Music Signal Analysis
And
Music Genre Classification

Anurag

16122003

B.Tech Engineering Physics

Introduction

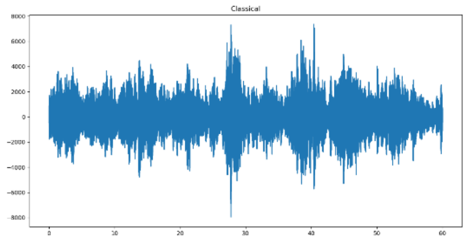
Analyzing music signal is in the core of Music app-based industry, which requires proper knowledge of signal & systems and core mathematics and programming knowledge. A large section of machine learning/deep learning community is working on music-related problems, and are making consumers life easier by classifying, tagging and recommending music to listeners without any human support.

But all of this requires proper mathematical background and knowledge of signal and systems.

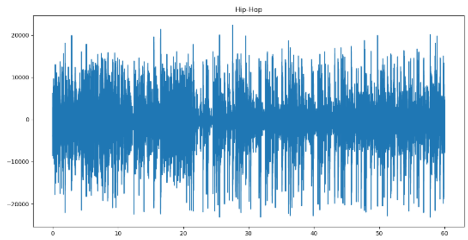
In the first section we will see how to analyze music signal based on i) FFT ii) Spectrogram ii) STFT iii) Q-transform iv) MFCC (feature of MIR community) and in the second section I am going to apply some machine-learning classification algorithms to classify genre of particular music. On the basis of first section we will choose our feature extraction method and we will train our model on that features which will predict future music's genre. In the sections I will use different feature functions to extract feature and train our classifier with it.

Music Signal Analysis

For simplicity, I am taking only two type signals to analyze (out of 10 class), Classical and Hip-Hop Genre of Music.



1) Classical Signal



2) Hip-Hop Signal

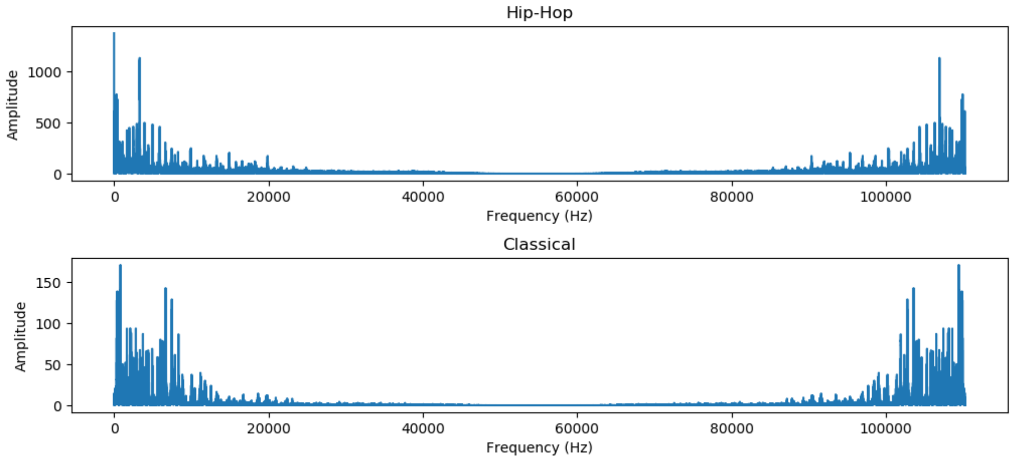
As, we can see the wave forms of different genres of music which are way different from each other, but it does not convey any useful information, we can only learn about the amplitude change with respect to time and it's also not beneficial to analyze full song in one go.

Transforms

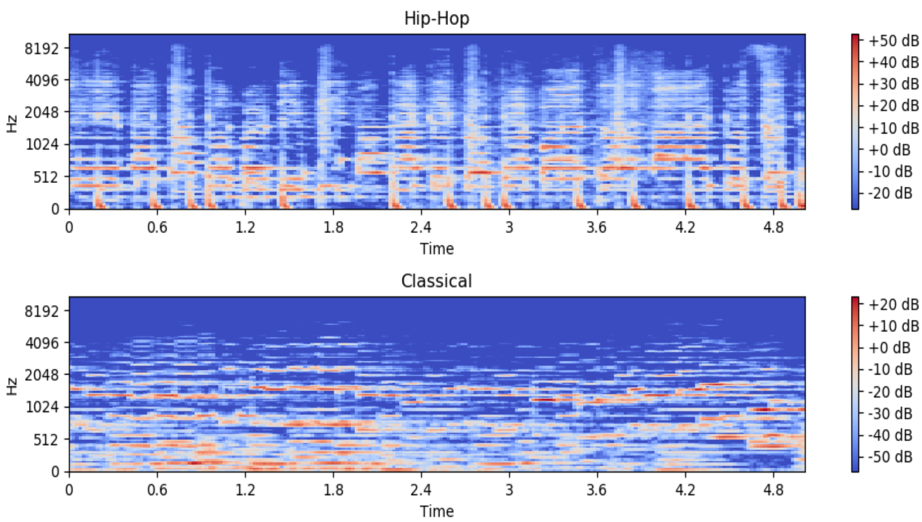
Simply plotting the wave form does not give any useful information, let's try Fourier Transform, which is most fundamental operations in applied mathematics and signal processing. It transforms our time-domain signal into frequency domain. Whereas the time domain expresses our signal as a sequence of samples, the frequency domain expresses our signal as a superposition of sinusoids of varying magnitudes, frequencies and phase offsets.

With assuming the reader is of mathematical background and has proper knowledge of Fourier Transform, I am skipping the mathematical part.

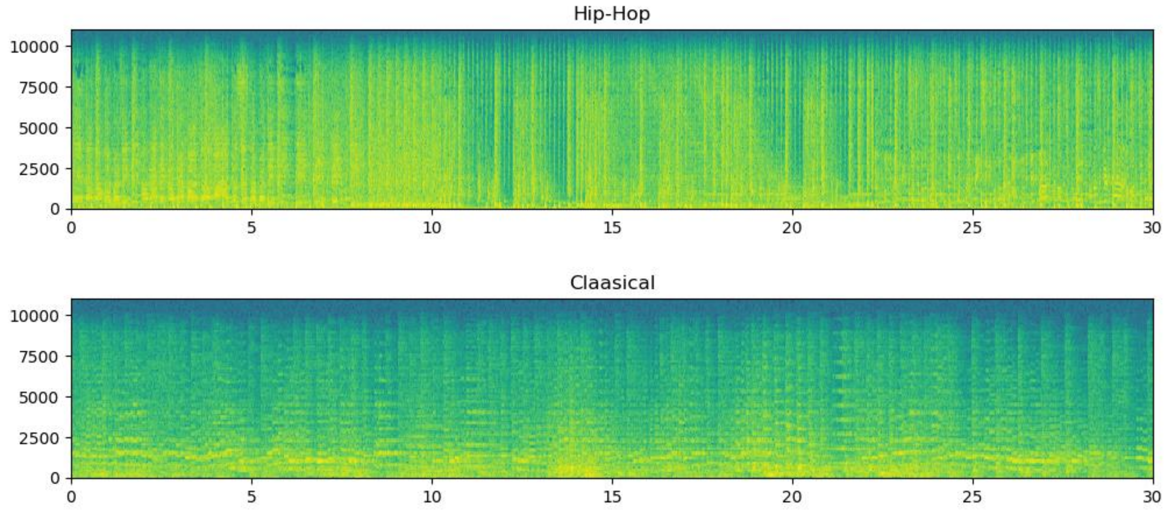
Fourier Transform of the songs:



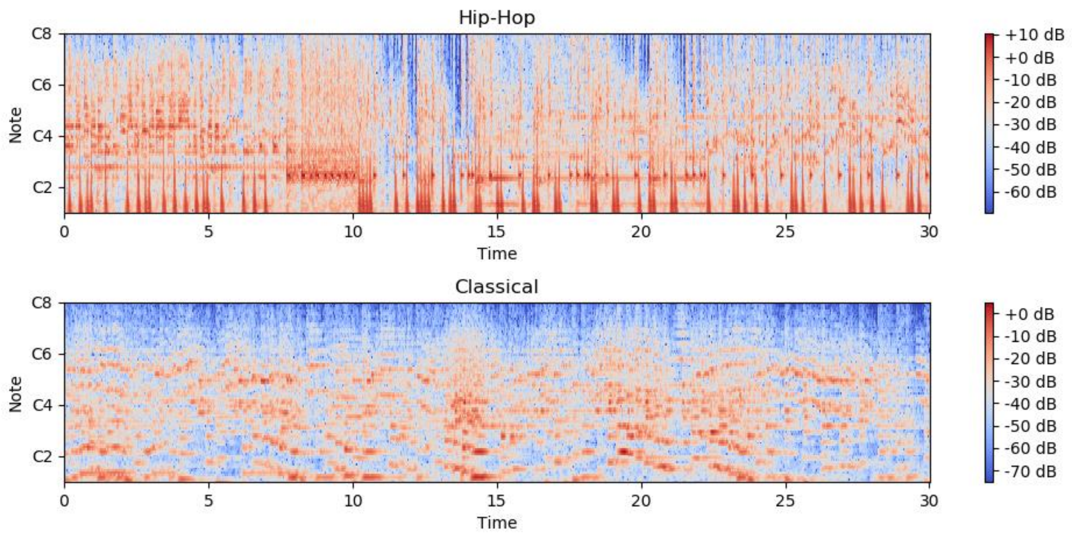
We have the frequency information of signals, which is much easier to analyze and has been used to make Classifier in next section, but we lost the information about time. There are many other challenges also, like, Musical signals are highly non-stationary, their statistics change over time. It would be rather meaningless to compute spectrum over an entire 10-minute song. The solution to this is Short-Time Fourier Transform (STFT). With STFT, we can break up the signal into discrete windows, each signal within a window is stationary signal. After applying FFT over each window we obtain Spectrogram of the signal.



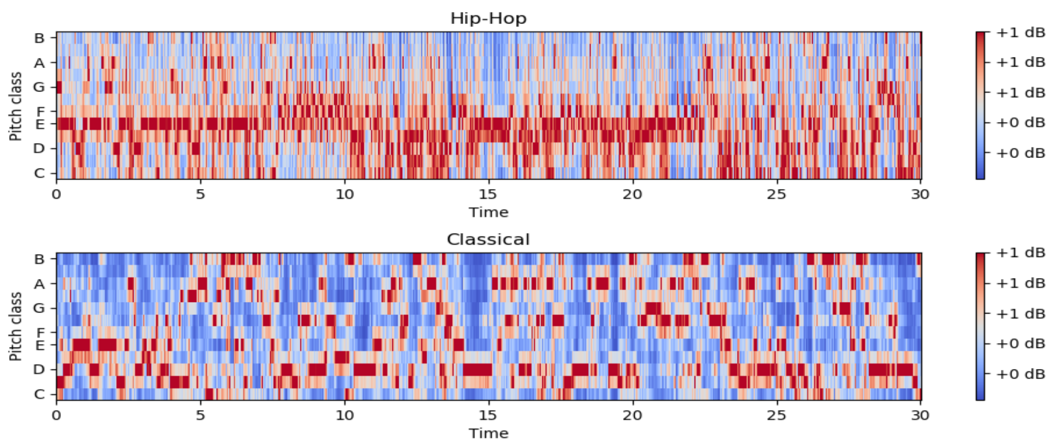
Now, lets see Spectrogram of music files, spectrogram is nothing but visual representation of frequency varying with time,



Q-transform, it is used to obtain information about the pitch values and it gives Chromagram for each song.



Constant Q-transform and Chromagram with pitch classes



As, from above signal analysis from different perspective, we can ensure that classification of such signals is possible and can be done with the use of machine learning algorithms, using array of features extracted from above transforms.

Music Genre Classification

We have created some kind of musical fingerprint of song using basic FFT. If we do this for a couple of songs, and manually assign their corresponding genres as labels, we have the training data that we can feed into our first classifier.

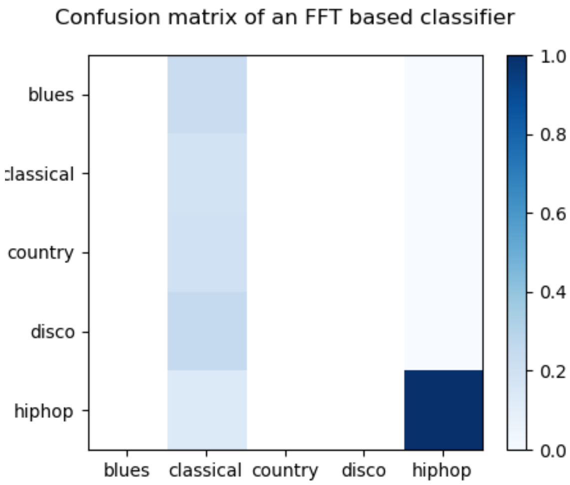
Although we have word “fast” in FFT, it is much slower than the creation of text-based features, we might have to think other way to speed up the whole feature-creation process.

We will cache the FFT, and read the cached FFT representation instead of wave file, fixing the no. of FFT component to first 1000, with our current knowledge we do not know this are the most important ones with regard to music genre classification, later on reader can vary the no. of components (less/more according to his convenience)

Let us use SVM to benchmark the Classification problem with FFT, the difficulty we are facing is that, we have to face multiclass classification and not conventional Binary Classification.

With multiclass problems, we should also not limit our interest to how well we manage to correctly classify the genres. In addition, we should also look at which genres we actually confuse with each other, which can be done by confusion matrix. It prints the distribution of labels that classifier predicated for the test set for every genre.

We have used the train/test split setup with the ratio 0.7/0.3 of total set.



I)SVM

For a perfect classifier, we would have expected a diagonal of dark squares from the left-upper corner to right-lower one, and light colors for the remaining area. In the above diagram we can see our classifier is 1000 miles away from being perfect.

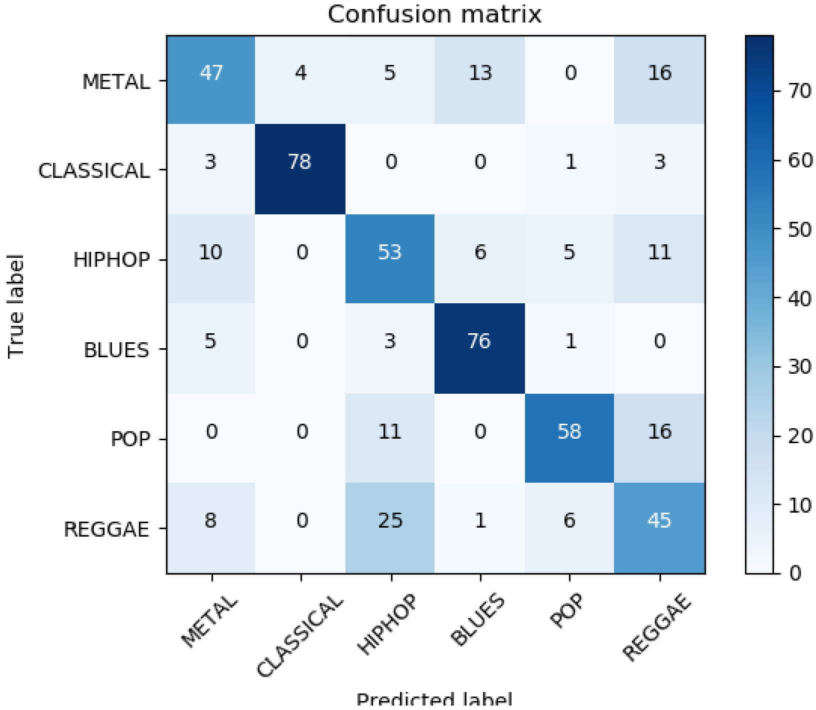
Obviously, using FFT points to the right direction, but it is not enough to get a decent classifier. Surely, we can play with the parameter tuning, we should more research. In which we find that FFT is indeed not a bad feature for genre classification – it is just not refined enough.

Going Beyond Fourier Transform

After much of literature survey of ISMIR and AMGC, I came across MFCC- Mel Frequency Cepstral Coefficients, which is nothing but the power spectrum of sound we saw in section I. It is calculated as the Fourier Transform of the logarithm of the signal's spectrum.

I am retrieving the Ceps, MFCC coefficient (first 20) and feeding it to our classifier. Which will then give us it's confusion matrix.

Similar to my work in FFT, I am using cache the once generated MFCC features and read them instead of recreating them each time I train my classifier.



SVM Classifer using Genre Classification

The classification of all the genres has been improved, confusion matrix looking much better than the previous one. We can see the diagonal element able to classify all the genres properly.

If we would want to improve on this, we can work on the non-white spot on the non-diagonal places. To fix this, we would probably need to dive deeper into songs and extract things, guitar pattern and similar genre-specific characteristics. Also, while going through ISMIR paper, I got to know about Auditory Filter bank Temporal Envelope features, which seem to outperform MFCC features in certain situations.

Future Developments

1)Using AFTE features, in search of perfect classifier.

Tools Used

1) Python 3.x
2) Scikit-learn
3) librosa
4) Matplotlib
5) Numpy
6) Scipy

All graphs, result and confusion matrix are obtained from python codes, and is reproducible, there efficiency depends on dataset used, in this case GZTAN (1.2 GB of audio files) data set was used.

References

1) Feature Selection in Automatic Music Genre Classification. C.N.Silla; A.L.Koerich; C.A.A.Kaestner
2) Automatic Music Genre Classification using Machine Learning. M.A.A.Zain ; Ahmed Siddiqui
3) Haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html

For code

https://github.com/rag9704/Music_Genre_Classification