European Journal of Computer Science and Information Technology

Vol.5, No.1, pp.11-32, February 2017

Published by European Centre for Research Training and Development UK (www.eajournals.org)

MELODY ANALYSIS FOR PREDICTION OF THE EMOTIONS CONVEYED BY SINHALA SONGS

M.G.W. Lakshitha, K.L. Jayaratne
University of Colombo School of Computing, Sri Lanka.

ABSTRACT: *This paper describes our attempt of assessing the capability of music melodies in isolation in order to classify music files into different emotional categories in the context of Sri Lankan music. In our approach, Melodies (predominant pitch sequences) are extracted from songs and the feature vectors are created from them which are ultimately subjected to supervised learning approaches with different classifier algorithms and also with classifier accuracy enhancing algorithms. The models we trained didn't perform well enough to classify songs into different emotions, but they always showed that the melody is an important factor for the classification. Further experiments with melody features along with some non-melody features showed us that those feature combinations perform much better, hence brought us to the conclusion that, even though, the melody plays a major role in differentiating the emotions into different categories, it needs the support of other features too for a proper classification.*

KEYWORDS*:* Music Information Retrival, Melody, Emotion Classification, Supervised Learning, Feature Selection, Ensemble

INTRODUCTION

Music can be categorized based on the emotions possessed by the songs, genres, cultural backgrounds, beats, rhythm and etc. These different categorizations play different roles when it comes to the application levels of them. With the recent advancements in music information retrieval field, there has been a vast interest towards identifying different emotions possessed by different pieces of music[1]. Music emotions are one of the top level music descriptors which encapsulate the semantic & intrinsic information of music files. Music can convey different emotions based on their melody (pitch), rhythm, beat, timbre etc.

Identification of the emotions and classification of music according to the emotions they convey has become a much needed application level requirements in the current field of music and other related sectors. The requirement of indexing methods for large music databases for much improved and efficient music searching and storing, can be addressed using the classification studies like classifying music into emotions. In the other hand, different music listeners may prefer different songs according to their current mood and it is a universal truth that music can feel you better and heal you. If you are in a happy mood and if you listen to the right song, it will increase your happiness. If you are nervous and need to be calm, the correct selection of the song will fix your mood. That is a kind of magic music is capable of. Even when it comes to music composing and music mixing stages in music field, music emotion classifiers (if possible and feasible) would become handy for musicians for composing songs with good standards

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

and proper qualities. They can get the assistance of such emotion classification systems in order to make good melodies while retaining the proper standards related with particular emotion categories.

There are some online music storages and websites (Allmusic.com[2], Aupeo.com[3], Last.fm[4] etc.) which provide songs categorized according to the moods/emotions they convey. In almost all of the cases, those categorizations have been done based on the user tags, but not according to the intrinsic features songs convey. It is true that, even the tags listeners put on, are based on the features the particular songs convey. But every listener may not stick the same emotion label to the same song since human perception towards the emotions are different from person to person. Therefore, their taggins/labelling may get differ according to their personal preferences and cultural, education and so many other backgrounds and the labeling a particular person does may differ time to time also.

Music melody plays a big role in presenting a required emotion through a song. If the melody is not suitable for the required emotion to be presented/conveyed by a song, the entire song will fail to deliver the meaning it supposed to deliver. Melody is a musicological concept which is based largely on the judgment of listeners, hence it has got different definitions and interpretations. In order to have a standard and a clear concept about the melody to be used in musicology studies, Music Information Retrieval (MIR) community has proposed a standard definition as "*.....the melody is the single (monophonic) pitch sequence that a listener might reproduce if asked to whistle or hum a piece of polyphonic music, and that a listener would recognize as being the 'essence' of that music when heard in comparison*"[5]. Since melody is one of the major contributors for the conveyed emotion by a song, when a song is considered, its melody should align with the emotions possessed by the lyrics of the song if the song is to be a success. A bad melody can even kill the spirit of a good set of lyrics. Therefore, composing a proper melody always needs a special consideration. If there is some sort of an automated way of assisting melody composers for letting them know about the alignment of their melodies/up to which extent their melodies agree with the emotions needed to be presented through a particular song, it would eventually have a big impact towards a sustainable future of proper music.

With that application requirement in mind, as an initial step towards that, in our study, we assess the capability of music melody in isolation to predict the emotions conveyed by songs. For that, we evaluate different machine learning algorithms in order to find the best feasible approach for our task. This paper presents our experiments with different algorithms along with their results. As the features for our classification task, we use some statistical descriptors calculated from the melodies extracted from polyphonic music files. We do experiments starting from the classifiers without any accuracy improving approaches and, then we proceed on with classifier ensemble approaches.

In this paper, our work has been categorized in following way. In Section 2, we describe some related work of music information retrieval and emotion classification together with some emotion models. Section 3 gives an overview of the methodology we utilized for our classification task which was carried out to assess the capacity of music melodies

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

to predict the emotions. Experimental results are given in Section 4, and conclusions and extensions in Section 5.

RELATED LITERATURE

A. Music Emotions

Some songs may not deliver a story or a message, but surely they will deliver an emotion or a combination of emotions[6]. But not each and every emotion is equally likely to be aroused by music. Only some mostly visible set of emotions are most likely to be triggered by music[6]. Music emotion studies attest for that and that fact has been studied by researchers as well[7].

How does music deliver emotions? Which attributes of music are most likely to have impacts on music emotions? Studies have shown that musical attributes like tempo, pitch or the melody, rhythm, timbre, harmony etc. are responsible for delivering the emotions through music files[8]. According to the variations of these attributes applications for music files, the emotion they convey differ. Number of emotion models have been proposed based on music psychology, and different emotions which are conveyed by music files have been modeled on them. Those models can basically be separated into two categories as, 1) Categorical models consist of set of discrete mood categories and 2) Scalar/Dimensional models on which emotions are positioned in a continuous multi-dimensional space[6][9]. The Hevner's taxonomy is one of the earliest music models(published in 1936) and it is still believed to be a one of best taxonomies in music psychology[6] and it is a categorical model for music emotions.

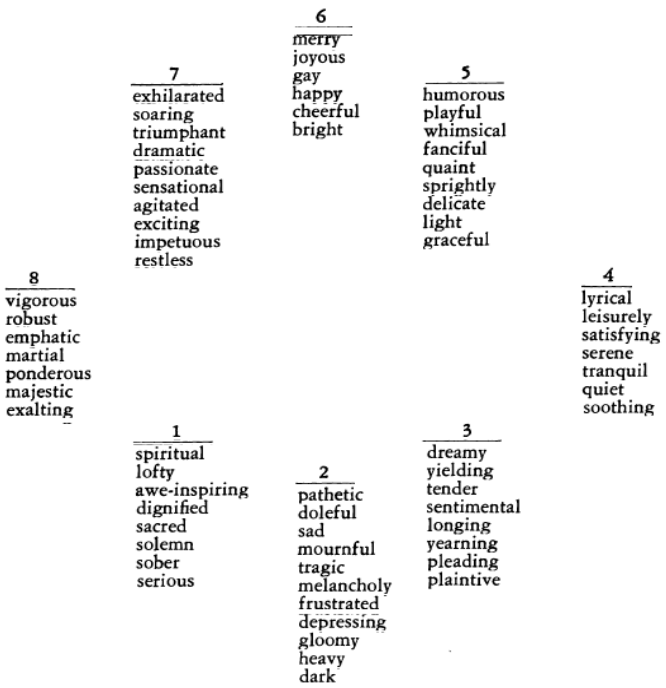

Figure 1. Hevner's emotion model for music[6]

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

Figure 1 shows the Hevner's taxonomy for music emotions. Russell's model[5] is a

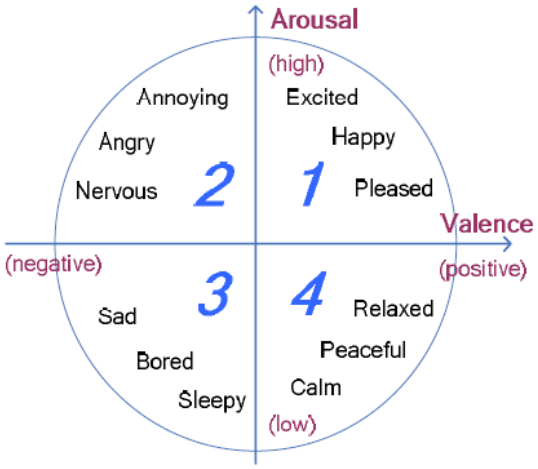

Figure 2. Thayer's emotion model for music[10]

dimensional emotion model in which the arousal and valance natures of music have been used the base of the music emotions. Thayer model[10] for music emotions is pretty much similar to the Russell's model. Both the models plot music emotions on a graph in terms of arousal (how exciting/calming) and valence (how positive /negative) phenomena. Figure **2** depicts the Thayer model for music emotions.

Emotion models which have been utilized change according to the properties of the studies. But the applicability of these models into real world scenarios is questionable since all these psychological models have been proposed in laboratory settings where social context is loosely applied[6].

B. Melody Extraction From Polyphonic Audio

Melody extraction is the process of obtaining a frequency sequence which represents the pitch of the dominant melodic line of a music file[11]. There have been different approaches proposed by researchers for this task. Goto has proposed a way of identifying a single, dominant periodicity over the audio signal which he called "Predominant F0 Estimation"[12]. In almost all of the of the research works, that F0 or the predominant pitch has been considered as the melody of the particular music file being considered[13][14]. Goto's method which is called PreFEst (Predominant F0 Estimation) is capable of detecting melody and bass lines in monaural audio signal which contains sounds from various sound sources/music instruments[12]. In Goto's method, a probability density function of the input audio signal is computed and then, the melody line is observed using that with the help of F0's probability density function. Goto's method is capable of extracting melody from polyphonic sound irrespective of the source of the sound. But in some of the research works, the melody extraction is considered as "single source predominant fundamental frequency estimation"[14][11]. For polyphonic music files, the solution they propose is to work with the music files that contain a clear lead singing voice or an instrument tone prevailing throughout the music piece being analyzed. Those approaches are quite okay for monophonic signals which contain only a one single source of sound.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

Thus, melody extracting methodologies/algorithms from polyphonic music face two major problems. The first one is identifying a set of candidate pitches which are present at a given period of time and the second one is identifying which of those identified pitches are related with the melody. Not all of those candidates do not belong to the melody of the particular segment of the music file being considered[14].Paiva[15] in his work, has proposed a way to extract the melody following the two steps mentioned above. With the intention of capturing a set of candidate pitches that constitute the melody, first he has carried out a Multi Pitch Detection (MPD) process in his algorithm. This step plays a big role in melody extraction since this is the step which extracts the pitch candidates which constitute the melody of the music file. Paiva[15] in his approach, has broken down the process in to four steps as,

- Conversion of the sound waveform into auditory nerve responses for each frequency channel, using a model of the human ear.
- Detection of the main periodicities in each frequency channel. A correlogram is produced.
- Detection of the global periodicities in the sound waveform by calculation of a summary correlogram.
- Detection of the pitch candidates in each time frame by looking for the most salient peaks in the correlogram. For each obtained pitch, a pitch salience is computed.

After the multi pitch detection process, the next step of paiva's approach is identifying the pitch values related with the melody by sequentially carrying out a stable pitch identifying step which indicate the presence of musical notes and a step for eliminating irrelevant musical notes. After selection of the most salient pitch values, Paiva has done a melody smoothing process in order to get a much better representation of the melody. That smoothing process is used to get rid of the spurious notes in the final melody if any[15].

Leon Fu and Xiangyang Xue[16] has adopted a bit different methodology of extracting melody when it comes to the approach of melody selection for their application requirement of query-by-humming. Their approach is to extract the feature descriptors related with the melody from frequency spectral information of the audio file. How they differ from others is, they have made use of the Modified Discrete Cosine Transform (MDCT) coefficients analysis on frequency spectral to extract pitches from songs.Matija Marolt’s approach for extracting melody is based on timbrel similarity of melodic fragments[17]. The approach is consist of the basic steps as; 1) finding predominant pitches using Goto's method[12], 2) tracking and merging continuous pitch contours for forming melody fragments, 3) finding fragments with well-defined melody using some predefined assumptions, 4) forming the final melody using the highly dominant clusters of melody contours identified in the previous steps.

In a research work done by Jinyu Han and Ching-Wei Chen[18], the probabilistic latent component analysis has been utilized in order for the melody extraction from polyphonic music. Their aim is to extract melody from the lead vocal of a song, hence they are based on the definition that melody is defined as the pitch contour of the lead vocal in a song. It is true that, for the parts of a song where the lead voice is present, the accompaniment

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

(instrument play) is also present. In order to remove that portion of instrument play, they have trained a probabilistic model using the accompaniment portion of the song where the voice is absent. But that is solely based on the assumption that the sound produced by the accompaniment is similar during both non-vocal and vocal parts of the song. That assumption is not that reasonable always for every type of music, hence it can be seen as a short coming of the approach of melody extracting.

Justin Salamon and Emilila Gomez has presented a novel system for the task of automated melody extraction from polyphonic music recordings[5]. Their approach is based on the creation and characterization of pitch contours which can be also recognized as time continuous sequences of pitch candidates grouped using auditory streaming cues. By defining a set of identified contour characteristics, they have defined set of rules the filter the melodic contours from non-melodic contours of the obtained whole pitch contour. They extract the melody using the voicing (voicing here refers to the instrumentation and vertical spacing and ordering of the pitches in a chord, not the instances where the human voice is not present) part of music files hence, this can be eventually used for voicing detection problems. This is a quite new study carried out and the good news is their methodology has outperformed the current state-of-the-art melody extraction systems in terms of the accuracy they come up with.

Just as the most of the melody extraction studies did, this study also based on the fact that the melody of a music file can be obtained by extracting the fundamental frequency (f0) values. The new method too, still based on the salience based melody extraction methods, but it is a novel approach of it. Justin and Emilia focuses on the pitch contour which can be defined as time continuous sequences of F0 candidates generated and grouped using heuristics based on auditory streaming ques such as harmonicity, pitch continuity, exclusive allocation. In their approach, given the audio signal, they compute the salience function describing pitch salience over time and then, peaks of the function are grouped over time using auditory streaming ques into pitch contours. By studying the features related with melody, they have defined a set of rules to filter out the pitch contours which are not related with the main melody line of the considered music file[5] whereas salience based methods directly picks the peaks of the salience function generated from the music file as the F0 candidates. From the remaining pitch contours after the filtering out is done, the melody is observed.

C. Music Emotion Classification

Music emotion classification studies based on their melodies are seem not to be abound. There are good amount of emotion classification researches have been conducted, but it is very rare(almost impossible) to see a study solely based on the melodic features of music. In fact, melodic features have been taken into the consideration, but, as a combination with other features such as timbre, rhythm etc.Using melodic features in isolation, a study has been carried out by Bruno Rocha, Renato Panda and Rui Pedro in order to do an emotion classification. The melody extraction process for the study has followed the approach proposed by Salamon et al. in 2012[5]. A data set of 903 30 second audio excerpts which are organized in 5 relatively balanced clusters (170, 164, 215, 191, 163 excerpts, respectively) have been utilized in this emotion classification study. From the extracted melody contours, a set of melodic features has been computed which are fallen under the categories of pitch and duration, Vibrato and contour typology

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

and in addition, the melody high and lowest pitches, the range between them, the ratio of contours with vibrato to all contours in the melody have been considered[19][20]. Other than the 51 features obtained through above steps, some more additional features have been obtained considering longer contours of the melody resulting 98 total amount of features. Classification task has been carried out using several supervised learning algorithms such as Support Vector Machines (SMO and LibSVM), K-Nearest Neighbors, C4.5, Bayes Network, Naïve Bayes, and Simple Logistic[19]. The experimental results have shown that the best results can be achieved using SVM classifiers. The authors have achieved 59.1% of accuracy using melodic features in isolation. But however, the best performance has been achieved by using melodic features together with standard audio features and accuracy reported is 64%.

Yi-Hsuan Yang et al.[21] have done a music emotion classification using a regression approach. They have used Thayer's arousal-valence emotion model for their study. They have first calculated the arousal and valence values (AV values) of each of the music samples provided and then those have been seen as a point in the emotion plane/model. According to the pointed plotted in the plane, a particular song can be labeled with the relevant emotion obtained from the Thayer's emotion plane. The primary focus of the authors of this study is to develop an effective method for computing the AV (arousal and valence) values. They have used a regression approach for that which can predict the AV values directly by training the implemented regression analysis models. Three types of regression algorithms have been adopted and compared in the study by Yi-Hsuan Yang et al. and they are 1) The multiple linear regression (MLR), 2) Support vector regression(SVR) and 3) AdaBoost.RT[21]. In order to train the regressors, ground truth values obtained by using a subjective test have been used. The support vector regression method has given the highest accuracy; 84% for arousal and 68% for valence[21].

A fuzzy approach has been utilized in order for the same requirement by the same set of researchers[22]. In this study too, they have adopted Thayer's model. The proposed system consists of two parts: "model generator" which generates a model according to the features of the training set and the "emotion classifier" which applies the resulting model of the "model generator" step in order to classify music files. The authors have separated a set of 195 songs into 4 different classes using a subjective study after emotion week songs are avoided. Then those labeled music files have been used for the feature extraction. Using the model observed in model generating phase, the relevant fuzzy vector for a given input is calculated. The maximum element in the vector is chosen as the final decision of the classification. It has been emphasized that, by using feature selection, the accuracy of the classification can be improved. 78.33% accuracy has been observed by using the Fuzzy Nearest-Mean classifier and Results have shown that the Fuzzy Nearest-Mean classifier outperforms the Fuzzy k-NN classifier[22].

For the emotion classification task, Naresh and Frank have utilized neural networks[23]. They have tried a Static feed forward Neural Network with supervised learning and back propagation in first attempt and then, an Elman neural network in second attempt. They use neural networks for predicting the arousal and valance models. In the first attempt, they have adopted a feed forward neural network with 13 input nodes, a hidden layer with 13 nodes and the output layer with 2 nodes for arousal and valence respectively.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

The neural network has done a good job predicting arousal and valence values with the error rate of only 14.3% indicating that the network accuracy was 85.7%.

In the second attempt, they have used a similar Elman network for arousal and valance value prediction. Note that the goal of this step had been to highlight the role of musical context during listeners' appraisal of emotional content within music. The results have shown that, the Elman network was more successful in capturing the trend of participant appraisals for arousal rather than valence[23].

METHODOLOGY

In order to assess the capability of the music melodies to identify the emotions possessed by the songs, we utilize machine learning approaches. The entire system is based on the design as given in the Figure **3**. The songs in our data set go through a pre-processing process and then the melodies are extract from the songs. Then some features representing the melody are calculated from the extracted melody and they are used to train a classifier. Following sub sections describes those steps in detail.

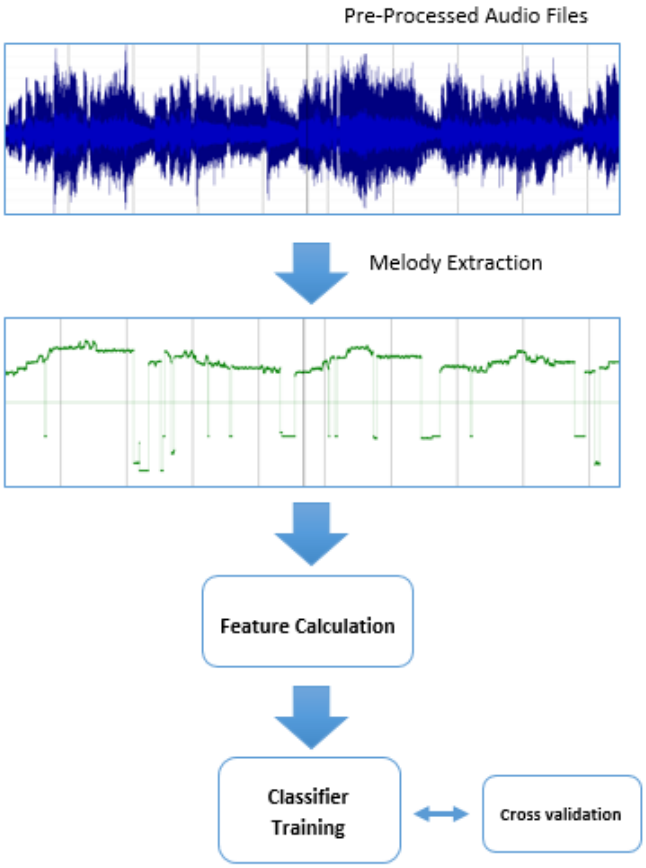

Figure 3. The steps of the emotion classification process

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

Data set and Pre-processing

Data set of a study plays a major role, regardless of the field or the scope of the study. A good data set will result in a good system, hence it will produce high accuracies. There are some general issues to be considered about and addressed to setup a good data set. In our music emotion classification system, in order to have a good classification outcome, an adequate amount of music files for each category has to be found and the classifier needs to be trained using them. The higher the good training data set, higher the accuracy of the classification it produces. In the other hand, the training set has to be labeled and it should be the ground truth of music emotions in our case. Since, the emotion is a subjective factor, a universal ground-truth for music emotion does not exist and getting reliable labels for the data is often a serious practical problem that researchers have to deal with.

In this study, the focus is about Sri Lankan music. Since there is no research work has been conducted for emotion classification of Sri Lankan music previously, no data sets are available. Since we do not have (in Sri Lankan context) any music storages in which the music files have already been labeled, we generated the data set from the scratch and labeled them. For the data set generation and the labeling, we got the assistance of music experts since they are the people who can witness the selection criteria of the songs into different categories since they are aware of how melodies are composed, on which bases particular melodies have been built upon and the other related factors.

When it comes to Sri Lankan music, some people tend to not to follow the standards or the proper qualities of music due to different reasons. They tend to deviate from the standards in large scale when composing melodies, when using the background instruments and the tones etc. Therefore, we have to agree with the fact that generating a labeled data set for a study related with Sri Lankan music needs some special expert touch and guidance.

Under the guidance of some music experts, we generated our data set which consisted of 158 songs representing 5 emotion categories which are most likely to be visible in Sri Lankan Sinhala music. The emotion categories we considered and the number of songs we had for each categories are,

- Happy - 33 songs
- Excited - 33 songs
- Sad - 33 songs
- Calm/Peaceful - 30 songs
- Heroic - 29 songs

We got the acceptance for each and every song to be included in above categories from skillful and reputed musicians. In our data set, each and every music piece was 90 seconds long starting from the chorus section. In other emotion classification studies, the authors have used much shorter excerpts[19][22][21]. The reason for using much longer excerpts is the difference of the structure of Sinhala songs. Starting from the chorus, up until the first verse is finished, the unique parts can be visible in Sinhala music. After that, the same things are getting repeated in almost all the songs. Therefore,

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

in order to not miss any unique parts and not to consider repetitive, we observed that, the 90 seconds is the most suitable length.

In order to make our data set uniform across all the songs we used, we converted them into monophonic WAV files which are 44100Hz in sampling rate and 16bits in bit depth. Stereophonic sounds have two channels to which the audio features are separated for conveying the output. If we consider only a one channel out of those two, we are not capturing all the audio features. Therefore in order to not miss any of the audio features, we converted all the stereophonic signals into monophonic signals. In that process, the two channels are fused together into a one single channel which convey all the audio features.

WAV and MP3 are the mostly visible audio formats today. Since we had songs from both WAVE and MP3 formats, in order to have the uniformity across all the songs, we converted all the MP3 files into WAV files. Other than that the other factors we considered were the audio sampling rate and the bit depth. We converted all the songs into Compact Disk audio quality in terms of bit depth and sampling rate. (44100Hz sampling rate and 16bit bit depth.)

Melody Estimation

Before extracting/calculating the features, we extracted the melody line of the music files we used. For that, we used the melody extraction methodology proposed by Justin Salamon and Emilia Gomez[5] which extracts the melody of a polyphonic music file irrespective of the source of the predominant frequency at a given time and it can be considered as the current state of the art methodology for melody extraction. The sequence of activities this method follows to extract the melody line has been described in the subsection B of this paper.

Feature Calculation

From the melodies estimated, some descriptors were calculated representing the melody line. Those descriptors which have been described below, were calculated from the output generated by the sonic-annotator[24] (a batch tool for feature extraction and annotation of audio files) using the vamp plugin – Melodia[25] which has been developed for the Justin Salamon's and Emilia Gomez's algorithm[5].

Each and every excerpt/music file/song has a melody line, which consist of a series of melody pitch contours. For each contour extracted from the above mentioned method, we retained following features which can be named as pitch and duration features.

- The starting and the ending pitch values.
- The highest and the lowest pitch values of the contour.
- The mean pitch value of the contour.
- The standard deviation of the pitch values of the contour.
- The skewness of the pitch values of the contour.
- The kurtosis of the pitch values of the contour.
- Pitch range - calculated by taking the absolute difference between lowest and highest pitch values.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

For each excerpt which were considered for our study, following features/descriptors were calculated using the features calculated above for the contours. (An excerpt consist of one or more contours).

- Pitch range features - calculated considering pitch ranges of all the contours
  - Pitch range mean
  - Pitch range standard deviation
  - Pitch range skewness
  - Pitch range kurtosis
  - Global pitch range
- Highest pitch features - calculated considering highest pitch values of all the contours
  - highest pitch mean
  - highest pitch standard deviation
  - highest pitch skewness
  - highest pitch kurtosis
  - Global highest pitch
- Lowest pitch features - calculated considering lowest pitch values of all the contours
  - lowest pitch mean
  - lowest pitch standard deviation
  - lowest pitch skewness range skewness
  - lowest pitch kurtosis
  - Global lowest pitch
- Pitch mean features - calculated considering mean pitch values of all the contours
  - mean pitch mean
  - mean pitch standard deviation
  - mean pitch skewness
  - mean pitch kurtosis
- Pitch Standard Deviation features - calculated considering mean pitch values of all the contours
  - Standard Deviation pitch mean
  - Standard Deviation pitch standard deviation
  - Standard Deviation pitch skewness
  - Standard Deviation pitch kurtosis
- Pitch interval features - Absolute difference between the mean pitch height of one contour and the previous contour
  - Interval mean
  - Interval standard deviation
  - Interval skewness
  - Interval kurtosis

For each excerpt, we calculated 27 features as listed above.

For our initial experiments, we used only these 27 features which are related only with the melody (Our research question is to explore the ability of melodic features in isolation to identifying/predicting the emotion conveyed through a Sinhala song). For subsequent experiments which produced better results, we used some more audio features which are not related with the melody - Rhythm features and Timbre features. The reason for us to select rhythm and timbre as features is that, those features have a big impact towards the emotions possessed by music files.(The reasons for doing some

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

further experiments considering the rhythmic and timbrel features will be explained in the 0 subsection.)

Rhythm is the timing pattern of musical sounds and silences in between[26]. Not only the melody, rhythm also has a strong effect on emotions[6][27][28]. Rhythm is composed of attributes like beat and the tempo. Beat is considered as a fundamental rhythmic elements of music[15]. Tempo is a derived attribute from beat and it is defined as the beats per minute (bpm)[1]. In this study, we extracted the average tempo from music files representing the rhythm of them. Timbre which can also be considered as the tone color or the quality of the tone, is the characteristic of a musical tone which distinguishes different sounds from each other. Timbre of sound is primarily determined by the spectral information of the sound[10]. Representing the timbre, we extracted 20 Mel-frequency cepstral coefficients (MFCC) from the music files we considered.
It is important to note that, for our initial experiments, we did not considered rhythm and timbre since our intention is to assess the capability melody in isolation to predict the emotions.

EXPERIMENT RESULTS

D. Experiments with Melody Features in Isolation

In order to analyze the capability of music melody for identification of the emotion, we initially did experiments using only the 27 melody features we calculated. We did all the experiments in Waikato Environment for Knowledge Analysis (Weka). For the evaluation task, the 10-fold cross validation was used. We tested our feature vector which contained 27 melody features from 5 relatively balanced emotion categories (Excited, Happy, Sad, Calm & Heroic) with different classification algorithms in WEKA. However no algorithm resulted in good classification outcomes. The best accuracy we obtained was 45.57% with 72 correctly classified instances out of 158 songs

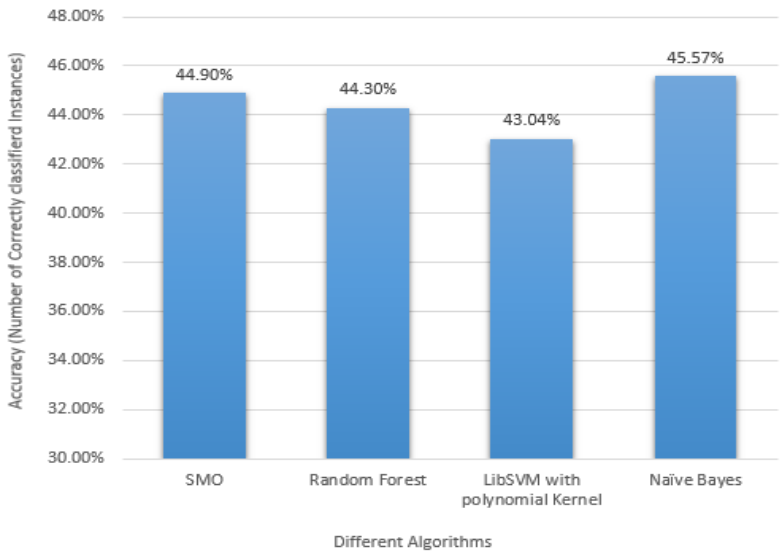

Figure 4. Accuracies from different algorithms using melody features

in our data set. That was with Naive Bayes classification algorithm. The algorithms which gave us considerable amounts of classification accuracies along with their

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

accuracies have been depicted in Figure **4**. Table 1 shows the confusion matrix which shows the relationships among actual and predicted classifications for the best performed algorithm which is Naive Bayes. Other than Naive Bayes, Sequential Minimal Optimization (SMO) with polynomial kernel, Random Forest with 700 tress in it and LibSVM with polynomial kernel machine learning algorithms gave some considerable accuracies.

In almost all the classes (other than the Heroic class), although the overall accuracy of classifying the data set into all the 5 emotion categories is low, the number of correctly classified instances are higher when compared with the misclassified instances for each of the other classes. This can be clearly seen in the Table 1. As an example, in Happy category, there are 11 instances which have been classified correctly. The misclassified instances have been separated into calm, excited, sad and heroic classes by 3, 7, 9 and 3 instances respectively. But still more than a half, has been misclassified. In this classification, the category "Excited" has shown the best classification accuracy with the true positive rate of 0.7 and Heroic category has shown the worst classification accuracy with true positive rate of 0.103.

As it can be seen in the Table 1, other than the heroic category, other emotion classes have shown some good results. In Calm category, 21 out of 30 instances, as a percentage 70% (True Positive rate is 0.7), have been correctly classified by Naïve Bayes. 5 instances have been misclassified into Sad category. The reason might be the fact that songs from Calm and Sad categories possess some similarities in some situations in their melodies in Sinhala music. Considering the results obtained, overall, the melody in isolation has shown some good accuracy in the Clam category. The important thing is that, it is noticeable that calm category has not been confused with the Excited category at all (considering the Calm category). When it comes to the Excited category too, only 5 out of 33 instances have been misclassified into Calm category. Therefore we can deduce that, melody is a good enough evidence to differentiate songs from Excited and Calm categories.

Excited category too, has shown a reasonable accuracy level. 60.6% of songs in that class have been correctly classified by the classification algorithm. As described in the previous paragraph also, 5 out of 33 songs has been misclassified into Calm category. Important thing in this category is, the classification between Excited and Sad categories. No instance has been misclassified into sad category. Therefore, we can get into a conclusion that, melody in isolation, can be used to differentiate Sinhala songs from Excited and Sad emotion categories. There is an evidence for that in the next row of the confusion matrix (Table 1). In that raw, we can see that, only one song from Sad category has been misclassified as a song which convey Excited emotion.

Heroic category has got the worst classification accuracy. Only 3 instances have been classified correctly out of 29. The songs with the heroic emotion have been confused among other songs with different emotions. The classifier has identified 9 instances which is the highest amount of instances fallen into a one category, as the songs with excited emotion. As we believe, that is because of the arousal nature possessed by the melodies of both the heroic and excited categories.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

According to the classification results we obtained, we have been unable to find a good classifier which is capable of doing a good enough emotion classification using melody features. But it is true that, melody has succeeded to differentiate songs among some of the categories; but not all.(e.g. Excited & Sad). We are looking for a good classification for all the emotion categories which can be mostly seen in Sinhala music. Therefore we carried out some further experiments.

Table 1. Individual classification into 5 emotion categories using melody features

	Calm	Happy	Excited	Sad	Heroic
Calm	**21**	2	0	5	2
Happy	3	**11**	7	9	3
Excited	5	4	**20**	0	4
Sad	5	9	1	**17**	1
Heroic	6	7	9	4	**3**

1) Experiments with Feature Selection

Since these initial experiment did not give us good enough classifications, we tried doing classifications with feature selections beforehand. We tried each and every classifier we did experiments for, in this step too. Our aim of doing experiments with a feature selection was to check whether we have considered more features than we need for a proper emotion classification. If so, after the feature selection is carried out, the classification accuracy of our emotion classification should be increased.

We used filter methods for the feature selection process. Under the filter methods, we used InfoGainAttributeEval attribute evaluator which evaluates the worth of an attribute by measuring the information gain with respect to the class[29]; together with the Ranker search method.

When we applied the InfoGainAttributeEval attribute evaluator along with Ranker search method, to obtain the set of attributes which can perform the classification task much better, it ranked following attributes with higher information gains out of the 27 input attributes.

- highest pitch mean
- mean pitch mean
- pitch range stdDev
- stdDev pitch stdDev
- lowest pitch mean
- stdDev pitch mean
- global lowest pitch

In Figure **5**, the results we obtained using the same set of machine learning algorithms we used for the previous step are shown (As a percentage of correctly classified instances).

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

It is clear that, all the accuracies have been decreased. The performances of all the classifiers are poor than they were without a features selection. Naive Bayes algorithm which gave the highest level of accuracy, has performed worst among all the algorithms tested here. Its accuracy has decreased noticeably. SMO and LibSVM has performed similarly with the selected features and those two algorithms are the best performed ones too. But still it is less than the best accuracy we obtained without any feature selection.

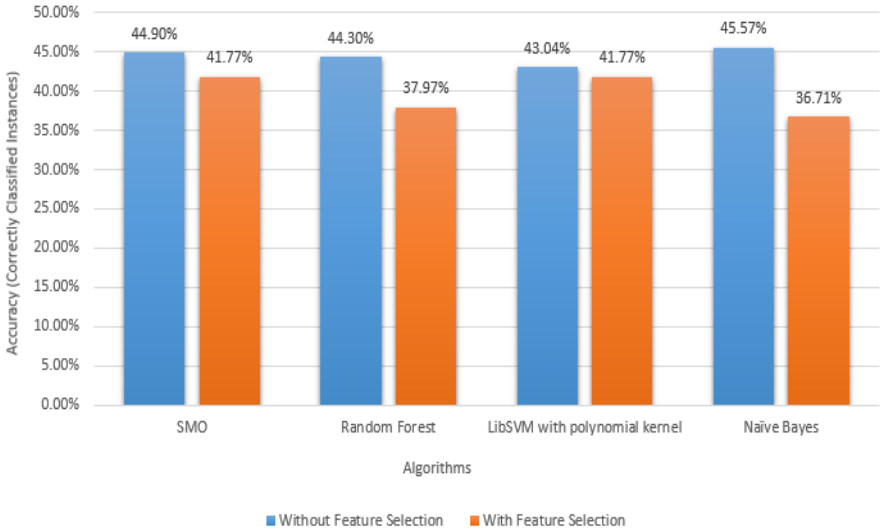

Figure 5. Difference of accuracies - before and after feature selection

Overall, feature selection using the InfoGainAttributeEval filter method, did not help us to increase the accuracy of the emotion classification. Although feature selection does decrease the time it takes to learn and complexity of the learning, it does not necessarily certify that it improves the accuracy. In our case, all the algorithms performed badly once the feature selection process is carried out beforehand.

Experiments using Ensemble Classification

Under ensemble methods, we used both Boosting and Bagging classification methods in order to improve the accuracy of emotion classification. We tested AdaBoostM1 boosting method available in WEKA along with the classifiers we tested as base classifiers for boosting. But Boosting did not improved accuracy level of any classification algorithm for emotion classification. Boosting may affect the classifiers to over learn the training data. In such cases, accuracy get decreased because of the over fitting.

Since Boosting was unable to improve the accuracy, we applied bagging ensemble methodology along with the classifiers we tested as base classifiers for boosting. Bagging improved the accuracy of LibSVM with a polynomial kernel to a level up to which any of the classifiers could not able to achieve in any experiment. Naive Bayes was the classifier which performed best up to this point with the accuracy of 45.57% for emotion classification, but LibSVM outperformed it with a classification accuracy of 46.2%. The accuracy of the Naive Bayes was reduced when it is used with Bagging.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

But 46.2% of accuracy is not a good enough classification result. Then, we used late fusion methodologies to combine different classifier outcomes in order to obtain a consensus output. We combined the two algorithms which gave us the highest accuracy levels up to now using different combination rules. The two classification methods which gave us the highest classification accuracies are LibSVM with a polynomial kernel as the base classifier in Bagging and the Naive Bayes algorithms. Both with all the 27 melody features we used.

The average of probabilities classifier combination rule along with LibSVM with a polynomial kernel as the base classifier in Bagging and the Naive Bayes gave us an accuracy level of 46.2%. That is the highest accuracy we obtained using late fusion methods for classifier combination. Like for all the other experiments, 10 fold cross validation is the methodology used for evaluation.

The Average of Probability classifier combination rule we used is an unweighed method which averages the posterior probabilities of predictions made by individual classifiers to different emotion categories. The final emotion category for a song is determined by averaging the prediction probabilities of individual classifiers for that particular song. The category which gets the highest average value wins. The confusion matrix for the above classification (LibSVM with bagging and Naive Bayes together with late fusion using Average of Probabilities) is given in the Table **2**.

As it can be clearly seen in the confusion matrix given in Table **2**, Excited & Sad categories compared to each other, have shown some promising accuracy. In Excited category, no instance has been classified as Sad and in Sad category, no instance has been classified as Excited. Even in the earlier experiments we did before this, those two categories showed better accuracy levels on each other. In this instance we can see the maximum level of accuracies those two categories resulted in. Those have been showed in color red in Table **2**.

Table 2. Confusion matrix for the highest accuracy obtained using melody features. – libsvm with a polynomial kernel as the base for bagging with naïve bayes – combined using average of probabilities rule

	Calm	Happy	Excited	Sad	Heroic
Calm	**20**	2	1	5	2
Happy	4	**9**	7	10	3
Excited	4	3	**22**	0	4
Sad	5	8	0	**17**	3
Heroic	5	6	9	4	**5**

DISCUSSION OF THE RESULTS FOR MELODY FEATURES

Doing all these experiments, our intention was to assess the ability of melodies of Sinhala songs for predicting the emotions they convey. We tried doing several

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

experiments starting from the classifications without using any ensemble approaches and then the classifications using ensemble approaches like Bagging and Boosting. Finally we combined several classifier algorithms in order to gain a better accuracy. But the highest classification accuracy we obtained was 46.2%. Therefore, the conclusion we have to attain is that when it comes to Sri Lankan Sinhala music, melody in isolation is not descriptive enough to do a successful emotion classification**.** Hence, we can infer that, melody alone cannot determine the emotion conveyed or possessed by Sinhala music. But it is clear that, melody features play a vital role and they are important to do an emotion classification because we obtained an accuracy level of 46.2% using melody features only.

In that case, we believe that, for a successful emotion classification for Sinhala music, melody needs the support of other musical features also. In other words, we believe that, melody is not capable enough to differentiate Sinhala songs among different emotions. In order to prove that presumption, we did some further experiments using melody features together with some non-melody features.

E*x*periments with Melody Features together with Non-Melody Features

As additional features, we extracted some features representing the Rhythm and the Timbre of songs. The features we extracted and how they were extracted have been described in subsection 0. Representing the rhythm of songs, we used the average tempo in beats per minutes and representing the timbre, we used 20 MFCC values as descriptors. Altogether, our feature vector contained 48 features.

- Melody - 27 features
- Rhythm - 1 feature - Tempo in beats per minute(bpm)
- Timbre - 20 features

The same dataset we used for the experiments with melody features, which contained 158 songs from 5 relatively balanced emotion categories (Excited, Happy, Sad, Clam & Heroic) was used for these experiments too. We used WEKA to do all the classifications and 10-fold cross validation as the evaluation. Figure **6** illustrates how different classifiers behave with melody features only and with all the features we considered in this step. (Melody features together with tempo representing the rhythm and MFCC representing timbre).

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

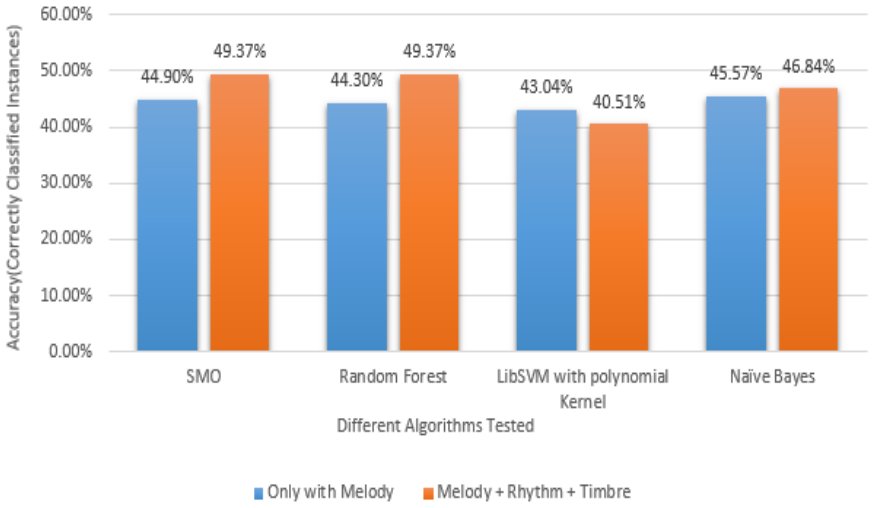

Figure 6. Comparison of the emotion classification accuracies between melody features alone and melody features together with tempo and MFCC

Other than LibSVM, all the other algorithms have performed better when it compared to the table 3classification done using only melody features. That is an evidence to prove our assumption that we had not considered adequate amount of features to carry out a good enough emotion classification in the previous step. SMO and the Random Forest are the algorithms which performed best this time with accuracy of 49.3671%. But the confusion matrices for those two classifiers are different.

We tried combining the two algorithms which give us the highest accuracy using late fusion methods. SMO and Random Forest are the best two algorithms which gave us the highest accuracies with all the features. The experiment with the classifier combination rule Maximum Probability resulted in with the accuracy of 50%. The confusion matrix for the classification is given in the Table 3.

Even in this classification, we cannot see a successful enough emotion classification. But it is clear that classification has given a better accuracy. In the confusion matrix (Table 3), we can see that Heroic emotion category has been classified better in this attempt. 13 out of 29 instances have been classifies correctly. This is a significant improvement of accuracy for that emotion category. MFCC should be the descriptive features which guided the classifier for a better classification of the heroic emotion category since when it comes to songs which convey the heroic emotion, those songs have some significant difference in tone color(the timbre) compared to other emotion categories.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

Even though we could improve the overall accuracy of our emotion classification mechanism, it is still not a good enough classification. But we can assure that by adding more features, this emotion classification can be improved. By adding non melody features into the feature set, our intention was to prove that the melody in isolation is unable to properly differentiate Sinhala songs among different emotion categories. By showing that melody features can do a better emotion classification together with rhythmic and timbre features, we could prove our assumption.

Table 3. Highest classification accuracy of the emotion classification: classifier combination with late fusion

	Calm	Happy	Excited	Sad	Heroic
Calm	**17**	1	1	7	4
Happy	6	**8**	6	9	4
Excited	4	2	**22**	1	4
Sad	5	6	1	**19**	2
Heroic	4	4	7	1	**13**

DISSCUSSION & CONCLUSION

Under ensemble methods, we used both Boosting and Bagging classification methods in order to improve the accuracy of emotion classification. We tested AdaBoostM1 boosting method available in WEKA along with the classifiers we tested as base classifiers for boosting. But Boosting did not improved accuracy level of any classification algorithm for emotion classification. Boosting may affect the classifiers to over learn the training data. In such cases, accuracy get decreased because of the over fitting.

In this paper, we tried to address a real world requirement in the field of music which is, assessing the ability of music melodies to predict the emotion conveyed by music files. If music melodies are able to do so, it would be advantageous in multiple ways. As our initial requirement asks for, an emotion classification system based on melody features (such system can be implemented, if our experiment is successful) will help musicians to predict the emotions conveyed by a newly composed melody, thus they can make sure that a proper melody is composed. In addition to that, a successful emotion classification system can be used as a proper indexing method for large music databases. Hence music searching and storing will become much easier.

For this study, we took Sri Lankan Sinhala music in to the consideration since at the present, it needs such a classification mechanism in order to avoid anomalies/conflicts which can be seen in current Sri Lankan music. We evaluated the ability of the melodies of Sinhala songs in terms of determining the emotion conveyed by them. We tried different classification algorithms together with different classifier combination methodologies. Doing all those experiments we tried to achieve the best feasible accuracy in order to draw a conclusion about whether the melodies in isolation are able to differentiate different emotions possessed by Sinhala songs.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

But even the best accuracy we obtained was not an adequate amount of accuracy which can act as an evidence for a good enough emotion classification system. Therefore the deduction we have to draw is, melody, in isolation is not able to differentiate different emotions conveyed by songs. We proved that hypothesis by doing some further experiments using some other features along with the melody features. We used Rhythm and Timbre features of songs together with the melody features we used for our initial experiments and carried out some classifications. Those classifications resulted in with higher accuracies than the accuracies obtained from the classifications conducted without the additional features. Therefore, our final deduction is, melody in isolation cannot differentiate among different emotions conveyed/possessed by Sinhala songs successfully; but together with other musical features, melody can do a such emotion classification. From the classification experiments, we noted that, even though the melody cannot do a proper emotion classification alone, melody is a very important factor for the intended task.

FUTURE RESEARCH

The study we explained throughout this dissertation, has shown some promising and interesting aspects of emotion classification. Many of them can be extended and improved as future works of this study.

For our dataset generation, we got the help and guidance from music experts. We propose that, as a future work, the dataset generation should be carried out using a subjective test. It will address the subjective nature of the music emotions more properly. Since Sri Lankan music is different from the music backgrounds which have been studied for in other music emotion classification related studies, the above proposed method will be a novel approach and it will surely produce some other emotion categories other than the ones we considered in this study.For labeling the training dataset, as a future work, Thayer's arousal-valance emotion model [10] can be utilized and then, then a regression approach can be used for the classification task.

Expanding the dataset which is used for the classifier training, is one other dimension of future works. The emotion categories should contain more music excerpts, thus a more representative learning for the emotion categories can be obtained by classifiers.

Here in this study, our focus was towards only the melody features. Other musical features also should be considered for a study which addresses the requirement of an emotion classification for Sri Lankan music (In fact that is our finding). We extracted the tempo representing the rhythm features and MFCC representing the timbre features. Some more features can be extracted representing those descriptors of music _les. Other than them, other low level music features which can be considered as candidates for determining the emotions of songs, also should be taken into the consideration in future studies. In that case, some special focus should be pointed towards the Heroic emotion category. Heroic is the emotion category which had the lowest classification accuracy of the classifications we did in this study. The songs from that category got confused with other categories in big manner. Therefore, by the time of the feature extraction process is carried out for a future study, some features which can differentiate the

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

attributes of Heroic emotion category from other categories should be identified and extracted.

REFERENCES

[1] B. Han, S. Ho, R. Dannenberg, and E. Hwang, "Smers: Music emotion recognition using support vector regression," 2009.
[2] "AllMusic | Record Reviews, Streaming Songs, Genres & Bands." [Online]. Available: http://www.allmusic.com/.
[3] "Aupeo." [Online]. Available: http://www.aupeo.com/.
[4] "Last.fm - Listen to free music and watch videos with the largest music catalogue online." [Online]. Available: http://www.last.fm/.
[5] J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *Audio, Speech, Lang. ...*, 2012.
[6] X. Hu, "Music and mood: Where theory and reality meet," 2010.
[7] M. Schoen and E. Gatewood, "The mood effects of music," *Eff. Music*, 1927.
[8] P. N. Juslin and J. A. Sloboda, *Music and emotion: Theory and research. Series in affective science.* .
[9] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.
[10] T. Zhang and C. Kuo, "Hierarchical system for content-based audio classification and retrieval," *Photonics East (ISAM, VVDC, ...*, 1998.
[11] J. Salamon, E. Gomez, D. P. W. Ellis, and G. Richard, "Melody Extraction from Polyphonic Music Signals: Approaches, applications, and challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 2, pp. 118–134, Mar. 2014.
[12] M. Goto, "A predominant-F 0 estimation method for CD recordings: MAP estimation using EM algorithm for adaptive tone models," *Acoust. Speech, Signal Process. 2001. ...*, 2001.
[13] C. Isikhan and G. Ozcan, "A survey of melody extraction techniques for music information retrieval," *Proc. 4th Conf. ...*, 2008.
[14] G. Poliner and D. Ellis, "Melody transcription from music audio: Approaches and evaluation," *Audio, Speech, ...*, 2007.
[15] R. Paiva, "An algorithm for melody detection in polyphonic recordings," *Proc. 2nd Music Inf. Retr. ...*, 2005.
[16] L. Fu and X. Xues, "A New Spectral-Based Approach to Query-by-Humming for MP3 Songs Database.," *WEC*, 2005.
[17] M. Marolt, "Audio melody extraction based on timbral similarity," *Proc. 2nd Music Inf. Retr. ...*, 2005.
[18] J. Han and C. Chen, "Improving melody extraction using probabilistic latent component analysis," ... *Signal Process. (ICASSP), 2011 IEEE ...*, 2011.
[19] B. Rocha, R. Panda, and R. Paiva, "Music Emotion Recognition: The Importance of Melodic Features," *5th Int. Work. ...*, 2013.
[20] J. Salamon, B. Rocha, and E. Gómez, "Musical genre classification using melody features extracted from polyphonic music signals," *Acoust. Speech Signal ...*, 2012.
[21] Y. Yang and Y. Lin, "Music emotion classification: A regression approach," *Multimed. Expo, 2007 ...*, 2007.
[22] Y. Yang, C. Liu, and H. Chen, "Music emotion classification: a fuzzy approach," *Proc. 14th Annu. ACM ...*, 2006.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

[23] N. Vempala and F. Russo, "Predicting emotion from music audio features using neural networks," ... *9th Int. Symp.* ..., 2012.
[24] "Sonic Annotator." [Online]. Available: http://vamp-plugins.org/sonic-annotator/. .
[25] "Melody Extraction." [Online]. Available: http://www.justinsalamon.com/melody-extraction.html.
[26] D. Chathuranga and L. Jayaratne, "Musical genre classification using ensemble of classifiers," ... *Intell. Model.* ..., 2012.
[27] F. Kuo, M. Chiang, M. Shan, and S. Lee, "Emotion-based music recommendation by association discovery from film music," *Proc. 13th Annu.* ..., 2005.
[28] L. Lu, D. Liu, and H. Zhang, "Automatic mood detection and tracking of music audio signals," *Audio, Speech, Lang.* ..., 2006.
[29] "InfoGainAttributeEval." [Online]. Available: http://weka.sourceforge.net/doc.dev/weka/attributeSelection/InfoGainAttributeEval.html. [Accessed: 06-Dec-2015].

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)