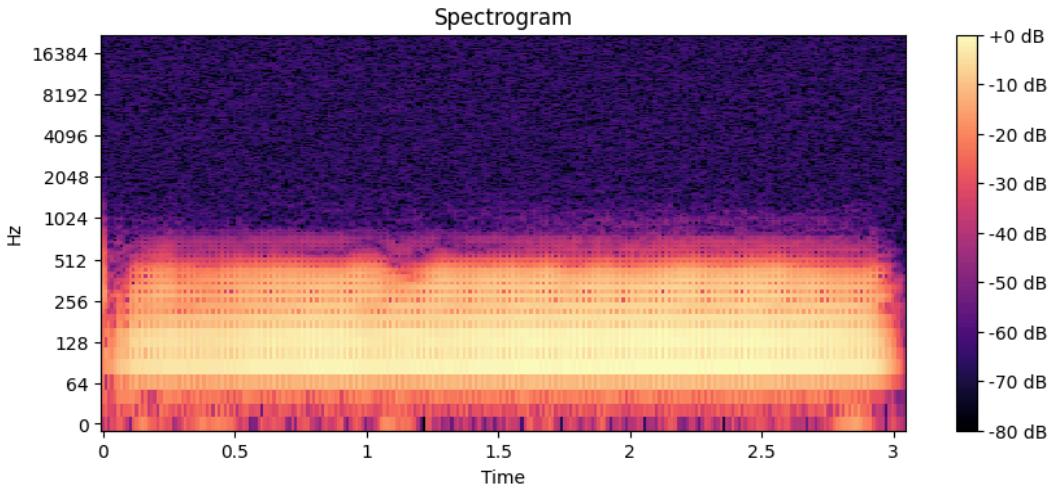
1. Meaning of Track ID: Cb-ord-A2-ff-1c-R100d

- **Cb:** Instrument abbreviation (Contrabass)
- **ord:** Technique abbreviation (ordinario, normal or standard playing technique)
- **A2:** Pitch
- **ff:** Dynamics (fortissimo, loud)
- **1c:** The string which is being bowed (On the first string)
- **R100d:** Resampled by decreasing pitch by a semitone, i.e. 100 cents (R100d)

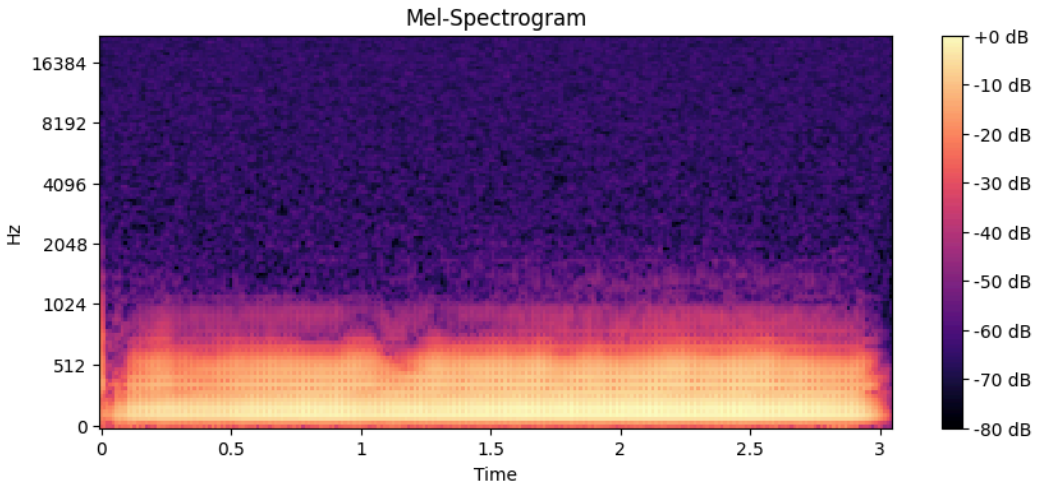
2. Find an audio file you like and plot its spectrogram, mel-spectrogram, and MFCC using the librosa functions. Please explain the differences between these features.

- **Spectrogram:** Visualize the spectrum of frequencies in a sound signal as they vary with time.

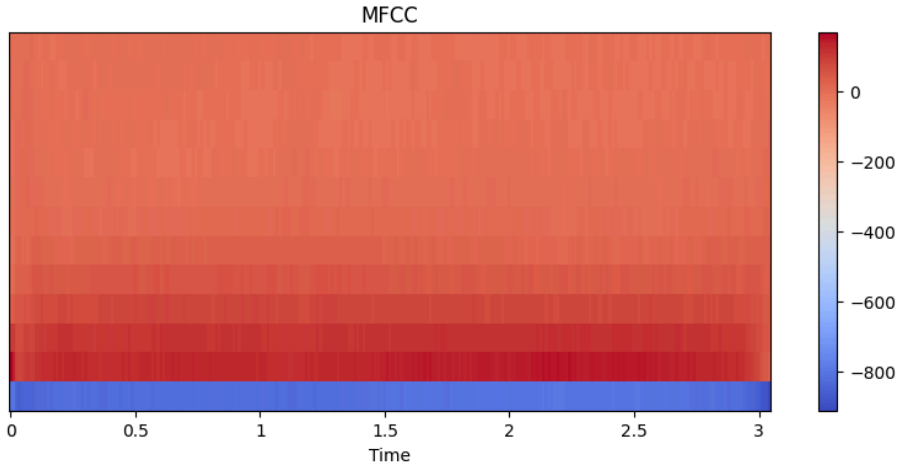
Song choosed:"BTb-ord-F#1-pp-N-N"



- **Mel-Spectrogram:** The image shows a darker part on higher frequency compared to the spectrogram above , implying the Mel-Spectrogram emphasizes lower frequencies more than Spectrogram, which is more aligned with human hearing.



- **MFCC:** Derived from the log mel-spectrogram, capturing the envelope of the spectrum.

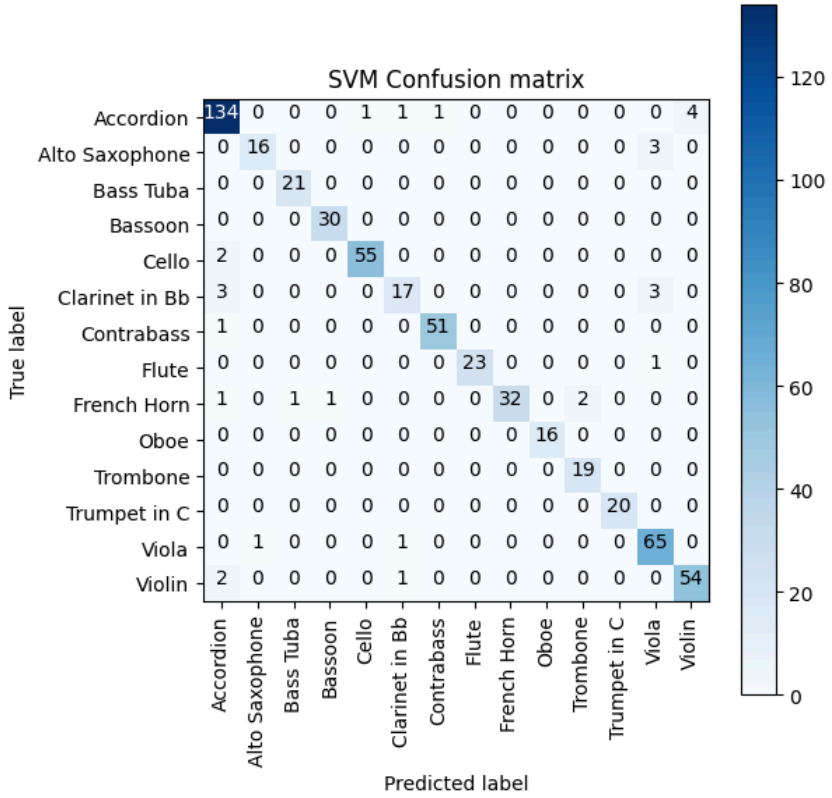


3. Instrument classification task.

a. SVM instrument classification performance:

Validation score 1	**0.9408740359897172**
Validation score 2	**0.9357326478149101**
Validation score 3	**0.9201030927835051**
Validation score 4	**0.9252577319587629**
Validation score 5	**0.9381443298969072**
Validation score 6	**0.9355670103092784**
Test score	***0.9485420240137221***
ACC	***0.949***
precision	***0.949***
recall	***0.949***

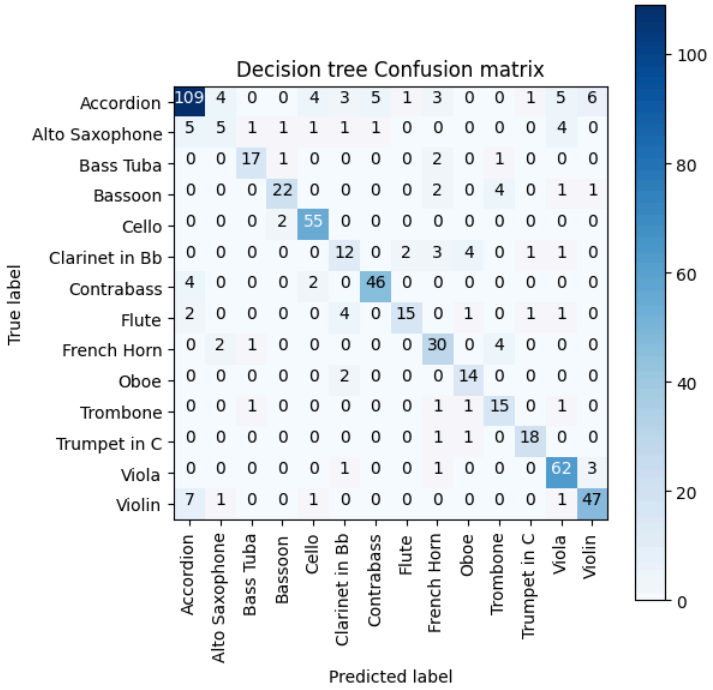
- I use Linear SVM to implement it, it demonstrates a high degree of diagonal concentration, indicating that it has high accuracy across most instrument classes. (I've also tried non-linear SVM (RBF), but the outcome was worse than linear SVM)
- Based on the confusion matrix below, The Bass Tuba, Bassoon, Oboe, Trombone and Trumpet in C are well-classified, maybe because their sound is unique.
- Many instruments are easily being misclassified to Accordion, maybe because there is more of an imbalance with the dataset, with more Accordion data in it.



b. Decision Tree instrument classification performance:

Validation score 1	**0.794344473007712**
Validation score 2	**0.7455012853470437**
Validation score 3	**0.7835051546391752**
Validation score 4	**0.8247422680412371**
Validation score 5	**0.788659793814433**
Validation score 6	**0.7422680412371134**
Test score	***0.8010291595197255***
ACC	***0.801***
precision	***0.801***
recall	***0.801***

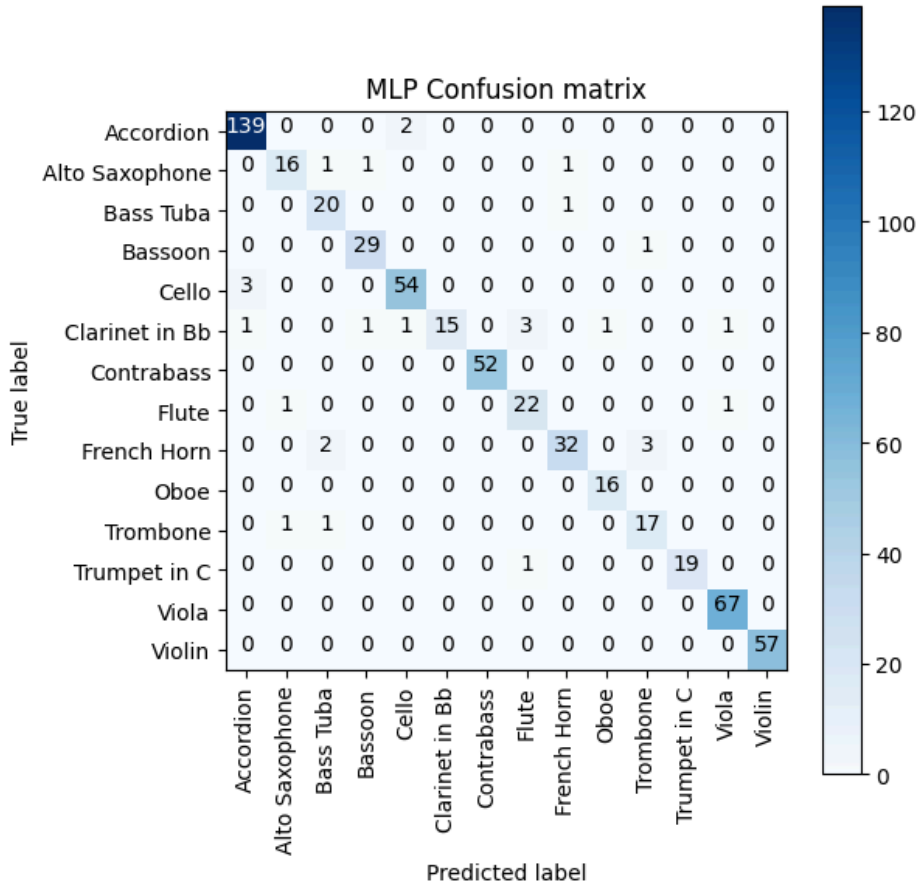
- The Decision Tree has a less concentrated diagonal, implying more misclassifications across the board compared to the SVM and MLP.
- Based on the confusion matrix below, the Accordion class has a notable number of misclassifications.
- Based on the confusion matrix below, the Accordion and Violin class are easily being confused.



c. MLP instrument classification performance:

Validation score 1	**0.9254498714652957**
Validation score 2	**0.9511568123393316**
Validation score 3	**0.9201030927835051**
Validation score 4	**0.9536082474226805**
Validation score 5	**0.9458762886597938**
Validation score 6	**0.9381443298969072**
Test score	***0.9519725557461407***
ACC	***0.952***
precision	***0.952***
recall	***0.952***

- The MLP shows a highly concentrated diagonal, suggesting a strong classification performance similar to the SVM.
- Based on the confusion matrix below, the Accordion, Contrabass, Viola and Violin classes have very high correct classification rates.



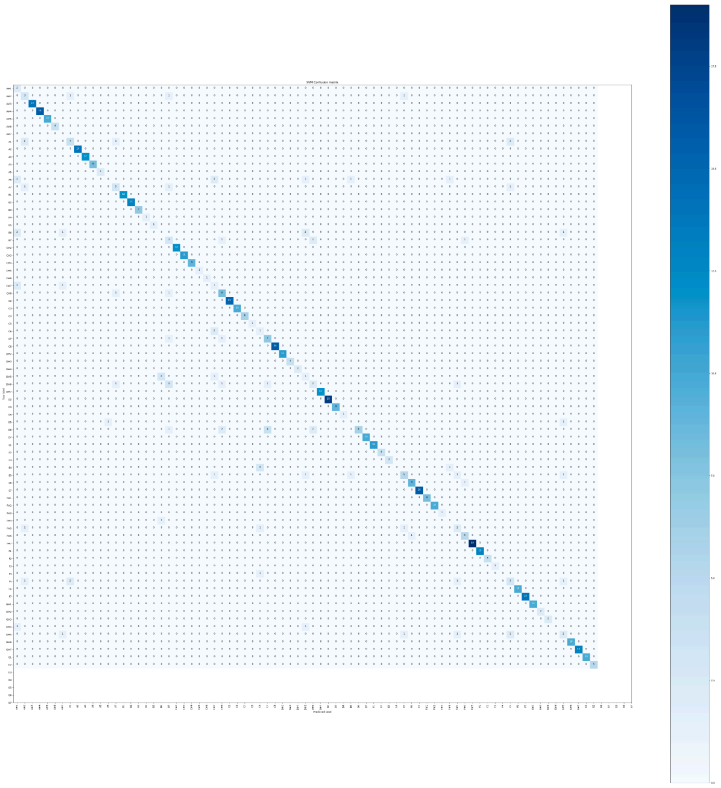
In summary, both the SVM and MLP models perform well compared to Decision trees, with both higher than 90% accuracy . Based the the confusion matrix results below, MLP seems to strike a good balance between classifying different instruments correctly and maintaining fewer misclassifications.

4. Pitch classification task.

a. SVM pitch classification performance:

Validation score 1	**0.8226221079691517**
Validation score 2	**0.8329048843187661**
Validation score 3	**0.8273195876288659**
Validation score 4	**0.845360824742268**
Validation score 5	**0.845360824742268**
Validation score 6	**0.8350515463917526**
Test score	***0.8507718696397941***
ACC	***0.851***
precision	***0.851***
recall	***0.851***

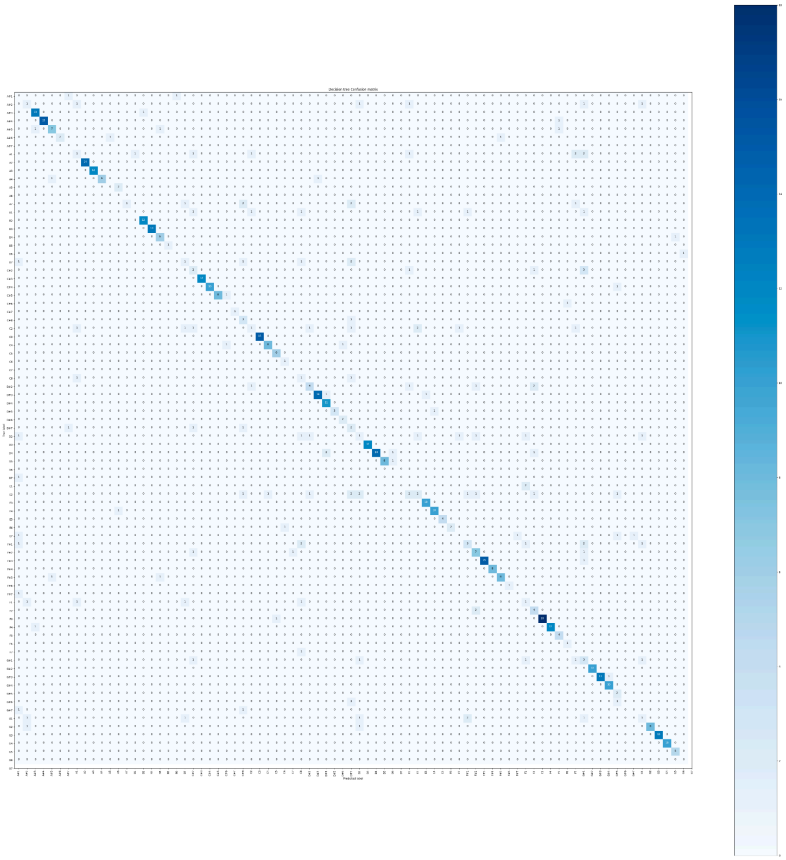
- The SVM has shown consistent and high validation scores around 0.82 to 0.84, indicating good generalization on the validation sets, and its test score of 0.851 suggests that the SVM has performed well on unseen data, maintaining accuracy close to its validation scores.
- The SVM confusion matrix shows a relatively dense diagonal, indicating a high number of correct predictions across different pitch classes.



b. DecisionTree instrument pitch performance:

Validation score 1	**0.6915167095115681**
Validation score 2	**0.7043701799485861**
Validation score 3	**0.6314432989690721**
Validation score 4	**0.7087628865979382**
Validation score 5	**0.6804123711340206**
Validation score 6	**0.6958762886597938**
Test score	***0.7358490566037735***
ACC	***0.736***
precision	***0.736***
recall	***0.736***

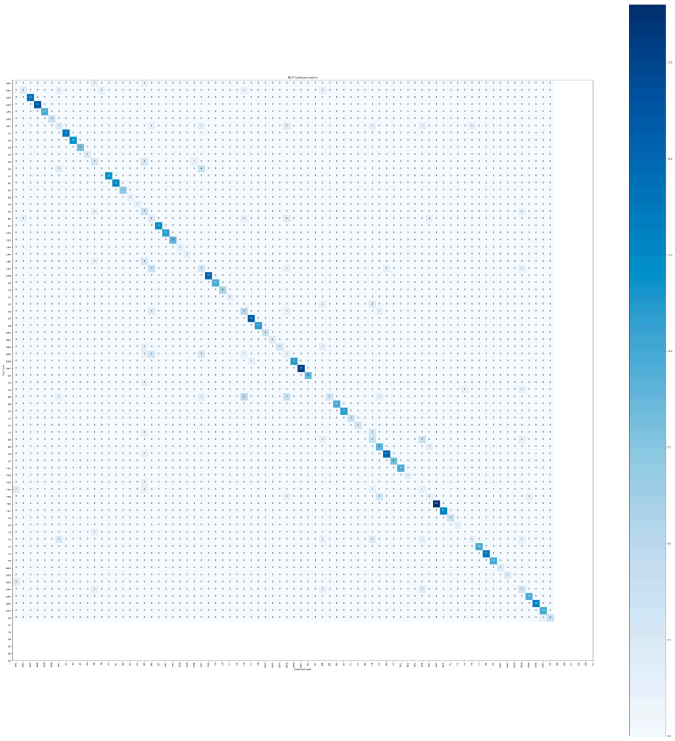
- Decision tree validation scores are noticeably lower and more varied, ranging from about 0.63 to 0.71. This suggests that the decision tree may be more sensitive to the specific splits of the data and may not generalize as well as SVM.
- The test score of 0.736 shows that the decision tree's performance dropped when confronted with the test set.
- The performance across different pitch classes seems to be less consistent compared to the SVM, with some classes having higher misclassification rates.



c. MLP instrument pitch performance:

Validation score 1	**0.7737789203084833**
Validation score 2	**0.794344473007712**
Validation score 3	**0.7963917525773195**
Validation score 4	**0.8195876288659794**
Validation score 5	**0.7989690721649485**
Validation score 6	**0.8015463917525774**
Test score	***0.8096054888507719***
ACC	***0.81***
precision	***0.81***
recall	***0.81***

- MLP's validation scores are more consistent than those of the decision tree but slightly lower than those of the SVM, with values ranging from approximately 0.77 to 0.82.
- The confusion matrix is similar to the SVM, the diagonal is quite dense, indicating a high number of correct classifications.



Based on the above results, the SVM model appears to be the best performing of the three for this pitch classification task. It has consistent validation and test scores. In contrast, the Decision Tree performs the worst among the three, with the lowest test score and greater variability in validation scores. The MLP falls between the two, with good performance but not quite reaching the level of the SVM.

For overall results of pitch classification performance compared to instrument classification performance, it seems that the MFCC spectrogram is better used for instrument classification, since it can capture timbral characteristics and the broad phonetic features of different audio signals. The Mel spectrogram may be considered better for pitch classification.

5. Please use GridSearch to find the best hyperparameters based on the cross-validation score for SVM, Decision Tree, and MLP models.

- **Tune the hyperparameters C and Gamma for SVM.**
  - Test score with best parameters is : 0.8593481989708405, which is slightly better than the previous score: 0.8507718696397941. Note that I found out that the gamma parameter did not affect the linear SVM but affected the RBF

SVM, so the gamma parameter is not considered in order to save time, since I trained linear SVM in the previous case.

- Best params: {'C': 0.1}
- A low C makes the decision surface smooth and simple, this can lead to a model that generalizes better to unseen data (better testing accuracy)

- **Tune hyperparameters Criterion and Maximum Depth of DecisionTree**
  - Best params: {'criterion': 'entropy', 'max_depth': 15}
  - Test score with best parameters: 0.7495711835334476, which is better than the previous score: 0.7358490566037735. Besides, the Best training score: 0.7231722453378563 is also better than all the previous training scores.
  - It chooses the split which has the highest information gain, and a max_depth of 15 was found to be optimal in the best performance on this particular dataset.
- **Tune the Hidden Layer Sizes and the strength of the L2 Regularization Term for MLP**
  - Test score with best parameters: 0.8353344768439108 is better than the previous Test score: 0.8096054888507719, with Best params: {'alpha': 0.1, 'hidden_layer_sizes': (100, 50)}.
  - A larger alpha results in more regularization and a simpler, smoother model that's less likely to overfit. The best structure of layers and the number of nodes in each layer was found to be (100, 50), meaning two hidden layers, the first with 100 neurons and the second with 50 neurons.