Progress Report of Course Project

Project title: Musical Instrument Detection

Group Members: Zishuo Zhao, Haoyun Wang

**Objective:**

To investigate timbre features of different musical instruments and model them via various existing methods for classification. As an extension, we can combine multiple models and/or develop new ones for better and more robust performance.

**Motivations:**

Humans can easily decide the instrument when listening to a piece of music, and if experienced, can even enumerate all instruments in a symphony, which is difficult for computers yet. On the other hand, music is a global language which can even be understood by the illiterate, and it can express delicate emotions in a mystic way. In this project, we are trying to enhance the understanding of music in the view of science.

**System Description:**

It is obvious that spectral characteristics are main features of a musical instrument. But in practice, we have noticed that different instruments have not only different spectral features, and different notes of the same instrument also vary in spectrum. So we use different models and try to determine the key features that human tell instruments apart.

**Methodology:**

Now used for feature extraction: STFT, cepstrum, MFCC, LP coefficient.

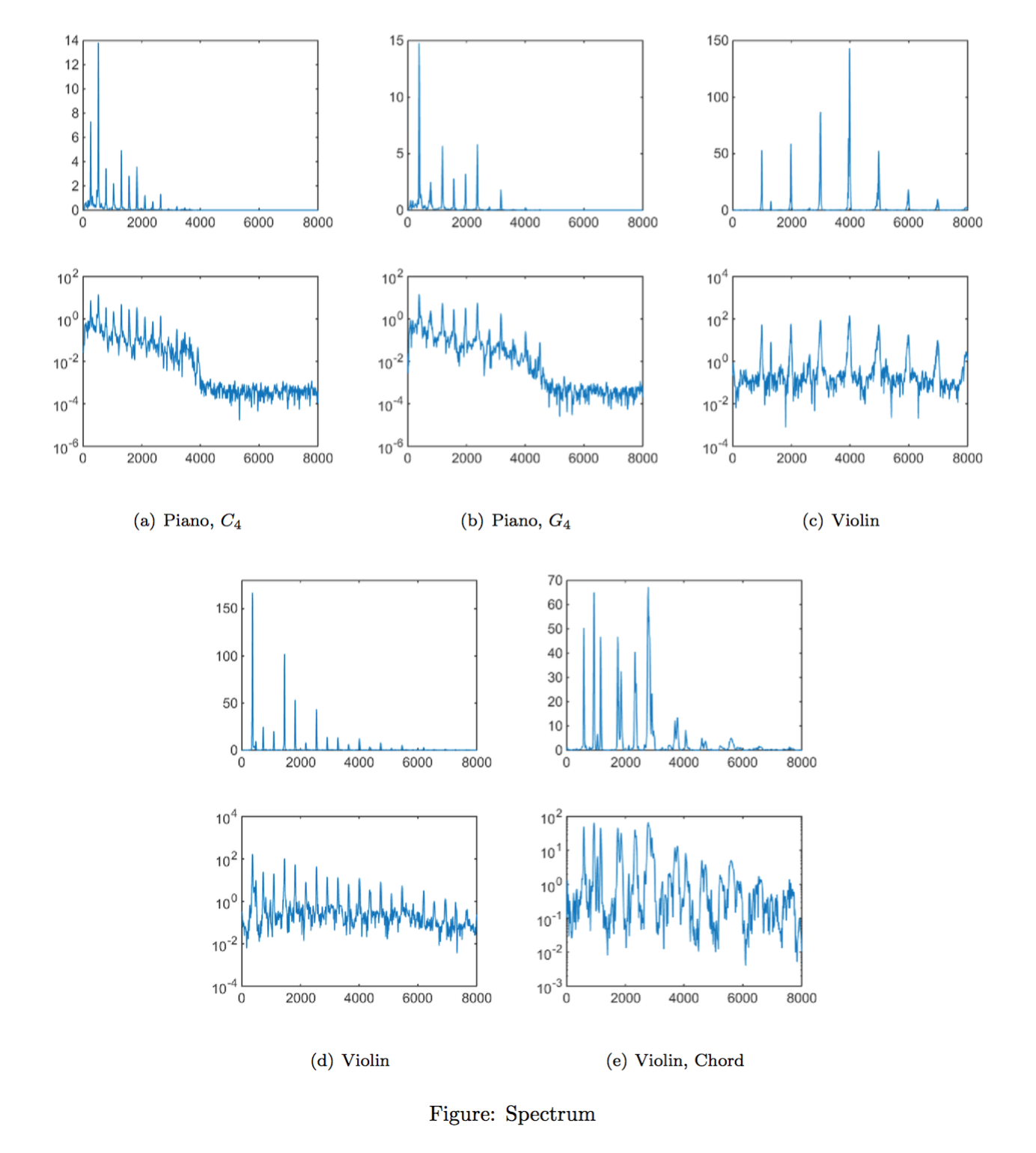
For classification: Not decided yet.

More models/methods may be added.

**Progress:**

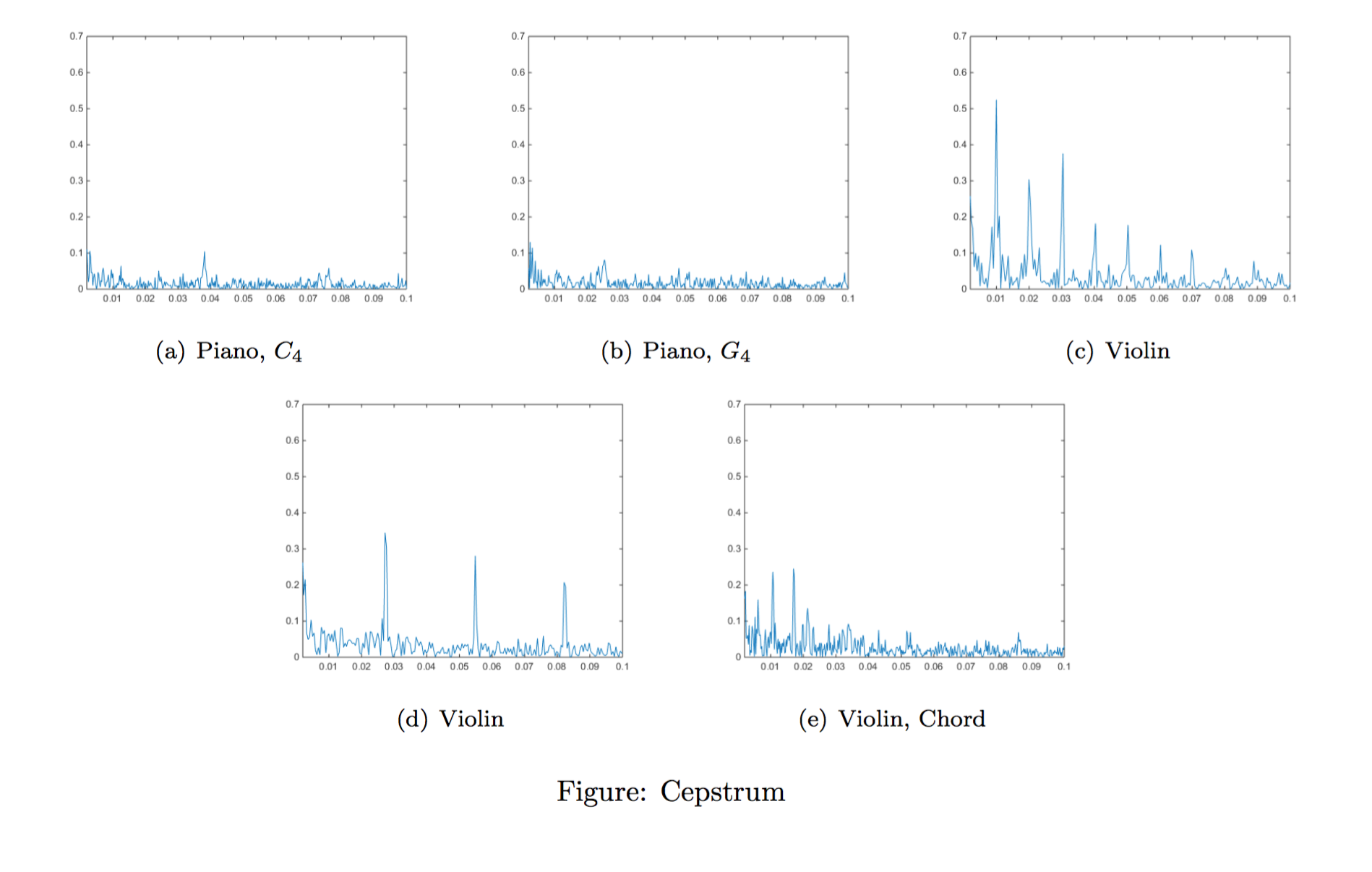
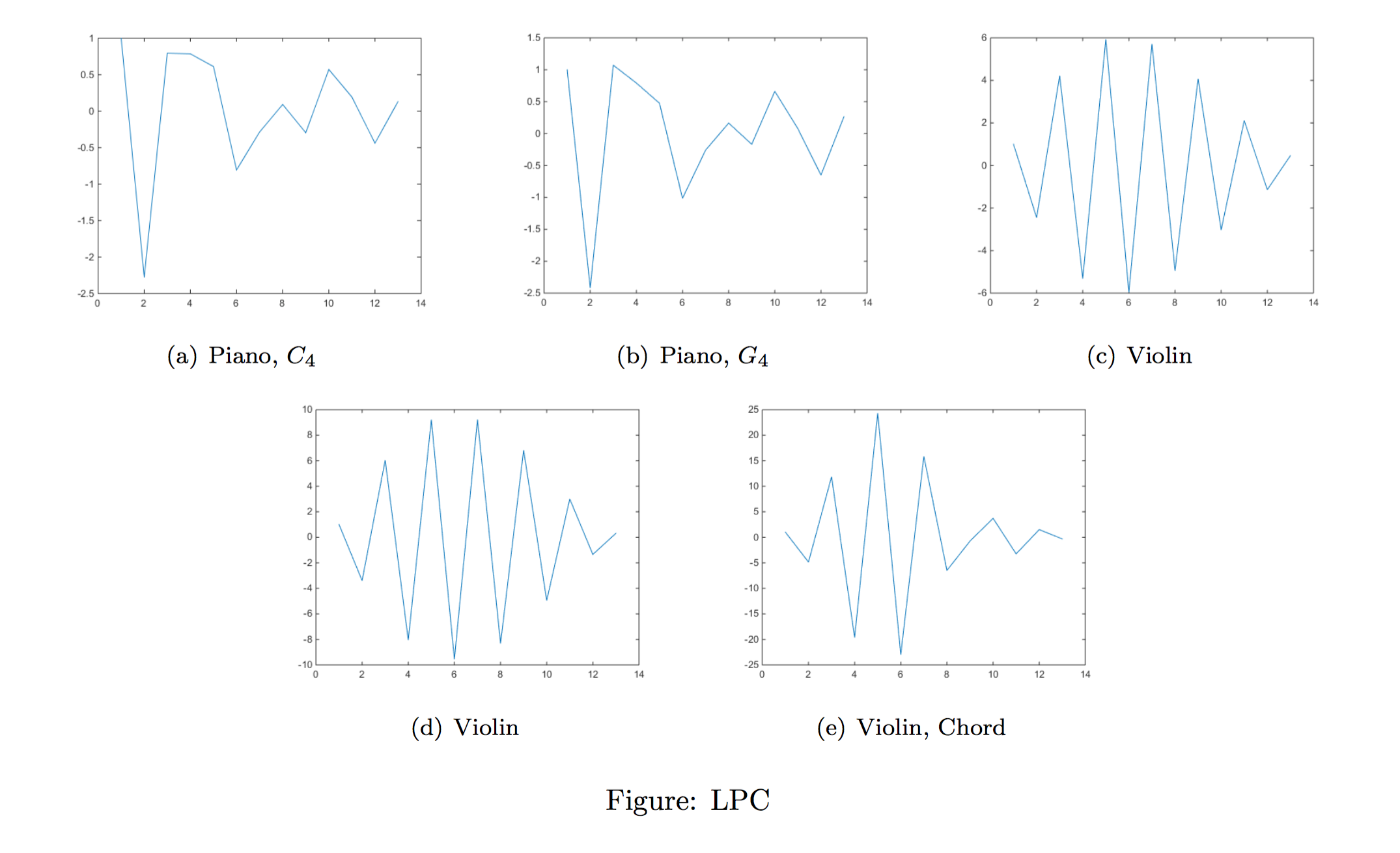
In preliminary experiments, we adopted different means to analyze features of two representative instruments (piano and violin) and five samples, and visually observe the difference for a perceptual impression of typical sound features. The five samples are: a) piano note C4, b) piano note G4, c) a violin note extracted from music, d) a lower violin note extracted from music, e) a violin chord extracted form music.

Firstly we used STFT and attempted to find the pattern of spectrums in both linear and logarithmic scales. In linear scale formants are more obvious to see, but more spectral features can be observed in the logarithmic scale. However, it is non-trivial for the computer. In addition, the spectrum pattern of the same instrument varies with pitch, and becomes confusing if chords are included. Therefore we need further study with other methods.

Then we tried the cepstrum, which performs another FT (precisely IFT) on the log-scale spectrum. Immediately we noticed the peaks much sharper for violin, but no additional information can be obviously seen, so cepstrum may not be clear enough if more kinds of instruments are included in the candidate set.

The LP Coefficient method works well on the samples, and it appears clearer and more robust because the shape is not distorted much in different pitches (although the magnitude varies). Even when chords are taken into consideration, it still roughly keeps its shape. Therefore, this model is probably more promising by now.

The Mel-Frequency Cepstrum Coefficient is widely used in processing human speeches, and many papers also use MFCC to deal with music. However, MFCC doesn’t well represent the features of our instruments, especially the violin, which is shown in the figure. Probably it is because that the violin audios we have are of worse quality, or because that the fundamental frequencies of the violin are so different that it affect the MFCC.



Later on we will try the delta of MFCC and the delta-delta to discover whether the changing rate of an audio among a single note will help us discover the type of the instrument.

**Schedule:**

5.1-5.14 Read more papers to see if there are other features, and try them to see whether they are feasible and effective.

5.15-5.22 Learn about clustering and classification methods.

5.23-6.10 Finish the program, do experiments and make optimizations.

