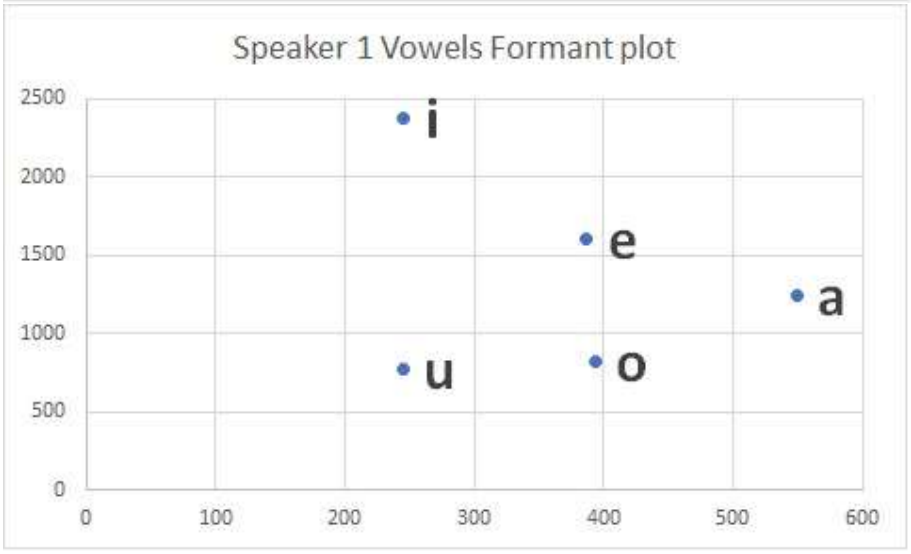


The graphs are Fourier magnitude spectrum plots. As can be seen, the peaks in the spectrum occur at the same frequency in both the plots. This confirms what we'd learnt in class while discussing the properties of the Fourier transform. I.e., convolution in time domain is similar to a multiplication operation in Fourier domain.

The magnitude of the peaks is different due to the associated scaling factor.

Question 2 & 3:





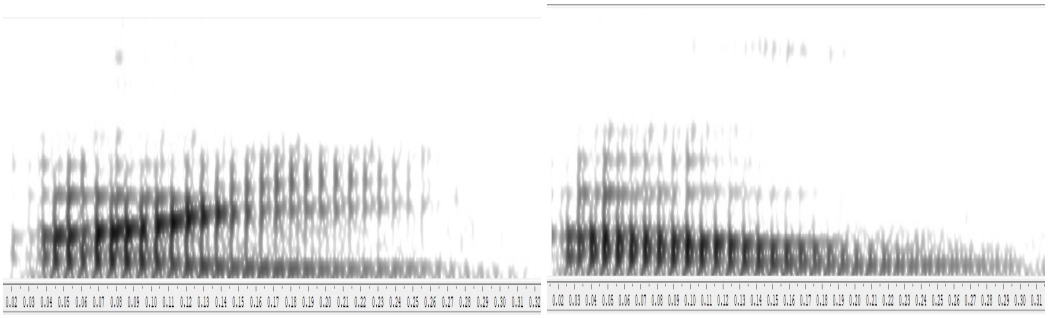
- These are F2 vs F1 peaks for isolated vowels recorded by the team members.
- These should ideally be triangles but perhaps due error in silence correction we are getting distorted shapes.

378.9447	2186.895	ka	unvoiced
305.8596	2158.956	ki	unvoiced
428.7381	1110.537	pa	unvoiced
407.0886	1390.3	ga	voiced
278.2983	2074.681	gi	voiced
481.4015	1037.827	ba	voiced

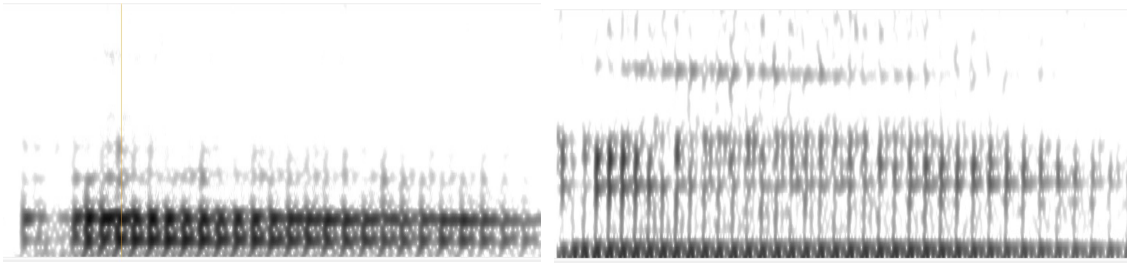
These are mean F1 & F2 values for some of the voiced and unvoiced vowels.

- Different unvoiced consonants change the formant frequencies differently as seen from the table above. However, the voiced consonants don't change the formant frequencies much when compared to the unvoiced ones. I.e., the deviation of formant frequencies of the vowel to that of the voiced consonant + vowel comparatively is small.

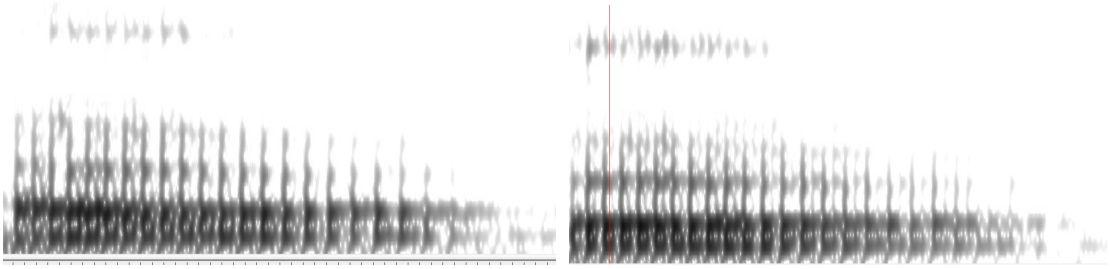
Question 4 & 5:



The above spectrograms are for **ai** and **ao.**



The above ones are for **a** and **i**.



And finally, for **sa** and **ha.**

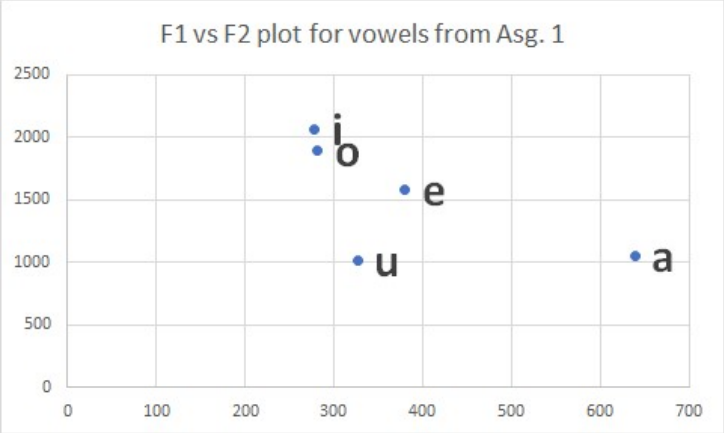
- As the window length is varied, we observe changes in the spectrogram as different lengths of the wav file is considered.

- For very small window lengths, we see vertical dark and bright strips as small changes in energy is also detected.
- For very large window lengths, horizontal dark and bright strips are observed due to the presence of different harmonics in the sound.
- In speech detection, we want the window size to be somewhere in between the two extremes as thick black strips which represent formant peaks are the most important for feature extraction.
- The distance between two adjacent vertical strips will be equivalent to the pitch period of the original sound.
- When the window size is two times pitch period instead of one, then the formant peaks will be more visible and the pitch information will be less visible.
- The code for finding pitch period from wav files has also been submitted.

Question 6:

- We find that as the frequency of the tone is increased the amplitude required for the sum of the noise and the tone also increases for it to be audible.
- As the frequency of the tone goes above a certain value (about 2000 Hz), no value of amplitude was enough for the sum to be audible.

Question 7:



- This formant plot is obtained from speaker 2 (Abhishek's) dataset.
- We observe that this plot forms a new formant triangle where there is a noticeable change in the second formant and not that drastic of a change in the first formant for most vowels.
- For the vowel O alone, there's a massive difference between the pure vowel's formants and when it's squeezed between consonants in the above plot.