

# **Architecture for Playing Songs using Audio Content Analysis according to First Chosen Song**

Chirag Juyal  
Department of CSE  
Graphic Era University  
Dehradun, India

R.H.Goudar  
Department of CSE  
Graphic Era University  
Dehradun, India

## **ABSTRACT**

Music is one of the basic human needs for recreation and entertainment. As song files are digitalized now a days and options for random play of songs are common in use as shuffle. Shuffle randomly picks a song and has a tendency to stick around mostly played songs. Thus there is a need to retrieve and recommend songs on the basis of one's mood just by his/her first made choice.

In this paper we will present a well-defined architecture to play songs on basis of song chosen using audio content analysis and audio detector. In audio content analysis we will use features such as intensity, timbre and rhythm to map related feature music. Finally, audio detector will detect and play similar featured songs.

## **KEYWORDS**

Shuffle's alternative, automatic playing, playing on one's mood.

## **1. INTRODUCTION**

Music is an art form whose medium is sound and silence. Its common elements are pitch (which governs melody and harmony), rhythm (associated concepts tempo, meter and articulation), dynamics and the sonic qualities of timbre and texture. Listening music is one of the oldest and easiest ways to entertain oneself and change one's mood. But as the digital libraries are increasing with myriad songs coming each year, it's becoming difficult to user to keep track on each song. There are several problems faced by listeners when they have the high freedom of choice as they favor upper 20% of the items against the other 80%; this is known as 80-20 rule(Pareto Principle). Thus there is a need for a system which retrieves and recommends songs to a simple yet important category i.e. mood.

A song comprises of two important components, music and lyrics. Thus the effect of both is important. There exist many models of human emotions and their relation with music and models that categorizes music on basis of category i.e. mood. The objective of the architecture is to use best possible model to find a way to retrieve music randomly belonging to similar features as that of currently or first selected song by the user rather than shuffle.

## **2. RELATED WORK**

Many models have been proposed in this direction like song on basis of mood (category), shuffle etc. Shuffle basically randomly picks any wanted else unwanted song and generally stick to mostly played songs and allow repetition of same songs too. Whereas model of categorizing music on basis of mood needs features extraction, lyrics online mapping and more online resources in order to map music in 2D –emotional space and hence categorizing music in sad, happy, etc. types of moods. Hence model presented here does not stick to mostly played songs, does not repeat any song if once played and need not any online resource and any specific categorization to serve user on basis of his/her choice. It's a new concept of playing song on user's first chosen song and hence judging his/her mood.

## **3. PROPOSED APPROACH**

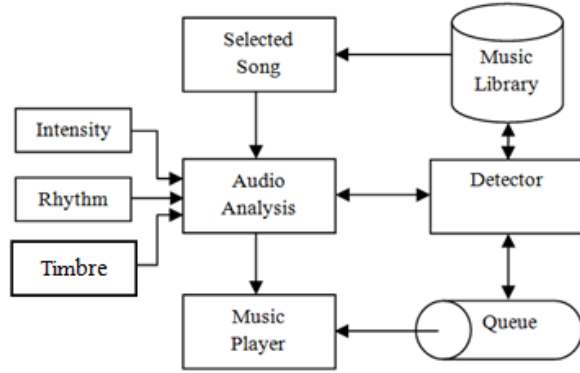
Approach is the method used or steps taken in setting about a task, problem and etc. Music gives a soul to the universe, wings to the mind, flight to the imagination, and life to everything. As music listeners do not have time to categorize music or some might not have online resources to categorize music on basis of mood. Approach proposed here in this model is such that it sticks around the similar music features of first chosen song by the user hence serve the songs according to user's mood by just judging the first selected song and need not any online source. This approach allows user to enjoy listening songs just on their choice of song, it play all songs rather than playing upper 20% newly item against the other 80%.

## **4. IMPLEMENTATION AND ARCHITECTURE**

Implementation is one of the important tasks of research; we accomplish this task by using best tools available for each problem. We have used Mat lab for extracting music features using MIR Toolbox, C-language for matching similar featured songs. Firstly song is chosen by the user and it then sent to Audio Analysis to extract features as Intensity, Timbre, and Rhythm using MAT lab Toolbox.

By the time the song gets played in the music player after it is passed by the Audio Analysis Detector which detects for the similar featured song from the music library and enqueue that song to the music player for next play.

Detector marks the currently playing or played song so that repetition can be avoided and Detector en-queue only those songs which are unmarked.



**Fig 1: Implementation of Architecture**

Music library is a collection of all songs and is a place from where Detectors detect the next song to be played.

Table-1 shown below gives a comparison between shuffle (M1) and proposed model (M2).

S – Songs Set

I – Intensity

R- Rhythm

T – Timbre

Table-1 shows the sequence of playing songs by considering both shuffle as well as our proposed model.

**Table 1: Comparison between Shuffle and Proposed Model**

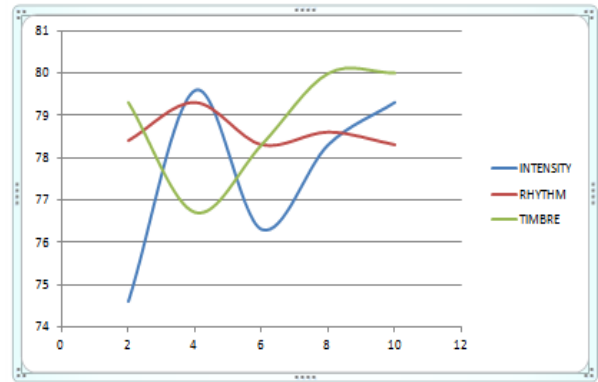
S	I	R	T	M1	M2
1	23.5	26.7	39.5	I, VI	
2	71.7	72.4	73.7	III	
3	76.6	78.3	78.3	IV	I
4	56.9	68.3	59.8	II	
5	78.9	79.3	78.7	V, X	III
6	67.7	67.8	65.8	VII	
7	78.8	73.8	75.9	VIII	II
8	80.0	79.6	78.7	IX	IV

According to below mentioned equation (1) each song will have values of every feature in between 1 to 100. As Shuffle plays the songs randomly but model proposed here will play songs of similar features (within the -3 to +3 range of first chosen song), as detector will mark the served song no

repetition is possible at all. Values given in table (see table-1) are statically assumed in order to show comparison.

We have tested the algorithm in 60% of Indian songs and reached an accuracy of 75%. The algorithm fails to depict those songs whose single feature is not in range of those ideal song features as for song 2 (refer to table-1).

Following plot (see Fig 2) depicts the Intensity, Rhythm, Timbre of ideal song (referred to table-1, ideal song is 3) with time (X-axis).



**Fig 2: Intensity, Rhythm, Timbre of ideal song**

## 5. AUDIO ANALYSIS

In this section we will discuss the algorithm to extract the music features such as intensity, timbre and rhythm. For extracting features, first all the music files are down-sampled to 16000Hz, 16 bits and mono-channel. Then a set of 10 features are extracted from music for every 10 seconds of the music file. Once all the features are extracted for a song on every 10 second basis then all features are normalized using Max-Min normalization so that each song has the uniform scale.

$$v' = \frac{v - \min(v)}{\max(v) - \min(v)} \times 100 \quad (1)$$

Thus each song will have values of every feature in between 0 to 100.

### 5.1 Feature Extraction

Feature extraction is necessary for knowing the type of music. It was indicated that mode, intensity, timbre and rhythm are of great significance in arousing different music moods (Hevner, 1935; Radocy, and Boyle, 1988; Krumhansl, 2002). We have used only three features viz intensity, timbre and rhythm.

#### 5.1.1 Intensity:

This feature has two sub-features: signal's RMS and low-energy. RMS of a signal is a more prominent feature out of the two. Low-energy of a Signal is the average of the values of the peak having values lower than a threshold which is generally very low. Thus low-energy will have values opposite to RMS i.e., for a song having high RMS, value of

low-energy would be low. Overall intensity is calculated using weighted average of the two sub-features:

$$\text{Intensity} = (\text{RMS}) \times 0.8 + (100 - \text{lowenergy}) \times 0.2$$

Low-energy is subtracted from 100 because using equation (see Eq. -1) we have scaled low-energy on a scale of 0-100. This overall Intensity is extracted for each 10 second of the song file.

### 5.1.2 Timbre:

Timbre is the quality of the music. Existing results show that the timbre of sound is determined primarily by the spectral information. We have used four features for extracting timbre:

- Zero-cross (Z)
- Centroid (C)
- Roll off (R)
- Entropy (E)

Overall timbre is calculated using weighted average of these four features:

$$\text{Timbre} = Z \times 0.2 + C \times 0.4 + R \times 0.3 + E \times 0.1 \quad (3)$$

### 5.1.3 Rhythm:

It is said that three features of rhythm are more important for music strength, tempo and regularity. The following features are extracted for calculating rhythm:

- Average Beat-spectrum (B)
- Average Tempo (T)
- Average Regularity (Ar)
- Average Positive Autocorrelation (Pa)
- Average Negative Autocorrelation (Na)

Out of the five features most important features are tempo, beat-spectrum and regularity. Overall rhythm is calculated using weighted average of these five features:

$$\text{Rhythm} = B \times 0.25 + T \times 0.45 + Ar \times 0.2 + (Pa + Na) \times 0.05$$

## 5.2 Scaling and Averaging

After calculating features for every 10 seconds of the song file, we have generated arrays of these three features. Now scaling is done to map the values of the features on a scale of (-10, 10). Before scaling we have passed each value of feature into the following equation:

$$v' = (v / \max(v)) * 100$$

By passing through this equation we have increased each value by a factor of the maximum value making the maximum value to 100%. After scaling, an average is taken of all the scaled values for each feature, the result would be the final mathematical values of the all three features of the song.

## 6. DETECTOR

This is important part of our model. As discussed earlier that audio analysis will give us the features (Intensity, Rhythm and Timbre) for the first chosen song, keeping these features as ideal for the next song that is going to be played. When the features get extracted, detector detects songs of similar features from the music library and places them in a queue. Then songs from the queue are served randomly. One more important function of detector is to mark songs those had been played and are placed in the queue in order not to create any repetition of same song again and again.

## 7 PROPOSED ALGORITHM

Following gives the algorithm to play songs on the basis of first chosen song.

Step I: Selected song is sent to Audio Analysis.

Step II: Intensity (I), Timbre (T), Rhythm (R) are extracted and song is sent to music player.

Step III: Detector fetches next song from library and matches whether that song's feature as (I', T', R') nearby equal to that (I, T, R).

Step IV: Matched songs are temporarily stored in a queue and then detector randomly serves the songs.

Step V: Served songs are then marked in order to avoid repeated serve.

Step VI: Stop.

Above given algorithm gives a concept of proposed model. User select the first song and its features are extracted by the audio analysis and the n detector fetches similar featured songs and placed them on a queue for random serve.

Below given algorithm is an implementation of proposed model.

ASSUMING RANGE OF FEATURES IS -3 AND +3.

1.  $S1 \leftarrow \text{Song1.}$
2.  $I \leftarrow \text{Intensity, } R \leftarrow \text{Rhythm, } T \leftarrow \text{Timbre}$
3. While ( $S2! = \text{Last Song}$ ) repeat Step IV-VI.
4.  $S2 \leftarrow \text{Song}$
5.  $I' \leftarrow \text{Intensity, } R' \leftarrow \text{Rhythm, } T' \leftarrow \text{Timbre}$
6. If ( $I' = I+3 \parallel I' = I-3 \ \&\& \ R' = R+3 \parallel R' = R-3 \ \&\& \ T' = T+3 \parallel T' = T-3$ ) go to Step VII.
7.  $\text{QUEUE} \leftarrow S2$
8. Stop.

## 8. CONCLUSION AND FUTURE WORK

In this paper we have discussed simple but detailed algorithm to analyze features of song. The architecture is not completely new but the way of implementation is novel. The result obtained here are more than satisfactory. The basic agenda of proposing this model is to provide music lover an easier and efficient way to enjoy listening to music. Recall could be much higher if features such as Rating, Lyrics Analysis, and Pitch etc. are taken into consideration results in efficient judging one's mood.

## 9. REFERENCES

- [1] Dan Liu, Lie Lu and Hong Jiang Zhang, Automatic Mood Detection from Acoustic Music Data, the Johns Hopkins University, ISMIR, 2003.
- [2] Yu-Hao Chen, Jin-Hau Kuo, Wei-Ta Chu, and Ja-Ling Wu, Movie Emotional Event Detection based on Music Mood and Video Tempo, IEEE, 2006.
- [3] Tao Li and Mitsunori Ogihara, Detecting Emotion in Music, the Johns Hopkins University, ISMIR, 2003.
- [4] Owen Craigie Meyers, A Mood-Based Music Classification and Exploration System, MS thesis, MIT, 2007
- [5] Sofia Gustafson-Cap ova, Emotions in Speech: Tag set and Acoustic Correlates, term paper, Stockholm University, 2001.
- [6] Campbell, J. P. (1997). Speaker recognition: a tutorial. Proceeding of the IEEE, 85 (9), 1437-1462.
- [7] Architecture for Automated Tagging and Clustering of Song files According To mood TSSN 2010.
- [8] Hevner, K. (1935). Expression in music: a discussion of experimental studies and theories. Psychological Review, 42, 186-204.
- [9] Masataka N. The origins of language and the evolution of music: A comparative perspective. Physics of Life Reviews 2009; 6:11–22.