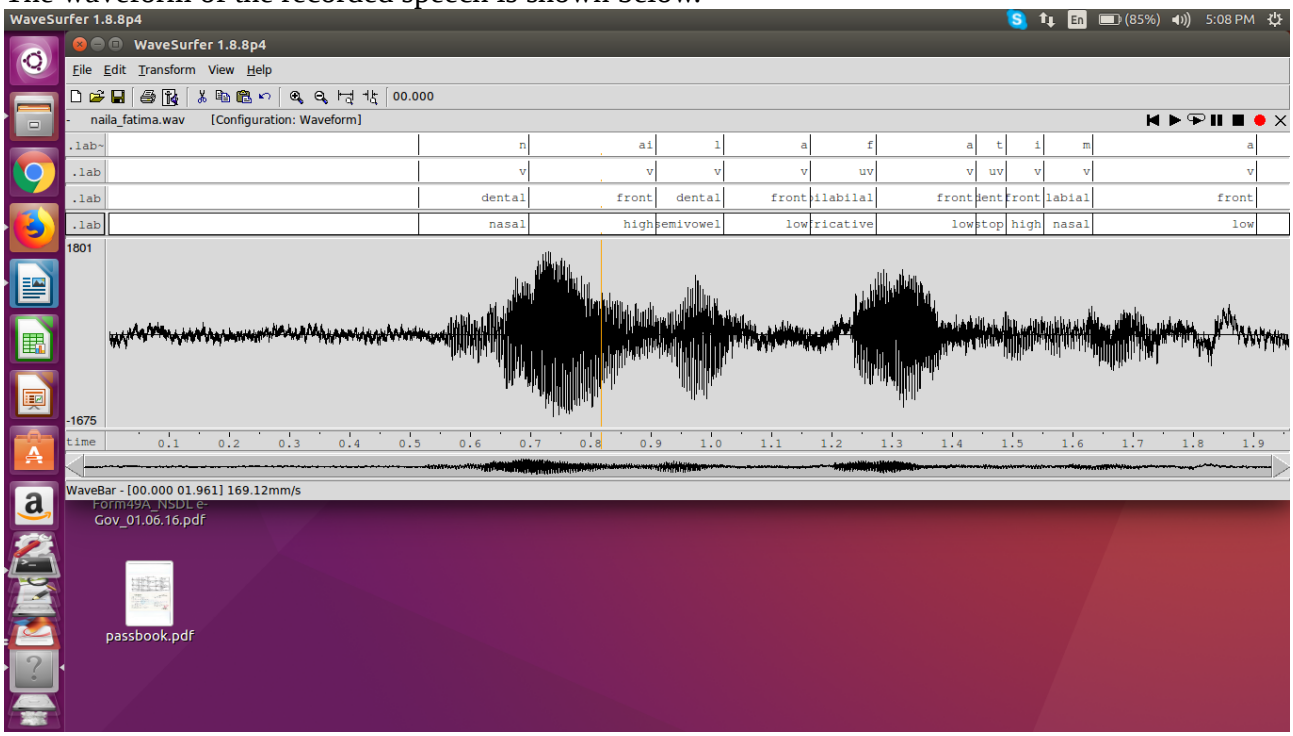
Assignment 1 of Speech Signal Processing
Naila Fatima
201530154

Question 1

The wav file, *naila_fatima.wav*, contains a wavesurfer recording of me saying my name. The waveform of the recorded speech is shown below.



Phonemes: After recording my name, I have distinguished 10 phonemes constituting my name. These phonemes include: the nasal ‘n’, the diphthong ‘ai’, the semivowel ‘l’, the vowel ‘a’, the unvoiced fricative ‘f’, the vowel ‘a’, the unvoiced stop ‘t’, the vowel ‘i’, the nasal ‘m’ and the vowel ‘a’. We can see that among these 10 phonemes, there are:

i) 4 consonants which include ‘n’, ‘f’, ‘t’ and ‘m’. We know that consonants include nasals, stops, fricatives and affricatives. In these categories, ‘n’ is a nasal, ‘f’ is a fricative, ‘t’ is a stop and ‘m’ is a nasal.

ii) 1 diphthong which includes ‘ai’. This diphthong is pronounced as the ‘ai’ in the word ‘buy’.

iii) 1 semivowel which is ‘l’.

iv) 4 vowels which include 3 instances of ‘a’ and one instance of ‘i’.

The phonemes have been transcribed and saved in the *phoneme.lab* file.

Voice / unvoiced decisions: The phonemes which are present in my name have further been classified as voiced and unvoiced. Voiced sounds occur when the vocal cords are tense and the air flow causes them to vibrate whereas unvoiced sounds occur when the vocal cords are relaxed and the air flow becomes turbulent by passing through a constriction or pressure which is build behind a point of closure within the vocal tract is suddenly released. It should be noted that all vowels and diphthongs are voiced as the oral cavity is open. Among the 10 phonemes in my name, 2 are unvoiced and 8 are voiced. The unvoiced phonemes are ‘f’ and ‘t’ which are a fricative and stop, respectively. The unvoiced fricative ‘f’ is produced by having a steady air flow become turbulent near the lips, which are constricted. The unvoiced stop ‘t’ is produced by a building up of pressure (the vocal cords remain relaxed) which is suddenly released.

The voiced phonemes include the phonemes ‘n’, ‘l’ and ‘m’ as well as the vowels ‘a’ and ‘i’ and the diphthong ‘ai’. The voiced and unvoiced decisions have been transcribed and saved in the *voiced.lab* file.

Manner of articulation for consonants/ high/low for vowels: Among the consonants, the phonemes ‘n’ and ‘m’ are nasal, ‘l’ is a semivowel, ‘f’ is a fricative and ‘t’ is a stop. Among the vowels and diphthongs, the ‘ai’ and ‘i’ phonemes are high whereas the ‘a’ phoneme is low.

The above information has been transcribed and saved in the *moa.lab* file.

Place of articulation for consonants/ front/mid/back for vowels: Among the consonants, the phonemes ‘n’, ‘t’ and ‘l’ are dental and ‘f’ and ‘m’ are bilabials (according to Wikipedia, ‘f’ is a voiceless labiodental fricative which means that it requires the use of the lower lip) . Among the vowels and diphthongs, the ‘ai’, ‘i’ and ‘a’ phonemes are all front.

The above information has been transcribed and saved in the *poa.lab* file.

Phonemes	Voiced/unvoiced	MOA	POA
‘n’	v	nasal	dental
‘ai’	v	Diphthong / high	front
‘l’	v	semivowel	dental
‘a’	v	Vowel / low	front
‘f’	uv	fricative	bilabial
‘a’	v	Vowel / low	front
‘t’	uv	stop	dental
‘i’	v	Vowel / high	front
‘m’	v	nasal	Bilabial
‘a’	v	Vowel / low	front

Question 4

Coarticulation: Coarticulation refers to the manner in which the articulation of speech sounds together causes each distinct sound to be influenced by its adjacent sounds. It is basically the manner in which an independent speech sound becomes more similar to the sounds adjacent to them when used along with them. For example, in the word ‘swing’, the ‘n’ phoneme tends to be alveolar whereas in words like ‘nap’, it tends to be dental in nature. The phoneme ‘n’ has changed itself according to the phonemes around it. The phenomenon of coarticulation makes it difficult for an individual to realize the beginning and ending of a phoneme when it is used in a syllable or word (in conjunction with other phonemes).

Phonation: Phonation refers to the manner in which speech sounds are produced. It is a process by which the quasi-static vibrations of the vocal fords present in the larynx (voicebox) causes the production of sounds. Phonation occurs when air passing through the windpipe causes the vocal cords to vibrate and depending on the positions of the tongue, lips and velum, different sounds are created.

Fundamental frequency: Fundamental frequency refers to the frequency at which the vocal cords are opening and closing. It is the frequency component which has the highest magnitude in the frequency spectrum. It can also be considered to be the lowest frequency of a periodic waveform. The other harmonic frequencies present in the waveform will be integral multiples of the

fundamental frequency. The fundamental frequency of the average adult male is 85-180 Hz whereas that for the average female is 165-255 Hz.

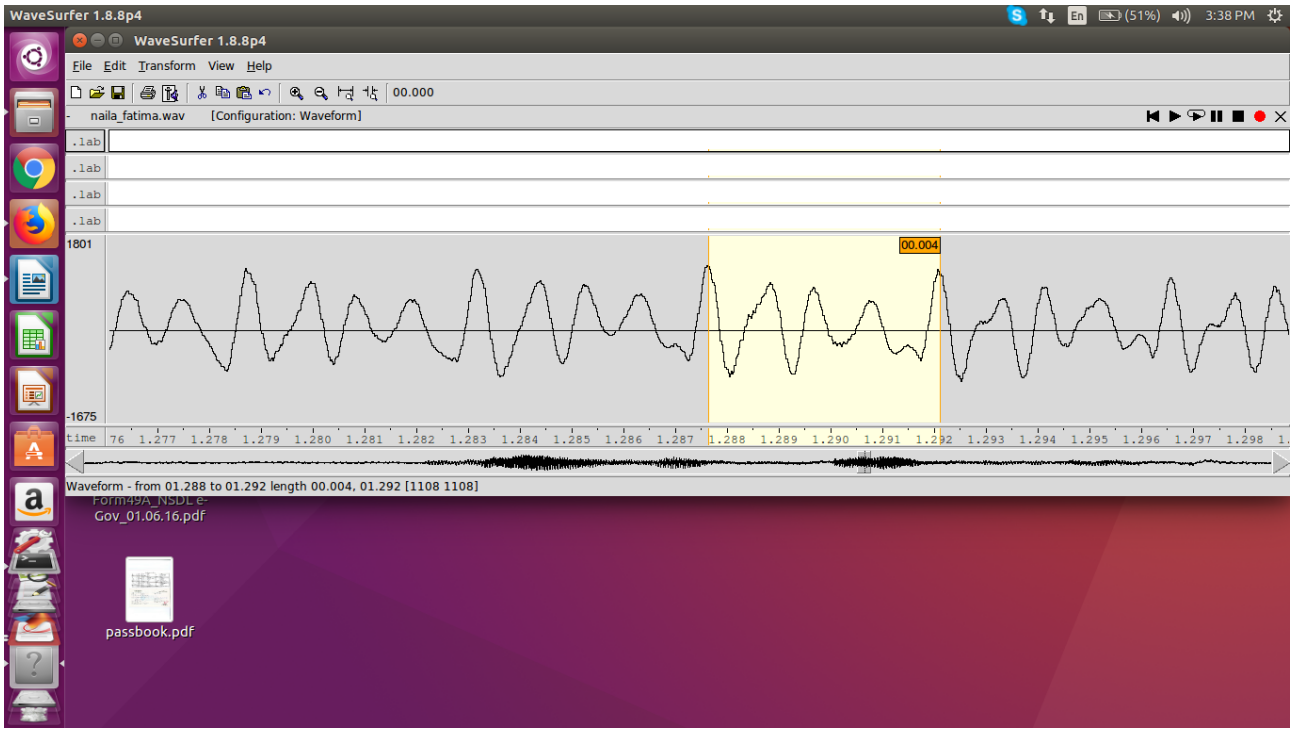
Epochs: Epoch refers to the significant excitation of the vocal tract system during the production of speech.

Formants: A formant refers to the concentration of acoustic energy around a particular frequency in the speech wave. A formant refers to a resonating frequency of the vocal tract which is indicated by a bump in the frequency response curve. The formant frequencies are indicated as F1, F2, F3 and so on and are such that the ratio F1:F2:F3 will always be 1:3:5 irrespective of the actual formant frequencies. The formants can be changed by changing the positions of the tongue and lips as this changes the frequencies at which the vocal tract vibrates at. Formants are indicated by the dark bands of a spectrogram. The darker a band, the stronger the corresponding formant.

Question 2

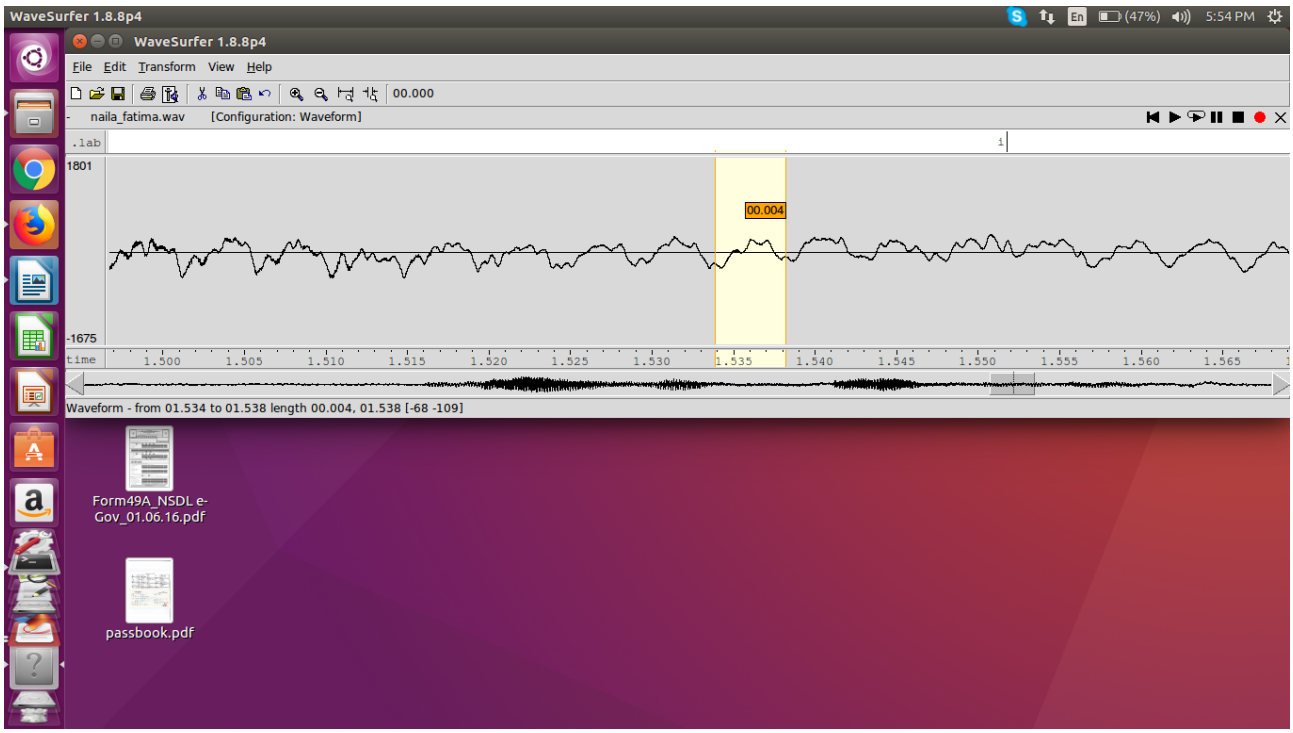
The pitch of a vowel region can be calculated as vowels are voiced phonemes which is why they show a quasi-static waveform. In order to calculate the pitch of a vowel region, one can notice the time period T of the waveform which repeats itself in that region. On observing T, we can calculate the pitch (which is the rate of vibration of the vocal folds) by calculating 1/T.

We can observe that for the first ‘a’ in ‘Fatima’, the highlighted waveform shown below, keeps repeating throughout the duration of that sound.



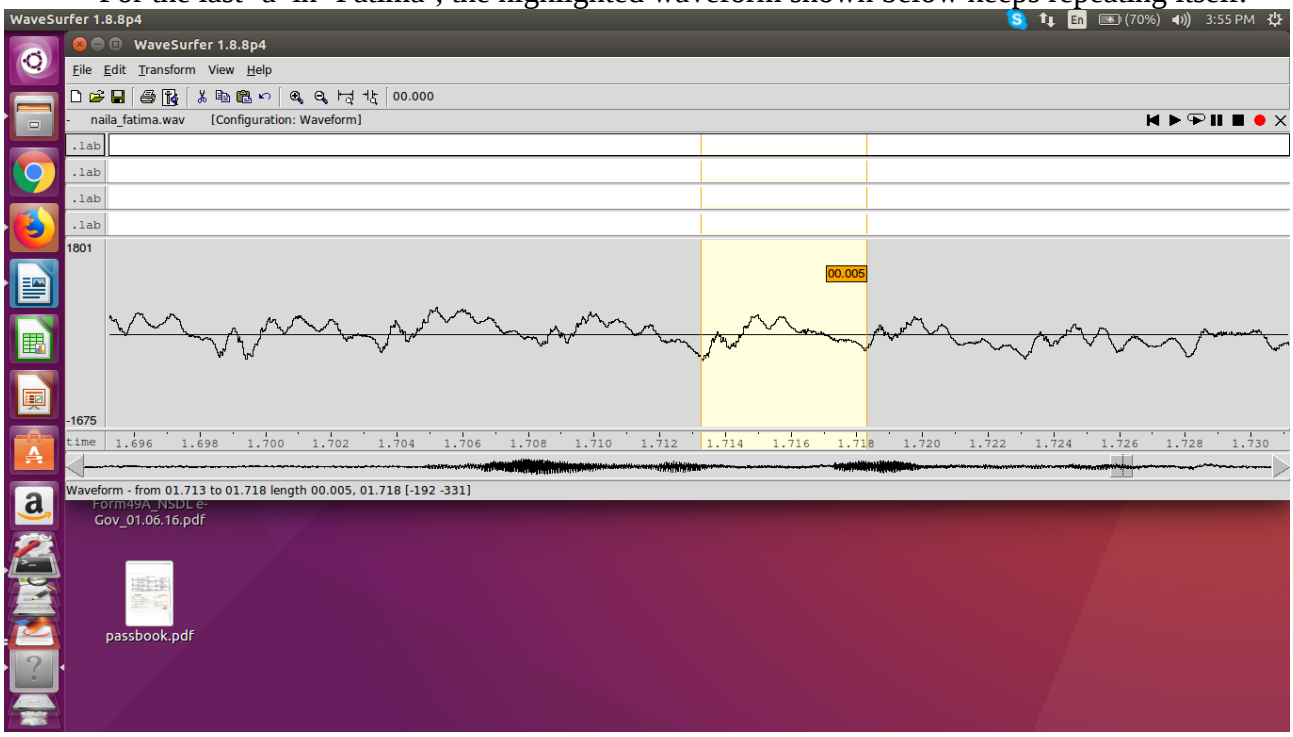
The time period for this repeating waveform is 0.004 seconds so that the pitch period for ‘a’ is 0.004 seconds and the pitch frequency is 1/T = 250 Hz.

For the ‘i’ present in ‘Fatima’, the highlighted waveform shown below keeps repeating bitself.



As we can observe, the time period T for the repeating waveform is 0.004 seconds so that the pitch period for ‘i’ is 0.004 seconds and the pitch frequency is 1/T = 250 Hz.

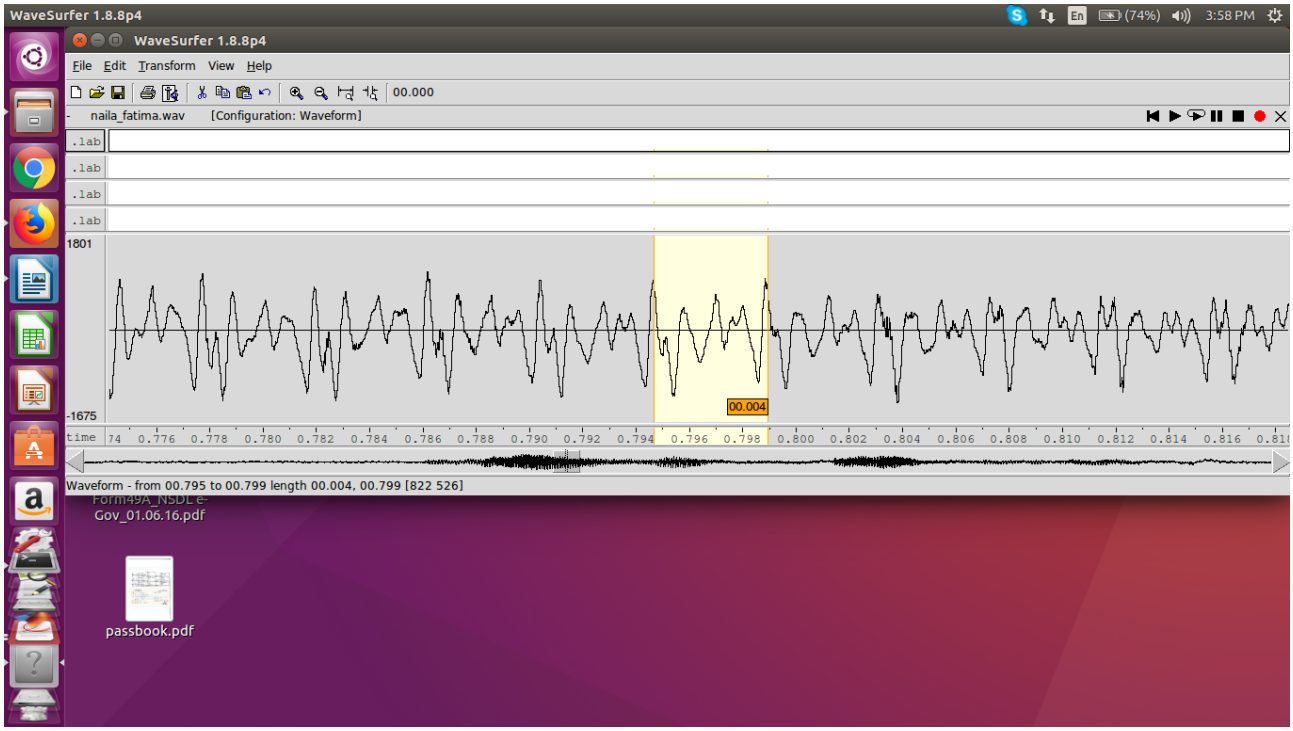
For the last ‘a’ in ‘Fatima’, the highlighted waveform shown below keeps repeating itself.



As we can see, the time period for this waveform is 0.005 seconds. The pitch period for the last ‘a’ will be 0.005 seconds while the pitch frequency for this sound will be 1/T = 200 Hz.

Bonus: ‘ai’ in ‘Naila’

As ‘ai’ is diphthong, it will be a voiced phoneme as all diphthongs are voiced. As it is voiced, there will be a repeating waveform which is as shown below.

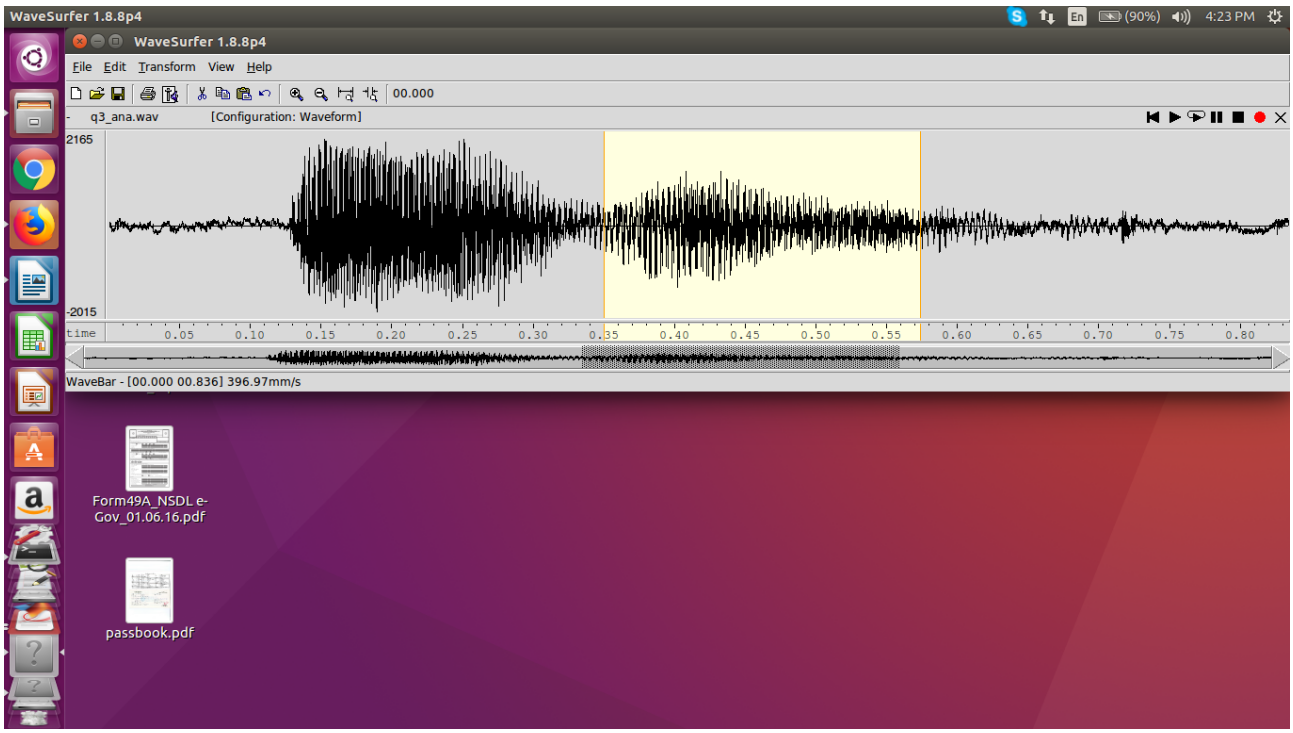


As wee can see, the time period for the repeating waveform is 0.004 seconds which is why the pitch period for ‘ai’ is 0.004 seconds while its pitch frequency is 250 Hz.

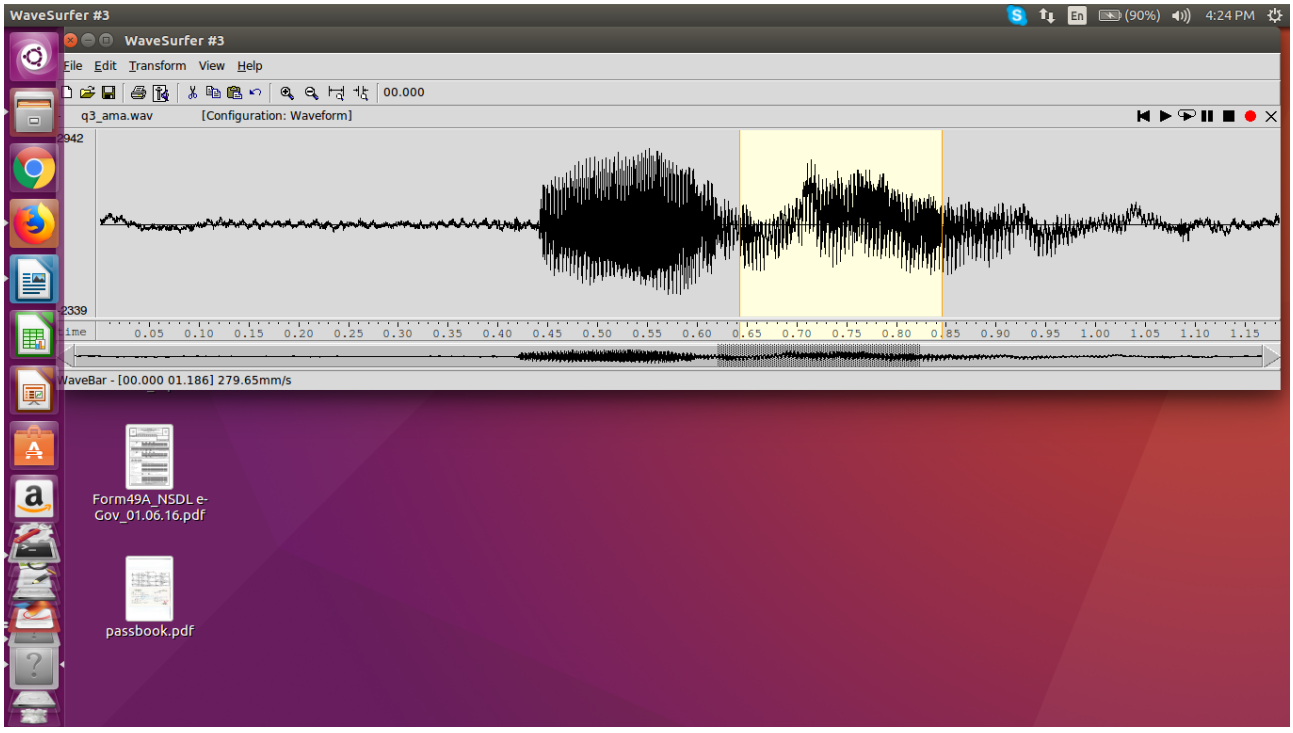
Observations: We can observe that the pitch frequencies for all the sounds were in the range of 200-250 Hz. This makes sense as the average range of female pitch lies between 210 to 220 Hz. Another observation is that the waveforms of the two ‘a’ s in the word ‘Fatima’ vary widely. This is probably due to the phenomenon of coarticulation where a speech sound becomes more similar to the speech sounds adjacent to it. We can also see that the vowel ‘i’ and the diphthong ‘ai’ have a higher pitch as compared to the vowel ‘a’.

Question 3

i)MOA is same, POA is different: For this case, I have recorded the VCV words ‘ana’ and ‘ama’ as both ‘n’ and ‘m’ are nasals (which have the same manner of articulation) but the place of articulation for ‘m’ is bilabial (requires the use of both lips) whereas that for ‘n’ is dental. The waveform for ‘ana’ is shown below.



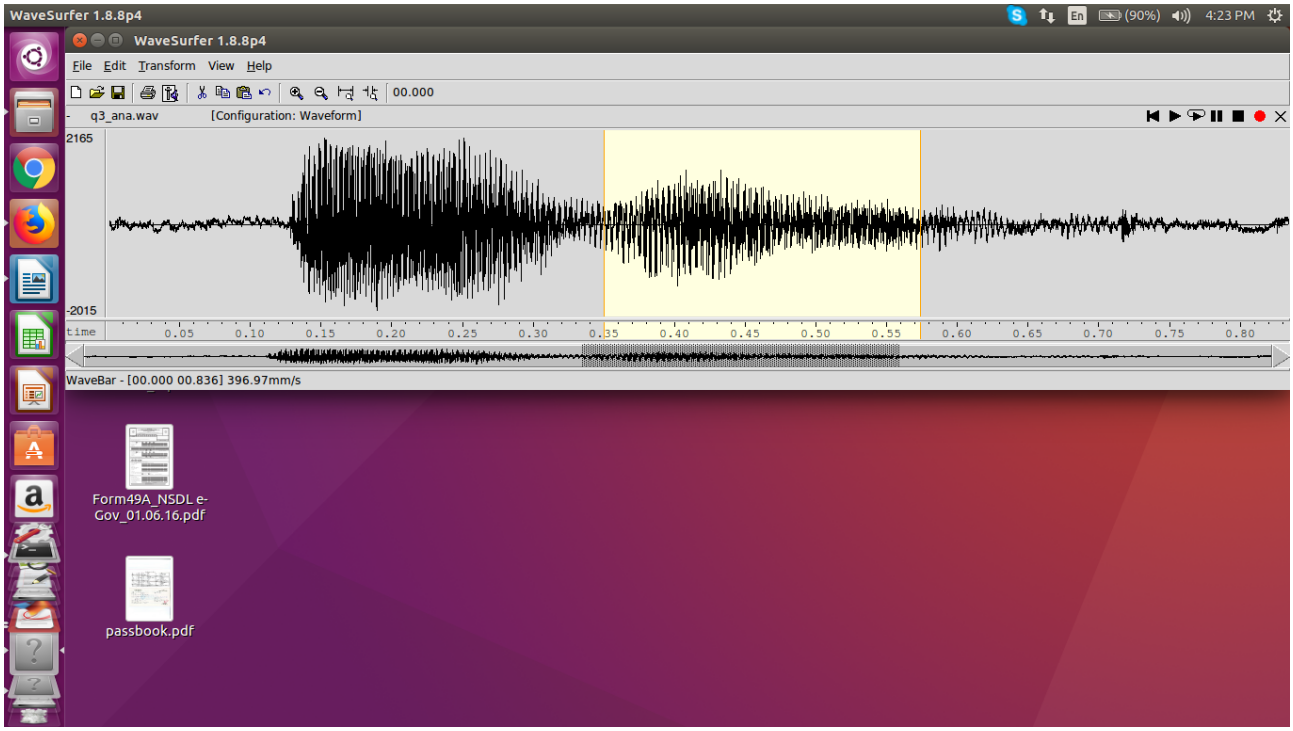
The waveform for ‘ama’ is shown below.



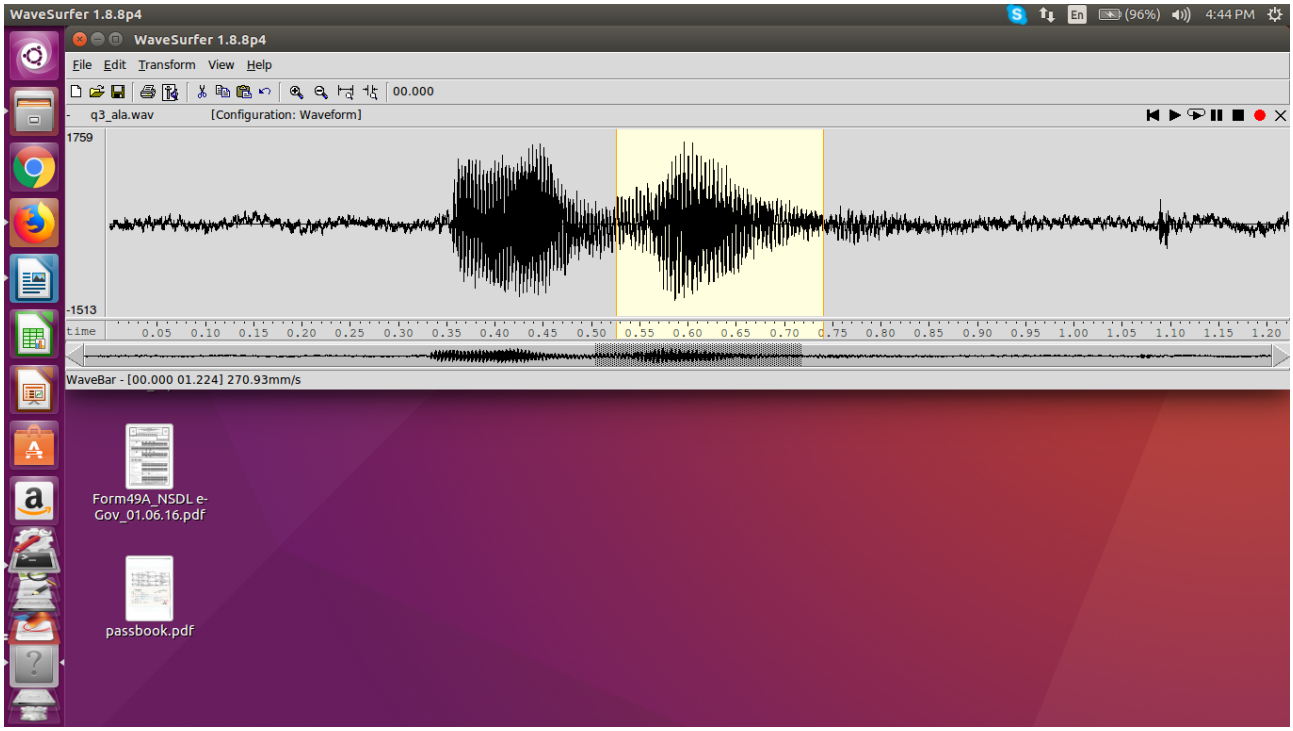
On comparing the two waveforms, we can see that the first part of the waveform which shows significant excitation appears to be similar for both the waveforms. We can ascertain by listening that this part of the waveform is the phoneme ‘a’. The highlighted portions of the two waveforms are the phonemes ‘n’ and ‘m’, respectively. On comparing these two phonemes, we can see the phoneme ‘n’ is a longer than ‘m’. We can also observe that ‘m’ shows a greater absolute magnitude as compared to ‘n’ as it is more deviated from the straight line which indicates no sound. We can also observe that the phonemes ‘m’ and ‘n’ are more clearly distinguished from the first ‘a’ of the sound as compared to the last ‘a’. We can also observe that the ‘ama’ sound shows a waveform which is more jittery as compared to the ‘ana’ waveform which is slightly more uniform in nature.

The wav file '*q3_ama.wav*' contains the 'ama' sound whereas the wav file '*q3_ana.wav*' contains the 'ana' sound.

ii) POA is same, MOA is different: For this case, I have recorded the VCV words 'ana' and 'ala'. We know that both the 'l' and 'n' phonemes are dental (which means that they require the use of the tongue against the upper teeth) so they have the same place of articulation. We should note that 'n' is a nasal while 'l' is a semivowel so this means that they have different manners of articulation. The waveform for 'ana' is shown below.



The waveform for 'ala' is shown below.



We can observe that the first part of both the waveforms which shows considerable excitation is similar. This part of the waveform can be shown to be the phoneme ‘a’ on listening. The highlighted portions of both waveforms refer to the phonemes ‘n’ and ‘l’, respectively. We can see that the ‘l’ phoneme looks to have greater magnitude/excitation when compared to the ‘n’ phoneme. We can also observe that the ‘l’ phoneme is considerably shorter than the ‘n’ phoneme. We can also notice that the phonemes ‘n’ and ‘l’ are more easily distinguished from the first ‘a’ as compared to the last ‘a’. We can also see that the ‘n’ phoneme is more uniform as compared to the ‘l’ phoneme. The ‘l’ phoneme shows greater excitation in the middle whereas the ‘n’ phoneme shows considerable excitation throughout.

The wav file for the ‘ala’ sound is ‘*q3_ala.wav*’ whereas that for the ‘ana’ sound is ‘*q3_ana.wav*’.