Junhao Wang, MUMT 621

Final Project Summary

Electric Guitar Transcription with Plucking Style Classification

## Objective

The objective of this project is to study and implement a guitar transcription system following the procedure proposed by Kehling et al. (2014). Specifically, automatic transcription is performed on isolated monophonic guitar recordings. Clean signals without any processing of audio effects are used. Besides general score-related parameters such as pitch and onset, the guitar-specific string number and plucking style will be estimated on the note level. The plucking styles considered in this project are fingerstyle, picked, and muted.

## Dataset Overview

The data used in this project are two subsets of the IDMT-SMT-GUITAR dataset (Kehling et al. 2014).

The first subset contains single note and chord recordings, each corresponding to one pluck. The single-pluck recordings have corresponding annotations regarding the pitch, onset, string number, and plucking style. However, all recordings in the first subset are played with a pick. There is thus only one plucking style involved, making it inappropriate for training and testing the plucking style classifier. Therefore, the monophonic portion of the first subset was used in two tasks: developing the string classifier and testing the pitch estimation algorithm.

The second subset contains realistic guitar phrases with note-level annotations regarding the pitch, onset, string number, and plucking style. It covers all the three plucking styles. According to the annotated onsets and offsets, the 261 phrases were split into a total of 4,812 single notes, which were then used for training and testing the plucking style classifier. Among the guitar phrases, some of them are monophonic solo-like performances, while the others have a polyphonic component to some extent. For plucking style classification, all 4,812 notes were used. For string number estimation and pitch estimation, phrases with significant polyphonic content were removed. The remaining 2,988 notes were considered monophonic and used for developing the string number classifier.

## Experiments and Results

This section presents the experiments conducted in this project. For onset detection and pitch estimation, existing implementations in the Essentia library (Bogdanov et al. 2013) were used. For the classification tasks, as a general procedure, support vector machine (SVM) classifiers with Gaussian kernel were applied. Before training the model, the dataset was split into the training set (80%) and the test set (20%). Ten-fold cross validation was performed with grid search to finetune the hyperparameters. The final accuracy reported is obtained on the test set.

### String Number Classification

For most solo performances, only single notes are played, and the guitar recordings can be roughly considered monophonic. In this case, one pluck involves only one string. In this project, we only consider the monophonic string number estimation, that is, only one string number is estimated for each single note recording.

As a preliminary test, an SVM classifier was trained on the non-chord portion of the first subset, which contains 312 monophonic single notes. The Mel frequency cepstrum coefficients (MFCC) were extracted on the frame level and aggregated to seven statistic values per note: maximum, minimum, mean, variance, median, skewness and kurtosis. The six classes (corresponding to the six strings) are perfectly balanced, each containing 52 notes.

Trained on the MFCC features only, the performance of the monophonic string number classifier was evaluated using ten-fold cross validation and an average accuracy of 97.57% was achieved with a standard deviation of 0.029. The test set accuracy was 93.65%.

It was also attempted to use the same method to train and test a separate string number classifier on the second subset, where a test set accuracy was achieved at 91.47%. Two reasons might explain the performance discrepancy. First, many notes in the second subset are played with special expression styles (e.g., harmonic, dead note), which considerably affect the timbre. This introduces noise to the features. Second, there are many overlaps in time between consecutive notes in the second subset. Although an effort was made to remove the polyphonic phrases from the dataset, some of the single note recordings obtained from splitting the phrases still contain a short polyphonic part at the beginning or the end.

Combining the two subsets, a third SVM classifier was trained and tested on the combined dataset. As a data cleansing step, we removed the notes whose pitch can only be played on one string. For example, if a note has a pitch of 40 (MIDI note number), it corresponds to the note E2, which can only be played on the lowest string on a typical guitar. In this case, the classifier can just learn the pitch and guess the string number. The resulting accuracy will not truly reflect whether the classifier has learned the subtle timbre difference between notes played on different strings. Therefore, the notes with a pitch higher than 83 (B5) or lower than 45 (A2) were removed. The remaining 3,235 notes were used to train and test the model. The class distribution is shown in Table 1.

Table 1. Class distribution of the combined dataset for string number classification

| String number | Number of notes | Percentage (%) |
|---|---|---|
| 1 | 212 | 6.55 |
| 2 | 614 | 18.98 |
| 3 | 579 | 17.90 |
| 4 | 823 | 25.44 |
| 5 | 645 | 19.94 |
| 6 | 362 | 11.19 |

The model achieved an accuracy of 91.50% on the test set. The confusion matrix is shown in Figure 1.

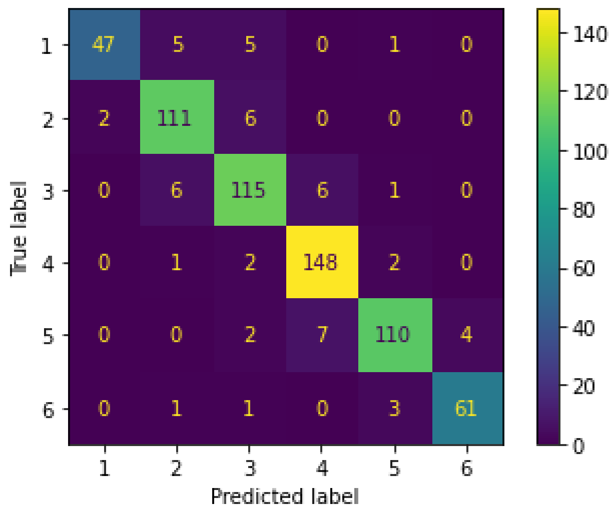


Figure 1. String number estimation accuracy results displayed in a confusion matrix

## Plucking Style Classification

The plucking style classifier differentiates between three classes: fingerstyle (FS), picked (PK), and muted (MU). The single notes obtained from the phrases in the second subset were used to train and test this classifier. The class distribution is shown in table 2.

Table 2. Class distribution of the dataset used for plucking style classification

| Plucking style | Number of notes | Percentage (%) |
|---|---|---|
| FS | 1318 | 27.39 |
| PK | 2090 | 43.43 |
| MU | 1404 | 29.18 |

Using the same set of features as extracted for the string number classifier, the model achieved an average accuracy of 90.99% in the 10-fold cross validation with a standard deviation of 0.0822. This model achieved 90.28% accuracy on the test set.

In seeking improvement to this result, the frame-level spectral centroid feature and the fundamental frequency of each note were added to the feature set. Using the new feature set, the cross validation average accuracy rose to 92.17% with a standard deviation of 0.0093. The test set accuracy also rose to 92.31%. The confusion matrix of the test set is shown in Figure 2. It is observed that using the current features, it is relatively more difficult to distinguish between FS and PK, while MU is relatively easy to distinguish from the other two plucking styles.

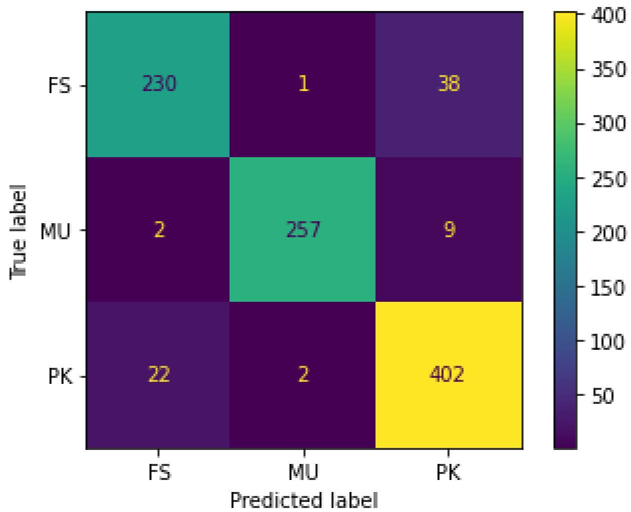


Figure 2. Plucking style classification accuracy results displayed in a confusion matrix

## Conclusion and Future Work

In this project, single note recordings were studied for guitar transcription and estimation of the string number and the plucking style. Onset detection and pitch estimation relied on existing implementations. For string number and plucking style estimation, SVM classifiers were trained, and the average accuracies were reported on the test set. For training and testing the classifiers, the MFCC features worked surprisingly well for both string number classification and plucking style classification. Adding the spectral centroid and fundamental frequency as new features only led to minor improvement.

In the current implementation for string number classification, the input single note recordings are assumed to be monophonic, and the data were filtered accordingly. Future work may focus on extending string number classification to a polyphonic setting, where multiple string numbers are returned for polyphonic recordings such as chords.

For plucking style classification, the algorithm achieved satisfactory performance. A more complete transcription system would require estimation of the expression styles (e.g., string bend) as well. We leave the exploration of expression styles for future work.

## Implementation Note

This project is implemented in Python on Google Colab. The source code in Jupyter Notebooks is available online at https://github.com/jwang44/Plucking-Style-Detection. Essentia (Bogdanov et al. 2013) was used for audio processing and feature extraction. The machine learning workflow and the SVM models were based on scikit-learn (Pedregosa et al. 2011).

## References


Kehling, Christian, Jakob Abeßer, Christian Dittmar, and Gerald Schuller. 2014. "Automatic Tablature Transcription of Electric Guitar Recordings by Estimation of Score- and Instrument-Related Parameters." In *Proceedings of the 17th International Conference on Digital Audio Effects, DAFx-14, Erlangen, Germany, September 1–5, 2014*, 219–26. http://www.dafx14.fau.de/papers/dafx14_christian_kehling_automatic_tablature_trans.pdf.

Bogdanov, Dmitry, Nicolas Wack, Emilia Gómez, Sankalp Gulati, Perfecto Herrera, Oscar Mayor, Gerard Roma, Justin Salamon, José R. Zapata, and Xavier Serra. 2013. "Essentia: An Audio Analysis Library for Music Information Retrieval." In *Proceedings of the 14th International Society for Music Information Retrieval Conference, ISMIR 2013, Curitiba, Brazil, November 4–8, 2013*, 493–98. http://www.ppgia.pucpr.br/ismir2013/wp-content/uploads/2013/09/177_Paper.pdf.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf