

ECE 2312

# Project 1

~ Mithil Amin

[GitHub Link](#)

Date of Submission: 2/16/2024

## Table of Contents

Question 1 .....	2
Question 2 .....	5
Question 3 .....	6
Question 4 .....	9
Question 5 .....	12
Question 6 .....	13
Question 7 .....	13

## Table of Figures

Figure 1 - Line 1 Plot .....	2
Figure 2 - Line 2 Plot .....	3
Figure 3 - Line 3 Plot .....	4
Figure 4 - Sound Segments .....	5
Figure 5 - Line 1 Spectrogram .....	6
Figure 6 - Line 2 Spectrogram .....	7
Figure 7 - Line 3 Spectrogram .....	8
Figure 8 - Line 1 .wav spectrogram .....	9
Figure 9 - Line 2 .wav spectrogram .....	10
Figure 10 - Line 3 .wav spectrogram .....	11

## Question 1

Record your audio and plot functions, produce three separate plots of the time domain representation of the following three phrases:

- “The quick brown fox jumps over the lazy dog.”
- “We promptly judged antique ivory buckles for the next prize.”
- “Crazy Fredrick bought many very exquisite opal jewels.”

Please make sure that the x-axis and y-axis are properly labeled, and the x-axis is scaled properly in term of seconds.

Plot 1 is: “The quick brown fox jumps over the lazy dog.”

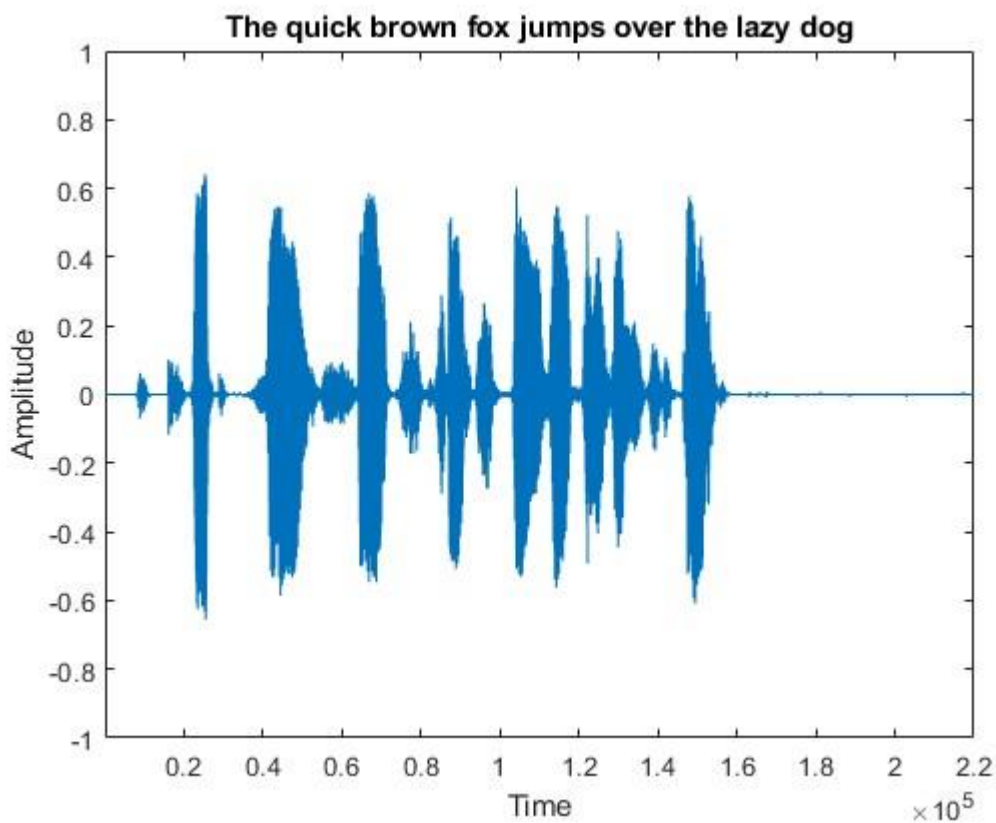


Figure 1 - Line 1 Plot

Plot 2 is: “We promptly judged antique ivory buckles for the next prize.”

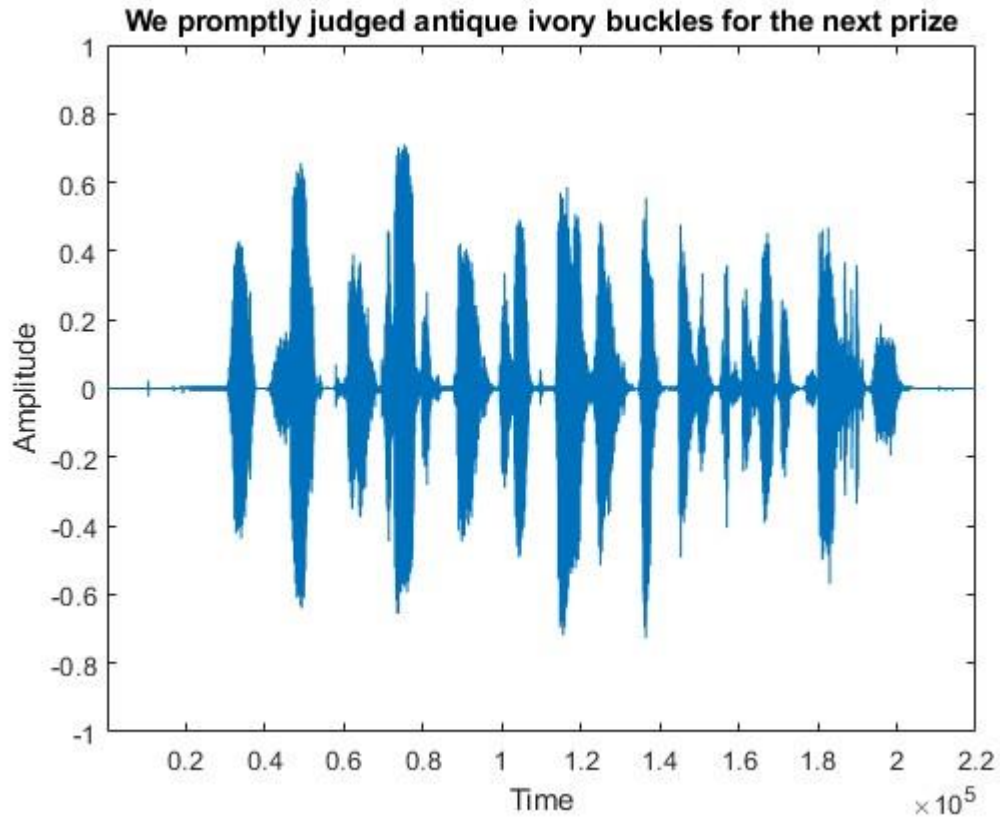


Figure 2 - Line 2 Plot

Plot 3 is: “Crazy Fredrick bought many very exquisite opal jewels.”

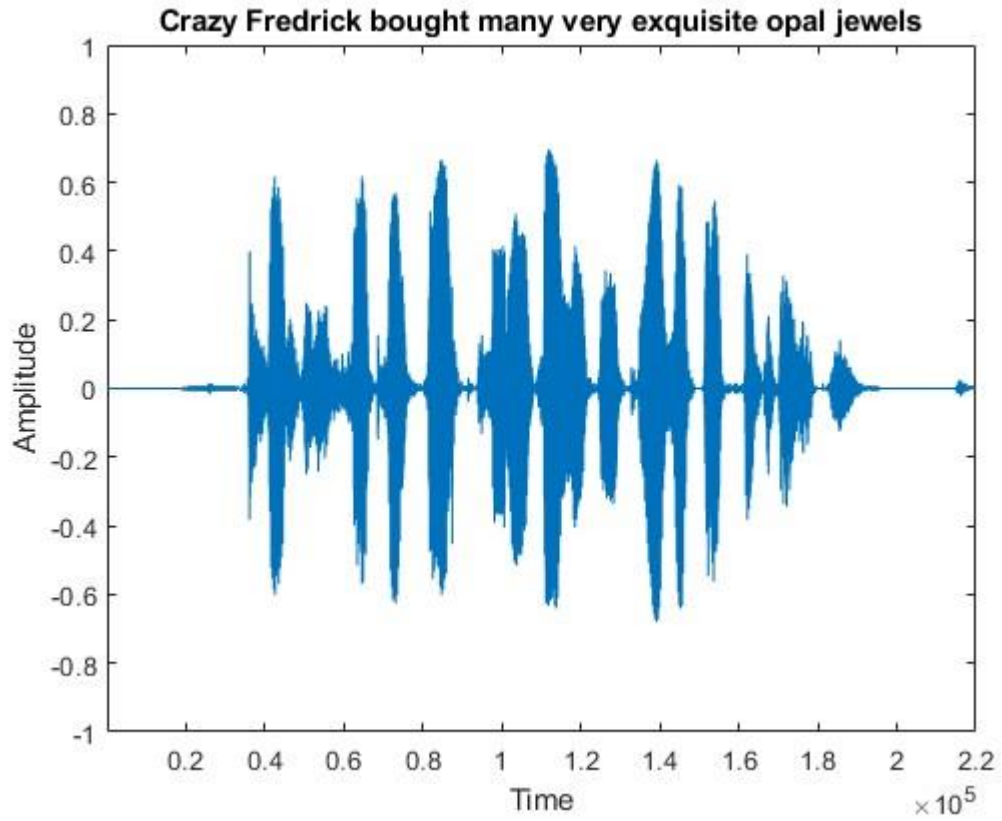


Figure 3 - Line 3 Plot

## Question 2

Based on the spectrogram shown in Figure 2, approximately determine which time segments belong to the sounds “do”, “re”, “mi”, “fa”, “so”, “la”, “ti”, and “do”. One way of accomplishing this task is by looking at the frequency behavior at each time instant of the speech signal. Vowel phenomes will have well-pronounced energy bands in frequency called formants, while sounds like “efff” and “shhh” (i.e., fricatives) will have almost all their energy located above 4000 Hz.

For this question, I have taken Figure 2 given in the project document, and annotated the different segments for “do”, “re”, “mi”, “fa”, “so”, “la”, “ti”, and “do”.

Figure 4 shown below is my annotation.

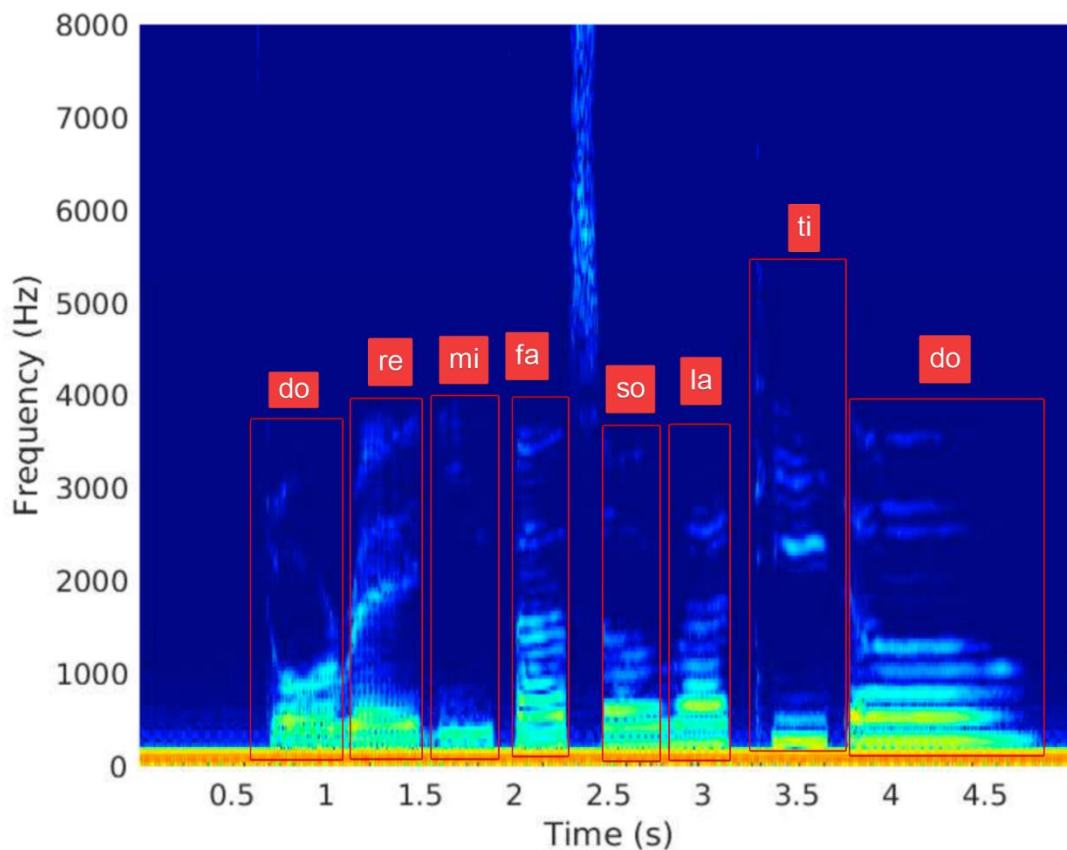


Figure 4 - Sound Segments

### Question 3

Using the recorded speech signals from the previous section, namely:

- “The quick brown fox jumps over the lazy dog.”
- “We promptly judged antique ivory buckles for the next prize.”
- “Crazy Fredrick bought many very exquisite opal jewels.”

Generate their corresponding spectrograms throughout their entire time duration and only across their first 8000 Hz of frequency. Furthermore, annotate these spectrograms with respect to highlighting any sounds, phonemes, and/or words that can be readily identifiable based on their spectral characteristics.

Sounds for words like “the” aren’t very definitive as while speaking I noticed that certain words are spoken quite quickly, and often without fully pronouncing them out.

Spectrogram 1 is: “The quick brown fox jumps over the lazy dog.”

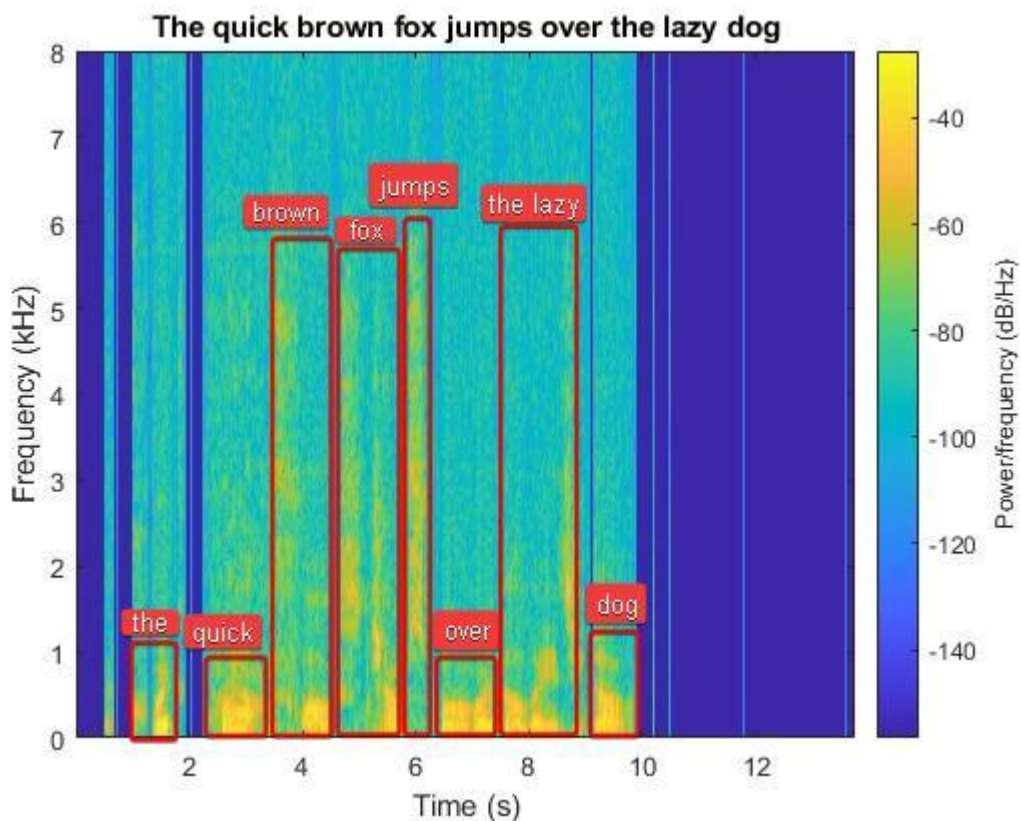


Figure 5 - Line 1 Spectrogram

Spectrogram 2 is: “We promptly judged antique ivory buckles for the next prize.”

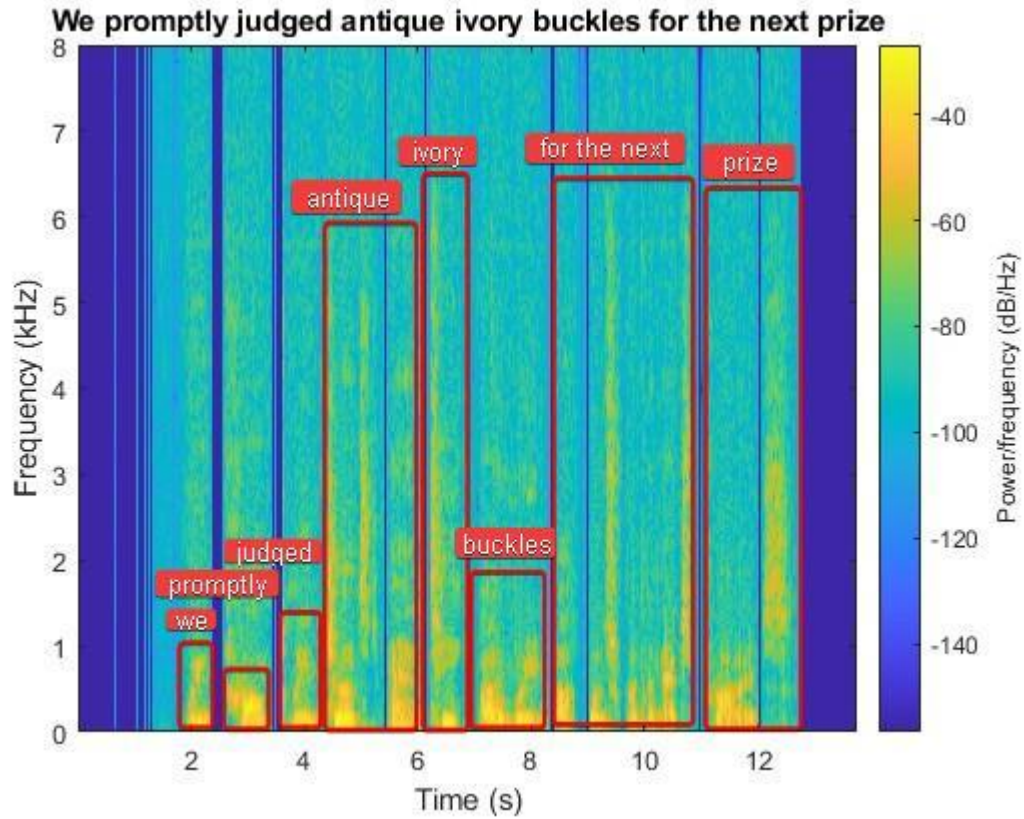


Figure 6 - Line 2 Spectrogram



Spectrogram 3 is: “Crazy Fredrick bought many very exquisite opal jewels.”

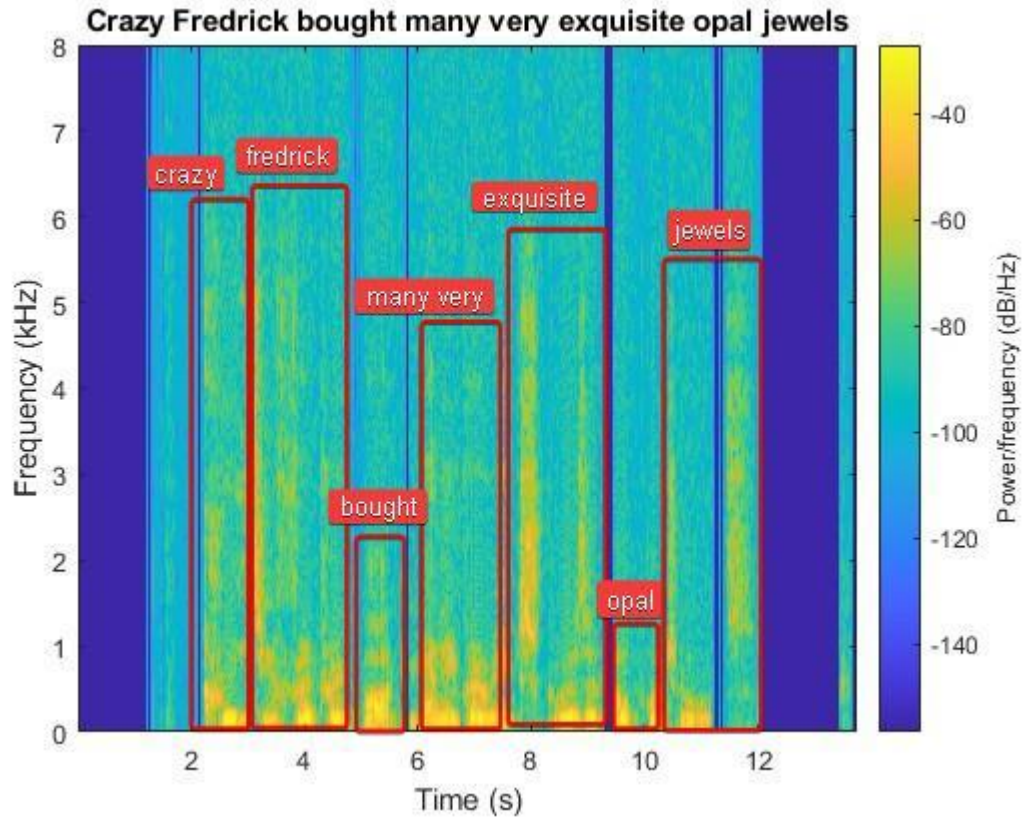


Figure 7 - Line 3 Spectrogram

## Question 4

Using audiowrite, store the recorded speech signals from the previous section to three separate WAV files, namely:

- “The quick brown fox jumps over the lazy dog.”
- “We promptly judged antique ivory buckles for the next prize.”
- “Crazy Fredrick bought many very exquisite opal jewels.”

Generate the corresponding spectrograms of these speech signals. Compare these side-by-side with your three speech signals and highlight any similarities between the two sets of speech signals.

Using audiowrite, we were able to store the three recorded speech signals in audio files with .wav extension with the following file names:

- “The quick brown fox jumps over the lazy dog.wav”
- “We promptly judged antique ivory buckles for the next prize.wav”
- “Crazy Fredrick bought many very exquisite opal jewels.wav”

The Spectrogram generated by the data stored in these .wav files is as follows.

Spectrogram for Line 1

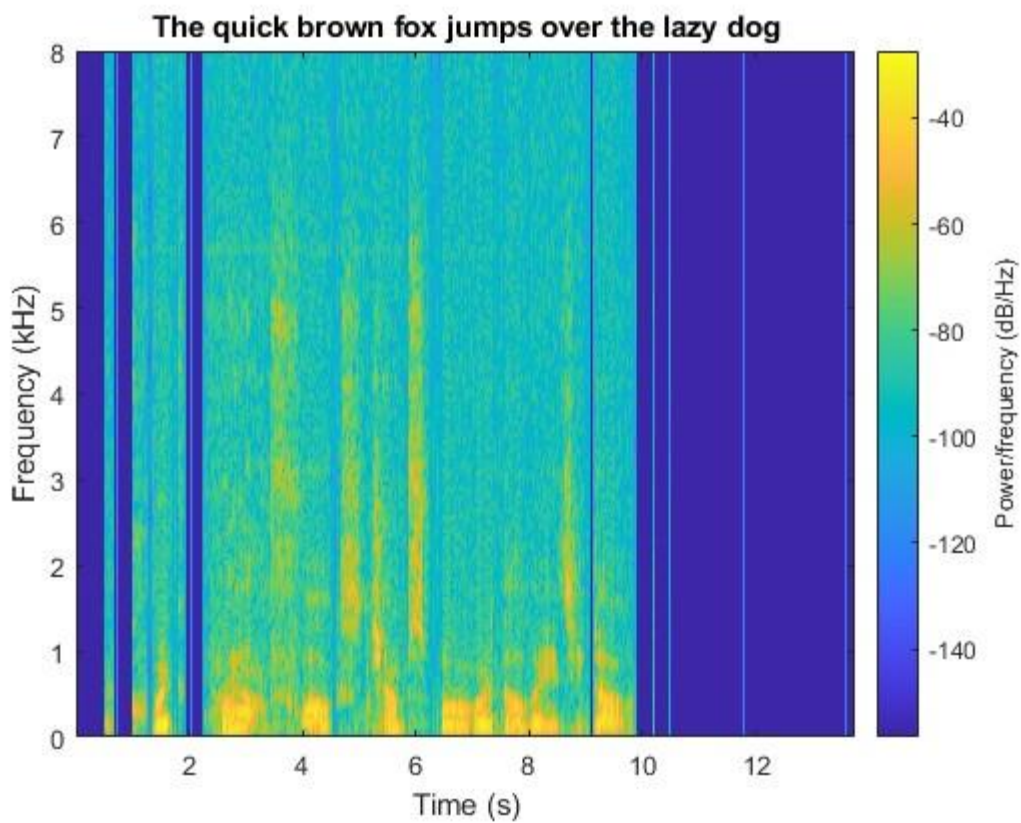
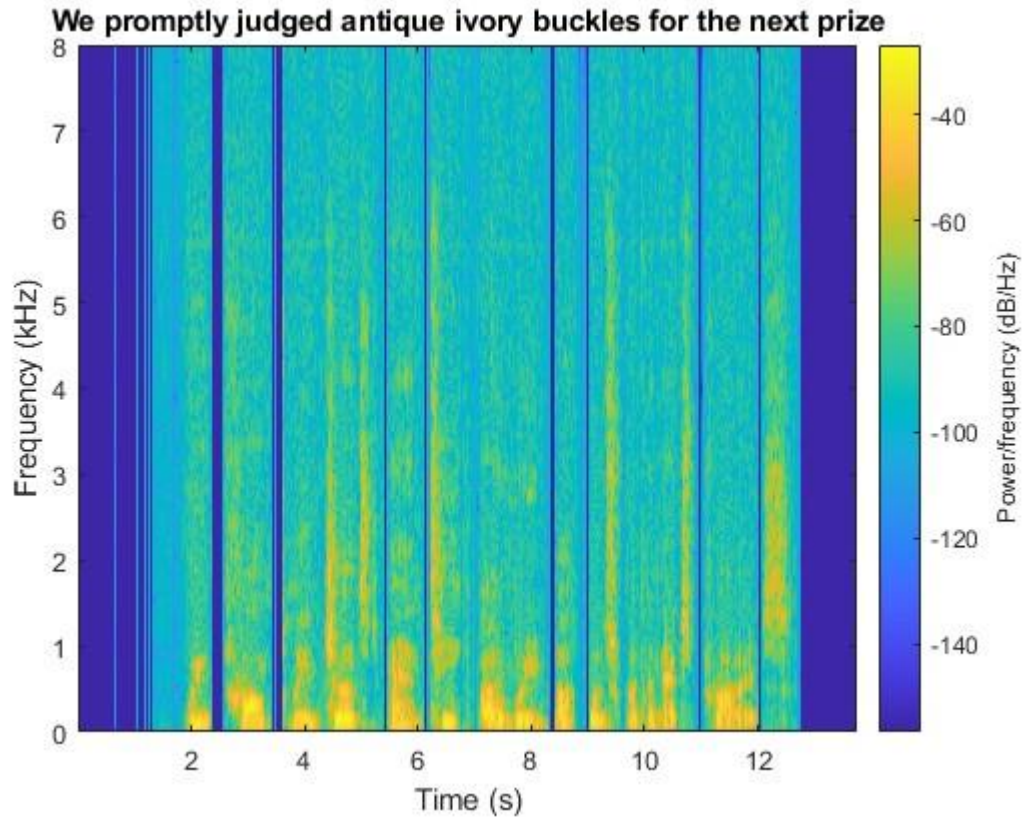


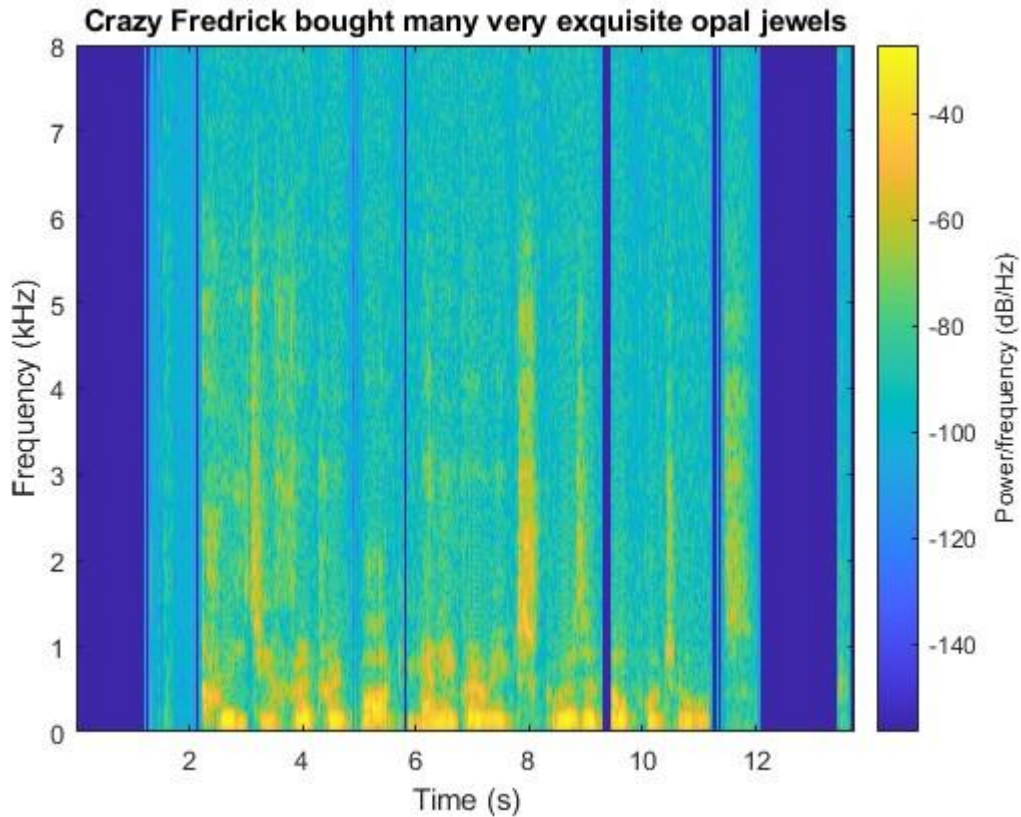
Figure 8 - Line 1 .wav spectrogram

## Spectrogram for Line 2



*Figure 9 - Line 2 .wav spectrogram*

### Spectrogram for Line 3



*Figure 10 - Line 3 .wav spectrogram*

When these spectrograms were compared side by side with their recorded counterparts, there was almost no visible difference. The main reason for the absence of any difference is due to the file extension used to store the audio data.

Audio data stored with .wav extension have lossless audio storage, thus with the parameters with which these spectrograms were generated, there is no visible difference.

## Question 5

Using any measurement method, measure the absolute distance (in meters) from one ear canal to the other for each team member. Use the speed of sound to calculate the time delay (in milliseconds) between the ears. Finally, calculate the number of samples that would occur in the time delay (hint: answer will depend on the sample rate of your audio device). Calculate the team average number of samples.

For this question, I used two spare pieces of wood and a measuring tape. By placing the tape flat on the table, I held the two wooden pieces to each ear. Then, I carefully placed my forehead on the table such that the wooden pieces were on the tape. Now, I lifted my head and took down the readings. After doing this a couple times, I came to an average of 0.16 meters of distance between my two ear canals.

This method has its limitations as if I don't hold the wooden pieces parallel to each other, the readings will be affected quite a bit.

With the speed of sound( $c$ ) as 343 meters per second and distance( $s$ ) as 0.16 meters. I was able to get the delay time period as 0.0004664723 seconds, and when multiplied by 1000, we get 0.4664723 milliseconds.

Now, to calculate the number of samples that would occur in this time delay, we first need to understand the Sampling rate. Sampling rate is the number of samples taken per second. With this information, I was able to deduce that multiplying the sampling rate with the time delay, we will have number of samples occurring. Now, since we can't take half a sample, it was important to use `ceil()` function when doing this calculation, this way we always made sure that the samples we were delaying by was accurate and complete samples.

Since I am attempting this project myself the average was same as the number of samples we found previously.

## Question 6

Using one of your previously recorded speech signals, copy the column to the right to create two identical columns, and save this signal to a WAV file. Label this file with the convention “team[[yourteam#]]-stereosoundfile-[[delay]].wav”. Then, delay the second column by the average number of samples calculated previously, and save this signal. Finally, delay the second column by 1ms, 10ms, and 100ms, and save these. Have each team member listen to the files and compare the acoustic experience of each of these files (0ms, avghead, 1ms, 10ms, 100ms).

In this question I was able to listen to each of the audio files and I was able to infer that higher delay between two channels, the clearer was the direction of the sound.

## Question 7

Attenuate the right channel of the oms delay audio by -1.5dB, -3dB, and -6dB, and save these signals to WAV files. Label these files with the convention “team[[yourteam#]]-stereosoundfile-[[delay]]-[[attenuation]].wav”. Perform the same operation to the average head delay file. Have each team member listen to the files and compare the acoustic experience of each of these files (0ms, oms-1\_5dB, oms-3dB, oms-6dB, avghead, avghead-1\_5dB, ...).

Which delay and attenuation values produced the most convincing? illusion that the sound source is to the left of the listener. How might you improve the basic model to improve the illusion or support sounds from other locations in 3D space?

For this question, I use the following gain multiplication for -1.5dB, -3dB, and -6dB respectively:

- 0.75
- 0.5
- 0.25

After attenuating this audio recording, I was able to listen to these files and the delay along with the attenuation increased the clarity of the direction of sound. Using both of these together definitely improves the experience.

To improve the illusion of sound location, we can further increase the delay and attenuation while recording the audio with higher sampling rate and bitrate for increased quality. We can also attempt to add to the audio illusion by actively attenuating sounds as they are being played to resemble the hearing of voice from a moving object or person.