MIR - HW0 -Report

彭麒任

cloud drive link: https://drive.google.com/drive/u/4/folders/1Di79YIJdo6KXicygmjk8I8uyP9Skp3xn

Agenda

- Task 1
- Task 2
- Task 3

Task 1

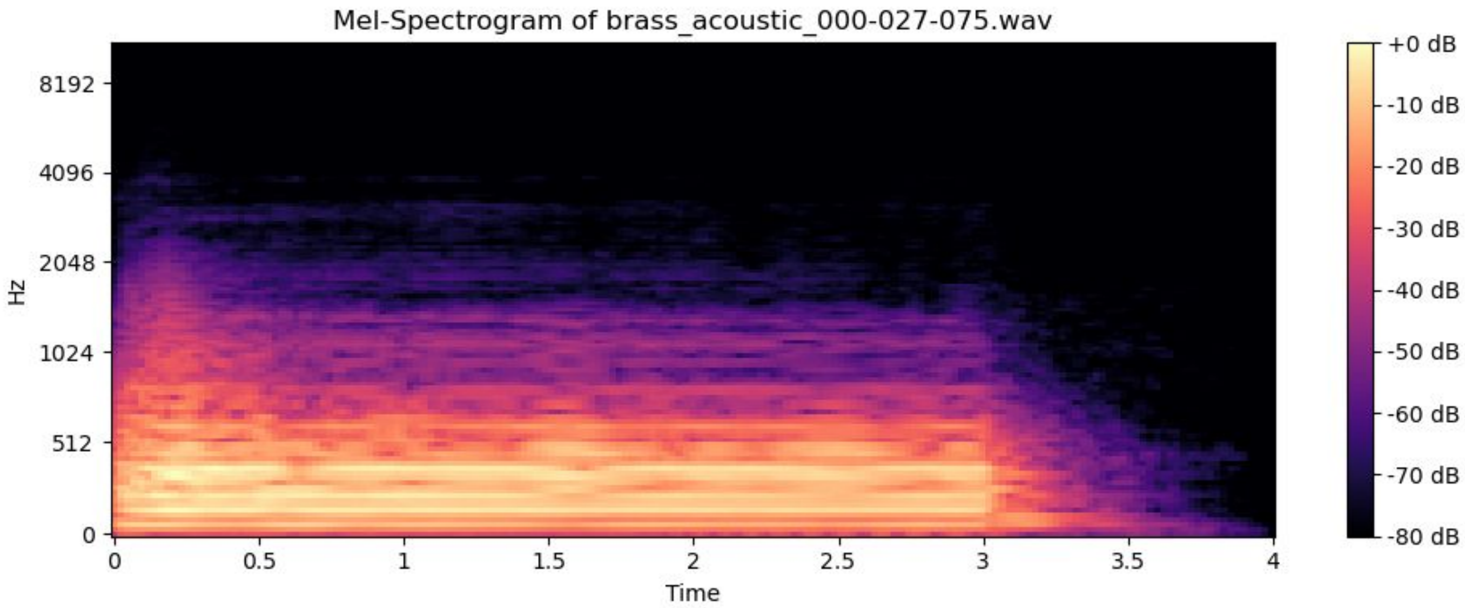
- Selected pitch: 27, 51, 23
- Selected instrument: brass, synth_lead, reed

- Selected .wav file:
  - brass_acoustic_000-027-075
  - synth_lead_synthetic_007-027-100
  - reed_acoustic_026-027-100
  - brass_acoustic_008-051-075
  - synth_lead_synthetic_005-051-050
  - reed_acoustic_059-051-025
  - brass_acoustic_030-023-050
  - synth_lead_synthetic_010-023-100
  - reed_acoustic_021-023-050

Task 1

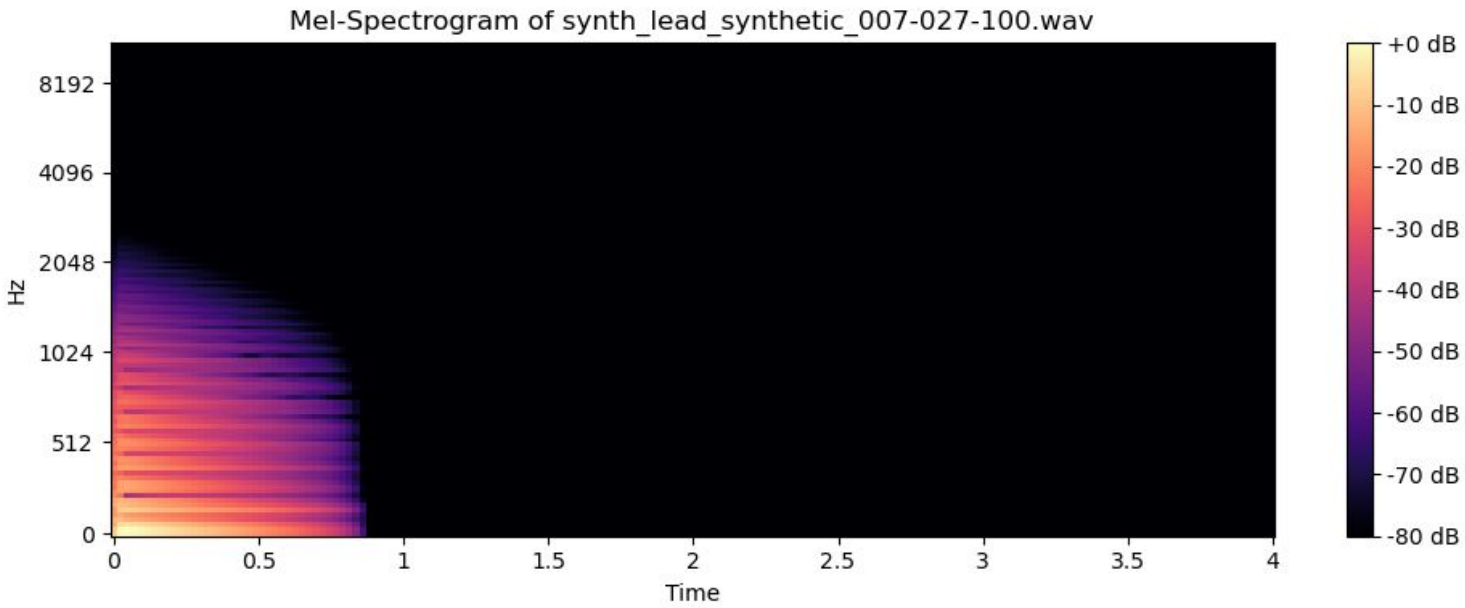
- use ***librosa.feature.melspectrogram*** to generate Mel Spectrogram and use ***matplotlib.pyplot*** to plot it

Task 1

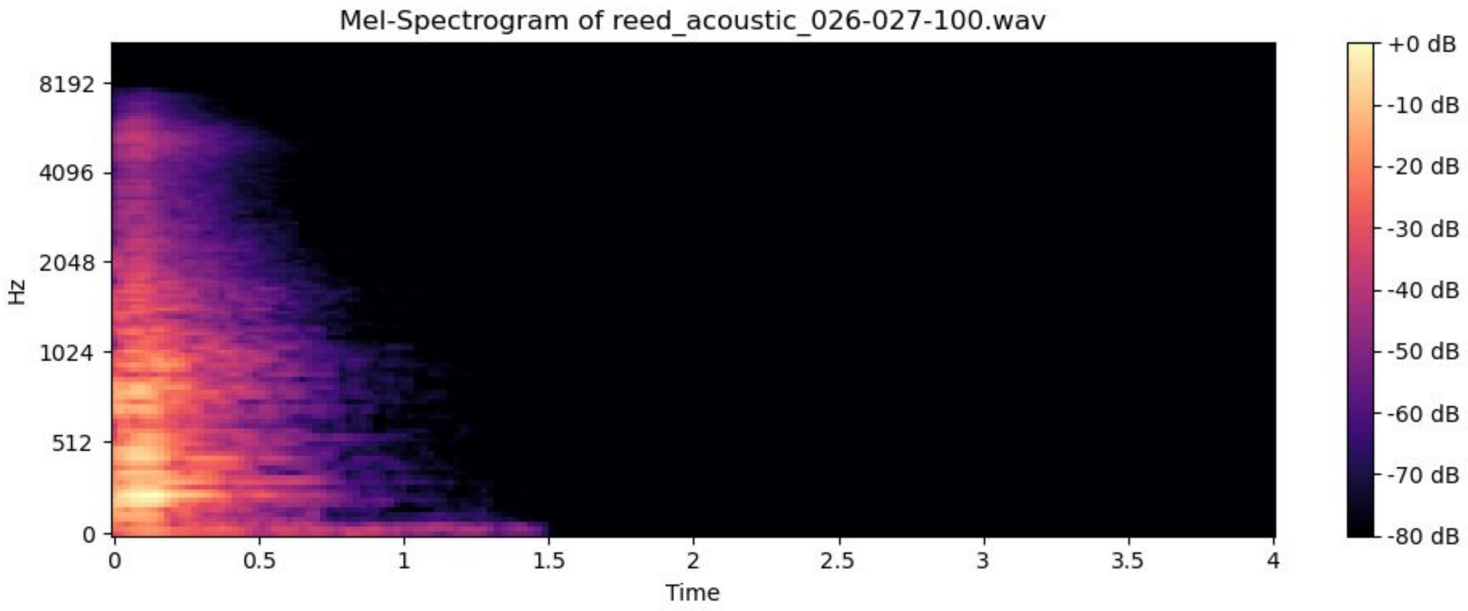


Mel-Spectrogram of brass_acoustic_000-027-075.wav
Hz
8192
4096
2048
1024
512
0
0
0.5
1
1.5
2
2.5
3
3.5
4
Time
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

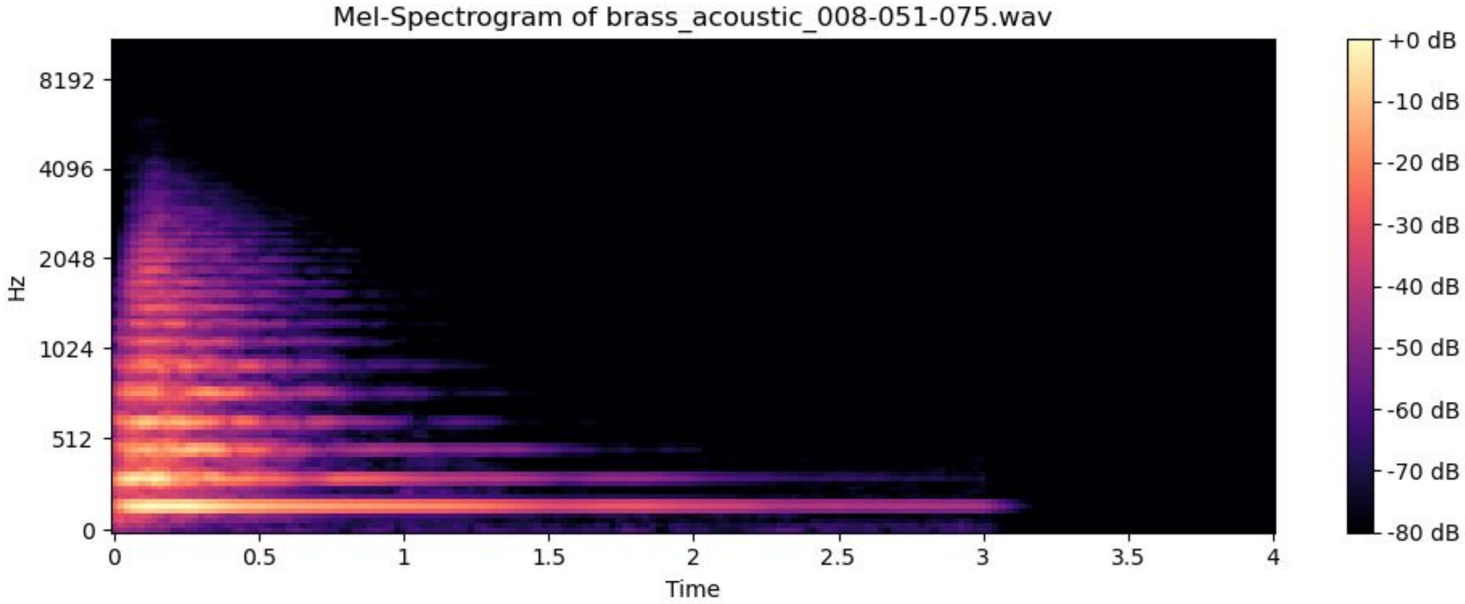
Task 1


Mel-Spectrogram of synth_lead_synthetic_007-027-100.wav
8192
4096
2048
1024
512
0
Hz
0
0.5
1
1.5
2
2.5
3
3.5
4
Time
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

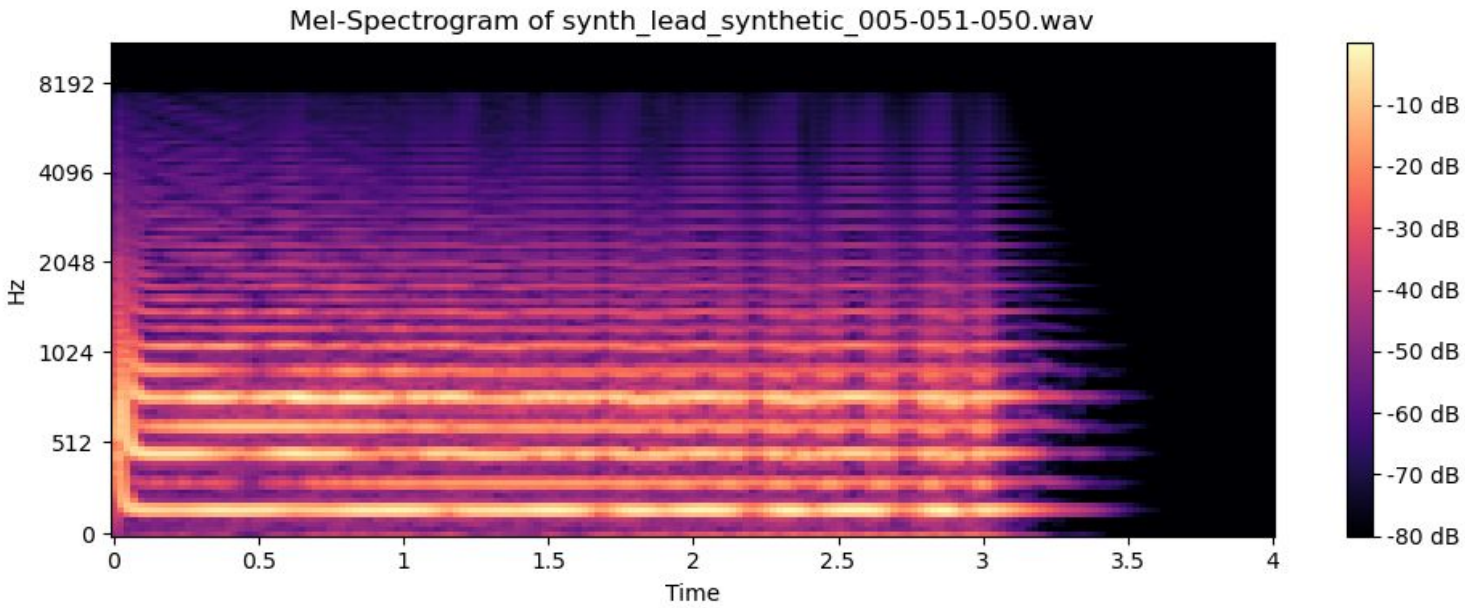
Task 1


Mel-Spectrogram of reed_acoustic_026-027-100.wav
Hz
8192
4096
2048
1024
512
0
0
0.5
1
1.5
2
2.5
3
3.5
4
Time
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

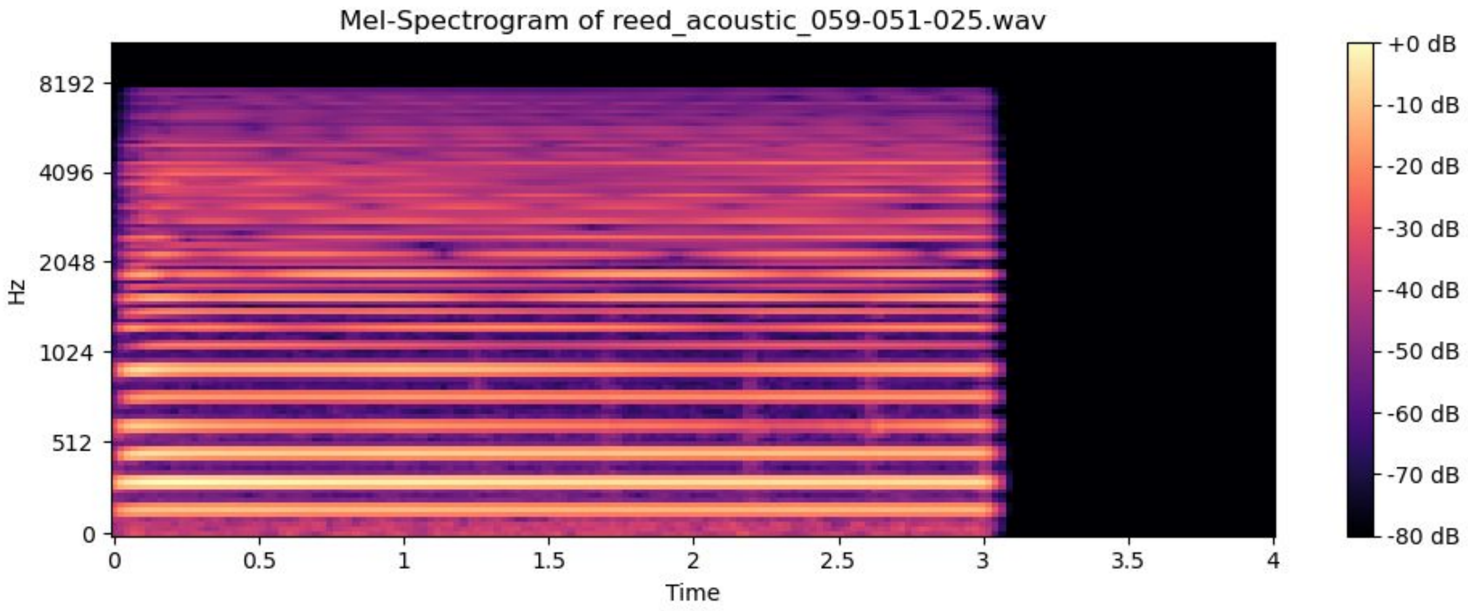
Task 1


Mel-Spectrogram of brass_acoustic_008-051-075.wav
Hz
8192
4096
2048
1024
512
0
0
0.5
1
1.5
2
2.5
3
3.5
4
Time
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

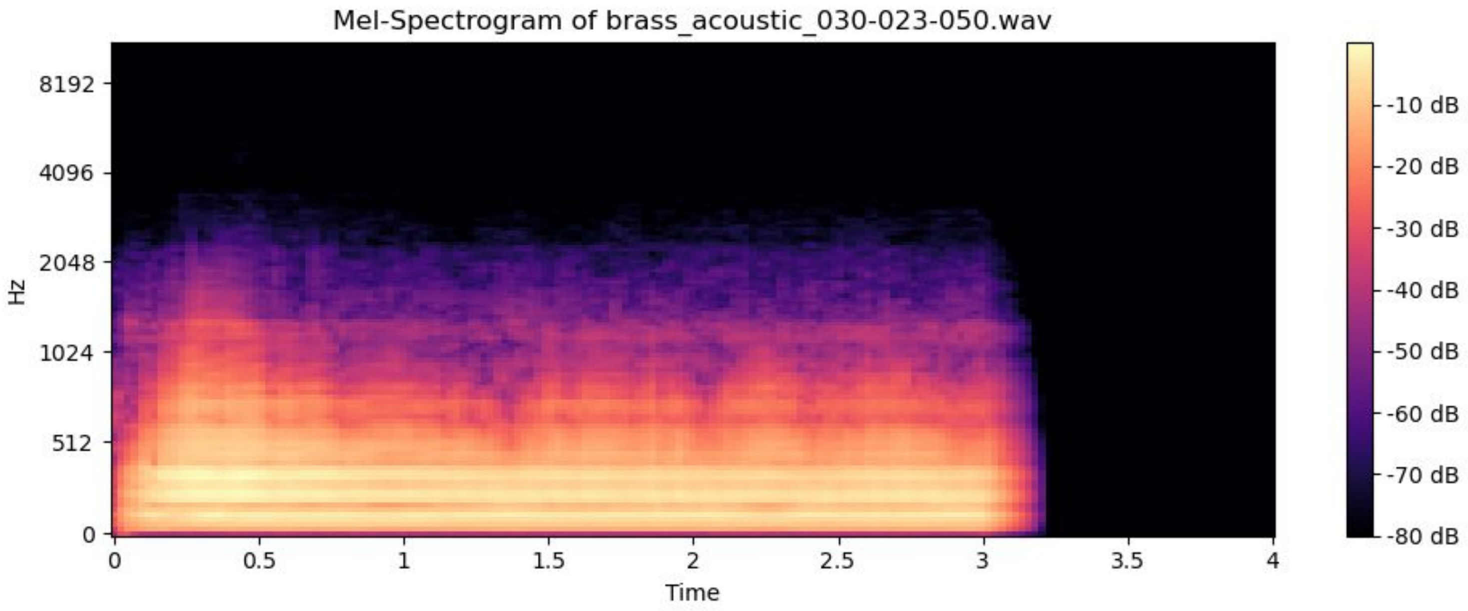
Task 1


Mel-Spectrogram of synth_lead_synthetic_005-051-050.wav
Hz
8192
4096
2048
1024
512
0
0
0.5
1
1.5
2
2.5
3
3.5
4
Time
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

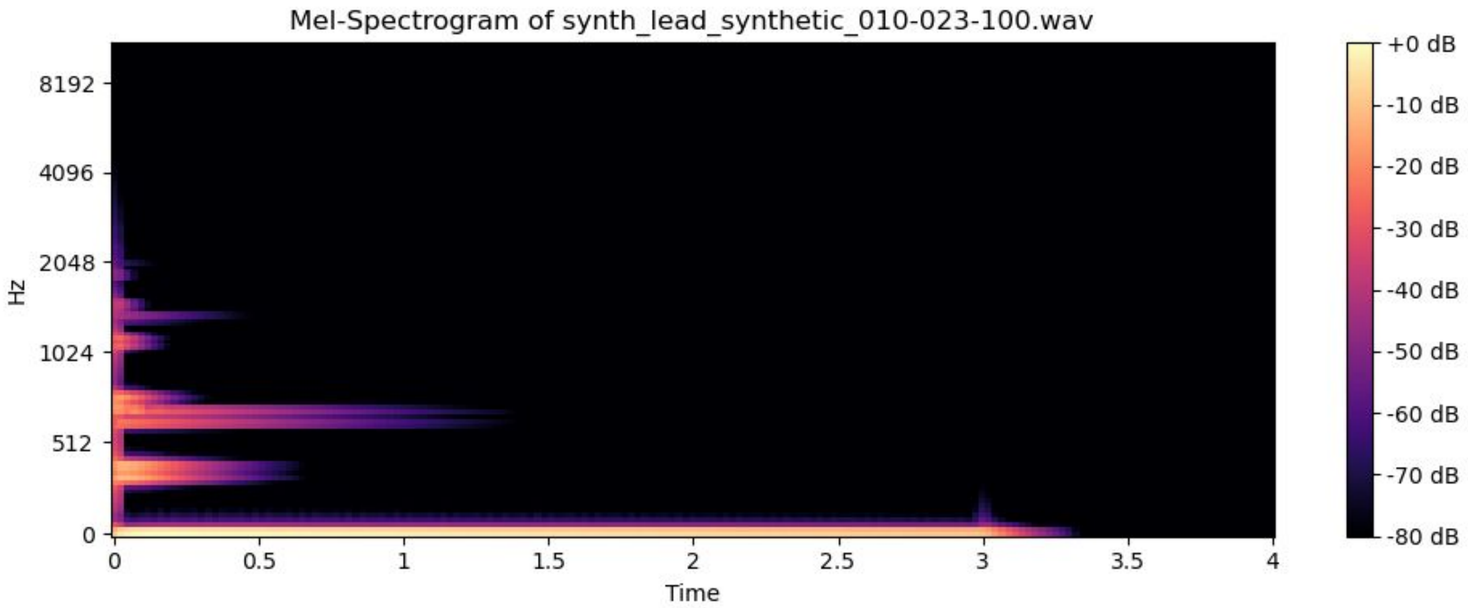
Task 1


Mel-Spectrogram of reed_acoustic_059-051-025.wav
Hz
8192
4096
2048
1024
512
0
Time
0
0.5
1
1.5
2
2.5
3
3.5
4
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

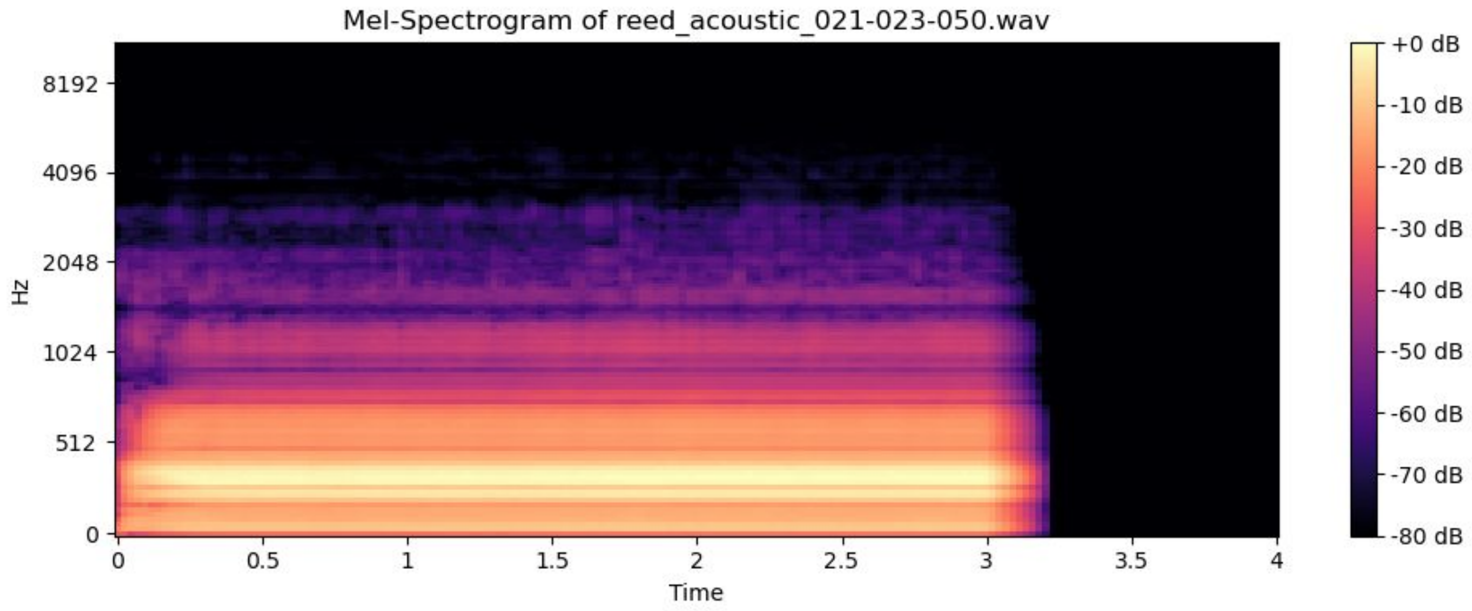
Task 1


Mel-Spectrogram of brass_acoustic_030-023-050.wav
Hz
8192
4096
2048
1024
512
0
0
0.5
1
1.5
2
2.5
3
3.5
4
Time
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Task 1


Mel-Spectrogram of synth_lead_synthetic_010-023-100.wav
8192
4096
2048
1024
512
0
Hz
0
0.5
1
1.5
2
2.5
3
3.5
4
Time
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Task 1



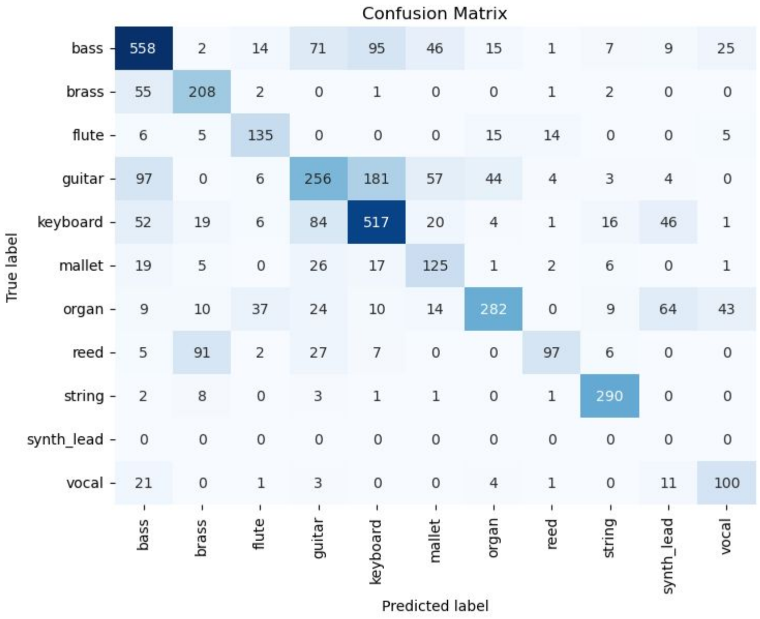
Mel-Spectrogram of reed_acoustic_021-023-050.wav
Hz
8192
4096
2048
1024
512
0
0
0.5
1
1.5
2
2.5
3
3.5
4
Time
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Task 2

- Using ***librosa.feature*** to extract two features:
  - **Spectral_contrast**: It accentuate the frequency difference between each instrument, I think it's a great way to differentiate between different instruments
  - **MFCC**: According to Wikipedia, it approximates the human auditory system's response more closely
- After extract the two features, they're normalized using ***StandardScaler()***. After that, I flatten them and concat them, creating a [1, 4671] vector for each .wav file. For validation data and test data, the same pre-processing is appied

Task 2

- Model: using ***knn***, with ***k = 3***
- **Top 1** accuracy: **0.626953125**
- **Top 3** accuracy: **Null**
    - It runs for 25 minutes but still not finished
    - It is likely because KNN is a lazy learner, and predicting the test data takes a long time, especially for labels that are far from the true label



Task 2

- What I found:
    - The model can distinguish vocals and guitar from other instruments very well
- Inprovements that can be done:
    - Try different model, such as SVM and random forest
    - Try larger ***k*** value in ***knn***, to whether it perform better or not

Task 3

- use ***librosa.feature.melspectrogram*** to extract Mel Spectrogram feature and use ***librosa.power_to_db*** to extract feature with log scaling
- Encode the ***.wav*** file into integer encoding, label from 0 to 10, imply different instruments
- Reference the suggested Short Chunk CNN model(Here), and modify some parts
  - Change the loss function from binary cross entropy loss to cross entropy loss
  - Modify the input, calculate the mel-spectrogram beforehand, instead of calculating in the the model

Task 3 (Mel Spectrogram, without log scaling)

- Top-1 Accuracy: 6.57%
- Top-3 Accuracy: 26.29%
- The perform is **POOR**, There must be something wrong with the model, or the input
- Same error happened with **Mel Spectrogram with log scaling**

```
Confusion Matrix:
[[  0   0   0   0   0   0   0   0   0   0 843]
 [  0   0   0   0   0   0   0   0   0   0 269]
 [  0   0   0   0   0   0   0   0   0   0 180]
 [  0   0   0   0   0   0   0   0   0   0 652]
 [  0   0   0   0   0   0   0   0   0   0 766]
 [  0   0   0   0   0   0   0   0   0   0 202]
 [  0   0   0   0   0   0   0   0   0   0 502]
 [  0   0   0   0   0   0   0   0   0   0 235]
 [  0   0   0   0   0   0   0   0   0   0 306]
 [  0   0   0   0   0   0   0   0   0   0 141]
 [  0   0   0   0   0   0   0   0   0   0   0]]
```

Task 3

- Inprovements that can be done:
    - Implement the model **properly**