MUSIC GENRE CLASSIFICATION

Diya Sankhla(B18BB011)
Suniket Das(B18EE062)
Shashi Deegwal(B18EE048)
Muskan Bathla(B18CSE036)
Nivedit Jain(B18CSE039)

Introduction

Music for a long time has served as art through which humans express their emotions. Music can lift anyone's mood up, it can act as a source of inspiration for someone or a kind of therapy for someone else.

Professor Gideon Nave, says "When people are given the task of getting to know each other, music is one of the first things that they tend to talk about.".

Genre classification

Automatic music genre classification can assist or replace the human user in this process and would be a valuable addition to music information retrieval systems

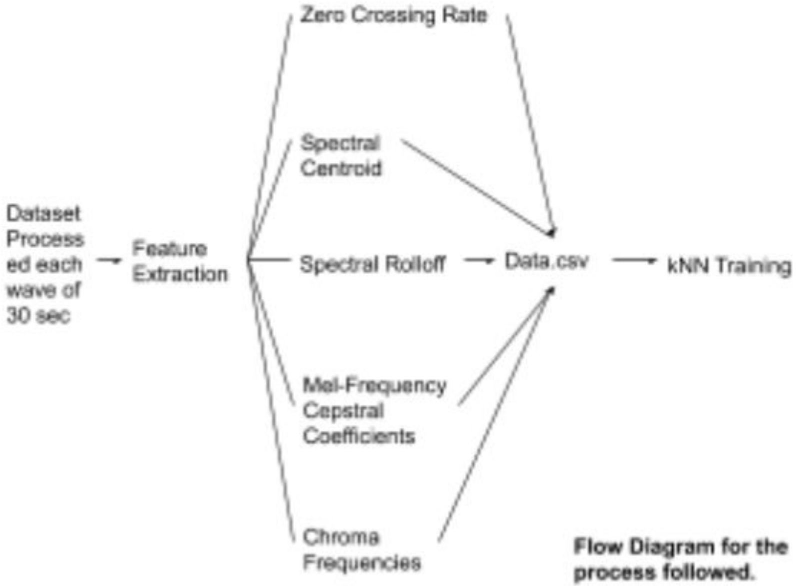
We have classified a given music file according to the following genres

- Blues
- Classical
- Pop
- Jazz
- Metal
- Disco
- Country
- Hip-Hop
- Reggae
- Rock

Dataset

- Used the famous **GTZAN DATASET**
- Dataset consists of 1000 audio tracks- each 30 seconds long
- It contains 10 genres, each represented by 100 tracks
- All track are 22050 Hz Mono 16-bit audio files in .wav format

Methodology



Flow Diagram for the process followed.

Average Zero Crossing Rate

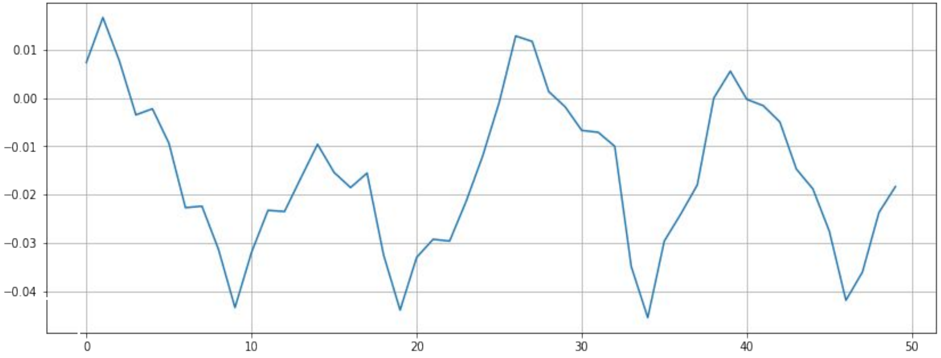
- Zero crossing rate- defined as rate at which signal changes its sign from positive to negative or back

- The Average Zero crossing rate- defined as the average of zero-crossing rates if the signal is divided into various parts of equal length

AVERAGE Zero Crossing Rate

$$\frac{1}{T-1}\sum_{t=1}^{T-1}\mathbb{I}\{s_t s_{t-1} < 0\}$$

Expression for calculating Zero Crossing Rate



The above graph has 5 zero crossings with a zero-crossing rate of 0.102

Spectral Centroid

- Represents the magnitude of the centroid of DTFT of an input signal
- This feature helps us to differentiate between signals of consistent values and with rapidly varying values

Spectral Centroid

$$Centroid = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$$

Formula for Centroid

Spectral Rolloff

- Basically measure the shape of the signal
- Represents the frequency at which high frequencies decline to 0
- Defined as the frequency below which 85% of the magnitude spectrum is located

Spectral Rolloff

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 * \sum_{n=1}^{N} M_t[n]$$

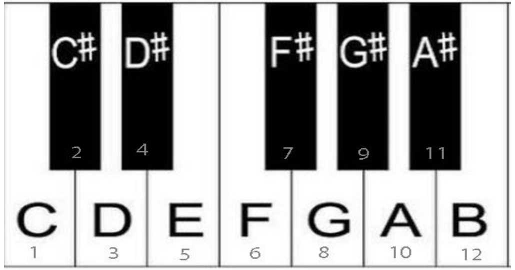
Here, Mt represents magnitude spectrum
And Rt represents the value of Spectral roll-off

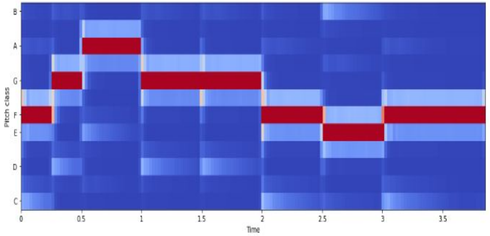
Formula for Spectral Rolloff

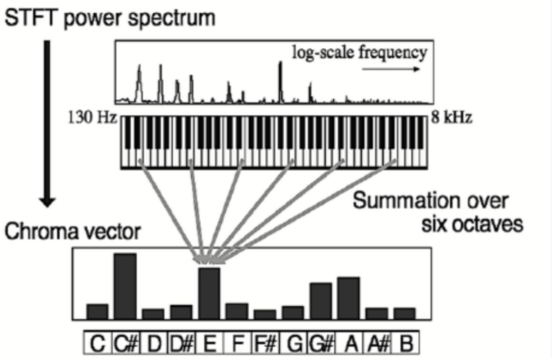
Chroma Frequencies

- Chroma features are closely related to twelve different classes of the pitch
- Chroma features help in identifying pitches that differ by an octave
- Help us gather information about the harmonic and melodic characteristics of music

Chroma Frequencies







12 Chromas and Chromatogram Representation and Algorithm

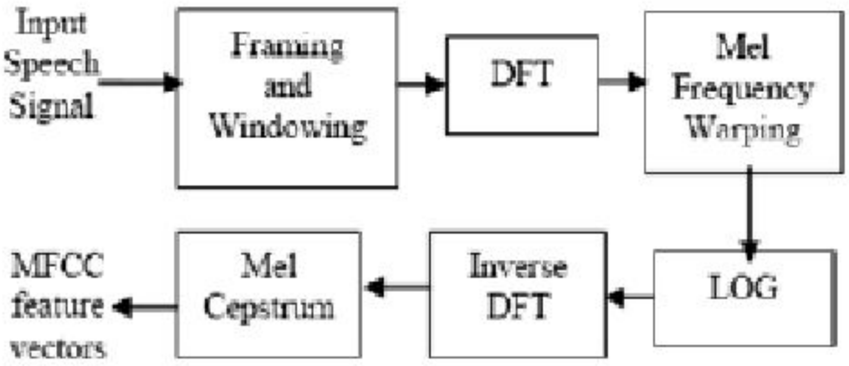
Fundamental Frequencies

Note	FF(Hz)	Note	FF(Hz)
C_0	16.35	$F_0^{\#}$	23.12
$C_0^{\#}$	17.32	G_0	24.50
D_0	18.35	$G_0^{\#}$	25.96
$D_0^{\#}$	19.45	A_0	27.50
E_0	20.60	$A_0^{\#}$	29.14
F_0	21.83	B_0	30.87

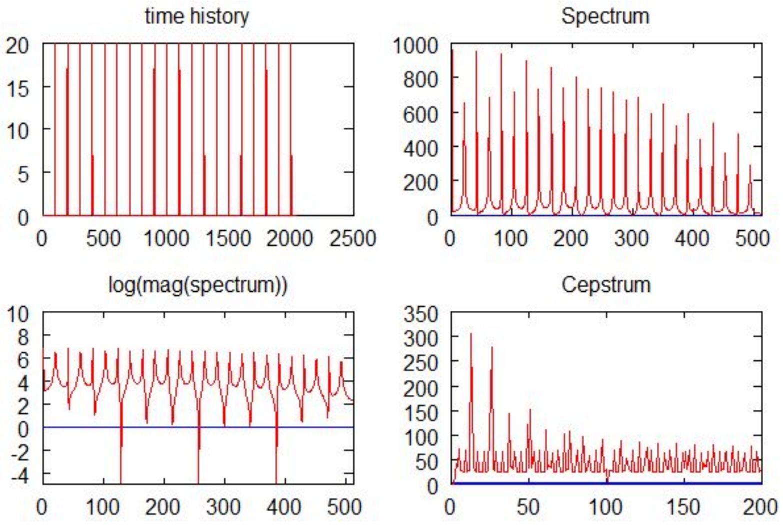
MFCCs

- MFCC stands for Mel frequency cepstral coefficients
- MFCCs of a signal are a small set of features that describes the shape of a spectral envelope
- Used in many state-of-the-art human voice recognition systems

MFCCs



MFCC Feature Extraction

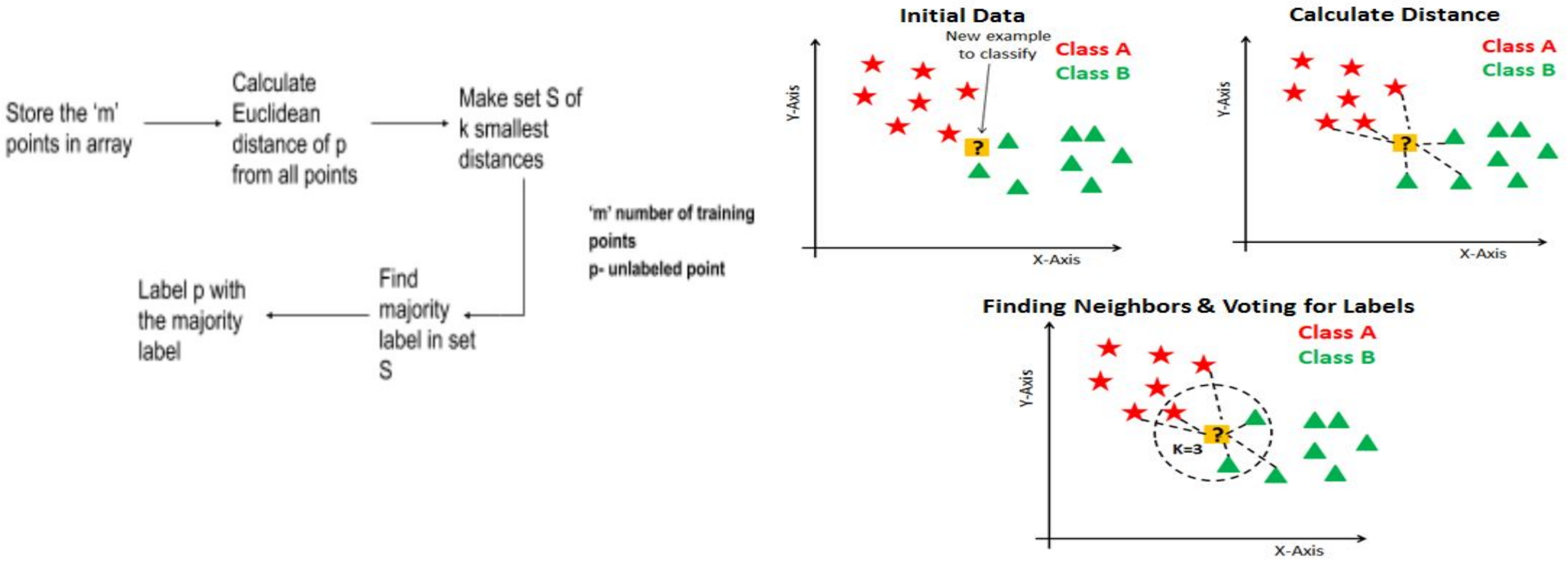


MFCC

Training kNN Model

- k-Nearest-Neighbours (kNN)- a machine learning model for classification
- A supervised machine learning algorithm that accepts some inputs and their corresponding outputs for training so as to classify new data
- Works by finding the ‘k’ closest labeled points (the nearest neighbours) to a given data point and assigning the majority label amongst them to the unlabeled data points into consideration

Training kNN Model



kNN Algorithm

Hypertuning Model Parameters

- Process of finding optimal parameters for the model used; in order to achieve better accuracy
- Optimal value for number of nearest neighbours of the unclassified point taken into consideration, is found out
- Training the model multiple times on a specified range of parameters and then figure out the optimal value which gives the best accuracy

Result And Observation

- Hypertuning model parameters

Parameter	Accuracy (%)	Parameter	Accuracy (%)
1	63.00	6	61.50
2	61.10	7	61.40
3	61.99	8	62.70
4	61.80	9	61.50
5	**63.70**	10	62.80

Conclusions

- We have successfully used various tools of Signals and Systems for feature extraction which were the basis of machine learning model training.
- We were successful in training the kNN machine learning model and have achieved a decent accuracy.

Limitations

- Limitations of our project include less optimization and high running time for feature extraction.
- Future works can be done on making faster algorithms and making better classification models to achieve better accuracy.
- One may try various other approaches like CNN on spectrogram, etc

References

- Nave, G., Minxha, J., Greenberg, D. M., Kosinski, M., Stillwell, D., & Rentfrow, J. (2018). Musical Preferences Predict Personality: Evidence From Active Listening and Facebook Likes. Psychological Science, 29(7), 1145–1158. https://doi.org/10.1177/0956797618761659
- G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," in *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293-302, July 2002. DOI: 10.1109/TSA.2002.800560
- Parul Pandey, “A Guide to analyzing Audio/Music signals in Python”. Accessed 15 Nov. 2019 https://towardsdatascience.com/music-genre-classification-with-python-c714d032f0d8
- Marsyas, GTZAN Genre Collection Dataset, http://marsyas.info/downloads/datasets.html. Accessed 14th November 2019.
- Archit Rathore, Margaux Dorido, Project Report, Accessed 19 November 2019, https://www.cse.iitk.ac.in/users/cs365/2015/_submissions/archit/report.pdf

Thank You !

Any questions?

If you want to find the secrets of the universe, think in terms of energy, frequency and vibration.

Nikola Tesla