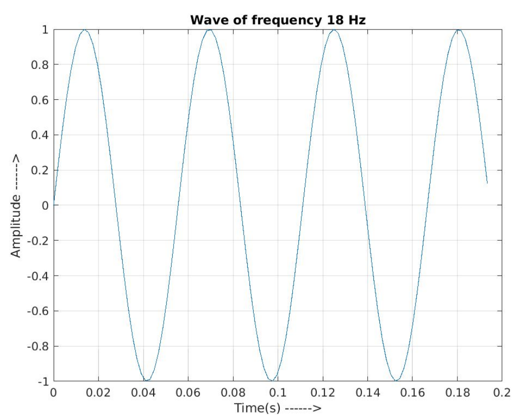
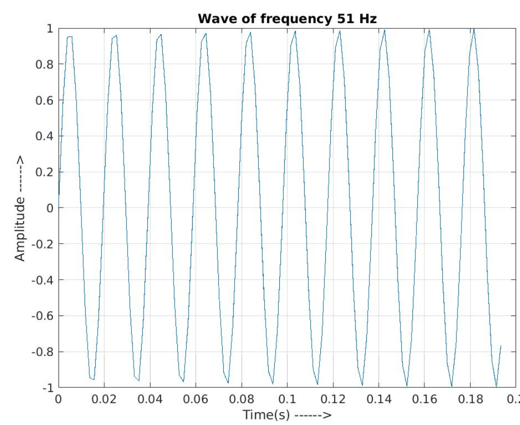
LAB ASSIGNMENT 2 & 3

V Sai Krishna : EE17B035 || Kommineni Aditya : EE17B047

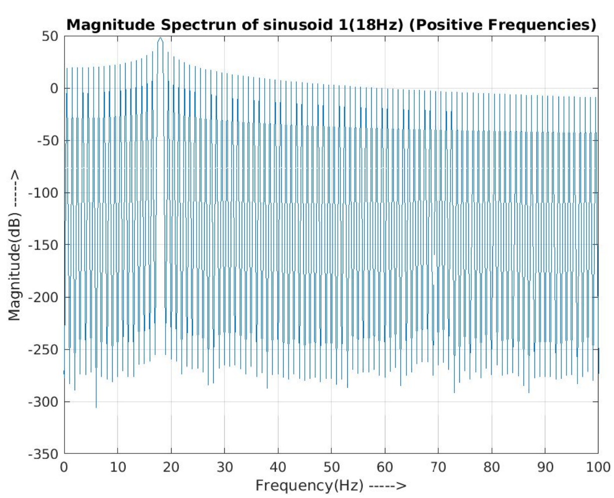
QUESTION 1

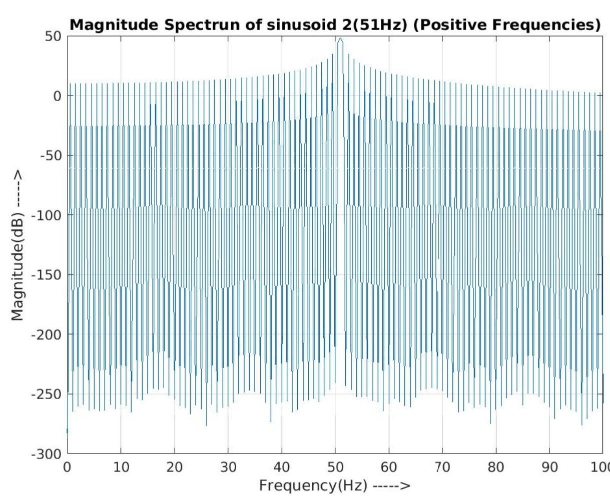
Below are the plots of the frequencies which have been used to calculate the convolution and the corresponding DFT's. Matlab uses the function fft() in order to calculate the circular DFT's of functions. Therefore, we need to keep in mind that adequate zero padding of signals should be done prior to calculating fft if we intend to calculate the linear convolution.



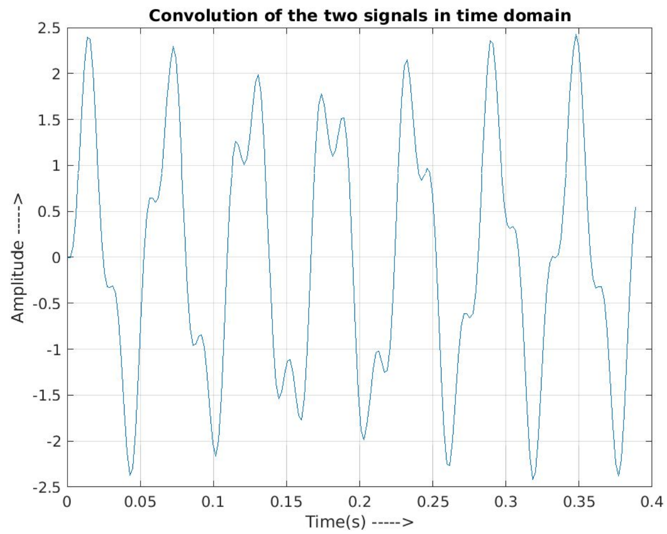


We sample both the signals at 512 data points as asked in the question. Then, we calculate DFT of both the signals above which have been padded with 512 zeros.

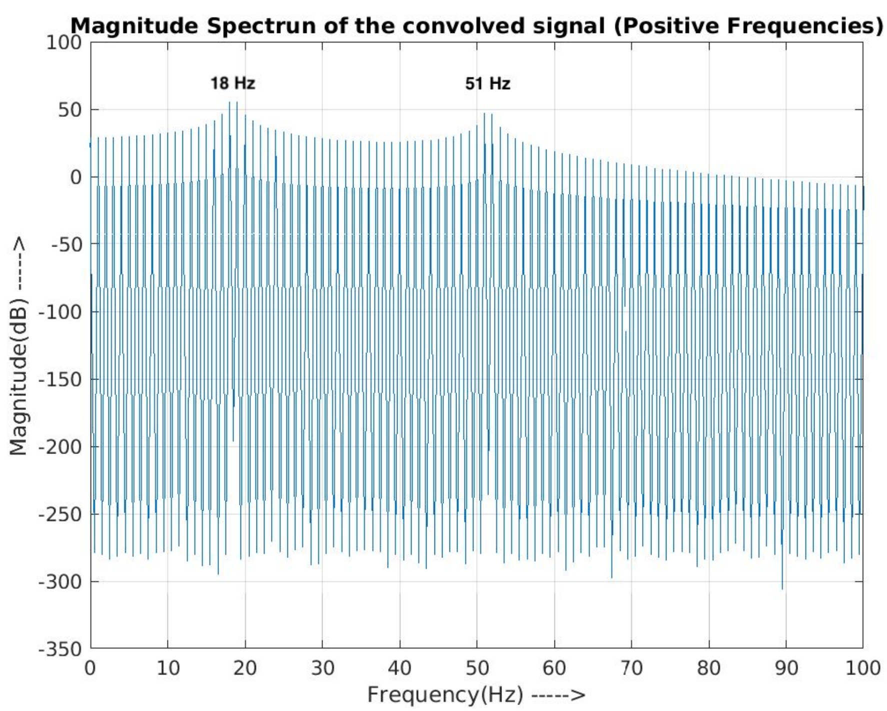




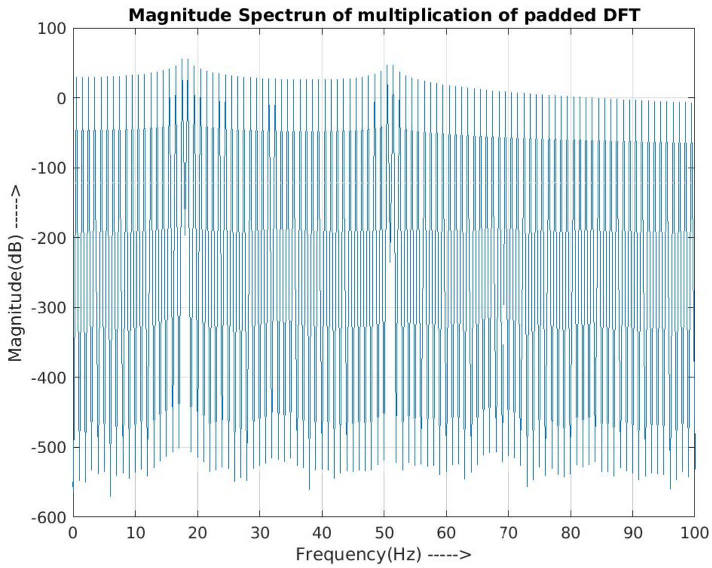
Below is the plot of the convolution of the two signals using the conv() function. The convolved signal is of the length 1023 points as is expected from a linear convolution.

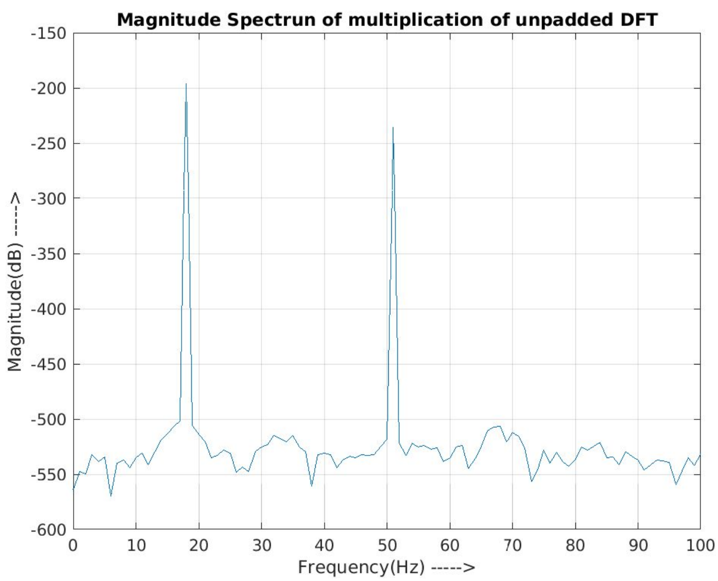


We then calculate the DFT of the convolved signal using the fft() function wherein we could mention the number of DFT points we need. Here, we take a 1024 point DFT as that is the nearest power of 2 to the convolved signal length. For the sake of illustration, the domain of frequency of interest in the graph has been reduced to [0,100] Hz.



Now, we calculate the DFT of the convolved signal by multiplying the DFT's of the padded signals. We can see by comparing both the pictures that when the sinusoids are padded then the DFT is in accordance with convolved signal DFT. However, when the unpadded signal DFT's are multiplied, they don't match with convolved signal DFT.



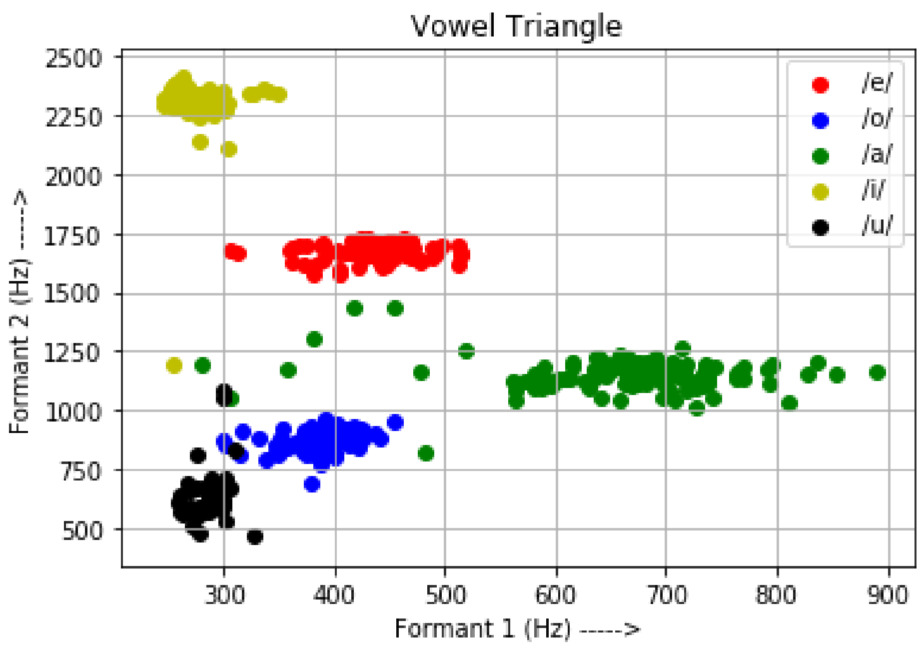


<u>INFERENCE:</u>

Although we expect the DFT of the sinusoid to be two impulses at the respective frequencies, in practice while calculating the DFT, since we window the sinusoids to a particular finite length, the DFT of the window i.e. sinc function brings in the lobes in the convolved DFT. In the un-padded case, the fft calculated is circular hence, we don't get the desired output.

QUESTION 2

As given in the question, we recorded both our voices and obtained the formants of the respective vowels using the wavesurfer tool. Below is the graph which shows the range of formants for various vowels.



From the above diagram, we can say that the rough estimates of the formants for the vowels are as given below:

Vowel	Formant 1 (Hz)	Formant 2 (Hz)
/a/	700	1100
/e/	430	1600
/i/	280	2225
/o/	390	825
/u/	280	650

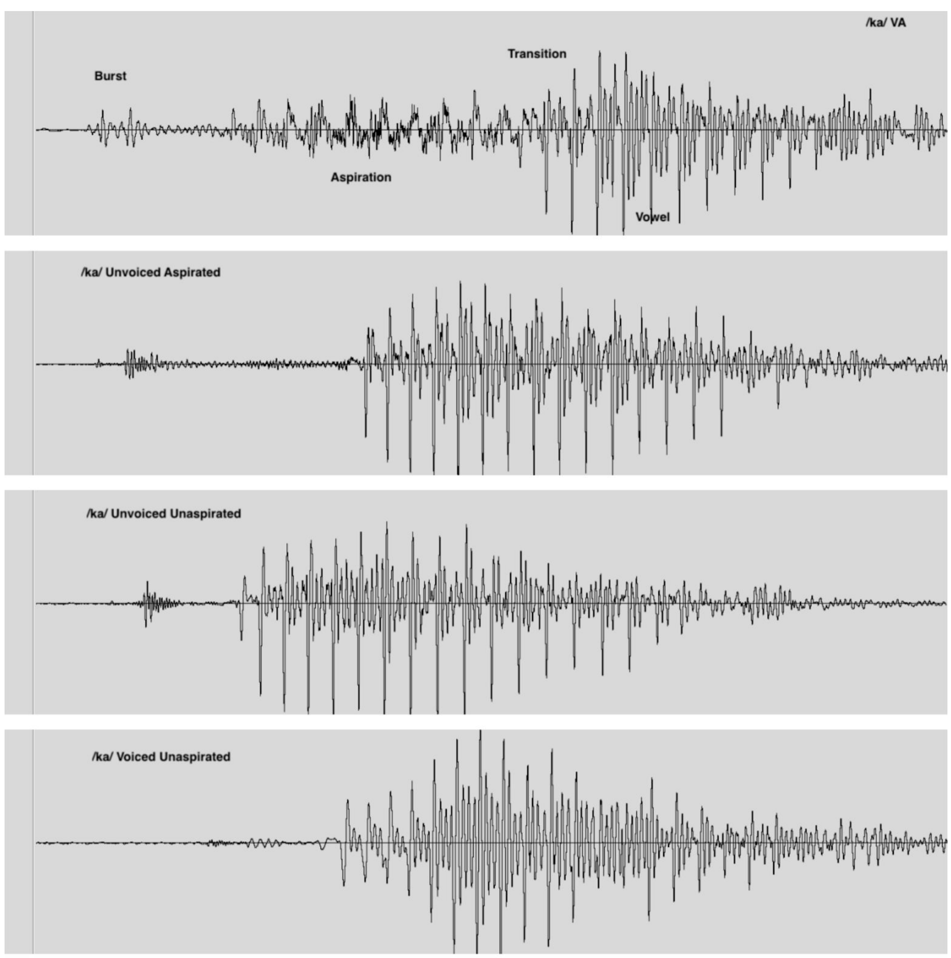
INFERENCE:

The above observations are in accordance with the vowel triangle plot which was provided in the slides. Therefore, the formants value could be used as a parameter for speech recognition. However, we must also take into consideration that there are a range of formant values from which the same phoneme can be muttered. This makes it difficult to identify which phoneme has been uttered as in the case of /o/ and /u/ wherein there is an overlap.

QUESTION 3

Below are the figures which depict the waveforms of the stop consonants /k/, /tch/, /t/, /th/, /p/ suffixed with the vowel /a/. The waveforms of the stop consonants can be divided into four parts, namely burst, aspiration, transition, vowel.

Figure composed /k/ waveforms



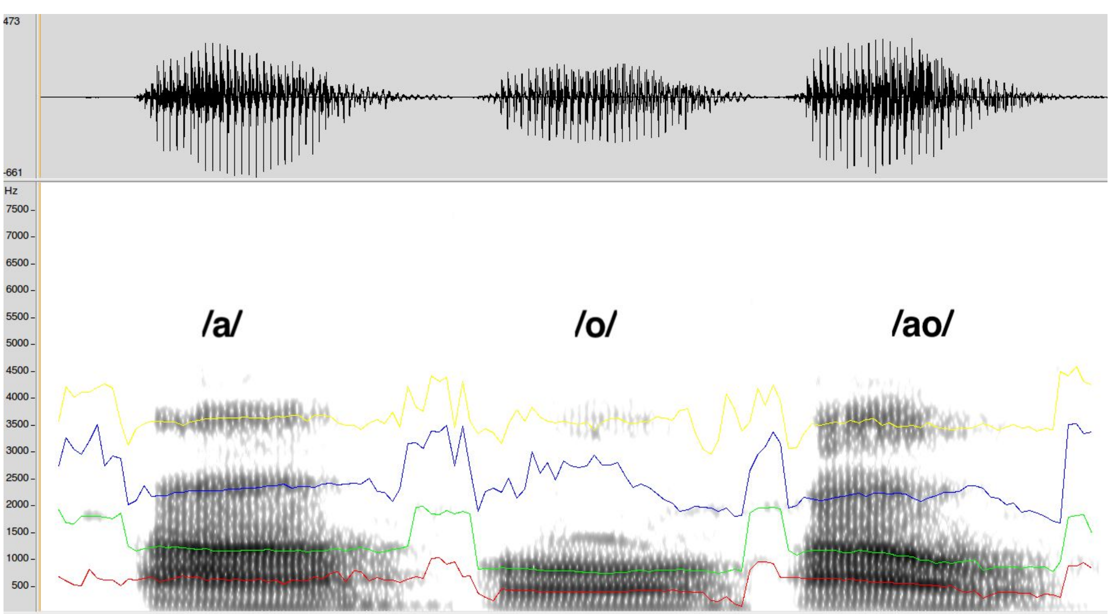
For other waveforms refer to Page numbers : 14-18

INFERENCE:

We can observe that the consonants aren't demarcated as clearly as the vowels are and the time duration of the consonants is small when compared to the vowels. Also, voiced stop consonants have a periodic burst whereas the burst is noisy in unvoiced. The burst is followed by silence in case of unaspirated consonants and it isn't in case of aspirated. Similar to the above observations, when suffixed with others vowels like /i/ and /u/, the burst, aspiration and transition appear to be similar to the one as suffixed with /a/. However, the vowel waveform appears to be completely different as is expected.

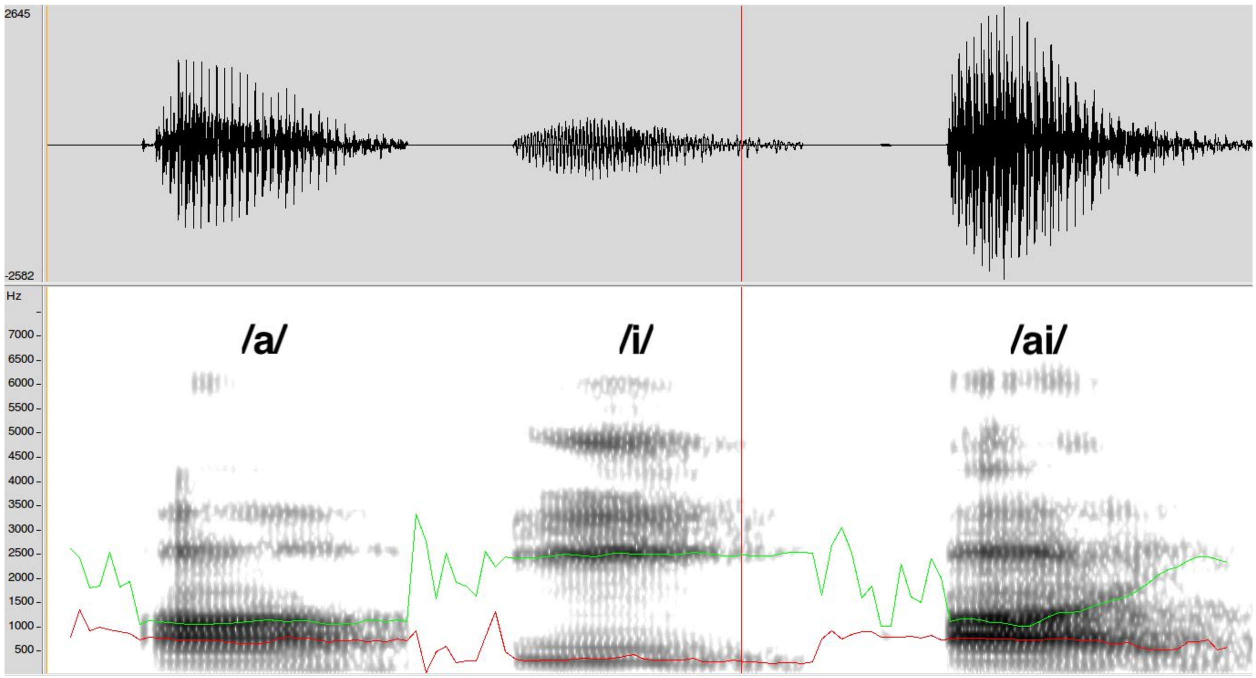
QUESTION 4

Below are the recordings of the individual vowels and the diphthong, their respective formant plots. From the figure below, we can observe that the first vowel in the diphthong /ao/ i.e. /a/ has an almost similar spectrogram plot. The formant frequencies as well are in agreement to the frequencies obtained in Question 2. However, the second vowel spectrogram is different from the original spectrogram. The transition from the first vowel to the second vowel is gradual and there is no abrupt change in the formant frequencies. This could be attributed to the inertial nature of the human vocal cords and vocal tract i.e. there are no sudden changes during the course of speech.



The figure above is the plot of the diphthong /ao/ and its spectrogram

In accordance to the arguments above, the diphthong /ai/ manifests its spectrogram similar to /ao/ i.e. the spectrogram of /a/ is similar when recorded alone and in the /ai/ whereas, spectrogram of /i/ is different from that of /i/ in /ai/. This again points to the fact that there is a period of transition from /a/ to /i/ which can't be characterised by any one vowel well.



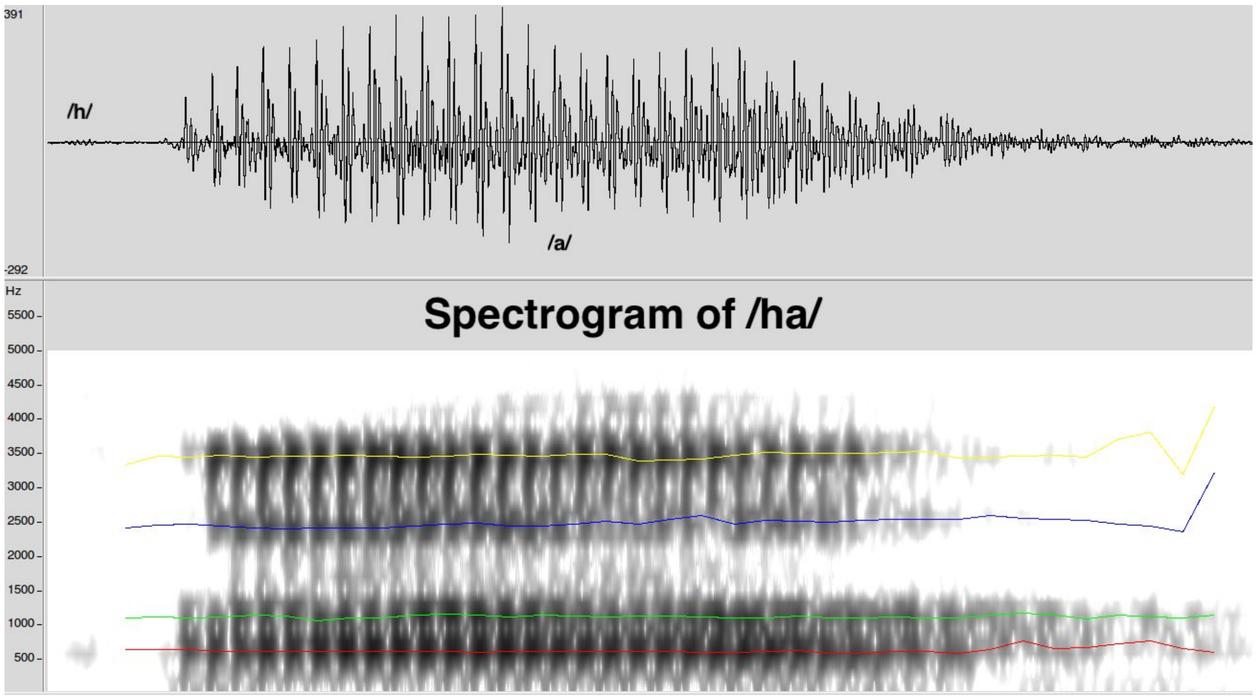
The figure above is the plot of the diphthong /ai/ and its spectrogram

<u>INFERENCE:</u>

From the above observations, we can conclude that the exact formant values of vowels may be lost in course of pronouncing a word i.e. we cannot expect to match the formant frequencies exactly to the frequencies obtained when recorded in isolation. This adds complexity to the speech recognition systems wherein formant identification could make the task relatively easier. When we changed the window size to cover one pitch alone, we could see that the formants were clearly visible as dark bands with minimal vertical striations. Opposed to the previous observations, when more than two pitch periods were included, there was a gradual increase in the vertical striations which led to distortion in the bands of formants and thereby couldn't be distinguished. This shows that an optimal window size is extremely important to capture the correct spectrograms and generate proper features for recognition.

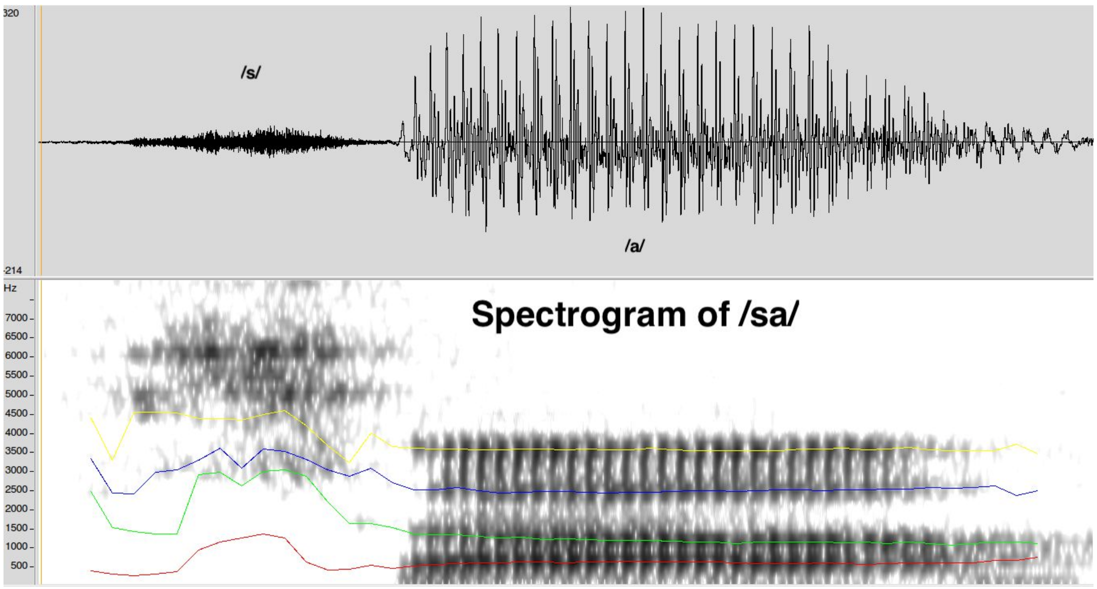
QUESTION 5

The spectrogram of both the fricatives /s/ and /h/ suffixed with the vowel /a/ are as depicted below.



As depicted in the picture above, there is no clear distinction between the formants of /h/ and /a/. The consonant and the vowel have almost continuous formants in the spectrogram. Moreover, in the time domain apart from the starting burst there is a very small transition which makes it more difficult to identify the /h/.

Unlike the above case, /sa/ has a clear demarcation between the consonant and vowel formants. The consonant formants are much higher than that of the vowel thereby giving a clear distinction between them.



INFERENCE:

From the above we can see that /s/ can be easily recognised to be distinct from the vowel as it has different set of formant ranges whereas /h/ can't be easily detected by a speaker recognition system employing formants alone as the formants for /h/ and /a/ are almost the same.

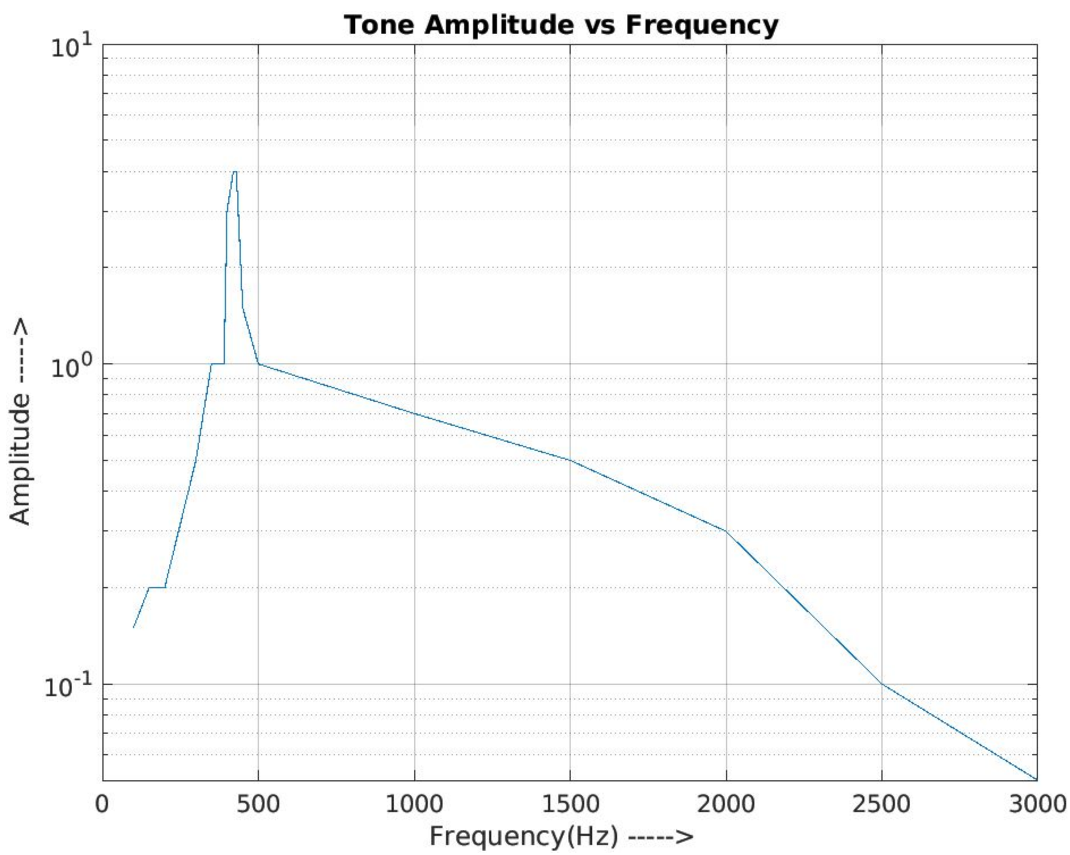
QUESTION 6

The below data has been recorded with one of the team member was unaware of the frequency of the tone and the amplitude, the other team member varying the parameters accordingly.

The amplitude of the noise is 1. The noise is a sum of sinusoids from the frequency 350 Hz to 440 Hz taken at regular intervals. Amplitude in the second column refers to the value at which the tone was clearly perceived.

Frequency of Tone (Hz)	Amplitude
100	0.15
150	0.2
200	0.2
300	0.5
350	1

390	1
400	3
420	4
430	4
450	1.5
500	1
1000	0.7
1500	0.5
2000	0.3
2500	0.1
3000	0.05

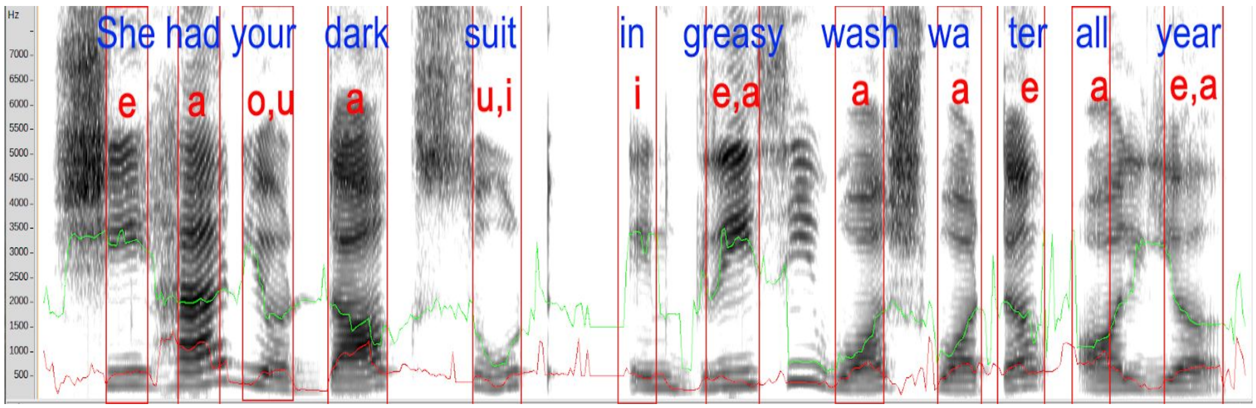


INFERENCE:

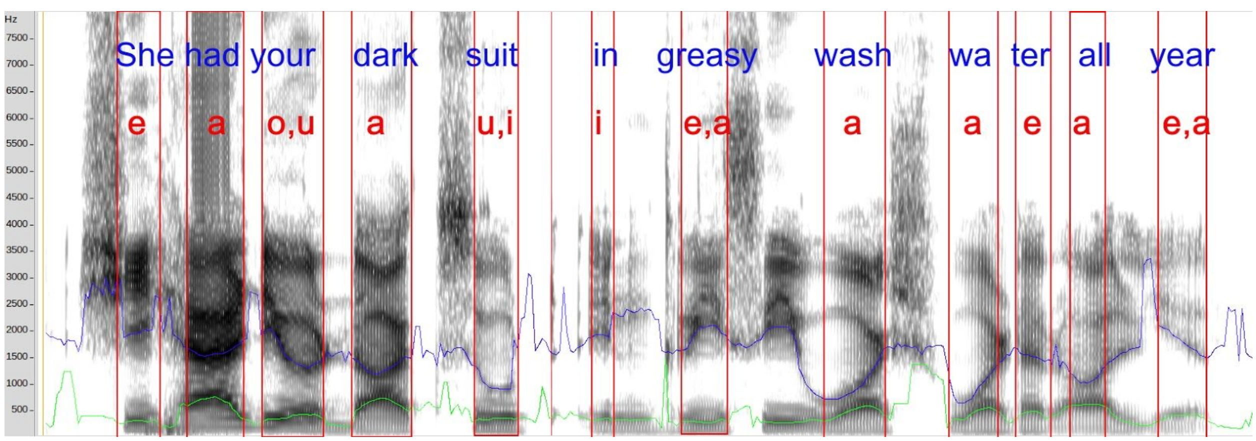
From the data and the corresponding graph as listed above, we can clearly observe that the amplitude of the tone required to clearly distinguish it from the noise is higher for the frequencies which lie in the range of the noise frequencies. The curve nearly resembles a bell shape.

QUESTION 7

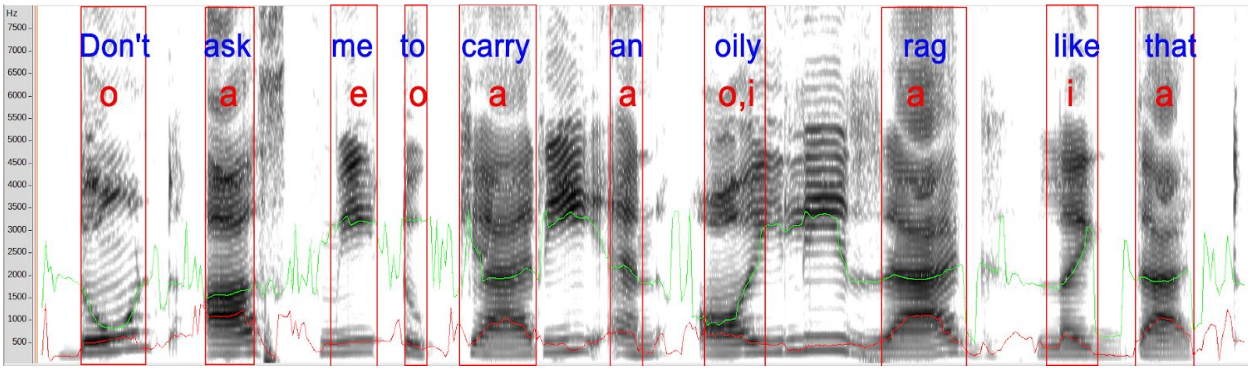
The following are the formant plot of two timid sentences SA1 and SA2 spoken by a male and a female speaker. In those the boundaries of the formants of the vowels have been marked.



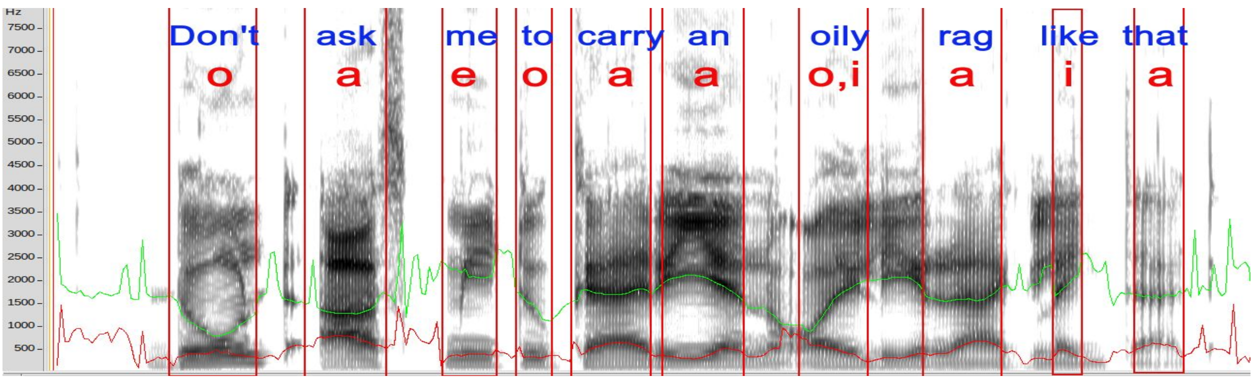
SA1 sentence by a female speaker



SA1 sentence by a male speaker

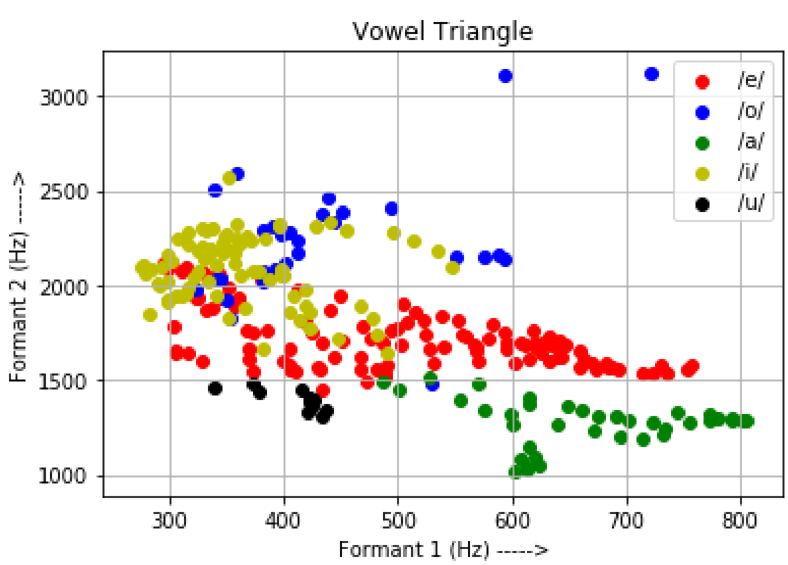


SA2 sentence by a female speaker



SA2 sentence by a male speaker

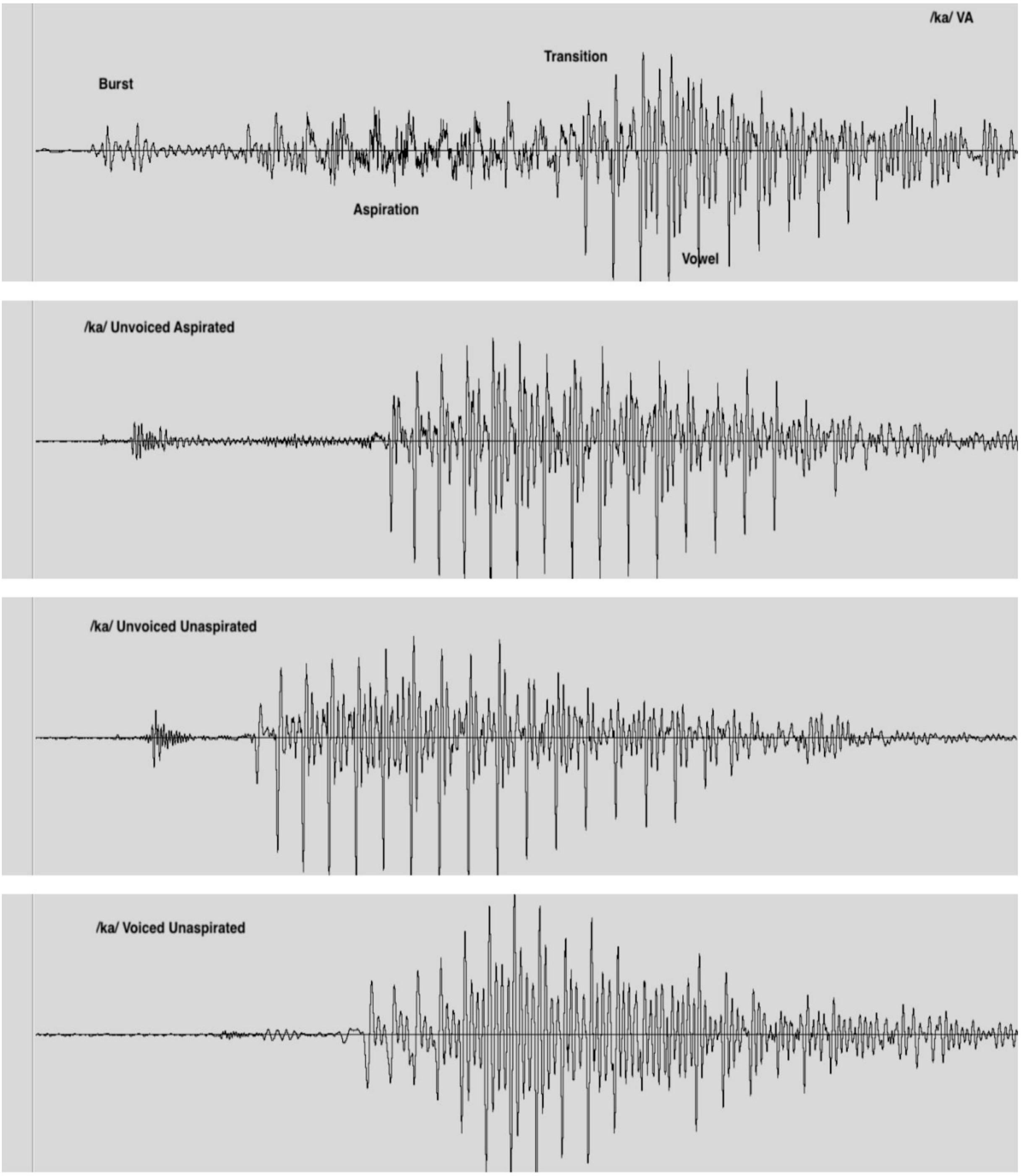
The extracted formant values for each of the vowels have been used to plot the vowel triangle which is as depicted below.



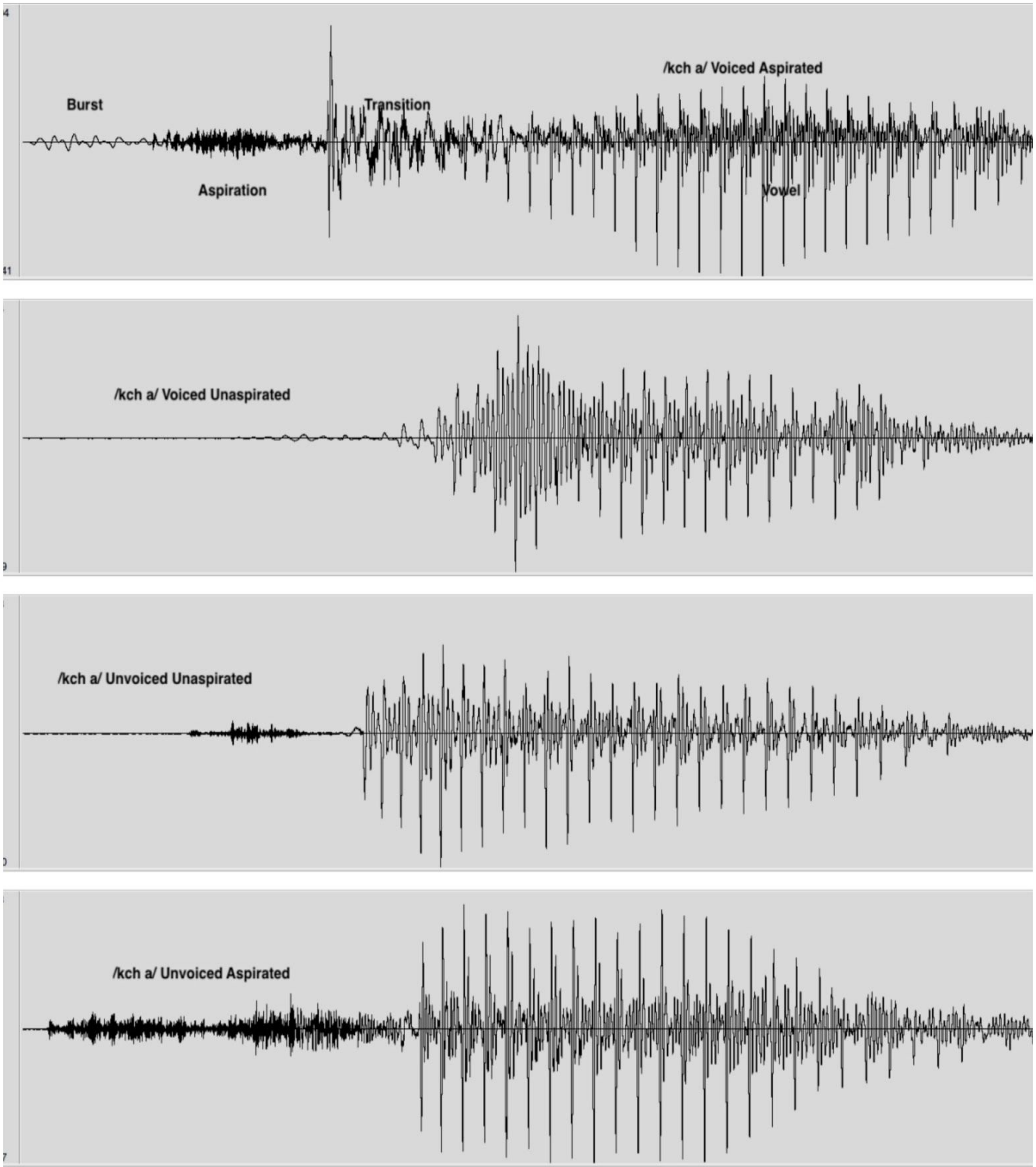
INFERENCE:

From the vowel triangle obtained, we can clearly observe that it has some discrepancies with that obtained in Question2 wherein those values were obtained from isolated recordings of the vowels. This is owing to the fact that vocal tract is an inertial system and the formants of vowels are affected by the previously uttered sounds.

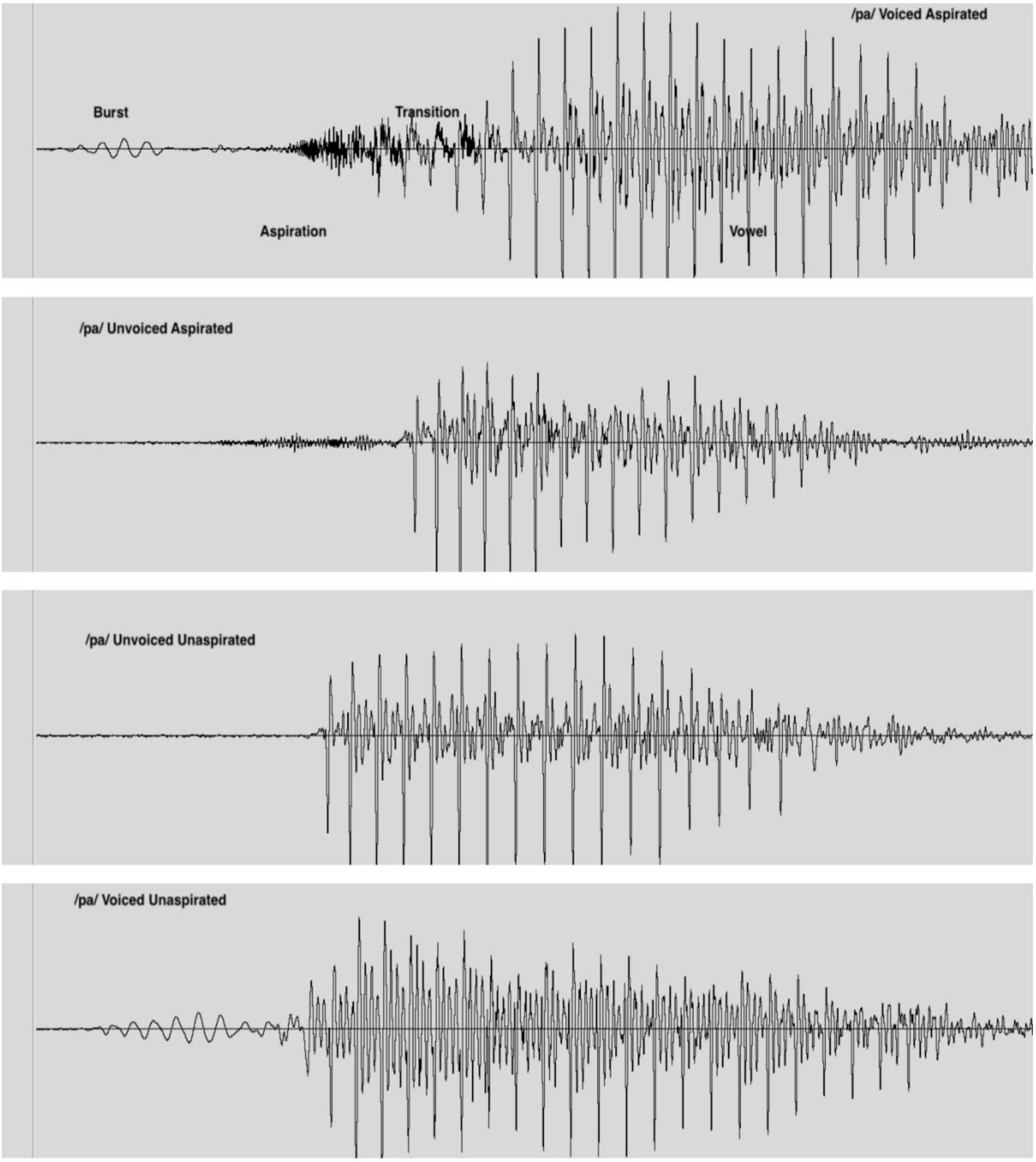
IMAGES OF THE WAVEFORMS OF STOP CONSONANTS (Question 3)



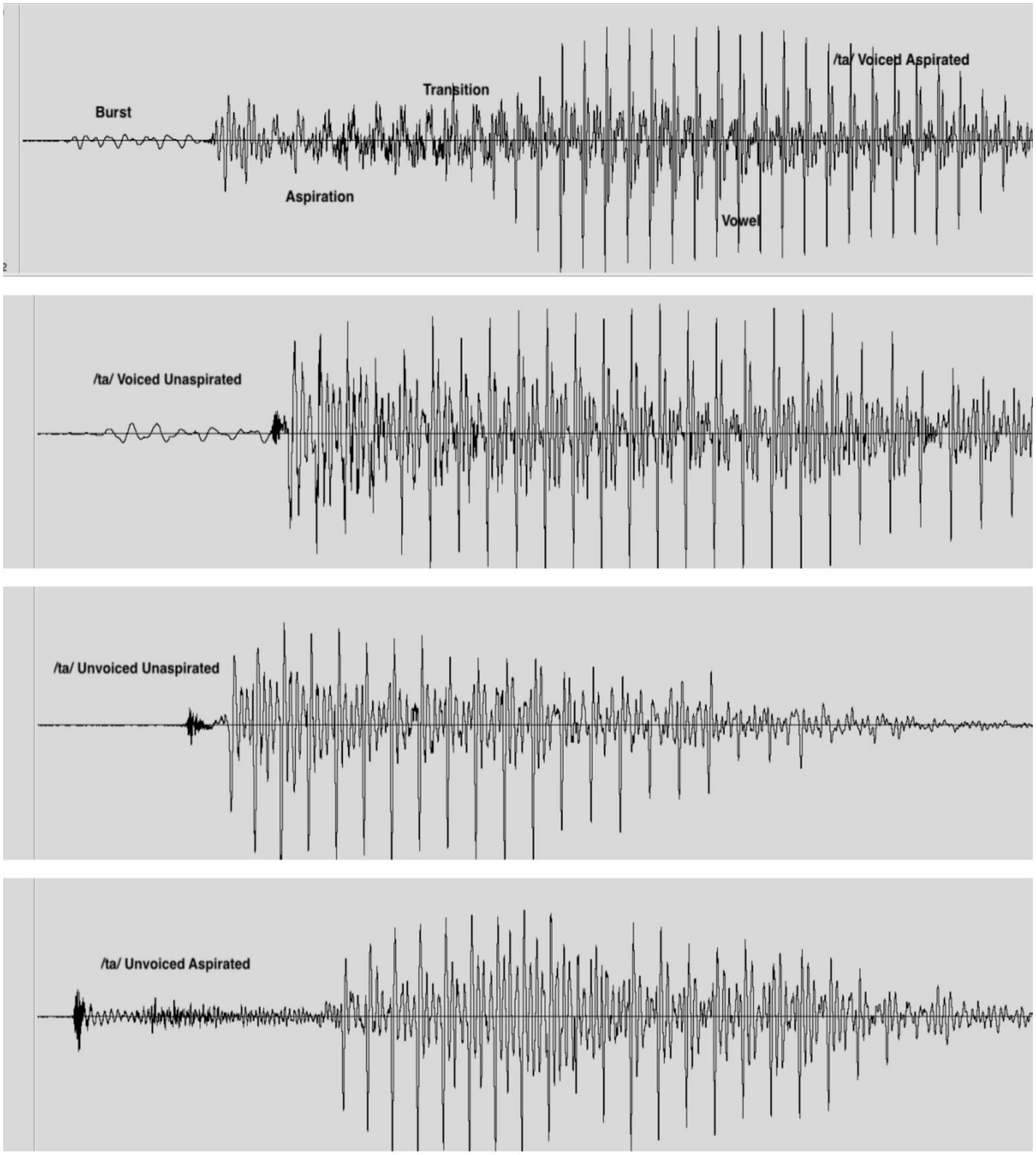
Waveforms of /k/



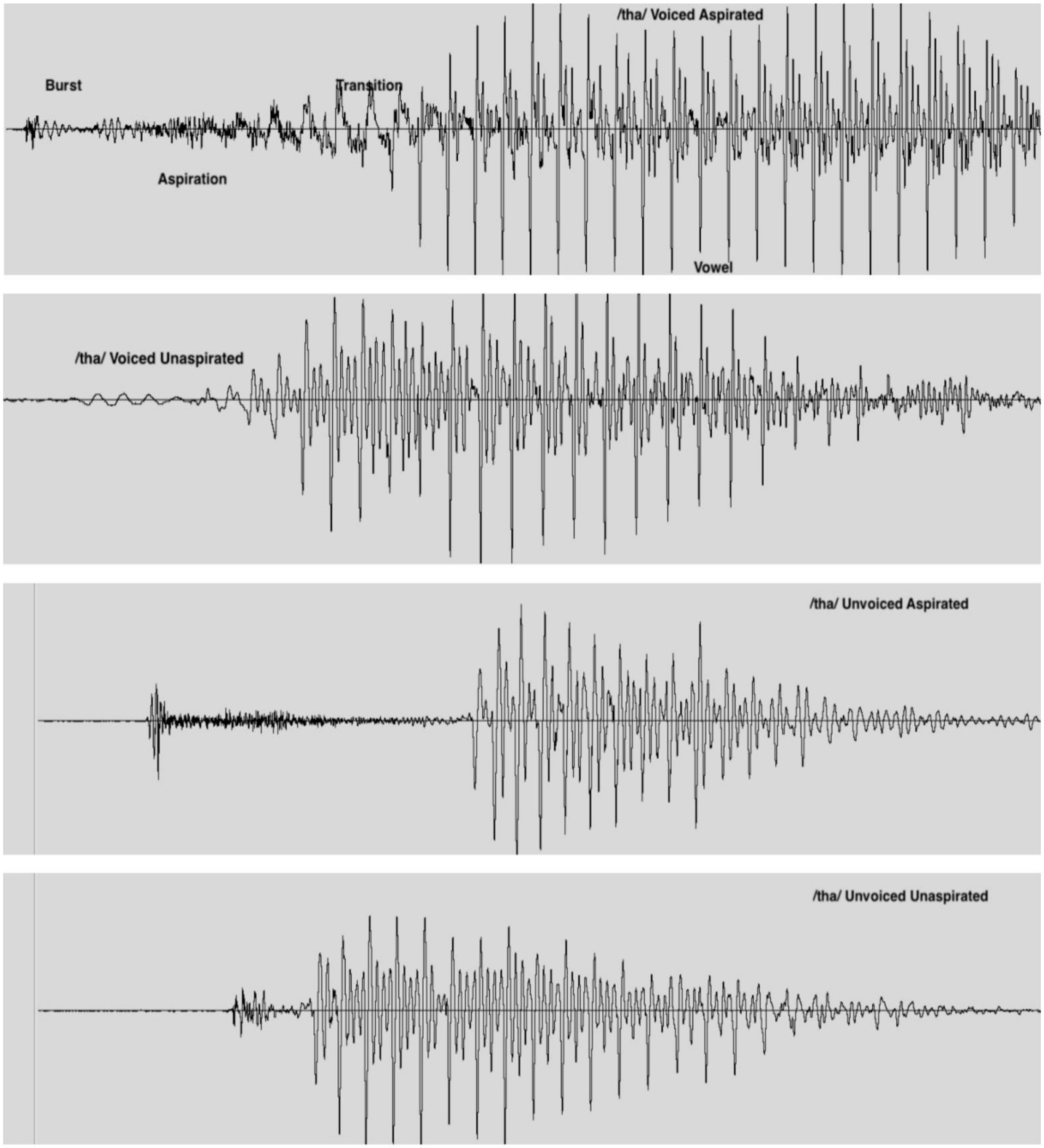
Waveforms of /tch/



Waveforms of /p/



Waveforms of /t/



Waveforms of /th/