An Independent Study on-

Music and Machine Learning (Report)

Guide: Prof. Sergey Bereg
Course: CS 6V81
Semester: Fall 2019

Submitted by –
Name: **Anant Vaibhav**
NetID: **axv180022**

Erik Jonsson School of Engineering and Computer Science
University of Texas at Dallas

Contents

Section 1

Music Theory

1.1 Introduction

Music is an omnipresent and indispensable piece of the lives of billions of individuals around the world. Melodic manifestations and exhibitions are among the most mind boggling and perplexing of our social curios, and the enthusiastic intensity of music can contact us in astonishing also, significant ways. Music traverses a gigantic scope of structures and styles, from straightforward, unaccompanied society tunes, to prominent and jazz music, to orchestras for full symphonies. The computerized upset in music dissemination and capacity has all the while filled enormous enthusiasm for and thoughtfulness regarding the ways that data innovation can be applied to this sort of substance. From perusing individual assortments, to finding new craftsmen, to overseeing and ensuring the privileges of music makers, PCs are currently profoundly associated with pretty much every part of music utilization, which isn't even to specify their essential job in quite a bit of the present music creation.

Regardless of the significance of music, music preparing is as yet a moderately youthful discipline contrasted and discourse handling, an exploration field with a long custom. As a matter of fact, a bigger research network spoke to by the International Society for Music Information Retrieval (ISMIR), which efficiently manages a wide range of PC based music examination, handling, and recovery themes, was framed in the year 2000. Customarily, PC based music explore has for the most part been led based on emblematic portrayals utilizing music documentation or MIDI portrayals. As a result of the expanding accessibility of digitized sound material and a blast of figuring power, computerized handling of waveform-based sound signals is presently progressively in the focal point of research endeavors.

A considerable lot of these exploration endeavors are coordinated towards the improvement of advancements that enable users to get to and investigate music in the entirety of its various features. For model, sound fingerprinting systems are these days incorporated into business items that help clients to compose their private music assortments. Music handling strategies are utilized in expanded sound players that feature the present measures inside sheet music while playing back an account of an orchestra. On request, extra data about melodic and symphonious movements or musicality and beat is consequently introduced to the audience. Intelligent music interfaces show auxiliary portions of the present bit of music and enable clients to straightforwardly hop to any key part, for example, the melody segment, the principle melodic subject, or a solo segment without repetitive quick sending and rewinding. Besides, audience members are outfitted with Google-like web indexes that empower them to investigate enormous music assortments in different ways. For instance, the client may make an inquiry by indicating a specific note group of stars, or some symphonious or cadenced example by whistling a song or tapping a musicality, or just by choosing a short section from a CD recording; the framework at that point gives the client a positioned rundown of accessible music passages from the assortment that are musically identified with the inquiry. In music handling, one primary objective is to contribute ideas, models, calculations, usage, and assessments for handling such kinds of examination and recovery issues.

For the purpose of this study, I have gone through few research papers and books which discuss about the Music Information Retrieval and how it is used in creating some really cool applications. We use the data extracted from the signal of the audio and then process it in such a way that they become relevant for the application. For example, Performing FFT of the Audio signal to get to convert the timed signal into Frequency to gain insights about the pitch of the song. There are many other features which can be extracted from the audio signal which can later be used to apply machine learning for clustering similar type of songs together. We can also generate a spectrogram of a song which converts time-amplitude domain into Frequency – Time domain to understand how the pitch varies with respect to time.

We will be going over following topics:

1) Music representation
2) Fourier analysis
3) Features of Music

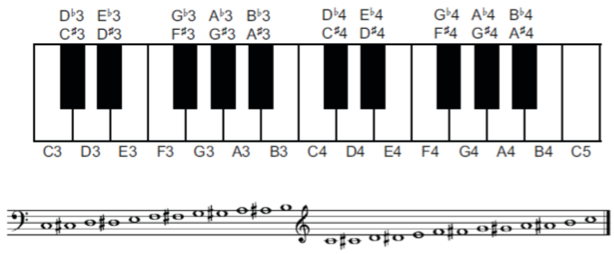
1.2. Music Representation

Music can be spoken to from various perspectives and arrangements. For instance, a writer may record a sythesis as a melodic score. In a score, melodic images are utilized to outwardly encode notes and how these notes are to be played by an artist. The printed type of a melodic score is additionally alluded to as sheet music. The first vehicle of this portrayal is paper, in spite of the fact that it is presently likewise open on PC screens through computerized pictures. For electronic instruments furthermore, PCs, music might be conveyed by methods for standard conventions for example, the generally utilized Musical Instrument Digital Interface (MIDI) convention, where occasion messages determine pitches, speeds, and different parameters to produce the proposed sounds. In this book, we utilize the term representative to allude to any machine readable information design that unequivocally speaks to melodic substances. These melodic substances may go from coordinated note occasions, similar to the instance of MIDI documents, to graphical shapes with connected melodic importance, just like the instance of music etching frameworks. In contrast to emblematic portrayals, sound portrayals, for example, WAV or MP3 documents don't unequivocally indicate melodic occasions. These documents encode acoustic waves, which are produced when a source (e.g., an instrument) makes a sound that movements to the human ear as pneumatic stress motions.

1.2.1 Sheet Music Representation

Sheet music, likewise alluded to as melodic score, gives a visual portrayal of what we normally allude to—specifically for Western old style music—as the "bit of music." Sheet music portrays a melodic work utilizing a conventional language in light of melodic images and letters, which are delineated in a graphical–printed structure. Perusing sheet music, a performer can make an exhibition by following the given guidelines. Playing out a piece from sheet music, be that as it may, not just requires an extraordinary type of education, i.e., the capacity to comprehend the music documentation, yet in addition includes an inventive procedure. A melodic score is once in a while played precisely. Artists may shape the progression of the music by fluctuating the rhythm, elements, and enunciation, in this way bringing about an individual understanding of the given melodic score. In this sense, as opposed to giving unbending determinations, sheet music can be considered as a manage for playing out a bit of music leaving space for various elucidations.

1.2.1.1. Musical Notes and Pitches

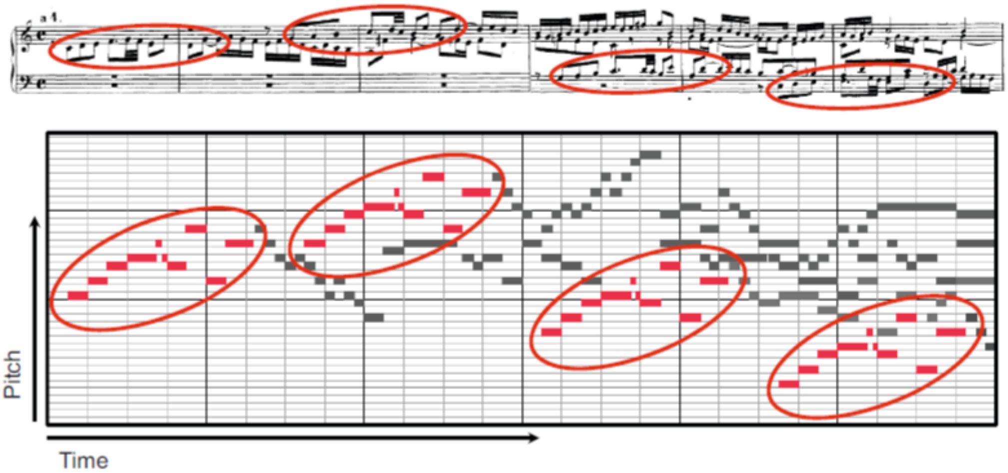


1.2.1.2. Western Music Notation

1.2.2. Symbolic Representations

Emblematic portrayals depict music by methods for substances that have an unequivocal melodic importance and, given in a few computerized position, can be parsed by a PC. Any sort of computerized information arrangement might be viewed as "representative" since it depends on a limited letters in order of letters or images. For instance, the pixels in an advanced picture record or the examples in a computerized sound document might be viewed as images or essential elements. Be that as it may, thinking about these elements independently, no melodic significance can be induced. In this way, neither filtered pictures nor digitized music chronicles are viewed as being emblematic music designs. So also, graphical shapes in vector designs portrayals are not considered to be melodic substances as long as no musically important detail of the shapes is given. All things considered, there is a wide scope of what might be considered as emblematic music. In this segment, we examine a few models including piano-move portrayals, MIDI portrayals, and other representative arrangements that encode sheet music. Moreover, we address optical music acknowledgment (OMR), which is the way toward changing over computerized sweeps of printed sheet music into emblematic portrayals.

1.2.2.1. Piano Roll Representation



1.2.2.2. MIDI Representation

The next symbolic representation we want to discuss is based on the MIDI standard, which stands for Musical Instrument Digital Interface

(a)

(b)

Time (Ticks)	Message	Channel	Note Number	Velocity
60	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	63	100
0	NOTE ON	2	51	100
0	NOTE ON	2	39	100
240	NOTE OFF	1	63	0
0	NOTE OFF	2	51	0
0	NOTE OFF	2	39	0

(c)

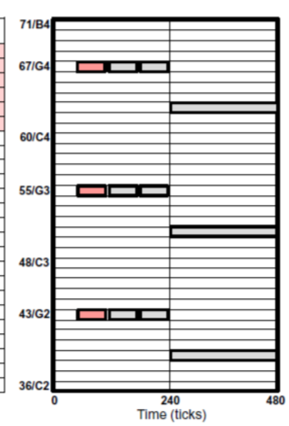


Fig. 1.13 Various symbolic music representations of the first twelve notes of Beethoven's Fifth. **(a)** Sheet music representation. **(b)** MIDI representation (in a simplified, tabular form). **(c)** Piano-roll representation.

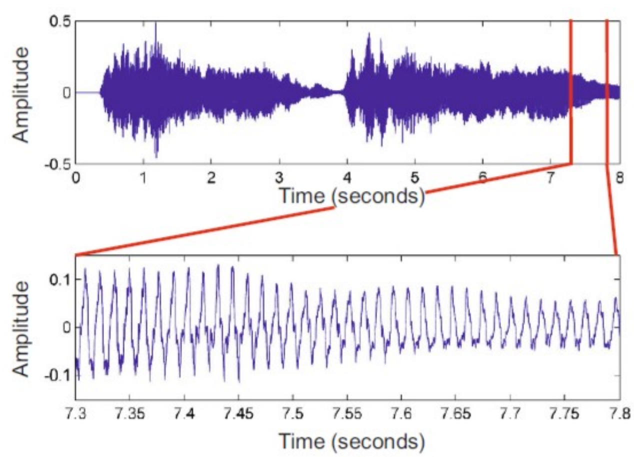
1.2.2.3. Score Representation

Fig. 1.14 Different sheet music representations corresponding to the same score representation of the beginning of Prelude BWV 846 (C major) by Johann Sebastian Bach. From top left to bottom right, a computer-generated, a handwritten, and two traditionally engraved representations are shown.

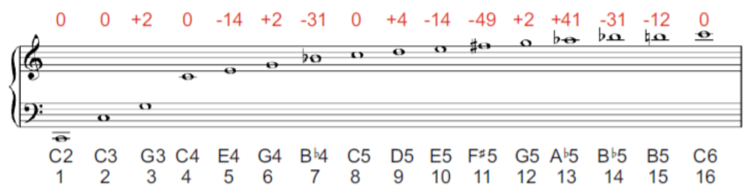
1.2.3. Audio Representation

Music is significantly more than an emblematic portrayal of the notes to be played. Music is tied in with making, making, and molding sounds. At the point when artists start diving into the music, the playing guidelines retreat away from plain sight. The melodic meter transforms into a cadenced stream, the distinctive note objects liquefy into consonant sounds and smooth song lines, and the instruments speak with one another. Artists get sincerely associated with their music and respond to it by ceaselessly adjusting beat, elements, and verbalization. Rather than playing precisely, they accelerate at a few and delayed down at others so as to shape a bit of music. Correspondingly, they constantly change the sound power and stress certain notes. The entirety of this brings about a one of a kind execution or an elucidation of the bit of music.

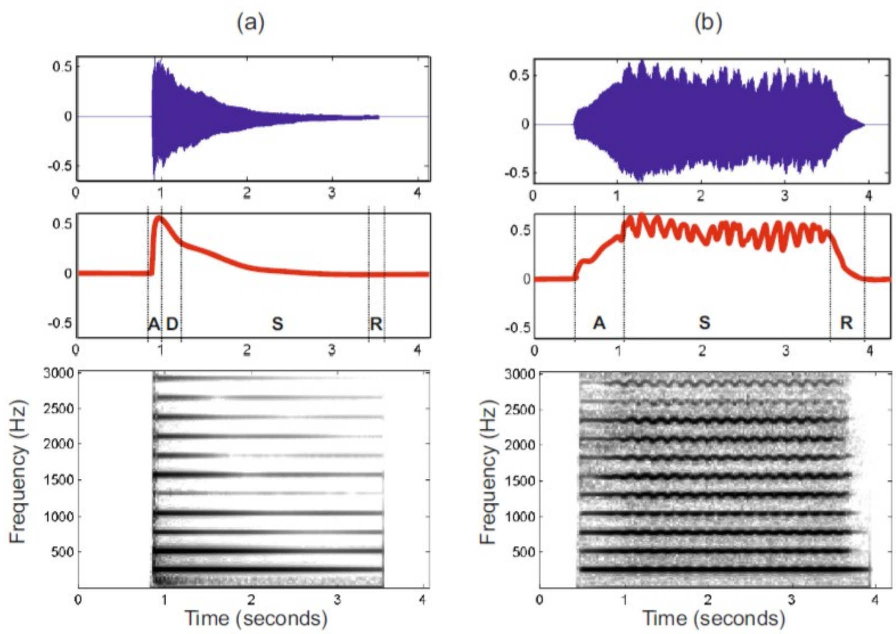
1.2.3.1. Waves and Waverforms



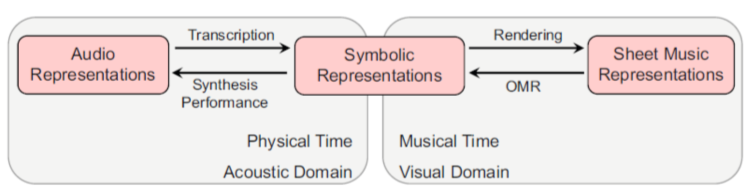
1.2.3.2. Frequency and Pitch



1.2.3.3. Timbre



All these 3 techniques of Music Representation can be transformed to one another. Befoew figure represents that:

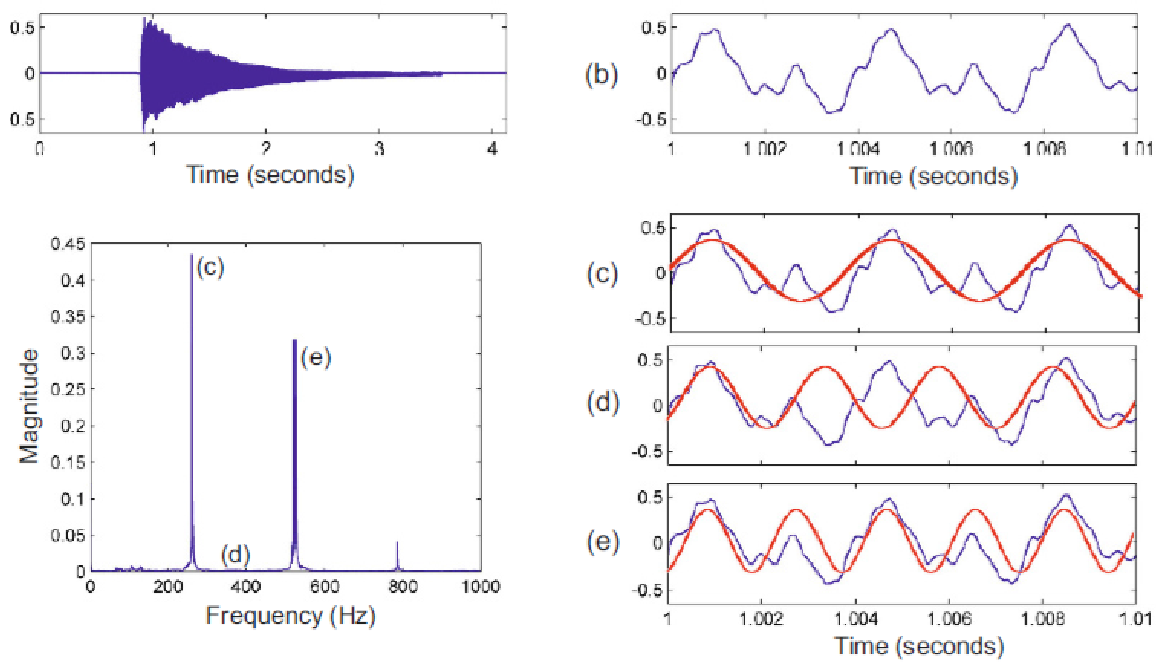


1.3. Fourier analysis of Signals

Music signals are commonly mind boggling sound blends that comprise of a large number of various sound segments. Along these lines intricacy, the extraction of musically applicable data from a waveform comprises a troublesome issue. An initial phase in better understanding a given sign is to disintegrate it into building hinders that are increasingly open for the ensuing handling steps. For the situation that these structure squares comprise of sinusoidal capacities, such a procedure is likewise called Fourier investigation. Sinusoidal capacities are unique in the feeling that they have an unequivocal physical significance as far as recurrence. As an outcome, the subsequent deterioration unfurls the recurrence range of the signal—like a crystal that can be utilized to split light up into its constituent otherworldly hues. The Fourier change changes over a sign that relies upon time into a portrayal that relies upon recurrence. Being one of the most significant devices in signal handling, we will experience the Fourier change in an assortment of music handling undertakings.

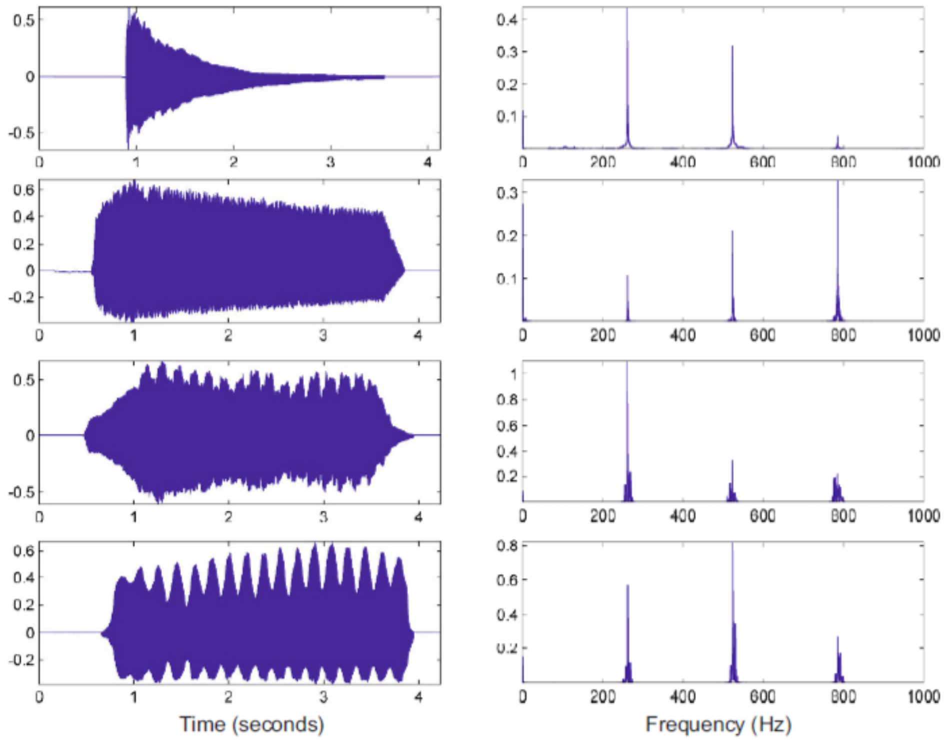
It basically converts the Sound from Time domain to Frequency Domain.

Here are few examples:



Fourier Transformation of Same Frequency but with different instruments:

(a)Piano. (b) Trumpet. (c) Violin. (d) Flute



1.4. Features of Music

There are few really important features associated to and audio signal. Once we can break down audio using a sapling rate, we can extract following features for our purpose:

1. **Zero Cross Rate**
   The zero crossing rate is the rate of sign-changes along a signal, i.e., the rate at which the signal changes from positive to negative or back. This feature has been used heavily in both speech recognition and music information retrieval. It usually has higher values for highly percussive sounds like those in metal and rock.
2. **Spectral Centroid**
   It indicates where the ”centre of mass” for a sound is located and is calculated as the weighted mean of the frequencies present in the sound. If the frequencies in music are same throughout then spectral centroid would be around a centre and if there are high frequencies at the end of sound then the centroid would be towards its end.
3. **Spectral Roll off**
   Spectral rolloff is the frequency below which a specified percentage of the total spectral energy, e.g. 85%, lies.
4. **MFCC**
   This feature is one of the most important method to extract a feature of an audio signal and is used majorly whenever working on audio signals. The mel frequency cepstral coefficients (MFCCs) of a signal are a small set of features (usually about 10–20) which concisely describe the overall shape of a spectral envelope.

There are few other features also which can be extracted.

5. Chroma Vector
6. Chroma Deviation
7. Energy
8. Entropy of Energy
9. Spectral Spread
10. Spectral Entropy
11. Spectral Flux

Section 2

MIR in Practice

2. MIR in Practice

As a part of this study, I have created few minor projects to understand the techniques of MIR in a better way.

1) Audio Equalizer
2) Tempo/Beat Tracking
   a. Onset Detection
   b. Tempo Estimation
   c. Beat Tracking
3) Genre Classification (Machine Learning)

The coding has been done in python, and below Libraries have been used:

libROSA,
mirEval,
pyAudio,
pandas,
numpy,
scipy

2.1. Audio Equalizer

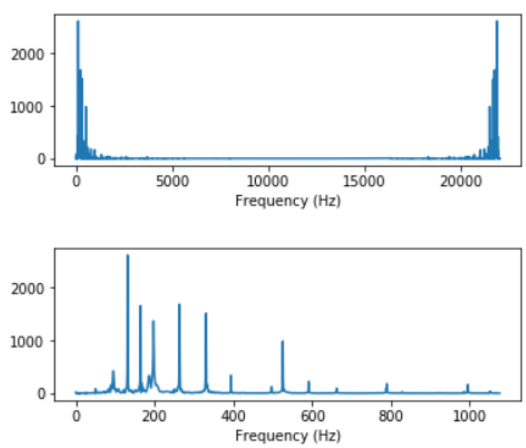
Equalization is the process of adjusting the balance between frequency components within an electronic signal. The most well known use of equalization is in sound recording and reproduction but there are many other applications in electronics and telecommunications. The circuit or equipment used to achieve equalization is called an equalizer.

Basic layout of the project:

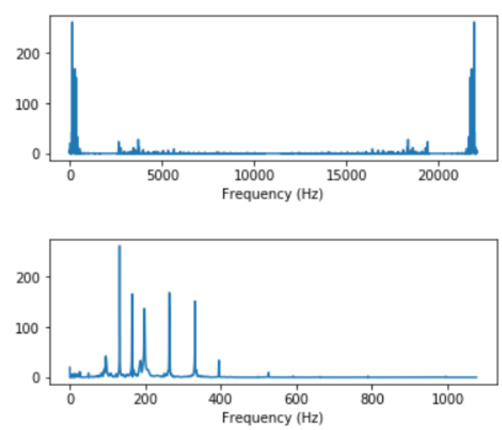
1) Import Audio
2) Convert it to Frequency domain using FFT
3) Modify the Frequency Amplitudes based on Low, Medium, High Frequency bins.
4) Convert signal back to time domain by IFFT
5) Export the Audio

Here are the results.

The freq spectrum based on original music file:

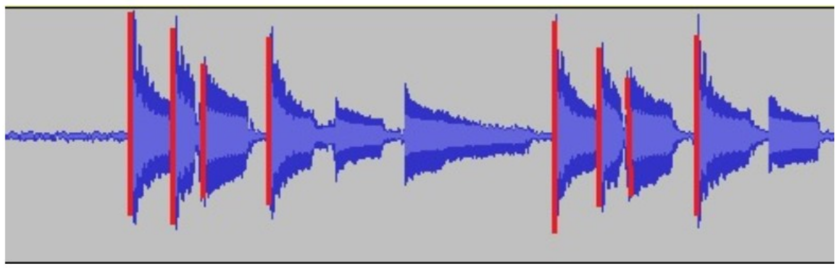


The freq spectrum based on equalized wave:



2.2. Temp/Beat Tracking

In tempo/Beat tracking, we have to use the concepts of Onset Detection, which means, for each ‘attack’ of the envelop of a measure, we want to detect when the onset happened. This becomes the base for tempo estimation.

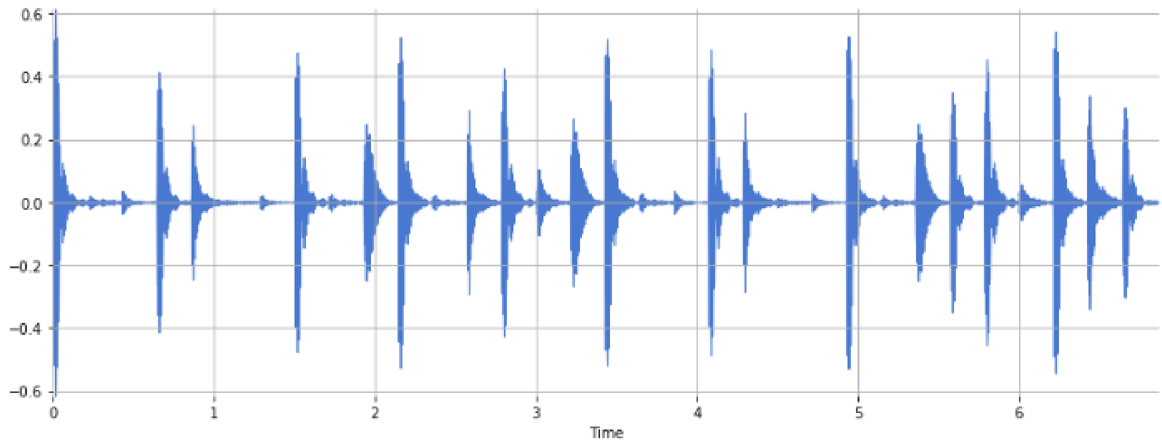


2.2.1. Onset Detection

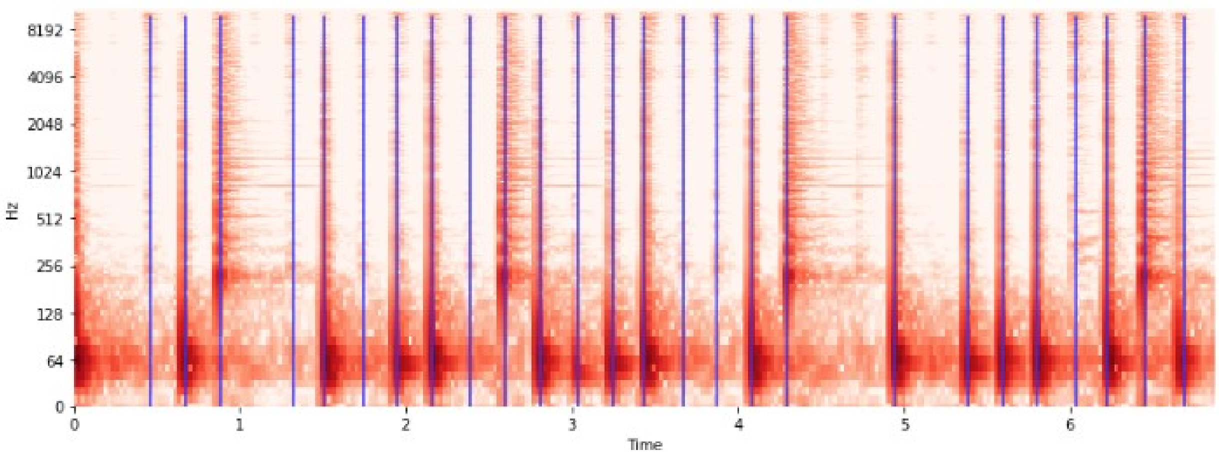
In this exercise, we take a drum beat and try to estimate its onsets. We use the liberary librosa.onset for this purpose.

Here are the results:

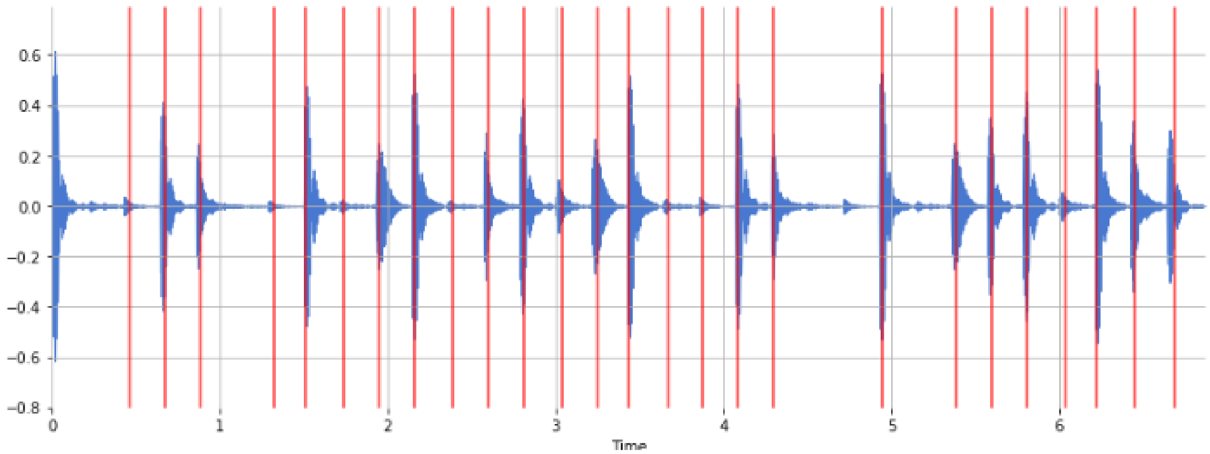
Original sound:



Plotting the onsets on a spectrogram



Plotting the onsets on the original sound wave, to over lap



After this process, we simply add a ‘click’ sound at these times to create another track of clicks which has the information of onsets.

2.2.2. Tempo Estimation

Tempo refers to the speed of a musical piece. More precisely, tempo refers to the rate of the musical beat and is given by the reciprocal of the beat period. Tempo is often defined in units of beats per minute (BPM).

For this experiment, we use a drum beat to estimate tempo, and use librosa library for a global tempo estimation.

Here is the result:

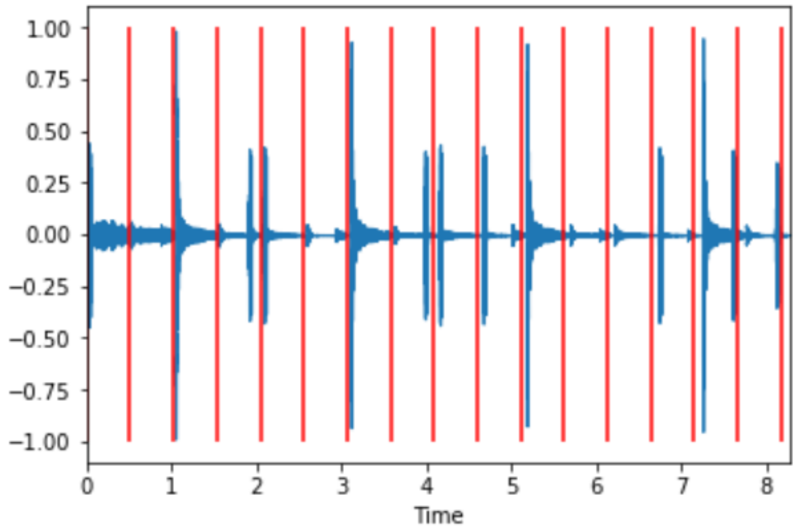
Estimate the tempo:

```
tempo = librosa.beat.tempo(x, sr=sr)
print (tempo)
```

```
[117.45383523]
```

In [17]:
```
librosa.display.waveplot(x)
plt.vlines(beat_times, -1, 1, color='r')
```

Out[17]: <matplotlib.collections.LineCollection at 0x24947716dd8>



2.2.3. Beat Tracking

The **Beat** is the regularly occurring pattern of rhythmic stresses in music. When we count, tap or clap along with music we are experiencing the Beat. Try tapping your finger along with different types of music and see what happens.

Here also, we use a drum beat to extract the beat of the audio signal:

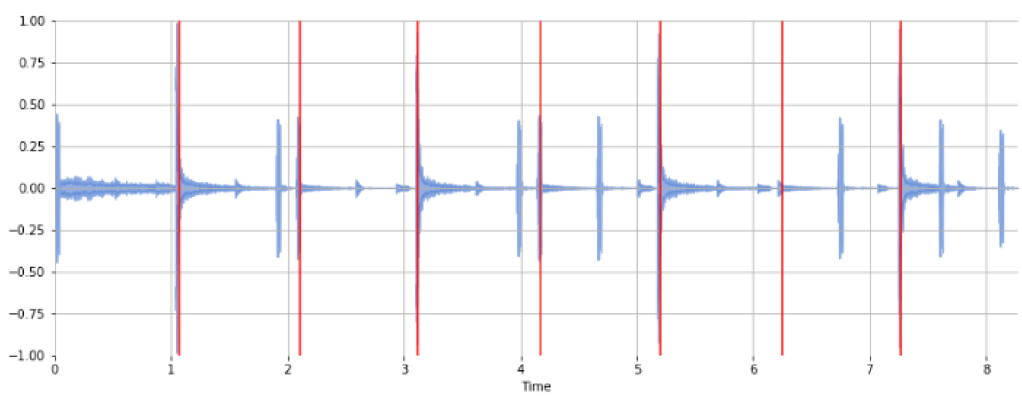
```
tempo, beat_times = librosa.beat.beat_track(x, sr=sr, start_bpm=60, units=
print(tempo)
print(beat_times)
```

```
58.72691761363637
[1.06811791 2.11301587 3.11147392 4.17959184 5.20126984 6.2461678
 7.2678458 ]
```

Plot the beat locations over the waveform:

In [5]:
```
plt.figure(figsize=(14, 5))
librosa.display.waveplot(x, alpha=0.6)
plt.vlines(beat_times, -1, 1, color='r')
plt.ylim(-1, 1)
```

Out[5]: (-1, 1)



2.3. Genre Recognition (Machine Learning)

A music genre is a conventional category that identifies some pieces of music as belonging to a shared tradition or set of conventions.[1] It is to be distinguished from musical form and musical style, although in practice these terms are sometimes used interchangeably.

In this exercise, we will be using few techniques of MIR to identify the Genre.
For simplicity, we are not naming the Genre, we are just considering 2 classes.
By listening to the Audio, we get to know, that 1 song is of '**Classical**' genre, and the other one belongs to '**Hip Hop**' Genre

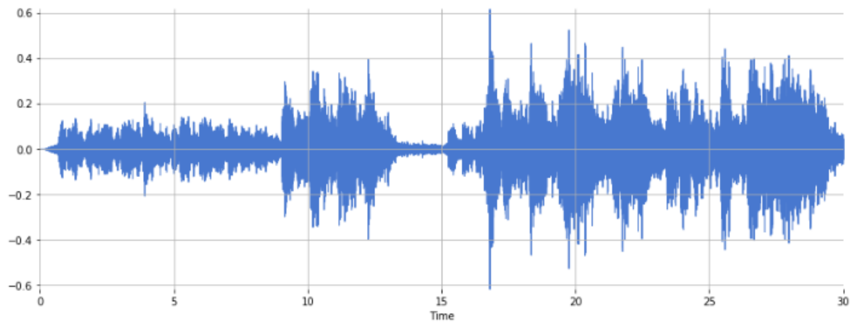
For this exercise, we will be extracting the most significant feature for Music signals – MFCC (Which we discussed previously).

Based on the results of the MFCC components, we create a Support vector machine to classify between these two classes.

1) We load 2 songs (for first 30 second):
    audio/brahms_hungarian_dance_5.mp3
    audio/busta_rhymes_hits_for_days.mp3

```
plt.figure(figsize=(14, 5))
librosa.display.waveplot(x_brahms, sr_brahms)
```

Out[15]: <matplotlib.collections.PolyCollection at 0x19e6e23a3c8>

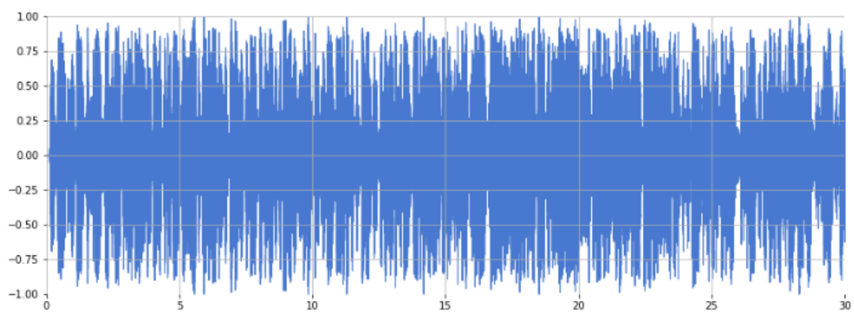


[16]:

Add tag

```
plt.figure(figsize=(14, 5))
librosa.display.waveplot(x_busta, sr_busta)
```

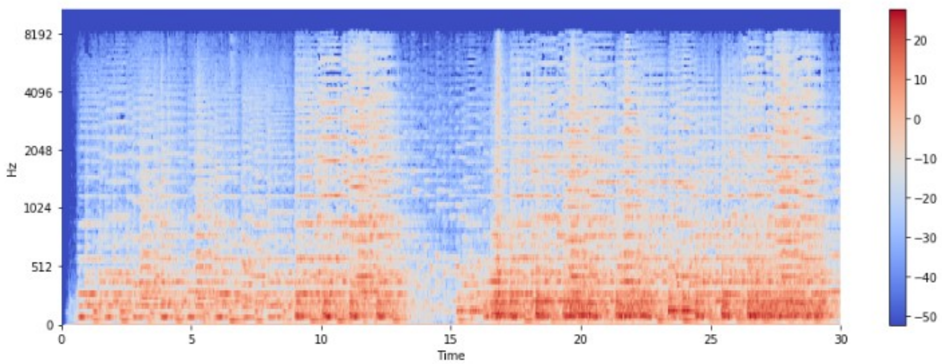
Out[16]: <matplotlib.collections.PolyCollection at 0x19e6e5a85f8>



We plot the Spectrogram for both the songs.

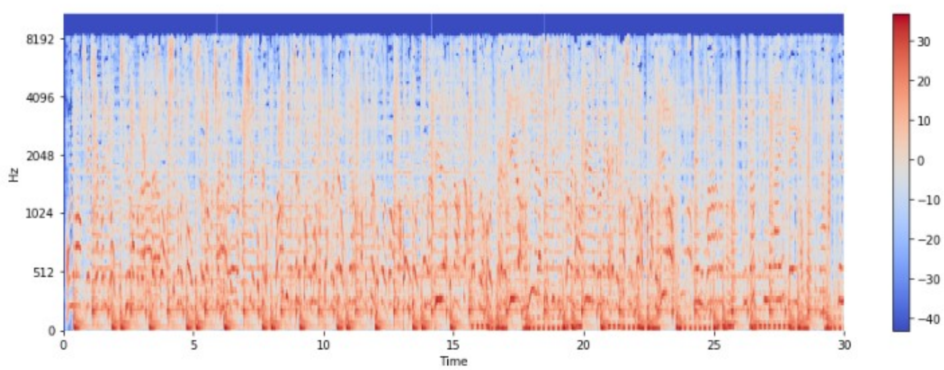
```
plt.figure(figsize=(15, 5))
librosa.display.specshow(Sdb_brahms, sr=sr_brahms, x_axis='time', y_axis='
plt.colorbar()
```

```
: <matplotlib.colorbar.Colorbar at 0x19e6e29cb38>
```

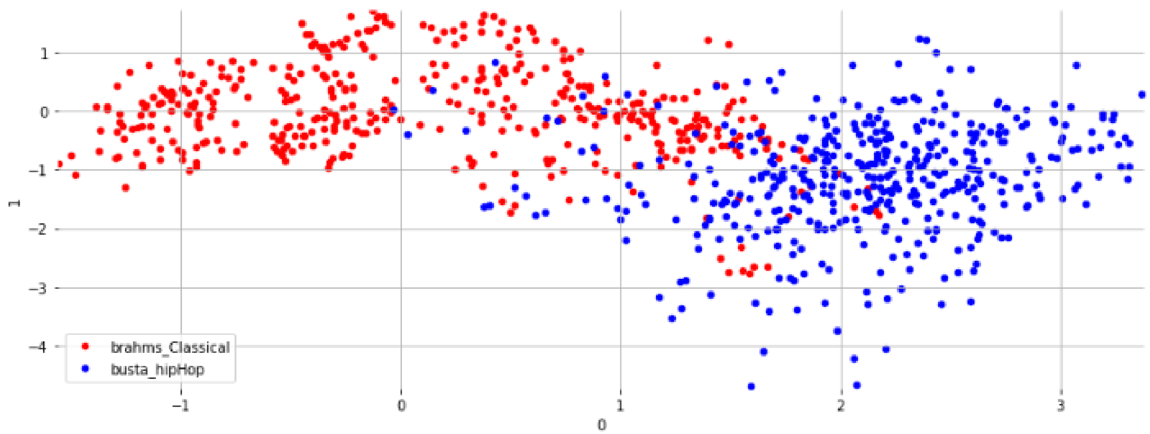



```
plt.figure(figsize=(15, 5))
librosa.display.specshow(Sdb_busta, sr=sr_busta, x_axis='time', y_axis='me
plt.colorbar()
```

```
: <matplotlib.colorbar.Colorbar at 0x19e6e36d240>
```



Just to have a visual indication, we make a scatter plot for 2 components of MFCC for both the songs:



We can see that for these samples, the data points look separable.

Further steps

2) We use libRosa to extract features of MFCC
3) We scale the MFCC results so that they lie between -1 and +1
4) Train the Classifier
5) For prediction, we use the 10 second song snippet with offset of 120 seconds from both the songs.
6) **We get the score of 97% accuracy for the Support Vector Machine for classification**.

References –

1. Fundamentals of Music Processing - Meinard Müller
2. Stanford MIR for Data Sets
3. mir_eval: A TRANSPARENT IMPLEMENTATION OF COMMON MIR METRICS – E. J. Humphery et. al. , ISMIR 2014.