MT Assignment#3 Report

Audio Analysis

資工三 406410035 秦紫頤

Setup

Components

There are:

- **2** folders: result, src
- **3** files: hw3.m, hw3_live.mlx, MT-hw3_report.pdf

in this assignment folder.

- **result:** This folder contains all the results generated from the program. There are **5** folders in it
  - **frequency_domain:** There are **4** images inside this folder, which are 4 audio files plotted in the frequency domain
  - **spectral_view:** There are **4** images inside this folder, which are 4 audio files plotted in the spectral view.
  - **time_domain:** There are **4** images inside this folder, which are 4 audio files plotted in the time domain
  - **compare_with_girl:** There are **8** images in this folder, which are 4 frequency component comparison bar graphs, and 4 spectral view comparison graphs. These are the comparison results of my 4 audio files and the other **female** classmate's audio files
  - **comparison_with_male:** There are **8** images in this folder, which are 4 frequency component comparison bar graphs, and 4 spectral view comparison graphs. There are the comparison results of my 4 audio files and the other **male** classmate's audio files.
- **src:** There are **3** folders inside this folder.
  - **myaudio:** There are **4** audio files in this folder which are all my own recordings. **"audio_1.wav"** is me talking, **"audio_2.wav"** is me coughing, **"audio_3.wav"** is me singing, and **"audio_4.wav"** is the combination of the previous 3 audio files.
  - **girl:** There are **4** audio files in this folder and all from another female classmate. **"speaking.wav**", **"noise.wav"**, **"singing.wav"**, and **"combination.wav"** can see the purpose of these audio files by the filename.
  - **rong_audio:** There are **4** audio files in this folder which are all from another male classmate. **"Speak.wav"**, **"Noise.wav"**, **"Sing.wav"**, and **"Mix.wav"** which can tell the purpose from the audio files filename.

- **hw3.m, hw3_live.mlx:** The program for this assignment

- **MT-hw3_report.pdf:** The report for this assignment

Prerequisite

1. MATLAB: recommended version R2019b

Execution

1. Move "**src**" folder **from part2 folder to part1** folder
1. Open MATLAB
2. Select **"406410035_hw3_v1_part1"** this folder to open in your MATLAB
3. Open hw3.m or hw3.mlx
4. Run the program
5. Go to **"406410035_hw3_v1/result/"** to find all the results generated from the program

Method Description

I split this assignment into 2 parts. The detail is as follow:

Part1: Analyze my own voice

1. Record my own voice: talking, coughing, singing
    a. I use QuickTime for audio recording. And the audio format is **.m4a**
2. Audio Conversion to .wav
    a. Use online-convert.com to convert from .m4a to .wav
    b. Concatenate the three audio files as the forth audio file using **QuickTime**
3. Open MATLAB and read wav files with **audioread()**
4. Time-domain plot
    a. Use only the first dimension of the audio data (total 2 dimensions)
    b. Get the derivative of time -> $\frac{1}{sample\_rate}$
    c. Get all the timestamps of the audio
    d. Draw the plot using **plot()**
    e. Save the plot as an image using **saveas()**
5. Spectral View
    a. Use only the first dimension of the audio data
    b. Set the window to **128**
    c. Set number of overlapped samples to half of the window size
    d. Set number of DFT points same as the window
    e. Show the spectral view using **spectrogram()**
    f. Save the spectral view as an image using **saveas()**
6. Frequency domain plot
    a. Get the original audio length by **length()**
    b. Get all the frequency point as the x-axis

c. Change the whole signal to the frequency domain by doing fast Fourier transform by **fft()** and rearranges the transform (normalize) by **fftshift()**
d. Draw the plot by **plot()**
e. Save the plot as an image by **saveas()**

Part2: Compare my voice with other people's voice

1. Compare in the frequency domain
   a. Find the frequency domain of other people's audio files -> same as **Part1 Step6 1~3**
   b. Find the frequency range of me audio and the other people's audio
   c. Get the intersection of the two frequency range and get sample points within the intersection
   d. Draw 3d plot as the comparison of my audio and the other people's audio using **bar3()**
   e. Make the figure larger so the plot looks clearer ->**set(gcf, 'Units', 'Normalized', 'OuterPosition', [0, 0.04, 1, 0.96]);**
   f. Set the labels and title of the plot -> **xticklabels(), xlabel(), yticklabels(), ytickangle(), zlabel()**
   g. Save the plot as an image by **saveas()**
2. Compare in spectral view
   a. Create subplot using **subplot()**
   b. On the first subplot, I drew the spectral view of my audio
   c. On the second subplot, I drew the spectral view of other people's audio as for comparison
   d. Save as image using **saveas()**

Results

a. Time-domain results

The results of the time domain are in the folder **"./result/time_domain"**. There are 4 images corresponding to the 4 audio files. The results are shown in Figure (a), Figure (b), Figure (c) and Figure (d)

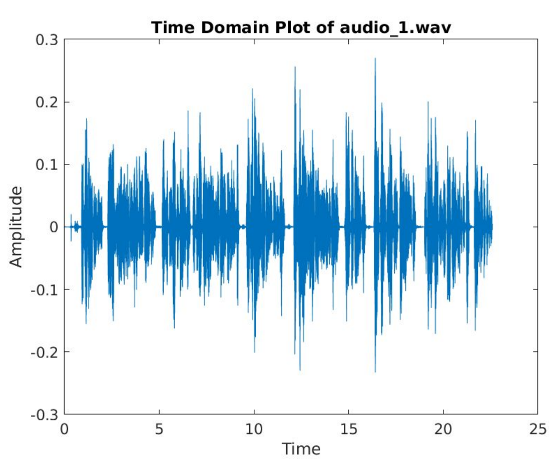


Figure (a) /time_domain/audio_1.png

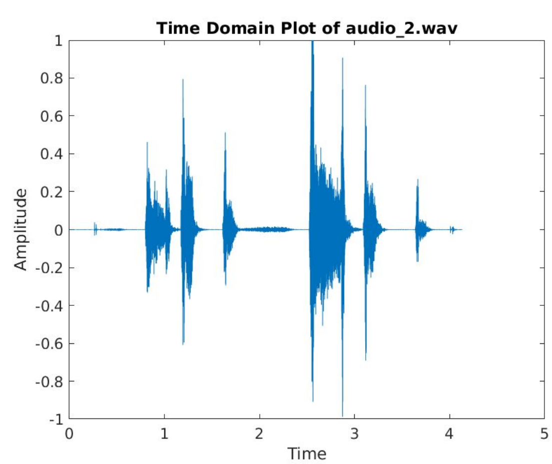


Figure (b) /time_domain/audio_2.png

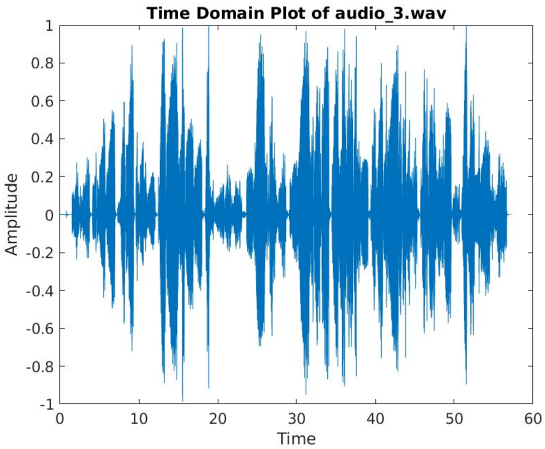


Figure (c) /time_domain/audio_3.png

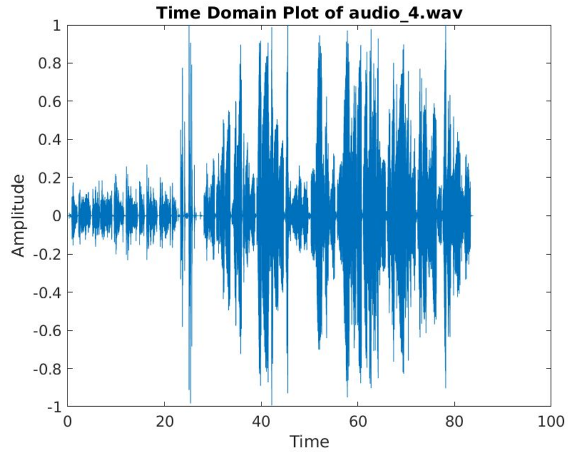


Figure (d) /time_domain/audio_4.png

b. Spectral View results

The results of the spectral view are in the folder **"./result/spectral_view"**.There are 4 images corresponding to the 4 audio files. The results are shown in Figure (e), Figure (f), Figure (g) and Figure (h)

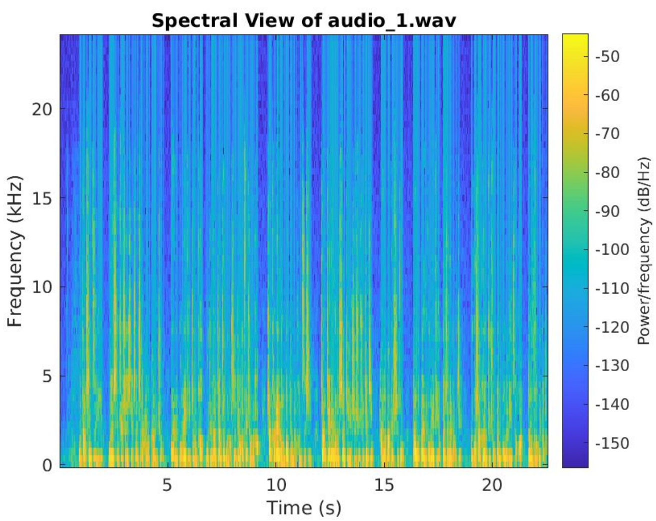


Figure (e) /spectral_view/audio_1.png

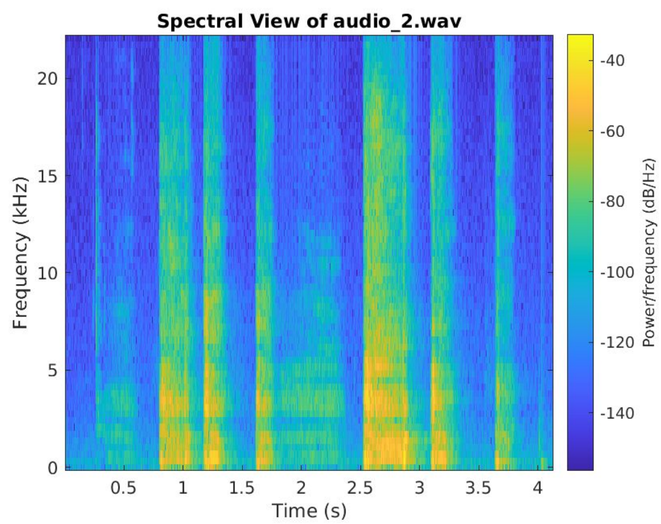


Figure (f) /spectral_view/audio_2.png

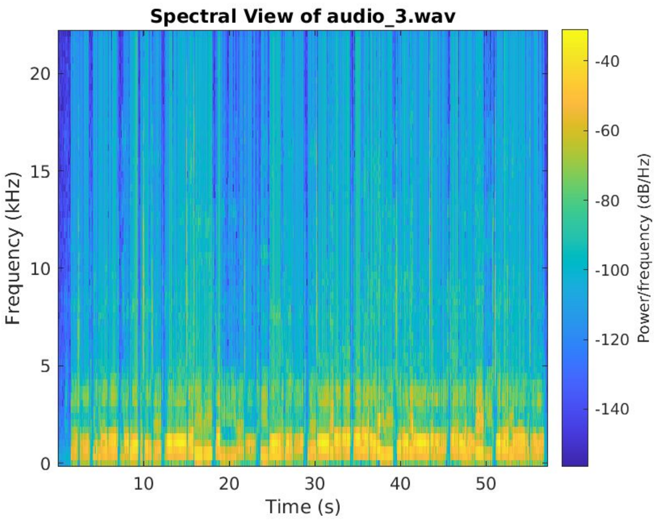


Figure (g) /spectral_view/audio_3.png

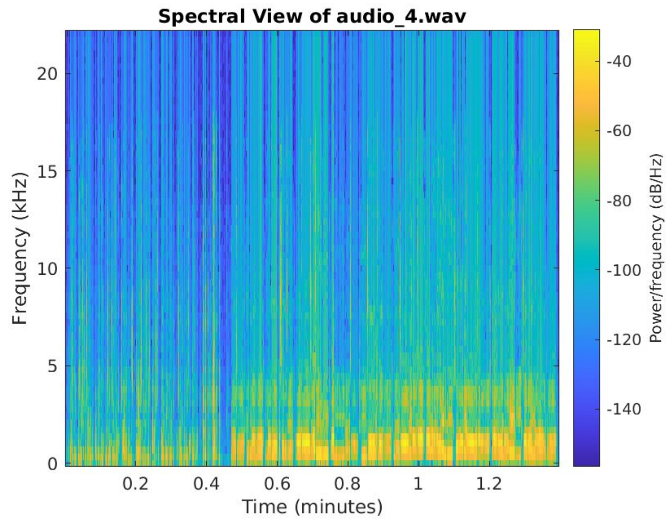


Figure (h) /spectral_view/audio_4.png

c. Frequency domain results

The results of the frequency domain are in the folder **"./result/frequency_domain"**. There are 4 images in this folder corresponding to 4 audio files. The results are shown in Figure (i), Figure (j), Figure (k), Figure (l) below:

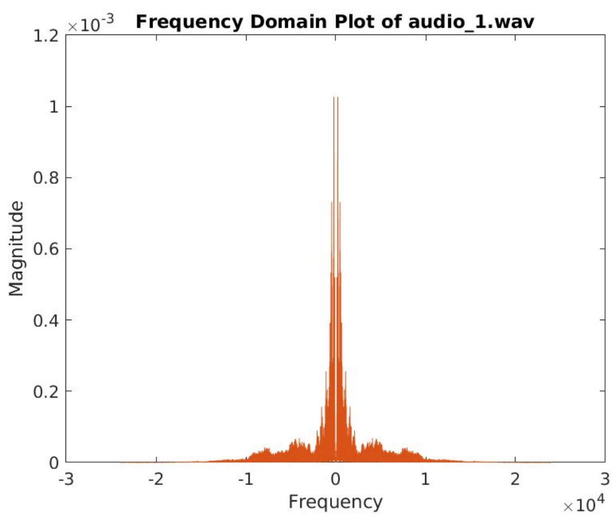


Figure (i) /frequency_domain/audio_1.png

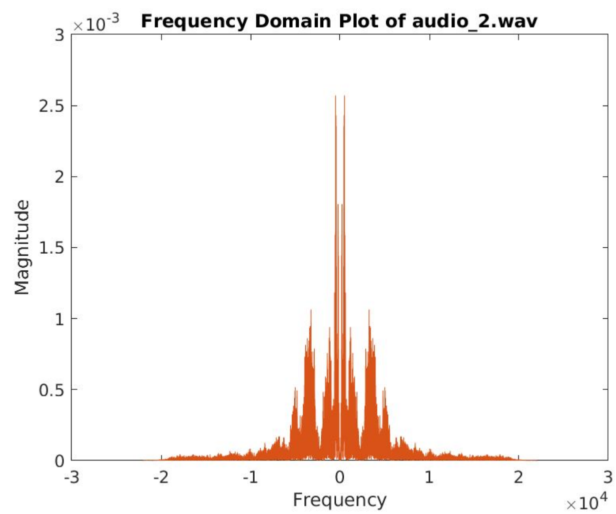


Figure (j) /frequency_domain/audio_2.png

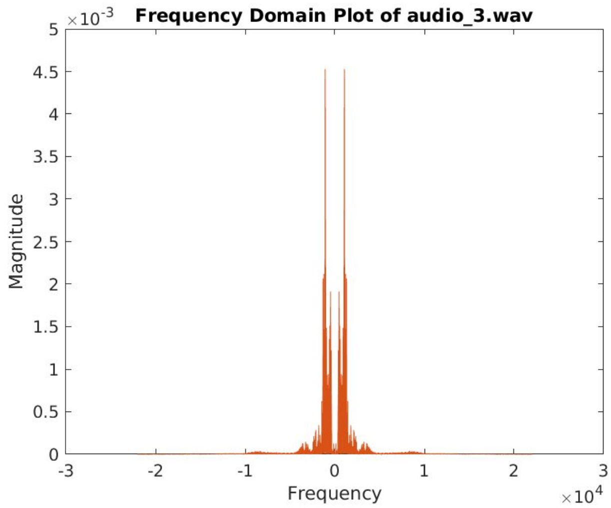


Figure (k) /frequency_domain/audio_3.png

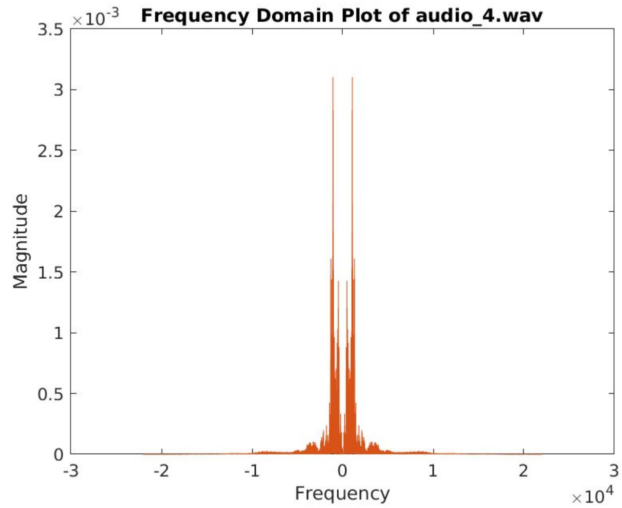


Figure (l) /frequency_domain/audio_4.png

d. Discuss the results of speaking, coughing and singing in different domains

1. **Talking --- audio_1.wav**

   First of all, in the time-domain, the curve is quite regular. This is because usually, voice tone in the regular scenario wouldn't change much so the curve looks regular. Second, in the spectral view, I can see the power of my speaking is quite normal, not so high, cause I didn't shout while speaking. Last, in the frequency domain, I can see the magnitude curve in the frequency domain is a normal distribution with mean when frequency equals 0. I think this is because purely talking won't have stunning tone affect the curve.

2. **Coughing --- audio_2.wav**

   First, in the time domain, I can see some huge amplitude change at some intervals. That is because when coughing I cough hard at these intervals. Second, in the spectral view, I can see the power in the intervals I coughed hard is quite strong. Last, in the frequency domain, I can see although the mean is still in frequency 0. But the curve isn't as focus in 0 as the audio_1. I think this is because when coughing there got a wider range in the frequency and the magnitude is much higher too.

3. **Singing --- audio_3.wav**

   First, in the time-domain, the curve is totally irregular and keep changing at every moment in different amplitude. This is because when singing my voice isn't in the same tone, it changed tremendously throughout the song. Second, in the spectral view, the power of my voice is always strong for the whole 57 seconds. This is because I'm signing Lascia Ch'io Pianga which is a pretty hard song which can't present properly if I don't sing it loud and strong. Last, in the frequency domain, my audio signal doesn't focus on frequency 0, this can tell from the magnitude when frequency equals 0 is nearly 0. Instead, there are higher magnitudes when the frequency is a bit away from 0 which is what I'm really surprised of. I don't know the reason for this.

e. The changes in different domains when the 3 audio files combine

In the time domain, the curve is simply the concatenation of the three audio files time-domain curves. This is same as in the spectral view curve. This is because the x-axis is based on time. But in the frequency domain, it is quite obvious that the curve isn't as simple as concatenating the three curves cause the x-axis isn't based on time. But from the curve, I can tell that audio_3 affect this curve the most cause this curve looks similar to audio_3's curve than the other curves. I think this is because singing takes 57 seconds but talking and coughing takes less time. This is just one of my guesses.

f. Frequency Component Comparison Results

1. The following is the comparison results of frequency component results between my own audio and the other female classmate's. They are all saved in **"./result/compare_with_girl"**. They will be shown in Figure (m), Figure (n), Figure (o), Figure (p) below:

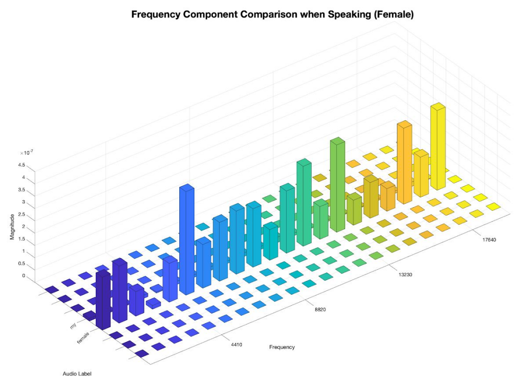


Figure (m) compare_with_girl/speaking.png

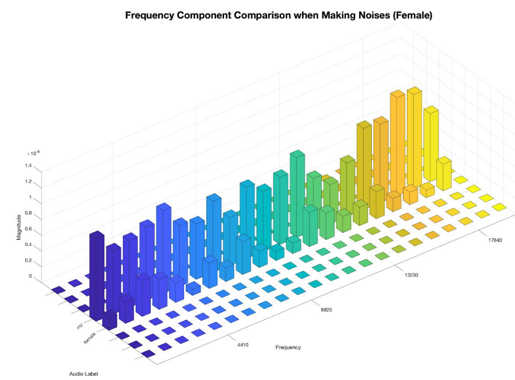


Figure (n) compare_with)girl/noise.png

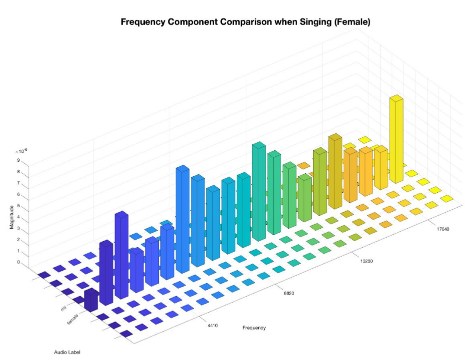


Figure (o) compare_with_girl/singing.png

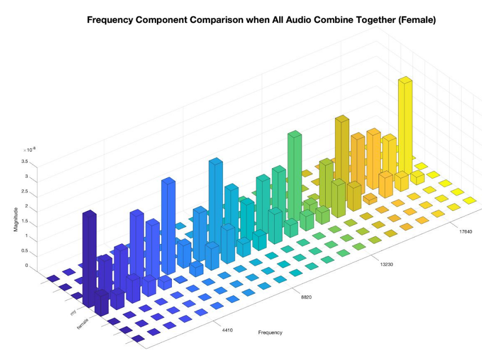
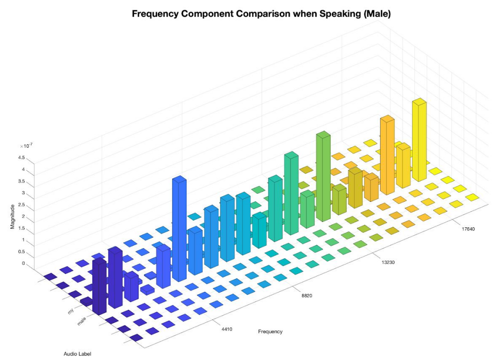


Figure (p) compare_with_girl/combination.png

2. The following is the comparison results of frequency component results between my own audio and the other male classmate's. They are saved in **"./result/compare_with_male"**. They will be shown in Figure (q), Figure (r), Figure (s), Figure (t) below:



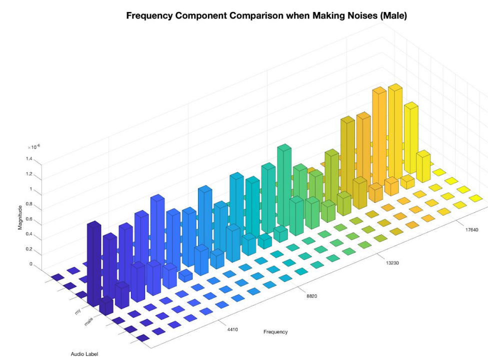
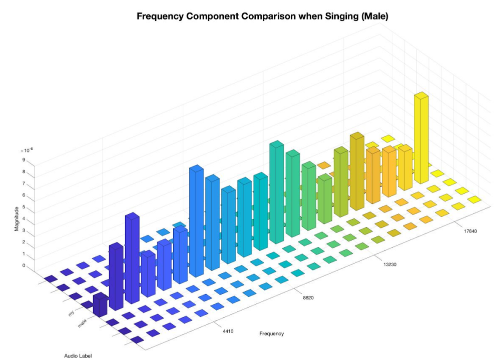


Figure (q) compare_with_male/speaking.png

Figure (r) compare_with_make/noise.png



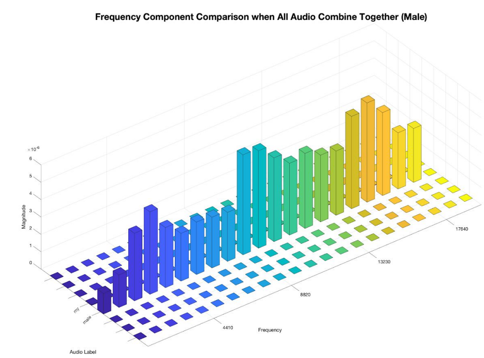
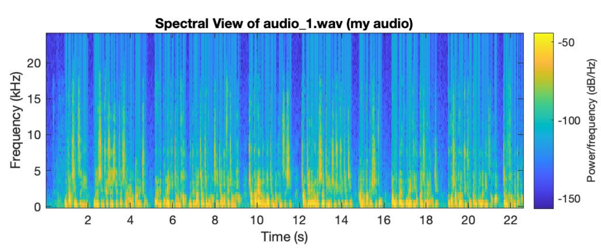


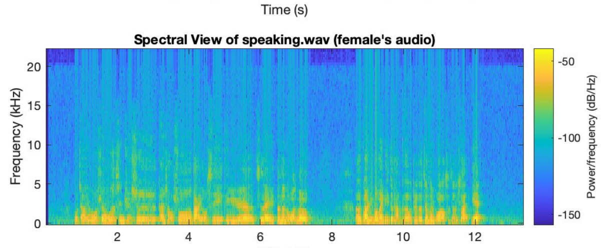
Figure (s) compare_with_male/singing.png

Figure (t) compare_with_male/combination.png

g. <u>Spectral view comparison results</u>

1. Comparison between my audio files and another female classmate's audio files. All the results can be found in **"./result/compare_with_girl"**.
   a. Speaking

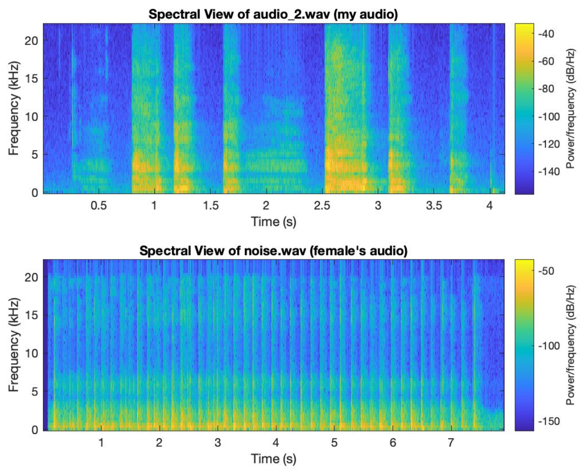




From the spectral view in Figure (u) I can see that when I am speaking my voice's frequency range is larger than the other person. But the other person's voice is more focus in 0-3kHz. As for mine, it isn't focus on specific frequency so it may sound more vivid is hearing the original audio.

Figure (u) compare_with_girl/spectral_speaking.png

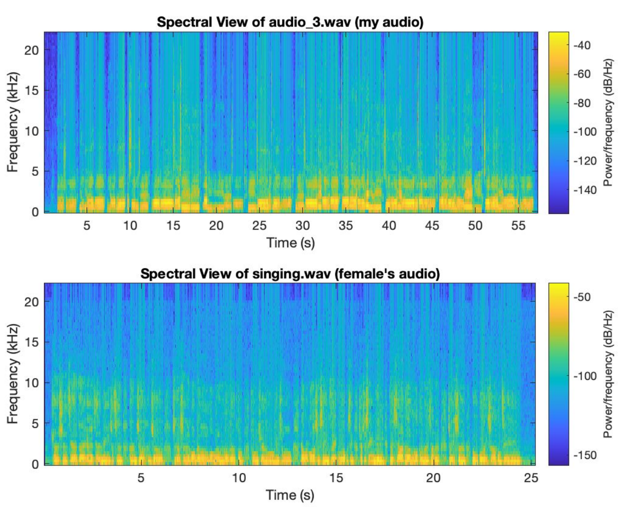
b. Noise



From the spectral view in Figure (v) I can see that the power I use to making noize seems stronger than another person. Especially in some interval the yellow sections looks quite obvious. Another person's noise seems regular and repetitive compared to mine.

Figure (v) compare_with_girl/spectral_noise.png

c. Singing



From Figure (w) I can see that when singing both me and another person's frequency range is larger. Cause we are both female and our voice can reach a wide range. But also I found out that my voice's range is slightly larger than the other person.

Figure (w) compare_with_girl/spectral_singing.png

d. Combination

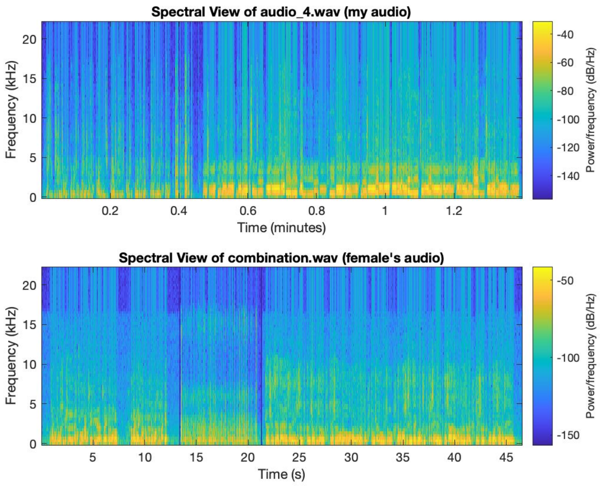
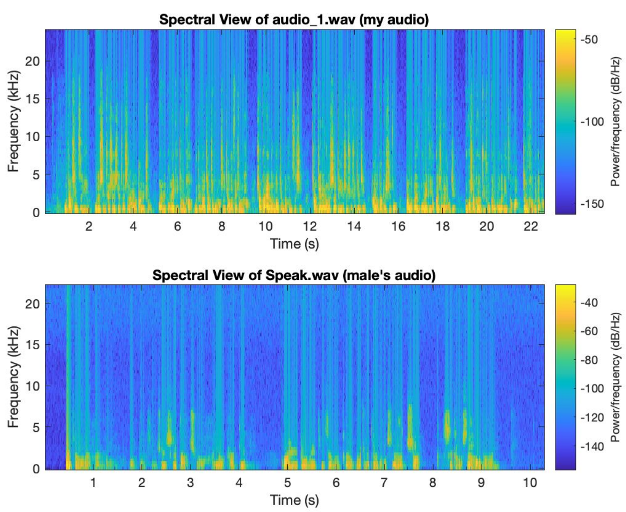


Figure (x) is the concatenation of the previous 3 spectrogram() so I think there is nothing to add.

Figure (x) compare_with_girl/spectral_combination.png

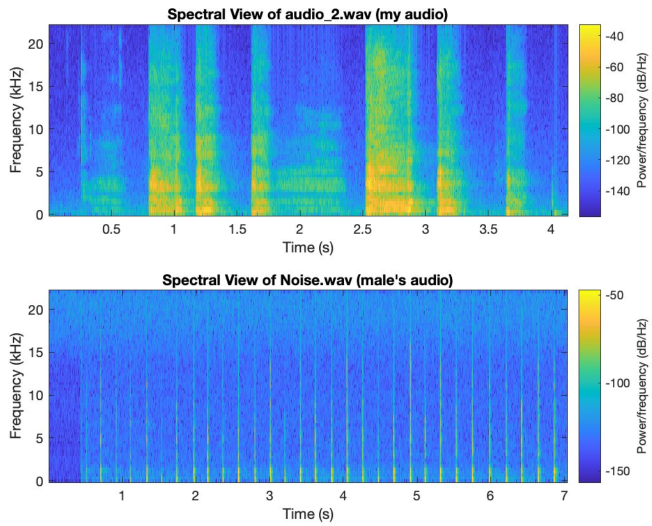
2. Comparison between my audio files spectral view and another male classmate's audio files. The results can be found in **"./result/compare_with_male"**.
   a. Speaking



From Figure (y) it is quite obvious that my voice's frequency range is larger than the other person. And I think the reason is that this is the difference between male and female voice. Female voice ranges larger in frequency compare to male.

Figure (y) compare_with_male/spectral_speaking.png

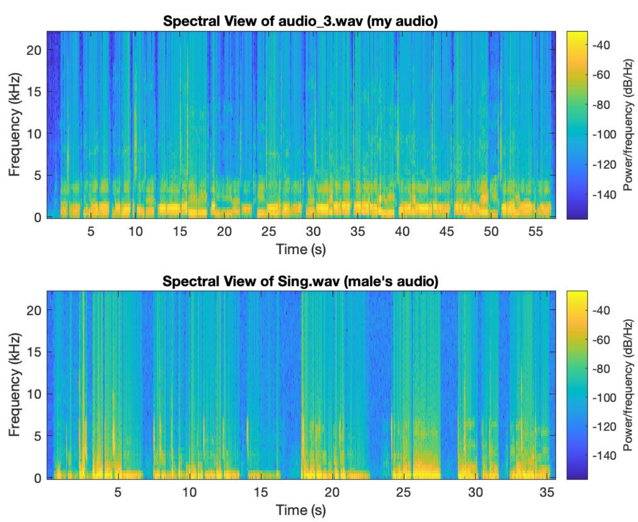
b. Noise



From Figure (z) I can see that the other person's noise is don't have much power compared to mine and the frequency is quite low too~ This is because making noise can take many different forms. This may be teeth shattering or the sound of knocking on items.

Figure (z) compare_with_male/spectral_noise.png

c. Singing



From Figure (aa) I have no comment about it. Because the two spectral view looks normal when singing the frequency range and power looks bigger. There is only one thing I'm surprised at that is although the other person is male he also got high-frequency voice too.

Figure (aa) compare_with_male/spectral_singing.png

d. Combination

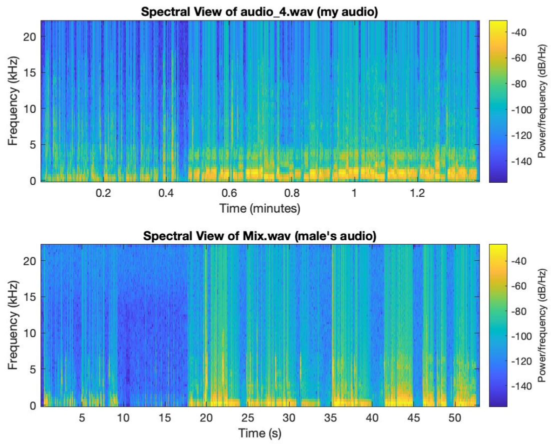


Figure (ab) is the combination of the previous 3 spectrogram() so I think there is nothing to add here.

Figure (ab) compare_with_male/spectral_combination.png

Discussion

I think throughout the whole assignment there is one thing worth discussing that is the result of the comparison of the frequency component. There are some results that make me confused. The weird results are shown in Figure (m), (o), (q), (s), (t). In these 5 figures, my voice magnitudes are always near 0 which I think it isn't possible. So I look into all the sample points I choose and find out that in these figures the resulting magnitude of my voice is like $10^{-8} \sim 10^{-7}$ and the other person is like $10^{-7} \sim 10^{-6}$. There is 10 times the gap which I don't think it's reasonable, especially when comparing with another male classmate. I think having this result is because of the way I choose samples. Maybe the chosen sample frequencies corresponding magnitude just small in coincidence. If **bar3()** have a better visualization when there are more sample points then I think I can choose more samples to see the difference between the two frequency component and have a more accurate result. I conclude the bad results of this assignment due to the lack of sample points.

Difficulties and Problems

There are 2 difficulties I encountered in this assignment.

1. Subplot showing when I didn't specify it
   In the comparison of the spectral view, I use subplot to show the two people's audio spectral view in the convenience to do the comparison. But when the program executes to the code after the subplot all my figures become subplot which this is not I want. Then I found out that when saving plot using **"gcf"** then the program will read the condition right now. If we didn't tell the program I don't want subplot manually then the program will always think I want it cause I set up before. The solution for this is to type **close all hidden** and **axes** to close the special plot specification whenever I want.
2. A great way to visualize the comparison of the spectral view

The spectral view got **3 pieces of information: in the time domain each frequencies power value.** I can't come up with a graph that can compare the 3 pieces of information simultaneously so I just show the two spectral view side by side and use the words as the discussion at last.