MUSIC MOOD RECOGNITION

Kunika Saraswat[1] Kavish Baranwal[2] Abhay Sharma[3] Gaurav Agarwal[4]

[1]
Student

Department of Computer science and engineering

ABES Institute of Technology, Ghaziabad, Uttar Pradesh

kunika.saraswat@gmail.com

[2]
Student

Department of Computer science and engineering

ABES Institute of Technology, Ghaziabad, Uttar Pradesh

kavi2101sh@gmail.com

[3]
Student

Department of Computer science and engineering

ABES Institute of Technology, Ghaziabad, Uttar Pradesh

abhaysharmagzb@gmail.com

[4]
Assistant Professor

Department of Computer science and engineering

ABES Institute of Technology, Ghaziabad, Uttar Pradesh

gaurav.agarwal@abesit.in

ABSTRACT

Music and human moods carry a psychological relationship. It is often seen that there is a relation in human's mood and the type of song one plays. Music is the art of "language of emotions". The central question that motivates this research paper: How the mood of a listener can be predicted using music? The project splits into two parts: extracting data from the music played determining its genre and using Artificial Intelligence (AI) and Machine Learning (ML) to determine emotion of the user. This paper surveys around detection of the mood of music on the basis of genres taking into account various features of music; harmony, timbre, lyrics, affecting the human mood profoundly. The music played is either stored in the system or is played in real-time. The type of song that the user plays determines the mood of the user and likewise creates a playlist of the same genre. Using the detected mood of the listener, final playlist is sorted out. Extraction and comparison of different features using different implementation.

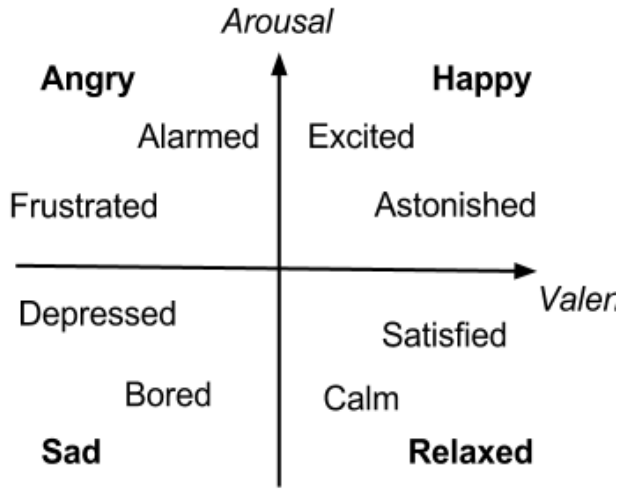
INTRODUCTION

There has been a psychological relationship between music and mood of the listener. Even, study of the same topic proves the same. Well said by Leo Tolstoy, "Music is the shorthand of emotion". There are many other sayings that finely describe the connection between music, human moods, physiological and psychological functions [2]. Five major elements that are used for musical composition are; rhythm, melody, pitch, harmony, and interval. There is a slight difference between emotion and music; therefore they can be used interchangeably in our project.

Mood recognition is an artificial intelligence application, to explain a music piece with a set of moods. This project is also related to Machine Learning. In this project the following moods are included: - Happy, Sad, Angry, Relax. Often times, we play songs to reflect the mood that we are in. However, it can be difficult to identify the songs with the right emotion to play in a particular situation given a large song library. We propose to create an application that will classify songs based on its emotion content. A person who is angry is more likely to listen heavy metal songs than a

person in happy mood who wants to listen light songs.

The project is divided into two parts: Music Analysis and Recognition of the mood of the user. The project is focused on analyzing the genre of the music that leads to knowing the mood of the listener. Music analysis is done by reading mp3 metadata of the data file or when the music is played in real-time. Recognition of the mood is done through various features that are extracted from the music.



MOTIVATION

Artificial Intelligence has proved to be a groundbreaking technology, Artificial Intelligence applications have proliferated in recent years. Computer Systems that are based on an effective user interaction will play a vital role in the next generation of computer vision systems. Artificial Intelligence acquires the capacity to detect mood of a person using facial gestures and emotions.

Generally, there are large numbers of songs in people's library. On listening to different genre of music user react differently, they experience change in blood pressure, heartbeat rate etc. Researchers on the topic of mood classification of music tend to vary to the types of features that are used to conduct the experiments. These features may be extracted from different types of sources. The song played is already stored in the system or it is played in real-time. The experiment focuses on better classification of songs than by just using traditional information like a genre [4].

There is a slight difference in emotion and mood, though at times they can be used interchangeably. In both of their comparison, a mood is considered to be much more long-lasting emotion. Emotions are less specific, less intense and less likely to be triggered by a particular stimulus or event [3].

There are various possible applications based on the experiment conducted [3]:

1) Intelligent media players – Primarily, they are going to take the mood of the user and modify the existing playlist. It is portable.

2) Intelligent systems – It is cloud-based system. Used to recommend songs based on the current mood of the user.
3) Cloud-based media players- It adapts the playlist of the user based on the mood of the user at different locations (office, home, recreational places etc).
4) Intelligent media player – They can be cloud based or portable. It determines the mood of the user and could create a playlist accordingly.

LITERATURE REVIEW

Artificial Intelligence and Machine Learning have led to various researches and development in Music Mood Recognition. The process of determination of the genre of music is a cross-disciplinary endeavor spanning. It also requires an understanding of auditory perception, psychology, and music theory [5].

Mood Recognition technique is still in its early stages of development. On studying the differences between perceptions and inductions of mood, it has been demonstrated that both can be subject to social context of the listening experience (such as audience and venue) and personal motivation (i.e., music used for relaxation, stimulation etc [6]. In this paper, we will focus more on the mood expressed rather than induced by music. The recent Music Information Retrieval Evaluation Exchange (MIREX) evaluates music mood classification and categorizes songs into one of five mood clusters, shown in Table 1. [7].

Clusters	Mood Adjectives
Cluster 1	passionate, arousing, confident, boisterous, rowdy
Cluster 2	rollicking, cheerful, fun, sweet, amiable/good natured
Cluster 3	literate, poignant, wistful, bittersweet, autumnal, brooding
Cluster 4	humorous, silly, campy, quirky, whimsical, witty, wry
Cluster 5	aggressive, fiery, tense/anxious, intense, volatile, visceral

The mood annotation task is approached essentially in three different ways. The first approach is very straightforward as one can use humans to annotate tracks by using surveys, social tagging mechanisms or annotation games. The second approach is to use contextual text information, which means *e.g.*, mining web documents, analyzing lyrics and social tags. And finally the

third approach is to evaluate the content itself, thus one can use signal processing to detect and analyze audio features. These methods can also be combined in to hybrid methods, to improve overall accuracy [5].

PROPOSED WORK

Music mood recognition is usually defined as a regression problem or classification problem where one tries to relate a music piece with a set of moods. Kim, Youngmoo E., et al. states in [5] that a music piece that is used for analysis can be an entire song or a section of a song (e.g., chorus, verse), a fixed length clip (e.g., 30-second song snippet [*sic*]), or a short-term segment (e.g., 1 second)" [5].

Here, the important thing which is to be decided is what approach is to be used and which model/s is to be followed. Then they acquire the data, prepare it, extract features and usually use a supervised machine learning technique or a combination of such techniques on these features. That could eventually be finally able to recognize mood in the *music piece*. The mood is the frame of human mind that is temporary. Various Supervised Learning models are used for mood recognition-

- Support Vector Machines (SVM)
- Neural Networks
- Label propagation (Semi-supervised)
- Linear regression: Partial Least Squares
- Gaussian Mixture Models (GGM)
- Multiple Linear Regression
- Naive Bayes
- Support Vector Regression (SVR)
- k-nearest neighbor (kNN)
- In this project, mood of the song is predicted by following Thayer's two-dimensional model of Emotion recognition, using following Supervised

Learning techniques:

- k- Nearest neighbor (kNN)
- Support Vector Machine (SVM)

This project revolves around recognition of the genre of music played, eventually leading to

determination of the mood of listener. Though there are various features of music, we are using five features to detect the mood.

1. Zero Crossing rate (ZCR) – It is the rate of sign-changes along a signal. It is the rate at which the signal changes from positive to negative or back. This is a key feature which is used in both speech recognition and music information retrieval to classify percussive sounds [8].

$$ZCR = \frac{1}{2N}\sum_{n=1}^{N} |sign(x[n]) - sign(x[n-1])|$$

2. Root Mean Square (RMS) – The continuous music power that amplifier can deliver is called Root Mean Square. It is an effective value of the total waveform. It is the area under the curve.

$$x_{\text{rms}} = \sqrt{\frac{1}{n}\left(x_1^2 + x_2^2 + \cdots + x_n^2\right)}$$

3. Mel-Frequency Cepstral Coefficients (MFCC) – Mel-frequency cepstral coefficients (**MFCCs**) are coefficients that are combined together to make up an MFC. They are derived from a type of cepstral representation of the audio clip.

4. Contrast – The difference between parts of sound or different instrumental sounds. The three types of contrast are: - Rhythmic Contrast, melody contrast, harmonic contrast.

5. Tempo – It can be defined as the pace or speed at which a section of music is played. The speed of the music is measured in beats per minute (BPM).

PRACTICAL IMPLEMENTATION

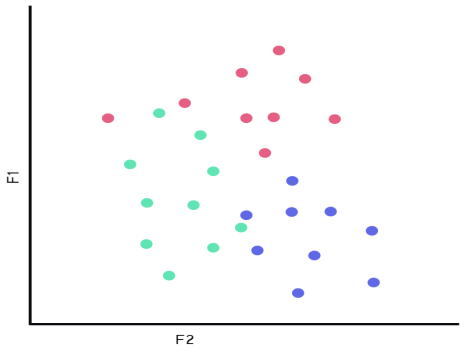
The project focuses on detection of mood of the listener by determining the genre of the music played. The genre is decided by extracting features of the music and implementing using learning techniques. The project is based upon Artificial Intelligence and Machine Learning; therefore Python 2.7 is used for implementation of the project. Python is very much familiar with Machine Learning. Various Python independent libraries for example – Librosa, Matplotlib, Scikitlearn, numpy, scipy, tkinter, pygame, csv are used.

Python comes with large standard libraries that support programming tasks such as reading and modifying files. Python's interactive mode makes it easy to test short snippets of code. A variety of different data types such as numpy array and tuples are available. For music and signal analysis, python has its separate library called Librosa. Comma Separated Value (CSV) file is used to store dataset and features that are extracted from the song.

kNN (k- nearest neighbors) Model

The *k*-nearest neighbors algorithm (*k*-NN) is a non-parametric method which is used for classification and regression. In both cases, the input consists of the *k* closest training examples in the feature space. The output depends on whether *k*-NN is used for classification or regression:

- In *k-NN classification*, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors.
- In *k-NN regression*, the output is the property value for the object. This value is the average of the values of its *k* nearest neighbors [9].

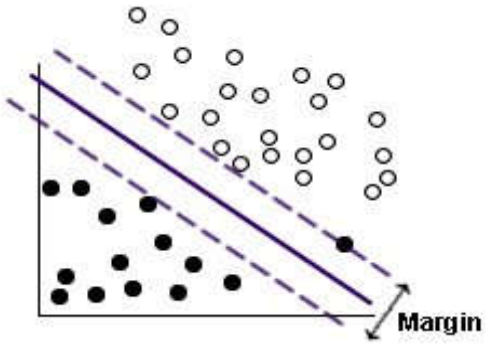


F2= instances (values in dataset)

F1= values corresponding to instances

SVM Model

Support Vector Machines are decision planes that define decision boundaries. A decision plane separates a set of objects having different class memberships. It is a classier method that performs classification tasks by constructing hyper planes in a multidimensional space that separates cases of different class labels. It is both regression and classification tasks and can handle multiple continuous and categorical variables [10].

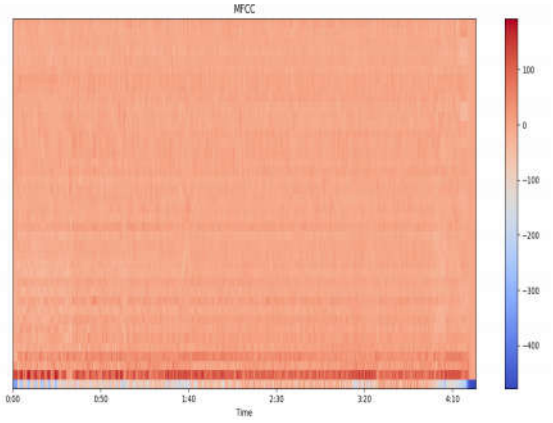


RESULT: The following is the result obtained by implementing using KNN and SVM model and predetermined expected values of some songs are following.

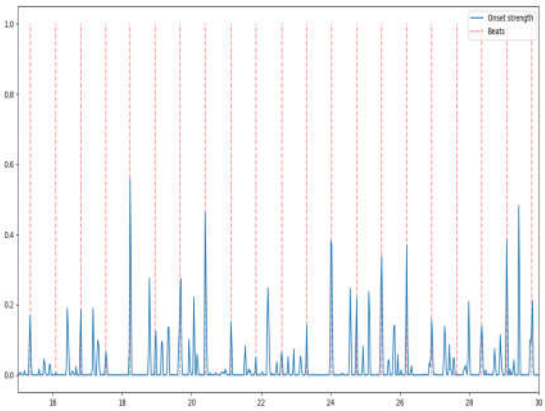
Song	KNN	SVM	Expected
Perfect- Ed Sheeran	Sad	Sad	Happy
Drag me down	Relax	Relax	Relax
Dangerous	Sad	Relax	Relax
Timbre	Angry	Relax	Angry
What makes you so Beautiful	Happy	Happy	Happy
A thousand years	Sad	Happy	Happy
Something Great	Happy	Happy	Sad
You and I	Sad	Happy	Relax
I don't wanna live forever	Sad	Relax	Happy
Love me like you do	Angry	Angry	Happy

Following are the results that are the presented by considering one song:

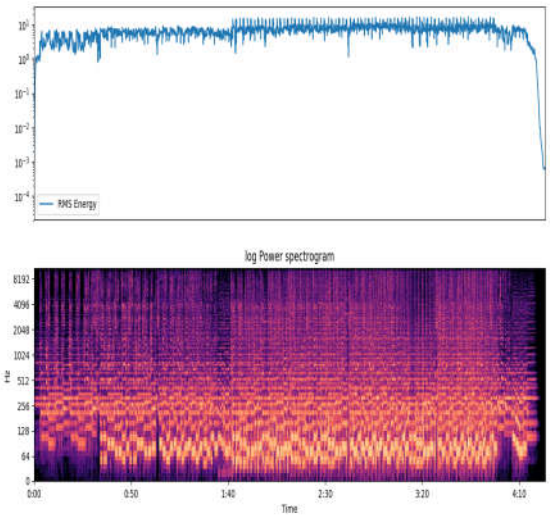
1. MFCC



2. Beats



3. RMS



CONCLUSION

We have presented an application that could detect the mood of the

listener by extracting the features of the music using Thayer's two-dimensional emotion recognition model following SVM and kNN model. Emotions are fuzzy, therefore chances of getting ambiguous results. But, on continuous working of the application, the chances of getting expected answers increases.
There is always scope for improvement in the accuracy of the mood recognition system. Within this project other modules of Artificial Intelligence and Machine Learning could also be added.

ACKNOWLEDGEMENT

First and foremost, praises and thanks to the God, for his blessings throughout our research work to complete the research successfully. We would like to express our sincere gratitude to our research supervisor, Dr. Dilkeshwar Pandey, ABESIT Institute of Technology. His sincerity, motivation, prompts response to queries and questions have deeply inspired us. It was a great privilege to work and study under his guidance. Finally, we would like to express our deep gratitude to our parents for their love, prayers, caring and sacrifices throughout our work.

REFERENCES

[1] Pratt, C. C. Music as the language of emotion. The Library of Congress, (1952)

[2] Carolyn J. Murrock,"Music and Mood," in *Psychology of Moods* 2005

[3] Vallabha Hampiholi, "A method for music classification based on perceived mood detection for Indian Bollywood music" 2012

[4] P. N. Juslin ; P. Laukka, "Communication of emotions in vocal expression

and music performance: Different channels, same code?" Psychol.

Bull.,vol. 129, no. 5, pp. 770814, 2003.

[5] Youngmoo E. Kim, Erik M. Schmidt, Raymond Migneco, Brandon G. Morton

Patrick Richardson, Jeffrey Scott, Jacquelin A. Speck, and Douglas Turnbull, "Music Emotion Recognition – A state of the art review" *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*

[6] X. Hu, J. Downie, C. Laurier, M. Bay, and A. Ehmann, "The

2007 MIREX audio mood classification task: Lessons learned," in *Proc. of the Intl. Conf. on Music Information Retrieval*, Philadelphia, PA, 2008.

[7]https://en.wikipedia.org/wiki/Zero-crossing_rate

[8]https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[9]http://www.statsoft.com/Textbook/Support-Vector-Machines